

# Računalno otkrivanje stilskih obilježja slikarskih djela utemeljeno na analizi značajki slike više razine

---

Cetinić, Eva

Doctoral thesis / Disertacija

2019

*Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj:* **University of Zagreb, Faculty of Electrical Engineering and Computing / Sveučilište u Zagrebu, Fakultet elektrotehnike i računarstva**

*Permanent link / Trajna poveznica:* <https://urn.nsk.hr/urn:nbn:hr:168:374132>

*Rights / Prava:* [In copyright](#)/[Zaštićeno autorskim pravom.](#)

*Download date / Datum preuzimanja:* **2024-07-25**



*Repository / Repozitorij:*

[FER Repository - University of Zagreb Faculty of Electrical Engineering and Computing repozitory](#)





Sveučilište u Zagrebu  
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

Eva Cetinić

**RAČUNALNO OTKRIVANJE STILSKIH  
OBILJEŽJA SLIKARSKIH DJELA  
UTEMELJENO NA ANALIZI ZNAČAJKI  
SLIKE VIŠE RAZINE**

DOKTORSKI RAD

Zagreb, 2019.





Sveučilište u Zagrebu  
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

Eva Cetinić

**RAČUNALNO OTKRIVANJE STILSKIH  
OBILJEŽJA SLIKARSKIH DJELA  
UTEMELJENO NA ANALIZI ZNAČAJKI  
SLIKE VIŠE RAZINE**

DOKTORSKI RAD

Mentor: Prof. dr. sc. Sonja Grgić

Zagreb, 2019.



University of Zagreb

FACULTY OF ELECTRICAL ENGINEERING AND COMPUTING

Eva Cetinić

**COMPUTATIONAL DETECTION OF  
STYLISTIC PROPERTIES OF PAINTINGS  
BASED ON HIGH-LEVEL IMAGE  
FEATURE ANALYSIS**

DOCTORAL THESIS

Supervisor: Professor Sonja Grgić, PhD

Zagreb, 2019

Doktorski rad izrađen je na Sveučilištu u Zagrebu Fakultetu elektrotehnike i računarstva, na Zavodu za radiokomunikacije, u suradnji s Centrom za informatiku i računarstvo Instituta Ruđer Bošković.

Mentor: prof. dr. sc. Sonja Grgić

Doktorski rad ima: 115 stranica

Doktorski rad br.: \_\_\_\_\_

## **O mentoru**

**Sonja Grgić** diplomirala je 1989., magistrirala 1992. i doktorirala 1996. godine na Fakultetu elektrotehnike i računarstva Sveučilišta u Zagrebu. Od 1989. godine radi na Zavodu za radiokomunikacije Fakulteta elektrotehnike i računarstva u Zagrebu, a trenutno je u zvanju redovite profesorice u trajnom zvanju. Njezino područje istraživanja uključuje postupke za obradu i analizu slike i videosignala te postupke za ocjenu kvalitete slike. Sudjelovala je kao istraživačica na više od deset domaćih i međunarodnih znanstvenih projekata. Bila je voditeljica projekta "Upravljanje kvalitetom slike u radiodifuziji digitalnog videosignala" koji je bio financiran od strane Ministarstva znanosti, obrazovanja i sporta Republike Hrvatske. Autorica je ili koautorica više od 190 radova objavljenih u znanstvenim časopisima i zbornicima radova s međunarodnih znanstvenih skupova. Aktivna je članica strukovnih udruga: IEEE, SMPTE, EURASIP, KoREMA i Elmar. Od 1998. godine članica je Akademije tehničkih znanosti Hrvatske (HATZ).

## **About the Supervisor**

**Sonja Grgic** received the B.Sc., M.Sc. and Ph.D. degrees in electrical engineering from University of Zagreb, Faculty of Electrical Engineering and Computing, Zagreb, Croatia, in 1989, 1992 and 1996, respectively. Since 1989, she has been with the Department of Wireless Communications, Faculty of Electrical Engineering and Computing, University of Zagreb. Currently she is a professor at the same Department. Her research interests include methods for image and video processing and analysis and picture quality evaluation. She has participated in more than 10 domestic and international scientific projects. She was the principal investigator of the project "Picture Quality Management in Digital Video Broadcasting" financed by the Ministry of Science, Education and Sports of the Republic of Croatia. She is the author or co-author of more than 190 papers published in scientific journal and proceedings of international scientific conferences. She is active member of IEEE, SMPTE, EURASIP, KoREMA and Elmar. Since 1998 she is a member of the Croatian Academy of Engineering (HATZ).

# **Zahvala**

Hvala mojoj obitelji na ljubavi, potpori i razumijevanju.

Osobito hvala mojim kolegama i prijateljima s Ruđera, posebno Tomislavu Lipiću na pomoći i poticaju u istraživačkom radu.

Veliko hvala mojoj mentorici Sonji Grgić koja je prva prepoznala istraživački potencijal ovog područja. Također, mnogo joj hvala na pomoći, ohrabrenju i podršci tijekom svih izazova mojeg doktorskog studija.

## Sažetak

Dostupnost velikih kolekcija digitaliziranih slikarskih djela otvorila je mogućnost novih istraživačkih pristupa u analizi likovne umjetnosti, zasnovanih na razvoju i primjeni metoda računalnog vida i strojnog učenja. Cilj je istraživanja ovog doktorskog rada ostvarenje metoda za računalnu detekciju i analizu stilskih obilježja slikarskih djela. Razvoju tih metoda pristupa se prema uzoru na povijesno-umjetničku analizu djela koja obuhvaća tri razine razmatranja: kategorizaciju, formalnu analizu i doživljajnu analizu. Metode za sve razine pristupa zasnivaju se na primjeni dubokih konvolucijskih neuronskih mreža. Prva razina pristupa odgovara problematici automatske klasifikacije slika. Provedena je usporedna analiza različitih postavki učenja mreža te su postignuti trenutno najbolji rezultati klasifikacijske točnosti za većinu predstavljenih zadataka klasifikacije slikarskih djela. Druga razina pristupa ostvaruje se razvojem metode kvantifikacije zastupljenosti specifičnih stilskih obilježja i predikcije vrijednosti tih obilježja zasnovana na učenju regresijskih modela konvolucijskih neuronskih mreža. Treća razina pristupa ostvaruje se metodom kvantifikacije subjektivnih aspekata estetske, afektivne i memorijske percepcije likovnog djela. U ovome radu prvi se puta provodi usporedna analiza predikcijskih vrijednosti navedenih percepcijskih obilježja dobivenih primjenom konvolucijskih neuronskih mreža na velikom skupu slikarskih djela. Kvantitativni i kvalitativni rezultati dobiveni primjenom predstavljenih metoda druge i treće razine pristupa sukladni su s povijesno-umjetničkim saznanjima, kao i s rezultatima ispitivanja ljudskih procjena zastupljenosti određenih obilježja u slici.

**Ključne riječi:** digitalna analiza slike, klasifikacija slikarskih djela, duboko učenje, računalni vid, značajke slike više razine, konvolucijske neuronske mreže, digitalna povijest umjetnosti

# Computational detection of stylistic properties of paintings based on high-level image feature analysis

Large-scale digitization efforts which took place in the last two decades led to a significant increase of online accessible fine art collections. The availability of those collections makes it possible to easily explore and enjoy artworks which are scattered within museums and art galleries all over the world. The increased visibility of digitized artworks is particularly useful for art history education and research purposes. Apart from the advantages of the visibility boost, the very translation of information, from the domain of the physical artwork into the digital image format, plays a key role in opening new research challenges in the interdisciplinary field of computer vision, machine learning and art history.

Fine art collections are a data source of historically relevant, as well as perceptually and emotionally intriguing visual information. Because of its manifold nature, the domain of fine art images represents a fruitful data source for formulating semantically relevant image analysis tasks. The majority of available online collections include some particular metadata, usually in the form of annotations done by art experts. Those annotations mostly contain information about the artist, style, genre, technique, date and location of origin, etc. Art experts can easily identify the artist, style and genre of a painting using their experience and knowledge of specific features. However, a great current challenge is to automate this process using computer vision and machine learning techniques. Generating metadata by hand is time consuming and requires the expertise of art historians. Therefore, automated recognition of artworks' characteristics would enable not only a faster and cheaper way of generating already existing categories of metadata such as style and genre in new collections, but also open the possibility of creating new types of metadata that relate to the artwork's content or its specific stylistic properties.

Stylistic properties of paintings are abstract attributes inherent to the domain of human perception. Analysing artworks is a complex task which involves understanding the form, expression, subject matter and meaning. Motivated by the successful performance of Convolutional Neural Networks (CNN) for a wide variety of computer vision tasks, this thesis explores the use of CNNs for learning features that are relevant for understanding stylistic properties of paintings. The proposed approach addresses three levels of analysing paintings: categorization, formal analysis and perceptual analysis. The first level of analysis corresponds to the task of automated image classification. The second level of formal analysis includes training CNN regression models to predict values of features that quantify specific stylistic properties relevant for art history. The third level of perceptual analysis involves quantitative approaches to highly subjective aspects of perceiving artworks. The structure of the thesis is organized in order to present the methodologies and research outcomes of each of the three levels of analysis as three

---

separate chapters (chapters 3, 4 and 5).

The first chapter of this thesis introduces the motivation of the research, as well as gives a comprehensive overview of related studies. The chapter concludes with the thesis objectives and structure. Because the developed methods for analysing artworks on all three levels are based on convolutional neural networks, the second chapter provides a brief introduction and conceptual overview of convolutional neural networks.

The third chapter presents the proposed approach and research results of the automated fine art classification challenge. This thesis focuses on exploring the applicability of CNNs for art-related image classification tasks. Extensive CNN fine-tuning experiments are performed and results for five different art-related classification tasks on three large fine art datasets are presented. Along with addressing the previously explored tasks of artist, genre, style and time period classification, a novel task of classifying artworks based on their association with a specific national artistic context is introduced. State-of-the-art classification results of the addressed tasks are achieved, signifying the impact of the proposed method on computational analysis of art, as well as other image classification related research areas. Different CNN fine-tuning strategies are explored in order to identify the best training setup for different art-related tasks and datasets, with a particular focus on exploring the impact of domain-specific weight initialization. The effects of domain-specific weight initialization are systematically compared by evaluating networks pre-trained for different tasks, varying from object and scene recognition to sentiment and memorability prediction. The experimental outcomes show that fine-tuning networks pre-trained for scene recognition and sentiment prediction yields better results than fine-tuning networks pre-trained for object recognition. Besides weight initialization, different fine-tuning scenarios are considered in order to find the most efficient solution, as well as avoid over-fitting. These scenarios include variations of the extent to which the error from the new task is being back propagated within the network or, in other words, how many of the transferred layers are kept frozen. Based on the accuracy results, the best scenario in most cases is to re-train all except the first convolutional layer. After determining the best performing training setup for each task, a further exploration of the task-specific classification is carried out by looking into the per-class classification performance. In comparison to other tasks, style classification has the lowest overall accuracy, most probably due to the high level of visual properties overlapping between classes, as well as to the great diversity of content depicted in the same style. On the other hand, the classes of the genre classification task are more uniform in terms of content and CNNs show a high ability to distinguish scenes and objects in paintings, regardless of the various artistic techniques and styles. In addition to exploring fine-tuning settings and task-specific classification performance, the practical usability of the fine-tuned models is addressed by analysing different aspects of image similarity. The presented research outcomes show that features derived from models fine-tuned for different tasks can be employed to retrieve diffe-



---

rently similar images. In particular, CNN models fine-tuned for genre can be used to retrieve images that are more similar in terms of content, by including specific objects and similar compositions. However, CNN fine-tuned for style recognition focuses more on style properties such as brushwork or level of details. These findings suggest that the proposed approach can serve as a basis for implementing a novel framework for refined retrieval of fine art images.

The fourth chapter describes the proposed methodology and research results related to the second level of understanding artworks based on the formal analysis of style. The majority of studies concerned with computational analysis of artworks focuses on automatic classification or recognizing objects in artworks, while not many attempts have been made to analyze artworks in terms of their specific stylistic properties. Particularly, regarding concepts which play an important role within art history such as the concepts defined by Heinrich Wölfflin in his book *Principles of Art History* (1915). Discussing the historical transformation of styles, particularly from Renaissance to Baroque, Wölfflin identifies five key visual principles. Each principle is defined by two contrasting visual schemes: (1) linear and painterly, (2) closed and open form, (3) planar and recessional, (4) multiplicity and unity and (5) absolute and relative clarity. The transformation of styles from Renaissance to Baroque corresponds to the conversion of the arrangement of visual elements from the first to the second visual scheme in each principle. Although Wölfflin's comparative principles were developed based on the differences between characteristics of artworks of 16th and 17th centuries, they became a standard method of formal analysis of art and a conventional approach in understanding changes of artistic styles. The main goal of this thesis is to quantify and predict the level of presence of two contrasting visual schemes for each of the five principles in an image. Each principle is defined as one high-level image feature that measures how much each of the contrasting concepts is present in an image. Research outcomes of one previous study suggest that convolutional neural networks trained to classify paintings according to different styles, implicitly learn features related to Wölfflin concepts. Based on those findings, convolutional neural networks models were trained to predict values of each concept. Concretely, the best performing CNN model for style classification introduced in the previous chapter of this thesis is fine-tuned using an existing small-sized dataset labelled with annotations of Wölfflin's pairs' values provided by art historians. For each of the five Wölfflin's concepts, one separate model is trained to predict the value of the corresponding feature. Several different fine-tuning approaches were tested in order to identify the optimal setting for each task. The experimental results are analysed and discussed from different viewpoints. Firstly, the performance of regression models trained for each of the five different tasks on the test sets is analyzed. Furthermore, to evaluate how the models generalize to a new dataset, the models are employed on a large unlabelled dataset. An analysis of the qualitative results is provided by comparing images with the highest and lowest prediction values for each feature. The outcome of the analysis suggests that the models learn

---

to discriminate meaningful features that correspond to the visual characteristics described by Wölfflin. Finally, an analysis how the predicted feature values change over time, as well as in relation to different styles and artists is provided. The outcome of this analysis shows that the results correspond to Wölfflin’s theoretical observation on the historical transformation of artistic styles. This indicates that the presented approach can be used to enable new ways of exploring fine art collections based on image features relevant and well-known within art history.

The fifth chapter describes the methods for analyzing perceptually relevant properties of fine art images. Deep learning based quantitative approaches to highly subjective aspects of perceiving images (aesthetics, memorability, sentiment) are employed for the first time in the domain of fine art images. These three different aspects of image perception have been studied by psychologists for a long time and have recently become an emerging subject of interest within the computer vision and machine learning community. Due to the higher availability of natural image datasets, most of the studies concerned with computationally addressing perceptual image features were done on photographs, while art images have not yet been systematically explored. The appearance of big and comprehensively annotated fine art datasets facilitates the analysis of those specific visual properties on a large scale. However, collecting ground-truth labels for attributes related to subjective perception of images is laborious and expensive because it requires complex experimental surveys. On the other hand, the concept of transfer learning and the transferability of pre-trained CNN models across different domains enable new ways of feature assessment. Several CNN models trained to predict aesthetic, sentiment and memorability scores of natural images are employed in order to explore those features in art images. For each task, three different CNN models trained on natural images have been employed. In order to investigate the consistency of predictions for each task, the scores obtained by three different CNN models on the large unlabeled fine art dataset are compared. Based on the correlation between predicted scores and human judgments, the best performing CNN model is selected for further analysis. After identifying the best performing model for each task based on the correlation between the predicted scores and human rating scores, the qualitative results of the predicted aesthetic, sentiment and memorability scores on the large unlabeled fine art dataset are evaluated by visually inspecting the images with the highest and lowest prediction values. Regarding the prediction of the aesthetic evaluation, the major difference between the high and low valued images lies in color and lighting. Highly rated images tend to include bright and intense colors, while low rated images are dim and pale. Among the properties which contribute to the evaluation of the positive or negative image sentiment, the obvious visual difference is the color choice, with positive images being bright and colorful and negative images darker. Regarding content, positively categorized images most commonly depict flowers, portraits of smiling people, couples and family portraits. Negatively categorized images often depict outdoor scenes, abstract images with sharp edges and strong contrast, as well as portraits of

---

sad or fearful faces. When looking into the images with the highest and lowest memorability scores, three dominant motifs occur within the most memorable images: abstract images with dot patterns, nude paintings and portraits. The least memorable images predominantly include landscapes. To better understand how different image properties correlate with the predicted aesthetic, sentiment and memorability scores, the correlation with different image features is analyzed. Decision tree based machine learning models are employed to analyze the relative importance of various image features related to content, composition and color in determining image aesthetics, visual sentiment and memorability scores. Outcomes of this analysis suggest that content and image lighting have significant influence on aesthetics, that color vividness and harmony strongly influence sentiment prediction, while object emphasis has a high impact on memorability. Additionally, the predicted aesthetic, sentiment and memorability scores are explored in the context of art history by analyzing their distribution in regard to different artistic styles, genres, artists and centuries. An interesting outcome of this analysis is that abstract styles tend to be more memorable, but have a lower average aesthetic and positive sentiment score. Furthermore, the genre-based distribution of scores, where the subject matter plays the most important role, corresponds to previous photography-related findings and demonstrates consistency between art and natural images, as well as compatibility with intuitive human presumptions. The overall qualitative and quantitative results suggest that CNN models pre-trained on natural images can extract meaningful aesthetic, sentiment and memorability features in the domain of fine art images. However, limitations emerge based on the choice of a particular task-specific model. Although the results obtained from differently trained models are consistent for the task of memorability, the aesthetic predictions are less consistent and depend more strongly on the choice of dataset and model architecture. Nevertheless, the scores predicted using CNNs represent an interesting finding and can serve as a basis for formulating initial hypotheses regarding the psychological relation of different high-level image properties. Furthermore, the presented approach enables new ways of exploring fine art collections based on highly subjective aspects of art.

Chapter six summarizes the main scientific contributions of the thesis. The chapter concludes with a brief discussion of future research directions. Particularly, regarding the aim to intensify interdisciplinary collaboration and investigate the applicability of the presented methods to concrete art history-related research challenges.

**Keywords:** digital image analysis, fine art classification, deep learning, computer vision, high-level image features, convolutional neural networks, digital art history

# Sadržaj

<b>1. Uvod</b>	1
1.1. Pregled dosadašnjih istraživanja	3
1.2. Cilj istraživanja	7
1.3. Struktura rada	8
<b>2. Konvolucijske neuronske mreže</b>	9
2.1. Struktura konvolucijskih neuronskih mreža	9
2.1.1. Slojevi konvolucijske neuronske mreže	10
2.1.2. Arhitektura konvolucijske neuronske mreže	14
2.2. Učenje konvolucijske neuronske mreže	15
2.2.1. Algoritam propagacije pogreške unatrag	15
2.2.2. Metode optimizacije i regularizacije	17
<b>3. Računalno raspoznavanje slikarskih djela</b>	20
3.1. Pozadina i motivacija istraživanja	20
3.2. Baze slika i klasifikacijski zadaci	22
3.3. Fino podešavanje parametara konvolucijske neuronske mreže	26
3.3.1. Razina dosega finog podešavanja parametara mreže	27
3.3.2. Domenski određene inicijalizacije parametara mreže	28
3.3.3. Postavke treniranja mreže	29
3.4. Ovisnost rezultata klasifikacije o postavkama finog podešavanja parametara mreže	30
3.4.1. Utjecaj domenski određene inicijalizacije parametara mreže	30
3.4.2. Utjecaj razine dosega finog podešavanja parametara mreže	32
3.5. Rezultati automatske klasifikacije slikarskih djela	34
3.5.1. Pregled i usporedba klasifikacijskih rezultata	34
3.5.2. Klasifikacija slikarskih djela prema stilu, žanru, autoru i nacionalnosti	35
3.5.3. Primjena naučenih modela za analizu različitih aspekata vizualne sličnosti	39
3.6. Osvrt na rezultate istraživanja	42

<b>4. Računalna analiza specifičnih stilskih obilježja slikarskih djela . . . . .</b>	<b>44</b>
4.1. Pozadina i motivacija istraživanja . . . . .	44
4.2. Baze slika i regresijski zadaci . . . . .	48
4.3. Fino podaševanja parametara mreže za učenje značajki Wöllflinovih koncepata	49
4.3.1. Postavke učenja modela . . . . .	49
4.3.2. Rezultati učenja modela . . . . .	50
4.4. Predikcije stilskih obilježja u kontekstu povijesti umjetnosti . . . . .	52
4.4.1. Linearnost i slikovitost . . . . .	53
4.4.2. Zatvorena i otvorena forma . . . . .	56
4.4.3. Ploha i dubina . . . . .	57
4.4.4. Mnoštvo i jedinstvo . . . . .	59
4.4.5. Jasnost i nejasnost . . . . .	60
4.5. Osvrt na rezultate istraživanja . . . . .	62
<b>5. Računalna predikcija percepcijskih obilježja slikarskih djela . . . . .</b>	<b>64</b>
5.1. Pozadina i motivacija istraživanja . . . . .	64
5.1.1. Računalna estetika . . . . .	66
5.1.2. Računalno predviđanje vizualnog sentimenta . . . . .	66
5.1.3. Računalno predviđanje pamtljivosti slike . . . . .	66
5.2. Postavke eksperimenta . . . . .	67
5.2.1. Baze slika . . . . .	67
5.2.2. Modeli za predikciju percepcijskih obilježja . . . . .	69
5.2.3. Ekstrakcija naučenih značajki slike više razine . . . . .	71
5.3. Analiza naučenih reprezentacija slike i aktivacijskih mapa . . . . .	72
5.4. Evaluacija predikcija percepcijskih obilježja slike . . . . .	75
5.4.1. Analiza konzistentnosti predikcija dobivenih različitim modelima . . . . .	76
5.4.2. Usporedba predikcijskih i stvarnih vrijednosti . . . . .	77
5.5. Odnos značajki slike i percepcijskih obilježja slike . . . . .	79
5.5.1. Kvalitativni rezultati predikcije percepcijskih obilježja . . . . .	79
5.5.2. Odnos značajki slike više razine i percepcijskih obilježja . . . . .	81
5.5.3. Korelacija značajki boje i percepcijskih obilježja . . . . .	83
5.6. Predikcije percepcijskih obilježja u kontekstu povijesti umjetnosti . . . . .	84
5.6.1. Odnos predikcija percepcijskih obilježja i stilskih pravaca . . . . .	84
5.6.2. Odnos predikcija percepcijskih obilježja i žanrova . . . . .	84
5.6.3. Odnos predikcija percepcijskih obilježja i autora . . . . .	86
5.6.4. Odnos predikcija percepcijskih obilježja i povijesnih razdoblja . . . . .	88
5.7. Osvrt na rezultate istraživanja . . . . .	88

<b>6. Zaključak . . . . .</b>	<b>90</b>
<b>Prilozi . . . . .</b>	<b>93</b>
<b>A. Pregled klasa za sve baze i zadatke klasifikacije slikarskih djela . . . . .</b>	<b>93</b>
<b>B. Kvalitativni rezultati ekstrakcije naučenih značajki slike više razine . . . . .</b>	<b>100</b>
<b>Literatura . . . . .</b>	<b>104</b>
<b>Životopis . . . . .</b>	<b>114</b>
<b>Biography . . . . .</b>	<b>115</b>

# Poglavlje 1

## Uvod

Intenzivna digitalizacija umjetničkih djela dovela je do pojave velikog broja digitaliziranih i online dostupnih zbirki slikarskih djela. Korisnost postojanja takvih digitaliziranih zbirki mnogostruka je i obuhvaća sve prednosti svojstvene slobodnoj dostupnosti povijesno-kulturnog sadržaja. Osim samorazumljivih prednosti kao što su očuvanje kulturne baštine i unaprjeđenje obrazovnih mogućnosti, pojava digitaliziranih zbirki slikarskih djela stvorila je preduvjete za razvoj novih mogućnosti istraživanja povijesti umjetnosti i slikarstva. Postupkom digitalizacije slikarsko djelo postaje matrica brojeva. Prijenosom informacije iz domene ljudskoga djelovanja koju karakterizira kompleksnost umjetničkog izražaja u domenu računala koja funkcionira na načelu numeričkoga prikaza informacija, otvara se prostor novih istraživačkih pristupa. Objedinjeni pod pojmom „digitalne humanistike“ ti novi istraživački pristupi uključuju primjenu različitih računalnih tehnologija u obradi i analizi podataka iz područja humanističkih znanosti i umjetnosti. Budući da se digitalna humanistika kao grana znanosti nalazi tek u začetku svoga razvoja, sav potencijal i sve prednosti takvog interdisciplinarnog pristupa tek se otkrivaju. U tom kontekstu osobito izazovno mjesto pripada analizi slikovnih podataka budući da se trenutno značajan broj digitalno-humanističkih istraživanja temelji na računalnoj analizi teksta i metapodataka. Iako je povijesno-sociološki kontekst nužan za cjelovito razumijevanje umjetničkog stvaralaštva određene epohe, postavljanje slike u središte istraživačkog interesa podrazumijeva stavljanje naglaska na formalnu analizu likovnog djela.

Formalna analiza likovnog djela uključuje razlikovanje, opis i interpretaciju glavnih izražajnih sredstava, odnosno osnovnih elemenata slikarskog djela kao što su boja, kompozicija, linija, osvjetljenje i dr. Određeni suodnos osnovnih elemenata slikarskog djela tvori specifična stilsku obilježja svojstvena određenom umjetniku ili stilskome pravcu. Shvaćena u širem smislu, stilsku obilježja, osim suodnosa vizualnih elemenata, uključuju i motive i tematske preokupacije karakteristične za određeno stilsko razdoblje ili određenog autora. Stilsku obilježja shvaćena kao zbir perceptivnih i semantičkih značajki slikarskog djela čine kompleksan skup atributa čija kvantifikacija, odnosno prevodivost u domenu računalno obradivih podataka predstavlja

temeljnu motivaciju istraživanja izloženog u ovome radu.

Oblikovanje predstavljenih metoda za računalnu analizu i detekciju stilskih obilježja zasnovano je na tri različite razine pristupa likovnom djelu. Prva razina odnosi se na kategorizaciju umjetničkih djela, odnosno raspoznavanje pripadnosti djela određenom autoru, određenim žanrovskim i stilskim kategorijama te određenom povijesno-geografskom kontekstu. Priroda zadatka prve razine pristupa u kontekstu računalnog vida odgovara problematici automatske klasifikacije slika i nadziranog učenja (engl. supervised learning). Druga razina pristupa odnosi se na formalnu analizu slikarskog djela, odnosno raščlanjivanje osnovnih elemenata i principa likovnog oblikovanja. Osnovni je izazov u problemu automatizacije formalne analize modeliranje značajki slike prema uzoru na umjetničke koncepte. Oblikovanje novih značajki kao mjera zastupljenosti određenih principa likovnog oblikovanja omogućuje istraživanje odnosa i zakonitosti tih principa u velikim skupovima slikovnih podataka. Treća razina pristupa likovnom djelu odnosi se na interpretativnu i doživljajnu dimenziju analize umjetničkog djela. Automatizacija ove razine pristupa posebno je izazovna jer se temelji na kvantifikaciji subjektivnih aspekata ljudske percepcije.

Ključnu ulogu u svim razinama pristupa imaju načini reprezentacije slike, odnosno značajke slike. S obzirom na njihovu razinu složenosti i interpretabilnosti, značajke slike možemo podijeliti na tzv. značajke slike niske razine (engl. low-level image features) i značajke slike visoke razine (engl. high-level image features). Određenje niske razine odnosi se na one značajke slike koje na temelju globalnih i lokalnih varijacija u vrijednostima intenziteta piksela detektiraju informacije o boji, teksturi, rubovima i osnovnim oblicima u slici. Informacije sadržane u značajkama slike niske razine najčešće se odnose na jedan aspekt opisa osnovnih vizualnih elemenata u slici, pri čemu je priroda tih informacija bitno različita od načina na koji ljudi interpretiraju informacije u slici. Integracijom primitivnih elemenata u složenije koncepte nastaju značajke slike više razine. Značajke slike više razine su kvantifikacije koncepata koji odgovaraju načinu na koji ljudi razumiju i/ili doživljavaju sliku, tj. onih informacija u slici koje se odnose na sadržaj i percepciju slike.

S obzirom na način nastanka, značajke slike mogu biti oblikovane (engl. engineered features) ili naučene (engl. learned features). Oblikovane značajke nazivaju se još i „ručno izrađenim“ značajkama (engl. hand-crafted image features) jer su one dizajnirane na temelju ljudskog razumijevanja matičnog prikaza slike te s obzirom na specifične zadatke obrade i analize slike. Postoji veliki skup oblikovanih značajki koji se najčešće kategorizira na temelju svojstava koja se analiziraju (značajke boje, značajke teksture) ili s obzirom na područje slike nad kojim se izvodi određena operacija (lokalne i globalne značajke slike). U posljednjih nekoliko godina u različitim domenama primjene metoda računalnog vida, oblikovane značajke zamijenjene su tzv. naučenim značajkama, odnosno reprezentacijama slike dobivenim iz slojeva konvolucijske neuronske mreže. Pokazano je da se upotrebom dubokih konvolucijskih neuronskih mreža



postiću mnogo bolji rezultati u problemima automatske klasifikacije i raspoznavanja sadržaja u slici nego kada se metode strojnog učenja primjenjuju nad skupom značajki ekstrahiranih iz slike nekom od metoda dobivanja oblikovanih značajki.

Teorijske osnove dubokih neuronskih mreža daleko prethode razdoblju njihove iznimne popularnosti kojem trenutno svjedočimo. Neki od razloga zbog kojih su duboke neuronske mreže tek relativno nedavno postale iznimno učinkovite jesu jeftina i dostupna računalna snaga grafičkih procesorskih jedinica te dostupnost velike količine podataka iz kojih mreža može naučiti velik skup parametara i postići bolju generalizaciju u rješavanju problema. Akceleracija primjene dubokih neuronskih mreža u rješavanju različitih problema računalnog vida dovela je do pojave velikog broja naučenih modela koji ne predstavljaju više samo sredstvo za rješavanje određenog problema, već postaju izvor za prijenos naučenog znanja iz jedne domene u drugu. Upravo se pitanje prijenosa znanja dubokih neuronskih mreža iz jedne domene u drugu problematizira u ovome radu različitim razinama pristupa računalnoj analizi slikarskih djela.

## 1.1 Pregled dosadašnjih istraživanja

Shvaćena tematski u najširem smislu, računalna analiza slikarstva obuhvaća širok spektar istraživanja čija je zajednička odrednica upotreba metoda računalnog vida za raspoznavanje i razumijevanje digitaliziranih slika likovnih djela. Svrha i pojedinosti upotrebe tih metoda razlikuju se s obzirom na cilj i ishod istraživanja. U ovome poglavlju prikazan je pregled različitih oblika primjenjivosti računalnih metoda u analizi slikarstva. Struktura pregleda prati kronološki tijek i tematske odrednice različitih istraživanja, pri čemu se osnovna podjela ovoga pregleda temelji na prekretnici koja je nastupila upotrebom dubokih neuronskih mreža. Na temelju te strukture, u prvome dijelu ovoga pregleda izložen je kratak opis istraživanja koja prethode upotrebi dubokih neuronskih mreža i temelje se na tzv. izrađenim značajkama slike (engl. hand-crafted image features), dok je ostatak pregleda posvećen istraživanjima koja za različite oblike analize koriste duboke neuronske mreže i naučene značajke.

Problem automatske klasifikacije slikarskih djela jedan je od najranije prepoznatih i najčešće zastupljenih istraživačkih izazova u području računalne analize slikarstva. Jedna od prvih metoda za automatsku klasifikaciju slikarskih djela prema autoru predložena je u [1], gdje autori korištenjem lokalnih značajki koje se temelje na koeficijentima diskretne kosinusne transformacije (DCT) i naivnog Bayesovog klasifikatora postižu klasifikacijsku točnost od 86% na bazi slika sastavljenoj od djela pet različitih slikara. Nešto kasnije predstavljen je ArtHistorian [2], prototip sustava za klasifikaciju slikarskih djela prema autoru i stilskim razdobljima u kojem je svaka slika opisana 6-dimenzionalnim vektorom koji uključuje globalne statističke značajke poput udjela tamnih piksela u slici, te različitih značajki histograma intenziteta. Na temelju reprezentacije slika tim skupom značajki, metodom potpornih vektora (engl. SVM support vector

machine) ostvarili su apsolutnu klasifikacijsku točnost na skupu od 290 slika od 12 različitih autora. Iste godine, Lombardi et al. [3] predložili su model koji treba poslužiti kao osnova interaktivnog sustava za pretraživanje slikarskih djela. Model se temelji na značajkama slike koje su podijeljene u dvije kategorije: značajke palete (engl. palette features) koje su povezane sa skupom boja u slici i značajke slikarskog platna (engl. canvas features) koje izražavaju frekvencijsku i prostornu distribuciju korištenih boja. Značajke palete čini ukupan broj jedinstvenih RGB (engl. red blue green) trojki na slici, dok u značajke platna spadaju minimumi, maksimumi, srednje vrijednosti, median i standardna devijacija za R, G i B komponentu slike te također za svaku komponentu HSV (engl. hue saturation value) prikaza slike. Sustav za pretraživanje funkcionira tako da korisnik odabere sliku na temelju koje mu sustav predloži deset najbližijih slika dobivenih izračunom Euklidske udaljenosti između skupova značajki.

Značajno veći skup značajki za reprezentaciju slike korišten je u [4] gdje je predložena metoda za klasifikaciju slikarskih djela prema autoru koja se zasniva na ekstrakciji globalnih i lokalnih značajki boje (HSV histogram), teksture (Gaborov filter) i kontura (Cannyjev detektor rubova). Za klasifikaciju 1080 slika u 25 različitih klasa korištena je neuronska mreža temeljena na radijalnoj baznoj funkciji (RBF), a postignuta točnost klasifikacije iznosi 69.7% Shamir et al. [5] predstavili su metodu za automatsko prepoznavanje devet različitih slikara i tri slikarska pravca (impresionizam, ekspresionizam i nadrealizam) na temelju velikog broja značajki. Skup od nekoliko tisuća značajki čine numerički deskriptori dobiveni primjenom različitih metoda ekstrakcije značajki iz izvorne slike i njenih transformacija, npr. značajke dobivene Chebyshevljevom i Radonovom transformacijom, značajke dobivene primjenom Gaborovih filtera, Prewitovog detektora rubova, teksturne značajke dobivene Haralickovom metodom itd. Točnost klasifikacije iznosi 77% za klasifikaciju prema slikarima i 91% za klasifikaciju prema slikarskim pravcima korištenjem klasifikacijske metode  $k$  – najbližih susjeda s težinskim faktorima. Na temelju iste metode u [6], uz korištenje proširene baze slika (34 slikara), izrađena je mapa sličnosti, odnosno dan je stablasti prikaz međuodnosa pojedinih slikara na osnovi Euklidske udaljenosti među skupovima značajki koje predstavljaju sliku. Sličan princip vizualizacije odnosa među slikarima na temelju izračuna udaljenosti između različitih skupova značajki predložen je u [7] i [8]. Većina ranijih radova posvećenih klasifikaciji slikarskih djela ima sličan metodološki princip. Varijacije različitih pristupa očituju se klasifikacijskom zadatku (autor, stil, žanr), u izboru skupa značajki, pri čemu se najčešće radi o značajkama boje i teksture te u odabiru različitih metoda klasifikacije, pri čemu se najčešće provodi analiza točnosti različitih tipova klasifikatora nad istim skupom značajki [9, 10, 11, 12, 13, 14, 15].

Različita istraživanja automatske klasifikacije slikarskih djela suočavala su se sa sličnim preprekama i izazovima, a najizraženija bila je nedostatak opsežne zajedničke baze slika kako bi se adekvatno uspoređivali klasifikacijski rezultati. U većini ranijih istraživanja korištene su relativno male baze slika, s arbitrarno odabranim i varirajućim skupovima klasa. Nedavni zna-

čajni napredak u području automatske klasifikacije potaknut je dvama uzrocima: s jedne strane pojavom velikih, detaljno označenih i online dostupnih baza digitaliziranih slikarskih djela, a s druge strane značajnim napretkom u području računalnog vida postignutog korištenjem dubokih konvolucijskih neuronskih mreža (engl. Convolutional Neural Networks - CNN).

Jedan od glavnih argumenata za nedavni uspjeh konvolucijskih neuronskih mreža u rješavanju mnogih problema u području računalnog vida jest pojava iznimno velikih baza slika s označenim sadržajem, kao što je ImageNet [16], baza od preko 15 milijuna digitalnih slika označenih oznakama 22 000 različitih kategorija. Od 2010. godine ImageNet projekt organizira ImageNet Large Scale Visual Recognition Challenge (ILSVRC), godišnje natjecanje u klasifikaciji i raspoznavanju objekata u slikama korištenjem podskupa cjelokupne ImageNet baze koji se sastoji 1000 slika za svaku od 1000 kategorija. Značajna prekretnica u rješavanju problema automatske klasifikacije dogodila se kada je na ILSVRC natjecanju održanom 2012. godine pobijedio model duboke konvolucijske neuronske mreže AlexNet [17], uz izrazito poboljšanje klasifikacijske točnosti u odnosu na prethodne godine. Pojava te metode uzrokovala je revoluciju u različitim područjima računalnog vida jer je pokazano da se korištenjem tzv. naučenih značajki, odnosno reprezentacija slike dobivenih iz slojeva duboke neuronske mreže, postižu značajno bolji rezultati nego s dotada korištenim tzv. oblikovanim značajkama (engl. engineered features) za različite klasifikacijske zadatke i u različitim domenama. Daljne poboljšanje u mnogim domenama dobiveno je treniranjem mreže nad domenski specifičnim skupom slika, uvođenjem metode finog podešavanja parametara (engl. fine-tuning) duboke konvolucijske neuronske mreže, kao i razvojem kompleksnijih i slojevitijih arhitektura konvolucijskih neuronskih mreža.

U kontekstu klasifikacije slikarskih djela, konvolucijske neuronske mreže uvedene su najprije kao ekstraktori značajki. Karayev et al. [18] predstavili su pristup koji se temelji na uzimanje izlaza dvaju različitih slojeva duboke konvolucijske neuronske mreže koja je prethodno naučena raspoznavati prirodne slike na temelju objekata u njima, kao vektora značajki za klasifikaciju umjetničkih slika prema stilu. Bitan doprinos toga rada jest što je pokazano da su značajke dobivene iz mreže naučene za raspoznavanje objekata, nadmoćne nad mnogim drugim oblikovanim značajkama slike čak i kad je u pitanju klasifikacija prema sasvim drugačijoj prirodi zadatka, odnosno prema raspoznavanju stila umjetničkog djela. Učinkovitost značajki dobivenih iz konvolucijskih neuronskih mreža, posebice u kombinaciji s drugim oblikovanim značajkama slike, potvrđena je za klasifikaciju slikarskih djela prema stilu [19], žanru [20, 21] i autoru [22].

Osim korištenja konvolucijskih neuronskih mreža kao ekstraktora značajki, dodatno poboljšanje klasifikacijske točnosti u mnogim domenama računalnog vida postignuto je metodom finog podešavanja parametara predtrenirane konvolucijske neuronske mreže korištenjem novog skupa podataka [23]. Superiornost ovog pristupa potvrđena je i za klasifikaciju slikarskih djela

s obzirom na različite oblike kategorizacije [24, 25, 26, 27]. Također, osim za problem klasifikacije, ovaj pristup pokazao se uspješnim i za druge zadatke kao što su raspoznavanje objekata u slikarskim djelima [28], dohvaćanje vizualnih poveznica u kolekcijama slikarskih djela [29], razlikovanje ilustracija od fotografija [30]. Preduvjet za korištenje metode podešavanja parametara predtreniranih konvolucijskih neuronskih mreža dostupnost je dovoljno velikog skupa za treniranje, odnosno dovoljno velike baze slikarskih djela.

Nedavna pojava online dostupnih baza slika poput WikiArt baze koja sadrži više od 130 000 označenih umjetničkih djela, omogućila je treniranje dubokih neuronskih mreža na takvom specifičnom skupu podataka. Također, pojava takve baze stvorila je jednoznačan okvir za usporedbu različitih klasifikacijskih metoda. Osim WikiArt baze koja je trenutno najveća i najčešće korištena baza u problemima klasifikacije [14, 18, 19, 22, 23, 24, 31], korištene su i druge baze slikarskih djela, poput Web Gallery of Art (WGA) baze koja sadrži više od 40 tisuća slika [29] ili Rijksmuseum Challenge baze umjetničkih djela [12, 27]. Nadalje, postojalo je nekoliko inicijativa izrade baza slikarskih djela prvenstveno namijenjenih rješavanju problema automatske klasifikacije kao što su Painting-91 baza sa 4266 slika od 91 različitog autora [32], zatim Pandora baza sa 7724 slika iz 12 različitih stilskih razdoblja [13] te nedavno predstavljena OmniArt baza s više od milijun slika umjetničkih djela agregiranih iz različitih muzejskih baza [33].

Osim rješavanja problema automatske klasifikacije, dostupnost velikih i kvalitetno označenih baza otvorilo je nove mogućnosti računalne analize slikarskih baza kojima je cilj ne samo računalom dostići razinu ljudskog ekspertnog znanja, već koristiti računalne metode za otkrivanje novih znanja. Pri tome je jedan od najvećih izazova modeliranje složenih i semantički relevantnih značajki koje odgovaraju konceptima formalne analize slikarskih djela. Ranija istraživanja temeljila su se na povezivanju oblikovanih značajki slike niže razine u složenije značajke, pa su tako predložene razne metode za oblikovanje viših značajki na temelju razrađene ontologije umjetničkih koncepata [34, 35], pri čemu najčešće prevladava usmjerenje na jedan aspekt stilskih obilježja kao što su modeliranje složenih značajki odnosa boja [36, 37, 38, 39, 40, 41], poteza kistom [42] i kompozicije [43, 44]. Superiornost dubokih neuronskih mreža u rješavanju mnogih problema računalnog vida potaknula je istraživanje novih metodologija razrade složenih značajki na temelju naučenih reprezentacija slike umjesto na niskim značajkama boje, teksture, rubova i dr. Elgammal et al. [45] pokazali su da postoji pozitivna korelacija između značajki ekstrahiranih iz neuronske mreže naučene za raspoznavanje umjetničkog stila i stručno procijenjenoj vrijednosti zastupljenosti određenih principa formalnog oblikovanja slike. Kim et al. [46] predstavili su metodu za semantičku interpretaciju naučenih značajki konvolucijske neuronske mreže na temelju povezivanja s reprezentacijama riječi koje označavaju različite vizualne atribute, dok su Brachmann et al. [47] koristili značajke neuronske mreže za analizu razlike između umjetničkih i prirodnih slika. Adekvatno oblikovane mjere zastupljenosti određenih koncepta omogućuju da se analizom velikih baza slikarskih djela utvrde tokovi promjena

različitih vizualnih obilježja kroz povijest [48, 49] te da se stvore novi sofisticirani sustavi za proučavanje i pretraživanje baza umjetničkih slika.

## 1.2 Cilj istraživanja

Cilj je ovog doktorskog rada ostvarenje metoda za računalnu analizu i detekciju stilskih obilježja slikarskih djela. Temeljna motivacija istraživanja predstavljenog u ovome radu ostvarenje je pomaka u smjeru prevladavanja semantičkog jaza između razumijevanja značajki slike u kontekstu računalnog vida i značajki slike u kontekstu povijesti umjetnosti. Oblikovanje značajki više razine koje predstavljaju kvantifikaciju zastupljenosti određenih umjetničkih koncepata u slici zasnovano je na primjeni dubokih konvolucijskih neuronskih mreža. Većina dosadašnjih istraživanja u području računalne analize slike ograničena je na primjenu dubokih konvolucijskih mreža isključivo sa svrhom klasifikacije likovnih djela. U ovome radu istražuju se mogućnosti njihove primjene u raspoznavanju složenih značajki koje po uzoru na stilska obilježja. Razvojem takvih značajki omogućuje se pretraživanje kolekcija slikarskih djela na temelju semantičkih, vizualnih i afektivnih dimenzija razumijevanja slikarskih djela. Osim istraživanja primjenjivosti konvolucijskih neuronskih mreža za razvoj metoda naprednog pretraživanja slika i potencijalnog otkrivanja znanja u domeni povijesti umjetnosti, istraživanje izloženo u ovome radu dotiče se i interpretabilnosti samih konvolucijskih neuronskih mreža istraživanjem utjecaja različitih postavki treniranja i domenski specifičnih inicijalizacija parametara mreže.

Zbog ostvarenja cilja istraživanja rezultati istraživanja ovog doktorskog rada daju odgovore na sljedeće hipoteze:

- Računalnom klasifikacijom moguće je kategorizirati slike sukladno povijesno-umjetničkim kategorijama stila, žanra i autorstva.
- Primjenom računalnih metoda moguće je raspoznati složene značajke stilskih obilježja slikarskih djela.
- Dobiveni rezultati računalne analize sukladni su ljudskim vrednovanjem zastupljenosti određenih koncepata u slici.

Ishodi istraživanja predstavljeni su kvalitativnim i kvantitativnim rezultatima. Budući da se u fokusu istraživanja nalazi računalna detekcija složenih obilježja slike koja nisu numeričke prirode i čija kvantifikacija predstavlja zaseban izazov, evaluacija ishoda istraživanja uvelike se oslanja na kvalitativne rezultate, odnosno vizualne primjere podskupova slika u kojima su zastupljena određena obilježja. Također, provjera dobivenih rezultati temelji se na usporedbi s teorijskim znanjem iz područja povijesti umjetnosti, kao i s rezultatima ispitivanja ljudskih procjena prisutnosti određenih obilježja u slici.

## 1.3 Struktura rada

Doktorski rad strukturiran je u šest poglavlja. U uvodnom poglavlju izložena je motivacija istraživanja te je opisan kontekst potreban za razumijevanja istraživačkog pristupa izloženog u ovome radu. Također je dan pregled dosadašnjih istraživanja u području računalne analize slikarskih djela te su istaknuti glavni istraživački ciljevi rada.

Budući da su metode predstavljene u ovom radu zasnovane na primjeni konvolucijskih neuronskih mreža, u drugom poglavlju dan je teorijski pregled potreban za razumijevanje dubokih konvolucijskih neuronskih mreža. Predstavljena su osnovna načela i temeljne sastavnice strukture konvolucijskih neuronskih mreža, kao i principi učenja te korištene metode optimizacije i regularizacije.

Struktura poglavlja u kojima su izloženi istraživački problemi i doprinosi prati tri različite razine pristupa likovnom djelu: kategorizaciju, formalnu analizu i doživljajnu analizu. Stoga je treće poglavlje posvećeno problemu kategorizacije slikarskih djela, odnosno raspoznavanju pripadnosti djela određenom autoru, žanru, stilu i povijesno-geografskom kontekstu. Prikazani su rezultati dobiveni metodom finog podešavanja parametra konvolucijskih neuronskih mreža za različite klasifikacijske zadatke i različite baze slika. Opisani su rezultati usporedne analize različitih postavki učenja mreža, s posebnim naglaskom na utjecaj domenski specifičnih inicijalizacija parametara mreže. Također su prikazane mogućnosti primjene modela naučenih za klasifikaciju u problemima analize vizualne sličnosti.

U četvrtom poglavlju predstavljene su rezultati računalne predikcije i analize specifičnih stilskih obilježja. Pod specifičnim stilskim obilježjima podrazumijevaju se koncepti usko vezani uz povijesno-umjetničko razmatranje razlike stilova i formalnu analizu likovnih djela. Predikcija navedenih obilježja ostvarena je učenjem regresijskih modela konvolucijskih neuronskih mreža te su opisane postavke treniranja i rezultati predikcije dobiveni primjenom modela. Izložene su analize kvantitativnih rezultata svakog obilježja te je razmotren odnos predikcijskih vrijednosti dobivenih obilježja u kontekstu povijesti umjetnosti.

Peto poglavlje posvećeno je doživljajnoj analizi likovnog djela te se u tom kontekstu razmatraju tri različita percepcijska obilježja slike: estetska evaluacija slike, osjećajni doživljaj slike i pamtljivost slike. Predstavljena je metodologija zasnovana na prijenosu znanja iz prirodnih slika na umjetničke slike korištenjem konvolucijskih neuronskih mreža. Izneseni su rezultati evaluacije različitih modela i analize naučenih značajki. Također, provedena je analiza odnosa dobivenih značajki percepcijskih obilježja s drugim složenim značajkama slike te njihov odnos prema pojedinim povijesno-umjetničkim kategorijama.

U posljednjem, šestom poglavlju, prikazan je osvrt na rezultate istraživanja i zaključno razmatranje znanstvenih doprinosa te su dane smjernice za buduća istraživanja.

## Poglavlje 2

# Konvolucijske neuronske mreže

Konvolucijske neuronske mreže poseban su oblik dubokih neuronskih mreža specijalizirane za podatke u kojima postoji prostorna korelacija među susjednim točkama, odnosno podatke čija topologija nalikuje rešetki (engl. grid like topology), kao što su vremenski slijedovi (1D) ili slike (2D). Budući da se metode i rezultati predstavljeni u ovome radu temelje na upotrebi (i modifikaciji) konvolucijskih neuronskih mreža za analizu i raspoznavanje digitaliziranih slikarskih djela, u ovome je poglavlju dan osnovni pregled temeljnih pojmova potrebnih za razumijevanje strukture i načina rada konvolucijskih neuronskih mreža.

### 2.1 Struktura konvolucijskih neuronskih mreža

Struktura svih umjetnih neuronskih mreža određena je njihovim temeljnim sastavnicama koje čine neuroni. Poznato je da su umjetne neuronske mreže nastale po uzoru na model ljudskog mozga, odnosno mreže međupovezanih bioloških neurona koji primaju signale i aktiviraju izlazne impulse ovisno o jačini podražaja te se osnovni princip rada umjetne neuronske mreže temelji na utjecaju jačine veza među neuronima. U umjetnoj neuronskoj mreži neuroni su međusobno povezani vezama kojima se dodjelju težine čija se vrijednost podešava prilikom učenja mreže. Postoji više različitih struktura umjetnih neuronskih mreža koje se razlikuju ovisno o funkciji i načinu rada mreže. Za problem klasifikacije, koji se svodi na preslikavanje ulaznih uzoraka u neku od definiranih klasifikacijskih kategorija, najčešće se koriste unaprijedne neuronske mreže (engl. feedforward networks). Struktura unaprijedne neuronske mreže određena je jednim ulaznim slojem čiji broj neurona odgovara dimenzionalnosti vektora ulaznih podataka, zatim jednim ili više skrivenih slojeva te izlaznim slojem u kojem je broj neurona jednak broju klasa. Broj skrivenih slojeva određuje dubinu mreže, stoga se pod pojmom „dubokih“ neuronskih mreža podrazumijeva mreža s većim brojem skrivenih slojeva čija je zadaća transformacija podataka dobivenih iz prethodnog sloja i slanje transformirane reprezentacije podataka u sljedeći sloj.

Transformacija podataka unutar mreže određena je vrijednošću težina dodjeljenih vezama među neuronima, kao i aktivacijskim funkcijama neurona. Svaki neuron određen je aktivacijskom funkcijom koja prima sumu umnoška ulaza neurona s pripadnim težinama veza te ih preslikava na izlaz neurona, što se sažeto može prikazati izrazom 2.1:

$$y_i = \sum_j f(w_{ij}x_j) = f(\mathbf{w}^T \mathbf{x}) \quad (2.1)$$

Pri čemu vektor  $\mathbf{x}$  predstavlja ulazne varijable, vektor  $\mathbf{w}$  vektor težina veza koje neuron povezuju s neuronima prethodnog sloja i čije se vrijednosti optimiraju postupkom učenja mreže (zbog pojednostavljenja pomak se promatra kao težina  $w_0$  koja je uvijek povezana s ulaznom vrijednošću 1), dok  $f$  predstavlja aktivacijsku funkciju neurona čija je uloga osigurati nelinearnost modela.

### 2.1.1 Slojevi konvolucijske neuronske mreže

Specifičnost konvolucijskih neuronskih mreža jest u tome da sadrži barem jedan konvolucijski sloj, odnosno sloj u kojemu se provodi operacija konvolucije. Konvolucija je operacija koja na temelju dvije funkcije vraća modificiranu verziju prve funkcije s obzirom na količinu preklapanja s drugom funkcijom. U kontekstu dvodimenzionalnih podataka, odnosno slika, konvolucija daje odziv ulazne slike ( $I$ ) i filtra ( $F$ ), koji predstavlja linearnu kombinaciju elemenata ulazne slike. U praktičnoj implementaciji konvolucijskih neuronskih mreža, najčešće se koristi izmijenjena verzija konvolucije, odnosno operacija kros-korelacijske funkcije, definirana izrazom 2.2:

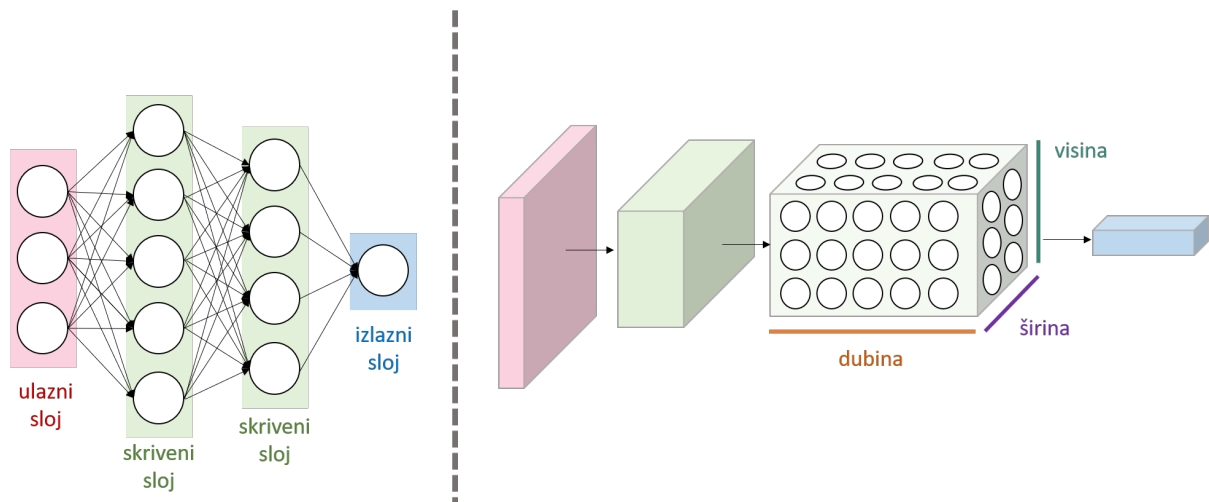
$$S(i, j) = (F * I)(i, j) = \sum_m \sum_n I(i + m, j + n)F(m, n) \quad (2.2)$$

Operacija kros-korelacije jednaka je operaciji konvolucije samo s obrnutim vrijednostima filtra. Obrtanje vrijednosti filtra nema utjecaja u praktičnoj primjeni budući da će mreža naučiti iste značajke bez obzira na poziciju filtra.

U kontekstu konvolucijskih neuronskih mreža, filter predstavlja skup težina neurona povezanih s prethodnim slojem. Osobitost konvolucijskih neuronskih mreža je u tome da su neuroni raspoređeni u tri dimenzije, kako je ilustirano prikazom na slici 2.1.

Sloj konvolucijske neuronske mreže transformira ulazni volumen u novi volumen aktivacija nizom diferencijalnih funkcija, pri čemu je svaki volumen aktivacija određen svojom dubinom, širinom i visinom. Struktura konvolucijskih neuronskih mreža sastavljena je od tri osnovna tipa slojeva: konvolucijski sloj, sloj sažimanja i potpuno povezani sloj. Svaki konvolucijski sloj sadrži unaprijed odabrani broj filtra, tzv. jezgri (engl. kernel). Neuroni konvolucijskog sloja raspoređeni su tako da dubina konvolucijskog sloja predstavlja broj različitih filtra. Neuroni





**Slika 2.1:** Lijevo: klasična višeslojna unaprijedna neuronska mreža (višeslojni perceptron) Desno: konvolucijska neuronska mreža s neuronima raspoređenim u tri dimenzije.

smješteni na istoj plohi određenoj visinom i širinom aktivacijskog volumena pojedinog konvolucijskog sloja dijele iste vrijednosti težina te prema tome zajednički predstavljaju jedan filter. Izlazi dobiveni iz neurona smještenih na istoj plohi predstavljaju odziv pojedinog filtra i ulaznog volumena i tvore aktivacijsku mapu, odnosno tzv. mapu značajki (engl. feature map). Sve aktivacijske mape dobivene od svih filtera jednog konvolucijskog sloja tvore izlazni aktivacijski volumen toga sloja i ulaz u sljedeći sloj. Svaki neuron unutar iste plohe određene visinom i širinom aktivacijskog volumena povezan je samo s jednim manjim podskupom neurona iz prethodnog sloja. Pri tome su svi neuroni koji se nalaze na istoj poziciji unutar plohe određene širinom i visinom, a na različitim dubinama volumena, povezani na isti podskup neurona iz prethodnog sloja. Na taj se način svaki filter proteže punom dubinom ulaznog volumena. Podskup neurona povezanih s prethodnim slojem određuje tzv. receptivno polje pojedinog sloja, odnosno broj neurona u podskupu odgovara veličini filtra.

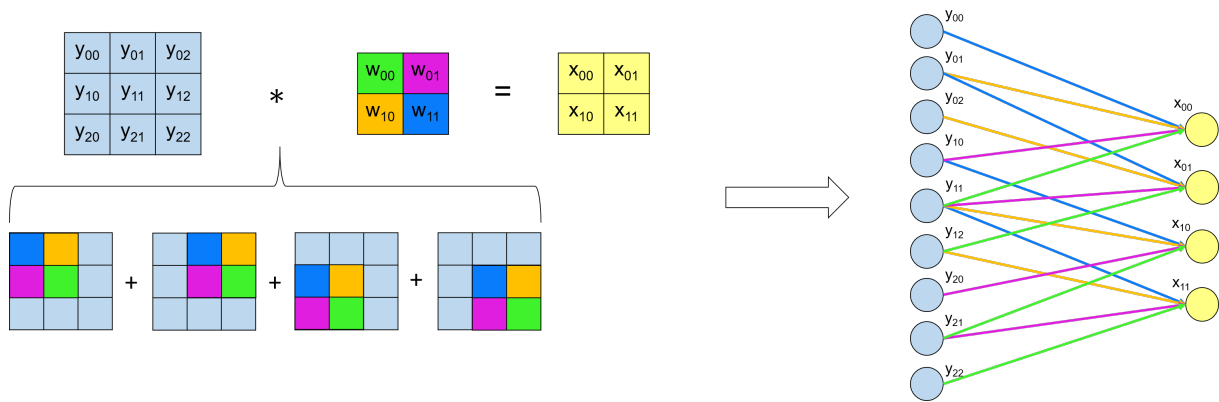
Veličina filtra ( $F$ ) jedan je od temeljnih hiper-parametara konvolucijskog sloja koji određuje dimenzije izlaznog volumena. Osim veličine filtra, ključnu ulogu ima i korak  $S$  (engl. stride) koji predstavlja broj piksela za koji se filter pomiče u horizontalnom i vertikalnom smjeru prilikom konvolucije filtra i ulazne mape. U slučaju da se radi o prvom konvolucijskom sloju, ulazna mapa je ili monokromatska slika ili jedna od triju R, G, B (red, green, blue) komponenti slike, pri čemu tada dimenzija dubine ulaznog volumena iznosi tri, tj. odgovara broju komponenti slike. Također, jedan od faktora koji određuje veličinu izlaznog aktivacijskog volumena je nadopunjavanje nulama (engl. zero padding). Nadopunjavanjem nulama dodaje se  $P$  redova nula sa svake strane izvorne ulazne mape. Osim što se nadopunjavanjem nulama omogućava kontroliranje veličine izlaznog aktivacijskog volumena, tim se postupkom osigurava da se ne izgube informacije koje se nalaze na rubovima ulazne mape, odnosno da se očuva odziv filtra i rubnih dijelova mape, tj. izvorne slike ukoliko se radi o prvom sloju.

S obzirom na spomenute parametre dimenzije filtra ( $F \times F$ ), veličinu ulazne mape ( $I$ ) pomaka ( $S$ ) i redova nadopunjavajih nulama ( $P$ ), dimenzije izlazne mape značajki iznosi  $N \times N$ , pri čemu je:

$$N = \frac{(I - F + 2P)}{S} + 1 \quad (2.3)$$

Dimenzija ukupnog izlaznog volumena iznosi  $N \times N \times D$ , pri čemu je  $D$  dubina volumena, odnosno broja različitih filtera u sloju.

U svrhu razumijevanja pojedinosti izvedbe konvolucije kao umnoška težina neurona, na slici 2.2 nalazi se pojednostavljen prikaz koji ilustrira konvoluciju jedne mape značajki dimenzije  $3 \times 3$  s jednim filtrom dimenzija  $2 \times 2$ .



**Slika 2.2:** Ilustracija ostvarivanja konvolucije mape značajki dimenzije  $3 \times 3$  s filtrom dimenzija  $2 \times 2$  povezivanjem neurona u konvolucijskom sloju.

Iz načina na koji se konvolucija ostvaruje unutar konvolucijske mreže, proizlaze određena specifična svojstva strukture konvolucijske neuronske mreže u odnosu na višeslojni perceptron. U višeslojnom perceptronu svaki neuron trenutnog sloja povezan je sa svakim neuronom prethodnoga sloja, pri čemu je svaka veza između dva neurona predstavljena jedinstvenom vrijednošću težine. U konvolucijskim slojevima više veza među različitim neurona posjeduje jednake vrijednosti težina, pa se na taj način ostvaruje svojstvo dijeljenja parametara unutar mreže. Također, iz slike 2.2 vidljivo je da izvedba konvolucije ne zahtijeva povezivanje svih neurona međusobno, odnosno konvolucijske slojeve karakterizira tzv. raspršena povezanost neurona.

Ulaz u neuron sloja  $l$  dobiven je kao zbroj umnožaka težina filtra  $w$  (dimenzija  $F \times F$ ) i pripadnih izlaza prethodnog sloja  $l - 1$ , te je definiran izrazom 2.4 koji odgovara operaciji konvolucije:

$$x_{ij}^l = \sum_{a=0}^{F-1} \sum_{b=0}^{F-1} w_{ab} y_{(i+a)(j+b)}^{l-1} \quad (2.4)$$

Na ulaz neurona  $x_{ij}^l$  zatim se primjenjuje aktivacijska funkcija:

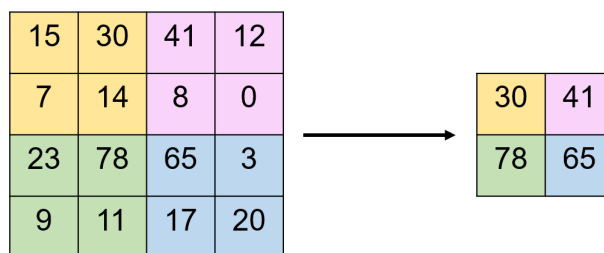
$$y_{ij}^l = f(x_{ij}^l) \quad (2.5)$$

Ovisno o aktivacijskoj funkciji neurona u sloju, pojedini slojevi obavljaju različite vrste transformacije podataka. Postoji više vrsta aktivacijskih funkcija koje se odabiru s obzirom na namjenu modela i položaj sloja u mreži. U konvolucijskim slojevima najčešće se koristi aktivacijska funkcija zglobnice, odnosno ReLu funkcija (engl. Rectified Linear Unit), definira kao:

$$f(x_{ij}^l) = \max(0, x_{ij}^l) \quad (2.6)$$

U praktičnoj se primjeni relu funkcija pokazala mnogo boljom od nekih drugih aktivacijskih funkcija, kao što su sigmoidna funkcija ili funkcija hiperbolnog tangensa, ponajprije zbog toga što se zbog njene jednostavnosti ubrzava algoritam učenja.

Nakon jednog ili bloka konvolucijskih slojeva, slijedi sloj sažimanja (engl. pooling layer). Svrha sloja sažimanja je smanjenje dimenzija mapa značajki dobivenih iz izlaza konvolucijskog sloja, uz što manji gubitak informacija. Postoji više vrsta metoda sažimanja ovisno o operaciji koja se provodi nad podacima, no najčešće je korištena metoda sažimanja maksimalnom vrijednošću (engl. max-pooling). Primjenom toga postupka određeno područje mape značajki zamjenjuje se najvećom vrijednošću u tome području. Ilustracija postupka maksimalnog sažimanja dana je na slici 2.3.



**Slika 2.3:** Ilustracija principa operacije sažimanja mape značajki dimenzija  $4 \times 4$  izborom maksimalnog elementa s filterom veličine  $2 \times 2$  i korakom 2.

Sažimanje se obično provodi bez preklapanja područja tako da se mapa značajku podijeli na različite regije te se svaka regija sažima u jednu vrijednost. Na taj se način mapa značajki smanjuje  $k$  puta, pri čemu je  $k$  veličina regije. Osim što se smanjuje dimenzionalnost aktivacijskih volumena, sažimanjem se postiže invarijantnost modela na male translacije u slici.

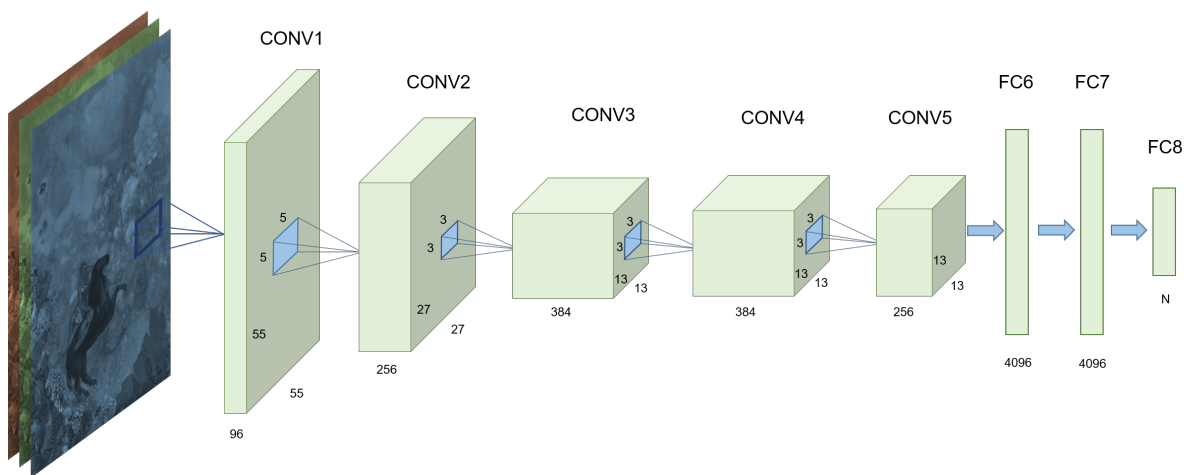
Na kraju konvolucijske neuronske mreže nalaze se potpuno povezani slojevi (engl. fully connected layers). U potpuno povezanim slojevima svaki je neuron povezan sa svakim neuronom aktivacijskog volumena iz prethodnog sloja. U praktičnoj implementaciji konvolucijskih mreža najčešće se na kraju mreže nalazi nekoliko potpuno povezanih slojeva, pri čemu je zadnji

potpuno povezani sloj ujedno i izlazni sloj. Broj neurona u izlaznom sloju odgovara broju klasa koje mreža uči raspoznavati, pri čemu se tada u neuronima najčešće koristi softmax aktivacijska funkcija koja daje probabilističku vrijednost pripadnosti uzorka određenoj klasi:

$$f(z)_j = \frac{e^{z_j}}{\sum_{k=1}^K e^{z_k}} \quad (2.7)$$

## 2.1.2 Arhitektura konvolucijske neuronske mreže

Broj i način slaganja konvolucijskih, sažimajućih i potpuno povezanih slojeva određuje arhitekturu mreže. Postoji više različitih arhitektura koje se razlikuju s obzirom na dubinu i organizaciju mreže. Bazna arhitektura eksperimenata predstavljenih u ovom radu je model AlexNet mreže [17] koja se sastoji od pet konvolucijskih slojeva i tri potpuno povezana sloja. Na slici 2.4 dan je pojednostavljeni prikaz AlexNet arhitekture s naglašenim veličinama filtera u svakom sloju i pripadajućih dimenzija aktivacijskih volumena.



**Slika 2.4:** Arhitektura AlexNet mreže.

Aktivacijska funkcija u svim skrivenim slojevima je ReLu aktivacijska funkcija, dok je u izlaznom sloju implementirana softmax aktivacijska funkcija. Budući da je AlexNet model prvi puta predstavljen na ImageNet ILSVRC natjecanju s ciljem klasifikacije slika u 1000 različitih kategorija objekata, izlazni sloj mreže sastoji se od 1000 neurona. Prilikom učenja mreže za neki drugi klasifikacijski problem, originalni izlazni sloj se zamjenjuje novim slojem čija dimenzionalnost odgovara broju klasa u novom klasifikacijskom problemu. Na ulazni sloj mreže šalje se slika dimenzija  $227 \times 227$ .

Osim AlexNet arhitekture, koja je jedna od najranijih i najpoznatijih arhitektura, postoje mnoge druge složenije i dublje arhitekture kao što su VGG [50], Google LeNet [51] ili ResNet [52].

## 2.2 Učenje konvolucijske neuronske mreže

Pod pojmom učenja mreže podrazumijeva se postupak pronalaženja optimalnih parametara s obzirom na ciljni zadatak i skup podataka. Nakon definiranja arhitekture konvolucijske mreže koja je prikladna za rješavanje određenog zadatka, potrebno je pronaći parametre mreže koji će osigurati najbolji mogući rezultat, pri čemu se pod parametrima podrazumijevaju vrijednosti težina među neuronima mreže. Uobičajeni postupak prilikom učenja mreže jest da se vrijednosti težina inicijaliziraju na slučajne vrijednosti koje se zatim podešavaju postupkom učenja. Najčešće korištena metoda učenja višeslojnih mreža je algoritam propagacije pogreške unatrag (engl. backpropagation) koji se koristi zajedno s optimizacijskim metodama kao što je metoda gradijentalnog spusta (engl. gradient descent).

### 2.2.1 Algoritam propagacije pogreške unatrag

Postupak nadgledanog učenja podrazumijeva da se na ulaz mreže dovodi niz poznatih uzoraka podataka, odnosno uzoraka za koje se zna kakav treba biti odziv mreže, tj. kojoj klasi pripadaju u slučaju učenja mreže za klasifikacijske zadatke. Zatim se na temelju odziva mreže na ulazni uzorak računa funkcija pogreške kao razlika dobivenog i očekivanog izlaza. Cilj je provođenja algoritma propagacije pogreške unatrag da se pogreška očekivanog i dobivenog izlaza širi unatrag kroz slojeve mreže kako bi se težine među neuronima ažurirale na način da se minimizira funkcija pogreške. Učenje mreže postiže se nizom interakcija koje se sastoje od propagacije unaprijed (računanje odziva na ulazni uzorak) i propagacije pogreške unatrag te ažuriranjem vrijednosti težina u svakoj iteraciji.

Propagaciju pogreške unatrag, od izlaznog sloja mreže do ulaznog sloja, moguće je izračunati primjenom lančanog pravila deriviranja. Provođenje algoritma propagacije pogreške unatrag u konvolucijskim neuronskim mrežama slično je njegovom provođenju u klasičnim unaprijednim višeslojnim mrežama, uz određene modifikacije s obzirom na specifično svojstvo dijeljenja težina među neuronima u konvolucijskim slojevima. U nastavku je opisan postupak računanja propagacije pogreške i ažuriranje težina između dva konvolucijska sloja primjenom metode gradijentalnog spusta.

Uz pretpostavku da znamo vrijednost pogreške na izlazu iz sloja, potrebno je izračunati gradijente funkcije pogreške prethodnoga sloja tako da se najprije izračuna gradijent funkcije pogreške s obzirom na ulaz u svaki neuron. Na temelju izraza za određivanje vrijednosti izlaza iz neurona konvolucijskog sloja danog u 2.6, gradijent funkcije pogreške  $E$  s obzirom na ulaznu vrijednost neurona konvolucijskog sloja  $l$  određen je kao:

$$\frac{\partial E}{\partial x_{ij}^l} = \frac{\partial E}{\partial y_{ij}^l} \frac{\partial y_{ij}^l}{\partial x_{ij}^l} = \frac{\partial E}{\partial y_{ij}^l} f'(x_{ij}^l) \quad (2.8)$$

Gradijent funkcije pogreške po ulazu u neuron jednak je umnošku parcijalne derivacije funkcije pogreške po izlazu neurona i parcijalne derivacije izlaza po ulazu koja odgovara derivaciji aktivacijske funkcije neurona.

Na temelju vrijednosti gradijenta funkcije greške po ulazu moguće je izračunati gradijente težina između dva sloja:

$$\frac{\partial E}{\partial w_{ab}} = \sum_{i=0}^{N-F} \sum_{j=0}^{N-F} \frac{\partial E}{\partial x_{ij}^l} \frac{\partial x_{ij}^l}{\partial w_{ab}} = \sum_{i=0}^{N-F} \sum_{j=0}^{N-F} \frac{\partial E}{\partial x_{ij}^l} y_{ij}^{l-1} \quad (2.9)$$

Težine se zatim ažuriraju u ovisnosti o stopi učenja:

$$w_{ab} \leftarrow w_{ab} - \eta \frac{\partial E}{\partial w_{ab}} \quad (2.10)$$

U postupku učenja mreže bitnu ulogu ima podešavanje vrijednosti stope učenja. Ako je vrijednost stope učenja suviše mala, dolazi do spore konvergencije gradijenta, dok su u slučaju da je vrijednost stope učenja prevelika, moguća preskakanja traženog minimuma. U mnogim se slučajevima postupno smanjivanje vrijednosti stope učenja tijekom treniranja pokazalo kao najbolje rješenje.

Propagacija pogreške na prethodni sloj provodi se sumiranjem utjecaja izlaza neurona prethodnog sloja na neurone trenutnog sloja s kojim je taj neuron povezan:

$$\frac{\partial E}{\partial y_{ij}^{l-1}} = \sum_{i=0}^{F-1} \sum_{j=0}^{F-1} \frac{\partial E}{\partial x_{(i-a)(j-b)}^{l-1}} \frac{\partial x_{(i-a)(j-b)}^{l-1}}{\partial y_{ij}^{l-1}} = \sum_{i=0}^{F-1} \sum_{j=0}^{F-1} \frac{\partial E}{\partial x_{(i-a)(j-b)}^{l-1}} w_{ab} \quad (2.11)$$

Iz izraza 2.11 vidljivo je da propagacija pogreške unatrag također odgovara operaciji konvolucije, samo što umjesto s  $x_{(i+a)(j+b)}$  filter množimo s  $x_{(i-a)(j-b)}$ , što odgovara obrtanju filtra za 180 stupnjeva. Također, izraz 2.11 primjenjiv je samo za pozicije  $i$  i  $j$  koje su udaljene za barem  $F$  od gornjeg i lijevog ruba, stoga se prilikom konvolucije gornji i lijevi rubovi nadopunjavaju nulama, što odgovara provođenju tzv. pune konvolucije (engl. full convolution).

Kod propagacije pogreške kroz slojeve sažimanja ne ažuriraju se parametri težina, već je samo potrebno propagirati pogrešku manje mape trenutnog sloja u veću mapu prethodnog sloja. U slučaju da slojevi sažimanja obavljaju operaciju maksimalnog sažimanja, tijekom računanja odziva mreže u slojevima sažimanja zapisuju se lokacije pojedinih maksimuma, te se prilikom propagacije unatrag pogreška zapisuje na tu zabilježenu lokaciju.

Među potpuno povezanim slojevima propagacija pogreške unatrag vrši se na jednak način kao i kod klasične višeslojne neuronske mreže i temelji se, kao i kod konvolucijskih slojeva, na lančanom pravilu deriviranja. Aktivacijska funkcija izlaznog sloja određuje koja funkcija pogreške će se računati na izlaznom sloju iz mreže. U slučaju da se u izlaznom sloju koristi softmax aktivacijska funkcija, na izlazni vektor najčešće se primjenjuje unakrsna entropija (engl. cross entropy) kao funkcija pogreške. Unakrsna entropija mjera je udaljenosti između dvije dis-

tribucije, u kontekstu neuronskih mreža to su distribucije očekivanog i dobivenog odziva mreže za jedan ulazni podatak.

Cilj učenja mreže je da se minimizira funkcija pogreške te se ono temelji se na izračunu gradijenta funkcije pogreške po svim parametrima mreže kako bi se optimalno ažurirale vrijednosti parametara. Ovisno o trenutku ažuriranja vrijednosti parametara razlikuju se različite varijante izvedbe metode gradijentnog spusta. Standardni gradijentni spust (engl. batch gradient descent) predstavlja varijantu u kojoj se potrebne promjene parametara izračunaju na temelju svih uzoraka iz skupa za učenje te se zatim ažuriraju. Korištenjem gradijentnog spusta konvergencija prema minimumu je stabilna ali spora jer zahtijeva računanje svih težina prije pomicanja prema minimumu, a postoji i opasnost od zapinjanja u lokalnom minimumu. S druge strane, varijanta stohastičkog gradijentnog spusta temelji se na ažuriranju težina nakon svakog uzorka te predstavlja aproksimaciju gradijentnog spusta inkrementalnom korekcijom težina. Ovakva varijanta otpornija je na zapinjanje u lokalnim minimumima zbog izrazite oscilacije gradijenta. U praktičnoj izvedbi najčešće se koristi metoda gradijentnog spusta koja predstavlja varijantu između standardnog i stohastičkog gradijentnog spusta, a temelji se na podjeli skupa za učenje na manje slučajno odabrane podskupove i ažuriranju težina nakon izračuna gradijenta svih uzoraka iz jednog podskupa (engl. minibatch gradient descent). U praktičnoj upotrebi, najčešće se navedena metoda upotrebljava pod nazivom stohastičkog gradijentnog spusta (engl. stochastic gradient descent - SGD).

### 2.2.2 Metode optimizacije i regularizacije

Izbor odgovarajuće arhitekture i hiper-parametara mreže ključan je za uspješnost učenja mreže. Također, postoje određene metode kojima se može dodatno ubrzati i poboljšati učenje, kao i postići bolja generalizacija naučenog modela. Metode optimizacije odnose se na poboljšanje učenja mreže i svladavanje slabe naučenosti (engl. underfitting), dok se metode regularizacije odnose na sprječavanje prenaučivosti mreže (engl. overfitting) koja uzrokuje slabu generalizaciju nad novim skupom podataka. Postoje brojne metode optimizacije i regularizacije, a u nastavku su pobliže objašnjene metode korištene u ovome radu.

Optimizacijska metoda uvođenja momenta u izvedbu algoritma stohastičkog gradijentnog spusta pokazala se vrlo djelotvornom u ubrzavanju postupka učenja neuronskih mreža. Naziv metode potječe od analogije s pojmom momenta u kretanju mase u fizici i temelji se na akumuliranju eksponencijalno padajućih aritmetičkih sredina kretanja prethodnih gradijenata. Primjenom ove metode uvodi se varijabla  $v$  koja se inicijalizira na 0 i predstavlja brzinu kretanja parametara te novi hiper-parametar  $\alpha$  iz intervala  $[0, 1)$  koji predstavlja moment i određuje koliko brzo prethodni gradijenti doprinose kretanju u smjeru minimuma. Na temelju tih varijabli

definira se novo pravilo ažuriranja stohastičkog gradijetnog spusta:

$$v_{ab}^{t+1} = \alpha v_{ab}^t + \frac{\partial E}{\partial w_{ab}} w_{ab}^t \quad (2.12)$$

$$w_{ab}^{t+1} = w_{ab}^t - \eta v_{ab}^{t+1}, \quad (2.13)$$

gdje  $t$  predstavlja iteraciju. Ažuriranjem težina na tako definiran način dobiva se bolja konvergencija prema točki minimuma i manje oscilacije oko točke minimuma te se bolje izbjegavaju točke lokalnih minimuma.

Osim postizanja optimalnog učenja mreže na skupu podataka za učenje, cilj treniranja mreže je postizanje dobrih performansi prilikom primjene mreže na novom skupu podataka. Jedan od osnovnih problema koji se pojavljuje kod učenja dubokih neuronskih mreža je prenaučenosť mreže. Prenaučena mreža postiže vrlo dobre rezultate na skupu podataka za učenje, no na skupu podataka za testiranje pokazuje loše rezultate. Za sprječavanje da se mreža previše prilagodi skupu za učenje koriste se različite metode regularizacije koje podrazumijevaju modifikacije algoritma učenja takve da utječu na smanjenje pogreške generalizacije, ali ne i na smanjenje pogreške učenja.

Jedna od najčešće korištenih metoda regularizacije je modifikacija funkcije pogreške uz pomoć  $L_p$  normi kako bi se spriječio kontinuirani rast težina. Funkcija pogreške modificira se dodavanjem nove vrijednosti:

$$E' = E + \lambda \|\mathbf{w}\|_p \quad (2.14)$$

gdje je  $\mathbf{w}$  vektor težina, a parametar regularizacije  $\lambda$  predstavlja doprinos koji norma  $L_p$  ima na funkciju pogreške. Dodana nova vrijednost služi za kažnjavanje funkcije pogreške u slučaju pojave težina velikih magnituda. Najčešće se koristi regularizacija  $L_2$  normom koja je definirana kao suma kvadrata komponenti vektora:

$$\|\mathbf{w}\|_2 = \sum_{i=1}^n w_i^2 \quad (2.15)$$

Primjena  $L_2$  norme dovodi do smanjenja vrijednosti težina i tako smanjuje varijancu modela te ga čini manje osjetljivim na male promjene u uzorcima ulaznih podataka.

Osim smanjivanjem vrijednosti težina, regularizacija mreže može se postići metodom slučajnog isključivanja neurona (engl. dropout) u potpuno povezanim slojevima [53]. U svakoj iteraciji unaprijednog prolaza kroz mrežu s uzorcima jednog podskupa skupa podataka za učenje i propagacije pogreške unatrag iz mreže se isključuju slučajno odabrani neuroni potpuno povezanih slojeva. Isključivanje određenog broja slučajno odabranih neurona tijekom svake iteracije odgovara učenju drugačije konfiguracije mreže u svakoj iteraciji. Isključivanje neurona



podrazumijeva postavljanje izlaza neurona na 0, a hiperparametar koji se uvodi korištenjem ove metode jest vjerojatnost gašenja neurona u sloju. Nakon generiranja skupa različitih konfiguracija mreže, predviđanje se temelji na prosječnoj vrijednosti svih konfiguracija čime se umanjuje varijanca i izbjegava prenaučenosť mreže.

Jedan od najčešćih razloga za prenaučenosť mreže nedovoljno je velik skup podataka u odnosu na broj parametara mreže. Budući da je skup podataka ograničene veličine, u svrhu regularizacije često se pribjegava različitim postupcima povećanja skupa podataka (engl. data augmentation). Određenim transformacijama postojećih podataka povećava se broj uzoraka u skupu za učenje. Primjena određenih transformacija treba biti prikladna domeni i problemu za koji se uči neuronska mreža. Transformacije postupcima kao što su rotacija, translacija ili zrcaljenje ulazne slike pokazale su se korisnima za mnoge slućajeve učenja mreža sa svrhom klasifikacije slika.

## Poglavlje 3

# Računalno raspoznavanje slikarskih djela

### 3.1 Pozadina i motivacija istraživanja

Poznavatelj likovne umjetnosti na temelju svoga znanja može prepoznati kojem autoru ili stilskome razdoblju pripada određeno djelo. Ljudske sposobnosti određivanja stilskog razdoblja i utvrđivanja autorstva nisu posljedica samo sjećanja na konkretnu ranije usvojenu informaciju, već rezultat cjelokupnog stečenog vizualnog iskustva iz područja slikarstva koje omogućuje povezivanje specifičnosti određenog slikarskoga djela s odgovarajućim autorom ili razdobljem. Procese klasifikacije i raspoznavanja čovjek provodi koristeći svoje iskustvo i kognitivne sposobnosti. Računalno se ti procesi najčešće ostvaruju tako da se čovjekovo vizualno iskustvo imitira korištenjem određene baze digitalno pohranjenih slika, pri čemu su karakteristike slikarskoga djela opisane numeričkim skupom podataka dobivenim metodama za ekstrakciju značajki slike, dok se moždane funkcije učenja i pamćenja ostvaruju upotrebom različitih metoda za strojno učenje. Velik dio istraživanja usmjerenih na problem automatske klasifikacije slikarskih djela pratio je navedenu klasifikacijsku metodologiju, dok se raznolikost pristupa rješavanju problema klasifikacije očitovala ponajprije u korištenju i ekstrakciji različitih vrsta značajki slike, zatim prema upotrebi veličinom i sadržajem različitih baza slika te primjenom različitih tipova klasifikatora. Prema tome, klasični se metodološki pristup klasifikaciji slika sastoji od tri temeljna koraka: prikupljanje baze slika, ekstrakcija značajki iz slika i treniranje modela za učenje na temelju dobivenih značajki.

Posljednjih nekoliko godina korištenjem dubokih konvolucijskih neuronskih mreža postignuti su mnogo bolji rezultati klasifikacije slika u različitim domenama. Budući da se model konvolucijske mreže trenira izravno na slikama, postupak ekstrakcije značajki automatiziran je i odvija se u konvolucijskim slojevima mreže. Primjenom konvolucijskih neuronskih mreža metodološki pristup sveden je na dva temeljna koraka: prikupljanje baze slika i učenje duboke konvolucijske neuronske mreže. U okviru takve metodologije temeljne modifikacije odnose se na postavke konvolucijske neuronske mreže, odnosno na odabir arhitekture i inicijalnog stanja

mreže, definiranje hiperparametara te primjenu različitih metoda optimizacije i regularizacije.

Budući da treniranje konvolucijske neuronske mreže zahtijeva veliki skup podataka za učenje, u razdoblju prijelaza s uobičajenog postupka ekstrakcije različitih oblikovanih značajki slike na treniranje konvolucijske mreže dostupnom bazom slika, konvolucijske neuronske mreže isprva su se koristile samo kao ekstraktori značajki. Takav pristup poznat je kao postupak tzv. učenja prijenosom (engl. transfer learning) koji podrazumijeva prijenos znanja naučenog treniranjem modela na skupu podataka iz jedne domene i na temelju jednog zadatka, na drugi zadatak i/ili drugu domenu. U kontekstu konvolucijskih neuronskih mreža, jedan od načina prijenosa naučenog znanja je da se odzivi slojeva mreže na ulaznu sliku koriste kao vektori značajki slike. Najčešće se kao vektori značajki uzimaju odzivi zadnjeg i/ili predzadnjeg sloja mreže, čija dimenzionalnost ovisi o arhitekturi mreže koja se koristi. U kontekstu klasifikacije slikarskih djela, modeli mreža naučeni za klasifikaciju fotografija na temelju objekata u njima, korišteni su za dobivanje vektora značajki, na temelju kojih se zatim vršila klasifikacija nekom od poznatih klasifikacijskih metoda. Pokazano je da se klasifikacijom tzv. naučenih značajki, tj. vektora odziva slojeva predtrenirane konvolucijske neuronske mreže, postiže veća točnost nego klasifikacijom oblikovanih značajki za različite klasifikacijske zadatke [18, 19, 21, 22].

Iako se klasičnom primjenom određenog klasifikatora nad skupom oblikovanih značajki dobiva slabija točnost modela u odnosu na korištenje konvolucijskih neuronskih mreža, interpretabilnost takvog pristupa je veća budući da je postupak dobivanja značajki dizajniran, a ne automatiziran. Upravo je interpretabilnost modela trenutno jedan od najvećih istraživačkih izazova u području dubokih neuronskih mreža. Ta činjenica motivirala je oblikovanje postupka treniranja konvolucijskih neuronskih mreža za različite probleme klasifikacije slikarskih djela predstavljenog u ovom radu. Osim želje za postizanjem što veće točnosti klasifikacije, razmotreno je i kako određene postavke mreže pridonose povećanju točnosti. U mnogim domenama primjene konvolucijskih neuronskih mreža za klasifikaciju slika pokazano je da se bolji rezultati postižu ako se modeli mreža koji su već predtrenirani za određeni zadatak treniraju postupkom finog podešavanja parametara (engl. fine-tuning) na novom skupu podataka za učenje, umjesto da se model mreže trenira od početka, tj. od stanja inicijalizacije slučajnim parametrima težina [54]. Metoda finog podešavanja parametara pokazala se najboljim pristupom za rješavanje različitih klasifikacijskih zadataka u domeni slikarskih djela [24, 25, 26]. U ovom radu dodatno poboljšanje klasifikacijske točnosti nastoji se postići istraživanjem utjecaja inicijalizacijskih postavki težina mreže, pri čemu svaka inicijalizacija predstavlja optimalno stanje mreže za neki domensko-specifični zadatak i određenu kolekciju slika. Na taj se način, osim optimizacije klasifikacijske točnosti mreže, analizira i potencijalna semantička korelacija između modela naučenih za klasifikaciju slikarskih djela i modela naučenih za druge zadatke i na drugačijim bazama slika.

Osim izazova razumijevanja procesa učenja konvolucijskih neuronskih mreža i njihove spo-

sobnosti učenja na kompleksnom skupu podataka kao što su slikarska djela, razmotreno je pitanje primjenjivosti naučenih modela za daljnje istraživačke i praktične primjene. Modeli naučeni za klasifikaciju slikarskih djela svoju praktičnu primjenu ponajprije nalaze u proširivanju pretraživačkih mogućnosti unutar kolekcija umjetničkih djela, tako da se osim pretrage ključnim riječima, kojim se pretražuju metapodaci, koriste vizualni upiti koji pretražuju same slike. Pojam vizualne sličnosti posebno je kompleksan u kontekstu umjetničkih djela budući da ona mogu biti slična na različitim razinama. U analizi primjenjivosti naučenih modela razmotreno je kako modeli naučeni za različite klasifikacijske zadatke mogu biti korišteni za određivanje mjera dvaju različitih vidova vizualne sličnosti: sličnost sadržaja (teme i motivi u slici) i sličnost stilskih obilježja (elementi djela kao što su tekstura, boje, kompozicija, itd.). Istraživanje i rezultati opisani u ovome poglavlju predstavljani su u [55].

## 3.2 Baze slika i klasifikacijski zadaci

U svrhu korištenja što većeg broja slikarskih djela i različitih klasifikacijskih zadataka, korištena su tri različita izvora za oblikovanje baza slikarskih djela i definiranje klasifikacijskih zadataka. Prvi izvor je WikiArt, trenutno najveća kolekcija digitaliziranih slikarskih djela dostupna na Internetu. WikiArt kolekcija sadrži veliki broj metapodataka kao što su informacije o autoru, stilu, žanru, tehnici, godini nastanka slike i sl. Kolekcija uključuje slikarska djela iz različitih stilskih razdoblja, s posebnim naglaskom na umjetnost devetnaestog i dvadesetog stoljeća. Zbog svoje veličine i detaljnih metapodataka, WikiArt kolekcija najčešći je izvor za oblikovanje baza u radovima posvećenim automatskoj klasifikaciji slikarskih djela te stoga je pogodna za usporedbu rezultata. Broj umjetničkih djela u kolekciji kontinuirano raste, a u trenutku dohvaćanja WikiArt kolekcije za potrebe ovog rada ukupan broj djela u kolekciji iznosio je 133220. Jedna od temeljnih kategorizacija umjetničkih djela unutar WikiArt kolekcije je kategorizacija na temelju žanra (engl. genre) u kojoj su uključene različite vrste likovnih djela kao što su slikarska djela, crteži, skulpture, ilustracije, grafiti, poster, itd. Međutim, pojam „žanra“ u likovnoj umjetnosti ima slojevitije značenje od onoga koje mu je dodijeljeno u organizaciji WikiArt kolekcije. Konkretno, u slikarstvu se kategorizacija na temelju žanra odnosi na razlikovanje slika s obzirom na motive, pa tako razlikujemo žanrovske kategorije kao što su portret, pejzaž, mrtva priroda, itd. Također, pojmom „žanr“ označava se i jedna žanrovska podkategorija, odnosno slike koje prikazuju prizore iz svakodnevnog života. Potrebno je istaknuti važnost pravilne upotrebe pojma „žanr“ budući da se u nekim radovima, koji se rukovode isključivo kategorizacijom na temelju WikiArt kolekcije, u kontekstu klasifikacije na temelju žanra uči automatski razlikovati kategorije kao što su ilustracije i skulpture [20], što je naravno manje zahtjevan problem od razlikovanje slikarskih djela na temelju motiva. Također, prije pojave WikiArt kolekcije, u mnogim ranijim radovima često su se terminološki pogrešno

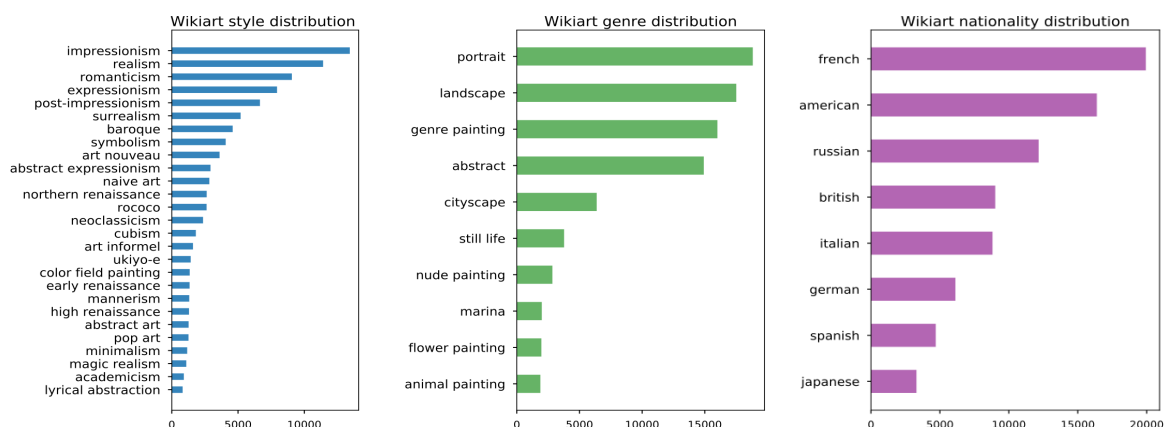
definirali klasifikacijski zadaci, pa se pod klasifikacijom na temelju žanra provodila klasifikacija na temelju stila [9, 10]. Sa svrhom oblikovanja klasifikacijskih zadataka koji su postojani s obzirom na vrstu likovnog djela i u skladu s kategorizacijama povijesti umjetnosti, iz WikiArt kolekcije korištena su samo digitalizirana slikarska djela. Na slici 3.1 dani su primjeri slika za deset različitih klasa odabranih za klasifikaciju prema žanru.



**Slika 3.1:** Primjeri slika iz deset različitih kategorija uključenih u podskup WikiArt baze za klasifikaciju prema žanru.

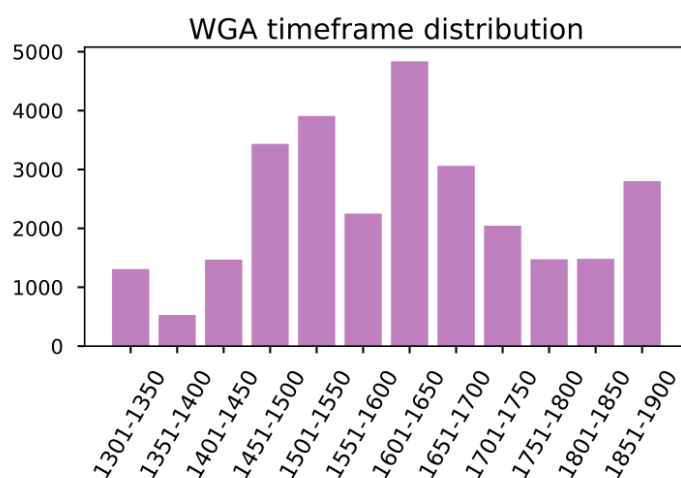
Ukupno su definirana četiri različita zadatka klasifikacije slikarskih djela iz WikiArt kolekcije koji uključuju klasifikaciju prema autoru, stilu, žanru i autorovoj nacionalnosti. Klasifikacija prema autoru, stilu i žanru predstavlja standardne klasifikacijske zadatke, dok se klasifikacija slikarskih djela prema nacionalnosti autora prvi puta uvodi u ovome radu. Uvođenjem takvog zadatka istražuje se mogućnost klasifikacije djela neovisno o žanru ili stilu, već s obzirom na specifičnosti područjem određenog kulturnog kruga.

Na temelju broja slikarskih djela dostupnih u pojedinim kategorijama, kao i prema broju i vrsti različitih klasa definiranih u prethodnim radovima [14], određen je podskup slika za svaki klasifikacijski zadatak. Za klasifikaciju slika prema autoru korišten je podskup od 23 slikara u kojem je svaki autor zastupljen s najmanje 500 slika. Za klasifikaciju prema stilu korišten je podskup od 27 različitih stilova u kojemu se za svaki stil nalazi barem 800 slika, dok je za klasifikaciju prema žanru korišten podskup od 10 različitih klasa, u kojemu svaka klasa ima najmanje 1880 slika, a za klasifikaciju prema nacionalnosti korišten je podskup od 8 različitih klasa s najmanje 3200 slika po klasi. Prikaz distribucija slika po klasama baze za klasifikaciju na temelju stila, žanra i nacionalnosti dan je na slici 3.2



**Slika 3.2:** Distribucija slika u podskupu WikiArt baze za klasifikaciju stila (lijevo), za klasifikaciju žanra (sredina) i klasifikaciju slika prema nacionalnosti autora (desno).

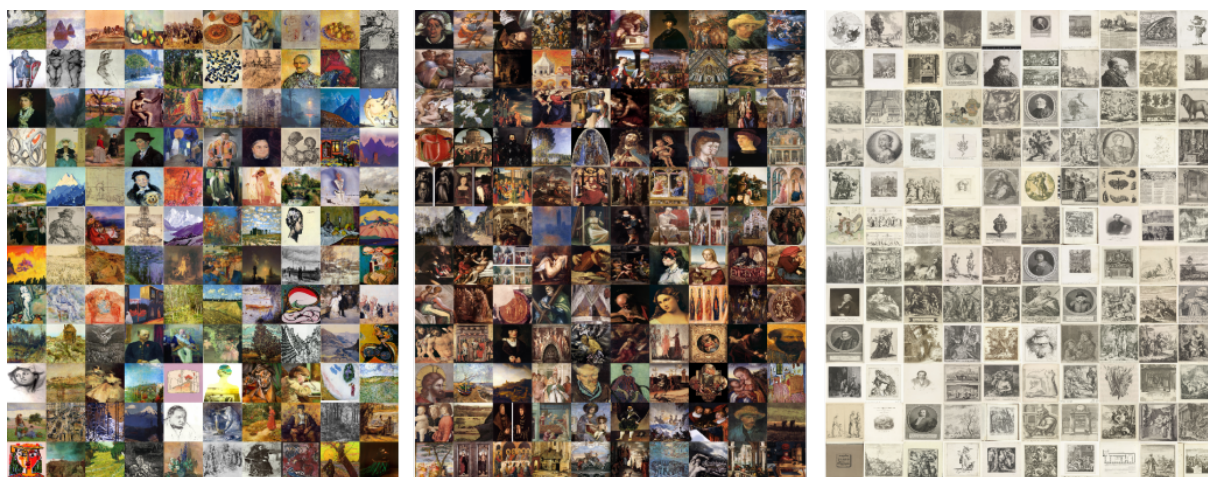
Osim WikiArt kolekcije, drugi izvor za oblikovanje baza je Web Gallery of Art (WGA). Ta kolekcija nije u radovima posvećenim automatskoj klasifikaciji korištena tako često kao WikiArt kolekcija, no zanimljiva je jer ima drugačiju distribuciju slikarskih djela s obzirom na povijesna razdoblja. U kolekciji se nalazi ukupno 28952 slikarskih djela, a zastupljen je velik broj djela iz razdoblja renesanse i baroka. Slično kao u WikiArt kolekciji, svakoj slici pripada oznaka autora, žanra, škole (definirane s obzirom na nacionalnost autora, odnosno pripadnosti autora nekom geografski određenom umjetničkom krugu) te vremenskog razdoblja nastanka djela. S obzirom na te oznake, definirana su četiri različita klasifikacijska zadatka: autor, žanr, vremensko razdoblje i nacionalnost. Klasifikacija prema vremenskom razdoblju može se smatrati sličnom klasifikaciji prema stilu, budući da su određena stilska obilježja karakteristična za određeni povijesni kontekst. Međutim, budući da su djela u WGA kolekciji s obzirom na godinu nastanka označena pripadajućim pedesetogodišnjim vremenskim razdobljem, moguća su preklapanja različitih stilova unutar određenog razdoblja. Na slici 3.3 prikazana je distribucija slika u WGA kolekciji s obzirom na vremensko razdoblje nastanka djela.



**Slika 3.3:** Distribucija slika u WGA kolekciji s obzirom na vremensko razdoblje nastanka djela.

Podskup slika za klasifikaciju djela prema vremenskom razdoblju nastanka sastoji se od 12 različitih klasa, pri čemu je svaka klasa zastupljena s najmanje 500 slika. Za klasifikaciju prema žanru definirano je 6 različitih klasa s više od 1000 slika po klasi, za klasifikaciju prema autoru 23 klase s više od 170 slika i za klasifikaciju prema nacionalnosti 8 klasa s više od 500 slika.

WikiArt i WGA kolekcije sadrže digitalizirana slikarska djela u boji, odnosno djela rađena ponajviše u tehnici ulja na platnu ili akvarela. Sa svrhom testiranja predložene klasifikacijske metode na drugačijem tipu umjetničkih slika, kao treći izvor korištena je kolekcija digitaliziranih grafika iz Rijksmuseuma. Radi se o tzv. TICC (Tilburg center for Cognition and Communication) bazi koja je prvi puta predstavljena u radu van Noorda [56]. Osim što se radi o akromatskim slikama, osobitost ove kolekcije je u tome što su sve slike visoke rezolucije i dobivene iz istog izvora, dok su slike u WikiArt i WGA kolekcijama prikupljene iz različitih izvora i različite su rezolucijske kvalitete. TICC baza sadrži 58630 digitaliziranih grafika koje su djela 210 različitih autora, pri čemu je svaki autor zastupljen s najmanje 96 djela.



**Slika 3.4:** Primjeri slika iz triju različitih izvora: WikiArt, WGA i TICC.

Kako bi se ilustrirala osobitost svakog od triju navedenih izvora, na slici 3.4 prikazani su slučajno odabrani primjeri slika iz svih triju izvora. U tablici 3.1 dan je popis klasifikacijskih zadataka, broja klasa za svaki zadatak i broja slika u svakom podskupu. Detaljan popis svih klasa iz svih podskupova dan je u Prilogu A.

Za svaki podskup slika određenog klasifikacijskog zadatka 70% ukupnog broja slika iz tog podskupa koristi se za treniranje modela, 10% za validaciju i 20% za testiranje. Također, sve slike preprocesiraju se tako da im se smanji dimenzionalnost na  $256 \times 256$  piksela.



**Tablica 3.1:** Broj slika i klasa za različite klasifikacijske zadatke i izvore baza slika.

zadatak	izvor	broj klasa	broj slika
autor	TICC	210	58,630
	WikiArt	23	20,320
	WGA	23	5,711
žanr	WikiArt	10	86,087
	WGA	6	26,661
stil	WikiArt	27	96,014
razdoblje	WGA	12	28,605
nacionalnost	WikiArt	8	80,428
	WGA	8	27,460

### 3.3 Fino podešavanje parametara konvolucijske neuronske mreže

Metoda finog podešavanja parametara konvolucijske neuronske mreže (engl. fine-tuning) temelji se na prijenosu znanja (engl. transfer learning) stečenog postupkom učenja mreže na jednom zadatku na učenje novog zadatka. Prilikom uobičajenog postupka učenja mreže, inicijalne vrijednost parametara, tj. vrijednosti težina među neuronima, postavljaju se na slučajne vrijednosti, te se zatim podešavaju dok ne bude zadovoljen kriterij zaustavljanja treniranja mreže. Budući da je broj parametara duboke neuronske mreže iznimno velik, potreban je dovoljno velik skup podataka kako bi parametri mreže optimalno konvergirali s obzirom na ciljani zadatak učenja. Pokazano je da parametri početnih konvolucijskih slojeva konvergiraju tako da nauče raspoznavati slične reprezentacije slike neovisno o klasifikacijskom zadatku za koji se mreža trenira. Radi o tzv. općenitim značajkama, odnosno o rubovima i jednostavnim oblicima koji su zajednički većini slika. S druge strane, reprezentacije naučene u zadnjem sloju mreže vrlo su specifične za određeni skup podataka i zadatak za koji se mreža trenira. Činjenica da početni slojevi uče općenite značajke koje su zajedničke različitim domenama, dovela je do zaključka da prilikom treniranja mreže za neki novi zadatak nije potrebno ponovno učiti parametre tih slojeva već se mogu iskoristiti parametri prethodno naučene mreže, dok se novi parametri uče samo za zadnji sloj mreže. Na toj ideji temelji se metoda finog podešavanja parametara mreže. Prilikom korištenja predtrenirane mreže, odnosno mreže koja je već naučena za određeni zadatak, za učenje nekog novog zadatka potrebno je zamijeniti zadnji sloj u kojem broj neurona odgovara broju klasa originalnog klasifikacijskog zadatka sa slojem u kojem je broj neurona jednak broju klasa u novom zadatku. S obzirom da se broj parametara koje je potrebno naučiti smanjuje, potreban je i manji skup podataka za učenje. Stoga se korištenjem metode finog po-



dešavanja parametara konvolucijske neuronske mreže mogu koristiti i za probleme kod kojih je skup podataka za učenje nedovoljno velik za treniranje cjelovite mreže od početka.

Osim što se metodom finog podešavanja parametara omogućuje korištenje konvolucijskih neuronskih mreža na manjem skupu podataka, pokazano je da se tom metodom postižu bolji rezultati čak i ako je skup podataka dovoljno velik za učenje svih parametara mreže [54]. Ako prilikom učenja mreže za neki novi zadatak i na novom skupu podataka inicijalne vrijednosti parametara mreže odgovaraju stanju mreže koja je naučena za neki drugi zadatak, postižu se bolji rezultati nego ako su inicijalni vrijednosti parametara postavljene na slučajno odabrane vrijednosti. Metoda finog podešavanja stoga se temelji na prijenosu znanja između predtrenirane bazne mreže i ciljne mreže. Pod baznom mrežom podrazumijeva mreža čiji su parametri optimizirani za rješavanje određenog domenski specifičnog problema, tj. za određeni klasifikacijski zadatak i određeni skup podataka. Ciljnom mrežom smatra se mreža čiji početni položaj odgovara baznoj mreži, a uči se za neki drugi domenski specifični problem. U okviru ovog rada ispitan je utjecaj korištenja različitih domenski specifičnih baznih mreža za učenje ciljnih mreža na klasifikacijskim zadacima razlikovanja slikarskih djela, kao i utjecaj razine dosega finog podešavanja parametara u slojevima mreža.

### 3.3.1 Razina dosega finog podešavanja parametara mreže

S obzirom na dostupnu veličinu skupa podataka za treniranje ciljne mreže na novom zadatku, potrebno je ispitati optimalan doseg finog podešavanja parametara mreže. Pod dosegom se podrazumijeva koliko slojeva bazne mreže će se modificirati učenjem na novom zadatku ciljne mreže. Metoda finog podešavanja parametara mreže sastoji se u tome da se kopiraju vrijednosti parametara slojeva bazne mreže, pa se zatim vrijednosti parametara određenog broja slojeva zamrzava dok se parametri preostalih slojeva ažuriraju propagacijom pogreške unatrag. S obzirom na broj slojeva koji se modificiraju u odnosu na broj zamrznutih slojeva, u ovome radu testirano je pet različitih scenarija finog podešavanja mreže:

- *all* - težine svih slojeva se modificiraju;
- *only last* - težine samo zadnjeg potpuno povezanog sloja (fc8) se modificiraju;
- *only last 3* - težine zadnja tri potpuno povezana sloja (fc6, fc7, fc8) se modificiraju;
- *skip first 2* - težine prvog i drugog konvolucijskog sloja (conv1, conv2) se zamrzavaju;
- *skip first* - težine prvoj konvolucijskog sloja (conv1) se zamrzavaju.

Ako je ciljni skup podataka za učenje mali, a broj parametara koji se modificiraju velik, fino podešavanje parametara može rezultirati prenaučenošću mreže (engl. overfitting). S druge strane, ako je ciljni skup podataka dovoljno velik u odnosu na broj parametara koji se modificiraju, neće doći do prenaučivosti mreže, no potrebno je odrediti potreban odnos zamrznutih slojeva i slojeva čiji se parametri modificiraju za postizanje optimalne naučenosti mreže.

### 3.3.2 Domenski određene inicijalizacije parametara mreže

U kontekstu ovoga rada, pod pojmom domene podrazumijeva se skup podataka i klasifikacijski zadatak za koji je konvolucijska neuronska mreža naučena. Primjerice, klasifikacija slika prema objektima sadržanima u slici predstavlja zadatak koji može biti naučen korištenjem različitih skupova ulaznih podataka. Budući da različiti skupovi najčešće sadrže različite klase slika, domena je određena ne samo zadatkom za koji se mreža uči, već i skupom podataka nad kojim se uči. Poznato je da uspješnost prijenosa znanja s bazne na ciljnu mrežu metodom finog podešavanja parametara ovisi o udaljenosti domena bazne i ciljne mreže. Primjerice, domena određena klasifikacijskim zadatkom raspoznavanja fotografija različitih pejzaža bliža je domeni raspoznavanja objekata u fotografijama nego domeni klasifikacije rendgenskih snimaka. Međutim, Yosinski et al. [54] pokazali su da ako se kao inicijalno stanje koriste parametri bazne mreže čija je domena vrlo udaljena od domene ciljne mreže, još uvijek se postiže bolja naučenost ciljne mreže nego postavljanjem parametara mreže na slučajne vrijednosti. Radi boljeg razumijevanja utjecaja inicijalizacije prilikom finog podešavanja parametara mreže, u ovom radu je analiziran utjecaj različitih domenski određenih inicijalizacija za različite zadatke klasifikacije slikarskih djela.

S ciljem istraživanja utjecaja domenski specifične inicijalizacije na učinkovitost finog podešavanja parametara mreže, analizirano je pet različitih baznih mreža koje imaju istu arhitekturu, no trenirane su za različite zadatke i na različitim skupovima podataka:

- CaffeNet je referentni BVLC model mreže [57] koji se temelji na AlexNet arhitekturi i treniran je podskupu ImageNet baze korištenom u ILSVRC-2012 natjecanju [58], a sastoji se od 1.2 milijuna slika i 1000 klasifikacijskih kategorija pri čemu je zadatak bio raspoznavanje slika na temelju objekata u njima.
- Hybrid-CNN mreža [59] je mreža trenirana za klasifikaciju slika na temelju objekata i scena, pri čemu se skup podataka za učenje sastojao od 3.5 milijuna slika iz 1183 različitih kategorija, a dobiven je kombiniranjem slika iz ImageNet i Places baze.
- MemNet mreža [60] dobivena je finim podešavanjem Hybrid-CNN bazne mreže korištenjem LaMem baze podataka koja se sastoji od 60 000 slika označenih faktorom pamtljivosti slike (engl. memorability score) koji predstavljaju vrijednost između 0 i 1, pri čemu 0 označava slabu pamtljivost, a 1 jaku pamtljivost slike. Faktori pamtljivosti slika prikupljeni su eksperimentalnom memorijskom igrom kojom se tijekom prikaza sekvence slika mjeri sposobnost ispitanika da zapamti pojedinu sliku. Budući da je faktor pamtljivosti realan broj iz intervala  $[0,1]$ , prilikom finog podešavanja mreže posljednji klasifikacijski sloj bazne mreže koji računa multinomijalnu logističku funkciju softmax ulaza, zamijenjen je regresijskim slojem koji računa srednji kvadratni gubitak očekivanog i dobivenog izlaza.
- SentiNet mreža [61] dobivena je finim podešavanjem bazne CaffeNet mreže za predikciju

emotivnog doživljaja slike korištenjem DeepSent baze slika [62], koja se sastoji od 1269 slika prikupljenih s Twittera i ručno označenih s obzirom na to izazivaju li slike pozitivan ili negativan osjećaj kod ispitanika.

- Flickr mreža [18] trenirana je za klasifikaciju digitalnih fotografija na temelju različitih vizualnih stilova korištenjem Flickr Style baze slika koja se sastoji od 80 000 fotografija označenih s 20 različitih kategorija vizualnog stila. U kontekstu učenja ove mreže, vizualni stil predstavlja različite stilske koncepte kao što su kompozicija, ugođaj i atmosfera slike.

Koncepti koji se uče unutar ovih pet različitih mreža izražavaju različite razine razumijevanja slike, od jednostavnog zadatka prepoznavanja predmeta do apstraktnih izazova predviđanja pamtljivosti ili osjećaja koje slika izaziva. Korištenjem tih pet različitih mreža nastoji se prije svega pronaći najbolje inicijalno stanje mreže za proces finog podešavanja parametara sa svrhom učenja različitih zadataka klasifikacije slikarskih djela. Osim toga, ovim pristupom nastoji se istražiti utječu li domenski specifične inicijalizacije na novi zadatak učenja tako da ishod odražava intrinzičnu povezanosti između koncepta za koji je naučena bazna mreža i koncepta umjetničkog stila, žanra i sl., za koje se uči ciljna mreža.

### 3.3.3 Postavke treniranja mreže

Osim postizanja najbolje moguće klasifikacijske točnosti, cilj je ovoga rada i istražiti utjecaj domensko specifične inicijalizacije mreže te je stoga izbor arhitekture baznih mreža sužen na korištenje AlexNet mreža opisane u poglavlju 2.1.2. Iako trenutno postoje mnogo dublje arhitekture konvolucijskih neuronskih mreža poput ResNeta ili GoogleLeNeta, kojima se postižu nešto bolji rezultati za određene klasifikacijske probleme, dostupni predtrenirani modeli takvih mreža najčešće su naučeni raspoznavati objekte na ImageNet bazi slika i ne pokrivaju raznolike domene. Budući da je AlexNet jedna od najstarijih i najviše korištenih arhitektura, dostupne su baze različitih predtreniranih modela\*.

Implementacija postupka učenja mreža provedena je korištenjem programskog okvira za duboko učenje Caffe<sup>†</sup>. Tijekom treniranja mreže korištene su jednostavne tehnike povećavanja skupa podataka kao što su horizontalno zrcaljenje i nasumično izrezivanje ulaznih slika (engl. random cropping). Tijekom finog podešavanja parametara mreže, korištena je metoda stohastičkog gradijentnog spusta s podjelom skupa za učenje na manje slučajno odabrane podskupove (engl. minibatch gradient descent) veličine 256 slika i mometumom 0.9. Korištena je L2 regularizacija s parametrom gubitka težina iznosa 0.0005 te isključivanje neurona (engl. dropout) s vjerojatnošću 0.5. Ispitan je utjecaj promjena postavki ovih hiper-parametara kori-

---

\*Caffe Model Zoo <https://github.com/BVLC/caffe/wiki/Model-Zoo>

<sup>†</sup>Caffe Deep Learning Framework <http://caffe.berkeleyvision.org>

štenjem različitih optimizacijskih strategija implementiranih u platformi SigOpt<sup>‡</sup> te je pokazano da promjene vrijednosti ovih hiper-parametara ne utječu značajno na poboljšanje klasifikacijske točnosti modela. Također, provedeno je iscrpno eksperimentalno ispitivanje sa svrhom pronalaska optimalnog broja epoha za treniranje pojedinog klasifikacijskog zadatka, kao i inicijalna vrijednost stope učenja te frekvenciju i faktor njene redukcije tijekom učenja.

### 3.4 Ovisnost rezultata klasifikacije o postavkama finog podešavanja parametara mreže

U ovom poglavlju prikazano je kako na ishod rezultata klasifikacije slikarskih djela utječu različite postavke finog podešavanja parametara konvolucijske neuronske mreže opisane u prethodnome poglavlju.

#### 3.4.1 Utjecaj domenski određene inicijalizacije parametara mreže

Kako bi se ispitao utjecaj različitih inicijalizacija, sve bazne mreže trenirane su pod istim uvjetima: provedba učenja mreže od 100 epoha, s konstantom stopom učenja od  $10^{-4}$  i mijenjanjem parametara svih slojeva (all scenarij u razinama dosega finog podešavanja mreže). Rezultati klasifikacijske točnosti za svaki klasifikacijskim zadatkom određeni skup podataka i svaku domenski određenu baznu mrežu dani su u Tablici 3.2.

Iz rezultata je vidljivo da se najbolja točnost za skoro sve zadatke postiže inicijalizacijom parametrima Hybrid-CNN mreže. Na slici 3.5 prikazane su krivulje validacijske točnosti različito inicijaliziranih modela za zadatak klasifikacije slikarskih djela prema autoru iz WikiArt baze prilikom ponovnog treniranja svih slojeva (lijeva slika) i samo zadnjeg sloja (desna slika), dok su na slici 3.6 prikazane krivulje validacijske točnosti različito inicijaliziranih modela za zadatak klasifikacije slikarskih djela prema autoru iz WGA baze.

Budući da je mreža Hybrid-CNN trenirana na vrlo velikom skupu podataka koji objedinjuje i slike različitih objekata (iz ImageNet baze) i slike različitih prostora (iz Places baze), pokazuje se kao bolji izbor bazne mreže za inicijalizaciju prilikom finog podešavanja parametara od CaffeNet mreže koja je trenirana samo za domenu razlikovanja različitih objekata u slici. Može se pretpostaviti da utjecaj širenja domene naučenosti mreže, konkretno od naučenosti na raspoznavanje objekata do naučenosti za raspoznavanje objekata i prostora, rezultira postavkama parametara mreže koji predstavljaju pogodnije inicijalno stanje za učenje različitih zadataka klasifikacije slikarskih djela.

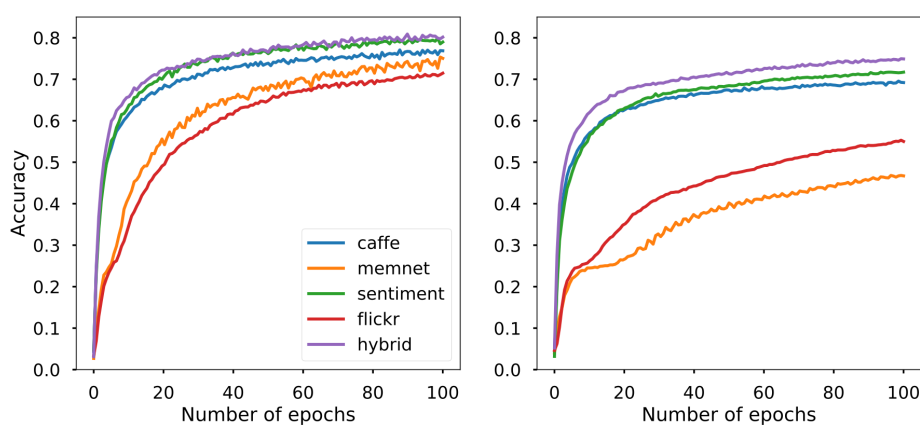
Zanimljiv rezultat predstavlja vrlo visoka klasifikacijska točnost postignuta finim podešavanjem SentiNet mreže, pogotovo imajući na umu da je SentiNet mreža dobivena finim pode-

---

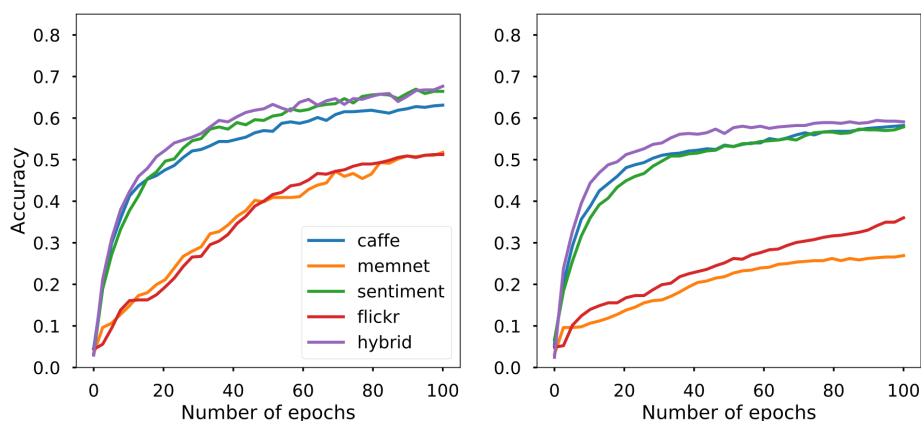
<sup>‡</sup>SigOpt <https://sigopt.com>

**Tablica 3.2:** Usporedba klasifikacijske točnosti dobivene na testnom skupu baza različitih zadataka s obzirom na različite domenski specifične inicijalizacije mreže prilikom finog podešavanja parametara.

Zadatak	Klasifikacijska točnost					Varijance točnosti
	hybrid	memnet	sentiment	caffe	flickr	
TICC_autor	<b>0.762</b>	0.666	0.738	0.719	0.678	$12.9 \times 10^{-4}$
WikiArt_autor	<b>0.791</b>	0.725	0.787	0.763	0.714	$9.92 \times 10^{-4}$
WikiArt_stil	<b>0.563</b>	0.526	0.558	0.542	0.507	$4.27 \times 10^{-4}$
WikiArt_žanr	<b>0.776</b>	0.759	0.774	0.772	0.755	$0.72 \times 10^{-4}$
WikiArt_nacionalnost	<b>0.583</b>	0.534	0.571	0.551	0.513	$6.31 \times 10^{-4}$
WGA_autor	<b>0.696</b>	0.551	0.686	0.655	0.569	$36.2 \times 10^{-4}$
WGA_razdoblje	<b>0.527</b>	0.482	0.526	0.506	0.469	$5.24 \times 10^{-4}$
WGA_žanr	0.796	0.779	<b>0.801</b>	0.787	0.765	$1.59 \times 10^{-4}$
WGA_nacionalnost	<b>0.656</b>	0.612	0.655	0.635	0.603	$4.70 \times 10^{-4}$

**Slika 3.5:** Krivulje učenja dobivene pri finom podešavanju parametara svih slojeva (lijevo) i samo zadnjeg sloja (desno) različito inicijaliziranih mreža na skupu za validaciju klasifikacije stila WikiArt baze.

šavanjem CaffeNet mreže na relativno malom skupu podataka (1269 slika). Iz toga se može zaključiti da su značajke koje mreža uči kao relevantne za predviđanje izaziva li određena slika pozitivnu ili negativnu emotivnu reakciju kod promatrača također značajne u raspoznavanju različitih kategorija slikarskih djela. S druge strane, niska klasifikacijska točnost dobivena korištenjem MemNet mreže pokazuje da značajke koje su bitne za predviđanje pamtljivosti mreže ne predstavljaju osobito povoljno inicijalno stanje za fino podešavanje mreže za različite zadatke klasifikacije slikarskih djela. Također, korištenjem Flickr mreže postižu se relativno slabe klasifikacijske točnosti u odnosu na druge bazne mreže. Iako je Flickr mreža naučena na zadatku



**Slika 3.6:** Krivulje učenja dobivene pri finom podešavanju parametara svih slojeva (lijevo) i samo zadnjeg sloja (desno) različito inicijaliziranih mreža na skupu za validaciju klasifikacije stila WGA baze.

raspoznavanja različitih stilova fotografija te bi se prema tome mogla pretpostaviti kao povoljna bazna mreža za učenje klasifikacije stila slikarskih djela, njena slaba performansa može biti posljedica nesklada između definicije stila u određivanju kategorija u Flickr bazi slika i definicije stila u kontekstu povijesti umjetnosti. Također, iako je trenirana na relativno velikom skupu podataka od 80 000 slika, Flickr mreža postiže relativno slabu točnost klasifikacije od 39% na originalnom zadatku koji uključuje raspoznavanje 20 stilskih kategorija, što može biti posljedica velikog preklapanja vizualnih značajki među slikama iz različitih kategorija.

Vrijednosti varijanca klasifikacijske točnosti pojedinog zadatka s obzirom na različite inicijalizacije prikazane u Tablici 3.2, ukazuju da je varijanca visoka za zadatak klasifikacije prema autoru u svim trima bazama, dok je varijanca najniža za zadatak klasifikacije prema žanru i za WikiArt i za WGA bazu. Iz toga se može zaključiti da inicijalizacija ima veći utjecaj pri finom podešavanju mreže za zadatke koji imaju velik broj klasa i manji broj primjera po klasi (kao što je zadatak raspoznavanja autora), nego za zadatke s malim brojem klasa i većim brojem primjera po klasi.

### 3.4.2 Utjecaj razine doseg a finog podešavanja parametara mreže

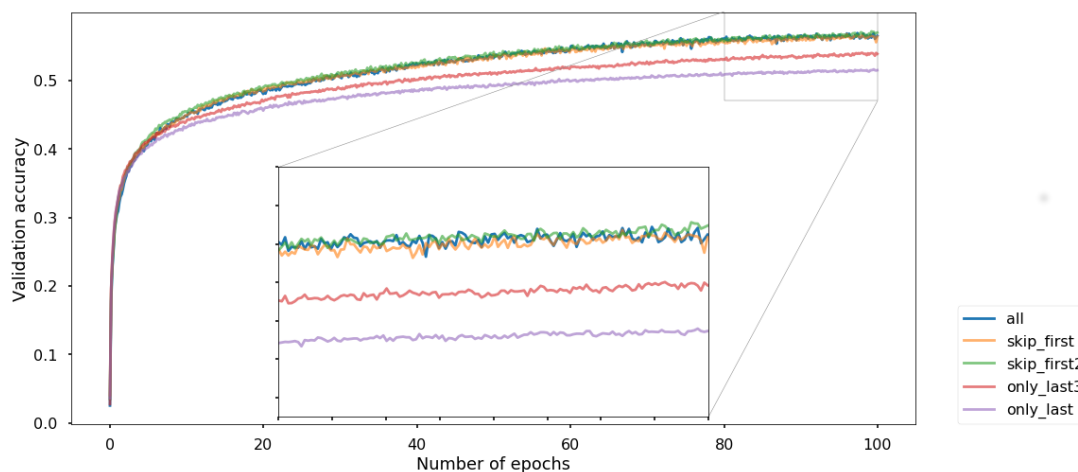
S ciljem utvrđivanja optimalnog odnosa zamrznutih slojeva i slojeva čiji se parametri modificiraju prilikom učenja bazne mreže na skupovima podataka za klasifikaciju slikarskih djela, ispitano je pet različitih scenarija učenja opisanih u poglavlju 3.3.1. Kako bi se utvrdio najbolji scenarij, sve ostale postavke učenja se fiksiraju. Kao bazna mreža koristi se HybridNet mreža, s obzirom na to da se pokazala kao najboljom domenski specifičnom inicijalizacijom, te se svaki model trenira 100 epoha s konstantnom stopom učenja. Rezultati ispitivanja različitih scenarija prikazani su u Tablici 3.3.

Na temelju točnosti klasifikacije može se zaključiti da je najbolji scenarij za većinu zadataka

**Tablica 3.3:** Usporedba klasifikacijske točnosti dobivene na testnom skupu baza različitih zadataka s obzirom na različite scenarije finog podešavanja parametara mreže.

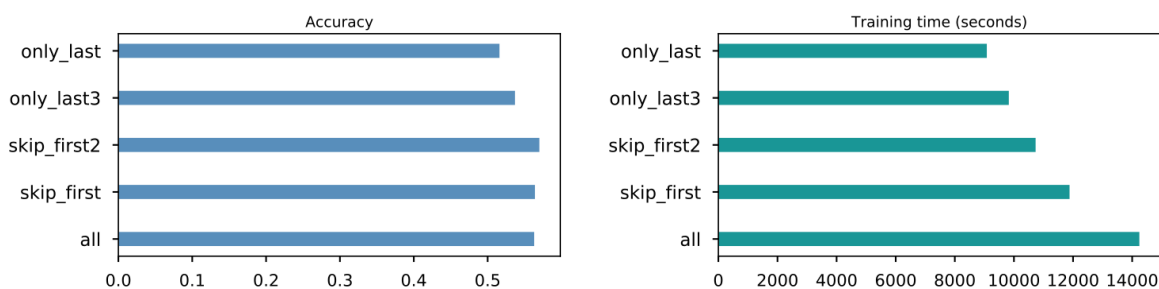
scenarij	Točnost klasifikacije								
	TICC	WikiArt				WGA			
	autor	autor	stil	žanr	nacionalnosti	autor	razdoblje	žanr	nacionalnosti
all	0.762	0.791	0.563	0.776	0.584	0.696	0.526	<b>0.796</b>	<b>0.656</b>
<i>skip_first</i>	<b>0.767</b>	<b>0.798</b>	0.564	0.774	<b>0.585</b>	<b>0.704</b>	<b>0.537</b>	0.792	0.651
<i>skip_first2</i>	0.765	0.795	<b>0.570</b>	<b>0.777</b>	0.583	0.689	0.524	0.790	0.652
<i>only_last3</i>	0.668	0.762	0.537	0.762	0.554	0.665	0.505	0.791	0.634
<i>only_last</i>	0.583	0.740	0.516	0.754	0.532	0.646	0.488	0.772	0.619

*skip first*, odnosno zamrzavanje parametara samo prvog sloja i novo učenje parametara svih ostalih slojeva. Također, vrlo dobri klasifikacijski rezultati postizu se i kada se zamrznu prva dva sloja bazne mreže. Time se pokazuje da su značajke naučene u prvom i drugom sloju Hybrid-Net mreže dovoljno općenite i relevantne za zadatke klasifikacije slikarskih djela. Na slici 3.7 prikazane su validacijske krivulje točnosti za klasifikaciju stila iz WikiArt baze iz kojih je vidljivo da se postižu vrlo slični rezultati kada se modificiraju svi slojevi i kada se zamrzavaju prvi i drugi sloj bazne mreže, dok se najslabiji rezultati postižu učenjem samo parametara zadnjeg sloja.



**Slika 3.7:** Krivulje učenja dobivene različitim scenarijima finog podešavanja mreže na skupu za validaciju klasifikacije stila WikiArt baze.

Usporedbom klasifikacijske točnosti (Slika 3.8, lijevo) i vremena potrebnog za učenje 100 epoha (Slika 3.8, desno) prilikom finog podešavanja mreže za zadatak klasifikacije stila WikiArt baze, može se zaključiti da se zamrzavanjem prva dva sloja dobiva najbolji omjer točnosti i utrošenog vremena treniranja mreže.



**Slika 3.8:** Klasifikacijska točnost (lijevo) i vrijeme potrebno za 100 epoha (desno) finog podešavanja parametara mreže za klasifikaciju stila WikiArt baze (lijevo) za različite scenarije.

## 3.5 Rezultati automatske klasifikacije slikarskih djela

### 3.5.1 Pregled i usporedba klasifikacijskih rezultata

Sa svrhom usporedbe utjecaja različitih postavki finog podešavanja parametara mreže svi ostali hiper-parametri mreže fiksirani su prilikom treniranja modela za različite zadatke i sve mreže su trenirane 100 epoha s konstantnom stopom učenja. Međutim, budući da se težina, kao i veličina skupa za učenje, pojedinih klasifikacijskih zadataka razlikuju, provedeno je iscrpno eksperimentalno ispitivanje kako bi se utvrdile optimalne postavke treniranja za postizanje maksimalne klasifikacijske točnosti svakog pojedinog zadatka. Za većinu zadataka maksimalna klasifikacijska točnosti postignuta je treniranjem mreže velik broj epoha s konstantnom stopom učenja  $10^{-4}$ . S druge strane, pokazuje se da ako se stopa učenja postepeno smanjuje tijekom treniranja, potreban je mnogo manji broj epoha za konvergenciju modela i dostizanje relativno visoke klasifikacijske točnosti. Primjerice, ako model za zadatak klasifikacije stila WikiArt baze treniramo 20 epoha, počevši sa stopom učenja  $10^{-3}$  i smanjivanjem stope učenja za faktor 10 svakih 5 epoha, postiže se klasifikacijska točnost od 53.02%, što je gubitak od samo 3% u odnosu na klasifikacijsku točnost postignutu treniranjem mreže 100 epoha s konstantnom stopom učenja  $10^{-4}$ . Konačni rezultati za svaki klasifikacijski zadatak, kao i usporedba s rezultatima dobivenim u drugim radovima, dani su u tablici 3.4.

U tablici 3.4, osim prikaza klasifikacijskih točnosti postignutih za svaki zadatak, također je dana usporedba s rezultatima dobivenim u pet drugih radova koji se ističu kao posebno relevantni za komparativnu analizu. Saleh i Elgamel [14] inicirali su korištenje WikiArt baze za formiranje skupa za učenje različitih zadataka klasifikacije slikarskih djela te odredili broj klasa za pojedine zadatke prema kojem su i u ovom radu odabrani podskupovi podataka. U njihovom radu ispitano je kako različite vrste značajki slike doprinose klasifikacijskoj točnosti. Najbolji rezultat postignut je fuzijom različitih značajki, uključujući i značajke ekstrahirane iz zadnjeg sloja predtrenirane konvolucijske neuronske mreže. Tan et al. [25] su koristeći metodu finog podešavanja parametara konvolucijske neuronske mreže temeljene na CaffeNet arhitekturi i pred-



**Tablica 3.4:** Usporedba rezultata za sve klasifikacijske zadatke i baze podataka.

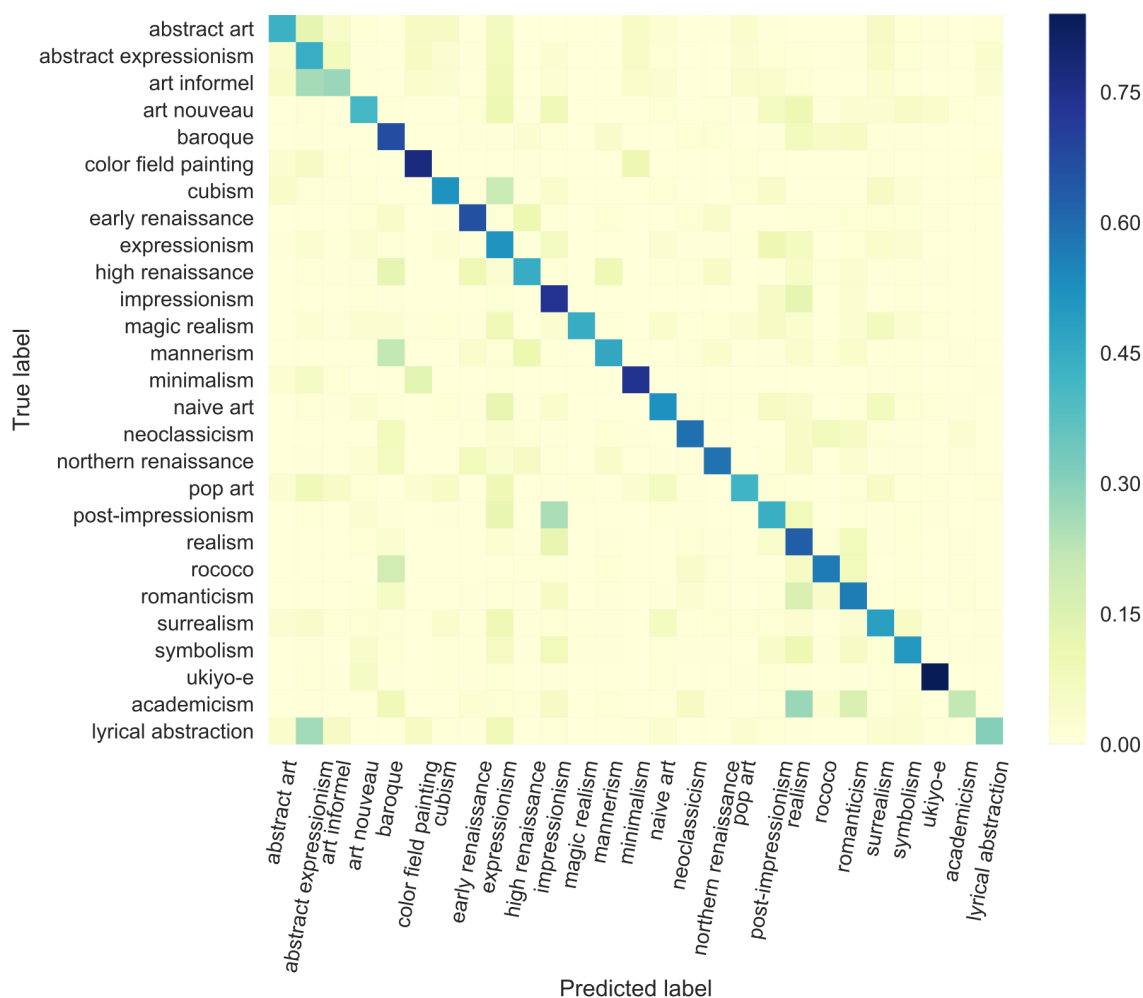
Referenca	Metoda	Baza podataka	Stil/Razdoblje		Žanr		Autor		Nacionalnost	
			broj klasa	točnost (%)	broj klasa	točnost (%)	broj klasa	točnost (%)	broj klasa	točnost (%)
	CNN fino podešavanje (CaffeNet)	TICC	-	-	-	-	210	80.42 (80.26 F-score)	-	-
		WGA	12	53.75	6	80.1	23	70.42	8	65.20
		WikiArt	27	56.43	10	77.6	23	81.94	8	58.35
B. Saleh et al.	fuzija značajki	WikiArt	27	45.97	10	60.28	23	68.25	-	-
Tan et al.	CNN fino podešavanje (AlexNet)	WikiArt	27	54.50	10	74.14	23	76.11	-	-
Hentschel et al.	CNN fino podešavanje (CaffeNet)	WikiArt	22	55.9 (MAP)	-	-	-	-	-	-
Lecoutre et al.	CNN fino podešavanje (ResNet50)	WikiArt	25	62.8	-	-	-	-	-	-
Noord et. al	višerezolucijski CNN (All-CNN)	TICC	-	-	-	-	210	77.01 (F-score)	-	-

treniranoj na ImageNet bazi postigli visoku klasifikacijsku točnost na zadacima klasifikacije slikarskih djela WikiArt baze i pokazali da se tim pristupom dobivaju bolji rezultati klasifikacijske točnosti nego kada se koriste samo značajke ekstrahirane iz predtrenirane konvolucijske mreže. Na su sličan način Hentschel et al. [24] pokazali da se finim podešavanjem parametara mreže postižu bolji rezultat klasifikacije slikarskih djela prema stilu nego korištenjem drugih značajki slike.

Rezultati pokazuju da se metodom predstavljenom u ovom radu postižu trenutno najbolji klasifikacijski rezultati za većinu zadataka klasifikacije slikarskih djela. Poboljšanje postignuto u odnosu na rezultate dobivene u [25] pokazuje da se promjenom domenski specifičnih inicijalizacija mreže i dosega treniranje mreže može postići dodatno značajno poboljšanje klasifikacijske točnosti pri finom podešavanju parametara mreže iste arhitekture. S druge strane, iako se odnose na manji broj klasa različitih stilova (25 umjesto 27), rezultati u radu Lecoutrea et al. [26] dobiveni finim podešavanjem parametara ResNet50 mreže pretrenirane na ImageNet bazi, ukazuju na to da se korištenjem mreže dublje arhitekture može postići dodatno poboljšanje klasifikacijske točnosti. Osim usporedbe s drugim izvedbama metode finog podešavanja, također se pokazuje da rezultati dobiveni u ovom radu predstavljaju poboljšanje klasifikacijske točnosti raspoznavanja autora u TICC bazi dobivene metodom višerezolucijskog treniranja konvolucijske neuronske mreže koju su predstavili van Noor i Postma [56].

### 3.5.2 Klasifikacija slikarskih djela prema stilu, žanru, autoru i nacionalnosti

Osim uvidom u sveukupnu klasifikacijsku točnost pojedinog zadatka, bolje razumijevanje ponašanja modela, kao i specifične problematike svakog zadatka, postiže se analizom klasifikacijske točnosti svake pojedine klase i najčešćih tipova pogreški. Na slici 3.9 prikazana je matrica grešaka (engl. confusion matrix) koja pokazuje omjer stvarnih primjera određene klase naspram



Slika 3.9: Matrica grešaka modela naučenog za raspoznavanje stilskih pravaca WikiArt baze.

broja primjera koji su u tu klasu svrstani modelom naučenog za zadatak klasifikacije stilskih pravaca WikiArt baze.

Iz navedenog prikaza vidljivo je da se najbolji rezultat klasifikacijske točnosti postiže za klasu Ukiyo-e (84%). Ukiyo je naziv za japansku školu slikanja koja je bila aktivna od 17. do 19. stoljeća i budući da predstavlja vrlo distinktivan stilski izričaj u odnosu na ostale stil-ske pravce, ne čudi da naučeni model upravo za tu klasu postiže najvišu točnost klasifikacije. Najslabija klasifikacijska točnost postiže se za stilski pravac akademizma (21.5%), a slike koje pripadaju tom pravcu najčešće su pogrešno klasificirane kao realizam. Ono što je svojstveno i akademizmu i realizmu jest realističan način slikanja, tj. slikanje kojim se teži postizanju iluzije zbilje, dok se razlika između realizma i akademizma očituje u izboru motiva. U širem smislu pojam akademizma označava praćenje ustaljenih i školskih načela oblikovanja umjetničkog djela, no u ovom kontekstu akademizam se ponajprije odnosi na osobiti stilski pravac u umjetnosti 19. stoljeća koji prati standarde prihvaćene na umjetničkim akademijama, posebice pariškoj likovnoj akademiji *Académie des Beaux-Arts*, a očituje se izborom povijesnih,



**Slika 3.10:** Primjeri slika koje pripadaju klasi akademizma i koje pripadaju klasi realizma.

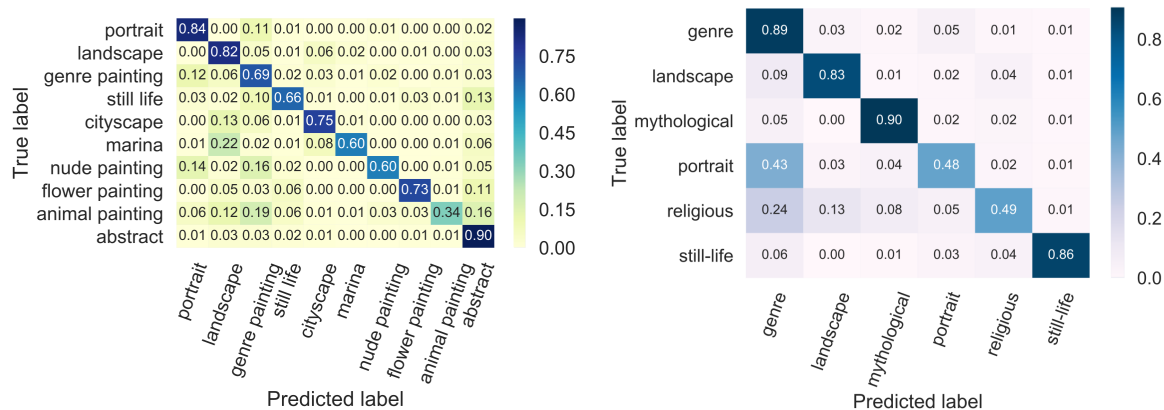
religioznih, mitoloških i alegorijskih motiva u likovnom stvaralaštvu. S druge strane, u fokusu realizma kao umjetničkog pravca zbivanja su svakodnevnog života, s posebnim naglaskom na vjerno prikazivanje društvenih pojava poput siromaštva i teškoga rada. Na slici 3.10 prikazano je nekoliko primjera slikarskih djela koja pripadaju akademizmu i koja pripadaju realizmu.

Navedeni primjer pogrešne klasifikacije ukazuje na složenost zadatka automatske klasifikacije stilskih pravaca. Pojam stilskega pravca nadilazi razlikovanje isključivo na temelju značajki i načela likovnog oblikovanja te ukazuje i na kompleksnu razinu razlikovanja sadržaja koja je često vrlo suptilna i ovisi o povijesno-društvenom kontekstu. Također, zbog preklapanja brojnih stilskih obilježja vrlo često ne postoji strogo razgraničenje stilskih pravaca kao što je to primjerice slučaj lirske apstrakcije i apstraktnog ekspresionizma, što je također vidljivo iz rezultata automatske klasifikacije.

Razlikovanje stilskih pravaca podrazumijeva prepoznavanje sličnih vizualnih obilježja u slikama različitog sadržaja, ali i prepoznavanje suptilnih razlika među sličnim vizualnim obilježjima, kao i među sličnim sadržajima. Budući da se radi o iznimno kompleksnom zadatku, ne čudi što je ukupna klasifikacijska točnost za zadatak klasifikacije stilskih pravaca niža u odnosu na točnosti postignute u drugim klasifikacijskim zadacima.

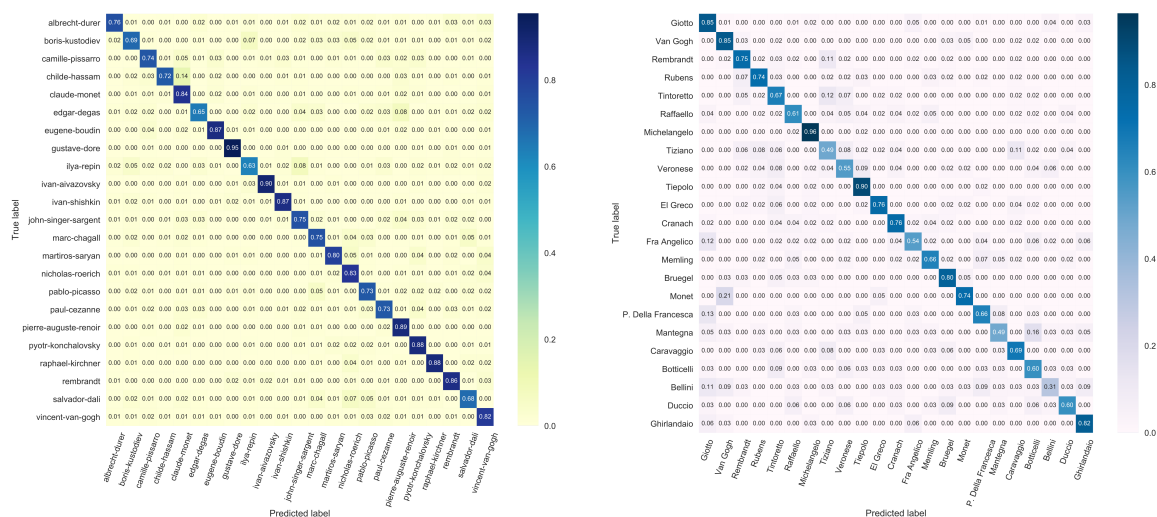
Problem klasifikacije slikarskih djela prema žanru temelji se na prepoznavanju sličnih sadržajnih motiva u različitim stilskim izvedbama. Pokazuje se da konvolucijske neuronske mreže naučene za raspoznavanje objekata i prostora u prirodnim slikama mogu vrlo dobro raspoznavati objekte i prizore u slikarskim djelima neovisno o stilskim značajkama oblikovanja i tehnici. Iz matrica grešaka prikazanih na slici 3.11 za klasifikaciju prema žanru WikiArt baze (lijevo) i WGA baze (desno), može se uočiti postojanje unutarnje logike klasifikacijskih pogrešaka. Primjerice, često su slike koje pripadaju klasi marina ili panorami grada klasificirane kao pejzaži budući da uključuju prizore vanjskih prostora, dok se aktovi često krivo klasificiraju kao portreti budući da često uključuju prikaz ljudskog lica.

U slučaju klasifikacije slikarskih djela prema autoru, ukupna klasifikacijska točnost vrlo je visoka za sve tri baze slika, posebno s obzirom na manji broj slika po klasi u odnosu na druge zadatke. Na slici 3.12 prikazane su matrice grešaka za zadatak klasifikacije prema autoru na WikiArt bazi (desno) i WGA bazi (lijevo). Stopa pogrešno klasificiranih slika u skladu je



Slika 3.11: Matrica grešaka modela naučenog za raspoznavanje žanra na WikiArt bazi (lijevo) i WGA bazi (desno).

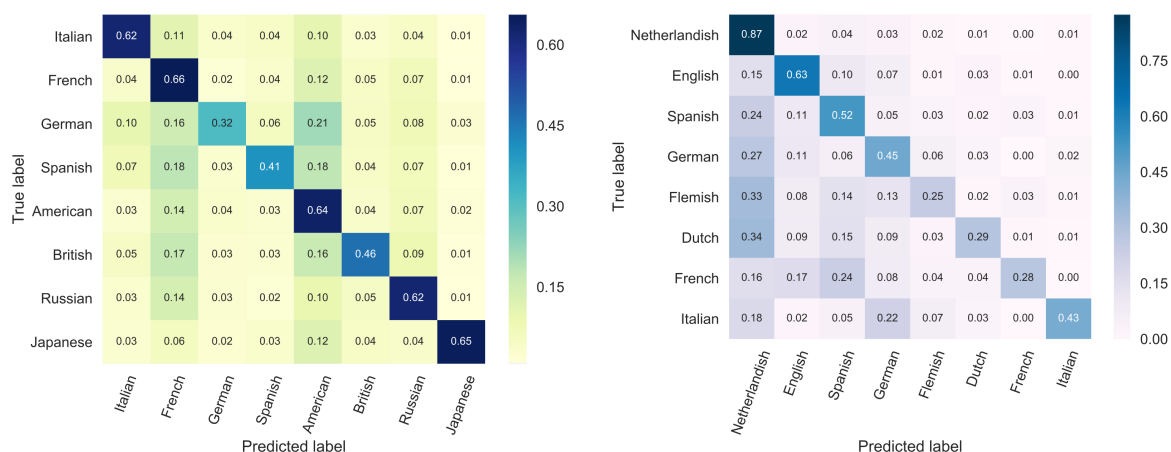
s općim shvaćanjem sličnosti među radovima pojedinih autora, pogotovo onih koji pripadaju istom stilskom razdoblju kao što su primjerice impresionisti Childe Hassam i Cloude Monet (WikiArt baza) ili talijanski slikari iz 16. stoljeća Tintoretto, Veronese i Tiziano (WGA baza).



Slika 3.12: Matrica grešaka modela naučenog za raspoznavanje autora na WikiArt bazi (lijevo) i WGA bazi (desno).

Klasifikacijske točnosti od 58.4% za WikiArt i 65.2% za WGA za zadatak klasifikacije slika prema nacionalnosti autora predstavljaju zanimljiv rezultat. Budući da je jedina zajednička poveznica u ovom klasifikacijskom zadatku pripadnost određenom nacionalnom umjetničkom krugu, tako visoka klasifikacijska točnost predstavlja relativno neočekivan ishod. Na slici 3.13 prikazane su matrica grešaka za zadatak klasifikacije prema nacionalnosti na WikiArt bazi (desno) i WGA bazi (lijevo).

Tako visoka klasifikacijska točnost za ovaj zadatak otvara nova pitanja i mogućnosti za daljnje istraživanje mjerljivosti vizualne sličnosti između slikarskih djela nastalih u određenom geografsko-povijesnim kontekstu, Također, korelacija među značajkama slika naučenih na



**Slika 3.13:** Matrica grešaka modela naučenog za raspoznavanje autora na WikiArt bazi (lijevo) i WGA bazi (desno).

mreži treniranoj za zadatak raspoznavanja slika prema nacionalnosti autora može ukazivati na dinamiku umjetničkih utjecaja, kao i na otkrivanje specifičnih uzoraka svojstvenih umjetničkoj baštini određenog područja.

### 3.5.3 Primjena naučenih modela za analizu različitih aspekata vizualne sličnosti

Pitanje svrhe učenja modela koje nadilazi isključivo analize postavki treniranja mreže i klasifikacijske točnosti očituje se u mogućnostima primjene naučenih modela, pri čemu se kao jedna od osnovnih mogućnosti nameće primjena naučenih modela u sustavima za automatsko pretraživanje slika na temelju sličnosti. Sustavi za pretraživanje slika na temelju sadržaja (engl. content-based image retrieval) uglavnom se zasnivaju na tome da se pretražuju slike koje sadrže iste motive, odnosno slične objekte ili prizore. No sličnost u kontekstu likovnog stvaralaštva može uključivati sličnost na temelju motiva, ali i sličnost na temelju specifičnih stilskih obilježja boje, kompozicije i linije. Budući da modeli trenirani za različite zadatke i skupove podataka uče reprezentacije slike specifične za određenu domenu, može se pretpostaviti da značajke iste slike ekstrahirane iz različito naučenih modela predstavljaju semantički različite reprezentacije slike. Stoga se značajke dobivene iz različito naučenih modela mogu koristiti za analizu različitih aspekata vizualne sličnosti.

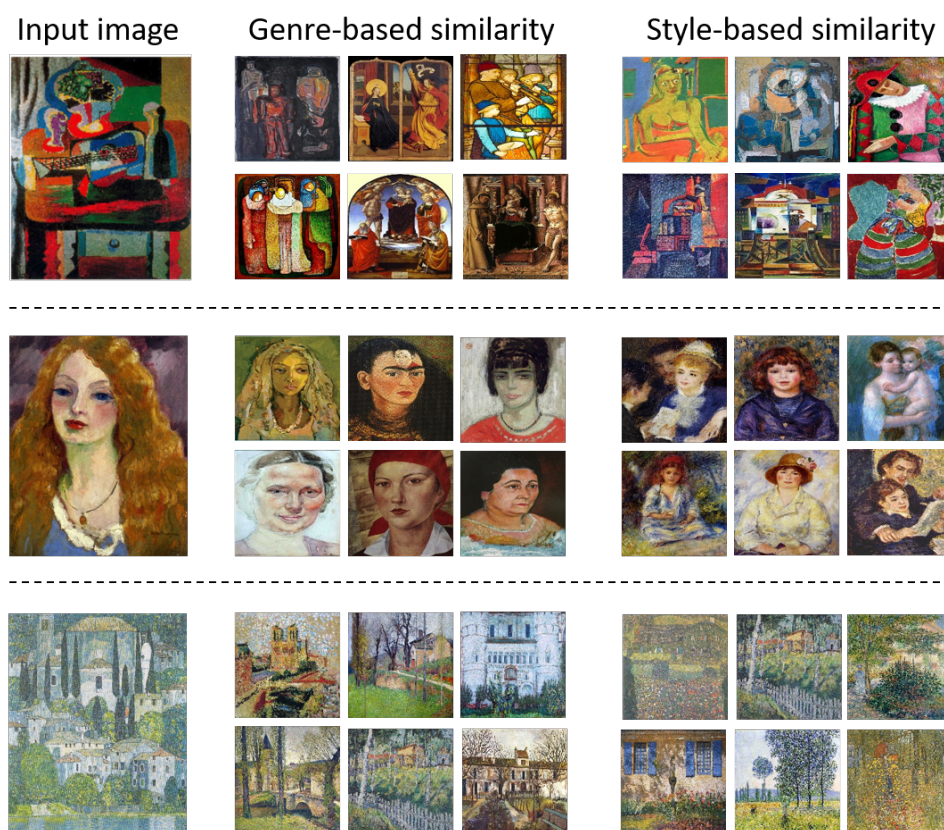
Modeli konvolucijskih neuronskih mreža koji su postupkom finog podešavanja parametara naučeni za klasifikaciju stila i klasifikaciju žanra na skupu podataka dobivenima iz WikiArt baze, koriste se za ekstrakciju značajki slika i definiranje dviju vrsta sličnosti - sličnost žanra i sličnost stila. Konkretno, aktivacije predzadnjeg potpuno povezanog sloja (fc7) koriste se kao 4096 dimenzionalni vektor značajki slike. Kao mjera sličnosti dvije vektorske reprezentacije



značajki koristi se kosinusna sličnost definirana kao:

$$K(x,y) = 1 - \frac{\sum_{i=1}^n x_i y_i}{\sqrt{\sum_{i=1}^n x_i^2} \sqrt{\sum_{i=1}^n y_i^2}} \quad (3.1)$$

Kosinusna udaljenost određena je kao 1.0 minus kosinus kuta između dva vektora. Ukoliko su vektori pozitivne vrijednosti, dobiva se pozitivna vrijednosti u rasponu od 0 i 1, pri čemu niže vrijednosti znače manje udaljenost među vektorima. Na temelju kosinusne udaljenosti među vektorima značajki određuje se stupanj sličnosti među različitim slikama u WikiArt bazi. Na slici 3.14 prikazani su primjeri slika koje su dobivene kao najbližnje ulaznim slikama kada se za ekstrakciju značajki slike koristio model naučen za raspoznavanje žanra i primjeri slika dobivenih kada se koristio model naučen za raspoznavanje stila. Navedeni primjeri ilustriraju upotrebu različito naučenih modela za dohvaćanje različito sličnih slika. Može se uočiti da slike koje su najbližnje ulaznoj slici na temelju značajki modela naučenog za klasifikaciju žanra pokazuju ponajprije sličnost sadržaja, odnosno prikaza motiva, kao sličnost kompozicije i rasporeda elemenata u slici. S druge strane, slike koje su najbližnje ulaznoj slici na temelju značajki modela naučenog za klasifikaciju stila pokazuju prvenstveno sličnost na temelju teksture određene specifičnim potezima kista te pripadaju istom stilskom pravcu.



**Slika 3.14:** Primjeri slika koje su dobivene kao najbližnje ulaznim slikama (lijevi stupac) kada se za ekstrakciju značajki slike koristio model naučen za raspoznavanje žanra (srednji stupac) i primjeri slika dobivenih kada se koristio model naučen za raspoznavanje stila (desni stupac).

Nadalje, navedena metoda za izračun sličnosti među slikama korištena je u svrhu ispitivanje sličnosti među značajkama različito naučenih modela. S tom svrhom korišten je podskup od 100 slučajno odabranih slika za koje su primjenom različito naučenih modela ekstrahirane fc7 značajke, odnosno 4096 dimenzionalni izlazi predzadnjeg potpuno povezanog sloja. Zatim je izračunata srednja vrijednosti kosinusnih udaljenosti među značajkama dobivenih iz različito naučenih modela te je dobivena matrica udaljenosti prikazana na slici 3.15

	caffenet	flickr	hybrid	memnet	sentiment	wiki_artist_caffe	wiki_artist_flickr	wiki_artist_hybrid	wiki_artist_memnet	wiki_artist_sentiment	wiki_genre_caffe	wiki_genre_flickr	wiki_genre_hybrid	wiki_genre_memnet	wiki_genre_sentiment	wiki_style_caffe	wiki_style_flickr	wiki_style_hybrid	wiki_style_memnet	wiki_style_sentiment
caffenet	-0.00	0.59	0.85	0.79	0.63	0.27	0.57	0.86	0.77	0.67	0.25	0.52	0.85	0.73	0.61	0.33	0.59	0.87	0.79	0.72
flickr	0.59	-0.00	0.89	0.85	0.79	0.67	0.48	0.90	0.83	0.81	0.67	0.46	0.89	0.81	0.79	0.69	0.48	0.91	0.85	0.83
hybrid	0.85	0.89	-0.00	0.70	0.90	0.86	0.84	0.30	0.56	0.89	0.85	0.82	0.26	0.49	0.88	0.87	0.84	0.38	0.59	0.90
memnet	0.79	0.85	0.70	-0.00	0.86	0.81	0.79	0.66	0.48	0.85	0.80	0.76	0.64	0.40	0.84	0.82	0.79	0.66	0.53	0.87
sentiment	0.63	0.79	0.90	0.86	-0.00	0.68	0.77	0.91	0.85	0.33	0.67	0.74	0.90	0.83	0.38	0.68	0.75	0.91	0.86	0.44
wiki_artist_caffe	0.27	0.67	0.86	0.81	0.68	-0.00	0.58	0.87	0.79	0.68	0.40	0.60	0.86	0.75	0.68	0.34	0.61	0.88	0.80	0.73
wiki_artist_flickr	0.57	0.48	0.84	0.79	0.77	0.58	-0.00	0.85	0.75	0.76	0.66	0.46	0.84	0.72	0.76	0.65	0.42	0.86	0.78	0.80
wiki_artist_hybrid	0.86	0.90	0.30	0.66	0.91	0.87	0.85	-0.00	0.53	0.90	0.86	0.83	0.40	0.52	0.88	0.88	0.85	0.38	0.56	0.91
wiki_artist_memnet	0.77	0.83	0.56	0.48	0.85	0.79	0.75	0.53	-0.00	0.83	0.78	0.73	0.56	0.39	0.82	0.80	0.76	0.58	0.45	0.85
wiki_artist_sentiment	0.67	0.81	0.89	0.85	0.33	0.68	0.76	0.90	0.83	-0.00	0.70	0.75	0.89	0.81	0.47	0.69	0.75	0.90	0.85	0.44
wiki_genre_caffe	0.25	0.67	0.85	0.80	0.67	0.40	0.66	0.86	0.78	0.70	-0.00	0.48	0.85	0.74	0.58	0.40	0.63	0.87	0.79	0.73
wiki_genre_flickr	0.52	0.46	0.82	0.76	0.74	0.60	0.46	0.83	0.73	0.75	0.48	-0.00	0.82	0.69	0.67	0.62	0.44	0.84	0.75	0.78
wiki_genre_hybrid	0.85	0.89	0.26	0.64	0.90	0.86	0.84	0.40	0.56	0.89	0.85	0.82	-0.00	0.42	0.88	0.86	0.84	0.42	0.58	0.90
wiki_genre_memnet	0.73	0.81	0.49	0.40	0.83	0.75	0.72	0.52	0.39	0.81	0.74	0.69	0.42	-0.00	0.79	0.76	0.72	0.54	0.44	0.82
wiki_genre_sentiment	0.61	0.79	0.88	0.84	0.38	0.68	0.76	0.88	0.82	0.47	0.58	0.67	0.88	0.79	-0.00	0.67	0.74	0.89	0.83	0.51
wiki_style_caffe	0.33	0.69	0.87	0.82	0.68	0.34	0.65	0.88	0.80	0.69	0.40	0.62	0.86	0.76	0.67	-0.00	0.59	0.88	0.81	0.70
wiki_style_flickr	0.59	0.48	0.84	0.79	0.75	0.61	0.42	0.85	0.76	0.75	0.63	0.44	0.84	0.72	0.74	0.59	-0.00	0.86	0.78	0.76
wiki_style_hybrid	0.87	0.91	0.38	0.66	0.91	0.88	0.86	0.38	0.58	0.90	0.87	0.84	0.42	0.54	0.89	0.88	0.86	-0.00	0.57	0.91
wiki_style_memnet	0.79	0.85	0.59	0.53	0.86	0.80	0.78	0.56	0.45	0.85	0.79	0.75	0.58	0.44	0.83	0.81	0.78	0.57	-0.00	0.86
wiki_style_sentiment	0.72	0.83	0.90	0.87	0.44	0.73	0.80	0.91	0.85	0.44	0.73	0.78	0.90	0.82	0.51	0.70	0.76	0.91	0.86	-0.00

**Slika 3.15:** Matrica srednjih vrijednosti kosinusnih udaljenosti među značajkama dobivenih iz različito naučenih modela na skupu od 100 slučajno odabranih slika.

Udaljenosti među značajkama slika dobivenih iz različito naučenih modela pokazuju koliko su ishodi učenja različitih reprezentacija različiti s obzirom na domenu za koju se model uči. Iz prikazane matrice udaljenosti vidljivo je da su nakon finog podešavanja parametara za određeni klasifikacijski zadatak značajke takvih modela najbližije značajkama dobivenima iz bazne mreže. Nadalje, prilikom analize najbližijih slika utvrđeno je da se za većinu slika srednja vrijednost kosinusne udaljenosti njima najbližih 100 slika kreće u rasponu od 0.3 do 0.4. S obzirom na iznose srednjih vrijednosti među značajkama istih slika dobivenih iz različito naučenih modela prikazanih na slici 3.15, zaključuje se da izbor modela znatno utječe na ishod dohvaćanja slika na temelju sličnosti.

### 3.6 Osvrt na rezultate istraživanja

U radu su prikazani rezultati provedenih eksperimenata finog podešavanja parametara konvolucijskih neuronskih mreža za različite klasifikacijske zadatke i različite baze slika. Provedena je usporedna analiza različitih strategija finog podešavanja mreže kako bi se odredile optimalne postavke za različite klasifikacijske zadatke te su postignuti trenutno najbolji rezultati klasifikacijske točnosti za većinu predstavljenih zadataka klasifikacije slikarskih djela. Također, prvi puta uveden je zadatak klasifikacije slikarskih djela prema pripadajućem nacionalnom umjetničkom kontekstu. S obzirom na neočekivano visoke rezultate klasifikacijske točnosti, pokazuje se da postoje osnove za daljnja istraživanja mjerljivosti vizualnih značajki u odnosu na tako visoko apstraktnu kategoriju kao što je pripadanje određenom prostorno i kulturološki određenom umjetničkom krugu.

Problem klasifikacije slikarskih djela predstavlja inicijalni korak u daljnjem razvoju metoda za računalnu analizu umjetničkih djela. Rezultati dobiveni u ovom radu ukazuju na nekoliko različitih budućih istraživačkih smjernica. Pokazano je da naučeni modeli mogu biti korišteni za pretraživanje kolekcija slika na temelju različitih oblika vizualne sličnosti, pa se time ocrta mogućnost njihove konkretne aplikativne svrhe. Nadalje, analizom vizualne sličnosti nad velikim skupovima slikovnih podataka, te posebice uzimajući u obzir odnos među likovnim opusima pojedinih autora ili stilskih razdoblja, moguće je provesti kvantificirano ispitivanje poznatih saznanja iz povijesti umjetnosti, a eventualno otkriti i neke nove obrasce povezanosti. Osim korištenja konvolucijskih neuronskih mreža za formiranje novih metoda analize u kontekstu povijesti umjetnosti, otvaraju se i daljnje smjernice za istraživanje ponašanja samih dubokih neuronskih mreža.

Poseban naglasak stavljen je na analizu utjecaja domenski specifične inicijalizacije bazne mreže prilikom postupka finog podešavanja parametara mreže za nove zadatke. Rezultati dobivene analize ukazuju na mogućnost određivanja semantičke udaljenosti među različitim domenama na temelju parametara različito naučenih modela, kao i na mogućnosti optimizacije postupka učenja odabirom drugačijih domenski određenih inicijalnih stanja mreže za različite zadatke klasifikacije slika. Međutim, kako bi se odredili konačni zaključci o semantičkim implikacijama domenski specifičnih težinskih inicijalizacija potrebna su daljnja istraživanja ovog područja koja podrazumijevaju i postojanje rezultata utvrđenih psihološkim eksperimentalnim ispitivanjem korelacije među određenim svojstvima slike. Primjerice, postojanje eksperimentalnim ispitivanje dobivenog zaključka o korelaciji između pamtljivosti slike i emocije koju ona izaziva, preduvjet je za evaluaciju semantičke sličnosti reprezentacija slike dobivenih iz različito naučenih modela konvolucijskih neuronskih mreža. Budući da je provedba ispitivanja koje uključuju prikupljanje oznaka subjektivne procjene određenih atributa slike složena i skupa, primjenom konvolucijskih neuronskih mreža mogu se postaviti preliminarne hipoteze o seman-



tičkoj povezanosti određenih domensko specifičnih klasifikacijskih zadataka, dok se naučene značajke mogu koristiti za nove oblike analize poznatog skupa podataka.

## Poglavlje 4

# Računalna analiza specifičnih stilskih obilježja slikarskih djela

### 4.1 Pozadina i motivacija istraživanja

Formalna analiza likovnog djela predstavlja jednu od osnovnih metoda povijesno-umjetničkog pristupa likovnog djelu. Iako nije u fokusu suvremenog povijesno-umjetničkog interesa, formalna analiza likovnih djela zauzima središnje mjesto u razvoju povijesti umjetnosti i razumijevanju stilskih mijena. Budući da se suvremena povijesno-umjetnička istraživanja sve više okreću uporabi računalnih metoda, a zadnjih se nekoliko godina razvija i nova istraživačka disciplina tzv. „digitalna povijesti umjetnosti“ (engl. Digital Art History), jedna od izranjajućih istraživačkih smjernica upravo je ostvarenje tradicionalnih povijesno-umjetničkih pristupa analizi umjetničkog djela korištenjem računalnih metoda. Razvoj metoda za kvantifikaciju zastupljenosti određenih principa oblikovanja i stilskih obilježja otvara mogućnost novih analitičkih perspektiva na već istraženu umjetničku baštinu ponajprije zbog mogućnosti obrade i analize velikog skupa podataka. U tom kontekstu bitnu ulogu igraju upravo tzv. značajke slike više razine, odnosno značajke slike koje služe kao mjere zastupljenosti određenih umjetničkih koncepata u slici. Koncepti čija se kvantifikacija istražuje u ovome radu, temelje se na pojmovnim parovima koje je početkom 20. stoljeća formulirao povjesničar umjetnosti Heinrich Wölfflin (1864.-1945.) u svome djelu *Temeljni pojmovi povijesti umjetnosti* [63]. Wölfflin uspostavlja pet pojmovnih parova kojima se služi u razlaganju prijelaza umjetničkog stila 16. stoljeća u stil 17. stoljeća. Iako se u svome djelu Wölfflin bavi isključivo razradom razlika umjetničkog oblikovanja renesanse i baroka, „temeljni pojmovi“ predstavljaju jednako važan okvir za tumačenje stilskih obilježja novijih umjetničkih stilova.

Postojanje baze s označenim značajkama slike preduvjet je za korištenje konvolucijskih neuronskih mreža sa svrhom učenja kompleksnih značajki. Za označavanje slika na temelju razlikovanja Wölfflinovih pojmovnih parova potrebno je stručno znanje pa se takve oznake teško

mogu prikupiti za velik broj slika i ispitivanjem velikog broja sudionika (engl. crowdsourcing). Značajan doprinos rada kojeg su predstavili Elgammal et al. [45] upravo je kreiranje baze slika s oznakama Wölfflinovih pojmovnih parova. Istraživanje i rezultati opisani u tom radu, predstavljaju osnovnu motivaciju za istraživanje predstavljeno u ovom poglavlju. U radu Elgammal et al. [45] predstavljeno je nekoliko modela konvolucijskih neuronskih mreža naučenih za klasifikaciju slike prema stilu. No osim samog problema klasifikacije, u tom je radu analizirana interpretabilnost naučenih reprezentacija slike. Nad naučenim reprezentacijama provedena je redukcija dimenzionalnosti korištenjem metode analize glavnih komponentata (engl. principal component analysis - PCA) te korelacijska analiza odnosa vrijednosti prvih dviju PCA komponenti i Wölfflinovih oznaka. Pokazano je da postoji značajna korelacija među vrijednostima prvih nekoliko PCA komponenti naučenih reprezentacija i vrijednosti zastupljenosti različitih Wölfflinovih koncepata. Zaključak toga istraživanja je da konvolucijske mreže naučene za klasifikaciju stila, uče stilove razlikovati prema značajkama koje odgovaraju povijesno-umjetničkim konceptima analize umjetničkog djela. Na temelju iznesenih zaključaka i dostupne baze slikarskih djela s oznakama Wölfflinovih koncepata, u ovome je radu predložena metoda učenja regresijskih modela konvolucijskih neuronskih mreža za raspoznavanje Wölfflinovih koncepata u slikarskim djelima.

U nastavku je prikazan kratak opis pet temeljnih Wölfflinovih koncepata. Za svaki koncept dan je primjer slika koje ilustriraju razliku pojmovnih parova. Na temelju razlike između umjetnosti renesanse i umjetnosti baroka, Wölfflin uspostavlja pet različitih pojmovnih parova. Tranzicija stilskih obilježja iz razdoblja renesanse u razdoblje baroka, podrazumijeva tranziciju od prvoga prema drugome pojmu:

- *Linearnost i slikovitost*

Linearni stil podrazumijeva ujednačeno i jasno omeđivanje oblika, istaknute rubove i čvrsto odvajanje pojedinih oblika, čime se postiže doživljaj opipljivosti ploha i plastičnosti prikaza. S druge strane, slikoviti stil usmjeren je više na optički privid stvari koji se postiže razaranjem neprekinutosti obrisa i stvaranjem privida pokretnosti elemenata u slici. Primjeri djela linearnog i slikovitog stila dani su na slici 4.1.

- *Zatvorena i otvorena forma*

Djela zatvorene forme (tektonski stil) odlikuje doživljaj nužnosti, reda i omeđenosti. Vertikale i horizontale dominantni su kompozicijski pravci, slikom upravlja opreka između horizontale i vertikale, dok se dijelovi slike raspoređuju oko središnje osi. Kod otvorene forme (atektonskog stila) glavno kompozicijsko usmjerenje je dijagonala, dok se oprečnost vertikale i horizontale zatamljuje. Kod slika otvorene forme motivi se presjecaju s rubom slike te se dobiva dojam prikaza prizora koji prolazi, a ne izdvojenog i omeđenog komadića svijeta. Na slici 4.2 dani su primjeri slika otvorene i zatvorene forme.



**Slika 4.1:** Lijevo - primjer slike linearnog stila, (Bronzino (1505-1572), *Eleonora od Toleda sa sinom Giovannijem*). Desno - primjer slike slikovitog stila (Velasquez (1599-1660), *Infantkinja Margareta Theresia u bijeloj haljini*). Primjeri su preuzeti iz knjige Heinricha Wölfflina *Temeljni pojmovi povijesti umjetnosti* [63].



**Slika 4.2:** Lijevo - primjer slike zatvorene forme, (Issenbrant (umro 1551), *Odmor na brijegu*). Desno - primjer slike otvorene forme (Rubens (1577-1640), *Marija sa svecima*). Primjeri su preuzeti iz knjige Heinricha Wölfflina *Temeljni pojmovi povijesti umjetnosti* [63].

- *Ploha i dubina*

Plošno povezivanje oblika podrazumijeva usporedno razmještanje elemenata u slici, takvo da se svi oblici uklapaju u glavnu plohu slike. Dojam dubine ostvaruje se razaranjem reljefnog karaktera prikaza i razmještanjem elemenata prema dubini. Također, u slikama naglašene dubine čest je dijagonalni potez u kompoziciji. Primjeri djela s naglašenom plošnom i dubinskom kompozicijom dani su na slici 4.3.

- *Mnoštvo i jedinstvo*

Dojam mnoštva postiže se isticanjem više pojedinačnih elemenata. Svaki element slike pojavljuje se na svome mjestu uvjetovan cjelinom, no pritom djeluje posve samostalno. Dojam jedinstva postiže se usmjeravanjem svih elemenata jednom glavnom motivu. Stilsko jedinstvo odlikuju sažimljujuće tendencije u oblikovanju elemenata, pri čemu se stvara privid protočnosti i isprepletenosti elemenata. Na slici 4.4 dan je primjer slika kod kojih se raspoznaje razlika između ostvarenja principa mnoštva i principa jedinstva.



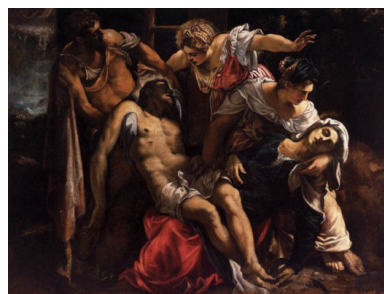
**Slika 4.3:** Lijevo - primjer slike plošnog oblikovanja, (Palma Vecchio (oko 1480-1528), *Adam i Eva*). Desno - primjer slike s naglašenom dubinom (Tintoretto (1518-1594), *Adam i Eva*). Primjeri su preuzeti iz knjige Heinricha Wölfflina *Temeljni pojmovi povijesti umjetnosti* [63].



**Slika 4.4:** Lijevo - primjer slike ostvarenja principa mnoštva, (Holbein (1488./98-1543), isječak iz slike *Poslanici Dinteville i de Selve*). Desno - primjer slike ostvarenja principa jedinstva (Velasquez (1599-1660), *Kardinal Borgia*). Primjeri su preuzeti iz knjige Heinricha Wölfflina *Temeljni pojmovi povijesti umjetnosti* [63].

- *Jasnost i nejasnost*

Jasnost, odnosno ono što Wölfflin naziva apsolutnom jasnoćom, očituje se u razgovijetnoj pojavnosti oblika. Svjetlo i sjena u službi su pojašnjavanja forme te se oštrinom bridova postiže dojam jasnoće predmetne pojavnosti. Nasuprot tome, nejasnost, odnosno relativna jasnoća, podrazumijeva prekrivanje i zasjenjivanje oblika. Na slici 4.5 prikazani su primjeri ostvarenja apsolutne jasnoće i relativne jasnoće.



**Slika 4.5:** Lijevo - primjer slike ostvarenja apsolutne jasnoće, (Joos van Cleve (1485-1540) *Oplakivanje Krista*). Desno - primjer slike ostvarenja relativne jasnoće (Tintoretto (1518-1594), *Oplakivanje Krista*). Primjeri su preuzeti iz knjige Heinricha Wölfflina *Temeljni pojmovi povijesti umjetnosti* [63].

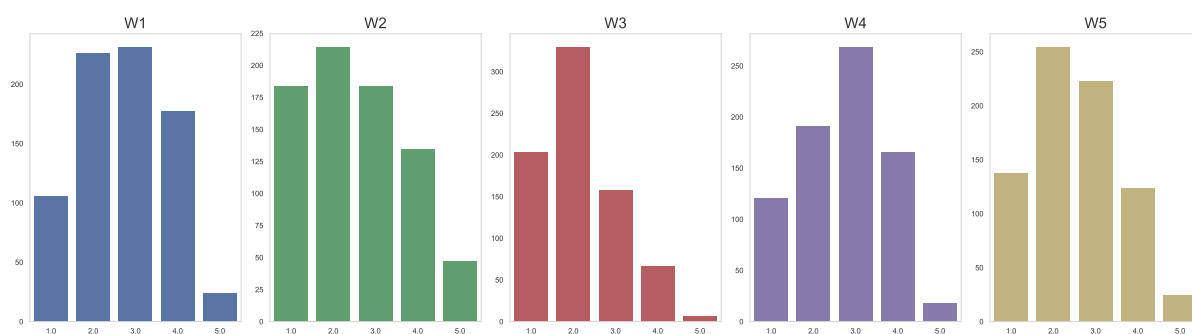
Navedeni primjeri preuzeti su iz knjige Heinricha Wölfflina *Temeljni pojmovi povijesti umjetnosti* [63] u kojoj autor definira karakteristike pojedinih pojmovnih parova na temelju razlika u likovnoj umjetnosti 16. i 17. stoljeća. S obzirom na kronološku bliskost, navedeni primjeri ilustriraju suptilnu razinu razlikovanja pojedinih obilježja, odnosno tek naznačuju stilsku tranziciju s prvog na drugi pojmovni par. U kasnijim stilskim razdobljima pojedina se obilježja ostvaruju u svojoj krajnjoj inačici. Primjerice, u razdoblju impresionizma slikovitost preuzima potpunu prevlast u odnosu na linearnost. Iako bi se izborom drugačijih primjera možda jasnije ilustrirala razlika pojedinih pojmovnih parova, navedeni primjeri odabrani su zbog želje da se pojmovni parovi objasne u skladu s izvornim opisom i da se naglasi suptilnost značajki stilskih obilježja i kompleksnost zadatka njihove računalne detekcije.

## 4.2 Baze slika i regresijski zadaci

Baza slika korištena za zadatak predikcije vrijednosti pojedinih Wölfflinovih koncepata, predstavljena je u radu Elgammala et al. [45]. Baza se sastoji od 765 slika preuzetih iz WikiArt kolekcije. Na temelju stručne procjene povjesničara umjetnosti, slikama su pridodane oznake vrijednosti za svaki Wölfflinov koncept. Za svaki koncept, odnosno pojmovni par, slikama su dodijeljene vrijednosti između 1 i 5. Primjerice, za koncept odnosa linearnosti i slikovitosti vrijednost 1 označava da je slika izrazito linearna, a vrijednost 5 da je izrazito slikovita. U nastavku rada mjere opisanih koncepata skraćeno se označavaju na sljedeći način:

- W1 - Linearnost i slikovitost
- W2 - Zatvorena i otvorena forma
- W3 - Ploha i dubina
- W4 - Mnoštvo i jedinstvo
- W5 - Jasno i nejasno

Na slici 4.6 prikazane su histogrami vrijednosti za svih pet pojmovnih parova (W1 - W5), dobivene stručnim označavanjem baze od 765 slikarskih djela.



**Slika 4.6:** Distribucija vrijednosti oznaka pojedinih pojmovnih parova dobivenih stručnim označavanjem baze od 765 slikarskih djela.

Baza s Wöllflinovim oznakama korištena je za učenje pet različitih modela konvolucijskih neuronskih mreža za predikciju vrijednosti navedenih obilježja u slici. Svaki koncept predstavlja jedan regresijski zadatak za koji je naučen određen model konvolucijske neuronske mreže. Za svaki je model skup slika podijeljen slučajnim odabirom tako da se 640 slika koristi za treniranje mreže, 65 za validaciju i 60 za testiranje.

Sa svrhom analize odnosa između predikcija pojedinih stilskih obilježja i povijesno-umjetničkih kategorija, modeli naučeni na manjem označenom skupu slika korišteni su za dobivanje predikcijskih vrijednosti pojedinih obilježja na velikom skupu slika iz WikiArt baze koja je detaljno opisana u poglavlju 3.2. Kao što je već istaknuto, u trenutku dohvaćanja WikiArt kolekcije za potrebe ovog rada, ukupan broj djela u kolekciji iznosio je 133220. Predikcija stilskih obilježja provedena je na pročišćenom skupu od 105121 slika. Korištene su isključivo slike koje su kategorizirane kao slikarska djela te su uklonjene slike koje pripadaju kategoriji fotografije, postera, grafita, itd. Također, zadržane su samo slike u boji, odnosno uklonjene su sve crno-bijele digitalne reprodukcije slika.

### **4.3 Fino podešavanja parametara mreže za učenje značajki Wöllflinovih koncepata**

Baza slika s označenim vrijednostima Wöllflinovih koncepata koristi se za učenje regresijskih modela konvolucijskih neuronskih mreža. Za svaki koncept, odnosno zadatak, uči se zasebni model. Budući da je baza označenih slika relativno mala, ne uči se cijela mreža već se radi fino podešavanja parametara pojedinih slojeva mreže te se ispituju optimalne razine dosega finog podešavanja. Pojedinih postavki finog podešavanja parametara mreža i dobiveni rezultati opisani su u nastavku.

#### **4.3.1 Postavke učenja modela**

U radu Elgamala et al. [45] pokazano je da postoji značajna korelacija između vrijednosti zastupljenosti različitih Wöllflinovih koncepata i vrijednosti PCA komponenti naučenih reprezentacija slike dobivenih iz modela konvolucijske neuronske mreže naučene za klasifikaciju stila. Zbog toga se u postupku finog podešavanja parametara modela konvolucijskih neuronskih mreža, s ciljem učenja značajki Wöllflinovih koncepata, kao bazna mreža koristi model naučen za klasifikaciju stilova dobiven kao rezultat istraživanja predstavljenog u poglavlju 3.5. Radi se o modelu koji je naučen raspoznavati 27 različitih klasa stilova WikiArt baze s klasifikacijskom točnošću od 56.45%. Budući da je to model naučen za klasifikaciju slika, zadnji softmax sloj mreže (fc8) zamijenjen je regresijskim slojem s jednim neuronom, čija je funkcija gubitka srednji kvadratni gubitak (engl. mean squared error) očekivanog i dobivenog izlaza.



**Tablica 4.1:** Srednja kvadratna pogreška i koeficijent determinacije  $R^2$  dobivene na testnim skupovima podataka primjenom najbolje naučenog modela za svaki zadatak predikcije pojedinog stilskog obilježja.

Zadatak	Srednja kvadratna pogreška	$R^2$ koeficijent determinacije
W1	0.0251	0.487
W2	0.0338	0.4179
W3	0.0185	0.475
W4	0.0258	0.399
W5	0.0369	0.3473

Prilikom finog podešavanja parametara modela ispitani su različiti scenariji s obzirom na doseg treniranja, odnosno s obzirom na to koliko slojeva bazne mreže ostaje zamrznuto, a koliko se modificira učenjem na novom zadatku. Eksperimentalnim ispitivanjem pokazano je da se za sve regresijske zadatke najbolji rezultati postižu za slučaj kada se modificiraju samo parametri zadnja tri potpuno povezana sloja (fc6, fc7 i fc8), dok parametri konvolucijskih slojeva odgovaraju parametrima bazne mreže, tj. parametrima modela naučenog za klasifikaciju stila. Ispitivanjem je pokazano da se najbolji rezultati postižu kada se mreža trenira 200 epoha, s fiksnom stopom učenja od  $10^{-4}$ . Implementacija postupka učenja mreža provedena je korištenjem programskog okvira za duboko učenje Caffe\*.

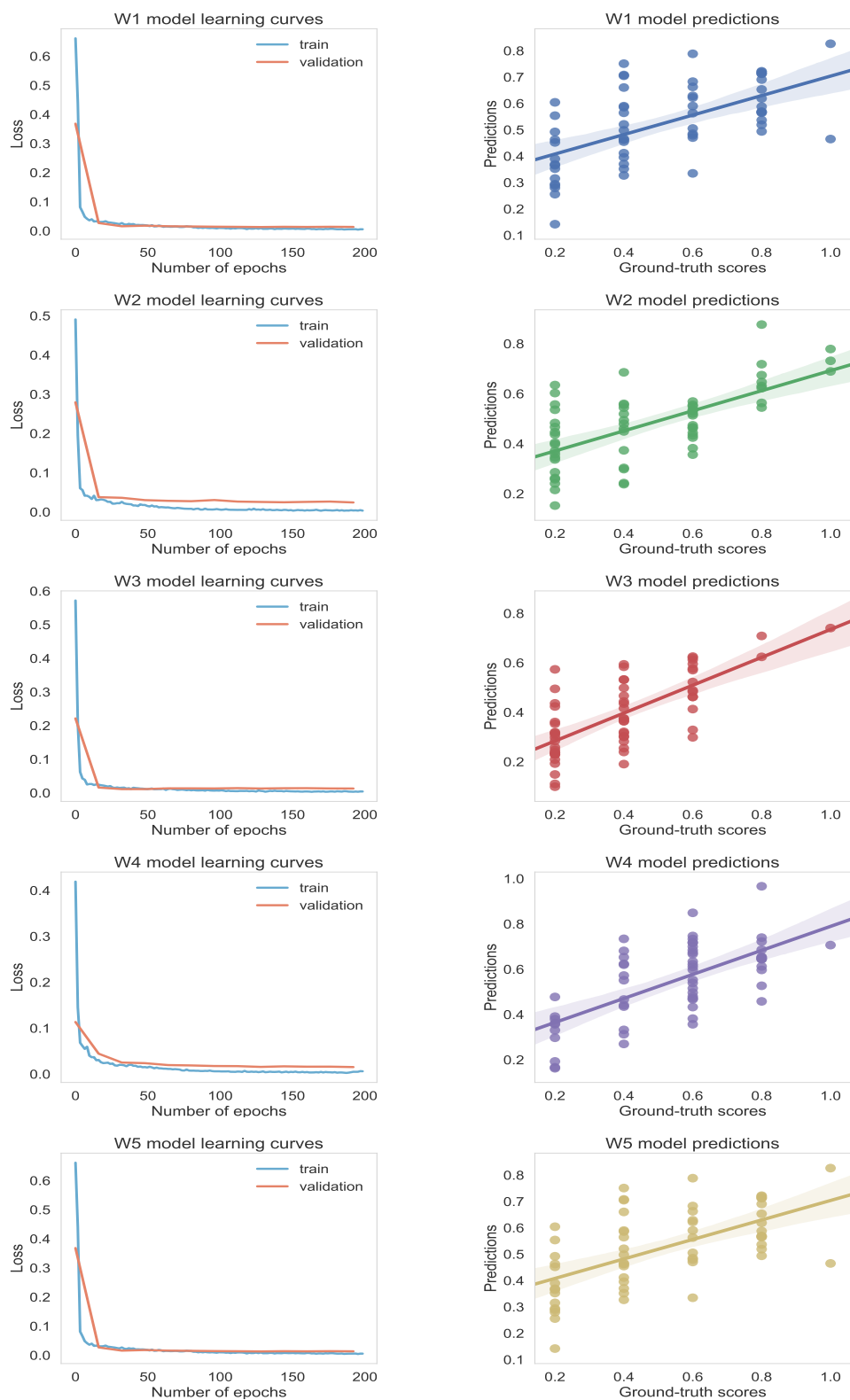
### 4.3.2 Rezultati učenja modela

Analizom krivulja učenja i ispitivanjem točnosti predikcije modela na testnim skupovima podataka utvrđene su optimalne postavke učenja za svaki model. U tablici 4.1 navedene su srednja kvadratna pogreška i koeficijent determinacije  $R^2$  dobivene na testnim skupovima podataka primjenom najbolje naučenog modela za svaki zadatak. Koeficijent determinacije je pokazatelj reprezentativnosti regresijskog modela budući da daje informaciju o tome koliki je udio promjena u mjerenim vrijednostima zavisne varijable objašnjen dobivenim modelom, a koliko otpada na rezidualno rasipanje. Koeficijent determinacije poprima vrijednosti iz intervala između 0 i 1, a model se smatra reprezentativnijim što je vrijednost koeficijenta determinacije bliže 1.

Rezultati mjere reprezentativnosti pojedinog regresijskog modela, tj. vrijednosti koeficijenta determinacije, pokazuju da se najbolji stupanj reprezentativnosti postiže kod modela naučenih za zadatke predikcije vrijednosti odnosa linearnosti i slikovitosti (W1) i odnosa plohe i dubine (W3).

\*Caffe Deep Learning Framework <http://caffe.berkeleyvision.org>





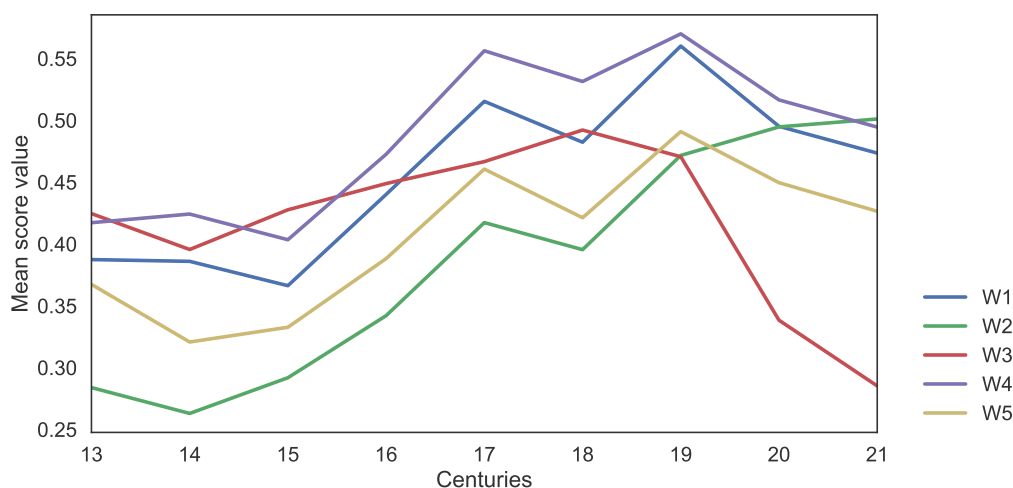
**Slika 4.7:** Krivulje učenja modela koje prikazuju vrijednost funkcije gubitka ovisno o broju epoha na skupu za treniranje i validaciju (lijevo) i korelacijski dijagram stvarnih vrijednosti i predikcijskih vrijednosti pojedinih stilskih obilježja na testnom skupu slika (desno).

Na slici 4.7 (lijevo) prikazane su krivulje učenja modela koje prikazuju vrijednost funkcije gubitka ovisno o broju epoha na skupu za treniranje i validaciju, pri čemu x os predstavlja broj

epoha, na lijevoj y osi prikazan je gubitak na skupu za treniranje (plavo) i gubitak na skupu za validaciju (crveno). Na slici 4.7 (desno) prikazan je korelacijski dijagram stvarnih vrijednosti pojedinih stilskih obilježja i predikcijskih vrijednosti dobivenih primjenom naučenih modela na testnom skupu slika.

## 4.4 Predikcije stilskih obilježja u kontekstu povijesti umjetnosti

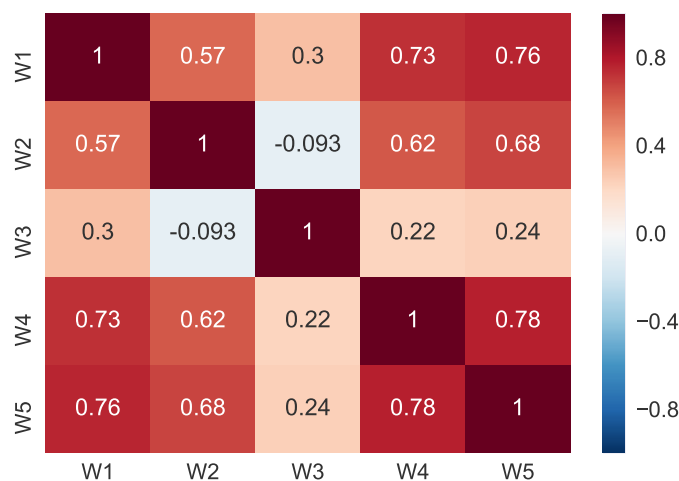
U ovom poglavlju razmatraju se kvalitativni rezultati predikcija pojedinih stilskih obilježja dobivenih finim podešavanjem parametara regresijskih modela konvolucijskih neuronskih mreža. Osim kvalitativnih rezultata, razmatra se kronološki odnos predikcijskih vrijednosti pojedinih obilježja te njihov odnos prema pojedinim povijesno-umjetničkim kategorijama slikarskih djela u WikiArt bazi. Sa svrhom ispitivanja kako se kronološki mijenjaju predikcijske vrijednosti pojedinih obilježja, koristi se podskup WikiArt baze koji sadrži oznake godine nastanka pojedinog djela. Podskup se sastoji od 82000 slika koje se s obzirom na godinu nastanka grupiraju prema stoljećima. Za sve slike koje pripadaju određenom stoljeću računa se srednja vrijednost predikcija pojedinog stilskog obilježja. Predikcijska vrijednost pojedinog obilježja određene slike izražena je brojem u intervalu između 0 i 1. Svako obilježje određeno je odnosom između dvaju komplementarnih pojmovnih parova opisanih u uvodnom poglavlju. Niske vrijednosti označavaju prisutnost obilježja izraženog prvim pojmovnim parom, a visoke vrijednosti obilježje izraženo drugim pojmovnim parom. Na slici 4.8 je graf koji prikazuje kako se kronološki mijenjaju srednje vrijednosti pojedinih stilskih obilježja.



**Slika 4.8:** Kronološki poredak srednjih vrijednosti predikcija stilskih obilježja (W1 - W5) na WikiArt bazi slika.

Kronološkim poretom srednjih vrijednosti predikcija Wölfflinovih stilskih obilježja utvrđuje se postojanje generalne konzistentnosti između promjena predikcijskih vrijednosti i opaža-

nja iznesenih u Wölfflinovoj knjizi *Temeljni pojmovi povijesti umjetnosti*. Vidljivo je da srednje vrijednosti predikcijske vrijednosti svih pet obilježja rastu od 15. do 17. stoljeća, što je u skladu s Wölfflinovim opažanjima o promjeni stilskih obilježja u umjetnosti renesanse i baroka. Osim konzistentnosti odnosa predikcijskih vrijednosti stilskih obilježja u razdobljima renesanse i baroka, njihovo opadanje početkom 19. stoljeća također je u skladu s Wölfflinovim navodima da umjetnost novog klasicizma, koja započinje oko 1800., skreće od slikovitosti prema plastičnosti (opadanje W1 vrijednosti). Iz kronoloških odnosa srednjih vrijednosti stilskih obilježja može se uočiti korelacija vrijednosti W1, W4 i W5 obilježja. Na slici 4.9 prikazane su vrijednosti korelacijskih koeficijenata između vrijednosti pojedinih stilskih obilježja dobivenih za sve slike iz WikiArt kolekcije.



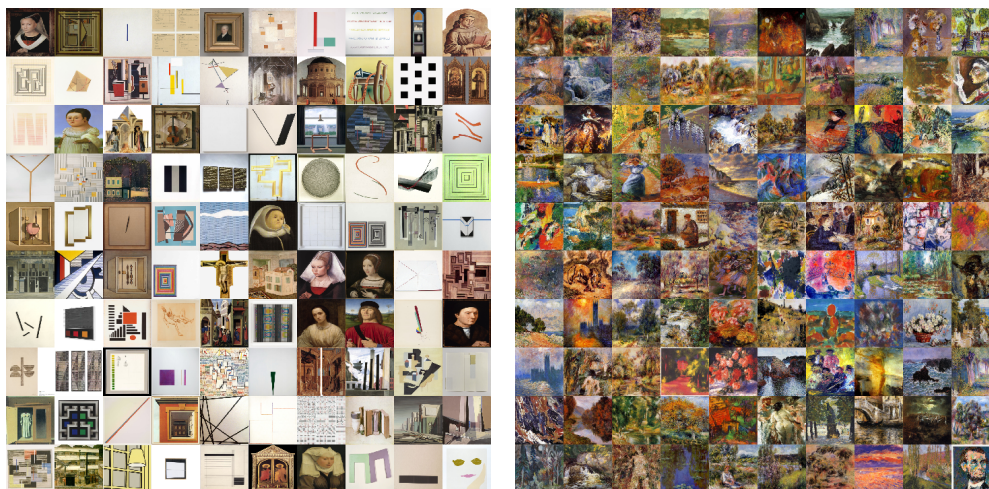
**Slika 4.9:** Toplinske mape (engl. heatmaps) vrijednosti Spearmanovih koeficijenata korelacije između predikcijskih vrijednosti stilskih obilježja dobivenih za slike iz WikiArt kolekcije ( $p$  - vrijednosti  $< 0.001$ ).

Vrijednosti Spearmanovih koeficijenata između predikcijskih vrijednosti pojedinih stilskih obilježja ukazuju na značajnu korelaciju između obilježja linearnosti i slikovitosti (W1) i mnoštva i jedinstva (W4) te apsolutne i relativne jasnoće (W5). S druge strane obilježje plohe i dubine (W3) slabije korelira s vrijednostima drugih obilježja, posebice s obilježjem otvorene i zatvorene forme (W2).

#### 4.4.1 Linearnost i slikovitost

Zbog provjere dobivenih predikcijskih vrijednosti s obzirom na definiciju značajki stilskog obilježja linearnosti i slikovitosti, analiziraju se kvalitativni rezultati zasnovani na prikazu slika s najvećim i najmanjim vrijednostima stilskog obilježja W1. Prema definiciji mjere obilježja W1, niske vrijednosti označavaju slike izrazito linearnog stila, a visoke vrijednosti označavaju do-

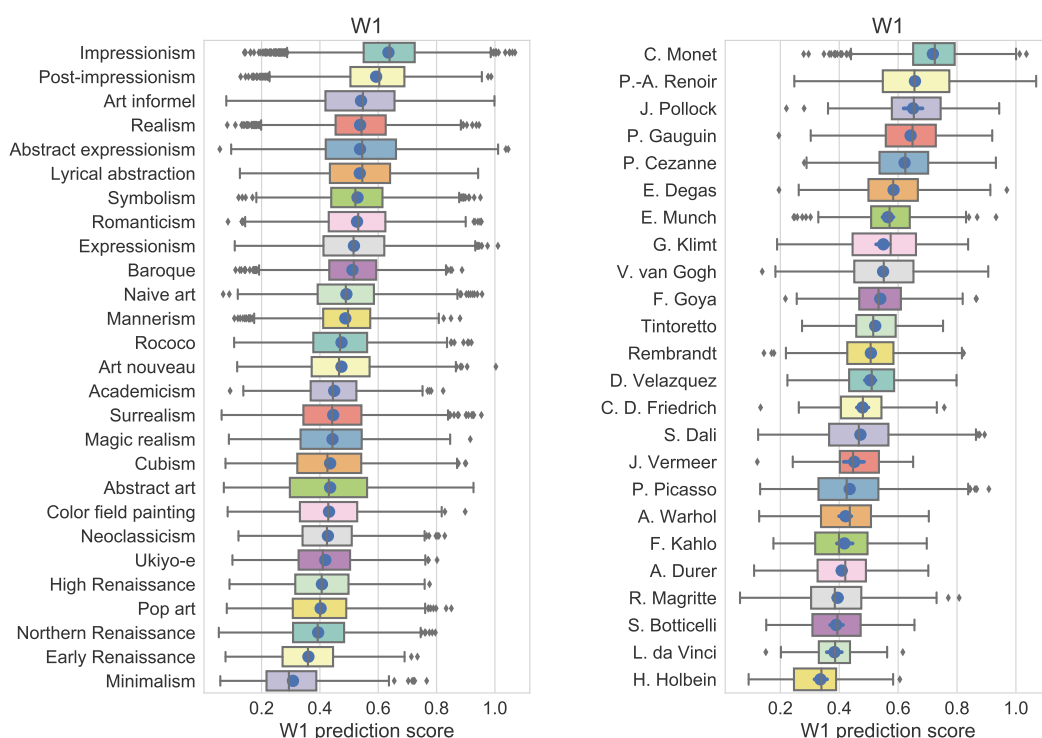
minantno slikoviti stil. Na slici 4.10 prikazano je 100 slika s najmanjim i 100 slika s najvećim vrijednostima predikcije stilskog obilježja odnosa linearnosti i slikovitosti.



**Slika 4.10:** 100 slika iz WikiArt baze s najvišim (lijevo) i najnižim (desno) predikcijskim vrijednostima obilježja linearnosti i slikovitosti (W1).

U slikama s najmanjom vrijednosti obilježja W1 upadljiva je jednostavnost, strogo omeđivane oblika i jasno istaknuti rubovi, što je u skladu s teorijskim opisom linearnog stila. Također, prisutne su slike iz različitih i kronološki udaljenih stilskih razdoblja, kao što su renesansa i minimalizam. Iako te slike pripadaju različitim stilovima i nisu vizualno slične, povezuje ih obilježje linearnosti. Takav ishod ukazuje na rubusnost i nepristranost naučenog modela s obzirom na sadržaj slike. Slike s najvećom predikcijskom vrijednošću obilježja W1 izrazito su slikovitog stila s obzirom na teksturu i potez kista kojim se postiže vizualni dojam koji Wölfflin naziva „treperavi privid“ [63]. Primjeri slika s najmanjim i najvećim vrijednostima obilježja W1 ukazuju na da je model regresijske konvolucijske neuronske mreže uspješno naučio razlikovati linearni i slikoviti stil, unatoč tome što je treniran na relativno malom skupu slika.

Dodatna evaluacija rezultata predikcija vrijednosti obilježja W1 provedena je ispitivanjem odnosa srednjih vrijednosti predikcija s obzirom na kategorije stila i autora slika. Kako bi se istražio odnos predikcijskih vrijednosti obilježja W1 i stilskih pravaca te pojedinih autora, odabrana su dva podskupa WikiArt baze. Prvi podskup sadrži slike 27 različitih stilskih pravca koji su korišteni prilikom učenja klasifikacijskog modela, pri čemu je svaki pravac zastupljen s najmanje 800 slika. Drugi podskup sastoji se od slika 24 različita autora. S obzirom na izbor autora, ovaj skup razlikuje se od podskupa korištenog za klasifikaciju. Odabir autora proveden je tako da obuhvati poznate autore koji pripadaju različitim stilovima, a zastupljeni su u WikiArt bazi većim brojem slika, kao i autore čije radove Wölfflin spominje kao reprezentativne primjere određenih obilježja. Na slici 4.11 prikazani su dijagrami pravokutnika (engl. box-plot) distribuciju predikcijskih vrijednosti stilskog obilježja W1 s obzirom na pojedine stilske pravce i autore, pri čemu je plavom točkom označena srednja vrijednost za svaku kategoriju.



**Slika 4.11:** Box plot distribucija i poredak srednjih vrijednosti predikcija obilježja W1 s obzirom na stilske pravce i opuse različitih autora.

Iz distribucija i srednjih vrijednosti predikcijskih vrijednosti W1 obilježja vidljivo je da su minimalizam i renesansa stilovi s najnižom srednjom vrijednošću predikcija W1 obilježja, dok su impresionizam i post-impresionizam stilovi s najvišom vrijednošću. Poredak srednjih vrijednosti u skladu je s kvalitativnim rezultatima, kao i s opažanjima iznesenim u Wölfflinovoj knjizi [63]. Posebno treba istaknuti poredak s obzirom na odnos renesanse i baroka. Srednja vrijednost predikcija obilježja u slikama rane renesanse iznosi 0.360, sjeverne renesanse 0.393 i visoke renesanse 0.407, dok srednja vrijednost slika koje pripadaju baroknom stilu iznosi 0.511. Poredak autora prema srednjim vrijednostima predikcijskog obilježja W1 također je u skladu s Wölfflinovim opažanjima koji djela Hansa Holbeina ističe kao reprezentativne primjere linearnog stila. Posebno je zanimljiv odnos srednjih vrijednosti slika Leonarda da Vincija (0.383) i Sandra Botticellija (0.389), budući da su i jedan i drugi predstavnici linearnog stila, no prema Wölfflinu „Leonardo je mnogo linearniji od Botticellija“ [63]. Time se pokazuje da je model naučio razlikovati suptilne razine obilježja na način koji je u skladu s povijesno-umjetničkim shvaćanjem linearnosti i slikovitosti.



#### 4.4.2 Zatvorena i otvorena forma

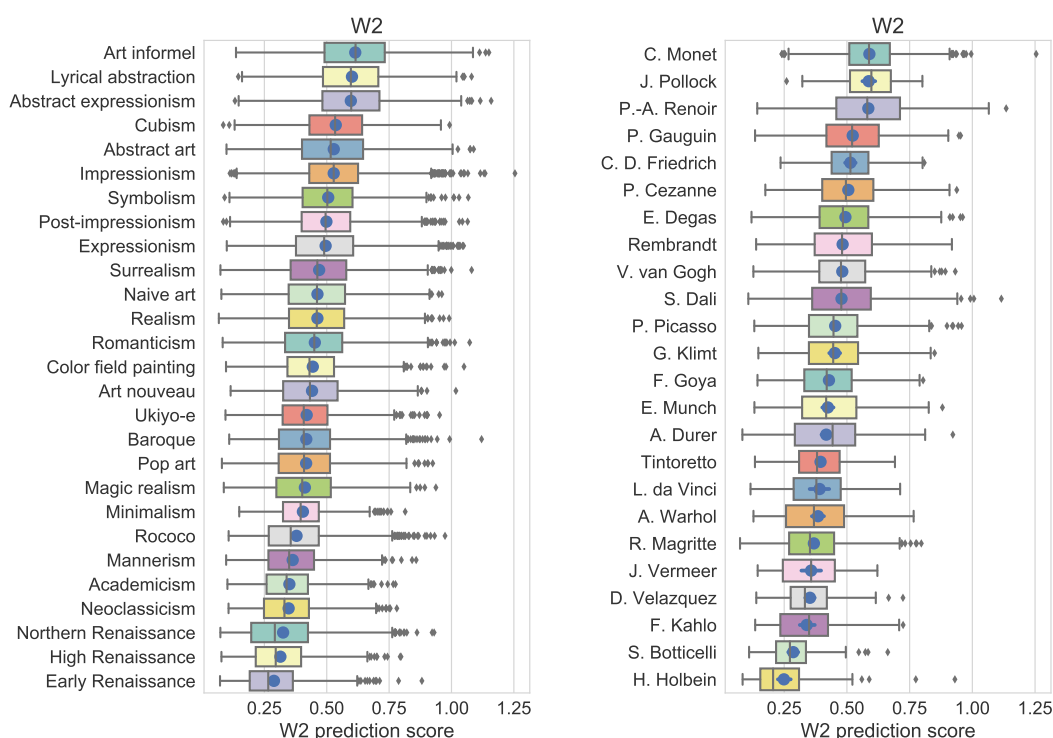
Prema definiciji mjere obilježja  $W_2$ , niske vrijednosti označavaju slike zatvorene forme (tektonski stil), a visoke vrijednosti označavaju slike otvorene forme (atektonski stil). Odnos tektonskog i atektonskog stila može se prije svega promatrati kao odnos omeđenosti i neomeđenosti slikarskog prikaza. Na slici 4.12 prikazano je 100 slika s najmanjim i 100 slika s najvećim vrijednostima predikcije stilskog obilježja odnosa zatvorenosti i otvorenosti forme.



**Slika 4.12:** 100 slika iz WikiArt baze s najvišim (lijevo) i najnižim (desno) predikcijskim vrijednostima obilježja zatvorene i otvorene forme ( $W_2$ ).

U primjerima slika s najmanjom vrijednosti obilježja  $W_2$  upadljiva karakteristika je uokvirenosti ili zaokruženost prikaza, kao i snažno izražene vertikalne i horizontalne linije te centralni smještaj glavnog motiva. U skupu slika s najvećom predikcijskom vrijednosti  $W_2$  obilježja prevladavaju apstraktna djela s dominantnim dijagonalnim usmjerenjem linija i raspored uzoraka koji stvara dojam kao da „izlazi“ iz slike, odnosno presijeca okvir slike. Iako se razlika slika s niskim i visokim  $W_2$  vrijednostima može promatrati i na razini razlike linearnosti i slikovitosti, usporedba primjera slika prikazanih na slici 4.10 i slika prikazanih na slici 4.12, ukazuje na dominantno obilježje omeđenosti kod slika s niskom predikcijskom vrijednosti  $W_2$  obilježja. Na slici 4.13 prikazane su box-plot distribucije predikcijskih vrijednosti stilskog obilježja  $W_2$  s obzirom stilske pravce i autore.

Srednje vrijednosti predikcija obilježja  $W_2$  očekivano su najniže za stilsko razdoblje renesanse, a najviše za apstraktne stilove poput informela (0.61), lirske apstrakcije (0.60) i apstraktnog ekspresionizma (0.59). Poredak autora sličan je kao i za  $W_1$  obilježje, gdje najveću srednju vrijednost obilježja imaju djela Moneta i Renoira, koje osim izražajne slikovitosti karakterizira i dojam zaustavljenosti prolaznog trenutka i prikaz prizora koji pripada širem kontekstu, što je vrlo svojstveno atektonskom stilu.



**Slika 4.13:** Box plot distribucija i poredak srednjih vrijednosti predikcija obilježja W2 s obzirom na stilske pravce i opuse različitih autora.

#### 4.4.3 Ploha i dubina

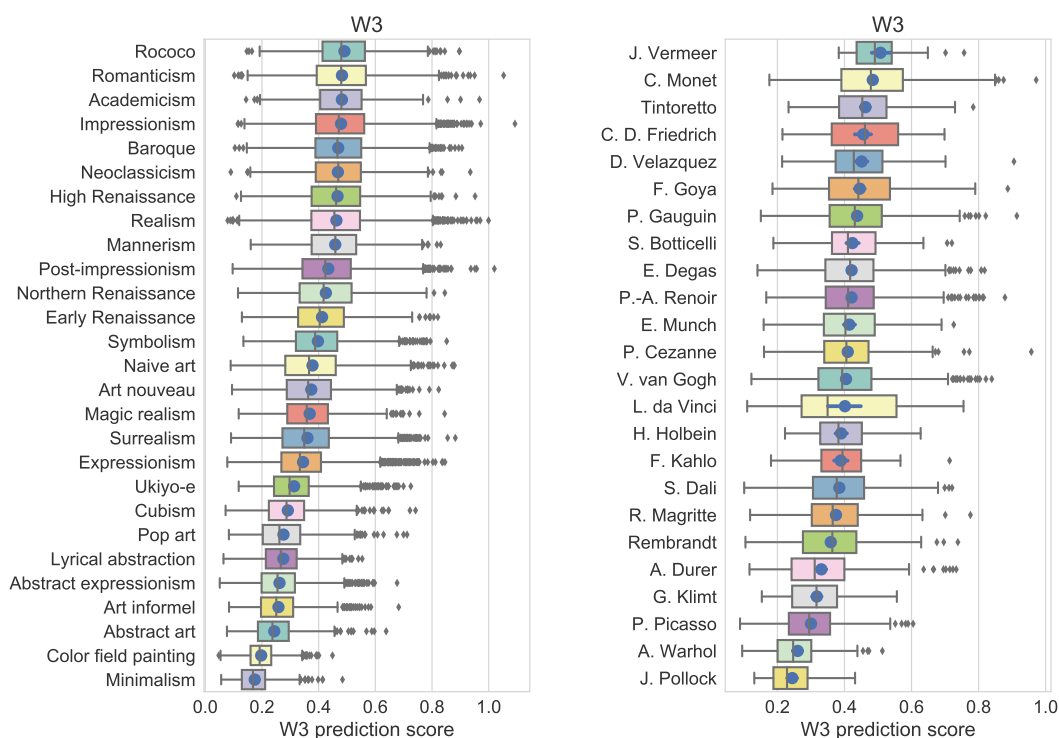
Obilježje označeno kao W3 predstavlja mjeru odnosa između plošnosti i dubine prikaza u slici, pri čemu niske vrijednosti znače izrazitu plošnost, a visoke snažno ostvaren dojam dubine. Na slici 4.14 prikazano je 100 slika s najmanjim i 100 slika s najvećim vrijednostima predikcije stilskog obilježja odnosa plošnosti i dubine.

Slike s najmanjom vrijednošću obilježja W3 posjeduju jedno zajedničko dominantno svojstvo koje se očituje u tome da su svi elementi smješteni na istoj ravnini. Takav ekstremni primjer plošnosti karakterističan je za apstraktne slike minimalizma i slikarstva obojenog polja. Nasuprot tome, kod slika s najvećom vrijednosti W3 obilježja prevladavaju pejzaži s izrazito naglašenim optičkim dojmom daljine.

Na slici 4.15 prikazane su box-plot distribucije predikcijskih vrijednosti stilskog obilježja W3 s obzirom na stilske pravce i autore. S obzirom na srednju vrijednost predikcija obilježja W3, poredak stilova od najnižeg prema najvišem prati razinu uslojavanja ravnina slike. Tako, primjerice, kod slika minimalizma i slikarstva obojenog polja postoji samo jedna ravnina. Prijelaz u dubinski stil ostvaruje se uslojavanjem u paralelne ravnine, dok se kod potpuno dubinskog stila gubi dojam razmještaja elemenata po ravninama i stvara se dojam prožetosti prikaza po



**Slika 4.14:** 100 slika iz WikiArt baze s najvišim (lijevo) i najnižim (desno) predikcijskim vrijednostima obilježja plohe i dubine (W3).



**Slika 4.15:** Box plot distribucija i poredak srednjih vrijednosti predikcija obilježja W3 s obzirom na stilske pravce i opuse različitih autora.

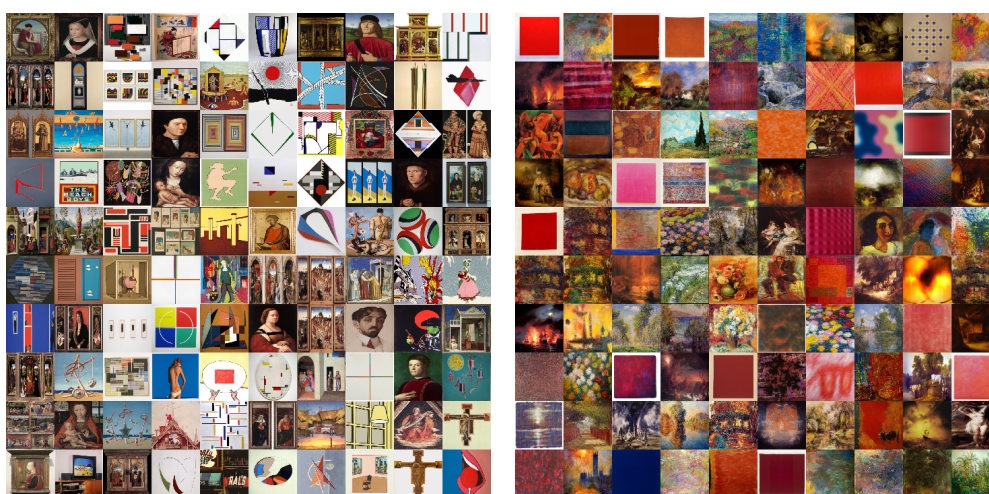
dubini. U razdoblju renesanse ne postoji zahtjev da sve mora biti smješteno na jednoj ravni, no glavni oblici moraju ležati na istoj ravni. Stoga su srednje vrijednosti stilova renesanse veće od srednjih vrijednosti apstraktnih stilova kod kojih prevladava potpuno odsuće uslojavanja. Poredak opusa pojedinih autora u skladu je s njihovim karakterističnim svojstvima odnosa plošnosti i dubine. To se naročito očituje u tome da djela Vermeera, kod kojih je dubinski efekt



posebno izoštren i naglašen, imaju najvišu srednju vrijednost predikcija W3 obilježja.

#### 4.4.4 Mnoštvo i jedinstvo

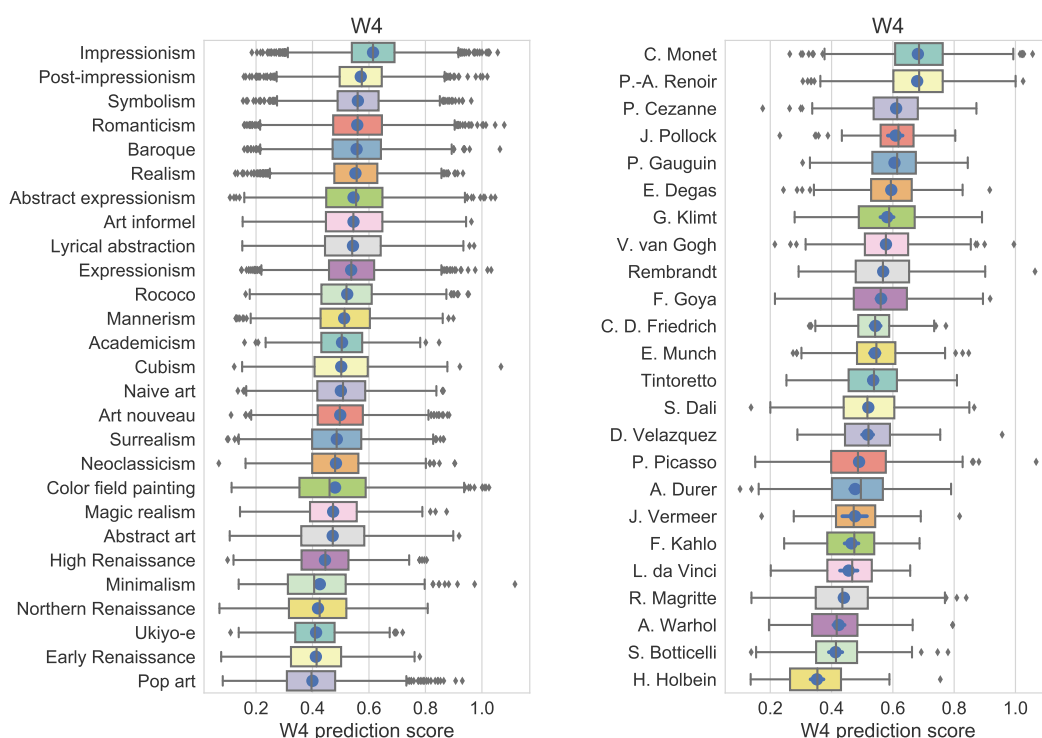
Obilježje označeno kao W4 predstavlja mjeru između mnoštva i jedinstva koja je definirana odnosom dijelova prema cjelini. Niske vrijednosti obilježja W4 karakteristične su za slike kod kojih prevladava princip mnoštva koji se očituje isticanjem većeg broja pojedinačnih elemenata. S druge strane, u slikama visoke vrijednosti obilježja W4 ukida se izdvojenost pojedinačnih oblika te se stvaranjem jedinstvenog toka elementi podređuju dominantnom motivu. Na slici 4.16 prikazano je 100 slika s najmanjim i 100 slika s najvećim vrijednostima predikcije stilskog obilježja odnosa mnoštva i jedinstva.



**Slika 4.16:** 100 slika iz WikiArt baze s najvišim (lijevo) i najnižim (desno) predikcijskim vrijednostima obilježja plohe i dubine (W4).

U primjerima slika s najmanjom vrijednosti obilježja W4 očituje se zajednička karakteristika razgovijetne razdvojenosti većeg broja pojedinačnih elemenata, dok se kod slika s visokom vrijednošću W4 obilježja uočavaju dvije vizualno različite kategorije slika kod kojih se na drugačiji način ostvaruje svojstvo jedinstva. Prva kategorija pripada impresionističkim slikama kod kojih se slikovitost i detaljna razrada teksture postavlja u službu stvaranja jedinstvenog toka likovnih elemenata. U drugoj su kategoriji slike slikarstva obojenog polja kod kojih se princip jedinstva ostvaruje minimalističkim prikazom jednog plošnog elementa. Distribucija predikcijskih vrijednosti stilskog obilježja W4 s obzirom na stilske pravce i autore prikazana je na slici 4.17.

Srednje vrijednosti predikcija obilježja W4 najniže su za djela koja pripadaju pop artu, koji s radovima rane renesanse dijeli karakteristično obilježje mnoštva pojedinačno istaknutih elemenata. Poništenje istaknutosti pojedinih oblika i podređivanje dijelova sustavu cjeline najočitije je u djelima impresionizma. Sukladno tomu je i srednja vrijednost predikcija W4 obilježja najviša upravo za slike impresionizma. Poredak stilova i autora za srednje vrijednosti obilježje W4



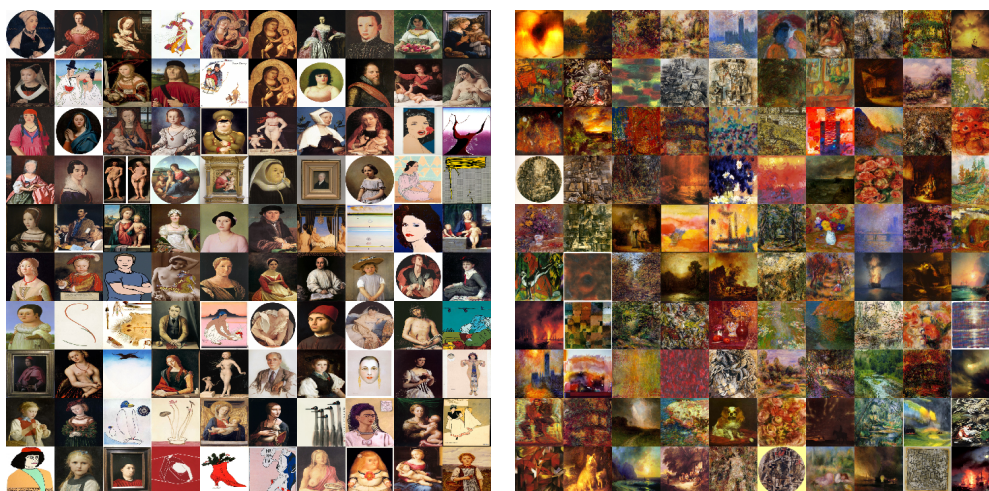
**Slika 4.17:** Box plot distribucija i poredak srednjih vrijednosti predikcija obilježja W4 s obzirom na stilske pravce i opuse različitih autora.

sličan je poretku za vrijednosti obilježja W1, što logično proizlazi iz visoke korelacije predikcijskih vrijednosti tih dvaju obilježja dobivene na djelima u WikiArt baze.

#### 4.4.5 Jasnost i nejasnost

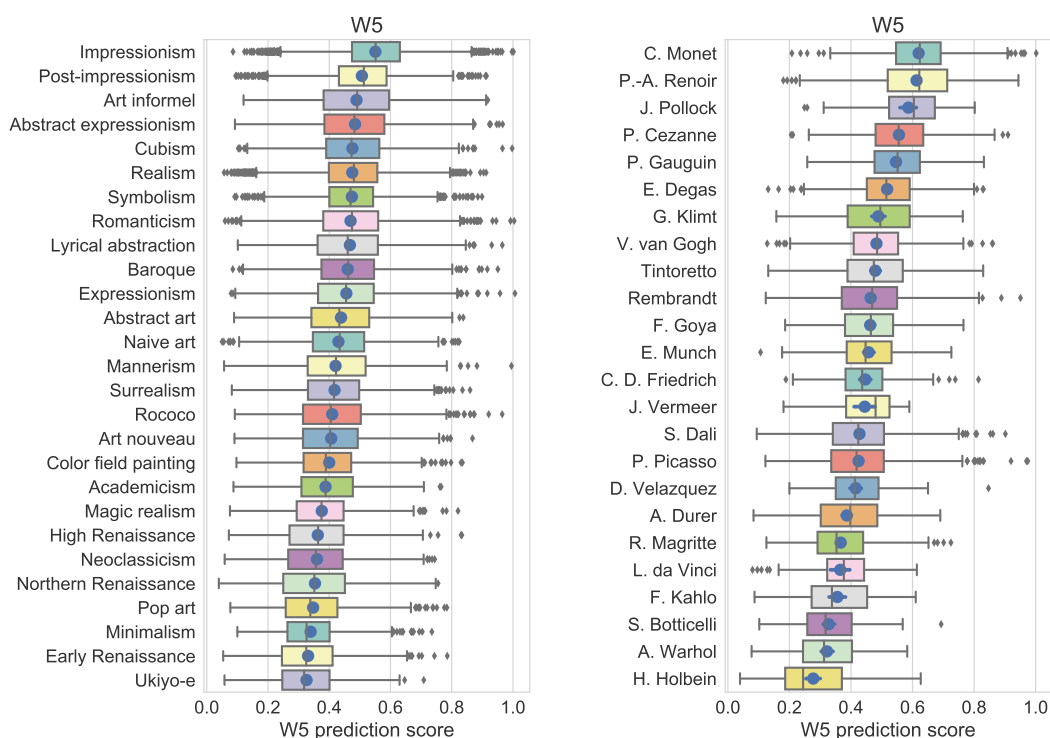
Obilježje jasnosti i nejasnosti određeno je oprekom prizora stvari kakve jesu i prikaza stvari kakvima se čine. Slike s niskom vrijednošću obilježja W5 odlikuje jasnoća i razgovjetnost motiva, dok se kod slika niske vrijednosti W5 obilježja pokazuju tek naznake motiva. Na slici 4.18 prikazano je 100 slika s najmanjim i 100 slika s najvećim vrijednostima predikcije stilskog obilježja odnosa jasnosti i nejasnosti.

Wölfflin ističe da se obilježje odnosa jasnosti i nejasnosti preklapa s obilježjem linearnosti i slikovitosti, što je vidljivo iz sličnosti primjera slika s najvećom i najmanjom vrijednošću tih obilježja. Svi slikoviti motivi nastaju određenim umanjivanjem jasnosti, pa su stoga slike s najvećom vrijednosti W5 obilježja izrazito pretežito impresionistička djela naglašeno slikovitog stila. Nasuprot tome, usporedba slika s najmanjim vrijednostima W1 i W5 obilježja pokazuje da, iako se u oba slučaja radi o slikama isključivo linearnog stila, načelo jasnosti dolazi do izražaja u tome što se kod slika s niskom vrijednošću W5 obilježja gotovo ne pojavljuju apstraktne



**Slika 4.18:** 100 slika iz WikiArt baze s najvišim (lijevo) i najnižim (desno) predikcijskim vrijednostima obilježja plohe i dubine (W5).

slike. Takav ishod kvalitativnih rezultata u skladu je s karakteristikama obilježja apsolutne jasnosti koje osim razgovijetnosti forme obuhvaća i jasnoću motiva. Distribucija predikcijskih vrijednosti stilskog obilježja W5 s obzirom stilske pravce i autore prikazana je na slici 4.19.



**Slika 4.19:** Box plot distribucija i poredak srednjih vrijednosti predikcija obilježja W5 s obzirom na stilske pravce i opuse različitih autora.

Poredak stilskih pravaca i opusa autora s obzirom na srednje vrijednosti predikcija obilježja

W5 sličan je poretku predikcija obilježja W1. Očekivano su niske vrijednosti za razdoblje renesanse, kao i za djela Hansa Holbeina, dok su najveće srednje vrijednosti dobivene za djela iz razdoblja impresionizma i djela iz Monetovog opusa.

## 4.5 Osvrt na rezultate istraživanja

U ovom poglavlju predstavljena je metoda kvantifikacije zastupljenosti specifičnih stilskih obilježja u slikarskim djelima i predikcija vrijednosti tih obilježja zasnovana na učenju regresijskih modela konvolucijskih neuronskih mreža. Pod specifičnim stilskim obilježjima u ovom kontekstu podrazumijevaju se značajke oblikovane po uzoru na umjetničke koncepte oblikovanja likovnog djela. U fokusu opisanog istraživanja nalaze se Wölfflinovi koncepti koji su opisani u knjizi *Temeljni pojmovi povijesti umjetnosti* [63] i predstavljaju ključne odrednice povijesno-umjetničkog razlikovanja stilova i formalne analize likovnih djela. Budući da se radi o semantički složenim i vizualno suptilnim razinama razlikovanja slika, računalno raspoznavanje i predikcija takvih obilježja u slici predstavlja novi izazov u kontekstu računalnog vida. U ovome radu pokazano je da se korištenjem dubokih konvolucijskih neuronskih mreža mogu naučiti raspoznavati takve složene značajke slike, čak i kada je skup slika za učenje relativno malen. S obzirom da se radi o značajkama za čije je označavanje potrebno stručno znanje, broj označenih slika je ograničen i modeli konvolucijskih mreža učeni su na relativno malom skupu slika. Budući da se radi o manjem skupu slika, treniranje mreže nije provedeno tako da se modificiraju svi parametri, kako ne bi došlo do prenaučivosti mreže, već su primjenom metode finog podešavanja parametara modificirani parametri samo posljednja tri potpuno povezana sloja modela mreže prethodno naučene za klasifikaciju stilova. Klasifikacijski sloj bazne mreže zamijenjen je regresijskim slojem te je za svaki koncept naučen zasebni model. Reprezentativnost naučenih modela ispitana je analizom kvantitativnih i kvalitativnih rezultata. Kvantitativni rezultati zasnovani su na izračunu koeficijenta determinacije i srednje kvadratne pogreške predikcijskih vrijednosti pojedinih obilježja i stvarnih vrijednosti na testnom skupu podataka. S obzirom na mali broj slika u skupu za testiranje, provjera reprezentativnosti modela dodatno je utvrđena provjerom kvalitativnih rezultata, odnosno analizom vizualnih obilježja slika s niskim i visokim predikcijskim vrijednostima pojedinih obilježja dobivenih na velikom skupu slika iz WikiArt baze. Također, provedena je analiza odnosa predikcijskih vrijednosti pojedinih stilskih obilježja i različitih stilskih pravaca i autora te je utvrđena konzistentnost dobivenih rezultata s Wölfflinovim teorijskim opažanjima. Ishod ovoga istraživanja ukazuje da se unatoč dostupnosti relativno malog skupa podataka za učenje, metodom finog podešavanja parametara konvolucijskih neuronskih mreža mogu postići vrlo dobri rezultati u raspoznavanju kompleksnih značajki slike. Time se otvaraju mogućnosti razvoja sustava za pretraživanje velikih kolekcija slikarskih djela na temelju složenih upita koji su usklađeni s povijesno-umjetničkom terminologijom.

Također, rezultati ovog istraživanja ukazuju na potencijalne smjernice razvoja discipline digitalne povijesti umjetnosti, posebice u smjeru povratka metodama formalne analize sa svrhom otkrivanja uzoraka u velikim skupovima podataka.

## Poglavlje 5

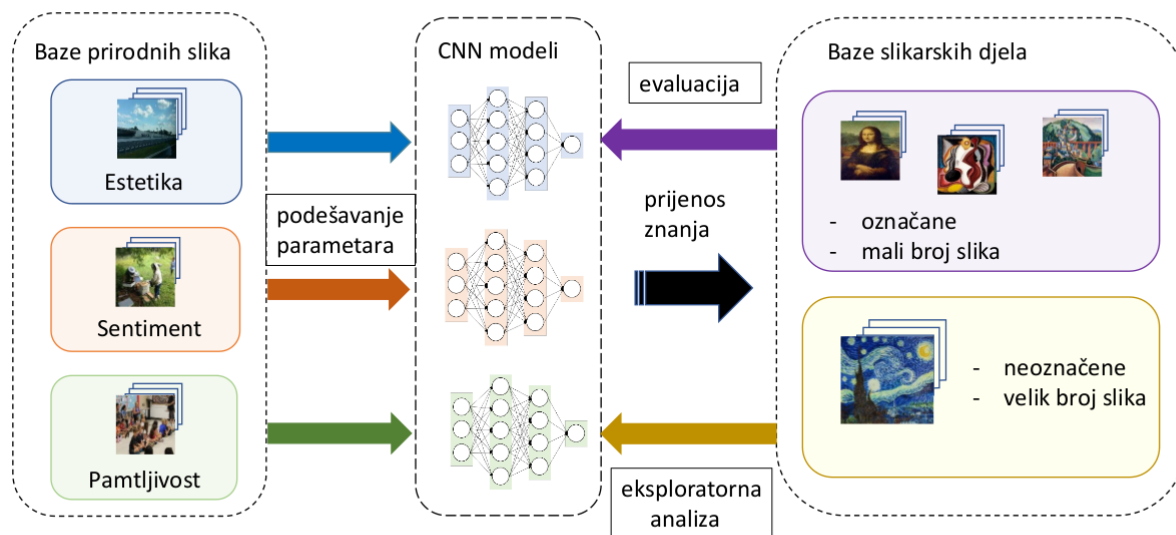
# Računalna predikcija percepcijskih obilježja slikarskih djela

### 5.1 Pozadina i motivacija istraživanja

Rezultati prikazani u trećem poglavlju ovog rada pokazuju da se korištenjem konvolucijskih neuronskih mreža postiže značajno bolja točnost u različitim zadacima klasifikacije slikarskih djela. Svladavanjem izazova automatske klasifikacije slikarskih djela postavlja se pitanje primjene konvolucijskih neuronskih mreža za probleme koji nadilaze razinu raspoznavanja i odnose se na razinu razumijevanja slike. Stoga se u okviru ovoga rada istražuje potencijal dubokih konvolucijskih neuronskih mreža u oblikovanju novih metoda istraživanja velikih skupova digitaliziranih slikarskih djela i otkrivanju uzoraka ponašanja kompleksnih značajki. S ciljem ispitivanja primjenjivosti računalnih metoda za analizu doživljajne i interpretativne razine slikarskih djela, istraživanje opisano u ovom poglavlju odnosi se na primjenu konvolucijskih neuronskih mreža u otkrivanju i predviđanju značajki percepcijskih obilježja slike. Istraživanje je usredotočeno na predviđanje značajki visoke razine apstrakcije koje se odnose na različite aspekte percepcije slike: estetsku evaluaciju slike, osjećajni doživljaj slike i pamtljivost slike. Ta tri aspekta percepcije slike već su dugi niz godina predmet različitih istraživanja iz područja psihologije, a nedavno su postali predmet interesa u području računalnog vida. Zbog veće dostupnosti velikih baza digitalnih fotografija, najveći dio istraživanja usmjeren na računalnu analizu ovih triju percepcijskih obilježja slika proveden je na prirodnim slikama, dok primjena računalnih metoda u otkrivanju tih obilježja u slikarskim djelima do sada još nije sustavno istražena. Pojava velikih i dobro označenih baza digitaliziranih umjetničkih djela otvara mogućnost za analizu percepcijskih obilježja slikarskih djela i njihovog odnosa sa specifičnim stilskim karakteristikama na velikom skupu podataka. Provođenje ispitivanja koje uključuju prikupljanje oznaka subjektivne procjene navedenih percepcijskih obilježja složeno je i skupo. Međutim, budući da su ispitivanja takve vrste već provedena na nekim skupovima prirodnih slika i da



su na temelju toga naučeni modeli dubokih neuronskih mreža koji mogu prepoznati određena percepcijska obilježja u prirodnim slikama, otvara se mogućnost primjene tih modela na digitaliziranim slikarskim djelima. Slika 5.1 ilustrira metodološki koncept prijenosa znanja usvojen u ovome radu. Istraživanje i rezultati opisani u ovome poglavlju predstavljani su u [64].



**Slika 5.1:** Konceptualni pregled metodologije prijenosa znanja.

Za dobivanje značajki percepcijskih obilježja korišteno je nekoliko različitih modela konvolucijskih neuronskih mreža naučenih za svaki od tri predikcijska zadatka – predviđanje estetskog vrednovanja, vizualnog sentimenta i pamtljivosti slike. Različiti modeli naučeni za predviđanje percepcijskih obilježja u prirodnim slikama primjenjeni su za dobivanje predikcijskih vrijednosti tih istih obilježja u slikarskim djelima. Sa svrhom provjere konzistentnosti dobivenih vrijednosti, provedena je korelacijska analiza vrijednosti dobivenih iz različitih modela naučenih za isti zadatak. Također, evaluacija različitih modela provedena je na temelju korelacijske analize između predikcijskih vrijednosti i stvarnih vrijednosti, dostupnih na manjim kolekcijama slikarskih djela. Pod „stvarnim“ vrijednostima podrazumijevaju se srednje vrijednosti subjektivnih ocjena percepcijskih obilježja slika dobiveni anketnim ispitivanjem određenog broja ispitanika. Na temelju navedene korelacijske analize, za svaki zadatak odabran je najbolji model za predikciju percepcijskih obilježja na velikom skupu slika. Zatim je provedena analiza odnosa dobivenih vrijednosti pojedinih percepcijskih obilježja i drugih značajki slike kako bi se utvrdilo koje značajke slike najviše doprinose određenim percepcijskim obilježjima. Također je istraženo kako se vrijednosti pojedinih percepcijskih obilježja odnose prema pojedinim povijesno-umjetničkim kategorijama slikarskih djela kao što su stil i žanr. U cilju boljeg razumijevanja istraživačkog konteksta, u nastavku je izložen kratak pregled triju područja istraživanja koji opisuju razvoj računalnih metoda za detekciju odabranih percepcijskih obilježja slike.

### 5.1.1 Računalna estetika

Računalna estetika (engl. computational aesthetics) rastuće je područje interesa u domeni računalnog vida i sve veći broj istraživanja usmjeren je na razvoj računalnih metoda koje mogu predvidjeti estetske prosudbe slika koje odgovaraju ljudskim procjenama. Većina istraživanja u ovome području usmjerena je na predikciju estetskog vrednovanja fotografija, dok mogućnosti primjene tih metoda na slikarska djela još nisu sustavno istražene. Kao i u većini drugih područja računalnog vida, metode računalne estetike ispočetka su se temeljile na ekstrakciji različitih oblikovanih značajki (engl. hand-crafted image features) i primjeni različitih metoda strojnog učenja [65, 66, 67]. Pri tome su se kao koristile baze slika koje su označene subjektivnim procjenama estetske vrijednosti slike dobivenima anketnim ispitivanjem većeg broja sudionika [68], pri čemu se uspješnost modela utvrđivala na temelju korelacije dobivenih prediktivnih vrijednosti i subjektivnih ocjena. Nedavnim istraživanjima pokazano je da se korištenjem značajki dobivenih iz konvolucijskih neuronskih mreža dobiva značajno poboljšanje predviđanja estetske vrijednosti slike [69], a posebno kada se koristi metoda finog podešavanja parametara konvolucijske mreže [70, 71]. Detaljan pregled razvoja metoda računalne estetike prikazan je u [72, 73].

### 5.1.2 Računalno predviđanje vizualnog sentimenta

S povećanjem broja digitalnih slika i sve većim utjecajem društvenih mreža analiza vizualnog sentimenta (engl. visual sentiment analysis) postaje sve aktivnije područje istraživanja u kontekstu računalnog vida. Vizualna analiza sentimenta podrazumijeva širok spektar istraživačkih radova, a razlikuju se dva temeljna usmjerenja: raspoznavanje osjećaja izraženog u izrazima lica ili tjelesnim gestama prikazanim na slici; ili predviđanje osjećaja koje određeni sadržaj u slici izaziva kod promatrača. Pod analizom vizualnog sentimenta u ovom radu se podrazumijeva ovo potonje, pri čemu se sentiment pojednostavljeno svodi na dvije polarnosti, odnosno na određivanja izaziva li slika negativan ili pozitivan osjećaj u promatraču. U posljednjih nekoliko godina razvijeno je nekoliko pristupa računalnoj predikciji tako shvaćenog vizualnog sentimenta. U većini prijašnjih radova uobičajeni se pristup temeljio na povezivanju značajki slike niže razine sa značajkama više razine koje predstavljaju složenije vizualne attribute [74] i treniranju binarnog klasifikatora na temelju eksperimentalno dobivenih subjektivnih oznaka vizualnog sentimenta [75]. Nedavne metode utemeljene na korištenju konvolucijskih neuronskih mreža za predikciju vizualnog sentimenta pokazuju se mnogo uspješnijima [76, 77, 78].

### 5.1.3 Računalno predviđanje pamtljivosti slike

Pamtljivost slike (engl. image memorability) je koncept koji se odnosi na pitanje koliko je jednostavno zapamtiti određenu sliku. Iako je pamtljivost subjektivna kategorija, rezultati psi-



holoških istraživanja pokazuju da većina ljudi pamti i zaboravlja slične slike. To ukazuje na to da su određene vizualne karakteristike univerzalno pamtljivije od drugih. Odnedavno je pitanje pamtljivosti slike postalo predmetom istraživanja u području računalnog vida kada su Isola et al. [79] prvi razvili model za predviđanje faktora pamtljivosti slike na temelju skupa različitih značajki slike. Preduvjet za mogućnost takvog istraživanja bilo je postojanje baze slika s oznakama mjere pamtljivosti koje su prikupljene eksperimentalnim ispitivanjem većeg broja sudionika kroz posebno oblikovanu memorijsku igru. Na temelju prikupljenih oznaka Isola et al. korištenjem su regresije potpornim vektorima (engl. support vector regression) predviđali faktore pamtljivosti na temelju reprezentacije slike koja je obuhvaćala veći broj različitih oblikovanih značajki. Također su istraživali odnos između značajki slike i faktora pamtljivosti te su utvrdili da ne postoji značajna korelacija između pamtljivosti slike i jednostavnih značajki kao što su deskriptori boje ili teksture i da sadržaj igra veliku ulogu u određivanju pamtljivosti slike. Naime, pokazano je da fotografije ljudi imaju znatno viši faktor pamtljivosti od fotografija pejzaža i prirode. Nakon rezultata predstavljenih u [79], uslijedili su drugi pristupi računalnoj predikciji faktora pamtljivosti slike korištenjem drugačijih značajki slike [80, 81, 82]. Sveobuhvatan pregled područja istraživanja računalnog predviđanja pamtljivosti slike dan je u [83]. Metoda predikcije pamtljivosti slike utemeljena na korištenju konvolucijskih neuronskih mreža predstavljena je u radu Khosle et al. [84]. U tom radu predstavljena je i do tada najveća baza slika koja sadrži oznake subjektivne procjene pamtljivosti slike i koja je korištena za fino-podešavanje konvolucijske neuronske mreže MemNet. Tim pristupom postignut je rezultat predikcije pamtljivosti slike s koeficijentom korelacije ranga 0.64, pri čemu koeficijent korelacije ranga među eksperimentalno prikupljenim ocjenama pamtljivosti slike između različitih sudionicima iznosi 0.68. Dodatno poboljšanje postignuto je AMNet modelom [85], kojim se dobiva koeficijent korelacije ranga 0.67 na LaMem bazi slika. Primjena konvolucijskih neuronskih mreža za predikciju pamtljivosti slikarskih djela prvi puta je predstavljena u [86].

## 5.2 Postavke eksperimenta

U ovome poglavlju opisane su osnovne postavke eksperimenta: baza slikarskih djela i modeli korišteni za predviđanje percepcijskih obilježja slika iz baze.

### 5.2.1 Baze slika

U ovome poglavlju izložen je detaljan opis svih baza prirodnih slika i likovnih djela korištenih u postupku predikcije i analize percepcijskih obilježja slikarskih djela. Različite baze slika korištene su u tri različite faze: (1) učenje modela konvolucijskih neuronskih mreža korištenjem velikih baza označenih prirodnih slika; (2) evaluacija predikcijskih vrijednosti percepcijskih

obilježja usporedbom sa stvarnim vrijednostima na manjim bazama označenih slikarskih djela; (3) analiza odnosa predikcijskih vrijednosti pojedinih percepcijskih obilježja, stilskih obilježja i povijesno-umjetničkih kategorija provedena na velikoj bazi neoznačenih slikarskih djela. Popis svih baza slika dan je u Tablici 5.1

**Tablica 5.1:** Popis svih baza slika korištenih u postupku predikcije i analize percepcijskih obilježja slikarskih djela. Za svaki skup slika označena je faza u kojoj se koristi, odgovarajuće obilježje, broj slika u bazi, broj prikupljenih subjektivnih ocjena po slici i tip slika u skupu podataka: umjetničke slike (U) ili prirodne slike (P).

Faza	Koncept	Baza slika	# slika	# ocjena po slici	Tip slika
	Estetika				
Analiza	Sentiment	WikiArt	105 121	-	U
	Pamtljivost				
Evaluacija	Estetika	JenAesthetic[87]	1568	~20	U
Evaluacija	Sentiment	MART [88]	500	20	U
Evaluacija	Sentiment	WikiEmotions [89]	3379	~10	U
Učenje mreže	Estetika	AADB [71]	10 K	5	P
Učenje mreže	Estetika	AVA [68]	250 K	~210	P
Učenje mreže	Estetika	FLICKR-AES [90]	40 K	5	P
Učenje mreže	Sentiment	Twitter DeepSent [77]	1269	5	P
Učenje mreže	Sentiment	Flickr Sentiment [91]	10 K	3	P
Učenje mreže	Pamtljivost	LaMem [84]	58 741	~80	P
Učenje mreže	Pamtljivost	SUN Memorability [92]	2222	~78	P

Analize odnosa između računalno predviđenih percepcijskih obilježja i stilskih obilježja slikarskih djela te povijesno-umjetničkih kategorija, provedena je nad WikiArt bazom slikarskih djela koja je detaljno opisana u poglavlju 3.2. Predikcija percepcijskih obilježja provedena su na pročišćenom skupu od 105121 slika. Pročišćeni skup sastoji se isključivo od slika koje su kategorizirane kao slikarska djela. Kako bi se mogla provesti analiza utjecaja značajki boje, zadržane su samo slike u boji, odnosno uklonjene su sve crno-bijele digitalne reprodukcije slika.

Za učenje različitih modela konvolucijskih mreža za predikciju vrijednosti pojedinog percepcijskog obilježja korišteno je nekoliko različitih baza prirodnih slika koje sadrže oznake prikupljenih subjektivnih procjena određenoga obilježja. U svrhu učenja modela za predikciju estetske vrijednosti slike korištene su baze AADB [71], AVA [68] i FLICKR-AES [90]. Za predviđanje vizualnog sentimenta korišteni su modeli naučeni na dva različita skupa podataka:

Twitter DeepSent [77] i Flickr Sentiment [91]; dok se za predviđanje pamtljivosti slike također koriste modeli naučeni na dvije baze slike: LaMem [84] i Sun Memorability [92].

Prikupljanje oznaka subjektivne procjene pojedinih percepcijskih obilježja složen je i zahtjevan postupak. Iako je dostupno nekoliko velikih baza prirodnih slika označenih subjektivnim procjenama percepcijskih obilježja u slici, ispitivanja takve vrste rijetko su provedena nad bazama slikarskih djela. Stoga trenutno postoji samo jedna baza slikarskih djela koja sadrži oznake estetske procjene većeg broja ispitanika, nekoliko baza s oznakama vizualnog sentimenta, te niti jedna s oznakama pamtljivosti slike. Također, te baze sadrže značajno manji broj slika nego baze prirodnih slika te stoga nisu pogodne za fino-podešavanje parametara konvolucijske neuronske mreže, već se koriste za evaluaciju predikcijskih vrijednosti dobivenih od modela naučenih na prirodnim slikama. Na temelju Spearmanovog koeficijenta korelacije, koji mjeri povezanost među vrijednostima poredanih varijabli, između različitih predikcijskih i stvarnih vrijednosti, odabiru se modeli za predikciju estetske vrijednosti i vizualnog sentimenta za daljnju analizu predikcijskih vrijednosti na WikiArt bazi slikarskih djela.

Za evaluaciju predikcija estetskog vrednovanja slika korištena je baza JenAesthetic [87, 93, 94, 95] koja sadrži 1568 digitaliziranih slikarskih djela iz jedanaest različitih stilskih razdoblja. Sudionici ispitivanja ocjenjivali su različita obilježja slike kao što su „estetska kvaliteta“, „boja“, „sadržaj“, „kompozicija“, itd. Svaku sliku ocijenilo je prosječno dvadeset ispitanika i medijan tih ocjena korišten je kao konačna vrijednost za svako obilježje.

Za evaluaciju predikcija vizualnog sentimenta korištene su dvije različite baze slika. Prva je MART [88] baza slika koja sadrži 500 isključivo apstraktnih slikarskih djela. Slikama je dodijeljena oznaka sentimenta koja ukazuje na to izaziva li slika pozitivan ili negativan osjećaj u promatraču. Ti podaci dobiveni su ispitivanjem u kojemu su sudionici vrednovali slike na skali od 1 do 7, pri čemu 1 znači izrazito negativan, a 7 izrazito pozitivan osjećaj. Svaka slika ocijenjena je od strane 20 ispitanika i prosječna vrijednost svih ocjena uzeta je kao konačna oznaka sentimenta. Budući da ta baza sadrži samo apstraktne slike, za evaluaciju korištena je još i WikiArt Emotion baza [89]. Ta baza predstavlja manji podskup slika dobivenih iz WikiArt.org kolekcije, točnije 3379 slikarskih djela iz 22 različita stilska razdoblja, koje sadrže oznake različitih kategorija emocija. Ispitanici su slikama dodijeljivali oznake određenih kategorija emocija, kao što su npr. „tuga“, „bijes“, „sreća“, „ljubav“, itd. te je svaku sliku ocijenilo najmanje deset ispitanika.

### 5.2.2 Modeli za predikciju percepcijskih obilježja

Učenje konvolucijskih neuronskih mreža postala je uobičajena praksa za rješavanje mnogih zadataka u području računalnog vida. To je rezultiralo pojavom rastućih zbirki modela predtreiniranih za različite domene i zadatke. Upotreba tih modela ne mora nužno biti jednokratna i ograničena na originalni zadatak, već takvi naučeni modeli predstavljaju izvor alata za prijenos

usvojenog znanja iz jedne domene u drugu. Različita istraživanja pokazala su da modeli naučeni na prirodnim slikama pokazuju vrlo dobre rezultate kada se primjenjuju na umjetničkim slikama sa svrhom detekcije objekata ili lica [33, 96]. Međutim, u ovome radu prvi puta se istražuje primjenjivost modela naučenih na prirodnim slikama s ciljem prepoznavanja značajki slikarskih djela koje nisu vezane uz sadržajna, već percepcijska obilježja slike. Sa svrhom predviđanja estetskog vrednovanja, vizualnog sentimenta i pamtljivosti slike, korišteno je nekoliko različitih modela konvolucijskih neuronskih mreža, predtreniranih na različitim bazama prirodnih slika. U Tablici 5.2 dan je popis svih modela. Za svaki model naznačeno je za koji zadatak je model naučen, koja je temeljna arhitektura modela, na kojoj je bazi slika model naučen, te je li model preuzet kao dostupan rezultat prijašnjih istraživanja ili je naučen u sklopu ovog rada.

**Tablica 5.2:** Popis svih modela korištenih u postupku predikcije i analize percepcijskih obilježja slikarskih djela. Za svaki model naznačeno je za koji zadatak je model naučen, koja je temeljna arhitektura modela, na kojoj bazi slika je model naučen i izvor dobivanja modela (je li preuzet iz prijašnjih istraživanja ili je naučen za potrebe ovog rada).

Zadatak	Naziv modela	Arhitektura	Baza slika	Izvor
Estetika	AestNet_1	AlexNet + atributni slojevi	AADB	[71]
Estetika	AestNet_2	GoogLeNet	AVA	[97]
Estetika	AestNet_3	ResNet50 + LSTM	FLICKR-AES	naš rad
Sentiment	SentiNet_1	AlexNet	Twitter	[78]
Sentiment	SentiNet_2	AlexNet	Flickr Sentiment	naš rad
Sentiment	SentiNet_3	ResNet50 + LSTM	Flickr Sentiment	naš rad
Pamtljivost	MemNet_1	AlexNet	LaMem	[84]
Pamtljivost	MemNet_2	ResNet50 + LSTM	SUN Memorability	[85]
Pamtljivost	MemNet_3	ResNet50 + LSTM	LaMem	[85]

AestNet\_1 je model koji je izvorno predstavljen u radu Kong et al. [71]. Temelji se na AlexNet arhitekturi koja je detaljno opisana u poglavlju 2.1.2. Originalna AlexNet arhitektura modificirana je dodavanjem regresijskog sloja umjesto softmax sloja i dodavanjem slojeva koji uče prepoznavanje određenih atributa u slici. Model je naučen na AADB bazi slika koja sadrži oznake srednje vrijednosti ocjena estetskog vrednovanja slike i oznake ocjena prisutnosti određenih atributa u slici. Budući da je ovaj model korišten i za ekstrakciju značajki određenih atributa u slici, detaljan opis tih atributa i slojeva mreže dan je u poglavlju 5.2.3. Osim navedenog modela, za predikciju estetske vrijednosti slike korišten je i AestNet\_2 model koji su predstavili Xin Jin et al. [97]. Model se temelji se na GoogleLeNet [51] modelu mreže koji je predtreniran za klasifikaciju slika prema sadržaju te zatim metodom finog podešavanja parametara naučen

za klasifikaciju slika na temelju visoke ili niske estetske vrijednosti slike, korištenjem AVA kolekcije prirodnih slika [68]. Treći model korišten za predikciju estetske vrijednosti slike je AestNet\_3 model koji se temelji na složenoj AMNet arhitekturi predstavljenoj u [85]. AMNet arhitektura sastoji se od ResNet50 [52] modela konvolucijske neuronske mreže predtreniranog na ImageNet bazi za klasifikaciju slika prema sadržaju, mehanizma vizualne pažnje (engl. soft attention mechanism) ostvarenog povratnom neuronskom mrežom s dugotrajnim kratkoročnim pamćenjem (engl. Long short-term memory network, LSTM) [98] te regresijske mreže. AMNet model je izvorno dizajniran sa svrhom predviđanja pamtljivosti slike, no implementiran je na generički način tako da se može naučiti i za druge regresijske zadatke. Stoga je AestNet\_3 model nastao kao rezultat treniranja AMNet modela na FLICKR-AES [90] bazi slika, pri čemu Spearmanov korelacijski koeficijent između predikcijskih i stvarnih vrijednosti iznosi 0.72 (p-vrijednost  $< 0.01$ ).

Slično kao i za predikciju estetske vrijednosti slike, AMNet model naučen je i za predikciju vizualnog sentimenta treniranjem na Flickr Sentiment [91] bazi. Rezultat toga učenja je SentiNet\_3 model koji postiže koeficijent korelacije 0.53 (p-vrijednost  $< 0.01$ ) na Flickr Sentiment skupu slika za testiranje. Osim tog modela, korištena su još dva modela naučena za raspoznavanje pozitivnog ili negativnog sentimenta u slici koji se temelje na AlexNet arhitekturi: SentiNet\_2 koji je također naučen na Flickr Sentiment bazi; te SentiNet\_1 koji je izvorno predstavljen u radu Campos et. al [78], te je već korišten za domensku inicijalizaciju prilikom finog podešavanja za mreže klasifikaciju slikarskih djela i opisan u poglavlju 3.3.2.

Za predikciju pamtljivosti slike također su korištena tri različita modela. Prvi je MemNet\_1 koji su predstavili Khosla et al. [84], temelji se na AlexNet arhitekturi, a naučen je na LaMem bazi slika. Budući da je već korišten sa svrhom domensko specifične inicijalizacije, detaljan opis toga modela dan je u poglavlju 3.3.2. Zatim su još korišteni izvorni AMNet modeli predstavljeni u [85], pri čemu je jedan naučen na SUN Memorability bazi slika (MemNet\_2), a drugi na LaMem bazi (MemNet\_3).

### 5.2.3 Ekstrakcija naučenih značajki slike više razine

Sa svrhom razumijevanja predikcijskih vrijednosti pojedinih percepcijskih obilježja, analizira se kako dobivene predikcijskih vrijednosti estetske procjene, sentimenta i pamtljivosti koreliraju s drugim značajkama slike. Pri tome se promatra odnos sa značajkama tzv. više razine (engl. high-level image features), odnosno značajkama koje predstavljaju mjeru zastupljenosti određenih stilskih obilježja i složenih vizualnih atributa. Po uzoru na tradicionalne principe oblikovanja slikarskih i fotografskih djela, značajke više razine predstavljaju mjeru zastupljenosti određenih svojstava prostorne i kromatske kompozicije te sadržaja slike. Postoje različite metode oblikovanja takvih značajki slike više [67, 74], no pokazano je da se primjenom konvolucijskih neuronskih mreža takve značajke mogu vrlo uspješno naučiti iz označenog skupa

podataka [70, 71]. Budući da se sposobnost konvolucijskih neuronskih mreža da raspoznaju složene uzorke u slici pokazala korisnom za učenje kompleksnih vizualnih atributa, za ekstrakciju značajki slike više razine korišten je atributno-adaptivni model predstavljen u radu Konga et al. [71]. Taj model je učen na AADB bazi slika, u kojoj su slike označene procjenama estetske kvalitete slike i procjenama prisutnosti određenih vizualnih atributa u slici. Na temelju agregacije procjena većeg broja ispitanika svakoj slici dodijeljena je vrijednost mjere prisutnosti određenog atributa. Arhitektura modela prilagođena je učenju atributa na način da je za svaki atribut dodano nekoliko potpuno povezanih slojeva. Konkretno, izlaz iz prethodnjeg potpuno povezanog sloja AlexNet mreže (fc7) grana se na novo dodane slojeve (fc8 i fc9) koji su specifično određeni za predikciju određenog atributa. Inicijalna ideja konfiguracije ovoga modela je da se u postupku treniranje mreže izlazi atributnih slojeva mogu aktivirati kao dodatna informacija sa svrhom predikcije estetske vrijednosti slike. Izlazi atributnih slojeva tako naučenog modela mogu se koristiti kao ekstraktori značajki više razine, odnosno kao mjere zastupljenosti određenih atributa u slici. Izlaz atributnog sloja je vrijednost u intervalu između 0 i 1, pri čemu više vrijednosti ukazuju na veću vjerojatnost prisutnosti određenog atributa u slici. Za usporedbu s predikcijskim vrijednostima percepcijskih obilježja odabrani su sljedeći atributi:

- kvaliteta sadržaja (engl. content) –određuje je li sadržaj slike pozitivno ocijenjen;
- naglašenost objekta (engl. object emphasis) –određuje sadrži li slika istaknuti objekt, tj. predmet u prvom planu;
- osvjetljenje (engl. lighting) –određuje je li osvjetljenje u slici pozitivno ocijenjeno;
- pravilo trećine (engl. rule of thirds) –određuje prati li kompozicija slike pravilo trećina;
- ponavljanje (engl. repetition) –određuje sadrži li slika ponavljajuće uzorke;
- simetrija (engl. symmetry) –određuje je li slika simetrična;
- harmonija boja (engl. color harmony) –određuje je su li kombinacije boja harmonične;
- izražajnost boja (engl. color vividness) –određuje je su li boje u slici izražajne.

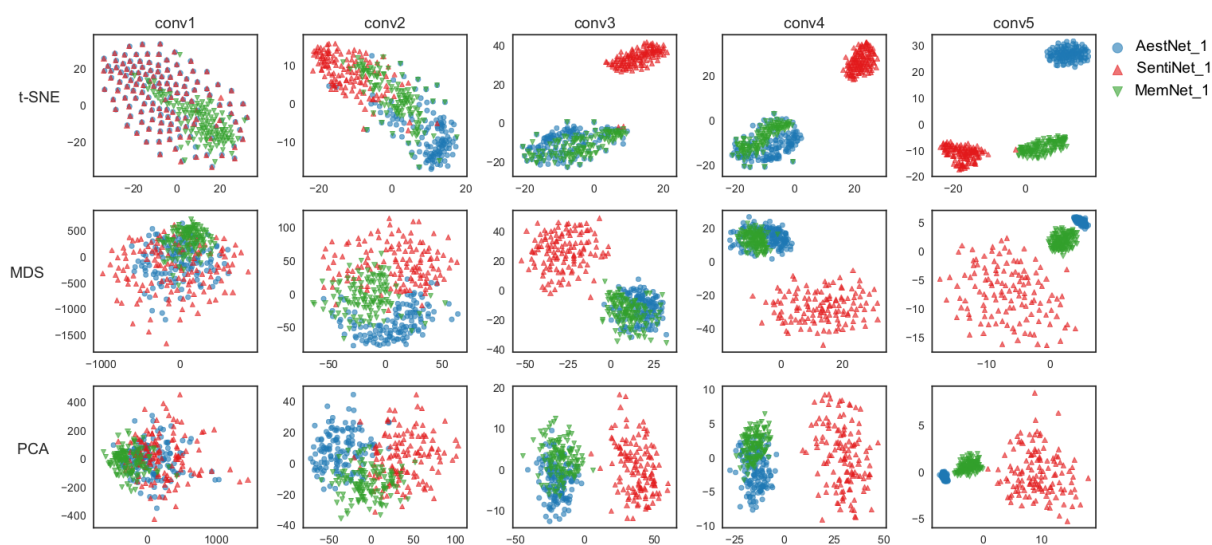
Kako bi se kvalitativnim rezultatima potkrijepio izbor navedenog modela za ekstrakciju značajki slike, u prilogu (Slika B.1-B.8) nalaze se prikazi 100 slika iz WikiArt baze s najvećim i 100 slika s najmanjim predikcijskim vrijednostima za svaki atribut.

### 5.3 Analiza naučenih reprezentacija slike i aktivacijskih mapa

Zbog razumijevanja svojstva naučenih reprezentacija slike analiziraju se aktivacije konvolucijskih slojeva modela neuronskih mreža koje imaju istu arhitekturu, ali su naučene za predikciju triju različitih percepcijskih obilježja. Izlaz svakog konvolucijskog sloja čini skup aktivacijskih mapa koje predstavljaju različito filtrirane verzije ulazne slike. S ciljem da se demonstrira kako slojevi tri različita modela koji imaju istu AlexNet arhitekturu i naučeni su za raspoznavanje različitih percepcijskih obilježja, postupno uče reprezentacije specifične za svako obilježje,

analizira se udaljenost među usrednjenim mapama značajki dobivenima kao odziv konvolucijskih slojeva naučenih modela na 100 nasumično odabranih slika iz WikiArt baze.

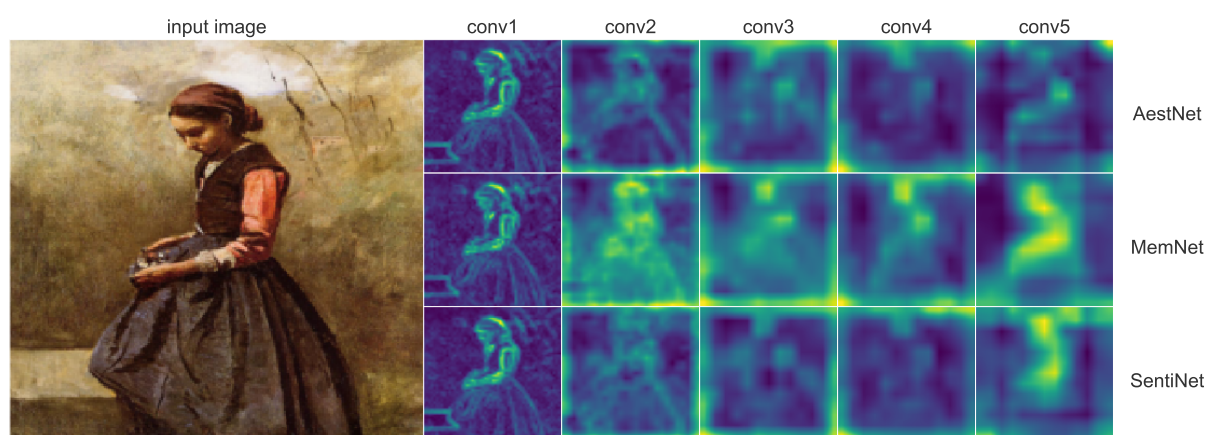
Nad usrednjenim mapama značajki provedena je redukcija dimenzionalnosti kako bi se svaka usrednjena mapa prikazala kao točka u 2D prostoru. Zbog provjere konzistentnosti rezultata, redukcija dimenzionalnosti provedena je korištenjem triju različitih metoda: t-SNE metoda (engl. t-Distributed Stochastic Neighbor Embedding, t-SNE), višedimenzionalno skaliranje (engl. multidimensional scaling, MDS) i analiza glavnih komponenti (engl. principal component analysis, PCA). Na slici 5.2 prikazana je distribucija redukcijama dimenzionalnosti dobivenih 2D uzoraka usrednjenih mapa značajki, dobivenih kao odziv različitih konvolucijskih slojeva triju različito naučenih modela na istih 100 slika.



**Slika 5.2:** Distribucije 2D uzoraka usrednjenih mapa značajki dobivenih primjenom t-SNE, MDS i PCA metoda redukcije dimenzionalnosti. Usrednjene mape značajki predstavljaju odzive različitih konvolucijskih slojeva AestNet\_1, SentiNet\_1 i MemNet\_1 mreža na 100 nasumično odabranih slika iz WikiArt baze.

Konfiguracije 2D točaka usrednjenih mapa značajki dobivenih primjenom različitih metoda redukcije dimenzionalnosti pokazuju konzistentnost s obzirom na stupanj formiranja grupa točaka dobivenih iz različitih konvolucijskih slojeva. Vidljivo je da se odzivi prvog konvolucijskog sloja ne razlikuju s obzirom na različito naučene modele, dok se distinktivne grupe točaka počinju stvarati među odzivima drugog konvolucijskog sloja. Odzivi trećeg i četvrtog sloja pokazuju preklapanje reprezentacija dobivenih od modela naučenih za raspoznavanje sentimenta i pamtljivosti, dok se konačno razdvajanje dobiva u odzivima petog konvolucijskog sloja. Vizualizacija postupne konvergencije usrednjenih 2D reprezentacija mapa značajki dosljedna je s dosadašnjim saznanjima o dinamici učenja konvolucijskih neuronskih mreža. Naime, poznato je da parametri nižih konvolucijskih slojeva konvergiraju tako da nauče raspoznavati slične reprezentacije slike, najčešće generičke oblike, dok su reprezentacije naučene u dubljim slojevima specifične za određeni skup podataka i zadatak za koji se mreža uči. U slučajevima

kada se mreža uči za zadatak klasifikacije slike prema objektima u slici, naglašena područja aktivacijskih mapa dubljih slojeva mreže odgovarat će uzorcima i oblicima koji su specifični za naučene objekte. Međutim, ako je model konvolucijske neuronske mreže naučen za apstraktnije i složenije zadatke raspoznavanja slika, kao što je raspoznavanje percepcijskih obilježja, interpretabilnost odziva dubljih slojeva predstavlja veći izazov. Na slici 5.3 prikazane su usrednjene aktivacijske mape dobivene kao odziv različitih konvolucijskih slojeva triju različito naučenih modela na istu ulaznu sliku. S obzirom na broj i konfiguraciju neurona u konvolucijskim slojevima, kao i na operacije sažimanja, dimenzionalnost aktivacijskih mapa varira za različite slojeve. Stoga su bilinearnom interpolacijom povećane rezolucije izlaznih aktivacijskih mapa kako bi dimenzionalnost svih mapa odgovarala dimenzijama ulazne slike (227 x 227 piksela).



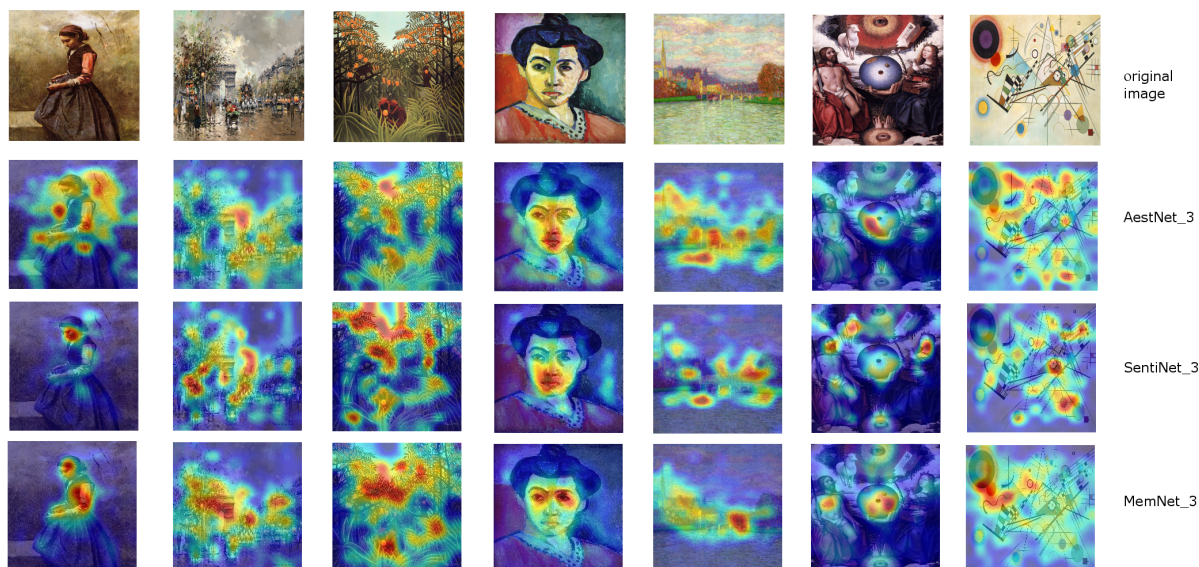
**Slika 5.3:** Usrednjene aktivacijske mape dobivene kao odziv pet konvolucijskih slojeva AestNet\_1, SentiNet\_1 i MemNet\_1 modela na istu ulaznu sliku (J.B.C. Corot, *Zamišljena djevojka*).

Aktivacijske mape prikazane na slici 5.3 konzistentne su s opažanjima koja proizlaze iz vizualizacije postupne konvergencije usrednjenih 2D reprezentacija mapa značajki, danoj u slici 5.3. Odnosno, da se aktivacijske mape prvog konvolucijskog sloja ne razlikuju značajno, te da se aktivacijske mape petog konvolucijskog sloja slične za SentiNet\_1 i MemNet\_1 model. Za odabrani primjer može se uočiti da su u aktivacijskim mapama zadnjeg konvolucijskog sloja modela naučenog za raspoznavanje slika prema sentimentu i pamtljivosti, naglašena područja ona koja odgovaraju područjima lica i tijela.

Sa svrhom detaljnije analize koja područja slike igraju odlučujuću ulogu u predikciji estetske vrijednosti, sentimenta i pamtljivosti slike, uspoređuju se aktivacijske mape dobivene iz modela s ugrađenim mehanizmom pažnje (engl. attention mechanism): AestNet\_3, SentiNet\_3 i MemNet\_3. Arhitektura tih modela temelji se na predtreniranoj ResNet50 mreži, nakon koje slijedi tzv. „meki“ mehanizam vizualne pažnje (engl. soft attention mechanism) ostvaren povratnom neuronskom mrežom s dugotrajnim kratkoročnim pamćenjem (engl. Long short-term memory network, LSTM) kroz sekvencu od tri koraka. U svakom LSTM koraku proizvodi se aktivacijska mapa na temelju stanja prethodnog koraka i vektora slike. Naglašene regije akti-



vacijske mape odgovaraju području pažnje, odnosno ukazuju na one elemente u slici koji su značajni za određivanje predikcijske vrijednosti određenog obilježja. Na slici 5.4 prikazane su primjeri aktivacijskih mapa dobivenih u posljednjem LSTM koraku za različite ulazne slike.



**Slika 5.4:** Primjeri aktivacijskih mapa dobivenih korištenjem AestNet\_3, SentiNet\_3 i MemNet\_3 modela.

Iz primjera u slici vidljivo je da su naglašena područja aktivacijskih mape modela naučenog za predikciju estetske vrijednosti slike češće raspršena, dok se za modele naučene za sentiment i pamtljivost izraženije lokaliziraju, pri čemu su osobito naglašena područja gdje se pojavljuju lica. Time se pokazuje konzistentnost između aktivacijskih mapa dobivenih iz konvolucijskih odziva mreža s AlexNet arhitekturom i aktivacijskih mapa dobivenih iz mreža s AMNet arhitekturom. Za bolje razumijevanje aktivacijskih područja, u Prilogu B dana je slika B.9 s usrednjenim aktivacijskim mapama za 300 slika s najvećim vrijednostima i 300 slika s najmanjim vrijednostima predikcija, koje pripadaju trima različitim žanrovima, za sva tri modela i sva tri LSTM koraka.

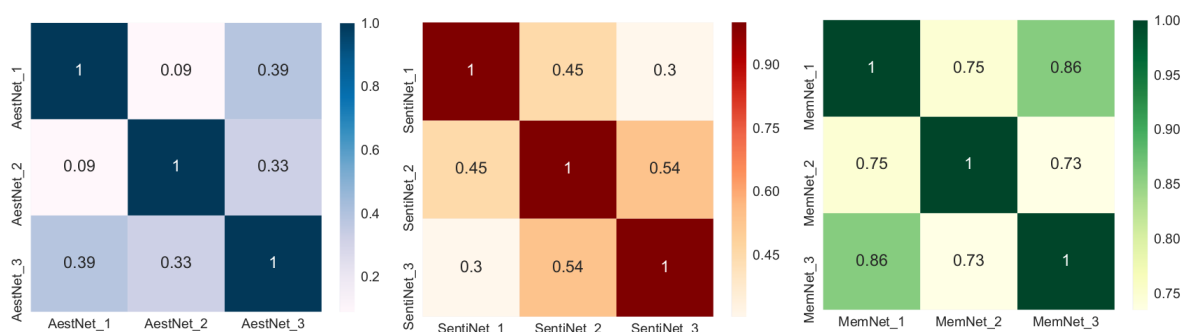
## 5.4 Evaluacija predikcija percepcijskih obilježja slike

Budući da su za predikciju svakog percepcijskog obilježja korištena tri različita modela, u ovome se poglavlju analizira konzistentnost rezultata dobivenih primjenom različitih modela. Također, analizira se korelacija predikcijskih i stvarnih vrijednosti na označenim skupovima slika te se odabiru modeli za daljnju analizu predikcijskih vrijednosti na WikiArt bazi slikarskih djela.

### 5.4.1 Analiza konzistentnosti predikcija dobivenih različitim modelima

Za predikciju svakog percepcijskog obilježja korištena su tri različita modela kako bi utvrdilo koji model daje najbolje rezultate te koliko su konzistentni rezultati dobiveni primjenom različito naučenih modela. Slike iz WikiArt baze koriste se kao ulaz u modele navedene u Tablici 5.2 te se za svaku sliku dobivaju tri predikcijske vrijednosti za estetsko vrednovanje slike, tri za sentiment i tri za pamtljivost slike. Predikcijska vrijednost, odnosno izlaz iz zadnjeg sloja mreže, broj je u intervalu između 0 i 1. Za predikciju estetske vrijednosti slike viša vrijednost dobivenog izlaza znači da model predviđa visoku estetsku kvalitetu slike. Slično je i za predikciju pamtljivosti, gdje viša vrijednost ukazuje veću pamtljivost slike. U slučaju predikcije sentimenta, viša vrijednost izlaza ukazuje da slika pobuđuje pozitivne osjećaje, a niža vrijednost da pobuđuje negativne osjećaje.

U svrhu analize konzistencije između predikcijskih vrijednosti dobivenih za isto obilježje korištenjem različitih modela, koristi se Spearmanov koeficijent korelacije. Na slici 5.5 prikazane su vrijednosti korelacijskih koeficijenata između vrijednosti pojedinih predikcijskih obilježja dobivenih za sve slike iz WikiArt kolekcije.



**Slika 5.5:** Toplinske mape (engl. heatmaps) vrijednosti Spearmanovih koeficijenata korelacije između različitih predikcija estetske vrijednosti (lijevo), sentimenta (sredina) i pamtljivosti slike (desno) ( $p$ -vrijednosti  $< 0.001$ ).

Korelacija među predikcijskim vrijednostima različitih modela najveća je za zadatak predikcije pamtljivosti slike, a najmanja za predikciju estetske vrijednosti slike. Visok koeficijent korelacije između različito dobivenih predikcija pamtljivosti slike ukazuje na konzistenciju odziva različitih modela i na robusnost rezultata predikcije pamtljivosti. S druge strane, relativno slab koeficijent korelacije među izlazima različitih modela naučenih za predikciju estetske vrijednosti slike ukazuje na složenost toga zadatka. Jedno od mogućih objašnjenja za takav ishod je pretpostavka da je estetsko vrednovanje slike subjektivnije od pamtljivosti slike te je time svako, pa i računalno, predviđanje toga obilježja složenije. Također, moguće je da slab koeficijent korelacije među predikcijskim vrijednostima estetskog vrednovanja slike proizlazi iz različitosti eksperimentalnih postavki pripreme kolekcije označenih slika (broj sudionika, tip slika, varijacije anketnih upitnika, itd.).

## 5.4.2 Usporedba predikcijskih i stvarnih vrijednosti

S obzirom na slabiju korelaciju među predikcijskim vrijednostima različitih modela naučenih za procjenu estetske vrijednosti i sentimenta slike i s obzirom na dostupnost manjih kolekcija slikarskih djela koje sadrže oznake anketnim ispitivanjem dobivenih subjektivnih vrednovanja određenih svojstava slike, sa svrhom odabira najboljeg modela za daljnju analizu, ispitan je odnos predikcijskih vrijednosti i stvarnih vrijednosti korištenjem tih postojećih baza. Potrebno je naglasiti da se izraz „stvarna vrijednost“ ovdje koristi zbog naglašavanja razlike između računalnih predikcija i oznaka dobivenih ljudskom procjenom. U kontekstu analize slike na razini percepcijskih obilježja koja su izrazito subjektivna i nisu kvantitativne prirode, ne mogu postojati “stvarne” vrijednosti kao takve. No, u ovom kontekstu, tim se terminom označavaju vrijednosti koje u načelu odgovaraju ljudskom konsenzusu, odnosno srednje ili medijan vrijednosti dobivene na temelju procjena većeg broja ispitanika.

Za odabir modela koji će se koristiti za analizu odnosa predikcija estetskih vrijednosti i drugih značajki na WikiArt bazi slika, ispitan je odnos predikcijskih vrijednosti i stvarnih vrijednosti na manjoj JenAesthetic bazi slika. U JenAesthetic kolekciji slika svaka slika sadrži oznake anketnim ispitivanjem dobivenih subjektivnih vrednovanja određenih svojstava slike. U tablici 5.3 navedeni su Spearmanovi korelacijski koeficijenti između predikcijskih vrijednosti dobiveni trima različitim modelima i usrednjenih vrijednosti subjektivnih vrednovanja određenih svojstava slike.

**Tablica 5.3:** Vrijednosti Spearmanovih korelacijskih koeficijenata između predikcijskih vrijednosti dobiveni korištenjem AestNet\_1, AestNet\_2 i AestNet\_3 i usrednjenih vrijednosti subjektivnih vrednovanja određenih svojstava slike na JenAesthetic bazi slika. (\* p-vrijednost < 0.01, \*\* p-vrijednost > 0.1).

	Estetska vrijednost	Boja	Kompozicija	Sadržaj
AestNet_1	0.040**	0.203*	0.029**	-0.003**
AestNet_2	0.166*	0.084*	0.131*	-0.076*
AestNet_3	0.267*	0.372*	0.287*	0.267*

Rezultati pokazuju da predikcijske vrijednosti dobivene AestNet\_3 modelom imaju najviši korelacijski koeficijent s ispitivanjem dobivenim procjenama estetske kvalitete slike, ali i drugih obilježja kao što su procjene vrednovanja boje, kompozicije i sadržaja slike. Na temelju tih rezultata, odzivi AestNet\_3 modela na slike iz WikiArt baze koriste se za daljnju analiza odnosa značajki i kategorija slikarskih djela s predikcijskim vrijednostima estetske kvalitete slike.

Sa svrhom evaluacije modela naučenih za predviđanje vizualnog sentimenta, analiziraju se odzivi modela na slike iz dvije različite baza. Prva je MART baza slika koja sadrži apstrak-

tne slike s binarno određenim oznakama sentimenta, tj. oznakama pobuđuje li slika pozitivan ili negativan osjećaj. Izračun Spearmanovih koeficijenata korelacije pokazuje da su predikcijske vrijednosti sva tri modela značajno pozitivno korelirane sa stvarnim vrijednostima oznaka sentimenta u MART bazi. Konkretno, koeficijent korelacije s predikcijskim vrijednostima SentiNet\_1 modela iznosi  $\rho = 0.531$ , s vrijednostima SentiNet\_2 modela  $\rho = 0.546$  i SentiNet\_3 modela  $\rho = 0.461$  (u sva tri slučaja je p-vrijednost  $< 0.01$ ).

**Tablica 5.4:** Vrijednosti Spearmanovih korelacijskih koeficijenata između predikcijskih vrijednosti dobiveni korištenjem SentiNet\_1, SentiNet\_2 i SentiNet\_3 i vrijednosti mjera prisutnosti određene emocije dobivene na temelju oznaka ispitanika u WikiArt Emotions bazi. Korelacija je statistički značajna za sve slučajeve (p-vrijednost  $< 0.01$ ), osim za \* (p  $< 0.1$ ).

	tuga	gađenje	strah	pesimizam	radost	ljubav	optimizam	povjerenje
SentiNet_1	-0.201	-0.024	-0.253	-0.195	0.056	0.028*	0.106	-0.054
SentiNet_2	-0.205	-0.086	-0.252	-0.202	0.230	0.061	0.223	0.004
SentiNet_3	-0.142	-0.217	-0.205	-0.178	0.396	0.246	0.268	0.289

Osim na MART bazi slika, evaluacija predikcijskih vrijednosti provedena je na WikArt Emotions bazi slika u kojoj su slike označene različitim kategorijama emocija. Budući da je više ispitanika ocjenjivalo slike s obzirom na emociju koju asociiraju sa slikom, mjerilo prisutnosti određene emocije u slici definirano je kao udio dodijeljenih oznaka određene emocije i ukupnog broja ispitanika koji su ocjenjivali sliku. U tablici 5.4 prikazani su Spearmanovi koeficijenti korelacije između predikcijskih vrijednosti sentimenta i mjere prisutnosti određene emocije dobivene na temelju oznaka ispitanika. Iz prikazanih rezultata vidljivo je da predikcijske vrijednosti dobivene SentiNet\_3 modelom imaju najveći pozitivni koeficijent korelacije s vrijednostima pozitivnih kategorija emocija. Također, rezultati korelacije na WikiArt Emotions bazi slika koja sadrži slike koje pripadaju različitim žanrovima i stilovima ukazuju da su predikcijske vrijednosti dobivene SentiNet\_3 modelom najkonzistentnije s obzirom na predznak zastupljenih kategorija emocija te stoga je taj model odabran za daljnju analizu.

S obzirom da trenutno ne postoji baza umjetničkih slika s oznakama pamtljivosti slike koje su dobivene eksperimentalnom ispitivanjem većeg broja sudionika, nije bilo moguće evaluirati predikcijske vrijednosti pamtljivosti slikarskih djela sa stvarnim vrijednostima. Međutim, budući da predikcijske vrijednosti dobivene korištenjem triju različitih modela pokazuju visok stupanj korelacije, odabir modela u ovome slučaju ne utječe značajno na ishod daljnje analize. Stoga je za analizu odnosa predikcijskih vrijednosti pamtljivosti odabran MemNet\_3 model, s obzirom da se njime postiže trenutno najbolja konzistentnost sa stvarnim vrijednostima pamtljivosti na bazama prirodnih slika.

## 5.5 Odnos značajki slike i percepcijskih obilježja slike

U ovom se poglavlju razmatraju kvalitativni rezultati predikcija te odnos predikcijskih vrijednosti percepcijskih obilježja slike i drugih značajki slike.

### 5.5.1 Kvalitativni rezultati predikcije percepcijskih obilježja

Kvalitativni rezultati predikcija percepcijskih obilježja temelje se na prikazu slika s najvećim i najmanjim vrijednostima predikcija pojedinih percepcijskih obilježja. Na slici 5.6 prikazano je 100 slika s najvećim i 100 slika s najmanjim vrijednostima predikcije estetske kvalitete slike dobivenih AestNet\_3 modelom na skupu slika iz WikiArt baze. Ovakav prikaz omogućuje da se brzo utvrde osnovne vizualne razlike i provjeri logičnost i smislenost dobivenih rezultata. Razlika između slika s visokim i niskim estetskim predikcijama očituje se ponajprije u izboru boja. Vidljivo je da među slikama s najvišim estetskim predikcijama dominiraju intenzivne i žarke boje, dok kod slika s niskim predikcijskim vrijednostima uglavnom prevladavaju tamne i monotone boje. Ovakav ishod kvalitativnih rezultata procjene estetske vrijednosti slika u skladu je s intuitivnim pretpostavkama, iako vrlo pojednostavljen, posebice u kontekstu estetskog vrednovanja umjetničkih djela.

U cilju razumijevanja obilježja slika koje utječu na predikciju vizualnog sentimenta, na slici 5.7 dan je prikaz 100 slika s najvišim i 100 slika s najnižim predikcijskim vrijednostima sentimenta dobivenih korištenjem SentiNet\_3 modela. Očita vizualna razlika ponovno je izbor boja, ali su također vidljivi uzorci motiva koji se ponavljaju između pozitivno i negativno ocijenjenih slika. Primjerice, u skupu pozitivnih slika često se mogu uočiti slike cvijeća, te nasmijanih portreta i obiteljskih portreta. U skupu negativnih slika učestalo se pojavljuju apstraktne slike u kojima dominiraju sivi tonovi i oštri rubovi, no također slike portreta koje prikazuju uznemirena ili tužna lica. Ovakvi kvalitativni rezultati ukazuju na to da je model naučio razlikovati suptilne nijanse u sadržaju, ali i da ne pokazuje pristranost prema samo jednom sadržajnom motivu, već raspoznaje vizualni sentiment neovisno o žanru i stilu.

Na slici 5.8 prikazano je 100 slika s najvišim i 100 slika s najnižim predikcijskim vrijednostima pamtljivosti slike dobivenih korištenjem MemNet\_3 modela. Razlikovanje slika prema pamtljivosti ponajprije se pokazuje na temelju sadržaja slike. Među slikama s najvećim predikcijskim vrijednostima prevladavaju aktovi i portreti te apstraktne slike s istaknutim i ponavljajućim uzorcima. S druge strane, slike s najnižim predikcijskim vrijednostima pamtljivosti su isključivo pejzaži. Ovakav odnos pamtljivosti slike i motiva potvrđen je analizom pamtljivosti slika prema žanrovima, predstavljenom u poglavlju 5.6.2.

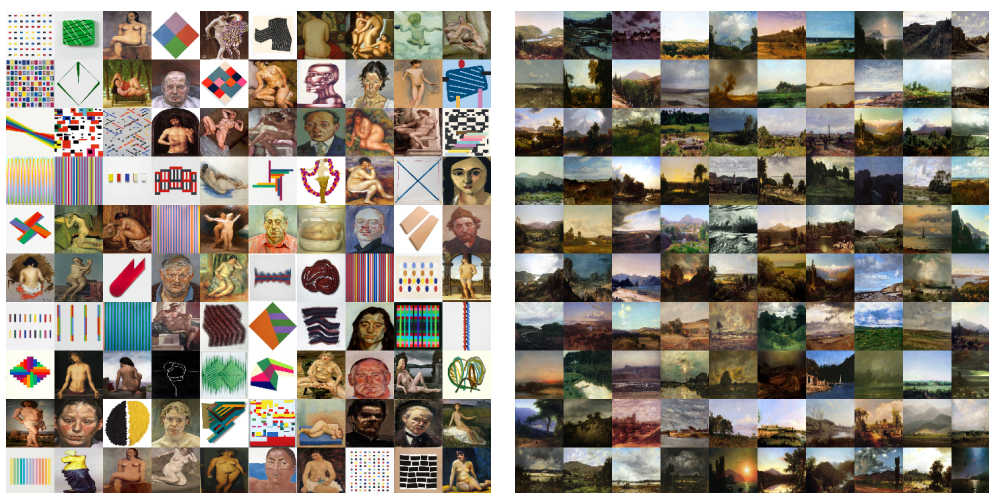




**Slika 5.6:** 100 slika iz WikiArt baze s najvišim (lijevo) i najnižim (desno) predikcijskim vrijednostima estetske procjene.



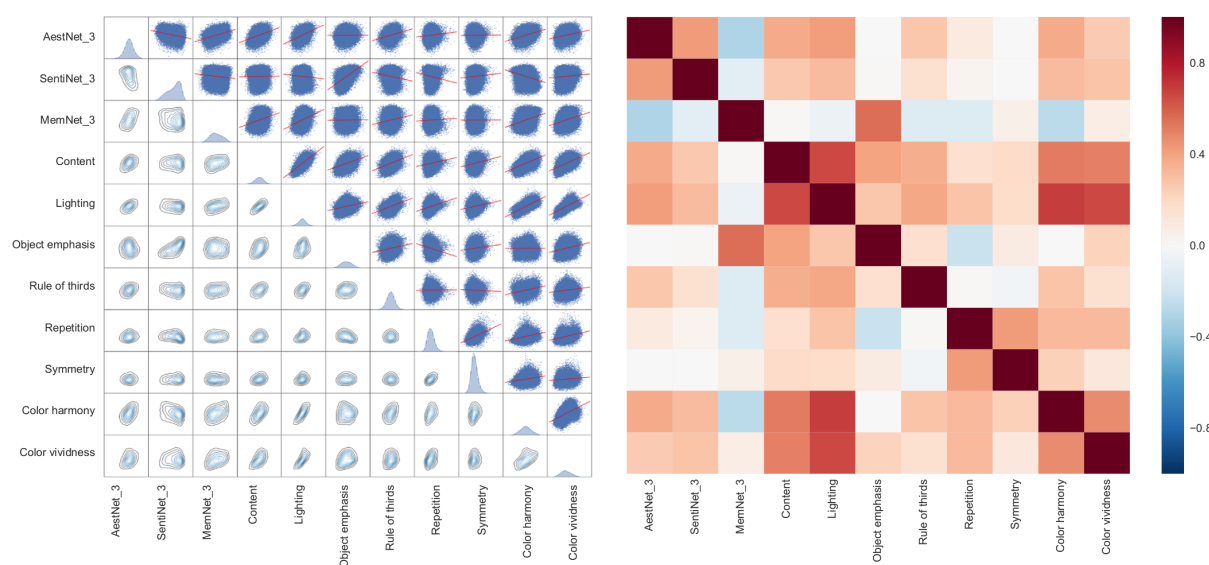
**Slika 5.7:** 100 slika iz WikiArt baze s najvišim (lijevo) i najnižim (desno) predikcijskim vrijednostima vizualnog sentimenta.



**Slika 5.8:** 100 slika iz WikiArt baze s najvišim (lijevo) i najnižim (desno) predikcijskim vrijednostima pamtljivosti slike.

## 5.5.2 Odnos značajki slike više razine i percepcijskih obilježja

Za bolje razumijevanje koje vizualne karakteristike slike doprinose predviđanju pojedinih percepcijskih obilježja slika, analizira se korelacija između predikcijskih vrijednosti percepcijskih obilježja i tzv. značajki slike više razine dobivenih metodom opisanom u poglavlju 5.2.3. Na desnoj strani slike 5.9 prikazana je toplinska mapa Spearmanovih koeficijenata korelacije između pojedinih predikcijskih obilježja i značajki slike ( $p$ -vrijednost  $< 0.01$ ). Na lijevoj strani slike 5.9 dan je detaljan prikaz međuodnosa značajki. U gornjem desnom trokutu prikazan je linearni odnos dviju varijabli, a u donjem lijevom trokutu jezgrena estimacija funkcije gustoće vjerojatnosti.



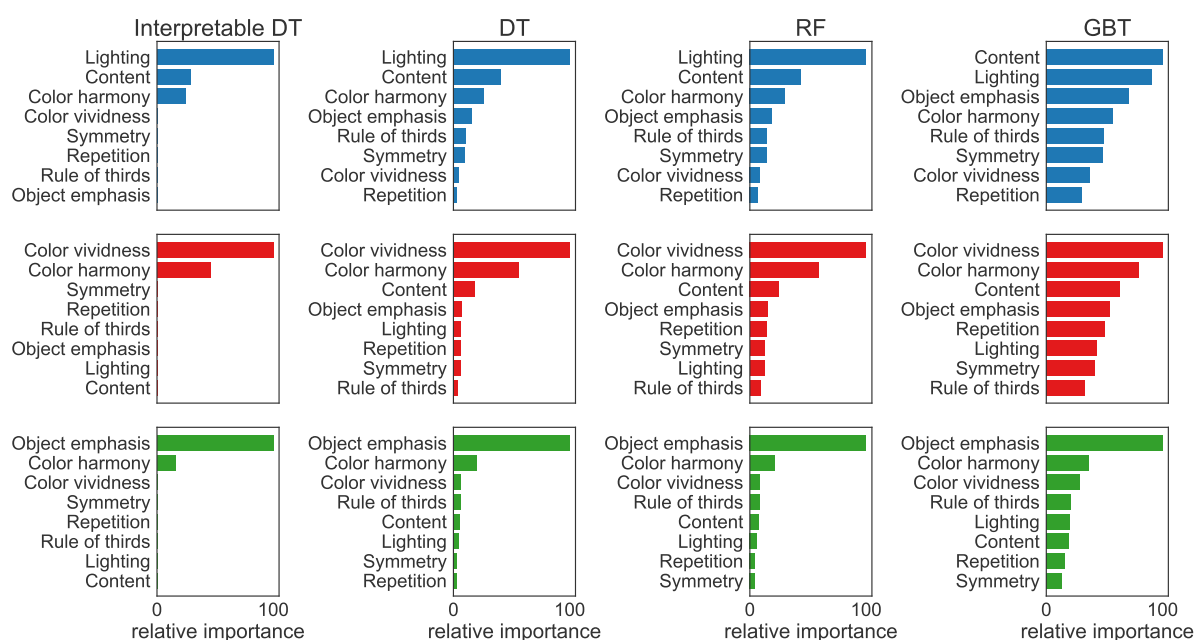
**Slika 5.9:** Međuosnos predikcija pojedinih percepcijskih obilježja i značajki slike.

Na temelju dobivenih rezultata vidljivo je da su predikcije estetske vrijednosti i sentimenta pozitivno korelirane ( $\rho = 0.424$ ,  $p$ -vrijednost  $< 0.01$ ), dok su obje negativno korelirane s predikcijskim vrijednostima pamtljivosti slike. Predikcijske vrijednosti estetske procjene slike pozitivno koreliraju sa značajkom kvalitete sadržaja, što ukazuje na to da ocjena sadržaja igra veliku ulogu u procjeni estetske vrijednosti slike. Također, predikcijske vrijednosti estetske procjene slike pozitivno su korelirane sa značajkama osvjetljenja i harmoničnosti te izražajnosti boja. S druge strane, predikcijske vrijednosti pamtljivosti pokazuju negativnu korelaciju s vrijednostima značajke harmoničnosti boje, no snažno su pozitivno korelirane s vrijednostima značajke naglašenosti objekta, što pokazuje da slike koje prikazuju određeni motiv u prvome planu imaju tendenciju da budu pamtljivije.

S ciljem da se istraži kako pojedine značajke ne samo koreliraju, već zajednički utječu na predikciju dobivenih vrijednosti estetske procjene, sentimenta i pamtljivosti slike, korištene su prediktivne tehnike rudarenja podataka temeljene na stablima odlučivanja (engl. decision trees). Stabla odlučivanja su prediktivni modeli koji na temelju podataka izvode veze s ciljem dobiva-



nja izlaznih vrijednosti. Nekoliko različitih regresijskih modela stabla odlučivanja korišteno je za dobivanje relativnog poretka važnosti značajki. Konkretno, korištena su stabla odlučivanja generirana CART algoritmom (engl. Classification And Regression Trees) [99, 100], te korištenjem metode slučajnih šuma (engl. random forest) [101] i Gradient boosting metode [102]. Regresijski zadatak formuliran je kao predikcija vrijednosti estetskih vrijednosti, sentimenta i pamtljivosti na temelju vrijednosti 8 različitih značajki. Evaluacija regresijskih modela provedena je na temelju srednje apsolutne pogreške (engl. mean squared error - MSE) koja daje prosjek razlike između izvornih vrijednosti i predviđenih vrijednosti. U ovom kontekstu se pod „izvornim“ vrijednostima smatraju predikcijske vrijednosti dobivene korištenjem AestNet\_3, SentiNet\_3 i MemNet\_3, koje se uspoređuju s vrijednostima predikcije tih percepcijskih obilježja dobivenih na temelju vrijednosti 8 različitih značajki. Na slici 5.10 prikazane su dobivene relativne važnosti značajki za svako percepcijsko obilježje korištenjem četiri različita modela.



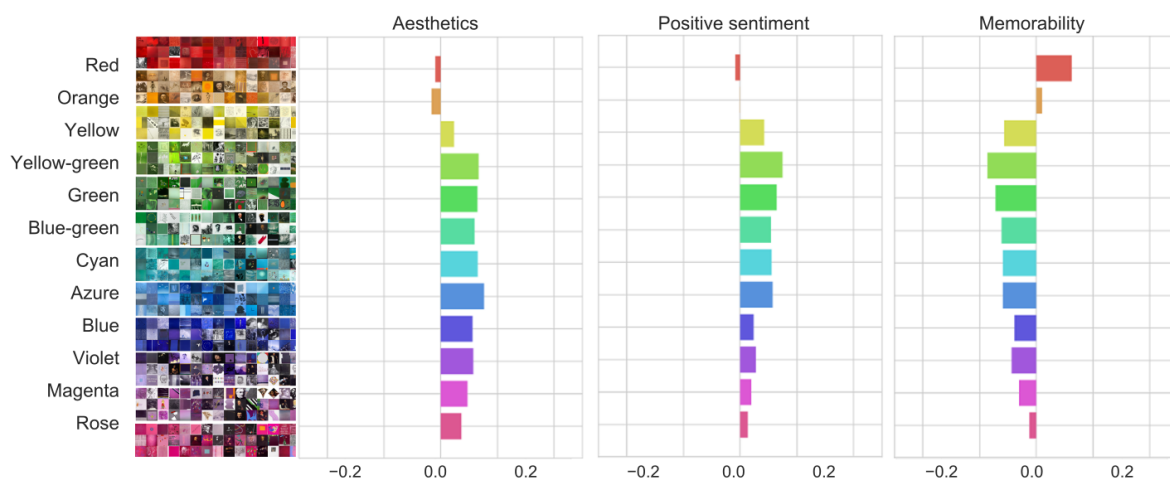
**Slika 5.10:** Relativne važnosti značajki za predikciju estetske vrijednosti slike (gornji red), sentimenta (srednji red) i pamtljivosti slike (donji red) korištenjem različitih regresijskih modela: stablo odlučivanja dobiveno CART algoritmom (DT) i njegova optimizirana verzija interpretabilnog stabla (Interpretable DT), slučajne šume (RT) i gradient boosting trees (GBT).

Poredak relativne važnosti značajki za predikciju pojedinih percepcijskih obilježja pokazuje se postojanim prilikom korištenja različitih modela, što ukazuje na robusnost rezultata. Također, dobiveni rezultati u skladu su s opažanjima proizašlim iz korelacijske analize. Primjerice, pokazuje se da značajka osvjetljenja igra veliku ulogu u predikciji estetske vrijednosti slike, dok su harmoničnost i izražajnost boja ključne za predikciju vizualnost sentimenta, a naglašenost objekta za predikciju pamtljivosti slike.



### 5.5.3 Korelacija značajki boje i percepcijskih obilježja

Budući da se značajke vezane uz boju (harmoničnosti i izražajnost boja) pokazuju značajnima u predikciji svih obilježja, u ovome poglavlju detaljnije se razmatra korelacija vrijednosti percepcijskih obilježja s pojedinim tonovima boja (engl. hue). Za svaku sliku izračunat je histogram boja s 12 razreda. Svaki razred odgovara rasponu od 30 stupnjeva u HSV krugu boja. Kvantizacija HSV prostora boja s obzirom na tonove boja najčešće se prikazuje kružnom paletom u kojoj je ton boje predstavljen kutom od 0 do 360 stupnjeva, pri čemu su primarne boje međusobno odvojene 120 stupnjeva. Podjelom toga prostora na 12 razreda dobiva se 12 temeljnih boja: crvena, narančasta, žuta, žuto-zelena, zelena, plavo-zelena, cijan, azurna, plava, ljubičasta, magenta i roza. Na slici 5.11 prikazan je graf vrijednosti Spearmanovih korelacijskih koeficijenata između predikcija pojedinih percepcijskih obilježja i udjela određene boje. Kako bi se naglasila konzistentnost naziva tona i pripadajuće boje, pokraj imena boja dan je umanjen prikaz slika iz WikiArt baze koje imaju najveću vrijednost udjela pojedine boje.



**Slika 5.11:** Vrijednosti Spearmanovih korelacijskih koeficijenata između predikcijskih vrijednosti estetske procjene, sentimenta i pamtljivosti i udjela različitih tonova boja u slici (p-vrijednost < 0.01).

Iako korelacija pojedinih boja i obilježja nije osobito snažna, ipak se mogu uočiti određene pravilnosti. Primjerice, predikcijske vrijednosti estetske procjene i sentimenta slike pokazuje relativno sličan odnos s različitim bojama koji je gotovo dijametralno suprotan odnosu predikcijskih vrijednosti pamtljivosti i boja. Pokazuje se da su korelacije predikcija estetske procjene i sentimenta s bojama konzistentne s općom pretpostavkom da ljudi preferiraju hladne boje (zelena, plava, cijan) u odnosu na tople boje (crvena, narančasta, žuta), potvrđenom raznim psihološkim ispitivanjima [103, 104].

## 5.6 Predikcije percepcijskih obilježja u kontekstu povijesti umjetnosti

U ovom poglavlju razmatra se odnos predikcijskih vrijednosti pojedinih percepcijskih obilježja prema pojedinim povijesno-umjetničkim kategorijama slikarskih djela u WikiArt bazi. Poglavlje je podijeljeno na nekoliko potpoglavlja s obzirom na sljedeće kategorije: stilski pravac, žanr, autor, povijesno razdoblje.

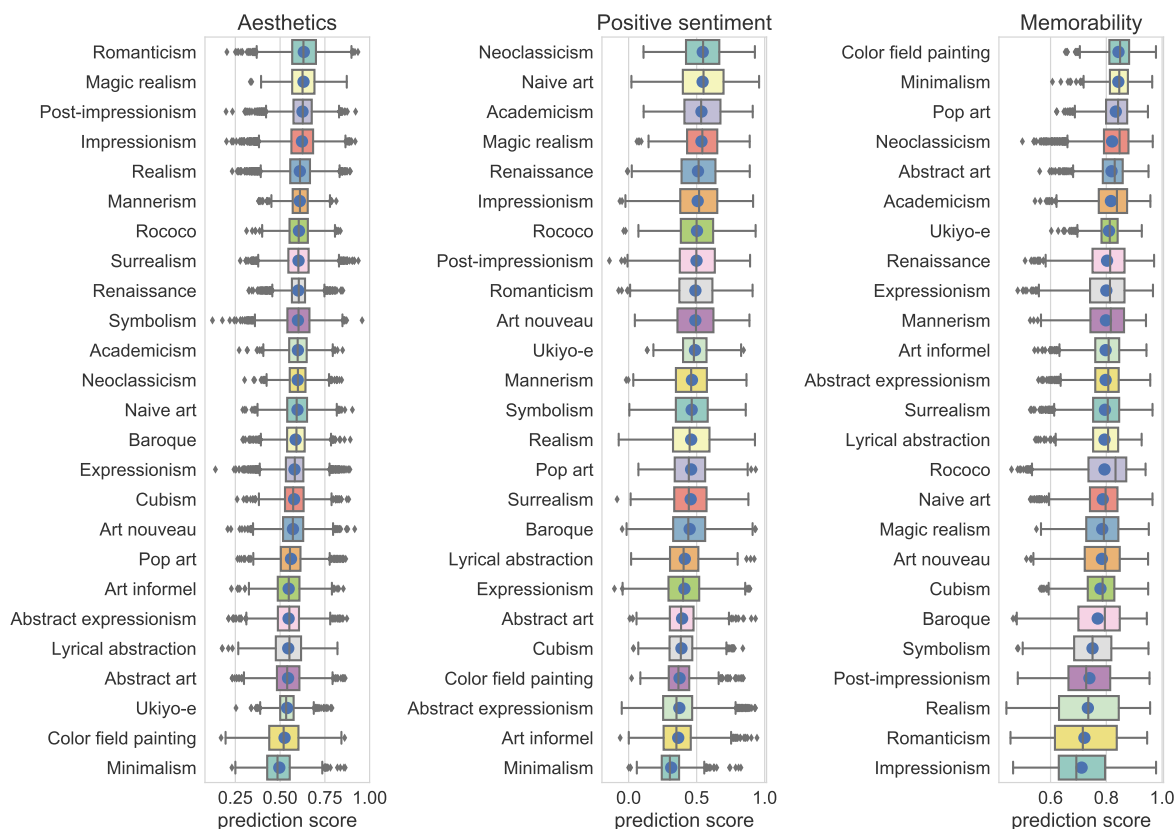
### 5.6.1 Odnos predikcija percepcijskih obilježja i stilskih pravaca

Stilski pravac kao određenje odnosi se na skup karakteristika slikarskog djela koja su specifične za pojedine umjetničke pokrete te je najčešće povezan s određenim povijesnim razdobljem. Kako bi se istražio odnos predikcijskih vrijednosti estetske procjene, sentimenta i pamtljivosti slike s pojedinim stilskim pravcima, odabran je podskup WikiArt baze koji sadrži slike iz 25 različitih stilskih pravaca, pri čemu je svaki pravac zastupljen s najmanje 800 slika. Za svako percepcijsko obilježje izračunava se njegova srednja vrijednost u podskupu slika pojedinog stilskog pravca. Dijagrami pravokutnika (engl. box-plot) na slici 5.12 prikazuju distribuciju predikcijskih vrijednosti percepcijskih obilježja s obzirom na pojedine stilske pravce, pri čemu je plavom točkom označena srednja vrijednost za svaki pravac.

Iz distribucija i srednjih vrijednosti predikcijskih vrijednosti estetske procjene slike, proizlazi da su romantizam i magični realizam stilski pravci s najvišom prosječnom vrijednosti (0.63), dok je minimalizam najniže rangiran stilski pravac (0.49). Budući da raspon srednjih vrijednosti različitih stilova nije osobito velik, teško se mogu donositi sigurni zaključci o odnosu stilova i predikcije estetike, no zanimljivo je da su niže rangirani pretežito apstraktni stilovi. Sličan trend se pokazuje i s obzirom na srednje vrijednosti sentimenta, gdje je minimalizam također najniže rangirani stilski pravac. Srednje vrijednosti pamtljivosti slike pokazuju pak suprotan slijed u poretku stilova. S obzirom na pamtljivost slike, minimalizam i slikarstvo obojenog polja (engl. color field painting) stilovi su s najvećom srednjom vrijednosti predikcija pamtljivosti slike, dok su romantizam i impresionizam najniže rangirani. Moguće objašnjenje takvog poretka je da u romantizmu i impresionizmu prevladavaju pejzaži, dok se sadržaji i kombinacije vizualnih atributa u apstraktnim slikama rijetko pojavljuju u našem svakodnevnom vizualnom iskustvu te stoga predstavljaju iznimku koja privlači pažnju.

### 5.6.2 Odnos predikcija percepcijskih obilježja i žanrova

Pojam žanra odnosi se na razlikovanje slika s obzirom na motive, pa tako razlikujemo žanrovske kategorije kao što su portret, pejzaž, mrtva priroda, itd. Za ispitivanje odnosa predikcijskih vrijednosti percepcijskih obilježja i žanra, koristi se podskup WikiArt baze slika s 13 različiti-



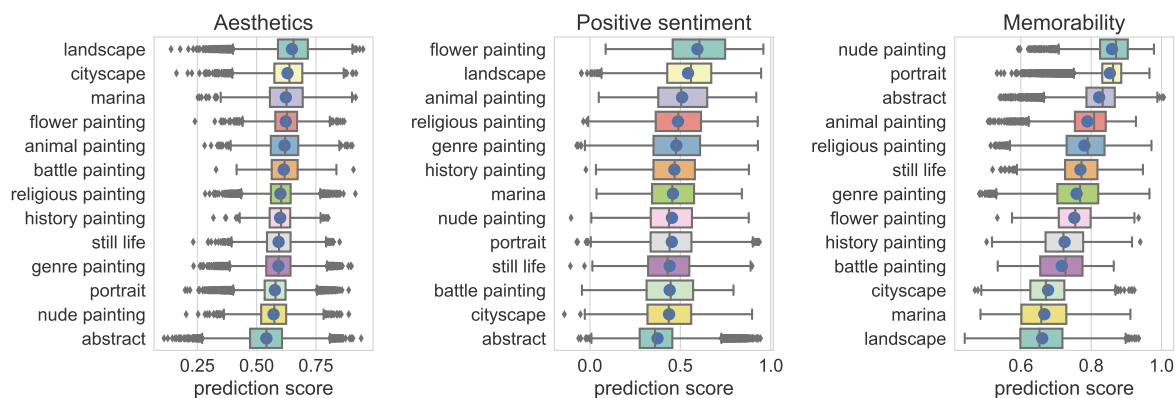
**Slika 5.12:** Box plot distribucija predikcijskih vrijednosti estetske procjene, sentimenta i pamtljivosti slike s obzirom na stilske pravce.

tih žanrovskih kategorija. Slika 5.13 prikazuje distribuciju i srednje vrijednosti predikcijskih vrijednosti percepcijskih obilježja s obzirom na pojedine žanrove.

Razmatranje odnosa predikcijskih vrijednosti estetske procjene i žanra potvrđuje zapažanja dobivena na temelju srednjih vrijednosti stilskih pravaca, budući da kategorija apstraktnih slika ima najnižu srednju vrijednosti. Slike koje sadrže motive otvorenih prostora (pejzaž, marina i gradska panorama) imaju najviše srednje vrijednosti estetske predikcije. Iako je pitanje estetike u umjetnosti mnogo sveobuhvatnije i složenije od kontekstualno neovisnog vrednovanja samog vizualnog prikaza, zanimljivo je da je žanrovski poredak dobivenih predikcijskih vrijednosti estetske procjene slike konzistentan s rezultatima ispitivanja generalnih ljudskih preferencija u umjetnosti. Rezultati opsežne krozkulturalne studije o estetskom vrednovanju različitih umjetničkih djela [104], pokazali su da ljudi općenito preferiraju figurativno slikarstvo u odnosu na apstraktno, te da su slike koje uključuju motive iz prirode najbolje ocijenjene.

Odnos predikcijskih vrijednosti sentimenta i žanra pokazuju se logičnim s obzirom na intuitivne pretpostavke, budući da su pejzaži i slike cvijeća kategorije s najvećom srednjom vrijednosti, dok su apstraktne slike i slike koje prikazuju bitke (engl. battle painting) najniže rangirani žanrovi.

Predikcijske vrijednosti pamtljivosti slike pokazuju da su pejzaži najslabije pamtljiva kate-



**Slika 5.13:** Box plot distribucija predikcijskih vrijednosti estetske procjene, sentimenta i pamtljivosti slike s obzirom na žanrove.

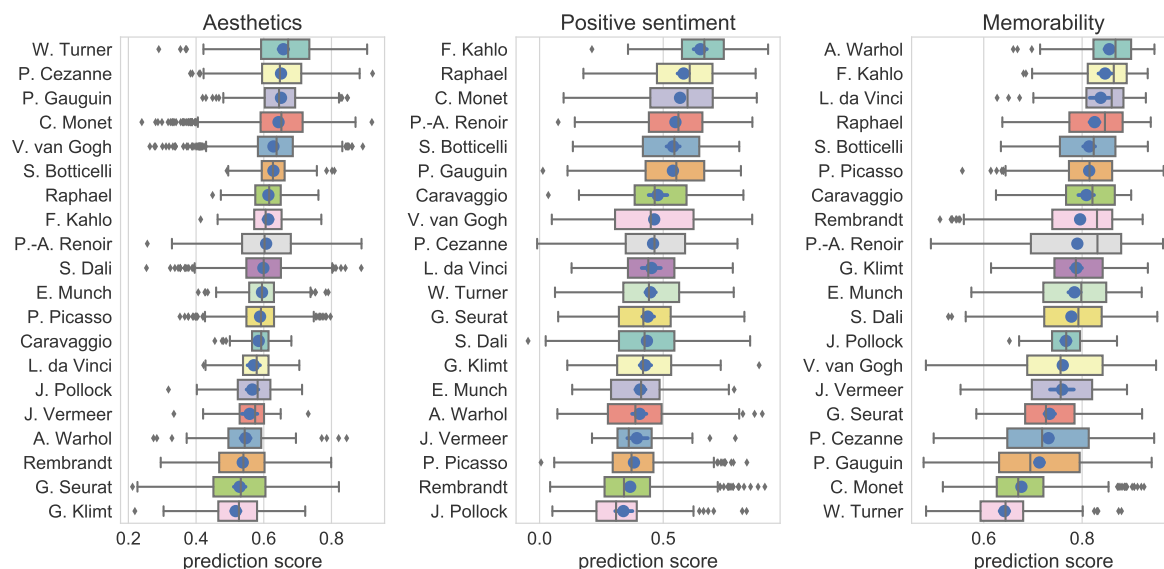
gorija slika, dok su aktovi, portreti i apstraktne slike najpamtljiviji. Ovakav žanrovski poredak s obzirom na pamtljivost slike u skladu je s rezultatima rada Isola et al. [79] koji pokazuje da su slike ljudi pamtljivije od slika prirode te se time pokazuje konzistentnost predikcije pamtljivosti u prirodnim i umjetničkim slikama.

### 5.6.3 Odnos predikcija percepcijskih obilježja i autora

WikiArt baza uključuje djela više od 200 različitih slikara. Sa svrhom ispitivanja odnosa percepcijskih obilježja slika, odabran je podskup od 20 poznatih autora koji pripadaju različitim stilskim razdobljima i zastupljeni su u bazi većim brojem slika. Slika 5.14 prikazuju distribuciju i srednje vrijednosti predikcijskih vrijednosti percepcijskih obilježja s obzirom na djela odabranih autora.

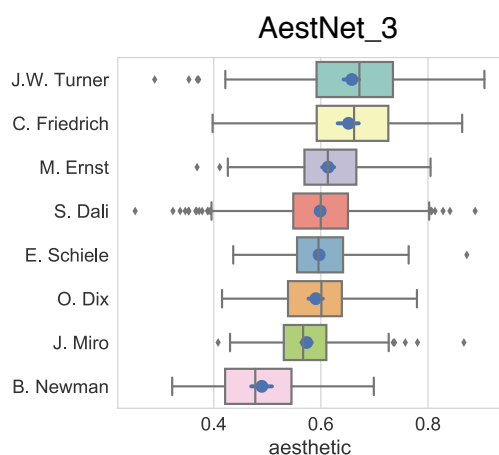
Arbitraran izbor autora onemogućuje donošenje generalnih zaključaka, no relativan odnos među srednjim vrijednostima predikcija percepcijskih obilježja u djelima odabranih autora pokazuje zanimljive rezultate. Primjerice, djela slikara Williama Turnera imaju najveću srednju vrijednost predikcija estetske procjene slike i najnižu srednju vrijednost predikcija pamtljivost. Budući da njegov opus sadrži velik broj morskih pejzaža, to bi objasnilo nisku srednju vrijednost pamtljivost. Zanimljivo je da je Turner također najbolje ocijenjen autor u studiji koja izlaže rezultate anketnog prikupljanja estetskog vrednovanja opusa pojedinih autora [105]. Budući da se u navedenoj studiji iznose podaci o estetskom rangiranju djela šest različitih autora koje je utemeljeno na online prikupljanju procjena većeg broja sudionika, rezultati toga istraživanja koriste se zbog evaluacije poretka dobivenog na temelju predikcija AestNet\_3 modela.

Na slici 5.15 prikazana je box plot distribucija i poredak srednjih vrijednosti djela šest odabranih autora dobiven na temelju predikcija modela (lijevo) te tablica iz [105] u kojoj je naveden poredak dobivenim anketnim ispitivanjem. Sličnost tih dvaju poredaka ukazuje na konzistentnost vrijednosti dobivenih modelom naučenom za predikciju estetske procjene slike i ljudskog



**Slika 5.14:** Box plot distribucija predikcijskih vrijednosti estetske procjene, sentimenta i pamtljivosti slike s obzirom opuse različitih autora.

vrednovanja estetske kvalitete slike.



artist	estimates	aesthetic
Joseph Turner	382	4.55
Caspar Friedrich	377	4.31
Egon Schiele	141	4.28
Salvador Dali	349	3.85
Otto Dix	363	3.85
Max Ernst	162	3.74
Joan Miro	374	3.11
Barnett Newman	359	2.89

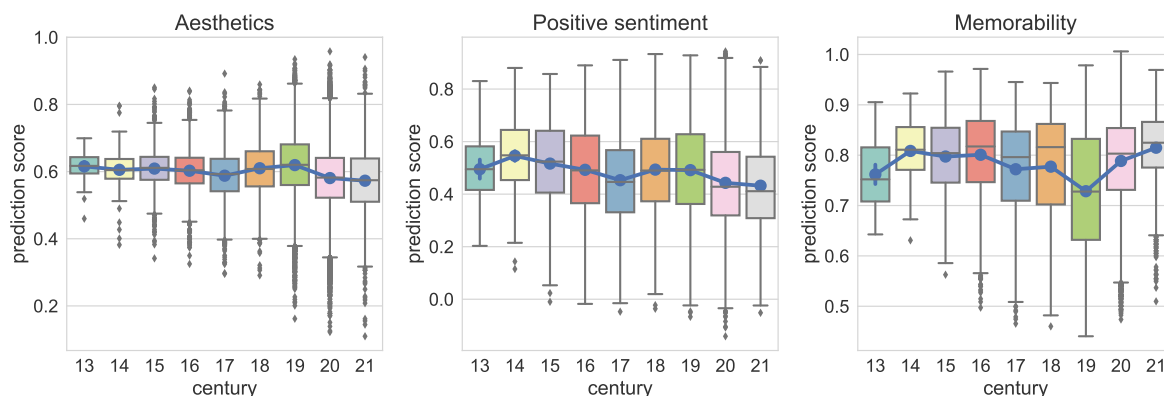
**Source:** Mandel, Igor. "Aesthetic, Art-Historical and Economic Values in Painting: Empirical Study.", 2018.

**Slika 5.15:** Box plot distribucija i poredak srednjih vrijednosti djela šest odabranih autora dobivena na temelju predikcija modela (lijevo), te tablica iz [105] u kojoj je naveden poredak dobivenim anketnim ispitivanjem.

Razmatranje odnosa opusa pojedinih autora s obzirom na predikcijske vrijednosti vizualnog sentimenta pokazuje značajno pozitivno odstupanje za djela Frida Kahlo, čija djela također imaju i visoku srednju vrijednosti pamtljivosti. Iako je razmatranje estetske vrijednosti, sentimenta i pamtljivosti u djelima pojedinih autora pojednostavljeno, ako se izuzme razumijevanje konteksta njihovog stvaranja, navedeni rezultati pokazuju da upravo stilska obilježja karakteristična za opuse pojedinih autora mogu značajno doprinijeti razumijevanju samih percepcijskih obilježja i mogućnosti njihove kvantifikacije i predikcije.

### 5.6.4 Odnos predikcija percepcijskih obilježja i povijesnih razdoblja

Kronološki poredak srednjih vrijednosti percepcijskih obilježja omogućuje lakše uočavanje određenih trendova i uzoraka u podacima. WikiArt kolekcija sadrži oznake godine nastanka za većinu slika u bazi. Za kronološku analizu koristi se podskup od 82000 slika za koje je poznata godina nastanka djela te se slike grupiraju prema stoljeću, počevši od 13. do 21. stoljeća. Na slici 5.16 prikazana je distribucija predikcijskih vrijednosti percepcijskih obilježja po stoljećima.



**Slika 5.16:** Kronološki poredak box plot distribucija predikcijskih vrijednosti estetske procjene, sentimenta i pamtljivosti slike.

Iz kronološkog poretka predikcijskih vrijednosti percepcijskih obilježja uočava se sličan trend u kretanju vrijednosti estetske procjene i sentimenta, s izraženim rastom u 19. stoljeću i padom u 20. i 21. stoljeću. Suprotno tome, predikcijske vrijednosti pamtljivosti imaju upravo najnižu srednju vrijednost za slike iz 19. stoljeća, čime se iznova potvrđuje već uočena suprotnost između vrijednosti estetike i sentimenta s jedne strane i pamtljivosti s druge strane.

## 5.7 Osvrt na rezultate istraživanja

U ovom poglavlju izloženi su rezultati računalne analize interpretativne i doživljajne dimenzije razumijevanja umjetničkih djela. Predstavljena je metoda kvantifikacije subjektivnih aspekata percepcije umjetničkih djela korištenjem dubokim neuronskih mreža. Predložena metoda temelji se na prijenosu znanja predikcije percepcijskih obilježja naučenog u domeni prirodnih slika na domenu slikarskih djela. Istraživanje obuhvaća kvantifikaciju triju različitih percepcijskih obilježja: estetsku evaluaciju slike, osjećajni doživljaj slike i pamtljivost slike. Provedena je usporedba različitih modela konvolucijskih neuronskih mreža naučenih za predikciju pojedinog percepcijskog obilježja te evaluacija rezultata različitih modela s obzirom na dostupnost anketnim ispitivanje dobivenih podataka o zastupljenosti navedenih obilježja u slikarskim djelima na manjim skupovima slika. Na temelju te evaluacije odabrani su najbolji modeli čije se

predikcijske vrijednosti koriste za opsežnu eksploratornu analizu velike baze slikarskih djela, sa svrhom ispitivanja odnosa percepcijskih obilježja i drugih složenih značajki slike. Također je provedena analiza distribucija predikcijskih vrijednosti percepcijskih obilježja s obzirom na različite povijesno-umjetničke kategorije.

Kvalitativni i kvantitativni ishodi opisanog istraživanja pokazuju da modeli naučeni za raspoznavanje kompleksnih značajki u prirodnim slikama, daju smislene rezultate kada se primjene na digitalizirane slike umjetničkih djela. Rezultati analiza odnosa predikcijskih vrijednosti pojedinih obilježja i drugih značajki i kategorija, pokazuju da su predikcijski dobivene vrijednosti u skladu s logičkim očekivanjima, kao i s ishodima nekih postojećih psiholoških istraživanja o ljudskim preferencijama u umjetnosti, dok konačna potvrda iznesenih rezultata iziskuje usporedbu s rezultatima dobivenim prikupljanjem procjena o zastupljenosti pojedinih percepcijskih obilježja na velikom broju slikarskih djela i na temelju odziva većeg broja ispitanika. Budući da su takva istraživanja složena i skupa, rezultati izneseni u ovome radu mogu poslužiti za definiranje inicijalnih hipoteza i za prepoznavanje smjernica budućih istraživanja percepcije umjetničkih djela. Također, mogućnost praktične primjene opisane metode nalazi se u oblikovanju sustava za napredno pretraživanje kolekcija slikarskih djela.

## Poglavlje 6

### Zaključak

Intenzivni procesi digitalizacije kulturne baštine i sve veća dostupnost digitaliziranih kolekcija slikarskih djela otvorili su mogućnosti novih istraživačkih pristupa zasnovanih na razvoju i primjeni metoda računalnog vida i strojnog učenja. Preduvjet je primjeni takvih metoda sužavanje semantičkog jaza između razumijevanja značajki slike u kontekstu računalnog vida i značajki slike u kontekstu povijesti umjetnosti. U cilju premošćivanja toga jaza potrebno je uzeti u obzir semantičku višedimenzionalnost kolekcija slikarskih djela kao izvora računalno obradivih podataka. Kolekcije slikarskih djela nisu tek još jedan domenski specifičan skup slikovnih podataka, već kolektivna baština u kojoj se povijesne, kulturološke i sociološke dimenzije isprepliću sa subjektivnim estetskim i afektivnim doživljajima. Na razini vizualne informacije ta se složenost, prije svega, očituje u varijacijama odnosa sadržaja i forme. No upravo se u slikarskim djelima odnos sadržaja i forme često očituje na takav način da već i samo njihovo razlikovanje postaje upitno. U tom smislu kolekcije slikarskih djela predstavljaju zanimljiv izvor za oblikovanje kompleksnih zadataka kojima se izazivaju dosezi umjetne inteligencije. Istraživanje izloženo u ovome radu usmjereno je prema razvoju metoda za računalnu analizu slikarskih djela, a potaknuto je motivacijom dvostruke naravi. S jedne strane računalne metode omogućuju analizu određenih značajki nad velikim skupom podataka, čime se otvaraju nove perspektive na već poznati sadržaj i omogućuje potencijalno otkrivanje novog znanja u kontekstu povijesti umjetnosti. S druge strane, kolekcije slikarskih djela predstavljaju izvor podataka koji zbog svoje semantičke višeslojnosti omogućuje istraživanje dosega određenih računalnih metoda, posebice u kontekstu danas vrlo aktualnih tendencija prema većem stupnju interpretabilnosti metoda dubokih neuronskih mreža.

Razumijevanje varijacija u odnosu forme i sadržaja likovnog djela oslanja se na interpretaciju stilskih obilježja slike. Cilj istraživanja opisanog u ovom doktorskom radu ostvarenje je metoda za računalnu detekciju i analizu stilskih obilježja. Po uzoru na povijesno-umjetničku analizu likovnog djela, metode predstavljene u ovom radu podijeljene su prema tri razine razumijevanja djela: kategorizaciju, formalnu analizu i doživljajnu analizu. Tri razine pristupa



određene su različitim stupnjem egzaktnosti rezultata analize, odnosno različitim razinama apstraktnosti i subjektivnosti značajki koje se koriste u određivanju obilježja djela. Prva razina je najegzaktnija, svodi se na raspoznavanje pripadnosti slikarskih djela određenim povijesno-umjetničkim kategorijama i odgovara problematici automatske klasifikacije slika. Problematika prve razine ujedno je i najčešće obrađivana istraživačka tema u području računalne analize slikarstva. U kontekstu te razine pristupa, znanstveni doprinosi ovoga rada uključuju detaljnu analizu različitih strategija metode finog podešavanja parametara konvolucijske neuronske mreže, s ciljem određivanja optimalnih postavki za različite klasifikacijske rezultate. Predložena je metoda ispitivanja utjecaja različitih domenski specifičnih inicijalizacija mreže na ishod učenja različitih klasifikacijskih zadataka. Također, prvi puta uveden je zadatak klasifikacije slikarskih djela prema pripadajućem nacionalnom umjetničkom kontekstu. Ishodi primjene konvolucijskih neuronskih mreža prema postavkama učenja istraženima u ovome radu predstavljaju trenutno vodeće rezultate za većinu klasifikacijskih problema i slikarskih baza korištenih u ovome radu.

Za razliku od većine dosadašnjih istraživanja koja se uobičajeno ograničavaju na primjenu dubokih konvolucijskih mreža isključivo zbog klasifikacije slikarskih djela, u ovome radu istražuju se mogućnosti njihove primjene u raspoznavanju složenih značajki. Izvorni doprinos ovoga rada je metodologija oblikovanja značajki slike više razine temeljena na kvantifikaciji zastupljenosti umjetničkih koncepata u slici zasnovanoj na povezivanju informacijskih doprinosa značajki boje i teksture slike. Takve složene značajke rezultat su računalnih metoda ostvarenih po uzoru na drugu i treću razinu pristupa razumijevanju slikarskog djela. Druga razina pristupa ostvaruje se razvojem metode za detekciju specifičnih stilskih obilježja slikarskog djela temeljene na analizi vrijednosti značajki slike više razine oblikovanih prema uzoru na umjetničke koncepte. Pod umjetničkim konceptima u kontekstu ovoga rada podrazumijevaju se, između ostalih, Wölfflinovi pojmovni parovi, koji čine okosnicu formalne analize likovnih djela i razmatranja razlike stilova. U ovome radu prvi su puta predstavljeni modeli za predikciju vrijednosti značajki oblikovanih po uzoru na Wölfflinove koncepte. Predikcija navedenih obilježja ostvarena je učenjem regresijskih modela konvolucijskih neuronskih mreža. Istražene su različite postavke učenja modela za različite značajke te su predstavljeni kvantitativni i kvalitativni rezultati predikcija dobivenih primjenom tih modela.

Treća razina pristupa ostvaruje se razvojem kvantifikacije subjektivnih značajki doživljajne razine slikarskog djela. Konkretno, razmatraju se estetska evaluacija slike, osjećajni doživljaj slike i pamtljivost slike. Znanstveni doprinos ovoga rada uključuje metodu kvantifikacije spomenutih obilježja u slikarskim djelima temeljenu na primjeni konvolucijskih neuronskih mreža. Također, u ovome radu prvi se puta provodi usporedna analiza predikcijskih vrijednosti tih obilježja na kolekciji slikarskih djela, kao i analiza odnosa navedenih obilježja s drugim značajkama slike i različitim povijesno-umjetničkim kategorijama. Metode druge i treće razine

pristupa slikarskom djelu odnose se na detekciju obilježja koja nisu egzaktna već očituju visok stupanj subjektivnosti. Konačni doprinos rada evaluacija je predloženih metoda detekcije stilskih obilježja temeljena na usporedbi rezultata računalne analize s rezultatima anketnog ispitivanja ljudskog ocjenjivanja zastupljenosti umjetničkih koncepata u slikarskim djelima.

Osnovna smjernica budućeg znanstvenog istraživanja zasniva se na razvoju interdisciplinarnе suradnje, prvenstveno s ciljem otkrivanja mogućnosti primjena ostvarenih metoda u kontekstu istraživačkih problema povijesti umjetnosti. Također, zbog svoje značenjske slojevitosti kolekcije slikarskih djela predstavljaju iznimno zanimljiv skup podataka za buduća istraživanja primjenjivosti i interpretabilnosti konvolucijskih neuronskih mreža.

# Prilog A

## Pregled klasa za sve baze i zadatke klasifikacije slikarskih djela

**Tablica A.1:** Lista 10 klasa s pripadajućim brojem slika za podskup WikiArt baze za klasifikaciju žanra.

KLASA	BROJ SLIKA PO KLASI
Portrait	18829
Landscape	17512
Genre painting	16006
Abstract	14920
Cityscape	6374
Still life	3773
Nude painting	2840
Marina	1993
Flower painting	1960
Animal painting	1880

**Tablica A.2:** Lista 23 klase s pripadajućim brojem slika za podskup WikiArt baze za klasifikaciju autora.

KLASA	BROJ SLIKA PO KLASI
Vincent van Gogh	1919
Nicholas Roerich	1837
Pierre-Auguste Renoir	1403
Claude Monet	1335
Pablo Picasso	1106
Salvador Dali	1065
Marc Chagall	1012
Pyotr Konchalovsky	925
Camille Pissarro	889
Albrecht Durer	829
John Singer Sargent	796
Rembrandt	770
Gustave Dore	752
Boris Kustodiev	645
Edgar Degas	612
Paul Cezanne	583
Martiros Saryan	578
Ivan Aivazovsky	576
Eugene Boudin	555
Childe Hassam	548
Ilya Repin	540
Raphael Kirchner	525
Ivan Shishkin	520

**Tablica A.3:** Lista 27 klasa s pripadajućim brojem slika za podskup WikiArt baze za klasifikaciju stila.

KLASA	BROJ SLIKA PO KLASI
Impressionism	13434
Realism	11418
Romanticism	9061
Expressionism	7942
Post-impressionism	6652
Surrealism	5194
Baroque	4595
Symbolism	4068
Art nouveau	3606
Abstract expressionism	2921
Naive art	2835
Northern renaissance	2634
Rococo	2627
Neoclassicism	2358
Cubism	1816
Art informel	1600
Ukiyo-e	1427
Color field painting	1353
Early renaissance	1344
Mannerism	1318
High renaissance	1299
Abstract art	1269
Pop art	1261
Minimalism	1162
Magic realism	1094
Academicism	908
Lyrical abstraction	818

**Tablica A.4:** Lista 8 klasa s pripadajućim brojem slika za podskup WikiArt baze za klasifikaciju slika prema nacionalnosti autora.

KLASA	BROJ SLIKA PO KLASI
French	19936
American	16383
Russian	12161
British	9014
Italian	8814
German	6123
Spanish	4698
Japanese	3299

**Tablica A.5:** Lista 23 klasa s pripadajućim brojem slika za podskup WGA baza za klasifikaciju autora.

KLASA	BROJ SLIKA PO KLASI
Giotto di Bondone	556
Gogh, Vincent van	329
Rembrandt Harmenszoon van Rijn	314
Rubens, Peter Paul	291
Tintoretto	287
Raffaello Sanzio	286
Michelangelo Buonarroti	278
Tiziano Vecellio	268
Veronese, Paolo	264
Tiepolo, Giovanni Battista	248
Greco, El	244
Cranach, Lucas the Elder	242
Angelico, Fra	241
Memling, Hans	208
Bruegel, Pieter the Elder	205
Monet, Claude	196
Piero Della Francesca	190
Mantegna, Andrea	186
Caravaggio	185
Botticelli, Sandro	178
Bellini, Giovanni	173
Duccio di Buoninsegna	172
Ghirlandaio, Domenico	170

**Tablica A.6:** Lista 8 klasa s pripadajućim brojem slika za podskup WGA baza za klasifikaciju prema nacionalnosti autora.

kLASA	BROJ SLIKA PO KLASI
Italian	12385
Dutch	4178
French	3934
Flemish	3115
German	1604
Spanish	1177
English	566
Netherlandish	501



**Tablica A.7:** Lista 6 klasa s pripadajućim brojem slika za podskup WGA baza za klasifikaciju žanra.

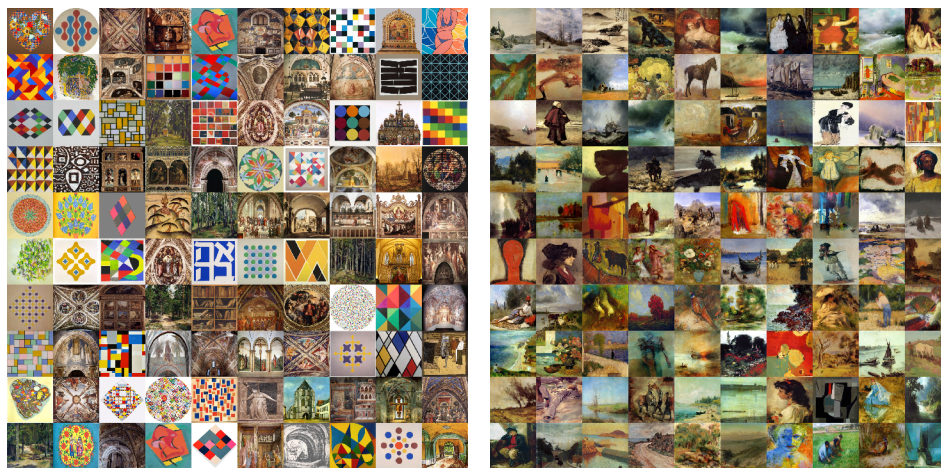
KLASA	BROJ SLIKA PO KLASI
Religious	12056
Portrait	4320
Landscape	3771
Mythological	2673
Genre painting	2533
Still life	1308

**Tablica A.8:** Lista 12 klasa s pripadajućim brojem slika za podskup WGA baza za klasifikaciju razdoblja.

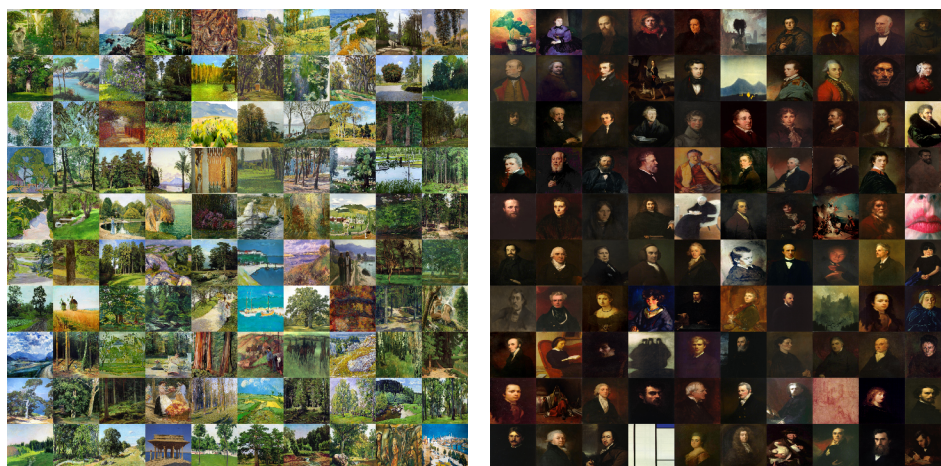
KLASA	BROJ SLIKA PO KLASI
1601-1650	4836
1501-1550	3909
1451-1500	3434
1651-1700	3062
1851-1900	2803
1551-1600	2251
1701-1750	2046
1801-1850	1483
1751-1800	1475
1401-1450	1468
1301-1350	1309
1351-1400	529

## Prilog B

# Kvalitativni rezultati ekstrakcije naučenih značajki slike više razine

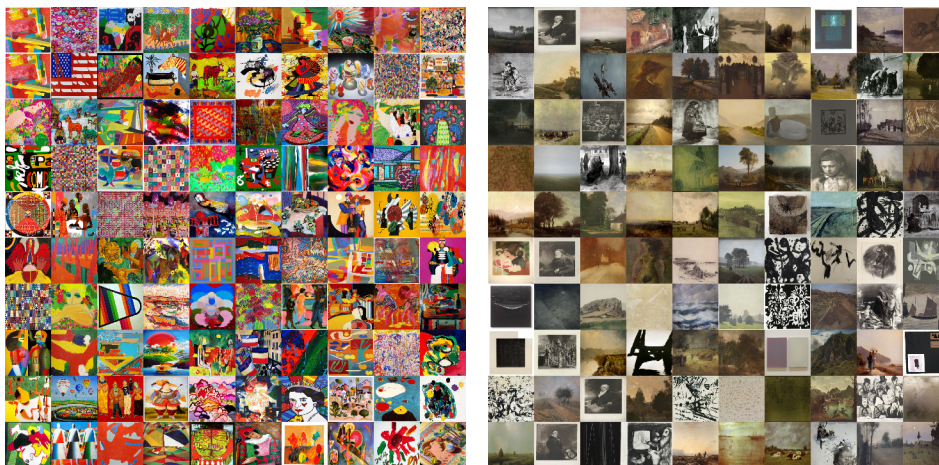


**Slika B.1:** Slike iz WikiArt baze s najvišim (lijevo) i najnižim (desno) vrijednostima značajke simetrije.

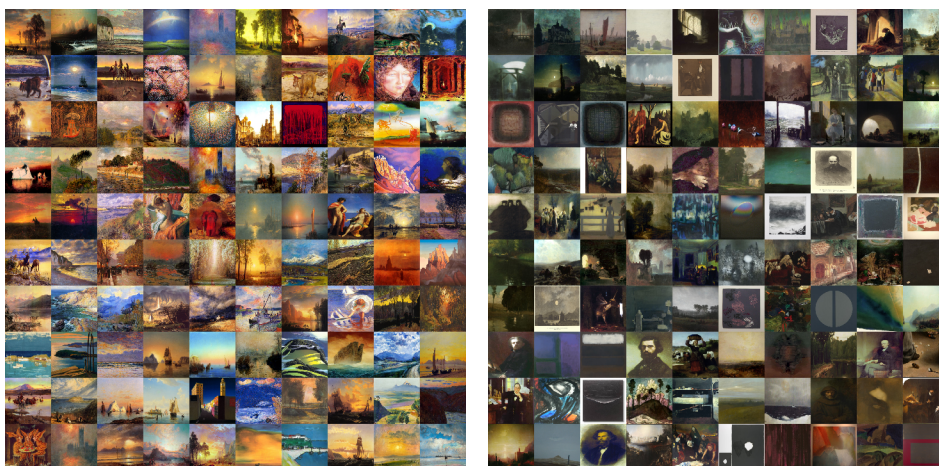


**Slika B.2:** Slike iz WikiArt baze s najvišim (lijevo) i najnižim (desno) vrijednostima značajke harmonije boja.





**Slika B.3:** Slike iz WikiArt baze s najvišim (lijevo) i najnižim (desno) vrijednostima značajke izražajnosti boja.

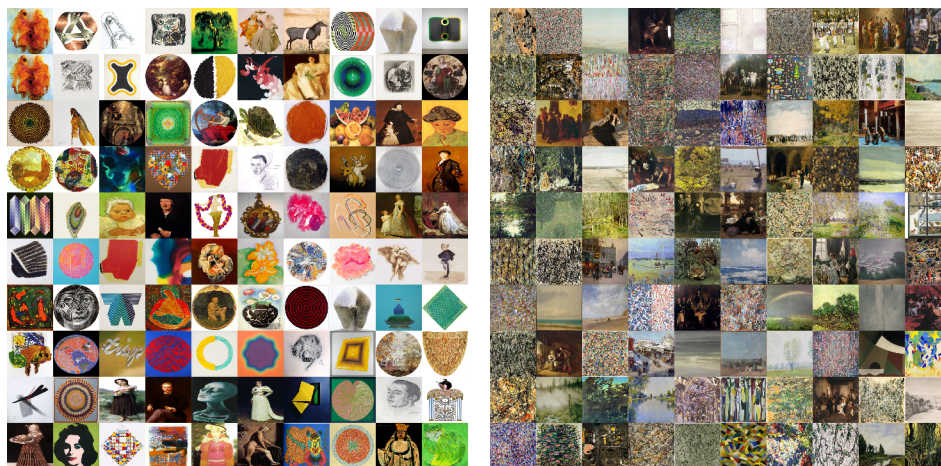


**Slika B.4:** Slike iz WikiArt baze s najvišim (lijevo) i najnižim (desno) vrijednostima značajke osvjetljenja.

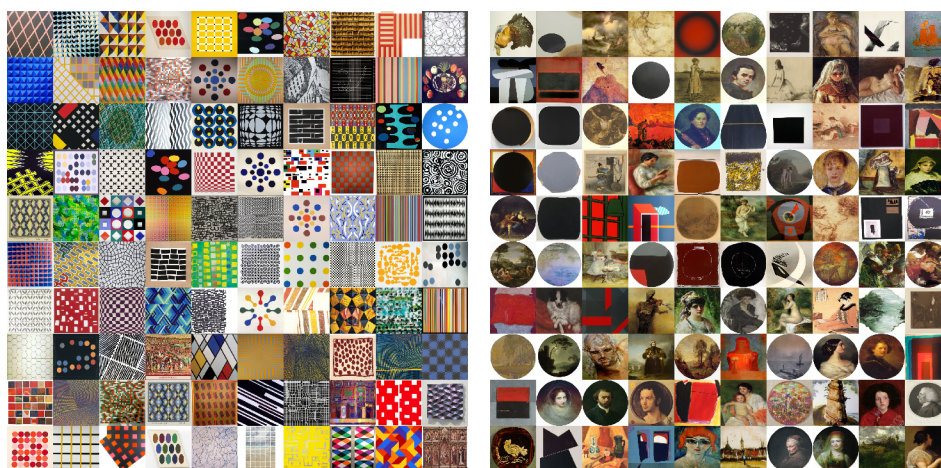


**Slika B.5:** Slike iz WikiArt baze s najvišim (lijevo) i najnižim (desno) vrijednostima značajke sadržaja.

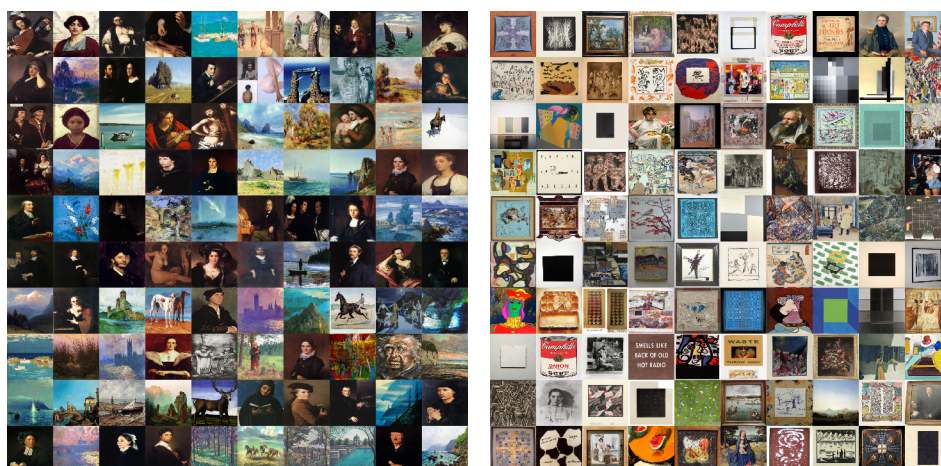




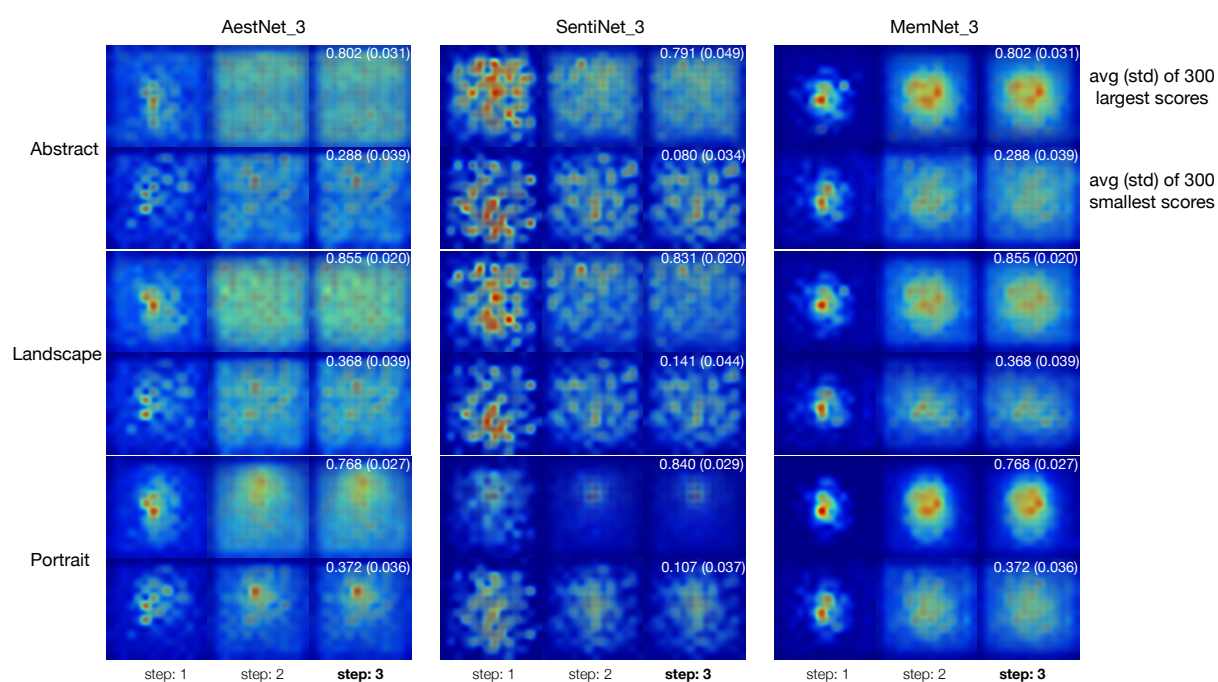
**Slika B.6:** Slike iz WikiArt baze s najvišim (lijevo) i najnižim (desno) vrijednostima značajke naglašenosti objekta.



**Slika B.7:** Slike iz WikiArt baze s najvišim (lijevo) i najnižim (desno) vrijednostima značajke ponavljanja uzoraka.



**Slika B.8:** Slike iz WikiArt baze s najvišim (lijevo) i najnižim (desno) vrijednostima značajke pravila trećine.



**Slika B.9:** Usrednjene aktivacijske mape za 300 slika s najvećim vrijednostima i 300 slika s najmanjim vrijednostima predikcija, koje pripadaju trima različitim žanrovima, za sva tri modela i sva tri LSTM koraka.

# Literatura

- [1] Keren, D., “Painter identification using local features and naive bayes”, in Pattern Recognition, 2002. Proceedings. 16th International Conference on, Vol. 2. IEEE, 2002, str. 474–477.
- [2] Günsel, B., Sariel, S., İcoğlu, O., “Content-based access to art paintings”, in Image Processing, 2005. ICIP 2005. IEEE International Conference on, Vol. 2. IEEE, 2005, str. II–558.
- [3] Lombardi, T., Cha, S.-H., Tappert, C., “A lightweight image retrieval system for paintings”, in Storage and Retrieval Methods and Applications for Multimedia 2005, Vol. 5682. International Society for Optics and Photonics, 2005, str. 236–247.
- [4] Shen, J., “Stochastic modeling western paintings for effective classification”, Pattern Recognition, Vol. 42, No. 2, 2009, str. 293–301.
- [5] Shamir, L., Macura, T., Orlov, N., Eckley, D. M., Goldberg, I. G., “Impressionism, expressionism, surrealism: Automated recognition of painters and schools of art”, ACM Transactions on Applied Perception (TAP), Vol. 7, No. 2, 2010, str. 8.
- [6] Shamir, L., Tarakhovskiy, J. A., “Computer analysis of art”, Journal on Computing and Cultural Heritage (JOCCH), Vol. 5, No. 2, 2012, str. 7.
- [7] Bressan, M., Cifarelli, C., Perronnin, F., “An analysis of the relationship between painters based on their work”, in Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on. IEEE, 2008, str. 113–116.
- [8] Wang, Y., Takatsuka, M., “A framework towards quantified artistic influences analysis”, in Digital Image Computing Techniques and Applications (DICTA), 2012 International Conference on. IEEE, 2012, str. 1–8.
- [9] Zujovic, J., Gandy, L., Friedman, S., Pardo, B., Pappas, T. N., “Classifying paintings by artistic genre: An analysis of features & classifiers”, in Multimedia Signal Processing, 2009. MMSP’09. IEEE International Workshop on. IEEE, 2009, str. 1–5.

- [10] Čuljak, M., Mikuš, B., Jež, K., Hadjić, S., “Classification of art paintings by genre”, in MIPRO, 2011 Proceedings of the 34th International Convention. IEEE, 2011, str. 1634–1639.
- [11] Arora, R. S., Elgammal, A., “Towards automated classification of fine-art painting style: A comparative study”, in Pattern Recognition (ICPR), 2012 21st International Conference on. IEEE, 2012, str. 3541–3544.
- [12] Mensink, T., Van Gemert, J., “The rijksmuseum challenge: Museum-centered visual recognition”, in Proceedings of International Conference on Multimedia Retrieval. ACM, 2014, str. 451.
- [13] Florea, C., Condorovici, R., Vertan, C., Butnaru, R., Florea, L., Vrânceanu, R., “Pandora: Description of a painting database for art movement recognition with baselines and perspectives”, in Signal Processing Conference (EUSIPCO), 2016 24th European. IEEE, 2016, str. 918–922.
- [14] Saleh, B., Elgammal, A., “Large-scale classification of fine-art paintings: Learning the right metric on the right feature”, International Journal for Digital Art History, Vol. 0, No. 2, 2016.
- [15] Cetinic, E., Grgic, S., “Automated painter recognition based on image feature extraction”, in ELMAR, 2013 55th International Symposium. IEEE, 2013, str. 19–22.
- [16] [dataset] Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., Fei-Fei, L., “Imagenet: A large-scale hierarchical image database”, in Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on. IEEE, 2009, str. 248–255.
- [17] Krizhevsky, A., Sutskever, I., Hinton, G. E., “Imagenet classification with deep convolutional neural networks”, in Advances in neural information processing systems, 2012, str. 1097–1105.
- [18] Karayev, S., Trentacoste, M., Han, H., Agarwala, A., Darrell, T., Hertzmann, A., Winnemoeller, H., “Recognizing image style”, in Proceedings of the British Machine Vision Conference. BMVA Press, 2014.
- [19] Bar, Y., Levy, N., Wolf, L., “Classification of artistic styles using binarized features derived from a deep neural network.”, in ECCV Workshops (1), 2014, str. 71–84.
- [20] Agarwal, S., Karnick, H., Pant, N., Patel, U., “Genre and style based painting classification”, in Applications of Computer Vision (WACV), 2015 IEEE Winter Conference on. IEEE, 2015, str. 588–594.



- [21] Cetinic, E., Grgic, S., “Genre classification of paintings”, in ELMAR, 2016 International Symposium, 2016, str. 201–204.
- [22] David, O. E., Netanyahu, N. S., “Deeppainter: Painter classification using deep convolutional autoencoders”, in International Conference on Artificial Neural Networks. Springer, 2016, str. 20–28.
- [23] Girshick, R., Donahue, J., Darrell, T., Malik, J., “Rich feature hierarchies for accurate object detection and semantic segmentation”, in Proceedings of the IEEE conference on computer vision and pattern recognition, 2014, str. 580–587.
- [24] Hentschel, C., Wiradarma, T. P., Sack, H., “Fine tuning cnns with scarce training data — adapting imagenet to art epoch classification”, in Image Processing (ICIP), 2016 IEEE International Conference on. IEEE, 2016, str. 3693–3697.
- [25] Tan, W. R., Chan, C. S., Aguirre, H. E., Tanaka, K., “Ceci n’est pas une pipe: A deep convolutional network for fine-art paintings classification”, in Image Processing (ICIP), 2016 IEEE International Conference on. IEEE, 2016, str. 3703–3707.
- [26] Lecoutre, A., Negrevergne, B., Yger, F., “Recognizing art style automatically in painting with deep learning”, in Asian Conference on Machine Learning, 2017, str. 327–342.
- [27] van Noord, N., Hendriks, E., Postma, E., “Toward discovery of the artist’s style: Learning to recognize artists by their artworks”, IEEE Signal Processing Magazine, Vol. 32, No. 4, 2015, str. 46–54.
- [28] Crowley, E. J., Zisserman, A., “In search of art”, in ECCV Workshops (1). Springer, 2014, str. 54–70.
- [29] Seguin, B., Striolo, C., Kaplan, F. *et al.*, “Visual link retrieval in a database of paintings”, in ECCV Workshops (1). Springer, 2016, str. 753–767.
- [30] Gando, G., Yamada, T., Sato, H., Oyama, S., Kurihara, M., “Fine-tuning deep convolutional neural networks for distinguishing illustrations from photographs”, Expert Systems with Applications, Vol. 66, 2016, str. 295–301.
- [31] Chu, W.-T., Wu, Y.-L., “Deep correlation features for image style classification”, in Proceedings of the 2016 ACM on Multimedia Conference. ACM, 2016, str. 402–406.
- [32] Khan, F. S., Beigpour, S., Van de Weijer, J., Felsberg, M., “Painting-91: a large scale database for computational painting categorization”, Machine vision and applications, Vol. 25, No. 6, 2014, str. 1385–1397.



- [33] Strezoski, G., Worring, M., “Omniart: A large-scale artistic benchmark”, *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, Vol. 14, No. 4, 2018, str. 88.
- [34] Leslie, L., Chua, T.-S., Ramesh, J., “Annotation of paintings with high-level semantic concepts using transductive inference and ontology-based concept disambiguation”, in *Proceedings of the 15th ACM international conference on Multimedia*. ACM, 2007, str. 443–452.
- [35] Yelizaveta, M., Tat-Seng, C., Irina, A., Jain, R., “Representation and retrieval of paintings based on art history concepts”, in *Multimedia and Expo, 2004. ICME'04. 2004 IEEE International Conference on*, Vol. 2. IEEE, 2004, str. 1023–1026.
- [36] Corridoni, J. M., Del Bimbo, A., Pala, P., “Retrieval of paintings using effects induced by color features”, in *Proceedings 1998 IEEE International Workshop on Content-Based Access of Image and Video Database*. IEEE, 1998, str. 2–11.
- [37] Colombo, C., Del Bimbo, A., Pala, P., “Semantics in visual information retrieval”, *IEEE MultiMedia*, Vol. 6, No. 3, 1999, str. 38–53.
- [38] Stanchev, P., Green Jr, D., Dimitrov, B., “High level color similarity retrieval”, dostupno na: <http://hdl.handle.net/10525/950> (10. rujna 2018.). 2003.
- [39] Yelizaveta, M., Tat-Seng, C., Irina, A., “Analysis and retrieval of paintings using artistic color concepts”, in *Multimedia and Expo, 2005. ICME 2005. IEEE International Conference on*. IEEE, 2005, str. 1246–1249.
- [40] Ivanova, K., Stanchev, P., “Color harmonies and contrasts search in art image collections”, in *Advances in Multimedia, 2009. MMEDIA'09. First International Conference on*. IEEE, 2009, str. 180–187.
- [41] Machajdik, J., Hanbury, A., “Affective image classification using features inspired by psychology and art theory”, in *Proceedings of the 18th ACM international conference on Multimedia*. ACM, 2010, str. 83–92.
- [42] Yelizaveta, M., Tat-Seng, C., Ramesh, J., “Transductive inference using multiple experts for brushwork annotation in paintings domain”, in *Proceedings of the 14th ACM international conference on Multimedia*. ACM, 2006, str. 157–160.
- [43] Tanaka, S., Kurumizawa, J., Inokuchi, S., Iwadate, Y., “Composition analyzer: support tool for composition analysis on painting masterpieces”, *Knowledge-based systems*, Vol. 13, No. 7-8, 2000, str. 459–470.

- [44] Wei-ning, W., Ying-lin, Y., Jian-chao, Z., “Image emotional classification: static vs. dynamic”, in Systems, Man and Cybernetics, 2004 IEEE International Conference on, Vol. 7. IEEE, 2004, str. 6407–6411.
- [45] Elgammal, A., Liu, B., Kim, D., Elhoseiny, M., Mazzone, M., “The shape of art history in the eyes of the machine”, in Thirty-Second AAAI Conference on Artificial Intelligence, 2018.
- [46] Kim, D., Liu, B., Elgammal, A., Mazzone, M., “Finding principal tics of style in art”, in Semantic Computing (ICSC), 2018 IEEE 12th International Conference on. IEEE, 2018, str. 156–163.
- [47] Brachmann, A., Barth, E., Redies, C., “Using cnn features to better understand what makes visual artworks special”, *Frontiers in psychology*, Vol. 8, 2017, str. 830.
- [48] Lee, B., Kim, D., Jeong, H., Sun, S., Park, J., “Understanding the historic emergence of diversity in painting via color contrast”, dostupno na: <https://arxiv.org/abs/1701.07164v1> (12. rujna 2018). 2017.
- [49] Sigaki, H. Y., Perc, M., Ribeiro, H. V., “History of art paintings through the lens of entropy and complexity”, *Proceedings of the National Academy of Sciences*, Vol. 115, No. 37, 2018, str. E8585–E8594.
- [50] Simonyan, K., Zisserman, A., “Very deep convolutional networks for large-scale image recognition”, dostupno na: <http://arxiv.org/abs/1409.1556> (15. siječnja 2016). 2014.
- [51] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A., “Going deeper with convolutions”, in Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, str. 1–9.
- [52] He, K., Zhang, X., Ren, S., Sun, J., “Deep residual learning for image recognition”, in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, str. 770–778.
- [53] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., “Dropout: a simple way to prevent neural networks from overfitting”, *The Journal of Machine Learning Research*, Vol. 15, No. 1, 2014, str. 1929–1958.
- [54] Yosinski, J., Clune, J., Bengio, Y., Lipson, H., “How transferable are features in deep neural networks?”, in Advances in neural information processing systems, 2014, str. 3320–3328.

- [55] Cetinic, E., Lipic, T., Grgic, S., “Fine-tuning convolutional neural networks for fine art classification”, *Expert Systems with Applications*, Vol. 114, 2018, str. 107–118.
- [56] van Noord, N., Postma, E., “Learning scale-variant and scale-invariant features for deep image classification”, *Pattern Recognition*, Vol. 61, 2017, str. 583–592.
- [57] Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., Darrell, T., “Caffe: Convolutional architecture for fast feature embedding”, in *Proceedings of the 22nd ACM international conference on Multimedia*. ACM, 2014, str. 675–678.
- [58] Deng, J., Berg, A., Satheesh, S., Su, H., Khosla, A., Fei-Fei, L., “Ilsvrc-2012, 2012”, URL <http://www.image-net.org/challenges/LSVRC>, 2012.
- [59] Zhou, B., Lapedriza, A., Xiao, J., Torralba, A., Oliva, A., “Learning deep features for scene recognition using places database”, in *Advances in neural information processing systems*, 2014, str. 487–495.
- [60] Khosla, A., Raju, A. S., Torralba, A., Oliva, A., “Understanding and predicting image memorability at a large scale”, in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, str. 2390–2398.
- [61] Campos, V., Jou, B., Giro-i Nieto, X., “From pixels to sentiment: Fine-tuning cnns for visual sentiment prediction”, *Image and Vision Computing*, Vol. 65, 2017, str. 15–22.
- [62] You, Q., Luo, J., Jin, H., Yang, J., “Robust image sentiment analysis using progressively trained and domain transferred deep networks.”, in *AAAI*, 2015, str. 381–388.
- [63] Wölfflin, H., *Temeljni pojmovi povijesti umjetnosti*. Kontura, Institut za povijest umjetnosti, Zagreb, 1998.
- [64] Cetinic, E., Lipic, T., Grgic, S., “A deep learning perspective on beauty, sentiment, and remembrance of art”, *IEEE Access*, Vol. 7, 2019, str. 73 694–73 710.
- [65] Hoenig, F., “Defining computational aesthetics”, in *Computational Aesthetics 2005: Eurographics Workshop on Computational Aesthetics in Graphics, Visualization and Imaging 2005*, Girona, Spain, May 18-20, 2005, 2005, str. 13–18.
- [66] Marchesotti, L., Perronnin, F., Larlus, D., Csurka, G., “Assessing the aesthetic quality of photographs using generic image descriptors”, in *IEEE International Conference on Computer Vision, ICCV 2011, Barcelona, Spain, November 6-13, 2011*, 2011, str. 1784–1791.

- [67] Dhar, S., Ordonez, V., Berg, T. L., “High level describable attributes for predicting aesthetics and interestingness”, in The 24th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2011, Colorado Springs, CO, USA, 20-25 June 2011, 2011, str. 1657–1664.
- [68] Murray, N., Marchesotti, L., Perronnin, F., “AVA: A large-scale database for aesthetic visual analysis”, in 2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, June 16-21, 2012, 2012, str. 2408–2415.
- [69] Dong, Z., Shen, X., Li, H., Tian, X., “Photo quality assessment with DCNN that understands image well”, in MultiMedia Modeling - 21st International Conference, MMM 2015, Sydney, NSW, Australia, January 5-7, 2015, Proceedings, Part II, 2015, str. 524–535.
- [70] Lu, X., Lin, Z., Jin, H., Yang, J., Wang, J. Z., “Rapid: Rating pictorial aesthetics using deep learning”, in Proceedings of the 22Nd ACM International Conference on Multimedia, ser. MM '14. New York, NY, USA: ACM, 2014, str. 457–466.
- [71] Kong, S., Shen, X., Lin, Z. L., Mech, R., Fowlkes, C. C., “Photo aesthetics ranking network with attributes and content adaptation”, in Computer Vision - ECCV 2016 - 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part I, 2016, str. 662–679.
- [72] Deng, Y., Loy, C. C., Tang, X., “Image aesthetic assessment: An experimental survey”, IEEE Signal Process. Mag., Vol. 34, No. 4, 2017, str. 80–106.
- [73] Brachmann, A., Redies, C., “Computational and experimental approaches to visual aesthetics”, Front. Comput. Neurosci., Vol. 2017, 2017.
- [74] Machajdik, J., Hanbury, A., “Affective image classification using features inspired by psychology and art theory”, in Proceedings of the 18th ACM International Conference on Multimedia, ser. MM '10. New York, NY, USA: ACM, 2010, str. 83–92.
- [75] Borth, D., Ji, R., Chen, T., Breuel, T., Chang, S.-F., “Large-scale visual sentiment ontology and detectors using adjective noun pairs”, in Proceedings of the 21st ACM International Conference on Multimedia, ser. MM '13. New York, NY, USA: ACM, 2013, str. 223–232.
- [76] Xu, C., Cetintas, S., Lee, K.-C., Li, L.-J., “Visual sentiment prediction with deep convolutional neural networks”, arXiv preprint arXiv:1411.5731, 2014.

- [77] You, Q., Luo, J., Jin, H., Yang, J., “Robust image sentiment analysis using progressively trained and domain transferred deep networks”, in Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, January 25-30, 2015, Austin, Texas, USA., 2015, str. 381–388.
- [78] Campos, V., Jou, B., Giró i Nieto, X., “From pixels to sentiment: Fine-tuning cnns for visual sentiment prediction”, *Image Vision Comput.*, Vol. 65, 2017, str. 15–22.
- [79] Isola, P., Xiao, J., Parikh, D., Torralba, A., Oliva, A., “What makes a photograph memorable?”, *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. 36, No. 7, 2014, str. 1469–1482.
- [80] Mancas, M., Meur, O. L., “Memorability of natural scenes: The role of attention”, in IEEE International Conference on Image Processing, ICIP 2013, Melbourne, Australia, September 15-18, 2013, 2013, str. 196–200.
- [81] Kim, J., Yoon, S., Pavlovic, V., “Relative spatial features for image memorability”, in Proceedings of the 21st ACM International Conference on Multimedia, ser. MM '13. New York, NY, USA: ACM, 2013, str. 761–764.
- [82] Goetschalckx, L., Vanmarcke, S., Moors, P., Wagemans, J., “Are memorable images easier to categorize rapidly?”, *Journal of Vision*, Vol. 17, No. 10, 2017, str. 98–98.
- [83] Amengual, X., Bosch, A., de la Rosa, J. L., “How to measure memorability and social interestingness of images: A review”, *International Journal of Pattern Recognition and Artificial Intelligence*, Vol. 31, No. 2, 2017, str. 1–35.
- [84] Khosla, A., Raju, A. S., Torralba, A., Oliva, A., “Understanding and predicting image memorability at a large scale”, in 2015 IEEE International Conference on Computer Vision, ICCV 2015, Santiago, Chile, December 7-13, 2015, 2015, str. 2390–2398.
- [85] Fajtl, J., Argyriou, V., Monekosso, D., Remagnino, P., “Amnet: Memorability estimation with attention”, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, str. 6363–6372.
- [86] Cetinic, E., Lipic, T., Grgic, S., “How convolutional neural networks remember art”, in 2018 25th International Conference on Systems, Signals and Image Processing (IW-SSIP). IEEE, 2018, str. 1–5.
- [87] “Jenaesthetics dataset”, dostupno na: <http://www.inf-cv.uni-jena.de/en/jenaesthetics> (30. listopada 2018.).
- [88] Yanulevskaya, V., Uijlings, J., Bruni, E., Sartori, A., Zamboni, E., Bacci, F., Melcher, D., Sebe, N., “In the eye of the beholder: employing statistical analysis and eye tracking for

- analyzing abstract paintings”, in Proceedings of the 20th ACM international conference on Multimedia. ACM, 2012, str. 349–358.
- [89] Mohammad, S., Kiritchenko, S., “Wikiart emotions: An annotated dataset of emotions evoked by art”, in Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC-2018), 2018.
- [90] Ren, J., Shen, X., Lin, Z., Mech, R., Foran, D. J., “Personalized image aesthetics”, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017.
- [91] Katsurai, M., Satoh, S., “Image sentiment analysis using latent correlations among visual, textual, and sentiment views”, in Acoustics, Speech and Signal Processing (ICASSP), 2016 IEEE International Conference on. IEEE, 2016, str. 2837–2841.
- [92] Isola, P., Xiao, J., Torralba, A., Oliva, A., “What makes an image memorable?”, in CVPR 2011. IEEE, 2011, str. 145–152.
- [93] Amirshahi, S. A., Denzler, J., Redies, C., “Jenaesthetics—a public dataset of paintings for aesthetic research”, in Poster Workshop at the European Conference on Computer Vision, 2013.
- [94] Amirshahi, S. A., Redies, C., Denzler, J., “How self-similar are artworks at different levels of spatial resolution?”, in Proceedings of the Symposium on Computational Aesthetics. ACM, 2013, str. 93–100.
- [95] Amirshahi, S. A., Hayn-Leichsenring, G. U., Denzler, J., Redies, C., “Jenaesthetics subjective dataset: analyzing paintings by subjective scores”, in Workshop at the European Conference on Computer Vision. Springer, 2014, str. 3–19.
- [96] Crowley, E. J., Zisserman, A., “In search of art”, in Computer Vision - ECCV 2014 Workshops - Zurich, Switzerland, September 6-7 and 12, 2014, Proceedings, Part I, 2014, str. 54–70.
- [97] Jin, X., Wu, L., Li, X., Zhang, X., Chi, J., Peng, S., Ge, S., Zhao, G., Li, S., “Ilgnet: inception modules with connected local and global features for efficient image aesthetic quality classification using domain adaptation”, IET Computer Vision, 2018.
- [98] Greff, K., Srivastava, R. K., Koutník, J., Steunebrink, B. R., Schmidhuber, J., “Lstm: A search space odyssey”, IEEE transactions on neural networks and learning systems, Vol. 28, No. 10, 2017, str. 2222–2232.
- [99] Breiman, L., Friedman, J. H., Olshen, R. A., Stone, C. J., Classification and Regression Trees. Wadsworth, 1984.

- [100] Loh, W., “Classification and regression trees”, Wiley Interdisc. Rev.: Data Mining and Knowledge Discovery, Vol. 1, No. 1, 2011, str. 14–23.
- [101] Breiman, L., “Random forests”, Machine Learning, Vol. 45, No. 1, 2001, str. 5–32.
- [102] Friedman, J. H., “Greedy function approximation: A gradient boosting machine.”, Ann. Statist., Vol. 29, No. 5, 10 2001, str. 1189–1232.
- [103] Palmer, S. E., Schloss, K. B., Sammartino, J., “Visual aesthetics and human preference”, Annual review of psychology, Vol. 64, 2013, str. 77–107.
- [104] Komar, V., Melamid, A., Painting by numbers: Komar and Melamid’s scientific guide to art. University of California Press, 1999.
- [105] Mandel, I., “Aesthetic, art-historical and economic values in painting: Empirical study”, dostupno na: <http://dx.doi.org/10.2139/ssrn.3160419> (15. veljače 2019.). 2018.

# Životopis

Eva Cetinić rođena je 1988. godine u Zagrebu. Diplomirala je 2012. godine na Fakultetu elektrotehnike i računarstva Sveučilišta u Zagrebu, smjer Informacijske i komunikacijske tehnologije. Iste godine upisuje doktorski studij na Fakultetu elektrotehnike i računarstva Sveučilišta u Zagrebu. Od 2015. godine zaposlena je kao stručni savjetnik na Institutu Ruđer Bošković u Centru za Informatiku i računarstvo. Tijekom rada na Institutu sudjelovala je na međunarodnim Obzor2020 projektima usmjerenima na uspostavu i diseminaciju tehnoloških rješenja u području digitalne humanistike. Njezin znanstvenoistraživački interes usmjeren je na razvoj i primjenu metoda obrade i analize slike i strojnog učenja u području humanističkih znanosti, posebice digitalne povijesti umjetnosti. Objavila je više znanstvenih radova i sudjelovala na različitim međunarodnim konferencijama.

## Popis objavljenih djela

### Rad u časopisima

1. Cetinic, E., Lipic, T., Grgic, S., “A Deep Learning Perspective on Beauty, Sentiment, and Remembrance of Art”, IEEE Access, 7, 2019, str. 73694-73710.
2. Cetinic, E., Lipic, T., Grgic, S., “Fine-tuning convolutional neural networks for fine art classification”, Expert Systems with Applications, Vol. 114, 2018, str. 107–118.

### Znanstveni radovi u zbornicima skupova s međunarodnom recenzijom

1. Cetinic, E., Lipic, T., Grgic, S., “How convolutional neural networks remember art”, in 2018 25th International Conference on Systems, Signals and Image Processing (IWSSIP). IEEE, 2018, str. 1–5.
2. Cetinic, E., Grgic, S., “Genre classification of paintings”, in ELMAR, 2016 International Symposium, 2016, str. 201–204.
3. Cetinic, E., Grgic, S., “Automated painter recognition based on image feature extraction”, in ELMAR, 2013 55th International Symposium. IEEE, 2013, str. 19–22.



# Biography

Eva Cetinić was born in 1988 in Zagreb. She received her M.Sc. in Information and communication technology at the Faculty of Electrical Engineering and Computing of University of Zagreb in 2012. In the same year she enrolled in the Ph.D. programme at the Faculty of Electrical Engineering and Computing of University of Zagreb. Since 2015 she is employed as a professional associate at the Centre for Informatics and Computing at the Rudjer Boskovic Institute. Within the scope of her work, she was actively involved in two Horizon 2020 projects by working on establishing and disseminating technological solutions for the digital arts and humanities research community. Her research interests focus on computer vision, image processing and machine learning and their application to digital art and humanities related data. She has published several scientific papers and participated at various international conferences.