

# Detekcija i klasifikacija pojedinih zuba iz panoramske rendgenske snimke maloljetnika

---

Oreč, Marija

Master's thesis / Diplomski rad

2024

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:168:956452>

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

Download date / Datum preuzimanja: **2025-03-22**



Repository / Repozitorij:

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



SVEUČILIŠTE U ZAGREBU  
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

DIPLOMSKI RAD br. 464

**DETEKCIJA I KLASIFIKACIJA POJEDINIH ZUBA IZ  
PANORAMSKE RENDGENSKE SNIMKE MALOLJETNIKA**

Marija Oreč

Zagreb, lipanj 2024.

SVEUČILIŠTE U ZAGREBU  
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

DIPLOMSKI RAD br. 464

**DETEKCIJA I KLASIFIKACIJA POJEDINIH ZUBA IZ  
PANORAMSKE RENDGENSKE SNIMKE MALOLJETNIKA**

Marija Oreč

Zagreb, lipanj 2024.

## DIPLOMSKI ZADATAK br. 464

Pristupnica: **Marija Oreč (0036518163)**

Studij: Računarstvo

Profil: Znanost o podacima

Mentor: prof. dr. sc. Marko Subašić

Zadatak: **Detekcija i klasifikacija pojedinih zuba iz panoramske rendgenske snimke maloljetnika**

### Opis zadatka:

Ispitati metode za detekciju i klasifikaciju objekata u slikama, uz posebni naglasak na metode dubokog učenja. Prikupiti odgovarajuću bazu panoramskih rendgenskih snimaka maloljetnih osoba s vidljivim trajnim i mliječnim zubima te označenim lokacijama pojedinih zuba. Implementirati metodu za detekciju pojedinih zuba. Ispitati točnost na prikupljenoj bazi slika uz analizu točnosti za svaki pojedini zub kao i za više različitih grupiranja zuba. Ispitati postupke augmentacije baze ulaznih slika s ciljem poboljšanja performansi. Za implementaciju koristiti programski jezik Python.

Rok za predaju rada: 28. lipnja 2024.

*Za mog didu*

## Sadržaj

Uvod .....	1
1. Inovacije u zdravstvu .....	2
1.1. Razvoj i rast tehnologije u zdravstvu .....	2
1.2. Umjetna inteligencija i dentalna medicina .....	3
2. Analiza slika .....	6
2.1. Digitalna slika.....	6
2.2. Neuronske mreže .....	8
2.2.1. Konvolucijske neuronske mreže.....	10
2.2.2. Modeli neuronskih mreža .....	14
3. Implementacija .....	19
3.1. Korišteni podaci.....	19
3.2. YOLO model .....	23
3.3. Detekcija.....	30
3.3.1. YOLO gubitak .....	30
3.3.2. Evaluacijske mjere.....	32
3.3.3. Testiranje .....	33
3.4. Klasifikacija.....	35
3.4.1. Analiza točnosti za vrstu zubi.....	38
3.4.2. Analiza točnosti po grupama .....	43
3.4.3. Analiza točnosti po godinama .....	44
3.4.4. Analiza točnosti za spol.....	47
Zaključak .....	48
Literatura .....	49
Sažetak.....	51
Summary.....	52

# Uvod

Evidentno je kako, u posljednjih nekoliko desetljeća, tehnologija postaje neizostavan dio našeg svakodnevnog života, a samim time i njezin utjecaj na medicinu i zdravstvo. Razvoj tehnologije omogućio je napredak u dijagnostici, liječenju i praćenju pacijenata, pružajući zdravstvenim djelatnicima dodatna sredstva koja značajno poboljšavaju i ubrzavaju njihov rad. Osim projektiranja i konstruiranja sve boljih zdravstvenih uređaja i opreme, ističe se primjena računarstva i umjetne inteligencije, koja donosi nove mogućnosti u obradi i analizi podataka, omogućavajući preciznije i brže donošenje medicinskih odluka. Sve smo svjesniji kako količina podataka konstantno raste, a prikupljamo ih iz različitih izvora kao što su laboratorijska istraživanja, zdravstvena literatura, različite medicinske slike, nalazi pacijenata itd. Shodno tome, metodama umjetne inteligencije, točnije strojnog učenja, želimo iskoristiti potencijal tih podataka i sve te informacije analizirati i razviti algoritme koji će prepoznati obrasce pomažući nam u budućim izazovima. Jedan od interesantnijih i opsežnije istraženih područja je zasigurno digitalna obrada i analiza slike, izričito detekcija i klasifikacija objekata na medicinskim slikama. Ovaj proces zahtjeva jako dobru pripremu i prikupljanje odgovarajuće baze podataka te pomoć zdravstvenih specijalista određenog područja tijekom razvoja. Primjena analize slika koje se dotičemo u ovom radu vezana je uz dentalnu medicinu, a konkretno ćemo istražiti korištenje metoda dubokog učenja, specifičnije neuronskih mreža, na skupu prikupljenih panoramskih rendgenskih snimaka maloljetnih osoba s vidljivim trajnim i mliječnim zubima te označenim lokacijama pojedinih zuba. Iako su mnogi i dalje skeptični prema primjeni programskih rješenja u medicini i zdravstvu, vjerujemo da bi ih uspješni rezultati, poput našeg slučaja, mogli uvjeriti u suprotno. Ovakva istraživanja i ostvarena rješenja, koja često nadmašuju zdravstvene djelatnike u točnosti dijagnoza, mogu uvelike pomoći u planiranju liječenja i smanjiti broj pogrešaka, čime bi stomatolozima i drugim dentalnim radnicima omogućili da više vremena posvete suradnji s pacijentima i pružanju kvalitetnije skrbi.

# 1. Inovacije u zdravstvu

## 1.1. Razvoj i rast tehnologije u zdravstvu

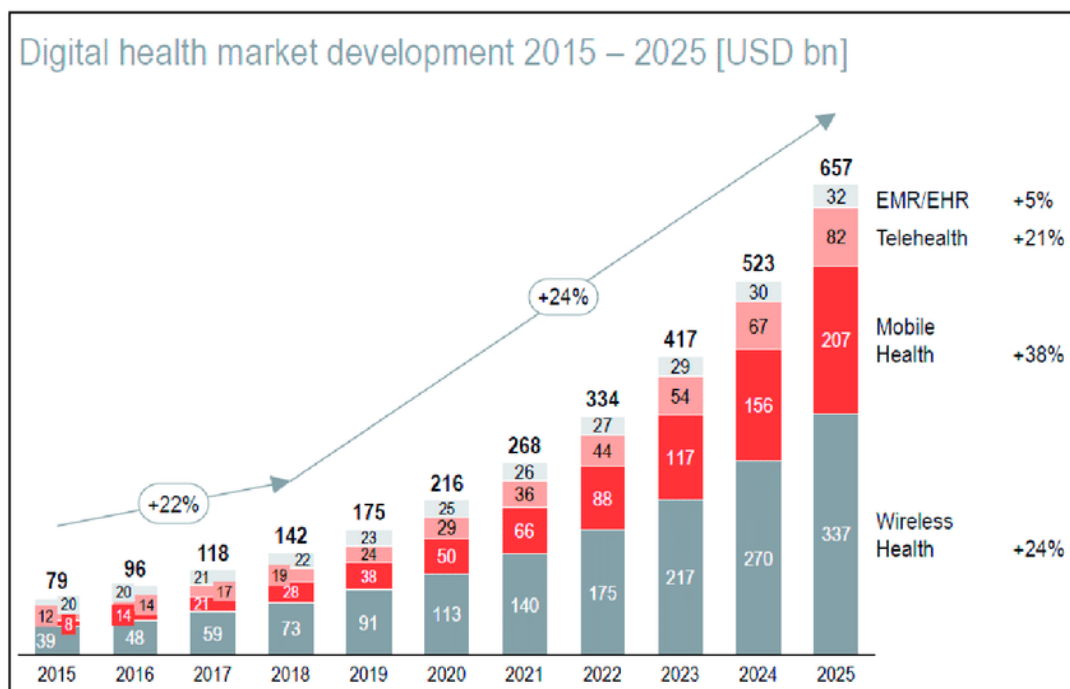
Razvoj tehnologije u zdravstvu doživio je nevjerojatan napredak tijekom posljednjih nekoliko desetljeća, transformirajući način na koji se pruža zdravstvena skrb. Početkom 20. stoljeća, otkriće rendgenskih zraka postavilo je temelje za modernu medicinsku dijagnostiku. Kasniji razvoj kompjuterizirane tomografije (CT) i magnetske rezonancije (MRI) omogućio je detaljnije preglede unutarnjih organa, što je značajno poboljšalo dijagnostičke mogućnosti liječnika. Ove tehnologije su bile revolucionarne jer su omogućile liječnicima da precizno i brzo identificiraju medicinske probleme koji su prije bili teško uočljivi [1].

U nešto novije vrijeme, sve češće čujemo kako umjetna inteligencija (AI) i strojno učenje donose preokret u analizi medicinskih podataka. Algoritmi AI-a sada mogu analizirati medicinske slike s velikom točnošću, često nadmašujući ljudske stručnjake. Ova tehnologija omogućuje bržu i precizniju dijagnozu, što je posebno korisno u ranoj detekciji bolesti poput raka. Uz to, AI se koristi za personalizaciju tretmana, predviđanje ishoda bolesti i otkrivanje novih lijekova [2].

Osim spomenutih, jedan od ključnih alata postala je i telemedicina, osobito tijekom pandemije COVID-19, omogućujući liječnicima i pacijentima interakciju na daljinu. Ovo je smanjilo potrebu za fizičkim posjetama zdravstvenim ustanovama, smanjujući rizik od širenja zaraze i poboljšavajući pristup zdravstvenoj skrbi u ruralnim i manje razvijenim područjima [3]. Telemedicina omogućuje kontinuiranu skrb za kronične bolesnike, smanjujući hospitalizacije i poboljšavajući kvalitetu života.

Uz prethodno navedene primjere postoji još mnoštvo područja u medicini i zdravstvu gdje uvodimo tehnološke napretke i sigurno će ih biti sve više. Kako se tržište digitalnog zdravstva kroz godine razvijalo, vidljivo je iz priloženog grafičkog prikaza (Sl. 1.1). Može se uočiti rast tržišta u milijardama USD od 2015. do danas te očekivani trend rasta i prema 2025. godini. To nam pokazuje koliko je digitalizacija zdravstva brzo napredovala i koliki potencijal ima za budući rast. Očekuje se da će taj trend rasta i dalje biti prisutan, potaknut kontinuiranim inovacijama i sve većom potrebom za efikasnijom i pristupačnijom zdravstvenom skrbi.





Sl. 1.1 Razvoj tržišta digitalnog zdravstva

Izvor: [https://www.researchgate.net/figure/Figure-e-Growth-of-the-global-digital-health-market-from-mmmm-to-oooo-Reprinted-with\\_fig1\\_362910287](https://www.researchgate.net/figure/Figure-e-Growth-of-the-global-digital-health-market-from-mmmm-to-oooo-Reprinted-with_fig1_362910287)

## 1.2. Umjetna inteligencija i dentalna medicina

Umjetna inteligencija (AI) predstavlja koncept koji obuhvaća duboko učenje, strojno učenje i, djelomično, podatkovnu znanost. Od navedenih, ona predstavlja najšire polje te obuhvaća tehnike i metode korištene za stvaranje inteligentnih strojeva u svrhu obavljanja zadataka koji inače iziskuju ljudsku inteligenciju (npr. vizualna percepcija, obrada prirodnog jezika,...), a može se podijeliti i po granama, poput stručnih sustava, robotike i sl. Unutar AI, kako je spomenuto, nalazi se područje dubokog učenja (DL) koje koristi analogiju ljudskog mozga u konstrukciji neuronske mreže kako bi se postiglo učenje iz velike količine podataka. Na ovaj način, algoritmi DL automatski prepoznaju i izdvajaju elemente poput slika, zvukova i teksta iz sirovih podataka te ih koriste u predviđanju i donošenju odluka. Primjena DL obuhvaća prepoznavanje slike, prepoznavanje glasa te obradu prirodnog jezika. Strojno učenje, kao drugo polje unutar AI, usmjereno je na razvoj statističkih modela i algoritama potrebnih za učenje računala iz podataka, iako prethodno nisu programirani u

ovu svrhu, a tehnike mogu biti nadzirane (u ovom su slučaju podaci označeni), nenadzirane (podaci su neoznačeni) i djelomično nadzirani (kombiniranje označenih i neoznačenih podataka). Primjena strojnog učenja se može pronaći u sustavima preporuka, otkrivanja prijevara i prediktivnim modelima. Podatkovna znanost kao područje djelomičnog preklapanja s AI je interdisciplinarno područje u kojem se kombiniraju statističke i računalne metode sa specifičnim znanjem unutar određene domene u svrhu dobivanja uvida u to područje te generiranja znanja iz podataka. Aktivnosti koje obuhvaća su prikupljanje, pročišćavanje i prethodnu obradu podataka, istraživačku analizu podataka, statističko modeliranje te strojno učenje, a koristi se u zdravstvu, financijama, društvenim medijima i e-trgovini [4].

Primjenom računala diljem svijeta stvorene su tijekom desetljeća u medicini i zdravstvu goleme baze podataka s velikom količinom različitih podataka o pacijentima, dijagnozama, zdravstvenim kartonima i povezanim laboratorijskim nalazima. Tu su i radiološki i klinički nalazi i slike, terapijski zahvati, ishodi liječenja i još mnogo toga. Zahvaljujući takvim opsežnim bazama podataka osiguran je jedan od temeljnih preduvjeta za strojno učenje i razvoj umjetne inteligencije u svim područjima medicine [4].

Prva primjena umjetne inteligencije u dentalnoj medicini datira iz kraja 20. stoljeća kada su istraživači počeli koristiti osnovne algoritme za analizu dentalnih slika i predviđanje dentalnih bolesti. Početkom 21. stoljeća, napredak u tehnologijama obrade slike i računalne snage omogućio je sofisticiranije primjene AI u stomatologiji. Na primjer, istraživači su razvili sustave za automatsko prepoznavanje karijesa, analizu ortodontskih slika i planiranje dentalnih implantata. Studije su pokazale da ti sustavi mogu postići ili čak premašiti točnost ljudskih stručnjaka u dijagnosticiranju dentalnih problema [5].

Korištenje umjetne inteligencije u dentalnoj medicini može biti od velike koristi u nekoliko područja. Prvo, dijagnostika može biti znatno unaprijeđena pomoću AI sustava koji analiziraju radiografske slike za otkrivanje raznih dentalnih bolesti. Ovi sustavi mogu pomoći stomatolozima da preciznije dijagnosticiraju i tretiraju pacijente, smanjujući mogućnost ljudske pogreške. Drugo, u ortodonciji, AI može pomoći u planiranju terapije, izradi prilagođenih aparatića i praćenju napretka tretmana. Sustavi za prepoznavanje slika mogu analizirati 3D skeniranja zuba i čeljusti, pružajući detaljne informacije za izradu optimalnih ortodontskih rješenja.

Nadalje, umjetna inteligencija može biti korisna u područjima poput planiranja dentalnih implantata, gdje algoritmi mogu analizirati CT snimke i predložiti najbolju poziciju za implantate uzimajući u obzir anatomske strukture pacijenta. U endodonciji, AI može pomoći u identifikaciji složenih anatomskih struktura korijenskih kanala i predložiti optimalne strategije liječenja. Također, AI sustavi mogu pomoći u razvoju personaliziranih planova liječenja na temelju analize povijesti bolesti pacijenta i drugih relevantnih podataka, pružajući stomatolozima alate za donošenje informiranih odluka koje su prilagođene specifičnim potrebama svakog pacijenta [6].

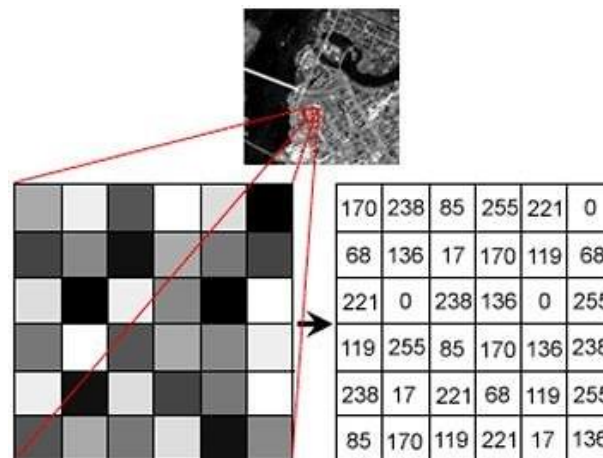
Osim kliničkih primjena, umjetna inteligencija može poboljšati administrativne i logističke aspekte dentalne medicine. Algoritmi za obradu prirodnog jezika mogu pomoći u automatizaciji administrativnih zadataka kao što su zakazivanje termina, vođenje evidencije i komunikacija s pacijentima. Ovo može smanjiti opterećenje stomatoloških timova i omogućiti im da se više posvete kliničkoj skrbi. Konačno, korištenje AI u obrazovanju stomatologa može unaprijediti edukacijske programe kroz simulacije i virtualne trenere koji pružaju povratne informacije u realnom vremenu, pomažući studentima da brže i učinkovitije steknu potrebne vještine [7].

Važno je istaknuti da je umjetna inteligencija u dentalnoj medicini još uvijek tehnologija u razvoju i da njezine prednosti uvelike ovise o pojedinom području primjene. Također je važno stvoriti odgovarajuću pravnu regulativu i nadzor kako bi se jamčila sigurnost, učinkovitost i etičnost u njezinu korištenju. I na kraju, važno je osigurati primjenu sustava umjetne inteligencije tako da bude stomatolozima i drugim stručnjacima potpora u radu, a nikako ne njihova zamjena [4].

## 2. Analiza slika

### 2.1. Digitalna slika

U današnje doba, sliku zasigurno smatramo jednim od glavnih izvora informacija. Sukladno tome, područje njenog istraživanja, posebice obrade i analize, neprestano razvija. Slika se može definirati kao dvodimenzionalna funkcija  $f(x, y)$ , gdje su  $x$  i  $y$  prostorne koordinate, a amplituda  $f$  na bilo kojem paru koordinata  $(x, y)$  naziva se intenzitetom ili sivom razinom slike u toj točki. Kada su  $x$ ,  $y$  i intenzitetne vrijednosti  $f$  konačne i diskretne veličine, sliku nazivamo digitalnom slikom. Digitalna slika sastoji se od konačnog broja elemenata, od kojih svaki ima određenu lokaciju i vrijednost. [8]. Navedeni primjer (Sl. 2.1) prikazuje nam jednu digitalnu sliku u kojoj razni stupnjevi boje, od bijele do crne, odgovaraju numeričkim vrijednostima intenziteta, od 0 do 255, u tabličnom (matričnom) prikazu. Upravo takav prikaz se često koristi kako bi lakše logički objašnjavali razvoj i provedbu određenih postupaka sa digitalnim slikama.



Sl. 2.1 Primjer digitalne slike.

Izvor: [https://www.researchgate.net/figure/Matrix-for-certain-area-of-a-grayscale-image-17\\_fig3\\_325569674](https://www.researchgate.net/figure/Matrix-for-certain-area-of-a-grayscale-image-17_fig3_325569674)

Poznajemo različite načine snimanja digitalnih slika od nama svakodnevno korištenih digitalnih kamera i skeniranja do medicinskih uređaja koji koriste rendgenske i ultrazvučne valove. Nakon spomenutog snimanja, slijedi proces koji uključuje pretvaranje analognih signala u digitalni oblik pomoću analogno-digitalnog pretvornika (ADP). Rezultat ovog

procesa su pikseli, najmanje jedinice digitalne slike, gdje svaki piksel ima određenu boju i intenzitet. Boja svakog piksela određena je količinom crvene, zelene i plave svjetlosti (RGB model), dok intenzitet određuje svjetlinu piksela [8].

Digitalne slike spremamo u raznim formatima od kojih su najčešće korišteni JPEG (Joint Photographic Experts Group) i PNG (Portable Network Graphics). Svaki format je specifičan po načinu kompresije, količini očuvanja kvalitete i veličini datoteke, a sama priroda zadatka ili budućeg rada nas usmjerava koji je oblik najpogodniji.

Shodno tome, područje obrade slike je zadnjih desetljeća iznimno istraživano kako bi kompresija i poboljšanje kvalitete slike bili što uspješniji. Također razvojem AI i sve naprednijih računala, interes za istraživanje i korištenje analize slika rapidno raste o čemu ćemo detaljnije pričati i u narednim poglavljima.

Obrađivanje slike i analiza slike su dvije blisko povezane, ali suštinski različite discipline u području računalne vizije. Obrađivanje slike odnosi se na manipulaciju slikovnih podataka s ciljem poboljšanja njihove kvalitete ili ekstrakcije korisnih informacija. Tipični zadaci u obradi slike uključuju filtriranje, poboljšanje kontrasta, smanjenje šuma i kompresiju slike [8].

Analiza slike, s druge strane, fokusira se na interpretaciju i razumijevanje slikovnih podataka. Cilj analize slike je automatsko prepoznavanje objekata, scena i uzoraka unutar slike. Dok obrađivanje slike transformira sirove podatke u poboljšane slike, analiza slike interpretira te poboljšane slike kako bi dobila korisne informacije [9].

Prvi pristupi analizi slike razvijeni su sredinom 20. stoljeća, kada su znanstvenici počeli koristiti računalne algoritme za prepoznavanje osnovnih oblika i uzoraka unutar slika. Ovi rani pristupi uključivali su detekciju rubova, segmentaciju slike i prepoznavanje oblika.

Razvojem računalne snage i naprednih algoritama, pristupi analizi slike značajno su se razvili. Današnji napredni pristupi uključuju: konvolucijske neuronske mreže CNN, strojno učenje i supe rezolucija.

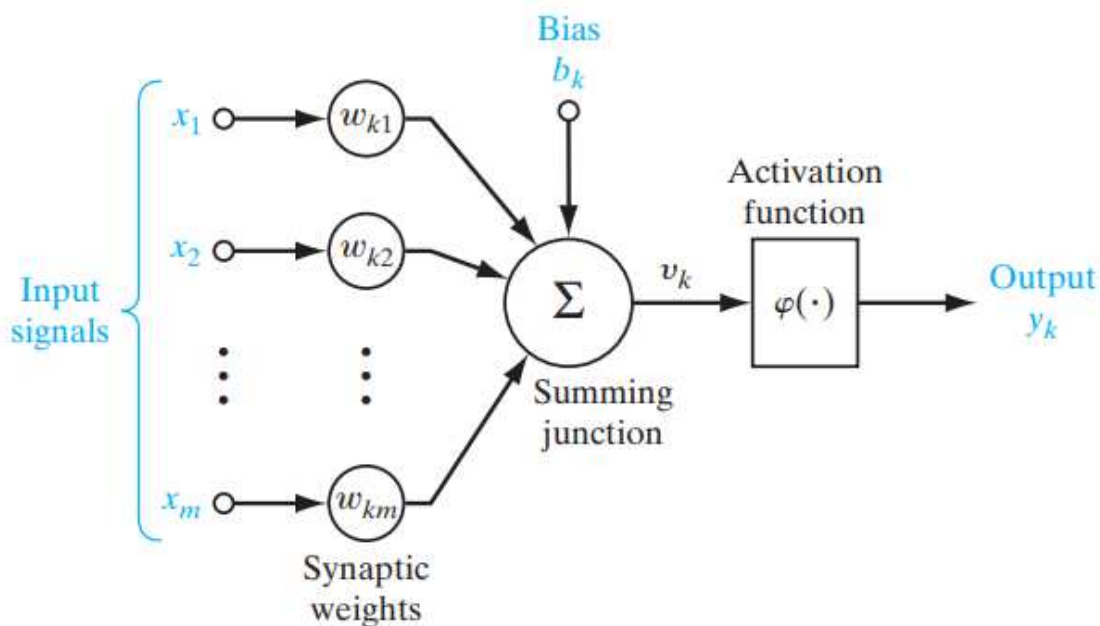
Analiza slike je kompleksno i dinamično područje koje se neprestano razvija. Od svojih početaka s jednostavnim metodama detekcije rubova i segmentacije, analiza slike je evoluirala u sofisticiranu znanstvenu disciplinu koristeći napredne tehnike dubokog učenja i umjetne inteligencije. Ovi napredni pristupi omogućuju automatsku interpretaciju i

razumijevanje slikovnih podataka, čineći analizu slike ključnom komponentom modernih tehnologija u raznim industrijama.

## 2.2. Neuronske mreže

Cijelo područje umjetne inteligencije je nastalo ljudskom željom da programi nauče razmišljati što sličnije ljudima. Sve većim brojem inovacija došlo je i do razvoja računalnog vida (Computer Vision CV) kao područja AI-a u kojem želimo razvijati programe koji vide kao ljudski mozak. Iako je vid za ljude jednostavna urođena radnja, digitalna slika zahtjeva složenije obrade i analize kako bi se objekti i informacije o njima ispravno prepoznali. Shodno tome, nastale su neuronske mreže inspirirane ljudskim neuronima tj. mozgom koji je zaslužan za našu sposobnost razmišljanja, učenja i donošenja odluka o stvarima koje promatramo oko nas.

Neuron je temeljna jedinica za obradu informacija unutar neuronske mreže. Njegov model, prikazan slikom (Sl. 2.2 Neuron), služi kao osnova za dizajn raznih neuronskih mreža. Glavni elementi ovog modela su sinapse, zbrajalo i aktivacijska funkcija. Sinapse su zapravo signali predstavljeni numeričkim vrijednostima  $x$  koje imaju svoju specifičnu težinu. Signal na ulazu svake sinapse množi se s odgovarajućom težinom  $w$ , pri čemu indeksi težine određuju koja je jakost pojedine sinapse. Za razliku od bioloških neurona, sinaptičke težine umjetnih neurona mogu biti pozitivne ili negativne. Zbrajalo sumira sve ulazne signale koji su ponderirani sinaptičkim težinama, što predstavlja linearni kombinator. Na kraju aktivacijska funkcija ograničava amplitudu izlaza neurona na određeni raspon, obično između  $[0,1]$  ili  $[-1,1]$ . Ova funkcija komprimira izlazni signal unutar tih granica. Model neurona također uključuje vanjski pomak *bias*  $b$  koji može povećati ili smanjiti ulaz aktivacijske funkcije, ovisno o tome je li pozitivan ili negativan [10].



Sl. 2.2 Neuron

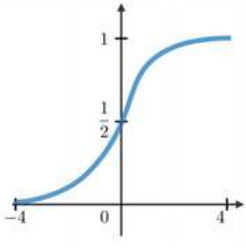
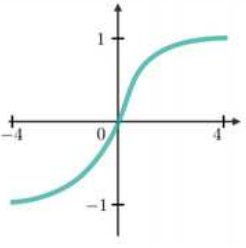
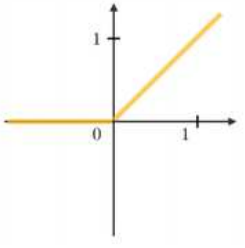
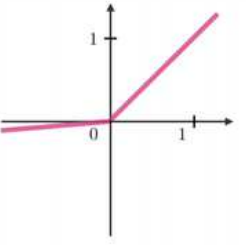
Izvor: [https://www.researchgate.net/figure/A-visualization-of-the-components-of-a-neural-network-Source-35\\_fig1\\_339228419](https://www.researchgate.net/figure/A-visualization-of-the-components-of-a-neural-network-Source-35_fig1_339228419)

Matematički se izlaz k-tog neurona o može opisati jednađbom koja uključuje m ulaznih signala x, sinaptičkih težina w, pomak b i aktivacijsku funkciju  $\varphi$  izrazom (2-1)

$$o_k = \varphi \left( \sum_{i=1}^m w_{ki} x_i + b_k \right)$$

(2-1)

Za aktivacijsku funkciju se koriste različite vrste, formule i izgled neke od najčešćih su prikazane sljedećom slikom (Sl. 2.3 Aktivacijske funkcije).

Sigmoid	Tanh	ReLU	Leaky ReLU
$g(z) = \frac{1}{1 + e^{-z}}$	$g(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}}$	$g(z) = \max(0, z)$	$g(z) = \max(\epsilon z, z)$ with $\epsilon \ll 1$
			

Sl. 2.3 Aktivacijske funkcije

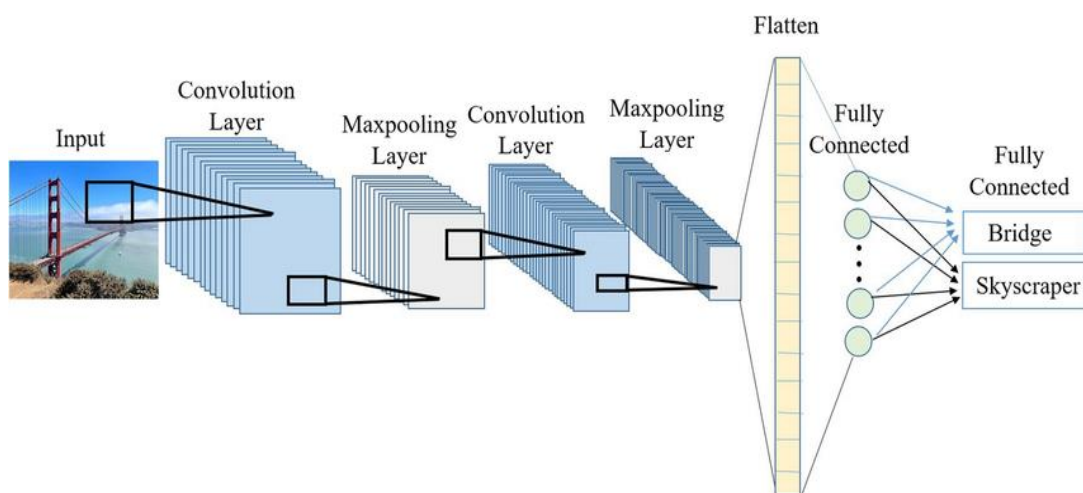
Izvor: [https://www.researchgate.net/figure/a-neuron-and-a-MLP\\_fig3\\_344954689](https://www.researchgate.net/figure/a-neuron-and-a-MLP_fig3_344954689)

Neuronske mreže mogu biti jednoslojne, višeslojne i duboke. Jednoslojne mreže sastoje se od samo jednog sloja neurona između ulaza i izlaza. Višeslojne mreže (MLP - Multi-Layer Perceptrons) imaju jedan ili više skrivenih slojeva između ulaznog i izlaznog sloja, što omogućuje složeniju obradu podataka. Duboke neuronske mreže (DNN - Deep Neural Networks) sadrže mnoge skrivene slojeve, što im omogućuje obradu vrlo složenih i višedimenzionalnih podataka. Ove mreže su ključne za napredak u prepoznavanju slika, govoru i obradi prirodnog jezika [11].

### 2.2.1. Konvolucijske neuronske mreže

Konvolucijska neuronska mreža (Convolutional Neural Network, CNN) je najčešće korištena umjetna neuronska mreža izgrađena od neurona koji se samooptimiziraju tijekom učenja. I dalje svaki neuron prima ulaz i izvršava operaciju te pomoću izračunatih težina daje izlaz, a u posljednjem sloju sadržava funkcije gubitka povezane s klasama. CNN koristimo za procesiranje nestrukturiranih podataka, a jedna od glavnih prednost je mogućnost prepoznavanja obrazaca unutar slika [12]. Osnovna struktura konvolucijske neuronske mreže sastoji se od nekoliko vrsta slojeva, a tu su ulazni sloj iza kojeg slijedi nekoliko konvolucijskih, aktivacijskih (nelinearnih) slojeva i slojeva sažimanja te pred kraj dolaze potpuno povezani slojevi kao što je prikazano slikom (Sl. 2.4).

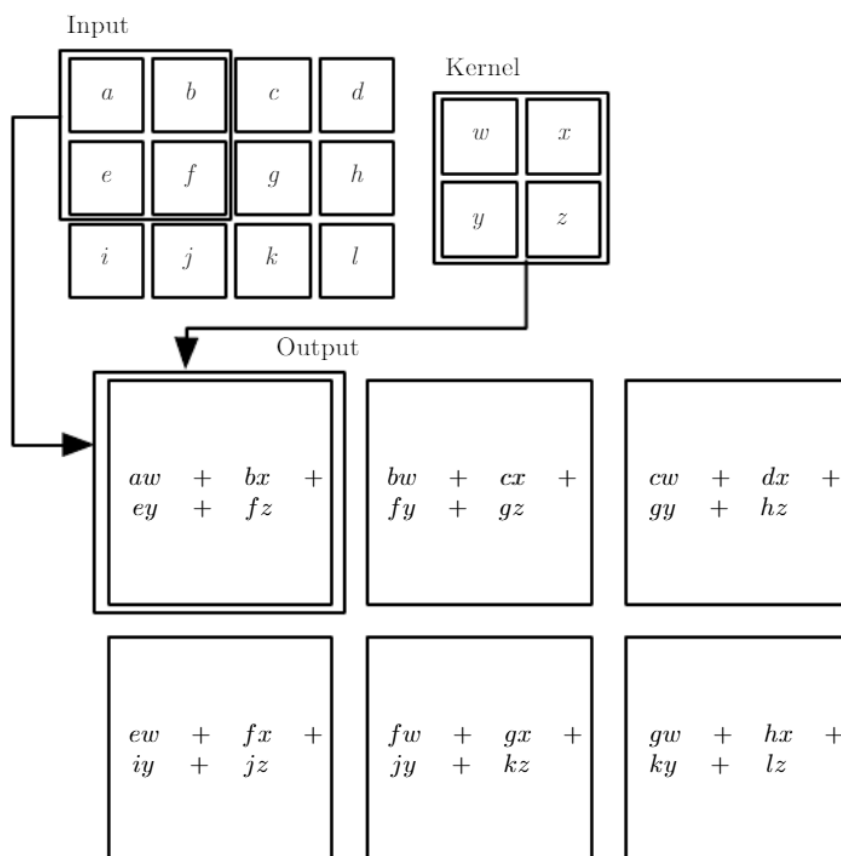




Sl. 2.4 CNN struktura

Izvor: [https://www.researchgate.net/figure/A-CNN-sequence-to-classify-an-image-to-bridge-or-skyscraper\\_fig1\\_359219555](https://www.researchgate.net/figure/A-CNN-sequence-to-classify-an-image-to-bridge-or-skyscraper_fig1_359219555)

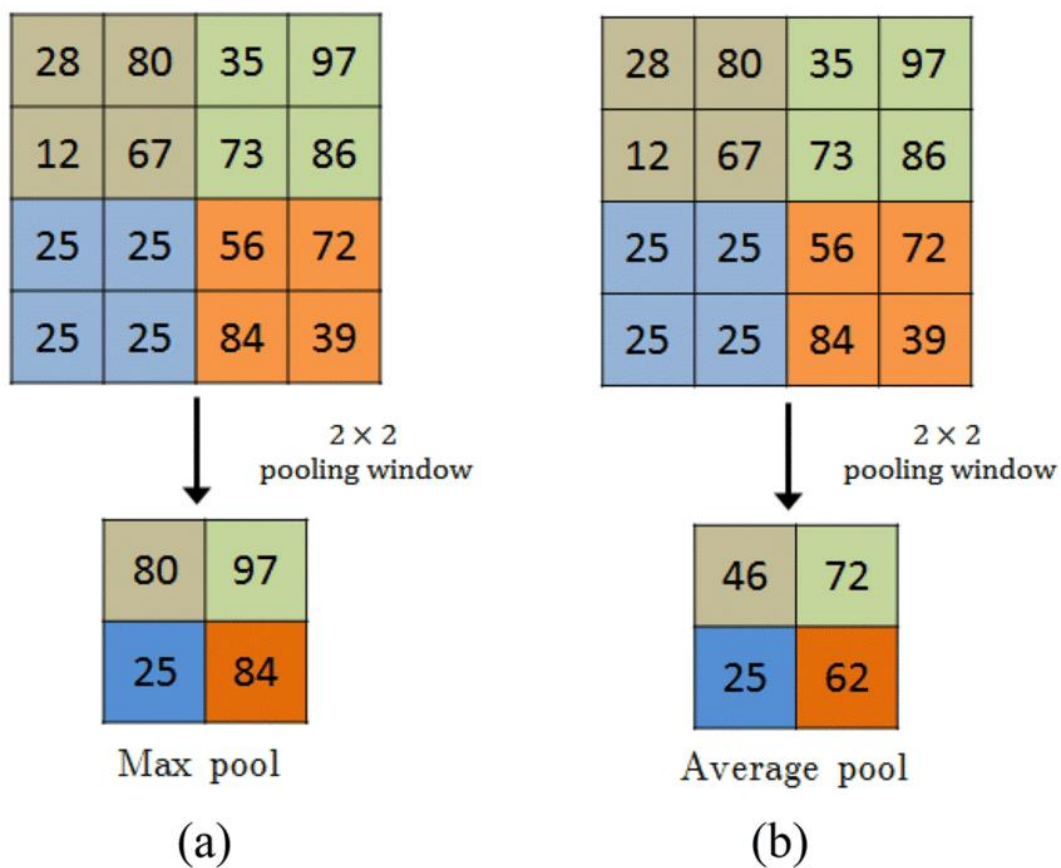
Konvolucijski sloj je ključan dio arhitekture CNN-a koji služi za izdvajanje značajki. Konvolucija predstavlja specifičnu vrstu linearne operacije u kojoj malo polje brojeva poznat kao jezgra (kernel) se primjenjuje preko ulaznih podataka, koji se sastoje od skupa brojeva poznatog kao tenzor. Na svakoj poziciji unutar tenzora, izračunava se umnožak elementa jezgre i odgovarajućeg elementa ulaznog tenzora te se ti rezultati zbrajaju kako bi se dobila izlazna vrijednost na odgovarajućoj poziciji izlaznog tenzora kao na slici (Sl. 2.5) [13].



Sl. 2.5 Konvolucija

Izvor: [https://www.researchgate.net/figure/An-example-of-2D-convolution-60-The-output-data-is-a-result-of-the-weighted-sum-of\\_fig2\\_349379059](https://www.researchgate.net/figure/An-example-of-2D-convolution-60-The-output-data-is-a-result-of-the-weighted-sum-of_fig2_349379059)

Slojevi sažimanja (pooling) smanjuju dimenzionalnost mape značajki, čime se postiže invarijantnost na translacije i manje promjene te se smanjuje broj parametara za učenje. Tu se koriste hiperparametri poput veličine filtra, koraka i popunjavanja. Najčešći oblik grupiranja je maksimalno grupiranje (max pooling), koje izvlači maksimalne vrijednosti iz segmenta ulaznih značajnih mapa i odbacuje ostale vrijednosti. Uobičajeno se koristi maksimalno grupiranje s filtrom veličine  $2 \times 2$  i korakom 2, što smanjuje dimenzionalnost značajnih mapa u ravnini za polovicu. Globalno prosječno grupiranje predstavlja drugi važan tip, gdje se značajne mape smanjuju na jedno polje veličine  $1 \times 1$  uzimanjem prosjeka svih elemenata u značajnim mapama, zadržavajući dubinu. Ova operacija se obično primjenjuje prije potpuno povezanih slojeva radi smanjenja broja parametara i omogućavanja prilagodbe varijabilnoj veličini ulaza [13]. Oba primjera sažimanja prikazana su slikom (Sl. 2.6).



Sl. 2.6 Sažimanje

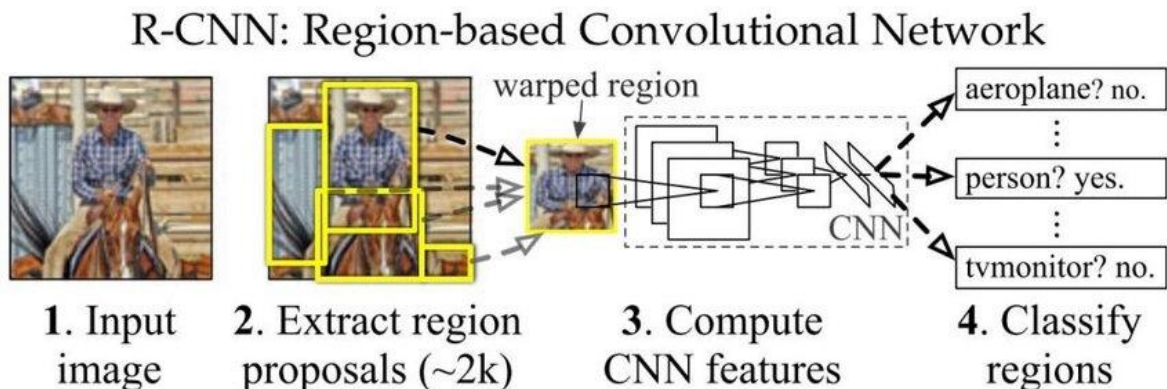
Izvor: [https://www.researchgate.net/figure/Visual-representation-of-pooling-operations-a-max-pooling-b-average-pooling\\_fig4\\_340255272](https://www.researchgate.net/figure/Visual-representation-of-pooling-operations-a-max-pooling-b-average-pooling_fig4_340255272)

Izlazi značajnih mapa iz posljednjeg konvolucijskog ili grupiranja sloja obično se izravnavaju u jednodimenzionalni niz brojeva ili vektor te se dalje povezuju s jednim ili više potpuno povezanih slojeva, poznatih i kao gusti slojevi. Nakon ekstrakcije značajki prethodnim slojevima, podaci se prenose preko potpuno povezanih slojeva do konačnih izlaza mreže, poput vjerojatnosti za svaku klasu u zadacima klasifikacije. Konačni potpuno povezani sloj obično ima isti broj izlaznih čvorova kao i broj klasa. Svaki potpuno povezani sloj zatim prolazi kroz nelinearnu aktivacijsku funkciju, kao što je ReLU, kako je opisano ranije [13].

## 2.2.2. Modeli neuronskih mreža

Kroz godine su se razvijali različiti algoritmi s ciljem poboljšanja detekcije objekata. Modele temeljene na CNN-u možemo podijeliti u dvije osnovne kategorije: pristup u dva koraka i pristup u jednom koraku [14].

R-CNN (Regions with CNN features) je prvi algoritam, prikazan (Sl. 2.7), koji opisujemo za detekciju objekata u dva koraka. Prvo se koriste metode za izdvajanje prijedloga regija (njih oko dvije tisuće) različitih veličina, nakon čega se na tim regijama provodi regresijska predikcija za određivanje objekata. Postigao je 58.5% srednje prosječne preciznosti (mAP) na VOC2007 skupu podataka, međutim R-CNN ima izazove poput dugotrajne pretrage regija i potrebe za odvojenim treniranjem više modela, što negativno utječe na brzinu. Također, R-CNN prilagođava veličinu ulazne značajne mape kako bi riješio probleme s različitim veličinama prijedloga regija, što može rezultirati gubitkom informacija iz originalne slike [15].

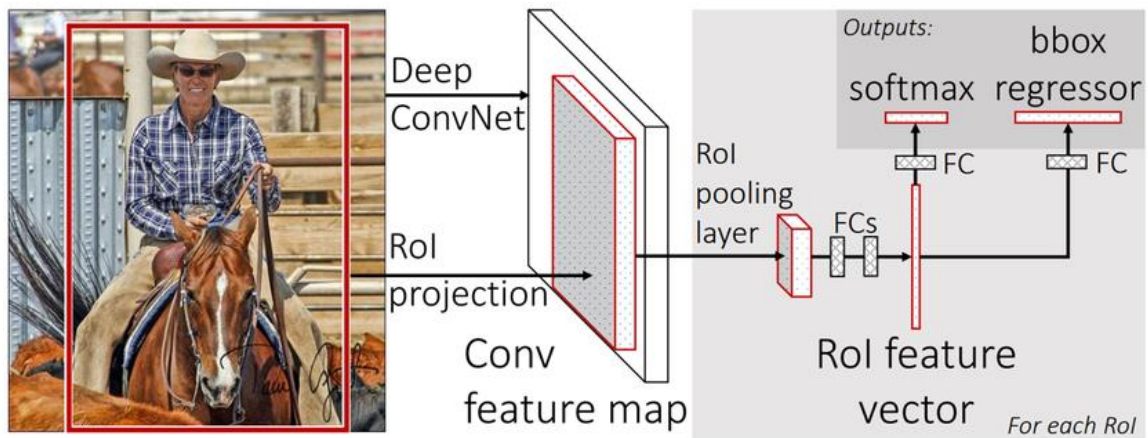


Sl. 2.7 R-CNN

Izvor: [https://www.researchgate.net/figure/Three-major-modules-of-a-Faster-R-CNN-network-Source-adapted-from-27\\_fig3\\_375778602](https://www.researchgate.net/figure/Three-major-modules-of-a-Faster-R-CNN-network-Source-adapted-from-27_fig3_375778602)

Fast R-CNN, prikazana slikom (Sl. 2.8), poboljšana je struktura mreže R-CNN-a u kombinaciji s SPPnet-om kako bi se predložio sloj bazena regija od interesa (RoI pooling). Ovaj pristup rješava problem ponovnog treniranja više predloženih regija u R-CNN-u i nerazmjernih veličina potpuno povezanih slojeva dobivenih SVD metodom. Testiranje na VOC2007 skupu podataka postiglo je mAP od 70%, a vrijeme obuke je približno 9 puta brže od R-CNN-a. Ipak, Fast R-CNN-a još uvijek ima nekoliko nedostataka. Prvo, korištenje selektivnih pretraživanja za generiranje velikog broja predloženih regija produžuje vrijeme

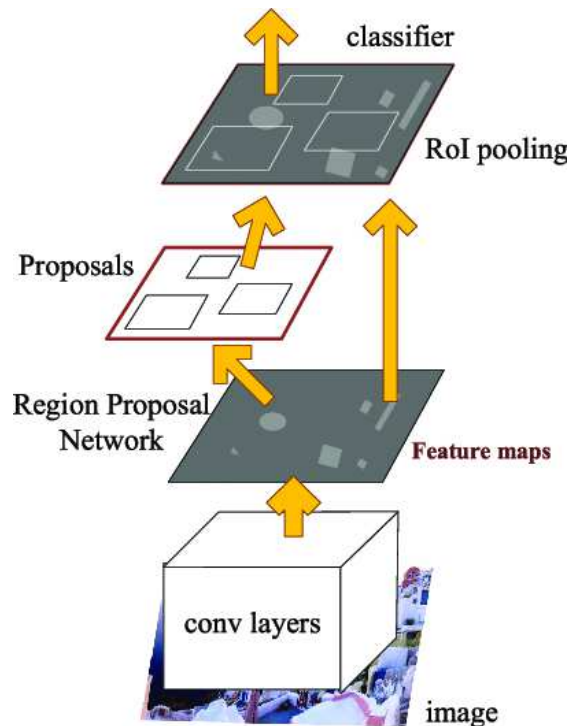
obuke i predviđanja. Drugo, korištenje više potpuno povezanih slojeva na kraju mreže izračunava se odvojeno, bez dijeljenja težina, što povećava broj parametara [15].



Sl. 2.8 Fast R-CNN

Izvor: [https://www.researchgate.net/figure/Fast-RCNN-object-detection-architecture-20\\_fig3\\_357401502](https://www.researchgate.net/figure/Fast-RCNN-object-detection-architecture-20_fig3_357401502)

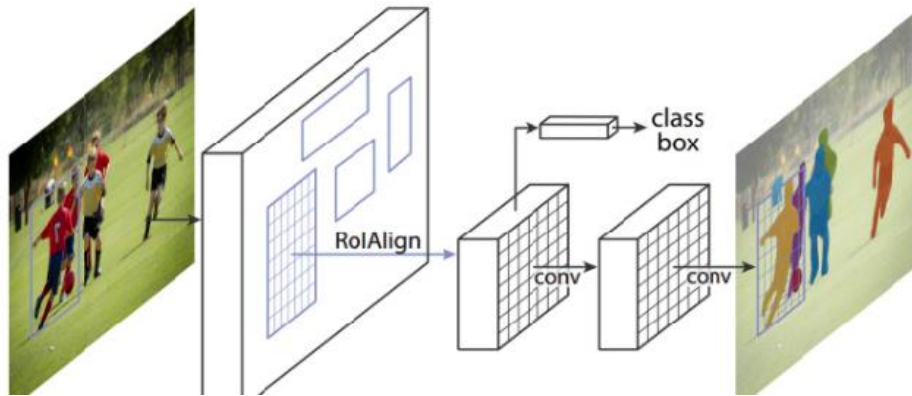
Već 2017. godine predložen je Faster R-CNN kako bi se riješio problem uskog grla u izboru predloženih regija tijekom dvofaznog algoritma. Novi mrežni sustav za predlaganje regija (RPN) zamjenjuje klizni prozor u Fast R-CNN-u, generirajući prijedloge regija i poboljšavajući brzinu detekcije kao na slici (Sl. 2.9). Za razliku od prethodnika, Faster R-CNN je postigao 73.2% mAP na VOC2007 uz brzinu detekcije od 5 FPS. Međutim, mreža je manje precizna za male objekte zbog velikog uzorkovanja i velikog broja parametara te ne zadovoljava zahtjeve za detekciju u stvarnom vremenu.



Sl. 2.9 Faster R-CNN

Izvor: [https://www.researchgate.net/figure/Algorithm-framework-of-faster-R-CNN\\_fig2\\_347265949](https://www.researchgate.net/figure/Algorithm-framework-of-faster-R-CNN_fig2_347265949)

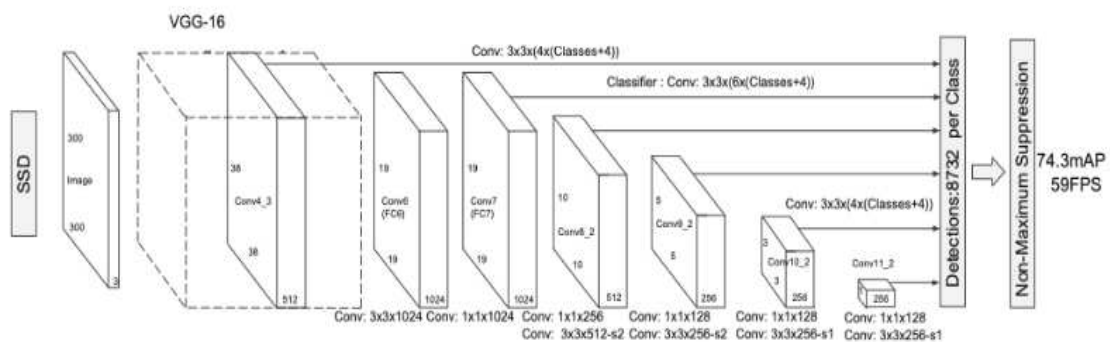
Za Mask R-CNN idu korak dalje i lociraju točne piksele svake instance objekta, umjesto samo ograničavajućih okvira. Budući da ovaj model neovisno maskira svaku instancu objekta, nazvali su ga Mask R-CNN, a koristi istu mrežu za predložak regija (RPN) kao Faster R-CNN što je prikazano slikom (Sl. 2.10). Autori su primijenili sloj RoIAlign na prijedloge regija umjesto RoI pooling sloja kako bi poravnali izvađene značajke s ulaznom lokacijom objekta. Poravnati RoI-ji se zatim dovode u posljednji dio kako bi generirali tri izlaza: oznaku klase, ofset okvira i binarnu masku objekta. Za maskiranje se koristi mala potpuno konvolucijska neuronska mreža na svakom RoI-u [14].



Sl. 2.10 Mask R-CNN

Izvor: [https://www.researchgate.net/figure/Mask-R-CNN-structure-He-et-al-2017\\_fig1\\_344980963](https://www.researchgate.net/figure/Mask-R-CNN-structure-He-et-al-2017_fig1_344980963)

SSD (Single Shot Multibox Detection) model, prikazan (Sl. 2.11), koristi cijelu sliku kao ulaz i prolazi kroz više konvolucijskih slojeva s različitim veličinama filtara ( $10 \times 10$ ,  $5 \times 5$  i  $3 \times 3$ ). Mape značajki iz tih slojeva koriste se za predviđanje ograničavajućih okvira. Dodatni konvolucijski slojevi s filtrom  $3 \times 3$ , nazvani dodatni slojevi značajki, generiraju konačne skupove okvira. SSD koristi sidrišne okvire s parametrima za centar, širinu i visinu. Model predviđa vjerojatnosti za svaku klasu objekata te omogućuje detekciju objekata različite veličine uz dobru ravnotežu brzine i točnosti [14].



Sl. 2.11 SSD

Izvor: [https://www.researchgate.net/figure/SSD-Network-model-Also-Figure-4-shows-the-determination-flow-by-the-two-hierarchical-AIs\\_fig2\\_336304666](https://www.researchgate.net/figure/SSD-Network-model-Also-Figure-4-shows-the-determination-flow-by-the-two-hierarchical-AIs_fig2_336304666)

Još jedan model u jednom koraku je YOLO algoritam koji je detaljno opisan u poglavlju 3.2.

Odabir odgovarajućeg modela ovisi o prirodi samog zadatka i raspoloživim resursima . Najčešće se gleda u kojoj mjeri model mora biti točan odnosno brz. Za slučaje kojima je najbitnija neka mjera uspješnosti kao na primjer točnost, Fast R-CNN i Faster R-CNN pokazuju se kao bolji izbor. Za razliku od njih modeli u jednom koraku su puno brži te upravo zbog brzine i jednostavnosti smo se u ovom radu odlučili za YOLO algoritam.

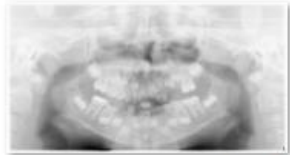


## 3. Implementacija

Zadatak ovog znanstvenog istraživanja je iskoristiti najprikladniji pristup analizi slika, nastavljajući se na sve prethodno proučeno ovo radu, kako bismo implementirali odgovarajući model za detekciju i klasifikaciju pojedinih zubi maloljetnika.

### 3.1. Korišteni podaci

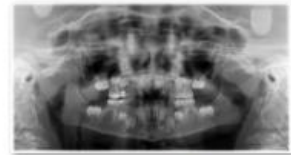
Kako sve kreće od podataka, tako je i ovaj proces započeo prikupljanjem podataka od strane Stomatološkog fakulteta koji nam je ustupio 600 panoramskih rendgenskih slika maloljetnika. Rendgenske slike zubi (dentalni rendgeni) nastaju prolaskom vrste elektromagnetskog zračenja, zvano rendgenske zrake, kroz različite strukture u ustima poput zubi i tkiva koje apsorbiraju zrake u različitoj mjeri. Na digitalnim slikama područja koja su apsorbirala više rendgenskih zraka pojavljuju se svjetlije, dok područja kroz koja je prošlo više izgledaju tamnija. Na primjeru nekoliko skupljenih dentalnih rendgena iz našeg skupa podataka (Sl. 3.1 Uzorak prikupljenih slika) možemo uočiti kako su područja zubi i kosti svjetliji, a šupljine i tkiva tamniji.



ImageFile\_92a0854c-bf  
a9-4389-a1cb-090b0ad  
61f23\_000014015.png



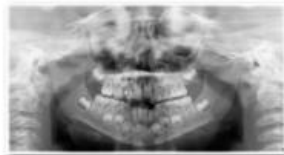
ImageFile\_93ba0756-e1  
6e-454d-977f-697dfc3c  
85a5\_000007082.png



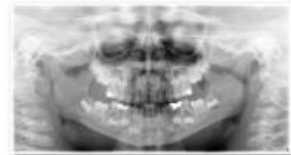
ImageFile\_95ebd68a-a8  
f1-4f8a-b5e8-7c947c59  
f8cd\_000036775.png



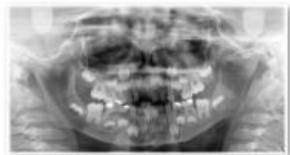
ImageFile\_55477a05-46  
cf-4fac-8b80-8df45096  
2e7f\_000205092.png



ImageFile\_67795f02-58  
05-4f7a-896b-7c20926  
6d991\_000206747.png



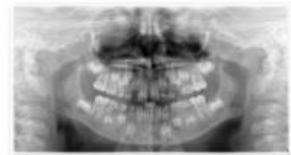
ImageFile\_86938e79-c5  
e2-4a7c-a19a-46d745a  
3d8de\_000237711.png



ImageFile\_c68a7924-93  
63-40ee-838b-f601b34  
dd3bb\_000008458.png



ImageFile\_d1fb08c9-e8  
9c-4363-abb7-7707ec3  
6c0dd\_000174254.png



ImageFile\_d70112fc-d2  
0e-47a0-a377-603de55  
bf74f\_000178277.png

Sl. 3.1 Uzorak prikupljenih slika

Također, uočavamo kako se pojavljuju različite vrste zubi, i trajni i mliječni, a njihovi položaji razlikuju u pojedinim slikama što čini upravo ovaj skup dječjih zubi posebno zanimljivim i izazovnim za testiranje.

Kako bismo znali izvući informacije o značajkama zubi, potrebno je bilo od strane dentalnih stručnjaka da svaku sliku anotiraju. Slijedom toga, dobili smo datoteku s anotacijama u XML (EXtensible Markup Language) formatu pisano jezikom za označavanje podataka kojem je ideja biti jednostavno čitljiv i ljudima i računalnim programima. U takvim formatima sadržaj se treba uokviriti odgovarajućim oznakama koje ga opisuju i imaju poznato, ili lako shvatljivo značenje [16]. Na prikazanom dijelu dokumenta s anotacijama (Sl. 3.2) vidimo da su naveli ime slike, odgovarajuće dimenzije iste te za svaki zub njegovu lokaciju i šest atributa.

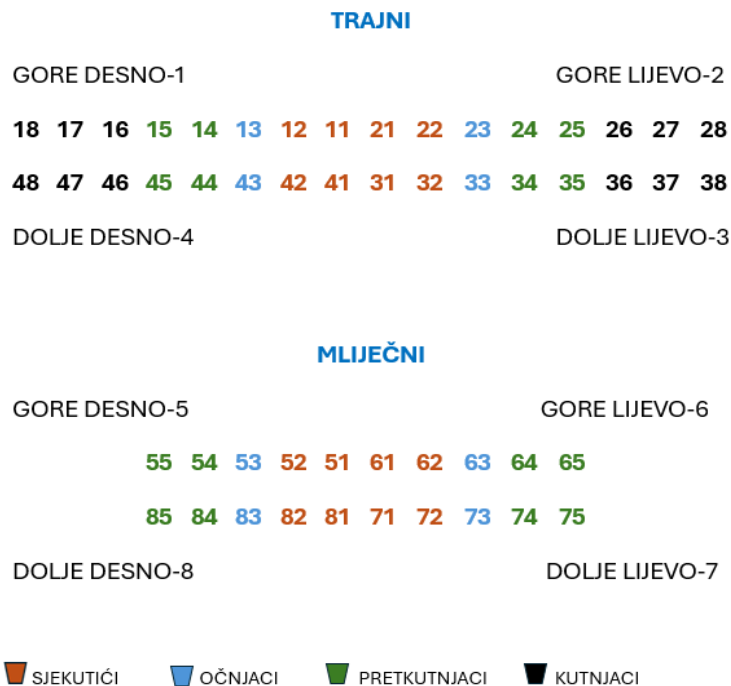
```

<image id="599" name="ImageFile_ffa8b7fd-c021-41ed-b26b-4007d3af7a4c_000101204.png" width="2775" height="1480">
  <box label="tooth" occluded="0" source="manual" xtl="1226.71" ytl="554.04" xbr="1310.08" ybr="835.36">
    <attribute name="quad">1</attribute>
    <attribute name="index">3</attribute>
    <attribute name="nezavršen rast korijena">false</attribute>
    <attribute name="završen rast korijena">true</attribute>
    <attribute name="iznikli zub">false</attribute>
    <attribute name="ne iznikli zub">true</attribute>
  </box>
  <box label="tooth" occluded="0" source="manual" xtl="1167.17" ytl="780.33" xbr="1293.41" ybr="923.49">
    <attribute name="quad">5</attribute>
    <attribute name="index">3</attribute>
    <attribute name="nezavršen rast korijena">false</attribute>
    <attribute name="završen rast korijena">true</attribute>
    <attribute name="iznikli zub">true</attribute>
    <attribute name="ne iznikli zub">false</attribute>
  </box>
  <box label="tooth" occluded="0" source="manual" xtl="1311.39" ytl="654.55" xbr="1416.20" ybr="784.55">
    <attribute name="quad">1</attribute>
    <attribute name="index">2</attribute>
    <attribute name="nezavršen rast korijena">false</attribute>
    <attribute name="završen rast korijena">true</attribute>
    <attribute name="iznikli zub">true</attribute>
    <attribute name="ne iznikli zub">false</attribute>
  </box>

```

### Sl. 3.2 Dio datoteke s anotacijama

Nama je za ovaj rad bilo potrebno koristiti atribute „index“ i „quad“ odnosno kvadranta i indeksa jer FDI sustav numeracije zubi (Federation Dentaire Internationale) koristi dvoznamenkasti brojčani sustav za označavanje zuba. Zubni lukovi su podijeljeni na četiri kvadranta koji su obilježeni brojevima 1 - 4 za stalnu denticiju, a četiri kvadranta mliječne denticije su obilježena brojevima 5 - 8. Drugi broj u sustavu označava položaj zuba u kvadrantu, polazeći od medijalne linije [17, 19]. Vrstu i količinu tako označenih trajnih i mliječnih vidimo na slici ispod (Sl. 3.3).



Sl. 3.3 Oznake zubi

S obzirom da se radi o stvarnim podacima, a ne unaprijed pripremljenom ili generiranom skupu za ovu vrstu analize, bilo je višestrukog pregledavanja sadržava li svaka slika anotacije, jesu li neke vrijednosti oznaka prazne ili neispravnih nepostojećih vrijednosti te su uspješno korigirani i pripremljeni za daljnje postupke.

Kako bismo informacije iz anotacija pravilno oblikovali za neuronsku mrežu, o kojoj detaljno pričamo u sljedećem poglavlju, bilo ih je potrebno, parsirajući XML, spremići u YOLO (You Only Look Once) format. Ovakav format anotacija koristi se za označavanje objekata unutar slike za zadatke poput detekcije objekata. U njemu svaka slika ima odgovarajuću tekstualnu datoteku koja sadrži anotacije. Nadalje, svaka linija u datoteci predstavlja jedan okvir oko objekta i sadrži klasu objekta, normalizirane koordinate središta objekta te širinu i visinu objekta [18] u obliku izraza (3-1).

$$[label, x\_center, y\_center, width, height]$$

(3-1)

Primjer linije anotacije u YOLO formatu:

*12 0.4386396 0.662645273 0.035729729 0.1388581*

Prvi broj dobiven je uzimajući redni broj klase zuba (krećući od 0), u ovom primjeru označen s 12, koja u skupu mogućih klasa odgovara onoj naziva 25 „gornji lijevi (2) pretkutnjak (5)“. Ostale vrijednosti predstavljaju normalizirane koordinate središta i dimenzije objekta izračunate na temelju atributa „x<sub>tl</sub>“, „y<sub>tl</sub>“, „x<sub>br</sub>“ i „y<sub>br</sub>“. Ti atributi su koordinate koje definiraju položaj i veličinu bounding box-a (ograničavajućeg okvira) gdje x<sub>tl</sub> (x top left) i y<sub>tl</sub> (y top left) predstavljaju x i y koordinate gornjeg lijevog vrha, a x<sub>br</sub> (x-bottom-right) i y<sub>br</sub> (y-bottom-right) x i y koordinate donjeg desnog vrha. Koristeći njih te visinu i širinu cijele slike koristeći izraz (3-2) dobivamo sve tražene vrijednosti.

$$\begin{aligned}x\_center &= ((float(xtl) + float(xbr)))/ (2 * float(width)) \\y\_center &= ((float(ytl) + float(ybr)))/ (2 * float(height)) \\box\_width &= ((float(xbr) - float(xtl)))/ float(width) \\box\_height &= ((float(ybr) - float(ytl)))/ float(height)\end{aligned}$$

(3-2)

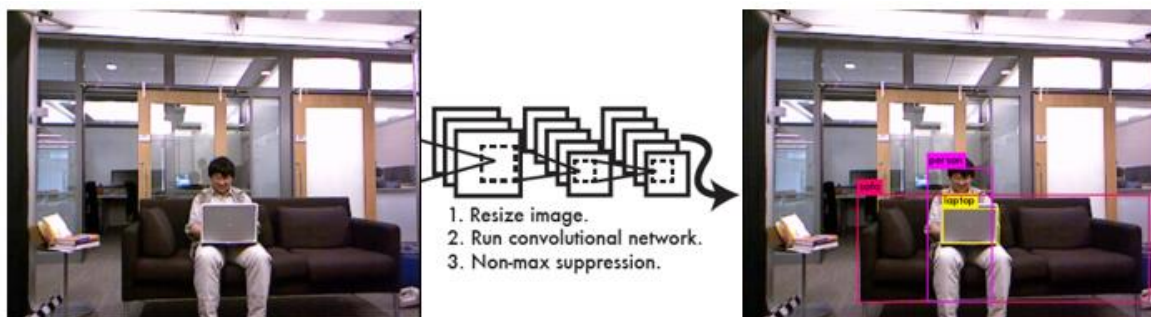
Kad smo imali sve dobro formatirane anotacije i odgovarajuće slike mogli smo krenuti na sljedeću fazu implementaciju modela koji će nad tim skupom podataka učiti i provesti testiranje svoje točnosti.

## 3.2. YOLO model

YOLO (You Only Look Once) donosi revolucionarni pristup detekciji objekata tretirajući cijelu sliku kao mrežu ćelija, od kojih svaka predviđa granice okvira i klase objekata u jednoj iteraciji. Ovaj jedinstveni model simultano predviđa granice okvira i klase objekata, omogućujući detekciju u stvarnom vremenu. Za razliku od tradicionalnih metoda koje koriste regijske predloške i višestupanjske modele, YOLO pristup omogućava bržu i učinkovitiju detekciju [20].

Ovaj algoritam predstavljen je 2015. radom "You Only Look Once: Unified, Real-Time Object Detection" autora Josepha Redmona, Santosha Divvle, Rossa Girshicka i Alija Farhadija uvodeći ideju regresije u detekciju objekata [20]. Pojednostavljenu ideju kako radi prikazuje primjer (Sl. 3.4)., a u nastavku navodimo kako su ga autori opisali.

YOLO koristi jednu konvolucijsku mrežu koja istovremeno predviđa više graničnih okvira i vjerojatnosti klasa za te okvire. Trenira se na cjelovitim slikama i izravno optimizira performanse detekcije.



Sl. 3.4 Pojednostavljeno YOLO rad

Izvor: [https://www.researchgate.net/figure/The-YOLO-detection-system-1-resizes-the-input-image-to-448-448-2-runs-a-single\\_fig3\\_346813584](https://www.researchgate.net/figure/The-YOLO-detection-system-1-resizes-the-input-image-to-448-448-2-runs-a-single_fig3_346813584)

Ovaj integrirani model ima brojne prednosti u odnosu na tradicionalne metode detekcije objekata. Prvo, YOLO je izuzetno brz jer detekciju postavlja kao regresijski problem, izbjegavajući potrebu za složenim procesima. Za testiranje, jednostavno se pokreće neuronska mreža kako bi se predvidjele detekcije. YOLO postiže dvostruko veću prosječnu preciznost u usporedbi s drugim sustavima detekcije. Dodatno, YOLO razmatra sliku globalno prilikom predviđanja, za razliku od kliznih prozora i prijedloga regija, što omogućuje implicitno kodiranje kontekstualnih informacija o klasama i njihovom izgledu. Zahvaljujući svojoj dobroj generalizaciji, YOLO je manje sklon kvarovima pri primjeni na nove domene ili neočekivane ulaze.

Unatoč svojoj brzini u prepoznavanju objekata na slikama, YOLO još uvijek zaostaje za najnovijim sustavima detekcije u pogledu preciznosti lokalizacije pojedinih objekata.

Nastavljajući se na prethodno, trebamo objasniti kako to YOLO predstavlja paradigmatički pomak u detekciji objekata konsolidacijom različitih komponenti ovog zadatka u jednu neuronsku mrežu. Tradicionalni pristupi detekciji objekata obično uključuju nekoliko odvojenih faza kao što su ekstrakcija značajki, predikcija graničnih okvira (bounding box) i klasifikacija. Suprotno tome, YOLO koristi značajke cijele slike za simultano predviđanje granica i klasa objekata, čime omogućuje globalno rasuđivanje o slici i njezinim objektima.

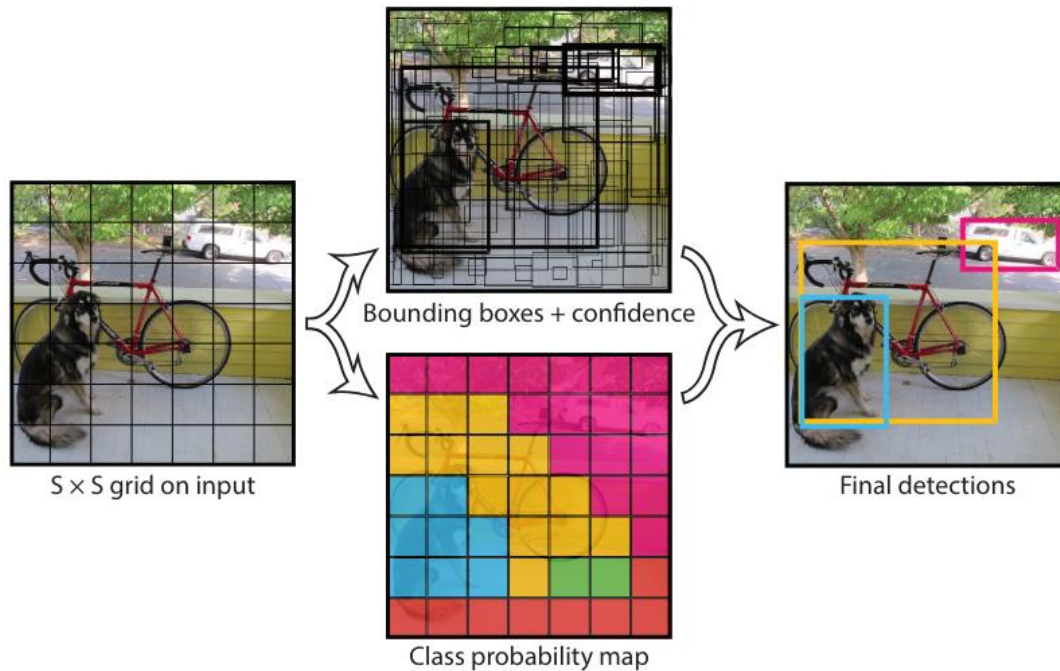
Sustav dijeli ulaznu sliku u mrežu dimenzija  $S \times S$ , pri čemu je svaka ćelija mreže odgovorna za detekciju objekata čije se središte nalazi unutar te ćelije. Takav način podjele omogućuje učinkovitije procesiranje i lokalizaciju objekata. Svaka ćelija mreže predviđa  $B$  graničnih okvira i pouzdanost, pri čemu pouzdanost rezultata izrazom (3-3) odražava koliko je model uvjeren da okvir sadrži neki objekt te kolika je točnost okvira.

$$Pr(Object) * IoU_{pred}^{truth} \quad (3-3)$$

Ako u toj ćeliji ne postoji nijedan objekt, pouzdanost bi trebala biti nula, a u suprotnom, želimo da bude jednako omjeru preklapanja (IoU). Nadalje, svaki granični okvir ima pet parametara:  $x, y, h, w$  i koeficijent povjerenja. Koordinate  $(x, y)$  predstavljaju središte okvira u odnosu na granice ćelije mreže, a  $w$  i  $h$  širinu i visinu u odnosu na cijelu sliku. Koeficijent povjerenja predstavlja omjer presjeka i unije (Intersection over Union - IoU) između predviđene granice i stvarne granice objekta, čime se mjeri preciznost predikcije. Svaka ćelija mreže također predviđa vjerojatnosti uvjetne klase  $Pr(Class_i | Object)$ . Vjerojatnosti su uvjetovane na mreži ćelije koja sadrži objekt. Predviđa se samo jedan skup vjerojatnosti klase po ćeliji mreže bez obzira na broj okvira  $B$ . Za vrijeme testiranja množe se vjerojatnosti uvjetne klase i predviđanje pouzdanosti za pojedinačne okvire po izrazu (3-4).

$$Pr(Class_i | Object) * Pr(Object) * IoU_{pred}^{truth} = Pr(Class_i) * IoU_{pred}^{truth} \quad (3-4)$$

Tako nam daje pouzdanost rezultata za svaku klasu okvira. Ovi rezultati kodiraju vjerojatnost te klase koja se pojavljuje u okviru i koliko dobro predviđeni okvir odgovara objektu na slici. Primjer opisanog sustava možemo vidjeti na slici (Sl. 3.5).



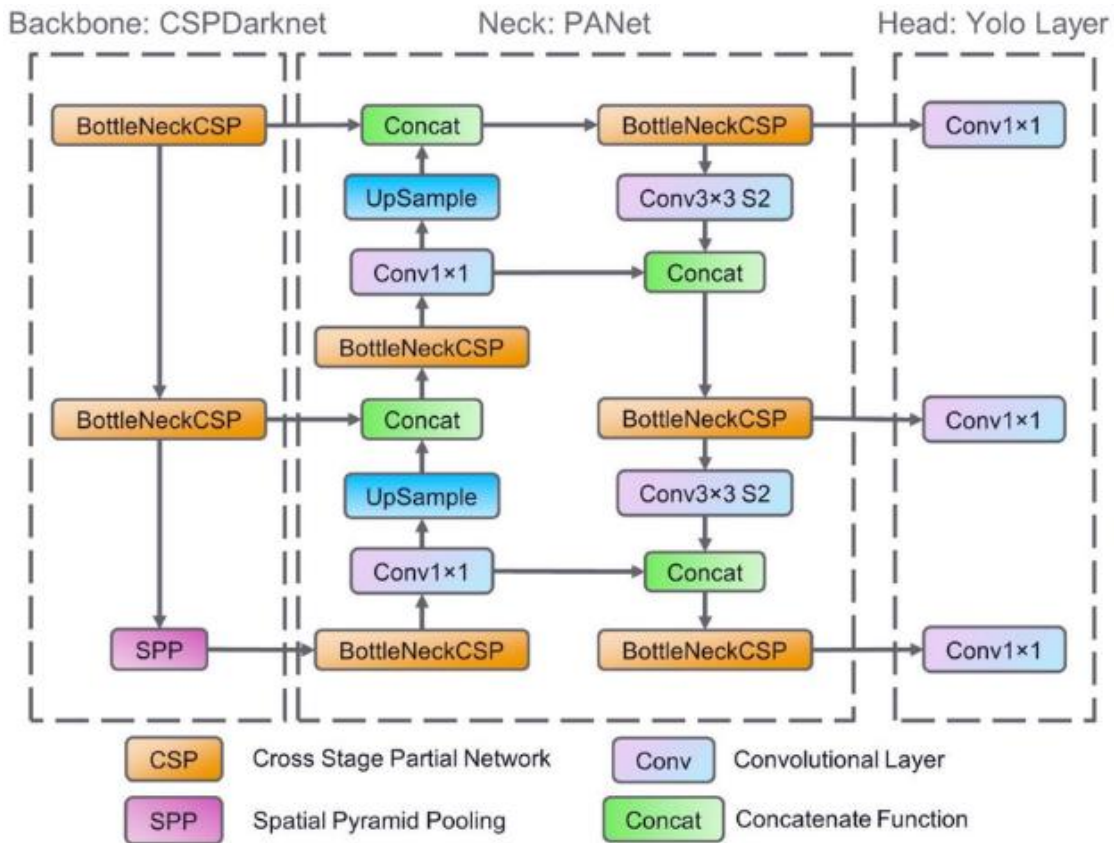
Sl. 3.5 Sustav za detekciju kao regresijski problem

Izvor: [https://www.researchgate.net/figure/YOLOv4-divides-the-image-into-regions-of-S-S-and-predicts-class-probabilities-along\\_fig1\\_351906130](https://www.researchgate.net/figure/YOLOv4-divides-the-image-into-regions-of-S-S-and-predicts-class-probabilities-along_fig1_351906130)

Od prvotne verzije YOLO modela, čiju smo ideju rada prethodno opisali, godinama nakon i dan danas su razvijane sve bolje verzije u kojima su se i autori izmijenili. Za naš rad smo se odlučili za YOLOv5 čiju ćemo arhitekturu u nastavku pobliže opisati temeljem službenog sažetka [21].

Struktura ove verzije mreže se sastoji od arhitekture modela CSPDarknet-53 kao kraljevnice mreže, zatim kao vrat slijede SPP i PANet sloj te na kraju dolazi YOLOv3 detekcijska glava. Ova struktura je prikazana skicom (Sl. 3.6)





Sl. 3.6 Arhitektura YOLOv5

Izvor: [https://www.researchgate.net/figure/The-neural-network-architecture-of-YOLOv5\\_fig4\\_375893271](https://www.researchgate.net/figure/The-neural-network-architecture-of-YOLOv5_fig4_375893271)

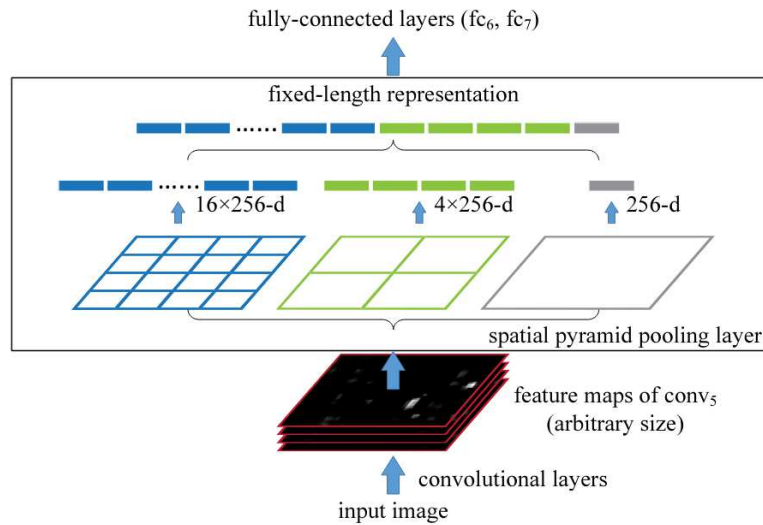
CSPDarknet-53 dio kombinira CSPNet i Darknet 53. Prema autorima, cilj dizajniranja CSPNet-a (Cross Stage Partial Network) je postići bogatiju kombinaciju gradijenata. Ova arhitektura postiže bolju efikasnost razdvajanjem mape značajki osnovnog sloja na dva dijela i njihovim spajanjem kroz hijerarhiju međufaza. Glavna ideja je da se tok gradijenata propagira kroz različite putanje mreže razdvajanjem gradijenata. Na taj način, potvrđeno je da propagirana informacija o gradijentu može imati veliku korelacijsku razliku promjenom koraka spajanja i prelaza. CSPNet također može značajno smanjiti količinu računanja i poboljšati brzinu inferencije, kao i točnost [22]. Darknet-53 je konvolucijska neuronska mreža korištena u YOLOv3 kao ekstraktor značajki. Sastoji se od 53 konvolucijska sloja, kombinirajući  $3 \times 3$  i  $1 \times 1$  konvolucijske slojeve i rezidualne veze kao na slici (Sl. 3.7). Ova arhitektura je učinkovitija i snažnija u usporedbi s prethodnim verzijama te postiže bolju ravnotežu između točnosti i brzine.

	Type	Filters	Size	Output
	Convolutional	32	3 × 3	256 × 256
	Convolutional	64	3 × 3 / 2	128 × 128
1x	Convolutional	32	1 × 1	128 × 128
	Convolutional	64	3 × 3	
	Residual			
	Convolutional	128	3 × 3 / 2	64 × 64
2x	Convolutional	64	1 × 1	64 × 64
	Convolutional	128	3 × 3	
	Residual			
	Convolutional	256	3 × 3 / 2	32 × 32
8x	Convolutional	128	1 × 1	32 × 32
	Convolutional	256	3 × 3	
	Residual			
	Convolutional	512	3 × 3 / 2	16 × 16
8x	Convolutional	256	1 × 1	16 × 16
	Convolutional	512	3 × 3	
	Residual			
	Convolutional	1024	3 × 3 / 2	8 × 8
4x	Convolutional	512	1 × 1	8 × 8
	Convolutional	1024	3 × 3	
	Residual			
	Avgpool		Global	
	Connected		1000	
	Softmax			

Sl. 3.7 Darknet-53

Izvor: [https://www.researchgate.net/figure/Darknet-53-Network-Architecture-24\\_fig3\\_344021825](https://www.researchgate.net/figure/Darknet-53-Network-Architecture-24_fig3_344021825)

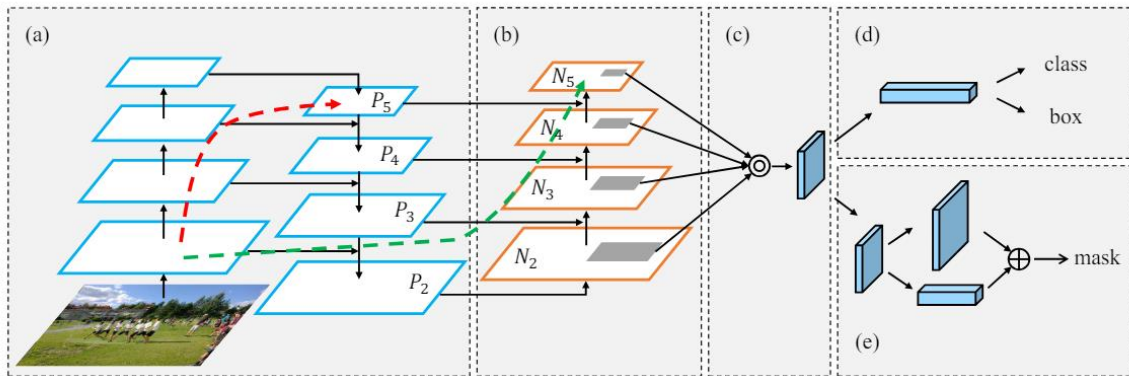
Slojevi prostorne piramidalne grupiranja (Spatial Pyramid Pooling - SPP) u mreži omogućuju obradu slika proizvoljnih veličina. Za razliku od ranijih metoda koje zahtijevaju fiksne dimenzije ulaznih podataka, SPP djeluje tako da grupira značajke iz prostornih binova prilagođenih veličini slike. Zamjenom tradicionalnog sloja grupiranja s SPP nakon posljednjeg konvolucijskog sloja, mreža proizvodi vektore značajki fiksne duljine. Ova fleksibilnost omogućuje mreži da obradi slike različitih skala bez promjene arhitekture ili veličine filtera, što poboljšava sposobnost ekstrakcije značajki na učinkovit način neovisno o veličini i mjerilima ulaznih slika. Primjer opisanog rada na slici (Sl. 3.8) [23].



Sl. 3.8 SPP

Izvor: [https://www.researchgate.net/figure/Network-structure-of-SPPNet-image-from-He-et-al-2015-256-denotes-the-number-of\\_fig4\\_369379297](https://www.researchgate.net/figure/Network-structure-of-SPPNet-image-from-He-et-al-2015-256-denotes-the-number-of_fig4_369379297)

Glavna ideja PANet-a (Path Aggregation Network) je poboljšati tok informacija u okviru za segmentaciju instanci temeljenom na prijedlozima kao na skici (Sl. 3.9). PANet poboljšava cijelu hijerarhiju značajki preciznim lokalizacijskim signalima u donjim slojevima pomoću augmentacije putanje odozdo prema gore, što skraćuje put informacija između donjih slojeva i najgornje značajke. Također uvode adaptivno sažimanje značajki koje povezuje mrežu značajki i sve razine značajki kako bi korisne informacije s svake razine mogle izravno proći do sljedećih podmreža prijedloga. Konačno, dodaje se grana za klasifikaciju i regresiju koja obuhvaća različite poglede svakog prijedloga kako bi se dodatno poboljšala predikcija maske [24].



Sl. 3.9 PANet

Izvor: [https://www.researchgate.net/figure/Pyramidal-feature-hierarchy-as-in-Tsung-Yi-et-al-2017-YOLOv4-by-Alexey-Chien-Yao\\_fig4\\_359200399](https://www.researchgate.net/figure/Pyramidal-feature-hierarchy-as-in-Tsung-Yi-et-al-2017-YOLOv4-by-Alexey-Chien-Yao_fig4_359200399)

Za programski dio, prethodno opisanog modela, koristili smo Colab, ili "Colaboratory", Googleov alat koji omogućuje pisanje i izvršavanje Python koda u web pregledniku bez potrebe za instalacijom. Posebno je koristan za strojno učenje, analizu podataka i edukaciju. Colab pruža besplatan pristup snažnim računalnim resursima, uključujući GPU-ove i TPU-ove, što omogućuje ubrzanje složenih izračuna. Korištenje Colab-a za neuronske mreže je iznimno učinkovito kao u našem primjeru di se YOLOv5 (s izvora: Wozniak, Alexey. "Ultralytics/yolov5: YOLOv5: PyTorch implementation for real-time object detection." GitHub, 2020, <https://github.com/ultralytics/yolov5>) može jednostavno implementirati, trenirati modele i testirati što znatno ubrzava process [25].

### 3.3. Detekcija

#### 3.3.1. YOLO gubitak

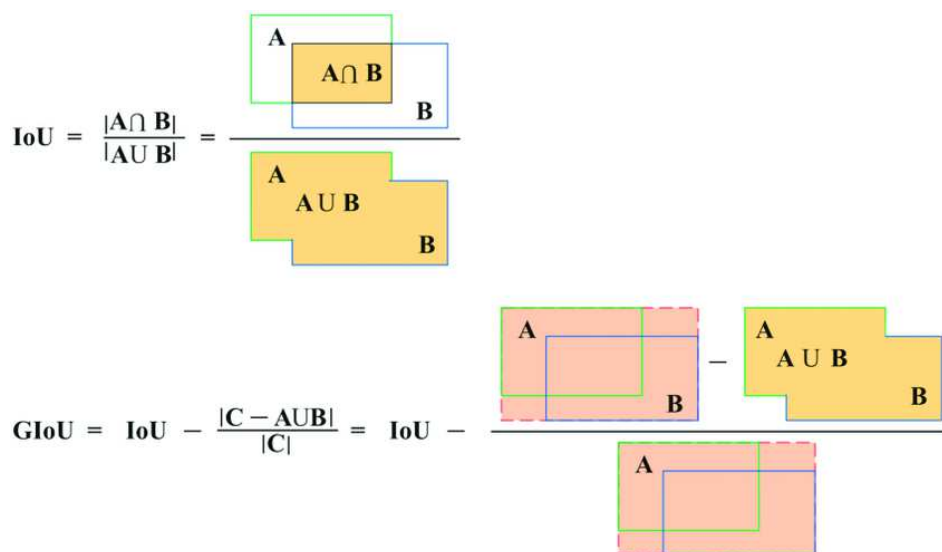
Za početak ćemo spomenuti, prema izvoru [21], kako se gubitak kod YOLOv5 sastoji od tri dijela:

- Gubitak klasa (BCE gubitak)
- Gubitak objektnosti (BCE gubitak)
- Gubitak lokacije (CIoU gubitak)

Gubici objekta za tri sloja predikcije računamo izrazom (3-5).

$$L_{obj} = 0.4 \cdot L_{obj}^{small} + 1.0 \cdot L_{obj}^{medium} + 0.4 \cdot L_{obj}^{large} \quad (3-5)$$

Complete Intersection over Union (CIoU) funkcija gubitka je poboljšanje u odnosu na tradicionalne funkcije gubitka Intersection over Union (IoU) prikazan slikom i Generalized Intersection over Union (GIoU) koje su prikazane slikom (Slika 3.10 IoU i GIoU).



Slika 3.10 IoU i GIoU

Izvor: [https://www.researchgate.net/figure/Diagrams-of-IoU-and-GIoU\\_fig3\\_367144033](https://www.researchgate.net/figure/Diagrams-of-IoU-and-GIoU_fig3_367144033)

CIoU uzima u obzir tri geometrijska faktora kako bi poboljšala performanse modela za detekciju objekata i segmentaciju instanci: površinu preklapanja, normaliziranu udaljenost središnje točke i dosljednost omjera stranica. Formula za CIoU je (3-6).

$$CIoU = S(A, B) + D(A, B) + V(A, B) \quad (3-6)$$

gdje:

- S je površina preklapanja, izračunata kao  $1 - IoU$
- D je normalizirana udaljenost središnjih točki

- V je dosljednost omjera stranica, koja uključuje složeniji izračun koji uključuje širinu i visinu stvarnog i predviđenog okvira.

Ovaj pristup nastoji pružiti precizniju i brže konvergentnu funkciju gubitka rješavajući ograničenja IoU i GIoU, koje uzimaju u obzir samo površinu preklapanja i često zahtijevaju više iteracija za konvergenciju, posebno u scenarijima bez preklapanja [26].

Završno, ukupan gubitak se onda računa kao zbroj ova tri prema izrazu (3-7).

$$Loss = \lambda_1 L_{class} + \lambda_2 L_{object} + \lambda_3 L_{location} \quad (3-7)$$

### 3.3.2. Evaluacijske mjere

Kako bismo mogli pričati o uspješnosti detekcije i klasifikacije potrebno je prvo definirati koje mjere ćemo u tu svrhu koristiti. Za izračun mjera učinkovitosti koristimo sljedeće vrijednosti:

- TP (true positive) točno pozitivni primjeri (okviri točni prema uvjetima IoU)
- FP (false positive) pogrešno pozitivni primjeri
- TN (true negative) točno negativni primjeri
- FN (false negative) pogrešno negativni primjeri (nisu prepoznati)

Prva mjera je preciznost (precision) koja nam govori koliki je udio točno pozitivno klasificiranih među svim pozitivno predviđenim prema izrazu (3-8).

$$P = \frac{TP}{TP + FP} \quad (3-8)$$

Odziv (recall) definiramo kao omjer točno pozitivno predviđenih i ukupno stvarno pozitivnih primjera koristeći izraz (3-9).

$$R = \frac{TP}{TP + FN} \quad (3-9)$$

Nadalje opisujemo prosječnu preciznost (average precision) koja je zapravo površina ispod krivulje preciznost-odziv (izraz (3-10)), a ta krivulja prikazuje preciznost izračunatu za određeni odziv.

$$AP = \int_0^1 p(r) dr \quad (3-10)$$

Na posljepku srednju prosječnu preciznost (mean average precision) računamo kao aritmetičku sredinu prethodno spomenute prosječne preciznosti za sve razrede objekta prema formuli (3-11).

$$mAP = \frac{1}{K} \sum_{i=0}^{K-1} AP_i \quad (3-11)$$

### 3.3.3. Testiranje

Koristeći podatke opisane u poglavlju 3.1 i YOLOv5 mrežu predstavljenu u 3.2 prvo provodimo detekciju zubi bez detaljne klasifikacije o kojoj je vrsti riječ.

Razdijelili smo skup slika i anotacija na nove skupove za treniranje, validaciju i testiranje tako da sadržavaju 80%, 10% i 10% podataka.

U *detect.yaml* datoteci smo definirali jednu klasu i nazvali je “tooth“ te koristeći predtrenirani yolo5s model sa 100 epoha kao što je prikazano na slici (Sl. 3.11) pokrenuli postupak treniranja i nakon testiranja dobili evaluacijske mjere (Sl. 3.12).

```
!python train.py --img 640 --cfg yolov5s.yaml --hyp hyp.scratch.yaml --batch -1 --epochs 100 --data detect.yaml
```

	all	60	1974	0.959	0.946	0.974	0.495
Epoch	GPU_mem	box_loss	obj_loss	cls_loss	Instances	Size	
91/99	12.6G	0.05323	0.1746	0	3942	640: 100% 8/8 [00:11<00:00, 1.43s/it]	
	Class	Images	Instances	P	R	mAP50	mAP50-95: 100% 1/1 [00:01<00:00,
	all	60	1974	0.955	0.939	0.972	0.5
Epoch	GPU_mem	box_loss	obj_loss	cls_loss	Instances	Size	
92/99	12.6G	0.05269	0.1691	0	3475	640: 100% 8/8 [00:11<00:00, 1.49s/it]	
	Class	Images	Instances	P	R	mAP50	mAP50-95: 100% 1/1 [00:01<00:00,
	all	60	1974	0.951	0.934	0.969	0.487
Epoch	GPU_mem	box_loss	obj_loss	cls_loss	Instances	Size	
93/99	12.6G	0.05326	0.1763	0	4080	640: 100% 8/8 [00:09<00:00, 1.22s/it]	
	Class	Images	Instances	P	R	mAP50	mAP50-95: 100% 1/1 [00:01<00:00,
	all	60	1974	0.954	0.948	0.974	0.506

Sl. 3.11 Treniranje detekcija

```
!python val.py --weights runs/train/teeth/weights/best.pt --data detect.yaml --task test
```

```
val: data=/content/yolov5/data/detect.yaml, weights=['runs/train/teeth/weights/best.pt'], batch_size=1, device='0', num_workers=0,
YOLOv5 v7.0-172-gc3c1304 Python-3.10.11 torch-2.0.1+cu118 CUDA:0 (Tesla T4, 15102MiB)
```

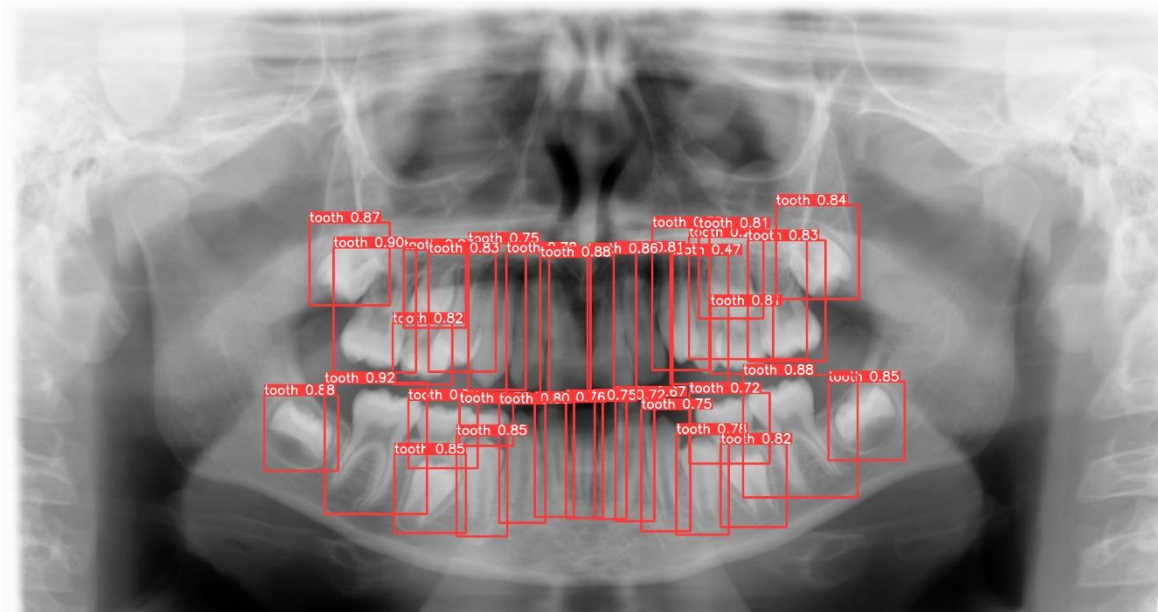
```
Fusing layers...
YOLOv5s summary: 157 layers, 7012822 parameters, 0 gradients, 15.8 GFLOPs
test: Scanning /content/yolov5/labels/test... 60 images, 0 backgrounds, 0 corrupt: 100% 60/60 [00:00<00:00, 1.00s/it]
test: New cache created: /content/yolov5/labels/test.cache
```

Class	Images	Instances	P	R	mAP50	mAP50-95: 100%
all	60	2049	0.945	0.934	0.952	0.506

Sl. 3.12 Rezultat detekcija

Jako dobri rezultati bili su očekivani zbog dobrih početnih težina modela i jednostavnosti zadatka zbog sadržavanja samo jedne klase. Ostvareni mali broj pogrešaka uzrokovan je lošijom kvalitetom pojedinih dijelova slike. Primjer kako izgleda rezultat procesa detekcije na jednoj od slika iz testnog skupa prikazan je ispod (Sl. 3.13).

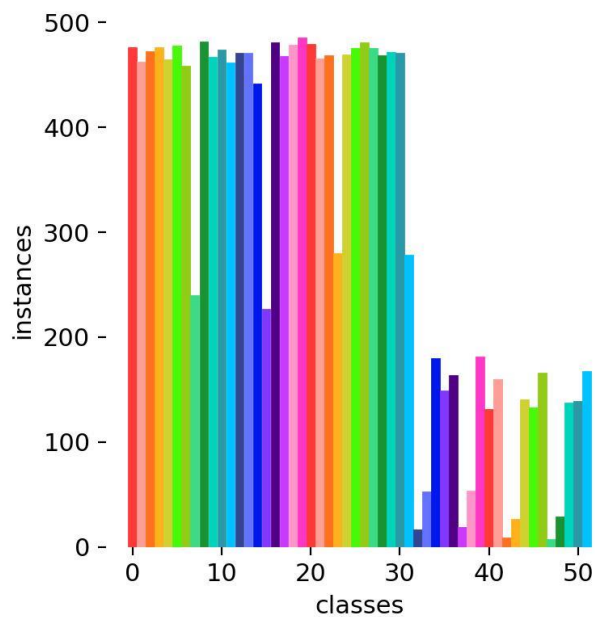




Sl. 3.13 Test slika detekcija

### 3.4. Klasifikacija

Nakon podosta uspješne detekcije potrebno je provesti zahtjevniju klasifikaciju vrste zubi koristeći anotacije koje nam ukazuju točnu vrstu pojedinih detektiranih okvira. Za očekivati je kako udio trajnih zubi u skupu podataka prevladava, što se vidi i na slici (Sl. 3.14).



Sl. 3.14 Klase zubi

Pošto je svaka vrsta zuba zasebna nova klasa, napravili smo i novi *teeth.yaml* file u kojem definiramo broj klasa *# number of classes nc: 52* , njihova imena *# class names: ['11', '12', ..., '85']* te *hyp.scratch.yaml* u kojem se korigiraju željeni hiperparametri modela.

Za sva testiranja smo koristili iste slike u skupu za treniranje, validaciju i testiranje te u istom omjeru 80:10:10 , a za razliku od detekcije umjesto 100 epoha koristimo 300 jer , iako vremenski zahtjevnije, pokazalo se kao znatno poboljšanje u točnosti.

Objasnit ćemo pet varijanti modela koje smo trenirali i testiranja uspješnosti u klasifikaciji koje smo proveli.

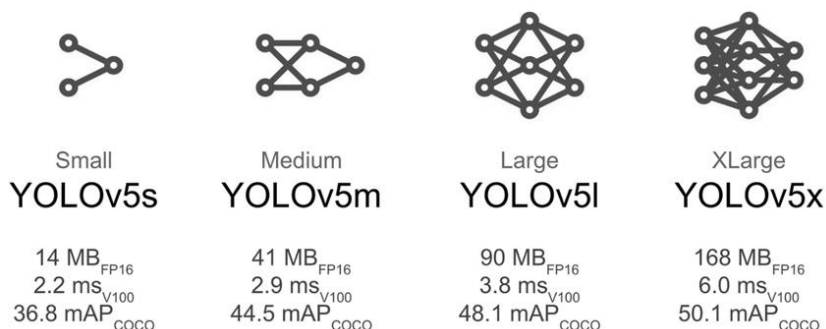
Prvo smo krenuli s istim modelom *S\_hp0* kao u klasifikaciju tj. yolo5s. Također, u treningu je korišten parametar *--img 640* čime je osigurano da model na ulazu dobiva sve slike jednakih veličina 640x640. Na posljetku hiperparametri su definirani kao na slici (Sl. 3.15).

```
copy_paste: 0.0 # segment copy-paste (probability)
lr0: 0.01 # initial learning rate (SGD=1E-2, Adam=1E-3)
lrf: 0.2 # final OneCycleLR learning rate (lr0 * lrf)
momentum: 0.937 # SGD momentum/Adam beta1
weight_decay: 0.0005 # optimizer weight decay 5e-4
warmup_epochs: 3.0 # warmup epochs (fractions ok)
warmup_momentum: 0.8 # warmup initial momentum
warmup_bias_lr: 0.1 # warmup initial bias lr
box: 0.05 # box loss gain
cls: 0.5 # cls loss gain
cls_pw: 1.0 # cls BCELoss positive_weight
obj: 1.0 # obj loss gain (scale with pixels)
obj_pw: 1.0 # obj BCELoss positive_weight
iou_t: 0.20 # IoU training threshold
anchor_t: 4.0 # anchor-multiple threshold
# anchors: 3 # anchors per output layer (0 to ignore)
fl_gamma: 0.0 # focal loss gamma (efficientDet default gamma=1.5)
hsv_h: 0.015 # image HSV-Hue augmentation (fraction)
hsv_s: 0.7 # image HSV-Saturation augmentation (fraction)
hsv_v: 0.4 # image HSV-Value augmentation (fraction)
degrees: 0.0 # image rotation (+/- deg)
translate: 0.1 # image translation (+/- fraction)
scale: 0.5 # image scale (+/- gain)
shear: 0.0 # image shear (+/- deg)
perspective: 0.0 # image perspective (+/- fraction), range 0-0.001
flipud: 0.0 # image flip up-down (probability)
fliplr: 0.5 # image flip left-right (probability)
mosaic: 1.0 # image mosaic (probability)
mixup: 0.0 # image mixup (probability)
```

Sl. 3.15 Hiperparametri 1

Oznaka *s* na kraju naziva modela yolo5s označava da se radi o maloj “small“ verziji, a naš drugi primjer *M\_hp0*, točnije yolo5m “medium“, sadržava sve jednake parametre treniranja

kao prethodnik, ali se radi o većem i složenijem modelu kao što je prikazano slikom (Sl. 3.16).



Sl. 3.16 Veličine modela

Izvor: [https://www.researchgate.net/figure/YOLOv5-different-model-sizes-where-FP16-stands-for-the-half-floating-point-precision\\_fig3\\_354846944](https://www.researchgate.net/figure/YOLOv5-different-model-sizes-where-FP16-stands-for-the-half-floating-point-precision_fig3_354846944)

Svakako treba uzimati u obzir da složeniji modeli troše više prostora i vremena za treniranje te naći balans u odabiru odgovarajuće veličine i željenih rezultata.

Za treću verziju  $M_{hp1}$  ostavili smo prethodni model zbog odličnih rezultata kao i veličinu slike, ali smo promijenili dva hiperparametara:

- *obj*: 0.80
- *mixup*: 0.2

Prvi hiperparametar *obj* (*obj loss gain*) tiče se gubitka objekta koji smo spominjali u 3.3.1 kao dio ukupne funkcije gubitka YOLO modela, a govori koliko je uvjerenje da je objekt unutar predviđenog okvira. Smanjivanjem ove vrijednosti u odnosu na prvotne smo smanjili utjecaj ovog gubitka na ukupni.

Nadalje, *mixup* (*image mixup*) je tehnika kojom stvaramo nove primjere za treniranje miješanjem parova slika i odgovarajućih oznaka. Povećanjem ovog hiperparametra na 0.2 stvaramo 20% šanse da će na primjeru za treniranje biti primijenjena ova tehnika. Ovo se često koristi kako bi poboljšali generalizaciju i robusnost modela.

Kako su se ove promjene pokazale dobre, u sljedećoj verziji  $M_{hp2}$  smo također probali promijeniti neke od hiperparametara:

- *lr0*: 0.001
- *lrf*: 0.01

Oba se tiču brzine učenja u procesu treniranja i to lr0 kao početna brzina učenja i lrf kao vrijednosti koja množi početnu da bi dobili konačnu brzinu. Njihovim smanjivanjem se pogoršala točnost te nismo nastavili u tom smjeru.

Za kraj mijenjanja parametra u M\_hp3 odlučili smo upitati AI (<https://chatgpt.com/>) uz malo konteksta koje bi parametre promijenio i to su bili sljedeći:

- lr0: 0.001
- lrf: 0.1
- warmup\_epochs: 5.0
- cls: 0.6
- iou\_t: 0.25
- degrees: 5.0

Osim već objašnjenih za brzinu učenja koje su ovdje smanjene, ostali redom služe za dodavanje epoha za zagrijavanje, kontrolu gubitka klasifikacije, prag za IoU te dodavanje rotacije, ali njihovim blagim povećanjem su postignuti lošiji rezultati.

Na posljertku, zadnjem primjeru *M\_hp1\_l* nismo mijenjali hiperparametre već smo uzeli *M\_hp1* verziju, a odlučili promijeniti veličinu slika na ulazu povećavajući dosadašnji parametar –img 640 na 1024. Iako veće slike zahtijevaju više memorije i računalne snage, htjeli smo ovakvim postupkom probati očuvati više detalja originalnih slika.

U sljedećim potpoglavljima slijede detaljni rezultati ovih pet provedenih implementacija.

### 3.4.1. Analiza točnosti za vrstu zubi

Glavna klasifikacija kojom smo određivali uspješnost pojedinog modela je ona za pojedinu vrstu zuba. Rezultati su u tablici (Tablica 1).

Tablica 1 Rezultati vrsta zubi

vrsta	broj	S_hp0	M_hp0	M_hp1	M_hp2	M_hp3	M_hp1_l
svi	1974	P	P	P	P	P	P
<b>11</b>	58	0.747	0.78	0.817	0.263	0.206	0.756
<b>12</b>	57	0.735	0.79	0.864	0.131	0.118	0.704

13	60	0.802	0.783	0.817	0.411	0.223	0.703
14	59	0.718	0.764	0.809	0.411	0.0438	0.607
15	60	0.775	0.799	0.847	0.292	0.171	0.688
16	62	0.755	0.824	0.866	0.237	0.185	0.73
17	55	0.835	0.779	0.838	0.242	0.208	0.654
18	32	0.78	0.709	0.77	0.315	0.239	0.621
21	60	0.588	0.755	0.835	0.249	0.209	0.811
22	58	0.672	0.828	0.825	0.158	0.173	0.721
23	61	0.749	0.823	0.828	0.458	0.243	0.813
24	62	0.71	0.66	0.717	0	0.0556	0.733
25	58	0.639	0.729	0.706	0.486	0.163	0.719
26	60	0.758	0.8	0.855	0.261	0.204	0.756
27	57	0.753	0.755	0.811	0.231	0.194	0.733
28	28	0.698	0.713	0.77	0.326	0.217	0.596
31	57	0.848	0.845	0.79	0.243	0.184	0.762
32	60	0.793	0.904	0.937	0.185	0.187	0.848
33	61	0.768	0.802	0.918	0.331	0.171	0.832
34	61	0.796	0.792	0.921	0.279	0.188	0.907
35	57	0.786	0.804	0.83	0.238	0.187	0.82
36	58	0.877	0.878	0.877	0.272	0.174	0.907
37	58	0.763	0.802	0.883	0.219	0.173	0.803
38	38	0.765	0.834	0.879	0.309	0.16	0.855
41	61	0.794	0.768	0.875	0.297	0.275	0.807
42	60	0.862	0.834	0.865	0.226	0.124	0.816
43	61	0.861	0.837	0.834	0.339	0.182	0.785
44	58	0.929	0.808	0.821	0.199	0.152	0.74
45	55	0.859	0.789	0.872	0.265	0	0.758
46	56	0.935	0.901	0.869	0.284	0	0.75
47	60	0.885	0.825	0.871	0.203	0	0.745
48	37	0.944	0.856	0.875	0.297	0.0833	0.724
51	1	1	1	1	0	1	1
52	5	1	0	0.287	0	0	0.632
53	19	0.964	0.84	0.704	1	0	0.685

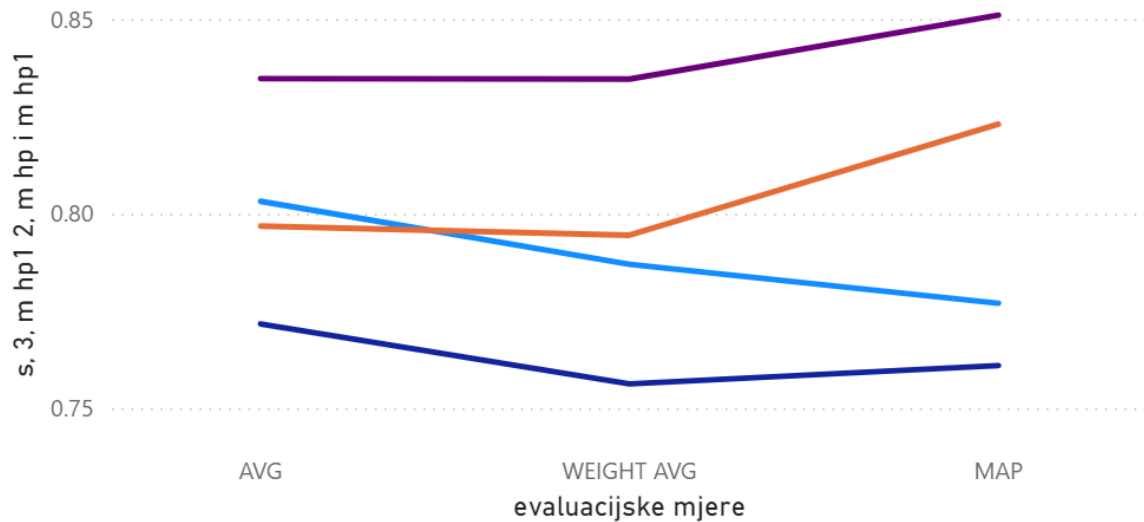
54	12	0.958	0.806	0.821	1	0.345	0.701
55	16	0.864	0.713	0.94	1	1	0.589
61	1	1	1	1	0	0	1
62	5	0	0.927	1	0	0	1
63	18	0.71	0.633	0.629	1	1	0.682
64	12	0.857	0.795	0.915	1	1	0.703
65	16	0.729	0.698	0.696	1	1	0.802
71	1	1	1	1	0	1	1
72	3	1	1	1	0	0	1
73	12	0.787	0.635	0.638	1	1	0.611
74	11	0.718	0.739	0.692	1	1	0.692
75	15	0.577	0.58	0.618	0.443	0	0.604
81	0	1	1	1	1	1	1
82	1	1	1	1	0	0	1
83	14	0.905	0.832	0.881	1	1	0.897
84	11	0.615	0.854	0.868	1	1	0.569
85	16	0.901	0.811	0.821	0.48	0	0.758
<b>AVG</b>		0.8032	0.7968	<b>0.8347</b>	0.3958	0.3161	0.7717
<b>WEIGHT AVG</b>		0.7870	0.7945	<b>0.8346</b>	0.3213	0.2021	0.7563
<b>MAP</b>		0.777	0.823	<b>0.851</b>	0.214	0.142	0.761

Ovdje je osim naziva vrste i stupac s brojem instanci i tu opet dolazimo do zaključka da mliječnih zubi ima dosta manje u odnosu na trajne te se može dogoditi da u testnom skupu imamo samo par primjera ili čak niti jedan od pojedine rjeđe klase. Gledajući prosjeke svih preciznosti i težinski koji uzima u obzir i broj instanci, vidimo da najbolje rezultate daje  $M_{hp1}$  sa rezultatima oko 83,5%. Uz njega su vrlo blizu  $M_{hp0}$  i  $M_{hp\_l}$ , a ostale modele ćemo u budućim pregledima izostaviti zbog loših rezultata. Također je zanimljivo da  $M_{hp1}$  i  $M_{hp0}$  imaju veći maP od AP-a dok ostali modeli bolje rezultate postižu u preciznosti.

Spomenutu situaciju bolje uočavamo na grafičkom prikazu evaluacijskih mjera za najbolja četiri modela prikazan slikom (Sl. 3.17).

s, 3, m hp1 2, m hp i m hp1 kategorije evaluacijske mjere

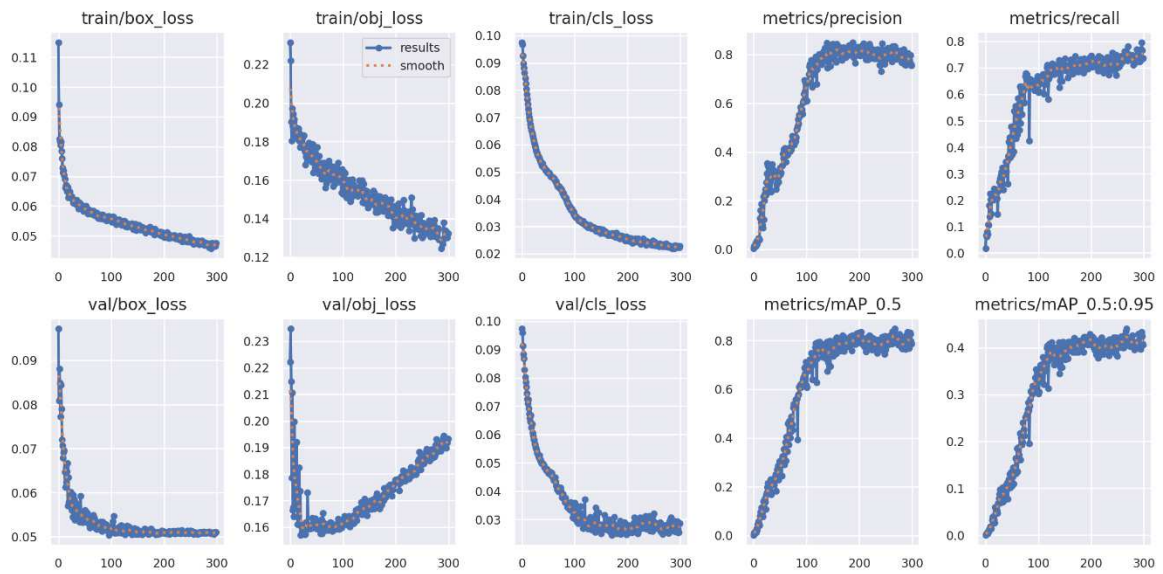
● s, 3 ● m hp1 2 ● m hp ● m hp1



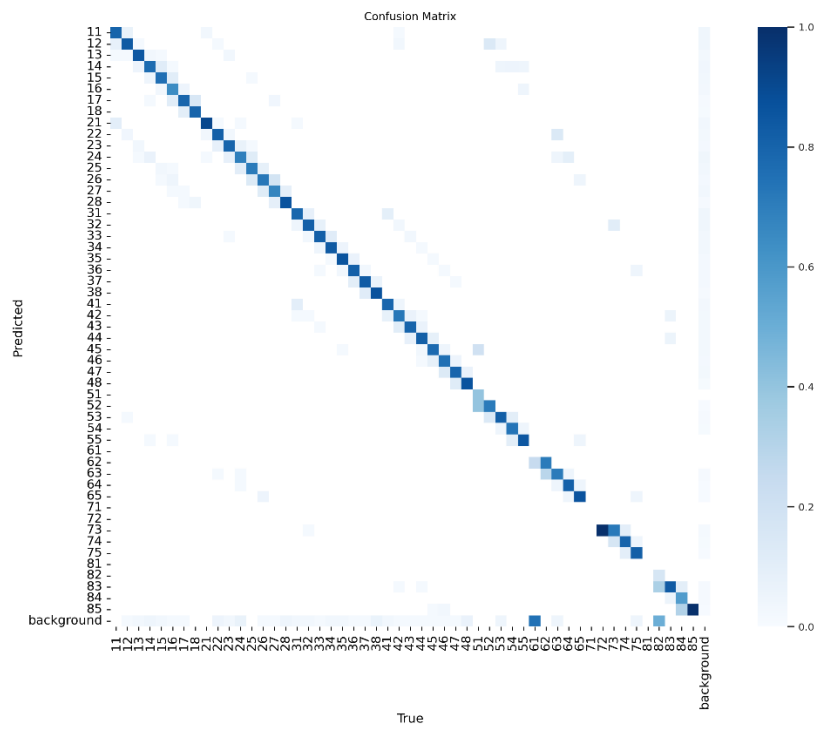
Sl. 3.17 Evaluacijske mjere za vrstu

Za  $M_{hp1}$  model vidimo u rezultantnoj slici treniranja (Sl. 3.18) kako se određene mjere mijenjale kroz broj epoha te testnu konfuzijsku matricu (Sl. 3.19) u kojoj vidimo da većina pogrešno klasificiranih zubi je zamijenjena sa susjednim ili s pozadinom što je uzrokovano smetnjama na originalnim snimkama.

Kako bismo trenutno najuspješniji model mogli koristiti i testirati bez ponovnog treniranja spremamo datoteku *best.pt* u kojoj su sadržane najbolje težine mreže.



Sl. 3.18 Treniranje za vrstu



Sl. 3.19 Konfuzijska matrica



### 3.4.2. Analiza točnosti po grupama

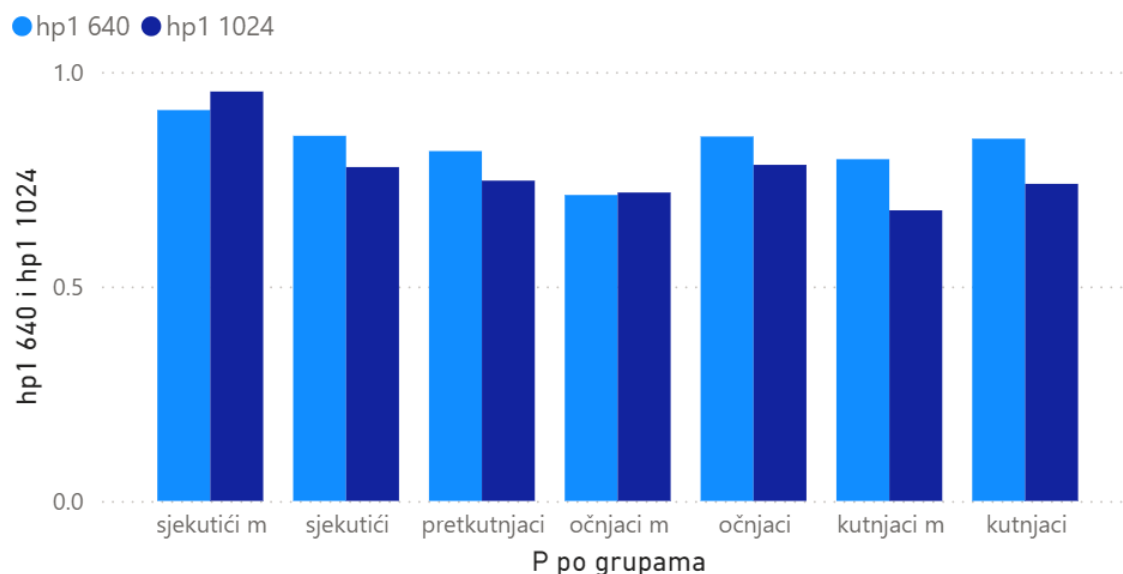
Zanimajući se postoji li određena grupa zubi koja se bolje ili lošije klasificira u provedenom testiranju, uzeli smo prosjek preciznosti za svaku grupu pri tom odvajajući zasebno trajne i mliječne. U tablici (Tablica 2 Rezultati po grupama) prikazane su rezultatne preciznosti za ovakve grupacije koristeći modele  $M_{hp1}$  i  $M_{hp1_l}$ . Ovi modeli imaju iste hiperparametre, ali se razlikuju po veličini ulaznih slika pa smo htjeli vidjeti ukoliko neka grupa je bolje prepoznavana koristeći veću rezoluciju. Vidimo da su općenito najbolje prepoznavani mliječni sjekutići, a zatim očnjaci.

Tablica 2 Rezultati po grupama

<i>P po grupama</i>	<b>hp1 640</b>	<b>hp1 1024</b>
<b>sjekutići</b>	0.851	0.7781
<b>očnjaci</b>	0.8493	0.7833
<b>pretkutnjaci</b>	0.8154	0.7465
<b>kutnjaci</b>	0.8442	0.739
<b>sjekutići m</b>	0.9109	0.9542
<b>očnjaci m</b>	0.713	0.7188
<b>kutnjaci m</b>	0.7964	0.6773

Koristeći vizualizaciju (Sl. 3.20 Usporedba po grupama) vidimo da se za pojedine grupe model s većim ulaznim slikama ( $M_{hp1_l}$ ) pokazao učinkovitijim i to za mliječne sjekutiće i mliječne očnjake koji su ujedno i najmanji zubi na našim snimkama.

### P po grupama za hp1 640 i hp1 1024



Sl. 3.20 Usporedba po grupama

### 3.4.3. Analiza točnosti po godinama

Osim slika i anotacija, tijekom istraživanja nam je ustupljena još jedna dodatna datoteka izgleda (Sl. 3.21) u kojoj za svaku pojedinu sliku imamo informacije o dobi i spolu osobe koja je na snimku.

A	B	C	D	E
42	ImageFile_f8b151cd-47	6.622861054	male	(6, 7]
43	ImageFile_14e59cdc-cf	6.461327858	male	(6, 7]
44	ImageFile_6c2abf4a-72	6.986995209	male	(6, 7]
45	ImageFile_b73ea157-d2	6.976043806	male	(6, 7]
46	ImageFile_944a798a-4b	6.877481177	male	(6, 7]
47	ImageFile_3d11be0d-c2	6.735112936	male	(6, 7]
48	ImageFile_6ce18004-9a	6.078028747	male	(6, 7]
49	ImageFile_f9eaa994-a1	6.663928816	male	(6, 7]
50	ImageFile_24861338-2f	7.874058864	female	(7, 8]
51	ImageFile_fd62bbe1-a6	7.570157426	female	(7, 8]
52	ImageFile_5a67aab5-20	7.389459274	female	(7, 8]
53	ImageFile_cf630bd9-4e	7.326488706	female	(7, 8]
54	ImageFile_8c4e99fb-31	7.753593429	female	(7, 8]
55	ImageFile_4be57d8f-ba	7.247091034	female	(7, 8]
56	ImageFile_c16c5b77-03	7.926078029	female	(7, 8]

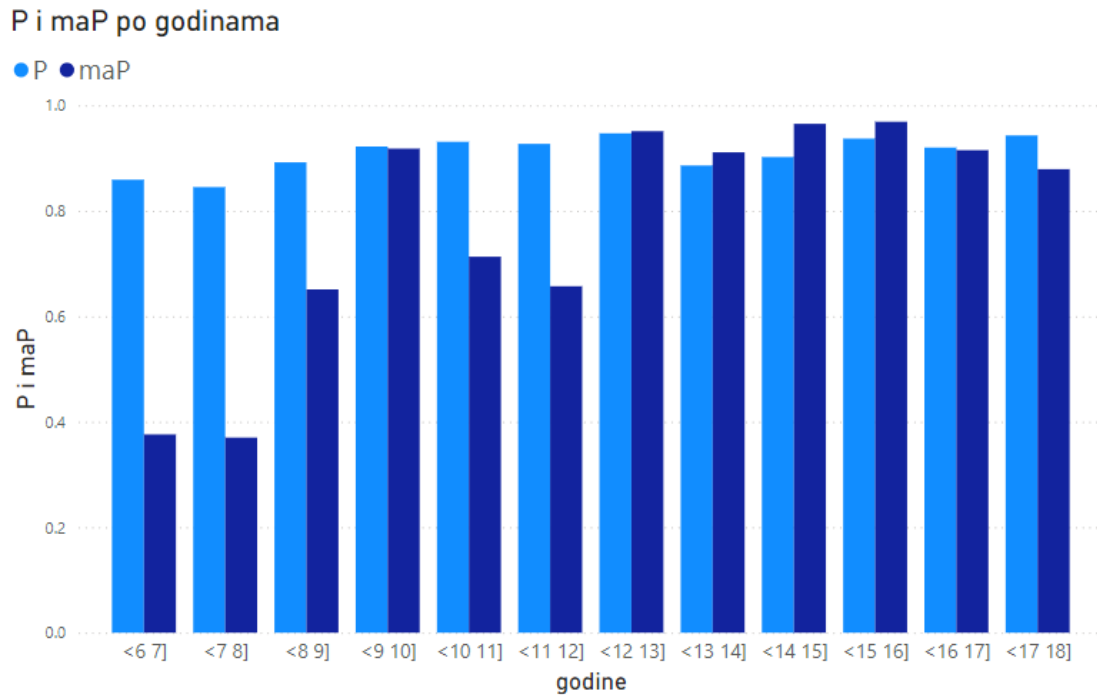
Sl. 3.21 Dodatne informacije o slikama

Zahvaljujući novim informacijama otkrivamo da je 600 slika od osoba starosti od 6 do 18 godina i to s jednakim udjelom svakog uzrasta, točnije po 50 slika za svaku grupu krenuvši od <6, 7] do <17, 18]. Napravili smo posebne nove mape za slike i anotacije za 12 starosnih grupa te iskoristili težine *best.pt* iz prethodno istrenirane mrežu *M\_hpl* kako bismo vidjeli rezultatne preciznosti i maP što je prikazano tablicom (Tablica 3 Rezultati po godinama). Globalno starija djeca imaju bolje rezultate klasifikacije zubi, a znamo da je s godinama sve manji udio mliječnih zubi što doprinosi takvom rezultatu. Najbolja preciznost se pokazala za grupu od 12. do 13., što bi moglo biti uzrokovano činjenicom da često nekima nisu izrasli kutnjaci koji se najlošije prepoznaju, a mliječni zubi su uglavnom ispalili. Odmah iza te grupe je ona najstarija od 17 do 18 godine gdje su zubi već razvijeni i očekivano dosta dobro prepoznati.

Tablica 3 Rezultati po godinama

godine	broj	P	maP
<6 7]	2017	0.859	0.376
<7 8]	1938	0.845	0.37
<8 9]	1880	0.892	0.651
<9 10]	1818	0.922	0.918
<10 11]	1672	0.931	0.713
<11 12]	1628	0.927	0.657
<12 13]	1579	<b>0.947</b>	0.951
<13 14]	1552	0.886	0.911
<14 15]	1556	0.902	<b>0.965</b>
<15 16]	1542	0.937	<b>0.969</b>
<16 17]	1547	0.92	0.915
<17 18]	1536	<b>0.943</b>	0.879

Uvijek je zanimljivo za usporediti i maP vrijednost po kojoj su najbolje grupe od 14 do 15 i od 15 do 16 godine, a kako se ta vrijednost razlikuje po kategorijama godina s obzirom na preciznost prikazano je grafički (Sl. 3.22 Evaluacijske mjere po godinama).



Sl. 3.22 Evaluacijske mjere po godinama

Vidimo da, za razliku od preciznosti koja je za većinu grupa konstantno visoka, maP ukazuje da bismo mogli detaljnije pogledati skupove od 10 do 11 i od 11 do 12 godine gdje se vidi dosta veliki pad u vrijednosti. Prikaz dijela testnog skupa za lošije ocjenjene grupe je dan slikom (Sl. 3.23) te je vidljiv udio lošijih snimaka koje otežavaju mreži postupak klasifikacije.



Sl. 3.23 Primjer lošijeg skupa

### 3.4.4. Analiza točnosti za spol

U prethodno spomenutoj dodatnoj datoteci (Sl. 3.21 Dodatne informacije o slikama) smo dobili informaciju i o spolu pacijenta. Naime u svakoj starosnoj skupini je jednaki udio djevojaka i dječaka, odnosno u cijelom skupu je jednak omjer. Iskorištavajući prethodno saznanje da slike osoba starosti od 15. do 16. godine imaju najbolju maP vrijednost u postupku testiranja, odlučili smo upravo tu grupu testirati s obzirom na spol. Gledajući rezultate u tablici (Tablica 4) vidimo da malu prednost imaju ženski primjeri u skupu iako zbog malog broja slika i mogućnosti da neke nisu skroz jasne sve rezultate gledamo samo kao ishod ovog testiranja, a ne i opće zaključke.

Tablica 4 Rezultati po spolu

	P	maP
muški	0.914	0.957
ženski	0.994	0.983

# Zaključak

Odabir YOLO algoritma za rješavanje problema detekcije i klasifikacije zubi pokazao se dobrim izborom dajući zadovoljavajuće rezultate uz veliku brzinu i jednostavnosti korištenja. Iako se 600 slika možda na početku činilo dovoljnim, kao što to često biva, pokazalo se da je za učenje poželjno imati još više podataka. Neke vrste zubi bile su vrlo slabo zastupljene, a dodatna razdvajanja po spolu i godinama rezultirala su malim brojem slika u testnom skupu. Također, zbog mogućih pogrešaka tijekom snimanja, neke slike nisu prikazivale određene dijelove zubi ili su ih prikazivale prilično nejasno, što je značajno utjecalo na mogućnost prepoznavanja pa samim time i evaluacijske mjere. Nastavno na spomenuto, takve okolnosti ukazuju na važnost postupka obrade slika, čime se u ovom radu, gdje smo naglasak stavili na sam postupak analize, nismo dovoljno bavili. Za buduće implementacije, preporučljivo je provesti postupke poboljšanja slike za lošije primjere iz baze te povećati broj slika koliko god je to moguće.

Postoji značajan prostor za napredak i u samom modelu. Jedan od mogućih smjerova je korištenje novijih verzija algoritma koje su već dostupne i koje će se sigurno još razvijati. Nadalje, ovo istraživanje pokazalo je da hiperparametri značajno utječu na performanse modela, pa bi bilo korisno dodatno istražiti njihov utjecaj i načine prilagodbe za uspješnije rezultate.

U budućem radu, bilo bi zanimljivo iskoristiti ostale označene atribute “završen rast korijena“ i “iznikli zub“ koje u ovom radu nismo koristili. Ovakve informacije mogu biti od velike koristi stomatolozima i ostalim dentalnim radnicima, koji bi odmah nakon snimanja bili upozoreni na prisutnost takvih primjera kod pacijenata.

Ovaj rad ima potencijal biti dobra baza za buduća istraživanja u području detekcije i klasifikacije zubi, posebice za skupove podataka koji se odnose na maloljetne zube, koji su rjeđe korišteni u dosadašnjim istraživanjima.

# Literatura

- [1] Ivanković, D., Pokupec, R. *Povijest radiologije i njezin utjecaj na dijagnostiku*. Hrvatski znanstveni radovi (2015).
- [2] Novak, I., Blažević, K. *Primjena umjetne inteligencije u analizi medicinskih slika*. Journal of Medical Technology, (2020).
- [3] Šimić, I., Kovačević, M. *Telemedicina tijekom pandemije COVID-19*. Acta Medica Croatica (2020).
- [4] Vodanović, M.; Subašić, M.; Milošević, D.; Savić Pavičin, I. *Artificial Intelligence in Medicine and Dentistry*. Acta Stomatol. Croat. (2023), 57, str. 70–84
- [5] Levin, M. A., Levin, I. T. *Artificial Intelligence in Dentistry and Oral Health*, Journal of Dental Research, 97, 12 (2018), str. 1304-1310.
- [6] Schwendicke, A., Göstemeyer, C. *Machine Learning in Dentistry: A Scoping Review*, Journal of Dental Research, 100, 3 (2021), str. 232-245.
- [7] Park, H. W., Kim, K. C. *Artificial Intelligence in Orthodontics*, The Korean Journal of Orthodontics, 50, 2 (2020), str. 61-70.
- [8] Gonzalez, Woods, *Digital Image Processing*, 4th Ed., Pearson, (2018)
- [9] Szeliski, R. , *Computer Vision: Algorithms and Applications*, London, Springer, (2010)
- [10] Haykin, S. *Neural Networks and Learning Machines*, 3rd ed.; Pearson Education, (2009), str. 11-20.
- [11] LeCun, Y., Bengio, Y., & Hinton, G. *Deep learning*, Nature, (2015), str. 436-444.
- [12] O’Shea K., Nash R. *An introduction to convolutional neural networks*, arXiv preprint arXiv:1511.08458, (2015).
- [13] Yamashita, R., Nishio, M., Do, R.K.G. *Convolutional neural networks: an overview and application in radiology*, (2018), str. 611–629.
- [14] Sultana, F.; Sufian, A.; Dutta, P. *A Review of Object Detection Models based on Convolutional Neural Network*, Intelligent Computing: Image Processing Based Applications. Advances in Intelligent Systems and Computing, Vol. 1157, (2020)
- [15] Wang, M.; Leelapatra, W. *A Review of Object Detection Based on Convolutional Neural Networks and Deep Learning*, International Scientific Journal of Engineering and Technology (ISJET), Vol. 6, No. 1, (2022)
- [16] "XML", Wikipedija, (2024, lipanj). Poveznica: <https://hr.wikipedia.org/wiki/XML>; pristupljeno 21. lipnja 2024.
- [17] "FDI World Dental Federation notation," Wikipedia, (2024, lipanj). Poveznica: [https://en.wikipedia.org/wiki/FDI\\_World\\_Dental\\_Federation\\_notation](https://en.wikipedia.org/wiki/FDI_World_Dental_Federation_notation); pristupljeno 21. lipnja 2024.
- [18] "YOLO: Real-Time Object Detection," pjredditie.com. Poveznica: <https://pjredditie.com/darknet/yolo/>; pristupljeno 21. lipnja 2024.

- [19] "Zub" Wikipedia, (2024, lipanj). Poveznica: <https://hr.wikipedia.org/wiki/Zub>; pristupljeno 21. lipnja 2024.
- [20] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, *You Only Look Once: Unified, Real-Time Object Detection*, 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, SAD, (2016), str. 779-788
- [21] "YOLOv5 (6.0/6.1) brief summary" github.com, Poveznica: <https://github.com/ultralytics/yolov5/issues/6998>; pristupljeno 22. lipnja 2024.
- [22] Wang, C. Y., Liao, H. Y. M., Wu, Y. H., Chen, P. Y., Hsieh, J. W., Yeh, I. H. *CSPNet: A New Backbone That Can Enhance Learning Capability of CNN*, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), (2020), str. 1571-1580
- [23] He, K., Zhang, X., Ren, S., & Sun, J. *Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition*, ArXiv preprint arXiv:1406.4729 (2014)
- [24] Liu, S., Qi, L., Qin, H., Shi, J., Jia, J. *Path Aggregation Network for Instance Segmentation*, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), (2018)
- [25] "Colaboratory" research.google.com, Poveznica: <https://research.google.com/colaboratory/faq.html>; pristupljeno 22. lipnja 2024.
- [26] Karthik. "Complete IoU," (2020, svibanj). Poveznica: <https://karthikziffer.github.io/journal/Complete-IoU.html>; pristupljeno 24. lipnja 2024.



## **Sažetak**

### **Detekcija i klasifikacija pojedinih zuba iz panoramske rendgenske snimke maloljetnika**

Cilj ovog rada je istražiti i implementirati metodu za detekciju i klasifikaciju pojedinih zuba iz panoramskih rendgenskih snimaka maloljetnika koristeći tehnike dubokog učenja. Prvo je proučeno područje problema te postojeće metode za detekciju i klasifikaciju objekata u slikama, s posebnim naglaskom na konvolucijske neuronske mreže. Zatim je prikupljena baza snimaka maloljetnih osoba te su konfigurirane odgovarajuće oznake vrste i lokacije pojedinog trajnog i mliječnog zuba. Implementirane su različite varijante YOLO mreže te su testirane na prikupljenim podacima uz dodatnu analizu uspješnosti za svaku grupu zubi te godine i spol osoba na slikama.

**Ključne riječi:** detekcija, klasifikacija, zubi, maloljetnici, neuronska mreža, yolo

# Summary

## Detection and classification of individual teeth from panoramic X-rays of minors

The aim of this paper is to research and implement a method for the detection and classification of individual teeth from panoramic X-ray images of minors using deep learning techniques. At first, the problem domain and existing methods for object detection and classification in images were studied, with a special emphasis on convolutional neural networks. Then, a database of X-ray images of minors was collected and appropriate annotations of the type and location of each permanent and deciduous tooth were configured. Various versions of the YOLO network were implemented and tested on the collected data, with additional performance analysis for each group of teeth, as well as the age and gender of the individuals in the images.

**Keywords:** detection, classification, teeth, minors, neural network, yolo