

Programska aplikacija za praćenje ljudskog pogleda

Jakovac, Jakov

Undergraduate thesis / Završni 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://um.nsk.hr/um:nbn:hr:168:977392>

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

Download date / Datum preuzimanja: **2025-04-01**



Repository / Repozitorij:

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



SVEUČILIŠTE U ZAGREBU
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

ZAVRŠNI RAD br. 1650

**PROGRAMSKA APLIKACIJA ZA PRAĆENJE LJUDSKOG
POGLEDA**

Jakov Jakovac

Zagreb, lipanj 2024.

SVEUČILIŠTE U ZAGREBU
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

ZAVRŠNI RAD br. 1650

**PROGRAMSKA APLIKACIJA ZA PRAĆENJE LJUDSKOG
POGLEDA**

Jakov Jakovac

Zagreb, lipanj 2024.

ZAVRŠNI ZADATAK br. 1650

Pristupnik: **Jakov Jakovac (0036535126)**
Studij: Elektrotehnika i informacijska tehnologija i Računarstvo
Modul: Računarstvo
Mentor: prof. dr. sc. Tomislav Burić

Zadatak: **Programska aplikacija za praćenje ljudskog pogleda**

Opis zadatka:

Algoritmi za praćenje ljudskog pogleda ključni su za razvoj interaktivnih sustava koji mogu interpretirati i reagirati na ljudske namjere i želje, koristeći se informacijama dobivenima iz smjera i fokusa pogleda korisnika. Cilj ovog rada je osmisliti i implementirati aplikaciju za praćenje ljudskog pogleda koristeći poznate algoritme. Rezultate je potrebno analizirati i prikazati na prikladan način.

Rok za predaju rada: 14. lipnja 2024.

Zahvale

Posebnu zahvalu htio bih dati prvotno svom mentoru prof. dr. sc. Tomislavu (Dr. π) Buriću koji mi je (između ostalog) omogućio slobodu rada na temama koje mene dubinski zanimaju i koji nikad nije bio udaljen više od jedne poruke za sve moje nedoumice i upite. Ovaj rad neka smatra kao poklon za svoj rođendan koji je za nekoliko dana.

Također bi želio zahvaliti svojim roditeljima i sestri koji su sve ove godine bili uz mene, odgajali me, podržavali i poticali. Bez obzira na to koje sam odluke tijekom godina donosio, nikad nisu sumnjali da znam što radim, da ništa nije nemoguće i da ću uspjeti u svemu čega se čvrsto uhvatim. Bez njihove podrške ne znam gdje bih sada bio, ali vjerojatno ni blizu prvostupničke diplome.

Još jednu posebnu zahvalu zaslužuju moji cimeri (Pavle Jugović i Ivan Kovačić) i docimeri (previše za nabrojati) u Studentskom domu Stjepan Radić koji su me ovih nekoliko proteklih godina podnosili. Provodili sa mnom u dobrom i u lošem duge neprospavane noći te time stvorili sjećanja koja nikad neću zaboraviti.

Hvala i svim drugima koji su me vodili, pomagali i oblikovali u osobu koju sam danas. Koliko god možda sitan, svi ste imali utjecaj te ste stoga dio ovog mog malog uspjeha.

Sadržaj

Uvod	1
1. Računalni vid.....	3
1.1. Povijest	6
1.2. Osnovne tehnike i zadaci	8
1.3. Primjene i trendovi	9
2. Prepoznavanje uzoraka	11
2.1. Treniranje i testiranje modela	13
2.2. Algoritmi	14
2.3. Raspoznavanje objekata	16
3. Praćenje pogleda.....	32
3.1. Povijest	32
3.2. Tehnike praćenja.....	34
3.3. Vizualizacija podataka.....	35
3.4. Primjene i trendovi	37
4. Programska implementacija	41
4.1. Primjer korištenja	42
Zaključak	47
Literatura	49

Uvod

U svakodnevnom životu koristimo vid kako bismo promatrali okolinu i prikupljali relevantne informacije. Iako nam se taj proces čini jednostavnim, mehanizmi koji omogućuju vid iznimno su složeni. Vid kod ljudi i drugih životinja temelji se na procesu detekcije i interpretacije svjetlosnih uzoraka iz okoline. Ovaj proces uključuje oči, koje prikupljaju vizualne informacije, te mozak, koji te informacije obrađuje i stvara slike koje prepoznajemo.

Računalni vid se bazira na složenim algoritmima i modelima umjetne inteligencije koji pokušavaju simulirati ljudsku vizualnu obradu, ali na način koji ne replicira biološku složenost našeg vida. On pokušava simulirati taj proces pomoću softvera i hardvera koji analiziraju i obrađuju vizualne podatke. Ključna karakteristika računalnog vida je potreba za velikim količinama podataka, koje sustav koristi kako bi prepoznao različite vizualne elemente i objekte. To uključuje sve korake, od prikupljanja i prijenosa podataka do njihove obrade, filtriranja i pohrane, kako bi se steklo razumijevanje vanjskog svijeta i prikupile relevantne informacije. Iako se računalni vid razvija sve brže, njegove metode i dalje su bitno različite od onih koje koristi ljudski vid.

Praćenje pogleda (en. gaze tracking) je tehnika koja se temelji na računalnom vidu i omogućava mjerenje točke na koju osoba usmjerava pogled ili praćenje pokreta oka u odnosu na glavu. Ova tehnologija se sve više koristi u različitim aplikacijama, poput asistivnih tehnologija za osobe s invaliditetom, gdje omogućava upravljanje uređajima poput invalidskih kolica ili robotskih ruku. Također, praćenje pogleda se istražuje kao alat za ranu dijagnostiku poremećaja poput autizma. Najčešći način mjerenja pokreta oka je putem video tehnologije koja detektira položaj oka.

Cilj ovog rada je istražiti ključne aspekte i primjene računalnog vida, s posebnim naglaskom na prepoznavanje objekata i praćenje pogleda.

Rad obuhvaća analizu povijesti razvoja računalnog vida, uz pregled temeljnih tehnika i algoritama koji se koriste u prepoznavanju uzoraka i objekata. Također, istražuju se suvremene metode praćenja pogleda, uključujući različite tehnike i alate za vizualizaciju

podataka dobivenih praćenjem pogleda. Poseban fokus stavljen je na izazove i napredak u tehnologijama dubokog učenja, kao i na etička pitanja koja proizlaze iz njihove primjene.

Također za cilj želi pružiti cjelovit uvid u trenutno stanje tehnologije računalnog vida, te identificirati ključne smjernice za buduća istraživanja i razvoj u ovom području, a za kraj će se dati kratki pregled u implementaciju jednog sustava za praćenje pogleda koristeći biblioteku Webgazer.js.

1. Računalni vid

Računalni vid (en. computer vision) je interdisciplinarno područje koje se bavi razvojem metoda pomoću kojih računala mogu steći visokorazinsko razumijevanje digitalnih slika ili videozapisa. S inženjerskog gledišta, cilj računalnog vida je automatizirati zadatke koje ljudski vizualni sustav prirodno obavlja.

Računalni vid kroz svoje zadatke obuhvaća razne metode za prikupljanje, obradu, analizu i razumijevanje digitalnih slika te izvlačenje visokodimenzionalnih podataka iz okoline kako bi se proizvela numerička ili simbolička informacija.

Znanstvena disciplina računalnog vida se specifično bavi teorijom koja stoji iza umjetnih sustava za izvlačenje informacija iz slika. Ti podaci dolaze u različitim oblicima, kao što su najobičnije slike, videozapisi, višedimenzionalni podaci dobiveni 3D skeniranjem, 3D oblaci točaka dobiveni LiDAR sensorima kao što to koriste iPhone-i i Teslini modeli auta ili pak medicinski skenovi. Tehnološka disciplina računalnog vida pokušava nadograditi znanje znanstvene discipline tako da primjeni te modele i razne teorije u izgradnji sustava za računalni vid.

Područja rada i primjene računalnog vida koje će se proći u daljnjim poglavljima uključuju rekonstrukciju scena, detekciju i raspoznavanje objekata i događaja, procjenu 3D položaja i gibanja te restauraciju slika.

Najbliže povezana područja računalnog vida su analiza i obrada slika (en. image analysis and processing), i strojni vid (en. machine vision).

Obrada slika i analiza slika uglavnom se fokusiraju na 2D slike te kako ih transformirati. Primjeri transformacija uključuju operacije kao što su poboljšanje kontrasta, označivanje i izdvajanje rubova ili uklanjanje šuma na slici, ili pak jednostavne geometrijske transformacije poput rotacije slike.

Strojni vid, podskup računalnog vida, odnosi se na primjenu niza tehnologija i metoda za automatsku inspekciju temeljenu na slikama, kontrolu procesa i vođenje robota u industrijskim primjenama. Strojni se vid obično fokusira na primjene u proizvodnji, npr. sustavi za vizualnu inspekciju, mjerenje ili uzimanje objekata (en. bin picking) i sl..

Postoji značajno preklapanje u ovim područjima. U obradi slika, ulaz je slika, a izlaz je također slika, dok u računalnom vidu, ulaz može biti slika ili videozapis, a izlaz može biti poboljšana slika, razumijevanje sadržaja slike ili čak ponašanje računalnog sustava temeljeno na tom razumijevanju.

Još jedna česta primjena računalnog vida postoji u prepoznavanju uzoraka (en. pattern recognition), područja koje koristeći razne metode za izvlačenje informacija iz raznih tipova signala, uglavnom temeljeno na statističkim pristupima i umjetnim neuronskim mrežama. Prepoznavanje uzoraka se također uvelike koristi u primjeni sličnih metoda i modela na slikovne podatke i njihovu obradu.

U mnogim primjenama mjeri se samo jedna vrijednost na svakoj točki uzorkovanja, koja predstavlja svjetlinu slike na toj točki; ta vrijednost naziva se razina sive (en. gray level) piksela. U primjenama koje zahtijevaju informacije o boji, mjeri se nekoliko vrijednosti na svakoj točki uzorkovanja, koje predstavljaju mjerenja svjetline u nizu spektralnih traka; tako svaki piksel ima k -tuple vrijednosti.

U računalnom vidu, ulaz je slika (ili više slika), a izlaz se sastoji od informacija o sceni koja je proizvela tu sliku. Na primjer, izlaz može biti neka vrsta opisa scene.

Procesi koji se koriste za detekciju dijelova u slici poznati su kao tehnike segmentacije ili detekcije značajki (en. feature detection). Detekcija značajki identificira posebne vrste lokalnih obrazaca u slici, primjerice, rubove gdje se razina sive naglo mijenja, linije ili krivulje, točke, kutove itd. Takvi obrasci mogu se detektirati uspoređivanjem slike s predlošcima (u svakoj poziciji i orijentaciji) ili izračunavanjem različitih kombinacija prvih i drugih derivacija u svakoj točki slike.

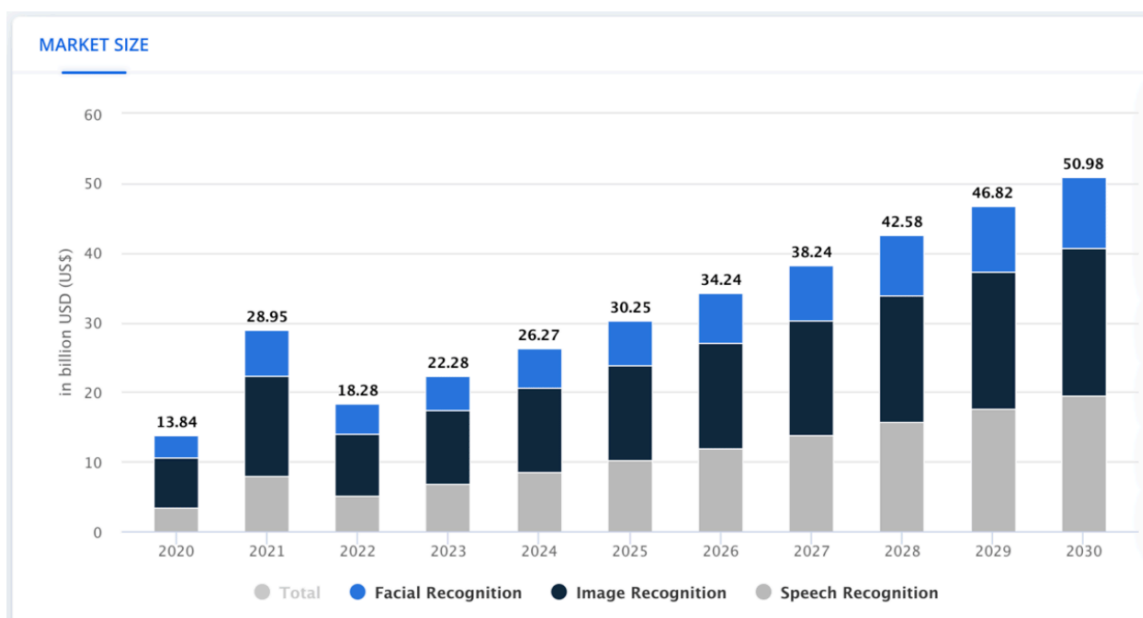
Segmentacija identificira prepoznatljive podskupine piksela ili dijeli sliku u povezane regije koje su "homogene" u nekom smislu. Procesi segmentacije i detekcije značajki mogu se smatrati procesima dodjeljivanja oznaka pikselima slike, što ukazuje na posebne klase kojima pikseli pripadaju. Tako se izlaz segmentacijskog procesa može predstaviti kao "simbolička slika" u kojoj su "vrijednosti" piksela oznake, a ne razine sive.

Neka poboljšanja u karakterizaciji objekata mogu se postići definiranjem objekata hijerarhijski, kao sastavljenih od dijelova koji su opet sastavljeni od poddijelova, itd. sve

do razine piksela, ako je potrebno. To je pristup koji se koristi u sintaktičkom prepoznavanju uzoraka (en. syntactic pattern recognition), gdje se "gramatička" pravila koriste za "parsiranje" slike prepoznavanjem konfiguracija primitivnih dijelova koji zadovoljavaju zadane uvjete.

Drugi smjer poboljšanja temelji se na promatranju da prirodne klase objekata nisu jasno definirane; one se moraju definirati nejasno ili probabilistički. Na primjer, možemo specificirati vjerojatnosne gustoće vrijednosti svojstava i odnosa za zadanu klasu objekta. Tada možemo koristiti Bayesov teorem za procjenu vjerojatnosti da promatrani skup vrijednosti potječe od objekta koji pripada zadanoj klasi. To je pristup koji se koristi u statističkom prepoznavanju uzoraka (en. statistical pattern recognition).

Razumijevanje računalnog vida također je važno i zbog njegovog ogromnog tržišnog potencijala. Ovo područje bilježi značajan rast, s projekcijama koje se donekle razlikuju među različitim izvorima. Prema istraživanju Grand View Research veličina tržišta iznosila je 14,10 milijardi dolara u 2022., a do 2030. godine bi trebala dosegnuti 58,29 milijardi dolara. Prema Allied Market Research u 2022. tržišna vrijednost iznosila je 15 milijardi dolara, a do 2032. Očekuju iznos od 82,1 milijardi dolara, dok prema Statisti tržišna veličina trebala bi dosegnuti 50,97 milijardi dolara do 2030.



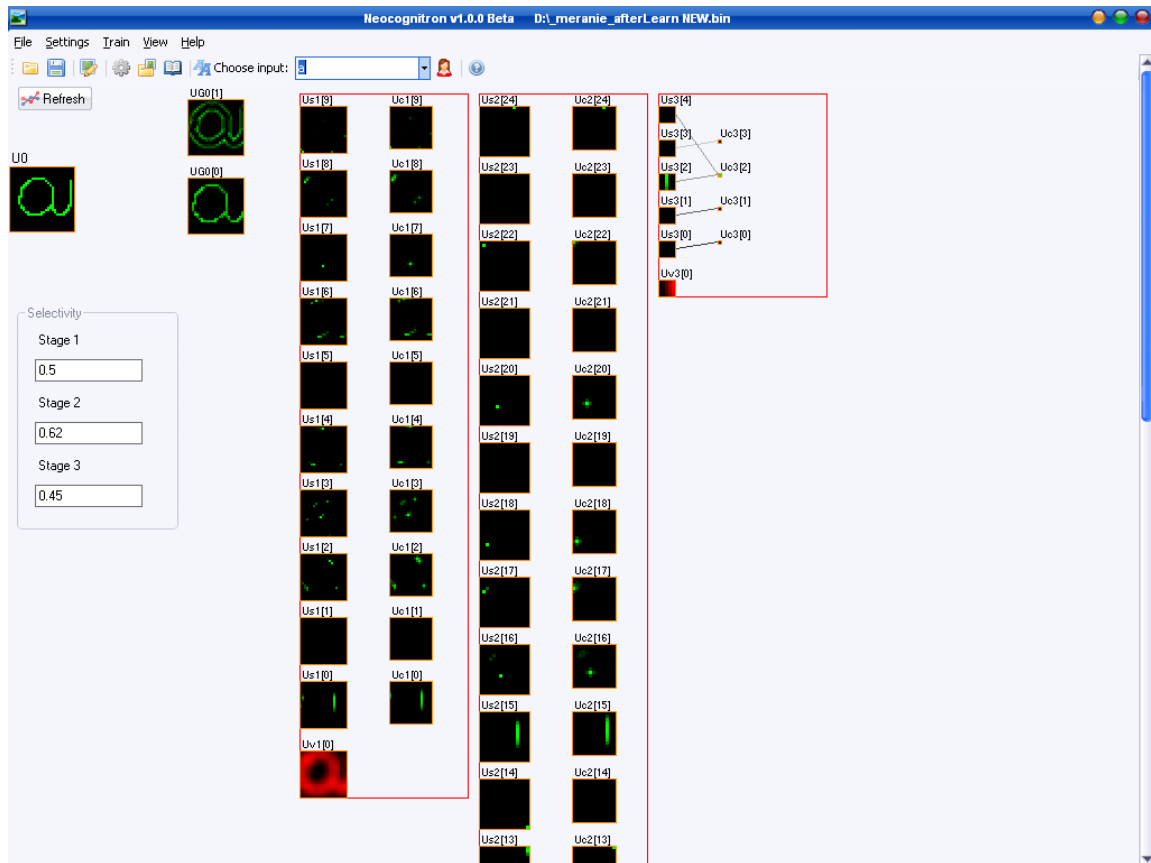
Slika 1.1 Projekcije veličine tržišta za računalni vid prema Statisti iz kolovoza 2023.

1.1. Povijest

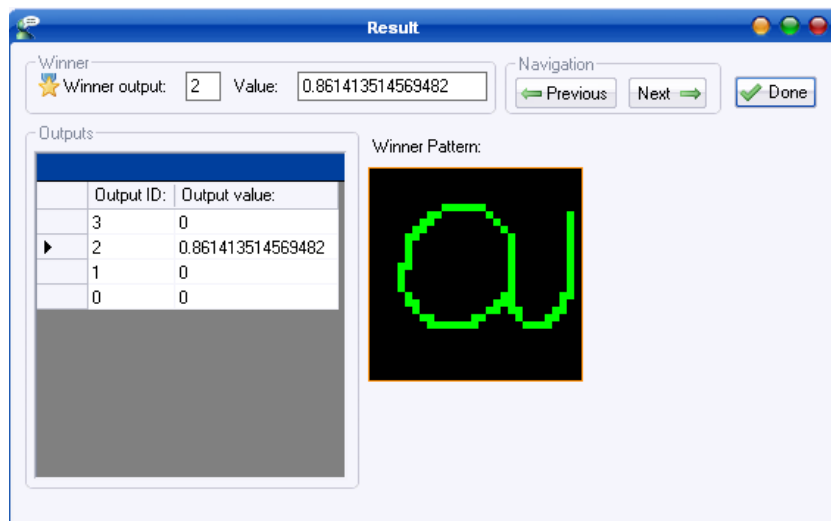
Računalni vid počeo se razvijati krajem 1960-ih godina na sveučilištima koja su bila pioniri u istraživanju umjetne inteligencije. U početku, cilj je bio oponašati ljudski vizualni sustav kao korak prema osposobljavanju robota za inteligentno ponašanje. Godine 1966. postojalo je uvjerenje da se to može postići kroz ljetni studentski projekt, jednostavno spajanjem kamere na računalo koje bi "opisivalo ono što vidi".

Sljedeće desetljeće donijelo je istraživanja temeljena na matematičkoj analizi i kvantitativnim stranama računalnog vida. Ta istraživanja uključivala su koncept prostora skale (en. scale-space), izvođenje oblika iz različitih indikatora poput sjenčanja, teksture i fokusa, te modele kontura poznate kao "zmije" (en. snakes). Istraživači ovog područja su ubrzo shvatili da se mnogi od ovih matematičkih koncepata mogu tretirati unutar istog okvira optimizacije kao što su regularizacija i Markovljeva slučajna polja (en. Markov random fields).

Neurobiologija je značajno utjecala na razvoj računalnog vida, posebno na algoritme. Tijekom 19. stoljeća provodila su se detaljna istraživanja struktura oka, neurona i ostalih sustava koji su zaduženi za obradu vizualnih informacija kod životinja. Ova su saznanja potaknula razvoj posebnog područja računalnog vida, u kojem su umjetni sustavi konstruirani kako bi imitirali biološke procese i ponašanja na različitim razinama složenosti. Također, brojne metode učenja, uključujući analizu slika i klasifikaciju putem neuronskih mreža i dubokog učenja, svoje temelje vuku iz istraživanja u neurobiologiji. Neocognitron, hijerarhijska, višeslojna umjetna neuronska mreža koju je 1979. godine predložio Kunihiro Fukushima, korištena je za prepoznavanje japanskih rukom pisanih znakova i druge zadatke prepoznavanja uzoraka, te je poslužila kao inspiracija za konvolucijske neuronske mreže (en. convolutional neural networks).



Slika 1.2 Primjer rada Neocognitrona



Slika 1.3 Primjer odabira odgovora Neocognitrona

S dolaskom metoda optimizacije za kalibraciju kamera, uočeno je da su mnoge ideje već bile istražene u teoriji prilagodbe skupa podataka (en. bundle adjustment) iz područja fotogrametrije. Ovo razdoblje također obilježava prvo praktično korištenje statističkih metoda učenja za prepoznavanje lica na slikama, što je vidljivo u konceptu Eigenface. Krajem 1990-ih došlo je do značajne promjene zbog sve veće interakcije između područja

računalne grafike i računalnog vida. To je uključivalo tehnike kao što su renderiranje temeljeno na slikama, morfiranje slika, interpolacija pogleda, spajanje panoramskih slika i rana renderiranja svjetlosnog polja (en. light-field rendering).

U posljednje vrijeme metode koje se temelje na prepoznavanju značajki ponovno su postale važne, osobito kada se kombiniraju s tehnikama strojnog učenja i naprednim optimizacijskim metodama. Razvoj dubokog učenja (engl. Deep Learning) značajno je unaprijedio područje računalnog vida. Algoritmi dubokog učenja sada postižu izvanredne rezultate na nekoliko referentnih skupova podataka za zadatke poput klasifikacije, segmentacije i praćenja optičkog toka, pri čemu su nadmašili prethodne metode.

1.2. Osnovne tehnike i zadaci

- **Raspoznavanje**

Raspoznavanje je klasičan problem u računalnom vidu, obradi slika i strojnom vidu, a odnosi se na prepoznavanje specifičnih objekata, značajki ili aktivnosti u slikovnim podacima. Prepoznavanje objekata (en. object recognition) uključuje identifikaciju unaprijed definiranih ili naučenih objekata ili klasa objekata u slici, obično uz određivanje njihovih 2D položaja u slici ili 3D položaja u sceni. Identifikacija se odnosi na prepoznavanje pojedinačne instance objekta, dok se detekcija odnosi na pretraživanje slike za specifične objekte i njihove lokacije.

Najbolji algoritmi za ove zadatke trenutno su temeljeni na konvolucijskim neuronskim mrežama (en. convolutional neural networks, CNNs), koje su postigle performanse bliske ljudskima na testovima kao što je ImageNet. Ipak, ove mreže još uvijek imaju poteškoća s prepoznavanjem malih ili tankih objekata te iskrivljenih slika, dok ljudi lakše prepoznaju takve elemente.

Prepoznavanje u računalnom vidu obuhvaća nekoliko specijaliziranih zadataka. Na primjer, pretraživanje slika prema sadržaju (en. content-based image retrieval) omogućuje pronalazak slika s određenim sadržajem unutar veće zbirke slika. Procjena položaja (en.

pose estimation) odnosi se na procjenu položaja i orijentacije objekta, što je korisno u industrijskim primjenama poput asistencije robotskoj ruci. Optičko prepoznavanje znakova (en. optical character recognition, OCR) koristi se za prepoznavanje znakova u slikama tiskanog ili rukom pisanog teksta. Prepoznavanje lica (en. facial recognition) omogućuje prepoznavanje i podudaranje lica u digitalnim slikama ili videozapisima s bazom podataka lica, dok prepoznavanje emocija (en. emotion recognition) predstavlja klasifikaciju ljudskih emocija na temelju analize lica. Prepoznavanje gesta (en. gesture recognition) odnosi se na prepoznavanje i interpretaciju ljudskog govora tijela.

- **Rekonstrukcija scene**

Rekonstrukcija scene uključuje izradu 3D modela scene na temelju jedne ili više slika ili videozapisa. Ova tehnika omogućuje detaljnu analizu i razumijevanje prostora snimljenog u slikama ili videozapisima.

- **Restauracija slike**

Restauracija slike odnosi se na proces popravljavanja izvorne slike koja je degradirana ili oštećena zbog vanjskih čimbenika poput loše pozicioniranog objektiva, smetnji u prijenosu, slabog osvjetljenja ili zamućenja uzrokovano pokretom. Ova tehnika je korisna za poboljšanje kvalitete slika kako bi bile jasnije i preciznije.

1.3. Primjene i trendovi

Jedna od najistaknutijih primjena računalnog vida jest u medicini, posebno u obradi medicinskih slika (en. medical image processing), gdje se iz slikovnih podataka izvlače informacije potrebne za dijagnosticiranje pacijenata. Primjeri takve primjene uključuju detekciju tumora, arterioskleroze ili drugih malignih promjena, kao i pružanje podrške tijekom kirurških zahvata te razvoj lijekova.

Druga važna primjena računalnog vida nalazi se u proizvodnoj industriji, gdje je poznat pod nazivom strojni vid (en. machine vision). U ovom kontekstu, računalni vid se koristi za prikupljanje informacija u svrhu podrške proizvodnim procesima. Primjer toga je kontrola kvalitete (en. quality control), gdje se automatski pregledavaju detalji ili gotovi proizvodi kako bi se otkrili eventualni nedostaci. Jedan od najvažnijih sektora u kojem se

ova tehnologija koristi jest industrija proizvodnje poluvodiča (en. semiconductor industry), gdje se svaki pojedini silicijski wafer mjeri i pregledava kako bi se spriječila proizvodnja neispravnih čipova, dok je drugi primjer mjerenje položaja i orijentacije komponenti s kojima mora interagirati robotska ruka.

Osim u proizvodnji, strojni se vid također široko koristi u poljoprivredi, primjerice u navođenju te u procesu optičkog sortiranja, gdje se uklanja neželjeni materijal iz rasutih proizvoda. Primjeri također uključuju kontrolu kvalitete i optimizaciju procesa u proizvodnji, praćenje usjeva, preciznu poljoprivredu i upravljanje stokom.

Nadalje, robotika koristi računalni vid za autonomnu navigaciju, gdje se planira putanja robota kroz određeno okruženje. Da bi robot mogao navigirati, potrebno je detaljno razumjeti okruženje, a računalni vid djeluje kao vizualni senzor koji pruža visoko kvalitetne informacije o okolini i samom robotu.

Jedno od novijih područja primjene računalnog vida nalazi se u autonomnim vozilima. Razina autonomnosti može se kretati od potpuno autonomnih vozila bez potrebe za ljudskim nadzorom, do vozila u kojima sustavi bazirani na računalnom vidu pomažu vozačima ili pilotima u određenim situacijama. Potpuno autonomna vozila koriste računalni vid za navigaciju, primjerice za simultanu lokalizaciju i mapiranje okoline (en. SLAM - Simultaneous Localization and Mapping), kao i za detekciju prepreka na cesti. Također se koriste za specifične zadatke, kao što su bespilotne letjelice koje pretražuju požare u šumama. Primjeri sustava za podršku uključuju sustave za upozoravanje na prepreke u vozilima, kamere i LiDAR senzore, te sustave za autonomno slijetanje zrakoplova. Svemirska istraživanja već koriste autonomna vozila s računalnim vidom, primjerice NASA-in Curiosity rover.

U drugim industrijama, računalni vid se koristi u maloprodaji za upravljanje zalihama, analizu ponašanja kupaca i samoposlužne blagajne. Očigledni primjeri vojne primjene uključuju detekciju neprijateljskih vojnika ili vozila te navođenje projektila. Automatska obrada podataka smanjuje složenost i omogućava fuziju informacija iz više senzora kako bi se povećala pouzdanost.

2. Prepoznavanje uzoraka

Prepoznavanje uzoraka (en. pattern recognition) je proces prepoznavanja obrazaca korištenjem raznih algoritama. Definira se kao klasifikacija podataka na temelju već stečenog znanja ili statističkih informacija izvučenih iz obrazaca i/ili njihove reprezentacije. Prepoznavanje uzoraka ima svoje korijene u statistici i inženjerstvu, a moderne metode sve češće uključuju primjenu strojnog učenja, zahvaljujući sve većoj dostupnosti velikih količina podataka (en. big data) i povećanoj procesorskoj snazi.

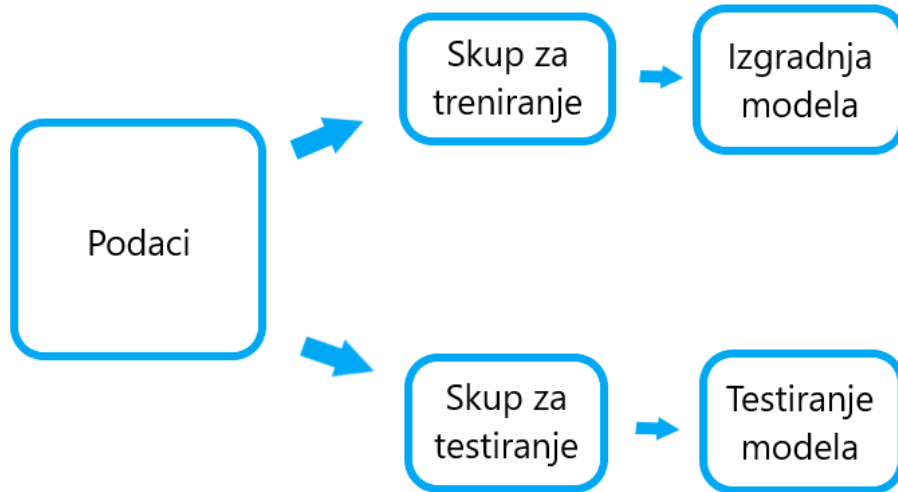
Primjene prepoznavanja uzoraka uključuju optičko prepoznavanje znakova (en. optical character recognition, OCR), statističku analizu podataka, obradu signala, analizu slika, pretraživanje informacija, bioinformatiku, kompresiju podataka, računalnu grafiku, strojno učenje, prepoznavanje govora, identifikaciju govornika i automatsku medicinsku dijagnostiku.

U tipičnoj primjeni prepoznavanja uzoraka, sirovi podaci se obrađuju i pretvaraju u oblik koji je prikladan za obradu računalom. Prepoznavanje uzoraka obuhvaća klasifikaciju i grupiranje (en. clustering) uzoraka. Klasifikacija dodjeljuje odgovarajuću klasu uzorku na temelju apstrakcije generirane iz skupa podataka za obuku ili domena znanja, dok grupiranje dijeli podatke u segmente što pomaže u donošenju odluka, čime ovaj postupak postaje osobito važan za nenadgledano učenje.

Algoritmi za prepoznavanje uzoraka obično nastoje pružiti razuman odgovor za sve moguće ulaze i izvršiti "najvjerojatnije" podudaranje unosa, uzimajući u obzir njihovu statističku varijaciju. Ovo se razlikuje od algoritama za prepoznavanje obrazaca, koji traže točne podudarnosti unosa s unaprijed definiranim obrascima, kao što je slučaj s algoritmima za podudaranje regularnih izraza (en. regular expression matching).

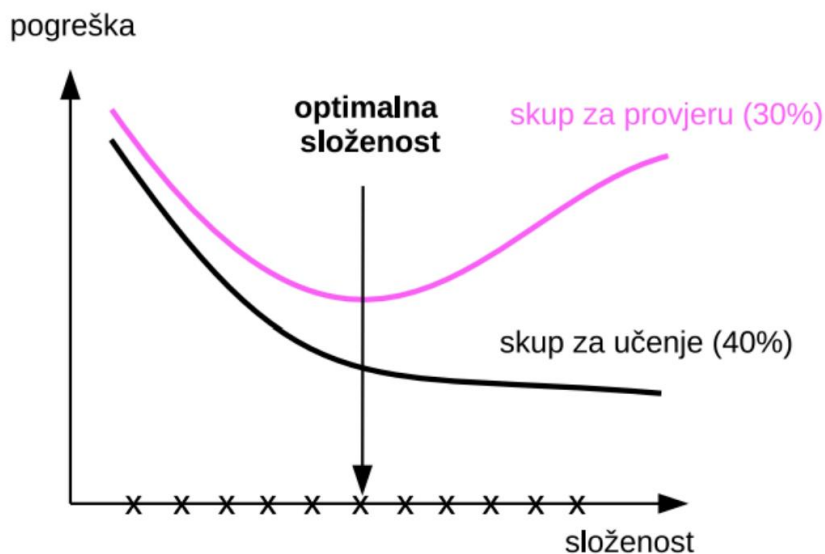
Učenje je proces kroz koji se sustav trenira i prilagođava kako bi pružio točne rezultate. Kvaliteta performansi sustava ovisi o algoritmima koji se koriste tijekom obrade podataka. Cijeli skup podataka obično se dijeli na dva dijela: skup za obuku (en. training set), koji se

koristi za treniranje modela, i skup za testiranje (en. testing set), koji se koristi za ispitivanje modela nakon obuke. Kada nisu dostupni označeni podaci, koriste se drugi algoritmi za otkrivanje prethodno nepoznatih obrazaca.



Slika 2.1 Raspodjela podataka na skupove za treniranje i testiranje

Glavni cilj u prepoznavanju uzoraka je osigurati da model dobro generalizira. Ako je model previše složen, može se vrlo dobro prilagoditi podatcima iz skupa za obuku, ali će davati loše predikcije na neviđenim podatcima, što dovodi do prenaučnosti (en. overfitting). Budući da neviđeni primjeri nisu dostupni, dio primjera iz skupa podataka obično se izdvaja da "glume" neviđene primjere kako bi se testirala sposobnost modela da generalizira.

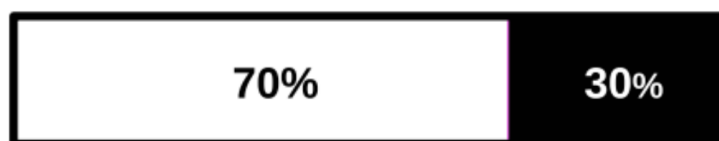


Slika 2.2 Vizualni prikaz prenaučnosti modela

2.1. Treniranje i testiranje modela

Skup za treniranje (en. training set) koristi se za izgradnju modela. Sastoji se od skupa slika ili podataka koji se koriste za obuku sustava. Pravila i algoritmi za treniranje koriste se kako bi se osigurale relevantne informacije o tome kako povezati ulazne podatke s odgovarajućim izlaznim odlukama. Općenito, 70% podataka iz skupa podataka uzima se za treniranje modela.

Skup za testiranje (en. testing set) koristi se za provjeru sustava nakon što je obučan. To je skup podataka koji se koristi za provjeru točnosti sustava, tj. za mjerenje koliko dobro sustav proizvodi točne izlaze nakon obuke. Na primjer, ako sustav točno prepozna sedam od deset kategorija cvijeća, točnost je 70%. Općenito, 30% podataka iz skupa podataka koristi se za testiranje.



Slika 2.3 Rasporedjela podataka na skupove za treniranje i testiranje u postocima

Polu-nadzirano učenje (engl. semi-supervised learning) istražuje se kroz kombinaciju treniranja i testiranja koja koristi označene i neoznačene podatke. Ova tehnika najčešće uključuje mali skup označenih podataka u kombinaciji s velikom količinom neoznačenih podataka, čime se optimizira proces učenja. Na taj način model može poboljšati svoje performanse koristeći neoznačene podatke kako bi dopunio informacije dobivene iz manjih označenih skupova podataka.

Ponekad se koriste različiti termini za opisivanje nadziranih i nenadziranih postupaka učenja za isti tip izlaza. Nenadzirana verzija klasifikacije obično se naziva grupiranje (en. clustering), što podrazumijeva grupiranje ulaznih podataka u klastere na temelju inherentne sličnosti (npr. udaljenost između instanci u višedimenzionalnom prostoru), umjesto da se svaka ulazna instanca dodjeljuje jednoj od unaprijed definiranih klasa.

Svaki podatak za koji se generira izlazna vrijednost formalno se naziva instanca. Instanca je formalno opisana vektorom značajki, koji zajedno čine opis svih poznatih karakteristika instance. Značajka je funkcija jednog ili više mjerenja, izračunata tako da kvantificira neka značajna svojstva objekta. Ovi vektori značajki definiraju točke u odgovarajućem višedimenzionalnom prostoru, a metode za manipulaciju vektorima u vektorskim prostorima mogu se primijeniti na njih, kao što su izračun skalarnih produkata ili kuta između dvaju vektora.

Značajke obično mogu biti kategorizirane (npr. spol: "muško" ili "žensko", krvna grupa: "A", "B", "AB" ili "O"), ordinalne (npr. veličine: "veliko", "srednje" ili "malo"), cjelobrojne (npr. broj pojavljivanja određene riječi u emailu) ili realne vrijednosti (npr. mjerenje krvnog tlaka).

Algoritmi za odabir značajki pokušavaju izravno eliminirati suvišne ili irelevantne značajke, ali zbog svoje složenosti, ovaj problem predstavlja optimizacijski izazov. S obzirom na n značajki, treba istražiti sve $2^n - 1$ podskupove značajki. Algoritam grananja i ograničavanja (en. Branch-and-Bound) smanjuje ovu složenost, ali je i dalje neizvodiv za srednje do velike vrijednosti n .

Algoritmi za ekstrakciju značajki pokušavaju smanjiti veliki vektor značajki na manji vektor koji je lakši za obradu i sadrži manje redundancije, koristeći matematičke tehnike kao što je analiza glavnih komponenti (en. principal components analysis, PCA).

2.2. Algoritmi

Algoritmi za prepoznavanje uzoraka ovise o vrsti izlazne oznake, o tome je li učenje nadzirano ili nenadzirano, te o tome je li algoritam statističke ili nestatističke prirode. Statistički algoritmi dodatno se mogu kategorizirati kao generativni ili diskriminativni.

Mnogi uobičajeni algoritmi za prepoznavanje uzoraka su probabilističke prirode, što znači da koriste statističku inferenciju kako bi pronašli najbolju oznaku za određenu instancu. Za razliku od drugih algoritama koji samo generiraju "najbolju" oznaku, probabilistički algoritmi često također daju vjerojatnost da instanca odgovara danoj oznaci.

Probabilistički algoritmi imaju mnoge prednosti u odnosu na ne-probabilističke algoritme. Oni pružaju vrijednost pouzdanosti povezanu sa svojim odabirom, mogu se suzdržati od davanja rezultata kada je pouzdanost previše niska, te se mogu učinkovitije uključiti u složenije zadatke strojnog učenja, na način koji djelomično ili potpuno izbjegava problem propagacije pogrešaka.

Prvi klasifikator uzoraka, linearni diskriminant koji je predstavio Fisher, razvijen je u tradiciji frekventističke statistike. S druge strane, Bayesova statistika vuče korijene iz grčke filozofije, gdje je već napravljena razlika između apriori i aposteriori znanja. Kasnije je Kant definirao razliku između onoga što se zna apriori – prije promatranja – i empirijskog znanja stečenog promatranjem. Iskustvo kvantificirano kao apriori vrijednosti parametara može se kombinirati s empirijskim opažanjima koristeći distribucije poput Beta Dirichlet.

Drvo odluke (en. decision tree) je hijerarhijski model potpore odlučivanju koji koristi stablo odluka i njihovih mogućih posljedica, uključujući ishode slučajnih događaja, troškove resursa i korisnost. U statistici, naivni Bayesovi klasifikatori (en. naive Bayes classifiers) su obitelj linearnih "probabilističkih klasifikatora" koji pretpostavljaju da su značajke uvjetno neovisne s obzirom na ciljnu klasu. Naivni Bayesovi klasifikatori su visoko skalabilni, zahtijevaju broj parametara linearan s brojem varijabli (značajki) u problemu učenja.

Neuronske mreže (engl. neural networks) sastoje se od povezanih jedinica, poznatih kao umjetni neuroni, koji su labavo modelirani prema biološkim neuronima. Svaki neuron prima signale od drugih povezanih neurona, obrađuje te signale i šalje rezultat dalje. Ovi signali su realni brojevi, a izlaz svakog neurona određuje se pomoću nelinearne funkcije ulaza poznate kao aktivacijska funkcija. Snaga signala, odnosno važnost svakog signala, određuje se težinama koje se prilagođavaju tijekom procesa učenja. Neuroni su obično organizirani u slojeve, s različitim slojevima koji provode različite transformacije ulaznih podataka.

Signal se prenosi od ulaznog sloja do izlaznog sloja, često prolazeći kroz jedan ili više skrivenih slojeva. Mreža se naziva "dubokom neuronskom mrežom" (engl. deep neural network) ako sadrži najmanje dva skrivena sloja. Ova dubina omogućava mreži da modelira složenije odnose u podacima, jer svaki sloj izvodi apstraktniju transformaciju ulaznih podataka. Ova struktura i prilagodba težina ključne su za sposobnost neuronskih mreža da "nauče" složene uzorke i predikcije na temelju podataka.

Rekurentne neuronske mreže (en. Recurrent Neural Networks, RNNs) su klasa umjetnih neuronskih mreža za obradu sekvencijalnih podataka. Za razliku od feed-forward neuronskih mreža, koje obrađuju podatke u jednom prolazu, RNN-ovi obrađuju podatke tijekom više vremenskih koraka. Osnovni gradivni blok RNN-a je rekurentna jedinica, koja održava skriveno stanje, odnosno oblik memorije, koje se ažurira pri svakom vremenskom koraku na temelju trenutnog ulaza i prethodnog skrivenog stanja. Ova povratna veza omogućuje mreži da uči iz prošlih ulaza i koristi to znanje u trenutnoj obradi.

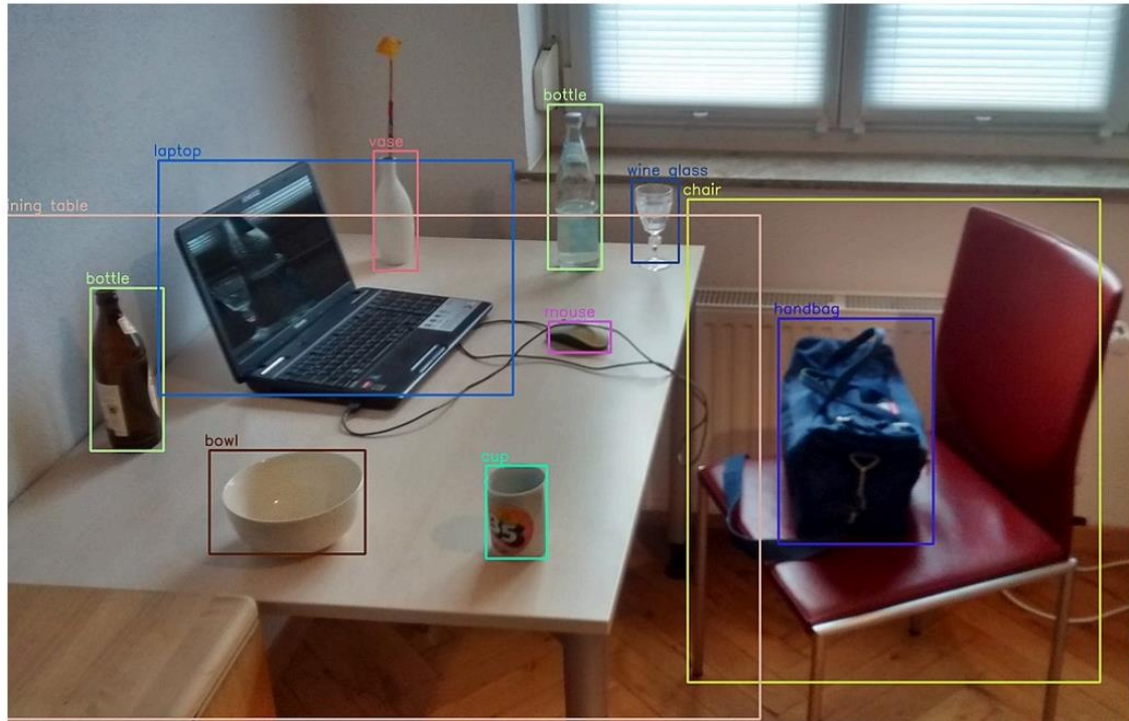
MEMM (Maximum Entropy Markov Model) ili uvjetni Markovljev model (Conditional Markov Model, CMM) je grafički model za označavanje sekvenci koji kombinira značajke skrivenih Markovljevih modela (HMMs) i modela maksimalne entropije (MaxEnt). Kao diskriminativni model, MEMM proširuje standardni model maksimalne entropije tako što povezuje nepoznate varijable u Markovljev lanac, omogućujući modeliranje sekvencijalnih podataka, umjesto tretiranja podataka kao neovisnih. Ova kombinacija se često koristi u područjima kao što su prepoznavanje govora i obrada prirodnog jezika.

2.3. Raspoznavanje objekata

Raspoznavanje objekata specifičan je podskup prepoznavanja uzoraka koji se fokusira na identificiranje i klasifikaciju objekata unutar slika ili videozapisa. Ova tehnika uključuje prepoznavanje specifičnih instanci objekata (poput automobila, ljudi, životinja) i određivanje njihovih lokacija u vizualnoj sceni. Prepoznavanje objekata najčešće se koristi u scenarijima gdje je vizualni kontekst ključan, poput tražilica slika, autonomnih vozila i sustava za prepoznavanje lica. Također se koristi za praćenje objekata, primjerice praćenje lopte tijekom nogometne utakmice, kretanja kriket palice ili praćenje osobe u videozapisu.

Većina suvremenih metoda prepoznavanja objekata oslanja se na tehnike dubokog učenja (en. deep learning) kako bi poboljšale svoje performanse. Duboko učenje odnosi se na klasu algoritama strojnog učenja u kojima se hijerarhija slojeva koristi za transformaciju ulaznih podataka u apstraktniju i složeniju reprezentaciju. Na primjer, u modelu za prepoznavanje slika, sirovi ulaz može biti slika. Prvi sloj može pokušati identificirati osnovne oblike poput linija i krugova, drugi sloj može sastaviti i kodirati raspored rubova, treći sloj može kodirati nos i oči, dok četvrti sloj prepoznaje da slika sadrži lice.

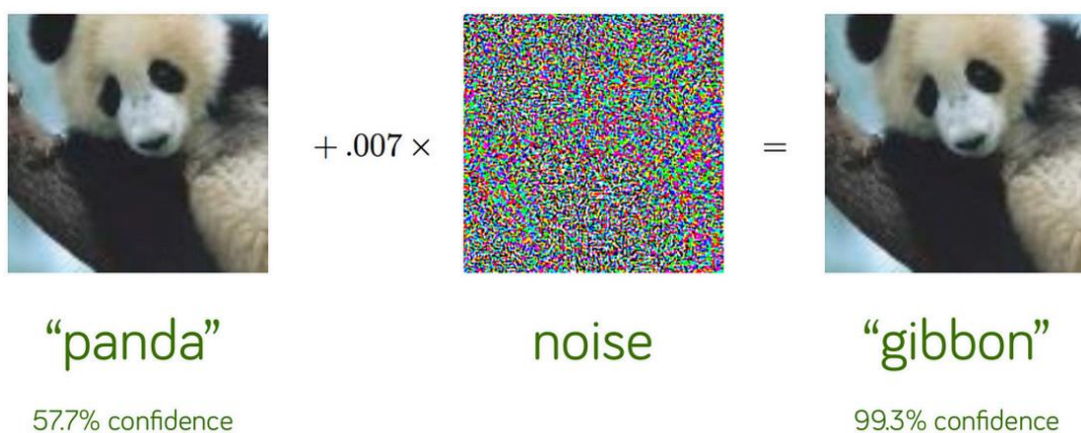
Ključna prednost dubokog učenja je u tome što postupak može samostalno naučiti koje značajke treba optimalno postaviti na kojoj razini. Prije razvoja dubokog učenja, tehnike strojnog učenja često su uključivale ručno osmišljavanje značajki kako bi se podaci transformirali u prikladniju reprezentaciju za rad klasifikacijskog algoritma. U pristupu dubokog učenja, značajke nisu ručno izrađene, već model automatski otkriva korisne reprezentacije značajki iz podataka. Ipak, ovo ne eliminira potrebu za ručnim podešavanjem; na primjer, variranje broja slojeva i veličina slojeva može pružiti različite stupnjeve apstrakcije.



Slika 2.4 Primjer raspoznavanja objekata koristeći YOLO algoritam

Kako se duboko učenje, koje se koristi u mnogim algoritmima unutar računalnog vida, prepoznavanja objekata i lica te praćenja pogleda, sve više primjenjuje izvan laboratorija u

stvarnom svijetu, istraživanja i iskustva pokazuju da su umjetne neuronske mreže (en. artificial neural networks, ANN) podložne hakiranju i obmanama. Identificiranjem obrazaca koje ovi sustavi koriste za svoje funkcioniranje, napadači mogu modificirati ulazne podatke u mreži na takav način da mreža prepozna podudarnost koju ljudski promatrači ne bi prepoznali. Primjerice, napadač može napraviti suptilne promjene na slici tako da neuronska mreža prepozna podudarnost, iako slika ljudskom oku izgleda potpuno različito od traženog cilja. Takva manipulacija naziva se adversarijalni napad (en. adversarial attack), a kroz njega neovlaštene osobe mogu koristiti umjetnu inteligenciju za lažno predstavljanje ili druge zlonamjerne svrhe.



Slika 2.5 Primjer adversarijalnog napada

• Povijest

Frank Rosenblatt je 1958. godine predložio višeslojni perceptron (en. multilayer perceptron, MLP) s tri sloja: ulaznim slojem, skrivenim slojem s nasumičnim težinama koje se nisu učile, i izlaznim slojem. Neki smatraju da je knjiga iz 1962. godine, koja je razvila i istražila sve osnovne elemente današnjih sustava dubokog učenja, postavila temelje ovog područja.

Metoda grupnog upravljanja podacima, način treniranja dubokih neuronskih mreža, objavljena je 1967. godine od strane Alexeya Ivakhnenka i Laphe, koji su ovu metodu smatrali generalizacijom Rosenblattovog perceptrona.

Termin "back-propagating error correction" (propagacija pogreške unatrag) uveo je Frank Rosenblatt 1962. godine, ali nije znao kako ga implementirati. Ipak, on je proučavao samo neurone čiji su izlazi bili diskretni nivoi, što je činilo propagaciju pogreške unatrag

nemogućim. Godine 1982., Paul Werbos primijenio je propagaciju pogreške unatrag na MLP-ove na način koji je postao standardan, a David E. Rumelhart neovisno je razvio i popularizirao ovaj algoritam u svojoj istraživačkoj zajednici.

Duboke arhitekture za konvolucijske neuronske mreže (en. convolutional neural networks, CNNs) s konvolucijskim slojevima i slojevima za smanjivanje započele su s Neocognitronom, kojeg je 1980. godine predstavio Kuniyuki Fukushima, iako nije bio treniran pomoću propagacije pogreške unatrag. On je 1969. godine također uveo ReLU (en. rectified linear unit) aktivacijsku funkciju, koja je postala najpopularnija aktivacijska funkcija za CNN-ove i duboko učenje općenito.

Pojam "duboko učenje" (en. deep learning) uveden je u zajednicu strojnog učenja od strane Rina Dechter 1986. godine, dok je Igor Aizenberg s kolegama 2000. godine uveo taj termin u kontekstu umjetnih neuronskih mreža za Booleove neuronske pragove.

Rekurentne neuronske mreže (en. Recurrent Neural Networks, RNNs) razvijene su 1980-ih godina za obradu sekvencijalnih podataka. Sepp Hochreiter je u svom diplomskom radu iz 1991. godine predložio neuralni kompresor povijesti i identificirao problem nestajanja gradijenta, što je dovelo do razvoja dugoročnih kratkotrajnih memorija (en. Long Short-Term Memory, LSTM) 1995. godine.

Revolucija dubokog učenja započela je oko CNN-ova i računalnog vida temeljenog na GPU-ima. Iako su CNN-ovi trenirani pomoću propagacije pogreške unatrag postojali desetljećima, brže implementacije na GPU-ima bile su ključne za napredak u računalnom vidu. U 2012. godini, AlexNet, kojeg su razvili Alex Krizhevsky, Ilya Sutskever i Geoffrey Hinton, pobijedio je na velikom natjecanju ImageNet, značajno nadmašivši pliće metode strojnog učenja.

Generativne suparničke mreže (en. Generative Adversarial Networks, GANs), koje je 2014. godine uveo Ian Goodfellow i suradnici, postale su vodeće u generativnom modeliranju, a Nvidia-ov StyleGAN (2018) postigao je izvanrednu kvalitetu slike. Ipak,

difuzijski modeli (en. diffusion models), poput DALL-E 2 (2022), Stable Diffusion (2022), Midjourney (2022) i Grok (2024), nadmašili su GAN-ove u generativnom modeliranju.

- **Izazovi**

Treniranje modela dubokog učenja zahtijeva veliku količinu slika kako bi se postigle konkurentne performanse. Međutim, prikupljanje slika često je skup i izazovan proces. Kako bi se ovaj problem ublažio, potvrđeno je da je augmentacija slika (en. image augmentation) učinkovita strategija. Razumijevanje postojećih metoda augmentacije slika ključno je za implementaciju odgovarajućih algoritama.

Jedan od glavnih izazova u prepoznavanju objekata su varijacije u slikama, koje mogu biti uzrokovane različitim uvjetima osvjetljenja ili deformacijama objekata. Još jedan izazov je neravnoteža klasa, gdje se različiti objekti pojavljuju s različitim učestalostima. Na primjer, u medicinskim slikama abnormalni slučajevi često se pojavljuju s niskom vjerojatnošću, a ovaj problem dodatno je pojačan pitanjima privatnosti. Kada se model trenira na neuravnoteženom skupu podataka, sklona je veća vjerojatnost da će normalne slučajeve prepoznati s većom sigurnošću. Neravnoteža klasa također može rezultirati u situaciji gdje ima malo slika iz više klasa, ali mnogo slika iz jedne klase.

Dodatni izazov predstavlja domenska razlika (en. domain shift), gdje skupovi podataka za treniranje i testiranje pokazuju različite distribucije. To je posebno uočljivo u kontekstu automatske vožnje, gdje su slike snimljene danju i noću vrlo različite. Često se modeli treniraju na slikama snimljenim tijekom dana, ali se očekuje da će se koristiti i noću, što predstavlja izazov. Novi problem uveden dubokim učenjem je memoriranje podataka. Općenito, veći broj parametara koji se mogu naučiti zahtijeva više podataka za treniranje, što se naziva strukturni rizik. Povećanjem broja parametara, model dubokog učenja može zapamtiti specifične podatke kada je broj slika za treniranje nedovoljan, što dovodi do problema s generalizacijom u obliku prenaučivosti.

Augmentacija slika koristi se u istraživanjima koja se bave klasifikacijom slika (gornji dio tablice) i detekcijom objekata (donji dio tablice). Primjeri različitih metoda augmentacije koje koriste neki od poznatih modela prikazani su u tablici 2.1.

Tablica 2.1 Metode augmentacije slika korišteni pri treniranju modela

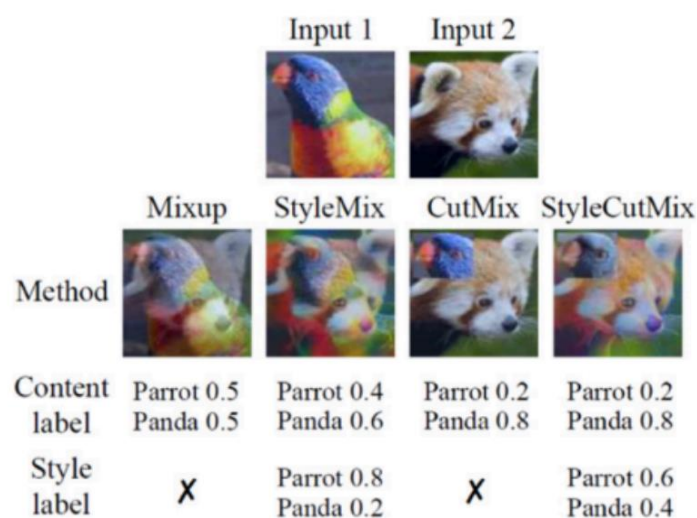
Paper	Image Augmentation Method
AlexNet	Translate, Flip, Intensity Changing
ResNet	Crop, Flip
DenseNet	Flip, Crop, Translate
MobileNet	Crop, Elastic distortion
NasNet	Cutout, Crop, Flip
ResNeSt	AutoAugment, Mixup, Crop
DeiT	AutoAugment, RandAugment, Random Erasing, Mixup, CutMix
Swin Transformer	RandAugment, Mixup, CutMix, Random Erasing
Faster R-CNN	Flip
YOLO	Scale, Translate, Color space
SSD	Crop, Resize, Flip, Color Space, Distortion
YOLOv4	Distortion, Scale, Color space, Crop, Flip, Rotate, Random erase, Cutout, Hide-and-Seek, GridMask, Mixup, CutMix, StyleGAN

Metode obrade slika, poput geometrijskih transformacija i manipulacije na razini piksela, mogu se iskoristiti za augmentaciju slika. Geometrijska transformacija nastoji promijeniti prostorni odnos između piksela, uključujući i elastične deformacije, dok se obrada slika u boji fokusira na promjenu boje ulazne slike. Posljednja metoda, koja uključuje promjenu dijelova slike, sve više privlači pažnju istraživača. Ostale metode uključuju generiranje prilagođenih slika i miješanje slika.

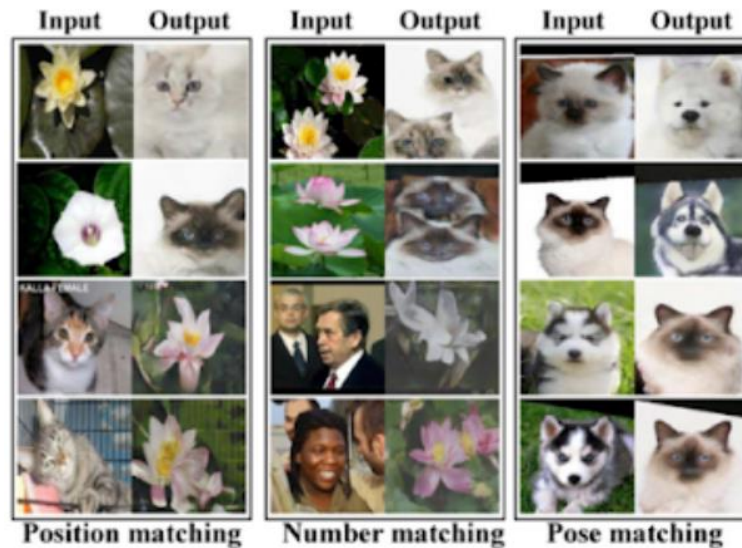


Slika 2.6 Primjeri geometrijskih transformacija slika

Skupovi podataka smatraju se ključnima za postizanje zadovoljavajućih performansi. Jedan način za generiranje prikladnog skupa podataka je putem algoritama za augmentaciju slika, koji su pokazali impresivne rezultate na različitim skupovima podataka i heterogenim modelima. Trenutno se široko koriste politike koje nisu vezane uz model, kao i one optimizacijske, dok je pristup temeljen na modelu aktivno istraživačko područje za specifične izazove, poput neravnoteže klasa i prilagodbe domeni. Iako većina algoritama čuva oznake, algoritmi koji mijenjaju oznake također su nedavno privukli pažnju, dok razvoj novih algoritama ostaje ključan za poboljšanje performansi dubokog učenja.



Slika 2.7 Augmentacija slika koristeći algoritme

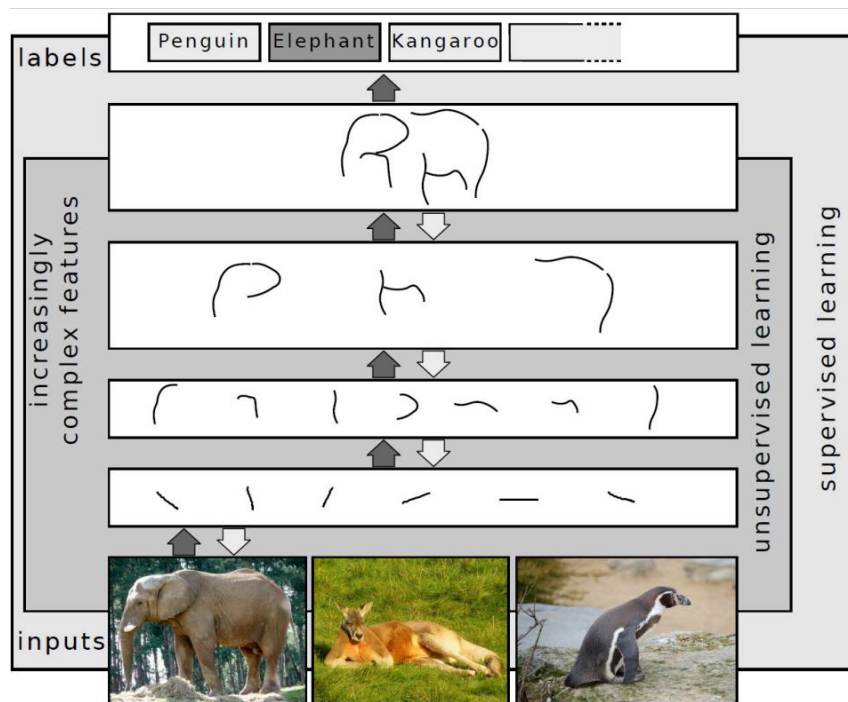


Slika 2.8 Augmentacija slika koristeći generativne modele

Prepoznavanje slika temeljeno na dubokom učenju u posljednjih je nekoliko godina postalo "nadjudsko", postigavši preciznije rezultate od ljudskih natjecatelja. Prvi put se to dogodilo 2011. godine u prepoznavanju prometnih znakova, a zatim 2014. u prepoznavanju ljudskih lica. Prepoznavanje objekata često koristi pristupe dubokog učenja, osobito konvolucijske neuronske mreže, koje su dizajnirane za analizu vizualnih podataka. Ove tehnike dovele su do izvanrednog napretka posljednjih godina, a često uključuju i tehnike poput regresije ograničavajućih okvira (en. bounding box regression) za lociranje objekata unutar slika.

• Algoritmi

Algoritmi za detekciju objekata obično se dijele na pristupe temeljene na neuronskim mrežama i one koji nisu temeljeni na njima. U pristupima koji nisu temeljeni na neuronskim mrežama, prvo je potrebno definirati značajke pomoću jedne od metoda poput onih navedenih u nastavku, a zatim koristiti tehniku poput stroja potpornih vektora (en. support vector machine, SVM) za klasifikaciju. S druge strane, pristupi temeljeni na neuronskim mrežama mogu provoditi detekciju objekata od početka do kraja bez potrebe za specifičnim definiranjem značajki, a obično se temelje na konvolucijskim neuronskim mrežama, kao što je već raspravljano u poglavlju o prepoznavanju uzoraka.



Slika 2.9 Primjer funkcioniranja konvolucijske neuronske mreže

Pristupi temeljeni na neuronskim mrežama uključuju modele kao što su Region Proposals (R-CNN, Fast R-CNN, Faster R-CNN, cascade R-CNN), YOLO (You Only Look Once), Retina-Net, deformabilne konvolucijske mreže, Single Shot MultiBox Detector (SSD), Single-Shot Refinement Neural Network for Object Detection (RefineDet) i drugi. Pristupi koji nisu temeljeni na neuronskim mrežama uključuju okvir za detekciju objekata Viola–Jones temeljen na Haar značajkama, Scale-Invariant Feature Transform (SIFT), Histogram of Oriented Gradients (HOG) značajke i slične metode.

Odabir najboljeg modela za detekciju objekata ovisi o nekoliko faktora, uključujući specifične zahtjeve aplikacije, prirodu podataka te kompromise između brzine, točnosti i računalnih resursa.

Neki od kriterija koji se često razmatraju pri odabiru najbolje modela za detekciju objekata uključuju točnost, brzinu te računalna učinkovitost.

Točnost, mjerena putem Mean Average Precision (mAP), koji pruža sveobuhvatnu mjeru točnosti uzimajući u obzir preciznost i prisjećanje (en. recall) kroz različite IoU pragove i klase objekata, ili putem Class-Level AP-a, koji provjerava prosječnu preciznost za pojedinačne klase kako bi se osiguralo da model dobro funkcionira za sve vrste objekata.

Brzina, mjerena kroz vrijeme izvođenja (en. inference time) i latenciju. Vrijeme izvođenja razmatra brzinu modela u obradi slika, osobito za aplikacije u stvarnom vremenu, dok latencija procjenjuje ukupnu latenciju, uključujući predobradu podataka, izvođenje modela i post-obradu.

Računalna učinkovitost procjenjuje računalne resurse potrebne za rad modela, poput zahtjeva za GPU/CPU i memorije, te skalabilnost kako bi se osiguralo da model može rasti s povećanjem podataka i računalnih zahtjeva.

Ostali specifični metrički pokazatelji uključuju sposobnost modela da obrađuje male i zaklonjene objekte, performanse pod različitim uvjetima, jednostavnost korištenja i integracije, dostupnost unaprijed treniranih modela, fleksibilnost i mogućnost prilagodbe, te karakteristike podataka za obuku.

Standardne evaluacijske metrike u zajednici uključuju Intersection over Union (IoU), koji se koristi za mjerenje preklapanja između predviđenih i stvarnih ograničavajućih okvira, osiguravajući točnu lokalizaciju. F1 score se koristi kao uravnotežena mjera preciznosti (točnost pozitivnih predikcija) i prisjećanja (sposobnost pronalaženja svih relevantnih instanci), osobito kada su kritični i lažno pozitivni i lažno negativni rezultati.

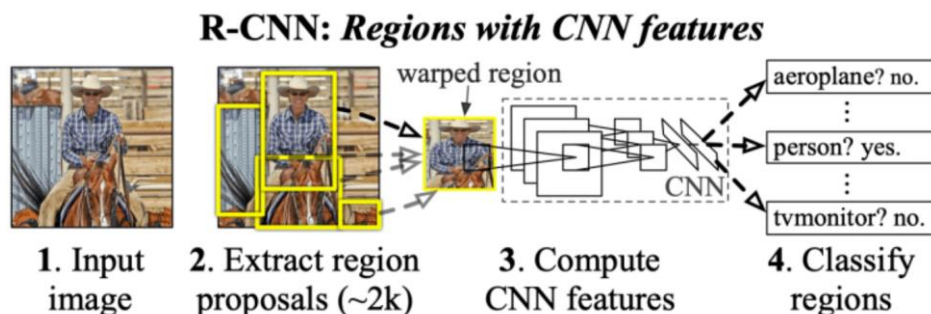
Tablica 2.2 Usporedba točnosti i brzina modela za raspoznavanje objekata

Model	Pascal 2007 mAP	Speed (fps)	Speed (s)
DPM v5	33.7	0.07 fps	14.3s
R-CNN	66	0.05 fps	20.0s
Fast R-CNN	70	0.5 fps	2.0s
Faster R-CNN	73.2	7 fps	0.14s
YOLO	69	45 fps	0.02s
100Hz DPM	16	100 fps	0.01s
Fast YOLO	16	155 fps	0.006s

Mask R-CNN

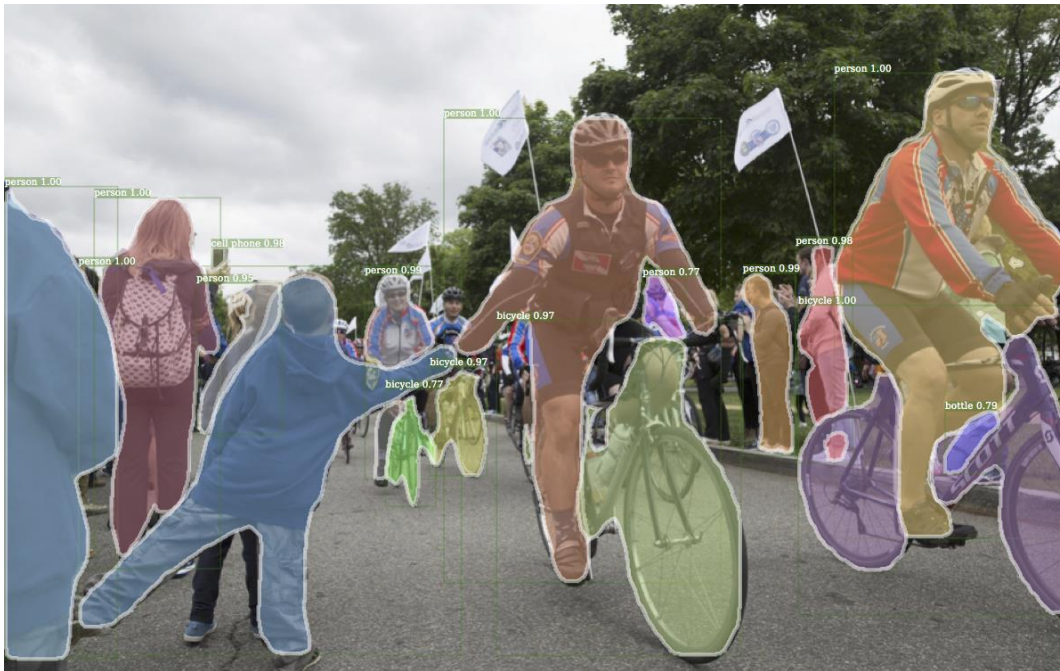
Region-based CNN (R-CNN) pristupi za detekciju objekata unutar ograničavajućih okvira (en. bounding-box) koriste tri odvojena modela: model za prijedlog regija, CNN za ekstrakciju značajki i stroj potpornih vektora (en. support vector machine, SVM) za svaku klasu. Ovi modeli se obično treniraju odvojeno. Iako su se razvijale mnoge metode temeljene na RNN pristupu od 2014. godine, osnovna ideja prijedloga regija uz klasifikaciju i dalje ostaje ključna.

Pristup R-CNN-a za detekciju objekata unutar ograničavajućih okvira je usmjeren na razmatranje ograničenog broja kandidata za regije objekata i procjenu konvolucijskih mreža na svakoj regiji interesa (en. Region of Interest, RoI). R-CNN je proširen kako bi omogućio razmatranje regija interesa na kartama značajki pomoću RoIPool, što je dovelo do brže brzine i bolje točnosti. Faster R-CNN unaprijedio je ovaj pristup učenjem mehanizma usmjeravanja pomoću mreže za prijedlog regija (en. Region Proposal Network, RPN). Faster R-CNN je fleksibilan i robusan prema mnogim nadogradnjama, te je trenutno vodeći okvir u nekoliko mjerila performansi.



Slika 2.10 Rad region-based konvolucijskih neuronskih mreža

Mask R-CNN nadograđuje Faster R-CNN dodavanjem dodatne grane koja služi za predviđanje segmentacijskih maski u svakoj regiji interesa, dok se istovremeno odvija klasifikacija i regresija ograničavajućeg okvira. Ova paralelna struktura omogućuje modelu da, osim što identificira objekte i njihovu poziciju unutar slike, predvidi i točne granice tih objekata. Mask R-CNN je široko primijenjen u projektima računalnog vida, uključujući Meta-in projekt Detectron, koji podržava istraživačke projekte i produkcijske aplikacije unutar Facebooka.

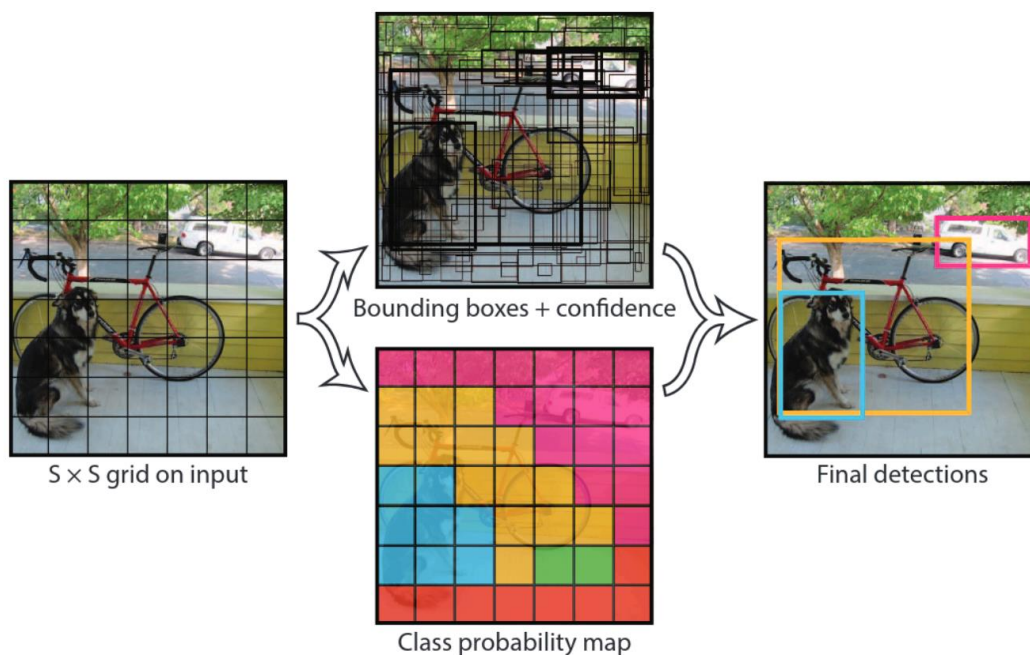


Slika 2.11 Primjer rada Detectrona koji koristi Mask R-CNN model

YOLO (You Only Look Once)

YOLO je popularan model za detekciju objekata u jednom koraku, poznat po svojoj brzini i točnosti. On obrađuje slike u stvarnom vremenu, što ga čini pogodnim za aplikacije koje zahtijevaju brzu detekciju objekata, jer detektira objekte u jednom prolazu kroz mrežu. Za razliku od pristupa poput deformabilnih filtera dijelova ili inspekcije regija kao u RCNN-u, YOLO pristupa slici samo jednom.

Prvi korak u YOLO modelu je prekrivanje ulazne slike mrežom sa S mrežnih ćelija na svakoj osi. Svaka mrežna ćelija tada proizvodi dva rezultata: set ograničavajućih okvira centriranih na točku unutar mrežne ćelije s pridruženim vrijednostima povjerenja da objekt postoji unutar svakog ograničavajućeg okvira, te kartu vjerojatnosti klase za svaku mrežnu ćeliju, koja ukazuje na to koja je klasa objekta najvjerojatnija u toj ćeliji, pod pretpostavkom da objekt uopće postoji u ćeliji. Konačno, model kombinira te informacije kako bi dobio detekciju objekata.

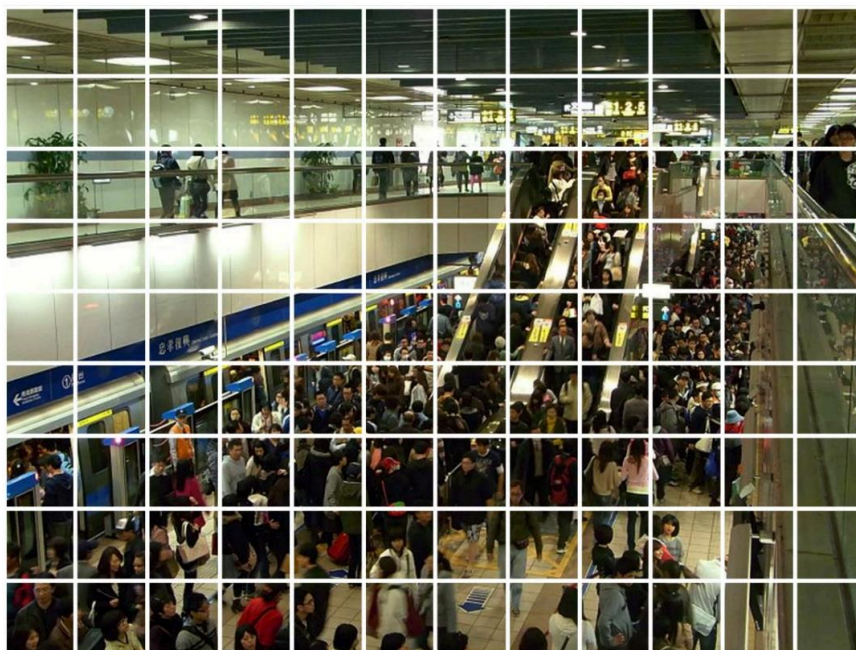


Slika 2.12 Pojednostavljeni prikaz rada YOLO modela

Za svaku mrežnu ćeliju, izračunava se set izlaznih vektora. Prvi vektor je vektor vjerojatnosti klasa, dok drugi set predstavlja parametre ograničavajućeg okvira, specificirajući X i Y poziciju centra okvira, kao i kvadratni korijen širine i visine okvira. Na kraju, tu je i vrijednost povjerenja koja označava sigurnost modela da objekt uistinu postoji unutar okvira.

Umjesto da imaju samo jedan ograničavajući okvir, svaka mrežna ćelija obično ima više okvira, što omogućuje ćeliji da detektira više objekata.

Neki od ograničenja prve verzije YOLO-a, koja su značajno poboljšana u kasnijim verzijama, uključuju činjenicu da svaka mrežna ćelija može sadržavati samo jednu klasu i ograničen broj ograničavajućih okvira, što može predstavljati problem u scenama s puno objekata, poput urbanih okruženja s mnogo ljudi u jednoj mrežnoj ćeliji.



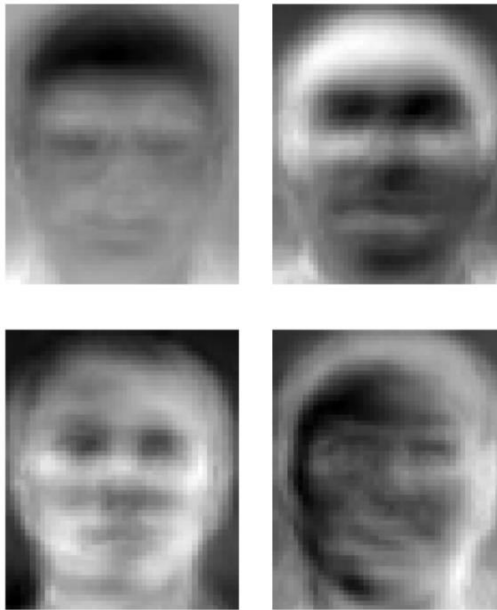
Slika 2.13 Primjer izazovnog okruženja za YOLO model

Osim toga, predviđanje ograničavajućih okvira kod YOLO-a može imati problema s neuobičajenim omjerima širine i visine objekata, a isti stupanj pogreške u malom ograničavajućem okviru tretira se ozbiljnije nego isti stupanj pogreške u velikom okviru.

- **Detekcija lica**

Detekcija lica je računalna tehnologija koja se koristi u raznim primjenama za identifikaciju ljudskih lica na digitalnim slikama, te se može smatrati specifičnim slučajem detekcije klasa objekata. Detekcija lica odgovara na dva osnovna pitanja: ima li ljudskih lica u prikupljenim slikama ili videu te gdje se lica nalaze.

Popularni pristupi detekciji lica temelje se na genetskom algoritmu u kombinaciji s tehnikom eigenface. Prvo se detektiraju moguće regije ljudskih očiju testiranjem svih dolinskih regija na slici u sivoj skali, a zatim se genetski algoritam koristi za generiranje svih mogućih regija lica koje uključuju obrve, šarenicu, nosnice i kutove usta. Vrijednost prikladnosti svakog kandidata mjeri se na temelju njegove projekcije na eigenface. Nakon nekoliko iteracija, svi kandidati za lica s visokim vrijednostima prikladnosti odabiru se za daljnju verifikaciju.



Slika 2.14 Primjer eigenface-ova

Pristup eigenface započeo je istraživanjem niskodimenzionalne reprezentacije slika lica. Lawrence Sirovich i M. Kirby pokazali su da se analiza glavnih komponenti (en. principal component analysis, PCA) može primijeniti na zbirku slika lica kako bi se formirao skup osnovnih značajki. Te osnovne slike, poznate kao eigenpictures, mogu se linearno kombinirati kako bi se rekonstruirale slike u izvornom skupu za obuku. Godine 1991., M. Turk i A. Pentland proširili su ove rezultate i predstavili metodu eigenface za prepoznavanje lica.

Popularniji pristup detekciji lica su algoritmi fisherface, koji se široko smatraju superiornijima u odnosu na druge tehnike, poput eigenface, zbog napora da se maksimizira razdvajanje između klasa tijekom procesa obuke. Prepoznavanje lica metodom fisherface ne samo da je sposobno obraditi testne slike lica s različitim komponentama boja u odnosu na slike za obuku i skicu izvorne slike, već je i otporno na šum i učinak zamućenja na slici. Većina slika koje ne uspijevaju u prepoznavanju obično ne uspijeva zbog dva faktora: skaliranja i položaja. Prvi faktor se može riješiti korištenjem boljeg skaliranja slike, dok se problem s položajem može prevladati dodavanjem više slika za obuku s različitim položajima.

Primjene i trendovi

Detekcija lica se koristi u sektorima poput hvatanja pokreta lica i prepoznavanja lica u biometriji, video nadzoru i sličnim aplikacijama. Također se koristi u fotografiji za automatsko fokusiranje, dok moderni uređaji koriste detekciju osmijeha za snimanje fotografija u pravom trenutku. U marketingu, web kamera može biti integrirana u televiziju i detektirati lica koja prolaze pored nje. Sustav tada izračunava rasu, spol i dobni raspon lica, te na temelju tih informacija prikazuje seriju reklama prilagođenih otkrivenim karakteristikama.

Detekcija lica također je našla svoje mjesto u sigurnosti putem automatiziranog čitanja s usana kako bi računala mogla odrediti tko govori. U zdravstvu, detekcija lica može se koristiti kao dio sustava za zaključivanje emocija, što može pomoći osobama s autizmom da razumiju osjećaje ljudi oko sebe.

U posljednjih nekoliko godina, detekcija emocija na licima uz pomoć umjetne inteligencije (AI) dobila je značajnu pažnju, koristeći različite modele za interpretaciju ljudskih emocionalnih stanja. OpenAI-ov CLIP model je primjer korištenja dubokog učenja za povezivanje slika i teksta, omogućujući nijansirano razumijevanje emocionalnog sadržaja. Na primjer, u kombinaciji s pristupom psihometrijske mreže, model se koristio za analizu političkih govora na temelju promjena u izrazima lica političara, iako još uvijek postoje izazovi u preciznom razlikovanju srodnih emocija i razumijevanju kulturnih nijansi.

3. Praćenje pogleda

Praćenje pogleda (en. eye-tracking) mjeri pokrete oka pomoću nekoliko metoda, koje se najčešće svrstavaju u tri glavne kategorije. Prva metoda uključuje praćenje pokreta objekta, poput posebne kontaktne leće, pričvršćene na oko. Druga metoda je optičko praćenje, koje ne zahtijeva izravan kontakt s okom i obično koristi infracrvenu svjetlost kako bi pratilo kretanje oka. Treća metoda temelji se na mjerenju električnih potencijala pomoću elektroda postavljenih oko očiju, a može se koristiti čak i u potpunom mraku ili kad su oči zatvorene.

Optičko praćenje, bilo pomoću infracrvene ili vidljive svjetlosti, posebno ono temeljeno na video snimanju, široko se koristi za praćenje pogleda te je popularno zbog svoje neinvazivnosti i relativno niske cijene.

Mnoga istraživanja posvećena su proučavanju mehanizama i dinamike rotacije oka, ali cilj praćenja pogleda najčešće je procijeniti smjer pogleda kako bi otkrili koje značajke slike najviše privlače pogled. Praćenje pogleda ne pruža apsolutni smjer pogleda, već mjeri samo promjene u smjeru pogleda. Kako bi se točno odredilo što subjekt gleda, potreban je neki postupak kalibracije u kojem subjekt gleda u točku ili niz točaka, dok eye-tracker bilježi vrijednosti koje odgovaraju svakoj poziciji pogleda.

3.1. Povijest

Tokom 19. stoljeća, glavnina istraživanja pokreta oka provodila su se izravnim promatranjem. Primjerice, Louis Émile Javal je 1879. godine uočio da čitanje ne uključuje glatko klizanje očiju duž teksta, kao što se ranije pretpostavljalo, već niz kratkih zaustavljanja (koji se nazivaju fiksacije) i brzih saccada. Taj niz fiksacija i saccada naziva se putanja pregleda (en. scanpath). Ovo opažanje otvorilo je važna pitanja o procesu čitanja koja su istraživana tijekom 20. stoljeća: na kojim se riječima oči zaustavljaju, koliko dugo traju te fiksacije, te kada i zašto dolazi do vraćanja očiju na već viđene riječi.

DANS, KÖN OCH JAGPROJEKT

På jakt efter ungdomars kroppsspråk och den "synkretiska dansen", en sammansmältning av olika kulturers dans, har jag i mitt fältarbete under hösten rörligt mig på olika arenor inom skolans värld. Nordiska, afrikanska, syd- och östeuropeiska ungdomar gör sina röster hörda genom sång, musik, skrik, skriff och gestaltat känslor och uttryck med hjälp av kroppsspråk och dans.

Den individuella estetiken framträder i kläder, frisyrer och symboliska tecken som förstärker ungdomarnas "jagprojekt" där också den egna stilen i kroppsrörelserna spelar en betydande roll i identitetsprövningen. Upphållsrummet fungerar som offentlig arena där ungdomarna spelar upp sina performance-liknande kroppsspråk.

Slika 3.1 Primjer saccade-ova prilikom čitanja teksta

Prvi neinvazivni uređaji za praćenje pogleda izgradio je Guy Thomas Buswell u Chicagu, koristeći snopove svjetlosti koji su se reflektirali na oku, a zatim snimali na film. U 1950-ima, Alfred L. Yarbus je proveo istraživanje praćenja pogleda, a njegova knjiga iz 1967. godine često se citira. Pokazao je da zadatak koji je zadan subjektu ima veliki utjecaj na pokrete očiju subjekta.

Zabilješke pokreta očiju pokazuju da je pažnja promatrača obično usmjerena samo na određene elemente slike. Pokreti očiju odražavaju ljudske misaone procese; stoga se do određene mjere može pratiti misao promatrača putem zabilješki pokreta očiju (misao koja prati pregledavanje određenog objekta). Iz ovih zapisa lako je utvrditi koji elementi privlače pogled promatrača (a time i njegovu misao), kojim redoslijedom i koliko često. Pažnju promatrača često privlače elementi koji ne daju važne informacije, ali koji, prema njegovom mišljenju, mogu to učiniti. Često će promatrač usmjeriti svoju pažnju na elemente koji su neobični u određenim okolnostima, nepoznati, neshvatljivi i slično.

Kraj 20. stoljeća donio je početak korištenja praćenja pogleda za odgovaranje na pitanja vezana uz interakciju između čovjeka i računala. Konkretno, istraživači su ispitivali kako korisnici traže naredbe u računalnim izbornicima. Osim toga, računala su omogućila

istraživačima da koriste rezultate praćenja pogleda u stvarnom vremenu, prvenstveno kako bi pomogli korisnicima s invaliditetom.

U novije vrijeme, došlo je do rasta u korištenju praćenja pogleda za proučavanje interakcije korisnika s različitim računalnim sučeljima. Najčešća pitanja koja istraživači postavljaju odnose se na to koliko su različita sučelja jednostavna za korisnike te kako to poboljšati radeći promjene u dizajnu sučelja. Još jedno nedavno područje istraživanja fokusira se na razvoj weba, uključujući kako korisnici reagiraju na padajuće izbornike ili gdje usmjeravaju svoju pažnju na web stranici kako bi se uočila najbolja mjesta za oglase.

U 21. stoljeću, korištenje umjetne inteligencije i umjetnih neuronskih mreža postalo je održiv način za obavljanje zadataka i analize praćenja pogleda. Posebno se konvolucijske neuronske mreže koriste za praćenje pogleda, jer su dizajnirane za zadatke vezane uz slike. Pomoću umjetne inteligencije, zadaci praćenja pogleda mogu pružiti dodatne informacije koje možda nisu bile otkrivene ljudskim promatračima. Praksa dubokog učenja također omogućuje da se neuronska mreža poboljša u određenom zadatku kada je izložena dovoljnoj količini uzoraka podataka.

Ovakva vrsta praćenja pogleda može se implementirati na jednostavnim sustavima poput osobnih prijenosnih računala, a takve primjene istražene su u sljedećem poglavlju.

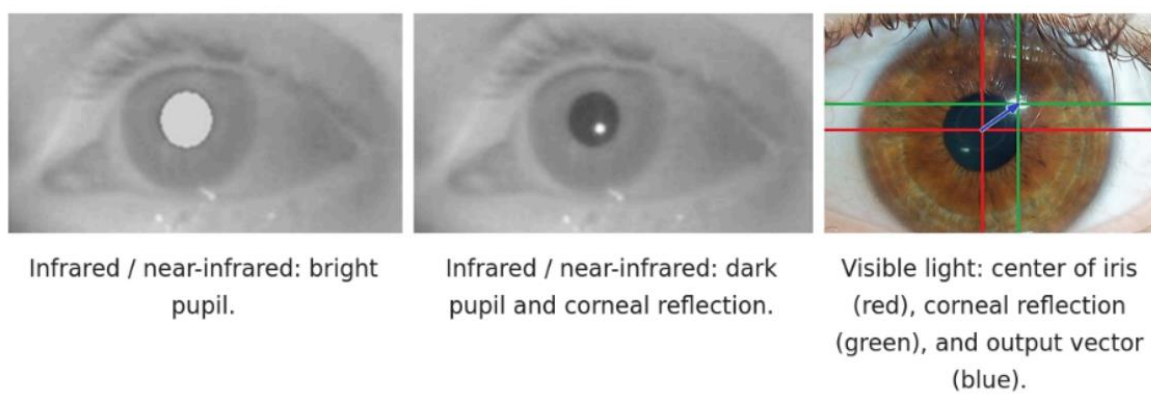
3.2. Tehnike praćenja

Najšire korišteni dizajni za praćenje pogleda danas su oni temeljeni na video tehnologiji. Kamera se fokusira na jedno ili oba oka i bilježi pokrete oka dok promatrač gleda neki podražaj. Većina modernih uređaja za praćenje pogleda koristi središte zjenice i infracrvenu svjetlost za stvaranje refleksija na rožnici. Vektor između središta zjenice i refleksije na rožnici može se koristiti za izračunavanje točke pogleda na površini ili smjera pogleda. Obično je potrebna jednostavna kalibracija pojedinca prije korištenja uređaja za praćenje pogleda.

Postoje dvije opće vrste tehnika praćenja pogleda koje koriste infracrvenu svjetlost (poznatu i kao aktivna svjetlost): praćenje svijetle zjenice i praćenje tamne zjenice. Razlika između njih temelji se na položaju izvora osvjetljenja u odnosu na optiku. Ako je izvor

osvjetljenja koaksijalan s optičkom putanjom, oko djeluje kao retroreflektor jer se svjetlost reflektira od mrežnice, stvarajući efekt svijetle zjenice sličan crvenom oku. Ako je izvor osvjetljenja pomaknut u odnosu na optičku putanju, zjenica izgleda tamno jer je retrorefleksija s mrežnice usmjerena dalje od kamere.

Praćenje svijetle zjenice stvara veći kontrast između šarenice i zjenice što omogućuje robusnije praćenje pogleda sa svim pigmentacijama šarenice, te znatno smanjuje smetnje uzrokovane trepavicama i drugim prekrivajućim značajkama. Također omogućuje praćenje u uvjetima osvjetljenja od potpune tame do vrlo jakog svjetla.



Slika 3.2 Tehnike praćenja pogleda koristeći infracrvenu i vidljivu svjetlost

Druga, manje korištena metoda poznata je kao pasivna svjetlost. Ona koristi vidljivu svjetlost za osvjetljavanje, što može uzrokovati određene smetnje za korisnike. Dodatni izazov ove metode je manji kontrast zjenice u usporedbi s metodama aktivne svjetlosti, zbog čega se za izračunavanje vektora koristi središte šarenice. Ova metoda zahtijeva detekciju granice šarenice i bijele bjeloočnice (en. limbus tracking), što predstavlja dodatni izazov za vertikalne pokrete oka zbog prekrivanja kopcima.

3.3. Vizualizacija podataka

Podaci zabilježeni uređajima za praćenje pogleda mogu se prikazati na različite načine, ovisno o potrebama aplikacije. Popularne vizualizacije mogu se podijeliti na statičke i dinamičke prikaze podataka.

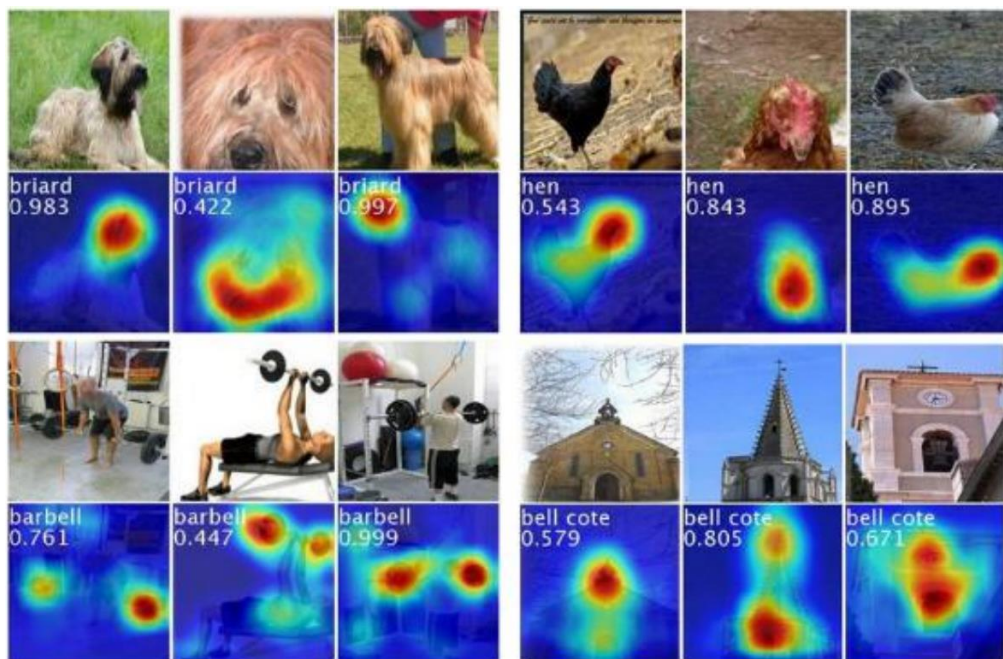
Animirane reprezentacije točke na sučelju koriste se kada se vizualno ponašanje ispituje pojedinačno, pokazujući gdje je korisnik fokusirao svoj pogled u svakom trenutku, uz malu putanju koja označava prethodne saccade pokrete.

Pojednostavljena verzija ovog prikaza su statičke reprezentacije putanje saccade, s razlikom da je ovo statična metoda koja zahtijeva višu razinu stručnosti za interpretaciju u usporedbi s animiranim prikazima.

Alternativni statički prikaz su toplinske karte (en. heat maps), koje se koriste uglavnom za skupnu analizu vizualnih obrazaca istraživanja u skupini korisnika. U ovim prikazima, "vruće" zone ili zone s većom gustoćom označavaju gdje su korisnici najčešće fokusirali svoj pogled (ne nužno i pažnju). Toplinske karte su najpoznatija tehnika vizualizacije u istraživanjima praćenja pogleda.

Karte istaknutih područja (en. saliency maps) su jednostavnija alternativa koja ilustrira područja fokusa tako što svijetlo prikazuje objekte koji privlače pažnju na crnoj podlozi. Što je veća usmjerenost na određeni objekt, to će svjetlije izgledati.

Pojednostavljena verzija toplinske karte koja prikazuje samo "slijepe" zone ili fokusne karte također se koristi, gdje su jasno prikazane zone koje su manje vizualno posjećene od strane korisnika, što omogućuje lakše razumijevanje najrelevantnijih informacija.



Slika 3.3 Primjer toplinskih karti za prikaz područja fokusa pogleda

3.4. Primjene i trendovi

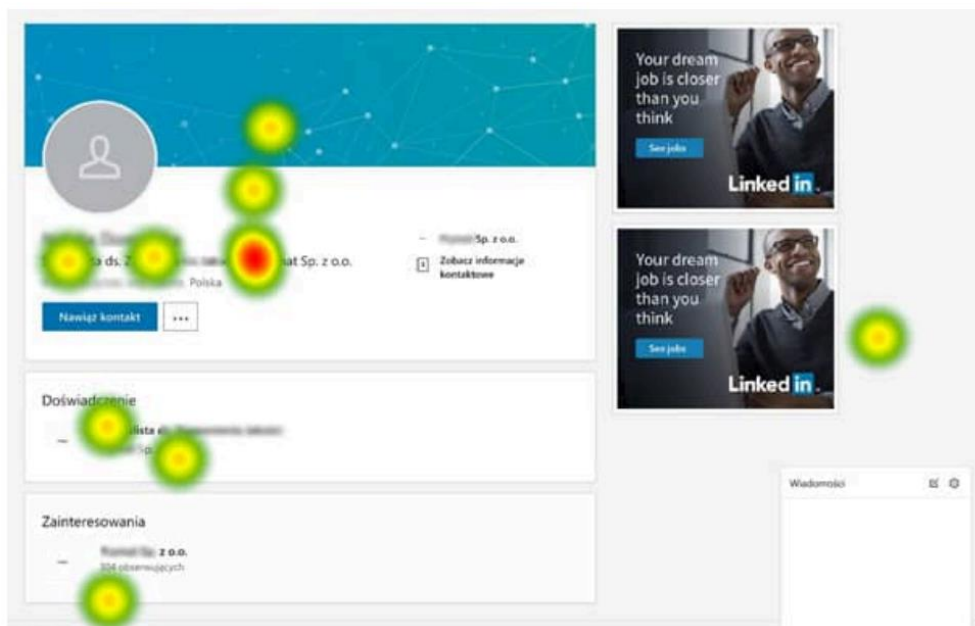
Tehnike praćenja pogleda koriste se u raznim područjima, uključujući kognitivnu znanost, psihologiju, interakciju čovjek-računalo, ergonomiju, marketing i medicinska istraživanja. Primjene su brojne, od praćenja pokreta oka prilikom čitanja, prepoznavanja aktivnosti, do procjene kognitivnog opterećenja vozača i pilota, te upravljanja računalima kod osoba s motoričkim poteškoćama. U virtualnoj stvarnosti, praćenje pogleda pomaže u optimizaciji grafičkog prikaza, tako što se detaljno renderira samo područje koje korisnik promatra.

Također u posljednjih nekoliko godina, sve veća sofisticiranost i dostupnost tehnologija za praćenje pogleda izazvale su veliki interes u komercijalnom sektoru. Primjene uključuju automobilsku industriju (npr. detekcija pospanosti i intoksikacije), analizu upotrebljivosti web stranica, oglašavanje, sponzorstva, dizajn pakiranja i asistivne tehnologije. Općenito, komercijalna istraživanja praćenja pogleda funkcioniraju tako da ispitanicima prezentiraju ciljne podražaje, poput web stranica, sportskih događaja, filmova, reklama, časopisa, novina, pakiranja, polica u trgovinama i slično, dok uređaj za praćenje pogleda bilježi aktivnost oka.

Dobiveni podaci mogu se statistički analizirati i grafički prikazati kako bi se identificirali specifični vizualni obrasci koji omogućuju istraživačima da donesu zaključke o učinkovitosti određenog medija ili proizvoda.

Iako su tradicionalne tehnike analize upotrebljivosti web stranica često vrlo učinkovite u pružanju informacija o obrascima klikanja i skrolanja, praćenje pogleda nudi mogućnost analize interakcije korisnika između klikova te vremena provedenog između njih. To pruža vrijedan uvid u to koji su elementi web stranice najupečatljiviji, koji uzrokuju zbunjenost, a koji se u potpunosti ignoriraju. Konkretno, praćenje pogleda može se koristiti za procjenu učinkovitosti pretraživanja, brendiranja, online oglasa, upotrebljivosti navigacije, ukupnog dizajna i mnogih drugih komponenti web stranice.

Jedan od primjera istraživanja temeljenih na praćenju pogleda dolazi iz područja zapošljavanja, gdje je jedno istraživanje analiziralo kako zapošljivači pregledavaju LinkedIn profile te je rezultate prikazalo putem toplinskih mapa (en. heat maps).



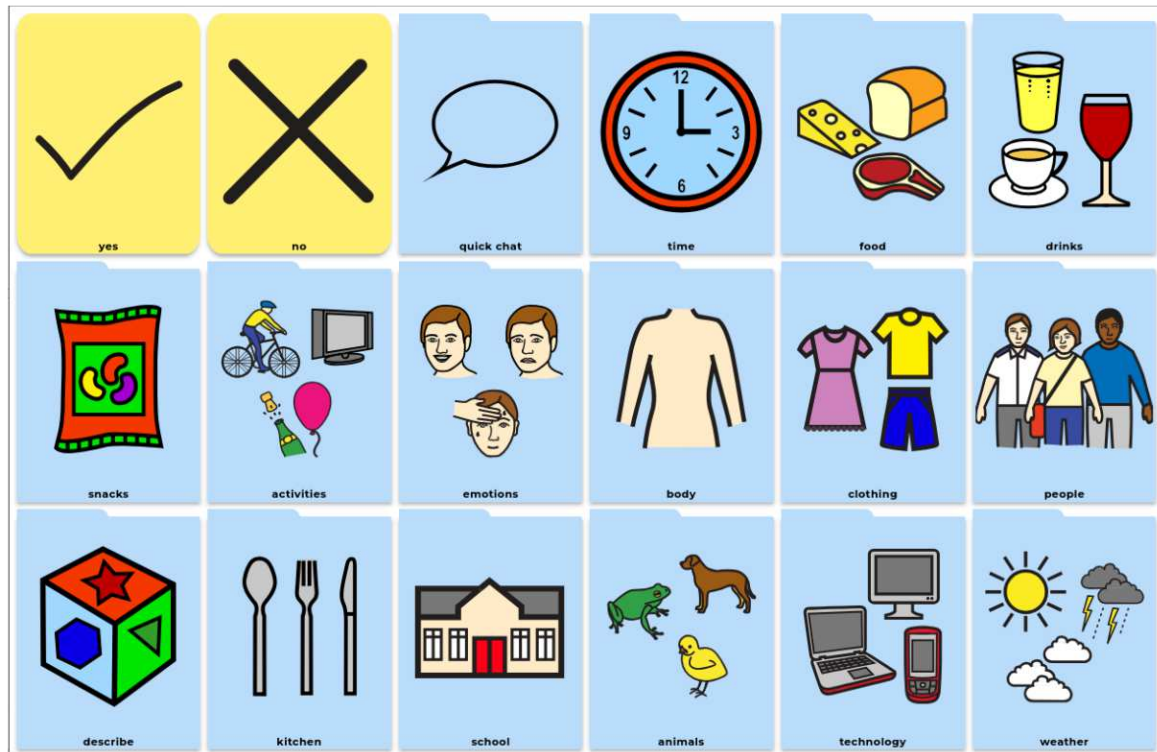
Slika 3.4 Područja fokusa pogleda zapošljivača na LinkedIn platformi

U istraživanju iz 2019. godine, konvolucijska neuronska mreža razvijena je kako bi identificirala pojedine šahovske figure na sličan način kao što druge CNN mreže prepoznaju crte lica. Zatim je mreža dobila podatke o praćenju pogleda od 30 šahista različitih razina vještine. Korištenjem tih podataka, CNN je koristio procjenu pogleda (en. gaze estimation) kako bi utvrdio na koje dijelove šahovske ploče igrači obraćaju posebnu pažnju te je generirao kartu istaknutih područja. Na kraju, CNN je kombinirao svoje znanje o ploči i figurama s kartom istaknutih područja kako bi predvidio sljedeći potez igrača. Bez obzira na podatkovni skup na kojem je mreža trenirana, predviđanje sljedećeg poteza bilo je točnije nego nasumično odabrani potez, a karte istaknutih područja nacrtane za određenog igrača i situaciju bile su više od 54% slične.

U industriji video igara, igra "Before Your Eyes" iz 2021. godine koristi treptanje igrača kao glavni način interakcije s igrom, što je još jedan primjer primjene praćenja pogleda.

Osobe s teškim motoričkim oštećenjima mogu koristiti praćenje pogleda za interakciju s računalima, što je brže od tehnika jedinstvenog pretraživanja te intuitivnije za korištenje.

Usmjeravanje pogleda na objekt interesa vrlo je intuitivan način komunikacije koji može normalizirati razgovore, ubrzati učenje i povećati motivaciju. Međutim, jedna od glavnih mana tehnika praćenja pogleda je njihova ovisnost o osjetljivosti osobe koja mora stalno pratiti na što osoba gleda.



Slika 3.5 Primjer komunikacijske ploče web aplikacije CBoard

U novije vrijeme, istraživači su istraživali i mogućnost kontrole robotskih ruku te motoriziranih invalidskih kolica putem praćenja pogleda. Praćenje pogleda također je korisno u analizi vizualnih obrazaca pretraživanja, otkrivanju prisutnosti nistagmusa (en. nystagmus) te u detekciji ranih znakova poteškoća u učenju kroz analizu kretanja očiju tijekom čitanja.

Tehnologija praćenja pogleda i tehnike analize podataka koriste se i za istraživanje razumljivosti koncepata u softverskom inženjerstvu. To uključuje razumljivost modela poslovnih procesa i dijagrama koji se koriste u softverskom inženjerstvu, poput UML dijagrama aktivnosti i EER dijagrama.

Kartografska istraživanja također su široko usvojila tehnike praćenja pogleda. Istraživači ih koriste za analizu načina na koji pojedinci percipiraju i interpretiraju karte. Na primjer,

praćenje pogleda korišteno je za proučavanje razlika u percepciji 2D i 3D vizualizacija, usporedbu strategija čitanja karata između početnika i stručnjaka ili studenata i njihovih profesora geografije te za procjenu kartografske kvalitete karata. Uz to, kartografi koriste praćenje pogleda za istraživanje različitih faktora koji utječu na čitanje karata, uključujući attribute kao što su boja ili gustoća simbola.

4. Programska implementacija

Kao jedan od primjera jednostavne implementacije praćenja pogleda je korištenje biblioteke WebGazer.js, koja omogućava praćenje pogleda u stvarnom vremenu koristeći obične web kamere. Ova tehnologija omogućuje integraciju unutar bilo koje web stranice bez potrebe za skupom opremom ili dodatnim hardverom, budući da koristi već ugrađene web kamere u računalima i laptopima. Osim što je vrlo dostupna, velika prednost WebGazer.js-a je što svi podaci ostaju lokalni na korisničkom uređaju, bez potrebe za prijenosom videozapisa na poslužitelje, čime se dodatno štiti privatnost korisnika.

WebGazer.js koristi kombinaciju algoritama za detekciju zjenice i regresijskih modela kako bi predvidio točku pogleda korisnika na zaslonu. Detekcija lica i očiju se odvija pomoću biblioteka kao što su MediaPipe Facemesh i clmtrackr, koje identificiraju područje očiju na temelju videozapisa s web kamere. Unutar tog područja, WebGazer.js detektira zjenicu, koristeći pretpostavke da je zjenica tamnija od okolnog tkiva i da je smještena približno u središtu šarenice. Algoritam zatim koristi te podatke za procjenu točke pogleda.

Jedna od ključnih komponenti sustava je obrada slike oka kako bi se dobile značajke koje se koriste za predikciju točke pogleda. WebGazer.js koristi 120-dimenzionalne (120D) vektore značajki oka, koje se stvaraju iz malih slika (patches) oka, veličine 6 x 10 piksela. Prvo se slika oka konvertira u sivu sliku, nakon čega se primjenjuje histogram izjednačavanja kako bi se poboljšao kontrast. Ove obrađene slike zatim se koriste za stvaranje vektora značajki koji opisuju karakteristike oka na način koji može biti koristan za predikciju pozicije pogleda.

Ovi 120D vektori omogućuju precizniju analizu karakteristika oka jer kvantificiraju informacije o teksturi, kontrastu i obliku zjenice i okolnog tkiva. Točnost modela za predikciju ovisi o preciznosti tih značajki, a povećanjem dimenzionalnosti vektora dobiva se bogatija informacija koja omogućuje bolju procjenu.

Ridge regresija je primarni algoritam koji se koristi za mapiranje značajki oka (poput pozicije zjenice) na koordinate zaslona. Ridge regresija je vrsta regularizirane linearne regresije, koja se koristi kako bi se spriječilo prenaučavanje modela na podacima s previše značajki. U WebGazer.js, ovaj model uzima vektore značajki oka kao ulaz i trenira model koji predviđa gdje korisnik gleda na zaslonu. Regularizacija omogućava sustavu da ostane robustan i precizan čak i s velikim brojem dimenzija, sprječavajući prekomjerno prilagođavanje modela specifičnim podacima za treniranje. Formula za ridge regresiju u modelu je također linearna što omogućuje efikasnu predikciju u stvarnom vremenu brzo može računanje predikcije na temelju ulaznih podataka.

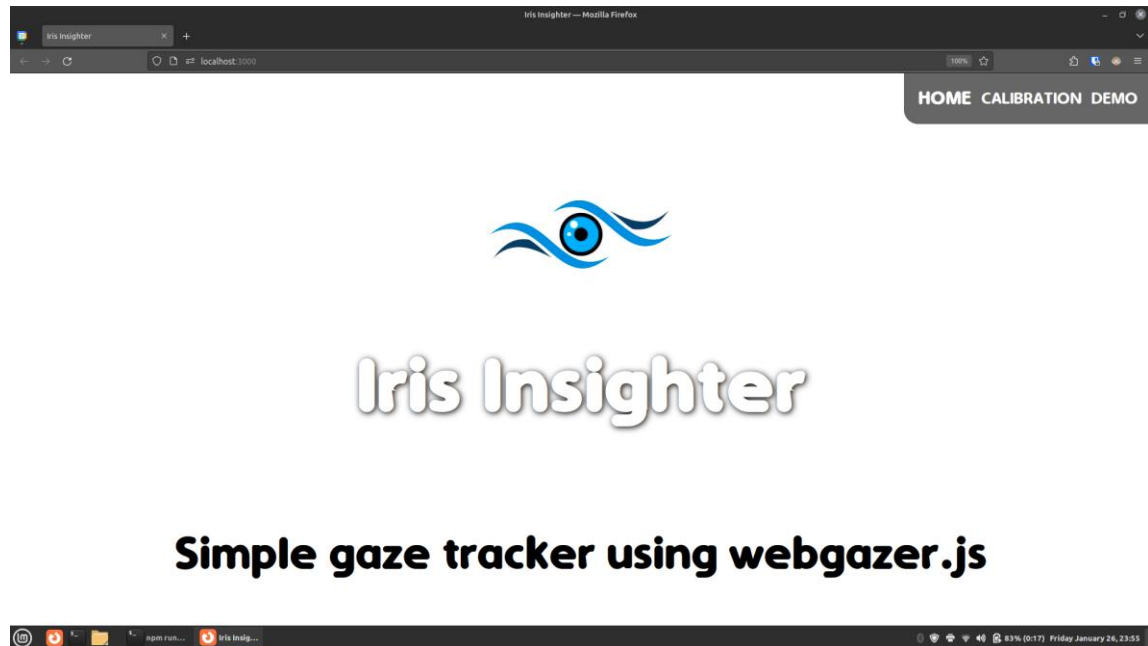
Jedna od ključnih prednosti WebGazer.js je kontinuirana samokalibracija, što znači da nije potrebna eksplicitna kalibracija korisnika na početku korištenja. Umjesto toga, model koristi prirodne interakcije korisnika s web stranicom, poput klikova mišem, za učenje i prilagodbu. Prilikom svakog klika, algoritam pretpostavlja da je korisnikov pogled usklađen s pozicijom kursora, dok se ta informacija koristi za treniranje modela. Kako korisnik više komunicira sa stranicom, model postaje sve precizniji, jer koristi sve više podataka o tome kako se pokreti oka odnose na poziciju kursora.

WebGazer.js omogućava vizualizaciju podataka o praćenju pogleda koristeći nekoliko različitih metoda. Toplinske karte prikazuju koja su područja zaslona korisnici najviše gledali, dok karte fokusa (en. saccade paths) prate putanju korisnikovog pogleda, pokazujući smjerove u kojima su se oči kretale tijekom navigacije stranicom. Ove vizualizacije pružaju vrijedne uvide u ponašanje korisnika i mogu se koristiti za analizu korisničkog iskustva, optimizaciju rasporeda elemenata na web stranicama i istraživanje interakcija korisnika.

4.1. Primjer korištenja

U sklopu rada izrađena je jednostavna web aplikacija za praćenje pogleda koja koristi WebGazer.js, dostupna na GitHubu pod licencom GNU GPL verzije 3 ili novije. Aplikacija je osmišljena kao alat koji korisnicima omogućuje kalibraciju i praćenje točke pogleda u stvarnom vremenu koristeći ugrađene web kamere u preglednicima. Aplikacija

nudi nekoliko osnovnih funkcionalnosti, uključujući početni zaslon, kalibraciju i mjerenje točnosti te demo način rada za testiranje predikcije pogleda.



Slika 4.1 Početna stranica Iris Insighter-a

Kada korisnik pokrene aplikaciju, dočekuje ga početni zaslon s jednostavnim logom aplikacije, nazivom i kratkim opisom funkcionalnosti. U gornjem desnom kutu nalazi se navigacijska traka koja omogućuje pristup stranicama za kalibraciju, demonstraciju i povratak na početni zaslon. Navigacija je jednostavna i intuitivna, čime se korisnicima olakšava korištenje aplikacije.

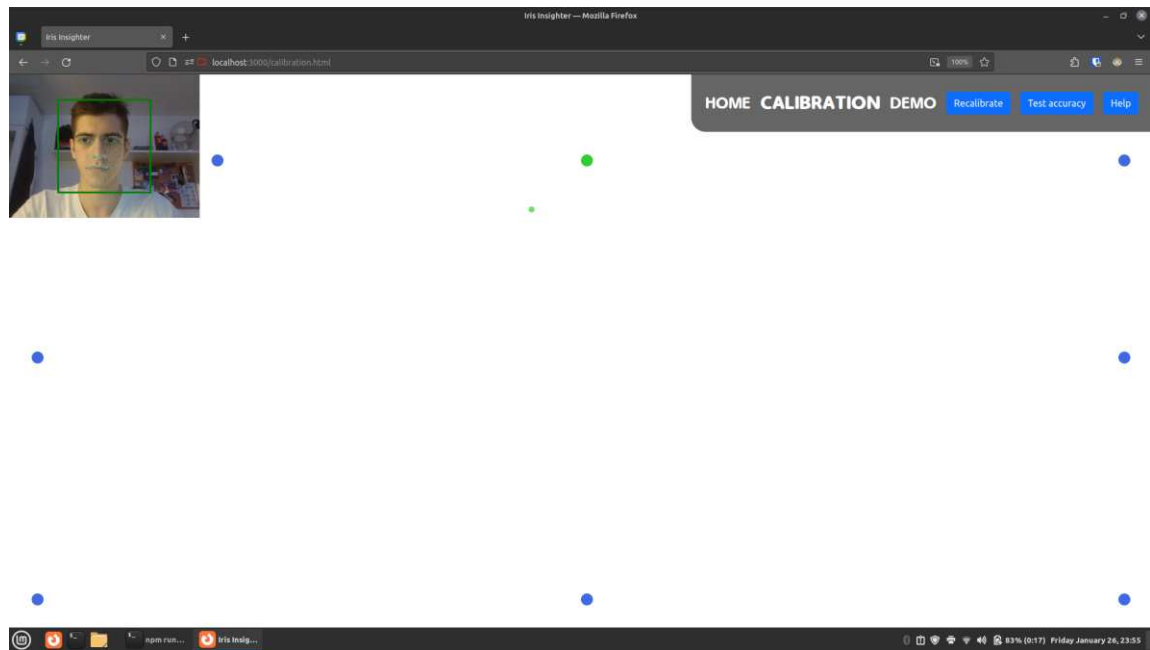
- **Kalibracija i testiranje točnosti**

Odabirom opcije Kalibracija u navigaciji, korisnik je preusmjeren na stranicu za kalibraciju, gdje se prikazuje kamera uživo korisnika s pravokutnikom u kojem sustav predviđa točku pogleda. Kada je korisnik pravilno pozicioniran ispred kamere, pravokutnik mijenja boju iz crvene u zelenu, čime aplikacija pokazuje da je spremna za preciznu predikciju pogleda. Također, na ovoj stranici korisniku je prikazana mala točka koja označava predviđenu točku pogleda u stvarnom vremenu.

U izborniku stranice za kalibraciju nalaze se dodatne opcije: Rekalibriraj, Testiraj točnost i Pomoć. Gumb Rekalibriraj omogućava brisanje postojećih podataka o kalibraciji i ponovno postavljanje kalibracijskih točaka na zaslonu. Testiraj točnost omogućava

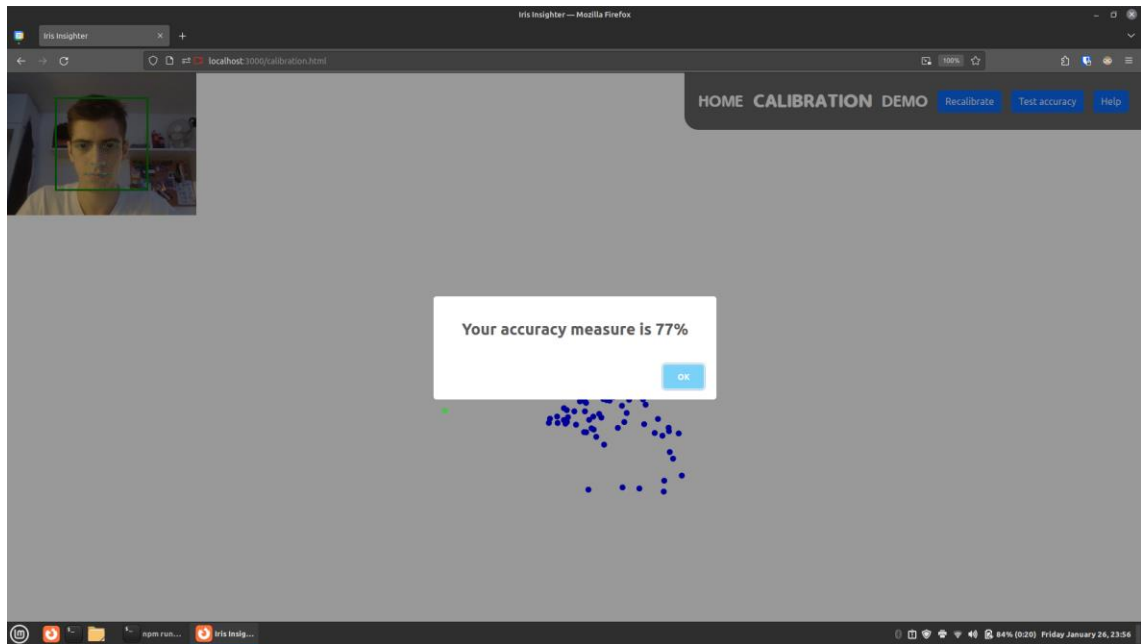
korisniku testiranje točnosti modela nakon kalibracije, dok gumb Pomoć otvara prozor s uputama za korištenje aplikacije i objašnjenjem postupka kalibracije.

Kalibracija se provodi tako što korisnik klikne na točke na ekranu dok gleda u njih. Nakon što korisnik uspješno kalibrira osam početnih točaka, prikazuje se središnja točka za završnu kalibraciju. Kada se kalibrira svih devet točaka, započinje testiranje točnosti.



Slika 4.2 Kalibracijska stranica s kalibracijskim točkama

Testiranje točnosti sastoji se od toga da korisnik fokusira pogled na središnju točku tijekom pet sekundi dok model predviđa gdje korisnik gleda. Na temelju prosječne Euklidske udaljenosti između predikcije i središnje točke, korisnik dobiva procjenu točnosti modela.

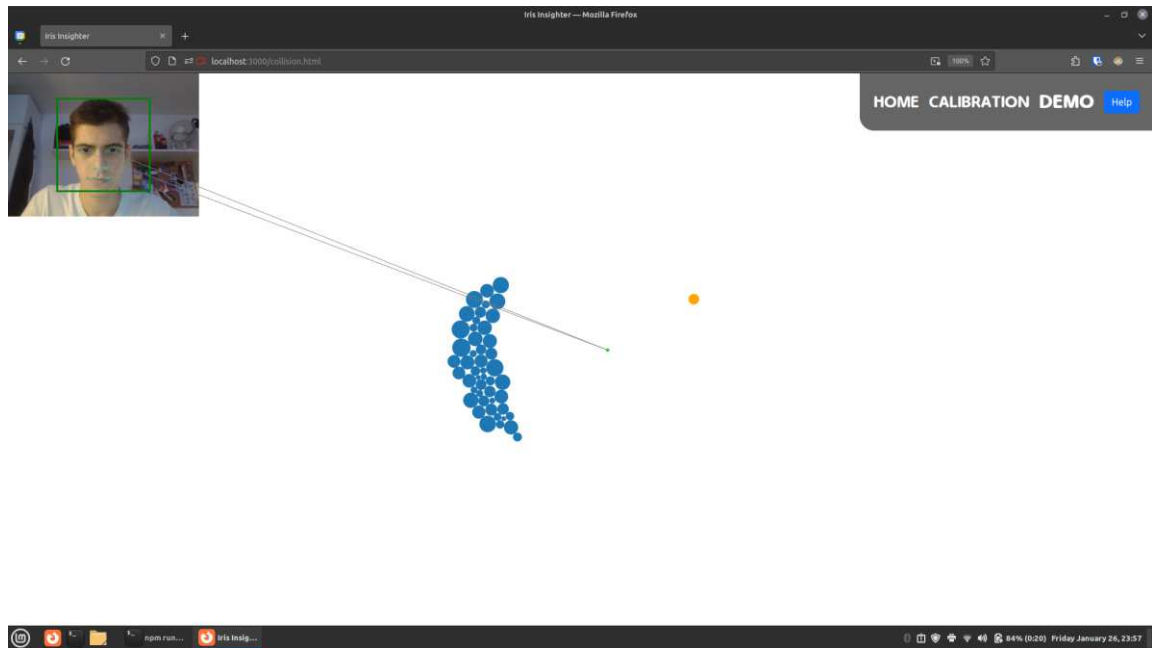


Slika 4.3 Kalibracijska stranica nakon testiranja točnosti modela

- **Demo**

Odabirom opcije Demo u navigaciji, korisnik je preusmjeren na stranicu za demonstraciju. Slično kao na stranici za kalibraciju, prikazuje se pregled kamere s pravokutnikom koji prikazuje predviđenu točku pogleda. Korisnik može ponovno koristiti gumb Pomoć kako bi dobio dodatne informacije o tome kako kalibrirati model i testirati demo funkcionalnosti.

Model se može kalibrirati klikom na različite dijelove zaslona dok korisnik gleda u pokazivač miša. Što više klikova korisnik napravi na različitim dijelovima ekrana, to će predikcija modela biti preciznija. Cilj demo načina rada je da korisnik pokušava razdvojiti kuglice na ekranu pomoću pogleda, dok se kuglice automatski vraćaju prema sredini ekrana.



Slika 4.4 Demo stranica

Za testiranje funkcionalnosti kuglica korisnik također može koristiti pokazivač miša umjesto pogleda, što omogućuje usporedbu između predikcije pogleda i kretanja kursora.

Zaključak

Računalni vid i njegove primjene, s naglaskom na prepoznavanja objekata i praćenja pogleda, postali su ključni alati u mnogim industrijama, transformirajući način na koji interagiramo s tehnologijom. Povijesni pregled pokazao je kako su se ove tehnologije razvijale, od jednostavnih opažanja pokreta očiju u 19. stoljeću, do sofisticiranih metoda dubokog učenja koje danas omogućuju "nadjudsko" prepoznavanje objekata.

Analiza tehnika praćenja pogleda pokazala je važnost optičkog praćenja u stvarnom vremenu te korištenje infracrvene svjetlosti za precizno praćenje pokreta oka. Vizualizacija podataka dobivenih praćenjem pogleda nudi važne uvide u ponašanje korisnika i omogućava optimizaciju dizajna računalnih sučelja.

Napredak u algoritmima dubokog učenja, poput Mask R-CNN i YOLO modela, omogućio je značajna poboljšanja u preciznosti i brzini prepoznavanja objekata, dok su tehnike augmentacije slika pokazale ključnu ulogu u rješavanju izazova poput neravnoteže klasa i domenskih razlika.

S obzirom na to da se praćenje pogleda predviđa kao uobičajena značajka u raznim potrošačkim elektroničkim uređajima, uključujući pametne telefone, prijenosna računala i uređaje za virtualnu stvarnost (en. virtual reality headsets), pojavile su se zabrinutosti oko utjecaja ove tehnologije na privatnost korisnika. Pomoću tehnika strojnog učenja (en. machine learning), podaci o praćenju pogleda mogu neizravno otkriti informacije o korisnikovoj etničkoj pripadnosti, osobnim karakteristikama, strahovima, emocijama, interesima, vještinama te fizičkom i mentalnom zdravlju. Ako se takvi zaključci izvode bez korisnikova znanja ili odobrenja, to se može smatrati napadom putem inferencija (en. inference attack). Aktivnosti očiju nisu uvijek pod svjesnom kontrolom, npr. pogledi uzrokovani vanjskim podražajima, širenje zjenica, tremor oka i spontano treptanje uglavnom se događaju bez svjesnog napora, slično kao probava i disanje. Stoga može biti teško za korisnike tehnologije praćenja pogleda procijeniti ili kontrolirati količinu

informacija koje otkrivaju o sebi. U konačnici ostaje važno kontinuirano preispitivati utjecaj ovih tehnologija na društvo.

Literatura

- [1] Wikipedia contributors, *Computer vision*, Wikipedia, The Free Encyclopedia, (2024., rujan). Poveznica: https://en.wikipedia.org/wiki/Computer_vision; pristupljeno 21. kolovoza 2024.
- [2] Wikipedia contributors, *NeoCognitron*, Wikipedia, The Free Encyclopedia, (2024., lipanj). Poveznica: <https://en.wikipedia.org/wiki/Neocognitron>; pristupljeno 21. kolovoza 2024.
- [3] Minarik, G., *NeoCognitron*, (2006). Poveznica: <http://neocognitron.euweb.cz/links/>; pristupljeno 21. kolovoza 2024.
- [4] Mazars, G., *CVPapers - Computer Vision Resource*. Poveznica: <http://www.cvpapers.com/index.html>; pristupljeno 23. kolovoza 2024.
- [5] Se, K., Kuka, V., *The History of Computer Vision on the Path to AGI*, Turing Post, (2024., travanj). Poveznica: <https://www.turingpost.com/p/cvhistory1>; pristupljeno 23. kolovoza 2024.
- [6] Papert, S., *The Summer Vision Project*, Massachusetts Institute of Technology, (1966., srpanj). Poveznica: <https://people.csail.mit.edu/brooks/idocs/AIM-100.pdf>.
- [7] A. Rosenfeld, *Computer vision: basic principles*, in *Proceedings of the IEEE*, (izdanje 76, broj 8), (1988., kolovoz), pp. 863-868. Poveznica: <https://ieeexplore.ieee.org/abstract/document/5961>.
- [8] M. Laad, R. Maurya, N. Saiyed, *Unveiling the Vision: A Comprehensive Review of Computer Vision in AI and ML*, 2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS), Chennai, India, (2024, travanj), str. 1-6. Poveznica: <https://ieeexplore.ieee.org/document/10533631>.
- [9] X. Li, Y. Shi, *Computer Vision Imaging Based on Artificial Intelligence*, 2018 International Conference on Virtual Reality and Intelligent Systems (ICVRIS), Hunan, China, (2018, kolovoz), str. 22-25. Poveznica: <https://ieeexplore.ieee.org/document/8531341>.
- [10] A. K. Sharma *et al.*, *Dermatologist-Level Classification of Skin Cancer Using Cascaded Ensembling of Convolutional Neural Network and Handcrafted Features Based Deep Neural Network*, IEEE Access (izdanje 10), (2022, veljača), str. 17920-17932. Poveznica: <https://ieeexplore.ieee.org/document/9706441>.
- [11] Ansari, S., *Pattern Recognition Introduction*, GeeksforGeeks (2024, travanj). Poveznica: <https://www.geeksforgeeks.org/pattern-recognition-introduction/>; pristupljeno 24. kolovoza 2024.
- [12] Bašić, B., D., Š, Jan, *Strojno učenje*, Uvod u umjetnu inteligenciju, Sveučilište u Zagrebu, Fakultet elektrotehnike i računarstva, (2020.). Poveznica: https://www.fer.unizg.hr/_download/repository/UI-2020-10-StrojnoUcenje.pdf.
- [13] Wikipedia contributors, *Pattern Recognition*, Wikipedia, The Free Encyclopedia, (2024., rujan). Poveznica: https://en.wikipedia.org/wiki/Pattern_recognition; pristupljeno 24. kolovoza 2024.

- [14] Bishop, C. M., *Pattern Recognition and Machine Learning*, Microsoft Research. Cambridge: Springer Science+Business Media, 2006. Poveznica: <https://www.microsoft.com/en-us/research/uploads/prod/2006/01/Bishop-Pattern-Recognition-and-Machine-Learning-2006.pdf>.
- [15] Quora AI, *What is the Difference Between Pattern Recognition and Object Recognition*, (2024., srpanj). Poveznica: <https://www.quora.com/What-is-difference-between-pattern-recognition-and-object-recognition>; pristupljeno 24. kolovoza 2024.
- [16] Wikipedia contributors, *Object Detection*, Wikipedia, The Free Encyclopedia, (2024., kolovoz). Poveznica: https://en.wikipedia.org/wiki/Object_detection; pristupljeno 25. kolovoza 2024.
- [17] Wikipedia contributors, *Deep Learning*, Wikipedia, The Free Encyclopedia, (2024., rujan). Poveznica: https://en.wikipedia.org/wiki/Deep_learning; pristupljeno 25. kolovoza 2024.
- [18] Wikipedia contributors, *Backpropagation*, Wikipedia, The Free Encyclopedia, (2024., rujan). Poveznica: <https://en.wikipedia.org/wiki/Backpropagation>; pristupljeno 25. kolovoza 2024.
- [19] LeCun, Y., Bengio, Y., Hinton, G., *Deep Learning*. Nature, (2015., svibanj). Poveznica: https://www.researchgate.net/publication/277411157_Deep_Learning.
- [20] Sejnowski, T., J., *The Deep Learning Revolution: Machine Intelligence Meets Human Intelligence*, Massachusetts Institute of Technology. Cambridge, Massachusetts: The MIT Press, 2018. Poveznica: [https://aitskadapa.ac.in/e-books/CSE/DEEP%20LEARNING/The%20Deep%20Learning%20Revolution_%20Machine%20Intelligence%20Meets%20Human%20Intelligence%20\(%20PDFDrive%20\).pdf](https://aitskadapa.ac.in/e-books/CSE/DEEP%20LEARNING/The%20Deep%20Learning%20Revolution_%20Machine%20Intelligence%20Meets%20Human%20Intelligence%20(%20PDFDrive%20).pdf).
- [21] Xu, M., Yoon, Y., Fuentes, A., Park, D., S., *A Comprehensive Survey of Image Augmentation Techniques for Deep Learning*, Pattern Recognition (izdanje 137), (2023, svibanj). Poveznica: https://www.sciencedirect.com/science/article/pii/S0031320323000481?ref=pdf_download&fr=RR-2&rr=8b9d15e6dc65c205#bib0022.
- [22] García-García, A., Orts-Escolano, S., Oprea, S., Villena-Martinez, V., Garcia-Rodriguez, J., *A Review on Deep Learning Techniques Applied to Semantic Segmentation*, arXiv, (2017., travanj). Poveznica: <https://arxiv.org/abs/1704.06857>.
- [23] Moharana, P., *Best Object Detection Algorithms*, Analytics India Magazine, (2024., kolovoz). Poveznica: <https://analyticsindiamag.com/topics/best-object-detection-algorithms/>; pristupljeno 26. kolovoza 2024.
- [24] Krizhevsky, A., Sutskever, I., Hinton, G. E., *ImageNet Classification with Deep Convolutional Neural Networks*, Advances in Neural Information Processing Systems (izdanje 25), (2012.). Poveznica: <https://proceedings.neurips.cc/paper/2012/hash/c399862d3b9d6b76c8436e924a68c45b-Abstract.html>.
- [25] He, K., Zhang, X., Ren, S., Sun, J., *Deep Residual Learning for Image Recognition*, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, (2016), str. 770–778. Poveznica: https://openaccess.thecvf.com/content_cvpr_2016/html/He_Deep_Residual_Learning_CVPR_2016_paper.html.

- [26] Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K. Q., *Densely Connected Convolutional Networks*, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, (2017.), str. 4700–4708.
- [27] Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., Adam, H., *Mobilenets: Efficient Convolutional Neural Networks for Mobile Vision Applications*, arXiv, (2017.). Poveznica: [http://refhub.elsevier.com/S0031-3203\(23\)00048-1/sbref0013](http://refhub.elsevier.com/S0031-3203(23)00048-1/sbref0013).
- [28] Zoph, B., Vasudevan, V., Shlens, J., Le, Q. V., *Learning Transferable Architectures for Scalable Image Recognition*, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, Utah, USA, (2018., lipanj), str. 8697–8710. Poveznica: [http://refhub.elsevier.com/S0031-3203\(23\)00048-1/sbref0015](http://refhub.elsevier.com/S0031-3203(23)00048-1/sbref0015).
- [29] Zhang, H., Wu, C., Zhang, Z., Zhu, Y., Lin, H., Zhang, Z., Sun, Y., He, T., Mueller, J., Manmatha, R., *et al.*, *ResNeSt: Split-Attention Networks*, arXiv, (2020.). Poveznica: <https://arxiv.org/abs/2004.08955>.
- [30] Touvron, H., Cord, M., Douze, M., Massa, F., Sablayrolles, A., Jégou, H., *Training Data-Efficient Image Transformers & Distillation Through Attention*, Proceedings of the 38th International Conference on Machine Learning, PMLR, (2021., srpanj), str. 10347–10357. Poveznica: <https://proceedings.mlr.press/v139/touvron21a>.
- [31] Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B., *Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows*, Proceedings of the IEEE/CVF International Conference on Computer Vision, (2021.), str. 10012–10022. Poveznica: https://openaccess.thecvf.com/content/ICCV2021/html/Liu_Swin_Transformer_Hierarchical_Vision_Transformer_Using_Shifted_Windows_ICCV_2021_paper.
- [32] Ren, S., He, K., Girshick, R., Sun, J., *Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks*, IEEE Transactions on Pattern Analysis and Machine Intelligence (izdanje 39, broj 6), (2017., lipanj) str. 1137-1149. Poveznica: <https://ieeexplore.ieee.org/abstract/document/7485869>.
- [33] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., Berg, A. C., *SSD: Single Shot Multibox Detector*, European Conference on Computer Vision, (2016.), str. 21–37. Poveznica: [http://refhub.elsevier.com/S0031-3203\(23\)00048-1/sbref0021](http://refhub.elsevier.com/S0031-3203(23)00048-1/sbref0021).
- [34] *R-CNN: Regions with CNN features*, Medium. Poveznica: https://miro.medium.com/v2/resize:fit:1400/0*EJN-MCrYGXaM3DOA; pristupljeno 21. kolovoza 2024.
- [35] He, K., Gkioxari, G., Dollár, P., Girshick, R., *Mask R-CNN*, arXiv, (2018.). Poveznica: <https://arxiv.org/abs/1703.06870>.
- [36] Facebook Research, Detectron, GitHub. Poveznica: <https://github.com/facebookresearch/Detectron>; pristupljeno 26. kolovoza 2024.
- [37] Redmon, J., Divvala, S., Girshick, R., Farhadi, A., *You Only Look Once: Unified, Real-Time Object Detection*, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, (2016.). Poveznica: https://www.cv-foundation.org/openaccess/content_cvpr_2016/papers/Redmon_You_Only_Look_CVPR_2016_paper.pdf.

- [38] Bochkovskiy, A., Wang, C., Liao, H., *YOLOv4: Optimal Speed and Accuracy of Object Detection*, arXiv, (2020., travanj). Poveznica: <https://arxiv.org/abs/2004.10934v1>.
- [39] Wang, C.-Y., Bochkovskiy, A., Liao, H.-Y., *YOLOv6: A Single-Stage Object Detection Framework for Industrial Applications*, arXiv, (2022., rujan). Poveznica: <https://arxiv.org/abs/2209.02976>.
- [40] Wang, C.-Y., Liao, H.-Y., *YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors*, arXiv, (2022., lipanj). Poveznica: <https://arxiv.org/abs/2207.02696>.
- [41] *YOLOv7: A Powerful Object Detection Algorithm (2024 Guide)*, Viso, (2023., studeni). Poveznica: <https://viso.ai/deep-learning/yolov7-guide/>; pristupljeno 26. kolovoza 2024.
- [42] Mirkhan, A., *YOLO Algorithm: Real-Time Object Detection From A to Z*, Kili Technology. Poveznica: <https://kili-technology.com/data-labeling/machine-learning/yolo-algorithm-real-time-object-detection-from-a-to-z>; pristupljeno 26. kolovoza 2024.
- [43] DeepBean, *How YOLO Object Detection Works*, YouTube. Poveznica: <https://www.youtube.com/watch?v=svn9-xV7wjk>; pristupljeno 26. kolovoza 2024.
- [44] Wikipedia contributors, *Face Detection*, Wikipedia, The Free Encyclopedia, (2024., kolovoz). Poveznica: https://en.wikipedia.org/wiki/Face_detection; pristupljeno 27. kolovoza 2024.
- [45] Wikipedia contributors, *Facial Recognition System*, Wikipedia, The Free Encyclopedia, (2024., kolovoz). Poveznica: https://en.wikipedia.org/wiki/Facial_recognition_system; pristupljeno 27. kolovoza 2024.
- [46] Wikipedia contributors, *Eigenface*, Wikipedia, The Free Encyclopedia, (2024., ožujak). Poveznica: <https://en.wikipedia.org/wiki/Eigenface>; pristupljeno 28. kolovoza 2024.
- [47] Anggo, M., Arapu, L., *Face Recognition Using Fisherface Method*, Journal of Physics: Conference Series, (2018., lipanj). Poveznica: <https://iopscience.iop.org/article/10.1088/1742-6596/1028/1/012119>
- [48] Wikipedia contributors, *Eye Tracking*, Wikipedia, The Free Encyclopedia, (2024., kolovoz). Poveznica: https://en.wikipedia.org/wiki/Eye_tracking; pristupljeno 30. kolovoza 2024.
- [49] Wikipedia contributors, *Gaze-Contingency Paradigm*, Wikipedia, The Free Encyclopedia, (2024., rujan). Poveznica: https://en.wikipedia.org/wiki/Gaze-contingency_paradigm; pristupljeno 30. kolovoza 2024.
- [50] Heatmap, DebuggerCafe. Poveznica: https://debuggercafe.com/wp-content/uploads/2021/05/cam_exmp_2.jpg; pristupljeno 30. kolovoza 2024.
- [51] Michalewski, M., *3 Seconds to Screen Candidate's LinkedIn Profile: Eye Tracking Research Results*, Element. Poveznica: <https://elementapp.ai/blog/3-seconds-to-screen-candidate-profile-biometric-research-results>; pristupljeno 31. kolovoza 2024.
- [52] Corno, F. and Farinetti, L., Signorile, I, *A cost-effective solution for eye-gaze assistive technology*, Proceedings of the IEEE International Conference on

- Multimedia and Expo (izdanje 2), Lausanne, Switzerland, (2002.), str. 433-436.
Poveznica: <https://ieeexplore.ieee.org/document/1035632>.
- [53] Maglio, P., Matlock, T., Campbell, C., Zhai, S., Smith, B., *Gaze and Speech in Attentive User Interfaces*, Lecture Notes in Computer Science, (2001., ožujak).
Poveznica:
https://www.researchgate.net/publication/2363407_Gaze_and_Speech_in_Attentive_User_Interfaces.
- [54] Sibert, L., Jacob, R., Templeman, J., *Evaluation and Analysis of Eye Gaze Interaction*. Poveznica:
https://www.researchgate.net/publication/2635298_Evaluation_And_Analysis_Of_Eye_Gaze_Interaction.
- [55] Galante, A., Menezes, P., *A Gaze-Based Interaction System for People with Cerebral Palsy*, Procedia Technology (izdanje 5), (2012.). Poveznica:
<https://www.sciencedirect.com/science/article/pii/S2212017312005300?via%3Dihub>
- [56] Assistive Communication, *CBoard*. Poveznica: <https://www.cboard.io/>; pristupljeno 01. rujna 2024.
- [57] Papoutsaki, A., Sangkloy, P., Laskey, J., Daskalova, N., Huang, J., Hays, J., *WebGazer: Scalable Webcam Eye Tracking Using User Interactions*. Poveznica:
<http://cs.pomona.edu/~apapoutsaki/papers/ijcai2016webgazer.pdf>.
- [58] Brown WebGazer Team, *WebGazer.js*. Poveznica: <https://webgazer.cs.brown.edu/>; pristupljeno 30. kolovoza 2024.
- [59] Jakovac, J., *Iris Insider*, Github. Poveznica: <https://github.com/OffCrazyFreak/Iris-Insighter>; pristupljeno 30. kolovoza 2024.
- [60] Wikipedia contributors, *Adversarial Machine Learning*, Wikipedia, The Free Encyclopedia, (2024., kolovoz). Poveznica:
https://en.wikipedia.org/wiki/Adversarial_machine_learning; pristupljeno 2. Rujna 2024.
- [61] Apostolidis, K. D., Papakostas, G. A., *A Survey on Adversarial Deep Learning Robustness in Medical Image Analysis*, Electronics (izdanje 10, broj 17), (2021.).
Poveznica: <https://www.mdpi.com/2079-9292/10/17/2132>

Programska aplikacija za praćenje ljudskog pogleda

Sažetak

Ovaj rad istražuje područje računalnog vida s naglaskom na tehnike prepoznavanja uzoraka i praćenja pogleda. U uvodnim poglavljima objašnjavaju se osnovni pojmovi računalnog vida, povijest razvoja, te ključne tehnike kao što su segmentacija, rekonstrukcija scene i obnavljanje slike. U dijelu o prepoznavanju objekata analiziraju se algoritmi poput Mask R-CNN i YOLO, dok se u poglavlju o praćenju pogleda opisuje povijest, tehničke metode te izazovi istoga. Za kraj je dodatno prikazana jednostavna web implementacija praćenja pogleda koristeći biblioteku WebGazer.js koja omogućava praćenje pogleda putem obične kamere.

Ključne riječi: Praćenje pogleda, Prepoznavanje lica, Algoritmi, Duboke neuronske mreže, Interaktivni sustavi, Obrada slike

Application for Eye Movement Tracking

Summary

This paper explores the field of computer vision with a focus on pattern recognition techniques and gaze tracking. The introductory chapters explain the basic concepts of computer vision, its historical development, and key techniques such as segmentation, scene reconstruction, and image restoration. The section on object recognition analyzes algorithms like Mask R-CNN and YOLO, while the gaze tracking chapter discusses its history, technical methods, and challenges. Lastly, the paper presents a simple web implementation of gaze tracking using the WebGazer.js library, enabling gaze tracking through a regular webcam.

Keywords: Gaze tracking, Face recognition, Algorithms, Deep neural networks, Interactive systems, Image processing