SVEUČILIŠTE U ZAGREBU

**FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA**

ZAVRŠNI RAD br. xxxx

**SUSTAV ZA DETEKCIJU I PRAĆENJE DLANA**

Vedran Ćutić

Zagreb, lipanj 2024. SVEUČILIŠTE U ZAGREBU

**FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA**

ZAVRŠNI RAD br. xxxx

**SUSTAV ZA DETEKCIJU I PRAĆENJE DLANA**

Vedran Ćutić

Zagreb, lipanj 2024.

# Uvod

# Detekcija i praćenje objekata

Detekcija i praćenje objekata je proces pronalaženja, klasificiranja i praćenja objekata na slikama i videima. Postupak se izvodi u nekoliko slijednih koraka:

1. Obrada slike i predprocesiranje
2. Ekstrakcija značajki i segmentacija i detekcija
3. Klasifikacija
4. Postprocesiranje
5. Sinteza

## Obrada slike i predprocesiranje

Svaka slika je sačinjena od tisuće piksela koji se u tehnikama klasifikacije tretiraju kao liste matrica, veličine koja ovisi o rezoluciji same slike. Svaki algoritam implementira neke metode koje će poboljšati kvalitete slike i pripremiti ih za daljnju analizu. Slikama se često mijenjaju dimenzije (visina i širina) ili izrezuju samo dijelovi slike radi uklanjanja nebitnih dijelova slike ili radi smanjenja veličine i kompleksnosti. Koristi se metoda normalizacije slike gdje se vrijednost svakog piksela podešava na normalnu distribuciju. Svakom pikselu se oduzima srednja vrijednost i dijeli se sa standardnom devijacijom vrijednosti piksela. Mnoge slike sadrže šum koji može negativno utjecati na efikasnost modela pa je potrebno koristiti neke od tehnika filtriranja slike poput Gaussovog filtriranja ili srednjeg filtriranja. Ove metode pospješuju uspješnost i kvalitetu modela. Nakon prve obrade slike ona je spremna za daljnju analizu.

## Ekstrakcija značajki i segmentacija i detekcija

Ekstrakcija značajki je bitan proces prepoznavanja uzoraka unutar slike koji će se koristiti za razlikovanje jednog objekta od drugog. Uzorci su obično specifični za svaki razred što rezultira jasnom razlikom razreda. U slučaju klasifikacije jabuka i banana postoje značajke poput oblika i boje koje će se koristiti za razlikovanje ta dva razreda. Ekstrakcija značajki uvelike poboljšava performanse modela fokusiranjem na najrelevantnije dijelove svake slike, u suprotnom bi se svaki put trebala analizirati cijela slika što utječe na kompleksnost i točnost modela. Otkrivanje rubova na slici korisna je i bitna metoda jer pomoću nje uočavamo granice između područja na slici. Najčešće metoda otkrivanja rubova na slici je korištenje operatora gradijenta koji određuje di dolazi do prijelaza između elemenata na slici. Analiza teksture analizira sliku i pronalazi ponavljajuće uzorke što se koristi za prepoznavanje različitih materijala ili površina objekata, ovime se prepoznaju tumori na medicinskim slikama jer se tekstura zdravog tkiva razlikuje od teksture kancerogenog tkiva.

Slike se sastoje od informacija bitnih za analizu i onih nebitnih. Koriste se mnogi načini podijele slike na bitne dijelove za analizu čime se rezultati postupka klasificiranja objekata na slici značajno poboljšava. Segmentacija je postupak podijele slike ili videa na područja interesa za identificiranje i razlikovanje objekata ili zanimljivih dijelova. Područja od interesa mogu biti bilo kakvi objekti koji se ističu na slici – ljudi, automobili, zrakoplovi, životinje. U kontekstu detekcije i praćenja dlana dlan i ruka su područja interesa. Najčešći oblik je semantička segmentacija koja svakom pikselu na slici daje oznaku.

Detekcija lokalizira određeni objekt unutar slike često pružajući granične okvire njegove lokacije, što olakšava i ubrzava proces klasifikacije objekta unutar slike. Segmentacija slike daje detaljnu analizu gdje je izlazna slika podijeljena piksel po piksel dok nam detekcija daje granične okvire objekata.

Slika na kojoj se prikazuje vanjski, plaža, voda, pas

Opis je automatski generiran

## Klasifikacija slika i videa

Kako bi model mogao odrediti je li se na slici nalazi dlan, potrebno je odrediti razrede segmentiranih i detektiranih objekata, to se izvodi postupkom *klasifikacije*. Klasifikacija slika i videa zadatak je strojnog učenja da identificira i odredi što slika ili video predstavljaju. Model se trenira na skupu podataka za učenje koji sadrži različite razrede i oznake. Budući da je video skup slika, mnoga se jednaka pravila koriste za klasifikaciju slika kao i za klasifikaciju videa. Slici se dodjeljuje jedna ili više oznaka ili razreda na temelju postojećih podataka za treniranje već označenih slika. Dobivamo rezultate koji nam govore nalaze li se na slici određeni objekti, atributi ili uzorci. Neki od mnogih slučaja upotrebe klasifikacije slika i videa su: Automatska inspekcija i kontrola kvalitete artikala, Prepoznavanje objekata u autonomnim vozilima, Prepoznavanje lica u sigurnosnim snimkama.

Napretkom strojnog učenja u zadnjih nekoliko godina brzo i napreduju modeli klasifikacije i detekcije objekata. Danas, za razliku od prije samo deset godina klasifikacija objekata je brzo izvediva, ne zahtjeva toliko resurse i jednostavna je za implementaciju.

### Vrste klasifikacije slika

Postoje različite metodologije klasifikacije slike koje se mogu koristit u rješavanju širokog opsega zadataka i problema.

* **Binarna klasifikacija** klasificira sliku u jednu od dva razreda. Slika će pripadati ili jednom ili drugom razredu, što ima mnoge i široke primjene, poput prepoznavanja tumora, uočavanje nedostataka određenih objekata ili nešto poput klasifikacije životinje u razrede pas ili mačka.
* **Višerazredna klasifikacija** slična je binarnoj, kategorizira sliku u tri ili više razreda i ima širu primjenu jer se njome slike mogu detaljnije i kompleksnije analizirati i kategorizirati, moguće je napraviti model koji će na temelju slike prepoznati koju to životinju predstavlja, a ne poput binarne samo jednu od dvije moguće klasifikacije.
* **Hijerarhijska klasifikacija** za zadatak ima organizaciju klasa u hijerarhijsku strukturu na temelju njihove sličnosti, moguće je klasificirati životinje na sisavce i ribe te zatim sisavce i ribe dodatno podijeliti na specifičnije članove tog skupa poput lav, tuna, i slično.

### Nadzirano učenje

Nadzirano strojno učenje algoritam je učenja na označenim skupovima podataka, sa već određenim odnosom ulaza i ispravnih izlaza. Skup podataka slika je već označen i klasificiran. Nakon dovršetka učenja, analizom novim ulaza algoritam na temelju podataka za učenje zaključuje i predviđa razrede novih slika. Algoritmi nadziranog učenja dijele se na klasifikaciju s jednim razredom koji slici dodjeljuju samo jedan razred poput dlana ruke i klasifikaciju s više razreda. Mnogi algoritmi implementiraju neku vrstu nadziranog strojnog učenja. Jedan od takvih je stablo odluke koja se lako implementira i široko je primjenjiva. Stablo odluke funkcionira poput dijagrama toka gdje se na svakom čvoru određuje u kojem će se smjeru dalje kretati sve dok se ne dođe do lista koji predstavlja ciljnu vrijednost.

### Nenadzirano učenje

Za razliku od nadziranog učenja, skupovi podataka u algoritmima nenadziranog učenja nisu označeni niti unaprijed klasificirani, ovi algoritmi pronalaze pravilnosti i sličnosti u podacima grupirajući, otkrivajući stršećih vrijednosti i smanjenjem dimenzionalnosti. Cilj nenadziranog učenja je istražiti i razumjeti prirodu danih podataka. Algoritam tumači podatke prema vlastitim uvjetima, prepoznaje uzorke i izvlači zaključke. Najčešći koncept koji koristi nenadzirano učenje je grupiranje podataka na temelju nekih sličnosti i uzoraka kojim dobijemo grupe podataka čiji će se razredi kasnije odrediti .

### Neuronske mreže

Poznato je da se ljudski mozak sastoji od mnoštva međusobno povezani neurona koji rade istovremeno. Izgrađuju se sustavi u*mjetnih neuronskih mreža* koji imitiraju rad ljudskog mozga. Umjetna neuronska mreža (ANN) replika je ljudskog mozga kojom se simuliraju postupci učenja i obrade podataka.

Neuroni su procesni elementi neuronske mreže inspirirani biološkim neuronima u mozgu. Biološki neuron se sastoji od tijela za obradu impulsa, *dendrite* za primanje impulsa i *akson* za prijenos impulsa do ostalih neurona. Arhitektura umjetnog neurona slična je biološkom, sadrži ulazni sloj neurona, skriveni sloj neurona koji predstavljaju tijelo biološkog neurona i izlazni sloj neurona koji računa izlazne vrijednosti.

Funkcionalnost umjetnog neurona temelji se na vrijednostima osjetljivosti i pristranosti svakog od međusobno povezanih neurona u mreži. Vrijednost sa svakog ulaza *xi* u neuron se množi osjetljivošću tog ulaza *wi* i akumulira u tijelu. Ukupnoj sumi dodaje se pomak *w0,* još poznat kao pristranost (eng. bias). Time je definirana akumulirana vrijednost net koja se propušta kroz prijenosnu funkciju čime nastaje izlazna vrijednost.

*Slika na kojoj se prikazuje dijagram, Font, skeč, bijelo

Opis je automatski generiran*

Prijenosna funkcija smanjuje linearnost i omogućuje rješavanje nelinearnih problema. Neke od često korištenih prijenosnih funkcija su funkcija identiteta, funkcija skoka, sigmoidalna funkcija, tangens hiperbolni.

Duboke neuronske mreže proširenje su umjetnih neuronskih mreža. Za razliku od konvencionalnih neuronskih mreža duboke neuronske mreže obično sadrže mnogo skrivenih slojeva i velik broj neurona čiji broj dostiže milijune u nekim kompleksnijim primjerima.

Duboke neuronske mreže su se pokazale da probleme računalnog vida podižu na još višu razinu točnosti i učinkovitosti, zahvaljujući konvolucijskim neuronskim mrežama (CNN), te zbog toga su najčešće rješenje problema klasifikacije slika. Konvolucijska neuronska mreža proširena je inačica umjetne neuronske mreže (ANN) koja se primarno koristi za procesuiranje podataka matričnog oblika, poput slika. CNN-ovi oponašaju neuronske mreže ljudskog uma u više slojeva: slojevi ulaznih podataka, konvolucijski slojevi, ReLU slojevi, slojevi udruživanja i sloj izlaznih podataka.

Konvolucijski slojevi izvode linearno-transformacijske operacije među dvije matrice, gdje je prva matrica filter koji se može naučiti, a druga matrica je ograničeni dio ulazne matrice. Filter klizi po visini i širini slike stvarajući slikovnu reprezentaciju čime se dobiva aktivacijska mapa, dvodimenzionalni prikaz slike. Treniranjem na podacima mreža će naučiti sve filtre.

Sloj ReLU (Rectified Linear Unit) postao je popularan u zadnjih par godina. ReLU sloj se generalno nalazi iza konvolucijskog sloja i on svaku negativnu vrijednost pretvara u nulu.

Sloj udruživanja smanjuje veličinu slika što čini izračun bržim te smanjuje memorijsku i vremensku kompleksnost. Neke od vrsta udruživanja su maksimalno udruživanje i prosječno udruživanje.

# Metoda

Problem detekcije i praćenja objekata na slici i videu nije jednostavan proces. Primjenjivost i efikasnost metoda ovisi o mnogim čimbenicima, na primjer kvaliteta slike, broj sličica po sekundi na videu, šumovi na podacima i slično. Složenost najvećim dijelom proizlazi iz samih uvjeta u kojima se vrši detekcija. Kretnja dlana i ruke nepredvidiva je, brza i često promjenjiva, osim tog i pozadina sama po sebi može biti nepredvidiva. Količina vanjskog svjetla, sjene i slično drastično utječu na efikasnost modela. Spoj ili kombinacija spomenutih svojstava definira uvjete u kojima se vrši detekcija, koje je nužno uzeti u obzir pri odabiru metoda.

Računalni vid danas je glavni alat za složenu analizu slike, prvenstveno u onim slučajevima kada je potrebno postići seciranje na onoj razini na kojoj to radi ljudski mozak. Računalni vid podvrsta je umjetne inteligencije gdje se promatraju vizualni podaci. Raznovrsnim se metodama pokušava postići viša razina razumijevanja vizualnih podataka kako bi se povećala primjenjivost u što većem broju slučajeva. Jedan od najčešćih zadataka računalnog vida je raspoznavanje objekata na slici ili videu.

Danas dolazi do ubrzanog razvoja ovog područja zbog novih specifičnih potreba za raspoznavanjem predmeta, stalnim napretkom, poboljšanjem starih metoda i stalni razvoji novih. Postoje mnogi sustavi temeljeni na odabranim algoritmima koji posjeduju ugrađene vrijednosti čije stanje ovisi o nizu ulaznih podataka u procesu treniranja modela. Ti sustavi su prilagodljivi različitim zadacima i problemima. Neki od poznatijih konkretnih primjera su YOLO (You Only Look Once), TensorFlow i EfficientDet.

U sklopu ovog rada u svrhu rješavanja problema detekcije i praćenja dlana korišten je YOLOv8 model.

## YOLOv8

YOLO (You Only Look Once) vrhunski je model računalnog vida kojeg razvija *Ultralytics,* platforma umjetne inteligencije za stvaranje, treniranje i implementiranje modela strojnog učenja sa sučeljima bez izvornog koda za jednostavnost i praktičnost uporabe. Ultralytics pojednostavljuje proces i pruža rješenja koja se lako implementiraju.

Arhitektura dostiže veliki ugled i raširenost zahvaljujući brzini raspoznavanja s relativno visokom točnosti. Ova svojstva omogućavaju primjenu u pravom vremenu i kvalitetnu analizu videozapisa. YOLO arhitektura se razlikuje od ostalih jer na prepoznavanje predmeta gleda kao na problem regresije do prostorno odvojenih poznatih okvira i pridruženih vjerojatnosti mogućih klasa. Tijekom analize se promatra čitava slika što pospješuje preciznost jer se predviđanja temelje na globalnom kontekstu.

YOLOv8 ažurirana je verzija popularnog YOLO algoritma za raspoznavanje objekata, koji je predstavljen 2016. Model sadrži brojne arhitektonske i razvojne promjene i poboljšanja u odnosu na prijašnje verzije. Utvrđeno je kako uspijeva pronaći i klasificirati objekte različitih razreda na slikama s visokom točnošću.