SVEUČILIŠTE JOSIPA JURJA STROSSMAYERA U OSIJEKU

FAKULTET ELEKTROTEHNIKE, RAČUNARSTVA I

INFORMACIJSKIH TEHNOLOGIJA OSIJEK

Sveučilišni diplomski studij

DETEKCIJA LICA RAZLIČITIM METODAMA

Raspoznavanje uzoraka i strojno učenje

Karla Fehir

Osijek, 2023.

# SADRŽAJ

[SADRŽAJ 2](#_Toc128085953)

[1. UVOD 1](#_Toc128085954)

[2. PREGLED PODRUČJA I PROBLEMATIKE 2](#_Toc128085955)

[1. 1. Viola-Jones algoritam 2](#_Toc128085956)

[1.2. Single Shot Detection (SSD) 5](#_Toc128085957)

[1.2.1. VGG-16 5](#_Toc128085958)

[1.3. Mediapipe 7](#_Toc128085959)

[3. OPIS ZADATKA S DOBIVENIM REZULTATIMA 8](#_Toc128085960)

[3.1. Korištene tehnologije i alati 8](#_Toc128085961)

[3.1.1. Python 8](#_Toc128085962)

[3.1.2. Streamlit 8](#_Toc128085963)

[3.2. Implementacija 8](#_Toc128085964)

[3.3. Evaluacija metoda za detekciju lica 10](#_Toc128085965)

[3.3.1. Usporedba brzina izvođenja metoda 10](#_Toc128085966)

[3.3.2. Preciznost i odziv metoda 11](#_Toc128085967)

[4. ZAKLJUČAK 14](#_Toc128085968)

[5. LITERATURA 15](#_Toc128085969)

# 1. UVOD

Detekcija lica lak je vizualni zadatak za ljudski vid, međutim ovaj zadatak nije lak i smatra se izazovom za sustav koji se temelji na računalnom vidu jer ima visok stupanj varijabilnosti u svom izgledu. Kako računala mogu otkriti više ljudskih lica prisutnih na slici ili videu sa složenom pozadinom? To je problem. Rješenje ovog problema uključuje segmentaciju, ekstrakciju i provjeru lica i crta lica iz složene pozadine.

U ovom projektnom zadatku će se istražiti i usporediti različite metode za detekciju lica korištenjem Python biblioteka i model duboke neuronske mreže, te će se njihov rezultat odnosno točnost i brzina algoritma implementirati na Web aplikaciju korištenjem Streamlit platforme.

.

# 2. PREGLED PODRUČJA I PROBLEMATIKE

## 1. 1. Viola-Jones algoritam

Algoritam za detekciju objekata Viola-Jones [1] je pristup strojnog učenja za detekciju objekata, koji su 2001. godine predložili Paul Viola i Micheal Jones s ciljem rješavanja problema efikasnosti detekcije lice koja tada još nije bila dovoljno razvijena. Ovaj se algoritam može trenirati za detekciju gotovo bilo kojeg objekta, ali prvenstveno rješava problem detekcije lica u stvarnom vremenu. Evaluacija algoritma sastoji se od četiri faze:

1. Odabir Haarovih značajki

Viola-Jones algoritam temelji se na posebnim značajkama koje klasificiraju slike za detekciju. Algoritam koristi četiri vrste Haarovih značajki, kao što je prikazano na slici 1.1. Vrijednost svake značajke dobiva se oduzimanjem ukupne vrijednosti piksela u značajci koji se nalaze ispod svake vrste pravokutnika. Na primjer, značajke označene slovom "A" na slici 1.1. oduzima se ukupna vrijednost svih piksela ispod bijelog pravokutnika od ukupne vrijednosti piksela ispod crnog pravokutnika. Na slici 1.2. U prikazana je specifična primjene na crtama lica. Ako značajku označenu slovom "C" primijenimo na područje nosa bilo kojeg lica, dobit ćemo slične rezultate jer razlika u ukupnim vrijednostima piksela ispod pravokutnika koji prekriva sredinu nosa i dva pravokutnika koji prekrivaju lijevu i desne regije su slične.

A picture containing square

Description automatically generated

Slika 1.1. Prikaz Haarovih značajki

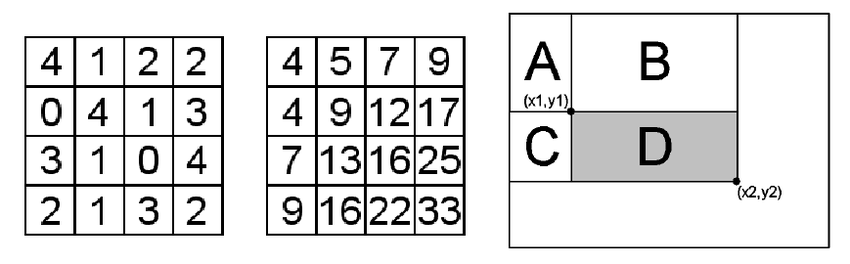
A picture containing text, television, screen

Description automatically generated

Slika 1.2. Primjena Haarovih značajki na licu

2. Integralni prikaz slike

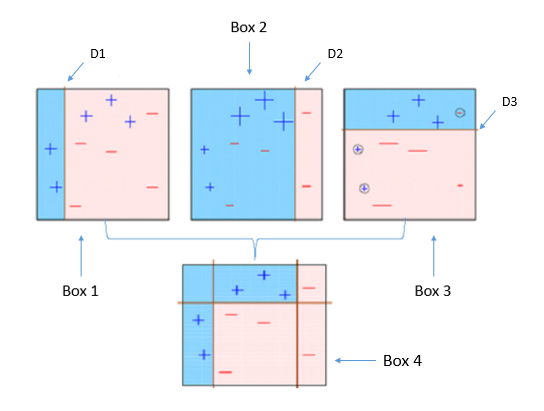
Vrijednost bilo koje točke na integralnoj slici je zbroj svih piksela iznad i lijevo od te točke. Integralna slika može se učinkovito izračunati u jednom prolazu preko slike. Na slici 1.3. nalazi se integralni prikaz slike.



Slika 1.3. Integralni prikaz slike

3. AdaBoost algoritam

Za okvir od 24x24 piksela može postojati oko 162 336 mogućih značajki koje bi bile vrlo skupe za procjenu. Stoga se AdaBoost algoritam koristi za obuku klasifikatora samo s najboljim značajkama. Na slici 1.4. prikazan je princip rada AdaBoost algoritma.



Slika 1.4 Prikaz AdaBoost algoritma

4. Arhitektura kaskadnog klasifikatora

Kaskadni klasifikator odnosi se na ulančavanje nekoliko klasifikatora poredanih uzastopnim redoslijedom. Sve bitne značajke detekcije grupirane su u nekoliko takvih faza u kojima svaka od njih evaluira određeni broj značajki. Zadaća svake od faza je da za određeni potprozor odredi da li se u njemu možda nalazi lice ili se sigurno ne nalazi. Detekcija lica biti će uspješna samo ako potprozor prođe sve faze kaskade klasifikatora tj. kada ga kaskada označi pozitivnim. Struktura kaskadnog klasifikatora prikazana je na slici 1.5.

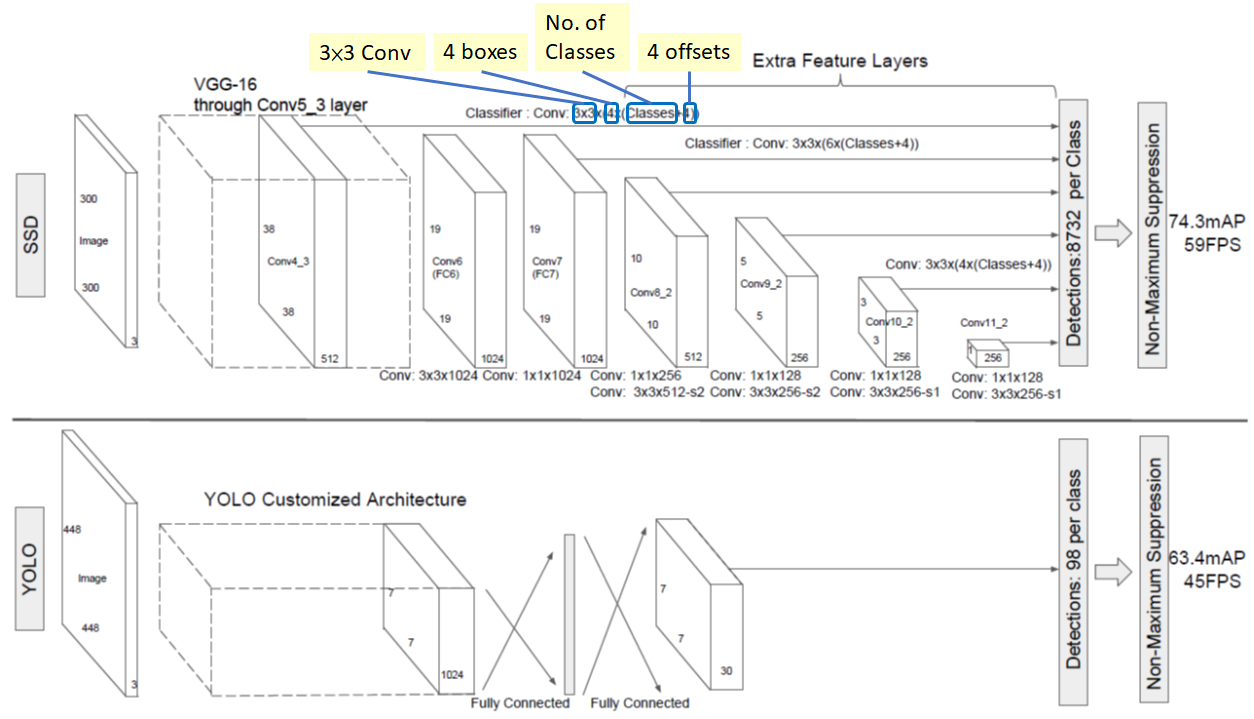
Diagram

Description automatically generated

Slika 1.5 Prikaz strukture kaskadnog klasifikatora

## 1.2. Single Shot Detection (SSD)

SSD (Single Shot Multibox Detector) [2] je tehnika koja se koristi za detekciju objekata na slikama pomoću jedne duboke neuronske mreže. Princip SSD-a je da ulazna slika prvo prođe kroz konvolucijsku neuronsku mrežu zbog izdvajanja mape značajki. Koristeći osnovnu arhitekturu VGG-16 arhitekture, SSD može nadmašiti druge detektore objekata kao što su YOLO i Faster R-CNN i u brzini i u točnosti. SSD obavlja detekciju objekata u samo jednom koraku za razliku od navedenih metoda. Usporedba arhitekture SSD i YOLO detektora prikazana je na slici 1.6.



Slika 1.6. Usporedba arhitektura SSD i YOLO metode

### 1.2.1. VGG-16

VGG-16 [3] je 16-slojna konvolucijska neuronska mreža, stoga je relativno opsežna mreža s ukupno 138 milijuna parametara - to je ogromno čak i prema današnjim standardima. Međutim, jednostavnost VGGNet16 arhitekture je njegova glavna atrakcija. VGGNet arhitektura uključuje najvažnije značajke konvolucijske neuronske mreže. VGG mreža sastoji se od malih konvolucijskih filtara. VGG-16 ima tri potpuno povezana sloja i 13 konvolucijskih slojeva. Arhitektura VGG mreže prikazana je na slici 1.7 i može se opisati na sljedeći način:

1. Ulaz - VGGNet prima ulaznu sliku veličine 224×224.

2. Konvolucijski slojevi - konvolucijski filtri VGG-a koriste najmanje moguće receptivno polje od 3×3. VGG također koristi 1×1 konvolucijski filtar kao ulaznu linearnu transformaciju.

3. ReLu aktivacija - sljedeća je komponenta Rectified Linear Unit Activation Function (ReLU), AlexNetova glavna inovacija za smanjenje vremena treniranja. ReLU je linearna funkcija koja daje odgovarajući izlaz za pozitivne ulaze i daje nulu za negativne ulaze. VGG ima postavljeni korak konvolucije od 1 piksela kako bi se očuvala prostorna rezolucija nakon konvolucije (vrijednost koraka odražava koliko se piksela "pomakne" filter da pokrije cijeli prostor slike).

4. Skriveni slojevi - svi skriveni slojevi VGG mreže koriste ReLU umjesto lokalne normalizacije odgovora kao AlexNet.

5. Slojevi udruživanja - Sloj udruživanja slijedi nekoliko konvolucijskih slojeva—ovo pomaže smanjiti dimenzionalnost i broj parametara mapa značajki stvorenih svakim korakom konvolucije. Objedinjavanje je ključno s obzirom na brzi rast broja dostupnih filtara sa 64 na 128, 256 i na kraju 512 u završnim slojevima.

6. Potpuno povezani slojevi - VGGNet uključuje tri potpuno povezana sloja. Prva dva sloja imaju po 4096 kanala, a treći sloj ima 1000 kanala, po jedan za svaku klasu.

Diagram

Description automatically generated

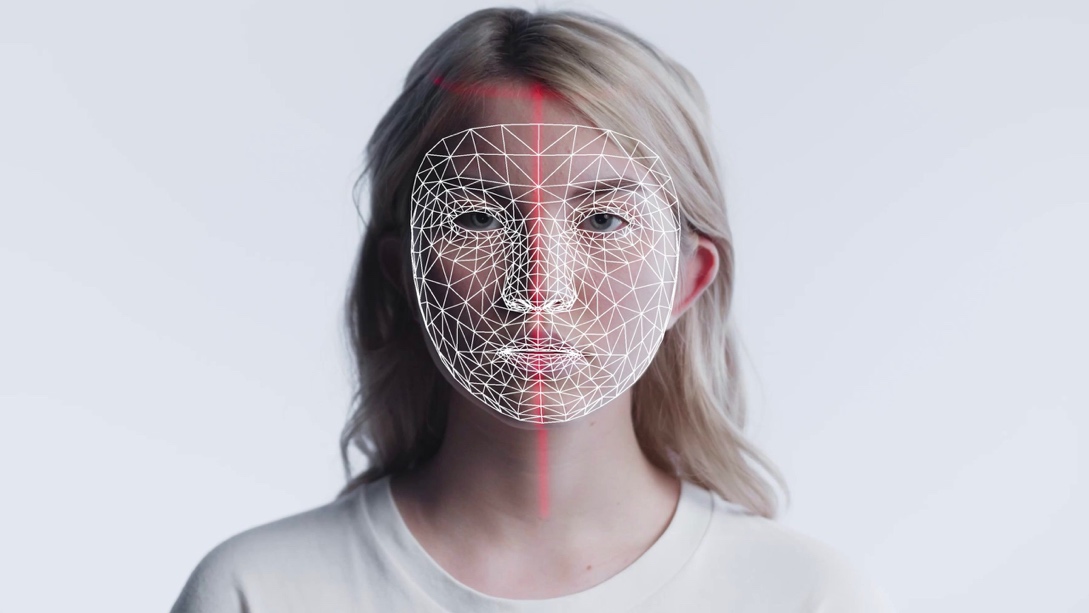
Slika 1.7. Arhitektura VGG-16 mreže

## 1.3. Mediapipe

Za detekciju lica ističe se Mediapipe biblioteka. MediaPipe[4] je Googleov projekt koji nudi opensource ML rješenja. Drugim riječima, MediaPipe pruža pristup širokom spektru snažnih modela strojnog učenja optimiziranih za rad na različitim platformama bez obzira na njihove performanse. Mediapipe se temelji na BlazeFace modelu kojeg je razvio Google i koji je vrlo efikasan model za detekciju lica.

Prvi korak je inicijalizacija modela iz mediapipe biblioteke što zahtijeva odabir tipa modela i postavljanje minimalne razine pouzdanosti detekcije. Razlikujemo dvije vrste modela, model 0 optimiziran je za detekciju lica do 2 metra udaljenosti od kamere, a model 1 za detekciju lica do 5 metara udaljenosti. Tijekom implementacije za detekciju odabran je model 1. Odabrani model se ne može mijenjati bez pristupa programskom kodu. Za detekciju lica mediapipe koristi ključne točke lica koje pokušava pronaći unutar svakog okvira. Ako postoji lice unutar provjerenog okvira, mediapipe funkcija vraća listu koja sadrži granični okvir svih pronađenih lica. Na slici 1.8 prikazane su ključne točke kod detekcije lica korištenjem Mediapipe biblioteke.

A picture containing text

Description automatically generated

Slika 1.8. Media pipe model za detekciju lica

# 3. OPIS ZADATKA S DOBIVENIM REZULTATIMA

## 3.1. Korištene tehnologije i alati

### 3.1.1. Python

Projekt je izrađen u programskom jeziku Python. Za implementaciju detektora lica koriste se biblioteke OpenCV i Mediapipe koje su vrlo poznate u području računalnog vida.

### 3.1.2. Streamlit

Streamlit[5] je Python biblioteka koja olakšava stvaranje i dijeljenje prilagođenih web aplikacija za strojno učenje i podatkovnu znanost. Omogućava jednostavnu izradu i implementaciju podatkovnih aplikacija. Na slici 3.1 prikazana je Streamlit aplikacija projektnog zadatka.

Graphical user interface, application, Teams

Description automatically generated

Slika 3.1. Streamlit aplikacija

## 3.2. Implementacija

Kao prvi algoritam za detekciju lica koristiti ćemo Haar Cascade s unaprijed treniranim XML datotekama koje implementaciju čine iznimno jednostavnom. Haar Cascade ne zahtijeva puno računanja; stoga je popularan za male uređaje s malom računskom snagom.

Biblioteka OpenCV upravlja spremištem na GitHubu za sve popularne Haar Cascade unaprijed trenirane datoteke koje se mogu koristiti za razne zadatke detekcije objekata, kao što su detekcija ljudskog lica, očiju, nosa, detekcija vozila i slično.

Text

Description automatically generated

Slika 3.2. Haar Cascade metoda za detekciju lica

Kao drugi algoritam koristit ćemo Single Shot Detector. Treniranje SSD modela od nule će zahtijevati puno podataka, pa možemo upotrijebiti unaprijed trenirane modele (Caffe Face Detector Model) koristeći OpenCV. Prikaz funkcije za detekciju lica korištenjem SSD algoritma prikazan je na slici 3.3.

Caffe je *framework* za duboko učenje koji su razvili Berkeley AI Research i suradnici. Caffe je razvijen kao brža i daleko učinkovitija alternativa drugim *framework*-ovima za izvođenje detekcije objekata.

Text

Description automatically generated

Slika 3.3. Single Shot Detector metoda za detekciju lica

Kao zadnji algoritam koristit ćemo Mediapipe model za detekciju lica. Prikaz funkcije za detekciju lica korištenjem Mediapipe biblioteke prikazana je na slici 3.4.

Text

Description automatically generated

Slika 3.4. Mediapipe metoda detekcije lica

## 3.3. Evaluacija metoda za detekciju lica

### 3.3.1. Usporedba brzina izvođenja metoda

Kako bi smo usporedili performanse svakog algoritma odnosno vrijeme potrebno za detekciju lica, računalni smo vrijeme izvođenja svake metode. Kako bismo ispitali i usporedili brzine navedenih metoda, testirali smo istih 50 slika na sva tri detektora te na kraju ispisali vrijeme izvođenja algoritma u sekundama. U tablici 3.1. prikazani su rezultati vremena izvođenja detektora.

Tablica 3.2. Rezultati vremena izvođenja detektora lica

|  |  |  |  |
| --- | --- | --- | --- |
| **Naziv detektora** | **Vrijeme izvođenja [s]** | **Prosječno vrijeme [s]** | **Standardna devijacija** |
| Haar Cascade | 12.52471 | 7,8451219333 | 3.36119 |
| Mediapipe | 4.7825098 |
| SSD | 6.228146 |

### 3.3.2. Preciznost i odziv metoda

Kako bismo u potpunosti ispitali kvalitetu ovih metoda u detekciji lica na slikama, koristit ćemo slike koje bi mogle izazvati problem u detekciji zbog varijacija u osvjetljenju, perspektive slikanja, izraza lica i slično.

Za izračun preciznosti i odziva, koristi se broj točnih pozitivnih detekcija (True Positive - TP), lažnih negativnih detekcija (False Negative - FN), lažnih pozitivnih detekcija (False Positive - FP) i točnih negativnih detekcija (True Negative - TN). Ako je na slici prikazano lice i treba ga detektirati te je detektirano, tada to predstavlja točnu pozitivnu detekciju. Ako je na slici prikazano lice i treba ga detektirati, a nije detektirano, tada to predstavlja lažnu negativnu detekciju. Ako je na slici na kojoj nije prikazano lice detektirano lice, tada to predstavlja lažnu pozitivnu detekciju. Ako na slici na kojoj nije prikazano lice nije detektirano lice, tada to predstavlja točnu negativnu detekciju. Za evaluaciju nije potrebno koristiti točno negativne detekcije s obzirom da one predstavljaju svaki dio slike na kojemu se ne nalazi lice i nije detektirano lice.

Kako bi ispitali preciznost detektora, koristiti ćemo 6 slika za koje je očekivan problem u detekciji zbog osvjetljenja, perspektive i izraza lica. U tablici 3.2. prikazani su rezultati testiranja detektora lica.

Tablica 3.2. Rezultati testiranja detektora lica

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Naziv detektora** | **Broj traženih lica** | **Broj detektiranih lica** | **Broj TP detekcija** | **Broj FN detekcija** | **Broj FP detekcija** |
| Haar Cascade | 50 | 42 | 40 | 10 | 2 |
| Mediapipe | 50 | 42 | 42 | 8 | 0 |
| Single Shot Detector | 50 | 35 | 35 | 15 | 0 |

Na temelju tablice 3.1. možemo izračunati preciznost i odziv detektora. Preciznost je udio točno klasificiranih primjera u skupu pozitivno klasificiranih primjera te se računa prema formuli 3-1.

(3-1)

Odziv je udio točno klasificiranih primjera u skupu svih pozitivnih primjera te se računa prema formuli 3-2.

(3-2)

Navedenim formulama računamo preciznost i odziv za sva tri detektora. Preciznost i odviz za Haar Cascade metodu iznosi:

Preciznost i odviz za Mediapipe metodu iznosi:

Preciznost i odviz za Single Shot Detector metodu iznosi:

Na slikama 3.5., 3.6. i 3.7. prikazani su rezultati i usporedba metoda za različite perspektive lica i osvjetljenja lica.

|  |  |  |
| --- | --- | --- |
| Haar Cascade | SSD | Mediapipe |
| A picture containing text, dark  Description automatically generated | A group of people with face paint  Description automatically generated with low confidence | A picture containing text, dark  Description automatically generated |

Slika 3.5. Usporedba preciznosti metoda za tamno osvjetljenje

|  |  |  |
| --- | --- | --- |
| Haar Cascade | SSD | Mediapipe |
| A collage of a person  Description automatically generated with medium confidence | A collage of a person's face  Description automatically generated | A collage of a person's face  Description automatically generated |

Slika 3.6. Usporedba preciznosti metoda za različite perspektive lica

|  |  |
| --- | --- |
| 0 | Haar Cascade |
| 0 | SSD |
| 0 | Mediapipe |

Slika 3.7. Usporedba preciznosti metoda za različita osvjetljenja

# 4. ZAKLJUČAK

Cilj ovog projektnog zadatka bio je upoznavanje s principom rada tehnologije detekcije objekata tj. detekcije lica na fotografijama. Kroz projekt susreli smo se s najpoznatijim metodama u području računalnog vida, poput Viola-Jones algoritma te dubokih neuronskih mreža. Svaka od navedenih metoda ima svoj jedinstveni pristup u detekciji objekata, te po tome se razlikuju i u brzini izvođenja i preciznosti. Detekcija lica u neograničenim uvjetima godinama je predstavljalo problem zbog raznih varijacija izraza lica, svjetline i rubova u boji te se pokazalo kako su novije metode poput Mediapipe puno preciznije u takvim uvjetima u odnosu na starije metode kao što je Haar Cascade.

# 5. LITERATURA

[1] Viola Jones Algoritam, <https://medium.datadriveninvestor.com/haar-cascade-classifiers-237c9193746b>

[2] Single Shot Detector, <https://www.cs.unc.edu/~wliu/papers/ssd.pdf>

[3] VGG-16 Arhitektura, <https://datagen.tech/guides/computer-vision/vgg16/>

[4] Mediapipe, <https://google.github.io/mediapipe/solutions/face_detection>

[5] Streamlit, <https://docs.streamlit.io/>