**UNIVERZITET SINGIDUNUM**

**Tehnički Fakultet**

Upotreba metoda dubokog učenja u sintezi sistema za prepoznavanje lica

- diplomski rad -

**Mentor: Kandidat:**

**prof. dr  *Milan Milosavljević Davor Jordačević***

**Beograd, *2020*.**

Sadržaj

[Uvod - 3 -](#_Toc49721053)

[Neophodni koncepti - 5 -](#_Toc49721054)

[1.1 2D Konvolucija - 5 -](#_Toc49721055)

[1.2 1x1 Konvolucija - 5 -](#_Toc49721056)

[1.3 Višeklasna unakrsna entropija (eng. Cross-entropy) funkcija gubitka - 6 -](#_Toc49721057)

[1.4 MobileNetV2 - 6 -](#_Toc49721058)

[1.4.1 Depthwise Separable Konvolucija - 7 -](#_Toc49721059)

[Detekcija lica - 9 -](#_Toc49721060)

[1.5 RetinaNet - 9 -](#_Toc49721061)

[1.6 RetinaFace - 10 -](#_Toc49721062)

[Poravnanje lica - 12 -](#_Toc49721063)

[Ekstrakcija vektora obeležja - 13 -](#_Toc49721064)

[Pretraga vektora obeležja (prepoznavanje) - 15 -](#_Toc49721065)

[Implementacija u programskom jeziku Python - 18 -](#_Toc49721066)

[Zaključna razmatranja - 29 -](#_Toc49721067)

[Literatura - 30 -](#_Toc49721068)

# Uvod

Glavni problem koji razmatram u ovom radu jesu komponente koje čine jedan sistem za prepoznavanje lica. Još od ranih dana sa pojavom kamera, pojavila se potreba za sistemima koji bi mogli da identifikuju osobe na njima. Tokom vremena su se razne metode smenjivale, od prepoznavanja pokreta, otisaka prstiju, do prepoznavanja lica. U ovom radu će biti reči upravo o tehnologijama koje se koriste za prepoznavanje lica, konrektno, fokusiraćemo se na tehnike dubokog učenja u sintezi sistema za prepoznavanje lica. Ovakvi sistemi su složeni i predstavljaju sintezu raznih tehnologija i metoda kako bi uspešno radili. Sam razvoj ovih sistema je zahtevan, i zahteva tim ljudi koji su sposobni za rešavanje kako matematičkih, tako i računarskih problema. Ovi problemi proizilaze iz potrebe za visokom tacnošću sistema, kao i od varijacija usled koriscenja razne opreme i alata. Svakoga dana se sve vise softvera zasniva upravo na ovoj tehnologiji, a to podrazumeva njenu ispravnost, robusnost i tačnost. Jedna od tehnika koja je omogućila nagli razvoj ove oblasti jesu duboke neuronske mreže. Pored ovoga, ubrzan razvoj tehnologija i računarskih komponenti omogućili su znatno brži razvoj i treniranje ovakvih mreža, o čemu će biti reči u nastavku ovog rada.

Glavni cilj mog istraživanja u ovom radu jeste sinteza sistema za prepoznavanje lica koristeći već postojeće metode za detekciju, ekstrakciju vektora obeležja, i njihovo upoređivanje radi dobijanja željenih rezultata.

U svom radu koristiću analitičke metode kako bi svaku celinu razlozio na delove i bolje objasnio. Sam sistem za prepoznavanje lica je jedna složena celina koja uključuje delove koji su se godinama razvijali i istrazivali. Ovo znači da se kombinacijom ovih podsistema mogu dobiti novi sistemi sa specifičnom namenom ili performansama. Ovi delovi u produkucionom sistemu ne mogu da rade jedan bez drugog, dok je prilikom istraiživanja moguće preskočiti neke od njih. Uzmimo za primer detekciju lica. Ukoliko je cilj sistema samo prepoznavanje, a za testiranje, razvoj i upotrebu se koriste slike koji sadrže samo detektovana lica, onda se sam korak detekcije moze preskočiti. S obzirom na to da ovo često nije slučaj, fokusiraćemo se na sintezu kompletog sistema. Treba uzeti u obzir da ovakav sistem mogu činiti i dodatne komponente, poput dela za predikciju da li je osoba stvarna ili lažna (face anti-spoofing).

U cilju testiranja performansi koristićemo podatke koji nisu često korišćeni, ali su javno dostupni. Reč je o slikama koje na prvi pogled čoveku deluju tesko za prepoznavanje.

Analiza lica predstavlja jedan od bitnih procesa u našim životima. Ljudi analizom lica prikupljaju bitne podatke o drugim osobama. Ovo uključuje podatke o broju godina, polu, rasnoj pripadnosti. Takodje možemo prepoznati da li je osoba srećna ili tužna, ili pak neku drugu emociju. Pokreti usana su vazni u oblasti prepoznavanja govora, kao i sve popularnijoj oblasti kao sto je generisanje lažnih snimaka. Metode analize lica nam mogu reći gde je usmeren pogled neke osobe, odnosno šta privlači njenu pažnju, i ovo može biti posebno interesantno u marketingu i šopovima, kazinima. U medicini ove metode mogu biti od koristi za prepoznavanje nekih bolesti, poput autizma koji se odlikuje time što osobe imaju poteškoće da iskazu svoje emocije. Sve navedene metode koriste kako ljudi, tako i računari.

U ovom radu ćemo se fokusirati na metode koje se koriste u procesu detekcije lica, njegove ekstrakcije, obrade i zatim prepoznavanja.

Na početku ovog rada je prvo bitno da uvrdimo sta je zapravo prepoznavanje lica. Svaka osoba ima karakteristično lice, i to je ono što nas čini unikatnima. Većna metoda se bazira upravo na ovoj činjenici, i njihov cilj je ekstrakcija ovih obelezja (features) za svaku osobu, a zatim i njihova klasifikacija na osnovu određenih parametara.

Treba razlikovati algoritme za prepoznavanje po vise kriterijuma. Osnovna klasifikacija je na algoritme zasnovane na geometriji (Geometry based) i algoritme zasnovane na šablonima (template based). Geometrijski zasnovani algoritmi analiziraju određena područja i geometrijske veze na njima. Zbog ovoga su i poznati kao algoritmi zasnovani na obelezjima. Sa druge strane su algoritmi zasnovani na šablonima, i u ovu grupu spadaju: metoda nosećih vektora (SVM), analiza glavnih komponenti (PCA), linearna diskriminantna analiza (LDA), kernel metode i još mnogo drugih.

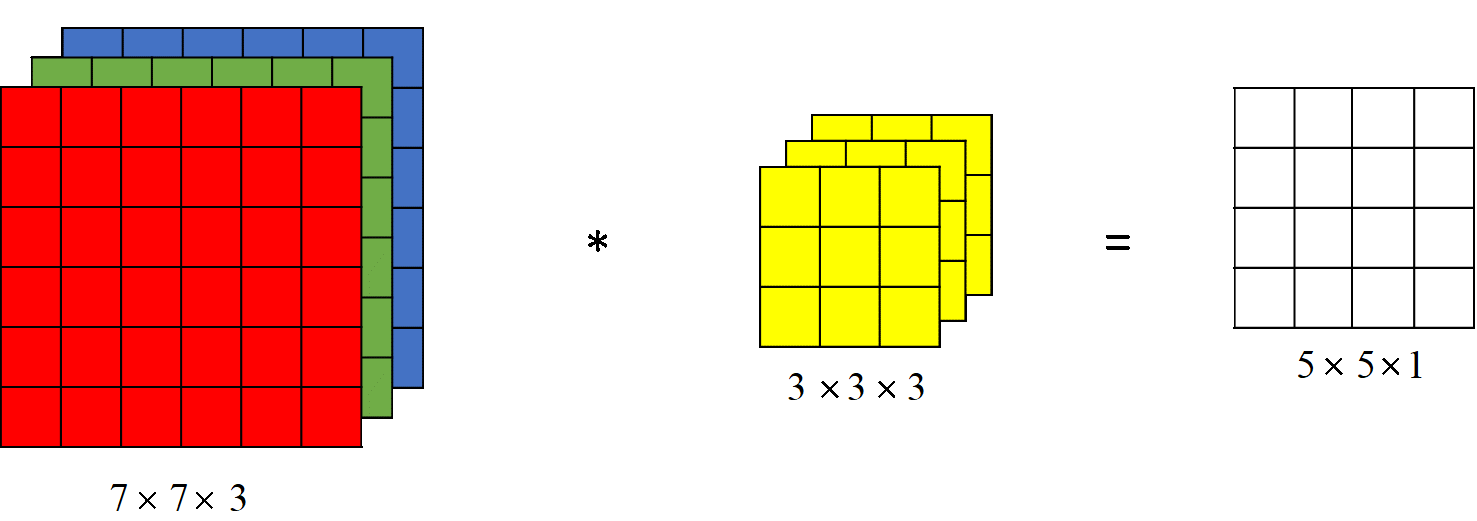
# Neophodni koncepti

Arhitekture dubokih neuronskih mreža o kojima će biti reči su zasnovane na već postojećim principa i njihovom usavršavanju. Premda su ovi principi dosta izmenjeni, u dobrom delu sistema ostaju isti.

## 2D Konvolucija

Konvolucija predstavlja matematički operator koji od dve funkcije proizvodi treću. U oblasti dubokog učenja se pod konvolucijom misli na primenu niza filtara na ulaznu sliku, kako bi se dobio željeni izlaz. Ukoliko želimo da sliku veličine Du x Du x M pretvorimo u izlaz dimenzija Dv x Dv x N, potrebno nam je N filtara, svaki dimenzija Dr x Dr x M.

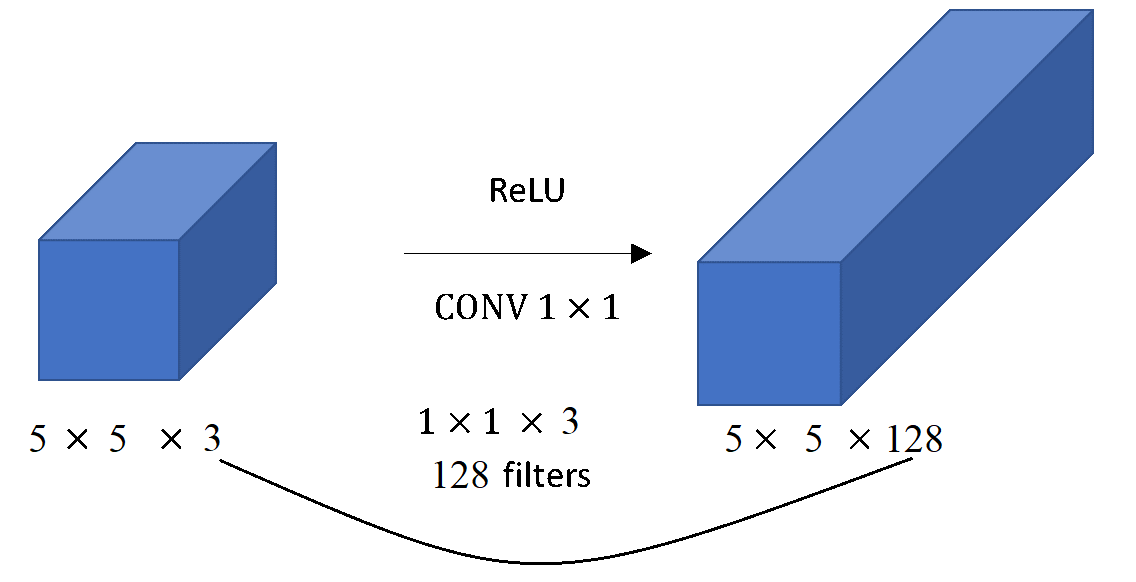
Veličina slike na izlazu iz mreže je (Du - Dr + 2P) / S + 1, gde P predstavlja padding, a S stride, dok je bias 1.



*Slika 1. Primer 2D konvolucije*

## 1x1 Konvolucija

Na prvi pogled se čini da 1x1 konvolucija nema smisla. Ovo je zapravo samo konvolucija sa filtrom velicine 1x1. U praksi, ako primenimo konvoluciju koristeći N filtra dimenzija 1x1 na sliku ulazne velicine Du x Du x M, rezultat je slika dimenzija Du x Du x N.



*Slika 2. Primer 1x1 konvolucije*

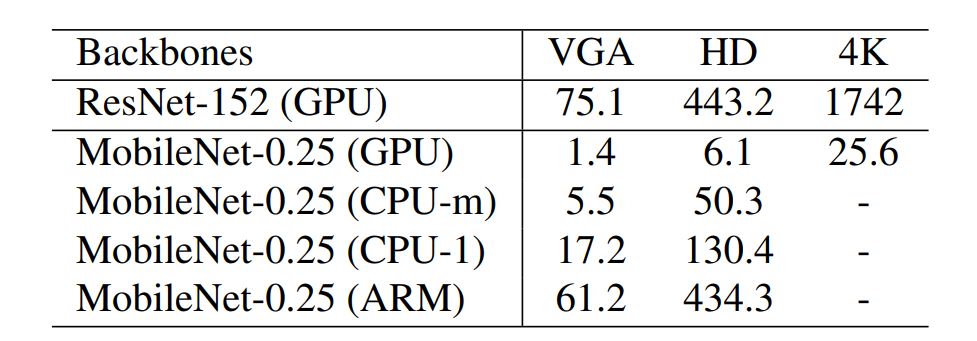
## Višeklasna unakrsna entropija (eng. Cross-entropy) funkcija gubitka

Unakrsna entropija se često koristi kao funkcija gubitka za optimizaciju klasifikacionih modela. Često se naziva i Softmax funkcija gubitka (eng. softmax loss). Predstavlja Softmax aktivacionu funkciju u vezi sa unakrsnom entropijom.

## MobileNetV2

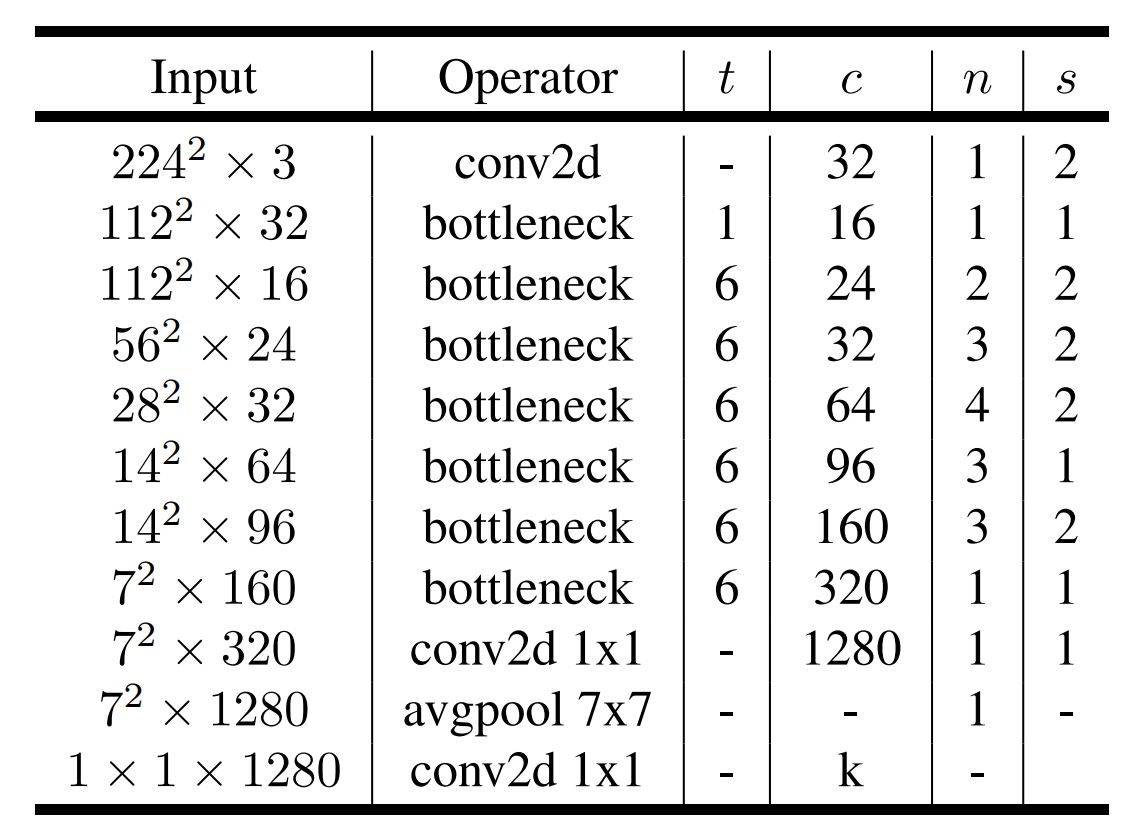
Modul za detekciju u ovom sistemu kao backbone mrežu koristi MobileNetV2, isto kao i modul za ekstrakciju vektora obeležja.

Po rečima autora RetinaFace (Deng, arXiv:1905.00641v2) modela, MobileNetV2 arhitektura postiže drastično kraće vreme inference kada je u pitanju detekcija lica.



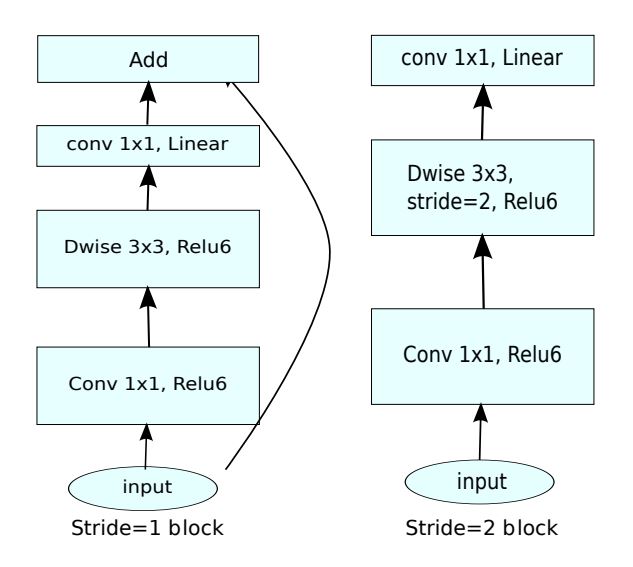
*Slika 3. Komparacija vremena inference RetinaFace modela za detekciju lica*

Arhitektura MobileNetV2 mreže je prikazana na slici.



*Slika 4. Arhitektura MobileNetV2 modela (Sandler, Mark , arXiv:1801.04381v4)*

Kao što se može videti na slici, MobileNetV2 arhitektura je zasnovana na botleneck slojevima. Botleneck sloj predstavlja kombinaciju 1x1 konvolucije sa Relu6 aktivacijom, 3x3 depthwise konvolucije sa Relu6 aktivacijom i linearne 1x1 konvolucije.

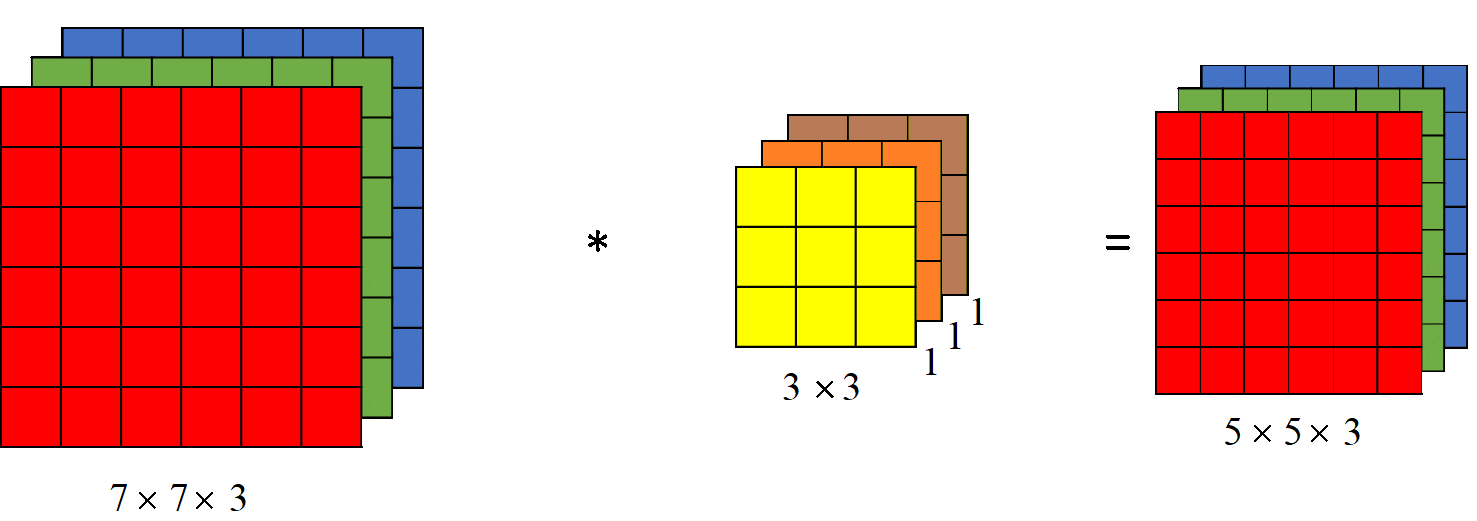


*Slika 5. Arhitektura botleneck slojeva u MobileNetV2 mreži*

### Depthwise Separable Konvolucija

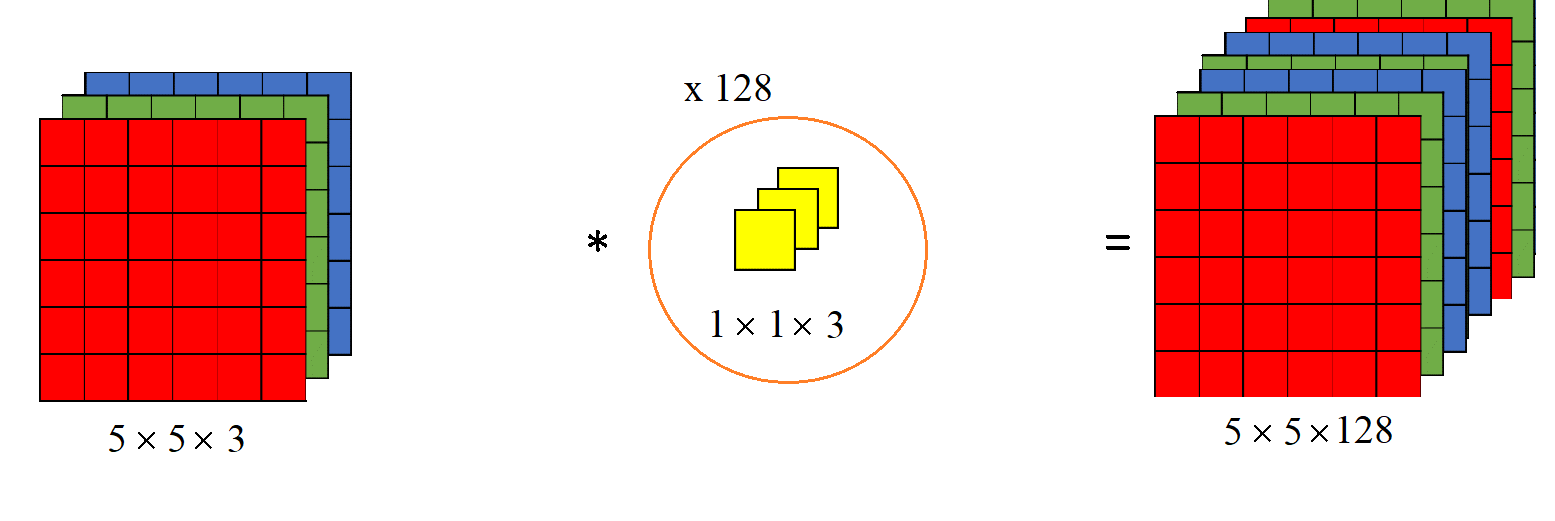
Potreba za pobljašanjem performanski dovela je i do nove vrste konvolucije. U prethodnom primeru sa 2D konvolucion smo mogli videti kako nas može dovesti od ulaznih dimenzija 7x7x3 do 5x5x1. Ukoliko bi primenili 128 filtra dimenzija 3x3x3, ovo bi nas dovelo do izlaza dimenzija 5x5x128. U nastavku će biti prikazano kako se ovo može postići korišćenjem deptwise separable konvulucije.

Umesto da koristimo jedan filter dimenzija 3x3x3, možemo iskoristiti 3 kernela odvojeno. Svaki filter veličine 3x3x1. Svaki kernel vrši konvoluciju sa jednim ulaznim kanalom slike, dajući izlaz dimenzija 5x5x1. Spajanjem ovih izlaza, ponovo dobijamo dimenzije 5x5x3.



*Slika 6. Primer separable konvolucije*

Kako bi dobili željene dimenzije, sledeći korak je primena 1x1x3 filtra. Primenom ovog filtra na ulaz dimenzija 5x5x3, dobijamo izlaz dimenzija 5x5x1. Primenom 128 ovakvih filtra, dobijamo izlaz dimenzija 5x5x128.



*Slika 7. Primer 1x1 konvolucije sa 128 filtra*

# Detekcija lica

Prvi korak u procesu prepoznavanja lica je njegova detekcija. Osnovna ideja procesa detekcije je pronalaženje svih lica na slici, odnosno njihovih koordinata. Ove koordinate služe za ekstrakciju lica sa slike, te se u nastavku radi samo sa slikama koje sadrže lice. Takođe se vrlo često koordinate čuvaju ukoliko je reč o radu sa video snimcima radi iscrtavanja. Pored ovog, čuvanjem originalne slike i koordinata (bounding box-a ili bbox-a) imamo mogucnost ponovne ekstrakcije lica sa raznim scale faktorima, što moze biti korisno kao ulaz u neke druge neuronske mreže. Primer ovoga su anti-spoofing mreže koje imaju za cilj predikciju da li je osoba na slici ili snimku prava (real) ili lažna (fake).

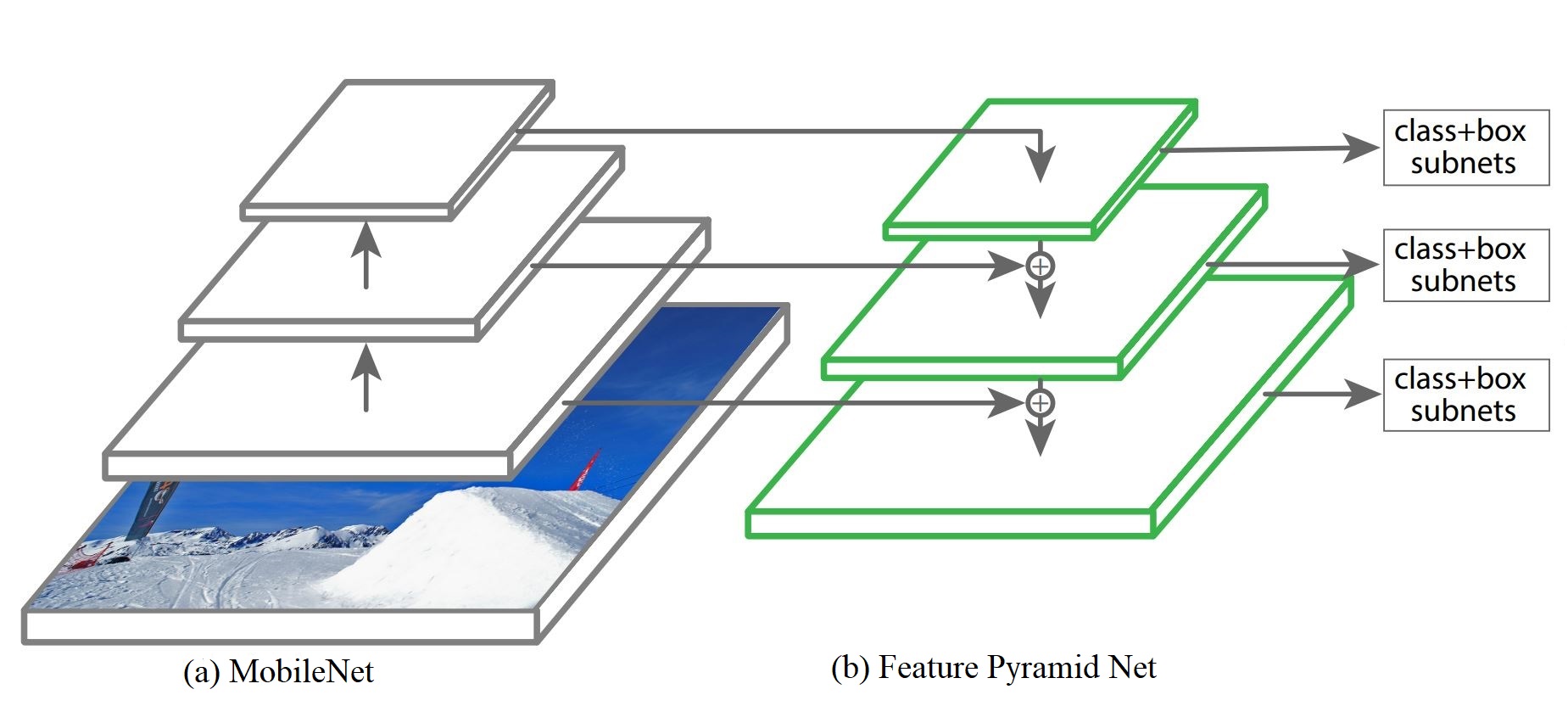
## RetinaNet

Postoji zaista veliki broj razvijenih ideja i projekata na temu detekcije, kako objekata, tako i lica. Jedna implementacija se izdvaja od drugih, reč je o RetinaNet mreži (Lin, Tsung-Yi, arXiv:1708.02002v2).

Osnovna ideja sa RetinaNet detektorom je bila u uraditi sve predickije u jednoj fazi. U 2019. godini, state-of-the-art (SOTA) detektori su bili bazirani na mehanizmu sa dve faze (takozvani two-stage). Prva faza je podrazumevala generisanje skupa kandidata za lokaciju na kojoj je moguće naći traženi objekat, dok se druga faza sastoji iz klasifikacije svakog kandidata u jednu od klasa.

RetinaNet je jednofazni (single stage) detektor koji se sastoji iz jedne mreže (bacbone) i dve mreze sa specifičnom funkcijom. Kao backbone mreža se moze koristiti bilo koja od poznathih arhitektura (VGG, MobileNet, ResNet) i osnovni cilj ove mreže je računanje konvolucione feature mape. Nakon toga, prva podmreža radi klasifikaciju na osnovu izlaza backbone mreže, dok druga podmreža proračunava bounding box regresiju.

Pored toga što je ovaj detektor single stage, podrzava koncept piramida. RetinaNet je zasnovan na konceptu piramida obeležja (Feature Pyramid Network - FPN) (Tang, Xu, arXiv:1803.07737v2). Ovo je omogućeno korišćenjem FPN mreže kao backbone mreže. Osnovni princip rada je da FPN mreža radi augmentaciju standardne konvolucione mreže sa vrha ka dnu sa lateralnim konekcijama (bočne veze).  Ovo omogućava mreži da efikasno konstruise piramidu sa različitim scale faktorima iz jedne slike. Svaki nivo piramide se može koristiti za detektovanje objekata u različitoj razmeri.



Slika 8 –RetinaNet arhitektura koja koristi FPN backbone mrežu preko MobileNet arhitekture

## RetinaFace

Jedan od velikih problema koji se pojavljivao je bila detekcija lica različitih veličina u nekotrolisanim uslovima (prirodi, gužvama u gradu). Pošto već poznati koncept piramida može reštiti probleme detektovanja sa različitim scale faktorima, sledeća mreža koristi iste principe. RetinaFace (Deng, Jiankang, arXiv:1905.00641v2) je mreža nešto novijeg porekla i zasnovna je na istim principima na kojima se zasniva i RetinaNet, ali je namenjena isključivo detekciji lica.

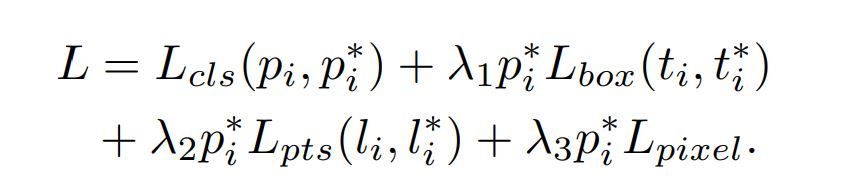
RetinaFace predstvalja single stage detektor lica. Tvorci ovog modela su uneli nove ili unapredili već postojeće metode korišćene u ovu svrhu, kao sto su multi-task obucavanje za istovremenu predikciju uverenja, bounding box-a, 5 ključnih tačaka na licu, i 3D poziciju (u originalnoj implementaciji).

Ono što je novo je 5 ključnih tačaka (keypoint-a) koji ce se kasnije koristiti za poravnanje lica.

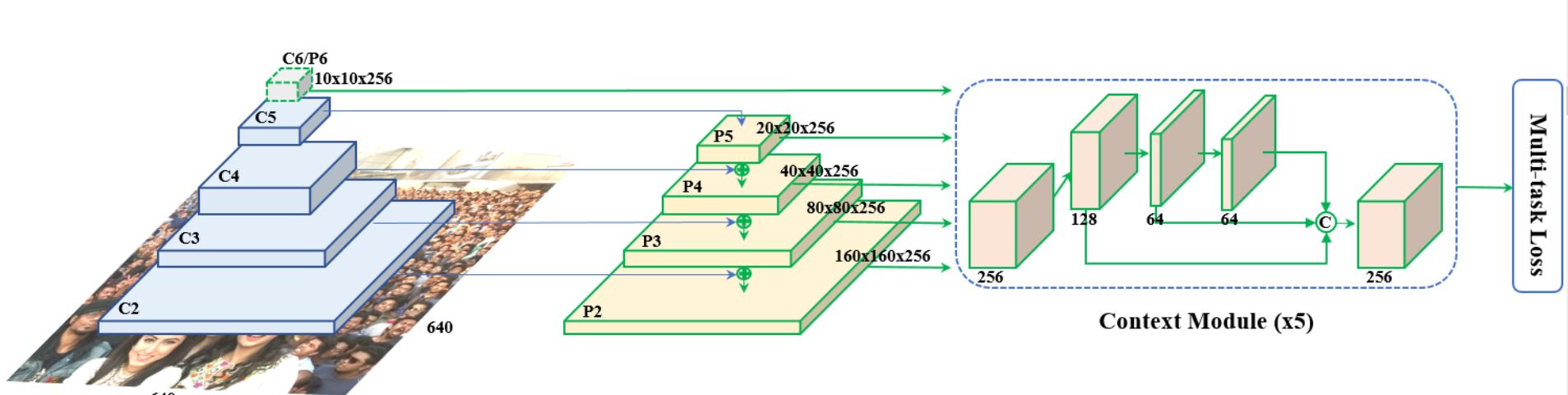
Kako bi se poboljšala detekcija lica, ili takozvanih Hard detekcija, korišćen je koncept modelovanja konteksta. Takozvana Hard lica su teska za detekciju zbog nedostatka vizualne konzistentosti, pozicije ili konteksta. Osnovna ideja je da mreža može da nauči ne samo obeležja koja su karakteristična za lice, već i kontekstualni deo kao što su vrat ili telo.

Kako bi se povećali efekti modelovanja nelinearnih (ne-rigidnih) transformacija (scaling, shearing) korišćeni su kontekst moduli. Geometrijske varijacije predstavljaju jedan od velikih problema u oblasti detekecije i prepoznavanja. Metoda koja se pokazala korisnom u prevazilazenju ovih problema je koriscenje deformabilne (eng. deformable)  konvolucije (Dai, Jifeng, arXiv:1703.06211v3).

Kao što je već pomenuto, RetinaFace mreža je zasnovana na pricipu multi-tast obučavanja. Samim tim se nameće koriscenje drugacije funkcije gubitka (loss funkcije). Loss funkcija koriscena u ovom slucaja je multi-task loss:

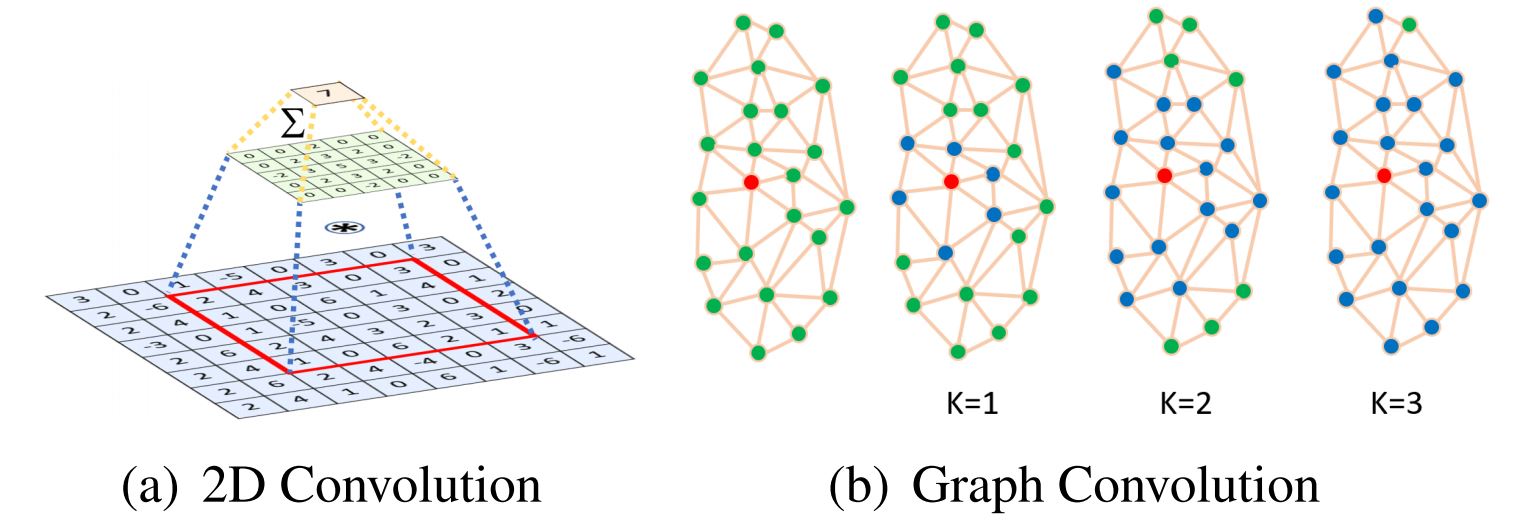


Ova funkcija se sastoji iz više delova, gde prvi deo *Lcls* predstavlja softmax gubitak binarne klasifikacije (ima lica/nema lica), drugi sabirak je gubitak regresije bounding box-ova, zatim sledi regresioni gubitak za predikciju 5 ključnih tačaka i dense regresioni gibitak.



*Slika 9. Prikaz RetinaFace arhitekture. (Levo) Backbone mreža (MobileNetV2). (Sredina) Feature Pyramid Netword. (Desno) Konteks modul.*

Kako bi ubrzao proces detekcije korišćen je takozvani mesh dekoder (mesh konvolucija i up-sampling). Ovo predstavlja vid graf konvolucionog metoda.

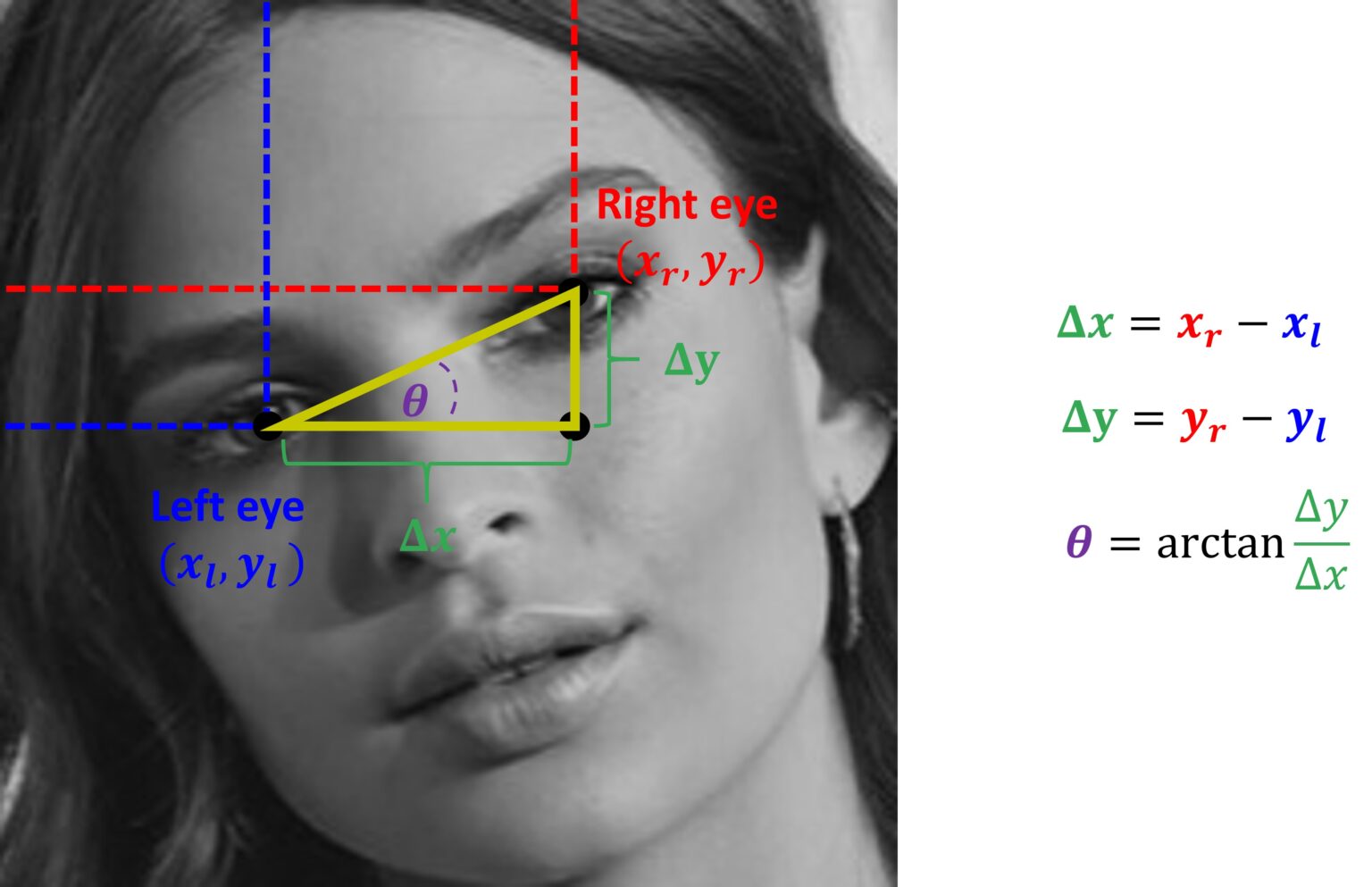


*Slika 10. (Levo) 2D konvolucija. (Desno) Graf konvolucija.*

# Poravnanje lica

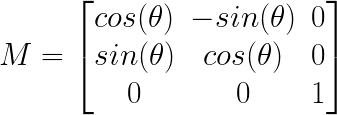
Nakon procesa detekcije i ekstrakcije lica i 5 ključnih tačaka, kako bi proces prepoznavanja bio što uspešniji, potrebno je uraditi poravnanje lica.

Jedan od jednostavnijih ali uspešnih metoda za ovo je pronalaženje arcus tangensa između dva oka (odnosno ugla između dva oka).



*Slika 11. Primer nalaženja ugla izmedju očiju preko arkus tangensa*

Ovde je bitno napomenuti da arctan funkcija u Numpy paketu vraća ugao u radijanima. Za dalju upotrebu je potrebno pretvoriti ga u stepene. Za ovo je samo potrebno pomnožiti sa 10 i podeliti sa PI. Nakon ovoga, potrebno je izracunati rotacionu matricu.



*Slika 12. Primer rotacione matrice*

Sledeći korak je upotreba afinih transformacija kako bi postigli željeni efekat. Ovakva vrsta transformacija, odnosno preslikavanja preslikva tačke u tačke, prave u prave, ravni u ravni. Kod ovakvih transformacija, par paralelnih pravi ostaje paralelan i nakon preslikavanja, ali uglovi između pravih ili razdaljine između tačaka ne moraju nužno da ostanu isti.

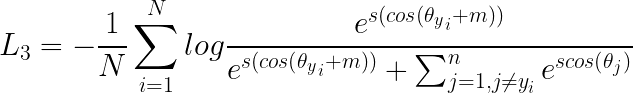
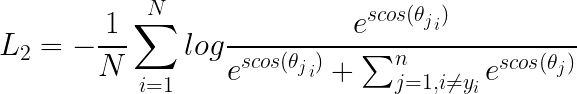
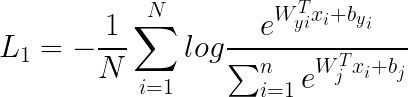
# Ekstrakcija vektora obeležja

Nakon što imamo sliku poravnanog lica, sledeći korak u procesu prepoznavanja je ekstrakcija vektora obelezja (feature vektora). Tokom godina su se smenjivale razne implementacije sa istom namenom, ali sa različitim metodama ekstrakcije vektora, kao  i veličine vektora. Ranije SOTA implementacije poput FaceNet mreže su bile zasnovane na 128-dimenenzionalnim vektorima. Novije metode imaju mogućnost ekstrakcije 512-dimenzionalnih obelezja, pa čak i 1024. U ovom slučaju, zadržaćemo se na 512.

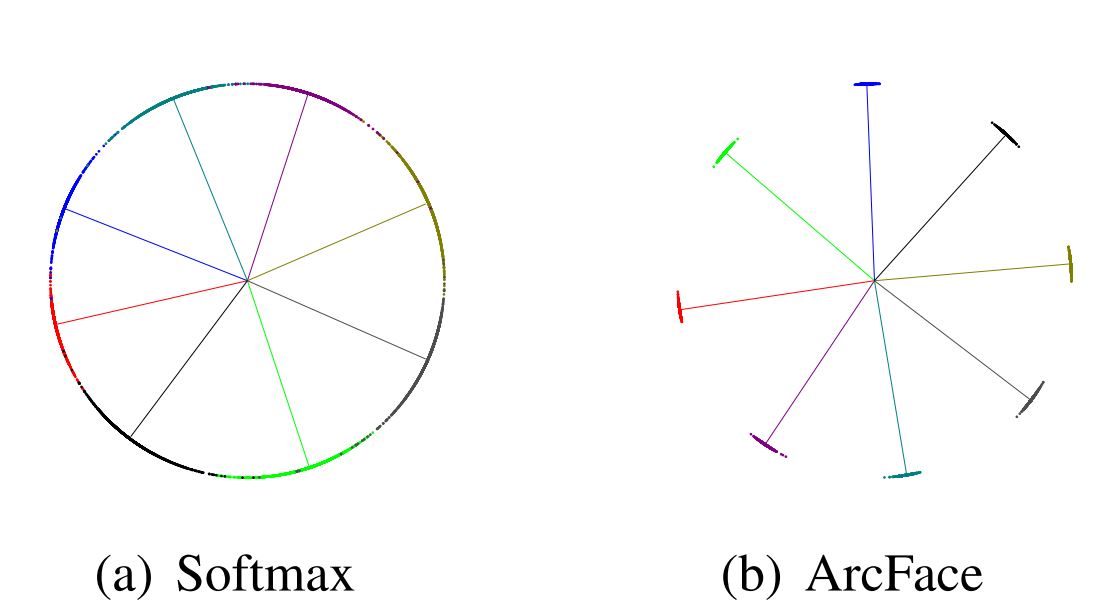
Metode poput FaceNet (Schroff, Florian, Dmitry Kalenichenko, and James Philbin, arXiv:1503.03832v3) mreže su za cilj imale direktno učenje vektora obeležja zasnovano na triplet loss funkciji. Ideja je minimiziranje distance između vektora iste osobe (positive), dok se vektori druge osobe (negative) udaljavalju. Ovo je podrazumevalo da u svakom trenutku tokom treninga imamo tri vektora (anchor, positive, negative). Proces odabiranja tripleta je jako zahtevan i spor, što je bio prvi nedostatak ovakvog i sličnih metoda.

Jedna od implementacija koja je postigla SOTA rezultate je ArcFace (Deng, Jiankang, arXiv:1801.07698v3). Osnovna razlika i ideja je bila napraviti klasifikator koji može da razdvoji različite identitete u trening setu na osnovu određene loss funkcije, kao i korišćenje 512-dimenzionalnih vektora obeležja. Problem sa triplet loss obučavanjem je i eksponencijalni skok broja kombinacija kod velikih setova podataka, dok je problem sa tradicionalnim funkcijama poput Softmax funkcije sto se linearna transformaciona matrica povecava linearno, što nije problem kod manjih setova podataka, ali je kod velikih setova ili produkcionih sistema neupotrebljivo.

Kako bi se povećala margina između klasa, predstavljena je nova funkcija gubitka. Rec je o ArcFace funkciji. U nastavku sledi transformacija iz Softmax u ArfFace funkciju.

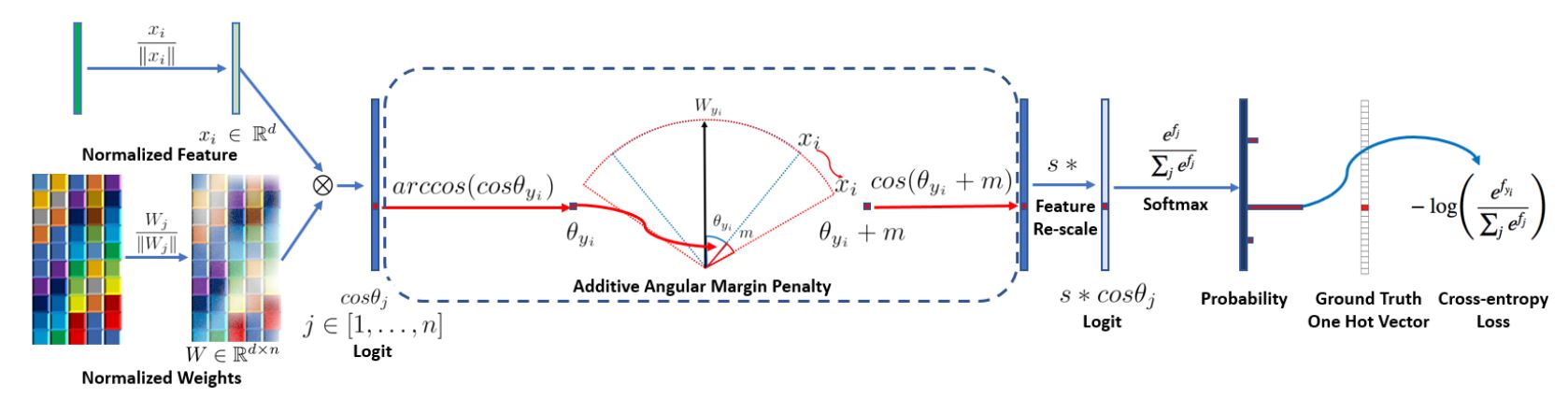


Na sledećoj slici je prikazano kako izgleda separacija 8 klasa korišćenjem Softmax i ArcFace funkcjie.



*Slika 13. Primer separacije 8 klasa korišćenjem Sotfmax i ArcFace funckije gubitka*

Kao backbone mreza je u originalnom radu koriscen ResNet, ali u nasem slucaju je izabrana lightweight arhitektura pod nazivom MobileNet.



*Slika 14. Proces treninga mreže korišćenjem ArcFace funkcije*

# Pretraga vektora obeležja (prepoznavanje)

Ranije imeplemetnacije sličnih sistema su koristile razne metode pretrage. Od najjednostavnihij poput brute force pretrage, k najblizih suseda (KNN), metode nosećih vektora (SVM), kd-trees, LSH, pa i neuronskih mreža za pretragu vektora i predickiju identitieta.

Pomenute metode su se pokazale kao spore, ili nedovoljno precizne. Stoga, bilo je potrebe za novim i bržim metodama pretrage velikih skupova podataka. Metoda koja se izdvaja je pretraga aproksimiranih k najblizih suseda korišćenjem Hierarchical Navigable Small World grafova (HNSW) (Malkov, Yury A., and Dmitry A. Yashunin, arXiv:1603.09320v4).

KNN algoritam za svaki element iz skupa podataka prvo definiše razdaljinu od susednih elemanta. Ta razdaljina može da predstavlja rastojanje između dva vektora u n dimenzionalnom prostoru i može se koristiti zeljena metrika. Parametar k biramo sami i on predstavlja broj suseda sa minimalnom razdaljinom od željene tačke, ili u našem slučaju vektora. Setovi podataka i njihove razmere u ovoj oblasti su veliki, kao i njihova dimenzionalnost. Stoga su metode kao što je linearno skeniranje spore. Treba napomenuti da sve novije metode ne garantuju vracanje korektnog rezultata osim takozvane exhaustive pretrage. Ovo je poznato kao “curse of dimensionality” problem.

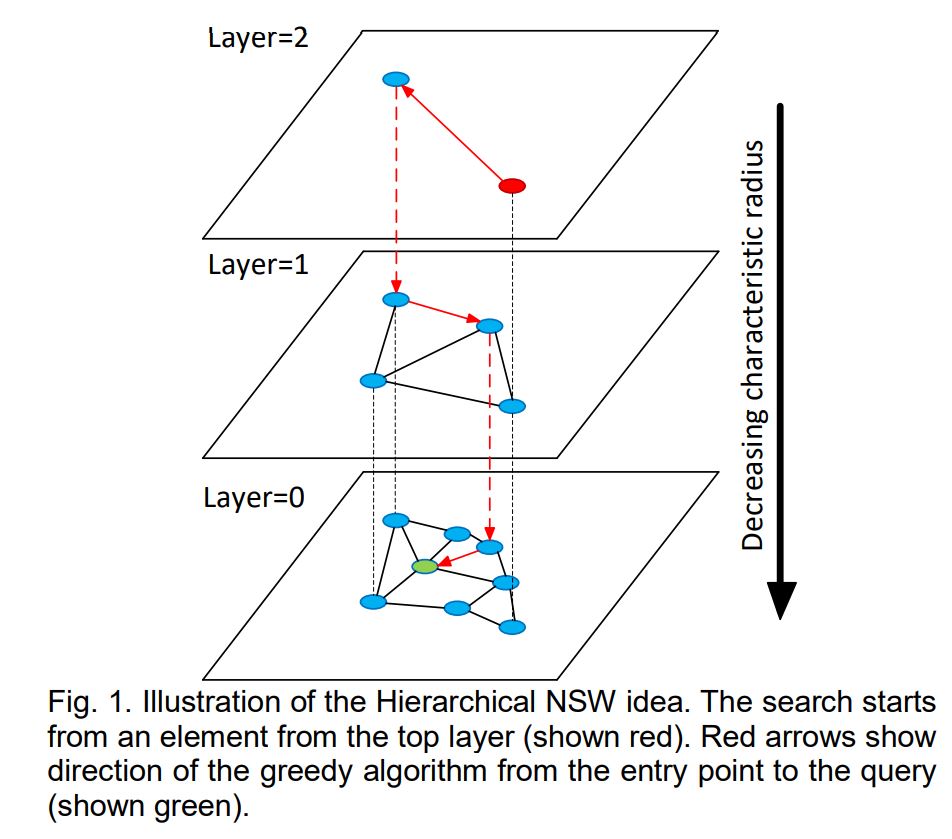
Kako bi se ovaj problem rešio, pojavljuje se metoda pod nazivom aproksimirani k najblizih suseda (Approximate Nearest Neighbors - ANN). Ova metoda dozvoljava mali broj grešaka, a kvalitet pretrage je definisan kao odnos tačno pronađenih suseda i parametra k.

Kako bi se ubrzala pretraga ANN algoritmom korišćene su tehnike transformacije vektora (smanjenje dimenzija, rotacija) pre indeksiranja, ali i enkodovanje vektora.

Vektori se mogu enkdovati koristeći stabla (Annoy), LSH, kvantizaciju, ali i grafove.

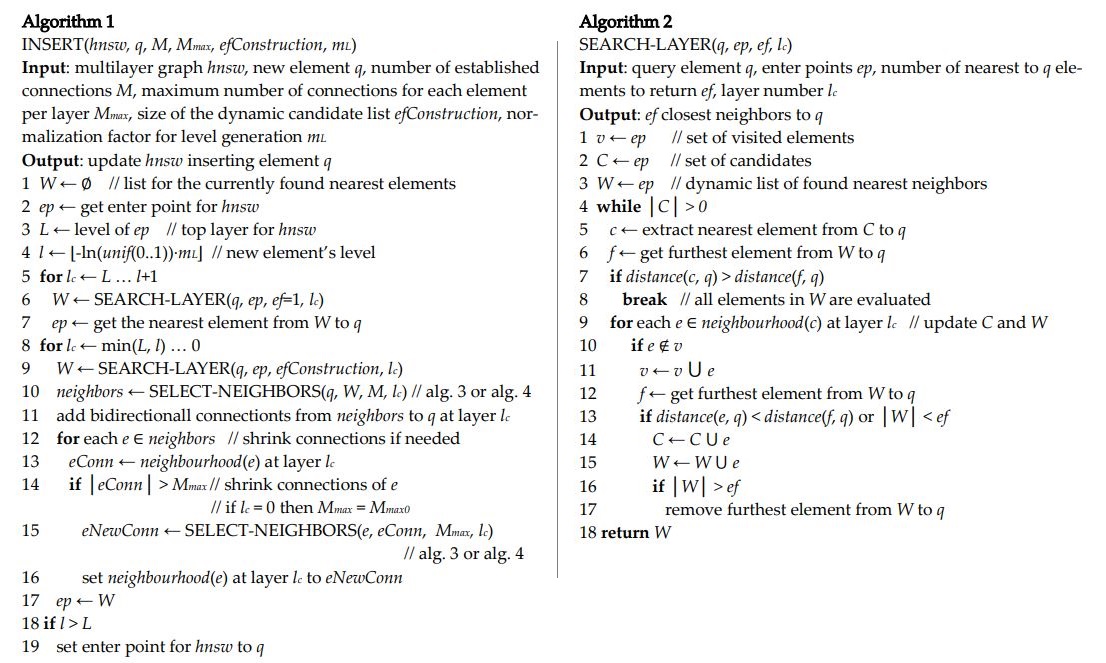
HNSW predstavlja implementaciju graf inkrementalnog ANN algoritma.

Teorija iza NSW i HNSW algoritma je previše obimna za ovaj rad, stoga će u nastavku ukratko biti predstavljeni algoritami.



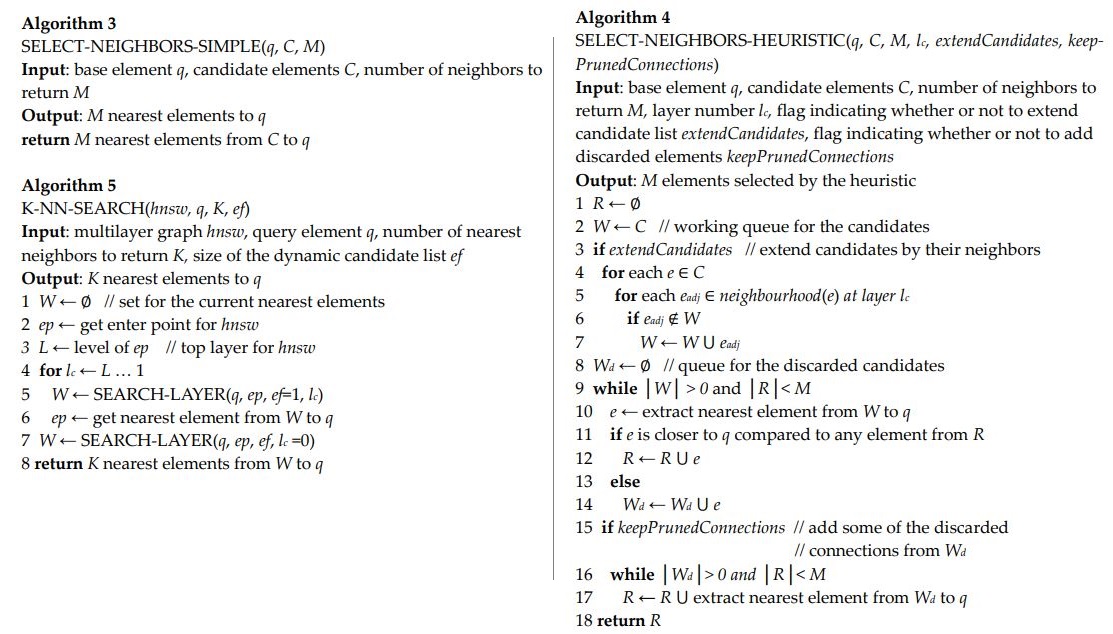
*Slika 15. Vizualizacija procesa pretrahe HNSW grafa*

Na slici 16 prikazani su algoritmi dodavanja čvora u graf, kao i pretraga jednog novia grafa.



*Slika 16. (Levo) Algoritam za dodavanje novog čvora u graf. (Desno) Pertraga sloja u grafu.*

Na slici 17 prikazani su jednostavni algoritam za pretrage komšija, kao i algoritam pretrage koristeći heuristike.



*Slika 17. (Levo) Pretraga suseda. (Desno) Pretraga suseda pomoću heuristike.*

Jedna od osnovnih stvari prilikom pretrage je izbor odgovarajuće distance. S obzirom na to da je ArcFace model treniran korišćenjem ArcFace funkcije, automatski se nameće korišćenje cos-inusne distance ili angular distance.

# Implementacija u programskom jeziku Python

Za implementaciju sistema za preponzavanje lica je korišćen programski jezik Python. Za upravljanje paketima je korišćena Conda.

U nastavku se nalazi lista paketa koji su korišćeni za implementaciju:

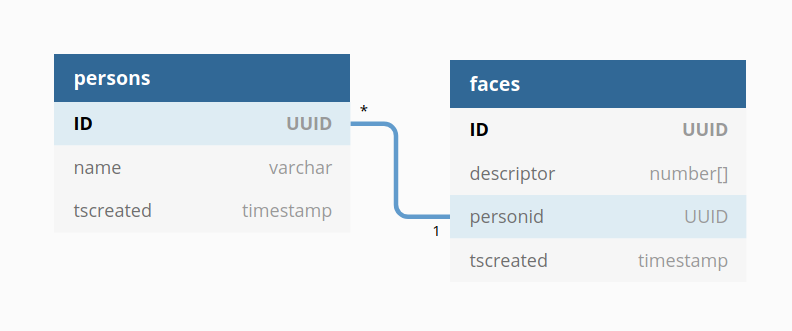
1. build-essential
2. python=3.7
3. tensorflow-gpu
4. opencv=4.4.0
5. psycopg2
6. pyfiglet
7. mtcnn
8. pillow
9. flask
10. flask-cors
11. Cython
12. n2
13. googledrivedownloader

Za kompletnu instalaciju sistema potrebno je instalirati tri odvojena dela, bazu, Python service i frontend servis. Instalaciju je moguće uraditi ručno, ili uz pomoć dostupnih skripti. Sistem je testiran samo na Linux operativnom sistemu, tačnije na Ubuntu 20.04 LTS.

Ukoliko su performance bitan factor, potrebno je instalirati TensorFlow sa podrškom za grafičku karticu, ukoliko ista postoji. Grafička kartica u ovom slučaju mora imati CUDA podršku, kao i instalirane adekvatne drajvere, CUDA, i CUDNN biblioteke. Moguće je konfigurisati i OpenCV biblioteku za rad sa grafičkom karticom, ali je zbog kompleksnosti i raznovrsnosti arhitektura preskočeno u instalacionoj skripti.

U nastavku neće biti prikazan ceo kod zbog svoje obimnosti, ali će biti dostupan. Fokisiraćemo se na delove koji su ključni i povezuju sve u jednu celinu. Treba napomenuti da se u sistemu nalaze i moduli koji nisu uključeni, na primer SSD detektor lica, MTCNN detektor lica, anti-spoofing modul. Promenom paramtera u JSON konfiguracionom fajlu moguće je menjati detektor, ili uključiti i isključiti eksperimentalni deo za anti-spoofing. Ovime je dobijeno na modularnosti sistema, novi moduli se mogu lako dodavati, kao i podesavati parametri vec postojećih. Ovo omogućava jednostavnije testiranje.

Za početak je bitno kreirati bazu podataka. U ovom slučaju je korišćena PostgreSQL baza, ali je vrlo jednostavno zameniti je. U bazi imamo samo dve tabele, tabelu persons gde čuvamo informaciju o imenu osobe, datum kreiranja i njen ID (UUIDv4) i drugu tabelu faces gde čuvamo encoding osobe (descriptor), ID (UUIDv4) zapisa, personid (FK na tabelu persons) i datum kreiranja zapisa.



*Slika 18. Model baze podataka*

U nastavku slede delovi koda neophodni za preoces prepoznavanja osobe.

Za čitanje podataka iz baze kreirana je funkcija read\_descriptors. Ona čita sve podatke iz tabele faces koji se kasnije koriste za kreiranja indeksa za pretragu.

|  |
| --- |
| def read\_descriptors(db):     """     read\_descriptors     The function for reading all face descriptors and ids from database     :param db: cursor     :return: {} or [], [], []     """      try:         db     except NameError:         print('Problem with the database connection')          return -1      query = 'SELECT "ID", descriptor, personid FROM public.faces'     db.execute(query)     records = db.fetchall()      ids = []     descriptors = []     persons\_ids = []      if records:         for r in records:             ids.append(r[0])             descriptors.append(r[1][0])             persons\_ids.append(r[2])          return ids, descriptors, persons\_ids     else:         return {'status': 'ERROR'} |

Nakon čitanja vektora iz baze, potrebno je dodati ih u HNSW graf. Ovo radimo metodom make\_base. Ova funkcija se nalazi u klasi RecognitionEngine. Ideja je dodati samo vektore i kreirati indeks. Ovo možemo sačuvati radi kasnijeg učitavanja ukoliko dođe do neke neželjene situacije. N2 paket podržava multithreading, tako da je ova opcija korišćena radi bržeg kreiranja indeksa.

|  |
| --- |
| def make\_base(self, descriptors: []) -> dict:     """     make\_base     The function used for adding data and building the index     :param descriptors: []     :return: img: numpy.array()     """      self.recognizer = HnswIndex(512, 'angular')      # add vectors to the ann     for d in descriptors:        self.recognizer.add\_data(np.array(d))      # build ann     self.recognizer.build(m=5, max\_m0=10, n\_threads=4)      self.recognizer.save('index.hnsw')     return {'status': 'SUCCESS'} |

Potrebno je inicijalizovati detektor lica koji se nalazi u klasi ImgProcessor. U ovom delu, učitavamo RetinaFace model sa diska i inicijalizujemo težine mreže vrednostima iz fajla.

|  |
| --- |
| if self.detector\_type == "RetinaFace":     # set config and checkpoints path     self.face\_det\_cfg\_path = load\_yaml(cfg['face\_det\_cfg\_path'])     self.face\_det\_checkpoints\_path = cfg['face\_det\_checkpoints\_path']     # load our serialized model from disk     # Here we need to read our pre-trained neural net     self.detector = RetinaFaceModel(self.face\_det\_cfg\_path,  training=False,  iou\_th=cfg["face\_det\_iou\_th"],  score\_th=cfg["face\_det\_score\_th"])  self.face\_det\_down\_scale\_factor = cfg["face\_det\_down\_scale\_factor"] # load checkpoint checkpoint\_dir = self.face\_det\_checkpoints\_path +  self.face\_det\_cfg\_path['sub\_name'] checkpoint = tf.train.Checkpoint(model=self.detector) if tf.train.latest\_checkpoint(checkpoint\_dir):     checkpoint.restore(tf.train.latest\_checkpoint(checkpoint\_dir)) else:      exit() |

Pored detektora, potrebno je inicijalizovati i modul za ekstrakciju vektora obeležja. Ovde se vrlo lako može zameniti backbone tip mreže sa MobileNet na ResNet.

|  |
| --- |
| # set config and checkpoints path self.face\_reco\_cfg\_path = load\_yaml(cfg['face\_reco\_cfg\_path']) self.face\_reco\_checkpoints\_path = cfg['face\_reco\_checkpoints\_path']  # initialize the ArcFace model self.model = ArcFaceModel(size=self.face\_reco\_cfg\_path['input\_size'],                                   backbone\_type=self.face\_reco\_cfg\_path['backbone\_type'], training=False)  # load model weights ckpt\_path = tf.train.latest\_checkpoint(self.face\_reco\_checkpoints\_path +  self.face\_reco\_cfg\_path['sub\_name']) if ckpt\_path is not None:     self.model.load\_weights(ckpt\_path) else:     exit() |

Za proces detekcije je zbog preglednosti prikazan je samo deo koji koristi RetinaFace detektor. U ovom delu, prvi korak je uraditi promenu veličine slike ukoliko je to potrebno, a nakon toga dodati padding kako bi se izbegao problem sa neodgovarajucim dimenzijama slike i ulaza u mrežu. Nakon inference, potrebno je ukloniti padding efekat kako bi imali početnu sliku.

U petlji prolazimo kroz svaku detekciju i uzimamo koordinate bounding box-a. Pored ovog, koristeći informacije o poziciji očiju vršimo poravnanje po već pomenutom algoritmu. Svaku sliku na kojoj postoji lice dodajemo u niz koji metoda prosledjuje narednoj metodi. Ukoliko je potrebno, sliku možemo sačuvati na disk promenom konfiguracionog fajla.

|  |
| --- |
| def detect(self, img: np.array([])) -> np.array([]):     """     detect     The actual function for performing detection based on configuration  parameter.     Crops all faces from the image and returns the numpy array  containing all of them.  :param img: numpy.array()     :return: numpy.array()     """      faces\_array = []      if self.detector\_type == "RetinaFace":        img\_height, img\_width, \_ = img.shape         if self.face\_det\_down\_scale\_factor < 1.0:            img = self.resize\_img(img, 0, 0,  self.face\_det\_down\_scale\_factor,  self.face\_det\_down\_scale\_factor,  cv2.INTER\_LINEAR)         # pad input image to avoid unmatched shape problem        img, pad\_params = pad\_input\_image(img,  max\_steps=max(self.face\_det\_cfg\_path["steps"]))  faces = self.detector(img[np.newaxis, ...]).numpy()        # recover padding effect        faces = recover\_pad\_output(faces, pad\_params)         for face in range(len(faces)):            # get coordinates            x1, y1, x2, y2 = int(faces[face][0] \* img\_width),  int(faces[face][1] \* img\_height),                              int(faces[face][2] \* img\_width),  int(faces[face][3] \* img\_height)        x1 = 0 if x1 < 0 else x1        y1 = 0 if y1 < 0 else y1         f = img[y1:y2, x1:x2]         if self.experimental:            if not self.is\_alive(f)['isAlive']:                continue         # landmark        if faces[face][14] > 0:            right\_eye\_x, right\_eye\_y = int(faces[face][4] \* img\_width),  int(faces[face][5] \* img\_height)           left\_eye\_x, left\_eye\_y = int(faces[face][6] \* img\_width),  int(faces[face][7] \* img\_height)             delta\_x = right\_eye\_x - left\_eye\_x            delta\_y = right\_eye\_y - left\_eye\_y            angle = np.arctan(delta\_y / delta\_x)            angle = (angle \* 180) / np.pi             h, w = f.shape[:2]            # Calculating a center point of the image            # Integer division "//"" ensures that we receive whole  # numbers            center = (w // 2, h // 2)            # Defining a matrix M and calling            m = cv2.getRotationMatrix2D(center, (angle), 1.0)            # Applying the rotation to our image using the            aligned\_face = cv2.warpAffine(f, m, (w, h))            faces\_array.append(aligned\_face)      if self.write == "true":         for face in faces\_array:             cv2.imwrite(str(uuid.uuid4())+'.jpg', face) |

Nakon što imamo poravnanu sliku lica, sledeći korak je uraditi inferencu koristeći ArcFace mrežu i dobiti vektore obeležja. Za ovo je kreirana metoda klase ImgProcessor pod nazivom encode. Pre inference je potrebno normalizovati sliku i pretvoriti je u float32 format. Nakon inference, L2 normalizacijom postižemo da koeficijenti vektora budu manji, i manje kompleksnosti.

|  |
| --- |
| def encode(self, img: np.array([])) -> np.array([]):     """     encode     Function for extracting face embeddings using ArcFace model     :param img: numpy.array()     :return: img: numpy.array()     """     img = self.resize\_img(np.array(img),  self.face\_reco\_cfg\_path['input\_size'],  self.face\_reco\_cfg\_path['input\_size'],  0, 0, cv2.INTER\_LINEAR)      img = img.astype(np.float32) / 255.     if len(img.shape) == 3:        img = np.expand\_dims(img, 0)     # extract face embeddings and normalize     embeds = l2\_norm(self.model(img))     return embeds |

Sledeći korak u procesu prepoznavanje je pretraga vektora i traženje onog sa najmanjom distancom. Za ovo koristimo već pomenutu klasu RecognitionEngine i metodu identification. Metoda kao ulazne parametre dobija vektor obeležja i spisak ID-jeva osoba. Nema potrebe za celim nizom vektora jer su oni već učitani korišćenjem make\_base metode.

|  |
| --- |
| def identification(self, descriptor: [], person\_ids: []) -> dict:     """     identification     Efficient and robust approximate nearest neighbor search using         Hierarchical Navigable Small World graph      :param descriptor: []     :param person\_ids: []      :return: dict     """     idx = self.recognizer.search\_by\_vector(  np.array(descriptor).flatten(),  2, 1, include\_distances=True  )      if idx[0][1] < self.threshold:        person\_id = person\_ids[idx[0][0]]     else:         person\_id = "Not recognized"     return {             'status': 'SUCCESS',             'personid': person\_id     } |

Ceo proces prepoznavanja se može povezati u jednu funkciju koja služi kao Rest endpoint.

|  |
| --- |
| # Create a URL route in our application for# "/@app.route('/identification')" # The purpose of this endpoint is to predict the person identity @app.route('/identification', methods=["POST"]) @cross\_origin() def predict\_rest() -> dict:     """     predict\_rest     The actual function for performing the recognition.     Extract all faces from the image, check if faces are real,     extract face descriptors and search the database using HNSW     and angular distance metric.     :return: dict     """     pil\_image = Image.open(request.files['image']).convert('RGB')     img = np.array(pil\_image)      img = cv2.resize(img, (640, 640))      # detect faces     faces = imgProcessor.detect(img)     response = []      global ids, descriptors, persons\_ids      start = time.time()     if persons\_ids:         for face in faces:             # encode face             descriptor = imgProcessor.encode(face)             # find personid in the database             person\_id = recEngine.identification(descriptor,  persons\_ids)             if person\_id['personid'] is not None:                 # find name in the database                 person = dbase.find\_person\_by\_id(db,  person\_id['personid'])                 response.append(person)             else:                 response.append(person\_id)     else:         response = 'Empty database'     logging.info('Identification time: ' + str(time.time()-start))     return {         'status': 'SUCCESS',         'response': response     } |

Ukoliko je osoba prepoznata, odnosno njen ID je vraćen kao rezultat funkcije, u bazi možemo pronaći njeno ime jednostavnim upitom. Ovo je prikazano u sledećem delu koda.

|  |
| --- |
| def find\_person\_by\_id(db, person\_id) -> dict:     """     find\_person\_by\_id     The function for search the database for name     :param db: cursor     :param person\_id: str     :return: dict     """     query = 'SELECT p.name FROM public.persons p WHERE "ID" = %s;'     db.execute(query, (person\_id,))     records = db.fetchall()     if records[0][0] is not None:         return {             'status': 'SUCCESS',             'name': str(records[0][0])         }     else:         return{'status': 'ERROR'} |

Proces prepoznavanje je beskoristan ukoliko je baza prazna. U ovom slučaju se ni indeks ne može kreirati, samim tim ni pretraga. Kako bi se ovo izbeglo, potrebna je funkcija koja vektore upisuje u bazu.

Već pomenute funkcije detect i encode se mogu iskoristiti. Nakon detektovanja lica, poravnanja, i ekstrakcije vektora obeležja, možemo preskočiti proces prepoznavanja i vektore upisati u bazu. Treba napomenuti da nakon upisivanja vektora u bazu, potrebno je ponovo kreirati indeks, odnosno pozvati make\_base metodu. Ovo se može izbeći inkrementalnim dodavanjem ukoliko biblioteka koja se koristi to podržava.

|  |
| --- |
| def receive\_descriptors(db, db\_conn, name, embeds) -> dict:     """     receive\_descriptors     The function for writing records in database     :param db: cursor     :param db\_conn: connection     :param name: str     :param embeds: np.array([])     :return: dict     """     query = 'INSERT INTO public.persons (name) VALUES (\'' + str(name) +  '\') RETURNING "ID";'      db.execute(query)     records = db.fetchall()     if records[0][0] != '':         person\_id = records[0][0]      for emb in embeds:         emb = np.array(emb).tolist()         query = 'INSERT INTO public.faces (descriptor, personid)  VALUES (%s, %s);'         db.execute(query, (emb, (person\_id,)))     db\_conn.commit()     return {'status': 'SUCCESS'} |

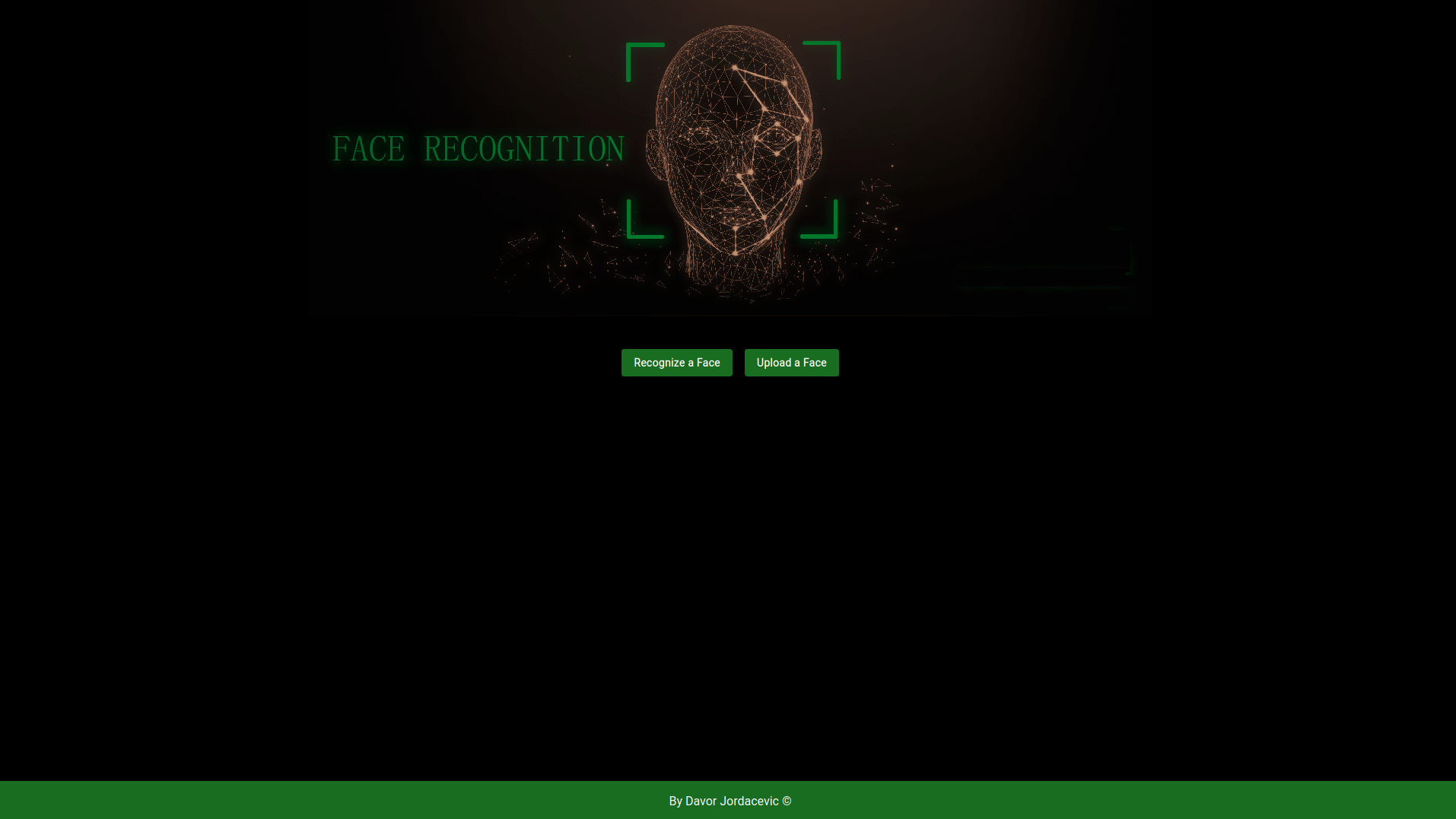
Kao i u slučaju prepoznavanja, i ovde se ceo proces enkodovanja i upisivanja u bazu može spojiti u jednu funkciji i izložiti eksterni endpoint.

|  |
| --- |
| # Create a URL route in our application for  # "/@app.route('/encodeAndInsert)" # The purpose of this endpoint is to encode the face @app.route('/encodeAndInsert', methods=["POST"]) def encode\_and\_insert() -> dict:     """     encodeAndInsert     The actual function for adding a new person into database.     After person is added successfully, make base is performed.     :return: dict     """     name = request.form.get('name')     uploaded\_files = request.files.getlist("images")      embeds = []     if not uploaded\_files:         return {"status": "ERROR"}     for image in uploaded\_files:         pil\_image = Image.open(image).convert('RGB')          img = np.array(pil\_image)          img = cv2.resize(img, (640, 640))         # detect face         faces = imgProcessor.detect(img)         if len(faces) != 1:             return {'status': 'ERROR'}         embeds.append(imgProcessor.encode(np.array(faces[0])))      result = dbase.receive\_descriptors(db, db\_conn, name, embeds)     if result['status'] != 'SUCCESS':         return {'status': 'ERROR'}      global ids, descriptors, persons\_ids     ids, descriptors, persons\_ids = dbase.read\_descriptors(db)     return recEngine.make\_base(descriptors) |

Pri pokretanju ovog servisa pokreće se main funkcija. Ono što je ovde bitno napomenuti je da se u njoj inicijalizuju instance svih klasa, i da se u njoj mora dozvoliti TenorFlow biblioteci da koristi grafičku karticu i da ne alocira memoriju unapred. TensorFlow podržava samo kartice sa CUDA podrškom, tako da je sistem namenjem karticama sa NVIDIA grafičkim karticama. U nastavku sledi kod main funkcije.

|  |
| --- |
| # Create the application instance app = Flask('FR APP') app.config['CORS\_HEADERS'] = 'Content-Type' app.config['UPLOAD\_EXTENSIONS'] = ['.jpg', '.png'] cors = CORS(app)  # initialize empty lists ids, descriptors, persons\_ids = [], [], []  if \_\_name\_\_ == '\_\_main\_\_':      # remove old log file     os.remove("FRAPP.log")      # set environment variables     os.environ['FLASK\_ENV'] = 'development'     os.environ['TF\_CPP\_MIN\_LOG\_LEVEL'] = '3'      # turn off tensorflow logger     logger = tf.get\_logger()     logger.disabled = True     logger.setLevel(logging.FATAL)      # open file for logging     logging.basicConfig(filename='FRAPP.log', level=logging.DEBUG,  format='%(asctime)s %(levelname)-8s  %(message)s')     # allow GPU memory grow    gpus = tf.config.experimental.list\_physical\_devices('GPU')    if gpus:        try:            os.environ['CUDA\_VISIBLE\_DEVICES'] = '0'           for gpu in gpus:                tf.config.experimental.set\_memory\_growth(gpu, True)       except RuntimeError as e:            logging.info(str(e))           print(e)      logging.info('Allow GPU memory grow successful.')      # parse arguements      parser = argparse.ArgumentParser(description='Process some  arguments.')      parser.add\_argument('--cdp', type=str, help='the path to config  file')      logging.info('Parsing arguments successful.')      args = parser.parse\_args()     config\_path = args.cdp      cfg = config.readConfig(config\_path)      db\_conn, db = dbase.db\_connect(cfg["host"], cfg["port"],  cfg["name"], cfg["user"], cfg["password"])     # initialize image processor    imgProcessor = ImgProcessor(cfg)    # initialize recognition engine    recEngine = RecognitionEngine(cfg['threshold'])   logging.info('All models initialized successfully.')      try:         # read ids, descriptors and person\_ids from database         ids, descriptors, persons\_ids = dbase.read\_descriptors(db)         logging.info('Read descriptors successful.')         recEngine.make\_base(np.array(descriptors))         logging.info('Make base successful.')     except:         logging.info('The database is empty.')      # Run the flask rest api     # This can be updated to use multiple threads or processors     # In addition, some type of queue should be used     # print starting text     ascii\_banner = pyfiglet.figlet\_format("F R     A P P", font="slant")     print(ascii\_banner)      logging.info('FR APP IS RUNNING.')     logging.info('---------------' \* 4)     # threaded=False, processes=3     app.run(debug=True, host='127.0.0.1', port=5000) |

Za potrebe demonstracije sistema, kreiran je frontend deo koristeći NodeJS programski jezik. Namena ovog servisa je da uplouduje sliku i pozove odgovarajući endpoint za prepoznavanje, ili da uplouduje sliku i pošalje ime osobe radi ubacivanja osobe u bazu podataka.



*Slika 19. Izled frontend-a*

# Zaključna razmatranja

U svom istraživanju u ovom radu došao/la sam do sledećih važnih nalaza.

RetinaFace, kao i ArcFace mreže su u ovom radu bazirane na MobileNet arhitekturi. Ova arhitektura pruža bolje performanse od prethodno pomenutog ResNet-a, ali to nije reč sa preciznošću. Prvi korak u postižanju tačnijeg sistema bi bio implementacija ResNet arhitekture. N2 paket koji je korišćen u ovom radu podržava jako mali broj dostupnih distanci. Premda angular distanca pruža bolje performanse, kod velikih vektora slika koji su jako blizu može doci do grešaka (slučaj kod većih setova podataka), pa je bolja opcija koristiti pakete koji podržavaju cos-inusnu distance, ili je implementirati i dodati u N2 paket.

Iako je MobileNet arhitektura brza, implementacija korišćena u ovom radu ne postiže SOTA rezultate, te se performanse mogu dodatno poboljšati korišćenjem originalne implementacije koja koristi dodatne trikove za optimizaciju.

RetinaFace mreža sa MobileNetV2 arhitekturom korišćena u ovom radu je kao trening set podataka koristila WiderFace dataset, dok je ArcFace mreža sa MobileNet arhitekturom trenirana na MS-Celeb-1M setu podataka. Premda ovi setovi podataka sadrže veliku količinu podataka, za najbolje rezultate je potrebno dodati setove koji sadrže uniformno distribuirane slike za sve rase. Ovo je jako bitno za jedan produkcioni sistem zbog takozvanog bias-a koji se javlja ukoliko setovi podataka nemaju uniformnu distribuciju.

Za jedan produkcioni sistem je bitno da koristi i anti-spoofing metode kako bi se zaštitili od napada. Sistem poput ovog je modularan, pa se ubacivanjem dodatnog modula između dela za detekciju i dela za prepoznavanje ovo može postići. Treba imati na umu da su implementacije anti-spoofing sistema kompleknse, i većina SOTA modela koristi RGB, depth i IR slike.

Delove koda koji se često ponavljaju, kao i delovi koji isključivo rade sa Numpy nizovima, mogu se prepraviti da koriste CuPy paket koji omogućava koriščenje grafičke kartice pri radu sa nizovima, ili iskoristiti brzinu C programskoj jezika korišćenjem Cython paketa i refaktorisanjem koda. Drugo rešenje može biti korišćenjem Numba kompajlera koji osim što prevodi kod u mašinski kod, omogućava u paralelizaciju koda.

Delovi koda se mogu prepraviti po konkurentnom principu kako bi se izbeglo zaključavanje procesa i čekanje odgovora. Takodje, premda lak za implementaciju, Flask razvojni okvir nije namenjen za komercijalne sisteme, te bi njegova zamena bila neophodna.

Trenutni sistem nije skalabilan, a kako bi se ovo omogućilo, potrebno je implementirati mehanizme koji će čuvati sve zahteve i raspoređivati nekom procesu po potrebi.

Sistemi dubokog učenja su zahtevni za deployment i odrzavanje. Rešenje koje je idealno za ovakve sisteme je korišćenje Docker-a, odnosno principa kontejnera. U ovom slučaju ne samo da bi se samo Python deo sistema morao prebaciti u kontejner, već i cela baza.

Do sada je bilo reči o radu sa slikama. Ukoliko ima potrebe raditi sa video snimcima, situacije postaje komplikovanija. Premda video snimak jeste samo niz slika, i moguće je uraditi prepoznavanje lica na svakoj slici (frejmu), ovo je prilično zahtevan posao za računar. U ovim slučajevima je potrebno koristiti tehnike praćenja (eng. tracking) i reidentifikacije.

# Literatura

1. Deng, Jiankang, et al. "Retinaface: Single-stage dense face localisation in the wild." *arXiv preprint arXiv:1905.00641* (2019).
2. Sandler, Mark, et al. "Mobilenetv2: Inverted residuals and linear bottlenecks." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018.
3. Tang, Xu, et al. "Pyramidbox: A context-assisted single shot face detector." *Proceedings of the European Conference on Computer Vision (ECCV)*. 2018.
4. Lin, Tsung-Yi, et al. "Focal loss for dense object detection." *Proceedings of the IEEE international conference on computer vision*. 2017.
5. Deng, Jiankang, et al. "Retinaface: Single-stage dense face localisation in the wild." *arXiv preprint arXiv:1905.00641* (2019).
6. Dai, Jifeng, et al. "Deformable convolutional networks." *Proceedings of the IEEE international conference on computer vision*. 2017.
7. Schroff, Florian, Dmitry Kalenichenko, and James Philbin. "Facenet: A unified embedding for face recognition and clustering." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015.
8. Deng, Jiankang, et al. "Arcface: Additive angular margin loss for deep face recognition." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2019.
9. Malkov, Yury A., and Dmitry A. Yashunin. "Efficient and robust approximate nearest neighbor search using hierarchical navigable small world graphs." *IEEE transactions on pattern analysis and machine intelligence* (2018).