

Raspoznavanje osoba na temelju slike lica

Valentino Janda
*Fakultet elektrotehnike
i računarstva*

Sveučilište u Zagrebu
Email: valentino.janda@fer.hr

Josip Maradin
*Fakultet elektrotehnike
i računarstva*

Sveučilište u Zagrebu
Email: josip.maradin@fer.hr

Filip Smolić
*Fakultet elektrotehnike
i računarstva*

Sveučilište u Zagrebu
Email: filip.smolic@fer.hr

Mia Krstičević
*Fakultet elektrotehnike
i računarstva*

Sveučilište u Zagrebu
Email: mia.krsticevic@fer.hr

Marin Boić

*Fakultet elektrotehnike
i računarstva*

Sveučilište u Zagrebu
Email: marin.boic@fer.hr

Nikolina Buzić

*Fakultet elektrotehnike
i računarstva*

Sveučilište u Zagrebu
Email: nikolina.buzic@fer.hr

Sažetak—Ovaj projekt se bavi prepoznavanjem osoba na temelju slike lica koristeći digitalnu obradu i analizu slike. Cilj je usporediti modele koji mogu točno identificirati pojedince na temelju njihovih fotografija. Istrenirali smo već postojeći YOLO model te pretrenirani MobileNetv2 model kako bismo usporedili njihovu učinkovitost. Tijekom treniranja modela koristili smo dataset koji se sastoji od 105 klasa s poznatim osobama. Slike su morale biti predprocesirane, uključujući promjenu dimenzija kako bi bile ujednačene. Naši rezultati pokazuju da je, uz dovoljno veliki dataset, moguće istrenirati model koji pouzdano prepoznaje ljudska lica. Modeli prepoznavanja lica imaju široku primjenu u različitim područjima, uključujući sigurnost, pristupne sustave i identifikaciju.

1. Uvod

Razvoj tehnologije prepoznavanja lica je postao ključan u suvremenom svijetu, pronalazeći primjenu u različitim područjima poput sigurnosti, pristupnih sustava i identifikacije. S porastom digitalne obrade slike i analize podataka, postavlja se pitanje kako različiti modeli mogu točno identificirati pojedince na temelju njihovih fotografija.

Cilj ovog istraživanja je usporediti različite modele prepoznavanja lica te evaluirati njihovu učinkovitost u prepoznavanju pojedinaca. Istrenirani su već postojeći YOLO model i pretrenirani MobileNetV2 model kako bi se usporedila njihova sposobnost prepoznavanja lica. Korišten je dataset koji sadrži podatke o 105 klasa s poznatim osobama, što je omogućilo opsežno treniranje modela.

Prepoznavanje lica nije samo tehnološki izazov, već i potencijalno sredstvo za poboljšanje sigurnosti i olakšavanje svakodnevnih procesa identifikacije. Kroz ovo istraživanje, želimo istražiti mogućnosti i granice trenutnih modela te istaknuti važnost razvoja pouzdanih sustava za prepoznavanje lica u digitalnom dobu.

2. Pregled postojećih pristupa i metoda raspoznavanja lica

Pregled postojećih pristupa u raspoznavanju lica uključuje i pregled povijesti ovog područja. Raspoznavanje lica ima dugu povijest koja seže od ranog istraživanja u akademskom kontekstu do širokog spektra primjena u stvarnom svijetu. Prvi pokušaji raspoznavanja lica datiraju još iz 1960-ih godina, kada su se istraživači počeli baviti ovim područjem kao područjem računalnog vida i obrade slika.

Od tada, raspoznavanje lica postalo je ključna tehnologija u raznim područjima, uključujući sigurnost, biometrijske sustave, analizu socijalnih medija i druge aplikacije. Danas se koristi u sustavima sigurnosti za identifikaciju i provjeru, u pristupnim sustavima za autentikaciju korisnika, u marketinškim alatima za prepoznavanje potrošača i u raznim drugim područjima.

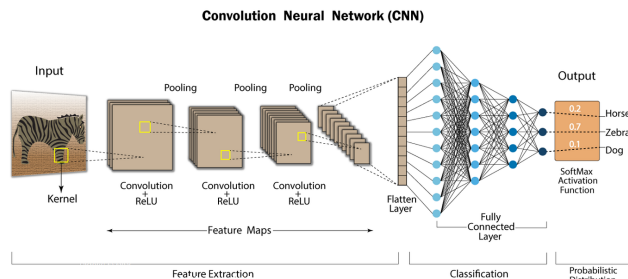
Pristupi raspoznavanju lica uključuju tradicionalne metode temeljene na karakteristikama lica poput geometrijskih značajki i statističkih modela te moderne pristupe koji se oslanjaju na duboko učenje i neuronske mreže. Kombinacija tih pristupa omogućava razvoj sve naprednijih tehnika prepoznavanja lica koje su ključne za različite aplikacije u suvremenom svijetu.

2.1. Moderni pristupi

U okviru modernih metoda prepoznavanja lica ističu se one koje se oslanjaju na duboko učenje i neuronske mreže. Ove tehnike omogućuju izgradnju složenih modela koji mogu samostalno učiti značajke iz velikih skupova podataka, što rezultira visokom točnošću u prepoznavanju lica.

Duboko učenje omogućuje modelima da automatski izvlače značajke iz slika lica na temelju slojeva neuronskih mreža, čime se eliminira potreba za ručnim definiranjem

značajki. Konvolucijske neuronske mreže (CNN) često se koriste u ovom kontekstu zbog svoje sposobnosti učenja prostornih značajki iz slika. Osim toga, tehnike poput transfernog učenja omogućuju prenošenje znanja iz jednog skupa podataka na drugi, što olakšava treniranje modela čak i s manjim skupovima podataka.



Slika 1. Arhitekturu dubokih neuronskih mreža (CNN)

Osim dubokog učenja, moderni pristupi uključuju i upotrebu generativnih suparničkih mreža (GAN), koje se koriste za generiranje realističnih slika lica. Ove tehnike mogu biti korisne u situacijama gdje su dostupni podaci ograničeni ili nepotpuni, jer omogućuju generiranje dodatnih primjera koji mogu poboljšati performanse modela.

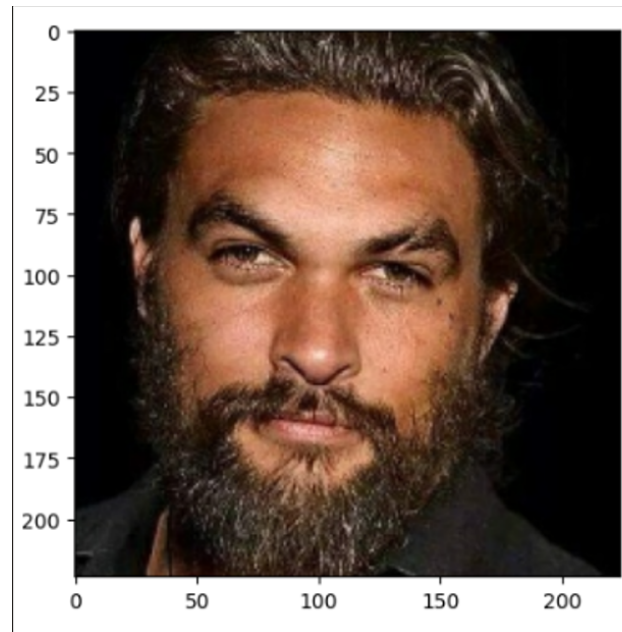
Uz korištenje modernih metoda poput dubokog učenja i generativnih suparničkih mreža, istraživači i inženjeri neprestano rade na razvoju novih tehnika i poboljšanju postojećih kako bi se postigla veća točnost i pouzdanost u prepoznavanju lica. Ovi napori doprinose razvoju sve naprednijih sustava prepoznavanja lica koji imaju široku primjenu u stvarnom svijetu.

3. Rješenje problema: Implementacija projektnog tima

S obzirom na složenost zadatka i raznolikost dostupnih pristupa, projektni tim je odabrao implementirati dva glavna modela za prepoznavanje lica: YOLO (You Only Look Once) model i MobileNetv2 model. YOLO model je poznat po svojoj brzini i sposobnosti istovremenog prepoznavanja i lokalizacije objekata na slici, dok je MobileNetv2 model poznat po svojoj efikasnosti na mobilnim uređajima i računalima s ograničenim resursima. Svaki od ovih modela ima svoje jedinstvene karakteristike, prednosti i ograničenja, što ih čini prikladnim za različite primjene prepoznavanja lica.

Koristili smo dataset koji se sastoji od 105 klasa s poznatim osobama za treniranje naših modela. Uz to, prilagodili smo parametre modela prema našim podacima kako bismo osigurali optimalno prilagođavanje modela našem specifičnom zadatku.

Tijekom procesa treniranja bilo je potrebno provesti određeno pretprocesiranje. Slike u našem skupu podataka nisu bile istih dimenzija, pa je bilo nužno izvršiti njihovo prilagođavanje na iste dimenzije. Nadalje, vrijedi istaknuti



Slika 2. Originalna slika

da je YOLO model samostalno provodio augmentaciju podataka, primjenjujući različite efekte na slike kako bi obogatio skup podataka.

Prilikom treniranja MobileNetv2 modela koristili smo različite transformacije slika, uključujući rotaciju, zoomiranje, translaciju i promjenu kontrasta kako bismo poboljšali sposobnost modela da prepoznaje lica u različitim uvjetima. Originalna slika te slika nakon provedenih transformacija mogu se vidjeti na Slikama 2 i 3.

U ovom poglavlju detaljno ćemo analizirati proces implementacije ovih modela, uključujući metodologiju, tehničke detalje, izazove s kojima se tim suočio tijekom implementacije te konačne rezultate i performanse postignute svakim od modela.

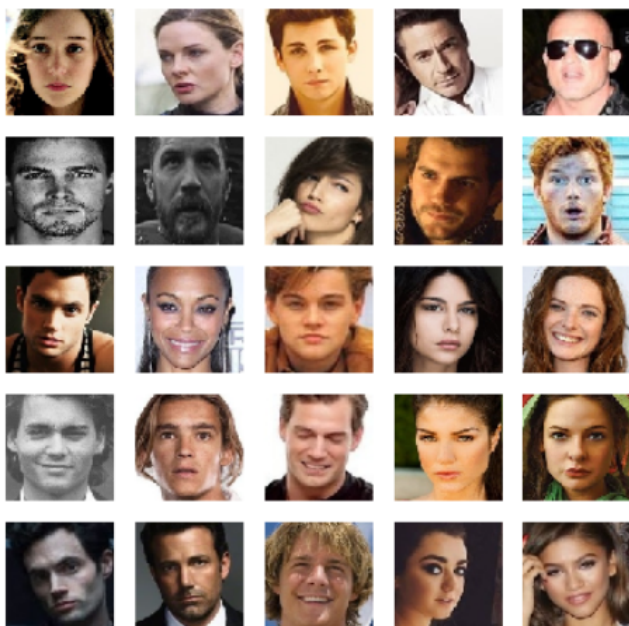
3.1. Priprema i obrada podataka

Pri implementaciji modela za prepoznavanje lica, prvi korak je detaljno proučavanje skupa podataka kako bi se osiguralo da su podaci pravilno strukturirani i pripremljeni za daljnju analizu. U ovom projektu koristimo opsežan skup podataka koji sadrži ukupno 17.534 slike, dok broj različitih osoba, odnosno izlaznih klasa, iznosi 105. Ovaj raznoliki i opsežni dataset ključan je za treniranje modela jer pruža dovoljno varijacija u slikama lica koje model mora naučiti prepoznati i klasificirati. Velika raznolikost u slikama uključuje različite kutove lica, osvjetljenja, izraze lica, te druge promjene koje model treba naučiti prepoznati i nositi se s njima.

Kako bismo olakšali pripremu podataka za model, prvo kreiramo DataFrame u kojem svaka osoba ima povezane sve njoj pripadajuće slike. DataFrame omogućuje jednostavno rukovanje i manipulaciju podacima, čime se ubrzava proces



Slika 3. Transformirane slike



Slika 4. Primjer fotografija iz korištenog skupa podataka

pripreme za modeliranje. Iz tog DataFrame-a stvaramo dvije ključne liste: images i labels. Lista images sadrži sve slike lica, dok lista labels sadrži oznake (klase) koje predstavljaju identitet osoba na slikama. Ove liste su potrebne kako bi podaci mogli biti uspješno ubačeni u model.

Kako bi model mogao pravilno raditi s izlaznim klasama, potrebno je provesti oznake (labels) kroz proces One-Hot Encoding. One-Hot Encoding pretvara kategorijske varijable (u ovom slučaju identitete osoba) u binarne vektore, gdje svaka klasa (osoba) ima jedinstveni vektor. Ovaj postupak je ključan za normalizaciju vrijednosti izlaznih klasa i omogućuje modelu da učinkovito nauči razlike među klasama. Na primjer, umjesto da se osobe označavaju brojevima (1, 2, 3, itd.), One-Hot Encoding stvara vektor koji je 1 na poziciji koja odgovara toj osobi, a 0 na svim ostalim pozicijama.

Osim toga, kako bi se osiguralo da sve slike imaju iste dimenzije, koristimo metodu resize. S obzirom na to da neuronske mreže zahtijevaju ulazne podatke uniformnih dimenzija, svaka slika se transformira na fiksnu veličinu od (224, 224) piksela. Ovaj korak je važan jer omogućuje modelu da dosljedno obrađuje slike i smanjuje varijacije koje bi mogle utjecati na njegovu učinkovitost.

Kvalitetan model prepoznavanja lica također zahtijeva adekvatnu podjelu podataka na skupove za treniranje i testiranje. Skup za treniranje koristi se za optimizaciju parametara modela, dok se skup za testiranje koristi za evaluaciju njegovih performansi na neviđenim podacima. Ova podjela je ključna za procjenu generalizacijske sposobnosti modela, odnosno njegove sposobnosti da točno klasificira nova, nepoznata lica.

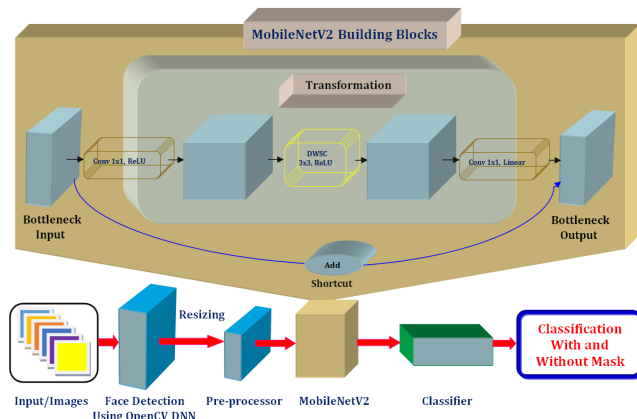
Uz to, kako bi model bio robusniji i otporniji na različite varijacije u slikama, primjenjujemo različite tehnike augmentacije podataka. Augmentacija podataka uključuje transformacije poput rotacije, zoomiranja, translacije i prilagodbe kontrasta prikazanih na slici 3. Možemo vidjeti da se na istu sliku može primijeniti i više od jedne transformacije, te da u slučaju translacije se dodaje dio slike, te nastaju dijelovi na kojima su dvije brade ili dva čela. YOLO model, kojeg koristimo u našem projektu, ima ugrađene funkcionalnosti za automatsku augmentaciju podataka, što dodatno poboljšava performanse modela. Augmentacija omogućuje modelu da se bolje nosi s različitim uvjetima osvjetljenja, položajima lica i drugim varijacijama, što je ključno za stvaranje pouzdanog sustava za prepoznavanje lica.

Sve ove faze pripreme podataka ključne su za uspješno treniranje modela i postizanje visokih performansi u prepoznavanju lica. Pravilna priprema podataka osigurava da model može učinkovito učiti iz skupa podataka i generalizirati svoja znanja na nove primjere.

3.2. MobileNetv2 model

MobileNetv2 je duboka konvolucijska neuronska mreža napravljena od Googlea 2018. godine. Originalna mreža se sastoji od 53 sloja, dok implementacija u TensorFlowu ili Kerasu uz sve pomoćne slojeve (npr. normalizacija, aktivacije i dr.) ima 154 sloja. Mreža je prednaučena na ImageNet

skupu podataka koji sadrži 1 400 000 slika podijeljenih u 1000 razreda. Ovaj model se često koristi u aplikacijama za prepoznavanje objekata ili lica te segmentaciji slika, a karakterizira ga visoka preciznost i efikasnost u radu s različitim vrstama podataka. Arhitektura MobileNetv2 modela može se vidjeti na slici 5.



Slika 5. Arhitektura MobileNetv2

Kako bi MobileNetv2 model radio za naš konkretan skup podataka, potrebno ga je ugoditi. Korišteni skup podataka Pins Face Recognition sadrži slike 105 poznatih osoba prikupljenih s Pinteresta i obrezanih na jednake dimenzije 224x224. Na arhitekturu modela još dodajemo nekoliko dodatnih slojeva kako bismo ga prilagodili našim podacima:

- GlobalAveragePooling2d - služi za pretvaranje 5x5 prostornih mapa značajki u jednodimenzionalne vektore
- Dropout - služi za sprječavanje prenaučivosti tako da nasumično postavlja težine određenih neurona na 0, rate=0.5
- potpuno povezani sloj - služi za pretvaranje dobivenih vrijednosti u jedinstvenu predikciju klase, korištenjem softmaxa

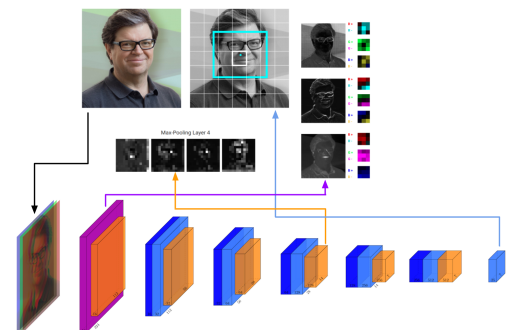
Naš skup podataka smo podijelili na skupove za treniranje, validaciju i testiranje u omjeru 70:20:10. Na svaku sliku, kako bismo dobili veću robusnost modela, nasumično primjenjujemo transformacije:

- RandomFlip - okreće sliku po x-osi
- RandomRotation - zakreće sliku za faktor=0.2, što znači da je kut zakretanja u intervalu $[-0.2 * 2\pi, 0.2 * 2\pi]$
- RandomZoom - uvećava ili smanjuje sliku za faktor=0.1 na x i y-osi
- RandomContrast - dodaje kontrast slici za faktor=0.2

Primjer originalne i transformiranih slika je prikazan na slikama 2 i 3. Na ulaz modela se uobičajeno predaju slike dimenzija 224x224, iako se mogu predati i slike proizvoljnih dimenzija, te vrijednosti piksela u intervalu $[-1, 1]$.

3.3. YOLO model

YOLO (You Only Look Once) je revolucionarni model za detekciju objekata u realnom vremenu koji koristi duboko učenje. Osnovna ideja YOLO-a je da se cijela slika analizira samo jednom kako bi se istovremeno otkrili objekti i njihovi okviri, što omogućuje bržu i učinkovitiju detekciju u usporedbi s tradicionalnim metodama koje zahtijevaju više prolaza kroz sliku. YOLO modeli su poznati po svojoj brzini i sposobnosti preciznog detektiranja objekata u realnom vremenu. Također, kao što je već spomenuto, YOLO modeli samostalno vrše transformacije slika kao što su rotacija, promjena veličine, i drugi postupci augmentacije podataka, što ih čini izuzetno prilagodljivima i učinkovitim za različite scenarije detekcije objekata. Arhitektura YOLO modela može se vidjeti na Slici 6.



Slika 6. Arhitektura YOLO

Kao i kod MobileNetv2 modela, izvršila se podjela na skup za treniranje i skup za validaciju. YOLO nakon svake epohe izvrši testiranje mreže slikama za validaciju nakon čega ispisuje rezultate određenih metrika.

4. Opis eksperimentalnih rezultata

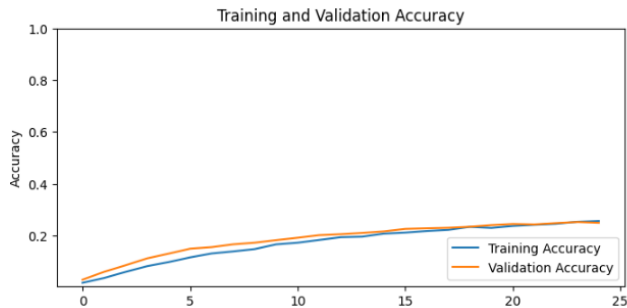
4.1. MobileNetv2

Na početku postavljamo parametar MobileNetv2 modela trainable=False, što sprječava ažuriranje parametara kod učenja, to radimo kako bismo samo trenirali naše dodatne slojeve. Nakon 25 epoha, s hiperparametrom stope učenja lr=0.0001 dobivamo rezultate prikazane na slikama 7 i 8.

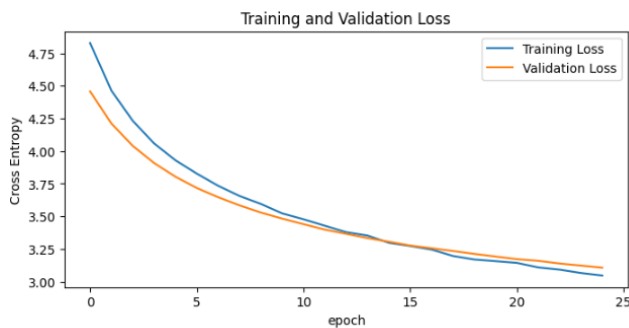
Nakon treniranja dodanih slojeva model postiže rezultate:

- točnost treniranja: 0.1563
- f1-mjera treniranja: 0.2493
- gubitak treniranja: 3.5407
- točnost validacije: 0.2185
- f1-mjera validacije: 0.1981
- gubitak validacije: 3.2951

Gubitak se jako sporo smanjuje te je model teško poboljšati daljnjim učenjem, sada krećemo na ugađanje pred-treniranog modela.



Slika 7. Točnost nakon treniranja novih slojeva



Slika 8. Gubitak nakon treniranja novih slojeva

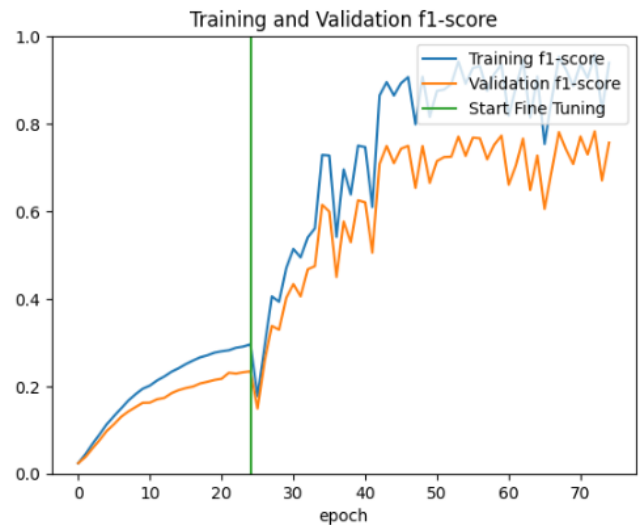
Ugađanje modela radimo na našem skupu podataka i postavljamo parametar trainable na True, no samo za posljednja 53 sloja. Razlog tome je što u dubokim konvolucijskim mrežama, što je sloj bliži izlazu je specijaliziraniji za određene detalje na slikama, stoga želimo ugoditi samo te slojeve, a slojeve koji služe za prepoznavanje općenitijih uzoraka ostaviti kakvima jesu. Nakon treniranja modela još 50 epoha s hiperparametrom $lr=1e-4$ dobivamo rezultate prikazane na slikama 9, 10 i 11.



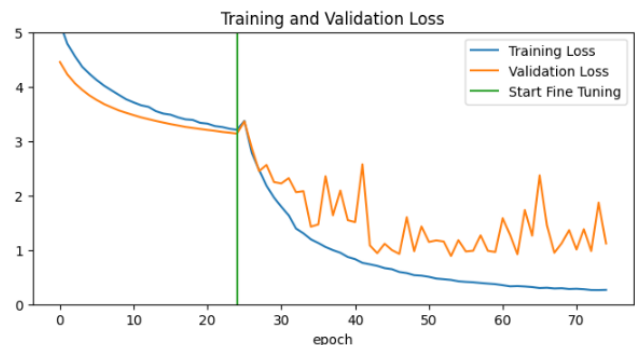
Slika 9. Točnost nakon ugađanja modela

Na kraju treniranja dobivamo rezultate modela:

- točnost treniranja: 0.9159
- f1-mjera treniranja: 0.9393
- gubitak treniranja: 0.2658
- točnost validacije: 0.7529



Slika 10. F1-mjera nakon ugađanja modela



Slika 11. Gubitak nakon ugađanja modela

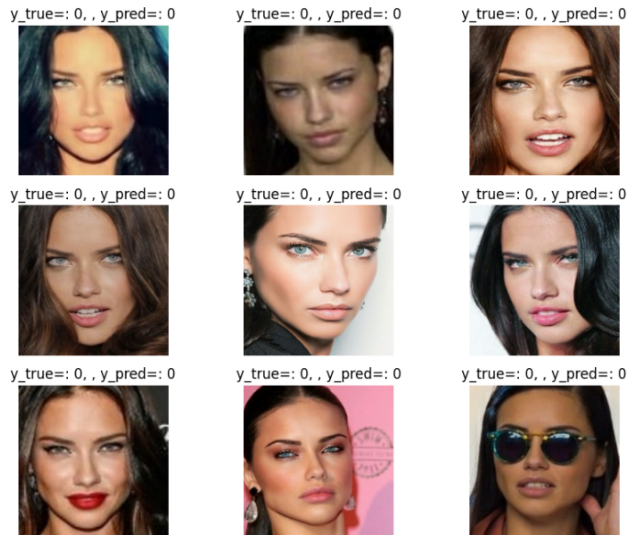
- f1-mjera validacije: 0.7566
- gubitak validacije: 1.1145

Kod validacije rezultati modela dosta variraju od epohe do epohe, razlog tome može biti zbog premalog skupa podataka, zbog transformacija primjenjenih na slike, zato što se dropout ne primjenjuje na skupu za validaciju, ili nekog drugog razloga.

A na skupu za testiranje:

- točnost testiranja: 0.8857
- f1-mjera testiranja: 0.8857
- gubitak testiranja: 0.8860

Predikcije za prvih 9 slika iz skupa za testiranje prikazane su na slici 12.



Slika 12. Predikcije modela za slike iz klase 0

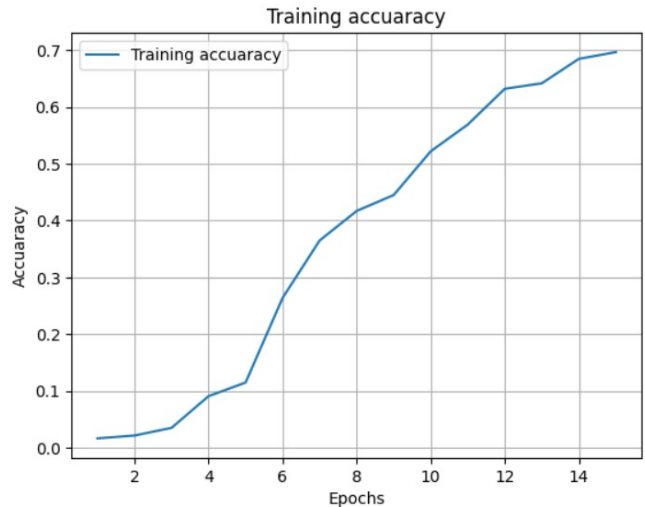


Slika 13. Gubitak treniranja i validacije po epohama.

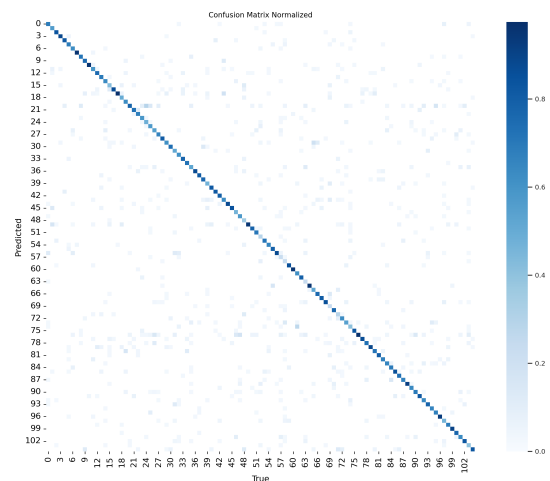
4.2. YOLO

Treniramo model "yolov8s-cls" koji je jedna od varijanti koji se mogu koristiti za klasifikaciju. Radi se o modelu manje kompleksnosti, sa 6.4 milijuna parametara (što je malo u usporedbi s većim varijantama, "yolov8x-cls" ima preko 57 milijuna parametara). Početni broj epoha je 1000, međutim treniranje teče jako sporo, zbog čega se model istrenirao samo kroz 15 epoha. Batch size je postavljen na 32. Slika 13 prikazuje kretanje funkcije gubitka treniranja i validacije tijekom svih 15 epoha.

Također, slika 14 prikazuje nagli porast točnosti klasifikacije sa svakom epohom, što ukazuje na to da bi uz veći broj epoha model imao visoke razine točnosti. Međutim, za potrebe ovoga projekta ograničeno je treniranje zbog velikog



Slika 14. Točnost klasifikacije po epohama



Slika 15. Normalizirana matrica zabune modela

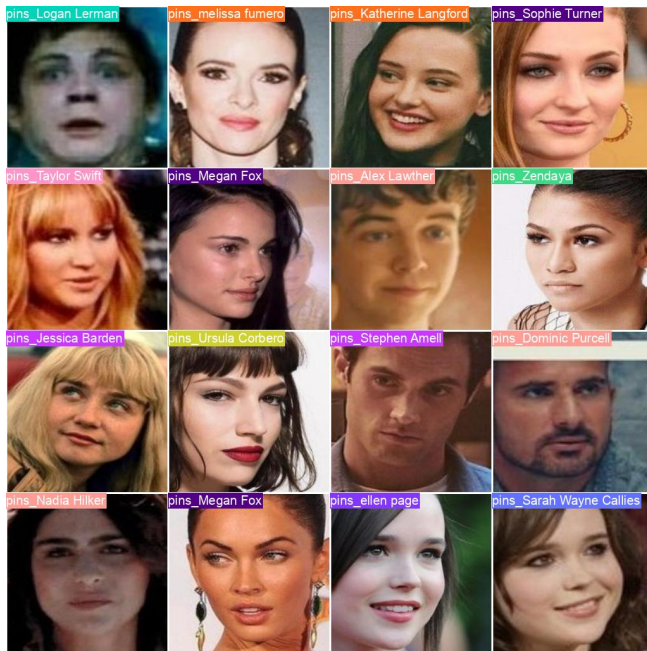
vremenskog utroška i ograničenosti resursa na platformi Google Colab.

Nakon treniranja, obavljeno je testiranje modela. Za evaluaciju su korištene dvije metrike:

- "top1" accuracy, koja predstavlja klasičnu točnost modela (0.696),
- "top5" accuracy, kada je jedna od 5 najvjerovatnijih predikcija modela točna (0.907).

Točnost se u ovom slučaju računa kao udio točno predviđenih predikcija u ukupnom broju predikcija. No, kvalitetu klasifikacije možemo procijeniti i iz tzv. matrice zabune. "Naglašenija" glavna dijagonala matrice sugerira višu preciznost modela, odnosno veći broj točnih klasifikacija.

Za kraj, slika 16 prikazuje izlaz odnosno predikcije modela za jedan "batch" slika.



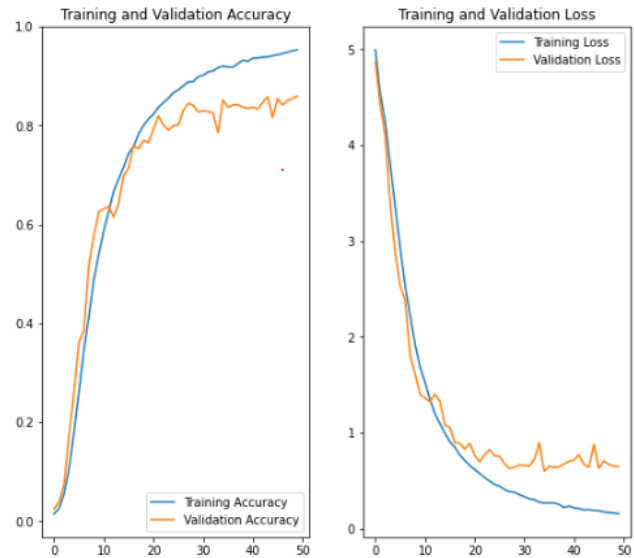
Slika 16. Predikcije modela na primjeru jednog batch-a.

5. Usporedba s rezultatima prethodnih istraživanja

Sličan projekt s istim skupom podataka nalazi se na sljedećoj poveznici [5]. U projektu se koristi model MobileNetv2 te je ostvarena točnost od 85% na skupu za validaciju. Projekt također koristi pretprocesiranje slika s razlikom da također koristi izduživanje slike po određenoj osi, te različite vrijednosti parametra faktor kod pojedinih transformacija. Razlika našeg i navedenog projekta je što se ne treniraju posebno dodatni slojevi, već se direktno ugođuje predtrenirani model, te se treniraju svi slojevi, umjesto onih bližih izlazu. Korištene vrijednosti hiperparametara su za dropout sloj 0.8, a za stopu učenja 0.0001. Dobiveni rezultati su prikazani na slici 17. Vidimo da također vrijednosti variraju na skupu za validaciju, ali u manjoj mjeri nego u našem slučaju.

6. Zaključak

Sustavi za prepoznavanje lica postaju sve važniji u širokom spektru aplikacija za obradu vizualnih podataka, a njihov značaj kao istraživačkog područja naglo raste. Ove implementacije nisu samo sredstvo za suzbijanje kriminala i nadzor, već postaju ključni alati u aktivnostima sigurnosti, verifikaciji identiteta i sličnim područjima. Njihova uloga u suvremenim tehnološkim inovacijama stvara novo razumijevanje ljudske interakcije s računalima te potiče razvoj naprednih sustava u različitim sektorima društva.



Slika 17. Rezultati na skupovima za treniranje i validaciju

Literatura

- [1] BuiltIn. (n.d.). Facial recognition technology explained. Retrieved from <https://builtin.com/articles/facial-recognition-technology-explained>
- [2] HowStuffWorks. (n.d.). How facial recognition systems work. Retrieved from <https://electronics.howstuffworks.com/gadgets/high-tech-gadgets/facial-recognition.htm>
- [3] Biometric Update. (2018). Facial recognition technology: The future of digital identity verification. Retrieved from <https://www.biometricupdate.com/201812/facial-recognition-technology-the-future-of-digital-identity-verification>
- [4] TensorFlow Tutorial, (updated 16.4.2024.) Transfer learning and fine-tuning https://www.tensorflow.org/tutorials/images/transfer_learning
- [5] Srireshram, (2020.) Celebrity face recognition <https://github.com/Srireshram/Celebrity-Face-Recognition>