



UNIVERZITET U NOVOM SADU
PRIRODNO-MATEMATIČKI FAKULTET
DEPARTMAN ZA MATEMATIKU I
INFORMATIKU



Generisanje slika sa *Generative Adversarial Network* modelima sa *CelebA* skupom podataka

-Praktični projekat iz predmeta Eksperimenti nad Neuronskim Mrežama 2-

Marko Mirković, 72m/24

Novi Sad, novembar 2025.

Sadržaj

1 Uvod.....	3
2 Modeli.....	4
2.1 Deep Convolutional Generative Adversarial Network (DCGAN).....	5
2.2 Wasserstein Generative Adversarial Network with Gradient Penalty (WGAN-GP).....	6
3 Eksperimentalna postavka.....	7
3.1 Skupovi podataka.....	7
3.2 Trening.....	8
3.3 Evaluacija.....	9
3.4 Programski jezik i biblioteke.....	10
4 Rezultati.....	11
4.1 Kvantitativna evaluacija.....	11
4.2 Stabilnost treniranja i vizuelna analiza.....	12
4.2.1 DCGAN.....	12
4.2.2 WGAN-GP.....	13
4.2.3 Razlike izmedju vizuelnih i numeričkih kvaliteta.....	15
5 Zaključak.....	16
6 Reference.....	17

1 Uvod

Generisanje novih slika korišćenjem algoritama mašinskog učenja predstavlja oblast koja je u poslednjoj deceniji doživela intenzivan razvoj. Napredak **dubokih neuronskih mreža** omogućio je modelima da uče **kompleksne vizuelne obrasce i strukture**, kao i da na osnovu toga kreiraju nove, do tada nepostojeće primere koji podsećaju na stvarne podatke. Posebno zanimljivo područje jesu modeli koji se primenjuju na **generisanje ljudskih lica**, jer ovaj domen zahteva razumevanje finih detalja, varijacija osvetljenja, pozicije, izraza i mnogobrojnih karakteristika koje čine ljudsko lice vizuelno prepoznatljivim.

Razvoj ovakvih modela ima široku praktičnu primenu. Sintetičke slike koriste se za **proširivanje i balansiranje skupova podataka, unapređenje performansi modela za detekciju i prepoznavanje, generisanje realističnih vizuelnih sadržaja**, kao i u **kreativnim primenama, digitalnoj umetnosti i simulacijama**. Pored praktičnog značaja, istraživanje generativnih tehnika pruža uvid u sposobnost neuronskih mreža da modeluju složene distribucije i da se nose sa izazovima kao što su varijabilnost podataka i visoka dimenzionalnost.

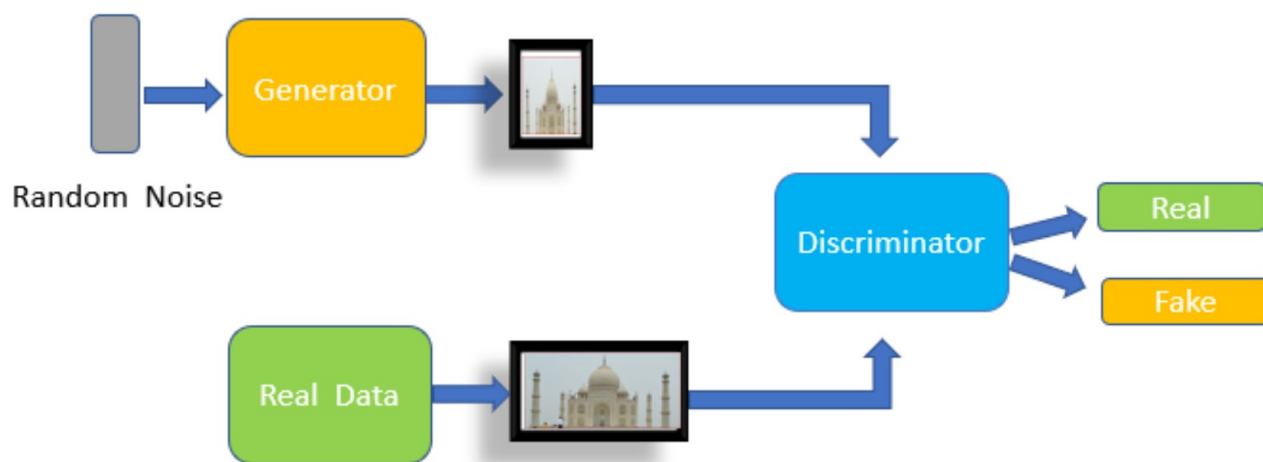
Cilj ovog rada jeste da se kroz eksperimente i uporednu analizu ispita kvalitet generisanja slika ljudskih lica pomoću **savremenih dubokih generativnih modela**. Poseban akcenat stavljen je na stabilnost treniranja, razlike u vizuelnim rezultatima i mogućnosti ovih modela da generišu raznovrsne i realistične uzorke. Kao osnovu za proveru performansi koristi se *CelebA* skup podataka, dok se evaluacija kvaliteta generisanih slika vrši kako vizuelno, tako i pomoću savremenih metričkih pristupa. Na ovaj način rad obezbeđuje širi uvid u mogućnosti i ograničenja generativnih modela pri radu sa kompleksnim slikovnim domenima.

2 Modeli

Generativni modeli zasnovani na dubokim neuronskim mrežama razvijeni su sa ciljem da uče **kompleksne distribucije podataka** i na osnovu naučene strukture **generišu nove**, realistične primere. Tokom poslednjih godina posebno su se izdvojili modeli koji se oslanjaju na **suparničko treniranje** dva odvojena podsistema, pri čemu jedan model generiše slike, dok drugi procenjuje njihovu autentičnost. Ovakav pristup omogućio je postizanje izuzetno uverljivih rezultata u različitim vizuelnim domenima, posebno u generisanju ljudskih lica.

U okviru ovog rada razmatrana su dva reprezentativna pristupa: klasični duboki konvolucionni generativni model, poznat pod nazivom *Deep Convolutional Generative Adversarial Network (DCGAN)*, i njegova **savremenija** varijanta zasnovana na korišćenju *Wasserstein distance* i penalizaciji gradijenata, poznata kao *Wasserstein GAN with Gradient Penalty (WGAN-GP)*. Prvi model predstavlja osnovnu, referentnu arhitekturu koja se često koristi kao polazna tačka za rad sa slikama, dok drugi model uvodi stabilniji način treniranja i omogućava znatno bolji kvalitet generisanih uzoraka.

U nastavku su detaljno objašnjeni principi rada oba modela, njihove arhitekture i ključne razlike koje utiču na stabilnost i kvalitet generisanih slika.



Slika 1: Dijagram GAN arhitekture (high-level)

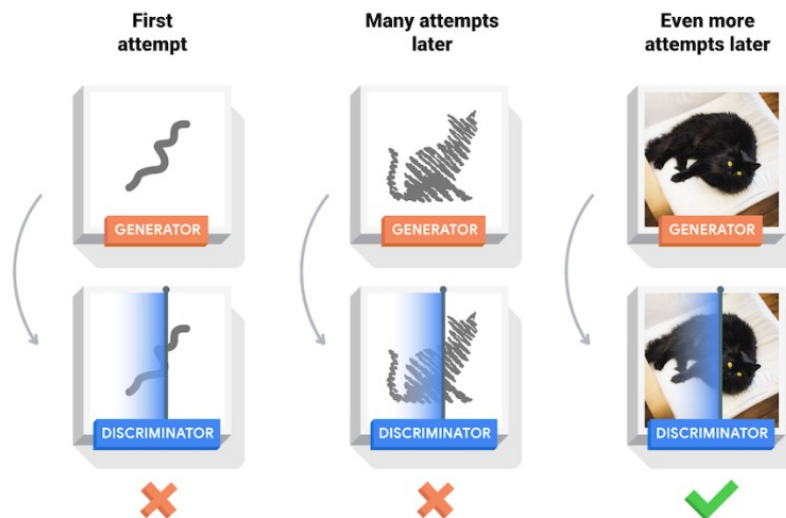
2.1 Deep Convolutional Generative Adversarial Network (DCGAN)

Deep Convolutional Generative Adversarial Network (DCGAN) predstavlja jednu od najuticajnijih i najčešće korišćenih arhitektura u oblasti generativnog modelovanja slika. Predstavljen 2015. godine, *DCGAN* je pokazao da je moguće trenirati stabilne *GAN* modele primenom konvolucionih i transponovanih konvolucionih slojeva, što je omogućilo znatno kvalitetnije rezultate u odnosu na dotadašnje pristupe zasnovane na potpuno povezanim mrežama.

DCGAN se sastoji iz dva podsistema: **generatora** i **diskriminatora**. **Generator** prima slučajni ulazni vektor niskih dimenzija i kroz niz transponovanih konvolucionih operacija ga transformiše u sliku željene rezolucije. Tokom ovog procesa koristi se kombinacija **normalizacije** (*Batch Normalization*) i **nelinearnosti** (*ReLU*), što doprinosi stabilnosti treniranja. Cilj generatora je da nauči mapiranje iz latentnog prostora na vizuelno uverljive slike koje će diskriminator pogrešno klasifikovati kao stvarne.

Discriminator je konvoluciona neuronska mreža koja prima ulazne slike i kroz niz konvolucionih operacija pokušava da odredi da li su one realne ili generisane. Za razliku od generatora, discriminator najčešće koristi *LeakyReLU* kao aktivacionu funkciju, dok se *Batch Normalization* primenjuje samo u dubljim slojevima, ne i na ulazu, kako bi se omogućilo stabilnije učenje. Tokom treniranja generator i discriminator se nalaze u **suparničkom odnosu**: generator unapređuje svoje sposobnosti stvaranja realističnih slika, dok discriminator poboljšava sposobnost razlikovanja realnih od sintetičkih uzoraka.

Ovakav način treniranja dovodi do specifičnih problema, kao što su **dominacija diskriminatora**, **eksplodirajući gradijenti** i **mode collapse**, situacija u kojoj generator proizvodi veoma ograničen broj različitih uzoraka. Uprkos tim izazovima, *DCGAN* je značajan jer predstavlja čvrstu osnovu za razvoj stabilnijih arhitektura i služi kao referentni model u različitim istraživanjima iz oblasti generativnog modelovanja.



Slika 2: Vizualizacija DCGAN-a

2.2 Wasserstein Generative Adversarial Network with Gradient Penalty (WGAN-GP)

Wasserstein Generative Adversarial Network (WGAN) uveden je kao odgovor na ograničenja klasičnih *GAN* arhitektura, posebno na probleme nestabilnog treniranja i pojave *mode collapse*-a. Ključna ideja *WGAN*-a jeste zamena standardne divergencije (npr. *Jensen-Shannon* divergencije koju implicitno koriste klasični *GAN*-ovi) sa *Wasserstein* (*Earth-Mover*) distancom, koja predstavlja znatno glatkiju i informativniju meru razlike između realne i generisane distribucije podataka. U praksi, ova distanca omogućava modelima stabilniji tok gradijenata tokom treniranja i time olakšava proces optimizacije.

Za procenu *Wasserstein* distance koristi se **kritičar** (*critic*), koji se strukturno ponaša slično diskriminatoru, ali umesto binarne klasifikacije proizvodi realne skalarne vrednosti. Tokom treniranja kritičar pokušava da maksimizuje razliku između srednjih vrednosti svojih ocena nad realnim i generisanim uzorcima, dok generator minimizuje tu razliku generišući slike koje kritičar rangira što sličnije realnim primerima.

Jedan od ključnih uslova za korektnu procenu *Wasserstein* distance jeste da kritičar mora biti *1-Lipschitz* funkcija. Originalni *WGAN* je ovo implementirao jednostavnim *clippingom* težina, ali je takav pristup pokazao brojne probleme, poput slabe reprezentativne moći mreže i oscilatornog treniranja. Da bi se ovaj problem prevazišao, uveden je *Wasserstein GAN with Gradient Penalty (WGAN-GP)*, koji uvodi znatno stabilniji i elegantniji način obezbeđivanja *Lipschitz* ograničenja.

Umesto sečenja težina, *WGAN-GP* dodaje penalizaciju gradijenata u funkciju gubitka kritičara. *Gradient penalty* term podstiče norme gradijenata kritičara da budu bliske vrednosti 1 na interpolacijama između realnih i generisanih slika. Time se obezbeđuje zadovoljenje *Lipschitz* uslova bez narušavanja kapaciteta mreže i bez potrebe za ručnim podešavanjem opsega vrednosti parametara. Rezultat je znatno **stabilnije** treniranje, glatkiji tok gradijenata, kao i manja sklonost ka *mode collapse*-u.

WGAN-GP se u praksi pokazao kao **pouzdanija** i **stabilnija** arhitektura za generisanje slika. U poređenju sa klasičnim *GAN* i *DCGAN* modelima, omogućava stabilnije učenje, manje oscilacije tokom optimizacije i generisanje vizuelno kvalitetnijih i raznovrsnijih uzoraka. Ove prednosti čine *WGAN-GP* pogodnijim izborom za generisanje **kompleksnih vizuelnih domena**, poput ljudskih lica u *CelebA* skupu podataka koji se koristi u ovom radu.

3 Eksperimentalna postavka

3.1 Skupovi podataka

Za potrebe ovog rada korišćen je **CelebA** (*CelebFaces Attributes Dataset*), jedan od poznatijih skupova podataka u oblasti generisanja slika ljudskih lica. Skup sadrži ukupno **202,599** slika sa oznakama za **10,177** različitih identiteta, kao i **40** anotiranih atributa (npr. pol, položaj glave, naočare, izraz lica). Slike su raspoređene u tri podskupa: trening, validacioni i test, pri čemu je najveći deo rezervisan za trening, što ga čini pogodnim za rad sa generativnim modelima.

Originalne slike u *CelebA* skupu su visoke rezolucije i variraju u pogledu osvetljenja, pozadine i izraza, što ovaj skup čini izazovnim, ali i veoma informativnim za učenje generativnih modela. Zbog širokog spektra varijacija u ljudskim licima, *CelebA* je postao standardni test za procenu kvaliteta *GAN* modela.

Preprocesiranje podataka sprovedeno je tako da se slike najpre centrično seku na dimenziju 178x178, čime se eliminišu delovi pozadine koji nisu informativni. Nakon toga se skaliraju na ciljnu rezoluciju 128x128, koja predstavlja kompromis između vizuelnog kvaliteta i računarske efikasnosti treniranja. Slike se zatim transformišu u tenzore i normalizuju na opseg vrednosti $[-1, 1]$, što je standardna procedura kod *GAN* modela zbog upotrebe funkcije *tanh* u izlaznom sloju generatora.

CelebA skup podataka, zahvaljujući svojoj raznovrsnosti i velikom broju uzoraka, obezbeđuje adekvatnu osnovu za učenje modela koji treba da generalizuju složene vizuelne obrasce i generišu nova, realistična lica. Time predstavlja relevantnu platformu za poređenje performansi *DCGAN*-a i *WGAN-GP* modela u okviru ovog rada.



Slika 3: Primer slika iz CelebA skupa podataka

3.2 Trening

DCGAN i *WGAN-GP* se razlikuju po funkciji gubitka, uslovima optimizacije i načinu obezbeđivanja stabilnog toka gradijenata.

Oba modela trenirana su korišćenjem **Adam** optimizatora, s uobičajenim hiperparametrima koji su standardno preporučeni za generativne modele. Za *DCGAN* se koristi funkcija gubitka zasnovana na **binarnoj klasifikaciji**, kojom discriminator uči da razlikuje realne slike od generisanih, dok generator pokušava da obmane discriminator maksimizovanjem verovatnoće da su generisani uzorci klasifikovani kao realni. Tokom treniranja **generator** i **discriminator** se smenjuju u ažuriranju parametara, pri čemu se discriminator ažurira nakon svakog novog para realnog i generisanog uzorka.

Trening *WGAN-GP* modela razlikuje se po načinu **procene udaljenosti** između realnih i generisanih distribucija. Umesto klasičnog diskriminatora koristi se kritičar, čiji je cilj da maksimizuje razliku između prosečnih ocena nad realnim i generisanim uzorcima. Kako bi se obezbedilo zadovoljenje *Lipschitz* uslova, u funkciju gubitka uvodi se penalizacija gradijenata, koja ograničava normu gradijenata kritičara na željeni opseg. U skladu sa preporukama iz literature, kritičar se trenira više puta u odnosu na generator (tipično **pet** iteracija kritičara na **jednu** iteraciju generatora), što doprinosi stabilnijem učenju i glatkijem toku gradijenata.

Tokom treniranja oba modela primenjena je i tehnika ***Exponential Moving Average (EMA)*** za parametre generatora. *EMA* omogućava da se paralelno održava „srednja“ verzija generatora, koja nije podložna kratkoročnim oscilacijama i često daje ujednačenije i vizuelno kvalitetnije rezultate. Za evaluaciju generisanih slika nakon završetka treniranja korišćena je upravo *EMA* verzija generatora.

Broj epoha treniranja podešen je tako da omogući konvergenciju modela **bez prenaučavanja** ili **mode collapse**-a. Korišćena veličina mini-batch-a iznosila je **64**, dok je latentni prostor modela dimenzije **128**. Nakon završetka treniranja generisano je **10,000** slika za potrebe kvantitativne evaluacije korišćenjem savremenih metrika kvaliteta.

3.3 Evaluacija

Evaluacija generativnih modela predstavlja poseban izazov, jer vizuelna uverljivost generisanih slika ne može uvek da se opiše jednostavnim numeričkim metrikama. Tokom poslednjih godina razvijeni su standardizovani metodi koji omogućavaju kvantitativno poređenje kvaliteta generisanih uzoraka sa realnim podacima. U ovom radu korišćene su dve najprihvaćenije metrike u savremenoj literaturi: *Fréchet Inception Distance (FID)* i *Kernel Inception Distance (KID)*. Pored njih, sprovedena je i vizuelna evaluacija koja omogućava kvalitativni uvid u sposobnost modela da generiše raznovrsne i realistične primere.

FID računa sličnost između dve distribucije na osnovu reprezentacija slika ekstrahovanih iz duboke konvolucione mreže (*Inception-V3*). Što je vrednost *FID* niža, to su uzorci prve distribucije sličniji podacima druge u smislu statističke strukture visoko-nivoijskih karakteristika. Ova metrika je posebno korisna jer **dobro korelira sa ljudskim opažanjem kvaliteta slika**.

KID predstavlja komplementarnu metriku koja, za razliku od *FID*-a, ne zahteva procenu kovarijanse i zasniva se na proceni maksimalnog srednjeg odstupanja pomoću polinomijalnog kernela. *KID* je ne-pristrasna mera i stabilnija je kod manjih skupova podataka, što je čini pogodnom za poređenje modela kada se radi sa ograničenim brojem generisanih slika. Niže vrednosti *KID*-a ukazuju na sličniju distribuciju generisanih i stvarnih podataka.

Za potrebe evaluacije, nakon završetka treniranja svakog modela generisano je **10,000** slika, koje su zatim poređene sa slučajno odabranim realnim uzorcima iz test skupa *CelebA*. Time su obezbeđeni uslovi za pouzdanu statističku procenu. Pored numeričkih metrika, vizuelne reprezentacije generisanih slika korišćene su kako bi se kvalitativno procenila raznovrsnost uzoraka, prisustvo artefakata, kao i eventualna pojava *mode collapse*-a.

Kombinovanjem rezultata dobija se celovita slika o ponašanju modela tokom učenja i o njihovoj sposobnosti da generišu realistične uzorke. Ovaj pristup omogućava objektivno poređenje *DCGAN* i *WGAN-GP* modela i pruža uvid u njihove prednosti i ograničenja u zadatku generisanja slika ljudskih lica.

3.4 Programski jezik i biblioteke

Implementacija modela i eksperimentalnih procedura realizovana je u programskom jeziku *Python*, koji se smatra standardom u oblasti **dubokog učenja**. Treniranju modela pristupljeno je kroz *Google Colab* i *Vast.ai* platforme, koje omogućavaju jednostavan pristup grafičkim procesorima (GPU). Na *Google Colab*-u je korištena *Nvidia Tesla T4*, dok je na *Vast.ai* korištene *Nvidia RTX 5080* i *RTX 5090*.

Za izgradnju generativnih mreža korišćen je *PyTorch*, jedan od najpopularnijih *framework*-a za duboko učenje. *PyTorch* pruža fleksibilno definisanje neuronskih arhitektura, efikasno izvođenje operacija nad tenzorima i automatsku računicu gradijenata, što ga čini pogodnim za rad sa kompleksnim modelima poput *DCGAN*-a i *WGAN-GP*-a. Biblioteka *torchvision* korišćena je za učitavanje i transformaciju slika, uključujući operacije centričnog sečenja, promene rezolucije i normalizacije. Za obradu i vizuelizaciju rezultata korišćene su biblioteke *matplotlib* i *torchvision.utils*, koje omogućavaju prikaz generisanih slika tokom i nakon treniranja modela.

Ovakav skup alata omogućio je stabilno, transparentno i reproduktivno izvođenje eksperimenata, pri čemu su sve faze rada – učitavanje podataka, treniranje modela, generisanje slika i evaluacija – realizovane u jedinstvenom programskom okruženju.

4 Rezultati

U ovom poglavlju prikazani su **kvantitativni** i **kvalitativni** rezultati dobijeni treniranjem *DCGAN* i *WGAN-GP* modela na *CelebA* skupu podataka rezolucije 128×128 , uz korišćenje *EMA* verzije generatora za generisanje slika nakon treniranja kako bi dobili što stabilnije slike. Za svaki model generisano je 10,000 slika koje su zatim upoređene sa podskupom od 10,000 realnih slika iz *CelebA* skupa.

4.1 Kvantitativna evaluacija

Kvalitet generisanih slika kvantitativno je procenjen korišćenjem *FID* i *KID* metrika. Niže vrednosti ovih metrika ukazuju na veću sličnost između distribucije generisanih slika i distribucije realnih podataka. Za potrebe evaluacije korišćeni su sledeći eksperimenti:

- Slike generisane *DCGAN*-om naspram realnih slika;
- Slike generisane *WGAN-GP*-em naspram realnih slika;
- Poređenje *DCGAN* generisanih i *WGAN-GP* generisanih slika međusobno.

comparison	FID	KID
DCGAN vs Real	76.931600	0.063693
WGAN-GP_100 vs Real	90.365100	0.084355
DCGAN vs WGAN-GP_100	39.624049	0.039140

Slika 4: Vrednosti *FID* i *KID* metrika po eksperimentu

Posmatrajući *FID* i *KID* vidi se da *DCGAN* postiže **numerički bolje rezultate** u odnosu na *WGAN-GP* model: *DCGAN* ima niži *FID* i niži *KID* u poređenju sa realnim podacima. To je **suprotno** očekivanjima zasnovanim na literaturi, gde *WGAN-GP* u pravilu daje stabilnije treniranje i kvalitetnije uzorke, ali oslikava konkretne uslove eksperimenta i ograničenja treniranja u ovom radu.

Dodatno, *FID* i *KID* između *DCGAN* i *WGAN-GP* modela pokazuju da se distribucije ova dva modela međusobno primetno razlikuju ($FID = 39.6$), što ukazuje da modeli „uče“ **različite reprezentacije unutar istog domena lica**.

4.2 Stabilnost treniranja i vizuelna analiza

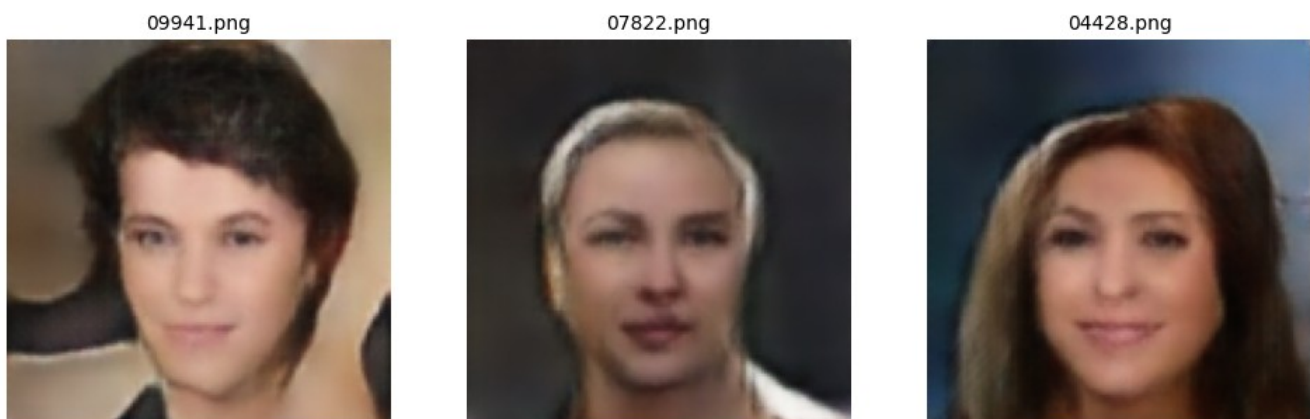
Za početak ćemo prikazati tri nasumično odabrane fotografije iz originalnog skupa podataka, kako bi mogli da ih poredimo sa generisanim slikama.



Slika 5: Slike iz realnog skupa podataka

4.2.1 DCGAN

DCGAN je treniran 16 epoha (inicijalno 15 epoha, uz dodatnu epohu naknadnog treniranja) sa Adam optimizatorom i standardnim hiperparametrima. Tokom treniranja uočava se relativno **brz** prelazak sa potpunog šuma na prepoznatljive konture lica, nakon čega se kvalitet slika **postepeno** poboljšava, ali **bez dramatičnih promena** u kasnijim epohama.



Slika 6: Slike iz skupa podataka generisanog DCGAN-om

Kod *DCGAN*-a se povremeno javljaju:

- zamućenja u pojedinim regionima slike,
- artefakti na ivicama lica ili u pozadini,
- povremeno nešto manja raznovrsnost poza i osvetljenja.

I pored toga, model **uspešno uči** osnovnu strukturu ljudskog lica (položaj očiju, usta, kose) i generiše slike koje su vizuelno ubedljive, što potvrđuju i relativno dobre *FID* i *KID* vrednosti.

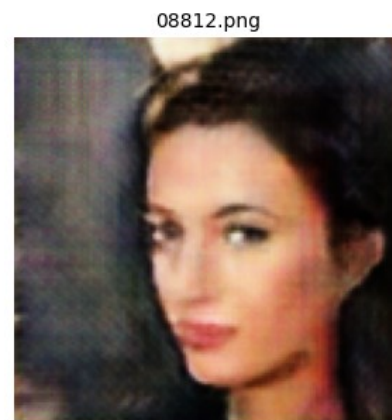
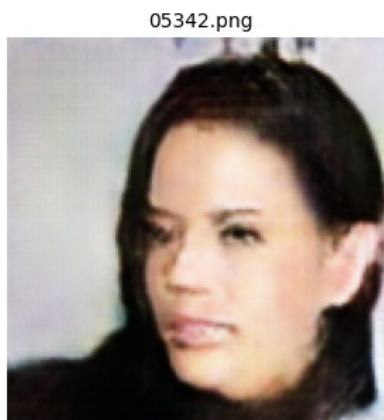
Pored solidnih vizuelnih rezultata, *DCGAN* je **veoma težak** model za treniranje usled njegove **velike nestabilnosti**. Vrlo lako dolazi do eksplozije gradijenata, *mode collapse*-a ili prenaučjenja. Ove negativne pojave se mogu desiti i u jednoj epohi, gde su rezultati jedne epohe skroz normalni dok u sledećoj slike postanu neupotrebljive.

Korisno je objasniti i *mode collapse* kao pojavu kod modela mašinskog učenja, dok je primer *GAN* arhitektura najinteresantniji. U ovom slučaju *mode collapse* je kada generator shvati da jedna specifična slika koju je generisao ima **ocenu kao realna** od diskriminatora i onda kontinualno generiše samo tu sliku. Ova pojava čini model **neupotrebljivim** jer su svi njegovi parametri okrenuti ka toj jednoj lošoj slici. Prilikom treninga preporučeno je generisati **više** slika kako bi primetili njihove razlike. U slučaju *mode collapse* ćete videti kako se sve slike pretvaraju u jednu sa veoma čudnim karakteristikama ili artifaktima.

4.2.2 *WGAN-GP*

WGAN-GP je treniran **znatno** duže - 100 epoha sa brzinom učenja $1e-4$, uz dodatnih 10 epoha tokom kojih je brzina učenja generatora smanjena na $5e-5$ kako bi se postiglo fino doterivanje i potencijalno „izoštavanje“ detalja. Kritičar je tokom treniranja ažuriran više puta po jednoj iteraciji generatora ($n_{critic} = 7$), u skladu sa preporukama za *WGAN-GP*. Sa aspekta stabilnosti treniranja, *WGAN-GP* pokazuje očekivano ponašanje:

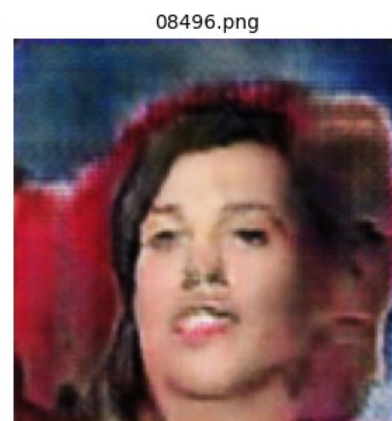
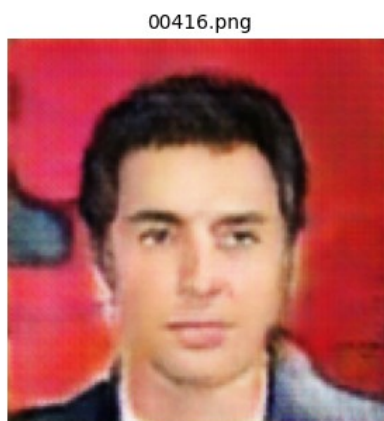
- gubitak kritičara se menja postepeno, bez naglih oscilacija;
- nema jasnih znakova ozbiljnog *mode collapse*-a;
- generisane slike u kasnijim epohama prikazuju uravnotežene konture lica i relativno raznovrsne primere.



Slika 7: Slike iz skupa podataka generisanog WGAN-GP-em

Vizuelno, uzorci *WGAN-GP* modela često deluju **ujednačeno** i **prirodno**: lica imaju dobro definisane globalne strukture, a raspored osvetljenja i pozadina je generalno **konzistentan**. Ipak, model i dalje pokazuje:

- poneka lokalna izobličenja (npr. u predelu očiju ili usta);
- blaga zamućenja kod složenijih pozadina;
- izvesnu „mekoću“ tekstura koja može negativno uticati na *FID/KID*, jer *Inception* mreža osetljivo reaguje na fine razlike u raspodeli lokalnih karakteristika.



Slika 8: Slike iz skupa podataka generisanog WGAN-GP-em sa blago izmenjenim hiperparametrima

4.2.3 Razlike između vizuelnih i numeričkih kvaliteta

Iako se vizuelno može steći utisak da *WGAN-GP* proizvodi barem **jednako dobre**, a ponekad i ubedljivije slike od *DCGAN*-a, numeričke metrike *FID* i *KID* **favorizuju** *DCGAN*. Postoji nekoliko razloga za ovakvu pojavu:

1. **Ograničeno podešavanje hiperparametara *WGAN-GP* modela:** *WGAN-GP* je osetljiv na izbor hiperparametara poput broja kritičevih koraka, težine gradient penalty termina i brzina učenja za generator i kritičar. U ovom radu korišćene su razumno postavljene, ali ne nužno optimalne vrednosti. Uz detaljnije pretraživanje hiperparametara i duže treniranje (ili različite šeme spuštanja brzine učenja), očekivano je da bi *WGAN-GP* dodatno poboljšao *FID/KID*.
2. **Osetljivost *FID/KID* metrika na sitne statističke razlike:** *FID* i *KID* se oslanjaju na reprezentacije iz *Inception* mreže trenirane na *ImageNet* skupu. Model koji generiše nešto širu raznovrsnost ili drugačiji raspored tekstura i pozadina može numerički izgledati „dalje“ od realnih *CelebA* slika, čak i kada su slike vizuelno uverljive. *DCGAN* može delimično „prilagoditi“ distribuciju svojih uzoraka specifičnim statistikama koje *Inception* ekstrahuje, što se odražava u nešto boljim vrednostima metrika.
3. **Ograničenja eksperimentalnog okvira:** *WGAN-GP* je treniran duže, ali je trening zaustavljen u trenutku kada su promene vizuelnog kvaliteta i metrika postale minimalne. Moguće je da je model ostao u lokalnom optimumu koji vizuelno daje zadovoljavajuće rezultate, ali još uvek nije statistički dovoljno približen realnoj distribuciji da bi nadmašio *DCGAN* u *FID/KID* smislu.

U celini, kvantitativna evaluacija pokazuje da *DCGAN* u ovom konkretnom eksperimentalnom okviru postiže **bolje numeričke metrike**, dok *WGAN-GP* pruža **stabilnije treniranje i vizuelno raznovrsne uzorke**, ali uz nešto veće vrednosti *FID* i *KID* metrika.

5 Zaključak

U radu su upoređene performanse dva savremena generativna modela – *DCGAN* i *WGAN-GP* – na zadatku generisanja slika ljudskih lica sa *CelebA* skupa podataka rezolucije 128×128 . Fokus je bio na stabilnosti treniranja, vizuelnom kvalitetu generisanih slika i kvantitativnoj evaluaciji korišćenjem *FID* i *KID* metrika.

Eksperimenti su pokazali da oba modela uspešno uče strukturu domena i generišu slike koje vizuelno podsećaju na realna lica. *DCGAN* relativno brzo dolazi do prepoznatljivih uzoraka, ali i dalje povremeno proizvodi zamućenja i artefakte. *WGAN-GP* obezbeđuje stabilniji tok treniranja, manju sklonost ka naglim oscilacijama i zadovoljavajuću raznovrsnost generisanih primera.

Iako *WGAN-GP* nije nadmašio *DCGAN* po numeričkim metrikama, analiza ukazuje da ovaj model i dalje nudi važne praktične prednosti – pre svega u pogledu stabilnosti učenja i opšteg kvaliteta uzoraka. U kontekstu šireg istraživanja, *WGAN-GP* ostaje perspektivan pristup, ali zahteva pažljivije projektovanje trening procedure da bi se u potpunosti iskoristio njegov potencijal.

Potencijalni pravci daljeg rada uključuju:

- Detaljnije pretraživanje hiperparametara *WGAN-GP* modela (broj koraka kritičara, težina *gradient penalty* termina, šeme spuštanja brzine učenja),
- Treniranje na većem broju epoha ili pri drugoj rezoluciji slika,
- Primenu naprednijih arhitektura, kao što su *StyleGAN* ili modeli sa dodatnim uslovljavanjem (npr. na attribute iz *CelebA* skupa),
- Poređenje više različitih metrika kvaliteta i ispitivanje njihove korelacije sa subjektivnim ljudskim opažanjem.

Na taj način, dobijeni rezultati predstavljaju dobru polaznu tačku za dalje istraživanje generativnih modela i njihovih primena u domenu generisanja slika ljudskih lica.

6 Reference

- [1] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative Adversarial Nets” <https://arxiv.org/abs/1406.2661>
- [2] A. Radford, L. Metz, and S. Chintala, “Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks” <https://arxiv.org/abs/1511.06434>
- [3] M. Arjovsky, S. Chintala, and L. Bottou, “Wasserstein GAN” <https://arxiv.org/abs/1701.07875>
- [4] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. Courville, “Improved Training of Wasserstein GANs” <https://arxiv.org/abs/1704.00028>
- [5] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, “GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium” <https://arxiv.org/abs/1706.08500>
- [6] M. Bińkowski, D. J. Sutherland, M. Arbel, and A. Gretton, “Demystifying MMD GANs” <https://arxiv.org/abs/1801.01401>
- [7] Syed Abdul Gaffar, “Generate Your Own Dataset using GAN” Analytics Vidhya, 07 Jan 2025. Available: <https://www.analyticsvidhya.com/blog/2021/04/generate-your-own-dataset-using-gan/>
- [8] TensorFlow Team, “Deep Convolutional Generative Adversarial Network (DCGAN)” TensorFlow Tutorials, 2024. Available: <https://www.tensorflow.org/tutorials/generative/dcgan>
- [9] Z. Liu, P. Luo, X. Wang, X. Tang, “Large-scale CelebFaces Attributes (CelebA) Dataset” Multimedia Laboratory, The Chinese University of Hong Kong, 2015. Available: <https://mmlab.ie.cuhk.edu.hk/projects/CelebA.html>