

Seminarski rad B

Osnovni principi neuralnih mreža

Super rezolucija slika sa SRGAN modelom

autor: Dejan Gjer

mentor: Miloš Racković

Univerzitet u Novom Sadu - Prirodno-matematički fakultet

10.7.2021.

Sadržaj

1 Uvod	2
2 Generativne suparničke mreže	2
3 Struktura modela	3
3.1 Generatorska mreža	3
3.2 Diskriminatorska mreža	3
4 Skup podataka	4
5 Implementacija	5
5.1 SRResNet	5
5.2 SRGAN	6
5.2.1 Konfiguracija sa MSE gubitkom	7
5.2.2 Konfiguracija sa kombinovanim gubitkom	8
6 Rezultati	9
7 Zaključak	10

1 Uvod

U ovom radu biće prikazan i implementiran model SRGAN i pojedinačno njegov generator SRResNet za rešavanje problema super-rezolucije. Super-rezolucija je problem dobijanja slika više rezolucije od slika niže rezolucije i za njegovo rešavanje pre naglog razvoja kompjuterske vizije koristili su se najčešće algoritmi filtriranja kao što su bilinearni, bikubični i Lanczos. Oni brzo generišu rezultate ali najčešće nisu uspevali da kreiraju sitne detalje sa slika i kao bolje rešenje pokazale su se metode iz dubokog učenja korišćenjem neuralnih mreža. Dataset koji će u radu biti korišćen je Celaba-hq, koji se sastoji od slika poznatih ličnosti, i fokus u radu će biti na skaliranju tih slika 4 puta (tj. 16 puta veći broj piksela) sa slika dimenzije 64x64 na slike dimenzija 256x256. Modeli koji su implementirani prate originalni SRGAN rad [1] iz 2017. godine.

2 Generativne suparničke mreže

Generativne suparničke mreže (skraćeno GAN) prvi put su se pojavile u radu 2014. godine [2]. One predstavljaju vrstu arhitekture za generativne modele, koje najčešće koriste duboke neuralne mreže za zadatke generisanja novih podataka. Gan-ovi se najčešće sastoje iz dva modela dubokih neuralnih mreža, generatora i diskriminatora. Generator služi za kreiranje podataka (u ovom radu slika), dok diskriminator ima za zadatak da odlučuje koje slike su realne, a koje lažne tj. generisane. U suštini generator i diskriminator igraju minimax igru, gde postoji suparnički gubitak u obliku:

$$E_x[\log D(x)] + E_z[\log(1 - D(G(z)))] \quad (1)$$

gde je:

- $D(x)$ je procena diskriminatora da je podatak x realan (verovatnoća)
- E_x je očekivana vrednost za sve realne vrednosti podataka
- $G(z)$ je izlaz generatora za ulaz z (z je vrlo često nasumični šum, a u ovom radu će biti slika male rezolucije)
- $D(G(z))$ je procena diskriminatora da je generisana vrednost iz generatora realna (verovatnoća)
- E_z je očekivana vrednost za sve moguće vrednost z

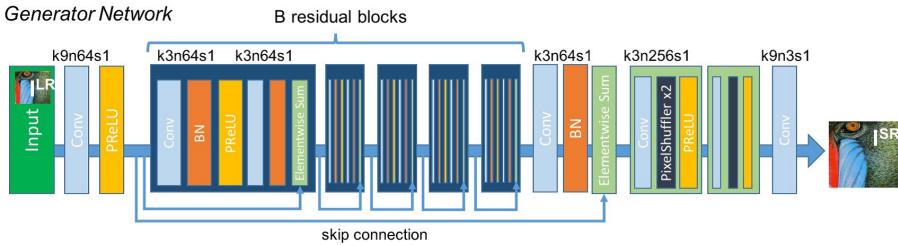
Diskriminator se trudi da ovaj gubitak bude blizak 0, a generator da umanji desni sabirak (pošto na levi ne može da utiče), tj. da maksimizuje izraz $\log D(G(z))$.

3 Struktura modela

SRGAN model, pošto predstavlja generativnu suparničku mrežu, sastoji se od generatora i diskriminatora. Generator je SRResNet duboka rezidualna mreža i može i sama da se koristi za dobijanje pristojnih rezultata za problem super rezolucije. Diskriminator je takođe duboka neuralna mreža koja služi za binarnu klasifikaciju slike na realne i generisane (lažne). Detaljnije o njihovim strukturama sledi u naredna dva poglavlja.

3.1 Generatorska mreža

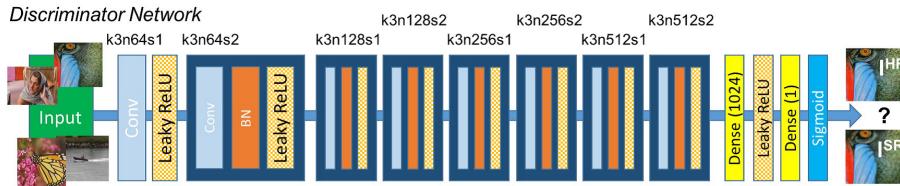
Generator je rezidualna mreža sa skip-konekcijama. Sastoji se iz B rezidualnih blokova sa identičnom strukturom. Svaki blok sastoji se iz dva konvolucionih sloja sa 64 filtera korakom 1 i dimenzijom kernela 3x3. Ovi slojevi praćeni su batch normalizacionim slojem, kao i ParametricReLU aktivacionom funkcijom. Svaki rezidualni blok je takođe skip konekcijom povezan sa izlazom prethodnog bloka. U mojoj implementaciji SRResNet mreže korišćeno je $B = 8$ rezidualnih blokova. Nakon njih slike se uvećavaju pomoću sub-piksel konolutivnih slojeva, gde svaki od njih uvećava sliku 2 puta. Prikaz strukture generatora možete videti na slici ispod.



Slika 1: Prikaz strukture generatora (SRResNet)

3.2 Diskriminatorska mreža

Diskriminator je takođe duboka neuralna mreža sastavljena iz 8 blokova, gde svaki sadrži jedan konvolucijski sloj dimenzije kernela 3x3, gde svaki neparni uvećava broj filtera 2 puta, a svaki parni pravi korake dužine 2. Ti slojevi praćeni su batch-normalizacijom i LeakyReLU aktivacionom funkcijom sa parametrom $\alpha = 0.2$. Na ovaj način se smanjuju dimenzije slike po 2 puta. Nakon blokova model sadrži potpuno povezane slojeve sa sigmoidnom aktivacionom funkcijom na kraju kako bi dobili verovatnoće da li je slika realna ili lažna. Prikaz strukture diskriminatora možete videti na slici ispod.

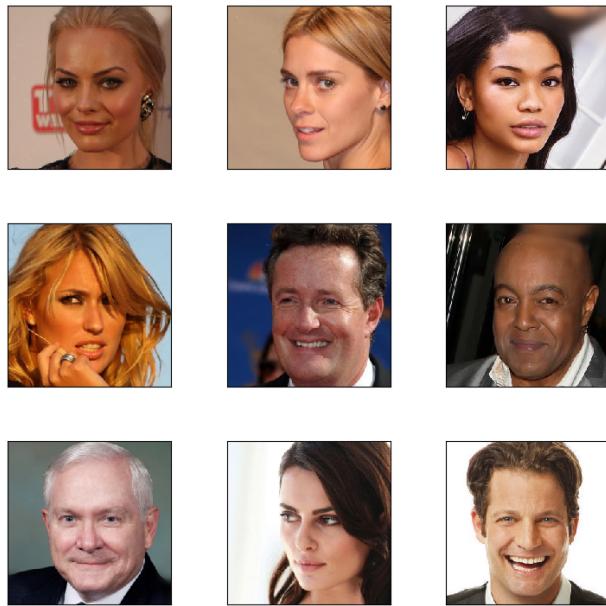


Slika 2: Prikaz strukture diskriminatora

4 Skup podataka

Skup podataka koji je korišćen u ovom radu je Celaba-hq dataset. On se sastoji od 30000 visoko kvalitetnih slika lica poznatih ličnosti. Za potrebljeno treniranje preuzet je dataset sa dimenzijama slika 256x256 (slike visoke rezolucije), koje su potom bilinearno smanjivane na dimenziju 64x64 (slike niske rezolucije). Ceo skup podataka podeljen je nasumično na 3 dela:

- trening skup - čine ga 28000 slika na kojima se model trenira
- validacioni skup - čini ga 1900 slika, na kojima model ne trenira, ali se proveravaju rezultati u toku i po završetku treninga, da bi se uverili da model nije preučio (overfitting)
- test skup - čini ga 100 slika koje su isključivo za prezentaciju



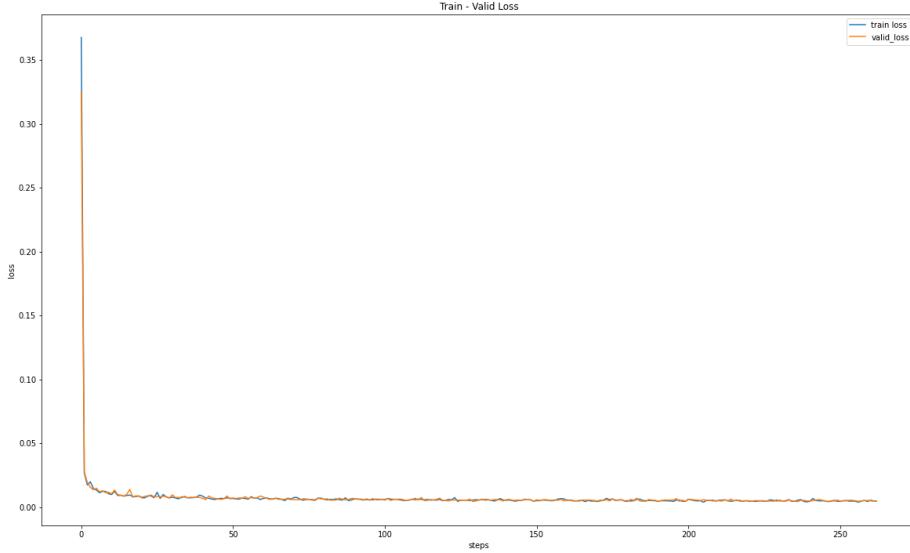
Slika 3: Primeri slika iz dataset-a

5 Implementacija

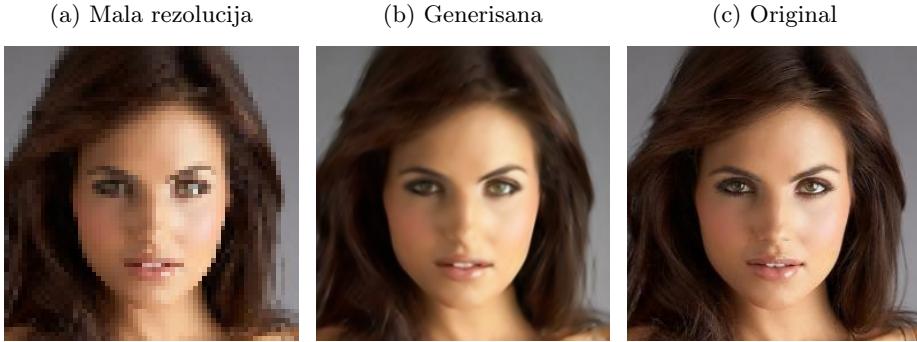
U ovoj sekciji biće navedeni implementacioni detalji za pojedinačan trening SRResNet-a, kao i celog SRGAN-a. Kod vezan za ovaj rad možete pogledati na github-u.

5.1 SRResNet

Ovaj model za ulaz prima RGB slike niske rezolucije dimenzija (64,64,3), gde svi pikseli trebaju da budu skalirani na opseg [0..1]. Zato je ulaz prethodno preprocesiran tako da je skaliran sa prirodnog opsega [0..255] na traženi. Kao optimizator ove mreže koristio se Adam sa koeficijentom učenja $\alpha = 0.0002$ i preporučenim parametrima $\beta_1 = 0.9$ i $\beta_2 = 0.999$. Ovakva konfiguracija optimizatora koristila se za treniranje svih modela u ovom radu. Funkcija gubitka za ovaj model je srednja kvadratna greška (mse - mean square error). Model je treniran 30 epoha, mada se ukupna greška posle nekoliko prvih epoha nije mnogo smanjivala, tako da je trening mogao da se prekine i ranije. Izlaz ovog modela su slike visoke rezolucije (256,256,3) skalirane na opseg [-1..1] jer je aktivacija poslednjeg konvolutivnog sloja hiperbolični tangens (tanh funkcija). Model je treniran po mini serijama veličine 64. Ovako istrenirana mreža koristi se kao generator za SRGAN model. Gubitke u toku treniga možete videti na slici 4, a primer rezultata na slici 5.



Slika 4: Grafik gubitaka na trening i validacionom skupu u zavisnosti od koraka pri treniranju



Slika 5: Prikaz slika za SRResNet model

5.2 SRGAN

SRGAN model za mrežu generatora koristi već prethodno istreniranu SRResNet mrežu. Diskriminatator je implementiran po modelu iz poglavlja 3.2. Ulazni podaci diskriminatatora su slike visoke rezolucije u opsegu $[-1..1]$, a izlaz je realan broj između 0 i 1. Broj blizu 1 predstavlja da je slika realna, a blizu 0 da je lažna (kao izlaz sigmoidne funkcije). Za funkciju gubitka koriste se binarna kro-sentropija, jer je u pitanju binarni klasifikator. SRGAN model je implementiran kao model koji se sastoji od generatora i diskriminatatora, gde je ulaz slika niske rezolucije, potom one prolaze kroz generator i dobijamo generisane slike visoke rezolucije. Posle se one šalju diskriminatoru koji proizvodi verovatnoće da su tako generisane slike realne. U ovom modelu gubitke delimo na dve vrste:

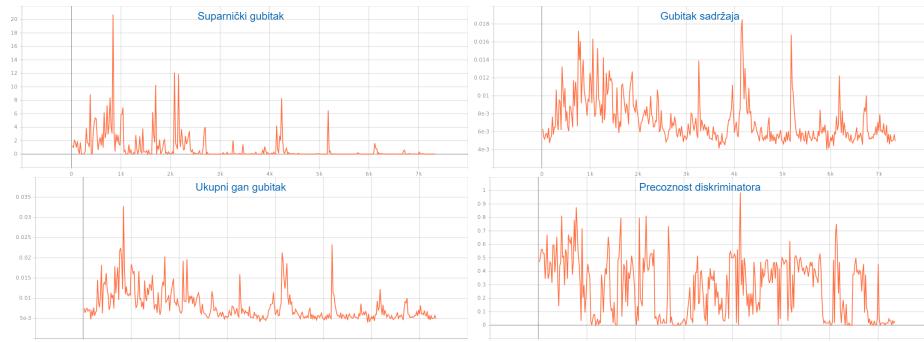
1. Gubitak sadržaja - Odnosi se na kvalitet generisane slike u odnosu na odgovarajuću realnu sliku. Za ovo su korištene dve metode:
 - Srednja kvadratna greška (mse loss) - Održava malu razliku u vrednostima piksela generisane i relane slike (suma po svim pikselima kvadrata razlike u vrednostima odgovarajućih piksela). Ipak nedostatak je manjak sitnih detalja na generisanim slikama i preterana glatkoća.
 - Greška bazirana na izlaznim odlikama dobijenim iz VGG mreže (vgg loss) - Realna i generisana slika puste se kroz pretreniranu VGG19 mrežu, kojoj su neki slojevi odstranjeni, i potom se generisane mape odlika upoređuju srednjom kvadratnom greškom. Ovde smo koristili mape odlika sloja $\phi_{5,4}$ (za objašnjenje notacije pogledati srgan rad [1]). Ovaj gubitak omogućava bolje generisanje detalja, pošto je VGG mreža već naučena za klasifikaciju raznih slika i u dubljim slojevima prepoznaće određene osobine.
2. Suparnički gubitak - Opisan formulom 1

U ovom radu implementirane su dve različite konfiguracije SRGAN-a, na osnovu gubitka sadržaja. Prva konfiguracija koristi samo srednju kvaratnu grešku

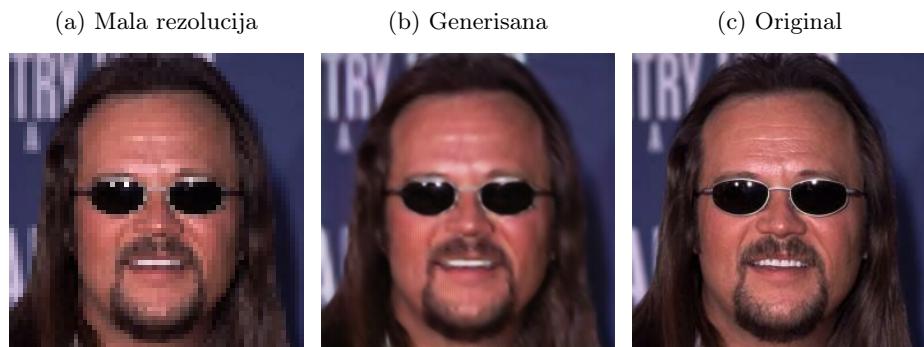
između generisanih i relanih slika (kao kod SRRResNet), dok druga koristi kombinovan gubitak (mse loss + vgg loss). U oba modela nazimenično su se trenirali diskriminator i generator, za svaku mini seriju podataka (koja se u ovom modelu sastoji iz 32 slike).

5.2.1 Konfiguracija sa MSE gubitkom

Kao što je prethodno navedeno, u ovoj konfiguraciji ulazne slike niske rezolucije prvo prolaze kroz generator, a potom diskriminator procenjuje validnost generisanih slika. Izlaz ovog modela se sastoji iz dva dela: prvi su izgenerisane slike, a drugi rezultat diskriminatora, i na osnovu njih se računa funkcija gubitka. Prvi izlaz upoređuje se sa realnim slikama visoke rezolucije, pomoću mse gubitka, a drugi sa jedinicama i preko binarne krosentropije formira suparnički gubitak (jer bi ako je diskriminator za sve generisane slike izbacio 1, to značilo da je greška generatora najmanja). Ovi gubici se težinski sabiraju, tako da je težina gubitka sadržaja 1, a suparničkog gubitka 0.001. Model je treniran 8 epoha. Gubitke u toku treniga možete videti na slici 6, a primer rezultata na slici 7



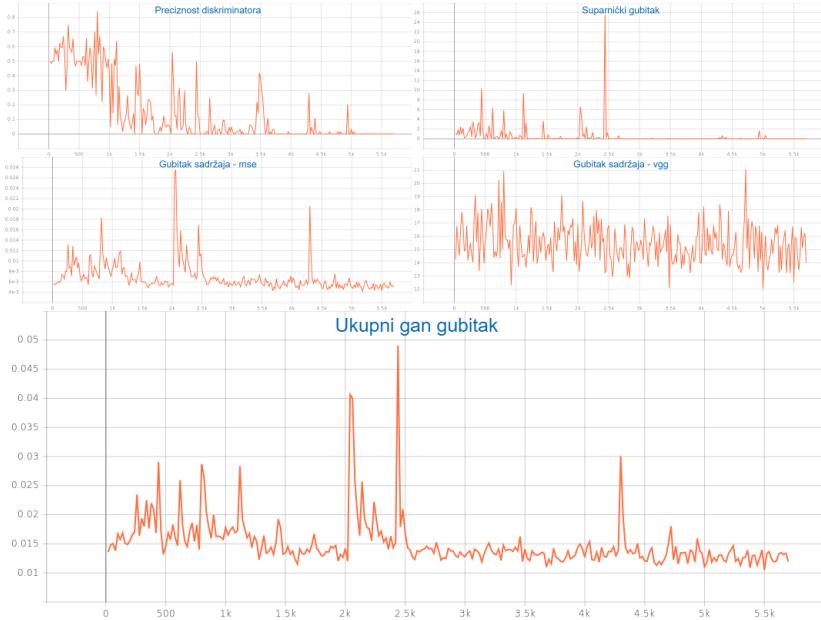
Slika 6: Grafici za SRGAN konfiguracije sa mse gubitkom



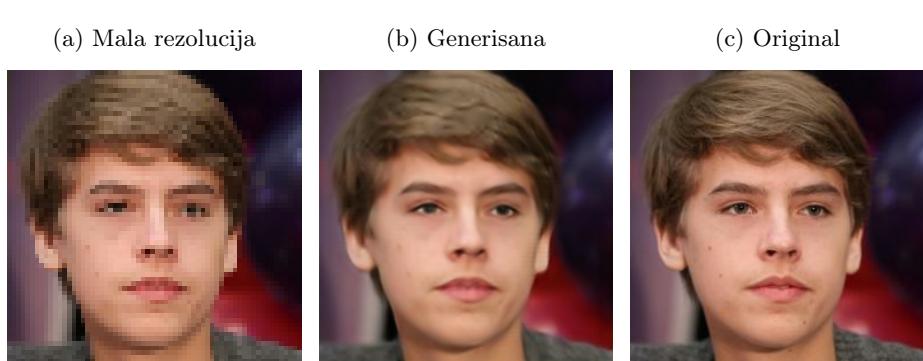
Slika 7: Prikaz slika za SRGAN - mse model

5.2.2 Konfiguracija sa kombinovanim gubitkom

Konfiguracija ovog modela slična je prethodnom uz nekoliko razlika. Pored prethodna dva gubitka dodat je i treći vgg gubitak. Model za izlaz šalje i mapu odlika generisanih slika i poredi ih sa mapom odlika stvarnih slika, mse gubitkom. Gubici se takođe težinski sabiraju i ovde je težina gubitka sadržaja 1 (mse) i 0.0005 (vgg), dok je težina suparničkog dubitka 0.001. Zbog korišćenja i VGG mreže trening ide još sporije, pa je zbog ograničenih resursa model treniran samo 6 epoha (sa većim brojem bi se verovatno dobili bolji rezultati). Gubitke u toku trenira možete videti na slici 8, a primer rezultata na slici 9



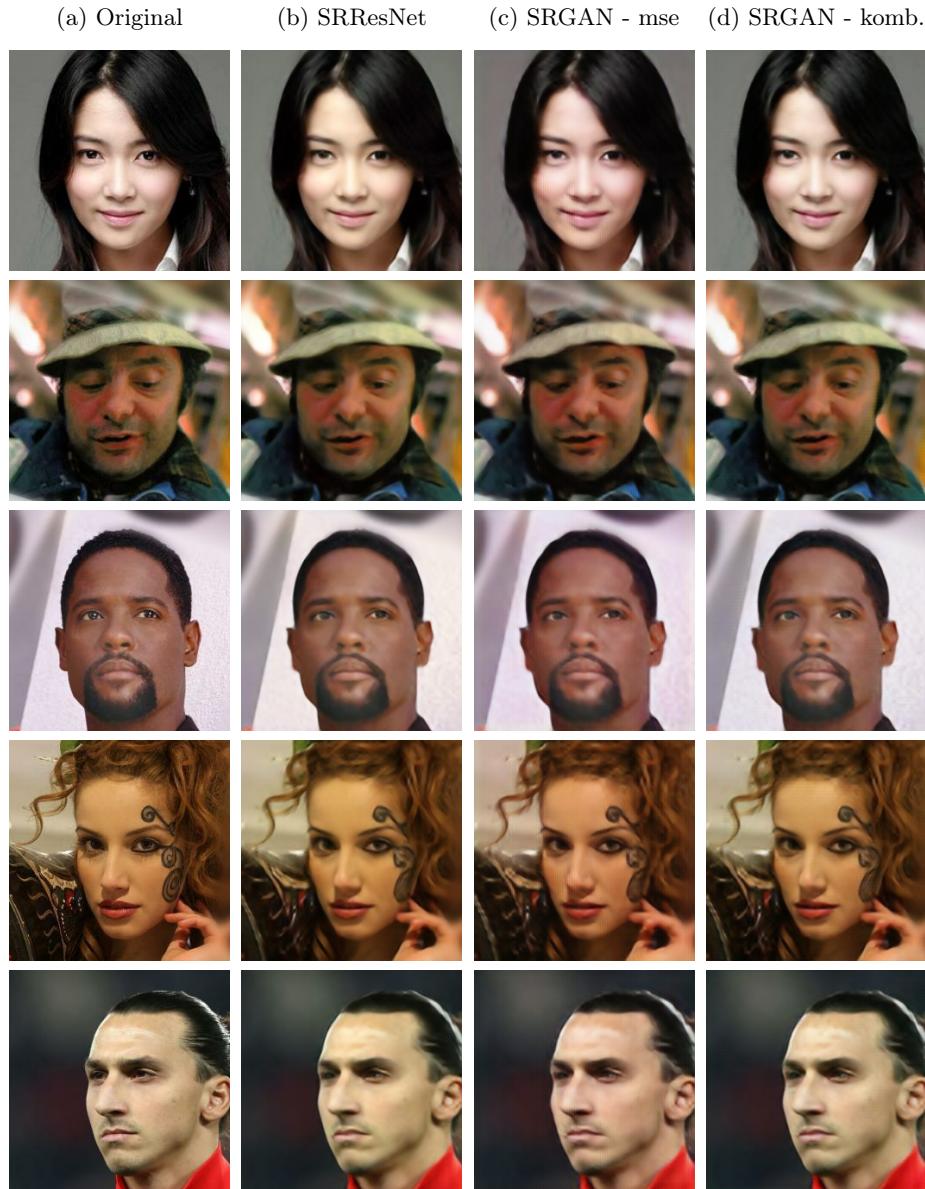
Slika 8: Grafici za SRGAN konfiguracije sa kombinovanim gubitkom



Slika 9: Prikaz slika za SRGAN - kombinovani model

6 Rezultati

Primeri generisanih slika pomoću različitih konfiguracija isključivo iz test skupa.



Slika 10: Prikaz prve test slike

PSNR i SSIM su metrike koje se često koriste za procenu kvaliteta slike npr.

	SRResNet	SRGAN - mse	SRGAN - komb
SSIM	0.8587	0.7901	0.8472
PSNR [dB]	22.5923	21.7244	22.7253

Tablica 1: Prikaz metrika za generisane slike

pri kompresiji. PSNR (Peak signal-to-noise ratio) se izražava u decibelima i veći rezultat znači bolju rekonstrukciju. SSIM (Structural similarity index) se često koristi kao metrika u video kompresiji i rezultati su mali pozitivni brojevi, gde je manji rezultat bolji. Mora se napomenuti i da ove metrike nisu najpouzadije merilo kvaliteta i da dobri rezultati ne moraju značiti vizuelno bolje slike. Rezultate na test skupu možete videti u tabeli 1

7 Zaključak

U ovom radu prikazan je SRGAN model i ispitani su rezultati neki od konfiguracija iz rada [1]. Pored toga isprobana je SRGAN model sa kombinovanim gubitkom koji je prikazao zadovoljavajuće rezultate. Može se zaključiti da i sam generator (SRResNet) daje pristojne rezultate. SRGAN sa mse gubitkom je dao rezultate malo slabijeg kvaliteta, dok SRGAN sa kombinovanim gubitkom izgenerisao slike kvaliteta približno korišćenju samo generatora. Uzrok toga je dosta teško treniranje SRGAN modela. Za to postoje nekoliko razloga. Prvi je samo trajanje treniranja i što ne postoji jasan trenutak kada treba da se stane. Takođe dešavalo se da u toku treninga se kvalitet rezultata značajno menja i generalno nije moralno da važi pravilo da se slike stalno poboljšavaju za razliku od SRResNet-a gde se minimizuje samo jedan gubitak. Pored toga bilo je teško uskladiti generator i diskriminator tako da njihova “borba” bude ujednačena. Uglavnom se dešavalo da diskriminatoru treba više vremena da se prilagodi promenama generisanih slika, pa je generator bio taj koji je najčešće nadjačavao protivnika. Ipak može se primetiti da su svi modeli naučili da prepoznaju delove lica, koji se na slikama niske rezolucije nekad nisu čak mogli ni uočiti.

Literatura

- [1] Christian Ledig, Lucas Theis, Ferenc Huszar, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, Wenzhe Shi.
Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network.
arXiv:1609.04802
- [2] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, Yoshua Bengio.
Generative Adversarial Networks.
arXiv:1406.2661