

Generisanje slika pomoću varijacionog autoenkodera

Igor Paunović

September 2023

1 Uvod

Varijacioni Autoenkoderi (VAE - Variational Autoencoder) su klasa generativnih modela za mašinsko učenje koji se odlikuju sposobnošću za hvatanje složenih obrazaca i generisanje novih uzoraka podataka. VAE-ovi su dizajnirani da nauče kompaktne i smislene reprezentacije visokodimenzionalih podataka, što ih čini korisnim alatima za zadatke kao što su kompresija podataka, generisanje slika i sinteza podataka. Ovi modeli se sastoje od enkodera koji mapira ulazne podatke u nižedimenzioni latentni prostor i dekodera koji rekonstruše uzorce podataka iz ovog latentnog prostora. VAE-ovi su posebno poznati po sposobnosti da generišu nove i raznolike uzorce podataka putem uzorkovanja iz naučenog latentnog prostora, što omogućava primene u generisanju slika, detekciji anomalija i mnogim drugim oblastima.

Cilj ovog rada je da demonstrira kako VAE može da primi set slika, nauči njihove fundamentalne osobine, i dozvoli nam da manipulišemo tim osobinama, i generišemo nove slike koje zadržavaju suštinu originalne slike, uz suptile i kontrolisane promene.

2 Podaci

Korišćen je skup podataka (celebA). Sadrži preko 202.599 slika ljudskih lica. Slike su u RGB formatu, dimenzije 178x218.



Figure 1: Izvorni skup slika

Pretprocesiranje

Korišćena je biblioteka cv2 za detekciju lica. Slike na kojima su lica lepo detektovana su prvo bitno kropovane na dimenziju 128x128, a potom su skalirane na dimenziju 64x64. Nakon toga, svaki piksel je podeljen sa 255.

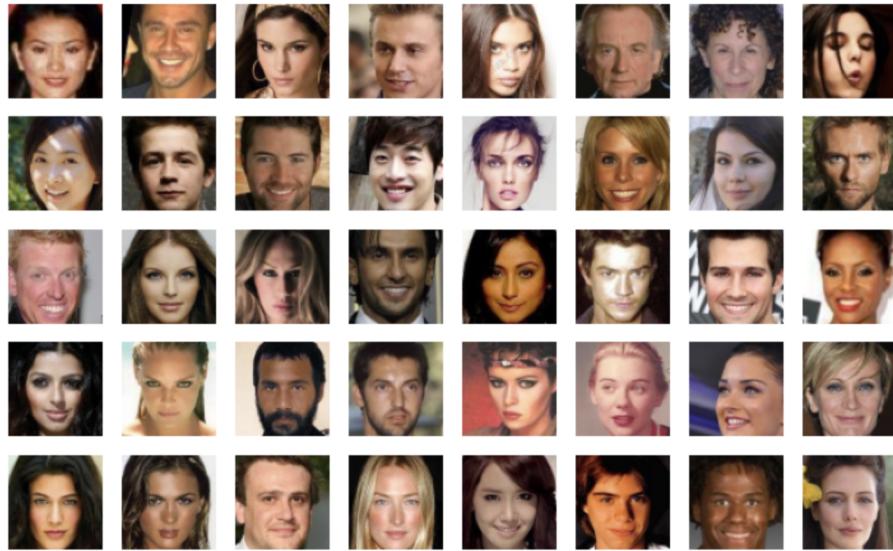


Figure 2: Pretprocesiran skup slika

3 Arhitektura

Arhitektura korišćenog varijacionalnog autoenkodera inspirisana je arhitekturom predstavljenom u radu [1].

Layer	Number of Filters	Kernel Size	Stride Size	Activation Function
Conv2D	32	4x4	2x2	ReLU
Conv2D	32	4x4	2x2	ReLU
Conv2D	32	4x4	2x2	ReLU

Table 1: Enkoder

Layer	Number of Filters	Kernel Size	Stride Size	Activation Function
Dense	256	N/A	N/A	ReLU
Dense	256	N/A	N/A	ReLU
Conv2DT	32	4x4	2x2	ReLU
Conv2DT	32	4x4	2x2	ReLU
Conv2DT	32	4x4	2x2	Sigmoid

Table 2: Dekoder

4 Treniranje

Funkcija greške

Za grešku rekonstrukcije korišćena je srednje kvadratna greška (MSE - Mean Squared Error) između piksela originalne i generisane slike. Za parametar regulizacije koristi se KL divergencija, koja meri "udaljenost" raspodele podataka u latentnom prostoru od željene raspodele (Uglavnom standardne normalne).

$$\mathcal{L}_{VAE}(x) = \underbrace{\frac{1}{2} \sum_{i=1}^N (x_i - \hat{x}_i)^2}_{MSE} + \beta \underbrace{\sum_{i=1}^N \left(\frac{1}{2} (\sigma_i^2 + \mu_i^2 - \log(\sigma_i^2) - 1) \right)}_{KL} \quad (1)$$

Korišćen je optimizator Adam sa brzinom učenja 0.0005. Batch veličina je 64, a dimenzija latentnog prostora 128. Parametar β je 1.0

Treniran je još jedan model, čija je izvorna arhitektura malo izmenjena, te umesto upsampling sloja i obične konvolucije koristi strided konvoluciju, što u ovom slučaju ne pravi veliku razliku. Od parametara se samo razlikuje dimenzija latentnog prostora koja je 100 i broj epoha treniranja, koji je 10. Reč je o PVAE (Plain Varational Autoencoder) modelu, više o njemu možete pročitati u radu [2].

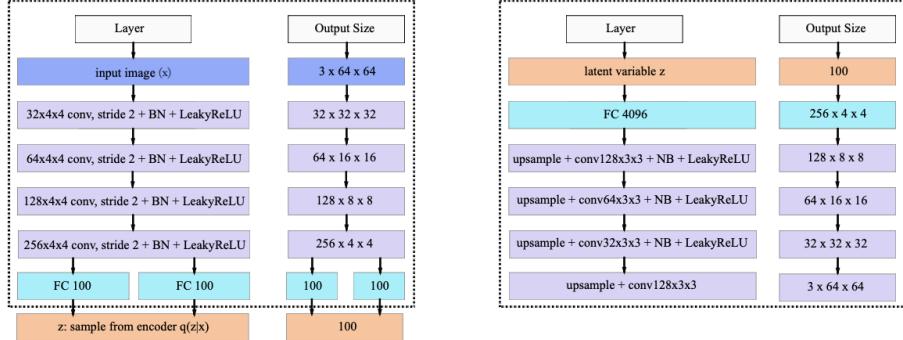


Figure 3: Arhitektura korišćena u radu [2]

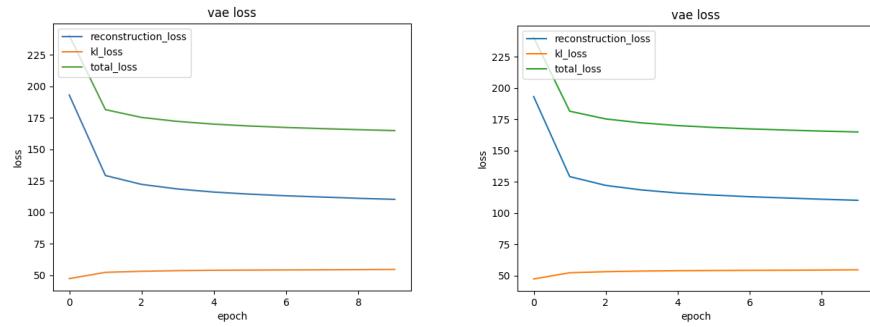


Figure 4: Promena funkcija greške po epohi za prvi i drugi model redom

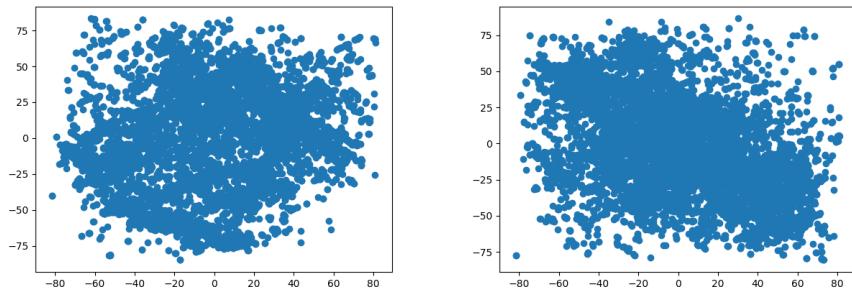


Figure 5: latentni prostor u 2 dimenzijskih reprezentacija za prvi i drugi model redom

Oba modela su se zaustavila na jako sličnim vrednostima MSE i KL divergencije.

Prvi model: loss: 165.0184 - reconstruction loss: 110.4022 - kl loss: 54.6771

Drugi model: loss: 164.9289 - reconstruction loss: 110.1697 - kl loss: 54.6052

5 Rezultati

Prvi model



Figure 6: Rekonstrukcija

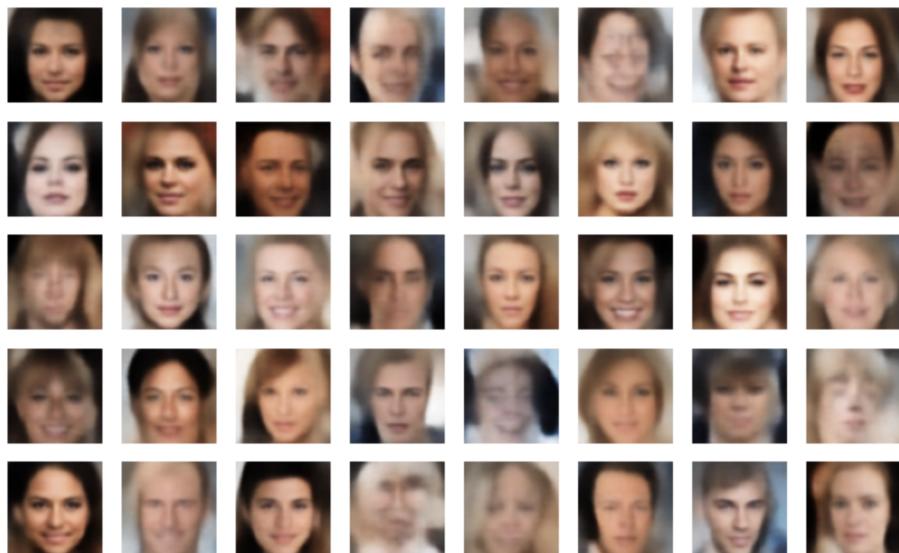


Figure 7: Generisanje

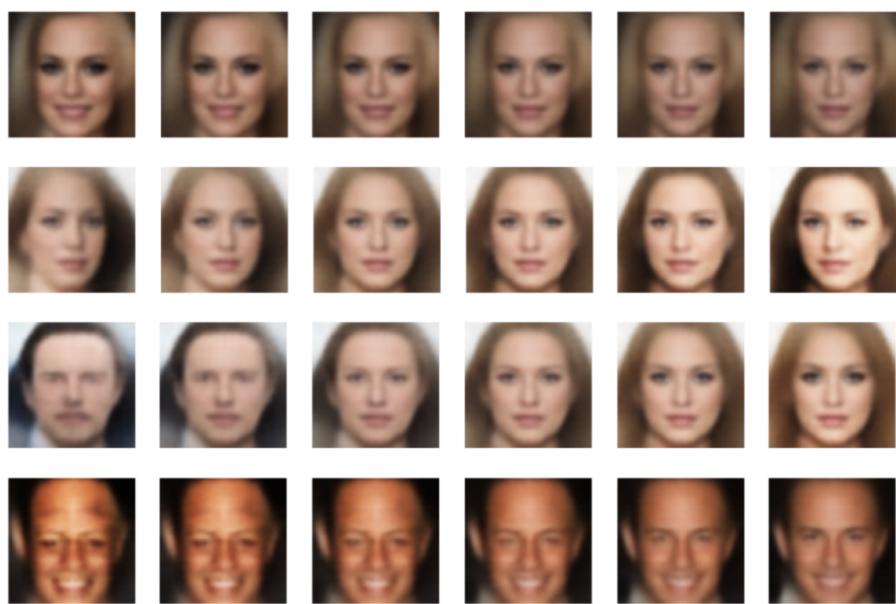


Figure 8: Kontrolisane promene

Drugi model

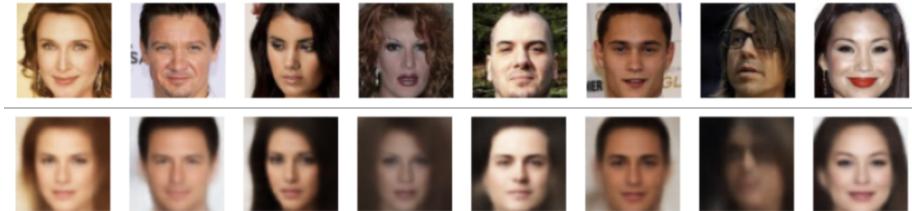


Figure 9: Rekonstrukcija

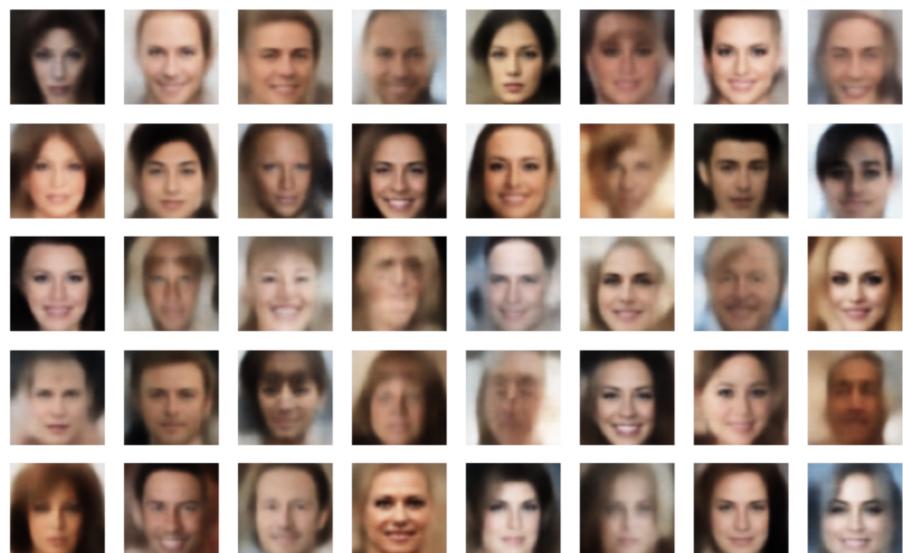


Figure 10: Generisanje

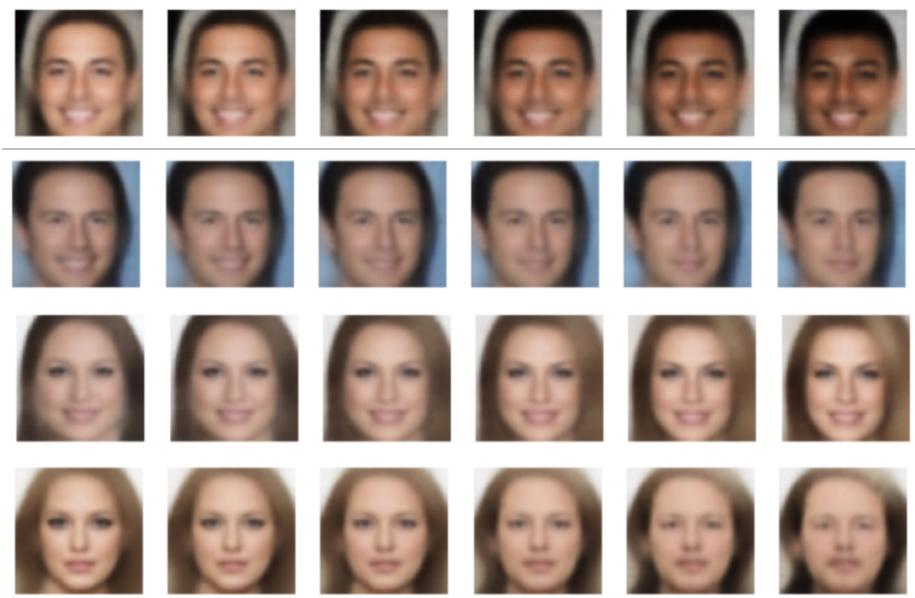


Figure 11: Kontrolisane promene

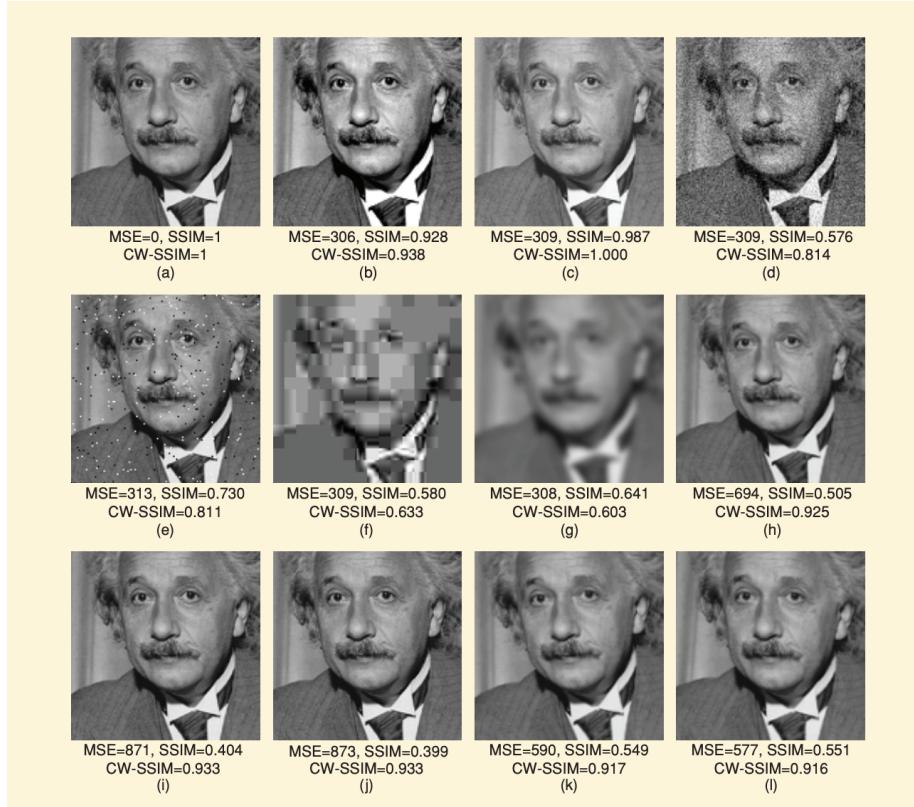
6 Problem

KL divergencija ne opada

Tokom treniranja primećeno je tek blago opadanje KL diverencije na poprilično skromnim arhitekturama, nevidljivo na grafiku. Na svim boljim arhitekturama na kojima je model treniran, KL neprekidno raste, i stabilizuje se nakon nekog trenutka. Zato se intezivnije treniranje čini kao neprestani kompromis. Pretpostavka je da bi, makar u nekom delu treninga, KL divergencija treba da opada.

MSE nije idealna funkcija greške rekonstrukcije

Dobra svojstva MSE jesu da se lako i efikasno računa i da je pogodna za analize i upoređivanja. Ipak, manja vrednost MSE ne znači nužno da je slika sličnija po parametrima ljudskog oka.



Na slici je prikazano slka uz transformacije slike, poput dodavanje šumova, zamagljivanje, kompresije, itd. Očigledno MSE ne uspeva da prati stvarnu sličnost među slikama za razliku o metrike SSIM (Structural Similarity Index Measure). Slika je preuzeta iz rada [3] u kome se može naći opširnija kritika na račun MSE.

7 Zaključak

Po malobrojnom broju naučnih radova na temu generisanja slike sa običnih varijacionim autoenkoderom, kritikama [3] na korišćenje MSE u procesu generisanja slika, i prikazanim rezultatima, možemo zaključiti da obični varijacioni autoenkoderi nisu dovoljni za kvalitetno generisanje slika, ali njihove varijante su jako moćne i vodeće u toj disciplini

Oba trenirana modela mogu rekonstrusati i generisati slike ljudskih lica, i uz promenu samo jednog broja, menja se svojstvo na slici sa kojim taj broj korespondira. Ove tačke su bile glavni cilj rada i ispunjene su.

Na ovom linku se nalazi repozitorijum sa kodom.

8 Literatura

- Miroslav Fil, Munib Mesinovic, Matthew Morris, Jonas Wildberger.
Beta-VAE Reproducibility: Challenges and Extensions
arXiv:2112.14278
pdf
- Xianxu Hou, Linlin Shen, Ke Sun, Guoping Qiu.
Deep Feature Consistent Variational Autoencoder
arXiv:1610.00291
pdf
- Z. Wang and A. C. Bovik, "Mean squared error: Love it or leave it? A new look at Signal Fidelity Measures," in IEEE Signal Processing Magazine, vol. 26, no. 1, pp. 98-117, Jan. 2009, doi: 10.1109/MSP.2008.930649.
pdf