



# Generisanje slika pomoću varijacionog autoenkodera

Igor Paunović  
Univerzitet u Beogradu, Matematički fakultet  
Septembar 2023

# Varijacioni autoenkoderi



Zamislamo da smo osoba koja je jako dobra u crtanju ljudskih lica. Inspektor nas zamoli da nacrtamo kriminalca koga nikada nismo videli, ali dobijemo informacije da je kriminalac ćelav, ima svetle plave oči, veliki nos, oštre crte lica i tamnu crnu bradu. I mi ga verno nacrtamo. Jedno je sigurno, ili smo jako dobar slikar ili smo VARIJACIONI AUTOENKODER.

Da bi varijacioni autoenkoder generisao sliku lica našeg kriminalca, njemu su potrebni brojevi koji kodiraju svaki od atributa lica - veliki broj za veličinu nosa, oštrinu lica i boju brade, a mali, ili negativan broj za dužinu kose i svetle plave oči. On je već trenirao na velikom skupu lica, i zna šta je ljudsko lice, pa od ovih brojki zavisi kakvo će generisano lice biti.



## Varijacioni autoenkoderi

- Generativni model za mašinsko učenje
- Dobar u učenju složenih obrazaca i generisanju novih podataka
- Sastoj se iz dva modela - Enkodera i Dekodera
- Enkoder - Preslikava sliku u nižedimenzioni prostor i uči raspodelu podataka u nižedimenzionom prostoru
- Dekoder - Preslikava tačku iz nižedimenzionog prostora u sliku
- Male promene nižedimenzione reprezentacije se oslikavaju malim promenama generisane slike



## Cilj rada

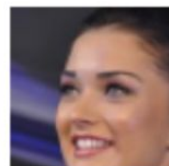
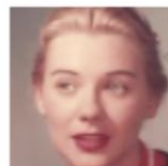
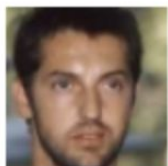
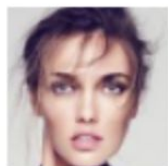
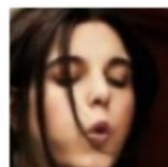
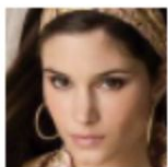
Demonstracija sposobnosi VAE da:

- Rekonstruiše datu sliku
- Generiše novu sliku
- Malom promenom latentne reprezentacije slike dobijemo suptilnu promenu na novogenerisanoj slici




## Skup podataka

- Korišćen je skup podataka **celebA**
- Sadrži 202.599 slika ljudskih lica u formatu 178x128x3
- Slike su kropovane i skalirane na format 64x64x3 tako da se sačuva najbitniji deo - lice
- Novodobijeni skup podataka sadrži 104.861 slika.



# Arhitektura



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



## Funkcija greške

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





# Hiperparametri

Optimizator: Adam

Brzina učenja: 0.0005

Dimenzija niskodimenzionog (latentnog) prostora: 128

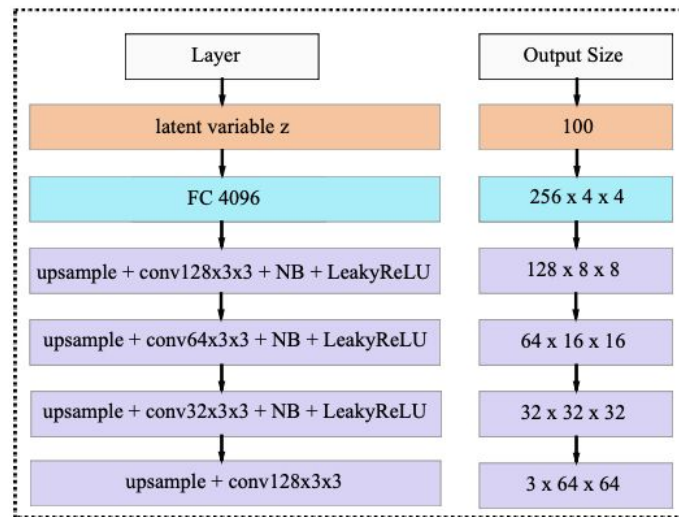
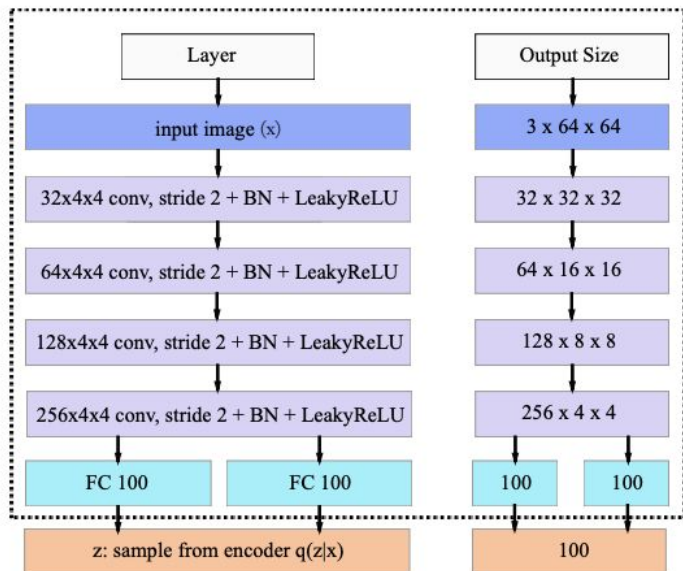
Batch\_size: 64

Epohe: 15

Beta (težina KL divergencije): 1.0

# Treniranje

Pored našeg, treniran je još jedan referentni model sledeće arhitekture, uz male izmene, treniran na 10 epoha i sa dimenzijom latentno prostora 100



# Rezultati treniranja

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

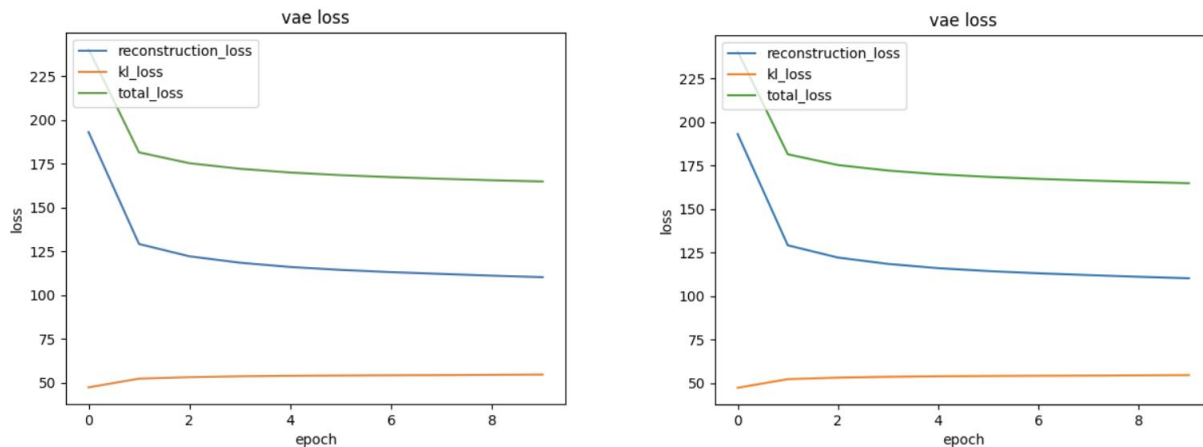


Figure 4: Promena funkcija greške po epohi za prvi i drugi model redom

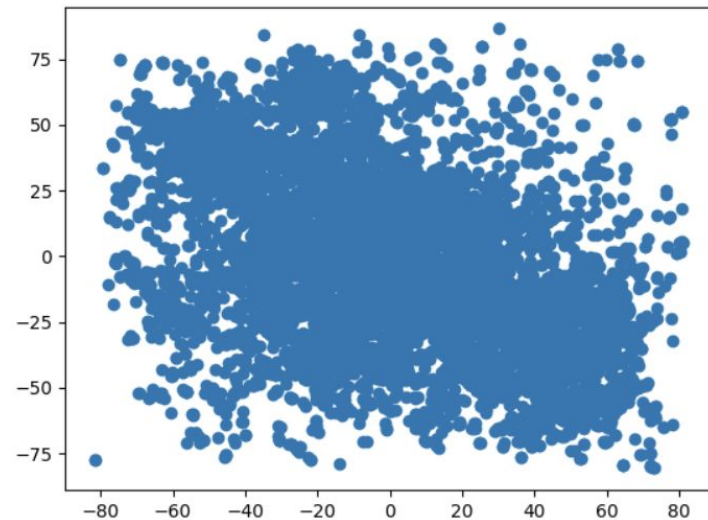
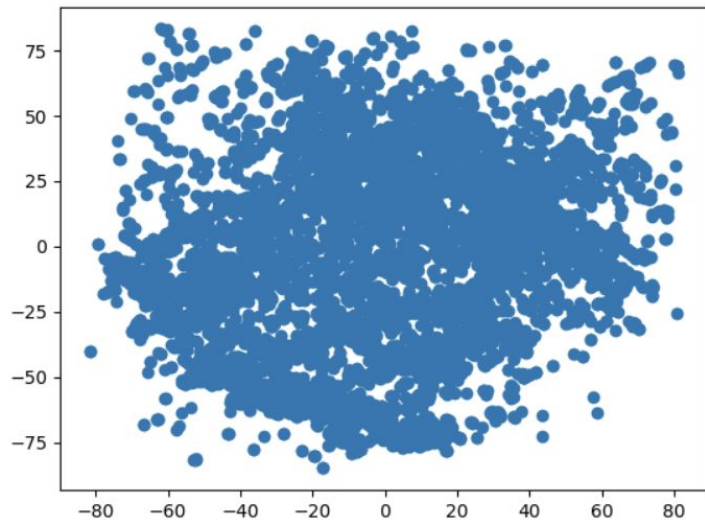


Figure 5: latentni prostor u 2 dimenzije za prvi i drugi model redom

# Rekonstrukcija



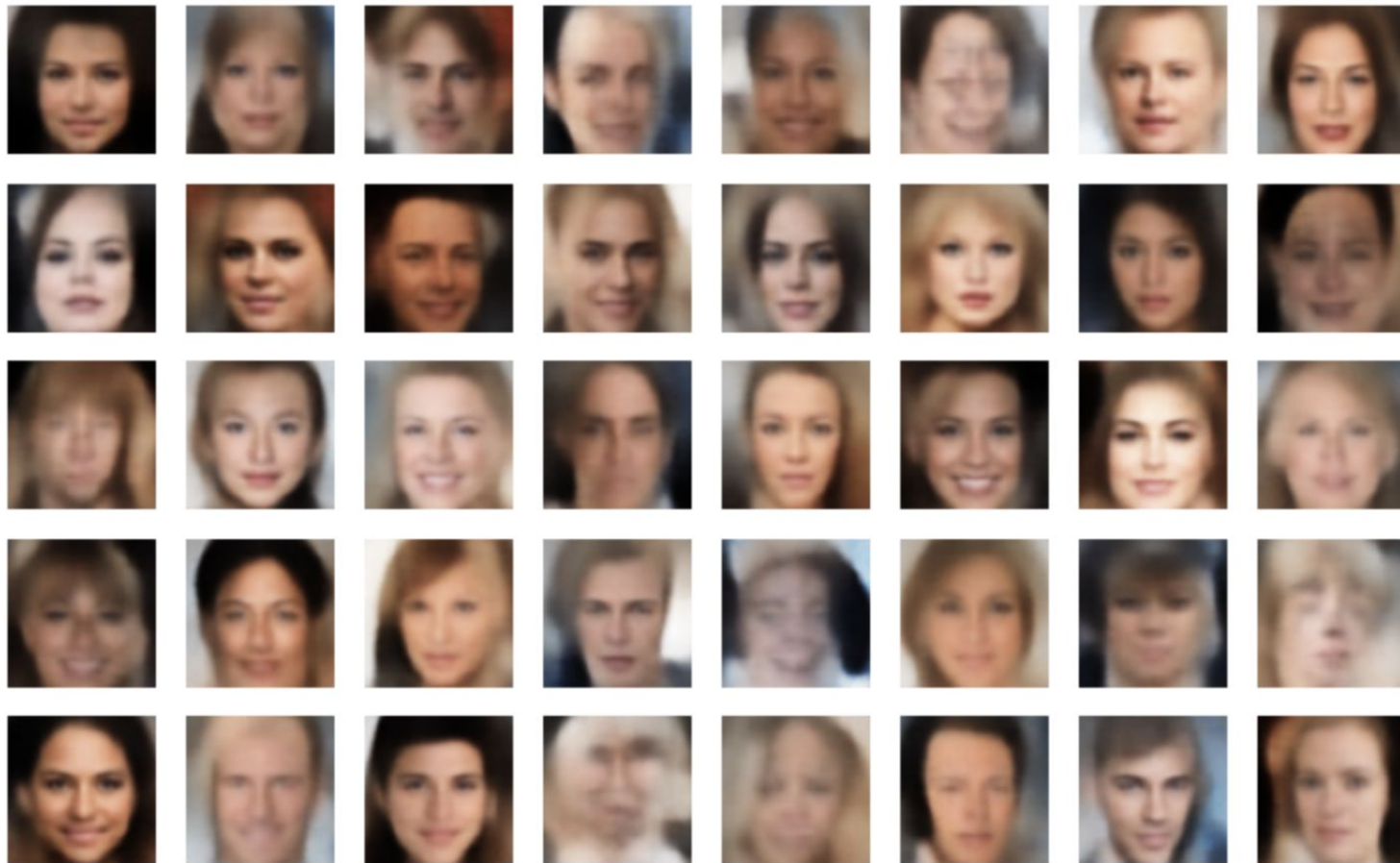
Prvi model



Ddrugi model

# Generisanje prvog modela

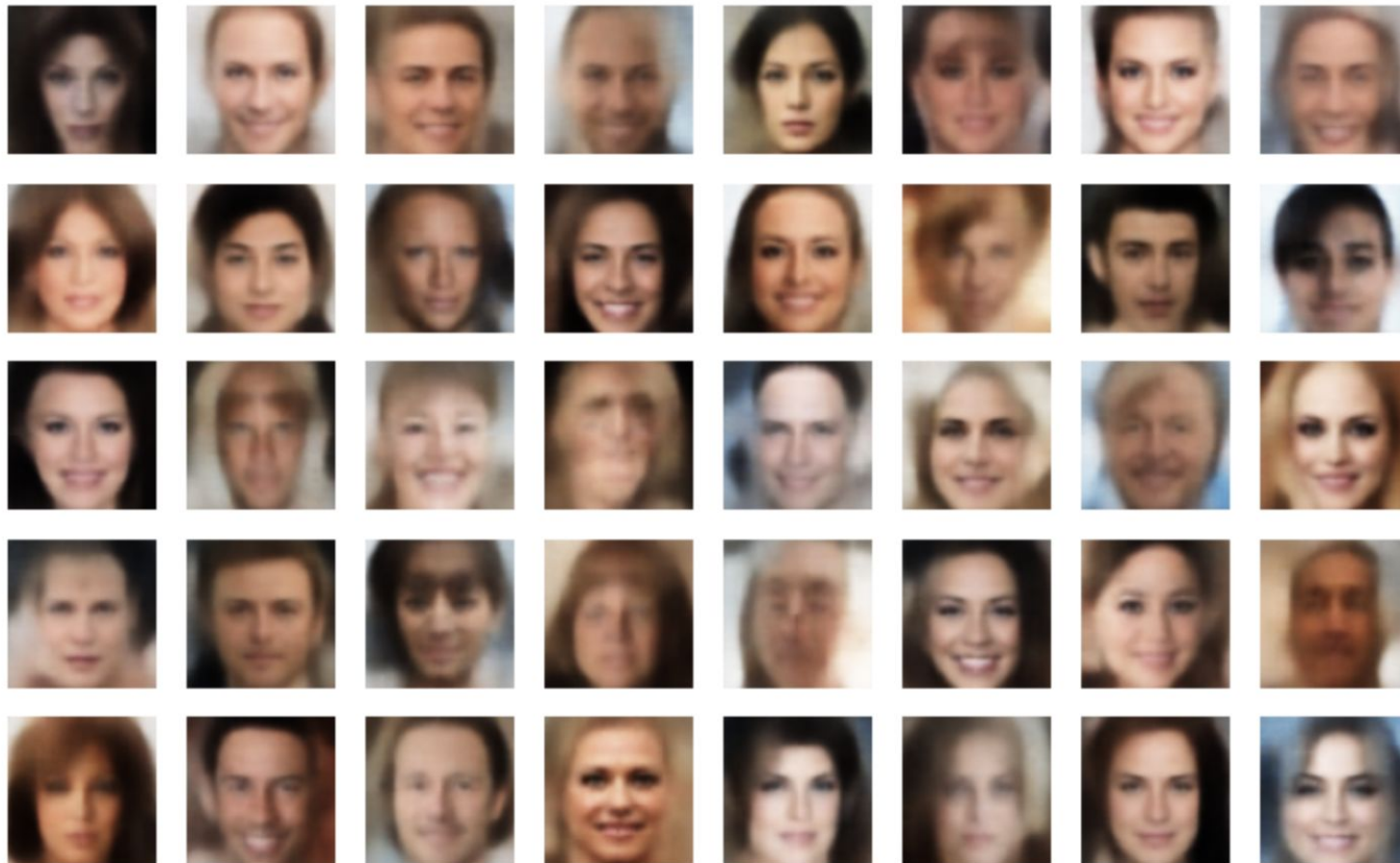
—



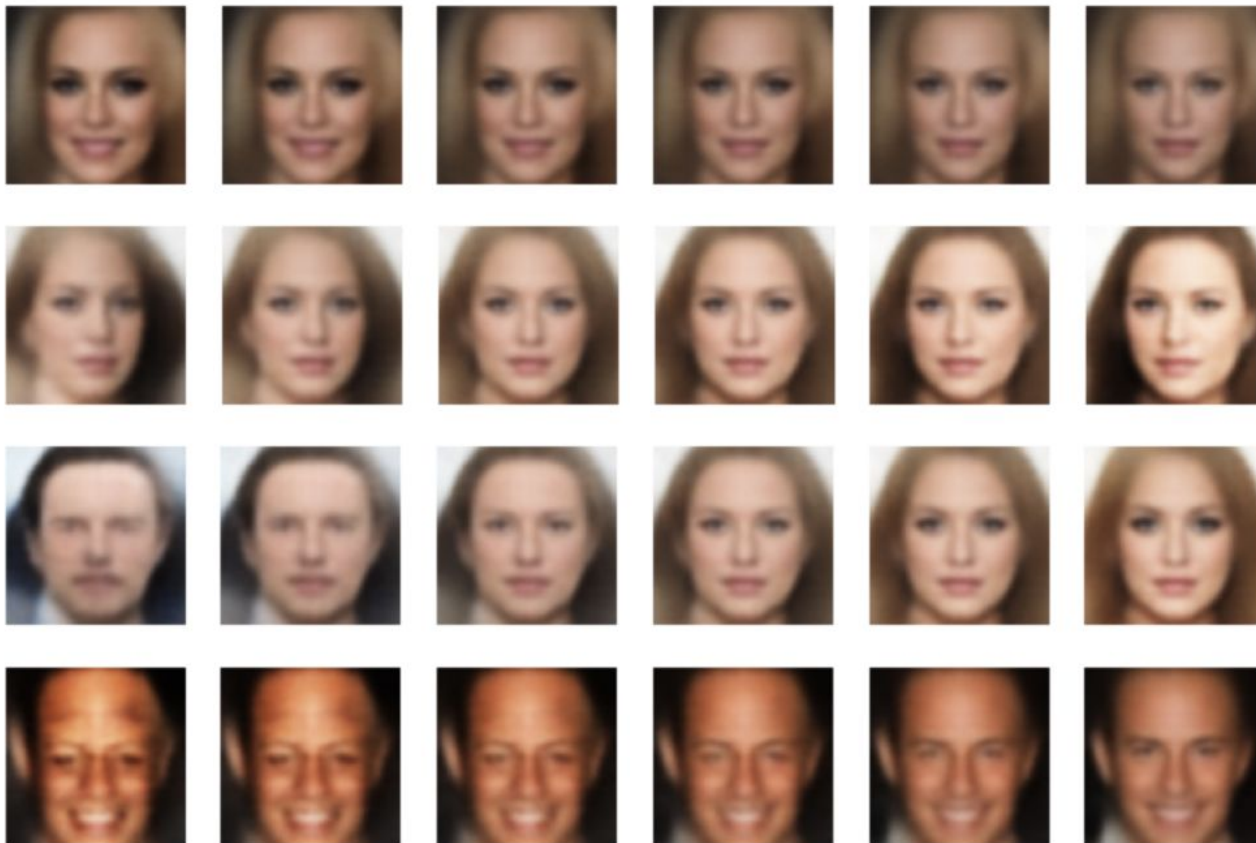


# Generisanje drugog modela

—



# Kontrolisana promena atributa prvog modela





# Kontrolisana promena atributa prvog modela





## Zaključak

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.

Obični varijacioni autoenkoderi nisu dovoljni za kvalitetno generisanje slika, ali njihove varijante su jako moćne i vodeće u toj disciplini.