

# Projekat iz predmeta masinsko ucenje - varijacioni autoenkoderi

Dragana Ninkovic 2023/3010

Decembar 2023

## 1 Opis koriscenog modela

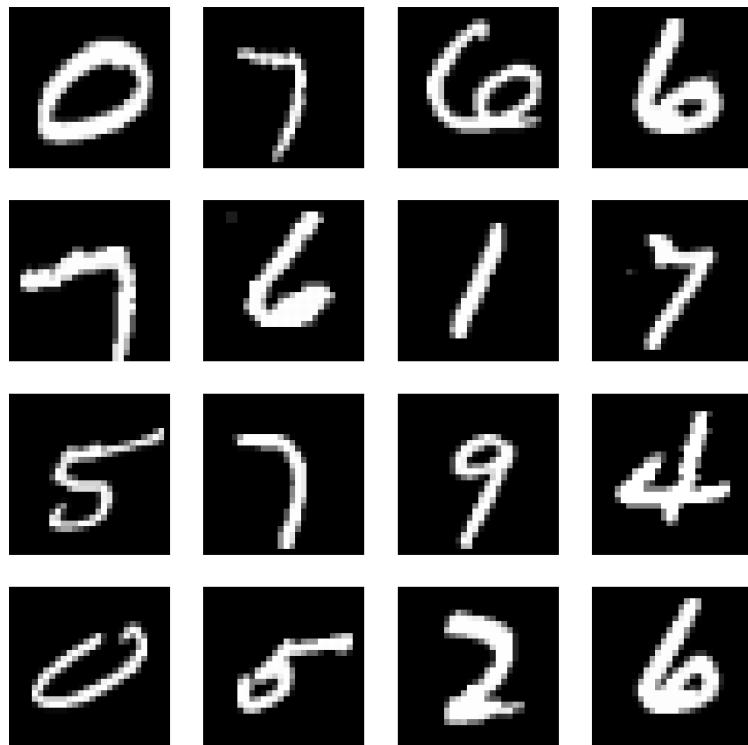
Tema ovog projekta su varijacioni autoenkoderi koji su prvi put spomenuti u radu "Auto-Encoding Variational Bayes" gde su prikazani rezultati rada ovog modela nad MNIST skupom podataka i Frey Face. U ovom radu koriscen je najobicniji perceptron za treniranje modela i prikazani su rezultati za vise dimenzija latentnog prostora. S obzirom da su konvolucione neuralne mreze mnogo podobnije za rad sa slikama u ovom projektu je koriscena konvolucionna neuralna mreza za enkoder i dekoder, u kombinaciji sa Adam optimizatorom za razliku od arhitekture koja je koriscena u radu. Batch size je uzet 128 (malo veci nego u originalnom radu gde je  $batch\_size$  100) a rezultati ce biti prikazani za velicine skrivenih slojeva 16, 100 i 200 u slucaju sivih slika. S obzirom da je za jednu epohu treniranja potrebno oko 3-5 minuta, model je treniran samo 30 epoha ali zahvaljujuci slozenosti modela cak i sa ovako malim brojem epoha daje dobre rezultate. Takodje, radi bolje vizuelizacije podataka za dimenziju latentnog prostora korisceno je  $N = 2$  za sive slike i bilo je dovoljno za dobijanje dobrih rezultata. U slucaju slika u boji, bila je potrebna znacajno veca dimenzija latentnog prostora pa on nije vizualizovan vec samo primeri random generisanih slika.

## 2 Korisceni skupovi podataka

U ovom projektu su pored MNIST skupa podataka korisceni FashionMNIST i SVHD(Street view house numbers) skupovi podataka koji se skidaju iz biblioteke torchvision.datasets.

### 2.1 MNIST

MNIST skup podataka je jedan od najosnovnijih i najlaksih skupova podataka za treniranje modela za problem kompjuterske vizije. Velicina trening skupa je 60 000 podataka a velicina testirajuceg skupa je 10 000 podataka. Sastoji se od sivih slika dimenzija 28x28 koje predstavljaju cifre napisane rukom od 0 do 9.



Slika 1 Primeri slika iz trening skupa za Fashion MNIST skup podataka

## 2.2 Fashion MNIST

Fashion MNIST je još jedan od cesto korišćenih skupova podataka za probleme u kompjuterskoj viziji. U njemu se nalaze sive slike dimenzija 28x28 odevnih predmeta - majca, pantalone, dzemper, haljina, kaput, sandale, kosulja, patike, torba i cizme. Veličina skupa je ista kao i prethodnog.



Slika 2 Primeri slika iz trening skupa za MNIST skup podataka

## 2.3 Fashion MNIST

## 2.4 SVHN (Street View House Numbers)

Za razliku od prethodnih skupova podataka, ovaj skup podataka je u boji, i sastoji se od 73257 slika u trening skupu i 26032 slika na test skupu. Slicno MNIST skupu podataka, sadrži cifre od 0 do 9 ali ovde su u pitanju prave slike cifara u boji iz stvarnog života poput isecaka cifara sa tablica/ brojeva zgrada/kuća u ulici i slično. Na slikama se cesto pored cifre nalaze još neke distrakcije što čini ovaj skup podataka znacajno težim za obucavanje od prethodna dva.



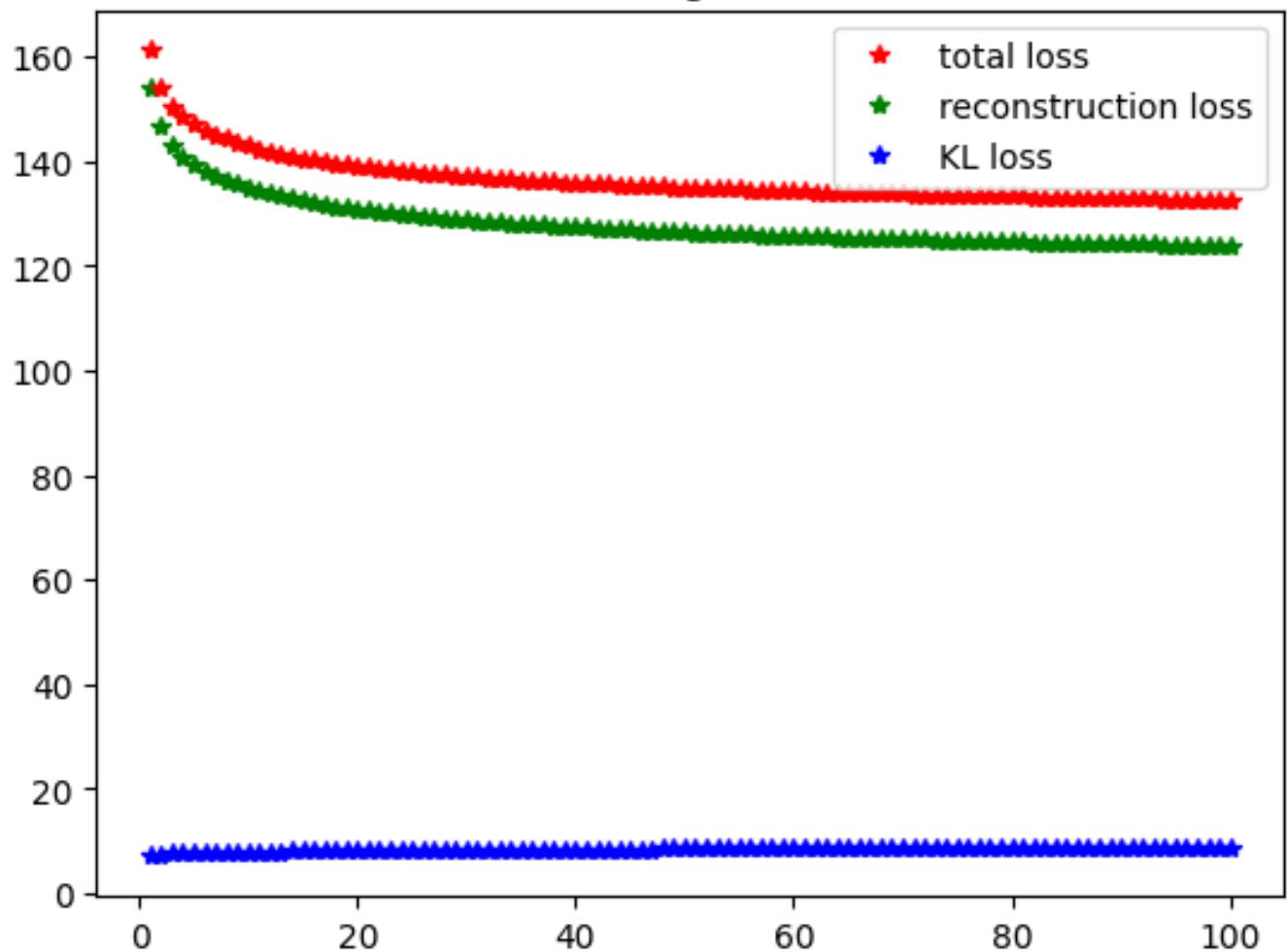
**Slika 3** Primeri slika iz trening skupa za SVHN skup podataka

### 3 Rezultati

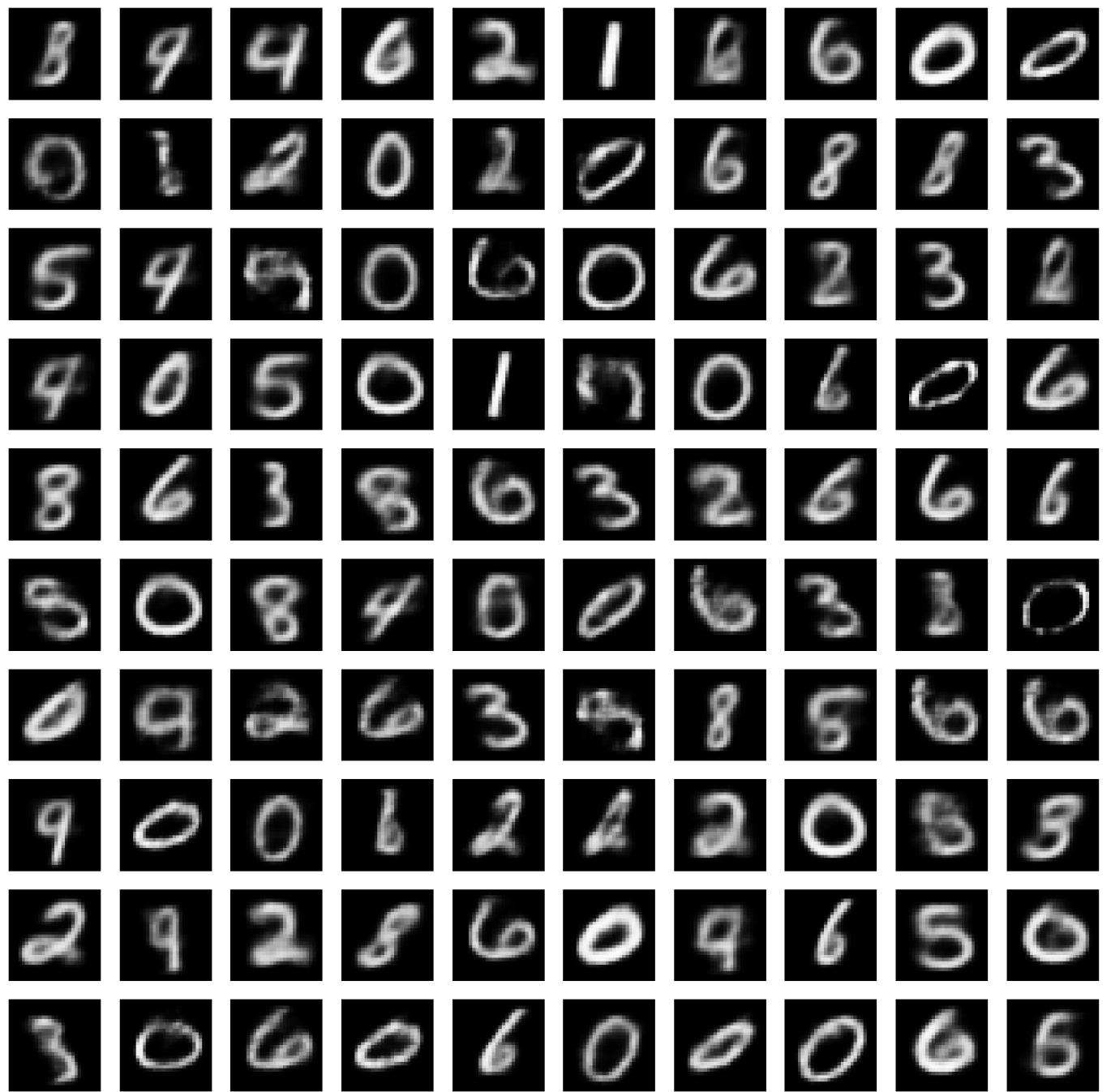
#### 3.1 Poredjenje sa rezultatima iz rada

U ovom poglavlju uzeti su svi parametri identicno parametrima iz rada za latentni prostor dimenzije 3 samo sto je umesto potpuno povezanih slojeva korisena konvolucionna mreza za enkoder i dekoder. U radu su za  $10^8$  iteracija dobili gresku izmedju 125 i 135, dok je nas model ostvario gresku oko 130 na trening skupu za  $6\ 000\ 000 < 10^7$  iteracija, sto su priblizno isti rezultati a manji broj iteracija se moze objasniti boljim performansama konvolucionih mreza od potpuno povezanih u slucaju slika.

### Training curve



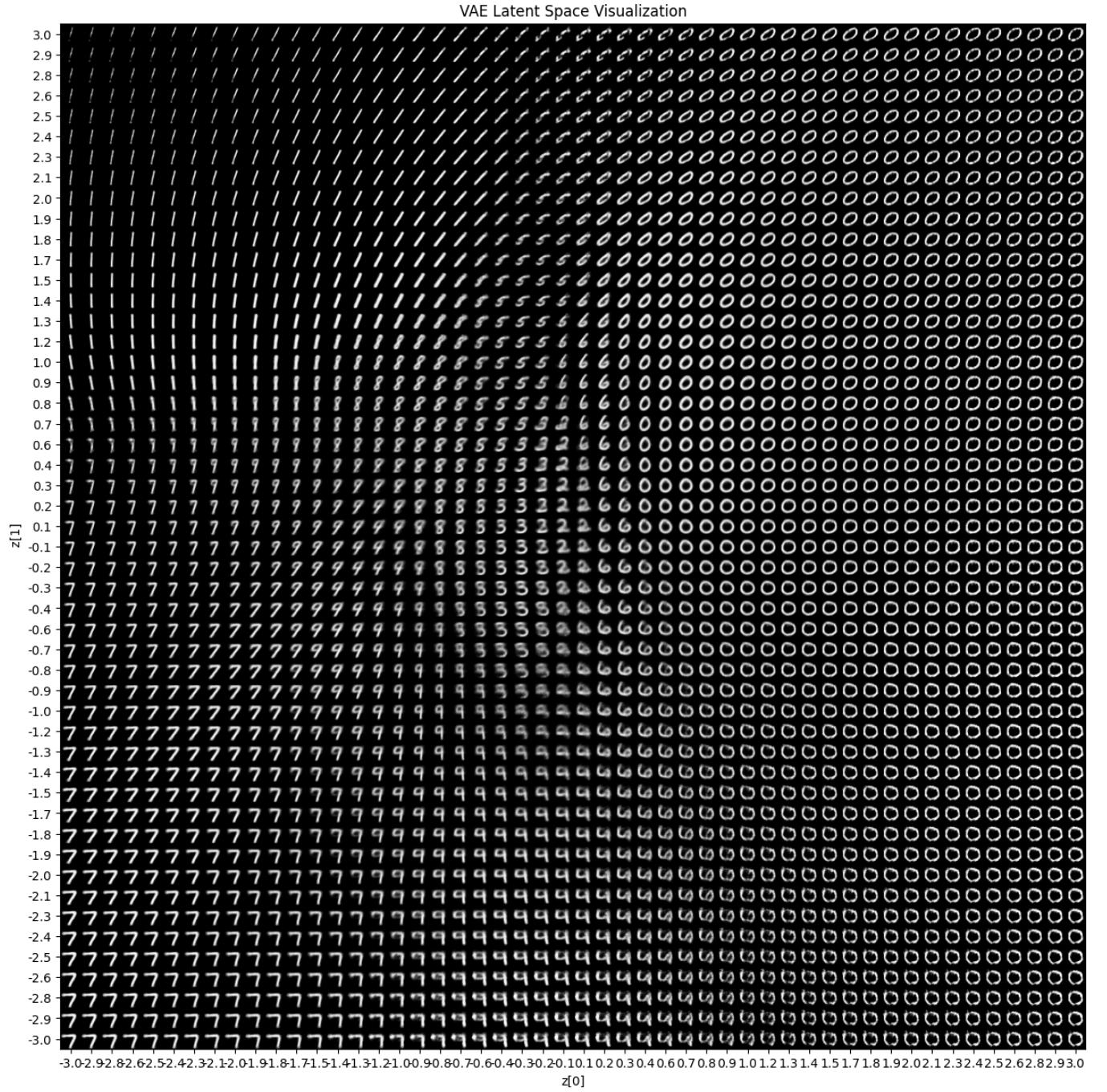
Slika 4 Prikaz treniranja



Slika 5 Prikaz generisanih slika od strane modela

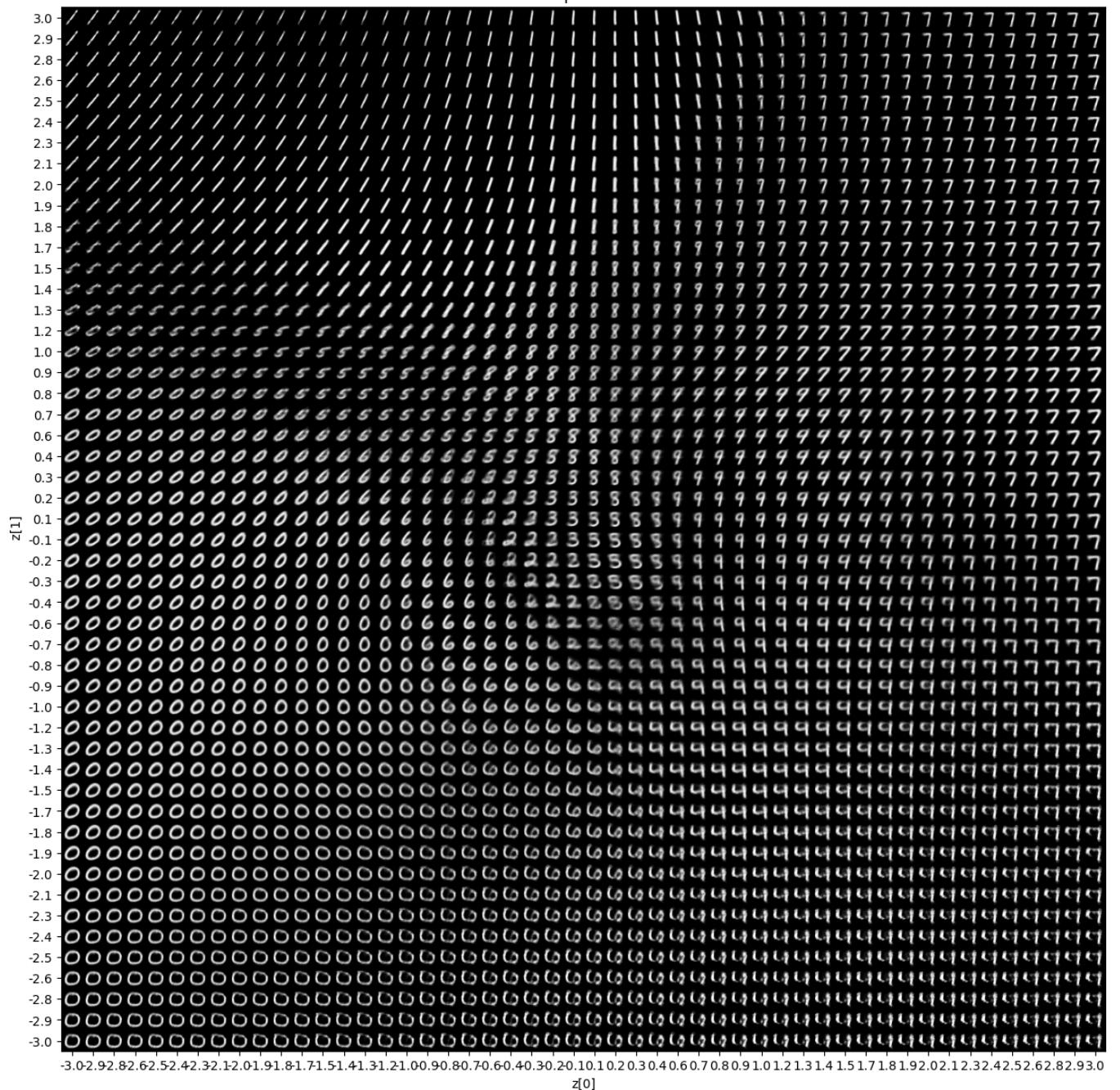
Rezultati na test skupu:  
Average total loss: 132.3857901529948  
Average reconstruction Loss: 123.628955859375  
Average KL Loss: 8.756834337361653

## 3.2 MNIST

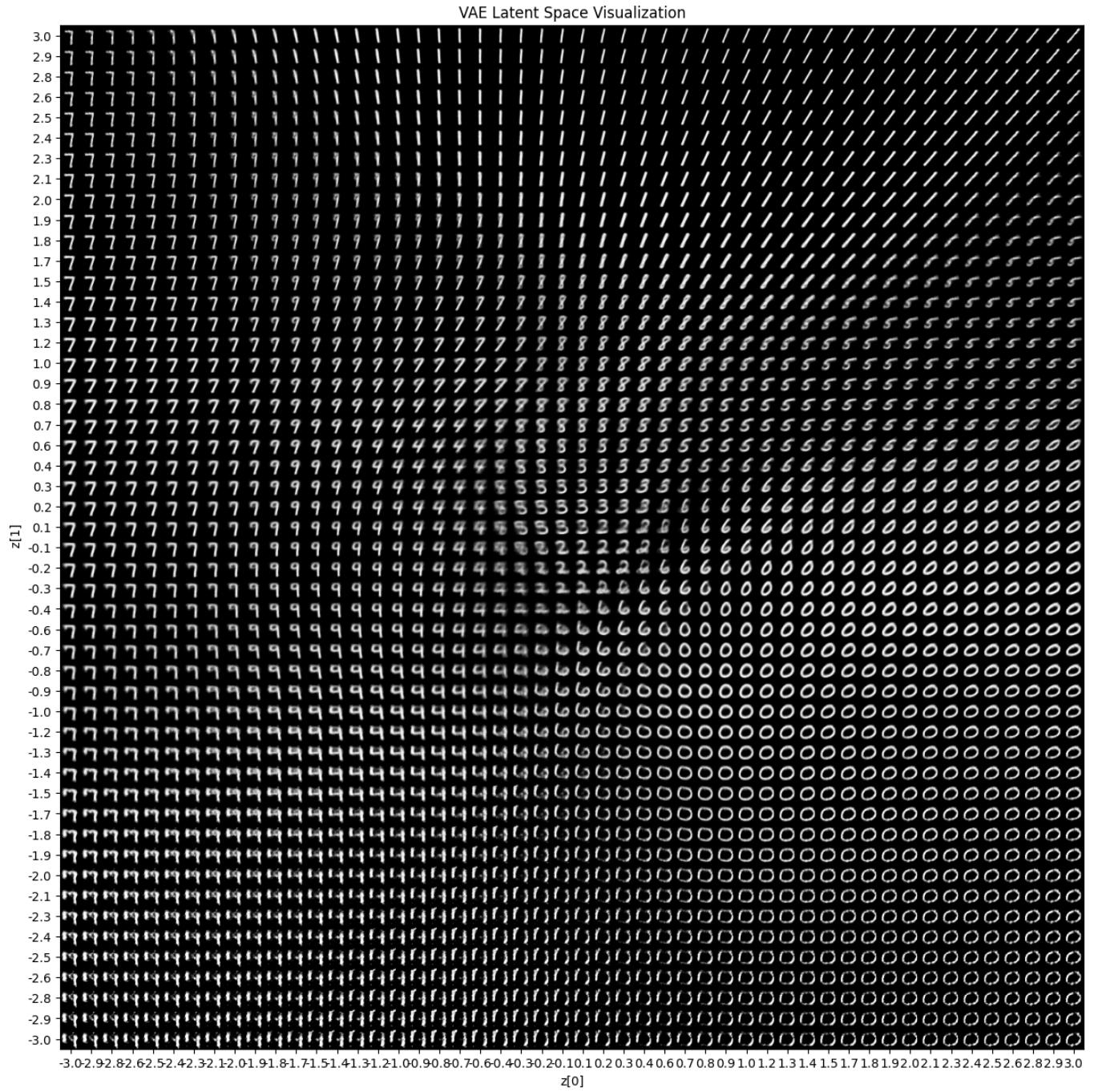


**Slika 6** Prikaz latentnog prostora za velicinu skrivenog sloja 16

VAE Latent Space Visualization

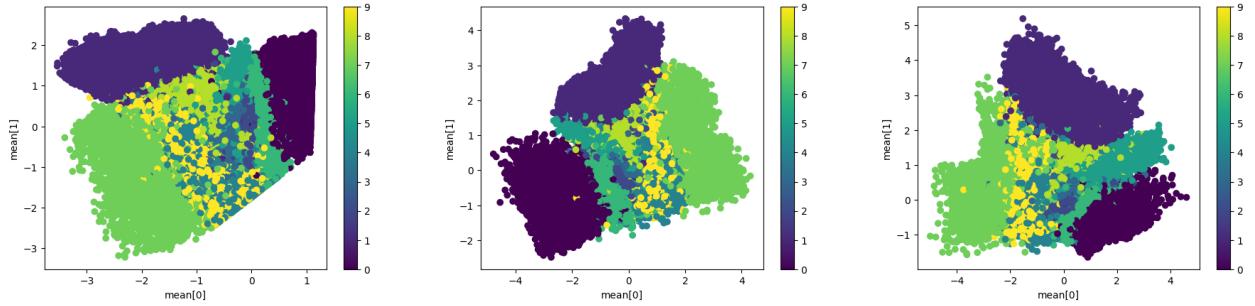


Slika 7 Prikaz latentnog prostora za velicinu skrivenog sloja 100



**Slika 8** Prikaz latentnog prostora za velicinu skrivenog sloja 200

Vidimo da za veće dimenzije skrivenog sloja dobijamo detaljnije generisane primere slika i realisticnije ali i da je razlika izmedju 16 i 100 ogrona, dok je izmedju 100 i 200 skoro pa neprimetna.



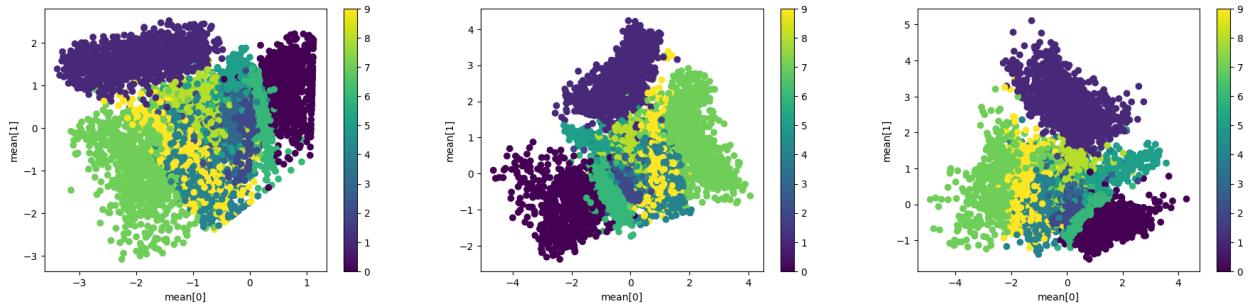
(a) velicina skrivenog sloja = 16

(b) velicina skrivenog sloja = 100

(c) velicina skrivenog sloja = 200

**Slika 9** Prikaz obucavajuceg skupa u latentnom prostoru

Vidimo da sa povecanjem dimenzija skrivenog sloja bolje odvajaju razlicite klase u latentnom prostoru.



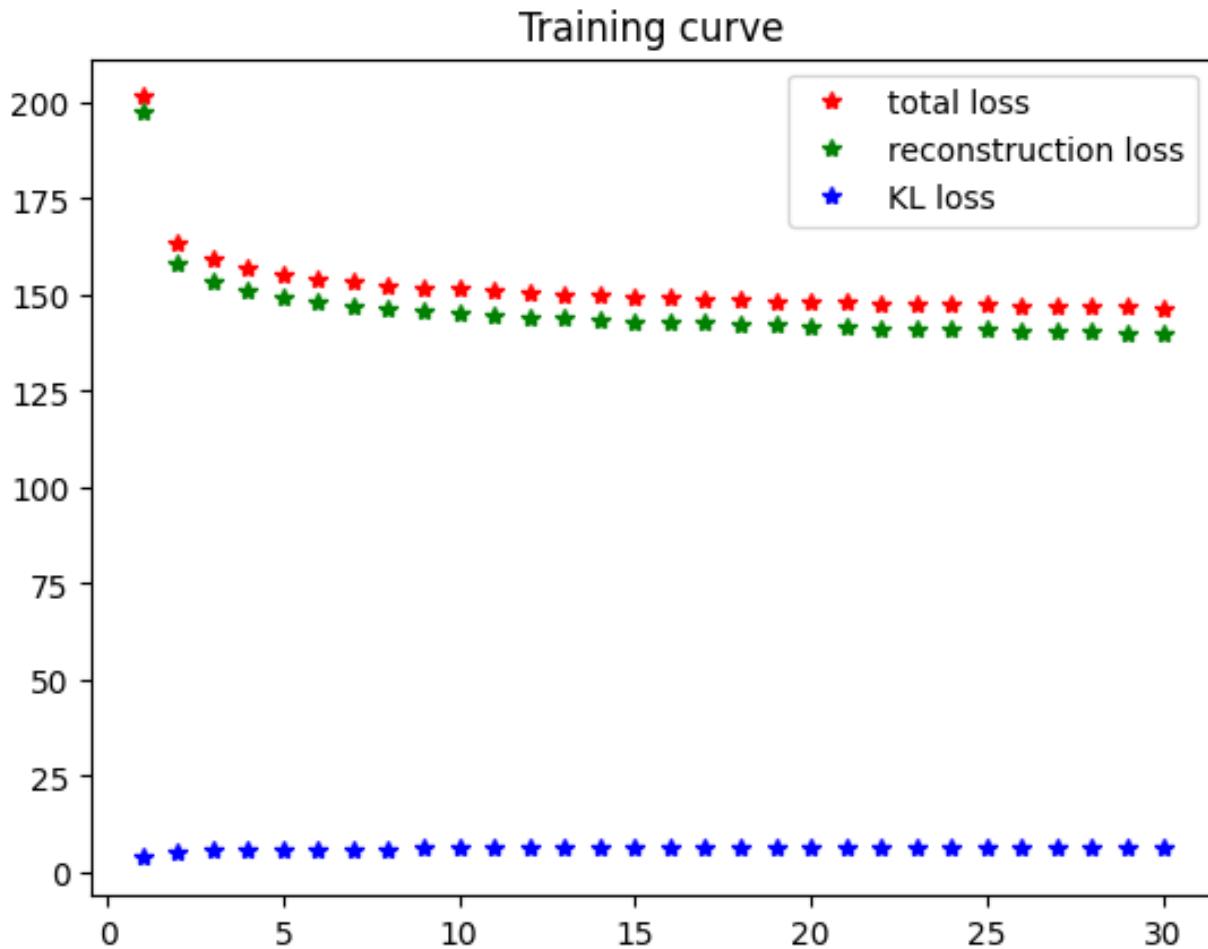
(a) velicina skrivenog sloja = 16

(b) velicina skrivenog sloja = 100

(c) velicina skrivenog sloja = 200

**Slika 10** Prikaz test skupa u latentnom prostoru

Vidimo da su klase slicno rasporedjene u latentnom prostoru za obucavajuci i test skup sto nam govori da nas model dobro generazlicuje nauceno.



**Slika 11** Prikaz treniranja za velicinu skrivenog sloja 100

Rezultati na test skupu:

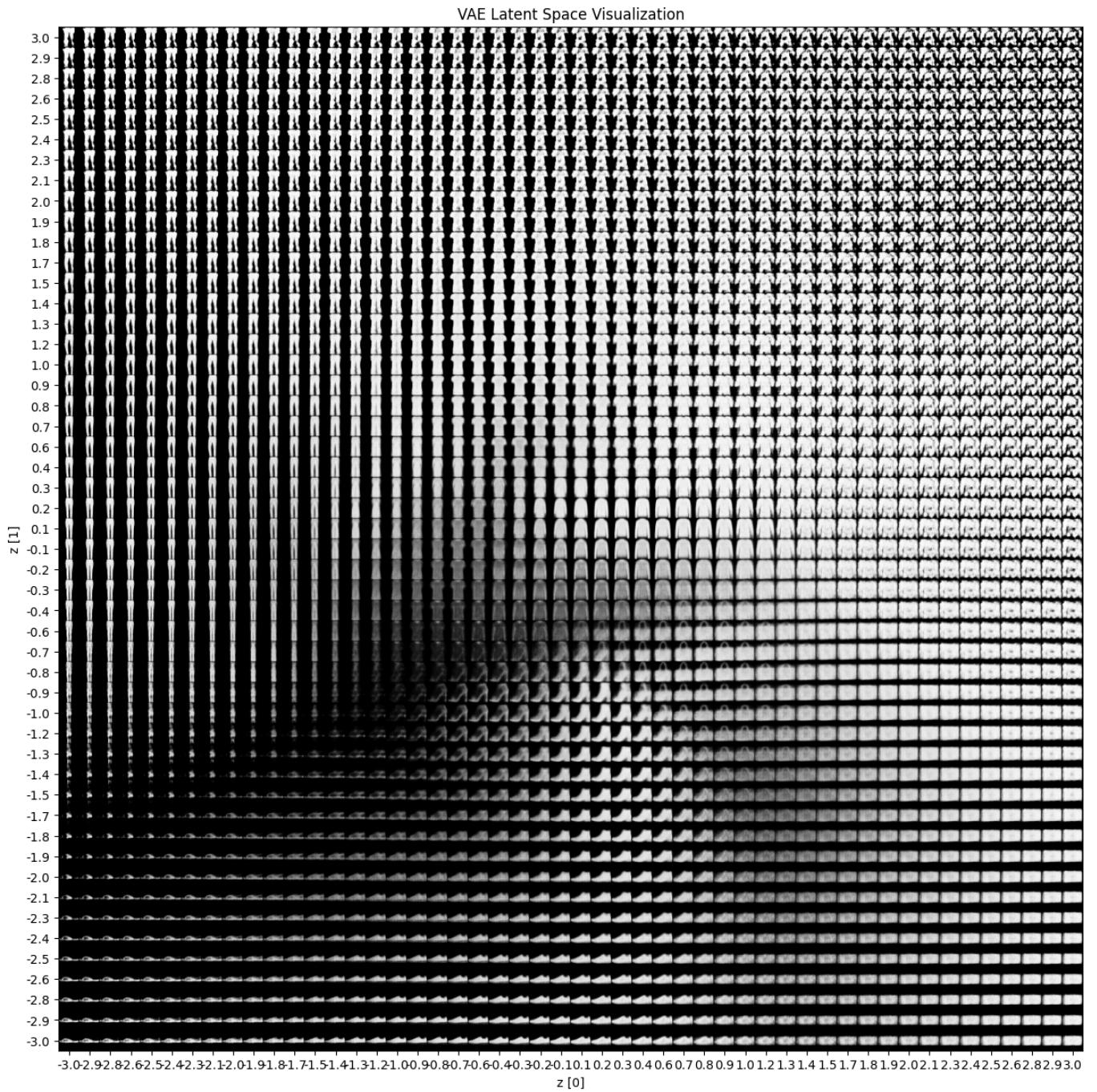
Average total loss: 146.53036391601563

Average reconstruction Loss: 139.8220814453125

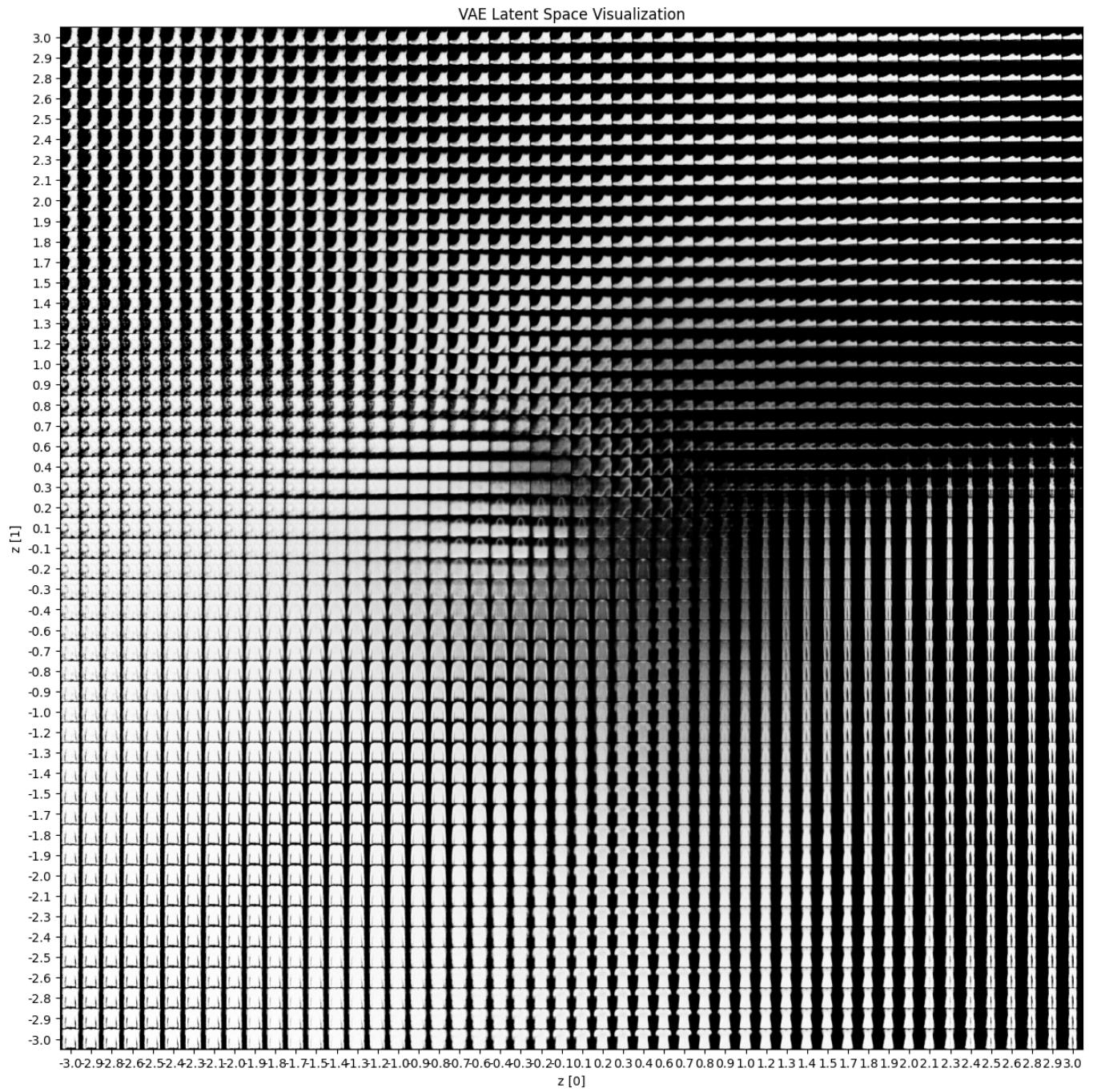
Average KL Loss: 6.7082821563720705

S obzirom da u samom radu nije prikazan gubitak za latentnu dimenziju 2 nije moguce uporediti tacnu vrednost gubitka ali s obzirom da je opseg vrednosti gubitka koje se dobijaju u radu za latentnu dimenziju 3 i odokativno isti broj iteracija treniranja oko 135-150 ovaj rezultat deluje prihvatljivo.

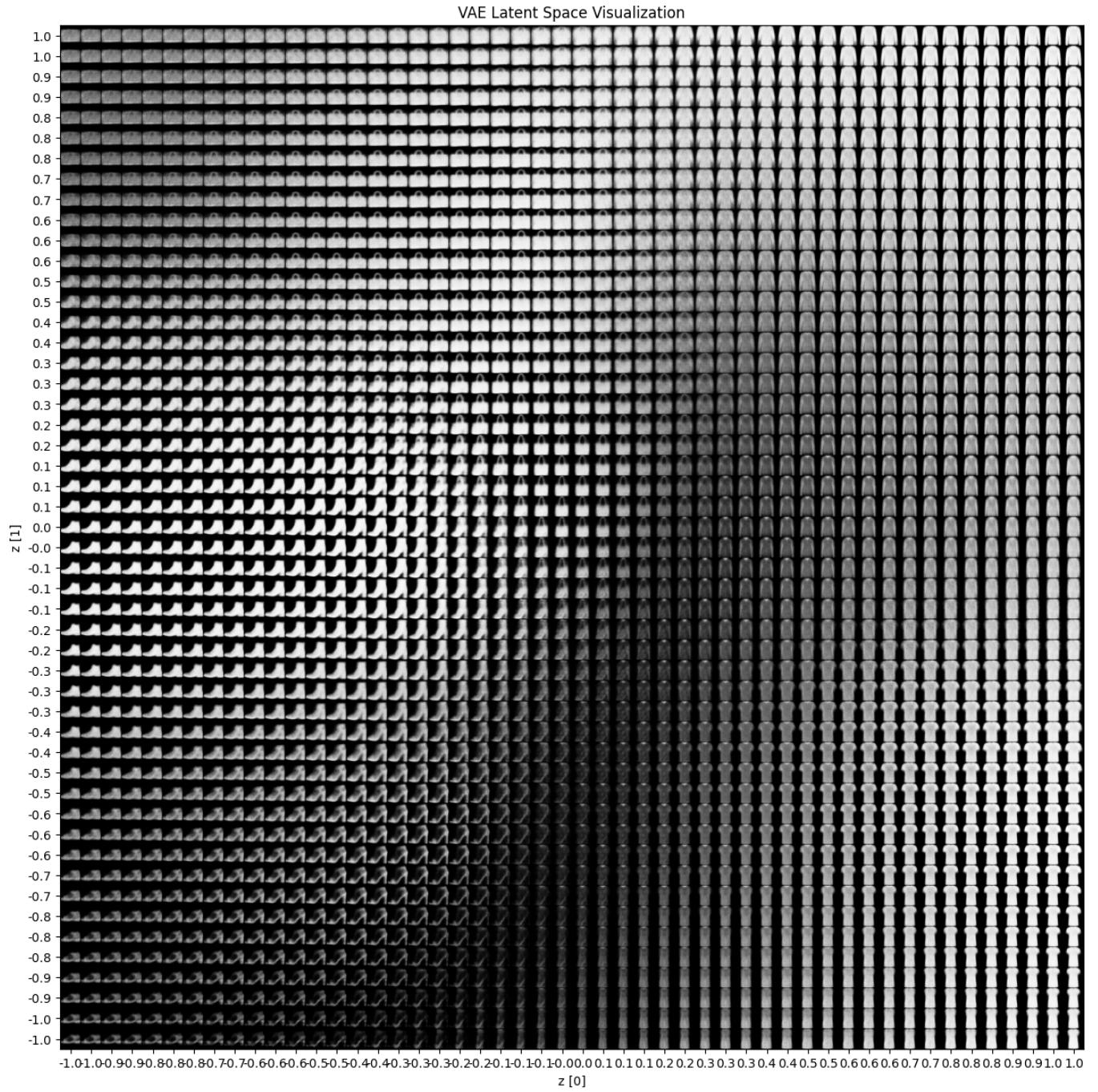
### 3.3 FashionMNIST



Slika 12 Prikaz latentnog prostora za velicinu skrivenog sloja 16

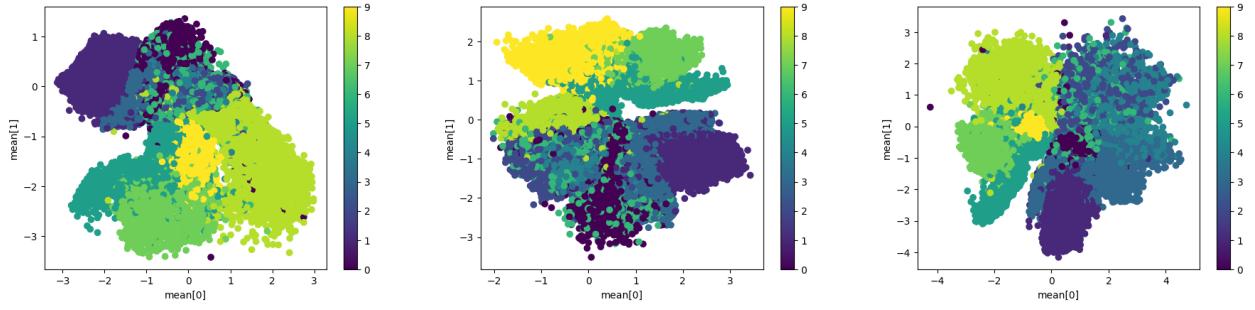


**Slika 13** Prikaz latentnog prostora za velicinu skrivenog sloja 100



**Slika 14** Prikaz latentnog prostora za velicinu skrivenog sloja 200

Vidimo da za veće dimenzije skrivenog sloja dobijamo detaljnije generisane primere slika i realističnije ali i da je razlika između 16 i 100 ogrona, dok je između 100 i 200 skoro pa neprimetna.



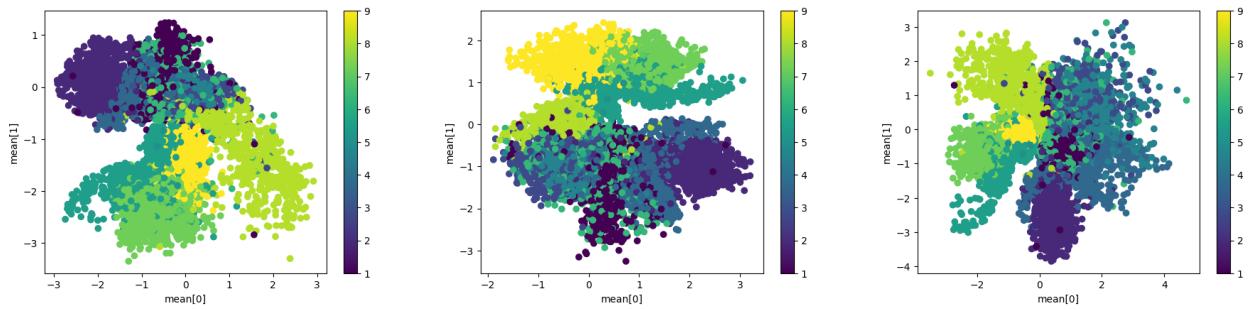
(a) velicina skrivenog sloja = 16

(b) velicina skrivenog sloja = 100

(c) velicina skrivenog sloja = 200

**Slika 15** Prikaz obucavajućeg skupa u latentnom prostoru

Vidimo da sa povećanjem dimenzija skrivenog sloja bolje odvajaju razlike klase u latentnom prostoru.



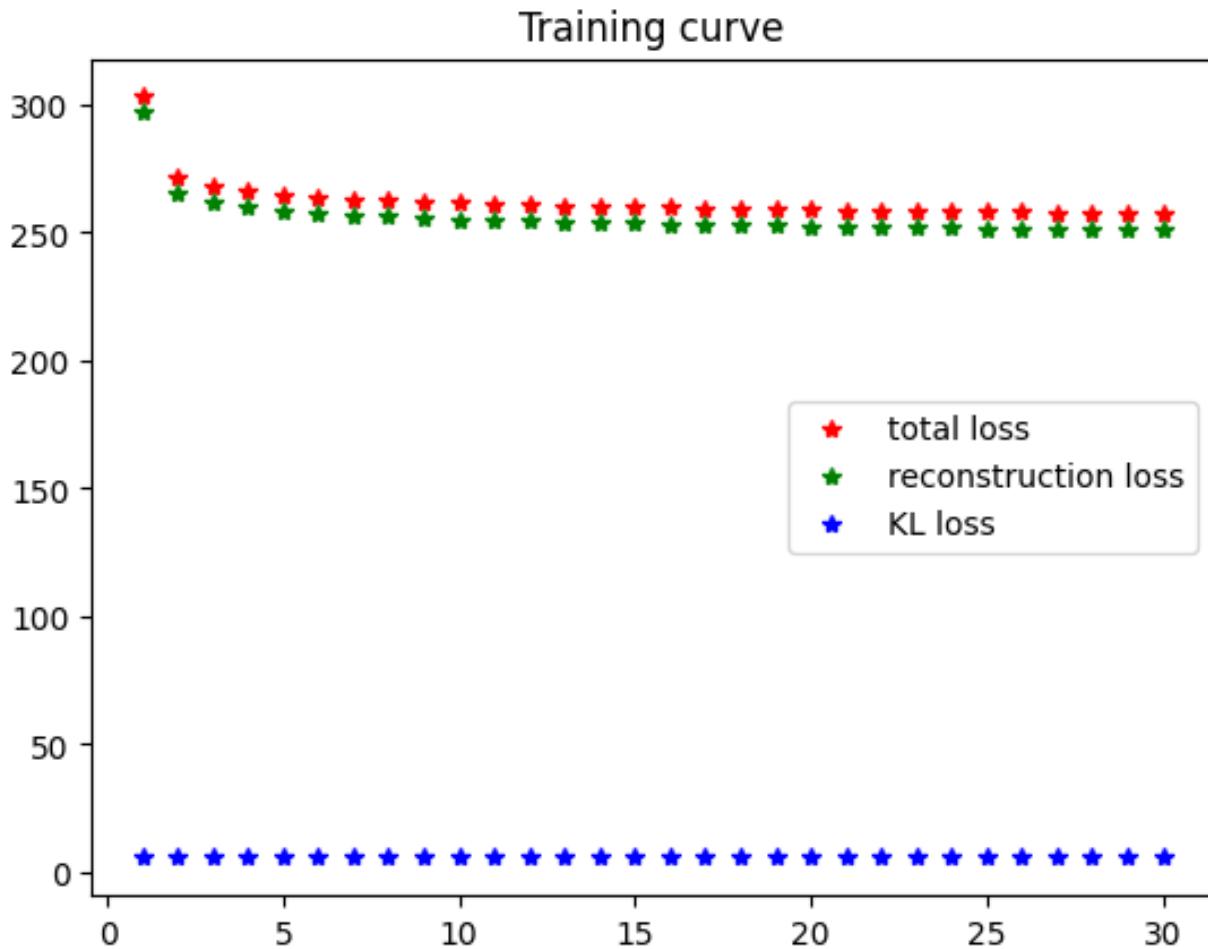
(a) velicina skrivenog sloja = 16

(b) velicina skrivenog sloja = 100

(c) velicina skrivenog sloja = 200

**Slika 16** Prikaz test skupa u latentnom prostoru

Vidimo da su klase slicno rasporedjene u latentnom prostoru za obucavajući i test skup sto nam govori da nas model dobro generazlicuje nauceno.



Slika 17 Prikaz treniranja za velicinu skrivenog sloja 100

Rezultati na test skupu:

Average total loss: 254.21143955078125

Average reconstruction Loss: 247.38248948567707

Average KL Loss: 6.828949353027344

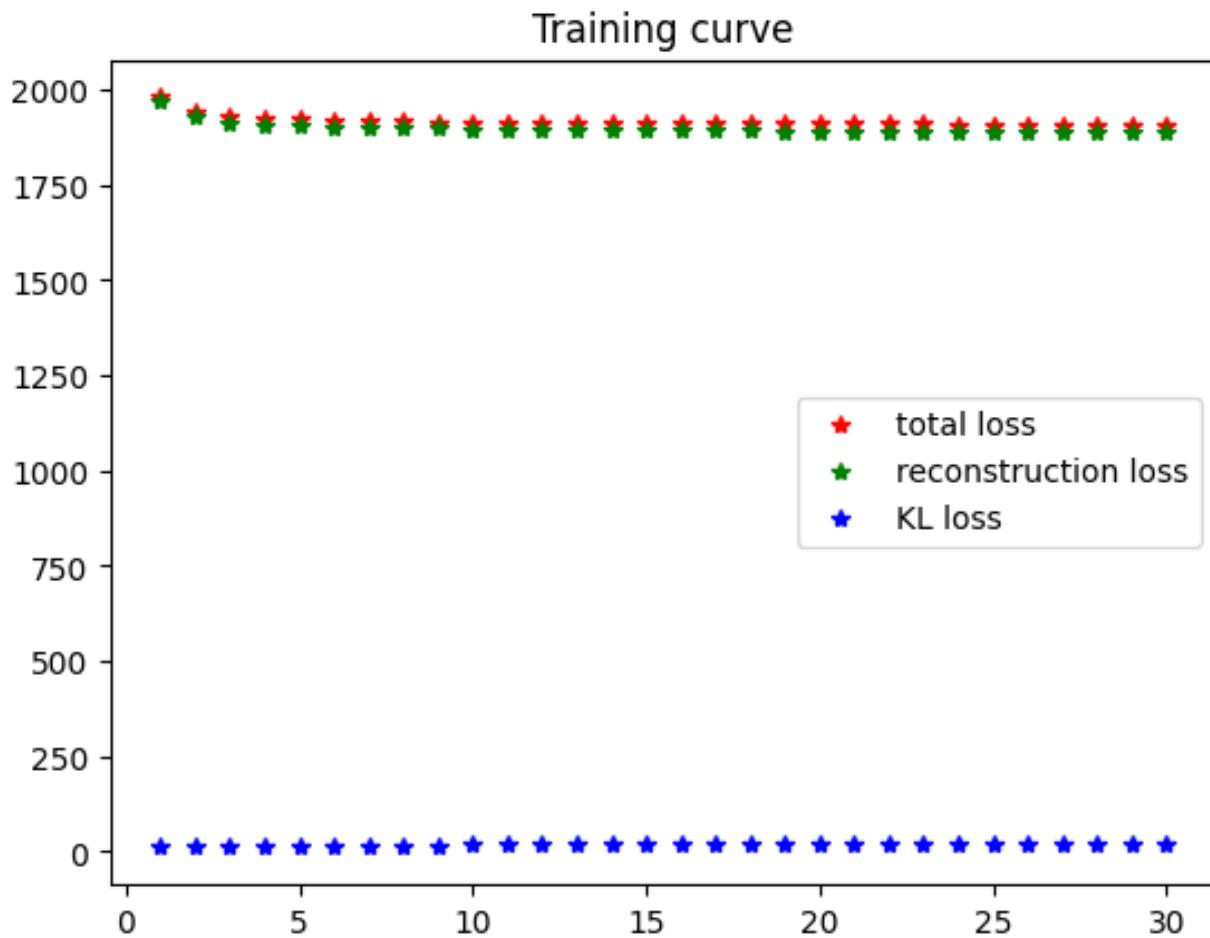
Ocekivano je sa porastom tezine problema porastao gubitak ali je red velicine idalje isti.

### 3.4 SVHN

S obzirom na veliku slozenost ovog problema velicina skrivenog sloja je povecana na 1024 a dimenzija latentnog prostora je podešena na 200. Sada nije moguce prikazati latentni prostor jer je latentni vektor prevelikih dimenzija.



Slika 18 Prikaz generisanih slika od strane modela



Slika 19 Prikaz treniranja

Vidimo da je greska dosta veca nego ranije. Greske su prikazane po slici, a kako slika trenutno ima 3 kanala ona se moze izraziti oko 600 po jednom kanalu sto je otprilike duplo vise nego na prethodnom skupu podataka. Ovaj rezultat ima smisla jer je sam problem mnogo kompleksniji od prethodnog. Rezultati na test skupu: Average total loss: 1908.6200860071392

Average reconstruction Loss: 1888.4970333466426

Average KL Loss: 20.123051950640168