

Name: Bileam Scheuven, Pankaj Bora  
Email: benedictbileam@gmx.de, bora.pankajt1@gmail.com  
Matr. Nr.:6983475, 6946375

## 1 Exercise 1

### 1.1 (a)

See code for implementation.

### 1.2 (b)

#### 1.2.1 LSTM

**Epoch 10:** Initialization: ‘but the night will be too sho...

Tolkien: — ‘but the night will be too short,’ said gandalf. ‘i have come back here, for i must have a little peace, alone. you should sleep,

LSTM model: — ‘but the night will be too shot te

**Observation:** Outputs repetitive meaningless sequences. LSTM hasn’t yet learned to model dependencies across time steps.

**Epoch 100:** Initialization: at last they came out of shado... Tolkien: — at last they came out of shadow to the seventh gate, and the warm sun that shone down beyond the river, as frodo walked in the gla

LSTM model: — at last they came out of shador and the store and the store of the store and the store of the store and the store of the store and

**Observation:** Can now capture basic word structures but still struggles with long term dependencies, leading to repetition.

**Epoch 482:** Initialization: pippin gazed in growing wonder...

Tolkien: — pippin gazed in growing wonder at the great stone city, vaster and more splendid than anything that he had dreamed of; greater and

LSTM model: — pippin gazed in growing wonder and stroke his head and seems that is a have begone the many, and the lord of the shadow of the city

**Observation:** Now outputs semantically plausible phrases. LSTM has learned to capture dependencies, and generates more coherent text. Mainly limited by its memory horizon.

### 1.2.2 TCN

**Epoch 10:** Initialization: ‘then why did you not say so a...

Tolkien: — ‘then why did you not say so at once?’ said bergil, and suddenly a look of dismay came over his face. ‘do not tell me that he has

TCN model: — ‘then why did you not say so ayy

**Observation:** TCN fails to generalize patterns and latches onto most common tokens.

**Epoch 100:** Initialization: ‘man!’ cried pippin, now thoro... Tolkien: — ‘man!’ cried pippin, now thoroughly roused. ‘man! indeed not! i am a hobbit and no more valiant than i am a man, save perhaps now

TCN model: — ‘man!’ cried pippin, now thoro!!!

 $((((??$ 

uu)))!!!

[illegible]

**Observation:** TCN’s architecture unable to capture long term dependencies needed for text generation leading to unstable and incoherent output.

**Epoch 482** Initialization: ‘that is good!’ said bergil. ‘...’

Tolkien: — ‘that is good!’ said bergil. ‘we boys are no longer allowed to pass the gate without an elder. now we shall see better.’

TCN model: — ‘that is good!’ said bergil. ‘j!u !! (followed by 100+ newlines)

**Observation:** TCN overfits or collapses during training. Observing the loss, it starts oscillating at around epoch 100 and never converges.

### 1.2.3 Transformer

**Epoch 10:** Initialization: ‘you are peregrin the halfling...

Tolkien: — ‘you are peregrin the halfling?’ he said. ‘i am told that you have been sworn to the service of the lord and of the city. welcome!’

Transformer model: — ‘you are peregrin the halfling the the the the the the the the the  
the the the the the the the the the t the the t the

**Observation:** Struggles to capture sequence order, possibly due to lack of positional embeddings.

**Epoch 100:** Initialization: pippin lifted it and presented...

Tolkien: — pippin lifted it and presented the hilt to him. ‘whence came this?’ said denethor. ‘many, many years lie on it. surely this is a b

Transformer model: — pippin lifted it and presented the hilt he him. ‘came the seany,’ said inet lan ore our of the ling oredred in pan in the win the l

**Observation:** Unable to effectively model dependencies between tokens in a sequence, causing semantic errors. Still generates contextually relevant phrases.

**Epoch 512:** Initialization: ‘can you find your way?’ said ...

Tolkien: — ‘can you find your way?’ said beregond at the door of the small hall, on the north side of the citadel, where they had sat. ‘it is

Transformer model: — ‘can you find your way?’ said beregond at the door of the small hall, on the north side of the citadel, where they had sat. ‘it is

**Observation:** Even without positional embedding, given that the training dataset is limited, the model basically overfits to the training data, able to almost always duplicate the original text.

## Exercise 2

### 1.3 (a)

See checkpoints for training results. VAE with  $\beta = 4$  was not trained as long as the others. For consistency we contrast only  $\beta \in [0, 20]$  but it can be reasonably assumed that 4 would perform somewhere in between. We unfortunately learned too late that we had to change the queue type from test to submit longer jobs, so we trained locally and were therefore slightly limited in our experiments.

## 1.4 (b)

Images of a VAE with higher beta seem to have more contrast and look sharper but a little less coherent/realistic, while the VAE with beta 0 produces blurry images of plausible looking faces. Due to computational constraints it is likely that neither model was trained to convergence, and it is thus unclear but likely that this effect would persist or even amplify with more training. The beta value enforces more distinct structure in the latent space, while a vanilla VAE is free to learn an average of all possible faces. Results can be seen in [4](#)

Figure 1: 16 Randomly sampled images with varying  $\beta$

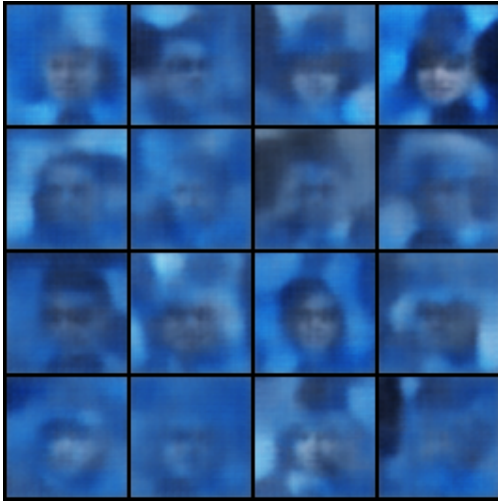


Figure 2:  $\beta = 0$



Figure 3:  $\beta = 20$

## 1.5 (c)

For the vanilla VAE varying individual dimensions produces drastically different faces, but it is rarely clear how to interpret this change as opposite faces have little in common. Additionally many dimensions are similar in their output. For the VAE with beta 20, the distinction is stronger and for many dimensions one can make an educated guess about what the dimension represents, sunglasses vs hats (dim 0 and 3), hair vs bald (dim 5) or skin color (dim 18). Results can be seen in [7](#)

## 1.6 (d)

Multiple dimensions contain some information about rotation/ facial orientation, but for our model, dimension 11 of the beta20 VAE was the most prominent. Still, negative perturbations have little effect (possibly due to the sigmoid activation clamping the negative rotation away), leaving the face mostly centered. Positive perturbations however rotate the face as desired. The face is not kept fully intact, in fact for drastic changes (like in the given example) the face is rotated correctly but barely resembles the original face. Results can be seen in [7](#).

Figure 4: Random latent vector with individually perturbed dimensions with varying  $\beta$

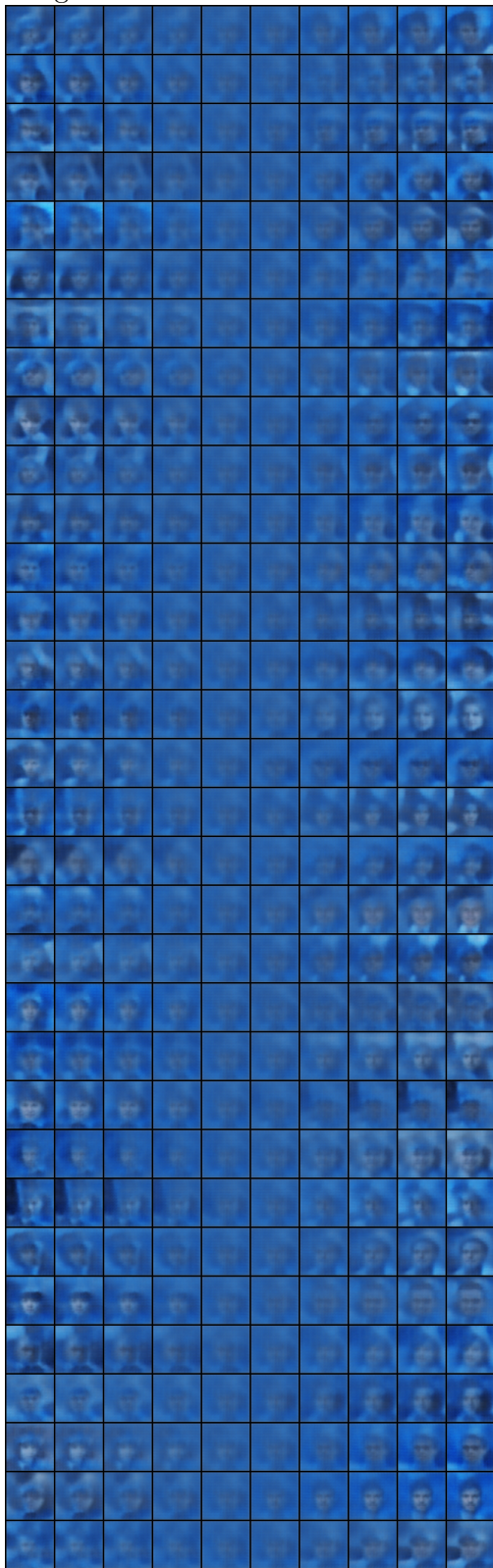


Figure 5:  $\beta = 0$

7



Figure 6:  $\beta = 20$

Figure 7: Image rotated left and right (positive and negative latent dimension)

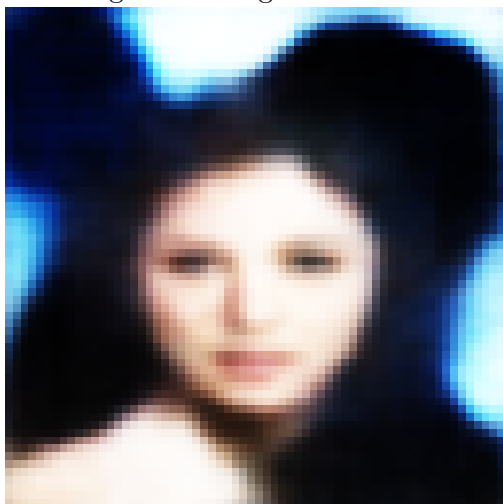


Figure 8: negative latent perturbation

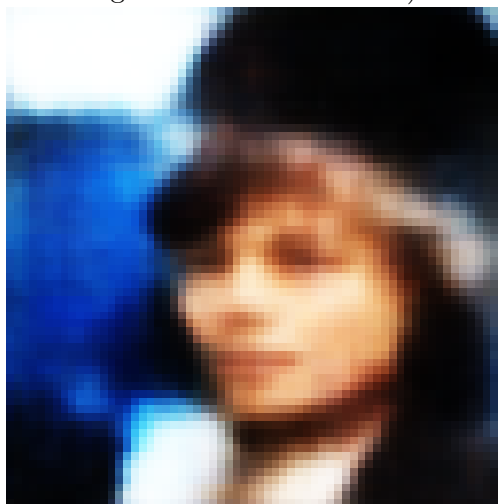


Figure 9: positive latent perturbation