Universität St.Gallen

GSERM - Deep Learning: Fundamentals and Applications (2023)

IaroHasNoFashionNet

Iaroslav Kriuchkov

Setup

- 4 CNNs: IaroHasNoFashionNet**1-4**
- Training included all possible combinations of the following learning rates and mini batch sizes:
  - Learning rates: 0.0005, 0.001, 0.005, 0.01
  - Mini batch sizes: 8, 16, 32, 64, 128, 256, 512
- 100 epochs for each training

Computer setup

- Aalto University high-performance computing cluster
- Nvidia DGX-1 machine
- 1 CPU, 12 GB RAM
- 1 NVIDIA V100 GPU (32 GB)

Model	Parameters	Job runtime using CUDA
IaroHasNoFashionNet1	21 352	08h 42min
IaroHasNoFashionNet2	885 470	10h 08min
IaroHasNoFashionNet3	2 114 490	10h 47min
IaroHasNoFashionNet4	2 273 170	11h 08min

Universität St.Gallen

Summary of models

Model	Parameters	Convolution layers	Max Pool layers	Fully connected layers
IaroHasNoFashionNet1	21 352	1	1	2
IaroHasNoFashionNet2	885 470	1	1	3
IaroHasNoFashionNet3	2 114 490	2	1	3
IaroHasNoFashionNet4	2 273 170	2	2	3

Universität St.Gallen

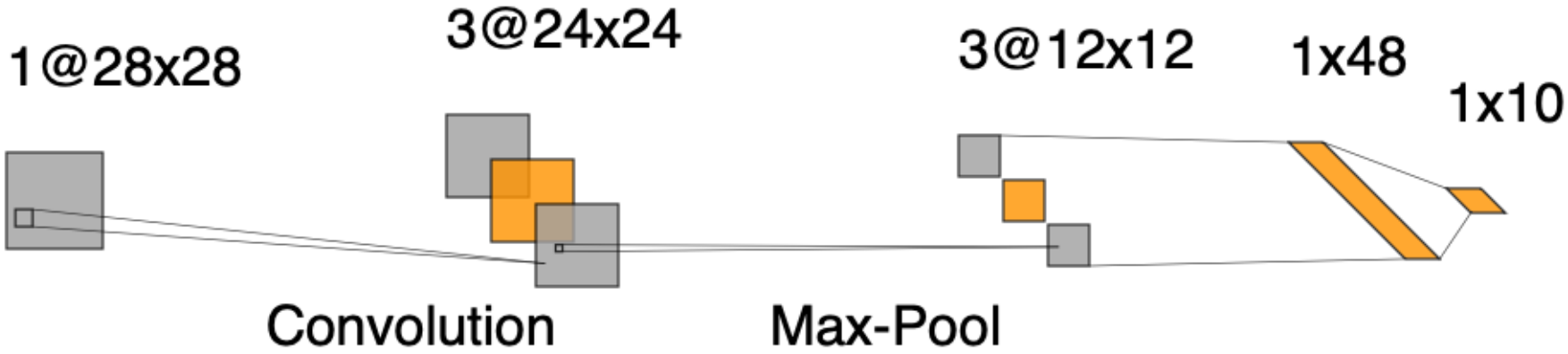
Summary of results

Model	Best evaluation accuracy	Learning rate	Mini batch size	Min training loss @ epoch	Evaluation loss	Time per epoch
IaroHasNoFashionNet1	0.8941	0.01	32	0.175 @ 99	0.329	11.38 sec
IaroHasNoFashionNet2	0.9253	0.005	8	0.00033 @ 99	0.538	23.51 sec
IaroHasNoFashionNet3	0.9220	0.01	8	0.000021 @ 99	0.789	26.12 sec
IaroHasNoFashionNet4	0.9234	0.01	8	0.000028 @ 99	0.718	30.49 sec

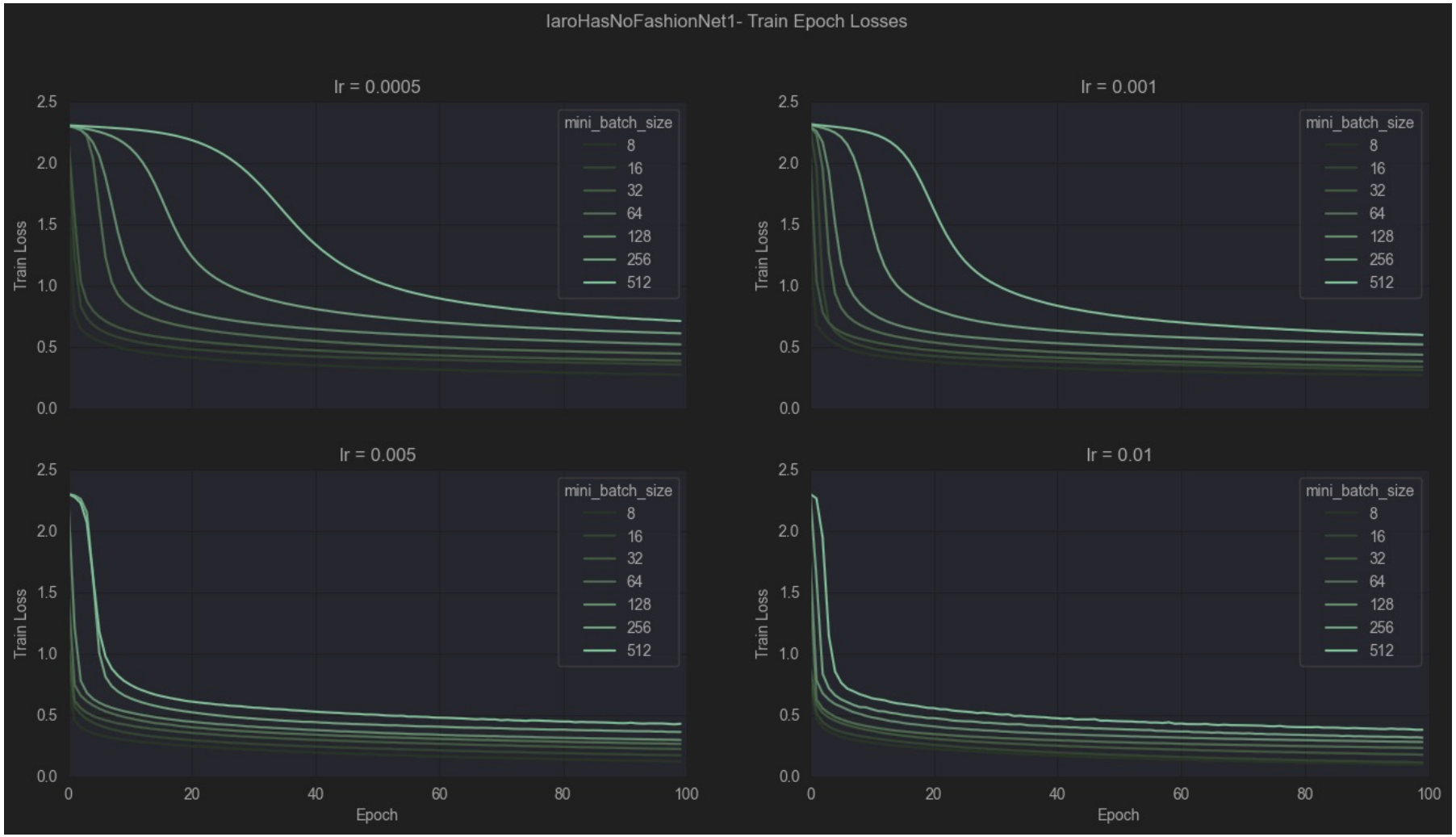
IaroHasNoFashionNet1

Universität St.Gallen

IaroHasNoFashionNet1 – Structure

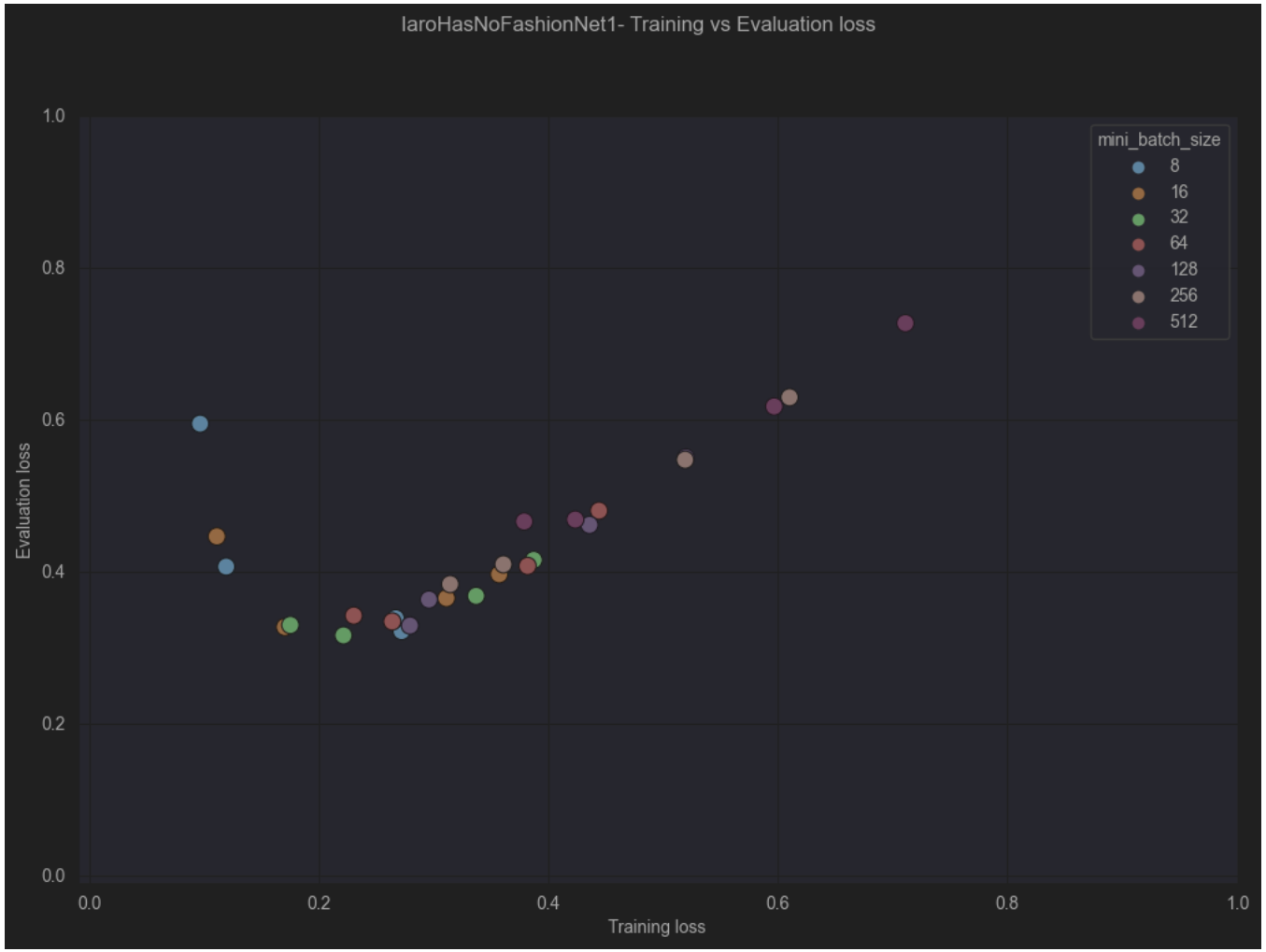


IaroHasNoFashionNet1 – Train loss



Main learning:
Lower learning rates combined with bigger mini-batch sizes require more epochs to start to converge, and it seems that 100 epochs are not enough. For higher learning rates and smaler mini-batches, at around 80 epochs training loss barely keeps getting lower.

IaroHasNoFashionNet1 – Train vs eval loss

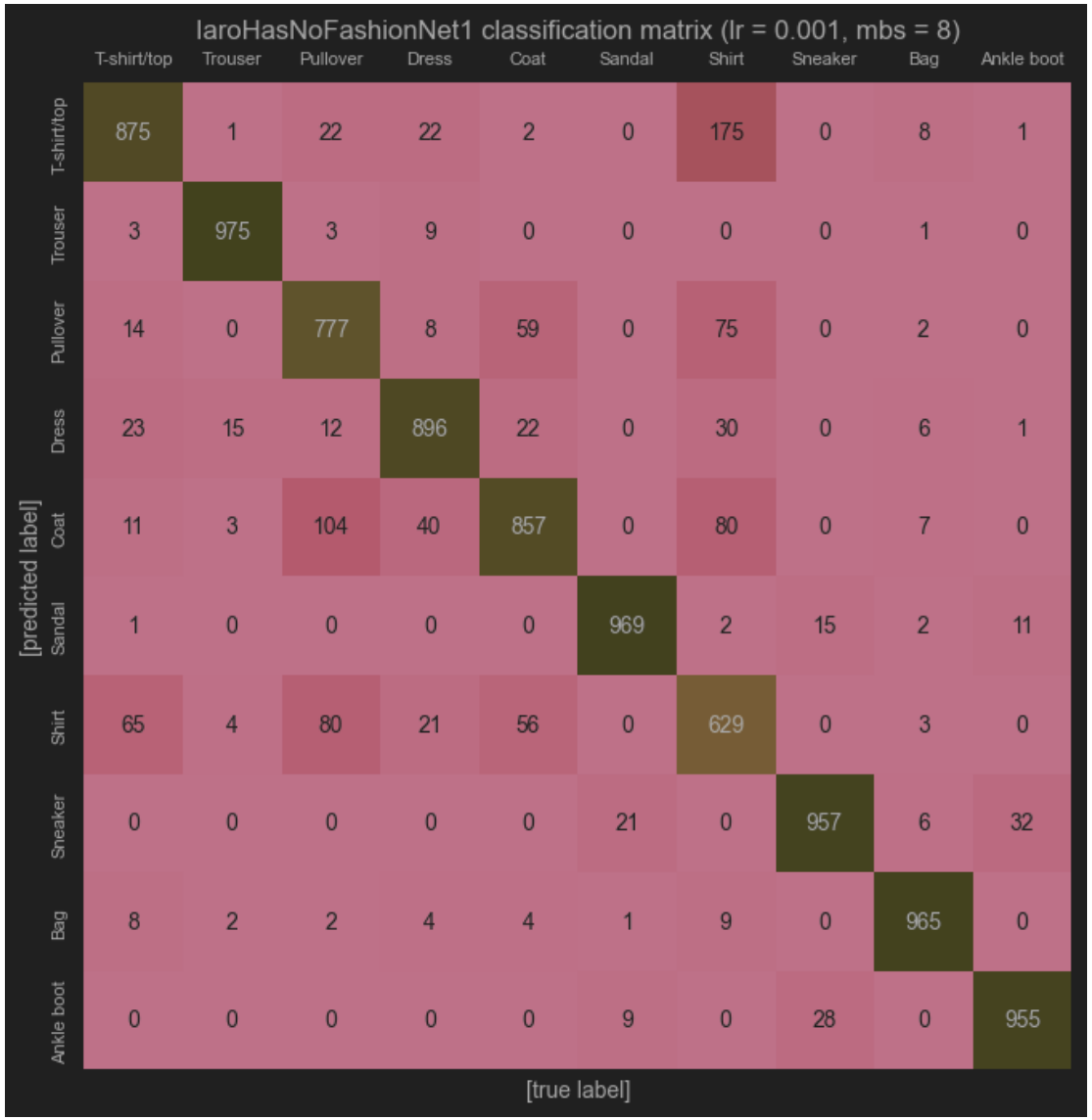


Main learning:
When the training loss goes below 0.4, models tend to overfit, as the evaluation loss is not comparable with the train loss. This can be especially observed among the models with the low mini-batch size (between 8 and 64): with the training loss below 0.2, the evaluation loss is 0.3 and beyond.

IaroHasNoFashionNet1 – Eval accuracy

Main learning:
Variance in evaluation accuracy is higher at lower learning rates. There is a tendency, that lower mini-batch sizes provide higher evaluation accuracy, however, with higher learning rate this difference is being neglected.

IaroHasNoFashionNet1 – Conf matrix



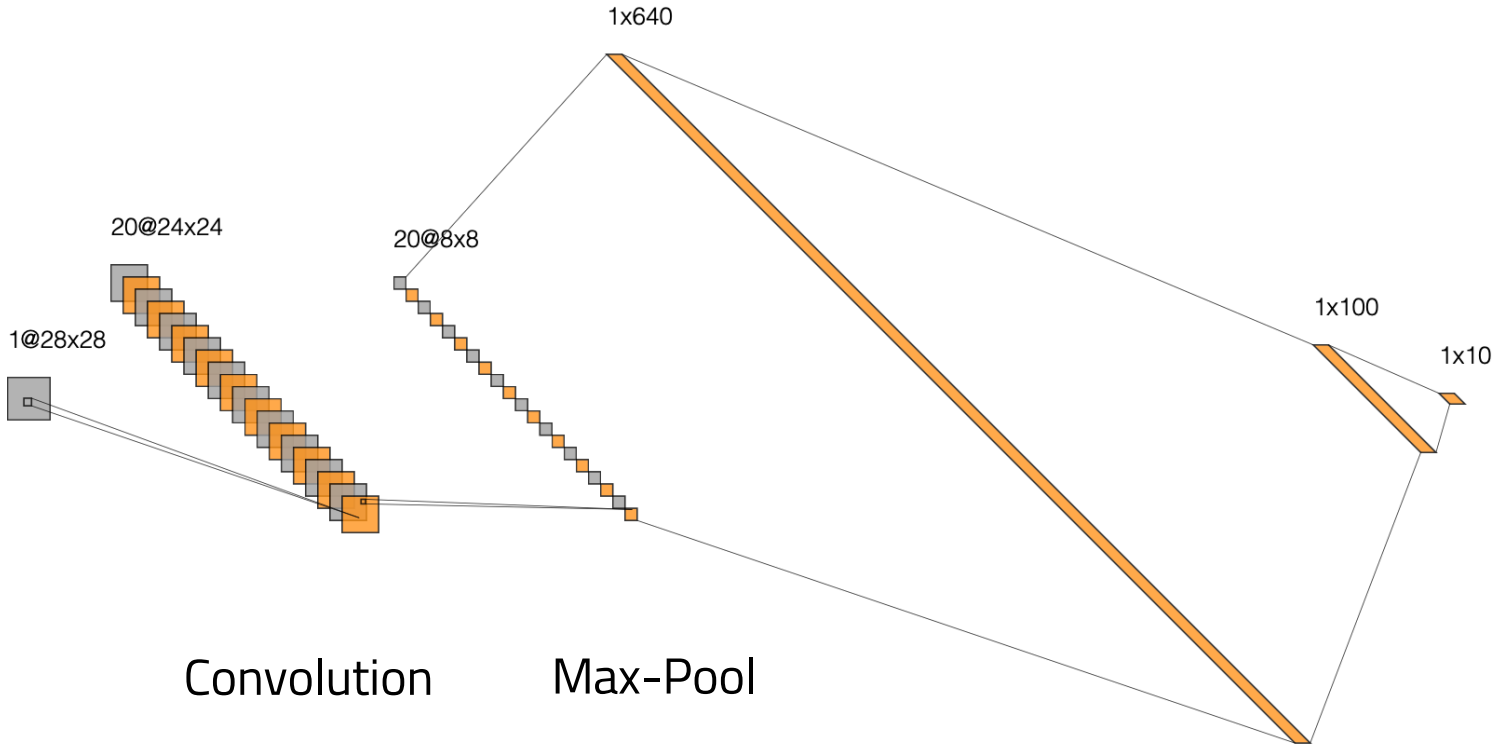
Main learning:

The main issue is in differentiating between shirts, T-shirts, coats and pullovers. This is the main factor to account for, as there are no problems distinguishing between different types of shoes.

IaroHasNoFashionNet2

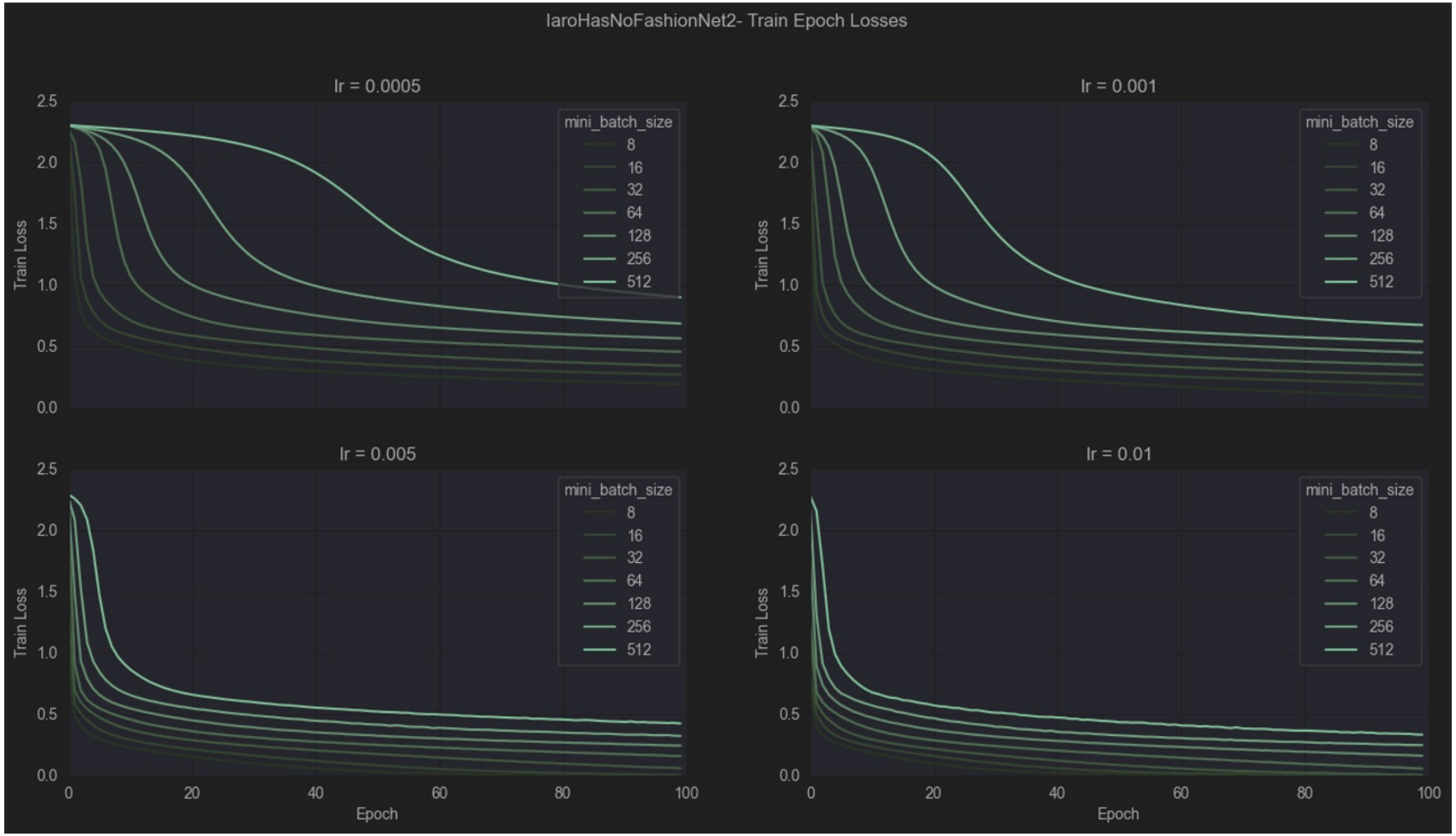
Universität St.Gallen

IaroHasNoFashionNet2 – Structure



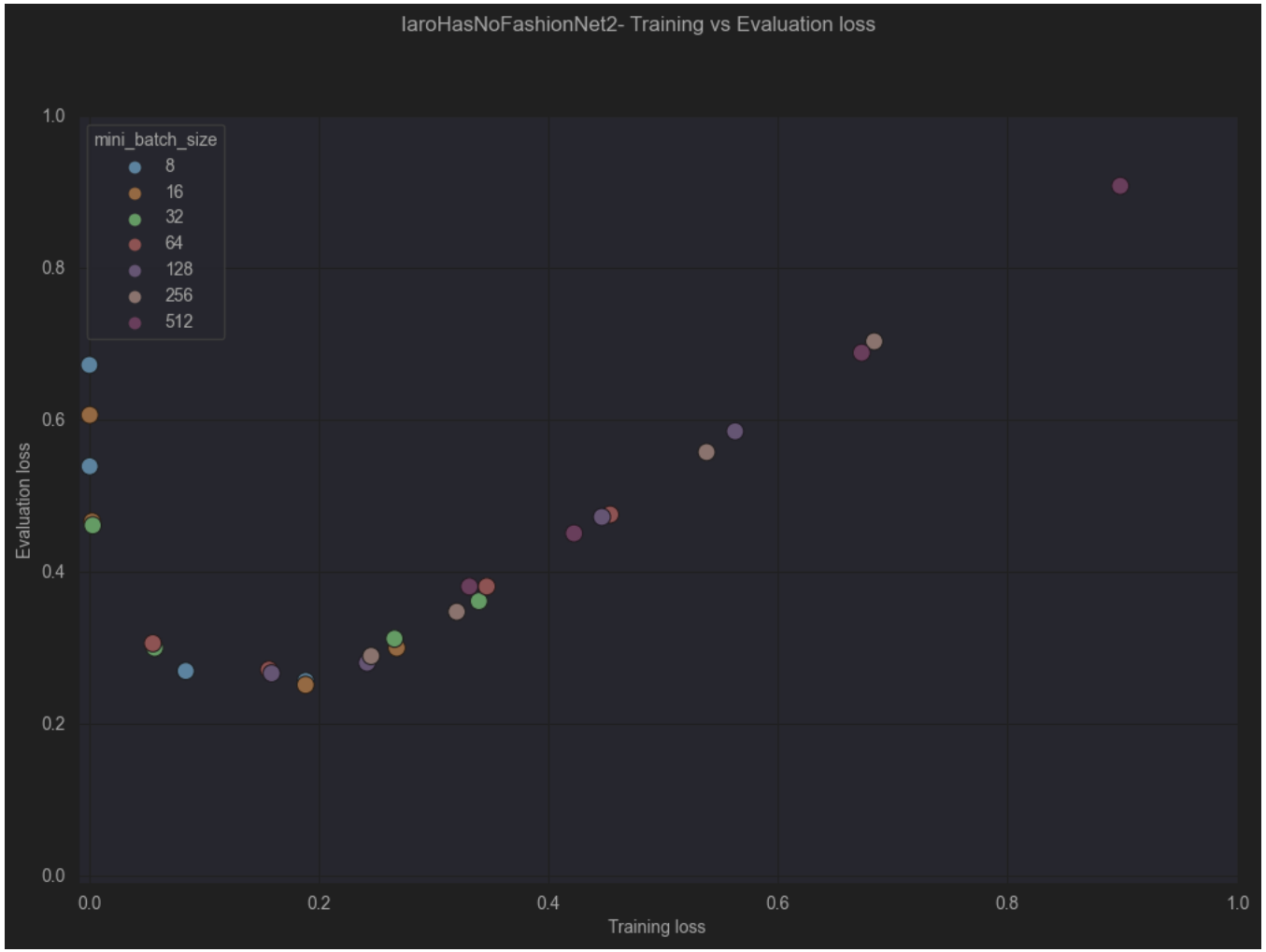
Universität St.Gallen

IaroHasNoFashionNet2 – Train loss



Main learning:
Lower learning rates combined with bigger mini-batch sizes require more epochs to start to converge, and it seems that 100 epochs are not enough. For higher learning rates and smaler mini-batches, at around 80 epochs training loss barely keeps getting lower.

IaroHasNoFashionNet2 – Train vs eval loss



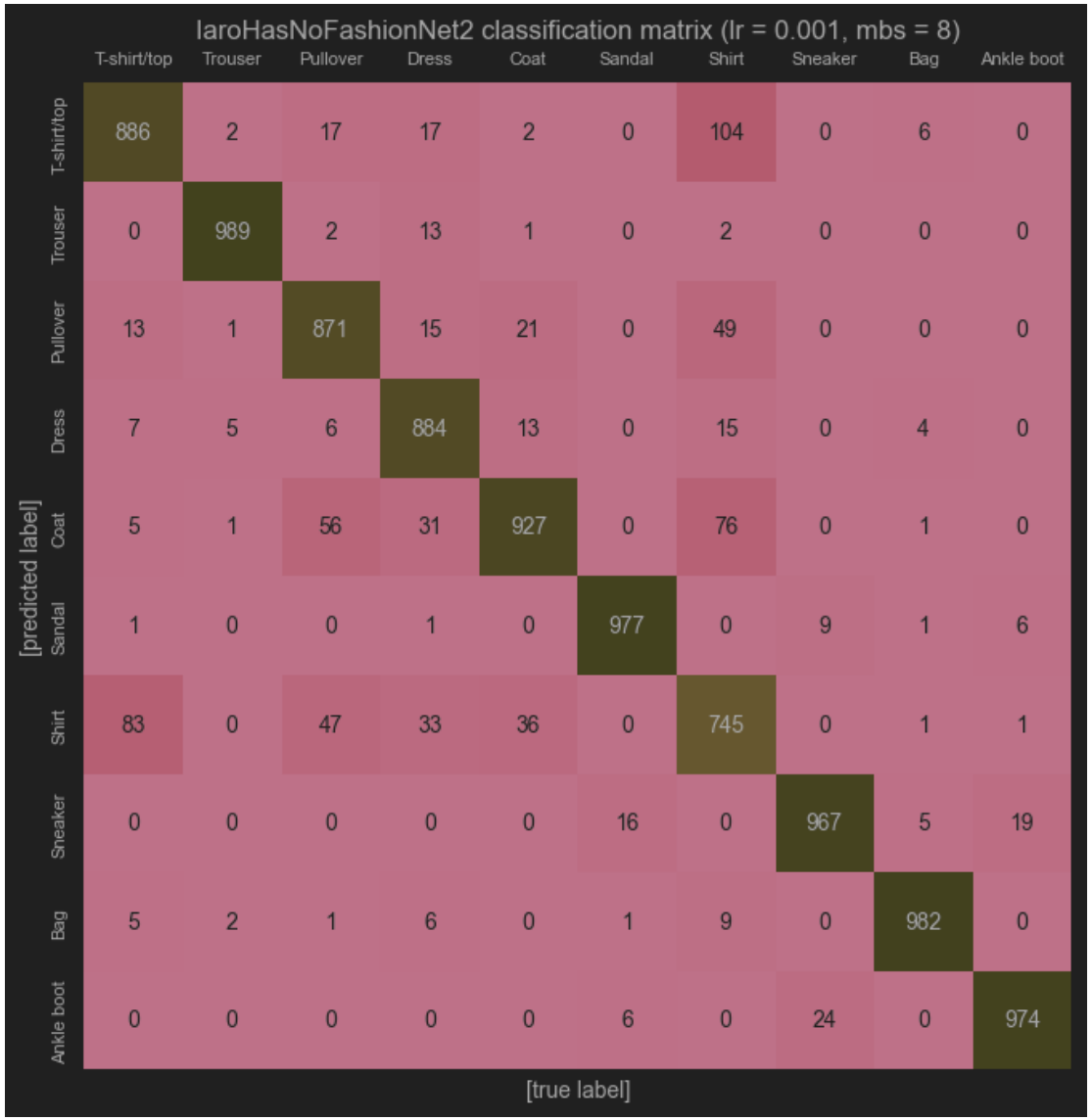
Main learning:
When the training loss goes below 0.4, models tend to overfit, as the evaluation loss is not comparable with the train loss. This can be especially observed among the models with the low mini-batch size (between 8 and 64): with the training loss below 0.2, the evaluation loss is 0.3 and beyond.

IaroHasNoFashionNet2 – Eval accuracy

Main learning:
Variance in evaluation accuracy is higher at lower learning rates. There is a tendency, that lower mini-batch sizes provide higher evaluation accuracy, however, with higher learning rate this difference is being neglected.

Universität St.Gallen

IaroHasNoFashionNet2 – Conf matrix



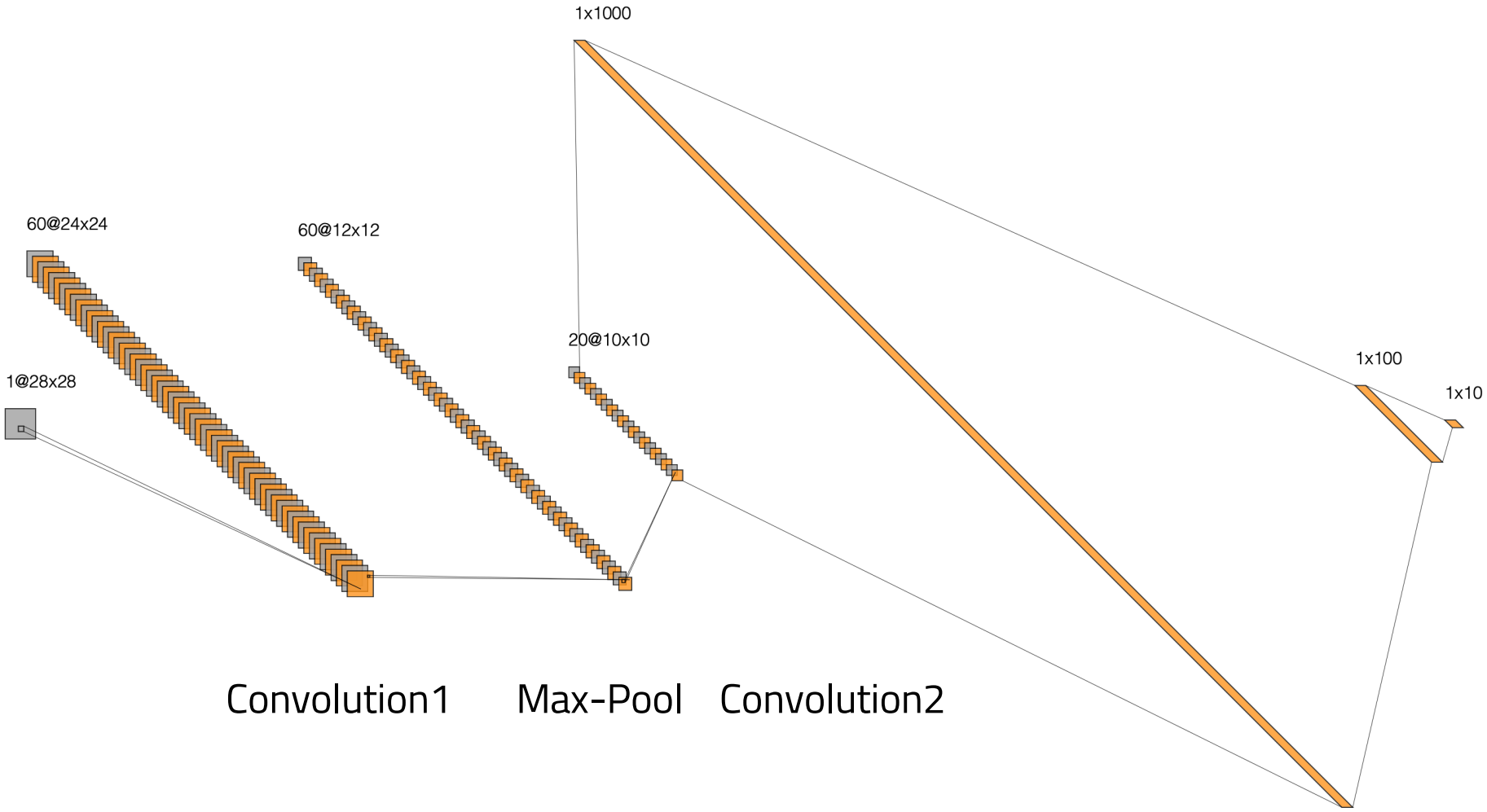
Main learning:

The main issue is in differentiating between shirts, T-shirts, coats and pullovers. This is the main factor to account for, as there are no problems distinguishing between different types of shoes.

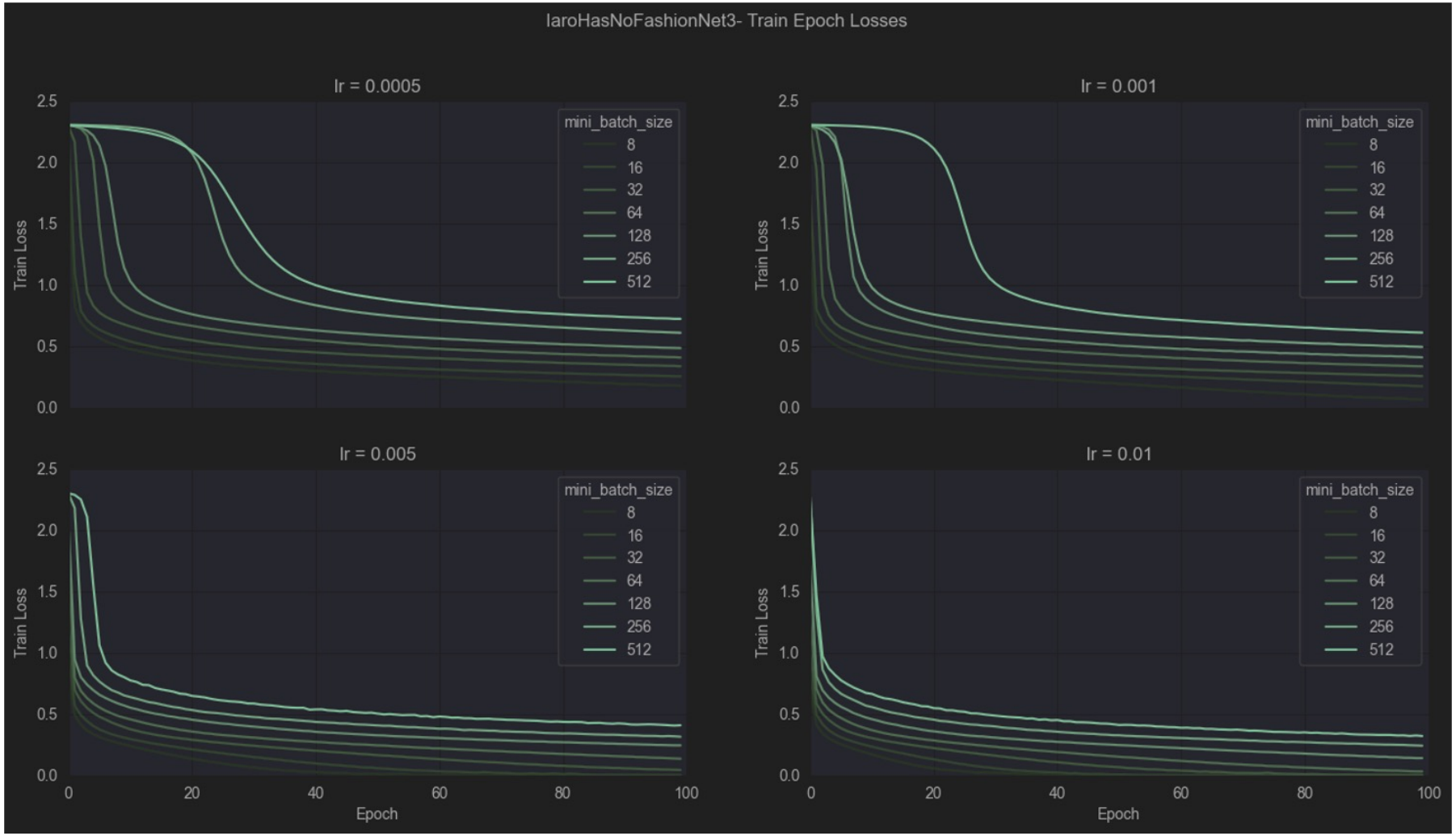
IaroHasNoFashionNet3

Universität St.Gallen

IaroHasNoFashionNet3 – Structure

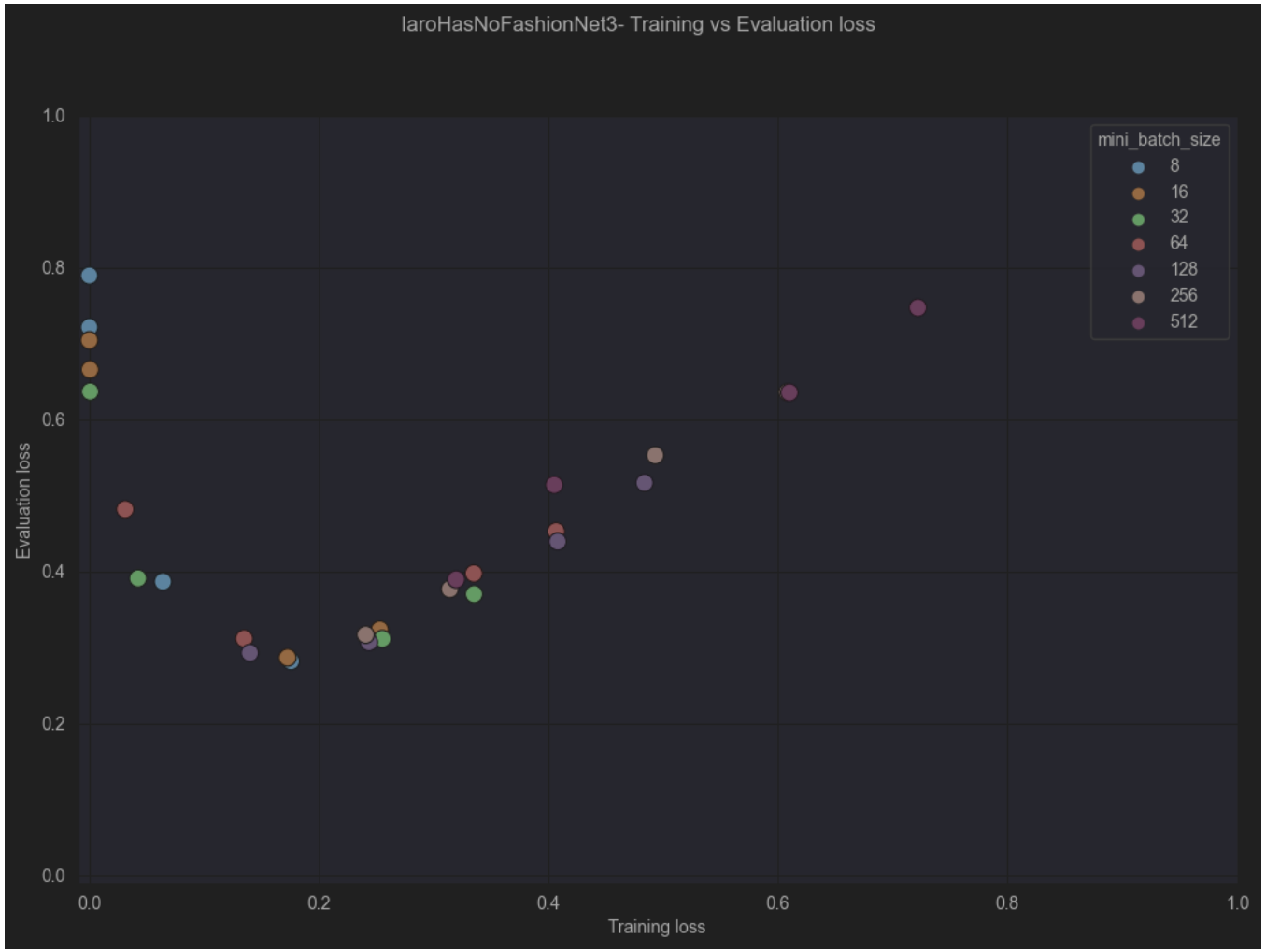


IaroHasNoFashionNet3 – Train loss



Main learning:
Lower learning rates combined with bigger mini-batch sizes require more epochs to start to converge, and it seems that 100 epochs are not enough. For higher learning rates and smaler mini-batches, at around 80 epochs training loss barely keeps getting lower.

IaroHasNoFashionNet3 – Train vs eval loss



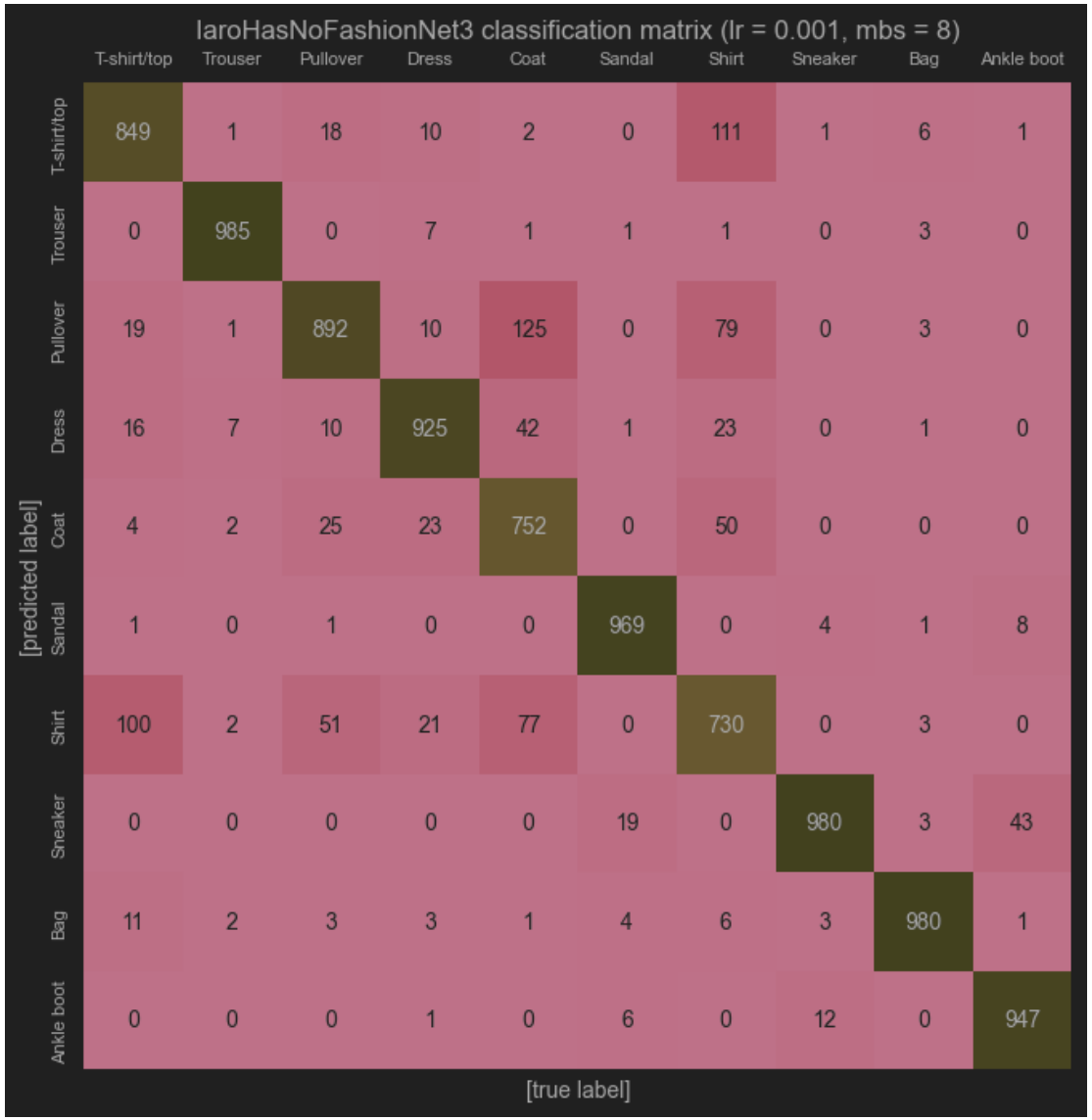
Main learning:
When the training loss goes below 0.4, models tend to overfit, as the evaluation loss is not comparable with the train loss. This can be especially observed among the models with the low mini-batch size (between 8 and 64): with the training loss below 0.2, the evaluation loss is 0.3 and beyond.

IaroHasNoFashionNet3 – Eval accuracy

Main learning:
Variance in evaluation accuracy is higher at lower learning rates. There is a tendency, that lower mini-batch sizes provide higher evaluation accuracy, however, with higher learning rate this difference is being neglected.

Universität St.Gallen

IaroHasNoFashionNet3 – Conf matrix



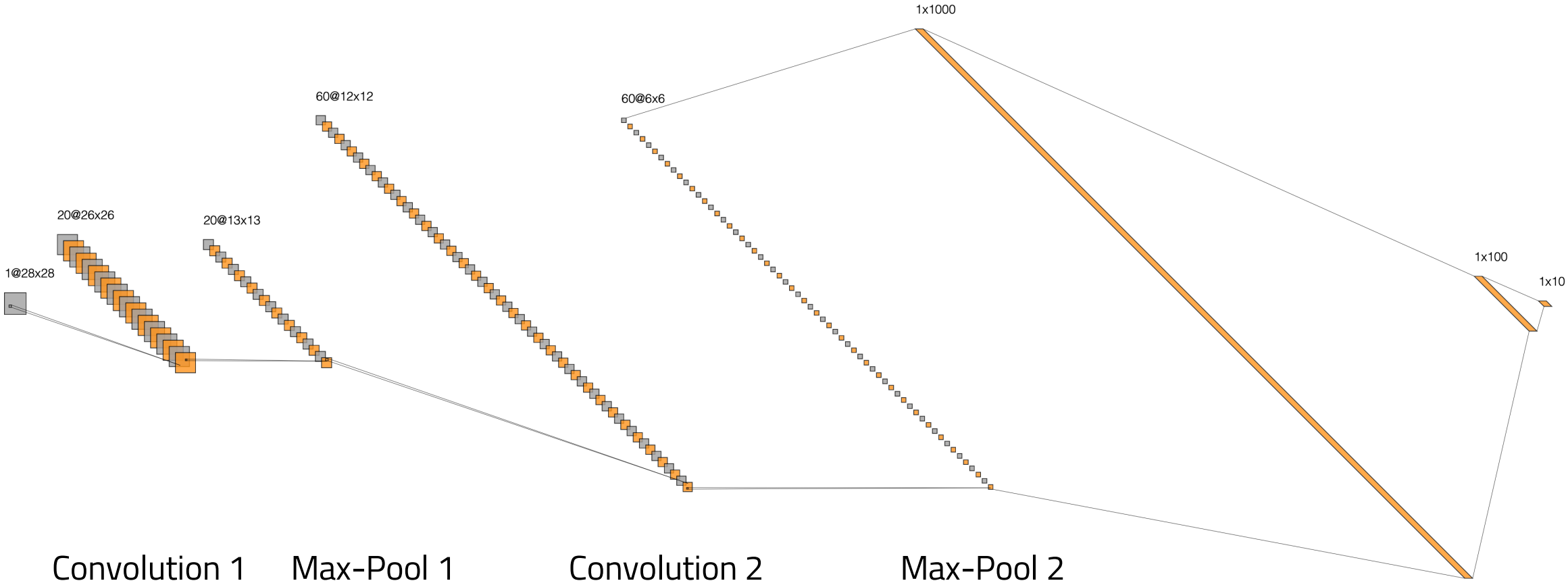
Main learning:

The main issue is in differentiating between shirts, T-shirts, coats and pullovers. This is the main factor to account for, as there are no problems distinguishing between different types of shoes.

IaroHasNoFashionNet4

Universität St.Gallen

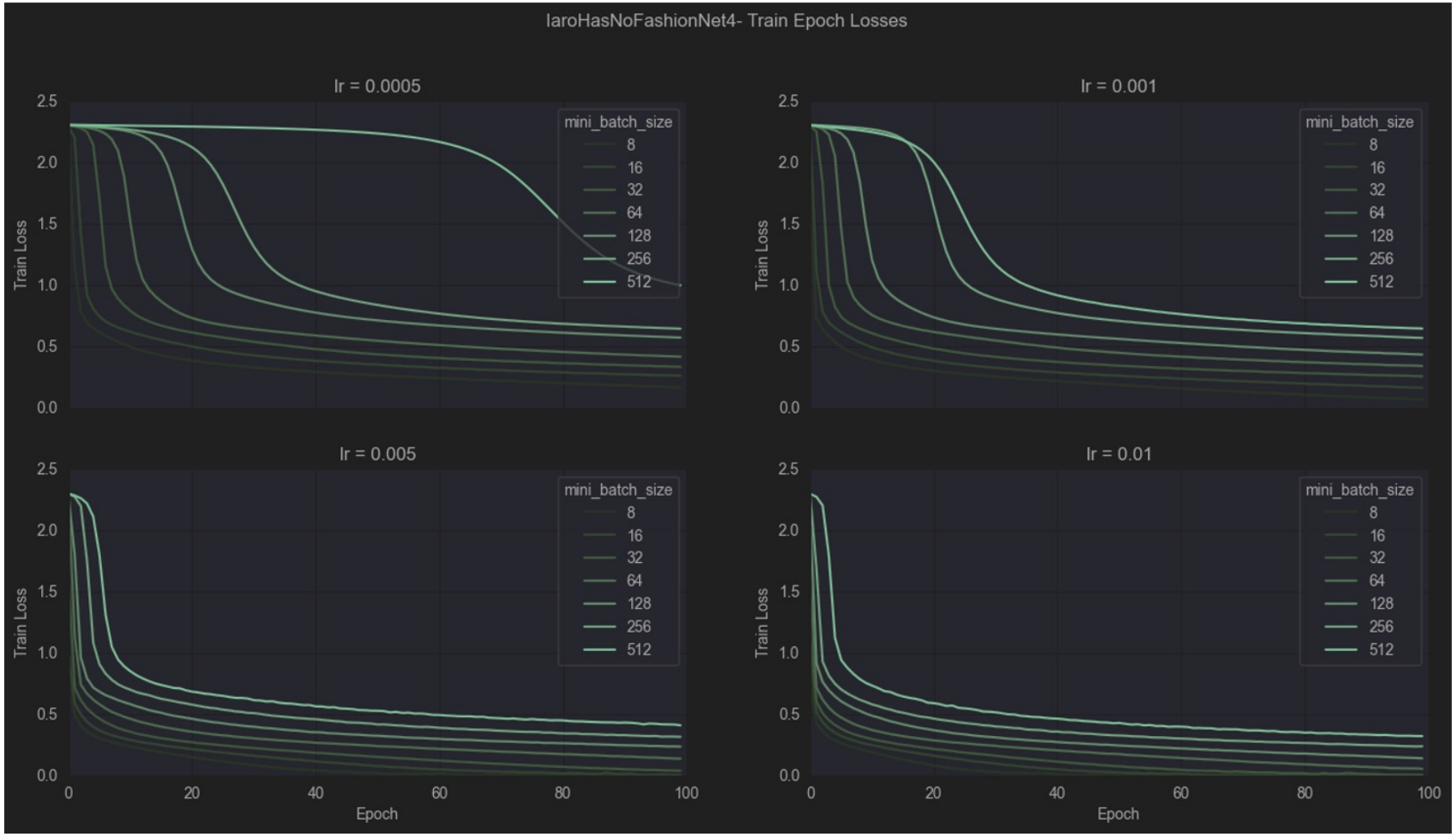
IaroHasNoFashionNet4 – Structure



Convolution 1 Max-Pool 1 Convolution 2 Max-Pool 2

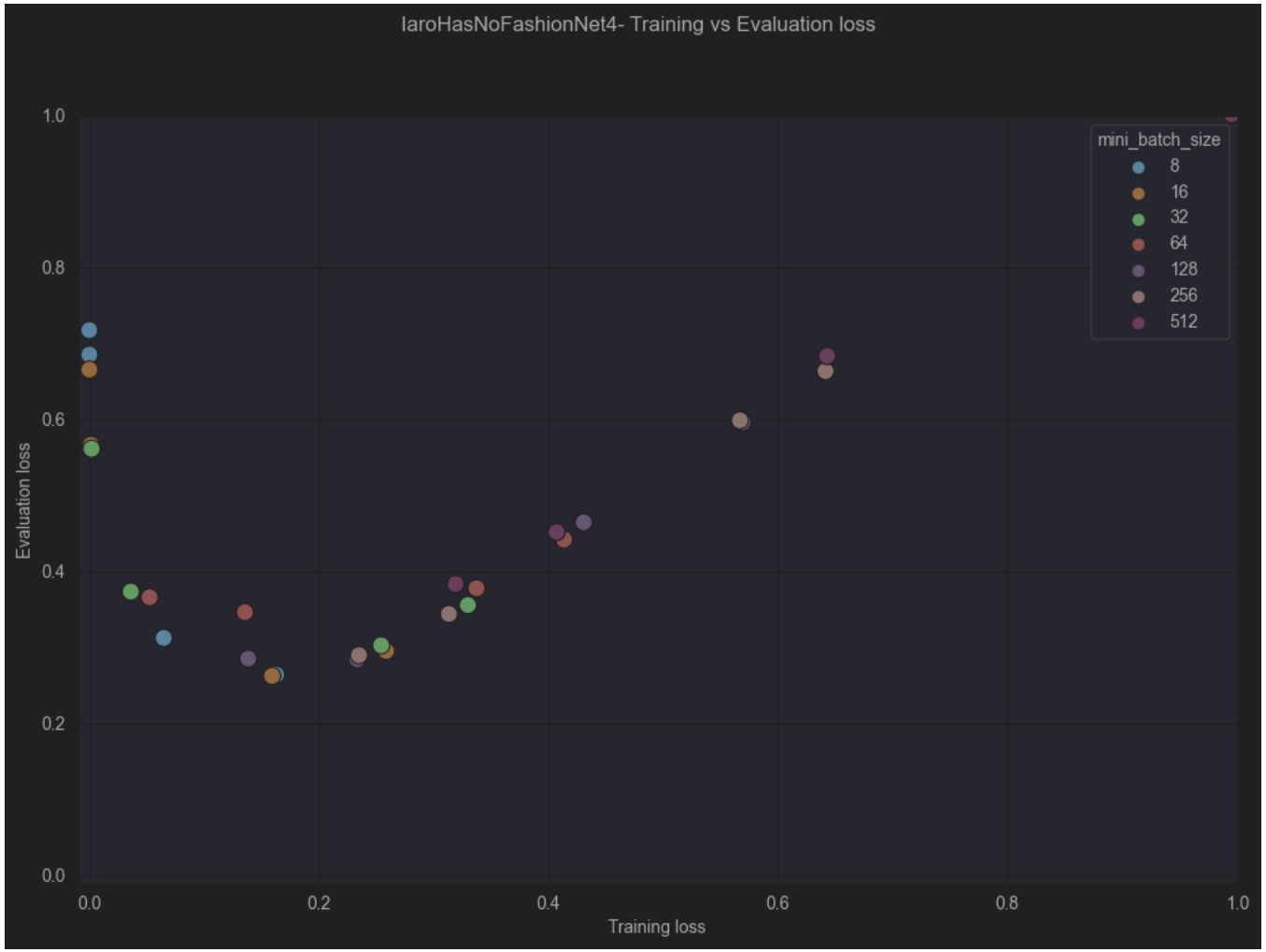
Universität St.Gallen

IaroHasNoFashionNet4 – Train loss



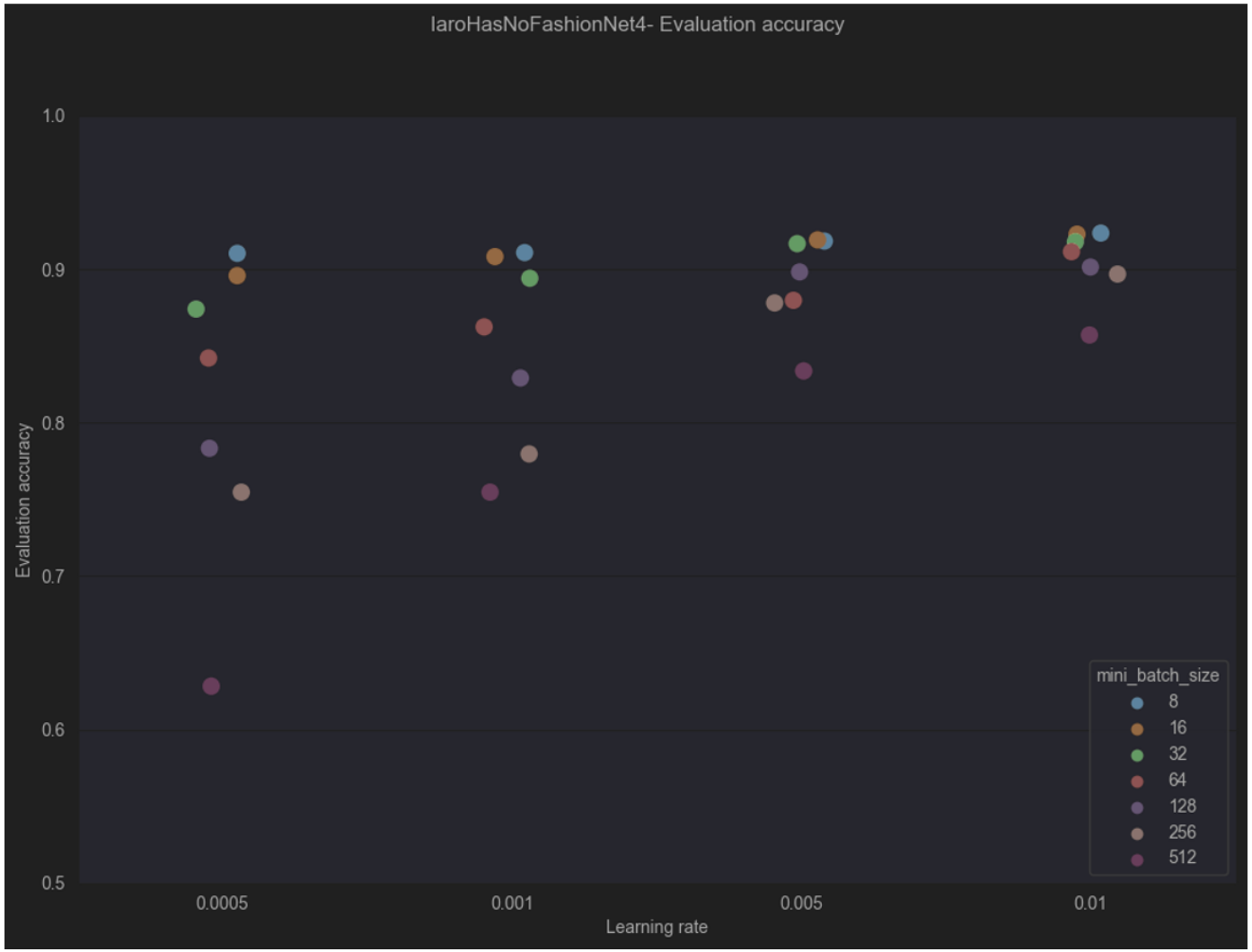
Main learning:
Lower learning rates combined with bigger mini-batch sizes require more epochs to start to converge, and it seems that 100 epochs are not enough. For higher learning rates and smaler mini-batches, at around 80 epochs training loss barely keeps getting lower.

IaroHasNoFashionNet4 – Train vs eval loss



Main learning:
When the training loss goes below 0.4, models tend to overfit, as the evaluation loss is not comparable with the train loss. This can be especially observed among the models with the low mini-batch size (between 8 and 64): with the training loss below 0.2, the evaluation loss is 0.3 and beyond.

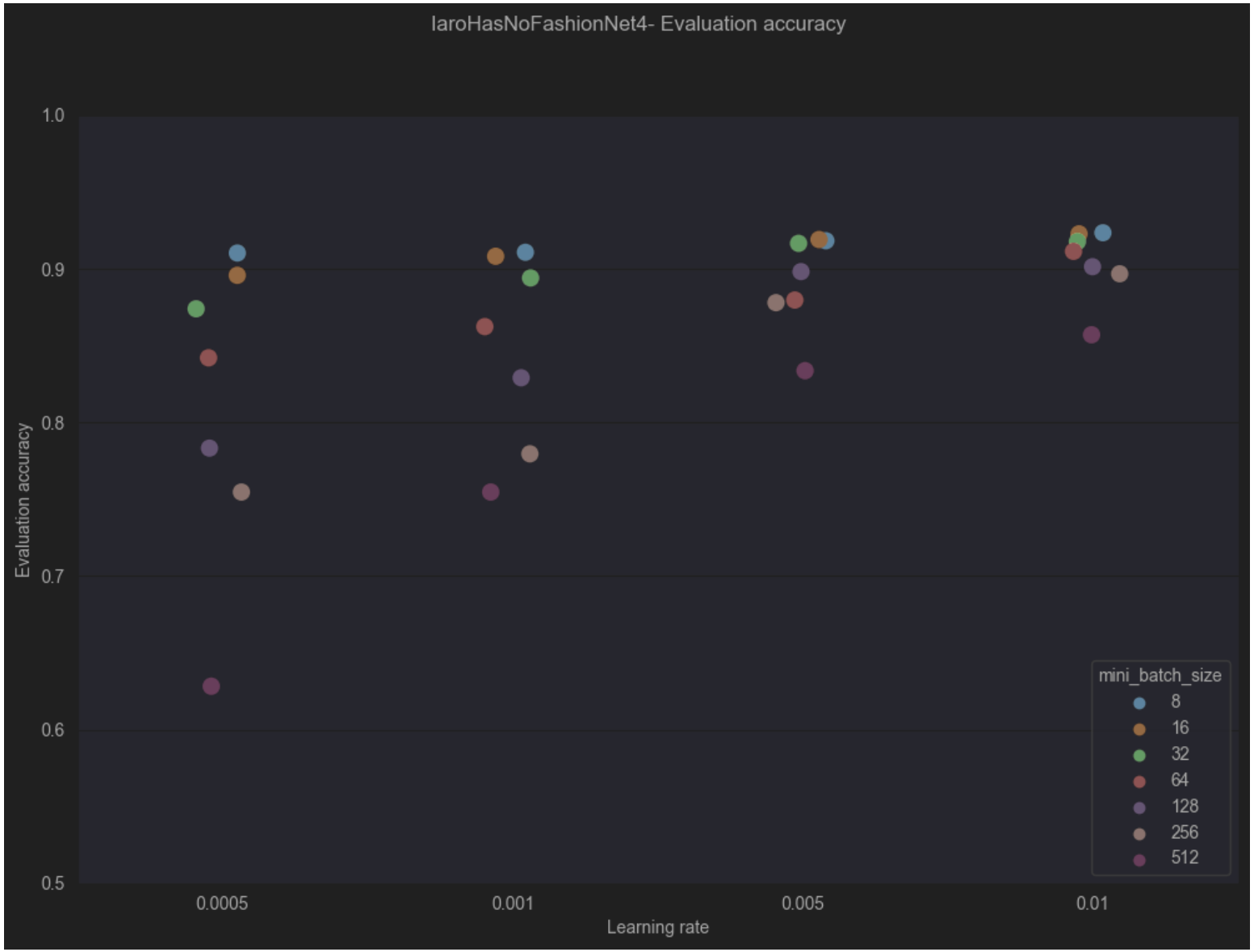
IaroHasNoFashionNet4 – Eval accuracy



Main learning:
Variance in evaluation accuracy is higher at lower learning rates. There is a tendency, that lower mini-batch sizes provide higher evaluation accuracy, however, with higher learning rate this difference is being neglected.

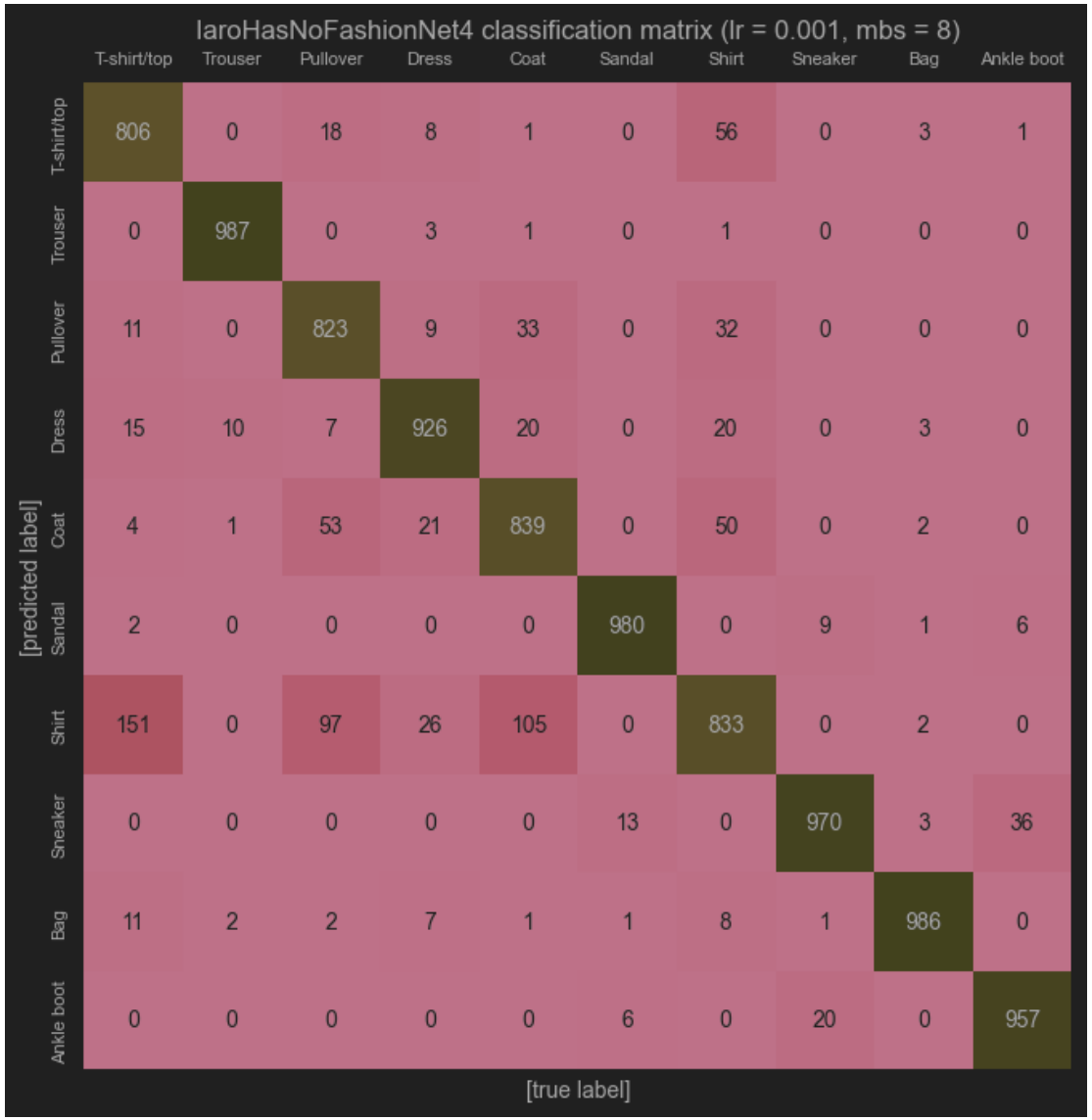
Universität St.Gallen

IaroHasNoFashionNet4 – Eval accuracy



Universität St.Gallen

IaroHasNoFashionNet4 – Conf matrix



Main learning:

The main issue is in differentiating between shirts, T-shirts, coats and pullovers. This is the main factor to account for, as there are no problems distinguishing between different types of shoes.

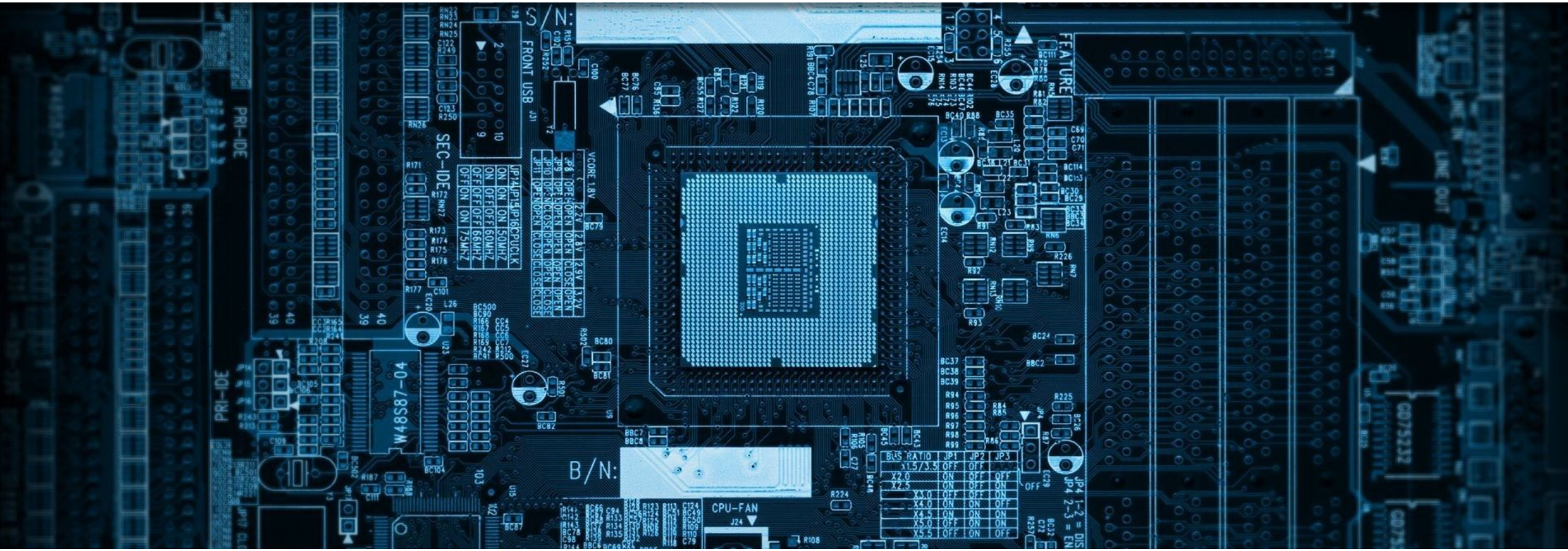
Summary

Looking into the performace of different models, I would prefer to use either the IaroHasNoFashionNet**2** or IaroHasNoFashionNet**4**. Model **2** has about 30% less parameters and requires 30% less time to be trained, when compared to model **4**. Overall accuracy is similar, however, model **4** is better when it comes to shirts, while model **2** works better with T-shirts and pullovers. If costs allow, model **4** would be better, while model **2** is a cost-saving option.

Looking in differences between training and evaluation from the perspective of different learning rates and mini-batch sizes, I would land on learining rates 0.001 and 0.005 and mini-batch sizes 64, 128 or 256. Those combinations are not prone to overfitting and balanced from the perspective of time performance and accuracy.

Additional examination of F-scores would provide additional food for thoughts, but I leave it out of scope for this work.

Universität St.Gallen



Questions?