

Deep Learning for Computer Vision

iMaterialist Challenge

Group 6

Francesca Mazzetti

Gaia Vaccarezza

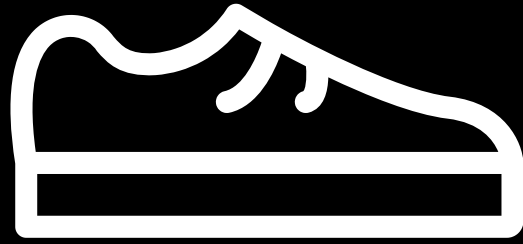
Michele Scarnera

Michelle Luijten

Agenda

- Data Description
- Multi-Task Learning
- Explanation of different models
- Results

Apparel Classes



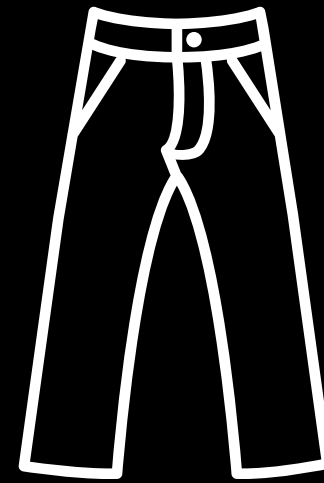
Shoe

14 tasks



Dress

11 tasks



Pants

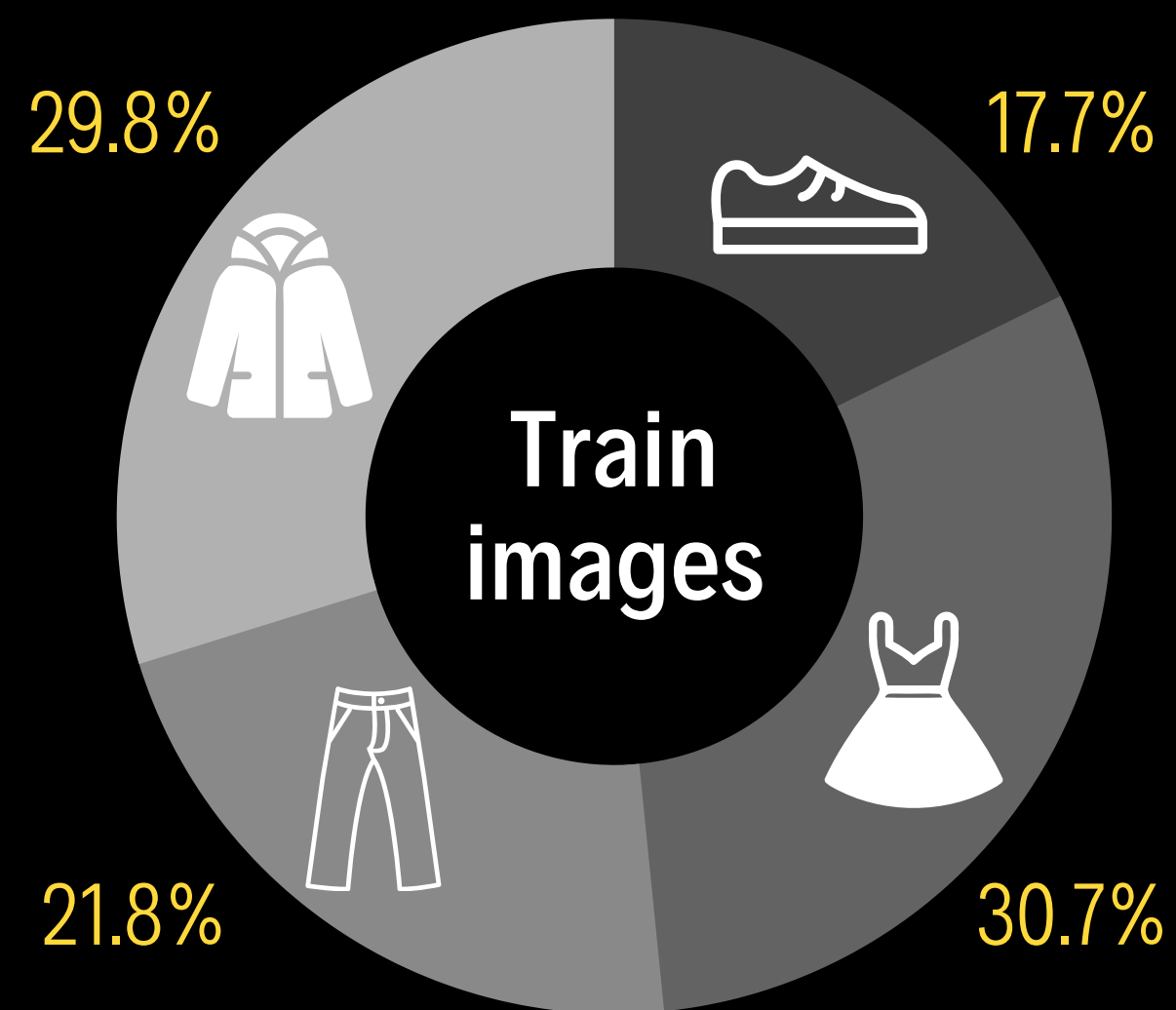
10 tasks



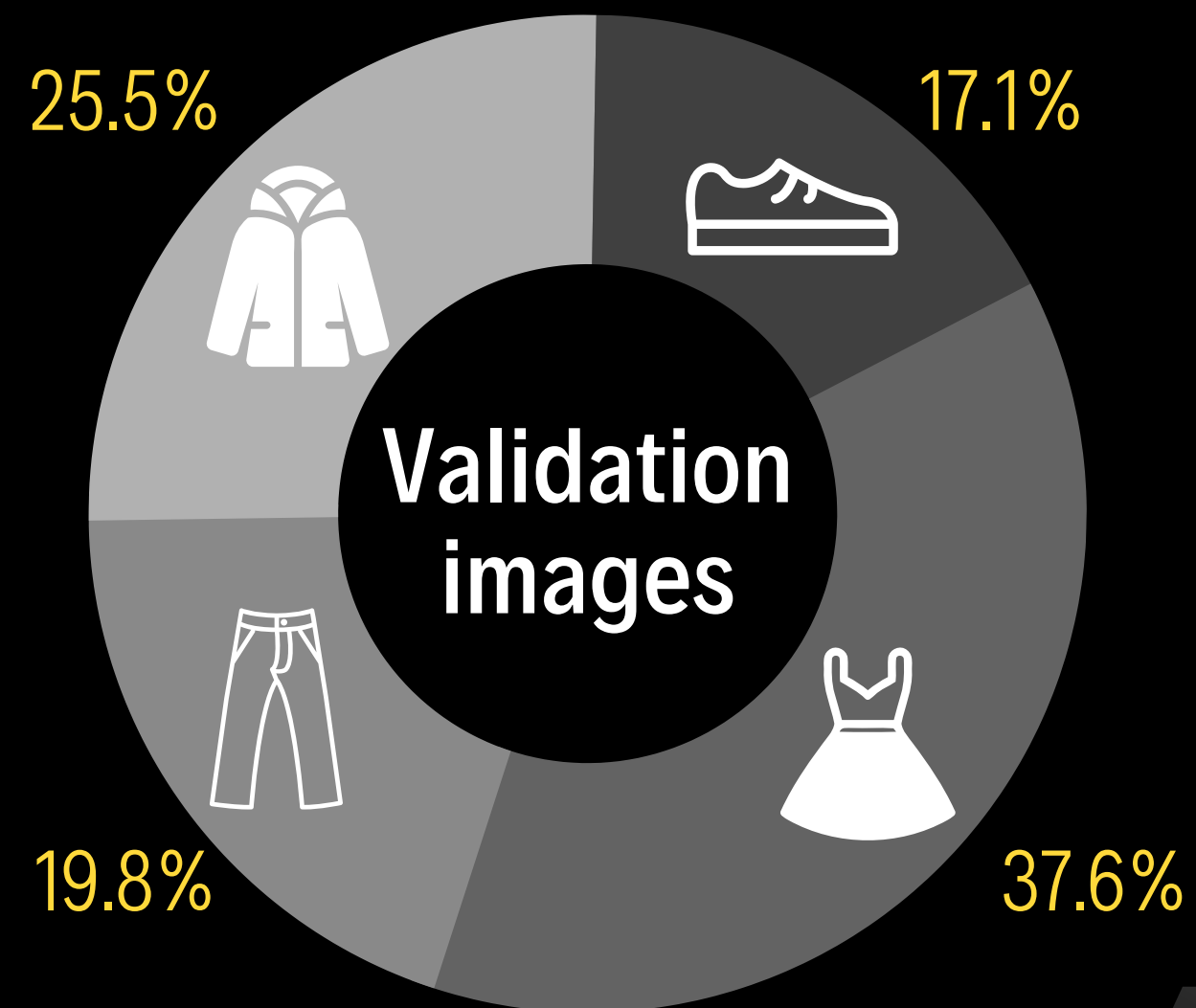
Outerwear

10 tasks

Distributions of the classes in the dataset

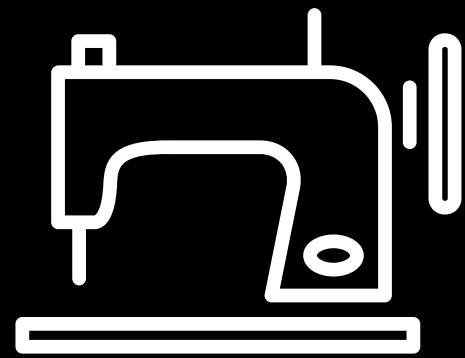


42 029 images → 24642 images



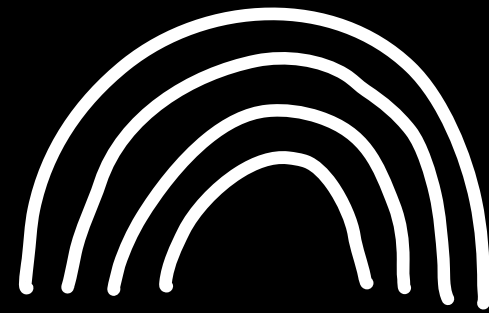
8 432 images → 3000 images

Common tasks among the apparel classes



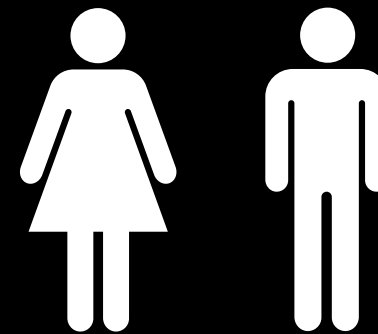
Material

70 labels



Color

68 labels



Gender

2 labels



Age

3 labels

The number of labels for *color* and *material* are reduced by grouping labels that only occurred $< 1\%$

Multi-Task Learning

Information contained in other tasks could help predicting the apparel class.

Improve learning efficiency and prediction accuracy

Share hidden layers while keeping task-specific output layers

- Reduce risk of overfitting
- Focus on features that matter
- Pattern recognition

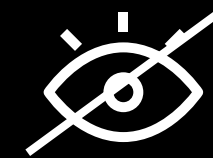
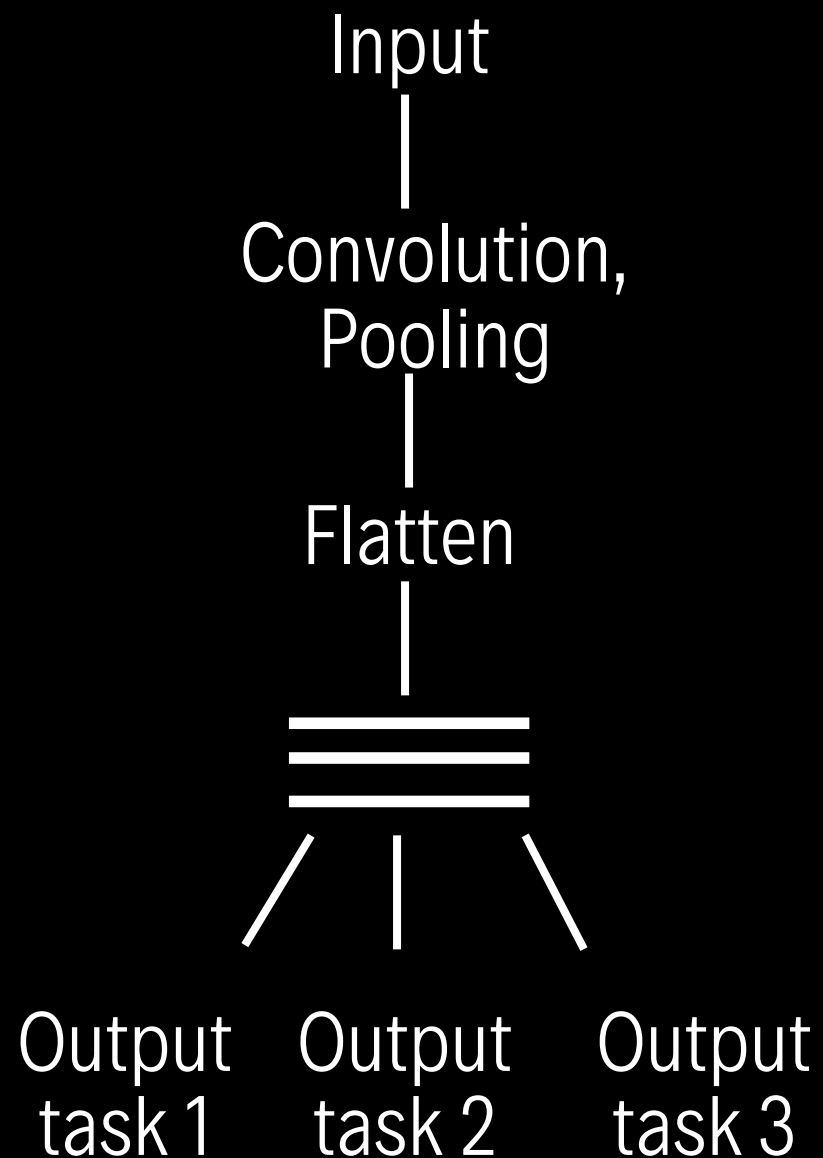


The Karate Kid

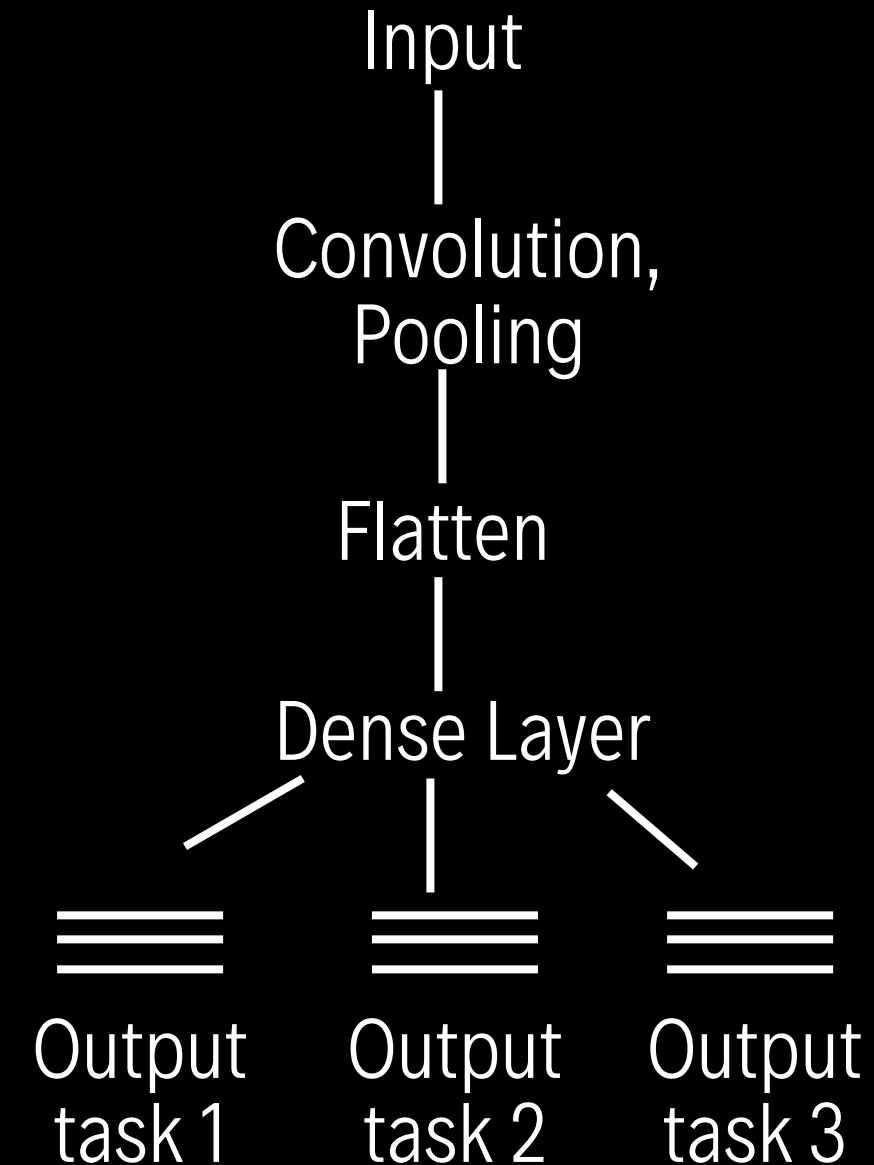
Two different models for predicting multiple tasks



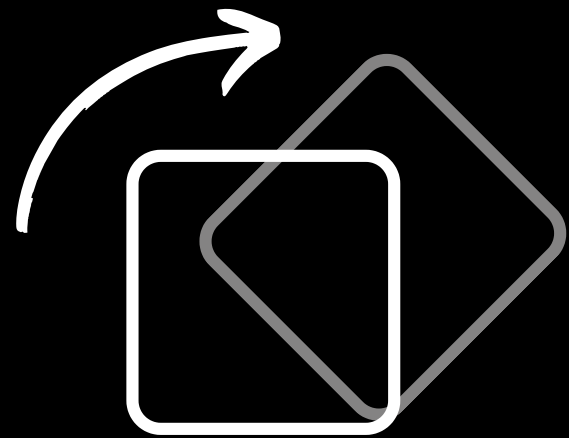
**Network with all
Public Layers**



**Network with some
Private Layers**

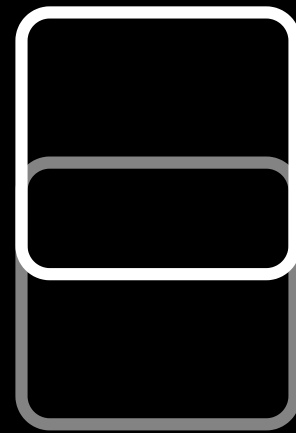


Data Preprocessing



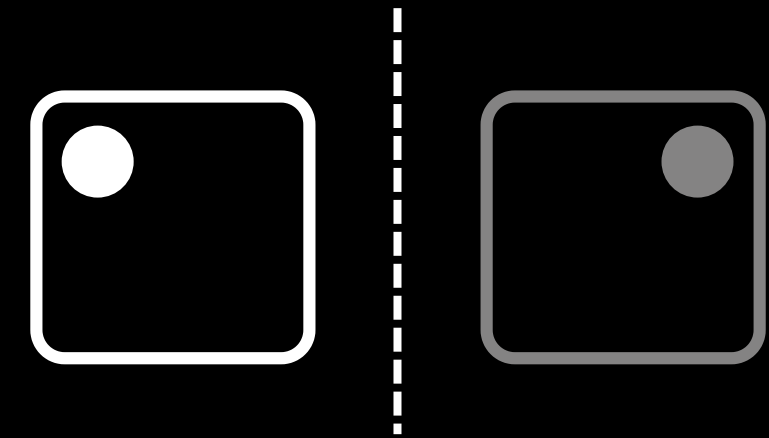
Rotation

45 degrees



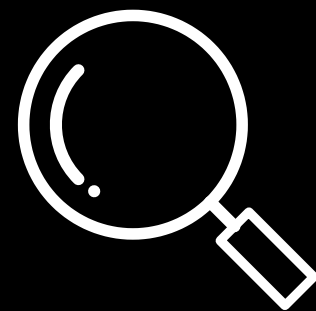
Height shift

0.25 of the size



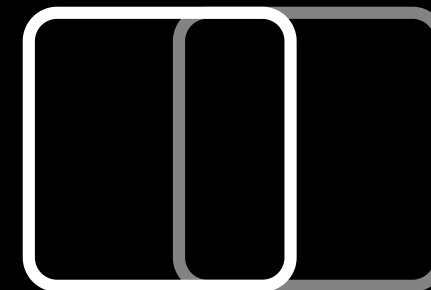
Horizontal flipping

half of the images



Zoom

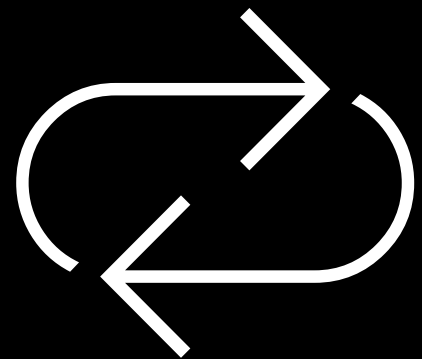
0.2



Width shift

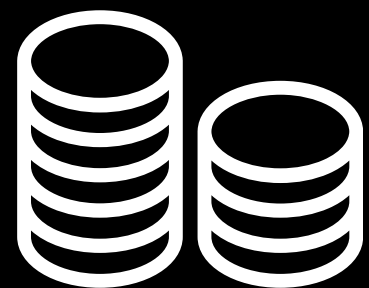
0.25 of the size

Training



Train each model 5 times

First apparel only, then apparel and one other task



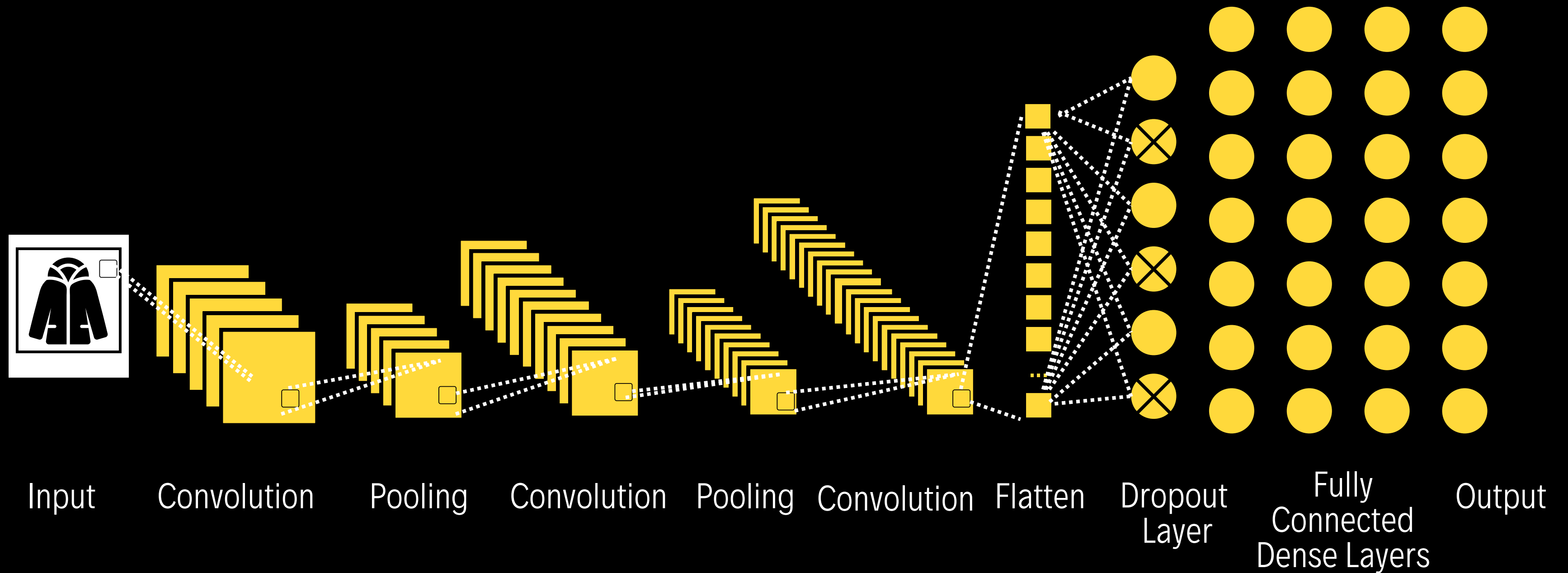
Batch size = 128, number of epochs = 20



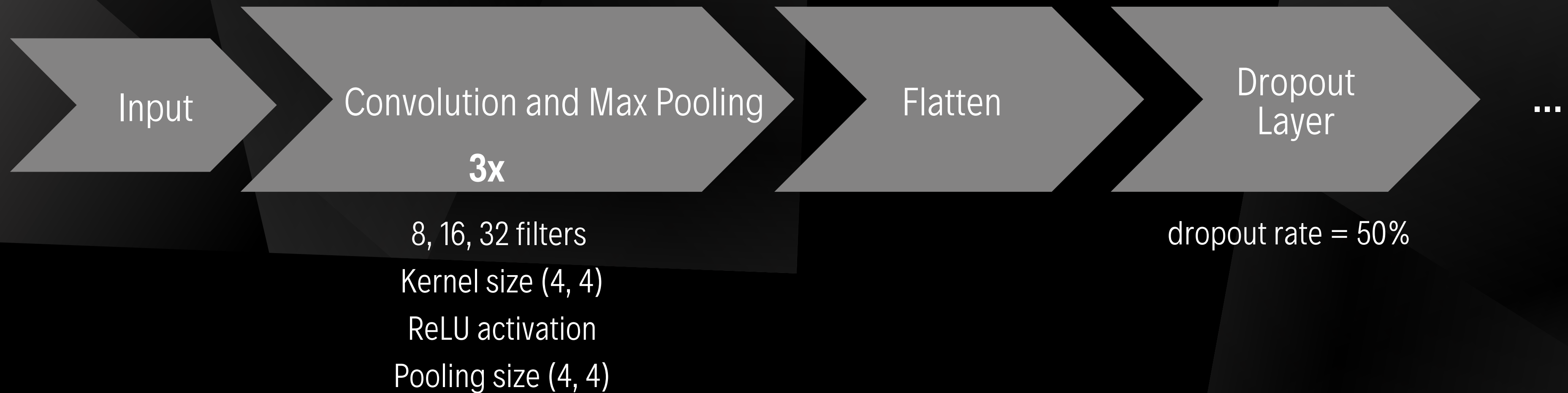
Early Stopping Method

When validation loss has a change of < 0.001 for more than $\text{int}(\max([6, 1 + \sqrt{\text{epochs}}]))$ epochs

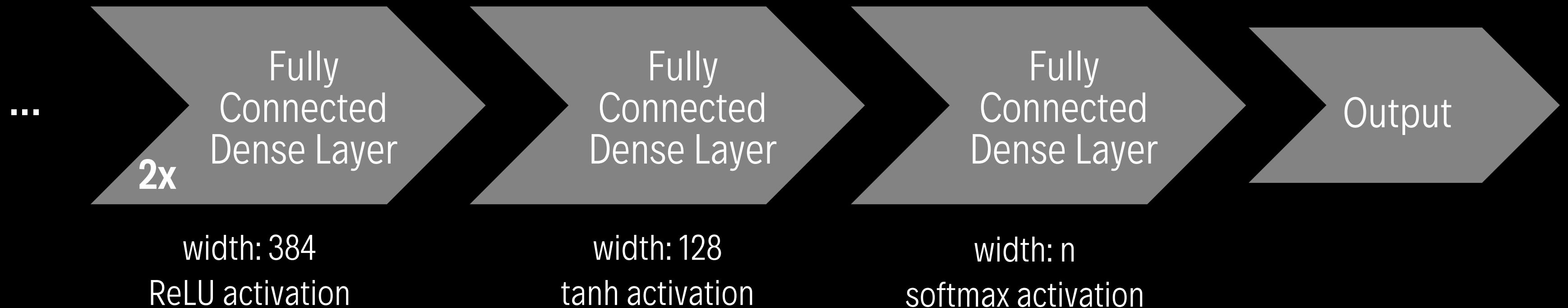
Configuration of the model with public layers



More details on the model with public layers



Model on all tasks performs best when using Public Layers and 3 million parameters



Optimizing Loss

Use of stochastic gradient descent method Adam

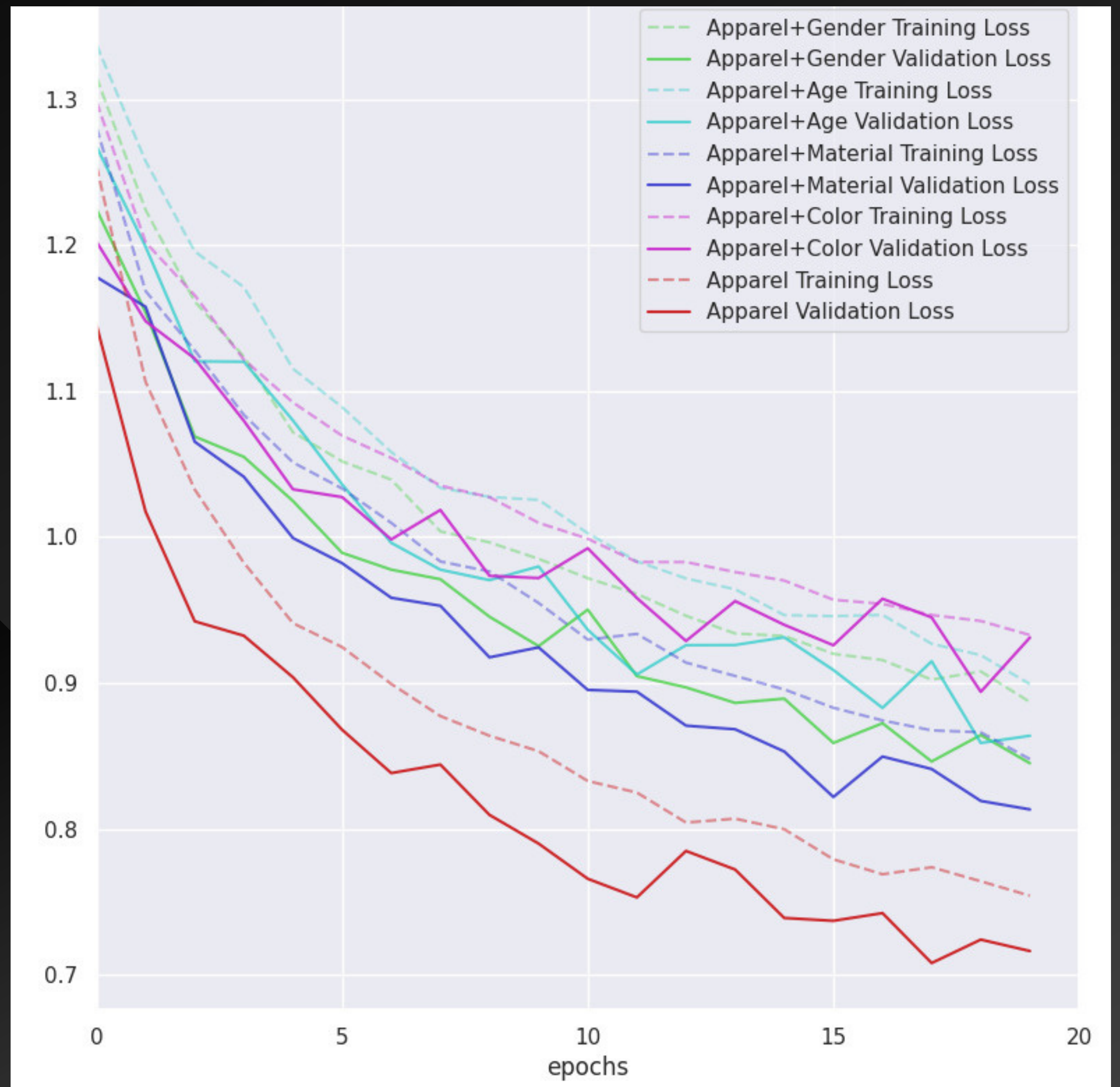
Apparel class has the highest weight, as it is marked most important

Usage of Sparse Categorical Cross Entropy function

Comparing validation loss and training loss

Validation loss of the different models

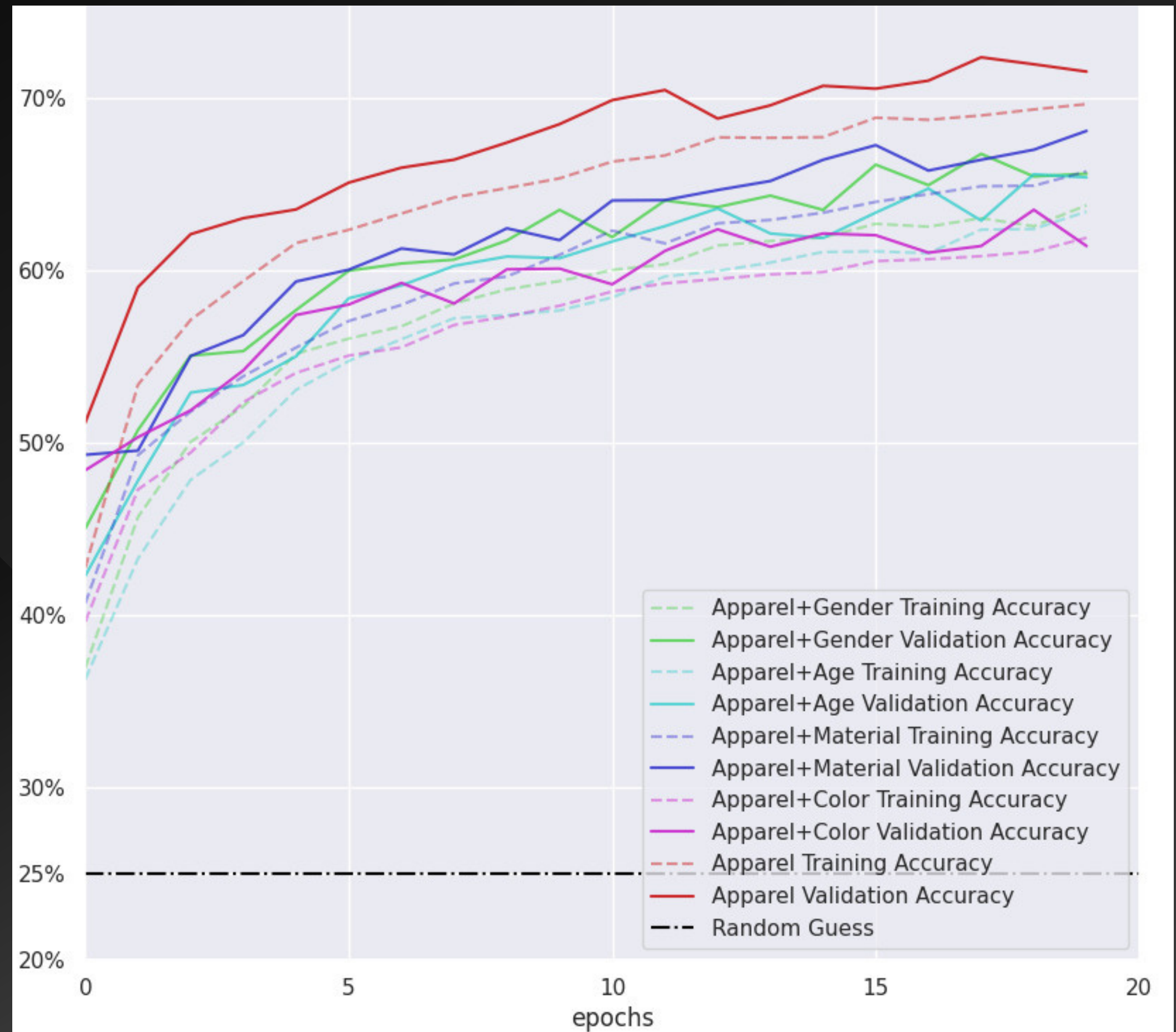
Apparel	Baseline model	0.716
Apparel and Material		0.813
Apparel and Color		0.931
Apparel and Gender		0.844
Apparel and Age		0.863



Examining validation accuracy

Validation accuracy of the different models

Apparel	Baseline model	0.715
Apparel and Material		0.680
Apparel and Color		0.614
Apparel and Gender		0.656
Apparel and Age		0.653



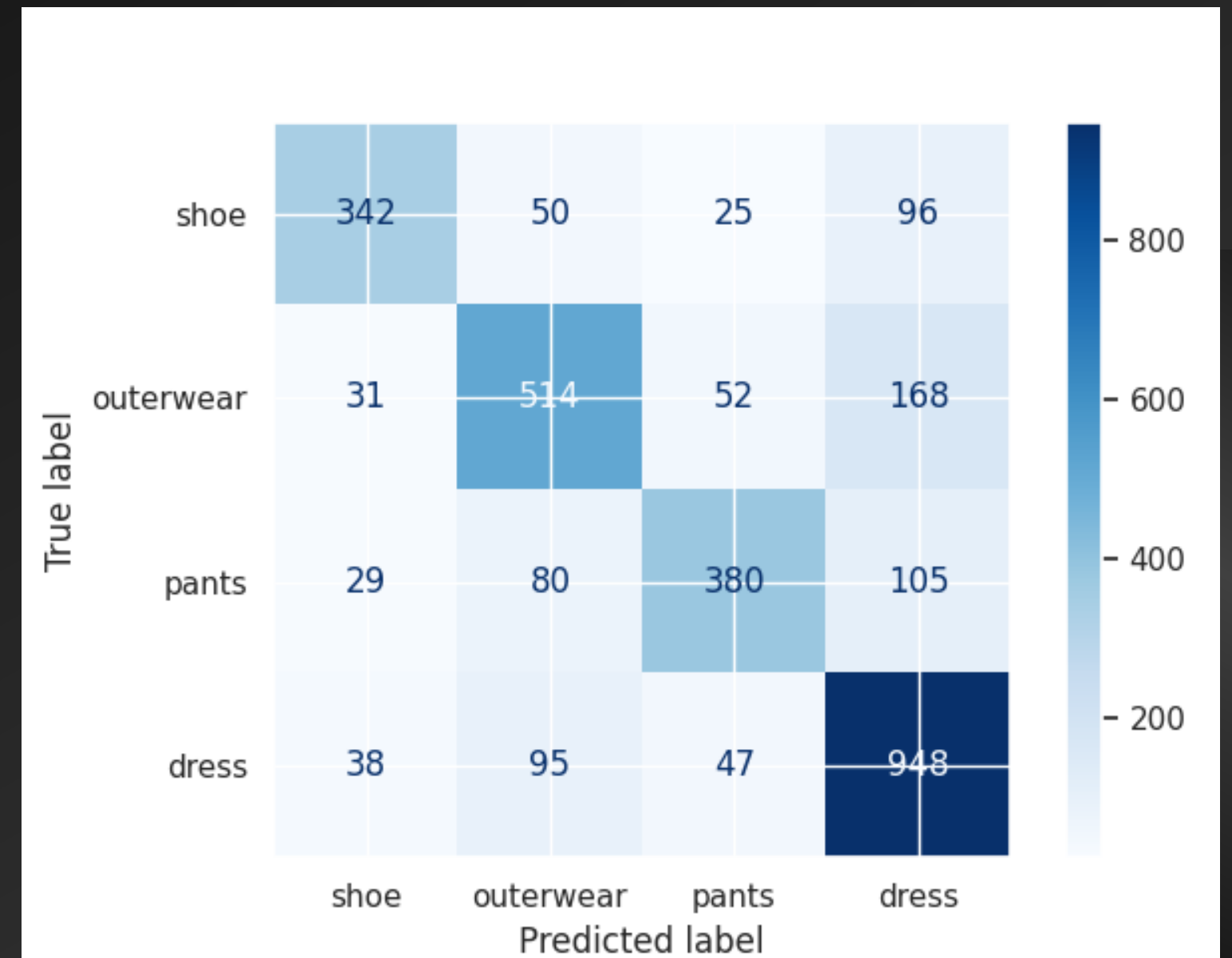
Using precision and recall measures to evaluate the model

Precision

Of all images that we predicted as dresses, how many actually dresses?


Recall

Of all the images that dress, how many did we label as dresses?



Apparel-only
Baseline model

Using F1 scores to evaluate the models

					
Precision	0.70–0.78	0.64–0.75	0.54–0.73	0.58–0.75	0.61–0.74
Recall	0.64–0.84	0.63–0.78	0.45–0.65	0.58–0.81	0.52–0.83
F1–score	0.72	0.67	0.61	0.63	0.64

The apparel-only model with 3 million parameters and public layers performs best.

Multi-task learning did not improve the goodness of fit of the model.

Future Research

Perform hyperparameter selection

Adding regularization, pruning or pre-trained models

Grouping labels using an auto-encoder

Thank you!

Are there any questions?