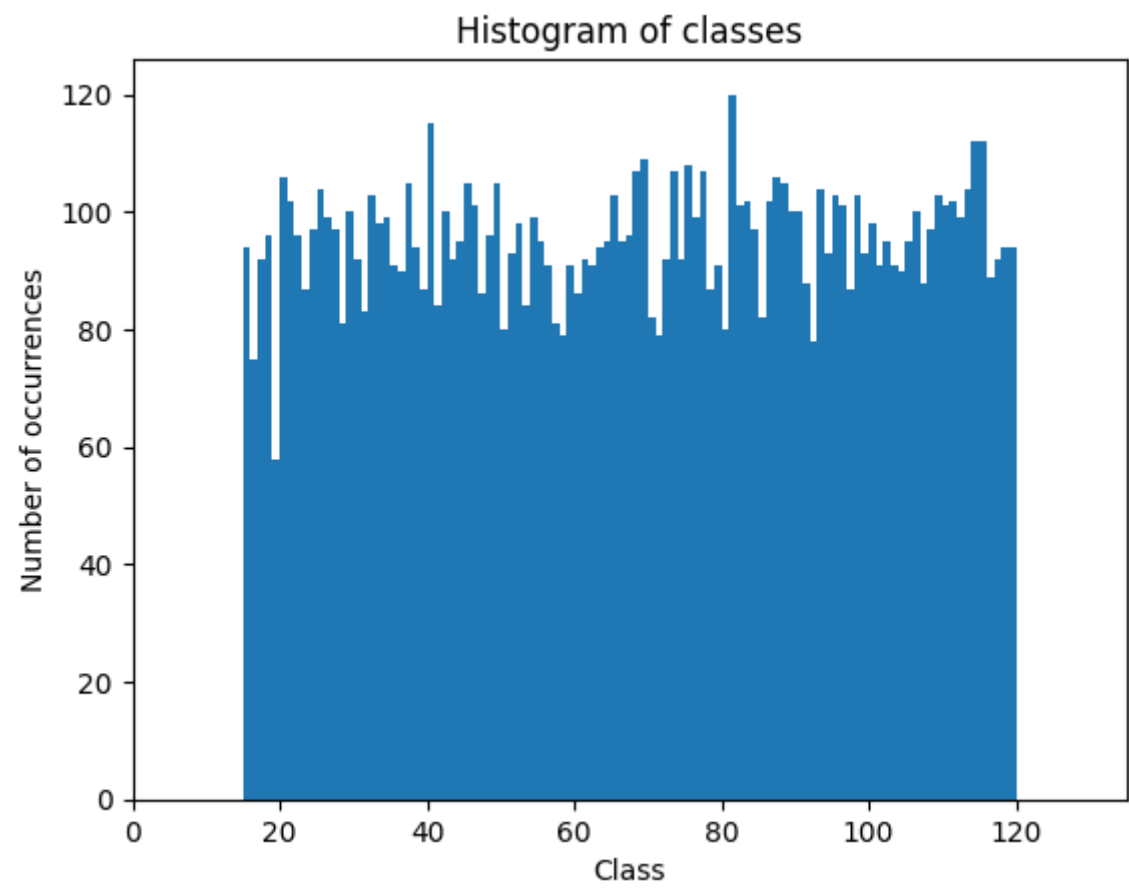# Wojciech Mierzejek **DNN Homework 1**

1. Exploratory data analysis

Figure 1 – histogram of classes

This plot shows the number of occurrences of every class (after converting labels from `data/labels.csv` to classification labels).

Figure 2 – table with regression data

	squares	circles	up	right	down	left
1	0	0	0	0	0	0
2	415	486	501	480	500	484
3	475	480	477	464	479	468
4	493	463	464	484	467	500
5	458	487	465	494	456	480
6	492	508	462	489	470	450
7	463	486	456	466	508	464
8	483	500	494	477	487	425

	squares	circles	up	right	down	left
9	0	0	0	0	0	0

From both figures we see that every image has at least 2 shapes of one kind and at most 8 of another. Thus the dataset contains samples from only 105 of the 135 theoretically possible classes.

2. Model architecture

The backbone described in the assignment is the core of the model. There are two heads: classification and regression.

Classification head

```
nn.Sequential(
    nn.Dropout(0.3),
    nn.Linear(256, 64),
    nn.SiLU(),
    nn.Dropout(0.1),
    nn.Linear(64, 135),
    nn.LogSoftmax(dim=1)
)
```

Regression head

```
nn.Sequential(
    nn.Dropout(0.3),
    nn.Linear(256, 64),
    nn.SiLU(),
    nn.Dropout(0.1),
    nn.Linear(64, 6),
)
```

The heads are similar; the differences are the output size and that the classification head applies `LogSoftmax` and returns log-probabilities. Both are minimal and lightweight.

3. Augmentations

I implemented five augmentations:

- random horizontal flip
- random vertical flip
- random 90° rotation
- random brightness and contrast
- Gaussian noise

Every augmentation has a probabillity that can be set when initializing training dataset. Function randomly choses at least one augmentation (when augmentations are enabled). I think that random brightness and contrast are the most important for generalization of the model. Flipping and rotating the images are easy ways for extending the dataset.

4. Experiment setups

All experiments use the same dataset with three augmentations (one original copy and two augmented). Patience = 15 for early stopping.

Experiment 1

Classification only, $\lambda\_cnt = 0$

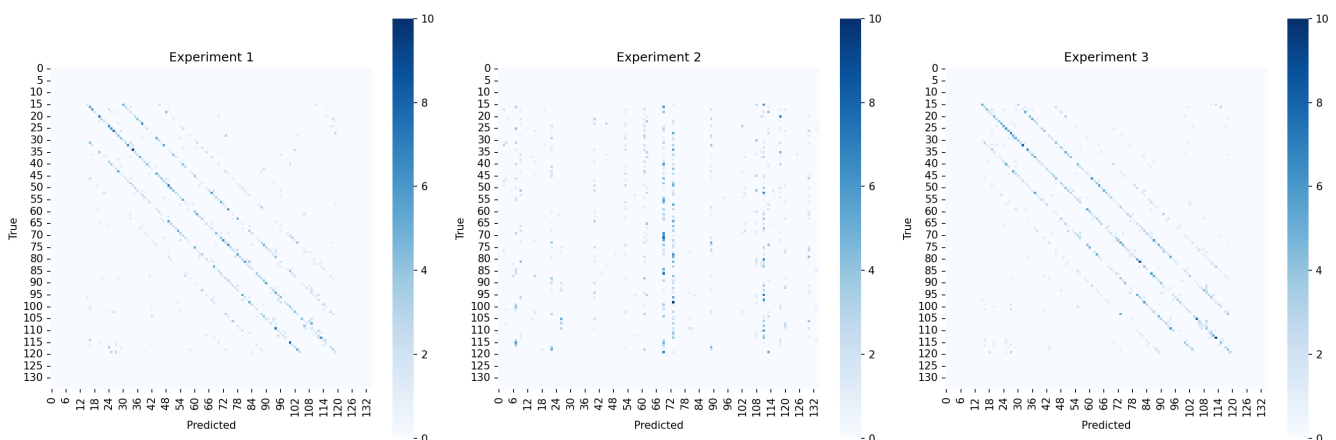Experiment 2

Regression-only, classification loss is ignored

Experiment 3

Multitask with $\lambda\_cnt = 5$. Chosen because the initial `NLLLoss` (classification) was about five times larger than the `SmoothL1Loss` (regression).

5. Final loss, accuracy, RMSE and MAE for each experiment

	test_loss	Top1_accuracy	per_pair_accuracy	RMSE_overall	MAE_overall
exp1	2.31404	0.315	0.792	2.97444	1.97195
exp2	0.397	0.008	0.044	1.09292	0.644991
exp3	4.19012	0.315	0.819	1.03876	0.626164

6. Confusion matrix heatmaps

In experiments 1 and 3, we observe stronger signals parallel to the diagonals, shifted by 15 or 30, which corresponds to good per-pair classification, but model misses the exact counts of the two selected shapes.

7. Learning curves
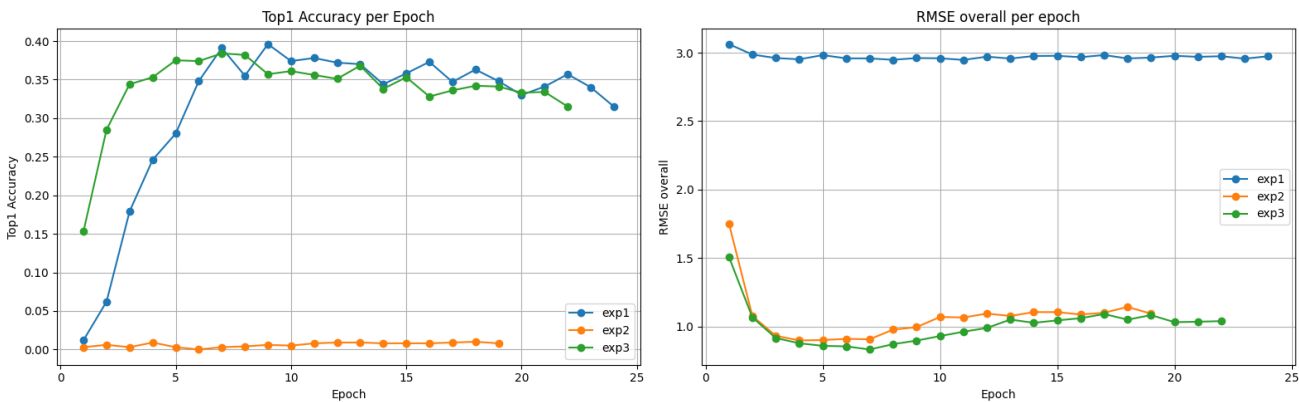
The multitask model outperforms both single-task models on both tasks. After some epochs single-task models get closer results, but still multitask dominates. This may be because both tasks answer the same question, so they also require finding the same patterns. That couses the shared backbone to benefit from combined signals, with every epoch it gets new data from both heads.