

## 1 Q1

Our multi-task CNN was trained with a 2-block shared backbone, an Adam optimizer, and a batch size of 64. We conducted five experimental runs, primarily tuning the joint loss weights  $\lambda_1$  (classification) and  $\lambda_2$  (regression).

Our key observation from `wandb` logs is that the model’s performance is **majorly dependent on the  $\lambda_1$  weight**. When  $\lambda_1 \geq \lambda_2$  (e.g., 2:1), the model excels at classification. When  $\lambda_1 < \lambda_2$  (e.g., 0.25:1), the model immediately solves the simpler regression task, as seen by a near-zero initial val/regloss.

This confirms a direct trade-off: run2lambda21 (prioritizing classification) achieved the highest accuracy, while run5lambda0.251 (prioritizing regression) achieved the lowest RMSE. Feature map visualizations confirmed the shared backbone learned low-level edges (good for regression) and high-level object parts (good for classification).

## 2 WandB Links

[Q1 Link](#)

[Q2A Link](#)

[Q2B Link](#)