

Figure 1: Cardinality distribution for top-k experts $\mathcal{K} = \{1, 2, 4, 8\}$ for the CIFAR-10 and CIFAR-100 datasets, analyzed under two different λ values. For a given dataset, increasing λ results in fewer samples with the largest cardinality of 8, and more samples with smaller cardinalities. This is because increasing λ amplifies the impact of cardinality in the cost functions. When comparing across different datasets, the distribution differs for the same λ due to the varying complexity levels of the classification tasks.



Figure 2: Comparison of hard and easy images as judged by humans in original quality on the CIFAR-10 dataset for top-*k* experts $\mathcal{K} = \{1, 2, 4, 8\}$. *Hard images* are predicted correctly by our algorithms with a cardinality of 8 but predicted incorrectly when using a cardinality of 4 instead. *Easy images* are correctly predicted by our algorithms with a cardinality of 1.