Findings

Image classifiers improve performance through an auxiliary task derived from the original labels:

- Accuracy improves for many SOTA architectures.
- Converges faster.
- Training loss and joint prediction boost performance without adding too many additional parameters.
- Predictions are more structured (hence explainable) because of the auxiliary criterion.

Methods

1. Use the original labels and group them (mutually exclusive and balanced)

   Use a confusion matrix from a pre-trained model as basis for a similarity metric.

2. Define a classifier for the original task, and an auxiliary one for the grouped labels

3. Train concurrently

   \[ L = \lambda_1 L_f + \lambda_2 L_g \]

   Context

4. Joint Prediction

   \[ f(x) \cdot g(x) \cdot (x) \]

   Softmax

   \[
   \begin{array}{cccc}
   \text{Original Labels} & \text{Group Labels} \\
   \text{chicken} & \text{cherry} & \text{orange} & \text{horse} & \text{chicken} & \text{horse} & \text{orange} & \text{cherry} \\
   \end{array}
   \]

Results

1. Higher accuracy of ResNet50 on:

   - CIFAR-100
     - Baseline: 78.9
     - SSAL (Ours): 80.6
   - Imagenet
     - Baseline: 75.5
     - SSAL (Ours): 76.9

2. Same architecture, trains faster:

   Using SSAL Loss

   Regular NLL Loss

3. Train concurrently

   \[ L = \lambda_1 L_f + \lambda_2 L_g \]

4. Ablation: more efficient use of additional parameters!

   - Baseline
     - Nr. Parameters: 11.2M
     - Val. Accuracy: 39.9%
   - Wider
     - Nr. Parameters: 25.3M
     - Val. Accuracy: 42.3%
   - Wider&Deeper
     - Nr. Parameters: 19.0M
     - Val. Accuracy: 43.7%

5. More structured (explainable) predictions thanks to the grouping objective

Future Work:

- Cluster based grouping
- Other criteria (different than grouping)
- Add SSAL to Neural Architecture Search

This work was supported by the BMBF projects ExplAIINN (Grant 01IS19074), DeFuseNN (Grant 01IW17002) and the NVIDIA AI Lab (NVAIL) program.