Summary:
Learning From Noisy Large-Scale Datasets With Minimal Supervision

Andreas Veit, Neil Alldrin, Gal Chechik, Ivan Krasin, Abhinav Gupta, Serge Belongie

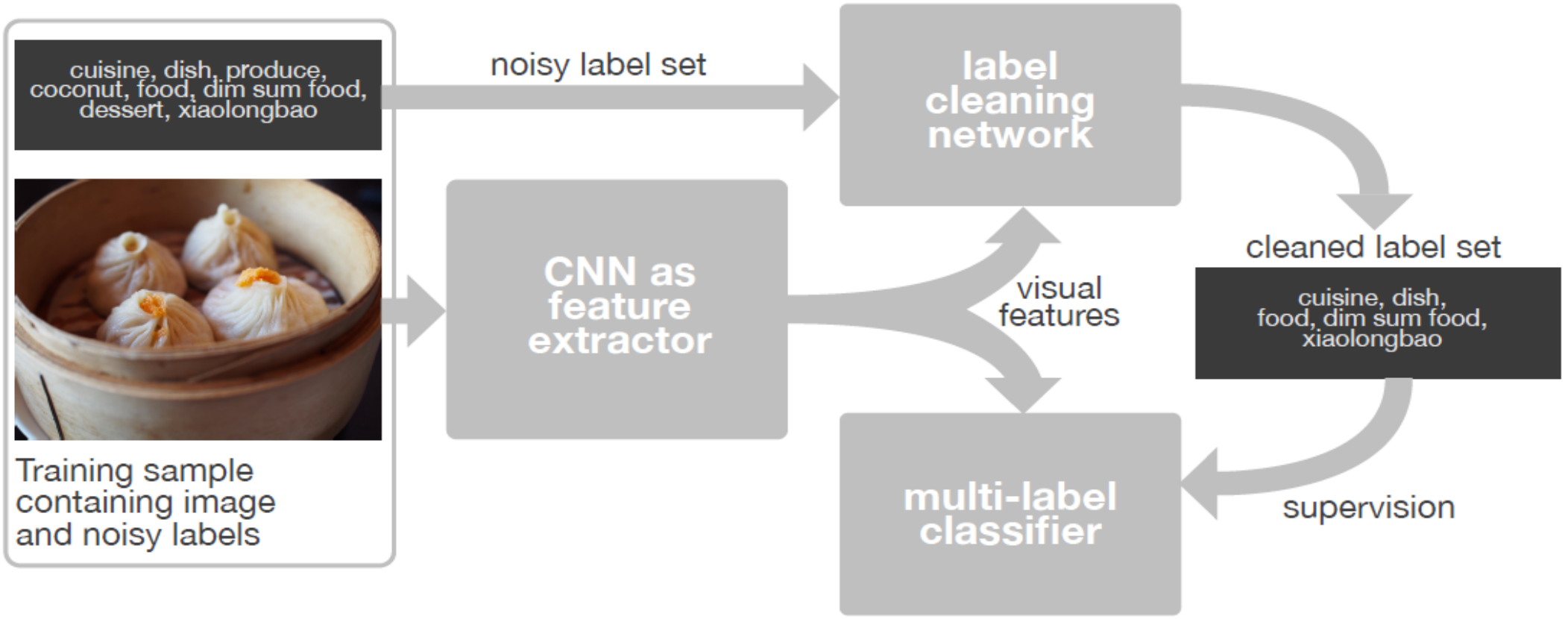
Abstract

- Use millions of images with noisy annotations
- Small subset of clean annotations
- Common approach: pre-train with noisy and fine-tune with clean
- Instead: use clean to reduce noise
- Train with both clean and full reduce-noise
- Learn to clean noisy annotations
- Open image dataset: 9M witn 6000 classes
- Clean set: 40 K

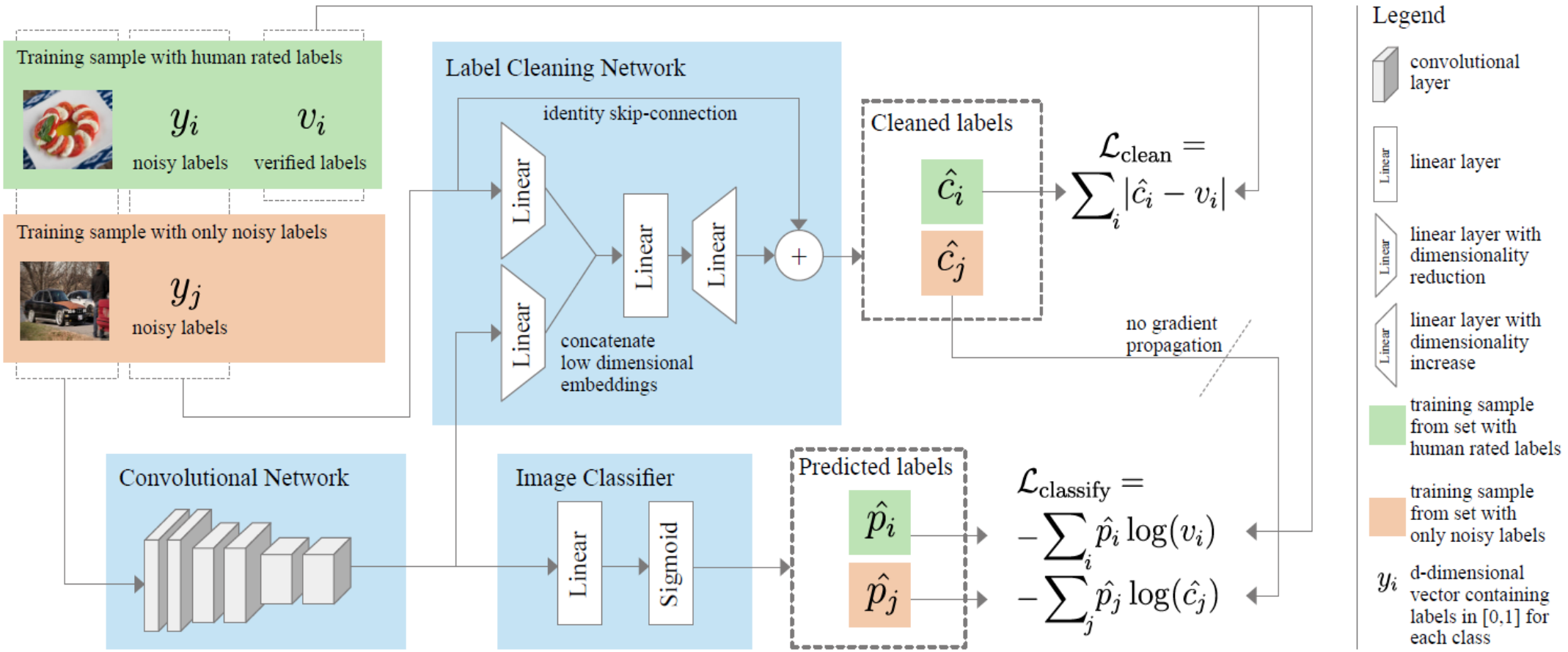
Introduction

- Bottleneck: need for massive and clean data collection
- Argument: Pre-train and then fine-tune is not efficient
- Learn a mapping between noisy and clean annotations
- Learn the pattern of noise
- Capture the structure
- Clean the noisy dataset
- Muti-task network to jointly learn to clean and classify
- Classes having 20%-80% false-positive annotations in training set

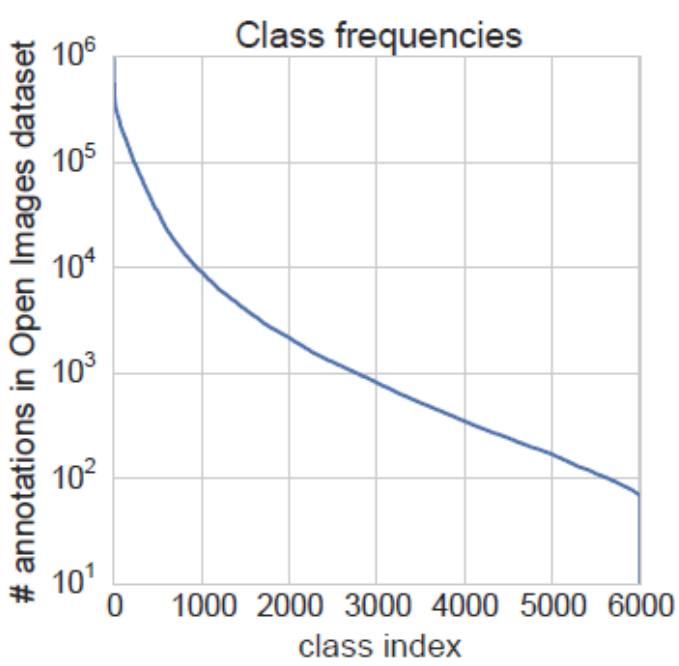
High-level Overview



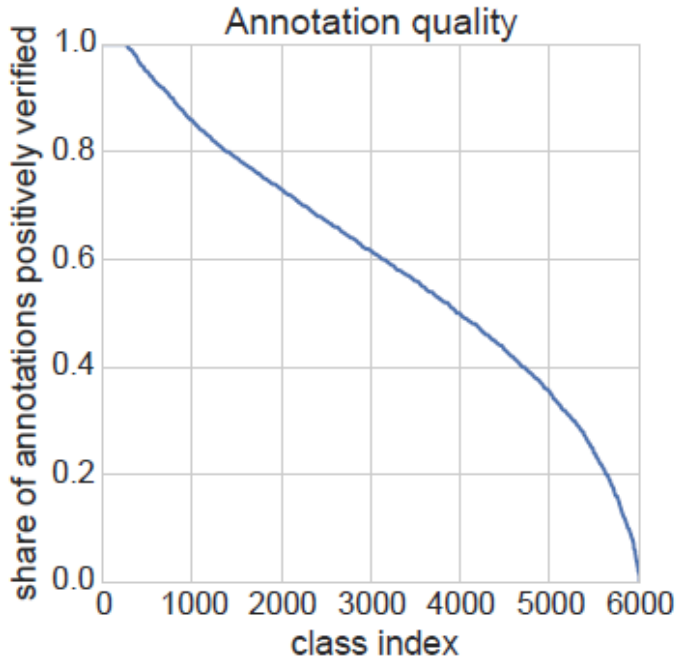
Block Diagram



Stats



(a) Class frequencies



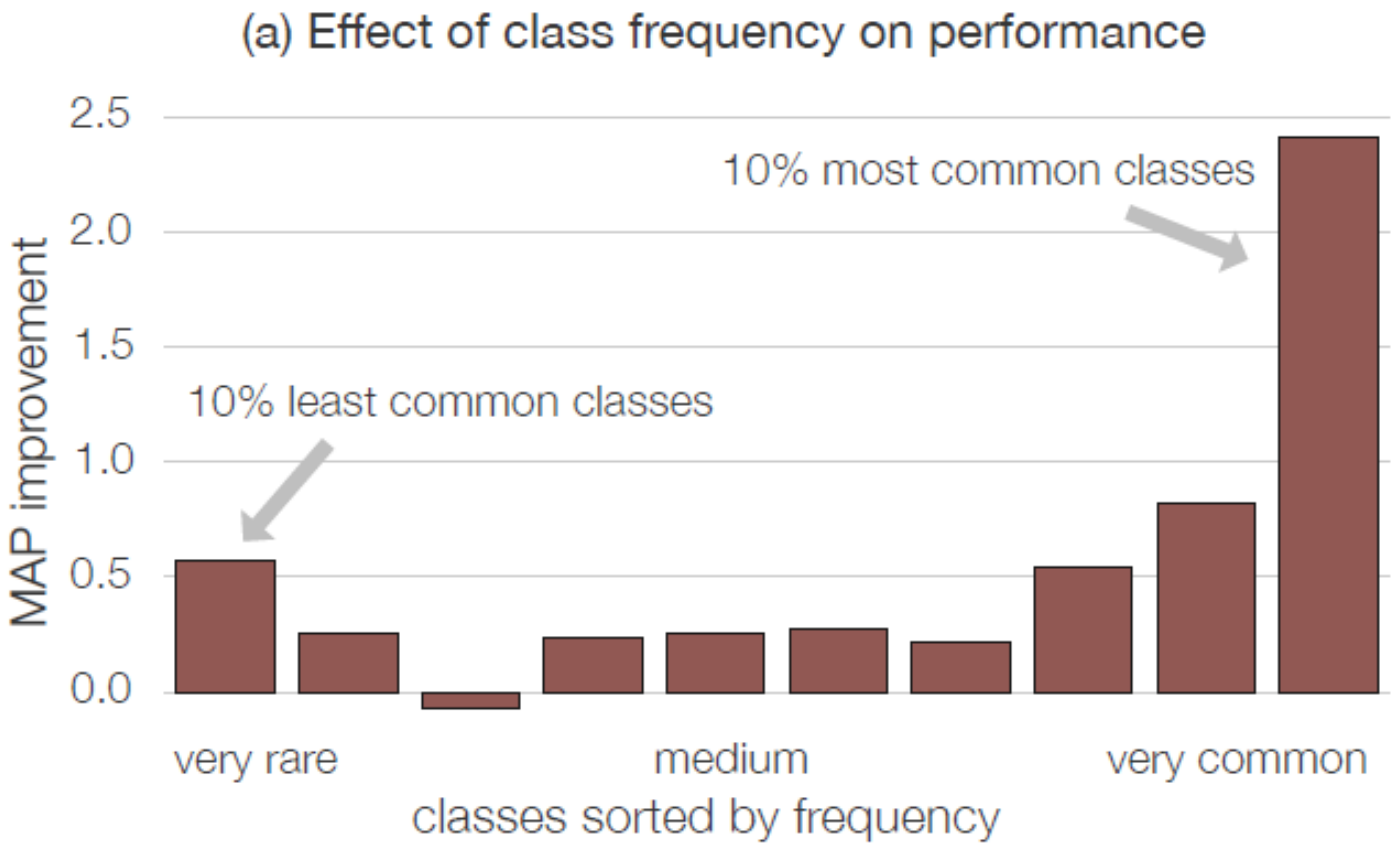
(b) Annotation quality

Results

Table 2. Comparison of models in terms of AP and MAP on the held out subset of the Open Images validation set. Our approach outperforms competing methods. See Sections 4.2 and 4.3 for more details on the metrics and model variants.

Model	AP_{all}	MAP
Baseline	83.82	61.82
Misra et al. [21] visual classifier	83.55	61.85
Misra et al. [21] relevance classifier	83.79	61.89
Fine-Tuning with mixed labels	84.80	61.90
Fine-Tuning with clean labels	85.88	61.53
Our Approach with pre-training	**87.68**	62.36
Our Approach trained jointly	87.67	**62.38**

Results


(a) Effect of class frequency on performance
MAP improvement
2.5
2.0
1.5
1.0
0.5
0.0
10% most common classes
10% least common classes
very rare
medium
very common
classes sorted by frequency

Results


(b) Effect of annotation quality on performance
MAP improvement
1.0
0.8
0.6
0.4
0.2
0.0
very noisy
medium
very clean
classes sorted by annotation quality

Summary

- The paper introduces a method for classification networks to take advantage of a large but noisy annotations to improve their accuracy. In this work, a dataset with 9M images with noisy annotations and only 40K images with cleaned annotations are used. The idea is to train a network to learn the noise pattern and then use the network to clean the noisy annotations. Then both the clean data and the reduced-noise data are used for training. The authors claim that their approach is more effective than the traditional pre-training with noisy data and then fine-tunning with the clean data. Their results supports the claim.