Self-supervised Visual Priors for Softmax Classification

Ishan Nigam
ishann@cs.utexas.edu

1. Proposal

Softmax classifiers are extremely effective at mapping visual structure to arbitrary labels [18]. However, softmax classifiers assume mutual exclusion in label space when this is clearly not the case as we move towards real-world long-tailed benchmarks [5].

Past attempts such as training independent binary cross-entropy classifiers, have not received much traction because they are (counter-intuitively) less effective even for multi-label classification problems [11].

We propose to reconcile top-down knowledge with bottom-up visual structure for interpretable visual recognition. We hope to supplement the softmax classifier with a self-supervised visual prior. Visual recognition should be treated as a multi-label classification problem. There is some indirect evidence to support this [1]. Apart from 1-hot encoded softmax objectives, we additionally view recognition as a structured output prediction problem in the label space based on the visual structure of the object.For example, atoy deer could be 0.5 toy and 0.5 deer, and a white horse could be 0.8 horse and 0.2 zebra.

2. Related Work

- Label aware margin distribution loss: [2].
- Knowledge distillation, label smoothening, subclass distillation: [7], [13], [12].
- Decoupling representation and classification: [9].
- The devil is in the tails: [15].
- Effective number of samples: [4].
- Label Refinery [1].

3. Technical Details

Notion of inter-class similarity Perform instance-aware self-supervised embedding learning. Cast embeddings onto unit hypersphere. Cosine distance of centroid embeddings is the notion of inter-class similarity.

Self-supervised Learning Self-supervised instance aware embedding learning has received quite a lot of attention in recent times [16, 19, 6, 3]. We intend to work with [19].

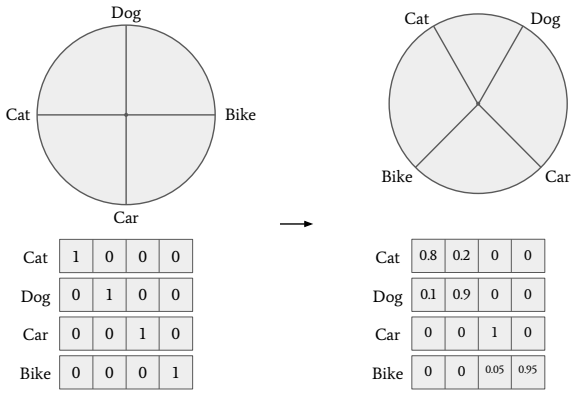

Figure 1. We propose to transform the categorical discrete softmax classifier (left) into a self-supervised instance embedding learning based probability distribution (right). Our proposed method can be interpreted as a prior on the cross-entropy objective function.

Benchmarks Imbalanced CIFAR-100 [10], iNaturalist [14].

Nearest Neighbors [17, 2, 9]

4. Expected Plan

We do not expect our proposed method to outperform the classical softmax classifier on traditional balanced benchmarks such as CIFAR-10 or ImageNet-1k. Our initial experiments indicate that a several hundred examples per class are usually sufficient to learn an effective representation. However, we do expect to see improvements on balanced benchmarks in terms of computational speedups, since we expect our proposed algorithm to converge faster.

However, we expect that our proposed method should be able to outperform the softmax classifier with imbalanced long-tailed distributions. We expect that the tail classes would be able to borrow information from the head classes.

The evaluation metric *precisely* measures softmax classification. Hence, we do not expect our method to replace softmax classification. Instead, we expect it to supplement softmax classification. Our proposed method will be more general at train time, and more interpretable at test time.

5. Mid-Term Progress Update

The following is a brief summary of the progress:

- I have written an imbalanced data loader for CIFAR-100 with a controllable parameter for the degree of data imbalance.
- I was trying to implement Momentum Contrast [6] for unsupervised pre-training to obtain self-supervised visual priors. This was quite a lot of work, and I am in the process of adapting the officially released code [8] (released a few weeks ago).
- Once I am able to obtain distributions over the classes for inter-class similarity, I am hoping to train an auxiliary BCE loss alongwith the regular cross-entropy loss for Resnet-18 classification.

6. Revised Timeline

- Unfortunately, I have not made as much progress as I anticipated. I expect that the officially released Momentum Contrast codebase will allow me to focus on incorporating the self-supervised visual prior with the standard cross-entropy loss.
- My aim is to have a working implementation on CIFAR-100 by the end of the month.
- If the results are encouraging, we can try to scale up the effort on a more rigorous benchmark such as ImageNet.

References

[1] Hessam Bagherinezhad, Maxwell Horton, Mohammad Rastegari, and Ali Farhadi. Label refinery: Improving imagenet classification through label progression. *arXiv preprint arXiv:1805.02641*, 2018. 1

[2] Kaidi Cao, Colin Wei, Adrien Gaidon, Nikos Arechiga, and Tengyu Ma. Learning imbalanced datasets with label-distribution-aware margin loss. In *Advances in Neural Information Processing Systems*, pages 1565–1576, 2019. 1

[3] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. *arXiv preprint arXiv:2002.05709*, 2020. 1

[4] Yin Cui, Menglin Jia, Tsung-Yi Lin, Yang Song, and Serge Belongie. Class-balanced loss based on effective number of samples. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 9268–9277, 2019. 1

[5] Agrim Gupta, Piotr Dollar, and Ross Girshick. Lvis: A dataset for large vocabulary instance segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5356–5364, 2019. 1

[6] Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. Momentum contrast for unsupervised visual representation learning. *arXiv preprint arXiv:1911.05722*, 2019. 1, 2

[7] Geoffrey Hinton, Oriol Vinyals, and Jeff Dean. Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531*, 2015. 1

[8] Facebook Inc. MoCo: Momentum Contrast for Unsupervised Visual Representation Learning, Apr. 2020. 2

[9] Bingyi Kang, Saining Xie, Marcus Rohrbach, Zhicheng Yan, Albert Gordo, Jiashi Feng, and Yannis Kalantidis. Decoupling representation and classifier for long-tailed recognition. In *International Conference on Learning Representations*, 2020. 1

[10] Alex Krizhevsky, Geoffrey Hinton, et al. Learning multiple layers of features from tiny images. 2009. 1

[11] Dhruv Mahajan, Ross Girshick, Vignesh Ramanathan, Kaiming He, Manohar Paluri, Yixuan Li, Ashwin Bharambe, and Laurens van der Maaten. Exploring the limits of weakly supervised pretraining. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 181–196, 2018. 1

[12] Rafael Müller, Simon Kornblith, and Geoffrey Hinton. Subclass distillation. *arXiv preprint arXiv:2002.03936*, 2020. 1

[13] Rafael Müller, Simon Kornblith, and Geoffrey E Hinton. When does label smoothing help? In *Advances in Neural Information Processing Systems*, pages 4696–4705, 2019. 1

[14] Grant Van Horn, Oisin Mac Aodha, Yang Song, Yin Cui, Chen Sun, Alex Shepard, Hartwig Adam, Pietro Perona, and Serge Belongie. The inaturalist species classification and detection dataset. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 8769–8778, 2018. 1

[15] Grant Van Horn and Pietro Perona. The devil is in the tails: Fine-grained classification in the wild. *arXiv preprint arXiv:1709.01450*, 2017. 1

[16] Zhirong Wu, Yuanjun Xiong, Stella Yu, and Dahua Lin. Unsupervised feature learning via non-parametric instance-level discrimination. *arXiv preprint arXiv:1805.01978*, 2018. 1

[17] Xueting Yan, Ishan Misra, Abhinav Gupta, Deepti Ghadiyaram, and Dhruv Mahajan. Clusterfit: Improving generalization of visual representations. *arXiv preprint arXiv:1912.03330*, 2019. 1

[18] Chiyuan Zhang, Samy Bengio, Moritz Hardt, Benjamin Recht, and Oriol Vinyals. Understanding deep learning requires rethinking generalization. *arXiv preprint arXiv:1611.03530*, 2016. 1

[19] Chengxu Zhuang, Alex Lin Zhai, and Daniel Yamins. Local aggregation for unsupervised learning of visual embeddings. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 6002–6012, 2019. 1