Self-Supervised Representation Learning by Rotation Feature Decoupling Supplementary Material

Zeyu Feng Chang Xu Dacheng Tao
UBTECH Sydney AI Centre, School of Computer Science, FEIT,
University of Sydney, Darlington, NSW 2008, Australia
zfen2406@uni.sydney.edu.au, {c.xu, dacheng.tao}@sydney.edu.au

A. Illustration of positive unlabeled learning

Figure 1 illustrates the positive unlabeled learning of image rotational ambiguity. For the task of predicting image rotation, the original images in dataset are labeled as 0 degree. Here we regard all these raw images in the dataset as being in default orientation. They are labeled as positive examples. The remaining data for predicting image rotation include all rotated copies. For some orientation ambiguous images, their rotated copies still look like being in the default orientation after being rotated. We regard these rotated copies as unlabeled examples. If we label unlabeled images all as negative examples, orientation ambiguous images will have noisy labels. We propose to weight each rotated image and reduce the relative loss of rotation ambiguous images in the unlabeled set.

B. ImageNet classification with a non-linear classifier

Following Noroozi & Favaro [7] and Gidaris *et al*. [5], we perform 1000-way ImageNet classification using self-supervised pre-trained network with weights from `conv1` up to certain layers being fixed. The rest of the network is retrained from scratch. In comparison to the previous experiment of linear classification on activations, this experiment is equal to training a non-linear classifier on top of fixed features (For `conv5`, it is a three-layer MLP, and for `conv4` is one convolutional layer plus an MLP). We train the non-linear classifier with batch normalization after convolutional and fully-connnected layers and use the open source protocol provided by Gidaris *et al*. [5]. We report top-1 classification accuracy on ILSVRC 2012 validation set using single crop.

In Table 1 we report the results of our approach and we compare it with other self-supervised learning methods. Our approach achieves significant improvement on both `conv4` and `conv5` layers (6.2 and 8.2 percentage points, respectively.)

Method\Layer	**conv4**	**conv5**
ImageNet-labels [6, 1]	59.7	59.7
Random [7]*	27.1	12.0
Doersch *et al*. (Context) [3]*	45.6	30.4
Noroozi & Favaro (Jigsaw) [7]*	45.3	34.6
Zhang *et al*. (Colorization) [10]	40.7	35.2
Donahue *et al*. (BiGANs) [4]	41.9	32.2
Bojanowski & Joulin (NAT) [1]	–	36.0
Noroozi *et al*. (Counting) [8]*	43.3	32.9
Gidaris *et al*. (RotNet) [5]	<u>50.0</u>	43.8
Noroozi *et al*. (CC+) [9]*	47.6	41.1
Noroozi *et al*. (CC+vgg-) [9]*	49.5	43.9
Caron *et al*. (DeepCluster) [2]	–	<u>44.0</u>
Ours	**56.2**	**52.2**

Table 1: Top-1 classification accuracies on ImageNet validation set using different pre-trained networks fixed until certain layers. * indicates results reported using ten-crop average.

C. Per class performance of object detection

Table 2 summarizes the per class detection performance on PASCAL VOC 2007 measured in average precision metric (mAP) as usual. The results of the ImageNet-labels entry come from Doersch *et al*. [3]. We observe that our approach substantially improves over the RotNet method and narrows the gap between self-supervised learned features and supervised learned features.

References

[1] Piotr Bojanowski and Armand Joulin. Unsupervised learning by predicting noise. In Doina Precup and Yee Whye Teh, editors, *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pages 517–526, International Convention Centre, Sydney, Australia, 06–11 Aug 2017. PMLR.

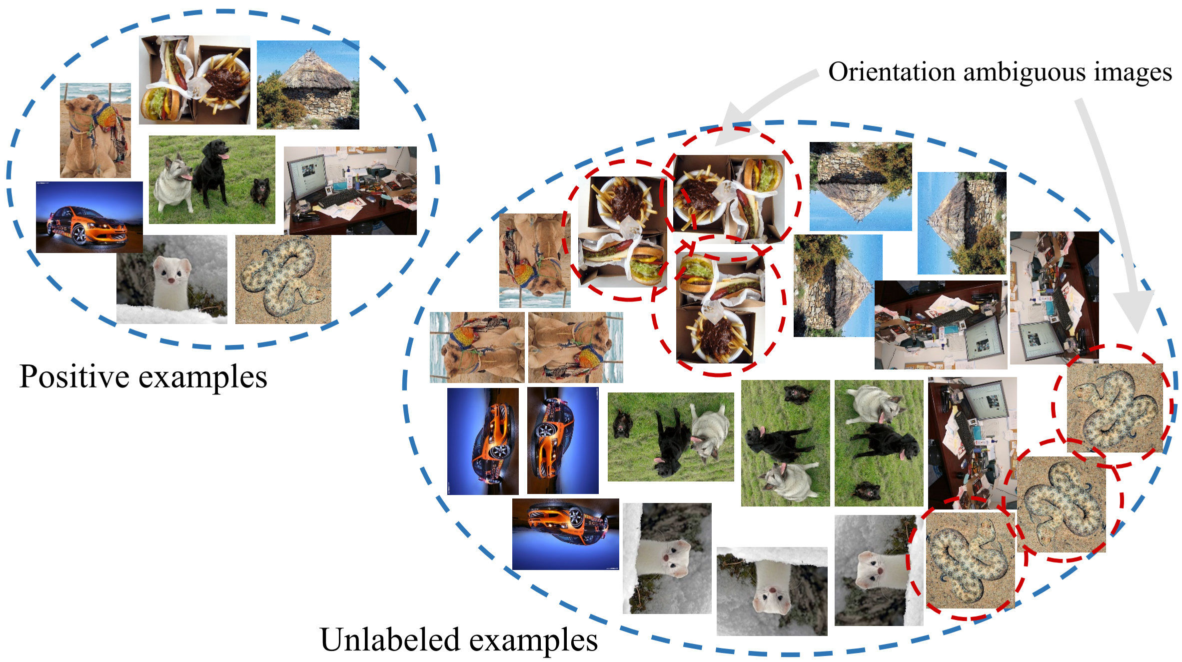

Figure 1: Positive unlabeled learning formulation of predicting image rotations.

Method\Classes	aero	bike	bird	boat	bottle	bus	car	cat	chair	cow	table	dog	horse	mbike	person	plant	sheep	sofa	train	tv
ImageNet-labels [6, 3]	64.0	69.6	53.2	44.4	24.9	65.7	69.6	69.2	28.9	63.6	62.8	63.9	73.3	64.6	55.8	25.7	50.5	55.4	69.3	56.4
Gidaris *et al*. (RotNet) [5]	65.5	65.3	43.8	39.8	20.2	65.4	69.2	63.9	30.2	56.3	62.3	56.8	71.6	67.2	56.3	22.7	45.6	59.5	71.6	55.3
Ours	68.3	70.8	50.8	40.8	25.8	71.4	70.3	68.0	32.1	58.2	61.5	61.7	73.8	69.3	57.9	28.6	50.4	59.2	73.2	58.9

Table 2: Per class performance on PASCAL VOC 2007 detection.

[2] Mathilde Caron, Piotr Bojanowski, Armand Joulin, and Matthijs Douze. Deep clustering for unsupervised learning of visual features. In Vittorio Ferrari, Martial Hebert, Cristian Sminchisescu, and Yair Weiss, editors, *Computer Vision – ECCV 2018*, pages 139–156, Cham, 2018. Springer International Publishing.

[3] Carl Doersch, Abhinav Gupta, and Alexei A. Efros. Unsupervised visual representation learning by context prediction. In *The IEEE International Conference on Computer Vision (ICCV)*, December 2015.

[4] Jeff Donahue, Philipp Krähenbühl, and Trevor Darrell. Adversarial feature learning. In *International Conference on Learning Representations*, 2017.

[5] Spyros Gidaris, Praveer Singh, and Nikos Komodakis. Unsupervised representation learning by predicting image rotations. In *International Conference on Learning Representations*, 2018.

[6] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 25*, pages 1097–1105. Curran Associates, Inc., 2012.

[7] Mehdi Noroozi and Paolo Favaro. Unsupervised learning of visual representations by solving jigsaw puzzles. In Bastian Leibe, Jiri Matas, Nicu Sebe, and Max Welling, editors, *Computer Vision – ECCV 2016*, pages 69–84, Cham, 2016. Springer International Publishing.

[8] Mehdi Noroozi, Hamed Pirsiavash, and Paolo Favaro. Representation learning by learning to count. In *The IEEE International Conference on Computer Vision (ICCV)*, Oct 2017.

[9] Mehdi Noroozi, Ananth Vinjimoor, Paolo Favaro, and Hamed Pirsiavash. Boosting self-supervised learning via knowledge transfer. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.

[10] Richard Zhang, Phillip Isola, and Alexei A. Efros. Colorful image colorization. In Bastian Leibe, Jiri Matas, Nicu Sebe, and Max Welling, editors, *Computer Vision – ECCV 2016*, pages 649–666, Cham, 2016. Springer International Publishing.