

A. Qualitative Results

In Fig. 1, we visualize the feature maps from ‘layer4’ of ResNet-18 pre-trained by PCD and vectorized variant of PCD, respectively. The feature maps of ResNet-18 pre-trained by PCD show clear outlines of input images.

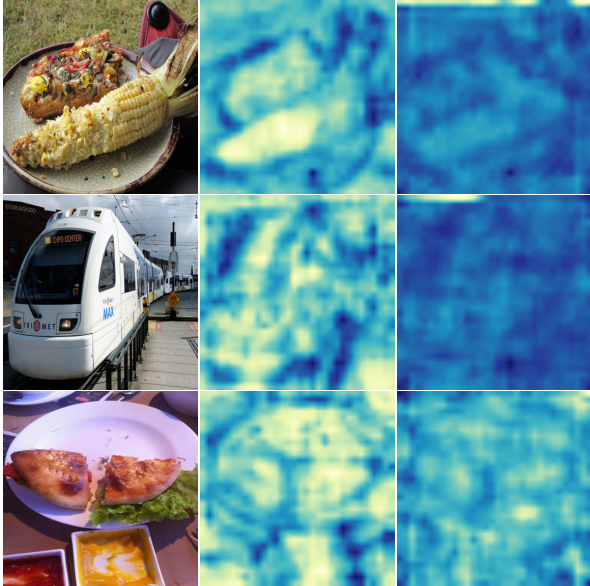


Figure 1: Visualization of feature maps. Images of the first column are from COCO val2017. The second and the third column depict the output feature maps of ‘layer4’ of ResNet-18 pre-trained by PCD and vectorized variant of PCD, respectively.

B. Details of Fine-Tuning Experiments

B.1. Fine-Tuning on CityScapes

The fine-tuning hyper-parameters are listed as follows:

Configuration	Value
optimizer	SGD
base learning rate	0.01
momentum	0.9
weight decay	1e-4
batch size	16
training steps	90000
learning rate schedule	WarmupMultiStepLR
warmup iters	1000
decay milestones	63000, 81000
shortest edge of resizing	[512, 768, ..., 2048]
max input size	4096
random cropping	True
random flipping	True

B.2. Fine-Tuning MobileNet v3 (Large)

MobileNet v3 (Large) [5] is not a stage-wise architecture like ResNet series [4]. We have to manually define the ‘stem’ and ‘res’ stages in MobileNet v3 (Large) to fit in Detectron2. The rules for partitioning the modules of MobileNet v3 (Large) are: i) modules with the same stride belong to the same partition; ii) res4 must be of stride 16. We show the partitioning results:

index	modules	total stride	partitions
1	conv	2	stem
2	IBN	2	
3	IBN	4	res2
4	IBN	4	
5	IBN	8	res3
6	IBN	8	
7	IBN	8	
8	IBN	16	res4
9	IBN	16	
10	IBN	16	
11	IBN	16	
12	IBN	16	
13	IBN	16	
14	IBN	32	res5
15	IBN	32	
16	IBN	32	
17	conv	32	

B.3. Distilling from Different Teachers

In Sec.4.3, we try using other models as teachers. Most of checkpoints are from the official repository, except for BYOL [3]. The ResNet-50 [4] pre-trained by BYOL are from the implementation of [6]. We list out the URLs for downloading these models:

- MoCo v3 (ResNet-50) [2]: <https://dl.fbaipublicfiles.com/moco-v3/r-50-1000ep/r-50-1000ep.pth.tar>
- MoCo v3 (ViT-Base) [2]: <https://dl.fbaipublicfiles.com/moco-v3/vit-b-300ep/vit-b-300ep.pth.tar>
- SwAV [1]: https://dl.fbaipublicfiles.com/deepcluster/swav_800ep_pretrain.pth.tar
- BYOL [3]: <https://drive.google.com/file/d/1-5-049vsro9YW9WTokSc8CoSrmjKfieB/view?usp=sharing>

- Barlow Twins [7]: <https://dl.fbaipublicfiles.com/barlowtwins/ljng/checkpoint.pth>

Barlow Twins has a projection head with an output dimension of 8192. We have to use the variant of PCD with asymmetric loss to distill knowledge from Barlow Twins due to limited memory.

References

- [1] Mathilde Caron, Ishan Misra, Julien Mairal, Priya Goyal, Piotr Bojanowski, and Armand Joulin. Unsupervised learning of visual features by contrasting cluster assignments. *Advances in Neural Information Processing Systems*, 33:9912–9924, 2020. [1](#)
- [2] Xinlei Chen, Saining Xie, and Kaiming He. An empirical study of training self-supervised vision transformers. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 9640–9649, 2021. [1](#)
- [3] Jean-Bastien Grill, Florian Strub, Florent Alché, Corentin Tallec, Pierre Richemond, Elena Buchatskaya, Carl Doersch, Bernardo Avila Pires, Zhaohan Guo, Mohammad Gheshlaghi Azar, et al. Bootstrap your own latent—a new approach to self-supervised learning. *Advances in neural information processing systems*, 33:21271–21284, 2020. [1](#)
- [4] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. [1](#)
- [5] Andrew Howard, Mark Sandler, Grace Chu, Liang-Chieh Chen, Bo Chen, Mingxing Tan, Weijun Wang, Yukun Zhu, Ruoming Pang, Vijay Vasudevan, et al. Searching for mobilenetv3. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 1314–1324, 2019. [1](#)
- [6] Junqiang Huang, Xiangwen Kong, and Xiangyu Zhang. Revisiting the critical factors of augmentation-invariant representation learning. In *European Conference on Computer Vision*, pages 42–58. Springer, 2022. [1](#)
- [7] Jure Zbontar, Li Jing, Ishan Misra, Yann LeCun, and Stéphane Deny. Barlow twins: Self-supervised learning via redundancy reduction. In *International Conference on Machine Learning*, pages 12310–12320. PMLR, 2021. [2](#)