

# Improving Pixel-Level Contrastive Learning by Leveraging Exogenous Depth Information : Supplementary Material

Ahmed Ben Saad<sup>1,2</sup> Kristina Prokopc<sup>2</sup> Josselin Kherroubi<sup>2</sup> Axel Davy<sup>1</sup>  
Adrien Courtois<sup>1</sup> Gabriele Facciolo<sup>1</sup>

<sup>1</sup>ENS Paris-Saclay, Centre Borelli <sup>2</sup>Schlumberger AI Lab

## More results for feature quality comparison

In this supplementary material we provide more examples of Representation Quality Evaluation experiment described in Section 4 of the main paper. Figure 1 shows extra results comparing on several images from ImageNet using networks pre-trained on a subset of the MIDAS training set. The depth maps were computed using MIDAS. Figure 2 shows the same experiment but on images from ScanNet while using a network pre-trained on ImageNet. The idea of this variations of training datasets is to use images that are not used during training for evaluation.

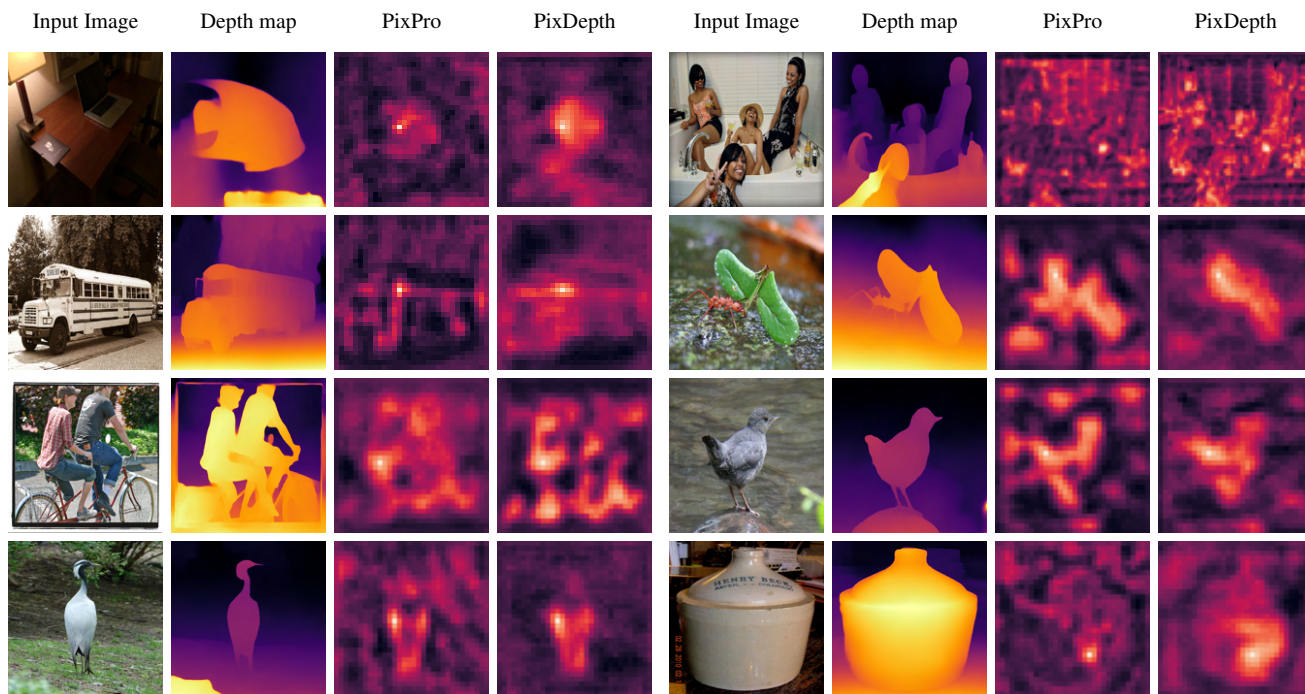


Figure 1. More comparison of the quality of the learned representations on ImageNet images. The bright pixels in the similarity maps correspond to the reference feature vector, which is compared to all other vectors forming the features map. Note that these images are part of the validation set (unseen during training), and that the depth map is given for comparison purposes, it is not fed into the encoder.

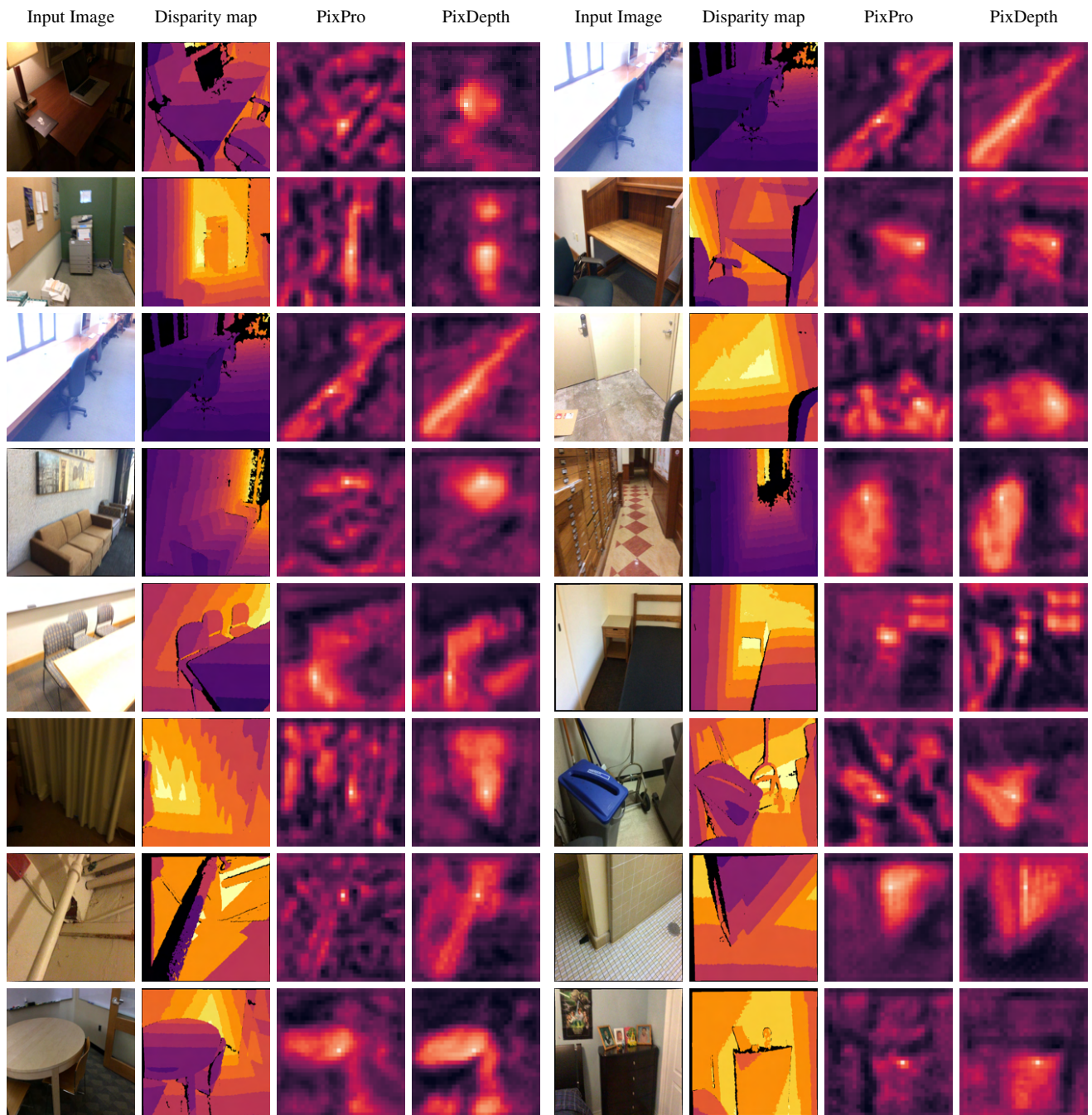


Figure 2. Comparison of the quality of the learned representations on images from the ScanNet dataset. The bright pixels in the similarity maps correspond to the reference feature vector, which is compared to all other vectors forming the features map. Note that these images are part of the validation set (unseen during training), and that the depth map is given for comparison purposes, it is not fed into the encoder.