Locating Noise is Halfway Denoising for Semi-Supervised Segmentation

Yan Fang[1,2] *, Feng Zhu[3], Bowen Cheng[4] †, Luoqi Liu[5], Yao Zhao[1,2], Yunchao Wei[1,2] ✉

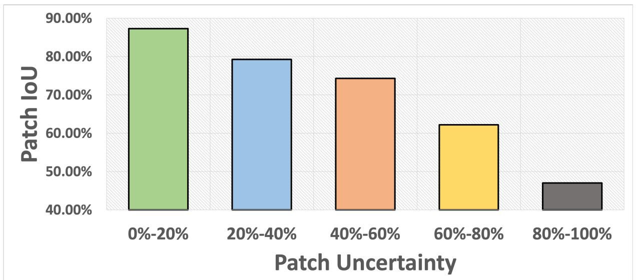

Figure 1: Statics of correspondence between patch uncertainty and patch accuracy. From left to right are patch groups with different uncertainty whereas the left ones are patches with lower uncertainty.

1. Appendix

1.1. Uncertainty Measurement

In the first part of our supplementary material, we analyze the correspondence between the uncertainty score and pseudo mask quality, and report our findings. As depicted in Figure 1, we measure the accuracy of different patches with varying levels of uncertainty. Our results demonstrate a relatively strong relationship between uncertainty and pseudo mask quality, which provides good evidence to support the accurate measurement role of our proposed Patch-wise Uncertainty approach.

1.2. Effects on Learning Hard Samples

UPC removes the Top-k regions with high uncertainty, but for some difficult samples, they may already have low confidence. So, it is possible for UPC to learn more simple samples, ignoring the truly meaningful regions. To answer this question, we brief analyze the learning of UPC on hard samples. We select 200 hard samples with low confidence, and find that top-2 uncertain regions (used in our redundant augmentation strategy) in the hard samples tend to be **harmful** instead of **meaningful**. Around **70%** pixels of these regions are given wrong pseudo labels. Using these noisy regions will result in model degradation instead of improvement. We further test the performance of UPC on two difficult categories: chair (baseline **32.13%** IoU) and sofa (**52.46%** mIoU). UPC achieves significant improvements on these two difficult categories (**3.39%** and **4.30%** respectively), which proves UPC is capbale of learning hard samples better with limited costs.

1.3. Scaling up to Larger Datasets

We have shown UPC has robust generalization on o.o.d unlabeled images by using Pascal VOC as labeled data and COCO as unlabeled data. To validate the effectiveness of this method on even larger annotated datasets, we deploy UPC on larger labeled dataset with ADE20K+COCO setting, using ADE20K as labeled data and COCO as unlabeled data. Result of our UPC on ADE20K+COCO is **3.6%** higher than SupOnly, demonstrating the effectiveness of our UPC on larger datasets.

Method	SupOnly	w/ MSCOCO
UPC	47.8	**51.4** ↑3.6

Table 1: Results on ADE20K+COCO setting. “SupOnly” stands for training only using ADE20K training data.

1.4. Compatibility with other augmentation methods

We further conduct ablation experiments on the compatibility of our proposed UPC with other augmentation methods. First we have to clarify that our method is a specific improvement over CutMix. Using UPC and CutMix alike methods together will lead to heavy damage on meaningful regions, which is the reason why UPC is not compatible with CutMix and Copy-Paste, as shown in Table 2. But it is still a good question whether UPC is compatible with other kinds of augmentation methods, such as Gaussian Blur and RandAugment. The results below demonstrate UPC is well compatible with Gaussian Blur and RandAugment.

UPC	w/ CutMix	w/ Copy-Paste	w/ Gaussian Blur	w/ RandAugment
79.47	77.60 ↓1.87	78.23 ↓1.24	**79.83** ↑0.36	**79.75** ↑0.28

Table 2: Compatibility with other data augmentations.