# A Simple Baseline for Semi-supervised Semantic Segmentation with Strong Data Augmentation: Supplementary Material

## S1. Semi-supervised learning Framework

We show our semi-supervised learning framework in Algorithm 1.

**Algorithm 1:** Semi-supervised Learning Framework

**Labeled images:** $n$ pairs of images $x_i$ and corresponding labels $y_i$
**Unlabeled images:** $m$ images without labels $\{(\widetilde{x}_1, \widetilde{x}_2, ..., \widetilde{x}_m\}$
**Step1:** Optimize the teacher model with the cross-entropy loss on labeled images;
**while** *iteration <maximum number of iterations* **do**
  **Step2:** Use the teacher model to generate hard pseudo-labels (one-hot encodings) for clean (*i.e.*, not distorted) unlabeled images with multi-scale and flip inference to get $m$ pairs of images with pseudo-labels $\{(\widetilde{x}_1, \widetilde{y}_1), (\widetilde{x}_2, \widetilde{y}_2), ..., (\widetilde{x}_m, \widetilde{y}_m)\}$
  **Step3:** Mix the $(\widetilde{x}, \widetilde{y})$ and $(x, y)$ to get a new dataset $\mathsf{M}$.
  **for** $step = 1, ..., n_{steps}$ **do**
    $x^{w}_{batch}$=Sample(M),
    $x^{s}_{batch}$=Sample(M),
    $x^{s}_{batch}$ = SDA($x^{s}_{batch}$)
    $x_{batch}$=[$x^{w}_{batch}$,$x^{s}_{batch}$],
    $[y^*, \widetilde{y}^*] = f([x, \hat{x}])$;
    loss $=$ CE($y^*$, $y$) $+$ SCL($\widetilde{y}^*$, $\widetilde{y}$);
    minimizing the loss and update student parameters $\theta_s$;
  **end**
  Re-initialize the teacher parameters $\theta_t = \theta_s$;
  Jump to **step 2**.
**end**

## S2. Ablation Study with DeepLabV2

We conduct experiments to explore the effectiveness of each proposed module under 1/8 labeled image settings. First, we establish the baseline for our experiments. We evaluate DeepLabV2 based on ResNet-101 on the validation set. Same as the experiments with the DeepLabV3Plus model, self-training adopts the baseline model as the teacher model and the teacher model generates pseudo-labels on the unlabeled dataset, then the student model which is not smaller than the teacher model is trained on pseudo labels and original labeled images. We use the scales including $\{0.5, 0.75, 1.0, 1.5, 1.75\}$ and mirror for the remaining images to generate pseudo labels. For fair comparison, the student model is the same model initialized by the teacher model. The main results are shown in Table 1.

| Model | s.t. | SDA | DSBN | SCL | iter. | mIoU |
|---|---|---|---|---|---|---|
| DeepLabV2 | | | | | | 60.6 |
| DeepLabV2 | ✓ | | | | | 62.1 |
| DeepLabV2 | ✓ | ✓ | | | | 62.2 |
| DeepLabV2 | ✓ | ✓ | ✓ | | | 63.8 |
| DeepLabV2 | ✓ | ✓ | ✓ | ✓ | | 64.5 |
| DeepLabV2 | ✓ | ✓ | ✓ | ✓ | ✓ | 67.6 |
| DeepLabV2-Full | | | | | | 70.1 |

Table 1. Ablation study on the proposed semi-supervised learning framework. Baselines are ResNet101-based Deeplabv2. 's.t.́denotes self-training with pseudo labels without strong augmentation. SDA means the strong data augmentation. DSBN means the distribution specify batch normalization. SCL is the proposed self correction loss. 'iter.' represents iterative training. The results are evaluated on the validation set with the single-scale input. Except for those with the full identifier, all experiments used $1/8$ labeled data and the remaining images in the training set are used as unlabeled data. '· Full' means the fully supervised setting.

We also report detailed performance under different iterations in Figure 1. We can see that more iterations in general help improve the performance, but the growth trend of the performance is gradually flattening.

## S3. Ablation Study with the SCL loss

We compare our proposed SCL loss with the popular weighted cross entropy(WCE) loss. Our proposed SCL loss applies different loss weight to each pixels according to the confidence, while the weighted cross entropy apply different for pixels belonging to different classes. The experiments are conducted on Cityscapes. We use $\frac{1}{8}$ labeled data, and $\frac{7}{8}$ pseudo labeled data. The results are shown in Table 2.

| Baseline | mIoU |
|---|---|
| W-CE | 72.4 |
| CE | 72.3 |
| SCL | 72.8 |

Table 2. Compared the proposed SCL with weighted cross entropy based on DeepLabv3plus.

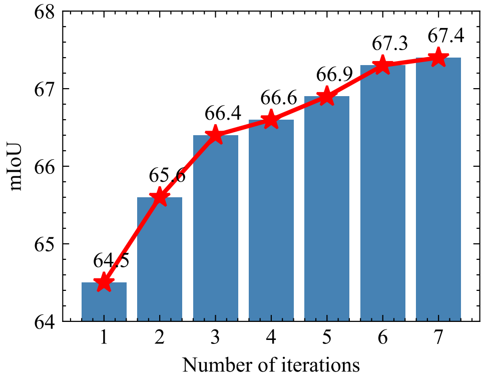
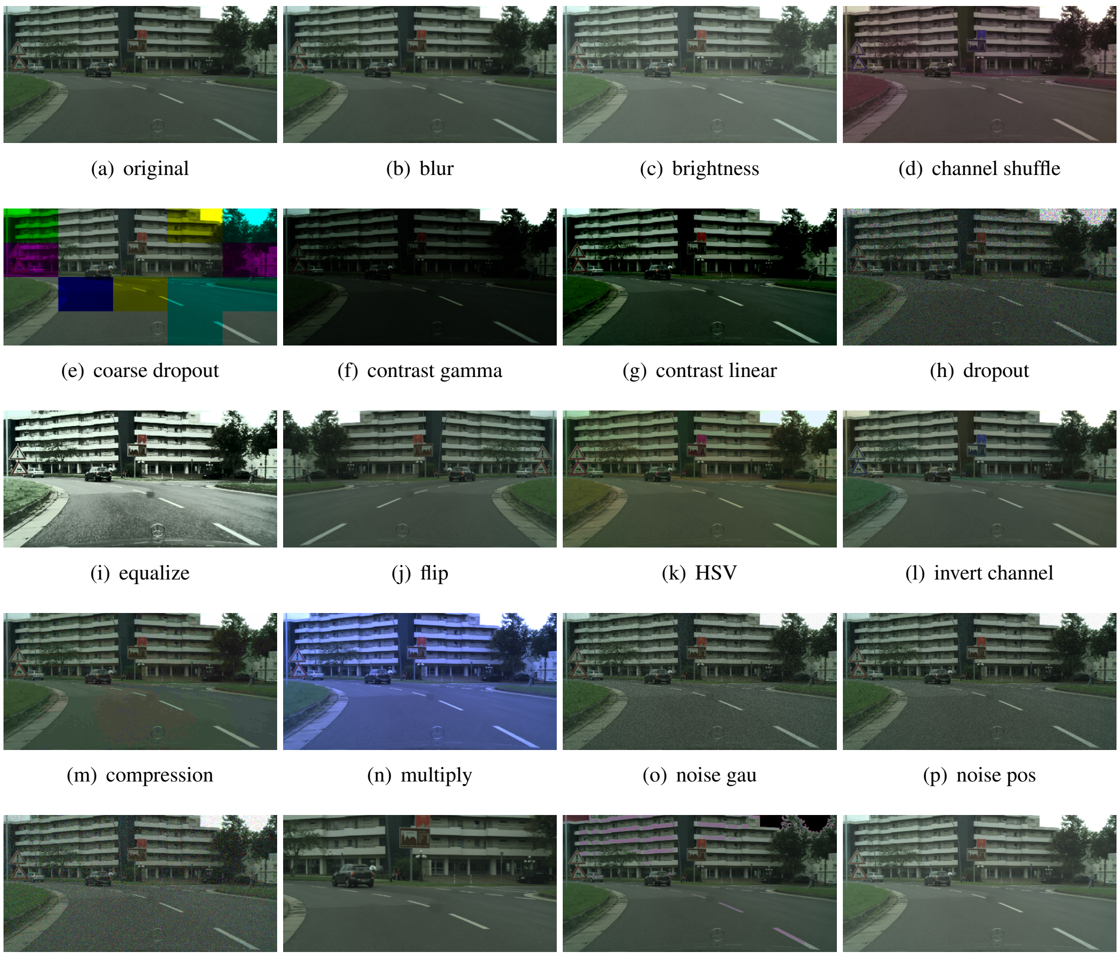



Figure 1. Performance vs. number of iterations, using the DeepLabV2 model.

## S4. Visualization Strong Augmentation

Several visualization augmentation images are shown in Figure 2. From Figure 2, we find that strong augmentation images are much different from the original image. This may cause the domain gap.



(a) original (b) blur (c) brightness (d) channel shuffle

(e) coarse dropout (f) contrast gamma (g) contrast linear (h) dropout

(i) equalize (j) flip (k) HSV (l) invert channel

(m) compression (n) multiply (o) noise gau (p) noise pos

(q) salt pepper (r) scale (s) solarized (t) brightness channel

Figure 2. Strong augmentation for semantic segmentation