Revisiting Weak-to-Strong Consistency in Semi-Supervised Semantic Segmentation [CVPR 2023]

We surprisingly notice that, so long as coupled with appropriate strong perturbations, FixMatch can indeed still exhibit powerful generalization capability in our scenario.

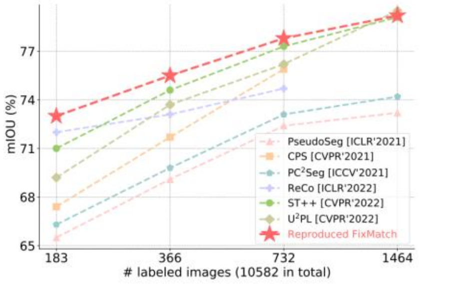


Figure 1. Comparison between state-of-the-art methods and our reproduced FixMatch [55] on the Pascal dataset.

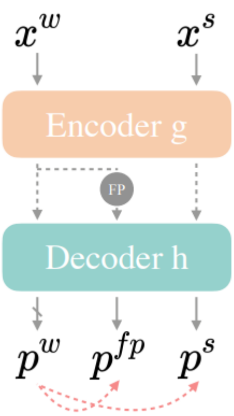
Method	# labeled images (10582 in total)				
	92	183	366	732	1464
w/o any SP	39.5	52.7	65.5	69.2	74.6
w/ CutMix	56.7	67.9	71.9	75.1	78.3
w/ whole SP	**63.9**	**73.0**	**75.5**	**77.8**	**79.2**

Table 1. The importance of image-level strong perturbations (SP) to FixMatch on the Pascal dataset. *w/o any SP*: directly utilize hard label of x^w to supervise its logits. *w/ CutMix*: only use CutMix [71] as a perturbation. *w/ whole SP*: strong perturbations contain color transformations from ST++ [68], together with CutMix.

(1) Expanding a broader perturbation space:
design a unified perturbation framework for both raw images and extracted features
(2) Sufficiently harvesting original perturbations:
dual independent strong views are randomly sampled from the perturbation pool

Unified Perturbations for Images and Features

In order to construct a broader perturbation space, built on top of FixMatch, we propose to inject perturbations on features of the weakly perturbed image.



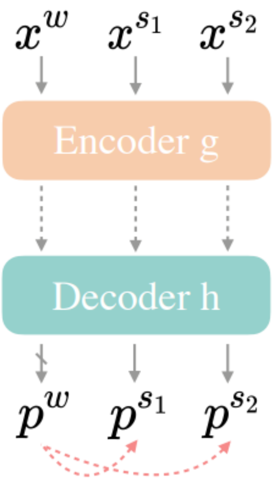
(a) UniPerb

$$\mathcal{L}_u = \frac{1}{B_u}\sum \mathbb{1}(\max(p^w) \geq \tau)\big(\mathrm{H}(p^w, p^s) + \mathrm{H}(p^w, p^{fp})\big).$$

Do not aim at proposing a novel feature perturbation approach in this work. Actually, an embarrassingly simple channel dropout (nn.Dropout2d in PyTorch) is well-performed enough.

Dual-Stream Perturbations

Rather than feeding a single p^s into the model, we independently yield dual-stream perturbations (x^{s1}, x^{s2}) from x^w by strong perturbation pool A^s

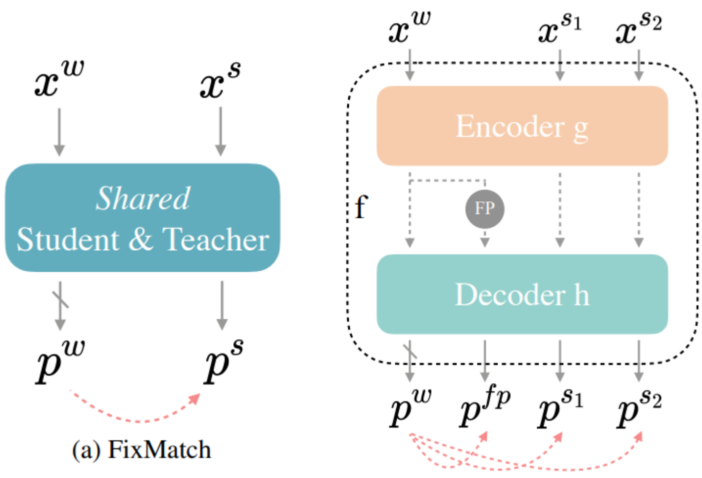


(b) DusPerb

$$\mathcal{L}_u = \frac{1}{B_u} \sum \mathbb{1}(\max(p^w) \geq \tau) \cdot \\ \left(\lambda \mathrm{H}(p^w, p^{fp}) + \frac{\mu}{2}\left(\mathrm{H}(p^w, p^{s_1}) + \mathrm{H}(p^w, p^{s_2})\right)\right).$$

Regularizing two strong views with a shared weak view can be regarded as enforcing consistency between these two strong views.

UniMatch



(a) FixMatch

(b) UniMatch

Figure 2. (a) The FixMatch baseline. (b) Our proposed unified dual-stream perturbations method (UniMatch). The FP denotes feature perturbation, and the dashed curves represent supervision.