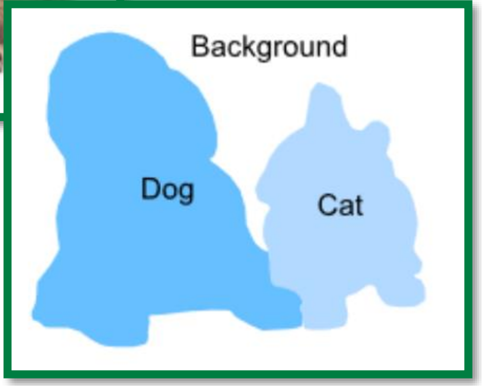
Semi-Supervised Semantic Segmentation with Cross Pseudo Supervision

Xiaokang Chen[1*] Yuhui Yuan[2] Gang Zeng[1] Jingdong Wang[2]
[1]Key Laboratory of Machine Perception (MOE), Peking University [2]Microsoft Research Asia

北京理工大学 BEIJING INSTITUTE OF TECHNOLOGY

Semantic Segmentation

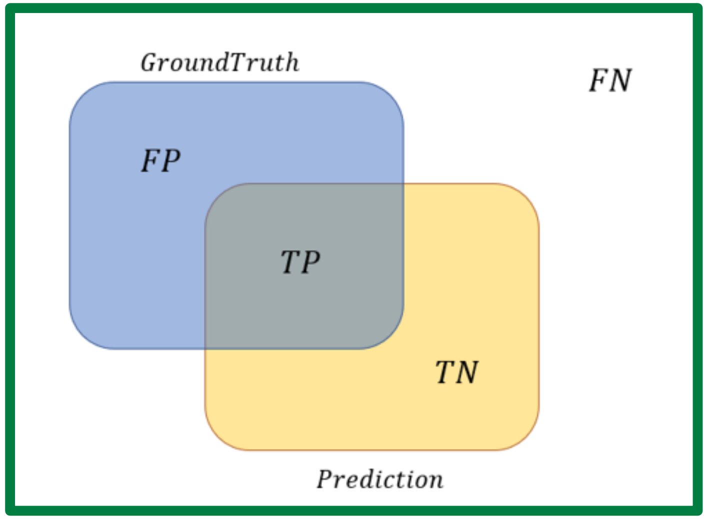
- Category-label prediction for each pixel
- A pixel-wise classification
- mask



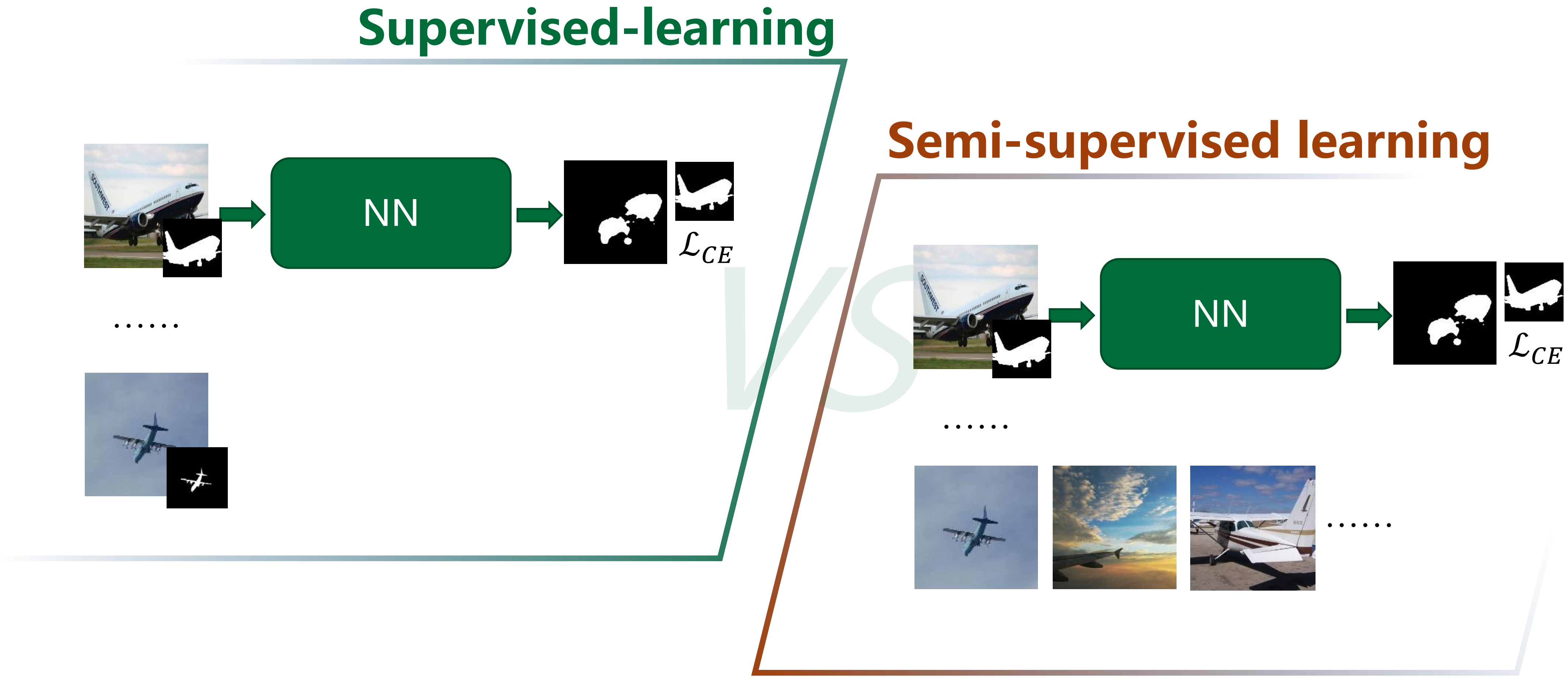
IoU

- $$\mathrm{IoU} = \frac{\mathrm{Prediction} \cap \mathrm{GroundTruth}}{\mathrm{Prediction} \cup \mathrm{GroundTruth}}$$

$$= \frac{\mathrm{T}P}{\mathrm{TP} + \mathrm{T}N + FP}$$



Semi-supervised learning



Self-training

Predict label for unlabeled sample & enlarge train set

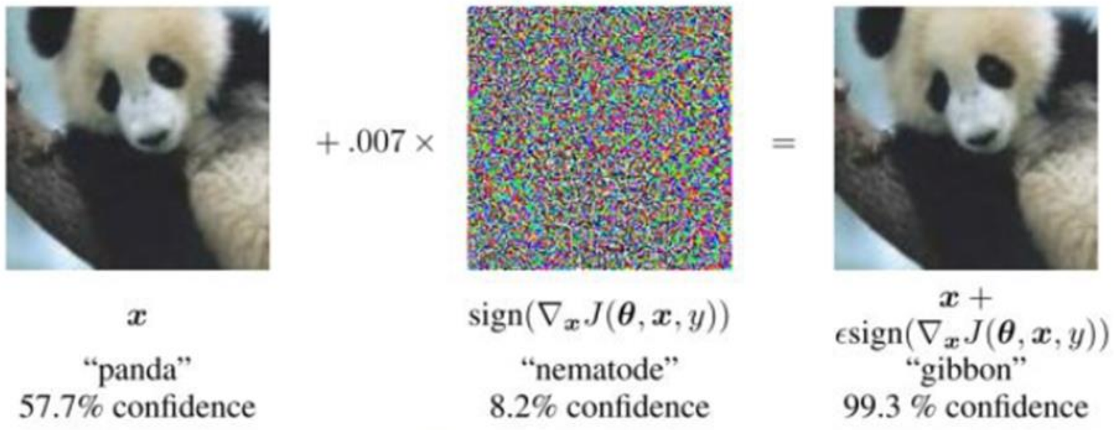
Algorithm 1 Self-training.

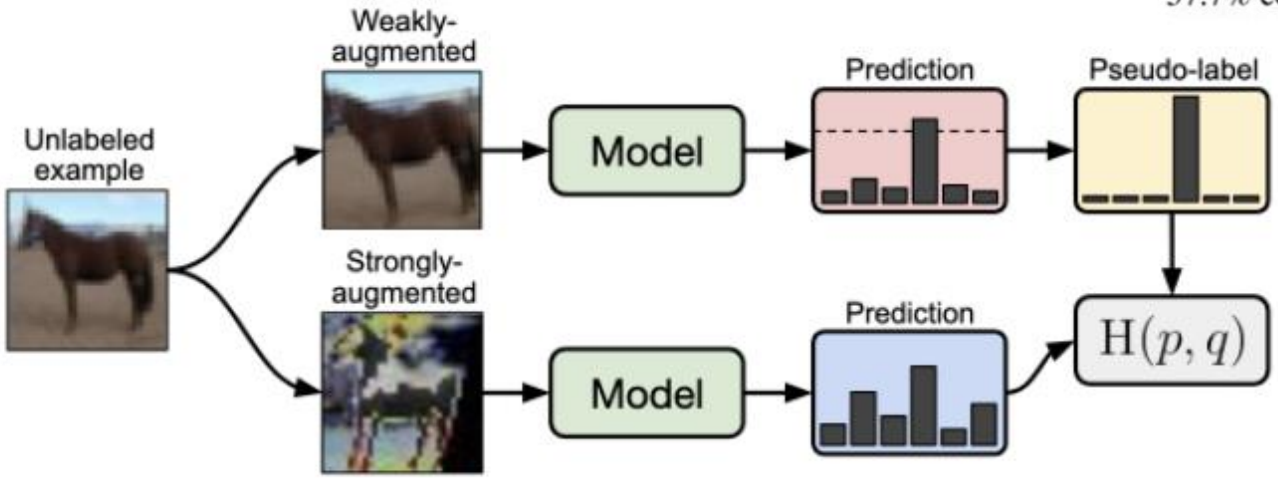
Initialize:
Given $(X_{train}, y_{train}) = (X_l, y_l)$
while stopping criteria not met **do**
  Train classifier C_{int} from (X_{train}, y_{train})
  Use C_{int} to predict class label y_u of X_u
  Select confidence sample (X_{conf}, y_{conf}); $(X_{conf}, y_{conf}) \in (X_u, y_u)$
  Remove selected unlabeled data $X_u \leftarrow X_u - X_{conf}$
  Combine newly labeled data $(X_{train}, y_{train}) \leftarrow (X_l, y_l) \cup (X_{conf}, y_{conf})$
end while

Methods

Consistency Regularization

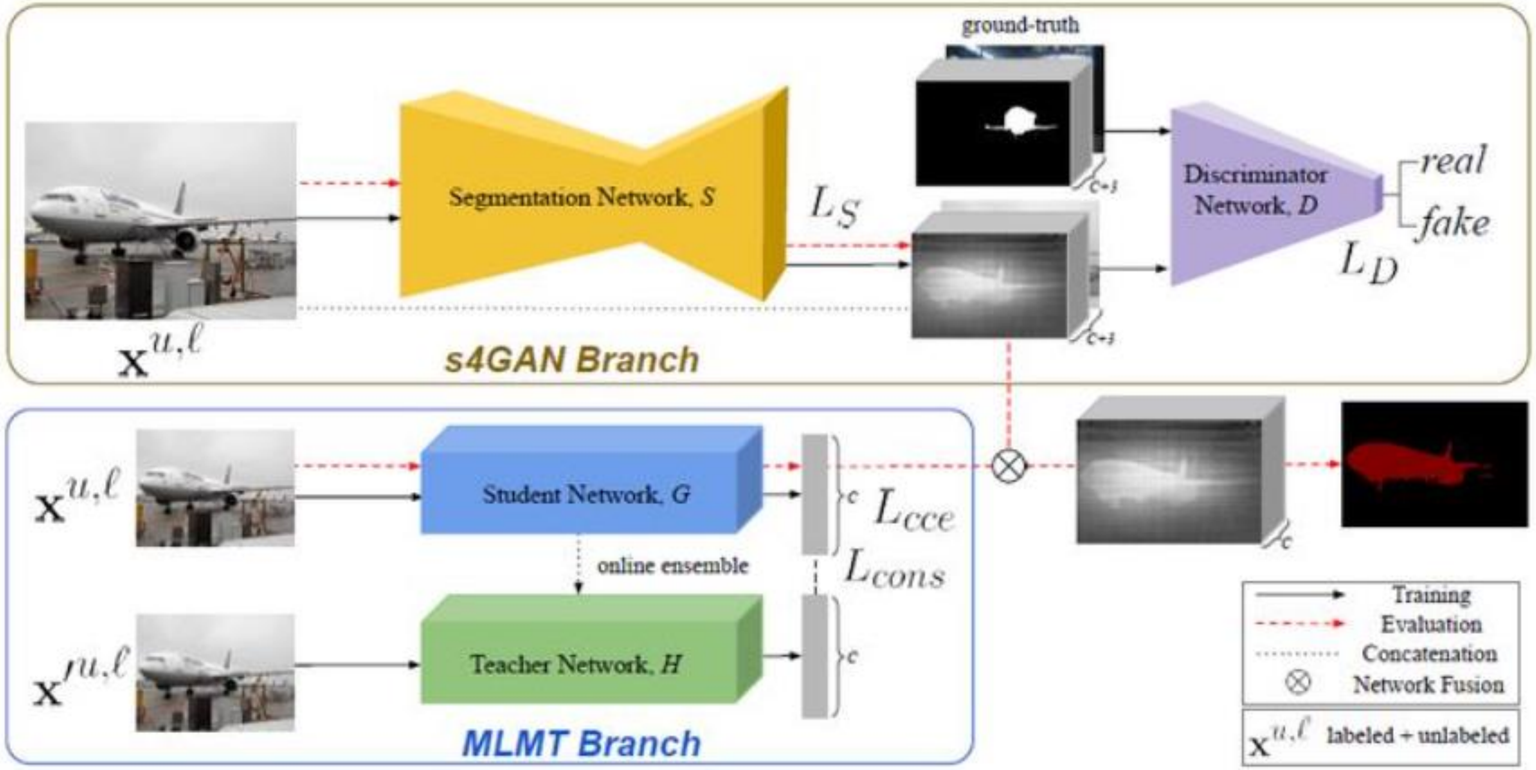
Enforce predictions consistency with various perturbations of input





[1]Goodfellow I J, Shlens J, Szegedy C. Explaining and harnessing adversarial examples[J]. arXiv preprint arXiv:1412.6572, 2014.
[2] Sohn K, Berthelot D, Li C L, et al. Fixmatch: Simplifying semi-supervised learning with consistency and confidence[J]. arXiv preprint arXiv:2001.07685, 2020.

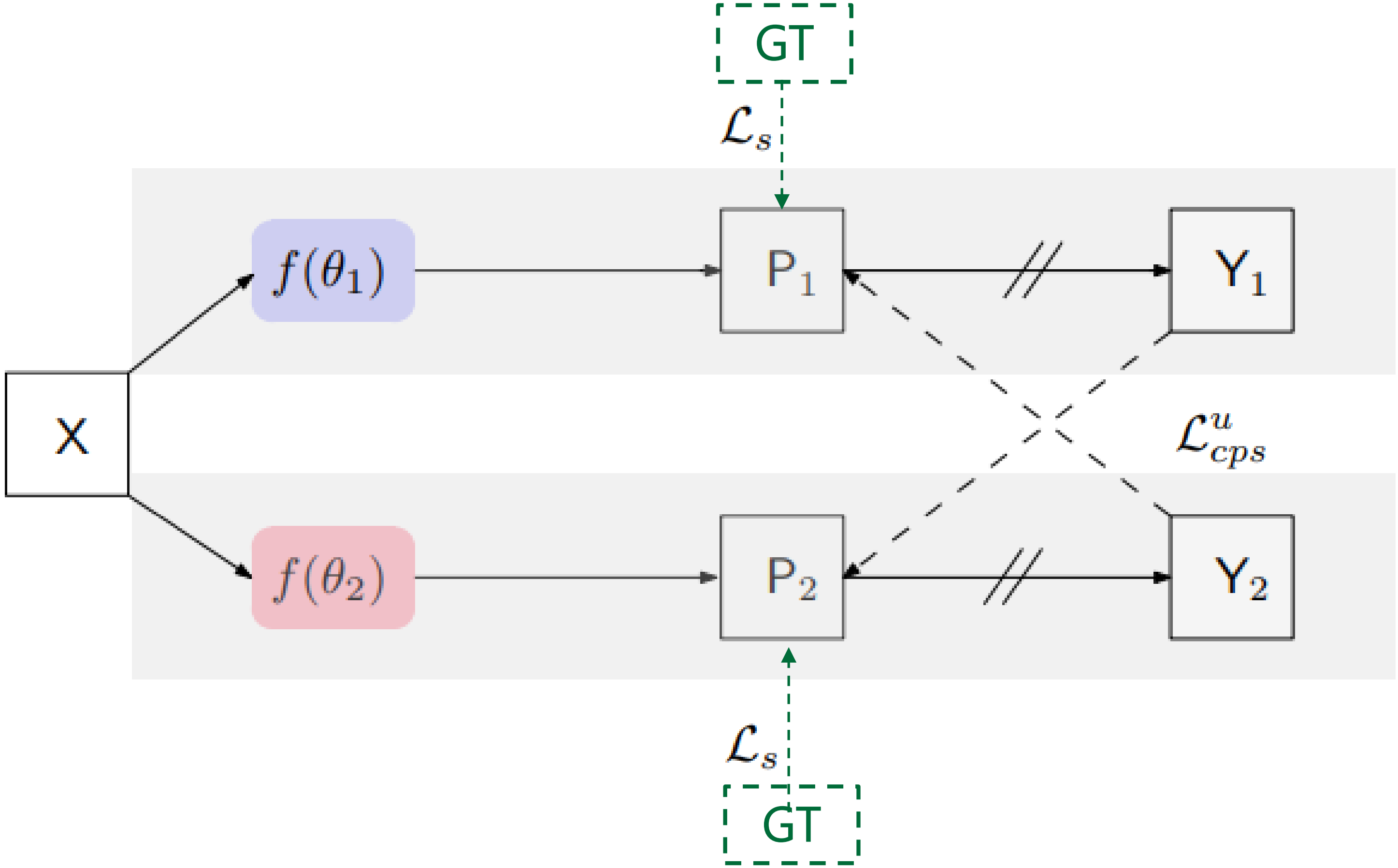
GAN-Based Methods



Generator:
Input: image
Output: C-channel mask

Discriminator:
Input: C-channel mask ⊕ Image
Output: real/fake

Cross Pseudo Supervision



)

$\boldsymbol{f(\theta_1), f(\theta_2)}$**:**
Same model architecture with different initialization
// : stop gradient

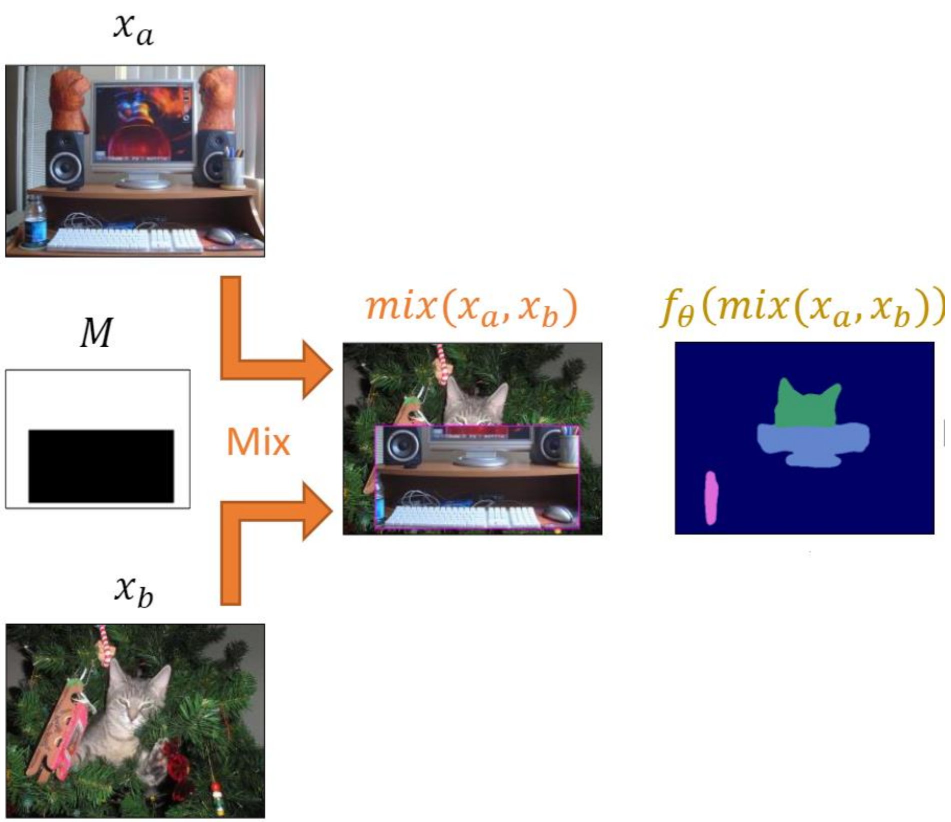
$$\mathcal{L} = \mathcal{L}_s + \lambda \mathcal{L}_{cps}$$

implement:
DeepLab v3+(ResNet)

Cutmix

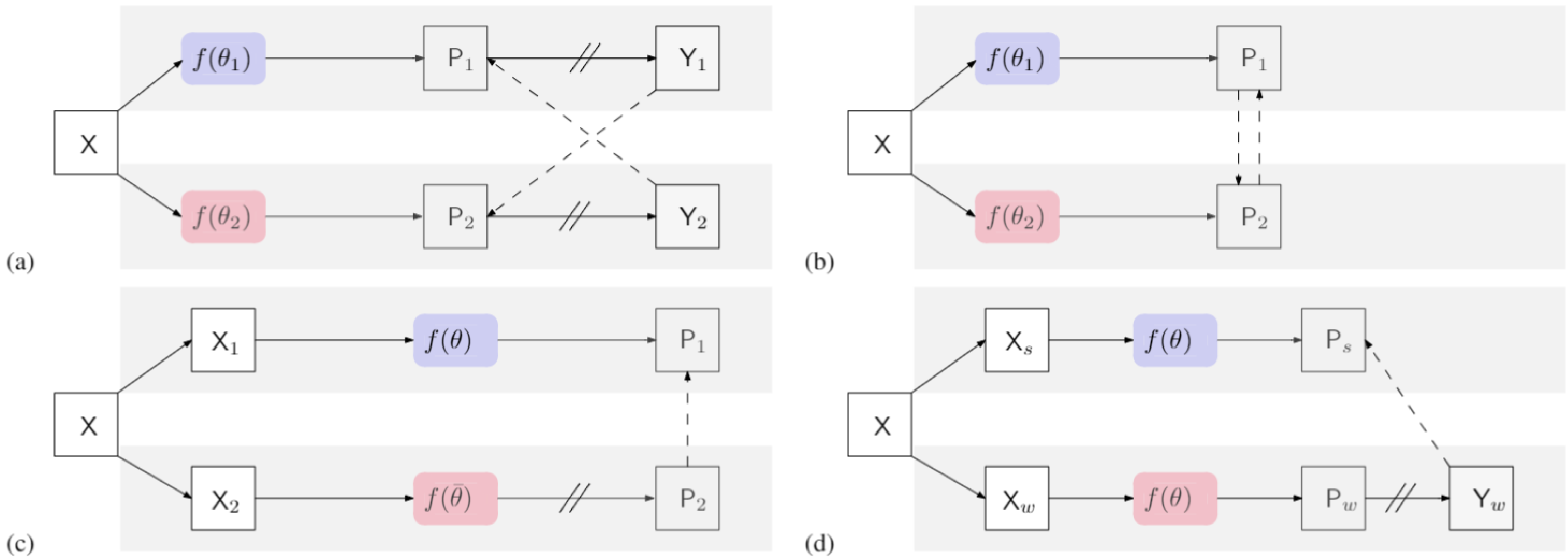
	ResNet-50	Mixup [48]	Cutout [3]	CutMix
Image				
Label	Dog 1.0	Dog 0.5 Cat 0.5	Dog 1.0	Dog 0.6 Cat 0.4
ImageNet Cls (%)	76.3 (+0.0)	77.4 (+1.1)	77.1 (+0.8)	**78.6** **(+2.3)**
ImageNet Loc (%)	46.3 (+0.0)	45.8 (-0.5)	46.7 (+0.4)	**47.3** **(+1.0)**
Pascal VOC Det (mAP)	75.6 (+0.0)	73.9 (-1.7)	75.1 (-0.5)	**76.7** **(+1.1)**

Cutmix in Semi-supervised Segmentation



[3]French G, Laine S, Aila T, et al. Semi-supervised semantic segmentation needs strong, varied perturbations[C] //British Machine Vision Conference. 2020 (31).

Comparation



Overview

Table 1: **Comparison with state-of-the-arts** on the PASCAL VOC 2012 val set under different partition protocols. All the methods are based on DeepLabv3+.

Method	ResNet-50				ResNet-101			
	1/16 (662)	1/8 (1323)	1/4 (2646)	1/2 (5291)	1/16 (662)	1/8 (1323)	1/4 (2646)	1/2 (5291)
MT [32]	66.77	70.78	73.22	75.41	70.59	73.20	76.62	77.61
CCT [27]	65.22	70.87	73.43	74.75	67.94	73.00	76.17	77.56
CutMix-Seg [11]	68.90	70.70	72.46	74.49	72.56	72.69	74.25	75.89
GCT [17]	64.05	70.47	73.45	75.20	69.77	73.30	75.25	77.14
Ours (w/o CutMix Aug.)	68.21	73.20	74.24	75.91	72.18	75.83	77.55	78.64
Ours (w/ CutMix Aug.)	**71.98**	**73.67**	**74.90**	**76.15**	**74.48**	**76.44**	**77.68**	**78.64**

Overview

(a) (b) (c) (d) (e)

Figure 6: **Example qualitative results from PASCAL VOC 2012.** (a) input, (b) ground truth, (c) supervised only, (d) ours (w/o CutMix Aug.), and (e) ours (w/ CutMix Aug.). All the approaches use DeepLabv3+ with ResNet-101 as the segmentation network.

Losses

Losses					PASCAL VOC 2012		Cityscapes	
$\mathcal{L}_s$	$\mathcal{L}^l_{cps}$	$\mathcal{L}^u_{cps}$	$\mathcal{L}^l_{cpc}$	$\mathcal{L}^u_{cpc}$	ResNet-50	ResNet-101	ResNet-50	ResNet-101
✓					69.43	72.21	70.32	72.19
✓	✓				69.99	72.98	71.73	73.08
✓		✓			73.00	75.83	73.97	75.28
✓	✓	✓			**73.20**	**75.85**	**74.39**	**75.71**
✓			✓	✓	71.23	74.01	72.03	73.77

With single model(pseudo label)

Method	1/16	1/8	1/4	1/2
SPS (w/o CutMix Aug.)	59.54	69.05	72.55	75.17
Ours (w/o CutMix Aug.)	**68.21**	**73.20**	**74.24**	**75.91**
SPS (w/ CutMix Aug.)	65.62	71.27	73.70	74.87
Ours (w/ CutMix Aug.)	**71.98**	**73.67**	**74.90**	**76.15**

With self-training

Method	ResNet-50		ResNet-101	
	1/4	1/2	1/4	1/2
PASCAL VOC 2012				
Ours	74.24	75.91	77.55	78.64
Self-Training	74.47	75.97	76.63	78.15
Ours + Self-Training	**74.96**	**76.60**	**77.60**	**78.76**