Team 20 Semi-Supervised Semantic Segmentation with Cross-Consistency Training

Le Viet Thanh Long (20190780), Nguyen Tuan Kiet* (20200734), Cao Viet Hai Nam (2020817), Ayhan Suleymanzade (20210784), Eugene Lee (20214195)

Problem Formulation

Semantic Segmentation

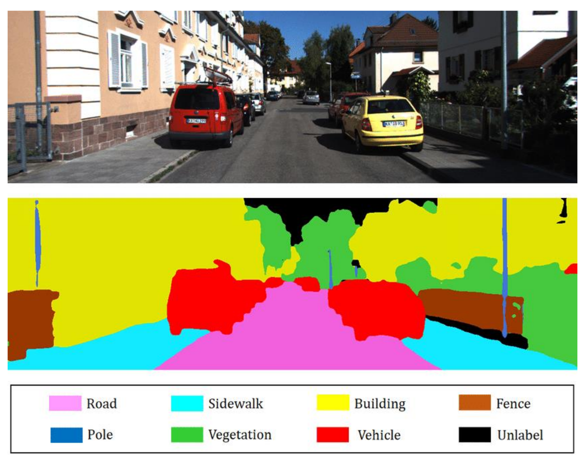


Fig 1. Semantic Segmentation example

- **Pixel-level** classification
- **PASCAL VOC** dataset:
  - 21 classes
  - Training: 1464 images
  - Validation: 1449 images

Semi-Supervised Learning (SSL)

- A **small** amount of **expensive labeled** data
- A **large** amount of **unlabeled** data

Related Work

Semi-supervised Learning

- **Temporal Ensembling**

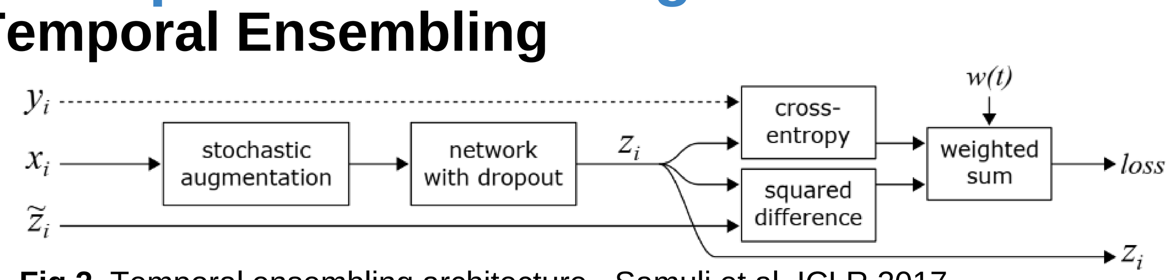


Fig 2. Temporal ensembling architecture - Samuli et al. ICLR 2017

Weakly-supervised Learning (WSL)

- **Pseudo-label**: utilize image-level label

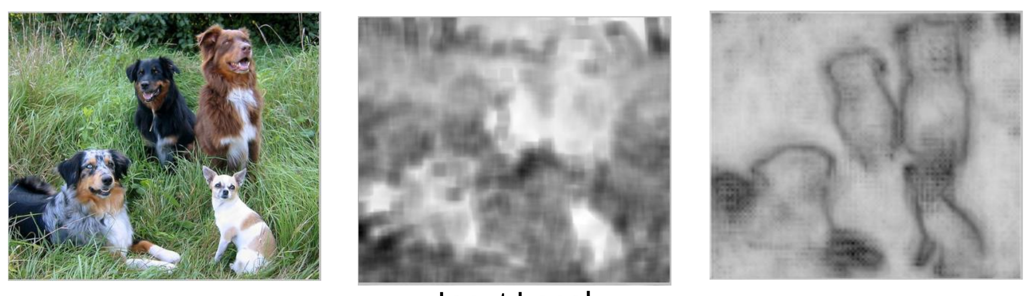
Semantic Segmentation

- **Generative Adversarial Network (GAN)**: provide guidance from a **discriminator**

Method

Cluster assumption

- Classes must be **separated** by **low density** regions



Input | Input Local Smoothness | Hidden representation

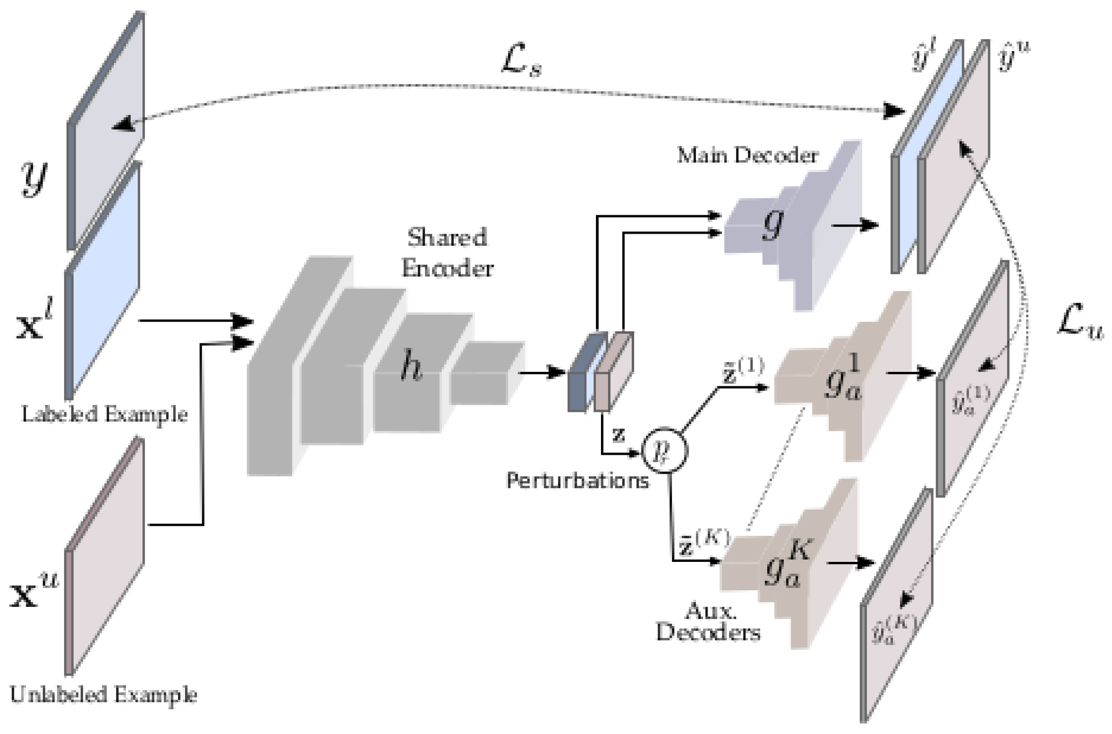


Fig 3. CCT architecture - Y Ouali et al. CVPR 2020

Consistency training

- Enforce **invariance** of model's predictions over small **perturbations** on hidden features

Segmentation Network

- Main **encoder**, main **decoder**

$$\mathcal{L}_s = \frac{1}{|\mathcal{D}_l|} \sum_{\mathbf{x}_i^l, y_i \in \mathcal{D}_l} \mathbf{H}(y_i, f(\mathbf{x}_i^l))$$

Cross-Entropy (CE) based supervised loss

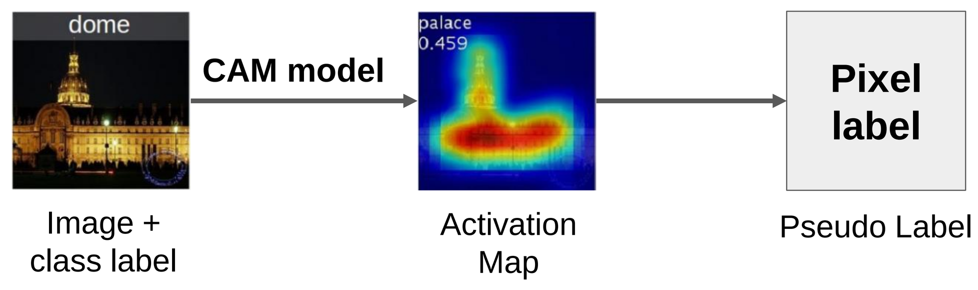
- Auxiliary decoders: **Cross-consistency** training (with unsupervised loss)

$$\mathcal{L}_u = \frac{1}{|\mathcal{D}_u|} \frac{1}{K} \sum_{\mathbf{x}_i^u \in \mathcal{D}_u} \sum_{k=1}^{K} \mathbf{d}(g(\mathbf{z}_i), g_a^k(\mathbf{z}_i))$$

Unsupervised loss

- **Perturbation** types: Feature-based, Prediction-based, random (Dropout)

Exploiting weak labels



- Assign class to pixel: **attention score**

$$\mathcal{L}_w = \frac{1}{|\mathcal{D}_w|} \frac{1}{K} \sum_{\mathbf{x}_i^w \in \mathcal{D}_w} \sum_{k=1}^{K} \mathbf{H}(y_p, g_a^k(\mathbf{z}_i))$$

Weakly supervised loss

Objective Function

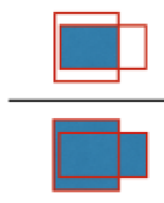
$$\mathcal{L} = \mathcal{L}_s + \omega_u \mathcal{L}_u + \omega_w \mathcal{L}_w$$

Evaluation

Metric

- **mIoU**: mean of class-wise intersection over union.

$$IoU = \frac{B_1 \cap B_2}{B_1 \cup B_2} =$$



Quantitative Results

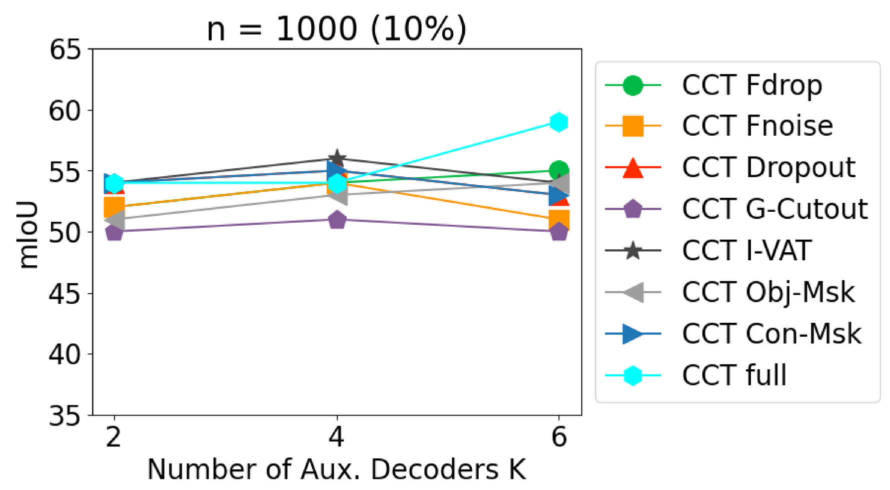
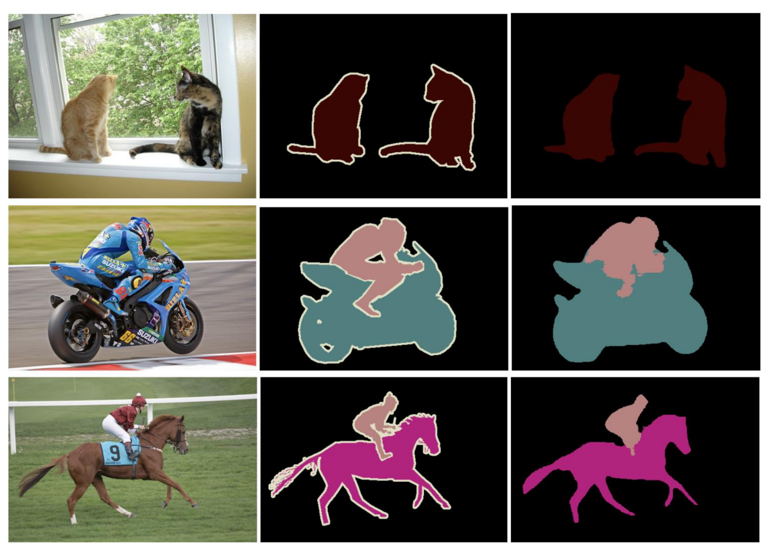


Fig 4. 1000 labeled examples using different perturbations and various numbers of auxiliary decoders K.

Method	Pixel-level Labeled Examples	Image-level Labeled Examples	Val - mIoU
WSSL	1.5k	9k	64.6
MDC	1.5k	9k	65.7
FickleNet	1.5k	9k	65.8
Hung et al	1.5k	-	68.1
CCT	1k	-	64.0
CCT	1.5k	-	**69.4**
CCT	1.5k	9k	**73.2**

Table 1. Comparison with the-state-of-the-art (in 2020).

Qualitative results



Input | Label | CCT (1.5k + wsl)

Improvement Approach

Different Encoder Backbone

Encoder Backbone	Pixel-level Labeled Examples	Image-level Labeled Examples	Val - mIoU
poolformer-m36	1.5k	-	70.8
convnext_base_ink22	1.5k	-	**72.5**

Table 2. Different **Encoders** results.

Seg-GAN network

- **Discriminator**: give signal to improve segmentation network

$$\mathcal{L}_D = -\sum_{h,w} (1 - y_n) \log(1 - D(S(\mathbf{X}_n))^{(h,w)}) + y_n \log(D(\mathbf{Y}_n)^{(h,w)})$$

- **Segmentation network:** generate better output to trick **discriminator**

$$\mathcal{L}_{seg} = \mathcal{L}_{ce} + \lambda_{adv}\mathcal{L}_{adv} + \lambda_{semi}\mathcal{L}_{semi}$$

- Incorporate **WSL** signal

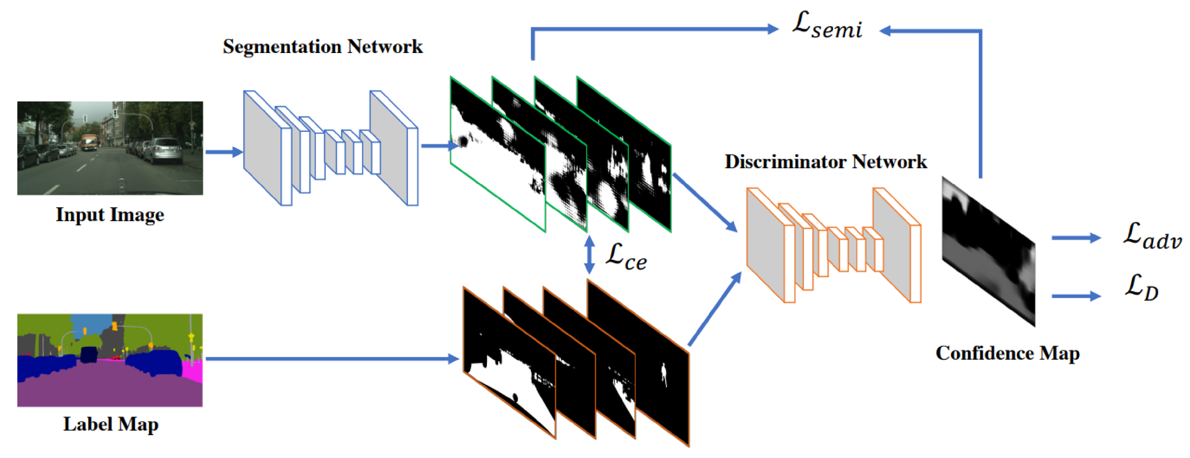


Fig 5. AdvSemiSeg architecture - Hung et al. BMVC 2018

Temporal Ensembling

- Weighted **moving average** of previous predictions as weak label

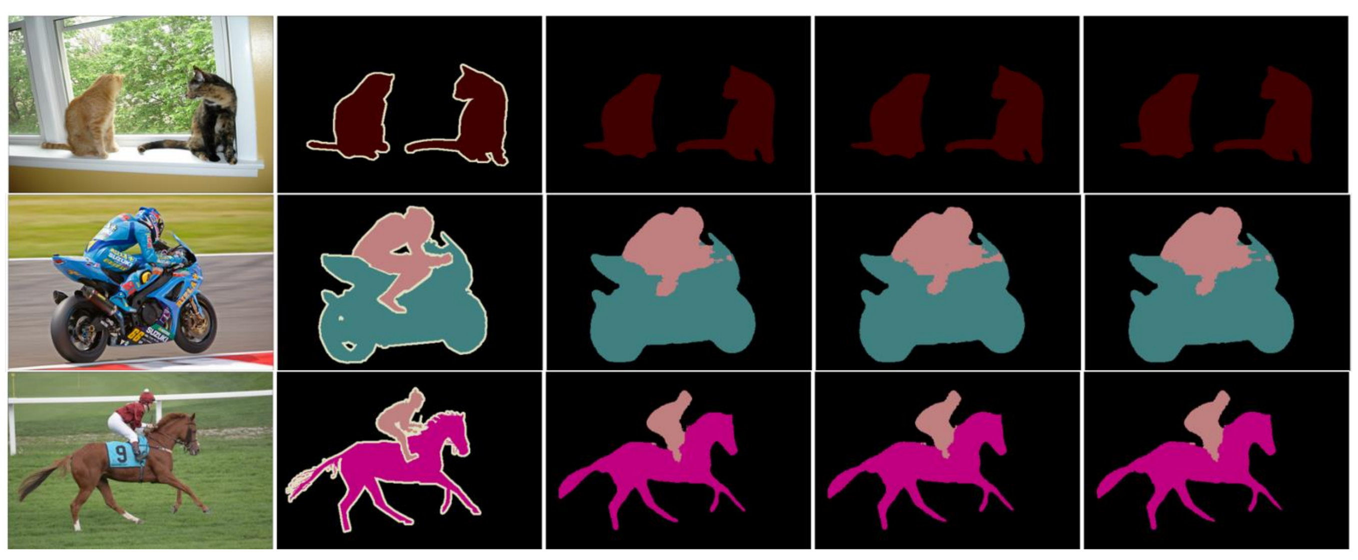
$$L = L_{sup} + w * L_{un\_sup}$$

- Require **large memory**
- Train on 1000 labeled data

Method	Pixel-level Labeled Examples	Image-level Labeled Examples	Continual Training Epoch	Val - mIoU
Seg-GAN	1.5k	-	10	73.30
Seg-GAN semi	1.5k	9k	10	73.46
Seg-GAN semi + WSL	1.5k	9k	10	**73.50**
Temporal Ensembling	1.0k	-	10	68.10

Table 3. Continual training result.

Qualitative results



Input | Label | Seg-GAN | Seg-GAN semi | Seg-GAN semi + WSL