STRUCT Group Paper Reading

# Augmentation Matters: A Simple-yet-Effective Approach to Semi-Supervised Semantic Segmentation

Zhen Zhao, Lihe Yang, Sifan Long, Jimin Pi, Luping Zhou*, Jingdong Wang*
University of Sydney, Baidu Inc., Nanjing University, Jilin University

CVPR 2023 Highlight

PRESENTER: JIAHANG ZHANG

2023/06/04

# Outline

# Outline

# Outline

# Background

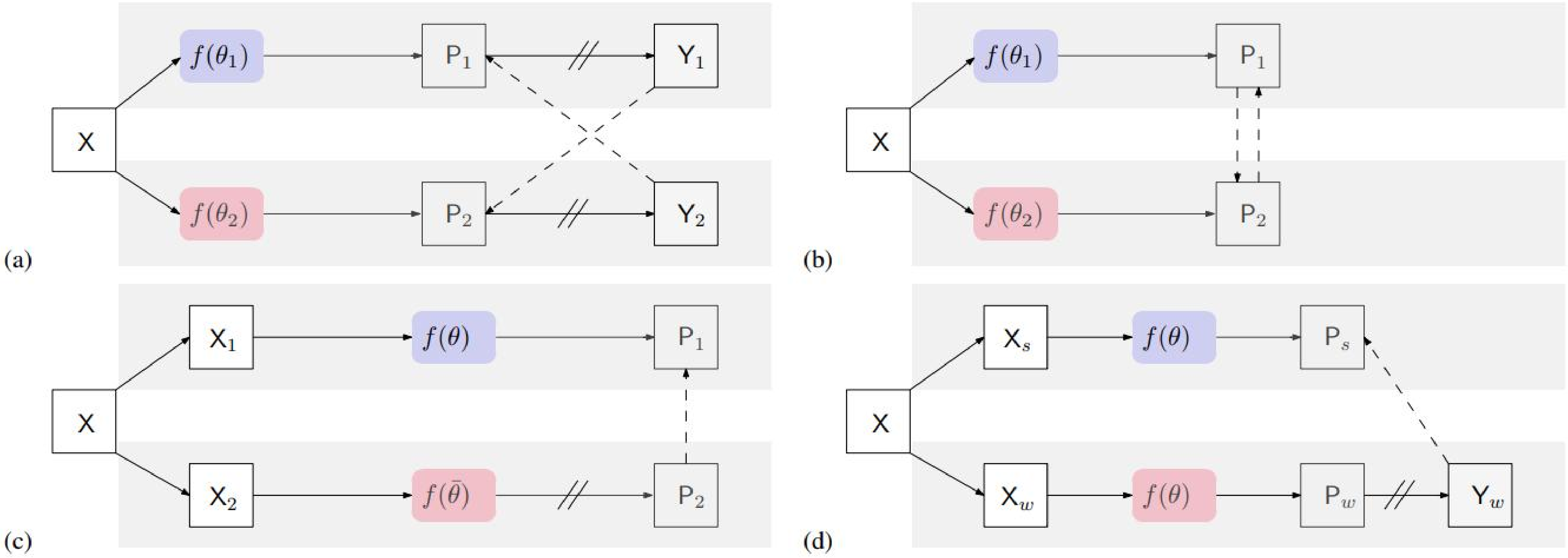
- Semi-Supervised Learning
  - labeled data for supervised learning
  - unlabeled data for unsupervised or pseudo supervised learning

# Background

- Semi-Supervised Learning
  - Multiple Branches
  - Unsupervised Contrast
  - Uncertainty / Attention
  - Co-training
  - Re-balancing
  - Correcting Network

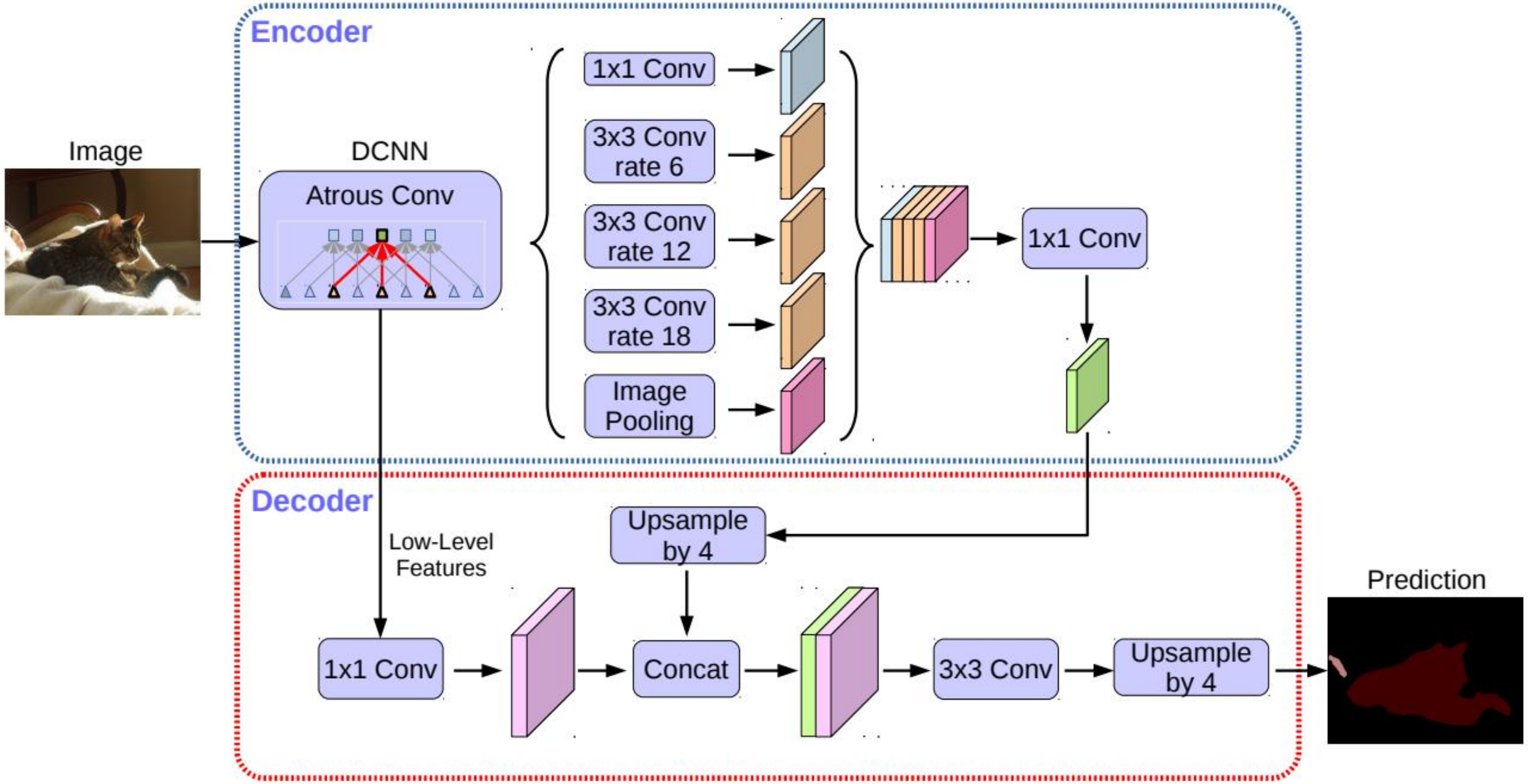
# Background

## ■ Semi-Supervised Learning



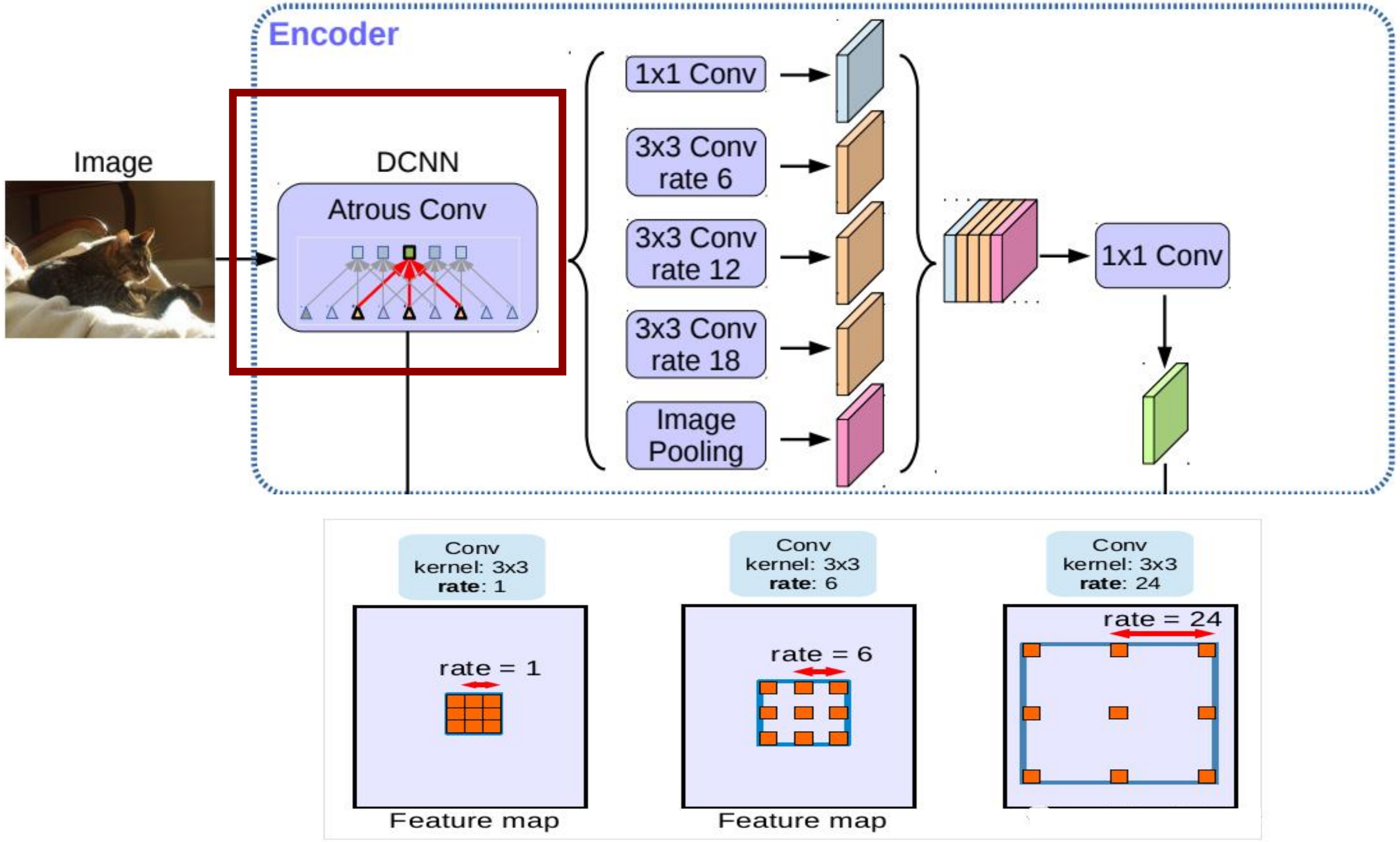
# Background

## ■ DeepLab V3+



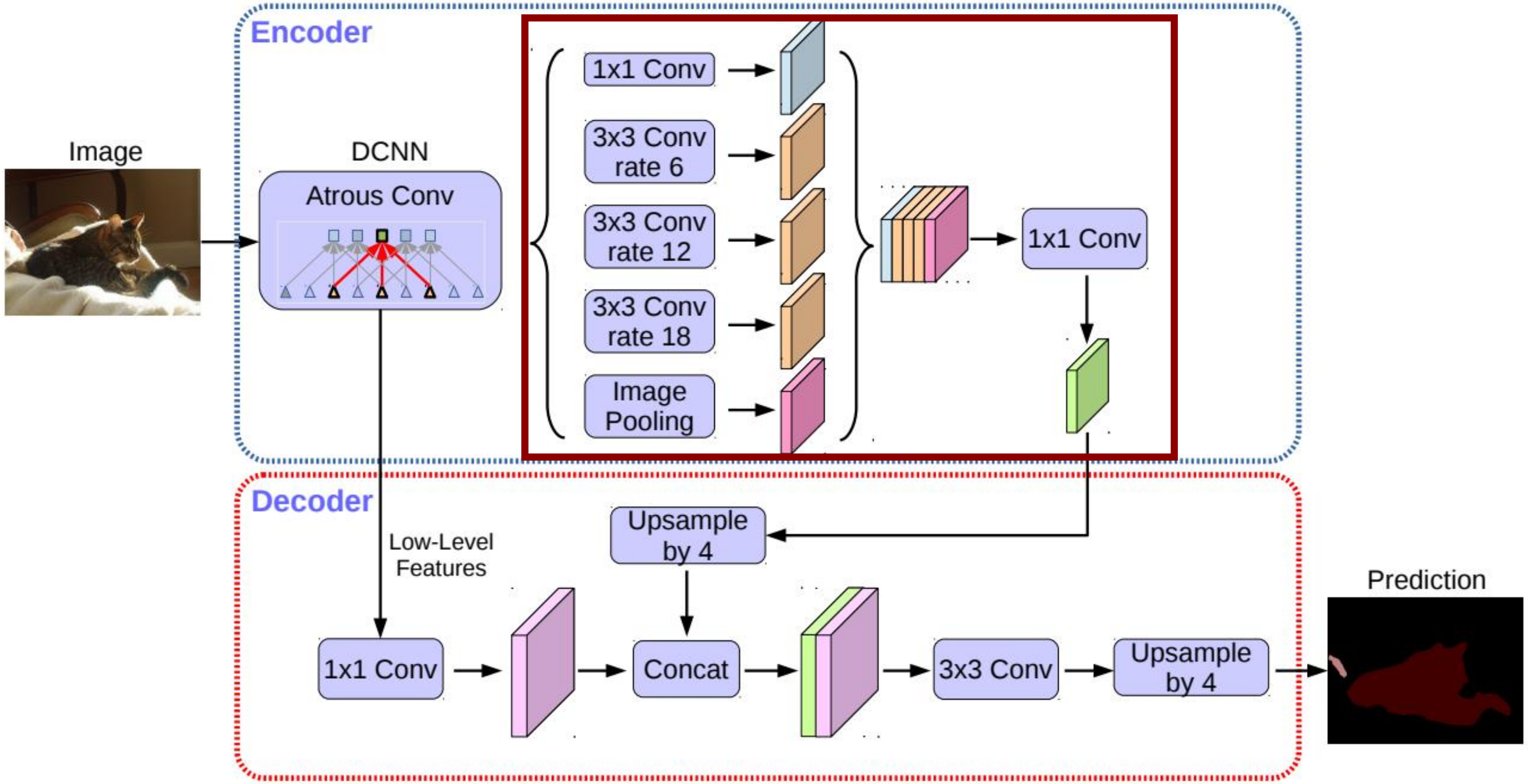
# Background

- **DeepLab V3+**



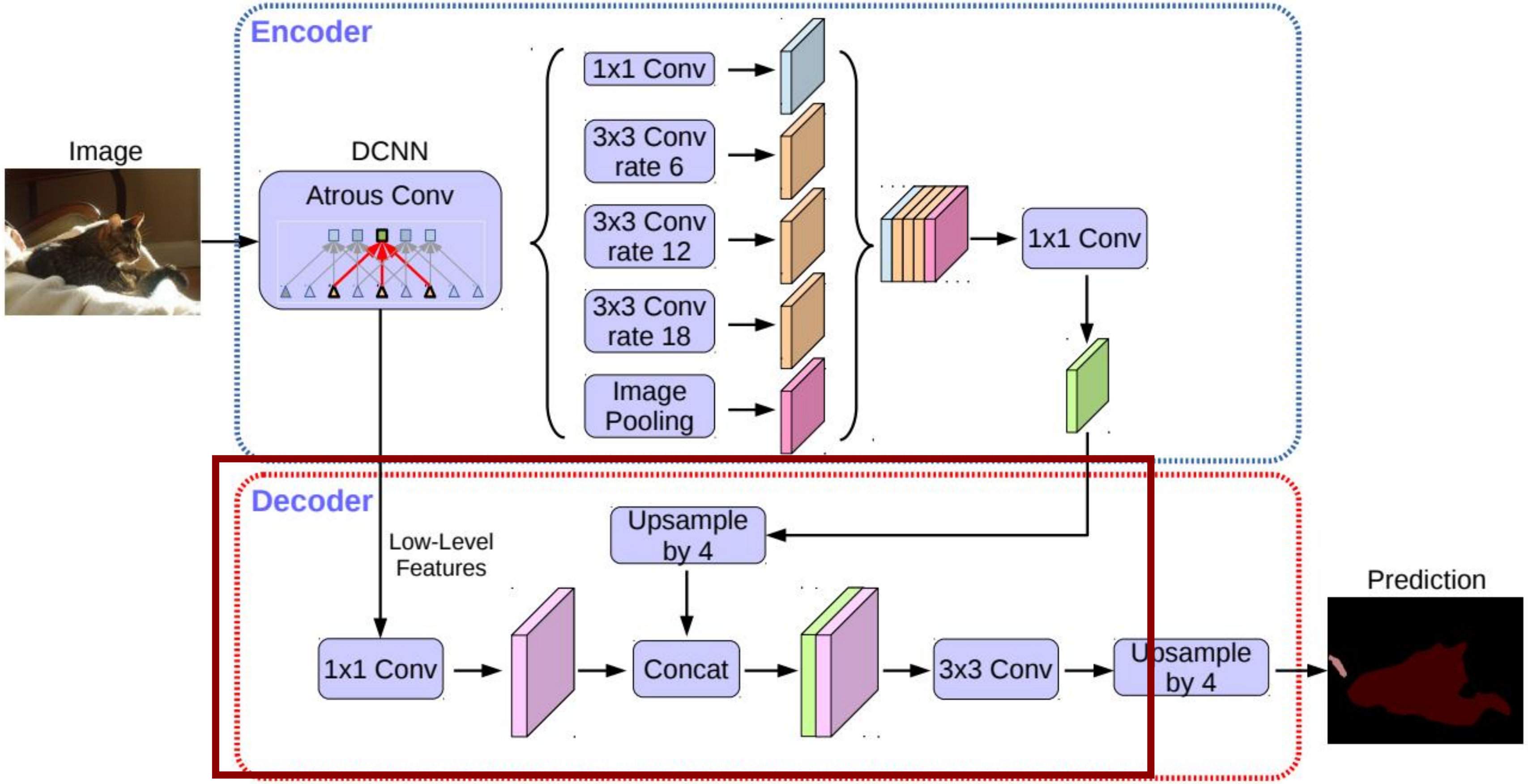
# Background

- **DeepLab V3+**



# Background

## ■ DeepLab V3+



# Outline

1 Authors

2 Background

3 **Method**

4 Experiments

5 Discussion

# Method

- **Motivation:**
  - **Propose a simple and clean approach for SSS.**
  - **Techniques**
    - **Only Augmentations!**
    - Multiple branches ✕
    - Unsupervised Contrast ✕
    - Uncertainty/Attention ✕
    - Re-balancing ✕
    - Correcting Network ✕

# Method

- **Problem Definition:**
  - **Labeled data** $(x_i, y_i)$
  - **Unlabeled data** $u_i$

- **Optimization objective:**

$$\mathcal{L} = \mathcal{L}_x + \lambda_u \mathcal{L}_u$$

# Method

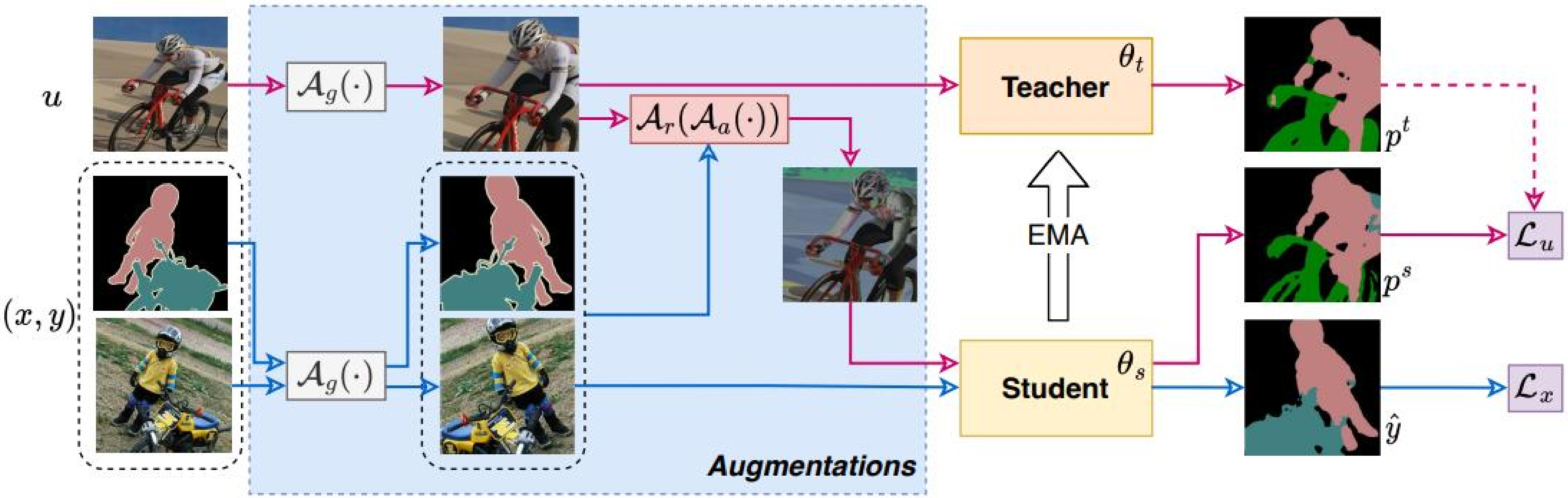
- **Problem Definition:**
  - **Labeled data** $(x_i, y_i)$
  - **Unlabeled data** $u_i$

- **Optimization objective:**

$$\mathcal{L} = \underbrace{\mathcal{L}_x}_{\text{Cross-Entropy Loss}} + \lambda_u \underbrace{\mathcal{L}_u}_{\text{Unsupervised Consistency Loss}}$$

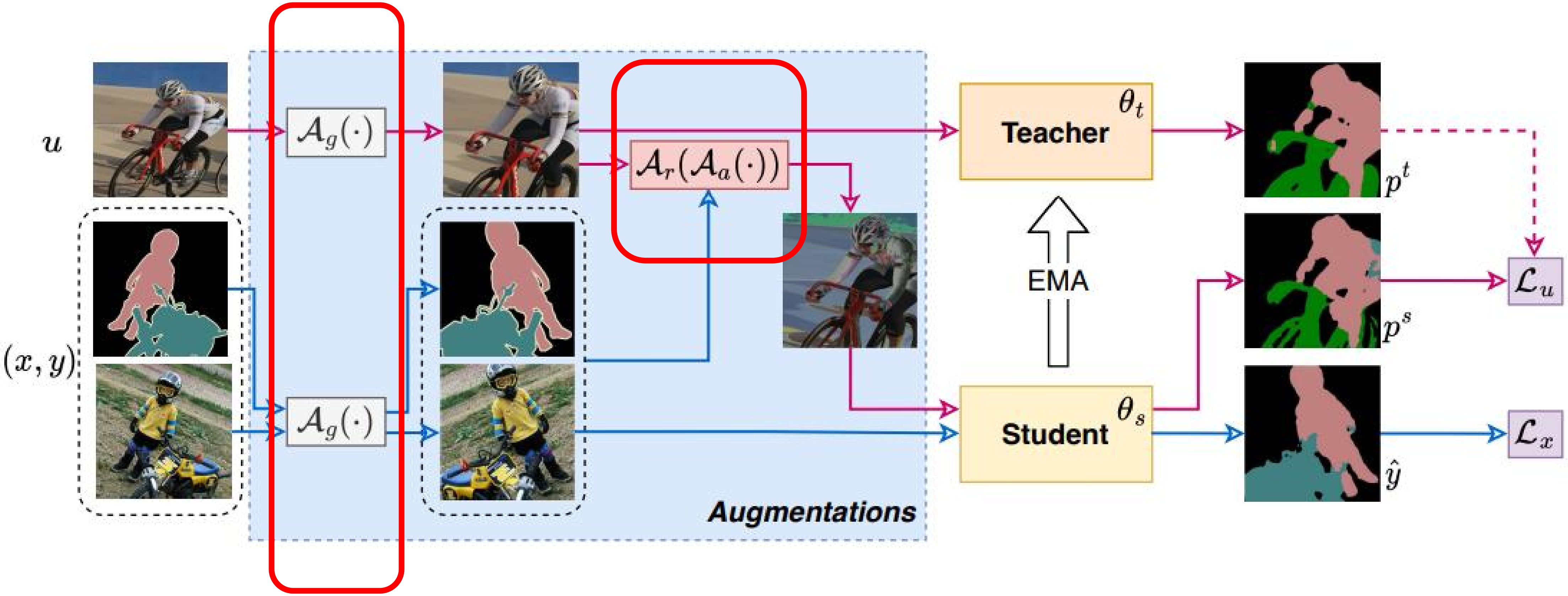
# Method

## Overview



# Method

## Overview



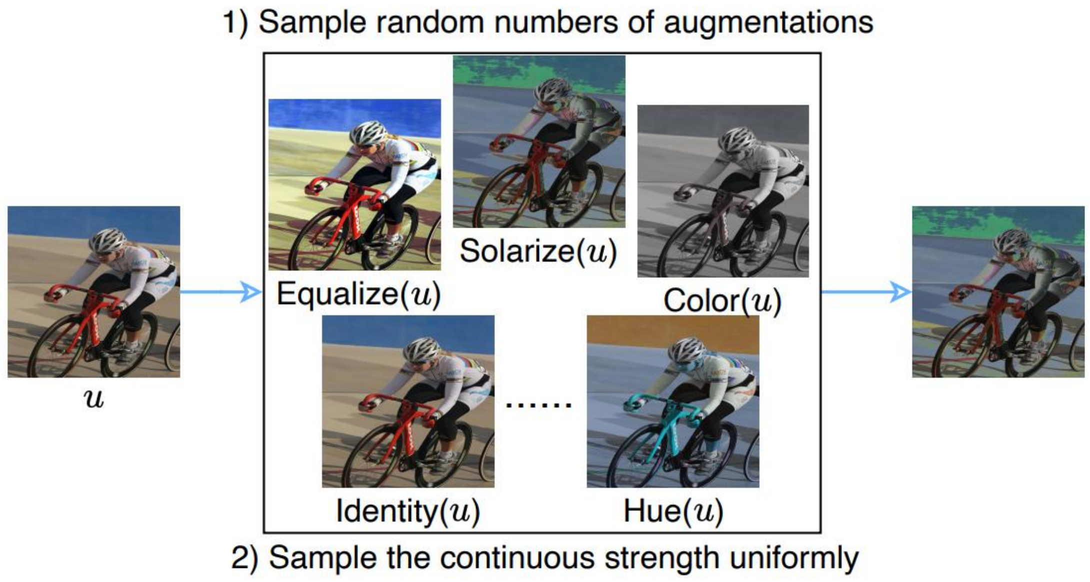
# Method

- "Weak " and "Strong" Augmentations
  - Weak aug. $\mathcal{A}_g(\cdot)$
    - Teacher branch
  - Strong aug. $\mathcal{A}_r(\mathcal{A}_a(\cdot))$
    - Student branch

| | |
|---|---|
| **Weak Geometrical Augmentation** - Apply all | |
| Random Scale | Randomly resizes the image by $[0.5, 2.0]$. |
| Random Flip | Horizontally flips the image with a probability of 0.5. |
| Random Crop | Randomly crops an region from the image. |
| **Random Intensity-based Augmentation** - Apply $k$ randomly | |
| Identity | Returns the original image. |
| Autocontrast | Maximizes (normalize) the image contrast. |
| Equalize | Equalize the image histogram. |
| Gaussian blur | Blurs the image with a Gaussian kernel. |
| Contrast | Adjusts the contrast of the image by [0.05, 0.95]. |
| Sharpness | Adjusts the sharpness of the image by [0.05, 0.95]. |
| Color | Enhances the color balance of the image by [0.05, 0.95] |
| Brightness | Adjusts the brightness of the image by [0.05, 0.95] |
| Hue | Jitters the hue of the image by [0.0, 0.5] |
| Posterize | Reduces each pixel to [4,8] bits. |
| Solarize | Inverts image pixels above a threshold from [1,256). |

# Method

- **Strong Augmentations (1/2)**
  - **Random Intensity-based Augmentations**
    - **Random distorting degree**
    - **Random augmentation number**



# Method

- **Strong Augmentations (2/2)**
  - **Adaptive Label-aided CutMix**
    - **Confidence Estimation**

$$\rho_i = \frac{1}{H \times W} \sum_{j=1}^{H \times W} \max(p_i^t(j))(1 - \frac{-\sum p_i^t(j) \log p_i^t(j)}{\log N})$$

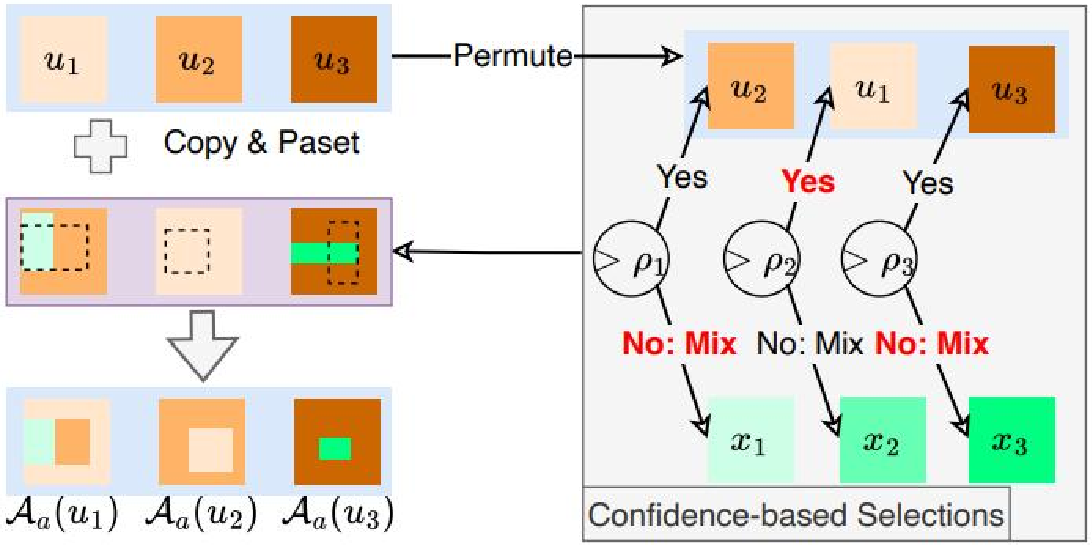
    - **Set it as the probability for CutMix**

# Method

- **Strong Augmentations (2/2)**
  - **Adaptive Label-aided CutMix**
    - **CutMix based on confidence**

$$u_n^{'} \leftarrow M_n \odot u_n + (\mathbf{1} - M_n) \odot x_n,$$

$$\mathcal{A}_a(u_m) \leftarrow M_m \odot u_m + (\mathbf{1} - M_m) \odot u_n^{'}$$



# Outline

# Experiments

## Main Results (1 / 2)

| Method | Encoder | 1/16 (92) | 1/8 (183) | 1/4 (366) | 1/2 (732) | Full (1464) |
|---|---|---|---|---|---|---|
| Supervised | R50 | 44.03 | 52.26 | 61.65 | 66.72 | 72.94 |
| PseudoSeg [58] | R50 | 54.89 | 61.88 | 64.85 | 70.42 | 71.00 |
| $PC^2Seg$ [56] | R50 | 56.90 | 64.63 | 67.62 | 70.90 | 72.26 |
| **AugSeg** | R50 | 64.22 | 72.17 | 76.17 | 77.40 | 78.82 |
| Supervised | R101 | 43.92 | 59.10 | 65.88 | 70.87 | 74.97 |
| CutMix-Seg [17] | R101 | 52.16 | 63.47 | 69.46 | 73.73 | 76.54 |
| PseudoSeg [58] | R101 | 57.60 | 65.50 | 69.14 | 72.41 | 73.23 |
| $PC^2Seg$ [56] | R101 | 57.00 | 66.28 | 69.78 | 73.05 | 74.15 |
| CPS [7] | R101 | 64.07 | 67.42 | 71.71 | 75.88 | - |
| PS-MT [36] | R101 | 65.80 | 69.58 | 76.57 | 78.42 | 80.01 |
| ST++ [50] | R101 | 65.20 | 71.00 | 74.60 | 77.30 | 79.10 |
| $U^2PL$ [47] | R101 | 67.98 | 69.15 | 73.66 | 76.16 | 79.49 |
| **AugSeg** | R101 | 71.09 | 75.45 | 78.80 | 80.33 | 81.36 |

Table 3. Compared with the state-of-the-art methods on classic Pascal VOC 2012 val set under different partition protocols. '1/n' means that '1/n' data is used as labeled dataset, and the remaining images are used as unlabeled dataset.

# Experiments

## Main Results (2 / 2)

| Method | ResNet-50 | | | ResNet-101 | | |
|---|---|---|---|---|---|---|
| | 1/16 (662) | 1/8 (1323) | 1/4 (2646) | 1/16 (662) | 1/8 (1323) | 1/4 (2646) |
| Supervised | 63.72 | 68.49 | 72.46 | 67.76 | 72.13 | 75.04 |
| MT [46] | 66.77 | 70.78 | 73.22 | 70.59 | 73.20 | 76.62 |
| CCT [44] | 65.22 | 70.87 | 73.43 | 67.94 | 73.00 | 76.17 |
| GCT [29] | 64.05 | 70.47 | 73.45 | 69.77 | 73.30 | 75.25 |
| CPS [7] | 68.21 | 73.20 | 74.24 | 72.18 | 75.83 | 77.55 |
| CPS w/ CutMix [7] | 71.98 | 73.67 | 74.90 | 74.48 | 76.44 | 77.68 |
| ST++ [50] | 72.60 | 74.40 | 75.40 | 74.50 | 76.30 | 76.60 |
| PS-MT [36] | 72.83 | 75.70 | 76.43 | 75.50 | 78.20 | 78.72 |
| **AugSeg** | 74.66 | 75.99 | 77.16 | 77.01 | 77.31 | 78.82 |
| Supervised‡ | 67.66 | 71.91 | 74.53 | 70.63 | 75.02 | 76.47 |
| $U^2$PL‡∗ [47] | 74.74 | 77.44 | 77.51 | 77.21 | 79.01 | 79.30 |
| **AugSeg‡** | 77.28 | 78.27 | 78.24 | 79.29 | 81.46 | 80.50 |

Table 4. Comparison with the state-of-the-art on the PASCAL VOCAug val set under different partition protocols. The VOCAug trainset consists of 10,582 labeled samples in total. ‡ means the same split as $U^2$PL, which prioritizes selecting high-quality labels from classic VOCs. Other methods use the same split as CPS. ∗ presents our reproduced results for $U^2$PL [47] using ResNet-50.

# Experiments

## Ablation Study

| AugSeg | | | mIoU | |
|---|---|---|---|---|
| MT | $\mathcal{A}_r$ | $\mathcal{A}_a$ | VOC (366) | Citys (744) |
| | | | 61.65 (supervised) | 74.14 (supervised) |
| ✓ | | | 69.06 (7.41↑) | 75.96 (1.82↑) |
| ✓ | ✓ | | 72.41 (10.76↑) | 77.29 (3.15↑) |
| ✓ | | ✓ | 74.33 (12.68↑) | 77.44 (3.30↑) |
| ✓ | ✓ | ✓ | **76.17** (14.52↑) | **78.76**(4.62↑) |

Table 6. Ablation studies on our AugSeg. "MT" means the standard mean-teacher semi-supervised framework. $\mathcal{A}_r$ and $\mathcal{A}_a$ represent the two main augmentation strategies, the random intensity-based and adaptive label-aided augmentations, respectively. Improvements over the supervised baseline are highlighted in blue.

# Experiments

## Ablation Study

| $\lambda_u$ | 0.0 | 0.5 | 1.0 | 1.5 | 2.0 |
|---|---|---|---|---|---|
| VOC (366) | 61.65 | 75.21 | 76.17 | 75.95 | **77.05** |
| Citys (744) | 74.14 | 77.02 | 78.76 | **78.99** | 78.68 |

Table 7. Ablations on the loss weight $\lambda_u$, set as 1.0 by default.

| $k$ | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| VOC (366) | 74.38 | 75.50 | 76.10 | 76.17 | **76.32** |
| Citys (186) | 71.26 | 72.10 | 73.42 | **73.73** | 73.03 |
| Citys (744) | 77.44 | 78.34 | 78.11 | **78.76** | 78.48 |

Table 8. Ablations on the maximum number of selected intensity-based augmentations, using R50 as the encoder. $k = 3$ by default.

# Experiments

## ■ Qualitative Results

Original Image GT Supervised Mean Teacher Ours

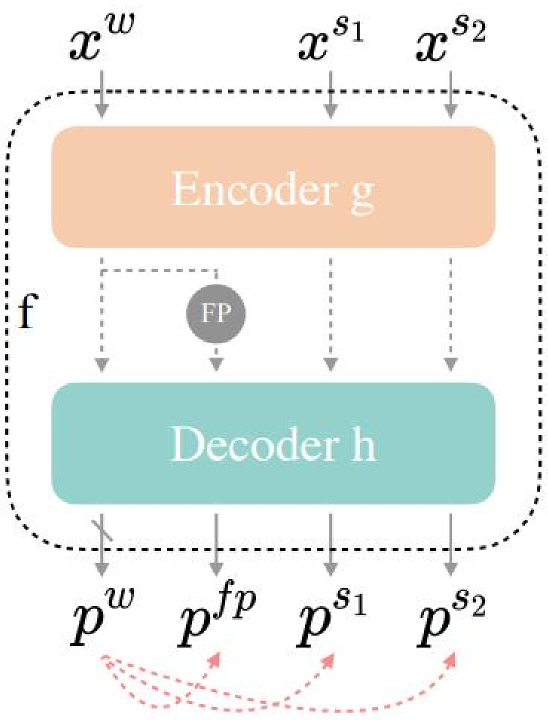
# Outline

# Discussion

- Review the proposed AugSeg framework
    - Simple two-branch model architecture
    - Weak-Strong augmentations
    - SOTA performance only by augmentations

# Discussion

- **Comparison to the concurrent work UniMatch, CVPR 23**
  - **Weak-Strong augmentations**
  - **More augmentations**




(b) UniMatch

**Revisiting Weak-to-Strong Consistency in Semi-Supervised Semantic Segmentation**

Lihe Yang[1] Lei Qi[2] Litong Feng[3] Wayne Zhang[3] Yinghuan Shi[1*]
[1]Nanjing University [2]Southeast University [3]SenseTime Research

# Thanks!