ECCV 2020 Workshop Paper Summary

1등부터 5등 + Hoya님의 솔루션 정리

1. Overview
(1) Overview

Overview

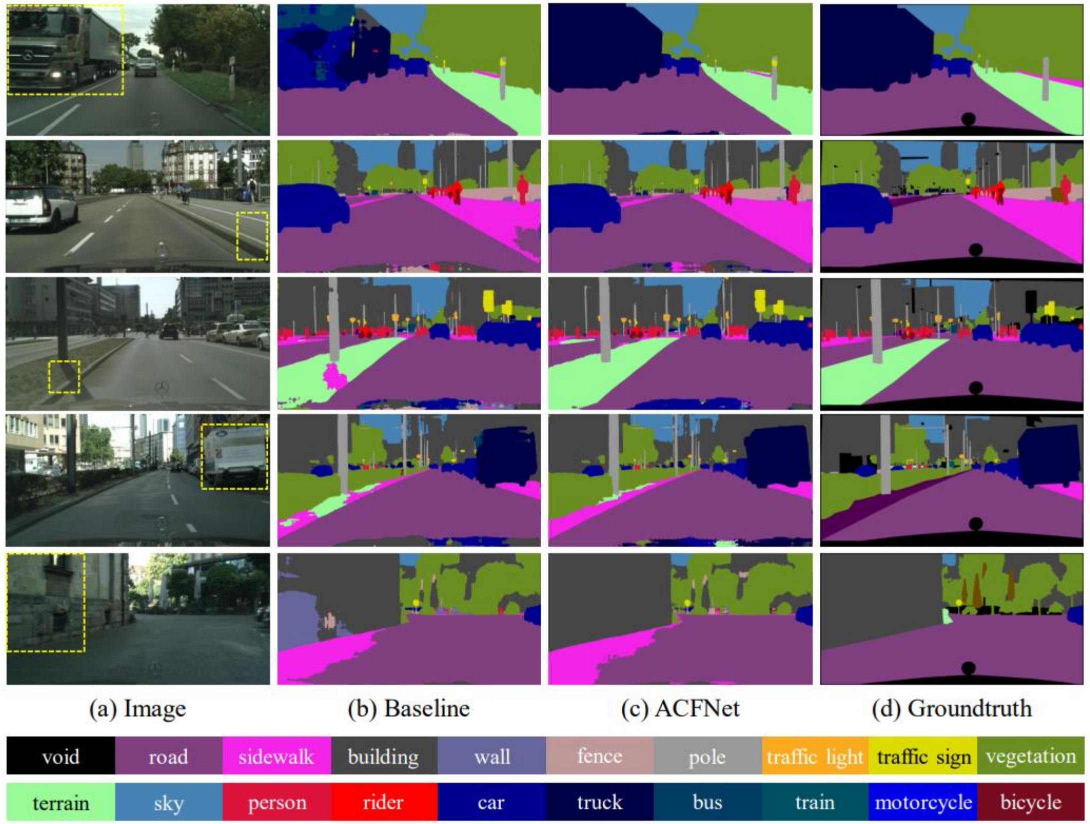


Figure 7. Visualization results of ACFNet based on ResNet-101 network on Cityscapes val set.

Original : Training Set 2975장, Validation Set 500장, Test Set 1525장

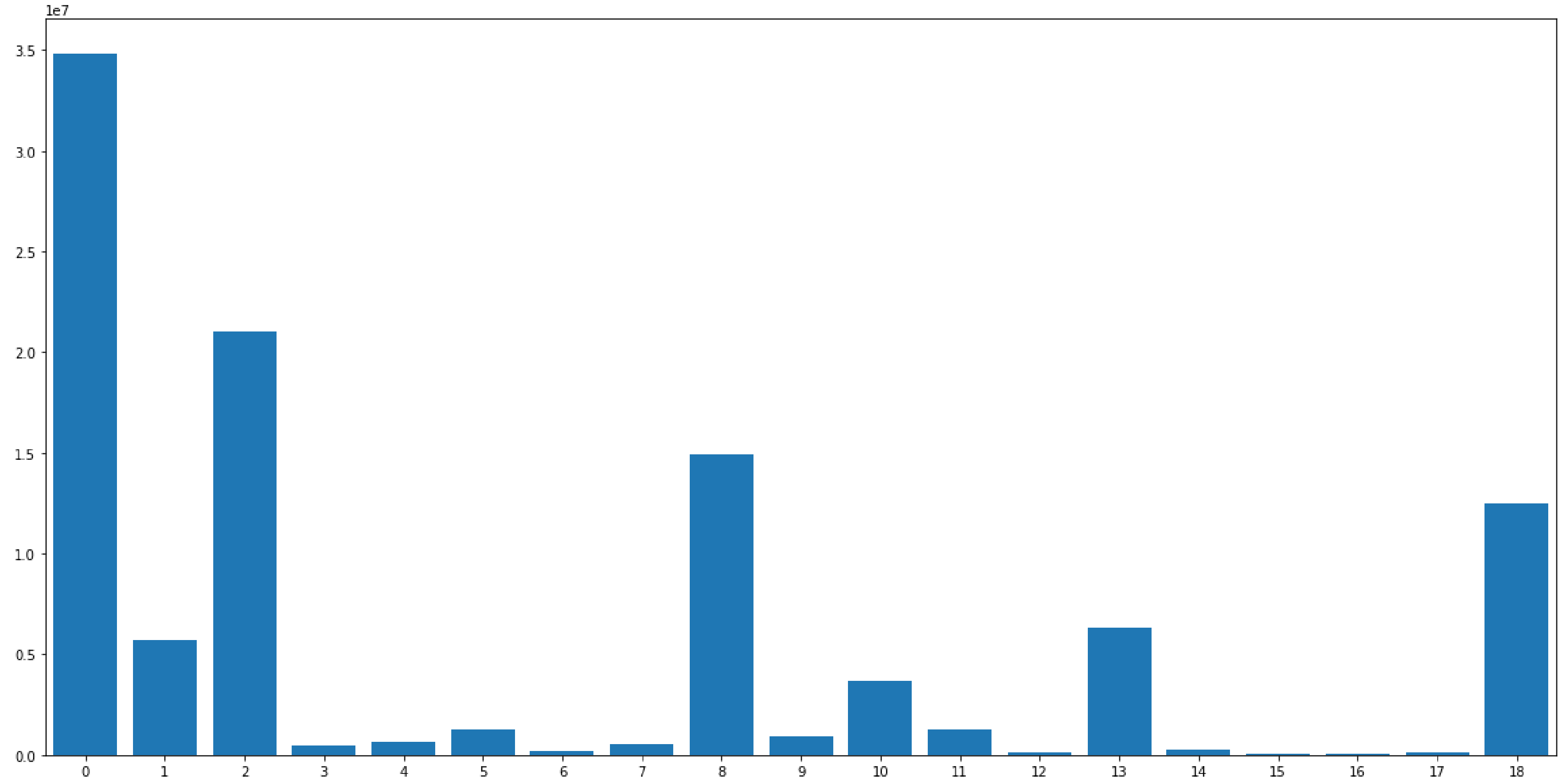
MiniCity : Training Set 200장, Validation Set 100장, Test Set 200장

출처 : https://hoya012.github.io/blog/segmentation_tutorial_pytorch/

1. Overview

(1) Overview

Overview



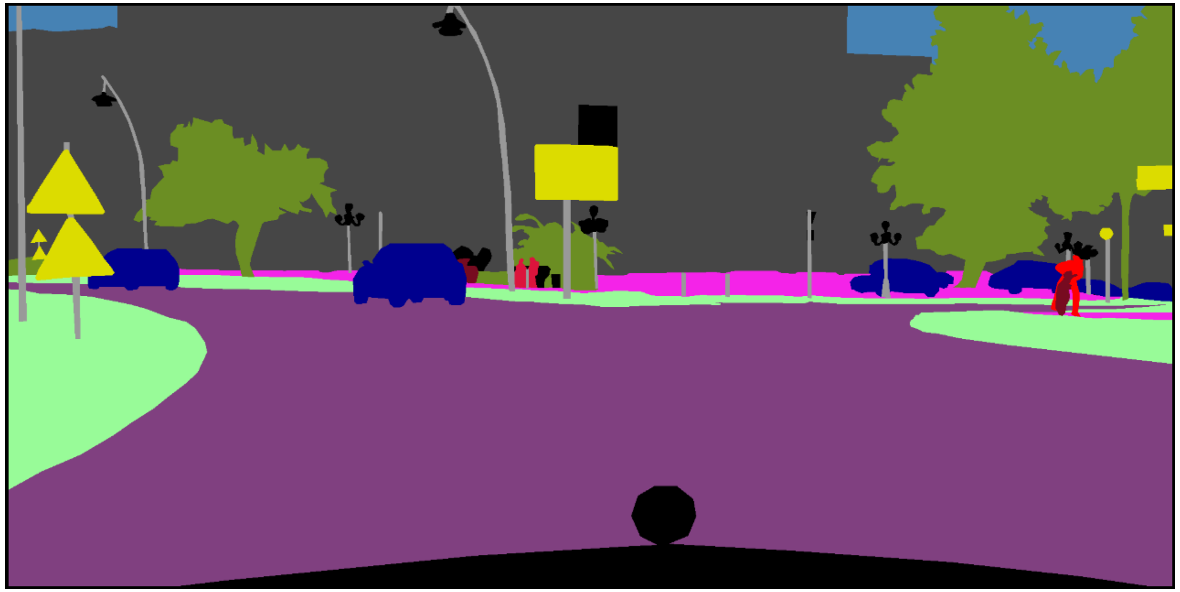
classes		IoU	nIoU
road	:	0.963	nan
sidewalk	:	0.762	nan
building	:	0.856	nan
wall	:	0.120	nan
fence	:	0.334	nan
pole	:	0.488	nan
traffic light	:	0.563	nan
traffic sign	:	0.631	nan
vegetation	:	0.884	nan
terrain	:	0.538	nan
sky	:	0.901	nan
person	:	0.732	0.529
rider	:	0.374	0.296
car	:	0.897	0.822
truck	:	0.444	0.218
bus	:	0.244	0.116
train	:	0.033	0.006
motorcycle	:	0.492	0.240
bicycle	:	0.638	0.439
Score Average	:	0.573	0.333

출처 : https://hoya012.github.io/blog/segmentation_tutorial_pytorch/

2. Hoya's Solution

(1) Training

Training



1024

2048

최근 SOTA

- Batch 1 1024X2048 학습

Hoya (2080ti 4대)

- Batch 8 : ResNet50 DeepLabv3 (crop 576 x 1152)
- Batch 8 : ResNet101 DeepLabv3 (crop 384 x 786)

Loss

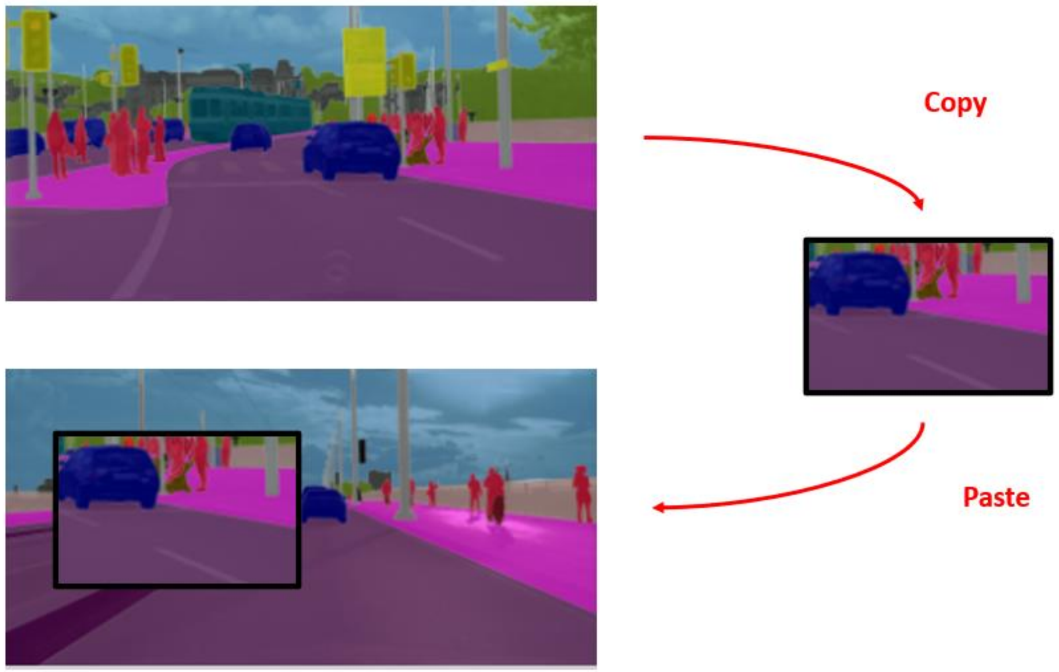
- CrossEntropy Loss

Unbalanced에 좋은 Loss

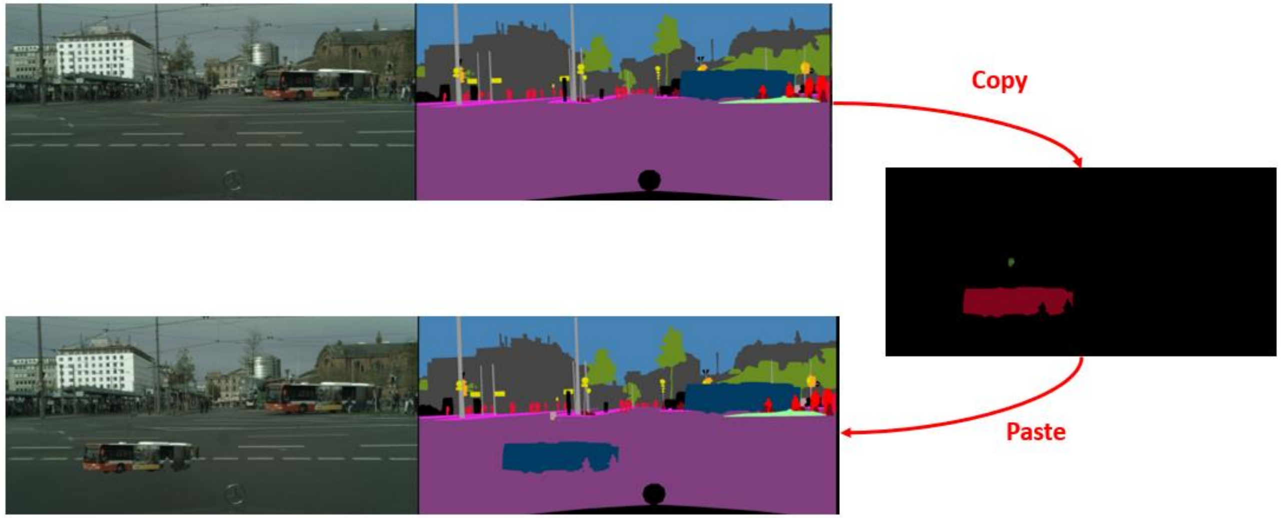
- Focal Loss
- Weighted Cross Entropy Loss

Augmentation

"CutMix" Augmentation



"Copy Blob" Augmentation



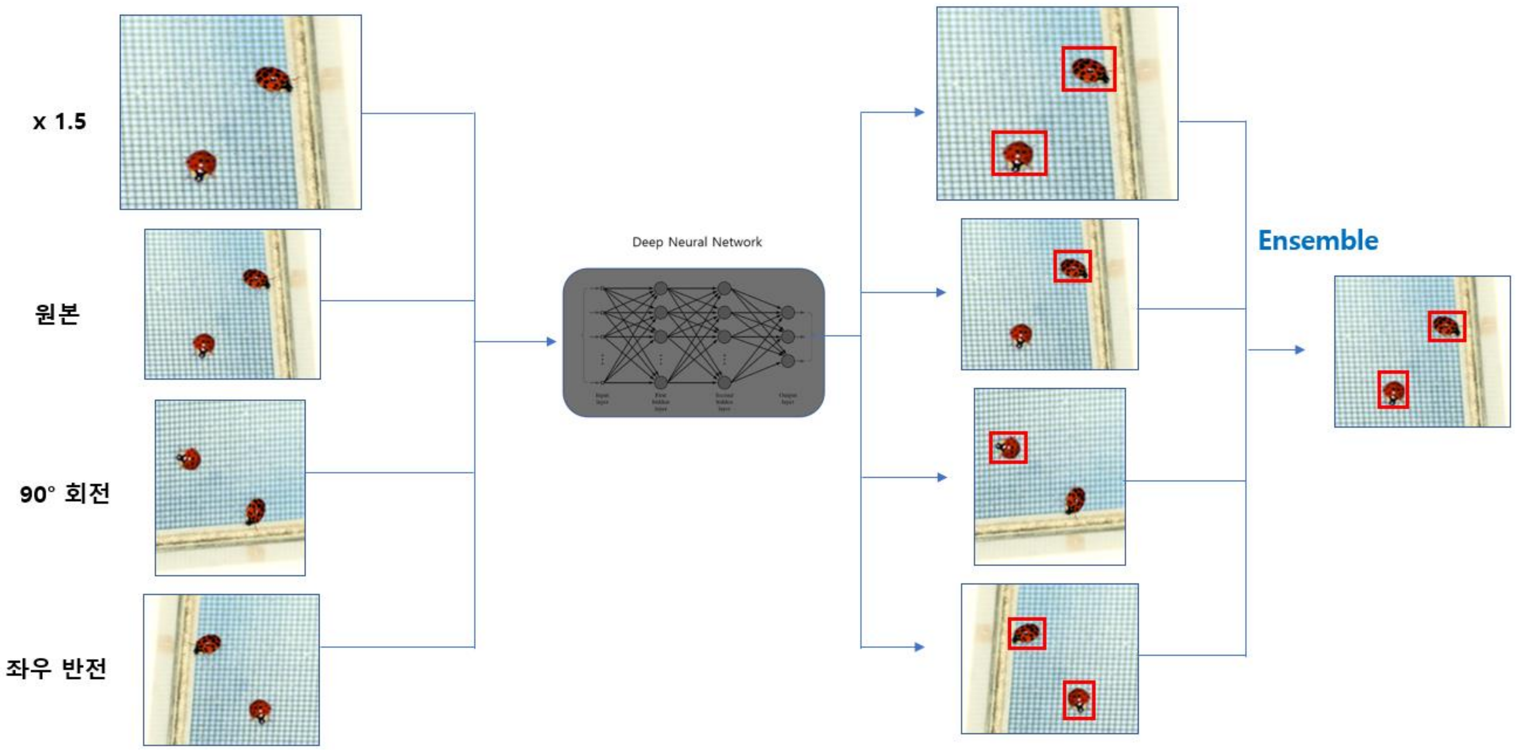
Visual Inductive Prior

Wall is located on the **sidewalk**
Fence is located on the **sidewalk**

Bus is located on the **Road**
Train is located on the **Road**

2. Hoya's Solution
(3) Inference

Multi Scale Inference



3. Top Solution

(1) 1st Solution

Multi-level tail pixel cutmix and scale attention for long-tailed scene parsing

문제
- Long-tailed Label Distribution (Unbalanced)

Previsious Work
- Re-sampling
- Re-weighting

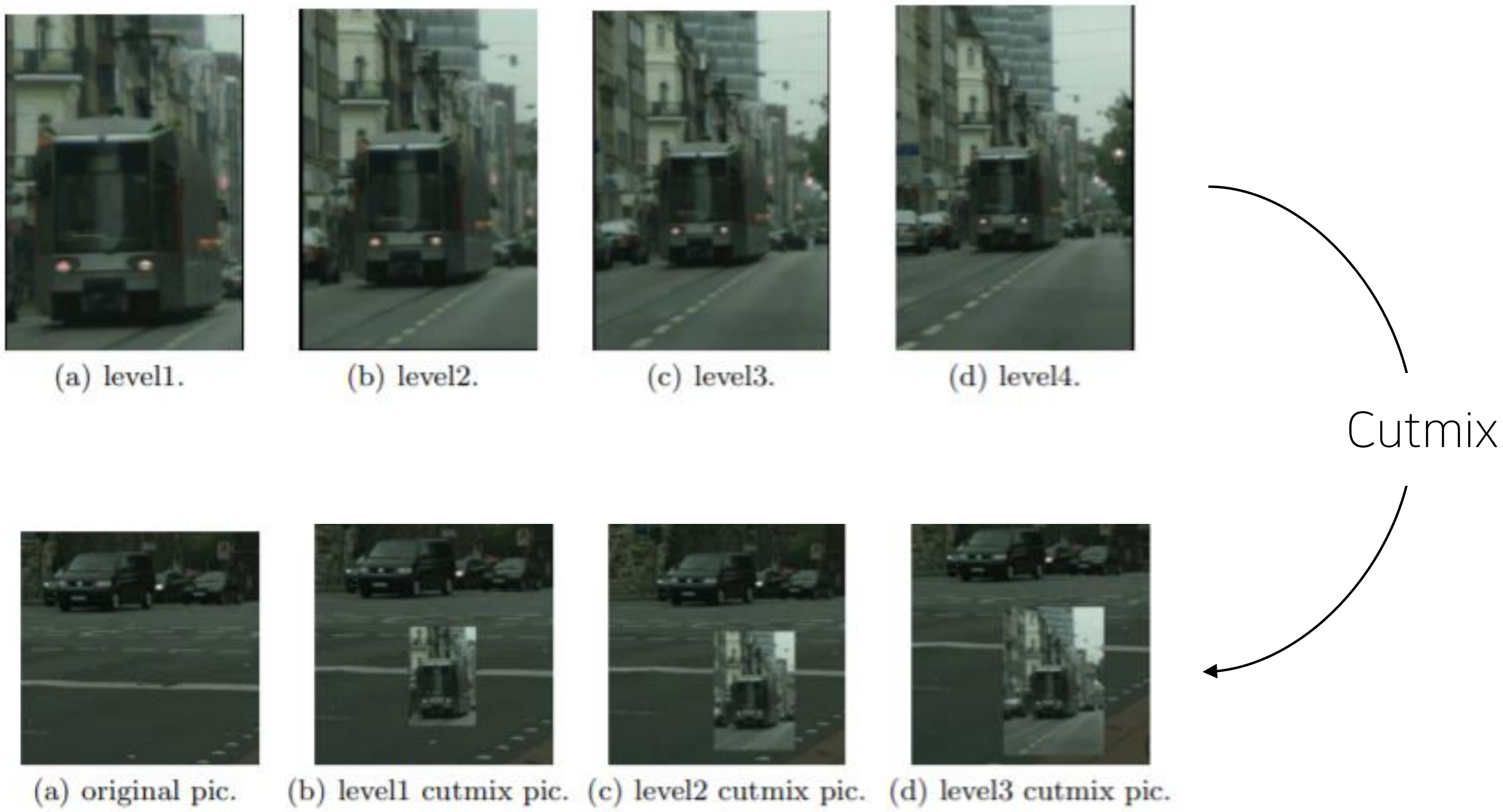
한계
- 전체 이미지에서 특정 픽셀에 대해서 sampling 하는 것이 힘듦 (balancing 불가 <- sampling)
- Weight 같은 경우도 Random Crop / Resize 등에 의해서 Weight Map이 변경되는 문제 발생

제안
- Long tail pixel distribution 을 이용한 Multi-level tail pixel cutmix
- New cutmix with scale attention model 사용

출처 : Weitao Chen, Zhibing Wang 1st Visual Inductive Priors for Data-Efficient Deep Learning workshop at ECCV 2020: semantic segmentation Challenge Track Technical Report: Multi-level tail pixel cutmix and scale attention for long-tailed scene parsing /

3. Top Solution
(1) 1st Solution

Multi-level tail pixel cutmix



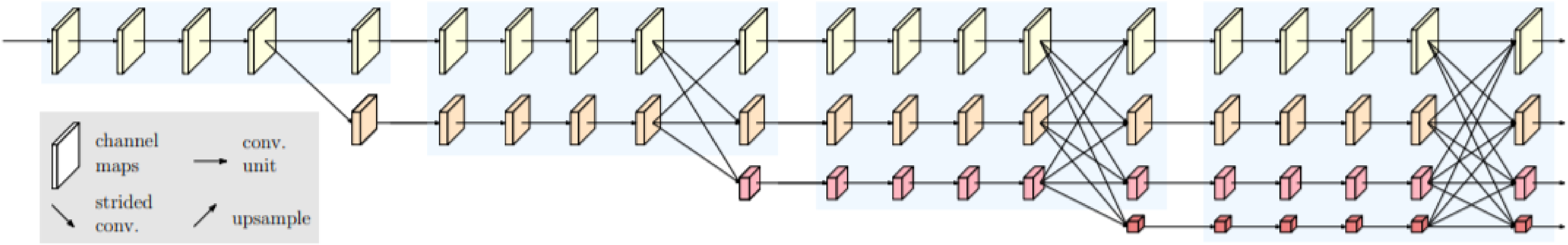
출처 : Weitao Chen, Zhibing Wang 1st Visual Inductive Priors for Data-Efficient Deep Learning workshop at ECCV 2020: semantic segmentation Challenge Track Technical Report: Multi-level tail pixel cutmix and scale attention for long-tailed scene parsing /

3. Top Solution

(1) 1st Solution

Scale Attention

출처 : Jingdong Wang, Ke Sun, Tianheng Cheng, Borui Jiang, Chaorui Deng, Yang Zhao, Dong Liu, Yadong Mu, Mingkui Tan, Xinggang Wang, Wenyu Liu, Bin Xiao, Deep High-Resolution Representation Learning for Visual Recognition



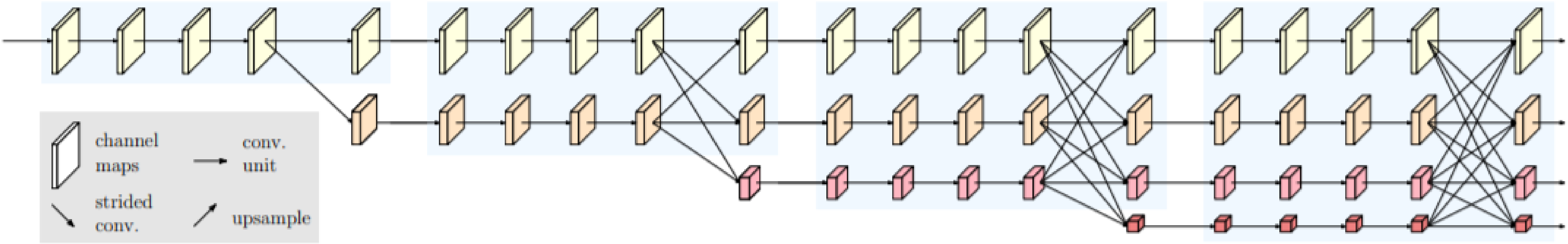
Convolution Layer -> 3x3Conv(BN) – ReLU – 3x3Conv(BN) – ReLU – 1x1Conv

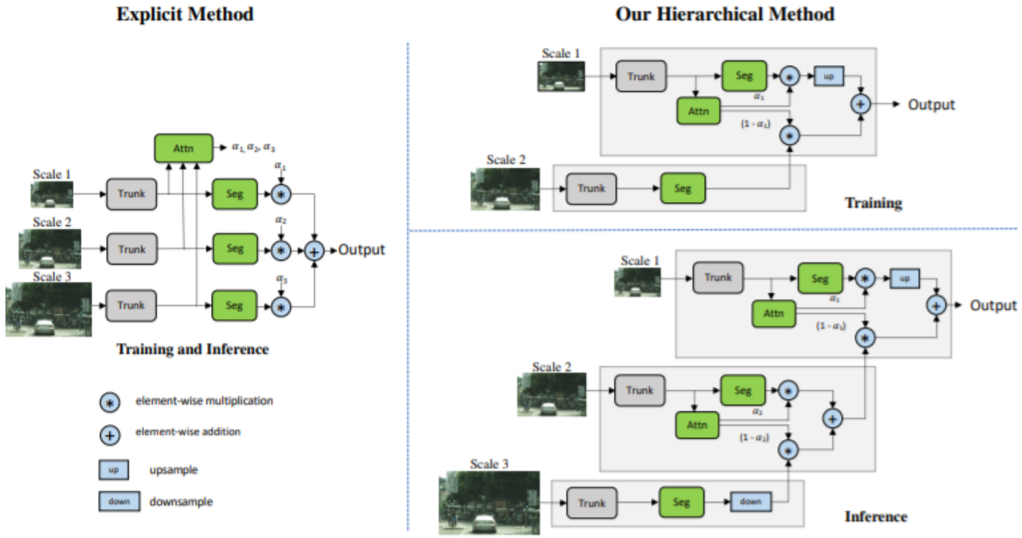
3. Top Solution

(1) 1st Solution

Scale Attention

출처 : Jingdong Wang, Ke Sun, Tianheng Cheng, Borui Jiang, Chaorui Deng, Yang Zhao, Dong Liu, Yadong Mu, Mingkui Tan, Xinggang Wang, Wenyu Liu, Bin Xiao, Deep High-Resolution Representation Learning for Visual Recognition





모든 클래스가 효과가 있었던 것은 아니여서 효과가 없는 클래스는 train만 Scale Attention 방법을 적용

출처 : Jingdong Wang, Ke Sun, Tianheng Cheng, Borui Jiang, Chaorui Deng, Yang Zhao, Dong Liu, Yadong Mu, Mingkui Tan, Xinggang Wang, Wenyu Liu, Bin Xiao, Deep High-Resolution Representation Learning for Visual Recognition

출처 : Weitao Chen, Zhibing Wang 1st Visual Inductive Priors for Data-Efficient Deep Learning workshop at ECCV 2020: semantic segmentation Challenge Track Technical Report: Multi-level tail pixel cutmix and scale attention for long-tailed scene parsing /

Diversification is All You Need : Towards Data Effcient Image Understanding

문제
- Lack of enough number of labeled images

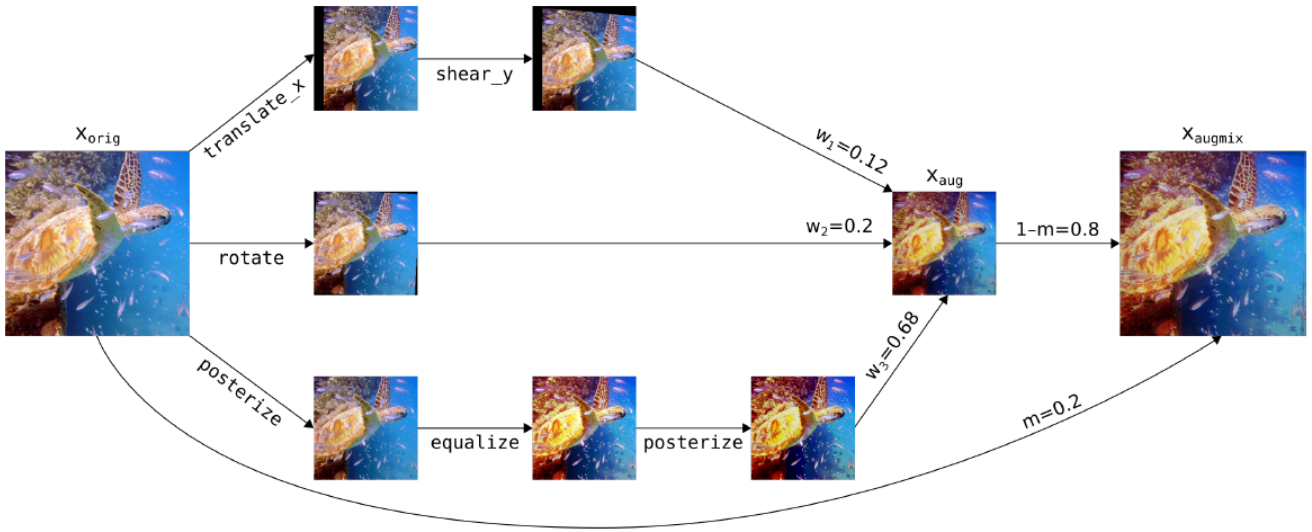
제안
- Augmix + mixup within class diversification and between class diversification
- seeAugmix : Extend Augmix
- Frequency weighted model ensemble

출처 : Diversification is All You Need: Towards Data Efficient Image Understanding

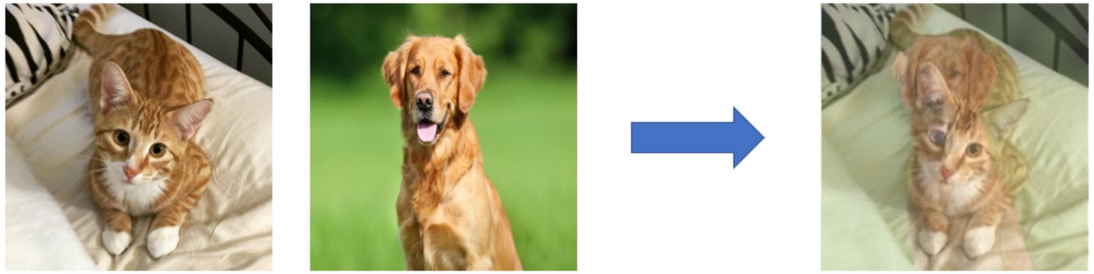
3. Top Solution

(2) 2nd Solution

Augmix + mixup within class diversification and between class diversification



[Aug mix]

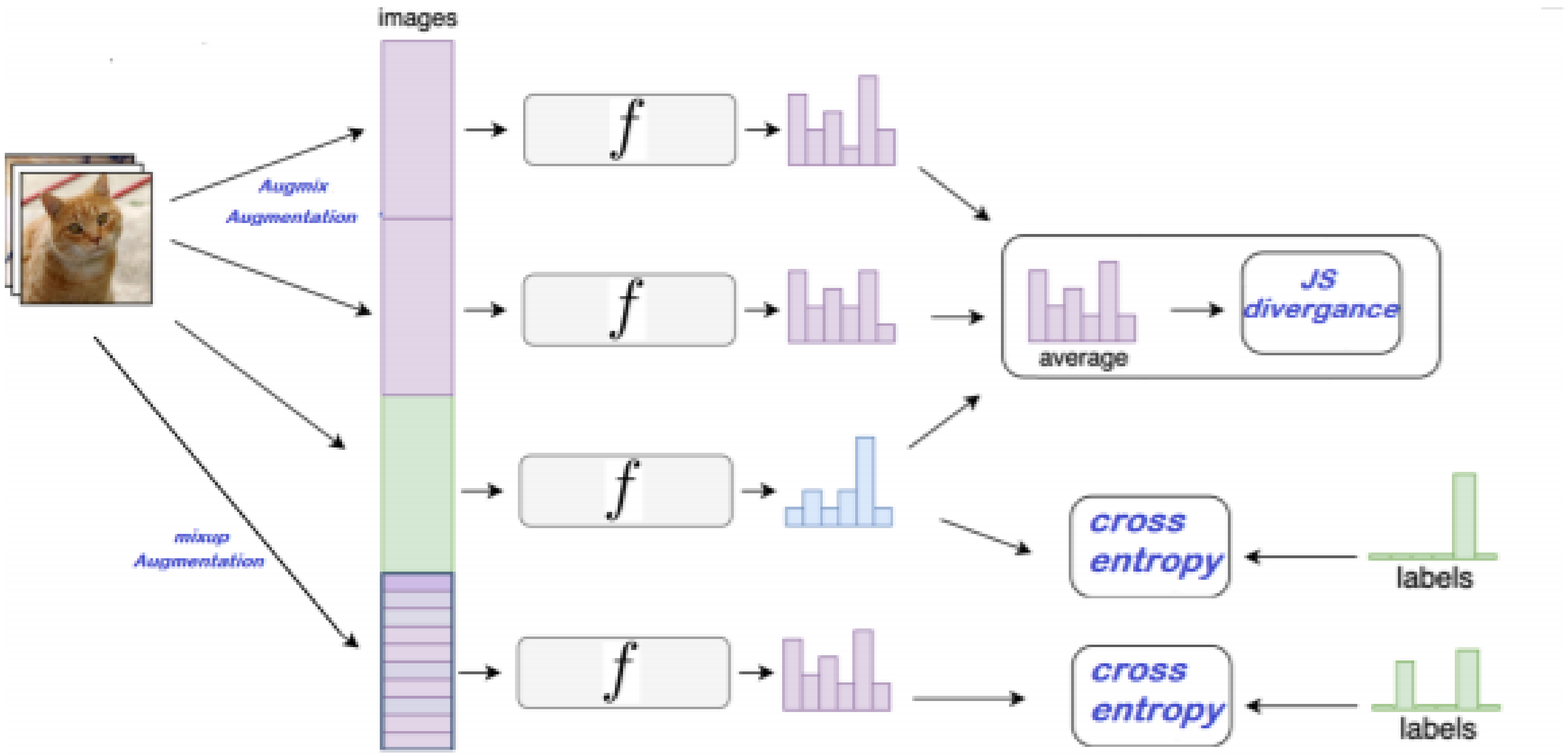


[Mixup]

3. Top Solution

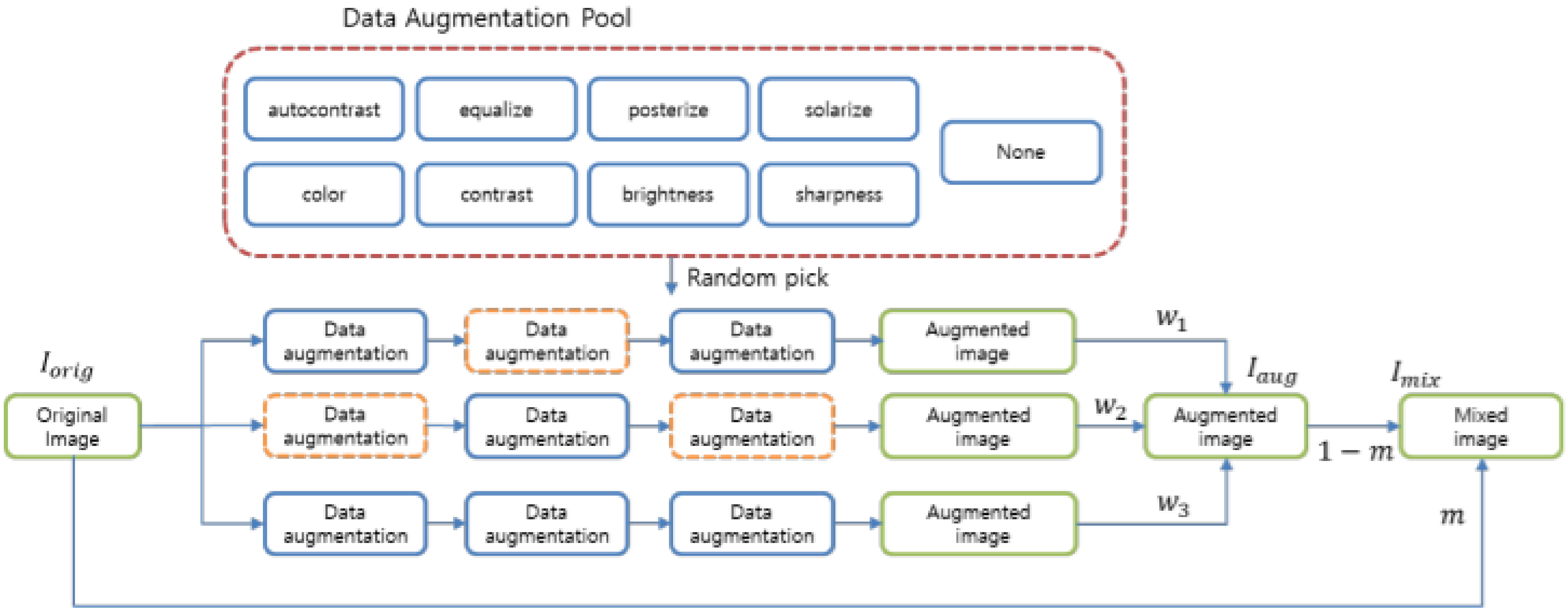
(2) 2^{nd} Solution

Augmix + mixup within class diversification and between class diversification



출처 : Diversification is All You Need: Towards Data Efficient Image Understanding

seeAugmix : Extend Augmix



출처 : Diversification is All You Need: Towards Data Efficient Image Understanding

Frequency Weighted Ensemble

Method	mIoU (%) validation dataset
HRNetV2-W48 + OCR	59.21
HRNetV2-W48 + OCR + OHEM	59.25
HRNetV2-W48 + OCR + seg-Augmix	59.84
Average ensemble of above	59.85
FW model ensemble	60.60

Table 7. The performance of HRNetV2-W48 + OCR, HRNetV2-W48 + OCR + OHEM, HRNetV2-W48 + OCR + seg-Augmix models, average model ensemble of them, and Frequency Weighted model ensemble of them on validation dataset of MiniCity.

출처 : Diversification is All You Need: Towards Data Efficient Image Understanding

3. Top Solution
(3) 3rd Solution

Edge-Preserving Guided Semantic Segmentation for VIPriors Challenge

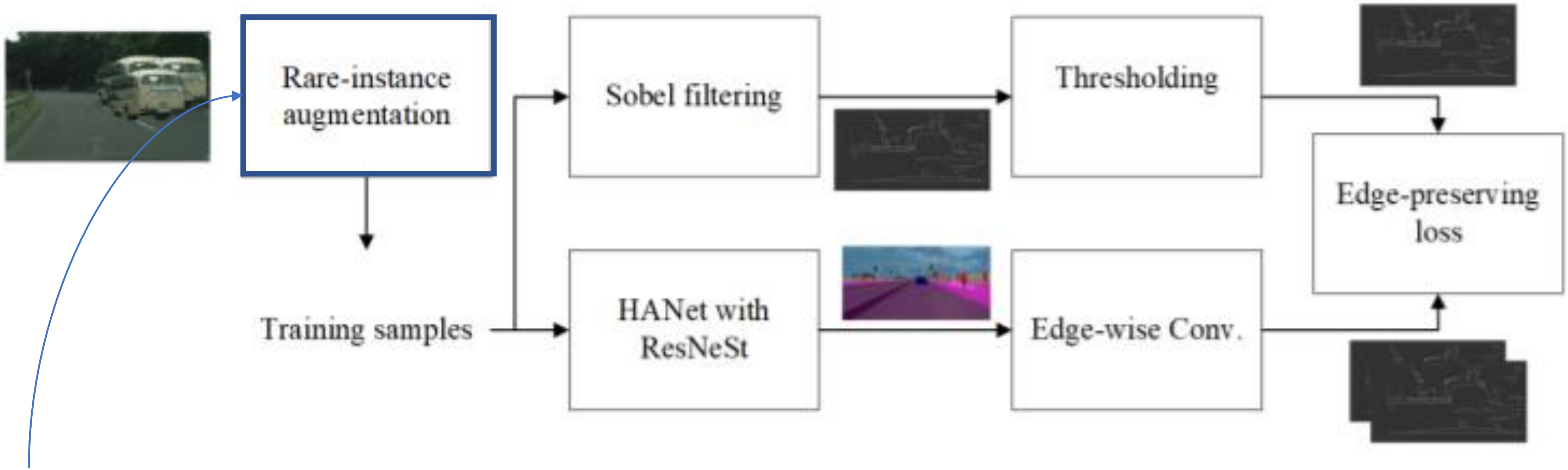
문제
- 학습시킬 데이터의 부족

제안
- Edge-preserving guidance to obtain the extra prior information (prior 정보로 edge 활용)
  - Two channeled convolution layer is concatenated to the last layer
  - Edge map is calculated from the ground truth by sobel operation
  - Sobel operation followed by concatenating a hard thresholding operation to indicate whether the pixel edge or not
  - Calculate the loss "Predicted edge map – edge preserving loss"

출처 : Chih-Chung Hsu, Hsin-Ti Ma Edge-Preserving Guided Semantic Segmentation for VIPriors Challenge

3. Top Solution
(3) 3rd Solution

Edge-Preserving Guidance to obtain the extra prior information



클래스의 라벨이 적은 "rare"한 instance만 augmentation 하는 방법

3. Top Solution
(4) 4th Solution

EfficientSeg: An Efficient Semantic Segmentation Network

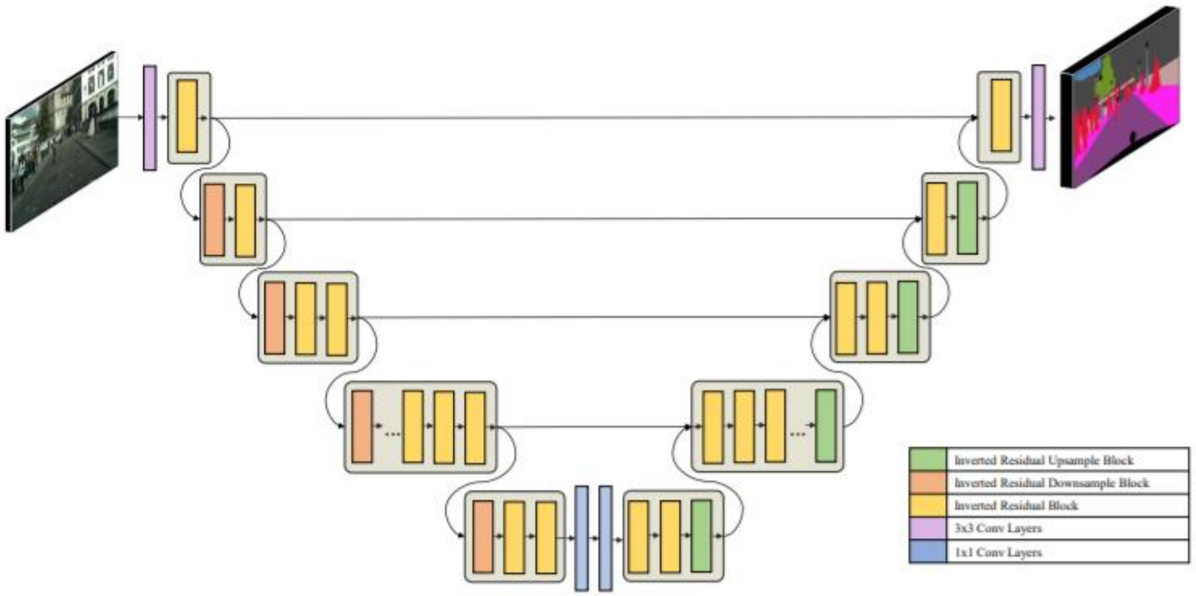


Fig. 2. EfficientSeg architecture. There are 5 different type of blocks. Inverted Residual Blocks are MobileNetV3 blocks described as in the paper. 1x1 and 3x3 blocks are standard convolution blocks which has activation and batch normalization. Downsampling operations are done with increasing the stride and for upsampling, linear interpolation is used.

3. Top Solution

(5) 5^{th} Solution

- Data-efficient semantic segmentation via extremely perturbed data augmentation

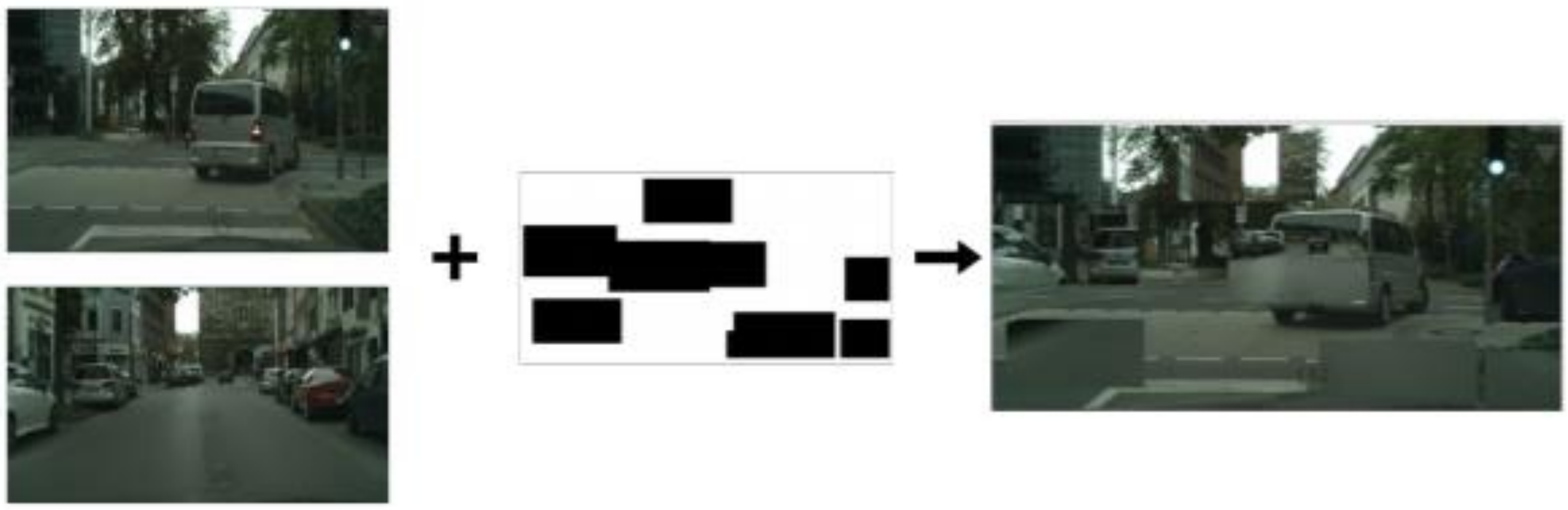


Fig. 3: Example of CutMix Sprinkles in semantic segmentation setup.