MATHÉMATIQUES VISION APPRENTISSAGE

école normale supérieure paris-saclay | université PARIS-SACLAY

Project Presentation

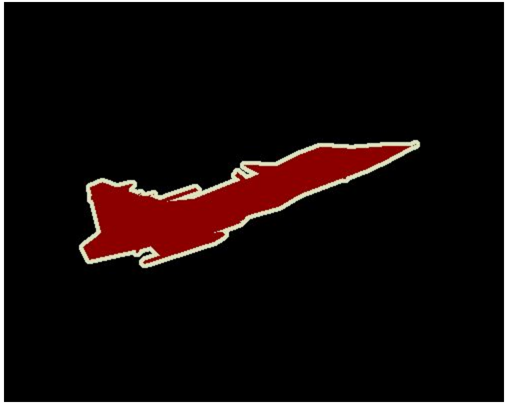
Weakly Supervised Semantic Segmentation

Vincent LIU liuvincent25@gmail.com
Supervised by V.LEPETIT & M.VAKALOPOULOU

Semantic Segmentation

Assign a category label to each pixel of an image.

Input

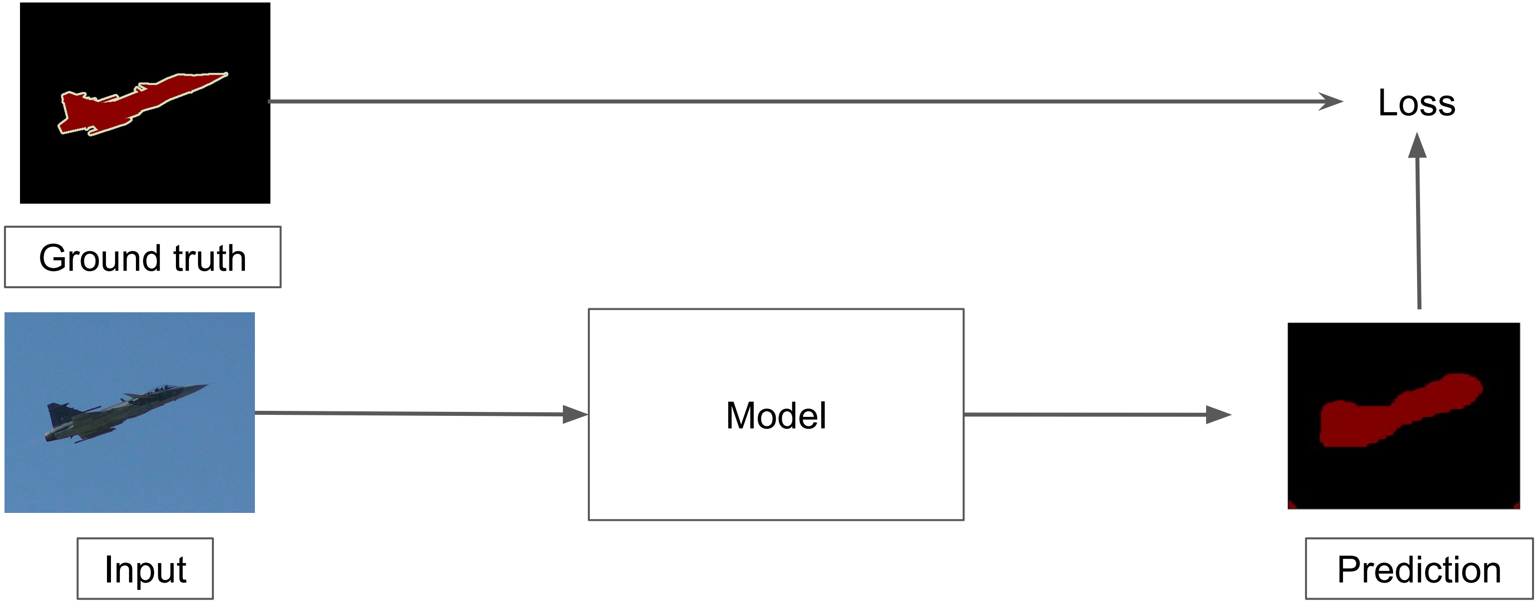


Target

Pictures taken from Pascal VOC 2012 Dataset [12].

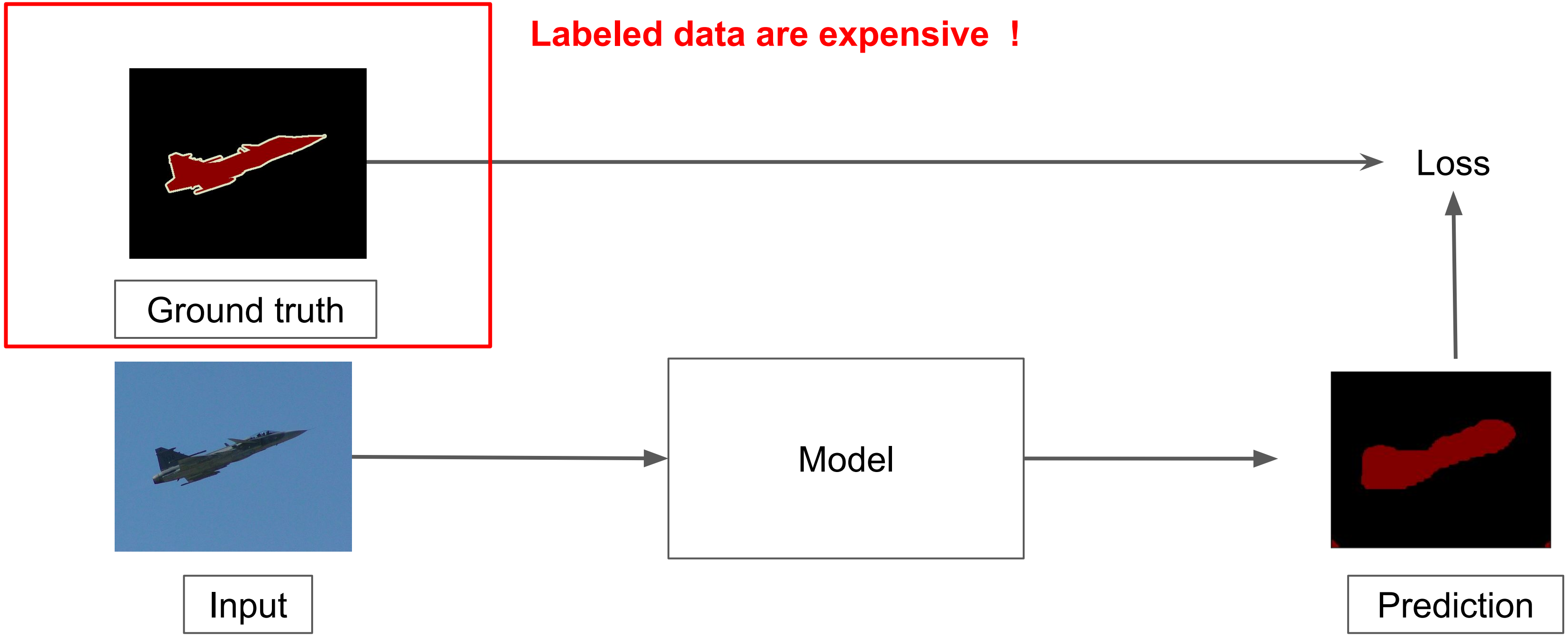
Supervised Semantic Segmentation

The model learns from the segmentation map ground truth.



Pictures taken from Pascal VOC 2012 Dataset [12].

Supervised Semantic Segmentation



Pictures taken from Pascal VOC 2012 Dataset [12].

Weakly Supervised Semantic Segmentation (WSSS)

Cheaper labels

Plane,
...

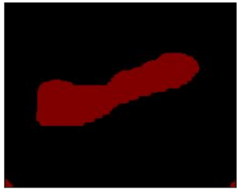
Ground truth

Loss

Input

Model

Plane: 80 %,
Person: 1 %,
...



Prediction

Pictures taken from Pascal VOC 2012 Dataset [12].

Weakly Supervised Semantic Segmentation (WSSS)

Cheaper labels

Plane,
...

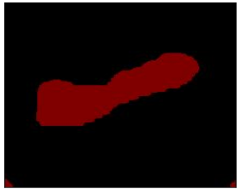
Ground truth

Loss

Input

Model

Plane: 80 %,
Person: 1 %,
...

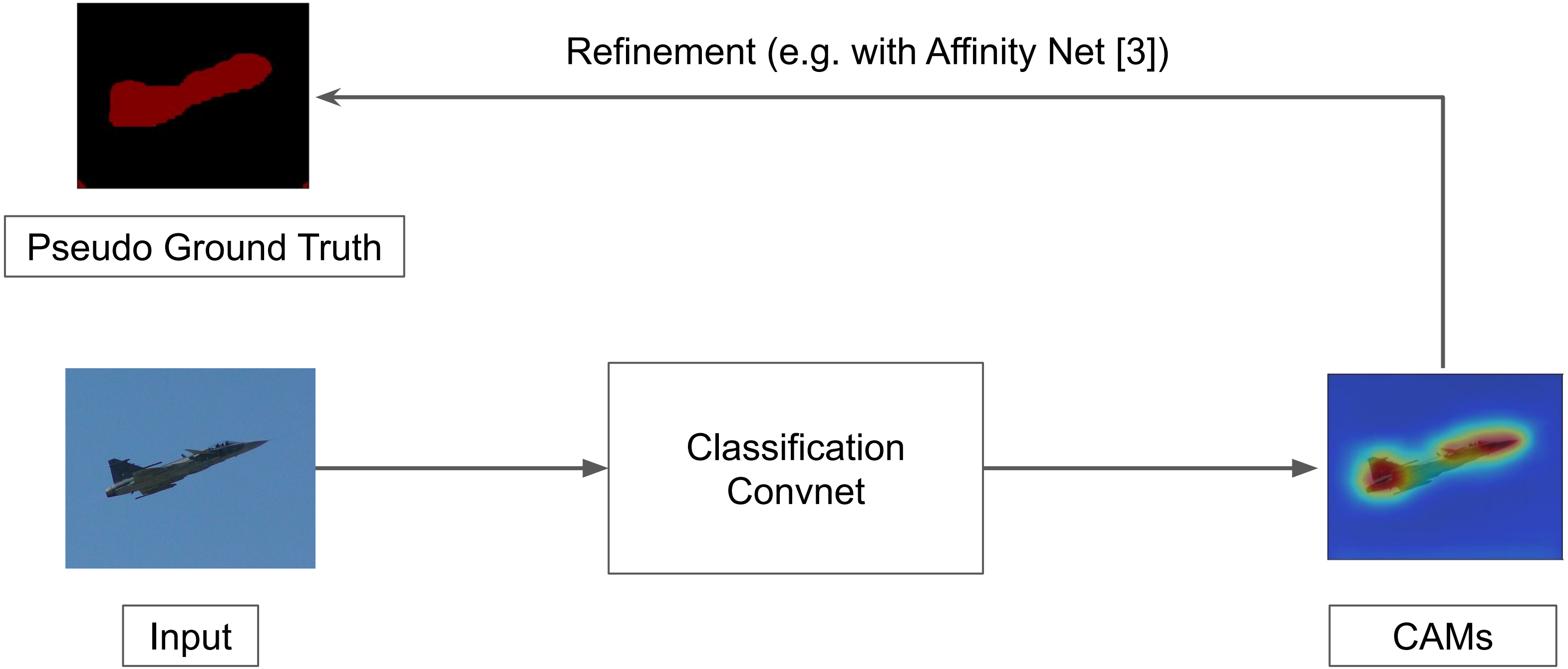


Prediction

End goal

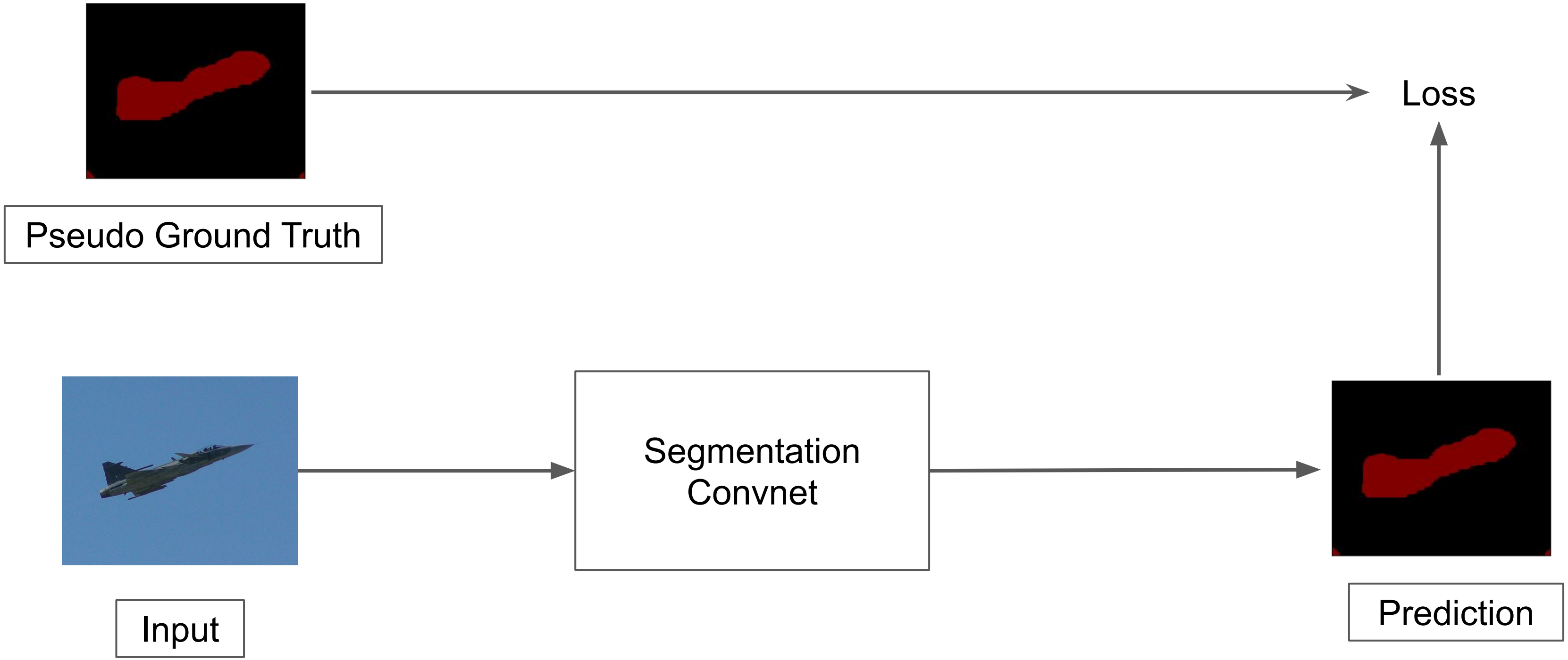
Pictures taken from Pascal VOC 2012 Dataset [12].

Main Idea: Use Class Activation Maps (CAMs) (1/2)



Pictures taken from Pascal VOC 2012 Dataset [12].

Main Idea: Use Class Activation Maps (CAMs) (2/2)



Pictures taken from Pascal VOC 2012 Dataset [12].

Problems

The initial CAMs focus only on the most discriminative part of the image.

On the literature, WSSS fall into two categories:

- Methods that improves the initial CAMs.
- Methods that refines the CAMs as a second step.

Pictures taken from Pascal VOC 2012 Dataset [12].

Goal of the project

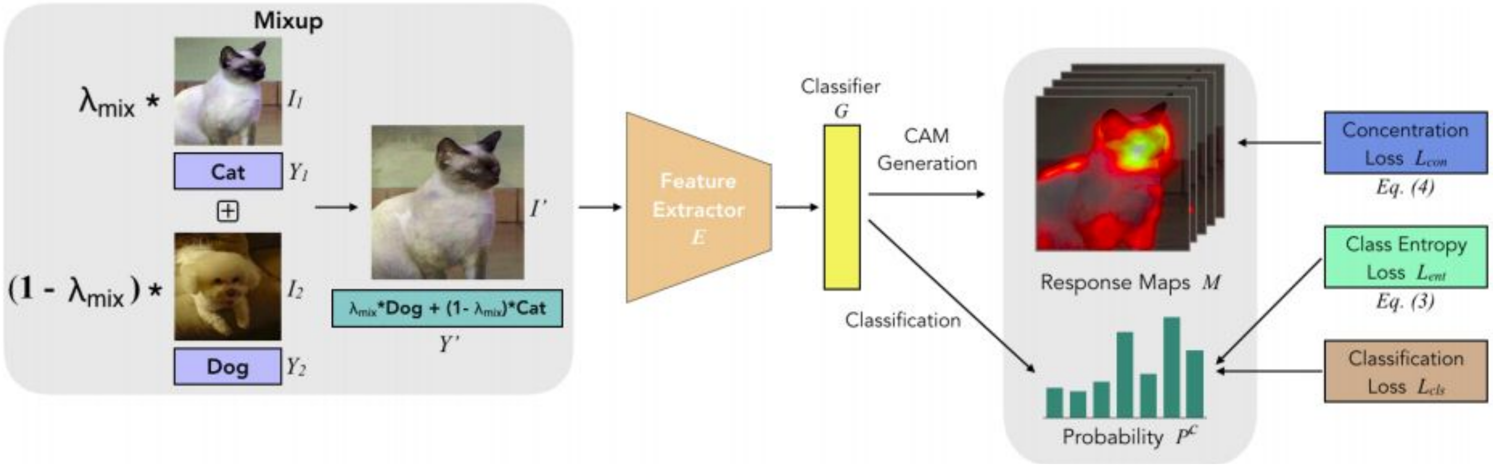
Study two papers, which aims at **improving the initial CAMs**:

- [1] Yu-Ting Chang, Qiaosong Wang, Wei-Chih Hung, Robinson Piramuthu, Yi-Hsuan Tsai, Ming-Hsuan Yang. **Mixup-CAM: Weakly-supervised Semantic Segmentation via Uncertainty Regularization**. 2020.
- [2] Yu-Ting Chang and Qiaosong Wang and Wei-Chih Hung and Robinson Piramuthu and Yi-Hsuan Tsai and Ming-Hsuan Yang. **Weakly-Supervised Semantic Segmentation via Sub-category Exploration.** CVPR, 2020.

Evaluate variants of these algorithms, which we named **FMix-CAM**, **Manifold Mixup-CAM** and **Sub-category Teaching**.

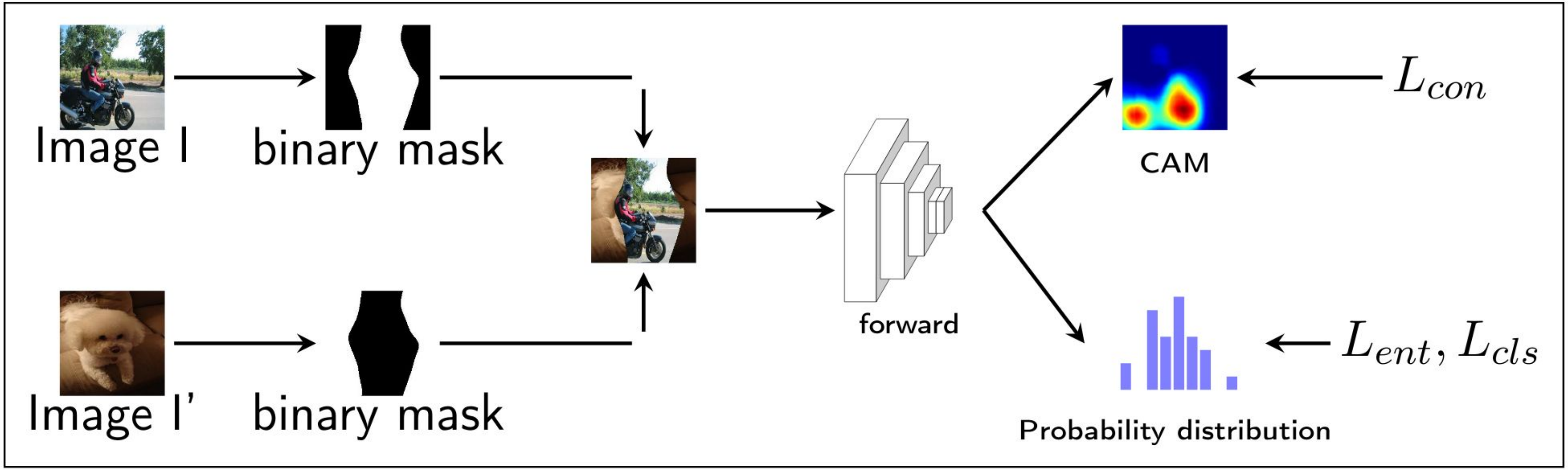
Paper [1]: Mixup-CAM

Figure taken from [1].



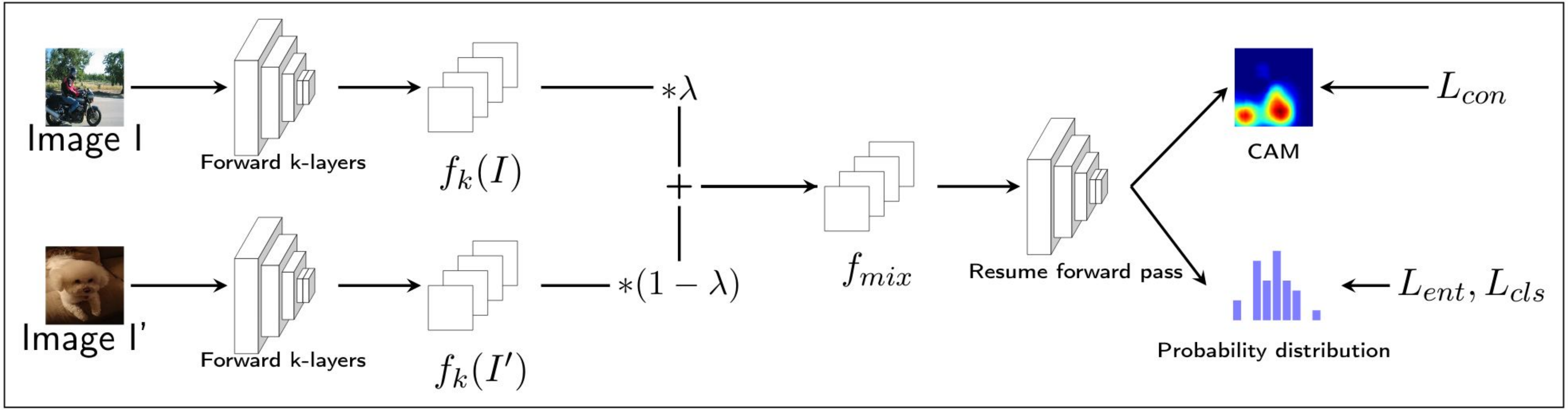
Our experiment: Fmix CAM

Mix training images according to a binary mask that has been sampled.
Based on Fmix [6].



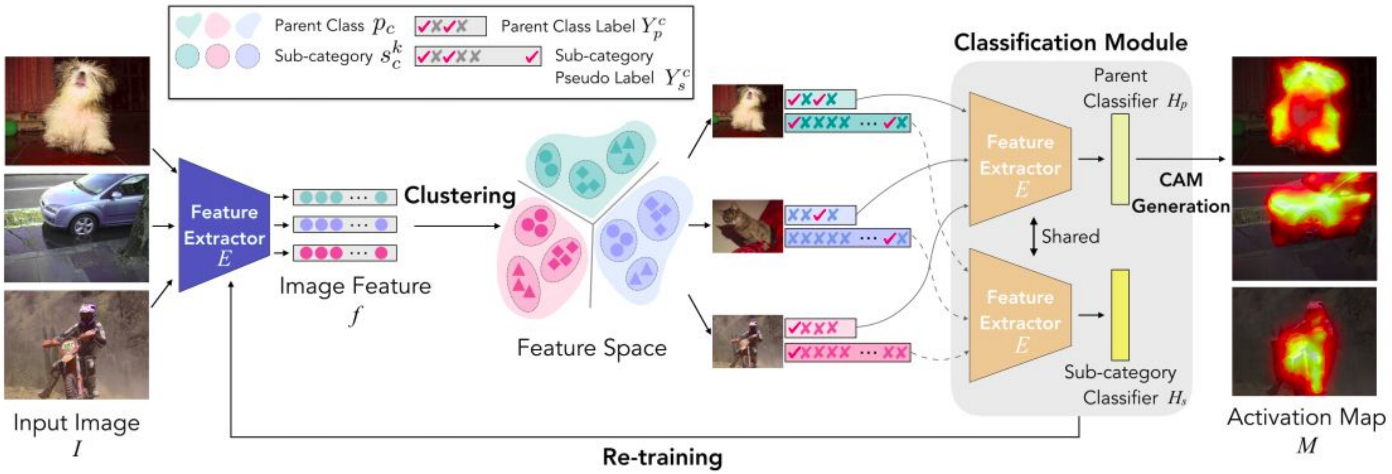
Our experiment: Manifold Mixup CAM

Mix feature maps during forward pass at random layer k. Based on Manifold Mixup [9].



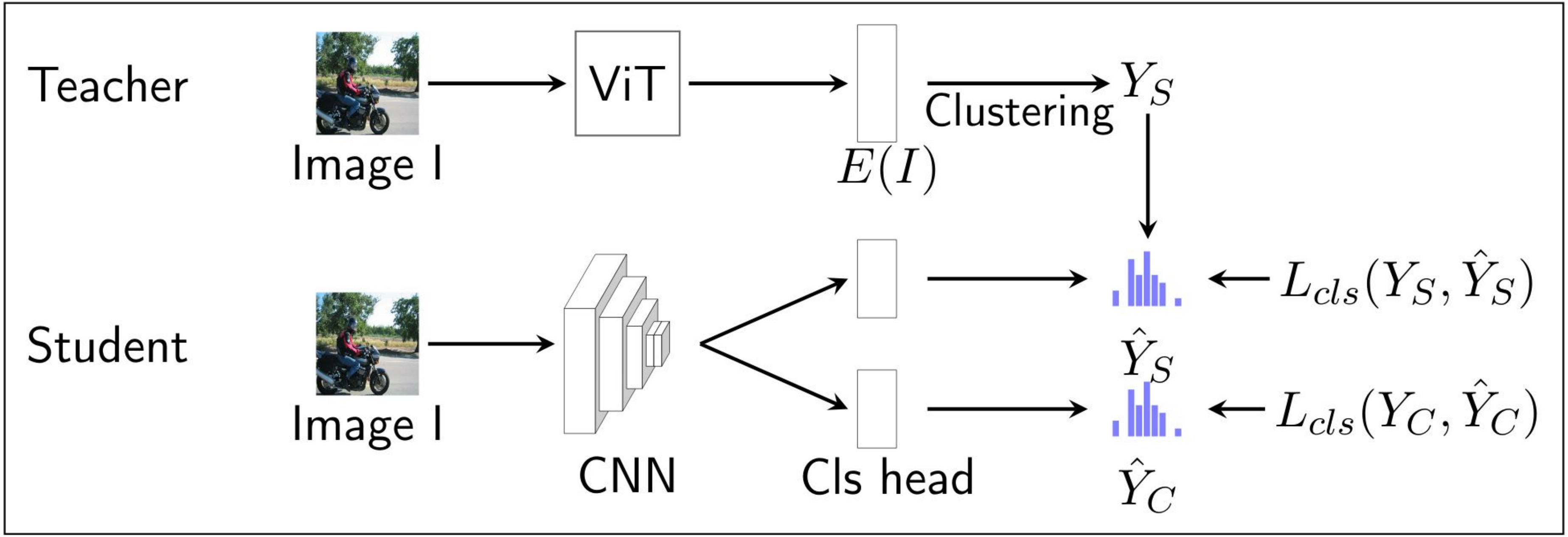
Paper [2]: Sub-category Exploration

Figure taken from [2].



Our experiment: Sub-category Teaching

Use an auxiliary network to distill knowledge to the student network. Based on [11] [13].



Quantitative results on subset of dataset

Encouraging results during the initial step. Bad results on final step (overfitting the segmentation network).

Methods	mIoU (%)	mIoU (%) (refined)
CAM (baseline)	45.0	55.6
Sub-Category Exploration [6]	46.5	56.7
Sub-Category Teaching (ours)	46.4	55.2
Mixup CAM [5]	46.8	55.8
Manifold Mixup CAM (ours)	46.8	56.6
Fmix CAM (ours)	**48.2**	**58.4**

Table 1. mIoU of class activation maps on *train* set during the initial step using only 1,464 training images

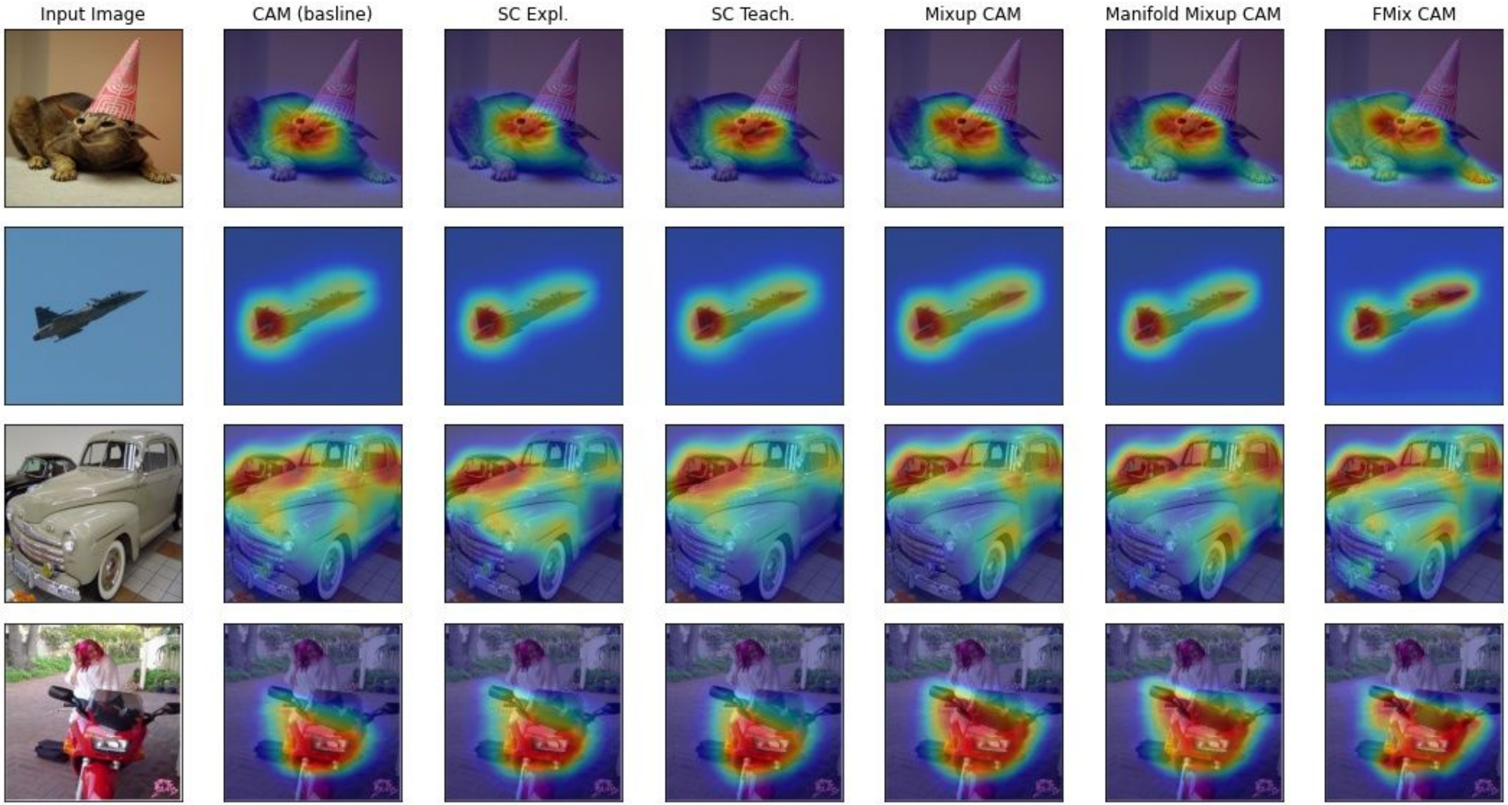
Initial step

Methods	mIoU (%)
CAM (baseline)	52.7
Sub-Category Exploration [6]	54.3
Sub-Category Teaching (ours)	53.0
Mixup CAM [5]	**55.1**
Manifold Mixup CAM (ours)	53.5
Fmix CAM (ours)	54.0

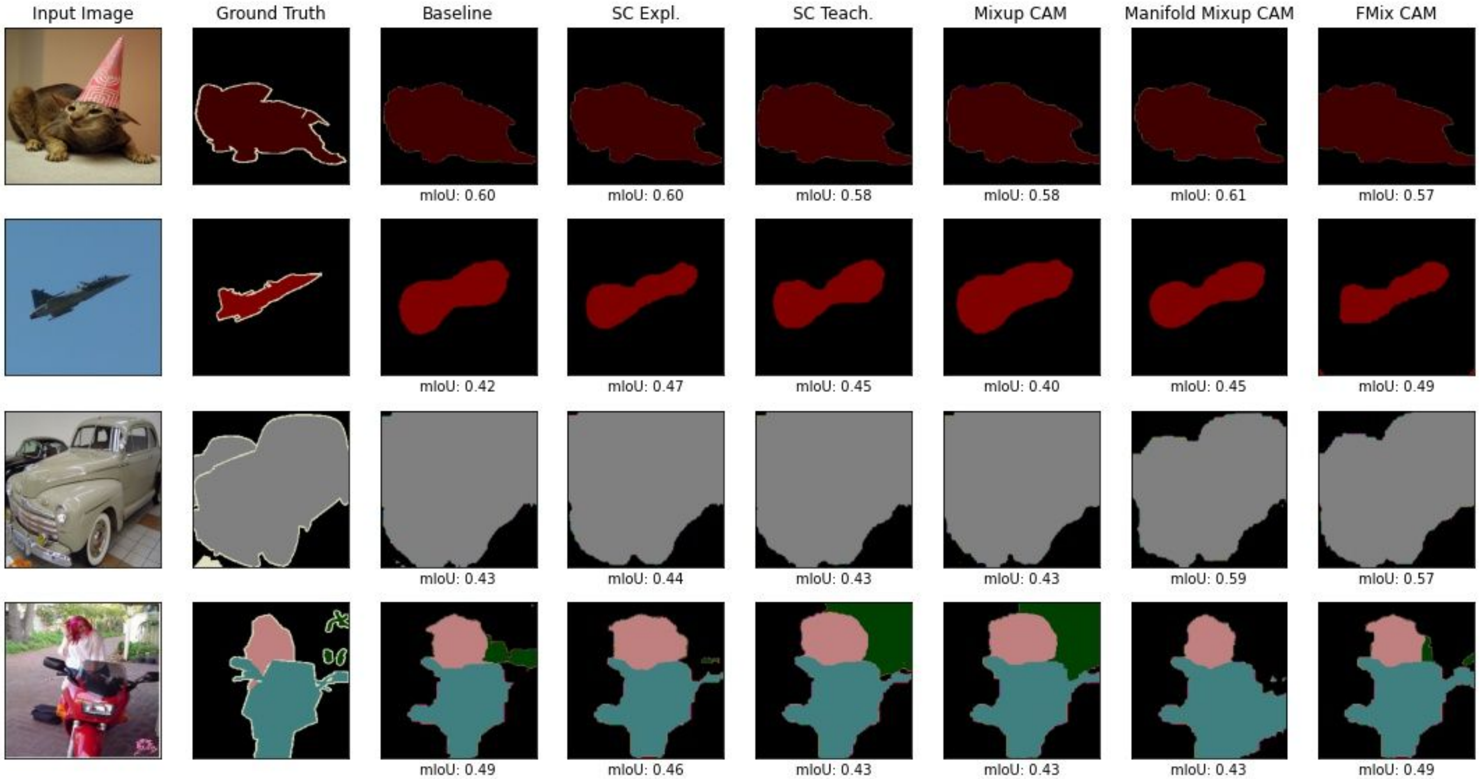
Table 2. Final mIoU of class activation maps on *val* set after training DeepLabv2 segmentation network using only 1,464 training images

Final step

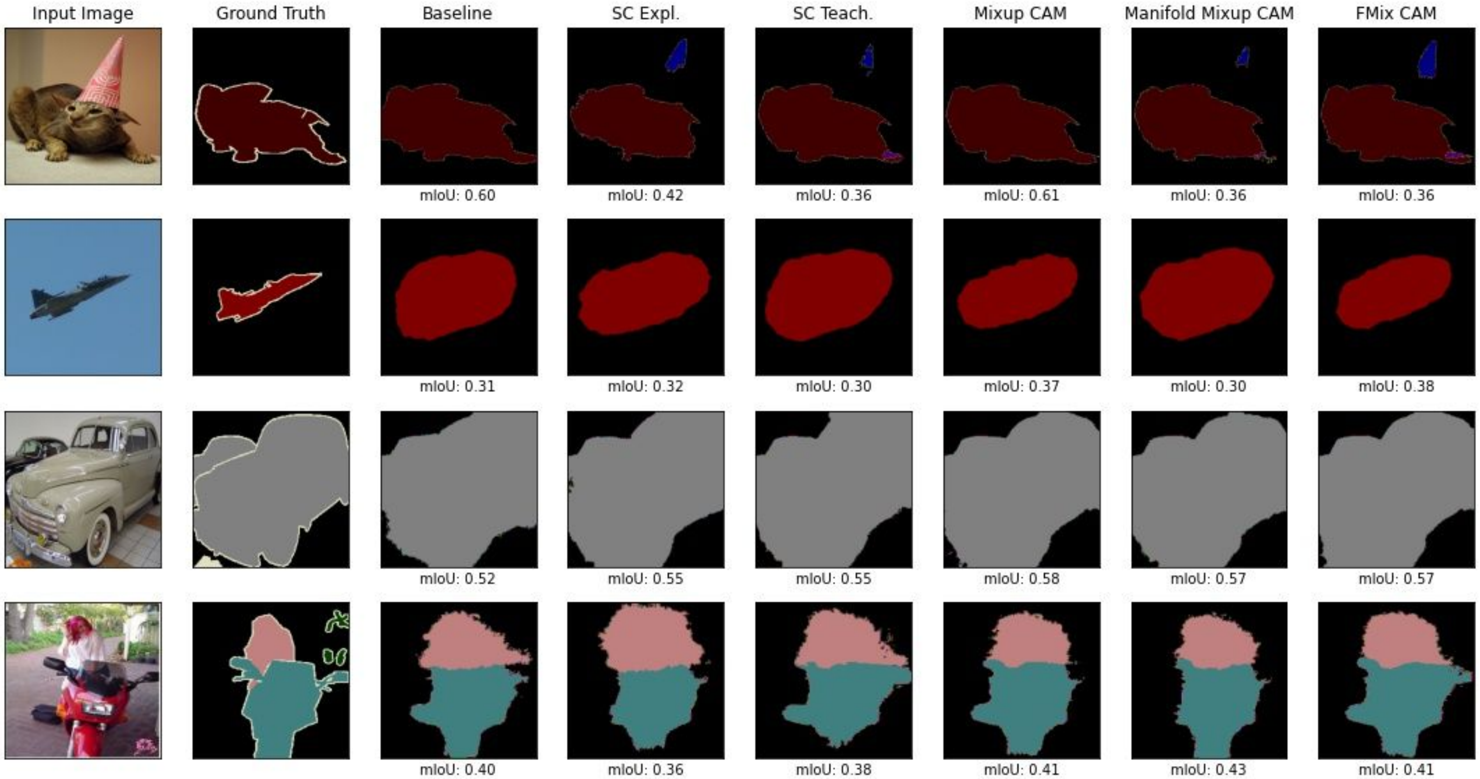
Qualitative results: initial CAMs



Qualitative results: Refined CAMs



Qualitative results: Segmentation network



Conclusion

Encouraging result with FMix CAM on the initial step.

On the final step, we got bad results compared to the published papers' results because we had overfitting issue to the noisy labels.

Please see my github repo: https://github.com/liuvince/mva-wsss.

References

1. Yu-Ting Chang and Qiaosong Wang and Wei-Chih Hung and Robinson Piramuthu and Yi-Hsuan Tsai and Ming-Hsuan Yang. Weakly-Supervised Semantic Segmentation via Sub-category Exploration. CVPR, 2020.
2. Yu-Ting Chang, Qiaosong Wang, Wei-Chih Hung, Robinson Piramuthu, Yi-Hsuan Tsai, Ming-Hsuan Yang. Mixup-CAM: Weakly-supervised Semantic Segmentation via Uncertainty Regularization. 2020
3. Jiwoon Ahn and Suha Kwak. Learning pixel-level semantic affinity with image-level supervision for weakly supervised semantic segmentation. CVPR, 2018.
4. Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. TPAMI, 2017.
5. Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. CVPR, 2016.
6. Ethan Harris, Antonia Marcu, Matthew Painter, Mahesan Niranjan, Adam Prügel-Bennett, Jonathon Hare. FMix: Enhancing Mixed Sample Data Augmentation. 2020.
7. M. Saquib Sarfraz, Vivek Sharma and Rainer Stiefelhagen. Efficient Parameter-free Clustering Using First Neighbor Relations. CVPR, 2019.
8. Liang-Chieh Chen, George Papandreou, Florian Schroff, Hartwig Adam. Rethinking Atrous Convolution for Semantic Image Segmentation. 2017.
9. Vikas Verma, Alex Lamb, Christopher Beckham, Amir Najafi, Ioannis Mitliagkas, Aaron Courville, David Lopez-Paz, Yoshua Bengio. Manifold Mixup: Better Representations by Interpolating Hidden States. ICML 2019.
10. Bolei Zhou, Aditya Khosla, Agata Lapedriza, Aude Oliva, and Antonio Torralba. Learning deep features for discriminative localization. In CVPR, 2016.
11. Geoffrey Hinton, Oriol Vinyals, Jeff Dean. Distilling the Knowledge in a Neural Network. NIPS 2014
12. M. Everingham, S. M. A. Eslami, L. Van Gool, C. K. I.Williams, J. Winn, and A. Zisserman. The pascal visual object classes challenge: A retrospective. International Journal of Computer Vision, 111(1):98–136, Jan. 2015.
13. Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov,Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner,Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition atscale, 2020.