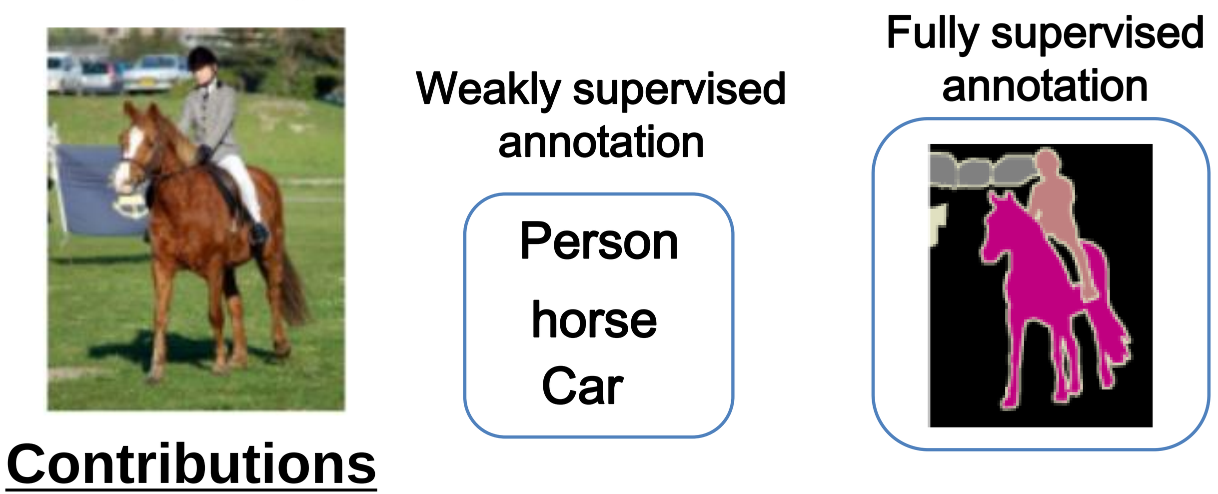
# Predicting Segmentation “Easiness” from the Consistency for Weakly-Supervised Segmentation

Wataru Shimoda and Keiji Yanai
The University of Electro-Communications, Tokyo, Japan

## Objective

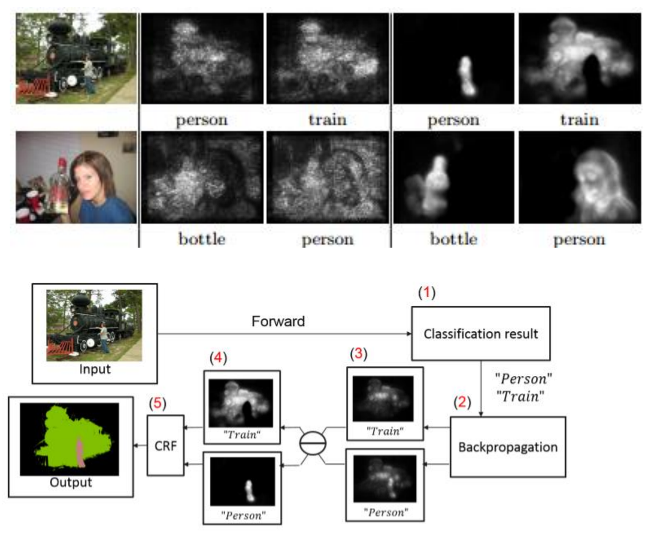
**Weakly supervised segmentation**
- Use only image-level annotation




**Contributions**
- Improved the method by Simonyan et al. [1] greatly
- Achieved state-of-the-art in weakly-supervised segmentation with PASCAL VOC 2012

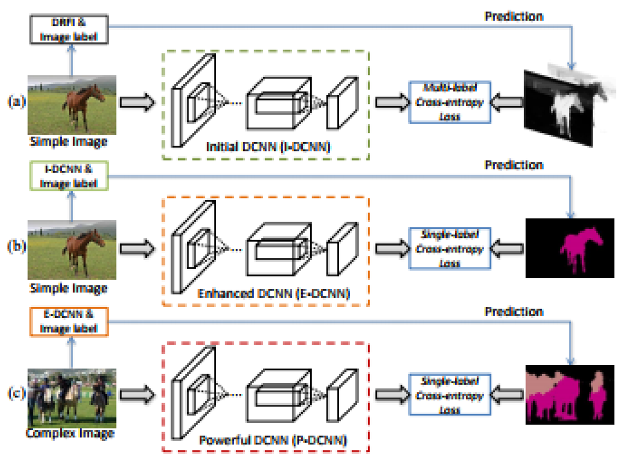
## Background

Distinct class specific saliency map (DCSM) [ECCV 2016]
- Visualization using subtraction for class specific map
- Improved method of Simonyan et al. [ICLRWS 2014]
- The method adapts visualization for multi class objects.
- Achieved high score on weakly supervised segmentation.




[Shimoda et al. ECCV 2016]

Simple to complex(STC) [PAMI 2016]
- Calculate seed area in advance
- Train full supervised model with the seed
- Repeat this process. (Self-pased learning)
- Kind of EM-algorithm -> initialization is crucial.




[Wei et al. PAMI 2016]

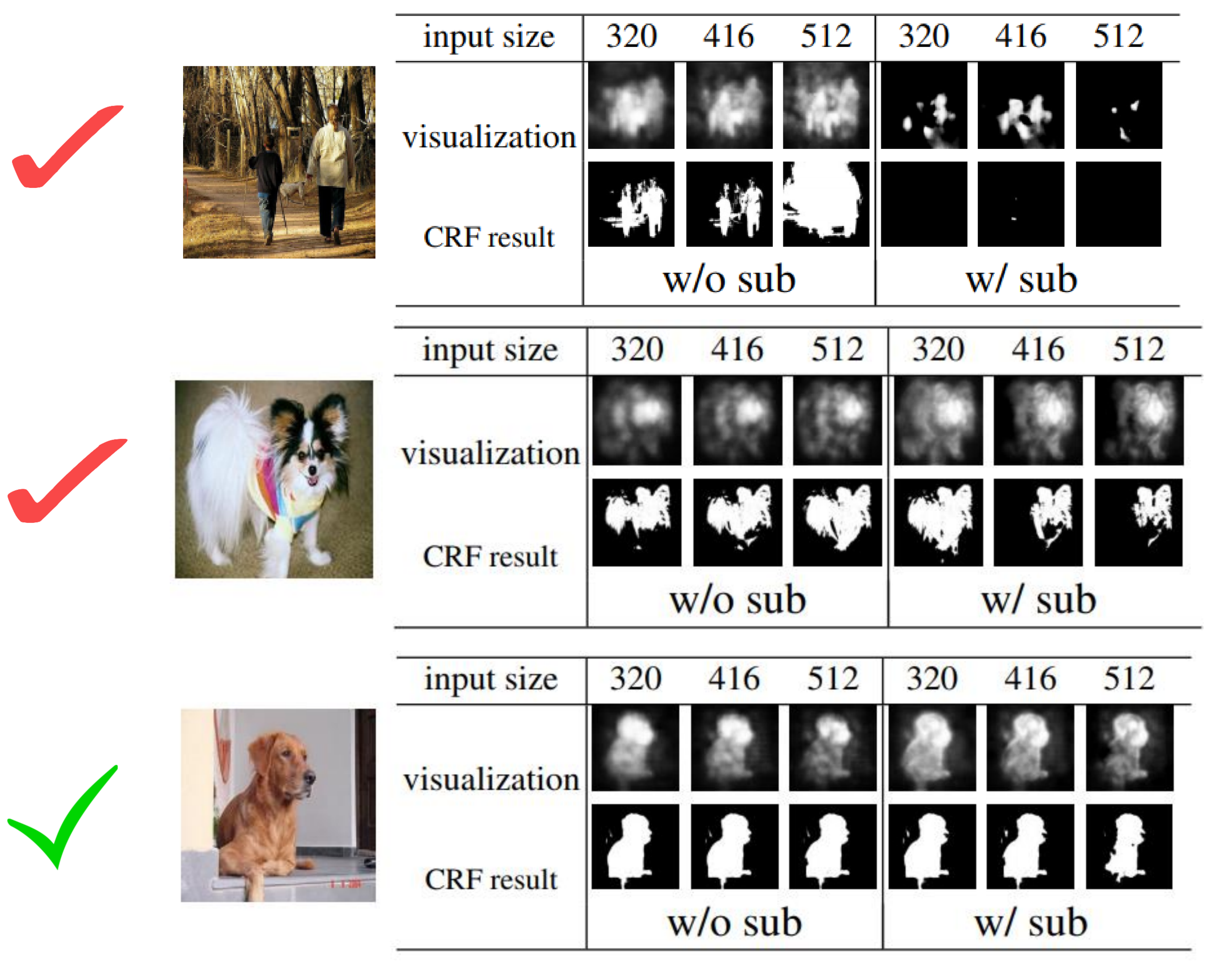
## Estimating “Easiness”

Reason of fault of segmentation using visualization
- Accuracy of classification
- Various object size
- Gap between the classification task and the segmentation task

Estimation of “Easiness” using consistency $R_{sub}(x) = \frac{1}{|\mathcal{C}|}\sum_{c \in \mathcal{C}} IoU(V_o^c(x), V_w^c(x))$

(1)Consistency of visualization
- The subtraction on DCSM generates clear class specific maps.
- We used defference between visualization with subtraction and witout subtraction
- This takes into account the classification result and image complexity.

(2)Consistency of receptive field $R_{size}(x) = \frac{1}{2|\mathcal{C}|}\sum_{c \in \mathcal{C}} IoU(V_o^b(x), V_o^c(x)) + IoU(V_w^b(x), V_w^c(x))$
- On the inference of pixel surrounding pixel information is very important.
- The receptive field size is related to the segmentation accuracy strongly.
- We used the change of receptive field on difference of input image size



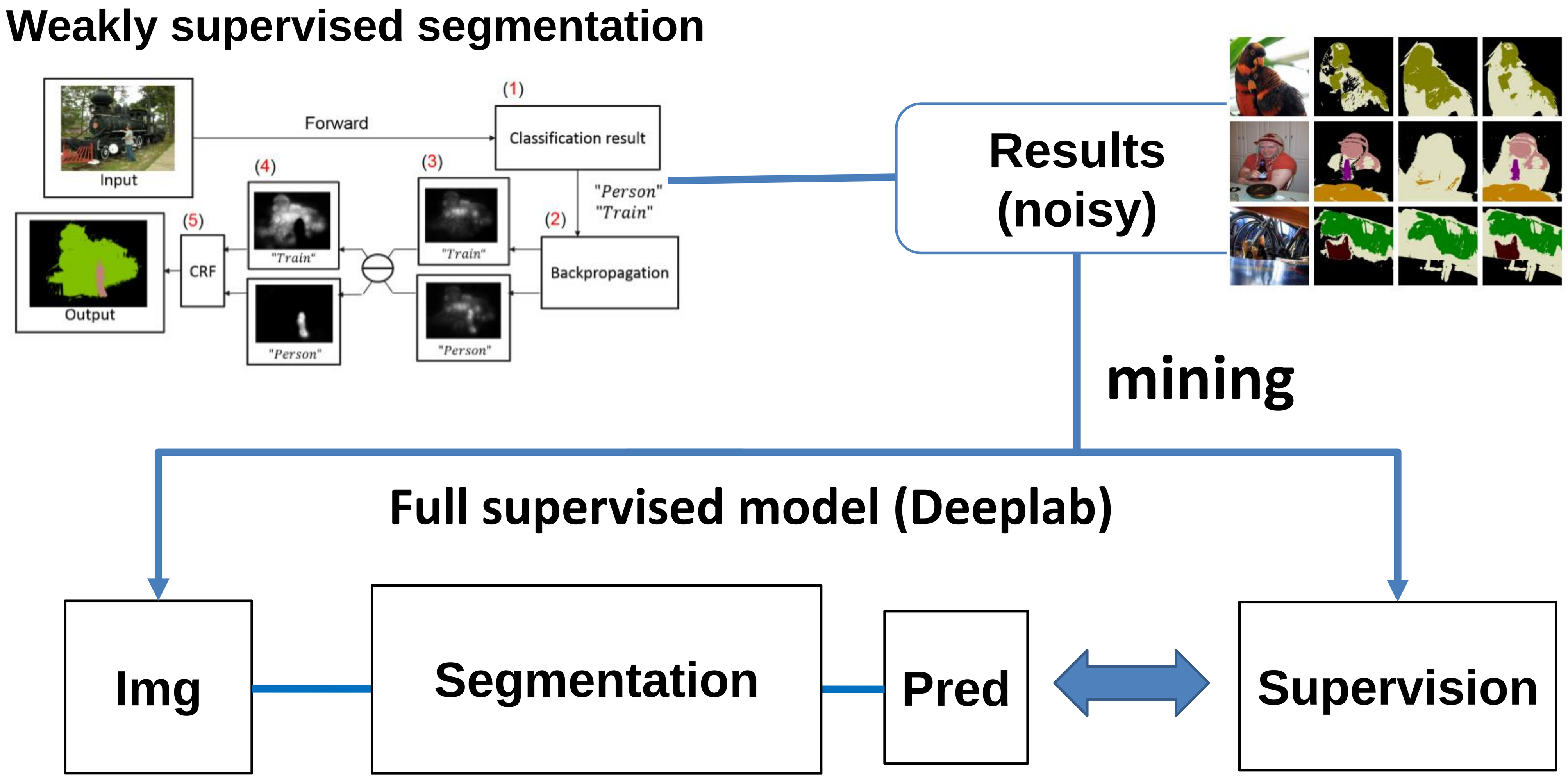

## Proposed Method

First step
- Obtain segmentation mask by existing weakly supervised segmentation method.

Second step
- Mine better results by our proposed method

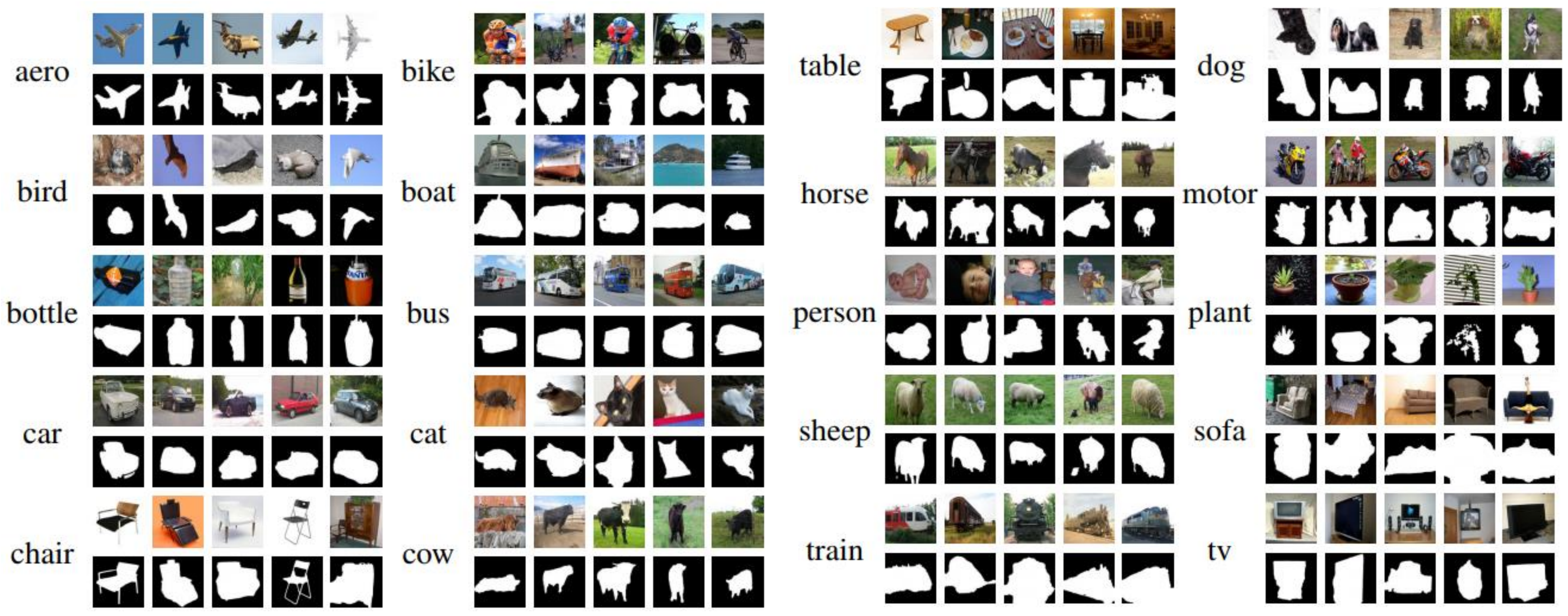
Third step
- Train full supervised segmentation model with mined supervision

**Weakly supervised segmentation**




## Experiments

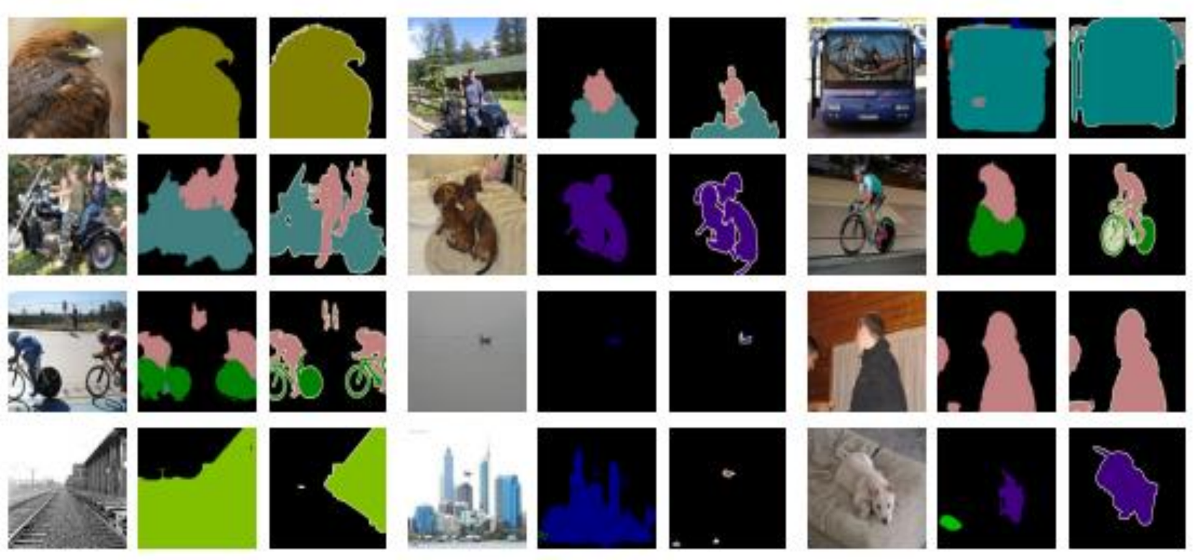
Top5 retrieval results obtained by our proposed method on Pascal VOC 2012 dataset
- Most of mined segmentation seeds are close to the ground truth.
- The results of sofa, chair and table include some noise. In general segmentation results of these class show low performance, hence the retrieved results are affected by the low quality of prediction directly




Pascal VOC 2012 validation set
Comnbination of “Easiness” with data augmentation

Table 1. Comnbination of “Easiness” with data augmentation

| setting | Base image N | Aug image N | mIoU |
|---|---|---|---|
| (a) | 8760 (th $\geq$ 0.3) | 730 (th $\geq$ 0.8) | 50.1 |
| (b) | 10582 (all) | 730 (th$\geq$0.8) | 48.9 |
| (c) | 8760 (th $\geq$ 0.3) | 2105 (th $\geq$ 0.7) | **51.3** |
| (d) | 10582 (all) | 2105 (th$\geq$0.7) | 49.9 |
| (e) | 8760 (th $\geq$ 0.3) | 8760 (th $\geq$ 0.3) | 49.7 |
| (f) | 10582 (all) | 10582 (all) | 48.8 |



Result on Pascal VOC 2012 test set

| Full supervised | Year | Acc |
|---|---|---|
| O2P | 2012 | 47.6 |
| SDS | 2014 | 52.6 |
| Deeplab | 2015 | 71.6 |

| method | Year | Acc |
|---|---|---|
| Using additional annotations | | |
| Point annotation | 2016 | 46.0 |
| Video | 2017 | 58.7 |
| Using only the image level label | | |
| Global pooling | 2015 | 24.9 |
| Simonyan (Backward) | 2014 | 33.8 |
| **Ours (Backward)** | **2016** | **45.1** |
| STC (re-training) | 2016 | 51.2 |
| SEC (re-training + unsupervised) | 2016 | 51.7 |
| **Ours (Backward + re-training)** | **2017** | **52.8** |
| Adversarial Erasing | 2017 | 55.7 |

*We used the same network architecture with deeplab model for training full supervised model.

**Conclusion**

Trade off exists between the number of training data and quality of training data.
The segmentation accuracy can be boost by data selection with data augmentation.