FickleNet:
Weakly and Semi-supervised Semantic Image Segmentation using Stochastic Inference

김태미

2021.06.04.

1. Introduction

Weakly-Supervised Segmentaion

- Segmentation using weakly labeled data

ex) only with image-level annotations without pixel-level annotations

2. Related Works

SRG (Seeded Region Growing)

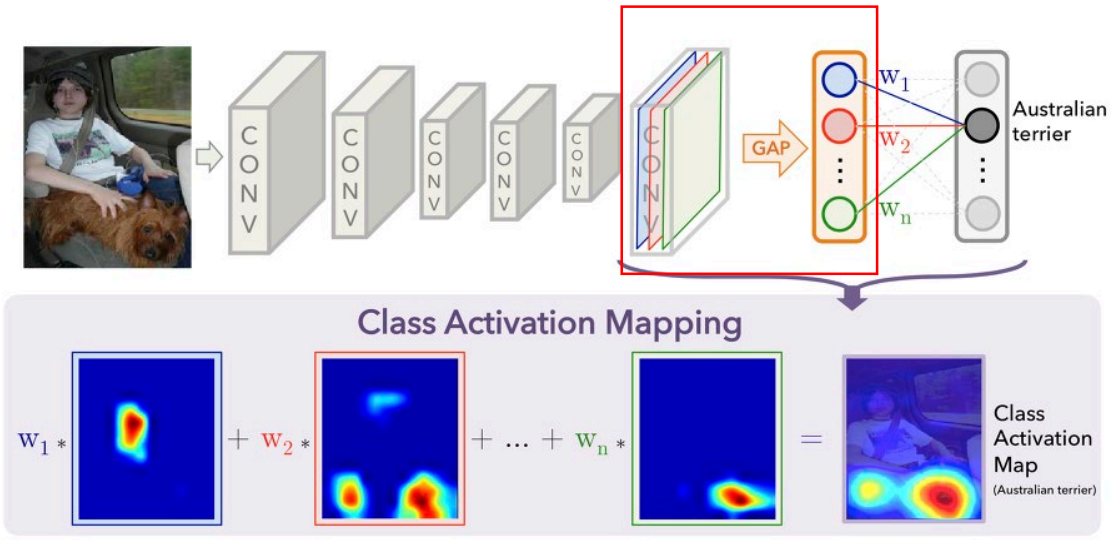
- Expanding the localization map from the seed

  ex) random seed, all pixels

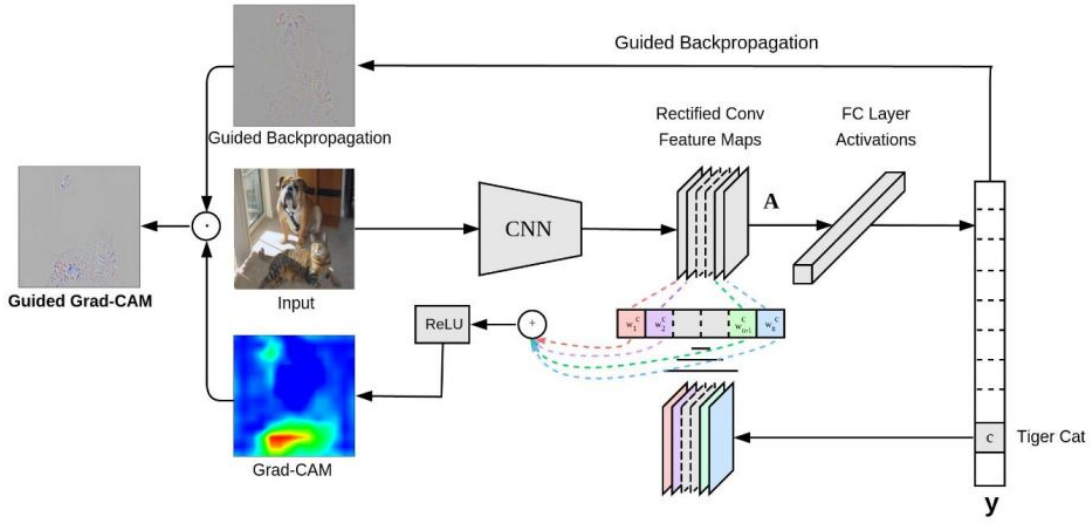
2. Related Works

CAM (Class Activation Map)

- The contribution of each hidden unit in a neural net to the classification score
- Discriminative region



Grad-CAM



2. Related Works

DSRG (Deep Seeded Region Growing)

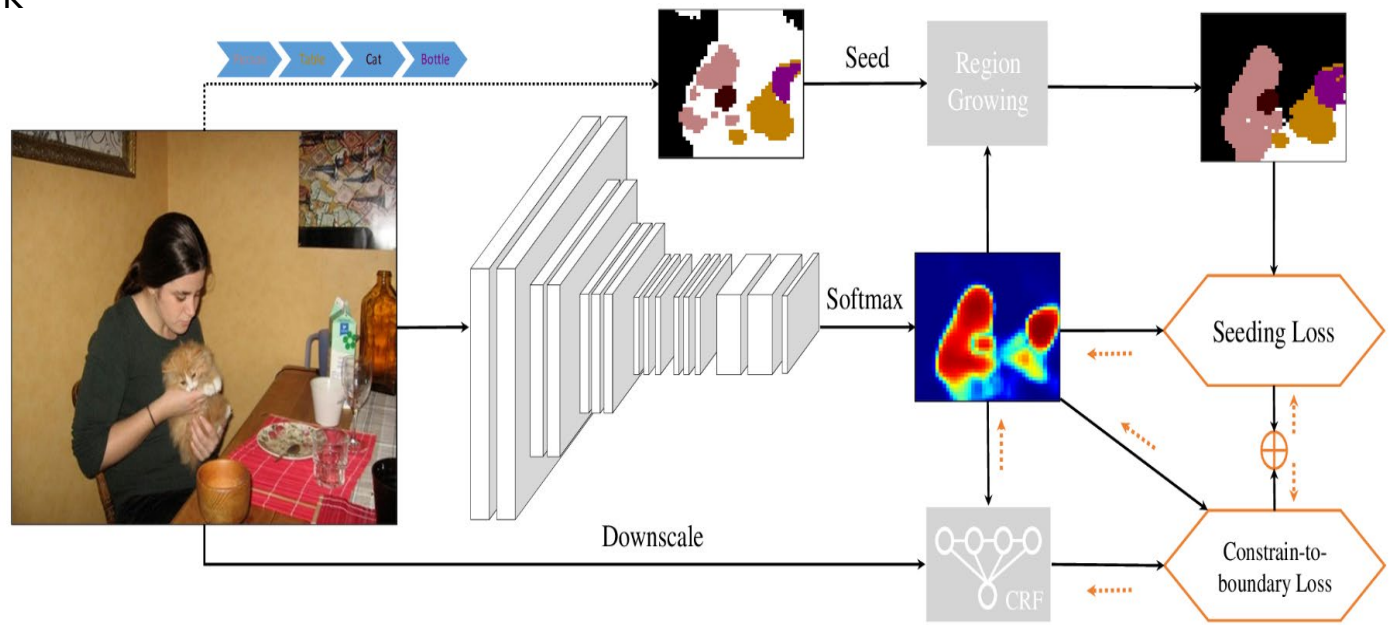
- Seed generation

  : VGG-16 + GAP layer -> extracting CAM

- Weakly-supervised segmentation network

  : $L = L_{seed} + L_{boundary}$

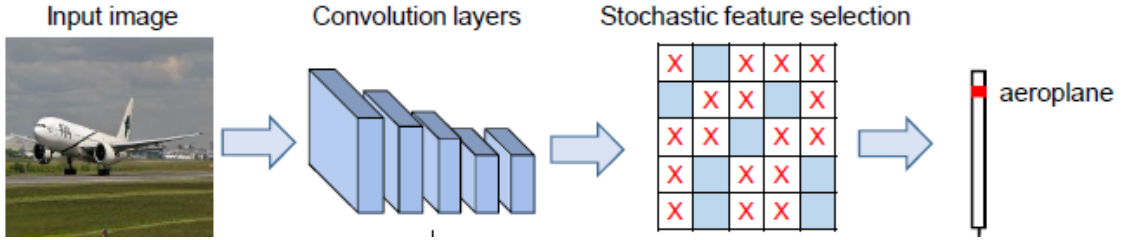
- Deep seeded region growing

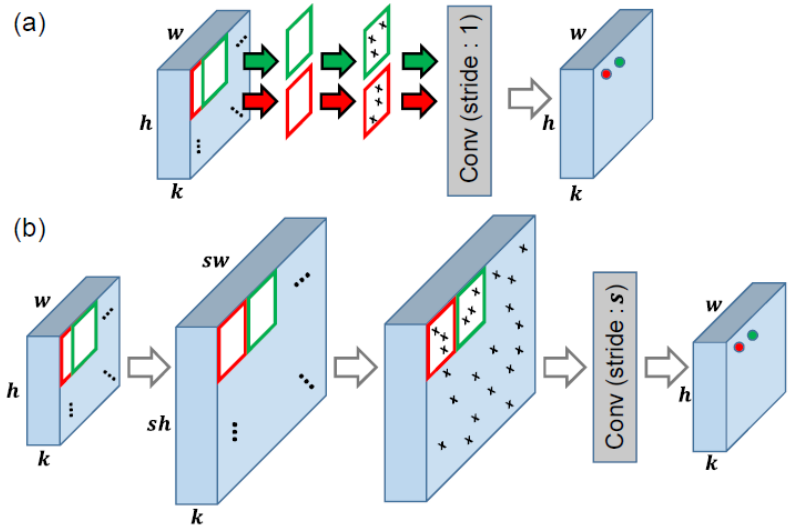


3. Method

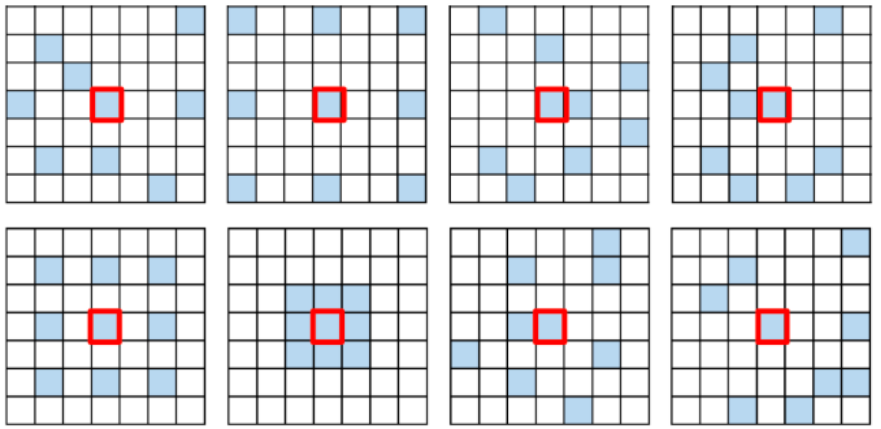
Stochastic Hidden Unit Selection

- Randomly select the feature maps
- Spatial dropout





Feature map expansion

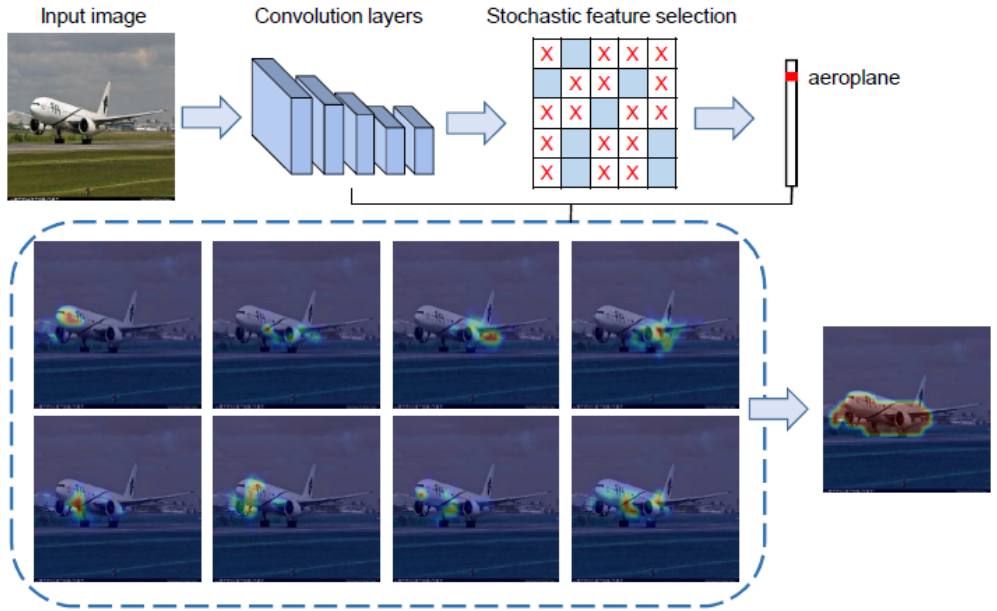


Center-preserving spatial dropout

3. Method

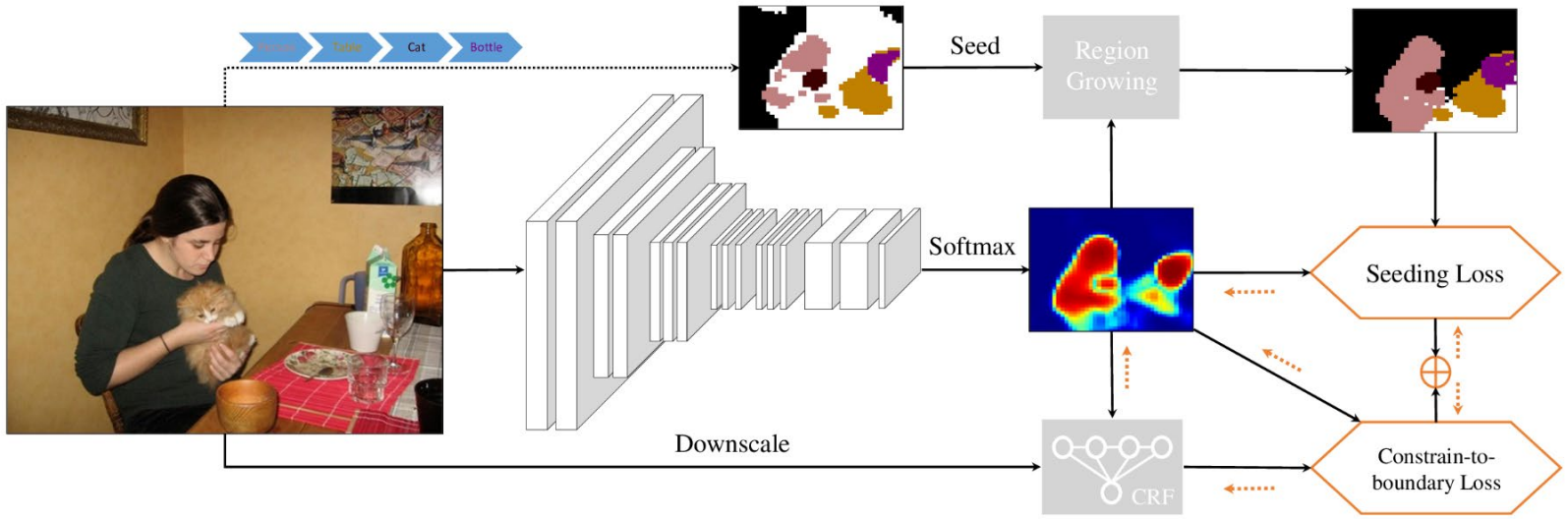
Inference Localization Map

- Localization map from Grad-CAM
- Localization map aggeregation



3. Method

Training the Segmentation Network



Conditional Random Field

$$L_{seed} = -\frac{1}{\sum_{c\in C}|S_c|}\sum_{c\in C}\sum_{u\in S_c} logH_{u,c} - \frac{1}{\sum_{c\in \bar{C}}|S_c|}\sum_{c\in \bar{C}}\sum_{u\in S_c} logH_{u,c}$$

Normalization

Foreground class

Background class

$$L_{boundary} = -\frac{1}{n}\sum_{u=1}^{n}\sum_{c\in C} Q_{u,c}(X, f(X)) log \frac{Q_{u,c}(X, f(X))}{f_{u,c}(X)}$$

Fully-connect conditional random field (Labeling을 위한 softmax regression)

4. Experiments

Table 1. Comparison of weakly supervised semantic segmentation methods on VOC 2012 validation and test image sets. The methods listed here use DeepLab-VGG16 for segmentation.

Methods	Training	*val*	*test*
Supervision: Image-level and additional annotations			
MIL-seg $_{\text{CVPR '15}}$ [23]	700K	42.0	40.6
STC $_{\text{TPAMI '17}}$ [32]	50K	49.8	51.2
TransferNet $_{\text{CVPR '16}}$ [9]	70K	52.1	51.2
CrawlSeg $_{\text{CVPR '17}}$ [10]	970K	58.1	58.7
AISI $_{\text{ECCV '18}}$ [11]	11K	61.3	62.1
Supervision: Image-level annotations only			
SEC $_{\text{ECCV '16}}$ [16]	10K	50.7	51.1
CBTS-cues $_{\text{CVPR '17}}$ [24]	10K	52.8	53.7
TPL $_{\text{ICCV '17}}$ [14]	10K	53.1	53.8
AE_PSL $_{\text{CVPR '17}}$ [31]	10K	55.0	55.7
DCSP $_{\text{BMVC '17}}$ [2]	10K	58.6	59.2
MEFF $_{\text{CVPR '18}}$ [8]	10K	-	55.6
GAIN $_{\text{CVPR '18}}$ [19]	10K	55.3	56.8
MCOF $_{\text{CVPR '18}}$ [30]	10K	56.2	57.6
AffinityNet $_{\text{CVPR '18}}$ [1]	10K	58.4	60.5
DSRG $_{\text{CVPR '18}}$ [12]	10K	59.0	60.4
MDC $_{\text{CVPR '18}}$ [33]	10K	60.4	60.8
FickleNet (Ours)	10K	**61.2**	**61.9**

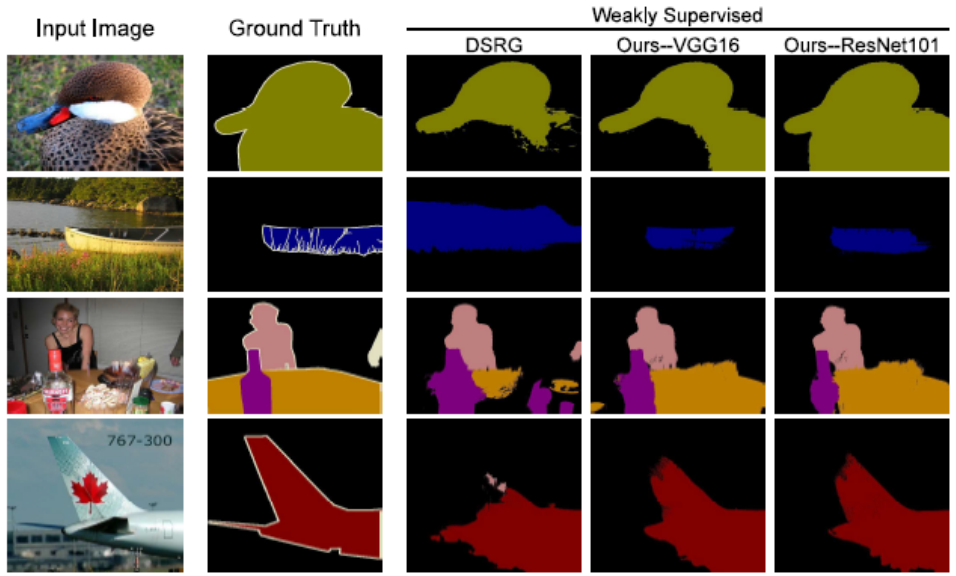
Table 2. Comparison of weakly supervised semantic segmentation methods on VOC 2012 validation and test image sets. The methods listed here use ResNet-based DeepLab for segmentation.

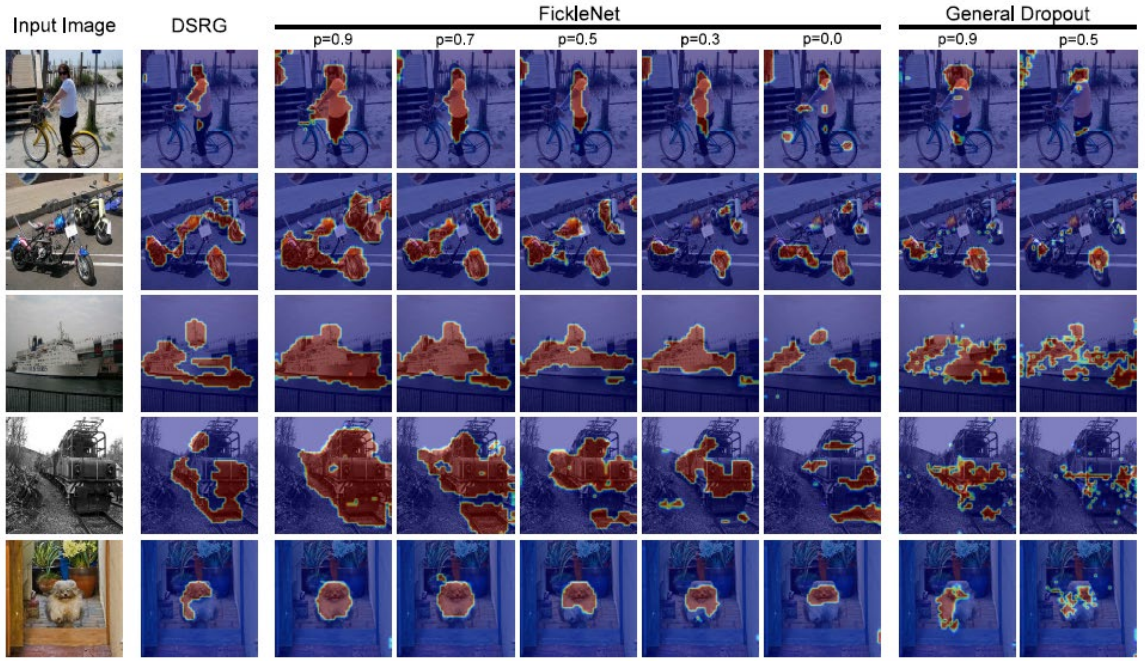
Methods	Backbone	*val*	*test*
MCOF [30]	ResNet 101	60.3	61.2
DCSP [2]	ResNet 101	60.8	61.9
DSRG [12]	ResNet 101	61.4	63.2
AffinityNet [1]	ResNet 38	61.7	63.7
FickleNet (ours)	ResNet 101	**64.9**	**65.3**

Table 5. Comparison of mIoU scores using different dropout rates (p) on PASCAL VOC 2012 validation images.

Methods	Dropout Rate (p)	mIoU
Deterministic	0.0	56.3
General Dropout	0.5	45.6
	0.9	49.1
FickleNet	0.3	58.8
	0.5	59.4
	0.7	60.0
	0.9	**61.2**

4. Experiments





Thank you