One-Shot Learning for Semantic Segmentation

Jia Zheng

SIST, ShanghaiTech

June 12, 2018

Outline

Outline

Problem Setup

Given a support set $S = \{(I^i, Y^i(l))\}_{i=1}^{k}$ of k image-binary mask pairs and query image I_q, the goal is to predict a binary mask $\hat{M}_q$ for semantic class l. Note that the semantic classes in train set and test set are mutually exclusive.

Outline

Overview

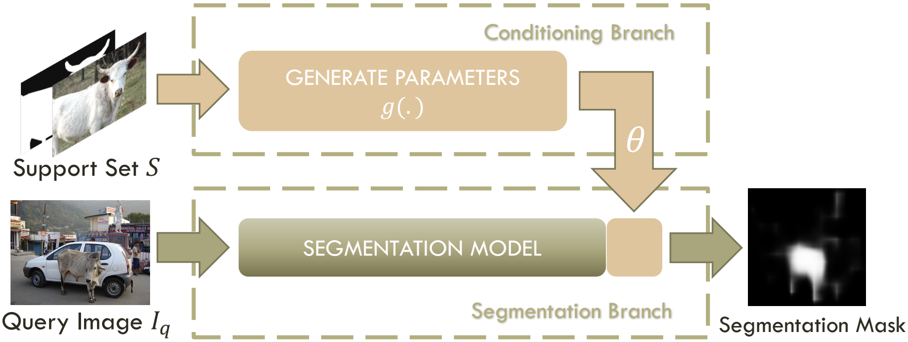


Figure: OSLSM [5]

Method

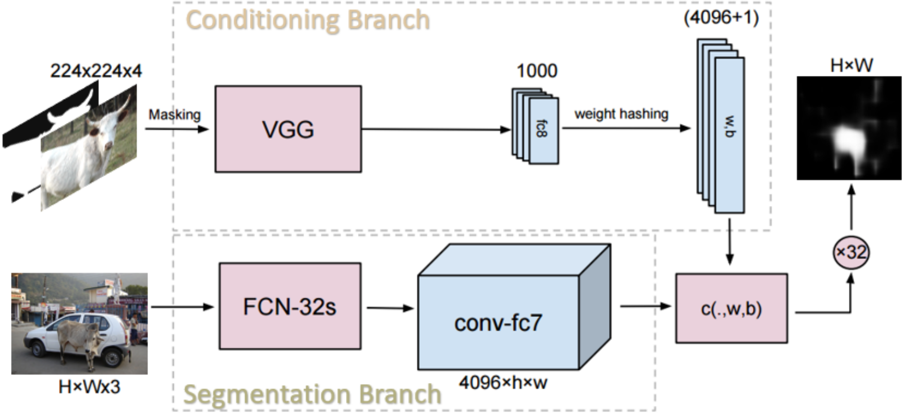
In the conditioning branch, we input the support set $S = \{I, Y(I)\}$ and produce a set of parameters,

$$w, b = g(S)$$

In the other branch, we extract a dense feature $F_q = \phi(I_q)$ from I_q, and use feature F_q and these parameters w, b to get the final mask

$$\hat{M}_q^{mn} = \sigma(w^\top F_q^{mn} + b)$$

Architecture



Conditioning Branch

Masking

Mask image with its corresponding label so it contains only the target object instead of the first layer to recieve the four channel image-mask pair as input.

Weight Hashing

To avoid over-fitting, employ a weight hashing layer from [2] to map the 1000-dimension vector output from the layer of VGG to the 4097 dimension of $\{w, b\}$.

Data sampling in Training

1. Sample an image-label (I_q, Y_q) uniformly from D_{train}
2. Sample a class $l \in L_{train}$ uniformly and use it to produce the binary mask $Y_q(l)$
3. Support set S is formed by picking one image-label pair at random from $D_{train} - \{(I_q, Y_q)\}$ with class l present

Extension to k-Shot

We use k labeled image to produce k sets of parameters. Each simple classifier has high precision but low recall. We ensemble these masks of classifiers by a logical OR operator.

Outline

Baseline

1. Base classifiers (1-NN and logistic regression)
2. Fine-tunning on support set (Suggested by [1])
3. Co-segmentation by Composition [3]
4. Siamese Network for One-Shot Dense Matching [4]

$i=0$	$i=1$	$i=2$	$i=3$
aeroplane, bicycle, bird, boat, bottle	bus, car, cat, chair, cow	diningtable, dog, horse, motorbike, person	potted plant, sheep, sofa, train, tv/monitor

Methods (1-shot)	PASCAL-5^0	PASCAL-5^1	PASCAL-5^2	PASCAL-5^3	Mean
1-NN	25.3	44.9	**41.7**	18.4	32.6
LogReg	26.9	42.9	37.1	18.4	31.4
Finetuning	24.9	38.8	36.5	30.1	32.6
Siamese	28.1	39.9	31.8	25.8	31.4
Ours	**33.6**	**55.3**	40.9	**33.5**	**40.8**

Methods (5-shot)	PASCAL-5^0	PASCAL-5^1	PASCAL-5^2	PASCAL-5^3	Mean
Co-segmentation	25.1	28.9	27.7	26.3	27.1
1-NN	34.5	53.0	**46.9**	25.6	40.0
LogReg	**35.9**	51.6	44.5	25.6	39.3
Ours	**35.9**	**58.1**	42.7	**39.1**	**43.9**

Table 1: Mean IoU results on PASCAL-5^i. **Top:** test classes for each fold of PASCAL-5^i. The **middle** and **bottom** tables contain the semantic segmentation meanIoU on all folds for the 1-shot and 5-shot tasks respectively.

Qualitative results

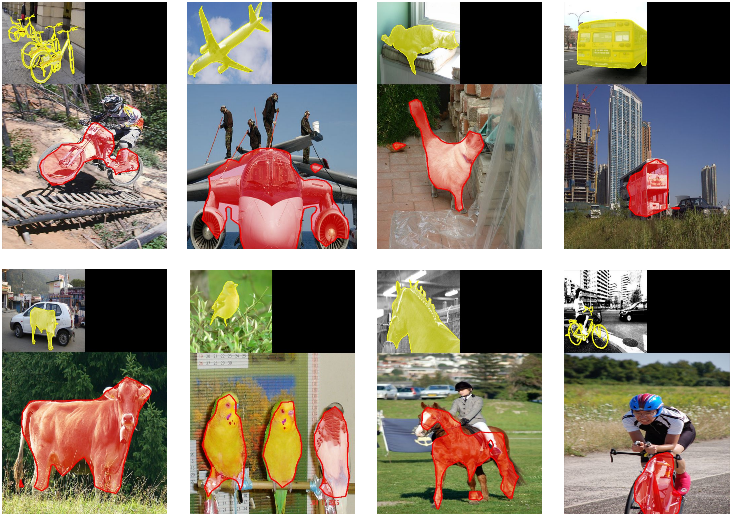


Illustration on conditioning effect

Effect of increasing the size of the support set

Reference

Sergi Caelles et al. “One-shot video object segmentation”. In: *CVPR*. 2017.

Wenlin Chen et al. “Compressing neural networks with the hashing trick”. In: *ICML*. 2015.

Alon Faktor and Michal Irani. “Co-segmentation by composition”. In: *ICCV*. 2013.

Gregory Koch. “Siamese Neural Networks for One-Shot Image Recognition”. PhD Thesis. University of Toronto, 2015.

Amirreza Shaban et al. “One-Shot Learning for Semantic Segmentation”. In: *BMVC*. 2017.

Thanks

Thanks for Attention!