Learning pixel objectness with Positive and Unlabelled examples

First Author
Institution1
Institution1 address
firstauthor@i1.org

Second Author
Institution2
First line of institution2 address
secondauthor@i2.org

July 12, 2017

Abstract

The ABSTRACT HERE

1 Introduction

The lack of "gold standard" annotations becomes the bottleneck for semantic segmentation.
The state-of-art deep learning algorithms for semantic segmentation [3] usually suffer from the lack of large-scale annotated dataset. Most of these methods assume the existence of precise, consistent and exaustive annotations. However, collecting such perfect segmentations manuallt in a large scale is expensive and time-consuming. Millions of images are available through Flikr, ImageNet and many other sources on the Internet but only a few of them were well annotated for semantic segmentation tasks. [1, 4, 2] A common method to overcome the lack of training samples is to use the pre-trained weights of the convolutional layers from the classification models. However, the neural network architecture design for semantic segmentation does not nessarily follows the architecture design of image classification. For instance, the segmentation models require less subsample pooling layers than the classification models do to keep more local information to locate the object. However, the difficulty for pre-training comparable representation of the images in the context of semantic segmentation arises from the lack of "gold standard" in large scale. We wanted to tackle the problem by alleviating the "gold standard" annotation assumption and scaling up the training dataset accordingly. This would require methods to learn good representation of images in the precense of annotation noise.
This paragraph should explain the difficulties for collecting segmentation annotation, including what could be the main errors in annotation.
The benchmark datasets usually provide perfect segmentations with all instance annotated (exaustive) and no misannotated instance (precise and consistent). However, it is natural for human being to make mistakes while annotating due to the lack of expertise, the intrisic ambiguity of tasks or unconscious bias. Huge efforts would be made to correct the mistakes being made, including double-checking the annotations over and over again and ensembling opinions from multiple annotators. Otherwise, these mistakes may lead to annotations that contain misannotated instances, misclassification of instances and unannotated instances.
This paragraph should reason the idea obtain pre-trained features by learning "objectness" with Positive and Unlabeled examples.
It is obvious that training with the noisy annotations would lead to higher segmenting errors than training with the correct annotations, if the validation samples had also correct annotations. It is nevertheless unclear how the annotation noises influence the learned image representation. With end-to-end models contained convolutional neural networks, the purpose of a learning objective is not only to train a classifier with minimum errors but also to learn satisfactory convolutaional filters.
That leads to our research questions:

1. How to compensate the classification bias introduced by the annoation noises with corresponding prior

knowledges.

2. How to learn satisfactory image representation in the presence of annotation noises.

Briefly formulate the problem.
The Semantic Segmentation problem can be considered as per-pixel classification. Each of the pixels is assigned a label of either 0, indicating a pixel for the background, or $k \in 1, \ldots, K$, indicating a pixel for an instance from one of the K categories. The aforementioned errors can be interpreted by the pixel label flipping: *misannotation* flipped from 0 to k, *inexaustive annotation* flipped from k to 0, and *misclassification* flipped from i to j, where $i, j, k \in 1, \ldots, K$.
Misannotating and misclassification
One of the main hypothesis made in this work is that the misannotation and misclassification of instances can still possibly provide information for learning visual representation of object, asumming the errors are not dorminant. Supposing an instance of a dog toy is annotated as a dog, given that "toy" is not one of the pre-defined categories whereas the "dog" is, the misannotation error would introduce bias to the classification layer but not necessarily to the convolutional layers, especially the bottom ones. The bottom-level features are believed to be shared among different categories and thus should be more robust to the misannotation error than the top-level features, assumming that the misannotated instance is still visually distinguishable and semantically meaningful.
Inexaustive annotating
The inexaustive annotations, on the other hand, can introduce bias to both the decoding layer and the encoding layers because they negatively contribute to the activations in all the layers. Therefore, the inexaustive annotation needs to be properly handled with the prior knowledge modeling the annotation missing pattern. Given that we treat any annotated instance as "reliable" annotation, an extra prior knowledge can be added that all the annotated instances, i.e.the foreground, are reliable and the unannotated pixels, i.e.the background, may contain missing instances. That satisfies a Positive and Unlabeled learning setup where the training dataset contain only the positive examples and unlabeled examples that could be either positive or negative.
Table of contents
In the next section, we review related works on deep learning with label noises. In Section 3 we judge the possibility of learning convolutional representation with misannotations by learning to predict the pixel objectness. Section 4 explored the methods to compensate the inexaustive annotations in a Positive and Unlabeled Learning setup. Features learned by predicting the pixel objectness with inexaustive annotations were then validated with experiments described in Section 5.

2 Related work

Positive and Unlabeled Learning

Deep Learning with Noisy Labels *Robustness analysis* Deep Learning is Robust to Massive Label Noise [7]

Entropy regularization Training deep neural networks on noisy labels with bootstrapping [6] Regularizing Neural Networks by Penalizing Confident Output Distributions [5]

Semi-supervised Learning

3 Pre-train features by learning "objectness"

4 Positive and Unlabeled Learning

One sentence summary of Positive and Unlabeled Learning

Formulation *This part should explain the Positive and Unlabeled Learning setup with mathematical representation when necessary.*

This part should discuss the linear model for observing positive conditioning on true positive and its relationship to changing the class weight.

This part should explain the influence of the imbalanced problem and how to overcome.

This part should explain why the exponential loss could perform better than the cross-entropy loss, potentially with a figure of 2D Gaussians.

Annotation	Loss	acc.	prec.	rec.	F_1
Complete	CrossEntropyU	0.87	0.88	0.82	0.85
50%P+50%N	CrossEntropyU	0.83	0.84	0.78	0.80
50%P+U	CrossEntropyU	0.61	0.92	0.30	0.38
50%P+U	WeightedU	0.66	0.93	0.39	0.51
50%P+U	ExponentialU	0.82	0.86	0.73	0.78
50%P+U	BootstrapHard	0.74	0.81	0.60	0.67
50%P+U	LinearNoiseModel				
50%P+U	DropoutRegularization				

Table 1: VGG8 CIFAR10

Annotation	Loss	pixel acc.	mean acc.	mean IU	f.w. IU
Complete	CrossEntropyU				
50%P+50%N	CrossEntropyU				
50%P+U	CrossEntropyU				
50%P+U	WeightedU				
50%P+U	ExponentialU				
50%P+U	BootstrapHard				
50%P+U	LinearNoiseModel				
50%P+U	DropoutRegularization				

Table 2: PASCAL VOC

This paragraph should explain why fade-in was introduced to avoid all-positive prediction

5 Results

6 Conclution

References

[1] Mark Everingham, SM Ali Eslami, Luc Van Gool, Christopher KI Williams, John Winn, and Andrew Zisserman. The pascal visual object classes challenge: A retrospective. *International journal of computer vision*, 111(1):98–136, 2015.

[2] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *European conference on computer vision*, pages 740–755. Springer, 2014.

[3] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3431–3440, 2015.

[4] Roozbeh Mottaghi, Xianjie Chen, Xiaobai Liu, Nam-Gyu Cho, Seong-Whan Lee, Sanja Fidler, Raquel Urtasun, and Alan Yuille. The role of context for object detection and semantic segmentation in the wild. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 891–898, 2014.

[5] Gabriel Pereyra, George Tucker, Jan Chorowski, Łukasz Kaiser, and Geoffrey Hinton. Regularizing neural networks by penalizing confident output distributions. *arXiv preprint arXiv:1701.06548*, 2017.

[6] Scott Reed, Honglak Lee, Dragomir Anguelov, Christian Szegedy, Dumitru Erhan, and Andrew Rabinovich. Training deep neural networks on noisy labels with bootstrapping. *arXiv preprint arXiv:1412.6596*, 2014.

[7] David Rolnick, Andreas Veit, Serge Belongie, and Nir Shavit. Deep learning is robust to massive label noise. *arXiv preprint arXiv:1705.10694*, 2017.

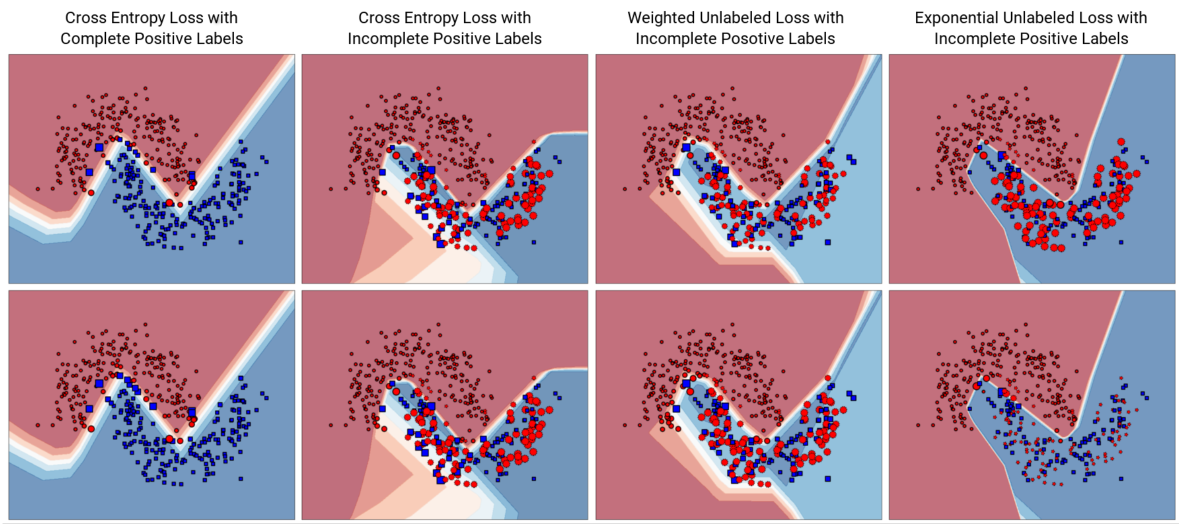

Figure 1: MOONS

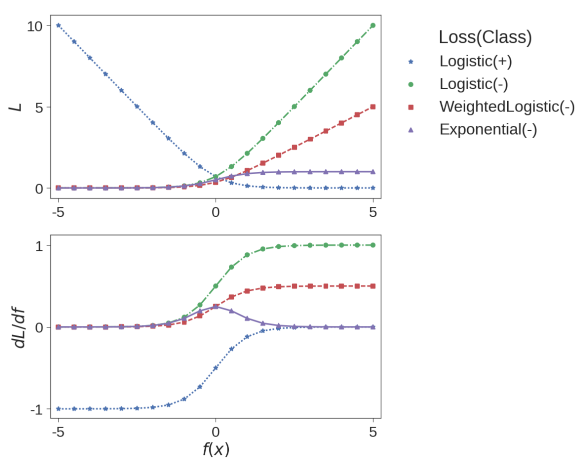

Figure 2: The Logistic Loss, Weighted Logistic Loss, Exponential Loss and their dirivatives.

Figure 3: VGG8 CIFAR10 with 10%, 20%, 50%, 80% and 100% of the positive examples annotated