

Sparse Labeling Augmentation for Dense Models Training

Íñigo Alonso, Ana C. Murillo

Afilación: Robótica, Percepción y Tiempo Real (ROPERT)
Instituto de Investigación en Ingeniería de Aragón (I3A)
Universidad de Zaragoza, Mariano Esquillor s/n, 50018, Zaragoza, Spain.
Tel. +34-976762707, e-mail: inigo@unizar.es

Abstract

This work proposes and validates a simple but effective approach to train dense semantic segmentation models from sparsely labeled data. Data and labeling collection is most costly task of semantic segmentation. Our approach needs only a few pixels per image reducing the human interaction required.

Problem Statement

Semantic segmentation is a challenging problem that can benefit numerous robotics applications, since it provides information about the content of the image per pixel. Solutions to this problem have recently witnessed a boost on performance and results thanks to deep learning approaches. Unfortunately, common deep learning models for semantic segmentation present several challenges which hinder real life applicability in many domains.

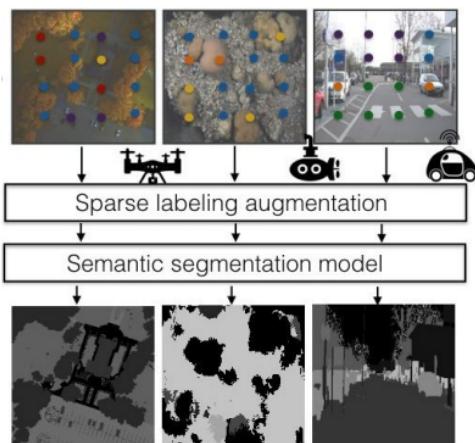


Fig. 1. Robotic platforms have enabled easy collection of plenty of monitoring datasets in different domains. This work demonstrates that augmenting sparse input data labels enables effective training of semantic segmentation models to process them with much lower labeling effort.

A significant one is the need of pixel level labeling on large amounts of training images, to be able to train those models, which implies a very high cost.

Labeling Augmentation

Our approach is based on augmenting the sparse annotation to a dense one with the proposed adaptive superpixel segmentation propagation. We show that this label augmentation enables effective learning of state-of-the-art segmentation models, getting similar results to those models trained with dense ground-truth. We demonstrate the applicability of the presented approach to different image modalities in real domains (underwater, aerial and urban scenarios) with publicly available datasets (see Fig. 1).

Superpixel segmentation

Our strategy for label augmentation is based on existing superpixel segmentation techniques. These techniques cluster the image pixels creating groups of similar connected pixels (known as superpixels). The basic single-level[1] superpixel based augmentation strategy is simple:

First, the image is segmented into superpixels. Then, the labeled pixels information is propagated following the superpixel segmentation, i.e., all pixels in each superpixel get the label that appears the most within that superpixel.

The multi-level superpixel segmentation proposed consists of applying iteratively the superpixel segmentation to the image, progressively decreasing the number of superpixels generated in each iteration. In the first iteration, the number of superpixels is very high, leaving a lot of unlabeled pixels in the augmented labeling but being able to find small regions. The following iterations continue increasing the size of the superpixels until they manage to fill all the unlabeled pixels (see Fig. 2).

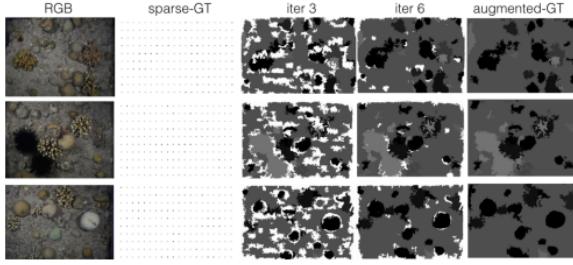


Fig. 2 Three examples of the multi-level superpixel augmentation algorithm. From left to right, the input RGB image, the available labels (sparse GT), and the augmented labels after a few iterations (3 and 6) and the final augmented labels (augmentedGT) after 10 iterations.

Semantic Segmentation

Semantic segmentation is formulated as a pixel classification problem. In this case, an end-to-end CNN architecture is trained from dense input ground truth labels to obtain directly the classification for each pixel, i.e., the final semantic segmentation. We consider the most common fully convolutional architectures for this problem: the FCN architecture and the symmetric encoder-decoder.

Both approaches consist of a classification problem, whose model is obtained by minimizing of the error between the predicted and expected value.

Some results can be seen in Figure 3.

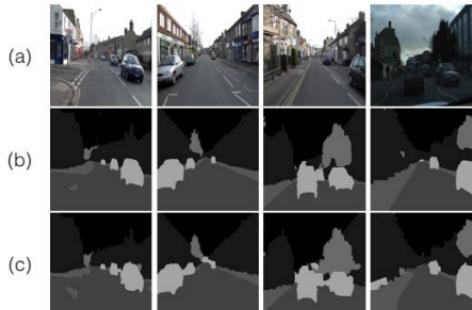


Fig. 3 Semantic segmentation of four examples using a model trained with the dense labeling (b) and a model trained with our proposed augmented labeling (c). Images (a) from Camvid dataset.

Conclusion

We have presented a novel approach to augment labeled data, at pixel level, to facilitate training semantic segmentation models. As shown, our approach enables state of the art architectures for semantic segmentation in scenarios where there are only sparse labels available on the training images.

More generally, it also benefits any scenario by lowering the labeling requirements to train new models on new domain datasets which still need to be labeled.

We have demonstrated that the proposed augmented labeling is effective to train encoder-decoder CNN models for segmentation, reaching comparable results to those obtained by training with much more costly to obtain dense ground-truth labels.

The use of grid-based sparse data was motivated because it is actually available in real world use case. As future steps though, we plan to explore different types of sparsity distributions and extend the applicability to other types of data like 3D information.

REFERENCIAS

- [1]. ALONSO, I. CAMBRA, A. MUÑOZ, A. TREIBITZ, T., MURILLO A.C. Coral-Segmentation: Training dense labeling models with sparse ground-truth. Proc. of the IEEE Int. Conf. On Computer Vision Workshops pages 2874-2882, 2017.