Politecnico di Torino

ScuDo
Scuola di Dottorato ~ Doctoral School
WHAT YOU ARE, TAKES YOU FAR

Doctoral Dissertation
Doctoral Program in Computer and Control Engineering (35^{th}cycle)

Learning New Classes from Limited Data in Image Segmentation and Object Detection

By

Fabio Cermelli

******

Supervisor:
Prof. Barbara Caputo

Doctoral Examination Committee:
Prof. Bernt Schiele, Referee
Prof. Diane Larlus, Referee

Politecnico di Torino
2023

Declaration

I hereby declare that, the contents and organization of this dissertation constitute my own original work and does not compromise in any way the rights of third parties, including those relating to the security of personal data.

Fabio Cermelli
2023

* This dissertation is presented in partial fulfillment of the requirements for **Ph.D. degree** in the Graduate School of Politecnico di Torino (ScuDo).

Learning New Classes from Limited Data in Image Segmentation and Object Detection

Fabio Cermelli

Image segmentation is a critical capability for autonomous systems to understand their surroundings. Although deep neural networks have enhanced image segmentation performance, they require expensive and massive datasets for training, and they cannot update their knowledge for new classes without experiencing catastrophic forgetting. In the first part of this thesis, we address the issue of catastrophic forgetting by analyzing a unique aspect of semantic segmentation that exacerbates it. At each training step, the annotation only covers the new classes, while other classes, such as the ones already learned by the model, may appear in the image as background. To solve this, we propose a simple yet efficient solution that revisits the knowledge distillation framework and explicitly models this peculiarity. We also extend this approach to incremental learning in object detection and instance segmentation. In the second part of the thesis, we investigate learning new classes by reducing the number of images needed. We introduce the incremental few-shot semantic segmentation setting, where the model must learn new classes using only a few images. We propose a method for this novel setting that combines prototype learning and knowledge distillation, effectively preventing the model from forgetting old classes and overfitting the few images. Additionally, we explore the extreme setting where no labels are available for novel classes, proposing a self-training solution that extracts supervision from the unlabeled pixels in the training set. Finally, in the last part of the thesis, we aim to learn a segmentation model without relying on expensive pixel-level annotations, using cheaper alternatives instead. We suggest a general loss function to learn from points and scribbles, exploiting the assumption that all pixels in the image must belong to one of the annotated classes. Furthermore, we investigate the use of image-level labels for incremental learning in semantic segmentation. We present a new setting where a pre-trained model is trained to predict new classes using image-level labels. Building on the knowledge distillation framework, we propose an approach that integrates a localizer to extract pixel-level pseudo-supervision from image-level labels, which trains the model on novel classes without forgetting old ones.