Abstract

Deep learning is the dominant approach in modern computer vision. However, its success mainly hinges on the availability of large scale annotated datasets for training, and capturing the infinite semantic diversity of the real world into one or more training sets is, unfortunately, unfeasible. Consequently, deep neural networks are forced to limit their understanding of the world to the restricted knowledge available during the training phase. In this thesis, we argue that, to develop deep neural networks capable of operating in the real world, it is vital to empower them with the capabilities of i) detecting previously unseen concepts and ii) incrementally integrating them in subsequent learning stages. In the first part of this thesis, we address the aforementioned challenges separately. We first address the anomaly segmentation problem, which involves identifying for each pixel of an image whether it belongs to a previously unseen category, *i.e*. an anomaly. We propose to segment anomalies using class-specific prototypes extracted from a cosine classifier, and to determine pixels to be anomalous when the highest matching score between a pixel and the set of known prototypes is below a certain threshold. We then address the challenges of incremental learning, which involves incrementally updating existing models as new categories become available. Despite advancements in the field, state-of-the-art semantic segmentation strategies still require supervision at pixel-level on new classes, which is often costly and time-consuming to acquire. In this thesis we present a new perspective to the field, showing how to incrementally extend the knowledge of a pre-trained segmentation model using only cheap image-level labels, which provide information only on the presence of a certain class but not on its appearance or location. We demonstrate that directly applying existing weakly-supervised segmentation strategies to the traditional incremental segmentation ones is sub-optimal, and we propose to use a localizer module to produce pseudo-labels and a distillation-based loss to prevent forgetting previously learned classes. In the second part of this thesis, we address the open world recognition (OWR) setting

to tackle the two challenges simultaneously. Differently from prior works, we demonstrate that learning a separate rejection threshold for each class is crucial to reduce the number of samples wrongly identified as never-seen-before ones. To achieve this, we shape the representation space to be semantically consistent through a global-to-local clustering approach, that enforces samples to be closer to the centroid of the respective class, while pushing away samples from other classes. The training sets, however, impose not only semantic limitations on agents, but also environmental ones, due to the inherent bias towards specific acquisition conditions that do not necessarily represent the high variability of the real world. Therefore, in the final part of the thesis, we investigate the impact of different training and test distributions (domain-shifts) on OWR frameworks. We introduce the first benchmark to assess OWR methods under domain-shifts, and we show that existing OWR strategies significantly suffer from performance degradation when the train and test distributions differ. We demonstrate that coupling OWR methods with domain generalization algorithms mitigates this degradation, but their simple integration is not sufficient to identify new and unknown categories in unfamiliar domains. We then highlight open challenges and future research directions, that serve as foundations towards developing agents capable of reliably operating in real open world environments.