Learn to Generalize and Adapt across Domains in Semantic Segmentation

Antonio Tavera

Artificial Intelligence (AI) is a rapidly evolving field that has the potential to transform our world in countless ways. A vital part of AI is Computer Vision, which focuses on developing systems and algorithms that can interpret and comprehend visual information from the world around us, such as through Semantic Segmentation - a technique of assigning a distinct class label to each pixel in an image, grouping all pixels that belong to the same object or region under the same label. Semantic Segmentation has many critical applications such as autonomous driving or aerial images understanding. In Autonomous Driving, it is used to accurately recognize and classify different objects and regions in images received from sensors to make informed decisions about safe navigation. In Aerial Images Analysis, it can be useful for a variety of tasks including mapping, land use planning, and disaster response to identify and map damaged infrastructure and impacted areas. Nevertheless, semantic segmentation has several limitations related to data availability and quality, such as a limited diversity in training data, lack of annotation, poor quality of annotation, and imbalanced classes. To address this challenges, the purpose of this thesis was to explore and develop solutions that would make the neural models more robust and capable of generalizing to different domains from the ones they were trained on. One way to overcome these issues is through the use of synthetic datasets, which are computer-generated images that can be generated in large quantities and do not require manual annotation. For this reason we present IDDA, the largest synthetic dataset for autonomous driving, with over 100 different scenarios that allow to assess the domain generalization capability of semantic segmentation models. However, the use of synthetic datasets can present a significant challenge when it comes to generalizing the model to real-world scenarios. Synthetic datasets lack the complexity and diversity and may not include the same types of noise, occlusions, and other factors that are present in a real-world data. To overcome this problem, domain adaptation techniques can be used. In particular, few-shot domain adaptation, which allows for a more efficient use of real-world annotated data, may be a potential solution. The PixDA technique that we present uses a limited amount of annotated real-world data to prioritize pixel alignment based on class imbalance and network classification confidence, resulting in increased accuracy. Despite its effectiveness in self-driving scenarios, understanding aerial scenes faces additional challenges such as severe camera angle distortions and a lack of reference points. To address these challenges, advanced techniques like the new loss that we present in AIAS can be used. In this context, our HIUDA framework presents a new mixing strategy that is specifically designed for aerial images, taking into account their specific challenges and helping to prevent elements from being placed in unnatural contexts. The effectiveness of the proposed solutions is evaluated using both real-world and synthetic datasets, showing their superiority in comparison to previous methods.