



Politecnico
di Torino

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

Doctoral Dissertation

Doctoral Program in Artificial Intelligence (37th cycle)

From pixels to masks: Toward open-set segmentation

By

Shyam Nandan Rai

Supervisor(s):

Prof. Barbara Caputo, Supervisor

Prof. Carlo Masone, Co-Supervisor

Prof. Zeynep Akata, Co-Supervisor

Doctoral Examination Committee (in alphabetical order):

Prof. Cyrill Stachniss, University of Bonn, *Referee*

Prof. Fatma Güney, Koç University

Prof. Pietro Zanuttigh, Università degli Studi di Padova

Prof. Siniša Šegvic, University of Zagreb, *Referee*

Politecnico di Torino

2025

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.

Shyam Nandan Rai
2025

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

Abstract

Autonomous navigation agents such as self-driving cars can operate without human intervention. A key challenge is ensuring these agents can navigate safely and reliably, which requires a deep semantic understanding of their surroundings. To achieve this, deep learning based scene understanding techniques are designed to enhance autonomous systems by enabling object localization and segmentation at the pixel level. However, these scene understanding methods fail to segment unseen or anomalous objects, encountered by autonomous agents in real-world scenarios.

This thesis explores the problem of open-set segmentation, with a particular emphasis on the related subtask of anomaly segmentation for road scenes. The key contributions can be summarized in three key areas. Firstly, we propose a paradigm change in existing anomaly segmentation architecture by shifting from a per-pixel classification to a mask classification. Our mask-based method, Mask2Anomaly demonstrates the feasibility of integrating a mask-classification architecture by providing better localization of anomalies with reduced false positives. Next, we extend Mask2Anomaly to jointly address anomaly segmentation, open-set semantic segmentation, and open-set panoptic segmentation. Secondly, we examine the issue of training parameter inefficiency ingrained in existing anomaly segmentation methods that require fine-tuning large transformer backbones. To address this, we present dual-stream adapters for open-set segmentation that provide a training parameter efficient approach for anomaly segmentation while preserving known class semantic segmentation performance. Finally, in the last part of this thesis, we hypothesize incorporating temporal information leads to more accurate anomaly segmentation as the dependencies between consecutive frames enable models to leverage contextual cues from neighboring frames. So, we introduce road-obstacle video segmentation, a task designed to identify and segment anomalies or obstacles in road scene videos.