Doctoral Dissertation
Doctoral Program in Electric, Electronic and Communication Engineering
(37^{th} cycle)

Efficient and Robust Deep Learning for Robotics

Towards real-time AI systems that generalize to unseen scenarios

By

Simone Angarano

******

Supervisor(s):
Prof. Marcello Chiaberge

Politecnico di Torino
2025

Abstract

In recent years, the advancement of deep learning research has driven the creation of large models capable of performing complex tasks with high accuracy. However, it is crucial to efficiently apply these advancements in fields like robotics, where computational constraints are crucial.

The integration of artificial intelligence in robotics has revolutionized the capabilities of autonomous systems across diverse domains - from domestic assistance to precision agriculture. However, deploying state-of-the-art deep learning models in robotic applications presents significant challenges. Computational constraints, strict latency requirements, and robust performance in real-world scenarios create a pressing need for efficient yet powerful AI solutions that can be practically implemented on resource-limited robotic platforms.

This dissertation addresses these challenges head-on, developing innovative deep learning architectures and training methodologies that bridge the gap between cutting-edge advancements and the practical realities of service robotics. The key focus is enhancing efficiency while preserving the remarkable capabilities that deep learning has demonstrated, particularly in computer vision tasks essential for intelligent perception and decision-making in unstructured environments.

A primary contribution of this work is the adaptation of computationally intensive neural network architectures, such as Transformers, to make them viable for real-time robotic applications. Through careful architectural optimization, we demonstrate that state-of-the-art models can be practical for resource-constrained platforms without sacrificing high performance. This new direction is exemplified by our development of an efficient transformer for real-time human action recognition, enabling natural human-robot interaction in populated environments. By

crafting efficient variants of contemporary models, we have opened the door for the deployment of powerful deep learning capabilities on robotic systems.

In addition to these architectural innovations, this dissertation also advances the field of knowledge distillation - a technique for compressing large, complex models into efficient student networks. For instance, we introduce a standardized ensemble knowledge distillation approach that not only reduces model size but also significantly enhances generalization across diverse scenarios, a critical requirement for real-world robotic deployment. This methodology has been particularly impactful in the context of crop segmentation, where we demonstrate robust performance across various agricultural settings. By effectively transferring knowledge from high-capacity teachers to compact student models, we start paving the way to the democratization of advanced deep learning capabilities in service robotics.

Beyond these contributions to efficient model design and training, this work presents a comprehensive deep learning-driven pipeline for autonomous navigation in crop fields. The system integrates satellite imagery processing, contrastive learning-based waypoint generation, and GPS-agnostic local planning using semantic segmentation. This end-to-end solution showcases the practical application of our developed deep learning methodologies, exhibiting robust performance across diverse environmental conditions. By seamlessly combining various neural network-powered components, we have created a powerful robotic system capable of navigating complex agricultural settings.

The techniques and findings presented in this dissertation have been extensively validated across multiple real-world applications. The demonstrated improvements over existing approaches highlight our work's significant impact in bridging the gap between state-of-the-art deep learning and the practical constraints of service robotics.