Context-Aware Learning for Enhancing Traffic Helmet Violation Detection

Phan Nguyen Huu Phong   Nguyen Tien Huy

University of Information Technology - National University of Vietnam

Abtract

In urban settings, vehicular accidents, particularly motorcycle incidents, significantly contribute to property damage and loss of life. This paper introduces a novel traffic helmet detection system leveraging context-aware learning to address limitations in existing methodologies under complex environmental contexts. We propose a Spatial-Channel Attention Module (SCAM) to focus on multi-scale objects, demonstrating competitive real-time detection outcomes with a lightweight model of only 8.9M parameters.

Introduction

Problem Statement: Increasing traffic accidents, especially in undeveloped countries, are often due to the non-utilization of helmets.

Challenge: Existing models are computationally intensive and not practical for real-world application.

Solution: Automated helmet detection system leveraging context-aware learning to adapt to different traffic conditions.

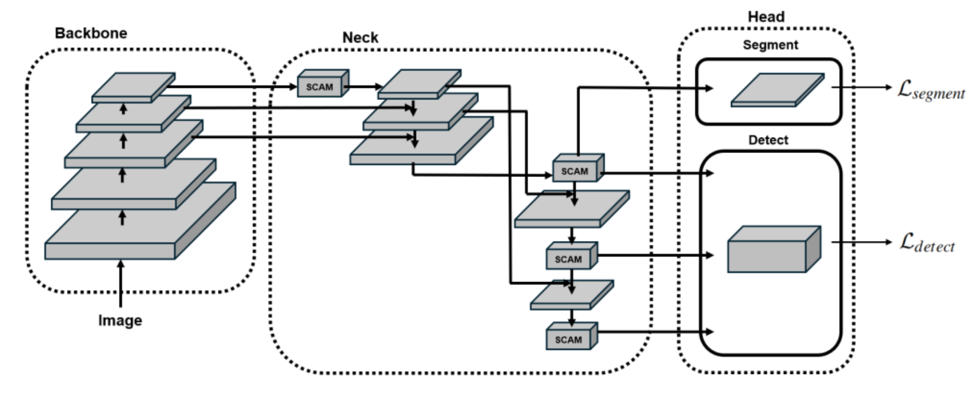
Methodology

1. Proposed Dataset

Dataset: 91,000 annotated images from 12 cities in Myanmar.

Automatic Segmentation: Using Mask2Former pre-trained on the CityScapes dataset to create segmentation maps.

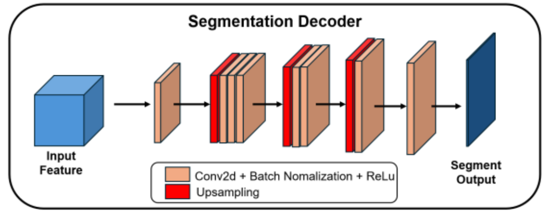
2. Network Architecture



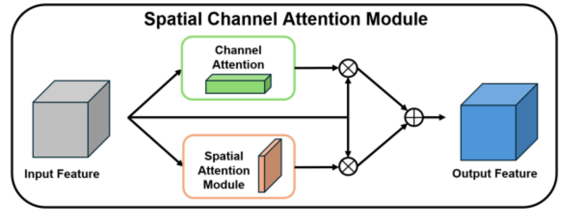
Backbone: Extracts image features at five levels.
Neck: Aggregates information from different levels using attention modules.
Head: Divided into Segment Decoder (for segmentation predictions) and Object Detector (for final detection outputs).

3. Context-Aware Learning:

Semantic segmentation is integrated into the network to provide contextual information, enhancing object detection performance.



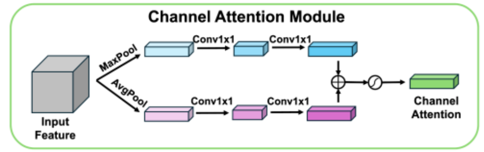
4. Spatial – Channel Attention Module (SCAM)



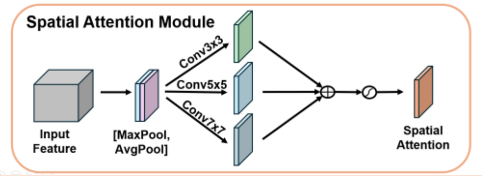
Combines channel and spatial attention mechanisms to focus on multi-scale objects.

Channel Attention Module

Recalibrates the weight of each channel and uses it to select significant channels.



Spatial Attention Module



Highlights important spatial regions, recognizing features across multiple scales.

Experiments:

1. Results

Method	Model	Backbone	mAP	Parameters
Transformers	Cascade R-CNN	Swin Transformer	30.4	>80M
	Faster R-CNN	Swin Transformer	27.6	>40M
	DETR	ResNet-50	27.2	41M
	Deformable DETR	ResNet-50	22.9	40M
CNN	RetinaNet	ResNet-50	25.9	34M
	YoloV7-L	ELANet	28.6	>70M
	PP-YOLOE	CSPResNet	18.3	52.2M
	PP-YOLOv2	ResNet-50	16.6	54.6M
Ours	Ours	Ours	**18.5**	**8.9M**

2. Ablation Study

Method	mAP (%)
without SCAM	17.4
with SCAM	18.5

1. Our model demonstrates competitive performance with fewer parameters compared to other models.
2. Effectiveness of SCAM showing improvements in mAP.