A Foveated Approach to Automated Billboard Detection

IARIA

By

Sayali Avinash Chavan **Dermot Kerr** **Sonya Coleman** **Hussein Khader***

Intelligent Systems Research Centre

University of Ulster, UK

Ulster University

*The Neuron, Amman, Jordan

email:{chavan-s, d.kerr, sa.coleman}@ulster.ac.uk, hussein.khader@theneuron.com

About Me: Sayali

GitHub in

PhD Researcher Ulster University, UK

- Computer Vision for Advertising Analytics
- Studentship: Digital Natives Postgraduate Scholarship (International Award)
- Publication: Billboard Detection in the Wild

Ulster University Business School

Seminar Support staff

Data Analytics

Business School Ulster University, UK

PG: MSc in Advanced Computer Science

Loughborough University, UK

- British Council Ambassador – Full Scholarship
- Dissertation : Machine Learning for Wildlife Conservation

Associate Fellowship by Advance HE

- Level 7 Postgraduate CPD Framework UK
- Award: First Steps to Supporting Learning and Teaching in Higher Education (FST)

[SPPU] Pune University

UG: Bachelors of Engineering in Electronics & Telecommunication

- Pune University - Dean list rank.

Application Development Analyst Accenture, India

Introduction

Billboards play a significant role in outdoor advertising, aiming to capture the attention of individuals and deliver brand messages effectively.

Detecting billboards automatically with computer vision can provide valuable insights for this purpose.

It allows advertisers and marketers to assess the impact and reach of their advertisement campaigns. Making informed decisions and optimise their advertising strategies for maximum impact.

Background

Challenges in Billboard Detection

- Localisation difficulties with multiple objects in the scene.
- Dependency on high-resolution cameras for accurate detection.

Focus on Text and Content Recognition

- Machine learning-based approaches for content analysis.
- Detection of image manipulation and adherence to perspective rules.
- Use of Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) for text recognition.
- Compliance-related research on identifying advertisements on buildings.

Limitations of Previous Approaches

- Lack of comprehensive approaches to detect billboards regardless of their content.
- Challenges with small object sizes, cluttered backgrounds, and low resolutions in billboard detection.
- Scalability (ADNet) and effectiveness concerns in diverse real-world scenarios.
- Previous research heavily reliant on high-resolution limited datasets.

Paper Contributions

Comprehensive approach to billboard detection using street level images.

We have developed a robust billboard detection system by leveraging state-of-the-art object detection models, such as YOLOv8, YOLOv5, Faster-RCNN and CenterNet resulting in high model accuracy.

We have introduced an innovative foveated approach, that applies a Gaussian function to assign weights to billboards to determine which is the most significant billboard based on a combination of confidence and location with respect to the image centre.

The approach demonstrates an improvement in overall accuracy of the detection process. **YOLOv8** experienced a high accuracy increase from **63.40** to **82.71** percent.

Dataset

Total number of Images: **3,437**

73% training

20% validation

7% test.

Fig. 1. Example of a low-quality or irrelevant image (blurry image) filtered during dataset cleaning process

Fig. 2. Training sample from image dataset for UK region showcasing billboards in real-life environments with surrounding vehicles and background scenery

Training Process

- Input images.
- Loss function - calculated based on the model predictions and ground truth.
- Model parameters - updated using gradient descent optimisation, minimising the loss.
- The training process iterates over the dataset multiple times depending on the number of steps/epochs to improve the model's performance.

Model Selection

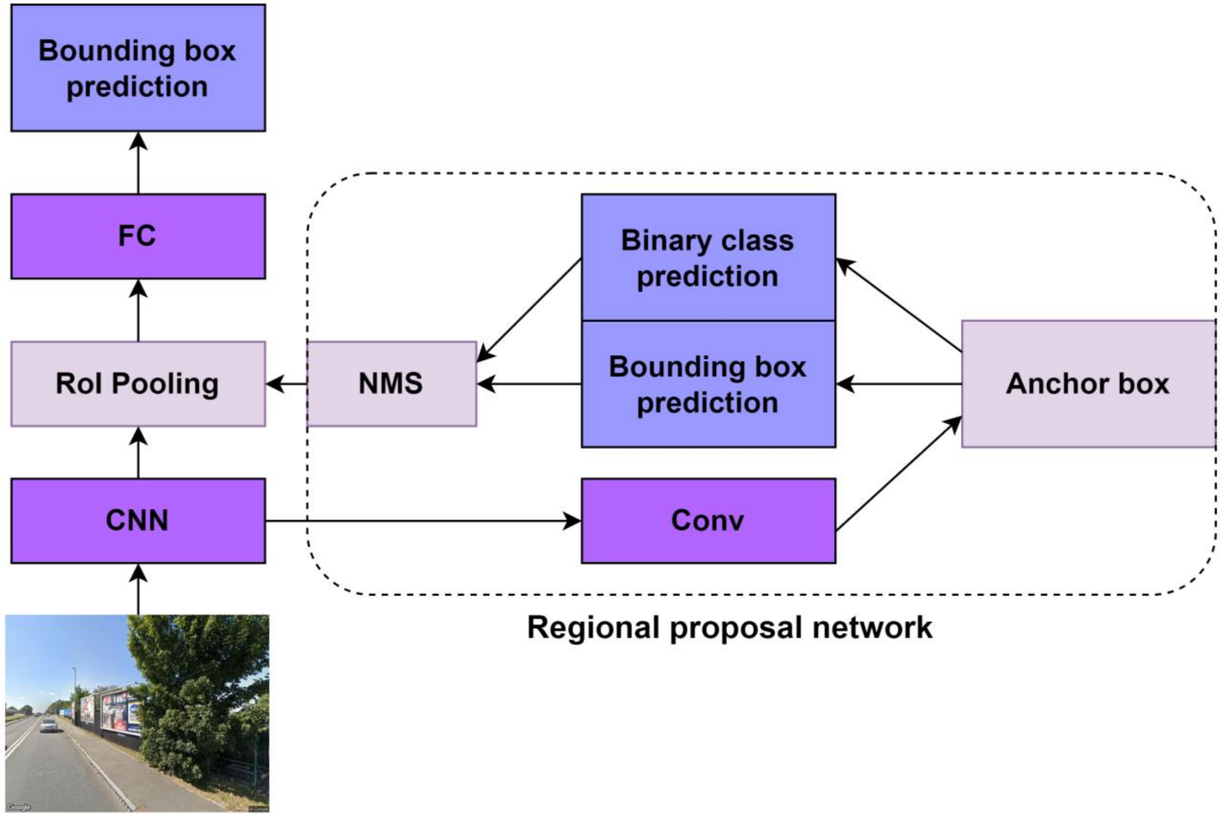
Faster R-CNN stands out for its exceptional accuracy in object detection. By utilising a region proposal network, it generates potential object locations and then fine tunes the model for improved localisation and recognition, which is essential for applications requiring reliable results.

CenterNet focuses on estimating the centre points of objects. Its centre point estimation approach allows it to handle and distinguish multiple objects efficiently, making it suitable for scenarios where multiple billboards may be present in the scene.

YOLO is renowned for its impressive speed in object detection in real-time. Moreover, **YOLOv8**, the latest version released in 2023, is designed to be computationally efficient enabling faster processing, which can be advantageous in scenarios where quick detection is crucial.

Faster R-CNN

Faster R-CNN (Region Convolutional Neural Network) that revolutionised the field by combining accuracy and efficiency.



- **Backbone Network:** Extracts feature maps from the input image.
- **Region Proposal Network (RPN):** Generates potential object regions.
- **RoI Align:** Fine-tunes the feature extraction process for accurate object localisation.
- **Classification and Regression:** Classifies detected objects and refines their bounding box coordinates.

- **Advantages**:

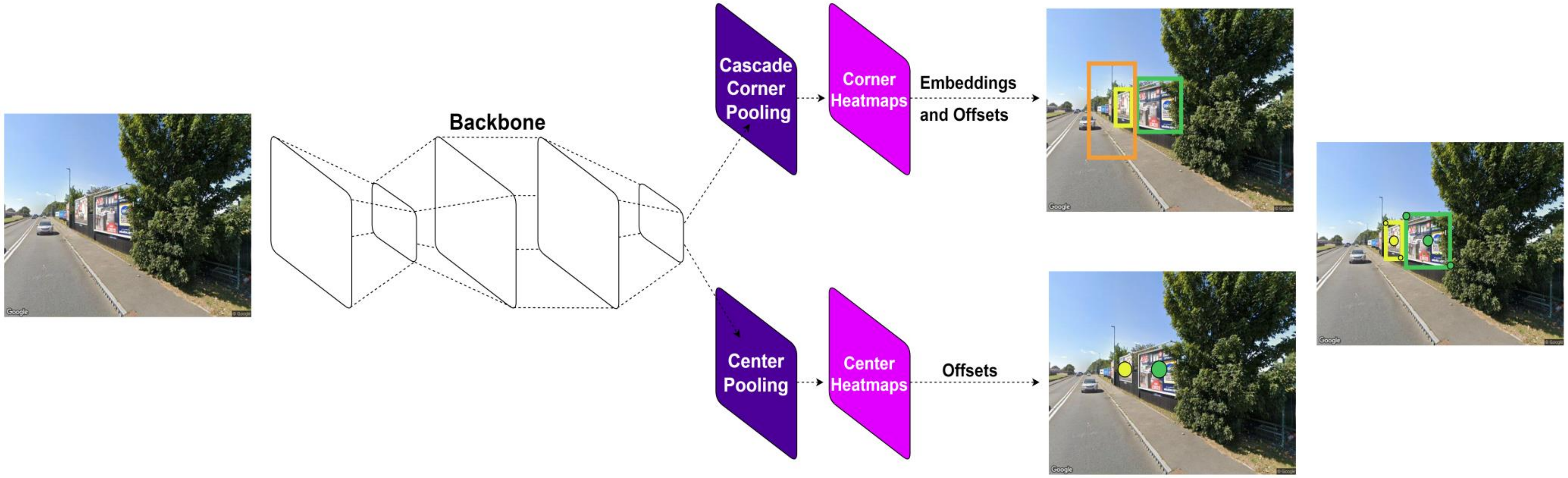
- High accuracy in object detection.
- Accommodates various object sizes and aspect ratios.
- Robustness in detecting complex scenes.

- **Limitations:**

- Relatively slower than some newer models.
- Can be memory-intensive due to multi-stage architecture.

CenterNet

CenterNet presents a unique approach to object detection by directly predicting object centers and their attributes, simplifying the detection pipeline.

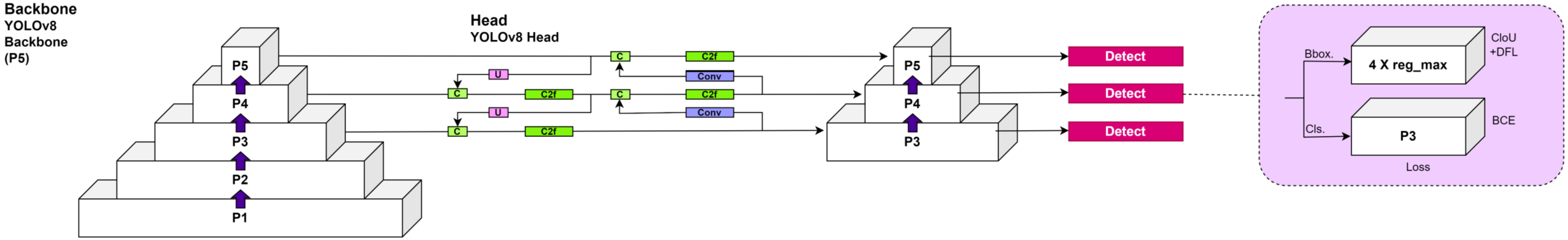


- **Keypoint Estimation:** CenterNet predicts object centers as keypoints, allowing for precise localization without complex bounding box regression.
- **One-Stage Detector:** simplifying training and inference.
- **Object Attributes:** Predicts additional attributes like object size and orientation alongside center points.

YOLO: YOLOv5, YOLOv8

YOLO, particularly YOLOv5 & latest YOLOv8, is a family of real-time object detection models known for their impressive speed and accuracy trade-off.

- **Grid-based Detection:** YOLO divides the image into a grid and predicts bounding boxes, object probabilities, and class labels within each grid cell.
- **Single Forward Pass:** YOLO performs object detection in a single forward pass, making it exceptionally fast.



- **Input Division:** Divides the image into a grid of cells.
- **Predictions:** For each cell, predicts bounding boxes, objectness scores, and class probabilities.
- **Non-Maximum Suppression:** Removes duplicate detections
- **YOLOv8 Improvements:** It combines the Feature Pyramid Network (FPN) and the Path Aggregation Network (PAN). FPN in YOLOv8 gradually reduces the spatial resolution of the input image while increasing the number of feature channels. This results in the creation of feature maps that can effectively detect objects at different scales and resolutions.

Hyper-Parameters

Model	Faster RCNN	CenterNet	YOLOv5	YOLOv8
Image Size	640	512	416	800
Batch Size	1	1	32	16
Optimizer	Momentum	Adam	SGD	SGD
Step Size/ Epochs	0-2000	0-2000	0-100	0-50
Warm Up Learning Rate	0.001	0.001	0.001	0.001
Step Size/ Epochs	2000-25,000	2000- 25,000	0-100	0-50
Final Learning Rate	0.004	0.004	0.001	0.001

- **Momentum Optimizer**: Accelerates convergence by accumulating velocity from previous updates.
- **Adam Optimizer**: Combines concepts from Momentum and RMSprop. Adapts learning rates for each parameter based on past gradients.
- **Stochastic Gradient Descent (SGD) Optimizer**: Updates parameters based on the gradient of the loss function computed using a small subset of training data

Gaussian Weighting Algorithm

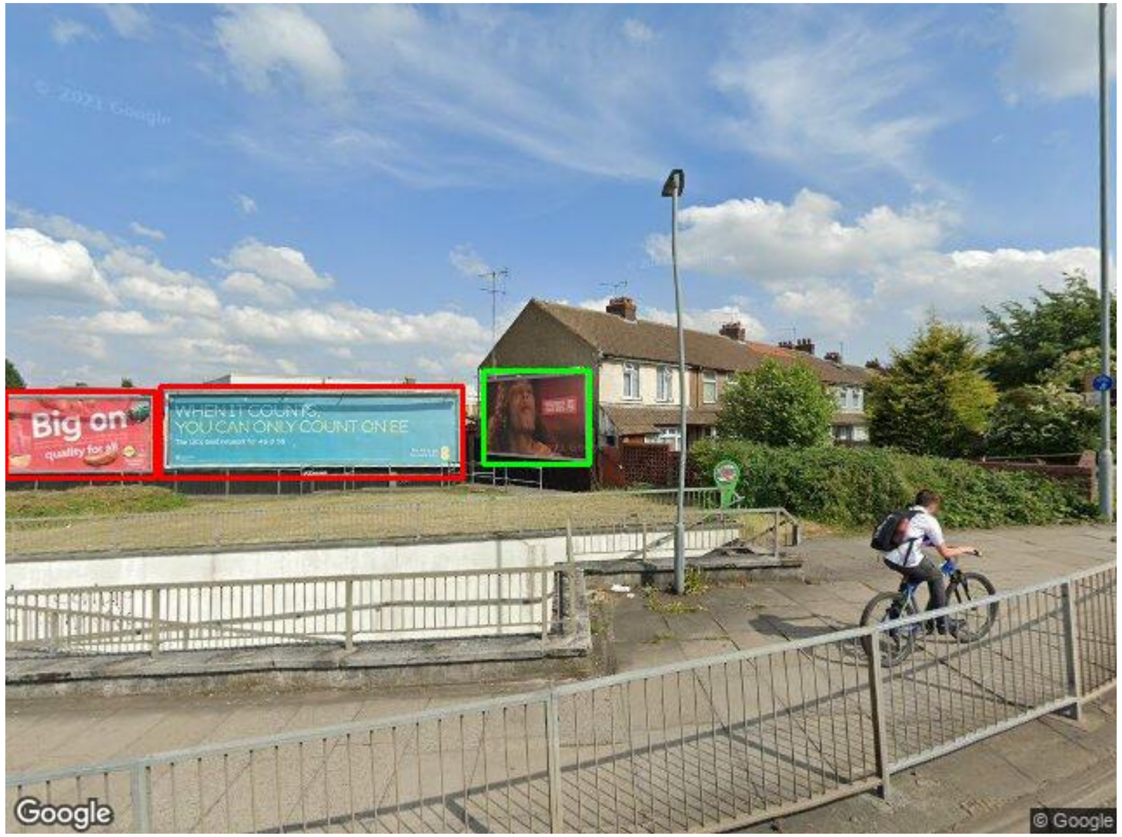
To further enhance the accuracy of billboard detection Gaussian weighting was applied to the detected object’s centre.

$$\textbf{Gaussian Distribution} = \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

- Gaussian weight is multiplied by computed object's confidence score, this combined score is compared to 50% threshold
- **Result** : Green bounding box drawn for scores above threshold to indicate Billboard is in the Center, Red bounding box drawn for scores below threshold.

Test Images

(a)



(b)

Result and Analysis

UK Data ->	Faster RCNN	CenterNet	YoloV5	YoloV8
Epoch/ Step size	25,000	25,000	100	50
Training Accuracy (mAP.50)	59%	75.7%	64.5%	**87.4%**
Testing Accuracy (mAP.50)	54.3%	57.2%	52.6%	63.4%
With Gaussian Testing Accuracy	65.43%	55.55%	80.24%	**82.71%**

Future Work

Could explore further refinements to optimise the proposed approach and extend it to real-time billboard detection systems, encompassing the task of verifying the relevance of displayed information.

This would involve ensuring that billboards continuously present accurate and up-to-date content, addressing scenarios where some billboards may no longer convey valid information.

Furthermore, exploring the billboard visibility based on environmental conditions, as well as the unique perspectives offered by different viewing angles, resulting in more effective outdoor advertising.

Conclusion

Challenges involved in this endeavor included finding optimal, hyperparameters, mitigating over-fitting, and efficiently managing computational resources during the training process, all of which we adeptly addressed and resolved during the development of these networks.

we introduced an innovative approach by applying a Gaussian weighting technique to determine the most central billboards.

This significantly improved the overall accuracy of the detection process, particularly in the case of YOLOv8, which achieved an impressive accuracy of 82.71%

Questions?

Thank You!