Comenius University in Bratislava
Faculty of Mathematics, Physics and Informatics

Evaluating the significance of outdoor advertising utilizing Computer Vision

Master's Thesis

2024

Bc. Carlos Andres Pizarroso Troncoso

Comenius University in Bratislava
Faculty of Mathematics, Physics and Informatics

Evaluating the significance of outdoor advertising utilizing Computer Vision

Master's Thesis

Study Programme:	Applied Computer Science
Field of Study:	Computer Science
Department:	Department of Applied Informatics
Supervisor:	RNDr. Zuzana Berger Haladová, PhD.

Bratislava, 2024
Bc. Carlos Andres Pizarroso Troncoso

22898864

Comenius University Bratislava
Faculty of Mathematics, Physics and Informatics

THESIS ASSIGNMENT

Name and Surname:	Carlos Andres Pizarroso Troncoso
Study programme:	Applied Computer Science (Single degree study, master II. deg., full time form)
Field of Study:	Computer Science
Type of Thesis:	Diploma Thesis
Language of Thesis:	English
Secondary language:	Slovak

Title: Evaluating the Significance of Outdoor Advertising utilizing Computer Vision

Annotation: The goal is to create and test a method that, based on a photograph (or video), will evaluate the importance of a billboard. The work will use deep neural networks and the BillboardLamac database

Aim:
1. State of the art review of deep learning object detectors and trackers
2. Choose 3 methods for object detection on image/videos and test them on the Billboard Lamac dataset
3. Create a new method for classification of billboard significance
4. Validate the propose method

Literature: Zou, Zhengxia, et al. "Object detection in 20 years: A survey." Proceedings of the IEEE (2023).
H. Marciano et al., “The effect of billboard design specifications on driving: a pilot study,” Accident Analysis & Prevention, vol. 104, pp. 174–184, 2017.
Z. Bylinskii, T. Judd, A. Oliva, A. Torralba, and F. Durand, “What do different evaluation metrics tell us about saliency models?” IEEE transactions on pattern analysis and machine intelligence, vol. 41, no. 3, pp. 740–757, 2018.

Supervisor:	RNDr. Zuzana Berger Haladová, PhD.
Department:	FMFI.KAI - Department of Applied Informatics
Head of department:	doc. RNDr. Tatiana Jajcayová, PhD.
Assigned:	06.10.2023
Approved:	14.11.2023

prof. RNDr. Roman Ďurikovič, PhD.
Guarantor of Study Programme

.... Student

.... Supervisor

22898864

Univerzita Komenského v Bratislave
Fakulta matematiky, fyziky a informatiky

ZADANIE ZÁVEREČNEJ PRÁCE

Meno a priezvisko študenta:	Carlos Andres Pizarroso Troncoso
Študijný program:	aplikovaná informatika (Jednoodborové štúdium, magisterský II. st., denná forma)
Študijný odbor:	informatika
Typ záverečnej práce:	diplomová
Jazyk záverečnej práce:	anglický
Sekundárny jazyk:	slovenský

Názov: Evaluating the Significance of Outdoor Advertising utilizing Computer Vision
Vyhodnocovanie dôležitosti billboardov z využitím počítačového videnia

Anotácia: Cieľom je vytvoriť a otestovať metódu, ktorá na základe fotografie (alebo videa) vyhodnotí dôležitosť billboardu. Práca bude využívať hlboké neurónové siete a databázu BillboardLamac.

Cieľ:
1. State of the art review of deep learning object detectors and trackers
2. Choose 3 methods for object detection on image/videos and test them on the Billboard Lamac dataset
3. Create a new method for classification of billboard significance
4. Validate the propose method

Literatúra:

Zou, Zhengxia, et al. "Object detection in 20 years: A survey." Proceedings of the IEEE (2023).
H. Marciano et al., “The effect of billboard design specifications on driving: a pilot study,” Accident Analysis & Prevention, vol. 104, pp. 174–184, 2017.
Z. Bylinskii, T. Judd, A. Oliva, A. Torralba, and F. Durand, “What do different evaluation metrics tell us about saliency models?” IEEE transactions on pattern analysis and machine intelligence, vol. 41, no. 3, pp. 740–757, 2018.

Vedúci:	RNDr. Zuzana Berger Haladová, PhD.
Katedra:	FMFI.KAI - Katedra aplikovanej informatiky
Vedúci katedry:	doc. RNDr. Tatiana Jajcayová, PhD.
Dátum zadania:	06.10.2023

Dátum schválenia: 14.11.2023

prof. RNDr. Roman Ďurikovič, PhD.
garant študijného programu

.... študent

.... vedúci práce

Acknowledgments:

Abstrakt

Bilbordy pri cestách sú prominentným médiom vonkajšej reklamy, avšak tento typ marketingu môže potenciálne viesť k rozptýleniu počas jazdy, čím sa zvyšuje pravdepodobnosť nehôd. Na základe predchádzajúcej práce zahŕňajúcej súbor údajov BillboardLamac sa súčasná práca zameriava na výskum detekcie a klasifikácie billboardov. Aby sa to dosiahlo, bol najprv vyvinutý rámec detekcie objektov založený na YOLO s použitím podmnožiny dátového súboru Mapillary Vistas na počiatočné školenie a dátového súboru BillboardLamac na jemné doladenie. Táto metóda dosiahla spoľahlivé výsledky detekcie, čo viedlo k dvom modelom, ktoré dosiahli presnosť 93 % a 94 %. Okrem toho sa vyvíja snaha o klasifikáciu trvania pohľadu vodiča na billboardy, pričom sa využíva ďalšia časť súboru údajov BillboardLamac. Počiatočné štádiá vývoja klasifikátora sa zameriavajú na kategorizáciu trvania pohľadu do vopred definovaných tried. Tento prebiehajúci výskum má za cieľ prispieť k pochopeniu a prepojeniu medzi viditeľnosťou billboardov a pozornosťou vodičov, pričom sa zdôrazňuje úloha hlbokého učenia pri hodnotení vplyvu reklám na cestách.

Kľúčové slová: Keywordsslová—detekcia billboardu, detekcia objektov, modely YOLO, pozornosť vodiča, klasifikácia reklamy

Abstract

Roadside billboards are a prominent medium for outdoor advertisement, however, this type of marketing can potentially lead to distractions while driving, increasing the probability of accidents. Building on a previous work involving the BillboardLamac dataset, the current work aims to research into billboard detection and classification. To achieve this, a YOLO-based object detection framework was first developed, using a subset of the Mapillary Vistas dataset for initial training and the BillboardLamac dataset for fine-tuning. This method achieved reliable detection results, leading to two models which achieved 93% and 94% of accuracy respectively. Furthermore, an effort to classify driver gaze durations towards billboards is under development, making use of another portion of the BillboardLamac dataset. Early stages of the classifier development aim to categorize gaze durations into predefined classes. This ongoing research has the goal to contribute to the understanding and connection between billboard visibility and drivers' attention, emphasizing the role of deep learning in evaluating roadside advertisements' impact.

Keywords: Keywordsbillboard detection, object detection, YOLO models, driver attention, advertisement classification, neural networks

Contents

List of Tables

List of Figures

Introduction

Neural networks have been constantly evolving and adapting to our current technologies and needs, to the point that they are now present in various fields as part of systems designed to assist people with different tasks. Computer vision is one of these fields that we are particularly interested in, especially because humans can benefit significantly from it. Therefore, the current work focuses on computer vision tasks applied to road safety.

Roadside billboards are one of the most popular media for outdoor advertising, used by businesses to capture the attention of both pedestrians and drivers. However, while this method is effective for marketing purposes, advertisements can pose potential risks, particularly for drivers. Studies have shown that billboards can become a source of distraction, increasing the probability of accidents. Therefore, understanding the interaction between billboards and drivers' behavior is critical.

Recent advancements in computer vision offer various opportunities to analyze the impact of billboards. However, despite the growing interest in understanding the effects of roadside billboards, current methods face different limitations. Traditional object detection approaches often struggle to achieve high accuracy in complex urban environments, where billboards appear in different sizes, shapes, and placements. Furthermore, while some studies on driver attention exist, there is limited research linking billboard detection with gaze duration classification, leading to a significant gap in understanding how billboard visibility is related to driver distraction.

This work aims to contribute to road safety by developing a robust pipeline that will consist of a YOLO-based object detector (trained and fine-tuned with the Mapillary Vistas and BillboardLamac datasets) and a classification system for driver gaze toward billboards, using the BillboardLamac dataset to categorize gaze into predefined classes. With this pipeline, we intend to investigate the connection between billboard detection results and driver attention metrics, providing insights into road safety and advertisement effectiveness.

The structure of this thesis is as follows: The first chapter will review previous works that address challenges related to our study. The second chapter will outline the methodology and implementation of the pipeline, detailing the training of the object detector and classification systems. Chapter 4 will present the results obtained after training each system, along with their metrics. Chapter 5 will provide a deep analysis of the results, examining their impact on solving the problem. Finally, the last chapter will summarize our approach, offer suggestions for future work, and provide a conclusion of our experiments.

Chapter 1

Literature Review

Under development, will be introduced later due to the constant adding new literature.

Chapter 2

Data

This chapter introduces the datasets used to train the models in our pipeline. Important insights about these datasets will be provided, along with the distribution of images into their subsets for training, validation, and testing of the object detector, as well as the predefined classes for the classification system. Additionally, details about the annotation and preprocessing of this data will be discussed.

2.1 Data Overview

The development of the pipeline of this thesis makes use of two datasets: the Mapillary Vistas dataset and the BillboardLamac dataset. These datasets were chosen due to their focus on street-level objects and billboards, aligning with the objectives of this study.

The Mapillary Vistas dataset is a large and diverse street-level imagery dataset, which is used for computer vision tasks such as semantic segmentation and object detection. This dataset consists of images from urban and rural areas, captured under different conditions such as lighting, weather, and geographic locations. It includes instance-specific human annotations, which include billboards. This dataset was chosen to train the object detection model since we believe it would be a suitable foundation to detect roadside advertisements in complex environments.

On the other hand, the BillboardLamac dataset focuses exclusively on roadside billboards and driver behavior. This dataset was created in a previous work, and it provides high-resolution images with billboard annotations. A good portion of the dataset is dedicated to the fine-tuning of the object detector for improving its accuracy. Additionally, the dataset includes driver gaze information that is currently being used for developing the classification system that categorizes the drivers' gaze duration into predefined classes.

2.2 Dataset Structure

To train the components of the pipeline effectively, the datasets were used as follows.

2.2.1 Data for Object Detection Task

The Mapillary Vistas Dataset contains a total of 25,000 high-resolution images. Due to GPU and storage constraints, a portion of the original dataset was used to train the object detector. The method of selecting images to create the subset is described in further subsections. The data is organized into training, validation, and testing subsets. Table 2.1 shows the details of the distributions of these subsets..

	Mapillary Vistas (Original)	**Mapillary Vistas Subset**
Training	18000	6000
Validation	2000	1200
Testing	5000	2500
Total	25000	9700

Table 2.1: Comparison of the full Mapillary Vistas dataset and its subset used for training our object detector.

After training the object detector with the Mapillary Vistas subset, the BillboardLamac Dataset was used to fine-tune the results. This dataset contains a specific subset of images exclusively for object detection. The details of this subset can be found in Table 2.

	BillboardLamac Dataset
Training	795
Validation	199
Testing	219
Total	1213

Table 2.2: BillboardLamac Dataset subsets for fine-tuning the object detector.

2.2.2 Data for Classification

The classification task is currently under development. This section uses a subset of the BillboardLamac dataset for classification, divided into four defined gaze duration classes. Each class contains a set of images, with each image assigned a unique ID. This ID is linked to additional information about the billboards in each image, such as their locations, sizes, and the images themselves. Various strategies are being employed to process this information effectively for training the classification system; details about these experiments will be added once they are concluded. The classification subset is detailed in Table 2.3.

Pre-Defined Classes	**BillboardLamac Dataset (Classification)**
Long	7
Medium	23
None	45
Short	70
Total	145

Table 2.3: BillboardLamac Dataset subset of images containing raw information.

2.3 Data Annotation

Annotations are a crucial part of the data needed to train a robust model. The Mapillary Vistas dataset includes high-quality, manually annotated images with detailed semantic segmentation masks. This annotation process involved using polygons to delineate individual objects. With this information, it was possible to obtain the bounding boxes for each billboard whenever it appeared in an image.

However, the original Mapillary Vistas dataset only included annotations for the training and validation data; there were no annotations for the testing subset. It was necessary to annotate this subset to properly evaluate the results after training the models. To annotate the images, both RoboFlow and Segment Anything tools were used, allowing for careful manual annotation of each billboard through their interfaces.

The BillboardLamac Dataset already included annotations from previous work when this dataset was created. The annotations in this dataset are provided in the form of bounding boxes under the YOLO format. This format will be discussed with more details in the next section.

2.4 Data Preprocessing

Preprocessing was performed to properly prepare the data for the YOLO-based object detector and the gaze classification model.

2.4.1 Data Preparation for Object Detection Task

As introduced earlier, the Mapillary Vistas Dataset includes annotations for its training and validation subsets. To create our training subset from the original dataset, we first selected 5000 images that contained billboards and added 1000 images that did not contain any billboards. Similarly, for the validation subset, we selected 1000 images with billboards and 200 images without them. Lastly, for the testing set, 2500 random images were chosen since this subset's goal is to test the object detector model.

To train the object detector using the YOLO pre-trained models, images with their respective labels are needed. These labels must be in a specific format, referred to as YOLO format for the purposes of this work. The YOLO format must consist of five numbers: *<class_id>* *<x_center>* *<y_center>* *<width>* *<height>*. The *<class_id>* value refers to the ID of the object class; in our case, we assigned the class "0" to billboards since the only task of our object detector is to detect billboards. The *<x_center>* and *<y_center>* values are the normalized x and y-coordinates of the center of the bounding box, respectively; they can be calculated as:

$$x_{\text{center}} = \frac{x_{\min} + x_{\max}}{2 \cdot \text{image\_width}} \tag{2.1}$$

$$y_{\text{center}} = \frac{y_{\min} + y_{\max}}{2 \cdot \text{image\_height}} \tag{2.2}$$

Similarly, *<width>* corresponds to the normalized width of the bounding box relative to the image width, and *<height>* is the normalized height of the bounding box relative to the image height. These values can be calculated as follows:

$$\text{width} = \frac{x_{\max} - x_{\min}}{\text{image\_width}} \tag{2.3}$$

$$\text{height} = \frac{y_{\max} - y_{\min}}{\text{image\_height}} \tag{2.4}$$

The Mapillary Vistas Dataset does not explicitly include these labels. We created a script that processes the segmentation label information and transforms it into YOLO format for each billboard whenever an image contains one. The testing subset did not contain this segmentation label information. To obtain the label-bounding boxes of the test images, the *RoboFlow* Annotate tool was primarily used, along with *Segment Anything* to support challenging scenarios where billboards were difficult to select. The annotation was manually performed by us using the interfaces of the mentioned tools. The advantage of using these tools is that we could export the results directly in YOLO format.

2.4.2 Data Preparation for Classification Task

This specific part of the thesis is under development. Currently, we are attempting to extract as many billboard images as possible using the BillboardLamac classification subset of images. These images will be classified according to the pre-defined classes introduced earlier, and the system will be trained using both a pre-trained YOLO classification model and RESNET. We will then compare both models to determine which one achieves better performance.

2.5 Challenges and Limitations

After preprocessing the Mapillary Vistas dataset, we faced some challenges that are worth mentioning. For instance, we created a script that displays images from the dataset and draws their respective bounding boxes in case billboards are present. We noticed that in some images, billboards were not selected, while in others, letters representing the names of buildings were incorrectly classified as billboards. This can be seen in more detail in figure 2.1.

This misclassification might be due to the fact that the authors and those responsible for annotating the original dataset did not explicitly focus on billboards, even though the "billboard" class is part of the original set of classes in the dataset. We believe this could become a limitation when training the object detector and, eventually, impact its accuracy. This will be discussed again in the results section.

(a) Billboard not selected (b) Miscslassification example

Figure 2.1: Examples of a billboard not being classified (left) and a misclassification case in the training set of the Mapillary Vistas dataset.

Chapter 3

Methodology

The current thesis is developed in a two-stage pipeline: a YOLO-based object detection model capable of detecting roadside billboards and a gaze duration classification system that will categorize driver attention. The object detector was trained using a specific subset of the Mapillary Vistas dataset and fine-tuned with the BillboardLamac dataset for improved accuracy. The image classifier utilizes a different subset of the BillboardLamac dataset to predict drivers' gaze duration into predefined classes: Long, Medium, None, and Short.

3.1 Object Detector

The object detector training is divided into two parts: base training and fine-tuning. We refer to the initial training of the model on the largest dataset as "base training" and the subsequent training on the smaller dataset as "fine-tuning." Datasets are neither merged nor combined in any way. All training experiments began with a pre-trained YOLO model. This pre-trained YOLO model (medium or large size) was initially trained exclusively on the Mapillary Vistas subset using the default hyperparameters, after which different hyperparameters were adjusted to observe their impact on the results. Both YOLOv8 and YOLO11 were the pre-trained models selected for these experiments. The first was chosen for its robustness in object detection tasks, while the second was selected because it is the most recent model in the pre-trained series; thus, both were chosen to determine whether they generate different results after training. Table 3.1 displays the most important default hyperparameters that were later altered for experimentation.

Once a base training was complete, the trained model was tested on the Mapillary Vistas testing subset. From all base training experiments, two of the best models were selected: one based on YOLOv8 and one based on YOLO11. To choose these models, five metrics were taken into account: Precision, Recall, mAP@50, mAP@50-95, and

Fitness. However, since our research focuses on evaluating detection quality and its connection to driver attention, mAP@50 and Recall were the most relevant metrics for determining which model performed the best.

Hyperparameter	**Default Value**
Training Epochs	50
Batch Size	16
Image Size	640
Optimizer	Stochastic Gradient Descent (SGD)
Cosine Learning Rate Scheduler	False
Initial Learning Rate	0.01
Final Learning Rate	0.01
Momentum	0.937
Weight Decay	0.00005
Data Augmentation: Rotation	0.0
Data Augmentation: Scaling	0.5
Data Augmentation: Flipping Upside Down	0.0
Data Augmentation: Flipping Left to Right	0.5

Table 3.1: YOLO pre-trained models default hyperparameters that were changed through different training experiments.

After selecting the two best base models, they were fine-tuned using the BillboardLamac Dataset. Similarly, after experimenting with different combinations of hyperparameters and testing each model on the BillboardLamac Dataset testing subset, two final models were selected based on their metric results.

3.2 Classifier

This part is currently under development.

The classification task involves training a classifier that can correctly classify driver gaze durations into four classes based on billboard features from the BillboardLamac classification images dataset. This aims to evaluate how long drivers focus on billboards and understand its potential impact on distraction. We plan to conduct training experiments using two pretrained models: YOLO-classification and RESNET50. The two best models (one from each pretrained model) will be selected based on their metric results. For this classification task, the metrics to consider will be Accuracy, Precision, Recall, and F1-score.

3.3 Pipeline Integration

This section will be completed once the classification system training is finished.

The final pipeline will integrate the object detection model with the classification model. Detected billboards from YOLO will be input into the gaze classifier, which will predict the duration of driver attention. Results will be stored for analysis of correlations between billboard visibility and distraction.

3.4 Resources

Initially, free resources were used to train our first models. We alternated between Google Colab's free GPU resources and Google Cloud Computing Services. The former offered daily free GPU usage for up to 3 hours in total, while Google Cloud services operated based on credits. Although these training methods served well as an introduction to our current work, they were very limited in terms of usage time and availability. As a result, we performed most of our training using the Department of Applied Informatics' GPU Computing Servers.

We started our experiments on the Uran server; however, after the addition of the Jupiter server, we switched to the latter due to its larger storage and increased RAM capabilities. Training experiments were conducted on 2 NVIDIA GeForce RTX 4080 Super GPUs, each with 16 GB of RAM.

All created scripts for data pre-processing, training, and results visualization are Python-based, using frameworks and libraries such as PyTorch, OpenCV, NumPy, and Matplotlib, among others. For image annotation, RoboFlow and Segment Anything online interface applications were used.

Chapter 4

Results

This chapter presents the results obtained after training the object detector and gaze classification models. The first section describes the evaluation process of the different YOLO-based detection models after completing their respective base training and fine-tuning. We will also present the metrics obtained and the criteria used to select the two best base training models, including visualizations of performance metrics. The second section will focus on displaying the results of the driver gaze classification system (currently under development).

4.1 Object Detection Results

A total of 30 base trainings were performed. Sixteen training models used the medium-sized YOLOv8 pre-trained model, twelve used the large-sized YOLOv8 pre-trained model, and two trainings were conducted using the large-sized YOLO11 pre-trained model for comparison purposes. Since the YOLO11 pre-trained model series is an update of the YOLOv8 pre-trained model family, two experiments were conducted using only the best combination of hyperparameters found from the YOLOv8 base trainings. The detailed results, including the most important hyperparameters changed during the base training process and the main metric results, can be found in Tables A.1–6 in Appendix A.

Every base training varied in at least one hyperparameter. After each training process was completed, the resulting model was tested on the Mapillary Vistas testing subset, and the obtained metric values were saved. At the end, the 30 results were plotted against the training experiments (training ID). Figures 4.1, 4.2, and 4.3 show how some metrics (Precision, Recall, and mAP@50) varied across the 30 different training experiments. The best two models were selected after considering a balance between these results, with priority given to the mAP@50 and Recall metrics.

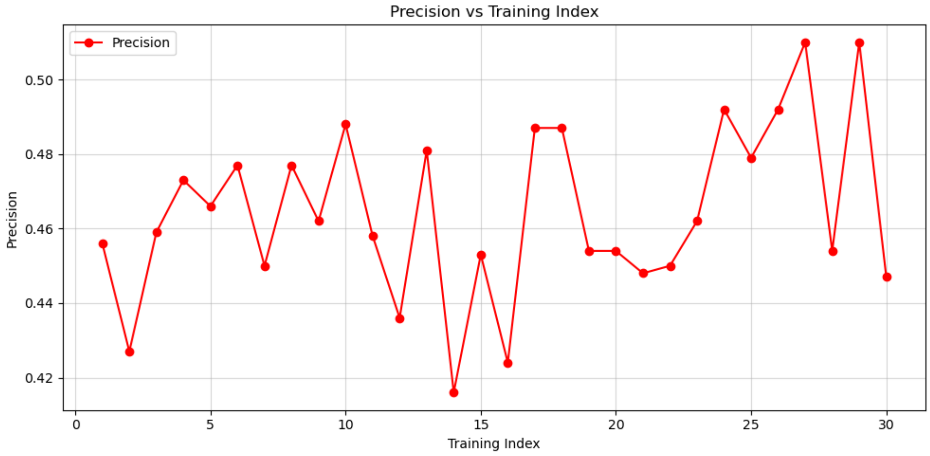

Figure 4.1: Base Training Precision vs Training Index

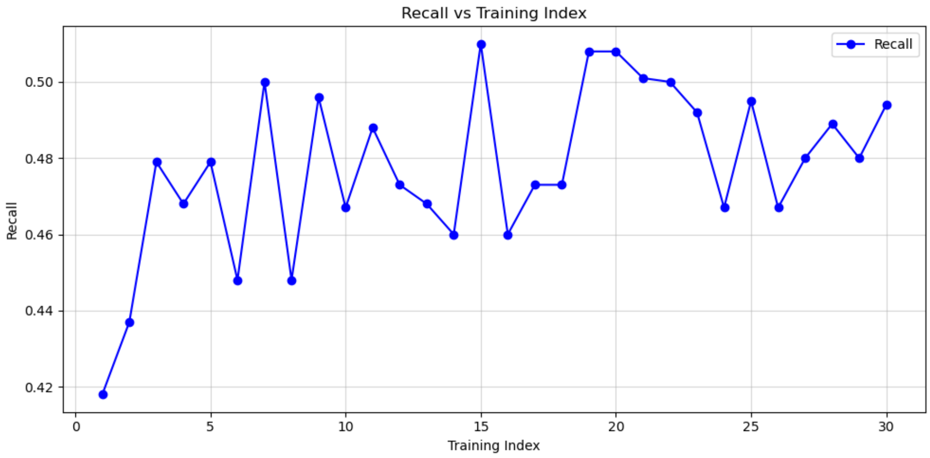

Figure 4.2: Base Training Recall vs Training Index

Figure 4.3: Base Training map@50 vs Training Index

After careful consideration, experiments 19 and 27 (from the YOLOv8 and YOLO11 pre-trained models, respectively) were chosen based on their performance metrics. The details of their metric results can be found in Table 4.1.

Metrics	Experiment 19 (v8)	Experiment 27 (11)
Precision	0.454	0.510
Recall	0.508	0.480
mAP@50	0.455	0.467
mAP@50-95	0.323	0.347
Fitness	0.337	0.347

Table 4.1: YOLOv8 and YOLO11 best base training experiments metric results.

These two models were then fine-tuned using the BillboardLamac Dataset. A total of 20 fine-tuning experiments were conducted, with 10 per model. Once each fine-tuning experiment was complete, the resulting model was tested on the testing subset of the BillboardLamac Dataset. The same considerations as for the base training were taken into account, and the best two models were selected based on their metric performances. The complete fine-tuning results can be found in Tables A.7–10 in Appendix A.

Images 4.4, 4.5, and 4.6 show how metrics varied across the experiments, and Table 4.2 shows the details of the metrics for the 7th and 20th experiments (YOLOv8 and YOLO11 originally-based models), which were the best models identified after reviewing their results.

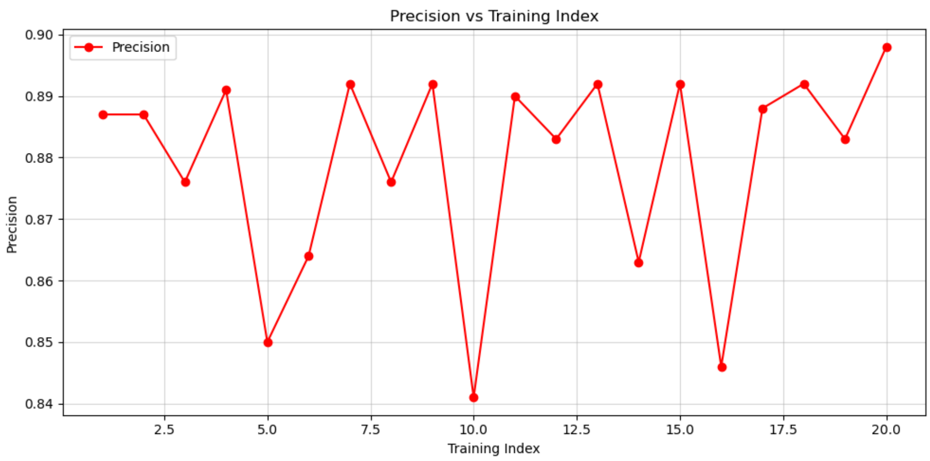

Figure 4.4: Precision vs Training Index

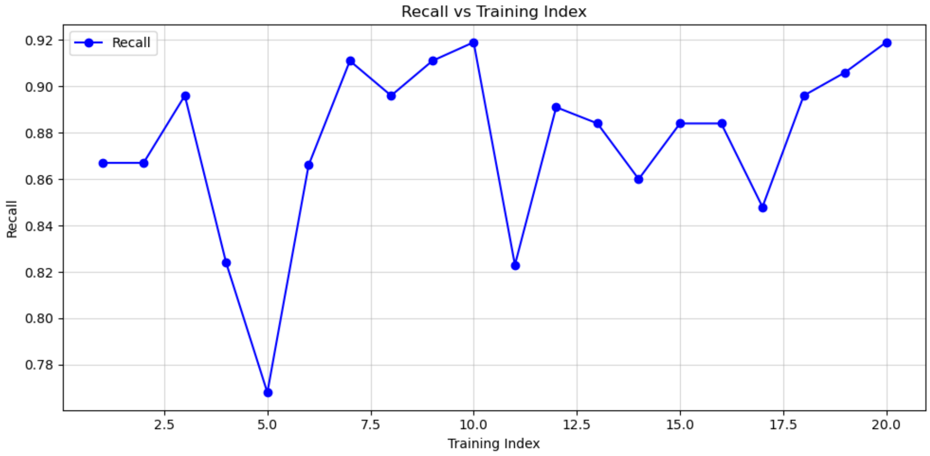

Figure 4.5: Recall vs Training Index

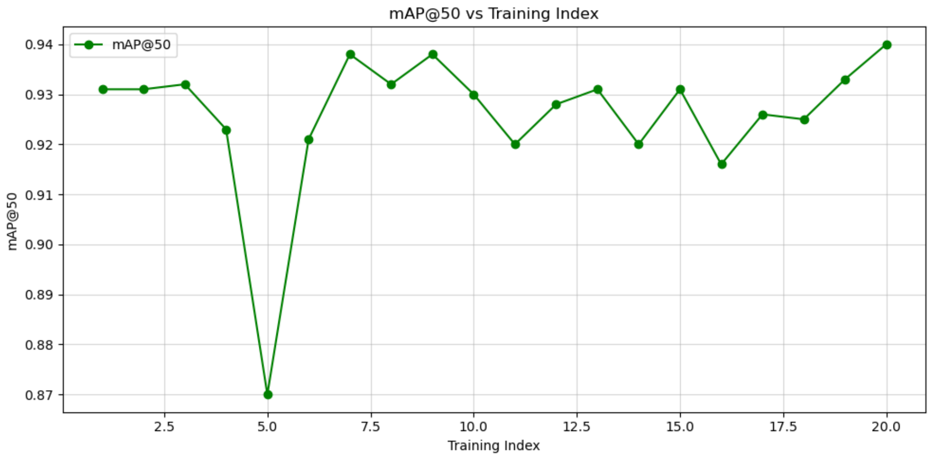

Figure 4.6: map@50 vs Training Index

Metrics	**Experiment 7 (v8)**	**Experiment 20 (11)**
Precision	0.892	0.898
Recall	0.911	0.919
mAP@50	0.938	0.940
mAP@50-95	0.771	0.794
Fitness	0.788	0.809

Table 4.2: YOLOv8 and YOLO11 best fine-tuning experiments metric results.

4.2 Classification Results

Under development.

Chapter 5

Discussion and Future Work

Under development.

Conclusion

Under development.

Bibliography

[1] David Crundall, Editha Van Loon, and Geoffrey Underwood. Attraction and distraction of attention with roadside advertisements. *Accident Analysis and Prevention*, 38(4):671–677, July 2006.

[2] Vasiliki Kondyli, Mehul Bhatt, Daniel Levin, and Jakob Suchan. How do drivers mitigate the effects of naturalistic visual complexity? *Cognitive Research: Principles and Implications*, 8(1):54, Aug 2023.

[3] Zuzana Berger Haladova, Michal Zrubec, and Zuzana Cernekova. A method for estimating roadway billboard salience. In *ACM Symposium on Applied Perception 2024*, SAP '24, New York, NY, USA, 2024. Association for Computing Machinery.

[4] Gerhard Neuhold, Tobias Ollmann, Samuel Rota Bulò, and Peter Kontschieder. The mapillary vistas dataset for semantic understanding of street scenes. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 5000–5009, 2017.

[5] Zuzana Černeková, Zuzana Berger Haladová, Ján Špirka, and Viktor Kocur. Evaluating the significance of outdoor advertising from driver's perspective using computer vision. *2023 World Symposium on Digital Intelligence for Systems and Machines (DISA)*, pages 53–58, November 2023.

[6] Sukriti Dhang, Atseosi Idogho, Mimi Zhang, and Soumyabrata Dev. Towards accurate billboard detection: an ablation and benchmarking study of deep learning models. *IET Conference Proceedings*, 2024:178–185, 2024.

[7] Ángel Morera, Ángel Sánchez, A. Belén Moreno, Ángel D. Sappa, and José F. Vélez. Ssd vs. yolo for detection of outdoor urban advertising panels under multiple variabilities. *Sensors*, 20(16), 2020.

[8] Sayali Chavan, Dermot Kerr, Sonya Coleman, and Hussein Khader. Billboard detection in the wild. In *Irish Machine Vision and Image Processing Conference Proceedings 2021 - DCU, Ireland*, pages 57–64, Ireland, September 2021. Irish Pattern Recognition and Classification Society. Irish Machine Vision and Image

Processing Conference 2021, IMVIP2021 ; Conference date: 01-09-2021 Through 03-09-2021.

[9] Sayali Chavan, Dermot Kerr, Sonya Coleman, and Hussein Khader. Deep learning for billboard classification. In Petre Dini, editor, *DIGITAL 2023*, pages 29–35, United States, October 2023. International Academy, Research, and Industry Association. Advances on Societal Digital Transformation DIGITAL 2023, DIGITAL 2023 ; Conference date: 25-09-2023 Through 29-09-2023.

Appendix A

Training Results

List of Abbreviations Used in this Appendix

- **LRS**: Learning Rate Scheduler.
- **WD**: Weight Decay.
- **CA**: Cosine Annealing.
- **AW**: AdamW Optimizer.
- **DO**: Dropout ($p = 0.2$).
- **DA1**: Data Augmentation 1 - (Flip Upside Down, Flip Left to Right, Scaling).
- **DA2**: Data Augmentation 2 - (DA1 + Hue, Saturation, Value).
- **X**: Extra Large Model.
- **M**: Momentum.
- **GA**: Gradient Augmentation.
- **ESA**: Early Stopping Activated.

Training ID	**Epochs**	**Batch**	**Image Size**	**Learning Rate**	**Extra Notes**
1	50	16	640	0.01	-
2	50	8	640	0.01	-
3	50	8	1024	0.01	-
4	50	8	1280	0.01	-
5	50	4	1280	0.01	-
6	100	8	1280	0.01	-
7	100	4	1280	0.01	-
8	100	8	1280	0.01	LRS
9	100	4	1280	0.01	WD
10	100	4	1280	0.01	CA
11	150	4	1280	0.01	WD+CA
12	150	8	1280	0.01	-
13	100	4	1280	0.01	WD+CA+AW
14	100	4	1280	0.01	AW
22	100	4	1280	0.0025	-
24	100	6	1280	0.0025	-

Table A.1: YOLOv8 medium size pre-trained model base training hyperparameters.

Training ID	Precision	Recall	mAP@50	mAP@50-95	Fitness	Training Time
1	0.456	0.418	0.381	0.245	0.259	45’
2	0.427	0.437	0.376	0.247	0.260	45’
3	0.459	0.479	0.435	0.301	0.315	1h 20’
4	0.473	0.468	0.432	0.300	0.314	2h 50’
5	0.466	0.479	0.447	0.311	0.325	2h 49’
6	0.477	0.448	0.426	0.300	0.313	5h 40’
7	0.450	0.500	0.448	0.312	0.325	5h 45’
8	0.477	0.448	0.426	0.300	0.313	5h 40’
9	0.462	0.496	0.442	0.306	0.319	5h 41’
10	0.488	0.467	0.440	0.307	0.320	5h 45’
11	0.458	0.488	0.428	0.297	0.310	8h 55’
12	0.436	0.473	0.409	0.285	0.297	9h 1’
13	0.481	0.468	0.440	0.306	0.320	5h 20’
14	0.416	0.460	0.386	0.262	0.275	5h 44’
22	0.450	0.500	0.448	0.312	0.325	6h 40’
24	0.492	0.467	0.452	0.314	0.328	6h 5’

Table A.2: YOLOv8 medium size pre-trained model base training results.

Training ID	Epochs	Batch	Image Size	Learning Rate	Extra Notes
15	100	4	1280	0.01	-
16	100	4	1280	0.01	WD+CA+AW
17	150	4	1280	0.01	-
18	150	4	1280	0.01	DO
19	100	4	1280	0.01	DA1
20	100	4	1280	0.01	DA2
21	100	4	1280	0.01	X
23	100	6	1280	0.01	DA1
25	150	6	1280	0.01	DA1
26	150	6	1280	0.00375	-
28	100	6	1280	0.00375	-
30	150	6	1280	0.00375	DA1

Table A.3: YOLOv8 large size pre-trained model base training hyperparameters.

Training ID	**Precision**	**Recall**	**mAP@50**	**mAP@50-95**	**Fitness**	**Training Time**
15	0.453	0.510	0.438	0.302	0.316	9 h 29’
16	0.424	0.460	0.391	0.266	0.279	9h 31’
17	0.487	0.473	0.443	0.310	0.323	13h 1’
18	0.487	0.473	0.443	0.310	0.323	13h 1’
19	0.454	0.508	0.455	0.323	0.337	8h 40’
20	0.454	0.508	0.455	0.323	0.337	8h 45’
21	0.448	0.501	0.449	0.317	0.330	13h 50’
23	0.462	0.492	0.449	0.316	0.329	9h 25’
25	0.479	0.495	0.458	0.323	0.336	14h 10’
26	0.492	0.467	0.452	0.314	0.328	6h 48’
28	0.454	0.489	0.438	0.304	0.318	9h 45’
30	0.447	0.494	0.439	0.306	0.319	13h 55’

Table A.4: YOLOv8 large size pre-trained model base training results.

Training ID	**Epochs**	**Batch**	**Image Size**	**Learning Rate**	**Extra Notes**
27	150	6	1280	0.01	DA1
29	150	6	1280	0.00375	DA1

Table A.5: YOLO11 large size pre-trained model base training hyperparameters.

Training ID	**Precision**	**Recall**	**mAP@50**	**mAP@50-95**	**Fitness**	**Training Time**
27	0.510	0.480	0.467	0.334	0.347	13h 1’
29	0.510	0.480	0.467	0.334	0.347	13h’ 14’

Table A.6: YOLO11 large size pre-trained model base training results.

Training ID	Epochs	Batch	Image Size	Learning Rate	Extra Notes
1	50	4	1280	0.01	-
2	50	4	1280	0.01	WD
3	100	4	1280	0.01	-
4	100	4	1280	0.01	DA1
5	150	4	1280	0.01	-
6	150	4	1280	0.01	WD+M
7	100	6	1280	0.01	-
8	100	4	1280	0.0025	-
9	100	6	1280	0.00375	-
10	100	4	1280	0.01	GA

Table A.7: YOLOv8-based fine-tuning hyperparameters.

Training ID	Precision	Recall	mAP@50	mAP@50-95	Fitness	Training Time
1	0.887	0.867	0.931	0.725	0.746	37’
2	0.887	0.867	0.931	0.725	0.746	37’
3	0.876	0.896	0.932	0.727	0.748	1h 2’
4	0.891	0.824	0.923	0.634	0.663	1h 1’
5	0.850	0.768	0.870	0.639	0.662	1h 20’
6	0.864	0.866	0.921	0.648	0.675	1h 47’
7	0.892	0.911	0.938	0.771	0.788	1h 13’
8	0.876	0.896	0.932	0.727	0.748	1h 11’
9	0.892	0.911	0.938	0.771	0.788	1h 12’
10	0.841	0.919	0.930	0.768	0.784	1h 2’

Table A.8: YOLOv8-based fine-tuning results.

Training ID	**Epochs**	**Batch**	**Image Size**	**Learning Rate**	**Extra Notes**
11	50	4	1280	0.01	-
12	100	4	1280	0.01	-
13	150	4	1280	0.01	-
14	200	4	1280	0.01	ESA
15	150	4	1280	0.0025	-
16	50	6	1280	0.01	-
17	100	6	1280	0.01	-
18	150	6	1280	0.01	-
19	200	6	1280	0.01	-
20	500	6	1280	0.01	ESA

Table A.9: YOLO11-based fine-tuning hyperparameters.

Training ID	**Precision**	**Recall**	**mAP@50**	**mAP@50-95**	**Fitness**	**Training Time**
11	0.890	0.823	0.920	0.702	0.724	37’
12	0.883	0.891	0.928	0.740	0.759	1h 10’
13	0.892	0.884	0.931	0.770	0.786	1h 35’
14	0.863	0.860	0.920	0.715	0.735	1h 10’
15	0.892	0.884	0.931	0.770	0.786	1h 40’
16	0.846	0.884	0.916	0.710	0.731	36’
17	0.888	0.848	0.926	0.720	0.741	1h 11’
18	0.892	0.896	0.925	0.744	0.762	1h 41’
19	0.883	0.906	0.933	0.781	0.796	2h 27’
20	0.898	0.919	0.940	0.794	0.809	4h 25’

Table A.10: YOLO11-based fine-tuning results.

Appendix B

Myproject source code