A method for estimating roadway billboard salience

Zuzana Berger Haladova
zuzana.bergerhaladova@fmph.uniba.sk
Faculty of Mathematics Physics and Informatics, Comenius University
Bratislava, Slovakia

Michal Zrubec
m.zrubec60@gmail.com
Faculty of Mathematics Physics and Informatics, Comenius University
Bratislava, Slovakia

Zuzana Cernekova
cernekova@fmph.uniba.sk
Faculty of Mathematics Physics and Informatics, Comenius University
Bratislava, Slovakia

ABSTRACT

Roadside billboards and other forms of outdoor advertising play a crucial role in marketing initiatives; however, they can also distract drivers, potentially contributing to accidents. This study delves into the significance of roadside advertising in images captured from a driver's perspective. Firstly, it evaluates the effectiveness of neural networks in detecting advertising along roads, focusing on the YOLOv5 and Faster R-CNN models. Secondly, the study addresses the determination of billboard significance using methods for saliency extraction. The UniSal and SpectralResidual methods were employed to create saliency maps for each image. The study establishes a database of eye tracking sessions captured during city highway driving to assess the saliency models.

CCS CONCEPTS

• **Computing methodologies → Object detection**; **Interest point and salient region detections**.

KEYWORDS

Eyetracking, Saliency, Visual advertising, Object detection

ACM Reference Format:
Zuzana Berger Haladova, Michal Zrubec, and Zuzana Cernekova. 2024. A method for estimating roadway billboard salience. In *ACM Symposium on Applied Perception 2024 (SAP '24), August 30–31, 2024, Dublin, Ireland*. ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3675231.3675238

1 INTRODUCTION

Outdoor advertising, particularly through roadside billboards, plays an integral role in contemporary marketing strategies, serving as a prominent medium for brand promotion. However, the widespread presence of these advertisements along roadways raises concerns about potential distractions for drivers, which could lead to safety hazards and accidents. Recognizing the dual impact of outdoor advertising on marketing effectiveness and road safety, this study investigates the nuanced significance of roadside advertising from a driver's perspective.

This study leverages the capabilities of neural networks. The YOLOv5 and Faster R-CNN models are evaluated for their effectiveness in detecting advertising spaces along roads. Furthermore, the study delves into the determination of billboard significance through saliency extraction methods. By employing the UniSal and SpectralResidual techniques, the research aims to create saliency maps for each image, providing insights into the key features that capture a driver's attention.

To ground the findings in real-world driving scenarios, the database of eye-tracking sessions conducted during city highway driving was created within the study. By utilizing eye-tracking data, the research seeks to evaluate the effectiveness of the saliency models in capturing the attention of drivers.

The paper is organized as follows. The second section provides an overview of the previous works in the area of billboard/object detection and saliency extraction as well as the research on the relevance and viewability of billboards. In the third section datasets used for training and evaluation of the proposed methods are discussed. In the fourth section, the proposed method is described. In the fifth section, the results of the acquired experiments are presented. Then the Conclusions are given.

2 RELATED WORK

This section delves into previous research on the relevance and viewability of billboards, billboard detection, and saliency extraction methods. In the area of billboard relevance, several factors such as gaze duration, placement, and content impact on driver attention and road safety were examined. The placement of advertising significantly affects its reach. Advertisements on the driver's side of the road attract more fixations than those on the opposite side [Costa et al. 2019]. Contrary to the assumption that the height of the advertisement placement prolongs fixation duration, studies reveal that drivers look more at advertisements at road level than those elevated three or more meters above the ground, except when drivers intentionally seek out higher-placed advertisements. Additionally, factors such as the angle of the advertising surface, the complexity of the traffic section, and the environment influence the visibility of advertisements [Costa et al. 2019; Crundall et al. 2006; Mollu et al. 2018; Zalesinska 2018].

Content is another crucial characteristic. Research shows that advertisements with longer text, a sexual undertone, or featuring human beings result in longer fixations [Harasimczuk et al. 2021; Maliszewski et al. 2019; Meuleners et al. 2020; Tarnowski et al. 2017]. Moreover, more complex advertisements tend to capture the driver's gaze for a longer duration [Marciano et al. 2017]. Advertisements eliciting negative emotions in drivers also lead to longer fixations and reduced speed, with negative content being more memorable according to post-drive questionnaires filled out by drivers [Chan and Singhal 2013]. While our study focuses solely on the saliency of billboards, it can be extended in the future to analyze other

SAP '24, August 30–31, 2024, Dublin, Ireland

ACM ISBN 979-8-4007-1061-2/24/08
https://doi.org/10.1145/3675231.3675238

important aspects that increase the duration of fixations among drivers.

2.1 Billboard detection

Since billboard detection is a special case of object detection, we will discuss it further in the following section. Convolutional Neural Networks (CNNs) for object detection typically use a universal backbone for extracting image features, paired with a framework to recognize object classes and generate bounding boxes. Two primary categories of object detectors have emerged: single-stage and two-stage detectors.

In two-stage detectors, a network first proposes objects, which are then verified in a second stage. The initial bounding boxes are refined in this second stage for greater precision. Faster R-CNN is a notable example of a two-stage detector, known for its accuracy [Ren et al. 2015]. Its extension, Mask R-CNN, also detects object masks [He et al. 2017].

Single-stage detectors refine bounding boxes in one pass through the network. This is exemplified by anchor-based methods like RetinaNet [Lin et al. 2017] and anchor-free approaches like YOLO [Redmon et al. 2016] and FCOS [Tian et al. 2019]. Keypoint-based methods detect key points at bounding box corners [Law and Deng 2018] or centers [Zhou et al. 2019], with regression outputs to define the boxes. ATSS offers a unique single-stage approach with adaptive training sample selection [Zhang et al. 2020].

There were also several studies dealing with the CNN methods for outdoor advertising detection, usually relying on transfer learning [Chavan et al. 2021; Hossari et al. 2018; Liu et al. 2018; Rahmat et al. 2019]. For instance, one approach [Rahmat et al. 2019] used AlexNet's Deep Convolutional Neural Network (DCNN) with inductive transfer learning, achieving high training accuracy for advertisement billboard detection. Other methods, such as the attention-based multi-scale feature fusion region proposal network (AM-RPN) based on Faster R-CNN [Liu et al. 2018], incorporated advanced techniques like adaptive non-maximum suppression to enhance detection performance in challenging conditions.

2.2 Saliency extraction

Much of the existing literature on visual saliency modeling focuses on predicting human visual attention mechanisms in static scenes. First models [Hou and Zhang 2007; Itti et al. 1998; Judd et al. 2009; Sun and Fisher 2003] primarily concentrate on low-level image features, such as intensity, contrast, color, and edges, referred to as bottom-up methods. One of them the Spectral Residue [Hou and Zhang 2007] is considered fast and robust. Recent advancements leverage deep neural networks, starting with Vig et al. [Vig et al. 2014]. Jiang et al. [Jiang et al. 2015] collected SALICON, a large-scale saliency dataset, facilitating exploration in deep learning-based saliency modeling. Other studies center around network architecture design, and varying model sizes. In [Droste et al. 2020] Droste et al. proposed UNISAL saliency extraction method achieves state-of-the-art performance on all video saliency datasets and was on par with the state-of-the-art for image saliency datasets, despite faster runtime and a 5 to 20-fold smaller model size compared to all competing deep methods. For a thorough review of saliency extraction methods see [Abraham and Kovoor 2023; Ullah et al. 2020].

Figure 1: Image with annotated billboards from our dataset

3 DATASET

This section discusses datasets used and created within this work.

In our solution, we use two datasets. The first dataset is the existing Mapillary Vistas Dataset [Neuhold et al. 2017], which is ideal due to its extensive street scenes and diverse object annotations. It contains 25,000 images with annotations for training and validation sets, divided into 18,000 training images, 2,000 validation images, and 5,000 test images.

The second dataset is custom, created using car dashboard imagery and an eyetracker. For eyetracking, we used the Tobii Pro Glasses 3. The roads were chosen to represent various driving scenarios, including highways and urban streets, aiming to cover a wide range of real-world conditions. This dataset has 1580 images, standardized to 1920×1080 pixels. Figure 1 shows an example.

Creating this dataset was challenging and time-consuming, requiring many images. All images were manually annotated, with each image having one associated file. We used Roboflow to label the billboards, focusing on large advertisements and excluding smaller ads, such as store names and distant billboards. This dataset will be available online (link omitted due to double-blind review). It was created with images from 3 drivers (2 males, and 1 female, ages 22-26) and divided into training (1350 images), validation (150 images), and test (80 images) sets for model training and evaluation.

4 PROPOSED METHOD

To address the challenge of determining the saliency of advertisements, we propose the following procedure.

(1) detection of the advertising areas in the input image,
(2) creating a saliency map for the input image,
(3) obtaining the average saliency of the advertisement areas,
(4) checking whether the average significance value of the ad space is large enough,
(5) classifying the saliency of the advertising spaces.

A scheme of the process can be seen in Figure 2 and a detailed explanation of the steps is provided in the subsequent text.

To detect billboard areas, we opted to test the YOLO[Redmon et al. 2016] and Faster R-CNN models[Ren et al. 2015]. These models achieve very good results in terms of speed while maintaining sufficient detection accuracy. During the evaluation, we use models

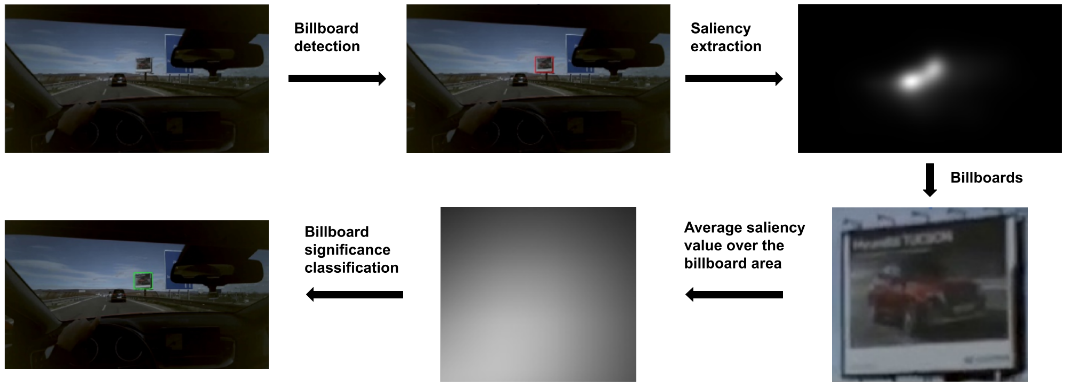

Figure 2: The scheme of the process of classification of the significance of billboards on the images

pretrained on the MS COCO dataset [Lin et al. 2014]. We evaluate the models and use the better model to detect the billboard areas for the following saliency detection.

4.1 Saliency extraction methods

After identifying regions of interest using object detection models, the next step involves generating a saliency map from the original image. To achieve this, we explored two distinct methods designed for saliency extraction: spectral residue[Hou and Zhang 2007] and UniSal [Droste et al. 2020]. The **Spectral Residue** method operates by breaking down the image into its phase and amplitude components using the Fourier transform. By smoothing the amplitude component, we emphasize outliers, which are areas significantly different from the average. Subtracting this smoothed amplitude from the original component reveals regions that stand out the most, highlighting salient areas effectively [Hou and Zhang 2007]. In contrast, the **UniSal** model [Droste et al. 2020] offers a more comprehensive approach. It is specifically engineered to handle saliency modeling in both images and videos. The architecture utilizes an encoder - RNN - decoder design tailored for salience modeling. The model first encodes the input image using MobileNet-V2 [Sandler et al. 2018], capturing its essential features. Then, a convolutional Gated Recurrent Unit (GRU) RNN processes this information to capture temporal dependencies, if applicable. This sophisticated design allows UniSal to effectively identify salient regions in diverse visual content

4.2 Billboard Significance Assessment

After obtaining the saliency maps, we determine the significance of billboards by assessing the average saliency value within the bounding box of each billboard area. Using the Spectral residual and UniSal methods, we generate saliency maps, focusing specifically on regions corresponding to advertising areas. In these areas, we normalized the saliency values of all pixels to an interval from 0 to 1. The average saliency value in each area is obtained by summing and dividing these normalized values by the number of pixels in the area. Similarly, we compute the average saliency value for all advertising areas using the same methodology. To classify the significance of an advertisement, we compared the average salience value of the advertising space and the average salience value of all advertising spaces. If the average salience value is greater, the advertising space is classified as salient. Subsequently, we validate the significance of billboards based on eyetracker data. A billboard is classified as significant if corresponding eye fixations are observed within the billboard area in the eyetracker image. This serves as a crucial validation step in confirming the impact and visibility of billboards.

5 RESULTS

The evaluation was performed in two phases. At first, we focused on the evaluation and comparison of the YOLOv5 and Faster R-CNN models, which are used to detect advertising areas. Then, we evaluated the methods for extracting saliency in the image and the method for classifying the significance of advertising spaces.

For testing purposes, we used the test dataset that consists of 80 images containing a total of 210 billboard areas. The dataset contains images with a resolution of 1920×1080, featuring billboard areas from both highways and cities. These images include diverse lighting conditions, weather scenarios, and instances where billboard areas may be occluded with other objects.

For the evaluation of the model, we have employed standard metrics used to determine the accuracy of a detector, the Intersection over Union (IoU). It expresses how much the bounding box of the predicted object overlaps with the actual bounding box of the object. Another widely used metric for evaluating object detectors is average precision (AP). It is used e.g. in COCO challenges. The average precision is evaluated over several IoU values ranging from 0.5 to 0.95 with a step of 0.05.

Table 1: Results of models for detecting billboards after fine-tuning on Mapillary Vistas dataset.

Model	AP@0.5	AP@0.5:0.95
YOLOv5	63.2%	48.1%
Faster R-CNN	55.8%	40.2%

In Table 1 we can see the results of the models after fine-tuning on Mapillary Vistas dataset. We experimented with various combinations of data augmentation techniques, and the most favorable outcomes were achieved when incorporating random rotation (ranging from -10° to 10°), random brightness adjustment (ranging from -10% to 10%), random Gaussian noise insertion, and random clipping.

Next, testing of the models fine-tuned on our dataset was performed. The results are shown in the Table 2.

Table 2: Results of models for detecting billboards after fine-tuning on our dataset.

Model	AP@0.5	AP@0.5:0.95
YOLOv5	96.6%	69.5%
Faster R-CNN	89.6%	58.4%

In Tables 1 and 2, we can notice that significantly better results were obtained on our dataset than on the Mapillary Vistas dataset. The reason for these differences is that in the Mapillary Vistas dataset, the images are of a much higher resolution (at most up to 5248×3936 pixels), where even very small and distant advertising areas are annotated. When training the models for object detection, the size of the images is reduced to a smaller size (1920×1080 pixels in our case) on input, and thus the very small advertising areas annotated on the high-resolution images are "faded out".

5.1 Comparing Saliency Maps to Fixation Maps

We analyzed the saliency maps generated by our models against the ground truth fixation maps gathered using the eyetracker. To evaluate the consistency between the saliency and fixation maps, we utilized two evaluation metrics: Area Under ROC Curve (AUC) and Normalized Scanpath Saliency (NSS) [Riche et al. 2013]. These metrics quantify the agreement between the predicted saliency and actual fixation points, providing complementary insights into the effectiveness of our saliency extraction methods.

The ideal value of the AUC-Judd metric is 1, which means that it represents a 100% detection accuracy. The random saliency map has a resulting value of 0.5, which means that using our methods, we should obtain a value greater than 0.5.

A positive NSS score indicates that the saliency value at a fixation point is above the mean saliency value, while a negative score indicates a saliency value below the mean. The mean used here is the average of all standardized saliency values at fixation points. For instance, an NSS score of 1 means that the saliency values at fixation points were 1 standard deviation above the mean saliency value. Unlike AUC, NSS works with actual saliency values and is more sensitive to false positives, providing a nuanced assessment of how well a saliency model aligns with human eye fixations.

Table 3: Results of saliency models.

Method	AUC	NSS
Unisal	0,926	2,287
SpectralResidual	0,915	2,921

In Table 3 we can see that both used methods performed well and achieved good results on the test dataset.

5.2 Significance of the billboards

In assessing the significance of billboards, we computed the average saliency values for all advertising spaces in the training set. The resulting average significance value, normalized to the interval from 0 to 1, was determined to be 0.416.

Table 4: The billboards' significance results.

Method	Accuracy	Sensitivity
Unisal	82,6%	74,0%
SpectralResidual	69,2%	78,3%

In Table 4, we can observe that our method for detecting the significance of billboard areas achieved an accuracy of 82.6% and a sensitivity of 74.0% when generating saliency maps using the UniSal method. When employing the SpectralResidual method, the accuracy was 69.2%, and the sensitivity was 78.3%. This implies that when using the UniSal method, there is a higher likelihood that predicted positive detections were indeed correct.

6 CONCLUSION AND DISCUSION

In this study, we introduced a method for classifying the significance of billboards. Our evaluation indicates that YOLOv5 outperforms Faster R-CNN in the task of detecting billboard areas, demonstrating higher accuracy and speed. The best results were achieved after fine-tuning both YOLOv5 and Faster R-CNN on our custom dataset. However, Faster R-CNN struggled with small and medium-sized objects even after fine-tuning.

For saliency extraction, we explored the UniSal and SpectralResidual methods, generating saliency maps for each image using ground truth fixation maps from an eyetracker. Both methods yielded satisfactory results, with UniSal showing potential for improvement through training on eyetracker-acquired data in diverse traffic scenarios.

Our proposed method for billboard significance classification was evaluated, revealing better performance on saliency maps generated using the UniSal method. The method has delivered promising and satisfactory results in the classification of billboard significance. The proposed method achieved an accuracy of 82.6% and a sensitivity of 74.0%. This affirms the effectiveness of our approach in addressing the task at hand, marking a significant step forward in the field of billboard detection and saliency analysis. These findings contribute valuable insights to the ongoing development of methods for accurate and efficient billboard classification in various real-world scenarios.

ACKNOWLEDGMENTS

This publication was supported by projects: (ACCORD, ITMS2014+: 313021X329), co-funded by the European Regional Development Fund, the TERAIS project, Horizon-Wider-2021 programme of the European Union under grant agreement number 101079338, and project APVV-23-0250.

REFERENCES

Shilpa Elsa Abraham and Binsu C Kovoor. 2023. Visual Saliency Modeling with Deep Learning: A Comprehensive Review. *Journal of Information & Knowledge Management* 22, 02 (2023), 2250066.

Michelle Chan and Anthony Singhal. 2013. The emotional side of cognitive distraction: Implications for road safety. *Accident Analysis & Prevention* 50 (2013), 147–154.

Miss Sayali Avinash Chavan, Dermot Kerr, Sonya Coleman, and Mr Hussein Khader. 2021. Billboard Detection in the Wild. In *Irish Machine Vision and Image Conference Proceedings 2021*. Irish Pattern Recognition and Classification Society., 57–64.

Marco Costa, Leonardo Bonetti, Valeria Vignali, Arianna Bichicchi, Claudio Lantieri, and Andrea Simone. 2019. Driver's visual attention to different categories of roadside advertising signs. *Applied ergonomics* 78 (2019), 127–136.

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

Richard Droste, Jianbo Jiao, and J Alison Noble. 2020. Unified image and video saliency modeling. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part V 16*. Springer, 419–435.

Justyna Harasimczuk, Norbert E Maliszewski, Anna Olejniczak-Serowiec, and Adam Tarnowski. 2021. Are longer advertising slogans more dangerous? The influence of the length of ad slogans on drivers' attention and motor behavior. *Current Psychology* 40 (2021), 429–441.

Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. 2017. Mask r-cnn. In *Proceedings of the IEEE International Conference on Computer Vision*. IEEE, 2961–2969.

Murhaf Hossari, Soumyabrata Dev, Matthew Nicholson, Killian McCabe, Atul Nautiyal, Clare Conran, Jian Tang, Wei Xu, and François Pitié. 2018. ADNet: A deep network for detecting adverts. *arXiv preprint arXiv:1811.04115* (2018).

Xiaodi Hou and Liqing Zhang. 2007. Saliency detection: A spectral residual approach. In *2007 IEEE Conference on computer vision and pattern recognition*. Ieee, 1–8.

Laurent Itti, Christof Koch, and Ernst Niebur. 1998. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on pattern analysis and machine intelligence* 20, 11 (1998), 1254–1259.

Ming Jiang, Shengsheng Huang, Juanyong Duan, and Qi Zhao. 2015. Salicon: Saliency in context. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 1072–1080.

Tilke Judd, Krista Ehinger, Frédo Durand, and Antonio Torralba. 2009. Learning to predict where humans look. In *2009 IEEE 12th international conference on computer vision*. IEEE, 2106–2113.

Hei Law and Jia Deng. 2018. Cornernet: Detecting objects as paired keypoints. In *Proceedings of the European Conference on Computer Vision*. 734–750.

Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. 2017. Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision*. IEEE, 2980–2988.

Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. 2014. Microsoft coco: Common objects in context. In *European conference on computer vision*. Springer, 740–755.

Gang Liu, Chuyi Wang, and Yanzhong Hu. 2018. RPN with the Attention-based Multi-Scale Method and the Adaptive Non-Maximum Suppression for Billboard Detection. In *2018 IEEE 4th International Conference on Computer and Communications (ICCC)*. 1541–1545. https://doi.org/10.1109/CompComm.2018.8780907

Norbert Maliszewski, Anna Olejniczak-Serowiec, and Justyna Harasimczuk. 2019. Influence of sexual appeal in roadside advertising on drivers' attention and driving behavior. *PloS one* 14, 5 (2019), e0216919.

Hadas Marciano et al. 2017. The effect of billboard design specifications on driving: a pilot study. *Accident Analysis & Prevention* 104 (2017), 174–184.

Lynn Meuleners, Paul Roberts, and Michelle Fraser. 2020. Identifying the distracting aspects of electronic advertising billboards: A driving simulation study. *Accident Analysis & Prevention* 145 (2020), 105710.

Kristof Mollu, Joris Cornu, Kris Brijs, Ali Pirdavani, and Tom Brijs. 2018. Driving simulator study on the influence of digital illuminated billboards near pedestrian crossings. *Transportation research part F: traffic psychology and behaviour* 59 (2018), 45–56.

Gerhard Neuhold, Tobias Ollmann, Samuel Rota Bulo, and Peter Kontschieder. 2017. The mapillary vistas dataset for semantic understanding of street scenes. In *Proceedings of the IEEE international conference on computer vision*. 4990–4999.

Romi Fadillah Rahmat, Dennis Dennis, Opim Salim Sitompul, Sarah Purnamawati, and Rahmat Budiarto. 2019. Advertisement billboard detection and geotagging system with inductive transfer learning in deep convolutional neural network. *TELKOMNIKA (Telecommunication Computing Electronics and Control)* 17, 5 (2019), 2659–2666.

Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. 2016. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. IEEE, 779–788.

Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. 2015. Faster r-cnn: Towards real-time object detection with region proposal networks. In *Advances in neural information processing systems*. 91–99.

N. Riche, M. Duvinage, M. Mancas, B. Gosselin, and T. Dutoit. 2013. Saliency and human fixations: State-of-the-art and study of comparison metrics. In *International Conference on Computer Vision*. IEEE, 1153–1160.

Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. 2018. Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 4510–4520.

Yaoru Sun and Robert Fisher. 2003. Object-based visual attention for computer vision. *Artificial intelligence* 146, 1 (2003), 77–123.

Adam Tarnowski, Anna Olejniczak-Serowiec, and Agnieszka Marszalec. 2017. Roadside advertising and the distraction of driver's attention. In *MATEC Web of Conferences*, Vol. 122. EDP Sciences, 03010.

Zhi Tian, Chunhua Shen, Hao Chen, and Tong He. 2019. Fcos: Fully convolutional one-stage object detection. In *Proceedings of the IEEE/CVF international conference on computer vision*. 9627–9636.

Inam Ullah, Muwei Jian, Sumaira Hussain, Jie Guo, Hui Yu, Xing Wang, and Yilong Yin. 2020. A brief survey of visual saliency detection. *Multimedia Tools and Applications* 79 (2020), 34605–34645.

Eleonora Vig, Michael Dorr, and David Cox. 2014. Large-scale optimization of hierarchical features for saliency prediction in natural images. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2798–2805.

Malgorzata Zalesinska. 2018. The impact of the luminance, size and location of LED billboards on drivers' visual performance—laboratory tests. *Accident Analysis & Prevention* 117 (2018), 439–448.

Shifeng Zhang, Cheng Chi, Yongqiang Yao, Zhen Lei, and Stan Z Li. 2020. Bridging the gap between anchor-based and anchor-free detection via adaptive training sample selection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 9759–9768.

Xingyi Zhou, Dequan Wang, and Philipp Krähenbühl. 2019. Objects as points. *arXiv preprint arXiv:1904.07850* (2019).