Title	Towards the Use of Computer Vision Techniques on Streetscape Imagery to Empower Citizens in the Planning Enforcement Process
Authors(s)	Lynch, Sam, Cuffe, Paul
Publication date	2023-09-01
Publication information	Lynch, Sam, and Paul Cuffe. “Towards the Use of Computer Vision Techniques on Streetscape Imagery to Empower Citizens in the Planning Enforcement Process,” September 1, 2023. https://doi.org/10.5281/zenodo.8223949.
Conference details	The 25th Irish Machine Vision and Image Processing Conference, University of Galway, Ireland, 30 August - 1st September 2023
Item record/more information	http://hdl.handle.net/10197/25722
Publisher's version (DOI)	10.5281/zenodo.8223949

Towards the Use of Computer Vision Techniques on Streetscape Imagery to Empower Citizens in the Planning Enforcement Process

Sam Lynch and Paul Cuffe

School of Electrical and Electronic Engineering, University College Dublin

Abstract

Urban streetscapes are often cluttered with intrusive advertising signage, which is typically erected without appropriate planning permission. This paper proposes the deployment of computer vision techniques to automatically identify this type of illicit signage within geotagged and timestamped digital images taken of an urban streetscape from a moving vehicle. Such object detection can underpin a semi-automated workflow for instigating planning enforcement complaints against offending signage at scale. The proposed method adapts deep learning models for object detection on a manually collated and labelled dataset of 1051 images containing illegal advertising signage. The system is evaluated on a batch of acquired streetscape images collected from various urban areas in Dublin, Ireland. These early results the broad feasability of automatically detecting non-compliant vinyl banners and property signs. The main research contribution of this paper is illustrating the potential for computer vision techniques to mediate new relationships between citizens and local authorities.

Keywords: Machine Learning (ML), Deep Learning (DL), Single-Shot Detector (SSD)

1 Introduction

Ever since the invention of the printing press, outdoor advertising has been widespread in cities. This can greatly harm the visual amenity of urban spaces: such advertisements have been termed as *visual pollutants*. Research shows that urban spaces free of visual pollution increase the quality of life of people living in that environment [Voronych, 2013], while, also, inspiring a sense of pride and belonging in their area [Jensen et al., 2014].

Figure 1: Unauthorised advertising signage in Dublin, which cheapens the visual amenity of this otherwise pleasant canal bank setting

The regulation of outdoor advertising varies from country to country; for example, in Southeast Asia there is little to nothing public administrations can do about what is built or assembled in urban spaces [Jana and De, 2015]. In many developed nations, such as in Ireland, to erect an advertising sign, one must obtain planning permission from the relevant local authority. Notwithstanding this, though, it is common sight to see large vinyl advertising banners hung on and around commercial premises, typically without any such planning permission.

Likewise, ubiquitous *For Sale* realty signage typically exceeds the modest dimensions allowed for small developments that are exempt from having to seek planning permission. For instance, typical requirements state that planning-exempt property signs are *"subject to a maximum area of 0.6sqm for a house and 1.2sqm for any other structure/land. There must be one sign only and it must be removed no later than 7 days after the sale/letting."*.

This research aims to tackle the problem of visual pollution by facilitating the reporting of these signs to authorities with greater scale and speed. In the Irish example, Local Authorities are bound by the *Planning & Development Act*, 2000 (section 152) to investigate all written allegations of unauthorised developments within six weeks of receiving same. This project aims to deploy modern image processing techniques to activate this legal provision at much greater scale, by automating the creation of written complaints against specific planning breaches.

The ambition of this project is to deploy modern computer vision techniques on batches of streetscape imagery to mediate a new relationship between citizens and the planning enforcement officers within Local Authorities who are legally empowered to act against unauthorised advertising displays.

1.1 State of the Art

As computing hardware and computing techniques improved as costs went down through the 1990s, developments in subsets of machine learning such as convolutional neural networks (CNNs) and deep learning, which are inspired by neurons in the human brain, improved. A major breakthrough in computer vision occurred in 2012 with the release of a deep neural network called AlexNet [Krizhevsky et al., 2012].

The rapid evolution of such computer vision techniques has seen the technology being utilised extensively in many sectors. Impressive applications have been seen in transportation [Cao et al., 2021], security [Moeslund and Granum, 2001], automatic speech recognition [Yu and Deng, 2016] and agriculture [Bhargava and Bansal, 2021].

Early research in advertising detection used computer vision techniques to detect billboards in sports TV. For instance [Watve and Sural, 2008], where soccer videos were deterministically processed to detect pitch-side advertisement billboards. This paper used Hue slicing and canny edge detection to find a region of interest where the billboard lies in a shot.

The problem with edge and colour detection techniques is that all edges of the signage must visible and the background must be plain with a contrasting colour to the sign. In a more complex environment, such as at street-level, a supervised machine learning approach is required for advertisement detection. Modern research has shown that CNNs produce accurate results for outdoor image classification tasks such as advertising billboard detection. Architectures within this space include AlexNet, VGG, GoogLeNet, NiN, DenseNet and ResNet [Alom et al., 2018].

Work in [Rahmat et al., 2019] retrained AlexNet for billboard detection. This paper successfully made use of the technique of transfer learning, where the weights and biases of pre-trained models can be transferred and used as initial weights and biases for a new (unseen) dataset. Work in [Moffett et al.,] used streetscape imagery for automated detection of tobacco advertising. This research used Faster-RCNN (Region-based CNN) [Chen and Gupta, 2017] to detect the tobacco advertisements and this produced good results.

Work in[Bochkarev and Smirnov, 2019] deals with the problem of illegal advertising on building façades. This research focused on the regional-level laws that dictate the conditions for advertising on building façades in Saint-Petersburg, developing a set of checks for detected advertising objects to check their legality. Thus, although the motivation of the paper is the same as the present one it few transferable insights on how to tackle the problem. Work in [Jiang et al., 2020] developed an algorithm to automatically detect illegal billboards in Hebei Province. This paper finds that a Faster R-CNN detection framework is successful for detecting illegal billboards.

A very attractive feature of CNNs for object detection is transfer learning, where a pre-trained neural network model can be repurposed for a new image classification task with a new dataset. There exists a range of mature neural network models that can be re-trained on a new dataset using an open-source framework such as TensorFlow Object Detection API [Abadi et al., 2016]. The two major architectural types of object detection model are: one-stage detection, such as You Only Look Once (YOLO) [Redmon et al., 2016], and Single Shot Detection (SSD) [Liu et al., 2015]; and two-stage detection models, such as Faster R-CNN [Ren et al., 2015]. The key difference between the two types of models is that in the two-stage detection model, the region of interest is, determined first and detection is performed on the region of interest only, resulting in a slower but generally more accurate process that is computationally expensive.

In a recent study comparing the SSD model and the YOLO model [Ángel Morera et al., 2020] propose a robust method for the automatic detection of urban advertising panels in outdoor images. Their work usef a Single Shot Multiblox Detector (SSD) for advertising detection. Their paper states that the SSD model is most suited for this type of application as it does not re-sample features for bounding box hypotheses

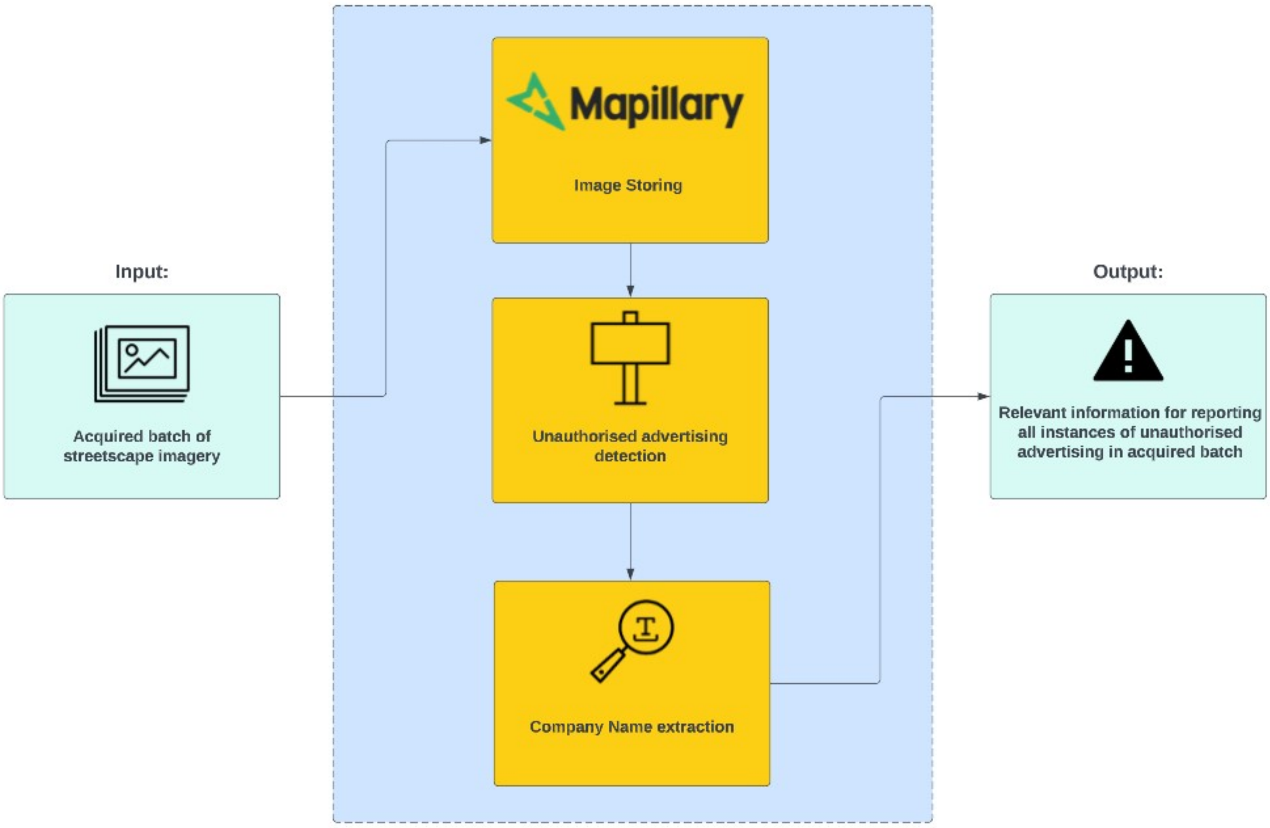

Figure 2: Broad flowchart of developed algorithm

Text extraction, to produce semantic data on the content of advertisements in scene images presents a more complex problem. Work in [Zhou et al., 2017] presents a fast and accurate solution for the detection of text in scenery with the EAST (Efficient and Accurate Scene Text) detector. With the regions of text in an image detected, the characters of the text in the region must be recognised to reveal what the text says. The Tesseract OCR engine [Smith, 2007] is used for the text recognition in this research.

2 Methodology

2.1 Image Acquisition

As outlined in figure 2, the input to the algorithm is a batch of streetscape imagery which is acquired on a drive through a city. The images acquired are stored in a suitable hosting infrastructure. This implementation utilises the open-source hosting infrastructure Mapillary. This platform is selected as it allows users to freely and conveniently upload images to Mapillary, where they are geotagged and stored with their stret address (see a comparative analysis in [Mahabir et al., 2020]). Through interactions with the Mapillary API, the longitude and latitude coordinates of a queried image may be obtained. This research uses interactions through Python to obtain the coordinates of an image where a potentially unauthorised advertisement is situated.

2.2 Unauthorised Advertising Detection

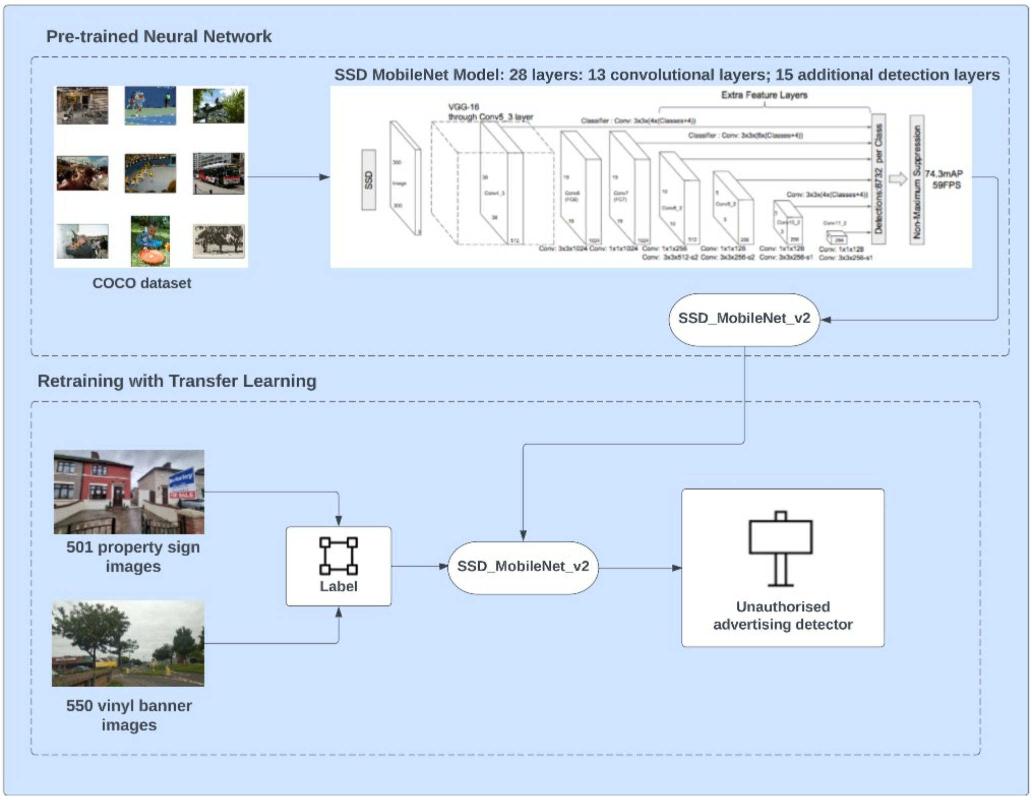

Figure 3: Process of transfer learning utilised to train unauthorised advertising detection model

2.2.1 Dataset Acquisition and Image Tagging

There are no public datasets of outdoor unauthorised advertising. Therefore, one task of this research endevour was to create a suitable dataset to train the object detection model. It is very important to obtain a large and diverse set of training images for each class that are to be detected, to ensure that the object detection model performs well on new, unseen, images.

The classes for detection in this research are the types of unauthorised developments discussed previously: vinyl banners and property signs. Addressing the former, 550 images with vinyl banners were acquired from Visual Genome's [Krishna et al., 2017] VG_100k dataset and the Mapillary Vistas dataset [Neuhold et al., 2017]. Acquiring the set of property sign images was more challenging as the property signs of interest are much less frequently found in the Mapillary Vistas dataset than vinyl banners. 501 images with property signs were acquired from Google and through the Mapillary API using Python, by filtering for advertisement detections. Combining these gives a total dataset with 1051 images.

Each image was labelled with object bounding boxes in Pascal VOC format by use of the open-source graphical image annotation tool LabelImg [Tzutalin, 2015]. The dataset was, then, split randomly into training images and test images with 80/20 Training/Test split.

2.2.2 Object Detection Model

The pre-trained SSD MobileNetV2 is implemented for this object detection task, on the TensorFlow Object Detection API framework. The MobileNet [Howard et al., 2017] is a object detector released in 2017 as an efficient CNN architecture designed for mobile and embedded vision application, it is chosen for this task due

Figure 4: Examples of model performance on test images

to it's lean network and high accuracy. SSD , introduced by Liu et al. [Liu et al., 2015], is based on a feed-forward convolutional network that scores bounding box regions based on a confidence score for the presence of the object class in that box. This is followed by a non-maximum suppression step to produce the detections. The architecture of an SSD network uses two steps for object detection: (1) extraction of feature maps; (2) application of convolution filters to detect objects. For the extraction of feature maps, SSD uses the very deep convolutional network for image recognition VGG16, the SSD MobileNetV2 model uses the MobileNetV2 deep CNN. The object detection is carried out using the Conv4_3 layer. Each prediction from the Conv4_3 layer comprises of a bounding box and 21 scores for each object class (including no class); then, simply, the class with the highest confidence score is selected as the bounded object class. Conv4_3 makes a total of $38 \times 38 \times 4$ predictions: four predictions per cell.

2.2.3 Training the Model

The TensorFlow Object Detection API was used to re-train the SSD MobileNetV2 model, in Python. TensorFlow's object detection pipeline includes a configured file with configuration selection optimised for efficient model training. The parameters to be chosen were batch size and number of training epochs. 12 was chosen as the default batch size and 50,000 was chosen as the number of training epochs. The configuration file was updated to include the path to the label map, and the training and test TFRecords. The model was trained on Google Colab and exported to a local machine once training was complete.

2.2.4 Company Name Extraction

Text detection and recognition is carried out on the cropped detections from the unauthorised advertising detector. From the recognised text, the name of the company responsible for the erection of the sign can be extracted manually.

2.3 Collation of Output Information For Reporting At Scale

The output of the developed algorithm in this research is the relevant information required to instigate the removal of any instances of unauthorised advertising detected in the acquired input sequence. Such information is: a clear image with bounding boxes around the unauthorised advertising detected; cropped image(s) of the unauthorised advertising detected in the image; the address where the image is taken; and, where possible, the name of the company responsible for erecting the advertising. Local authorities outline enforcement measures for the removal of unauthorised advertising from both private and public areas, for example in the Dublin City Development Plan 2022-2028 [Dev, 2022]. Thus, the findings gathered from the algorithm can form many emails of complaint to the local authority reporting all detected instances and requesting their removal.

The output findings may include multiple detections of the same sign and may contain some incorrectly detected instances, e.g. square road signs. A script was written to automatically filter out any duplicate detections of the same sign. If any two detections are found in images taken closer than a pre-defined threshold of 5 metres, calculated by subtracting the longitude and latitude coordinates, the histograms of the cropped images are compared using OpenCV's correlation comparison method. If the histograms are found to be sufficiently similar, meaning the same sign is pictured in both, the detection with higher confidence score is kept and the other detection is discarded.

The output detections should, then, be manually parsed to filter out any duplicate detections missed and any incorrect detections to eliminate the risk of redundant emails of complaint.

3 Results

The scripts implementing the functionality in Fig. 2, and the training image set, are available in an enduring online repository at [Cuffe and Lynch, 2023].

An example batch of streetscape imagery was acquired by use of the Mapillary app on a drive of approx. 16 km^2 through South County Dublin [1]. This batch contained 644 images.

The acquired imagery was uploaded to Mapillary where geotagging and privacy blurring was applied to each image. The computer vision scripts were executed locally against these with a confidence threshold of 60% set for detections.

Figure 5: Map of the area surveyed with pins depicting location of each potential property sign detected by algorithm

After duplicates were automatically detected and filtered out, the output contained 77 potential instances of property signs and 45 potential instances of vinyl banners.

As shown in Table 1, analysis of this test acquisition shows that 56 property signs were correctly detected and classified, 21 instances were incorrectly classified as property signs and 8 property signs were not detected. This means that the algorithm precision for property signs is 72.7% and more importantly, due to the manual filtering of the output,the recall for property signs is 87.5%.

30 vinyl banners were correctly detected and classified, 15 instances were incorrectly classified as vinyl banners and 7 vinyl banners were not detected. Resulting in an algorithm precision for vinyl banners of 66.67% and more importantly a recall for vinyl banners of 81.1%.

	True Positives	**False Positives**	**False Negatives**	**Precision**	**Recall**
Property Signs	56	21	8	72.7%	87.5%
Vinyl Banners	30	15	7	66.7%	81.1%

Table 1: Property sign and vinyl banner algorithm results summary for test acquisition

4 Conclusion

This research describes a tool to empower ordinary citizents to initiate enforcement actions at scale against the proliferation of unauthorised advertising sigange. The results show that the object detection model can accurately detect and locate instances of property signs and vinyl banners in streetscape images. The test acquisition obtained, surveyed an area of approximately 16 km^2. In this batch of

Figure 6: Map of area surveyed with pins depicting location of each potential vinyl banner detected by algorithm

[1] The journey can be viewed online on Mapillary itself

imagery the tool presented in the research detected 87.5% of property signs, 56 in total, and 81.1% of vinyl banners, 30 in total, in the area, which highlights the potential scale of the tool.

This tool allows citizens to have a considerable impact on the protection of the visual beauty of their city by reporting unauthorised advertising at scale. Before this research, the only way to report unauthorised advertising was to take a photo of any potentially unauthorised advertising witnessed, and report it in an email, manually recording the address of the advertising and the time the photo was taken. Consequently, reporting instances at a large scale is a very labour intensive process which is very time-consuming and impractical to execute. The automated workflow presented in this research ensures that the manual effort involved in the reporting of an instance has decreased significantly, due to automatic geotagging, detection and timestamping.

References

[Dev, 2022] (2022). Dublin City Development Plan 2022-2028 Volume 2: Appendices.

[Abadi et al., 2016] Abadi, M. et al. (2016). TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. *arXiv preprint arXiv:1603.04467*.

[Alom et al., 2018] Alom, M. Z., Taha, T., Yakopcic, C., Westberg, S., Hasan, M., Esesn, B., Awwal, A., and Asari, V. (2018). The History Began from AlexNet: A Comprehensive Survey on Deep Learning Approaches.

[Bhargava and Bansal, 2021] Bhargava, A. and Bansal, A. (2021). Fruits and vegetables quality evaluation using computer vision: A review. *Journal of King Saud University-Computer and Information Sciences*, 33(3):243–257.

[Bochkarev and Smirnov, 2019] Bochkarev, K. and Smirnov, E. (2019). Detecting advertising on building façades with computer vision. volume 156, pages 338–346. Elsevier B.V.

[Cao et al., 2021] Cao, J., Song, C., Song, S., Xiao, F., Zhang, X., Liu, Z., and Ang, M. H. (2021). Robust object tracking algorithm for autonomous vehicles in complex scenes. volume 13. MDPI AG.

[Chen and Gupta, 2017] Chen, X. and Gupta, A. (2017). An Implementation of Faster RCNN with Study for Region Sampling.

[Cuffe and Lynch, 2023] Cuffe, P. and Lynch, S. (2023). Scripts and data from "Towards the Use of Computer Vision Techniques on Streetscape Imagery to Empower Citizens in the Planning Enforcement Process". *DOI: 10.6084/m9.figshare.23703423.v1*.

[Howard et al., 2017] Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., and Adam, H. (2017). MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications.

[Jana and De, 2015] Jana, M. K. and De, T. (2015). Visual pollution can have a deep degrading effect on urban and suburban community: a study in few places of Bengal, India, with special reference to unorganized billboards. *European Scientific Journal*.

[Jensen et al., 2014] Jensen, C. U., Panduro, T. E., and Lundhede, T. H. (2014). The vindication of Don Quixote: The impact of noise and visual pollution from wind turbines. *Land economics*, 90(4):668–682.

[Jiang et al., 2020] Jiang, X.-H., Feng, H.-L., and Dong, Y.-J. (2020). Application of Neural Network in Image Detection of Illegal Billboards.

[Krishna et al., 2017] Krishna, R., Zhu, Y., Groth, O., Johnson, J., Hata, K., Kravitz, J., Chen, S., Kalantidis, Y., Li, L.-J., Shamma, D. A., et al. (2017). Visual genome: Connecting language and vision using crowdsourced dense image annotations. *International journal of computer vision*, 123:32–73.

[Krizhevsky et al., 2012] Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. In Pereira, F., Burges, C., Bottou, L., and Weinberger, K., editors, *Advances in Neural Information Processing Systems*, volume 25. Curran Associates, Inc.

[Liu et al., 2015] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., and Berg, A. C. (2015). SSD: Single Shot MultiBox Detector.

[Mahabir et al., 2020] Mahabir, R., Schuchard, R., Crooks, A., Croitoru, A., and Stefanidis, A. (2020). Crowdsourcing street view imagery: A comparison of mapillary and openstreetcam. *ISPRS International Journal of Geo-Information*, 9(6):341.

[Moeslund and Granum, 2001] Moeslund, T. B. and Granum, E. (2001). A survey of computer vision-based human motion capture. *Computer vision and image understanding*, 81(3):231–268.

[Moffett et al.,] Moffett, C., Abunku, P., Zhu, J., Chaturvedi, I., Chen, G., and Dobler, G. Capstone Project Report CUSP NYU 2018 Automated Detection of Street-Level Tobacco Advertising Displays.

[Neuhold et al., 2017] Neuhold, G., Ollmann, T., Bulo, S. R., and Kontschieder, P. (2017). The Mapillary Vistas Dataset for Semantic Understanding of Street Scenes. volume 2017-October, pages 5000–5009. Institute of Electrical and Electronics Engineers Inc.

[Rahmat et al., 2019] Rahmat, R. F., Dennis, Sitompul, O. S., Purnamawati, S., and Budiarto, R. (2019). Advertisement billboard detection and geotagging system with inductive transfer learning in deep convolutional neural network. *Telkomnika (Telecommunication Computing Electronics and Control)*, 17:2659–2666.

[Redmon et al., 2016] Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 779–788.

[Ren et al., 2015] Ren, S., He, K., Girshick, R., and Sun, J. (2015). Faster R-CNN: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, 28.

[Smith, 2007] Smith, R. (2007). An Overview of the Tesseract OCR Engine.

[Tzutalin, 2015] Tzutalin, D. (2015). LabelImg. *GitHub repository*, 6.

[Voronych, 2013] Voronych, Y. (2013). Visual pollution of urban space in lviv. *Przestrzeń I Forma*, (20):309–314.

[Watve and Sural, 2008] Watve, A. and Sural, S. (2008). Soccer video processing for the detection of advertisement billboards. *Pattern Recognition Letters*, 29:994–1006.

[Yu and Deng, 2016] Yu, D. and Deng, L. (2016). *Automatic speech recognition*, volume 1. Springer.

[Zhou et al., 2017] Zhou, X., Yao, C., Wen, H., Wang, Y., Zhou, S., He, W., and Liang, J. (2017). EAST: An Efficient and Accurate Scene Text Detector.

[Ángel Morera et al., 2020] Ángel Morera, Ángel Sánchez, Moreno, A. B., Ángel D. Sappa, and Vélez, J. F. (2020). SSD vs. YOLO for detection of outdoor urban advertising panels under multiple variabilities. *Sensors (Switzerland)*, 20:1–23.