https://dx.doi.org/10.4314/dujopas.v10i2b.17
ISSN (Print): 2476-8316
ISSN (Online): 2635-3490

Assessing Selected CNN Models for Efficient Feature Extraction in SSD for Text Detection in Advertisement Images

[1]Aanuoluwa O. Adio*, [2]Caleb O. Akanbi, [2]Adepeju A Adigun, [3]Abdulwakil Kasali, [4]Solagbade Adisa

[1]Department of Computer Science,
Redeemer's University,
Ede,
Nigeria.

[2]Department of Information and Communication Technology,
Osun State University,
Osogbo,
Nigeria.

[3]Department of Computer Engineering Technology,
Federal Polytechnic,
Ede,
Nigeria.

[3]Information and Communication Technology Unit,,
Federal Polytechnic,
Ede,
Nigeria.

Email: adioa@run.edu.ng

Abstract

Digital advertisement promotes goods and services using digital media and technology. These digital advertisement images contain important information on the product and services being advertised and seek to persuade potential customers to take specific actions toward contacting the advertiser. Manual extraction of information from the advertisement images is tedious and prone to errors. The literature on text detection from images, billboards, and signposts using Single-shot detection (SSD) is vast. However, the literature has not explored its performance for text detection on advertisement images. Therefore, there is a need to evaluate the performance of these models on advertisement images. The performance of three selected Convolutional Neural Network (CNN) models (Resnet-50, Mobilenetv2, and Resnet-101) with SSD for text detection in advertisement images was evaluated. A total of 400 digital advertisement images were manually collected and annotated for use in this study. Results of comparing the performance of selected CNN models with the SSD architecture for text detection from advertisement images showed that Resnet-50 performed well with the detection of small texts with a mean Average Precision (mAP) of 0.736, AP(small) of 0.692 and AR(small) of 0.781.

Keywords: Text Detection, Single shot detection, Advertisement Images, Convolutional Neural Network, Feature Extraction.

Author for Correspondence

INTRODUCTION

Advertising is a technique for media distribution of information that helps a business gather in more clients and improve its turnaround according to Ilyas & Nayan (2020). The advertisement aims to inform and enlighten potential customers about the goods sold and services rendered by the advertiser according to Anbukkarasi et al. (2023). Once the potential customer's attention is captured, it is expected that the customer contacts the advertiser using the contact information provided on the advertisement according to Sayali et al (2021). A digital advertisement image has complex background and contains lots of texts with varying sizes. The literature on text detection is wide and vast. Single shot detection (SSD) according to Liu et al (2016) has been proven to perform well in balancing between speed and accuracy Kang & Park, (2023). Since the detection of texts from the advertisement images needs to be done swiftly and also as accurate as possible, this informed the choice of SSD for text detection from advertisement images. However, the choice of a suitable CNN model that extracts relevant features needed for accurate detection of texts in advertisement images needs to be explored. Hence, this study aims to assess the performance of selected CNN models for feature extraction for use by SSD for text detection from advertisement images. To achieve this, a dataset containing advertisement images was created and annotated.

The studies on text detection from images are enormous Raisi et al. (2020), as their real-life application must be supported (Surana et al., 2022). Its application improves disaster response, plate number detection, traffic management, advertisement, and announcements. Several approaches for image text detection have been proposed in the literature, like R-CNN, EAST, YOLO, and SSD. Text and object detection from billboards was proposed by Intasuwan et al, (2018). The study suggests a system that automatically connects an advertisement with the product’s website using Tesseract OCR. It achieved a mean AP of 0.85 across five categories of text. Anbukkarasi et al, (2023) designed an enhanced feature model-based hybrid neural network for text detection and signboard, billboard, and News Tickers. The study uses both CNN and RNN for text recognition utilizing low-level, mid-level, and high-level features for enriched feature extraction. Results showed that the developed model achieved 90% precision, 91% recall, and 91% F-score on the Tamil news tickets images dataset. Yu & Zhang, (2021) worked on effectively recognizing texts in English billboards using deep learning. Progressive Scale Expansion Network detected texts in the English billboard dataset, and CRNN was used to identify the detected texts. The result of the study highlighted that the recognition of significant texts performed worse than average font texts. Sayali et al, (2021) designed a system to detect billboards in the wild. The study created a novel dataset of annotated billboard images from various geographical locations and used the Single shot detector architecture to detect the billboards. Borisyuk et al, (2018) suggested developing a reliable and precise optical character recognition system that could process hundreds of millions of photos every day in real time. The architecture of the system is two-step. Using Faster-RCNN, the first stage of word detection is carried out. In the second stage, a fully convolution model with CTC loss is used to recognize words. Each of the two models receives separate training. The system's outstanding productivity in terms of scale computing time and model accuracy was demonstrated by the results.

METHODOLOGY

Data Collection and Model Design

A total of 400 advertisement images were collected as screenshots from various social media platforms such as Twitter, WhatsApp and Instagram with the TECNO Pouvoir 4 mobile phone. The images were resized to 512 x 512 pixels. The images were annotated for text detection using the LabelImg annotation tool. SSD is a one-stage object detector which aim to detect objects in an image in a single step. The architecture makes use of bounding boxes with varying aspect ratios which helps to detect objects of various sizes. The Architecture is as shown in Figure 1. The SSD architecture comprises of the base network which is basically a CNN model and extra features layers as prediction layers. The advertisement images along with the annotation (.xml) files were the input to the architecture. For that base network, Mobilenetv2, Resnet-50 and Resnet-101 were selected to evaluate its performance.

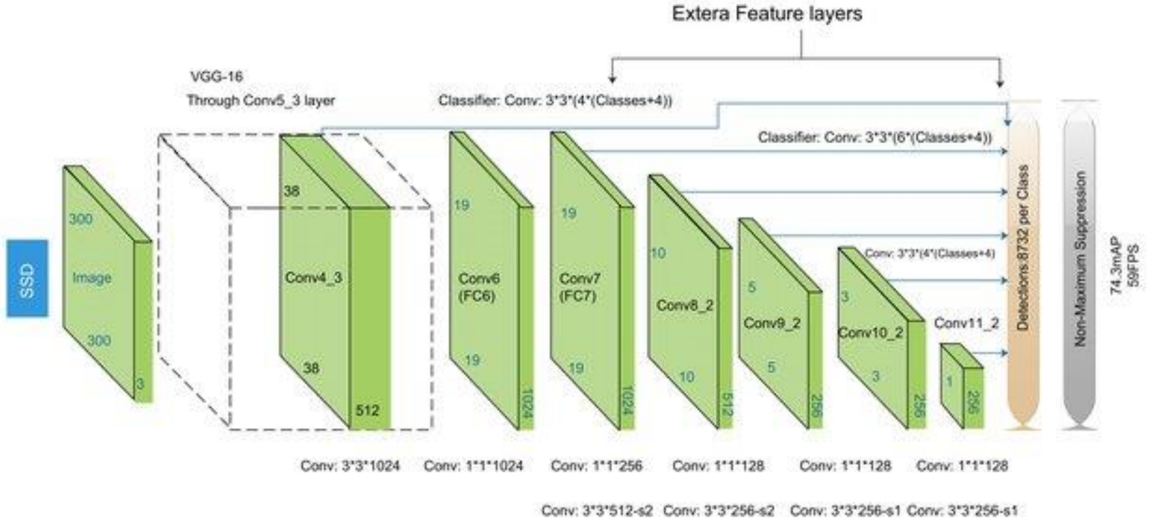

Figure 1: Single Shot Detection Architecture (Liu et al., 2016)

Experimental Setup

The text detection model is implemented in python programming language using the tensor flow framework in the Google Colab environment. The performance of the models was evaluated using the following performance metrics: mean average precision (mAP), average precision over varying sizes of IoU, average recall, average recall over varying sizes of IoU.
The experiments were taken in the Google Colab environment. Google Colab offers free GPU services for experimental purposes. The pertained models available in the Tensor flow Model zoo were also used. The advertisement images were divided into 80% training and 20% testing. The images used for training were augmented using various preprocessing techniques such as contrast enhancement, brightening, darkening, Addition of hue and saturation.

RESULTS AND DISCUSSION

Result of SSD-Resnet-50

A total step of 25000 and a warm up step of 2000, Results of the experiments carried out with Resnet-50 showed a mAP of 0.736, AP @50 of 0.874, AP@75 of 0.805, AP(small) of 0.692, AP(medium) of 0.857, AP(large) of 0.769, AR(all) of 0.764, AR(small) of 0.723, AR(medium) of 0.865, AR(large) of 0.781.

Result of SSD-Mobilenetv2

With a total step of 50000, and batch size of 12, Mobilenetv2 performed well with detection of large texts in terms of precision and recall. Result of the experiments carried out showed a mAP of 0.691, AP @50 of 0.863, AP@75 of 0.741, AP(small) of 0.629, AP(medium) of 0.853, AP(large) of 0.783, AR(all) of 0.726, AR(small) of 0.685, AR(medium) of 0.862, AR(large) of 0.807.

Result of SSD-Resnet-101

Results of the experiments carried out with Resnet-101 showed a mAP of 0.755, AP @50 of 0.878, AP@75 of 0.823, AP(small) of 0.718, AP(medium) of 0.854, AP(large) of 0.762, AR(all) of 0.780, AR(small) of 0.756, AR(medium) of 0.860, AR(large) of 0.769. The summary of the results is shown in Table 1. In Table 1, AP (s), AP(m), and AP(l) stands for AP(small), AP(medium), and AP(large) respectively. AR(s), AR(m), and AR(l) stands AR(small), AR(medium) and AR(large) respectively.

Table1: Result of the performance of selected CNN models for text detection

Model	**mAP**	**AP(.5)**	**AP(.75)**	**AP(s)**	**AP(m)**	**AP(l)**
Resnet-50	0.736	0.874	0.805	0.692	0.857	0.769
Mobilenetv2	0.691	0.863	0.741	0.629	0.853	0.783
Resnet-101	0.755	0.878	0.823	0.718	0.854	0.762
Model	**AR(all)**	**AR(s)**	**AR(m)**	**AP(l)**		
Resnet-50	0.764	0.735	0.865	0.781		
Mobilenetv2	0.726	0.685	0.862	0.807		
Resnet-101	0.780	0.756	0.860	0.769		

Testing inference time on 20 images showed that it took Resnet-50, Mobilenetv2, and Resnet-101 59.78s, 13.4s, and 83.89s respectively.

The result of the performance of Single shot detection with Resnet-101 on the advert images showed that the model performed well in terms of Average Precision and Recall, followed by Resnet-50. However, Mobilenetv2 did not perform so well like the other models for text detection especially when detecting small texts. From the experiments carried out, it should be noted that it took Mobilenetv2 less time to detect text on the test images compared to the other models. One of the issues to consider is the speed of detection and Resnet-101 took the longest time. This makes it not suitable to use when considering the design of a mobile application. Resnet-50 strikes a good balance between speed and accuracy.

CONCLUSION

Text detection from advertisement images is an important research in computer vision as it enables the advertiser achieve the aim of the advertisement as it ensures the prompt and accurate detection of texts on advertisement images that the potential customer will use to contact the advertiser. Despite the extensive research on object and text detection from images, billboards and sign posts, the research on text detection from advertisement image has not been explored. For future work, we plan to increase the number of images in the advertisement dataset. Also, explore other models to improve the detection accuracy of small texts.

REFERENCES

Anbukkarasi, S., Sathishkumar, V. E., Dhivyaa, C. R., & Cho, J. (2023). Enhanced Feature Model based Hybrid Neural Network for Text Detection on Signboard, Billboard and News tickers. *IEEE Access*. https://doi.org/10.1109/ACCESS.2023.3264569

Borisyuk, F., Gordo, A., & Sivakumar, V. (2018). Rosetta: Large scale system for text detection and recognition in images. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 71–79. https://doi.org/10.1145/3219819.3219861

Ilyas, N. A., & Nayan, S. M. (2020). Advertisement For Growing Business. *Journal of*

Undergraduate Social Science & Technology, 2(2), 242–254.

Intasuwan, T., Kaewthong, J., & Vittayakorn, S. (2018). Text and object detection on billboards. *Proceedings of 2018 10th International Conference on Information Technology and Electrical Engineering: Smart Technology for Better Society, ICITEE 2018*, 6–11. https://doi.org/10.1109/ICITEED.2018.8534879

Kang, S. H., & Park, J. S. (2023). Aligned Matching: Improving Small Object Detection in SSD. *Sensors*, *23*(5), 2589. https://doi.org/10.3390/s23052589

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., & Berg, A. C. (2016). SSD: Single shot multibox detector. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *9905 LNCS*, 21–37. https://doi.org/10.1007/978-3-319-46448-0_2

Raisi, Z., Naiel, M. A., Fieguth, P., Wardell, S., & Zelek, J. (2020). Text Detection and Recognition in the Wild: A Review. *ArXiv Preprint ArXiv:2006.04305*. http://arxiv.org/abs/2006.04305

Sayali Avinash Chavan, M., Kerr, D., Coleman, S., & Khader, H. (2021). Billboard Detection in the Wild. *Irish Machine Vision and Image Processing Conference 2021*, 57.

Surana, S., Pathak, K., Gagnani, M., Shrivastava, V., Mahesh, T. R., & Madhuri G., S. (2022). Text Extraction and Detection from Images using Machine Learning Techniques: A Research Review. *Proceedings of the International Conference on Electronics and Renewable Systems, ICEARS 2022*, 1201–1207. https://doi.org/10.1109/ICEARS53579.2022.9752274

Yu, E., & Zhang, Z. (2021). English billboard text recognition using deep learning. *Journal of Physics: Conference Series*, *1994*(1), 12003. https://doi.org/10.1088/1742-6596/1994/1/012003