Traffic Sign Recognition from Digital Images by Using Deep Learning

Jiawei Xing, Ziyuan Luo, Minh Nguyen, and Wei Qi Yan(✉)

Auckland University of Technology, Auckland 1010, New Zealand
weiqi.yan@aut.ac.nz

Abstract. Traffic signs are essentially needed to obey the traffic rules. Once a driver ignores the signs, especially those critical signs, due to the complexity of actual traffic scenes or the influence of inclement weather conditions, it will lead to violating traffic regulations or traffic accidents, causing casualties and property losses. Therefore, Traffic Sign Recognition (TSR) is an essential part of autonomous vehicles and has important academic significance. The main contributions of this paper are as follows: (1) We apply an algorithm to the dark channel prior, and we also provide a guided image filtering algorithm for image defogging. Our results show that the guided image filtering method is very effective in image defogging. (2) This paper presents a number of deep learning solutions towards the aforementioned problems. Based on the experiments conducted, we discover that YOLOv5 is very suitable for real-time TSR.

Keywords: TSR · Defogging · Faster R-CNN · YOLOv5

1 Introduction

With the continuous growth of populations and the development of our society, cars have become an indispensable means of transportation for people. While people relish the convenience of cars, the new technology also brings a series of inevitable problems to road traffic. The pressure on the urban transportation system is gradually increasing. The most important one is the frequent occurrence of traffic accidents [1]. In order to reduce accidents, advanced accident-avoiding systems have become popular [2]. With the development of traffic sign recognition (TSR), computer-aided systems have undergone a significant leap. Computers have been applied to simulate and process large amounts of visual data. An intelligent driving assistance system simulates the mechanism of the human visual system to complete visual object detection and recognition [3, 4] as well as other applications [5].

In a nutshell, the identification and detection of traffic signs are essential components, especially the research project on the identification of traffic speed limit signs in haze weather, which meets the needs of current automobile development and is conducive to the promotion of scientific and technological knowledge. Therefore, the research work on traffic sign detection under severe weather has important and practical significance. Thus, the goal of this paper is to collect foggy images as our dataset and compare two

H. Wang et al. (Eds.): PSIVT 2022, LNCS 13763, pp. 37–49, 2023.
https://doi.org/10.1007/978-3-031-26431-3_4

different deep learning methods. At the same time, the influence of fogs on visual object recognition is investigated, and finally a TSR method with fast speed and high precision is developed.

The remaining part of this paper is organized as follows: We have our literature review in Sect. 2, and our methods will be addressed in Sect. 3. Our results will be shown in Sect. 4. Our conclusion will be drawn in Sect. 5.

2 Literature Review

2.1 External Factors Affecting Traffic Sign Recognition

With the development of deep learning in artificial intelligence, TSR is being developed rapidly. Many high-end models are now equipped with TSR driver assistance systems to help drivers safely. The current TSR only has a higher accuracy rate when it is sunny and traffic signs are unobstructed, or in severe weather (e.g., foggy, rainy, snowy, etc.), and challenging lighting conditions (e.g., night, direct sunlight, etc.). If a traffic sign is obscured, false recognition or unrecognition may occur. Thus, we describe how these external factors affect TSR and what the differences are; in rainy weather, accurate TSR is also very difficult because fog is static and there is no apparent movement, while raining or snowing has dynamic motion characters [6]. This is very challenging for TSR, because TSR always needs real-time visual object detection. Lighting is also very important for TSR; during the daytime, if sunshine directly lights on a traffic sign or camera, it will cause overexposed. It is often too dark at night, at the same time, the lights will be reflected on traffic signs while driving at night. This may result in incorrect or unrecognizable traffic signs.

In haze weather, due to a number of disordered particles in the air, ambient light will be severely scattered and result in a blurry image; no relevant operations such as feature extraction are used [7]. Throughout the analysis of plenty of haze images and clear images of the same scene, we see that the haze image has specific characteristics. Analyzing these image features helps us identify the haze by using pattern classification [8], the imaging process is conducted on a foggy day by establishing the corresponding atmospheric scattering model and mathematical model [9].

Severe weather, such as haze, directly lead to the decline of image contrast, the reduction of grayscale dynamic range, the reduction of blurring, and the coverage of detailed information. Therefore, it is necessary to study how to restore a clear and fog-free image. At present, the research work on foggy removal is qualitatively grouped into two-fold: One is image restoration, which is based on an established mathematical model and previous knowledge of image degradation and reverses the model to calculate clear and fog-free Images [10]. The other direction is image enhancement which is based on human visual requirements through highlighting image details, filtering noises, and restoring clear images [11]. The difference between them is that image restoration is to improve the understanding of images from the perspective of image essence, image enhancement is based on human visual meaning to improve visual effect of the images to meet the needs of human vision. These algorithms could be applied to traffic sign location detection and traffic sign recognition.

2.2 Related Algorithms

Convolutional neural networks (CNNs) were inspired by cat visual cortex in physiology. There are neuronal cells that are extremely sensitive to external rays of light, which is called receptive field [12]. Since then, visual cortex began to enter our research area and attracted people's attention. Turnaround neural network was firstly applied to handwritten digit recognition. The model [13] cooperated with CNN and achieved excellent results in experiments through setting a precedent for a wide spectrum of applications in the fields of visual object detection, face recognition, speech recognition, and so on. Meanwhile, artificial neural network is improved and promoted. The neural network consists of five parts: Input layer, output layer, convolutional layers and pooling layers, fully connected layer [12]. Based on these neural networks, the workflow of TSR mainly includes three parts. The first is image preprocessing, which usually includes image enhancement, image scaling, and other image processing operations. The second part is traffic sign detection, which includes three important steps: (1) Extract candidate regions, (2) confirm traffic signs, (3) classify traffic signs.

3 Our Method

In this paper, we firstly make use of the guided image filtering algorithm to dehaze images and then compare the images before and after the dehazing process. Then, we introduce how to select the networks, including YOLOv5 and the improved YOLOv5. Finally, we state the evaluation methods of our experiment.

3.1 Dark Channel Prior Defogging Algorithm

The dark channel prior defogging algorithm [14] was firstly proposed in 2009. In the defogging algorithm, a consortium of outdoor sunny images are analysed. In the non-sky part of these images, one or two of the three RGB colour channels of each image have very low intensities [20]. There are four reasons why we take use of this method: The shadow of all kinds of glass, the projection of natural objects, the brightly coloured surface of visual objects, a dull surface of visual objects. As explained, the dark channel is shown as Eq. (1).

$$J^{dark}(x) = \min_{y \in \Omega(x)} (\min_{c \in [r,g,b]} J^c(y)) \tag{1}$$

where a colour channel is presented by r, g, or b as the components of c, J^{dark} is the image dark channel. At the same time, from the prior knowledge, we know that the grayscale intensity of a pixel in the dark channel is very low, namely, J^{dark} tends to 0.

In Eq. (1), atmospheric light is assumed to be the known variable. In fact, for any input image, 0.10% of the maximum grayscale intensity of pixels in the dark channel image corresponds to the average grayscale of pixels in the corresponding position of each channel of the original image, so as to obtain the atmospheric light of each channel.

On the premise that atmospheric light is assumed to be known, the atmospheric scattering model is transformed into

$$\frac{I^c(x)}{A^c} = t(x)\frac{J^c(x)}{A^c} + 1 - t(x) \tag{2}$$

where c means that each channel needs to be tackled separately. At the same time, we consider the light transmission *t(x)* as a constant, both sides of Eq. (2) are filtered twice to give the minima,

$$\min_{y\in\Omega(x)}(\min_c(\frac{I^c(x)}{A^c})) = \tilde{t}(x)\min_{y\in\Omega(x)}(\min_c(\frac{J^c(x)}{A^c})) + 1 - \tilde{t}(x) \tag{3}$$

where $\tilde{t}(x)$ is a constant, the minimum value $\tilde{t}(x)$ is calculated, J represents the original image, from the previous dark channel prior, we see that J^{dark} is close to 0. Combined with Eq. (1), we have

$$\min_{y\in\Omega(x)}(\min_c\left(\frac{J^c(y)}{A^c}\right)) = 0. \tag{4}$$

By substituting Eq. (4) to Eq. (3), the estimated value of transmittance $\tilde{t}(x)$ is obtained. The calculation is conducted as follows,

$$\tilde{t}(x) = 1 - \min_{y\in\Omega(x)}(\min_c\frac{I^c(y)}{A^c}) \tag{5}$$

In real cases, even in sunny days with good line of sight, there will still be tiny droplets and aerosol particles in the atmosphere. If all the fog is removed, it will have an impact on the realism of images. Therefore, a factor ω is introduced into Eq. (5), whose value is between [0, 1.00], Eq. (5) thus becomes:

$$\tilde{t}(x) = 1 - \omega\min_{y\in\Omega(x)}(\min_c\frac{I^c(y)}{A^c}) \tag{6}$$

where ω is generally set as 0.95. If *I(x)* is very small, the value of J will be too large, resulting in a lot of noises in the whole image. Therefore, a threshold t_0 should be set, if *t(x)* is less than t_0, let $t(x) = t_0$, the final one is shown as Eq. (7),

$$\tilde{t}(x) = 1 - \min_{y\in\Omega(x)}(\min_c\frac{I^c(y)}{A^c}) \tag{7}$$

3.2 Guided Image Filtering

Image defogging is an important pre-processing for the haze removal, which enhances visual effects such as edges and contours. Figure 1 is a pipeline of the haze removal process by using the dark channel prior algorithm.

Image filtering algorithm adopts an image to guide and filter the target image so that the final output image roughly resembles to the target image, the texture is akin

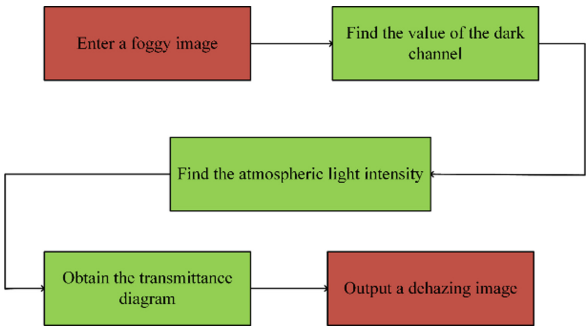

Fig. 1. The workflow of fog removal algorithm

to the guiding image. The guiding or reference image is either a different one or the same one as the input image itself. If the guiding image is equivalent to the input image, the filtering becomes an edge-preserving operation, which is able to be used for image reconstruction. By using visual features from the guided image filtering, haze image processing for traffic signs achieved the ideal results after image denoising [19], image smoothing, and fog removal.

3.3 YOLOv5 Model for Traffic Signs Recognition

In haze weather, the problem of traffic sign images will lead to a decline in the recognition accuracy of deep learning models, which poses a threat to traffic safety requirements. At the same time, the angle and size of the traffic sign will lead to a decrease in recognition accuracy. The rapidity of real-time TSR also has high requirements on computational speed of the model. Thus, we improve YOLOv5 model, which has great advantage in small object detection, while taking into account of TSR accuracy rate and speed, so as to better complete TSR in haze weather. At the same time, we have also improved the YOLOv5 model for TSR by using satellite images, another auxiliary landmark detection is proposed so as to achieve better results.

YOLO is a fast and compact open-source object detection model [15–17]. Compared with other nets, it has strong performance with very good stability. YOLO framework treats visual object detection as a regression problem, which is the first one that harnesses deep neural network in the end-to-end way that predicts the class and bounding box of visual objects. At present, YOLOv5 has faster recognition speed and smaller network size than YOLOv4 [18]. While model training by using different datasets, YOLOv3 and YOLOv4 need a program to calculate the initial anchor box, YOLOv5 automatically calculate the best anchor box for multiple datasets. In YOLOv5, we have fine-tuned the parameters, set the learning rate as 1.20×10^{-3}, the momentum as 0.95, the batch size as 16, and the epoch is assigned to 200.00 according to the batch size.

4 Our Results

4.1 Data Sources and Data Collection

Our dataset contains a total of 3,105 images and 5,536 instances. Pertaining to the experiments for TSR, we took use of our own dataset, where each image was manually labelled with a traffic sign. There are 12 classes of traffic signs which were included in this database. However, the visual data we collected did not include foggy images. Thus, we also utilized FRIDA, FRIDA2, and FROSI databases. FRIDA is made up of 90 images of 18 urban road scenes, while FRIDA2 consists of 330 images of 66 different road scenes. They were viewed from the same perspective as a driver with a variety of fogs in original images: Homogeneous fog, heterogeneous fog, cloudy fog, and cloudy heterogeneous fog, as well as the traffic signs like give-way, watch-out for pedestrians, etc. The FROSI dataset contains fog visibility from 50 m to 400 m with 1,620 traffic signs at various locations. By using these datasets, we are able to train our YOLOv5 model and Faster R-CNN model much comprehensively. Amongst them, 60% images were used for training, 20% for validation, and 20% for testing.

4.2 Comparison and Analysis of Two Defogging Model

In this section, we analyse and compare the defogging results by using the dark channel algorithm and the guided image filtering method. Figure 2 shows the output of each defogging algorithm.

Fig. 2. The results of defogging methods for various scenes.

We see from the results that the dehazing algorithm based on guided image filtering is robust, the dehazing outcome is stable for multiple scenes, it obtains a better defogging result with less colour distortion or darkening of the image. On the contrary, it plays a key role in enhancing the colour images.

4.3 TSR Experiments

In our experiments, we optimized our parameters of Faster R-CNN, we set momentum as 0.90, learning rate as 0.01, maximum epochs as 200, batch size as 24, and weight attenuation as 3.00×10^{-4}. At the same time, we took use of the fully connected layer and ReLU activation function to extract visual features of objects from given images.

Table 1. Experimental results of Faster R-CNN in various whether conditions

Weather	Precision	Recall	mAP@0.5
Sunny	0.97	0.97	0.97
Foggy	0.89	0.90	0.90
Weather	Precision	Recall	mAP@0.5

Table 2. Experimental results of Faster R-CNN with guided image filtering

Weather	Precision	Recall	mAP@0.5
Sunny	0.96	0.96	0.96
Foggy	0.93	0.93	0.93
Weather	Precision	Recall	mAP@0.5

In our experiment, we grouped our data into sunny days and foggy days, and derived the difference between recall, precision, and mAP with and without guided image filtering, as shown in Table 1 and Table 2. These two tables compare the accuracy and recall rates before and after using guided image filtering on sunny and foggy days. Faster R-CNN has higher accuracy and recall rates.

In Table 1 and Table 2, we compare the accuracy and recall rates before and after using guided image filtering on sunny and foggy images. The Faster R-CNN has a higher precision and recall rates. We took use of a guided image filtering method to defog the foggy day images. By using guided image filtering, the TSR accuracy based on sunny images is much higher than that on foggy images. Throughout using guided image filtering method to defog the given images, the accuracy rate by using the images with sunny days is 0.70% lower, the accuracy on foggy days is 3.00% higher, because the guided image filtering method not only removes the fog from the foggy images, but also adds a small number of noises to the sunny images.

Table 3 shows the TSR results using YOLOv5 without guided image filtering. We see from Table 3 that YOLOv5 has higher recognition accuracy and recall rate by using the images shot on sunny days, but the accuracy decreases by using the images on foggy days. At the same time, the recall rate dropped by 6.50%, which is significantly lower than the accuracy and recall rates of images on sunny days. In order to improve the TSR accuracy of images on foggy days, the experimental results of YOLOv5 after using guided image filtering are shown in Table 4. The TSR accuracy rate of images on sunny

days has dropped by 0.30%. The reason is that we added a plethora of noises to the traffic signs whilst removing the fogs, but the accuracy of images on foggy days has increased 3.90% compared with previous results, which effectively improve the accuracy of the visual object recognition with the images taken on foggy days.

YOLOv5 model, which is good at small targets and multi-target detection, conducts multitarget detection in complex scenes such as landmark images by improving the optimization ability of loss function in the adaptive balance of foreground and background loss. In other words, the improved loss function makes the model paying much attention on the image with small-size objects after defogging, which focuses on a variety of traffic signs, and accurately recognizes visual objects from complicated road conditions.

Table 3. Experimental result of Faster R-CNN with guided image filtering.

Weather	Precision	Recall	mAP@0.5
Sunny	0.96	0.96	0.96
Foggy	0.88	0.89	0.89
Weather	Precision	Recall	mAP@0.5

Table 4. Experimental result of Faster R-CNN with guided image filtering.

Weather	Precision	Recall	mAP@0.5
Sunny	0.95	0.96	0.96
Foggy	0.92	0.93	0.93
Weather	Precision	Recall	mAP@0.5

4.4 Result Comparisons

In this section, we compare the experimental results of YOLOv5 and other models. Figure 3 shows our TSR in sunny weather conditions, Fig. 4 and Fig. 5 indicate the results from videos with fogs. Figure 6 shows the results with different datasets.

In Fig. 3, we see that the TSR results reveal that Faster R-CNN often misses or fails to detect if the signs are far away from our camera. In contrast, YOLOv5 has a much higher recognition accuracy and computing speed while recognizing small objects or objects that are moving faster. Figure 4 and Fig. 5 show the recognition results on a foggy day, the results are roughly similar to those on a sunny day.

The video we tested is composed of 2,590 frames after processing. YOLOv5 takes 9.00×10^{-3} s to cope with each frame. Under the same accuracy rate, YOLOv5 has a faster recognition speed. Because TSR is often used for real-time object detection and recognition with high requirements of computing speed, YOLOv5 is more suitable for TSR. Figure 8 shows the recognition results of the two methods in the FRIDA dataset.

From Fig. 4 and Fig. 5, we see that both methods achieve accurate detection of traffic signs against a complex fog background. YOLOv5 method did so perfectly. It also benefits from a lighter model size and computing speed. Overall, YOLOv5 performs better if recognizing small objects (Fig. 7).

Fig. 3. The TSR results on sunny days (a) Faster R-CNN (b) YOLOv5.

(a) (b)

Fig. 4. The TSR results by using Faster R-CNN (a) and YOLOv5 (b) on the same scene.

(a)

(b)

Fig. 5. The TSR results on foggy days (a) R-CNN (b) YOLOv5.

(a)

(b)

Fig. 6. The TSR results based on FRIDA dataset (a) Faster R-CNN (b) YOLOv5.

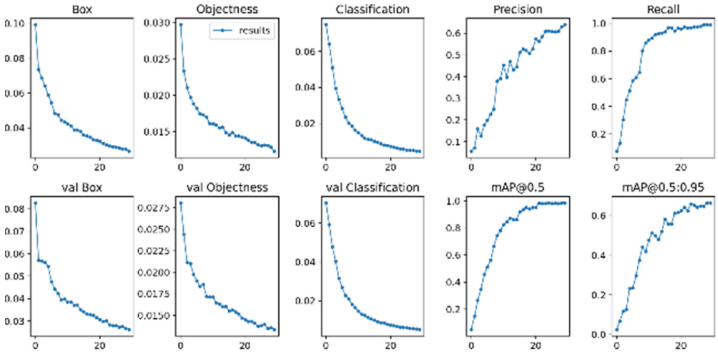

Fig. 7. TSR results by using YOLOv5

4.5 Analysis and Discussions

In the experiments, Fig. 8 shows how each metric changes as the number of iterations increases. In the case where the prediction of bounding boxes decreases with the increase of iterations currently, mAP increases with the growth of iterations at this time. It shows

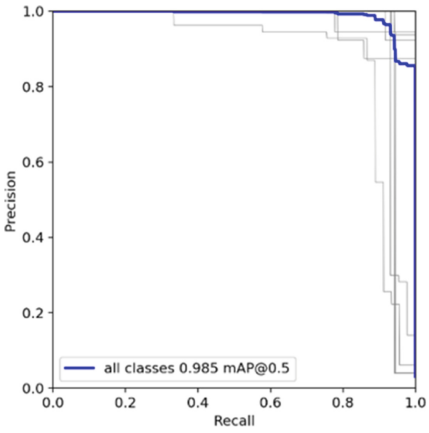

Fig. 8. The PR curve of YOLOv5.

that as the number of iterations rises, the proposed net in this paper is getting better and better. The accuracy and recall will increase with the iterations of the network training. This indicates that the number of true positive samples in the detection also grows as the number of iterations boosts up. Figure 8 is the PR curve of the test results of our experiment, y-axis is the accuracy rate, x-axis is the recall rate. The PR curve is very close to the top-right corner, which indicates that the model is effective. Therefore, the TSR based on YOLOv5 is better which has been well developed.

5 Conclusion

In this paper, we choose the best network for TSR by using different backbone networks. Then we utilize cross-layer links and activation functions to construct feature maps more efficiently, followed by feature extraction. We provide the method of YOLOv5 to detect traffic signs. We improved the loss function to improve the performance of YOLOv5. In our experimental comparisons, we see that YOLOv5 is very important. However, with similar accuracy, real-time TSR usually requires faster recognition. This confirms that YOLOv5 is a better choice for TSR.

Our future work includes three aspects. Firstly, we will continue to expand our dataset by adding more samples in various lighting conditions, such as fog and rain. Second, we compare more object recognition and detection methods in TSR. Finally, we will use more evaluation methods to evaluate our model, which will be able to intuitively discover the shortcomings of our model and make our model more robust and powerful.

References

1. Litman, T., Burwell, D.: Issues in sustainable transportation. Int. J. Global Environ. Issues **6**(4), 331–347 (2010)
2. Berkaya, S.K., Gunduz, H., Ozsen, O., Akinlar, C., Gunal, S.: On circular traffic sign detection and recognition. Expert Syst. Appl. **48**, 67–75 (2016)

3. Shi, X., Fang, X., Zhang, D., Guo, Z.: Image classification based on mixed deep learning model transfer learning. J. Syst. Simul. **28**, 167 (2016)
4. Ma, X., Fu, A., Wang, H., Yin, B.: Hyperspectral image classification based on deep deconvolution network with skip architecture. IEEE Trans. Geosci. Remote Sens. **56**, 4781–4791 (2018)
5. Pan, C., Sun, M., Yan, Z., Shao, J., Wu, D., Xu, X.: Vehicle logo recognition based on deep learning architecture in video surveillance for intelligent traffic system. In: International Conference on Smart and Sustainable City (2013)
6. Garg, K., Nayar, S.K.: Detection and removal of rain from videos. In: IEEE CVPR (2004)
7. Li, B., Wang, S., Zheng, J., Zheng, L.: Single image haze removal using content-adaptive dark channel and post enhancement. IET Comput. Vis. **8**(2), 131–140 (2014)
8. Peng, J., Liu, B., Dong, W., Wang, J., Wang, Y.: Method of image enhancement based on multi-scale retinex. Laser Infrared **38**(11), 1160–1163 (2008)
9. Nayar, S.K., Narasimhan, S.G.: Vision in bad weather. In: IEEE International Conference on Computer Vision (1999)
10. Huang, D., Huang, W., Gu, P., Liu, P., Luo, Y.: Image super-resolution reconstruction based on regularization technique and guided filter. Infrared Phys. Technol. **83**, 103–113 (2017)
11. Feng, X., Li, J., Hua, Z.: Low-light image enhancement algorithm based on an atmospheric physical model. Multimed. Tools Appl. **79**(43–44), 32973–32997 (2020). https://doi.org/10.1007/s11042-020-09562-6
12. Hubel, D.H., Weisel, T.N.: Receptive fields and functional architecture of monkey striate cortex. J. Physiol. **195**(1), 215–243 (1968)
13. Ripley, B.D.: Pattern Recognition and Neural Networks. Cambridge University Press, Cambridge (1996)
14. He, K., Sun, J., Tang, X.: Single image haze removal using dark channel prior. IEEE Trans. Pattern Anal. Mach. Intell. **33**(12), 2341–2353 (2011)
15. Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You only look once: unified, real-time object detection. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 779–788 (2016)
16. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. In: International Conference on Learning Representations (2015)
17. Rezatofighi, H., Tsoi, N., Gwak, J., Sadeghian, A., Reid, I., Savarese, S.: Generalized intersection over union: a metric and a loss for bounding box regression. In: IEEE/CVF CVPR, pp. 658–666 (2019)
18. Luo, Z., Nguyen, M., Yan, W.Q.: Sailboat detection based on automated search attention mechanism and deep learning models. In: International Conference on Image and Vision Computing New Zealand (IVCNZ), pp. 1–6 (2021)
19. Liu, Z., Yan, W.Q., Yang, M.L.: Image denoising based on a CNN model. In: International Conference on Control, Automation and Robotics (ICCAR), pp. 389–393 (2018)
20. Wang, X., Zhang, J., Yan, W.Q.: Gait recognition using multichannel convolution neural networks. Neural Comput. Appl. **32**(18), 14275–14285 (2019). https://doi.org/10.1007/s00521-019-04524-y

Youtube Engagement Analytics via Deep Multimodal Fusion Model

Minh-Vuong Nguyen-Thi[1,2], Huy Le[1,2], Truong Le[1,2], Tung Le[1,2], and Huy Tien Nguyen[1,2(✉)]

[1] Faculty of Information Technology, University of Science, Ho Chi Minh, Vietnam
{18120265,18120182,21c11038}@student.hcmus.edu.vn,
{lttung,ntienhuy}@fit.hcmus.edu.vn
[2] Vietnam National University, Ho Chi Minh City, Vietnam

Abstract. As popularity of video-sharing platforms, content creators have a high demand to produce content which attracts the large amount of viewers. There are many factors related to engagement: visual, sound, transcript, title etc. To take into account of these factors, we propose a deep multi-modal hybrid fusion for YouTube video engagement. Our architecture allows us to be easy to adapt state-of-the-art models for a particular task or variety of modalities, then fuse them to obtain more information aim to classify better. A proposed residual block as a simple neuron architecture search is used to get better features extracted. Our work is at the forefront of classifying YouTube video engagement and promises to broaden the research community's reach. Through detailed experiments, we proved that the model is the state-of-the-art in problem YouTube video engagement analytics.

Keywords: YouTube · Video Engagement · Multi-modal · Data Fusion · Architectural Search · Big Data · Data Mining · Social Media

1 Introduction

With the development of the Internet and social network, online communication through famous platforms such as Facebook, Twitter, Quora, TikTok, or YouTube become popular. These platforms provide clients with abilities to create and share information, news, and personal experiences with others every time, everywhere, even to become an official job. Brand owners choose them as a core place to build marketing campaigns for their products.

Established in 2005, YouTube [1] is a content-sharing platform made by users, and currently the biggest one in the world. Persons making shared clips are content creators, or also known as YouTubers. Viewers and YouTubers connect each other based on subscriber models. In particular, subscribers receive notifications of new content and updates from Youtubers. To earn more money, spread the videos to more people, and increase the number of subscribers, Youtubers need to maximize popularization of their videos' content and constantly improve the

H. Wang et al. (Eds.): PSIVT 2022, LNCS 13763, pp. 50–60, 2023.
https://doi.org/10.1007/978-3-031-26431-3_5