Image Translation of Vehicle Front Camera Frame Failures based on CycleGAN

[1]Dian Ning, [2]Dong Seog Han
School of Electronic and Electrical Engineering
Kyungpook National University
Daegu, Republic of Korea
[1]ningdian@knu.ac.kr, [2]dshan@knu.ac.kr

***Abstract*—More and more kinds of sensors are used including cameras in the vehicle to proactively address safety issues, either directly or indirectly. Camera failures, such as abnormal frames caused by muzzy, obstruction, and flutter, can lead to system exceptions and even traffic accidents because of their important role in the vehicle's system. We hope to reduce those exception faults by recovering abnormal frames. Therefore, in this paper, we first collect the video from the front-facing camera and define the abnormal frames. Then, this dataset is learned by a cycle generative adversarial network (CycleGAN) to generate more abnormal frames because sufficient samples are needed for better training. Moreover, CycleGAN can also restore the abnormal frames to normal frames, which reduces the system faults. This method can mitigate the consequence of camera failures and also works as a generator of corresponding failure frames.**

***Index Terms*—Fault Detection, GAN, Vehicle Inpainting, Camera Faulty**

I. Introduction

The autonomous industry has always been a pioneer in integrating various technologies, such as artificial intelligence, information communications, and so on. Autonomous driving has been gradually developing mature standards and systematic modules based on rich sensors and deep learning technologies [1]. The autonomous vehicle system is integrated with four core functional modules: application, perception, planning, and control. The application layer, as the basis for all other layers, includes a variety of sensors and outputs plenty of data to other layers. For example, camera sensors are placed at the front, rear, and even inside vehicles for different purposes. One or two cameras can be deployed in front of the car for different requirements [2].

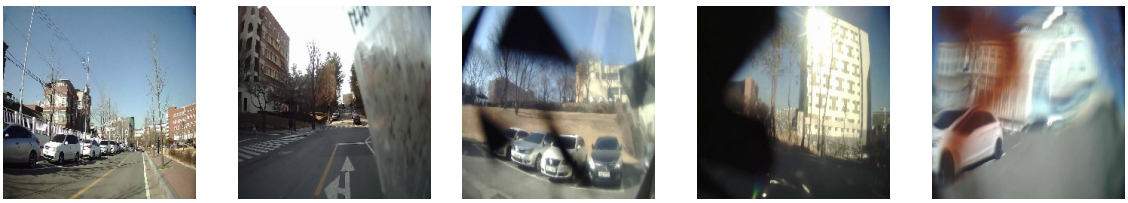
Front cameras have been used to detect inter-vehicle spacing for safety. The original algorithm focuses on traditional image features, such as color, shadows, and edges [3, 4]. As hardware and software are improving, machine learning methods have been introduced to camera applications, such as fog detection [5] and front vehicle detection [6]. Convolutional neural network (CNN) is one of the representative deep learning models used in vehicle cameras for lane detection [7], pedestrian detection, and even aggregating the data with other sensors to work on vehicle perception [8] and localization.

To keep safety, we should pay attention to potential malfunctions in the camera. The frame failure is defined as a frame that exhibits compromised output quality and therefore results in poor performance of vision-based algorithms using those frames. Frame failure can be produced by various factors, including jitter caused by camera instability, partial loss due to occlusion, and issues related to camera hardware integrity. Collecting failed camera frames is relatively difficult because of diversity, so we use cycle generative adversarial network (CycleGAN) [9] to establish a connection between normal and abnormal frames. CycleGAN is able to generate various abnormal frames, which make a more sufficient dataset. CycleGAN can link the source and target data in an unpair correspondence. We collect the normal and failed frames separately as source and target data. Experimental results indicate that image translation is effective for this dataset. To a considerable extent, the frames are reliably generated and reconstructed.

II. Dataset Collection and Finalization

A. Data Collection

There is no available public dataset for our purpose. However, a suitable dataset is necessary for training a model. We construct a new scale dataset using objects of different shapes, colors, light transmittance, and materials for practical cases. This dataset was taken using an RGB camera with



(a) (b)

Fig. 1. The dataset sample. (a) A normal frame. (b) The failure frames of different occlusion rates from 30% to 100%.

a resolution of 1920×1080 and a frame rate of 30 fps, mounted on the vehicle's windscreen. Dataset is collected in both the daytime environment of the kyungpook national university(KNU) campus, city centre, and motorways. Some of the videos are sourced from YouTube.

B. Data Finalization

This study divides the dataset into normal and failure classes. In the failure class, the camera is obstructed by objects of different sizes and materials placed in front of it, as shown in Fig. 1. We convert collected video to images as size of 256×256 at 1 fps. There are 8000 images per class with 3500 training and 600 testing images per class.

III. Proposed Model

A. Loss Function

We use CycleGAN for our unpaired dataset. CycleGAN is an adversarial network for situations without paired training data. In this model, there are two domains: a source domain X and a target domain Y. Between X and Y are two mapping generators G: X → Y and F: Y → X to translate images shown in the Fig. 2.

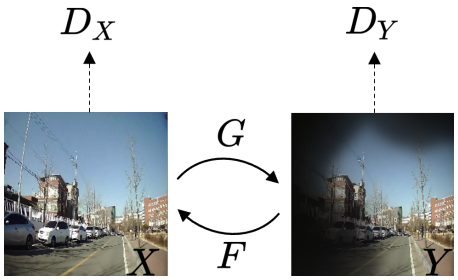


Fig. 2. The architecture of CycleGAN. The model has two mappings: G and F, between the source domain X and the target domain Y. Each domain has its own discriminator D.

In each mapping, the objective function is given by,

$$\begin{aligned} &L_G\left(G, D_X, X, Y\right) \\ &= E_{y \sim P_{data}\,(y)}\left[\log D_Y(y)\right] \\ &+E_{x \sim P_{data}\,(x)}\left[\log\left(1 - D_Y\left(G(x)\right)\right)\right], \end{aligned} \tag{1}$$

$$\begin{aligned} &L_G\left(F, D_Y, X, Y\right) \\ &= E_{x \sim P_{data}\,(x)}\left[\log D_X(x)\right] \\ &+E_{y \sim P_{data}\,(y)}\left[\log\left(1 - D_X\left(F(y)\right)\right)\right], \end{aligned} \tag{2}$$

where adversarial loss L_G is the L1-loss function that helps the Discriminator D distinguish fake images and the Generator G to improve its output.

To increase the relationship between the source and target domains in the unpaired dataset, we compare the original image and the restored image obtained through two mapping cycles for a comparison called cycle consistency loss. The sample of cycle consistency: x → G(x) → F(G(x)) as shown in Fig. 3.

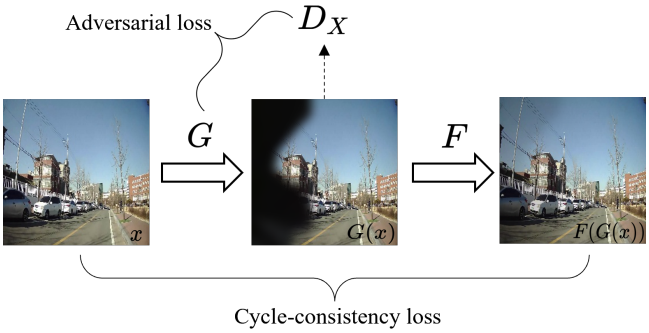


Fig. 3. The input normal image x, output failure images G(x) and the reconstructed images F(G(x)).

Cycle consistency loss $L_C\left(G, F\right)$ which is calculated as follows, as follows:

$$\begin{aligned} &L_C\left(G, F\right) \\ &= E_{x \sim P_{data}\,(x)}\left[\left\|F\left(G(x)\right) - x\right\|_1\right] \\ &+E_{y \sim P_{data}\,(y)}\left[\left\|G\left(F(y)\right) - x\right\|_1\right]. \end{aligned} \tag{3}$$

So the full loss function as below,

$$
\begin{aligned}
&L_{full}\left(G, F, D_X, D_Y\right) \\
&= L_G\left(G, D_X, X, Y\right) \\
&+L_G\left(F, D_Y, X, Y\right) \\
&+\lambda L_C\left(G, F\right),
\end{aligned} \tag{4}
$$

where λ is the weight of the cycle consistency loss.

B. Model Architecture

The CycleGAN training involves both a generator and a discriminator, as shown in the TABLE I and II. The generator is construed with Conv2D layers as an encoder, ResNet blocks [10] as a converter and the ConvTranspose layer as a decoder. The discriminator consisted of 5 Conv2D layers.

TABLE I: The generator structure.

Input : 256 × 256 × 3				
Layer	Filter	Stride	Channel	Activation
Conv	7 × 7	1 × 1	64	ReLu
Conv	3 × 3	2 × 2	128	ReLu
Conv	3 × 3	2 × 2	256	ReLu
9 × ResNet				
ConvTranspose	3 × 3	2 × 2	128	ReLu
ConvTranspose	3 × 3	2 × 2	64	ReLu
ConvTranspose	7 × 7	1 × 1	1	Tanh
Output : 256 × 256 × 3				

TABLE II: The discriminator structure.

Input : 256 × 256 × 3				
Layer	Filter	Stride	Channel	Activation
Conv	4 × 4	2 × 2	64	LeakyReLu
Conv	4 × 4	2 × 2	128	LeakyReLu
Conv	4 × 4	2 × 2	256	LeakyReLu
Conv	4 × 4	1 × 1	512	LeakyReLu
Conv	4 × 4	1 × 1	512	LeakyReLu
Output : 32 × 32 × 1				

IV. Experiments and Results

A. Experiments

The experimental data has shown in Section II-B. The generator and discriminator optimizers use the Adam algorithm with parameters β_1 of 0.5 and β_2 of 0.999. The decaying learning rate starts at 0.002 and decreases linearly. The training epoch is 50.

B. Results

The experimental results are shown in Fig. 4. By comparing the six sets of images, we can obtain that in the cycle consistency: $x \rightarrow$ G(x) $\rightarrow$ F(G(x)), the model learned the failure characteristics under high occlusion rates, but its ability to learn partial occlusion characteristics is relatively weak. In the reverse cycle: $y \rightarrow$ G(y) $\rightarrow$ G(F(y)), the obscured edges are easier to restore in the failed images; The unobstructed areas are also trained and merged with images from other driving scenes.

(a) (b) (c)

Fig. 4. Example results. (a) Input images. (b) Generated images with single mapping. (c) Reconstructed images with double mapping.

V. Conclusion

In this paper, we investigate the task of image transfer between normal and failure states in vehicle cameras. The goal is to generate a failed frame from a normal image and then restore the failed

part to a normal frame. We separately collect normal and failure frames as our source and target data. Our results show that the differences between normal and failure frames can be addressed using an image style transformer. We gain experience in video inpainting and restoration for vehicle cameras.

ACKNOWLEDGMENT

This research was supported by the Core Research Institute Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2021R1A6A1A03043144).

REFERENCES

[1] L. Liu, S. Lu, R. Zhong, B. Wu, Y. Yao, Q. Zhang, and W. Shi. Computing systems for autonomous driving: State-of-the-art and challenges, 2020.

[2] S. Indu, S. Srivastava, and V. Sharma. Optimal camera placement and orientation of a multi-camera system for self driving cars. In *Proceedings of the 2020 4th International Conference on Vision, Image and Signal Processing*, ICVISP 2020, New York, NY, USA, 2021. Association for Computing Machinery.

[3] Z. Sun, G. Bebis, and R. Miller. On-road vehicle detection: a review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(5):694–711, 2006.

[4] S. Tuohy, D. O'Cualain, E. Jones, and M. Glavin. Distance determination for an automobile environment using inverse perspective mapping in opencv. In *IET Irish Signals and Systems Conference (ISSC 2010)*, pages 100–105, 2010.

[5] M. Pavlić, H. Belzner, G. Rigoll, and S. Ilić. Image based fog detection in vehicles. In *2012 IEEE Intelligent Vehicles Symposium*, pages 1132–1137, 2012.

[6] S. Sivaraman and M. M. Trivedi. Real-time vehicle detection using parts at intersections. In *2012 15th International IEEE Conference on Intelligent Transportation Systems*, pages 1519–1524, 2012.

[7] J. Kim, Emeršič, and D. S. Han. Vehicle path prediction based on radar and vision sensor fusion for safe lane changing. In *2019 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*, pages 267–271, 2019.

[8] D. Ning and D. S. Han. Radar fault detection via camera-radar branches learning network. In *2023 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*, pages 463–467, 2023.

[9] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks, 2020.

[10] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition, 2015.