CycleGAN-based Image Transformation for Severe Weather Self-Driving

Bo Yang
yang.6113@buckeyemail.osu.edu

1 Motivation

Autonomous driving technology mainly uses lasers and cameras to detect the surrounding environment. Lidar provides depth and distance information from an object to a vehicle by creating a dot cloud of the environment. The camera handles all perceptual tasks such as detecting traffic signs and signals, and detecting track lines. Stereo cameras can also extract information from depth. Both of these sensors work well under normal driving conditions. However, under adverse weather conditions such as rain, fog and snow, the performance of the laser radar is significantly reduced and affected by a phenomenon called scattering, as it works to reflect the laser beam in the surrounding environment.

Autonomous vehicle technology is currently moving from L4 to L5. One of the main goals for this breakthrough is to improve the accuracy of the detection of autonomous vehicles in adverse weather conditions, and to enable safe navigation. At the same time, the continuous progress of sensor technology is very important to improve road safety. Data from these sensors can warn of potential safety risks and take action to prevent accidents. Bad weather causes about 1.2 million traffic accidents each year in the United States [1]. Bad weather accidents account for one-fifth of all accidents in the United States. Therefore, it is reasonable to solve the problem that bad weather degrades the performance of the detector.

The main goal of current autonomous companies is to perform sensing operations with as few sensors as possible without compromising safety, such as autonomous driving giant Tesla relying only on cameras to observe its surroundings. Since companies like Tesla and others use only cameras (stereo) and radar to generate higher resolution data, these systems are cost effective compared to other autonomous driving companies that use expensive lidar. As a result, major companies have invested a lot of resources into developing computer vision-based methods to overcome the problem of camera performance degradation in bad weather. One of the challenges of current research is that there are no paired datasets. It is difficult to collect both severe weather images and clear images in the same driving environment. Therefore, it is difficult to solve the problem of image transformation based on supervised learning methods. In this paper, a CycleGAN method based on deep learning is proposed to reduce the influence of adverse weather conditions on the image acquired by the on-board camera. We use the unpaired dataset to train the model for unsupervised learning.

2 Related Work

2.1 Severe weather self-driving related works

Zang et al. [2] mentioned that intensity fluctuations introduced by rain and snow in images captured by autonomous driving cameras blur the edges of various objects in the images. Typically, rain decreases the image intensity, while heavy snow increases it. Foggy conditions reduce the image contrast and increase the difficulty of object edge recognition. Kodieswari et al. [3] proposed three image restoration methods based on algebraic methods. The inverse filter can be used when we fully understand the blur function that causes image degradation. If we have only partial knowledge of the blur function, the Weiner filter is used. If we do not have any prior knowledge of the blur function, a blind recovery filter is used. Image restoration based on algebraic methods has limitations in terms of accuracy and cannot be used in dynamic scenes with changing weather conditions. The AI-based image restoration methods have achieved better results in terms of the quality and accuracy of the recovered images. Eigen et al. [4] proposed common convolutional neural networks for removing dirt or rain. Wang et al. [5] proposed DehazeNet, a deep network for removing the impact of heavy rain weather. Park et al. [6] proposed SPANet, a spatial attention network that removes the effect of rain.

2.2 CycleGAN application in autonomous driving

Ostankovich et al. [7] applied CycleGAN to nighttime autonomous driving scenario. In their study, the perception module was defined as a combination of object detection and road segmentation. The results showed that the road segmentation task was improved after using CycleGAN augmentation. Qu et al. [8] used CycleGAN for low-light enhancement of object detection in autonomous driving applications. In their work, they tested the proposed method using the Oxford Robotic Car dataset. They found that the proposed method was able to significantly improve the detection accuracy in low-light environments and increase the number of detected targets. Wang and Chuan Tan [9] used the HSI preprocessing method and the CycleGAN method to improve the visual object detection of mobile robots under dynamic illumination changes. They found that the effect of target detection can be easily affected by external environmental factors such as uneven illumination and target occlusion. Compared to the HSI preprocessing method, the brightness transfer method based on CycleGAN was able to achieve better detection results in robot vision.

3 Model Design

3.1 CycleGAN for unpaired image transformation

Paired images refer to a set of images that have a one-to-one correspondence with each other. In other words, each image in the set has a corresponding image that shows the same scene or object from a different perspective, viewpoint, or condition. Paired images are often used in image-to-image translation tasks, where a model is trained to convert an image from one domain into an image in another domain. Unpaired images, on the other hand, are images that do not have a corresponding image in the other domain. In other words, they are not part of a set of paired images. Unpaired images are often used in unsupervised learning tasks, where a model is trained to learn the underlying structure of the data without the need for explicit labels or paired examples.

In the context of self-driving cars, it is often difficult or impractical to collect paired datasets of images that show the same scene or object under different conditions. For example, it may be difficult to find two images of the same road scene that were taken at different times of day, or in different weather conditions. In these cases, it may be more practical to use an unpaired dataset, where the images are not part of a set of paired images. This allows the model to learn the underlying structure of the data without the need for explicit labels or paired examples. Additionally, using unpaired images may also make it possible to train the model on a larger dataset, since it is not necessary to find corresponding images for each example.

Zhu et al. [10] proposed that for unpaired images, CycleGAN is the first choice. CycleGAN is a type of generative model that is specifically designed to handle unpaired image-to-image translation tasks. In other words, it is able to transform one type of image into another without the need for a corresponding image in the other domain. This makes it well-suited for tasks where it is difficult or impractical to collect paired datasets of images. Additionally, CycleGAN uses a unique training approach that involves two generator networks and two discriminator networks, which helps to ensure that the generated images are realistic and of high quality. This makes it a powerful tool for solving a wide range of unpaired image-to-image translation tasks.

3.2 CycleGAN structure

The overall structure of a CycleGAN model consists of four main components: the generator networks, the discriminator networks, the reconstruction loss function, and the adversarial loss function. The generator networks are responsible for converting images from one domain to another, and are trained to produce high-quality images that are difficult for the discriminator networks to distinguish from real images. The discriminator networks, on the other hand, are trained to differentiate between real and generated images. The reconstruction loss function helps to ensure that the generated images are similar to the input images, while the adversarial loss function helps to ensure that the generated images are realistic and of high quality.

Overall, the structure of a CycleGAN model is designed to allow it to learn complex image-to-image translation tasks without the need for large amounts of labeled data. This makes it a powerful tool for solving a wide range of problems in computer vision and image processing.

We have a 70 x 70 patch-GAN as the discriminator. A PatchGAN is a type of convolutional neural network that is commonly used for image-to-image translation tasks, such as generating high-resolution images from low-resolution inputs. It works by dividing the input image into a grid of patches and then predicting the likelihood that each patch belongs to a certain class. By doing this, the PatchGAN is able to capture both local and global features of the input image, allowing it to generate high-quality images that are faithful to the original input. The discriminator uses IN to represent instance normalization per layer. From top to bottom, our discriminator models are: Image Input, CNN 64, CNN 128 IN, CNN 256 IN, CNN 512 IN, Patch Output. The generator model is divided into downscaling layer and upscaling layer, the downscaling layer is denoted by uN, and the upscaling layer is denoted by dN. The most important block in the generator model is the Resnet block, which mainly aims to overcome the gradient vanishing problem in deep CNN. The connections of Resnet are hopping and the information from the top is concatenated with the output. Our generator models from top to bottom are: dN128 IN, dN256 IN, Resnet 256, uN128 IN, uN64 IN.

4 Evaluation

4.1 Dataset selection

We chose the radiant dataset Sheeny et al.[11]. This dataset does not suffer from feature similarity issues. The feature similarity problem will have an irreducible noise effect on the final effect of GAN. We use two types of data; 1) Ground truth data, basically automatic driving data without adverse conditions and noise. This data is highly accurate and well characterized. 2) Unfavorable data Unfavorable and noisy features for autonomous driving conditions.

4.2 Results

We used different resolution images: 128 x 128, 256 x 256 and 512 x 512. The three outputs are very similar. In all three configurations, the network is able to filter out the water drop noise in the image with rain, but the network also filters out vehicles from the image. In the configuration with an image resolution of 512, the model is relatively faster. So we shape the input image as 512 x 512.

Tuning the generator factor involves increasing and decreasing the number of layers of the current layer and the filter size of each layer. The filter sizes we use are :3 x 3, 5 x 5 and 7 x 7. We set the filter size to 5 x 5 because there are no more than two vehicles in the graph, and this setting can stop filtering out vehicles.

We tried to fine-tune the discriminator by increasing the number of convolutional layers and the results showed significant improvement. The model is able to distinguish between droplets and vehicles. The model also stopped masking the vehicle with droplets, as shown in Figure 1.

Figure 1: Model filter out the almost water drop noise

5 Issues

One of the main issues with CycleGAN is that it can sometimes produce distorted or unrealistic images, especially if the input images are very different from the target domain. This can be caused by a number of factors, including the quality of the training data, the complexity of the image-to-image translation task, and the ability of the model to learn the underlying structure of the data.

Additionally, CycleGAN can also be sensitive to hyperparameters, such as the learning rate and the number of training iterations, which can affect the performance of the model.

Figure 2: Model produces distorted image

Another issue with CycleGAN is that it can sometimes produce images with visible artifacts, such as blurriness or noise. This can be caused by the loss functions used during training, which may not always be able to adequately capture the structure of the data. This can result in generated images that do not accurately reflect the input images, and can lead to poor performance on downstream tasks.

Figure 3: Model produces artifacts like trees

6 Conclusion

Our main contribution is to apply CycleGAN to the problem of image translation for autonomous driving in bad weather and obtain remarkable results. While the previous image translation problem solving for autonomous driving in bad weather mainly uses paired image datasets, we accomplish this task using unpaired image datasets. The model learns stable features using smaller pairs of data, e.g. a cloudy sky with dark clouds becomes a clear sky after recovery. The quality of the generated images is mainly affected by the size and location of the droplets, with larger droplets leading to failure in image restoration. The network does not completely remove droplets if they cover the vehicle. To optimize this network, we need more different datasets of different vehicles in the future so that the generator generates vehicles that are completely occluded by droplets.

References

[1] How Do Weather Events Impact Roads? https://ops.fhwa.dot.gov/weather/q1_roadimpact.htm#:~:text=On%20average%2C%20there%20are%20over,1%2C235%2C000%20%2D%20are%20weather%2Drelated.

[2] Shizhe Zang, Ming Ding, David Smith, Paul Tyler, Thierry Rakotoarivelo, and Mohamed Ali Kaafar. The impact of adverse weather conditions on autonomous vehicles: How rain, snow, fog, and hail affect the performance of a self-driving car. *IEEE Vehicular Technology Magazine*, 14(2):103–111, 2019.

[3] A. Kodieswari, V. Parameshwari, and S. Sruthi. Reduction of rain and snow within the image using image processing. 2021.

[4] David Eigen, Dilip Krishnan, and Rob Fergus. Restoring an image taken through a window covered with dirt or rain. *2013 IEEE International Conference on Computer Vision*, pages 633–640, 2013.

[5] Keping Wang, Yumeng Duan, and Yi Yang. Single image dehazing algorithm based on pyramid mutil-scale transposed convolutional network. *Systems Science & Control Engineering*, 9(sup1):150–160, 2021.

[6] Yeachan Park, Myeongho Jeon, Junho Lee, and Myungjoo Kang. Mcw-net: Single image deraining with multi-level connections and wide regional non-local blocks. *Signal Processing: Image Communication*, 105:116701, 2022.

[7] Vladislav Ostankovich, Rauf Yagfarov, Maksim Rassabin, and Salimzhan Gafurov. Application of cyclegan-based augmentation for autonomous driving at night. In 2020 International Conference Nonlinearity, Information and Robotics (NIR), pages 1–5, 2020.

[8] Yangyang Qu, Yongsheng Ou, and Rong Xiong. Low illumination enhancement for object detection in self- driving. In *2019 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, pages 1738–1743, 2019.

[9] Feng Wang and Jeffrey Too Chuan Tan. Improving deep learning based object detection of mobile robot vision by hsi preprocessing method and cyclegan method under inconsistent illumination conditions in real environment. In *2019 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM)*, pages 583–588, 2019.

[10] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A. Efros. Unpaired image-to-image translation using cycle- consistent adversarial networks, 2017.

[11] Marcel Sheeny, Emanuele De Pellegrin, Saptarshi Mukherjee, Alireza Ahrabian, Sen Wang, and Andrew Wallace. Radiate: A radar dataset for automotive perception in bad weather, 2020.