Parameterized Defogging Network for Object Detection in Adverse Weather Conditions

Rishie Raj
Maryland Applied Graduate Engineering
University of Maryland
College Park, MD 20740
rraj27@umd.edu

Uthappa Madettira
Maryland Applied Graduate Engineering
University of Maryland
College Park, MD 20740
uthu@umd.edu

Nathan Nussbaumer
Department of Computer Science
University of Maryland
College Park, MD 20740
nussbau@terpmail.umd.edu

Sahaj Singh
Department of Computer Science
University of Maryland
College Park, MD 20740
sahaj10@terpmail.umd.edu

Lucas Leitao
A. James Clark School of Engineering
University of Maryland
College Park, MD 20740
luke.leitao@gmail.com

Abstract

Efficient and accurate object detection in adverse weather conditions pose a significant challenge in autonomous driving. This paper introduces a small convolutional neural network model designed to predict the parameters of differentiable image processing functions with the aim to defog the input images. By transforming obscured images into clear visuals, our model facilitates more robust object detection. We trained and evaluated our model using a dataset of foggy images and obtained the output parameters for the differentiable image processing function (DIP). Our results indicate improvement in detection speeds and significantly improved results on the foggy images. The source code can be found at `https://github.com/nussbau/CMSC472_Final`

1 Motivation

Autonomous driving is an active field of research that could significantly reduce the amount of traffic accidents that occur. Autonomous driving systems rely heavily on accurate object detection mechanisms to navigate safely. However adverse weather conditions such as rain, fog significantly impact the object detection capabilities of the autonomous driving systems and pose a serious risk to the passengers. The image captured in an adverse weather condition can be decomposed into a clean image and its corresponding weather specific information. The image if properly enhanced can reveal more latent information and the misidentified objects can be recovered.

There is a need for an innovative solution that can swiftly enhance the images captured during adverse weather conditions without sacrificing details. Addressing this, we introduce a small convolutional neural network model that predict the parameters of differentiable image processing functions. This model, specifically designed for applications in autonomous driving provide good results on foggy images and also has a good detection speed when used with the original YOLO. By enabling clear vision during adverse weather conditions we aim to make autonomous vehicles safer and more reliable in a wider range of environments.

2 Background

2.1 Object detection

Object detection plays a crucial role in autonomous driving systems, enabling the vehicles to detect and classify its surroundings. Two main strategies are commonly employed for object detection, namely, the two-stage and single-stage approaches. Two-stage approaches, such as the region-based convolutional neural network (RCNN) and the Faster R-CNN, are widely adopted due to their high accuracy. However, they often fall short of meeting the real-time processing requirements of object detection in vehicular environments.

On the other hand, single-stage approaches, exemplified by you only look once (YOLO) and the single shot detector (SSD), are renowned for their speed. These methods directly predict object classes and bounding box coordinates from a single evaluation of the network. They achieve real-time performance but may sacrifice some accuracy in certain scenarios. Despite their advantages, detecting objects under challenging driving conditions remains a significant challenge for these single-stage approaches

2.2 Object detection in adverse weather conditions

Research on object detection under adverse weather conditions has been relatively limited compared to general object detection. A common approach involves preprocessing the image using classical dehazing technique. Although this method improve image quality, they do not necessarily enhance detection performance.

To address this, image enhancement and detection were combined to minimize the impact of weather-related distortions. Furthermore, recognizing the domain shift between images taken in normal and adverse conditions, some studies have adopted domain adaptation strategies. These methods adopt the domain adaptation principles and focus on aligning the features of two distributions, and the latent information which can be obtained in the process of weather-based image restoration is usually ignored.

To address these limitations, we have implemented the Image-Adaptive YOLO for Object Detection in adverse weather conditions.

2.3 Our approach

(a) Before passing through our defogger

(b) After passing through our defogger

Figure 1: For the images captured in adverse weather conditions, our defogger model outputs clear images aiding improved object detection

To address the issues of object detection in adverse weather conditions, we implement a detection framework that adapts to varying weather conditions by filtering out weather-specific interference and uncovering hidden details. The model integrates three main components: a CNN-based parameter predictor (CNN-PP), a differentiable image processing module (DIP), and a detection network. The CNN-PP is trained on 367 images, out of which 2/3rds are foggy images. The input images are resized to 256 × 256 and processed through the CNN-PP, which estimates the parameters for the DIP. CNN-PP output gives parameters for classical differentiable defogging, white balancing, gamma correction, toning, contrasting, and sharpening methods. The image, once processed by the DIP module, serves as the input for the YOLO detector. A hybrid end-to-end training approach is incorporated, which enables the CNN-PP to optimally adjust the DIP parameters for improved image clarity in object detection. This method allows for effective image enhancement and improved object detection for images captured in adverse weather conditions.

3 Model Architecture

Our model integrates three components to achieve the task of object detection in adverse weather conditions. They are

- CNN-based parameter predictor (CNN-PP).
- Differentiable image processing module (DIP).
- Detection network module

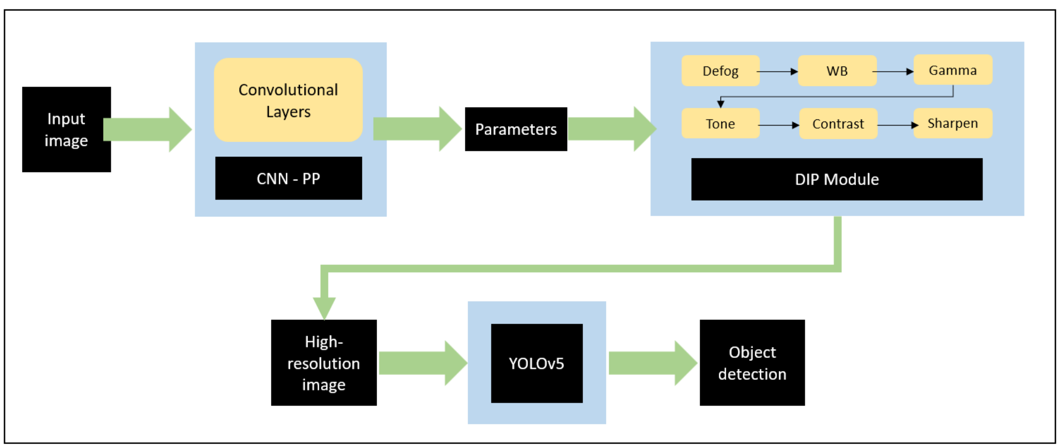


Figure 2: Architecture of the integrated model for object detection in adverse weather conditions. The model integrates a CNN-based Parameter Predictor, a Differentiable Image Processing module, and a YOLOv5 detection network to enhance detection capabilities under challenging conditions.

3.1 CNN based parameter predictor

The CNN-PP (Convolutional Neural Network for Parameter Prediction) module is a crucial component of our system, designed to efficiently extract meaningful features from input images. This module leverages convolutional neural networks (CNNs) due to their proven capability in handling image data and extracting hierarchical features.

3.1.1 Design

The CNN-PP module consists of a series of convolutional blocks, each followed by a non-linear activation function, specifically Leaky ReLU, to introduce non-linearity into the model, thus allowing it to learn more complex patterns in the data.

3.1.2 Layers

The module initiates with a convolutional layer with a kernel size of 3x3, stride of 2, and padding of 1. This layer maps the 3-channel input image to a 16-channel feature map, effectively doubling the depth while reducing the spatial dimensions by half, considering the stride. Subsequent layers follow a similar pattern, gradually increasing the channels from 16 to 32 while maintaining the kernel size and padding, thus focusing on extracting finer details as the network deepens. The use of Leaky ReLU with a negative slope of 0.01 ensures that the network retains gradient flow during training, which can often vanish in deep networks using standard ReLU. Following the convolutional layers, the feature maps are flattened into a vector and passed through fully connected layers. The first dense layer transforms the flattened features to a 128-dimensional space, followed by a Leaky ReLU to maintain non-linearity. The final output layer consists of a linear transformation to the desired number of outputs, which in this context is set to 16, representing specific parameters required for the subsequent DIP (Deep Image Processing) module.

3.2 Differentiable Image Processing Module

The hyperparameters predicted by the CNN-PP module are then passed to the DIP module. The DIP module consists of six differentiable filters with adjustable hyperparameters, including Defog, White Balance(WB), Gamma, Contrast, Tone and Sharpen.

3.2.1 Image Defogging

This filter attempts to remove fog from an image by estimating the scene's transmission and atmospheric light. The atmospheric light is estimated by finding the brightest pixels in the image. The transmission map is estimated and refined to recover the clear image from the foggy input image.

3.2.2 Image white balance adjustment

This filter adjusts the white balance of the image to correct color cast of images taken under different light sources. The filter scales each color channel by a factor, which is derived from the learned parameters. **Additionally we have added a *Sigmoid* function to squeeze all values for white balance between 0 & 1. This is so that no value goes to 0 and a gradient is maintained everywhere.**

3.2.3 Image gamma correction

This filter adjusts the gamma values of the image which are useful in correcting the exposure of the image. It applies a power-law transformation to each pixel of the image and adjusts the brightness according to the gamma parameter predicted by the CNN-PP. **In this filter, we have added a *Relu* function so that there are no negative values whose exponents are evaluated.**

3.2.4 Image tone mapping

This filter maps the tones of the high dynamic range image to the displayable range, maintaining the details in both bright and dark areas. The tone mapping function adjusts the intensity of the image based on the tone parameters, effectively compressing the dynamic range of the image.

3.2.5 Image contrast enhancement

This filter enhances the contrast of the image by scaling the intensity of the pixels. The image contrast is adjusted by scaling the pixel values with a factor alpha and then adding a bias beta, both of which are learned parameters from the CNN-PP.

3.2.6 Image sharpening

This filter enhances the edges and details of the image by applying a high-pass filter. A Gaussian blur is applied to the image first, and the difference between the original and blurred image is scaled by a factor lambda and added back to the original image to enhance sharpness.

3.3 Detection Network Module

The detection network incorporated is YOLOv5, a one-stage detector commonly used in the field of autonomous driving and other real time applications. The architecture of YOLOv5 is modular, consisting of a backbone, a neck, and a head, which are optimized for speed and accuracy.

The backbone of YOLOv5 is based on the CSPDarknet53 architecture. This backbone is enhanced with Cross Stage Partial networks (CSP), which help reduce the model's complexity and improve its computational efficiency. The neck of the architecture employs the PANet (Path Aggregation Network) design. This component aggregates the different levels of feature information, ensuring that the high-resolution context from early layers can be combined with the deeper, more semantic features from later layers. The head of the YOLOv5 model makes the final predictions. It processes the aggregated feature maps to predict bounding boxes, object classes, and prediction scores.

4 Implementation

Our implementation of the Image-Adaptive YOLO model involves the following:

- Building the CNN-PP and DIP modules from scratch.
- Using the YOLOv5 architecture from Ultralytics to perform Object Detection.
- Training the entire model on a custom dataset containing Foggy + Underwater + Coco images.

We have taken a variety of images in our dataset so that the model can be trained to perform detection is a wide range of environments.

5 Results

To gauge the effectiveness of our model, the outputs are compared with that of the original YOLOv5. We can observe significantly improved outcomes on foggy and underwater images when utilizing YOLO trained with the DIP module. These results can be seen in Figure 3.

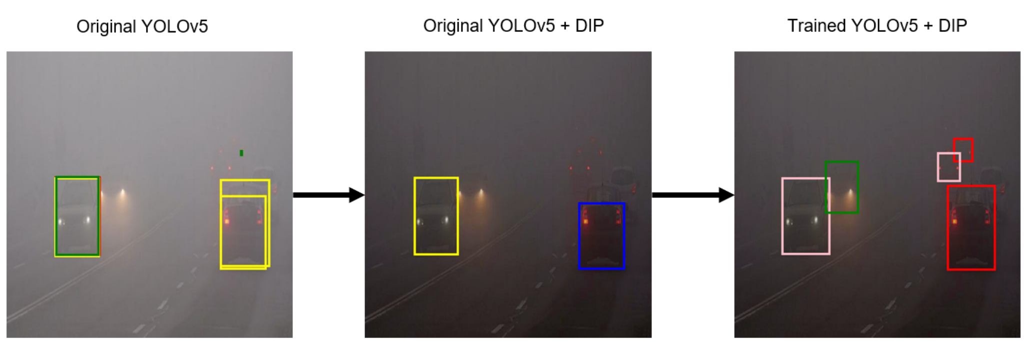
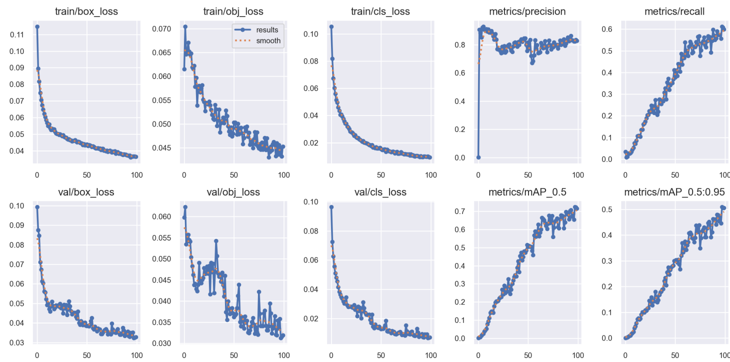


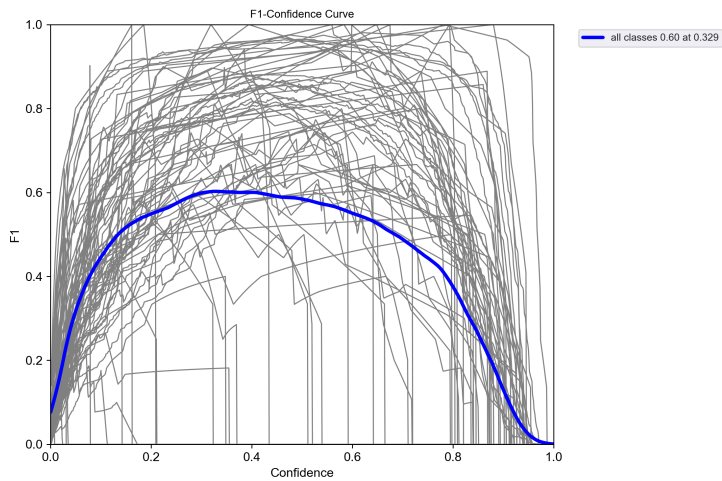
Figure 3: The above image shows the difference between the predictions of the different models. It can be seen that the YOLO which is trained on our custom dataset along with the DIP module predicts larger number of objects.

The training of our model was also quite stable, with losses reducing as expected. We also got high values on the performance metrics of the model such as mean Average Precision and F1 score. The original YOLO gave a *mAP50* of 0.98 and *mAP50-95* of 0.8.

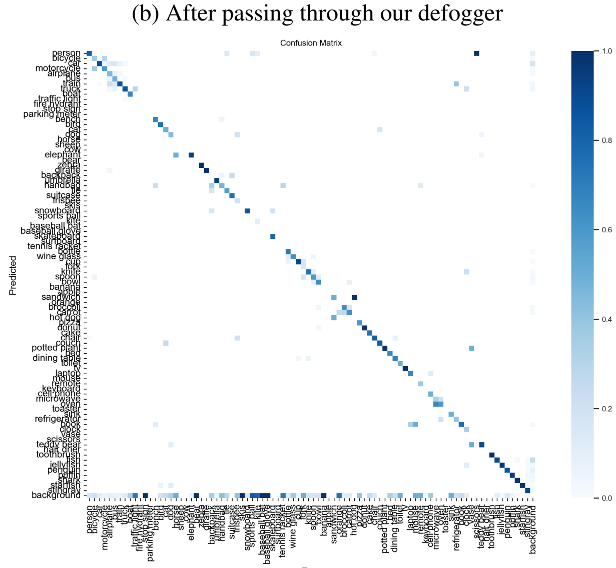
The metrics were slightly reduced when compared with training only on foggy data because the underwater dataset we used had a lot of light reflection. Due to this, the detection performance was hindered. However, the model has managed to train fairly smoothly. As it can be seen below, the *mAP50* has reduced to 0.738 and the *mAP50-95* to 0.5.



(a) Before passing through our defogger



(b) After passing through our defogger



(c) After passing through our defogger

Figure 4: For the images captured in adverse weather conditions, our defogger model outputs clear images aiding improved object detection

6 Conclusion

We have successfully implemented the Image-Adaptive YOLO architecture as proposed by *Liu et al. (2022)* and trained it on a wide range of images which include foggy, underwater and Coco-style images. The model achieved a mean Average Precision score of approximately 0.7. It was able to detect almost 50% of the classes with high accuracy. The outputs of the DIP module showed that our implementation was able to correctly defog the images based on the parameters learnt by the CNN-PP module during training.

7 References

[1] Liu, W. & Ren, Gaofeng (2022) Image-Adaptive YOLO for Object Detection in Adverse Weather Conditions *In: arXiv:2112.08088*.

[2] Liu, W., Wonteng, L., Zhu, J & Zhang, L. (2023) Improving Nighttime Driving Scene Segmentation via Dual Image-adaptive Learnable Filters *In: arXiv:2207.01331*.

[3] He, K., Sun, J. &Tang, X. (2009) Single image haze removal using dark channel prior. *In: Proceedings of IEEE/CVF Conference Computer Vision Pattern Recognition (CVPR)*.