2024 IEEE 27th International Conference on
Intelligent Transportation Systems (ITSC)
September 24- 27, 2024. Edmonton, Canada

RAF-RCNN: Adaptive Feature Transfer from Clear to Rainy Conditions for Improved Object Detection

Chuheng Wei, *Student Member, IEEE*, Guoyuan Wu, *Senior Member, IEEE*, Matthew J. Barth, *Fellow, IEEE*,

***Abstract*— In the challenging realm of object detection un- der rainy conditions, visual distortions significantly hinder accuracy. This paper introduces Rain-Adapt Faster RCNN (RAF-RCNN), an innovative end-to-end approach that merges advanced deraining techniques with robust object detection. Our method integrates rain removal and object detection into a single process, using a novel feature transfer learning approach for enhanced robustness. By employing the Extended Area Structural Discrepancy Loss (EASDL), RAF-RCNN enhances feature map evaluation, leading to significant performance improvements. In quantitative testing of the Rainy KITTI dataset, RAF-RCNN achieves a mean Average Precision (mAP) of 51.4% at IOU [0.5, 0.95], exceeding previous methods by at least 5.5%. These results demonstrate RAF-RCNN's potential to significantly enhance perception systems in intelligent transportation, promising substantial improvements in reliability and safety for autonomous vehicles operating in varied weather conditions.**

I. INTRODUCTION

Object detection is crucial in the development of autonomous driving technologies, a key focus area in intelligent transportation systems [1], [2]. However, the reliability of these systems is often compromised under adverse weather conditions, such as rain, which is a critical challenge for autonomous vehicles [3]. This paper introduces Rain Adapt Faster R-CNN *(RAF-RCNN)*, a groundbreaking approach designed specifically for intelligent driving systems, that adapts features from clear to rainy conditions, enhancing the robustness and reliability of autonomous vehicle perception in challenging environments.

By integrating the deraining process into the object detection pipeline, the RAF-RCNN model creates an end-to-end solution that is seamless and end-to-end. In addition to retaining the essential features, this innovative approach bypasses the traditional two-stage rainy day object detection process, thereby increasing the efficiency of detection, as shown in Fig. 1-(c). Moreover, by leveraging feature corrective transfer learning [4], RAF-RCNN adeptly adjusts to the variances of rainy conditions, using real-world affected datasets for training. This enables precise detection and identification under severe rain, where visibility is notably reduced.

The current technologies for object detection in rainy conditions can be broadly categorized into two approaches. The first approach is a two-stage method, as illustrated

Chuheng Wei, Dr. Guoyuan Wu, and Dr. Matthew J. Barth are with the College of Engineering, Center for Environmental Research and Technology, University of California at Riverside, Riverside, CA, 92507. (Corresponding author: Chuheng Wei, e-mail: chuheng.wei@email.ucr.edu)

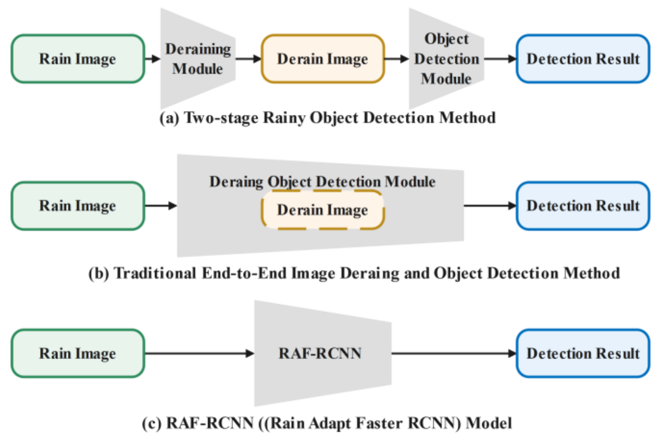

Fig. 1: Pipeline for Previous Deraining Object Detection Methods and RAF-RCNN

in Fig. 1 -(a). This method initially employs a deraining module to process rainy images, producing relatively rain-free images. Subsequently, an object detection module is used to recognize objects in the derained images. The second approach, which has been the focus of recent research efforts [5], is an end-to-end deraining module as shown in Fig. 1 -(b). This method utilizes a single neural network to simultaneously achieve deraining and object detection in one training process. However, even in this approach, the derained images are considered an interim outcome of the model and are evaluated using a loss function to assess the deraining effectiveness.

Nevertheless, it's important to question the necessity of derained images in the neural network's process for rain-affected object detection. It might be feasible and potentially more efficient to rely solely on the neural network's capacity to interpret and detect objects directly from rainy images [6], [7]. This approach challenges the conventional methodology and suggests a paradigm shift, leveraging the advanced learning capabilities of neural networks to bypass the deraining step entirely.

A. Research Significance and Potential Impacts

A key assumption of this research is that perfect visual clarity, similar to human vision, is not necessary to detect objects effectively. Convolutional Neural Networks (CNNs), the backbone of modern object detection frameworks, are the basis for this argument. Unlike human vision, deep neural

2024 IEEE 27th International Conference on Intelligent Transportation Systems (ITSC) | 979-8-3315-0592-9/24/$31.00 ©2024 IEEE | DOI: 10.1109/ITSC58415.2024.10920096

networks possess the unique ability to recognize relevant features from images, even if the effects of these images differ somewhat from those observed by the human eye [8]. It has been demonstrated that neural networks do not require neural networks to produce high-quality, human- friendly images for visual tasks, especially object detection tasks [8], [9].

In contrast to the traditional paradigm, this approach suggests that object detection might not require the pursuit of derained images as a prerequisite. By bypassing the intermediary step of rain removal, RAF-RCNN is intended to achieve favorable detection results directly from rainy images.

Furthermore, the application of transfer learning in RAF-RCNN aids in aligning features between rainy and clear weather conditions, enhancing the model's ability to learn from and adapt to these different scenarios. This feature alignment is critical in ensuring the robustness and accuracy of the model in diverse environmental conditions.

B. Contributions

With the complexity of object detection in rainy conditions, the development of RAF-RCNN is an important advancement. Moving forward, we delineate the principal contributions of RAF-RCNN as follows:

- **End-to-End Rainy Object Detection Models:** We present RAF-RCNN, the first model of its kind that disregards the human visual perception aspect in deraining. Unlike traditional methods that evaluate deraining effectiveness at the image level, our model integrates the deraining process within the object detection framework. This integration not only simplifies training complexity but also significantly enhances detection performance in rainy conditions.
- **Feature Corrective Transfer Learning in Rainy Object Detection:** We incorporate feature corrective transfer learning technique [4] specifically tailored for rainy condition object detection. By utilizing sunny realistic datasets, we facilitate the adaptation of the model at the feature map level, enhancing its ability to detect objects in rainy conditions with enhanced accuracy.
- **Novel Loss Function - EASDL:** We introduce a novel loss function, the Extended Area Structural Discrepancy Loss (EASDL), designed to evaluate and optimize feature map alignment in the context of transfer learning. This loss function plays a crucial role in fine-tuning the model's ability to distinguish and detect objects amidst rain-distorted images challenges.

II. Related Work

A. Single Image Deraining

Deep Neural Networks for Deraining: The field of single image deraining has benefited significantly from deep learning technologies. Fu et al. [10] introduced DerainNet, a novel approach that utilizes a guide filter to decompose an image into a base and a detail layer, where the detail layer is processed through a residual network to remove rain effects. This approach effectively clears the image of rain. Additionally, Qian et al. [11] delved into the potential of Generative Adversarial Networks (GANs), augmented with visual attention mechanisms. As a result, DeRaindrop was developed, a specialized GAN designed specifically for the detection and elimination of raindrops. Fu et al. [12] further developed the Deep Detail Network (DDN), which uses CNNs to accurately predict and reduce discrepancies between rainy and clear images.

Progressive Deraining Networks: Progressive deraining methods have revolutionized the approach to removing rain from images by introducing more dynamic and effective techniques. [13], who developed PReNet. This innovative approach entails the recursive removal of rain from a single image. This technique progressively derains the original image through iterations, each improving the rain removal, and ultimately results in a high-quality rain-free image. Building on these concepts, Wang et al. [14] introduced the Rain Convolution Dictionary Network (RCDNet), which autonomously identifies and processes rain elements, thus improving the deraining quality. Zamir et al. [15] further contributed by proposing MPRNet, a multi-stage architecture that simplifies the restoration process and improves the learning of restoration functions incrementally.

B. Object Detection in Rainy Conditions

Two-stage Object Detection Approaches in Rainy Condition: In deep learning, two-stage object detection methods have been developed to handle rainy and adverse weather conditions. These methods typically involve an initial stage of image restoration followed by object detection. Huang et al. [16] pioneered the development of a dual-subnet network (DSNet), which bifurcates its mechanism into two separate subnetworks dedicated to image recovery and object detection, respectively. Further augmenting this field, they introduced the Selective Features Absorption Network (SFA-Net) [17]. The SFA-Net consists of three integral parts: a feature selection subnetwork, an object detection subnetwork, and a feature absorption subnetwork. It is especially beneficial in rainy conditions due to its enhanced detection capabilities.

In a similar vein, on tasks, enhancing training data with synthetic rainfall variations to improve object detection robustness. Sindagi et al. [18] presented a domain adversarial object detection framework that uses pre-adversarial training to address various weather conditions. Moreover, Appiah and Mensah [19] enhanced images with the ESRGAN algorithm before detection using the YoloV7 algorithm to cope with rain impacts effectively.

End-to-End Object Detection Approaches in Rainy Conditions: Additionally, recent advancements include end-to-end object detection systems that integrate deraining and object detection into a single neural network. Wang et al. [20] developed an end-to-end network that integrates deraining and object detection into cascading modules, treating the tasks as distinct yet sequential stages within one framework. This approach starts with deraining before object detection,

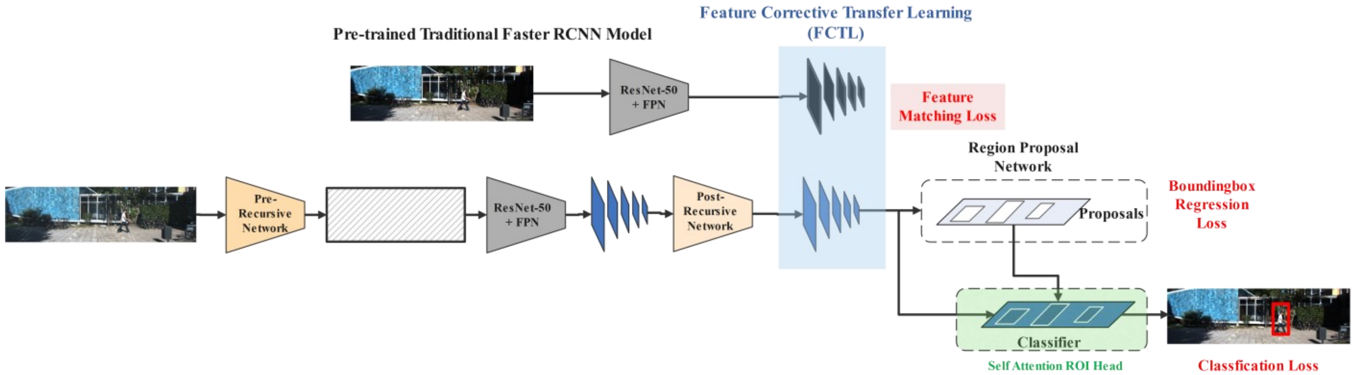

Fig. 2: Architectures of RAF-RCNN (Rain-Adapt Faster RCNN)

maintaining separate focuses within a unified system. Liu et al. [3] introduced the Image-Adaptive YOLO (IA-YOLO) which combines a small CNN with YOLOv3 and a differentiable image processing (DIP) module. This configuration preprocesses rain-distorted images for detection, still modifying images to reduce weather impacts before the detection phase.

Using the concept of Feature Corrective Transfer Learning (FCTL), Wei et al. [4] developed a model for object detection in non-ideal conditions, termed NITF RCNN. Through the use of this model, real-time detection is enabled, eschewing any processes which create idealized images either directly or indirectly. However, the NITF RCNN model, due to its generalist nature, lacks specific adaptations for detecting objects in images captured during rain.

Adverse weather cross-domain object detection (CDOD) has been widely studied. Recent works, such as those by Li et al. [21] and Chen et al. [22], focus on domain adaptation to improve detection under various adverse weather condi- tions without requiring interim outcomes. These approaches emphasize the importance of adapting models to different weather scenarios, yet still leave room for methods specifically designed to handle the unique challenges posed by rainy weather. Despite these advancements, our approach, RAF-RCNN, integrates feature transfer learning directly into the detection framework without relying on deraining evaluation, simplifying the model architecture and enhancing detection performance under rainy conditions.

III. Methodology

A. Overview of RAF-RCNN Framework

The RAF-RCNN, as illustrated in Fig. 2, is a novel object detection framework tailored for the intricacies of rainy conditions. It meticulously extends the capabilities of the standard Faster R-CNN [23] by incorporating specialized structural modifications. These adaptations are critical in enabling the RAF-RCNN to maintain high fidelity in feature representation and object localization, even in the presence of visual noise introduced by rain. The model's end-to- end architecture encapsulates advanced preprocessing, robust feature extraction, feature refinement, and attention-enhanced detection mechanisms. This integration facilitates a seamless transition of feature learning from clear to rain-affected scenarios, thereby ensuring consistent detection performance across varied weather conditions.

The RAF-RCNN architecture, tailored for rain-distorted conditions, features a dual-stage backbone upgrade and a self-attention-equipped ROI head. The **Pre-Recursive Net- work** initially processes inputs to alleviate rain's visual noise, streamlining them for the **ResNet-50 FPN** to extract multi-scale features crucial for accurate detection. Post- extraction, the **Post-Recursive Network** fine-tunes these features, enhancing the model's interpretive capabilities in adverse weather (Fig. 4). Complementing the backbone, the ROI head incorporates a **self-attention mechanism**, focusing the model's predictive power on salient features for robust object localization.

RAF-RCNN's cutting-edge feature corrective transfer learning, depicted in Fig. 2, directly aligns rain-affect feature maps with their clear-weather counterparts, harnessing the traditional Faster R-CNN's strengths. Due to the introduction of a transfer learning structure, the loss function does not only include the bounding box regression loss L_{BBox} and classification loss L_{Class} covered by Faster R-CNN [23], but also incorporates a feature match lossL_{Feat}. Therefore, the modified total loss formula is:

$$L_{total} = \lambda \cdot L_{Feature} + \alpha \cdot L_{BBox} + \beta \cdot L_{Class}. \quad (1)$$

.

This feature alignment is quantified and optimized through the **Extended Area Structural Discrepancy Loss (EASDL)**, a loss function that critically evaluates and min- imizes the structural discrepancies between feature maps, bolstering the model's detection accuracy in challenging rain conditions.

B. Backbone Enhancements

The enhanced backbone architecture in the Recursive Faster R-CNN [23] is innovatively designed to potentially improve the model's performance in processing rain-distorted images. Based on PReNet [13] and other recursive deraining

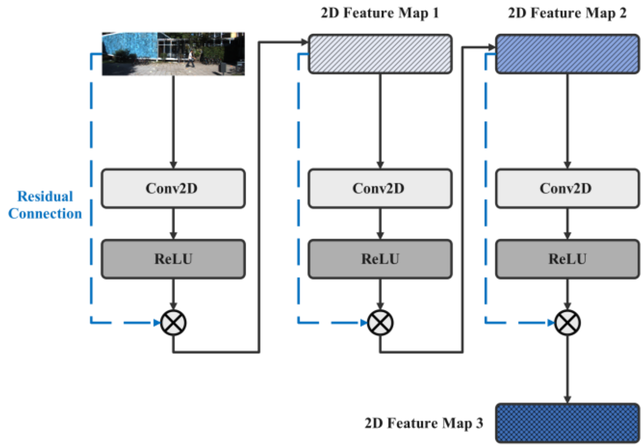

Fig. 3: Architecture of Pre-Recursive Network in Backbone

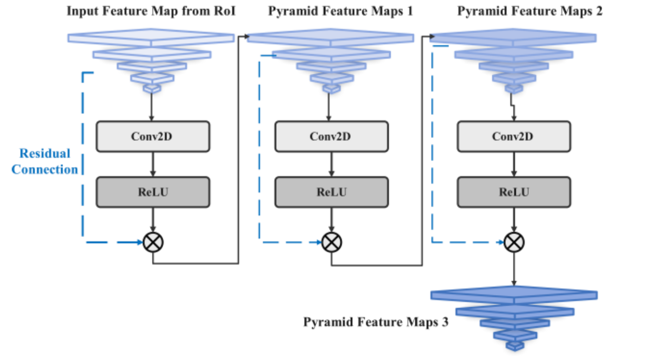

Fig. 4: Architecture of Post-Recursive Network in Backbone

methods [5], [14], [15], this modified backbone is organized into three distinct but interconnected sub-networks. Each sub-network is strategically positioned to contribute towards the potential improvement of feature representation in rain-affected scenarios.

1) Initial Stage: Pre-Recursive Network: To combat rain-distorted images, the Pre-Recursive Network operates as an advanced preprocessing unit that targets the visual anomalies that result from rain. It is important to note that this network plays an important role in adapting to the intricate details of RGB images affected by rain as part of the initial processing in order to ensure that distortions caused by rain are fully addressed.

Structurally, this network consists of several convolutional layers, each followed by a Rectified Linear Unit (ReLU). The ReLU activation [24] is crucial for introducing non-linearity, enabling the network to capture the complex patterns associated with rain-distorted images. In order to ensure that essential image details are preserved, each convolutional layer extracts rain-specific features while the residual connection ensures the extraction of essential features. Fig. 3 shows the architecture of the Pre-Recursive Network. In order to ensure that critical contextual information is retained, this combination of feature extraction and preservation is crucial in preparing the image for further processing by the FPN backbone.

2) Core Feature Extraction: FPN Backbone: Based on the robust ResNet50 model [25], the architecture is centered around the Feature Pyramid Network (FPN) Backbone [26]. In rain-distorted environments, the ResNet50 FPN backbone is capable of extracting multi-scale features, which is essential for the detection of objects at different size scales. To detect objects obscured or altered by rain, the FPN must be able to create hierarchical representations of features at various scales. It generates a comprehensive feature map that captures both macro and micro details, making it easier to detect objects that may be partially obscured or distorted as a result of rain.

3) Refinement Stage: Post-Recursive Network: The Post-Recursive Network serves as a critical refinement stage, strategically designed to align the feature maps from the FPN closer to those typical of clear weather conditions. By processing the multi-scale feature maps generated by the FPN, this network aims to fine-tune the rain-distorted features, enhancing their suitability for more accurate object detection in adverse weather.

Similar to the Pre-Recursive Network, it consists of several convolutional layers with ReLU activations [24] tailored to the FPN output scale, which is showen in Fig. 4. In each layer, we refine the feature maps, adjusting the rain-affected features subtly while preserving the integrity of the core data. The targeted post-processing of the model is essential in order to enable it not only to detect objects in rainy conditions, but also to do so with a feature representation that bridges the gap to clear weather conditions, thereby facilitating a successful transfer of learning.

C. Integration of Attention Mechanism in ROI Head

By incorporating the Transformer module [27] into the Region of Interest (ROI) header of the Faster R-CNN framework [23], some assistance in the field of object detection can be provided. We believe that, especially under challenging conditions of rainwater distortion, improvements in ROI heads can help the model focus more on prominent features in RoI to improve the detection process, which is critical to maintaining performance in harsh weather conditions.

As displayed in Fig. 5, the modified ROI header based on attention mainly includes the following steps.

Input Feature Map Reception: The journey starts with the input feature map from the RoI, which contains essential information for object detection.

Attention Module Operation:

- The feature map is first passed through a linear layer that performs down-sampling, reducing its dimension- ality to focus on critical features.
- A ReLU activation function follows, introducing non-linearity to enhance feature representation.
- An up-sampling linear layer then restores the feature dimension, preparing it for attention weight application.

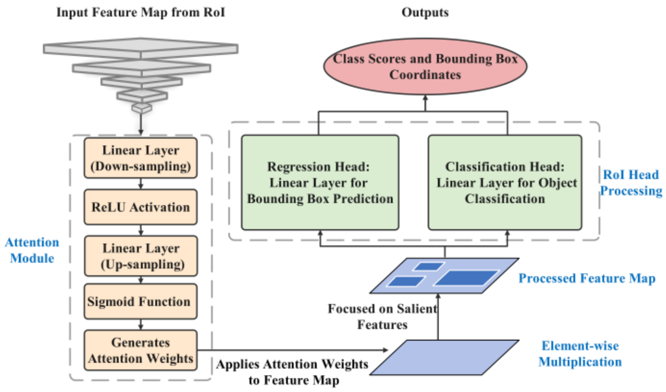

Fig. 5: Artecture of Modified Self-Attention ROI Head

- A Sigmoid function generates attention weights, which are crucial for highlighting significant aspects of the feature map while diminishing less important parts.

Application of Attention Weights: An element-wise multiplication is performed between the attention weights and the feature map. This step selectively amplifies salient features, ensuring that the model focuses on the most relevant aspects for object detection.

Processed Feature Map: The resultant feature map, now refined with focused attention, encapsulates the crucial

features needed for accurate object identification and localization.

RoI Head Processing: The classification head, comprising a linear layer, processes the feature map to predict object classes. Concurrently, the regression head, also a linear layer, predicts bounding box coordinates, crucial for object localization.

Output Generation: The final output includes class scores, indicating the likelihood of each class, and bounding box coordinates, specifying the location of detected objects within the RoI.

D. Feature Corrective Transfer Learning Strategy

Our research utilizes a transfer learning approach at the feature map level, which is particularly effective in rain-distorted image scenarios [4]. In this approach, the structural and spatial characteristics of feature maps generated from rain-affected images are aligned with those of feature maps generated under clear weather conditions. The alignment is hypothesized to facilitate the extraction of weather-agnostic features, increasing the model's generalization capabilities across a variety of environmental conditions.

Feature Map Generation: For a given rain-distorted image, the enhanced backbone of our RAF RCCN model generates a feature map, denoted as A. As a result, the same image in clear weather conditions is processed through a traditional Faster RCNN backbone to produce B as a feature map.

Objective of Transfer Learning: The core objective is to ensure that A, despite being derived from rain-distorted inputs, is closely similar to B in terms of structural sim- ilarity

for utilizing the clear weather detection capabilities of the traditional Faster R-CNN model [23].

E. Extended Area Structural Discrepancy Loss: EASDL

To evaluate and learn from the differences between A and B at the feature level, we propose the Extended Area Structural Discrepancy Loss (EASDL). In addressing the challenges posed by complex visual environments in tasks such as object detection, the need arises for a loss function capable of effectively capturing and quantifying structural differences between feature maps.

The EASDL is formulated to capture not only local gradient discrepancies at individual pixel locations but also gradient consistency in a broader surrounding area. This dual consideration enables the model to be sensitive to both immediate and expansive structural variances, which is particularly useful in detecting subtle yet critical spatial discrepancies in feature maps. Unlike traditional approaches that rely on pixel-based comparisons, EASDL focuses on the structural aspects and area consistency of the feature maps.

1) Mathematical Formulation: Given two feature maps A and B with dimensions *[batch size, channels, width, height]*, the EASDL is defined as:

$$\mathrm{EASDL}(A,B) = \frac{1}{WH}\sum_{x=1}^{W}\sum_{y=1}^{H} \exp(-\Delta S(x,y)) \times \Delta S(x,y) + \lambda \cdot \Omega(A,B,x,y,r) \quad . \qquad (2)$$

where:

- $\Delta S(x,y)$ represents the difference in local gradient magnitudes at position (x,y), computed as the absolute difference between the gradients of A and B.
- λ is a weighting factor that balances the contribution of the extended area gradient term.
- $\Omega(A,B,x,y,r)$ denotes the extended area gradient consistency term, defined over a neighborhood of radius r around each pixel.

2) Implementation Details: The gradient function $G(\cdot)$, a core component of the EASDL, is designed to extract the structural characteristics of feature maps. We implement this function using the Sobel operator, a prominent method for edge detection, to compute the gradient at each pixel within the feature maps. The operator functions by convolving the feature map with two separate 3x3 kernels, each designed to detect edges along specific orientations.

The Sobel operator is defined using the following convolutional kernels:

$$S_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}, \quad S_y = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix}. \qquad (3)$$

and spatial consistency. This resemblance is crucial

The horizontal gradient (Sobel-x) detects vertical edges, and the vertical gradient (Sobel-y) detects horizontal edges. These gradients are computed as follows:

$$G_x(A) = A * S_x, \quad G_y(A) = A * S_y. \tag{4}$$

The overall gradient magnitude at each pixel is then calculated by combining these individual gradients:

$$G(A, x, y) = \sqrt{G_x(A, x, y)^2 + G_y(A, x, y)^2}. \tag{5}$$

The extended area gradient term $\Omega(A, B, x, y, r)$, evaluates the consistency of gradient changes within a neighborhood around each pixel. For a specified radius r, it considers a square window around each pixel:

$$\Omega(A, B, x, y, r) = \frac{1}{(2r+1)^2} \sum_{i=-r}^{r} \sum_{j=-r}^{r} \big| (G(A, x, y) - G(A, x+i, y+j)) - (G(B, x, y) - G(B, x+i, y+j)) \big| . \tag{6}$$

The integration of local gradient differences and extended area consistency in EASDL enables the model to capture both immediate pixel-wise discrepancies and broader spatial patterns, facilitating the learning of robust and generalizable features in complex visual environments.

IV. Experiments

A. Datasets

In our study, we strategically use the KITTI Object Detection Dataset [28] for its diverse real-world driving scenarios and the Rainy KITTI Dataset [29] for its synthetic rain conditions, effectively capturing the performance of our model in both clear and adverse weather scenarios.

KITTI Object Detection Dataset: The KITTI dataset, originating from Karlsruhe, Germany, is a cornerstone in the automotive field, renowned for benchmarking perception tasks like object detection [28]. It consists of 7,481 images, featuring diverse urban, rural, and highway scenes with detailed annotations of various entities, including vehicles, pedestrians, and cyclists among others.

Rainy KITTI Object Detection Datasets: Rainy KITTI is a variation of the original KITTI dataset that includes artificially synthesized rain that mimics real-world rainy conditions for object detection [29]. As part of this dataset, virtual generated images with life-like rain effects are created by overlaying generated rainy condition on the original images utilizing physical particle simulators, scene illumination estimation, and accurate photometric modeling of rain.

In our experiments, the Rainy KITTI dataset serves as the primary resource for training and validation. Each image in the KITTI dataset has seven different levels of rainfall conditions, each corresponding to a different intensity of rainfall (1mm, 5mm, 17mm, 25mm, 50mm, 75mm, 100mm, 200mm). For our dataset assembly, we randomly selected one image from these seven rainfall conditions for each original KITTI image, thereby ensuring an equal number of training data for comparison between clear and rainy conditions. Additionally, this approach allows a more realistic simulation of the impact of varying rain densities on the model, ensuring that the model is capable of adapting to diverse real-world scenarios. Throughout the experiment, 80% of the data is allocated for training and 20% for validation.

B. Quantitative Results

In this study, all models were developed, trained, and evaluated using the PyTorch framework in Python. In order to maximize model performance, 50 training epochs were used with the stochastic gradient descent (SGD) optimization method [30]. SGD optimizer parameters included a learning rate of 0.0005, a momentum of 0.9 and a weight decay of 0.0001.

TABLE I: Evaluation Results of Different Models on Rainy KITTI Dataset After 50 Training Epochs

Method	***mAP***		
	@0.5	***@0.75***	***@[0.5,0.95]***
Faster R-CNN [23]	73.5%	45.1%	42.3%
Retina Net [1]	54.6%	28.8%	30.1%
DETR [31]	59.7%	29.8%	31.2%
YOLOv4 [32]	65.5%	35.5%	35.1%
Dynamic R-CNN [33]	71.8%	47.5%	44.4%
YOLOx [34]	62.6%	33.7%	34.1%
YOLOv8 [35]	74.3%	44.2%	40.4%
NITF-RCNN [4]	78.2%	47.8%	45.9%
RAF-RCNN	82.1%	54.5%	51.4%

Moreover, three metrics were utilized to evaluate model performance: mAP@0.5, mAP@0.75, and mAP@[0.5:0.95], reflecting mean average precision at varying intersection over union (IoU) thresholds. These metrics range from basic (mAP@0.5) to more stringent (mAP@[0.5:0.95]), averaging performance across thresholds from 0.5 to 0.95. After 50 epochs of training and assessment, results for two distinct models are summarized in Table I, reported as mean average precision (mAP) according to the COCO benchmark [36].

The quantitative evaluation, as summarized in Table I, demonstrates the superior performance of RAF-RCNN across all metrics when compared to a variety of contemporary object detection models tested under similar rainy conditions. Notably, RAF-RCNN achieved an mAP@[0.5,0.95] of 51.4%, which is 5.5% higher than the next best model, NITF-RCNN [4], and a significant 9.0% improvement over Dynamic R-CNN [33], the best performing conventional model. This precision enhancement emphasizes RAF-RCNN's capability to maintain detection accuracy under challenging weather conditions, highlighting its effectiveness in adverse environments.

C. Qualitative Results

For a comprehensive analysis, quantitative evaluation metrics alone cannot fully demonstrate the effectiveness of the algorithm. Therefore, we also opted for direct observation

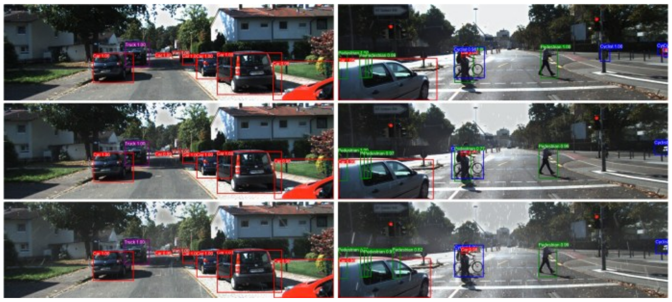

Fig. 6: Detection Performance of RAF-RCNN on Rainy KITTI Dataset under Various Rainfall Conditions (5mm, 25mm, 100mm)

TABLE II: Ablation Study Results of Each Improvement Component in RAF-RCNN

Component				*mAP*		
A	*B*	*C*	*D*	*@0.5*	*@0.75*	*@[0.5:0.95]*
×	✓	✓	✓	78.3%	52.0%	48.3%
✓	×	✓	✓	79.6%	53.1%	49.8%
✓	✓	×	✓	78.5%	48.3%	46.3%
✓	✓	✓	×	80.3%	52.7%	49.4%
✓	✓	✓	✓	82.1%	54.5%	51.4%

Note: A = Pre-recursive network in backbone, B = Post-recursive network in backbone, C = Feature corrective transfer learning module, D = Modified self-attention ROI head.

of the detection results. Using the Rainy KITTI dataset, we selected two representative images (000031, 000383) and simulated rainfall conditions (5 mm, 25 mm, and 100 mm). Fig. 6 displays the intuitive results of our proposed object detection algorithm applied to these six images, none of which were included in the training set.

Our algorithm maintains a relatively good detection performance in sparse object conditions across a wide range of rainfall intensities, as demonstrated by the left three images in Fig. 3. With confidence levels close to 1.00, it is evident that the bounding boxes are consistently and accurately positioned.

However, the right three images reveal a less ideal detection outcome when objects have a high degree of overlap. In scenarios with multiple overlapping objects, there is noticeable deviation in the bounding boxes positioning. As the intensity of simulated rainfall increases, the incidence of false positives and missed detection also rises.

D. Ablation Experiments

An ablation study was conducted to validate the significance of each component of the RAF-RCNN by selectively disabling components of the model. The components, denoted as A, B, C, and D, correspond to the Pre-Recursive Network in the Backbone (A), the Post-Recursive Network in the Backbone (B), the Feature Corrective Transfer Learning correction module (C), and the Modified Self-Attention ROI Head (D), respectively. As we removed each component in turn, we were able to investigate their individual contributions to the performance of the model. As shown in Table II, the complete model (including all components) achieved the highest performance with an mAP@[0.5:0.95] of 51.4%. In contrast, removing component A resulted in a decrease to 48.3%, highlighting its critical role in enhancing detection under rainy conditions. Similarly, the removal of components B, C, and D also led to varying decreases in performance, demonstrating their importance in the model's overall effectiveness.

To validate the effectiveness of the EASDL function proposed in this paper, we also assessed the impact of different loss functions on the performance of RAF-RCNN. In particular, we compared the EASDL with other standard loss functions, such as Mean Square Error (MSE), Struc- tural Similarity Index (SSIM), and Cosine Similarity. The results, presented in Table III, show that EASDL signifi- cantly outperforms the other evaluated functions, with an mAP@[0.5:0.95] of 51.4%, compared to 46.1% for the next best performing loss function (Cosine Similarity). Accordingly, EASDL is effective in aligning feature maps from different weather conditions, which enhances the model's ability to detect objects in rainy weather.

TABLE III: Performance Comparison of Different Loss Functions as Feature Loss in RAF-RCNN's Feature Corrective Transfer Learning Mechanism on the Rainy KITTI Dataset

Loss Function	*mAP*		
	@0.5	*@0.75*	*@[0.5,0.95]*
MSE	74.7%	45.1%	44.3%
SSIM	76.8%	44.5%	43.8%
Cosine Similarity	78.8%	49.0%	46.1%
EASDL	82.1%	54.5%	51.4%

V. Conclusions and Future Work

In this work, we presented RAF-RCNN, an innovative approach to object detection in rainy conditions that integrates a deraining process directly into the detection framework. Our model leverages a novel feature transfer learning tech- nique and the Extended Area Structural Discrepancy Loss (EASDL), outperforming traditional models on the challenging Rainy KITTI dataset. The success of RAF-RCNN not only demonstrates its efficacy in adverse weather but also its potential to improve autonomous systems' reliability, such as vehicles and surveillance in variable environments.

Aside from its current application, RAF-RCNN has the potential to provide a multitude of benefits and contributions. As a result of its ability to detect features in complex weather conditions without requiring pristine images, it represents a significant advancement in autonomous vehicle technology. Combined with its efficient feature transfer learning, the model's adaptability paves the way for future research into other adverse conditions, including fog, dust, and nighttime. This method can be used for a variety of complex image signal processing tasks, providing a broader range of applications for deep learning. During future research, we intend to minimize the occurrence of false positives and missed detections in highly cluttered environments, improve the feature transfer learning process to handle even more challenging conditions, and extend our approach to a wider range of environments.

References

[1] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dolla´r, "Focal loss for dense object detection," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2980–2988.

[2] D. P. Bavirisetti, H. R. Martinsen, G. H. Kiss, and F. Lindseth, "A multi-task vision transformer for segmentation and monocular depth estimation for autonomous vehicles," *IEEE Open Journal of Intelligent Transportation Systems*, 2023.

[3] W. Liu, G. Ren, R. Yu, S. Guo, J. Zhu, and L. Zhang, "Image- adaptive yolo for object detection in adverse weather conditions," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, no. 2, 2022, pp. 1792–1800.

[4] C. Wei, G. Wu, and M. J. Barth, "Feature corrective transfer learning: End-to-end solutions to object detection in non-ideal visual conditions," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2024, pp. 23–32.

[5] T. Wang, K. Wang, Q. Li, and H. Cecotti, "Progressive rain removal based on the combination network of cnn and transformer," *Intell. Neuroscience*, vol. 2022, jan 2022. [Online]. Available: https://doi.org/10.1155/2022/5067175

[6] C. Wei, G. Wu, M. Barth, P. H. Chan, V. Donzella, and A. Huggett, "Enhanced object detection by integrating camera parameters into raw image-based faster r-cnn," in *2023 IEEE International Intelligent Transportation Systems Conference (ITSC)*. IEEE, 2023.

[7] Y. Wang, P. H. Chan, and V. Donzella, "A two-stage h. 264 based video compression method for automotive cameras," in *2022 IEEE 5th international conference on industrial cyber-physical systems (ICPS)*. IEEE, 2022, pp. 01–06.

[8] P. H. Chan, C. Wei, A. Huggett, and V. Donzella, "Raw camera data object detectors: an optimisation for automotive processing and transmission," *Authorea Preprints*, 2023.

[9] C. Wei, "Vehicle detecting and tracking application based on yolov5 and deepsort for bayer data," in *2022 17th International Conference on Control, Automation, Robotics and Vision (ICARCV)*. IEEE, 2022, pp. 843–849.

[10] X. Fu, J. Huang, X. Ding, Y. Liao, and J. Paisley, "Clearing the skies: A deep network architecture for single-image rain removal," *IEEE Transactions on Image Processing*, vol. 26, no. 6, pp. 2944–2956, 2017.

[11] R. Qian, R. T. Tan, W. Yang, J. Su, and J. Liu, "Attentive generative adversarial network for raindrop removal from a single image," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 2482–2491.

[12] X. Fu, J. Huang, D. Zeng, Y. Huang, X. Ding, and J. Paisley, "Removing rain from single images via a deep detail network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 3855–3863.

[13] D. Ren, W. Zuo, Q. Hu, P. Zhu, and D. Meng, "Progressive image deraining networks: A better and simpler baseline," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 3937–3946.

[14] H. Wang, Q. Xie, Q. Zhao, and D. Meng, "A model-driven deep neural network for single image rain removal," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 3103–3112.

[15] S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan, M.-H. Yang, and L. Shao, "Multi-stage progressive image restoration," in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021, pp. 14 816–14 826.

[16] S.-C. Huang, T.-H. Le, and D.-W. Jaw, "Dsnet: Joint semantic learning for object detection in inclement weather conditions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 8, pp. 2623–2633, 2021.

[17] S.-C. Huang, Q.-V. Hoang, and T.-H. Le, "Sfa-net: A selective features absorption network for object detection in rainy weather conditions," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 34, no. 8, pp. 5122–5132, 2023.

[18] V. A. Sindagi, P. Oza, R. Yasarla, and V. M. Patel, "Prior-based domain adaptive object detection for hazy and rainy conditions," in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Part XIV 16*. Springer, 2020, pp. 763–780.

[19] E. O. Appiah and S. Mensah, "Object detection in adverse weather condition for autonomous vehicles," *Multimedia Tools and Applications*, pp. 1–27, 2023.

[20] K. Wang, T. Wang, J. Qu, H. Jiang, Q. Li, and L. Chang, "An end-to-end cascaded image deraining and object detection neural network," *IEEE Robotics and Automation Letters*, vol. 7, no. 4, pp. 9541–9548, 2022.

[21] Y.-J. Li, X. Dai, C.-Y. Ma, Y.-C. Liu, K. Chen, B. Wu, Z. He, K. Kitani, and P. Vajda, "Cross-domain adaptive teacher for object detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 7581–7590.

[22] M. Chen, W. Chen, S. Yang, J. Song, X. Wang, L. Zhang, Y. Yan, D. Qi, Y. Zhuang, D. Xie *et al.*, "Learning domain adaptive object detection with probabilistic teacher," *arXiv preprint arXiv:2206.06293*, 2022.

[23] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 6, pp. 1137–1149, 2016.

[24] A. F. Agarap, "Deep learning using rectified linear units (relu)," *arXiv preprint arXiv:1803.08375*, 2018.

[25] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[26] T.-Y. Lin, P. Dolla´r, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *IEEE conference on computer vision and pattern recognition*, 2017, pp. 2117–2125.

[27] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.

[28] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision meets robotics: The kitti dataset," *The International Journal of Robotics Research*, vol. 32, no. 11, pp. 1231–1237, 2013.

[29] S. S. Halder, J.-F. Lalonde, and R. d. Charette, "Physics-based rendering for improving robustness to rain," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 10 203–10 212.

[30] L. Bottou, "Large-scale machine learning with stochastic gradient descent," in *Proceedings of COMPSTAT'2010: 19th International Conference on Computational StatisticsParis France, August 22-27, 2010 Keynote, Invited and Contributed Papers*. Springer, 2010, pp. 177–186.

[31] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko, "End-to-end object detection with transformers," in *European conference on computer vision*. Springer, 2020, pp. 213–229.

[32] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "Yolov4: Op- timal speed and accuracy of object detection," *arXiv preprint arXiv:2004.10934*, 2020.

[33] H. Zhang, H. Chang, B. Ma, N. Wang, and X. Chen, "Dynamic r- cnn: Towards high quality object detection via dynamic training," in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XV 16*. Springer, 2020, pp. 260–275.

[34] Z. Ge, S. Liu, F. Wang, Z. Li, and J. Sun, "Yolox: Exceeding yolo series in 2021," *arXiv preprint arXiv:2107.08430*, 2021.

[35] G. Jocher, A. Chaurasia, and J. Qiu, "Ultralytics YOLO," Jan. 2023. [Online]. Available: https://github.com/ultralytics/ultralytics

[36] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dolla´r, and C. L. Zitnick, "Microsoft coco: Common objects in context," in *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Part V 13*. Springer, 2014, pp. 740–755.