WARMOS: Enhancing Weather-Affected Referred Moving Object Segmentation

Prafulla saxena[1] Dinesh Kumar Tyagi[1] Santosh Kumar Vipparthi[2] Subrahmanyam Murala[2,3]

[1] Malaviya National Institute of Technology Jaipur
[2] Indian Institute of Technology Ropar
[3] SCSS Trinity College Dublin
2020rcp9580@mnit.ac.in

Abstract. Environmental noise, such as haze and rain, poses significant challenges in video surveillance, in tasks like referred video object segmentation. These weather-related disturbances introduce excessive pixel variance, making moving object segmentation more complex. In this work, we focus on addressing the issue of adverse weather conditions by simulating the effects of haze and rain in videos and employing a robust noise removal model. The model effectively reduces pixel variance caused by environmental factors. This enhanced framework is precious for referred moving object segmentation, where objects identification done based on text queries. By integrating our noise removal module, we ensure better alignment of features, which enhances the precision of referred moving segmentation. Our approach maintains temporal consistency, making object segmentation more reliable under challenging weather conditions while preserving the original video quality by removing weather noise. We have employed separate noise removal modules for haze and rain environmental noise. A ResNet based classifier model trained to identify the noise class on the fly. To demonstrate the effectiveness of our methodology, we selected an ROV benchmark to assess segmentation performance. Experiments on the DAVIS 2017 dataset show that our proposed methodology performs well on weather-affected videos, significantly improving the benchmark metrics Jaccard (J) and F-measure (F) indices after removing weather noise. Using the benchmark SgMg model for referred segmentation, the mean J&F score is 63.64 without environmental noise. When haze is introduced to the dataset, the mean J&F score drops to 58.71. After applying WARMOS approach, the mean J&F improves to 60.50. A similar pattern is observed for rain: when rain is introduced, the mean J&F score is 61.00, and after applying WARMOS, it improves to 61.06. This highlights our approach's significance in mitigating the impact of environmental noise.

Keywords: Moving Object Segmentation · Referred segmentation · Weather noise · Haze and rain removal · Deep learning

1 Introduction

Weather disturbances such as haze, rain, and fog pose significant challenges to video surveillance tasks, including referred video object segmentation (RVOS) [1]. These weather-related noises degrade the quality of video frames, making it difficult to accurately capture the visual features needed for effective object segmentation. The visibility of objects is often compromised due to blurred or occluded regions, which can severely affect the performance of segmentation models. Figure-1 demonstrates the effects of haze and rain in real world images with simulation. In tasks like moving object segmentation (MOS) [2–8], where temporal and spatial accuracy is crucial, the interference caused by weather noise can lead to substantial errors in identifying and tracking objects.

Fig. 1: Visual degradation of a video frame after environmental noise like haze and rain. Due to haze and rain quality of an image degrades and compromised the object visibility. Images from DAVIS 2017 [9] dataset

Referred video object segmentation (RVOS) [1,10–13] is an emerging field at the intersection of computer vision and natural language processing. It focuses on segmenting an object based on a language expression, combining both visual and linguistic information. While standard object segmentation primarily focuses on the spatial domain within a single frame, RVOS introduces a new layer of complexity by adding natural language as a reference. However, some of approaches often ignore temporal information [14], which is essential in video-based tasks. Standard moving object segmentation (MOS) [2,15], methods, in contrast, excel in capturing temporal changes across frames to identify moving objects but fail to incorporate semantic understanding or contextual information.

When weather disturbances like rain or haze are present in captures videos and images, these challenges are exacerbated in video segmentation applica-

tions [16]. The combination of low visibility, dynamic backgrounds, and weather noise hampers the model's ability to accurately target specific objects described through text. In scenarios like Referred video object segmentation, where precise spatiotemporal features are required for identifying referred objects, weather disturbances can severely degrade performance. Objects may become partially or fully obscured, leading to incorrect or incomplete segmentation. This further complicates the task, as spatiotemporal information and text-based embedding must work in conjunction, while weather noise interferes with both.

Furthermore, weather noise like haze and rain significantly impacts the effectiveness of moving object segmentation, adding complexity to Referred moving segmentation as well. Techniques for removing such weather-related noise, like dehazing [17–20] and rain removal [21–23], are essential to improving the quality of video frames. By enhancing visibility and removing these disturbances, the model can more accurately focus on the relevant object characteristics, leading to improved performance in referred video segmentation.

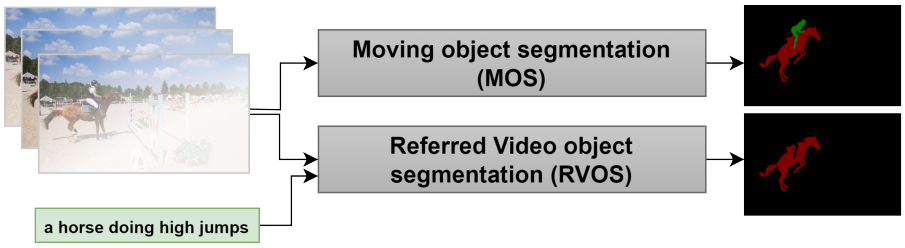

Fig. 2: In RVOS, the segmentation of moving objects depends on a text query. Whether in MOS task, all moving objects are segmented irrespective of the text query.

Though several learning-based approaches have been proposed to improve the MOS task by designing various deep learning models, including dealing with challenges like illumination changes, low visibility, and dynamic backgrounds, addressing weather noise remains a critical area of focus. When referring to moving objects through language expressions, the segmentation task becomes even more complex, requiring advanced handling of cross-modal sources like vision and language. So in this work we have proposed an approach to enhance the Referred moving object segmentation in weather-affected scenarios. Figure-2 shows the referred object segmentation in weather affected scenario. We propose the following contribution to this paper.

- Proposed WARMOS, a weather noise removal framework for improving the existing referred video object segmentation performance.
- Shows the adverse effect of weather disturbance in referred segmentation task by simulating haze and rain in real world dataset.
- Utilises a benchmark SgMg, referred video segmentation model and improves its performance in weather affected scenarios.

– Evaluate on Ref-DAVIS17 benchmark dataset and verify the proposed methodology's effectiveness with quantitative and qualitative analysis.

2 Related work

RVOS focuses on segmenting specific objects in a video based on a text reference. In this task, the representation of objects is influenced by the natural language query used to refer to them. Various approaches, such as those in [10–12], handle the segmentation of referred objects frame by frame as the video progresses. However, when referring to moving objects, a deeper understanding of their movement is necessary. Segmenting a moving object referred to by a text query requires the integration of both temporal and visual features [13,24]. Therefore, the combined use of spatio-temporal and linguistic features is crucial for the RVOS task. Additionally, language embedding must be incorporated into the dataset for RVOS, which results in a limited number of available databases in the literature. The RVOS benchmark [25] introduces a dataset and evaluation metrics where segmentation is guided by a natural language query.

Environmental disturbances like rain or haze removal in video frames is a crucial task for improving visual quality in video surveillance tasks, especially in outdoor environments. These weather disturbances degrade visibility, making it challenging for models to accurately detect and segment objects in various general computer vision tasks like anomaly detection [26–30], video segmentation etc. Numerous statistical and deep learning models have been developed to address these issues. Traditional statistical approaches focus on physical models that account for atmospheric scattering effects. For example, the dark channel prior (DCP) [16] is a widely-used statistical method for haze removal, based on the observation that haze-free images contain pixels with low intensity in at least one color channel. DCP has been extended and refined in various works [18] to improve dehazing performance.

In recent years, deep learning-based models have shown remarkable success in both haze and rain removal. Methods like DehazeNet [19] leverage convolutional neural networks (CNNs) to automatically learn haze-relevant features from training data. This method bypasses the need for handcrafted features, making it more flexible in diverse conditions. Similarly, rain removal techniques have also evolved, with early methods focusing on low-level image processing [21], and more recent approaches utilizing deep networks to separate rain streaks from background scenes. The deep detail network (DDN) [22] is a notable method that splits the image into two layers: a rain-free background and a rain-streak layer, using deep CNNs to refine the background.

These deep learning methods significantly outperform traditional techniques, especially in complex scenarios with heavy haze or rain. Hybrid approaches that combine physical models with deep learning [23] have also shown promise by leveraging the strengths of both worlds. The ongoing development of larger and more diverse datasets, such as RESIDE [20] for haze removal and Rain800 [31] for rain removal, has also contributed to the rapid advancement of weather noise

removal technologies. These databases and methods provide improved benchmarks and evaluation metrics, helping researchers better understand and tackle the challenges posed by weather-induced noise in videos.

3 Proposed Work

In this work we have utilised a state-of-the-art SgMg [1] referred video segmentation model for evaluation. SgMg model is a deep learning based end-to-end framework which is trained on ref-DAVIS 2017 dataset. We have simulated haze and rain environmental noise to the dataset and evaluate the performance on weather-affected dataset first. Degradation in the performance motivates us to develop towards weather noise removal framework. In the later sections we have elaborated the weather simulation method from which environmental noise has been introduced in the dataset.

3.1 Simulation method

We have utilised procedural image generation techniques that use randomness and geometric principles to simulate haze and rain effects.

Haze Generation Methodology In the context of haze generation, we rely on the standard haze formation model, where the observed hazy image $I(x)$ is a combination of the scene radiance $J(x)$ and the atmospheric light A. The transmission map $t(x)$ controls the blending of these two components. The mathematical representation is as follows:

Hazy Image Model

$$I(x) = J(x) \cdot t(x) + A \cdot (1 - t(x)) \tag{1}$$

where:

- $I(x)$ is the observed hazy image,
- $J(x)$ is the scene radiance (haze-free image),
- $t(x)$ is the transmission map,
- A is the global atmospheric light.

Transmission Map (Haze Map) Generation The transmission map $H(x)$ is generated using a random gradient based on image coordinates, modeled as:

$$H(x) = \alpha_1 \cdot xv + \alpha_2 \cdot yv \tag{2}$$

where:

- xv and yv are 2D gradient fields based on the image coordinates,
- α_1 and α_2 are random weights controlling the gradient's direction and intensity.

Final Hazy Image The final hazy image is computed by blending the original image with the haze layer as follows:

$$I_{hazy}(x) = I(x) \cdot (1 - H(x)) + 255 \cdot H(x) \tag{3}$$

where:

– 255 represents the white haze layer.

Rain Generation Method For rain generation, we simulate stick-like raindrops by calculating their positions and orientations. Gaussian blurring is applied to enhance the raindrop streaks.

Raindrop Endpoint Calculation To generate a raindrop, we compute its endpoint based on its length and orientation angle:

$$x_1 = x_0 + L \cdot \cos(\theta) \tag{4}$$

$$y_1 = y_0 + L \cdot \sin(\theta) \tag{5}$$

where:

– (x_1, y_1) is the endpoint of the raindrop,
– L is the length of the raindrop,
– θ is the angle of orientation.

Gaussian Blur for Raindrop Streaks To simulate the appearance of raindrop streaks, we apply Gaussian blur, which is mathematically represented as:

$$G(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}} \tag{6}$$

where σ is the standard deviation that controls the spread of the raindrop streaks.

Final Rainy Image The final rainy image is computed by blending the original image with the raindrop layer:

$$I_{rainy}(x) = I(x) \cdot (1 - \beta) + R(x) \cdot \beta \tag{7}$$

where:

– $I(x)$ is the original image,
– $R(x)$ is the generated raindrop layer,
– β is the blending factor that controls the intensity of the rain effect.

3.2 Enhancement of weather-affected referred video object segmentation (WARMOS) framework

Referred video object Segmentation architecture In this work, we utilize the Spectrum-guided Multigranularity (SgMg) [1] approach, which performs direct segmentation on encoded features and leverages visual details to refine the segmentation masks. SgMg employs spectrum-guided intra-frame global interactions in the spectral domain to enhance multimodal representation, allowing for more accurate segmentation across diverse input types practicality in complex video analysis tasks. The SgMg performance has been shown in the Table- 1 on Ref-DAVIS 2017 dataset.

Method	**Backbone**	**J&F**	**J**	**F**
SgMg[[1]])	Video-Swin-B	63.64	61.22	66.06

Table 1: Quantitative results of SgMg [1] on Original Ref-DAVIS-2017. **J**: Jaccard, **F**: F-measure, **J&F** is average of J and F

Once the weather noise like haze and rain have been introduced to the dataset we directly evaluate them with Benchmark SgMg framework. The noise degraded the video quality hence we got sub-optimal results. Table-2 and Table-3 show the weather-affected results on the Ref-DAVIS 2017 dataset.

Method	**Backbone**	**J&F**	**J**	**F**
SgMg[[1]])	Video-Swin-B	58.71	56.48	60.95

Table 2: Quantitative results of SgMg [1] on Ref-DAVIS-2017 affected by weather noise **Haze**. **J**: Jaccard, **F**: F-measure, **J&F** is average of J and F

Method	**Backbone**	**J&F**	**J**	**F**
SgMg[[1]])	Video-Swin-B	61.00	58.03	63.69

Table 3: Quantitative results of SgMg [1] on Ref-DAVIS-2017 affected by weather noise **Rain**. **J**: Jaccard, **F**: F-measure, **J&F** is average of J and F

Weather noise removal framework To reduce the adverse effects of environmental noise, we used a convolutional neural network-based method Light-DehazeNet (LD-Net) [17]. LD-Net works by estimating both the transmission map and atmospheric light through a modified atmospheric scattering model. Additionally, a color visibility restoration technique is applied to avoid color distortion in the dehazed image. We fine-tuned the LD-Net model separately for haze and rain removal, training two different models with the same architecture. The improved results highlight the importance of this noise removal framework, which boosts the performance of RVOS in weather-affected situations. A ResNet [32] based classifier is trained to know the environmental noise class on the fly. Once the noise class is recognized, the corresponding noise removal model is enabled.

4 Experimental Setup

We have evaluated WARMOS on DAVIS-2017 [9] datasets which are widely used benchmarks in video object segmentation tasks. To assess the performance of the proposed method, quantitative results of the Jaccard and the F-measure similarity index have been compared with state-of-the-art methods. The Jaccard index (J), also known as Intersection over union (IoU), measures the similarity between predicted output and ground truth mask by computing the ratio of Intersection over union among two classes. Jaccard Index can be represented in eq. 8

$$J = \frac{\text{Area of Overlap}}{\text{Area of Union}} = \frac{|A \cap B|}{|A \cup B|} \tag{8}$$

where A and B are the sets that represent the predicted output and group truth, and the F-measure index is evaluated as a harmonic mean of precision and recall. It provides a balanced evaluation of the model performance. F-measure can represented as in eq. 9

$$F - measure = 2 \times \frac{precision \times recall}{precision + recall} \tag{9}$$

where precision is the ratio of predicted true positives (TP) to the total predicted positives (TP + FP), and recall is the ratio of true positives predicted to the total actual positives.

$$precision = \frac{TP}{TP + FP} \text{ , } recall = \frac{TP}{TP + FN}$$

4.1 Dataset:

The DAVIS-2017 [9] dataset consists of 90 videos, widely used as benchmarks for Video Object Segmentation (VOS). Of these 90 videos, 60 are designated for training, while the remaining 30 are set aside for testing. To adapt the dataset for Referred Video Object Segmentation (RVOS) tasks, it is augmented with Regular Expressions (REs), which act as text queries to identify specific objects in the videos. Ground-truth maps are then dynamically generated for each frame based on the input text query. If a video contains multiple objects, the relevant objects are masked according to the corresponding text query.

4.2 Training details:

SgMg framework We have leverages SgMg [1] framework for RVOS task. This is a PyTorch based implementation. We utilised trained weights available to us by the authors for the inference purpose.

Weather noise removal framework We have fine-tuned ResNet model for classification purpose. The classifier accuracy is 99%. We employed Light-DehazeNet (LD-Net) and utilizes torch-1.13 framework over NVIDIA A100 GPU with torch-1.13 and Cuda 11.7 to train and fine-tune our haze and rain removal network. The objective of the weather noise removal with RVOS task is to generate accurate binary mask that highlights relevant objects in video frames despite weather disturbance in original video frames. To enhance the robustness of our model, we applied various data augmentations during training, such as random flips, rotations, random crops, etc. Our training process spanned 60 epochs.

5 Results and discussion

In this section we have discussed the performance of our proposed method and shows its effectiveness with quantitative and qualitative results. We have reported the mean Intersection over Union also known as Jaccard (J) and average F-measure (F) to evaluate the segmentation performance of our model.

5.1 Quantitative Analysis

We evaluate our method on the Ref-Davis 2017 dataset and shown the results in Table-4 and Table-5. The quantitative results of WARMOS demonstrate satisfying performance and improved the results when compared with Table-2 and 3 that evaluated in weather affected scenarios. These results highlight the robustness and efficacy of our approach in handling complex video object segmentation guided by text query.

Method	**Backbone**	**J&F**	**J**	**F**
SgMg[[1]])	Video-Swin-B	60.50	58.00	63.00

Table 4: Quantitative results of SgMg [1] on Ref-DAVIS-2017 after Haze noise removal **J**: Jaccard, **F**: F-measure, **J&F** is average of J and F

Method	**Backbone**	**J&F**	**J**	**F**
SgMg[[1]])	Video-Swin-B	61.05	58.58	63.52

Table 5: Quantitative results of SgMg [1] on Ref-DAVIS-2017 after Rain noise removal. **J**: Jaccard, **F**: F-measure, **J&F** is average of J and F

5.2 Qualitative Analysis

In Figure-3 we have shown qualitative results of our WARMOS approach. Results indicates that our approach performs well in segmenting objects referred by text query. In Figure-3 two set of examples are considered from DAVIS-2017 dataset from dog-jumps and black-scooter category. Label-A shows actual image, label-B

shows ground truth, label-C shows Weather simulated frame, haze effect in left dog-jump category and rain effect in right scooter-black category. Label-D shows the output of our approach which segments referred object well in challenging weather scenario.

5.3 Discussion

Environmental noise compromises visibility in video frames, thereby negatively affecting the segmentation task. This work proposes the WARMOS approach, which removes weather noise such as haze and rain to enhance segmentation output. We utilized state-of-the-art deep learning-based haze and rain removal methods to mitigate the negative effects of noise. We simulated haze and rain conditions on the DAVIS-2017 dataset to train and fine-tune the models. The addition of weather noise removal techniques improved the performance of the state-of-the-art SgMg segmentation model. Our experiments on the DAVIS-2017 dataset showed that while the SgMg model's performance was hindered by environmental noise, WARMOS successfully restored video clarity, improving the mean J&F scores from 58.71 to 60.50 in haze conditions and from 61.00 to 61.05 in rain-affected scenarios. This improvement highlights the importance of noise removal in maintaining segmentation accuracy in adverse weather conditions.

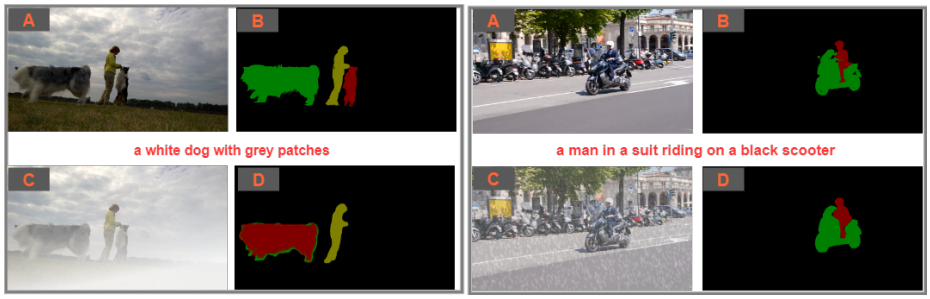

Fig. 3: A: DAVIS-2017 Video frame, B: Ground truth, C: Weather noise simulated image, Haze in left and Rain in right, D: Output of WARMOS approach based on text query highlighted in red color.

6 Conclusion

In this work, we have demonstrated the challenges posed by environmental noise such as haze and rain in the task of Referred Video Object Segmentation (RVOS). Weather conditions introduce pixel variance, which hampers the segmentation performance by obscuring object appearances and disrupting temporal and spatial consistency. To address this, we simulated haze and rain effects in video sequences and proposed a robust noise removal framework to mitigate

these disturbances, thereby improving segmentation performance. Our approach leverages the state-of-the-art Spectrum-guided Multigranularity (SgMg) model for video object segmentation, which is guided by text-based queries. We observed that the introduction of environmental noise in the DAVIS-2017 dataset resulted in significant degradation in segmentation accuracy. To combat this, we employed Light-DehazeNet (LD-Net) method that effectively removes weather noise. Fine-tuning LD-Net for haze and rain removal significantly improved the segmentation performance on weather-affected videos. The effectiveness of our WARMOS framework was demonstrated through both quantitative and qualitative results. In conclusion, our approach effectively mitigates the adverse effects of environmental noise on RVOS tasks by ensuring that the performance remains optimal even under challenging real-world conditions.

References

1. Bo Miao, Mohammed Bennamoun, Yongsheng Gao, and Ajmal Mian. Spectrum-guided multi-granularity referring video object segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 920–930, 2023.
2. Murari Mandal, Prafulla Saxena, Santosh Kumar Vipparthi, and Subrahmanyam Murala. Candid: Robust change dynamics and deterministic update policy for dynamic background subtraction. In *IEEE/ICPR*, pages 2468–2473, 2018.
3. Prafulla Saxena, Kuldeep Biradar, Dinesh Kumar Tyagi, and Santosh Kumar Vipparthi. Richex: A robust inter-frame change exposure for segmenting moving objects. In *IEEE CVF/ICIP*, pages 2172–2176. IEEE, 2022.
4. Prashant W Patil, Kuldeep M Biradar, Akshay Dudhane, and Subrahmanyam Murala. An end-to-end edge aggregation network for moving object segmentation. In *proceedings of the IEEE/CVF CVPR*, pages 8149–8158, 2020.
5. Prashant W Patil and Subrahmanyam Murala. MSFgNet: A novel compact end-to-end deep network for moving object detection. *IEEE Transactions on Intelligent Transportation Systems*, 20:4066–4077, 2018.
6. Thangarajah Akilan, Qingming Jonathan Wu, Amin Safaei, Jie Huo, and Yimin Yang. A 3D CNN-LSTM-based image-to-image foreground segmentation. *IEEE Transactions on Intelligent Transportation Systems*, 21:959–971, 2019.
7. Jingchun Cheng, Yi-Hsuan Tsai, Wei-Chih Hung, Shengjin Wang, and Ming-Hsuan Yang. Fast and accurate online video object segmentation via tracking parts. In *Proceedings of the IEEE CVPR*, pages 7415–7424, 2018.
8. Ping Hu, Jun Liu, Gang Wang, Vitaly Ablavsky, Kate Saenko, and Stan Sclaroff. Dipnet: Dynamic identity propagation network for video object segmentation. In *Proceedings of the IEEE/CVF WACV*, pages 1904–1913, 2020.
9. Jordi Pont-Tuset, Federico Perazzi, Sergi Caelles, Pablo Arbeláez, Alex Sorkine-Hornung, and Luc Van Gool. The 2017 davis challenge on video object segmentation. *arXiv preprint arXiv:1704.00675*, 2017.
10. Miriam Bellver, Carles Ventura, Carina Silberer, Ioannis Kazakos, Jordi Torres, and Xavier Giro-i Nieto. A closer look at referring expressions for video object segmentation. *Multimedia Tools and Applications*, 82(3):4419–4438, 2023.
11. Jiannan Wu, Yi Jiang, Peize Sun, Zehuan Yuan, and Ping Luo. Language as queries for referring video object segmentation. In *Proceedings of the IEEE/CVF CVPR*, pages 4974–4984, 2022.

12. Adam Botach, Evgenii Zheltonozhskii, and Chaim Baskin. End-to-end referring video object segmentation with multimodal transformers. In *Proceedings of the IEEE/CVF CVPR*, pages 4985–4995, 2022.
13. Prafulla Saxena, Susim Mukul Roy, Dinesh Kumar Tyagi, Santosh Kumar Vipparthi, Subrahmanyam Murala, and R Balasubramanian. Refmos: A robust referred moving object segmentation framework based on text query. In *2024 IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pages 1–7. IEEE, 2024.
14. Linwei Ye, Mrigank Rochan, Zhi Liu, and Yang Wang. Cross-modal self-attention network for referring image segmentation. In *IEEE/CVF conference on computer vision and pattern recognition*, pages 10502–10511, 2019.
15. Pierre-Luc St-Charles, Guillaume-Alexandre Bilodeau, and Robert Bergevin. Subsense: A universal change detection method with local adaptive sensitivity. *IEEE Transactions on Image Processing*, 24(1):359–373, 2014.
16. Kaiming He, Jian Sun, and Xiaoou Tang. Single image haze removal using dark channel prior. *IEEE transactions on pattern analysis and machine intelligence*, 33(12):2341–2353, 2010.
17. Hayat Ullah, Khan Muhammad, Muhammad Irfan, Saeed Anwar, Muhammad Sajjad, Ali Shariq Imran, and Victor Hugo C de Albuquerque. Light-dehazenet: a novel lightweight cnn architecture for single image dehazing. *IEEE transactions on image processing*, 30:8968–8982, 2021.
18. Gaofeng Meng, Ying Wang, Jiangyong Duan, Shiming Xiang, and Chunhong Pan. Efficient image dehazing with boundary constraint and contextual regularization. In *IEEE international conference on computer vision*, pages 617–624, 2013.
19. Bolun Cai, Xiangmin Xu, Kui Jia, Chunmei Qing, and Dacheng Tao. Dehazenet: An end-to-end system for single image haze removal. *IEEE transactions on image processing*, 25(11):5187–5198, 2016.
20. Boyi Li, Wenqi Ren, Dengpan Fu, Dacheng Tao, Dan Feng, Wenjun Zeng, and Zhangyang Wang. Benchmarking single-image dehazing and beyond. *IEEE Transactions on Image Processing*, 28(1):492–505, 2018.
21. Kshitiz Garg and Shree K Nayar. Vision and rain. *International Journal of Computer Vision*, 75:3–27, 2007.
22. Xueyang Fu, Jiabin Huang, Delu Zeng, Yue Huang, Xinghao Ding, and John Paisley. Removing rain from single images via a deep detail network. In *IEEE conference on computer vision and pattern recognition*, pages 3855–3863, 2017.
23. He Zhang and Vishal M Patel. Density-aware single image de-raining using a multi-stream dense network. In *IEEE conference on computer vision and pattern recognition*, pages 695–704, 2018.
24. Dezhuang Li, Ruoqi Li, Lijun Wang, Yifan Wang, Jinqing Qi, Lu Zhang, Ting Liu, Qingquan Xu, and Huchuan Lu. You only infer once: Cross-modal meta-transfer for referring video object segmentation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 1297–1305, 2022.
25. Ning Xu, Linjie Yang, Yuchen Fan, Dingcheng Yue, Yuchen Liang, Jianchao Yang, and Thomas Huang. Youtube-vos: A large-scale video object segmentation benchmark. *arXiv preprint arXiv:1809.03327*, 2018.
26. Kuldeep Biradar, Sachin Dube, and Santosh Kumar Vipparthi. Dearest: deep convolutional aberrant behavior detection in real-world scenarios. In *2018 IEEE 13th international conference on industrial and information systems (ICIIS)*, pages 163–167. IEEE, 2018.

27. Kuldeep Marotirao Biradar, Ayushi Gupta, Murari Mandal, and Santosh Kumar Vipparthi. Challenges in time-stamp aware anomaly detection in traffic videos. *arXiv preprint arXiv:1906.04574*, 2019.
28. Kuldeep Marotirao Biradar, Murari Mandal, Sachin Dube, Santosh Kumar Vipparthi, and Dinesh Kumar Tyagi. Triplet-set feature proximity learning for video anomaly detection. *Image and Vision Computing*, 150:105205, 2024.
29. Kuldeep Marotirao Biradar, Ayushi Gupta, Murari Mandal, and Santosh Kumar Vipparthi. Challenges in time-stamp aware anomaly detection in traffic videos. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2019.
30. Sachin Dube, Kuldeep Biradar, Santosh Kumar Vipparthi, and Dinesh Kumar Tyagi. Mag-net: A memory augmented generative framework for video anomaly detection using extrapolation. In *International Conference on Computer Vision and Image Processing*, pages 426–437. Springer, 2021.
31. He Zhang, Vishwanath Sindagi, and Vishal M Patel. Image de-raining using a conditional generative adversarial network. *IEEE transactions on circuits and systems for video technology*, 30(11):3943–3956, 2019.
32. Brett Koonce and Brett Koonce. Resnet 50. *Convolutional neural networks with swift for tensorflow: image recognition and dataset categorization*, pages 63–72, 2021.