Adversarial Weather-Resilient Image Retrieval: Enhancing Restoration using Captioning for Robust Visual Search

Prem Shanker Yadav, Kushall Singh, Dr.Dinesh Kumar Tyagi, and Dr.Ramesh Babu Battula

Malaviya National Institute of Technology, Jaipur, 302017, Rajasthan, India

Abstract. Accurate image retrieval in real-world scenarios is often hampered by degraded or noisy images, particularly those affected by adverse weather conditions such as rain, fog, or snow. Traditional retrieval methods that rely solely on feature extraction struggle to handle these degraded inputs and image captioning models are similarly limited in their ability to interpret distorted images. To address these challenges, we propose a novel framework that integrates image restoration with image captioning to create a robust image retrieval system capable of handling images degraded by adverse weather. Additionally, we introduce an integrated loss function to optimize restoration and captioning processes for degraded images. Our system enhances retrieval performance in challenging weather conditions by leveraging improved visual content alongside semantic context. Evaluations on Flicker8k dataset demonstrate that our approach significantly outperforms traditional image retrieval systems, particularly in scenarios where weather-induced degradation presents a challenge.

Keywords: Feature Differences loss · Unseen Changes · Textual Significance

1 Introduction

In recent years, advancements in image retrieval systems have significantly improved the ability to search for and retrieve images based on their content. However, these systems often face challenges such as adverse weather conditions, which can severely impact image quality and the effectiveness of traditional image retrieval methods. Uncertain conditions like rain, blur, fog, noise, snow, etc. can obscure key features, making it difficult for conventional models to accurately interpret and retrieve relevant images [26].

Image restoration techniques have become a potential solution to the problem of improving the visual quality of degraded images and enhancing the performance of image retrieval systems. Recent studies have shown that image restoration is effective in mitigating the effects of adverse weather conditions, allowing for clearer and more accurate image analysis [30]. In particular, image restoration methods that use deep learning models have shown promising results in reconstructing high-quality images from degraded inputs [13, 17].

In parallel, image captioning has proven to be a valuable tool for understanding and describing the content of images. By generating textual descriptions of images, image captioning systems facilitate more nuanced image retrieval by enabling semantic search based on the content of the images [15, 27]. The combination of image restoration and captioning offers a novel approach to enhance image retrieval systems, particularly in challenging conditions [1]. Figure 1 demonstrates how textual descriptions of images can facilitate the retrieval of scenes under adverse conditions. Captioning significantly improves the accuracy of this retrieval process.

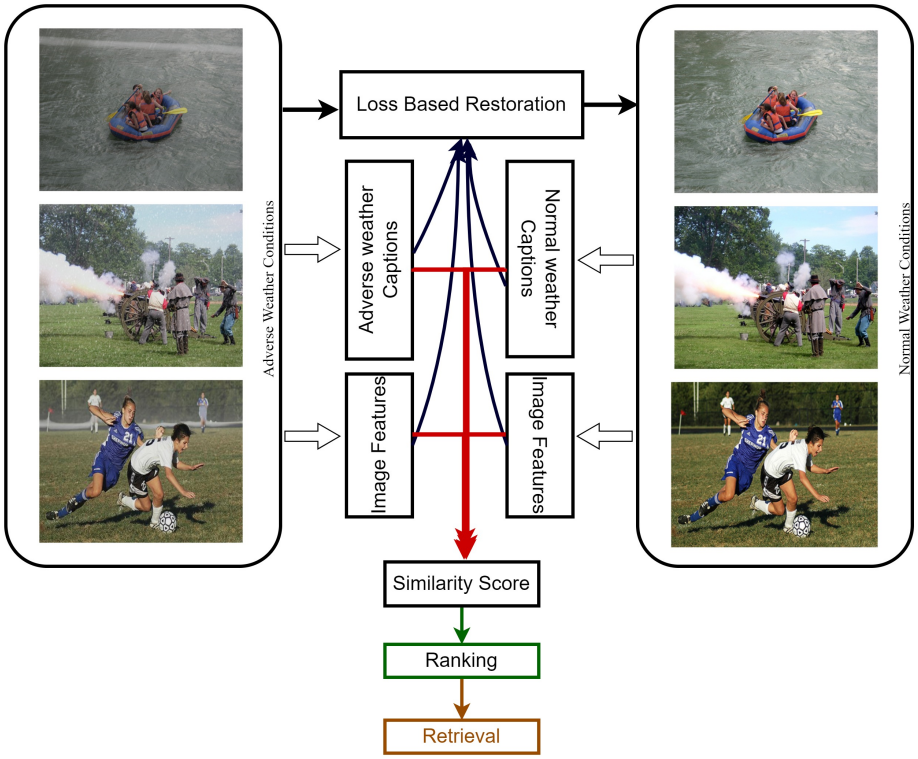

Fig. 1: Image retrieval using restoration by captioning and using Loss Based Restoration on weather-degraded images

This work proposes a novel framework that integrates image restoration with image captioning to improve image retrieval performance under adverse weather conditions.

Our approach involves the following key contributions:

- **Integrated Loss Functions for Image Restoration:** We utilize the integrated loss function to restore images degraded by adverse weather conditions, thereby improving image quality, clarity, and feature visibility.
- **Proposed Captioning-based Restoration:** We integrate deep learning-based image captioning to retrieve natural images with minimal synthetic image generation.
- **Integrated Retrieval-Restoration Framework:** We propose a framework that merges the restored images with their generated captions, thereby improving retrieval accuracy, particularly in challenging scenarios where conventional methods are less effective.

By integrating these components, our framework aims to provide a robust solution for image retrieval in adverse weather conditions, offering significant improvements over existing methods. Our proposed approach addresses the limitations of current systems and provides a more effective means of retrieving relevant images in challenging environments.

We organize the remaining sections in the following order: Section 2 describes a previously devised literature review. In Section 3, The designed framework for the restoration and integration with image retrieval is presented. Section 4 presents the outcome of the developed technique. Section 5 presents the conclusion.

2 Literature Review

The development of image retrieval systems has been an important field of study in computer vision. Traditional image retrieval systems have used feature extraction techniques such as (Scale-Invariant Feature Transform) and HOG (Histogram of Oriented Gradients) to represent visual content. However, these algorithms are constrained when photos are damaged by external variables such as poor weather, resulting in erroneous or partial feature extraction [34]. The paper [21] presented a method for estimating hurricane rain rates using SAR images analyzed through an Artificial Neural Network (ANN). The ANN helps interpret complex radar data to provide accurate rain rate measurements, which are important for weather forecasting and understanding hurricane dynamics. The paper [23] presented a novel approach for retrieving nighttime videos using a Temporal Weighting Appearance-Aligned Network. This network improves the retrieval process by addressing temporal variations and appearance changes in low-light conditions. By incorporating these aspects, the proposed method aims to enhance the accuracy and relevance of video retrieval in challenging nighttime scenarios. Extreme noise or low-quality nighttime videos could diminish the network's effectiveness, and it may struggle to adapt to novel scenarios not covered in the training data. Deep learning-based image retrieval models have recently evolved, capable of learning improved feature representations using Convolutional Neural Networks (CNNs) and Vision Transformers. Despite their gains, these approaches still struggle in difficult situations such as rain, fog, and snow,

when critical visual features are frequently concealed, resulting in poor retrieval performance [8].

Adverse weather conditions, such as rain, fog, and snow, can significantly degrade the visual quality of images by introducing distortions like occlusions, motion blur, and low contrast, which can obscure important features. Traditional image retrieval systems often fail to properly interpret images affected by these factors, leading to misclassifications or the failure to retrieve relevant results. Studies have shown that weather-induced degradation reduces the performance of many feature extraction and classification techniques in image retrieval [5]. As a result, there is an increasing demand for more robust retrieval systems capable of restoring image clarity and mitigating weather-induced distortions to enhance the accuracy of retrieval [12, 33].

Image restoration has been extensively studied as a solution to address image degradation caused by adverse weather conditions. Techniques such as dehazing [3], rain removal [7], and super-resolution have shown their effectiveness in restoring images to a clearer state, making them more suitable for subsequent image analysis tasks, including image retrieval. Early works, like He et al.'s dark channel prior method [11] for image dehazing, paved the way for more advanced restoration techniques. This method introduced the dark channel prior, a simple yet effective image prior for single-image haze removal. By leveraging the observation that haze-free images tend to have very low-intensity pixels in at least one color channel, the method estimates haze thickness and recovers high-quality, haze-free images, with the added benefit of producing a useful depth map as a byproduct.

The field of image restoration has significantly evolved from CNN-based methods [25, 29, 32] to transformer-based approaches [6, 20], which excel in capturing long-range dependencies and improving restoration performance. Techniques like window-based approaches [18] and transposed attention [28] have been employed to manage computational efficiency while maintaining restoration quality. Unified models, such as IPT [4] and AirNet [16], aim to handle multiple degradations, though their effectiveness remains limited across diverse tasks [22, 31]. On the other hand, prior-based methods, that utilize external information like high-resolution images or pre-trained models have been instrumental in improving restoration results [14, 19]. More recent approaches, such as text-based priors combined with image-based priors, have introduced a new paradigm, enabling models to learn degradation information at a textual level, thereby providing clean guidance for enhanced restoration performance [2].

Parallel to image restoration, image captioning has become an invaluable tool for describing the content of images in natural language. Image captioning models generate textual descriptions that capture the objects, scenes, and activities depicted in an image, facilitating the semantic search for image retrieval tasks. Traditional image retrieval systems rely heavily on low-level features, while captioning introduces a higher-level semantic understanding of image content. Karpathy and Fei-Fei [15] introduced a method for aligning image regions with natural language descriptions, which greatly improved the accuracy of image

caption matching. This work laid the foundation for more advanced models like the Show and Tell model [24] and the attention-based Show, Attend and Tell model [27], which further enhanced the capability of captioning systems to provide meaningful, context-aware descriptions of images. The integration of such captioning techniques into image retrieval systems enables more robust search capabilities, even under challenging conditions like those posed by adverse weather.

The combination of image restoration and captioning represents a novel approach to image retrieval. By first restoring the visual quality of images affected by weather conditions and then generating detailed captions describing the content, such systems can significantly improve retrieval accuracy in scenarios where traditional methods fall short. Anderson et al. [1] introduced a bottom-up and top-down attention mechanism for image captioning, which has the potential to be combined with restoration techniques to enhance the retrieval process. Wang et al. [25] explored the integration of restoration techniques with image recognition systems, suggesting that restoring image quality can improve both recognition and retrieval outcomes. However, this combined approach has yet to be fully explored in the context of adverse weather conditions, presenting a gap in the current literature. While considerable progress has been made in both image restoration and captioning individually, the integration of these two approaches for robust image retrieval under adverse weather conditions has received limited attention. Existing image retrieval models primarily focus on improving feature extraction but often fail to address image degradation caused by environmental factors. Similarly, while image captioning models can provide semantic descriptions, their performance is hindered when operating on degraded images.

Our work focuses on addressing these limitations by proposing a novel framework that integrates image restoration with deep learning-based image captioning to enhance image retrieval in adverse weather conditions.

3 Methodology

Our framework combines advanced image restoration techniques with deep learning-based image captioning to improve image retrieval in challenging weather conditions. This approach enhances both image clarity and semantic understanding, resulting in more precise and meaningful retrieval outcomes.

The complete architecture of our framework is depicted in Figure 2.

3.1 Dataset Preparation

We use a dataset comprising normal ($\mathcal{I}_n$) and weather-degraded ($\mathcal{I}_w$) images, along with their respective captions ($\mathcal{C}$). Each pair $(\mathcal{I}_n, \mathcal{I}_w)$ forms the basis for learning the difference in feature distributions caused by weather degradation. The dataset is represented as follows:

$$\mathcal{D} = (\mathcal{I}_n, \mathcal{I}_w, \mathcal{C}) \tag{1}$$

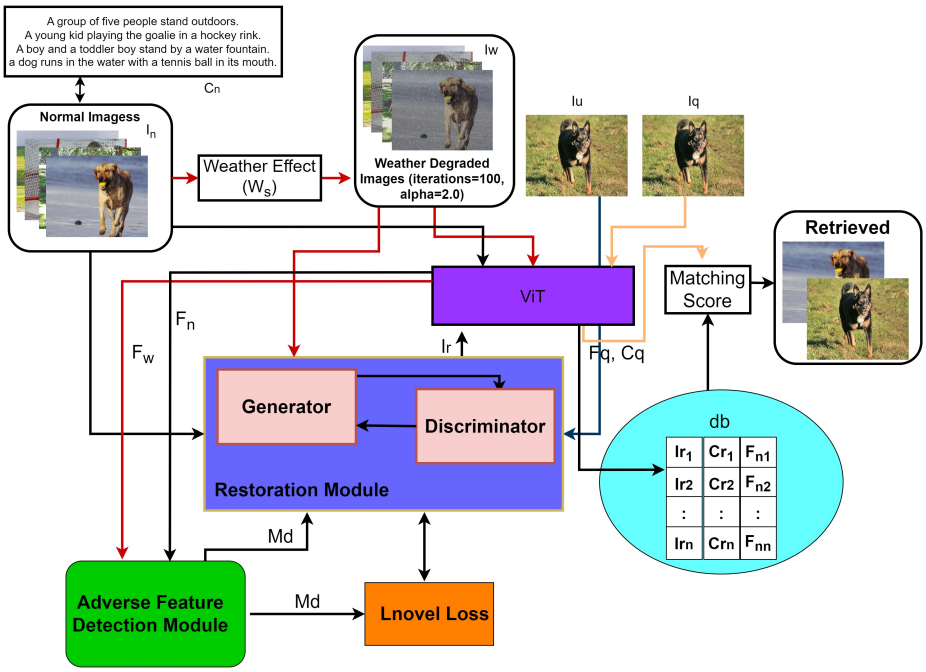

Fig. 2: Proposed Framework for image retrieval using restoration and captioning

To get adverse weather conditions, We apply various weather simulation effects [9] $\mathcal{W}_s$ to normal images. This generates synthetic weather-degraded images:

$$\mathcal{I}_w = \mathcal{W}_s(\mathcal{I}_n), \quad \mathcal{W}_s \in \text{rain}, \text{snow}, \text{fog} \tag{2}$$

where $\mathcal{I}_n$ and $\mathcal{I}_w$ are the images, and $\mathcal{C}$ is the corresponding caption. Figure 3 shows normal and degraded images.

Fig. 3: Normal and degraded images due to fog, rain, and snow

3.2 Feature Extraction

We utilize a dual-input feature extraction module that processes $\mathcal{I}_n$ and $\mathcal{I}_w$. A Vision Transformer (ViT) extracts feature representations for each image:

$$\mathcal{F}_n = \text{ViT}(\mathcal{I}_n), \quad \mathcal{F}_w = \text{ViT}(\mathcal{I}_w) \tag{3}$$

Where $\mathcal{F}_n$ and $\mathcal{F}_w$ are the feature vectors of the normal and weather-degraded images, respectively. The Vision Transformer captures high-level features crucial for semantic understanding and restoration. It significantly contributes to the embedding process and enhances the quality of caption generation. We are using BLIP (Bootstrapping Language-Image Pre-training) to generate captions.

3.3 Adverse Feature Detection

Using the extracted features $\mathcal{F}_w$, the Adverse Feature Detection Module identifies specific regions or distortions related to adverse weather effects. Adverse Feature Detection Module is a layered structure of convolution shown in the Figure 4. This produces an adverse degradation map $\mathcal{M}_d$:

$$\mathcal{M}_d = \text{AdverseNet}(\mathcal{F}_w) \tag{4}$$

Where $\mathcal{M}_d$ highlights the areas of $\mathcal{I}_w$ most affected by weather conditions, guiding the restoration process.

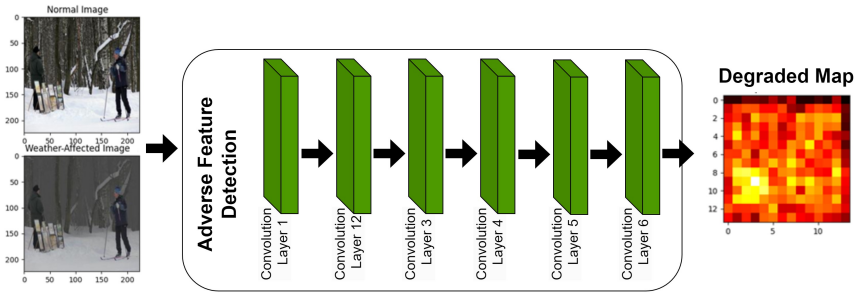

Fig. 4: Layered structure of AdverseNet

3.4 Image Restoration

The restoration process is driven by a dynamic restoration network $\mathcal{R}$ Generative Adversarial Networks (GAN-UNet), which restores $\mathcal{I}_w$ using the degradation map $\mathcal{M}_d$ and outputs a restored image $\hat{\mathcal{I}}w$. The network optimizes a novel composite loss function $\mathcal{L}$novel composed of the following key components:

– **Feature Loss** ($\mathcal{L}_f$): Ensures the restored image $\hat{\mathcal{I}}_w$ retains the visual characteristics of the original image $\mathcal{I}_n$.

$$\mathcal{L}_f = |\mathcal{F}_n - \text{ViT}(\hat{\mathcal{I}}_w)|^2 \tag{5}$$

– **Captioning Loss** ($\mathcal{L}_c$): Aligns the generated caption $\hat{\mathcal{C}}_w$ of the restored image with its ground truth caption $\mathcal{C}$.

$$\mathcal{L}_c = \text{CrossEntropy}(\mathcal{C}, \hat{\mathcal{C}}_w) \tag{6}$$

– **Adverse Feature Loss** ($\mathcal{L}_a$): Minimizes weather-induced distortions in the restored image based on the degradation map $\mathcal{M}_d$.

$$\mathcal{L}_a = \|\mathcal{M}_d\|^2 \tag{7}$$

– **Semantic Consistency Loss** ($\mathcal{L}_{\text{sc}}$): A novel loss component ensuring that the restored image retains not only visual but also semantic consistency with the original caption. It measures the similarity between captions generated for both $\mathcal{I}_n$ and $\hat{\mathcal{I}}_w$:

$$\mathcal{L}_{\text{sc}} = 1 - \text{CosineSimilarity}(\hat{\mathcal{C}}_n, \hat{\mathcal{C}}_w) \tag{8}$$

– **Adaptive Restoration Loss** ($\mathcal{L}_{\text{ar}}$): An adaptive loss that dynamically adjusts restoration based on the severity of weather degradation. This loss weights the restoration error inversely to the clarity level of $\mathcal{I}_w$:

$$\mathcal{L}_{\text{ar}} = \frac{\|\hat{\mathcal{I}}_w - \mathcal{I}_n\|^2}{1 + \|\mathcal{M}_d\|} \tag{9}$$

The total novel loss function used to train the restoration model is:

$$\mathcal{L}_{\text{novel}} = \lambda_f \mathcal{L}_f + \lambda_c \mathcal{L}_c + \lambda_a \mathcal{L}_a + \lambda_{\text{sc}} \mathcal{L}_{\text{sc}} + \lambda_{\text{ar}} \mathcal{L}_{\text{ar}} \tag{10}$$

where λ_f, λ_c, λ_a, λ_{sc}, and λ_{ar} are hyperparameters controlling the weight of each loss component.

3.5 Image Captioning and Feedback Mechanism

After restoration, a transformer-based image captioning model generates a detailed caption $\hat{\mathcal{C}}_w$ for the restored image $\hat{\mathcal{I}}_w$. The caption feedback loop ensures that the generated caption aligns with the original context, helping to further refine the restoration process. The caption feedback is incorporated into the total loss function:

$$\mathcal{L}_{\text{novel}} \leftarrow \mathcal{L}_{\text{novel}} + \lambda_c \cdot \text{CaptionFeedback}(\mathcal{C}, \hat{\mathcal{C}}_w) \tag{11}$$

This feedback loop dynamically adjusts the restoration process, improving semantic consistency and contextual accuracy.

3.6 Feature-Driven Adverse Detection for Unseen Images

For unseen images $\mathcal{I}_u$, the model predicts the presence of adverse weather effects by comparing the features of the unseen image $\mathcal{F}_u$ with learned normal image features $\mathcal{F}_n$:

$$\mathcal{F}_u = \text{ViT}(\mathcal{I}_u) \tag{12}$$

The model generates a prediction of adverse features $\hat{\mathcal{M}}_d$ if weather degradation is detected. The restoration network then restores the image accordingly:

$$\mathcal{I}_r u = \mathcal{R}(\mathcal{I}_u, \mathcal{F}_u, \hat{\mathcal{M}}_d) \tag{13}$$

3.7 Image Retrieval

Once the restoration is complete, the restored image I_r and its caption $\mathcal{C}_r$ are stored in an index (db). During the retrieval phase, a query image $\mathcal{I}_q$ is matched with stored images based on the similarity of their features $\mathcal{F}_q$ and captions $\mathcal{C}_q$. The Matching-Score is calculated as:

$$\text{Similarity}(\mathcal{I}_q, \mathcal{I}_{\text{db}}) = \alpha \cdot \text{cosine\_similarity}(\mathcal{F}_q, \mathcal{F}_{\text{db}}) + \beta \cdot \text{Similarity\_Score}(\mathcal{C}_q, \mathcal{C}_{\text{db}}) \tag{14}$$

where α and β are weighting parameters, ensuring a balanced emphasis on feature and caption similarity. Similarity-Score is calculated using Term Frequency-Inverse Document Frequency (TF-IDF) and Bag of Words.

4 Experiments and Result Analysis

This section describes the experiments conducted to evaluate the performance of the proposed image retrieval framework under adverse weather conditions. The experiments are designed to assess the effectiveness of image restoration, caption generation, and retrieval quality using both quantitative metrics and qualitative analysis.

Dataset

The experiments were conducted on the Flicker8k dataset, which includes a diverse collection of images with corresponding captions. The dataset features images captured in various contexts, with captions provided to evaluate the semantic accuracy of the generated descriptions. There are 8,092 images and five captions describing each image. Since weather conditions are not explicitly labeled; we applied approaches [9] and implemented them to generate adverse weather conditions in Flicker8k images.

Evaluation Metrics We employ a range of metrics to evaluate the effectiveness of the image retrieval framework by BLUE, Precision, Recall, Peak Signal-to-Noise Ratio (PSNR), and Structural Similarity Index (SSIM).

4.1 Results and Analysis

Image Captioning Effect in Retrieval Table 1 presents a comparative analysis of the image captioning effect in image retrieval in adverse conditions. We used BLUE, Precision, Recall and F1-score to present the accuracy and relevance of the image retrieval system.

Table 1: Image Captioning Affected Retrieval.

Method	BLEU	Precision	Recall	F1-score
CNN+LSTM	20.02	0.743	0.703	0.717
CNN+Bi-LSTM	21.20	0.764	0.744	0.754
Deep CNN+LSTM [10]	22.3	-	-	-
Show and Tell [24]	27.1	-	-	-
ViT (our)	40.10	0.842	0.807	0.824

Image Restoration Performance Table 2 shows the quantitative performance of our proposed framework on image restoration tasks. The results are measured using PSNR and SSIM.

Table 2: Effect of various losses on Restoration Performance.

Method	PSNR	SSIM
Proposed Framework	27.94	0.91
$(\mathcal{L}_f)$	23.5	0.81
$(\mathcal{L}_f)$+$(\mathcal{L}_c)$+ $(\mathcal{L}_{sc})$	25.9	0.85

Ablation Study We conduct an ablation study to examine the contributions of individual components in our loss function. Table 3 presents the results of removing specific components and their impact on captioning performance.

Table 3: Ablation Study on loss Captioning Performance (BLEU).

Configuration	BLEU
Without Adaptive Restoration Loss $(\mathcal{L}_{\text{ar}})$	37
Without Captioning Loss $(\mathcal{L}_c)$	34
Without Semantic Consistency $(\mathcal{L}_{\text{sc}})$	33
Without Feature Loss $(\mathcal{L}_f)$	32

The ablation results confirm that each loss component plays a crucial role in improving the overall captioning performance. The Captioning Loss ($\mathcal{L}_c$) and Feature Loss ($\mathcal{L}_f$) contribute significantly to the quality of the generated captions.

Qualitative Analysis Figure 5 showcases a qualitative comparison of retrieved images under various weather conditions. The proposed framework retrieves images with high visual and semantic similarity, even in challenging weather conditions, compared to the baseline methods.

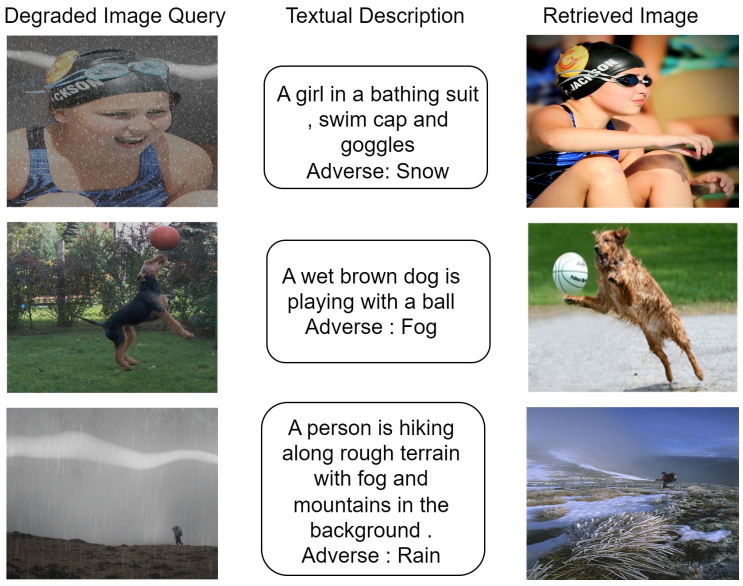

Fig. 5: The proposed model retrieves images that better match both the visual content and the captions under adverse weather conditions.

4.2 Discussion

The experimental results demonstrate the effectiveness of our network. The proposed framework significantly improves image quality by effectively mitigating weather-induced distortions. The high PSNR and SSIM values reflect this. Including the Caption Feedback Mechanism and maintaining semantic consistency in the loss function improves image retrieval quality. The model successfully retrieves semantically relevant images even in challenging weather scenarios. The ablation study confirms the importance of the key components of our loss function, highlighting the importance of feature and caption loss in achieving optimal performance.

Our framework integrates image restoration and deep learning-based image captioning to enhance image retrieval in adverse weather conditions. This

methodology improves both image clarity and semantic comprehension, which leads to more accurate and meaningful retrieval results.

5 Conclusion

We proposed a novel framework for robust image retrieval in adverse weather conditions, integrating image restoration and caption generation. By utilizing weather-degraded and normal image pairs, our approach restores images affected by conditions such as rain, snow, and fog while simultaneously generating accurate captions that preserve semantic consistency. Through extensive experiments on the Flickr8k dataset, we demonstrated that our model significantly enhances image retrieval performance in weather-affected images. The restoration component of the framework effectively mitigates weather distortions, while the transformer-based captioning model ensures that the generated captions align with the visual content. The combination of these two processes enables more accurate and semantically relevant image retrieval.

Future work will focus on fine-tuning the model for specific weather conditions and expanding the framework to larger and sequential event-based datasets for better generalization. Additionally, exploring weakly supervised learning could enhance the system's efficiency and applicability across various domains.

References

1. Anderson, P., He, X., Buehler, C., Teney, D., Johnson, M., Gould, S., Zhang, L.: Bottom-up and top-down attention for image captioning and visual question answering. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 6077–6086 (2018) 2, 5
2. Bai, Y., Wang, C., Xie, S., Dong, C., Yuan, C., Wang, Z.: Textir: A simple framework for text-based editable image restoration. arXiv preprint arXiv:2302.14736 (2023) 4
3. Cai, B., Xu, X., Jia, K., Qing, C., Tao, D.: Dehazenet: An end-to-end system for single image haze removal. IEEE transactions on image processing **25**(11), 5187–5198 (2016) 4
4. Chen, H., Wang, Y., Guo, T., Xu, C., Deng, Y., Liu, Z., Ma, S., Xu, C., Xu, C., Gao, W.: Pre-trained image processing transformer. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 12299–12310 (2021) 4
5. Cui, G., Ma, Q., Zhao, J., Yang, S., Chen, Z.: Image dehazing algorithm based on optimized dark channel and haze-line priors of adaptive sky segmentation. JOSA A **40**(6), 1165–1182 (2023) 4
6. Dosovitskiy, A.: An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929 (2020) 4
7. Fu, X., Huang, J., Zeng, D., Huang, Y., Ding, X., Paisley, J.: Removing rain from single images via a deep detail network. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 3855–3863 (2017) 4
8. Gkelios, S., Boutalis, Y., Chatzichristofis, S.A.: Investigating the vision transformer model for image retrieval tasks. In: 2021 17th International Conference on Distributed Computing in Sensor Systems (DCOSS). pp. 367–373. IEEE (2021) 4

9. Gupta, H., Kotlyar, O., Andreasson, H., Lilienthal, A.J.: Robust object detection in challenging weather conditions. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 7523–7532 (2024) 6, 9
10. Gupta, N., Jalal, A.S.: Integration of textual cues for fine-grained image captioning using deep cnn and lstm. Neural Computing and Applications **32**(24), 17899–17908 (2020) 10
11. He, K., Sun, J., Tang, X.: Single image haze removal using dark channel prior. IEEE transactions on pattern analysis and machine intelligence **33**(12), 2341–2353 (2010) 4
12. He, X., Jia, T., Li, J.: Learning degradation-aware visual prompt for maritime image restoration under adverse weather conditions. Frontiers in Marine Science **11**, 1382147 (2024) 4
13. Hodges, C., Bennamoun, M., Rahmani, H.: Single image dehazing using deep neural networks. Pattern Recognition Letters **128**, 70–77 (2019) 1
14. Jiang, Y., Chan, K.C., Wang, X., Loy, C.C., Liu, Z.: Robust reference-based super-resolution via c2-matching. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2103–2112 (2021) 4
15. Karpathy, A., Fei-Fei, L.: Deep visual-semantic alignments for generating image descriptions. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 3128–3137 (2015) 2, 4
16. Li, B., Liu, X., Hu, P., Wu, Z., Lv, J., Peng, X.: All-in-one image restoration for unknown corruption. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 17452–17462 (2022) 4
17. Li, R., Cheong, L.F., Tan, R.T.: Heavy rain image restoration: Integrating physics model and conditional adversarial learning. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 1633–1642 (2019) 1
18. Liang, J., Cao, J., Sun, G., Zhang, K., Van Gool, L., Timofte, R.: Swinir: Image restoration using swin transformer. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 1833–1844 (2021) 4
19. Lin, X., He, J., Chen, Z., Lyu, Z., Dai, B., Yu, F., Ouyang, W., Qiao, Y., Dong, C.: Diffbir: Towards blind image restoration with generative diffusion prior. arXiv preprint arXiv:2308.15070 (2023) 4
20. Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B.: Swin transformer: Hierarchical vision transformer using shifted windows. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 10012–10022 (2021) 4
21. Liu, Z., Ai, W., Zhao, X., Hu, S., Guo, C., Wang, L., Feng, M.: Retrieval of hurricane rain rate from sar images based on artificial neural network. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing (2024) 3
22. Potlapalli, V., Zamir, S.W., Khan, S.H., Shahbaz Khan, F.: Promptir: Prompting for all-in-one image restoration. Advances in Neural Information Processing Systems **36** (2024) 4
23. Ruan, W., Tao, Y., Ruan, L., Shu, X., Qiao, Y.: Temporal weighting appearance-aligned network for nighttime video retrieval. IEEE Signal Processing Letters **29**, 2008–2012 (2022) 3
24. Vinyals, O., Toshev, A., Bengio, S., Erhan, D.: Show and tell: A neural image caption generator. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 3156–3164 (2015) 5, 10
25. Wang, Y., Li, Y., Wang, G., Liu, X.: Multi-scale attention network for single image super-resolution. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5950–5960 (2024) 4, 5

26. Warburg, F., Miani, M., Brack, S., Hauberg, S.: Bayesian metric learning for uncertainty quantification in image retrieval. Advances in Neural Information Processing Systems **36** (2024) 1
27. Xu, K., Ba, J., Kiros, R., Cho, K., Courville, A., Salakhudinov, R., Zemel, R., Bengio, Y.: Show, attend and tell: Neural image caption generation with visual attention. In: International conference on machine learning. pp. 2048–2057. PMLR (2015) 2, 5
28. Zamir, S.W., Arora, A., Khan, S., Hayat, M., Khan, F.S., Yang, M.H.: Restormer: Efficient transformer for high-resolution image restoration. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 5728–5739 (2022) 4
29. Zamir, S.W., Arora, A., Khan, S., Hayat, M., Khan, F.S., Yang, M.H., Shao, L.: Learning enriched features for real image restoration and enhancement. In: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXV 16. pp. 492–511. Springer (2020) 4
30. Zhang, H., Patel, V.M.: Densely connected pyramid dehazing network. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 3194–3203 (2018) 1
31. Zhang, J., Huang, J., Yao, M., Yang, Z., Yu, H., Zhou, M., Zhao, F.: Ingredient-oriented multi-degradation learning for image restoration. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5825–5835 (2023) 4
32. Zhang, K., Zuo, W., Chen, Y., Meng, D., Zhang, L.: Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. IEEE transactions on image processing **26**(7), 3142–3155 (2017) 4
33. Zhang, Z., Zhang, S., Wu, R., Zuo, W., Timofte, R., Xing, X., Park, H., Song, S., Kim, C., Kong, X., et al.: Ntire 2024 challenge on bracketing image restoration and enhancement: Datasets methods and results. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 6153–6166 (2024) 4
34. Zheng, L., Yang, Y., Tian, Q.: Sift meets cnn: A decade survey of instance retrieval. IEEE transactions on pattern analysis and machine intelligence **40**(5), 1224–1244 (2017) 3