RESEARCH PAPER

Sensor-Guided Image Dehazing Using Self-Supervised Learning

Kavitha P[1*], Dr.M.Santhalakshmi[2], and Mr. N. Saravanan[3]

[1]*Department of Mathematics, Amrita School of Physical Sciences Coimbatore, Amrita Vishwa Vidyapeetham, India*
[2]*Assistant Professor, School of CS and IT, JAIN (Deemed-to-be) University, Bangalore-560069*
[3]*Mr. N. Saravanan, Assistant Professor, School of CS and IT, (Deemed-to-be) University, Bangalore-560069*
***Corresponding Author:** Kavitha P*

Received: 28th Feb, 2026; Revised: 6th March 2026; Accepted: 7th April, 2026; Available Online: 20th April, 2026

ABSTRACT

The problem of haze removal in real images is still challenging due to the limitations of the prior-based methods and the poor generality of the learned models on synthetic images. In this paper, the problem is addressed based on the relationship between visual haze and environmental factors. The Vision Transformer model is pre-trained via the self-supervised DINO approach on a large number of unlabeled hazy images; thus, the model learned the haze representations in a realistic way without the need for annotations.

In addition, the regression model is learned for the prediction of air quality factors such as PM2.5 and PM10 from the images. The estimated values are then used as a proxy for the intensity of the haze and are utilized in the proposed atmospheric scattering model to make the dehazing process adaptive. This method is based on image-based sensor information and thus provides a more adaptive approach.

The experimental results show that the proposed model is effective in estimating the level of pollution.

Keywords: Image dehazing, self-supervised learning, Vi- sion Transformer, DINO, air pollution estimation, PM2.5, PM10, atmospheric scattering, domain generalization.

How to cite this article: Kavitha P, Santhalakshmi M, Saravanan N. Sensor-Guided Image Dehazing Using Self-Supervised Learning. Int J Drug Deliv Technol. 2026;16(64s):1011-1020. DOI: 10.25258/ijddt.16.64s.109
Source of support: Nil.
Conflict of interest: None

I. INTRODUCTION

Atmospheric haze is one of the image degradation effects. This happens due to the scattering and absorption of light by the dust, smoke, and water present in the atmosphere. This scattering of light results in an image with low contrast and distorted color information. This results in a significant reduc- tion in the visibility of the image [1]. This image degradation is not only for human vision but also for computer vision. So, it is of great importance to remove the haze from the image, especially for the vision of self-driving cars.

The formation of the hazy image can be explained by the Atmospheric Scattering Model. This model was first proposed by McCartney [2] and then extended by Narasimhan and Nayar [3]. The image formed by the haze can be represented by the following equation: as:

$$I(x) = J(x) \cdot t(x) + A \cdot (1 - t(x)) \quad (1)$$

where $J(x)$ represents the scene radiance, $t(x)$ denotes the transmission map indicating the amount of light reaching the imaging sensor, and A represents the global atmospheric light. The transmission map is defined as:

$$t(x) = e^{-\beta \cdot d(x)} \quad (2)$$

where β represents the atmospheric attenuation coefficient and $d(x)$ represents the scene depth. The dehazing process can therefore be viewed as the problem of estimating the transmission map and recovering the underlying clear image. Traditional dehazing approaches generally involve the use of manually defined priors such as the Dark Channel Prior [4], which impose statistical constraints on natural images. Although these approaches have shown promising results in different conditions, they have also shown limitations in handling bright conditions or those that are not consistent with these priors. In recent years, deep learning-based approaches have shown promising results by directly learning the mapping from a hazy image to a clear image. However, these meth- ods are generally trained on synthetic datasets where haze is artificially generated. As a result, they suffer from poor generalization when applied to real-world hazy scenes due to

the domain gap between synthetic and natural data.

However, in recent times, self-supervised learning has been recognized as a new alternative for the effective learning of informative visual representations. For example, Vision Transformers have been recognized as having good potential for the effective learning of global structures and semantic relationships within images, especially when learned through approaches such as DINO. This is due to their application in the modeling of

*Author for Correspondence: Kavitha P

complex real-world phenomena such as haze in the atmosphere.

In this piece of work, we are trying to extend the domain of self-supervised learning by creating a connection with the environment. The pre-trained Vision Transformer model with the DINO approach [12] is used along with a large number of unlabelled hazy images obtained from the environment, which helps the model learn representations to describe the characteristics of the hazy scenes.

Using these learned representations, regression models can be learned to directly predict the air quality metrics like PM2.5, PM10, etc., from the images. These metrics, typically measured by environment sensors, offer quantitative measures for the concentration of particles within the atmosphere, di- rectly related to the haze.The direct prediction of these metrics from images shows their relevance to environment factors [16]. The major contributions of this work are summarized as

follows:

- A self-supervised learning framework based on Vision Transformers for effective learning of haze representations from real-world data.
- A regression model for predicting air quality indicators (PM2.5 and PM10) directly from images.
- A sensor-guided dehazing approach that integrates predicted air quality values into the atmospheric scattering model.
- Improved generalization performance on real-world hazy images compared to traditional and synthetic-data-driven methods.

II. RELATED WORKS

The task of image dehazing has always been a popular area of research. However, the earlier work in the area of image dehazing was related to the physics of the formation of images and the statistical properties of images. But the most popular and commonly used technique in the area of image dehazing is the Dark Channel Prior (DCP) technique, which was proposed by He et al. [4]. This technique was based on the assumption that in the case of natural outdoor images, there exist pixels that have low intensity in at least one of the color channels. However, the limitations of the technique in the areas of the sky have restricted the usage of the technique in the areas of image dehazing.

In order to overcome the limitations of the technique, alternative techniques related to prior have also been proposed in the area of image dehazing. For example, Zhu et al. [5] proposed a technique related to the Color Attenuation Prior for the estimation of the hazy images. However, the limitations of the technique in the areas of image dehazing are still a problem in the area of image dehazing. However, with the advent of deep learning techniques, CNNs are also used for image dehazing problems. One of the first CNNs used for image dehazing is DehazeNet [6], in which the haze features are learned directly from the image. Later, AOD-Net [7] and FFA-Net [8] models are used for end-to-end learning of the dehazing process. Although the performance of the model is good compared to existing image dehazing techniques, these models are trained using synthetic datasets like RESIDE [9], in which artificial haze is added. Hence, the performance of the model is poor for real haze images since there is a domain gap between synthetic and real haze images.

Recently, transformer-based models are also used for image dehazing problems to enhance the performance of image dehazing compared to CNN-based models. For this, in the proposed model of ”DehazeXL” by Chen et al. [17], the image dehazing problem is solved by applying a technique of ”patch tokenization” and ”global context fusion” of images. A Deep Learning Model Combining ResNet, DenseNet, and Attention Mechanisms by Umesh Kumar Lilhore, P. Kavitha [18], the image dehazing problem is solved by applying a combination of ”physical prior” and ”deep learning” techniques.

At the same time,”self-supervised learning” has also emerged as a prominent technique in learning image repre- sentation without requiring large amounts of labeled datasets.

Recently, in the study of ”SimCLR” [10] and ”transformer- based vision transformers” [11] in combination with ”DINO” [12], promising results have been achieved in learning global structures and semantics of images. The most prominent advantage of SSL-based approaches is that they do not require labelled datasets, making it highly suitable for practical appli- cations. However, the application of SSL-based approaches in the context of haze-related applications has just started. In the context of estimating the severity of the haze, which is considered one of the prominent tasks in the context of adaptive dehazing, the application of SSL-based approaches in the context of learning the representation of images has not been fully explored. In the context of conventional approaches for estimating the severity of the haze, the conventional features such as contrast, saturation, and transmission map statistics obtained by prior-based approaches are considered insufficient in explaining the characteristics of the haze.

Apart from the above, the relationship between the visual visibility and the air quality has also been extensively studied in the context of IoT-based environment monitoring systems. In this context, the conventional approaches usually employ a distributed sensor network that monitors the atmospheric conditions, including the PM2.5 and PM10 levels, in real-time. Such IoT-based approaches can effectively monitor the air quality and understand the occurrence of haze and its impact on the visual quality.

In this context, the conventional approaches have also been extended to employ machine learning approaches in under- standing the relationship between the environment

and the visual quality degradation. In this context, Li et al. [20] have effectively demonstrated the estimation of air quality indica- tors using images by employing machine learning approaches. This demonstrates a strong relationship between the particulate matter levels and the degradation in the images.

The results obtained in the context of the above problem statement suggest that the occurrence of haze has a strong relationship with the environment and cannot be considered a degradation in the images. Although the conventional ap- proaches employ IoT-based vision, the work in the paper is motivated to incorporate the air quality-inspired features in the context of the conventional approaches.

Based on these observations, this paper proposes a framework for self-supervised representation learning and environ- mental awareness. The proposed framework aims to bridge the gap between physical environments and visual perceptions by learning and predicting air quality factors such as PM2.5 and PM10 from images and incorporating them in dehazing.

III.METHODOLOGY

The proposed framework has three key components: self-supervised representation learning with Vision Transformers, air quality estimation with regression, and sensor-informed dehazing with the atmospheric scattering model. The overall process of the proposed method is conceptually depicted as follows: Image → SSL Encoder → PM Estimation →

Adaptive Dehazing. The overall framework of the proposed method is illustrated in Fig. 1.

A. Datasets Used

The framework that is proposed makes use of three different datasets, which have different roles in the framework's training.

Air Pollution Image Dataset: The air pollution image dataset, which contains data collected from India and Nepal, is used for learning the relationship that exists between the image and the sensor value. The images in the dataset are taken under certain environmental conditions like PM2.5 and PM10.

RESIDE Dataset: The RESIDE dataset is used for learning the dehazing process of images and the level of haziness. The dataset contains synthetic images, which are hazy, along with the ground truth images.

Scraped Web Images: The images collected using this method are used for self-supervised learning. The images collected using this method contain real images of haze.

B. Self-Supervised Representation Learning

For the input image $I(x)$, which is hazy, a Vision Trans- former, known as ViT, is used. The ViT is pre-trained using a self-supervised learning approach, which is known as DINO. Unlike the normal supervised learning approach, this approach helps the network learn the features without any labels. The labels are not needed because the network is using a student- teacher consistency.

Let $f_\theta(\cdot)$ denote the encoder function of the Vision Transformer. The output feature representation for an input image is given by:

$$z = f_\theta(I) \qquad (3)$$

These learned representations include the details of both the global structures and haze-related characteristics found in the real-world images.

C. Air Quality Estimation using Regression Head

In order to establish the relationship between visual haze and environmental factors, a regression model will be created to directly predict the air quality, such as PM2.5 and PM10, from the input images.

1) Data Preparation: The regression model is trained on the Air Pollution Image Dataset, which was obtained from In- dia and Nepal. Real images are associated with corresponding PM2.5 and PM10 values. Images are resized to a 224 × 224 dimension and are augmented by random flipping and color jittering for better generalization. To make the training process stable, standard score normal- ization is applied to the target variables:

2) Feature Extraction using SSL Encoder: To extract features from images, a convolutional neural network-based encoder is adopted. In this case, the ResNet-18 model is used along with pre-trained weights from the self-supervised learning (SSL) model.The model is referred to as:

$$z = f_\theta(I) \qquad (5)$$

where I is the input image and z is the feature embedding.

In order to keep the general representations learned in the pre-training of SSL and allow adaptation to specific tasks, the deeper layers of the encoder are used to perform the fine- tuning. The initial layers of the network are frozen to avoid overfitting and improve training efficiency. The lightweight MLP model was used as a regression model to project the features to air quality indicators.

The regression model used several fully connected layers with nonlinear activation functions, batch normalization, and dropout regularization. The structure of the regression model used in this paper is as follows:

- Fully connected layer (embedding → 256) + ReLU
- Batch normalization + dropout
- Fully connected layer (256 → 64) + ReLU
- Fully connected layer (64 → 2)

3) Training Strategy: The model is optimized by the Smooth L1 Loss Function, which makes the model robust to the presence of any outlier values:

L_{reg} = SmoothL1(y, $y\hat{}$) (6)

The AdamW optimizer with weight decays is used to optimize the model and avoid overfitting. Mixed precision training is also used to reduce the computation and memory footprint of the model on GPUs. To reduce the training time of the model, the initial few encoder layers of the model are frozen, and only the last few encoder and regression layers of the model are trained.

4) *Output and Integration:* The output of the optimized model is the normalized PM2.5 and PM10 values. This is then scaled to the actual scale and is used as the descriptor of the environment. This is then used as the adaptive parameter in the image restoration process.

D. Stage 3: Dual-Head Severity Fusion and Dehazing

To achieve robust dehazing under real-world conditions, a dual-head architecture is proposed that integrates both visual cues and environmental sensor information. The framework combines (i) a visual severity estimator derived from self- supervised learning and (ii) a sensor-based severity predictor derived from air quality measurements. These two sources are adaptively fused to guide a conditional dehazing network.

1) *Overview:* Given a hazy image I, the model estimates

$$\tilde{y} = \frac{y - \mu}{\sigma} \quad (4)$$

haze severity using two complementary branches:

- **Visual severity** $\hat{s}_v$: obtained from a frozen SSL-based

where y is the PM values, and μ and σ are the mean and standard deviations of the dataset.

encoder trained in Stage 2 using image features and transmission priors.

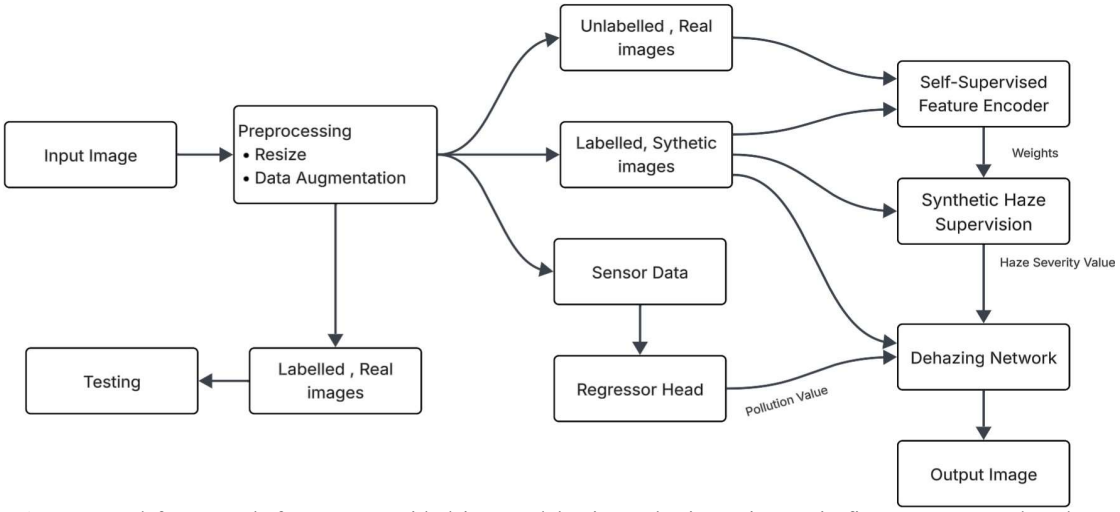

Fig. 1: Proposed framework for sensor-guided image dehazing. The input image is first preprocessed and passed through a self-supervised feature encoder that was trained on unlabelled real-world images. The regression module estimates air quality indicators such as PM2.5 and PM10. In parallel with the previous steps, synthetic haze supervision is used to inform the dehazing network. The estimated pollution levels are used as adaptive parameters in the dehazing process.

- **Sensor severity** $\hat{s}_s$: predicted using a regression head that takes air quality indicators (PM2.5, PM10) along with image features. These two estimates are combined using a learned gating mechanism to produce a final severity value $\hat{s}_f$, which condi- tions the dehazing network.

2) *Sensor-Based Severity Estimation:* The sensor head is a trainable regression module for predicting the level of haze based on both environmental and visual cues. Let's assume the feature representation learned from the SSL encoder to be represented by the variable z and the normalized values of PM2.5 and PM10 concentration to be represented by the variable $p = [p_{2.5}, p_{10}]$.

The sensor head computes:

$$\hat{s}_s = \sigma\, g_s([z, d, \phi(p)]) \quad (7)$$

where d denotes DCP-based features, $\phi(\cdot)$ is a projection of sensor inputs, and σ is the sigmoid activation ensuring output in $[0, 1]$.

3) *Adaptive Severity Fusion:* To integrate both visual and sensor-based predictions, an adaptive fusion gate is proposed. The gate will learn to assign weights to each of the sources based on feature confidence and sensor availability.

The fused severity is given by:

$$\hat{s}_f = w_v \cdot \hat{s}_v + w_s \cdot \hat{s}_s \quad (8)$$

where w_v and w_s are learned weights such that:

$$w_v + w_s = 1 \quad (9)$$

The weights are learned by a gating network that uses the SSL features, severities, and a flag for sensor availability as input. In the absence of sensors, the model will fall back to visual-based severity estimation.

4) *Severity Embedding and Conditioning:* The severity, when combined, is represented by a high-dimensional space through the sinusoidal function and the MLP:

$$e_s = \psi(\hat{s}_f) \quad (10)$$

The representation is then used by the conditioned dehazing network through the Feature-wise Linear Modulation (FiLM) and cross-attention mechanisms, allowing the network to adapt its operation according to the severity of the haze.

5) *Dehazing Network:* A U-Net-based architecture is employed for image restoration. The network takes the hazy image as input and is conditioned on:

- Severity embedding e_s
- SSL feature representation z
- DCP-based features d

The network outputs the restored image $\hat{J}$:

$$\hat{J} = F(I, e_s, z, d) \quad (11)$$

The cross-attention and FiLM layers interact with the global features of the SSL and local feature maps, respectively.

6) *Loss Function:* The model is trained end-to-end using a composite loss function defined as:

B. Result Analysis of Stage 2 (Haze Severity Regression)

The performance of the Stage 2 model is evaluated by using various regression performance metrics. The performance met-

$$L = \lambda_1 L_{pixel} + \lambda_2 L_{perc} + \lambda_3 L_{freq} + \lambda_4 L_{ssim} + \lambda_5 L_{sensor} + \lambda_6 L_{sev}$$

rics of the Stage 2 model are as follows: The Mean Absolute

where:

- L_{pixel} is the L1 reconstruction loss,
- L_{perc} is the perceptual loss using VGG features,
- L_{freq} enforces frequency-domain consistency,
- L_{ssim} improves structural similarity,

(12)

Error of the Stage 2 model is 0.0524. The R^2 score of the Stage 2 model is 0.6968. The Spearman rank correlation coefficient of the Stage 2 model is 0.8419.

The low value of the Mean Absolute Error indicates that the predicted values of the haze severity are closely correlated with the actual values. In addition, the positive value of the

- L_{sensor}, supervises the sensor head using ground-truth R^2 score indicates that a high proportion of the variance of the severity,
- L_{sev} aligns fused severity with transmission-based esti- mates.

To handle uncertainty, each sample is weighted inversely proportional to its predicted uncertainty.

7) *Training Strategy:* The Stage 2 encoder is frozen, and the representations are kept the same, while the sensor head, fusion gate, severity embedding, and dehazing network are optimized together with the AdamW optimizer and mixed precision training.

8) *Inference:* During inference, the model can operate in two modes:

- **With sensor data**: both visual and sensor severity are fused.
- **Without sensor data**: the model relies solely on visual estimation.

The final output is the dehazed image along with estimated severity and confidence measures.

IV. MODEL EVALUATION

A. Result Analysis of Stage 1 (Self-Supervised Learning)

During Stage 1, the self-supervised learning model is employed for the learning of significant visual representations from real-world hazy images without the use of label infor- mation. During this stage, the model is trained for 60 epochs with a loss of 3.02.

The training of the model by the SSL model cannot be considered under the evaluation parameters of accuracy or error rate; however, the rate at which the loss converges can be considered indicative of the success of the model in learning significant visual representations. The learning of significant visual representations by the model includes learning significant features related to haze images.

The feature embeddings of the learned feature representations can be considered indicative of the success of the model in learning significant visual representations for the haze severity estimation problem during Stage 2. The generalization of the model for a variety of real-world haze images can be considered indicative of the success of the self-supervised pre- training model in bridging the gap between synthetic and real- world data.

haze severity values is explained by the Stage 2 model across various images. Moreover, the high value of the Spearman correlation coefficient indicates that the relative order of the haze severity values is well preserved by the Stage 2 model.

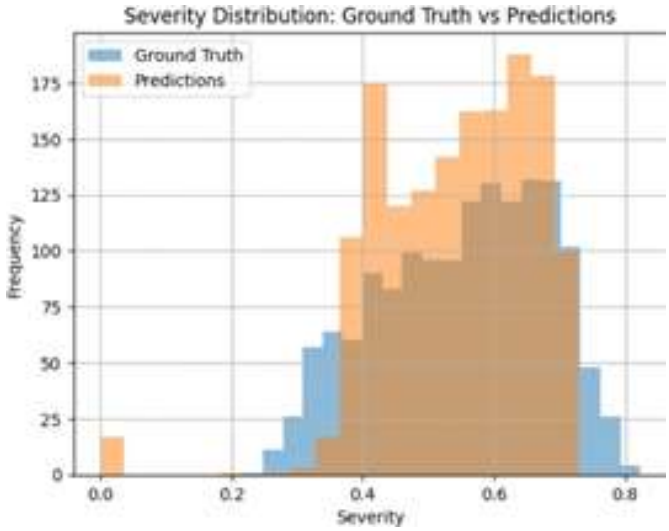

Fig. 2: Distribution of Ground Truth vs Predicted Severity

1) *Severity Distribution Analysis:* Fig. 2shows a compari- son of the distribution of the predicted severity and the ground truth. From the figure, it is evident that the two distributions have a similar trend, thus implying that the model has managed to capture the statistical nature of the haze severity. Though there is a slight variation in the two distributions, it is evident that the predicted distribution is similar to the ground truth, thus implying that there is no problem of prediction bias or collapse.

2) *Training Convergence Analysis:* Fig. 3 The graph above depicts the training and validation MAE for the epochs. From the graph, it is observed that the training and validation MAE show a declining trend, which is a good sign of learning. There is no major change in the training and validation loss, which is a good sign of generalization. There is no sign of overfitting in the model.

3) *Prediction Correlation Analysis:* Fig. 4 shows the scatter plot of the predicted severity and the ground truth. It is observed that the points are closely aligned with the diagonal line. This is a clear indication of the consistency between the predicted and the actual severity of the haziness. This alignment of the points with the diagonal line is an evidence of the model's capability of accurately estimating the severity of the haziness.

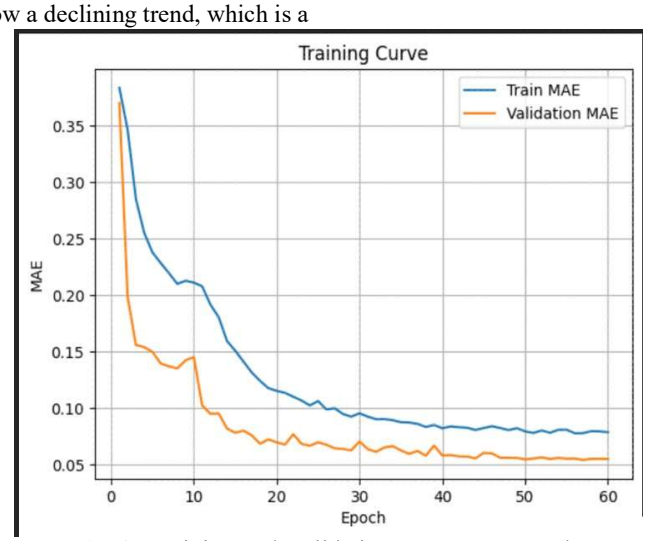

Fig. 3: Training and Validation MAE over Epochs

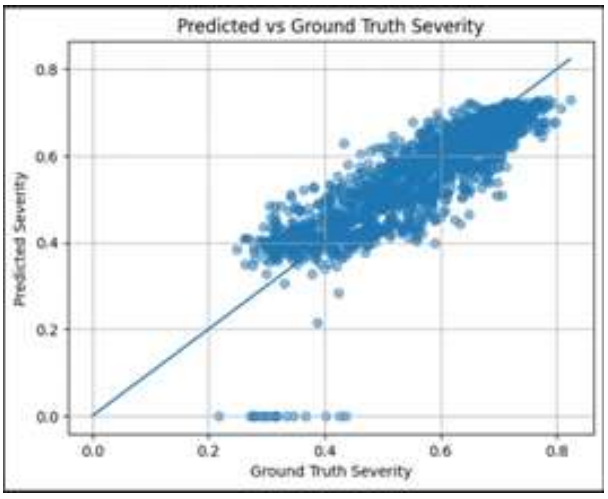

Fig. 4: Predicted vs Ground Truth Severity

Some level of dispersion is observed in the range of medium severity levels. This is understandable because of the inherent ambiguity in the perception of haziness.

4) *Discussion:* However, in order to comprehend the stability in the prediction, the standard deviation of the predicted value of the severity is compared to the standard deviation of the ground truth. The standard deviation of the predicted value of the severity is 0.1180, and the standard deviation of the ground truth is 0.1274. This clearly explains that the prediction collapse is not occurring; instead, a reasonable output is being generated by the prediction.

The results obtained from the prediction indicate that the regression model is able to comprehend the representation of the haze and is providing a reasonable prediction for the severity of the haze. The regression model can be utilized as a conditioning signal in the subsequent stages of the dehazing process.

C. Result Analysis of Sensor-Based AQI Regression

Apart from this, the regression model is also employed to directly calculate the air quality indices such as PM2.5 and PM10. The model is initialized with self-supervised weights. The model is then fine-tuned with the image-sensor pairs. The training of the model is carried out for 15 epochs with 12,240 data samples. From the graph, it can be observed that the model has good convergence characteristics.

The training loss of the model is decreasing constantly from the first epoch, in which the training loss of the model is 0.2347, to the last epoch, in which the training loss of the model is 0.0616. At the same time, the validation loss of the model is also decreasing constantly from 0.1794 to 0.0518.

From the graph, it can be observed that the training and validation loss of the model is decreasing constantly with minor changes in a few epochs. From the graph, it can also be observed that the training and validation loss of the model is almost similar. It can also be concluded that the model is not under any kind of overfitting conditions.

This is further confirmed by the low value of the MAE, which shows that the haze severity values are highly correlated with the predicted values. In addition, the positive value of the R^2 score shows that a large proportion of the variance of haze severity is explained by the Stage 2 model for different images. Moreover, the high value of the Spearman correlation coefficient shows that the relative order of haze severity is well preserved by the Stage 2 model.

The Smooth L1 Loss is helpful for achieving the robustness of the model with respect to outliers in the sensor data. In addition, the early encoder layers being frozen are helpful for retaining the information that has already been learned during the pre-training phase. Furthermore, the use of data augmentation techniques such as flipping and jittering is helpful for improving the performance of the model.

The results show that the model has learned the mapping between the visual features and air quality indicators well. Therefore, the model can be considered a good model for directly estimating the environmental conditions by using images. It can be considered a good source of information for haze-aware image enhancement as well.

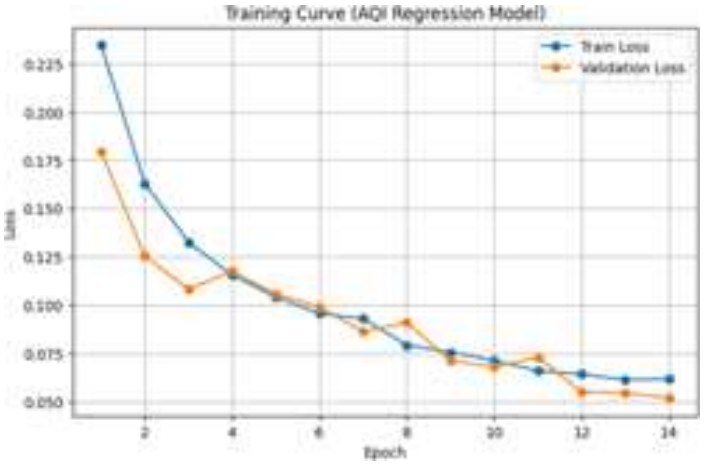

Fig. 5: Training and validation loss curves for the AQI regression model. The steady decrease in loss values indicates stable convergence, while the close overlap between training and validation curves suggests good generalization without significant overfitting.

D. Proposed Model Result Analysis

The final model, which combines the dual regressor head along with the self-supervised pretraining, showed good results

in the dehazing task. The model was able to achieve a Peak Signal-to-Noise Ratio (PSNR) between 22-28 dB, along with a Structural Similarity Index Measure (SSIM) ranging from

0.80 to 0.95.

The results showed that the proposed framework is able to achieve good results in dehazing images while maintaining the quality of the dehazed images. The high range of the SSIM index showed that the model was able to maintain the structural details of the images, while the PSNR results showed that the model was able to effectively remove the haze and noise. The use of the dual regressor mechanism allowed the model to effectively use the visual features and the sensor severity, which made the dehazing task more adaptive. The self-supervised pretraining stage made the model more generalizable, which allowed the model to perform well even in different real-world hazy conditions.

The proposed model was able to achieve a good trade-off between the results, which made the model applicable for the dehazing task.As shown in Table I, the proposed model achieved good results in all the evaluation metrics.

Table I: Performance of the Proposed Dual-Regressor De- hazing Model

Metric	Value
PSNR (dB)	19.77
SSIM	0.8131
MAE (Severity)	0.0630
R^2 Score	0.6968
Spearman Correlation (ρ)	0.8419

Fig. 6: Qualitative comparison of the proposed dehazing model. The dehazed output shows improved contrast, color restoration, and structural detail compared to the hazy input, closely matching the ground-truth image.

Fig. 6 some qualitative results of the proposed model are shown for outdoor and indoor scenes. In all scenes, the input image is hazy and has low visibility, contrast, and color information due to the scattering of light in the atmosphere.

The dehazed images show a considerable improvement in contrast and visual information. The structural details are well preserved by the model. For example, object boundaries and texture information are well recovered by the model. In addition, the severity level is close to the ground truth.

In terms of quantitative performance, the obtained results have PSNR values between 21 dB and 24 dB, and SSIM values

greater than 0.89. This shows that there is a good perceptual similarity between the obtained results and the ground-truth images.

The results obtained in this paper have confirmed that the incorporation of self-supervised feature learning and severity- aware conditioning is effective for dehazing images.

AUTHOR CONTRIBUTIONS

Kavitha. P: Conceptualization, Data curation, Methodology, Supervision, Writing–original draft, Writing–review & editing, methodology, Data curation, Model Implementation, Writing–original draft, Result interpretation, Writing– review & editing. All authors have read and approved the final manuscript.

ACKNOWLEDGEMENTS

The authors would like to thank Amrita Vishwa Vidyapeetham for providing the computational resources and research environment necessary to carry out this work. We also acknowledge the authors of the RESIDE dataset for providing large-scale synthetic hazy image data, and the contributors of the Air Pollution Image Dataset for making real-world air quality and image data publicly available, which supported the development and evaluation of the models used in this study.

REFERENCES

[1] K. Tang, J. Yang, and J. Wang, "Investigating haze-relevant features in a learning framework for image dehazing," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2014, pp. 2995–3002.

[2] E. J. McCartney, *Optics of the Atmosphere: Scattering by Molecules and Particles*. New York, NY, USA: Wiley, 1976.

[3] S. G. Narasimhan and S. K. Nayar, "Contrast restoration of weather degraded images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 6, pp. 713–724, 2003.

[4] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 12, pp. 2341–2353, 2011.

[5] Q. Zhu, J. Mai, and L. Shao, "A fast single image haze removal algorithm using color attenuation prior," *IEEE Transactions on Image Processing*, vol. 24, no. 11, pp. 3522–3533, 2015.

[6] B. Cai, X. Xu, K. Jia, C. Qing, and D. Tao, "DehazeNet: An end-to-end system for single image haze removal," *IEEE Transactions on Image Processing*, vol. 25, no. 11, pp. 5187–5198, 2016.

[7] B. Li, X. Peng, Z. Wang, J. Xu, and D. Feng, "AOD-Net: All-in- one dehazing network," in *Proc. IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 4770–4778.

[8] X. Qin, Z. Wang, Y. Bai, X. Xie, and H. Jia, "FFA-Net: Feature fusion attention network for single image dehazing," in *Proc. AAAI Conference on Artificial Intelligence*, vol. 34, no. 7, 2020, pp. 11908–11915.

[9] B. Li, W. Ren, D. Fu, D. Tao, D. Feng, W. Zeng, and Z. Wang, "Benchmarking single-image dehazing and beyond," *IEEE Transactions on Image Processing*, vol. 28, no. 1, pp. 492–505, 2019.

[10] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A simple framework for contrastive learning of visual representations," in *Proc. International Conference on Machine Learning (ICML)*, 2020, pp. 1597–1607.

[11] A. Dosovitskiy et al., "An image is worth 16x16 words: Transformers for image recognition at scale," in *Proc. International Conference on Learning Representations (ICLR)*, 2021.

[12] M. Caron et al., "Emerging properties in self-supervised vision trans- formers," in *Proc. IEEE International Conference on Computer Vision (ICCV)*, 2021, pp. 9650–9660.

[13] C. O. Ancuti, C. Ancuti, R. Timofte, and C. De Vleeschouwer, "O- HAZE: A dehazing benchmark with real hazy and haze-free outdoor images," in *Proc. IEEE CVPR Workshops*, 2018, pp. 754–762.

[14] C. Ancuti, C. O. Ancuti, and R. Timofte, "I-HAZE: A dehazing benchmark with real hazy and haze-free indoor images," in *Proc. ACIVS*, 2018, pp. 620–631.

[15] C. Ancuti, C. O. Ancuti, M. Sbert, and R. Timofte, "NH-HAZE: An image dehazing benchmark with non-homogeneous hazy and haze-free images," in *Proc. IEEE CVPR Workshops*, 2020, pp. 444–445. in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), 2025.

[16] Z. Zhang et al., "From visibility to air quality: A deep learning approach," *Atmospheric Environment*, 2018.

[17] J. Chen, X. Yan, Q. Xu, and K. Li, "Tokenize Image Patches: Global Context Fusion for Effective Haze Removal in Large Images,"

[18] Umesh Kumar Lilhore, P. Kavitha, Swetha G.,Sunder R. Sarita Simaiya, Ehab Seif Ghith, Shimaa A. Hussien "Early Detection of Potato Leaf Diseases with PotatoNet-X: A Deep Learning Model Combining ResNet, DenseNet, and Attention Mechanisms", Potato Research, Volume 69, article number 45 (2026)

[19] H. Zhang, et al., "Contrastive Learning-Driven Image Dehazing with Multi-Scale Feature Fusion and Hybrid Attention," Journal of Imaging, vol. 11, no. 9, 2025.

[20] X. Li et al., "Estimating Air Quality from Images Using Deep Learning," IEEE Transactions on Image Processing, 2020.