Department of Computer and Software Engineering

NUST CEME

PROJECT REPORT

Advanced Coding in Python

NOOR-UL-HUDA
539083

Objective

Single-image dehazing is a fundamental low-level computer vision task aimed at restoring clear images from hazy or foggy observations

In this project, a deep learning–based single-image dehazing model implemented in Python using the PyTorch framework is utilized. The first phase of the work focuses on successfully running the provided baseline code and validating that the model trains, validates, and tests correctly on a small sample dataset.

In the second phase, the same model is trained and evaluated on multiple real-world benchmark dehazing datasets. Quantitative evaluation is performed using standard image quality metrics, and qualitative evaluation is conducted through visual comparison of hazy inputs, dehazed outputs, and ground truth images. Finally, the baseline architecture is enhanced using modern deep learning techniques to improve perceptual and numerical performance.

Dataset Description

To comprehensively evaluate the robustness and generalization ability of the dehazing model, five publicly available benchmark datasets are used. These datasets cover both synthetic and real-world haze conditions and vary in scene types, haze density, and imaging environments.

- The NYU2 dataset is an indoor scene dataset originally designed for depth estimation tasks. It consists of RGB-D images captured using a Microsoft Kinect sensor across various indoor environments such as homes, offices, and classrooms. For the dehazing task, synthetic hazy images are generated using the clean RGB images.
  Due to its synthetic nature, NYU2 serves as a strong baseline dataset for evaluating the learning capability of dehazing networks under well-defined conditions.
  The NYU Depth V2 dataset consists of approximately 1,449 RGB-D images. For image dehazing tasks, synthetic hazy images are generated from the clean RGB images using corresponding depth maps and the atmospheric scattering model.
  The dataset contains paired data, where each hazy image has a corresponding clean ground truth image. The data type used is synthetic haze with paired supervision.

- The RESIDE dataset is a large scale benchmark dataset for single-image dehazing. The synthetic indoor and outdoor training subsets together contain over 13,000 hazy images generated from clean images using varying atmospheric parameters. Each hazy image is paired with its corresponding ground truth clear image. It also includes real-world hazy images without ground truth; however, in this project, only the paired synthetic subset are used for training and quantitative evaluation. The data type is synthetic haze**,** paired RGB images, making it suitable for supervised deep learning.

- The O-HAZE dataset contains 45 real-world outdoor scenes, where each scene includes one hazy image and one corresponding haze-free ground truth image, resulting in a total of 45 paired image samples. Haze is physically generated using professional haze machines, ensuring realistic atmospheric effects. All images are high-resolution RGB images captured with fixed camera setting. The dataset represents real homogeneous haze, meaning the haze density is approximately uniform across the image. The data type is real haze with paired ground truth, which is particularly valuable for testing real-world generalization.

- The NH-HAZE dataset consists of 55 real-world image pairs, where each pair contains a hazy image and its corresponding haze-free ground truth. Similar to O-HAZE, haze is generated physically; however, NH-HAZE focuses on non-homogeneous haze, where haze density varies spatially across the scene.The images are outdoor RGB images captured with professional equipment under controlled lighting. The data type is real non-homogeneous haze with paired supervision, making this dataset significantly more challenging than synthetic datasets and homogeneous haze benchmarks.

- The I-HAZE dataset includes 35 real-world indoor scenes, each consisting of a hazy image and a corresponding clear ground truth image, for a total of 35 paired samples. Haze is generated using professional haze machines in indoor environments such as rooms and corridors. All images are high-quality RGB images with controlled camera settings. The dataset represents real indoor haze and provides paired supervision. The data type is real haze**,** paired indoor RGB images, allowing evaluation of dehazing performance in indoor real-world scenarios.

Model Improvements:

- To improve the dehazing performance of the baseline model, the original residual blocks were replaced with enhanced residual blocks incorporating dilated convolutions, which expand the receptive field without increasing parameter count, enabling better modeling of large-scale haze patterns common in outdoor scenes such as those in the RESIDE dataset. Second, channel attention (CA) and pixel attention (PA) mechanisms were integrated directly inside each residual block, allowing the network to selectively emphasize informative feature channels and spatial regions affected by haze at every depth level, rather than applying attention only at the fusion stage. Additionally, the network depth was increased by using more residual blocks, improving its capacity to learn complex haze-related transformations. For feature aggregation, the original fixed group-based fusion strategy was replaced with a dynamic weighted fusion attention module, which adaptively computes importance weights for multi-level features extracted at different depths of the network. This allows the model to balance low-level detail preservation and high-level semantic information based on scene content. Finally, residual learning was preserved at both block and network levels through skip connections, ensuring stable training and efficient gradient flow.

- Compared to the initial model, the modified architecture replaces standard residual blocks with attentional residual blocks that integrate channel attention (CA) and pixel attention (PA) within each block, enabling finer feature recalibration throughout the network instead of only at the fusion stage. The original pointwise and standard convolutions were replaced with depthwise separable convolutions, significantly reducing computational cost while preserving representational capacity. The fixed group-based weighted fusion mechanism was redesigned into a multi-stage feature fusion strategy, where features from early, middle, and deep layers are concatenated and adaptively fused using channel attention, improving multi-scale haze representation. Additionally, the feature extraction pipeline was simplified by using a unified backbone structure and a cleaner residual learning formulation, making the model more stable and better suited for complex indoor haze scenarios in the NYU-V2 dataset

  .

- In the modified architecture, the original residual blocks were replaced with Feature Attention Blocks (FABlocks) that jointly integrate channel attention and pixel attention within each block, enabling more precise feature enhancement at both channel and spatial levels throughout the network. The fixed three-stage group fusion strategy used in the initial model was redesigned into a multi-level feature fusion scheme, where intermediate features are periodically extracted and fused using a 1×1 convolution, improving the aggregation of shallow and deep haze-related features. Additionally, LeakyReLU activations were introduced to improve gradient flow, and a global residual learning formulation was adopted to directly learn haze components and refine the input image. These changes make the model more adaptive and better suited for the complex, non-uniform haze characteristics of the I-HAZE dataset

- For training on the O-HAZE dataset, the original attention-driven residual network was redesigned into an encoder–decoder architecture to better handle dense, spatially varying outdoor haze. The multi-stage attention fusion mechanism (channel and pixel attention with weighted group fusion) was removed and replaced with hierarchical downsampling and upsampling paths, enabling the model to capture both global haze structure and fine image details. A residual bottleneck with multiple residual blocks was introduced at the lowest resolution to enhance deep feature learning while maintaining training stability. Additionally, skip connections between encoder and decoder layers were employed to preserve low-level spatial information and improve reconstruction quality, making the architecture more suitable for large-scale, real-world haze patterns present in O-HAZE scenes.

- To better address the non-homogeneous and spatially varying haze present in the NH-HAZE dataset, the original multi-group attention fusion architecture was simplified and restructured. The explicit group-based feature aggregation (GPS) and pixel attention fusion were replaced with enhanced residual blocks containing internal spatial attention, allowing the network to focus on locally dense haze regions. A CBAM-style spatial attention mechanism was introduced within each residual block to generate spatial masks that adaptively emphasize haze-affected areas. Additionally, the multi-stage feature concatenation was removed in favor of global channel attention applied to the final fused features, reducing complexity while improving robustness to uneven haze distribution. The overall design maintains residual learning for stable training while improving spatial sensitivity for real-world non-uniform haze.

Performance Metrics

- **Training AOD-Net with the provided code**

```
Batch 25 - Validate PSNR: 28.8986
Batch 25 - Validate SSIM: 0.9552
Batch 26 - Validate PSNR: 29.7405
Batch 26 - Validate SSIM: 0.9488
Validate PSNR: 27.1656
Validate SSIM: 0.9512
Time consume: 0.69 h
```

- **<u>Performance metrics on Reside Dataset:</u>**

```
Epoch 1 Loss: 0.0898
Epoch 2 Loss: 0.0738
Epoch 3 Loss: 0.0701
Epoch 4 Loss: 0.0675
Epoch 5 Loss: 0.0650

--- Validation Results ---
MSE: 0.0084 | PSNR: 23.22 | SSIM: 0.8334 | LPIPS: 0.1141
Epoch 6 Loss: 0.0631
Epoch 7 Loss: 0.0611
Epoch 8 Loss: 0.0598
Epoch 9 Loss: 0.0584
Epoch 10 Loss: 0.0576

--- Validation Results ---
MSE: 0.0073 | PSNR: 23.76 | SSIM: 0.8450 | LPIPS: 0.1026
Epoch 11 Loss: 0.0566
```

```
--- Final Test Results ---
MSE: 0.0073 | PSNR: 23.24 | SSIM: 0.8837 | LPIPS: 0.0591
```

<u>Performance metrics on Reside with modified architecture:</u>

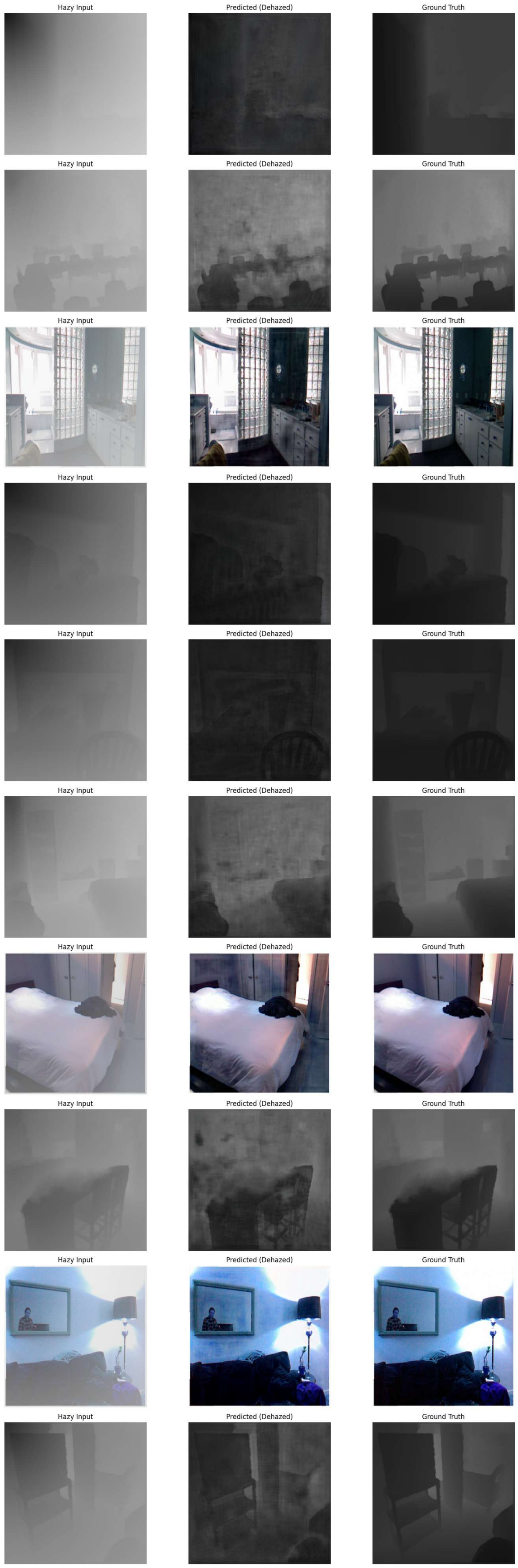
```
Epoch 1/10 | Loss: 0.0996 | LR: 0.000098
Epoch 2/10 | Loss: 0.0908 | LR: 0.000090
Epoch 3/10 | Loss: 0.0884 | LR: 0.000079
Epoch 4/10 | Loss: 0.0834 | LR: 0.000065
Epoch 5/10 | Loss: 0.0767 | LR: 0.000050

--- Validation Results ---
MSE: 0.0152 | PSNR: 21.84 | SSIM: 0.8425 | LPIPS: 0.1145
Epoch 6/10 | Loss: 0.0731 | LR: 0.000035
Epoch 7/10 | Loss: 0.0689 | LR: 0.000021
Epoch 8/10 | Loss: 0.0664 | LR: 0.000010
Epoch 9/10 | Loss: 0.0636 | LR: 0.000002
Epoch 10/10 | Loss: 0.0622 | LR: 0.000000

--- Validation Results ---
MSE: 0.0092 | PSNR: 23.04 | SSIM: 0.8435 | LPIPS: 0.1104
```

```
--- Final Test Results ---
MSE: 0.0063 | PSNR: 23.59 | SSIM: 0.8848 | LPIPS: 0.0669
```

- **<u>Performance metrics on Nyu2v dataset</u>**

```
Starting Training...
[KeepAlive] Training still running... 1 min elapsed
[KeepAlive] Training still running... 2 min elapsed
[KeepAlive] Training still running... 3 min elapsed
[KeepAlive] Training still running... 4 min elapsed
[KeepAlive] Training still running... 5 min elapsed
[KeepAlive] Training still running... 6 min elapsed
[KeepAlive] Training still running... 7 min elapsed
[KeepAlive] Training still running... 8 min elapsed
[KeepAlive] Training still running... 9 min elapsed
[KeepAlive] Training still running... 10 min elapsed
[KeepAlive] Training still running... 11 min elapsed
[KeepAlive] Training still running... 12 min elapsed
[KeepAlive] Training still running... 13 min elapsed
[KeepAlive] Training still running... 14 min elapsed
[KeepAlive] Training still running... 15 min elapsed
[KeepAlive] Training still running... 16 min elapsed
[KeepAlive] Training still running... 17 min elapsed
[KeepAlive] Training still running... 18 min elapsed
[KeepAlive] Training still running... 19 min elapsed
[KeepAlive] Training still running... 20 min elapsed
...
   Val PSNR: 30.38 | Best PSNR: 30.38
--- New Best Model saved with PSNR: 30.3804 ---
```

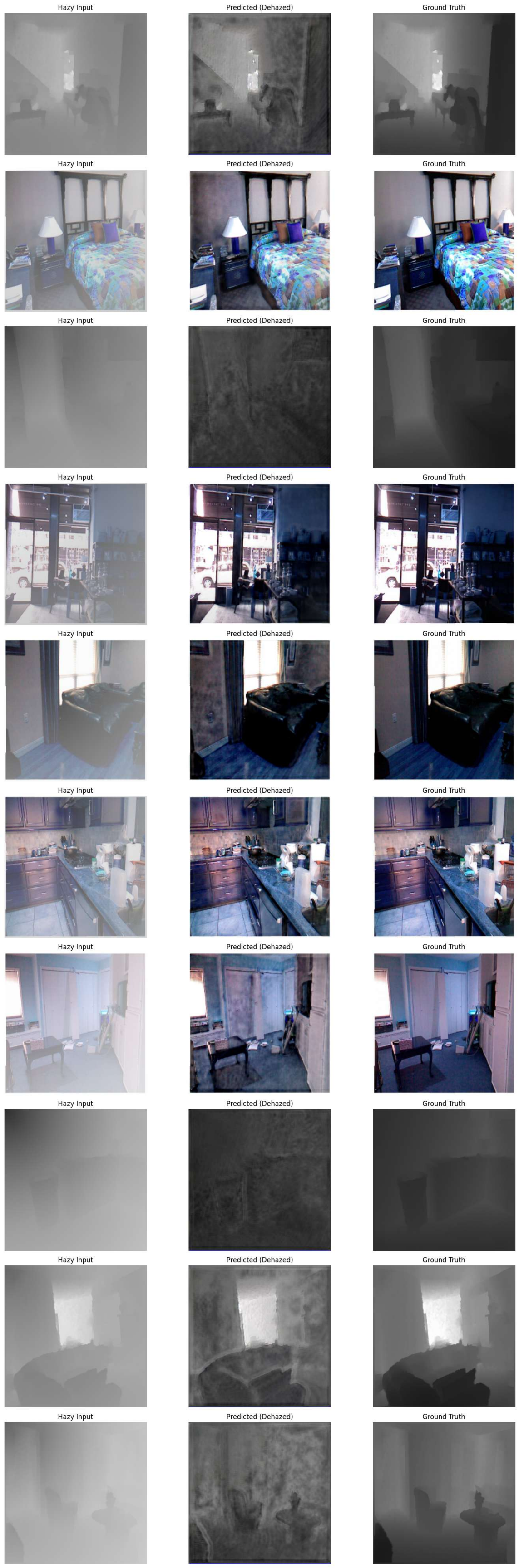
```
Evaluating on Test Set...
FINAL TEST RESULTS:
MSE:   0.00108
PSNR:  30.4600
SSIM:  0.9076
LPIPS: 0.1788
```


Hazy Input
Predicted (Dehazed)
Ground Truth

- **Performance metrics on NYU2v dataset with modified architecture**

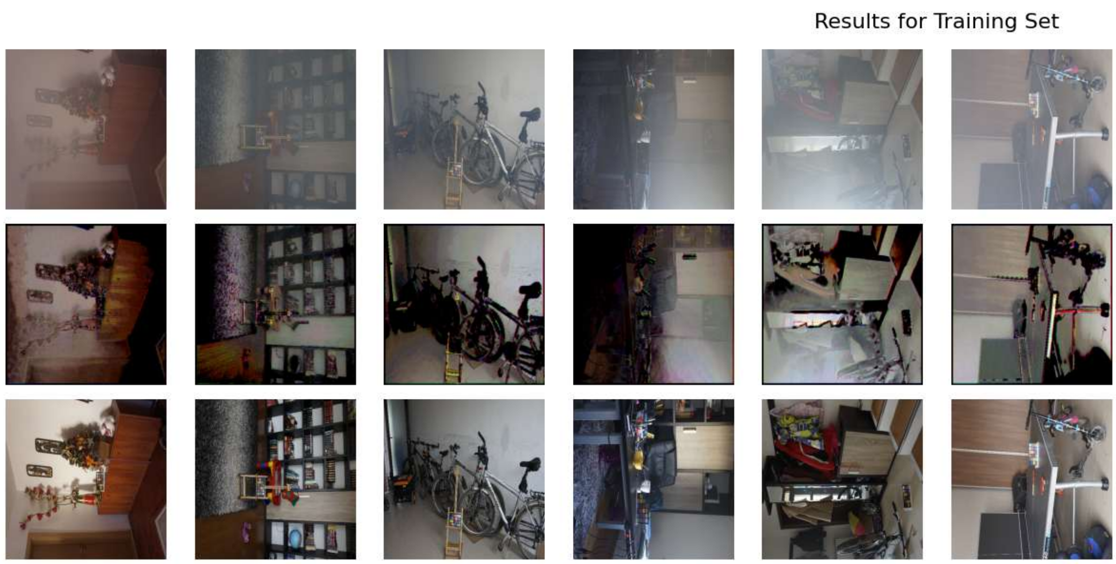
```
[KeepAlive] Training still running... 1 min elapsed
[KeepAlive] Training still running... 2 min elapsed
[KeepAlive] Training still running... 3 min elapsed
[KeepAlive] Training still running... 4 min elapsed
[KeepAlive] Training still running... 5 min elapsed
[KeepAlive] Training still running... 6 min elapsed
[KeepAlive] Training still running... 7 min elapsed
[KeepAlive] Training still running... 8 min elapsed
[KeepAlive] Training still running... 9 min elapsed
[KeepAlive] Training still running... 10 min elapsed
[KeepAlive] Training still running... 11 min elapsed
[KeepAlive] Training still running... 12 min elapsed
[KeepAlive] Training still running... 13 min elapsed
[KeepAlive] Training still running... 14 min elapsed
[KeepAlive] Training still running... 15 min elapsed
[KeepAlive] Training still running... 16 min elapsed
[KeepAlive] Training still running... 17 min elapsed
[KeepAlive] Training still running... 18 min elapsed
[KeepAlive] Training still running... 19 min elapsed
[KeepAlive] Training still running... 20 min elapsed
..
VAL   | PSNR: 24.09 | SSIM: 0.814 | MSE: 0.00440 | LPIPS: 0.245
--> [New Best Model] PSNR improved to 24.09
```

```
Evaluating on Test Set...
FINAL TEST RESULTS:
MSE:   0.00436
PSNR:  24.1471
SSIM:  0.8143
LPIPS: 0.2471
```

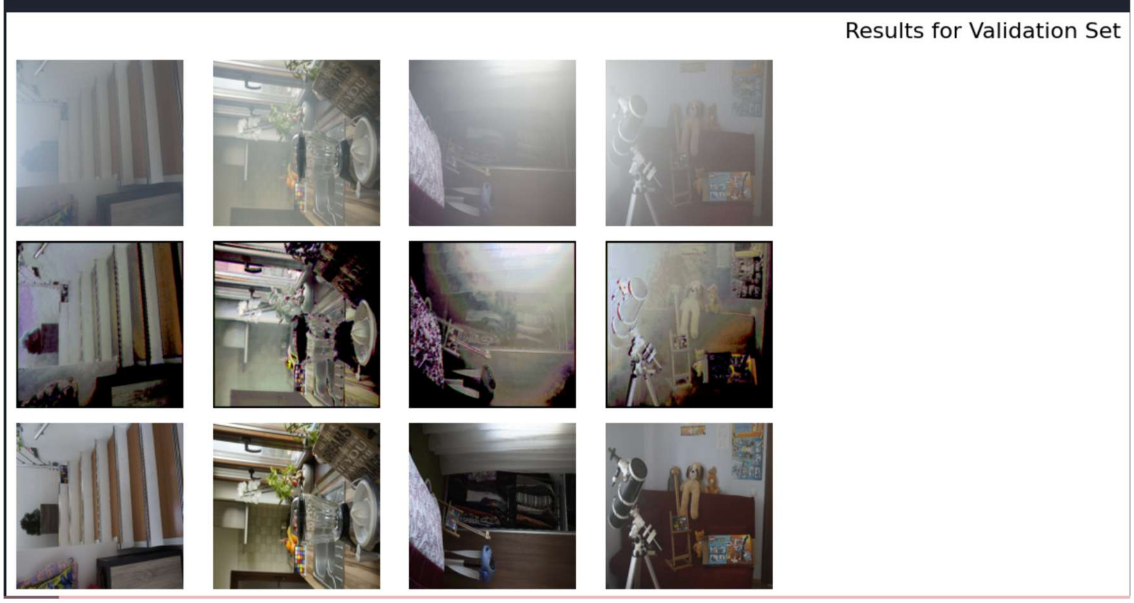

Hazy Input
Predicted (Dehazed)
Ground Truth
Hazy Input
Predicted (Dehazed)
Ground Truth
Hazy Input
Predicted (Dehazed)
Ground Truth
Hazy Input
Predicted (Dehazed)
Ground Truth
Hazy Input
Predicted (Dehazed)
Ground Truth
Hazy Input
Predicted (Dehazed)
Ground Truth
Hazy Input
Predicted (Dehazed)
Ground Truth
Hazy Input
Predicted (Dehazed)
Ground Truth
Hazy Input
Predicted (Dehazed)
Ground Truth
Hazy Input
Predicted (Dehazed)
Ground Truth

- **Performance metrics on I-Haze dataset**

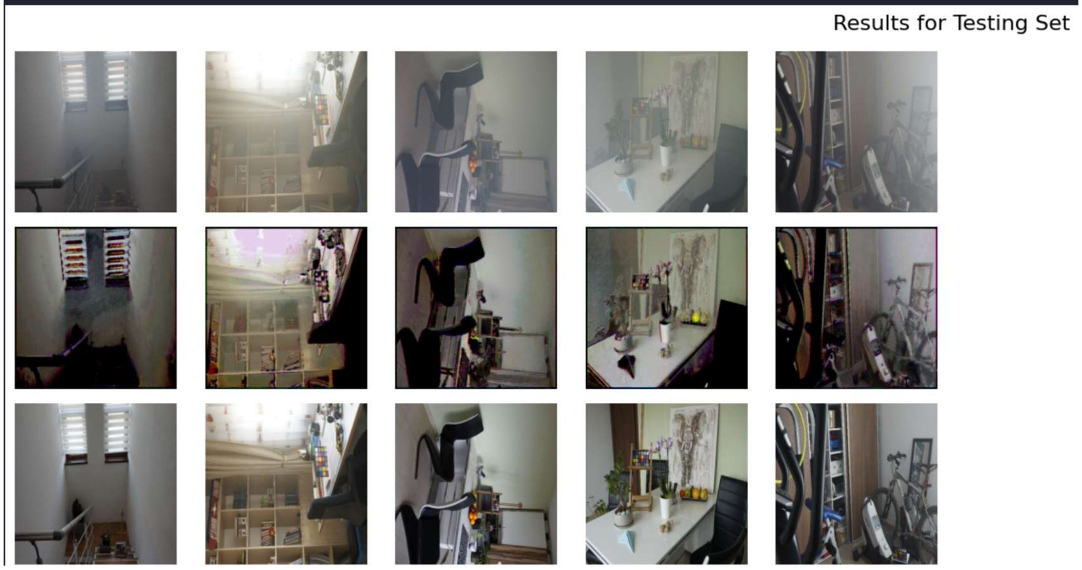
```
Starting Training...
Epoch 1 Loss: 0.2491
Epoch 2 Loss: 0.2098
Epoch 3 Loss: 0.1482
Epoch 4 Loss: 0.1669
Epoch 5 Loss: 0.1647
Epoch 6 Loss: 0.1502
Epoch 7 Loss: 0.1457
Epoch 8 Loss: 0.1562
Epoch 9 Loss: 0.1306
Epoch 10 Loss: 0.1557

--- Training Metrics ---
MSE: 0.051338 | PSNR: 13.0840 | SSIM: 0.3415 | LPIPS: 0.5904
```




```
--- Validation Metrics ---
MSE: 0.059306 | PSNR: 12.5874 | SSIM: 0.3202 | LPIPS: 0.6165
```

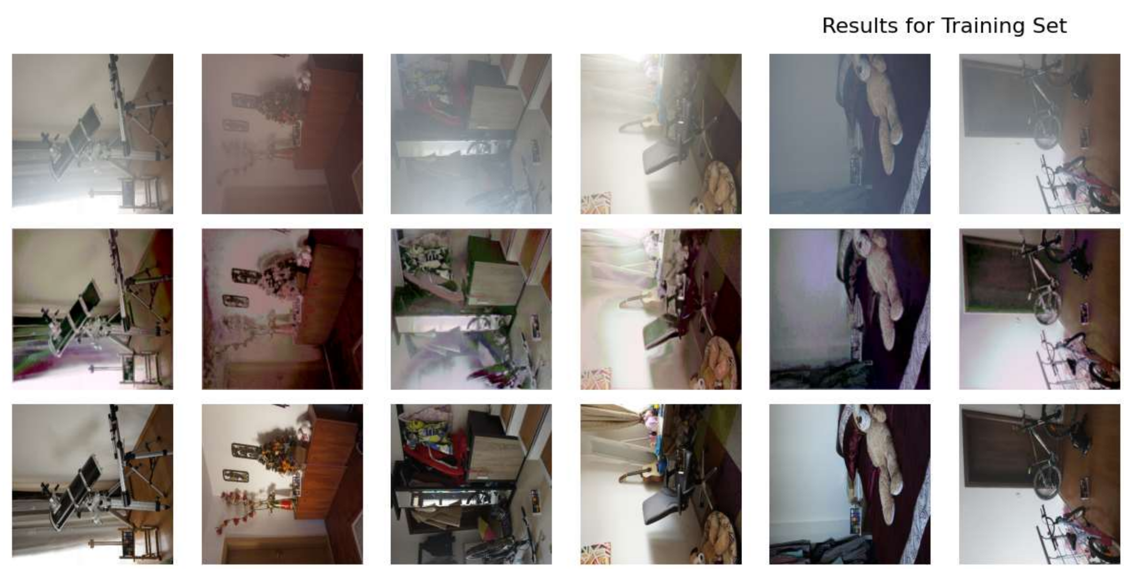


```
--- Testing Metrics ---
MSE: 0.042620 | PSNR: 13.7700 | SSIM: 0.3537 | LPIPS: 0.5733
```



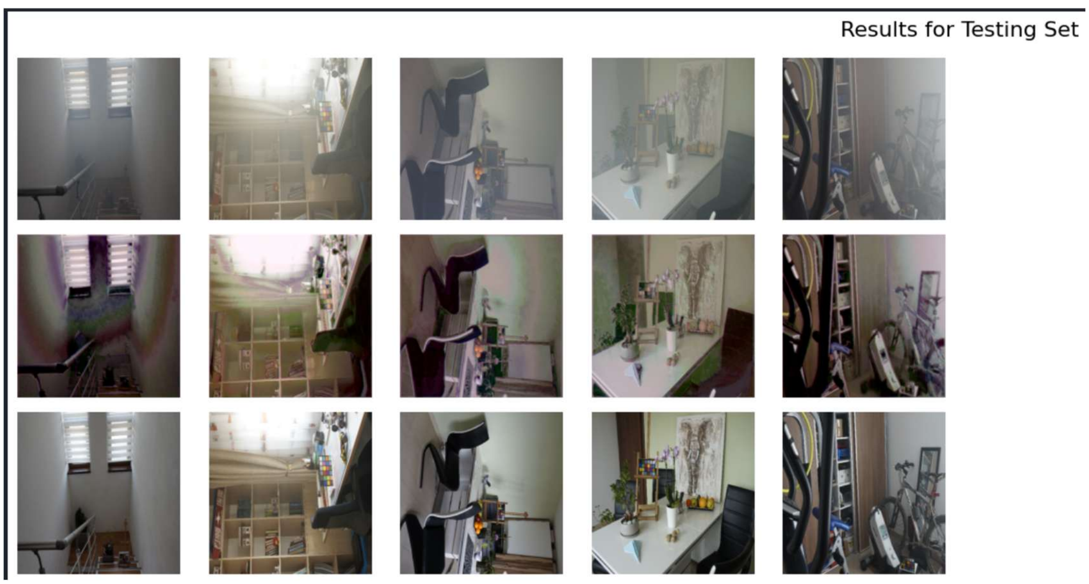
Performance metrics on I-Haze with modified architecture:

```
Starting Training...
Epoch 1 Loss: 0.1913
Epoch 2 Loss: 0.2059
Epoch 3 Loss: 0.2313
Epoch 4 Loss: 0.1437
Epoch 5 Loss: 0.1193
Epoch 6 Loss: 0.1598
Epoch 7 Loss: 0.0804
Epoch 8 Loss: 0.1981
Epoch 9 Loss: 0.1119
Epoch 10 Loss: 0.1647

--- Training Metrics ---
MSE: 0.020260 | PSNR: 17.1729 | SSIM: 0.5099 | LPIPS: 0.5085
```




```
--- Validation Metrics ---
MSE: 0.032309 | PSNR: 15.4153 | SSIM: 0.5113 | LPIPS: 0.5114
```


```
--- Testing Metrics ---
MSE: 0.017046 | PSNR: 18.0156 | SSIM: 0.5328 | LPIPS: 0.4870
```



- **<u>Performance metrics on O-Haze dataset</u>**

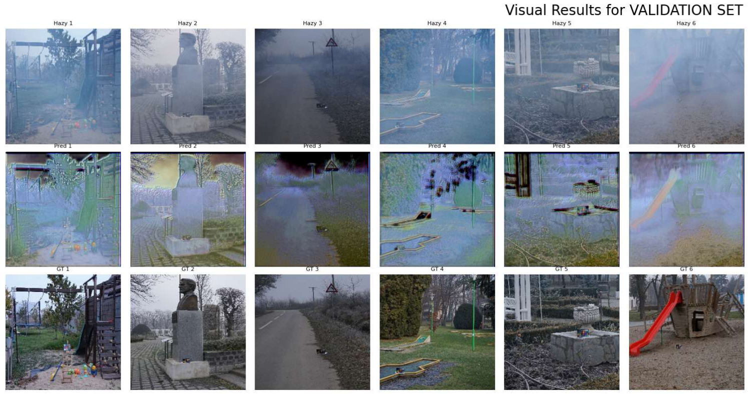
```
Epoch [1/10] - Average L1 Loss: 0.129008
Epoch [2/10] - Average L1 Loss: 0.186045
Epoch [3/10] - Average L1 Loss: 0.202379
Epoch [4/10] - Average L1 Loss: 0.208463
Epoch [5/10] - Average L1 Loss: 0.216212
Epoch [6/10] - Average L1 Loss: 0.211334
Epoch [7/10] - Average L1 Loss: 0.215623
Epoch [8/10] - Average L1 Loss: 0.213159
Epoch [9/10] - Average L1 Loss: 0.213598
Epoch [10/10] - Average L1 Loss: 0.209870
```

```
==============================
 PERFORMANCE METRICS: TRAINING SET
==============================
MSE:   0.068298
PSNR:  11.7190
SSIM:  0.4966
LPIPS: 0.4773
```

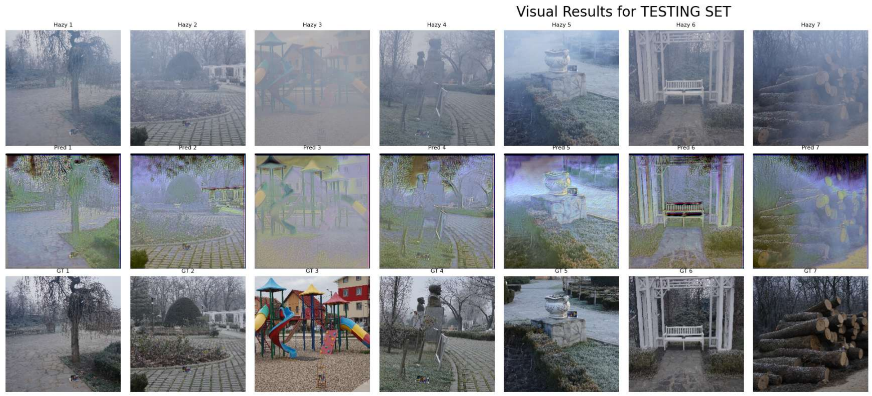
Visual Results for TRAINING SET

```
==============================
 PERFORMANCE METRICS: VALIDATION SET
==============================
MSE:   0.076383
PSNR:  11.3047
SSIM:  0.4956
LPIPS: 0.4975
```

Visual Results for VALIDATION SET



```
==============================
 PERFORMANCE METRICS: TESTING SET
==============================
MSE:   0.061707
PSNR:  12.2080
SSIM:  0.5358
LPIPS: 0.4483
```



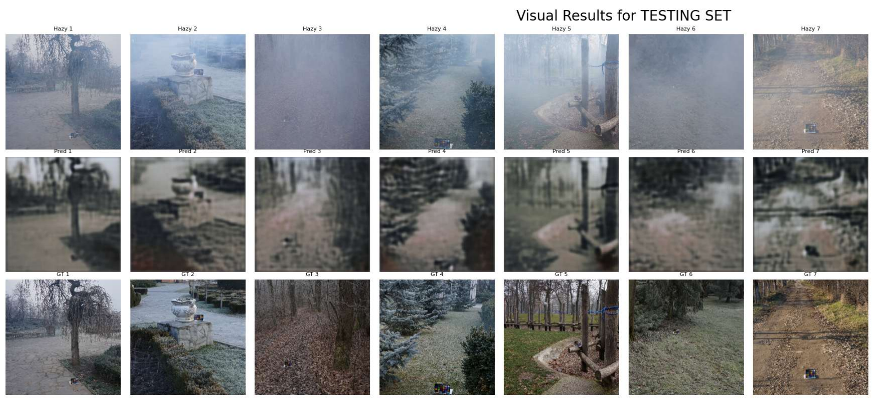
Performance metrics on O-Haze dataset with modified architecture:

```
Epoch [1/10] - Average L1 Loss: 0.167072
Epoch [2/10] - Average L1 Loss: 0.129515
Epoch [3/10] - Average L1 Loss: 0.123347
Epoch [4/10] - Average L1 Loss: 0.116086
Epoch [5/10] - Average L1 Loss: 0.117295
Epoch [6/10] - Average L1 Loss: 0.112176
Epoch [7/10] - Average L1 Loss: 0.114375
Epoch [8/10] - Average L1 Loss: 0.108420
Epoch [9/10] - Average L1 Loss: 0.108345
Epoch [10/10] - Average L1 Loss: 0.109750
```

```
==============================
 PERFORMANCE METRICS: TRAINING SET
==============================
MSE:   0.017176
PSNR:  17.7105
SSIM:  0.3048
LPIPS: 0.7309
```


```
==============================
 PERFORMANCE METRICS: VALIDATION SET
==============================
MSE:   0.023069
PSNR:  16.7248
SSIM:  0.2772
LPIPS: 0.7501
```


```
==============================
 PERFORMANCE METRICS: TESTING SET
==============================
MSE:   0.019635
PSNR:  17.1261
SSIM:  0.2437
LPIPS: 0.7734
```



- **<u>Performance metrics on N-Haze dataset</u>**

- **<u>Performance metrics on N-Haze dataset with modified architecture:</u>**


```
--- TRAINING Metrics ---
MSE:   0.045263
PSNR:  13.76 dB
SSIM:  0.4237
LPIPS: 0.5131

--- VALIDATION Metrics ---
MSE:   0.049201
PSNR:  13.46 dB
SSIM:  0.3594
LPIPS: 0.6133

--- TESTING Metrics ---
MSE:   0.044325
PSNR:  13.78 dB
SSIM:  0.3958
LPIPS: 0.5100
```

Performance Evaluation Metrics

To objectively evaluate the dehazing performance, four widely accepted quantitative metrics are used:

Mean Squared Error (MSE) measures the average squared difference between the dehazed output and the ground truth image, reflecting pixel-level reconstruction accuracy.

Peak Signal-to-Noise Ratio (PSNR) quantifies reconstruction quality in decibels and is commonly used to assess image restoration performance. Higher PSNR values indicate better reconstruction.

Structural Similarity Index (SSIM) evaluates perceptual image quality by measuring structural similarity between the predicted and ground truth images, taking luminance, contrast, and structure into account.

Learned Perceptual Image Patch Similarity (LPIPS) measures perceptual similarity using deep neural network features and correlates well with human visual perception. Lower LPIPS values indicate better perceptual quality.

All metrics are computed separately for training, validation, and testing sets for each dataset

Final Comparison and Conclusion

The final stage of the project presents a clear comparison between the original baseline model and the improved architecture across all five datasets. Quantitative results are summarized in tables, and qualitative improvements are highlighted through visual examples. This comparison demonstrates the effectiveness of the proposed enhancements and provides insights into the strengths and limitations of deep learning–based image dehazing models.