

Fusion of Visual and Near Infrared Images with Wavelets

Gustavo Fonseca
Email: gvf3@illinois.edu

Abstract—There has been a lot of work done on using near-infrared (NIR) images to enhance visual (VIS) images. For instance, important information in VIS images may be obscured by shadows or hidden due to strong saturation. NIR images can enhance VIS images because they capture additional wavelengths. An interesting method for enhancing VIS images with NIR was proposed by Zhang et. al in [1], which used the wavelet transform within their enhancement algorithm. However, they only used a one-level Haar wavelet decomposition within their algorithm. This study considers the benefit of applying multiple levels of the Haar wavelet decomposition to create the final enhanced image. This study further evaluates Zhang’s one-level decomposition and our multilevel decomposition with qualitative and quantitative evaluations.

I. INTRODUCTION

There has been significant interest in the fusion of VIS and NIR images for many applications, such as remote sensing, medical imaging, and surveillance. The NIR image contains details that are often absent in the VIS image, particularly from shadows, strong saturation, or low contrast.

One issue with this type of fusion is keeping the visual quality of the image consistent with the original VIS image and the scene being represented. This can be difficult as an NIR image is only defined by intensity and is grayscale.

While researching this topic, I found a promising method proposed by Zhang et. al [1], which used the wavelet transform to enhance VIS images with NIR images.

Zhang’s method of fusion only uses one level of the Haar wavelet decomposition, which is computationally efficient. However, using multilevel wavelet decomposition was unexplored by Zhang.

The goal of this study is to explore the impact of using multilevel wavelet decomposition for enhancing VIS images with NIR images. We modified Zhang’s method to incorporate the multiple levels. We see a very large potential in increasing the number of levels for enhancement of image quality especially for highly saturated regions. This method of image enhancement using the wavelet transform could also be impactful for other image processing applications for non-NIR image fusion.

We studied the effects of our method compared to Zhang’s single-level wavelet decomposition method qualitatively and quantitatively with various image quality metrics.

II. METHODOLOGY

A. Single Level Wavelet Algorithm Overview

To put my work in context, it is important to first provide an overview of Zhang’s algorithm [1]. We can see the overall

workflow for Zhang’s algorithm in Fig. 1. Zhang’s algorithm are performed only in the brightness layer of the VIS image and its corresponding NIR image. The hue and saturation layers of the VIS image are left unchanged.

In Zhang’s workflow, we can see three main aspects: The Weighted Region Mask (WRM), Transfer Contrast (TC), and Transfer Texture (TT).

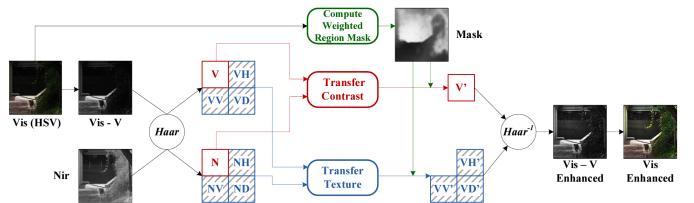


Fig. 1: The 3 main steps of Zhang’s method (Figure reproduced from [1]).

1) *Weighted Region Mask*: The goal of the Weighted Region Mask (WRM) is to indicate which sections of the image need the most enhancement. The WRM is computed solely from the VIS image.

In [1], Zhang reasoned that the parts of the image that lost the most detail would be the areas with the highest and lowest brightness and low saturation. To model which parts of an image should be enhanced the most, Zhang used in (1) and (2), where s_{ij} and v_{ij} are the saturation and brightness of each pixel respectively, and p_s and p_v are the probabilities that this specific value of s_{ij} or v_{ij} appear in the image.

Zhang got p_s and p_v by using a normalized histogram of the image’s S (saturation) and V (brightness) components. With this, Zhang obtained the final WRM (W) using (3).

$$W_{ij}^{(s)} = 1 - e^{-p_s|s_{ij}-1|}, \quad p_s \in [0, 1], s_{ij} \in [0, 1], \quad (1)$$

$$W_{ij}^{(v)} = 1 - e^{-p_v|v_{ij}-0.5|}, \quad p_v \in [0, 1], v_{ij} \in [0, 1], \quad (2)$$

$$W = W^{(s)} \cdot W^{(v)}. \quad (3)$$

2) *Transfer Contrast*: The goal in this step is to transfer the contrast of the NIR Image to the enhanced VIS image. Zhang proposes that, before trying to transfer the contrast, one applies a **one-level Haar wavelet decomposition** to the VIS image’s brightness component (V), and to the NIR image (N) as shown in Fig. 1. Let V_{LL} and N_{LL} be the LL subband of each of the one-level Haar wavelet decomposition respectively.

TC then uses V_{LL} and N_{LL} as follows. The algorithm starts by applying a bilateral filter (bf) separately to V_{LL} and N_{LL} . The output of this filter is defined as the large-scale layer for the image, resulting in images $V_{LL,l}$ and $N_{LL,l}$. These images are blurred versions of V_{LL} and N_{LL} but the edges are maintained more than a Gaussian blur.

From the large scale layers, the method produces the detail layers, $V_{LL,d}$ and $N_{LL,d}$, obtained as shown in (4) and (5).

$$V_{LL,l} = bf(V_{LL}), \quad V_{LL,d} = V_{LL}/V_{LL,l}, \quad (4)$$

$$N_{LL,l} = bf(N_{LL}), \quad N_{LL,d} = N_{LL}/N_{LL,l}. \quad (5)$$

From here, Zhang's method matches the histogram of $V_{LL,l}$ to the histogram of $N_{LL,l}$, resulting in $V'_{LL,l}$. Although Zhang applies a global histogram matching, we used local histogram matching as it provided better results in our testing. The final resulting LL subband of the to-be enhanced image is defined by (6):

$$V'_{LL} = W \cdot (V'_{LL,l} \cdot V_{LL,d}) + (1 - W) \cdot V_{LL}. \quad (6)$$

This new V'_{LL} will serve as the new LL subband when performing the inverse wavelet transform, as illustrated in Fig. 1. It is interesting to mention that the NIR image is used indirectly in (6), through the $V'_{LL,l}$.

3) Transfer Texture: The goal in this step is to transfer the texture of the NIR image to the enhanced VIS image. The TT step in Zhang's approach is straightforward: use the WRM for alpha-blending of the LH, HL, and HH subbands of NIR and VIS images; and, together with the LL subband given by V'_{LL} , obtain the enhanced image through the inverse wavelet decomposition, as indicated in Fig 1. More precisely, the LH, HL, and HH subbands are obtained with:

$$V'_{LH} = W \cdot N_{LH} + (1 - W) \cdot V_{LH} \quad (7)$$

$$V'_{HL} = W \cdot N_{HL} + (1 - W) \cdot V_{HL} \quad (8)$$

$$V'_{HH} = W \cdot N_{HH} + (1 - W) \cdot V_{HH} \quad (9)$$

and the brightness layer of the enhanced image is obtained by the inverse wavelet transform of the coefficients in V'_{LL} , V'_{LH} , V'_{HL} , and V'_{HH} .

Lastly, the hue and saturation layers from the original VIS image were recombined with the resulting brightness layer to produce the final enhanced VIS color image.

B. Our Multi-Level Wavelet Enhancement

Using a one-level wavelet decomposition seemed like an oversight to us. There are a variety of details from the NIR image that could be extracted further by using more than one wavelet decomposition level. In order to extract these details, we augmented the algorithm to enhance every V_{LL} subband until a given level.

To explain our method, let us consider just a two-level wavelet decomposition.

As seen in Fig. 2, we compute the first level Haar wavelet decomposition just like in [1]. However, instead of immediately computing the TC and TT steps with the first level components, we use $V_{LL,1}$ and $N_{LL,1}$ as input to the next level

of the Haar wavelet decomposition, producing the second level subband components: $V_{LL,2}$, $N_{LL,2}$, and their corresponding 2D wavelet coefficients. With the second subband components, we compute the TC and TT steps to produce $V'_{LL,2}$ and its associated 2D wavelet coefficients. We refer to the TC and TT steps at the second level as TC_2 and TT_2 .

At this point, we take the inverse Haar wavelet decomposition using $V'_{LL,2}$ and its associated 2D wavelet coefficients as inputs, producing $V'_{LL,1}$, which is used in TC_1 and TT_1 steps to produce $V'_{LL,1}$ and its associated 2D wavelet coefficients. Lastly, such coefficients are used in the inverse Haar wavelet transform to produce the final enhanced image V^* .

We also note that $V_{LL,1}$ is used to produce WRM_2 which is used in TC_2 and TT_2 steps. This mirrors how V is used to produce WRM_1 which is used in TC_1 and TT_1 steps like in the original method from [1].

To extend this method to more than 2 levels is straightforward: Assume enhancement of K levels of the wavelet decomposition. Each level of the Haar wavelet transform has $V_{LL,k}$ and $N_{LL,k}$ and its 2D wavelet coefficients. From here, we take the Haar wavelet transform of $V_{LL,k}$ and $N_{LL,k}$ to produce $V_{LL,k+1}$ and $N_{LL,k+1}$ and their 2D wavelet coefficients. Subsequent wavelet transforms are performed until $V_{LL,K}$ and $N_{LL,K}$ and their 2D wavelet coefficients are produced. We now go back up the stack, using $V_{LL,K}$ and $N_{LL,K}$ and their 2D wavelet coefficients as input to TC_K and TT_K to produce $V'_{LL,K}$ and its associated 2D wavelet coefficients, which are used as input to the inverse Haar wavelet transform. The result is $V'_{LL,K-1}$, which is the enhanced image at the $K-1$ level. $V'_{LL,K-1}$ replaces $V_{LL,K-1}$ as the input to TC_{K-1} and TT_{K-1} ; and the process repeats until TC_1 and TT_1 produce $V'_{LL,1}$ and its associated 2D wavelet coefficients, which are used as input to the last inverse Haar wavelet transform, producing the final enhanced image V^* .

We note that a K -level enhancement has K Haar wavelet transforms followed by K steps of TC and TT and K inverse Haar wavelet transforms. Also, each level has an associated WRM_k that is obtained from $V_{LL,k-1}$.

A benefit of our approach is that it can be easily implemented in a recursive manner.

III. RESULTS & DISCUSSION

We selected 28 images from [2], a dataset of VIS and NIR images, to run our multilevel decomposition algorithm and evaluate the results. We handpicked images that had more information in the NIR image than in the VIS image, which generally occurs in images with a lot of shadows or saturation. We evaluated our algorithm both qualitatively and quantitatively.

A. Qualitative Results & Discussion

Let us focus on one main example for this section, looking at how the visual quality of the image looks using multiple different levels for the wavelet decomposition.

The original VIS and NIR images are given in Fig 3a and 3b. We can see that the wall is extremely saturated in the VIS

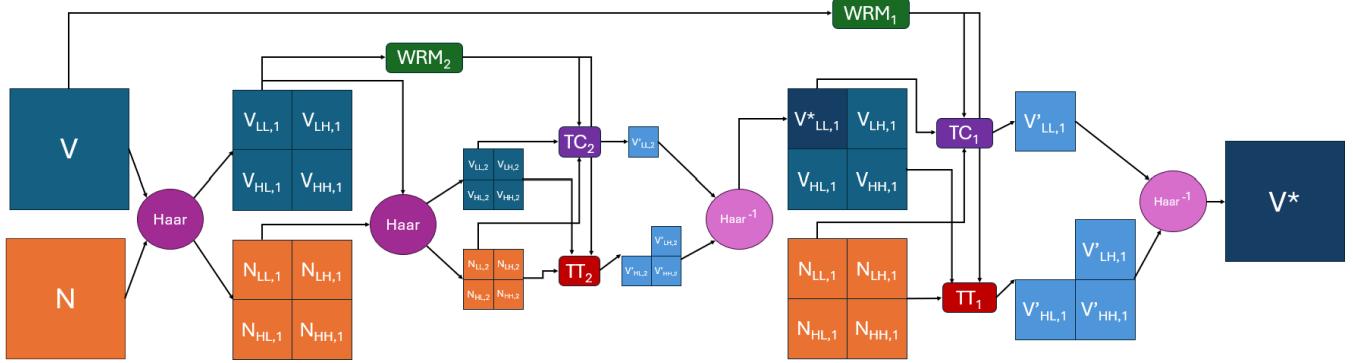


Fig. 2: Illustration of the proposed multilevel decomposition for two levels.

image, with the camera unable to capture any of the details of it. In contrast, we can see that the NIR image was almost unaffected by the harshness of light reflecting from the wall, as it was able to capture significantly more detail than the VIS image in this section.



(a) Visual spectrum image from [2].



(b) Near-Infrared spectrum image from [2].

Fig. 3: Sample VIS and NIR Image form [2].

We see the resulting enhanced image after applying the algorithm from [1] in Fig. 4a. Looking at the saturated wall, we can see that areas of the wall have darkened somewhat; however, the details on the wall from the NIR image are still largely missing. When we use two and three levels for the wavelet decomposition, as seen in Fig. 4b and 4c respectively, we can see that the details progressively improves throughout the region of interest. The highest levels of change can be seen on the right edge of the saturated wall. This shows that using more than one level of wavelet decomposition can improve the

qualitative aspects of an image.

However, increasing the number of enhanced decomposition levels past a point causes a particular kind of degradation. As seen in Fig. 4d, which is the output of six levels of decomposition, the wall color is significantly darkened and we can begin to see artifacts from enhancing each level of the wavelet transform. We can see that, at the edges of the wall, the average intensity of the area changes significantly in each distinct region, reflecting the effects of modifying lower frequency subbands of the wavelet transform. Specifically, regions of enhancements from the WRM at high levels of decomposition are enlarged when doing the inverse Haar wavelet transform.

The trend of the enhanced image worsening significantly after a few levels is not shared throughout all images tested. Many images that did not have much added information in the NIR image did not significantly change even after several levels of the wavelet transform was enhanced. An example of this is shown in Fig. 5, as the contrast in the sky is the largest change in the enhanced image, with the rest of the image relatively unaffected.

B. Quantitative Results & Discussion

To test the quality of our algorithm, we implemented 3 metrics: restore edges [3], colorfulness [3], and no-reference perceptual quality assessment [4]. We chose these metrics as they have been previously used for the evaluation of enhancement using NIR images in [3] and [5]. These metrics are summarized below:

- **Restore Edges (e):** Measures the amount of edges that were restored from an image that either did not have the edge before or was not strongly defined. A score above zero means that more edges were recovered than lost in the enhanced image with reference to the original VIS image.
- **Colorfulness (CF):** Measures the amount of color and saturation within an image. An image with large areas of saturation typically has a smaller CF score. The larger a CF value is, the more colorful or vibrant the image is.
- **no-reference perceptual quality assessment (NQA):** Quality metric that does not require a reference image for

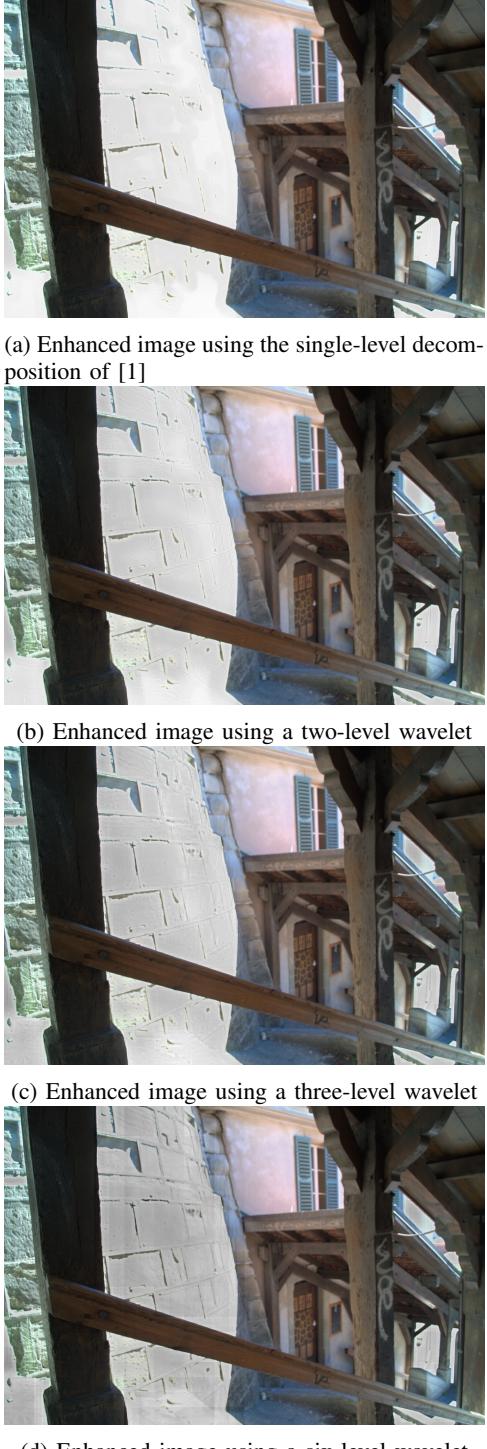


Fig. 4: Qualitative evaluation of multilevel decomposition using sample image of [2]

calculation. Typically used for calculations of distortions within JPEG compressed images. This is applied to the grayscale version of the image. The higher the NQA of the image the higher the quality of the image.

We then applied these metrics to 28 VIS and NIR image

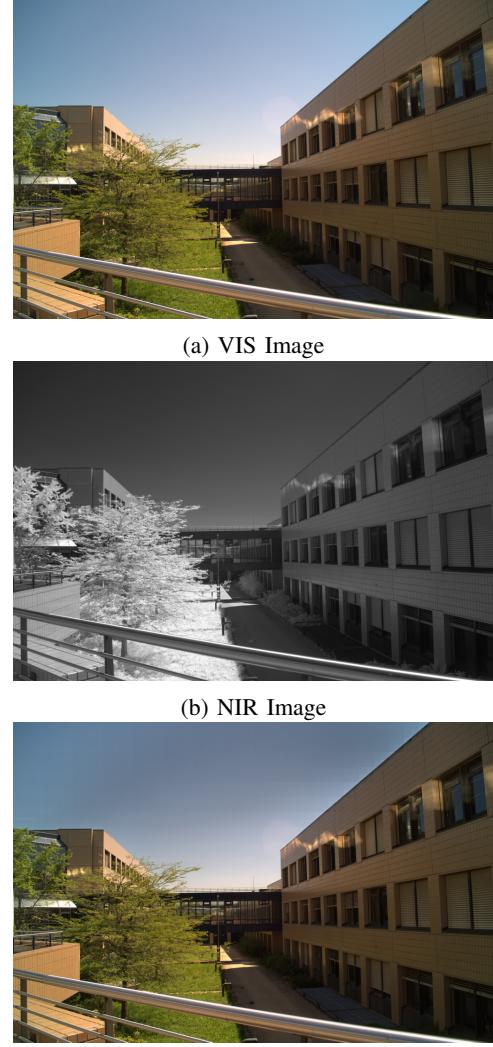


Fig. 5: Image minimally effected by enhancement [2]

pairs from [2] that were all enhanced using levels one through eight of our algorithm. The summary of the results can be found in TABLE I, where ΔCF and ΔNQA are the difference between the CF and NQA of the enhanced image and the original VIS image. We can see that generally the metrics are unfavorable to the algorithm's results, with ΔCF and ΔNQA being negative values. However, it is promising that all three metrics generally increased with more levels of the wavelet transform enhanced. After seven subband levels were enhanced, the highest value of e and the first value of a positive ΔCF were achieved. Thus, we believe that seven levels produced the best quantitative results across all tested images based on these metrics.

The results of restoring edges are encouraging, as we see that the average of the restore edges metric (e) were improved using every level. It is very logical that this is the case, as this algorithm's main job is to increase the amount of information or edges from the NIR image into the VIS image. To visualize

| # of Levels | e | ΔCF | ΔNQA |
|-------------|----------|-------------|--------------|
| 1 Level | 0.02018 | -0.466735 | -0.051064 |
| 2 Levels | 0.025593 | -0.350590 | -0.036992 |
| 3 Levels | 0.026966 | -0.244404 | -0.015142 |
| 4 Levels | 0.027529 | -0.154091 | -0.010905 |
| 5 Levels | 0.028441 | -0.091787 | -0.011551 |
| 6 Levels | 0.029164 | -0.041928 | -0.011598 |
| 7 Levels | 0.030211 | 0.025835 | -0.016411 |
| 8 Levels | 0.030149 | 0.108838 | -0.022916 |

TABLE I: Mean of e , ΔCF , and ΔNQA for each level of the wavelet transform

this better, Fig. 6 shows a histogram for the restore edges metric (e) obtained in each of the 28 images. We can see that most of the values for the algorithm from [1] is around 0. Assuming that the outlier images stay the same throughout each level, the shift is much more pronounced on the outliers on both sides, with the few images that had reduced edges with the algorithm from [1] lost even more edges with more levels in some cases. This could happen in some images if the VIS image actually contains more information than the NIR image, as is the case with transparency of objects in water.

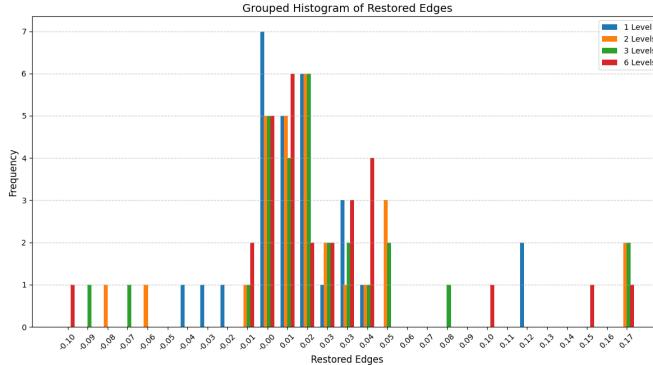


Fig. 6: Histogram of restore edges metric (e).

We believe that ΔCF may be an unfavorable metric for this type of image enhancement as it measures aspects that may be counterintuitive to optimize for in the cases we are interested in. In more details, the authors in [3] used this metric when trying to eliminate haze using NIR images, reasoning that haze has a low saturation and removing it would increase the image's colorfulness. In many of the images that we analyzed, there were many areas of extreme saturation that we are trying to remove, which counterintuitively lowers the colorfulness score. This can be seen in the ΔCF between Fig. 3a and 4c, which is -0.3386.

In the case of NQA, the authors in [4] created this quality assessment metric for measuring distortion in JPEG images. A key point for calculating distortion is the homogeneity of the image. This could be why we see that the ΔNQA goes up with more levels in the wavelet transform enhancement, as the lower level subbands are creating a more homogenetic effect across the image as it is at a lower frequency. Thus, the higher frequency changes at the higher levels adhere to the homogeneity of the image, making NQA improve.

Overall, these metrics are not optimal to evaluate the quality of image enhancement with NIR. As mentioned above, CF is counterintuitive in many cases, as there are many cases where a lower CF is actually better. Although we believe that restore edges is the best metric for this type of analysis, it fails to account for the information within NIR at all, only looking at the lack of information in the VIS image. Similarly, NQA falls short of fully encompassing the scope of resources available, as the concept of assessing image quality without a reference fails to take into account that we could use the regions of the NIR and VIS as reference.

IV. CONCLUSION

Overall, this study shows that using multiple levels of wavelet transform can improve image enhancement algorithms with NIR images. Enhancing around three or four subband levels gave the best qualitative results, while seven levels gave the best quantitative results. While the overall quality of the images could use improvement, the potential for this algorithm is very promising as we continue our research into this field.

As we progress with this research, we would like to improve our algorithm and create better metrics to test it. As seen in Fig.4d, the enhanced regions of the images were darkened dramatically with more subband levels enhanced. This could be due to applying the same weighting scheme to all levels equally, possibly causing a stronger intensity distortion. Additionally, as seen with Fig.5, not all images are affected as much as we would have liked. Implementing different algorithms for different types of images, such as high or low saturation images, may help with this. Finally, after extensively looking for better metrics for assessment in this field, we believe that development of robust metrics to measure the quality of the image receiving the NIR details into account is extremely important.

In terms of applications, this type of image enhancement has many applications. For instance, self driving cars need to recognize objects obstructed by shadows to increase their reliability.

REFERENCES

- [1] X. Zhang, T. Sim, and X. Miao, "Enhancing photographs with near infrared images," in *2008 IEEE conference on computer vision and pattern recognition*. IEEE, 2008, pp. 1–8.
- [2] E. IVRL, "RGB-NIR Scene Dataset," <https://www.epfl.ch/labs/ivrl/research/downloads/rgb-nir-scene-dataset/>, accessed: 2024-10-10.
- [3] M. A. Herrera-Arellano, H. Peregrina-Barreto, and I. Terol-Villalobos, "Color outdoor image enhancement by v-nir fusion and weighted luminance," in *2019 IEEE International Autumn Meeting on Power, Electronics and Computing (ROPEC)*. IEEE, 2019, pp. 1–6.
- [4] Z. Wang, H. R. Sheikh, and A. C. Bovik, "No-reference perceptual quality assessment of jpeg compressed images," in *Proceedings. International conference on image processing*, vol. 1. IEEE, 2002, pp. I–I.
- [5] A. V. Vanmali, S. G. Kelkar, and V. M. Gadre, "A novel approach for image dehazing combining visible-nir images," in *2015 Fifth national conference on computer vision, pattern recognition, image processing and graphics (NCVPRIPG)*. IEEE, 2015, pp. 1–4.