5
RDC-UNet: A UNet Based Model for Low Quality Fundus Image Enhancement

In the previous chapters, we have discussed the importance of fundus IQA and enhancement as a pre-processing step for a trustworthy diagnosis. Also, the significance of the "fair" category of quality in fundus IQA was also discussed.

As mentioned in Chapter 4, the proposed fundus IQA method classifies a fundus image into three categories of quality: good, fair, and poor. Here, good quality indicates fundus images having all its structural features intact, leading to reliable diagnosis. Similarly, the fair category includes fundus images with a few distortions that may cause erroneous results in automated diagnosis. On the other hand, poor quality images are unusable for diagnostic purposes. An example of fundus images from good, fair, and poor category is shown in Fig. 5.1. It can be observed that poor category images can rarely be enhanced for diagnosis purposes. However, fair category images hold the scope of enhancement to make them fit for a reliable diagnosis. In addition, recent advancements in image acquisition technologies lead to capture of fundus images using portable devices [Bourouis *et al.*, 2014]. In the era of telemedicine [Shi *et al.*, 2015], it provides an easy way to capture, store, and share them with the ophthalmologist. However, due to ease of use, such devices are more susceptible to various distortions in comparison to a conventional setup. Therefore such images demands strict quality assessment and enhancement if required before using the same in CAD system-based diagnosis.

As already discussed in Chapter 2, the limitations of the state-of-the-art fundus enhancement methods are as follows:

- Histogram equalization based fundus enhancement methods are not effective over: (i) controlling the level of enhancement and (ii) high intensity uneven illumination distortions.
- The use of Gaussian noise, Impulse noise, and Multiplicative noise in learning based methods are not strictly relevant to the fundus images.
- No relevant data-set is available to benchmark the methods. Also, many challenges exist, like manpower, time, and financial constraints, to create a large data-set of retinal images that are sufficient to train heavy deep learning models.

In this chapter, we have addressed the above mentioned limitations and proposed a new RDC-UNet model for fundus image enhancement. The structure of the rest of the chapter is as follows. Section 5.1 contains the detailed implementation descriptions of the algorithms to create artificial degradations that are closely resembling the naturally appearing degradations. Section 5.2 provide details of the proposed RDC-UNet model, including a brief introduction to the UNet and Densely connected residual blocks. Section 5.3 provides a detailed analysis of the experimental results. It analyzes the performance of the proposed model over both the naturally and artificially degraded retinal images. Also, the effectiveness of the obtained results is shown with the help of blood vessel segmentation application. In Section 5.4 a detailed discussion over the insights of the proposed work is presented. Finally, Section 5.5 concludes the chapter.

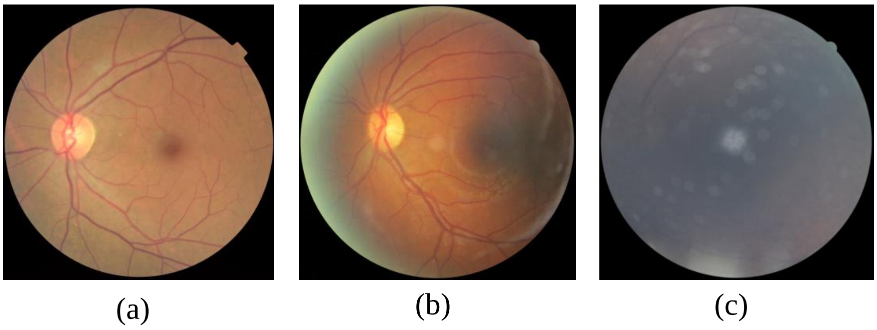


Figure 5.1 : Fundus images from three categories of image quality: (a) Good, (b) Fair, and (c) Poor.

5.1 IMPLEMENTATION OF FREQUENTLY OCCURRED DISTORTIONS

This section presents a detailed description of the proposed methods for the distortions that frequently occur in fundus images. The identification of these distortions was done using the visual assessment of EyeQ [Fu *et al.*, 2019a] data-set. As already mentioned, the EyeQ dataset contains three classes of retinal image quality. A careful study of the literature [Raj *et al.*, 2019] and observation of the EyeQ data-set led to the inference that improper luminance, uneven illumination, and haze are frequently occurring distortions in retinal images. Improper luminance leads to highly bright or dark fundus images, while uneven illumination mostly affects the macular and border areas of such images. Here, the primary reasons behind the occurrence of these distortions are the following: (i) dust and dirt on camera lenses, (ii) improper light conditions, and (iii) haze events [Raj *et al.*, 2019]. To create these distortions, a total of *1000* good quality images were randomly selected from the EyeQ data-set. Fundus images hold a considerably large area of a dark background. Such redundant information adversely affects the training accuracy. Therefore, all images were cropped to the fundus region boundary. The boundary location was determined by finding the nearest pixel coordinates with a value close to zero (i.e. black) from the centre of the image in each of the four axial directions. Additionally, each image was resized to the dimension of 512×512. The algorithms mentioned here were proposed to distort these images resembling the distortions that appear naturally.

- **MUI:** The macular region is one of the areas affected by uneven illumination distortion, as shown in Fig. 5.2 (a). As can be observed, it creates a dark region around the macula. It is worth mentioning that most of the fundus images used for the experiments were macula-centred. Therefore, the MUI distortion was created starting from the centre of the image, assuming that the macula is located near the centre. The spatial location of the MUI could be anywhere around the macular region and in any direction. Therefore, to increase the model's robustness towards the equiprobable spatial directions, a circular area around the macula is selected. A darkness intensity (DI) value is chosen heuristically to induce a darkness effect in the region. To gradually decrease the darkness effect away from the centre, the scaled DI value is subtracted from each pixel I(x,y) within the region. The scaling factor was proposed as the ratio of the square of the circle radius R and the squared Euclidean distance of the pixel from the centre, as provided in Algorithm 1, and it decreased away from the centre. Furthermore, during the initial experimentation, various DI values in the range of 60–150 were tested. However, after careful discussion with ophthalmologists, DI in the range between 80 and 120 with an interval of 10 was found to be satisfactory. Additionally, two radius values of 100 and 120 pixels were chosen to increase the degradation variability. The selected radius values are 20-25% of the resolution of the images. A total of 5000 samples of such distorted fundus images were created.

- **BUI:** Another common distortion in the retinal image is the appearance of green colour shade over

Algorithm 1 MUI

```
$I : Input\ Fundus\ Image$
$R : Radius\ of\ Circle : 100\ and\ 120\ pixels$
$DI : Darkness\ Intensity$
procedure MUI($I$)                                    ▷ MUI: Macula Uneven Illumination
    $r \leftarrow row(I)$
    $c \leftarrow column(I)$
    $center(x,y) \leftarrow (r/2, c/2)$
    for i = -R to R do
        for j = -R to R do
            $dist = i^2 + j^2$
            $if(dist < R^2)$
            $distNorm = dist/(R^2)$
            $scaleFactor = 1 - distNorm$
            $I_N(x+i, y+j) = I(x+i, y+j) - DI \times scaleFactor$
        end for
    end for
    return $I_N$                                       ▷ Output fundus image
```

Algorithm 2 BUI

```
$I : Input\ Fundus\ Image$
$hI : Pixel\ value\ with\ highest\ frequency\ near\ border$
$gI : Heuristically\ chosen\ intensity\ value$
procedure BUI($I$)                                    ▷ BUI: Border Uneven Illumination
    $r \leftarrow row(I)$
    $c \leftarrow column(I)$
    $center(x,y) \leftarrow (r/2, c/2)$
    for i = 1 to r do
        for j = 1 to c do
            $if(I(i,j) > 0)$
            $dist = (x-i)^2 + (y-j)^2$
            $if(x^2 >= dist >= (x/5)^2)$                ▷ Border Region
            $scaleFactor = dist/(x^2 + y^2)$
            $Inew(i,j) = I(i,j) + scaleFactor \times gI$
            $Inew(i,j) = min(Inew(i,j), hI)$
        end for
    end for
    return $Inew$                                      ▷ Output fundus image
```

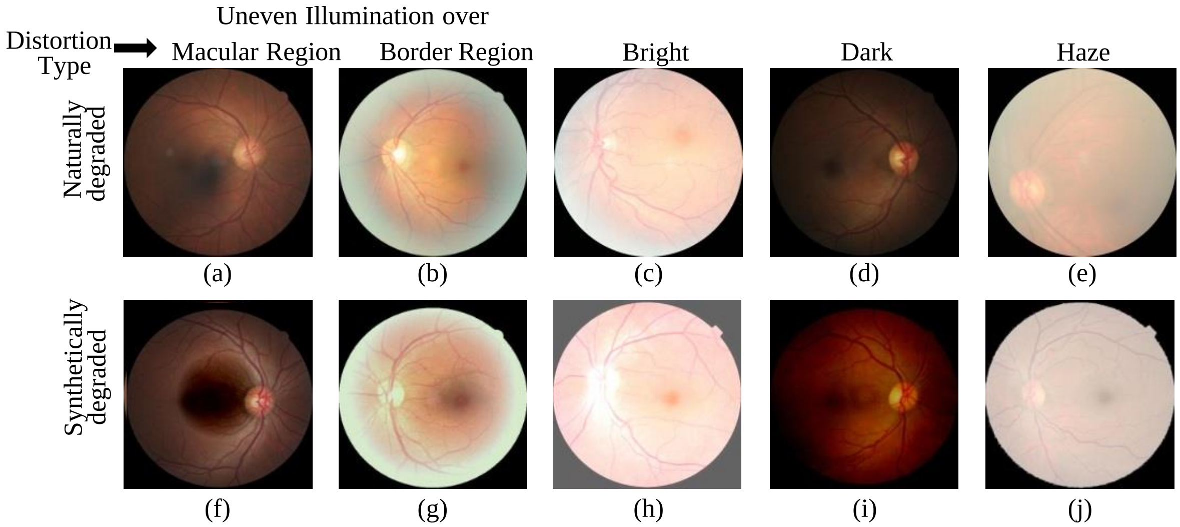


Figure 5.2 : Samples of naturally distorted and corresponding synthetically distorted fundus images. Here pair (a,f) represents MUI, (b,g) represents BUI, (c,h) represents bright, (d,i) dark, and (e,j) haze.

the border region, as shown in Fig. 5.2 (b). It was noticed that its intensity is high near the border and gradually decreases towards the centre. To create this distortion, the intensity values of the pixels around the border region of such various naturally degraded fundus images were analysed. Here, the border region was considered empirically within 50 pixels (about 10% of the image resolution) distance from the fundus boundary. The histogram of the pixel intensities within the selected region was obtained, and the pixel value (hI) with the highest frequency was identified. Now, to implement the distortion, the scaled value of a heuristically chosen intensity (gI) was added to each of the pixels within the fundus region, with the condition that no intensity value becomes greater than hI. Here, the value of the scaling factor decreased away from the fundus boundary, and it was calculated similarly to the method used previously for the MUI distortion. It is to be mention that after a careful discussion with doctors, the gI value was set in the range of 50 and 100 with an interval of 10. The detailed procedural steps are provided in Algorithm 2. A total of 5000 samples of such distorted fundus images were created.

- **Improper Luminance:** In this case, the entire image region becomes highly bright or dark, as shown in Fig. 5.2 (c) and Fig. 5.2 (d), respectively. Here the brightness and darkness are not confined to a specific region, in contrast with the case depicted in Fig. 5.2 (b). To induce a brightness effect, the mean value was added to each of the pixel intensities of the images, while for darkness effect, the same was subtracted. Here, the overflow and underflow problems of > 255 and < 0 were addressed by forcing the values to 255 and 0, respectively. A total of 1000 samples of both bright and dark fundus images were created.

- **Haze:** is the last important distortion included in this study, as shown in Fig. 5.2 (e). Initially, the maximum value (v_1) of input fundus image (I) was obtained, and then, a heuristically selected haze intensity value (h) was added to it. Finally, the haze effect was obtained by multiplying the intensity value of the input image I with a factor $k = v1/(v1+h)$. In order to introduce variability in the degradation, two different levels of (h) values, 300 and 400, were heuristically selected. A total of *2000* such distorted fundus images were created.

Fig. 5.2 shows the resemblance between natural and synthetic degradations. Here, Fig. 5.2 (a), Fig. 5.2

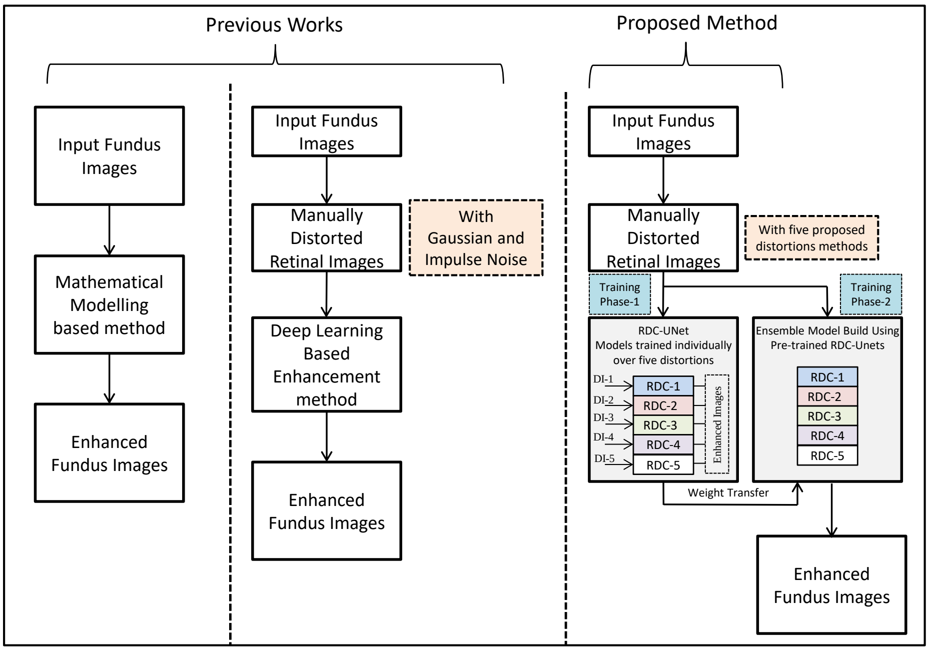
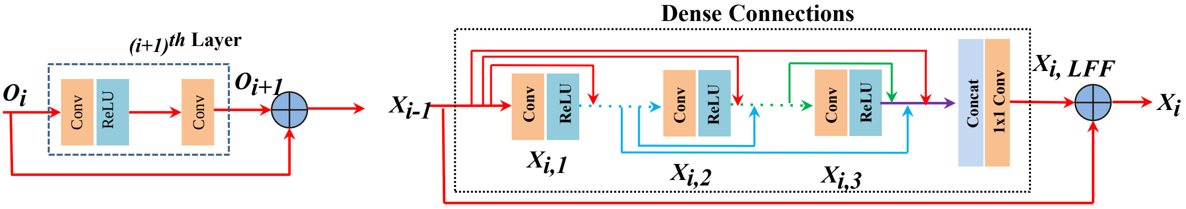

Previous Works
Proposed Method
Input Fundus Images
Mathematical Modelling based method
Enhanced Fundus Images
Input Fundus Images
Manually Distorted Retinal Images
With Gaussian and Impulse Noise
Deep Learning Based Enhancement method
Enhanced Fundus Images
Input Fundus Images
Manually Distorted Retinal Images
With five proposed distortions methods
Training Phase-1
Training Phase-2
RDC-UNet Models trained individually over five distortions
DI-1
DI-2
DI-3
DI-4
DI-5
RDC-1
RDC-2
RDC-3
RDC-4
RDC-5
Enhanced Images
Ensemble Model Build Using Pre-trained RDC-Unets
RDC-1
RDC-2
RDC-3
RDC-4
RDC-5
Weight Transfer
Enhanced Fundus Images

Figure 5.3 : Comparison Flow Chart of the state of the art fundus enhancement methods and the proposed method.

(b), Fig. 5.2 (c), Fig. 5.2 (d), and Fig. 5.2 (e) represent the naturally degraded fundus images and the images given in the second row, Fig. 5.2 (f), Fig. 5.2 (g), Fig. 5.2 (h), Fig. 5.2 (i), and Fig. 5.2 (j), are the synthetically generated images.



(a) Residual Block

(b) Residual Dense Block

Figure 5.4 : Architecture of (a) residual block with single skip connection. Here, O_i represents the output obtained from the i^{th} layer, and (b) residual dense block (RDB) with 3 conv layers. For an i^{th} RDB block, X_{i-1} and X_i represent the input and output, and $X_{i,n}$ represents the n^{th} conv layer. $X_{i,LFF}$ represents the reduced feature map obtained after applying the 1×1 conv layer. Here, LFF: local feature fusion.

5.2 PROPOSED MODEL

In our proposed model, the merits of UNet and residual dense blocks (RDB) containing residual and dense connections are exploited for the effective enhancement of fundus images. A visual comparison between the previous fundus enhancement works and the proposed model is illustrated in Fig. 5.3. It can be observed that the proposed model addresses the two major limitations of the previous works: (i) absence of a controlling factor in HE based methods and (ii) effectively handling the presence of frequently appearing distortions using CNN based methods. This section contains a precise introduction of the proposed RDC-UNet model.

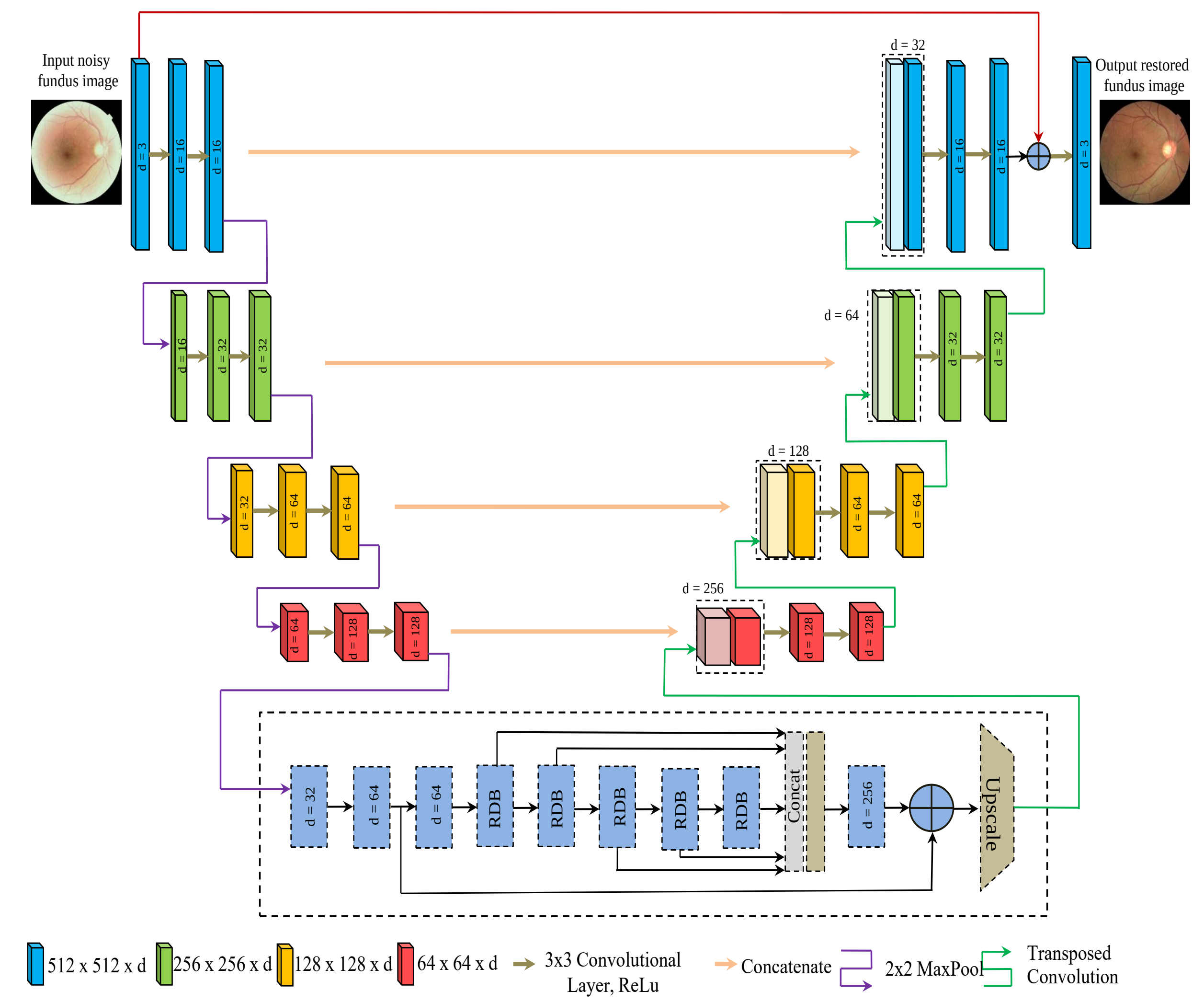


Figure 5.5 : Architecture of the proposed fundus image denoising model. RDB: Residual Dense Block, d: Depth of the feature map at each layer of the network model.

5.2.1 Preliminaries

- **UNet:** Ronneberger *et al.* [2015] proposed a CNN based model for medical image segmentation called UNet. It is one of the most successful CNN models for medical image segmentation problems. UNet mainly consists of three sections: encoder, decoder, and bottleneck. The encoder section performs the feature extraction process similar to other classification models such as AlexNet [Krizhevsky *et al.*, 2017], DenseNet [Huang *et al.*, 2017], etc. The feature extraction encoder part helps in capturing the contextual information corresponding to the output. For this, it performs two 3x3 convolutions followed by down-sampling using a pooling layer, repetitively. The second part is the decoder that performs the task of localising the captured features and then using the up-sampling operation to map it to the desired output. Here, after each up-sampling operation, the feature map is concatenated with the same scale of the channel corresponding to the encoder. Bottlenecks are the CNN blocks that make a model learn the compressed representation of the input data. The objective is to carry only such useful information that is sufficient for reconstructing the input image.

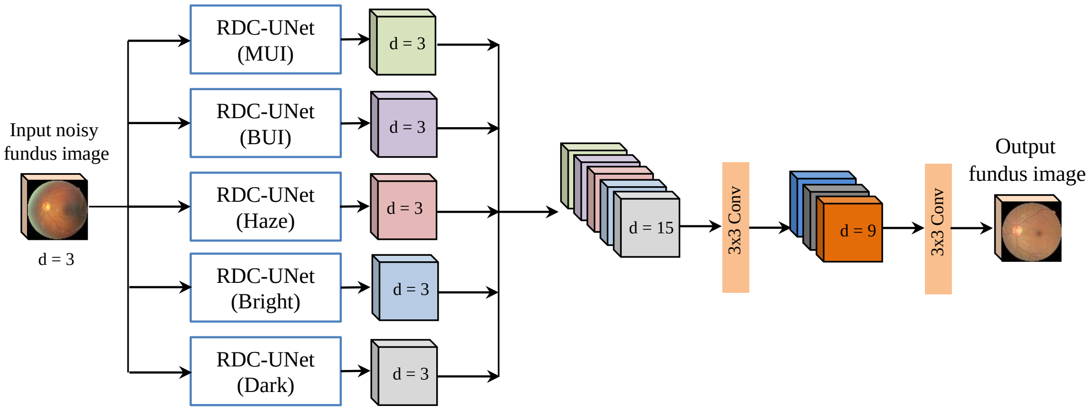


Figure 5.6 : Architecture of the proposed ensemble model built using the proposed RDC-UNet architecture (shown in Fig. 5.5). Here, d indicates the depth of the feature map at each layer of the network model.

Table 5.1 : Information of the number of images in each distortion category along with its train and test split.

Distortion	# of Images	Train	Test
MUI	5000	4000	1000
BUI	5000	4000	1000
Haze	2000	1600	400
Bright	1000	800	200
Dark	1000	800	200

- **Residual and Dense Connection:** Residual and dense connections are widely used to facilitate the vanishing gradient problem. The residual neural network uses the skip connections by re-utilising the weights from the previous layer for this purpose. In a typical residual block, the output of the current layer gets added to the output of its subsequent layer. A residual block with a single skip connection is shown in Fig. 5.4 (a).

  On the other hand, a densely connected network block has a direct connection between each layer and all other layers in the forward direction, as shown in Fig. 5.4 (b). It shows that all the feature maps obtained from the previous layers are utilised as input for every subsequent layers. Unlike ResNets, instead of summation, dense networks concatenate the obtained feature maps.

- **RDBs:** The RDB [Zhang *et al.*, 2018], illustrated in Fig. 5.4 (b), is one of the essential building blocks of the proposed RDC-UNet network architecture. RDBs are used to extract the significant local features using densely connected convolutional layers. It has three components: contiguous memory (CM) mechanism, local feature fusion (LFF), and local residual learning (LRL). It enables a direct connection from the previous RDB to each layer of the current RDB. This setting of connections is termed as the CM mechanism. For a i^{th} RDB, let X_{i-1} and X_i represent the input and output respectively. Then the output $X_{i,n}$ derived from the n^{th} convolutional (Conv) layer of the i^{th} RDB can be represented as follows:

$$X_{i,n} = \delta(W_{i,n}[X_{i-1}, X_{i,1}X_{i,2}........, X_{i,n-1}]) \tag{5.1}$$

Here, $W_{i,n}$ represents the weights of the n^{th} Conv layer, and δ indicates the ReLU activation function. The concatenation of the feature maps derived by the $(i-1)^{th}$ RDB is expressed as $[X_{i-1}, X_{i,1}X_{i,2}........, X_{i,n-1}]$. It enables the architecture to extract the local dense features. Further, LFF is used to reduce the number of features derived after the concatenation operation. It applies a 1×1 conv layer to the concatenated features derived previously, as given below.

$$X_{i,LFF} = C^{i}_{LFF}([X_{i-1}, X_{i,1}X_{i,2}........, X_{i,n-1}]) \tag{5.2}$$

Here, C^{i}_{LFF} represents the conv layer on i^{th} RDB. Further, LRL is introduced in the RDB to further improve the information flow by adding skip connections.

5.2.2 Proposed RDC-UNet for Fundus Image Denoising

For the denoising task, the incorporation of both global and local information is highly beneficial [Park *et al.*, 2019]. The proposed residual densely connected UNet (RDC-UNet) model is an enhanced version of standard UNet architecture proposed for the medical image segmentation task. The RDB block in our proposed architecture effectively facilitates abundant local features extraction and fusion. Additionally, unlike UNet architecture, based on the local features, we were able to construct hierarchical features due to the presence of shortcut connections between the different layers in RDB blocks. We also apply global residual learning [He *et al.*, 2016] in the proposed architecture between the input and final output block to generate output images by pixel-wise addition of learned residual information to the input images. Like UNet, the RDC-UNet model has three sections: encoding, decoding, and bottleneck. A detailed architecture of the RDC-UNet is shown in Fig. 5.5. Two 3x3 convolutions, ReLU activation [Nair and Hinton, 2010], followed by a max-pooling layer, were used at every stage in the encoding section. In the decoding section, transposed convolution [Dumoulin and Visin, 2016] was used, followed by a concatenation operation with the feature map obtained from the corresponding encoding section. Two 3x3 convolutions were further performed over the feature map obtained after the concatenation.

The lowest bottleneck level is considered one of the most crucial parts of the architecture. It is useful in the extraction of features that capture non-local image information. However, in the primitive UNet model, the bottleneck layer couldn't make full use of the hierarchical features, obtained from the previous layers, using simple convolution. Because of this, the accuracy and effectiveness of the model decreases. As a result of this drawback, we propose a residual densely connected UNet architecture for the task of denoising. In the proposed model, we used a series of RDBs in the bottleneck layer for extracting local dense features. After that, the extraction of global features was done by fusing features from all RDBs. This helps in exploiting hierarchical features available in a global form. Moreover, a global residual connection was made between the initial and final block, as shown in Fig. 5.5. It helps in creating a smooth flow of gradient in the overall network. We tried various architectures by varying the number of convolutional layers in the RDB and the number of RDBs in the overall architecture. The best setting was found with three convolutional layers in RDB and five RDBs in the overall architecture. The number of filters at each level can be observed from Fig. 5.5 itself.

The RDC-UNet is trained in a supervised manner for each of the individual distortions, explained in section 5.1. However, in the case of naturally distorted fundus images, it is highly challenging to identify the exact type of distortion present in the image. Furthermore, there can be multiple distortions present in the image. To address this challenge, we utilised a popular machine learning concept called ensemble learning [Liu and Yao, 1999]. The objective of ensemble learning is to achieve higher prediction results by combining multiple models. A new multi-channel CNN is proposed that leverages the advantages of ensemble learning by combining each individually trained RDC-UNet, illustrated in Fig. 5.6. This architecture consists of multiple convolution blocks that take concatenation of predictions from the five distortions models. It has two convolutional blocks with a number of feature maps of varying sizes. The overall model is trained in two stages. Firstly, the five models are trained individually using the same pre-processing operation, and later, the ensemble model is trained with the predictions obtained from the model in the first stage of training.

5.2.3 Implementation Details

- **Pre-processing**: All the images before training were normalised in the range of [-1,1]. The images were normalised using the following formula:

  $$I_n = \frac{I - I_{max/2}}{I_{max/2}} \tag{5.3}$$

  Here, I_n and I represent the normalised and a single channel (RGB) of the input fundus image, respectively. Also, $I_{max/2}$ represents the half of the maximum possible intensity. Here, for each channel(8-bit) of a RGB image, the value of $I_{max/2}$ would be 127.5.

- **Loss Function**: For training and validation, a hybrid loss function was derived using the sum of two loss metrics: mean absolute error (MAE) and structural similarity index (SSIM) [Zhou Wang *et al.*, 2004] loss. The mean absolute error (MAE) is better at training a CNN model with reduced average error between the input and predication. Therefore, to train a CNN model for image enhancement tasks, it is beneficial to use MAE loss function to predict the enhanced image statistically (in terms of pixels values) close to the reference image. On the other hand, the SSIM is a perceptual metric that quantifies image quality degradation by analyzing the structural change occurring in the image due to some processing. So, for distilling the advantage of both loss functions for the image enhancement, we used summation of losses for training. It resulted in an image which appears to be of higher quality.

  The mathematical representation of the MAE is as follows:

  $$L_{MAE} = \frac{1}{m \times n} \sum_{i=1}^{m} \sum_{j=1}^{n} |(\hat{x}(i,j) - x(i,j))| \tag{5.4}$$

  Here, $x(i,j)$ and $\hat{x}(i,j)$ represent the input and the enhanced image, respectively, with a resolution of $m \times n$. The MAE is preferred over the mean square error (MSE) because MSE is prone to being affected by outliers or wrong predictions, as it gives high weightage to large errors in comparison to small errors.

  Next, SSIM is a quality assessment metric proposed by Wang *et al.* [Zhou Wang *et al.*, 2004]. It quantifies the level of degradation in the image by extracting its structural information. The SSIM score between a reference image *(r)* and an enhanced image *(p)* can be represented as $SSIM(r,p)$.

  $$SSIM(r,p) = \left(\frac{2\mu_r \mu_p + x_1}{\mu^2{}_r + \mu^2{}_p + x_1} \right) \cdot \left(\frac{2\sigma_r \sigma_p + x_2}{\sigma^2{}_r + \sigma^2{}_p + x_2} \right) \tag{5.5}$$

Here, μ_r and μ_p are the mean values, σ_r and σ_p represent the standard deviation values, and $\sigma^2{}_r$ and $\sigma^2{}_p$ is the covariance values of r and p, respectively. Additionally, x_1 and x_2 are the small positive constant values added to ignore the numeric instability. It generates value in the range of $(0,1]$, where a higher value indicates better quality and a lower value vice versa. Therefore, the loss value derived using SSIM can be given as follows:

$$L_{SSIM} = 1 - SSIM(r, p) \tag{5.6}$$

finally the overall loss (L) is derived using the summation of L_{MAE} and L_{SSIM}:

$$L = L_{MAE} + L_{SSIM} \tag{5.7}$$

- **Computational Set-up:** All the models were trained on a computer system of 2.0 GHz CPU with NVIDIA V100 GPU of 32 GB memory. The adaptive moment estimation (ADAM) [Kingma and Ba, 2014] optimisation method was used for error minimisation with a learning rate of 5×10^{-4}. The ADAM was performed for 1000 epochs with a mentioned batch size of 24 images during the training process and inference time was 0.083 sec. Moreover, the weight decay regularisation method with a value of 10^{-6} was applied after every 100 epochs. All deep learning models were implemented using the Python programming language and Keras library [Chollet, 2015]. The pre-processing (resize and crop) and data loading tasks were achieved using the PIL Image library.

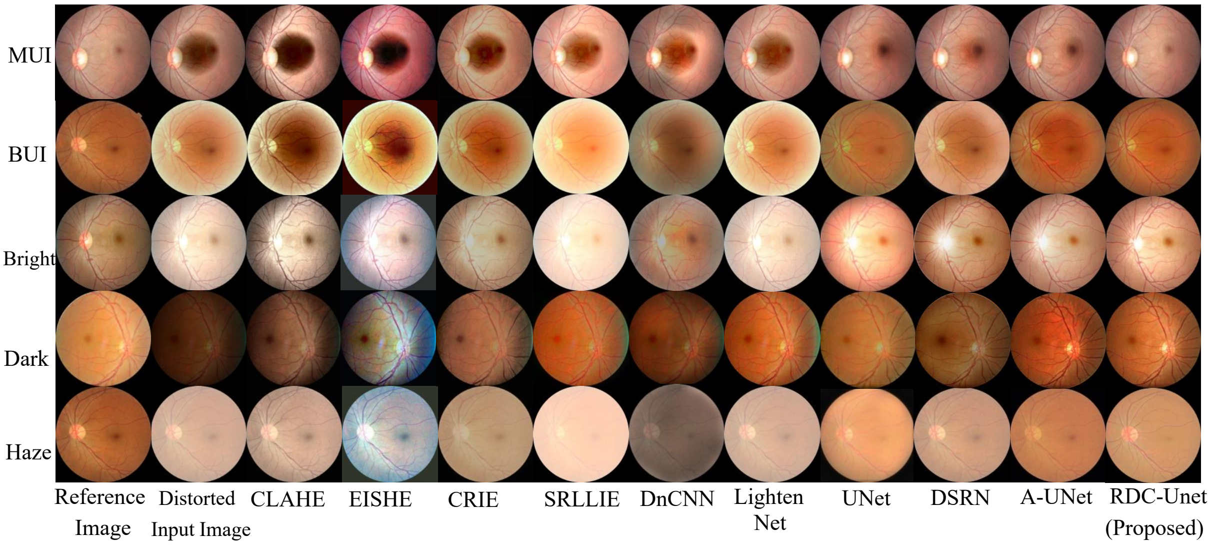


Figure 5.7 : Visual comparison of the results obtained from the proposed model using state-of-the-art methods with synthetically distorted images.

5.3 RESULTS AND ANALYSIS

5.3.1 Data

As mentioned earlier in section 5.1, a total of 1000 good quality fundus images were randomly selected as reference images. Thereafter, a total of 14000 manually distorted images were created with the distortions mentioned earlier in the paper. To train and test the proposed denoising model, the images were split into an 80:20 ratio in a disjointed manner. Table 5.1 contains information about the number of images belonging to each distortion category and their respective training and testing split. In addition to the synthetically generated data-set, the performance of the proposed model illustrated in Fig. 5.6 was also tested over a total of 1000 naturally distorted fair quality fundus images taken from the EyeQ data-set.

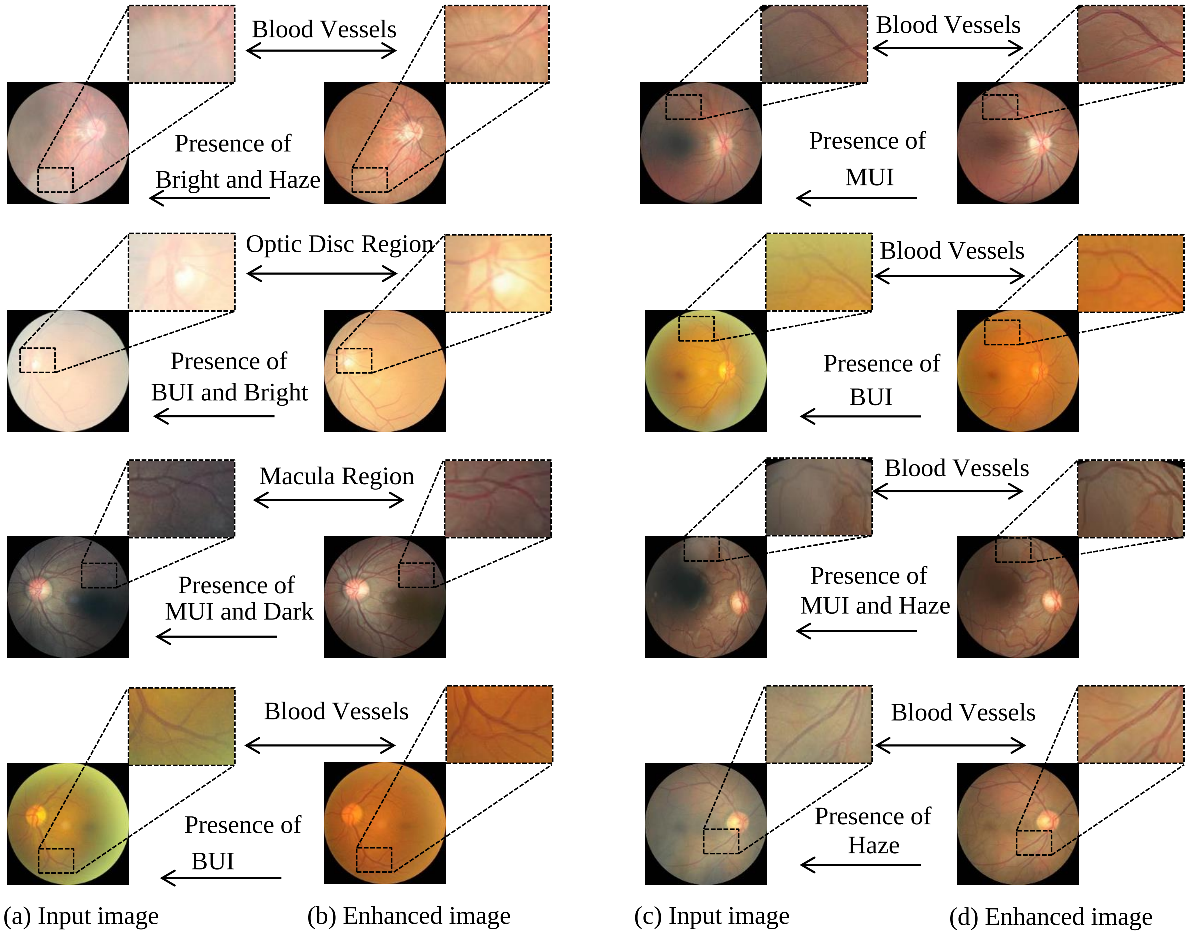


Figure 5.8 : Performance of the proposed ensemble model in terms of visual clarity. Here (a) and (c) represent naturally distorted images containing multiple distortions, and (b) and (d) are the corresponding enhanced images.

5.3.2 Evaluation Methodology

The performance evaluation of the proposed method was carried out on both synthetically and naturally distorted fundus images. In the case of synthetically degraded fundus images, for every input, a corresponding clean image was present. Therefore, full reference image quality assessment (FR-IQA) methods were applied for the performance evaluation. We employed the two most widely used FR-IQA metrics [Bosse *et al.*, 2016; Kim and Lee, 2017], namely, peak signal to noise ratio (PSNR) and SSIM. Here, PSNR determines the ratio between the maximum power (pixel intensity) of the reference image and corrupted image. SSIM quantifies the quality of an input image by analyzing its structural similarity with the reference image. The case of a natural image falls under the category of no-reference (NR) IQA. Therefore, we employed a MvRCNN model [Raj *et al.*, 2020] for fundus IQA for the quality evaluation of the enhanced output images. Additionally, for a comparative analysis, the performance of nine widely used and state-of-the-art methods were also included in this study. Out of nine, four are histogram equalisation-based methods ESIHE [Singh and Kapoor, 2014], CLAHE [Shome and Vadali, 2011], CRIE [Zhou *et al.*, 2018], and SRLLIE [Li *et al.*, 2018], and the other five are deep learning based methods DnCNN [Zhang *et al.*, 2017b], LightNet [Li *et al.*, 2018], UNet [Ronneberger *et al.*, 2015], DSRN [Das *et al.*, 2021], and A-UNet [Oktay *et al.*, 2018].

Further, to demonstrate the effectiveness of the proposed methods, application-based

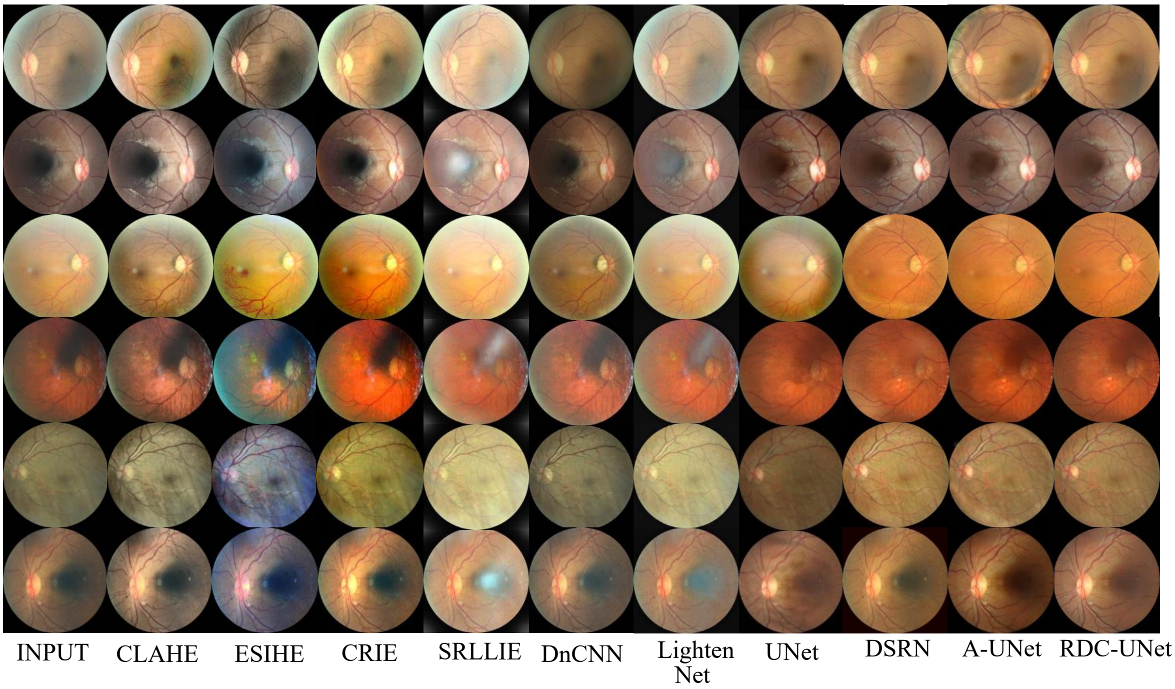


Figure 5.9 : Comparative performance of RDC-UNet based ensemble model with state-of-the-art methods in terms of visual observation over naturally degraded fundus images.

experiments were also conducted. These included the analysis of the effectiveness of the proposed model using the blood vessel segmentation task. For this purpose, the DRIVE [Staal *et al.*, 2004] image dataset was chosen, which provides ground truth for the annotations of the blood vessel. The standard UNet model was taken as a baseline method for the evaluation.

5.3.3 Performance analysis over synthetically degraded fundus images

The PSNR and SSIM values of the proposed RDC-UNet model, together with these methods, have been provided in Table 5.2. In addition, an illustration of the performance of the proposed model with these methods has also been shown in Fig. 5.7. The individual performance of the first three methods over dark degradation is relatively better in comparison to other degradations. However, in terms of overall performance, these methods are not satisfactory, and the same can be observed from Table 5.2. As mentioned earlier, histogram equalisation-based methods have no mechanism to control the level of enhancement. Due to this, they do not work effectively on the images containing significantly dark or bright regions. Further, the DnCNN model extracts the noise component from the degraded image and then subtracts from it. It is useful in extracting the noises such as the Gaussian that follows normal distribution. In case of uneven illumination, it is a challenging task to extract the noise component. This is especially so when image contains dark or bright regions confined to particular regions. The results obtained from the DnCNN are better in comparison to histogram equalisation-based methods but not satisfactory. The accuracy of the basic UNet model was also tested and reported the Table 5.2. The basic UNet model performs better than the other methods, but it lacks in performance with high bright and dark distortions. Additionally, the results obtained from the basic UNet model suffers from blur problems. Additionally, two other advance hybrid UNet models DSRN and Attention-UNet are also tested. The DSRN model uses the dilation mechanism and the Attention-UNet uses the attention mechanism. Both the models has achieved results better than the standard UNet

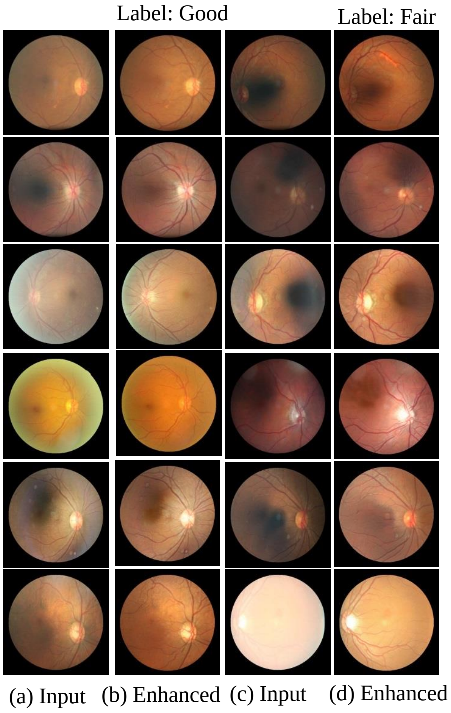


Figure 5.10 : Samples of the quality evaluation results obtained from the MvRCNN model. Here, column (a,c) represents naturally distorted fundus images, (b) shows the predicted images labelled as good, and (d) shows the images labelled as fair quality.

architecture. However, the proposed RDC-UNet with the RDB as a bottleneck layer has outperform all the models and shown the best results. The investigation for the reason behind the obtained results yields the limitation of both the dialation and attention mechanism. We found two issues with dilation in particular. First, it is not efficient towards extracting the local information. Second, after a certain level it results in a weak correlation between the neighbouring units. Also, the output obtained of the attention gate subsequently gets smaller by each higher layer. It results in to the loss of local information of the image.

5.3.4 Performance analysis of naturally degraded fundus images

The proposed ensemble model was tested on *1000* naturally degraded fundus images. The visual quality of the obtained enhanced images were found to be satisfactory by ophthalmologists. For visual clarity and comparison purposes, a few sample images are shown in Fig. 5.8 and Fig. 5.9. The Fig. 5.8 contains a total of *eight* pairs of naturally degraded images with multiple distortions and their respective enhanced fundus images. It also shows a magnified clip of the one of the affected areas containing blood vessels, the optic disc, and the macula in the input image and its enhanced version. Here, Fig. 5.8 (a) contains samples of fundus images distorted with bright-haze, BUI-bright, MUI-dark, and BUI from top to bottom, respectively, and the corresponding enhanced images are shown in Fig. 5.8 (b). Similarly, Fig. 5.8 (c) contains samples of fundus images distorted with MUI, GUI, MUI-haze, and haze distortions from top to bottom, respectively. Their corresponding enhanced images are shown in Fig. 5.8 (d). It can be observed that the our proposed model effectively handles the presence of multiple such distortions. In addition, for comparative performance analysis purposes, in Fig. 5.9 the output fundus images obtained from the proposed method are shown along with other competitive methods, mentioned in subsection 5.3.3. It can be observed that the histogram equalisation-based methods have suffered from over-enhancement issues. Fundus images with extra dark and bright regions remain

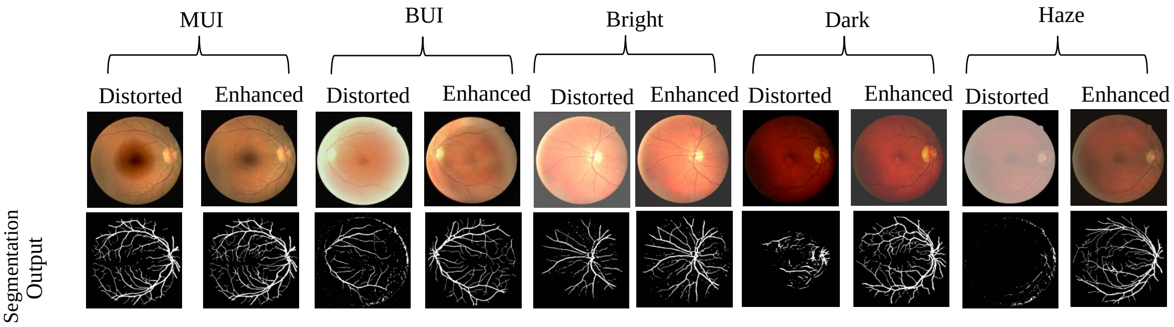


Figure 5.11 : Predicted segmentation map results obtained for the synthetically distorted and respective enhanced images from the DRIVE dataset.

unaffected. Further, learning-based methods DnCNN, UNet models effectively handled such issues but lacked with regard to uneven illumination and multiple distortion issues. However, the proposed model is able to efficiently address such distortions. This is because the training of the model was performed by combining the optimal feature maps that are obtained by training the RDC-UNet individually for each type of distortion.

5.3.5 Quality evaluation

The quality of the enhancement results for naturally degraded images was evaluated using the recently reported MvRCNN model [Raj *et al.*, 2020]. Out of 1000 such images, the model categorised 417 as of good quality while the other 583 remained as fair quality. It is important to mention that to ensure reliable diagnosis, the sensitivity of MvRCNN towards good quality is high. Due to this factor, images containing even a small amount of distortions were rejected to be labelled as good quality. In addition, many of the images labelled of fair quality in the EYEQ data-set contained comparatively high distortions. These were the border cases of poor and fair quality. The proposed RDC-UNet model recovered such types of images up to a certain extent but still remains short in pushing them into good quality. An illustration of the images labelled as good and average are shown in Fig. 5.10. It can be observed that the input images shown in Fig. 5.10 (c) are highly distorted ones, labelled as fair quality in the said database, and an enhanced version of the same is given in Fig. 5.10 (d). The enhancement is significant but not enough to be labelled as good quality to be used for reliable diagnosis.

5.3.6 Blood Vessel Segmentation

The ultimate objective of retinal image enhancement work is to enhance real clinical tasks. Therefore, we performed experiments on retinal blood vessel segmentation to demonstrate the effectiveness of the proposed methods. The popular DRIVE [Staal *et al.*, 2004] dataset was used where annotations for the blood vessels are provided. The UNet [Ronneberger *et al.*, 2015] model was considered as the baseline method for the segmentation task. Initially, the UNet model was trained using the images provided in DRIVE dataset with an area under the curve (AUC) value 0.97. Furthermore, the tested images were synthetically distorted with each of the five distortions. Thereafter, enhanced images were obtained using the proposed RDC-Unet model. Finally, the segmentation output for each distorted and their respective enhanced images were obtained. For synthetically distorted images, the average value of the obtained AUC is 0.493. However, for the respective enhanced images the obtained AUC vlaues is 0.856. It can be observed that there is a significant increase (1.7 times approx.) in the performance of the segmentation model. The visual representation of the obtained results are presented in Fig. 5.11. In addition, we also tested the accuracy of the segmentation model on naturally

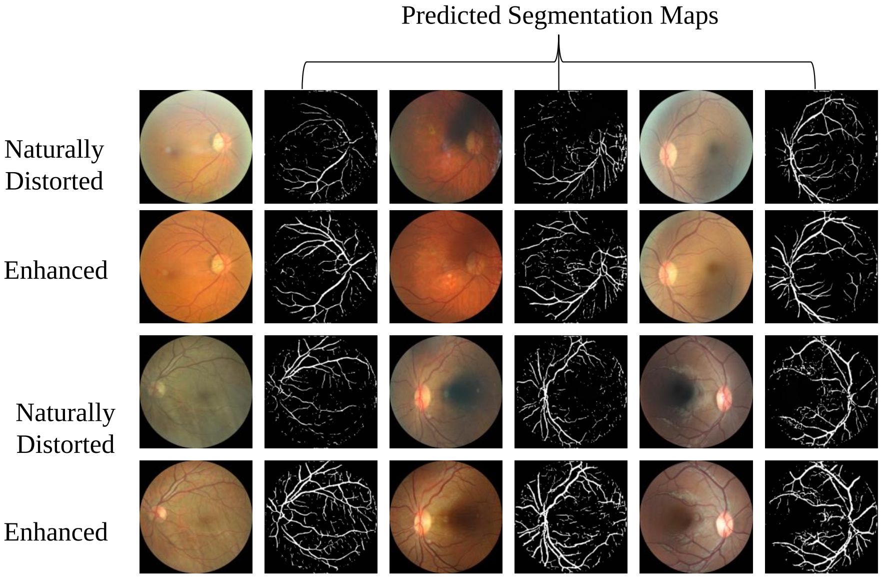


Figure 5.12 : Predicated segmentation map results obtained for naturally distorted and respective enhanced images.

distorted fundus images. In this case, no ground truth available for the quantitative analysis. However, for visual assessment, the obtained results are shown in Fig. 5.12. It can be observe that the blood vessel maps obtained for the enhanced images are significantly better than the maps of original images. The results demonstrate the efficiency of the proposed RDC-UNet model for the retinal image enhancement model.

5.4 DISCUSSION

A good retinal image enhancement method is expected to effectively suppress the presence of distortions that occur particularly in fundus images. The histogram processing based methods are effective in handling the natural distortions. However, such methods often suffer with over enhancement problem that is caused due to the absence of a controlling factor to balance the level of required enhancement. CNN based learning methods effectively address such limitations. For the enhancement task, use of supervised learning based models is advisable due for better error correction. Such methods require a reference image for each distorted image. However, w.r.t the naturally appearing distortions in retinal images, it is difficult to have reference images. One solution is to capture a low quality retinal image simultaneously with a good quality image by creating some noisy image acquisition environment. However, it is certainly not a cost effective approach in term of time, money, and human efforts. Through this paper we address the above mentioned limitations and below a detailed insights are provided for the proposed work.

- **Distortion Generation:** First, a total of five common degradations occurring in fair quality fundus images were identified: (i) uneven illumination over macula, (ii) uneven illumination over border region, (iii) bright, (iv) dark, and (iii) haze. Thereafter, algorithms proposed to create distortions that closely resemble the above mentioned distortions. A total of 1000 good quality images were randomly chosen as reference images from the EyeQ data-set. Now, With the help of the proposed algorithms, a data-set of 14000 degraded fundus images were created. We would like to

Table 5.2 : Comparative performance analysis in terms of PSNR and SSIM values. P: PSNR, S: SSIM.

Distortion →	**Bright**		**Dark**		**BUI**		**MUI**		**Haze**	
Method ↓	P	S	P	S	P	S	P	S	P	S
ESIHE [Singh and Kapoor, 2014]	12.22	0.47	11.49	0.49	9.42	0.40	14.91	0.42	9.87	0.46
CLAHE [Shome and Vadali, 2011]	13.28	0.58	14.16	0.50	10.65	0.51	16.83	0.60	10.05	0.49
CRIE [Zhou *et al.*, 2018]	16.37	0.71	17.99	0.65	21.66	0.66	22.00	0.64	9.87	0.44
SRLLIE [Li *et al.*, 2018]	13.45	0.49	17.23	0.54	16.03	0.53	25.71	0.67	15.41	0.51
LightNet [Li *et al.*, 2018]	17.41	0.55	22.10	0.57	24.21	0.63	34.07	0.72	13.14	0.46
DnCNN [Zhang *et al.*, 2017b]	24.41	0.73	21.80	0.68	28.51	0.73	33.87	0.79	24.94	0.72
DSRN [Das *et al.*, 2021]	26.75	0.68	26.18	0.67	29.21	0.75	33.71	0.83	28.81	0.78
UNet [Ronneberger *et al.*, 2015]	29.45	0.74	28.22	0.76	32.71	0.83	37.67	0.89	30.48	0.83
A-UNet [Oktay *et al.*, 2018]	33.15	0.83	30.01	0.85	39.17	0.91	41.49	0.88	30.80	0.84
Proposed	36.53	0.89	34.41	0.88	42.74	0.91	50.91	0.93	37.43	0.87

mention that, in this work most of the fundus images used are macula centered. Therefore, the macular uneven illumination (MUI) distortion is created starting from the center of the image, assuming that the macula is located near the center. Now, considering the experiences of the ophthalmologists and observatory experiments yields that the spatial location of the MUI could be anywhere around the macular region and also in any direction. There exist various possible combinations of shapes and directions while creating the MUI distortion. Therefore, to increase the model's robustness towards the equiprobable spatial directions, a circular area around the macula is selected.

- **RDC-UNet:** Furthermore, we proposed a residual dense connection based modified UNet architecture, as shown in Fig. 5.5. The proposed model has two stage training procedure. The first stage of training is achieved by training the RDC-UNet indiviually for each of the five distortions. Here, it is to mention that before finalizing the individual training idea, we implemented a UNet model based on a single shared encoder and separate decoder paths for each noise type. However, the obtained results were not satisfactory. The primary reason behind such results was that the model was expected to handle the many-to-one output condition. Using a single network was a situation where the network was trained to map many inputs to a single output. As in our case, there were a total of five distortions generated from a single fundus image. In addition, there also exist multiple levels of each distortions. This input condition was confusing the network for performing the correct mapping from a noisy input image to a clean output image. This lead to the bad performance of the model. Therefore, we opted for the individual training approach to solving the problem. The second stage of training was done over the porposed ensemble model, as shown in Fig. 5.6. The proposed two stage training effectively suppresses the presence of multiple such distortions in a naturally degraded retinal image.

It is also important to mention that our intuition to use RDC blocks in the bottleneck region of UNet proved to be beneficial for the enhancement task. As it effectively captures both the local and global information from the images. Experiments conducted with applying RDC with the encoder and decoder section as well. However, increasing the number of parameters in the model in terms of RDC block were not benefiting the overall enhancement accuracy in comparison with the proposed RDC-UNet. Based on the above reason we moved with the proposed architecture of the RDC-UNet.

- **Model Performance:** The number of parameters in the base UNet model is 29.24M, whereas the number of parameters in RDC-UNet is 41.26M. As mentioned in subsection 5.2.2 several experiments were conducted for the optimal setting of the number of RDB Blocks. With the mentioned specifications, the performance of the proposed RDC-UNet model was found to be better than other methods. In addition, our prime objective was to get the best possible quality scores for effective enhancement, leading to reliable diagnosis. It is important to note that further increasing the number of parameters did not increase the quality score (PSNR and SSIM) on the evaluation metric for the proposed RDC-UNet architecture. In addition, complex and larger models have different effects on relative improvements over accuracy as well as runtime. Making the model larger merely by increasing the number of channels (filters/ or layers) may have a linear improvement up to some extent, but later the model starts overfitting, and there were sharp reductions observed in testing accuracies. In the proposed architecture, the global and local artifacts in the images are addressed simultaneously using residual dense blocks.

5.5 SUMMARY

- Through this work, a novel approach is proposed to address the fundus image enhancement challenge. First, a total of five commonly appearing distortions in fundus images were identified: (i) MUI, (ii) BUI, (iii) high bright, (iv) extra dark, and (v) haze. Thereafter, algorithms are proposed to synthetically create the distortions closely resembling the same.

- Further, a modified version of UNet architecture RDC-UNet containing residual dense connections is proposed for the enhancement task. Initially, the proposed RDC-UNet was trained individually for each of the mentioned distortions. The experimental results demonstrate that the proposed RDC-UNet model achieves a high PSNR (50.91, 42.74, 36.53, 34.41 and 37.43) and SSIM (0.93, 0.91, 0.89, 0.88 and 0.87) values for (i), (ii), (iii), (iv), and (v), respectively, which are significantly higher than those of other state-of-the-art methods. Furthermore, naturally degraded fundus images might contain multiple such distortions at a time.

- Therefore, a new ensemble learning-based model is proposed to make the model work efficiently over naturally distorted fundus images. It was built using RDC-UNet trained individually for each of the above mentioned distortions. It helped to reduce the multiple distortions effectively by capturing the relevant information from these five models.

- The performance of the model is tested over a total of 1000 naturally degraded fair-quality images. The quality of the obtained enhanced images is tested using the MvR-CNN model. The obtained results show that the proposed ensemble model has recovered approximately 41% images.

- In addition, the clinical significance of the model is also demonstrated with the help of blood vessel segmentation application. Standard UNet based segmentation model is trained over DRIVE dataset. The segmentation model is tested over both distorted and its respective enhanced images. The AUC values obtained for the enhanced images are 1.7 times higher than the AUC values of distorted images.

- The performance evaluation results show that the proposed approach effectively suppresses multiple distortions present in the images. All the experimental results indicate the effectiveness of the proposed approach that could potentially fill the gap of the unavailability of a labeled dataset.

…