University of Glasgow

SINGAPORE INSTITUTE OF TECHNOLOGY

B.Eng (Hons)
Mechatronics Engineering
Final-Year Project 2019/20

TITLE: Removal of rain using Wavelet CycleGAN

NAME: Teo Lim Fong

UoG MATRICULATION No.: 2427466T

FIRST SUPERVISOR: MR Tang Lai Meng

SECOND SUPERVISOR: Dr Liu Xiao Lei

School of Engineering

Acknowledgements

I would like to express my sincere thanks to Mr Tang Lai Meng for his continuous guidance. I am honoured to assist him with his PhD research. Without his weekly guidance and support, I will not be able to implement a novel algorithm. Thanks to Mr Tang for sharing his experience and for explaining the utility of machine learning in the future. I deeply appreciated every meeting I had with him. I am lucky to have Mr Tang as my first supervisor as he always made time for me.

Thanks to my parents for their support and finance. They contributed by affording me a super computer. Without them, I will not able to train my program efficiently.

Lastly, I also want to thank anyone whom helped me directly or indirectly on my project.

Table of Contents

List of Figure

List of Table

1. Introduction

With the advent of technology, self-driving mobile also known as Autonomous Vehicle (AV) will gradually find its way into our daily lives in the coming years. However, environmental noise such as rain will disrupt the path of the AV which may result in accidents. To eliminate such risks, many researchers have been innovating and implementing new methods to remove rain. The most recent rain removal algorithms are the ID CGAN (Image De-raining using Conditional Generative Adversarial Network) [33] & CycleGAN [32].

In article [32], *Lai Meng Tang* has proven that CycleGAN as shown in *Figure 2* is capable of removing rain more efficiently than ID CGAN using synthetic data sets. However, CycleGAN produced discoloration when removing the rain and the low frequency noise as shown in *Figure 3* can be further enhanced and eliminated. Therefore, Wavelet CycleGAN was implemented to enhance the existing CycleGAN by removing rain more efficiently and preserving the colours of the image.

| Figure 1 | Figure 2 | Figure 3 |

Figure 1: Real Rain Image , Figure 2: Rain Removal Result by CycleGAN, Figure 3: A zoom-in rain streak in Figure 2

1.1 Objective and Aims

My objective is to create and implement a novel algorithm called the Wavelet CycleGAN by applying the concept of wavelet transform into the existing CycleGAN to:

- Remove rain efficiently
- Enhance the structure of the images
- Retain the original colour of the images
- Achieve a lower NIQE (Naturalness Image Quality Evaluator) value

2. Literature Review

2.1 Cycle Generative Adversarial Networks (GANs)

CycleGAN as shown in *Figure 4* is revised to improve the accuracy and prediction of GANs [1] by integrating two Generators and two Discriminators into the neural network. [2] In CycleGAN, unpaired training dataset is used for training, From the article [3], unpaired training dataset is proven to be better than paired training dataset as it does not require any information from Generated X [2]. On the other hand, paired dataset is expensive and has limited resources [4] such as semantic segmentation [5].

The purpose of having two domains in the network is to successfully learn a mapping using the given training datasets. For example, Generated X (synthetic rain) will learn a mapping function from G :X → Y (Generator X to Y) to become indistinguishable from Generated Y (rain-free). Adversarial loss [2][6] is applied to both mode G :X → Y and its Discriminator (D_Y) to calculate the minimum and maximum value. The formula is given by:

$$\mathcal{L}_{GAN}(G, D_Y, X, Y) = \mathbb{E}_{y \sim data(y)}[log D_Y(y)]$$
$$+ \mathbb{E}_{x \sim data(x)}[log(1 - D_Y(G(x)))] \text{ , ----- } Equation\ (1)$$

where G is the Generator X to Y, D_Y is the Discriminator in Generated Y, X is the Generated X (synthetic rain) and Y is the Generated Y (rain-free).
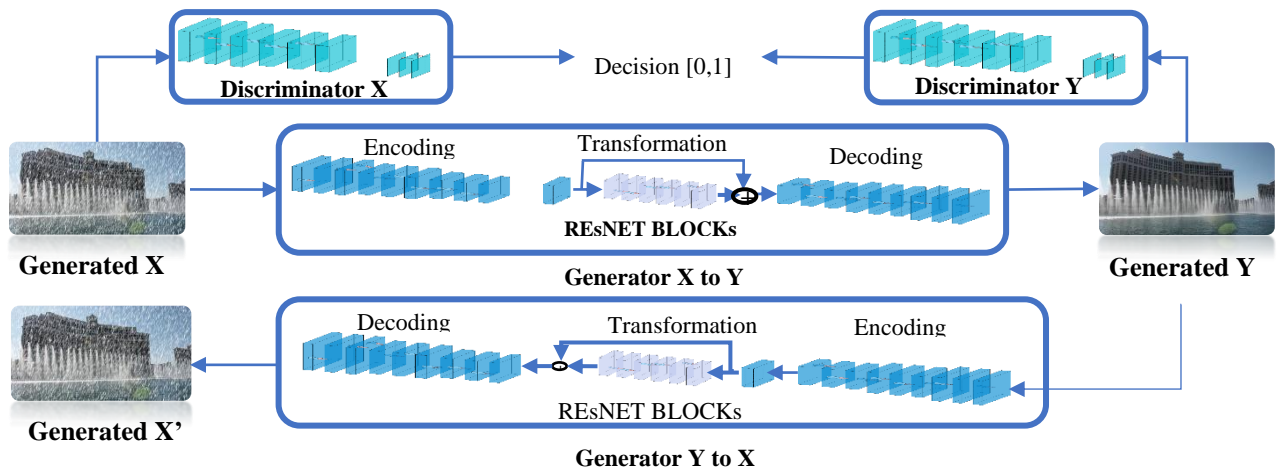
Figure 4: CycleGAN Neutral Network

In *Equation 1*, Generator (G) tried to make Domain Y indistinguishable by minimizing the adversarial loss however, Discriminator (D_Y) tried to make it distinguishable by maximising the adversarial loss. [2] Cycle Consistency loss can be found by mapping the inverse mapping function F :X → Y on Generated Y. The aim of cycle consistency loss is to force Generated X' to be Generated X. As such the combination of Cycle Consistency loss and adversarial loss can yield the overall result in CycleGAN.

Generator consists of Encoding, Transformation and Decoding.

Encoding
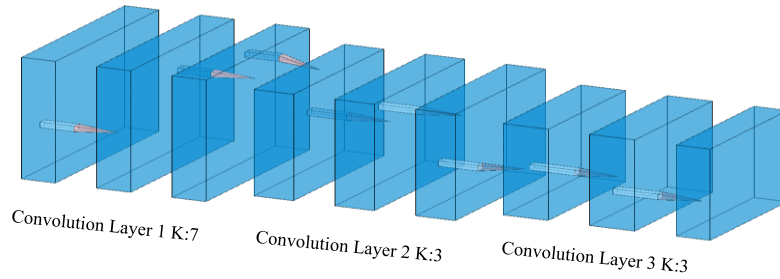
Figure 5: Encoding phase in Generator

Encoding as shown in *Figure 5* is the first process in the Generator. The input is assumed to be RGB images. Within the encoding process, multiple filter windows will move across the images to extract the different features. The first convolution layer consists of 256x 256 x 64 activation size and 64 x 7 x 7 kernel size for the convolution layer. For the subsequent convolution layer, it will extract higher level features from the image. [7]

Transformation

REsNET BLOCKs
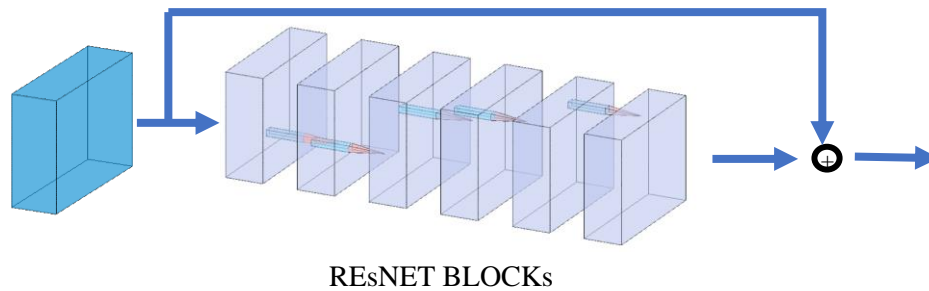
Figure 6: Transformation phase in Generator

Transformation as shown in *Figure 6* is the process that convert convolution layer into feature vector. It also consists of two convolution layers where the residue of input is added to the output. This is done to ensure the properties of the input are available for later layers, so the output will not deviate much from its original input.

Decoding
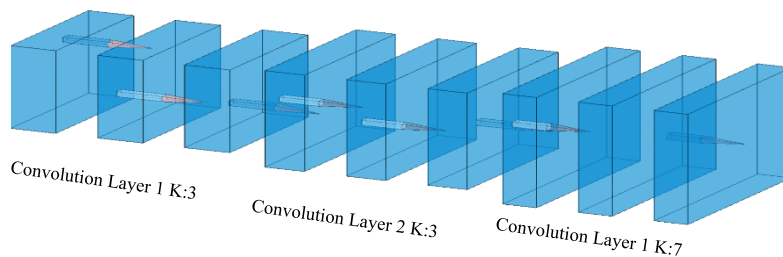
Figure 7: Decoding phase in Generator

Decoding as shown in *Figure 7* is the last process in the Generator. Decoding is the opposite of Encoding which is called Deconvolution. It converts the new feature vector from Transformation back to low level feature and slowly up sample it back to RGB images.

2.2 Wavelet Transform

Wavelet Transform (WT) as shown in *Figure 8* contains superior frequency information and time resolution when compared to Fourier Transform and Short-Time Fourier Transform. [10][15]
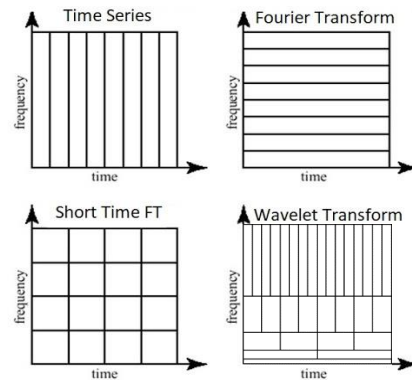
Figure 8: Comparing Wavelet Transform time resolution and frequency information with other algorithms

Fourier Transform (FT) is a great tool for data analysis. However, it does not represent abrupt changes efficiently. This is due to the fact that Fourier Transform does not localize in time. On the other hand [13], Short-Time Fourier Transform (STFT) is a time-frequency transform [9] that measures fixed time-frequency windows. This operation is not ideal due to the inflexible time-frequency windows resulting in inaccurate data analysis that is related to changes in time.
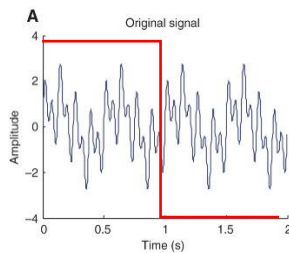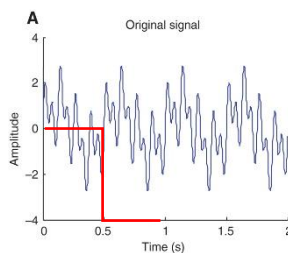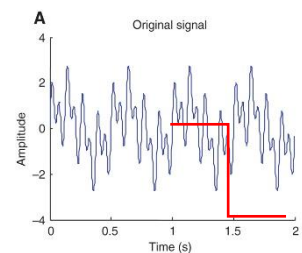
Figure 9   *Figure 10*   *Figure 11*

Figure 9: A large dynamic scaling window covered the signal, Figure 10: A down sampled dynamic scaling window to extract more features ,Figure 11: Translation - The dynamic scaling moving across the signal

In Wavelet Transform, a dynamic scaling window is used to solve the signal-cutting problem. The window is shifted along the signal and captured every different position of the spectrum. The size of the window and the shifting along the signal is often called the Scaling and Translation respectively as shown in *Figure 9* [10].

Scaling and Translation are two important parameters in Wavelet Transform. Scaling factor can be represented by α and Translation can be represented by **b** [14]. When the scaling factor is large as shown in *Figure 9*, the stretched wavelet will help to capture the low frequency components and the slow varying changes in the signal. Low frequency components contain great frequency information but is not localised in time as shown in *Figure 8*. When the scaling factor is reduced as shown in *Figure 9*, the compressed wavelet will help to capture the high frequency components by moving along the signal and the abrupt changes in the signal. High frequency components contain poor frequency information but is localised in time as shown in *Figure 8* [10].

There are two types of wavelet transform – Continuous Wavelet Transform (CWT) and Discrete Wavelet Transform (DWT). The difference is based on how the scaling and translation is being discretized [11].

2.2.1 Continuous Wavelet Transform

CWT is normally used for time frequency analysis and filtering of time localized frequency components [9][12]. Example of CWT are Morse Wavelet, Analytic Morlet Wavelet [15]. It is defined by a mother wavelet function $\Phi(t)$. In Continuous Wavelet Transform, the scaling factor is discretized even more finely by a fractional power of two. [12] When the signals get finer, it also means there are more coefficients. More output coefficients will result in imperfect reconstruction; which is not ideal for image processing.

2.2.2 Discrete Wavelet Transform

DWT is normally used in denoising and compression of signals and images, which is also known as sparse representation. [10][16] Example of DWT is Haar Wavelet. In DWT, it consists of mother wavelet $\Phi(t)$ and father wavelet $\varphi(t)$ [12][16]. Mother wavelet also known as wavelet function is dilated and moving across the signal to construct different feature in the signal. Father wavelet is also known as the scaling function that filters the lowest level of the signal. The subsequent level will be divided by two to ensure all the signal are covered.

Haar Wavelet is used in the proposed solution. Haar wavelet [13][17], one of the simplest WT uses an orthogonal and square function on the intervals of [0,1] to capture the frequency components in every spectrum. The input and the output coefficients are equal, thus, is recommended for image reconstruction.
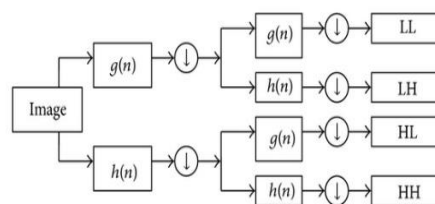
Figure 12 Level 1 Wavelet Decomposition Tree

In DWT, signals are filtered by the low pass and high pass filter denoted as g(n) and h(n) shown in *Figure 12*. The two frequency sub bands are then further decomposed by applying low pass and high pass filter on g(n) and h(n) to produce four different down sampled frequency components. They are the LL (Approximation), LH (Horizontal), HL (Vertical), HH (Diagonal) as shown in *Figure 13*. The LL (low-low) is the only low frequency component and the other three are high frequencies components. Each of the coefficients extract different frequency components from the signals. [10][12][16][17]

| (a) | (b) | (c) | (d) |
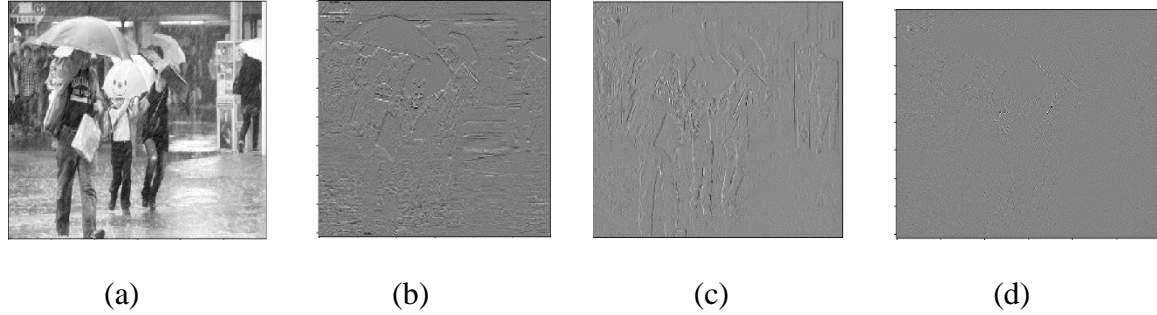
Figure 13: Four coefficients – (a) LL, (b) LH, (c) HL & (d) HH after applied DWT on an image.

Discrete Wavelet Transform is also a multiresolution analysis. Multiresolution analysis can be obtained by further decomposition from the LL as shown in *Figure 14*. Each time the coefficients is decomposed, it will be down sampled by a factor of 2 to extract more features from the signal. However, further decomposition also means the loss of all the frequency components but it would remain localised in time resolution. The more down sampled component will result in great time resolution. However, the objective is to remove rain in images and images do not depend on time resolution since it is a space resolution. Therefore, applying further decomposition will not improve the frequency or the overall performance in the images. [11][18]
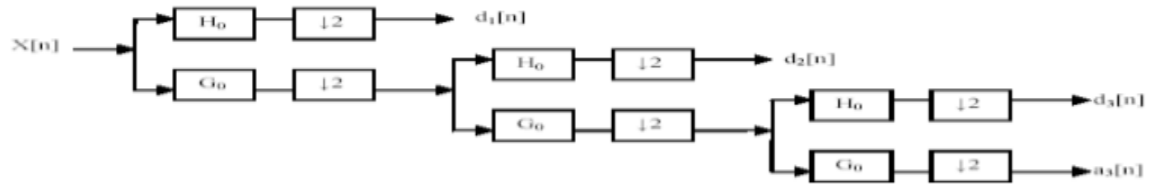
Figure 14: Multiple Level Wavelet Decomposition Tree

2.3 HSV Colour Space

HSV, which stands for Hue, Saturation and Value, separates chromatic and achromatic information into three individual channels which is proven better than RGB in term of human vision of colours. [19] HSV colour model is represented by a hexacone [21][22] as shown in *Figure 15*, where the height is the amount of luminance, the surface area is the saturation and the radius is the amount of hue. HSV model is formed by tilting the RGB cube vertically to ensure that black is at the bottom and white at the top. The other six corners consist of all the additive colour and subtractive colour. Additive colours contain RGB – Red, Green and Blue. Subtractive colours contain CMY – Cyan, Magenta and Yellow. As such, all the colours can be defined in the hexagon projection. [20][21]
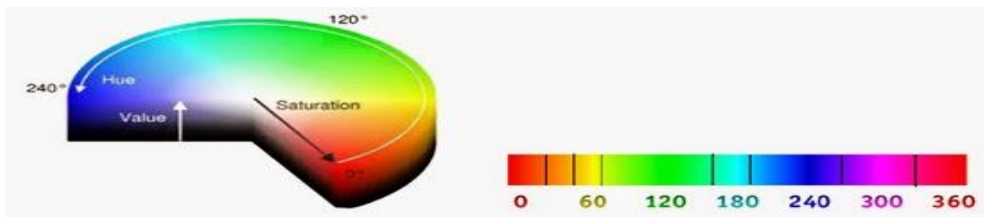
Figure 15   *Figure 16*

Figure 15: Individual location of HSV

Figure 16: Additive and Subtractive colours at respective degree

Hue, which consists of chromatic information [22] can be defined as the quality of a colour which depends on its dominant wavelength. Originally, colours are measured in range from [0 – 255] in RGB colour space. However, in HSV colour space hue is measured in degree [**0°** - **360°**]. It will start with red at **0°**, yellow at **60°**, green at **120°**, cyan at **180°**, blue at **240°**, magenta at **300°** and back to red at **360°** as shown in *Figure 16 & Figure 17*. [21]

Saturation can be defined as the amount of bandwidth of light from the source. The source mainly refers to the hue. Saturation is measured in range from [0 to 1] as shown in *Figure 18*. When the saturation is at its minimum value, assuming the intensity is at its maximum, it will produce white colour. When the saturation value starts to change from 0 to 1, the colour will change from dim to the purest form of the hue accordingly. [20][21][22]

Value, which consists of achromatic information, refers to the brightness of the colour. Value is a measurement of the amount of black and white in the hue which can be measured in the range of [0, 1] as shown in *Figure 19*. In the proposed solution, luminance channel (V) is used in removal of rain as human vision is sensitive to luminance changes. [20][21]

Figure 17: Hue at 180 degree – Cyan
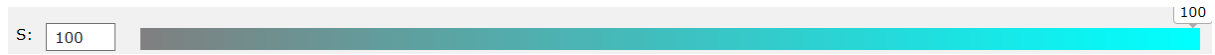
Figure 18: Saturation at its Maximum corresponding to hue

Figure 19: Value at its Maximum corresponding to hue

With all the separated chromatic and achromatic information, the proposed solution can mainly concentrate on the luminance channel while preserving the original hue to avoid any discolouration and distortion.

2.4 Gamma Correction

Gamma Correction is an exponential function that mirrors the non-linearity in human vision. The purpose of the exponential function is to remove the non-linearity [23][24][26].

The formula given for Gamma Correction, which is also known as power law, can be expressed as [26]:

$$I_{enhance} = (\frac{I}{255})^{\frac{1}{\gamma}} ; \text{-----------------------------} Equation\ (2)$$

where I is the input images and γ is the gamma.

Figure 20

Figure 21

Figure 22

Figure 20: Decoding gamma (0.9γ), Figure 21: Original Image (1.0γ), Figure 22: Encoding gamma (1.1γ)

When γ is more than 1, it is called encoding gamma and when γ is less than 1, it is called gamma decoding. Encoding gamma is used to enhance the bits per channel by optimizing the usage of bits as shown in *Figure 22* [25][26]. Gamma Correction has proven that the increment of gamma will result in a brighter image accordingly as shown in *Figure 22*. However, the drawback of Gamma Correction is that it does not have a specific gamma value. Every image has different luminance requirement. Is there a way to find the unique gamma value that can apply to different type of images (luminance)?

2.4.1 Adaptive Gamma Correction (AGC)

Adaptive Gamma Correction (AGC) [27] uses histogram analysis to analyse individual weight distribution to provide a correct gamma value in each image. AGC uses the probability density function (PDF) and cumulative distribution function (CDF) in histogram analysis to determine the weight distribution in each image [27].

As such, AGC has proven the enhancement on the luminance in different images. By using histogram analysis, it improved the weight distribution by averaging out the distribution instead of lumping all the weight at one side.

Reason Gamma Correction is selected instead of AGC

Adaptive Gamma Correction algorithm is not used in the proposed solution. It will contrast the images significantly even though AGC had adjusted the weight distribution. The result is shown in section 4.2.1.

To conclude, gamma correction is used in the proposed solution with multiple experiments tested in Section 4.2.2 & 4.3.2. In section 4.2.2, gamma correction is applied to the luminance channel (V) while in section 4.3.2 gamma correction is applied on the saturation channel to enhance the chromatic information.

2.5 Image Quality Comparison

Image quality comparison is important in image enhancement as it determines the quality of the image. Image quality comparison consists of quantitative comparison [32] and qualitative analysis. In machine learning, quantitative comparison refers to a comparison with the ground truth as shown in *Figure 23* while qualitative analysis does not require the ground truth as shown in *Figure 24*. Quantitative comparison is used in training data set [33] while qualitative analysis can be used in both training and testing data set. Examples of quantitative techniques that can be used are the following: SSIM (Structural Similarity) whilst NIQE (Naturalness Image Quality Evaluator) can be employed for qualitative analysis.

Synthetic Rain

Neural Network#

Rain-free
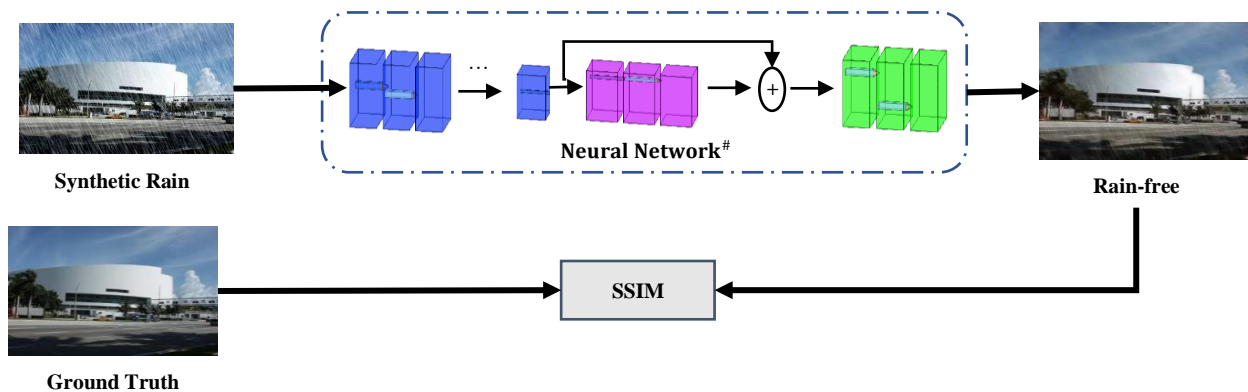
SSIM

Ground Truth

Figure 23: SSIM is used to find qualitative comparison using training datasets

SSIM [29] is a measurement of a ground truth with a processed image (synthetic rain). It measures the similarity between the ground truth and the processed image with a range from 0 to 1, where 1 is the ideal quality measurement. SSIM is used in training dataset (ground truth) and the rain-free images to compare how well rain is removed in Wavelet CycleGAN. The mean of 50 different training dataset SSIM is used to compare between different algorithm [32]
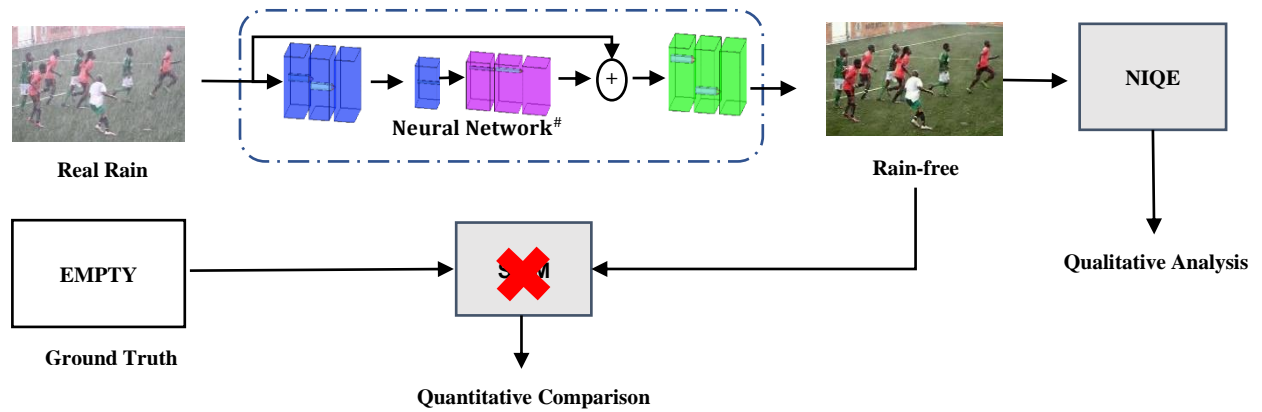

Figure 24: NIQE is used to find qualitative analysis using testing datasets

NIQE [31] is the qualitative analysis which is also known as the 'completely blind' image quality assessment. It does not need any dependency to predict the quality of the image. It is mainly used on the testing images as it does not have any ground truth to be compared with.

NIQE is invented based on space domain natural scene statistic (NSS) model. The NSS model extract features from natural images to capture the low order statistics. Patches in the images will be selected and applied to the NSS model. The patches that are selected are those that are most unlikely to be distorted and are rich in information. It is then used to construct a model of the natural image patches. A formula is formed to find the local sharpness $\delta(b)$ by using the summation of the size of the patches. After finding the local sharpness, it is compared with the threshold (T). Only those local sharpness $\delta(b)$ that is greater than threshold will be selected. [31]

To conclude, the proposed solution will be using SSIM as a quantitative comparison on the training dataset as shown in *Figure 23* and NIQE as a qualitative analysis on the testing dataset as shown in *Figure 24*.

3. Methodology

The proposed solution, Wavelet CycleGAN, is modified from the CycleGAN [34]. There are two sections in CycleGAN as well as Wavelet CycleGAN which are the training and testing. Training Section is used to train a rain-free model using 700 synthetic rain datasets [33] and 700 ground truth as shown in *Figure 25*. Testing Section is to apply the rain-free model on 52 testing real rain images to remove the rain. The target of the proposed solution must be better than the output of the CycleGAN which is also known as the residual rain removal.

(a)   (b)
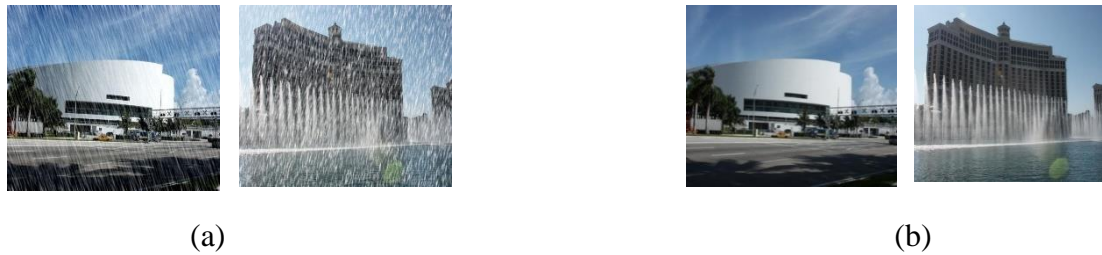
Figure 25: Synthetic Rain and Ground Truth as shown in (a) and (b) respectively.

Wavelet CycleGAN is separated into two sections – Frequency Sub Band and Residual Rain Removal. Frequency Sub Band extract different features from different location in the datasets while residual rain removal removes rain globally.

Frequency Sub Band uses the concept in this article [35] to mainly enhance noisy RGB images by converting it to HSV color space and using Discrete Wavelet Transform on the luminance(V) channel. It uses PWL (Piecewise Linear Filter) to sharpen the high frequency components and smoothen the low frequency components. The new V is obtained by using IDWT (Inverse Discrete Wavelet Transform) on the enhanced high frequency and low frequency and then converted back to RGB.

Instead, for the proposed solution, the Frequency Sub Band section is used to enhance the local frequency in different location. Following the steps in the article [35], in training, the Frequency sub band can be obtained by converting RGB images into HSV color space shown in *Figure 26*. HSV color space is then spilt into three individual channel – Hue, Saturation and Value. DWT is used on the Value channel to obtain all the four coefficients sub bands – LL, LH., HL and HH as shown in *Figure 27*. Depending on the configuration model, the coefficients sub bands will be integrated into Wavelet CycleGAN to remove rain. The enhanced V, obtained by using IDWT on the enhanced coefficients, will be reconstructed with the Saturation and Hue and converted back into the RGB color space. In residual rain removal section, the output of the Frequency sub band section will be integrated into CycleGAN to remove rain globally.

In testing section, additional gamma correction is applied to Saturation and the Value's low frequency (LL) to enhance the quality of the image in the Frequency Sub Band section before integrating it into CycleGAN.
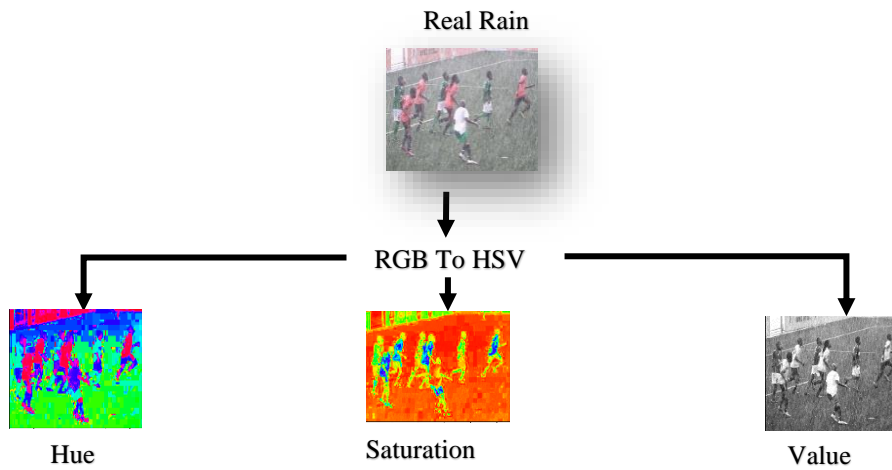
Figure 26: Real Rain (RGB) convert to HSV colour space
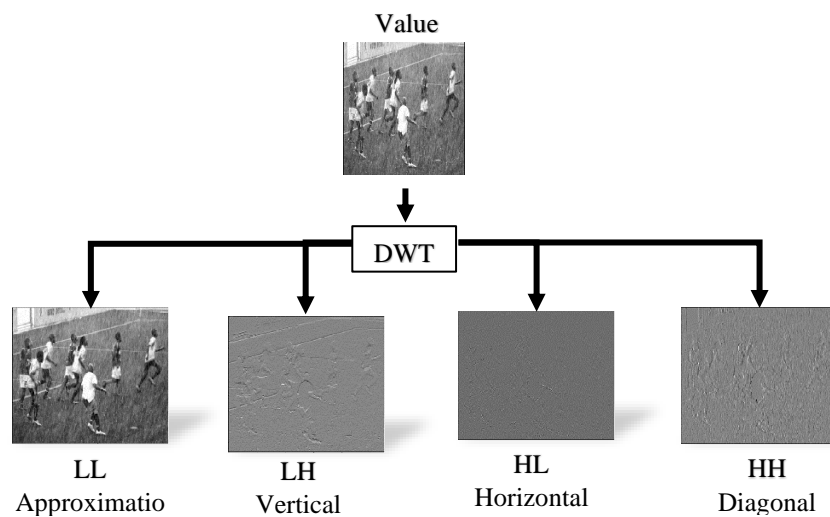
Figure 27: Four frequency sub bands are obtained by using DWT on the Value Channel

4. Experimental Results

Experimental results consist of four sections. The first section is to experiment different combinations and different level of decompositions between the coefficients sub bands – LL, LH, HL and HH from the luminance (V) channel. The second section is to experiment different gamma values on the luminance (V) low frequency component (LL). The third section is to further enhance Wavelet CycleGAN - the combination of Frequency Sub Band and Residual Rain Removal by applying gamma on the Saturation(S) channel and adjusting different parameters such as the number of datasets. The last section presents the flowchart of the finalized configuration of Wavelet CycleGAN. The NIQE is used as an image quality metrics. The lower the NIQE, the better the quality of the image. A detailed flowchart of each configuration requirements is drawn in **_Annex_**. All experiments are done using NVIDIA RTX 2060 and 16GB RAM shown in **_Annex F_**.

4.1 Selecting Correct Configuration for Frequency Sub Band

4.1.1 Configuration 1 (#S1_V14NTR & #S1_V14NTST)

In configuration 1, #S1_V14NTR and #S1_V14NTST networks as found in *Annex B1 & Annex C1* respectively are designed to integrate all the four coefficients – LL, LH, HL and HH in luminance (V) channel into the neural network for training and testing as reference from - Deep joint rain and haze removal from single images [36].

The article [36] separated rain removal and haze removal into two sections. For rain removal, DWT is applied to each of the RGB channels – (3 channel x 4 coefficients) 12 channels. These 12 coefficients will be integrated into the neural network so as to remove rain. For haze removal, the RGB images is converted into grayscale to integrate it into the neural network to remove the haze.

Based on the article, integrating all the 4 coefficients into the neural network can result in rain removal which is applied to configuration 1.

A detailed explanation of the numerical model networks can be found in **_Annex A_**.

Table 1: Result of #S1_V14NTST configuration

Examples of Rain Images	Output of # S1_V14NTST
(a)	(d)

(b)	(e)
(c)	(f)

Table 2: Qualitative analysis (NIQE) in Table 1

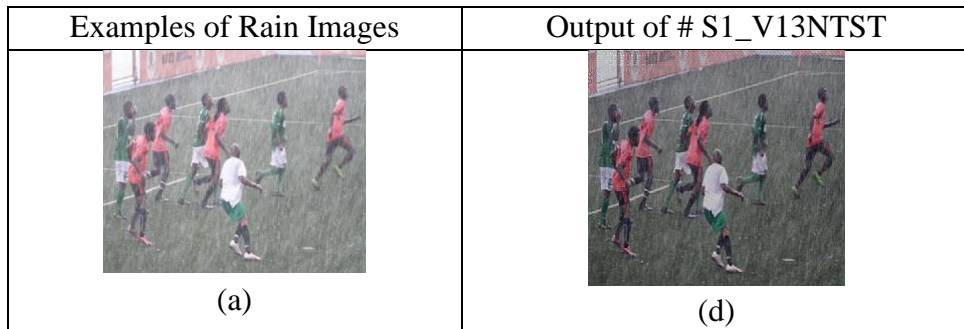|	(a)	(b)	(c)	(d)	(e)	(f)
NIQE	7.3045	5.8153	4.8687	6.3472	6.0850	5.6575

From the result, the output # S1_V14NTST produced an overly contrasted image due to the low frequency component (LL) in the luminance channel. In general, the low frequency component (LL) containing the property of each spectral band from the respective channel is not recommended in the neural network. This statement only applies to HSV colour space. The individual weight in HSV colour space is not the same as in RGB colour space. For luminance(V) channel, the LL varies its brightness according to the number of epochs. As such, when the epochs reached 200, the excess brightness will worsen the LL coefficient. Hence, resulting in a contrasted image.

4.1.2 Configuration 2 (#S1_V13NTR & #S1_V13NTST)

Configuration 2, #S1_V13NTR & #S1_V13NTST, is revised to improve on the luminance(V) in the testing datasets in which LL component is removed from model network for training and testing as found in *Annex B2 & Annex C2* respectively. It can be reference from - Automatic Single-Image-Based Rain Streaks Removal via Image Decomposition [37]. This article mainly integrated only the high frequency components into neural network to remove the rain. High frequency components extract detail feature such as the rain streak in the image. As such the neural network will learn the detail feature and remove the rain.

Based on the article, integrating only the high frequency components into the neural network can result in rain removal which is applied to configuration 2.

Table 3: Result of #S1_V13NTST configuration

Examples of Rain Images	Output of # S1_V13NTST
(a)	(d)

| | (b) | | (e) |
| | (c) | | (f) |
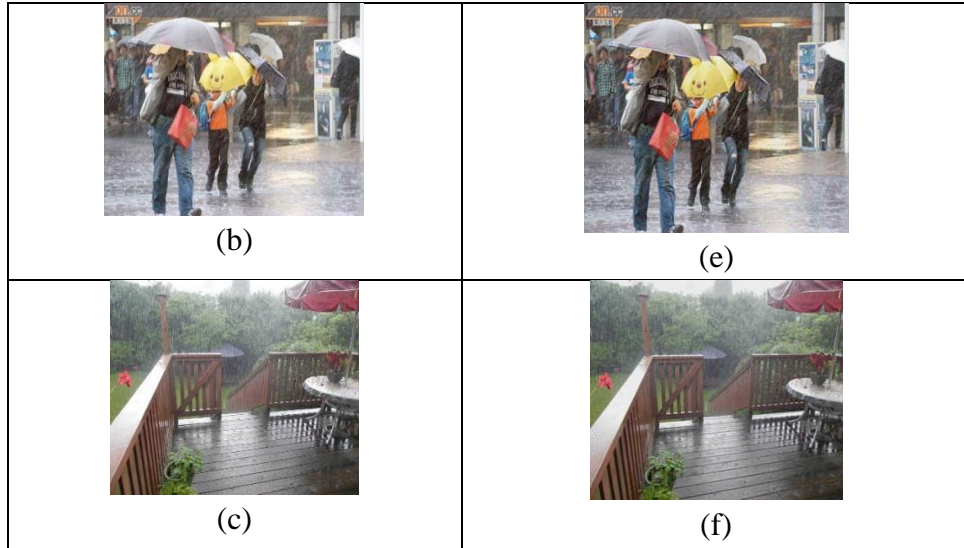
Table 4: Qualitative analysis (NIQE) in Table 3

	(a)	(b)	(c)	(d)	(e)	(f)
NIQE	7.3045	5.8153	4.8687	4.9750	4.0052	3.2592

The result shown looks more realistic and the NIQE is much lower than the original raining images. The only downside of this configuration is that it dimmed the images. This is because when the Generators in the frequency sub bands try to remove the rain, it also degraded the achromatic information (luminance). Multi-decomposition will then be used on the next configuration to improve the achromatic information.

4.1.3 Configuration 3 (#S1_V23NTR & #S1_V23NTST)

In configuration 3, two level decomposition, #S1_V23NTR & #S1_V23NTST networks as found in *Annex B3 & Annex C5* respectively, are designed to integrate the high frequency components in luminance (V) channel from each level into neural network for training and testing as reference from - Image Reconstruction Using Discrete Wavelet Transform [18]. This article has proven that Wavelet multi-decomposition can preserve the signal even when the image is being compressed unlike Fourier Analysis. As such, it is used to compressed large images file. Also, Wavelet multi-decomposition have good time resolution which can be used to enhance video motion. In theory, the more level of decomposition occurs, the more frequency information within the image is lost. However, what will happen if it is applied to a space domain such as an image?

Based on the article, integrating all the multiple level of decomposition into the neural network can result in enhancing the time resolution and perfect reconstruction which is applied to configuration 3.

Table 5: Result of #S1_V23NTST configuration

Examples of Rain Images	Output of # S1_V23NTST

Table 6: Qualitative analysis (NIQE) in Table 7

	(a)	(b)	(c)	(d)	(e)	(f)
NIQE	7.3045	5.8153	4.8687	5.3661	4.1640	5.5896

The result in #S1_V23NTST is unsatisfactory. It does not enhance the achromatic information but instead causes the images to become dimmer and more blur than #S1_V13NTST. This is mainly the cause of multiple high frequency components in #S1_V13NTST. It has doubled high frequency components than in #S1_V23NTST resulting in dimmer output. Using level two decomposition will result in better time resolution but loss in frequency information. However, images are in space domain and great time resolution does not affect any enhancement. Instead, it has lost some frequency information in the second layer (Value 2 up sampled*) when it tried to reconstruct with the first layer (Value 1*) which results in blurry images.

 (*Please refer to the *Annex C5*).
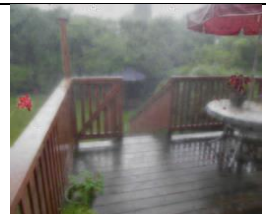
4.1.4 Configuration 4 (#S1_V33NTR & #S1_V33NTST)

In configuration 4, three level decomposition, #S1_V33NTR & #S1_V33NTST networks as found in *Annex B4 & Annex C6* respectively, are used to integrate the high frequency components in luminance (V) channel from each level into the neural network for training and testing as reference - Lightweight Pyramid Networks for Image Deraining [38]. This article has proven success using multiple scaling to extract various features from the high frequency components which is also known as the Laplacian coefficients to remove rain.

Based on the article, integrating all the multiple level of decomposition into the neural network can result in rain removal which is applied to configuration 3.

Table 7: Result of #S1_V33NTST configuration

Examples of Rain Images	Output of # S1_V33NTST
 (a)	 (d)
 (b)	 (e)
 (c)	 (f)

Table 8: Qualitative analysis (NIQE) in Table 7

	(a)	(b)	(c)	(d)	(e)	(f)
NIQE	7.3045	5.8153	4.8687	5.0921	4.3355	4.4887

The result in #S1_V33NTST is also unsatisfactory. It is more blurred and has discolouration compared to #S1_V23NTST. The concept in [38] does not apply to wavelet transform. As mention above, the more level of decomposition occurs, the more frequency information within the image is lost.

To conclude, enhancing time resolution in space domain (#S1_V23NTST & #S1_V33NTST) will distort the images in certain ways. Therefore, #S1_V13NTST is more ideal for space domain as it preserved frequency information.

4.2 Apply Gamma Correction in (V) channel

4.2.1 Configuration 5 (#S1_V13NTR & #S1 V13ATST)

In configuration 5, #S1_V13NTR network as found in *Annex B2*, are used to integrate only the high frequency components in luminance (V) channel into the neural network for training. For testing, #S1_V13ATST network as found in *Annex C4*, are used to integrate high frequency components into a rain-free model and reconstruct it with a contrast enhanced low frequency component (LL) in luminance channel as reference [39] to obtain a new V channel. The new V channel is then reconstructed with the original S and H and converted back to RGB.

The contrast enhanced low frequency (V) component can be obtained by applying Adaptive Gamma Correction [27] on the low frequency component in the V channel to balance the weight distribution. The gamma value can be determined by the CDF (Cumulative Distribution Function) which is found from the individual weight distribution.

Based on the article, applying Adaptive Gamma Correction on the low frequency can result in contrast enhanced image which is applied to configuration 5.

Table 9: Result of #S1_V13ATST Configuration

Type of Configurations	Output of #S1_V13ATST in the testing images		
Real Rain images	(a)	(b)	(c)
#FSB V112 TST	(d)	(e)	(f)

Table 10: Qualitative analysis (NIQE) in Table 9

	(a)	(b)	(c)	(d)	(e)	(f)
NIQE	7.3045	5.8153	4.8687	8.2026	4.7150	4.7066

The images are indeed contrast enhanced. However, some parts of the images are still overly contrasted. For example, for the soccer image in (d) the whole image is badly contrasted while the yellow umbrella image in (e) looks satisfactory. The algorithm in [39] consists of constant high frequency components. However, in the proposed solution, the high frequency components dimmed the luminance. This may be the reason that causes the unbalance weight distribution.

4.2.2 Configuration 6 (#S1_V13NTR & #S1_V13G-γTST)

In configuration 6, #S1_V13G-γTST network as found in *Annex C3*, is revised to improve the luminance in LL component in (V) channel by removing the histogram [28] from Adaptive Gamma correction in the testing section. By removing the histogram, the gamma value will be unknown. Therefore, a range of gamma values will be used on this configuration. The range of gamma values can be estimated from [26] where encoding gamma is more than 1 and decoding gamma is less than 1.

Also, this article [28] enhanced the image quality by optimizing local contrast using only gamma correction.

Based on the article, local contrast can be enhanced by applying gamma correction. Therefore, in this configuration, a range from 1.1 to 1.4 gamma values will be applied to the low frequency components in V channel.

Table 11: Result of #S1_V13G-γTST Configuration

Different gamma value	Output of respective Gamma Correction in the testing images		
Real Rain images	(a)	(b)	(c)
1.1	(d)	(e)	(f)

1.2	 (g)	 (h)	 (i)
1.3	 (j)	 (k)	 (l)
1.4	 (m)	 (n)	 (o)

Table 12: Qualitative analysis (NIQE) in Table 11

	(a)	(b)	(c)	(d)	(e)	(f)
	7.3045	5.8153	4.8687	6.6272	5.1717	5.1781
	(g)	(h)	(i)	(j)	(k)	(l)
NIQE	4.1108	4.6633	4.2122	4.4161	4.5591	4.2575
	(m)	(n)	(o)			
	5.2493	4.7183	4.7662			

The results look satisfactory. Comparing #S1 V13NTST and #S1 V13G-γTST models, gamma correction has proven an enhancement on the luminance in the image. Looking at the Table 11, #S1 V13G-1.2TST model, 1.2 gamma value applied to low frequency in V channel, resulted in the lowest NIQE value. Therefore, #S1 V13G-1.2TST model, is selected.

In conclusion, #S1 V13G-1.2TST model is selected on the first stage of Wavelet CycleGAN which is also known as the Frequency Sub band.

4.3 Enhancement on Wavelet CycleGAN

In this section, CycleGAN (Stage 2) will be added into the Frequency Sub band (Stage 1) to remove rain globally. The combination of the two stages is referred to as Wavelet CycleGAN.

4.3.1 Configuration 7 (#S2_V13NTR & #S2_V13G-1.2TST)

In configuration 7, #S2_V13NTR & #S2_V13G-γTST networks as found in *Annex D1 & Annex E1*, are used to integrate the output of the Frequency Sub Band into CycleGAN for training and testing respectively.

Table 13: Result of #S2_V13G-1.2TST Configuration

Type of Configurations	Output of #S2 V13G-1.2TST in the testing images		
Real Rain images	(a)	(b)	(c)
#S2_V13G-1.2TST	(d)	(e)	(f)

Table 14: Qualitative analysis (NIQE) in Table 13

	(a)	(b)	(c)	(d)	(e)	(f)
NIQE	7.3045	5.8153	4.8687	3.9483	3.6271	3.4942

Wavelet CycleGAN, #S2_V13G-1.2TST, produced rain-free images shown in Table 13. However, there is still minor discoloration in the image and the NIQE values shown in Table 14 is not ideal. In the next section, gamma correction will be applied to the saturation (S) channel for further enhancement.

4.3.2 Configuration 8 ((#S2_V13NTR & #S2_V13G-γ S-γ TST)

In Configuration 8, #S2_V13G-γ Sγ TST model network as found in *Annex E2*, was revised to achieve further enhancement by applying gamma correction on the saturation (S) channel as referenced from [35]. This article used histogram equalization on the S channel and reconstructed with enhanced V and original H and converted it back to RGB. As mentioned above, applying histogram equalization in the proposed solution will distort the images. *Figure 28* shown distorted images after applying the histogram equalization on the S channel. Therefore, gamma correction is used instead on the saturation (S) channel.

| (a) | (b) | (c) |

Figure 28: Distorted images after applying histogram equalization on the S channel

The reason why applying histogram equalization in the proposed solution will distort the images is mainly due to the Wavelet CycleGAN. Wavelet CycleGAN is made up of complicated neural networks that not all proven algorithms can be successfully incorporated. There are too many factors that affect the output. For example, the high frequency will dim the images after training. Adaptive Gamma Correction will enhance the luminance by calculating its weight distribution but when incorporated into Wavelet CycleGAN, it does not work out. With such a limited time, gamma correction is used to further enhance the quality of the image. A range of 1.03 – 1.1 gamma value is applied to saturation (s) channel.

Table 15: Result of #S2 V13G-1.2 Sγ TST Configuration

Different Gamma Value for S	Output of respective in the testing images		
0	(a)	(b)	(c)
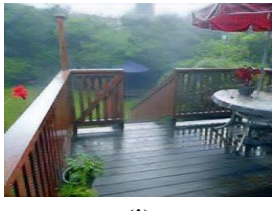
1.03	 (d)	 (e)	 (f)
1.04	 (g)	 (h)	 (i)
1.05	 (j)	 (k)	 (l)
1.1	 (m)	 (n)	 (o)

Table 16: Qualitative analysis (NIQE) in Table 15

	(a)	(b)	(c)	(d)	(e)	(f)
	3.9483	3.6271	3.4942	3.7443	3.3489	3.1726
NIQE	(g)	(h)	(i)	(j)	(k)	(l)
	3.6974	3.3447	3.4467	3.5109	3.3149	3.3134
	(m)	(n)	(o)			
	3.5039	3.2757	3.7687			

The additional enhancement on the saturation channel indeed improved the quality of the image. Table 15 above show that those images with additional enhancement on the saturation channel resulted in a better NIQE value. #S2_V13G-1.2 S1.05 TST model network is selected as it has a good NIQE value and preserved the colour well. Even though #S2_V13G-1.2 S1.1 TST model network has the best NIQE value, the images do not look realistic.

In conclusion, #S2_V13G-1.2 S1.05 TST model network is used in Wavelet CycleGAN as it can preserve colours and remove rain efficiently.

4.3.3 Different Parameters

In machine learning, different parameters also affect the outcome. The different parameters usually refer to the number of datasets, changes in learning rate, number of epochs and even the batch size. However, the proposed solution will only be concentrating on the number of datasets and changes in learning rate.

Number of datasets

Initially, 700 datasets [33] are used for training a model in Wavelet CycleGAN. Many articles [36][38] utilized a few thousand datasets to train a model. Theoretically, more datasets will result in a better prediction. Therefore, 4000 datasets [8] are used for training using the latest #S2_V13G-1.2 S1.05 TST model network.

Table 17: Result of using 700 and 4000 datasets using #S2_V13G-1.2 S1.05 TST

Number of datasets	Output of different amount of training datasets in the testing images		
700 datasets	(a)	(b)	(c)
4000 datasets	(d)	(e)	(f)

Table 18: Qualitative analysis (NIQE) in Table 17

	(a)	(b)	(c)	(d)	(e)	(f)
NIQE	3.5109	3.3149	3.3134	2.8680	3.2996	2.8656

The output of 4000 datasets is superior. The images are smooth and the NIQE value is low. However, these images have defects as shown in Table 19. In fact, previously all the images

have defect. Those images that were trained using 700 datasets presented with defects that were fuzzy and not noticeable.

Table 19: Defect in 4k datasets using #S2_V13G-1.2 S1.05 TST

No.	Result in 4000 datasets (Before)	Zoom-in defect
1		
2		
3		

In order to minimise the defect, different learning rates were tested. The cause of the defect is mainly from the Frequency Sub Band section because the existing program has been modified. One way is to adjust the learning rate so as to slow down the predication of the rain streak. As a result, it will reduce the chances of having a defect. However, how much should the learning rate be reduced by? The default learning rate is $2e^{-4}$. $1e^{-4}$ & $1.5e^{-4}$ will be used for testing.

Changes of learning rate

Wavelet CycleGAN consists of two stages – Frequency Sub Band and Residual Rain Removal. Each stage consists of its own individual learning rate. The Table below shown different learning rate from each stage. As such, for convenience purposes, Frequency Sub Band is referred to as Stage 1 and Residual Rain Removal as Stage 2.

Table 20: Result of different learning rate in Frequency Sub Band & Residual Rain Removal

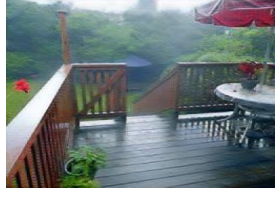|	
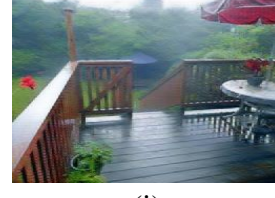
Learning Rate	Output of respective learning rate in the testing images		
Stage 1 :$2e^{-4}$ Stage 2: $2e^{-4}$	(a)	(b)	(c)
Stage 1 :$1e^{-4}$ Stage 2: $2e^{-4}$	(d)	(e)	(f)
Stage 1 :$1.5e^{-4}$ Stage 2: $2e^{-4}$	(g)	(h)	(i)
Stage 1 :$1e^{-4}$ Stage 2: $1e^{-4}$	(j)	(k)	(l)

Table 21: Qualitative analysis (NIQE) in Table 20

NIQE	(a)	(b)	(c)	(d)	(e)	(f)
	3.5109	3.3149	3.3134	2.7537	3.0171	2.8223
	(g)	(h)	(i)	(j)	(k)	(l)
	3.5039	3.2757	2.8882	4.3657	3.5757	3.7535

The stage 1 $1e^{-4}$ & stage 2 $2e^{-4}$ shown a superior image. The NIQE value shown below in Table 21 (d), (e) and (f) is even lower than the one trained using 4000 datasets and it does not have any defect. Stage 2 learning rate must stay at its default $2e^{-4}$. Any reduced learning rate in Stage 2 will not remove the rain efficiently as shown in Table 21 (j), (k) & (l).

Table 22: Result of using 700 datasets with learning rate of Stage 1: $1.5e^{-4}$ and Stage 2: $2e^{-4}$ shown no defect

No.	Result in 700 datasets (After)	Zoom-in defect
1		
2		
3		

As such, the proposed solution #S2_V13G-1.2 S1.05 TST model network uses the parameters of 700 datasets, 200 epochs, 1 batch size and learning rate for Frequency Sub Band is $1e^{-4}$ and Residual Rain Removal is $2e^{-4}$.

4.4. Wavelet CycleGAN

4.4.1 Neural Network (Training)
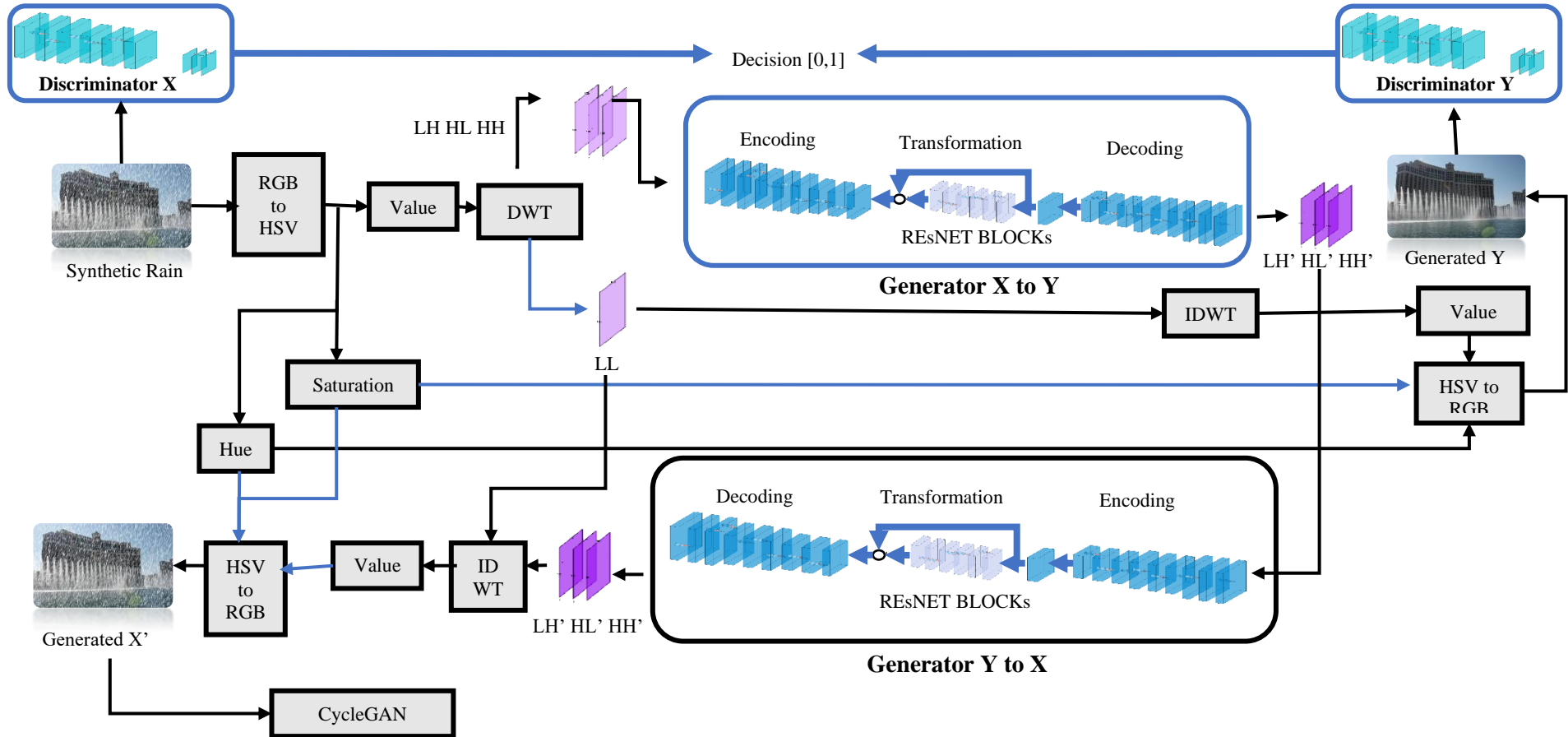

Figure 29: Wavelet CycleGAN Training Architecture
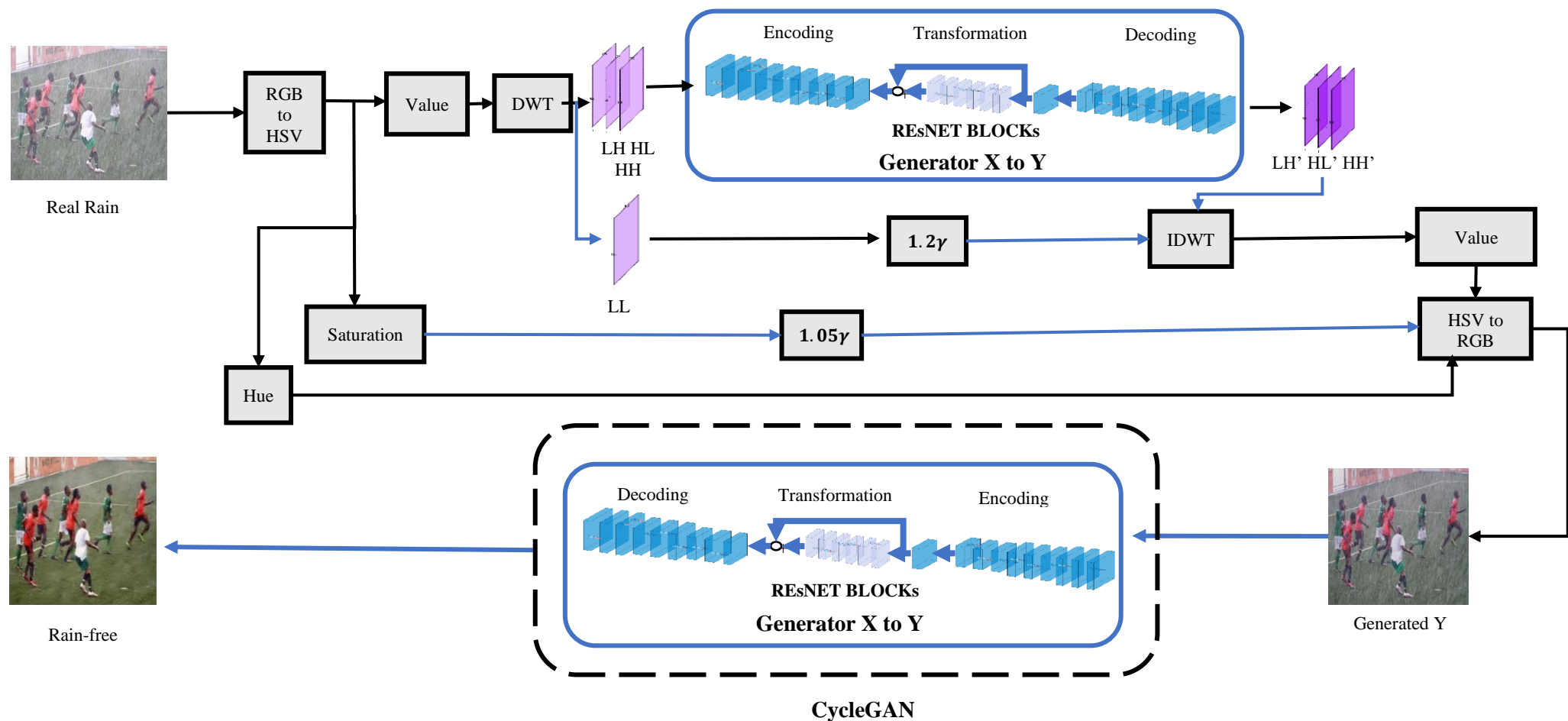
4.4.2 Neural Network (Testing)

Figure 30: Wavelet CycleGAN Testing Architecture
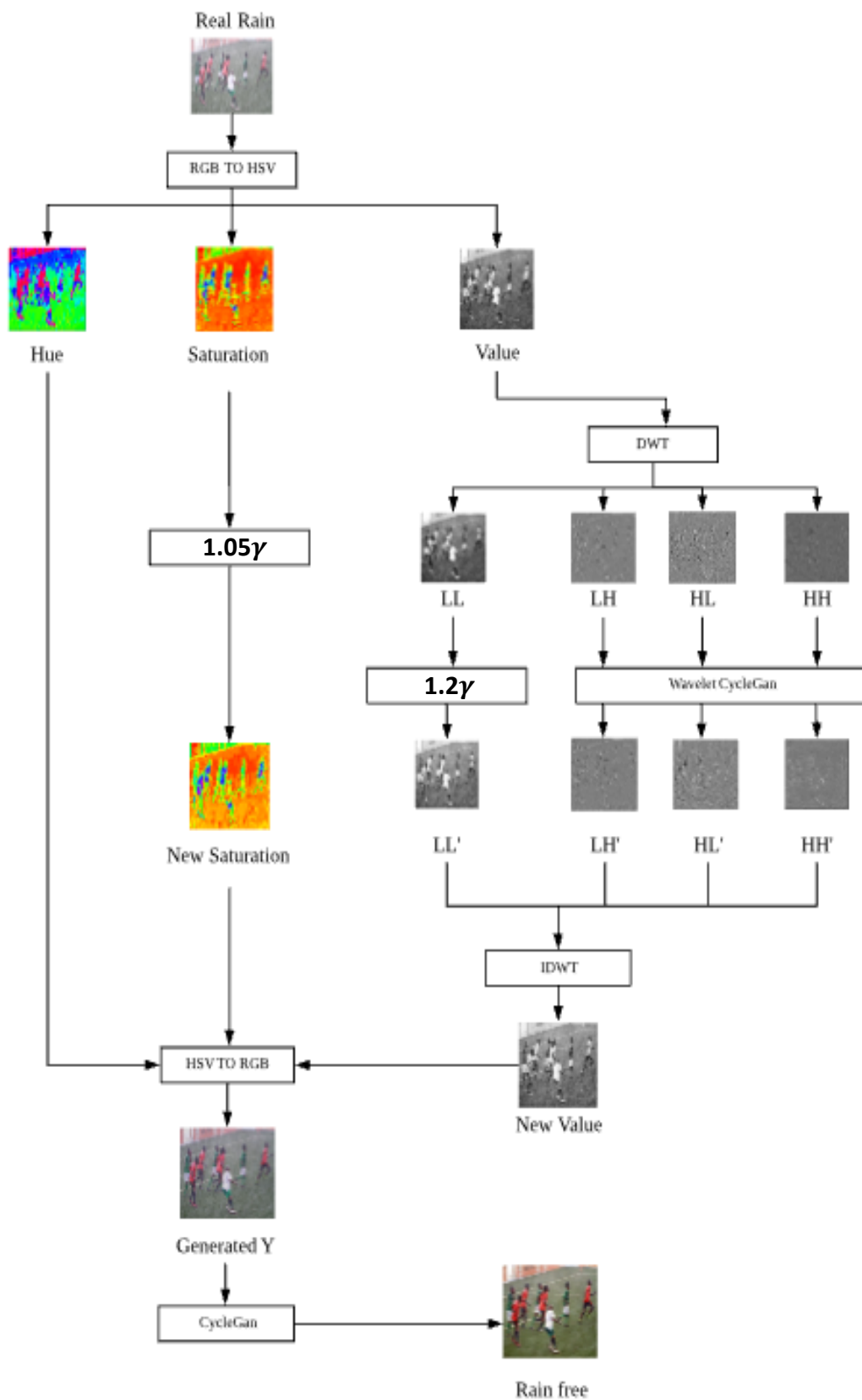
4.4.3 Flow Chart
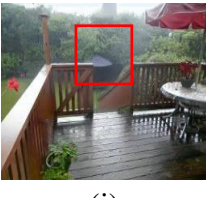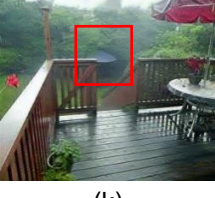
Figure 31: A detail flowchart of Wavelet CycleGAN

5 Image Comparison

5.1 Qualitative Analysis

Table 23: Result of different algorithms to remove rain

Different Algorithm to remove rain	Real Rain	ID-CGAN	CycleGAN	Wavelet CycleGAN
Output of different algorithms in rain removal	(a)	(b)	(c)	(d)
Zoom in rain streak				
Output of different algorithms in rain removal	(e)	(f)	(g)	(h)
Zoom in rain streak				
Output of different algorithms in rain removal	(i)	(j)	(k)	(l)

Zoom in rain streak				
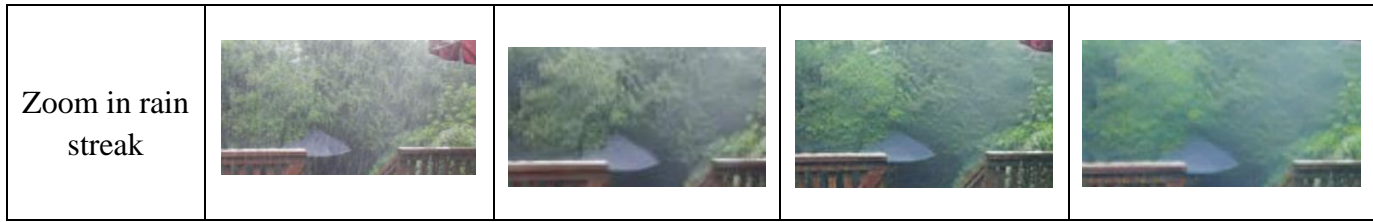
Table 24: Qualitative analysis in Table 23

	Real Rain	ID-CGAN	CycleGAN	Wavelet CycleGAN
NIQE	(a) - 7.3045	(b) - 3.7611	(c) - 3.0532	(d) - 2.7537
	(e)- 5.8153	(f) - 3.1650	(g) - 3.4610	(h) - 3.0171
	(i)- 4.8687	(j) - 3.1297	(k)- 3.2674	(l)- 2.8223

The Wavelet CycleGAN algorithm has completed its objective by preserving the original colors in the image remove rain efficiently. The proposed solution when compared with other existing solutions resulted in a better qualitative numerical value.

5.2 Quantitative Comparison

Table 25: Result of Image Quality Comparison using mean of SSIM and NIQE

	ID CGAN	CycleGAN	Wavelet CycleGAN
Quantitative Comparison (SSIM)	0.81333*	0.8980	0.908
Qualitative Analysis (NIQE)	4.01884*	3.95859	3.71742

*reference from [32]

In Table 25, 50 training datasets was used to find the mean of the SSIM for quantitative comparison and 52 testing datasets was used to find the mean of the NIQE for qualitative analysis. Wavelet CycleGAN shown an outstanding 0.908 quantitative values and 3.71742 qualitative value compared to other algorithms.

6. Discussion

The current learning rate for Frequency Sub Band and Residual Rain Removal is not ideal. The hardest part of this project is to determine the cause of the defect. In neural networks, there are too many factors that can cause the defect such as incorrect learning rate, filter window size, the number of REsNET BLOCKs in the Generator and maybe irregular rain streak in training datasets. Also, there are limited resources available regarding the defect. Therefore, for future improvement, more experiments on different parameters are needed.

Finding the correct gamma value for each image can be used for future improvement. The purpose of using gamma (V) is to enhance the luminance in LL to balance the dimmed high frequency components after training. If there is a way to find the loss in luminance in high frequency components and applied the losses into LL, then it will be much more reliable than estimating a gamma value.

7. Conclusion

Implementing a novel algorithm is never easy. Fixing unexpected errors and problems can easily occupy 2/3 of the time. Nevertheless, the proposed solution still greatly enhanced the image and achieved all the objective.

The proposed solution, compared to other algorithms, has proven to be superior in both rain removal and preservation of colors of the images. It has shown better result by using 700 datasets with 200 epochs, 0.001 learning rate for Frequency Sub Band and 0.002 learning rate for Residual Removal Rain. This combination reduced the defect significantly.

In addition, the Frequency Sub Band has solved the discoloration issue in CycleGAN by extracting local directional frequency. High frequency components in the V channel removed rain locally. Gamma Correction enhanced the chromatic and achromatic information in the S and LL in V channel respectively. The output of the Frequency Sub Band is integrated into CycleGAN to remove rain globally. The mean of SSIM and NIQE has been calculated for quantitative comparison and qualitative analysis respectively.

References

[1] Junbo Zhao, Michael Mathieu and Yann Le Cun. ENERGY-BASED GENERATIVE ADVERSARIAL NETWORKS. ICLR 2017

[2] Jun-Yan Zhu, Taesung Park, Phillip Isola and Alexei A. Efros. Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks. Nov 2018 2

[3] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. Image-to-image translation with conditional adversarial networks. In CVPR, 2017

[4] Saining Xie and Zhouwen Tu. Holistically-nested edge detection. In ICCV, 2015

[5] Marius Cordts1, Mohamed Omran, Sebastian Ramos, Timo Rehfeld1, Markus Enzweiler, Uwe Franke1, Stefan Roth2, and Bernt Schiele. The cityscapes dataset for semantic urban scene understanding. In CVPR, 2016

[6] IJ. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. 2014

[7] Fei Wu. Overview of CycleGAN architecture and training.

[8] https://github.com/hezhangsprinter/DID-MDN

[9] Ali N. Akansu. Richard A. Haddad. Multiresolution Signal Decomposition (Second Edition), 2001

[10] Nasser Kehtarnavaz. Digital Signal Processing System Design (Second Edition), 2008

[11] Mallat, Stephane. "A wavelet tour of signal processing. 1998

[12] Olivier Rioul and Pierre Duhamel. Fast Algorithms for Discrete and Continuous Wavelet Transforms. IEEE March 1992

[13] GILBERT STRANG. Wavelet Transforms Versus Fourier Transforms. April 1993

[14] Martin Vetterli, Jelena Kovaˇceviˊ. Wavelets and Sub band Coding. 1995

[15] Richard Harang, Guillaume Bonne. WAVOS: a MATLAB toolkit for wavelet analysis and visualization of oscillatory systems. BMC Research Notes 26 March 2012.

[16] Resende, J.W., Chaves, M.L.R., Penna, C. Identification of power quality disturbances using the MATLAB wavelet transform toolbox. 2001.

[17] Piotr Porwik, Agnieszka Lisowska. The Haar–Wavelet Transform in Digital Image Processing: Its Status and Achievements. Nov 2014.

[18] G.Shruthi1, Radha Krishna A.N. Image Reconstruction Using Discrete Wavelet Transform. IOSR Journal May – June 2013

[19] Pascale, Danny. A Review of RGB color spaces...from xyY to R'G'B". Retrieved 2008-01-21.

[20] Shamik Sural, Gang Qian and Sakti Pramanik. Segmentation and Histogram Generation using the HSV Colour Space for Image Retrieval. IEEE 2002

[21] George H. Joblove and Donald Greenber. COLOR SPACES FOR COMPUTER GRAPHICS. Aug 1978

[22] Allan Hanbury. Constructing Cylindrical Coordinate Colour Spaces. 2008

[23] Alvy Ray Smith. Gamma Correction. Technical Memo 9 1 Sep 1995

[24] Hany Farid. Blind Inverse Gamma Correction. IEEE 2001

[25] David R. Bull. 4.4 Gamma correction. 2014

[26] Adrian Rosebrock, Open CV Gamma Correction. Oct 2015.

[27] Yi-Sheng Chiu, Fan-Chieh Cheng, and Shih-Chia Huang. Efficient Contrast Enhancement Using Adaptive Gamma Correction and Cumulative Intensity Distribution. IEEE 2011.

[28] Alessandro CAPRA, Alfio CASTORINA, Silvia CORCHS. Dynamic Range Optimization by Local Contrast Correction and Histogram Image Analysis. IEEE 2006.

[29] Zhou Wang, Alan C. Bovik, Hamid R. Sheikh and Eero P. Simoncelli. Image Quality Assessment: From Error Visibility to Structural Similarity. IEEE 4 April 2004

[30] Peter Ndajah, Hisakazu Kikuchi, Masahiro Yukawa, Hidenori Watanabe and Shogo Muramatsu. An Investigation on The Quality of Denoised Images. 2011

[31] Anish Mittal, Rajiv Soundararajan and Alan C. Bovik. Making a 'Completely Blind' Image Quality Analyzer. IEEE 2012

[32] Lai Meng Tang, Li Hong Lim and Paul Siebert. Removal of Visual Disruption Caused by Rain using Cycle-Consistent Generative Adversarial Networks. 2018

[33] He Zhang, Member, Vishwanath Sindagi and Vishal M. Patel. Image De-raining Using a Conditional Generative Adversarial Network. June 2019

[34] https://github.com/cy-xu/simple_CycleGAN

[35] G. Saravanan, G.Yamuna & R. Vivek. A Color Image Enhancement based on Discrete Wavelet Transform. NCETICT 2013

[36] Liang Shen, Zihan Yue, Quan Chen, Fan Feng and Jie Ma. Deep joint rain and haze removal from single images. 21 Jan 2018

[37] Li-Wei Kang, Chia-Wen Lin, Yu-Hsiang Fu. Automatic Single-Image-Based Rain Streaks Removal via Image Decomposition. April 2012.

[38] Xueyang Fu, Borong Liang, Yue Huang, Xinghao Ding and John Paisley. Lightweight Pyramid Networks for Image Deraining. 16 May 2018.

[39] G.Padma Priya, T.Venkateswarlu. An optimal Gamma Correction Based Image Contrast Enhancement Using DWT-SVD. Aug 2018.

Annex A (Explanation of Model Network)

Example:

> #S1 V13GγTST

Table 26: Explanation of Model Network

Refer to the stages	Refer to the channel being used in modification.	No. of decomposition level	No. of coefficients used after applied DWT	Different enhancement methods	Gamma Value	Training (TR) Or Testing (TST)
S1	V	1	3	G	γ	TST

Refer to the stages.
S1 = Stage 1 (Frequency Sub Band), S2 = Stage 2* (Wavelet CycleGAN)

*Stage 2 consists of Frequency Sub Band and Residual Rain Removal

Refer to the channel being used in modification.
H = Hue, S=Saturation, V= Value.

Number of decomposition level
1 = Level 1 Decomposition
2 = Level 2 Decomposition
3 = Level 3 Decomposition

Number of coefficients used after applied DWT (Discrete Wavelet Transform)
4 = All 4 coefficients. [LL, LH, HL and HH]
3 = Only high frequency coefficients are used [LH, HL and HH]

Different enhancement methods
N = No enhancement
G = Gamma Correction
A =Adaptive Gamma Correction

Gamma value = How much gamma is applied to the channel/coefficients

TR = Training, TST = Testing

Therefore, #S1 V13 G-1.4**TST** means:

TeSTing Section - In Stage 1(Frequency Sub Band), V channel applied one level decomposition to obtain four coefficients – LL, LH, HL and HH. 3 high frequency coefficients are integrated into the network. 1.4 Gamma value is used to enhance the LL coefficient.

Annex A1

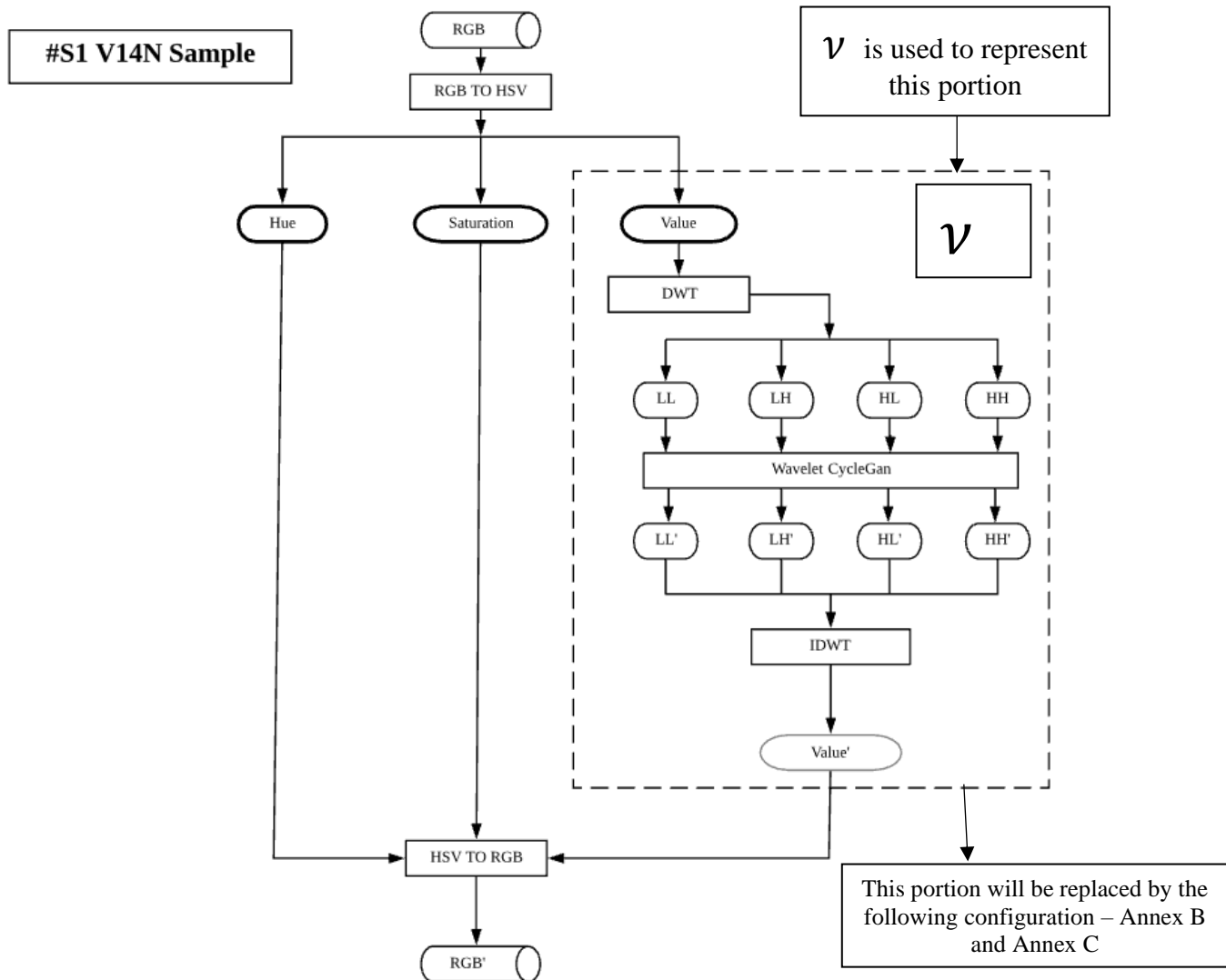#S1 V14N Sample

ν is used to represent this portion

ν

This portion will be replaced by the following configuration – Annex B and Annex C

Figure 32: A flowchart adapted from G. Saravanan [35]
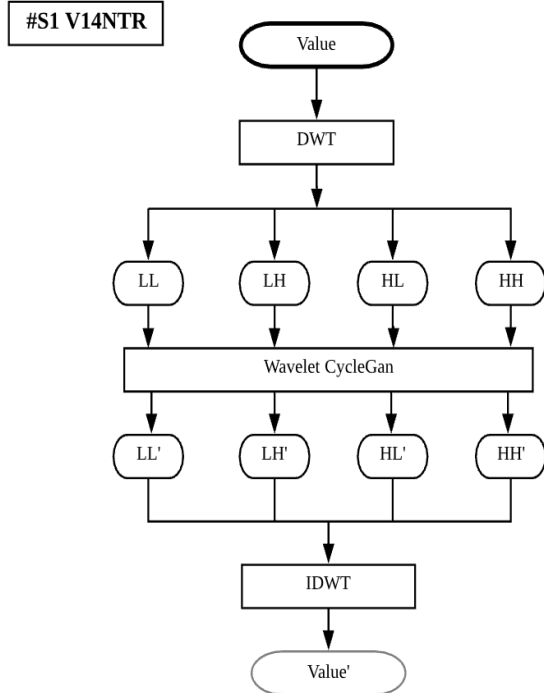
Annex B1:

#S1 V14NTR

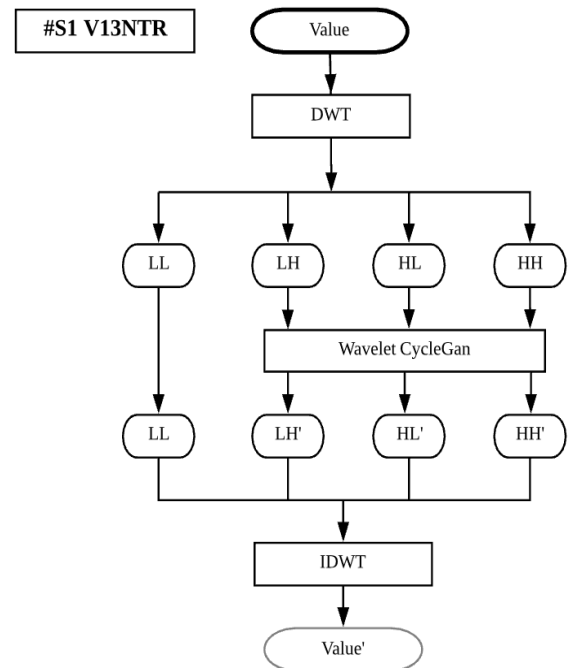*Figure 33:#S1V14NTR, An alternative ν from Figure 32*

Annex B2:

#S1 V13NTR

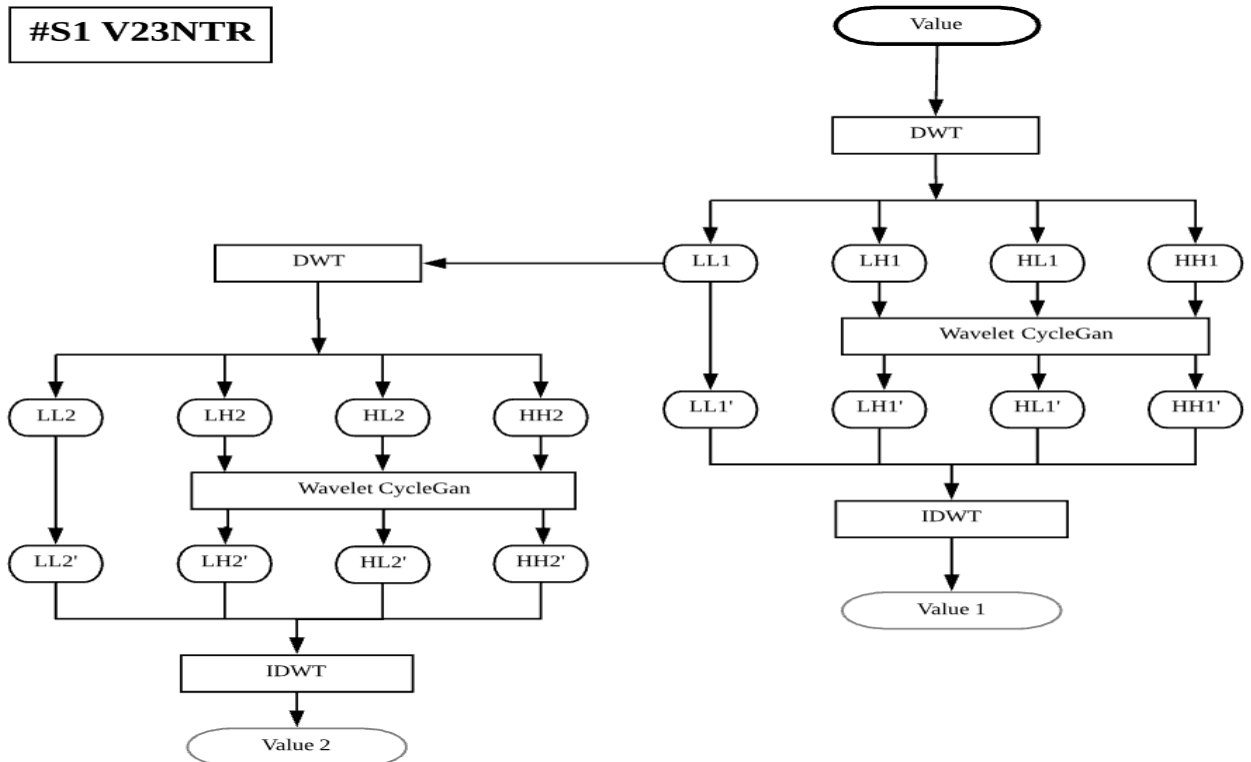*Figure 34 : #S1V13NTR,Analternative ν fromFigure 32*

Annex B3:

#S1 V23NTR

Figure 35: #S1V23NTR, An alternative ν from Figure 32

Annex B4:
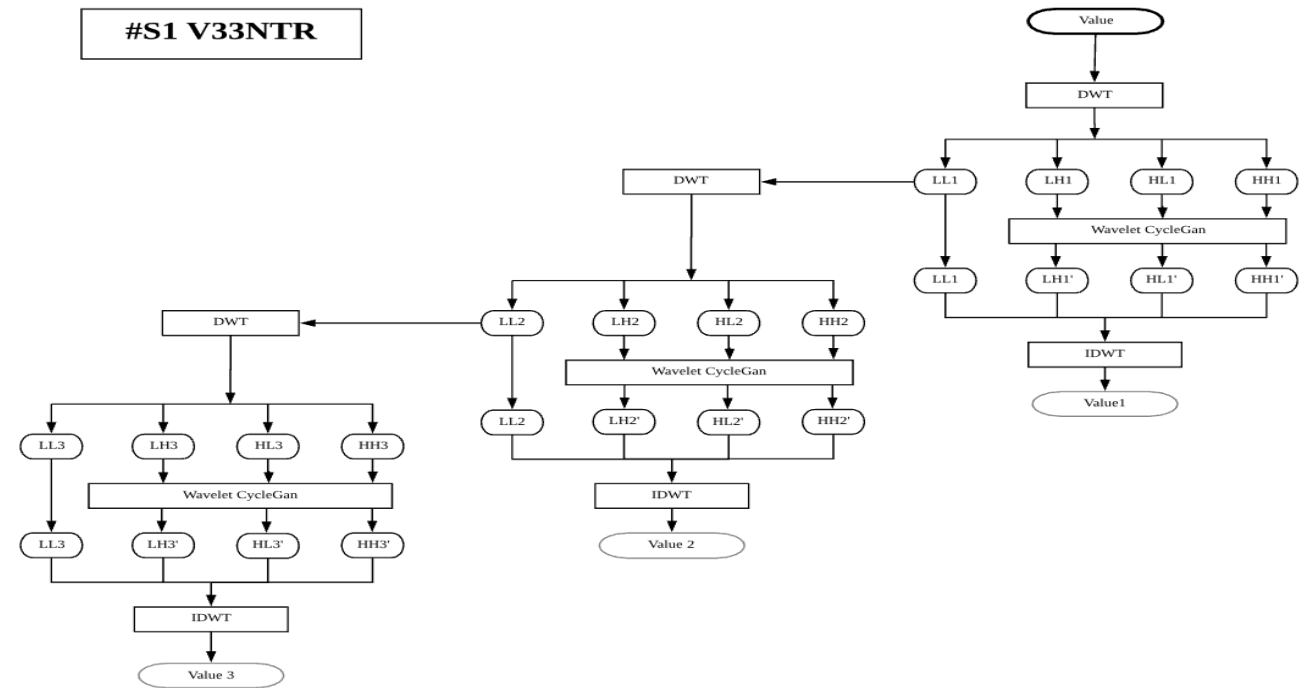
#S1 V33NTR

Figure 36: #S1V33NTR, An alternative v from Figure 32

Annex C1

#S1 V14NTST
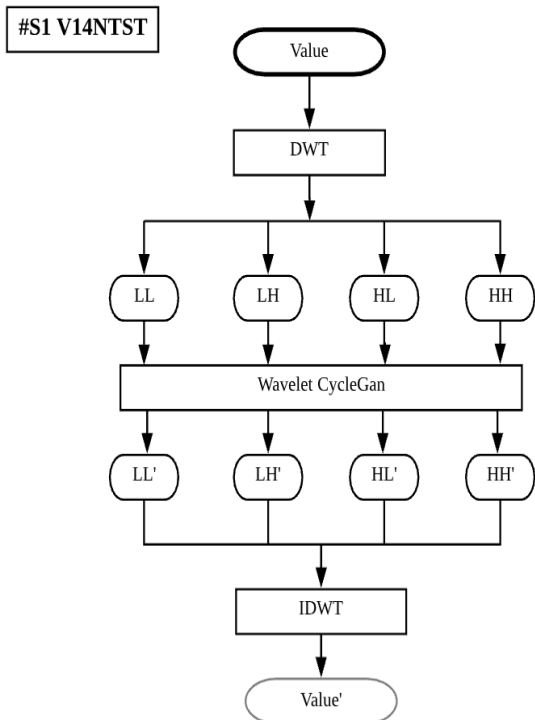
Annex C2

#S1 V13NTST

Figure 37: #S1V14NTST, An alternative v from Figure 32   *Figure 38: #S1V13NTST, An alternative v from Figure 32*
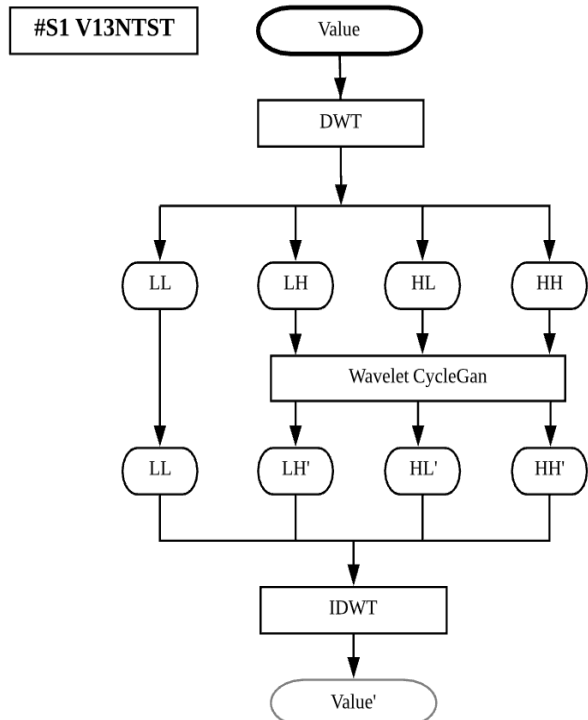
Annex C3

Figure 39: #S1V13G-γTST, An alternative ν from Figure 32

Annex C4

Figure 40:#S1V13ATST, An alternative ν from Figure 32

Annex C5

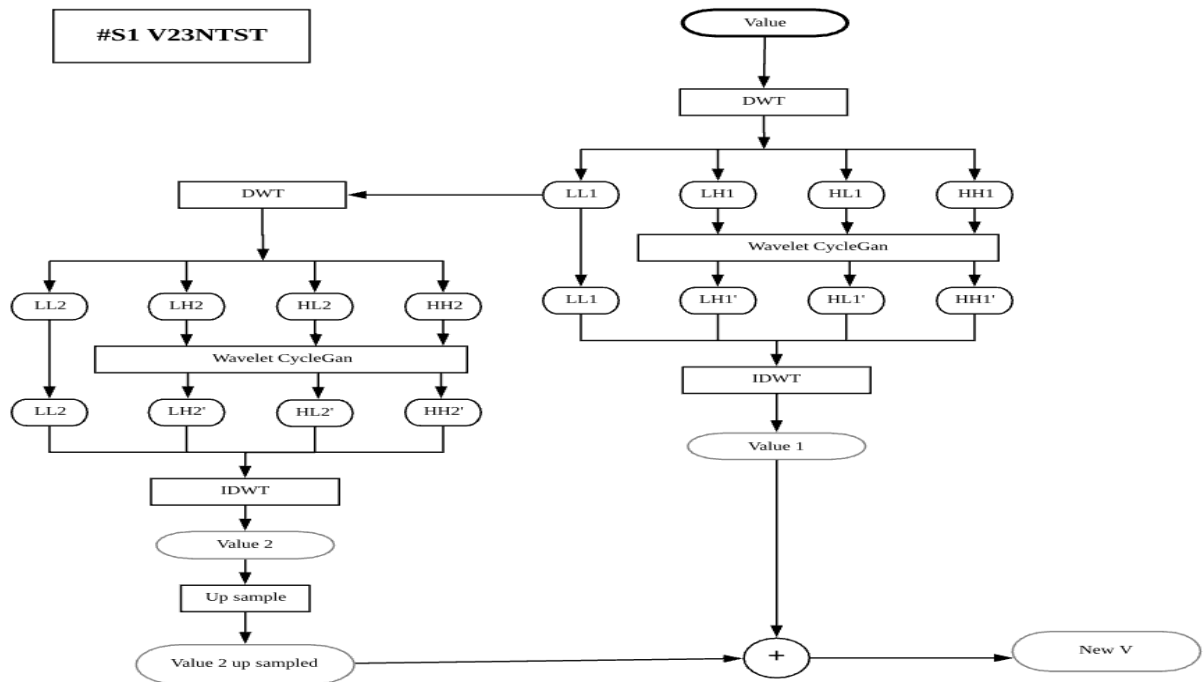*Figure 41: #S1V23NTST, An alternative ν from Figure 32*
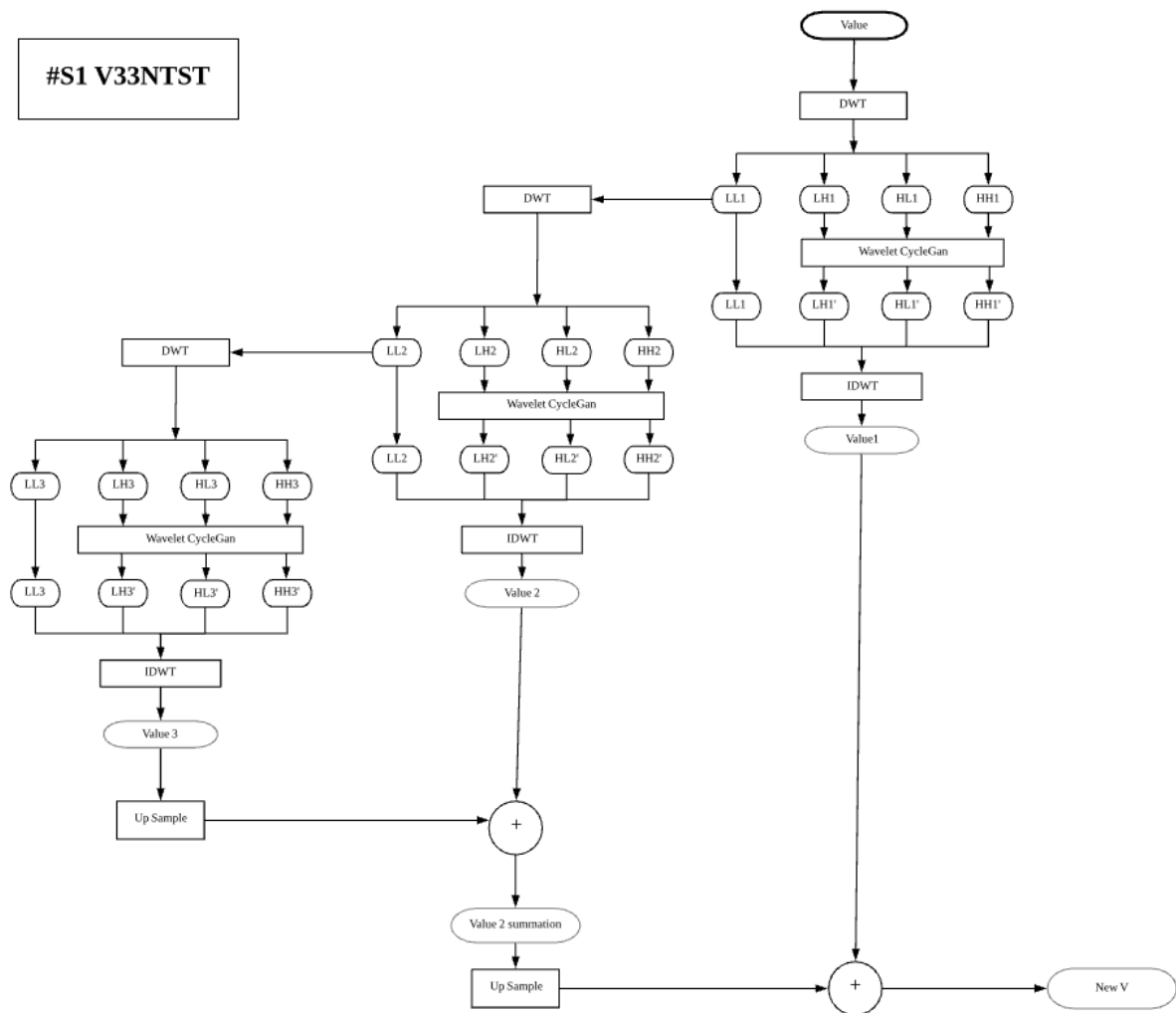
Annex C6

Figure 42: #S1V33NTST, An alternative v from Figure 32

Annex D1

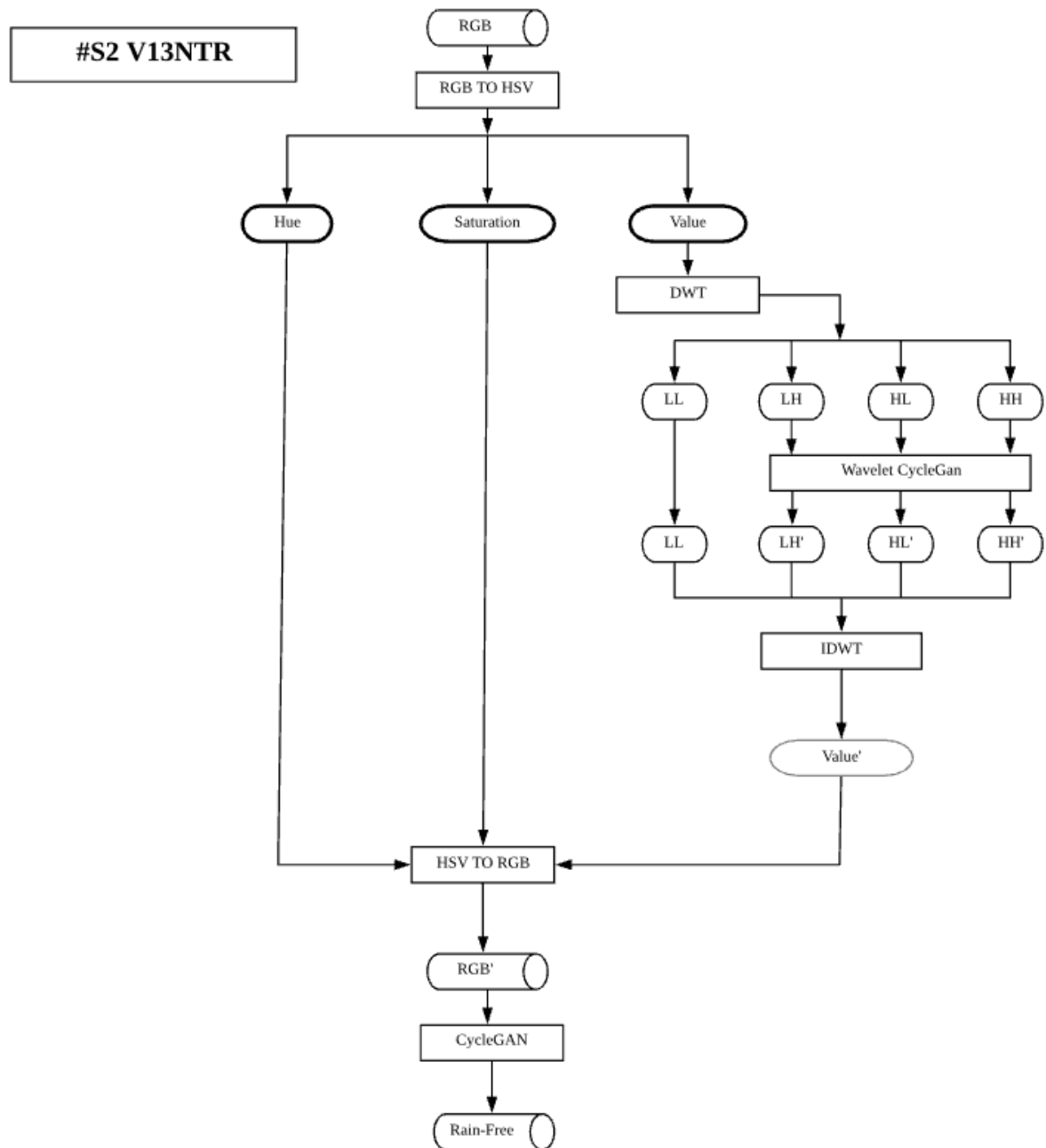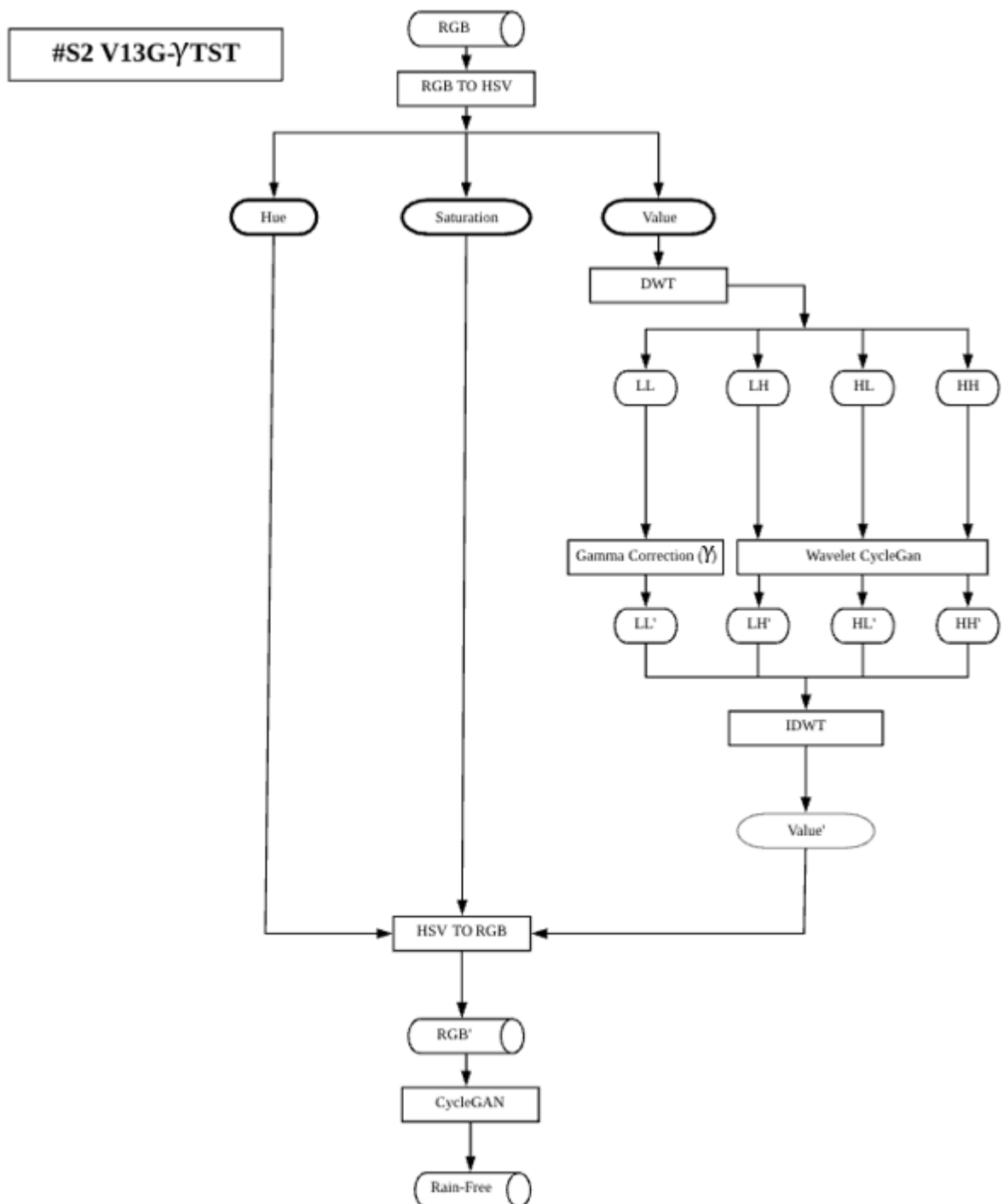*Figure 43: A flowchart of #S2 V13NTR*

Annex E1

#S2 V13G-ɣTST

Figure 44: A flowchart of #S2 V13G-γTST

Annex E2

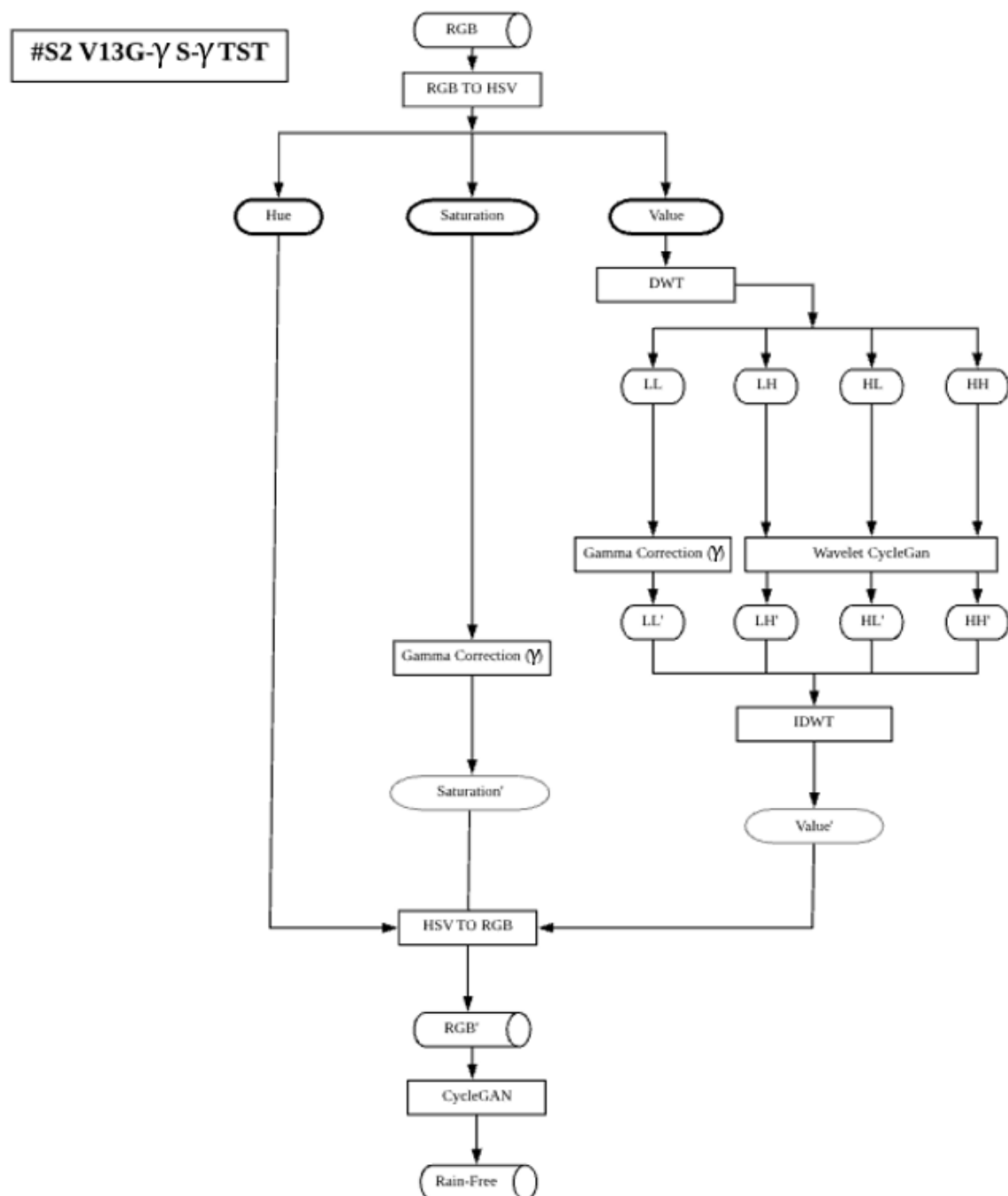*Figure 45: A flowchart of Wavelet CycleGAN*

Annex F

Hardware

Operating System: Window 10 Home-64-bit

Processor: AMD Ryzen 7 3700x-8-Core Processor (16 CPUs), ~3.6GHz

GPU: NVIDIA GeForce RTX 2060

Software

Pytorch – Training and Testing

Matlab – NIQE & SSIM

Training Parameters

Number of training datasets – 700

Number of epochs – 200

Learning rate – Frequency Sub Band is $1e^{-4}$ and Residual Rain Removal is $2e^{-4}$

Testing Parameters

Number of testing datasets – 52

Ideal Gamma Correction for LL in (V) – 1.2

Ideal Gamma Correction in S – 1.05