February 2023
Master's Degree Thesis

A study on old image and video artifact and noise generation using generative adversarial network (GAN)

Graduate School of Chosun University

Department of Information and Communication Engineering

Sadat Hossain

A study on old image and video artifact and noise generation using generative adversarial network (GAN)

February 24, 2023

Graduate School of Chosun University

Department of Information and Communication Engineering

Sadat Hossain

A study on old image and video artifact and noise generation using generative adversarial network (GAN)

적대적 생성망을 이용한 고전 영상 노이즈 생성 기법에 관한 연구

Advisor: Prof. Bumshik Lee

This thesis is submitted to Chosun University in partial fulfillment of the requirements for a Master's degree

October 2022

Graduate School of Chosun University

Department of Information and Communication Engineering

Sadat Hossain

This is to certify that the master's thesis of

Sadat Hossain

has been approved by the examining committee for the thesis requirement for the master's degree in Engineering.

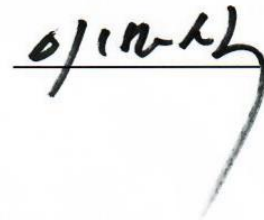**Committee Chairperson: Prof. Chanki Kim** _____

Committee Member: Prof. Young-Sik Kim _____

Committee Member: Prof. Bumshik Lee _____

December 2022

Graduate School of Chosun University

Table of contents

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

A study on old image and video artifact and noise generation using generative adversarial network (GAN)

Sadat Hossain

Advisor: Prof. Bumshik Lee

Dept. of Information and Communication Engineering

Chosun University

In an early age, a large number of images as well as videos were taken and stored in unfavorable conditions. As a result, the old images and videos have uncertain and different noise patterns compared to the modern ones. Denoising for old images is an effective technique that helps to reconstruct a clean image containing crucial information. However, it is difficult and challenging to obtain noisy-clean image pairs for supervised learning, and preparing such a pair is exorbitant and burdensome even if existing denoising approaches require a considerable amount of noisy-clean image pairs. To address this issue, this study proposes a robust noise-generation generative adversarial network (NG-GAN), which utilizes unpaired datasets to replicate the noise distribution of degraded old images based on the CycleGAN model. In the proposed method, the Perception-based Image Quality Evaluator (PIQE) metric to effectively control noise generation is utilized. An unpaired dataset is generated by selecting clean images with matching features from old images to train the proposed model. Experimental results show that the dataset generated by the proposed NG-GAN can better train the state-of-the-art denoising models through denoising the old videos effectively. The difference

in terms of the peak signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM) are 0.37 dB and 0.06 in average, respectively.

Keywords: Generative adversarial network, image denoising, recurrent residual channel and spatial attention, noise generation, perception-based image quality evaluator.

초 록

적대적 생성망을 이용한 고전 영상 노이즈 생성 기법에 관한 연구

호세인 사다트

지도 교수: 이범식

정보통신공학과

조선대학교

  수십년 전에는 사진뿐만 아니라 수많은 동영상을 촬영해 불리한 조건에서 오랜 기간 동안 보관했다. 결과적으로, 오래된 영상과 비디오는 현대의 것과 비교하여 불확실하고 다른 노이즈 패턴을 가지고 있다. 오래된 이미지에 대한 노이즈 제거는 중요한 정보를 포함하는 깨끗한 이미지를 재구성하는 데 도움이 되는 효과적인 기술이다. 그러나 지도 학습을 위한 노이즈-고화질 이미지 쌍을 얻는 것은 어렵고 도전적이며, 기존 노이즈 제거 방법이 많은 양의 노이즈-고화질 이미지 쌍을 요구하더라도 이러한 쌍을 얻는 것은 많은 어려움이 존재한다. 이 문제를 해결하기 위해 본 학위 논문에서는 성능이 저하된 오래된 이미지의 노이즈 분포를 복제하기 위해, 쌍을 이루지 않는 데이터 세트를 활용하는 CycleGAN 모델을 기반으로한 강력한 노이즈 생성 적대 네트워크(Noise Generation Generative Adversarial Network, NG-GAN)를 제안한다. 제안된 방법에서는 노이즈 생성을 효과적으로

제어하기 위한 PIQE (Perception-based Image Quality Evaluator) 메트릭을 사용한다. 제안된 모델을 훈련시키기 위해 오래된 이미지에서 특징이 일치하는 깨끗한 이미지를 선택하여 쌍을 이루지 않은 데이터 세트를 생성한다. 실험 결과는 제안된 NG-GAN에 의해 생성된 데이터 세트가 이전 비디오의 노이즈를 효과적으로 제거하여 최신 노이즈 제거 모델을 더 잘 훈련시킬 수 있음을 보여준다. 평균적으로 제안 방법으로 생성한 데이터셋은 PSNR (Peak Signal-to-Noise Ratio)과 SSIM (Structural Similarity Index Measure) 각각 0.37dB와 0.06의 성능 향상을 보였다.

키워드: 생성적 적대 네트워크, 이미지 노이즈 제거, 반복 잔류 채널 및 공간 주의, 노이즈 생성, 인식 기반 이미지 품질 평가기.

1. INTRODUCTION

1.1. Overview

Image denoising primarily focuses on eliminating unwanted signals from the given noisy observations. Considerable research has been conducted in this field, which is considered one of the most fundamental vision issues [1, 2, 3]. Significant advances have been made in image denoising with the advent of deep learning. Although deep convolutional neural networks (CNNs) for image enhancement have shown promising results [4–8], [34–40], several crucial obstacles prohibit their deployment in real-world applications. Because learning-based techniques are typically data-driven, training on a given dataset does not always ensure generalization to real-world scenarios. For various reasons, noise that instantiates from a camera pipeline differs from the theoretical noise assumption. For example, common additive white Gaussian noise (AWGN) implies that the given term is signal-independent [9, 10], which differs from actual noise. Hence, when a denoising algorithm is trained on synthetic data, such as AWGN, generalizing it to image restoration is difficult. Executing learning-based algorithms on a significant number of high-quality datasets is crucial. Most conventional learning-based denoising methods focus on the traditional Gaussian denoising problem and pay more attention to the architecture design of deep learning networks because creating a pair of noisy and noise-free images is simple using additive synthetic noise.

In [11-12], well-aligned noisy and clean image pairs with real-world noise were collected, allowing denoising algorithms to be learned in a supervised manner. Although such a technique successfully addresses real-world noise, obtaining large-scale pairings remains challenging due to two main practical difficulties. First, this is because of the lack of denoised or enhanced versions

of old images. In addition, old images are more likely to degrade in a more complicated manner than modern images. Second, no degradation model can accurately depict the artifacts of old images because the network cannot approximate them because of the domain disparity between synthetic and actual old images.

Generation-based techniques have been developed to address these issues [13, 14]. These methods employ noisy target images to train a noise generator, producing pseudo-noisy images coupled with clean images that are then used for training a denoising model. Following the success of conventional synthetic noise reduction technologies, attempts have recently been made to adapt this technology to real-world noise [13]. However, no generation-based solution which properly replicates real-world noise is proposed without supplying associated clean pictures to the target noisy images.

Gaussian and digital camera noise are insufficient for creating noise for the old film; generating global noise artifacts that can alter the contrast and brightness of the entire frame must be possible, as well as local noise that affects only a small area of the image. Actual old images are significantly more difficult to generate accurately because they frequently suffer severe deterioration from various unknown degradations. Furthermore, with technological advancements, current digital cameras are considerably more advanced in capturing the subtle characteristics of images than old cameras. Thus, images captured with modern cameras are unlikely to contain similar noise, distortion, or artifacts to those of old images. Hence, the collection of datasets for paired old and clean images is a challenging task.

This study proposes a Noise-Generation Generative Adversarial Network (NG-GAN), a noise-generation framework that can be trained without paired datasets. Using the Perception-Based Image Quality Evaluator (PIQE) [24]

2

metric with a clean image, noisy images can be generated in a more elaborately controlled manner.

1.2. Research Objective

Image-to-image translation methods such as Pix2Pix, CycleGAN and DualGAN is well-known unsupervised image translation methods. The basic working principle is that the models learn the translation using paired and unpaired images from distinct domains. When the models are utilized to generate realistic old image noise, they tend to focus on generating general translation, such as image color. However, they do not focus on generating detailed information such as noise information and texture of old photos, which are different from the synthetic dataset. Images generated by these models lose significant noise information and variation in the noise pattern. To overcome these limitations, the generator architecture is designed carefully by providing additional information along with clean images, add loss functions as well as modify the discriminator architecture to be able to focus on generating realistic-looking old images. This work proposes a noise generation generative adversarial network (NG-GAN), a novel noise generation framework that can be trained without using a paired dataset. By passing the PIQE metric [24] with the clean image, noisy images can be generated in a controlled manner. The following is a summary of the contributions of this work:

i. A robust framework, NG-GAN, has been introduced, which successfully imitates the noisy pattern of the degraded images.
ii. A noise-generation framework for old images and videos using a no-reference PIQE metric and an unpaired clean image to generate a noisy image based on the value of the PIQE metric is proposed.

iii. A recurrent residual convolutional and attention mechanism-based generator architecture is designed in the proposed NG-GAN method, which successfully imitates the noisy pattern of degraded images.

iv.   When state-of-the-art (SOTA) video restorers are trained on the datasets generated by the NG-GAN, they can effectively produce clean videos from noisy ones in terms of the peak signal-to-noise (PSNR) and structural similarity index measure (SSIM).

1.3. Thesis Layout

There are five chapters in this thesis. Following the introduction in Chapter 1, Chapter 2 gives an overview of deep learning algorithms for noisy image and video generation. The suggested method and workflow are explained in Chapter 3, notably the NG-GAN architecture and fusion with the convolutional block attention module. The experimental results are compared with the data of recent algorithms in Chapter 4. Finally, Chapter 5 wraps up with a brief overview of the findings of the thesis.

2. RELATED WORKS

2.1. Noisy Image Generation

In generative adversarial network (GAN)[16], a minmax game between the generator and the discriminator is defined. The purpose of the generator is to provide compelling samples that deceive the discriminator, allowing the generated samples to be distinguished from the ground truth. Then, GAN is used for visual enhancing and restoration, such as super resolution [42], image inpainting [43], and style transfer [44]. The first widely used GAN-based paired image-to-image translator is Pix2Pix GAN [19], unpaired image-to-image translator is CycleGAN [18], and DualGAN [41], which converts images from one domain to the other. Though they are used to map images from one domain to the other, they show the difficulties of generating fine noisy images for given a clean image. Instead of employing a single model, generation-based approaches generally use a two-stage pipeline to solve the denoising problem [9, 13, 14 ,15]. First, an unsupervised noise generator is learned to replicate the distribution of provided actual noisy samples, allowing any clean picture to be translated to pseudo-noisy data. The synthesized input and target pairs may then be used to train a denoiser in a straightforward manner. This GAN, like most others, aims to approximate the probability distribution of real-world noisy images by treating images as samples. This image-level GAN produces coarse learning of the real noise distribution because it ignores the fact that each pixel of a real noisy image is a random variable and that the real noise is spatio-chromatically associated. The NTGAN [21] approach illustrates that noise maps created with a certain camera response function may be employed in the denoiser. The GAN2GAN [9] approach makes use of improved noisy-patch extraction to provide better realistic noisy images for training a denoising model. The DA-Net [20] also

learns the denoiser and generator translations by comparing them to the real noisy-to-clean joint distribution, but this model requires paired data to do so. All the papers mentioned above have worked with digital camera-captured images, whereas this study has worked to generate noisy images that match the old images noise and degradation.

Image-to-image translation methods, such as Pix2Pix, CycleGAN, and DualGAN, are well-known unsupervised image translation methods. The basic working principle is that the models learn the translation using paired and unpaired images from different domains. When such models are utilized to generate realistic old image noise, they tend to focus on generating general translations, such as image color. However, they fail to generate detailed information, such as the noise and texture of old images, which significantly differ from that in the synthetic dataset. Therefore, images generated by these models lose significant noise information and variation in the noise pattern. To overcome these limitations, the generator architecture is carefully designed by providing additional information with clean images, added loss functions, and modified the discriminator architecture to focus on generating realistic-looking old images.

2.2. Channel and Spatial Attention

At present, deep learning uses attention algorithms frequently to improve feature extraction [45]. ECANet [46], which employs a local cross-channel connection method without downscaling and adaptive kernel selection for one-dimensional convolutional networks. There are several dual-attention mechanisms in addition to these single-channel ones. Using a channel attention and a spatial attention mechanisms, a convolutional block attention module (CBAM) [29] was introduced to enhance relevant information and eliminate pointless information.

3. PROPOSED METHOD

3.1. Problems in Old Degraded Images

Figure 3-1 shows the histogram comparison between AWGN and realistic old image noise. A smooth region is extracted from the AWGN-added image and the actual old noisy image and then plotted the histogram to show the difference between the distributions. The histogram shows that the AWGN has a bell-shaped distribution, whereas the old image noise does not have a smooth distribution and contains small peaks in the distribution. The smooth region is defined as pixel areas where the mean pixel value in the region approximates the pixel value itself. That is, R is a region in the image defined by $R \in R^{M \times N}$, and, if the intensity values of R are denoted by $I_R(x, y)$, a smooth region is defined as any region satisfying $\sum_{x=1}^{M} \sum_{y=1}^{N} |E(I_R) - I_R(x, y)| \approx 0$. Assume that regions R_1, R_2 and R_3 are smooth regions corrupted by a certain type of noise in the old images, and, in the AWGN-added image, they are corrupted by Gaussian noise. Distribution of noise in these regions is approximated using a histogram because these regions provide us with noise information.

| AWGN added image | Smooth region (R_1) | Smooth region (R_2) | Smooth region (R_3) | Noisy image histogram (R_1) | Noisy image histogram (R_2) | Noisy image histogram (R_3) |

| Real old noisy image | Smooth region (R_1) | Smooth region (R_2) | Smooth region (R_3) | Noisy image histogram (R_1) | Noisy image histogram (R_2) | Noisy image histogram (R_3) |
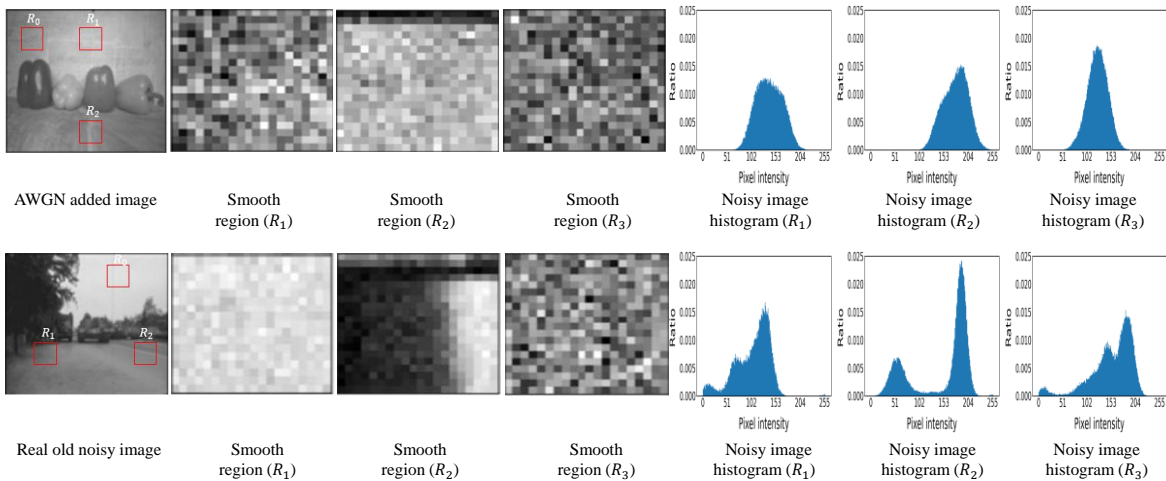
Figure 3-1. Histogram comparison between AWGN and realistic old image noise.

It is important to know the attributes and statistical characteristics of complex real noise for developing realistic noise using deep learning networks. Noise in old image usually emanates from sources in low-performance cameras in the early stage, such as artifacts related to compression, quantization, photon noise, and in-camera amplifiers. When all of these components are combined, the pixel-wise distortion is blended with a baseline clean signal to produce a noisy distorted image as:

$$I^n = I^c + y \tag{1}$$

Where I^n is noisy image, I^c is clean image and y is pixel-wise distortion. The noise component y is commonly approximated as an AWGN in traditional deep denoising approaches [4, 5]. In [11], although the noise model can offer a reasonable estimate of actual noise, many investigations have r that actual scenarios are significantly more intricate [22, 23]. Therefore, a learning-based strategy is used to imitate real-world noise rather than handmade approaches to solve the problem without employing paired data. To imitate the pattern of real-world noise, our architecture fully exploits the benefit of unsupervised learning.

3.2. Proposed Network Architecture

Figure 3-2. Summary of NG-GAN framework. It contains two generators and two discriminators. G_1 generates degraded images, whereas G_2 reconstructs the clean version of the image. D_1 and D_2 determines how much clean or fake a given input is.

A denoising network attempts to recover the clean data from the given noisy image if sufficient data pairs are used in supervised learning. However, for old

image denoising, collecting clean-old noisy image pairs is challenging. First, clean images were collected from multiple sources, such as the REDS, PASCAL VOC, and DIV2K datasets 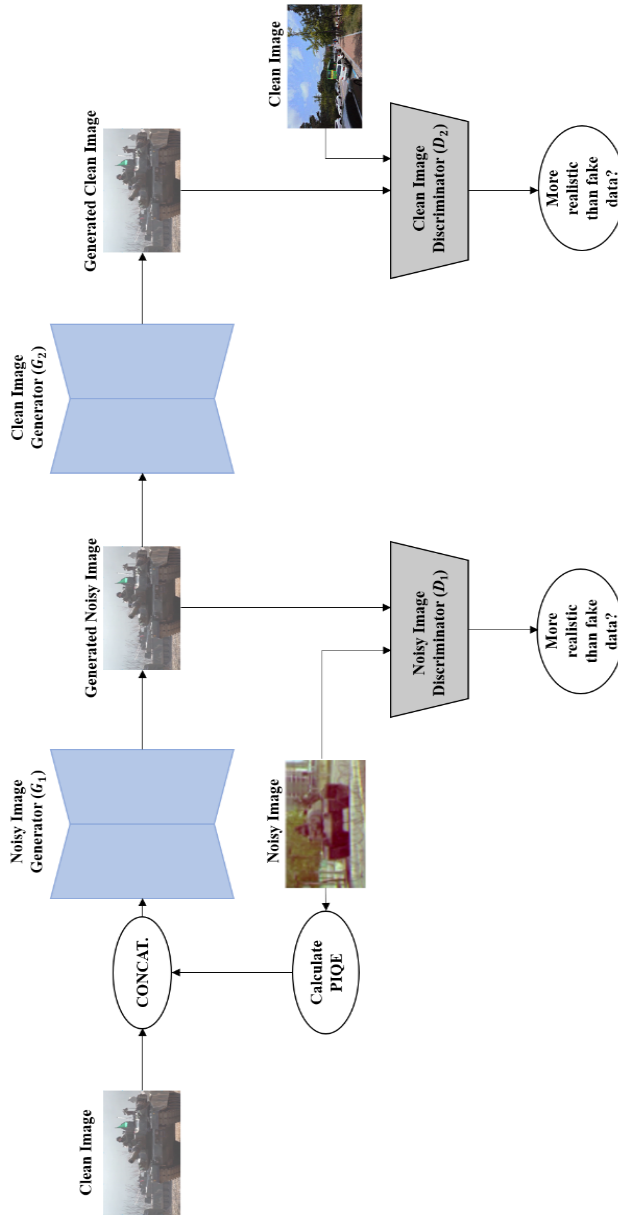[49, 51, 52]. Then, our proposed NG-GAN model is used to generate the target noise distribution, which can be obtained from the actual old images. In our proposed method, old images were also collected from the frames of old movies, such as D.O.A. (1949), Midnight Intruder (1938), A Matter of Life and Death (1946), and Bonjour Tristesse (1958).

The proposed NG-GAN is inspired by the CycleGAN framework [18]. Figure 3-1, shows the overall framework of the proposed NG-GAN. CycleGAN has shown promising performance in color transformation and image transformation from one domain to another, such as sketch-to-photo photograph-to-Monet applications, as well as object transfigurations, such as in transfiguring a horse into zebra. In addition, CycleGAN helps obtain paired datasets using unpaired datasets. However, when CycleGAN was applied to generate old noisy images, our experimental investigation observed that the generated image showed a lack of variety in noise patterns and was likely to change the image geometry from the original image. It also struggled to separate an object from the context owing to its generator architecture and loss functions [55]. To eliminate the problem of unpaired image generator networks in generating realistic-looking noisy images, as mentioned above, the PIQE metric [24], a no-reference Perception-based Image Quality Evaluator, to guide the network on the degradation quality of the old images has been utilized. VGG-19 [27] and SSIM [28] loss were used to guide the network in generating noisy images while maintaining the visual quality and structure of the images. Recurrent residual network strategy is used to aid in deep architectural training and better feature representation is ensured by

10

feature accumulation with recurrent residual convolutional layers. Additionally, Convolutional Block Attention Module (CBAM) [29] is used in place of skip connections to prevent the network from learning unnecessary background information, and to force it to learn and concentrate more on the key information. Moreover, this enables our network to accurately capture various features to pay attention to the most informative features to generate degraded images.

First, PIQE values [24] of the noisy images are obtained. The PIQE uses block-wise distortion estimation to compute the no-reference quality score of an image. Initially, the Mean Subtracted Contrast Normalized (MSCN) coefficient for every pixel in an image is calculated. Then, the image is divided into uniform-sized 16×16 blocks. Based on the variance of the MSCN coefficients, high spatially active blocks are identified. Then, utilizing the recognized spatially active blocks, an activity mask is generated, representing the regions of input image areas with higher levels of spatial variability caused by noise and compression artifacts.

After that, it uses the MSCN coefficients to analyze distortion caused by blocking artifacts and noise in each block. A threshold criterion is used to differentiate distorted blocks with blocking artifacts, undistorted blocks, and distorted blocks with Gaussian noise. Later, the spatial quality mask of noticeable artifacts is generated from the distorted blocks with blocking artifacts and the spatial quality mask of Gaussian noise from the distorted blocks with Gaussian noise. Finally, the PIQE score of the input image is computed as the mean of scores in the distorted blocks.

The computed PIQE score of the noisy image is spatially replicated through all pixel positions of I^C. The noisy image generator G_1 generates a blurry noisy version of the clean image depending on the PIQE value. The higher the

PIQE value, the higher the amount of noise it generates. The PIQE value ranges from [0-100]. The clean image generator G_2 reconstructs the clean image from the fake noisy image generated by G_1. Two discriminators D_1 and D_2 provides an approximation of how real or fake the generated noisy and clean image is, respectively. The losses to train NG-GAN can be represented as:

$$l_1 = |I^c - G_2(I^g)| \tag{2}$$

$$l_{VGG/i.j}^{Rec} = \frac{1}{W_{i,j}H_{i,j}} \sum_{x=1}^{W_{i,j}} \sum_{y=1}^{H_{i,j}} \left(\phi_{i,j}(I^c)_{x,y} - \phi_{i,j}\big(G_2(I^g)\big)_{x,y}\right)^2 \tag{3}$$

$$l_{SSIM} = 1 - SSIM\,(I^{cl}, G_2(I^g)) \tag{4}$$

Equations (2), (3), and (4) show the losses required to train the noise generator architecture. The generated noisy images should be as close as possible to the clean input images in terms of structure. Hence, in this study adopt the $l_1, l_{VGG/i.j}^{Rec}, l_{SSIM}$, loss, were adopted, where l_1 is a content loss that measures the l_1 norm distance in between reconstructed image $G_2(I^g)$ and the original clean image I^c. $l_{VGG/i.j}^{Rec}$ loss is based on the pre-trained 19-layer VGG network's ReLU activation layers. The indices i and j indicate that the ith maxpooling layer and j^{th} convolution (after activation) within the VGG19 network, respectively.

The Euclidean distance between the features extracted from a reconstructed image and the reference one is then defined as the VGG loss. Mean Squared Error treats every pixel as a separate entity, ignoring all spatial interactions between pixels. As a result, SSIM is utilized as a loss between I^c and $G_2(I^g)$. It was implemented and tested utilizing perceptual quality metrics in connection to visual perception in the human brain. Human subject ratings

provided validation for it. SSIM assesses picture quality from the standpoint of human visual perception, making it more suitable for loss function. The SSIM index is derived using common-size windows x and y between the pictures. Combining Equations (2), (3), and (4), the total loss for the generator is optimized as:

$$L_G = \lambda_{l1} l_1 + \lambda_{per} l_{VGG/i.j}^{Rec} + \lambda_{PIQE} l_{PIQE} + \lambda_{SSIM} l_{SSIM} + \mathcal{L}_{G^{Ra}} \qquad (5)$$

where $\mathcal{L}_{G^{Ra}}$ is the adversarial loss, which is discussed in Section 2.4, and λ_{l_1}, λ_{per}, and λ_{SSIM} are the coefficients used to balance the various loss terms.

3.3. Proposed Generator Architecture

The generator architecture of R2U-Net [26] is modified and used in our method. Figure 3-3, shows the overall architecture of the generator in the proposed NG-GAN. Similar to the GAN application [48], a random gaussian noise from $N(0, 1^2)$ has been sampled, which is zero mean and one standard deviation, then add to pixel coordinates of the clean image to produce a random distribution that generates various noisy photos of the same scenario. Two recurrent residual convolutional blocks were proposed within the Recurrent Residual Convolutional Layer (R2CL) of the proposed generator. In the encoding path, five R2CL blocks has been designed to extract features from a given image, within each R2CL block, there are two blocks, each containing a 3x3 convolutional layer followed by a batch normalization and a ReLU activation,features extracted from the first convolutional layer are passed through a channel attention block, which contains a global average pooling layer followed by a 1-D channel attention layer, it is used to effectively capture channel correlation and prevent information loss in the features. The attention calculated features are the passed to the second convolutional layer to extract more feature maps. A recurrent convolutional block with a residual unit

13

without a channel attention block was used in the decoding stages. Second, the CBAMs are used for adaptive feature refinement instead of skip connections.

Finally, in the up-sampling process, batch normalization (BN) was employed to improve the stability of the network and accelerate convergence [30]. Every stage in the encoding process includes a recurrent residual convolutional unit In addition, residual connections were introduced to construct more efficient and deeper models. The set of feature maps was doubled, and the size was reduced by half each time a recursive residual convolutional unit was processed. In the R2CL, recurrent convolutional layers are applied in discrete time steps, as specified by the recurrent convolutional neural network (RCNN).
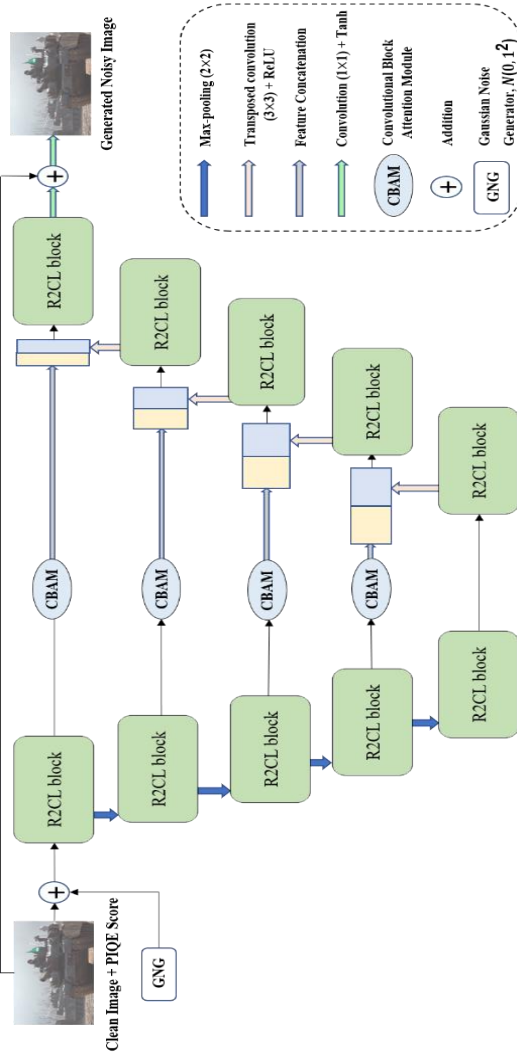
Figure 3-3. Generator architecture containing convolutional encoding and decoding units that are based on recurrent residual convolutional layer (R2CL) and Convolutional Block Attention Module (CBAM).

The recurrent convolutional block with the residual unit without channel attention block is utilized in the decoding stages. Second, for adaptive feature refinement, the skip connections are replaced with CBAM. Finally, BN [30] is employed in the up-sampling process to improve the neural network's stability and quicken its convergence speed. Every stage in the encoding process includes a recurrent residual convolutional unit, which is made up of two 3×3 convolutions and incorporates recurrent connections to every convolutional layer to improve the model's capacity to integrate contextual data. To construct

better efficient and deep models, residual connections are also introduced. The set of feature maps doubles and the size is reduced in half each time a recursive residual convolutional unit is processed. In R2CL, the recurrent convolutional layers are applied in discrete time steps as specified by RCNN. Consider the p_l as an input sample at the *l*-th layer of a block in the recurrent residual convolutional layer (R2CL) and (i, j) as a pixel located in an input sample of the *k*-th feature map in the recurrent convolutional layer (RCL). The output $X_{ijk}^l(t)$ at time step *t* is denoted as:

$$X_{ijk}^l(t) = \left(w_k^f\right)^T p_l^{f(i,j)}(t) + (w_k^r)^T p_l^{r(i,j)}(t-1) + b_k \qquad (6)$$

here, $p_l^{f(i,j)}(t)$ and $p_l^{r(i,j)}$ are standard convolutional layers and the input sample to the *l-th* RCL, respectively. The RCL generated from the *k*-th feature map and the standard convolutional layers are weighted by w_k^r and w_k^f, respectively, where bias is denoted b_k. The standard ReLU function activates the output of RCL as (7).

$$F(p_l, w_l) = f\left(X_{ijk}^l(t)\right) = max\left(0, X_{ijk}^l(t)\right) \qquad (7)$$

The output generated by the R2CL unit is given as:

$$p_{l+1} = p_l + F(p_l, w_l) \qquad (8)$$

where the input of the R2CL layer is denoted as p_l and p_{l+1} which represent both the results derived from the downsampling layer as well as a result from upsampling layer from the encoding and decoding path, respectively. The basic building blocks of the R2CL layer are illustrated in Figure 3-4.

The upsampling operation related to the output derived from the R2CL unit is performed by each phase of the decoding path. After implying the up-sampling technique, the feature maps will be reduced by fifty percent, and the

16

size is increased twice. The feature map size is reconstructed to the actual input image size in the final layer of the decoding path.
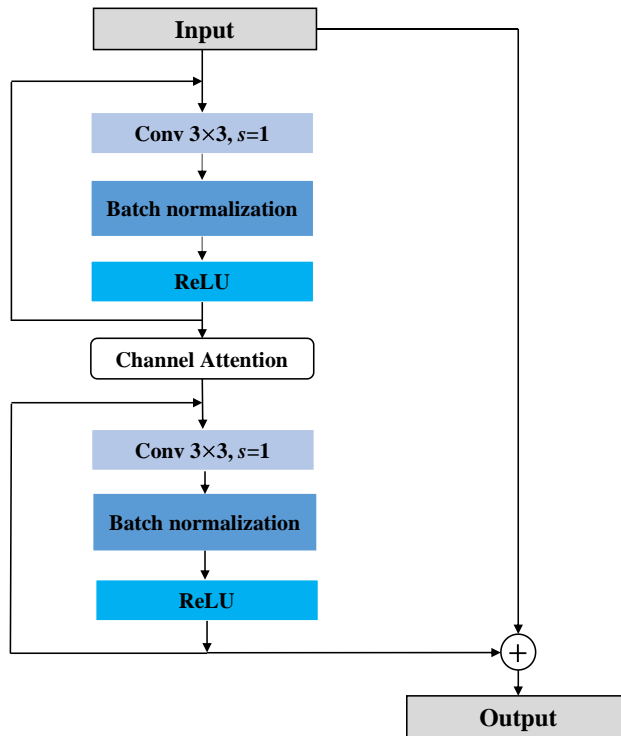
Figure 3-4. The R2CL block.

As shown in Figure 3-4, the result from the BN layer is fed onto the convolutional block attention module (CBAM) [29]. The CBAM consists of two sequential modules: the channel and spatial modules. In order to inform the model about where to focus, the CBAM builds attention maps for spatial and channel information from the given input features and then then the attention map is multiplied by the input feature maps. The intermediary feature map is refined using the CBAM module in each block of the deep network. The refined feature map is then concatenated with the feature maps obtained from the transpose convolution operation, as shown in Figure 3-5.

17
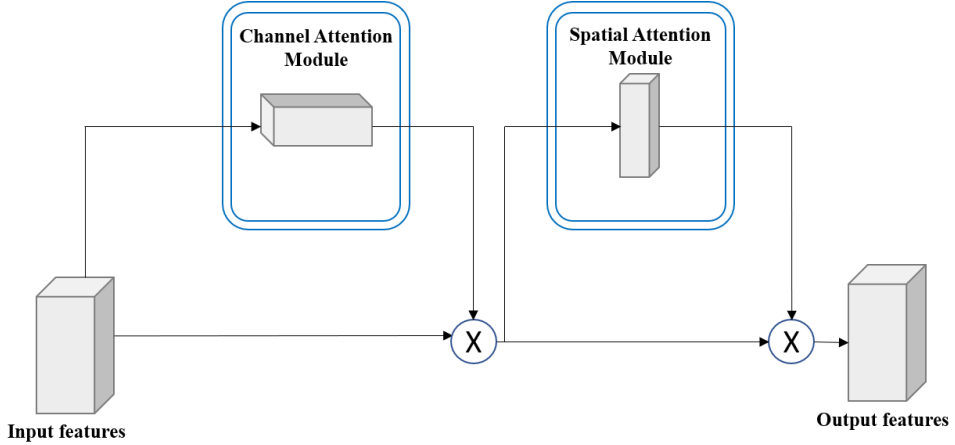
Figure 3-5. The Convolutional Block Attention Module (CBAM) block. It consists of channel and spatial modules. The features maps from encoding layers are refined through the CBAM block.

3.4. Discriminator Architecture

In the proposed architecture, the discriminator architecture was improved using a relativistic GAN [31], which differs from the standard discriminator D. This was used to improve the discriminator performance. A relativistic discriminator aims to estimate the likelihood that a real image is more realistic than a fake one better than the conventional discriminator D, which predicts whether the likelihood that an input image is real. The relativistic discriminator aims to estimate the likelihood that real image i_r is more realistic than generated image i_f. The architecture is shown in Figure 3-6.

The standard discriminator is expressed as $D(x) = \sigma(C(x))$, where $C(x)$ is the non-transformed output from the discriminator and σ is the sigmoid function. The relativistic average discriminator D_{Ra} is formulated as:

$$D_{Ra}(i_r, i_f) = \sigma\left(C(i_r) - \mathbb{E}_{i_f}\left[C(i_f)\right]\right) \tag{9}$$

18

in Equation 9, i_r represents the real noisy image, i_f is the fake noisy image and $\mathbb{E}_{i_f}$ is an average operator on all generated data in a mini-batch, respectively the discriminator loss is defined as:

$$\mathcal{L}_{D^{Ra}} = -\mathbb{E}_{i_r}\left[\log\left(D_{Ra}(i_r, i_f)\right)\right] - \mathbb{E}_{i_f}\left[\log\left(1 - D_{Ra}(i_f, i_r)\right)\right] \quad (10)$$

the generator adversarial loss is defined as:

$$\mathcal{L}_{G^{Ra}} = -\mathbb{E}_{i_r}\left[\log\left(1 - D_{Ra}(i_r, i_f)\right)\right] - \mathbb{E}_{i_f}\left[\log\left(D_{Ra}(i_f, i_r)\right)\right] \quad (11)$$

The adversarial loss of generator includes both i_r and i_f. Consequently, in adversarial training, the generator updates itself according to the discriminators' output of both fake and actual data.
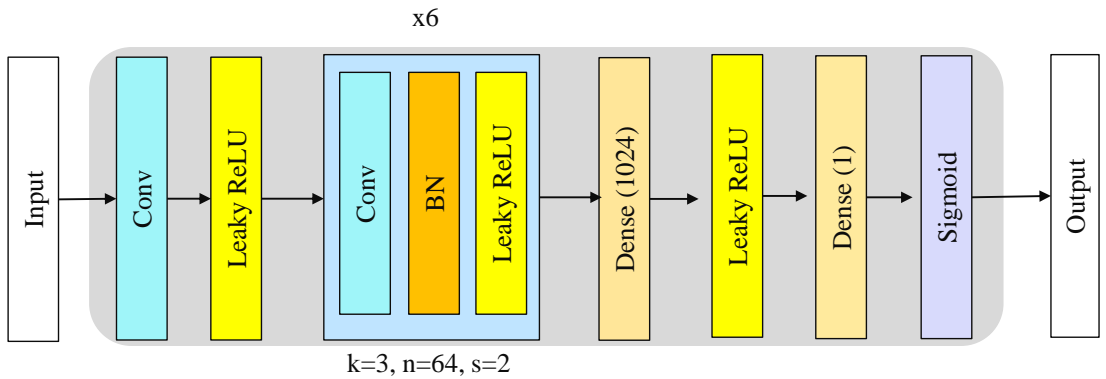

Figure 3-6. Discriminator architecture.

4. EXPERIMENTAL ANALYSIS

4.1. Hyperparameters

Table 1. List of simulation parameters.

Hyperparameters	Value
λ_{l_1}	5.0
λ_{per}	0.08
λ_{PIQUE}	0.3
λ_{SSIM}	0.1
β_1	0.5
β_2	0.999
Image size	64×64
Learning rate	1×10^{-5}

In this thesis, the values of the coefficient are set to $\lambda_{l_1}= 5.0$, $\lambda_{per} = 0.08$, $\lambda_{PIQE} = 0.3$ and $\lambda_{SSIM} = 0.1$. All the sub-modules are trained with the Adam optimizer where $\beta_1 = 0.5$ and $\beta_2 = 0.999$. Images are cropped to size 64×64 and fed into the model. The batch size is set to 1. 17,000 image patches were cropped with a size of 64×64 pixels from clean and noisy images and sampled those images to train the model; horizontal and vertical flips and random rotations $90 \times \theta$, where $\theta = 0, 1, 2, 3$, were performed for data augmentation. To collect more noisy images, noisy patches were extracted from old noisy images, and those noisy patches were added to the clean images. Noisy patches were extracted using the noise block extraction algorithm [50]. In the training stage, the learning rate was given as 1×10^{-5}. After every 14 epochs, the learning rate was reduced by multiplying with 0.8

for model stabilization. All models were trained on a GeForce RTX3090 GPU, G6X memory of 24 GB. The list of all the simulation parameters are given in Table 1.

4.2. Datasets

To train the proposed model, high-quality clean images from REDS, PAS-CAL VOC, and DIV2K datasets were used. REDS contains 240 videos, each video with 100 frames. Hence it contains a total of 24000 clean images. The PASCAL VOC dataset contains 17,125 high-quality clean images, and DIV2K contains 800 high-quality clean images. Noisy images were collected by extracting frames from old movies from the 1920s - 1970s as noisy samples, and clean images were also distorted by adding Gaussian blur, JPEG compression, and adding the noisy patches that were extracted from old videos using the noise estimation method [50].

Figure 4-1, shows five old noisy images collected from movies from the 1920s – 1970s. The old images in the film are contaminated with complicated degradation noise, which is different from synthetic noise and difficult to model mathematically because of the uneven distribution of the noise. The noise types in the old movies include compression artifacts from compression algorithms, blur noise that occurs due to improper camera lens alignment, unstructured defects such as film grain, color fading, and structured defects, e.g., scratches and dust spots. Hence, replicating these noisy patterns is more difficult compared to the digital noise in modern images.
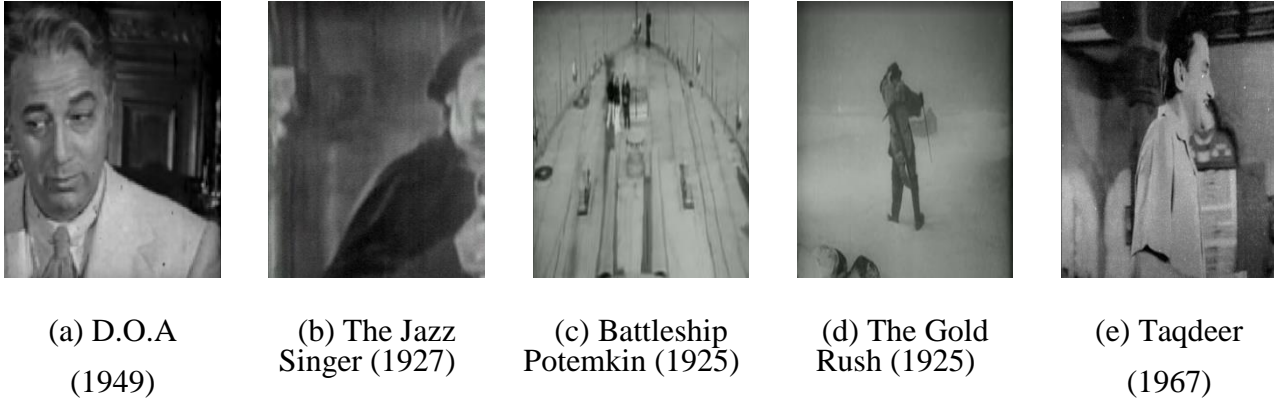
(a) D.O.A
(1949)

(b) The Jazz
Singer (1927)

(c) Battleship
Potemkin (1925)

(d) The Gold
Rush (1925)

(e) Taqdeer
(1967)

Figure 4-1. Examples of old noisy video frames collected from old movies
from the 1920s – 1970s.

4.3. Qualitative Comparison of Denoised Video

The datasets generated by C2N [48], CycleGAN [18], and the proposed NG-GAN were validated using SOTA denoising networks: BasicVSR [32] and BasicVSR++ [47]. These two SOTA denoisers exhibit the best performances in image denoising. The effectiveness of the architectures was validated through a qualitative comparison of the PSNR and SSIM values. For comparison, C2N, CycleGAN, and NG-GAN were trained under the same datasets and conditions, and the same number of paired datasets from each generating architecture was obtained. Finally, BasicVSR and BasicVSR++ were trained using the generated datasets, and the old videos were tested on the BasicVSR and BasicVSR++ trained by C2N, CycleGAN, and NG-GAN, as well as the pretrained BasicVSR.

22

| (a) Input old noisy video frames PIQE: 26.47 | (b) Results on pre-trained BasicVSR PIQE: 20.65 | (c) Results on pre-trained BasicVSR++ PIQE: 19.38 | (d) Results on BasicVSR trained on NG-GAN PIQE: **14.85** | (e) Results on BasicVSR++ trained on NG-GAN PIQE: **15.73** |

| (f) Input old noisy video frames PIQE: 22.09 | (g) Results on pre-trained BasicVSR PIQE: 18.64 | (h) Results on pre-trained BasicVSR++ PIQE: 16.20 | (i) Results on BasicVSR trained on NG-GAN PIQE: **12.58** | (j) Results on BasicVSR++ trained on NG-GAN PIQE: **11.17** |
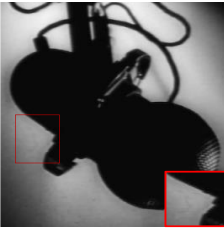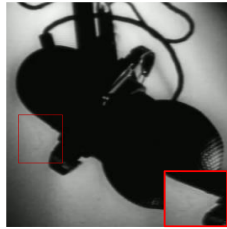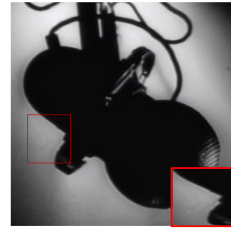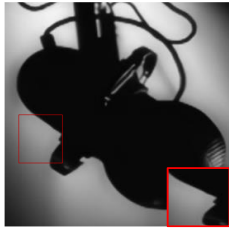
Figure 4-2. Examples of de-oldifying old videos using pretrained SOTA methods and SOTA methods trained on NG-GAN generated dataset. (a) Old video frames, (b) de-oldified output from pretrained BasicVSR, (c) de-oldification output from pretrained BasicVSR++, (d) de-oldified output from BasicVSR trained on the C2N-generated datasets, and (e) de-oldification output from BasicVSR++ trained on NG-GAN-generated datasets.

23

Figures 4-2 (a) and (f) are input images, and Figures 4-2 (b) and (g) are generated noisy images. Figure 4-2 shows that the video restorers trained on our model-generated datasets can produce significantly better-denoised images than those trained on REDS, which include synthetic noise with a Gaussian distribution. As shown in Figure 4-2, BasicVSR and BasicVSR++ trained on datasets generated by the proposed NG-GAN can retain the texture, details, and edges of the given images, whereas the pretrained models show lower-quality results, as shown in Figures 4-2 (b) and (c). This is because the pretrained models were trained using synthetic Gaussian and Poisson noise models, which do not reflect the actual old image noise and artifact patterns. Thus, they fail to capture the noise distribution of the old videos well. The marked region in Figure 4-2 highlights the restored region from the pretrained BasicVSR and BasicVSR++ and BasicVSR and BasicVSR++ trained on the datasets generated by the NG-GAN. The highlighted region in the first row clearly shows the delineation of the ear and neck region, maintaining edges and other structures intact. Notably, the restorers trained on our dataset generated by the NG-GAN can achieve smooth and highly denoised images compared with those pre-trained (Figure 4-2).

4.4. Quantitative Comparison of Denoised Video

In Figure 4-3, experiments were performed to test the results of the denoiser, trained using various datasets, including REDS, the C2N-generated, the CycleGAN-generated, and the proposed NG-GAN-generated datasets. The metrics used to measure the quality of the datasets are the PSNR and SSIM values. The images in Figures 4-3 (b), (h), and (n) show noisy images produced by the proposed NG-GAN model. The results in Figures 4-3 (c), (i), and (o) show the images denoised by the pretrained BasicVSR model. Then, these denoised outputs by the pretrained BasicVSR model were compared with the

outputs denoised by the BasicVSR trained on the datasets generated by the NG-GAN. in terms of the PSNR and SSIM metrics. As shown in the third and the sixth columns, the outputs trained using our NG-GAN-generated datasets show outperforming results in PSNR and SSIM values. Likewise, the outputs trained using the datasets using C2N-generated and CycleGAN-generated datasets show lower PSNR and SSIM values. In addition, the output images trained on the NG-GAN datasets show subjectively better results, as shown in Figures 4-3 (f), (l), and (r). This shows the effectiveness of the NG-GAN-generated dataset.
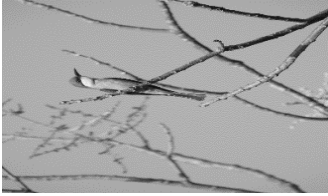
(a) Input clean image

(b) Generated noisy image by NG-GAN
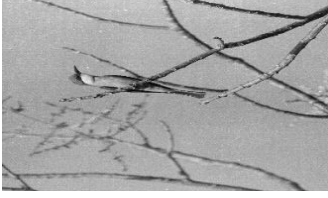
(c) Results on pre-trained BasicVSR PSNR: 28.34 dB SSIM: 0.581

(d) Results on BasicVSR trained on CycleGAN PSNR: 28.601 dB SSIM: 0.785

(e) Results on BasicVSR trained on C2N PSNR: 29.361 dB SSIM: 0.875

(f) Results on BasicVSR trained on NG-GAN PSNR: 29.915 dB SSIM: 0.899

(g) Input clean image

(h) Generated noisy image by NG-GAN

(i) Results on pre-trained BasicVSR PSNR: 27.95 dB SSIM: 0.773

(j) Results on BasicVSR trained on CycleGAN PSNR: 28.15 dB SSIM: 0.783

(k) Results on BasicVSR trained on C2N PSNR: 29.71 dB SSIM: 0.791

(l) Results on BasicVSR trained on NG-GAN PSNR: 30.18 dB SSIM: 0.795

(m) Input clean image   (b) Generated noisy image by NG-GAN   (c) Results on pre-trained BasicVSR PSNR: 25.41 dB SSIM: 0.758   (d) Results on BasicVSR trained on CycleGAN PSNR: 25.44 dB SSIM: 0.782   (e) Results on BasicVSR trained on C2N PSNR: 26.60 dB SSIM: 0.814   (f) Results on BasicVSR trained on NG-GAN PSNR: 26.71 dB SSIM: 0.843
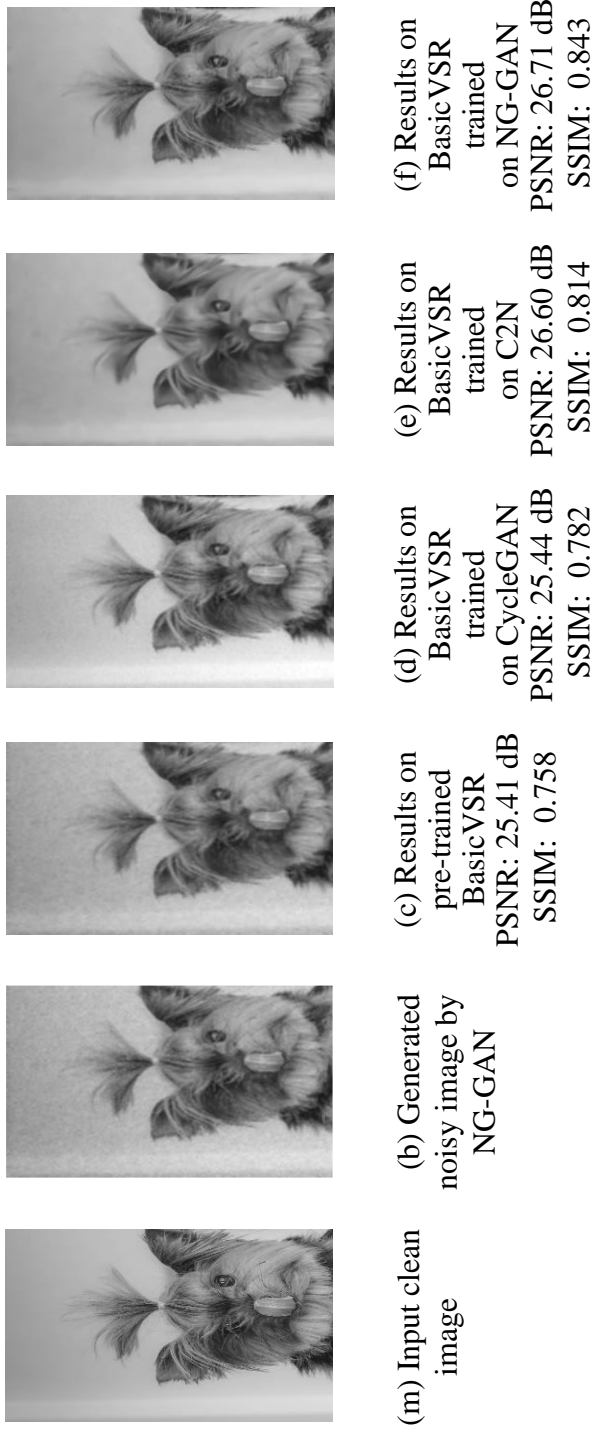
Figure 4-3. Examples of oldifying and de-oldifying videos using pre-trained SOTA methods, SOTA methods trained on C2N and NG-GAN generated dataset. (a), (g) and (m) High quality images form Flickr 2k image dataset, (b), (h) and (n) Distorted noisy frames generated by NG-GAN (c), (i) and (o) De-oldification output from pre-trained BasicVSR, (d), (j) and (p) De-oldification output from BasicVSR++ trained on CycleGAN generated dataset, (e), (k) and (q) De-oldification output from BasicVSR++ trained on C2N generated dataset, (f), (l) and (r) De-oldification output from BasicVSR++ trained on NG-GAN generated dataset.

Table 2. Comparison of old images denoised by state-of-the-art denoisers and image restorers trained on the dataset generated by our model.

Models		PSNR (dB)	SSIM
BasicVSR	Pretrained BasicVSR [32]	24.91	0.703
	BasicVSR (CycleGAN) [18]	24.93	0.698
	BasicVSR (C2N) [48]	25.27	0.736
	BasicVSR (Proposed NG-GAN)	**25.48**	**0.739**
BasicVSR++	Pretrained BasicVSR++ [47]	25.21	0.727
	BasicVSR++ (CycleGAN) [18]	25.03	0.705
	BasicVSR++ (C2N) [48]	25.81	0.768
	BasicVSR++ (Proposed NG- GAN)	**25.89**	**0.781**
Others	GCBD [50]	24.22	0.726
	UIDNet [14]	25.17	0.694

The average performance of the SOTA denoising methods has been evaluated, BasicVSR, BasicVSR++, GCBD, and UIDNet on datasets generated by the proposed NG-GAN, C2N, CycleGAN. BasicVSR and BasicVSR++ are known as the best-performing denoiser among the supervised denoising architectures, and GCBD and UIDNet are the unsupervised denoiser to show the best results. This experiment is to investigate how the generated datasets can train the denoiser well. Table 1 shows PSNR and SSIM values on average for each denoising method when they are trained using various datasets. As shown in Table 2, BasicVSR and BasicVSR++ trained using the NG-GAN generated datasets achieve significantly better PSNR and SSIM values.

28

(a) Input clean
image

(b) Generated noisy
image
PIQE input: 10
PIQE generated: 11.87

(c) Generated noisy
image
PIQE input: 20
PIQE generated: 26.47

(d) Generated noisy
image
PIQE input: 30
PIQE generated: 32.71

(e) Input clean
image

(f) Generated noisy
image
PIQE input: 10
PIQE generated: 12.85

(g) Generated noisy
image
PIQE input: 20
PIQE generated: 24.27

(h) Generated noisy
image
PIQE input: 30
PIQE generated: 37.36

(i) Input clean
image

(j) Generated noisy
image
PIQE input: 10
PIQE generated: 13.52

(k) Generated noisy
image
PIQE input: 20
PIQE generated: 24.48

(l) Generated noisy
image
PIQE input: 30
PIQE generated: 35.85

Figure 4-4. Effectiveness of various PIQE values as input for the generator.

The impact of the PIQE value on the REDS dataset [33] has been investigated
by testing various PIQE values. Figure 4-4 shows that, with a value of 10, the

image shows a controlled amount of noise and distortion in Figures 4-4 (b), (f), and (j). With an increase in the PIQE value, the distortion and noise increased in proportion to the input PIQE value, as shown in Figure 4-4 (c), (g), (k) and Figures 4-4 (d), (h), (l). This is because the NG-GAN was trained using the PIQE value extracted from the old, degraded images. This helps the model to learn noise generation better. The PIQE values are provided to the generator as input with the clean images, resulting in the PIQE values of the generated distorted image.
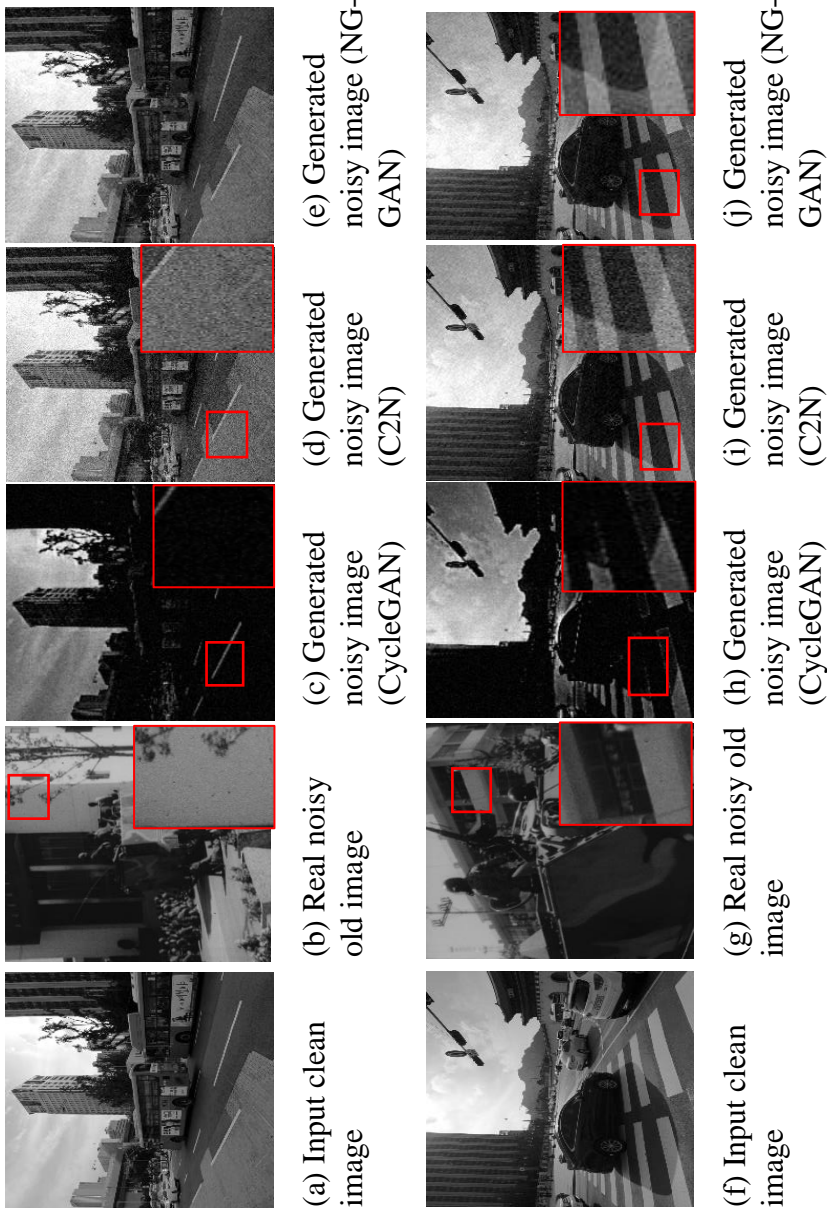
Figure 4-5. Visual comparison of noisy images generated by CycleGAN, C2N, and NG-GAN. (a) and (f) are high-quality images from the REDS dataset, (b) and (g) are noisy images generated by Cy-cleGAN, (c) and (h) are noisy images generated by C2N, (d) and (i) are noisy images generated by the NG-GAN, and (e) and (j) are real noisy images from old videos.

Figure 4-5 shows examples of noisy images generated by CycleGAN, C2N, and the proposed NG-GAN method, respectively. The images in Figures 4-5 (a) and (f) are clean input images, Figures 4-5 (b) and (g) are actual old noisy images, and the images in Figure 4-5 (c), (d), (e), (h), (i) and (j) are the images generated by CycleGAN, C2N, and NG-GAN, respectively. As can be seen in Figure 4-5, the proposed NG-GAN can generate more realistic-looking old image noise, whereas other noise generation networks fail to generate old image noise with actual noisy patterns in the given clean images. The image generated by CycleGAN shows unclear output, and the edges of the objects are not retained well, as shown in the red-marked region in Figure 4-5.

Table 3. Average Kullback-Leibler (KL) divergence values between generated and real noisy images.

Metric	CycleGAN	C2N	NG-GAN (Proposed method)
KL-divergence	0.3436	0.2195	**0.1879**

Table 3 shows the KL-divergence [54] values calculated between the noise map of old images and the noise map of images generated by CycleGAN, C2N, and our proposed method. In general, it evaluates how unique one probability distribution P is from a second reference distribution. Q. Hence, the lower value of KL-divergence indicates a higher similarity between the two populations of images. As shown in Table 3, it is observed that our proposed method achieves the lowest KL-divergence between the actual old noisy images and generated noisy images. The lower KL-divergence indicates that the proposed model successfully generates the old image noise pattern.

An experiment is conducted considering both PSNR and SSIM metrics to measure the superiority of the oldification. It is clear from the plot Figure 4-3, that the generator is able to produce images maintaining acceptable PSNR and SSIM scores. The qualitative results presented in Figure 4-2 and 4-3, supports the fact of oldification.
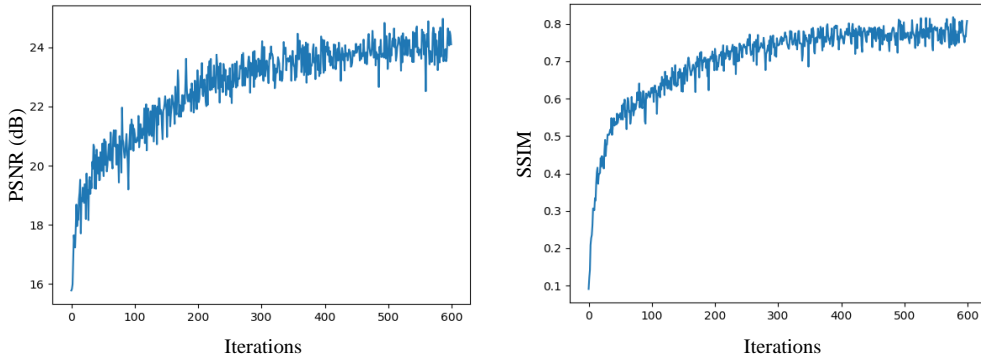
Figure 4-6. PSNR and SSIM over iterations images by the NG-GAN.

5. CONCLUSION

The main goal of this study is to build a model which can effectively produce old image noise with near-perfect noise distribution of ancient images. Since the amount of paired datasets of old images is quite scarce, it is very challenging to denoise such images. Thus, most of the existing studies did not consider solving this problem. Using datasets generated by the propsed model, video restorers can learn to denoise old degraded images better. It is shown that BasicVSR and BasicVSR++ can achieve a higher PSNR and SSIM value compared to the pretrained models. This approach can be a key solution to successfully imitate crucial degraded noise patterns for generating accurate noise from clean images.

BIBLIOGRAPHY

[1] Antoni Buades, Bartomeu Coll, and J-M Morel. A non-local algorithm for image denoising. CVPR, 2005.

[2] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian. Image denoising by sparse 3-d transformdomain collaborative filtering. TIP, 16(8):2080–2095, 2007.

[3] Shuhang Gu, Lei Zhang, Wangmeng Zuo, and Xiangchu Feng. Weighted nuclear norm minimization with application to image denoising. CVPR, 2014.

[4] Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. TIP, 26(7):3142–3155, 2017.

[5] Kai Zhang, Wangmeng Zuo, and Lei Zhang. FFDnet: Toward a fast and flexible solution for cnn-based image denoising. TIP, 27(9):4608–4622, 2018.

[6] Shi Guo, Zifei Yan, Kai Zhang, Wangmeng Zuo, and Lei Zhang. Toward convolutional blind denoising of real photographs. CVPR, 2019.

[7] Chang Chen, Zhiwei Xiong, Xinmei Tian, and Feng Wu. Deep boosting for image denoising. In ECCV, 2018.

[8] Yang Liu, Saeed Anwar, Liang Zheng, and Qi Tian. GradNet image denoising. In CVPRW, 2020.

[9] Sungmin Cha, Taeeon Park, Byeongjoon Kim, Jongduk Baek, and Taesup Moon. GAN2GAN: Generative noise learning for blind denoising with single noisy images. ICLR, 2021.

[10] Alexander Krull, Tim-Oliver Buchholz, and Florian Jug. Noise2Void-learning denoising from single noisy images. CVPR, 2019.

[11] Abdelrahman Abdelhamed, Stephen Lin, and Michael S Brown. A high-quality denoising dataset for smartphone cameras. CVPR, 2018.

[12] Tobias Plotz and Stefan Roth. Benchmarking denoising algorithms with real photographs. CVPR, 2017.

[13] Jingwen Chen, Jiawei Chen, Hongyang Chao, and Ming Yang. Image blind denoising with generative adversarial network-based noise modeling. CVPR, 2018.

[14] Zhiwei Hong, Xiaocheng Fan, Tao Jiang, and Jianxing Feng. End-to-end unpaired image denoising with conditional adversarial networks. AAAI, 2020.

[15]   Abdelrahman Abdelhamed, Marcus A Brubaker, and Michael S Brown. Noise flow: Noise modeling with conditional normalizing flows. ICCV, 2019.

[16]   Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. NeurIPS, pages 2672–2680, 2014.

[17]   Zili Yi and Hao Zhang and Ping Tan and Minglun Gong. DualGAN: Unsupervised Dual Learning for Image-to-Image Translation. ICCV, 2017.

[18]   Jun-Yan Zhu and Taesung Park and Phillip Isola and Alexei A. Efros. Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks. ICCV, 2017.

[19]   Phillip Isola and Jun-Yan Zhu and Tinghui Zhou and Alexei A. Efros. Image-to-Image Translation with Conditional Adversarial Networks. CVPR, 2017.

[20]   Zongsheng Yue, Qian Zhao, Lei Zhang, and Deyu Meng. Dual adversarial network: Toward real-world noise removal and noise generation. ECCV, 2020.

[21]   Rui Zhao, Daniel PK Lun, and Kin-Man Lam. NTGAN: Learning blind image denoising without clean reference. BMVC, 2020.

[22]   Kaixuan Wei, Ying Fu, Jiaolong Yang, and Hua Huang. A physics-based noise formation model for extreme low-light raw denoising. CVPR, 2020.

[23]   Gerald C Holst. Ccd arrays, cameras, and displays. 1998.

[24]   Venkatanath N, Praneeth D, Maruthi Chandrasekhar Bh, S. S. Channappayya and S. S. Medasani, "Blind image quality evaluation using perception basedperception-based features," 2015 Twenty First National Conference on Communications (NCC), 2015.

[25]   Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P. Image Quality Assessment: from Error Visibility to Structural Similarity. IEEE Transactions on Image Processing 13, 4 (2004), 600–612.

[26]   Alom, Md. Zahangir & Hasan, Mahmudul & Yakopcic, Chris & Taha, Tarek & Asari, Vijayan. (2018). Recurrent Residual Convolutional Neural Network based on U-Net (R2U-Net) for Medical Image Segmentation.

[27]   K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. International Conference on Learning Representations (ICLR), 2015.

[28]   H. Zhao, O. Gallo, I. Frosio and J. Kautz, "Loss Functions for Image Restoration With Neural Networks," in IEEE Transactions on Computational Imaging, vol. 3, no. 1, pp. 47-57, March 2017.

[29]   Sanghyun Woo, Jongchan Park, Joon-Young Lee, and In So Kweon, CBAM: Convolutional Block Attention Module. ECCV, 2018.

[30]   S. Ioffe and C. Szegedy, "Batch normalization: accelerating deep network training by reducing internal covariate shift," in Proceedings of the 32nd International Conference on Machine Learning, pp. 448–456, Lille, France, July 2015.

[31]   Jolicoeur-Martineau, "The relativistic discriminator: a key element missing from standard GAN", in International Conference on Learning Representation, 2019.

[32]   K. C. K. Chan, X. Wang, K. Yu, C. Dong, and C. C. Loy, "BasicVSR: The Search for Essential Components in Video Super-Resolution and Beyond." arXiv, 2020. doi: 10.48550/ARXIV.2012.02181.

[33]   S. Nah et al., "NTIRE 2019 Challenge on Video Deblurring and Super-Resolution: Dataset and Study," 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). IEEE, Jun. 2019.

[34]   H. C. Burger, C. J. Schuler, and S. Harmeling, "Image denoising: Can plain neural networks compete with bm3d?," in CVPR, 2012.

[35]   S. Lefkimmiatis, "Non-local color image denoising with convolutional neural networks," in CVPR, 2017.

[36]   Zhang, K.; Zuo, W.; Gu, S.; Zhang, L. Learning Deep CNN Denoiser Prior for Image Restoration. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21-26 July 2017.

[37]   Zhang, Y.; Tian, Y.; Kong, Y.; Zhong, B.; Fu, Y. Residual Dense Network for Image Restoration. IEEE Trans Pattern Anal Mach Intell 2021, 43(7), pp. 2480-2495.

[38]   Y. Cai, Z. Wang, Z. Luo, B. Yin, A. Du, H. Wang, X. Zhou, E. Zhou, X. Zhang, and J. Sun, "Learning delicate local representations for multi-person pose estimation," in ECCV, 2020.

[39]   Z. Luo, Z. Wang, Y. Cai, G. Wang, L. Wang, Y. Huang, E. Zhou, T. Tan, and J. Sun, "Efficient human pose estimation by learning deeply aggregated representations," in ICME, 2021.

[40] Y. Cai, J. Lin, X. Hu, H. Wang, X. Yuan, Y. Zhang, R. Timofte, and L. V. Gool, "Mask-guided spectral-wise transformer for efficient hyperspectral image reconstruction," in CVPR, 2022.

[41] X. Gong, S. Chang, Y. Jiang, and Z. Wang, "Autogan: Neural architecture search for generative adversarial networks," in ICCV, 2019.

[42] X. Hu, H. Wang, Y. Cai, X. Zhao, and Y. Zhang, "Pyramid orthogonal attention network based on dual self-similarity for accurate mr image super-resolution," in ICME, 2021.

[43] C. Zheng, T.-J. Cham, and J. Cai, "Pluralistic image completion," in CVPR, 2019.

[44] Li, C.; Wand, M. Combining Markov random fields and convolutional neural networks for image synthesis. In 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2016, pp. 2479-2486. doi:10.1109/CVPR.2016.272.

[45] Niu Z, Zhong G, Yu H. A Review on the Attention Mechanism of Deep Learning. Neurocomputing (2021) 452:482021 452:48–62. doi:10.1016/j.neucomdoi: 10.1016/j.neucom.2021.03.091.

[46] Wang, Q.; Wu, B.; Zhu, P.; Li, P.; Zuo, W.; Hu, Q. ECA-Net: Efficient channel attention for deep convolutional neural net-works. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Seattle, Washington, IEEE/ CVF Conference on Computer Vision and Pattern Recognition (CVPR) IEEE (2020). pp. 11531–9.

[47] K. C. K. Chan, S. Zhou, X. Xu, and C. C. Loy, "BasicVSR++: Improving Video Super-Resolution with Enhanced Propagation and Alignment." arXiv, 2021. doi: 10.48550/ARXIV.2104.13371.

[48] Jang, G.; Lee, W.; Son, S.; Lee, K. C2N: Practical generative noise modeling for real-world denoising. In Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Virtual, 11-17 October 2021.

[49] S. Nah et al., "NTIRE 2019 Challenge on Video Deblurring and Super-Resolution: Dataset and Study," 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2019, pp. 1996-2005, doi: 10.1109/CVPRW.2019.00251.

[50] J. Chen, J. Chen, H. Chao and M. Yang, "Image Blind Denoising with Generative Adversarial Network Based Noise Modeling," 2018

IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 3155-3164, doi: 10.1109/CVPR.2018.00333.

[51]   M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The Pascal Visual Object Classes (VOC) Challenge," International Journal of Computer Vision, vol. 88, no. 2. Springer Science and Business Media LLC, pp. 303–338, Sep. 09, 2009. doi: 10.1007/s11263-009-0275-4.

[52]   E. Agustsson and R. Timofte, "NTIRE 2017 Challenge on Single Image Super-Resolution: Dataset and Study," 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2017, pp. 1122-1131, doi: 10.1109/CVPRW.2017.150.

[53]   Z. Zhang, Q. Liu, and Y. Wang, "Road Extraction by Deep Residual U-Net," arXiv, 2017, doi: 10.48550/ARXIV.1711.10684.

[54]   Q. Zuo, S. Chen, and Z. Wang, "R2AU-Net: Attention Recurrent Residual Convolutional Neural Network for Multimodal Medical Image Segmentation," Security and Communication Networks, vol. 2021. Hindawi Limited, pp. 1–10, Jun. 10, 2021. doi: 10.1155/2021/6625688.

Acknowledgment

This thesis presents the conclusion of my master's degree at the department of Information and communication, Chosun University. It has been an amazing journey, and I had a chance to learn much during this challenging period. My task has been highly exciting to deal with, however the work itself was sometimes hard and challenging, the numerous individuals who helped during this period are much helpful and supportive.

At first, I would like to express my earnest gratitude to my advisor, for his encouragement, guidance and valuable knowledge shared throughout the development of this thesis. The exposure I have gained during my research would be a valuable treasure of my life.

I extend my appreciation and thanks to all my professors, seniors, lab members, friends, and family, who have endured my partial absence, yet have been there for me.

Finally, I would like to thank REDS and its collaborators for their great efforts, large amounts of work and willingness to share their data, without which this thesis and the original work described herein would not be possible.