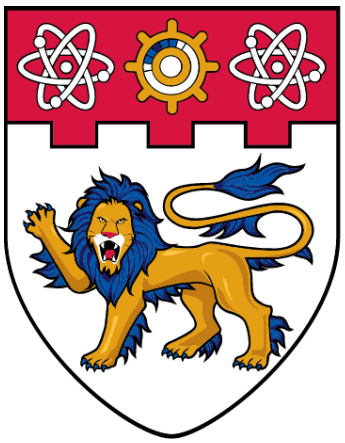
Automated image generation

Goh, Shan Ying

2023



NANYANG TECHNOLOGICAL UNIVERSITY

SINGAPORE

SCSE22-0656

Automated image generation

Submitted in Partial Fulfilment of the Requirements

for the Degree of Bachelor of Engineering (Computer Science)

of the Nanyang Technological University

by

Goh Shan Ying (U1921497C)

Name of Supervisor : A/P Lu Shijian

School of Computer Science and Engineering

2023

Abstract

Image translation techniques have gained significant attention in recent years, particularly CycleGAN. Traditionally, building image-to-image translation models requires the collection of extensive datasets with paired examples, which can be complicated and costly. However, CycleGAN's automatic training approach eliminates the need for such paired samples, thus simplifying the training process while enhancing the potential of image translation, allowing for imaginative and lifelike adjustments. For instance, CycleGAN can effortlessly transform styles like cats into dogs and vice versa, extending to practical domains like art, fashion, and medical imaging.

Nevertheless, CycleGAN's applicability in real-world scenarios is limited by its current constraint to a relatively small set of available styles. This compels us to explore more practical alternatives. This study introduces new styles into the framework, assessing their practical effectiveness and addressing concerns about potential loss in image quality. Results show the promising potential of these improved CycleGAN variants for various domains and applications.

Keywords: CycleGAN, Style Transfer, Image Translation, Diverse Aesthetics, Creative Applications, Content Preservation

Acknowledgment

I want to express my gratitude to Dr. Lu Shi Jian and Ph.D. student Zhang Jiahui for their invaluable support and unwavering guidance throughout my project. Their expertise, patience, and dedication have significantly influenced my research and deepened my understanding of the subject matter.

Furthermore, I'd like to express my gratitude to the developers and maintainers of the open-source libraries and tools that played a pivotal role in the successful execution of this project. Their dedication to providing accessible resources has been instrumental in achieving the project's goals.

I'm also deeply thankful to all those who offered their support in various ways during the completion of this project.

Tables Of Contents

List of Figures

List of Tables

1. Introduction

1.1 Background

Image translation is gaining momentum in the world of computational techniques, ushering in a new era of creative versatility. This technique allows us to morph images from one style or context to another while preserving their core content and structure [1]. Imagine having two images showcasing vibrant cityscapes with buildings and trees and another featuring black and white wildlife snapshots. Image translation involves training a computer algorithm to change images from the first set into images that embody the style and characteristics of the second set. For instance, it can transform colorful urban scenes into grayscale images with animals while retaining the distinctive elements that define each scene.

During image translation, we typically need two images for each transformation. One is the starting point, while the other represents the desired outcome. This pair of images is known as "paired data", similar to having a "before" picture and an "after" picture.

However, gathering and preparing these pairs of images can be intricate and costly.

1.2 Motivation

This is where CycleGAN, a generative adversarial network (GAN), enters the stage, capturing substantial attention. Unlike traditional image translation approaches relying on paired images, CycleGAN thrives on "unpaired data." Simply put, it learns to change styles without necessitating exact pairs of matching images.

CycleGAN's ability to achieve this without paired data has propelled its popularity. In image processing and manipulation, CycleGAN has emerged as a transformative force. It adeptly transitions styles across different types of images [1]. This significance extends beyond artistic endeavors, finding practical use in real-world scenarios [2]. CycleGAN's capacity to

independently learn connections between image domains has prompted its application in fields like art, medical imaging, and video synthesis [3], fundamentally reshaping our perception and utilization of images.

For example, it can turn summer landscapes into winter landscapes and vice versa, showing its power [1].

However, CycleGAN's potential is hindered by limited style diversity and constraining adaptation to artistic, design, and cultural variations. By broadening its style spectrum, we enable adaptability that enhances image adaptation and data augmentation and addresses data challenges by facilitating style transfer without strict pairings [4]. This advancement significantly benefits tasks ranging from classification to object detection, offering a more vibrant and versatile toolkit for real-world challenges.

1.3 Objective

This study aims to improve CycleGAN's abilities and broaden its real-world uses by introducing a more comprehensive range of artistic styles for transforming images. This will increase CycleGAN's adaptability across different industries and inspire creativity. Our main objective is to strike the right balance between CycleGAN's existing strengths and the integration of new artistic styles. This effort continues to advance the image alteration field, providing various visual interpretations.

1.4 Scope

This research project will primarily focus on the challenging task of translating Thermal Infrared (IR) images into RGB (color) images, particularly utilizing "white hot" representations, where 'white hot' displays warmer objects in white and cooler objects in black [10]. The translation of RGB to Thermal IR images, including "white hot" styles, has

significant real-world applications, ranging from military and surveillance use to night vision technology, industrial inspections, and wildlife protection [11-12][18].

Figure 1: a sample "white hot" thermal infrared (IR) image designed for finding and observing wildlife.

This scope is driven by the compelling need to harness the benefits of Thermal IR imaging for a broader range of practical scenarios.

Thermal IR imaging has gained prominence due to its ability to capture temperature differences and provide visibility in conditions where visible light is limited, such as darkness or through smoke and fog [12]. Notably, there have been few successful integrations of RGB to Thermal IR image style transfer, including "white hot" representations, into CycleGAN or similar frameworks. Therefore, our research represents a pioneering effort in exploring this new field, including translating RGB images to "white hot" thermal IR representations.

1.5 Limitations

Throughout the research, significant limitations have impacted the study's scope and outcomes.

The research faced limitations in computational resources. The absence of a dedicated Graphics Processing Unit (GPU) on the local laptop and limited access to school-provided GPU resources for large-scale image training constrained the efficiency and speed of training, potentially impacting the model's performance.

Another significant challenge in this study pertained to the data used. Dealing with thermal IR data, which is primarily sensitive, especially when seeking publicly available datasets for training, posed a considerable hurdle. Unfortunately, the availability of thermal IR data in terms of quantity and quality is quite limited. The dataset procured mainly consisted of small and noisy images, making it less representative of the intended domains. Additionally, locating thermal IR images proved to be an additional challenge, as such data was scarce and not easily accessible.

In summary, our research faced limitations in computational resources and access to high-quality thermal IR data. Nevertheless, we strived to extract valuable insights to enhance the understanding of thermal IR transformations with CycleGAN.

1.6 Source of Data

Our data source for this study was the VAP Trimodal People Segmentation Dataset, obtained from the Visual Analysis and Perception Lab (VAP Lab) [13]. This dataset comprises paired thermal IR with a "white hot" thermal style.

1.7 Organization of the Report

This section gives an overview of the following chapters:

Chapter 1 provides an overview of the project and its objectives.

Chapter 2 reviews relevant research on introducing new styles on top of CycleGAN.

Chapter 3 discusses the various phases and tasks completed within the project timeline.

Chapter 4 details the resources used, including material/equipment resources and costs.

Chapter 5 introduces the methodology and process of fine-tuning.

Chapter 6 summarizes the main contributions and critical findings of the project.

Chapter 7 discusses potential areas for further improvement in the project.

Chapter 8 discusses potential areas for further improvement in the project.

2. Related Works

This section introduces CycleGAN's current capabilities, highlighting its capacity to accommodate various styles. We will also explore the impact of GANs like StyleGAN on enhancing CycleGAN's effectiveness and expanding its range of styles.

While our project scope is on RGB ↔ Thermal IR style transfers, it's essential to acknowledge the broader landscape of unexplored stylistic transformations. In this literature review, we not only delve into the existing research on RGB to thermal IR style transfers but also shed light on various other styles that have yet to receive comprehensive attention within the realm of style transfer.

2.1 CycleGAN's Current Stylistic Capabilities

The styles that CycleGAN currently supports encompass a wide range of artistic genres and visual contexts [1]. It can translate artworks evoking Van Gogh's brushwork into images resembling the photographic realism of diverse styles. Apart from paintings, it transitions between photographic compositions and bold canvas-like strokes. Furthermore, CycleGAN effortlessly shifts between landscapes, morphing summer scenes into winter panoramas and vice versa.

Similarly, it can proficiently translate satellite images to maps or vice versa, proving valuable for cartography and geographic analysis. This adaptability extends to common subjects as well. For instance, it can transform images of horses to take on the appearance of zebras.

These illustrative instances collectively emphasize CycleGAN's capacity to redefine and traverse an array of distinctive styles [1].

2.2 StyleGAN

StyleGAN is renowned for its impressive prowess in crafting a wide array of high-quality images with remarkable diversity. While it shares its fundamental GAN architecture with CycleGAN, StyleGAN introduces several key innovations that elevate its performance.

One of the most notable enhancements is its ability to generate images with unparalleled realism and intricate details by disentangling the latent space into different levels of control over features such as facial expressions, hairstyles, and lighting conditions [5].

Unlike CycleGAN, which primarily focuses on domain-to-domain translation, StyleGAN excels at creating new images that exhibit a wide range of styles, allowing users to manipulate the generated content's global and local features of the generated content [6]. This capacity to generate highly customizable and diverse imagery has made StyleGAN a pivotal tool in various creative applications, including art, fashion, and design [3].

2.3 Sketch-and-Paint GAN (SAPGAN)

SAPGAN is a cutting-edge tool designed to recreate the intricate styles found in traditional Chinese landscape paintings [7]. Its innovation stems from training on an untapped dataset of these classical artworks, enabling it to intricately mimic the unique visual elements that define this revered genre.

SAPGAN can adapt various artistic styles, resulting in a diverse array of visually captivating landscapes that honor the classical aesthetics of traditional Chinese art [7]. This remarkable skill has practical implications across fields such as museums and cultural preservation. By harnessing SAPGAN's capabilities, museums can rejuvenate and reimagine historical artworks, offering modern audiences fresh perspectives.

This technology infuses new vitality into aged masterpieces, making them relevant to contemporary times while upholding their timeless appeal. Beyond museums, SAPGAN empowers artists to engage with and reinterpret traditional techniques, driving artistic discourse forward while preserving a solid link to cultural heritage.

2.4 Enhanced Styles in Other Projects

In addition to the advancements brought forth by StyleGAN and SAPGAN, there exist other noteworthy GAN architectures that have leveraged the foundational principles of CycleGAN to introduce novel stylistic dimensions. Notably, StarGAN and Multimodal Unsupervised Image-to-image Translation (MUNIT) GAN have emerged as prominent examples, each contributing unique approaches to expanding the artistic horizon.

StarGAN redefines the landscape by enabling a single model to generate images across multiple domains [8]. For instance, given a dataset containing images of felines and canines, StarGAN can generate images depicting diverse dog breeds or even entirely novel animal species [8].

On the other hand, MUNIT introduces disentangled content and style representations [9]. This means you could combine the brushstrokes of famous painters with real-life photos or merge different artistic ideas into a single piece. With MUNIT GAN, artists have a fresh way of blending styles and creating something new.

2.5 Styles Yet to Be Explored

In addition to the artistic effects achievable through CycleGAN and its variations, numerous other styles remain unexplored, each holding considerable potential.

Here are some examples:

- Atmospheric Enhancement: Refine mood and lighting in images, intensifying their atmospheric impact and eliciting heightened emotions.

- Colorization: Leveraging CycleGAN to introduce vivid hues to grayscale images sparks new dimensions of creative expression, enabling novel visual narratives.
- Diverse Stylistic Adaptation: Beyond emulating Van Gogh's style, CycleGAN's potential to adopt various artistic approaches, like Fauvism, Pointillism, and more, enriches creative experimentation.

In summary, there are numerous unexplored possibilities for artistic expression. The mentioned untapped art styles have the potential to enhance practical applications and ignite fresh creative concepts significantly.

2.6 Thermal IR

Thermal imaging has proven a valuable tool in various applications due to its unique discriminative properties. Notably, thermal images excel in highlighting warm bodies, including humans, animals, and hot vehicles, making them compelling subjects of interest for a wide range of surveillance and detection tasks. However, the great adoption of thermal cameras in certain applications is hindered by the aid of a set of inherent boundaries. Firstly, the price associated with the thermal digital camera era remains surprisingly excessive, restricting its accessibility for many users. Additionally, thermal cameras tend to possess a narrow field of view, which may constrain their applicability in scenarios requiring comprehensive situational awareness.

Moreover, thermal cameras are unable to fulfill all perception tasks; for instance, they are unable to correctly read traffic lights, an essential component of real-world traffic monitoring and control. Consequently, in many practical applications, integrating RGB cameras alongside thermal imaging technology is essential to achieve comprehensive perception tasks, particularly in the realm of semantic segmentation and environmental analysis. This combination allows for a more cost-effective and versatile approach, capitalizing on the strengths of thermal imaging while compensating for its limitations. In this manner, a fusion of thermal and RGB cameras presents a promising avenue for enhancing the robustness and applicability of perception systems in various domains.

2.6.1 Attempts to Introduce RGB ↔ Thermal IR Style

Numerous prior attempts have been made to introduce Thermal IR style transfers, with varying degrees of success and limitations. Among these endeavors, ThermalGAN has emerged as a noteworthy achievement.

ThermalGAN is a GAN framework built on top of CycleGAN that excels in color-to-thermal image translation for the purpose of person re-identification in multispectral datasets [19]. ThermalGAN has successfully bridged the gap between color and thermal imaging, particularly benefiting person re-identification (REID) in surveillance and security applications [19].

In Figure 2, we witness a visual representation of ThermalGAN's successful transformation of RGB images into thermal IR counterparts.

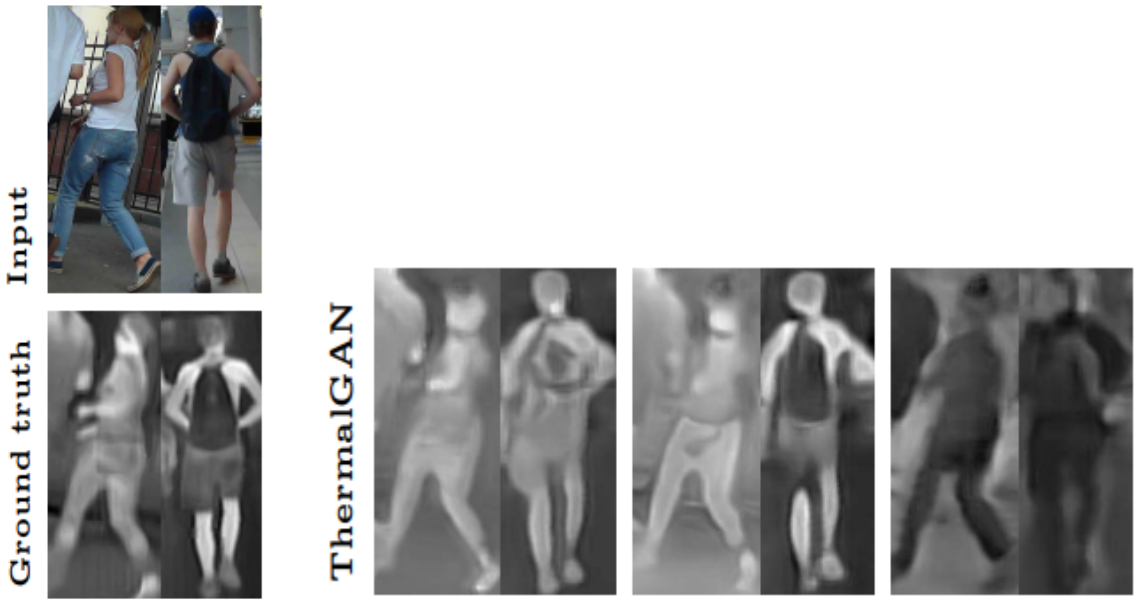


Figure 2: the successful conversion of RGB images into thermal IR images by ThermalGAN [19]

Going beyond style translation, ThermalGAN offers a wide array of functionalities, including pedestrian tracking across multiple camera streams, enabling efficient detection and monitoring of people across interconnected camera feeds. These capabilities empower a more comprehensive analysis of dynamic scenarios. One key contributor to ThermalGAN's achievement is its methodical use of annotated paired pictures during training, extensively boosting its proficiency in photo transformation and analysis responsibilities.

Other attempts have been made to introduce thermal style transfers for various applications, such as cars and humans. An illustrative example of such an attempt can be found in a project by Liik [20], where CycleGAN was employed for thermal image generation from RGB inputs. However, the outcomes of this attempt, as depicted in Figures 3 and 4, differ significantly from the success achieved by ThermalGAN.

Figure 3: gibberish results generated

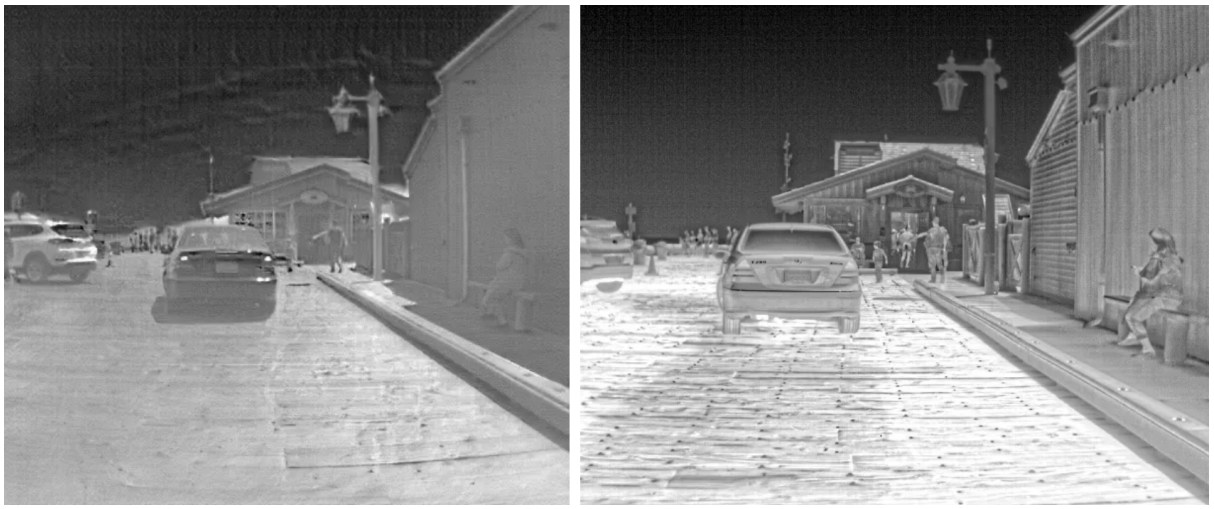


Figure 4: CycleGAN Prediction vs. Ground Truth (Not Aligned)

CycleGAN's behavior, as observed in Figure 3, resulted in less than-useful outputs that could be described as generating gibberish. This example underscores the challenges and limitations encountered in pursuing effective thermal style transfers in diverse applications.

3. Project Schedule

Table 1 presented a structured project timeline designed to align with the project's specific scope and limitations, detailing each stage and its allotted timeframes.

Table 1 Structured project timeline

Stages	Tasks	Time Allocation
Research and planning	• Investigate existing CycleGAN implementations and best practices for introducing new styles. • Identify the necessary software tools, frameworks, and datasets required for style transfer tasks.	2 months
Access GPU and Software Setup	• Apply for GPU access and set up the development environment, such as Google Colab. • Install and configure the necessary software, such as PyTorch for deep learning.	2 months
Dataset Finding and Preparation	• Conduct brainstorming sessions to assess the suitability of the data for your style transfer project and make any necessary adjustments. • Search for and evaluate datasets that align with the transfer objectives. Consider factors such as data quality, size, and suitability. • Preprocess the selected datasets	1 month
Style Implementation	• Develop Python scripts for implementing and fine-tuning new styles in CycleGAN. • Refine and optimize the models based on style transfer results.	3 months

Documentation and Reporting	● Create comprehensive documentation detailing the implemented styles, training processes, and results. ● Prepare a final project report summarizing the project's objectives, methodologies, and outcomes. ● Ensure that all project materials, including the dataset details and evaluation findings, are well-documented for reference.	1-month
Final adjustments	● Implement any final changes or optimizations.	1-month

4. Resource Utilization

This section provides an overview of the hardware and software resources used in the project, including GPU access, memory allocation, and programming language choices.

4.1 Hardware Resources

GPU Access: The project relied on GPU resources for deep learning tasks, accessed through Google Colab, and remote access to NTU (Nanyang Technological University) GPUs. Specifically, A100 and V100 GPUs were utilized based on availability and performance needs.

Memory: A minimum of 25 GB of memory was allocated to handle the VAP Trimodal People Segmentation dataset and model training, ensuring adequate capacity for large datasets and deep learning models.

4.2 Software Resources

Python3 was the primary programming language chosen for its robust library and framework support.

Google Colab, a cloud-based platform with free GPU access, was extensively utilized for code execution and deep learning experiments.

Jupyter Notebooks was the interactive coding environment for parameter configuration and algorithm testing.

5. Methodology

In this section, we will first introduce how CycleGAN works to provide a better understanding of the experimental results. Subsequently, we will delve into integrating the new style transfer technique into the capabilities of the CycleGAN model. This integration was meticulously orchestrated through a carefully designed series of steps, each customized to tackle the specific challenges encountered in our initial experiments.

5.1 CycleGAN: Generator and Discriminator

This subsection offers a quick introduction to how CycleGAN works by summarizing the roles of its generator and discriminator components [1][16].

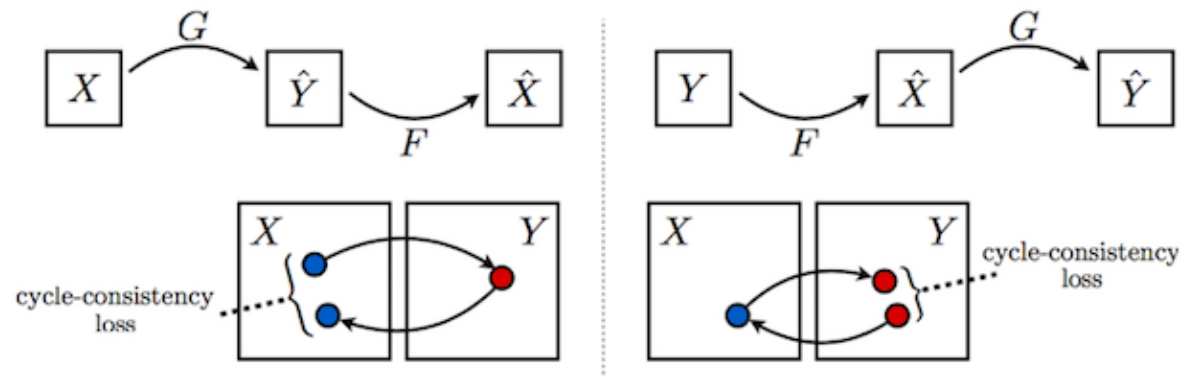


Figure 5: CycleGAN cycle consistency [1]

Generator: The Generator is like an artist. It creates data that resembles a certain distribution. Think of it as an artist creating realistic-looking imitation art.

Discriminator: The Discriminator is like an art critic. It's responsible for telling real data from the original dataset apart from fake data generated by the Generator. Imagine it as an expert art critic carefully evaluating artworks for authenticity.

In CycleGAN, there are two pairs of Generators and Discriminators. One pair transforms data from domain A to domain B, and the other does the opposite, from domain B to domain A.

These Generators and Discriminators compete to generate and evaluate data for their respective domains.

To improve the model's performance, CycleGAN uses "cycle consistency" [1]. This means that when data is transformed from one domain to another and back again, the resulting data should closely match the original data [1]. This ensures that the transformations between domains are meaningful and coherent.

5.2 Data Preprocessing

The VAP Trimodal People Segmentation Dataset comprises around 12,000 paired images. It's important to note that this dataset is characterized by the presence of three distinct scenes [13], each identified as follows:

- Meeting room full depth
- Meeting room constrained depth
- Canteen

In the initial stages of experimentation, I made the decision to utilize a combination of these scenes within the dataset. This choice was made to explore the model's potential to handle a mixture of scenes, thereby increasing its versatility and applicability.

To ensure the compatibility of the dataset with the model's architecture, a series of preprocessing steps were meticulously implemented:

- Resize: Given that the generator architecture in CycleGAN involves a series of downsampling and upsampling operations, it's essential to ensure that the size of the input and output images align. To achieve this, all images underwent meticulous resizing to a uniform dimension. Specifically, the command --resize_or_crop none --loadSize 480 was employed, ensuring that all training images were resized consistently [14]. This consistency was vital for seamlessly integrating images with

varying dimensions into the network architecture.

- Data Splitting: The dataset was effectively shuffled, and the data was thoughtfully divided into training, validation, and testing sets. This division followed the commonly adopted split ratio of 70% for training, 20% for validation, and 10% for testing [15]. Such partitioning is essential to facilitate model training, validation, and evaluation with rigor and statistical significance.

- Domain-Specific Folders: To streamline the training process, separate directories were meticulously created for domains A and B. In this case, domain A represented RGB images, while domain B encompassed thermal IR images. The division into train, test, and validation sets was consistently maintained for both domains, ensuring that each set contained the necessary images for model training, evaluation, and validation.

5.2.1 Initial Experimental Results

Following meticulous data preparation, the model was trained using 12,000 paired images. However, the results fell short of expectations due to various challenges.

One of the primary concerns with the generated images was the presence of blurriness and a failure to capture the "white hot" details accurately. Visual inspection was carried out at different intervals during training, and it was observed that no noticeable improvements were observed in image quality after approximately 60 to 80 epochs.

In Figures 6 and 7, a side-by-side comparison illustrates the differences between real images (left) and generated fake images (right). This visual representation highlights key disparities in image quality and can be used to assess the performance of the image generation model. Both real and generated images appear very blurred and fail to capture the details of the items in both scenes.

Notably, the images did not exhibit any improvement beyond the 80-epoch mark, and they all appear the same with no noticeable enhancements in terms of sharpness, details, and color accuracy.

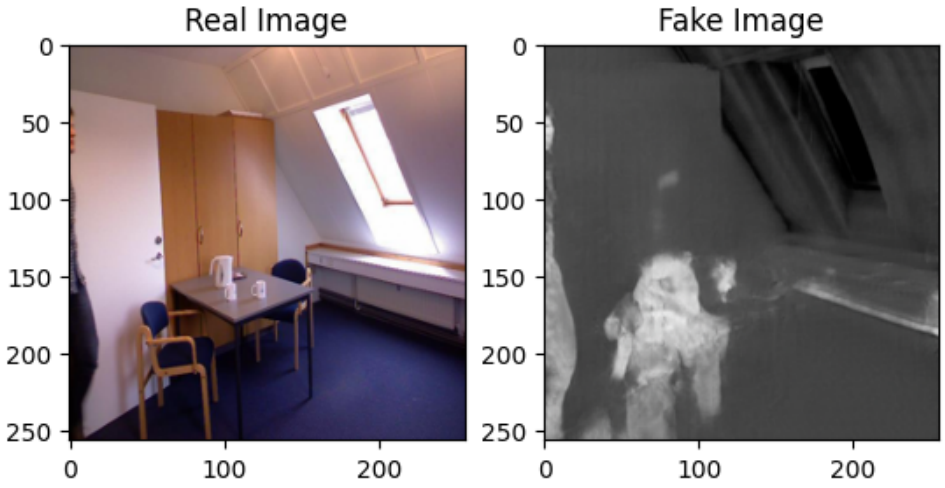


Figure 6: Comparison of Real and Generated Thermal IR Images at 80 Epoch

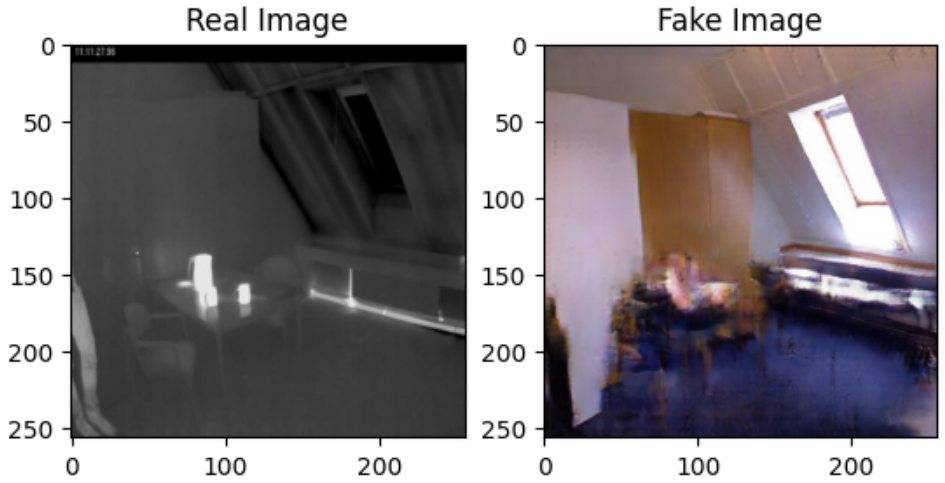


Figure 7: Comparison of Real Thermal IR Image and Generated RGB Images at 80 Epoch

To gain deeper insights into the factors contributing to the suboptimal training outcomes, we investigate the loss curves of the initial training. Figure 8 displays the generator and discriminator losses for both transfers from RGB to Thermal IR (denoted by A_Loss_curves) and Thermal IR to RGB (denoted by B_Loss_curves).

According to the authors, discriminator (D) and generator (G) loss curves exhibit oscillations. In an ideal scenario, the D loss should approach 0, while the optimal range for G loss depends on the application and data type. Figure 8 demonstrates that the loss curves follow a typical pattern.

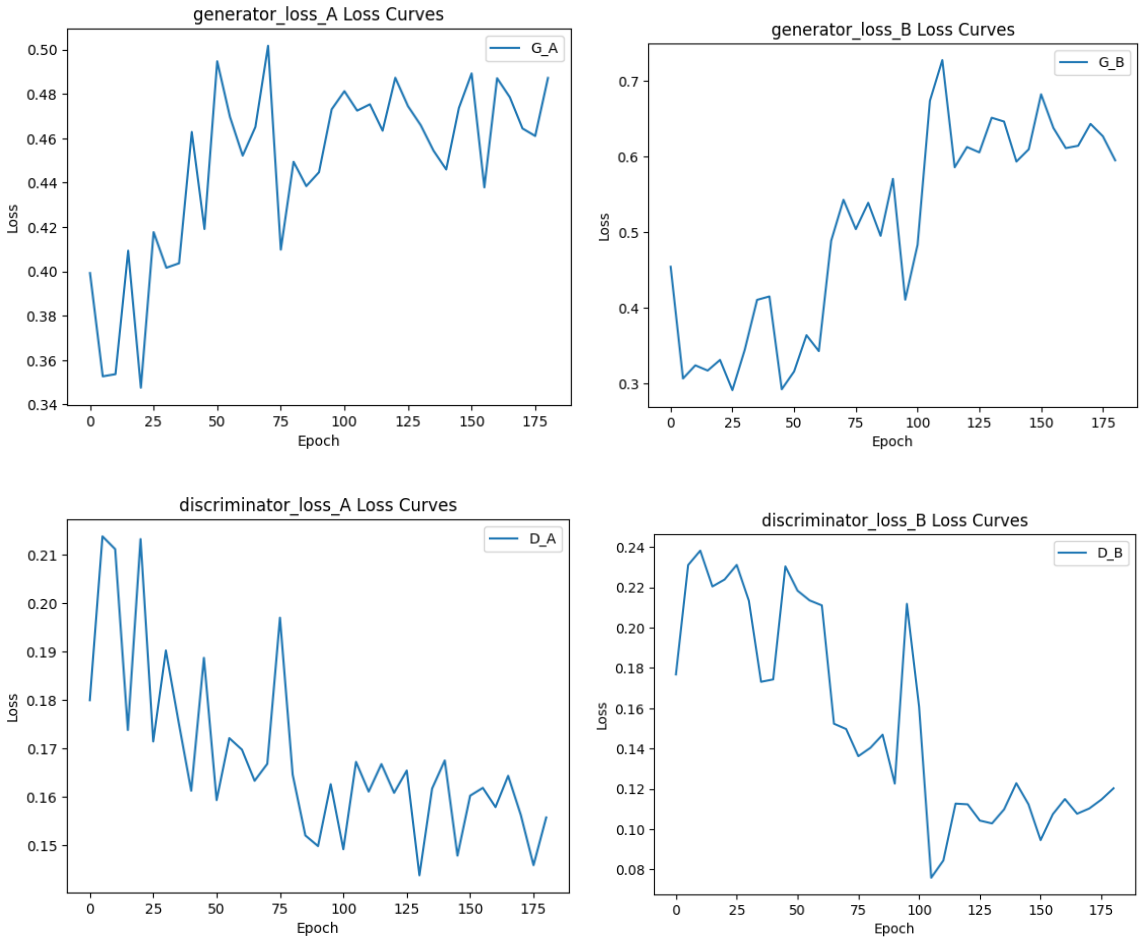


Figure 8 Loss curves for initial training

Since the loss curve doesn't reveal any abnormal patterns, we employed SSIM, a widely recognized metric for image quality assessment [22], to determine whether training exhibited improvement beyond 80 epochs.

Table 2 presents the SSIM scores at 10-epoch intervals from 60 to 150, highlighting that after 80 epochs, there was no substantial improvement in SSIM. This metric measures the structural similarity between generated and target images, and the persistently stagnant SSIM values signify limited progress in enhancing image quality [22].

Table 2 SSIM scores for the initial experiment

Epoch	SSIM Score (Thermal)	SSIM Score (RGB)
60	0.4527086357494565	0.4136598223226676
70	0.6104599668341557	0.5537221752907157
80	0.6110520506282978	0.5547964070730498
90	0.5977662208575732	0.5472226876774037
100	0.5463739647090523	0.5028143657651495
110	0.5917243868357263	0.550075507744895
120	0.5864292338556983	0.5447662542193822
130	0.6034873476180342	0.5540493768976488
140	0.593879230116054	0.5519895181152206
150	0.5991113267662409	0.5621314135620368

Figure 9 serves as the baseline for testing to assess the accuracy and performance of the model due to its inclusion of multiple objects and a human in the scene.

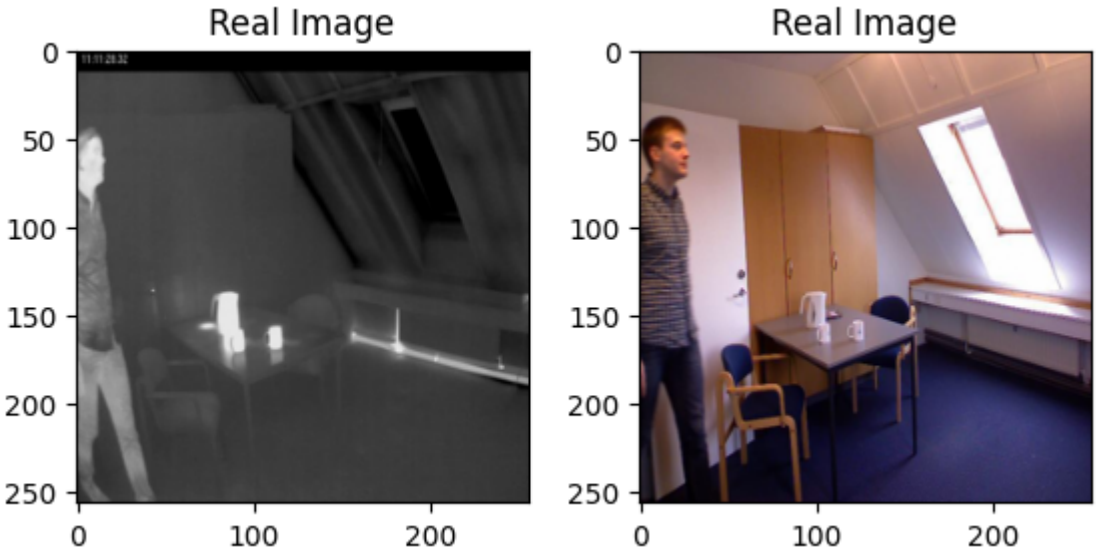


Figure 9: Thermal IR and RGB image used as baseline for SSIM testing

Therefore, we conclude that the challenges encountered during the training of the CycleGAN model, including issues like pronounced blurriness and an inability to capture "white hot" features, likely result from a combination of factors, possibly including data imbalances, where some scenes or objects may have been overrepresented or underrepresented [17]. Such imbalances could have adversely impacted model performance.

5.3 Scene Reduction

The data used for training doesn't cover a wide range of scenarios or has some built-in biases, making it tough for models to handle different scenes and objects effectively. We saw this when we first started training our model. To address this issue and improve the model's performance, we'll explore strategies such as reducing dataset size while maintaining diversity and making selective data subset choices.

As a result, we have narrowed our dataset to just 4k data, concentrating on a single scene within a room.

5.3.1 Experimentation Results

We decided to see what would happen if we used a smaller dataset that only focused on one specific scene in a room. Surprisingly, we encountered some unexpected issues.

At epoch 17 and before, all the images generated seemed to be able to capture the "white hot" features better than the initial experiment. This shows that reducing the scenes enables the model to have more focused training, hence better results. However, it is observed that after 17 epochs, the generated images seem to display checkered patterns.

Checkered patterns in images during GAN training are often a manifestation of a common problem known as mode collapse or the checkerboard artifact [24]. This artifact is

characterized by a grid-like or checkerboard pattern of alternating colors or textures in the generated images, which can be undesirable and disrupt the quality of the generated content.

Below are Figures 10 and 11 showing the generated images from both domains, which display better results as compared to the initial experiment.

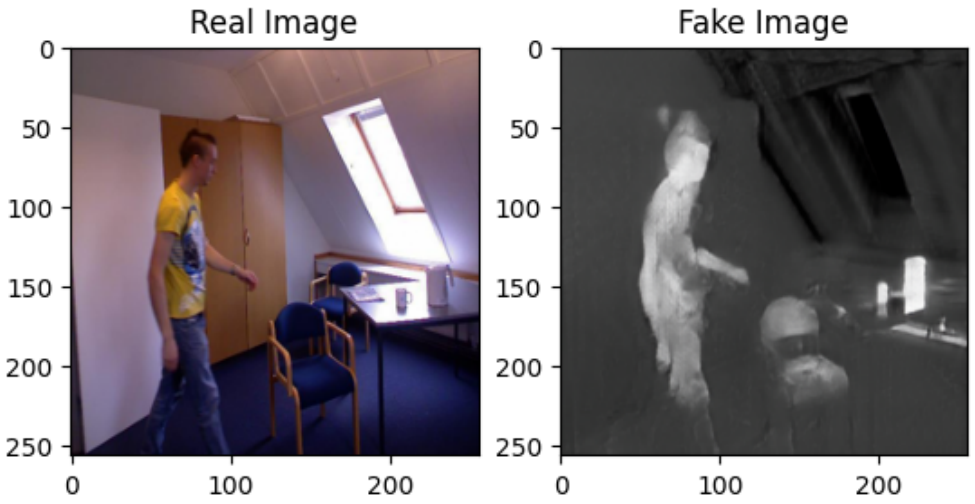


Figure 10: Comparison of Real and Generated Thermal IR Images at Epoch 17 (Second Experiment)

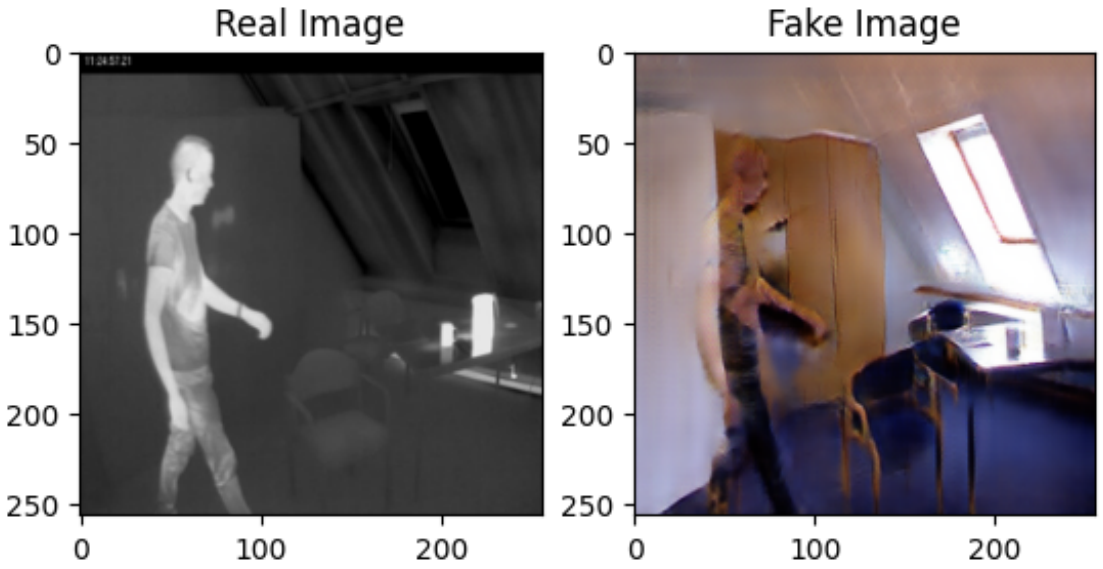


Figure 11: Comparison of Real Thermal IR Images and Generated RGB Images at 17 Epoch(Second Experiment)

Figures 12 and 13 display the result of generated images from both domains where checkered patterns are observed, especially when we translate thermal IR to RGB images.

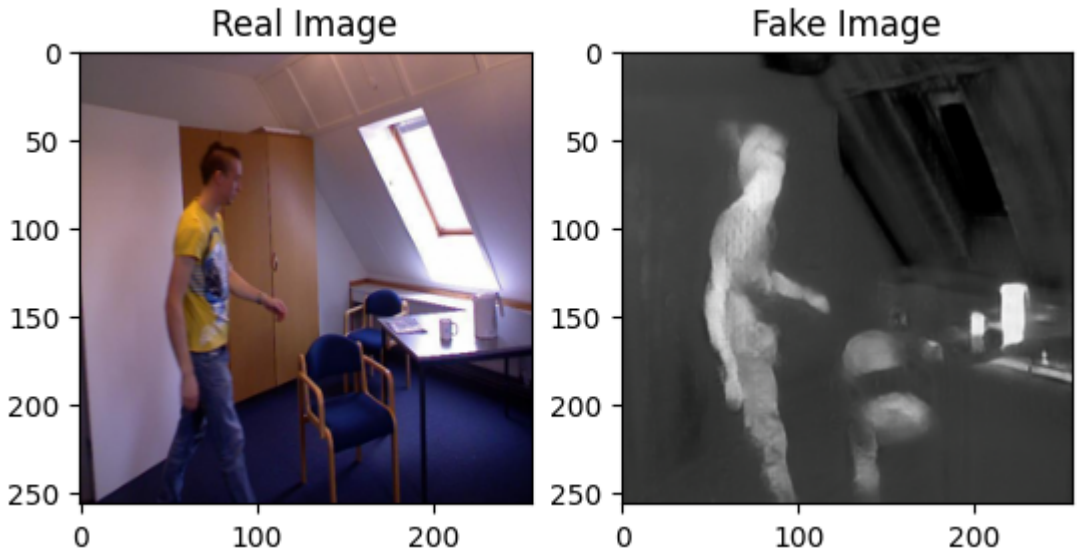


Figure 12: Comparison of Real and Generated Thermal IR Images at Epoch 25 (Second Experiment)

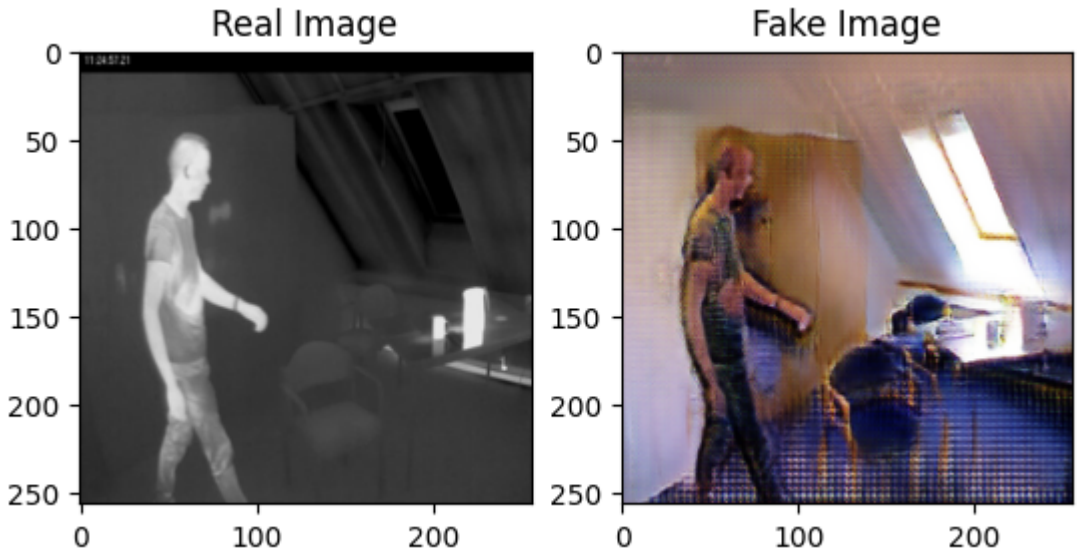


Figure 13: Checkerboard Patterns observed after Epoch 25(Second Experiment)

This issue was also discussed in a GitHub thread [21], revealing that the reduction in dataset size initially led to an improvement in training. On the contrary, it appeared to exacerbate the

problem with the checkered pattern. The exact reasons for this phenomenon are still under investigation, and potential factors being considered include insufficient sampling, which occurs due to a lack of data or a dataset that is not representative of the target distribution [23].

5.4 Mitigate Checkerboard Patterns

When we decreased the number of training samples, fake images generated began displaying checkerboard patterns. Consequently, we implemented data augmentation methods to improve the model's performance, allowing us to augment the number of images featuring the identical scene. Since we were working with thermal infrared (IR) images that lack color information, and adjusting image colors wasn't a feasible choice, our primary emphasis was on utilizing flips for augmentation.

We introduced more images by flipping images, both horizontally and vertically, providing an effective means to introduce variability into the training data. Additionally, researchers have proposed using bilinear upsampling to address the issue of checkered effects [24], which we will explore in the next section.

5.5 Bilinear Upsampling

A substantial modification in our methodology involved transitioning from using Transposed Convolution layers for upscaling in the generator network to the adoption of bilinear upsampling. This shift was prompted by the need to enhance the quality of generated images and address issues such as checkerboard artifacts, which are often associated with transposed convolutions. Bilinear upsampling introduced a more natural and visually appealing upscaling mechanism, significantly contributing to the overall fidelity of style transfer.

In our network architecture, this modification was realized by replacing the conventional Transposed Convolution layers with bilinear upsampling. Instead of using Transposed Convolution, we incorporated bilinear upsampling as a crucial step in our generator network. This change involved replacing the conventional Transposed Convolution layers with bilinear upsampling in our network architecture. We integrated the 'nn.Upsample' layer with a scale factor of 2 and the 'mode' set to 'bilinear'. Additionally, we added 'nn.ReflectionPad2d' and 'nn.Conv2d' layers to refine the network.

In summary, these modifications were essential in improving the performance and quality of our Cyclegan style transfer model. Our methodology now provides a strong foundation for achieving more accurate and visually appealing style transfers within specific scene contexts.

5.5.1 Experimentation Results

We proceeded to conduct experiments following the incorporation of bilinear upsampling. The results were exceptional, exhibiting a remarkable visual resemblance to real images.

Through visual inspections conducted at various training stages, we observed substantial improvements in the quality of generated images, successfully resolving the previously encountered checkered pattern issue. In Figure 14, we present a loss curve chart for the discriminator and generator, representing Domain A (RGB to Thermal IR) and Domain B (Thermal IR to RGB).

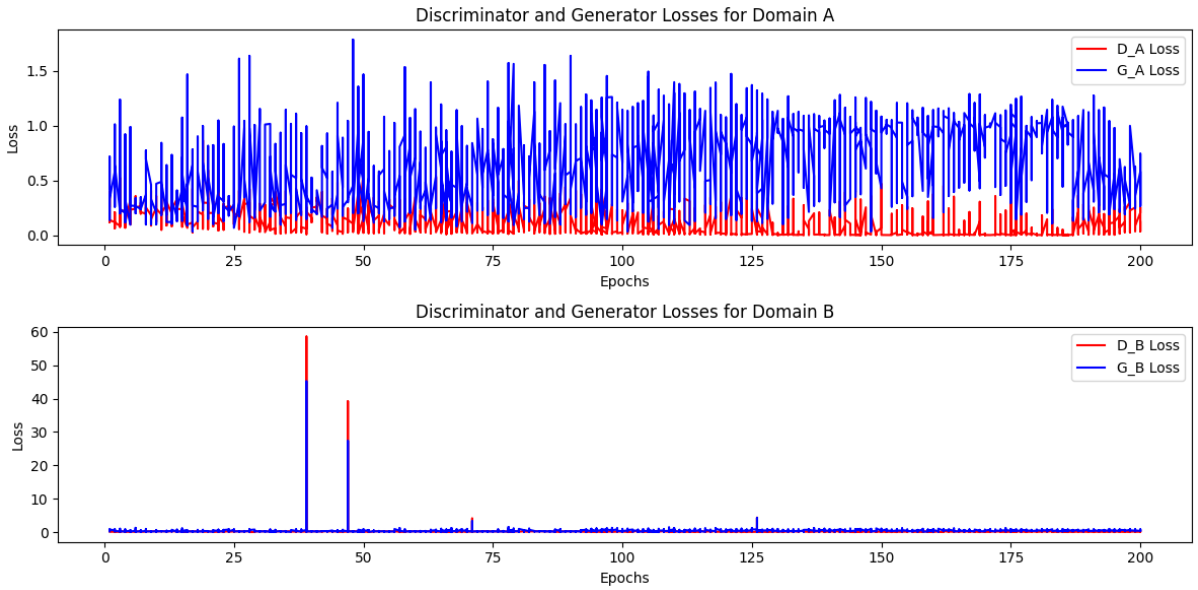


Figure 14: Discriminator and Generator Losses for Third Experiment

An important observation from Figure 14 is that in the initial training phases, the discriminator (D) loss surpasses the generator (G) loss. This indicates the accurate identification of real and fake samples. As the training progresses, the D loss gradually decreases, while the G loss experiences a slight increase before stabilizing. This pattern indicates that the generator becomes more proficient in generating realistic samples. Notably, we observed this convergence of D and G loss at approximately 160-180 epochs, signifying a critical point in the training process. This dynamic, where the discriminator loss initially exceeds the generator loss, and subsequently, the generator becomes more adept at producing realistic samples, is a key factor contributing to the success of the training process.

Table 3 presents the SSIM scores obtained during the experiment with bilinear upsampling, with Figure 15 serving as the baseline for testing to assess the accuracy and performance of the model due to its inclusion of multiple objects and a human in the scene.

Table 3 SSIM scores for experiment with bilinear upsampling

Epoch	SSIM Score (Thermal)	SSIM Score (RGB)
155	0.768052272148336	0.7730335182182819
160	0.7807392557192254	0.7569300399771453
165	0.773554927970701	0.7610954650873961
170	0.7259941222516224	0.7666506899519995
175	0.743281284492439	0.7639719953539491
180	0.7367696898665373	0.7515180461896966
185	0.727109545147844	0.7532073067282977
190	0.7710603803326982	0.7500704613516012
195	0.7772932162658809	0.7484290109011051
200	0.7764023127947229	0.7489500122531175

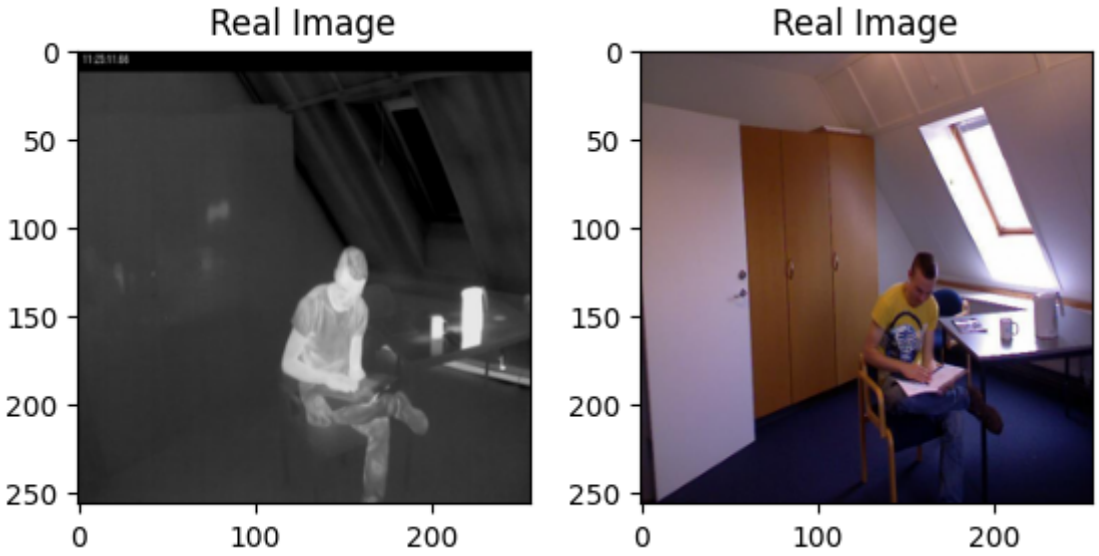


Figure 15: Thermal IR and RGB image used as baseline for SSIM testing

From Table 3, it's evident that the SSIM scores for Figure 15 peaked around epochs 155-160. In this context, a score greater than 0.7 is considered relatively high, indicating a significant level of similarity between the two objects being compared. It's worth noting that the variations between different epochs are relatively small, typically differing by just a few percentage points. Therefore, any of these epochs should be capable of generating fake images of similar quality.

Here are some examples of generated images, showcasing the conversion from RGB to Thermal IR and vice versa. Additional test results can be found in the appendix

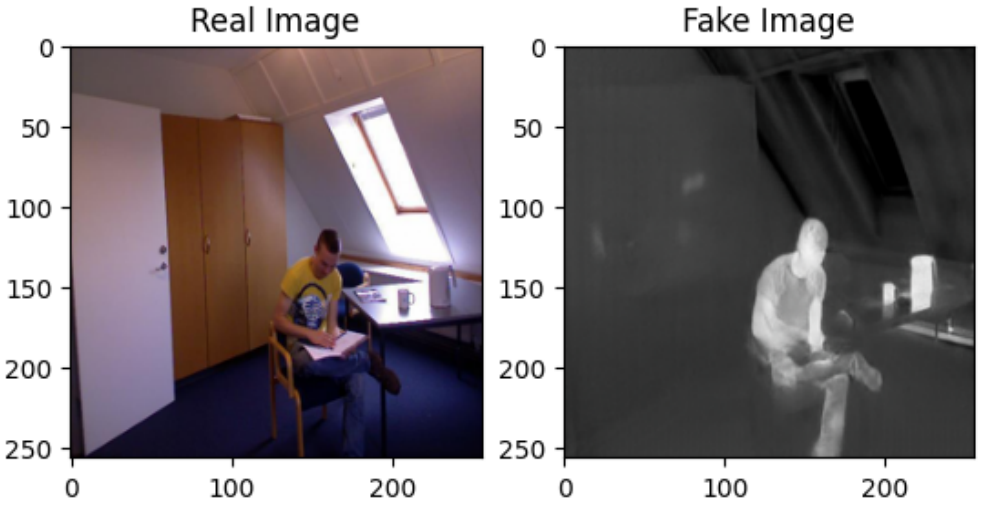


Figure 16: Comparison of Real and Generated Thermal IR Images at 155 Epoch

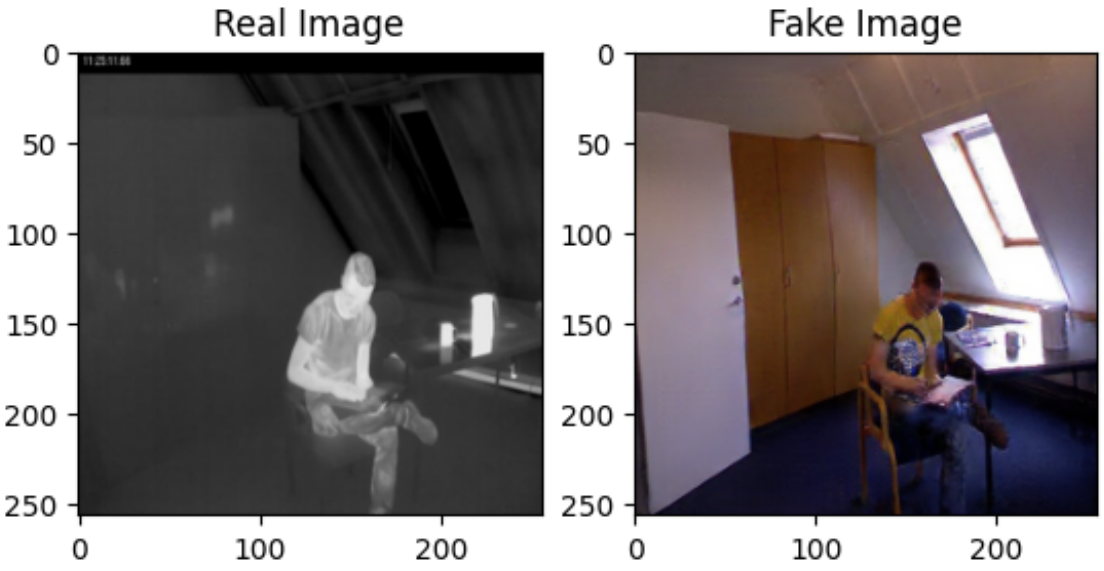


Figure 17: Comparison of Real and Generated Thermal IR Images at 155 Epoch

6. Contributions and Findings

While our proposed approach has demonstrated impressive results in many scenarios, it's important to acknowledge inherent limitations. Figure 17 highlights some challenges in facial restoration, where the model falls short in preserving fine details such as facial expressions and shirt design. Despite achieving a commendable near-0.8 SSIM score for the generated images, issues persist in this specific aspect, possibly stemming from an insufficient training dataset with limited images of individuals.

Nevertheless, our method consistently excels in colorizing thermal infrared images, particularly in indoor environments, covering a wide range of objects such as cups, books, and chairs, as observed in our test cases.

7. Future Works

As we advance in our CycleGAN style transfer research in thermal imaging, our current experiment successfully demonstrates our model's capacity to accurately capture objects and human subjects in regular and thermal imagery. This breakthrough holds immense potential for diverse applications across industries. To make the most of this potential and ensure it benefits a wide range of sectors, here are some recommendations for future works:

Dataset Expansion: Our existing dataset, though thoughtfully assembled, represents a treasure trove of untapped possibilities. By broadening the dataset's horizons to encompass an eclectic range of scenes and a richer, more diverse color palette, we set the stage for our model's adaptability to thrive in multiple industries.

Richer Color Representation: As our current dataset primarily consists of images within the same scene, we are limited to a relatively narrow color palette. It would greatly benefit our model's completeness and versatility for various real-world applications if we were to collect

data containing more vibrant colors. This is particularly important when we aim to generate RGB images from thermal IR images.

Recommendation for Industry-Specific Training: We strongly advise focusing on industry-specific training to harness the transformative potential of our research. In sectors like agriculture, healthcare, and security, our technology offers the opportunity for precise crop temperature monitoring, early disease detection, enhanced thermal imaging for surveillance, and anomaly detection. These applications exemplify the far-reaching impact our work can have within specialized industries.

To sum up, our future work goes beyond the lab and into various industries ready for innovation. This effort aims to change how we use thermal imaging in different fields.

8. References

[1] J.-Y. Zhu, T. Park, P. Isola and A. A. Efros, "Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks.," in IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 2017.

[2] Y. Choi, M. Choi, M. Kim, J.-W. Ha, S. Kim and J. Choo, "StarGAN: Unified Generative Adversarial Networks for Multi-Domain Image-to-Image Translation," in IEEE Computer Vision and Pattern Recognition (CVPR), 2018.

[3] A. Chartsias, J. Thomas, G. Mario Valerio , and T. Sotirios A, "Multimodal MR synthesis via modality-invariant latent representation," IEEE Transactions on Medical Imaging, pp. 525-534, 2018.

[4] X. Wang and A. Gupta, "Generative Image Modeling using Style and Structure Adversarial Networks," in European Conference on Computer Vision (ECCV), 2017.

[5] T. Karras, S. Laine, and T. Aila, "A Style-Based Generator Architecture for Generative Adversarial Networks," in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2019.

[6] A. B. e. al., "Large Scale GAN Training for High Fidelity Natural Image Synthesis," in International Conference on Learning Representations (ICLR), 2019.

[7] A. Xue, "End-to-End Chinese Landscape Painting Creation Using," in IEEE Winter Conference on Applications of Computer Vision (WACV) 2021, 2020.

[8] Y. Choi, M. Choi, M. Kim, J.-W. Ha, S. Kim and J. Choo, "StarGAN: Unified Generative Adversarial Networks for Multi-Domain Image-to-Image Translation," IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 8789-8797, 2017.

[9] H. Xun, M.-Y. Liu, S. Belongie and J. Kautz, "Multimodal Unsupervised Image-to-Image Translation.," 2018.

[10] R. Dulski et al. Enhancing image quality produced by IR cameras. Electro-Optical and Infrared Systems: Technology and Applications VII. Vol. 7834. International Society for Optics and Photonics.2010,p.783415.

[11] Shuo Liu, Mingliang Gao, Vijay John, Zheng Liu, and ErikBlasch. Deep learning thermal image translation for night vision perception. ACM Transactions on Intelligent Systems and Technology, 12:1–18, 12 2020

[12] J. Cilulko, P. Janiszewski, M. Bogdaszewski and E. Szczygielska, "Infrared thermal imaging in studies of wild animals," European Journal of Wildlife Research 59, 2012.

[13] C. Palmero, A. Clapés, C. Bahnsen, A. Møgelmose, T. B. Moeslund, and S. Escalera, "Multi-modal RGB–Depth–Thermal Human Body Segmentation," in International Journal of Computer Vision, pp. 1-23, 2016.

[14] J.-Y. Zhu, "Image-to-image translation with conditional adversarial networks - Issue #206," GitHub, Issue #206, pytorch-CycleGAN-and-pix2pix, 2019. [Online]. Available: https://github.com/junyanz/pytorch-CycleGAN-and-pix2pix/issues/206. [Accessed: October 9, 2023].

[15] R. Simon and K. Dobbin, "Optimally splitting cases for training and testing," BMC Med Genomics, vol. 4, p. 31, 2011.

[16] Z. Sun, J. Gui, Y. Wen, D. Tao, and J. Ye, "A Review on Generative Adversarial Networks," Journal of Latex Class Files, vol. 14, p. 2, 2015.

[17] P. Kumar, R. Bhatnagar, K. Gaur, and A. Bhatnagar, "Classification of Imbalanced Data: Review of Methods and Applications," IOP Conference Series: Materials Science and Engineering, vol. 1099, pp. 012077, 2021. DOI: 10.1088/1757-899X/1099/1/012077.

[18] Gade R, Moeslund T B. Thermal cameras and applications: a survey[J]. Machine vision and applications, 2014, 25(1): 245-262.

[19] V. V. Kniaz, V. A. Knyaz, J. Hladuvka, W. G. Kropatsch, and V. A. Mizginov, "ThermalGAN: Multimodal Color-to-Thermal Image Translation for Person Re-Identification in Multispectral Dataset," in Computer Vision -- ECCV 2018 Workshops, Springer International Publishing, 2018.

[20] H. Liik, "Thermal Image Generation from RGB," [Online]. Available: https://medium.com/@hannesliik/thermal-image-generation-from-rgb-b152efa66cc2. [Accessed: 1/09/2023].

[21] "GitHub Issue #190," GitHub, 190, Available: https://github.com/junyanz/pytorch-CycleGAN-and-pix2pix/issues/190. [Accessed: 11/10/2023]

[22] P. Kancharla and S. S. Channappayya, "Quality Aware Generative Adversarial Networks," in Proceedings of NeurIPS, 2019, pp. 3.

[23] R. Bayat, "A Study on Sample Diversity in Generative Models: GANs vs. Diffusion Models," Tiny Papers @ ICLR 2023, Mar. 02, 2023

[24] A. Odena, V. Dumoulin, and C. Olah, "Deconvolution and Checkerboard Artifacts," Oct. 17, 2016. [Online]. Available: https://distill.pub/2016/deconv-checkerboard/ [Accessed: 1/09/2023]

Appendix A: Supplementary Charts

In this section, we provide supplementary charts that support the findings presented in the main text.

Chart A1: Loss Curve for Initial Experiment

The following chart displays the cycle_consistency and identity loss curves for the initial experiment in Section 5.2.1:

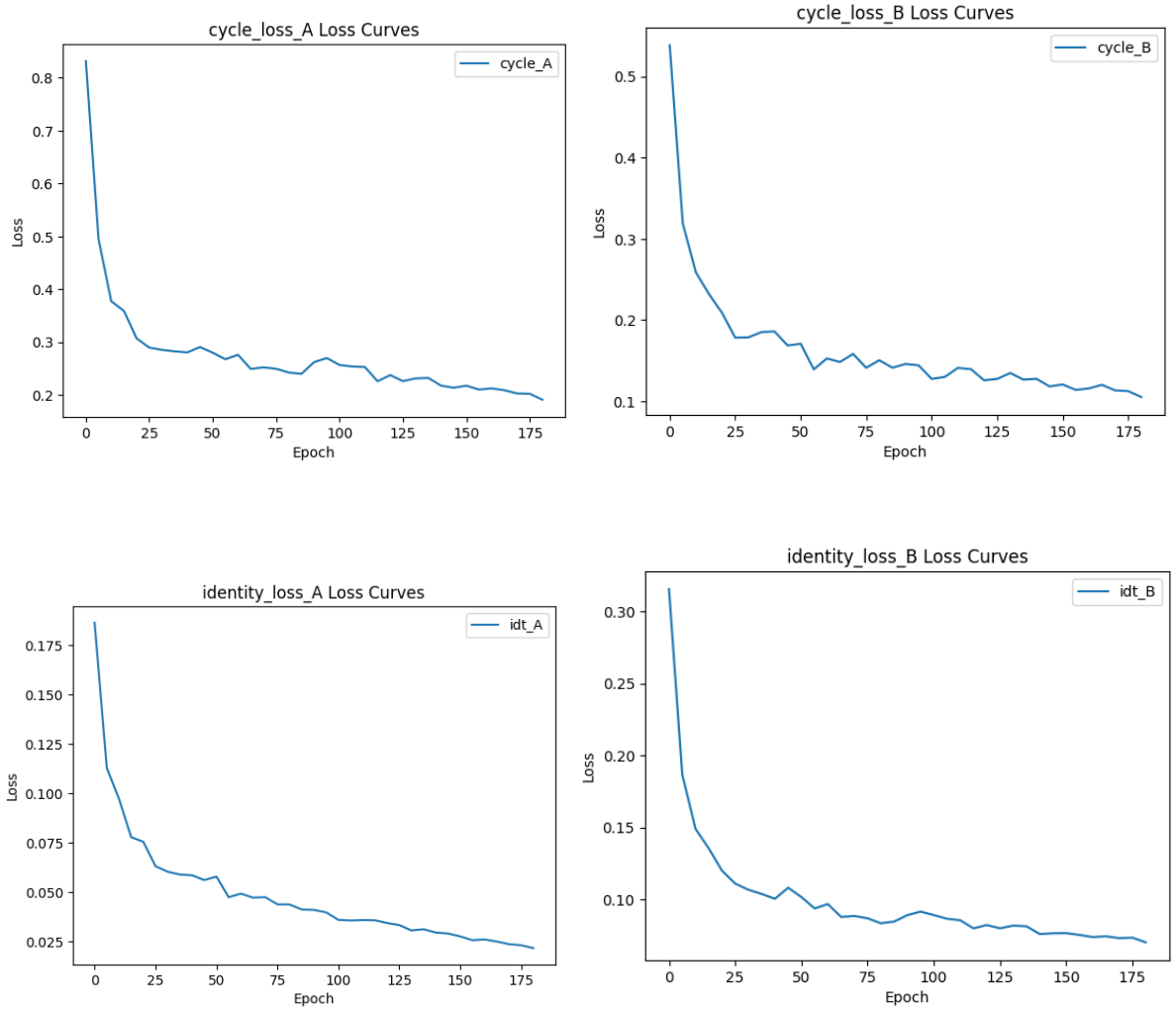
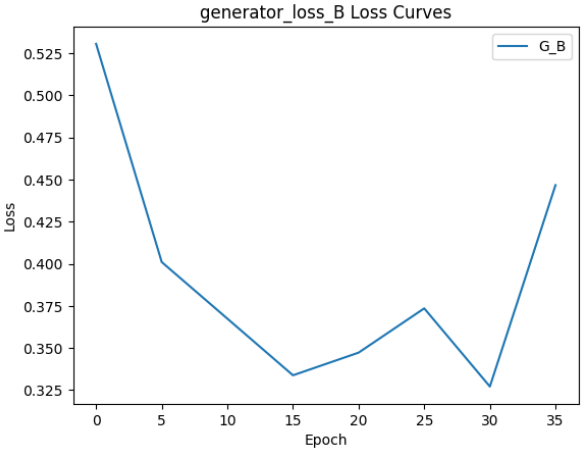
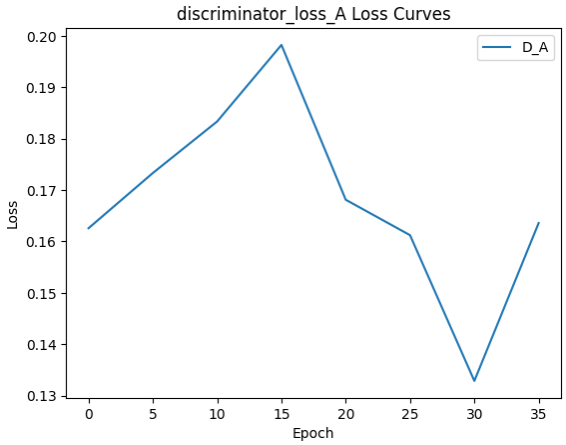


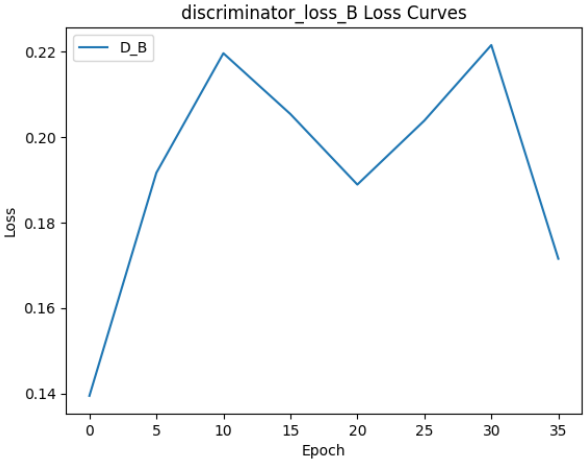
Chart A2: Loss Curve for Second Experiment

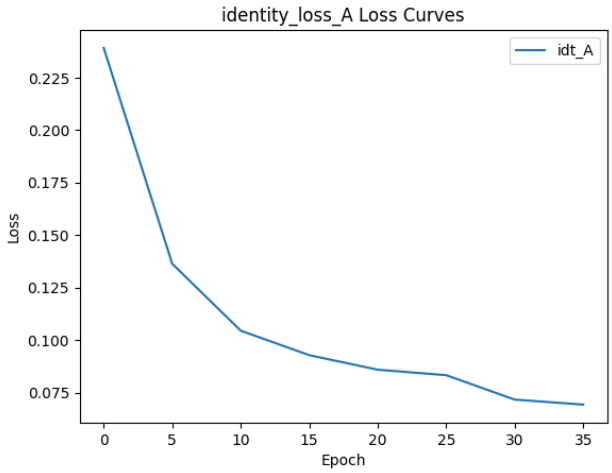
The following chart displays the loss curves for the second experiment in Section 5.3.1:

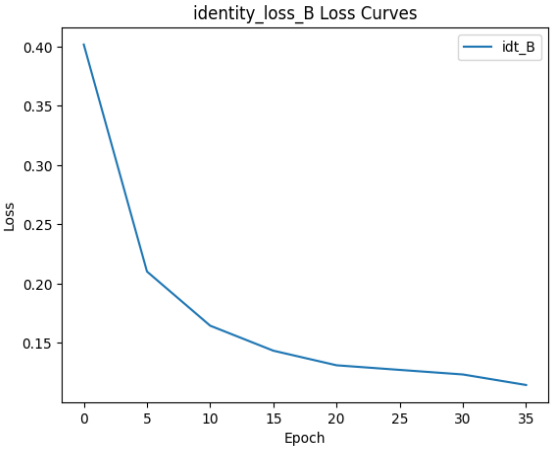

generator_loss_A Loss Curves
G_A
Loss
Epoch

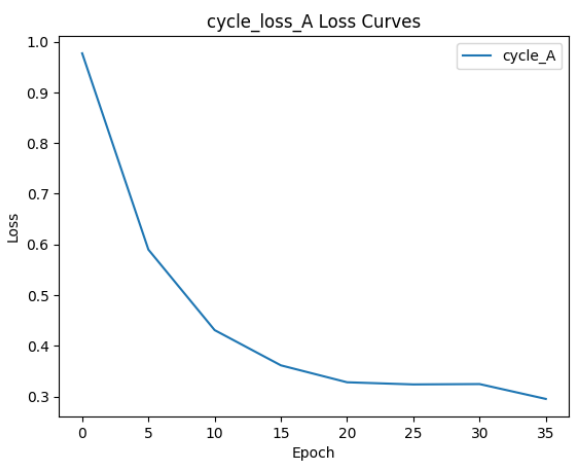

generator_loss_B Loss Curves
G_B
Loss
Epoch


discriminator_loss_A Loss Curves
D_A
Loss
Epoch


discriminator_loss_B Loss Curves
D_B
Loss
Epoch


identity_loss_A Loss Curves
idt_A
Loss
Epoch


identity_loss_B Loss Curves
idt_B
Loss
Epoch



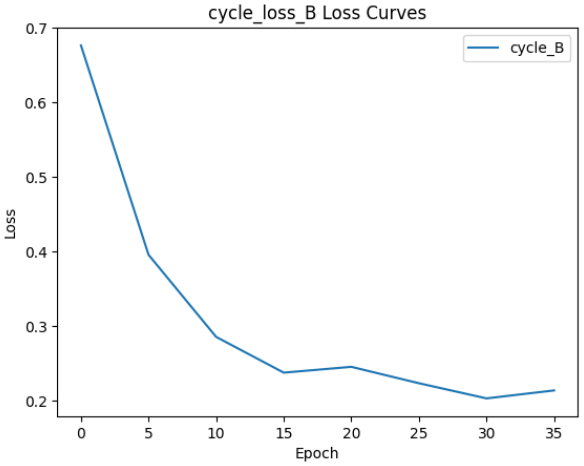
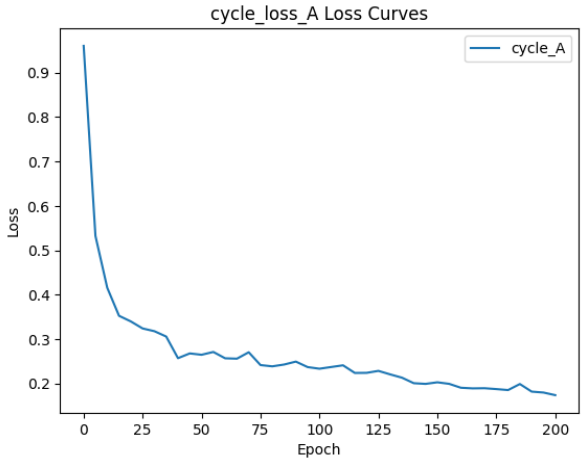
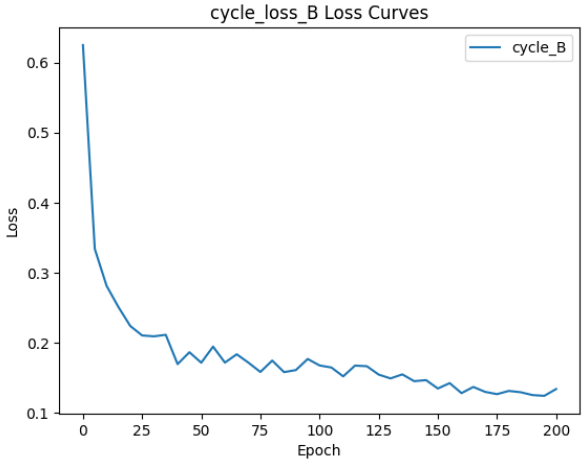
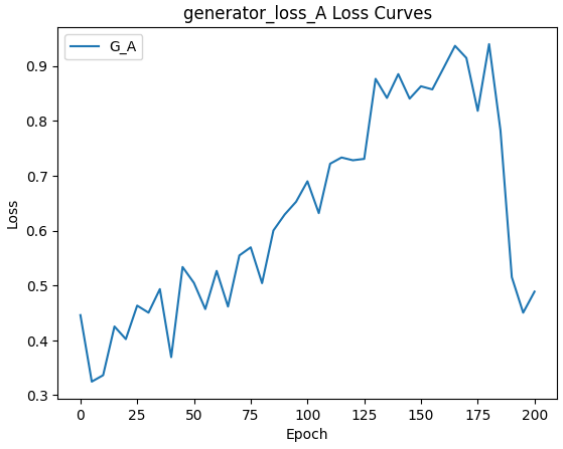


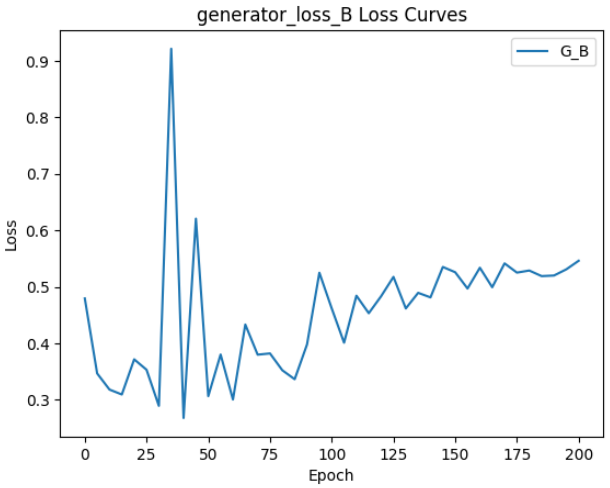
Chart A3: Loss Curve for Third Experiment

The following chart displays the loss curves for the second experiment in Section 5.5.1:









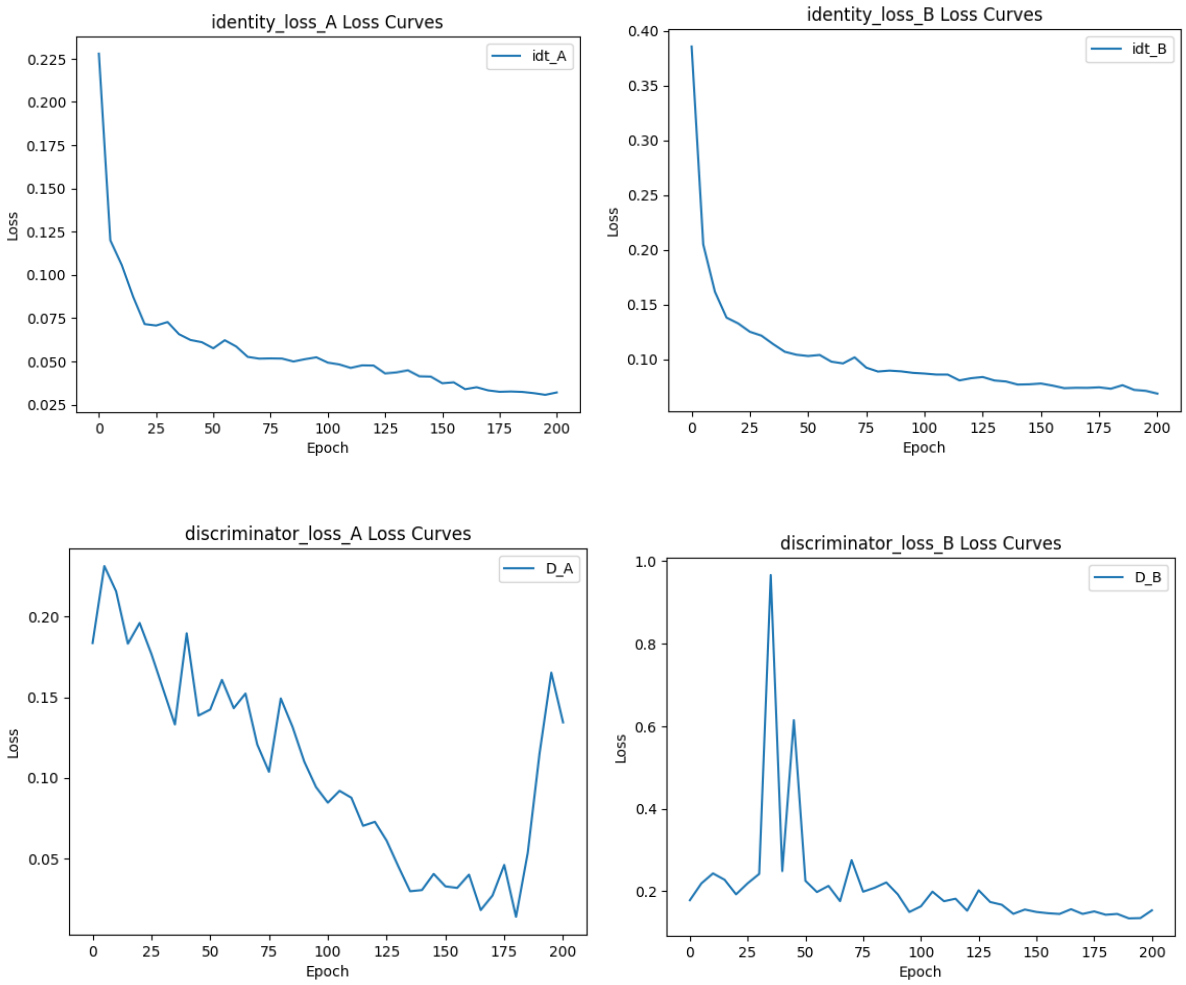


Figure A4: Extra Test Results for Initial Experiment

Below are test images that illustrate the comparison between real and generated images at the 80th epoch.

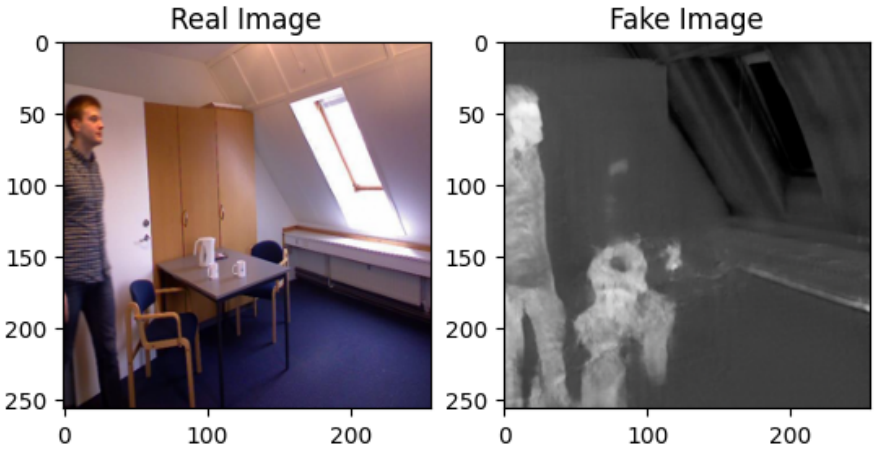


Figure A5: Extra Test Results for Experiment with Checked Patterns

Below are test images that illustrate the comparison between real and generated images at the 25th epoch.

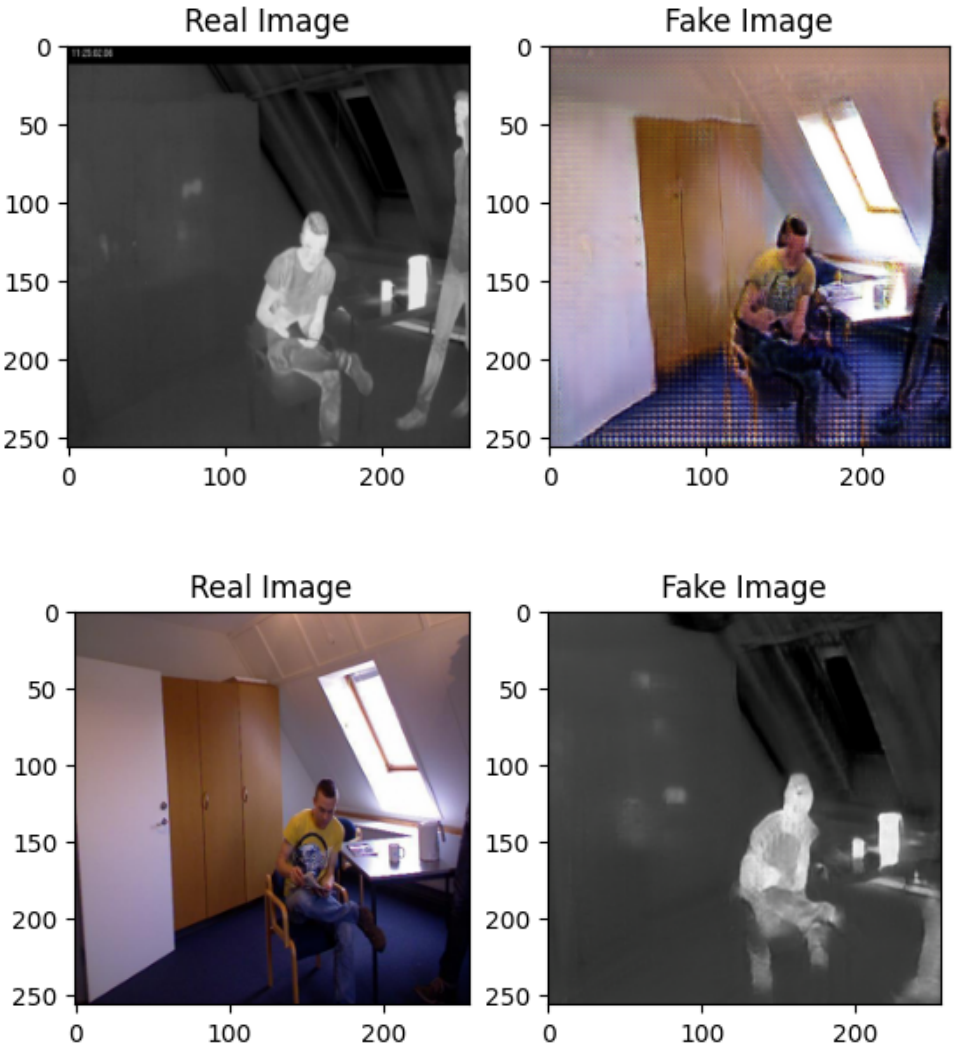
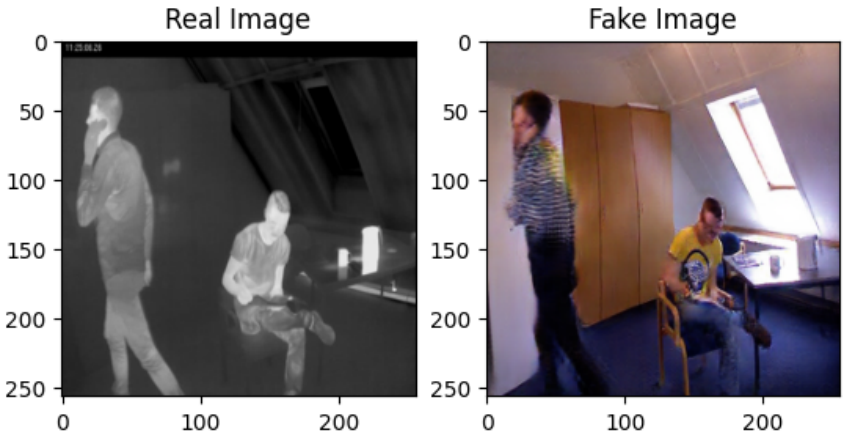
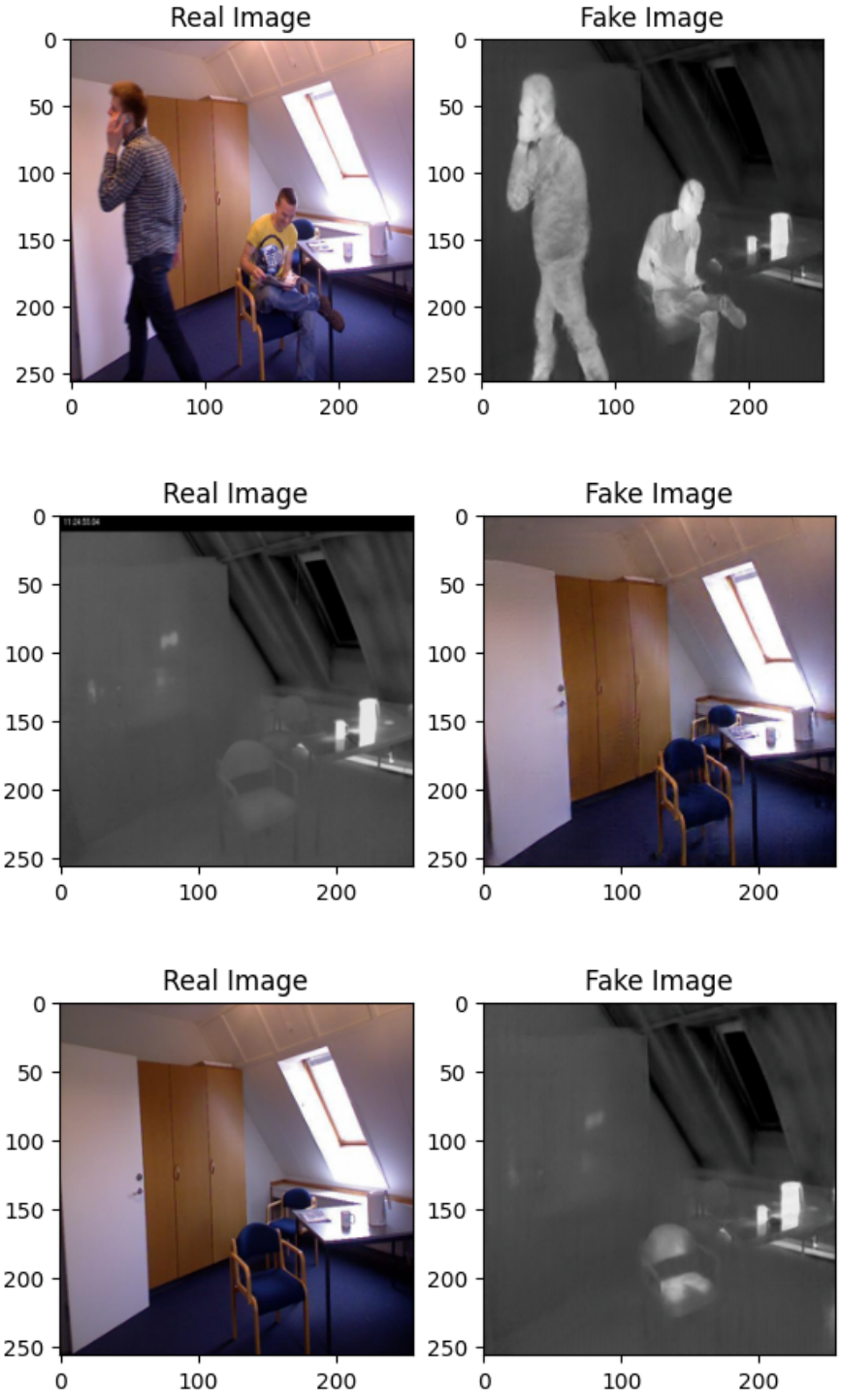


Figure A6 Extra Test Results for Experiment with Bilinear Upscaling

Below are test images that illustrate the comparison between real and generated images at the 155th epoch.




Real Image
Fake Image
Real Image
Fake Image
Real Image
Fake Image