Electronics and Computer Science
Faculty of Engineering and Physical Sciences
University of Southampton

Georgi Iliev

April 30, 2024

Enhanced Generative Image Transformation Tool (EGITT)

Supervisors
Supervisor: Hikmat Farhat
Second Examiner: Danesh Tarapore

A project report submitted for the award of MEng Computer Science with Artificial Intelligence

Abstract

Generative models, involving techniques such as image generation and transformation, are central to the development of artificial intelligence, offering remarkable possibilities for creating and synthesising new visual content. Despite their potential, these models face challenges such as low resolution, colour distortions, and style fidelity.

This study introduces the Enhanced Generative Image Transformation Tool (EGITT), a project designed to harness these models and bridge the gap between human and AI-generated images. It aims to address current limitations and advance applications of Generative Modelling across various domains, such as art and media, by developing comprehensive evaluation frameworks and improving model functionalities via systematic experimentation.

It paves the way for future research, development, and understanding of artificial intelligence in creative domains.

Statement of Originality

Statement of Originality

- I have read and understood the ECS Academic Integrity information and the University's Academic Integrity Guidance for Students.
- I am aware that failure to act in accordance with the Regulations Governing Academic Integrity may lead to the imposition of penalties which, for the most serious cases, may include termination of programme.
- I consent to the University copying and distributing any or all of my work in any form and using third parties (who may be based outside the EU/EEA) to verify whether my work contains plagiarised material, and for quality assurance purposes.

You must change the statements in the boxes if you do not agree with them.

We expect you to acknowledge all sources of information (e.g. ideas, algorithms, data) using citations. You must also put quotation marks around any sections of text that you have copied without paraphrasing. If any figures or tables have been taken or modified from another source, you must explain this in the caption and cite the original source.

I have acknowledged all sources, and identified any content taken from elsewhere.

If you have used any code (e.g. open-source code), reference designs, or similar resources that have been produced by anyone else, you must list them in the box below. In the report, you must explain what was used and how it relates to the work you have done.

I have not used any resources produced by anyone else.

You can consult with module teaching staff/demonstrators, but you should not show anyone else your work (this includes uploading your work to publicly-accessible repositories e.g. Github, unless expressly permitted by the module leader), or help them to do theirs. For individual assignments, we expect you to work on your own. For group assignments, we expect that you work only with your allocated group. You must get permission in writing from the module teaching staff before you seek outside assistance, e.g. a proofreading service, and declare it here.

I did all the work myself, or with my allocated group, and have not helped anyone else.

We expect that you have not fabricated, modified or distorted any data, evidence, references, experimental results, or other material used or presented in the report. You must clearly describe your experiments and how the results were obtained, and include all data, source code and/or designs (either in the report, or submitted as a separate file) so that your results could be reproduced.

The material in the report is genuine, and I have included all my data/code/designs.

We expect that you have not previously submitted any part of this work for another assessment. You must get permission in writing from the module teaching staff before re-using any of your previously submitted work for this assessment.

I have not submitted any part of this work for another assessment.

If your work involved research/studies (including surveys) on human participants, their cells or data, or on animals, you must have been granted ethical approval before the work was carried out, and any experiments must have followed these requirements. You must give details of this in the report, and list the ethical approval reference number(s) in the box below.

My work did not involve human participants, their cells or data, or animals.

ECS Statement of Originality Template, updated August 2018, Alex Weddell aiofficer@ecs.soton.ac.uk

Acknowledgements

I would like to thank my project supervisor, Hikmat Farhat, and my second examiner, Danesh Tarapore, for their guidance and support throughout this project.

Hikmat consistently gave his time to help direct the progress of the project and always promptly replied to any query I had. His kindness, and willingness to give me an appropriate amount of freedom has inspired me to progress within the field of this research.

Danesh was more than willing to meet with me and give his perspective on which areas of the project are valuable to a reader, this really helped enhance the quality of the report.

Contents

List of Figures

List of Tables

1 Introduction

Generative models, containing image generation and image transformation, are an increasingly popular domain of artificial intelligence. Using analysis and extraction techniques enables computers to generate new images by identifying patterns, understanding context, and applying knowledge from vast datasets. There are broad applications, as this project, named Enhanced Generative Image Transformation Tool (EGITT), focuses on celebrities and art. The increasing usage of such techniques narrows down the line between human and computer-generated imagery, resulting in new forms of creativity and challenges in distinguishing between the two.

Problem: Despite advanced technology, images produced via generative models suffer from numerous problems, such as low resolution, a lack of fidelity to the original style or content, blurred features, colour distortions, and visible artefacts. Every application is well-suited to a specific model. Therefore, choosing an accurate one contributes massively to the final output. However, the only way to develop a sophisticated predictive model is through trial and error until the desired result is generated, which is not practical. Having a standardised approach for every application, such as Generative Adversarial Networks (GANs) and Style Transfer, sets the path for numerous applications and further advancements, one of which is this third-year project.

Goals: The primary aim of this research is to choose appropriate generative models for developing evaluation frameworks for image generation and image transformation. It also aims to address present issues in generative models by suggesting novel functionalities. This project includes several learning and development steps:

- Familiarisation with various GAN techniques and their processes of development, training, testing, and evaluation. Investigate limitations and identify potential room for improvement.
- Implementation of different GAN models, including a standard GAN and several versions of a Wasserstein GAN. Apply evaluation metrics for comparison of the obtained results.
- Adaptation of image transformation models focusing on real-time style transfer and combining the content of one image with the style of another by using different networks.
- Production of tangible results comparable to well-known papers published by respected authors and institutions. Examine obstacles in current style translation methodologies and offer advancements.
- Evaluation and optimisation inspection of methods and strategies for implemented models. Do qualitative and quantitative assessments of constructions to ensure user acceptance.
- Identification of correlations between the resulting images. Showcase the impact and significance of this paper in the realm of artificial intelligence.

Scope: The scope of this study is to advance the field of generative models, namely

image generation and transformation. Key aspects of the scope include model exploration and selection, development and improvement of models, an evaluation framework, potential applications, and impact, together with the limitations and challenges faced during the process. Additionally, this scope includes a meaningful contribution to model transparency and practical applications of generative models in art and media.

2 Literature review

The literature review section embarks on a comprehensive exploration of generative modelling with its subdomains, image generation, and transformation. It examines their implications and transformative effects on the way machines perceive and interpret visual information, pushing the boundaries of creativity and innovation. This section explores both pioneering works and cutting-edge research. It captures the dynamic interaction between technological advancements and theoretical insights forming the current landscape of AI. By reviewing key methodologies, applications, and challenges, this paper underpins its position within the broader context of ongoing research and future directions of artificial intelligence. Moreover, it sets the stage for subsequent analysis and discussions.

2.1 Generative Modelling

At the core of artificial intelligence advances lies generative modelling, "a key process that provides machines with the remarkable capacity to create and synthesise new entities" [1][2]. This field, covering a spectrum of methodologies including GANs and style transfer as depicted in *Fig. 1*, demonstrates a step forward in computational creativity.

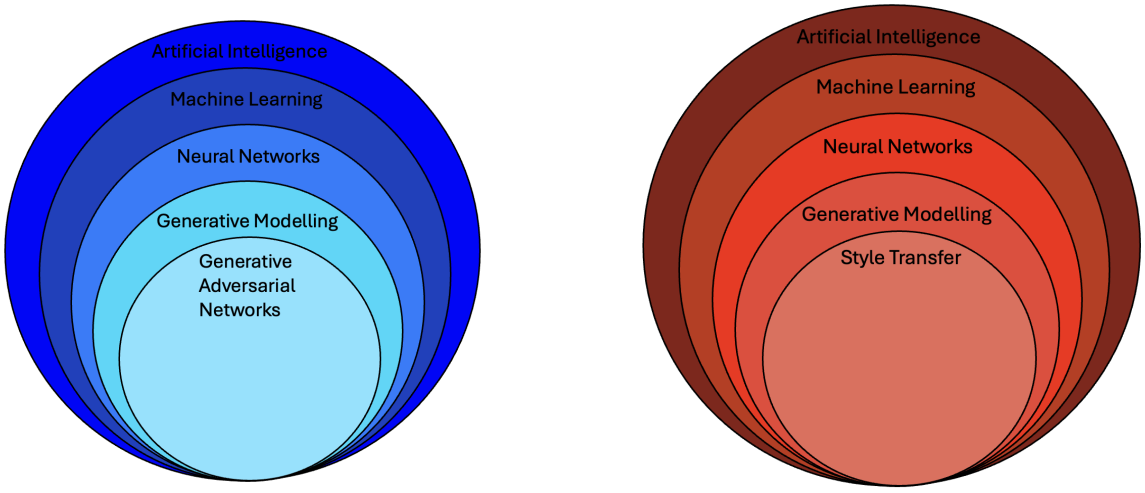


Figure 1: *Hierarchy of Generative Modelling*

Generative modelling works by using deep neural networks to mimic the distribution of original data, thus generating new data instances similar to the authentic dataset [11].

This process tries to capture and replicate complex patterns and nuances of input data, facilitating improved performance of ML models [2].

Generative modelling is applied in many real-world applications. Its evolution is characterised by transitioning from simple pattern recognition to comprehensive model architectures capable of understanding and reproducing complex data structures [28].

Despite its significant advancements, generative modelling faces several notable limitations due to the inherent design of modifying a single real image [1]. Such shortcomings significantly constrain the generalisation ability of models, positioning them as a focus area for research and development.

Overcoming these limitations requires continuous innovation in model architecture, loss function design, and learning methodologies [28]. Future research shall focus on improving the generativity of generative models and broadening their applicability and effectiveness. This will not only improve their practicality in real-world scenarios but also enrich the AI field by enabling more creative and diverse applications of machine-generated content.

2.2 Generative Adversarial Networks (GANs)

Generative Adversarial Networks, introduced by Goodfellow et al. in 2014 [3], are a significant paradigm within generative models. They embody a novel approach that addresses inherent challenges and uses an adversarial game theory framework between two distinct neural networks. This dynamic interaction, illustrated in *Figure 2*, aims to enhance the overall quality of the generated data [3].

The operational nature of GAN is grounded in the symbiotic, yet adversarial relationship between two components: the generator, tasked with deceiving by creating convincing samples from a random noise vector [4], and the discriminator, aiming to distinguish between real and artificial data by classifying samples as real or fake [5]. Effectiveness is reliant on adversarial loss, which showcases the learning regime and ensures the obtained results are indistinguishable from ground-truth images [6].

Besides their fundamental image generation capabilities, GANs are central to numerous applications, ranging from photo-realistic image synthesis and facial recognition to image enhancement and style transfer [3]. Such versatility highlights their transformative potential across different domains, allowing new levels of creativity and data interpretation.

Despite their tremendous impact, GANs face challenges limiting their effectiveness and scope of application. Key issues include achieving Nash equilibrium [7], mitigating the vanishing gradient problem [8], preventing mode collapse [22], and the lack of evaluation metrics [22]. These challenges require extensive research and innovative solutions to unfold GANs' fullest potential. As technology advances, ongoing exploration and GAN refinement remain paramount [3] not only academically but also for harnessing GANs' transformative power in practical applications. This fosters innovation and creativity in the realm of artificial intelligence [35]

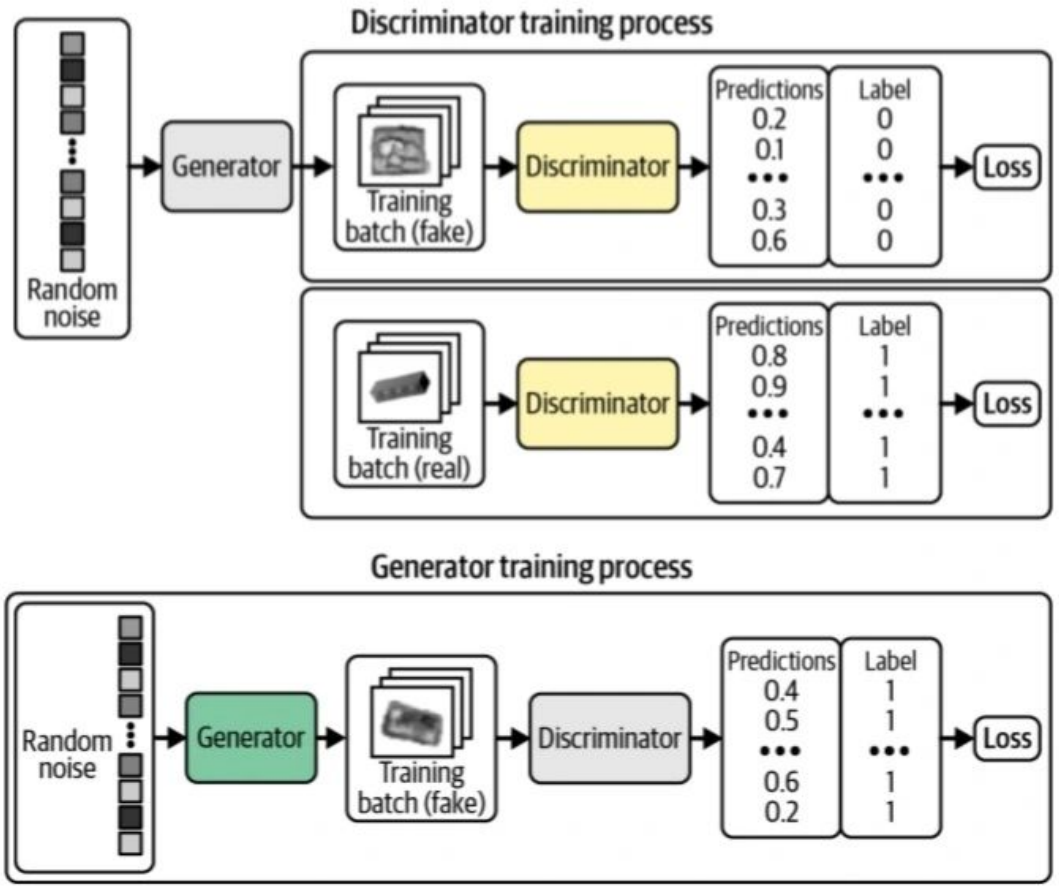


Figure 2: *GAN Modus operandi [11]*

2.3 Neural Style Transfer and Real-Time Style Transfer

Style transfer can be described as the process of producing a stylized image p that retains the content of a particular content image c while adopting the style of a different style image s [26] and is formulated as:

$$p = \Psi(c, s)$$

. The final goal of style transfer is to minimise the total loss, which is the sum of content loss $\mathcal{L}_c(p)$ and style loss $\mathcal{L}_s(p)$ functions in equation

$$\min_p \left(\mathcal{L}_c(p) + \mathcal{L}_s(p)\right)$$

Real-time style transfer,on the other hand, is a specialised form of this technique that is designed to operate rapidly enough to apply effects instantaneously [12]. Its objective is to determine the parameters θ for a feed-forward convolutional network T. This network includes residual connections between down-sampling and up-sampling layers and traverses across numerous content images [26]. The training uses the following formula as a loss function:

$$\min_\theta \left(\mathcal{L}_c(T(c)) + \mathcal{L}_s(T(c))\right)$$

The foundational technique of using convolutional neural networks (CNNs) for neural style transfer is initially introduced by Gatys and colleagues [1][2]. This breakthrough leverages the depth and power of the VGG network to segment and gather the stylistic and content features of images. Despite the promising artistic and technical sophistication of the novel technique, its computational demands—requiring iterative forward and backward transitions through the network—pose significant challenges to scalability and real-time applications [10]. This challenge sparks a wave of research integrating style transfer with deep learning. Initially, the focus is on optimisation-based methods, which, despite their effectiveness, are impractical for real-time applications due to the significant time requirements [29]. Therefore, the research emphasis shifts towards faster, feed-forward parametric approaches, which are significantly faster—about three orders of magnitude [10]. This paves the way for comprehensive systems that operate in real-time, as the one in *Fig.4*, pioneered by Justin Johnson, Alexandre Alahi, and Li Fei-Fei in their Real-Time Style Transfer and Super-Resolution" paper [12].

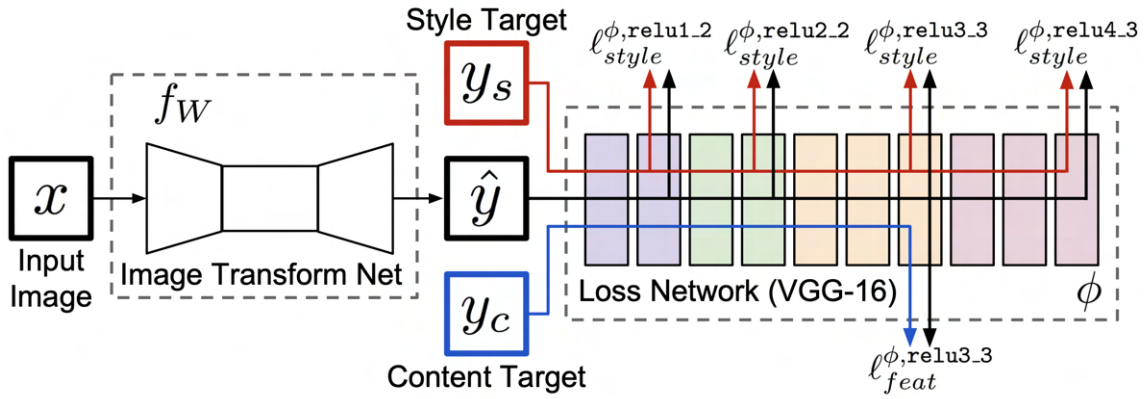


Figure 3: *Overview of a real-time system [12]*

The evolution of style transfer technologies broadens their application, ranging from augmenting digital media with artistic effects to enriching real-world entities with virtual textures. Such expansion, however, is not without its challenges. Current methodologies exhibit a bias towards certain styles, and the search for an optimal loss function that accurately captures the essence of both content and style remains a daunting task. Such limitations highlight the preciseness required when developing style transfer models [10].

As the field of style transfer continues to evolve, future research efforts are focused on exploring adaptive learning techniques covering a wider range of styles with minimal manual intervention. Furthermore, the development of more sophisticated models capable of more accurately determining the quality of style transfer presents a crucial opportunity for advancement [9]. These efforts not only improve flexibility and efficiency but also contribute to the broader crossover of technology and art, encouraging new forms of creative expression through the lens of artificial intelligence [23].

3 Methodology

The methodology of this third-year project follows a two-stage approach, aligned with the respective academic semesters and offering a structured framework for systematic exploration and development. It combines both supervised and unsupervised learning methods, recognising the diversity and complexity within these domains.

During the initial phase, corresponding to the first semester, the research focus was on learning GANs and diving into the fundamentals of image transformation. These processes included intensive experimentation, hands-on construction, and critical analysis, all of which laid the foundations for understanding the imperatives of advancing within these fields. The methods of trial and error, parameter tuning, and synthesis of new images acted pivotally in the development of these architectures.

Moving to the second semester, the focus shifted to the realm of Style Transfer, a fascinating domain that narrows the gap between artificially and artistically generated images. In this stage, a real-time style transfer model was developed and applied. The goal was to achieve high-quality stylistic transformations of images, potentially applied practically in real-world scenarios. However, the challenge was not only preserving the content and integrity of the ground-truth images but also doing a smooth implementation of artistic styles in a real-time context, which required innovative solutions and methodologies.

Following the completion of these two intensive research and development phases, the project continued its advancement with a comparative analysis stage, which was crucial for gathering insights and recognising the strengths and limitations of each domain. This stage included evaluating the quality of the obtained results, the performance of the models, and their applicability to real-world scenarios.

Ultimately, the methodology of this study outlines a path for research and implementation of advanced AI technologies. Furthermore, it reflects the precise and methodical approach of this project and results in pushing the boundaries of what is possible with current artificial intelligence in the realm of generative models.

3.1 Data and Preprocessing

This section discusses the various datasets used throughout the evolution of this project, each of which has undergone a rigorous ethical approval process by the university. Primarily, the CelebA dataset [13] acted as a benchmark when carrying out the experiments for standard GAN, Wasserstein GAN with weight clipping, Conditional Wasserstein GAN, and Wasserstein GAN with Fréchet Inception Distance. The intended mock Image Transformer model was supposed to employ the pix2pix architecture alongside its Kaggle dataset [14] for an effective image translation process. However, because model implementation was not successful, the dataset went unused. Nevertheless, this foundational work paved the way for choosing the COCO-2017 dataset [15] when building the real-time style transfer model. All datasets are subject to careful pre-processing steps, which set the stage for accurate recognition, data processing, and therefore optimal

performance from the models.

3.1.1 CelebA dataset

Chosen for its compatibility with Generative Adversarial Networks, the CelebA dataset [13] stands out due to its large scale and well-annotated collection of 202,599 celebrity face images. It is highly diverse, including complex backgrounds and 40 binary attributes that cover age, gender, and hair characteristics. This dataset serves as a universal tool for many computer vision tasks, such as face detection, attribute recognition, and facial landmark localization [13]. The processed samples from the dataset are transformed into three-channel images with 64x64 resolution samples and suggest a solid foundation for evaluative and analytical purposes.

3.1.2 COCO-2017 dataset

Microsoft's COCO-2017 (Common Objects in Context) dataset [15] comes in handy for the development of the real-time style transfer model. It covers over 330k images, 200k of which are annotated for various tasks, including object detection, instance segmentation, and keypoint detection. It is segmented into three primary subsets: Train2017, Val2017, and Test2017 [15]. The broad use of COCO in training and evaluating deep learning models in numerous domains highlights its value. Concretely, for this project, the dataset supports the development of a style transfer model. The resulting images are rendered at 256x256 resolution, offering a highly detailed canvas for stylistic innovation and transformation.

4 GAN Implementation and Experiments

Following the dataset selection, the subsequent phases involve the creation and implementation of the respective Generative Adversarial Networks, mock Image-to-Image transformation model, and Real-time Style Transfer model.

GAN experiments had a crucial contribution during the development trajectory of this project. Selecting an appropriate loss function determines the quality of output and has the potential to provide a unified approach that applies to different tasks. Various loss functions were examined during the first semester, ranging from cross-entropy to Wasserstein loss. However, while loss functions can be compared as an internal measure that evaluates a model's accuracy in approximating a desired output, performance metrics were employed post-training to evaluate the model's accuracy in predicting unseen data sets. Such metrics were Precision and Recall, Inception Score (IS), Fréchet Inception Distance (FID) [33]. By applying such procedures, different evaluation criteria were explored during the development of the project, which led to comprehensive testing and assessment of the built models.

4.1 Standard GAN

Implementation:

In the initial experiment, a standard generative adversarial network was used. It consists of two main elements: a generator and a discriminator. The generator network starts with a noisy input and refines it through multiple convolutional layers, each of which is complemented with batch normalisation and ReLU activation to produce images. Meanwhile, the discriminator network evaluates these images, striving to distinguish between real images and fakes created by the generator [3]. The training process utilises Adam Optimizer, set to a learning rate of 2e–4 for both components. The training loop includes a novel adversarial process, where the discriminator maximises the probability of correctly identifying real and fake images, while the generator aims to produce images that the discriminator classifies as real. The misalignment between the discriminator's estimates and the actual labels is measured using the binary cross-entropy loss function, which is defined in *Figure 4*. Essentially, the function calculates a portion of the loss based on the actual label and the predicted probability. If the prediction is accurate (i.e., the predicted probability is close to the actual label), the loss is small. If the prediction is inaccurate, the loss is high [30]. The network is trained on the CelebA dataset [13]. The image sourcing using custom resizing and normalisation transformations is driven by a DataLoader. This process results in efficient clustering and shuffling during training, inspired by PyTorch's guidelines for custom datasets and transformations [27].

$$-\frac{1}{N}\sum_{i=1}^{N} y_i \cdot \log(p(y_i)) + (1 - y_i) \cdot \log(1 - p(y_i))$$

Figure 4: *Binary cross-entropy loss function: N = number of samples, p(y_i) = probability of i-th sample belonging to the class, y_i = actual label [3]*

Results:

The evolution of the standard GAN's performance is captured through snapshots of different training stages presented in *Figures 5, 6* and *7*. Initially, as depicted in Figure 5, the output of the generator deviates significantly from the authentic images, producing faces that lack definition and realism, which is expected during the early training phase.

Towards the middle of training, illustrated in Figure 6, a noticeable improvement in the quality of the generated images is observed. The images begin to feature more distinct human-like characteristics, including improved facial structures, although some imperfections suggest the ongoing training process of the generator.

Upon concluding the training, Figure 7 reveals significant improvements in the detail and variety of the generated images, including realistic facial expressions and features. The decreased discriminator loss indicates an increased challenge in differentiating between real and generated images, marking the success of the generator in producing convincing fake images.

Figure 5: *Initial state*

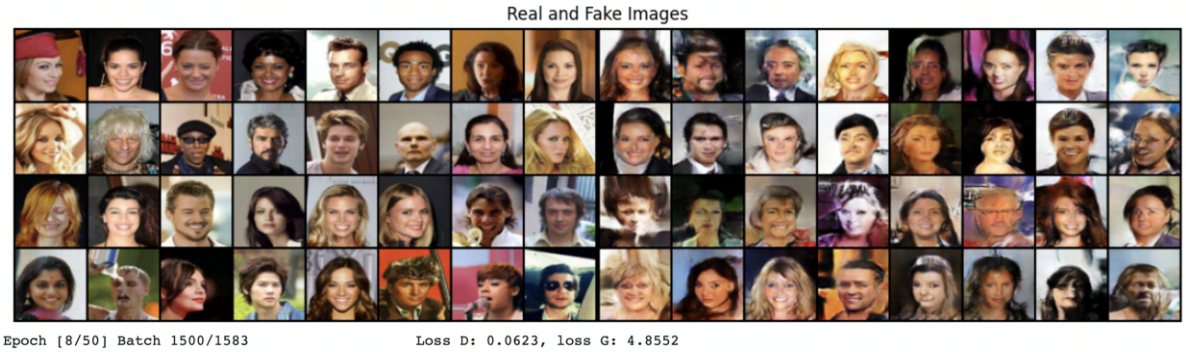


Figure 6: *Progress state*

Figure 7: *Final state*

4.2 Wasserstein GANs

The exploration of Generative Adversarial Networks was further extended with the implementation of advanced variants: the Wasserstein GAN with weight clipping, Conditional Wasserstein GAN and Wasserstein GAN with Gradient Penalty. While the initial experiments with standard GANs yielded promising image quality, the integration of critic functions, such as the 1-Lipschitz constraint in Wasserstein GANs, introduced a strategic approach to overcome some of the inherent challenges [34].

These challenges, including mode collapse and training instability and often leading to difficulties in standard GAN implementations, were addressed by applying heuristics such as weight clipping, spectral normalisation, and gradient penalty [16]. These techniques aim at not only increasing the robustness and reliability of the training process but also improving the overall quality of the generated images [33]. This is achieved with a more stable and controlled model learning environment.

After an extensive examination of various loss functions during the first semester, the prioritisation felt on Wasserstein/Earth Mover's Distance. This was due to its use of similarity-preserving learning and quantization error control. The formula is defined in *Figure 8*, and it represents the minimum required cost for transforming a single probability distribution into another [18].

$$W(P_r, P_g) = \inf_{\gamma \in \Pi(P_r, P_g)} \mathbb{E}_{(x,y)\sim\gamma}[\|x - y\|]$$

Figure 8: *Earth-Mover (Wasserstein) distance: P_r = real data distribution, P_g = model generated data, inf = infimum, Expected value E = computation of mean distance between paired data points x and y[18]*

4.2.1 Wasserstein GAN with Weight Clipping

Implementation:

The Wasserstein GAN with weight clipping framework consists of two main components: a generator and a discriminator. The generator initiates the process with a latent vector z that expands into a full-dimensional image by using transposed convolution layers. This expansion process is improved by batch normalisation and ReLU activation for all layers except the last one, where Tanh activation is applied instead. As opposed, the discriminator evaluates the input images and provides a scalar value, indicating their assumed authenticity. This evaluation involves multiple convolutional layers, each of which is coupled with LeakyReLU activation to introduce nonlinearity, along with batch normalisation to ensure robust learning. A key aspect of this model is integrating the Lipschitz constraint, which is essential for the computation of the Wasserstein distance achieved by weight clipping [18]. The formulation of the process is depicted in *Fig 9* below. The training alternates between refinement of the discriminator and generator and employs the RMSprop optimizer. This technique adjusts each gradient by the square root of the moving average, set with a learning rate of 5e-5 for this project. The weight initialization used follows a normal distribution to avoid suboptimal starting points. During the training process, periodic inspections of real and fake images are carried out to qualitatively assess the performance of the generator. Such checkpoints are essential for verifying the gradual refinement of generated images and ensuring they increasingly reflect the attributes of the CelebA dataset [13] as training progresses.

$$\min_{G} \max_{\|f\|_L \leq 1} \mathbb{E}[f(x)] - \mathbb{E}[f(\tilde{x})]$$

Figure 9: *Cost Function of a WGAN with Weight Clipping [18]*

Results:

The results obtained from training the Wasserstein GAN with weight clipping did not fully meet expectations. Challenges were encountered in terms of optimisations, with the technique sometimes leading to complications. Moreover, the relationship between the weight constraint and the loss function often results in either vanishing or exploding gradients unless thoroughly corrected [16]. This highlighted a significant drawback of the weight-clipping approach.

The training progress of the model mirrors that of standard GAN, with *Figure 10* showing the early results, *Figure 11* illustrating the intermediate stages, and *Figure 12* displaying the final results. Significant progress is observed, especially after 50 epochs, indicating some level of success. Nevertheless, the ongoing shortcomings necessitate further improvements in the architecture and training methodology.

Figure 10: *Early results*

Figure 11: *Intermediate results*

Figure 12: *Final results*

4.2.2 Conditional Wasserstein GAN

The development of a Conditional Wasserstein GAN was an exploratory step to incorporate specific features and broaden the research horizons of the project. This variation enriches the experiment by introducing additional inputs into the model, such as class labels or other data types. This allows the architecture to be "conditioned" by this additional information, granting enhanced control over the characteristics of the resulting images [16]. While the underlying cost function mirrors that of Wasserstein GAN with weight clipping, it is augmented with a conditional vector c that helps to tailor the data generation process, as shown in *Figure 13*. The model's formula features the same objective function as a min-max game between the generator and discriminator, but it also contains the conditioning variable c, which allows the architecture to control the generated data based on known labels or data attributes [18].

$$\min_{G} \max_{\|f\|_L \leq 1} \mathbb{E}_{x,c}[f(x,c)] - \mathbb{E}_{\tilde{x},c}[f(\tilde{x},c)]$$

Figure 13: *Cost Function of Conditional WGAN*

Two distinct activation functions are tested within the architecture to evaluate their performance. The first, ReLU, faces challenges related to zero-centeredness due to its inability to handle negative values. This leads to a condition where the neuron's weights fail to update, referred to as "dying ReLU ." To mitigate this issue, Leaky ReLU is introduced as an alternative. Leaky ReLU offers a benefit over its predecessor as it allows a small gradient when the neuron's input is negative. Thereby, it supports the activation of neurons that might otherwise become dormant.

Results:

Despite these modifications, the Conditional Wasserstein GAN does not outperform the weight-clipping Wasserstein GAN in terms of image quality. Nonetheless, this iteration is invaluable for benchmarking, offering insights into the applicability of conditional models to the CelebA dataset [13]. The comparison between the two activation functions is

depicted in *Figure 14*. This experimentation also allows a familiarisation with vanishing and exploding gradients, as illustrated in *Figure 15*.

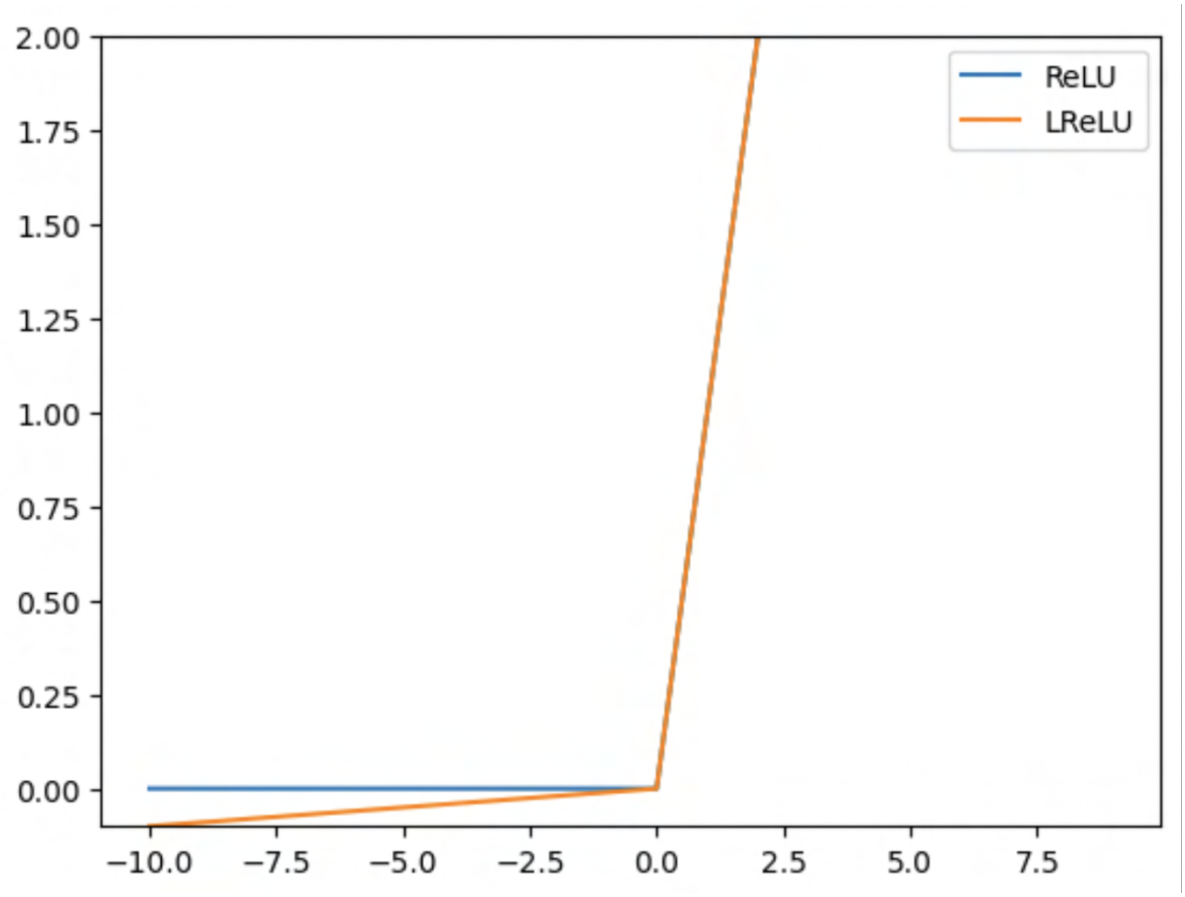


Figure 14: *ReLU and Leaky ReLU activation functions*

```
Epoch [8/10] Batch 1500/1583          Loss D: -178907.6094, loss G: 81721.8438
Epoch [9/10] Batch 0/1583           Loss D: -175913.5000, loss G: 82521.8594
Epoch [9/10] Batch 100/1583           Loss D: -187195.8750, loss G: 85896.7344
Epoch [9/10] Batch 200/1583           Loss D: -183791.7812, loss G: 84875.2969
Epoch [9/10] Batch 300/1583           Loss D: -189654.4219, loss G: 87523.7656
Epoch [9/10] Batch 400/1583           Loss D: -189019.1562, loss G: 88741.2656
Epoch [9/10] Batch 500/1583           Loss D: -125108.5938, loss G: -41000.6562
```

Figure 15: *Vanshing and Exploding gradients in Conditional WGAN*

This exploration emphasises the need to search for more sophisticated solutions, which led to a switch to Wasserstein GAN with a Gradient Penalty (WGAN-GP). This shift was driven by its reputation as a higher-quality image generator, suggesting a promising direction for overcoming the limitations observed in this and previously described models.

4.2.3 Wasserstein GAN with Gradient Penalty

Implementation:

While the Wasserstein GAN with weight clipping improves training stability, in many cases, it still generates poor samples or fails to converge. This challenge led to the adoption of the gradient penalty mechanism, resulting in the creation of the Wasserstein GAN with Gradient Penalty (WGAN-GP). This model uses the Wasserstein loss to compute the Earth-Mover distance but also introduces a gradient penalty over weight clipping to enforce the Lipschitz condition [16]. The formula for the architecture is illustrated in *Figure 16*. This combination facilitates smoother network learning than its predecessors and minimises the need for extensive hyperparameter adjustments [16].

$$\mathcal{L} = \mathbb{E}_{\tilde{x}\sim P_g}[f(\tilde{x})] - \mathbb{E}_{x\sim P_r}[f(x)] + \lambda \mathbb{E}_{\hat{x}\sim P_{\hat{x}}}\left[(\|\nabla_{\hat{x}} f(\hat{x})\|_2 - 1)^2\right]$$

Figure 16: *Loss function for gradient penalty: First term = expected value for generated data, second = real data, third = gradient norm maintaining Lipschitz continuity [18]*

This modification improves the robustness of the training and circumvents problems such as vanishing or exploding gradients. It ensures that the critic is well trained by supplying the generator with useful gradients. Moreover, it avoids common pitfalls such as mode collapse seen in traditional GANs [16]. Similar to other models, it includes a discriminator built on a convolutional neural network and a generator using a transposed convolutional neural network. By integrating a gradient penalty term, WGAN-GP develops a GAN learning methodology, offering a robust and consistent approach that outperforms both original GAN and WGAN with weight clipping. The inclusion of summary writers and checkpoints is pivotal for tracking progress. Consequently, this facilitates ongoing training adjustments, which is essential for refining the model to produce better results [16]. The model took inspiration from a GitHub implementation [19].

Results:

The outcomes derived from training WGAN with a Gradient Penalty demonstrate a significant improvement in visual attractiveness compared to the results from WGAN with weight clipping and Conditional WGAN. These results are methodically shown in three distinct stages of the learning process: the initial output, depicted in *Figure 17*, the mid-training output, presented in *Figure 18*, and the final output, illustrated in *Figure 19*. This structured representation allows for clear monitoring of model progress over time.

WGAN-GP marks a key achievement in terms of providing a more robust training regime and enhancing the overall quality of generated images. This advancement in generative modelling illustrates the model's ability to produce more realistic and visually appealing

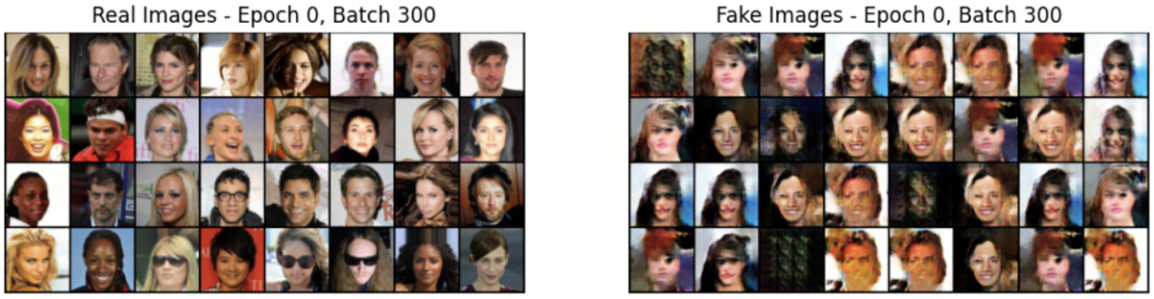


Figure 17: *Initial output*

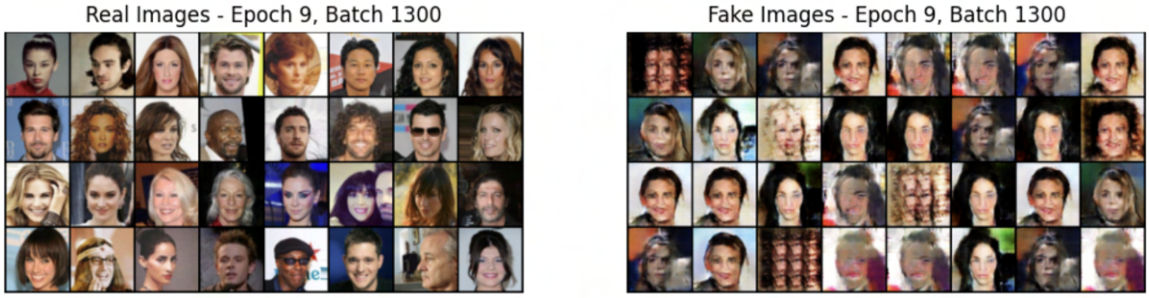


Figure 18: *Mid-training output*

images, setting a new benchmark for GAN architectures. Despite these improvements, it is important to note that the quality of the images, while noticeably better, is still not completely accurate and is indistinguishable from the actual images. This gap in achieving absolute realism highlights the ongoing challenges in the field of generative adversarial networks [16]. All this leads to experimenting with a more comprehensive and precise measurement of the similarity between generated and authentic images, namely Fréchet Inception Distance (FID). Consequently, new GAN models incorporating this metric were developed to further push the boundaries of this study.

5 WGAN using Fréchet Inception Distance

Implementation:

The Fréchet Inception Distance (FID) stands as a pivotal metric for assessing GANs' performance. It measures the Wasserstein-2 distance between two multivariate Gaussian distributions that represent the feature spaces of both real and synthetic images. The fundamental principle of FID states that an effective GAN should replicate feature distributions of authentic images [17]. These features should derive from a pre-trained classification model. The formula for the metric is depicted in *Figure 20*. A lower FID result means a closer approximation to the true images, highlighting the accuracy of the model. However, FID is not without its limitations, including sensitivity to the classifier used for feature extraction and the inability to fully capture the diversity in the image set [17].

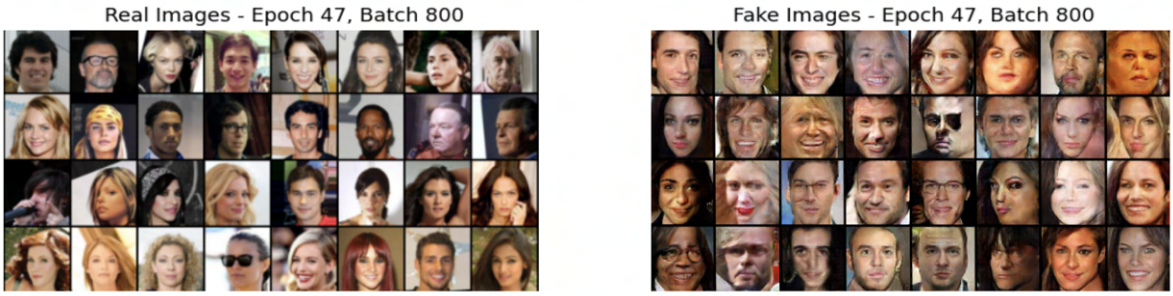


Figure 19: *Final output*

$$\mathrm{FID}(r, g) = \|\mu_r - \mu_g\|_2^2 + \mathrm{Tr}\left(\Sigma_r + \Sigma_g - 2\sqrt{\Sigma_r \Sigma_g}\right)$$

Figure 20: *Fréchet Inception Distance equation: L2 = squared Euclidean norm, a trace of the sum of the two covariance matrices minus twice the square root of their product [17]*

Within the context of this paper, applying the Fréchet Inception Distance to evaluate the Wasserstein GAN yields insightful and measurable results. The model developed during this research accurately reflects the quality of images found in the CelebA dataset [13], demonstrating its ability to generate realistic images. The architecture of this model integrates a discriminator and a generator, each of which is designed with specific blocks to proficiently process image data. Initialising the model weights according to a normal distribution with a mean of zero and a standard deviation of 0.02 proves essential for the stable training process of the model. It ensures a balanced learning regime and helps prevent vanishing or exploding gradients. Furthermore, the adoption of batch image processing via DataLoader, incorporating essential transformations like resizing, cropping, normalisation, and tensor conversion, plays a key role in ensuring a high-quality model output. The training methodology includes vital steps such as backward propagation for both networks, which ensures that they learn from their errors and improve over time. Furthermore, the incorporation of weight clipping for the discriminator ensures a robust and effective learning process.

Results:

Upon completing the training phase, it is observed that the model architecture becomes more sensitive to variations in the input noise. This intensive training process spans a significant number of epochs, with strategic checkpoints integrated at various stages of the process. Consequently, this facilitates an overall assessment of the model's evolution and performance.

Figure 21 presents a graphical representation of the training dynamics, showing the performance of the model over time, together with the findings from Heusel et al. [17], applying the Fréchet Inception Distance (FID) across different models and datasets.

Remarkably, both graphs demonstrate that the FID results converge below a threshold of 100. This indicates a high degree of similarity between the generated images and the original ones. Thus, it confirms the effectiveness of the model in capturing the underlying data distribution.

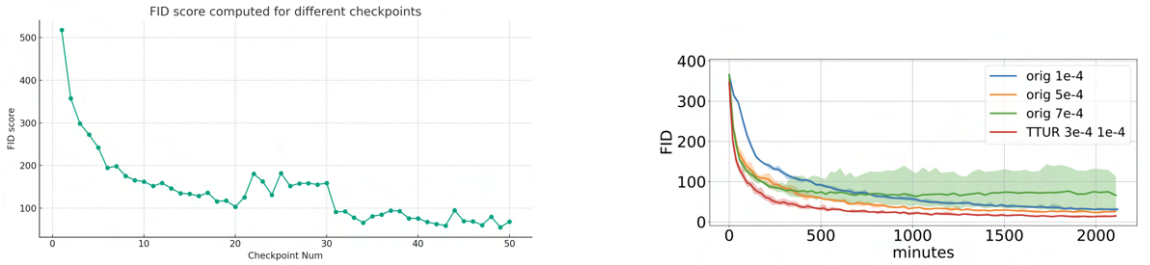


Figure 21: *Achieved training dynamics of model (left) and findings from Heusel et al. [17] (right)*

These promising results are also subjected to a comparison with the findings of a supervisor's research, ensuring accuracy and reliability. The consistency in results confirmed the efficacy of applying the FID metric to the WGAN, marking a significant achievement in the project. Such application not only underscores the model's ability to produce authentic-looking images but also establishes a benchmark for evaluating generative models. This directly contributes valuable insights into the fields of deep learning and artificial intelligence and highlights the potential of the model for future research and applications.

6 Image-to-image tranformer

Building upon the foundational knowledge gained from explorations with GANs, the development trajectory shifts towards the construction of an image-to-image transformation model. This strategic switch aimed at broadening the landscape of the project by introducing new capabilities and expanding its impact on the domain of image processing.

Implementation Challenges and Innovations:

The initial attempt to implement a transformer model encountered significant obstacles, making efforts unsuccessful. It draws inspiration from the image-to-image translation in the Conditional Adversarial Networks paper [4], undergoing modifications to enhance its capabilities. The intended output after constructing the model is illustrated in *Figure 22*. To achieve such output, the transformer is designed to go beyond the standard generator and discriminator configuration, incorporating additional elements such as a variational autoencoder to extend its analytical depth. The implemented loss function is carefully constructed to include several components: GAN loss for adversary training, feature matching loss to ensure consistency between different layers, VGG loss to estimate the Euclidean distance between the feature maps of the generated and real images, and KL divergence loss to measure the difference between the encoder output distributions.

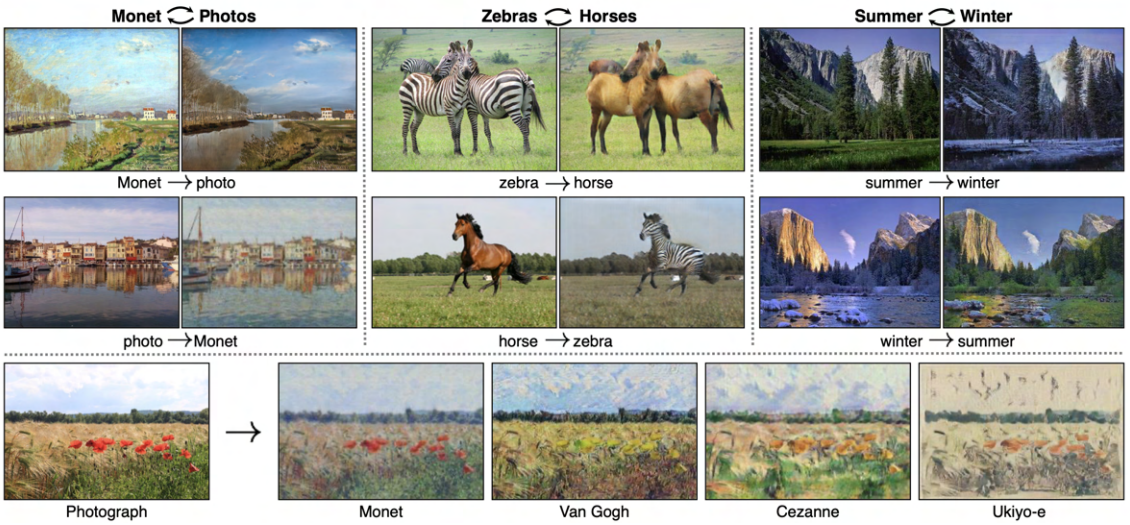


Figure 22: *Intended Image-to-image transformer output [4]*

The training process was fine-tuned using Adam optimizers, each with distinct learning rates and betas. The preprocessing phase involved transferring tensors to the GPU and transforming label maps into one-hot encoding. In addition, useful functions were developed for tasks such as extracting edges from tensors, reparameterization in the variational autoencoder, and checking GPU availability. A class was created to facilitate visualisation and summarization of the results. A dedicated testing loop ensured thorough evaluation, enhancing the model's comprehensiveness.

Despite the significant strides made during the development, the mock image-to-image transformer failed to produce visual results. The challenge was to effectively merge all the various components required for successful training. This led to the recognition that a complete redesign and a new approach were required. However, such realisation came at a time when academic deadlines and the upcoming exam period detracted from further development of the transformer model.

After the exam period ended, the supervisor suggested shifting the focus to a neural-style transformer architecture, a suggestion that was eagerly accepted. Consequently, the image translator architecture was suspended indefinitely. This decision, although difficult, was made on the basis of a detailed analysis of the obstacles encountered and a tight academic schedule that left little room for necessary troubleshooting and rework.

However, it is important to recognise the value of the experience gained during this development. The insights obtained from navigating the challenges that have arisen and significantly contributed to a deeper understanding of image transformation technologies and laid a strong foundation for future research in this area. The knowledge gained from this experience was essential for furthering the broad goals of the project, highlighting the importance of dedication to research and adaptability in the face of setbacks.

7 Style Transfer

During the second semester, the student was fully engaged in developing a sophisticated real-time style transfer model. The process of experimentation was extensive and careful, going through many stages until satisfactory results were obtained.

7.1 Implementation and Experiments

Before proceeding with any experimental work, it was crucial to establish a solid foundation in the field of style transfer. The research paper "Perceptual Losses for Real-Time Style Transfer and Super-Resolution" by Justin Johnson, Alexandre Alahi, and Li Fei-Fei [12] provides critical insights necessary for this endeavor. Likewise, the COCO-2017 dataset [15] is used during the development and evaluation of the model for both training and testing purposes.

Upon gaining a comprehensive understanding of the mentioned paper, the preliminary attempt at creating a model fell short of producing desirable outcomes. This attempt was characterised by a real-time style transfer model employing a transformer network framework. Using a pre-trained VGG network for style transfer, with a structure depicted in *Figure 23* below, is beneficial due to its proven ability to accurately extract content and style details across various layers. The good performance of the VGG network in image classification tasks highlights its powerful feature extraction abilities. This makes it ideal for applications requiring real-time style transfer.

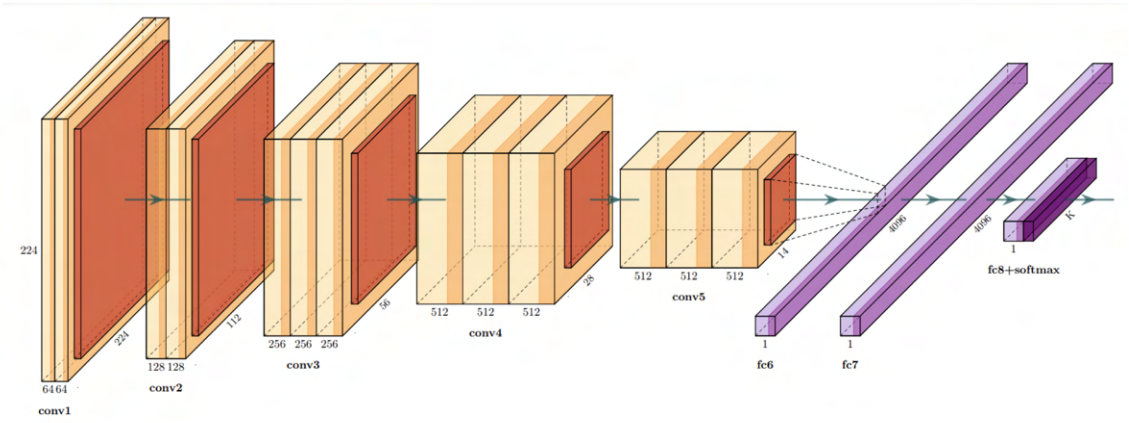


Figure 23: *VGG feature extraction structure. It uses only 3x3 convolutional layers stacked on top of each other in increasing depth. Reducing volume size is handled by max pooling [24].*

The design aims at converting input images into a stylized format by using a predetermined style image. This process involves a series of complex steps, including the use of convolutional layers, instance normalisation,and residual blocks, which do not incorporate explicit segmentation or specialised attention mechanisms. Despite its rigorous training regimen, including data processing, transformation, loss analysis, and result visualisation, the mock model fails to yield meaningful results. This shortcoming is

due to suboptimal parameter selection, which necessitates the design of an entirely new architecture.

This subsequent architecture presents a comprehensive real-time style transfer framework, utilising PyTorch for the deep learning components. It borrows concepts from the seminal work "Image Style Transfer Using Convolutional Neural Networks" by Gatys et al. [9], supplemented by additional components for training and implementing the model. This approach is also influenced by an existing GitHub repository [20]. The breakdown of the model architecture, training methodology, and auxiliary functions is defined as follows:

7.1.1 Model Architecture

- **TransformerNet:** Acts as a core unit responsible for applying the style transfer. It starts with convolutional layers for encoding the input images. Then, it incorporates residual blocks for maintaining performance at deeper levels. Furthermore, it utilises expansion layers for decoding the style-transfer features back to the original image size. Finally, it employs nonlinearity (ReLU) for introducing nonlinear transformations.
- **ConvLayer:** A custom convolution layer with reflection padding to minimise edge distortions, followed by a standard convolution procedure. This indicates an attempt to enhance the quality of the output by preserving the integrity of the image's edges.
- **ResidualBlock:** Facilitates the training of deep neural networks by providing an effective gradient flow, which suggests a lower likelihood of training instabilities as the network grows deeper. Moreover, by adding the block input to its output, also known as a "skip connection" or "shortcut," the network learns to modify the input by a residual amount rather than having to reconstruct the desired output from scratch [38].
- **UpsampleConvLayer:** An upsampling layer, followed by convolution, implies a mitigation of introducing undesirable artifacts. This process is preferred over ConvTranspose2d to eliminate checkerboard patterns in the output images.
- **LossNetwork:** Built upon VGG19, this network computes content and style loss by extracting and utilising intermediate features from the VGG19 network for the style transfer. This results in the creation of visually appealing and artistically enriched outputs.

7.1.2 Utility Functions and Training Process

- Computing the Gram matrix from input feature maps to assess the style loss, taking into account the correlations between the feature maps.
- Implementation of a tensor normalisation function based on the ImageNet dataset's mean and standard deviation for image normalisation.

- Image reconstruction converts the normalised tensors back to image format by applying the inverse normalisation obtained during preprocessing.
- The training regimen iterates over the COCO-2017 dataset [15], processing each image through TransformerNet and computing content, style, and regularisation losses. Style loss is derived from the Gram matrix of the LossNetwork's feature maps, while content loss is assessed through the mean squared error between the feature representations of the transformed and content images. The total loss is a weighted sum of these losses, with model parameter adjustments aimed at reducing this cumulative loss.
- Image preprocessing modifies images by resizing, cropping, and normalising them before passing them to the TransformerNet. As opposed, the postrocessing step recovers the image from its tensor representation. These steps ensure that images are optimally prepared for processing by TransformerNet and that their output is usable [37].
- The architecture incorporates sections for experimental training with defined hyperparameters, periodic updates on learning progress,visualisation of results, and saving of the model for future reference.
- The handling of the dataset is facilitated through specialised functions, ensuring an efficient interaction between the architecture and dataset. Consequently, this allows extensive training of the model on a wide and diverse dataset, demonstrating its adaptability to various scenarios.

7.2 Results

The results of the initial stage of development fell short of meeting anticipated standards. Unfortunately, the imagery produced at this stage fails to capture the expected realism and is categorised as poor quality. It exhibits noticeable defects, including distortion artefacts and checkerboard patterns. The framework allows limited control over the scope and specifics due to the significant influence of convolutional neural network (CNN) layers and the distribution of loss function weights. The model effectively executes style transfer on black-and-white images. However, when processing a colour image, it tends to produce the above-mentioned imperfections. Visual documentation of these initial findings is presented in *Figure 24* and *Figure 25*. The evident imperfections and the overall dissatisfaction with the result necessitate a complete re-evaluation and refinement of the real-time style transfer methodology to improve the visual quality of the results.

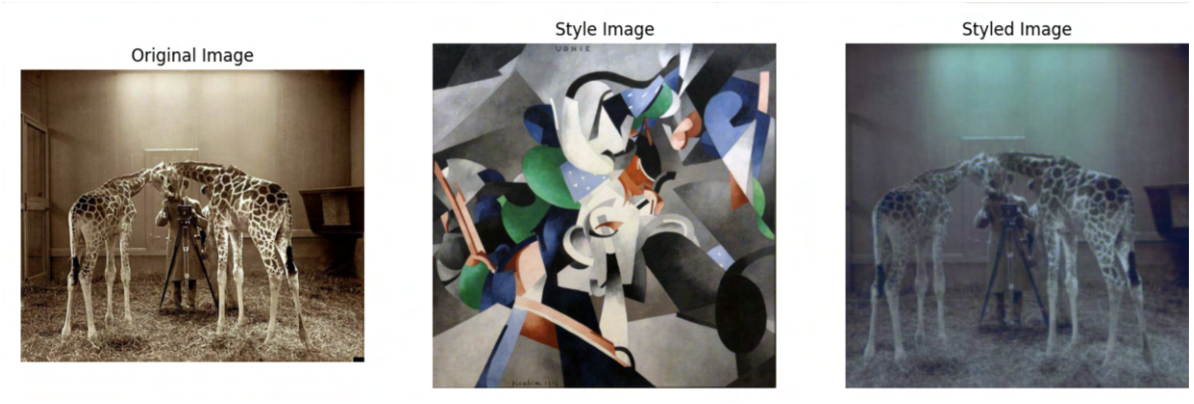


Figure 24: *Mock style transfer architecture result on black and white image using "Udnie" style image [21]*

Figure 25: *Mock style transfer architecture result on colour image using "Composition" style image*

Driven by the superior outcomes demonstrated in the GitHub repository [20], the revised real-time style transfer model results in a significant performance improvement. These refinements effectively address the shortcomings identified in the initial attempt by raising the image quality to a high standard, given the limitations of computational resources. By fine-tuning the model parameters, a delicate equilibrium is achieved between preserving the core characteristics of the original image and incorporating them with stylistic elements of the reference. This precise adjustment preserves intricate detail in the images, ensuring no vital visual information is lost.

The knowledge gained during this process enriches the understanding of the style transfer technique and its practical applications. The ability of the refined model to produce results that are consistent with the findings described in the pioneering research paper "Perceptual Loss for Real-Time Style Transfer and Superresolution" by Justin Johnson, Alexander Alahi, and Li Fei-Fei [12] demonstrates the efficacy of the refinement process. These accomplishments are captured in *Figure 26* and *Figure 27*. They highlight the

significant improvement achieved through iterative development and optimisation of the real-time style transfer model.

Figure 26: *Final style transfer architecture result using "Mosaic" style image*

Figure 27: *Final style transfer architecture result using "Picasso" style image*

The model is designed for real-time style transfer rather than extensive training periods. *Figure 28* illustrates outcomes after 1000 iterations. As opposed, *Figure 29* presents results after 300,000 iterations, which is the longest executed training period and is added to Appendix C. The model does not need a huge amount of training as it reaches its peak performance between 10,000 and 30,000 iterations, with only minimal improvements beyond this point. This characteristic highlights the model's capacity to quickly achieve peak performance compared to traditional neural style transfer models. Although there may be a slight trade-off in quality, the significant reduction in generation time is crucial for a real-time style transfer model.

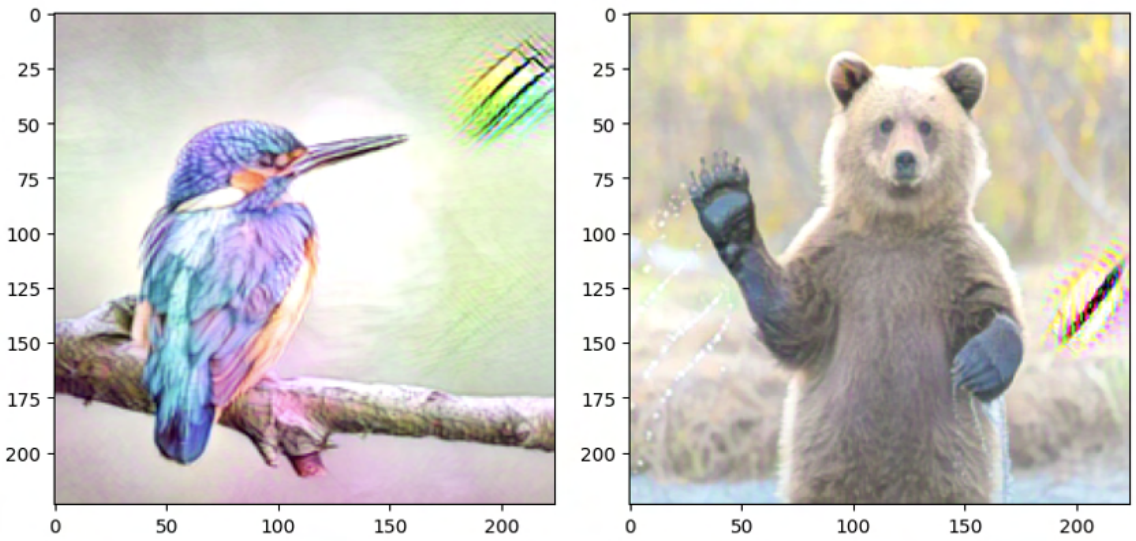


Figure 28: *Generated images after 30,000 iterations*

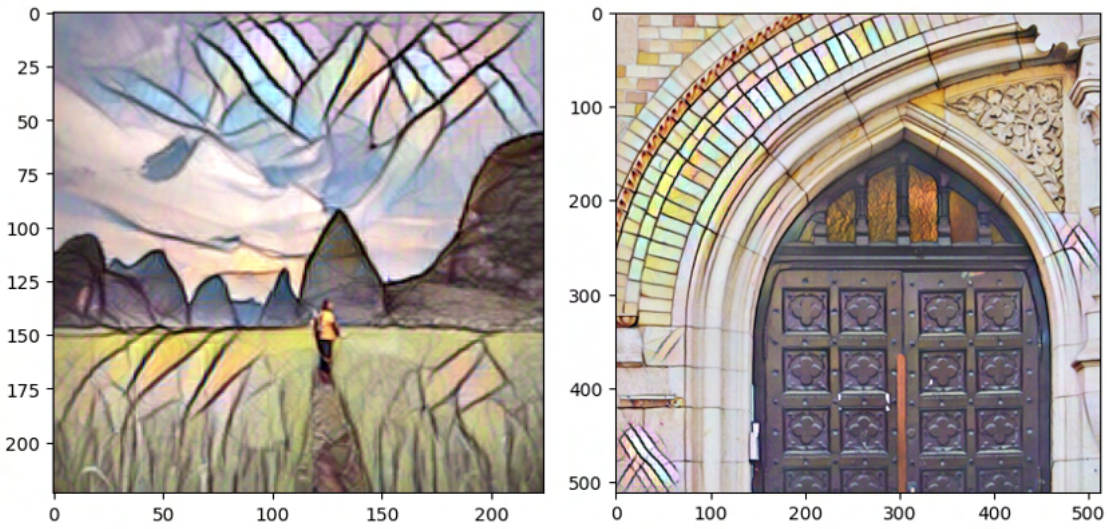


Figure 29: *Generated images after 300,000 iterations*

This process highlights the vital role of determination and adaptation in addressing initial setbacks, ultimately leading to a deeper understanding of the technology and its capacity to generate visually appealing and technically robust images.

8 Testing and Evaluation

This study explores a range of evaluation and testing methods tailored for both image generation and image transformation projects. The analysis covers both qualitative

assessments and quantitative evaluations, reflecting a holistic approach to determining the performance and accuracy of generated and transformed images.

8.1 Quantitative evaluation techniques

For quantitative assessment and evaluation, this project employs sophisticated metrics such as the Earth Mover's Distance (EMD) and Fréchet Inception Distance (FID). These metrics are essential for the statistical evaluation of the efficacy of generative models. EMD measures the minimum effort required to transform one image distribution into another, while FID evaluates the similarity between the feature distributions of authentic and synthetically generated images.

Furthermore, the stability and progress of the models are monitored by loss tracking, which includes mechanisms such as weight clipping and gradient penalty. Improvements include the use of GAN, feature matching, and VGG-19 loss to further refine the models. GAN loss evaluates how effectively the generator fools the discriminator. Feature Matching Loss ensures the internal features of generated images match those of the target images. VGG-19 Loss calculates content and style loss using a gram matrix to measure correlations and estimate perceptual similarity.

All experimental activities and results are systematically recorded via Comet and are illustrated in *Figures 30, 31, 32, 33* and *34*. *Figure 30* displays a standard GAN experiment panel over a few epochs, highlighting discriminator and generator losses. A significant increase in the discriminator loss is observed around epoch 2, indicating potential instability. This problem is resolved in the following epochs, as indicated by a decrease in loss. Meanwhile, the generator loss demonstrates a generally decreasing trend with some instability. The loss is lower in epoch 6 compared to epoch 0, suggesting an optimisation in deceiving the discriminator. Nevertheless, there are training instabilities indicating the need for more stable training methods, such as adjusting the learning rate, using a gradient penalty, or exploring different architectures to prevent such instabilities.

Figure 31 documents the extensive training progression of a Conditional Wasserstein GAN with Gradient Penalty. The critic's loss begins with a negative value and has an upward trend, indicating an improving ability to distinguish real from generated images. However, such variability suggests potential instability or variable difficulty levels throughout the training. In contrast, the generator loss shows a decreasing trend, reflecting improvement with regards to the adversarial nature of the training and suggesting a cautious interpretation of the WGAN-GP setup. The combined loss can be seen in Appendix D.

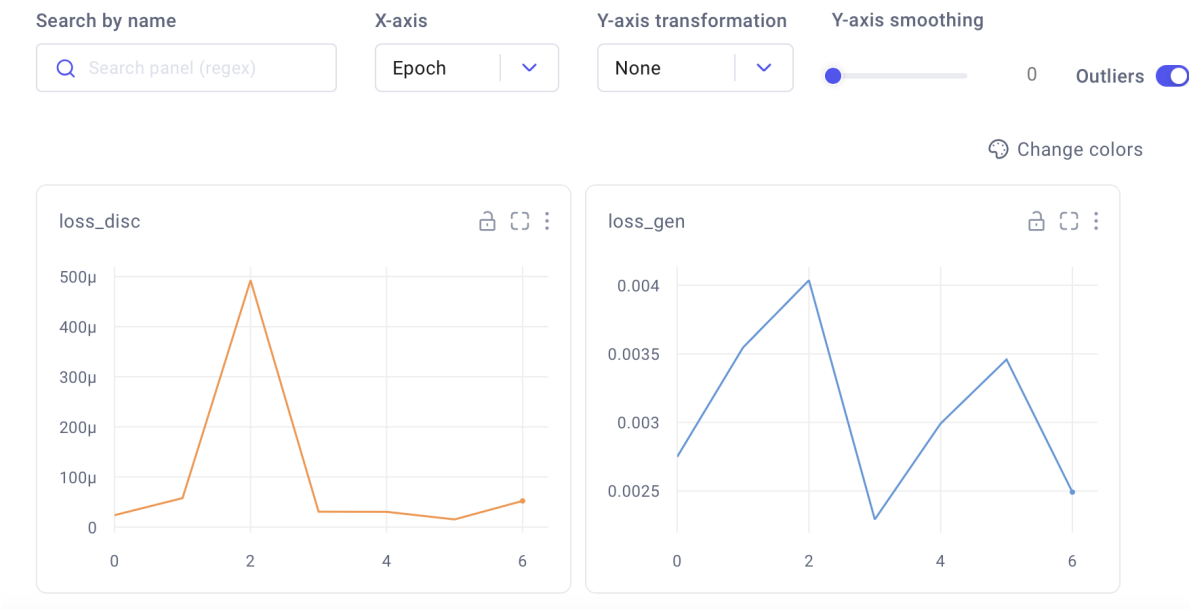


Figure 30: *Documenting epoch checkpoints for standard GAN*

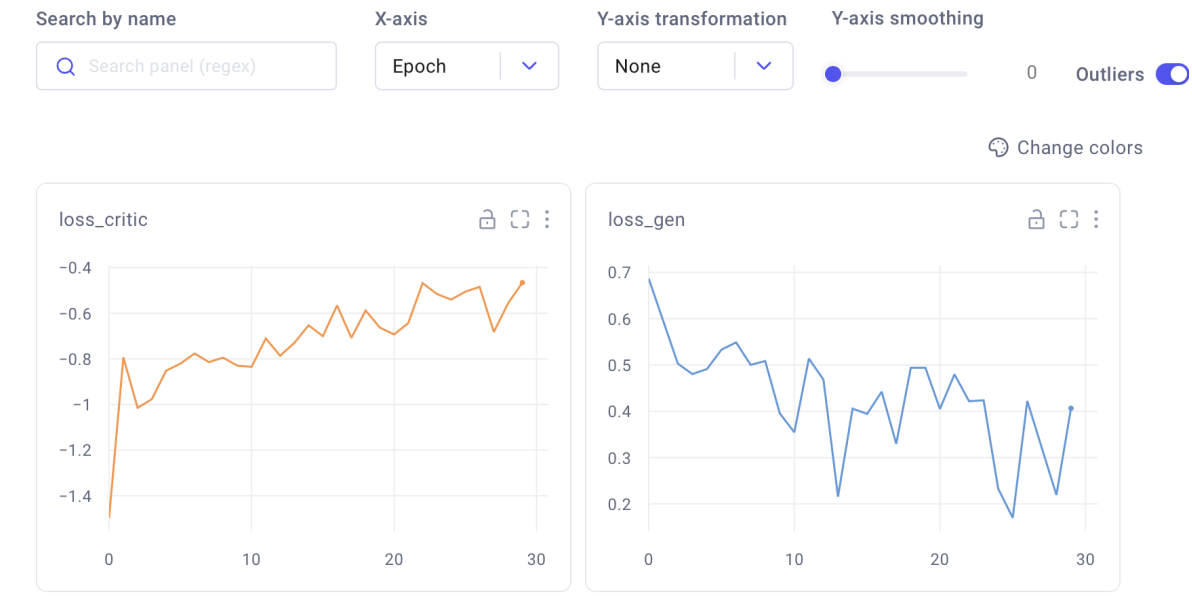


Figure 31: *Second documentation of epoch checkpoints for Conditional Wasserstein GAN with Gradient Penalty*

Figure 32 compares additional metrics such as Inception Score (IS) and FID, revealing initial differences in image quality that improve significantly with training. FID begins with high values but then drops sharply and stabilises at a lower value. This indicates that generated images become more similar to real ones over time. IS shows relatively

stable trends, reflecting steadiness and reliability in the diversity and perceived quality of the images. The stability of both metrics after initial changes suggests an approaching convergence, where further training may not significantly improve the diversity and quality of the images.

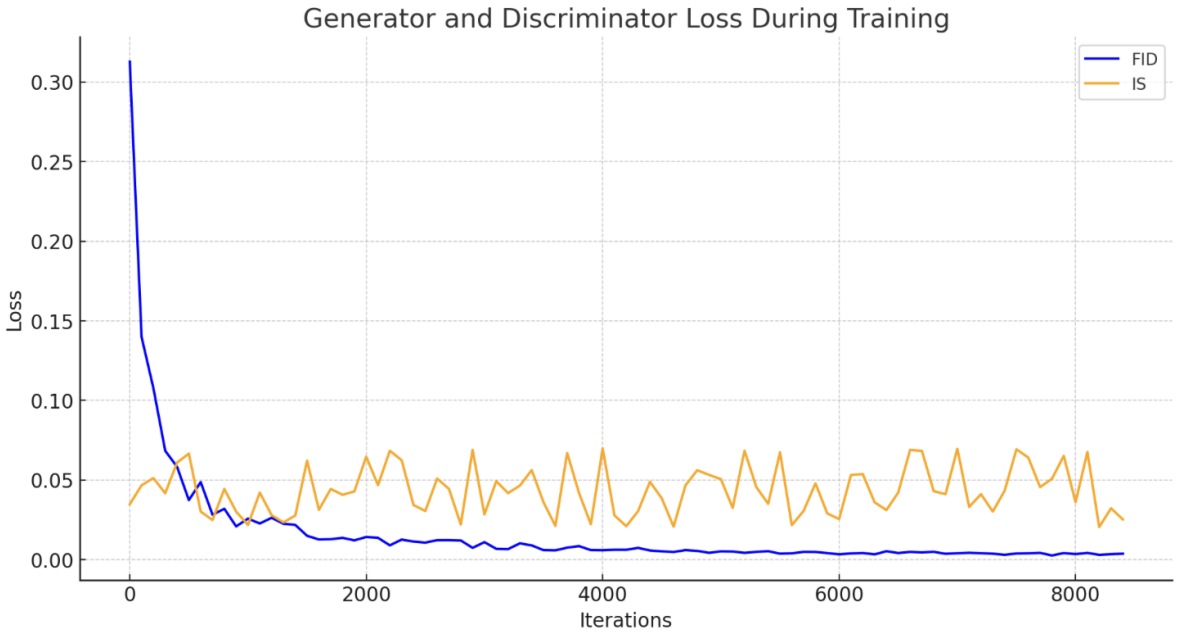


Figure 32: *Comparing trends of Inception Score and Fréchet Inception Distance*

The project also incorporates regularisation techniques that support the architecture. These techniques stabilise the learning process and ensure the stylized outputs maintain the desired aesthetic qualities, increasing the flexibility and capabilities of the model. Training progress is logged in Comet, detailed in *Figures 33* and *34*. *Figure 33* presents a substantial increase in critic or discriminator loss over epochs, indicating improvements in the generator's ability to fool the critic. Conversely, *Figure 34* displays the generator's loss, which shows fluctuations but tends to decrease overall, indicating progress in generating more convincing images.

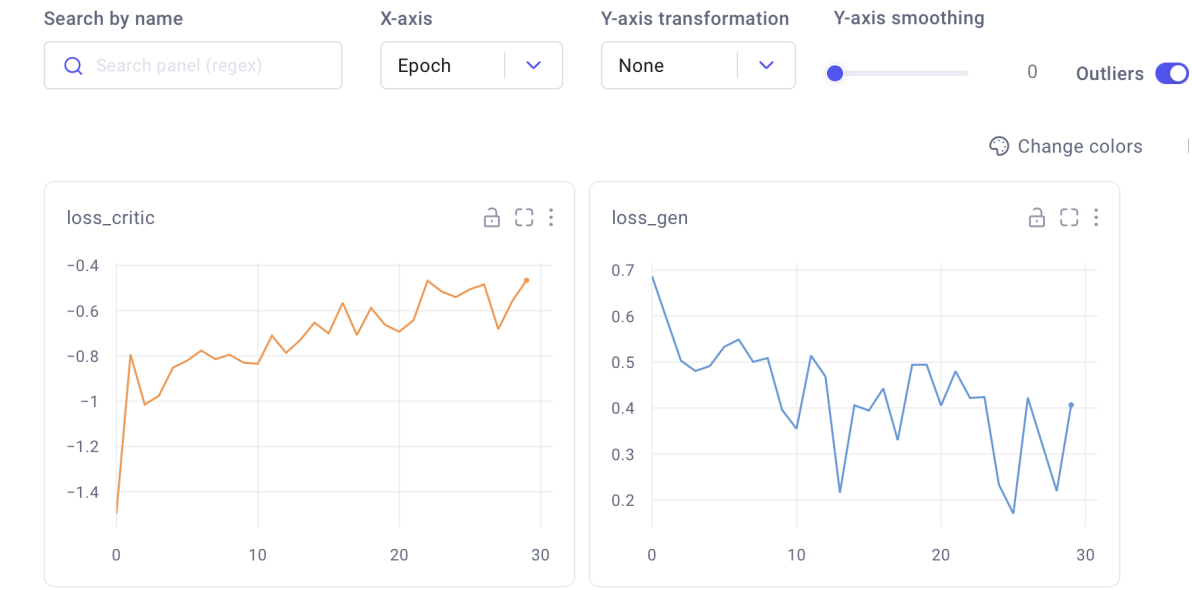


Figure 33: *First documentation of epoch checkpoints for Real-Time Style Transfer model*

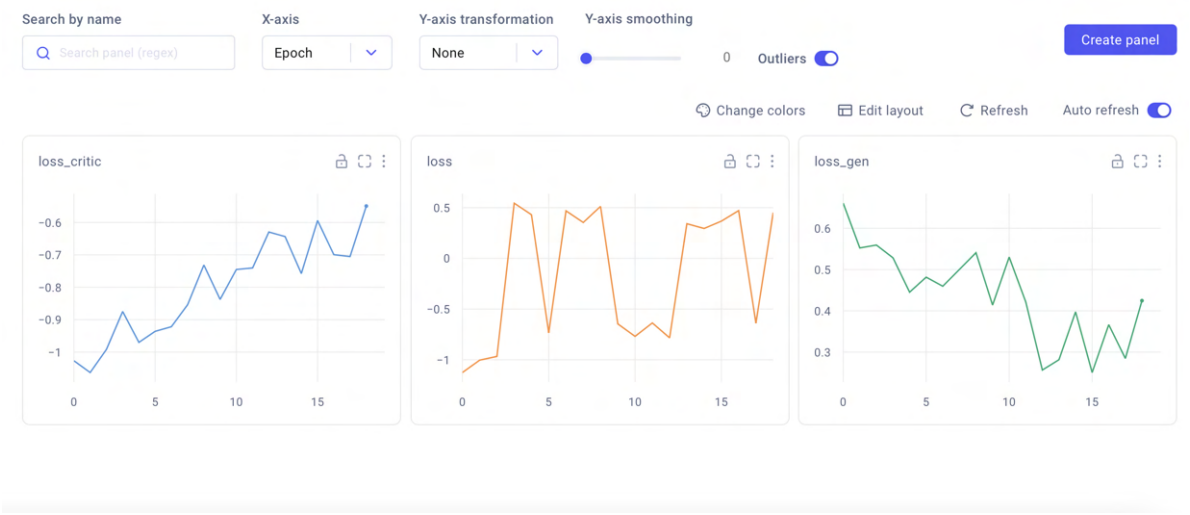


Figure 34: *Second documentation of epoch checkpoints for Real-Time Style Transfer model*

Post-processing involves the denormalization of image tensors. It uses the inverse normalisation parameters to return images to their natural state. Dynamic learning rate adjustments are also applied to optimise the network behaviour for better convergence and output quality. These adjustments are critical for the iterative learning process between the generator and the discriminator, which is fundamental to the style transfer process.

All experiments are designed to observe learning behaviour, thus allowing better visual and statistical feedback from the architectures.

8.2 Qualitative evaluation techniques

For qualitative evaluation, this project uses visual inspections to subjectively assess the quality, realism, and diversity of images. Each experiment presented in the figures includes a comparative analysis by displaying the input, output, and reference images side by side. This arrangement facilitates an in-depth visual examination to determine how well the transformations adhere to the ground-truth images. All observations gathered from these evaluations are carefully reviewed with the project supervisor. Such a process ensures investigations meet established expectations and goals.

Regarding diversity, the project evaluates the diversity of images produced by the generative models to confirm that they demonstrate adequate variability. This also ensures that the models avoid mode collapse, where the model repetitively produces a limited set of outputs. For real-time style transfer models, it is essential to maintain a consistent style across different outputs. Feature visualisation techniques are used to analyse and identify which aspects of style are effectively captured and reproduced by these models.

Additional qualitative evaluations involve the use of peers' photography as a basis for applying style images and are depicted in Appendix C. Subsequent feedback enables the artist to assess the quality of the images that the real-time style transfer model has generated from their photos. Such evaluations are extremely important as they provide insights into the artistic effectiveness of the models, where artistic value is a major factor.

Although the project covers a wide range of qualitative and quantitative evaluations, its experimental nature limits the depth of qualitative analysis possible. Nevertheless, the various evaluations carried out are crucial in ensuring that the project achieves both qualitative and quantitative standards. This offers a comprehensive exploration of the capabilities and effectiveness of the models.

8.3 User Study

Conducting a user study allows for further evaluation of the subjective attractiveness and functional characteristics of the models created. Adopting a user-centred evaluation approach ensured a comprehensive validation of both quantitative and qualitative assessment methods, providing critical insights for model refinement.

Selected images generated by the models are categorised into two groups: those produced by GANs and those from the real-time style transfer experiments. This categorization allowed for a clear comparison between the two different processes.

8.3.1 Survey Result and Analysis

The aim of the survey was to reach at least 25 participants. Ultimately, responses were collected from 28 university students and staff. The survey consisted of ten questions, beginning with a consent acknowledgment and ending with a prompt for additional comments on the methodologies. Each participant had to review a participant information sheet and a consent form, which are presented in Appendix E, together with the survey's questions and accept them before participating in the survey. All participants did accept the consent, which is illustrated in *Figure 35*.

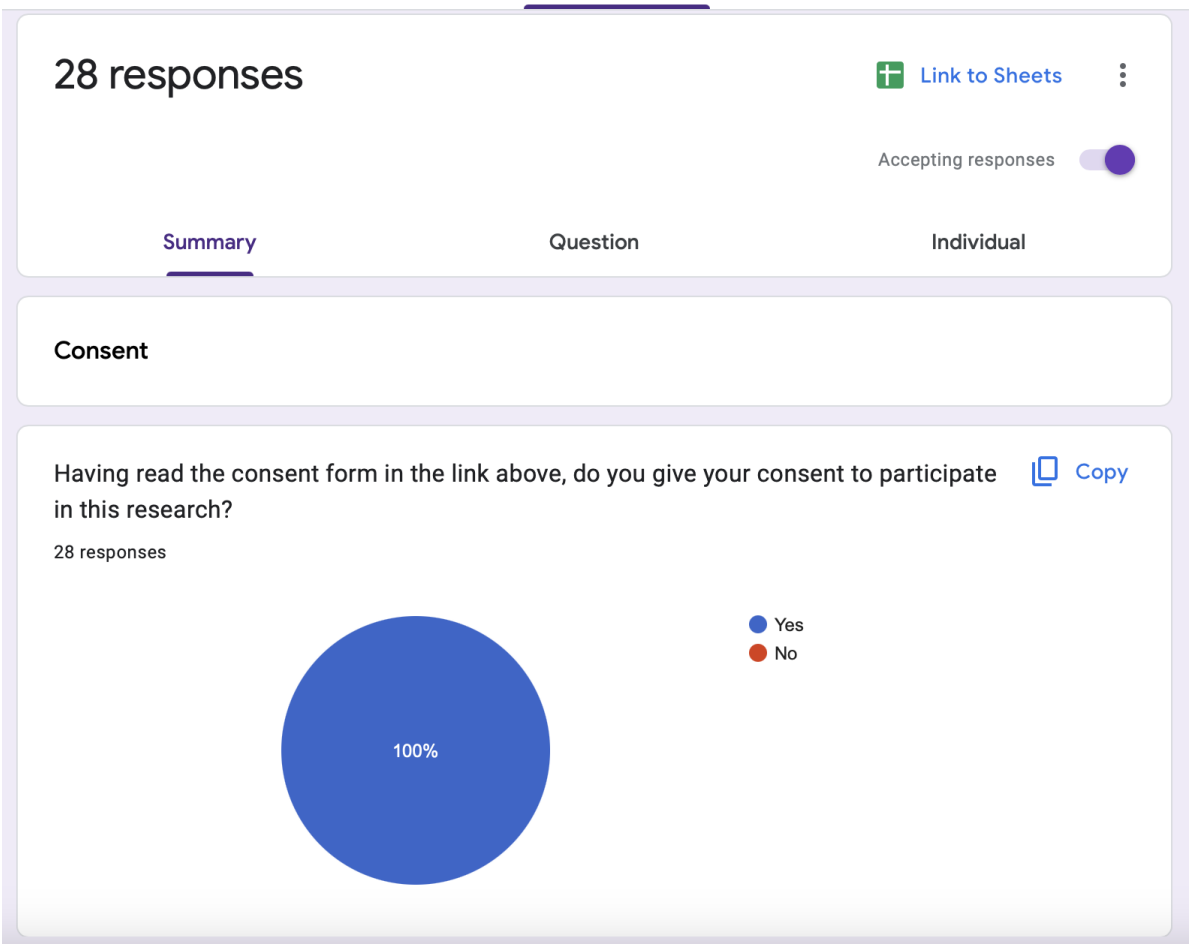


Figure 35: *Consent agreement from participants*

The survey highlighted two of the techniques applied: GANs and real-time style transfer, and showed their respective outcomes. The GAN evaluation focused on attributes such as overall image quality, realism, and textural details, while the real-time style transfer was evaluated based on the quality of style fusion, artistic impression, and uniformity.

The final part of this survey involved a comparative assessment to determine the preferred method and gather suggestions for further model improvements. An open feedback section invited participants to share their thoughts on the images and methods

used.

- The GANs' overall quality assessment ranged from 2 to 5, with 3 being the most common score, indicating respondents perceived the quality to be average. No top-tier ratings (9 or 10) were given, suggesting that the images did not reach a level of excellence in the eyes of the evaluators. These findings are documented in *Figure 36*.

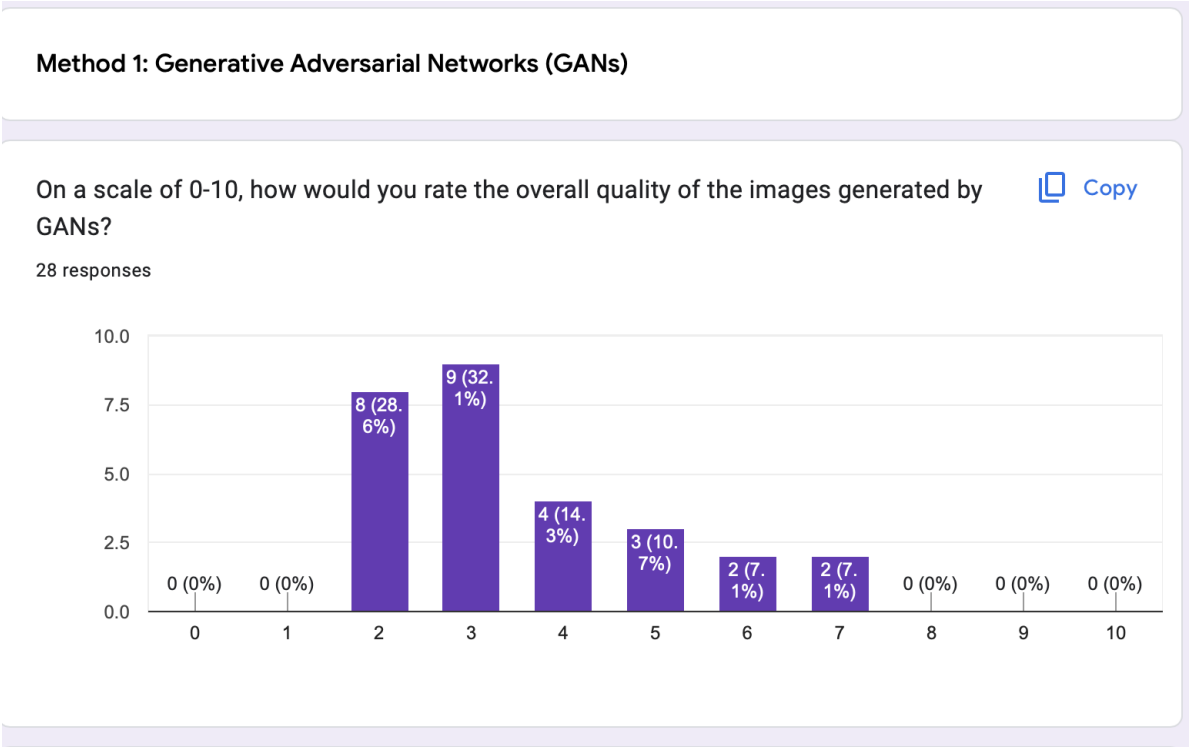


Figure 36: *GAN's overall quality assessment chart*

- The realism category showed more favourable responses, with the majority assigning a score of 6, reflecting a moderate sense of authenticity to the images generated by GAN. The concentration of scores in the middle suggests some degree of success in crafting believable visuals. These outcomes are detailed in *Figure 37*.

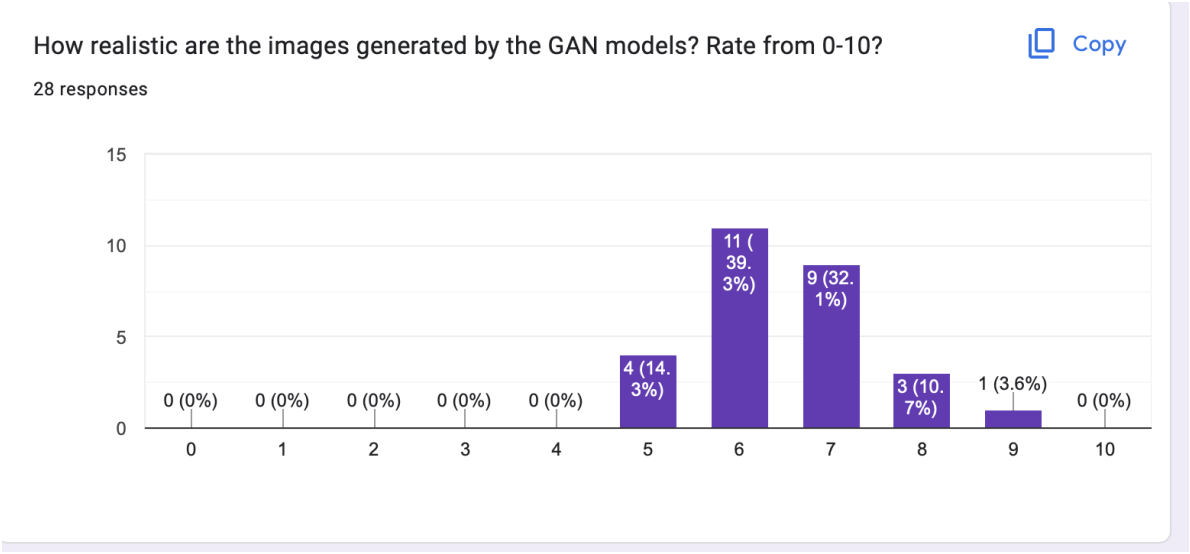


Figure 37: *GAN's realism chart*

- When it came to detail and texture fidelity, there was a notable trend towards the top scale, with most giving a score of 8. This indicates that the details and textures are generally appreciated and considered to be of high quality. These observations are described in *Figure 38*.

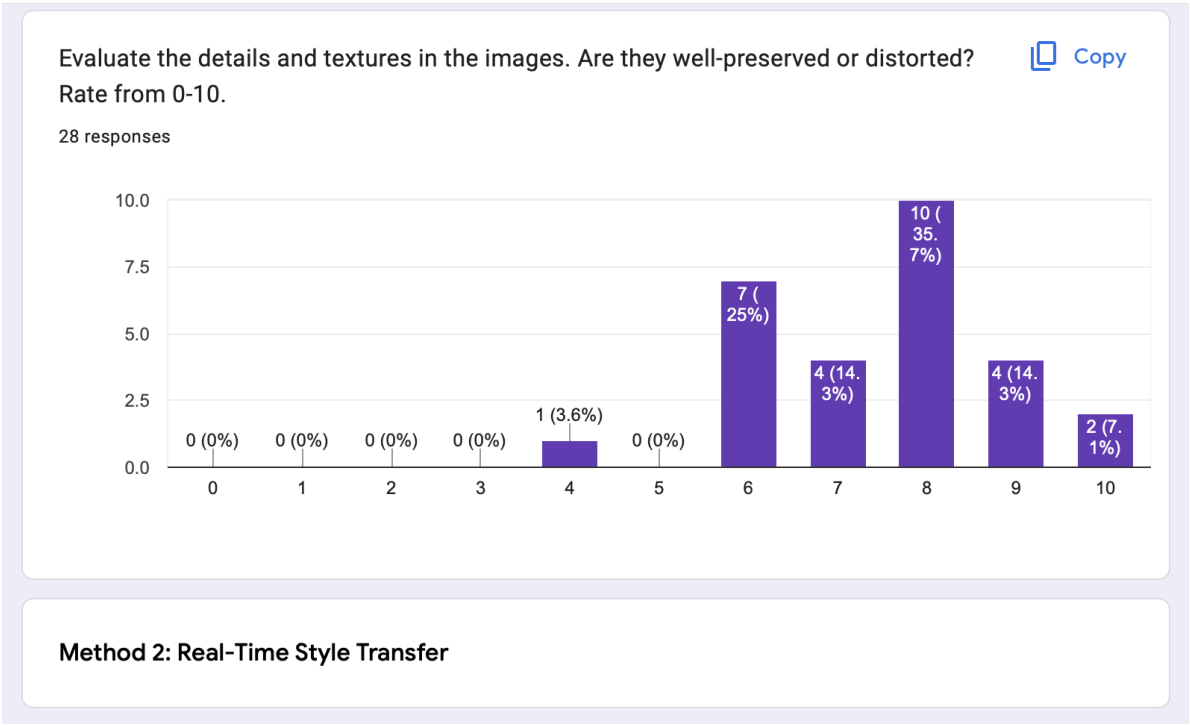


Figure 38: *GAN's detail and texture fidelity chart*

- Ratings for style integration in real-time style transfer samples tend towards the

higher end, with 9 and 10 being commonly selected. This data is shown in *Figure 39*.

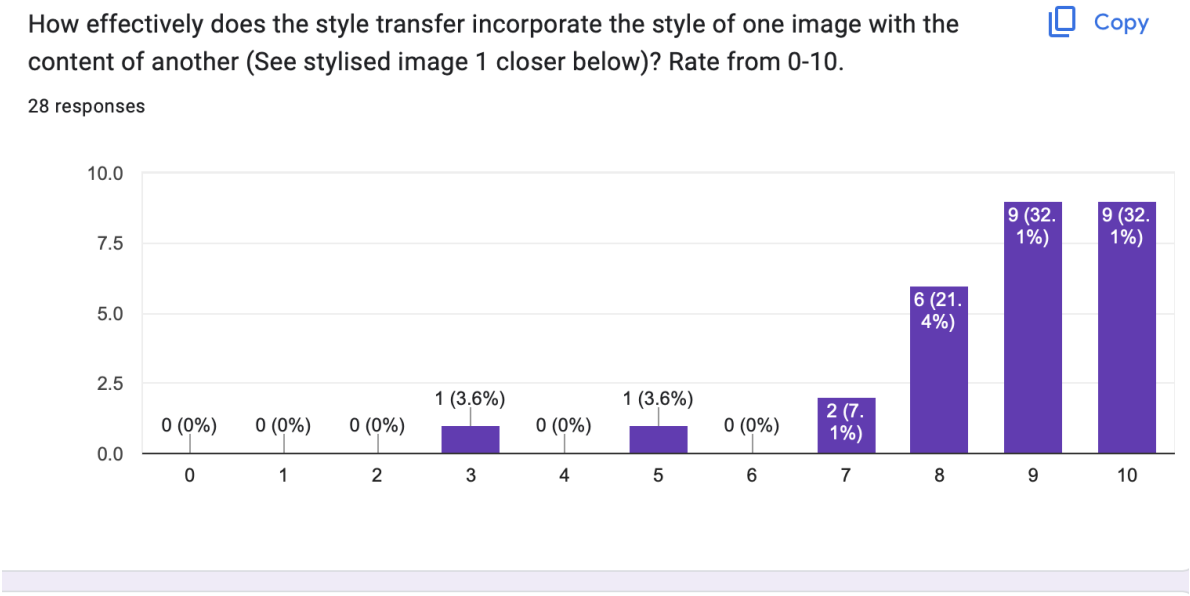


Figure 39: *Real-Time Style Transfer's style integration chart*

- The artistic impression also received high scores, with a peak score of 7, suggesting a favourable perception of the artistic elements in the style-transferred images. These results are captured in *Figure 40*.

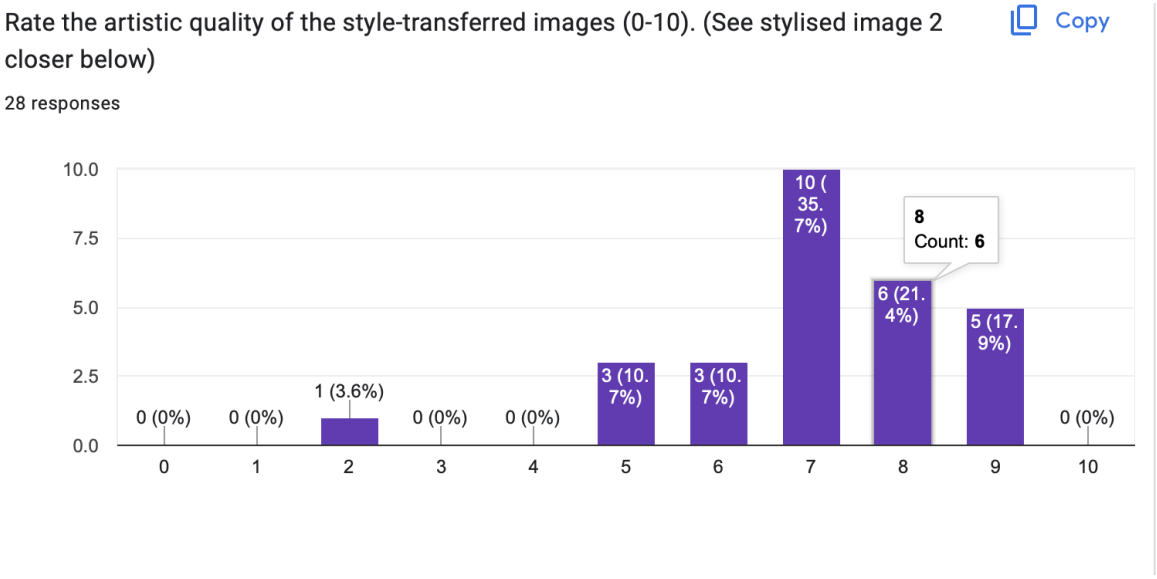


Figure 40: *Real-Time Style Transfer's artistic impression chart*

- Consistency scores vary, but the majority tend towards a higher score, with 9

being a popular choice, indicating an equal standard of quality. The findings are illustrated in *Figure 41*.

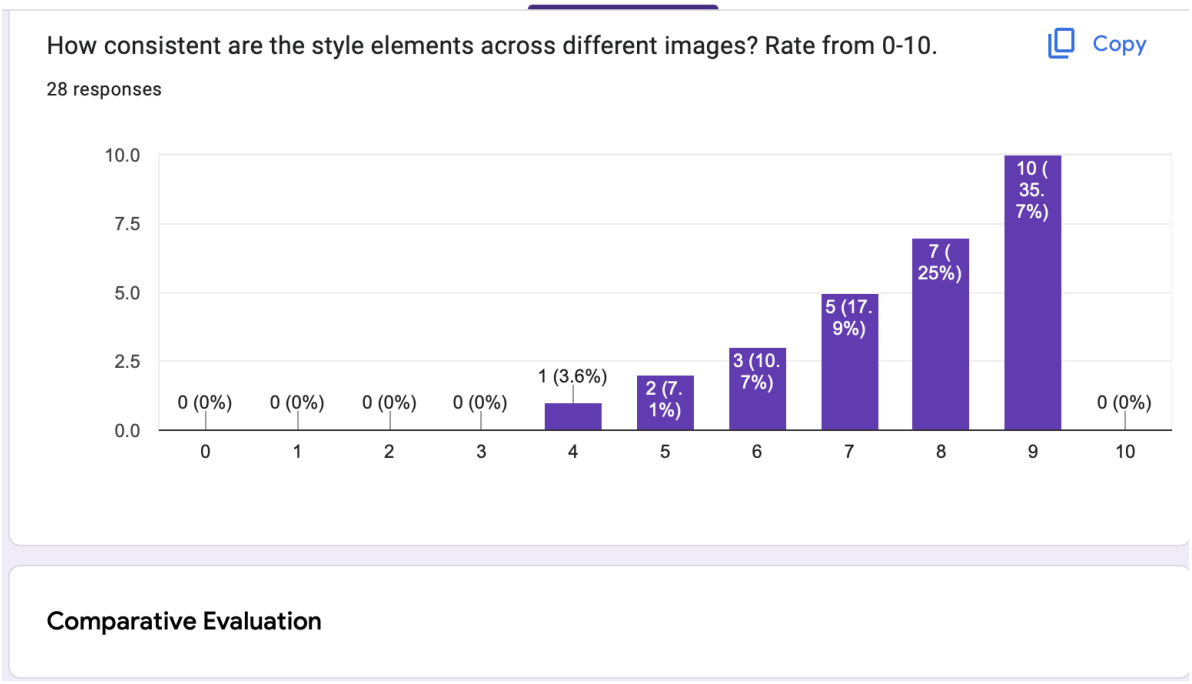


Figure 41: *Real-Time Style Transfer's consistency scores chart*

- Preference leaned towards real-time style transfer, chosen by nearly two-thirds of participants. This preference may indicate a greater appreciation for the artistic integration of style transfer compared to the realistic focus of GANs. Data supporting this conclusion is presented in *Figure 42*.

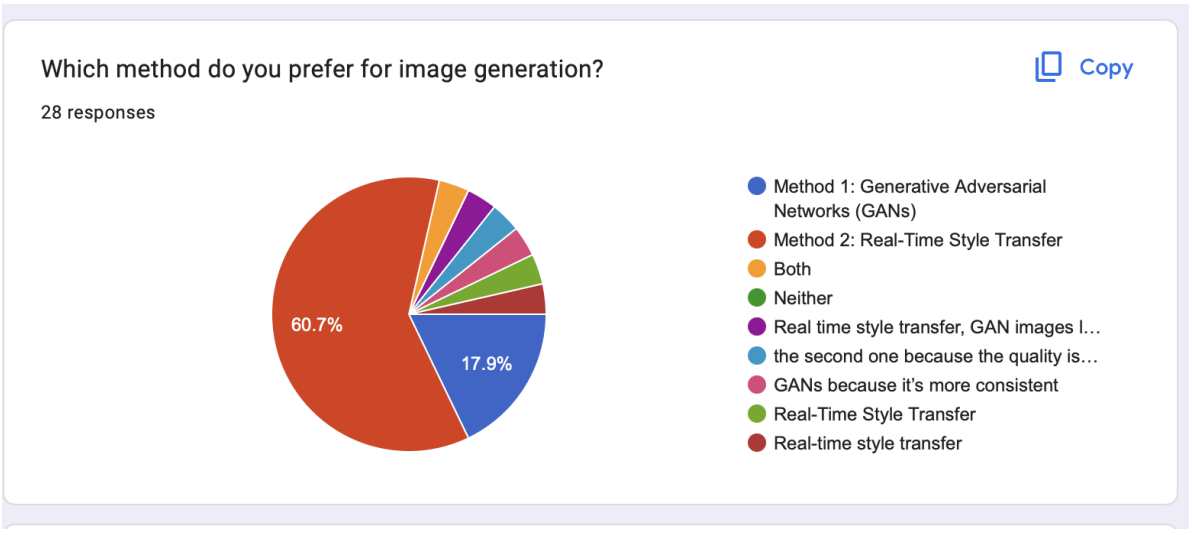


Figure 42: *Method's preference chart*

- Open-ended responses illuminated the reasons behind preferences, with some participants citing the unique artistic quality of style transfer images while others valued the iterative refinement and consistency of GANs. Suggested improvements included expanding the GAN training datasets to improve facial feature rendering and focusing on preserving details in style transfer for improved visual impact. The description of this feedback is summarised in *Figure 43*.

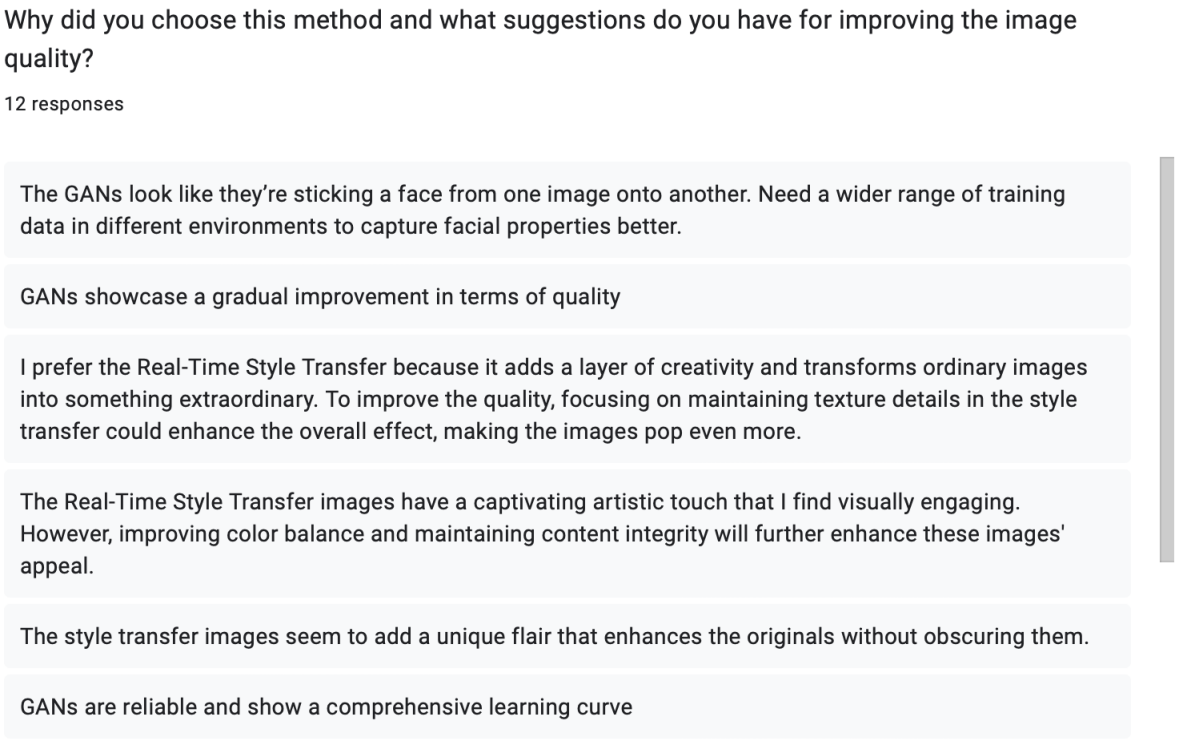


Figure 43: *Open-ended responses chart*

In conclusion, while GANs are recognised for being realistic and preserving detail, real-time style transfer is preferred for its artistic quality and effective style integration. Participants have provided constructive feedback, indicating areas for potential improvement in both methods. The preference for Real-Time Style Transfer suggests that the aesthetic appeal of the output images is a significant factor for users.

9 Progress and Project Management

This section delves deep into the progress and management of the project. It focuses on methodologies, tools, and strategies utilised, together with lessons learned. The primary aim is to reflect on understanding and provide insights that could assist future students embarking on similar academic journeys.

9.1 Project Management

The approach to managing this project involves detailed planning and agile responses to emerging challenges. It begins with a comprehensive plan and a well-articulated project brief, which can be seen in Appendix A. Although the final model deviates slightly from the original design expectations, the core aims and objectives remain in line with the overall project goals. The brief also identifies challenges that are tackled effectively.

As the project evolved, ongoing adjustments were made in response to the results of experiments and feedback from supervisors, especially as the focus shifted towards more specific areas such as style transfer. Initial plans to investigate diffusion models were set aside due to time constraints, stability concerns, an extensive training process, and limited research availability. The development process generally maintained a good pace relative to peers, often leading in terms of planning and execution.

9.2 Time Management

Effective time management was crucial to this project, with a detailed schedule outlining the trajectory of each stage. The first semester was dedicated to a literature review and experimentation with generative models. The second semester focused on reviewing literature, constructing a mock image transformer, and developing real-time style transfer models. Significant milestones include completing the progress report and preparing for the final report, with sufficient time buffers to handle unforeseen delays such as exam periods and holiday breaks.

9.3 Work Breakdown Structure (WBS)

The project was broken down into distinct phases: project planning, literature review, Generative Adversarial Networks, Image transformation models, and report writing. Each phase was subdivided into specific tasks such as setting clear goals and objectives, communicating with the supervisor, conducting extensive experimentation, and ensuring a systematic approach to the project. The work breakdown structure is illustrated in Appendix B.

9.3.1 Tools Used

Several advanced tools were used within the project scope to support model development and testing. Google Collab Pro is the primary software used due to its computational capabilities. Additionally, Kaggle and high-performance lab computers equipped with NVIDIA GeForce RTX 3070 GPUs were instrumental in handling intensive modelling and computation.

9.3.2 Skills Used

The project was both a technical and learning challenge, especially considering the lack of prior experience in the area. Therefore, it required learning new skills in genera-

tive models. These skills included working with complex modelling software, coded in PyTorch, and implementing sophisticated techniques. Moreover, applying theoretical concepts from the field of generative models benefits framework development over the course of this project. All skills possessed prior to the development of the project are presented in *Table 1* from Appendix B.

9.4 Risk Assessment

Identified risks include challenges with PyTorch, limitations in computational resources, potential delays in experimentation due to academic coursework, and concerns about data quality. A comprehensive risk assessment is laid out in *Table 2* from Appendix B. The initial risk assessment, used for the project report is also depicted in Appendix B.

Mitigation strategies involve relentless commitment to learning, access to relatively high-quality computational resources, regular meetings with the supervisory team, and maintaining an adaptive project plan. Additionally, the scope of the project is continually reassessed to allow for adaptation to necessary changes.

9.5 Gantt Chart

A Gantt chart was used for tracking the project timeline and key milestones. It also provides a visual overview of the schedule, tasks, and adjustments, which is essential for planning and monitoring this study's progress. The initial version of the Gantt chart used in the progress report is depicted in Appendix B. However, according to the supervisor's feedback, this version was not descriptive enough. This led to the construction of a more detailed version, illustrated in *Figure 48* and *Figure 49* from Appendix B.

10 Conclusion, Further Work, and Ideas

In conclusion, this paper project not only achieves its research objectives but also provides a robust framework for managing a complex and dynamic research project. This section presents the conclusion along with suggestions for further advancements.

10.1 Overview and Contributions

This paper, entitled "Enhanced Generative Image Transformation Tool (EGITT)," aims to advance the field of generative models, namely image generation and transformation, by exploring and applying Generative Adversarial Networks (GANs), Style Transfer, and Image Transformation techniques. During this project, a comprehensive approach is adopted to understand and implement various AI techniques that enhance the quality and efficiency of image generation and style transfer processes. Key contributions to this work include:

- **Development and comparison of GAN Models:** Several GAN models have been successfully implemented within the project, including a standard GAN, a Wasserstein GAN, and a Wasserstein GAN with a gradient penalty. These models provide insights into operational efficiency and limitations. This comparative analysis extends the understanding of model suitability depending on specific image characteristics and desired outcomes.
- **Advancements in Style Transfer:** By using style transfer techniques, the project refines the practical application of merging the content of one image with the style of another. The achieved advancements not only improve image quality but also increase the speed of transformation, making real-time applications more feasible.
- **Quantitative and Qualitative Assessments:** The project adopts a comprehensive evaluation framework, utilising both quantitative metrics like EMD and FID and qualitative analyses through visual inspections and peer reviews. This dual approach allows for rigorous testing of model performance and provides a balanced view of their effectiveness.
- **Usage of Sophisticated Technologies and Tools:** Utilising advanced computational tools and platforms such as Google Colab Pro and PyTorch enables working with complex models and large datasets. It also demonstrates the effective usage of modern AI development tools in academic research.
- **Project Management:** The integration of systematic project management practices, coupled with effective risk management and the use of advanced computational tools, lays a strong foundation for successfully overcoming challenges associated with advanced academic research.
- **Ethical Considerations and AI Transparency:** The research also highlights the importance of ethical considerations and transparency in AI developments, setting a precedent for future projects in the domain.

10.2 Limitations

Despite its successes, the project encountered several limitations:

- **Computational Resource Constraints:** While appropriate for the scope of the project, computational resources were sometimes limited. Therefore, this affected the speed and scope of certain experiments, particularly those involving high-resolution image transformations.
- **Model Generalisation:** While the models perform well on training datasets, their ability to generalise to radically different datasets is not fully tested. This might limit their applicability in diverse real-world scenarios.
- **Depth of Style Transfer Customisation:** While improvements are made, the depth of customisation in style transfer is still somewhat limited by current

technology. This is specifically evident when trying to achieve consistent quality across varying styles and complex images.

10.3 Future Work

To address these limitations and build on the current project's achievements, future research could focus on several areas:

- **Improved Model Robustness:** Future work could explore advanced techniques for training generative models to increase their stability and robustness. This would allow further improvement of their generalisation across different datasets and their applicability in real-world scenarios.
- **Enhanced Computational Efficiency:** Investigating more efficient algorithms and network architectures could reduce computational costs and improve the ability to deploy these models in real-time applications.
- **Enhancement of Real-Time Capabilities:** Developing more efficient algorithms and strategies could make real-time style transfer more feasible and practical for a variety of applications.
- **Cross-Domain Application Testing:** Testing models in different domains, such as medical imaging or real-time video processing, could significantly increase their applicability.
- **Ethical AI Development:** Continuing with focusing on the ethical aspects of AI image generation and manipulation, ensure that models are used responsibly and do not contribute to misinformation.
- **Integration with Other AI Technologies:** Combining the developed models with other AI technologies, such as natural language processing or robotics, can open up new opportunities for interdisciplinary research and applications.

In conclusion, although this study has made substantial contributions to the field of AI-driven image processing, the road ahead is filled with opportunities for further research that could revolutionise how machines understand and manipulate visual data. This work lays a solid foundation for future research and advances in the field of artificial intelligence.

11 Bibliography

References

[1] Gatys, L. A., Ecker, A. S., & Bethge, M. (2015b). Texture synthesis and the controlled generation of natural stimuli using convolutional neural networks. *CoRR*, abs/1505.07376.

[2] Gatys, L. A., Ecker, A. S., & Bethge, M. (2015a). A neural algorithm of artistic style. *CoRR*, abs/1508.06576.

[3] Brain, I. G. G., Goodfellow, I., & others. (2014, November 1). Generative Adversarial Networks. *Communications of the ACM*. Retrieved from `https://dl.acm.org/doi/10.1145/3422622`

[4] Isola, P., Zhu, J.-Y., Zhou, T., & Efros, A. A. (2018, November 26). Image-to-image translation with conditional adversarial networks. *arXiv*. Retrieved from `https://arxiv.org/abs/1611.07004`

[5] Jiang, S., Tao, Z., & Fu, Y. (2019, January 23). Segmentation guided image-to-image translation with adversarial networks. *arXiv*. Retrieved from `https://arxiv.org/abs/1901.01569`

[6] Waite, J. (2020). Generative adversarial networks (dissertation). *Electronics and Computer Science, Faculty of Physical Sciences and Engineering, University of Southampton*.

[7] Ratliff, L., Burden, S., & Sastry, S. (2013). Characterization and computation of local Nash equilibria in continuous games. In *Annual Allerton Conference on Communication, Control and Computing* (pp. 917–924).

[8] Borji, A. (2019). Pros and cons of gan evaluation measures. *Computer Vision and Image Understanding*, 179, 41–65.

[9] Gatys, L. A., Ecker, A. S., & Bethge, M. (2016). Image style transfer using convolutional neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 2414–2423).

[10] Ruta, D., Gilbert, A., Collomosse, J., Shechtman, E., & Kolkin, N. (2023, April 11). Neat: Neural artistic tracing for beautiful style transfer. *arXiv*. Retrieved from `https://arxiv.org/abs/2304.05139`

[11] Foster, D. (2022). Chapter 4. In *Generative Deep Learning* (2nd ed., pp. 105–110). O'Reilly Media, Inc.

[12] Johnson, J., Alahi, A., & Fei-Fei, L. (2016, March 27). Perceptual losses for real-time style transfer and super-resolution. *arXiv*. Retrieved from `https://arxiv.org/abs/1603.08155`

[13] Liu, Z., Luo, P., Wang, X., and Tang, X. (2021, September 10). Large-scale celeb faces attributes (celeba) dataset. *CelebA Dataset*. Retrieved from `https://mmlab.ie.cuhk.edu.hk/projects/CelebA.html`

[14] Tiwari, V. (2018, July 4). Pix2pix dataset. *Kaggle*. Retrieved from `https://www.kaggle.com/datasets/vikramtiwari/pix2pix-dataset`

[15] Lin, T.-Y., Patterson, G., Ronchi, M. R., & Cui, Y. (2017). Common objects in context (COCO). *COCO*. Retrieved from `https://cocodataset.org/#home`

[16] Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., & Courville, A. (2017). Improved training of Wasserstein GANs. *arXiv*. Retrieved from http://arxiv.org/pdf/1704.00028.pdf

[17] Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., & Hochreiter, S. (2018, January 12). GANs trained by a two time-scale update rule converge to a local Nash equilibrium. *arXiv*. Retrieved from https://arxiv.org/abs/1706.08500

[18] Arjovsky, M., Chintala, S., & Bottou, L. (2017). Wasserstein GAN. *arXiv e-Print archive*. Retrieved from https://arxiv.org/pdf/1701.07875.pdf

[19] Yik, J. (2023, December 1). GAN implementations/models/wgangp.py. *GitHub*. Retrieved from https://github.com/JayYik/GAN_implementations/blob/master/models/WGAN_GP.py

[20] Lee, C. (2017, July 27). Pytorch implementation of perceptual losses for real-time style transfer and super-resolution. *GitHub*. Retrieved from https://github.com/ceshine/fast-neural-style

[21] Udnie. *Centre Pompidou*. (n.d.). Retrieved from https://www.centrepompidou.fr/en/ressources/oeuvre/A6Zfdvn

[22] Goodfellow, I. (2016). NIPS 2016 tutorial: Generative Adversarial Networks. https://www.semanticscholar.org/paper/NIPS-2016-Tutorial:-Generative-Adversarial-Networks-Goodfellow/2c740e574eea66fdcf473e15ed2c228baef2eccd

[23] Liu, S., Ye, J., and Wang, X. (2023a, April 20). Any-to-any style transfer: Making Picasso and da vinci collaborate. arXiv.org. https://arxiv.org/abs/2304.09728

[24] Guo, Z., Wang, K., Li, W., Qian, Y., Arandjelović, O., & Fang, L. (2024, January 18). Artwork protection against neural style transfer using locally adaptive adversarial color attack. *arXiv*. Retrieved from https://arxiv.org/abs/2401.09673

[25] Gong, Y. (2023). Neural style transfer: A neural style transfer model based on semantic preservation of images and a quantitative evaluation method (dissertation).

[26] Babaeizadeh, M., & Ghiasi, G. (2018, November 21). Adjustable real-time style transfer. *arXiv*. Retrieved from https://arxiv.org/abs/1811.08560

[27] Chilamkurthy, S. (2017, June 10). Writing custom datasets, DataLoaders and transforms. *PyTorch Tutorials*. Retrieved from https://pytorch.org/tutorials/beginner/data_loading_tutorial.html

[28] Song, Y., & Ermon, S. (2020, October 10). Generative modeling by estimating gradients of the data distribution. *arXiv*. Retrieved from https://arxiv.org/abs/1907.05600

[29] Majumdar, S., Bhoi, A., & Jagadeesan, G. (2018, June 3). A comprehensive comparison between neural style transfer and Universal style transfer. *arXiv*. Retrieved from https://arxiv.org/abs/1806.00868

[30] Terven, J., Cordova-Esparza, D. M., Ramirez-Pedraza, A., and Chavez-Urbiola, E. A. (2023a, September 6). Loss functions and metrics in deep learning. arXiv.org. https://arxiv.org/abs/2307.02694

[31] He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2018, January 24). Mask R-CNN. *arXiv*. Retrieved from `https://arxiv.org/abs/1703.06870`

[32] Cao, Z., Hidalgo, G., Simon, T., Wei, S.-E., & Sheikh, Y. (2019, May 30). OpenPose: Realtime multi-person 2D pose estimation using part affinity fields. *arXiv*. Retrieved from `https://arxiv.org/abs/1812.08008`

[33] Jabbar, A., Li, X., & Omar, B. (2020, June 9). A survey on generative adversarial networks: Variants, applications, and training. *arXiv*. Retrieved from `https://arxiv.org/abs/2006.05132`

[34] Peters, T., & Farhat, H. (2023, November 21). High-resolution image-based malware classification using multiple instance learning. *arXiv*. Retrieved from `https://arxiv.org/abs/2311.12760`

[35] Niu, C., Zauner, K.-P., & Tarapore, D. (2023, January 31). End-to-end learning for visual navigation of Forest Environments. *MDPI*. Retrieved from `https://www.mdpi.com/1999-4907/14/2/268`

[36] Bousmalis, K., Silberman, N., Dohan, D., Erhan, D., & Krishnan, D. (2017, August 23). Unsupervised pixel-level domain adaptation with generative Adversarial Networks. *arXiv*. Retrieved from `https://arxiv.org/abs/1612.05424`

[37] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2023, August 2). Attention is all you need. *arXiv*. Retrieved from `https://arxiv.org/abs/1706.03762`

[38] He, K., Zhang, X., Ren, S., & Sun, J. (2015, December 10). Deep residual learning for image recognition. *arXiv*. Retrieved from `https://arxiv.org/abs/1512.03385`

12 Appendix A: Project Brief

The project brief outlines foundational objectives and scope, setting the stage for subsequent phases.

Part III Project Brief

Project Name:	Enhanced Generative Image Transformation Tool (EGITT)
Student Name and ID:	Georgi Iliev - gdi1u21
Supervisor Name and ID:	Hikmat Farhat - hf1g22

Background:

Artificial Intelligence has significantly transformed visual content generation, manipulation, and interpretation. The realm of generative AI has been making strong strides recently within various domains including Art and Design, Healthcare, Marketing and Education.

Problem:

Despite advancements, limitations remain in terms of computational cost and resources, deficiency in feature extraction [1] and bottlenecks, indicating a room for an improvement. This project will address these limitations by improving current overall capabilities of Image to Image (I2I) translation and providing comparable performance between Generative Adversarial Networks (GANs) and Diffusion Models with reduced computational cost [2].

Aims:

This third-year project aims to explore and advance the prominent and evolving field of unsupervised Image to Image (I2I) translation. It will involve translation of images from one domain to another This endeavor would empower users to gain practical experience on this emerging technological tool.

Objectives:

The primary purpose of this project is to create a cutting-edge tool that showcases the latest advances in generative AI. It will be executed using PyTorch framework and will be structured as follows:

1. Comparing efficiency of Generative Adversarial Networks (GANs) with Diffusion Models in image transformation. GANs employ adversarial training, while Diffusion Models transform noise into data iteratively.
2. Identify prevalent baseline and dataset, then conduct both qualitative and quantitative evaluations to determine and compare their efficiency and effectiveness
3. Explore appropriate metrics for comparing image transformations such as Frèchet Inception Distance (FID) where lower value indicates superior performance [3] and Kernel Inception Distance (KID) where generated images are closer to actual samples [4] to facilitate comparison and determine the potential quality of output.

By utilising these actions, introducing specific case functions and mitigating the current limitations, this project will broaden the generative AI landscape and contribute to the wider field of artificial intelligence such that the final product aligns with the initial vision.

Figure 44: *Project Brief*

The completion of most initial objectives is evidence of well-structured project planning and consistent progress throughout the project's life cycle.

13 Appendix B: Project Management

13.1 Work Breakdown Structure

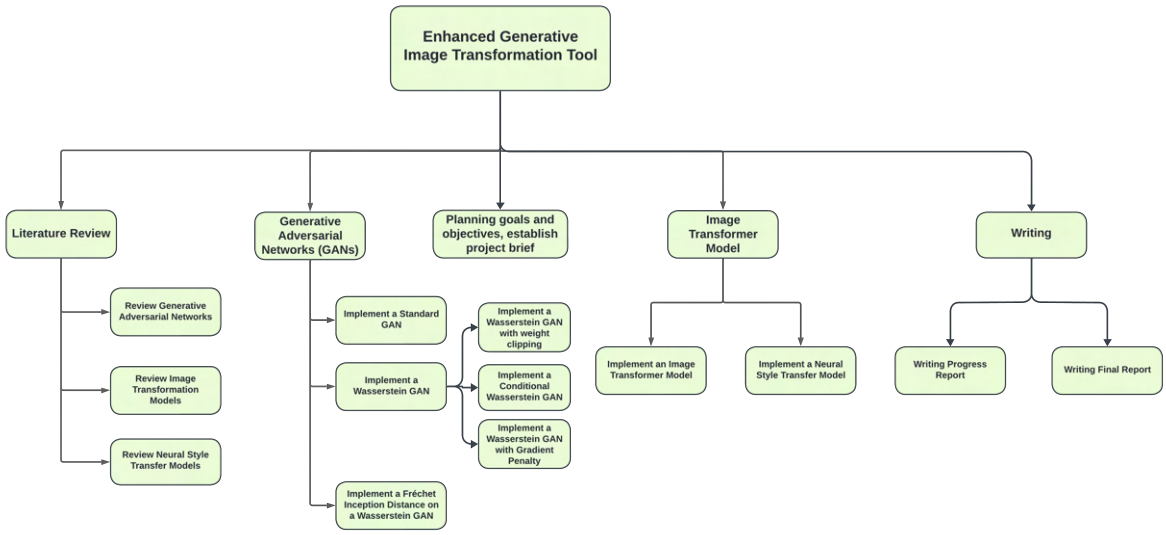


Figure 45: *Work Breakdown Structure (WBS)*

13.2 Skill Assessment prior to project

Skills	Rating (1-5)	Remarks
Data	1.8	Needs improvement
Gathering	3	Sufficient
Static feature extraction	1	Beginner level
Analysis	2	Basic understanding
Visualisation	2	Basic understanding
Processing	1	Beginner level
Computer Vision	1	Fundamental knowledge required
PyTorch	1	Basic familiarity
Generative Adversarial Networks	1	Introductory knowledge
Image Transformer Models	1	Introductory knowledge
Style Transfer Models	1	Introductory knowledge
Documentation	3.3	Competent
Literature review	3	Adequate
Planning	4	Good skills
Report writing	3	Adequate

Table 1: Personal skills assessment prior to the project

13.3 Risk Assessment

Table 2: A Risk Assessment summary

Risk name	Severity	Likelihood	Risk	Mitigation
Difficulty with PyTorch Development and Generative Models	4	3	12	Gather an in-depth understanding of PyTorch through early study and practice. Seek guidance from the project supervisor. Maintain regular progress monitoring and debugging.

Continued on next page

Table 2 Continued from previous page

Risk name	Severity	Likelihood	Risk	Mitigation
Not gaining access to a powerful enough computer for training, tuning, and testing the model	4	4	16	Investigate available computing resources from the university in advance. Consider alternative computing resources or cloud-based solutions.
Other modules take a significant amount of study time	3	5	15	Establish a detailed time management plan, allocating time for each module. Prioritize project-related tasks and maintain a flexible schedule to accommodate unforeseen academic demands; Seek help from the supervisor or reduce workload if required.
Data Availability and Quality	4	3	12	Conduct extensive research on available datasets. Create strategies for supplementing data. Consider data collection as needed. Apply data quality measures.
Over-ambitious Scope	3	3	9	Define clear project boundaries and objectives. Continuously review the project scope and consider expansion only after the initially set objectives have been achieved.
Failure to see supervisor regularly	2	2	4	Set up assigned weekly meetings with the first supervisor and maintain contact with the second examiner.

Continued on next page

Table 2 Continued from previous page

Risk name	Severity	Likelihood	Risk	Mitigation
User Acceptance and Feedback	2	4	8	Plan user testing and feedback acquisition to guarantee the tool meets user expectations. Make incremental improvements based on user feedback. Engage with potential users and experts for guidance.

Risk assessment

Problem	Loss	Probability	Risk	Mitigation
Difficulty with PyTorch Development and generative models	4	3	12	Gather in-depth understanding of PyTorch through early study and practice; Seek guidance from the project supervisor; Maintain regular progress monitoring and debugging
Not gaining access to a powerful enough computer for training, tuning, and testing the model	4	4	16	Investigate available computing resources from the University in advance; Consider alternative computing resources or cloud - based solutions
Other modules take significant amount of study time	2	5	10	Establish a detailed time management plan allocating time for each module; Prioritise project-related tasks and maintain flexible schedule to accommodate unforeseen academic demands; Seeking help from supervisor or reducing workload if required
Data Availability and Quality	4	3	12	Conduct extensive research on available datasets; Create strategies for supplementing data; Consider data collection as needed; Apply data quality measure
Over-ambitious Scope	3	3	9	Define clear project boundaries and objectives; Continuous reviewing of the project scope and consideration of expansion only after the initially set objectives have been achieved
User Acceptance and Feedback	2	4	8	Plan user testing and feedback acquisition to guarantee tool meets user expectations; Make incremental improvements based on user feedback; Interact with potential users and experts for guidance.

Figure 46: *Progress Report's Risk Assessment*

13.4 Gantt Charts

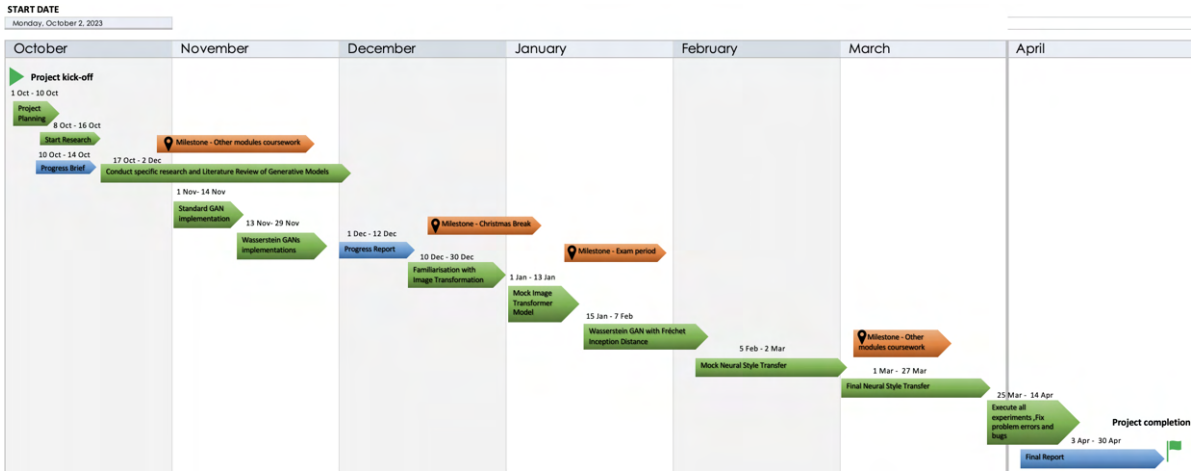


Figure 47: *Initial Gantt Chart*

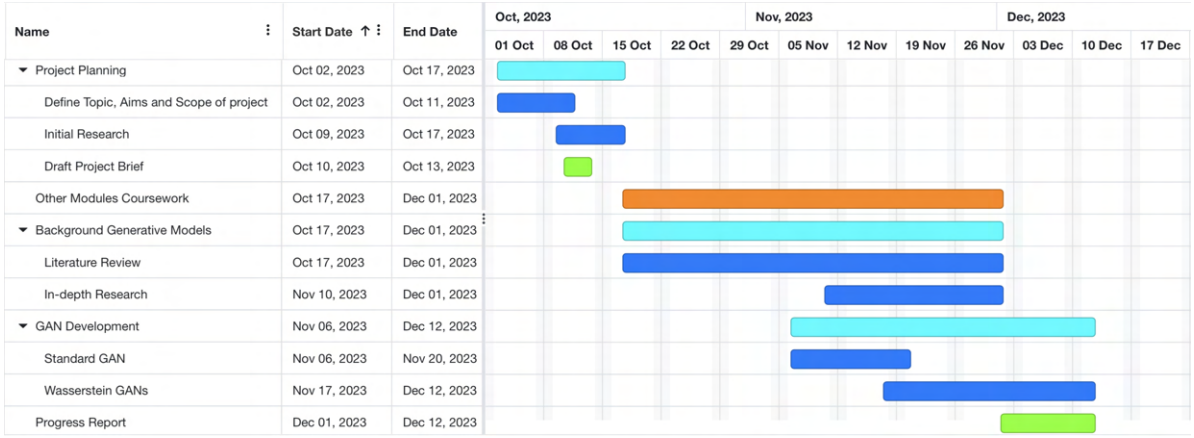


Figure 48: *Redefined Gantt Chart*

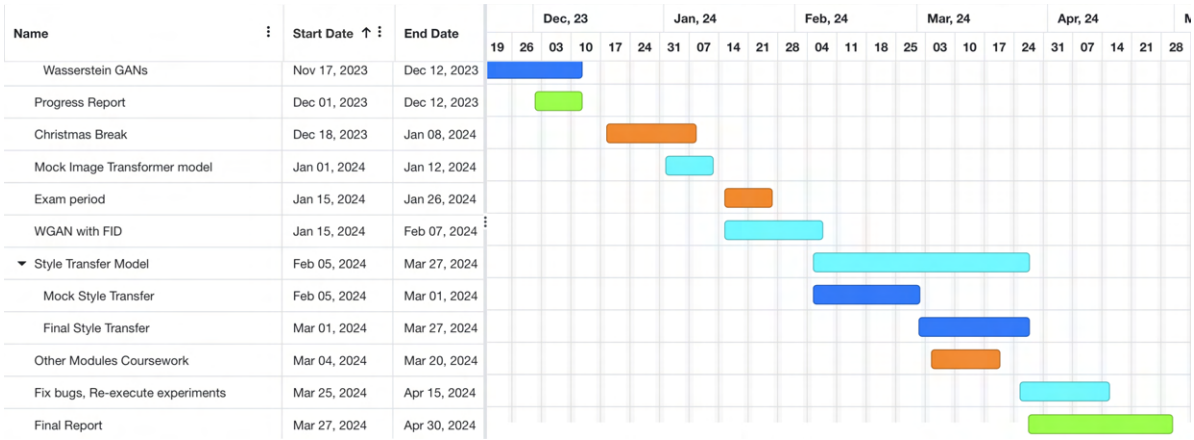


Figure 49: *Redefined Gantt Chart*

14 Appendix C: Style Transfer Results

14.1 Longest training process

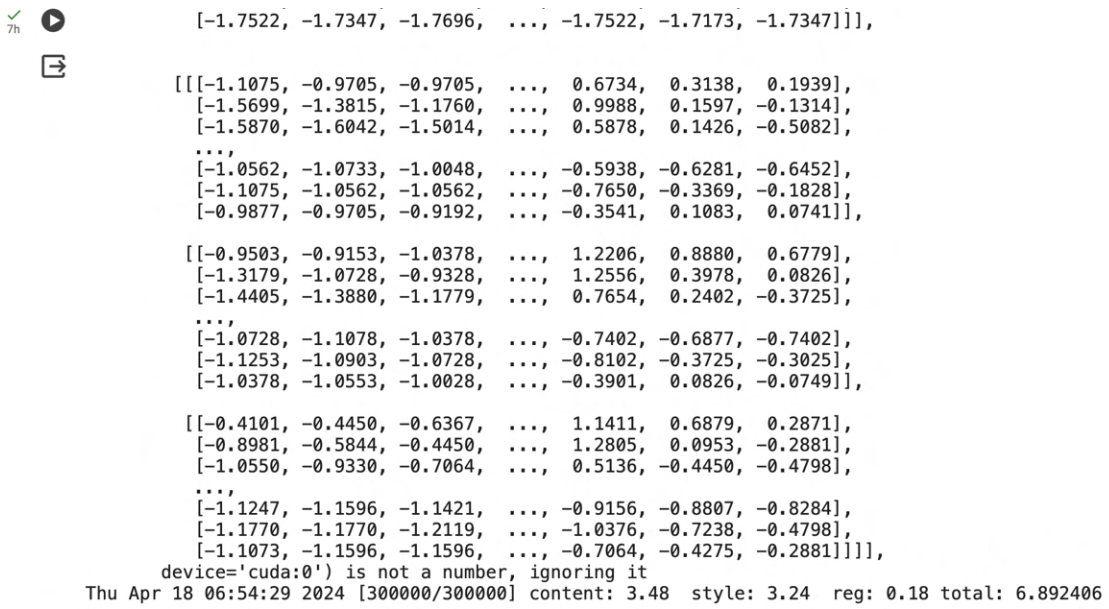


Figure 50: *Longest training process containing 300,000 iterations*

14.2 Qualitative Evaluations Using Peer's Photography

Figure 51: *Real-time style transfer model results using the "Mosaic" style image*

Figure 52: *Real-time style transfer model results using the "Mosaic" style image*

Figure 53: *Real-time style transfer model results using the "Picasso" style image*

This section includes styles applied to photographs of beautiful spots in Bulgaria. The images are taken by a peer who gave their permission to use them for the purposes of this study.

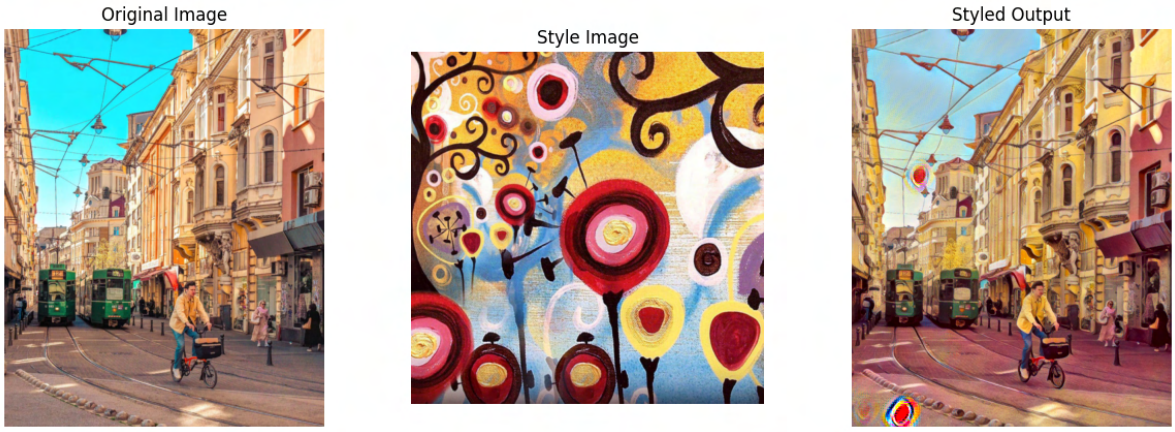


Figure 54: *Real-time style transferred image using peer's photography and the "Candy" style image*

Figure 55: *Real-time style transferred image using peer's photography and the "Starry night" style image*

Figure 56: *Real-time style transferred image using peer's photography and the "Composition" style image*

15 Appendix D: Documentation of epoch checkpoints for various GANs

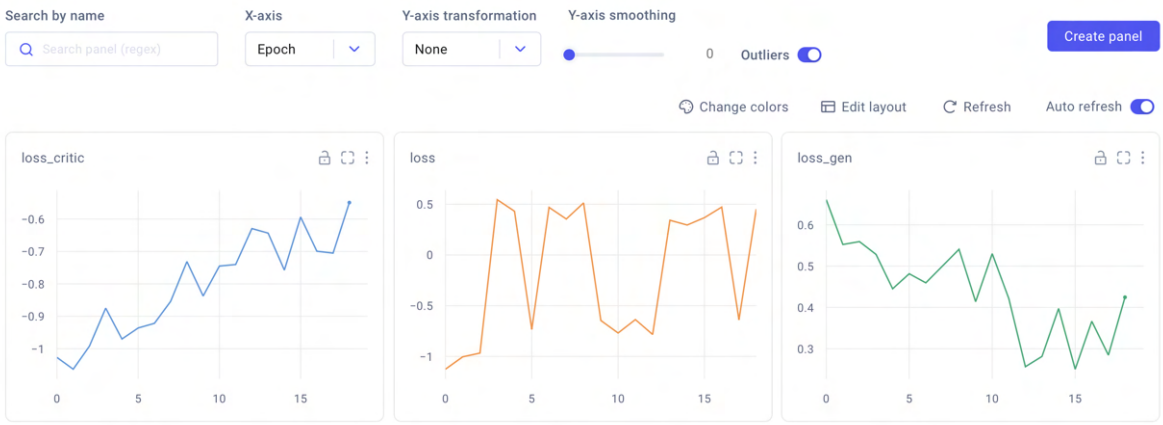


Figure 57: *First documentation of epoch checkpoints for Conditional Wasserstein GAN with Gradient Penalty*

16 Appendix E: User Study

16.1 Participant Information Sheet

Appendix (i) Participant Information

Participant Information

Ethics reference number: **ERGO**/FEPS/**78677.A1**	Version: 1	Date: 08/08/2023
Study Title: Student project for COMP3200		
Investigator: **Georgi Iliev**		

Please read this information carefully before deciding to take part in this research. If you are happy to participate you will be asked to indicate your consent to take part either verbally or by selection if an online questionnaire. Your participation is completely voluntary.

What is the research about? This research project is part of the COMP3200 project. The research will only involve a:

Questionnaire

Why have I been chosen? You have been approached because you are known to the student or because you have been identified by the student as being appropriate for the research

What will happen to me if I take part? You will take part in a short

Questionnaire

Are there any benefits in my taking part? The study will add to current knowledge, as well as being valuable practical learning for the COMP3200 project student

Are there any risks involved? No sensitive issues will be discussed and there are no risks beyond that which would normally be experienced in everyday life.

Will my data be confidential? The project student will collect anonymous data suitable for evaluating their project; this will be retained until the end of the project. No video or audio recording will occur. *Please do not give any identifiable personal information.*

What happens if I change my mind? You may withdraw at any time and for any reason. You may decline to give your consent and not take part in the study without penalty.

What happens if something goes wrong? If you have any concern or complaint, contact Georgi Iliev or Hikmat Farhat at gdi1u21@soton.ac.uk and h.farhat@soton.ac.uk otherwise please contact Research Governance Office (02380 595686, Rgoinfo@soton.ac.uk).

Figure 58: *Participant Information Sheet presented to every participant prior to starting with the survey. Only upon agreement, the participant can begin with the survey.*

16.2 Consent Form

Appendix (ii) Consent Form

Consent Form

Ethics reference number: **ERGO/**FEPS**/78677.A1**	Version: 1	Date: 25/03/2024
Study Title: Enhanced Generative Image Transformation Tool (EGITT)		
Investigator: **Georgi Iliev**		

Please read the following and indicate by selection on an online questionnaire if you agree with the following statements:

I have read and understood the Participant Information and have had the opportunity to ask questions about the study.

I agree to take part in this study.

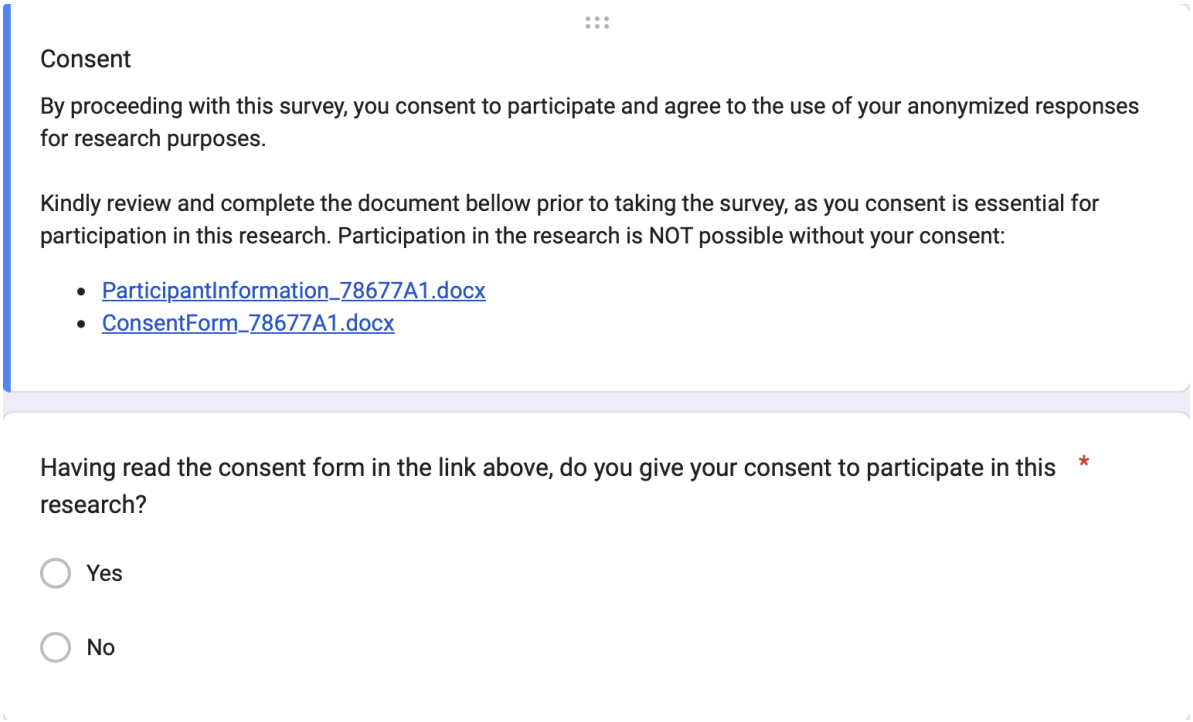
I understand my participation is voluntary and I may withdraw at any time and for any reason.

ONLY If the participant has verbally or by selection on an online questionnaire agreed to the above, and consented to take part in the research, the study may commence.

Figure 59: *Consent Form presented to every participant prior to starting with the survey. Only upon agreement, the participant can begin with the survey.*

16.3 Survey

Figure 60: Survey



Consent

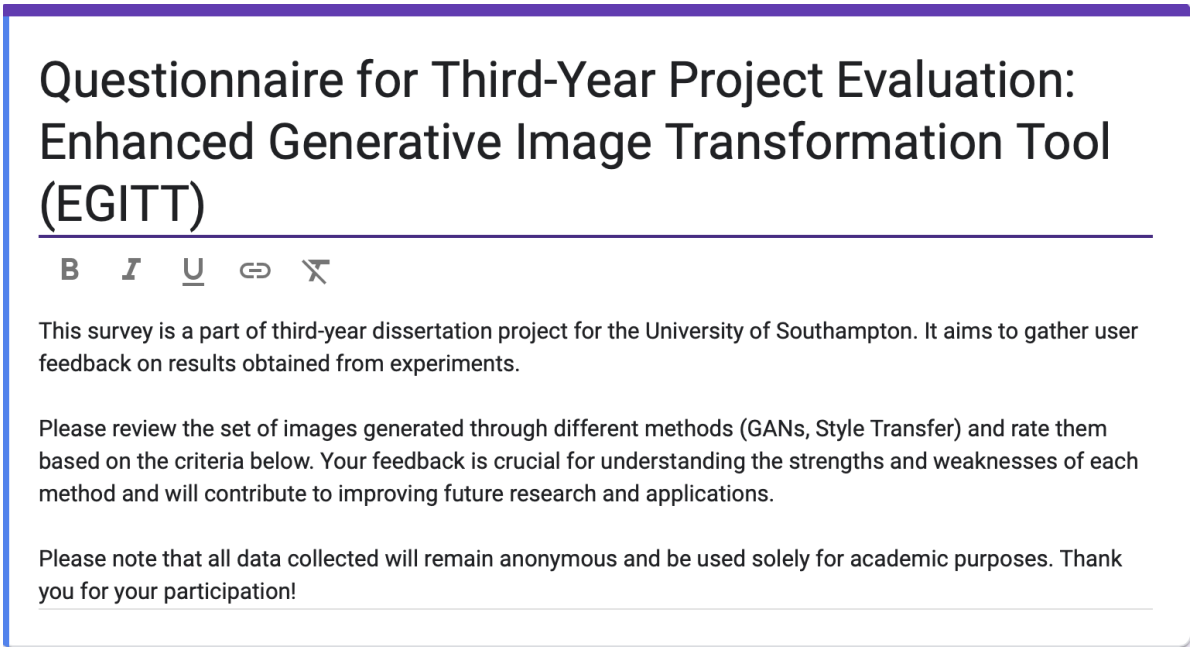
By proceeding with this survey, you consent to participate and agree to the use of your anonymized responses for research purposes.

Kindly review and complete the document bellow prior to taking the survey, as you consent is essential for participation in this research. Participation in the research is NOT possible without your consent:

- ParticipantInformation_78677A1.docx
- ConsentForm_78677A1.docx

Having read the consent form in the link above, do you give your consent to participate in this research? *

- Yes
- No



Questionnaire for Third-Year Project Evaluation: Enhanced Generative Image Transformation Tool (EGITT)

This survey is a part of third-year dissertation project for the University of Southampton. It aims to gather user feedback on results obtained from experiments.

Please review the set of images generated through different methods (GANs, Style Transfer) and rate them based on the criteria below. Your feedback is crucial for understanding the strengths and weaknesses of each method and will contribute to improving future research and applications.

Please note that all data collected will remain anonymous and be used solely for academic purposes. Thank you for your participation!

Method 1: Generative Adversarial Networks (GANs)

For the purposes of our survey, we ask you to rate:

- the realism and believability of the generated images.
- The quality of detail and texture that GANs manage to reproduce or update.
- The overall aesthetic and visual appeal of the images.

Evaluate the details and textures in the images. Are they well-preserved or distorted? Rate from 0-10. *

	0	1	2	3	4	5	6	7	8	9	10	
Very Poor	○	○	○	○	○	○	○	○	○	○	○	Excellent

How realistic are the images generated by the GAN models? Rate from 0-10? *

	0	1	2	3	4	5	6	7	8	9	10	
Not Realistic	○	○	○	○	○	○	○	○	○	○	○	Extremely Realistic

Evaluate the details and textures in the images. Are they well-preserved or distorted? Rate from 0-10. *

	0	1	2	3	4	5	6	7	8	9	10	
Very Poor	○	○	○	○	○	○	○	○	○	○	○	Excellent

Method 2: Real-Time Style Transfer

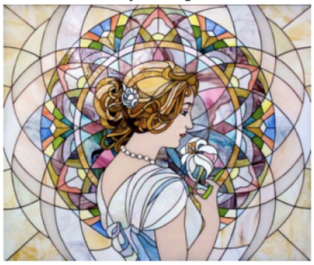
Your evaluation should consider:

- Effectiveness of Style Integration: How seamlessly and convincingly the style is integrated into the content image.
- Artistic Quality: The overall artistic appeal of the stylized images, including color harmony, texture fidelity, and creative expression.
- Performance and Consistency: The ability of the system to maintain consistent quality in real-time, assessing any variations in style application across different images.

How effectively does the style transfer incorporate the style of one image with the content of another (See stylised image 1 closer below)? Rate from 0-10. *

Original Image

Style Image



Styled Output

	0	1	2	3	4	5	6	7	8	9	10	
Ineffective	○	○	○	○	○	○	○	○	○	○	○	Seamless Integration

Stylised Image 1

Rate the artistic quality of the style-transferred images (0-10). (See stylised image 2 closer below) *

Original Image

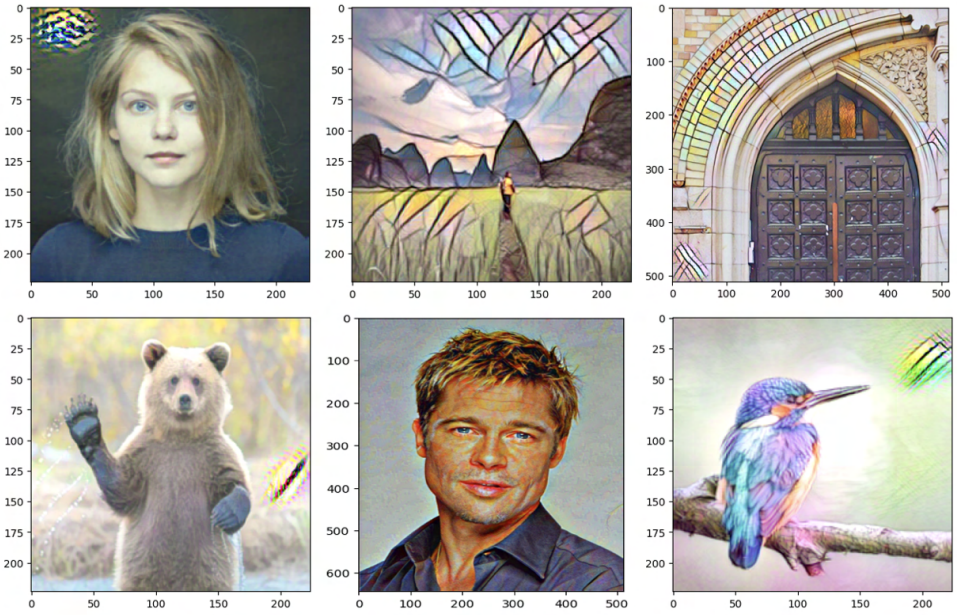
Style Image

Styled Output

	0	1	2	3	4	5	6	7	8	9	10	
Poor Quality	○	○	○	○	○	○	○	○	○	○	○	Exceptional Quality

Stylised Image 2

How consistent are the style elements across different images? Rate from 0-10. *



0 1 2 3 4 5 6 7 8 9 10

Comparative Evaluation

Description (optional)

Which method do you prefer for image generation? *

- ○ Method 1: Generative Adversarial Networks (GANs)
- ○ Method 2: Real-Time Style Transfer
- ○ Both
- ○ Neither

Why did you choose this method and what suggestions do you have for improving the image quality?

Short-answer text

Additional Feedback

Description (optional)

Please provide any additional comments or insights you have about the generated images or the methods used.

Long-answer text