Article

Stylit

Daniel Carreira [1,*], Fábio Gaspar [1] and Carlos Costa [1]

[1] Computer Science and Communications Research Centre, School of Technology and Management, Polytechnic University of Leiria, 2411-901 Leiria, Portugal
* Correspondence: daniel.s.carreira@ipleiria.pt, fabio.c.gaspar@ipleiria.pt, carlos.m.costa@ipleiria.pt

Abstract: This paper presents the development of a mobile app that combines text-to-image synthesis and style transfer using generative artificial intelligence techniques. The app aims to empower users to generate realistic images directly from textual descriptions and explore various artistic styles for customization. The key contributions include a unified mobile application with state-of-the-art deep learning techniques in image generation on the cloud, as well as a robust style transfer algorithm on the device. The paper discusses the related work in generative AI, text-to-image synthesis, and style transfer, and outlines the methodology employed for developing the mobile app. In the paper, we present the performance evaluation of our mobile app, highlighting the impressive results achieved. Additionally, we delve into potential future developments and research directions to further enhance the app's capabilities. By offering an intuitive and accessible platform, our app empowers users to unlock their creative potential in art, design, and entertainment.

Keywords: Generative AI; Image synthesis; Mobile application; Style Transfer; Text-to-Image

Citation: Carreira, D.; Gaspar, F.; Costa, C. Stylit. *Journal Not Specified* **2023**, *1*, 0.

Received:
Accepted:
Published:

Copyright: © 2024 by the authors. Submitted to *Journal Not Specified* for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

In recent years, generative artificial intelligence (AI) has made amazing strides, transforming a number of fields like image synthesis, natural language processing, and creative design. The creation of realistic visuals from verbal descriptions, or text-to-image (T2I) synthesis, is an intriguing application of generative AI. Applications for this capacity may be found in virtual reality, computer-aided design, and entertainment.

Alongside text-to-image synthesis, style transfer (ST) has emerged as a popular technique to manipulate and transform images according to desired artistic styles. By enabling the transfer of artistic attributes from a reference image to a target image, ST has become a powerful tool for artistic expression and visual customization.

In this paper, we present the development of a mobile app that leverages generative AI techniques to address the T2I synthesis problem and provides ST functionality. Our app aims to empower users with the ability to generate realistic images directly from textual descriptions and explore various artistic styles for customization. By combining these two capabilities into a user-friendly mobile app, we aim to democratize generative AI and provide an accessible platform for creative expression.

The key contributions of this work include a unified mobile application with the latest state-of-the-art deep learning techniques in image generation, on the cloud. Additionally, we incorporate a robust style transfer algorithm, on the device, that enables users to seamlessly transfer artistic styles to their generated images. We also highlight the challenges faced during the app development process and discuss potential future developments and research directions.

Through this work, we aim to unlock the creative potential of users by providing an intuitive and accessible platform that combines text-to-image synthesis and style transfer. By enabling users to effortlessly generate personalized images with desired styles, our app paves the way for innovative applications in art, design, and entertainment.

This paper is organized as follows: In Section 2, we discuss the related work in the field of generative AI, text-to-image synthesis, and style transfer. Section 3 outlines the methodology employed for developing the mobile app, including the text-to-image synthesis model, style transfer algorithm, and app development process. The results of our app's performance are presented in Section 4, followed by a the conclusion of the paper in Section 5, by summarizing our contributions and emphasizing the significance of the developed mobile app for generative AI.

2. Related Work

The integration of mobile applications and artificial intelligence has become increasingly important in our daily lives, as it allows us to access advanced computational capabilities on the go. From text-to-image synthesis to generative AI and style transfer, these technologies have revolutionized the way we interact with digital content and have opened up new opportunities for creative work where the unique constraints refer to the limited computational resources and energy efficiency requirements, presenting challenges that must be addressed for its effective implementation.

The work in [1] research adversarial text-to-image synthesis, where it discussed how the advent of generative adversarial networks (GANs) has made it possible to synthesize images from textual descriptions turning this approach significant progress for the next years in terms of visual realism, diversity, and semantic alignment. The authors also critically examine current strategies to evaluate text-to-image synthesis models and identify new areas of research such as the development of better datasets and evaluation metrics as well as possible improvements in architectural design and model training.

The work in [2] provide discusses advancements in the field of image synthesis, particularly in generating high-quality images, with a focus on faces. The authors conducted a quantitative comparison of three popular systems: Stable Diffusion, Midjourney, and DALL-E 2 with the goal to assess their ability to generate photorealistic faces in real-world scenarios.

The work presented in [3] is dedicated to the automatic synthesis of realistic images from textual descriptions, leveraging advancements in deep neural network architectures. Although it was presented in 2016, a time when AI systems had not yet achieved the goal of generating highly realistic images from text, the paper acknowledges notable progress made in two crucial areas: recurrent neural networks (RNNs) for text feature representation and GANs for image generation.

The work described in [4] provides focus on addressing the challenge of creating realistic visuals from hand-drawn sketches proposing the use of generative adversarial networks (GANs) which can be applied in numerous fields such as law enforcement and entertainment. The work in [5] also focuses on a face photo-sketch synthesis technique but primarily focuses on the challenge of enhancing the image quality in the generated faces.

The work in [6] introduces BlazeStyleGAN which is a remarkable contribution to the field of GANs and image synthesis, specifically focused on running StyleGAN models on mobile devices. While StyleGAN models have gained significant popularity for generating and editing face images, their implementation on smartphones has been limited. This work addresses this limitation by introducing BlazeStyleGAN, which is claimed to be the first StyleGAN model capable of real-time operation on smartphones.

The work in [7] focuses on a technique to transfer the style of a reference headshot photo onto a new one, allowing users to easily replicate the look of renowned artists in their own headshot portraits. The work explains that headshot portraits require advanced skills that may not be possessed by casual photographers and propose a technique to reach a compelling visual style in these photos. It introduces a multiscale and local approach to robustly transfer the local statistics of an example portrait onto a new one, while maintaining the integrity and uniqueness of the subject's face.

This paper in [8] presents a new method for transferring artistic styles to photographs using deep learning and introducing a constraint that ensures the transformation from input to output remains locally consistent in colorspace.

This work in [9] introduces a unified end-to-end pipeline for style transfer for face Swapping and reenactment. The authors propose a novel approach that outperforms existing state-of-the-art methods in generating realistic face images, without requiring subject-specific training. This approach holds potential for various domains such as entertainment, and deep fake detection.

The work in [10] focuses on the deployment of deep learning models on embedded devices, addressing the challenges associated with privacy, data limitations, network connectivity, and the need for fast model adaptation. The paper discusses the efforts by Google to overcome these challenges by incorporating an experimental transfer learning API into TensorFlow Lite, their machine learning library. To demonstrate the issue, experiments were conducted using a simple transfer learning model and an Android application was developed to showcase the limitations of transfer learning in these devices.

Large language models (LLMs) [11] are very much used in text-to-image thanks to their ability to process human-like text which is a very important feature in this domain. The paper in [12] introduces LLMScore, a new evaluation metric for text-to-image synthesis. The authors re-evaluate existing model-based metrics and highlight the effectiveness of large language models in this task. This research opens up possibilities for a more adaptable text-to-image evaluation approach that can follow human instructions by capturing the multi-granularity compositionality between the synthesized images and the text prompt.

In mobile development, various approaches have been used to implement artificial intelligence on mobile devices. One such approach is cloud computing, where the AI models are hosted on remote servers, and the mobile applications communicate with these servers to access the AI capabilities. This enables mobile devices to leverage the power of advanced AI models without overloading their limited computational resources. Such an example is the work in [13] that addresses the challenge, specifically focusing on GANs in the Cloud. The discriminative part of the GAN is trained on the user's devices using their data, while the generative model is trained remotely on a server. Another approach is the use of techniques like federated learning [14].

3. Methodology

This section presents a comprehensive overview of the methodology used to develop a mobile app for style transfer and image generation. The methodology comprises three key components: Text-to-Image (T2I) Synthesis Model, Style Transfer (ST) Model, and App Development. Each component is described in detail below.

3.1. Text-to-Image Synthesis Model

In the ever-evolving technological landscape, machines face a formidable challenge in achieving effective natural language understanding. Nevertheless, the once elusive concept of transforming a simple text input into a visually coherent image has now become a reality. Some of the more well-known systems/models in the community are:

- **Stable Diffusion**[1] is an open-source T2I model created by Stability AI. Launched in August 2022, this model has been trained on an open dataset, offering a powerful tool to bring your ideas to life.
- **DALL-E** [2] is a private T2I model created by OpenAI. OpenAI launched DALL-E in January 2021, capturing the attention of researchers, artists, and enthusiasts alike. Following the success of DALL-E, other models and projects have been inspired by its capabilities. One such example is DALL-E Mini [3], developed by Crayon and launched in 2021.

[1] https://stablediffusionweb.com/
[2] https://openai.com/dall-e-2
[3] https://www.craiyon.com/

- **Midjourney** [4] is a private T2I model, introduced in July 2022. It provides the same features as the rest of the competitors, however, it has overall superior quality and more realistic images.

Although all of these models share the same objective, Stable Diffusion shines due to its open access to the model and dataset, making it the choice for this project [2].

3.1.1. Dataset

The training of Stable Diffusion utilizes the 2b English language label subset of the LAION 5b dataset created by the German charity LAION [15]. Figure 1 provides a simple example.

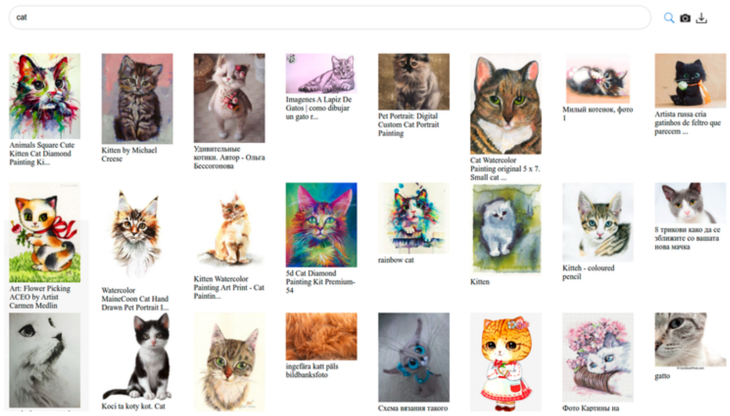


Figure 1. Laion-5b batch example - Taken from [16]

3.1.2. Architecture

Stable Diffusion comprises three primary components in its architecture [17]:

Text encoder: Transformer language model, which serves as the text encoder in a CLIP model that converts textual information into a numerical representation to capture the essence of the ideas contained in the text. The input text is transformed into a list of numbers that correspond to each word in the text, producing a vector for each token. Stable Diffusion models use a pre-trained ClipText model released by OpenAI.

CLIP, trains using a dataset containing a variety of images and their corresponding descriptions. This process combines the power of both an image encoder and a text encoder where each element is encoded separately using the image and text encoders and compared using similarity. In order to enhance the model's performance, an iterative process is employed to continually update the image and text encoders. The encoders eventually learn to build integration where an image of a dog and the associated phrase "a picture of a dog" have a strong representation using this method. To further reinforce the model's understanding, it is also necessary to include negative examples in the training process to assign them with low similarity scores.

Image information creator Generates image information within the latent space. Composed of a UNet neural network and a scheduling algorithm, this component goes through a "diffusion" process to gradually process the information, leading to the creation of a high-quality image in the format of a tensor (multi-dimensional array).

To incorporate text (prompts) into the image generation process, we need to modify the denoising Unet to include the text as an input. These text inputs latent spaces are generated using the previously trained text encoder which will create a noise input to the Unet that is responsible to make the diffusion process and get the representation of the image.

[4] https://www.midjourney.com/

Image decoder: The image decoder takes the information generated by the image information creator and utilizes it to generate a visual representation. This component runs only once at the conclusion of the process, producing the final image with its individual pixels (Figure 2).

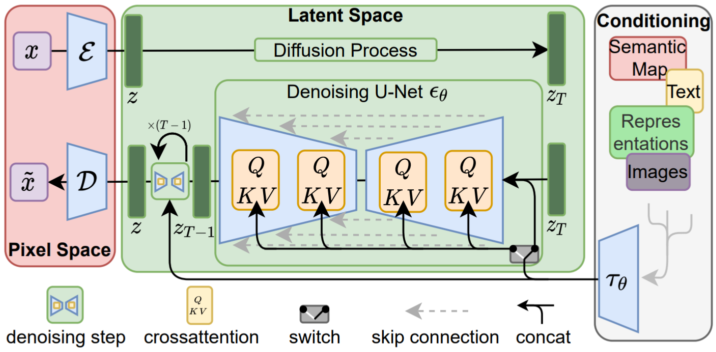


Figure 2. Stable diffusion model - Taken from [18]

3.2. Style Transfer Model

Style transfer is a technique that combines the content of one image with the artistic style of another, resulting in visually stunning compositions while preserving its structure. This process unlocks a world of possibilities and can be used to produce captivating visual art that merges different artistic styles and content for various domains, such as generating artistic images, producing visually appealing graphics, and even reimagining photographs with different artistic styles.

3.2.1. Dataset

In this project, we utilized the ImageNet dataset, which consists of 1000 classes, along with the WikiArt dataset containing 198 classes, for retraining purposes. In traditional transfer learning, it is common to freeze the conventional layers of a large model, preserving their feature extraction capabilities, and adapt only the classification layer to the specific task at hand. However, in the context of style transfer, where we are solely interested in the feature maps of specific convolutional layers, this approach doesn't offer much benefit. Therefore, we opted to initialize the weights of a VGG19 model with pre-trained weights from ImageNet, and subsequently train it using the WikiArt dataset (Figure 3).

Figure 3. WikiArt batch example.

3.2.2. Architecture

The style transfer model architecture (Figure 4) was inspired by a TensorFlow Example [5] that involves two main stages: style prediction and style transformation.

In the first stage, known as style prediction, a VGG-19 model is used to extract the style information from a style image. The VGG-19 model is a convolutional neural network (CNN) that has been trained on a large dataset for various image recognition tasks. By

[5] https://www.tensorflow.org/lite/examples/style$_{t}ransfer/overview$

feeding the style image through the VGG-19 model, the network analyzes the image and captures its style characteristics.

The output of this analysis is a style bottleneck vector, which represents a compressed representation of the style features extracted from the style image. This vector contains information about the textures, colors, and patterns that define the unique style of the image.

In the second stage, known as style transformation, another model utilizes the style bottleneck vector obtained from the previous stage, along with a content image, to generate a stylized output. This model takes the content image as input and applies the style characteristics captured in the style bottleneck vector to transform the appearance of the content image.

By combining the content image with the style bottleneck vector, the model adjusts the content image's features, such as colors, textures, and patterns, to match the style defined by the style image. The result is a stylized output that retains the structure of the original image but adopts the style of the style image.

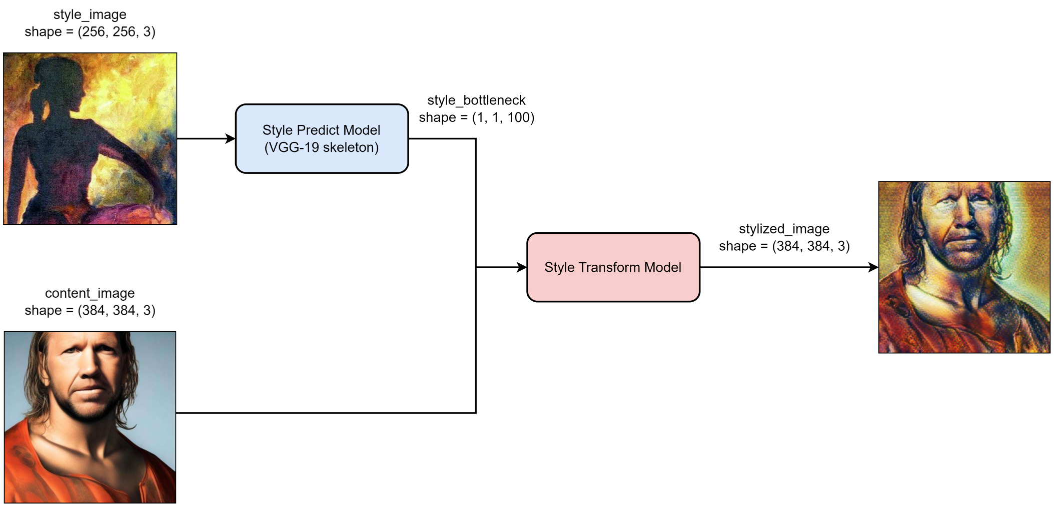


Figure 4. Style Transfer Model Architecture.

3.3. Application Development

This subsection provides comprehensive insights into the application, covering everything from the mockups and employed architecture to the final functionality overview offering a complete understanding of the application's design, structure, and features.

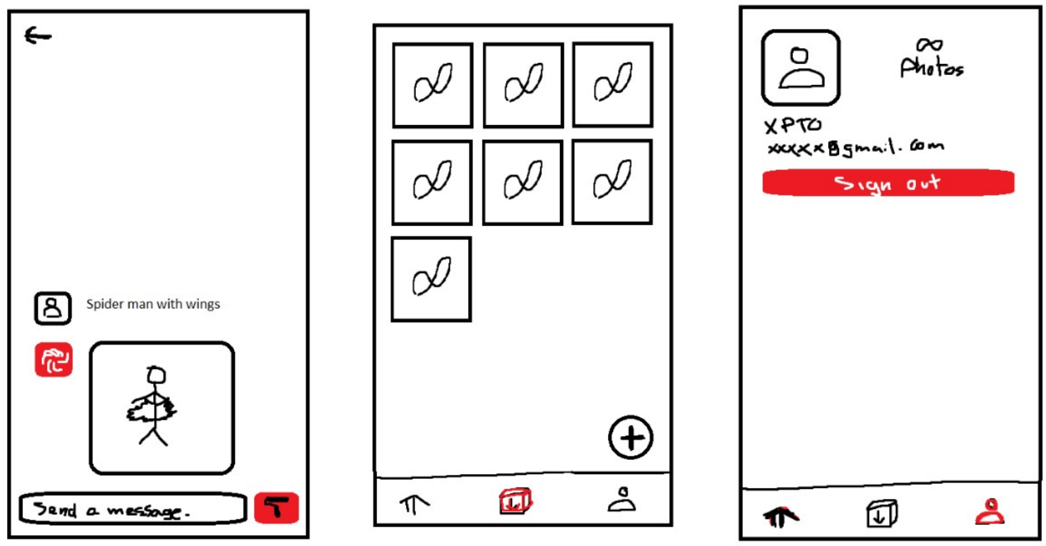
3.3.1. Mockups

To facilitate understanding and guide the design process of the application, a series of mockups have been meticulously created serving as visual representations that showcase the intended appearance and functionality of the application. The following Figure 5 images showcase the mockups.

3.3.2. Architecture

To enable the execution of machine learning models for style transfer functionality, we have deployed the model on a device equipped with TensorFlow Lite — a lightweight version of TensorFlow known for its efficient performance.

However, the text-to-image stable Diffusion requires significant computational power, and as a result, we have implemented a backend service hosted on the cloud. By utilizing the cloud infrastructure, we ensure a faster and more seamless user experience. Apart from generating images with text-to-image, all other functionalities and resources are exclusively allocated within the mobile application.



(a) Home screen. **(b)** Archive screen. **(c)** Profile.

Figure 5. Stylit application mockups

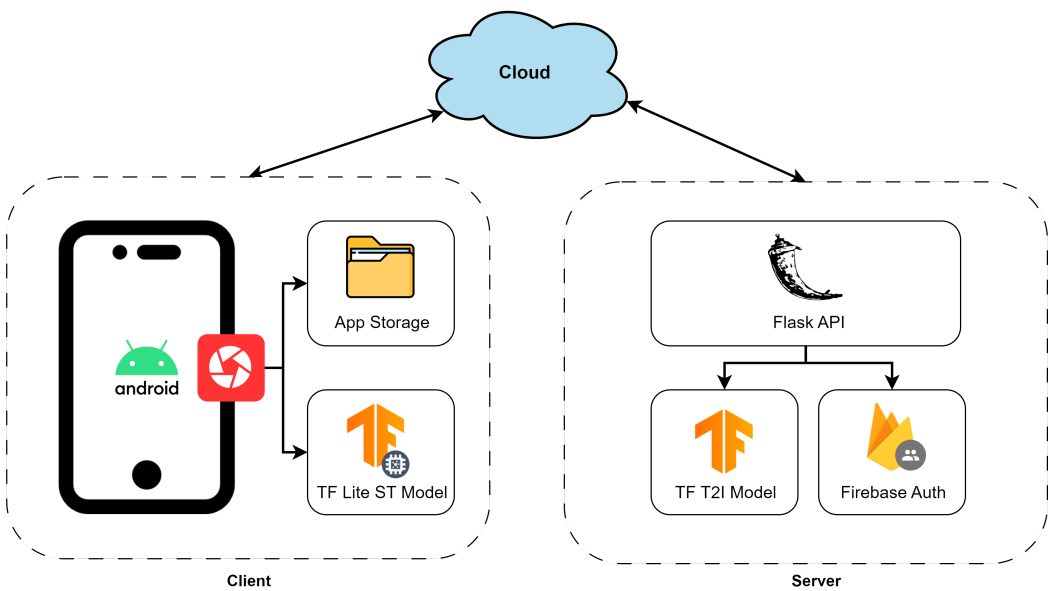


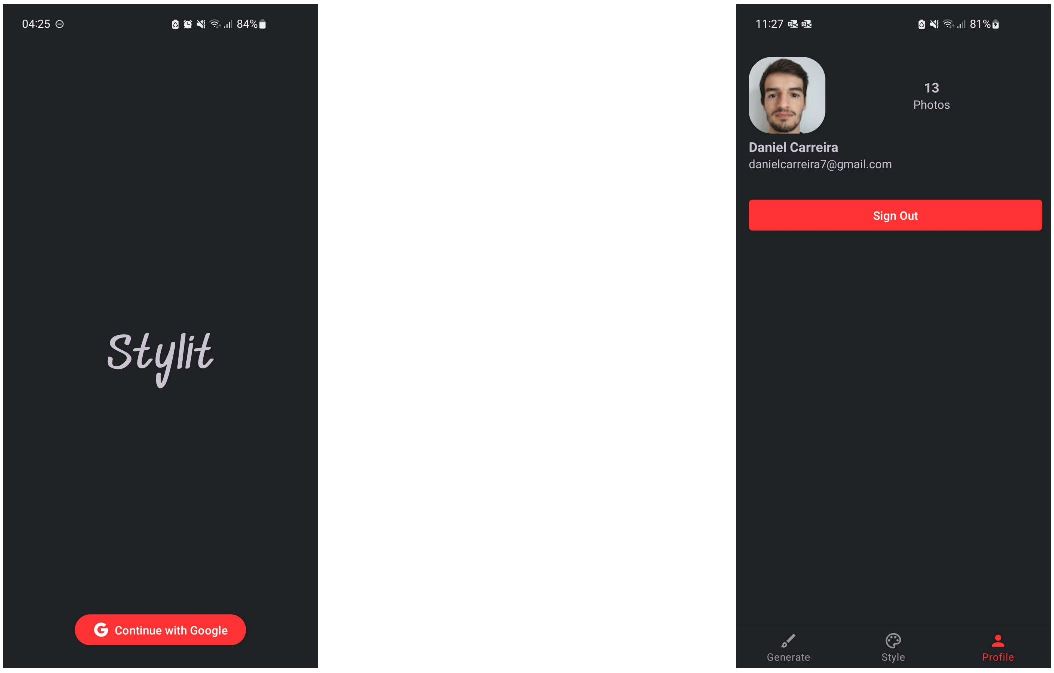
Figure 6. System overview diagram.

3.3.3. Functionality Overview

When the user opens the application for the first time, they are required to authenticate by logging in with their Google account. This authentication process is essential to ensure the security and proper utilization of the application's functionalities (Figure 7).

When the user completes his authentication, the user is greeted with the primary feature: the generation functionality, which is responsible for Text-to-Image Synthesis. To utilize this functionality, the user needs to enter their desired prompt into the designated text field and once the prompt and proceed by clicking on the "send" button, triggering the image generation process. Upon completion, the generated image is presented to the user where they have the option to open the image and to save it to their device storage (Figure 8).

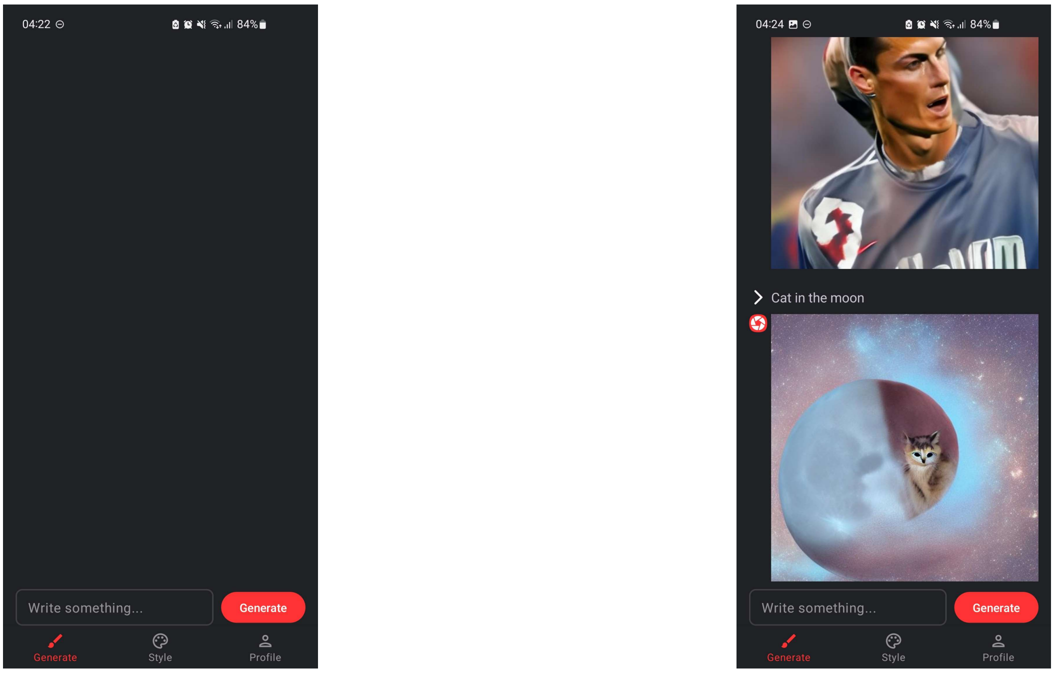
The next functionality of the application is the Style Transfer feature, which can be accessed by clicking on the "Style" option in the navigation bar. Upon entering the Style Transfer view, users are presented with a collection of saved images from the Generate



(a) Login View.

(b) User Profile View.

Figure 7. Text-to-Image Synthesis on the mobile application.



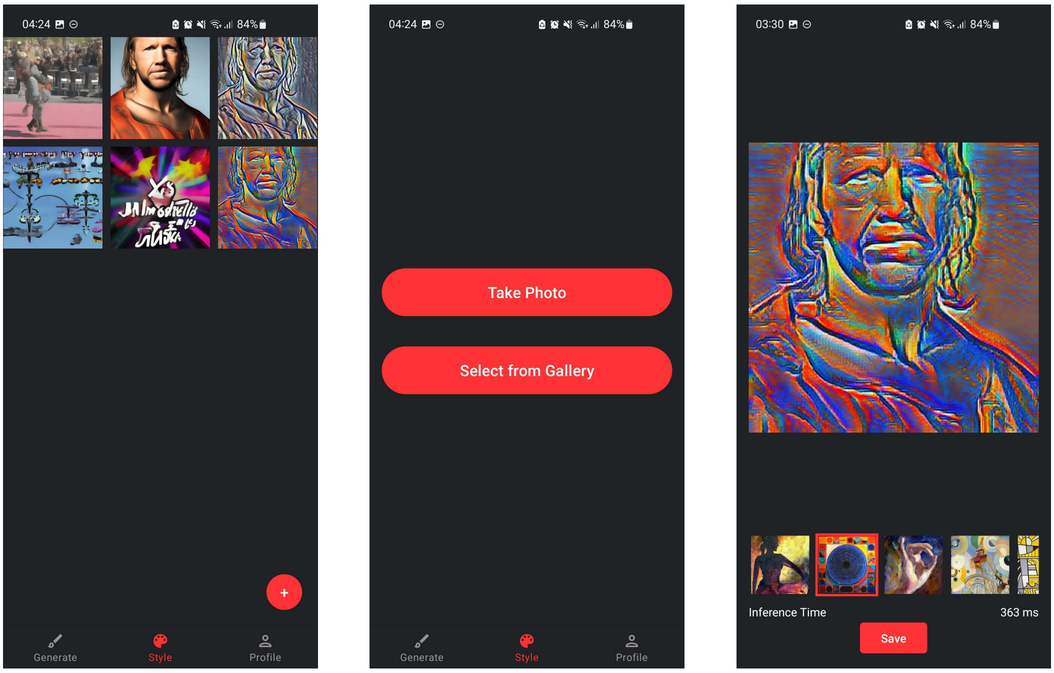
(a) Text-to-image Generator View.

(b) Image Results From Inserted Texts.

Figure 8. Style transfer on the mobile application.

feature. These images serve as the starting point for the style transfer process. Additionally, users have the flexibility to import images from their device's gallery or even capture a photo in real-time using the device's camera by clicking the button with the plus icon in the bottom right of the view. Once a image is selected or captured, users are presented with a diverse range of artistic styles to choose from which can range from famous artworks, such as paintings. to abstract or modern artistic interpretations. By selecting a specific style, users can apply it to their chosen image, instantly transforming its visual appearance.

After applying the desired style, users have the option to save the stylized image so that it can be easily accessed later within the app or saved to the device's gallery for external use (Figure 9).



(a) Gallery of saved images. **(b)** Import images. **(c)** Applying Style transfer.

Figure 9. Style transfer on the mobile application.

By clicking on the profile page, users have the opportunity to access and personal profile. Within the profile page, users can view the number of images they have already saved and make a log out from their account.

4. Results

All tests were executed on an RTX 3080 10GB and 32GB of RAM, with the TensorFlow framework.

4.1. Text-to-Image Synthesis Model

In this section, we will explore the performance of the chosen text-to-image synthesis model, Stable Diffusion, and the quality of the images generated by the prompts inserted by the user.

4.1.1. Performance

The performance of the Stable Diffusion text-to-image synthesis model was assessed through comprehensive evaluations. Both qualitative and quantitative assessments were conducted to measure the effectiveness and accuracy of the model in transforming textual inputs into visually coherent images.

There is a concept called inference steps, and this means the number of steps que inference will run per image, the more run you execute, the more resolution you will get. The starting point we used was 5 inference steps, and we additionally tested the impact of increasing this parameter, and how it affects the inference time and image quality.

Table 1. Stable Diffusion Inference Times.

Inference Steps	Time (s)
5	0.632
10	0.971
50	5.172

Given the results presented on Table 1, we can conclude the obvious, running the model for more inference loops will result in a higher-quality image in exchange for inference time. Our solution implemented 10 inference steps since was, to our eyes, the perfect balance.

4.1.2. Prompts

The prompts used in the evaluation of the Stable Diffusion text-to-image synthesis model played a crucial role in assessing the model's performance and understanding its ability to generate visually coherent images based on textual input. The evaluation of the text-to-image synthesis model involved conducting tests based on various prompt texts (Figure 10).

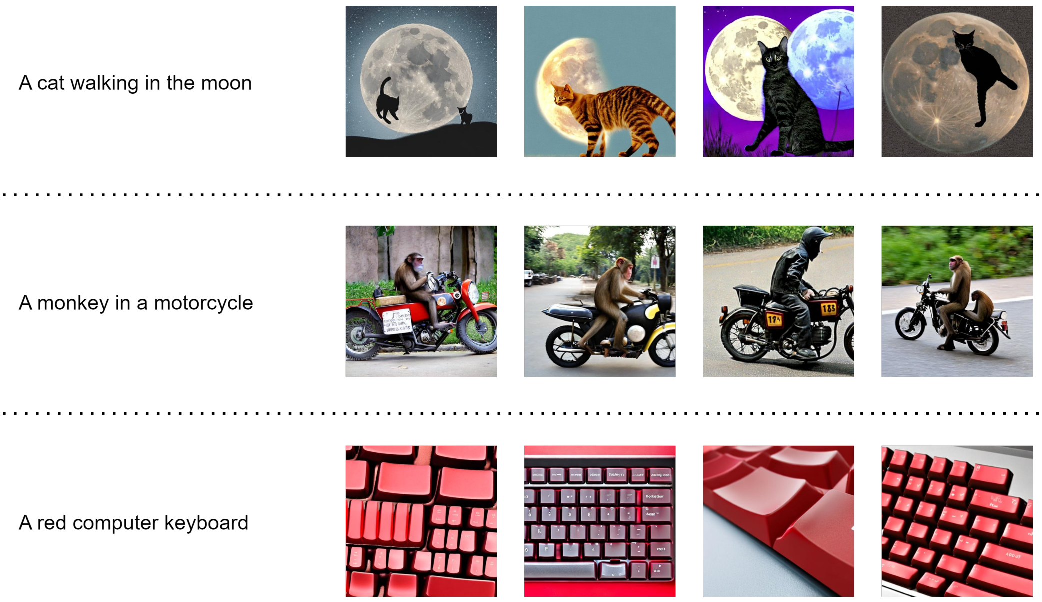


Figure 10. T2I Stable Diffusion - Prompt samples.

The images generated from the model, as shown in Figure 10, demonstrated consistency with the provided text prompts. For example, when given the prompt "A cat in the moon," the generated image depicted a cat and a moon as mentioned in the text. However, in terms of realism, the generated images fell short. On the other hand, when the prompt "A monkey in a motorcycle" was given, the resulting image was exceptionally well-done, capturing the essence of a monkey riding a motorcycle flawlessly.

4.2. Style Transfer Model

In this section, we will assess the performance of the style transfer model applied to the photos in our gallery. We will examine how well the model successfully transfers artistic styles to the images and evaluate the overall effectiveness of the style transfer process.

4.2.1. Performance

Within this section, we aim to provide an extensive evaluation of the style transform model algorithm employed in our mobile application. The analysis encompasses the model's performance across various scenarios, including low-quality and high-quality images. Additionally, we conducted tests to measure the time it takes to apply different styles to the same image. For this comparison, we used the following hardware: "Samsung S20 FE", "OnePlus 8T", and "Samsung S10".

The application of styles to various types of images, whether they are of low or high quality, takes nearly the same amount of time. Additionally, different styles were tested on the same image, and the time required to apply each style was found to be nearly identical.

With an average duration of approximately 300-350 ms per image, the process of applying styles to different images showcases remarkable efficiency and speed. This quick turnaround time not only demonstrates the effectiveness of the style transfer process but also offers users a convenient way to explore and visualize various styles within a short timeframe, considering that the model is running on the device.

4.2.2. Style variations

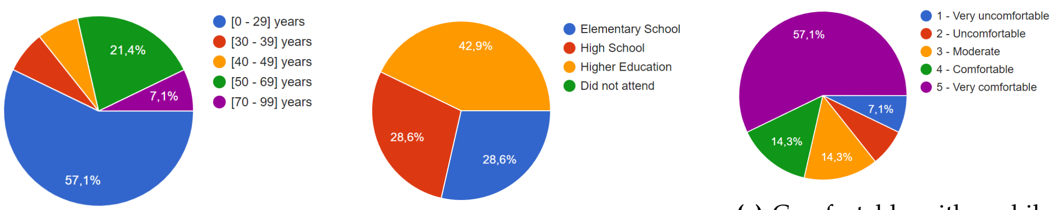
The objective of this subsection was to analyze and evaluate the performance of various style transfer models when applied to both the same image in our gallery and different images with similar styles. By comparing the results obtained from different style transfers, we aimed to gain information into their effectiveness and determine their compatibility with specific image characteristics.

Our observations revealed that the choice of style significantly influenced the outcome, depending on the unique characteristics of each image. We found that images featuring simpler details and more illustrative elements exhibited superior results when paired with certain style transfers. Conversely, images that captured high-quality real environments and people were only subtly affected or enhanced by various style transfers.

Additionally, we discovered that the suitability of a particular style for a given image depended on the level of detail and color composition present in the user-provided image. These factors played a crucial role in determining the success of style transfer and the fidelity of the final result.

4.3. Stylit App

In order to thoroughly test the usability of our application, we have undertaken a comprehensive approach. As part of this process, we have designed a usability test form to gather valuable feedback from a diverse range of individuals representing various domains and backgrounds.



(a) Age Range. **(b)** Education. **(c)** Comfortable with mobile devices.

Figure 11. User variety - Usability testing.

The data analysis in Figure 11 indicates a diversified user base, with responses spanning a wide variety of age groups, educational backgrounds, and mobile device comfort levels. This age range variety gives important information regarding the mobile application's potential performance and usability across generations. Furthermore, the range of educational levels implies the necessity for a user interface that caters to people from various educational backgrounds, assuring inclusivity and simplicity of use. Furthermore, the majority of respondents showing a high degree of comfort with mobile devices suggests that the target audience is well-equipped to easily navigate the app's features and functionalities.

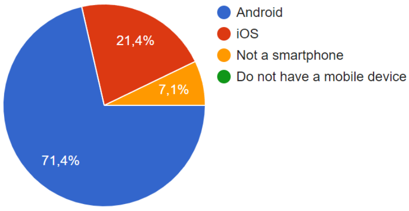
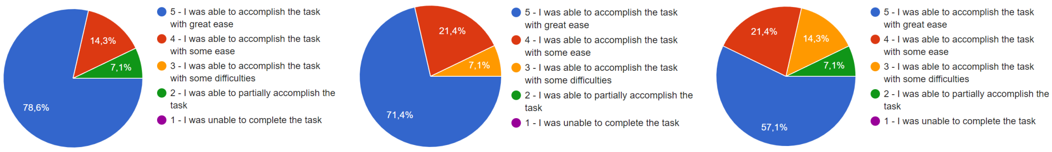


Figure 12. Most commonly used mobile operating system - Usability testing

Among the respondents, Android is the most commonly used mobile operating system, followed by iOS. Additionally, there were users who indicated not having a smartphone (Figure 12).



(a) Login task. **(b)** Take picture task. **(c)** Generating images.

Figure 13. Task to complete the usability test

The data in Figure 13 reflects positive user experiences regarding task performance across all four tasks. The majority of respondents reported accomplishing the tasks with great ease, as indicated by ratings of 4 or 5. This suggests that the mobile application's user interface and interaction design are effective in guiding users through these tasks, allowing them to complete them successfully. It indicates that the app provides a user-friendly experience and meets the expectations of users in terms of task completion.

5. Conclusion

As we conclude our paper, we want to emphasize the profound significance of the developed mobile app for generative AI and its incredible potential to revolutionize the realm of creative expression. Through the utilization of text-to-image synthesis and style transfer, our app serves as a dynamic platform empowering users to explore their artistic potential and actively engage in the transformative field of generative AI.

Our mobile app offers an intuitive and accessible interface, designed to seamlessly guide users in unleashing their artistic imagination by generating personalized images with customizable styles and applying styles to their desired pictures.

Looking ahead, we envision an exciting future characterized by further advancements in image generation techniques, refined style transfer algorithms, and expanded application domains for our mobile app, such as video style transfer.

References

1. Frolov, S.; Hinz, T.; Raue, F.; Hees, J.; Dengel, A. Adversarial Text-to-Image Synthesis: A Review. *CoRR* **2021**, *abs/2101.09983*, [2101.09983].
2. Borji, A. Generated Faces in the Wild: Quantitative Comparison of Stable Diffusion, Midjourney and DALL-E 2, 2023, [arXiv:cs.CV/2210.00586].
3. Reed, S.; Akata, Z.; Yan, X.; Logeswaran, L.; Schiele, B.; Lee, H. Generative Adversarial Text to Image Synthesis. In Proceedings of the Proceedings of The 33rd International Conference on Machine Learning; Balcan, M.F.; Weinberger, K.Q., Eds.; PMLR: New York, New York, USA, 2016; Vol. 48, *Proceedings of Machine Learning Research*, pp. 1060–1069.
4. Gunjate, S.; Nakhate, T.; Kshirsagar, T.; Sapat, Y.; Guhe, S. Sketch to Image using GAN.
5. Peng, C.; Zhang, C.; Liu, D.; Wang, N.; Gao, X. Face photo–sketch synthesis via intra-domain enhancement. *Knowledge-Based Systems* **2023**, *259*, 110026. https://doi.org/https://doi.org/10.1016/j.knosys.2022.110026.
6. Jia, H.; Wang, Q.; Tov, O.; Zhao, Y.; Deng, F.; Wang, L.; Chang, C.L.; Hou, T.; Grundmann, M. BlazeStyleGAN: A Real-Time On-Device StyleGAN. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, 2023, pp. 4689–4693.

7. Shih, Y.; Paris, S.; Barnes, C.; Freeman, W.T.; Durand, F. Style Transfer for Headshot Portraits. *ACM Trans. Graph.* **2014**, *33*. https://doi.org/10.1145/2601097.2601137.
8. Luan, F.; Paris, S.; Shechtman, E.; Bala, K. Deep Photo Style Transfer. In Proceedings of the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.
9. Ngo, L.M.; de Wiel, C.a.; Karaoglu, S.; Gevers, T. Unified Application of Style Transfer for Face Swapping and Reenactment. In Proceedings of the Proceedings of the Asian Conference on Computer Vision (ACCV), 2020.
10. Demosthenous, G.; Vassiliades, V. Continual Learning on the Edge with TensorFlow Lite. *CoRR* **2021**, *abs/2105.01946*, [2105.01946].
11. Zhao, W.X.; Zhou, K.; Li, J.; Tang, T.; Wang, X.; Hou, Y.; Min, Y.; Zhang, B.; Zhang, J.; Dong, Z.; et al. A Survey of Large Language Models, 2023, [arXiv:cs.CL/2303.18223].
12. Lu, Y.; Yang, X.; Li, X.; Wang, X.E.; Wang, W.Y. LLMScore: Unveiling the Power of Large Language Models in Text-to-Image Synthesis Evaluation, 2023, [arXiv:cs.CV/2305.11116].
13. Kortoçi, P.; Liang, Y.; Zhou, P.; Lee, L.H.; Mehrabi, A.; Hui, P.; Tarkoma, S.; Crowcroft, J. Federated Split GANs; Association for Computing Machinery: New York, NY, USA, 2022; FedEdge '22, p. 25–30. https://doi.org/10.1145/3556557.3557953.
14. Lim, W.Y.B.; Luong, N.C.; Hoang, D.T.; Jiao, Y.; Liang, Y.C.; Yang, Q.; Niyato, D.; Miao, C. Federated Learning in Mobile Edge Networks: A Comprehensive Survey. *IEEE Communications Surveys & Tutorials* **2020**, *22*, 2031–2063. https://doi.org/10.1109/COMST.2020.2986024.
15. Schuhmann, C.; Beaumont, R.; Vencu, R.; Gordon, C.; Wightman, R.; Cherti, M.; Coombes, T.; Katta, A.; Mullis, C.; Wortsman, M.; et al. LAION-5B: An open large-scale dataset for training next generation image-text models, 2022, [arXiv:cs.CV/2210.08402].
16. Zhou, I. Paper explained - laion-5b, 2022.
17. Rombach, R.; Blattmann, A.; Lorenz, D.; Esser, P.; Ommer, B. High-Resolution Image Synthesis with Latent Diffusion Models. *CoRR* **2021**, *abs/2112.10752*, [2112.10752].
18. Alammar, J. The Illustrated Stable Diffusion, 2022.