**Q1: How long does it take to train the model, and what strategies did you use to optimize training time?**

A1: Training takes considerable time, often more than 30 hours. To optimize this, we ensured that our model's architecture was as efficient as possible. Additionally, we ran training sessions in parallel when feasible and utilized the best hardware accelerators available on Colab.

**Q2: How do you balance the weights between adversarial loss, cycle consistency loss, and identity loss?**

A2: The balancing of these weights is critical and often determined empirically. We selected the recommended values from the paper.

**Q3: Can you explain more about the identity loss and its impact on the results?**

A3: Identity loss is crucial for preserving color composition and the overall structure of the target style images. When the generator receives an image from the target domain, we want it to minimize changes. This helps the model to learn not only the transformation but also the retention of the style within the domain, resulting in more stable and coherent translations.

**Q4: Why did you choose the fourth layer to calculate the content loss in NST?**

A4: The fourth layer is typically where higher-level features representing the image's content are captured in a VGG network. By calculating content loss at this layer, we ensure that the essence of the original image's content is preserved while still allowing for stylistic modifications.

**Q5: Is there a particular reason for using VGG as the network for NST, and have you considered other networks?**

A5: We used VGG because it's well-known for capturing rich feature representations and has been widely successful in style transfer tasks. However, we are open to exploring other networks that could offer improved performance or efficiency, which could be an area for future work.

**Q6: What inspired your team to choose the Ghibli style for your photo transformation project?**

A6: We were inspired by the unique and visually stunning characteristics of Studio Ghibli's films, such as their meticulous attention to detail, hand-drawn animation, and emotionally resonant storytelling. We wanted to explore how these distinct qualities could be applied to transform ordinary photos into artworks resembling Ghibli's style.

**Q7: Can you explain the role of CycleGAN and NST in your project?**

A7: CycleGAN is the main model for style transfer in our project. NST is a kind of baseline or other comparison approach.

**Q8: What future improvements or expansions do you envision for this project?**

A8: Future improvements could include optimizing the training process to reduce time and computational requirements, exploring more advanced loss functions for better quality transfers, finding other feasible metrics, and expanding the dataset to include a wider range of Ghibli styles and themes.

**Q9: Why didn’t you choose a numerical metric?**

A9: We selected FID, but FID will not capture the artistic nuances and fine detail of style that is unique to Ghibli artwork since it relies on features extracted by Inception v3, which is trained on ImageNet and may, therefore, not be very sensitive to stylistically different features spaces of Ghibli art.

**Q10: Why was human evaluation chosen alongside numerical metrics?**

A10: Human evaluation was chosen to complement numerical metrics because of the project's artistic nature. While numerical metrics like FID can provide objective data on the similarity between generated and real images, they cannot fully capture the subjective and aesthetic qualities unique to Ghibli's artwork. Human evaluation allows us to assess the emotional and artistic impact of the style transfer, which is central to the project's goals.

**Q11: How was the human evaluation conducted?**

A11: The evaluation was done by having each team member independently rate a set of images generated by the models based on the predefined rubric. Scores for each aspect were then averaged for each image and across all evaluators to provide a holistic view of the model's performance in achieving the desired style transfer.

**Q12: Does the Loss curve make sense?**

A12: Despite some fluctuations, the loss curves showed a general downward trend, which indicates the model was learning as intended. However, due to the extensive time needed for training, we limited the training to only 100 epochs. We expect a more pronounced downward trend if we extend the number of epochs for the model.

**Q13: What were the main sources of your data?**

A13: The main sources of our data were varied, merging images sourced directly from Ghibli's official websites and some ready-made datasets, especially for landscape images and Ghibli-style images from "The Wind Rises."

Why batch size = 1

CycleGAN is designed for unpaired image-to-image translation tasks, where there is no one-to-one correspondence between images in the source and target domains. Training with a batch size of 1 is naturally suited to this setup, as it does not require batching together corresponding pairs of images, which may not exist.

How many residual blocks?

9, 6

How do you separate the dataset

10%

Do you need the Validation set?

No, unsupervised, loss is self-validation

Is this kind of auto-encoder architecture?

Similar but not same, for dimensional compression  
Is this kind of diffusion model architecture?

No, from noise