# NAAN MUTHALVAN

**NAME :** MUTHULAKSHMI R
**REG. NO :** 712221104009
**YEAR & SEM :** 3rd yr & 06
**COLLEGE :** PARK COLLEGE OF ENGINEERING AND TECHNOLOGY

## GENERATIVE AI

# UNPAIRED- IMAGE-TO-IMAGE TRANSLATION-CYCLEGAN

# AGENDA

1. Problem Statement
2. Project Overview
3. End Users
4. Our Solution and Proposition
5. Key Features
6. Modelling Approach
7. Results and Evaluation
8. Conclusion

# PROBLEM STATEMENT

**1.Domain Shift:** Unpaired image translation often involves domains with significant differences in visual appearance, such as day to night conversion or photo to painting transformation. Domain shift refers to the misalignment between the distributions of source and target domains, leading to poor translation quality and artifacts in the generated images.

**2.Mode Collapse:** CycleGANs are prone to mode collapse, where the generator fails to capture the entire distribution of the target domain, resulting in limited diversity and repetitive patterns in the generated images. Mode collapse reduces the richness and realism of the translated images, undermining the quality of the translation.

**3.Semantic Ambiguity**: In unpaired image translation, preserving semantic content while changing visual attributes is crucial for maintaining the integrity and meaning of the images. However, existing methods struggle to disentangle content from style, leading to semantic ambiguity and distortion in the translated images

# PROJECT OVERVIEW

The project focuses on advancing the state-of-the-art in unpaired image-to-image translation using Cycle-Consistent Generative Adversarial Networks (CycleGANs). Unpaired image translation has immense potential for various applications but faces significant challenges such as domain shift, mode collapse, and semantic ambiguity. This project aims to tackle these challenges and enhance the performance of CycleGANs for unpaired image translation tasks.

**Objectives:**

1. Identify and understand the key challenges in unpaired image-to-image translation with CycleGANs.

2. Develop novel techniques and methodologies to address domain shift, mode collapse, and semantic ambiguity.

3. Design robust evaluation metrics and frameworks for comprehensive performance assessment.

4. Enhance the scalability, adaptability, and generalization capabilities of CycleGAN architectures.

5. Evaluate the proposed approaches on benchmark datasets and real-world applications to demonstrate their effectiveness.

## Approach:

The project adopts a multi-faceted approach to tackle the challenges in unpaired image-to-image translation with CycleGANs:

1.Research: Conduct an in-depth literature review to understand the existing methodologies and challenges in unpaired image translation. Identify gaps and opportunities for innovation.

2.Methodology Development: Develop novel techniques and methodologies to address domain shift, mode collapse, and semantic ambiguity. This involves exploring domain alignment strategies, diversity promotion techniques, semantic preservation mechanisms, and robust architecture designs.

3.Implementation: Implement the proposed approaches using state-of-the-art deep learning frameworks such as TensorFlow or PyTorch. Develop efficient training pipelines and experimental setups.

4.Evaluation: Design robust evaluation metrics and frameworks to quantitatively assess the performance of the proposed methods. This includes developing novel perceptual assessment tools and validation protocols.

5.Validation: Validate the effectiveness of the proposed approaches through extensive experiments on benchmark datasets and real-world applications. Compare the results with existing state-of-the-art methods to demonstrate improvements.

**Expected Outcomes:**

1. Novel techniques and methodologies for addressing challenges in unpaired image-to-image translation with CycleGANs.
2. Improved performance in terms of translation quality, diversity, and semantic preservation.
3. Robust evaluation metrics and frameworks for comprehensive performance assessment.
4. Insights into the underlying mechanisms of unpaired image translation with CycleGANs.
5. Open-source implementations and benchmark datasets for the research community.

# END USERS

1. **Entertainment Industry:** Professionals in the entertainment industry, including filmmakers, animators, and game developers, who require advanced image editing and manipulation tools to create immersive and visually stunning experiences for their audiences.
2. **Social Media Influencers:** Individuals who have a significant following on social media platforms and rely on visually appealing content to engage and attract their audience, such as fashion bloggers, lifestyle influencers, and travel enthusiasts.
3. **Personal Users:** Everyday users who enjoy editing and enhancing their personal photos for sharing with friends and family, creating custom artwork, or simply exploring their creativity in a fun and accessible way.

# OUR SOLUTION AND PROPOSITION

**1.Innovative Technology:** Our platform harnesses the cutting-edge power of unpaired image-to-image translation with CycleGANs, a state-of-the-art technology that enables seamless transformation of images across diverse domains. By leveraging advanced deep learning algorithms, we provide users with unprecedented flexibility and creativity in their image editing endeavors.

**2.User-Centric Approach:** At the heart of our platform is a commitment to delivering an exceptional user experience. We understand that every user is unique, with their own preferences, goals, and creative vision. That's why we've designed our platform to be intuitive, customizable, and accessible to users of all skill levels, ensuring that everyone can unleash their creativity with ease.

**3.Quality Assurance:** We take pride in delivering high-quality results that exceed our users' expectations. Our platform utilizes advanced quality assurance mechanisms to ensure that every transformation maintains the integrity, clarity, and fidelity of the original image. From color accuracy to texture preservation, we go above and beyond to deliver professional-grade results that elevate your images to new heights.

**4.Continuous Innovation:** We're committed to staying at the forefront of technological innovation and pushing the boundaries of what's possible in image transformation. Our team of researchers, developers, and designers are constantly exploring new techniques, algorithms, and features to enhance our platform and empower our users with even more creative possibilities.

**5.Empowering Creativity:** Our platform is more than just a tool for image editing – it's a catalyst for creativity and self-expression. Whether you're a professional artist, an amateur photographer, or simply someone who loves to create, our platform provides you with the tools and inspiration you need to bring your vision to life. With endless possibilities at your fingertips, the only limit is your imagination.

**6.Community Engagement:** We believe in the power of community and collaboration to fuel creativity and innovation. That's why we've built a vibrant community of users who share their ideas, experiences, and creations with one another. From sharing tips and tricks to collaborating on projects, our community is a place where creativity thrives and inspiration flourishes.

**7.Accessible Pricing:** We believe that creativity should be accessible to everyone, regardless of budget or resources. That's why we offer a range of pricing options to suit every need and budget, from free basic plans to premium subscriptions with advanced features. Whether you're a hobbyist, a professional, or somewhere in between, there's a plan that's right for you.

**8.Environmental Sustainability:** We're committed to minimizing our environmental impact and promoting sustainability in everything we do. From optimizing our algorithms to reduce energy consumption to partnering with eco-conscious suppliers and vendors, we're constantly striving to make a positive difference for the planet. With our platform, you can create with confidence, knowing that you're supporting a company that cares about the environment.

**9.Transparency and Trust:** We believe in building trust with our users through transparency, honesty, and integrity. That's why we're committed to being open and transparent about our practices, policies, and data handling procedures. From clear and concise terms of service to transparent pricing and billing practices, we strive to earn and maintain the trust of our users every step of the way

# KEY FEATURES

**1.Intuitive Interface:** Dive into the world of image transformation with ease through our intuitive user interface. With simple navigation and user-friendly controls, you'll feel right at home as you explore the possibilities of transforming your images.

**2.Diverse Transformation Options:** From changing the season of your landscapes to giving your portraits a vintage film vibe, our platform offers a diverse array of transformation options to suit every creative vision. Whether you're aiming for subtle enhancements or bold artistic statements, the possibilities are endless.

**3.High-Quality Results:** We understand that quality matters. That's why our platform utilizes advanced algorithms and cutting-edge techniques to ensure that every transformation maintains the integrity and clarity of your original images. Say goodbye to pixelation and artifacts, and hello to professional-grade results.

**4.Real-Time Preview:** Visualize your transformations in real-time and fine-tune your edits with precision. Our platform provides instant feedback, allowing you to experiment freely and make adjustments on the fly until you achieve the perfect look.

**5.Seamless Sharing:** Share your transformed images with the world with just a few clicks. Whether you're showcasing your latest creations on social media, printing them for display, or incorporating them into your projects, our platform makes it effortless to share your vision with others.

# MODELLING APPROACH

## Image-to-Image Translation

- Image-to-image translation is a class of vision and graphics problems.
- Our goal is to to learn the mapping between an input image and an output image using a training set of aligned image pairs.

# CNNs for Image-to-Image Translation

- CNNs are used in a wide variety of image prediction problems
- CNNs learn to minimize a loss function { used to score the quality of results
- But, What should we minimize?
- We take a naive approach and ask CNN to minimize Euclidean distance between predicted and ground truth pixels.
- But, it will produce blurry results.
- Coming up with loss functions that does what we really want - e.g., sharp and realistic images - is hard.

# Generative Adversarial Networks

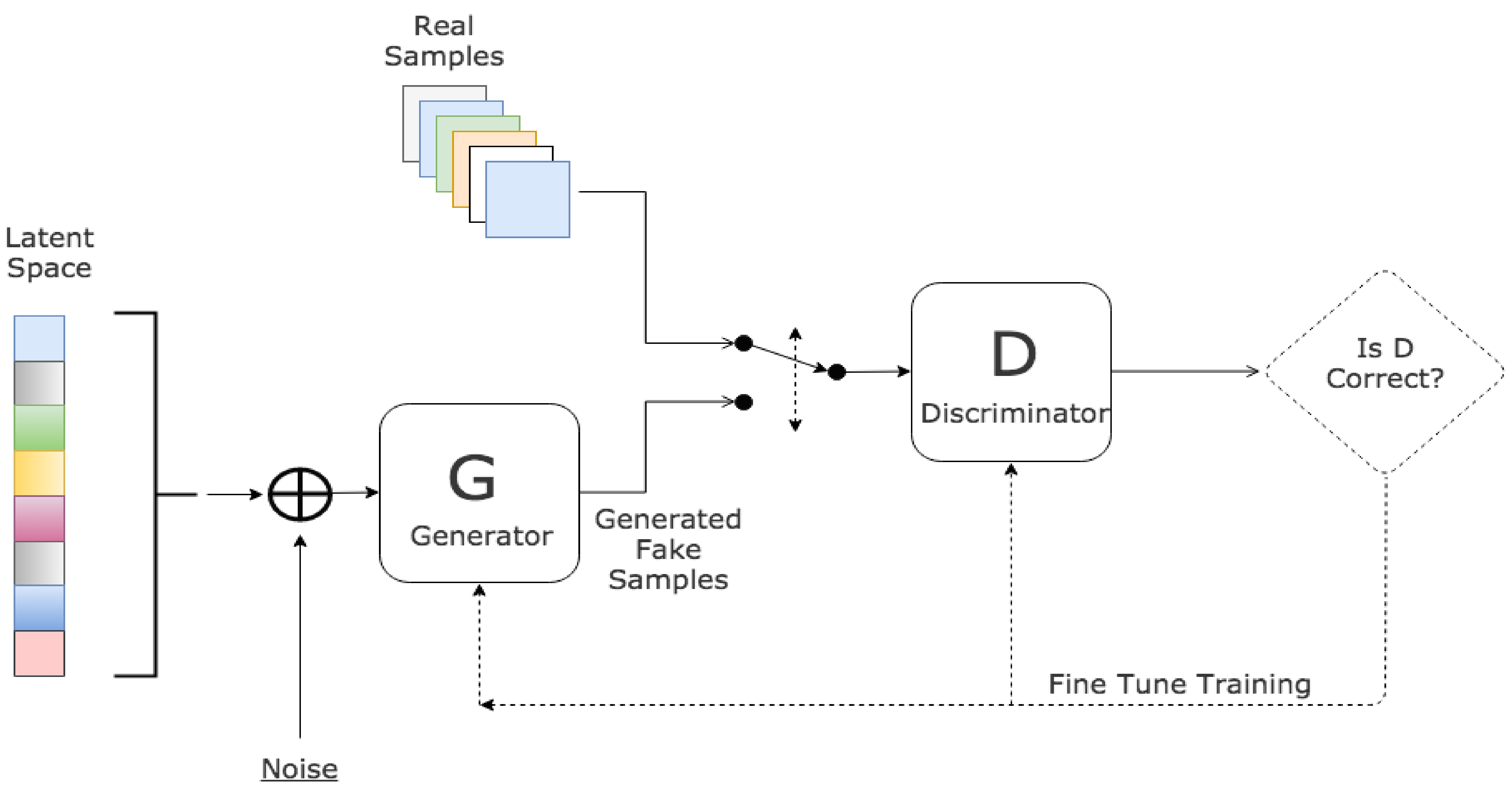
- GANs allows to specify high-level goal. For example, "make the generated images indistinguishable from real images".

- GANs simultaneously learn two models: a discriminative model D that tries to classify if the output image is real or fake, anda generative model G that captures the data distribution.

- Two neural networks contest with each other in a game.

- Yann LeCun described GANs as "the coolest idea in machine learning in the last twenty years".

# Generative Adversarial Networks



# Math behind GANs

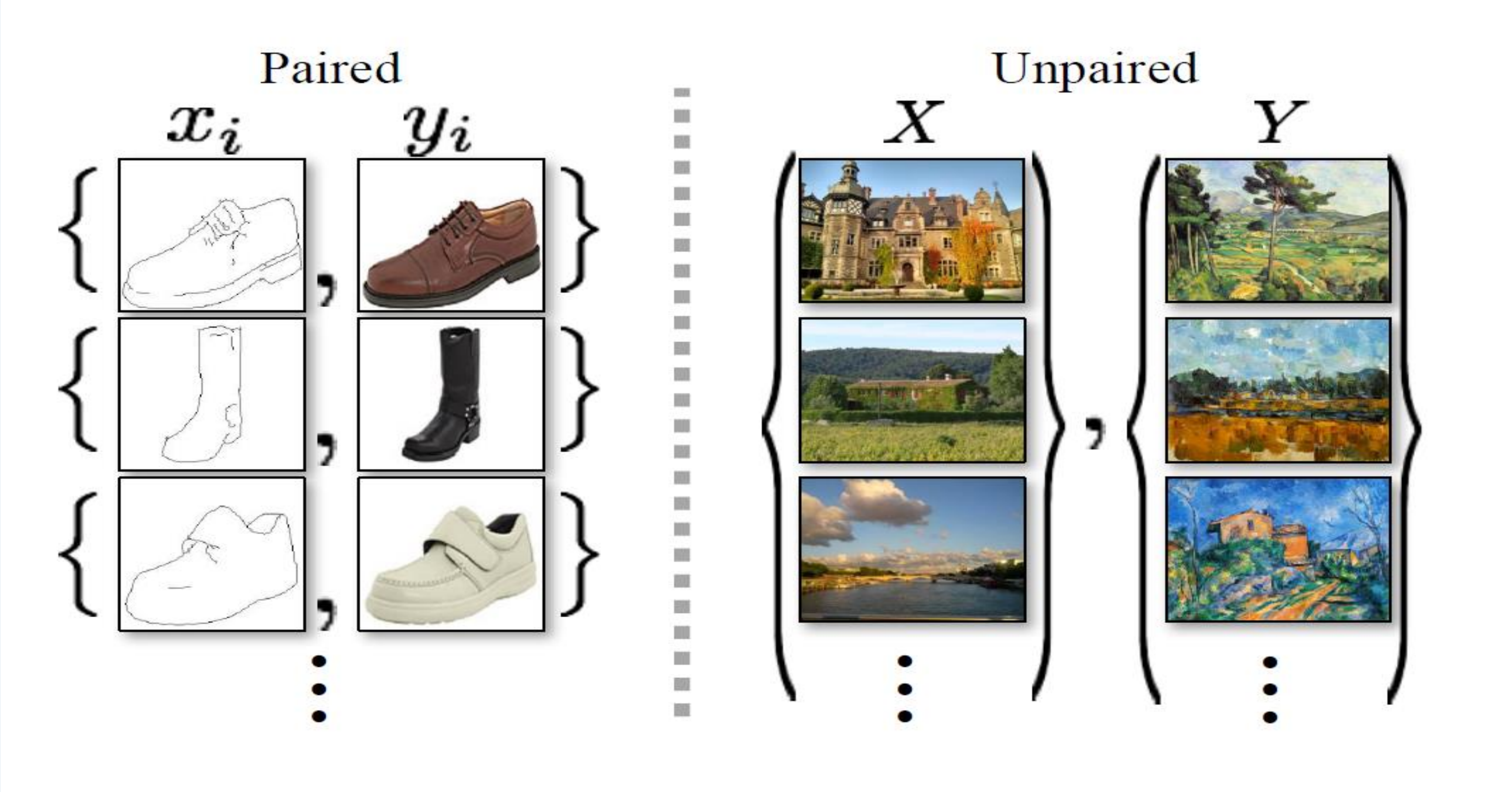
- Given a real data, we want D to maximize Expdata(x)[logD(x)].
- Meanwhile, given a fake sample G(z), we want D to maximize Ezpz (z)[log(1 − D(G(z)))] where D(G(z)) is the discriminator's estimate of the probability that a fake instance is real..
- However, G is trained to minimize log(1 − D(G(z))) simultaneously.

## Loss Function

In short, G and D are playing zero-sum game with following loss function:

$$\min_{G} \max_{D} V(D, G) = \mathbb{E}_{\boldsymbol{x} \sim p_{\text{data}}(\boldsymbol{x})}[\log D(\boldsymbol{x})] + \mathbb{E}_{\boldsymbol{z} \sim p_{\boldsymbol{z}}(\boldsymbol{z})}[\log(1 - D(G(\boldsymbol{z})))] \quad (2)$$

# Unpaired Image-to-Image Translation



# Problems with Paired Image-to-Image Translation

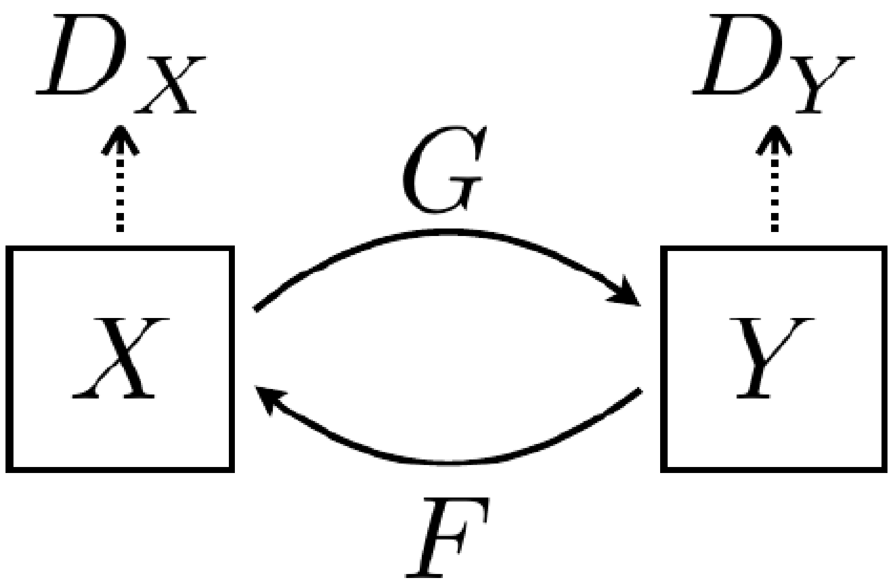
- Paired training data is difficult to obtain
- For object transfiguration (e.g., zebra$\leftrightarrow$horse), the desired output is not even well-defined

# Formulation

Goal: Learn mapping functions between two domains $X$ and $Y$ given training samples $\{x_i\}_{i=1}^{N}$ where $x_i \in X$ and $\{y_j\}_{j=1}^{M}$ where $y_j \in Y$. We denote the data distribution as $x \sim p_{data}(x)$ and $y \sim p_{data}(y)$.

# Adversarial Loss

Model includes two mappings $G : X \to Y$ and $F : Y \to X$. In addition to that, two adversarial discriminators $D_X$ and $D_Y$, where $D_X$ aims to distinguish between images $\{x\}$ and translated images $\{F(y)\}$; in the same way, $D_Y$ aims to discriminate between $\{y\}$ and $\{G(x)\}$.



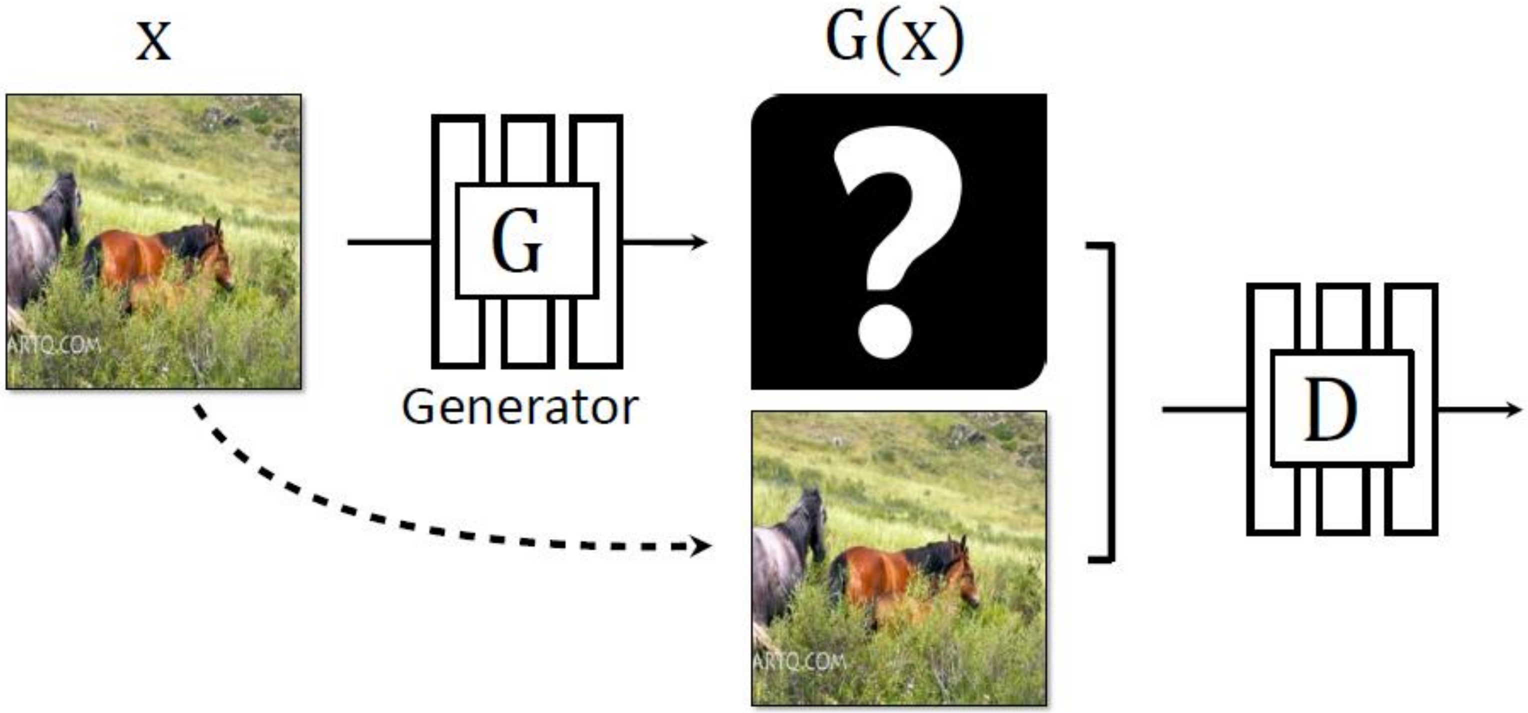
# Adversarial Loss

For the mapping function $G : X \rightarrow Y$ and its discriminator $D_Y$, we can write adversarial loss as:

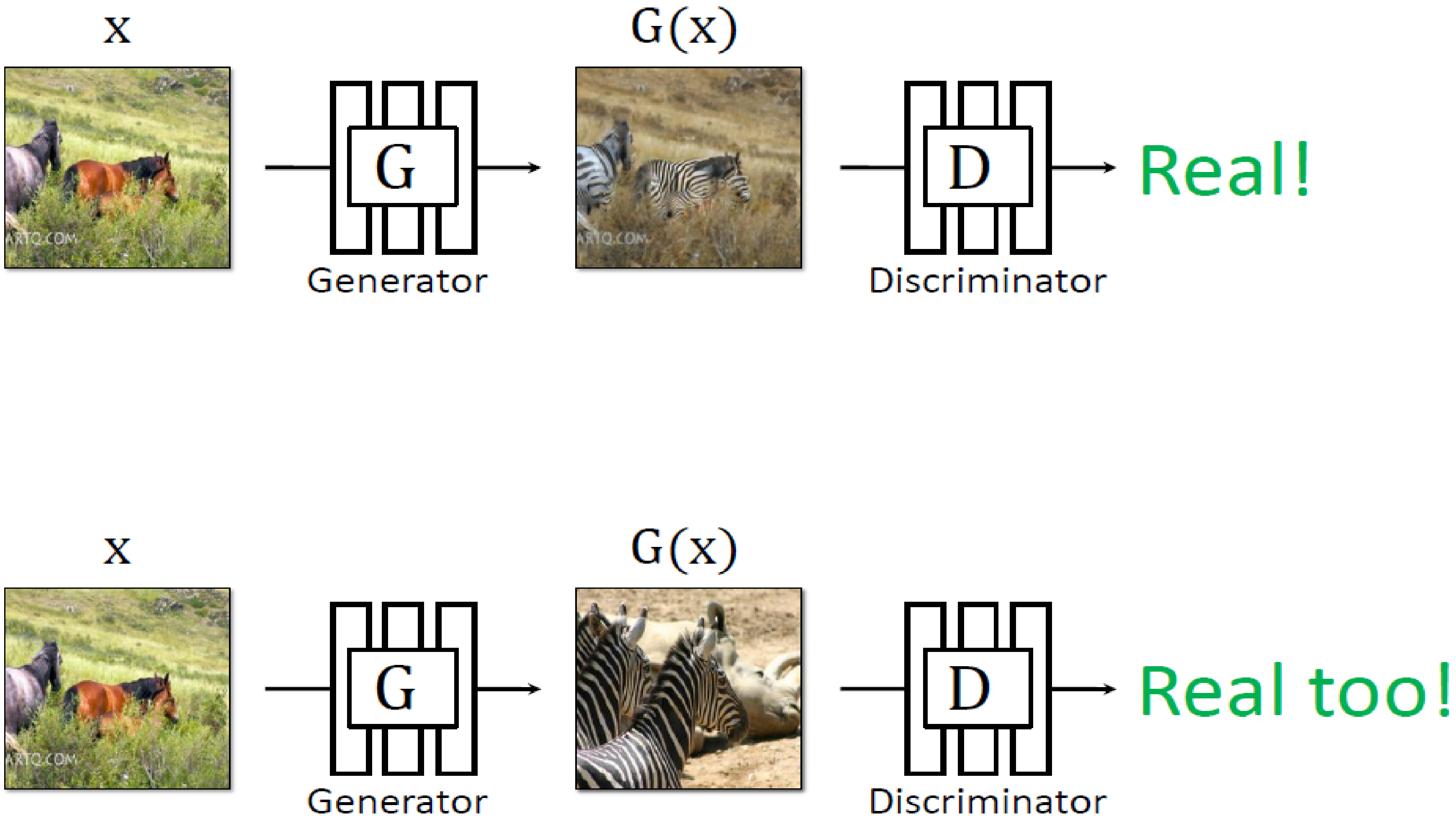
$$\mathcal{L}_{\mathrm{GAN}}(G, D_Y, X, Y) = \mathbb{E}_{y \sim p_{\mathrm{data}}(y)}[\log D_Y(y)] + \mathbb{E}_{x \sim p_{\mathrm{data}}(x)}[\log(1 - D_Y(G(x))], \quad (3)$$

where $G$ will try to generate images indistinguishable from $Y$, on the other hand $D_y$ learns to distinguish between fake and real images.

# Mode Collapse

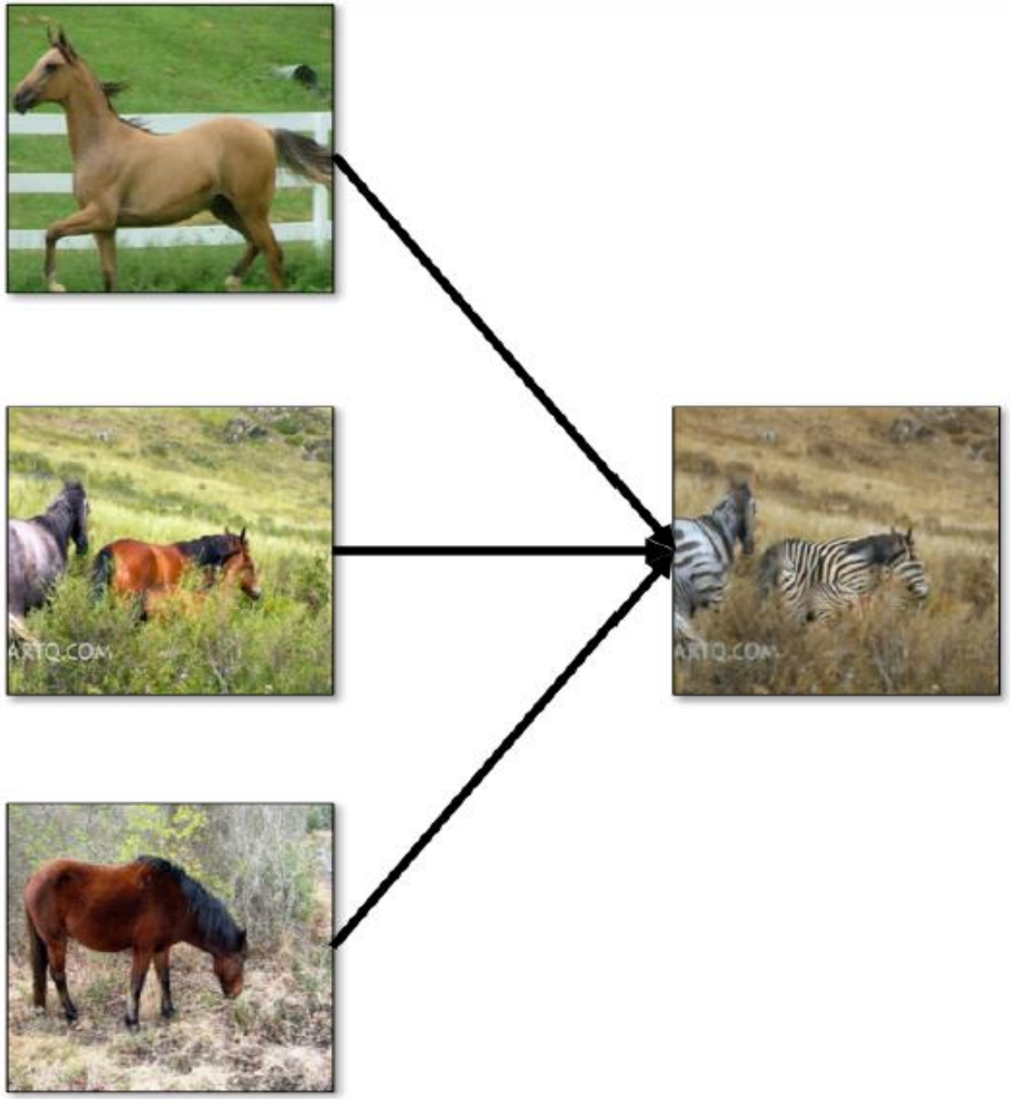


# Mode Collapse

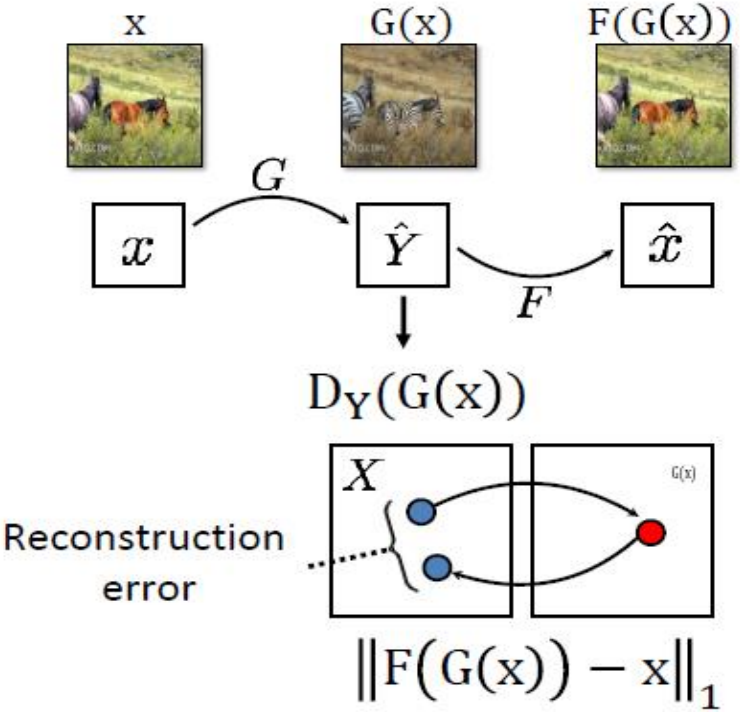


# Mode Collapse

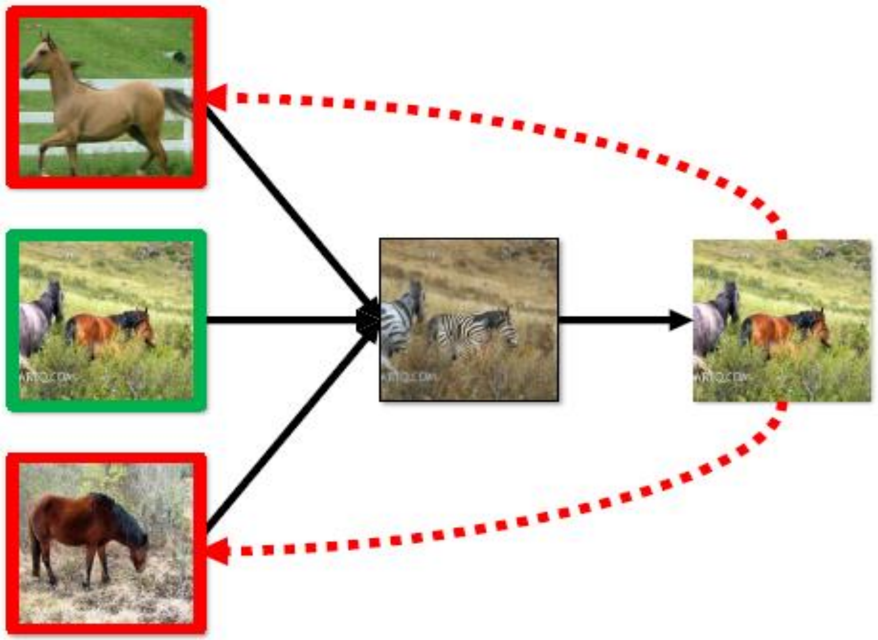
Mode Collapse!
GANs do **not** force output to correspond to input.



# Cycle Consistency Loss

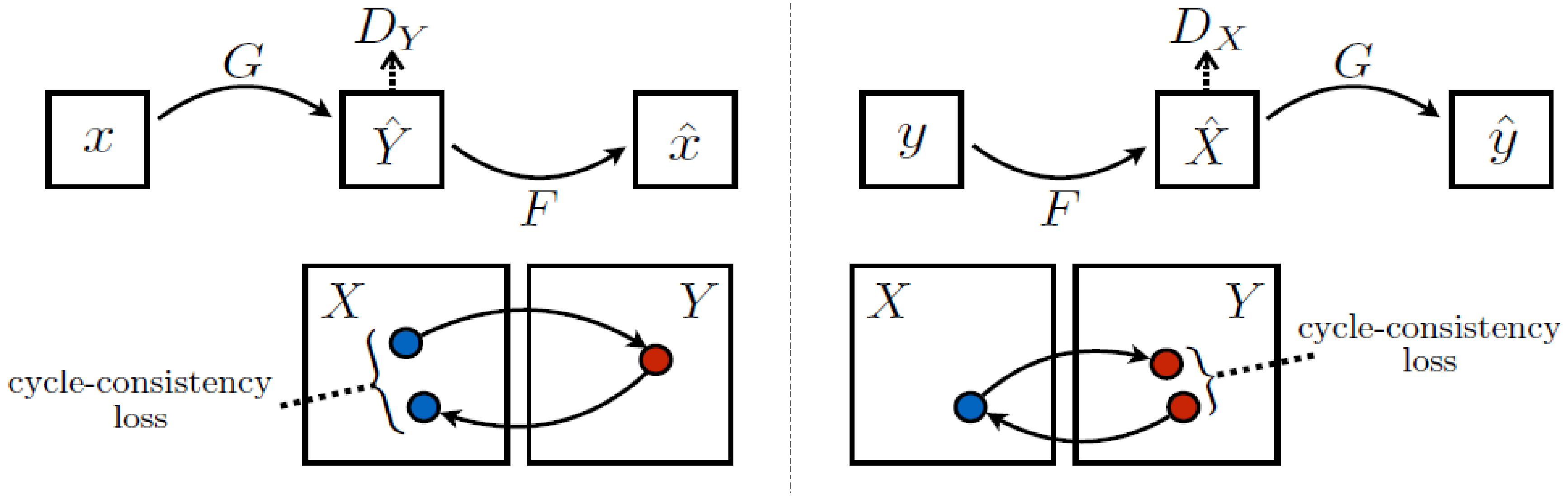



Large cycle loss



[Zhu*, Park*, Isola, and Efros, ICCV 2017]

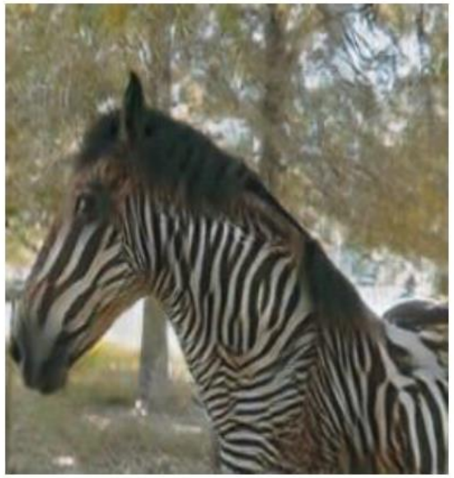
# Cycle Consistency Loss




For each image $x$ from domain $X$, the image translation cycle should be able to bring $x$ back to the original image, i.e., $x \rightarrow G(x) \rightarrow F(G(x)) \approx x$. This is called *forward cycle consistency*.

$$\begin{aligned} \mathcal{L}_{\text{cyc}}(G, F) = \; & \mathbb{E}_{x \sim p_{\text{data}}(x)}[\|F(G(x)) - x\|_1] \\ & + \mathbb{E}_{y \sim p_{\text{data}}(y)}[\|G(F(y)) - y\|_1], \end{aligned} \tag{4}$$

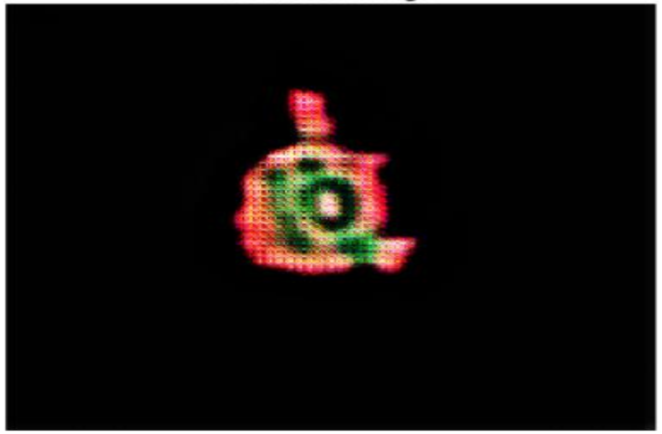
# RESULTS AND EVALUATION




```
/usr/local/lib/python3.6/dist-packages/PIL/Image.py:914: UserWarning: Palette images with Transparency   expressed in bytes s
hould be converted to RGBA images
  'to RGBA images')
{'epoch': 0, 'loss_G': 2.2598636150360107, 'loss_G_identity': 0.5154487490653992, 'loss_G_GAN': 0.6386906504631042, 'loss_G_c
ycle': 1.1057241559028625, 'loss_D': 0.4595997631549835}
```

Input Image

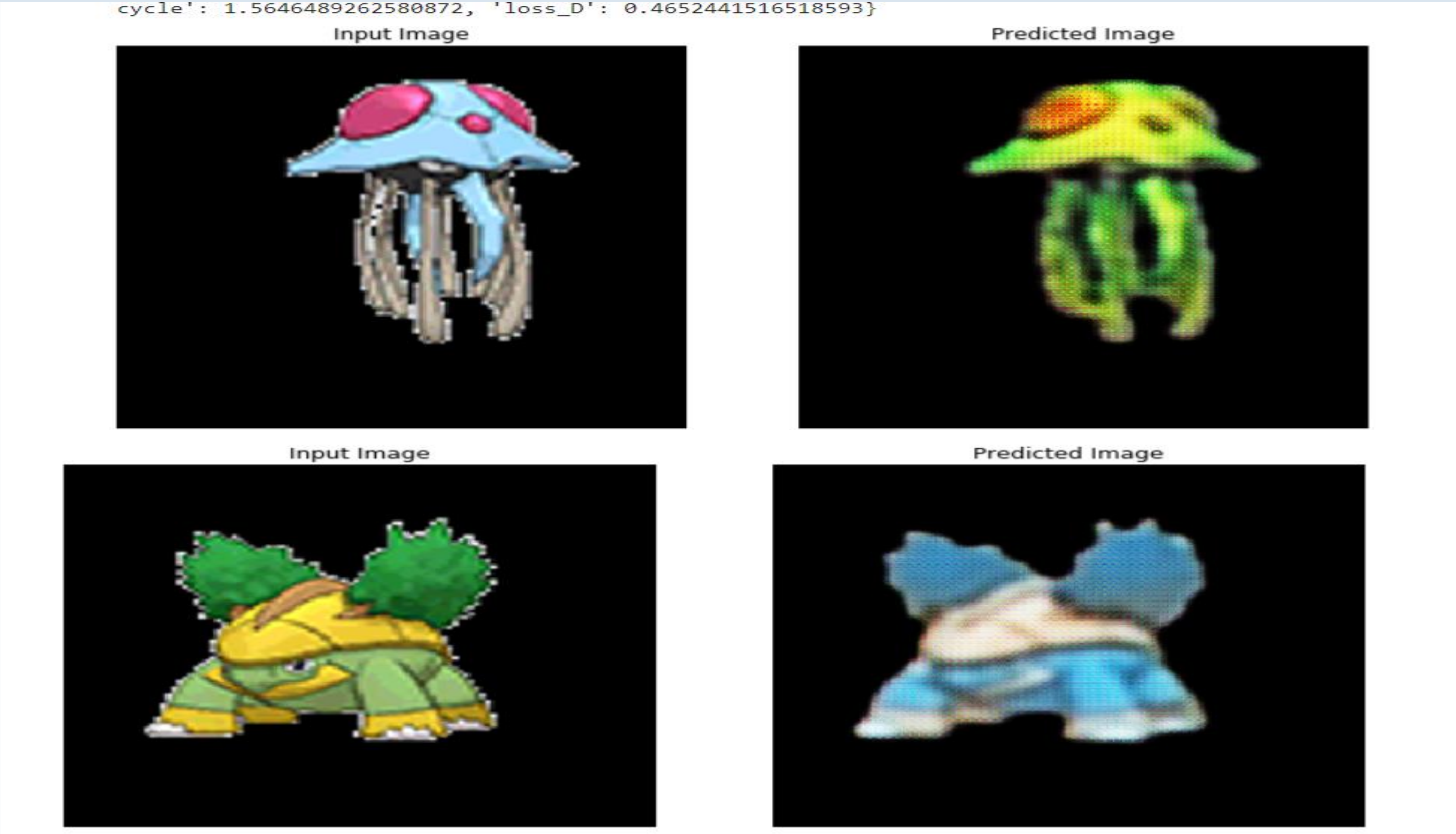
Predicted Image



Input Image

Predicted Image

```
Epoch : 1
/usr/local/lib/python3.6/dist-packages/PIL/Image.py:914: UserWarning: Palette images with Transparency   express
hould be converted to RGBA images
```


cycle': 1.5646489262580872, 'loss_D': 0.4652441516518593}
Input Image
Predicted Image
Input Image
Predicted Image

# CONCLUSION

**1.Ethical Implications:** As with any advanced technology, it's important to consider the ethical implications of unpaired image translation. This includes issues such as privacy concerns, potential misuse for deceptive purposes (e.g., deepfakes), and the impact on cultural and societal norms.

**2.Domain Adaptation:** Unpaired image translation with CycleGANs has shown promise for domain adaptation tasks, where the goal is to transfer knowledge learned from one domain to another. This has applications in areas such as medical imaging, where models trained on data from one hospital or imaging modality can be adapted to work with data from another.

**3.Interactive Editing:** Exploring ways to make the image translation process more interactive and user-guided could enhance the user experience. This might involve allowing users to provide feedback or guidance during the translation process to achieve more tailored results.

**4.Multi-Modal Translation:** Beyond simple style transfer, there's potential to explore multi-modal translation, where multiple attributes of an image can be modified simultaneously. For example, changing the season of a landscape while also adjusting the time of day or weather conditions.

**5.Transfer Learning:** Investigating techniques for leveraging pre-trained models or auxiliary datasets to improve the performance of CycleGANs, particularly in cases where limited labeled data is available for a specific task or domain.

# FAILURE CASES

- Caused by the distribution characteristics of the training datasets
- Fails at varied and extreme transformations, especially geometric changes

# Thank you...