



# 2D to 3D Animation Style Transfer

Timothy Le  
tle7@stanford.edu

Swathi Iyer  
swathii@stanford.edu

## Project Overview

### Motivation

3D animation is becoming increasingly popular amongst designers and gamers due to its realistic nature, which allows for a more engaging and immersive experience. However, it also suffers from high development cost in terms of time and resources, relative to 2D animation. For our project, we wanted to explore how we might transform low-cost 2D animation to 3D using style transfer methods.

### Problem Statement

Using several frames of both 2D and 3D animated clips, we will explore using feedforward neural style transfer, pix2pix, and CycleGAN style transfer methods to transfer 3D animation style onto 2D animation images.

## Data

- For the Feedforward Neural Style Transfer, we used an image from Monsters Inc (left) as our style image and an image from The Lion King as our content image (right)



- For the cycleGAN method, train: 53 frames from The Little Mermaid: Ariel's Beginning movie and 48 frames from Finding Dory; test: 45 Ariel, 81 Dory
  - Both movies take place in the ocean, and cycleGAN suggests having similar environments for best results
- The pix2pix model required us to have exact frames in 2 different styles, in order to learn the mappings between two different images. To do this, we found an animator's video of his manual recreation of the 2D animated show Rick and Morty in 3D animation, and extracted the frames.
  - Train/Val/Test split: 80/10/10
- We collected the data using a frame capturing tool called VLC media player
- Preprocessing: We used an online photo editing tool called Birme to standardize image dimensions

## Methods

### Feedforward Neural Style Transfer

- Uses a network based on the image classification VGG network to output an image  $x$ , which contains the content of image  $p$  in the style of image  $a$

$$L_{total}(\vec{p}, \vec{a}, \vec{x}) = \alpha L_{content}(\vec{p}, \vec{x}) + \beta L_{style}(\vec{a}, \vec{x})$$

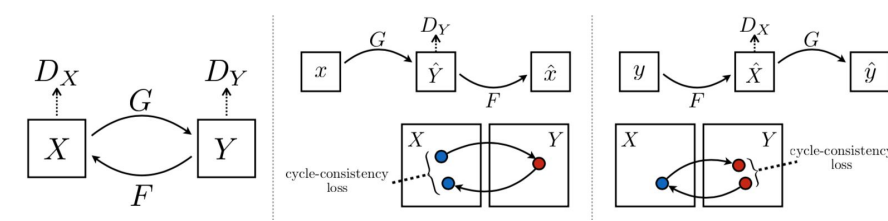
### Pix2Pix

- Conditional GAN
- Generator  $G$  trained to produce outputs that can't be distinguished from "real images, and discriminator  $D$  detects generator's "fakes"

$$G^* = \arg \min_G \max_D \mathcal{L}_{cGAN}(G, D) + \lambda \mathcal{L}_{L1}(G).$$

- Generators is an encoder followed by a decoder (conv net followed by a u-net)
- CycleGan
  - Seeks to find a  $G^*: X \rightarrow Y, F^*: Y \rightarrow X$  that minimizes the max over possible discriminators  $D_X$  and  $D_Y$  for:

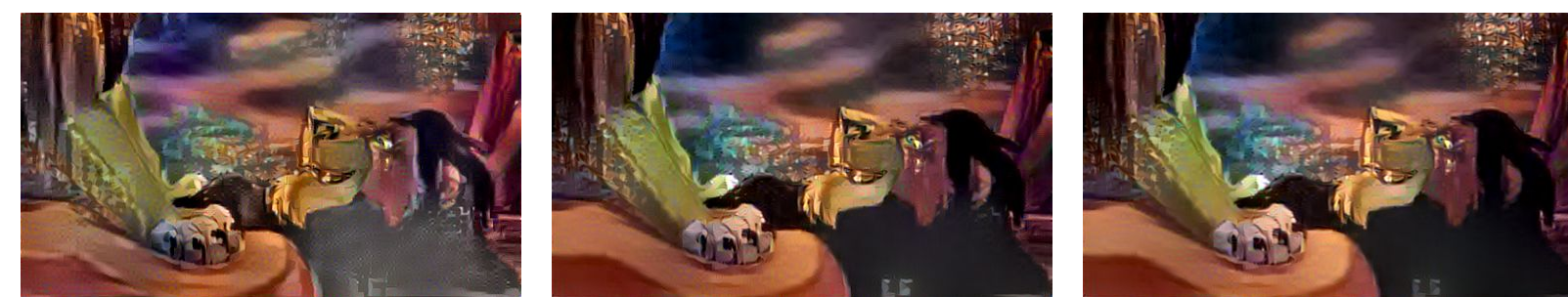
$$L(G, F, D_X, D_Y) = L_{GAN}(G, D_Y, X, Y) + L_{GAN}(F, D_X, Y, X) + \lambda L_{cyc}(G, F)$$



## Results

### Feedforward Neural Style Transfer

- Lion King/ Monster's Inc.*: Outputs after 300, 700, 1000 iterations (final output); final content loss: 565131.914



- Rick and Morty*: content 2D image(left), generated 3D image (middle), style 3D image(right), loss: 548,076; 1000 iterations



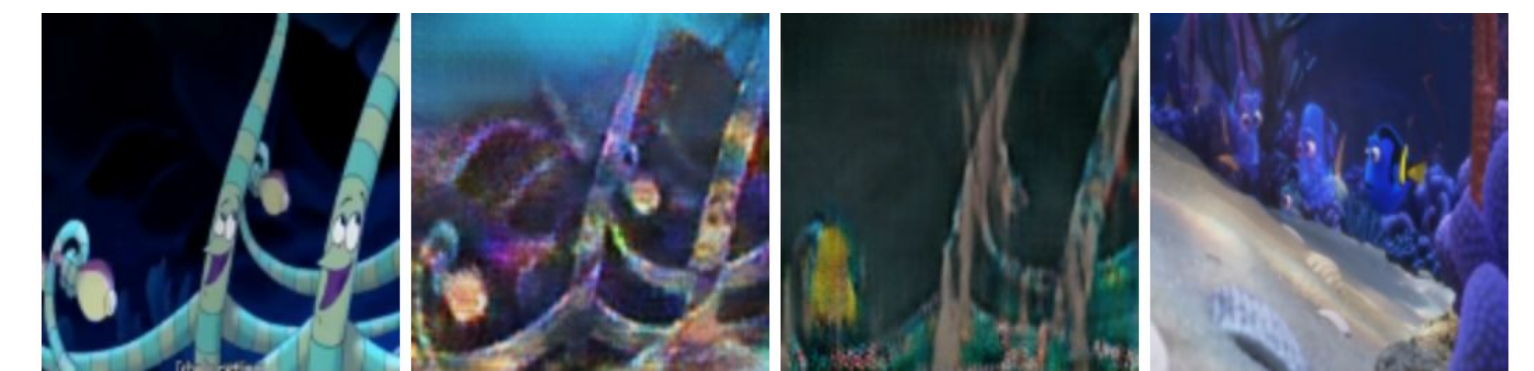
### Pix2Pix

- Losses after 200 epochs:
  - L1: 10.206, G\_GAN: 1.616, D\_real: 0.0688, D\_fake: 0.351



### CycleGAN

- left to right: content img, generated img, recovered img, content img



## Discussion

- Pix2Pix performed best at capturing style overall
- Feedforward does not perform well when style image is very different from the content image
  - Testing feedforward with style images from the same movie but in 3D, we found that feedforward was able to capture the 3D style very well
- Feedforward and CycleGAN have advantage over pix2pix in that it does not require paired data
- Feedforward has advantage over CycleGAN in that it does not require the two images to be from similar domains, although we did find that it performed better when they were

## Future Work

- Train with more data collected from more varied sources for pix2pix model, to learn more generalizable features
- Train with better aligned input and output images for pix2pix
- Paste style transferred frames together to test video quality
- Try data augmentation techniques on training images

## References

- Gatys, Leon A., Alexander S. Ecker, and Matthias Bethge. "Image style transfer using convolutional neural networks." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.
- Zhu, Jun-Yan, et al. "Unpaired image-to-image translation using cycle-consistent adversarial networks." *Proceedings of the IEEE international conference on computer vision*. 2017.
- Isola, Phillip, et al. "Image-to-image translation with conditional adversarial networks." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017.