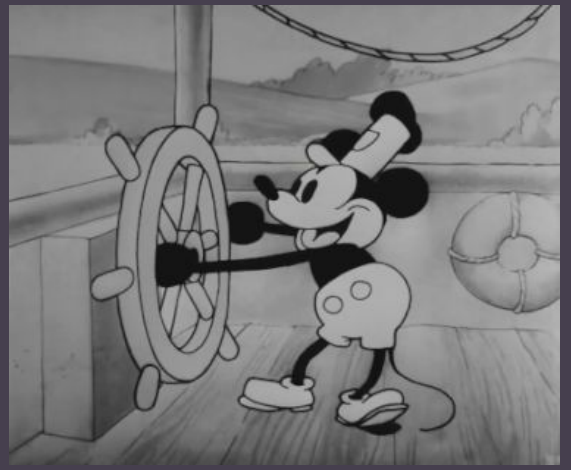
Applying Cartoony Style Transfer to Realistic Videos

Chloe Toda
Systems Lab, Dr. Yilmaz, Period 3
May 21, 2025

The Problem

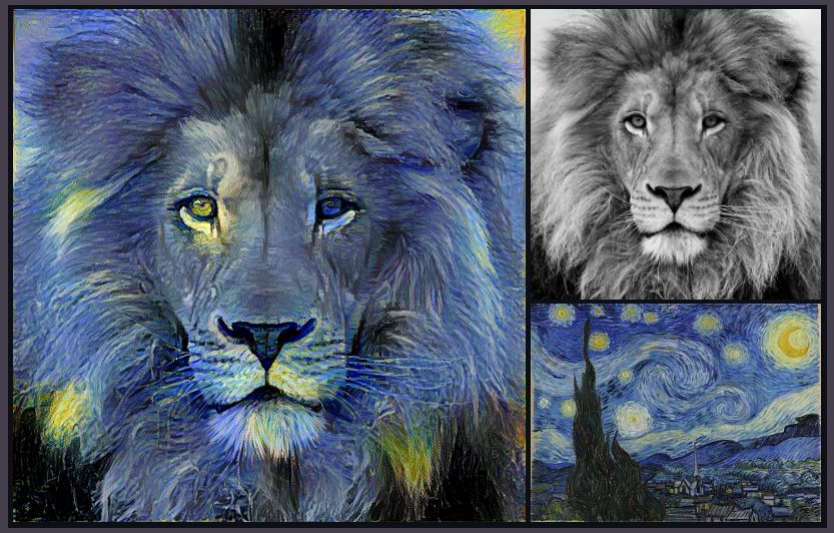
- **Cartoons are a popular form of media**
  - Take large amounts of time and effort to produce them
  - Over 20 frames per second



https://en.wikipedia.org/wiki/Walt_Disney_Animation_Studios#Production_logo

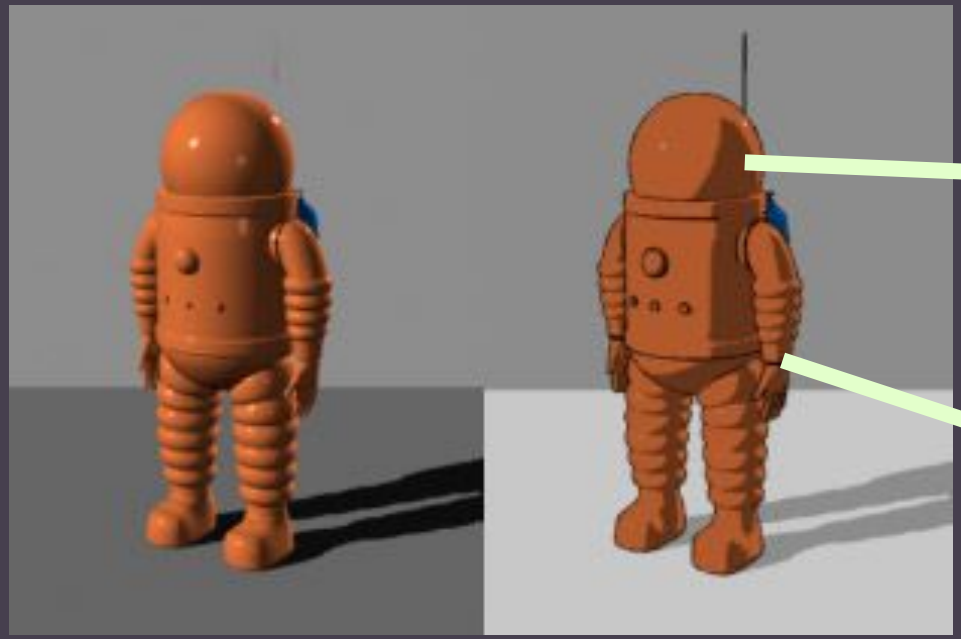
Background

- **What is Neural Style Transfer?**
  - Style and content images
- **CNNs and GANs**
  - Image recognition/ creation



https://github.com/cysmith/neural-style-tf?tab=readme-ov-file

Defining the Style



https://en.wikipedia.org/wiki/Non-photorealistic_rendering#/media/File:Toon_Shader.jpg

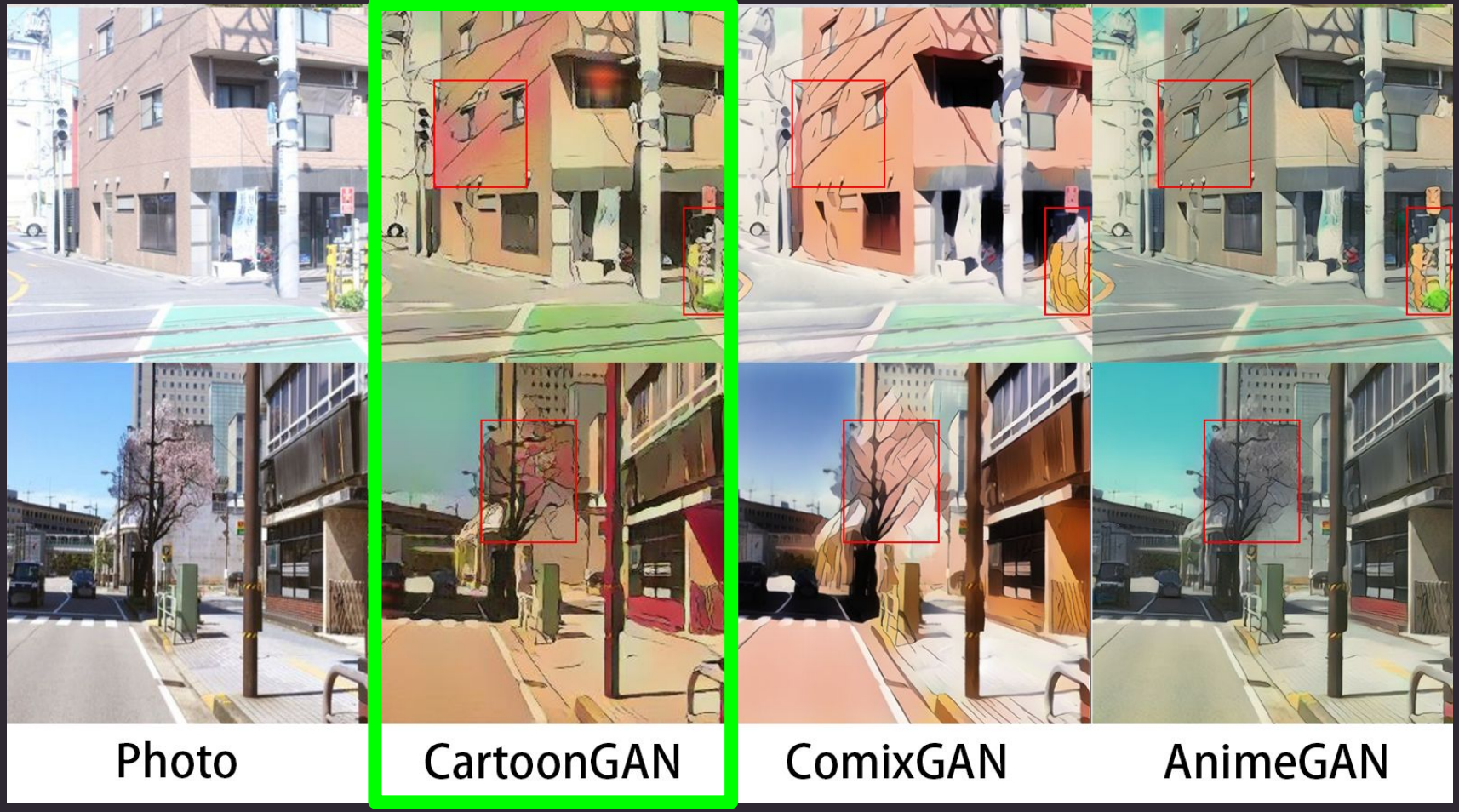
Similar Solutions

CartoonGAN [6]

- Anime style transfer
- Landscape input
- GAN

PortraitNET [7]

- Style transfer using non-anime cartoons
- People input
- Semantic network



https://github.com/TachibanaYoshino/AnimeGAN/blob/master/doc/Chen2020_Chapter_AnimeGAN.pdf

PortraitNET

https://www.sciencedirect.com/science/article/pii/S0925231221012856#f0060

Why is my solution better?

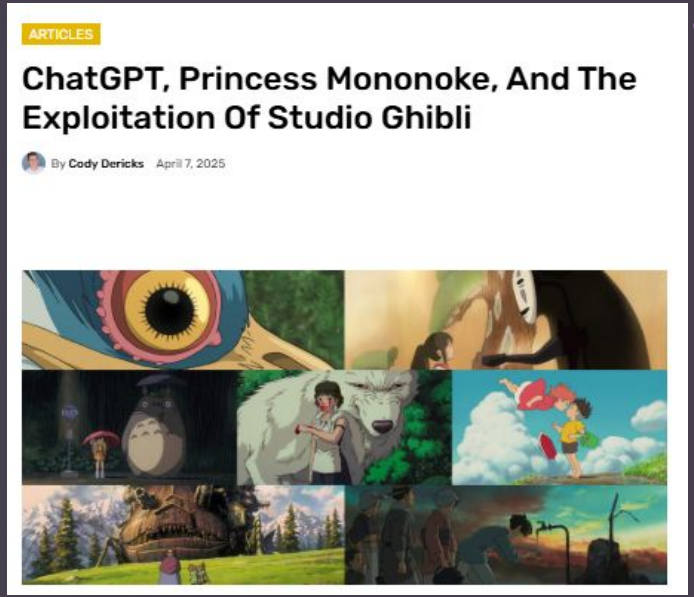
Video-based style transfer

Lack of cartoon style transfer besides using anime styles

Human input, not scenery input

Impact

- **Possible applications:**
  - Tool for animators
  - Other UGC
  - Better understanding of CV
- **Own personal enjoyment and understanding**
  - Style transfer is controversial



https://nextbestpicture.com/chatgpt-princess-mononoke-and-the-exploitation-of-studio-ghibli/#google_vignette

Training Data

https://mmlab.ie.cuhk.edu.hk/projects/CelebA.html

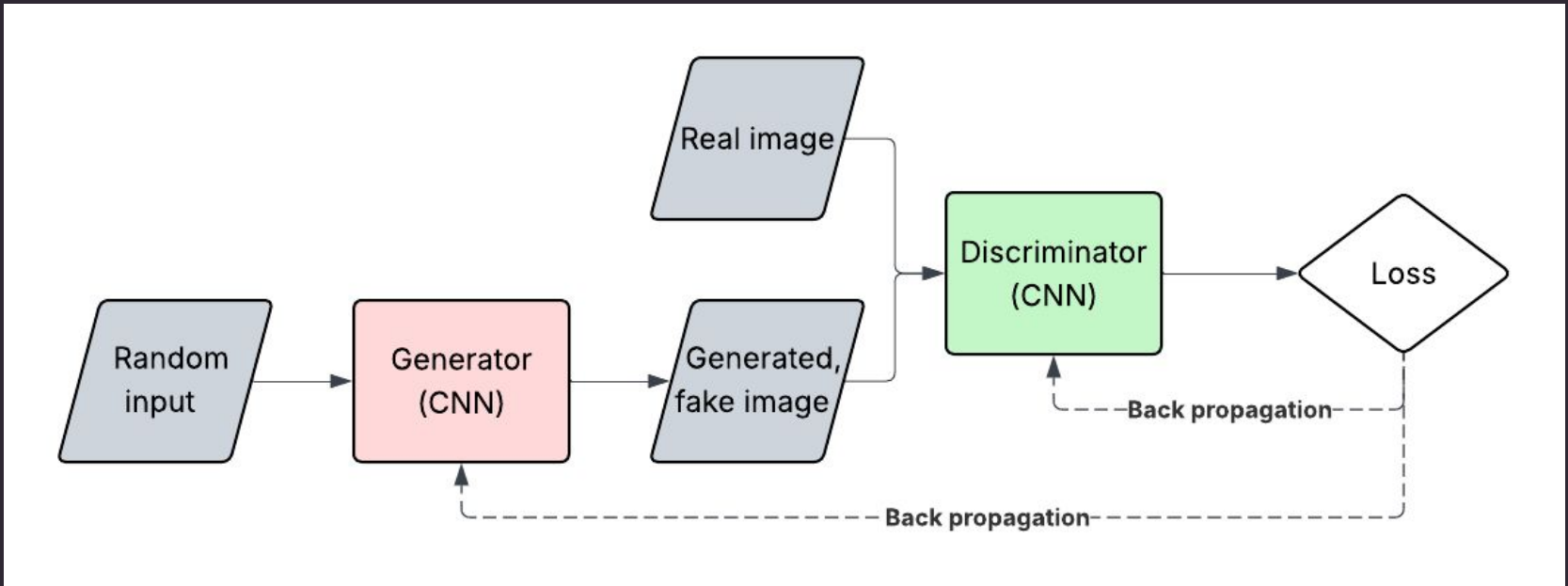
- Two parts: content and style
- Content training data
  - CelebA dataset [16]

Training Data

- Style images: Disney screencaps [11]
- 3,458 images

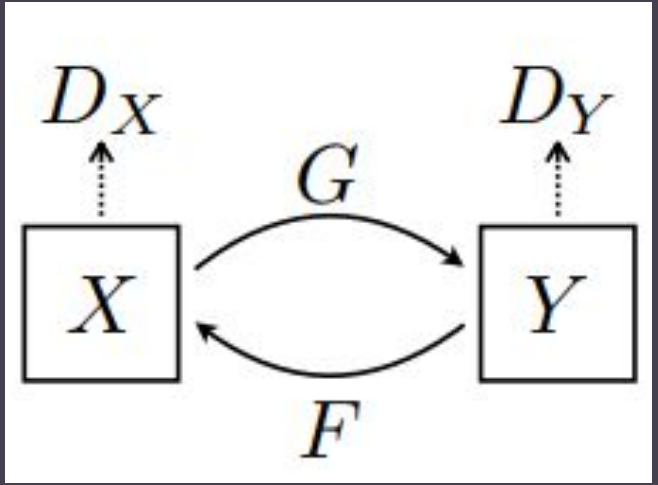
https://animationscreencaps.com/

Generative Adversarial Networks



Systems Architecture

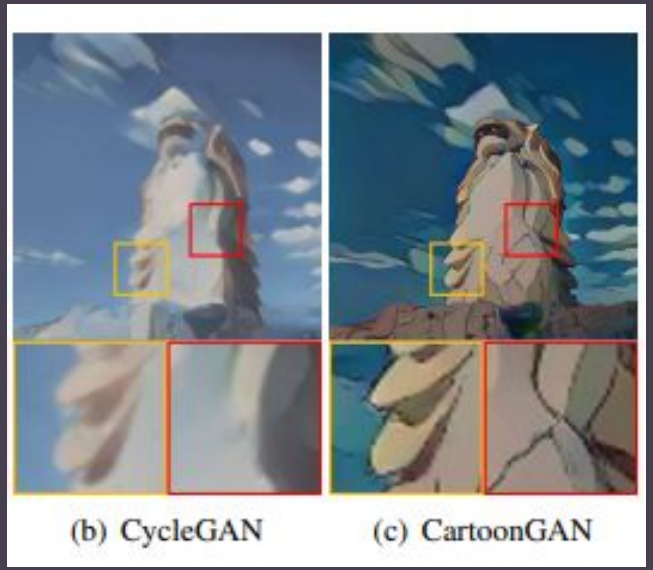
- CycleGAN [15]
- Image-to-image translation (style transfer)
- Two generators, two discriminators



https://arxiv.org/pdf/1703.10593

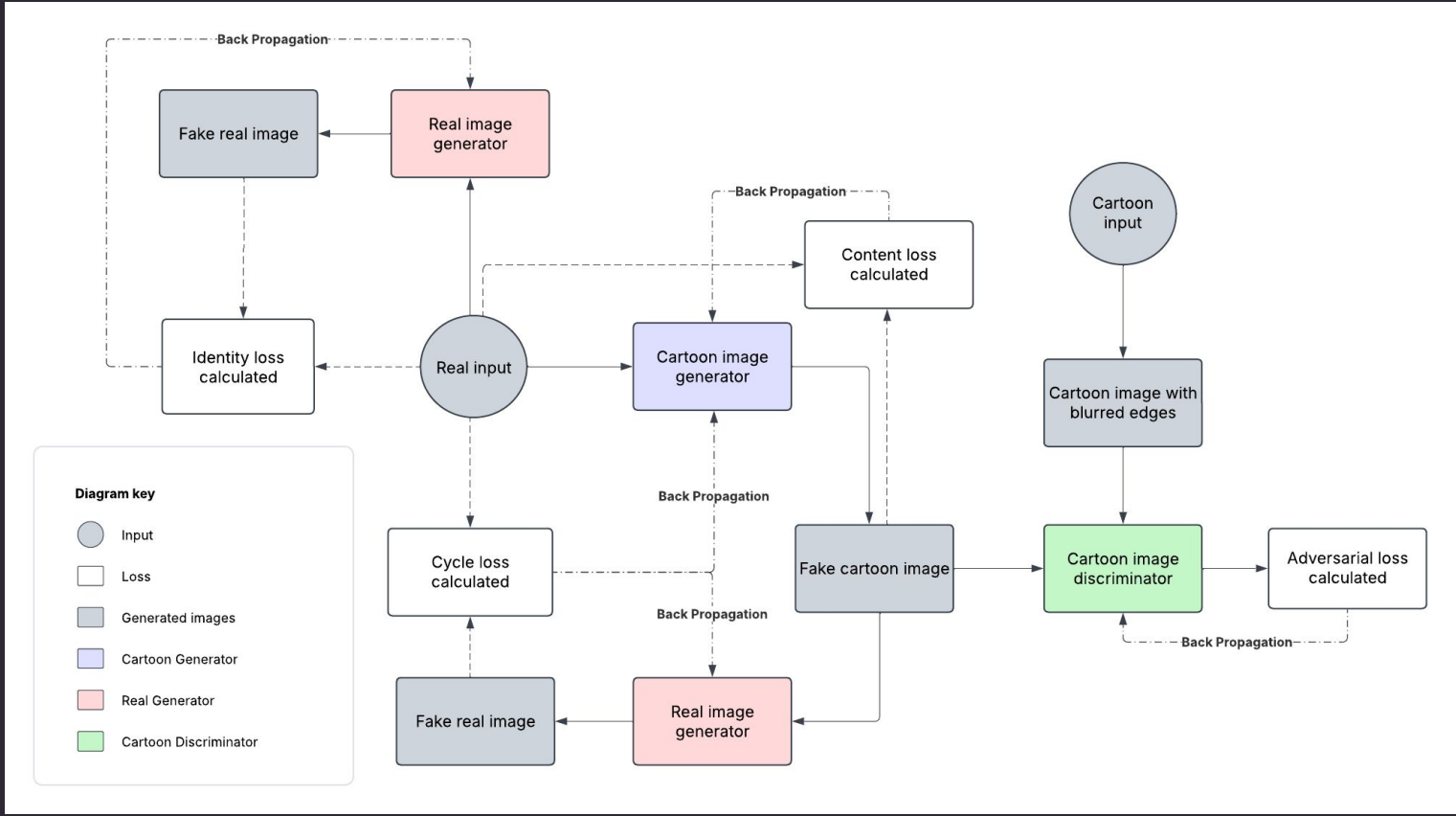
CartoonGAN [14]

- Inspired by CycleGAN [15]
- Maximizes cartoon features and content preservation
  - Initialization
  - Edge loss

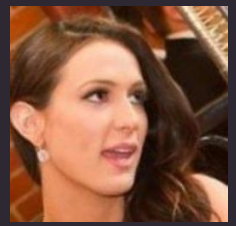


https://openaccess.thecvf.com/content_cvpr_2018/papers/Chen_CartoonGAN_Generative_Adversarial_CVPR_2018_paper.pdf

Updated Systems Architecture



Training Results



Epoch 28



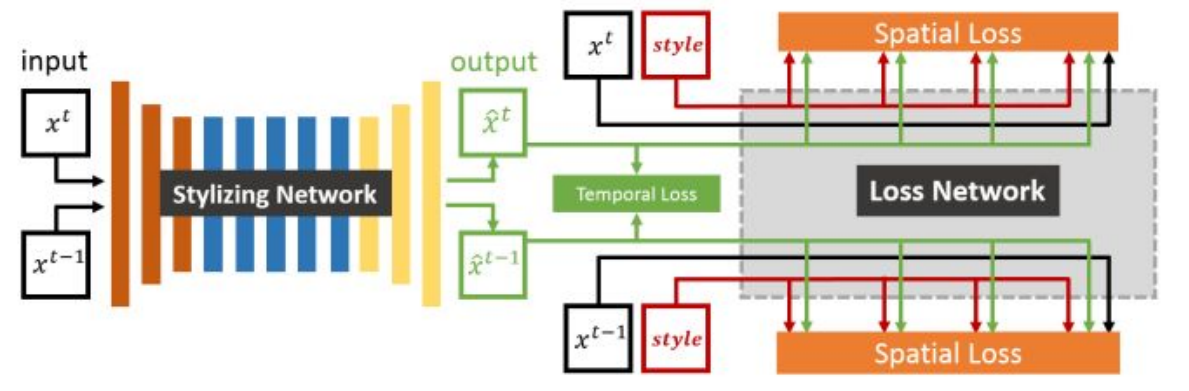
Epoch 35



Epoch 40

- ~10 epochs of initialization
- ~45 epochs of training with most successful model

Video Style Transfer [17]



https://openaccess.thecvf.com/content_cvpr_2017/papers/Huang_Real-Time_Neural_Style_CVPR_2017_paper.pdf

- Minimize artifacts in video frames
- Temporal loss: information between frames
- Optical flow: predict next frame

Temporal Loss Comparison

Demo

Limitations

- Large amount of resources required to train GANs
- JupyterLab servers overheating
- Not enough time to train data (1 epoch = 1 hour)

Conclusion and Future Work

Goal of this project:

- Convert realistic videos into cartoony ones
- Not the most successful with given limitations

Extensions following TJStar:

- Work with longer video clips
- Implement optical flow
- Apply different types of cartoon styles to videos

References

[1] Overview of GAN Structure - https://developers.google.com/machine-learning/gan/gan_structure

[2] Neural Style Transfer (NST) — theory and implementation - https://medium.com/@ferlatti.aldo/neural-style-transfer-nst-theory-and-implementation-c26728cf969d

[3] Rohan & Lenny #2: Convolutional Neural Networks - https://ayearofai.com/rohan-lenny-2-convolutional-neural-networks-5f4cd480a60b

[4] Neural Style Transfer Tutorial -Part 1 - https://docs.google.com/document/d/1_rO6fUkeTygqb067LyXZ7DKZ8Ar9kiTFa5Xt-e_eeDk/edit#heading=h.k74521sa5aoy

[5] Deep Visualization Toolbox - https://www.youtube.com/watch?v=AgkfIQ4IGaM&t=222s

[6] AnimeGAN: A Novel Lightweight GAN for Photo Animation - https://link.springer.com/chapter/10.1007/978-981-15-5577-0_18

References

[7] PortraitNET: Photo-realistic portrait cartoon style transfer with self-supervised semantic supervision - https://www.sciencedirect.com/science/article/abs/pii/S0925231221012856

[8] Convolutional Neural Networks (CNN) explained step by step - https://www.youtube.com/watch?v=sgL7RrqhGKI

[9] A Neural Algorithm of Artistic Style - https://arxiv.org/pdf/1508.06576

[10] Artistic style transfer for videos - https://www.youtube.com/watch?v=Khuj4ASIdmU

[11] Animation Screencaps - https://animationscreencaps.com/

[12] Neural Style Transfer Tutorial with Tensorflow and Python in 10 Minutes - https://www.youtube.com/watch?v=bFeltWvzZpQ&t=662s

[13] Human Faces (Object Detection) - https://www.kaggle.com/datasets/sbaghbidi/human-faces-object-detection/data?select=images

References

[14] CartoonGAN: Generative Adversarial Networks for Photo Cartoonization - https://openaccess.thecvf.com/content_cvpr_2018/papers/Chen_CartoonGAN_Generative_Adversarial_CVPR_2018_paper.pdf

[15] Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks - https://arxiv.org/pdf/1703.10593

[16] Large-scale CelebFaces Attributes (CelebA) Dataset - https://mmlab.ie.cuhk.edu.hk/projects/CelebA.html

[17] Real-Time Neural Style Transfer for Videos - https://openaccess.thecvf.com/content_cvpr_2017/papers/Huang_Real-Time_Neural_Style_CVPR_2017_paper.pdf

Thank you!

Any Questions?