

Image Animation

Problem Statement

The primary goal of this project is to use advanced deep learning techniques to bring static real-life photographs to life by animating them. By harnessing the power of cutting-edge algorithms such as Generative Adversarial Networks (GANs), Convolutional Neural Networks (CNNs), and image-to-image translation, our aim is to bridge the gap between reality and the captivating world of cartoons.

Our fascination lies in seamlessly transforming static images into dynamic and animated versions. Through the application of these state-of-the-art techniques, our objective is to create a harmonious fusion of the real world and the enchanting artistry found in cartoons.

To accomplish this, we will train deep learning models using extensive datasets comprising both cartoons and real-life photographs. By uncovering the distinct patterns and characteristics that differentiate cartoons from real-world images, we can develop a robust and versatile model capable of accurately capturing and enhancing the essential elements that define the unique cartoon aesthetic within any given photograph.

The potential applications of this project are vast and diverse. We envision the ability to transform personal photos into personalised cartoon avatars or create engaging animations for storytelling, advertising, and entertainment purposes. This endeavour opens up a realm of limitless possibilities, inviting boundless creativity and self-expression through animated pictures.

Dataset

Source of the Dataset:

"AnimeGAN: A Novel Lightweight GAN for Photo Animation" a conference paper by Jie Chen and Gang Liu have provided the dataset for open use to researchers and academics to pursue their own project. Dataset will be split into 3 parts:

- Training Set : Photos and anime data
- Validation set : Photos with different resolutions
- Testing set

Possible Approaches

1. **Generative Adversarial Networks (GANs):** Implement and train a GAN-based model to generate animated versions of real-life photographs. This approach involves training a generator network to transform input images into cartoon-like outputs, while simultaneously training a discriminator network to distinguish between real cartoons and generated images. The adversarial training process helps refine the generator's ability to produce more realistic and high-quality animated outputs.
2. **Image-to-Image Translation Models:** Explore image-to-image translation architectures such as Pix2Pix or CycleGAN, which are designed to learn the mapping between input and output images. Train the model using paired examples of real-life photographs and corresponding cartoon images, allowing the model to learn the mapping and generate cartoon-like versions of the input photographs.
3. **Style Transfer Techniques:** Investigate style transfer algorithms, such as Neural Style Transfer or Arbitrary Style Transfer, to transfer the visual style of cartoons onto real-life photographs. By leveraging pre-trained style transfer models or designing custom architectures, this approach aims to imbue the input photographs with the distinctive cartoon style while preserving the underlying content.

Experimentation Tracking Process

1. **Spreadsheet**
2. **NeptuneAI :** With Neptune AI, we can store all experimental data, including datasets, model architectures, hyperparameters, and evaluation metrics in one centralised location, making it easier for us to access and share experiment results with our team members.

Evaluation Metric

To assess the performance and effectiveness of the deep learning model in animating real-life photographs, the following evaluation metrics can be considered:

1. **Structural Similarity Index (SSIM):** SSIM measures the similarity between the generated animated image and the ground truth cartoon image. A higher SSIM value indicates better fidelity and resemblance to the cartoon style.
2. **Peak Signal-to-Noise Ratio (PSNR):** PSNR is a widely used metric to measure the quality of reconstructed images. It calculates the ratio between the

maximum possible pixel value and the mean squared error between the generated and ground truth images. A higher PSNR value indicates better image quality and fidelity.

3. **Frechet Inception Distance (FID):** FID is a perceptual metric that measures the similarity between the distribution of features extracted from the generated animated images and those from the ground truth cartoon images. A lower FID value signifies a closer match in terms of visual appearance and distribution.
4. **Human Subjective Assessment:** Conduct user studies or surveys where human observers evaluate the animated outputs based on visual quality, likeness to a cartoon, and overall satisfaction. Collecting subjective opinions and ratings from individuals can provide valuable insights into the perceptual appeal and authenticity of the generated animations.

Expected Outcome

1. Successful transformation of real-life photographs into animated versions with cartoon-like aesthetics.
2. Comparable or superior performance in quantitative evaluation metrics, demonstrating the effectiveness of the proposed deep learning approach.
3. Exploration of the limitations and challenges of the proposed method, providing insights for future improvements and research directions.
4. Potential for practical applications, such as personalised cartoon avatars, storytelling animations, advertising, and entertainment purposes.