

CV Final Project Proposal

Problem Definition

In the field of image transformation, there is a pressing need for models that can accurately convert images from one style to another, such as turning sketches into realistic photos or colorizing black-and-white images. Traditional methods often lack the ability to produce high-quality, realistic results due to limitations in capturing the complex relationships between input and target images. The challenge lies in developing a model that can learn these mappings effectively. Pix2pix, an image-to-image translation model based on Generative Adversarial Networks (GANs), addresses this by employing a generator to create transformed images and a discriminator to evaluate their realism. However, it requires paired datasets for supervised learning, which are not always readily available, limiting its applicability in some scenarios. The problem is to harness the capabilities of pix2pix to improve image transformation tasks while finding solutions to overcome the dependency on extensive paired datasets.

Motivation

Our idea is to apply pix2pix technology to creating sketches of suspects. Often, eyewitnesses don't have strong drawing skills, which makes it challenging to convey accurate information to law enforcement effectively. While most people can handle drawing outlines and facial proportions without much issue, the difficult part usually lies in detailing the facial features. Therefore, if we can use pix2pix to generate images based on rough outline sketches, we could achieve better results to aid eyewitnesses' recollection and assist detectives in identification.

Related work

1. <https://github.com/NVIDIA/pix2pixHD> (<https://github.com/NVIDIA/pix2pixHD>). Building upon pix2pix, NVIDIA introduced pix2pixHD (Wang et al., 2018), which addressed the challenge of generating high-resolution and photorealistic images. Pix2pixHD have introduced some important

innovation. First, by using generators and discriminators at multiple scales, the model are able to capture global structures and fine-grained details. Besides, it using Feature Matching loss function to stabilize the training procexx and improve the quality of the generated images. Last, it also introduce Perceptual Loss, enabling the generation of images with more realistic textures and details. By these effort, pix2pixHD can a higher resolution image than original pix2pix, and is far more photorealistic photo.

2. <https://github.com/junyanz/pytorch-CycleGAN-and-pix2pix>

(<https://github.com/junyanz/pytorch-CycleGAN-and-pix2pix>).

CycleGAN is a model based on pix2pix. They both use generators and discriminators, but they differ in data requirement and architecture.

CycleGAN extends original concept but works with unpaired datasets. It uses two sets of generators and discriminators to map images between two different domains. It also incorporates cycle consistency loss, ensuring that an image translated from one domain to another can be returned to its original form.