

Tutorial 4: Image-to-Image Translation

Image-to-image translation refers to the problem of generating a new image based on an input image. Many tasks in image processing can be formulated as an image-to-image translation problem, including image super-resolution, image inpainting, and style transfer. In this short course, we will cover both of the basics and applications of image-to-image translation.

We will start the course by reviewing the generative adversarial network (GAN) framework proposed by Goodfellow et. al., which is a popular generative model and is the backbone model of various state-of-the-art image-to-image translation methods thanks to its extraordinary capability in generating crispy sharp images. Various variants of GANs exist. We will cover several popular ones. We will also cover their conditional extensions.

Next, we will give a formal definition of the image-to-image translation problem. We will unify the notation and categorize existing works based on their learning settings, including the supervised setting (input-output relationship is observed), the unsupervised setting (input-output relationship is not observed), the semi-supervised setting, the multimodal setting, and the few-shot setting. We will discuss details of the representative works in each setting. We will cover their network designs, objective functions, training strategy, and limitations. Applications to various image processing tasks will be discussed.

We will then move to the video-to-video translation problem, which is a natural extension of the image-to-image translation problem. We will discuss techniques for generating convincing visual dynamics as well as techniques for ensuring temporal consistency. We will then present an integration of an existing 3D rendering engine and a video-to-video translation model for creating a new form of computer graphics. We will talk about its strength and weakness as a graphic rendering engine.

Finally, we will conclude the course by discussing the conditions needed for image-to-image translation to work. We will talk about practical methods for meeting the conditions, including how to collect training data and troubleshooting tips. We will outline the remaining challenges and potential research problems. More information about the tutorial will be available in <https://github.com/mingyuliutw/icip2019-image-translation-tutorial>

Speakers:



Ming-Yu Liu
NVIDIA



Ting-Chun Wang
NVIDIA, USA