VFX Final - X2Face

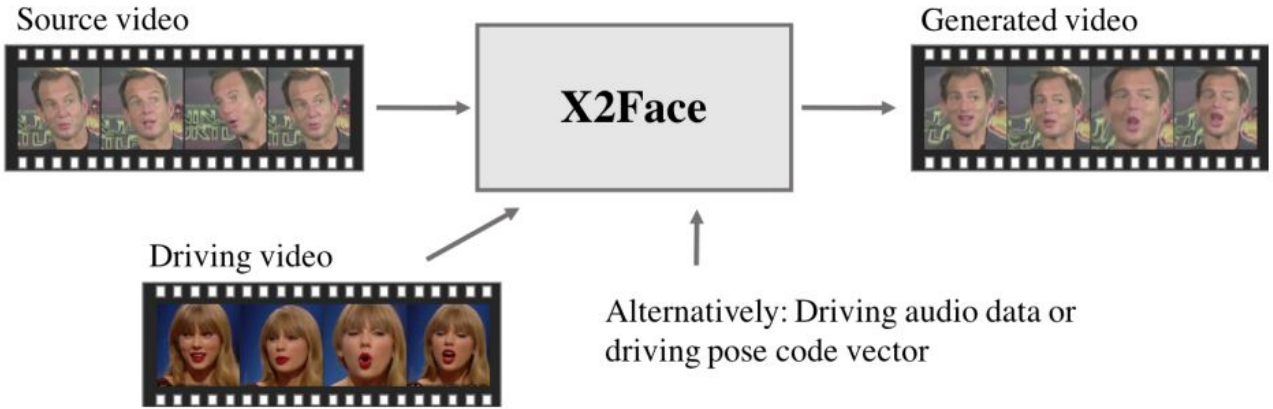
Team 4
R11922025 陸信丞 R11922036 陳靖元
R11944005 薛博文 R11944024 黃秉茂

Outline

- Introduction
- Methodology
- Implementation
- System
- Research
- Demo
- Reference

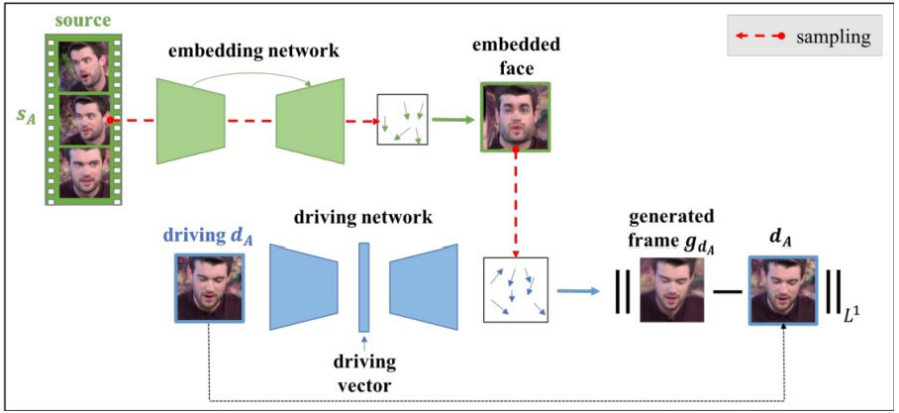
Introduction

- Image editing
- Source - Identity and Hairstyle
- Driving - Properties, Pose, and Expression



Methodology

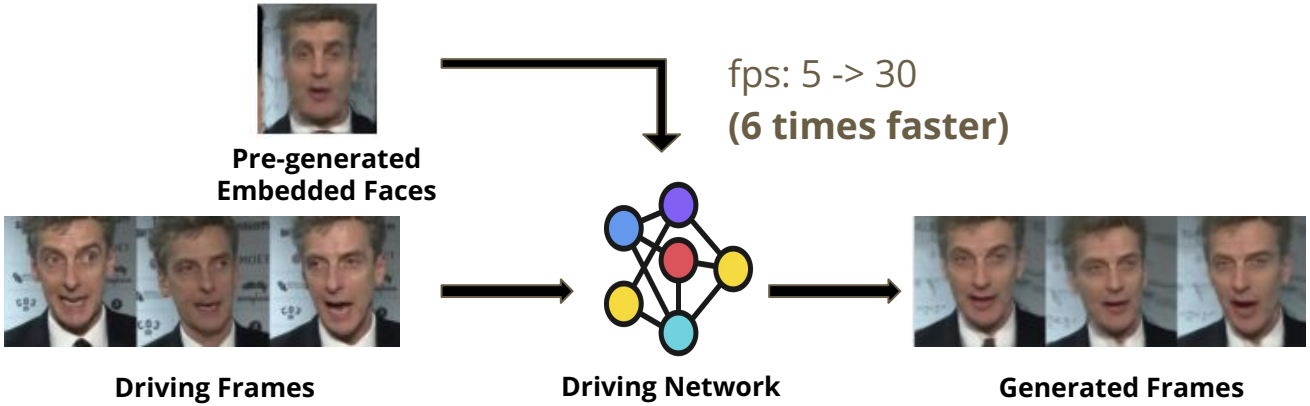
- SSL
- Embedding network
  - source frame to a face representation
  - U-Net and pix2pix
- Driving network
  - transform pixels from the embedded face to produce the generated frame
  - encoder-decoder



Implementation

- Embeded faces generation

- Faster inference with pre-generated embedded faces (only need to inference driving network)



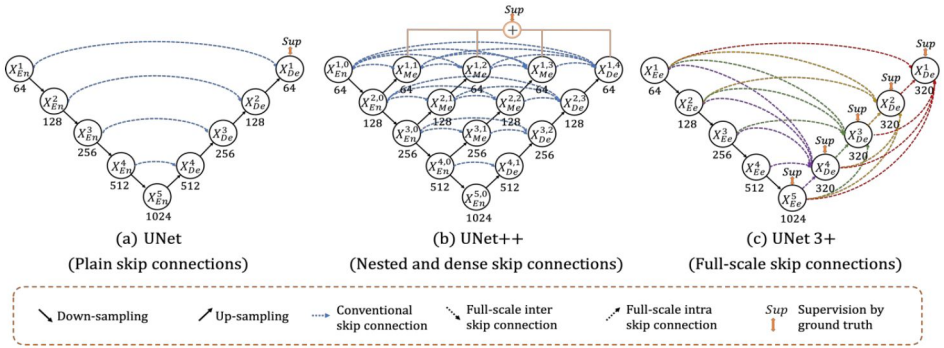
System and Library

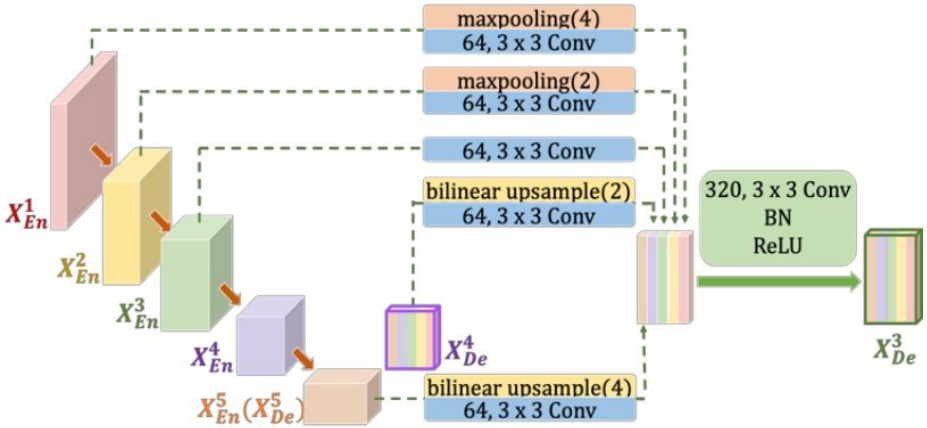
- Training
- Inference
- Real-time
- Video to Images
- Images to Video

Research

- Model Backbone
  - Unet
  - Unet 3+
    - Full-scale Skip Connections
    - Full-scale Deep Supervision
    - Reduce the network parameter
    - More full-scale information
- Loss
  - L1 loss
  - SSIM loss

$$\ell_{ms-ssim} = 1 - \prod_{m=1}^{M} \left(\frac{2\mu_p\mu_g + C_1}{\mu_p^2 + \mu_g^2 + C_1} \right)^{\beta_m} \left(\frac{2\sigma_{pg} + C_2}{\sigma_p^2 + \sigma_g^2 + C_2} \right)^{\gamma_i}$$





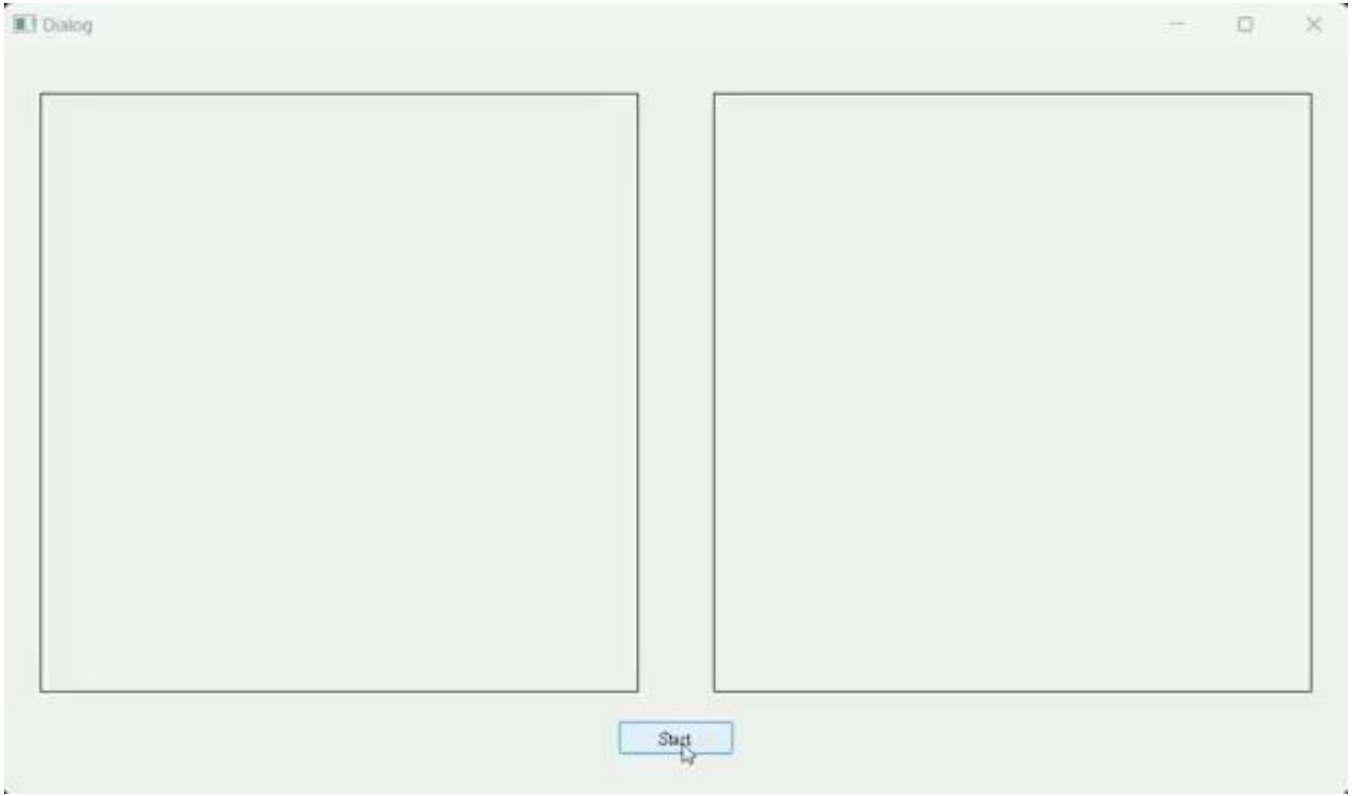
Demo - Offline

Source

Driver

Result

Demo - Online



Reference

- X2Face - WILES, Olivia; KOEPKE, A.; ZISSERMAN, Andrew. X2face: A network for controlling face generation using images, audio, and pose codes. In: Proceedings of the European conference on computer vision (ECCV). 2018. p. 670-686.
- Unet 3+ - HUANG, Huimin, et al. Unet 3+: A full-scale connected unet for medical image segmentation. In: ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2020. p. 1055-1059.