2019-2 A.I. & Security

Video-to-Video Synthesis

NVIDIA Corp. with MIT

KIM JEONG HYUN

Abstract: Video-to-video

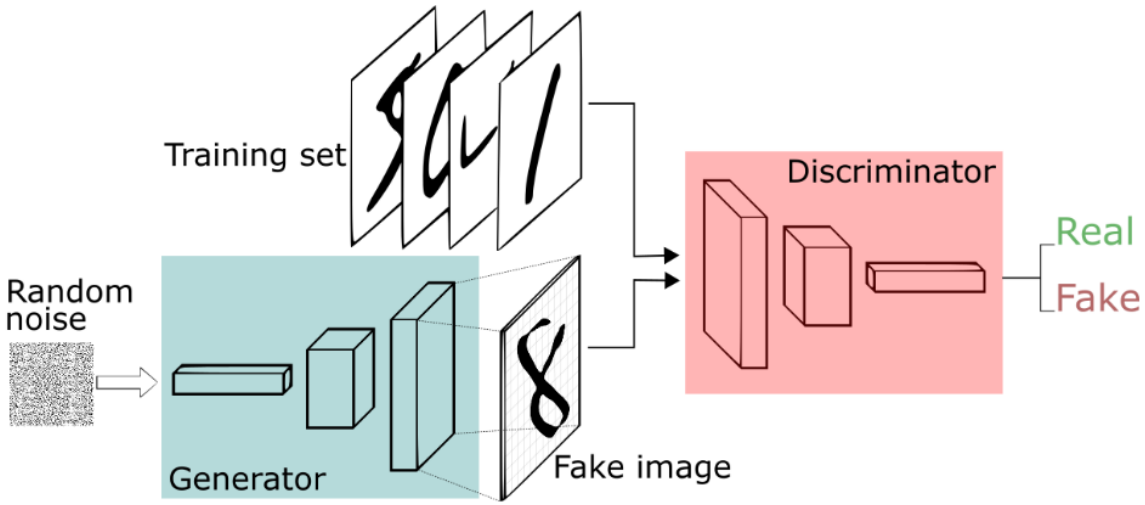
- 입력 영상(Input source video)을 통해 새로운 영상을 매핑
- Image-to-image 분야에 비해 많이 연구되지 않았음
- Generative Adversarial learning framework 방법 이용 (GANs)
- 기존 Image-to-image 기법을 통해 생성한 이미지 프레임을 연속적으로 나열한 것에 비해 월등히 자연스럽고 생생함

Introduction Video

https://youtu.be/5zlcXTCpQqM

GANs (Generative Adversarial Networks)

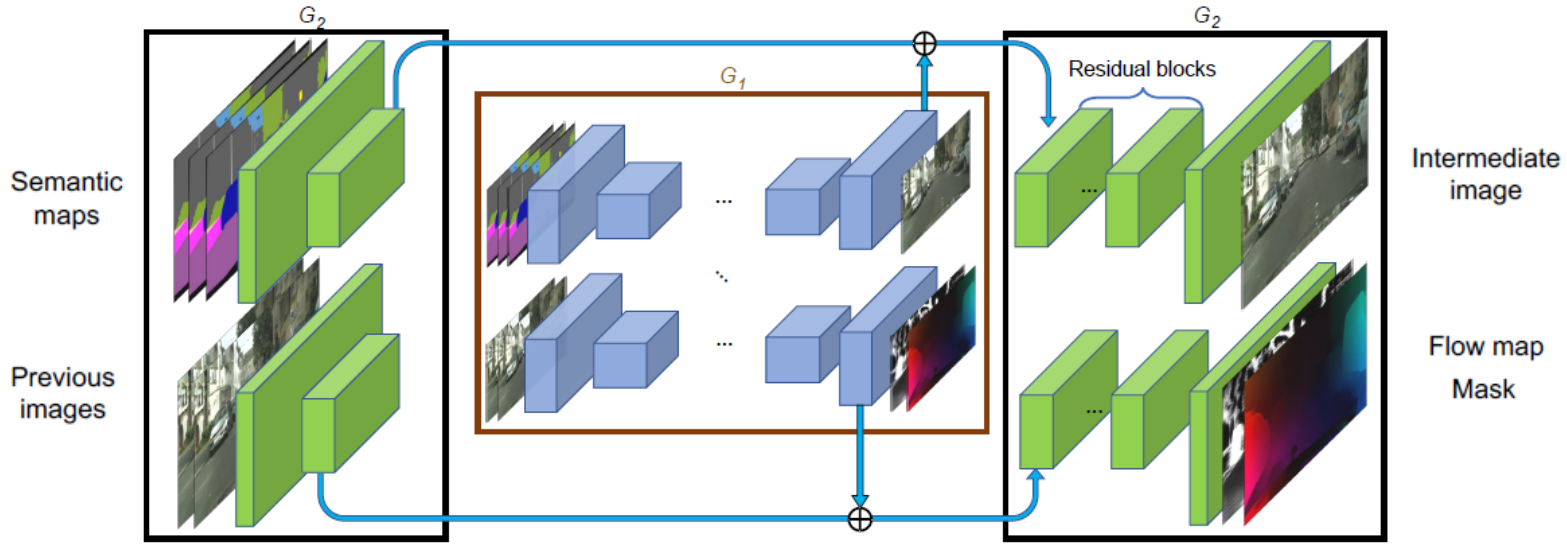
- GANs는 두 개의 신경망으로 구성 1)generator 2)discriminator
- *Generator*
  - 새로운 데이터 인스턴스 생성
- *Discriminator*
  - G가 생성한 데이터의 진위 평가(0: 가짜 / 1: 진짜)
- **핵심 개념 – 이중 피드백**
  - G는 생성한 가짜 이미지가 진짜처럼 보이기를 원하고
  - D는 전달된 이미지를 가짜로 식별하는 목표를 가짐



GANs in Vid2vid

$$\max_D \min_G E_{(\mathbf{x}_1^T, \mathbf{s}_1^T)}[\log D(\mathbf{x}_1^T, \mathbf{s}_1^T)] + E_{\mathbf{s}_1^T}[\log(1 - D(G(\mathbf{s}_1^T), \mathbf{s}_1^T))],$$

- Generator



$$p(\tilde{\mathbf{x}}_1^T|\mathbf{s}_1^T) = \prod_{t=1}^{T} p(\tilde{\mathbf{x}}_t|\tilde{\mathbf{x}}_{t-L}^{t-1}, \mathbf{s}_{t-L}^t). \longrightarrow F(\tilde{\mathbf{x}}_{t-L}^{t-1}, \mathbf{s}_{t-L}^t) = (\mathbf{1} - \tilde{\mathbf{m}}_t) \odot \tilde{\mathbf{w}}_{t-1}(\tilde{\mathbf{x}}_{t-1}) + \tilde{\mathbf{m}}_t \odot \tilde{\mathbf{h}}_t,$$

$$\tilde{\mathbf{w}}_{t-1} = W(\tilde{\mathbf{x}}_{t-L}^{t-1}, \mathbf{s}_{t-L}^t)$$
$$\tilde{\mathbf{h}}_t = H(\tilde{\mathbf{x}}_{t-L}^{t-1}, \mathbf{s}_{t-L}^t)$$
$$\tilde{\mathbf{m}}_t = M(\tilde{\mathbf{x}}_{t-L}^{t-1}, \mathbf{s}_{t-L}^t)$$

Results

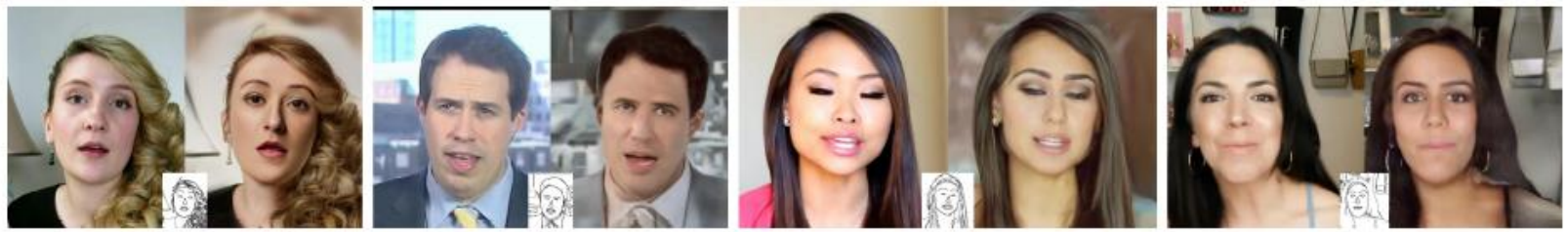


Figure 5: Example face→sketch→face results. Each set shows the original video, the extracted edges, and our synthesized video. *The figure is best viewed with Acrobat Reader. Click the image to play the video clip.*

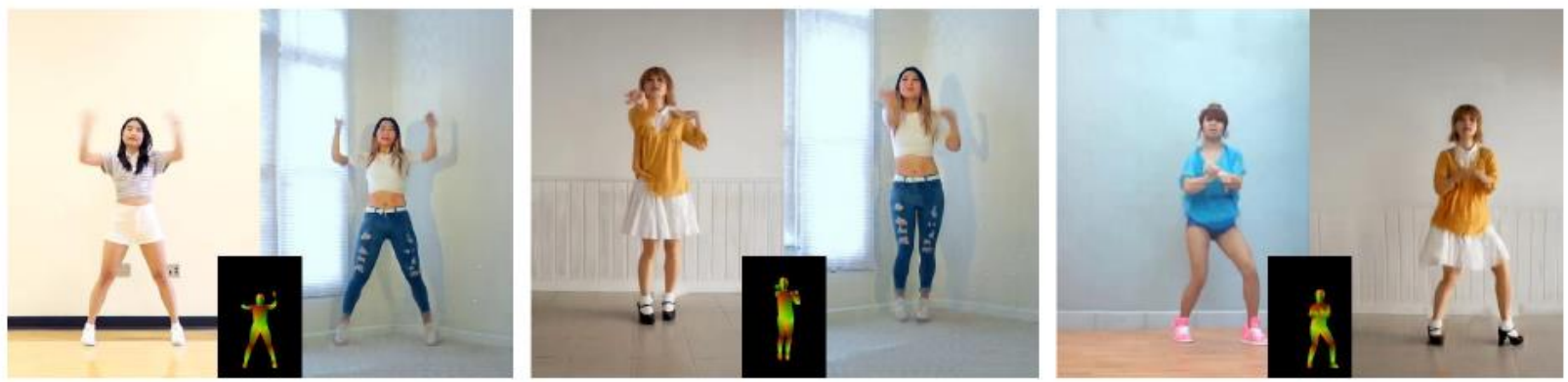


Figure 6: Example dance→pose→dance results. Each set shows the original dancer, the extracted poses, and the synthesized video. *The figure is best viewed with Acrobat Reader. Click the image to play the video clip.*