SSL | Smart Systems Lab.
YONSEI UNIVERSITY, SEOUL, KOREA

Edge Connect: Generative Image Inpainting with Adversarial Edge Learning

Kamyar Nazeri, Eric Ng, Tony Joseph, Faisal Z. Qureshi, Mehran Ebrahimi (2019)

석사과정 3학기 박희성

목차

Smart Systems Lab.
YONSEI UNIVERSITY, SEOUL, KOREA

목차

1. Edge Connect Introduction

- **Image Inpainting**
  - Image Inpainting, Image Completion은 이미지의 부족한 부분을 채우는 Task

- **Traditional Approaches**
  - Diffusion-based Methods : 주변 pixel value를 missing regions에 propagation 시키는 방식
  - Path-based Methods : 가장 유사한 부분의 path를 missing regions에 붙이는 방식

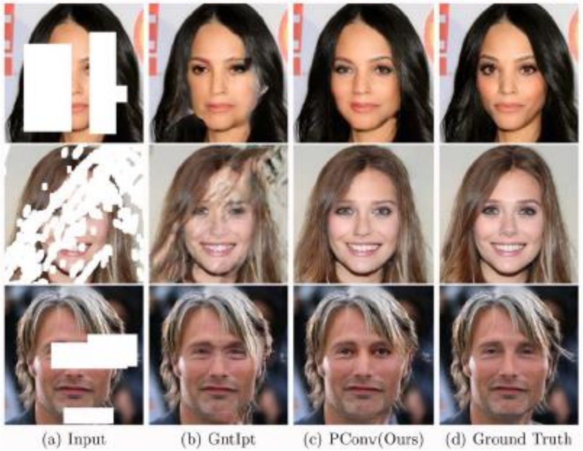
    > Over-smoothed and/or blurry 문제가 공통적으로 발생함.

(a) Ground Truth (b) Masked Image (c) Yu *et al.* [53] (d) Iizuka *et al.* [22] (e) Ours (f) Ours (Canny)

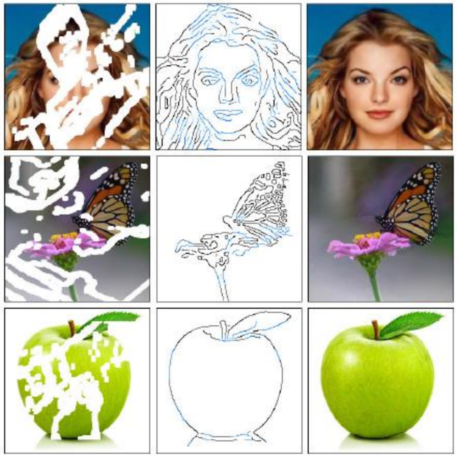
1. Edge Connect Introduction

SSL Smart Systems Lab. YONSEI UNIVERSITY, SEOUL, KOREA

- **Proposing Method**
  - Sketches의 중요성을 강조하며 ‘lines first, color next’ 접근법을 제안
  - ‘**Edge Generator**’ and ‘**Image Completion**’으로 나뉘는 two stage process로 구성되어 있음



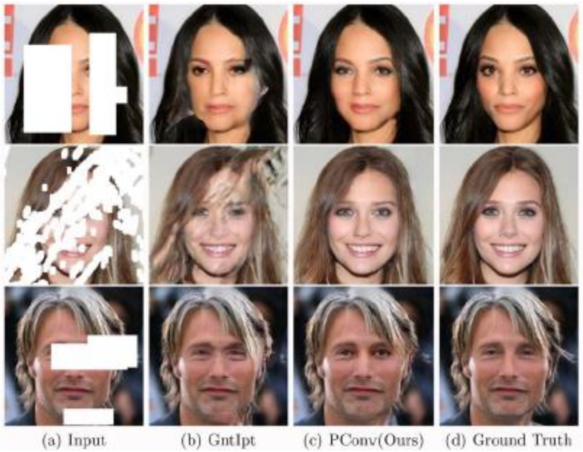
<기존 방법론>



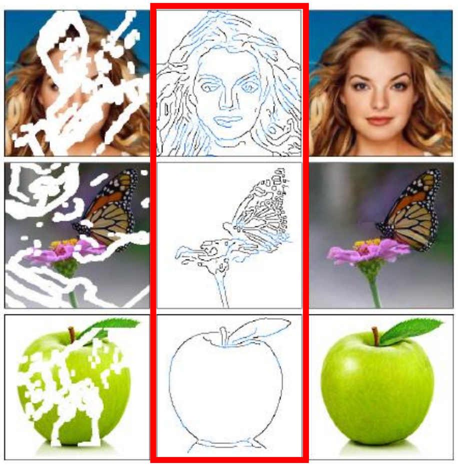
<제안 방법론>

1. Edge Connect Introduction

- **Proposing Method**
  - Sketches의 중요성을 강조하며 'lines first, color next' 접근법을 제안
  - '**Edge Generator**' and '**Image Completion**'으로 나뉘는 two stage process로 구성되어 있음



<기존 방법론>



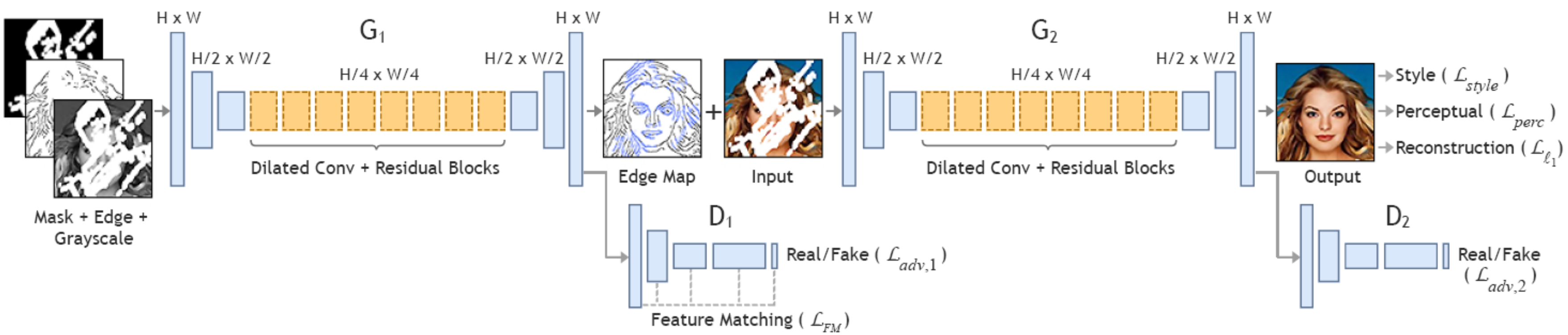
<제안 방법론>

1. Edge Connect Introduction

- Proposing Method Overview

- Sketches의 중요성을 강조하며 'lines first, color next' 접근법을 제안
- '**Edge Generator**' and '**Image Completion**'으로 나뉘는 two stage process로 구성되어 있음

<Edge Connect Model Architecture>

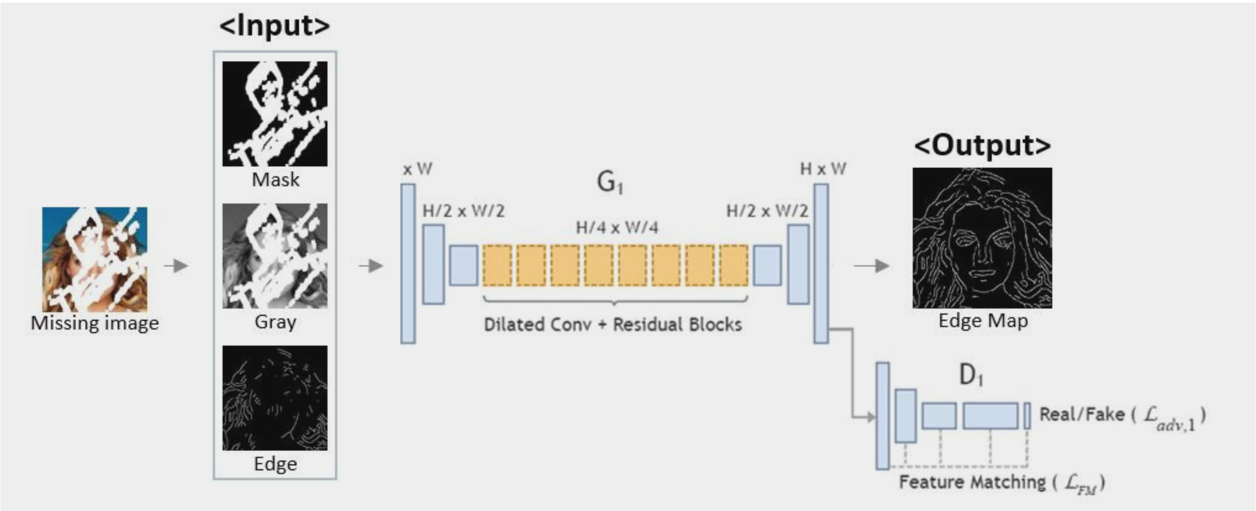


Edge Generator (G_1, D_1)

Image Completion Network (G_2, D_2)

2. Model Structure

< Edge Generator (G_1, D_1) >

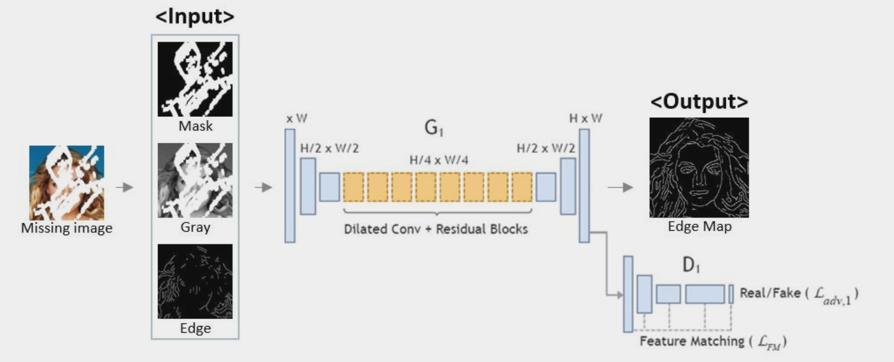


- Input & Output
  - ✓ $C_{pred} = G_1(M, \tilde{I}_{gray}, \tilde{C}_{gt})$
- Notation
  - ✓ Ground truth image, edge map, grayscale : I_{gt}, C_{gt}, I_{gray}
  - ✓ Mask : **M** (1 for the missing region, 0 for background)
  - ✓ Masked grayscale image : $\tilde{I}_{gray} = I_{gray} \odot (1 - M)$
  - ✓ Masked edge map : $\tilde{C}_{gt} = C_{gt} \odot (1 - M)$

  - 틸드가 붙은 Notation은 지워지지 않은 영역을 나타낸 것(Masked)

2. Model Structure

< Edge Generator (G_1, D_1) >



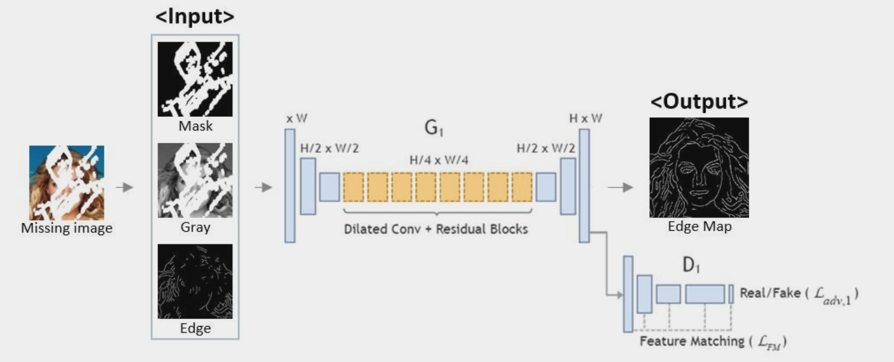
- Objective Function

$$\min_{G_1} \max_{D_1} \mathcal{L}_{G_1} = \min_{G_1} \left(\lambda_{adv,1} \max_{D_1} (\mathcal{L}_{adv,1}) + \lambda_{FM} \mathcal{L}_{FM} \right)$$

2. Model Structure

Smart Systems Lab.
YONSEI UNIVERSITY, SEOUL, KOREA

< Edge Generator (G_1, D_1) >



- Objective Function

$$\min_{G_1}\max_{D_1}\mathcal{L}_{G_1} = \min_{G_1}\left(\boxed{\lambda_{adv,1}\max_{D_1}(\mathcal{L}_{adv,1})} + \lambda_{FM}\mathcal{L}_{FM}\right)$$

➢ Adversarial Loss

$$\mathcal{L}_{adv,1} = \mathbb{E}_{(\mathbf{C}_{gt},\mathbf{I}_{gray})}\left[\log D_1(\mathbf{C}_{gt},\mathbf{I}_{gray})\right] + \mathbb{E}_{\mathbf{I}_{gray}}\log\left[1 - D_1(\mathbf{C}_{pred},\mathbf{I}_{gray})\right]$$

최대화↑

최대화↑

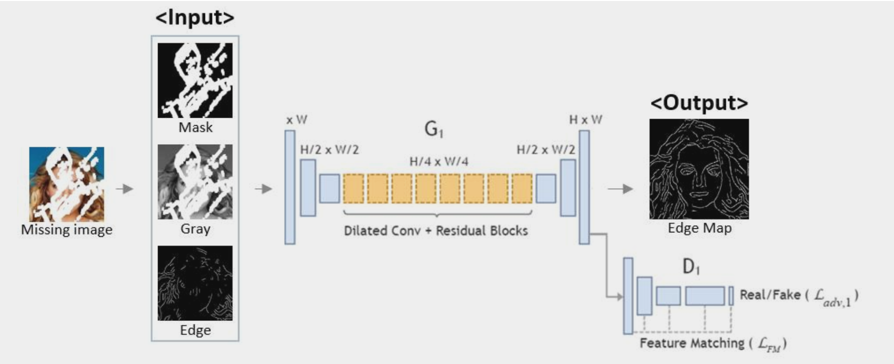
최소화↓

1에 가까워져야 하므로
GT는 real로 분류하도록

0에 가까워져야 하므로
Pred는 fake로 분류하도록

2. Model Structure

< Edge Generator (G_1, D_1) >



- Objective Function

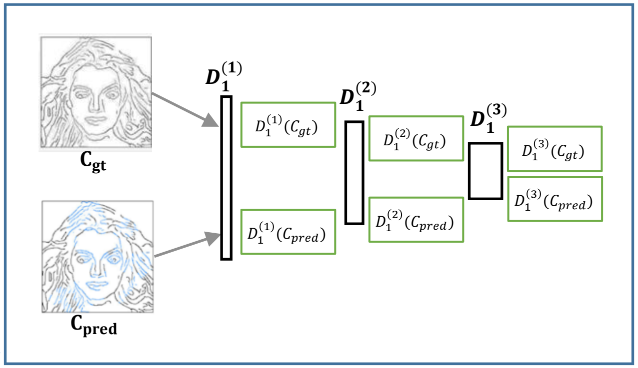
$$\min_{G_1}\max_{D_1}\mathcal{L}_{G_1} = \min_{G_1}\left(\lambda_{adv,1}\max_{D_1}(\mathcal{L}_{adv,1}) + \boxed{\lambda_{FM}\mathcal{L}_{FM}}\right)$$

➢ Feature-Matching Loss

$$\mathcal{L}_{FM} = \mathbb{E}\left[\sum_{i=1}^{L}\frac{1}{N_i}\underbrace{\left\|D_1^{(i)}(\mathbf{C}_{gt}) - D_1^{(i)}(\mathbf{C}_{pred})\right\|_1}_{\text{GT와 Pred가 유사해지도록}}\right]$$

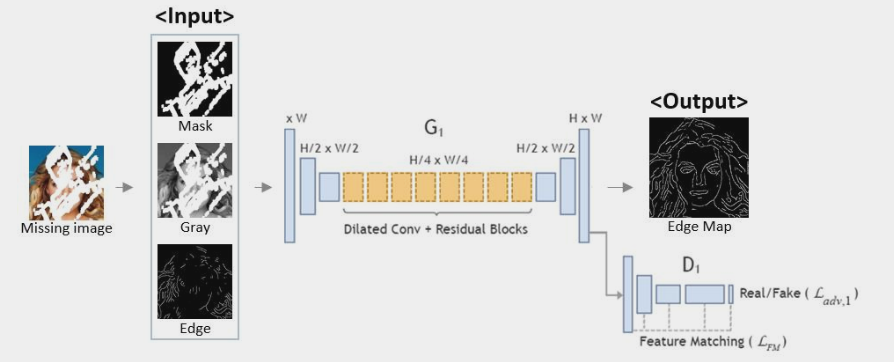
최소화↓

Discriminator의 Layer를 거친 Activation Map들을 각각 비교



2. Model Structure

< Edge Generator (G_1, D_1) >



- Objective Function

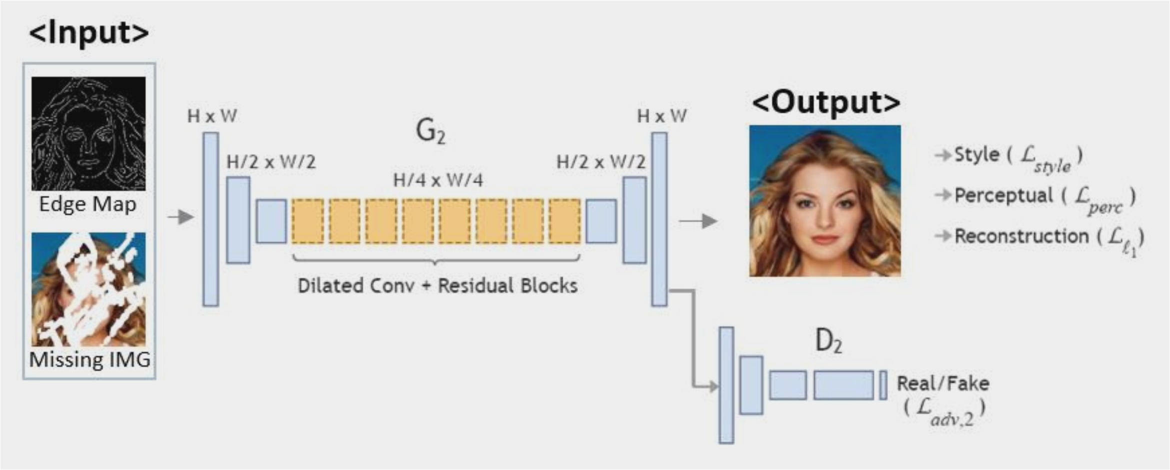
$$\min_{G_1} \max_{D_1} \mathcal{L}_{G_1} = \min_{G_1} \left(\lambda_{adv,1} \max_{D_1} (\mathcal{L}_{adv,1}) + \lambda_{FM} \mathcal{L}_{FM} \right)$$

✓ Regularization Parameter Setting

$\lambda_{adv,1}$ = 1

λ_{FM} = 10

< Image Completion Network (G_2, D_2) >



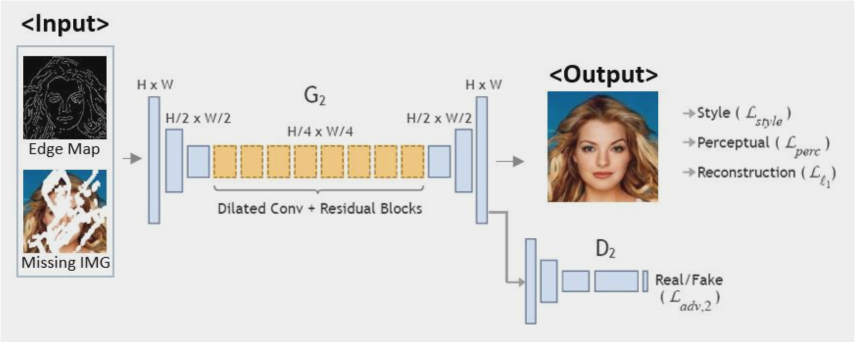
- Input & Output
  - ✓ $\mathbf{I_{pred}} = \mathbf{G_2}(\mathbf{\tilde{I}_{gt}}, \mathbf{C_{comp}})$

- Notation
  - ✓ **Comp**osite edge map : $\mathbf{C_{comp}} = \mathbf{C_{gt}} \odot (\mathbf{1} - \mathbf{M}) + \mathbf{C_{pred}} \odot \mathbf{M}$

2. Model Structure

SSL Smart Systems Lab. YONSEI UNIVERSITY, SEOUL, KOREA

< Image Completion Network (G_2, D_2) >



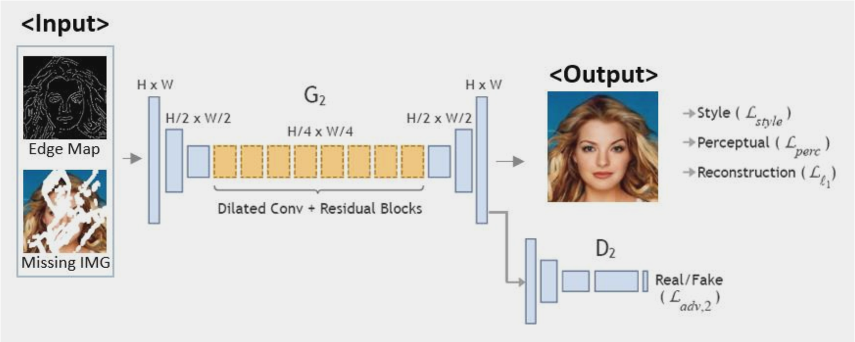
- Loss Function

$$\mathcal{L}_{G_2} = \lambda_{\ell_1}\mathcal{L}_{\ell_1} + \lambda_{adv,2}\mathcal{L}_{adv,2} + \lambda_p\mathcal{L}_{perc} + \lambda_s\mathcal{L}_{style}$$

2. Model Structure

Smart Systems Lab. YONSEI UNIVERSITY, SEOUL, KOREA

< Image Completion Network (G_2, D_2) >



- Loss Function

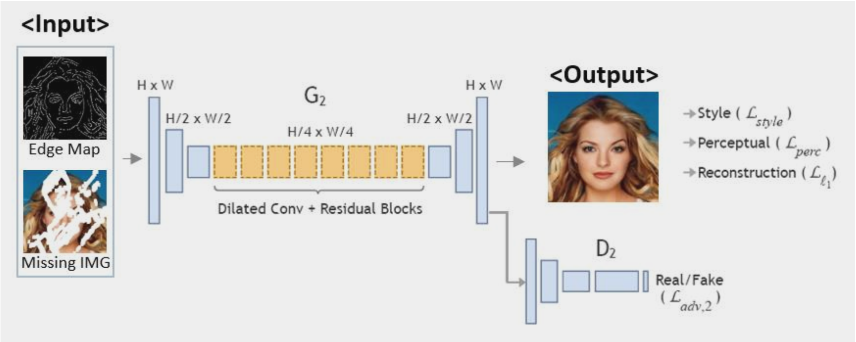
$$\mathcal{L}_{G_2} = \boxed{\lambda_{\ell_1}\mathcal{L}_{\ell_1}} + \lambda_{adv,2}\mathcal{L}_{adv,2} + \lambda_p\mathcal{L}_{perc} + \lambda_s\mathcal{L}_{style}$$

➢ L1 Loss

$$\mathcal{L}_{\ell_1} = \mathbb{E}\left(\left|\boldsymbol{I}_{\mathbf{gt}} - \boldsymbol{I}_{\boldsymbol{pred}}\right|\right)$$

2. Model Structure

< Image Completion Network (G_2, D_2) >



- Loss Function

$$\mathcal{L}_{G_2} = \lambda_{\ell_1}\mathcal{L}_{\ell_1} + \boxed{\lambda_{adv,2}\mathcal{L}_{adv,2}} + \lambda_p\mathcal{L}_{perc} + \lambda_s\mathcal{L}_{style}$$

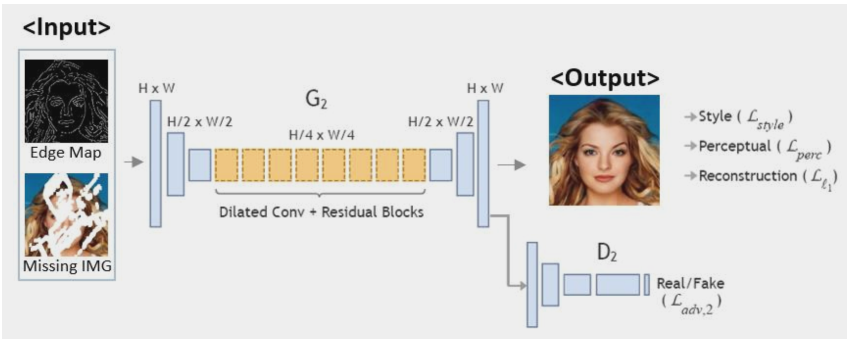
➢ Adversarial Loss

$$\mathcal{L}_{adv,2} = \mathbb{E}_{(\mathbf{I}_{gt},\mathbf{C}_{comp})}\left[\log D_2(\mathbf{I}_{gt},\mathbf{C}_{comp})\right] + \mathbb{E}_{\mathbf{C}_{comp}}\log\left[1 - D_2(\mathbf{I}_{pred},\mathbf{C}_{comp})\right]$$

Edge Generator의 Adversarial Loss와 유사하며,
Input만 완성 이미지(I_{gt}, I_{pred})로 변경

2. Model Structure

< Image Completion Network (G_2, D_2) >



- Loss Function

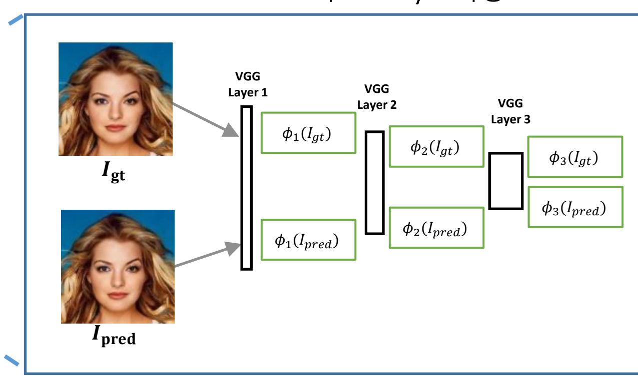
$$\mathcal{L}_{G_2} = \lambda_{\ell_1}\mathcal{L}_{\ell_1} + \lambda_{adv,2}\mathcal{L}_{adv,2} + \boxed{\lambda_p\mathcal{L}_{perc}} + \lambda_s\mathcal{L}_{style}$$

➢ Perceptual Loss

$$\mathcal{L}_{perc} = \mathbb{E}\left[\sum_i \frac{1}{N_i}\underline{\|\phi_i(\mathbf{I}_{gt}) - \phi_i(\mathbf{I}_{pred})\|_1}\right]$$

미리 학습해 놓은 다른 네트워크(VGG-19)를 활용해 얻은 Activation map 사이의 손실을 비교
(ϕ_i is the activation map of the i'th layer)

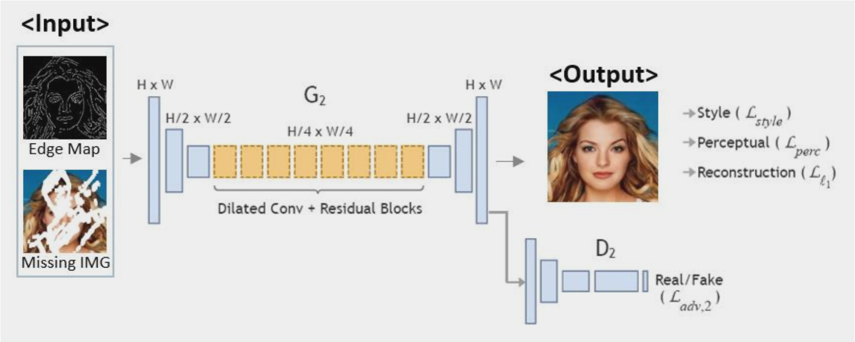
ImageNet dataset으로 Pre-train된 VGG-19의 1~5 Layer 사용



2. Model Structure

Smart Systems Lab. YONSEI UNIVERSITY, SEOUL, KOREA

< Image Completion Network (G_2, D_2) >



- Loss Function

$$\mathcal{L}_{G_2} = \lambda_{\ell_1}\mathcal{L}_{\ell_1} + \lambda_{adv,2}\mathcal{L}_{adv,2} + \lambda_p\mathcal{L}_{perc} + \boxed{\lambda_s\mathcal{L}_{style}}$$

➢ **Style Loss**

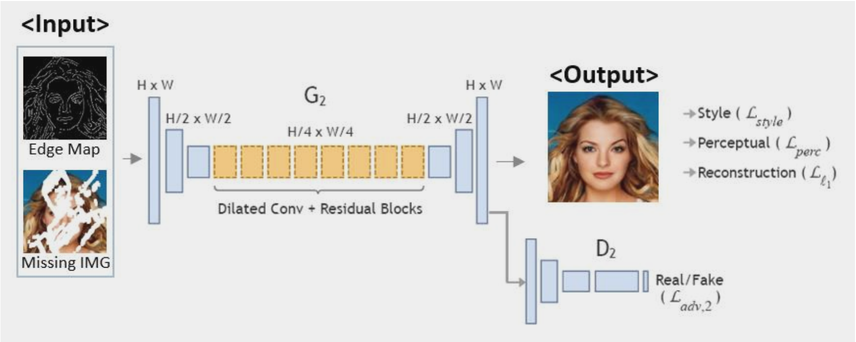
$$\mathcal{L}_{style} = \mathbb{E}_j\left[\|G_j^{\phi}(\tilde{\mathbf{I}}_{pred}) - G_j^{\phi}(\tilde{\mathbf{I}}_{gt})\|_1\right]$$

앞선 Perceptual Loss와 유사하게 Pred, GT를 VGG-19에 입력하여 나온
j'th Layer의 Activation map을 **Gram Matrix**로 변형하여 차이를 Loss로 사용

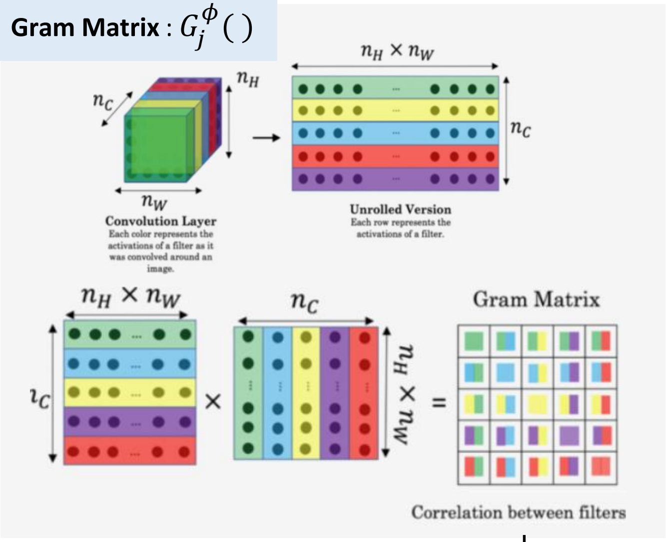
2. Model Structure

SSL Smart Systems Lab. YONSEI UNIVERSITY, SEOUL, KOREA

< Image Completion Network (G_2, D_2) >



Gram Matrix : $G_j^{\phi}(\,)$



n_c X n_c 행렬은 해당 Activation map 안의 서로 다른 두 지점에 있는 특징들 간의 co-occurrence를 담고 있음.

- Loss Function

$$\mathcal{L}_{G_2} = \lambda_{\ell_1}\mathcal{L}_{\ell_1} + \lambda_{adv,2}\mathcal{L}_{adv,2} + \lambda_p\mathcal{L}_{perc} + \boxed{\lambda_s\mathcal{L}_{style}}$$

➢ Style Loss

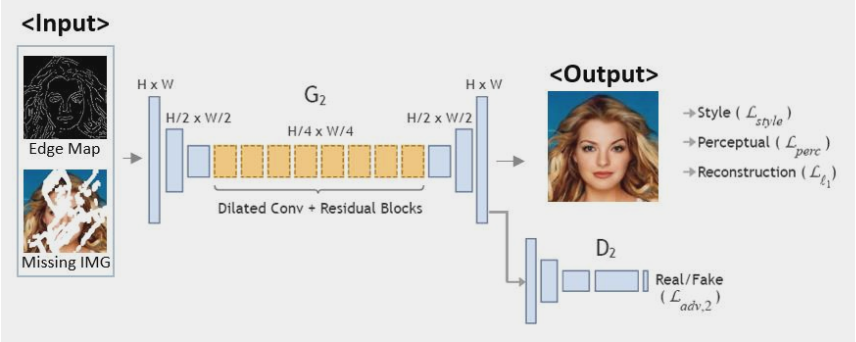
$$\mathcal{L}_{style} = \mathbb{E}_j\left[\|G_j^{\phi}(\tilde{\mathbf{I}}_{pred}) - G_j^{\phi}(\tilde{\mathbf{I}}_{gt})\|_1\right]$$

앞선 Perceptual Loss와 유사하게 Pred, GT를 VGG-19에 입력하여 나온
j'th Layer의 Activation map을 **Gram Matrix**로 변형하여 차이를 Loss로 사용

2. Model Structure

SSL Smart Systems Lab. YONSEI UNIVERSITY, SEOUL, KOREA

< Image Completion Network (G_2, D_2) >



- Loss Function

$$\mathcal{L}_{G_2} = \lambda_{\ell_1}\mathcal{L}_{\ell_1} + \lambda_{adv,2}\mathcal{L}_{adv,2} + \lambda_p\mathcal{L}_{perc} + \lambda_s\mathcal{L}_{style}$$

✓ Regularization Parameter Setting

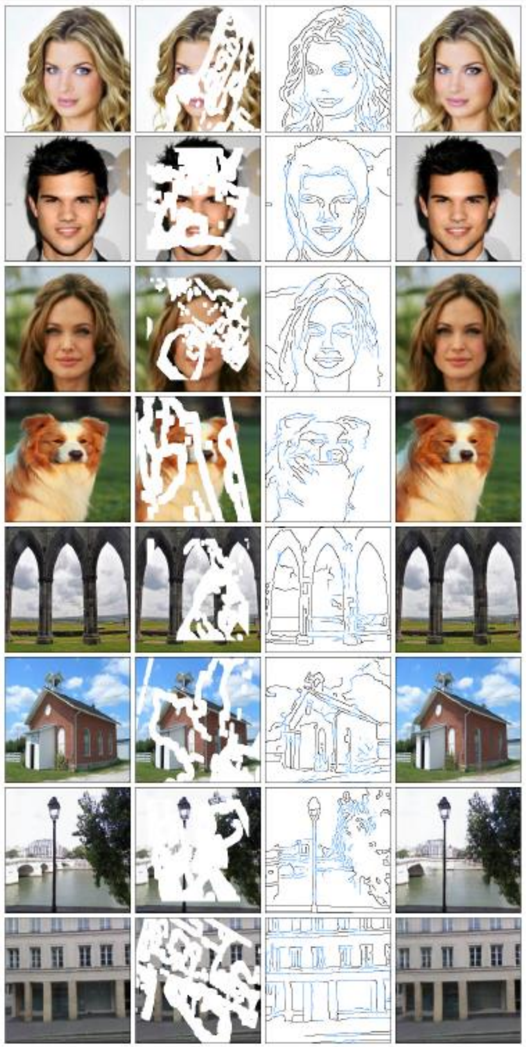
λ_{ℓ_1} = 1

$\lambda_{adv,2}$ = λ_p = 0.1

λ_s = 250

- **Experiments Setup**
  - Train & Test image size : 256 X 256
  - Canny edge detector 사용 ($\sigma \approx 2$)
  - Dataset : CelebA / Places2 / Paris StreetView
  - Baseline Models
    1) Contextual Attention (CA)
    2) Globally and Locally Consistent Image Completion (GLCIC)
    3) Partial Convolution (Pconv)
  - Evaluation Metrics
    1) Relative ℓ_1
    2) Structural Similarity Index (SSIM) : 시각적 화질 차이 및 유사도 측정 (휘도, 대비, 구조)
    3) Peak Signal-to-Noise Ratio (PSNR) : 최대신호대잡음비, 이미지 품질 측정
    4) Frechet Inception Distance (FID) : Feature space representation의 차이 측정

3. Experiments

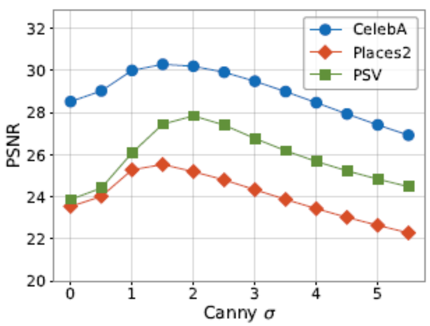


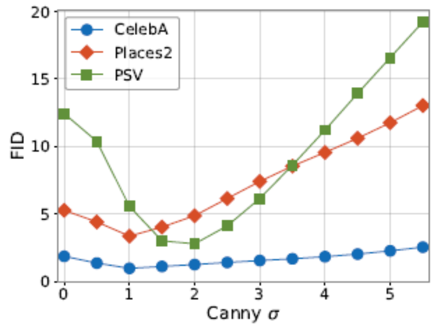
	Mask	CA	GLCIC	PConv*	Ours	Canny
ℓ_1 (%)†	10-20%	2.41	2.66	**1.14**	1.50	1.16
	20-30%	4.23	4.70	**1.98**	2.59	1.88
	30-40%	6.15	6.78	**3.02**	3.77	2.60
	40-50%	8.03	8.85	**4.11**	5.14	3.41
	Fixed	4.37	4.12	-	**3.86**	2.22
SSIM*	10-20%	0.893	0.862	0.869	**0.920**	0.941
	20-30%	0.815	0.771	0.777	**0.861**	0.902
	30-40%	0.739	0.686	0.685	**0.799**	0.863
	40-50%	0.662	0.603	0.589	**0.731**	0.821
	Fixed	0.818	0.814	-	**0.823**	0.892
PSNR*	10-20%	24.36	23.49	**28.02**	27.95	30.85
	20-30%	21.19	20.45	24.90	**24.92**	28.35
	30-40%	19.13	18.50	22.45	**22.84**	26.66
	40-50%	17.75	17.17	20.86	**21.16**	25.20
	Fixed	20.65	21.34	-	**21.75**	26.52
FID†	10-20%	6.16	11.84	-	**2.32**	2.25
	20-30%	14.17	25.11	-	**4.91**	3.42
	30-40%	24.16	39.88	-	**8.91**	4.87
	40-50%	35.78	54.30	-	**14.98**	7.13
	Fixed	8.31	8.42	-	**8.16**	3.24

3. Experiments

▪ Edge Detection 관련 실험

- Effect of σ in Canny Edge Detctor





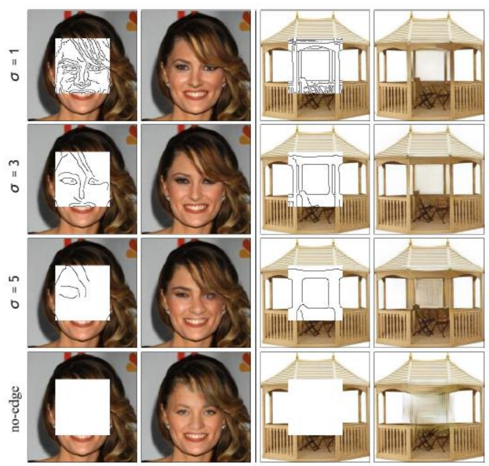


Figure 7: Effect of σ in Canny edge detector on inpainting results. Top to bottom: $\sigma = 1, 3, 5$, no edge data.

- Alternative Edge Detection Systems

1) Canny Edge Detector : speed, robustness, ease of use
2) Holistically-nested Edge Detector(HED)
   > NN 기반, Noise 적음, 윤곽선 위주
3) Element-wise multiplication

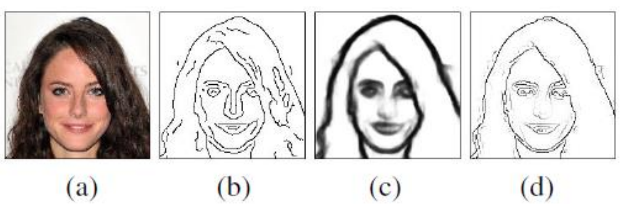
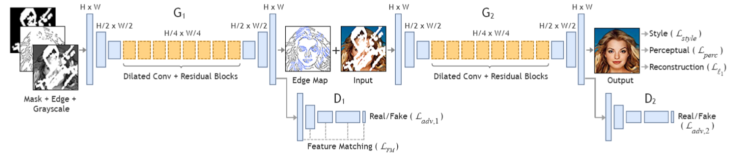
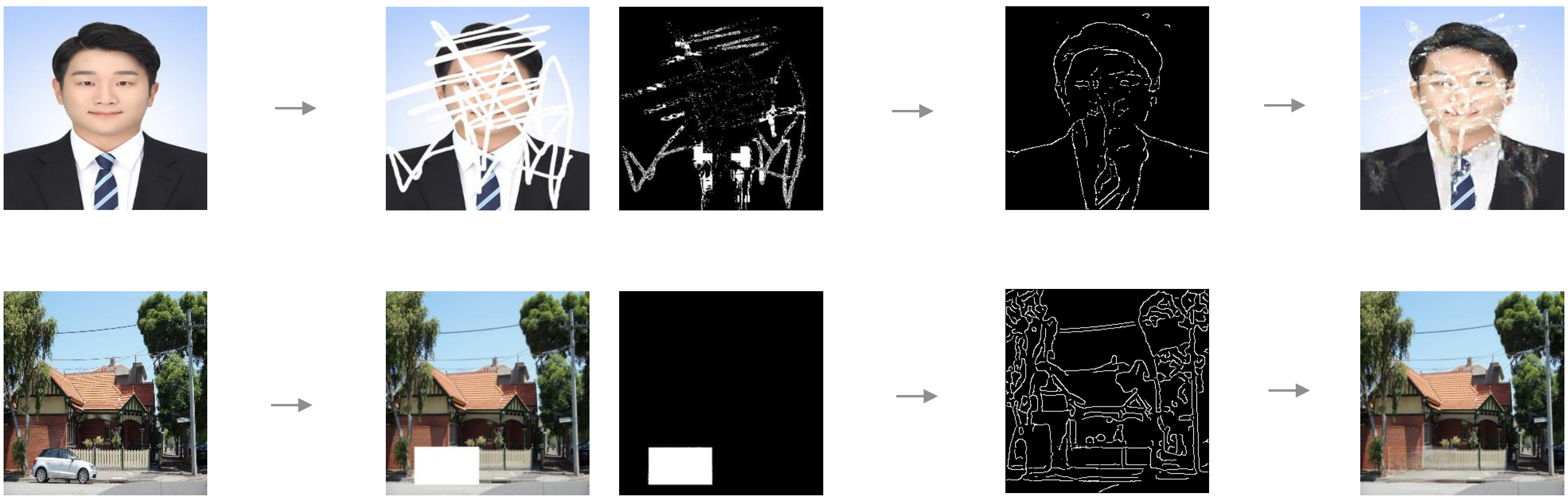


Figure 8: (a) Image. (b) Canny. (c) HED. (d) Canny$\odot$HED.

3. Experiments

- **Custom Data Test**





원본 이미지

Input

Edge Generator Result

Image Completion Result

감사합니다!