SagaGAN: Style Applied using Gram matrix Attribution based on StarGAN v2

Yongseon Yoo[1], Sunggyu Kim[2], Jong-Min Lee[1,2,3]

1 Department of Artificial Intelligence, Hanyang University, Seoul, Republic of Korea
2 Department of Electronic Engineering, Hanyang University, Seoul, Republic of Korea
3 Department of Biomedical Engineering, Hanyang University, Seoul, Republic of Korea

BMVC 2024

HYU HANYANG UNIVERSITY SEOUL, KOREA

INTRODUCTION

Background

- Image-to-image translation aims to convert images between domains while preserving content.
- Existing methods like StarGAN v2 use AdaIN for style transfer, but may not fully capture complex style characteristics.

$$AdaIN(x, y) = \sigma(y) \cdot \frac{x - \mu(x)}{\sigma(x)} + \mu(y)$$

- where x is the content feature map, y is the style feature map, μ and σ represent the channel-wise mean and standard deviation, respectively.
- Previous methods using the gram matrix could learn more style information than AdaIN by leveraging correlations between feature maps. However, the iterative inference process at that time was time-consuming.

$$G_{ij} = \sum_k F_{ik} \cdot F_{jk}$$

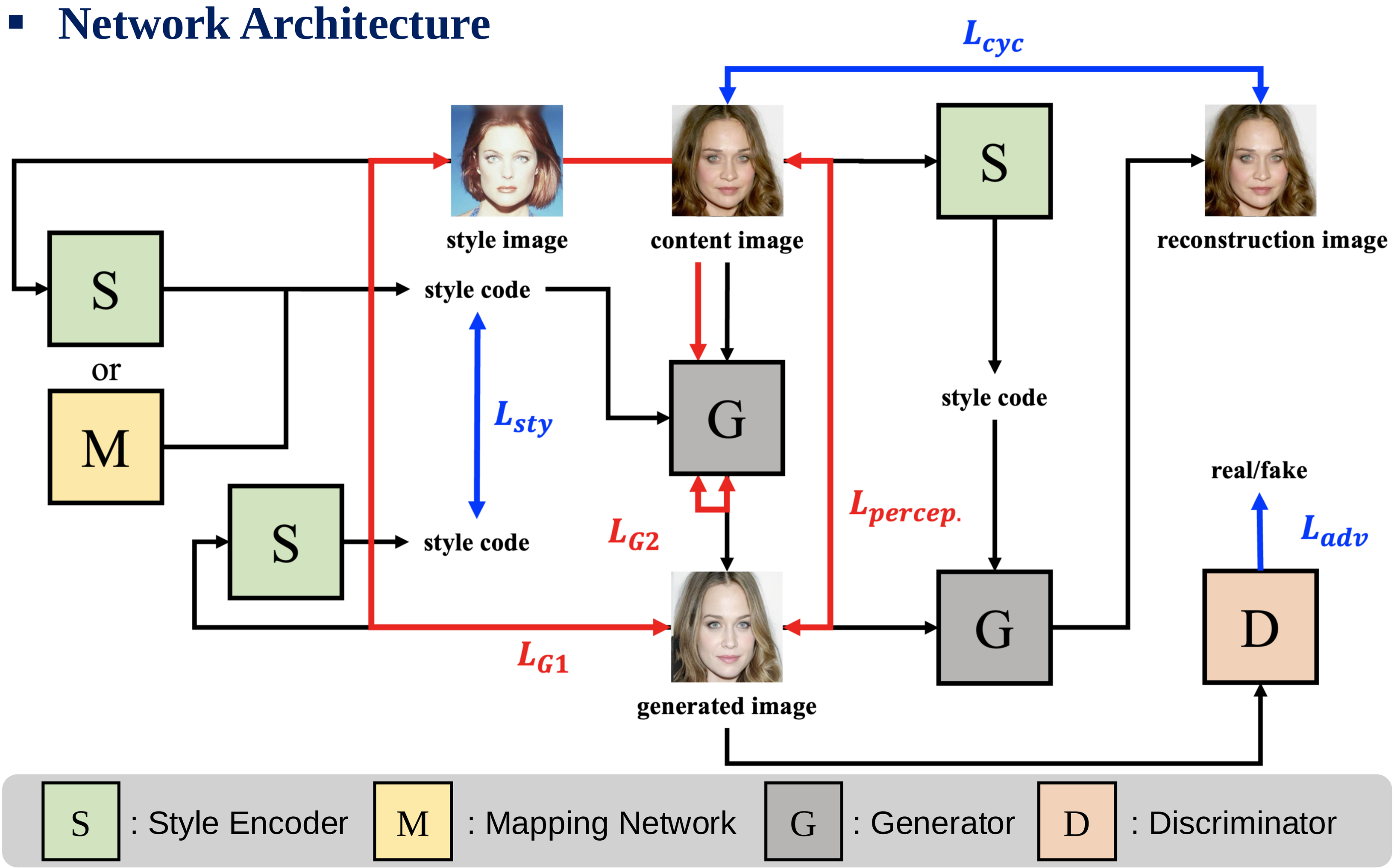
- The Gram matrix $G \in R^{C \times C}$ of a given feature map $F \in R^{C \times HW}$, where C is the number of channels, H is the height, and W is the width.
- where F_{ik} is the activation of the i-th channel at position k in the vectorized feature map, and F_{jk} is the activation of the j-th channel at the same position.

Research Purpose

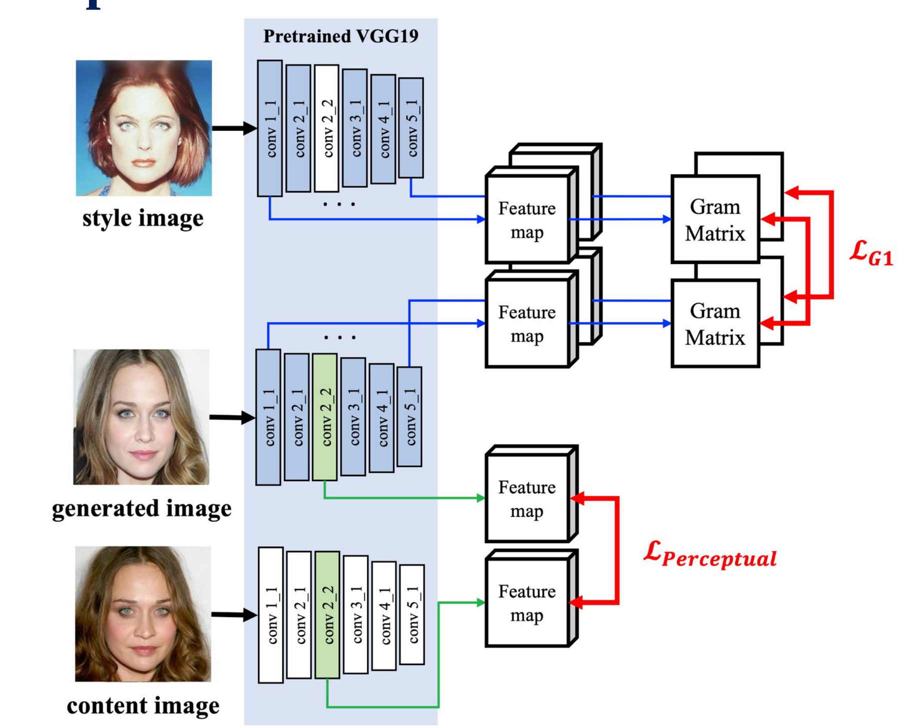
- We propose SagaGAN, a novel approach combining the gram matrix with AdaIN to enhance style transfer capabilities.
- Our goal is to improve the quality and diversity of generated images in multi-domain image-to-image translation tasks while maintaining content preservation.

METHOD

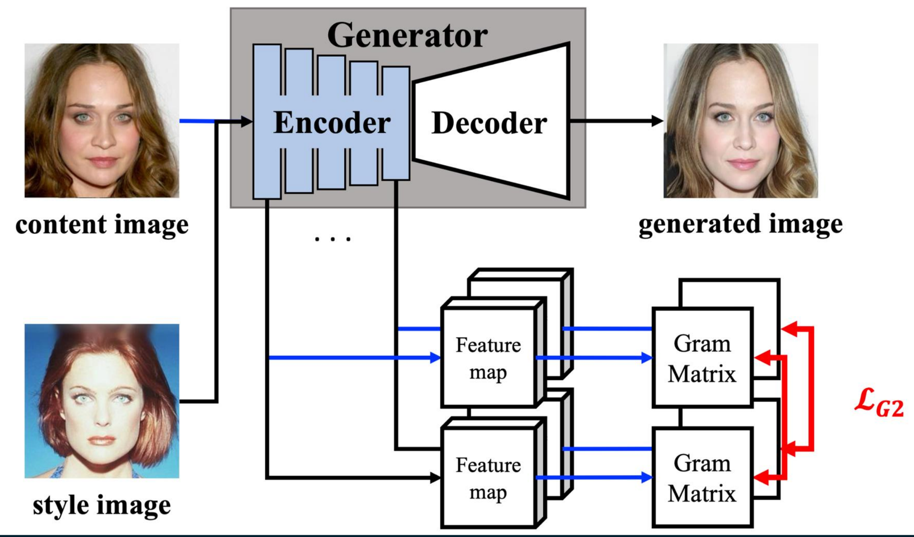
Network Architecture



G1 Loss, Perceptual Loss

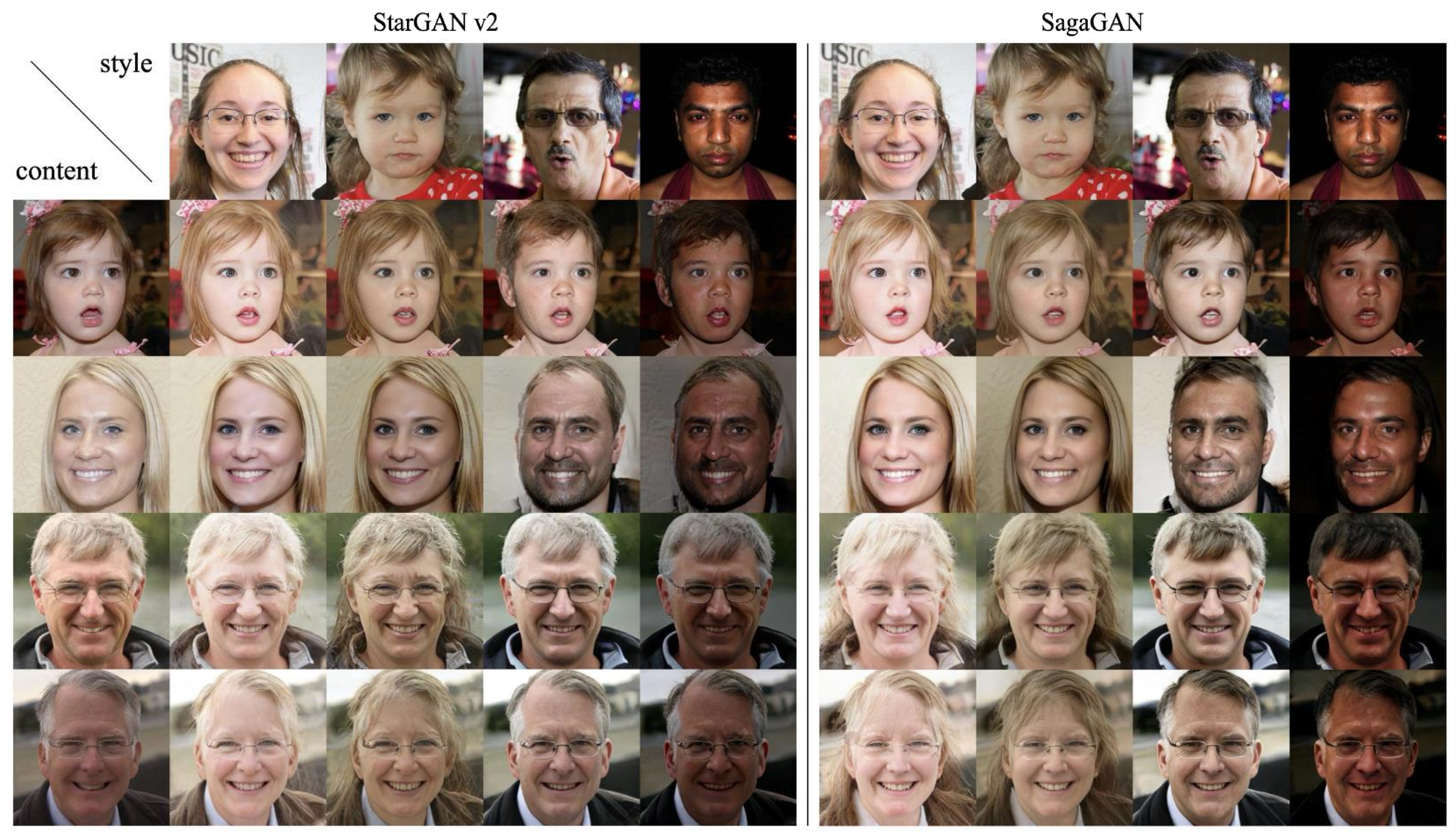


G2 Loss



RESULTS

Qualitative Results



- Qualitative results for each dataset CelebA-HQ(top), AFHQ(middle), FFHQ(bottom).

Quantitative Results

Dataset	Method	Task	latent		Reference	
			FID↓	LPIPS↑	FID↓	LPIPS↑
CelebA-HQ	StarGANv2	male→female	10.55	**0.444**	**19.27**	0.374
		female→male	18.62	0.460	26.70	0.401
		Mean	14.59	**0.452**	22.98	0.388
	SagaGAN	male→female	**9.73**	0.436	19.49	**0.380**
		female→male	**17.32**	**0.462**	**26.15**	**0.401**
		Mean	**13.53**	0.449	**22.82**	**0.391**
AFHQ	StarGAN v2	dog→cat	**6.67**	0.411	7.01	0.413
		cat→dog	37.97	0.413	41.53	0.440
		wild→cat	**8.31**	**0.456**	**7.73**	**0.416**
		cat→wild	33.97	0.452	39.10	0.421
		wild→dog	**31.91**	0.445	**36.89**	0.432
		dog→wild	33.84	0.434	40.56	0.408
		Mean	25.44	0.435	28.80	0.422
	SagaGAN	dog→cat	7.31	**0.420**	**6.99**	**0.418**
		cat→dog	**35.60**	**0.458**	**38.60**	**0.443**
		wild→cat	8.59	0.416	7.81	0.415
		cat→wild	**16.65**	**0.462**	**18.04**	**0.444**
		wild→dog	33.50	0.452	37.62	**0.433**
		dog→wild	**16.14**	**0.464**	**19.70**	**0.444**
		Mean	**19.63**	**0.445**	**21.46**	**0.433**
FFHQ	StarGAN v2	male→female	22.36	0.067	21.78	0.092
		female→male	26.08	0.061	25.83	0.081
		Mean	24.22	0.064	23.80	0.086
	SagaGAN	male→female	**18.78**	**0.133**	**18.42**	**0.137**
		female→male	**23.17**	**0.136**	**23.50**	**0.134**
		Mean	**20.98**	**0.135**	**20.96**	**0.136**

ACKNOWLEDGEMENT

This work was supported by Institute of Information communications Technology Planning Evaluation (IITP) grant funded by the Korea government(MSIT) (No.RS-2020-II201373, Artificial Intelligence Graduate School Program(Hanyang University))