Font Style Transfer

Presenter: 차성국

Team Font @ MINDs Lab/Brain-Vision

Content

- Introduction:이것은 무엇이고, 왜 해야하며, 무슨 가치가 있는가?
- Benchmarks: 기존의 연구들은 어떤가?

Introduction

폰트는 아주 중요한 시각적 요소(비언어적 표현)이다.

그런데 폰트를 만드는 것은 아주 비싼 작업이다.

영어의 경우 대소문자 52자면 충분하지만, 한글이나 한자같은 합성자는 수만 글자로 이루어져있어 굉장히 어려운 작업이다.

폰트를 만드는 알고리즘을 만든다면 정말 awesome할거야!

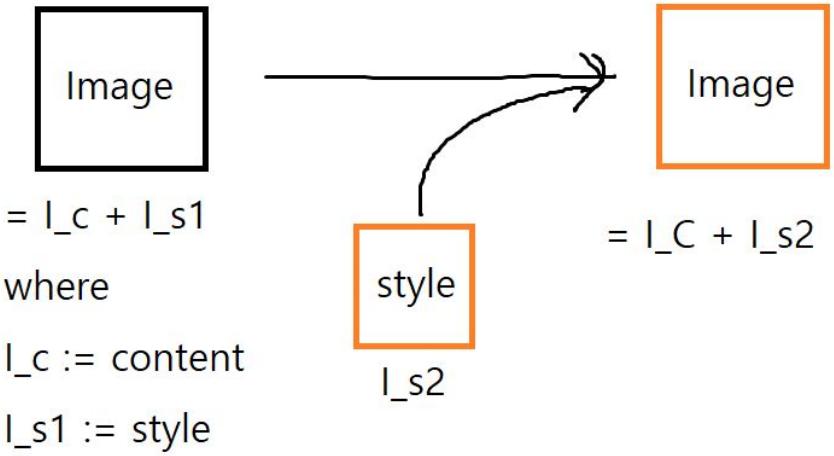
Inter-language style transfer 가 포함되었으면 좋겠다.

Introduction

유의미한 키워드인지는 모르겠음.
자주 붙기는 함

Font style transfer as few-shot image-to-image translation

Image-to-image translation



Benchmarks

EMD (CVPR’18)

FUNIT (ICCV’19)

DMFont (ECCV’20)

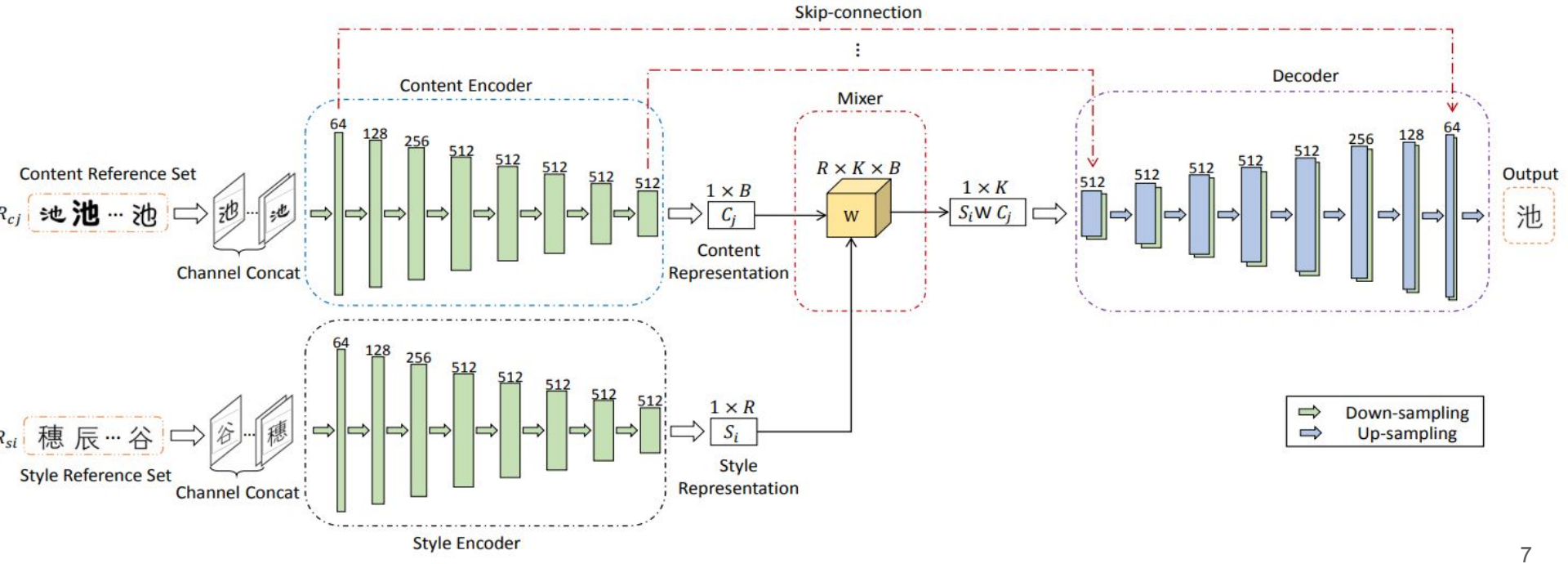
FTransGAN (WACV’21)

LFFont (AAAI’21)

Benchmark Objectives

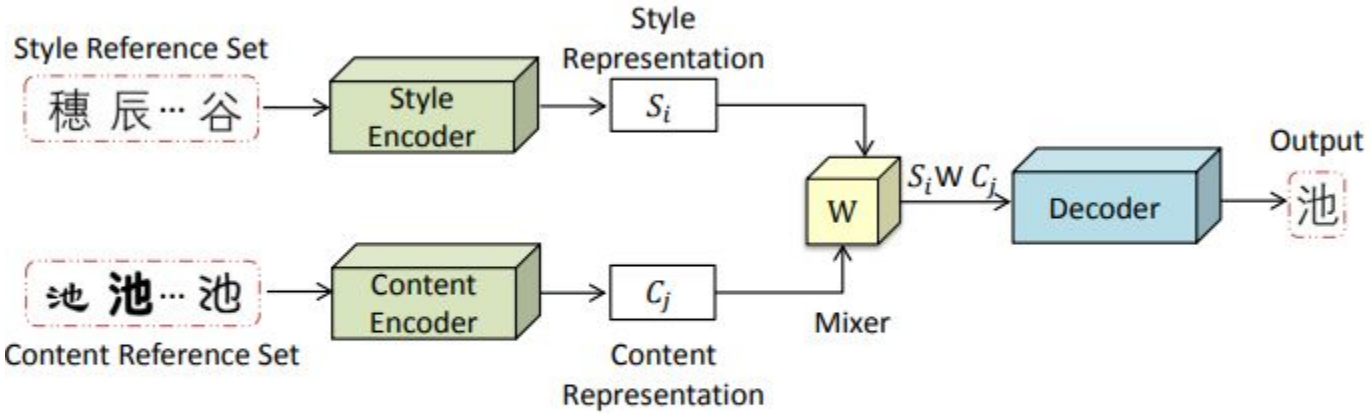
Model	Objective
EMD (CVPR'18)	$\theta = \arg\min_{\theta} \sum_{I_{ij} \in \mathcal{D}_t} L(\hat{I}_{ij}, I_{ij} \vert \mathcal{R}_{S_i}, \mathcal{R}_{C_j})$
FUNIT (ICCV'19)	$\min_{D} \max_{G} \mathcal{L}_{\text{GAN}}(D,G) + \lambda_{\text{R}} \mathcal{L}_{\text{R}}(G) + \lambda_{\text{F}} \mathcal{L}_{\text{FM}}(G)$
DMFont (ECCV'20)	$\min_{G,C} \max_{D} \mathcal{L}_{adv(font)} + \mathcal{L}_{adv(char)} + \lambda_{l1}\mathcal{L}_{l1} + \lambda_{feat}\mathcal{L}_{feat} + \lambda_{cls}\mathcal{L}_{cls}$
LFFont (AAAI'21)	$\min_{\substack{E_c,E_{s,u},G,\\ F_s,F_u,Cls}} \max_{D} \mathcal{L}_{adv(font)} + \mathcal{L}_{adv(char)} + \lambda_{L1}\mathcal{L}_{L1} + \lambda_{feat}\mathcal{L}_{feat} + \lambda_{cls}\mathcal{L}_{cls} + \lambda_{consist}\mathcal{L}_{consist}$
FTransGAN (WACV'21)	$L = \lambda_1 L_1 + \lambda_s L_{\text{style}} + \lambda_c L_{\text{content}},$

Separating Style and Content for Generalized Style Transfer (CVPR'18) EMD



EMD 3줄요약

- (당시) 기존의 방법들은 '특정 style' → '특정 style' transformation이 되었었다.
- 이 논문은 'generalized' style transfer network를 제안한다.



- 새로운 multi-task learning scenario를 제안한다.
  - Training with a triplet <R_{S_i}, R_{C_j}, I_{ij}>
  - where I_{ij} denotes the target image of style S_i and content C_j
  - R_{S_i} and R_{C_j} are respectively the style and content reference sets, each consisting of *r* random images.

EMD Objective

$$\theta = \arg\min_{\theta} \sum_{I_{ij} \in \mathcal{D}_t} L(\hat{I}_{ij}, I_{ij} | \mathcal{R}_{S_i}, \mathcal{R}_{C_j})$$

$$L(\hat{I}_{ij}, I_{ij} | \mathcal{R}_{S_i}, \mathcal{R}_{C_j}) = W_{st}^{ij} \times W_b^{ij} \times ||\hat{I}_{ij} - I_{ij}||$$

$$W_{st}^{ij} = 1/N_b^{ij}$$ about the size and thickness of characters

$$W_b^{ij} = \frac{exp(\text{mean}_{ij})}{\sum_{I_{ij} \in \mathcal{D}_t} exp(\text{mean}_{ij})}$$ about darkness of characters

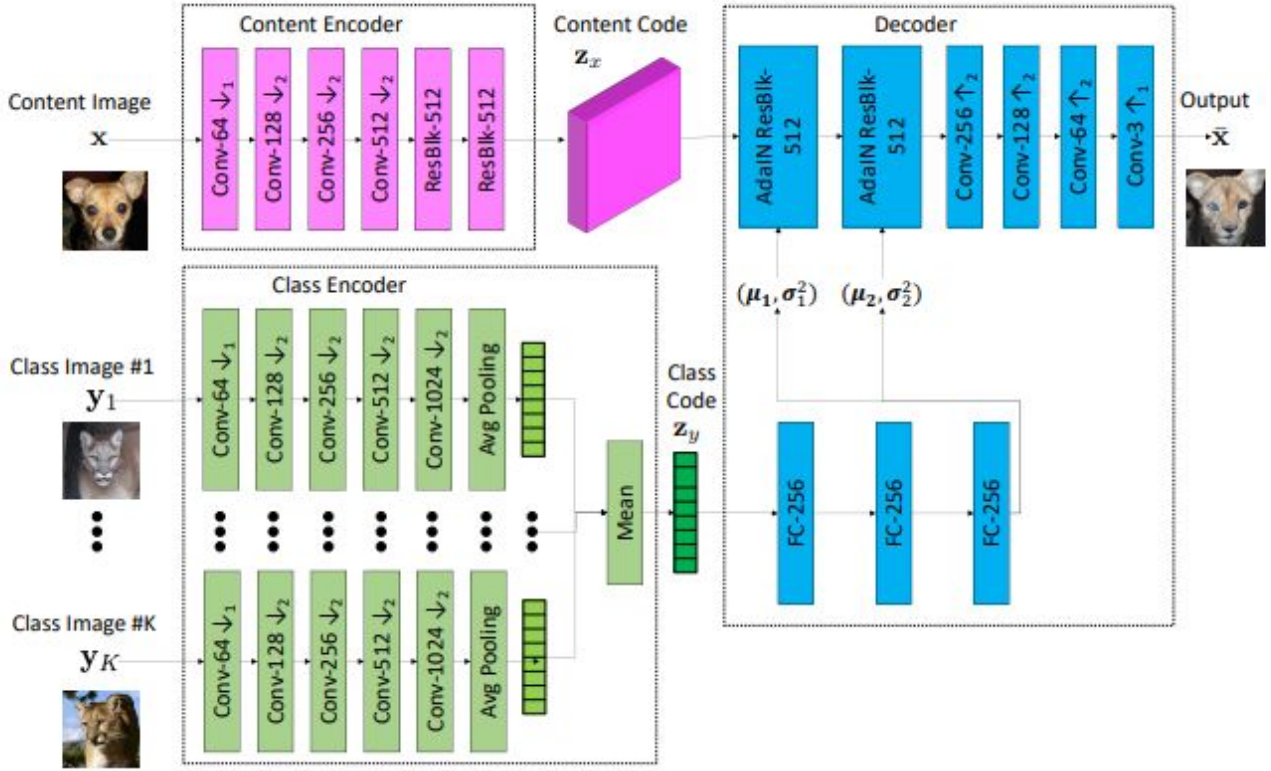
L1 loss in font supervision

generated

ground truth

L1 = Pixel-level loss sum

Few-Shot Unsupervised Image-to-Image Translation (ICCV'19) FUNIT



FUNIT Objective

$$\min_D \max_G \mathcal{L}_{\text{GAN}}(D,G) + \lambda_{\text{R}}\mathcal{L}_{\text{R}}(G) + \lambda_{\text{F}}\mathcal{L}_{\text{FM}}(G)$$

$$\mathcal{L}_{\text{GAN}}(G,D) = E_{\mathbf{x}}\left[-\log D^{c_x}(\mathbf{x})\right] + E_{\mathbf{x},\{\mathbf{y}_1,\ldots,\mathbf{y}_K\}}\left[\log\left(1 - D^{c_y}(\bar{\mathbf{x}})\right)\right]$$

Reconstruction Loss

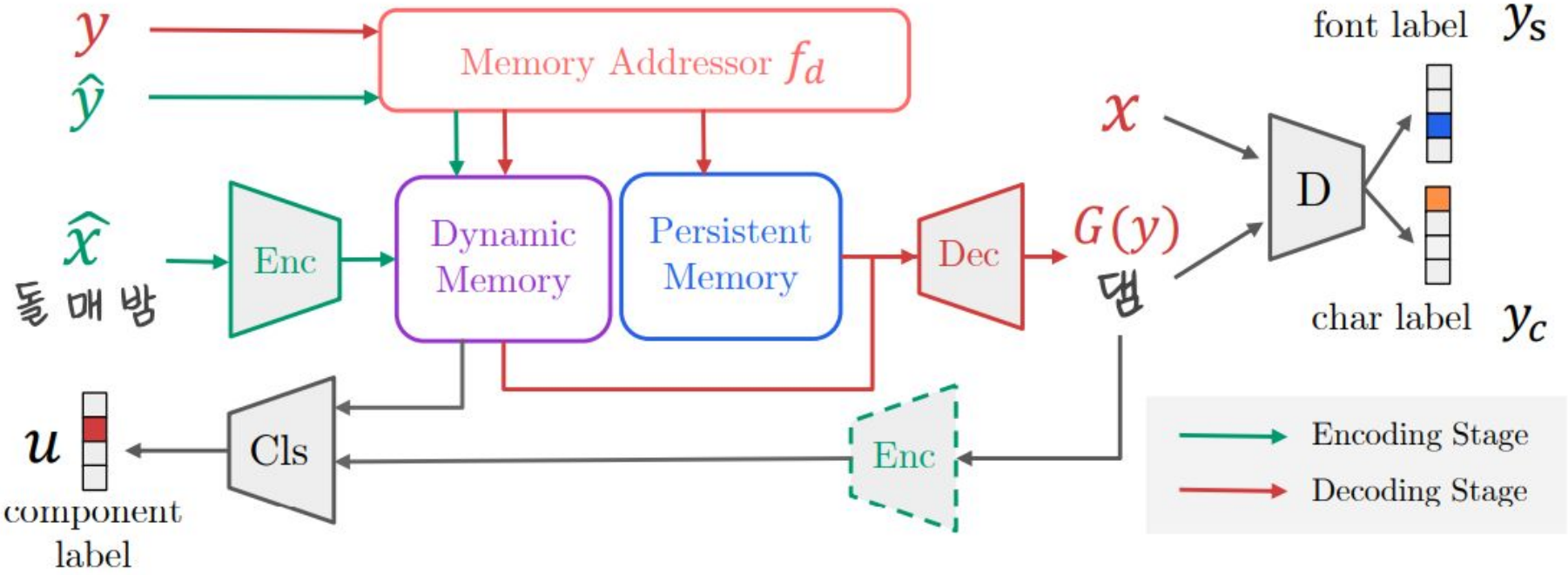
$$\mathcal{L}_{\text{R}}(G) = E_{\mathbf{x}}\left[||\mathbf{x} - G(\mathbf{x},\{\mathbf{x}\})||_1^1\right]$$

$$\bar{\mathbf{x}} = G(\mathbf{x},\{\mathbf{y}_1,\ldots,\mathbf{y}_K\})$$

Feature regularization

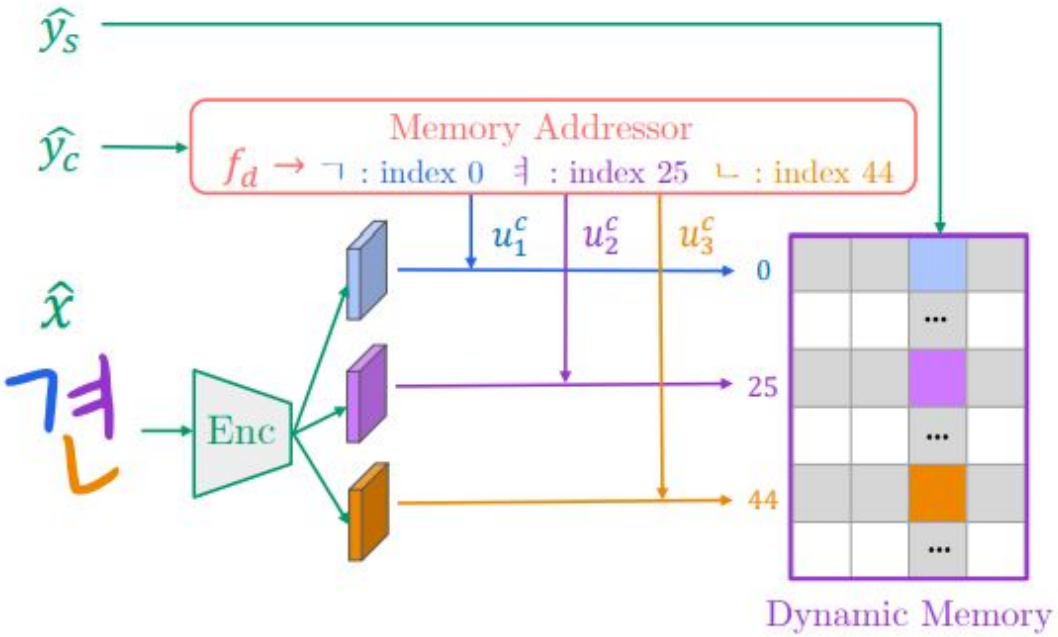
$$\mathcal{L}_{\text{F}}(G) = E_{\mathbf{x},\{\mathbf{y}_1,\ldots,\mathbf{y}_K\}}\left[||D_f(\bar{\mathbf{x}})) - \sum_k \frac{D_f(\mathbf{y}_k)}{K}||_1^1\right]$$

Few-shot Compositional Font Generation with Dual Memory (ECCV'20) DMFont



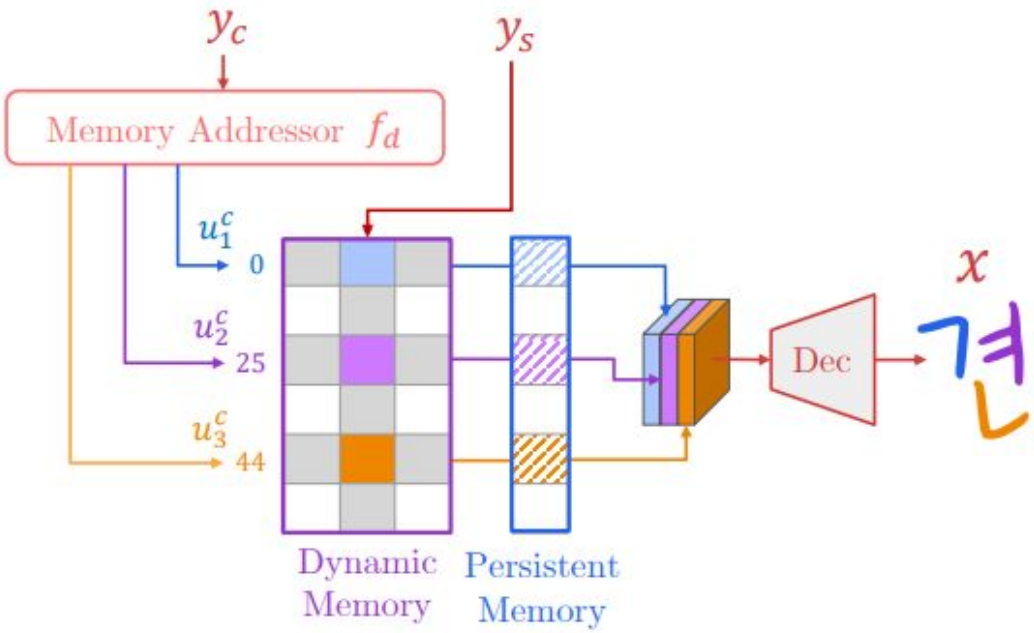
(a) Architecture overview.

DMFont



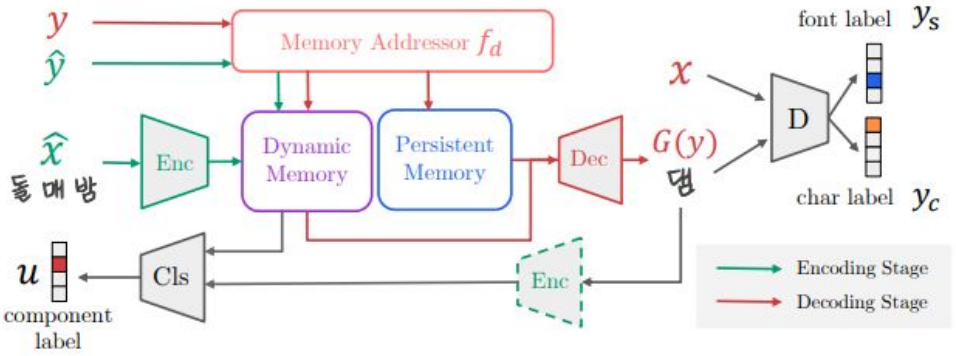
(b) Encoding phase detail.

DMFont

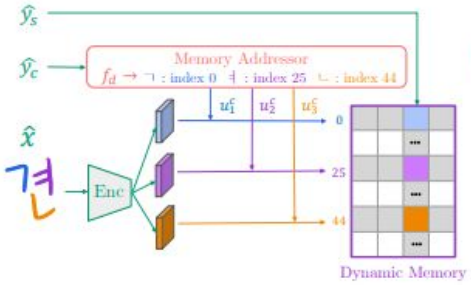


(c) Decoding phase detail.

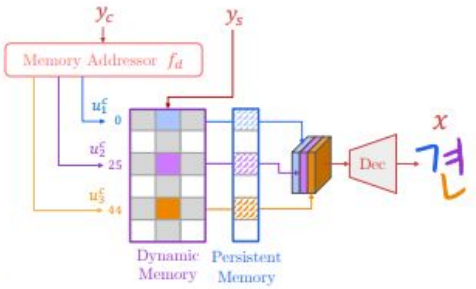
DMFont



(a) Architecture overview.

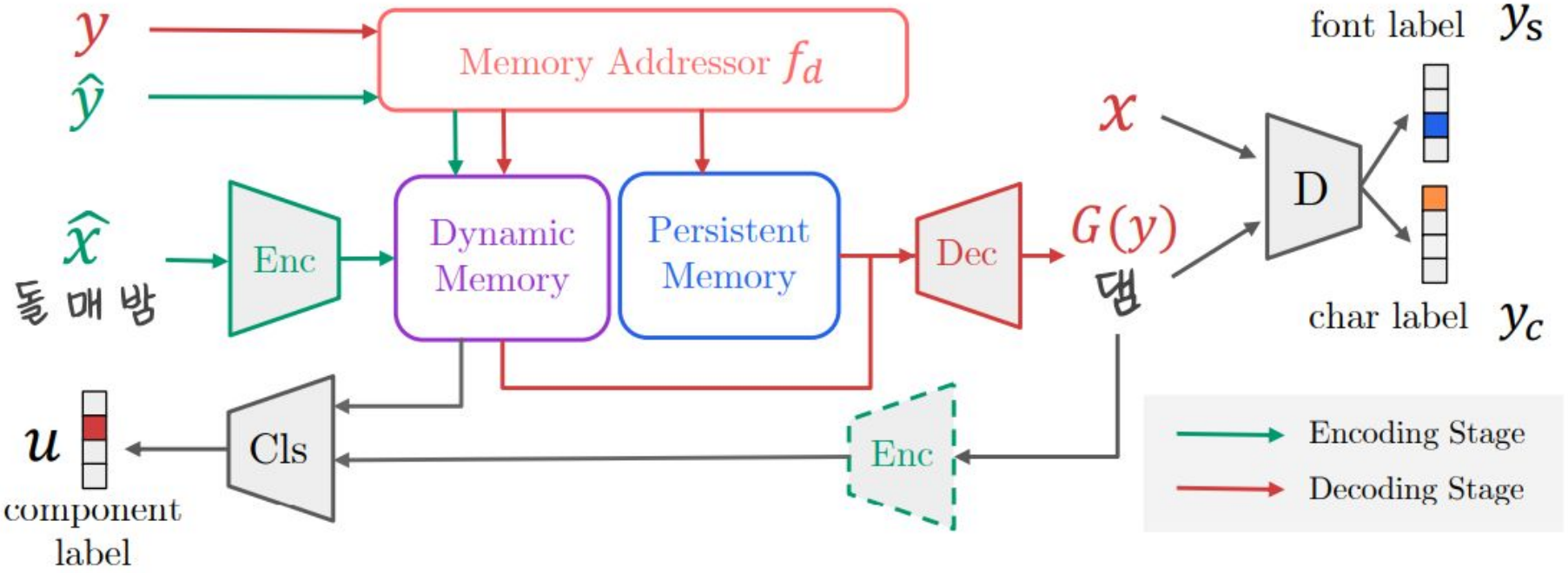


(b) Encoding phase detail.



(c) Decoding phase detail.

DMFont



(a) Architecture overview.

DMFont Objective

$$\min_{G,C} \max_{D} \mathcal{L}_{adv(font)} + \mathcal{L}_{adv(char)} + \lambda_{l1}\mathcal{L}_{l1} + \lambda_{feat}\mathcal{L}_{feat} + \lambda_{cls}\mathcal{L}_{cls}$$

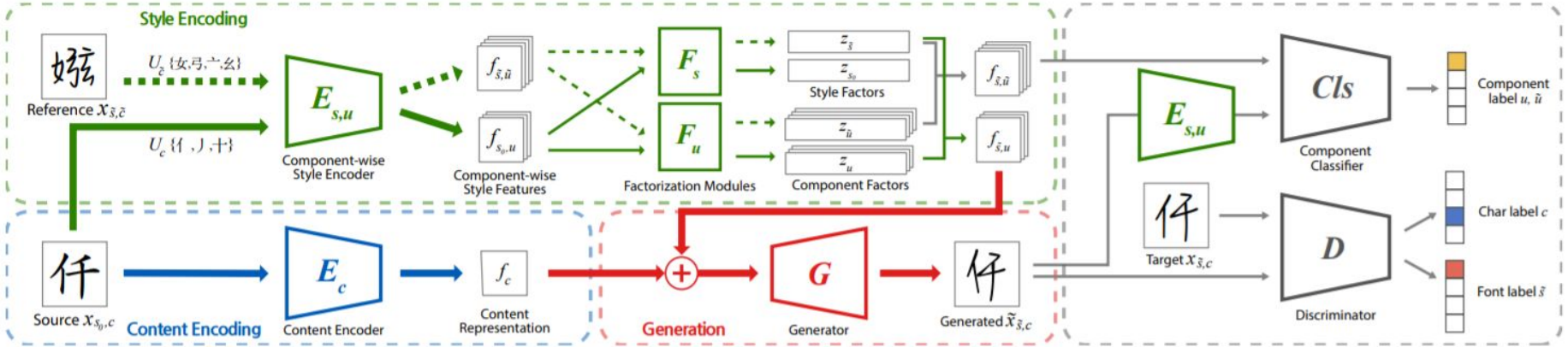
Classification loss: the same as which of image classification

$$\mathcal{L}_{cls} = \mathbb{E}_{x,y}\left[\sum_{u_i^c \in f_d(y_c)} \mathrm{CE}(Enc_i(x), u_i^c)\right] + \mathbb{E}_{y}\left[\sum_{u_i^c \in f_d(y_c)} \mathrm{CE}(Enc_i(G(y_c, y_f)), u_i^c)\right]$$

GT 이미지 가지고
classification

생성된 녀석 가지고
classification

Few-shot Font Generation with Localized Style Representations and Factorization (AAAI'20)
LFFont



	Style s	Character c	Components U_c
夏	s_1	夏	{一,目,丶,夂}
冬	s_1	冬	{夂,冫}
冬	s_2	冬	{夂,冫}

LFFont

$$x_{\tilde{s},c} = G(f_{\tilde{s}}, f_c),$$
$$f_{\tilde{s}} = E_s(\mathcal{X}_r) \text{ and } f_c = E_c(x_{s_0,c})$$

$$x(\tilde{s}, c) = G(f_{\tilde{s},c}, f_c), \quad f_c = E_c(x_{s_0,c}),$$
$$f_{\tilde{s},c} = \sum_{u \in U_c} f_{\tilde{s},u} = \sum_{u \in U_c} E_{s,u}(x_{\tilde{s},\tilde{c}_u}, u),$$

LFFont Objective

$$\min_{\substack{E_c,E_{s,u},G,\\ F_s,F_u,Cls}} \max_{D} \mathcal{L}_{adv(font)} + \mathcal{L}_{adv(char)} + \lambda_{L1}\mathcal{L}_{L1} + \lambda_{feat}\mathcal{L}_{feat} + \lambda_{cls}\mathcal{L}_{cls} + \lambda_{consist}\mathcal{L}_{consist},$$

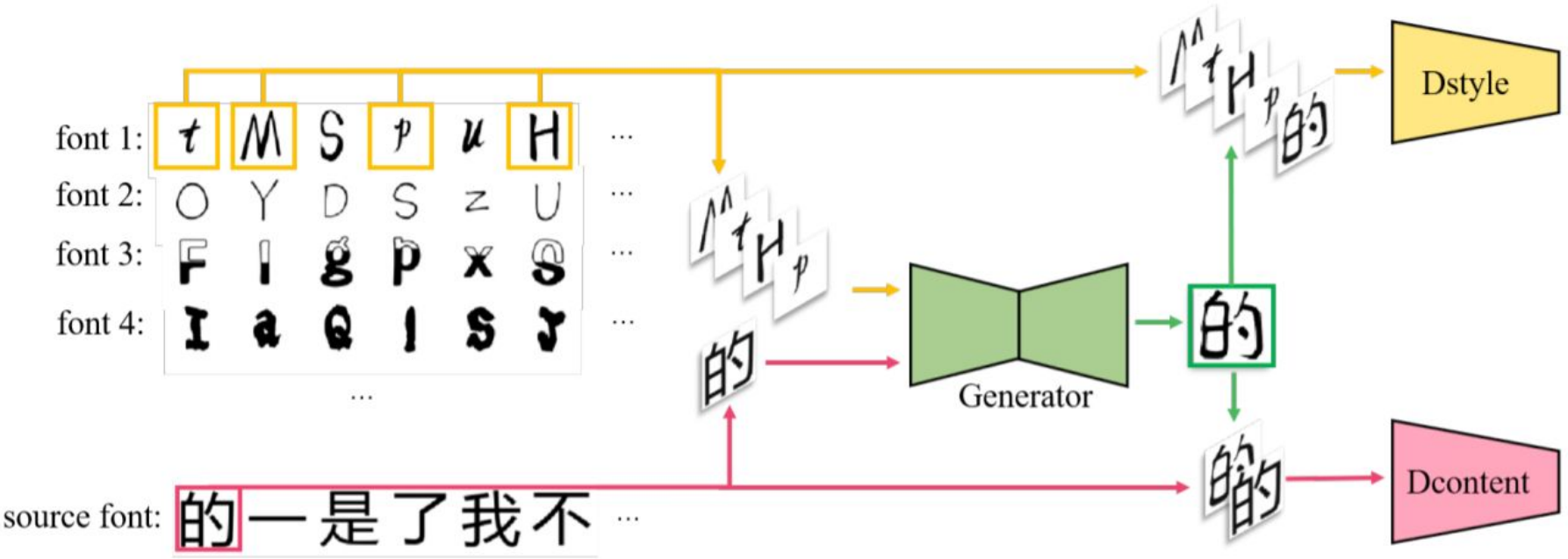
DFFont

Factorization loss...

$$\mathcal{L}_{consist} = \sum_{s\in\mathcal{S}}\sum_{u\in\mathcal{U}} \|F_s(f_{s,u}) - \mu_s\|_2^2 + \|F_u(f_{s,u}) - \mu_u\|_2^2$$

$$\mu_s = \frac{1}{|\mathcal{U}|}\sum_{u\in\mathcal{U}} F_s(f_{s,u}), \quad \mu_u = \frac{1}{|\mathcal{S}|}\sum_{s\in\mathcal{S}} F_u(f_{s,u}).$$

Few-shot Font Style Transfer between Different Languages (WACV'21) FTransGAN



FTransGAN

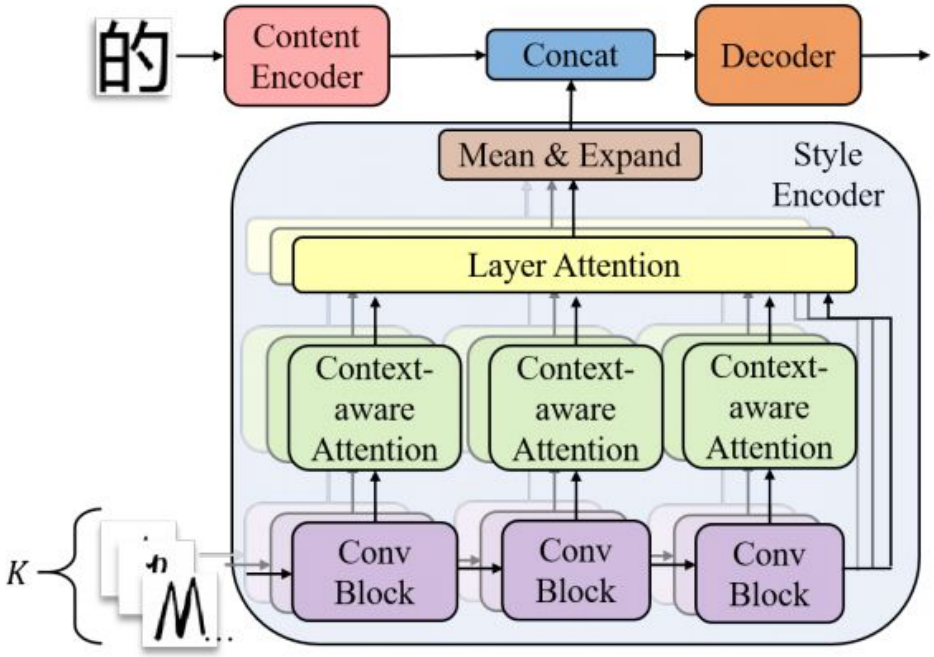


Figure 3. Overview of the proposed Generator G.

FTransGAN

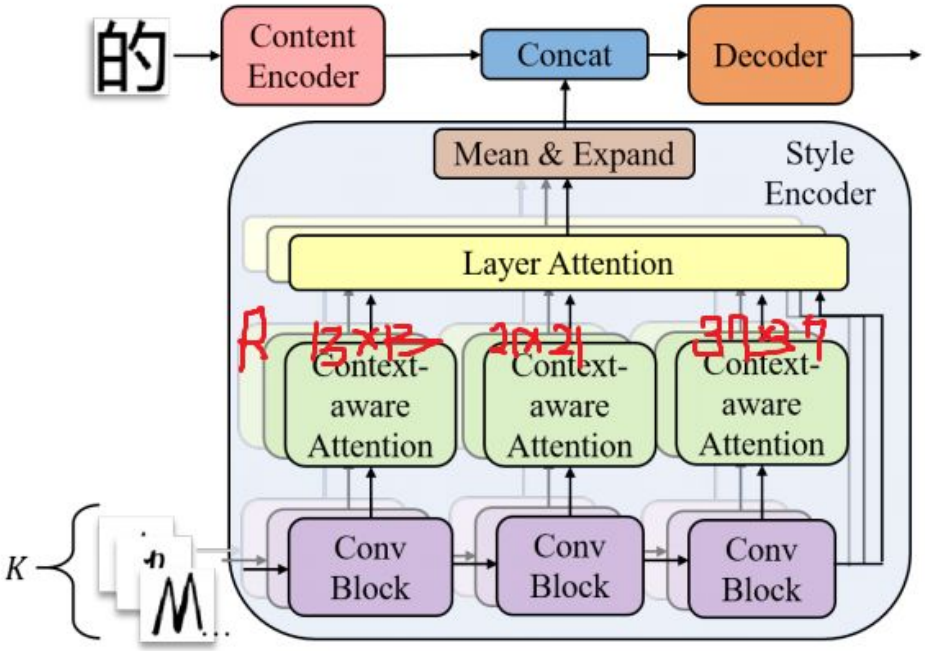


Figure 3. Overview of the proposed Generator G.

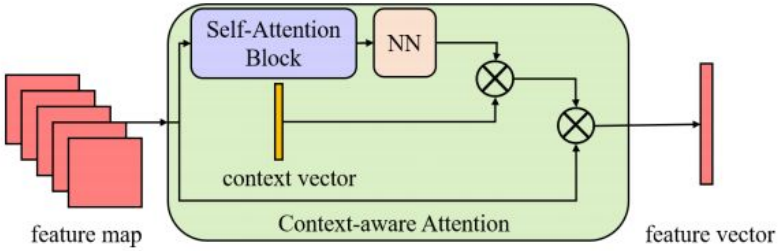


Figure 4. Architecture of the proposed Context-aware Attention Network.

non local neural network (He K.)
non local block

FTransGAN

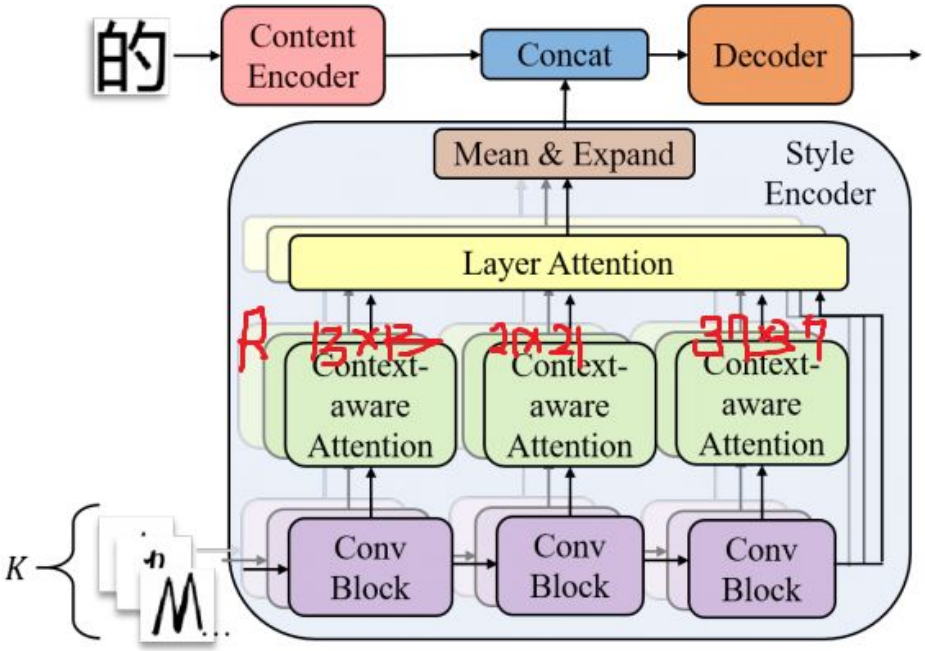


Figure 3. Overview of the proposed Generator G.

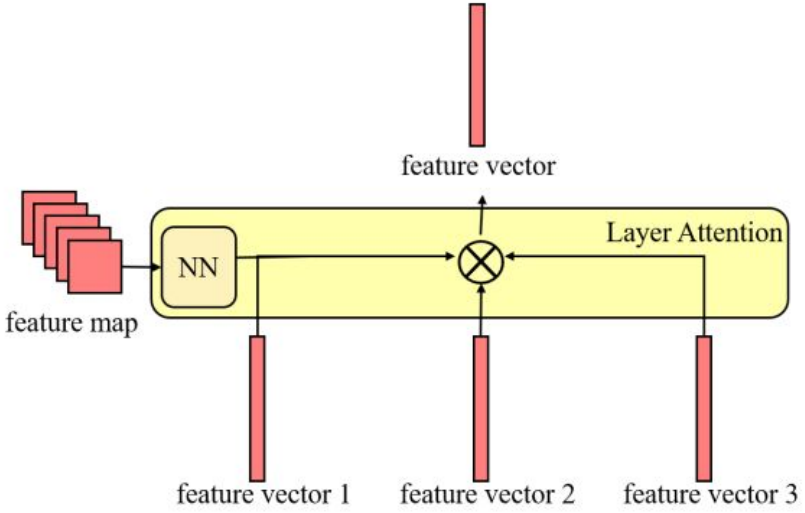


Figure 5. Architecture of the proposed Layer Attention Network.

FTransGAN Objective

$$L = \lambda_1 L_1 + \lambda_s L_{\text{style}} + \lambda_c L_{\text{content}}$$

Benchmark Objectives

Model	Objective
EMD (CVPR'18)	$\theta = \arg\min_{\theta} \sum_{I_{ij} \in \mathcal{D}_t} L(\hat{I}_{ij}, I_{ij} \vert \mathcal{R}_{S_i}, \mathcal{R}_{C_j})$
FUNIT (ICCV'19)	$\min_{D} \max_{G} \mathcal{L}_{\text{GAN}}(D, G) + \lambda_{\text{R}} \mathcal{L}_{\text{R}}(G) + \lambda_{\text{F}} \mathcal{L}_{\text{FM}}(G)$
DMFont (ECCV'20)	$\min_{G,C} \max_{D} \mathcal{L}_{adv(font)} + \mathcal{L}_{adv(char)} + \lambda_{l1}\mathcal{L}_{l1} + \lambda_{feat}\mathcal{L}_{feat} + \lambda_{cls}\mathcal{L}_{cls}$
LFFont (AAAI'21)	$\min_{\substack{E_c, E_{s,u}, G, \\ F_s, F_u, Cls}} \max_{D} \mathcal{L}_{adv(font)} + \mathcal{L}_{adv(char)} + \lambda_{L1}\mathcal{L}_{L1} + \lambda_{feat}\mathcal{L}_{feat} + \lambda_{cls}\mathcal{L}_{cls} + \lambda_{consist}\mathcal{L}_{consist}$
FTransGAN (WACV'21)	$L = \lambda_1 L_1 + \lambda_s L_{\text{style}} + \lambda_c L_{\text{content}},$

Implementation

Approach	#Style Shot	#Content Shot
EMD	10	**10**
DMFont	30 Korean 40 Thai	1
LFFont	8	1
FTransGAN	6	1

Datasets

Paper	Language	#Fonts	#Chars	#Train fonts	#Train chars	#Test fonts	Size
EMD	Chinese	832	1732	75%	75%		80x80
DMFont	Korean Handwriting (Expert refined)	86	2448	80%	90%		
	Thai-printing	105	?	80%	90%		
	Korean-unrefined	88	150				
LFFont	Chinese from web	482	19514 (total) 6654 (average) 371 (components)	467		15	
FTransGAN	ENG2CHN	847					64x64