ACM
Multimedia 2022
Lisbon, Portugal | 10-14 October

MF-Net: A Novel Few-shot Stylized Multilingual Font Generation Method

Yufan Zhang, Junkai Man, Peng Sun

Duke Kunshan University

昆山杜克大学
DUKE KUNSHAN
UNIVERSITY

Agenda

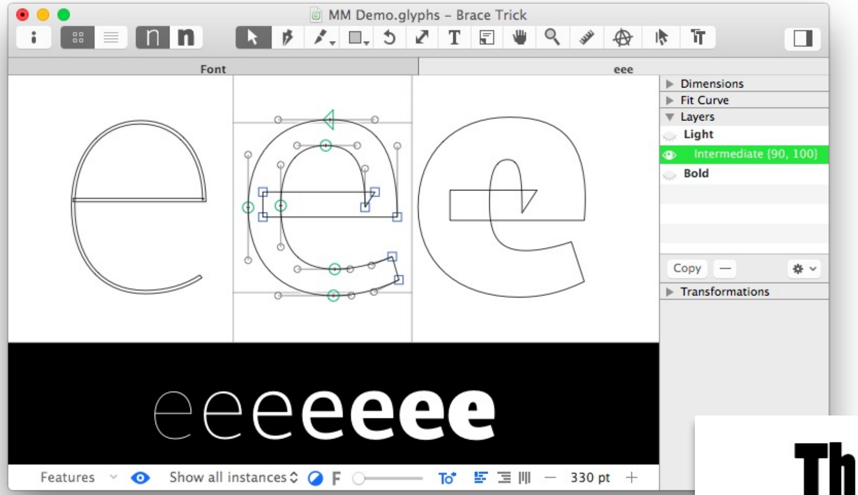
- Introduction

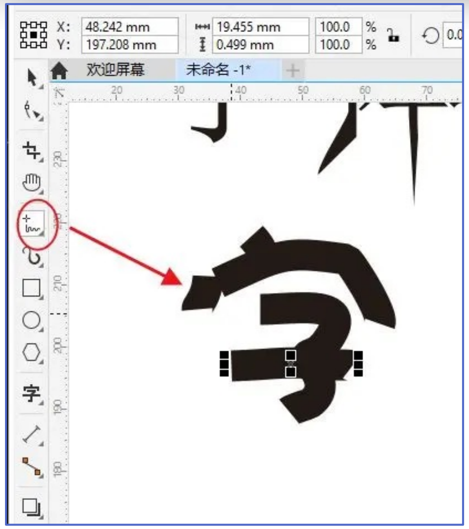
- The proposed method

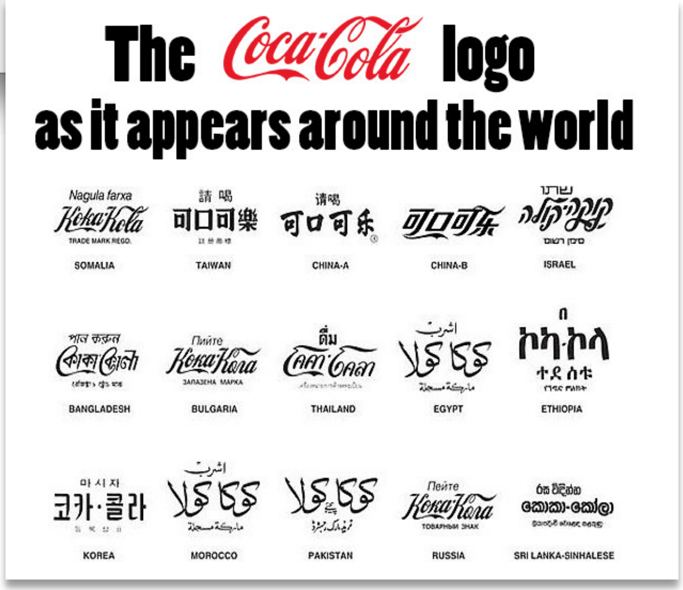
- Performance evaluation

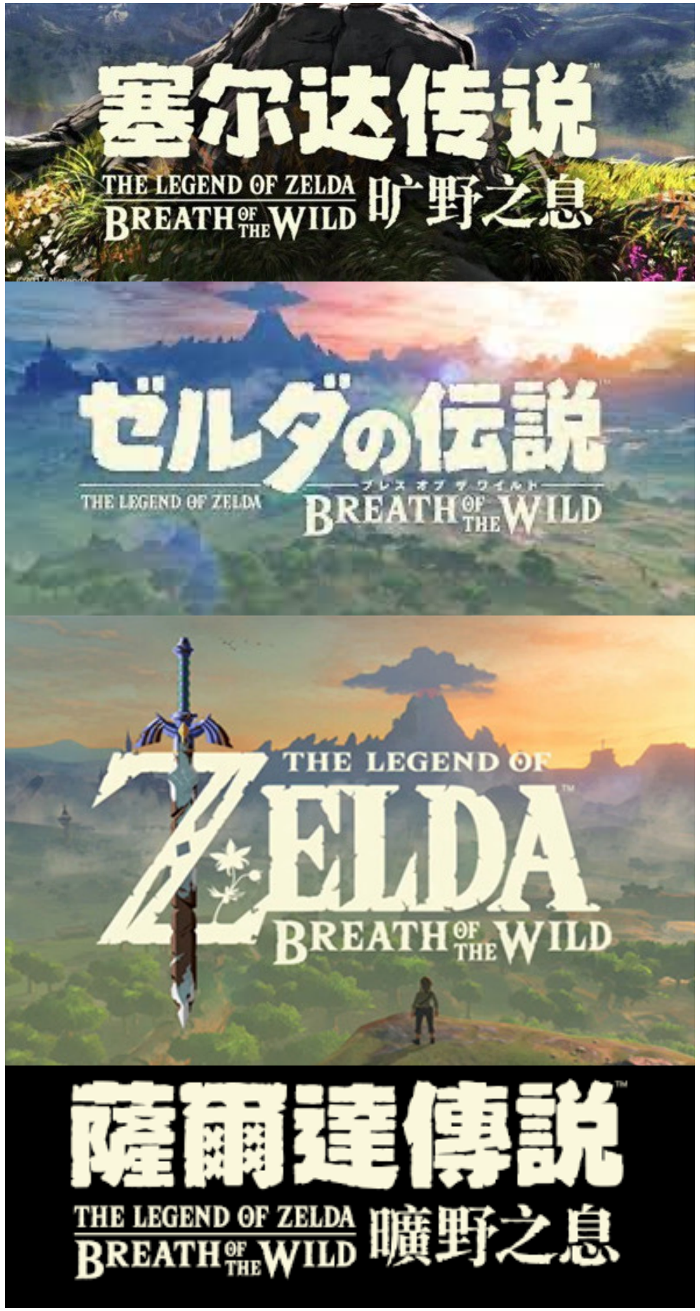
- Conclusion

Introduction

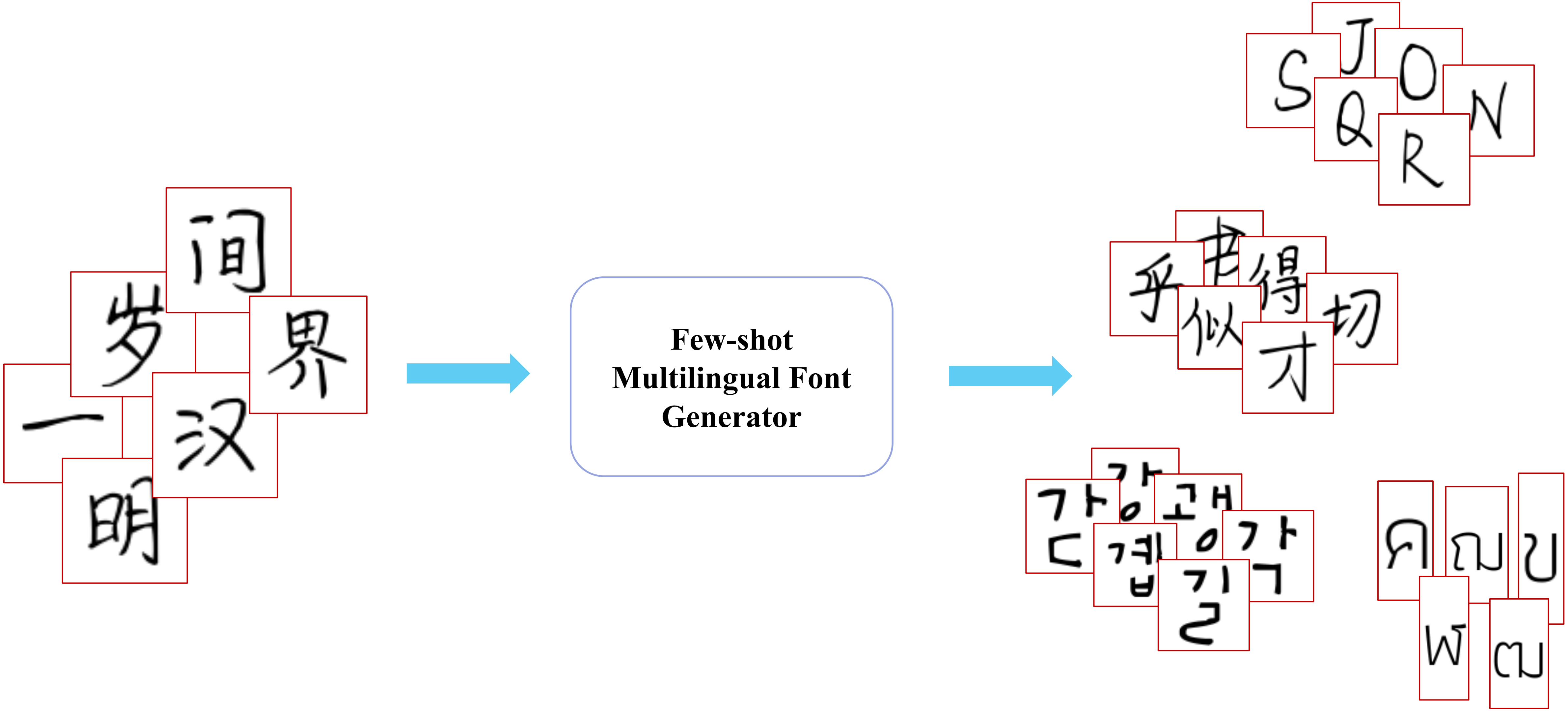








Introduction



Few-shot Multilingual Font Generator

Introduction

Existing methods for font style transfer

- Some models need a large number of input reference images of the target style.
- Some models need to fine-tune the pre-trained model with the style reference images to get the generated stylized font images.
- Some models only focus on the font style transfer within the same language or between two different languages that the model is trained on (dual-lingual).

MF-Net

- In a few-shot learning fashion
- Support font style transfer between untrained languages (multilingual)
- Generate target images by direct inference

Introduction

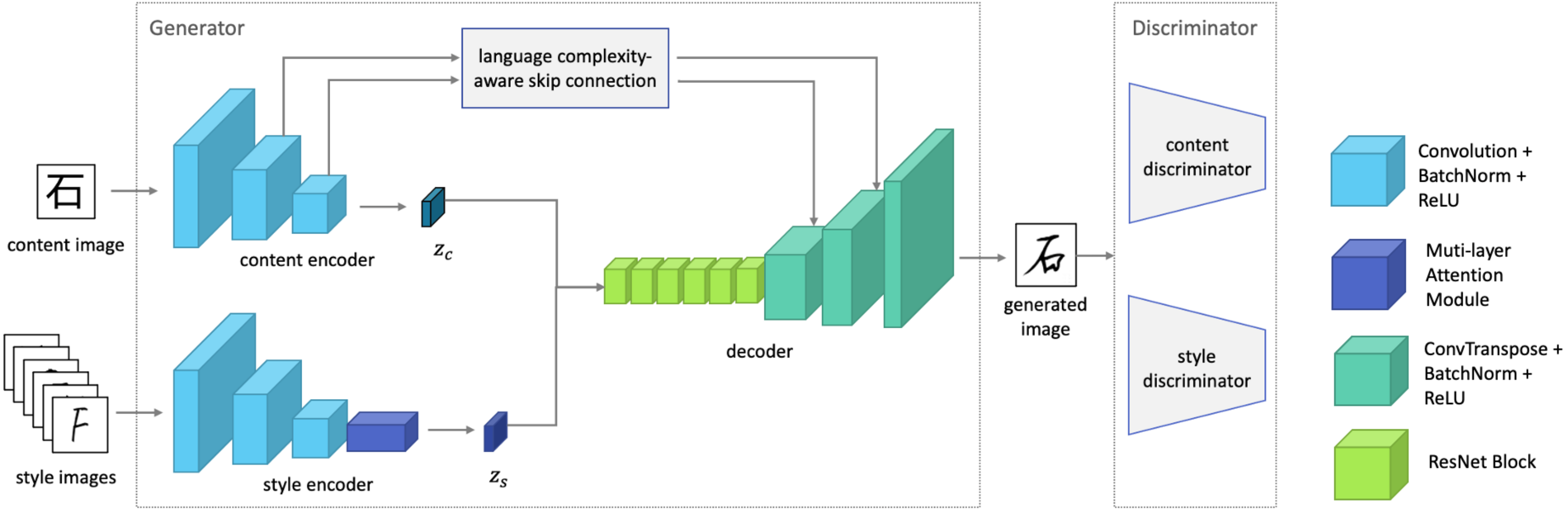
Main contributions of our work

- We propose the challenging task of few-shot stylized multilingual font generation and build a validation dataset for it.
- We propose a novel GAN-based model, MF-Net, which first presents a deep learning solution to font style transfer to characters of unseen languages.
- We design a novel language complexity-aware skip connection to adaptively adjust the structural information of the content to be preserved.
- We introduce a novel loss function, namely encoder consistent loss, to better disentangle the content and style features.

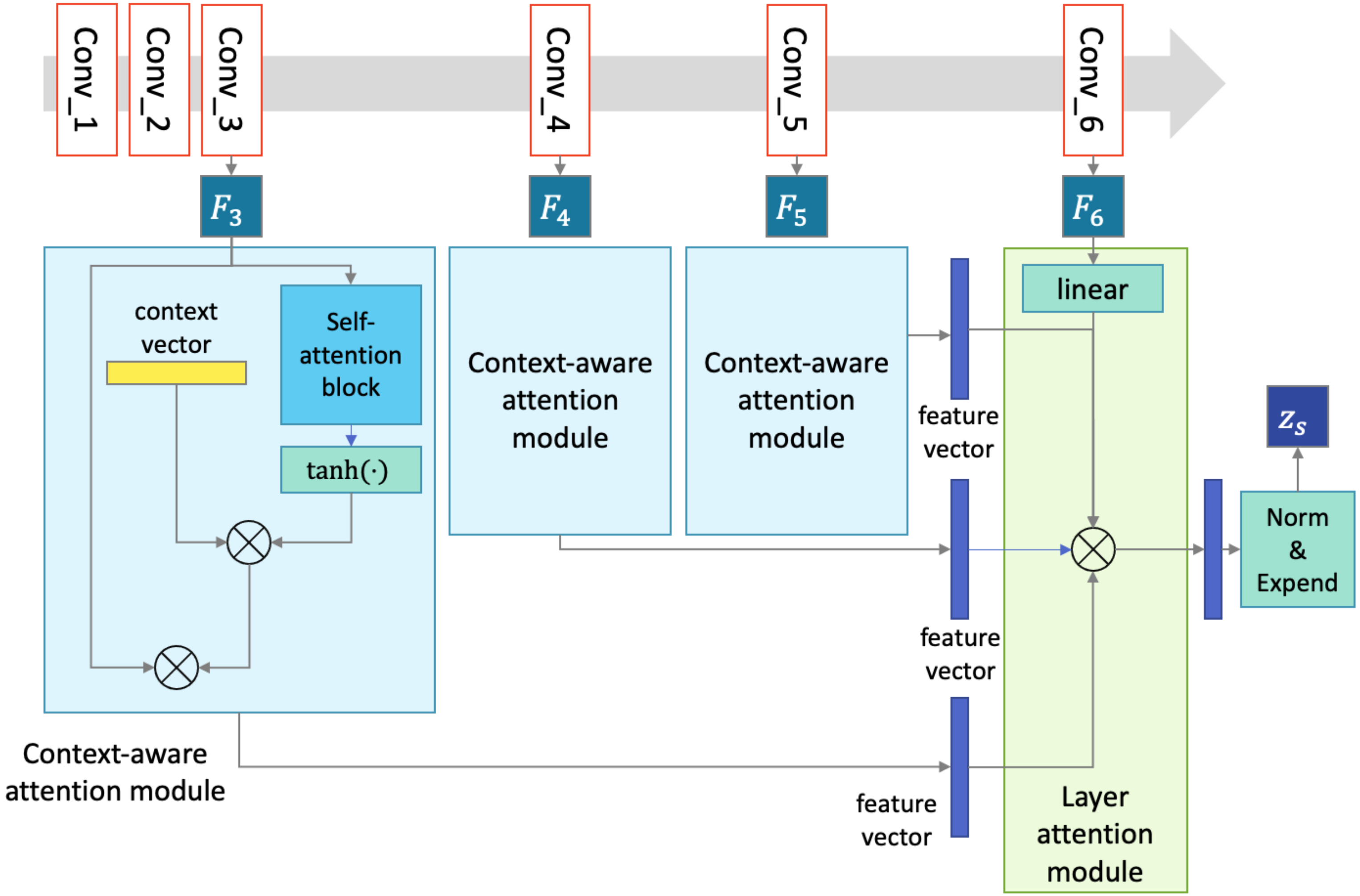
Agenda

- Introduction
- The proposed method
  - Network overview
  - Style encoder
  - Language complexity-aware skip connections.
  - Loss function
- Performance evaluation
- Conclusion

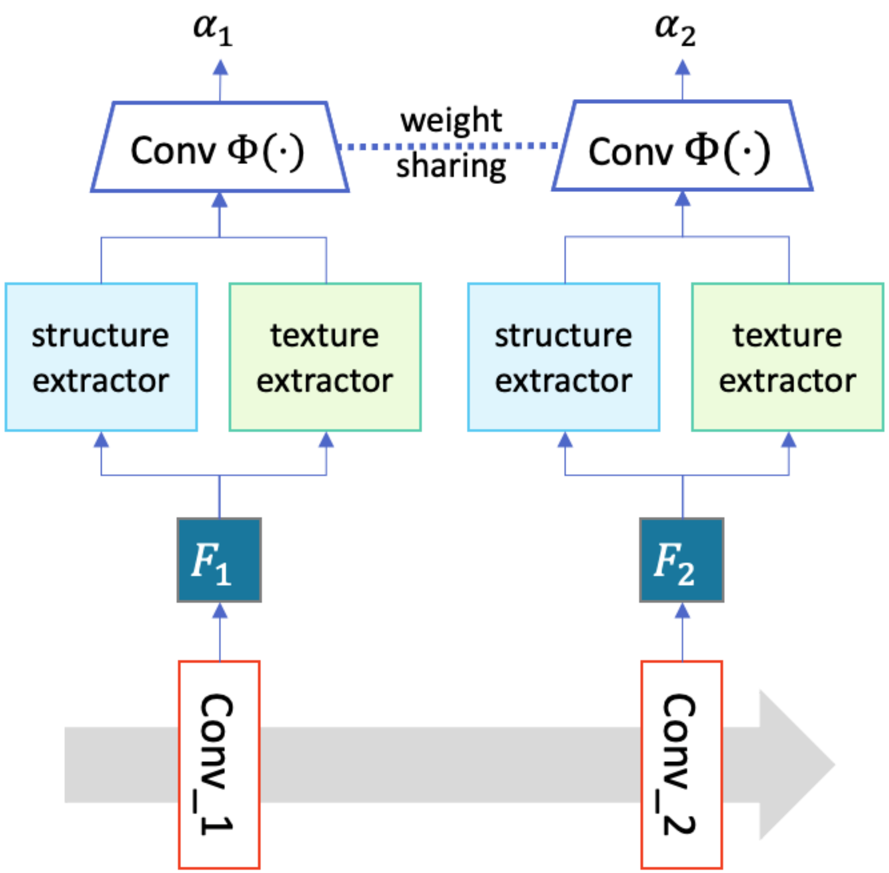
Network Overview



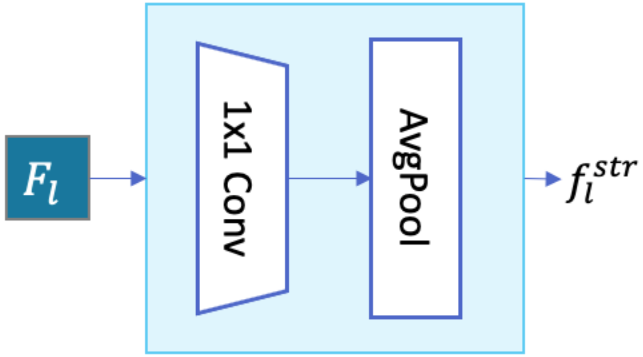
Style Encoder



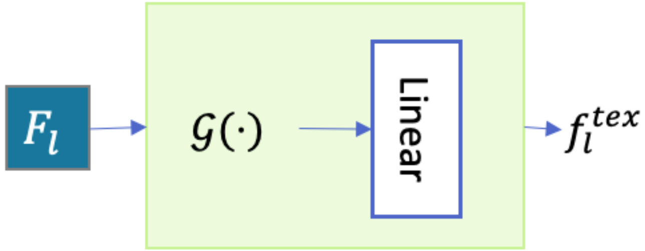
Language Complexity-aware Skip Connection



(a) Language complexity indicator



(b) Structure extractor



(c) Texture extractor

Loss Function

$$\mathcal{L} = \lambda_{adv}\mathcal{L}_{adv} + \lambda_{L1}\mathcal{L}_{L1} + \lambda_{enc}\mathcal{L}_{enc} + \lambda_{rec}\mathcal{L}_{rec} + \lambda_{lcc}\mathcal{L}_{lcc}$$

Adversarial loss

$$\begin{aligned}\mathcal{L}_{adv} &= \mathcal{L}_{advc} + \mathcal{L}_{advs},\\ \mathcal{L}_{advc} &= \max_{D_c}\min_{G}\mathbb{E}_{I_c\in P_c, I_s\in P_s}[\log D_c(I_c) + \log(1 - D_c(\hat{x}))],\\ \mathcal{L}_{advs} &= \max_{D_s}\min_{G}\mathbb{E}_{I_c\in P_c, I_s\in P_s}[\log D_s(I_s) + \log(1 - D_s(\hat{x}))],\end{aligned}$$

L1 loss

$$\mathcal{L}_{L1} = \mathbb{E}_{x,\hat{x}\in P_{(x,\hat{x})}}\ ||x - \hat{x}||_1.$$

Encoder consistent loss

Using two separate encoders: decouple the content and style information of a given font image

$$f_c(I_{c_1}) = f_c(I_{c_2}), \quad f_s(I_{s_1}) = f_s(I_{s_2}),$$

$$\begin{aligned}\mathcal{L}_{enc} &= \mathcal{L}_{enc_c} + \mathcal{L}_{enc_s},\\ \mathcal{L}_{enc_c} &= \mathbb{E}_{I_c}||f_c(I_c) - f_c(x)||_1,\\ \mathcal{L}_{enc_s} &= \mathbb{E}_{I_s}||f_s(I_s) - f_s(x)||_1.\end{aligned}$$

Loss Function

Domain reconstruction loss

To perpetuate the information from the content and style domain

$$\mathcal{L}_{rec_c} = \mathbb{E}_{I_c}||Ic - G(Ic, Ic)||_1,$$
$$\mathcal{L}_{rec_s} = \mathbb{E}_{I_s}||Is - G(Is, Is)||_1,$$
$$\mathcal{L}_{rec} = \mathcal{L}_{rec_c} + \mathcal{L}_{rec_s}$$

Language complexity classification loss

The binary cross-entropy to make the indicator learn the language complexity

$$\mathcal{L}_{lcc_1} = \mathbb{E}_{I_c^{ch}}[log(1 - \gamma(I_c^{ch})] + \mathbb{E}_{I_c^{en}}[log(\gamma(I_c^{en})],$$
$$\mathcal{L}_{lcc_2} = \mathbb{E}_{I_c^{ch}}[log(\gamma(I_c^{cn})] + \mathbb{E}_{I_c^{en}}[log(1 - \gamma(I_c^{en})],$$
$$\mathcal{L}_{lcc} = \mathcal{L}_{lcc_1} + \mathcal{L}_{lcc_2}$$

Agenda

- Introduction
- The proposed method
- Performance evaluation
- Conclusion

Experiments

Dataset

- Chinese and Latin as the training language pair
- Unseen languages: Japanese, Korean, Arabic, Devanagari, Cyrillic, and Thai languages

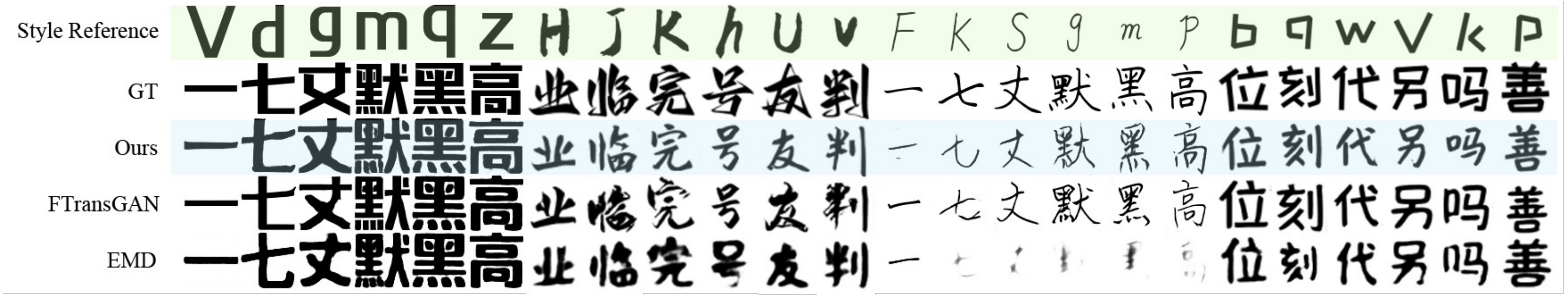
Models for comparison

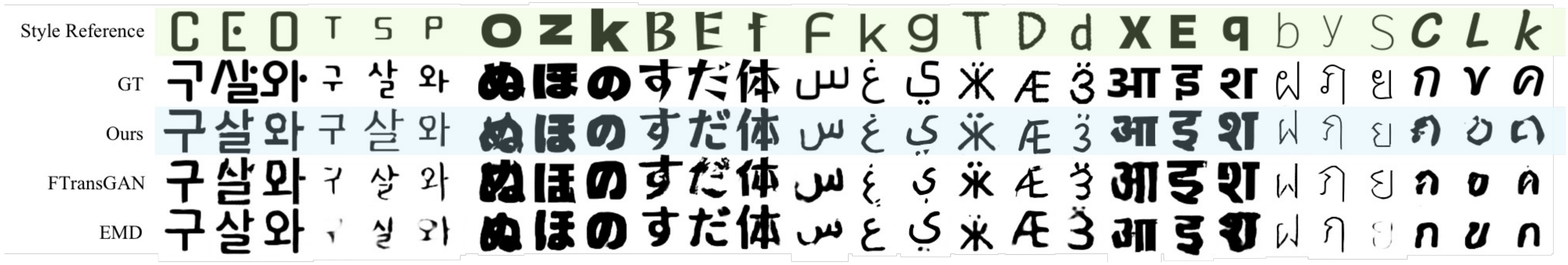
- EMD
- FTransGAN

Evaluation Metrics

- Quantitative: Image Distance (MAE, SSIM), Feature Distance (mFID)
- Visual: Survey
- Latency

Model Evaluation





Model Evaluation

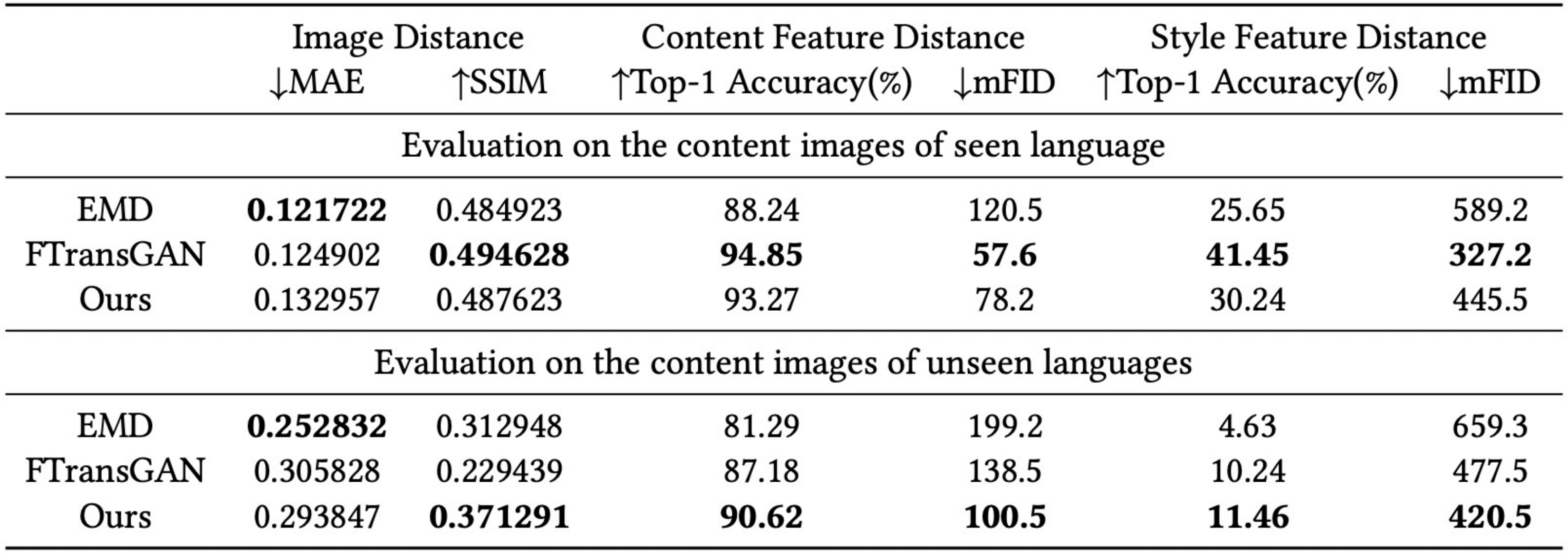
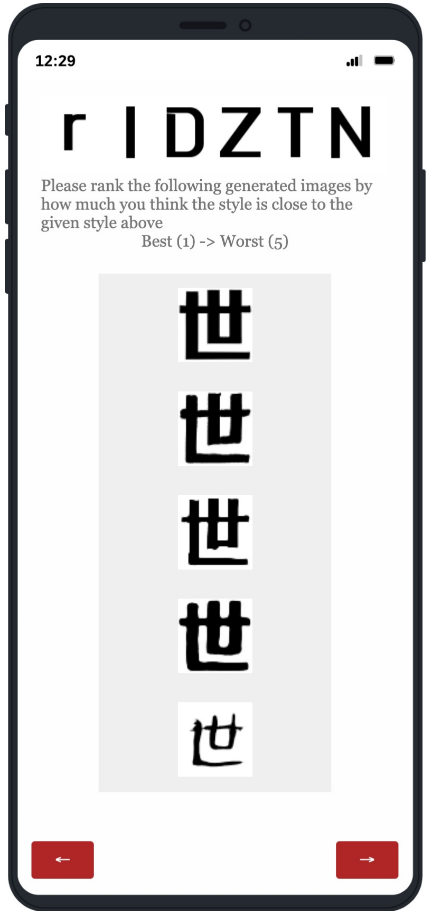


	Image Distance		Content Feature Distance		Style Feature Distance	
	↓MAE	↑SSIM	↑Top-1 Accuracy(%)	↓mFID	↑Top-1 Accuracy(%)	↓mFID
Evaluation on the content images of seen language						
EMD	**0.121722**	0.484923	88.24	120.5	25.65	589.2
FTransGAN	0.124902	**0.494628**	**94.85**	**57.6**	**41.45**	**327.2**
Ours	0.132957	0.487623	93.27	78.2	30.24	445.5
Evaluation on the content images of unseen languages						
EMD	**0.252832**	0.312948	81.29	199.2	4.63	659.3
FTransGAN	0.305828	0.229439	87.18	138.5	10.24	477.5
Ours	0.293847	**0.371291**	**90.62**	**100.5**	**11.46**	**420.5**

Table 1: Quantitative comparison among EMD [23], FTransGAN [15], and the model we propose. ↓ means the lower the better and ↑ means the higher the better. The best value for each comparison is stylized in bold.

Model Evaluation

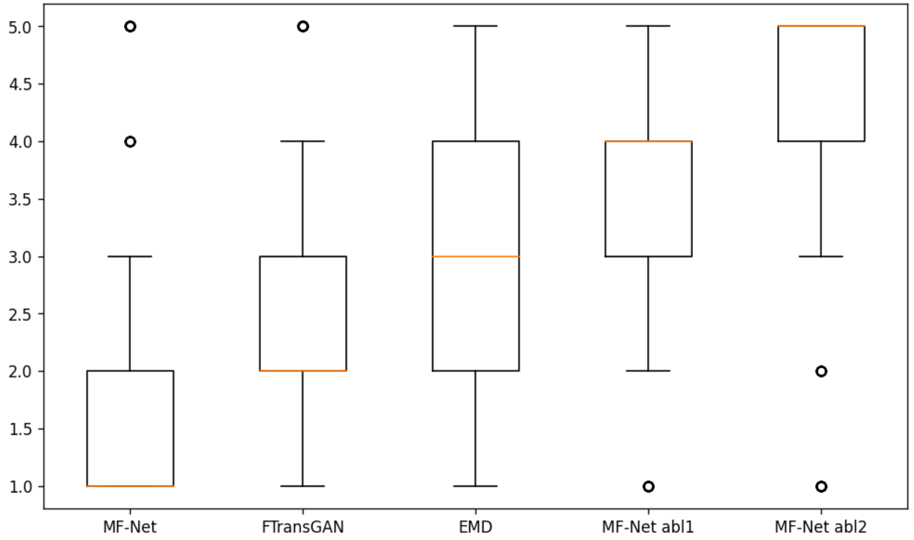


12:29

r I D Z T N

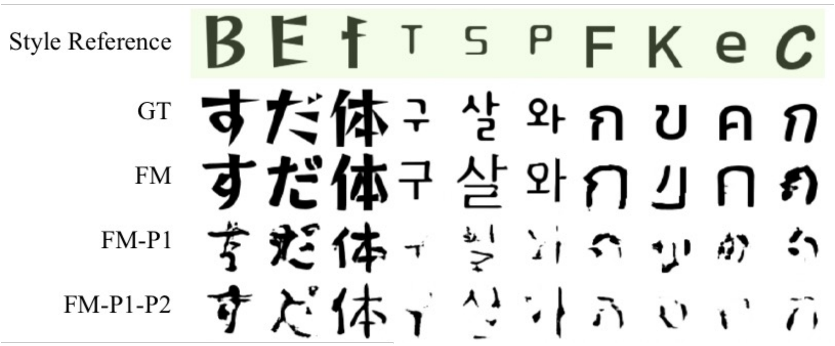
Please rank the following generated images by how much you think the style is close to the given style above

Best (1) -> Worst (5)



MF-Net	35.78% ± 23.4%
FTransGAN	52.51% ± 21.8%
EMD	55.54% ± 25.0%
MF-Net abla1	70.03% ± 20.2%
MF-Net abla2	86.12% ± 22.3%

Ablation Study



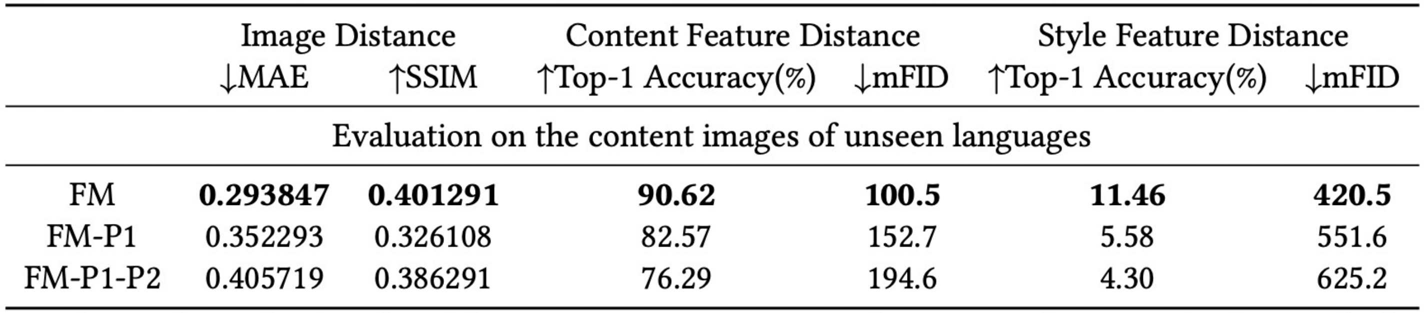


	Image Distance		Content Feature Distance		Style Feature Distance	
	↓MAE	↑SSIM	↑Top-1 Accuracy(%)	↓mFID	↑Top-1 Accuracy(%)	↓mFID
Evaluation on the content images of unseen languages						
FM	**0.293847**	**0.401291**	**90.62**	**100.5**	**11.46**	**420.5**
FM-P1	0.352293	0.326108	82.57	152.7	5.58	551.6
FM-P1-P2	0.405719	0.386291	76.29	194.6	4.30	625.2

Table 2: Ablation Study on the task of stylized font generation on unseen languages. ↓ means the lower the better and ↑ means the higher the better. The best value for each comparison is stylized in bold.

P1: Encoder consistent loss
P2: Language complexity-aware skip connections

Conclusion

Novelties

- Few-shot
- Multilingual

Prospects

- Accelerate the professional font design process
- Generate more copyright-free fonts
- Real-time AR translation

Thank You