One-Shot multilingual Font Generation Via ViT

+ Jiarui Liu + Zhiheng Wang

UMassAmherst | Manning College of Information & Computer Sciences

Overview

- Font design poses unique challenges for languages like Chinese, Japanese, and Korean (CJK), which contains thousands of unique characters.
- Our goal is automatic generate fonts in **multi language**, even for **unseen font**, **unknown and made up characters.**
- We Introduces a Bi-encoder Vision Transformer (ViT) based model with Combined loss pretraining on MAE task. Our model not rely on **base font**.

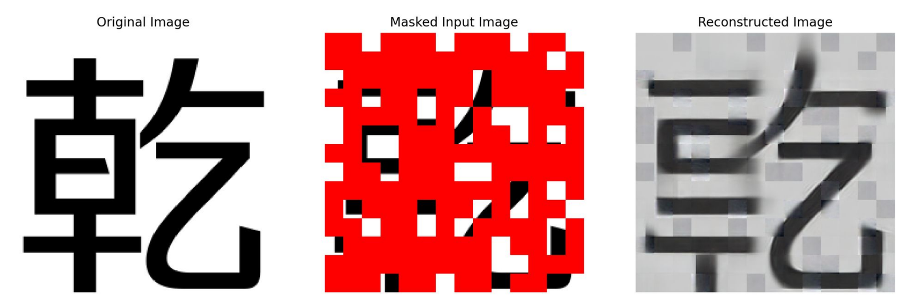
Dataset & Task & Eval


- 308 Fonts in CJK + EN
- 2.2M training samples

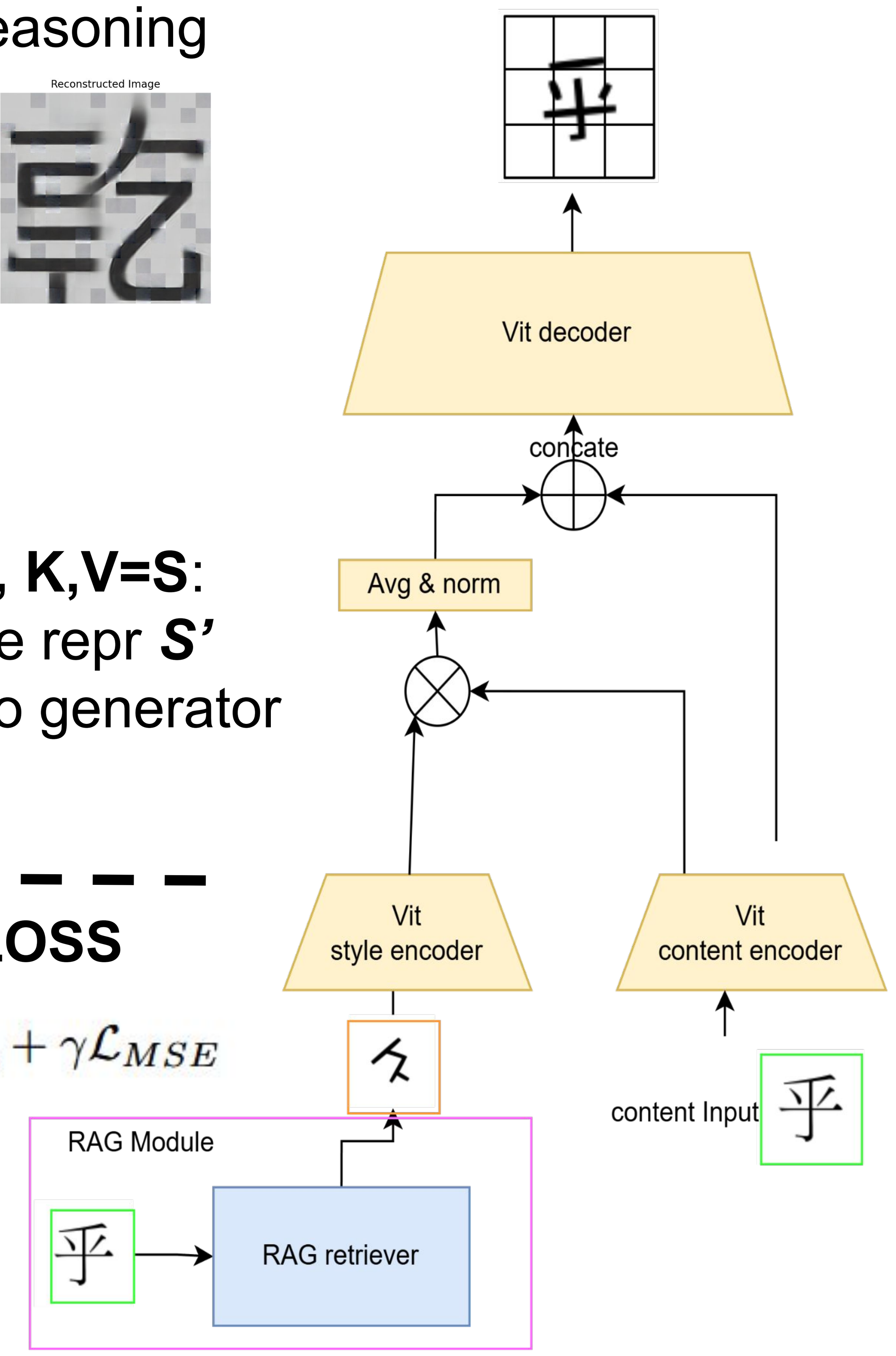
- Evaluation on pixel-level and perceptual metrics

Method

- **Pretrain** encoder & decoder on MAE task for spatial reasoning



- From 2 encoders, get:
  - content repr ***C***
  - Style repr ***S***
- **Cross-attention Q=C, K,V=S**:
  - content specific style repr ***S'***
- **Concate *S'*, *C*,** send to generator
- Get output



---

1. **Train on Combined LOSS**

$$\mathcal{L}_{total} = \alpha\mathcal{L}_{content} + \beta\mathcal{L}_{style} + \gamma\mathcal{L}_{MSE}$$

2. **Refine on L1+MSE**

+ Retriever module

1. Compute image embedding use *content encoder*
2. For each style, build *FAISS* Index for known chars
3. Search most suitable style expamlar during inference

RAG not contribute to eval matrices, but address specific failure cases and improve performance on hard examples

Conclusion

- Our model demonstrates the ability to transfer styles in different languages and even made-up characters.
- Our model achieves better result then DiffusionFont and delivers competitive performance with other approaches, all **while offering enhanced generalizability, usability, and scalability.**

Results

Unseen Settings					Metrics				
SS	SC	CS	CC		L1 Loss ↓	RMSE ↓	SSIM ↑	LPIPS ↓	FID ↓
✓	×	×	×	DG-Font	0.07841	0.2442	0.6853	0.1198	27.98
✓	×	×	×	CF-Font	0.07394	0.2354	0.7007	0.1182	26.51

- Ours also work well on unseen style, **unknow or even self-craft characters**

Unseen Settings				Metrics				
SS	SC	CS	CC	L1 Loss ↓	RMSE ↓	SSIM ↑	LPIPS ↓	FID ↓
✓	×	×	×	0.19858	0.56531	0.65361	0.20315	26.49332
✓	✓	×	×	0.19875	0.56551	0.65507	0.20319	26.37470
×	×	✓	×	0.19812	0.56434	0.65462	0.20121	26.61482
×	×	✓	✓	0.19930	0.56612	0.65337	0.20307	27.13604