JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

Disentangling Writer and Character Styles for Handwriting Generation

Gang Dai*, Yifan Zhang*, Qingfeng Wang, Qing Du, Zhuliang Liu, Shuangping Huang
Email: eedaigang@mail.scut.edu.cn Code: https://github.com/dailenson/SDT
South China University of Technology
NUS National University of Singapore

Handwriting generation

Online handwriting generation is to generate handwritten characters with **controllable content and style**, widely used in **writing robot** and **font design**

Challenges

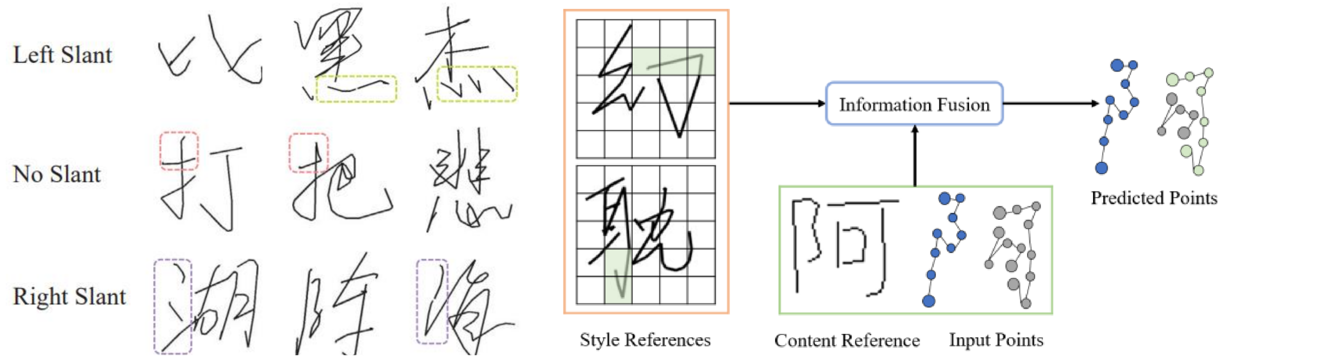
- It is non-trivial to obtain an exact writing style from a limited number of stylized samples
- It is hard to effectively integrate the extracted writing style with specific content for generation

Motivations

Previous RNN-based methods perform poorly:

- Mainly focus on the overall writing style (*e.g.*, *glyph slant*), but neglect the detailed style inconsistencies (*e.g.*, *stroke curvature*) of characters
- Naively concatenate the content and style results in undesired artifacts, *e.g.*, *extra stroke paddings*

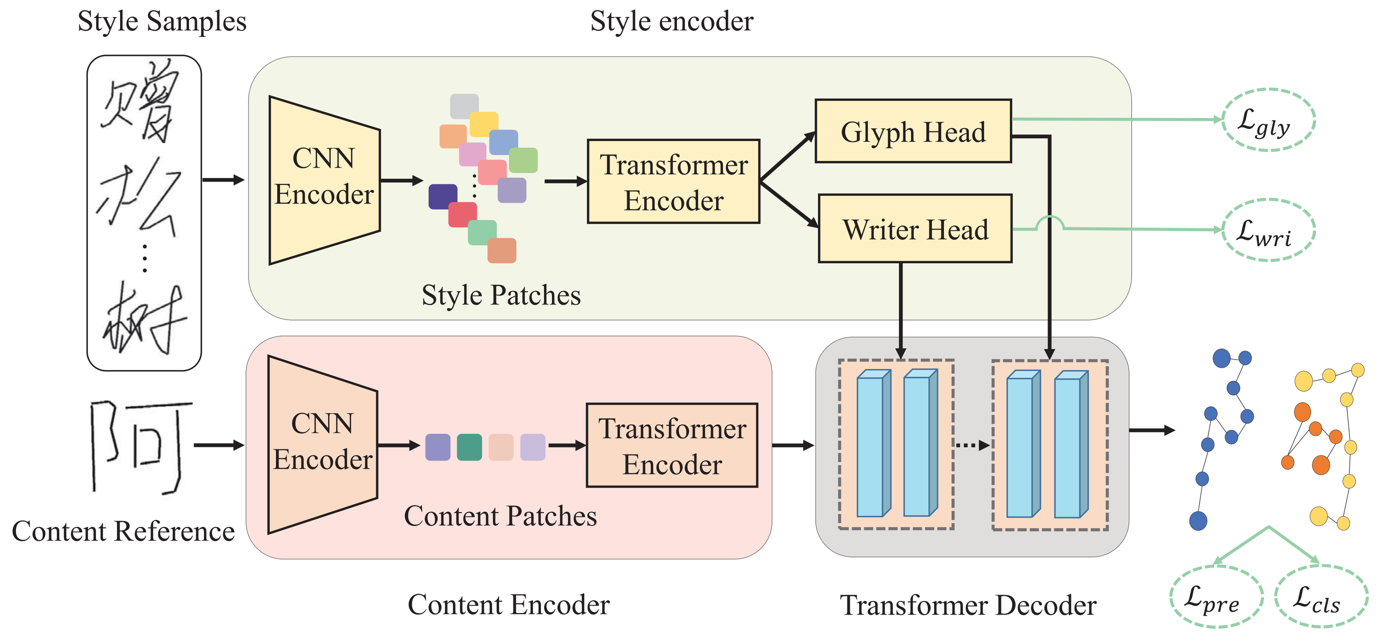
To address these, we disentangle individual handwritings into writer and character styles, and adaptively aggregate these styles with contents



Method Overview

SDT consists of a dual-head style encoder, a content encoder, and a **transformer decoder**

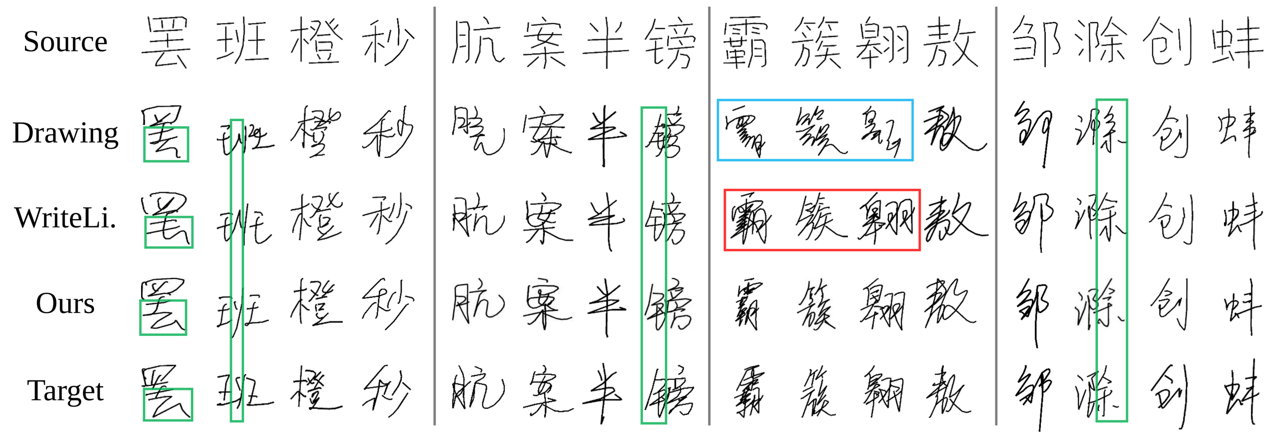
- **The dual-head style encoder** seeks to **disentangle** writer-wise and character-wise style features via complementary contrastive objectives
- The transformer decoder effectively integrates the content and style information with **adaptive information fusion**



Chinese Script Generation

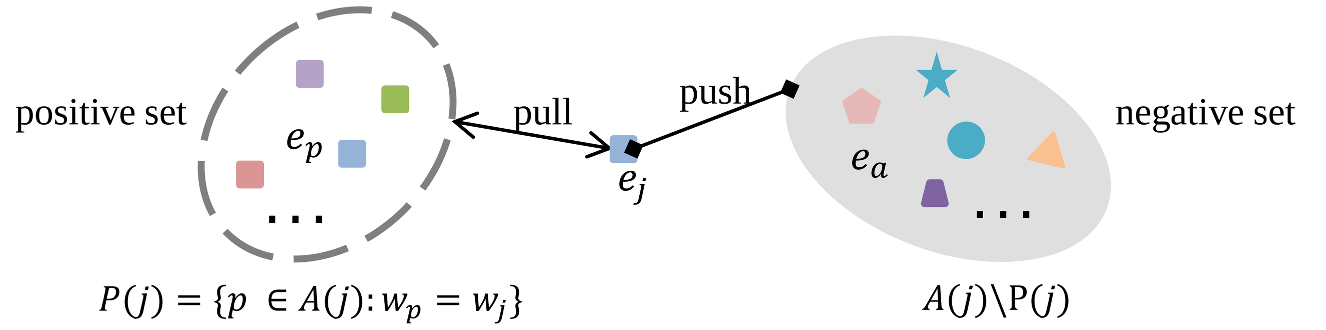
- SDT yields high-quality Chinese handwrittings

Method	Style Score↑	Content Score↑	DTW↓	User Prefer. (%)↑
Drawing	35.83	78.15	1.1813	3.53
FontRNN	46.14	92.18	1.0448	7.07
DeepImitator	50.67	90.92	1.0622	7.99
WriteLikeYou-v1	71.09	93.98	0.9832	11.67
WriteLikeYou-v2	72.37	96.44	0.9289	13.07
SDT(ours)	**94.50**	**97.04**	**0.8789**	**56.67**



Disentanglement of Two style representations

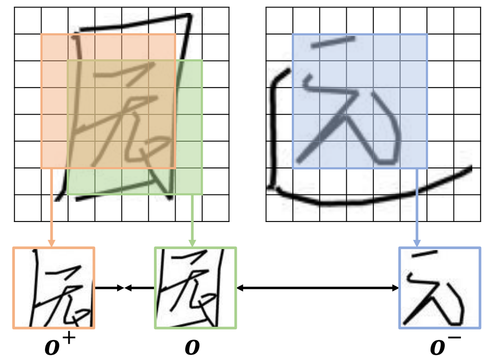
❶ **Writer-wise contrastive learning**:
Align the style features from the same writer



$$\min -\frac{1}{N}\sum_{j\in M}\frac{1}{|P(j)|}\sum_{p\in P(j)}\log\frac{\exp\left(\text{sim}(e_j,e_p)/\tau\right)}{\sum_{a\in A(j)}\exp\left(\text{sim}(e_j,e_a)/\tau\right)}$$

❷ **Character-wise contrastive learning**:

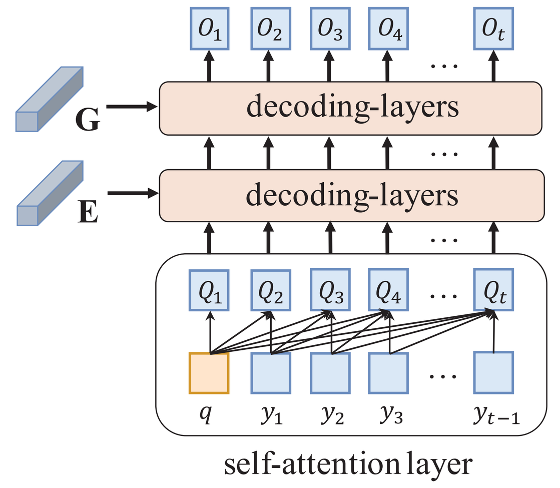
Maximize the mutual information between **diverse views of a character**, thus enforcing learning the detailed style pattern



$$\min -\log\frac{\exp\left(\text{sim}(o,o^+)/\tau\right)}{\exp\left(\text{sim}(o,o^+)/\tau\right)+\sum_{j=1}^{B-1}\exp\left(\text{sim}(o,o_j^-)/\tau\right)}$$

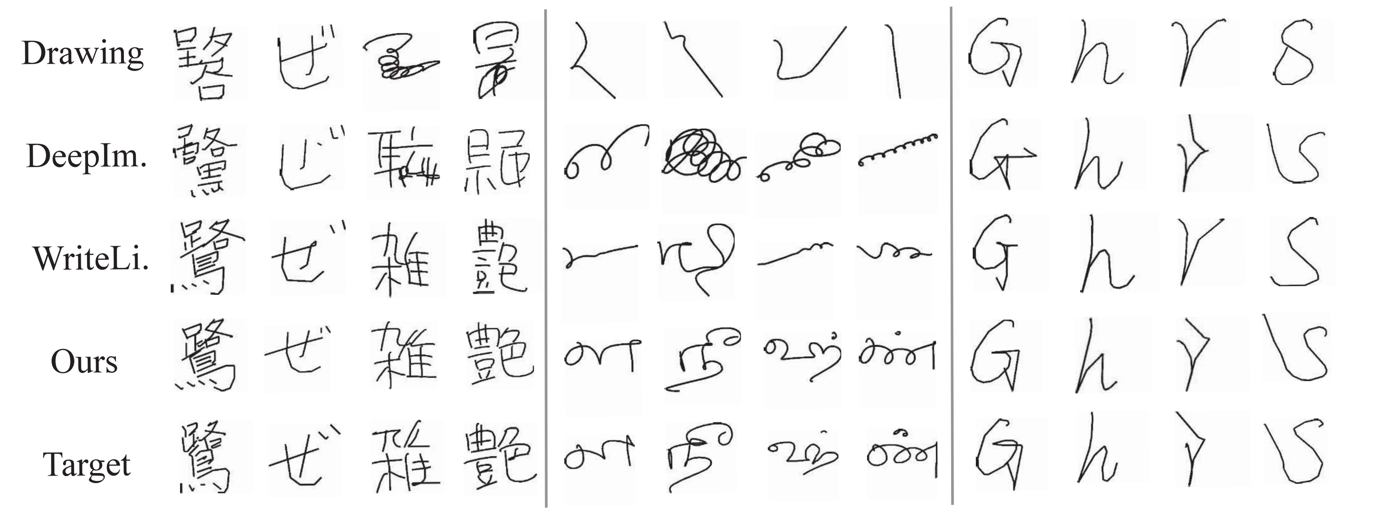
Adaptive Information Fusion

- Combine content feature q with past generated points $[y_1, \ldots, y_{t-1}]$ into the content context at decoding step t
- The context serially attends to the writer-wise style E and character-wise G



Application to other scripts

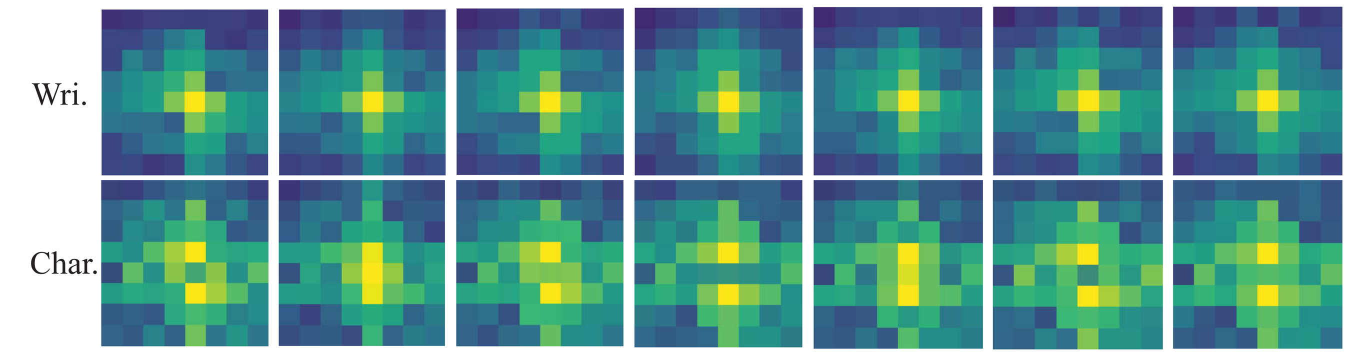
- SDT can generate handwritten characters in **different languages** well



(a) Japanese Script (b) Indic Script (c) English Script

Analysis

- Writer-wise style representations capture more low-frequency information, while character-wise ones capture more high-frequency information



- The writer head imitates the overall style (*e.g.*, *glyph slant*), while the glyph head captures the detailed style (*e.g.*, *stroke curvature*)

writer-wise	character-wise	Generated Samples	Style Score↑	FID↓	DTW↓
		领 解 朝	85.52	27.75	0.8941
✓		领 解 朝	91.38	26.38	0.8841
	✓	领 解 朝	90.31	26.89	0.8803
✓	✓	领 解 朝	94.50	25.46	0.8789
Ground Truth		领 解 朝			