# Beyond Isolated Words: Diffusion Brush for Handwritten Text-Line Generation

Gang Dai*, Yifan Zhang*, Yutao Qin*, Qiangya Guo, Shuangping Huang†, Shuicheng Yan

Email: eedaigang@mail.scut.edu.cn    Code: https://github.com/dailenson/DiffBrush

NUS National University of Singapore

MiroMind

## TEXT-LINE GENERATION

**Handwritten text-line generation** is to produce high-quality handwritten text-line with precise control over both style and content, rather than isolated word generation

Text content    日照香炉生紫烟，遥看瀑布挂前川。

Style sample    参考手写样库

Generated results    日照香炉生紫烟，遥看瀑布挂前川。

## LIMITATIONS OF EXISTING METHODS

- Most focus on isolated handwritten word generation, but neglect alignment and spacing between them
- Some achieve text-line generation but perform poorly:
1) Jointly optimizing content and style from the same model output leads to interference between the two
2) Optimizing content loss at the text-line level may fail to preserve individual character accuracy

## CHALLENGES AND MOTIVATIONS

- Modeling of complex style patterns encompassing both intra- and inter-word relationships is non-trivial
- Ensuring the readability of entire text lines with numerous characters remains challenging

To address these, we propose to: 1) decouple the learning of style and content to enable efficient style representation, 2) enhance content accuracy at both global and local levels.



(a) Style reference    (b) Random masking

(c) Column-wise masking    (d) Row-wise masking
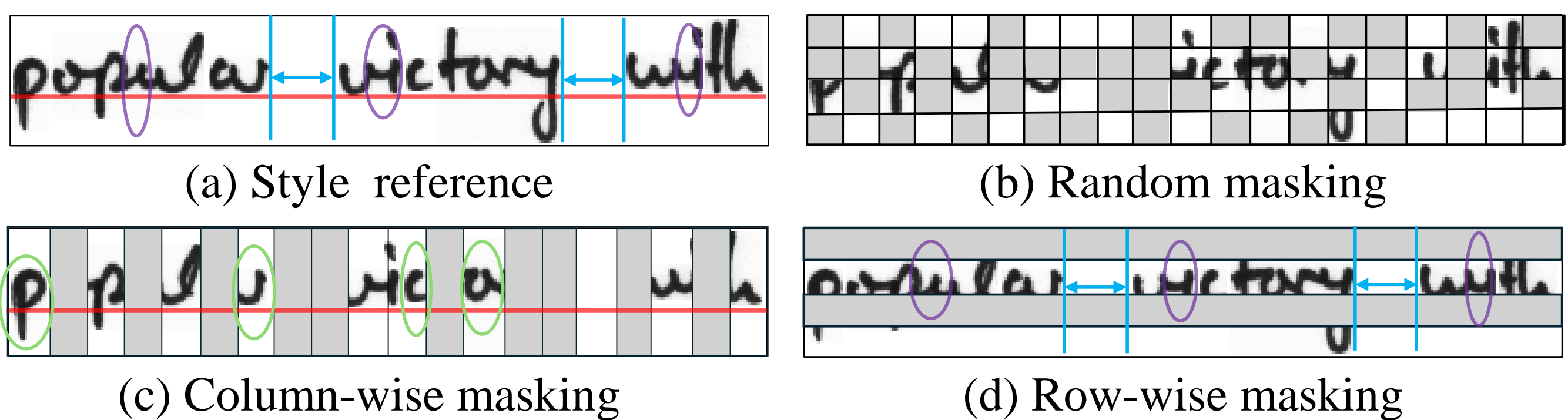
## METHOD OVERVIEW

**DiffBrush** has a **content-decoupled style module**, a **diffusion generator**, and **multi-scale content discriminators**

- The style-enhanced module seeks to **preserve key style patterns** while **disrupting content information** by selectively using two novel masking strategies
- The multi-scale content discriminators provide **finer-grained content supervision** at **both line and local levels** for content adversarial training
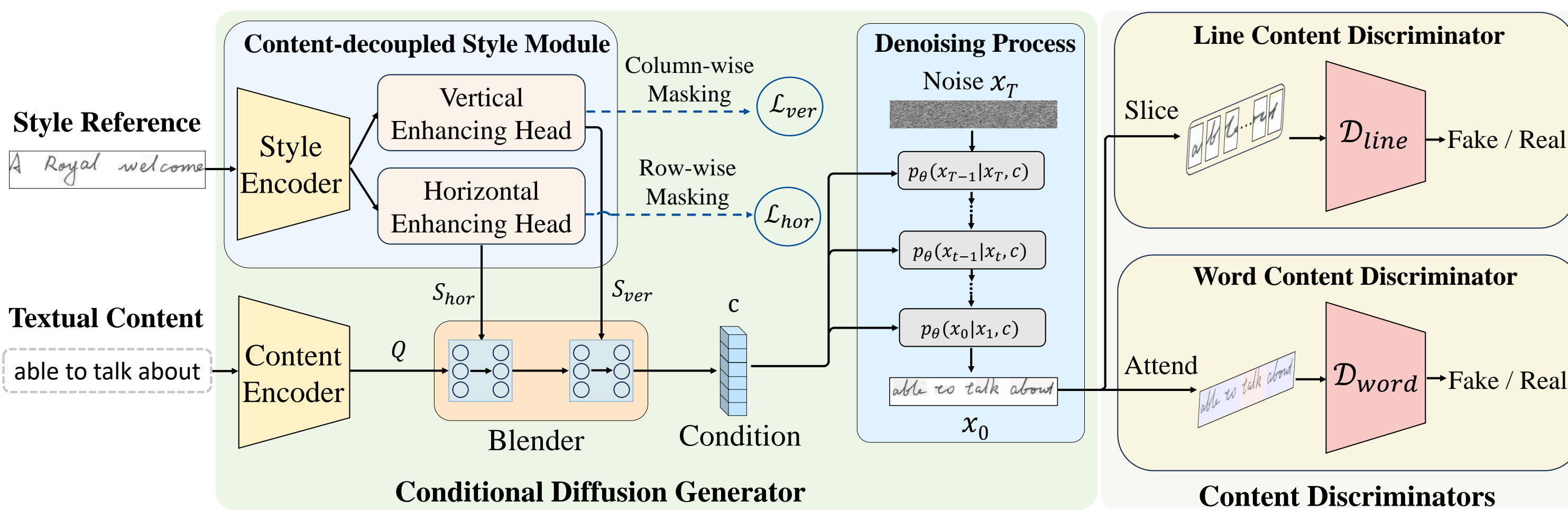


## STYLE ENHANCING BY MASKING



(a) Vertical Style Enhancing via Column-wise Masking



(b) Horizontal Style Enhancing via Row-wise Masking
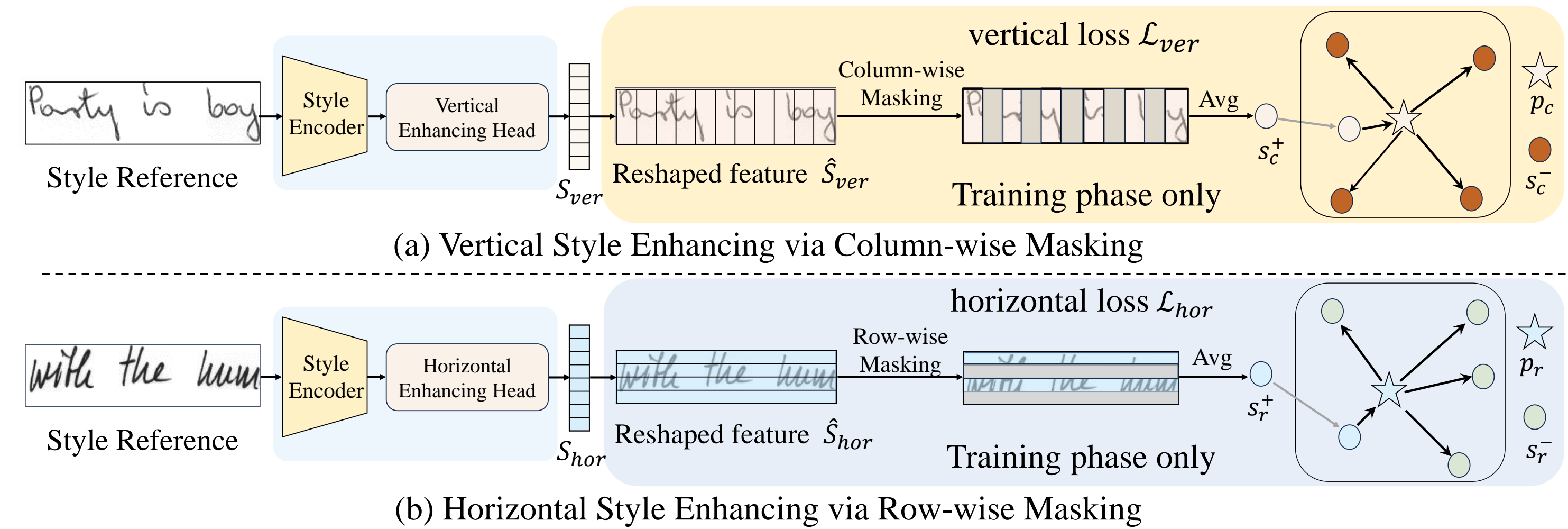
❶ **Vertical style enhancing via column-wise masking**: Assign a proxy per writer to cluster column-masked style features and separate features of different writers

$$\mathcal{L}_{ver} = \frac{1}{|P_c^+|} \sum_{p_c \in P_c^+} \log\left(1 + \sum_{s_c \in S_c^+} e^{-f_c^+}\right) + \frac{1}{|P_c|} \sum_{p_c \in P_c} \log\left(1 + \sum_{s_c \in S_c^-} e^{f_c^-}\right).$$

❷ **Horizontal style learning via row-wise masking**: Pull row-masked style features of the same writer together

$$\mathcal{L}_{hor} = \frac{1}{|P_r^+|} \sum_{p_r \in P_r^+} \log\left(1 + \sum_{s_r \in S_r^+} e^{-f_r^+}\right) + \frac{1}{|P_r|} \sum_{p_r \in P_r} \log\left(1 + \sum_{s_r \in S_r^-} e^{f_r^-}\right).$$

## MULTI-SCALE CONTENT LEARNING

❶ **Line content discriminator**: Determine whether the **overall character order** in $x_0$ matches that in guidance $I_{line}$ to ensure **global content correctness**

$$\mathcal{L}_{line} = \log(\mathcal{D}_{line}(I_{line}, x_{real})) + \log(1 - \mathcal{D}_{line}(I_{line}, x_0)).$$

❷ **Word content discriminator**: Ensure that the text structure is correctly generated, which provides **word-level content feedback** to refine the **local content readability**

$$\mathcal{L}_{word} = \sum_{i=1}^{T} \log(\mathcal{D}_{word}(I_{word}^i, x_{real}^i)) + \sum_{i=1}^{T} \log(1 - \mathcal{D}_{word}(I_{word}^i, x_{word}^i)).$$



(a) Line Content Discriminator    (b) Word Content Discriminator
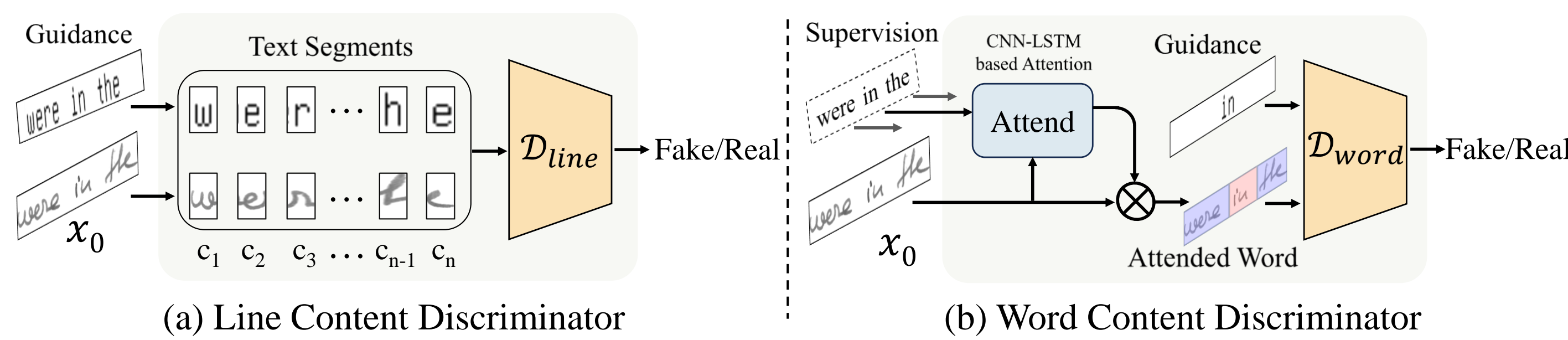
## MAIN RESULTS

- DiffBrush generates superior-quality handwritten text-line images with **readable content** and **consistent style**

| Datasets | Method | Shot | HWD↓ | $D_{CER}$↓ | $D_{WER}$↓ | FID↓ | IS↑ | GS↓ |
|---|---|---|---|---|---|---|---|---|
| IAM | TS-GAN [9] | one | 2.11 | 44.20 | 87.13 | 16.76 | 1.76 | $2.87 \times 10^{-2}$ |
| | CSA-GAN [23] | few | 2.25 | 42.27 | 84.14 | 13.52 | 1.74 | $1.62 \times 10^{-2}$ |
| | VATr [44] | few | 1.87 | 28.80 | 71.77 | 12.51 | 1.69 | $1.45 \times 10^{-2}$ |
| | DiffusionPen [40] | few | 1.72 | 54.75 | 84.70 | 10.24 | 1.83 | $6.42 \times 10^{-3}$ |
| | One-DM [7] | one | 1.80 | 20.91 | 54.27 | 10.60 | 1.82 | $8.42 \times 10^{-3}$ |
| | Ours | one | **1.41** | **8.59** | **28.60** | **8.69** | **1.85** | $2.35 \times 10^{-3}$ |
| CVL | CSA-GAN [23] | few | 1.72 | 41.64 | 72.02 | 8.71 | 1.48 | $6.71 \times 10^{-2}$ |
| | VATr [44] | few | 1.50 | 38.49 | 66.33 | 9.04 | 1.44 | $1.43 \times 10^{-1}$ |
| | DiffusionPen [40] | few | 1.32 | 55.94 | 88.36 | 11.90 | 1.59 | $5.08 \times 10^{-2}$ |
| | One-DM [7] | one | 1.47 | 32.42 | 63.35 | 11.95 | 1.46 | $1.29 \times 10^{-1}$ |
| | Ours | one | **1.06** | **20.92** | **36.38** | **7.57** | **1.70** | $2.96 \times 10^{-2}$ |



## ANALYSES AND DISCUSSIONS

- Vertical style head enhances **vertical word alignment**, while horizontal head improves the **horizontal spacing**

| Style module in DiffBrush | | HWD↓ | $D_{CER}$↓ | $D_{WER}$↓ |
|---|---|---|---|---|
| Single style encoder | *This is a difficult question* | 1.78 | 11.34 | 36.29 |
| Ver. enhancing head only | *just dusty-grey but muddy* | 1.63 | 10.92 | 35.04 |
| Hor. enhancing head only | *just dusty-grey but muddy* | 1.58 | 10.66 | 33.64 |
| Our style module | *just dusty-grey but muddy* | **1.41** | **8.59** | **28.60** |

- Our discriminators enhance content readability while **preserving style mimicry performance**

| Style sample | *she had been sufficiently* | HWD↓ | $D_{CER}$↓ | $D_{WER}$↓ |
|---|---|---|---|---|
| Base+$\xi_{style}$ | *gave a bugh cough raucous* | 1.47 | 54.64 | 44.33 |
| Base+$\xi_{style}$+$\mathcal{D}_{word}$ | *gave a long raucous cough cough* | 1.42 | 14.61 | 43.93 |
| Base+$\xi_{style}$+$\mathcal{D}_{line}$ | *gave a raucous cough* | 1.44 | 15.28 | 43.31 |
| Base+$\xi_{style}$+$\mathcal{D}_{line}$+$\mathcal{D}_{word}$ | *gave a long raucous cough* | **1.41** | **8.59** | **28.60** |

- CTC recognizer harms the learning of style and leads to **easily recognizable outputs with simplified styles**

| Style sample | *advantages the codex was fragile and* | HWD↓ | $D_{CER}$↓ | $D_{WER}$↓ |
|---|---|---|---|---|
| Base+$\xi_{style}$+CTC | *'We know much' repeated the Chief* | 1.67 | 16.53 | 50.48 |
| Base+$\xi_{style}$+$\mathcal{D}$ | *'We know much' repeated the Chief* | **1.41** | **8.59** | **28.60** |

## APPLICATION TO OTHER LANGUAGES

- DiffBrush generates **Chinese** handwritten text-lines well

| Text content | 武和平19日在公安部召开的新闻发布会上透露 |
|---|---|
| Style sample | 实行居住地户口登记制度的改革目标，这一 |
| One-DM | 武和平19日在公安部召开的新闻发布会上透露 |
| Ours | 武和平19日在公安部召开的新闻发布会上透露 |

| Text content | NASA当天在华盛顿国家航空航天博物馆门前展示 |
|---|---|
| Style sample | 姆斯韦布太空望远镜的同尺寸模型，这一大型望远镜 |
| One-DM | NASA当天在华盛顿国家航空航天博物馆门前展示 |
| Ours | NASA当天在华盛顿国家航空航天博物馆门前展示 |