BengaliDiff: Diffusion Model for One-Shot Bengali Font Generation

Md Bilayet Hossain[1,2], Honghui Yuan[1], Shabnur Anonna Akhy[1,2] and Prof.Keiji Yanai[1]

Department of Informatics, The University of Electro-Communications, Tokyo, Japan[1]
Department of Computer Science and Engineering, Daffodil International University, Dhaka, Bangladesh[2]

国立大学法人
電気通信大学
The University of Electro-Communications

YANAI Lab

Motivation

- Bengali is spoken by over 200 million people but has very **few high-quality fonts**.

- Existing font generation models (like for Chinese or Latin) don’t work well for Bengali.

- Creating Bengali fonts by hand is **slow and complex.**

- We aim to build an automatic system that can **generate Bengali fonts** from just a few examples.

Challenges

- **Challenge:** Existing models (e.g. **FontDiffuser** [1]) fail to generate on Bengali font.
- **FontDiffuser** struggles with:
  - Distorted matras and ligatures
  - Inconsistent stroke thickness
  - Style mismatch and loss of detail

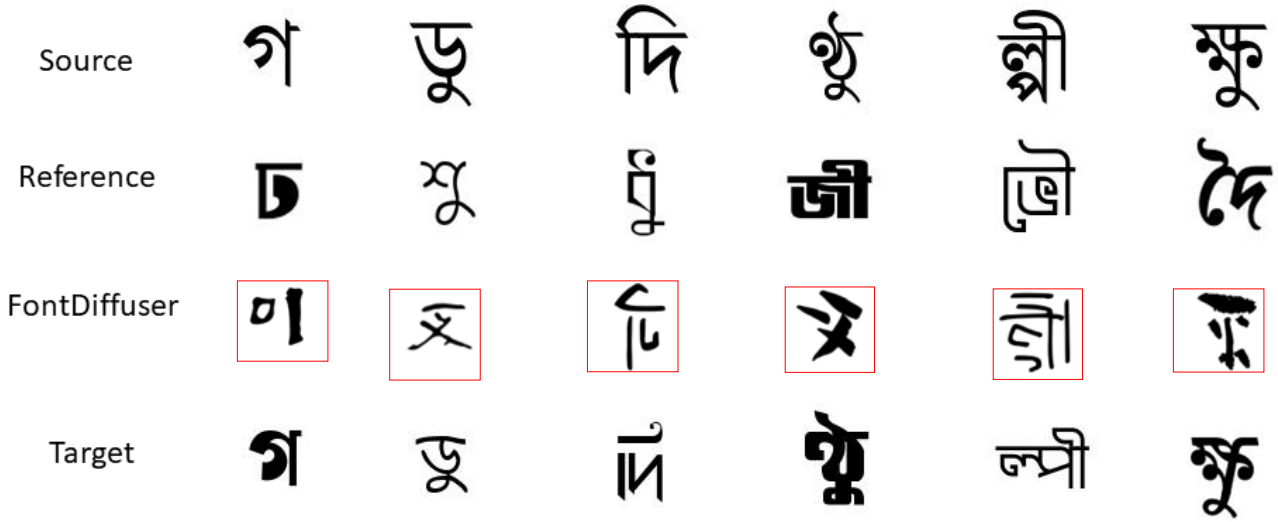

Figure 1: The results generated by FontDiffuser using Bengali font as the input image.

References: 1. Yang et al. (2024) — *FontDiffuser: One-shot font generation using diffusion with content aggregation and style contrastive learning*. AAAI Conf. on Artificial Intelligence, Vol. 38, pp. 6603–6611.

Base Architecture: FontDiffuser

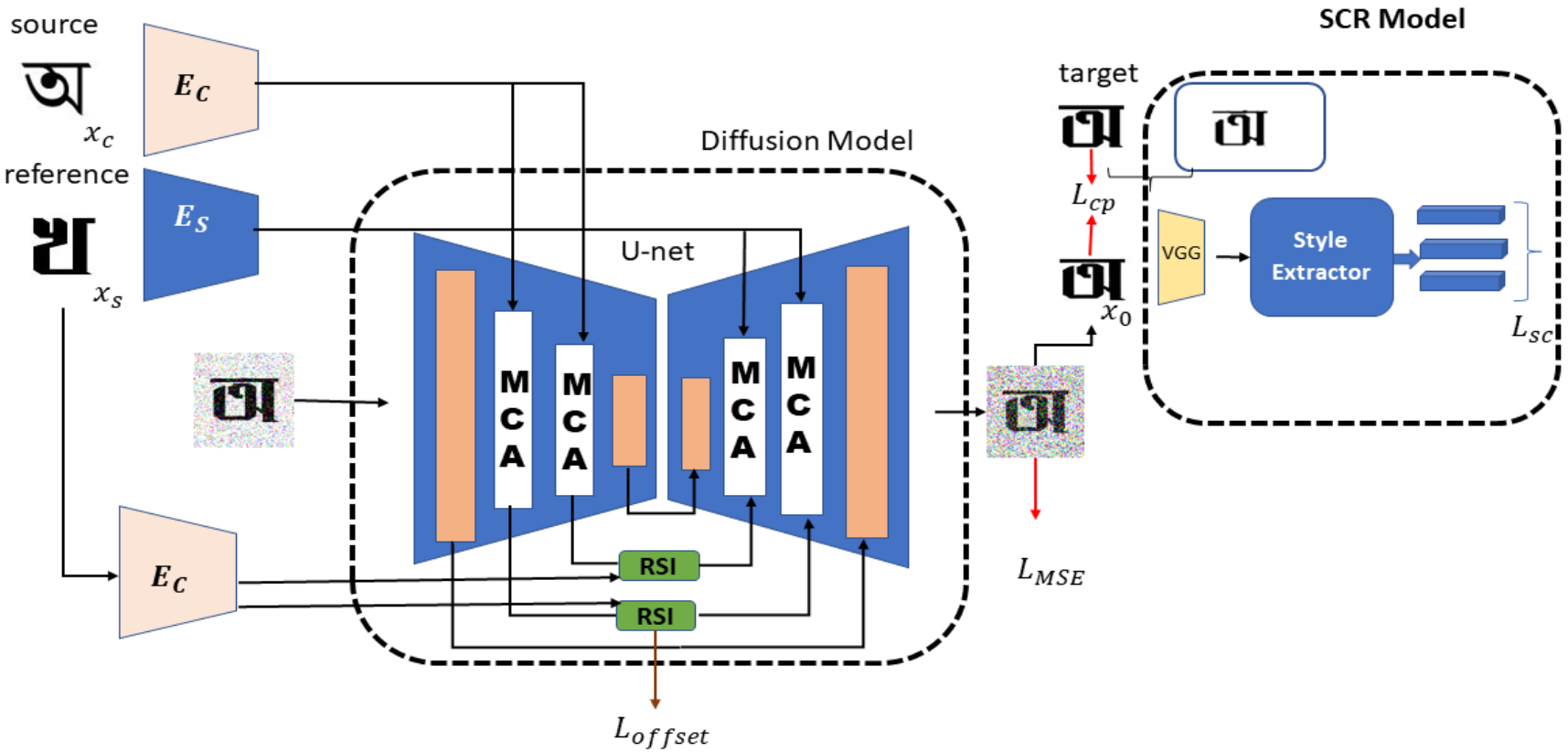

Figure 2: The framework of the base method FontDiffuser [1] .

- **FontDiffuser uses U-Net with:**
  - Content Encoder (structure)
  - Style Encoder (appearance)

- **Two key modules:**
  - MCA: Multi-Scale Content Aggregation
  - RSI: Reference Structure Interaction

References: 1. Yang et al. (2024) — *FontDiffuser: One-shot font generation using diffusion with content aggregation and style contrastive learning*. AAAI Conf. on Artificial Intelligence, Vol. 38, pp. 6603–6611.

Our Proposed Method “BengaliDiff: Overview”

- **First diffusion-based model** tailored for Bengali font generation.

- Our Key innovations:

- Cross-Attention Content Fusion
- Patch-level Adversarial Discriminator

- Synthesizes high-quality fonts from few references

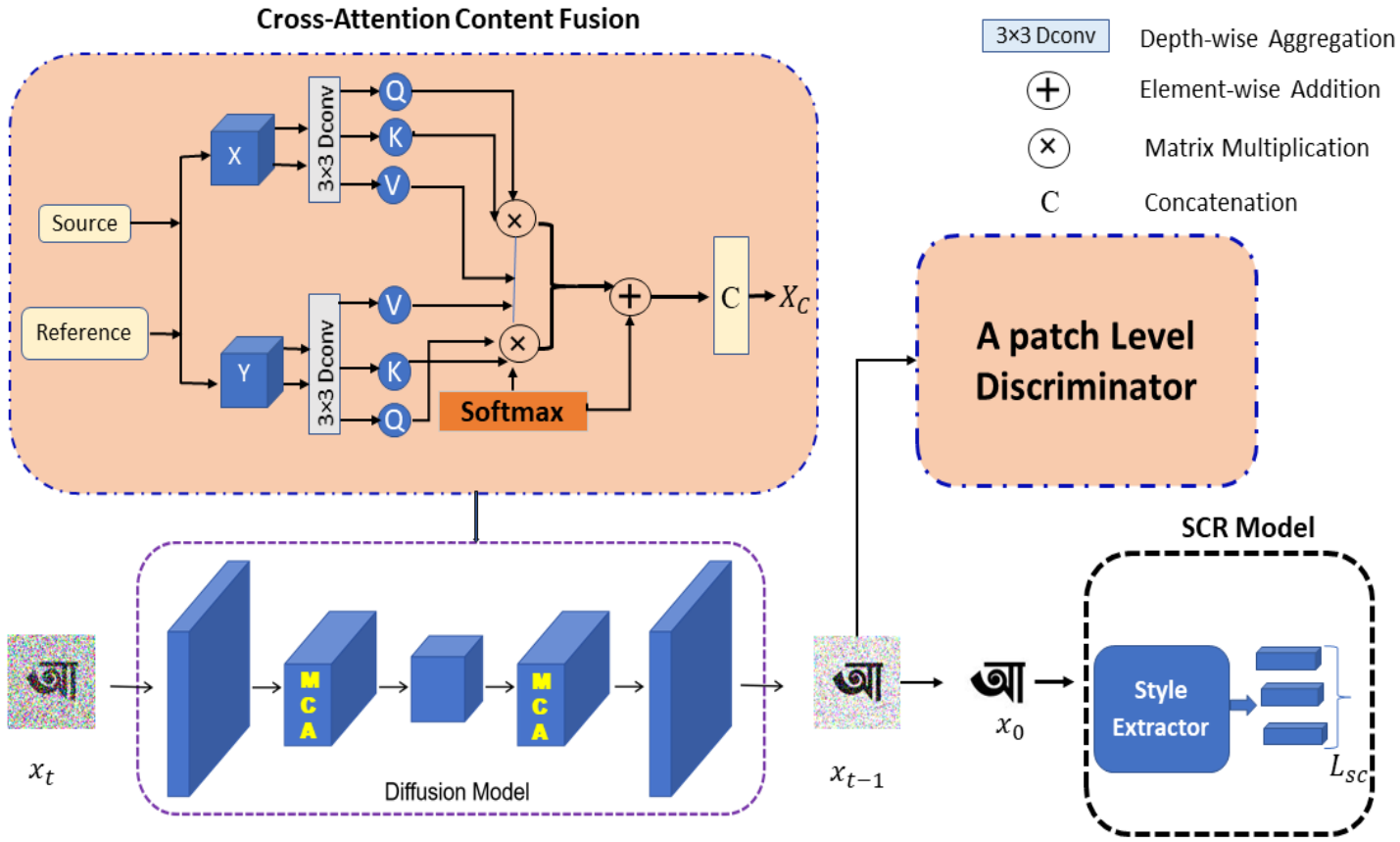

Figure 3: Overview of our proposed method. The Cross-Attention Content Fusion and A Patch Level Discriminator in the blue dashed line boxes.

Qualitative Results Comparison with FontDiffuser and Our Method

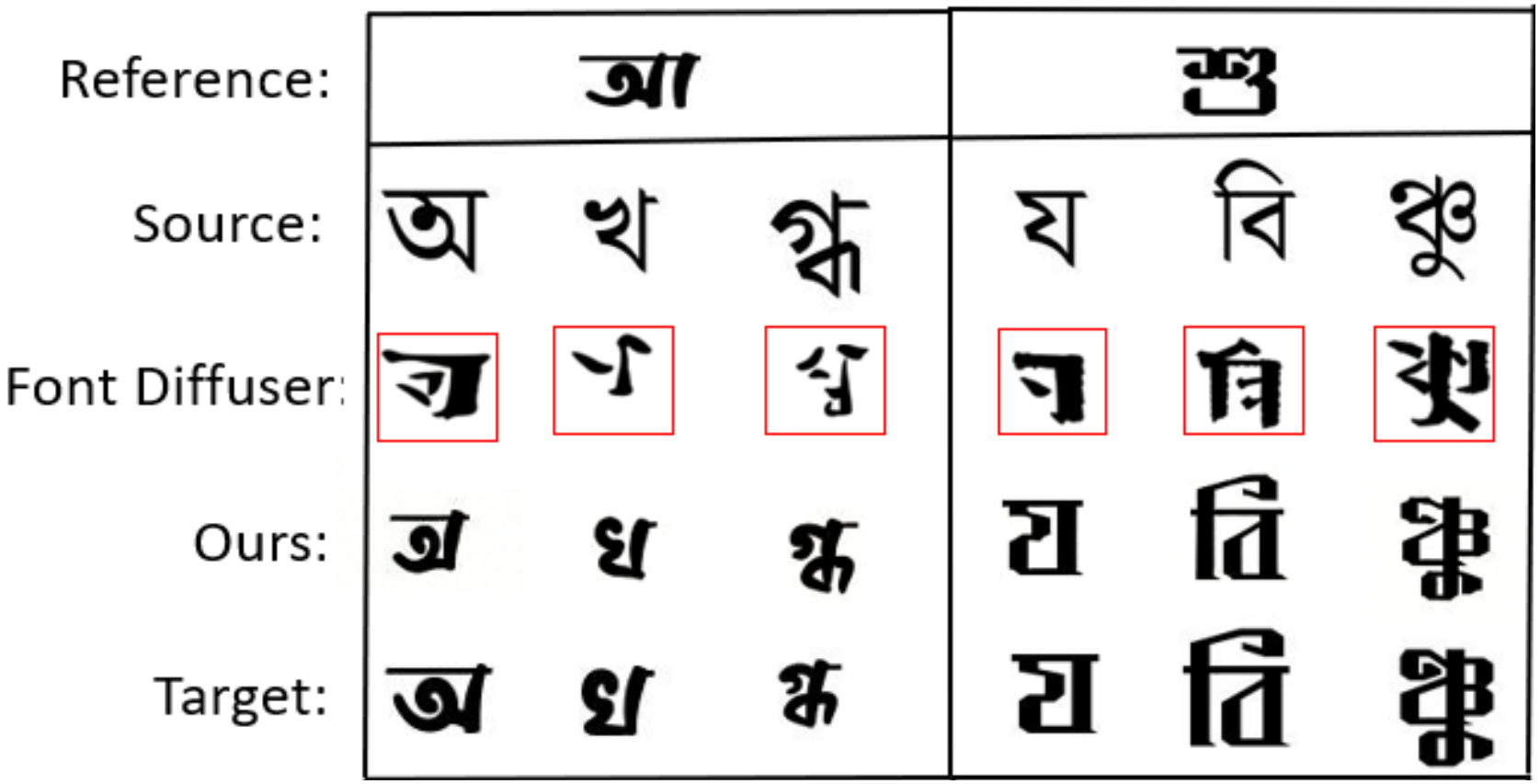

Figure 4: Qualitative Results Comparisons on SFUC between our method and the previous state-of-the-art method of FontDiffuser.

Seen Font Unseen Characters:

- BengaliDiff **outperforms** FontDiffuser
- Achieves **better content preservation**
- **Sharper stroke** details
- More accurate **style transfer**

Qualitative Results Comparison with FontDiffuser and Our Method

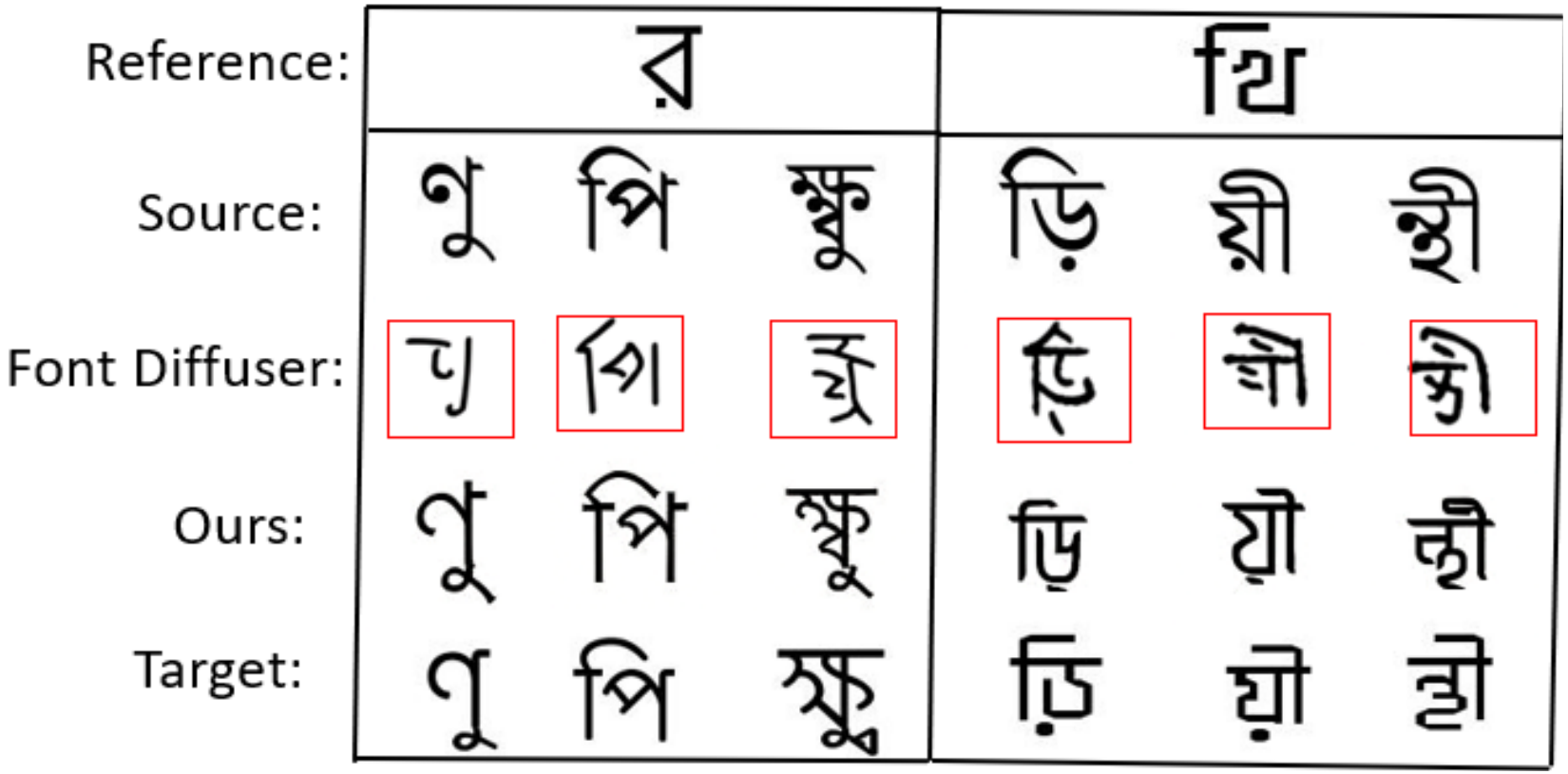

Figure 5: Qualitative Results Comparisons on UFUC between our method and the previous state-of-the-art method of FontDiffuser.

Unseen Font Unseen Characters:

- BengaliDiff maintains **structural accuracy**
- **Style transfer** can still be slightly

Quantitative Results Comparison with FontDiffuser and Our Method

Table 2:Quantitative results Comparisons with FontDiffuser and BengaliDiff

Model	FID ↓	LPIPS ↓	L1 ↓	RMSE ↓	SSIM ↑
FontDiffuser (SFUC)	0.8685	0.3568	**0.2377**	**0.3151**	**0.7131**
BengaliDiff (SFUC)	**0.7063**	**0.3223**	0.2547	0.3240	0.6751
FontDiffuser (UFUC)	0.9573	0.3738	**0.2419**	**0.3142**	**0.6855**
BengaliDiff (UFUC)	**0.7280**	**0.3406**	0.2706	0.3357	0.6491

- BengaliDiff achieves **lower FID and LPIPS** scores, indicating more realistic and perceptually similar images.

Ablation Study

- Tested 3 configurations:

- Baseline (FontDiffuser)
- + Cross-Attention (CA)
- + CA + Discriminator (Full BengaliDiff)

- Each module improves **visual fidelity and structure**

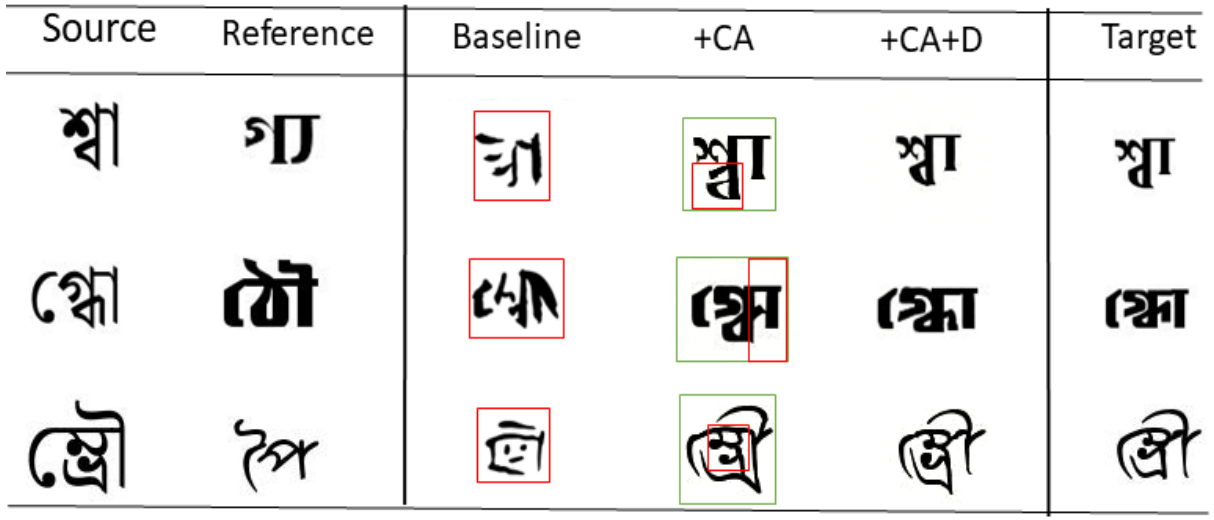

Figure 6: Qualitative evaluation results of ablation studies. An illustration of several modules. CA and D represent Cross-attention and Discriminator, respectively. Red boxes represent the missing strokes, while green represents the corresponding improvements.

Conclusion

- **BengaliDiff** generates high-quality Bengali fonts from few samples.
- Preserves **complex character structure** and ensures **style consistency.**
- **Cross-attention and discriminator** improve style and structure.
- Outperforms FontDiffuser in seen font test sets.
- Useful for **OCR, typography, and font design.**

Limitations & Future Work

Limitations

- Trained only on digital fonts
- Poor generalization to artistic/handwritten styles
- Limited control over font style

Future Work

- Add support for handwritten glyphs
- Improve style transfer to unseen
- Enable editable style vectors
- Support dynamic ligatures

Thank you

Question & Answer

Please Feel Free to Ask Question

Email: h2495009@gl.cc.uec.ac.jp