Enhancing GAN-Based Handwriting Generative Model: Handwriting Feature Extraction through LSTM and Transformer

Group 7
Lu Li 121090272
Yiqu Yang 121090711
Jingxuan Wu 121090607
Caijun Wang 121090532

School of Data Science,
The Chinese University of Hong Kong, Shenzhen

April 25, 2024

Outline

Background

- Individual handwriting variations make feature extraction crucial for imitation and enhancing recognition and signature authentication.
- However, there are two main challenges:
  1. Annotated handwriting datasets with varied styles are **labor-intensive** to acquire.
  2. Individual variability in **calligraphic styles** like character shape, stroke thickness, writing slant, and ligature is **difficult to be represented in data**.
- Facing the above two challenges,
  1. Scholars usually use **the largest handwriting dataset IAM[1]** for training their generative models.
  2. **CNN-base style encoders** are typically used to extract features from handwriting images (e.g. HiGAN[2], TextStyleBrush[3], GANwriting[4]).

Objective

To enhance the performance of handwriting feature extraction, our objectives include:

1. **Enriching the dataset through an automated processing pipeline.** This approach not only saves labor and time but also rapidly acquires a wealth of annotated handwriting word images essential for training models.
2. **Experimenting with alternative frameworks for the style encoder to optimize handwriting feature extraction.** This allows us to explore innovative methods to enhance feature accuracy and adaptability, potentially leading to more robust and versatile handwriting analysis models.

Outline

Data Collection Process

- **IAM dataset[1]**: created by having around 400 participants handwrite sections from the LOB corpus onto forms. These forms were then scanned to produce the dataset.
- **Our pipeline**: Pros & Cons:
  1. **Automated labeling**.
  2. Efficient image processing process.
  3. Scalable to large data sets.
  4. Primary limitation: demands manual intervention to correct OCR mislabeling, particularly when high accuracy is critical.

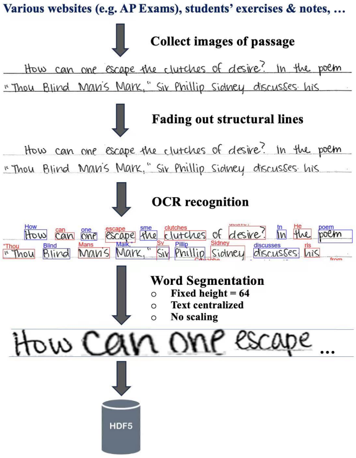


Figure: Our Data Collection Pipeline

Overview of Two Datasets

- **IAM dataset[1]**: Selected data contains 63401 word-level images from 500 writers.
- **Our dataset**: **22514** word-level images from **385** writers.
- We **merge** the IAM dataset with ours for training GAN models.

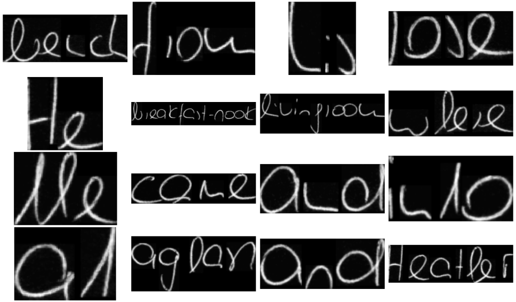


Figure: IAM dataset

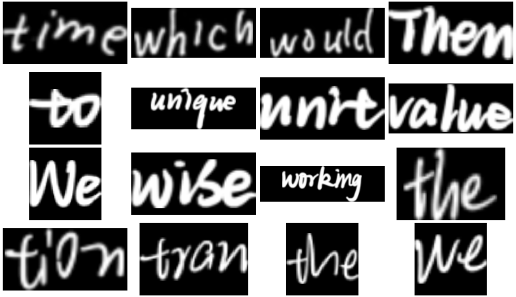
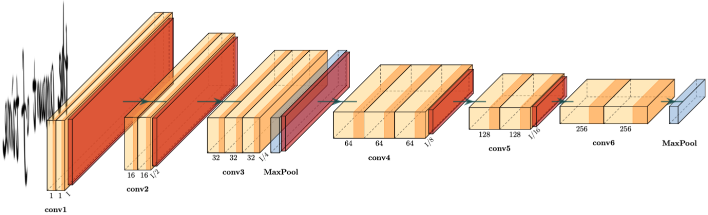


Figure: Our dataset

Outline

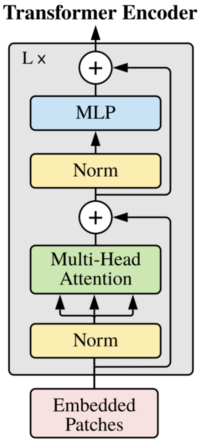
Methodology

- **Traditional CNN feature extraction network**: Use convolution to analyse and extract the local features, and usually constructs a series of residual blocks that allow for deeper networks by enabling training without severe degradation in performance.
- **Shortcomings of pure CNN**: hard to correlate length-variant data
- **Our Solution**: Experiment the RNN-based method **LSTM** and the more advanced structure **Transformer** to extract content-independent features from handwriting



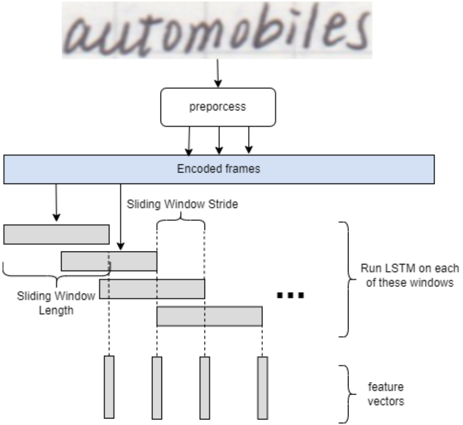
Methodology

- **Vision Transformer**: Using Transformer to learn the features of the images.
- **Conv Transformer**: Also using the same blocks, wanting to act as a part of Recognizer.
- **Transformer Block Framework**:
  - Embedded Patches: The input image is converted into a series of embedded vectors, representing small patches in the image.
  - Layer Norm: Stabilize the training process and speed up convergence to prevent gradient explosion.
  - Multi-head: Using multi-head self-attention to learn the features.
  - Add: The same principle as Resnet.
  - MLP: Increase the nonlinear processing capability of the model.



Methodology

- **RNN Structure: LSTM**
  In Long Short-Term Memory (LSTM) networks, the hidden state is an important component that captures information about the sequence processed so far.
- **Input**
  Encoded frames were obtained from the preprocessing of the handwriting sample (in our experiment, a simple CNN network).
- **Feature representation**
  Take the hidden state of the LSTM network of the last frame as the feature vector.



Outline

Experiment

- We tested our feature extraction model on a writer classification task.
  1. HiGANplus(2022)[2] + IAM dataset
  2. HiGANplus(2022)[2] + IAM & Our dataset
  3. Our Feature Extraction Model + IAM & Our dataset

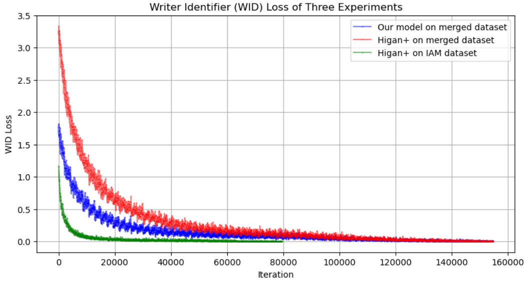


Figure: Accuracy Results

Method	Accuracy
Our model on merged dataset	91.15%
Higan+ on merged dataset	90.07%
Higan+ on IAM dataset	87.08%

Table: Comparison of methods

Experiment

- Apply to Handwriting Generative Model Based on GAN
  - Pipeline
    - Style Encoder & Writer Identifier: **Our feature extraction network**
    - Generator & Discriminator: Generate images according to the input letters and the style, then distinguish between real and fake handwriting images. GAN part structure modified from HiGAN+[2].
    - Recognizer: OCR module, evaluating the accuracy of the generated image.
  - Training
    - Pre-train Writer Identify and Recognizer on our data set.
    - Discriminator and Generator act the same roles in GAN and solve the minimax problem.

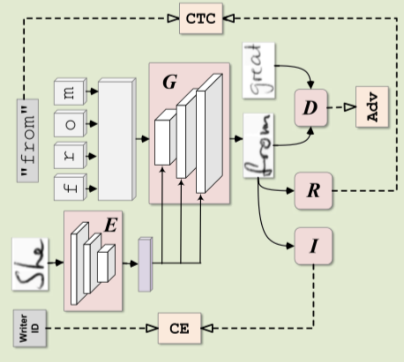
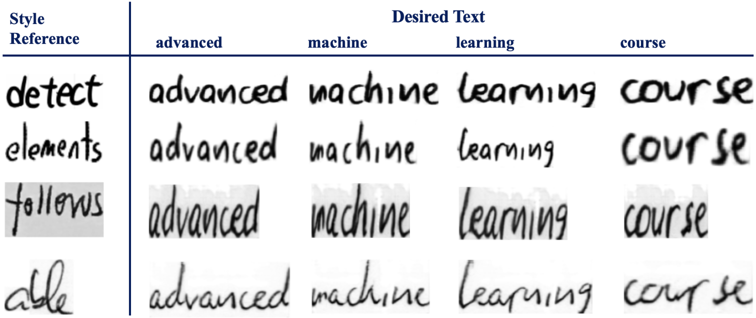


Figure: Part of HiGAN+ Framework

Experiment

- Representative generated results:



Remark: These results are preliminary and subject to further validation

Outline

Conclusion

Completed Work

- With our OCR-based pipeline, we successfully collected tens of thousands of word-level annotated images, enriching the existing handwriting dataset.
- With the LSTM-based style encoder, the modified HiGAN+ model successfully generated realistic handwriting images with desired calligraphic styles.

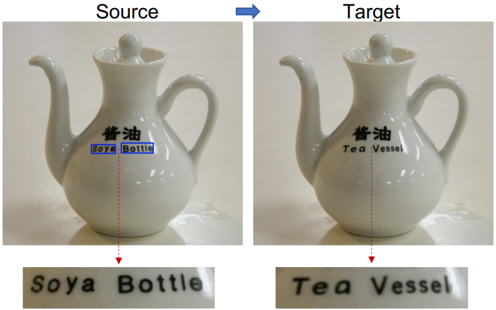
Ongoing Work

- Augment the epoch count, thereby facilitating deeper convergence.
- Refine the Transformer architecture, concomitantly delving into its interpretative facets to unravel underlying mechanisms.

Outline

Future Work

- Extract writing styles from images with intricate background details, facilitating text transfer.



- Retain the RGB information of images when extracting features of handwriting text or even scene text images.

References

[1] Marti ZU-V and Bunke Horst. The iam-database: An english sentence database for offline handwriting recognition. *International Journal on Document Analysis and Recognition*, 5(1):39–46, 2002.

[2] Ji Gan, Weiqiang Wang, Jiaxu Leng, and Xinbo Gao. Higan+: Handwriting imitation gan with disentangled representations. *ACM Trans. Graph.*, 42(1), 2022.

[3] Praveen Krishnan, Rama Kovvuri, Guan Pang, Boris Vassilev, and Tal Hassner. Textstylebrush: Transfer of text aesthetics from a single example, 2021.

[4] Lei Kang, Pau Riba, Yaxing Wang, Marçal Rusiñol, Alicia Fornés, and Mauricio Villegas. Ganwriting: Content-conditioned generation of styled handwritten word images, 2020.


Thank you
Thank you
Thank you
Thank you
thank you
thank you
thank you
thank you