Learning Generative Structure Prior for Blind Text Image Super-resolution

Xiaoming Li[1]   Wangmeng Zuo[2]   Chen Change Loy[1]

[1]Nanyang Technological University, Singapore   [2]Harbin Institute of Technology, Chine

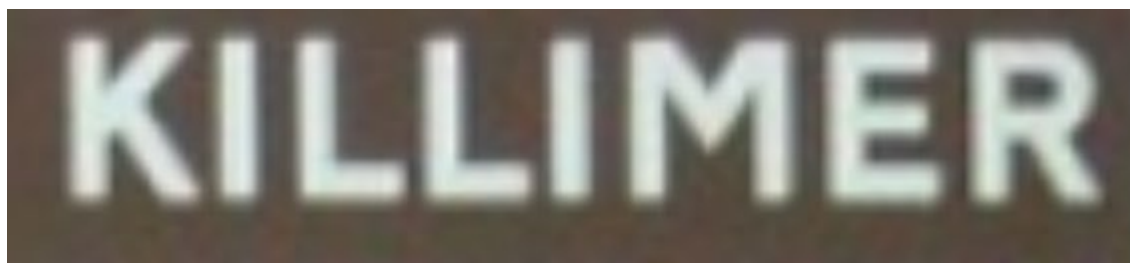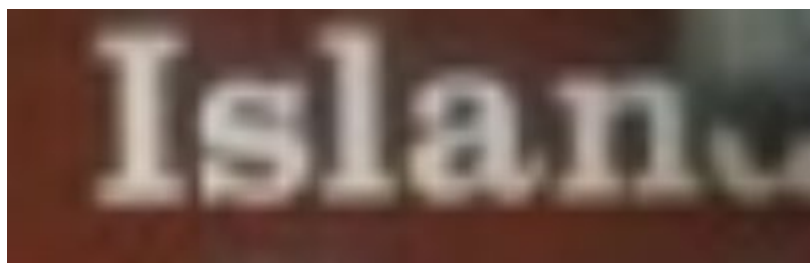Project Page: https://github.com/csxmli2016/MARCONet
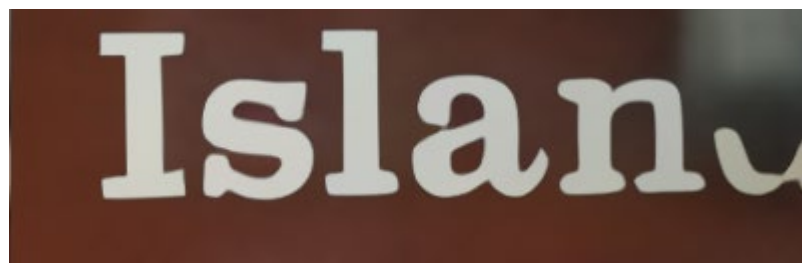
WED-AM-179

Blind text image super-resolution

- If the structure of the character is simple:

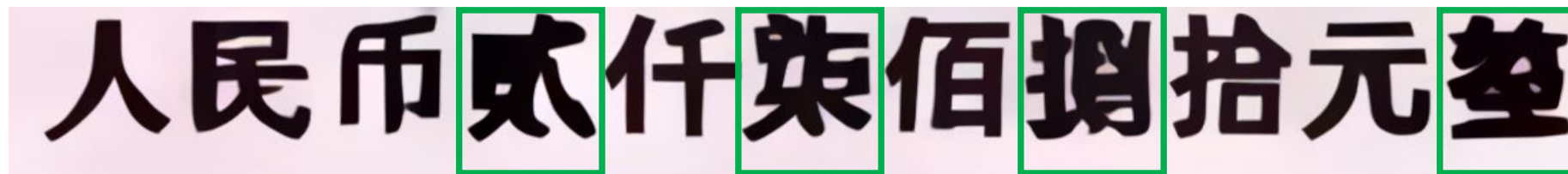Real-world LR Text Segment   BSRGAN Retrained on Text Image

Blind text image super-resolution

- If the structure of the character is complex (e.g., Chinese):
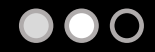

Real-world LR Text Segment

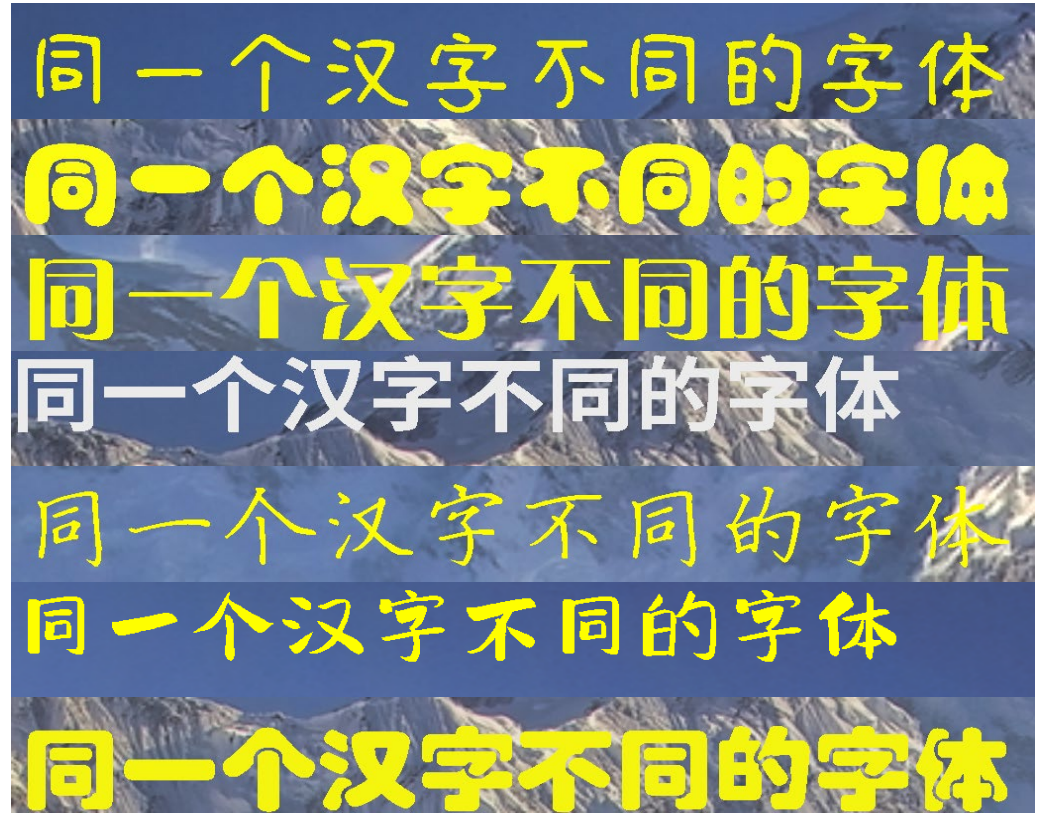

BSRGAN Retrained on Chinese Data

Structure Image from Human Perspective
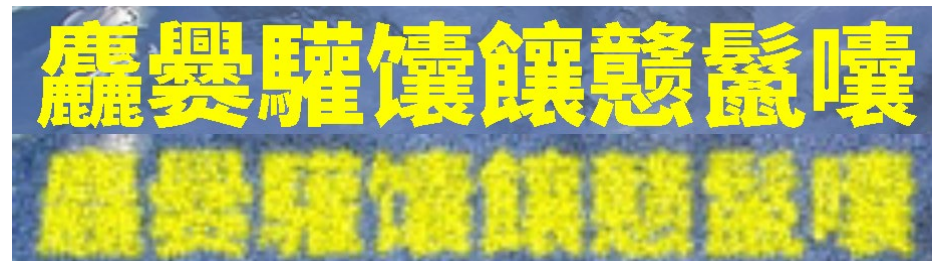
Challenge:

- Complex structure

- Diverse font styles
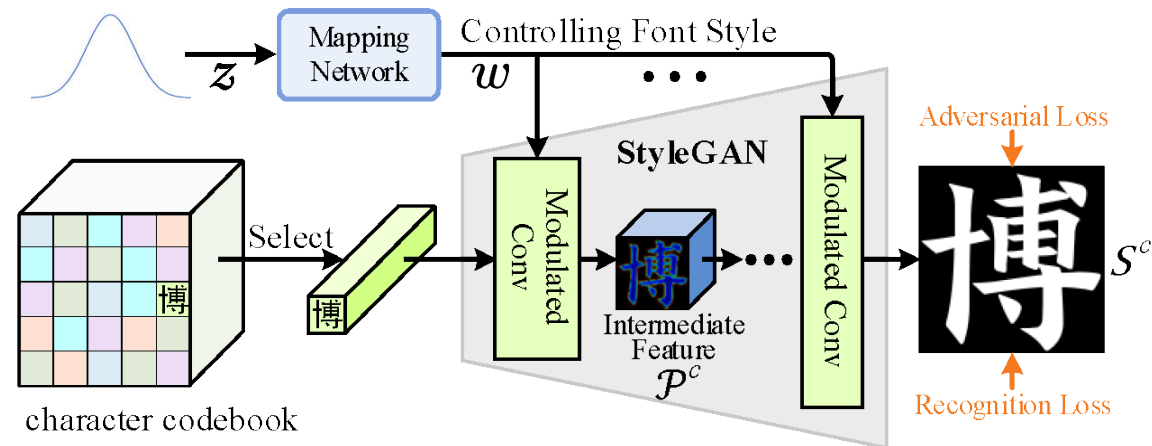
- Unknown degradation types on real-world scenarios

Original StyleGAN:

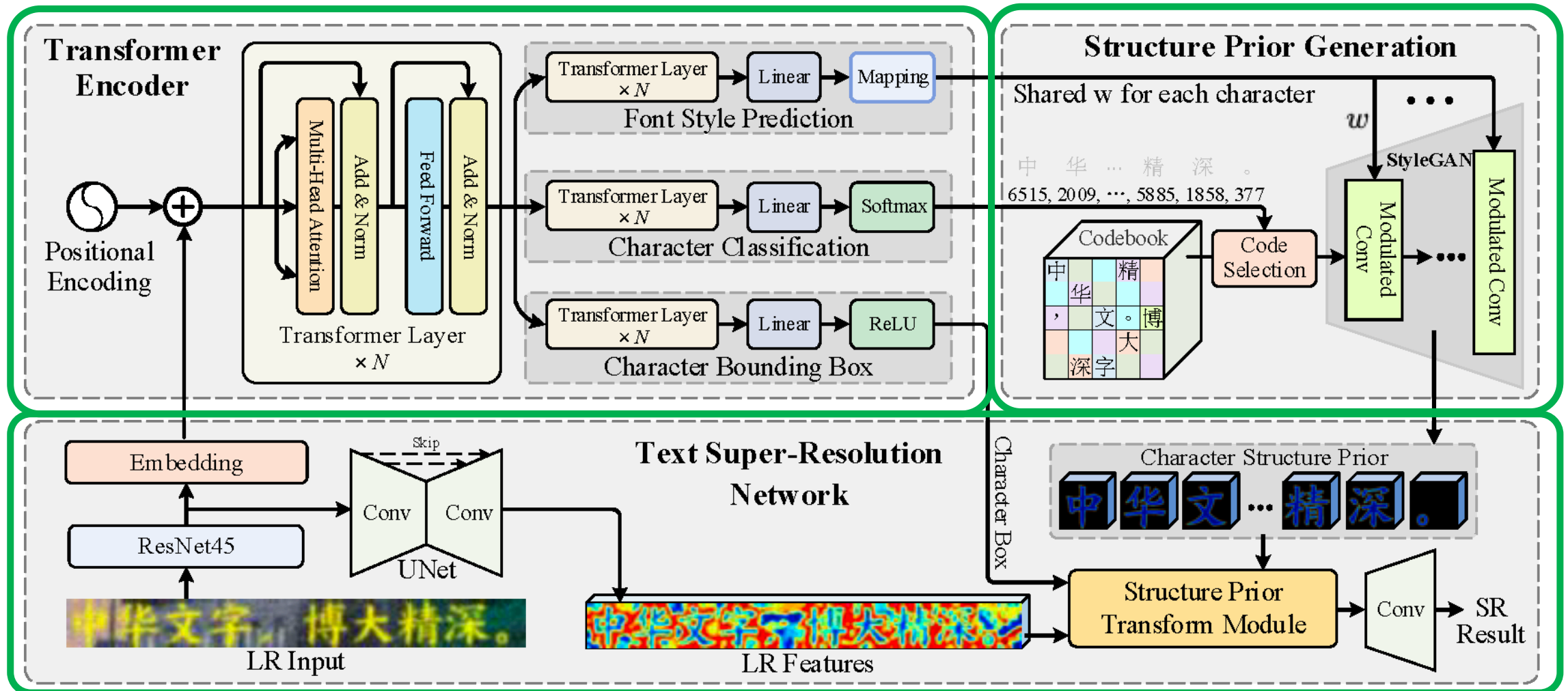- Powerful generation ability
- $\approx$ infinite representation

Reformulated StyleGAN for Text:

- Codebook for saving each character
- W controls the diverse font style

Pre-training of generative structure prior for each character

The whole pipeline. It contains three parts, *i.e.*, (i) Transformer encoder for predicting the font style, classification and bounding boxes of each character from LR input, (ii) structure prior generation with pre-trained StyleGAN for generating reliable structure prior for each character, and (iii) the SR process for reconstructing the SR output with the incorporation of each characters' structure prior.
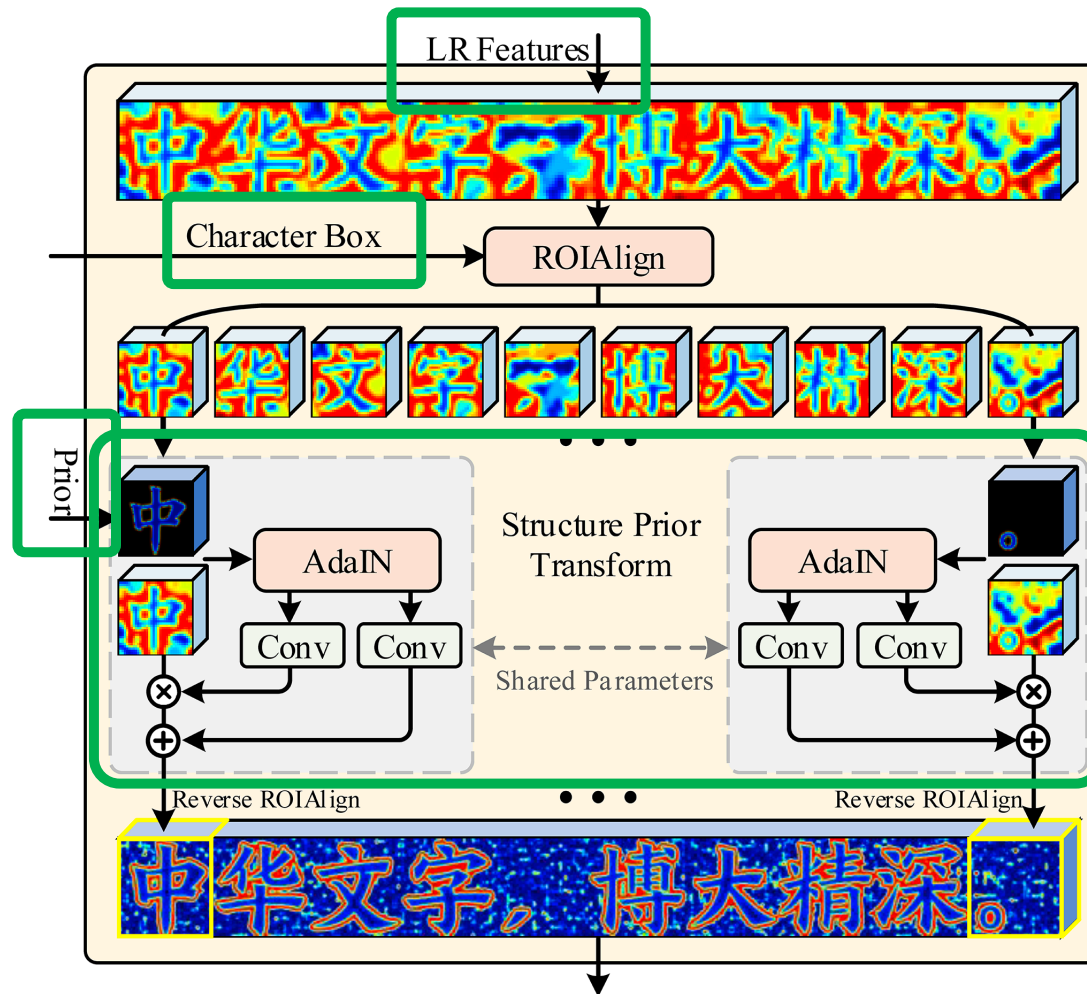
Structure prior transform module.

Pure Synthetic Text Images Using PIL package:
- More than 100 font styles;
- Background image is obtained from DF2K dataset

Ground-truth Image I_{HR}							128
Structure Image of Each Character S_{HR}^c							
Index in Coodbook of Each Character	4868	6252	5692	6545	4318	4776	1035	1361
Bounding Box of Each Character	(6, 117)	(132, 252)	(259, 382)	(389, 506)	(515, 637)	(643, 763)	(771, 893)	(898, 1018)
Synthetic LR Input							64 for X2 SR task

|	
Real LR	上海市定额专用发票 品名规格
Detected Bounding Box	上海市定额专用发票 品名规格
SR Result	上海市定额专用发票 品名规格
Structure Prior Image	上海市定额专用发票 品名规格
Real LR	纳税人识别号： 价税合计（大写）
Detected Bounding Box	纳税人识别号： 价税合计（大写）
SR Result	纳税人识别号： 价税合计（大写）
Structure Prior Image	纳税人识别号： 价税合计（大写）

More SR results on real-world LR scenarios

Wrong recognition when the degradation is severe.
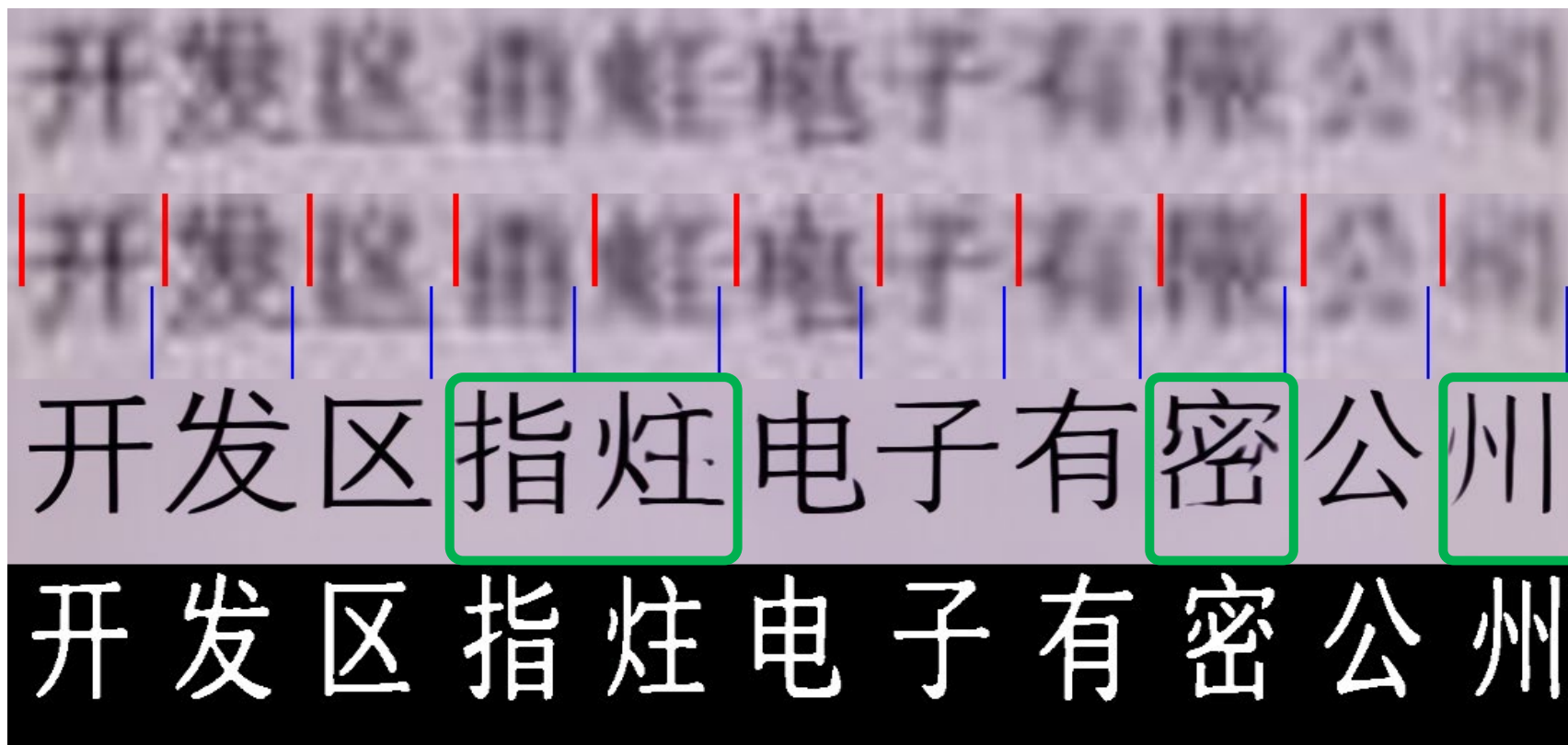
Manually correct the character recognition result.
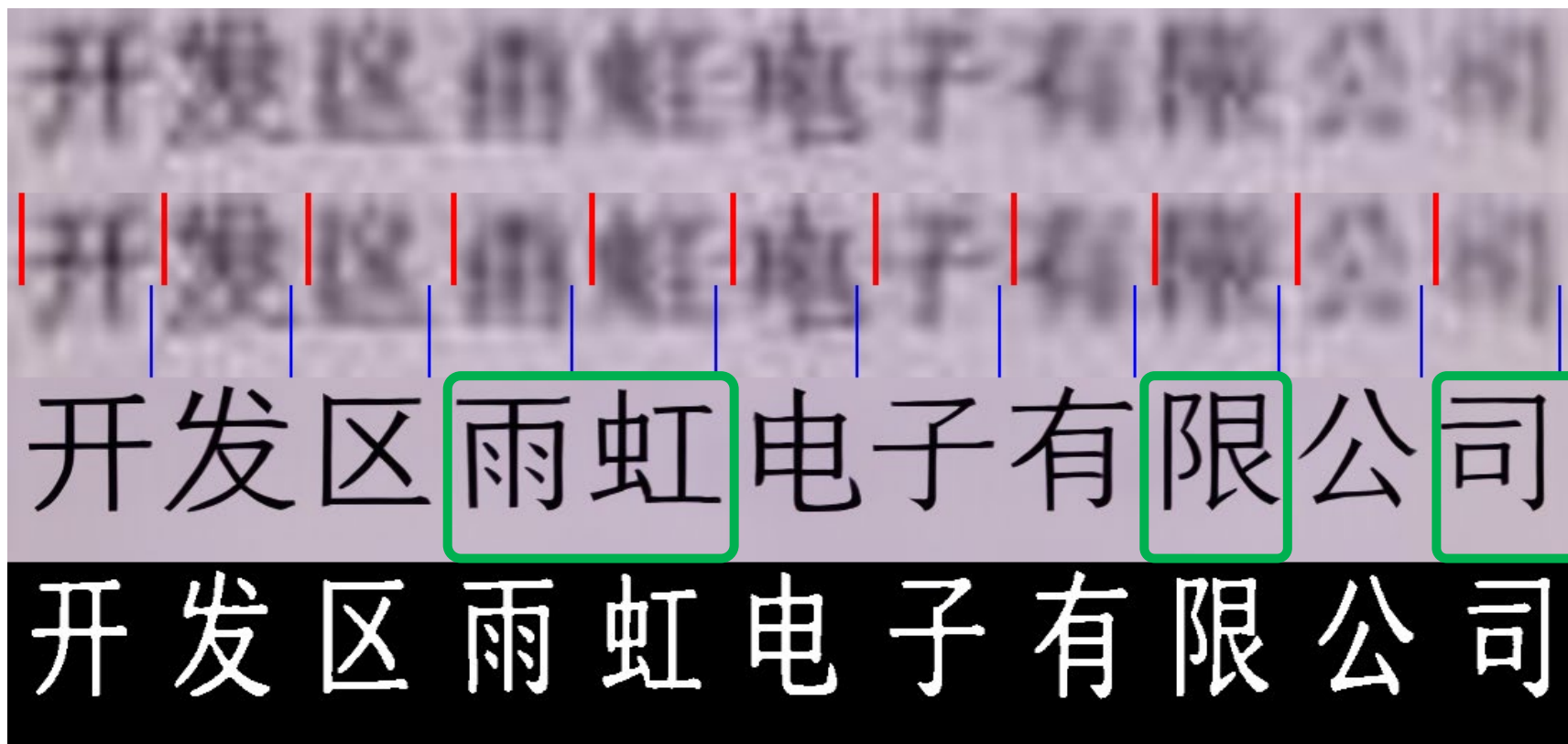
W interpolation for two text images with different font styles.

W interpolation for two text images with different characters.
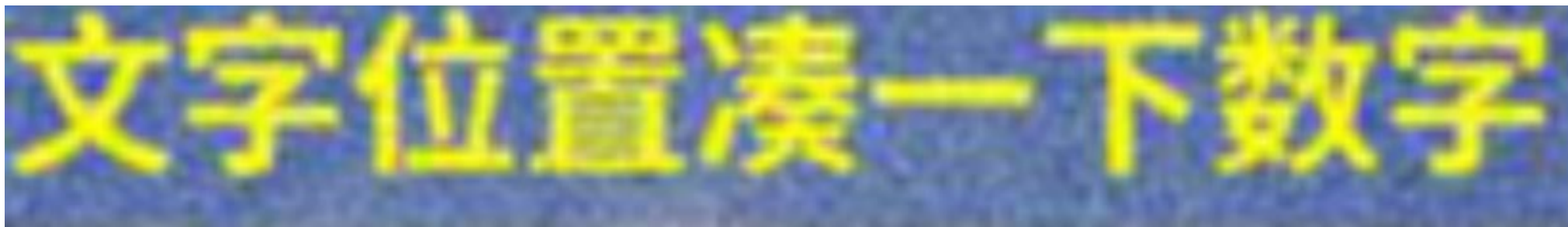
W interpolation for two text images with different locations:

Conclusion

- Embedding the generative structure prior for blind SR of text images
- The combination of a codebook for storing distinctive character specific codes and a retrofitted StyleGAN for controlling font style cope well with complicated structures
- Potentially extending to:
  - ➤ few-shot font generation
  - ➤ text image completion for ancient documents
  - ➤ font style transformation

A simple text image super-resolution package for post-processing text region:

```
pip install textbsr
```

Run on the terminal command

```
textbsr -i [LR_TEXT_PATH] –b [BACKGROUND_SR_PATH] -s
```

Run on the python environment

```
from textbsr import textbsr
textbsr.bsr(input_path='./LQs', bg_path='./RealESRGANResults', save_text=True)
```

Project Page

Whole LR Image

Any Blind Image SR Method

Post-process using our textbsr