Distilling GRU with Data Augmentation for Unconstrained Handwritten Text Recognition

Reporter: Zecheng Xie

South China University of Technology

August 6，2018

Outline

- **Problem Definition**
- **Multi-layer Distilling GRU**
- **Data Augmentation**
- **Experiments**
- **Conclusion**

Outline

- **Problem Definition**
- Multi-layer Distilling GRU
- Data Augmentation
- Experiments
- Conclusion

Problem Definition

Motivation

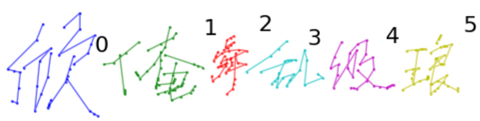
- Handwritten texts with various styles, such as horizontal, overlapping, vertical, and multi-lines texts, are commonly observed in the community.
- Most existing handwriting recognition methods only concentrate on one specific kind of text style.

⇨ **The new unconstrained online handwritten text recognition problem**

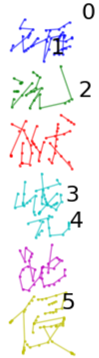
Problem Definition

The New Unconstrained OHCTR Problem

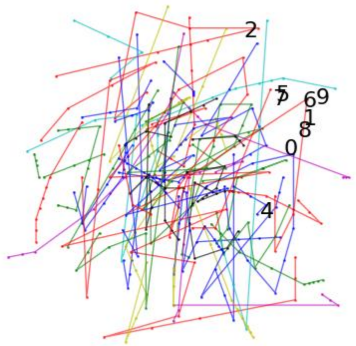
Horizontal



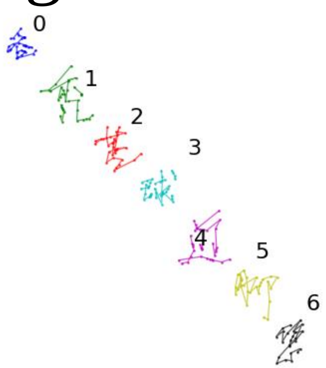
Vertical



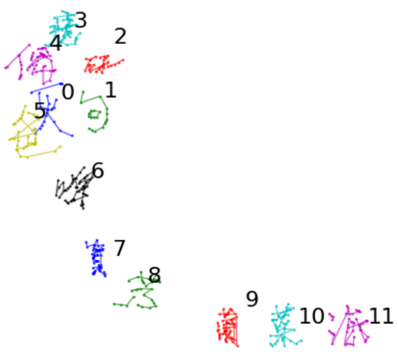
Overlap



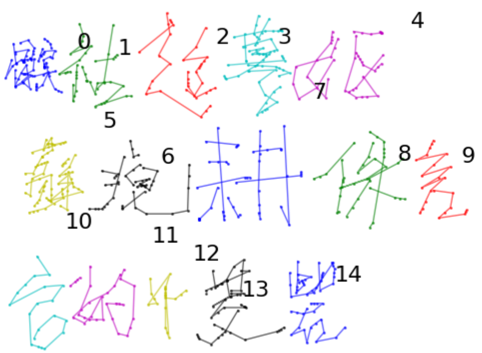
Right-Down



Screw-Rotation



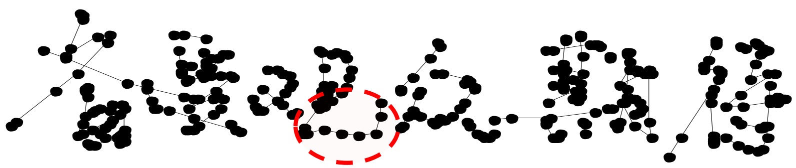
Multi-line



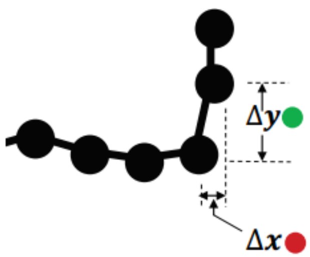
Problem Definition

Novel Perspective

Why not focusing on the variation between adjacent points[14,15].



More stable than the pen-tip coordinate—distribute between a specific bound for most situations.



$$\Delta y_t = y_{t+1} - y_t$$

$$\Delta x_t = x_{t+1} - x_t$$

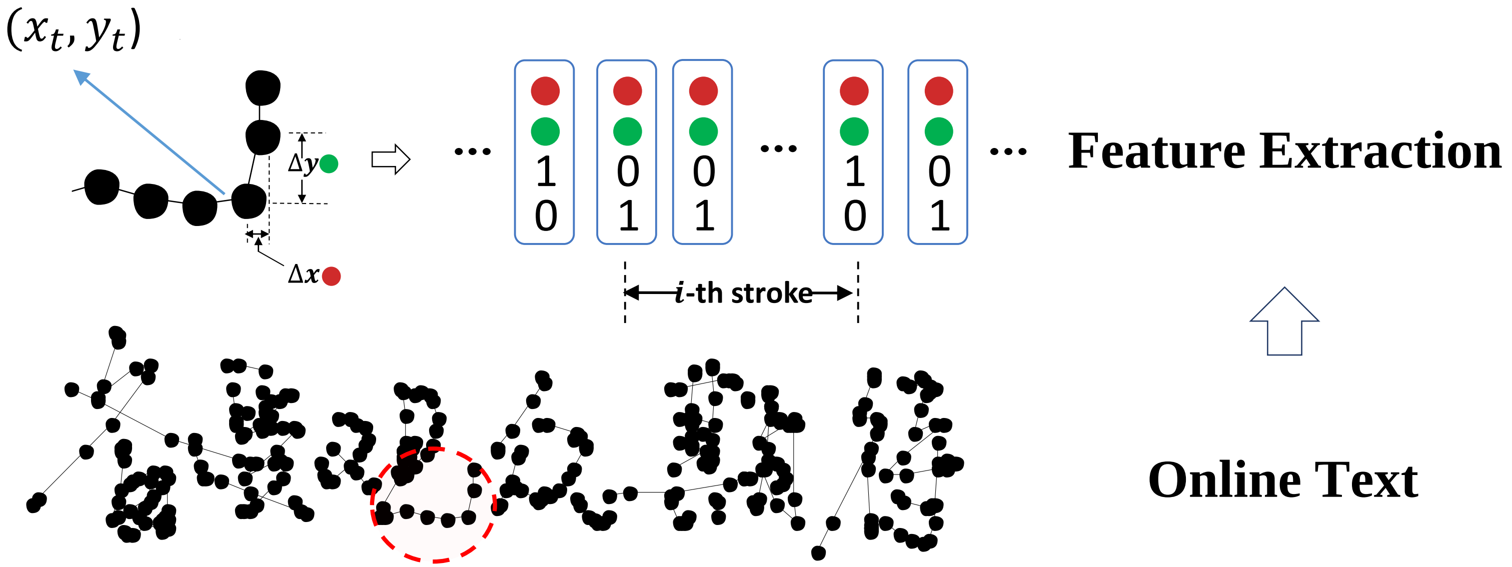
The unconstrained text of multiple styles share a very similar feature pattern, the only difference between different text styles is the pen-tip movement between characters.

[14] X. Zhang, *et al.* "Drawing and recognizing Chinese characters with recurrent neural network," IEEE transactions on pattern analysis and machine intelligence, 2018.
[15] L. Sun, *et al.* "Deep lstm networks for online Chinese handwriting recognition, in ICFHR 2016.

Outline

- Problem Definition
- **Multi-layer Distilling GRU**
- Data Augmentation
- Experiments
- Conclusion

Feature Extraction



Sampling Points→ Pen-tip Movement → Pen down\up state

Multi-layer Distilling GRU

Distilling GRU

- GRU can only output feature sequence with the same time step as that of the input data
  - greatly burden the framework if directly applied in text recognition problem.

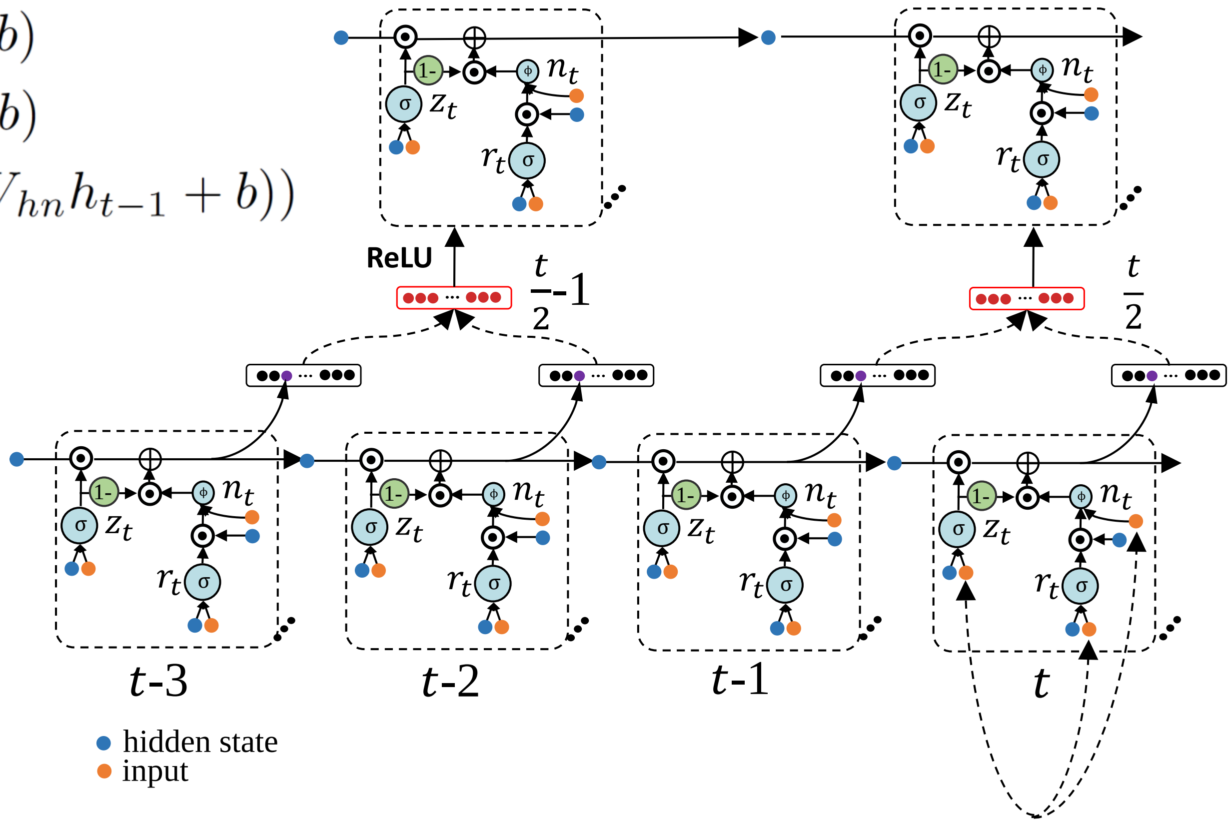
⇨ How to accelerate the training process while not sacrifice performance.

Distilling GRU

$$r_t = \delta(W_{ir}x_t + W_{hr}h_{t-1} + b)$$
$$z_t = \delta(W_{iz}x_t + W_{hz}h_{t-1} + b)$$
$$n_t = \tanh(W_{in}x_t + b + r_t(W_{hn}h_{t-1} + b))$$
$$h_t = (1 - z_t)n_t + z_t h_{t-1}$$

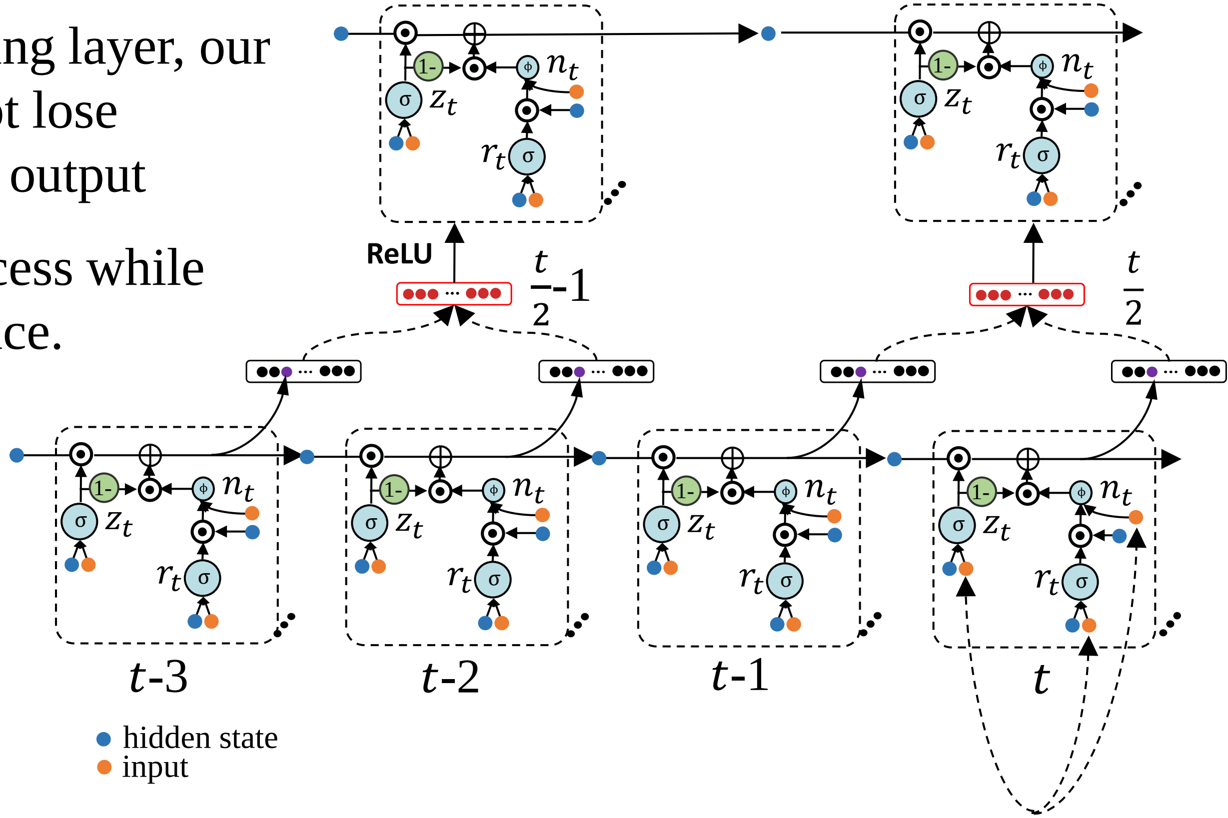
$$h = (h_1, h_2, \dots, h_T)$$
$$h' = (h'_1, h'_2, \dots, h'_{T/N})$$

$$\overline{h}_t = relu(W_{h'_t\overline{h}_t} h'_t)$$



Distilling GRU

- Unlike the traditional pooling layer, our distilling operation does not lose information from the GRU output
- Accelerate the training process while not sacrifice any performance.



Multi-layer Distilling GRU

Transcription

'blank'	…	0.907	0.349	…	0.1	0.82	…	0.02	…		
观	…	0.001	0.001	…	0.789	0.1	…	0.003	…		
…	…	0.003	0.003	…	0.08	0.007	…	0.004	…		
…	…	⋮	⋮	…	⋮	⋮	…	⋮	…		
期	…	0.002	0.001	…	0.001	0.001	…	0.8	…		
…	…	⋮	⋮	…	⋮	⋮	…	⋮	…		
…	…	0.001	0.0015	…	0.002	0.002	…	0.001	…		

$\boldsymbol{\pi}$: _备_受_观观_众_期期_待_

$\boldsymbol{\pi}$: _备_受_观_众_期_待

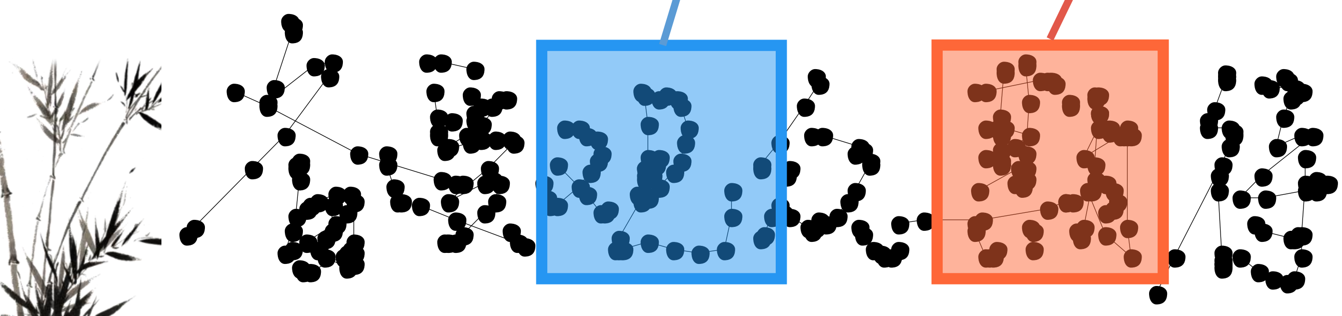
$\boldsymbol{\pi}$: _备_受_观_众_期期期_待

… …

$\mathfrak{B}$

备受观众期待

$$Pr(\boldsymbol{l}|\boldsymbol{s}) = \sum_{\boldsymbol{\pi}:\mathfrak{B}(\boldsymbol{\pi})=\boldsymbol{l}} Pr(\boldsymbol{\pi}|\boldsymbol{s})$$



Multi-layer Distilling GRU

Multi-layer Distilling GRU

$$L(Q) = -\ln \prod_{(l,z)\subset Q} p(z|l) = -\sum_{(l,z)\subset Q} \ln p(z|l)$$

⇧

$$\overline{h}_t = relu(W_{h'_t \overline{h}_t} h'_t)$$

$$h' = (h'_1, h'_2, \dots, h'_{T/N})$$

⇧

$$r_t = \delta(W_{ir}x_t + W_{hr}h_{t-1} + b)$$

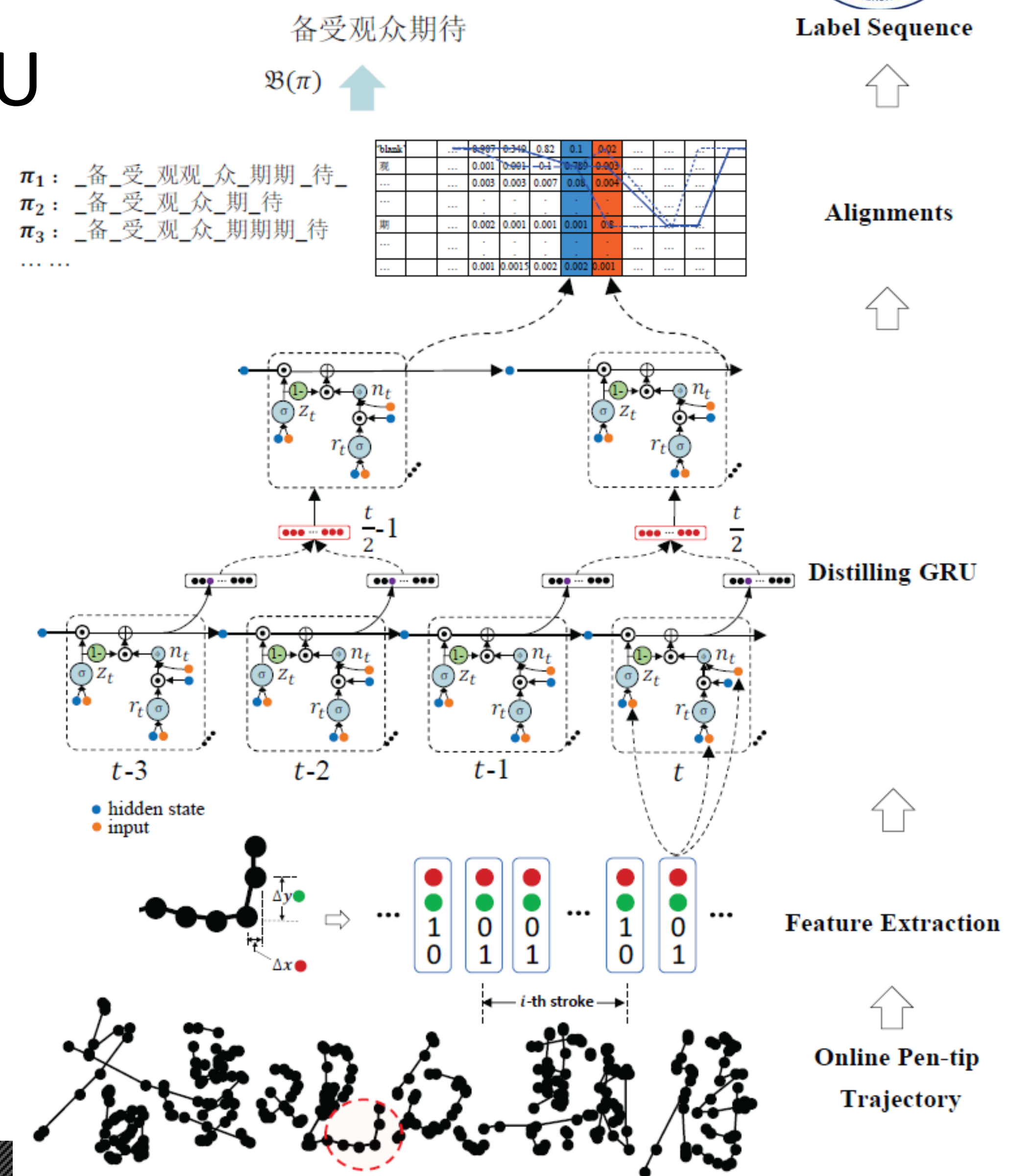
$$z_t = \delta(W_{iz}x_t + W_{hz}h_{t-1} + b)$$

$$n_t = \tanh(W_{in}x_t + b + r_t(W_{hn}h_{t-1} + b))$$

$$h_t = (1 - z_t)n_t + z_t h_{t-1}$$

⇧

$$I = \{(\Delta x_t, \Delta y_t, \mathbb{I}(s_t \neq s_{t+1}), \mathbb{I}(s_t = s_{t+1}))\}$$



Outline

- Problem Definition
- Multi-layer Distilling GRU
- **Data Augmentation**
- Experiments
- Conclusion

Data Augmentation

Horizontal
$$\begin{cases} \Delta x_i^h = (x_i^{max} - x_i^l) + (x_{i+1}^{min} - x_{i+1}^f) + \Delta x_r \\ \Delta y_i^h = y_{i+1}^f - y_i^l + \Delta y_r \end{cases}$$

Vertical
$$\begin{cases} \Delta x_i^v = x_{i+1}^f - x_i^l + \Delta x_r \\ \Delta y_i^v = (y_i^{max} - y_i^l) + (y_{i+1}^{min} - y_{i+1}^f) + \Delta y_r \end{cases}$$

Overlapping
$$\begin{cases} \Delta x_i^o = (x_i^{min} - x_i^l) + (x_{i+1}^f - x_{i+1}^{min}) + \Delta x_r \\ \Delta y_i^o = (y_i^{min} - y_i^l) + (y_{i+1}^f - y_{i+1}^{min}) + \Delta y_r \end{cases}$$

Multi-lines
$$\begin{cases} \Delta x_i^s = \Delta x_i^h cos(r) + \Delta y_i^h cos(r) \\ \Delta y_i^s = -\Delta x_i^h sin(r) + \Delta y_i^h cos(r) \\ r_t = r_{t-1} + \frac{\pi}{2\sqrt{t}}, r_0 = \frac{\pi}{5} \end{cases}$$

Screw rotation
$$\begin{cases} \Delta x_i^r = (x_i^{max} - x_i^l) + (x_{i+1}^{min} - x_{i+1}^f) + \Delta x_r \\ \Delta y_i^r = (y_i^{max} - y_i^l) + (y_{i+1}^{min} - y_{i+1}^f) + \Delta y_r \end{cases}$$

Right-down
$$\begin{cases} 1) \begin{cases} \Delta x_i^m = (x_i^{max} - x_i^l) + (x_{i+1}^{min} - x_{i+1}^f) + \Delta x_r \\ \Delta y_i^m = y_{i+1}^f - y_i^l + \Delta y_r \end{cases} \\ 2) \begin{cases} \Delta x_i^m = -\Delta x_{line} + \Delta x_r \\ \Delta y_i^m = (y_i^{max} - y_i^l) + (y_{i+1}^{min} - y_{i+1}^f) + \Delta y_r \end{cases} \end{cases}$$

$\Delta x_i, \Delta y_i$: pen movement between the *i* and *i* + 1-th characters.

x_i^{min}, x_i^{max} :the minimum and maximum x-coordinate value of the i-th character.

x_i^f, x_i^l: the x-coordinate values of the first and last points of the i-th character.

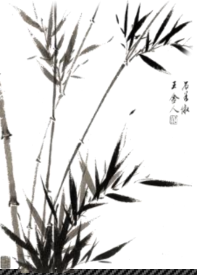
Δx_r:a random bias generated from an even distribution between (-2, 13).

Δx_{line} :text line length that can be adjusted according to practical situation.

All the abovementioned definitions also apply for the Y-axis.

Outline

- Problem Definition
- Multi-layer Distilling GRU
- Data Augmentation
- **Experiments**
- Conclusion



Experiments

- **Training Data**

CASIA-OLHWDB2.0-2.2[1]

Synthetic Unconstrained Data by CASIA-OLHWDB1.0-1.2[1]

- **Testing Data**

ICDAR2013 Test Dataset[2]

Synthetic Unconstrained Data by CASIA-OLHWDB1.0-1.2[1]

- **Network**

2-Layers Distilling GRU， Distilling Rate=0.25

- **Hardware**

GeForce Titan-X GPU

Convergence time 208h→95h

[1] C. Liu., *et al*, “CASIA online and offline Chinese handwriting databases,” 2011 International Conference on Document Analysis and Recognition (ICDAR), pp. 37–41, 2011

[2] Yin F., *et al*, “ICDAR 2013 Chinese handwriting recognition competition,” ICDAR2013 , pp. 1464–1470.

Experiments

TABLE I

EFFECT OF DISTILLING GRU AND SYNTHETIC SAMPLE.

Methods	Accuracy Rate	Training Time (hour)
baseline	88.31	208
+distilling GRU	88.33	95
+horizontal text	91.36	102

TABLE II

SYNTHETIC UNCONSTRAINED TEXT SAMPLES OF VARIOUS STYLE.

Text Styles	Casia	Casia+'hvo'	Casia+'hvorsm'
ICDAR2013	88.33	90.57	90.62
horizontal	30.67	91.93	92.61
vertical	0.76	93.32	93.30
overlap	1.41	92.23	91.98
right-down	23.31	91.40	93.74
screw-rotation	16.00	87.89	92.89
multi-line	24.62	90.26	91.94

Corresponding text style synthetic data is used for result marked with blue color.

Experiments

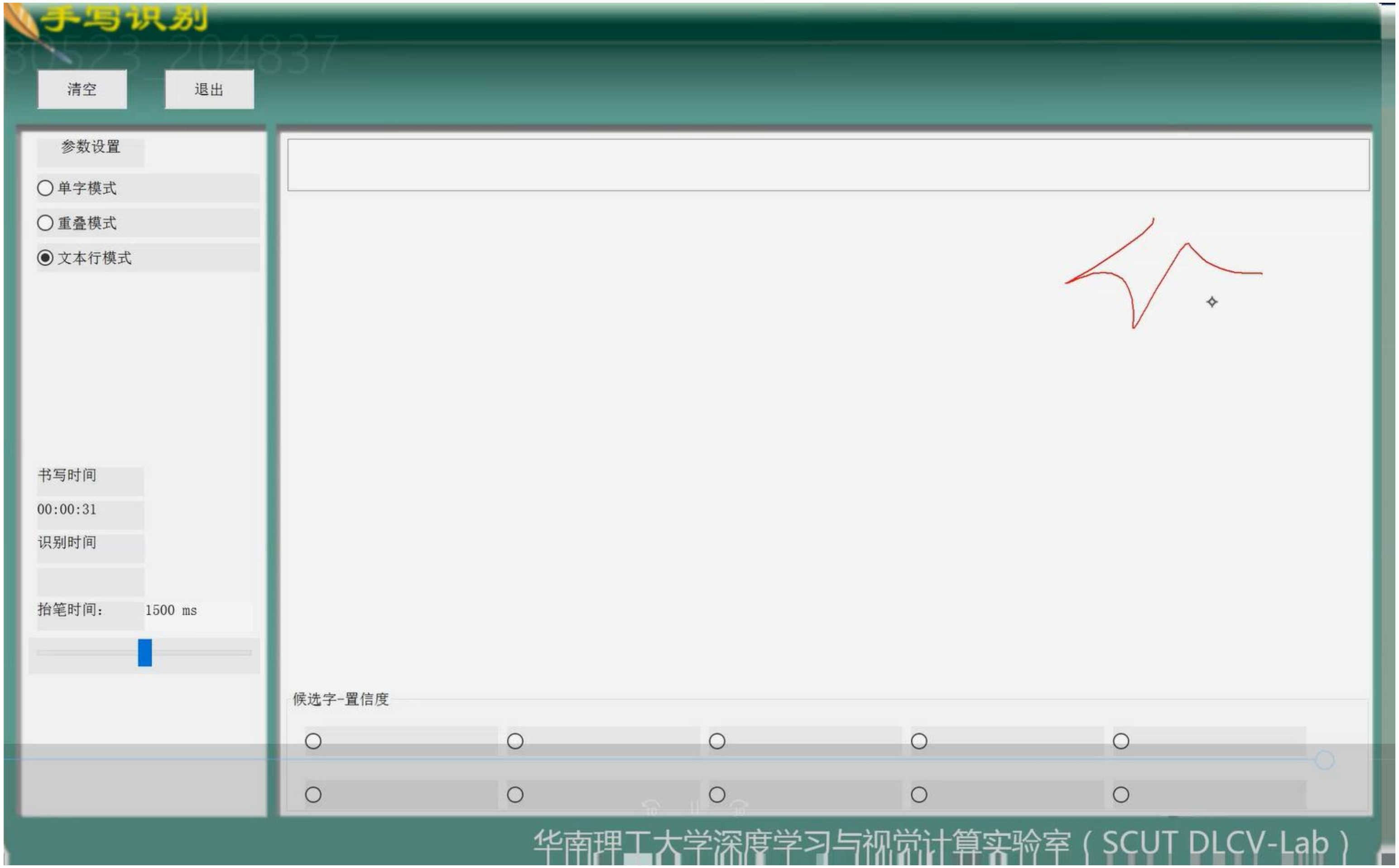
TABLE III
COMPARISON WITH PREVIOUS METHODS

Method	w.o. LM		with LM	
	CR	AR	CR	AR
Zhou et al., 2013 [3]	-	-	94.62	94.06
Zhou et al., 2014 [4]	-	-	94.76	94.22
VO-3 [24]	-	-	95.03	94.49
xie et al., 2017 [29]	90.17	88.88	94.51	93.45
Chen et al.,2017 [30]	85.17	84.61	**96.71**	**96.46**
ours	**92.37**	**91.36**	95.70	94.89

[3] X. Zhou., *et al*, IEEE TPAMI, vol. 35, no. 10, pp. 2413–2426, 2013.
[4] X. Zhou., *et al,* Pattern Recognition[J], 2014, 47(5): 1904-1916
[29] Z. Xie., *et al,* IEEE TPAMI, 2017
[30] K. Chen, *et al,* in ICDAR 2017, vol. 1. IEEE, 2017, pp. 1068–1073.

Experiments

Demo



Conclusion

- The new unconstrained text recognition problem is suggested to advance the handwritten text recognition community.
- A special perspective of the pen-tip trajectory is suggested to reduce the difference between texts of multiple styles.
- A new data augmentation method is developed to synthesize unconstrained handwritten texts of multiple styles
- A Multi-layer distilling GRU is proposed to process the input data in a sequential manner
- Achieves state-of-the-art results on ICDAR2013 text competition dataset but also shows robust performance on our synthesized handwritten test sets.

Q & A

Thank you!

Lianwen Jin(金连文), Ph.D, Professor

eelwjin@scut.edu.cn lianwen.jin@gmail.com

Zecheng Xie(谢泽澄), Ph.D, student

Manfei Liu(刘曼飞), Master, student

http://www.hcii-lab.net/