自动化前沿热点讲堂
2021年4月15日

跨模态零样本文字识别

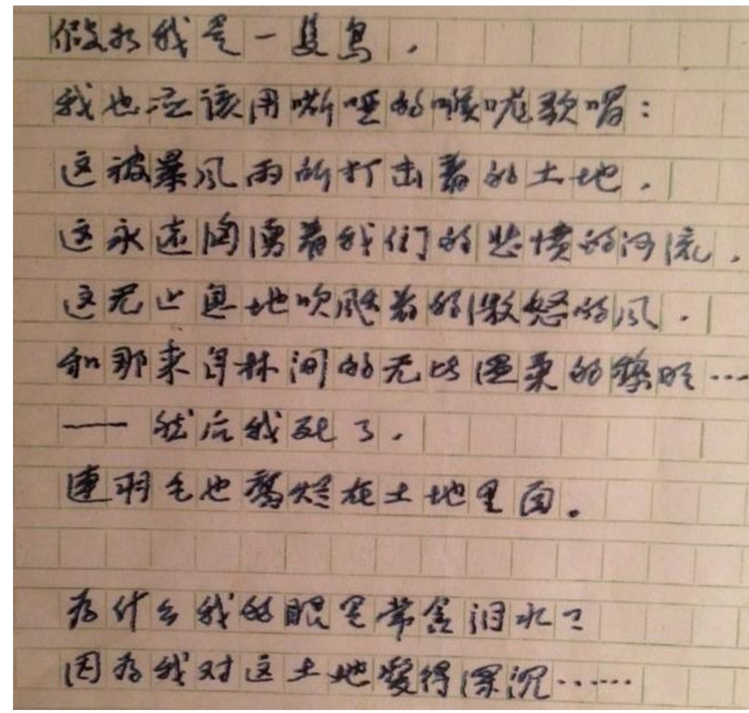
刘成林
中国科学院自动化研究所
模式识别国家重点实验室
liucl@nlpr.ia.ac.cn
http://www.nlpr.ia.ac.cn/liucl

提 纲

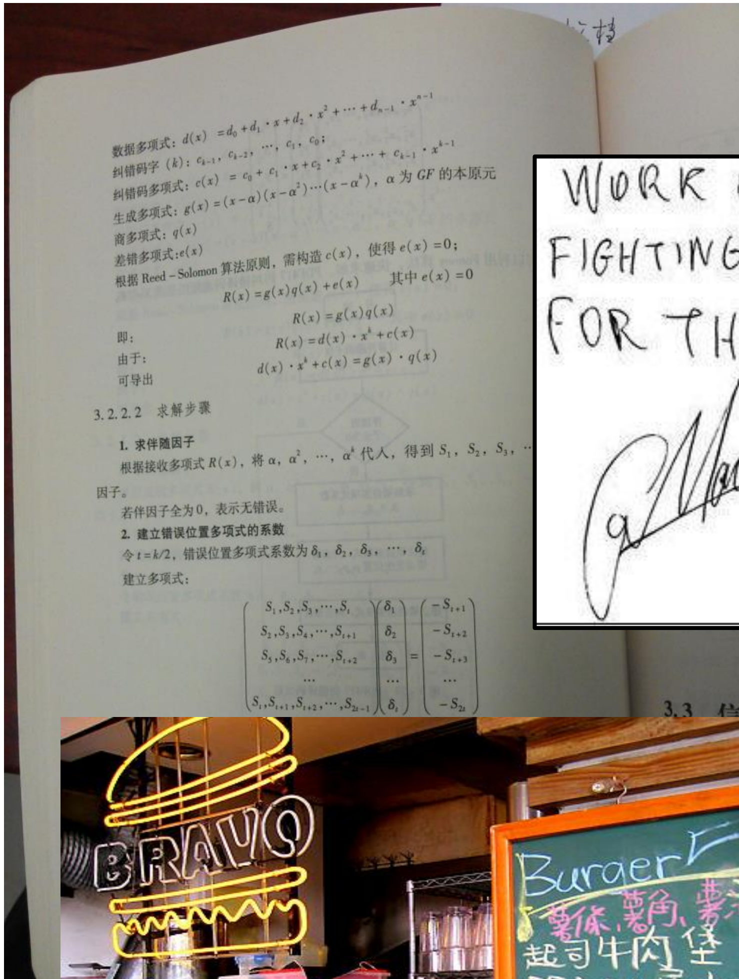
- 文字识别背景
- 单字识别研究意义
- 零样本文字识别
- 基于印刷体原型的手写汉字识别
- 基于跨模态度量学习的甲骨文字识别
- 讨论与展望

文字识别背景

- 文字识别(Character Recognition)
  - 字符图像转换为符号代码
- 文档分析(Document Analysis)
  - 从文档图像提取文本信息
  - 包括文本分割、识别、上下文处理、语义信息提取等
- 文档分析的意义
  - 数据压缩
  - 内容理解/语义提取



泛在的文档图像，内容丰富



数据多项式: $d(x) = d_0 + d_1 \cdot x + d_2 \cdot x^2 + \cdots + d_{n-1} \cdot x^{n-1}$

纠错码字 (k): $c_{k-1}, c_{k-2}, \cdots, c_1, c_0$;

纠错码多项式: $c(x) = c_0 + c_1 \cdot x + c_2 \cdot x^2 + \cdots + c_{k-1} \cdot x^{k-1}$

生成多项式: $g(x) = (x-\alpha)(x-\alpha^2)\cdots(x-\alpha^k)$，$\alpha$ 为 GF 的本原元

商多项式: $q(x)$

差错多项式: $e(x)$

根据 Reed – Solomon 算法原则，需构造 $c(x)$，使得 $e(x)=0$;

$$R(x) = g(x)q(x) + e(x) \quad 其中\ e(x)=0$$

即:
$$R(x) = g(x)q(x)$$

由于:
$$R(x) = d(x) \cdot x^k + c(x)$$

可导出
$$d(x) \cdot x^k + c(x) = g(x) \cdot q(x)$$

3.2.2.2 求解步骤

1. 求伴随因子

根据接收多项式 $R(x)$，将 $\alpha, \alpha^2, \cdots, \alpha^k$ 代入，得到 $S_1, S_2, S_3, \cdots$

因子。

若伴因子全为0，表示无错误。

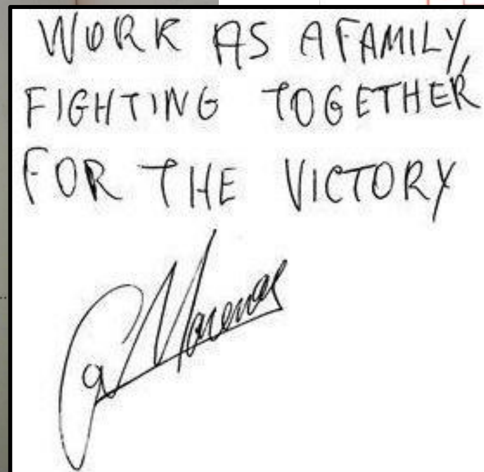
2. 建立错误位置多项式的系数

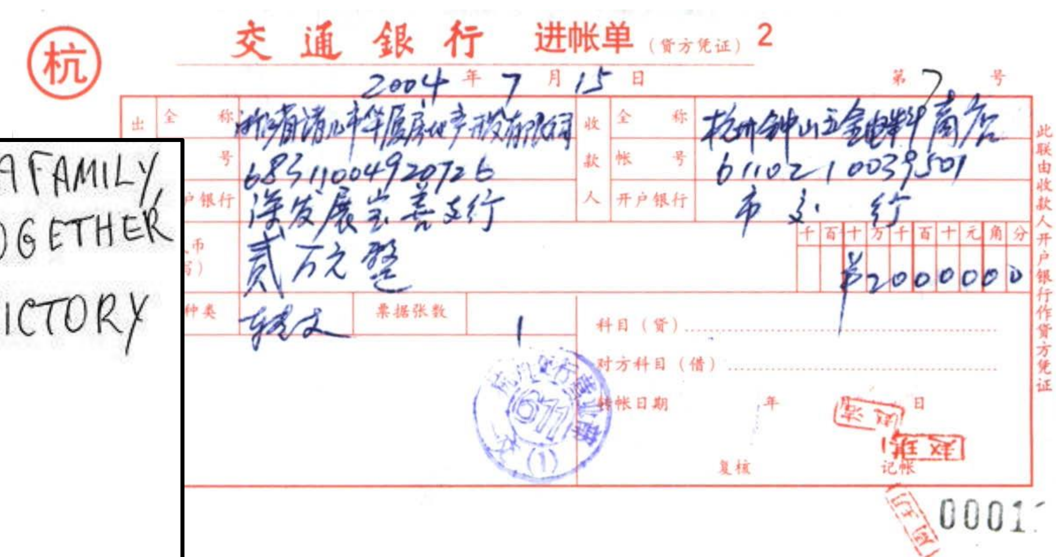
令 $t=k/2$，错误位置多项式系数为 $\delta_1, \delta_2, \delta_3, \cdots, \delta_t$

建立多项式:

$$\begin{pmatrix} S_1, S_2, S_3, \cdots, S_t \\ S_2, S_3, S_4, \cdots, S_{t+1} \\ S_3, S_4, S_5, \cdots, S_{t+2} \\ \cdots \\ S_t, S_{t+1}, S_{t+2}, \cdots, S_{2t-1} \end{pmatrix} \begin{pmatrix} \delta_t \\ \delta_{t-1} \\ \delta_{t-2} \\ \cdots \\ \delta_1 \end{pmatrix} = \begin{pmatrix} -S_{t+1} \\ -S_{t+2} \\ -S_{t+3} \\ \cdots \\ -S_{2t} \end{pmatrix}$$

3.3 信





杭 交通银行 进帐单（贷方凭证） 2

2004 年 7 月 15 日 第 号

出票人 全称: 浙江省诸暨市华厦房地产股份有限公司

账号: 68311004920726

开户银行: 浦发展宝善支行

收款人 全称: 杭州钟山五金电料商店

账号: 61102100039501

开户银行: 市支行

人民币（大写）: 贰万元整

千百十万千百十元角分: ¥2000000

票据种类: 转支 票据张数: 1

科目（贷）

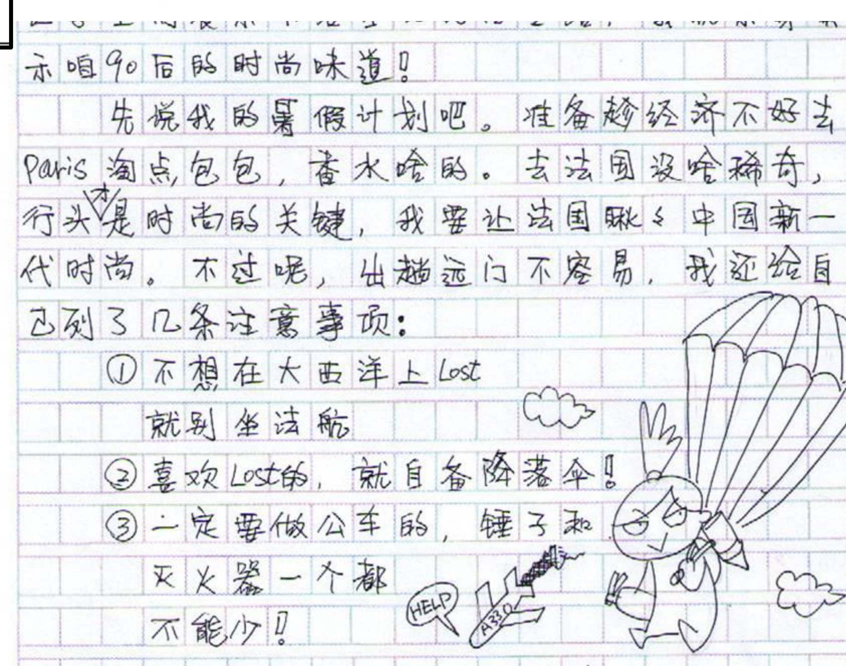
对方科目（借）

转帐日期 年 月 日

复核 记帐

此联由收款人开户银行作贷方凭证

0001



示咱90后的时尚味道!

先说我的暑假计划吧。准备趁经济不好去Paris淘点包包，香水啥的。去法国没啥稀奇，行头才是时尚的关键，我要让法国瞅瞅中国新一代时尚。不过呢，出趟远门不容易，我还给自己列了几条注意事项:

①不想在大西洋上Lost 就别坐法航

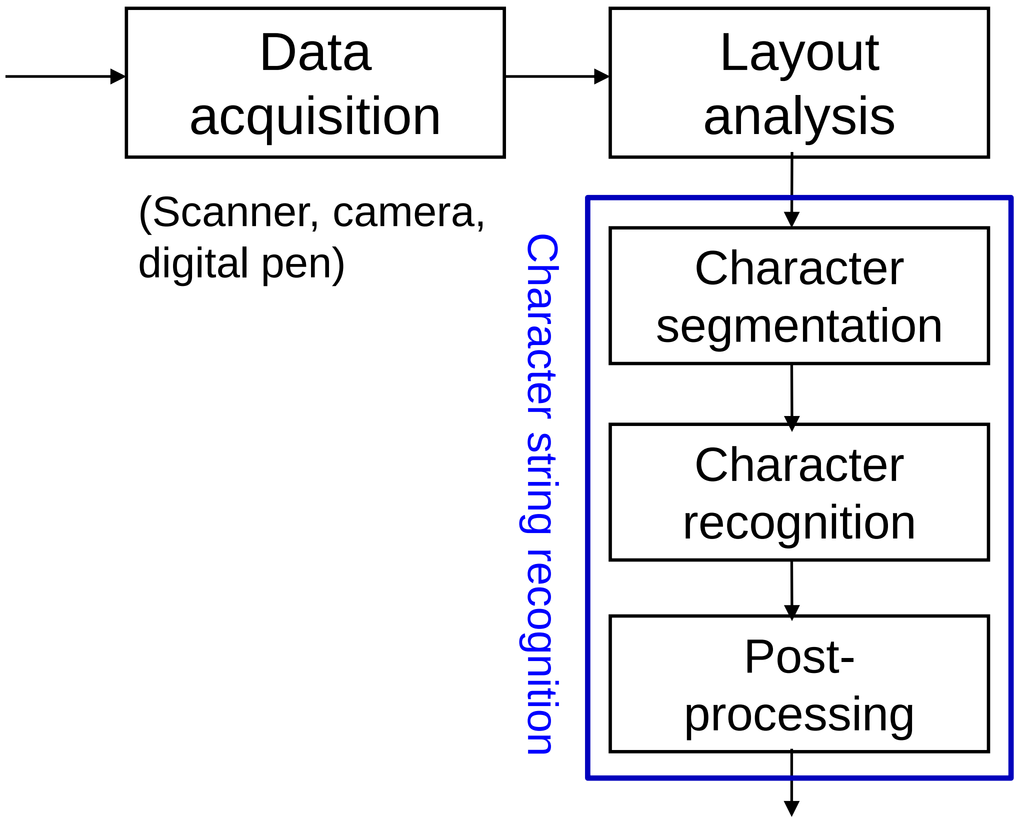
②喜欢Lost的，就自备降落伞!

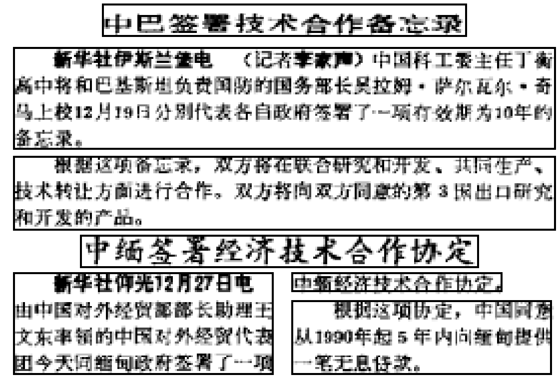
③一定要做公车的，锤子和灭火器一个都不能少!

HELP

A330

文档识别流程





中巴签署技术合作备忘录

新华社伊斯兰堡电 （记者李家声）中国科工委主任丁衡高中将和巴基斯坦负责国防的国务部长吴拉姆·萨尔瓦尔·奇马上校12月19日分别代表各自政府签署了一项有效期为10年的备忘录。

根据这项备忘录，双方将在联合研究和开发、共同生产、技术转让方面进行合作。双方将向双方同意的第3国出口研究和开发的产品。

中缅签署经济技术合作协定

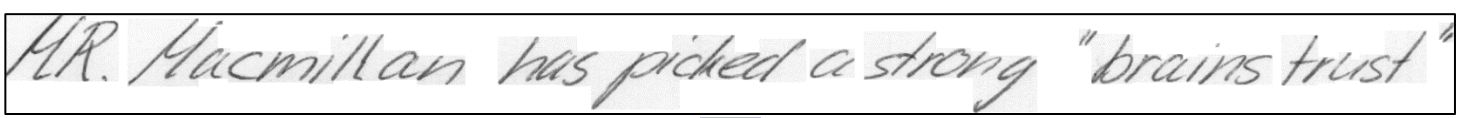
新华社仰光12月27日电 由中国对外经贸部部长助理王文东率领的中国对外经贸代表团今天同缅甸政府签署了一项

中缅经济技术合作协定。

根据这项协定，中国同意从1990年起5年内向缅甸提供一笔无息贷款。

中国对外经货部

中国对外经贸部



Text line recognition, with or w/o word segmentation

MR. Macmillan has picked a strong “brains trust”

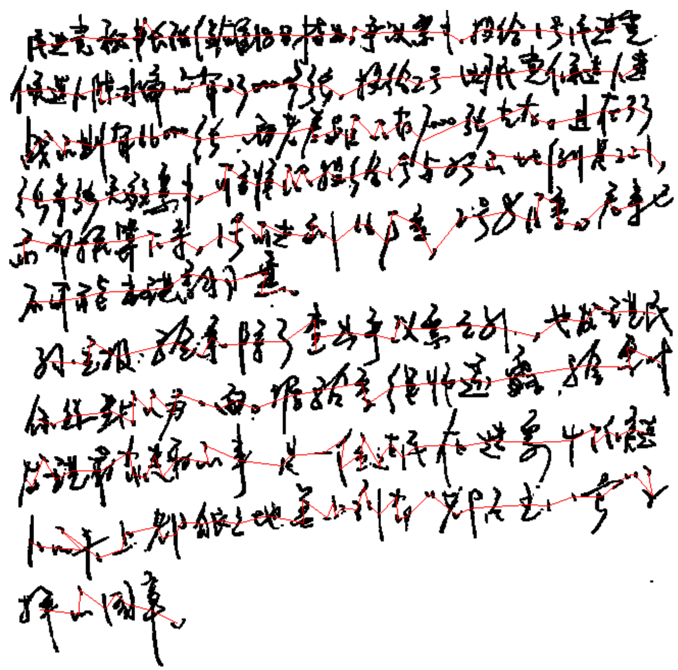
文档识别研究简史

Time	Methods	Target/Application	Events
1920s	Optical template matching	Printed digits/letters	1st patent on OCR
1950s-1960s	Correlated matching, simple structural analysis	Printed digits/letters Printed Chinese (1966)	1st PR Workshop in 1966
1970s-1980s	Feature matching (normalization, feature extraction), Structural matching, Statistical PR	Handprinted digits/letters Printed/handprinted words Handprinted Japanese/Chinese	1st ICPR in 1973 IAPR founded in 1978
1990s	Research of various issues, including layout analysis and segmentation, HMM	Practical applications in various areas (document entry, mail sorting, forms, business cards, text input)	PC got popular Internet 1st IWFHR/ICDAR/DAS in 1990/91/94
2000s	Re-inventing existing methods (e.g., HMM) Borrowing from ML and CV (e.g., BoW, deep learning, RNN)	Remaining hard problem Improve existing apps Explore new apps (e.g., camera-based, ink documents)	Google, Baidu Facebook, twitter Smart phone, GPU Mobile Internet Weibo, Weixin

进展：复杂文档版面分割

- Modern: to handle variable complex documents
  - Deformable models
  - Graph-based clustering
  - Structured prediction, such as CRF (conditional random graph), GNN (graph neural network)
  - FCN (fully convolutional network)

Active contour

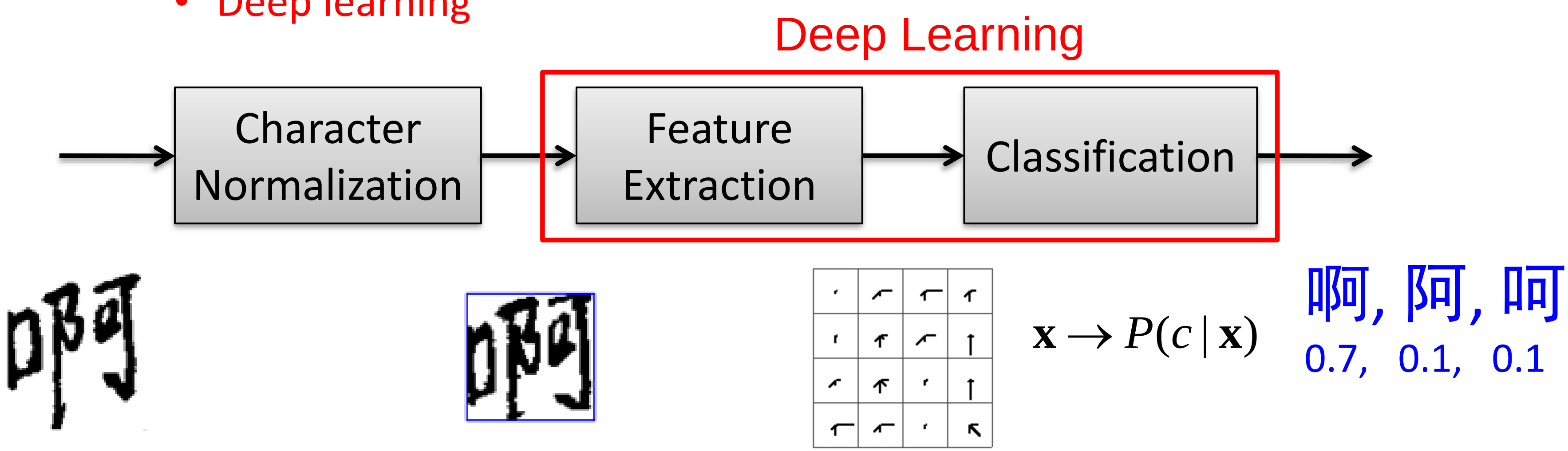


MST clustering

FCN based page seg

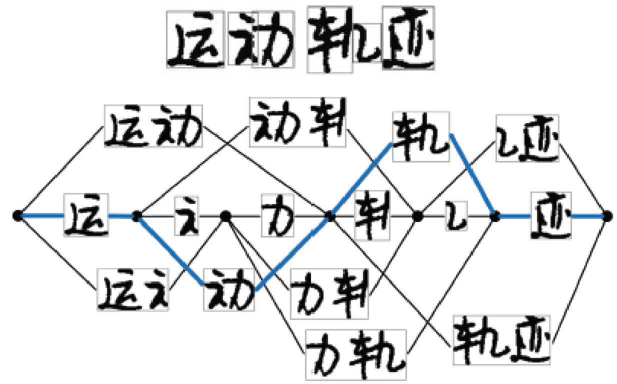
进展：文本识别

- Early research mostly focused on isolated character recognition, esp. for Chinese characters (large category)
  - Character recognition
    - Normalization: linear, moment-based, nonlinear, pseudo 2D
    - Feature extraction: direction histogram, Gabor, structural
    - Dimensionality reduction: PCA, FDA, DFE (discriminative)
    - Classification: statistical, neural (MLP, RBF, polynomial), SVM
      - Large category set: MQDF, LVQ, hierarchical
    - Deep learning

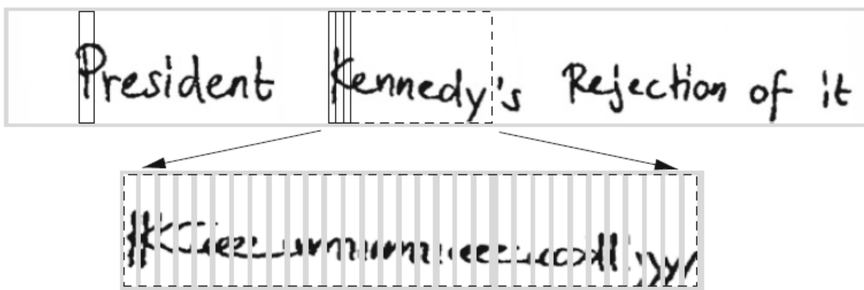


Multiple candidate classes for integration in string recognition incorporating linguistic context

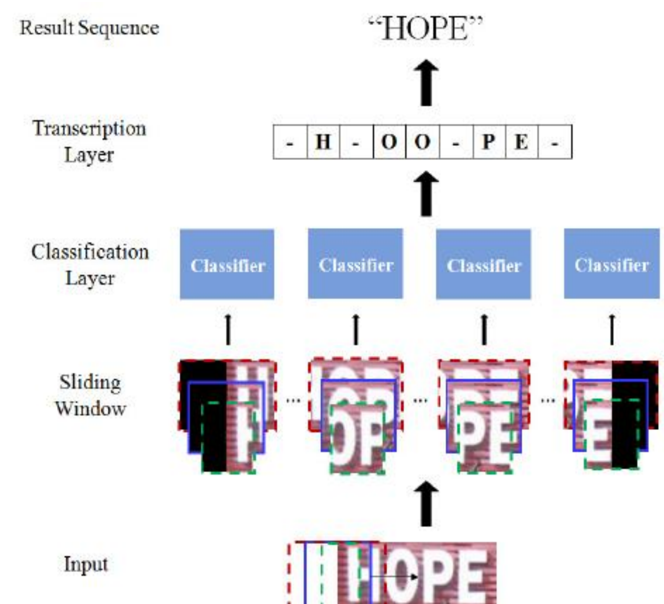
- Text (word/line) Recognition
  - Explicit/over segmentation
    - Relevant to human cognition
    - Good for fusing contexts and knowledge



  - Implicit segmentation: sliding window
    - HMM

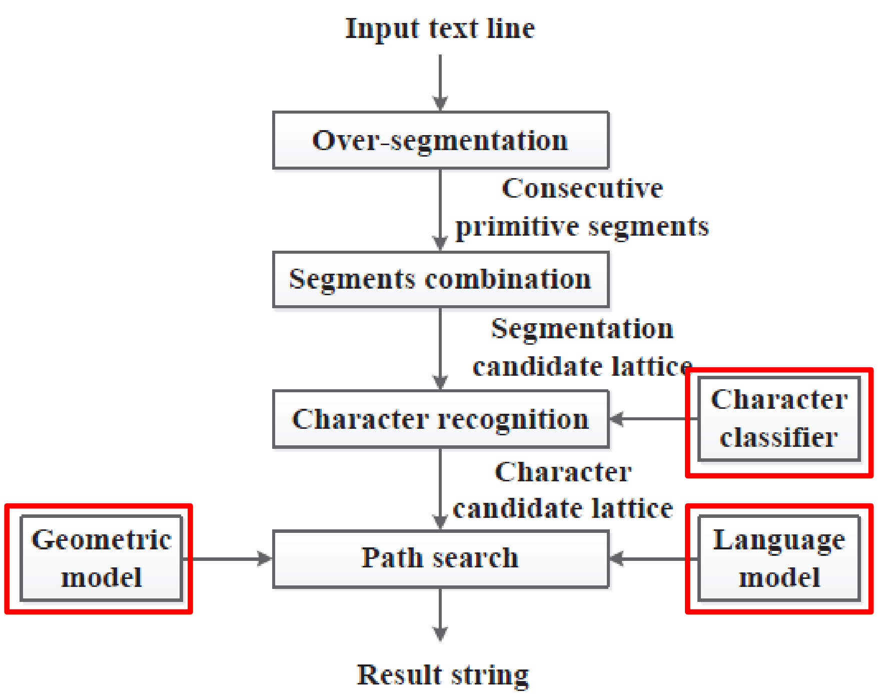


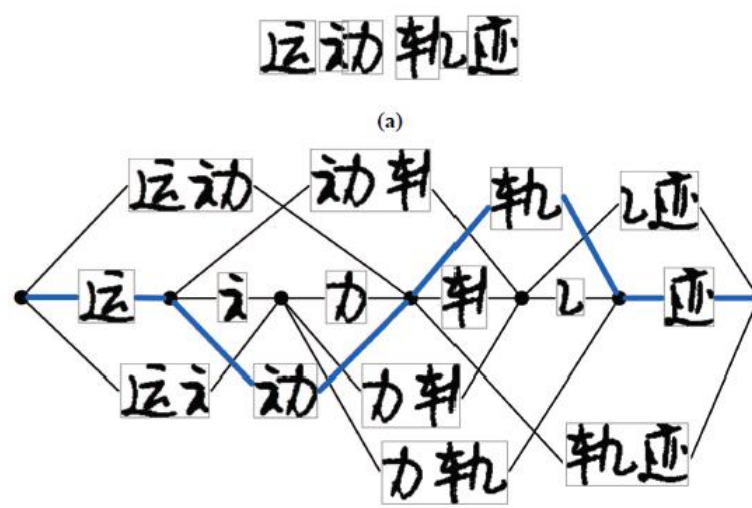
    - RNN, BLSTM
      (bidirectional long short-term memory)
    - BLSTM combined with CNN (CRNN)
    - Sliding window classifier
      (Applicable to large category set)



基于过切分的手写文本行识别

- Candidate segmentation-recognition path evaluation
- CNN for cut detection, character classification, geometric context
- RNN-based language model (character based)

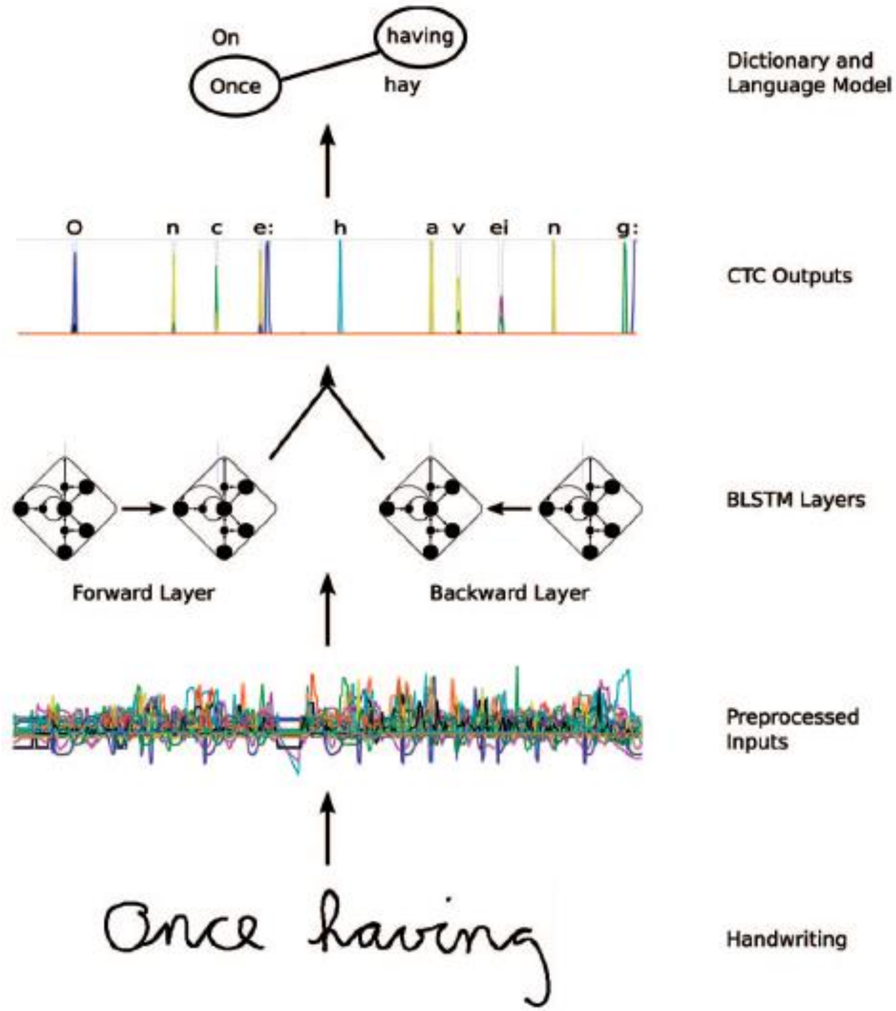




Y.-C. Wu, F. Yin, C.-L. Liu, Improving Handwritten Chinese Text Recognition Using Neural Network Language Models and Convolutional Neural Network Shape Models, *Pattern Recognition*, 2017.

基于循环神经网络的文本识别

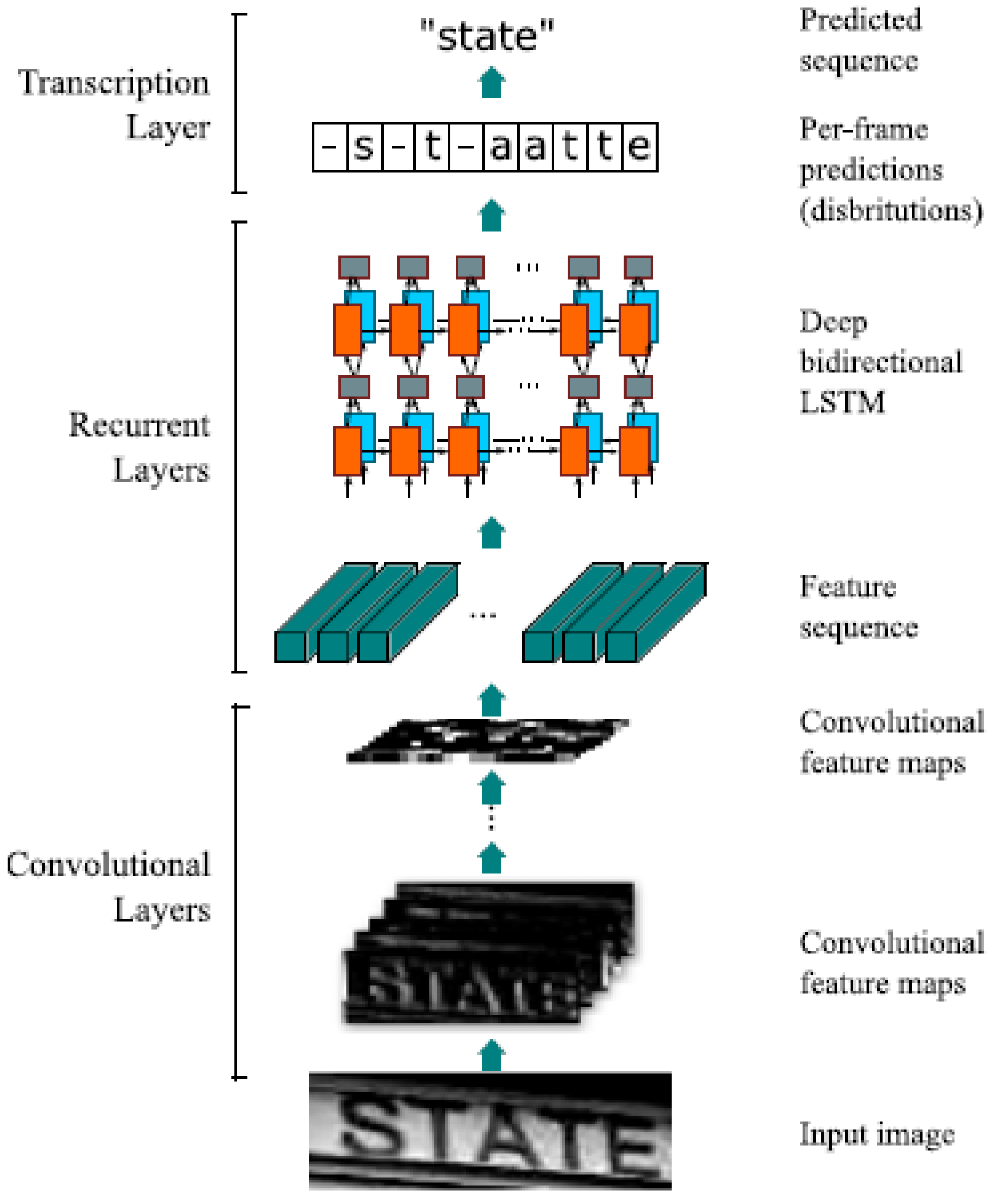
- LSTM (long-short-term memory) units to better model long-range dependency
- Decoding by CTC (connectionist temporal classification)
- Superior performance in text recognition of various styles (online/offline handwriting, printed, scene texts)



A. Graves, M. Liwicki, S. Fernandez, R. Bertonami, H. Bunke, J. Schmidhuber, [A novel connectionist system for unconstrained handwriting recognition](), *IEEE Trans. PAMI*, 2009.

基于卷积循环神经网络的文本识别

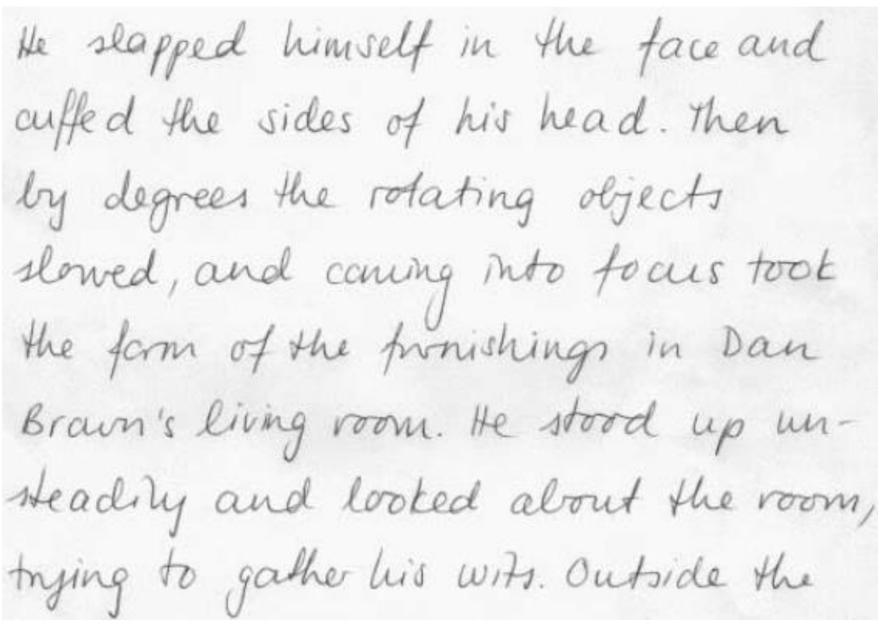
Conv-RNN (LSTM) now dominates in text recognition

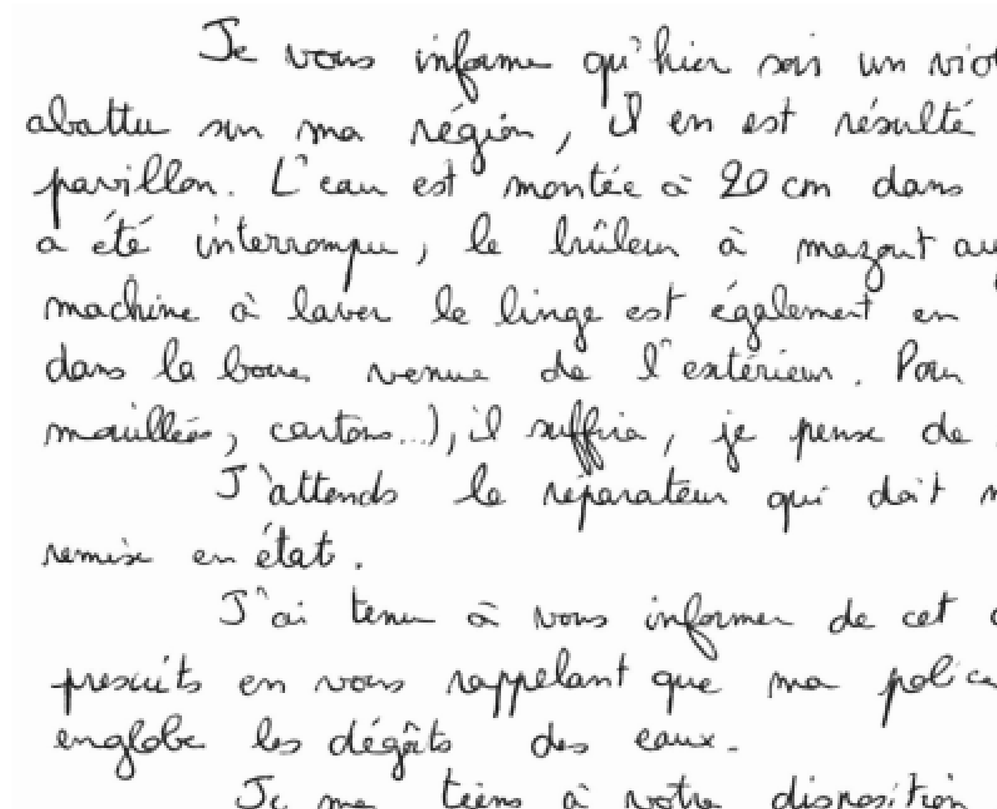


B. Shi, X. Bai, C. Yao, [An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition](), *IEEE Trans. PAMI*, 2017.

手写文本识别性能

- Datasets
  - IAM (University of Bern, Switzerland)
    - English paragraphs, 6486/972/2915 lines in training/validation/test
  - RIMES Database (French handwriting)
    - 12,093 lines





Results on IAM Dataset

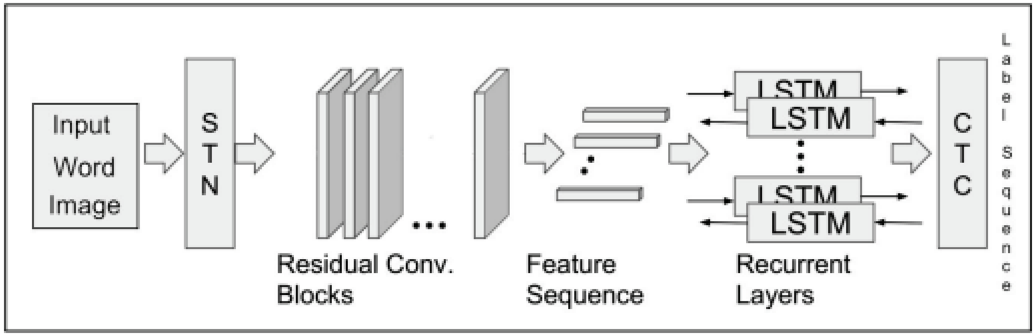
Method	Seg.	Decoding	WER	CER
Krishnan et al. [35]	Word	Unconstrained	16.19	6.34
Wigington et al. [18]			19.07	6.07
Sueiras et al. [14]			23.8	8.8
This Work			**12.61**	**4.88**
Sun et al. [15]		Full-Lexicon	11.51	-
Wigington et al. [18]			5.71	3.03
Stuner et al. [25]			5.93	2.78
Poznanski et al. [20]			6.45	3.44
This Work			**4.80**	**2.52**
Sueiras et al. [14]		Test-Lexicon	12.7	6.2
Wigington et al. [18]			4.97	2.82
Krishnan et al. [21]			6.69	3.72
Krishnan et al. [35]			5.10	2.66
This Work			**4.07**	**2.17**
Pham et al. [16]	Line	Unconstrained	35.1	10.8
Puigcerver et al. [19]			18.4	5.8
Chen et al. [17]			34.55	11.15
Krishnan et al. [35]			32.89	9.78
This Work			**17.82**	**5.7**

Results on RIMES Dataset

Method	Seg.	Decoding	WER	CER
Wigington et al. [18]	Word	Unconstrained	11.29	3.09
Sueiras et al. [14]			15.9	4.8
This Work			**7.04**	**2.32**
Wigington et al. [18]		Comp. Lexicon	2.85	1.36
Sueiras et al. [14]			6.6	2.6
Stuner et al. [25]			3.48	1.34
Poznanski et al. [20]			3.90	1.90
This Work			**1.86**	**0.65**
Pham et al. [16]	Line	Unconstrained	28.5	6.8
Chen et al. [17]			30.54	8.29
Puigcerver et al. [19]			**9.6**	**2.3**
This Work			14.70	5.07

Lexicon matters a lot!

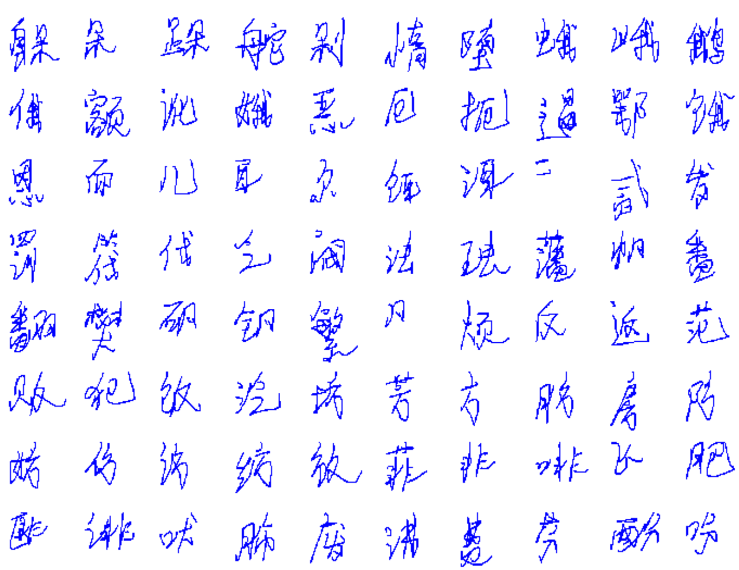
Improvements: Pre-processing, pre-training, data augmentation

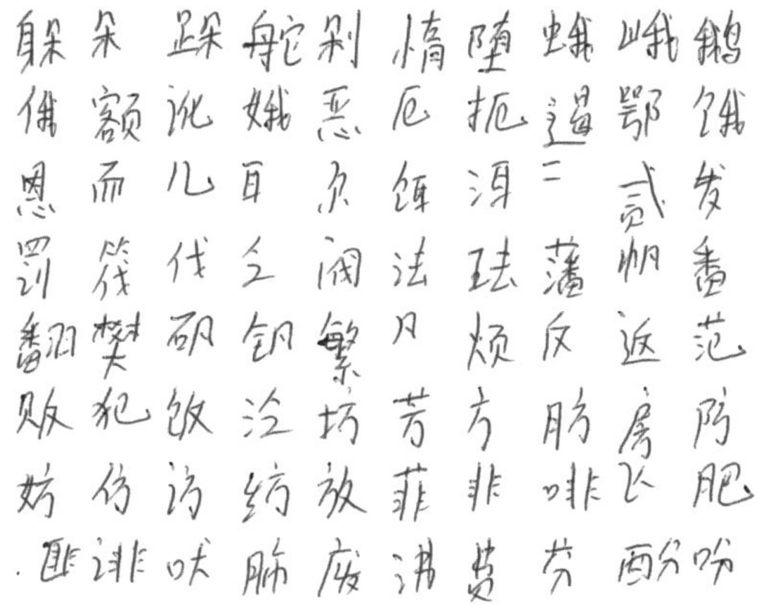


K. Dutta, P. Krishnan, M. Mathew, C.V. Jawahar, Improving CNN-RNN hybrid networks for handwriting recognition, ICFHR 2018.

中文手写样本数据集

- Handwritten Chinese Characters
  - CASIA OLHWDB/HWDB





  - ICDAR 2013 competition
    - Isolated: 3,755 classes
    - HWDB1.0+HWDB1.1 for training
    - Data of 60 writers in testing

- # 手写汉字识别性能
  - ICDAR 2013 competition

Table 4. Results of online character recognition (%).

System	CR (1)	CR (10)	Ave time	Dic size
UWarwick	**97.39**	**99.88**	355ms	37.8M
VO-3	96.87	99.67	15.3ms	87.6M*
VO-2	96.72	99.61	4.1ms	36M*
VO-1	96.33	99.61	1.6ms	10M*
HIT	95.18	99.39	2.3ms	120M
USTC-2	94.59	99.14	3.8ms	5.25M
USTC-1	94.25	99.06	2.0ms	3.19M
TUAT	93.85	99.24	5.3ms	96.2M
Faybee	92.97	98.87	0.5ms	4.48M
Ref [1]	95.31			
Human	95.19			

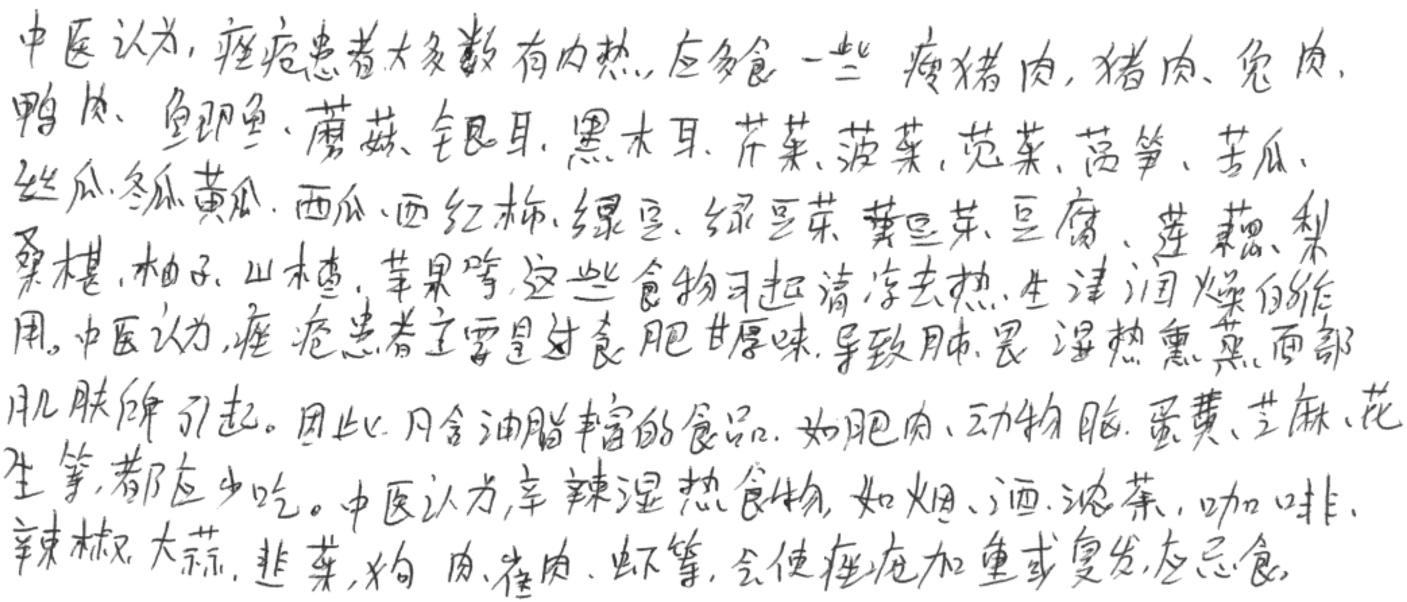
CASIA (PR2017)	dirMap+CNN	97.55	295ms	23.5M
	Ensemble-3	97.64		
SCUT (PRL2017)	CNN+DD+PS	97.55	295ms	23.5M
	Model acerage	97.64		
CASIA (PAMI2017)	RNN	97.89		10.38M
	Ensemble-6	**98.15**		78.11M

Offline character recognition

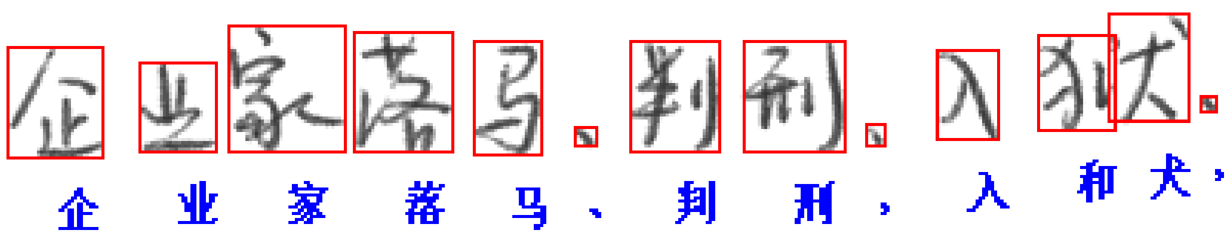
	System	CR (%)	Speed (ms)
ICDAR2013 Competition	Fujitsu, CNN	94.77	55 (GPU)
	IDSIAnn (8)	94.42	315 (CPU)
	IDSIAnn-1	94.24	197 (CPU)
	HIT	92.62	4.6 (CPU)
	Human	96.13	
IDSIA Tech Rep 05-13	CNN	94.47	3.03 (GPU)
	Multi-CNN (8)	95.78	22.04 (GPU)
Fujitsu (ICFHR2014)	ATR-CNN	95.04	
	CNN voting	96.06	
CASIA (PR2017)	dirMap+CNN	96.95	298 (CPU)
	Ensemble-3	97.12	
SCUT (PR2017)	CNN	97.30	1368
	compressed	97.09	9.7
CASIA (PR'19)	Lightweight CNN	97.19	2.8

Isolated character recognition is solved very well based on deep learning.

- 中文手写文本识别
  - ICDAR2013 competition: given text line segmentation



Performance metric: character correct rate (CR), accurate rate (AR)



ICDAR2013 Competition on Chinese Handwritten Text Recognition

Offline

Table 5. Results of offline text recognition (%).

	CR	AR	Ave time	Dic size
HIT-2	**88.76**	**86.73**	1.2s	309M
HIT-1	86.15	83.58	0.64s	111M
THU	82.92	79.81	0.85s	102M
SCUEC	42.05	35.14	0.15s	442M
Ref[6]	90.22	89.28		
Wang&Du	93.27		DNN-HMM	
ICFHR'16	94.86		Writer adaptation	
Fujitsu'16	95.53	94.02	Over-seg, CNN	
CASIA'17	96.32	96.20	Over-seg, CNN	
Jin et al'19	**96.70**	**96.22**	CRNN, ACE	

Online

Table 6. Results of online text recognition (%).

	CR	AR	Ave time	Dic size
VO-3	**95.03**	**94.49**	1.72s	56M*
VO-2	94.94	94.37	1.23s	37.9M*
VO-1	93.11	92.57	0.72s	20.8M*
TUAT	88.49	87.66	1.42s	246M
USTC	82.20	81.57	0.25s	29.3M*
Ref [29]	94.62	94.06		
* Su et al'16	94.43	93.40	Deep BLSTM	
Jin et al'17	**96.58**	**96.09**	MC-FCRN	

Over-segment and NN classification

研究现状总结

- 深度学习带来识别性能的突破
- 单字识别：精度提升空间有限，更可靠的识别结果依赖上下文和知识
- 文本行识别：字符切分+识别已不是问题
- 遗留问题
  - 大样本学习不是总有条件
  - 学习样本不是一次提供，数据风格变化时需要自适应
  - 增量学习时类别数动态增加
  - 有些类别样本极少甚至没有
  - 字符识别置信度、结构解释

单字识别研究意义

- 文本行识别中的作用
  - 基于过切分/滑动窗的识别方法：单字识别器性能起决定作用
  - 切分后进一步提高识别精度、结构解释
- 文本行数据集中类别集有限
  - 大量类别只有单字样本
  - 比如，CASIA-HWDB单字类别7356，文本行字符类别2703
- 类别增量学习、零样本学习
  - 在单字识别器上进行更方便，学习后融入文本行识别器

零样本文字识别

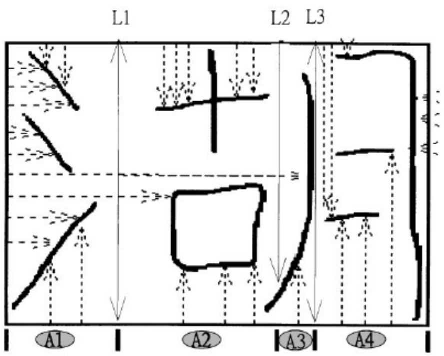
- 训练数据集中类别数有限
  - GB2312-80
    - Level-1: 3,755
    - Level-2: 6,763
  - CASIA-HWDB: 7,185+171
- 古籍
  - 大量字符没有样本
  - 大量异体字

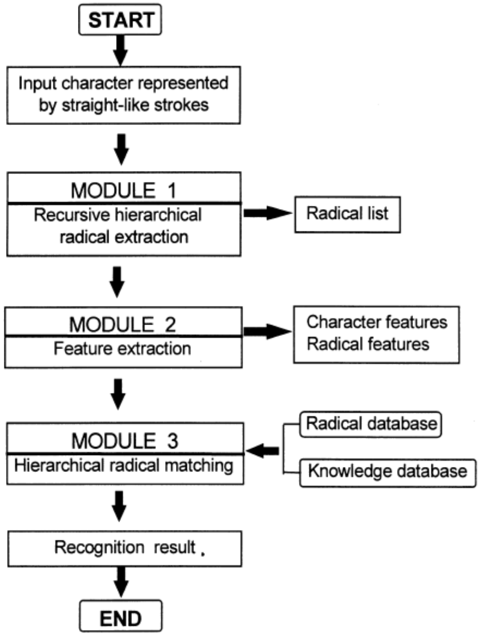
- 零样本文字识别
  - 用部分类别（比如简体字）的样本训练识别器，推广到识别新类别（比如繁体字）样本

基于部首检测的方法

- 传统方法
  - 笔划分析，基于规则的匹配

A.-B. Wang, K.-C. Fan, Optical recognition of handwritten Chinese characters by hierarchical radical matching method, *Pattern Recognition*, 2001.



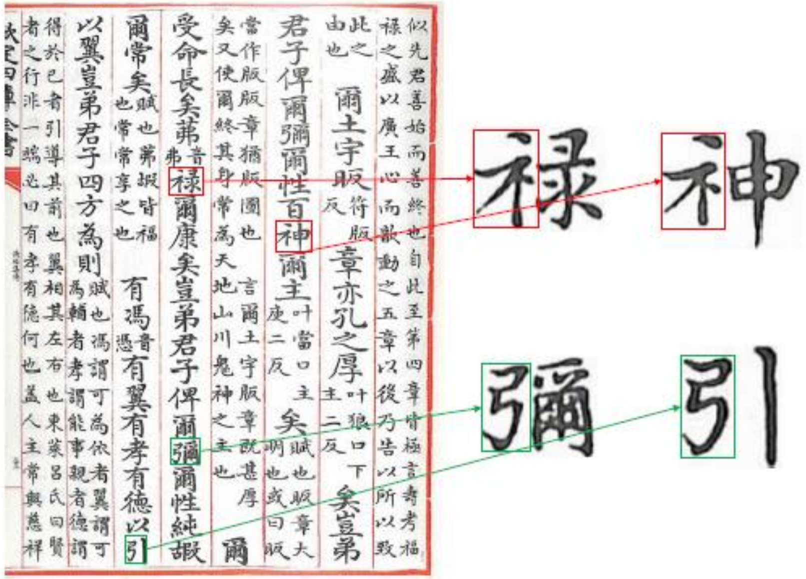


  - 难点
    - 笔划提取困难，抗形变能力差
    - 规则知识要人工设定
- 基于学习的方法
  - 基于深度神经网络的部首检测

基于CNN和多标记学习的部首检测

训练样本中标出位置相关部首类别(不需要标出位置)。部首类别较多，因为相同部首在不同位置视作不同类别（如：郁，随）。

- Currently experimented on printed characters
- Residual network by multi-labeled learning for position-dependent radical detection



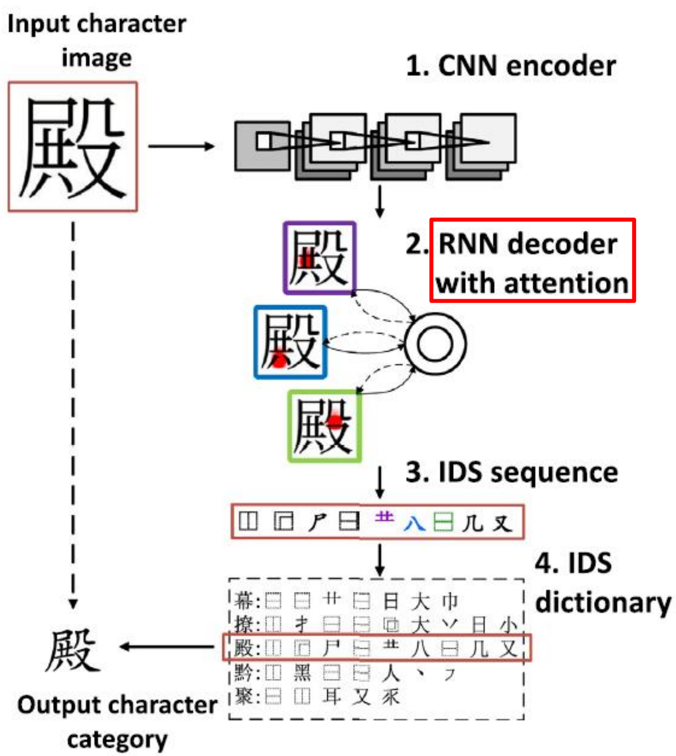
Model Version	Character Set	Character Number	Radical Number	Training Set Scale	Testing Set Scale	Accuracy	Recall	Precision
RD-Net-v1	HF SE	3755	1718	901290	225213	99.43%	99.58%	99.63%
RD-Net-v2	HF SE LF	6363	1869	1427125	375643	99.10%	99.15 %	99.17%
RD-Net-v3	HF SE LF RU	7697	1889	1675464	389544	98.92%	99.01%	99.04%
RD-Net-v4	US	9820	2018	2386513	558762	98.85%	98.90%	98.94%

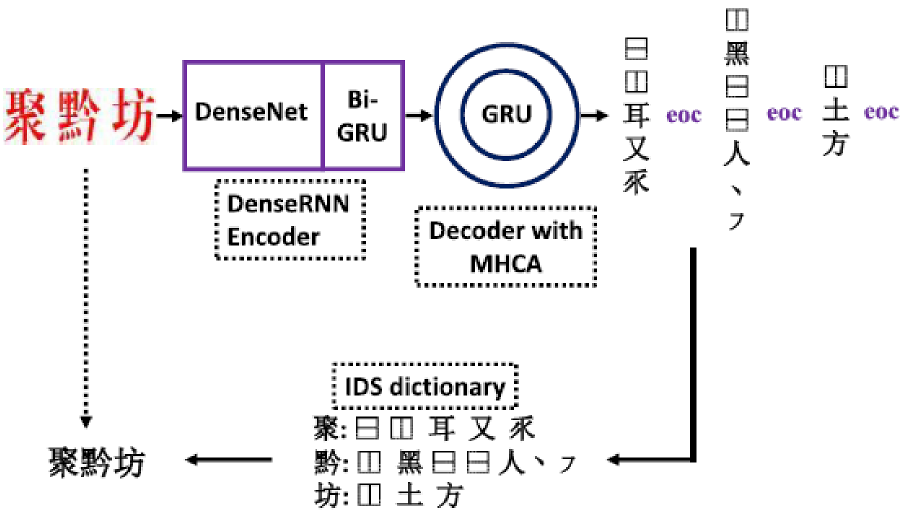
Testing on novel characters (zero-shot learning)

Net Version	Training Set	Testing Set	Testing Set Scale	Accuracy	Recall	Precision
RD-Net-v1	HF SE	LF RU	126372	98.67%	98.73%	98.71%
RD-Net-v2	HF SE LF	RU TC	117875	97.35%	97.52%	97.40%
RD-Net-v3	HF SE LF RU	TC	79024	97.93%	98.11%	97.95%

Wang & Liu, Radical-Based Chinese Character Recognition via Multi-Labeled Learning of Deep Residual Networks, ICDAR 2017.

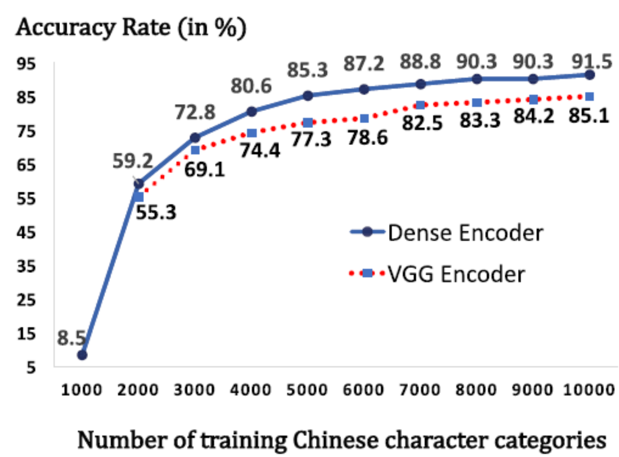
Radical Analysis Network (RAN)





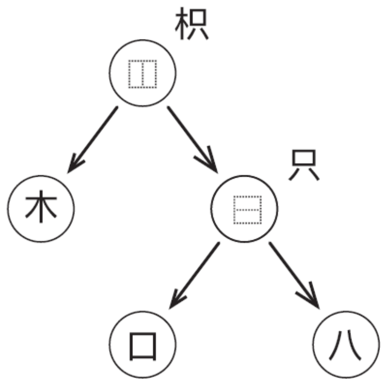
Extension to text line recognition

Experiments: 27,533 characters composed of 485 radicals, 10,000 in training, the others (unseen) in testing

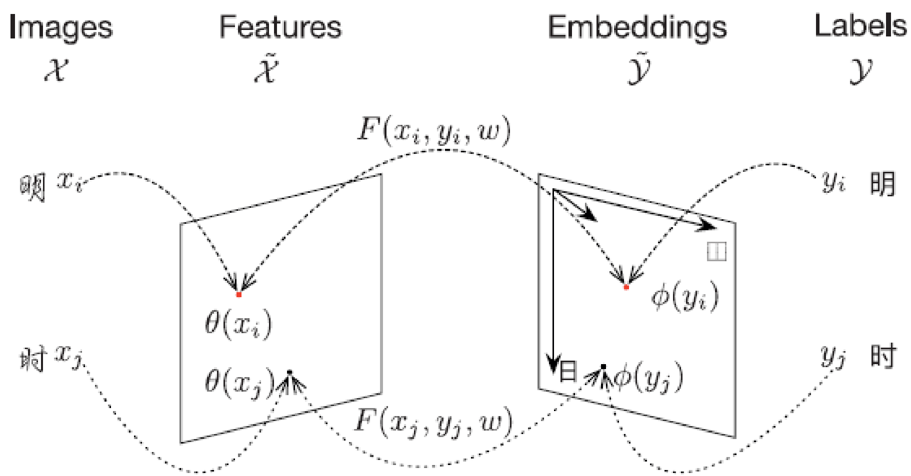


J. Zhang, J. Du, L. Dai, [Radical analysis network for learning hierarchies of Chinese characters](), *Pattern Recognition*, 2020.

基于树结构嵌入的方法



Tree layout of primitives



- **Embedding**: one-hot encoding for each node, then combine all vectors in one tree to the hierarchical decomposition embedding (HDE).
- Convolutional Neural Network (CNN) based framework to learn both radicals and structures of characters via the semantic vector.
- Can recognize unseen characters given the tree layout.

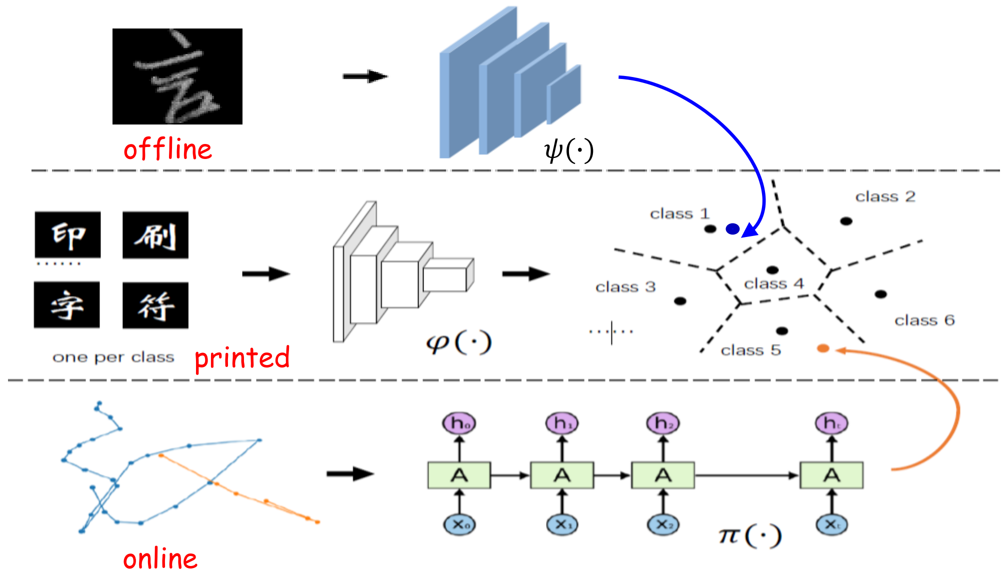
Performance comparisons on level-1 unseen handwritten characters with different methods.

train	test	DenseRAN [8]	FewshotRAN [4]	Ours
500	1000	1.70%	33.6%	***33.71%***
1000	1000	8.44%	41.5%	***53.91%***
1500	1000	14.71%	63.8%	***66.27%***
2000	1000	19.51%	70.6%	***73.42%***
2755	1000	30.68%	77.2%	***80.95%***

Z. Cao, J. Lu, S. Cui, C. Zhang, Zero-shot handwritten Chinese character recognition with hierarchical decomposition embedding, *Pattern Recognition*, 2020.

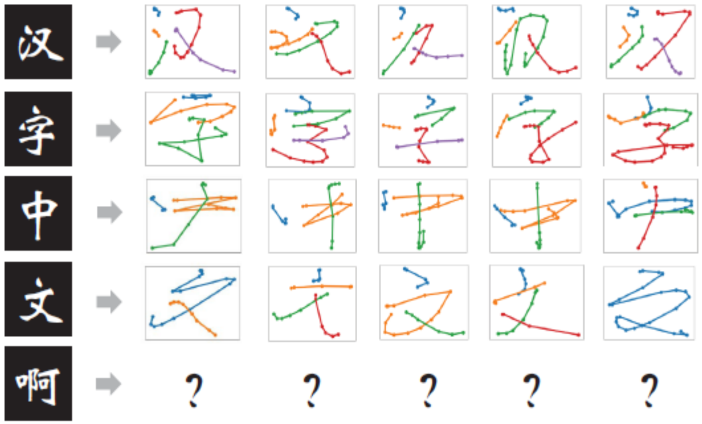
基于印刷体原型的手写汉字识别

- Large set Chinese characters, hard to get samples for all categories
- Recognize new-class handwritten samples using printed template
- Feature space learning to make handwritten samples close to printed template

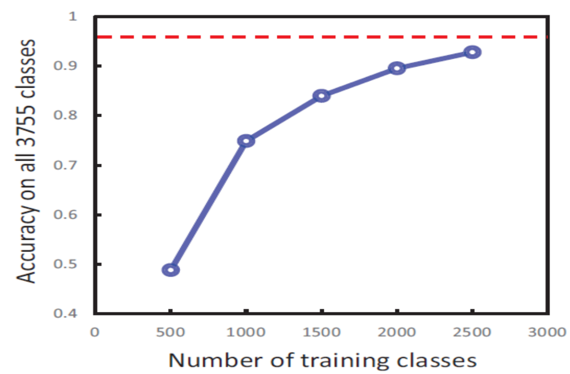


  - 对新类别，只需提供一个印刷样本，无需手写样本训练就能识别手写字符
  - 根据最近原型准则来分类 $x \in \arg\min_{I \in P} \|\pi(x) - \varphi(I)\|_2^2$

Ao, Zhang, Liu, Cross-modal prototype learning for zero-shot handwriting recognition, ICDAR 2019.



Results of bi-modal (printed, online) learning



Results of tri-modal (printed, online, offline) learning

N Training classes	Seen		Unseen		All	
	Online	Offline	Online	Offline	Online	Offline
500	0.92	0.86	0.70	0.72	0.73	0.74
1000	0.95	0.89	0.84	0.81	0.87	0.83

Unseen: un-trained classes, printed template only.

此方法可推广到古籍文字识别

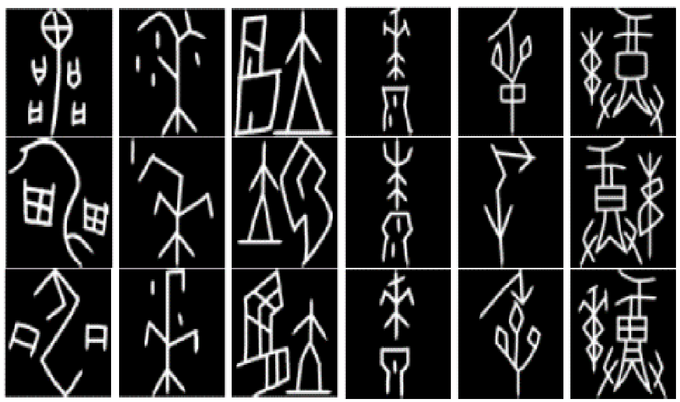
甲骨文字识别

- 中国文字起源和演化

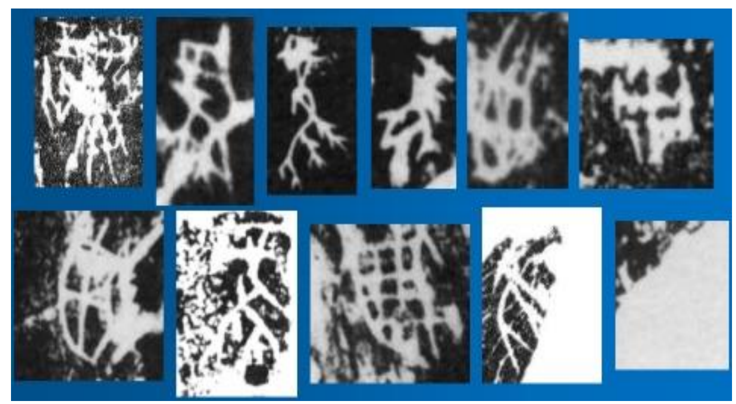
- 甲骨文
  – 镌刻、书写于龟甲与兽骨上而得名，殷商时期流传，内容主要为卜辞
  – 1899年首次考古发现
  – 截至2012年，发现有大约15万片甲骨，4500多个单字。已识别2500多个单字
- 研究甲骨文字识别的意义
  – 帮助现代人识读
  – 文字考古：需要结合大量历史文化知识

基于跨模态度量学习的甲骨文字识别

- 甲骨文字识别
  - 真实（拓片）样本少，手描（临摹）样本多
  - 用临摹甲骨文样本辅助拓片甲骨文识别
  - 识别无训练样本的拓片甲骨文字（新类识别）



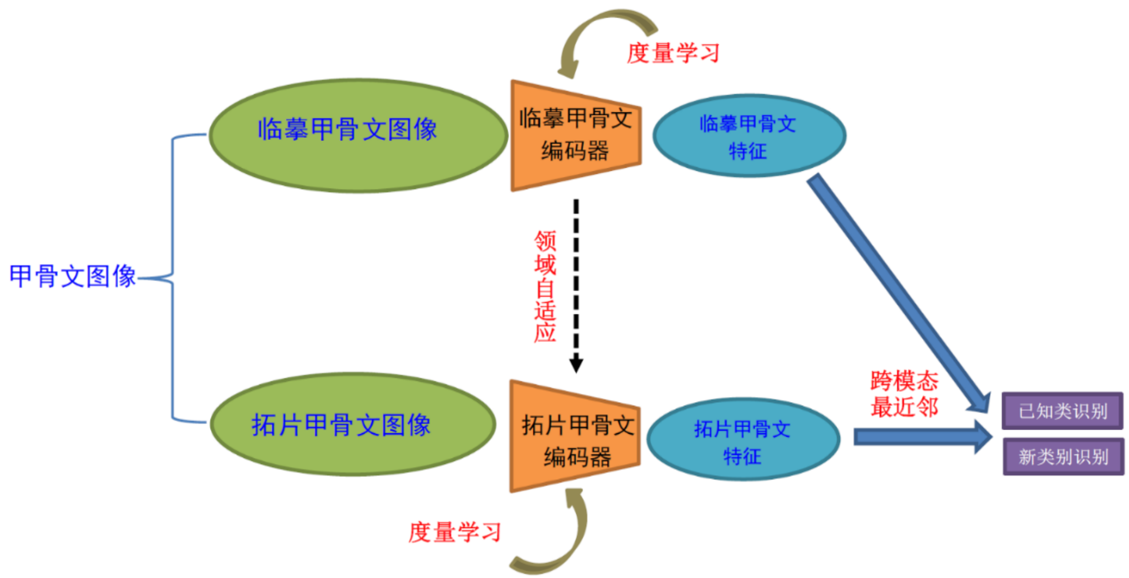
临摹甲骨文字（样本充足）
类内差异大(同一列为一个字)



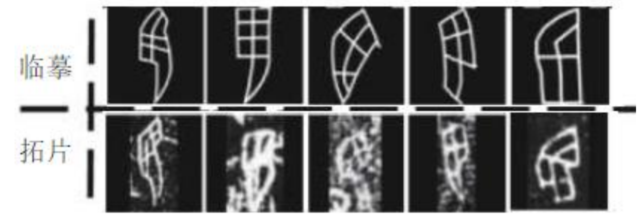
拓片甲骨文（样本少）

张颐康、张恒、刘永革、刘成林，基于跨模态深度度量学习的甲骨文字识别，自动化学报，2021年第4期
（合作单位：安阳师范学院甲骨文信息处理教育部重点实验室）

系统流程



领域自适应后
同类样本特征相近

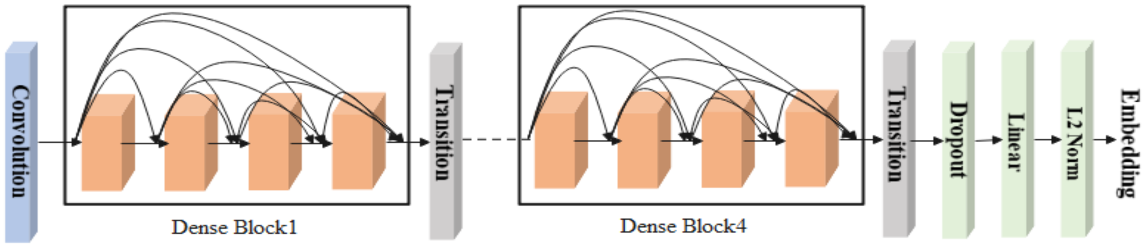


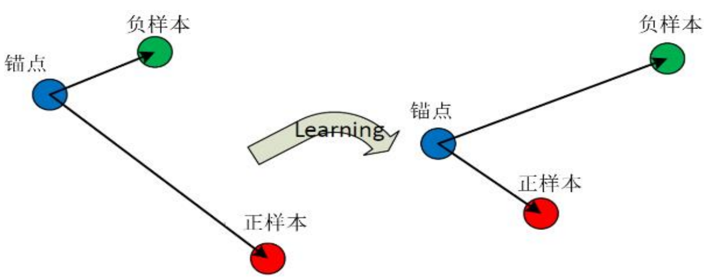
主要步骤：

1. 临摹甲骨文字特征编码（度量学习）
2. 领域自适应（对抗训练）
3. 拓片甲骨文字特征修正（度量学习）
4. 跨模态最近邻分类

- 临摹甲骨文字特征编码器训练
  - DenseNet作为甲骨文字特征编码器
  - 度量学习：三元组损失函数对模型进行参数优化

编码器





$$\left\|f(x_a^i)-f(x_p^i)\right\|_2^2+\alpha<\left\|f(x_a^i)-f(x_n^i)\right\|_2^2$$

$$\forall\left(f(x_a^i),f(x_p^i),f(x_n^i)\right)\in\tau$$

三元组损失：使得同类样本（正样本）尽可能靠近，不同类样本（负样本）尽可能远离

- 拓片甲骨文字特征编码器训练

（1）基于对抗训练的领域自适应：不同模态同类别特征距离更近

– 对抗训练：Wasserstein GAN框架，判别器与生成器迭代训练，使每类样本拓片文字和临摹文字特征尽可能靠近

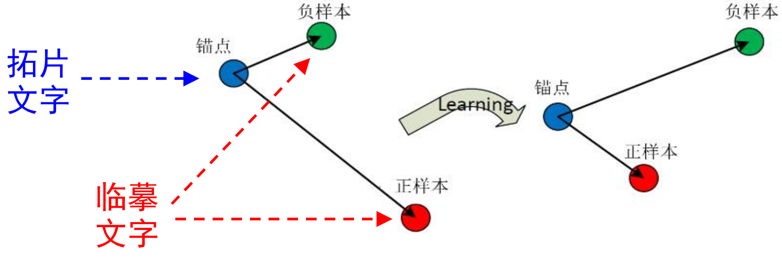
对抗损失（判别器） 第c类拓片文字和临摹文字后验概率分别为P^c_g, P^c_r

$$L_{dis}^{c}=E_{\tilde{x}\sim P_{g}^{c}}\left[D^{c}(\tilde{x})\right]-E_{\mathrm{x}\sim P_{r}^{c}}\left[D^{c}(\mathrm{x})\right]+\underbrace{\lambda E_{\hat{x}\sim P_{\hat{x}}^{c}}\left[\left(\left\|\nabla_{\hat{x}}D^{c}(\hat{x})\right\|_{2}-1\right)^{2}\right]}_{D_c(x)\text{在特征空间平滑地变化}}$$

生成损失 $$L_{\mathrm{gen}}^{c}=-E_{\tilde{x}\sim P_{g}^{c}}\left[D^{c}(\tilde{x})\right]$$

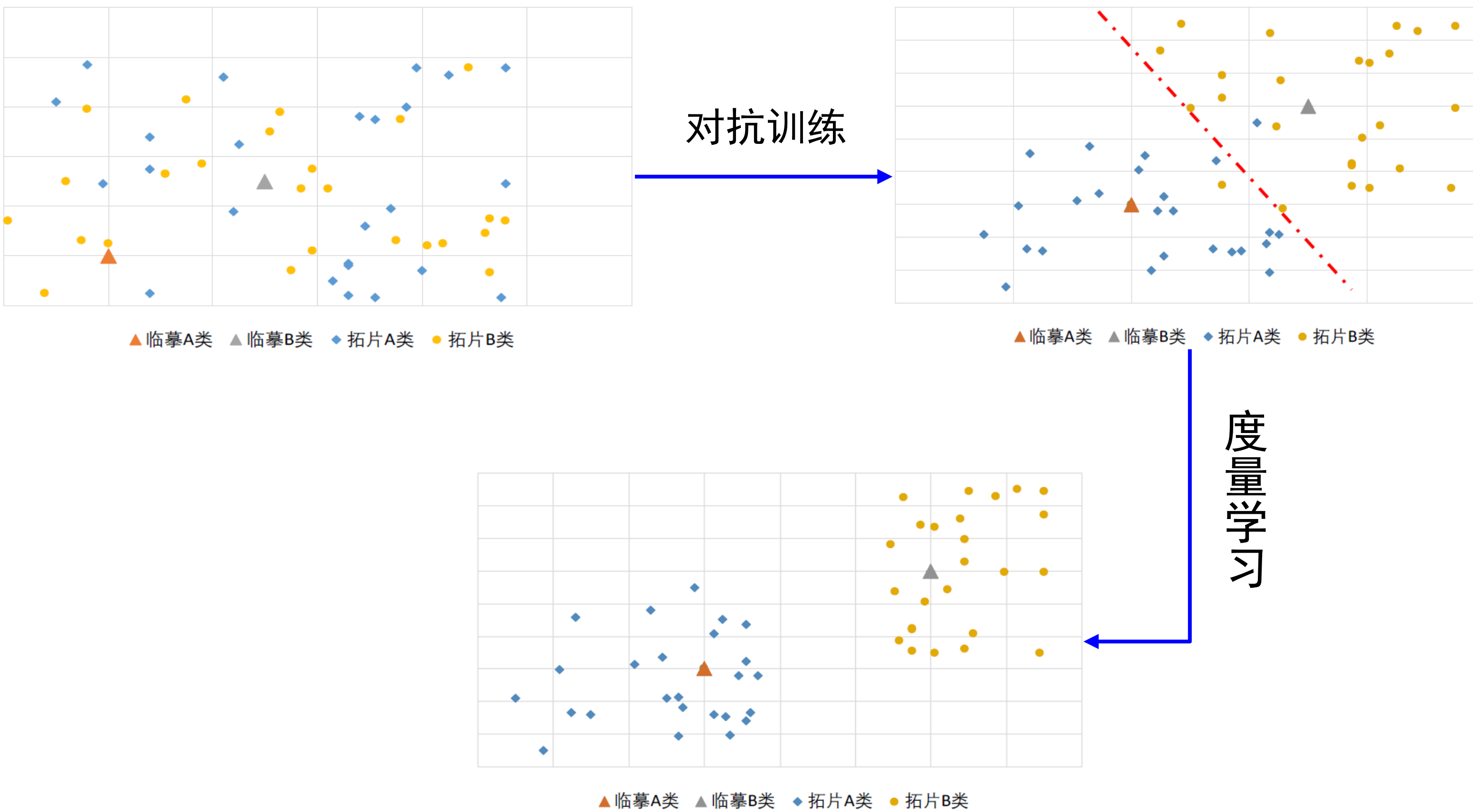
（2）跨模态度量学习：不同模态的特征中同类更近、异类更远

拓片甲骨文字为锚点样本，正样本和负样本来自临摹甲骨文字



- # 拓片甲骨文字特征编码器训练

效果示意图



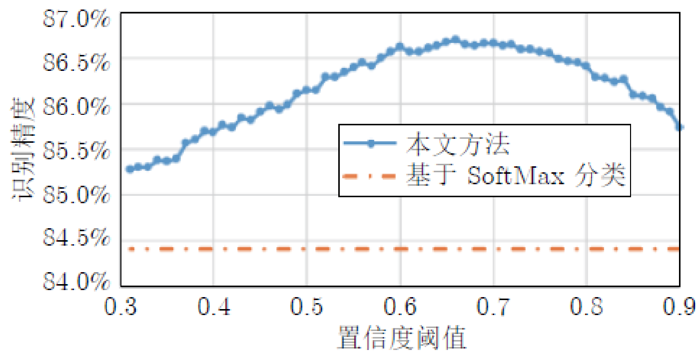
• 实验结果

- 拓片甲骨文字241类，295,466个样本，每类最少16个，最多25898个
- 临摹甲骨文字2583类，39062个样本，每类最少2个，最多287个。使用其中241类样本实验。

拓片甲骨文字分类精度对比

方法	识别率(%)
单模态最近邻	74.14
单模态CNN	84.40
跨模态最近邻	82.10
融合跨模态信息的CNN	**86.70**

融合方法：当CNN输出置信度小于一个阈值时，用跨模态K紧邻分类



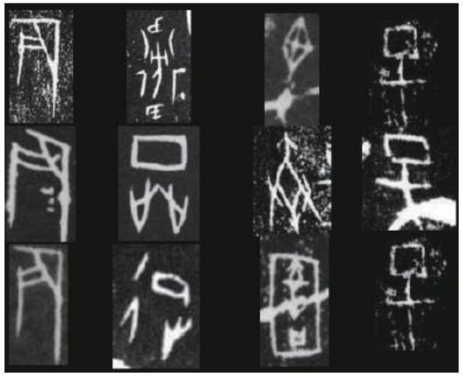
• 实验结果

零样本识别：200类训练，41类样本作为未知类

新类别拓片甲骨文字识别

特征学习方法	跨模态近邻分类精度(%)
度量学习+领域自适应	43.67
度量学习+领域自适应+特征修正	**62.10**

识别错误的主要原因：
- 拓片文字图像质量差
- 同一类样本写法变化多



每一列为同一类文字

未来研究方向：
- 综合利用部首和跨模态信息
- 同一类不同写法的细致标注
- 利用更多的语义知识（如字义解释）

讨论与展望

- 单字识别仍有研究意义
- 小样本、增类学习、零样本学习有需求
  - 古籍识别尤其如此
- 汉字零样本识别取得初步成效
  - 部首检测，部首结构嵌入，跨模态距离度量学习
  - 识别精度还不高
- 未来研究方向
  - 结合多种辅助信息或先验知识的零样本识别
  - 文本行识别中的小样本、增类学习、零样本识别
  - 文字考古：结合大量历史文化知识，高度复杂的语义分析和推理问题

谢谢聆听！

敬请批评指正