International Summer School on Document Analysis
Islamabad, Pakistan, August 19-23, 2019

Introduction to the Field of Document Analysis and Recognition

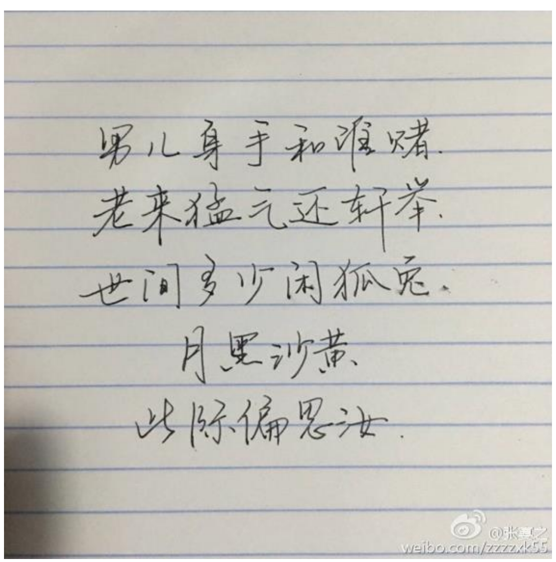
Cheng-Lin Liu
National Laboratory of Pattern Recognition
Institute of Automation of Chinese Academy of Sciences
liucl@nlpr.ia.ac.cn
http://www.nlpr.ia.ac.cn/liucl/

Outline

- DAR Introduction
- DAR Problems and History
- Academic Resources
- Major Approaches
  - Layout Analysis
  - Scene Text Detection
  - Text Line Recognition
  - Graphics Recognition
- Status of Performance
- Future Directions

Document Analysis and Recognition (DAR)

- Objective
  - Convert images to texts to enable semantic understanding
  - Image data reduction
- Related Problems
  - Layout analysis (page segmentation)
  - Character and text recognition
  - Document retrieval, semantic analysis and applications

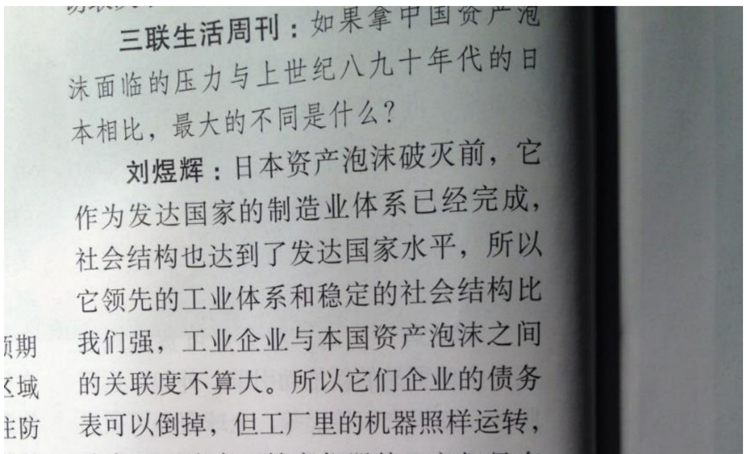


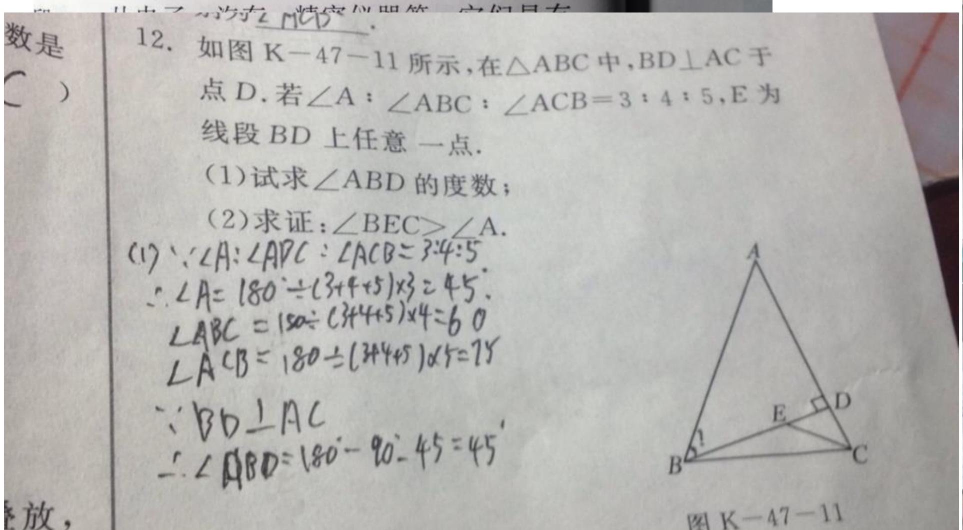
DAR

男儿身手和谁赌
老来猛气还轩举
世间多少闲狐兔
月黑沙皇
此际偏思汝

Many Sources of Document Images

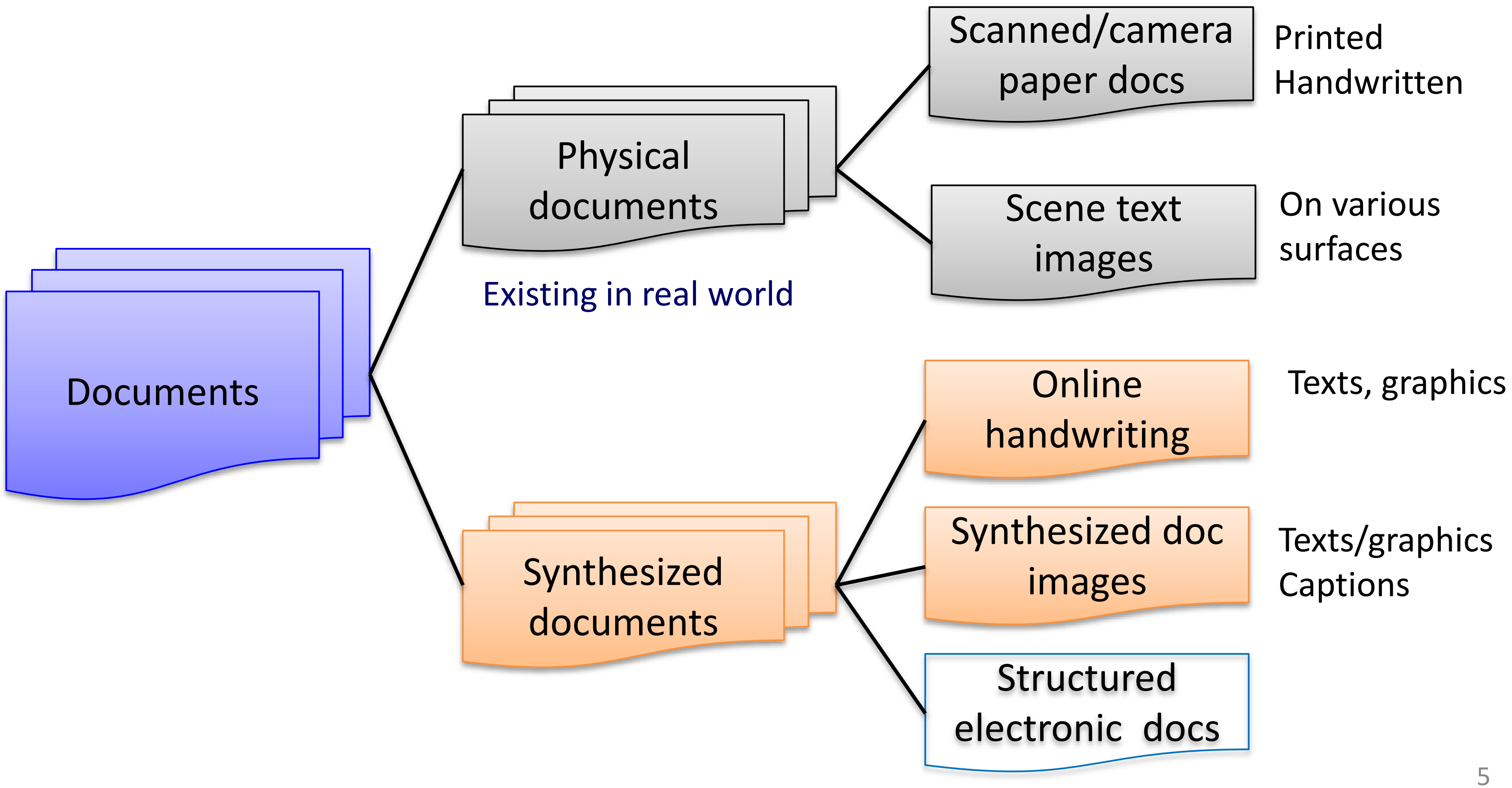
- Printed, online/offline handwritten
- Scanned, camera captured (scene text), born-digital





Categories of Documents

- What is Document
  - Media (paper, image or file) carrying texts and symbols



Application Needs

- Pervasiveness of Document Images
  - Books, magazines, newspapers, letters
  - Web data, emails, mini blogs (Weibo), Weixin
  - Sign boards, license plate, street numbers
  - Forms, certificates, business cards, id numbers
  - Pen-based writings and drawings
- When CR/DA needed
  - Text input to computers (Online)
  - Document image converted to text (Offline)
  - Semantics extraction from documents

- Online Character Recognition
  - Pen-based input, particularly suitable for mobile devices without keyboard
  - No disturbance compared to speech
  - Frontiers: continuous handwriting recognition, mixed text/graphics

Tablet PC

Anoto Pen

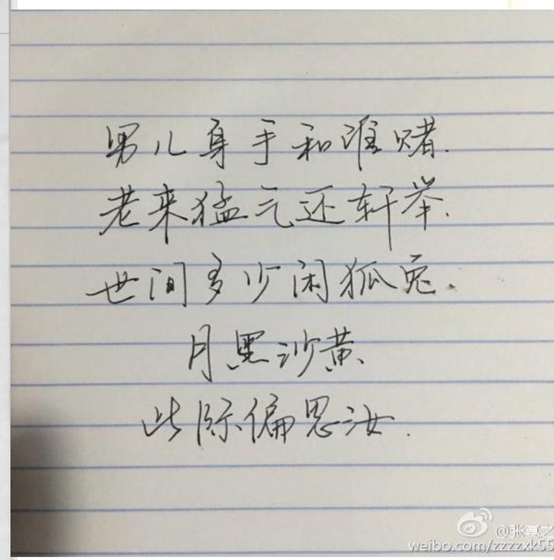
- Offline Document Recognition
  - Printed: not completely solved
    - Challenges: complex layout, degraded image, mathematics/flowchart, multi-lingual
  - Handwritten: un-solved
    - Layout analysis
    - Character segmentation and recognition

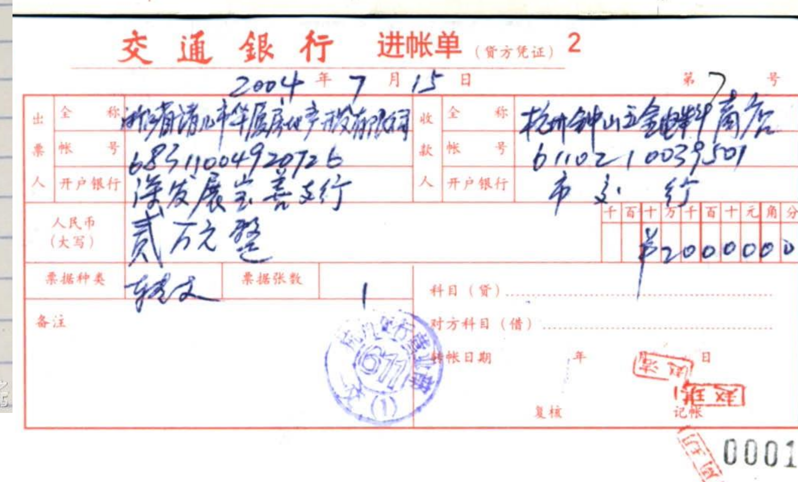
28 000 KILOMETERS

Two-Laser Lithography Shrinks Transistors on the Cheap

A new microscopy technique gets adapted for chipmaking

news briefs



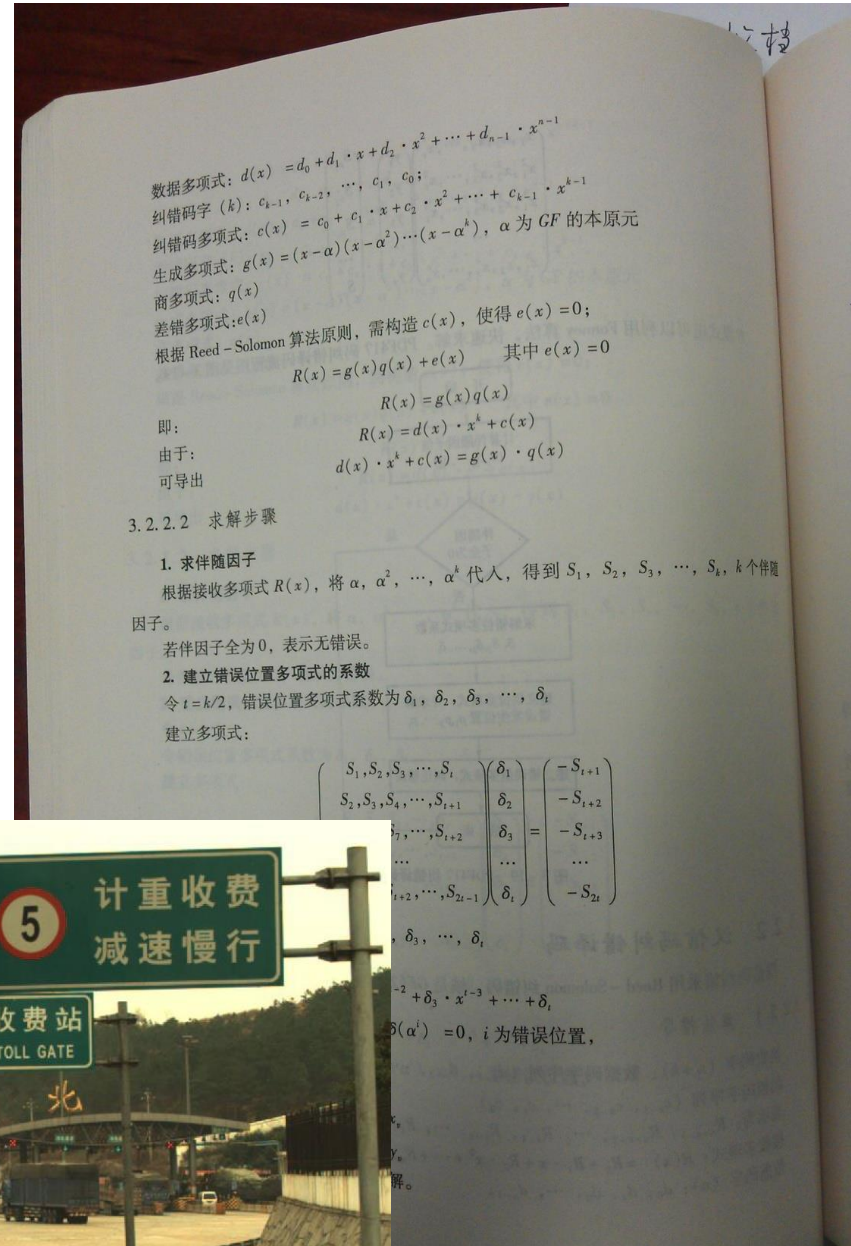


Large volumes of postal forms, bank forms and medical records

Camera-captured documents

- Increasingly captured by mobile phones and submitted to Internet
- Many challenges

Vehicle id number, Chattopadhyay et al., DAS 2012



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

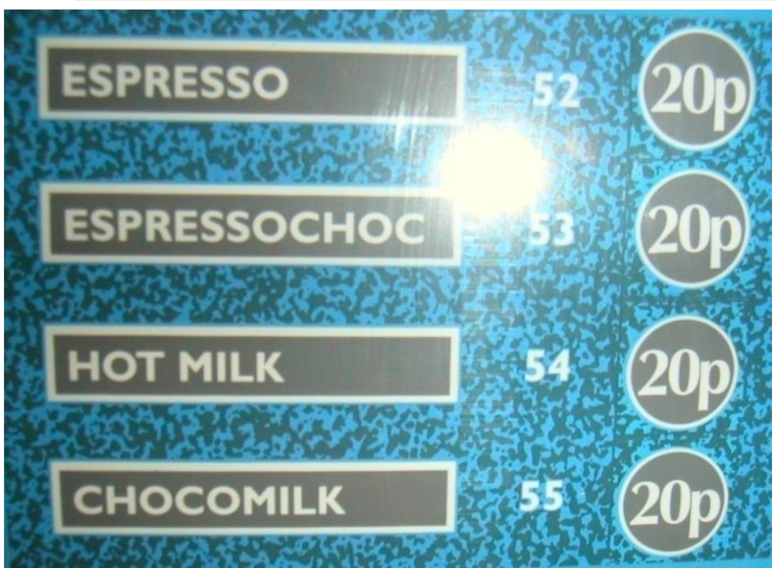
根据接收多项式 $R(x)$，将 α，α^2，…，α^k 代入，得到 S_1，S_2，S_3，…，S_k，k 个伴随因子。

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 δ_1，δ_2，δ_3，…，δ_t

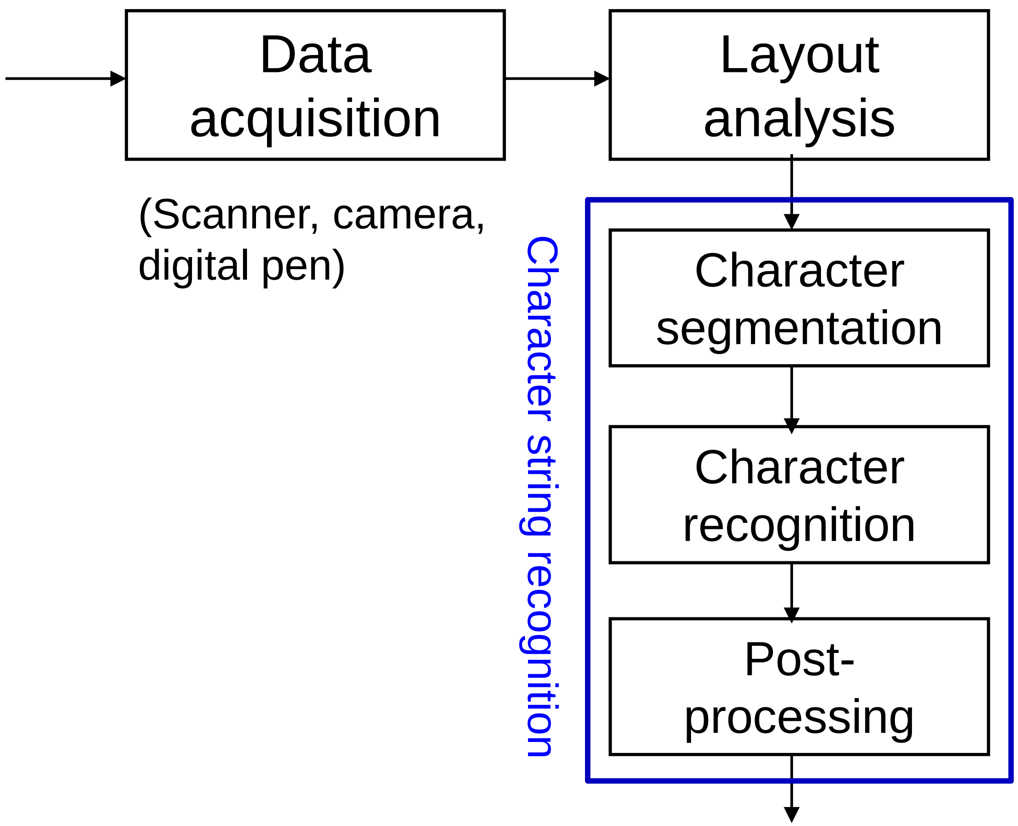
建立多项式：

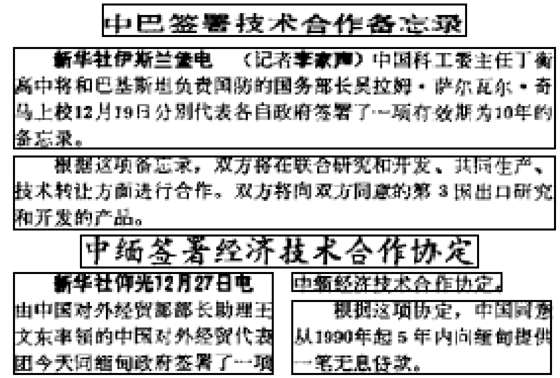


Webdata: many document images to be converted to text

Images from Weibo and Weixin (WeChat)

Document Analysis Pipeline





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。

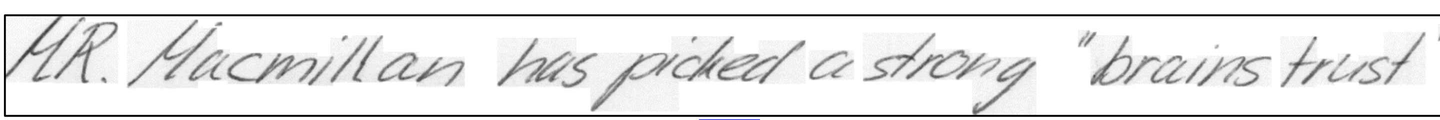
中缅签署经济技术合作协定

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”

Document Analysis Issues

- From Image to Semantics

Image Processing → Layout Analysis → Contents Recognition → Semantic/ Application

Image Processing

- Pre-classificat
  - Doc/Non-doc
- Enhancement
  - Contrast
  - denoising
- Rectification
  - Illumination
  - Skew
  - Perspective
- Binarization
- Frame removal

Layout Analysis

- Region separation
- Zone classification
- Text localization
- Text line segmentation
- Hand/print separation
- Table/form analysis
- Signature/logo/ stamp extraction

Contents Recognition

- Text recognition
  - Character segment
  - Feature extraction
  - Classifier design
  - Sequence model
  - Linguistic/geometric contexts
- Graphics/symbol
  - Diagrams
  - Eng. drawings
  - Musical scores
  - Mathematics
  - Physical/chemicals
- Style authentication
  - Font identification
  - Script identification
  - Writer identificat
  - Signature verificat

Semantic/ Application

- Structural understanding
  - Logical relation btw zones
  - Reconstruction
- Retrieval
  - Keyword spotting
  - Content-based
  - Structure-based
- Semantic analysis
  - Categorization
  - Summarization
  - Translation

Document Analysis Issues

- From Image to Semantics

→ Image Processing → Layout Analysis → Contents Recognition → Semantic/ Application →

Image Processing

- Pre-classificat
  - Doc/Non-Doc
- Enhancement
  - Contrast
  - denoising
- Rectification
  - Illumination
  - Skew
  - Perspective
- Binarization
- Frame removal

Layout Analysis

on
tion
alization
e
ation
rint
on
orm
re/logo/
traction

Contents Recognition

- Text recognition
  - Character segment
  - Feature extraction
  - Classifier design
  - Sequence model
  - Linguistic/geometric contexts
- Graphics/symbol
  - Diagrams
  - Eng. drawings
  - Musical scores
  - Mathematics
  - Physical/chemicals
- Style authentication
  - Font identification
  - Script identification
  - Writer identificat
  - Signature verificat

Semantic/ Application

- Structural understanding
  - Logical relation btw zones
  - Reconstruction
- Retrieval
  - Keyword spotting
  - Content-based
  - Structure-based
- Semantic analysis
  - Categorization
  - Summarization
  - Translation

Document Analysis Issues

- From Image to Semantics

Image Processing → Layout Analysis → Contents Recognition → Semantic/ Application →

Image Processing
- •Enhancement
- - Contrast
- - denoising
- •Rectification
- - Illumination
- - Skew
- - Perspective
- •Binarization
- •Frame removal

Layout Analysis
- •Region separation
- •Zone classification
- •Text localization
- •Text line segmentation
- •Hand/print separation
- •Table/form analysis
- •Signature/logo/stamp extraction

Contents Recognition

tion
ment
ction
gn
del
metric
bol
s
icals
cation
ation
ication
Writer identificat
- Signature verificat

Semantic/ Application
- •Structural understanding
- - Logical relation btw zones
- - Reconstruction
- •Retrieval
- - Keyword spotting
- - Content-based
- - Structure-based
- •Semantic analysis
- - Categorization
- - Summarization
- - Translation

Document Analysis Issues

- From Image to Semantics

Image Processing → Layout Analysis → … → …antic/ …ation →

Image Processing
- Enhancement
  - Contrast
  - denoising
- Rectification
  - Illumination
  - Skew
  - Perspective
- Binarization
- Frame removal

Layout Analysis
- Region separation
- Zone classification
- Text localization
- Text line segmentation
- Hand/print separation
- Table/form analysis
- Signature/logo/ stamp extraction

- Text recognition
  - Character segment
  - Feature extraction
  - Classifier design
  - Sequence model
  - Linguistic/geometric contexts
- Graphics/symbol
  - Diagrams
  - Eng. drawings
  - Musical scores
  - Mathematics
  - Physical/chemicals
- Style authentication
  - Font identification
  - Script identification
  - Writer identification
  - Signature verification

…ling
…lation
…ction
…spotting
…ased
…-based
…analysis
…ation
…ation
…n

Document Analysis Issues

- From Image to Semantics

→ Image Processing → Layout Analysis → Contents Recognition → Semantic/ Application →

Image Processing

- Enhancement
  - Contrast
  - denoising
- Rectification
  - Illumination
  - Skew
  - Perspective
- Binarization
- Frame removal

Layout Analysis

- Region separation
- Zone classification
- Text localization
- Text line segmentation
- Hand/print separation
- Table/form analysis
- Signature/logo/ stamp extraction

Contents Recognition

- Text recog
  - Characte
  - Feature e
  - Classifier
  - Sequenc
  - Linguistic contexts
- Graphics,
  - Diagrams
  - Eng. drav
  - Musical s
  - Mathem
  - Physical/
- Style aut
  - Font ider
  - Script ide
  - Writer identificat
  - Signature verificat

Semantic/ Application

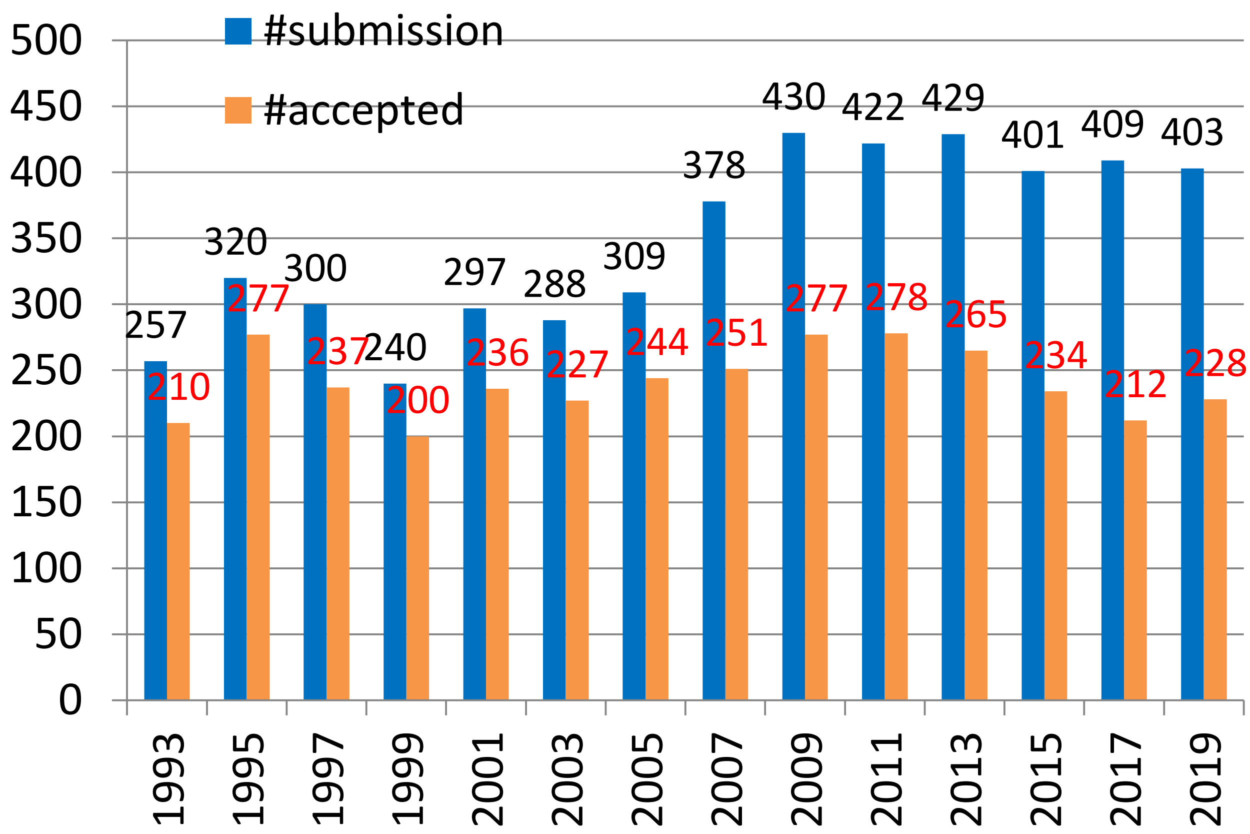
- Structural understanding
  - Logical relation btw zones
  - Reconstruction
- Retrieval
  - Keyword spotting
  - Content-based
  - Structure-based
- Semantic analysis
  - Categorization
  - Summarization
  - Translation

DAR Brief History

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, Neural networks	Handprinted digits/letters Printed/handprinted words Handprinted Japanese/Chinese	1st ICPR in 1972 IAPR founded in 1978
1990s	Research of various issues, including layout analysis and segmentation	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 New Boom	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, historical, ink documents)	Google, Baidu Facebook, twitter Smart phone Mobile Internet Weibo, WeChat

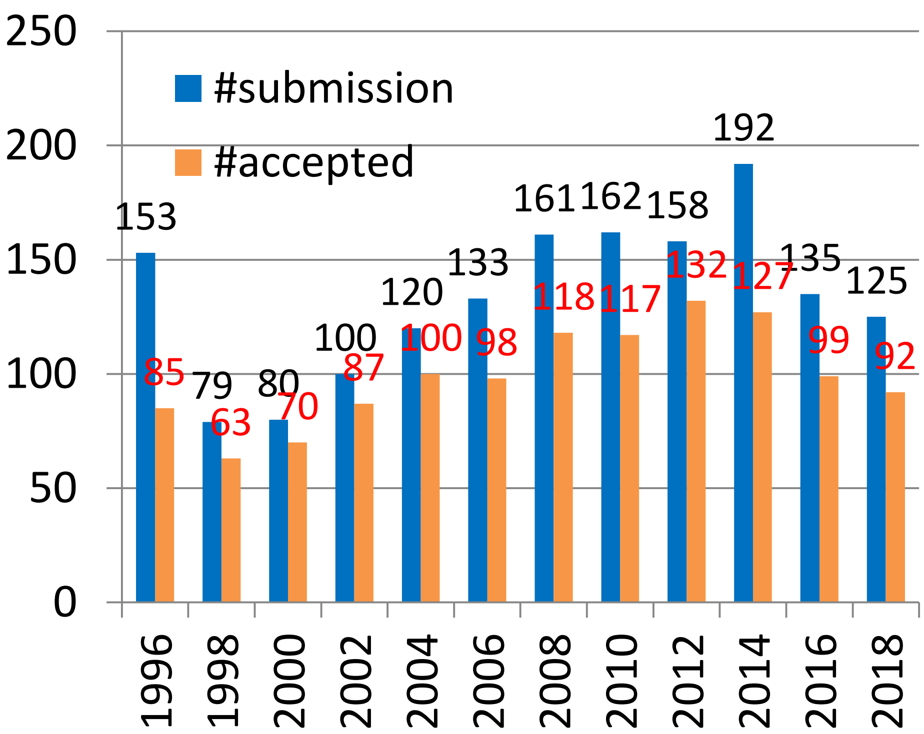
Conferences

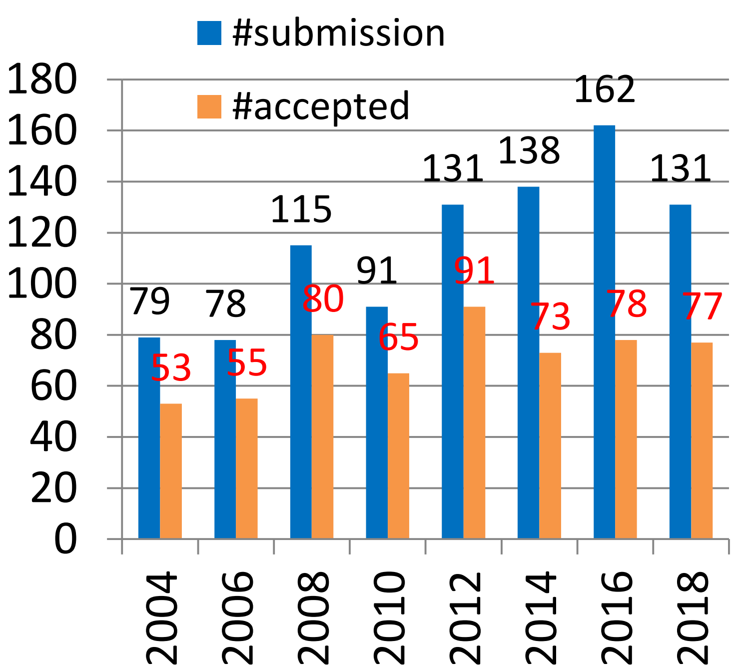
- ICDAR: Int'l Conf. on Document Analysis and Recognition (bi-ennial from 1991)



Conferences

- ICFHR: Int'l Conf. on Frontiers of Handwriting Recognition (from 1990, formerly IWFHR)
- DAS: IAPR Int'l Workshop on Document Analysis Systems (bi-ennial from 1994)





ICDAR2019 Competitions

We are pleased to announce that the ICDAR2019 will organize a set of competitions dedicated to a large set of document analysis problems. You are cordially invited to participate to this scientific event that will be a very good opportunity to objectively compare the quality of algorithms on different categories of challenges. You will find below the different categories of competitions, and the URL of their respective website, that will allow you to get all the required information for participating:

Category: Handwritten Historical Document Layout Recognition

- ICDAR 2019 Competition on Historical Book Analysis
- ICDAR 2019 Competition on Digitised Magazine Article Segmentation (historical documents)
- ICDAR 2019 Competition on German-Brazilian Newspaper Layout Analysis
- ICDAR 2019 Competition on Baseline Detection and Page Segmentation

Category: Historical Handwritten Script Analysis

- ICDAR 2019 Competition on Recognition of Historical Arabic Scientific Manuscripts
- ICDAR 2019 Historical Document Reading Challenge on Large Structured Family Records
- ICDAR 2019 Competition on Image Retrieval for Historical Handwritten Documents

Category: Document Recognition (Layout analysis & Text Recognition)

- ICDAR 2019 Competition on Table Detection and Recognition in Archival Documents
- ICDAR 2019 Competition on Table Recognition
- ICDAR 2019 Scanned Receipts OCR and Information Extraction
- ICDAR 2019 Competition on Form Understanding in Noisy Scanned Documents
- ICDAR 2019 Competition on Recognition of Documents with Complex Layouts
- ICDAR 2019 Competition on Recognition of Early Indian Printed Documents

Category: Handwriting recognition

- ICDAR 2019 Competition on Recognition of Handwritten Mathematical Expressions and Typeset Formula Detection

Category: Document Image Binarization

- ICDAR 2019 Competition on Binarization of Handwritten, printed, or mobile captured Documents
- ICDAR 2019 Competition on Document Image Binarization

Category: Robust Reading

Got popular from 2011

- ICDAR 2019 Competition on Robust Text Reading from Large-scale Street View Images with Partial Labels
- ICDAR 2019 RRC on Scene Text Visual Question Answering
- ICDAR 2019 RRC on Arbitrary-shaped scene text detection and recognition
- ICDAR 2019 RRC on Reading Chinese text on signboard
- ICDAR 2019 RRC on Multi-lingual scene text detection and recognition

Category: Post-OCR Correction

- ICDAR 2019 Competition on Post-OCR Text Correction

Category: Chart Parsing

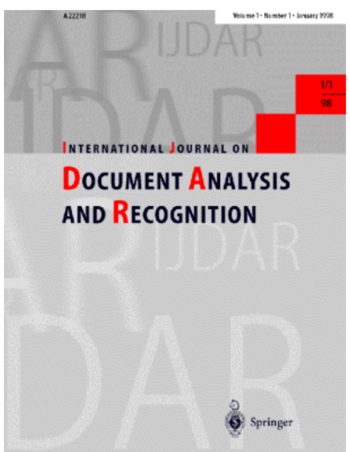
- ICDAR 2019 Competition on Chart Elements Parsing
- ICDAR 2019 Competition on Harvesting Raw Tables from Infographics

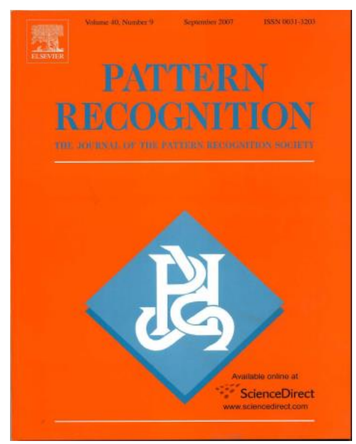
Category: Miscellaneous Competitions

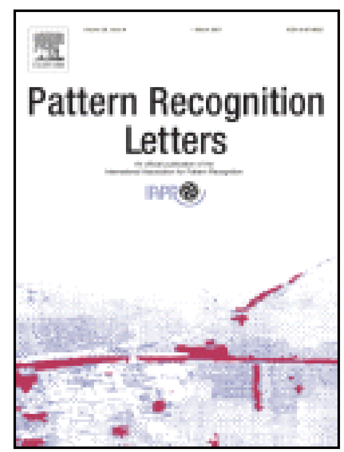
- ICDAR 2019 Competition on Fine-Grained Classification of comic characters
- ICDAR 2019 Competition on Object Detection and Recognition in Floorplan images
- ICDAR 2019 Competition on Signature Verification based on an On-line and Off-line Signature Dataset

Journals

- Int. J. Document Analysis and Recognition (IJDAR), Springer, 1998-
- IEEE Trans. Pattern Analysis & Machine Intelligence (PAMI), 1979-
- Pattern Recognition (PR), Elsevier, 1968-
- Pattern Recognition Letters (PRL), Elsevier, 1980-
- Int. J. Pattern Recognition & Artificial Intelligence (IJPRAI), World Scientific, 1987-

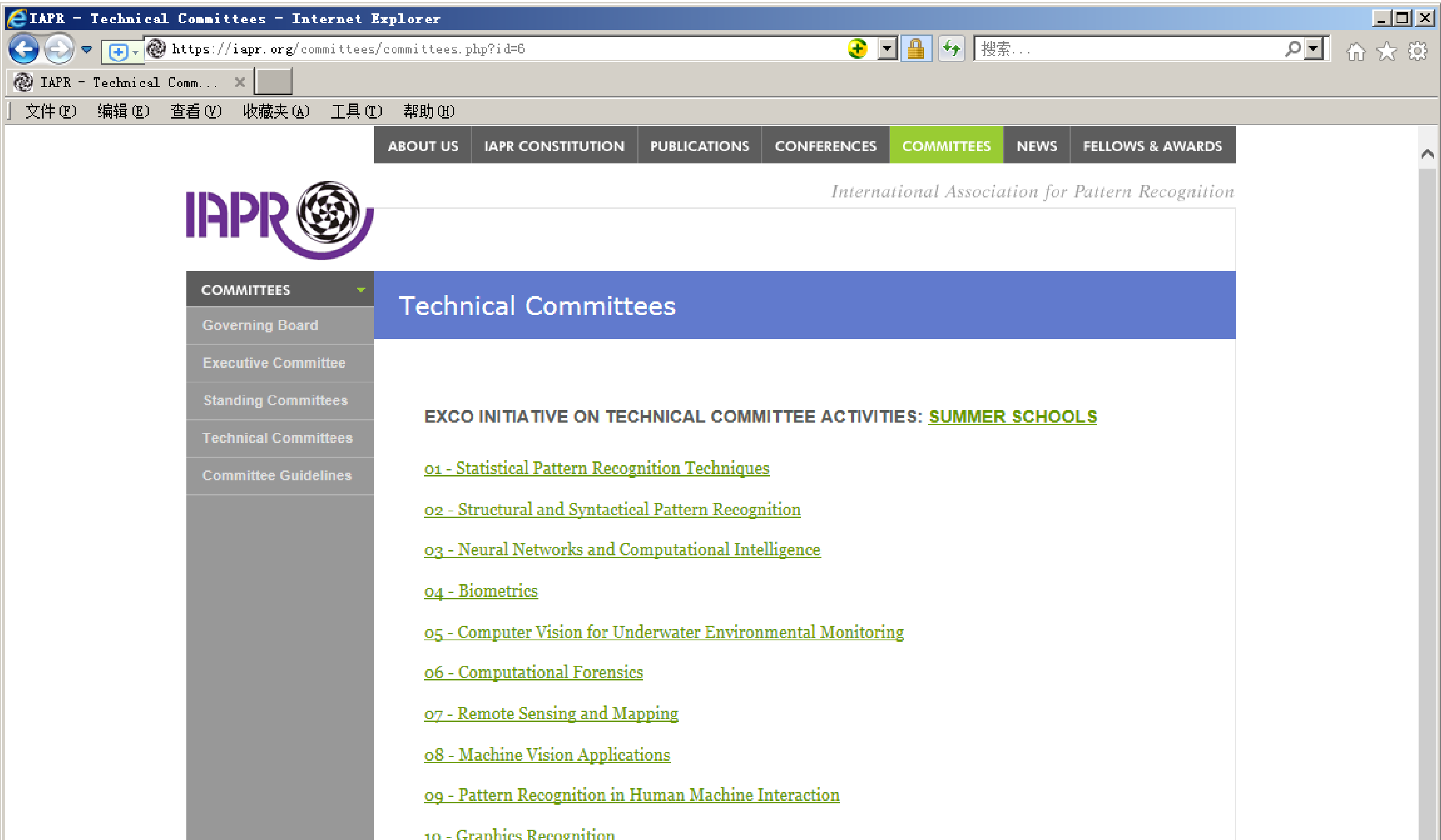




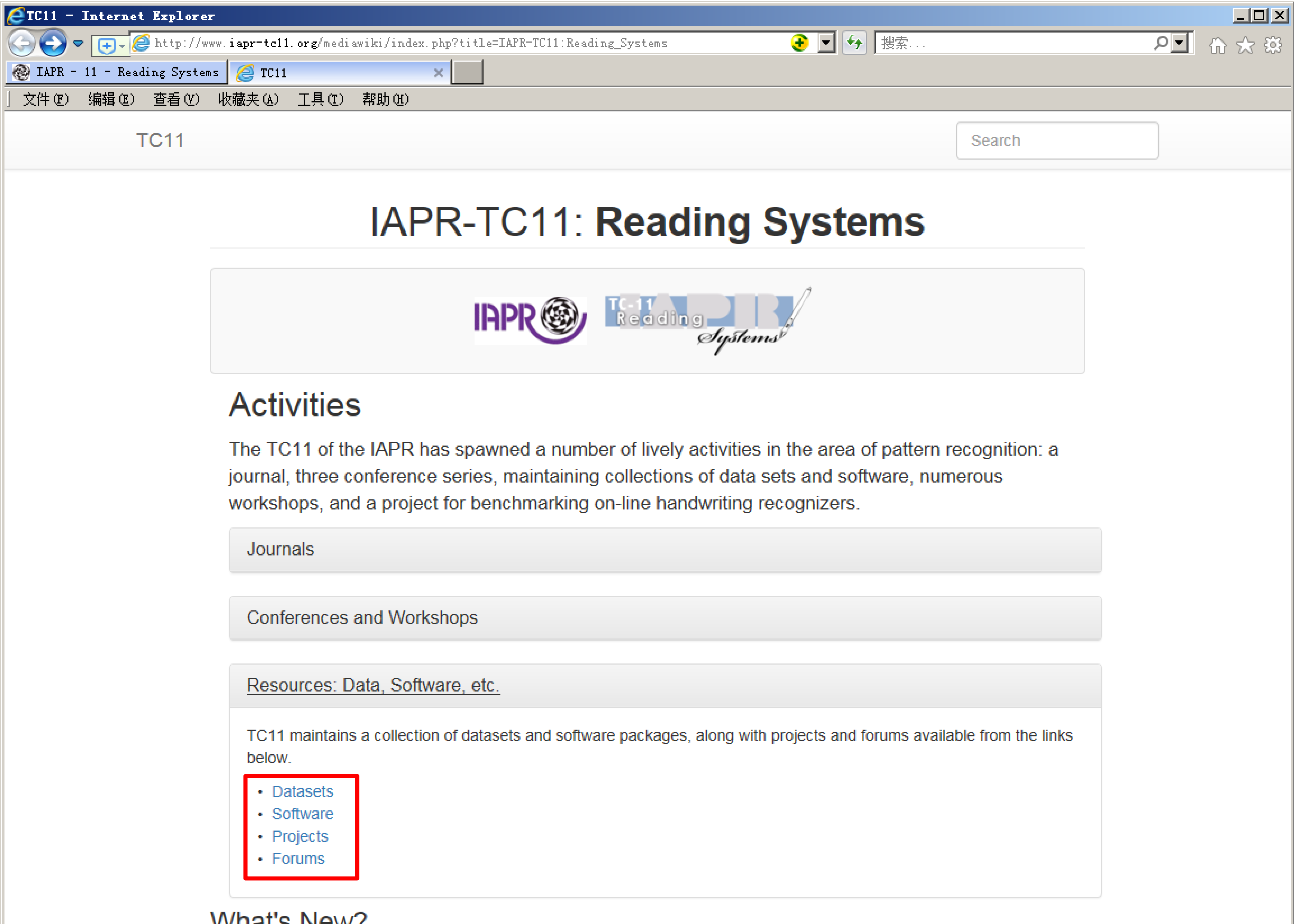


Societies

- International Association for Pattern Recognition (IAPR), 1978-
  – www.iapr.org
  – IAPR TC-10 (Graphics Recognition)
  – IAPR TC-11 (Reading Systems)



Resources: Datasets, Software



Datasets per Topic

Find [ ] Find

TradeMarks Image Database
24-06-2014 (v. 1) by Conny Gu

A Dataset for Arabic Text Detection, Tracking and Recognition in News Videos - AcTiV
16-03-2016 (v. 1) by Oussama Zayene

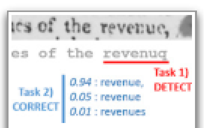
Ground Truth: Global xml file
16-03-2016 (v. 1) by Oussama Zayene

Ground Truth: Detection Ground-truth files
16-03-2016 (v. 1) by Oussama Zayene

Task: Text Detection in Arabic NewsVideo Frames
16-03-2016 (v. 1) by Oussama Zayene

Task: Text Tracking in Arabic NewsVideo
16-03-2016 (v. 1) by Oussama Zayene

Task: Text Recognition in Arabic NewsVideo Frames
16-03-2016 (v. 1) by Oussama Zayene



Dataset for the competition on Post-OCR Text Correction 2017
28-05-2019 (v. 1) by Guillaume Chiron

Ground Truth: Transcription for the competition on Post-OCR Text Correction 2017
28-05-2019 (v. 1) by Guillaume Chiron

D I B

Document Image Binarization Platform
27-11-2017 (v. 1) by Rafael Dueire Lins

Ground Truth: Ground Truth for DIB Platform
05-12-2017 (v. 1) by Rafael Dueire Lins

ICDAR 2009 Signature Verification Competition (SigComp2009)
23-02-2015 (v. 1) by Muhammad Imran Malik

Ground Truth: Ground Truth Information for the ICDAR 2009 Signature Verification competition (SigComp2009)

News

- 01/13/2013
  The new TC11 Online Resources site is open for testing

Datasets List - TC11 - Internet Explorer

http://www.iapr-tc11.org/mediawiki/index.php/Datasets_List

IAPR - 11 - Reading Systems | Datasets List - TC11 | Softwares - TC11

文件(F) 编辑(E) 查看(V) 收藏夹(A) 工具(T) 帮助(H)

查找: Chinese 上一个 下一个 选项

Old Dataset Repository

Search

Datasets List

NOTICE: TC11 datasets will be soon moved to the new Web portal at http://tc11.cvc.uab.es This page will remain available but will not be updated from January 2015 onwards.

Datasets -> **Datasets List**

Last updated: 2015-001-23

See the datasets sorted according to the Journal / Conference they first appeared in.

Complex Text Containers

Scene Text

- MSRA Text Detection 500 Database (MSRA-TD500)
- The Street View Text Dataset
- The Street View House Numbers (SVHN) Dataset
- NEOCR: Natural Environment OCR Dataset
- KAIST Scene Text Database
- ICDAR 2003 Robust Reading Competitions
- ICDAR 2005 Robust Reading Competitions

Machine-printed Documents

- Table Ground Truth for the UW3 and UNLV datasets
- The DocLab Dataset for Evaluating Table Interpretation Methods
- [the IMPACT data base] The dataset contains more than half a million representative text-based images compiled by a number of major European libraries. Covering texts from as early as 1500, and containing material from newspapers, books, pamphlets and typewritten notes, the dataset is an invaluable resource for future research into imaging technology, OCR and language enrichment.
- PRImA Layout Analysis Dataset
- DFKI Dewarping Contest Dataset (CBDAR 2007) The dataset, that was used in the CBDAR 2007 Dewarping Contest, contains 102 camera captured documents with their corresponding ASCII text ground-truth. Additionally, text-line level ground-truth was also prepared to benchmark curled text-line segmentation algorithms. Part of the dataset (76 out of 102

Softwares - TC11 - Internet Explorer

http://www.iapr-tc11.org/mediawiki/index.php?title=Softwares

IAPR - 11 - Reading Systems | Datasets - TC11 | Softwares - TC11

文件(F) 编辑(E) 查看(V) 收藏夹(A) 工具(T) 帮助(H)

查找: Chinese 上一个 下一个 选项

TC11

Search

Softwares

(Redirected from Software)

Last updated: 2018-12-21

On-line handwriting

Better HMM tool: KALDI

- HTK - Hidden Markov Model Toolkit
- Implementation of Bidirectional Long-Short Term Memory Networks (BLSTM) combined with Connectionist Temporal Classification (CTC) - including examples for Arabic recognition
- SRILM - A Toolkit for generating language modeles
- Torch5 - A Toolkit for HMM and GMM and many other machine learning algorithms
- **uptools:** Tools for reading and processing files in the UNIPEN file format.
- Comparison Tools for Handwriting Recognizers using the UNIPEN format (Gene Ratzlaff, IBM)

Off-line handwriting

- **HUE:** a software toolkit which supports the rapid development and re-use of handwriting and document analysis systems (Univ. of Essex, UK).

OCR

- OCRopus - The OCRopus(tm) open source document analysis and OCR system
- NHocr - OCR engine for Japanese language
- Public domain OCR software (Univ. of Maryland, USA)
- Source code at the DIMUND server (Univ. of Maryland, USA)
- Optical Character Recognition sources
- RWTH OCR - The RWTH Aachen University Optical Character Recognition System

Many good HR/OCR methods are not in open source

Pixels vs Vectors

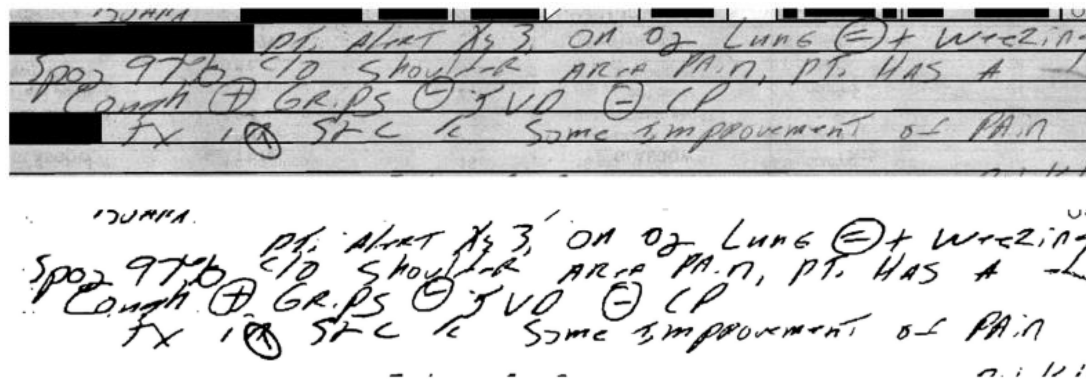
- AutoTrace bitmap to vector conversion

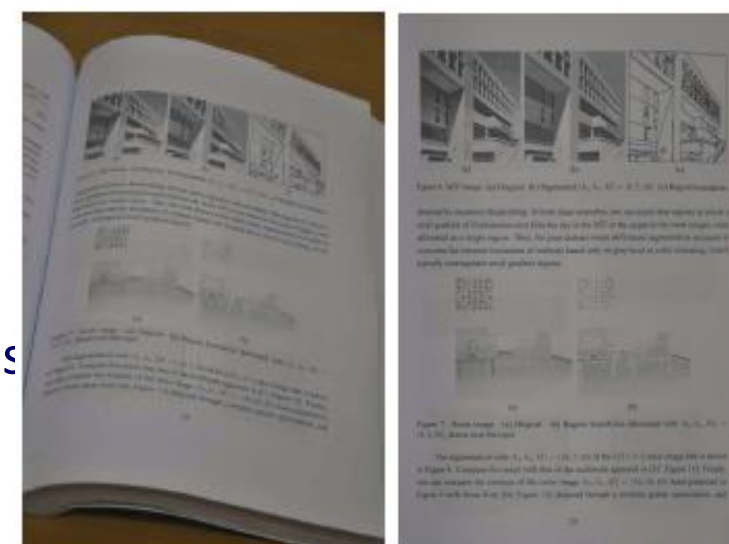
Main Approaches

- Image Pre-processing
- Layout Analysis
- Scene Text Detection
- Text Line Recognition
- Graphics Recognition

Image Pre-Processing

- Enhancement/denoising
  - MRF (Markov random field)
  - Morphology
  - Deblurring
- Binarization
  - Local/adaptive
  - Stroke edges
  - Classification-based, MRF, CRF (Conditional random field)
  - Full convolutional network (FCN)
- Rectification
  - 3D shape modeling
  - Cylindrical surface reconstruction
  - Polynomial curve fitting with text lines





Layout Analysis

- Tasks
  - Page segmentation (geometric layout analysis)
    - Segmenting page image into regions (zones) of homogeneous class (text, handwritten/print, graphics, header, signature, logo, table, mathematics)
      - Text region segmented into text lines/words
  - Layout Understanding (logical layout analysis)
    - Labeling the semantic class, logical order and relationship of regions
    - Layout reconstruction
    - Possibly incorporating cues from text recognition

Page segmentation: intensively studied, not solved

Layout Analysis Problem

- Task scope

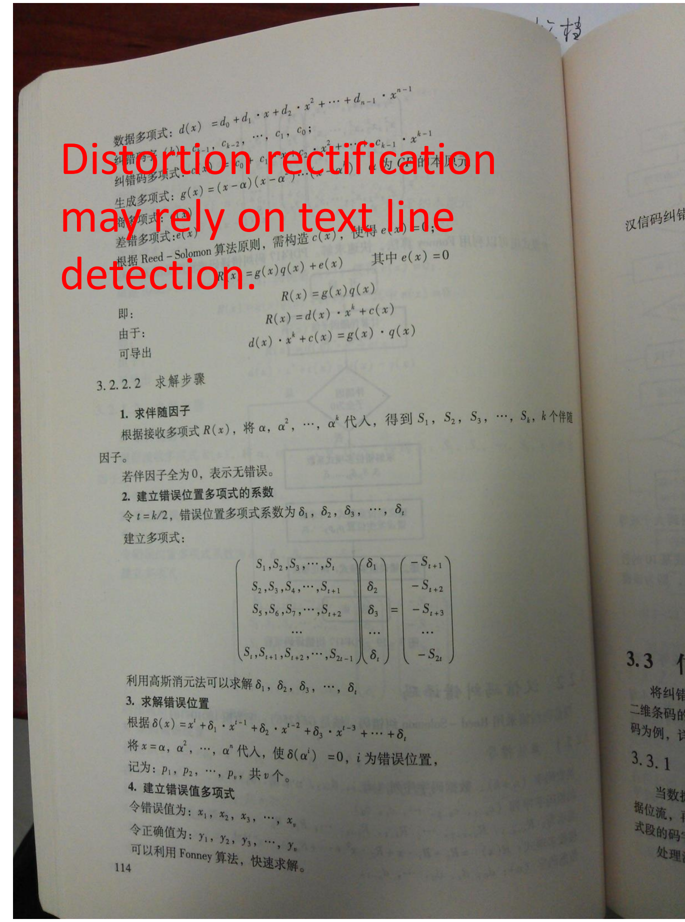
- Layout Analysis
  - Content classification
    - Text
      - Semantic classes (header, section title, caption, paragraph, etc
      - Script identification
        - Chinese
        - English
        - Arabic
      - Writing type separation
        - Handwritten
        - Printed
    - Non-text
      - Picture
      - Graphics
      - Decoration
  - Page segmentation
    - Normal Text
      - Word
      - Text line
      - Paragraph
      - Article
    - Structured Text
      - Formula
      - Table
      - Flow chart
    - Non-text
      - Picture
      - Graphics
      - Decoration

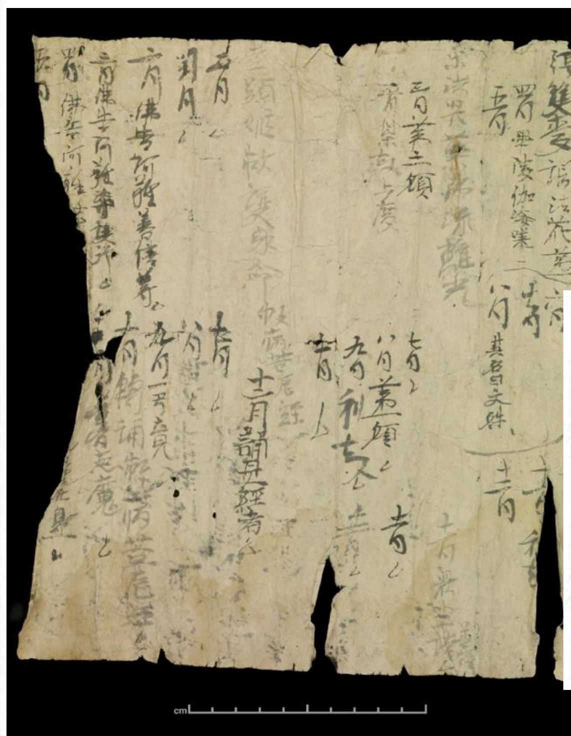
Page Segmentation Difficulties

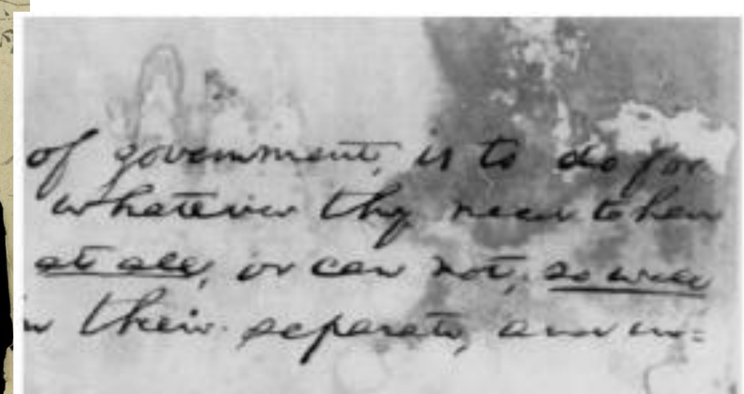
- Layout Complexity
  - Non-rectangular (non-Manhatton) structure
  - Multiple region types
  - Separation between close regions
  - Irregular shape (e.g., handwritten)
  - Complex background

Images from: M. Benjelil, S. Lanoun, R. Mullot, A.M. Alimi, Complex documents images segmentation based on steerable pyramid features, *IJDAR* 2010.

- Imaging Quality
  - Noise, low contrast
  - Paper contamination
  - Camera-based distortion
    - Rotation, perspective
    - Curvilinear surface
    - Background of paper

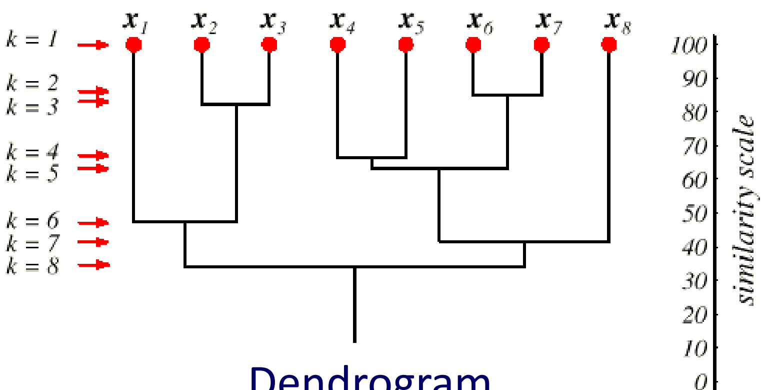




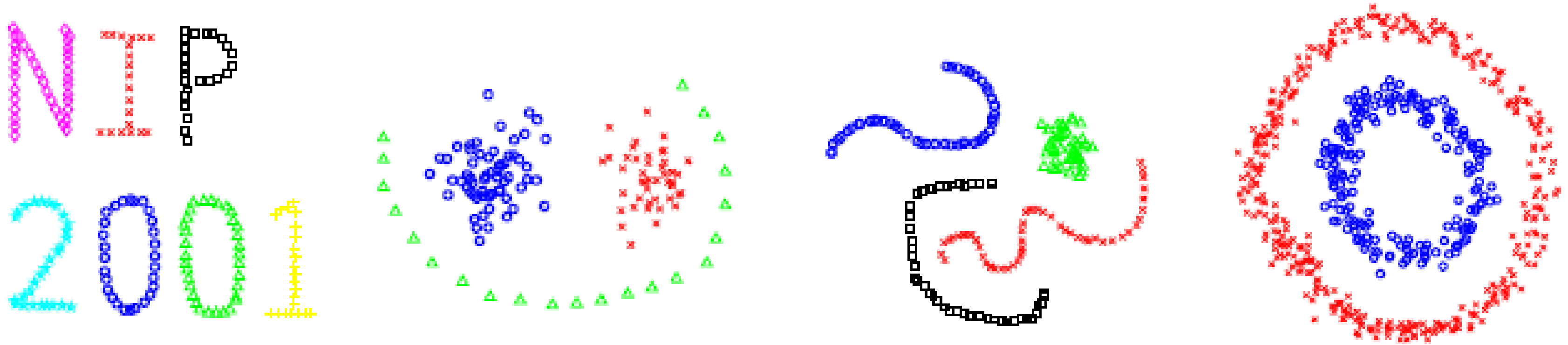


Page Segmentation as Clustering

- Clustering Methods
  - Partitional
    - K-means, Gaussian mixture density (EM)
  - Hierarchical
    - Divisive (top-down)
    - Agglomerative (bottom-up)
      - Single-link, complete-link
  - Graph theoretic
    - Spectral clustering



Dendrogram



Categorization of Methods

- Typology of Methods
  - Processing direction
    - ✓ Top-down methods
    - ✓ Bottom-up methods
    - ✓ Hybrid methods
  - Learning based or not
    - Heuristic rule based methods
    - Machine learning based methods
    - Hybrid methods
  - Layout segmentation limitation
    - Algorithm constrained methods
    - Parameter constrained methods
    - Potentially unconstrained methods

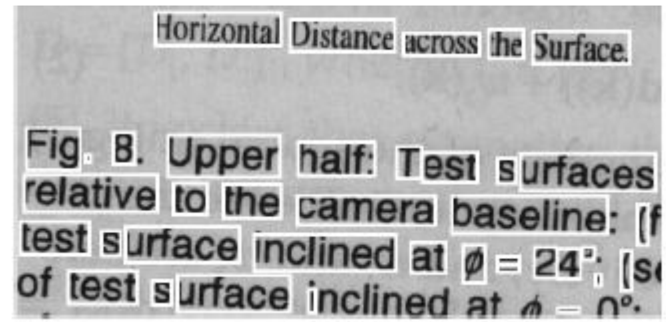
Bottom-up is more effective, but computationally expensive

Traditional Methods

- Pre-Processing
  - Binarization, noise removal
  - Connected component (CC) analysis
  - Rotation correction
- Classic Layout Analysis Methods
  - Projection profile analysis
  - Recursive x-y cuts
  - Whitespace analysis

    Top-down
  - Run-length smearing algorithm (RLSA)
  - Document spectrum
  - Voronoi diagram-based algorithm

    Bottom-up
  - Texture-based (feature-based) methods
  - Line adjacency graph (LGA) based methods



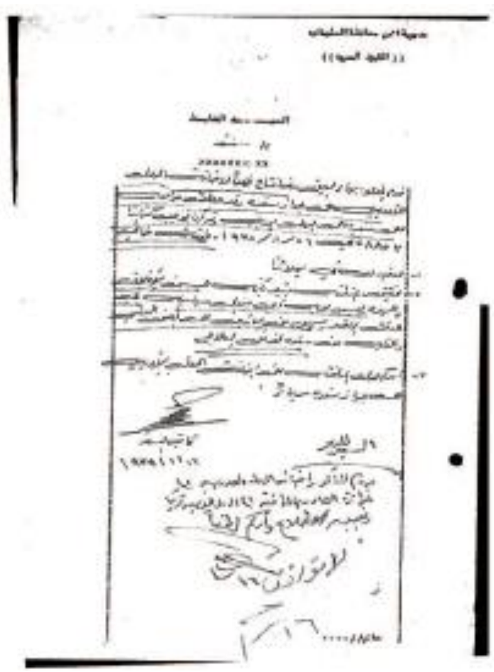
(Jain & Yu, 1998)

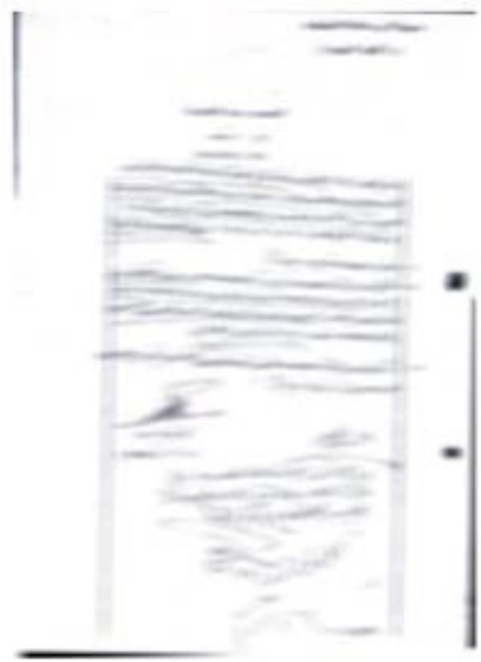
Latest Methods

- Objective: to handle variable complex documents
- Methods Based on Deformable Models
  - Level set, active contour, seam carving
- Methods by Graph-Based Clustering
  - Minimum spanning tree (MST) clustering
- Methods Based on Structured Prediction
  - Page segmentation using conditional random field (CRF)
- Methods Based on FCN
  - FCN for text line segmentation and detection
  - Multi-task layout analysis using FCN
  - Learning to extract semantic structure using multimodal FCN

Text Line Segmentation Using Level Set Method

- Script-independent, segmenting curvilinear, close and touching text lines
- The level set method is exploited to determine the boundary of neighboring text lines by evolving an initial estimate



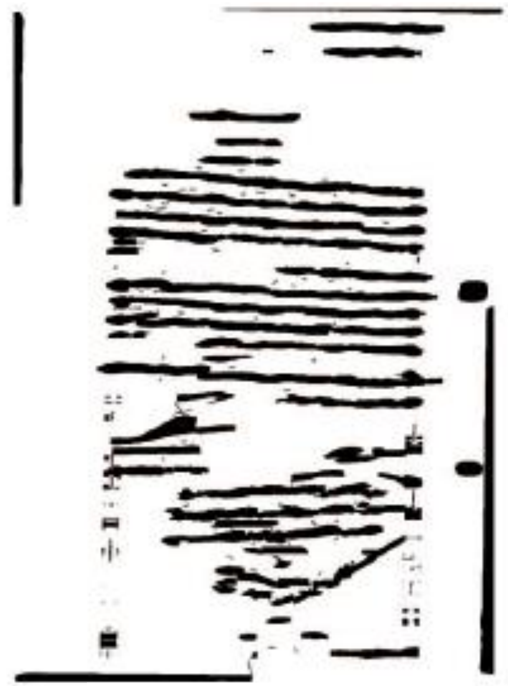


Density estimation using anisotropic kernel

Initial estimate of text lines (pixels of high density)

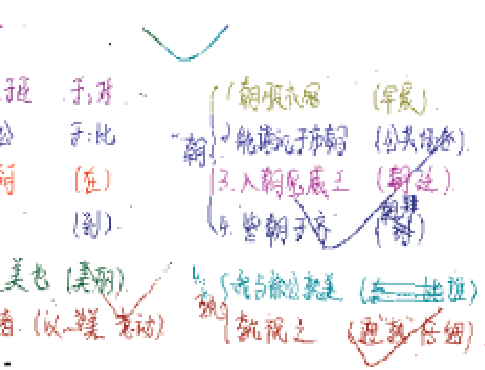
Y. Li, Y. Zheng, D. Doermann, S. Jaeger, <u>Script-independent text line segmentation in freestyle handwritten documents</u>, *IEEE Trans. PAMI*, 2008.

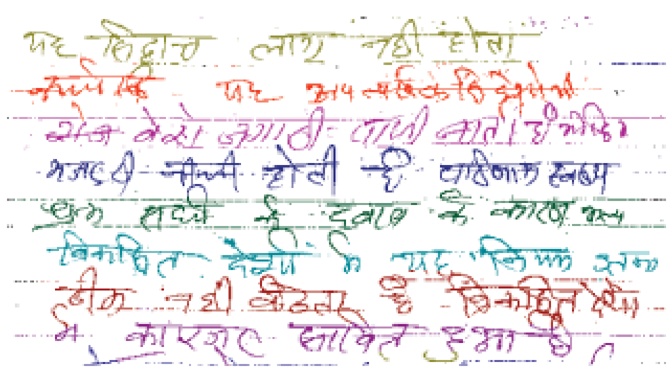
After 10 iterations of evolution using level set

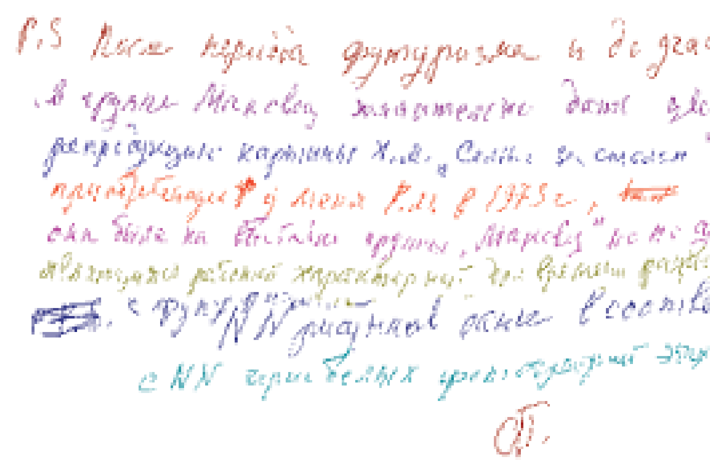


After connecting broken text lines

Final text line segmentation



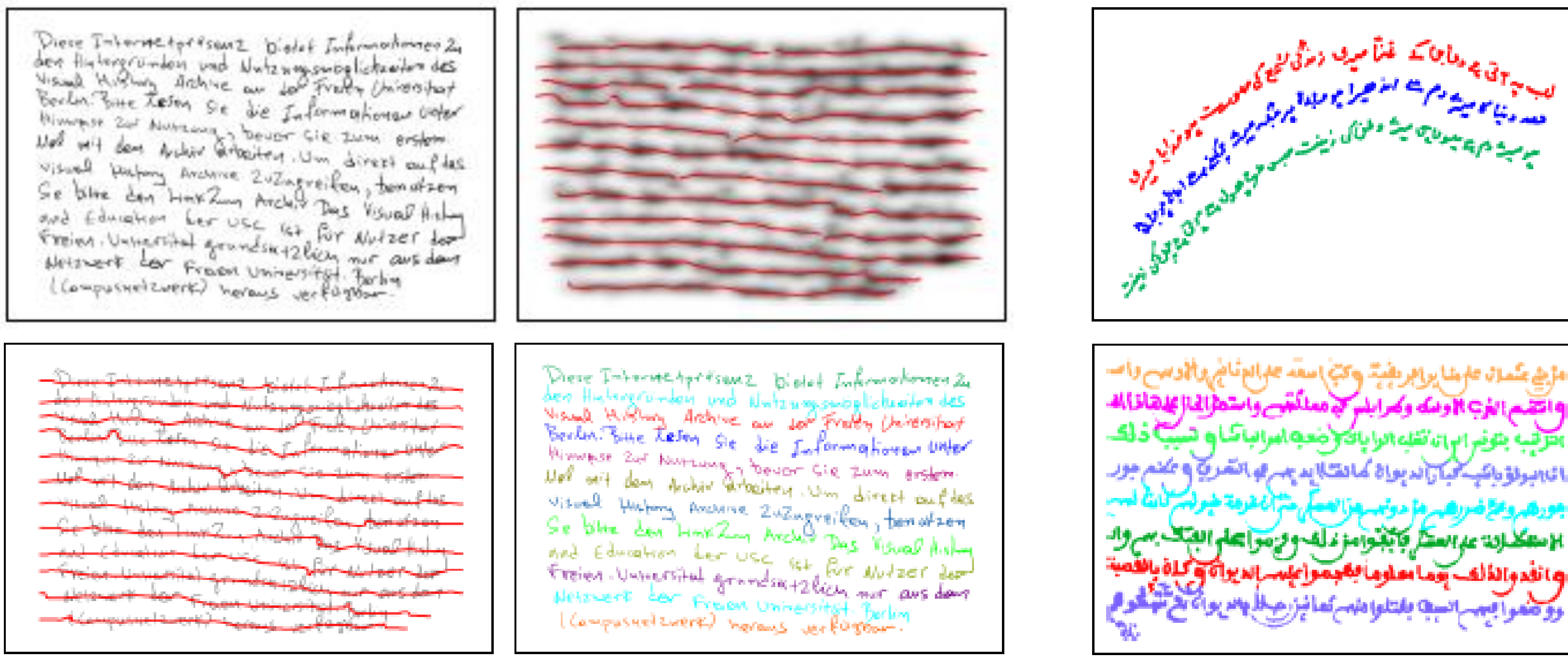




Textline Segmentation Using Active Contour

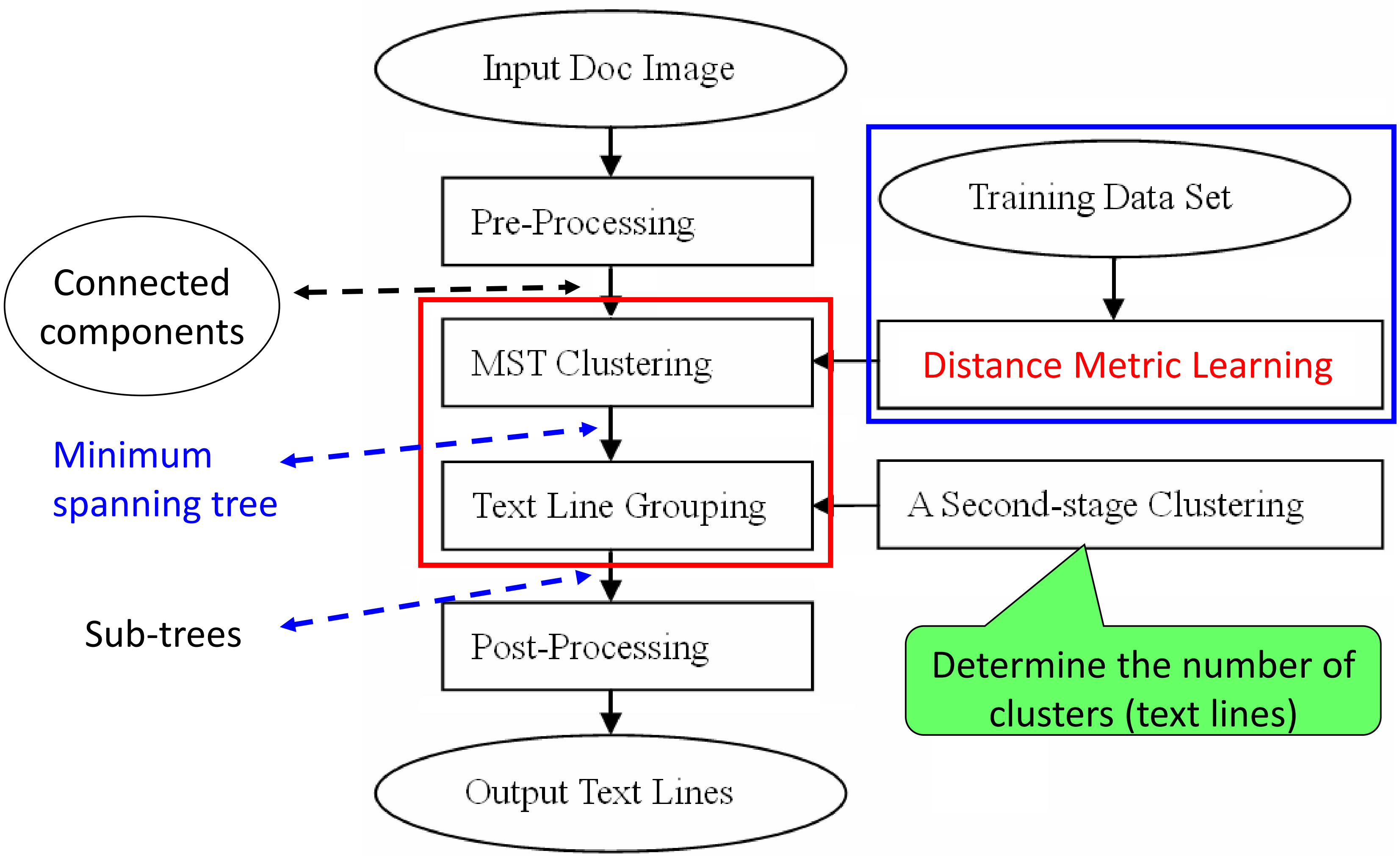
- Multi-oriented textlines smoothing using matched filter bank
- Central line approximation using Horn-Riley based ridge detection method (find zero crossing of directional derivatives of smoothed image)
- Adaptation of active contours (snakes) over ridges

Active contour: minimize $E = \int_0^1 [E_{int}\{S(s)\} + E_{ext}\{(S(s)\}ds$



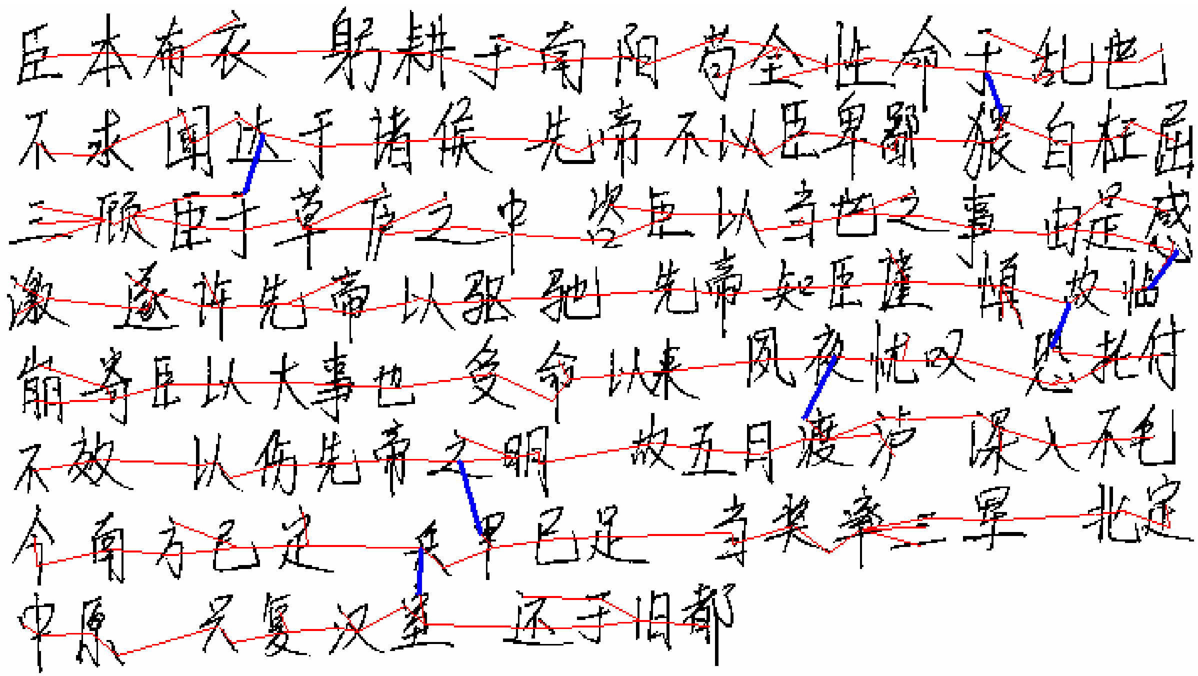
S.S. Bukhari, F. Shafait, T. Breuel, Script-Independent Handwritten Textlines Segmentation using Active Contours, ICDAR 2009.

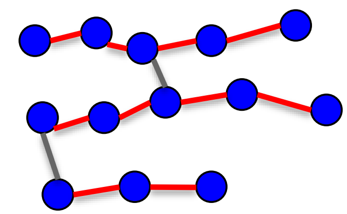
Handwritten Text Line Segmentation by MST Clustering



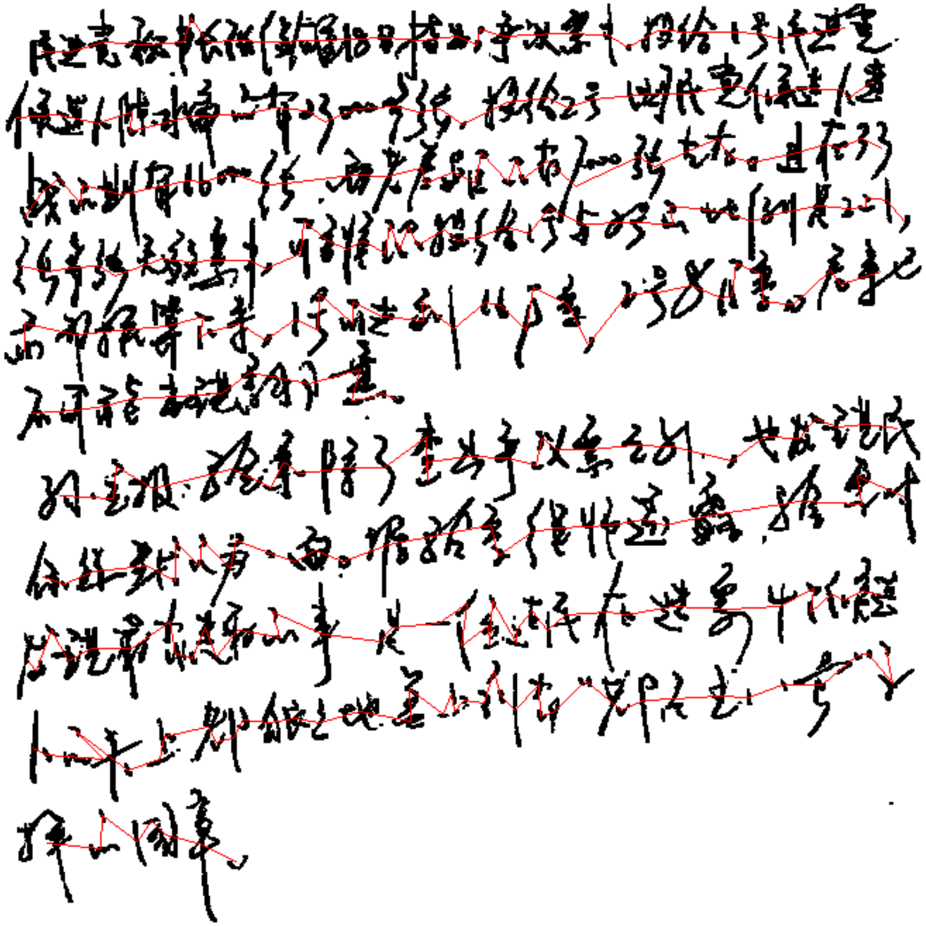
F. Yin, C.-L. Liu, Handwritten Chinese text line segmentation by clustering with distance metric learning, *Pattern Recognition*, 2009.

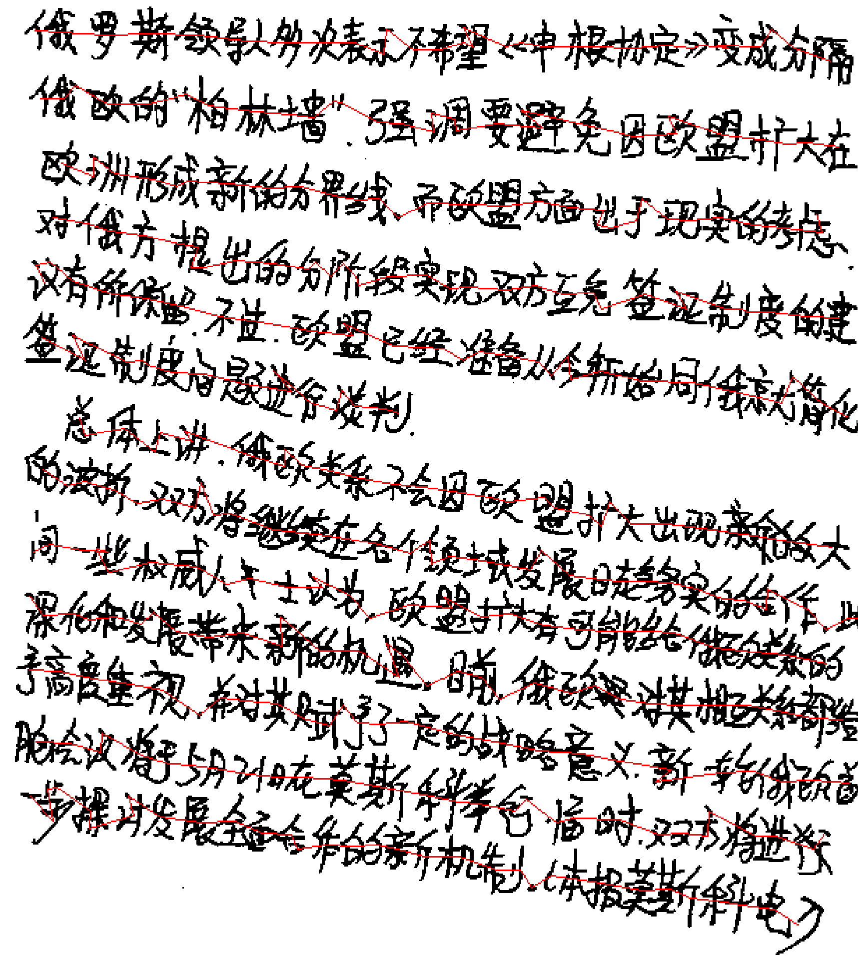
- Effect of Distance Metric Learning
  - Components in same line mostly connected
  - Less between-line edges





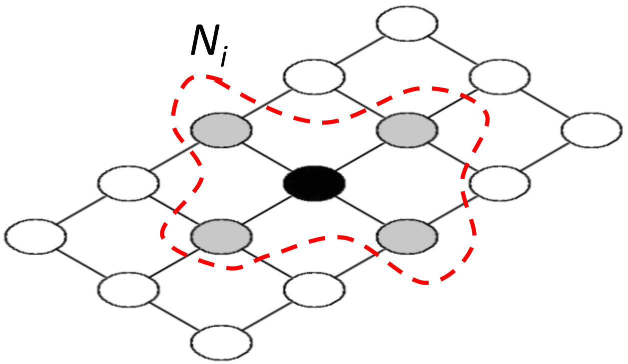
Results: Proposed Method





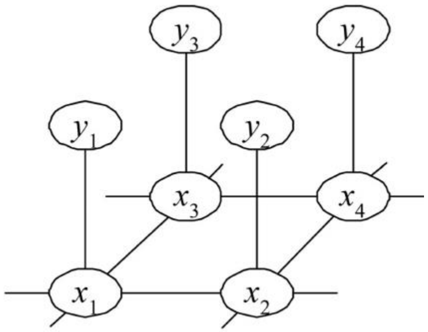
Layout Analysis as Structured Prediction

- Structured Prediction
  - Labeling multiple related objects/parts jointly
  - Markov random field (MRF), conditional random field (CRF), max-margin Markov network (M^3N)
- Deep Structured Model
  - Deep learning for potential functions or part models
  - Possibly trained end-to-end



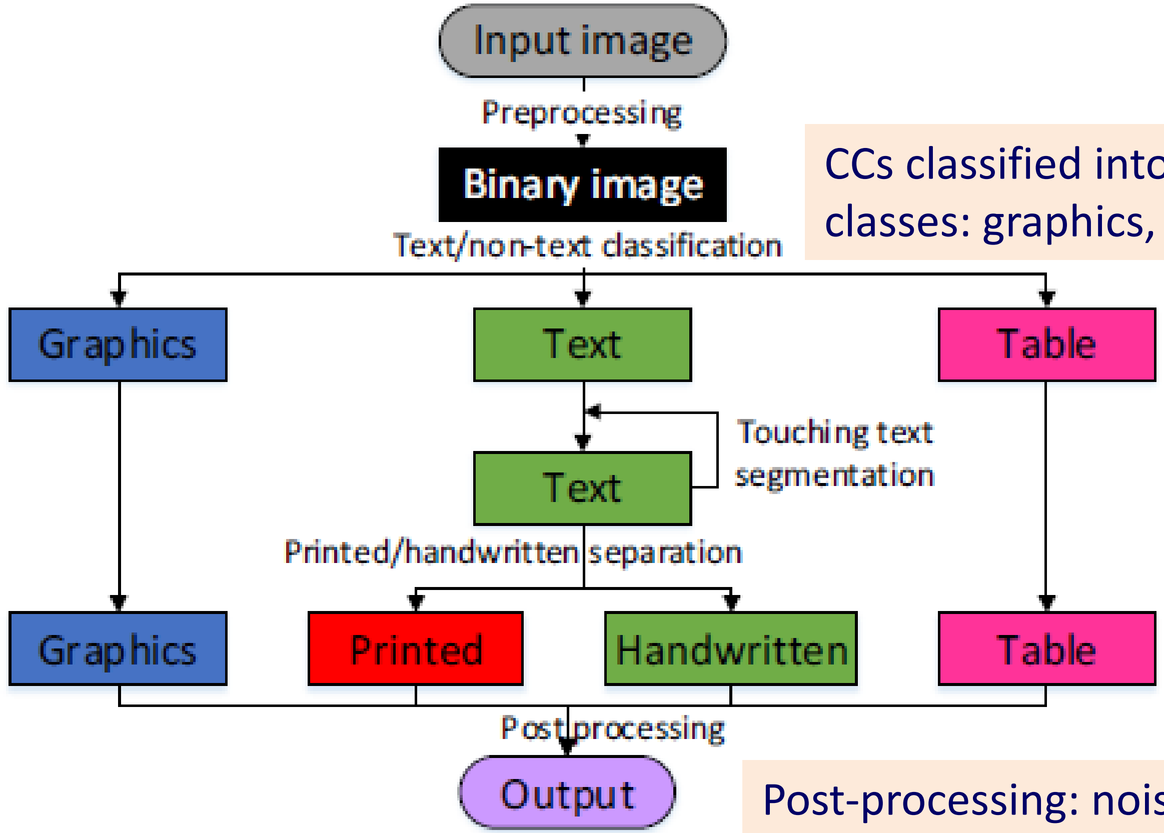
$$p(\mathbf{y}|\mathbf{x}) = \frac{1}{Z(\mathbf{x})} \prod_{C_p \in \mathcal{C}} \prod_{\Psi_c \in C_p} \Psi_c(\mathbf{x}_c, \mathbf{y}_c; \theta_p)$$

$$\Psi_c(\mathbf{x}_c, \mathbf{y}_c; \theta_p) = \exp\left\{ \sum_{k=1}^{K(p)} \lambda_{pk} f_{pk}(\mathbf{x}_c, \mathbf{y}_c) \right\}$$



Page Segmentation Using CRF

Connected components (CCs) classification with Conditional Random Field (CRF) for exploiting spatial context

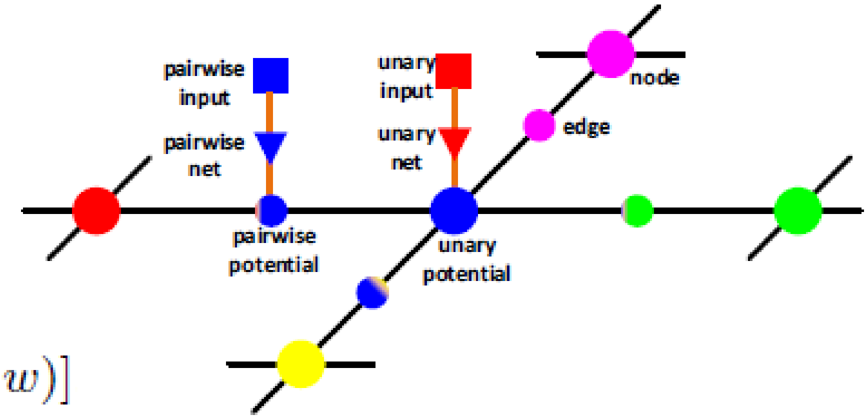


CCs classified into three classes: graphics, text, table

Post-processing: noise CCs assignment according to k-NN rule

X.-H. Li, F. Yin, C.-Lin Liu, Printed/Handwritten Texts and Graphics Separation in Complex Documents using Conditional Random Fields, DAS 2018.

CRF input: set of CCs x
Output: labels of CCs y* by MAP inference



$$P(y|x;w) = \frac{1}{Z(x;w)} exp[-E(y,x;w)]$$

$$Z(x;w) = \sum_{y} exp[-E(y,x;w)]$$

$$E(y,x;w) = \sum_{p \in N_U} U(y_p, x_p, ; w_U) + \sum_{(p,q) \in S_V} V(y_p, y_q, x_{pq}; w_V)$$

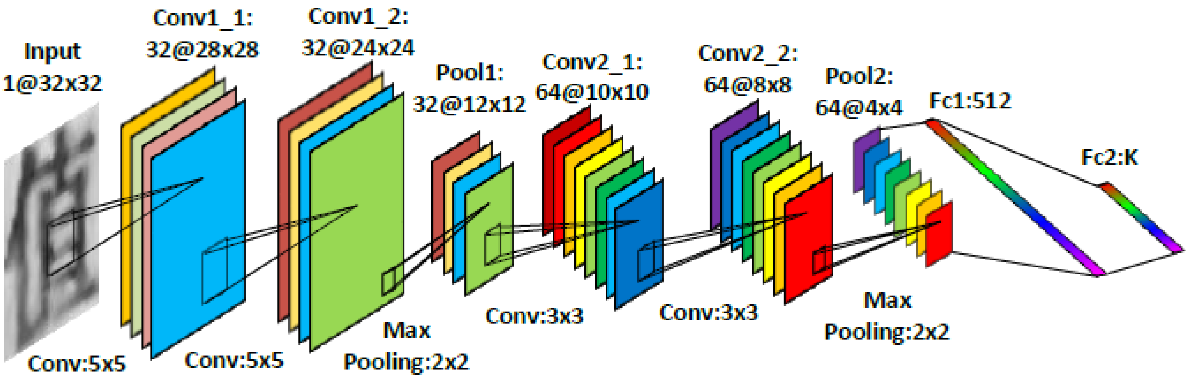
Unary potential $U(y_p, x_p; w_U) = \sum_{k=1}^{K} -\lambda_k \delta(k = y_p) z_{p,k}(x; w_U)$

Pairwise potential $V(y_p, y_q, x_{p,q}; w_V) = \sum_{k_p=1}^{K} \sum_{k_q=1}^{K} -\lambda_{k_p,k_q} \delta(k_p = y_p)$

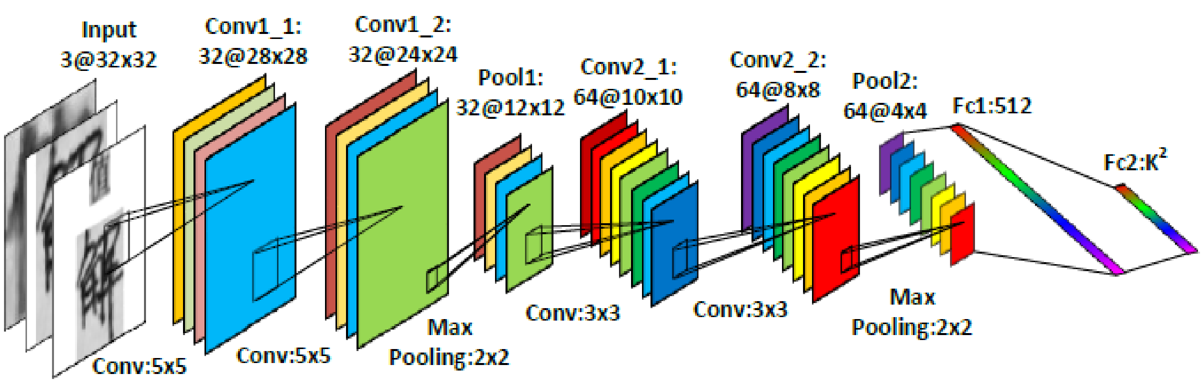
$$\delta(k_q = y_q) z_{p,k_p,q,k_q}(x; w_v),$$

Inference $y^* = arg \max_{y} P(y|x;w)$

$$= arg \max_{y} \frac{1}{Z(x;w)} exp[-E(y,x;w)]$$

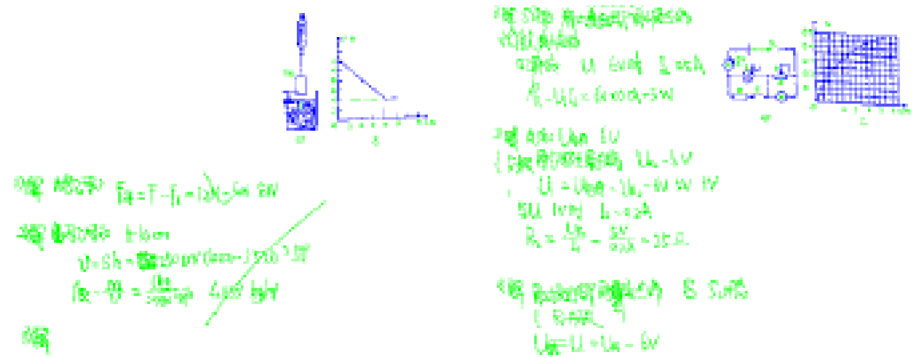


Unary potential CNN

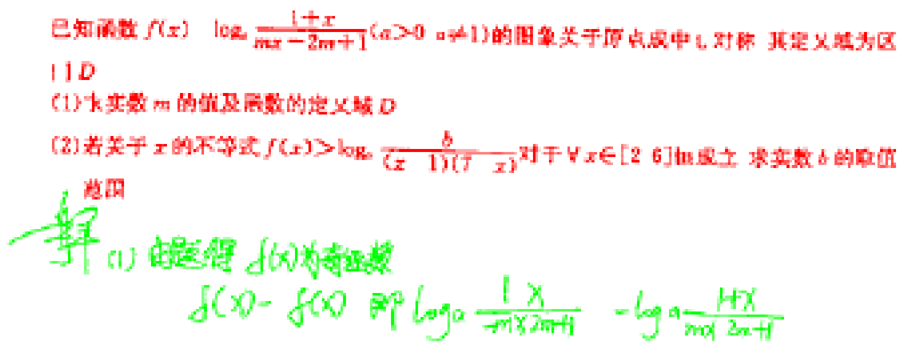


Pairwise-Net CNN

Text/non-text separation in test paper document



Printed/handwritten separation in test paper document



Printed/handwritten separation in Maurdor Dataset

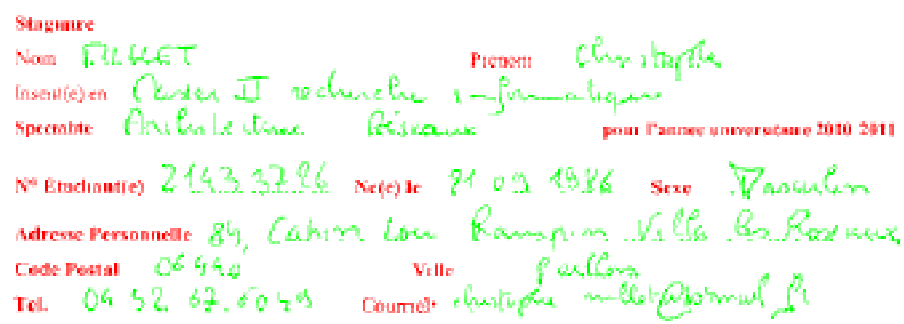
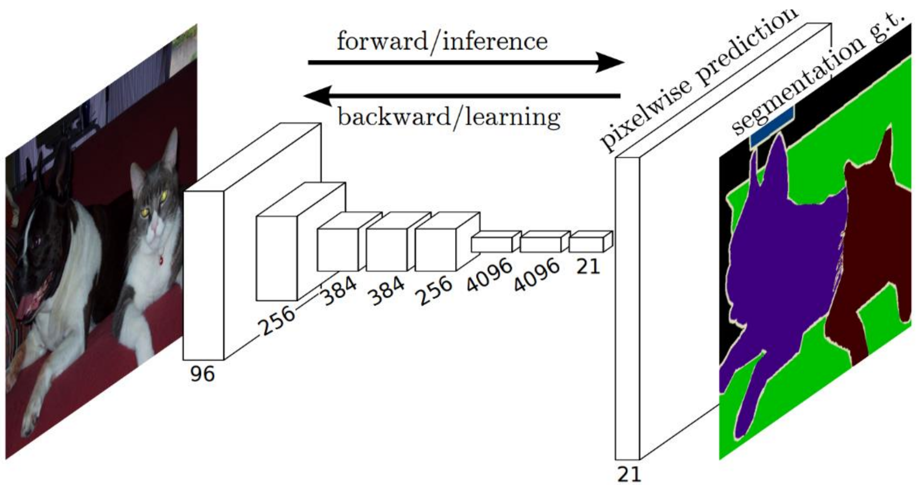
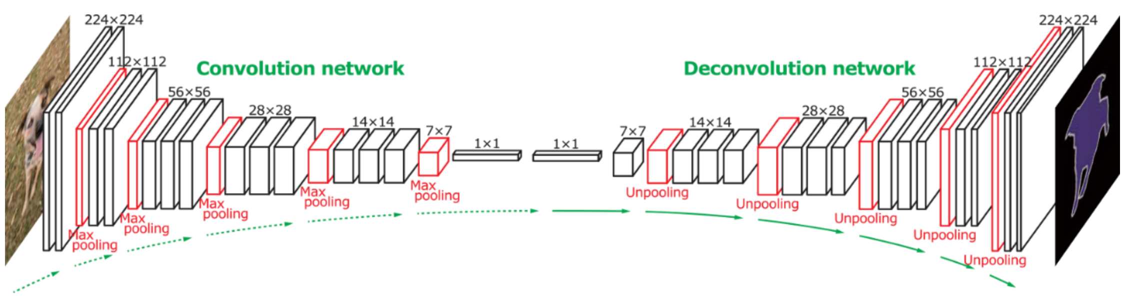
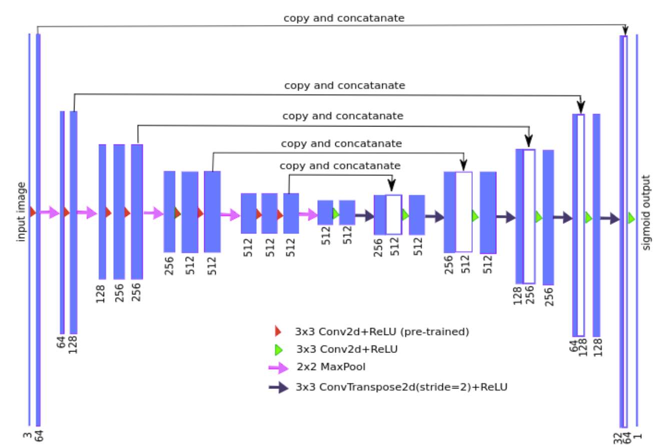


Image Segmentation Using FCN

- Fully Convolutional Network (FCN)
  - Pixel-wise prediction
  - Success in semantic segmentation

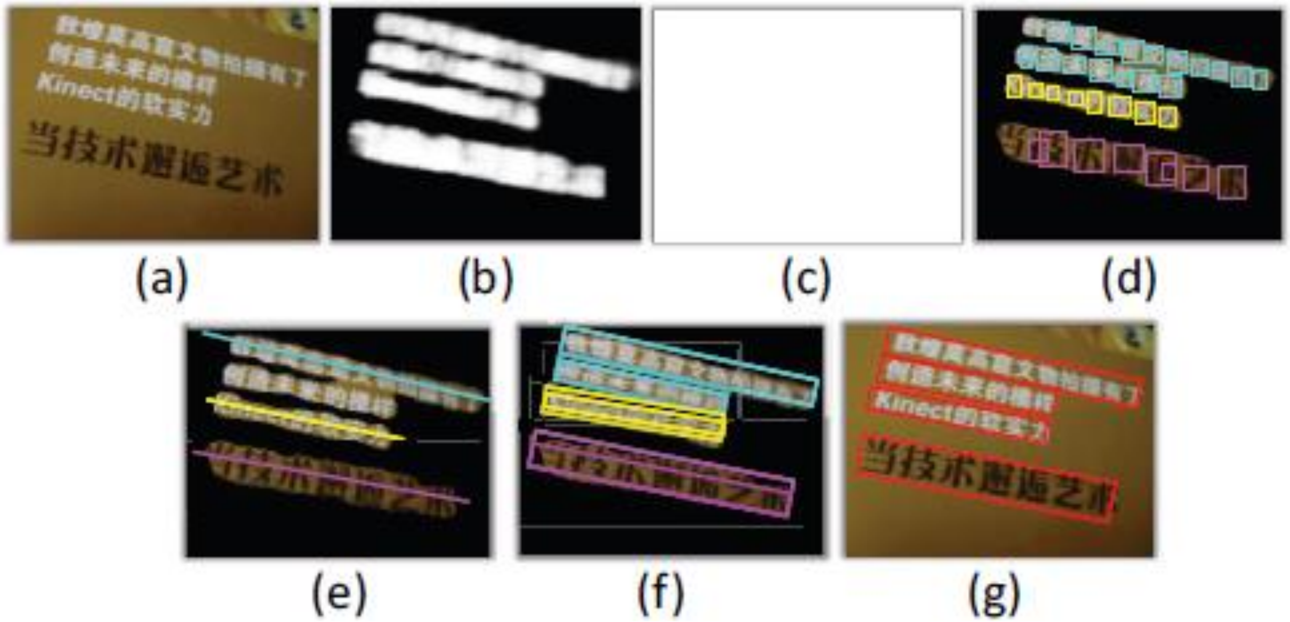






U-Net

FCN used in scene text detection



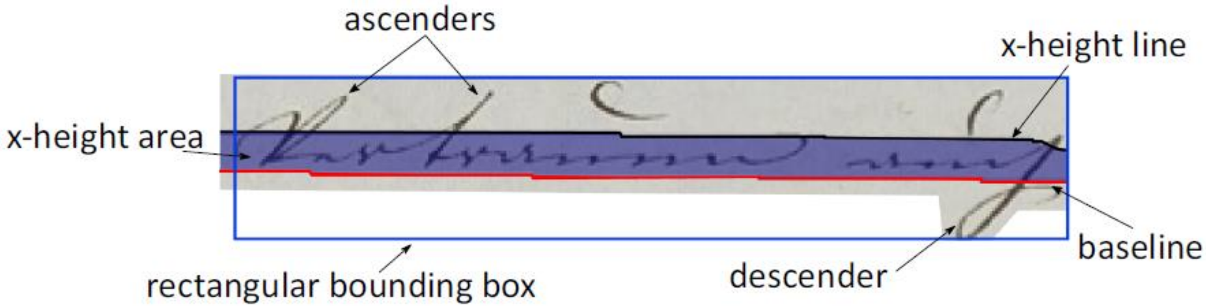
(b) Saliency map prediction by text-block FCN; (d) Candidate character component extraction; (e) Orientation estimation by projection; (f) Text line candidates generation; (g) Final result using character-centroid FCN for removing false hypotheses.

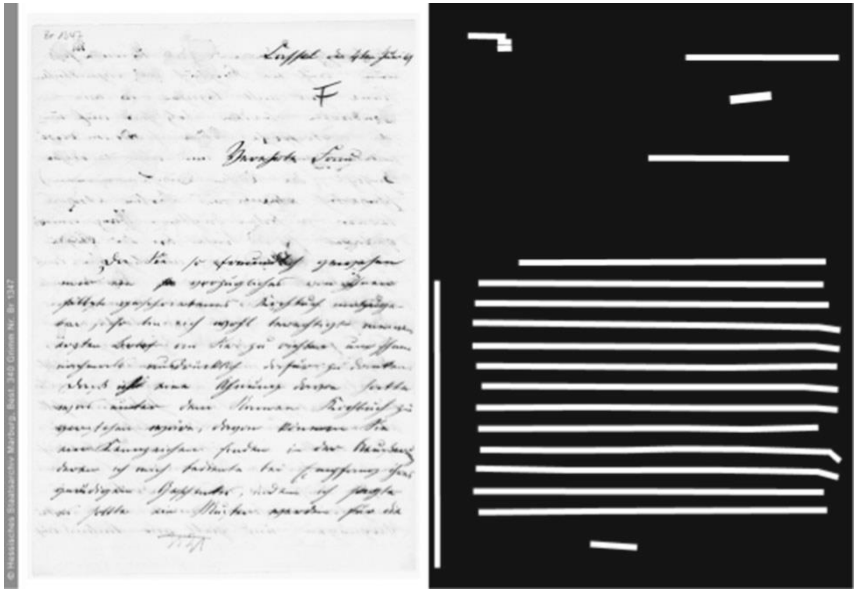
Ground truths for training text-block FCN: pixels within bounding boxes
Ground truths for training character centroid FCN (small version of text-block FCN): pixels within a distance from character centroid.

Z. Zhang, C. Zhang, W. Shen, C. Yao, W. Liu, X. Bai, Multi-oriented text detection with fully convolutional networks, CVPR 2016.

Handwritten Text Line Segmentation Using FCN with Dilated Convolutions

- Text line core pixels prediction

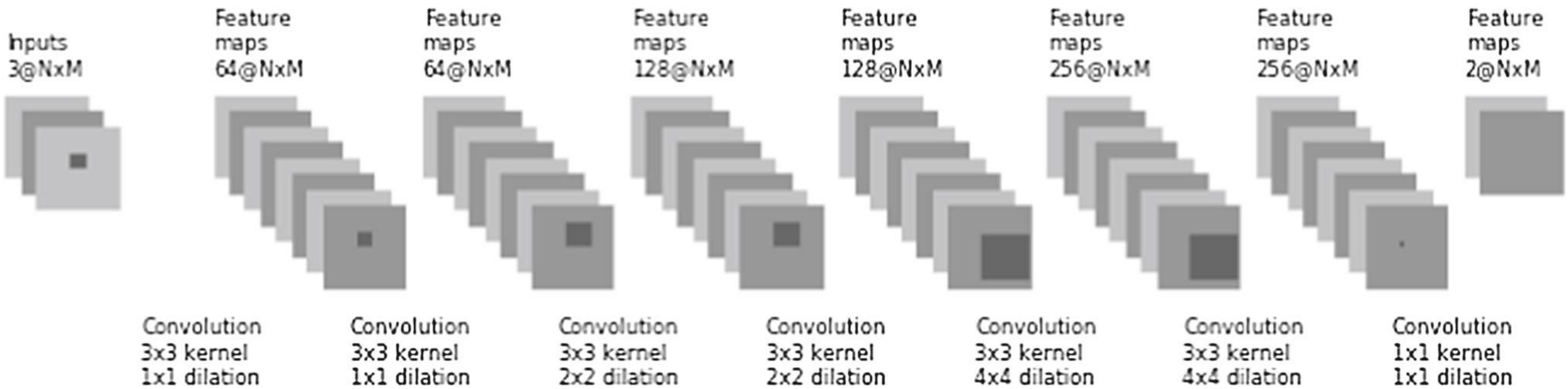




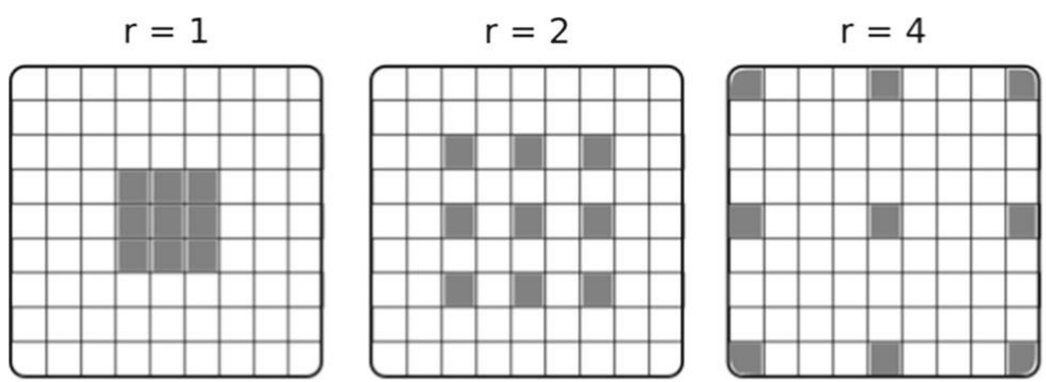
Example of
x-height labeling

G. Renton, et al. Fully convolutional network with dilated convolutions for handwritten text line segmentation, IJDAR, 2018.

FCN with dilated convolutions



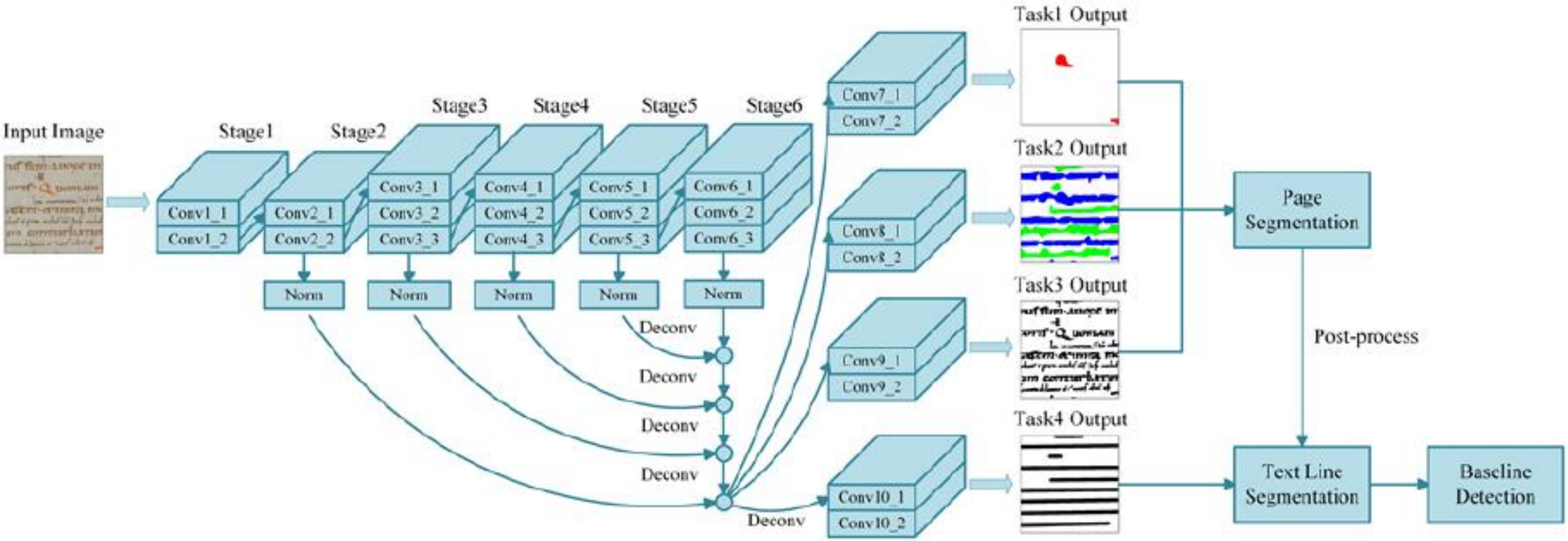
The input resolution is always the same and the receptive fields are increased due to the dilation



Receptive field of dilated convolution for different dilation rate r

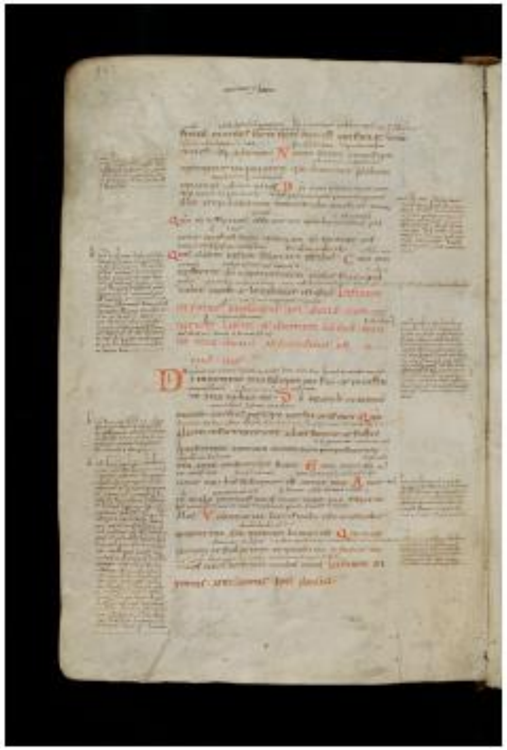
Multi-Task Layout Analysis Using FCN

FCN with three tasks: 1) region segmentation (text, background, comment, decoration); 2) text line contour extraction; 3) baseline detection



Four output branches: 1) decoration detection; 2) text and comment detection, coarse text line contour; 3) text/background separation; 4) center line detection

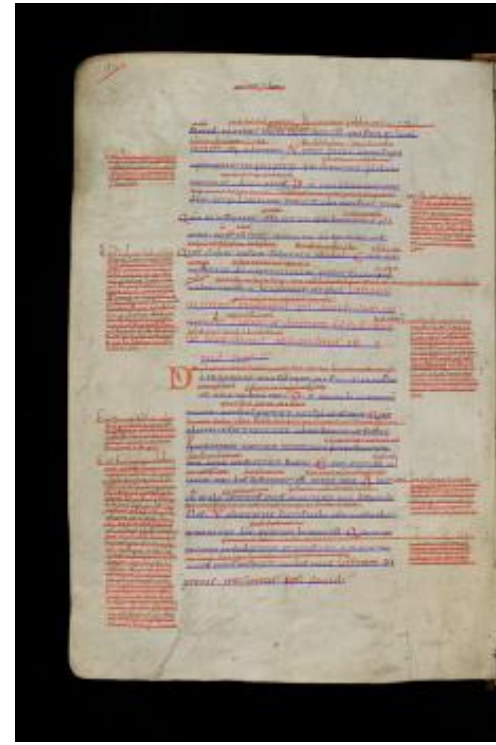
Y. Xu, F. Yin, Z. Zhang, C.-L. Liu, Multi-task Layout Analysis for Historical Handwritten Documents Using Fully Convolutional Networks, IJCAI 2018.



Input image

Page segmentation

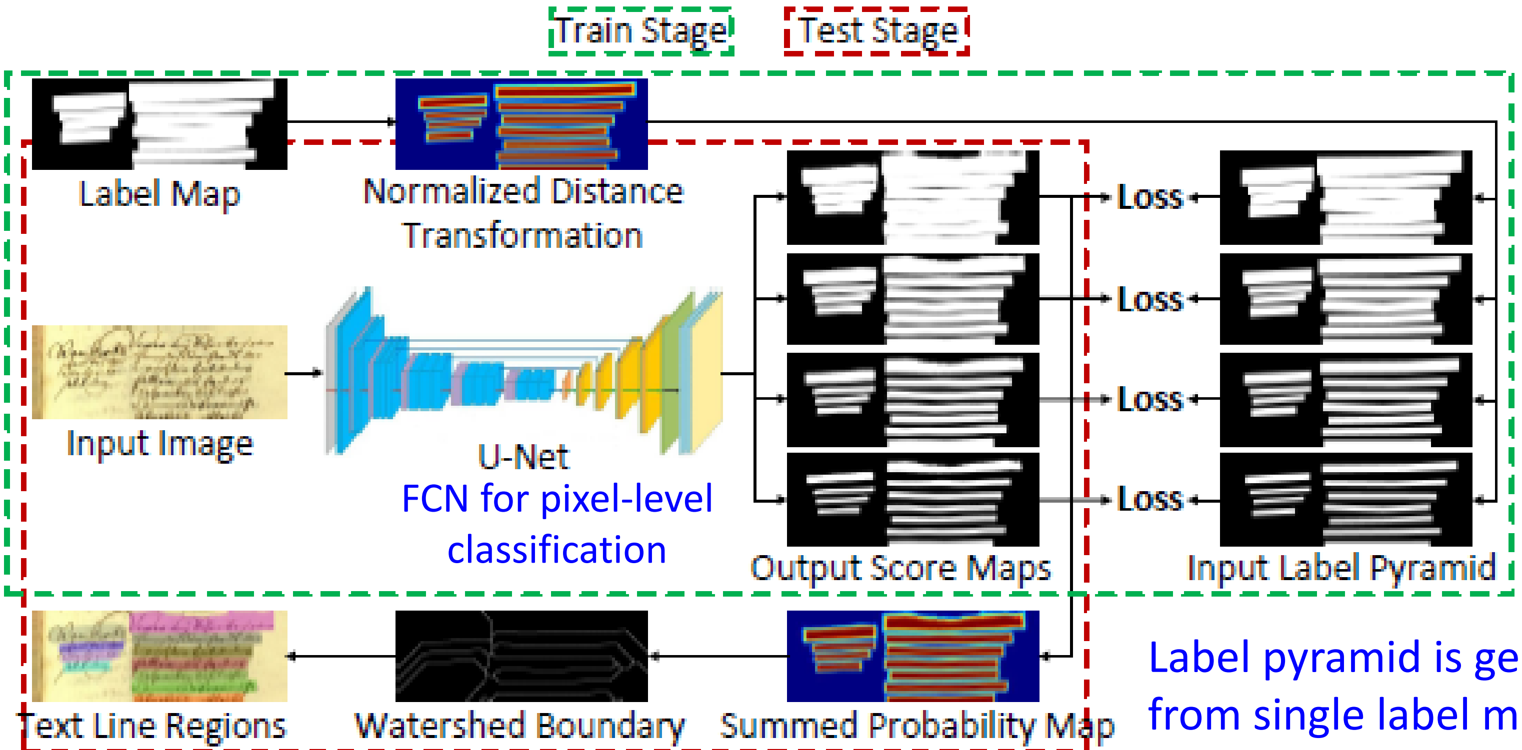
Text line segmentation



Baseline detection

Page Segmentation Using Label Pyramid Network

- To overcome the ambiguous boundary between text lines, by exploiting the hierarchical label information

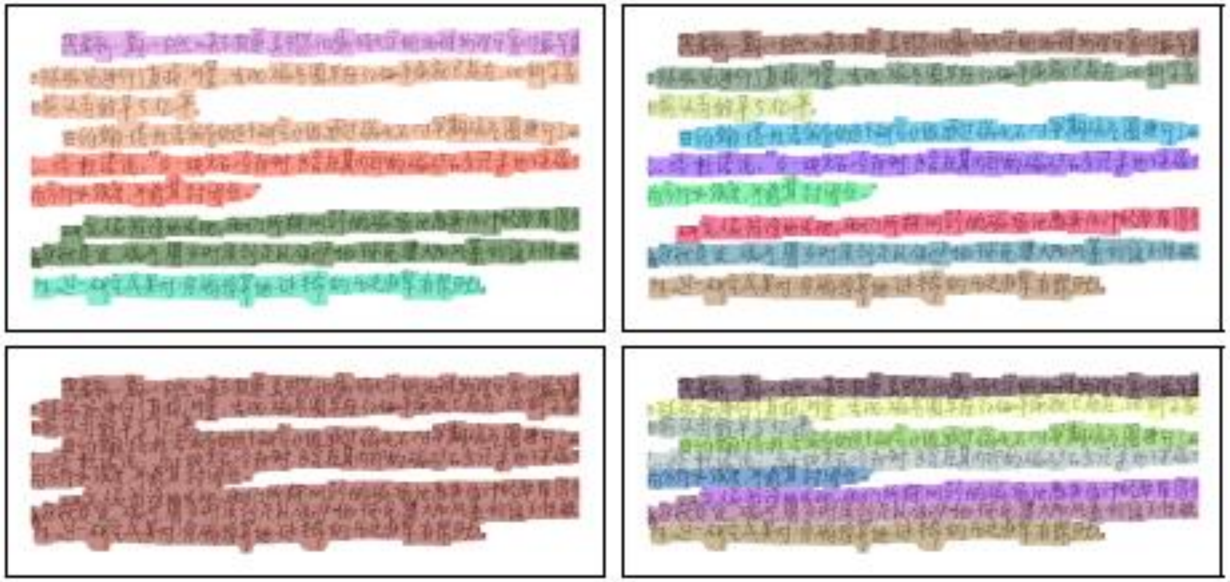


Label pyramid is generated from single label map by distance transform

X.-H. Li, et al., Instance Aware Document Image Segmentation using Label Pyramid Networks and Deep Watershed Transformation, submitted to ICDAR 2019.

Original image

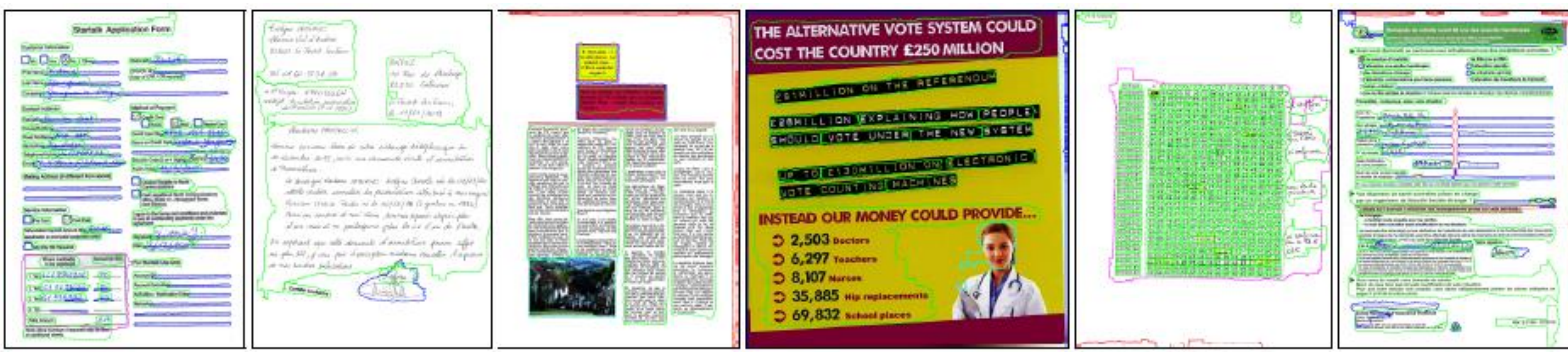
Image with line gap narrowed



LPN able to separate very close text lines

Left: FCN

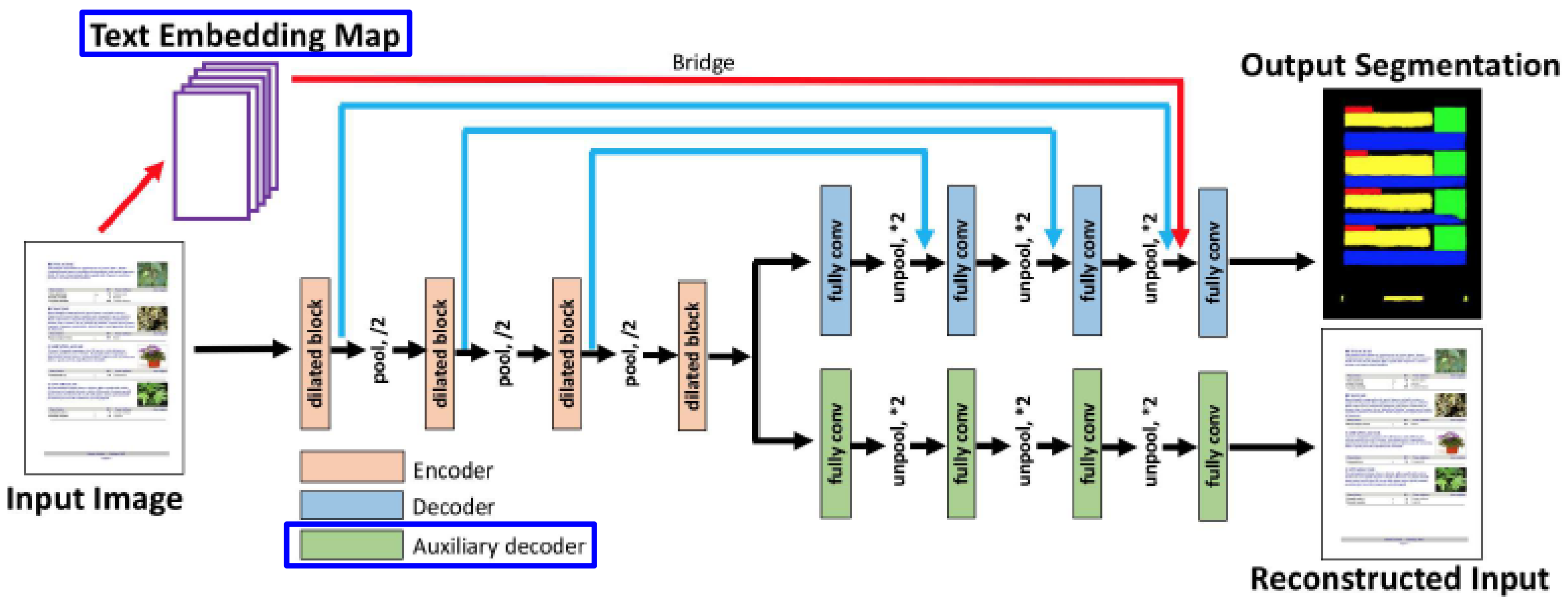
Right: LPN



Examples of region segmentation and text line segmentation

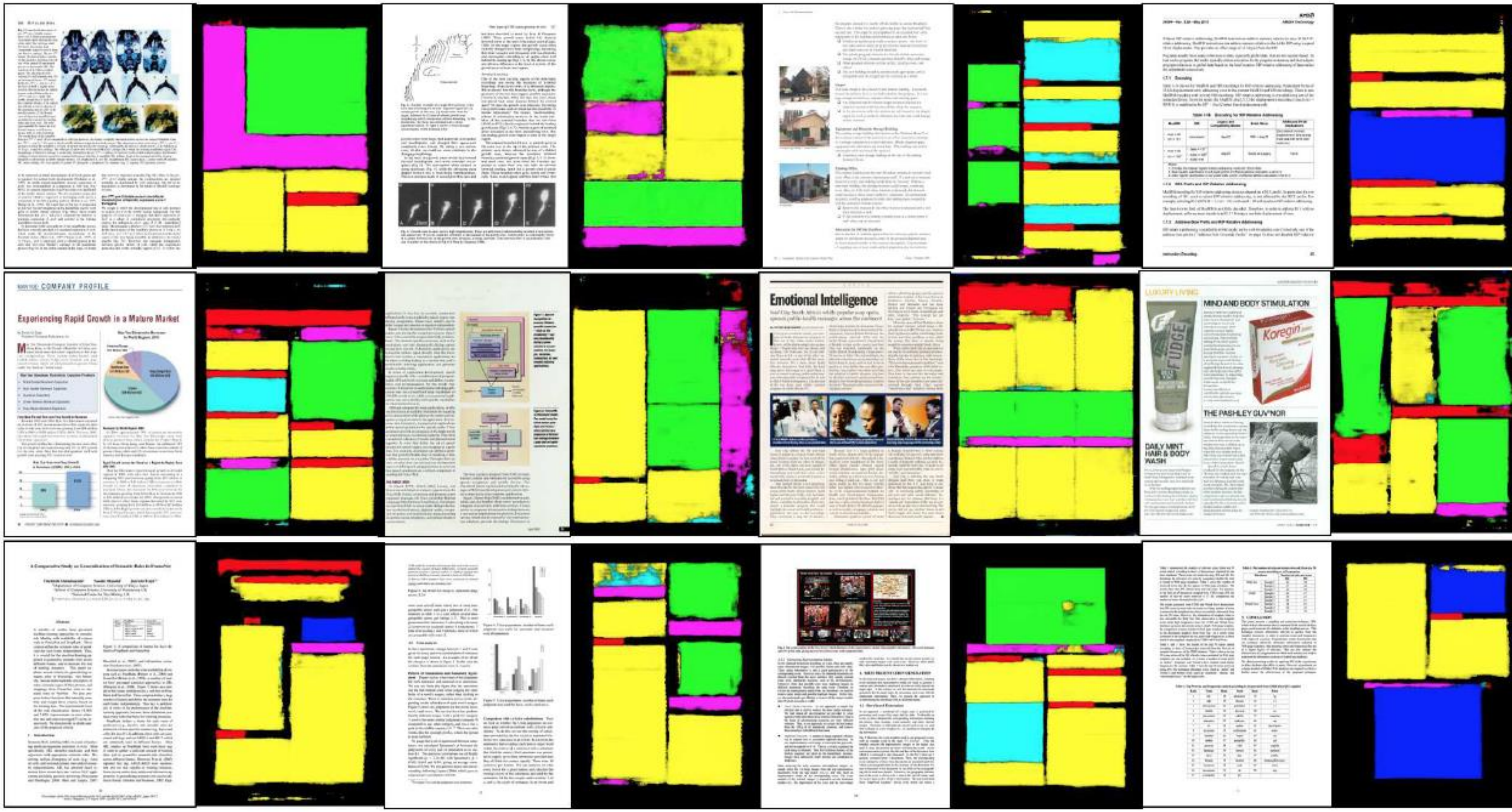
Learning to Extract Semantic Structure Using Multi-Modal FCN

- Document Semantic Structure Extraction (DSSE) as pixel-wise segmentation
  - Appearance-based and semantics-based classes
  - Using text embedding and unsupervised tasks to improve performance
  - Synthetic document generation



Unsupervised tasks: reconstruction, within-object consistency

X. Yang, et al., Learning to extract semantic structure from documents using multimodal fully convolutional neural networks, CVPR 2017.



Example real documents and their corresponding segmentation.
Top: DSSE-200. Middle: ICDAR2015. Bottom: SectLabel.
Segmentation label colors are: figure , table , section heading , caption , list and paragraph

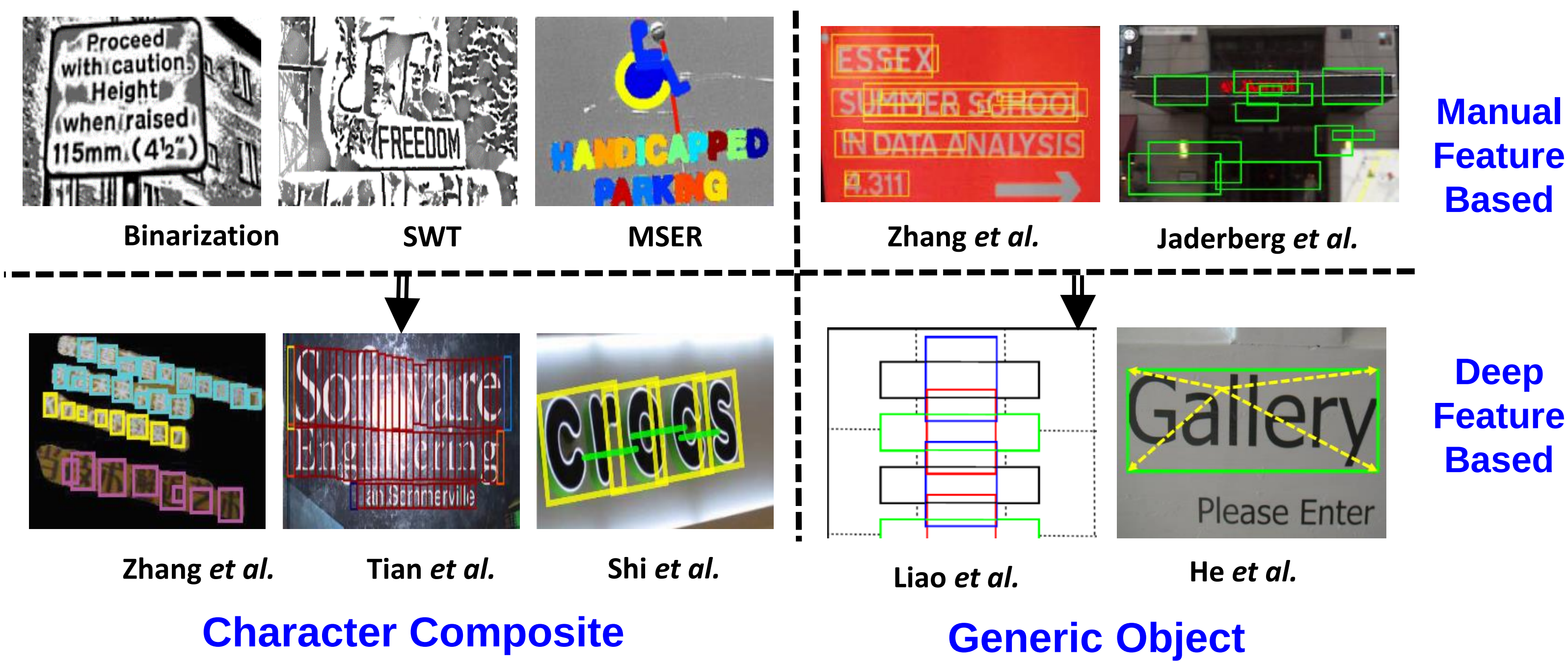
Scene Text Detection

- Difficulties
  - Complex background, change of illumination and perspective
  - Multi-oriented text, arbitrary shaped text
- Research History
  - Started from 1990s
  - Prevalent from 2011
    - Robust reading competitions in ICDAR 2011, 2013, 2015
  - Deep learning from 2016
  - Horizontal→Multi-oriented (2015)→Arbitrary shape (2017)
- Research Tasks
  - Text detection
  - Text recognition, mostly following generic text line recognition
  - End-to-end text detection-recognition (a.k.a. text spotting)
    - Joint model (e.g., shared feature extraction), multi-task learning, saved memory, improved performance

Text Detection Methods

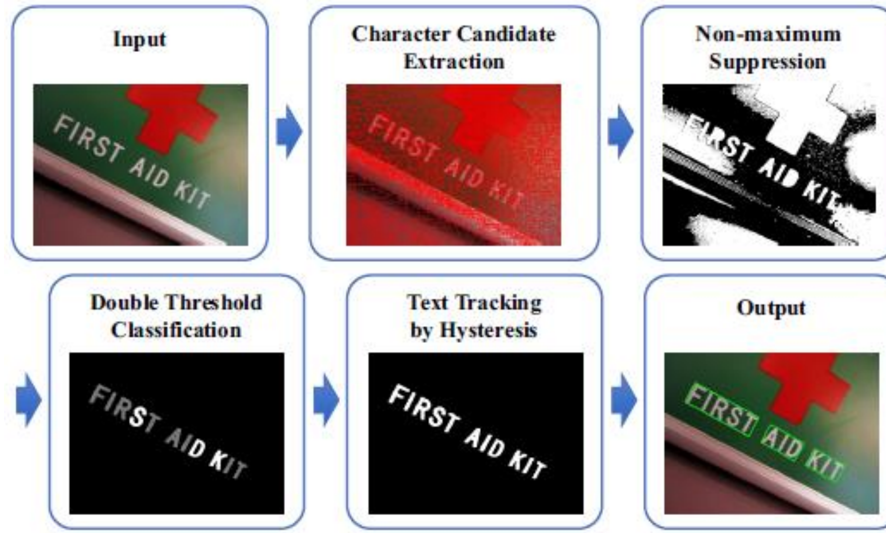
- Character based
  - MSER, SWT, …
  - Sliding window, text-line block/slice
- Word/line based (generic object)
  - Manually designed (hand-crafted) feature
  - CNN based object detection

DL-based approaches: largely motivated by object detection methods

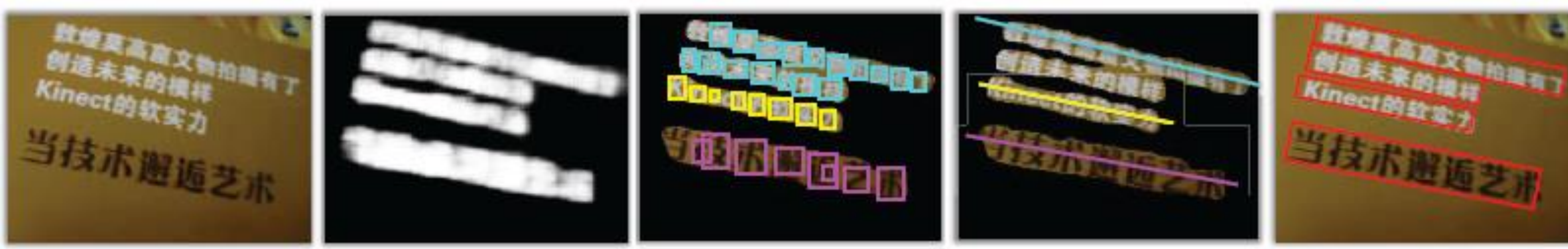


- Character based Scene Text Detection
  - Component-based: SWT, MSER (ER), FCN
  - Text Block/Slice

Linking segments, Shi et al., *CVPR* 2017



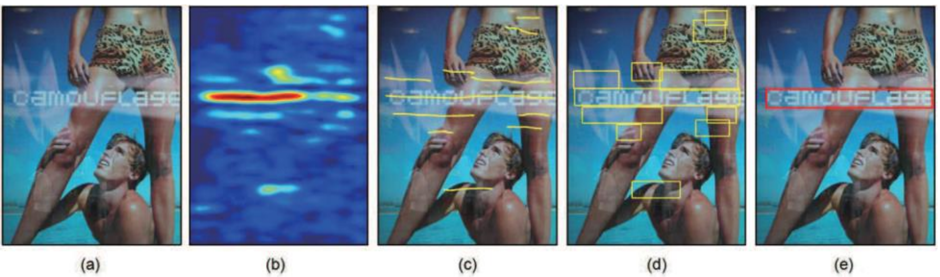
Canny text detector, Cho et al., *CVPR* 2016



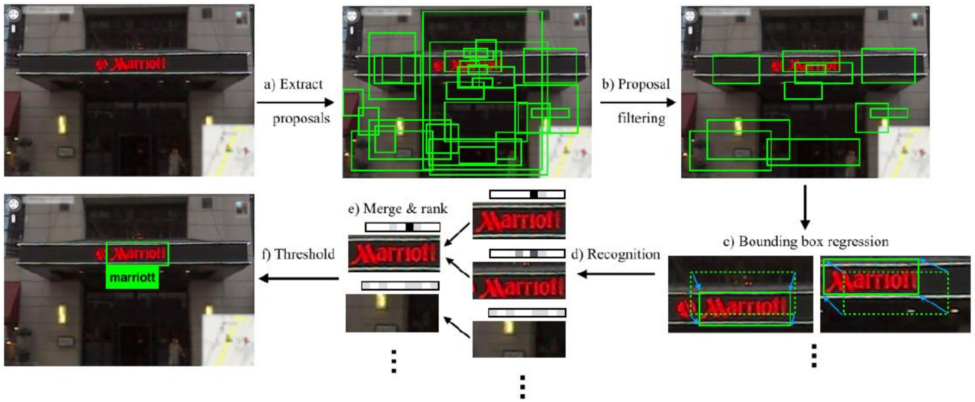
Saliency map predicted by text-block FCN, Zhang et al., CVPR 2016

(FCN to predict the saliency map of texts, character hypothesis)

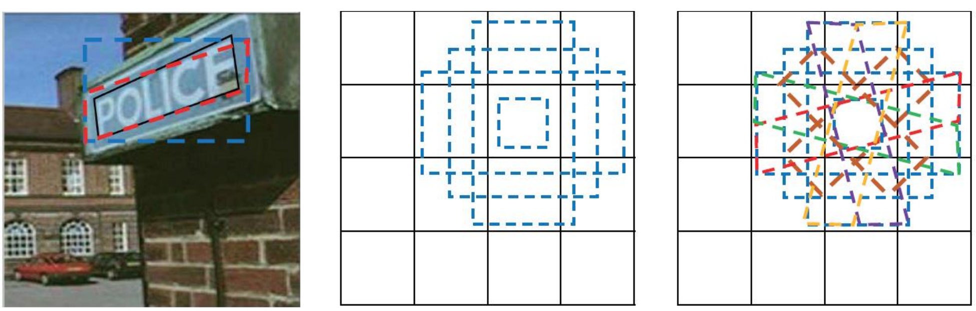
- Word/Line based Scene Text Detection
  - Manually designed features
  - CNN based features



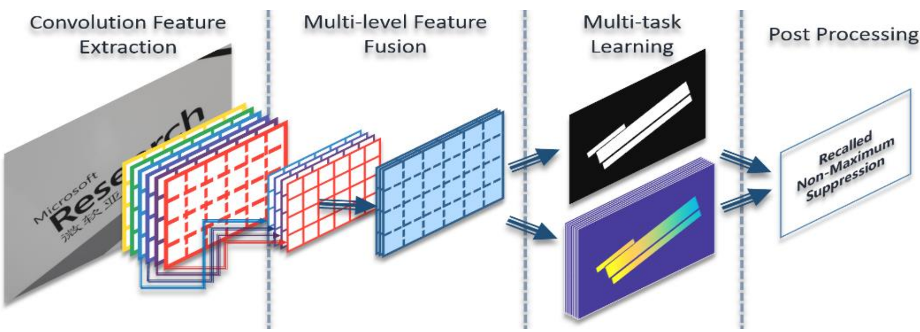
Symmetry Feature, Zhang et al., *CVPR* 2015



ACF and Edge Boxes, Jaderberg et al., *IJCV* 2016



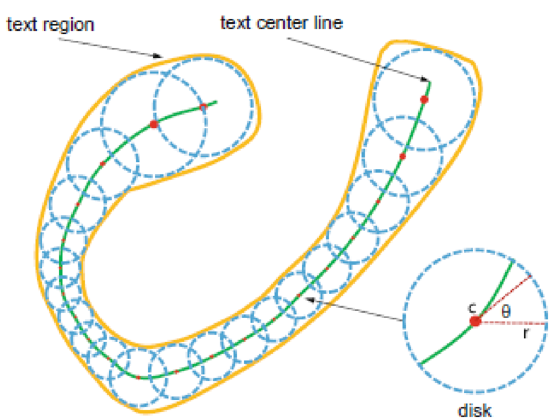
Rotated Anchors (RRPN), Liu et al., *CVPR* 2017

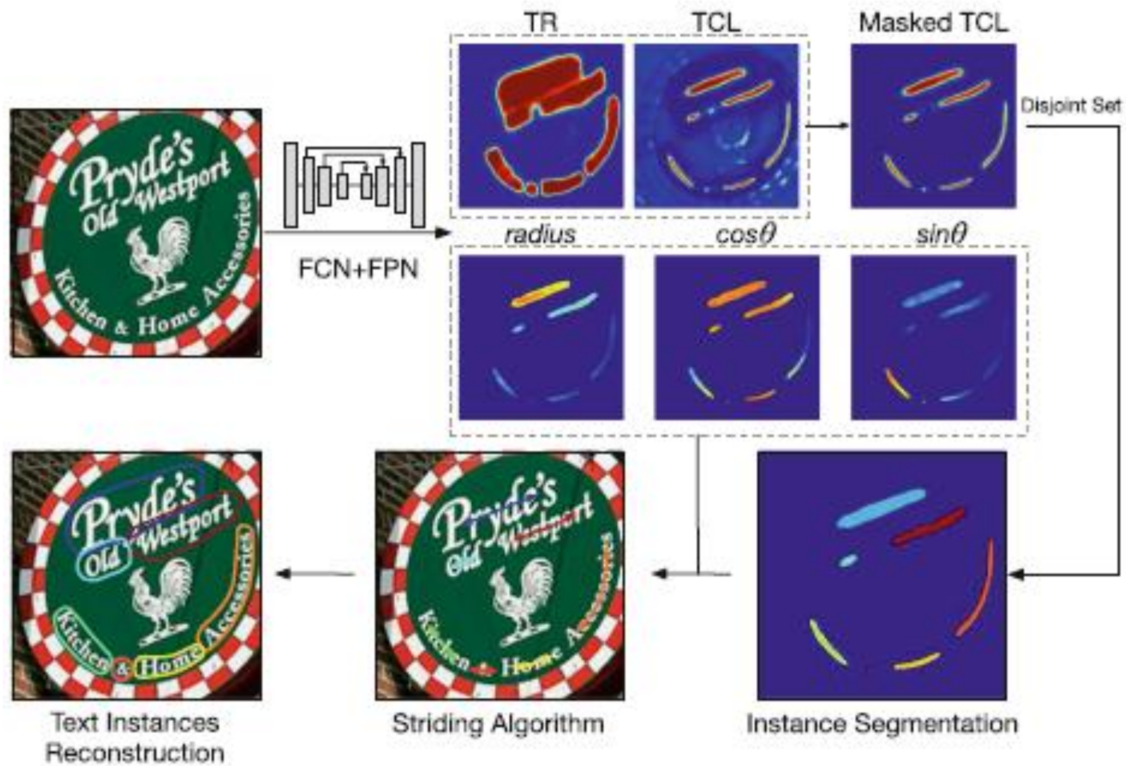


Direct Regression, He et al., *ICCV* 2017

Arbitrary Shape Text Detection

- TextSnake: local region prediction and reconstruction
  - Text as sequence of ordered overlapping disks
  - Score maps of text center lines (TCLs) and text regions (TRs)
  - Instance segmentation, central axis points extraction





S. Long, J. Ruan, W. Zhang, X. He, W. Wu, C. Yao, TextSnake: A flexible representation for detecting text of arbitrary shapes, *ECCV 2018*, LNCS 11206, pp.19-35, 2018.

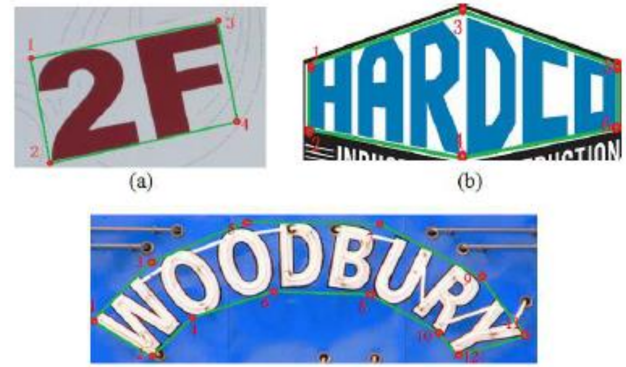
Yellow: detected text boundary
Green: ground-truth annotation

Red: score map for TR (text region)
Yellow: score map for TCL (center line)

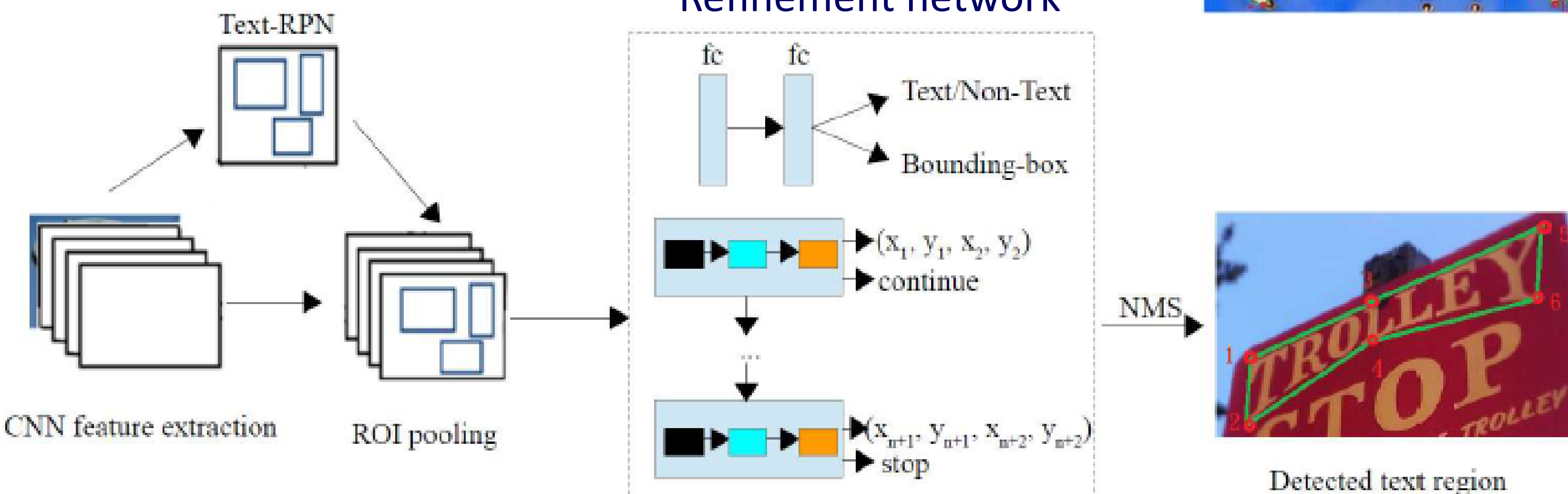
The approach works for both curved and multi-oriented texts.

Arbitrary Shape Text Detection

- Adaptive Text Region: flexible number of boundary points



Refinement network



Region proposal network (RPN)

RNN based adaptive region representation: coordinates of pairwise points as regression target

X. Wang, Y. Jiang, Z. Luo, C.-L. Liu, H. Choi, S. Kim, Arbitrary shape scene text detection with adaptive text region representation, *CVPR 2019*.


Phil's
MARKET
RED CROWN
GASOLINE
Firestone
Most Miles per Dollar
Arby's
ROAST
BEEF
Sandwich

(a)


CANNERY ROW
MALL
STREET
TEA HOUSE
WINE TASTING
GASTRONOMY

(b)


Summer's
Here!
CO-OPERATIVE

(c)


TOAST BOX

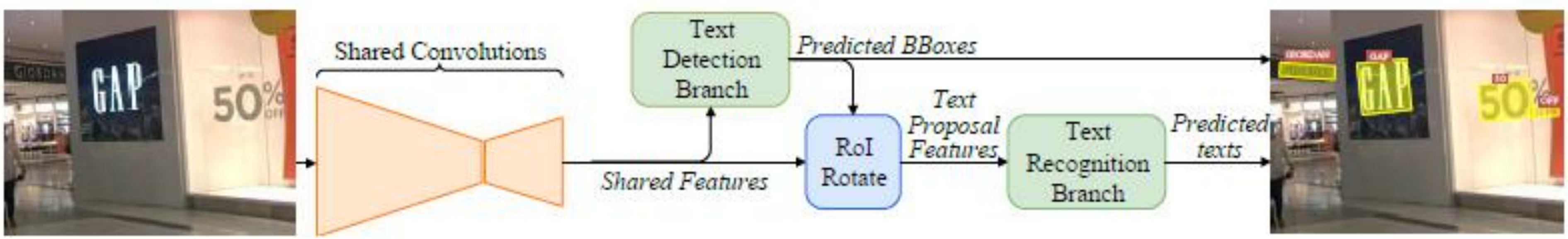
(d)


丹棱街
海淀中街
香埠街
P
电子大厦
国机集团
剩余车位

(e)

End-to-End Text Detection-Recognition

- Cascaded system vs joint model
- FOTS: Fast oriented text spotting
  - ROIRotate to share convolutional features between detection and recognition
  - Text detection branch: FCN, binarization, NMS
  - Text recognition branch: LSTM+CTC
  - Multi-task training



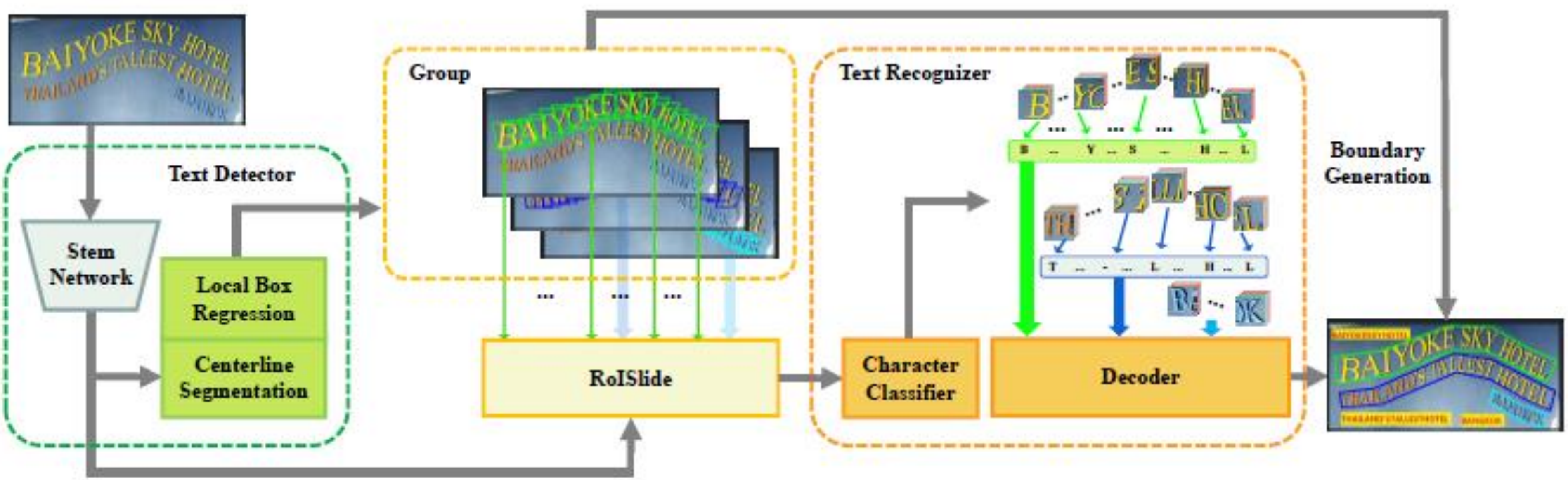
X. Liu, D. Liang, S. Yan, D. Chen, Y. Qiao, J. Yan, FOTS: Fast oriented text spotting with a unified network, CVPR 2018.

(a) ICDAR 2015 (b) ICDAR 2017 MLT (c) ICDAR 2013

Both detection and recognition results

End-to-End for Arbitrary Shape Text

- TextDragon
  - Local box detection, centerline prediction
  - ROISlide to extract convolutional features for recognition
  - Sliding character classifier based text recognition



W. Feng, W. He, F. Yin, X.-Y. Zhang, C.-L. Liu, TextDragon: An end-to-end framework for arbitrary shaped text spotting, ICCV 2019.

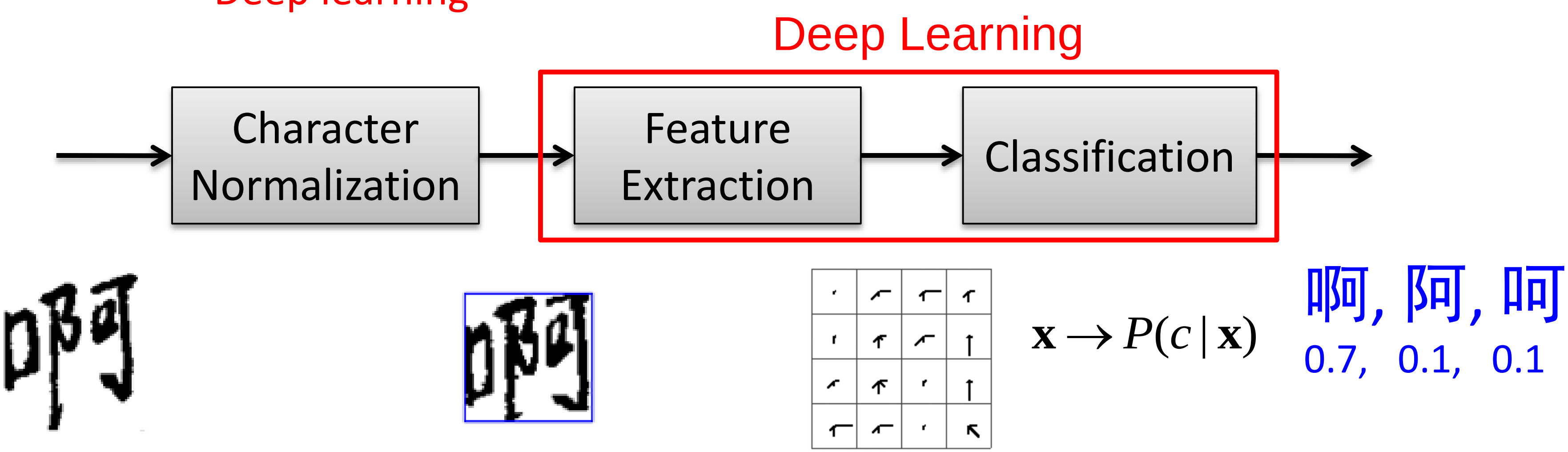
End-to-end training helps text detection

Detection and Recognition results on three datasets

CTW1500 Total-Text ICDAR 2015

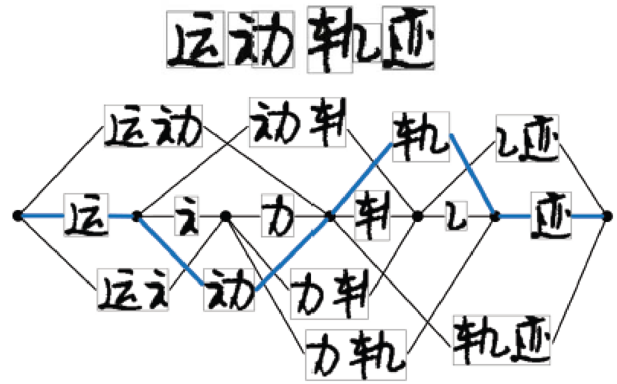
Text Recognition

- 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

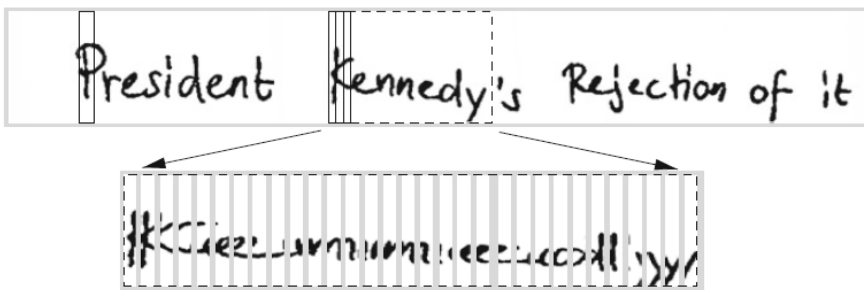


Multiple candidate classes for integration in string recognition incorporating linguistic context

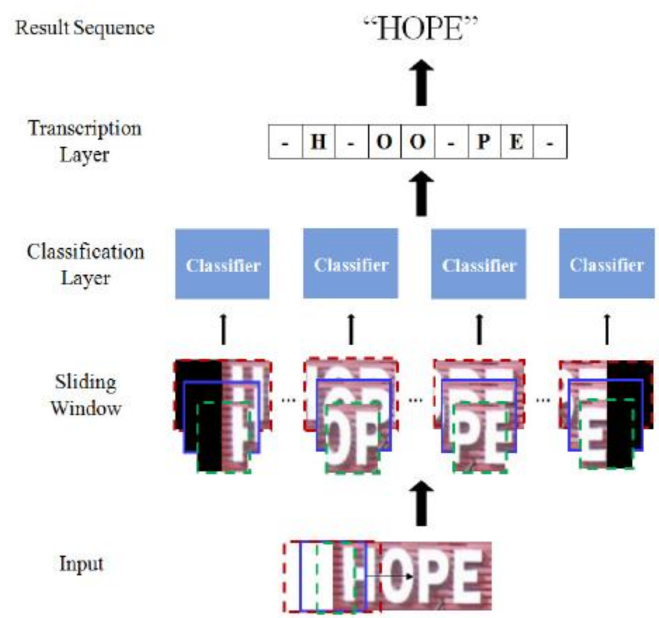
- Text (word/line) Recognition
  - Explicit/over segmentation
    - Relevant to human cognition
    - Good for fusing contexts and knowledge



  - Implicit segmentation: sliding window
    - HMM

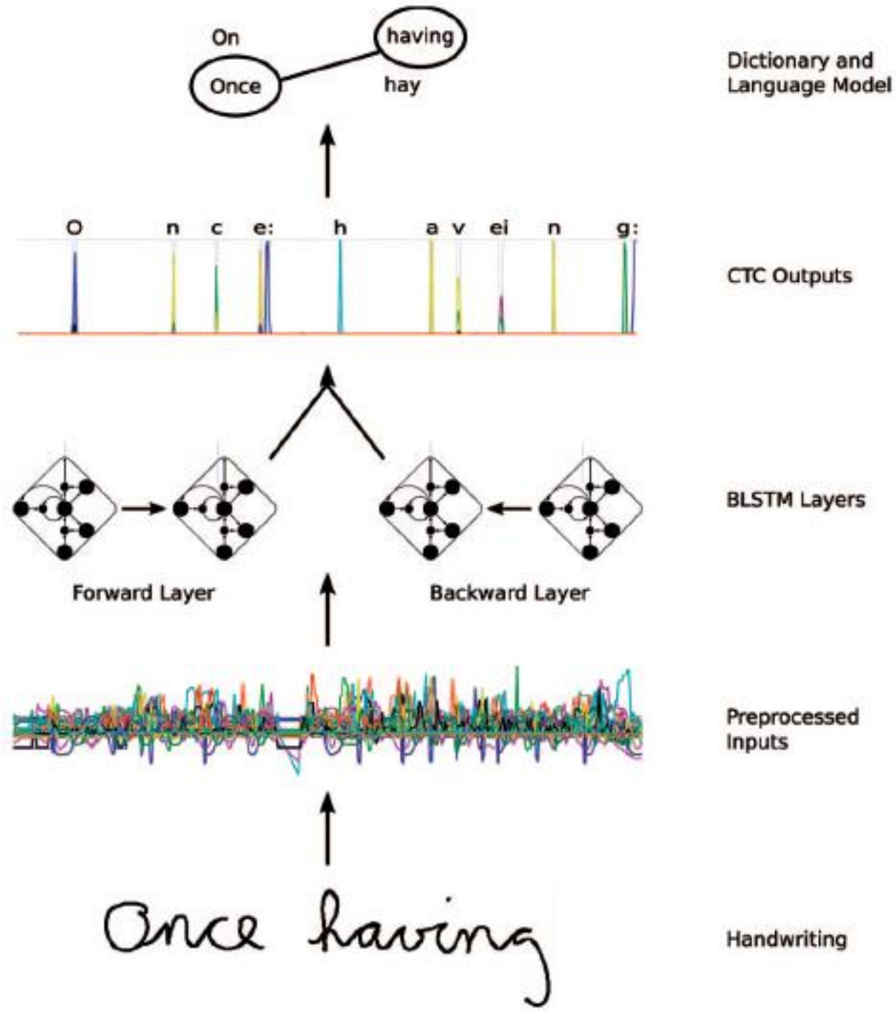


    - RNN, BLSTM
      (bidirectional long short-term memory)
    - BLSTM combined with CNN (CRNN)
    - Sliding window classifier
      (Applicable to large category set)



Handwriting Recognition Using RNN-LSTM

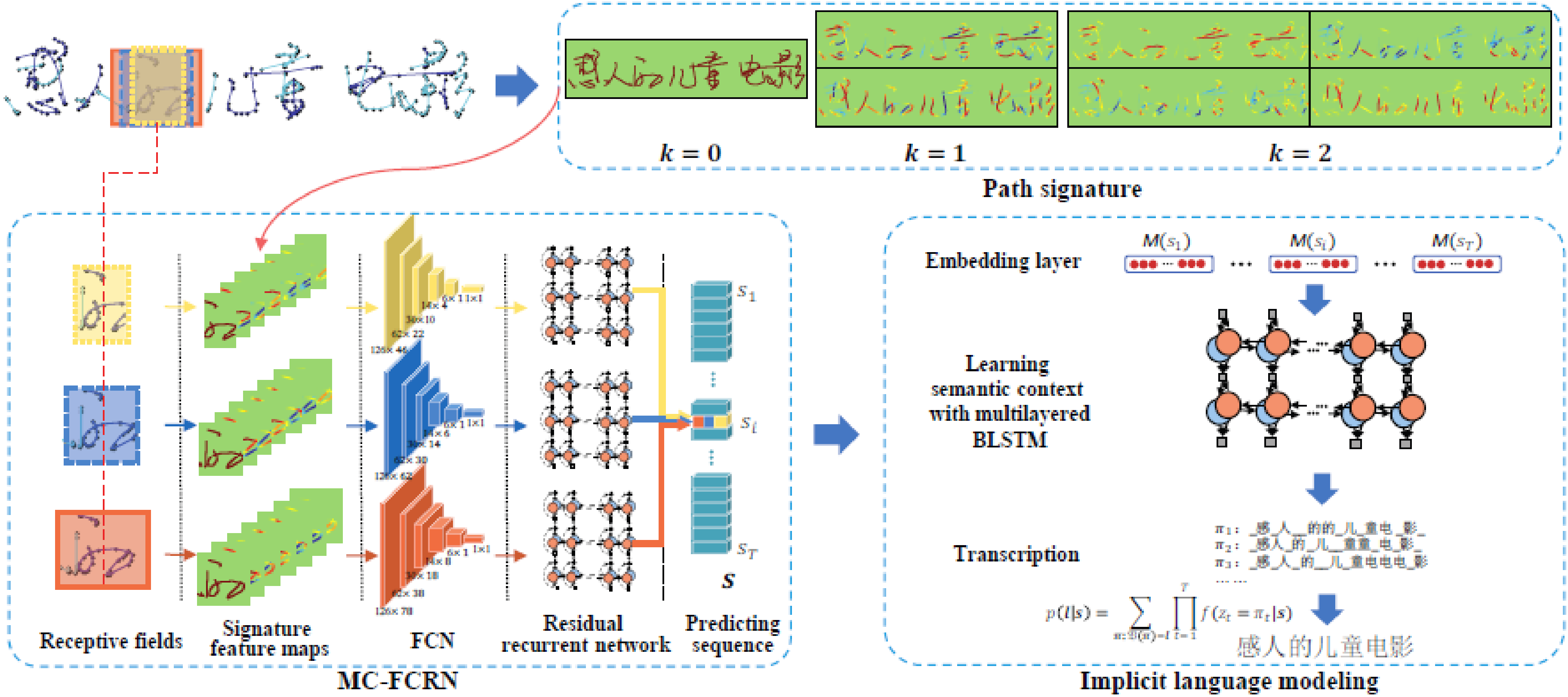
- LSTM (long-short-term memory) units to better model long-range dependency.
- Superior performance in text recognition of various styles (online/offline handwriting, printed, scene texts)



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

Online Handwritten Text Recognition with Convolutional RNN

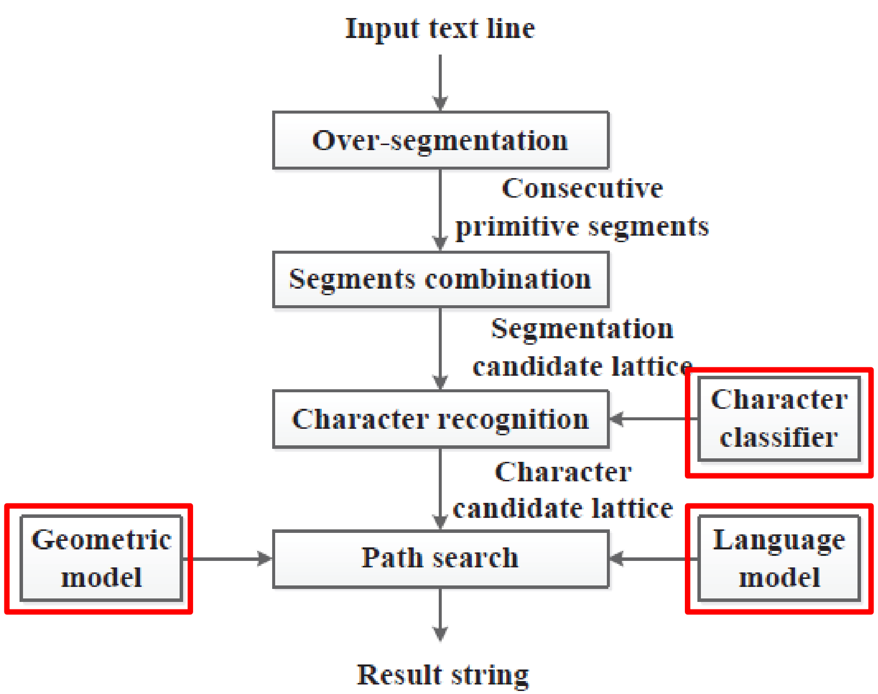
Multi-Spatially-Context Fully Convolutional Recurrent Network (MC-FCRN)

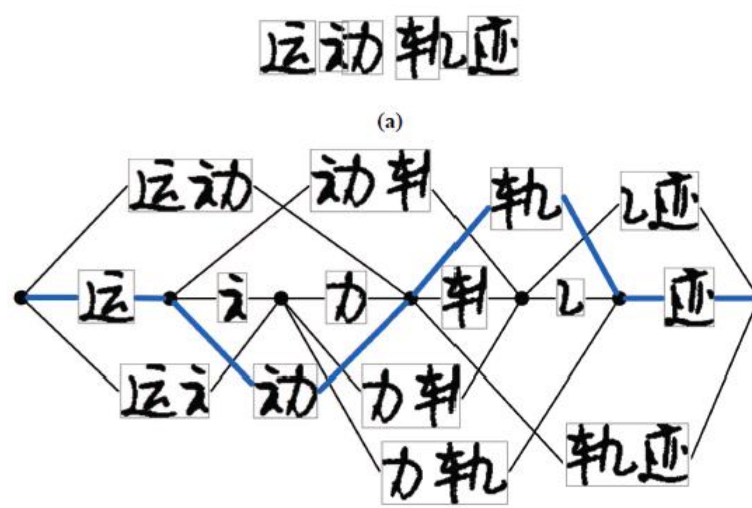


Z. Xie, Z. Sun, L. Jin, H. Ni, T. Lyons, [Learning spatial-semantic context with fully convolutional recurrent network for online handwritten Chinese text recognition](), *IEEE Trans. PAMI*, 2018.

Offline Handwritten Text Recognition Based on Over-Segmentation

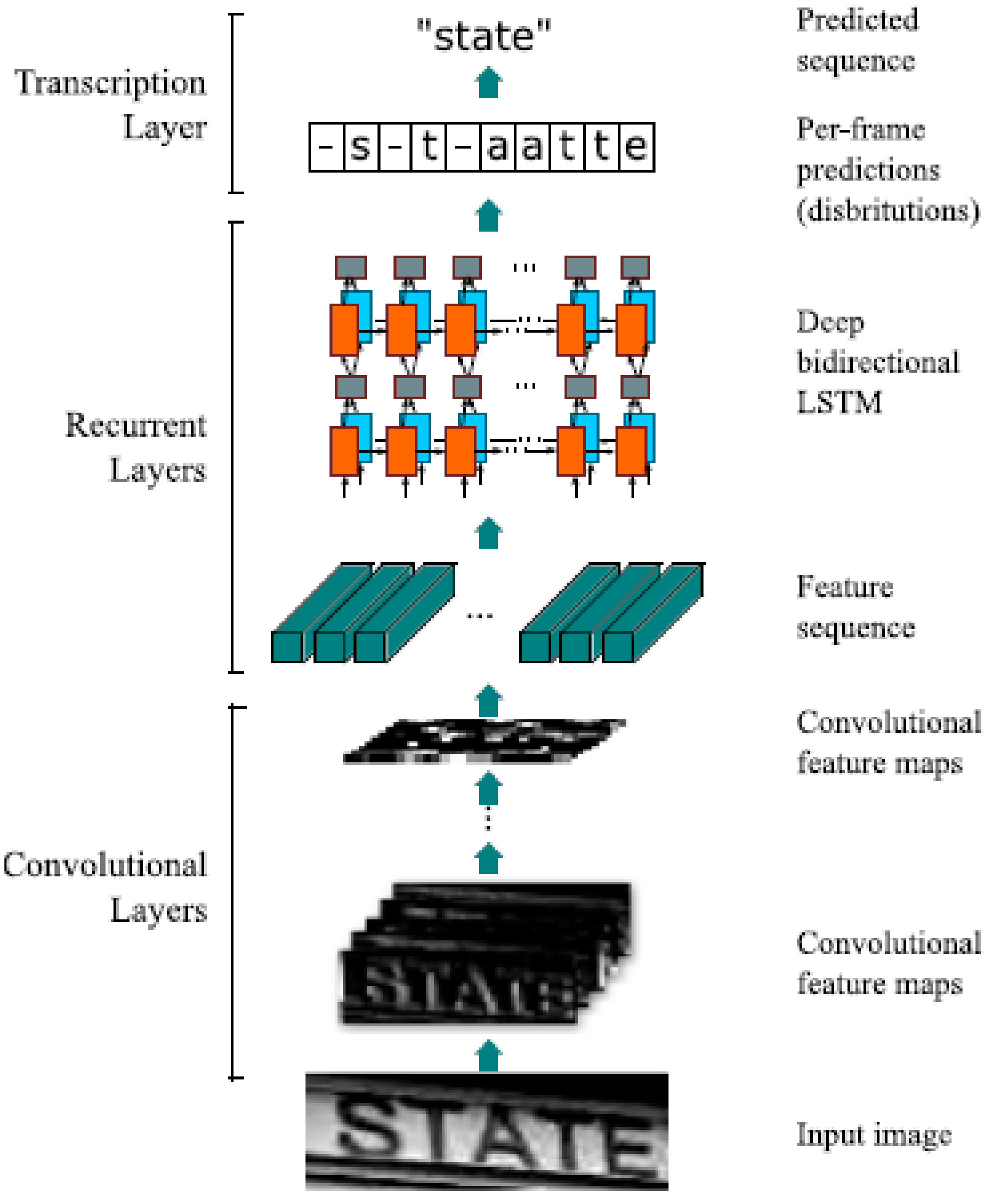
- Candidate segmentation-recognition path evaluation
- CNN for cut detection, character classification, geometric context
- RNN-based language model (character based)





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.

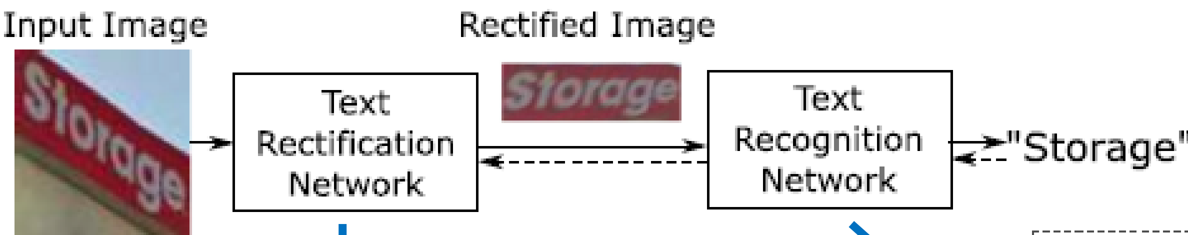
Scene Text Recogn Using Conv-RNN

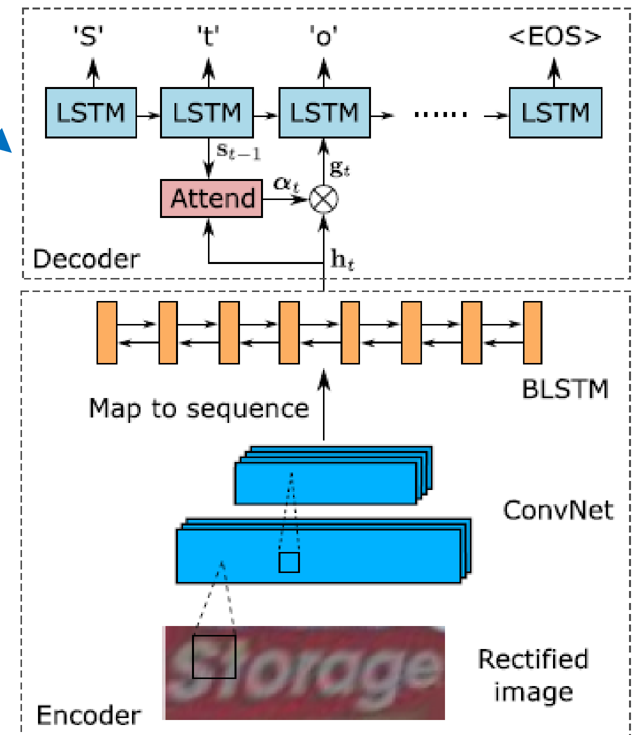


Conv-RNN (LSTM) now dominates in text recognition

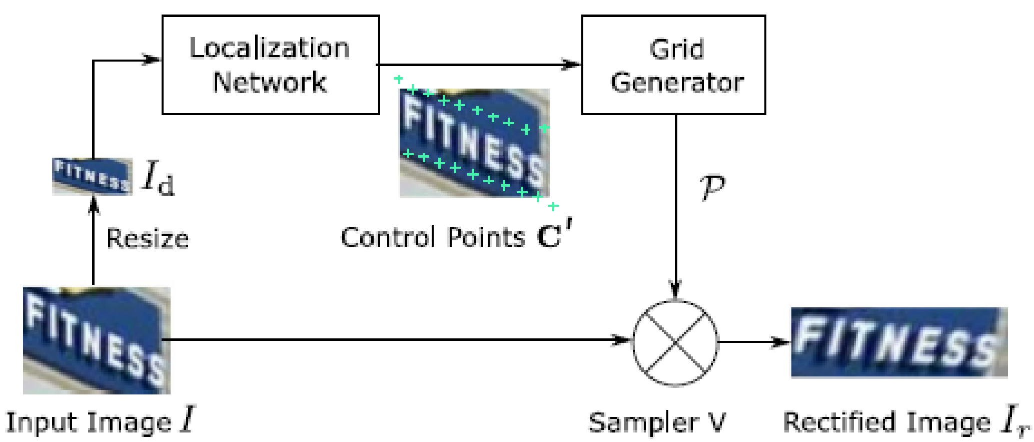
B. Shi, X. Bai, C. Yao, An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition, *IEEE Trans. PAMI*, 2017.

Curved Text Recognition with Attentional Network



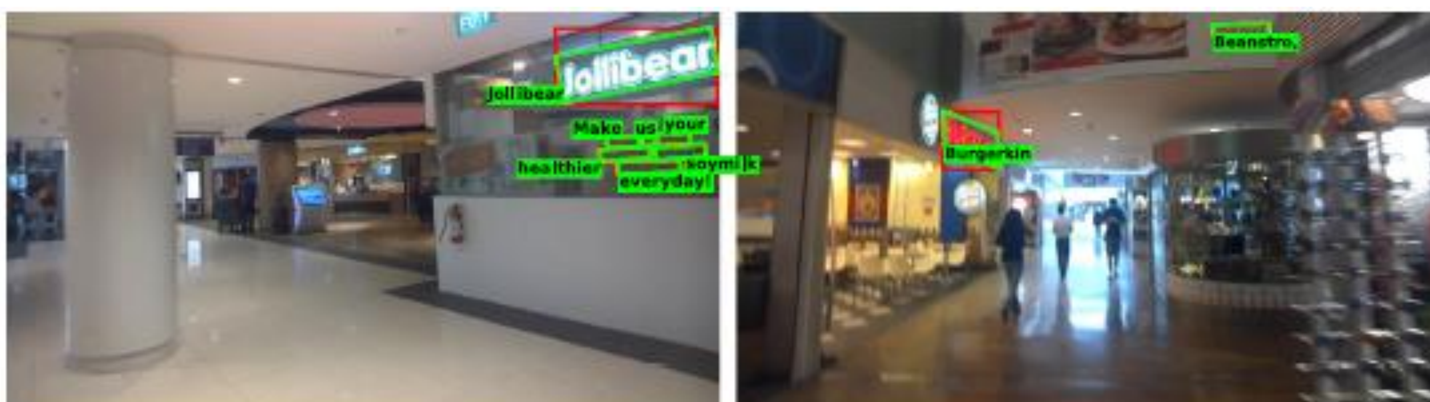


Rectification network uses thin plate splines for rectifying text shape.



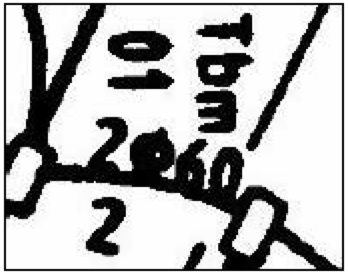
B. Shi, M. Yang, X. Wang, P. Lyu, C. Yao, X. Bai, ASTER: An Attentional Scene Text Recognizer with Flexible Rectification, *IEEE Trans. PAMI*, 2019.

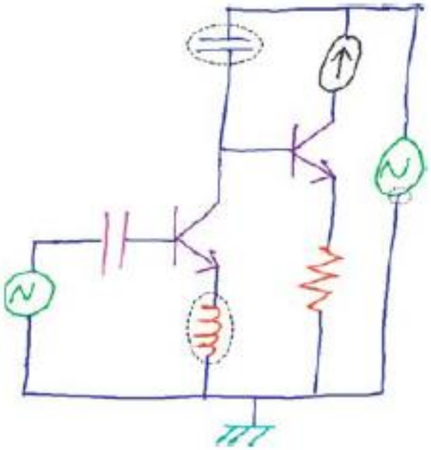
Detection using TextBoxes and rectification-recognition using ASTER



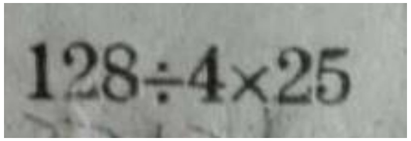
Graphics/Symbol Recognition

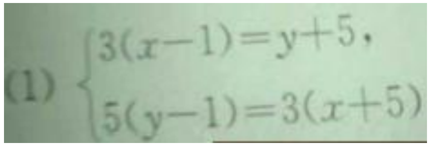
- Graphics/symbols in many documents, though less than texts
  - Important for many applications
    - Mathematic expressions/flowchart: education
    - Signature: forensics
    - Tables: business forms
  - Re-drawing attention because text recognition works well now
- Approaches
  - Engineering drawings
    - Primitive extraction
    - Graph matching
  - Flowchart
    - Stroke labeling: MRF, CRF
    - Rule-based interpretation

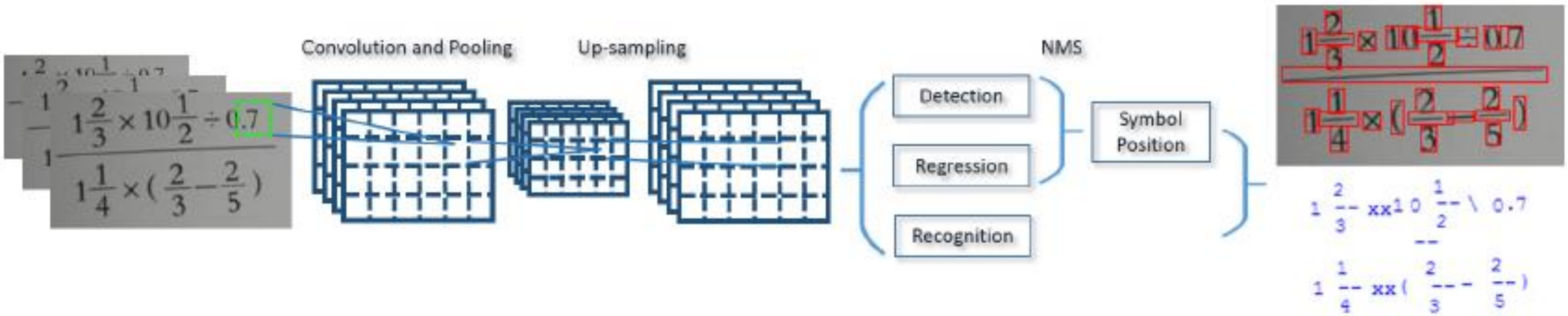




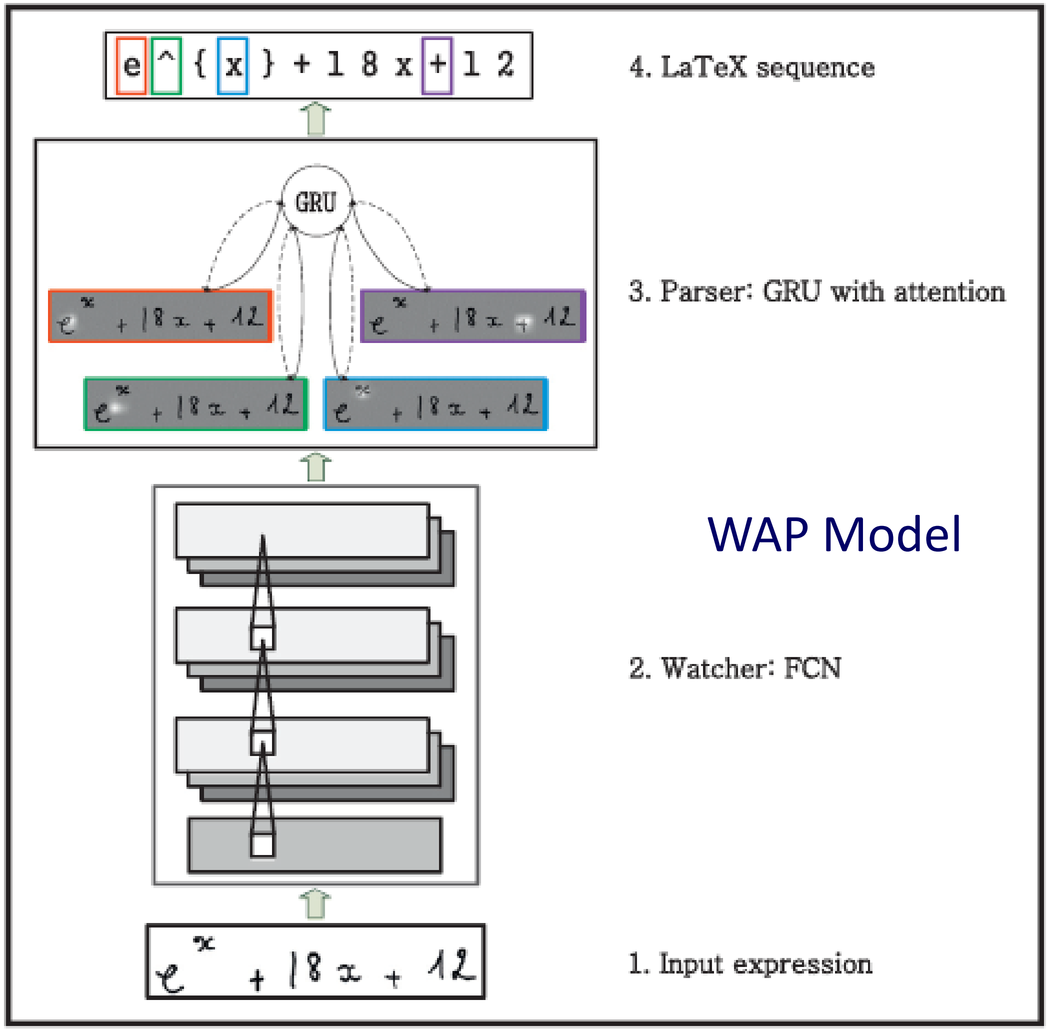
- Approaches
  - Logo recognition/retrieval
    - Similar to generic object detection/recognition
  - Signature verification
    - Feature extraction-matching
    - Deep learning based: Siamese network
  - Mathematic expressions
    - Symbol segmentation
    - Symbol recognition
    - Graph/grammar/rule-based interpretation
    - Fully convolutional network for simultaneously symbol detection ad recognition



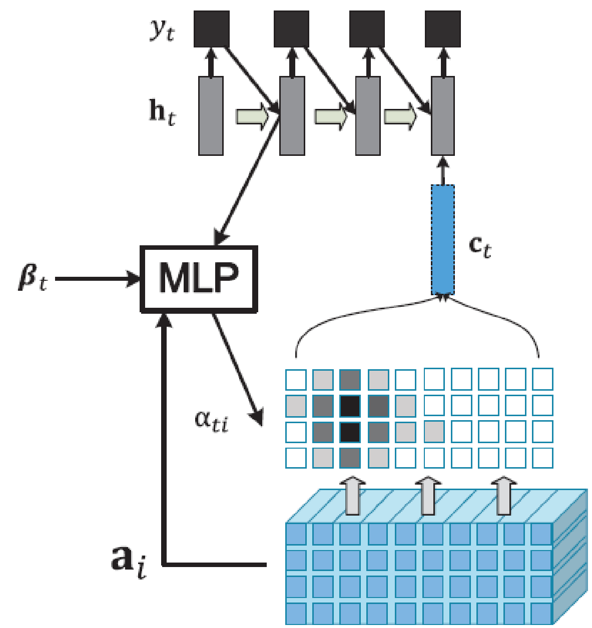




Attentional Network for HME Recognition



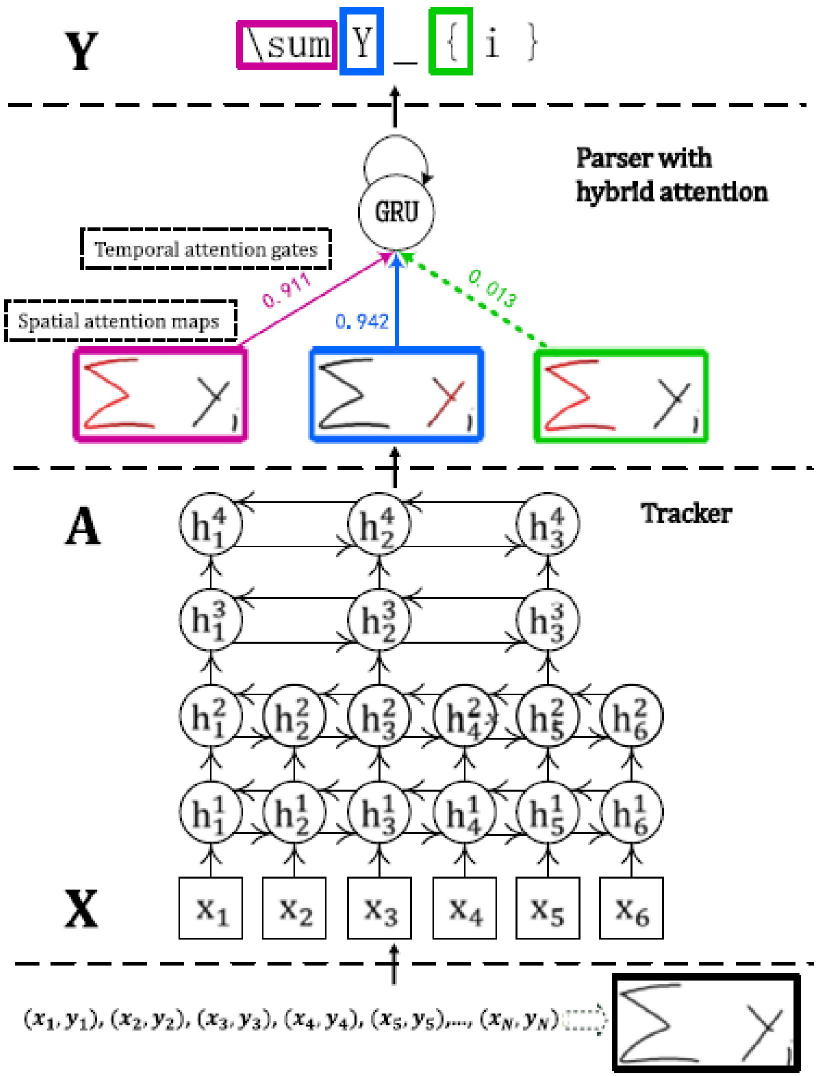
End-to-end recognition, from 2D image to character sequence in Latex format



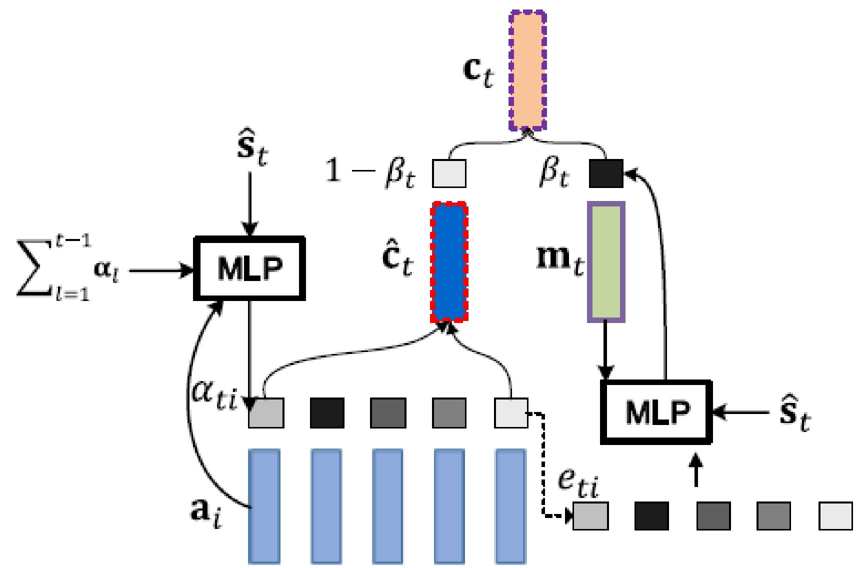
Coverage vector based attentional model

J. Zhang, J. Du, et al., Watch, attend and parse: An end-to-end neural network based approach to handwritten mathematical expression recognition, *Pattern Recognition*, 2017.

TAP Network for online HME Recognition



Utilize online sequence information



Hybrid attention

J. Zhang, J. Du, L. Dai, [Track, attend, and parse (TAP): An end-to-end framework for online handwritten mathematical expression recognition](), *IEEE Trans. Multimedia*, 2019.

Status of Performance

- Layout Analysis
- Numeral Recognition
- Handwriting Recognition
- Chinese Character and Text
- Scene Text Detection and Recognition
- Mathematics Recognition

Layout Analysis

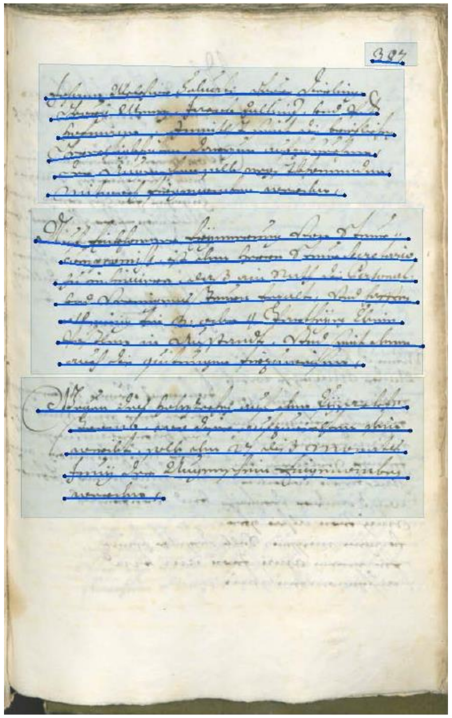
- Datasets
  - University of Washington (UW) document image databases: UW-I, UW-II, UW-III
  - ICDAR Competition on Recognition of Documents with Complex Layouts (RDCL) 2001-2017
  - ICDAR Handwriting Segmentation Contest (2007, 2009, 2013)
  - READ-BAD database
  - DIVA-HisDB
  - Maurdor database

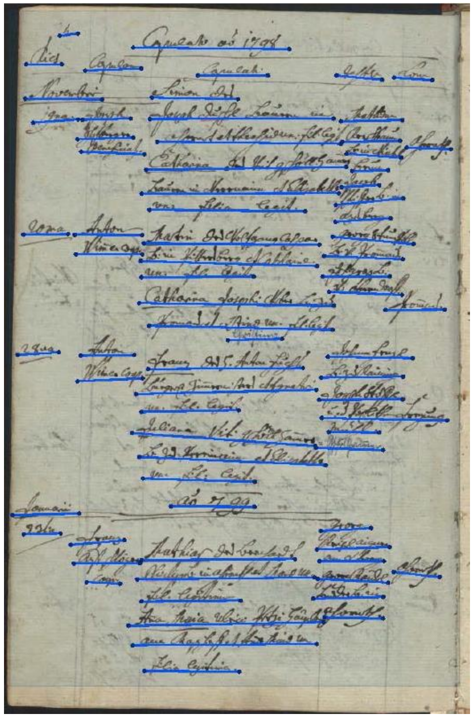
- # READ-BAD Database

For baseline detection

Dataset	Training set	Test set
Track A (Simple document)	216 pages	539 pages
Track B (Complex document)	270 pages	1010 pages

Samples with ground truth





Evaluation metric: Recall, Precision, F-value

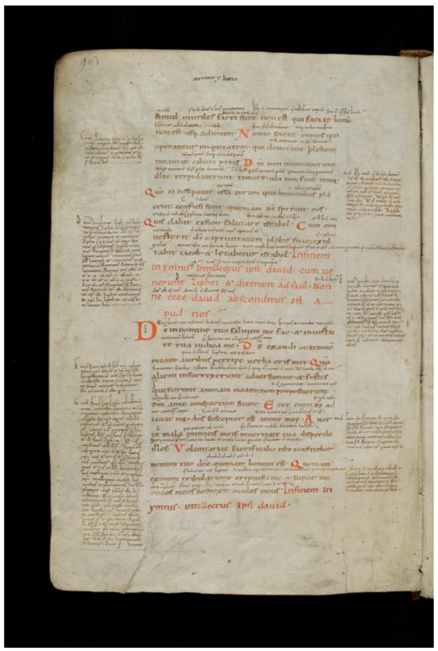
Reference baselines (ground-truth) annotated as polygonal chains

M. Diem, et al. cBAD: ICDAR2017 competition on baseline detection, ICDAR 2017.

- # DIVA-HisDB

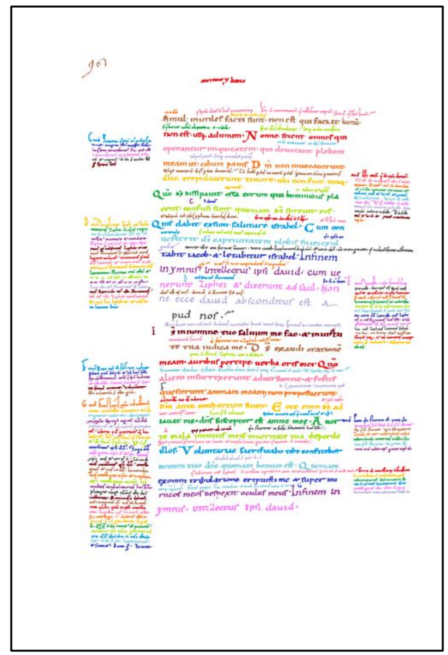
Historical: Medieval manuscripts, pixel-level annotation

Dataset	Training	Validation	Test
CB55	20 pages	10 pages	10 pages
CS18	20 pages	10 pages	10 pages
CS863	20 pages	10 pages	10 pages

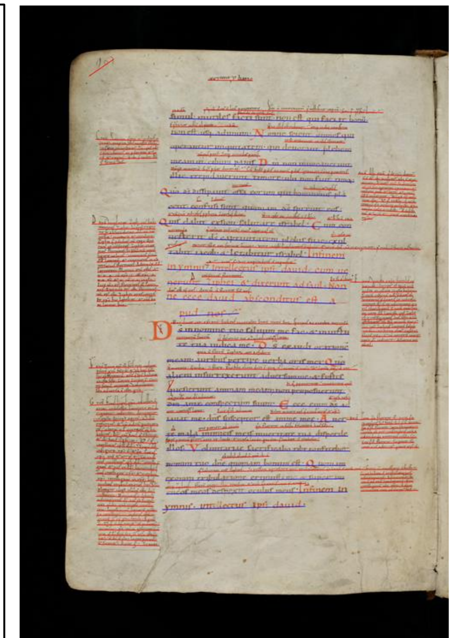


Original image

Task-1:
Layout analysis



Task-3:
Text line segmentation



Task-2:
Base line detection

F. Simistira, ICDAR2017 competition on layout analysis for challenging medieval manuscripts, ICDAR 2017.

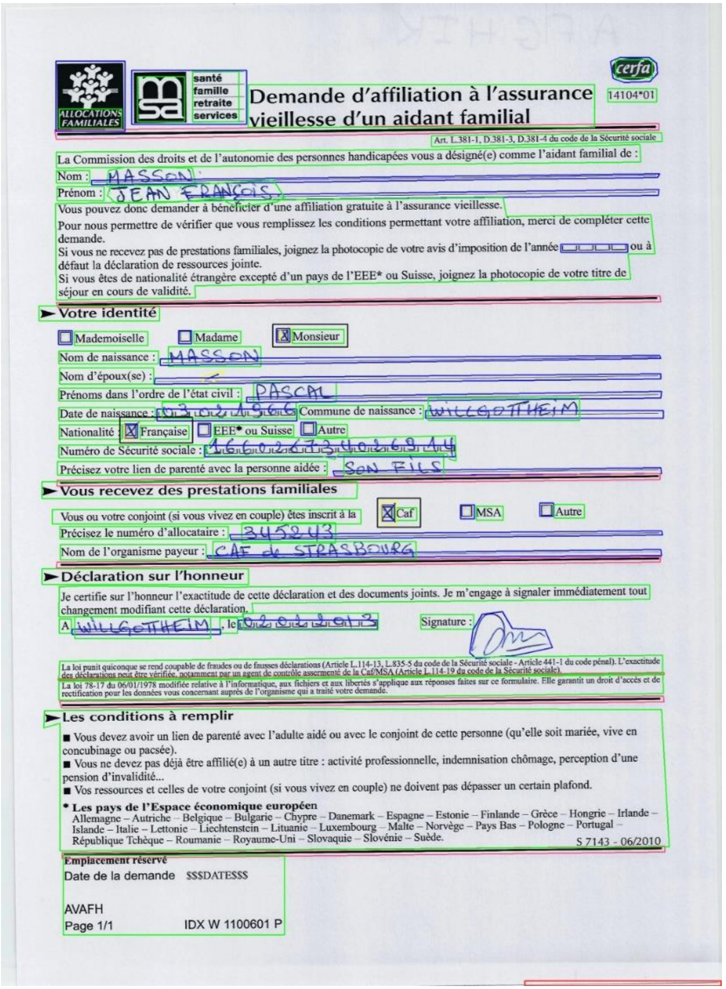
- The Maurdor Database
  - Multi-lingual, mixed printed and handwritten

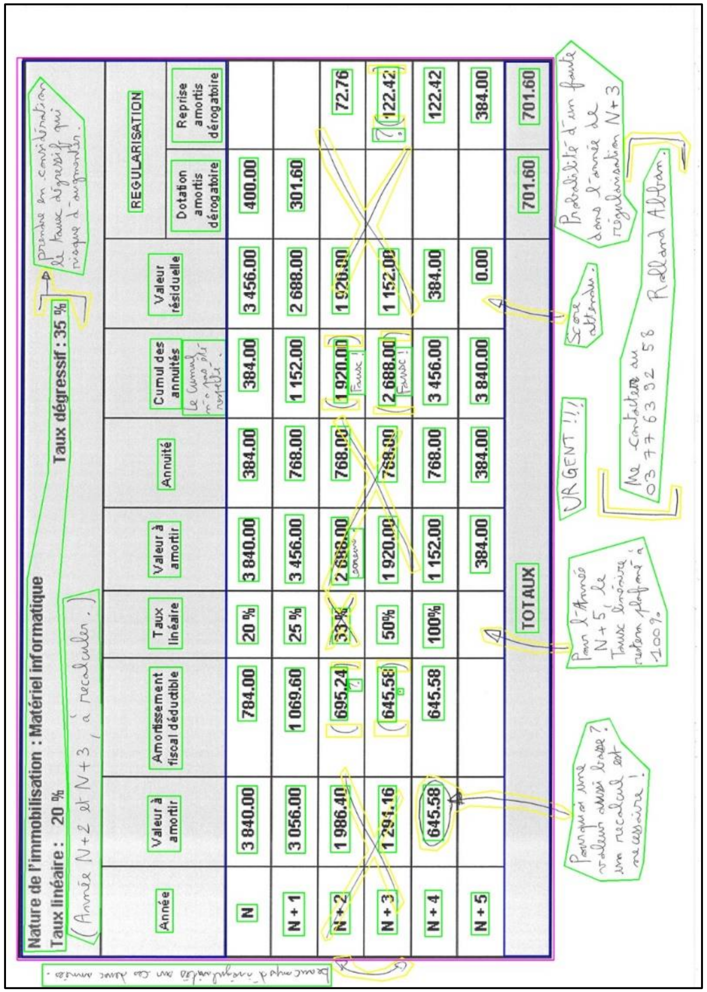
<table>
<tr><th rowspan="3">Set</th><th rowspan="3">Pages</th><th colspan="6">Zones</th></tr>
<tr><th colspan="3">Printed zones</th><th colspan="3">Handwritten zones</th></tr>
<tr><th>French</th><th>English</th><th>Arabic</th><th>French</th><th>English</th><th>Arabic</th></tr>
<tr><td rowspan="3">Train2</td><td rowspan="3">6 592</td><td colspan="6">141 683</td></tr>
<tr><td colspan="3">105 002</td><td colspan="3">36 681</td></tr>
<tr><td>57 821</td><td>25 773</td><td>21 263</td><td>18 417</td><td>8 530</td><td>9 729</td></tr>
<tr><td rowspan="3">Dev2/Test1</td><td rowspan="3">1 110</td><td colspan="6">25 663</td></tr>
<tr><td colspan="3">19 205</td><td colspan="3">6 458</td></tr>
<tr><td>9 908</td><td>5 124</td><td>4 122</td><td>2 857</td><td>1 765</td><td>1 835</td></tr>
<tr><td rowspan="3">Test2</td><td rowspan="3">1 072</td><td colspan="6">25 180</td></tr>
<tr><td colspan="3">18 907</td><td colspan="3">6 273</td></tr>
<tr><td>11 519</td><td>4 131</td><td>3 210</td><td>3 241</td><td>1 450</td><td>1 582</td></tr>
<tr><td rowspan="3">Total</td><td rowspan="3">8 774</td><td colspan="6">192 526</td></tr>
<tr><td colspan="3">143 114</td><td colspan="3">49 412</td></tr>
<tr><td>79 248</td><td>35 028</td><td>28 595</td><td>24 515</td><td>11 745</td><td>13 146</td></tr>
</table>

B. Moysset, et al. The A2iA multi-lingual text recognition system at the second Maurdor evaluation, ICFHR 2014.

- # The Maurdor Database

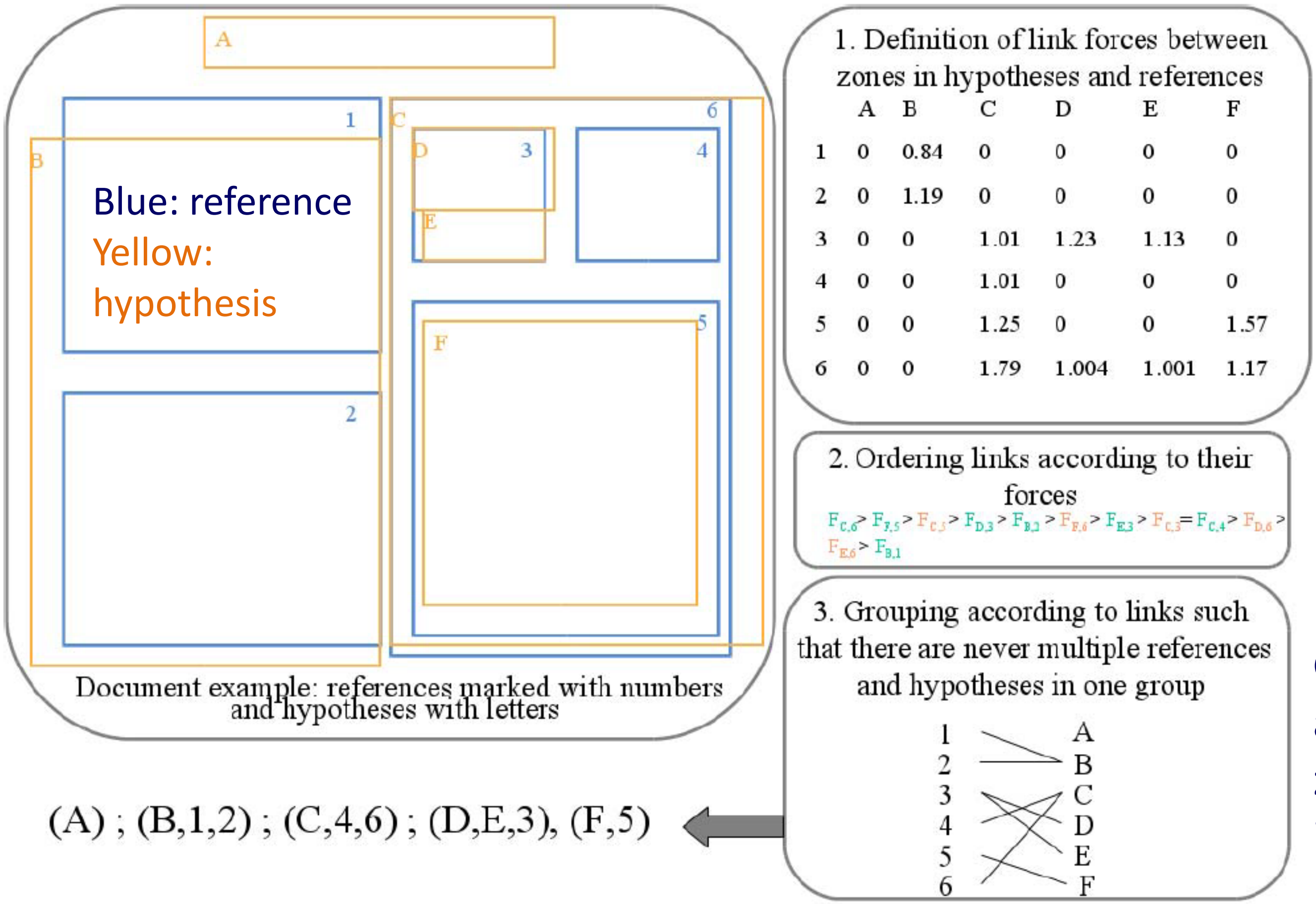
Samples with ground truth





- # The Maurdor Database:

Evaluation metric: ZoneMap

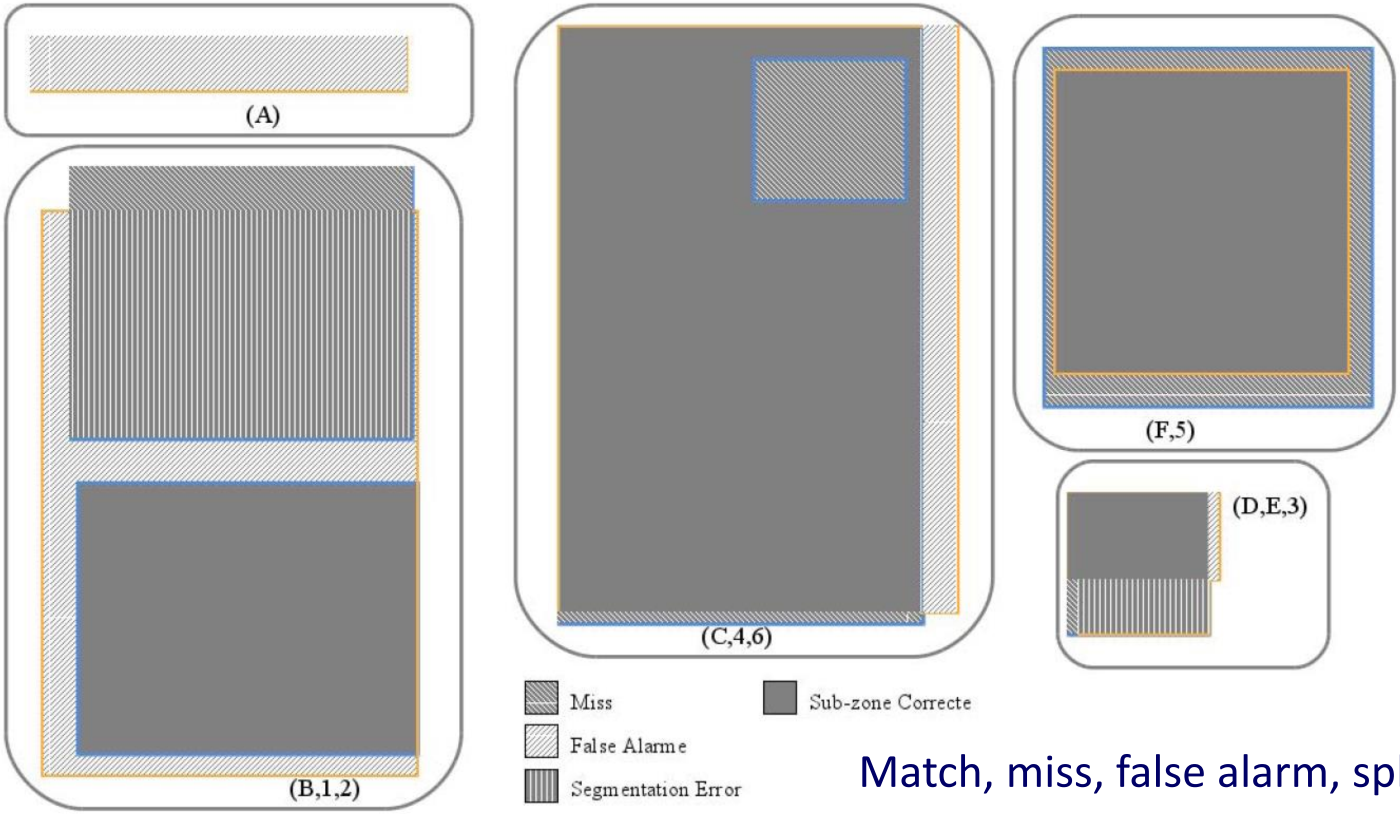


Group reference and hypothesis zones incrementally

Mapping process for calculating ZoneMap

O. Galibert, J. Kahn, I. Oparin, The ZoneMap metric for page segmentation and area classification in scanned documents, *Proc. ICIP 2014*

Evaluation metric: ZoneMap



Match, miss, false alarm, split

$$E_{ZoneMap} = \frac{\sum_{i=1}^{N} E_i}{Area(R)}$$

For each group $E = (1 - \alpha_c)E_S + \alpha_c E_C$

E_S: zone segmentation error; E_C: zone classification error

α_C: weight

Performance: Some Representative Results

Handwritten Text Line Segmentation Using FCN with Dilated Convolutions (G. Renton et al., IJDAR 2018)

Comparison of dilated FCN compared to the main submitted systems on the cBAD database (ICDAR2017 competition)

Method	Precision	Recall	F-measure
DMRZ	97.3	97.0	97.1
This work (11 layers)	94.9	88.1	91.3
This work (7 layers)	89.7	89.9	89.8
UPVLC	93.7	85.5	89.4
BYU	87.8	90.7	89.2
IRISA	88.3	87.7	88.0

*DMHZ: GmbH, Vienna, Austria. Convolutional U-net.
This work: light attention of post-processing

Multi-Task Layout Analysis Using FCN (Y. Xu, et al, 2018)

Category-average metric (%) of page segmentation on DIVA-HisDB Dataset
(Proposed*: without combining low-level feature)

	IU_{mean}	$F1_{mean}$	P_{mean}	R_{mean}
Proposed	**95.47**	**97.52**	**99.00**	96.52
Proposed*	94.67	96.55	97.93	96.53
Rank1	94.90	96.81	97.58	**97.20**
Rank2	93.95	96.04	96.55	97.10

Text line segmentation. Line-level IoU (LIU). PIU considers all pixels, MPIU only takes the pixels within matched lines. CSG0018/CSG0863/CB0055: three types of manuscripts.

	CSG0018			CSG0863			CB0055			Total		
	LIU	PIU	MPIU	LIU	PIU	MPIU	LIU	PIU	MPIU	LIU	PIU	MPIU
Proposed	**99.01**	**98.97**	**99.24**	**99.83**	**98.74**	**98.80**	**99.38**	97.84	97.89	**99.41**	**98.51**	**98.64**
Proposed*	98.32	98.12	98.29	95.30	98.12	98.54	99.28	**98.22**	**98.54**	97.63	97.97	98.17
Rank1	94.90	94.47	96.24	96.75	90.81	92.29	99.33	93.75	94.02	96.99	93.01	94.18
Rank2	69.57	75.31	92.28	90.64	93.68	96.07	84.29	80.23	88.82	81.50	83.07	91.27

Baseline detection (%)

	CSG0018	CSG0863	CB0055	Total
Proposed	**99.48**	**99.89**	**99.36**	**99.57**
Proposed*	98.79	99.51	98.51	98.94
Rank1	98.53	97.16	98.96	98.22
Rank2	98.79	98.30	95.97	97.68

Page Segmentation Using Label Pyramid Network (X.-H. Li, et al., 2019)

Table II: Baseline detection on Bozen and cBAD-TrackB.

Method	Bozen			cBAD-TrackB		
	P	R	F	P	R	F
DMRZ [19]	–	–	–	0.8540	0.8630	0.8590
Multi-Task [9]	0.9580	0.9910	0.9740	0.8480	0.8540	0.8510
dhSegment [10]	–	–	–	0.8260	0.9240	0.8720
ARU-Net [11]	0.9765	0.9734	0.9750	**0.9260**	0.9180	**0.9220**
Proposed	**0.9948**	**0.9986**	**0.9967**	0.8864	**0.9509**	0.9176

Table III: Document region segmentation results on Maurdor.

Method	ZoneMap			Jaccard
	$\alpha_c = 0.0$	$\alpha_c = 0.5$	$\alpha_c = 1.0$	
S1	90.0	107.1	124.1	0.150
S2	60.1	75.9	91.8	0.315
S3	31.2	57.3	83.4	0.190
S5	52.2	62.4	72.7	0.287
FCN	22.90	29.61	36.32	**0.8656**
LPN	**17.81**	**23.57**	**29.32**	0.8647

S1, S2, S3, S5: Four previous systems

Text/non-text classification and printed/handwritten separation using CRF (X.-H. Li, et al., 2018)

Table II: CC-level text/non-text classification results on TestPaper1.0 dataset

Method	Text			Graphics			Table			GP
	P	R	F-m	P	R	F-m	p	R	F-m	
MLP	99.96	99.97	99.96	91.53	87.66	89.55	91.91	93.28	92.59	99.91
CRF_MLP	**99.96**	**99.99**	**99.98**	**98.59**	**90.91**	**94.59**	**96.15**	**93.28**	**94.70**	**99.95**
CNN	99.96	99.92	99.94	81.41	89.61	85.31	93.99	93.28	93.63	99.87
CRF_CNN	99.95	99.98	99.97	94.48	88.96	91.64	95.38	92.54	93.94	99.93

Table V: Region-level writing type separation results on Maurdor dataset

System	Printed			Handwritten			GP	SR
	P	R	F-m	P	R	F-m		
Maurdor2013-S2	92.43	95.61	93.99	83.07	73.33	77.90	90.55	6.56
Maurdor2013-S5	93.96	92.59	93.27	78.88	82.30	80.56	90.00	0.02
Maurdor2014-S2	94.93	96.23	95.57	88.10	84.46	86.24	93.30	0.15
Maurdor2014-S5	96.92	98.09	97.50	93.18	89.35	91.23	96.11	11.12
CRF_CNN_Vote*	98.18	97.24	97.71	91.84	94.52	93.16	96.57	0
CRF_CNN_Vote	98.18	97.26	97.72	91.89	94.51	93.18	96.58	0.02
CRF_CNN_Vote	98.18	97.35	97.76	92.13	95.50	93.30	96.65	0.15
CRF_CNN_Vote	98.61	98.89	98.75	**96.25**	95.35	95.80	98.07	6.56
CRF_CNN_Vote	**98.63**	**98.87**	**98.75**	96.22	**95.42**	**95.82**	**98.08**	11.12

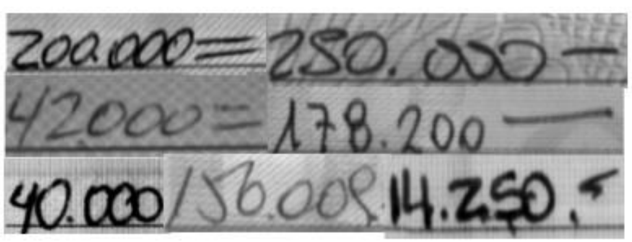
Handwritten Numeral Recognition

- Isolated: MNIST

Liu et al. 2003	Error%	#param	Time
MLP	0.60	63.31K	0.44ms
Polynomial	0.58	38.86K	0.76ms
SVC-poly	0.55	913K	5.90ms
SVC-rbf	0.42	1.61M	21.9ms

DNN	Error (%)
Simard et al. 2003	0.40
Ciresan et al. 2010 (IDSIA)	0.35
Wu et al. 2014 (Fujitsu)	0.254

- Numeral strings
  - ICFHR2014 HDSRC

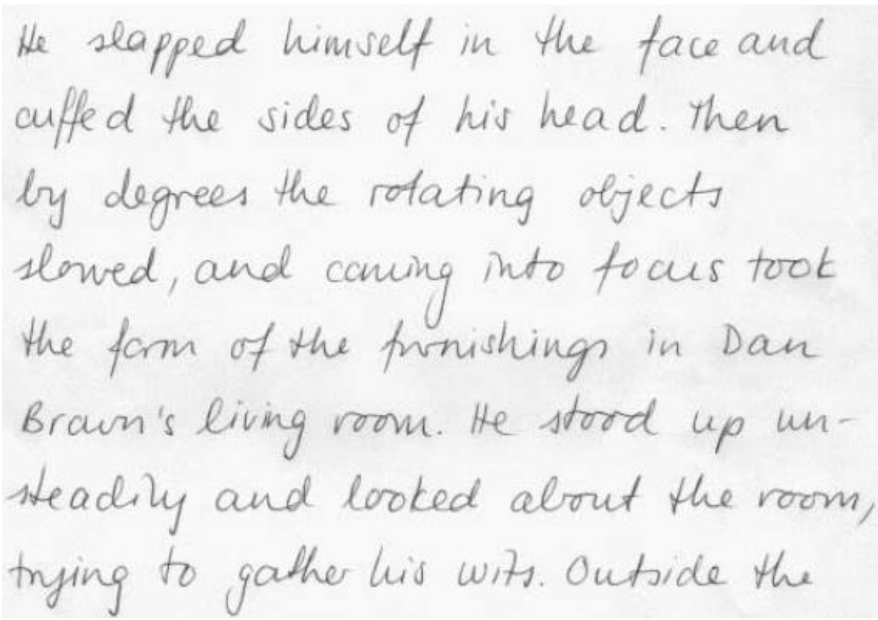


Submission	Guesses	CAR A	CAR B	CVL	Mean
Tébessa I	TOP-1	0.3705	0.2662	0.5930	0.4099
	TOP-2	0.4559	0.3401	0.6575	0.4845
	TOP-3	0.4720	0.3568	0.6690	0.4993
Tébessa II	TOP-1	0.3972	0.2772	0.6123	0.4289
	TOP-2	0.4477	0.3137	0.6527	0.4714
	TOP-3	0.4818	0.3411	0.6824	0.5018
Singapore	TOP-1	0.5230	0.5960	0.5040	0.5410
	TOP-2	0.6180	0.6770	0.6060	0.6337
	TOP-3	0.6540	0.7130	0.6540	0.6737
Pernambuco	TOP-1	0.7830	0.7543	0.5860	0.7078
	TOP-2	0.8916	0.8746	0.6850	0.8171
	TOP-3	0.9199	0.9009	0.7234	0.8481
Beijing	TOP-1	0.8073	0.7013	0.8529	0.7872
	TOP-2	0.8634	0.7638	0.9128	0.8467
	[illegible]	[illegible]	[illegible]	[illegible]	.8555
					.4217
					.4633
					.4849

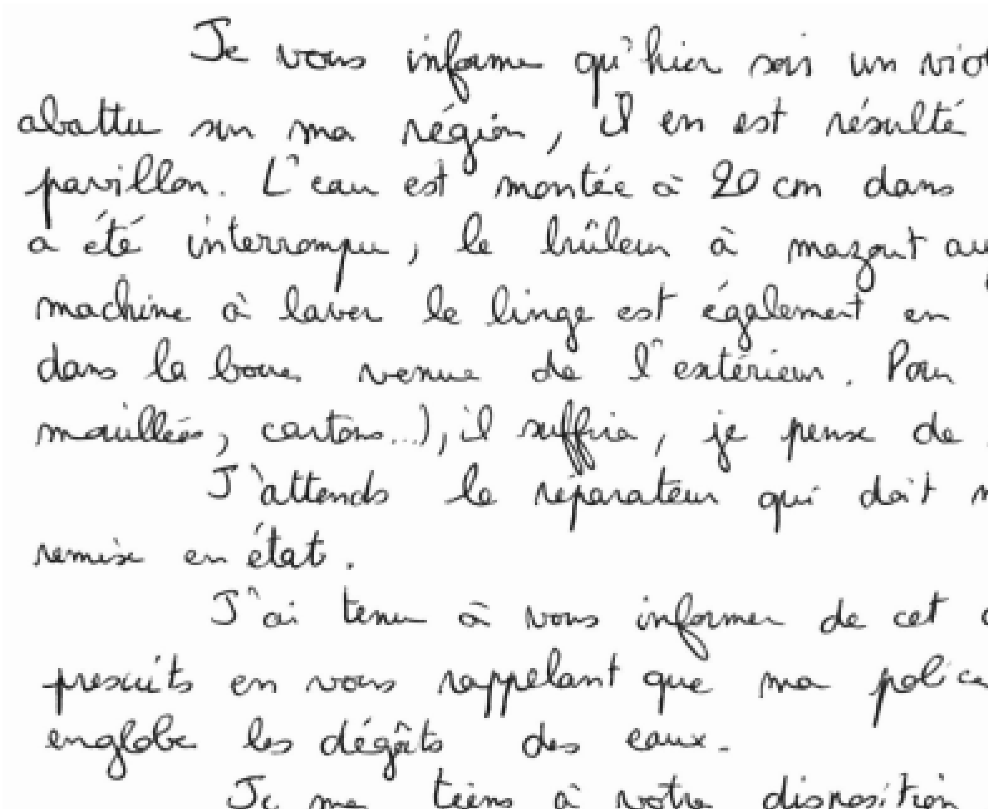
Not latest, should be solved well given enough training data.
Many string (text line) recognition methods available.

Handwriting Recognition

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



He slapped himself in the face and
cuffed the sides of his head. Then
by degrees the rotating objects
slowed, and coming into focus took
the form of the furnishings in Dan
Brown's living room. He stood up un-
steadily and looked about the room,
trying to gather his wits. Outside the



Je vous informe qu'hier soir un
abattu sur ma région, il en est résulté
pavillon. L'eau est montée à 20 cm dans
a été interrompu, le brûleur à mazout
machine à laver le linge est également en
dans la boue venue de l'extérieur. Pour
mouillés, cartons...), il suffira, je pense de
J'attends le réparateur qui doit
remise en état.
J'ai tenu à vous informer de cet
produits en vous rappelant que ma
englobe les dégâts des eaux.
Je me tiens à votre disposition

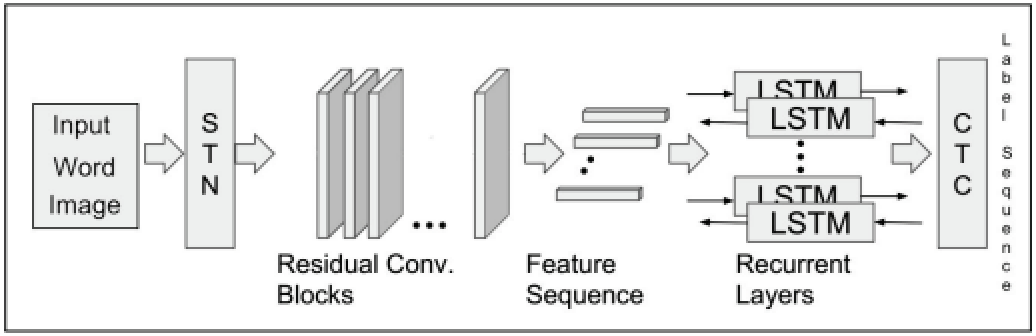
Results on IAM Dataset

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

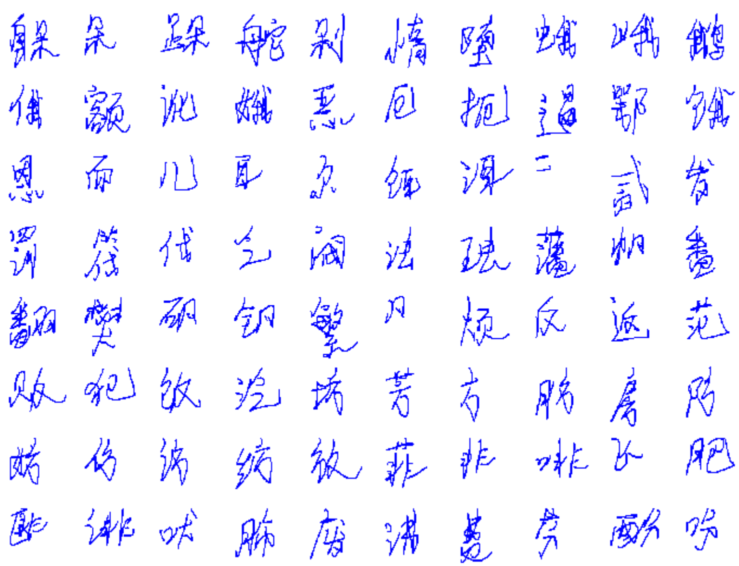
Improvements: Pre-processing, pre-training, data augmentation

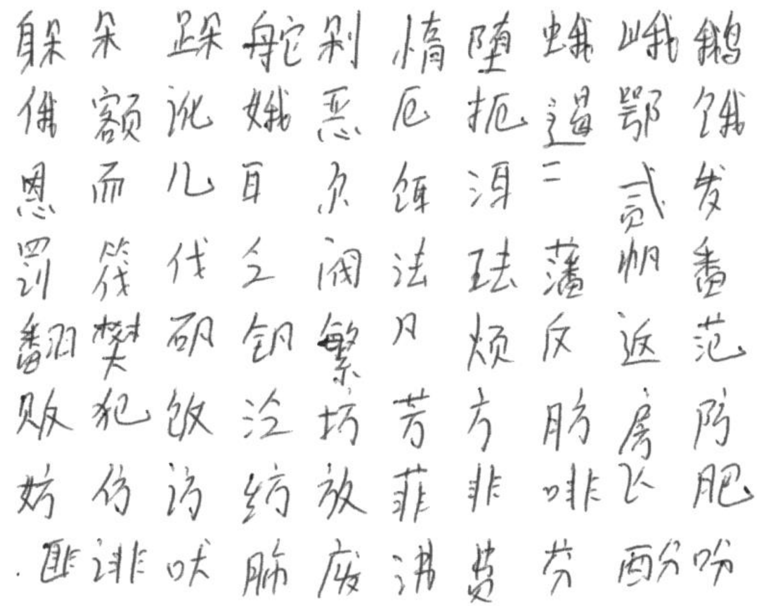


K. Dutta, P. Krishnan, M. Mathew, C.V. Jawahar, Improving CNN-RNN hybrid networks for handwriting recognition, ICFHR 2018.

Chinese Characters and Text

- 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

- # Handwritten Chinese Characters
  - ## 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			

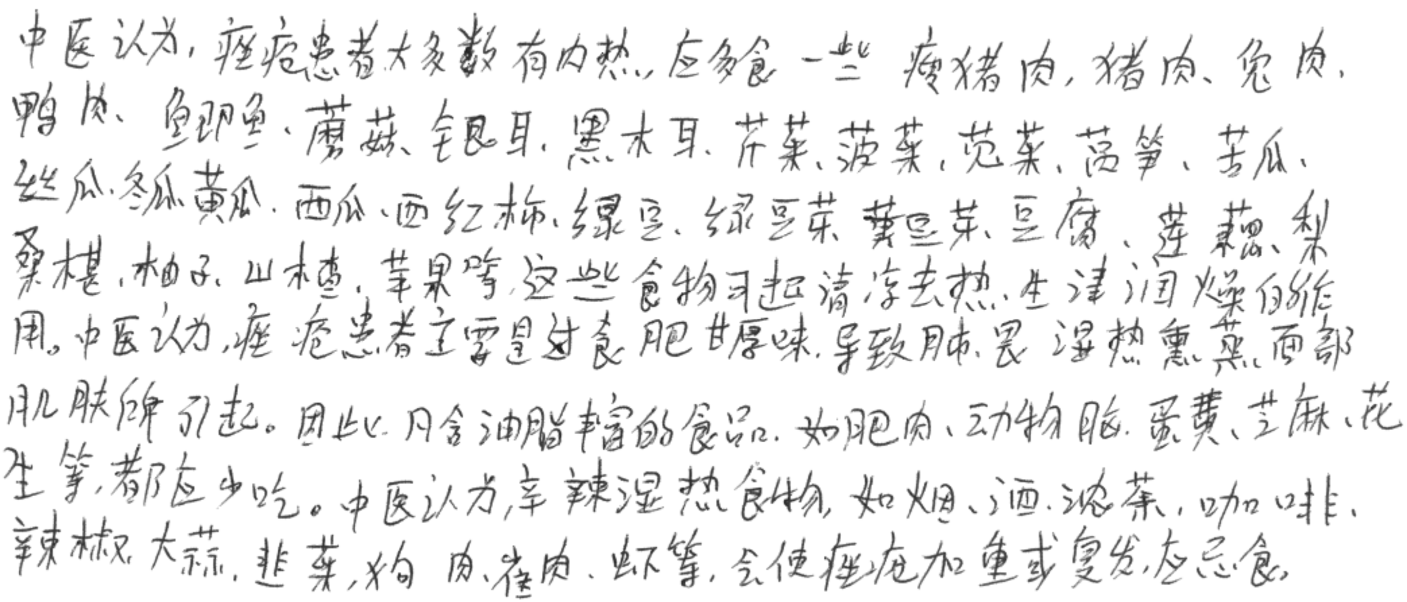
CASIA (PR2017)	dirMap+CNN	97.55	295ms	23.5M
	Ensemble-3	97.64		
SCUT (PRL2017)	CNN+DD+PS	97.55	295ms	23.5M
	Model acerage	97.64		
CASIA (PAMI2017)	RNN	97.89		10.38M
	Ensemble-6	**98.15**		78.11M

Offline character recognition

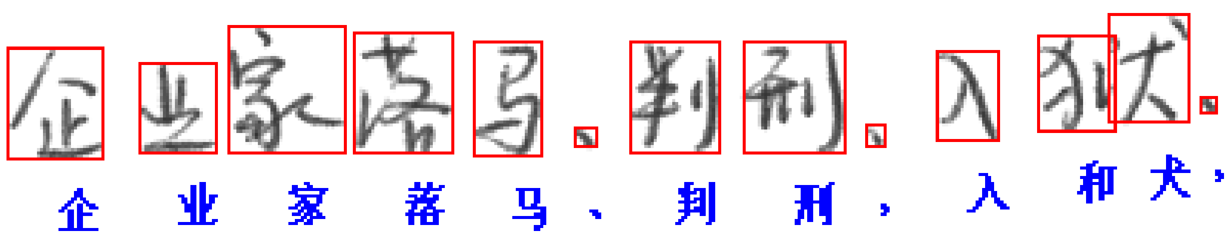
	System	CR (%)	Speed (ms)
ICDAR2013 Competition	Fujitsu, CNN	94.77	55 (GPU)
	IDSIAnn (8)	94.42	315 (CPU)
	IDSIAnn-1	94.24	197 (CPU)
	HIT	92.62	4.6 (CPU)
	Human	96.13	
IDSIA Tech Rep 05-13	CNN	94.47	3.03 (GPU)
	Multi-CNN (8)	95.78	22.04 (GPU)
Fujitsu (ICFHR2014)	ATR-CNN	95.04	
	CNN voting	96.06	
CASIA (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.

- Handwritten Chinese Texts
  - 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	

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

Scene Text Detection

- Datasets
  - ICDAR 2013: horizontal
  - ICDAR 2015: incidental
  - MSRA-TD500: multi-oriented
  - MLT-17: 9 languages, 18,000 images
  - RCTW-17: Chinese text, 8346 train, 4229 test
  - CASIA-10k: Chinese text, 7000 train, 3000 test
  - COCO-Text: loosely annotated, not used widely
- Performance metrics
  - Word/line level Recall, Prevision, F-value

Scene Text Image Datasets

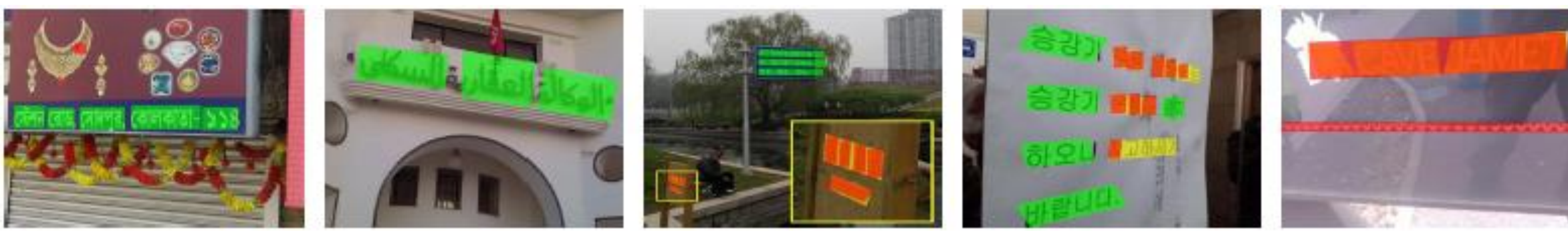
ICDAR2013 Focused Images

ICDAR2015 Incidental Images

COCO-Text Dataset

Fig. 1: Example images and annotations of the CTW-12k dataset.

RCTW-17



MLT-17

Multi-Oriented Text Detection Results

Results on ICDAR 2015

Algorithm	Precision	Recall	F-measure
Proposed (VGG-16)	0.85	**0.80**	**0.82**
Proposed (S-VGG)	0.84	0.79	0.81
Zhou *et al.* [50]	0.83	0.78	0.81
Proposed (ResNet-50)	**0.89**	0.73	0.80
Shi *et al.* [33]	0.73	0.77	0.75
Liu *et al.* [24]	0.73	0.68	0.71
Tian *et al.* [39]	0.74	0.52	0.61
Zhang *et al.* [48]	0.71	0.43	0.54
StradVision2 [18]	0.77	0.37	0.50
StradVision1 [18]	0.53	0.46	0.50
NJU-Text [18]	0.70	0.36	0.47
AJOU [18]	0.47	0.47	0.47
HUST_MCLAB [18]	0.44	0.38	0.41

Results on MSRA-TD500

Algorithm	Precision	Recall	F-measure
Proposed*	**0.91**	**0.81**	**0.86**
Shi *et al.* [33]	0.86	0.70	0.77
Proposed	0.85	0.70	0.76
Zhou *et al.* [50]	0.87	0.67	0.76
He *et al.* [12]	0.77	0.70	0.74
Zhang *et al.* [48]	0.83	0.67	0.74
Yin *et al.* [45]	0.81	0.63	0.71
Kang *et al.* [17]	0.71	0.62	0.66
Yao *et al.* [42]	0.63	0.63	0.60

Results on ICDAR 2013

Algorithm	Precision	Recall	F-measure
Proposed	**0.95**	**0.89**	**0.91**
He *et al.* [12]	0.92	0.81	0.86
Shi *et al.* [33]	0.88	0.83	0.85
Liao *et al.* [21]	0.88	0.83	0.85
Zhang *et al.* [48]	0.88	0.78	0.83
He *et al.* [11]	0.93	0.73	0.82
Tian *et al.* [38]	0.85	0.76	0.80

W. He, X.-Y. Zhang, F. Yin, C.-L. Liu, Multi-oriented and multi-lingual scene text detection with direct regression, IEEE T-IP, 2018

Results on CTW-17

Team Name	Precision	Recall	F-measure
Foo & Bar	0.7439	**0.5948**	**0.6611**
NLPR_PAL (Proposed)	0.7717	0.5729	0.6576
gmh	0.7064	0.5784	0.6360
SCUT_MBCNN	0.7361	0.5184	0.6084
IVA	0.6610	0.5522	0.6017
CCFLAB	0.7406	0.4713	0.5760
CAS_HotEye	**0.7915**	0.4417	0.5670
Baseline [33]	0.7603	0.4044	0.5278
XMU_SuperLab	0.7222	0.4133	0.5258
Image Search Team	0.6544	0.3996	0.4962
SCUT_DLVC	0.7058	0.3656	0.4817

Results on MLT-17

Method	Precision	Recall	F-measure
Proposed (Validation)	0.8266	0.7253	0.7726
Proposed (Test)	0.7669	0.5794	**0.6601**
SCUT_DLVClab	**0.8028**	0.5454	0.6496
Sensetime_OCR	0.5693	**0.6943**	0.6256
SARI_FDU_RRPN_v1	0.7117	0.5550	0.6237
TH-DL	0.6775	0.3478	0.4597
linkage-ER-Flow	0.4448	0.2559	0.3249
IDST_CV	0.3181	0.2602	0.2863

Results on CASIA-10K

Algorithm	Precision	Recall	F-measure
Proposed	**0.8128**	**0.7048**	**0.7550**
SegLink	0.7275	0.6967	0.7118
EAST	0.7771	0.5327	0.6321

Arbitrary Shape Text Detection

Method	Recall	Precision	Hmean
SegLink [24]	40.0	42.3	40.8
EAST [34]	49.1	78.7	60.4
DMPNet [16]	56.0	69.9	62.2
CTD [17]	65.2	74.3	69.5
CTD+TLOC [17]	69.8	77.4	73.4
TextSnake [19]	**85.3**	67.9	75.6
Proposed	80.2	**80.1**	**80.1**

Table 4. Results on CTW1500.

Method	Recall	Precision	Hmean
SegLink [24]	23.8	30.3	26.7
EAST [34]	36.2	50.0	42.0
DeconvNet [2]	44.0	33.0	36.0
Mask Textspotter [20]	55.0	69.0	61.3
TextSnake [19]	74.5	**82.7**	78.4
Proposed	**76.2**	80.9	**78.5**

Table 5. Results on TotalText.

CTW 1500

Total-Text

X. Wang, et al., CVPR 2019

ICDAR2017 Competition on Reading Chinese Text in the Wild (RCTW-17)

Task 1 - Text Localization Leaderboard

F-measure Rank	Team Name	Team Member	F-meausre	Precision	Recall	Institute
1	Foo & Bar	Zheqi He,Yongtao Wang	0.661054	0.743876	0.594827	Peking University
2	NLPR_PAL	Wenhao He, Fei Yin, Da-Han Wang, Cheng-Lin Liu	0.657598	0.771675	0.572905	NLPR,CASIA
3	gmh	Minghao Guo	0.636024	0.706367	0.578422	Tsinghua University
4	SCUT_MBCNN	Jinrong Li, Zijian Zhou, Shuangping Huang	0.608396	0.736135	0.518434	South China University of Technology

Task 2 - End-to-End Recognition Leaderboard

AED-Rank	Team Name	Team Member	Average Edit Distance	Institute
1	NLPR_PAL	Yan-Fei Lv, Wenhao He, Fei Yin, Cheng-Lin Liu	20.21967368	NLPR,CASIA
2	SCUT_DLVC	Lianwen Jin, Yuliang Liu, Zenghui Sun, Canjie Luo, Zhaohai Li, Lele Xie, Fan Yang	28.3078742	South China University of Technology
3	CCFLAB	Dai Yuchen, Huang Zheng, Gao Yuting	32.129818	Shanghai Jiao Tong University

http://mclab.eic.hust.edu.cn/icdar2017chinese/?from=timeline&isappinstalled=0

Scene Text Recognition

Free lexicon

Curved text

Methods	ConvNet, Data	IIIT5k			SVT		IC03			IC13	IC15	SVTP	CUTE
		50	1k	0	50	0	50	Full	0	0	0	0	0
Wang et al. [60]	-	-	-	-	57.0	-	76.0	62.0	-	-	-	-	-
Mishra et al. [44]	-	64.1	57.5	-	73.2	-	81.8	67.8	-	-	-	-	-
Wang et al. [62]	-	-	-	-	70.0	-	90.0	84.0	-	-	-	-	-
Bissacco et al. [7]	-	-	-	-	-	-	90.4	78.0	-	87.6	-	-	-
Almazan et al. [2]	-	91.2	82.1	-	89.2	-	-	-	-	-	-	-	-
Yao et al. [67]	-	80.2	69.3	-	75.9	-	88.5	80.3	-	-	-	-	-
Rodríguez-Serrano et al. [52]	-	76.1	57.4	-	70.0	-	-	-	-	-	-	-	-
Jaderberg et al. [29]	-	-	-	-	86.1	-	96.2	91.5	-	-	-	-	-
Su and Lu [56]	-	-	-	-	83.0	-	92.0	82.0	-	-	-	-	-
Gordo [16]	-	93.3	86.6	-	91.8	-	-	-	-	-	-	-	-
Jaderberg et al. [26]	VGG, 90k	97.1	92.7	-	95.4	80.7	98.7	98.6	93.1	90.8	-	-	-
Jaderberg et al. [25]	VGG, 90k	95.5	89.6	-	93.2	71.7	97.8	97.0	89.6	81.8	-	-	-
Shi et al. [54]	VGG, 90k	97.8	95.0	81.2	97.5	82.7	98.7	98.0	91.9	89.6	-	-	-
*Shi et al. [55]	VGG, 90k	96.2	93.8	81.9	95.5	81.9	98.3	96.2	90.1	88.6	-	71.8	59.2
Lee et al. [36]	VGG, 90k	96.8	94.4	78.4	96.3	80.7	97.9	97.0	88.7	90.0	-	-	-
Yang et al. [64]	VGG, Private	97.8	96.1	-	95.2	-	97.7	-	-	-	-	75.8	69.3
Cheng et al. [11]	ResNet, 90k+ST$^+$	99.3	97.5	87.4	97.1	85.9	99.2	97.3	94.2	93.3	70.6	-	-
ASTER-A	VGG, 90k	98.1	95.7	81.7	97.6	85.5	98.7	97.3	92.2	88.6	67.6	73.2	63.9
ASTER-B	ResNet, 90k	98.7	96.3	83.2	99.2	87.6	99.1	97.6	92.4	89.7	68.9	75.4	67.4
ASTER	ResNet, 90k+ST	99.6	98.8	93.4	99.2	93.6	98.8	98.0	94.5	91.8	76.1	78.5	79.5

B. Shi, et al., T-PAMI 2019.

End-to-End Scene Text Recognition

Table 2. Results on CTW1500 test set.

Method	Detection			End-to-End	
	P	R	F	None	Full
SegLink [37]	42.3	40.0	40.8	-	-
EAST [49]	78.7	49.1	60.4	-	-
DMPNet [30]	69.9	56.0	62.2	-	-
FOTS [29]	79.5	52.0	62.8	21.1	39.7
CTD [31]	74.3	65.2	69.5	-	-
CTD+TLOC [31]	77.4	69.8	73.4	-	-
TextSnake [32]	67.9	**85.3**	75.6	-	-
Our Two-Stage	79.5	81.0	80.2	37.2	69.9
With RoIRotate	80.7	83.4	82.3	38.6	70.9
With LSTM	84.3	81.8	83.0	39.2	71.5
TextDragon	**84.5**	82.8	**83.6**	**39.7**	**72.4**

Table 3. Results on Total-Text test set.

Method	Detection			End-to-End	
	P	R	F	None	Full
SegLink [37]	30.3	23.8	26.7	-	-
Ch'ng *et al.* [6]	40.0	33.0	36.0	-	-
EAST [49]	50.0	36.2	42.0	-	-
FOTS [29]	52.3	38.0	44.0	32.2	35.9
Liao *et al.* [27]	62.1	45.5	52.5	36.3	48.9
Mask TextSpotter [33]	69.0	55.0	61.3	**52.9**	71.8
TextSnake [32]	82.7	74.5	78.4	-	-
Our Two-Stage	84.5	74.2	79.0	46.1	70.6
With RoIRotate	**86.0**	74.3	79.7	47.1	73.6
With LSTM	85.2	75.7	80.2	48.3	74.7
TextDragon	85.6	**75.7**	**80.3**	48.8	**74.8**

W. Feng, et al., ICCV 2019.

End-to-end recognition also benefits detection.

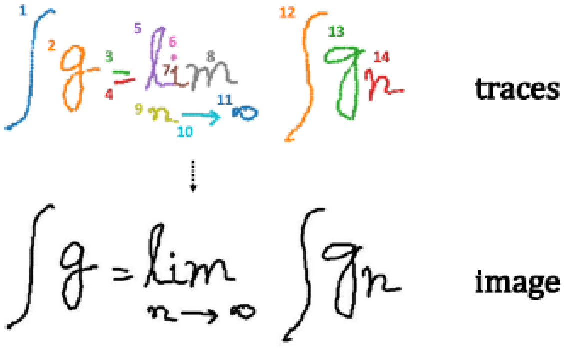
End-to-end recognition results on ICDAR 2015 Test set.

Method	Detection			Method	End-to-End			Word Spotting		
	P	R	F		S	W	G	S	W	G
SegLink [37]	74.74	76.50	75.61	Baseline OpenCV3.0 [23]	13.84	12.01	8.01	14.65	12.63	8.43
EAST [49]	83.27	78.33	80.72	Stradvision [23]	43.7	-	-	45.9	-	-
He *et al.* [15]	82.0	80.0	81.0	TextProposals [8, 18]	53.3	49.6	47.2	56.0	52.3	49.7
TextSnake* [32]	84.9	80.4	82.6	HUST_MCLAB [37, 38]	67.9	-	-	70.6	-	-
PixelLink [7]	85.5	82.0	83.7	Deep text spotter [3]	54.0	51.0	47.0	58.0	53.0	51.0
Mask TextSpotter* [33]	91.6	81.0	86.0	Mask TextSpotter* [33]	79.3	73.0	62.4	79.3	74.5	64.2
He *et al.* [14]	87.0	86.0	87.0	He *et al.* [14]	82.0	77.0	63.0	85.0	80.0	65.0
FOTS [29]	91.85	**87.92**	**89.84**	FOTS [29]	**83.55**	79.11	65.33	**87.01**	**82.39**	67.97
Our Detection	84.82	81.82	83.05	Our Two-Stage	75.23	73.15	53.04	77.03	75.11	54.51
With RoIRotate	92.18	82.93	87.31	With RoIRotate	82.51	**79.21**	**65.37**	86.20	82.03	**68.14**
TextDragon	**92.45**	83.75	87.88	TextDragon	82.54	78.34	65.15	86.22	81.62	68.03

S/W/G: Strong/Weak/Generic lexicon.

Mathematics Recognition

- Datasets
  - Printed expressions
    - IM2Latex-100K (Y. Deng et al., ICML 2017)
  - CROHME (Competition on Recognition of Online Handwritten Mathematical Expression) 2013, 2014, 2016
    - CROHME 2014: train set of 8836 math expressions (86K symbols), test set of 986 math expressions (6K symbols).
    - CROHME 2013 test set for validation set
    - CHROME 2016: test set of 1147 expressions, training set same as 2014



TAP: J. Zhang, et al., 2019
WAP: J. Zhang, et al., 2017

PAL: J. Wu, et al., ECML 2018

Results on CROHME 2014 test set.

System	Correct(%)	$\leq$ 1(%)	$\leq$ 2(%)	$\leq$ 3(%)
I	37.22	44.22	47.26	50.20
II	15.01	22.31	26.57	27.69
IV	18.97	28.19	32.35	33.37
V	18.97	26.37	30.83	32.96
VI	25.66	33.16	35.90	37.32
VII	26.06	33.87	38.54	39.96
Ours	**61.16**	**75.46**	**77.69**	**78.19**
WAP (offline)	46.55	61.16	65.21	66.13
PAL	39.66	56.80	65.11	70.49
PAL*	47.06	63.49	72.31	78.60

Results on CROHME 2016 test set.

	Correct(%)	$\leq$ 1(%)	$\leq$ 2(%)	$\leq$ 3(%)
Wiris	49.61	60.42	64.69	–
Tokyo	43.94	50.91	53.70	–
São Paolo	33.39	43.50	49.17	–
Nantes	13.34	21.02	28.33	–
Ours	**57.02**	**72.28**	**75.59**	**76.19**
WAP (offline)	44.55	57.10	61.55	62.34

Discussions

- Main Approaches
  - Layout Analysis
    - Top-down vs bottom-up, structured prediction (graphical model), FCN
  - Text recognition
    - Isolated: deep classifier (CNN)
    - Text line: over-segmentation based, HMM, RNN, conv-RNN
      - Character model based: gives clear character segmentation, applicable to large category set and multi-language
  - Scene text detection and recognition
    - Detection: deep learning (CNN), multi-task (pixel classification, boundary regression), with or w/o proposal generation
    - Recognition: conv-RNN becomes popular, special methods proposed for curved text recognition
    - End-to-end: joint model with shared feature extraction (CNN)
    - Local region (character) prediction better satisfy curved text
  - Mathematic expressions recognition
    - Attentional network (encoder-decoder) performs well, but weak interpretation (e.g., symbol segmentation)

Discussions

- State of the Art
  - Big progresses in multiple tasks of DAR benefitted from deep learning
  - Outperform human (really?) when training with big data
- Remaining Problems
  - Layout analysis: complex layout and background, divergent format
  - Text recognition: divergent styles, style drift, multi-language, small sample
  - Learning from mixed and weakly labeled data, continuous learning
  - Structural and semantic understanding
  - Real applications: multi-type, hybrid contents, degraded image

Future Directions

- Fundamental Theory & Methodology
  - Learning for small/imbalanced data, weakly labeled data
  - Fusion of multi-level/source contexts, global optimization
  - Online learning/adaptation
  - Cognitive mechanisms: from visual cues to high-level knowledge
- Document Image Processing
  - Image capturing paradigm/device
  - Complex layout analysis, logical structure
  - Joint layout analysis and text recognition
- Character Recognition
  - Structural analysis
  - One-shot learning, zero-shot learning

- Text Line Recognition
  - Sequence classification model and learning
  - Contexts modeling, fusion, and adaptation
  - Multi-lingual documents, especially mixed languages
  - Retrieval and semantic analysis
  - End-to-end page recognition without line segmentation
- Application Oriented
  - Modeling of interactive transcription
  - Confidence and reliability (to reject)
  - Objects beyond texts
    - Mathematics: offline printed, online handwritten
    - Table, symbols, diagrams, stamp, signature
  - Document authentication, writer identification
  - Character interpretation and verification
  - New applications
    - Human interface, robot, archeology, education, travel aid, impaired person assistance, etc.

Thank You for Your Attention!