Deterministic Routing between Layout Abstractions for Multi-Scale Classification of Visually Rich Documents

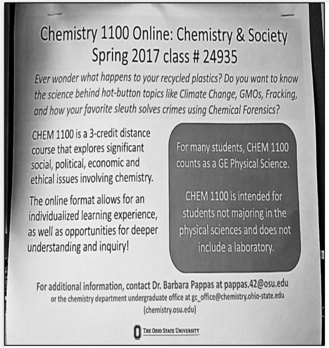
Ritesh Sarkhel*, Arnab Nandi
Department of Computer Science and Engineering

Outline

- Visually Rich Documents
- Challenges
- Our contributions
  - Spatial Pyramid Model
  - Deterministic Routing Scheme
- End-to-end architecture for classification
- Experimental results
  - Tobacco-Litigation dataset
  - Outperforms previous state-of-the-art by 4.73%

Visually Rich Document

- Visual cues to highlight distinct semantic entities
    - Color
    - Font size
    - Text orientation and positioning
    - Symmetry etc.
- Classification is a precursor of many document understanding tasks
    - Indexing
    - Information extraction



Chemistry 1100 Online: Chemistry & Society
Spring 2017 class # 24935

Ever wonder what happens to your recycled plastics? Do you want to know the science behind hot-button topics like Climate Change, GMOs, Fracking, and how your favorite sleuth solves crimes using Chemical Forensics?

CHEM 1100 is a 3-credit distance course that explores significant social, political, economic and ethical issues involving chemistry.

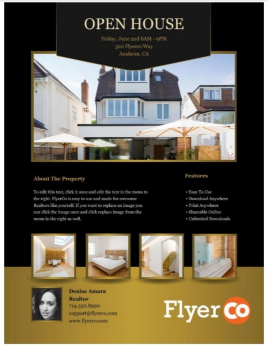
The online format allows for an individualized learning experience, as well as opportunities for deeper understanding and inquiry!

For many students, CHEM 1100 counts as a GE Physical Science.

CHEM 1100 is intended for students not majoring in the physical sciences and does not include a laboratory.

For additional information, contact Dr. Barbara Pappas at pappas.42@osu.edu or the chemistry department undergraduate office at gc_office@chemistry.ohio-state.edu (chemistry.osu.edu)

The Ohio State University



OPEN HOUSE

Friday, June 2nd 8AM - 9PM
320 Flyerco Way
Anaheim, CA

About The Property

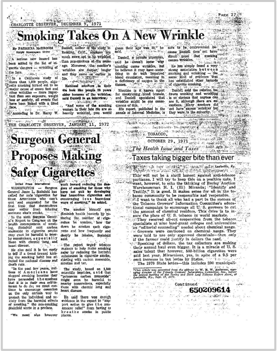
To edit this text, click it once and edit the text in the menu to the right. FlyerCo is easy to use and made for awesome Realtors like yourself. If you want to replace an image you can click the image once and click replace image from the menu to the right as well.

Features

- Easy To Use
- Download Anywhere
- Print Anywhere
- Shareable Online
- Unlimited Downloads

Denise Amara
Realtor
714.350.8290
support@flyerco.com
www.flyerco.com

FlyerCo



Page 27

CHARLOTTE OBSERVER, DECEMBER 9, 1971

Smoking Takes On A New Wrinkle

By PATRICIA McBROOM
Knight Newspapers Writer

A serious new hazard has been added to the list of cigarete dangers — wrinkling of the face.

In a California study of more than 1,000 people, cigarette smoking turned out to be major cause of crows feet and other wrinkles — more important than sun exposure, weight loss or anxiety, all events that have been linked with a lined face.

According to Dr. Harry W. Daniell, author of the study in Redding, Calif., smokers are much more apt to be wrinkled than non-smokers of the same age. Moreover, the smoker's wrinkles are deeper, longer and they come on earlier in life.

Habitual smokers in their 40s look like people 20 years older because of the wrinkles, said Daniell in an interview.

"And some of the smoking women in their late 30s are so heavily wrinkled, you would guess their age was 55," he said.

Daniell, a private physician, said he doesn't know why smoking cause wrinkles, but he believes it may have something to do with impaired blood circulation, resulting in a deficiency of oxygen in the tissues.

Nicotine is a known agent for constricting blood bessels, and Daniell speculated that wrinkles might be one consequence of this.

His report, published in the annals of Internal Medicine, is sure to be controversial because Daniell does not have direct proof that smoking causes wrinkles.

He has simply found a very strong association between smoking and wrinkling — the same kind of evidence that has established other hazards of cigarette smoking.

Daniell said the relation between smoking and wrinkling is so obvious that anyone can see it, although there are exceptions. Many smokers did not have excess wrinkles but they were in the minority.

THE CHARLOTTE OBSERVER, JANUARY 11, 1972

Surgeon General Proposes Making Safer Cigarettes

United Press International

WASHINGTON — Surgeon General Jesse L. Steinfeld has called for safe cigarettes for those Americans who can't quit and suggested for the first time that nonsmokers may suffer from inhaling someone else's smoke.

In the sixth Surgeon General's Report to Congress on the health consequences of smoking, Steinfeld said carbon monoxide in cigarette smoke may even be harmful to nearby nonsmokers, especially those with chronic lung and heart disease.

He also said it is too early to tell if a trend toward kicking the smoking habit has affected the national disease and death rate.

"In the past few years, millions of Americans have stopped smoking because they have persuaded themselves that it is in their own self-interest to do so; we must continue to encourage cessation as the only certain way to protect the individual and society from the harmful effects of smoking," the non-smoking Steinfeld wrote in a preface.

"We must also however work toward reducing the dangers of smoking for those who have not quit by developing less hazardous cigarettes and encouraging less hazardous ways of smoking," he added.

The smoker himself can diminish health hazards by reducing the number of cigarettes he smokes, how far down he smokes each cigarette and how frequently and deeply he inhales, Steinfeld said.

The report urged tobacco makers to help make smoking safer by reducing the harmful substances in cigarette smoke, starting with carbon monoxide, nicotine and tar.

The study, based on 1,500 scientific inquiries, said that "poisonous carbon monoxide" might even be harmful to nearby nonsmokers, especially those with chronic lung and heart disease.

He said there was enough evidence in the report to "support action to give the nonsmoker relief" from having to breathe smoke in public places.

TOBACCO,
OCTOBER 29, 1971

The Health Issue and Taxes

Taxes taking bigger bite than ever

Louisville, Ky.

This will not be a shrill lament against anti-tobacco partisans. I will try to keep this on a positive note. I want, however, to echo the thinking of Burley Auction Warehouseman H. I. (Hi) Miranda; "Identify and Testify." It is good. It makes sense for all in the tobacco community to be responsible and responsive.

I want to thank all who had a part in the success of the Tobacco Growers' Information Committee's educational campaign to encourage all U. S. growers to cut the amount of chemical residues. This drive is to insure the place of U. S. tobacco in world markets.

They received all-out cooperation from the tobacco specialists at nine land-grant colleges and universities on "editorial counseling" needed about chemical usage.

Growers were cautioned on chemical usage. They were told to use only approved chemicals—then only if the farmer could justify in dollars the need.

Speaking of dollars, the tax collectors are making their annual haul even bigger. It is a miracle of U. S. sales talent that however, 538-billion cigarettes were sold last year. Miraculous, yes, in spite of a 9.5 per cent increase in tax levies by States.

The 1970 State levies—this includes 200 municipali-

*This article was excerpted from the address by W. H. W. Anderson, managing director of the Tobacco Growers' Information Committee, Inc., before the annual meeting of the Burley and Dark Leaf Tobacco Export Assn., at Louisville, Ky., Sept. 27, 1971.

Continued

650209614

Our objective

1. Maximize classification accuracy using visual (dis)similarity in layouts
   - Multi-scale features
     - Encode local invariant patterns from multiple resolutions of the document
     - More robust than single-scale features: Leads to better classification accuracy
2. Minimize end-to-end inference turnaround
   - Suitable for modern interactive workflows

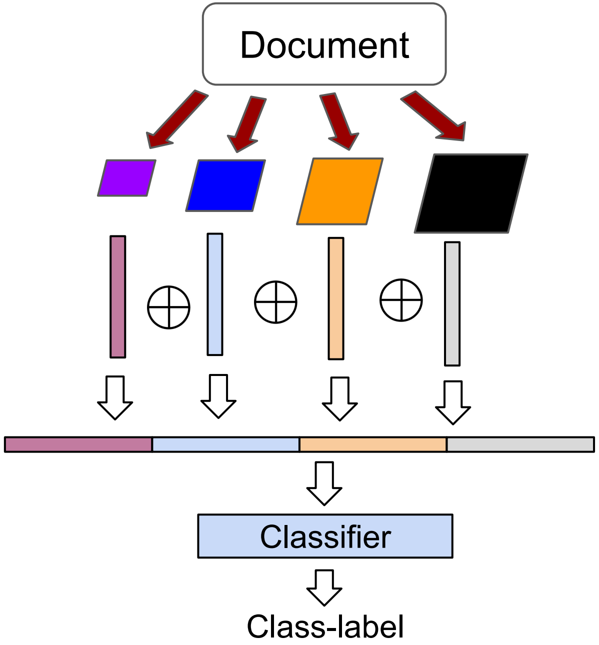
Assumption: A soft constraint on the relative positioning of components in documents belonging to the same class

Outline

- Visually Rich Documents
- Challenges
- Our contributions
  - Spatial Pyramid Model
  - Deterministic Routing Scheme
- End-to-end architecture for classification
- Experimental results
  - Tobacco-Litigation dataset
  - Outperforms previous state-of-the-art by 4.73%

Challenges

- Multi-scale: Aggregation based methods
  - Increased overhead in inference turnaround
  - Marginal gain in accuracy from single-scale counterparts
- Infeasible when bound by near real-time latency
  - e.g. Interactive workflow on edge-devices

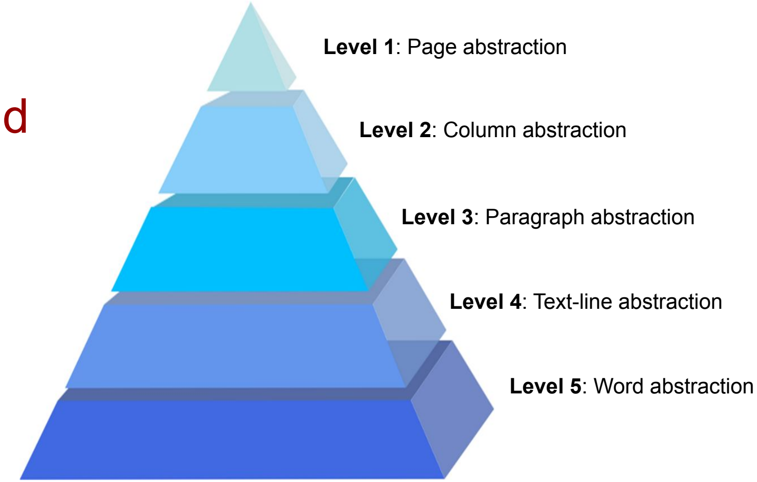


Outline

- Visually Rich Documents
- Challenges and observations
- Our contributions
    - Spatial Pyramid Model
    - Deterministic Routing Scheme
- Neural architecture for classification
- Experimental results
    - Tobacco-Litigation dataset
    - Outperforms previous state-of-the-art by 4.73%

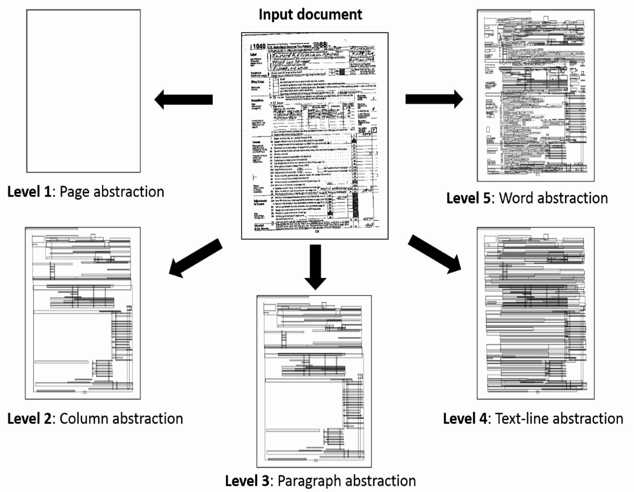
A Spatial Pyramid Model for document representation

- Objective: Increase accuracy by extracting highly discriminative features
- Represent the document as a spatial pyramid
  - Each level corresponds to an abstraction in layout hierarchy:
    - Words
    - Text-line
    - Paragraph
    - Column
    - Page


Level 1: Page abstraction
Level 2: Column abstraction
Level 3: Paragraph abstraction
Level 4: Text-line abstraction
Level 5: Word abstraction

Layout abstraction at each level of the Spatial Pyramid Model

- Each pyramid level corresponds to an abstraction at that level of the layout hierarchy
  - Binary Image with same dimensions as the input document
  - e.g. Word abstraction
    - Bounding box corresponding to each word of the document
- Layout hierarchy derived using an open-source page segmentation algorithm [1,2]


Input document
Level 1: Page abstraction
Level 2: Column abstraction
Level 3: Paragraph abstraction
Level 4: Text-line abstraction
Level 5: Word abstraction

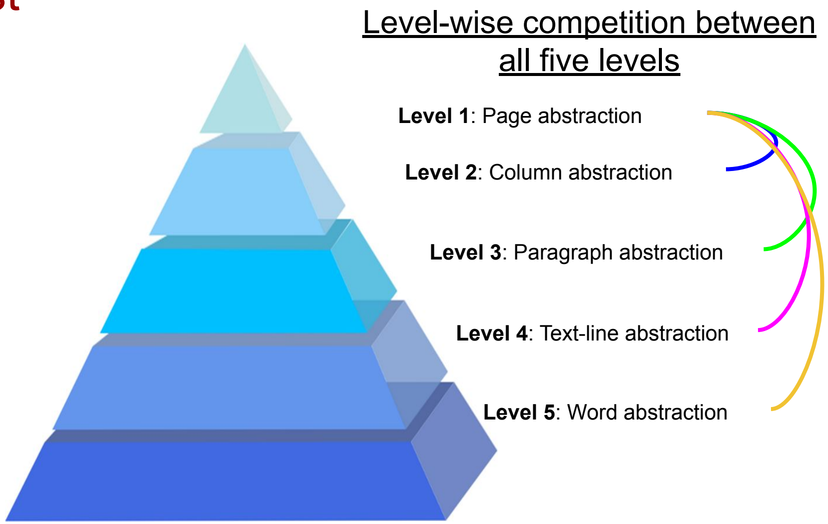
[1] https://github.com/tesseract-ocr/tesseract
[2] Smith, Ray. "An overview of the Tesseract OCR engine." In *Ninth International Conference on Document Analysis and Recognition (ICDAR 2007)*, vol. 2, pp. 629-633. IEEE, 2007.

Outline

- Visually Rich Documents
- Challenges and observations
- Our contributions
  - Spatial Pyramid Model
  - **Deterministic Routing Scheme**
- Neural architecture for classification
- Experimental results
  - Tobacco-Litigation dataset
  - Outperforms previous state-of-the-art by 4.73%

A Deterministic Routing Scheme for feature selection

- Attention-like operation to select the most discriminative pyramid level
- Introduce level-wise competition
  - Select the most discriminating pyramid level
- The winner represents the document for final prediction



Level-wise competition between all five levels

Level 1: Page abstraction

Level 2: Column abstraction

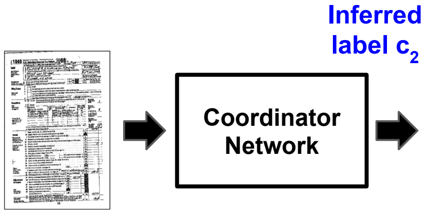
Level 3: Paragraph abstraction

Level 4: Text-line abstraction

Level 5: Word abstraction

Supervised coordinator network for Deterministic Routing

- LadderNet: Softmax classifier network
  - Extended the MobileNetV2 architecture [1,2]
  - Fast: Factorized convolution
  - Pretrained on ImageNet, fine-tuned on training corpus
- Softmax label to query the multiplex table
  - Single entry for each class-label
  - Route to the corresponding pyramid level



Class label	Pyramid level
c_1	l_5
c_2	l_4
….	….
c_N	l_4

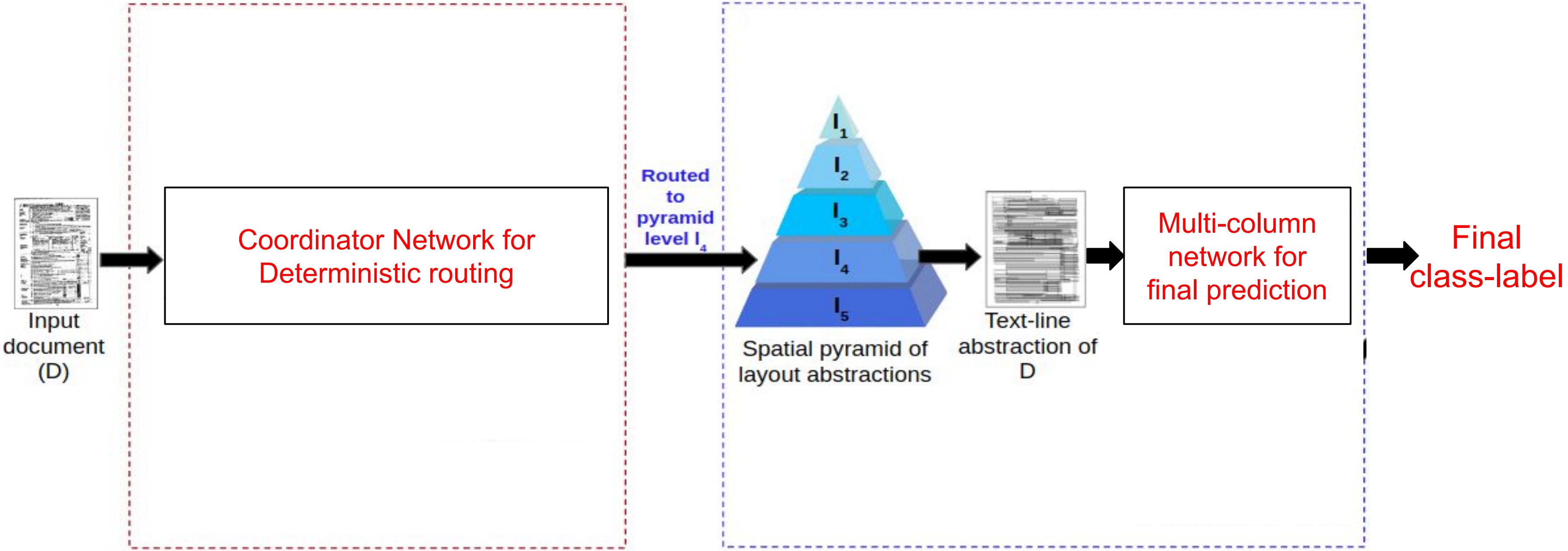
[1] Howard, Andrew G., Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. "Mobilenets: Efficient convolutional neural networks for mobile vision applications." *arXiv preprint arXiv:1704.04861* (2017).

[2] Sandler, Mark, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. "Mobilenetv2: Inverted residuals and linear bottlenecks." In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4510-4520. 2018.

Outline

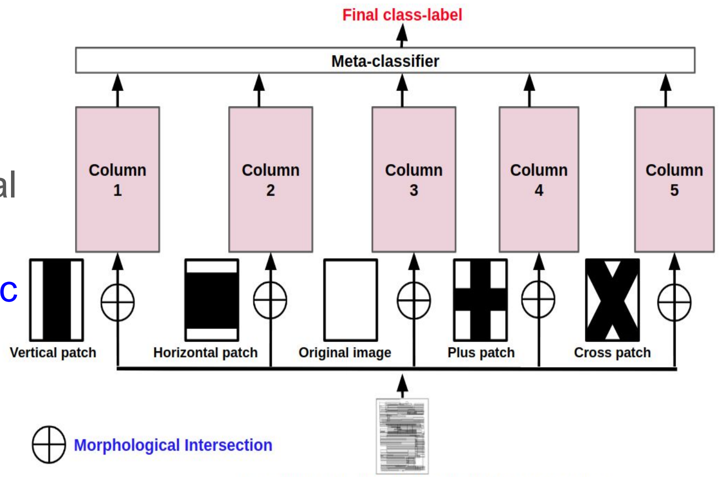
- ❖ Visually Rich Documents
- ❖ Challenges and observations
- ❖ Our contributions
  - ➢ **Spatial Pyramid Model**
  - ➢ **Deterministic Routing Scheme**
- ❖ End-to-end architecture for classification
- ❖ Experimental results
  - ➢ **Tobacco-Litigation dataset**
  - ➢ **Outperforms previous state-of-the-art by 4.73%**

End-to-end classification workflow


Input document (D)
Coordinator Network for Deterministic routing
Routed to pyramid level l_4
l_1
l_2
l_3
l_4
l_5
Spatial pyramid of layout abstractions
Text-line abstraction of D
Multi-column network for final prediction
Final class-label

A Multi-column Network for Final prediction

- Multi-column neural architecture
  - Depth-separable convolutional columns
  - Input: Morphological transformation of the document
  - Combine predictions from each column to compute final class label
    - Meta-classifier: Simple-avg, weighted-avg, logistic regression and MLP



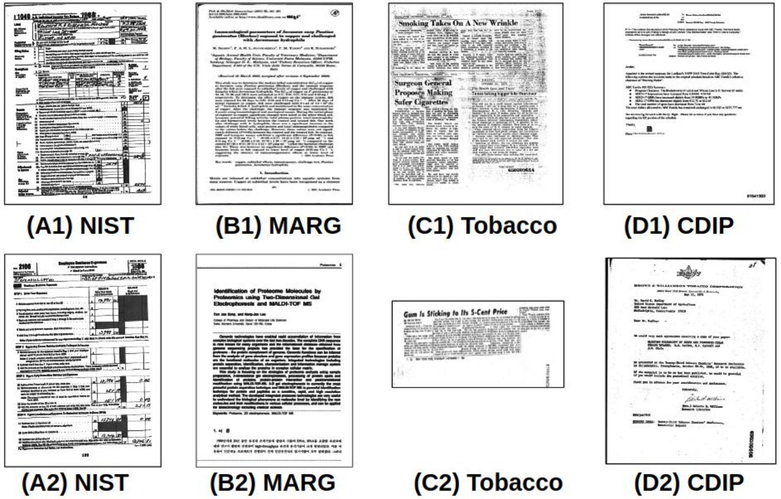
Text-line abstraction of the input document

Outline

- ❖ Visually Rich Documents
- ❖ Challenges and observations
- ❖ Our contributions
    - ➢ Spatial Pyramid Model
    - ➢ Deterministic Routing Scheme
- ❖ End-to-end architecture for classification
- ❖ Experimental results
    - ➢ **Tobacco-Litigation dataset**
    - ➢ **Outperforms previous state-of-the-art by 4.73%**

Experiments

- Four publicly available datasets of single-page documents
  - NIST Special dataset-6 (easy)
  - Medical Article Records Groundtruth or MARG dataset (medium)
  - Tobacco Litigation dataset (hard)
  - IIT-CDIP dataset (hard)
- Evaluation metrics
  - Classification accuracy
  - Inference turnaround time



(A1) NIST (B1) MARG (C1) Tobacco (D1) CDIP

(A2) NIST (B2) MARG (C2) Tobacco (D2) CDIP

Result highlights

- Tobacco Litigation dataset
  - High inter-class and low intra-class layout similarity
  - 3482 documents, 10 document classes (e.g. ‘letter’, ‘memo’, ‘email’ etc.)
- Median classification accuracy over 25 trials := 82.78%
  - Absolute improvement over previous state-of-the-art := 4.73%
- Average inference turnaround time := 362 (±10.27) ms
  - Average speed-up factor > 6

More results and ablation study in paper !!

Takeaways

- Multi-scale classifier for visually rich documents
  - Spatial pyramid model
    - Highly discriminative features → Increase in accuracy
  - Level-wise competition for optimal document representation
    - Attending to the most discriminative layout abstraction improves accuracy
- Inference turnaround time < interactive latency (≅ 500 ms)
- Robust performance on four publicly available datasets

Thank you!