UTS
UNIVERSITY OF TECHNOLOGY SYDNEY

ICFHR 2016: Panel Discussion

15:00-16:00 October 26th, Shenzhen

Panel Members:

Youbin Chen

Gernot Fink

Qiang Huo

Christopher Kermorvant

Lambert Schomaker

Michael Blumenstein (**Chair**)

15th International Conference on Frontiers in Handwriting Recognition

UTS
UNIVERSITY OF TECHNOLOGY SYDNEY

Opening Remarks

- Economic downturn or not – continue your research! (Nakagawa, 2016)
- Handwriting – *still* a natural interface for humans
- But…is handwriting recognition still popular?
  - are there sufficient new applications?
  - or do we need to change research directions?
- Deep Learning…everywhere…but for how long?

technische universität
dortmund

fakultät für
informatik

ICFHR 2016 Panel
New Frontiers in Handwriting Recognition

Gernot A. Fink

TU Dortmund University, Germany

Shenzhen, October 26, 2016

A Unification of Methods?

We have methods/devices that read ...

- Text in the wild
- Online handwriting
- Mathematical formulas
- Historical documents
- ...

When will we see methods that read ANY text?

Limits of Learning by Example?

Learning by example is nice and powerful, but ...

- It needs TONS of labeled data!
- Learned models generalize only MODERATELY beyond the seen examples!

When will we see ROBUST self-learning methods?

When will he be able to create models that generalize from printed to handwritten to artistic writing to ...?

A Word of Warning

Never declare a problem solved ...
(in public / to the media / to the funding organizations)

... when you see nice results on CURRENT benchmarks!

There will always be more challenging tasks ahead that nobody thought about so far!

New Frontiers in Handwriting Recognition

Qiang Huo

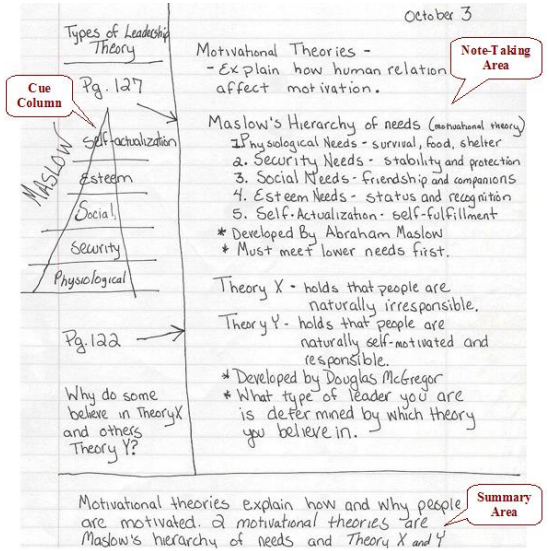
Speech Group

Microsoft Research Asia, Beijing, China

(qianghuo@microsoft.com)

Panel discussion in ICFHR-2016, October 26, 2016

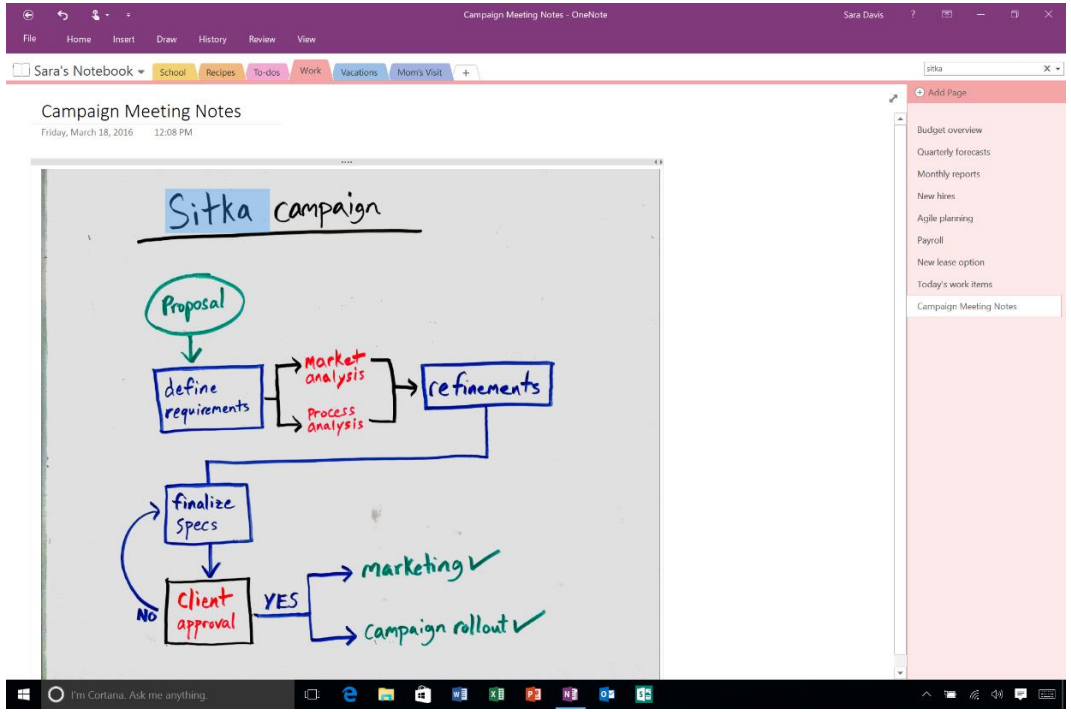
Handwriting OCR for Augmented Intelligence

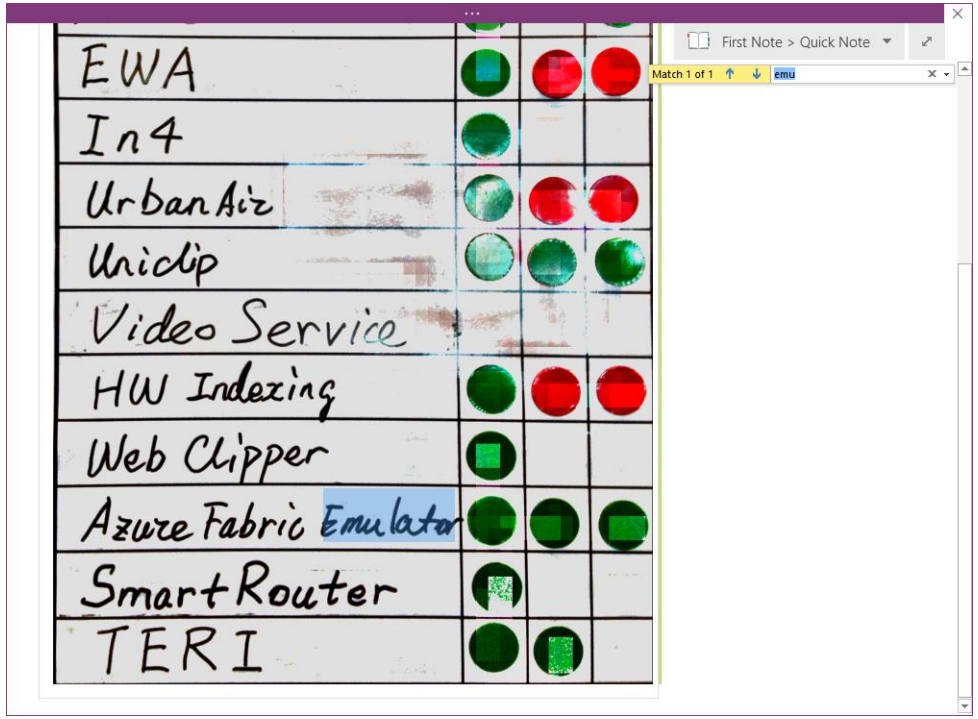


- Index and search for digital memory
- Entity recognition and linking for insights
- Task completion for improved productivity

Search Handwritten Text in Images For OneNote

- Announced publicly in March 2016

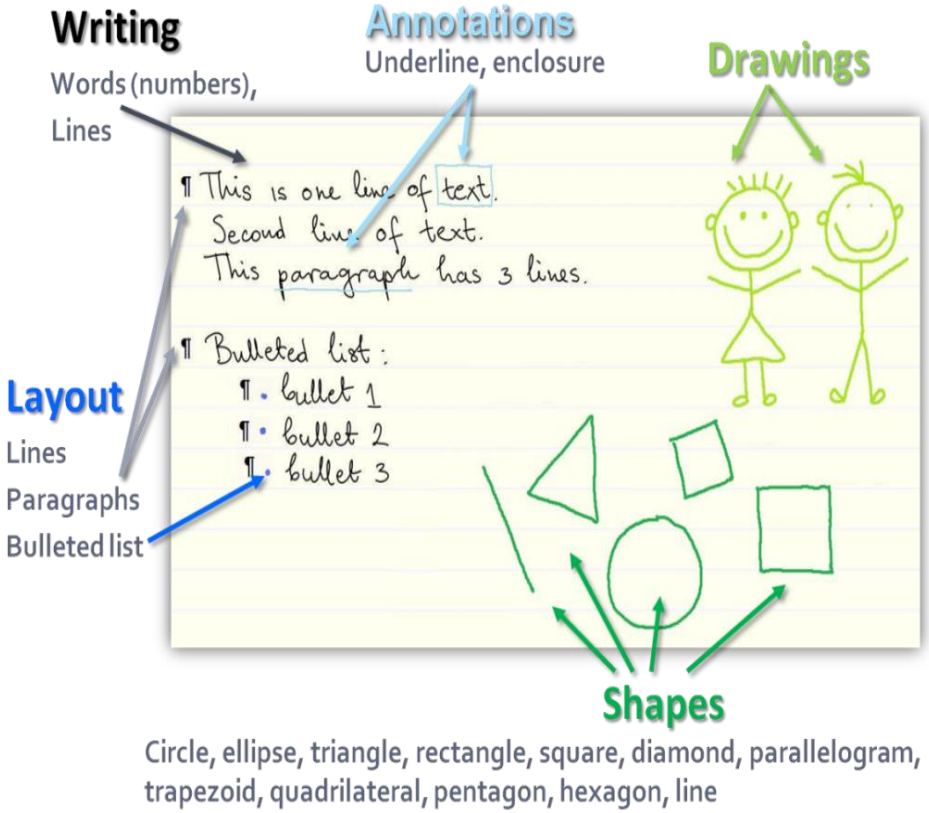




Images added to OneNote using Office Lens, OneNote Clipper, me@onenote.com, etc.

Challenges

- Rectification of distorted image
- Robust text detection
  - Large skew, multi-orientations, curved text lines
  - Long ascenders/descenders, touched text lines
  - Annotations (e.g., underline, enclosure, etc.)
  - Complex layout
- Intelligent layout analysis
  - Text, shapes, math, layout, annotations, unclassified drawings, etc.
  - Language ID of each text line
- Out of vocabulary (OOV) word problem
- Confidence measure
- Universal or customized language model
- Data, Data, Data



Pervasive Pen

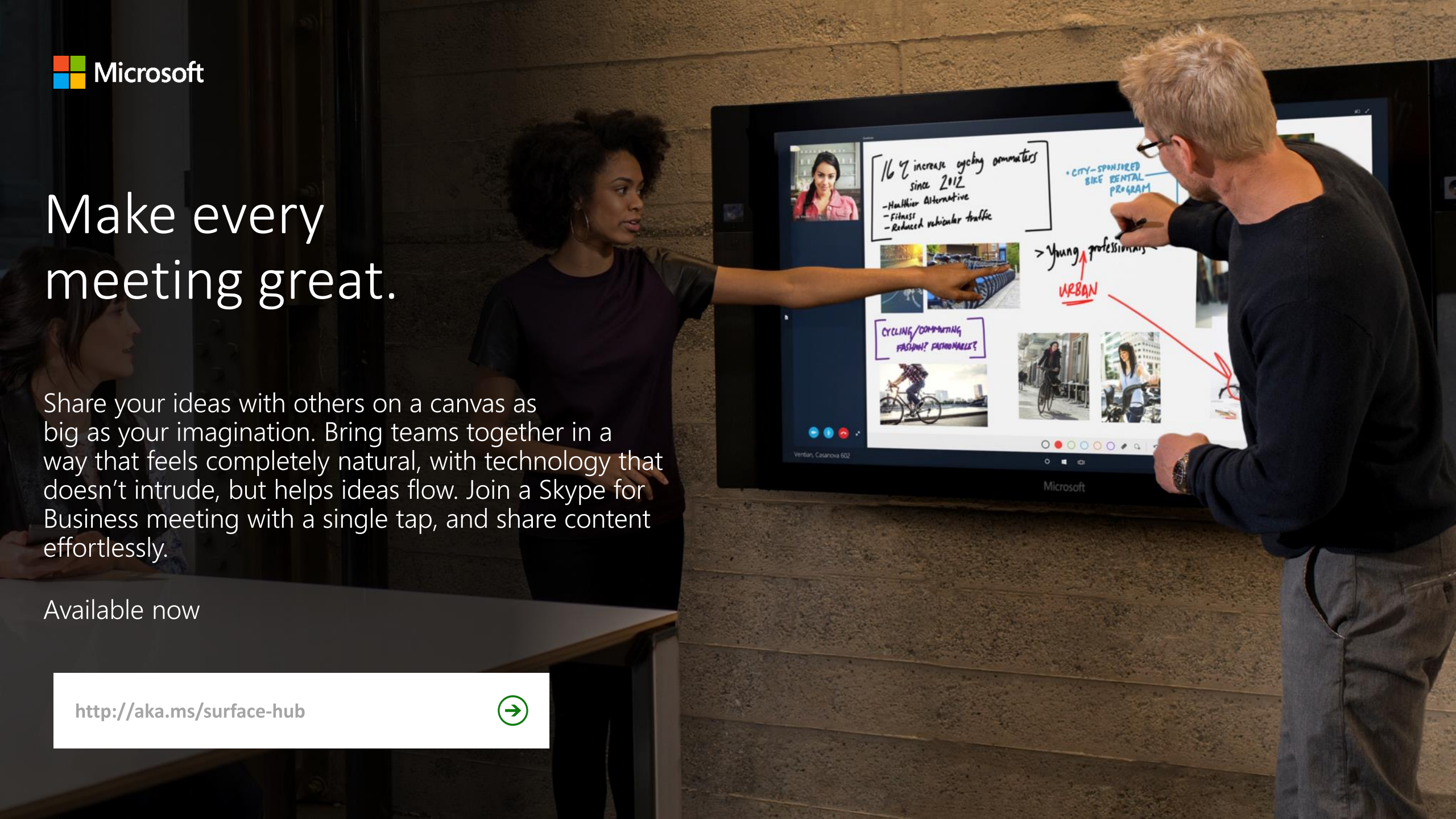
Microsoft

Make every meeting great.

Share your ideas with others on a canvas as big as your imagination. Bring teams together in a way that feels completely natural, with technology that doesn't intrude, but helps ideas flow. Join a Skype for Business meeting with a single tap, and share content effortlessly.

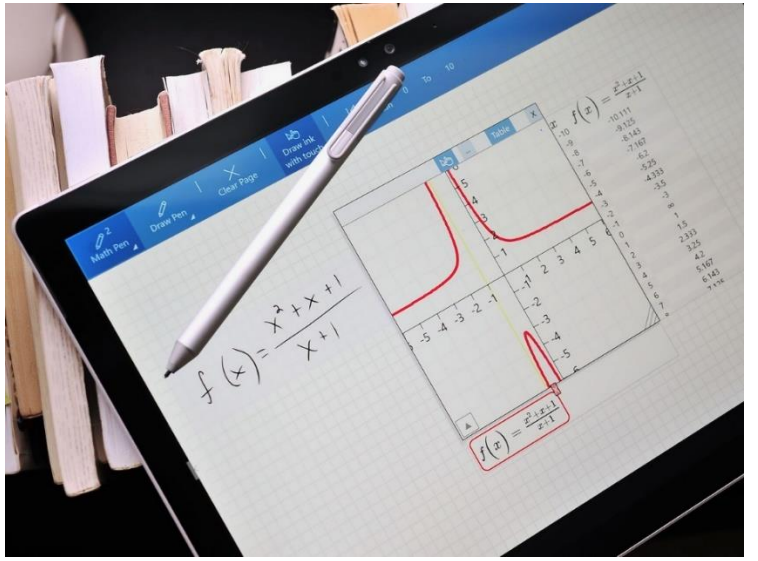
Available now

http://aka.ms/surface-hub



Why Pervasive Pen

- Give users **more reasons to pick up the pen and keep using it**
- **Do more with Windows Ink** than you can with pen and paper
- **Empower users through the magic of SmartInk** recognizing math, text, shapes and more



Need better technologies for ink analysis and recognition to improve user experience!

Windows Ink

ICFHR 2016

PANEL DISCUSSION

Christopher Kermorvant

Teklia
Machine Learning Agency

SEVERAL CONVERGENCES ARE OBSERVED AT ICFHR 2016 :

Convergence on models :

- Deep Learning : 50% of the papers contains « Deep »
- Chinese recognition : from MQDF to CNN/RNN
- Word spotting : features replaced by CNN
- Deep Learning for writer identification
- Handwritten text recognition : READ competition, 100% of competitors use LSTM/RNN

CONVERGENCE TO DEEP LEARNING

Consequence ?

- less diversity in the methods
- better for rapid adoption of proposed improvements

Good, because if the target is **to solve handwriting** recognition, the community is **too small** to work on many different methods

CONVERGENCE ON DATABSES

Convergence on models+databases :

- Word Spotting papers use recognition database
- more papers using several databases on different languages

Soon, direct comparisons of word spotting/recognition approaches

You can not say anymore « Arabic/Chinese/Bangla/… » is difficult/different

CROSS-DOMAIN CONVERGENCE

Convergence Handwriting/Speech recognition:

- Is handwriting harder than speech recognition ?

From the recurrent neural network point of view, they are the same

But still differences regarding the number/size of available databases, that might explain why handwriting recognition is less advanced

ICFHR 2016 Panel Discussion

Lambert Schomaker

Discussion topics

1. Single-trick frozen ponies vs active learning systems?

2. Separate linguistic post-processing pipeline?
   or *end-to-end* training, including semantics?

3. These terrible, handcrafted deep networks

4. If you already assume a Titan GPU, there must be other things to do besides endless training

5. How to keep & attract researchers during the Machine-Learning revolution?

1. Single-trick frozen ponies vs active learning systems?

- Neural networks were once (in the 80'ies) heralded as the replacement of rule-based systems that had to be programmed in detail. The dream was that a computer would adapt itself to a changing and complex world …

- 2016: Deep learning has yielded very high performances, but lab-based training is ever more complicated, even requiring special hardware. It only yields a frozen solution for a particular training set. Performance on unseen data cannot be predicted well and there is no adaptation in the operational stage.

Where are the active-learning systems?

2. Separate linguistic post-processing stage or *End-to-end* training, including semantics?

- In an integrated, multilevel information integration NN, you don't know what causes the current performance. Is it the good visual architecture or the context expectancy? Reuse in an other application would require lengthy retraining, at all levels.

- A separate post-processor is modular and reusable, but does it get all the information from the visual stage that it needs?

3. These terrible, handcrafted deep networks

The derogatory reference to handcrafted features is bit strange for a field that is completely submerged in manual design and fine tuning of complicated network architectures.

- Watch it: you risk the same fate in a few years
- Also: One should be proud of engineering in the first place

- Do you know why your network behaves as it does?
- Is there a much simpler design,
  that does not do much worse?

4. If you already assume a Titan GPU, there must be other things to do besides endless training

Massive computer power also allows, e.g, for on-line morphing and image correlation (2D elastic matching) during operation. No need to save large NNs!

BTW: bookkeeping hundreds of NNs in a dynamic world with changing class definitions each, is a complex endeavor

Good algorithms give a higher % if the CPU gets faster, without retraining or code rewriting. For instance: max.-depth search can be deeper with the same timeout in [s].

5. How to keep & attract researchers during the Machine-Learning revolution?

Even MSc students are bought away by companies, jeopardizing their graduation in return for a (fixed) contract

The same goes for PhDs.

Handwriting isn't exactly as 'useful' and impressive as ML in genomics, pharmacology and logistics.

What are you doing my son | my daughter?
"*I am in handwriting recognition!*"

vs

"*I am involved in Deep Learning*"

Answer to question – What is most important?

- Already mentioned was the data starvation. We need labeled data because automatic data augmentation does not fully cover all variations. At the same time, we as a community are understandably reluctant to use transductive labeling (promoting high-confidence recognition results to the next-stage training set) without human supervision. Therefore fresh data are always needed, first to show what performances to real unseen data are, then to add them to the training set.

Next important thing?

- It struck me that Machine Learning can learn a lot from current robotics. Whereas in ML the tendency is only to arrive at higher performances (and then forgetting about the explanations for them), researchers in robotics (cf. Boston Dynamics) currently kick their robots, once these are standing upright. The idea would be to test (a) on real unseen data, or (b) **distort the input quality** of test sets etc., to find out when and where an approach fails: perf(angle)?, perf(scale)? Also, testing on images of out-of-vocabulary words should be used, for instance using Edit distance to find out whether a recognizer provides **intuitive** results to humans.

  Also performance prediction is not often done (See Isabelle Guyon's performance prediction benchmarks at NIPS). It is better to predict 75% and obtain 75% in reality than claiming 90% and getting 60% when you demonstrate a system to a company with their own fresh data.

Use of human reading behavior

- Angelo Marcelli made the point that we could make more use of human eye movement data. The group of Andreas Dengel already did some work in this respect (Busch et al, 2008). I fully agree. For instance, although CNN is supposed to be convolutional, it is only convolutional **after** the segmentation of the proverbial 256x256 pixel square from a big input image and taking it as raw CNN input. Finding objects on a whole page by a convolution with a 256x256 mask will be very expensive with current image sizes, which may be e.g., 7000x4000 pixels. In such cases it is useful to use knowledge on salience, both for regular computer vision and for layout analysis. Also the general behavior of humans analyzing a page is an important inspiration.

[note: SIFT was designed by Itti & Koch 2000) for this purpose, but there is more information than SIFT can deliver. For instance, there are also symmetry-detecting kernels (Kootstra, 2009].