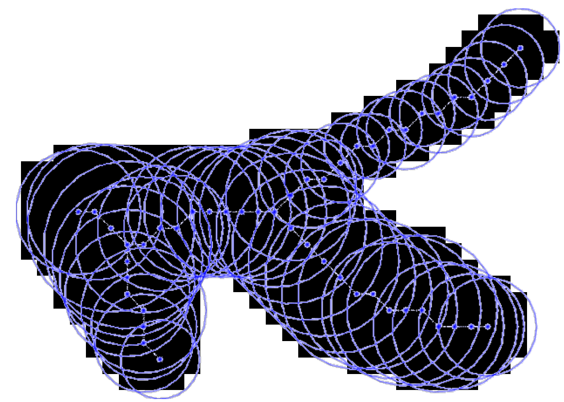
Inkball Models for Character Localization and Out-of-Vocabulary Word Spotting

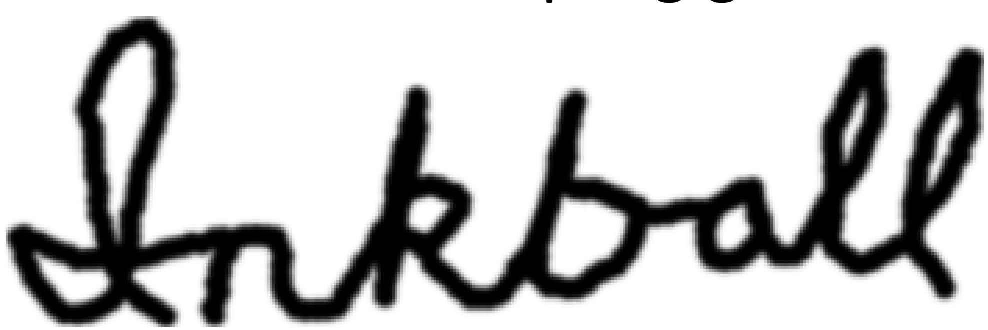
Nicholas Howe

SMITH COLLEGE

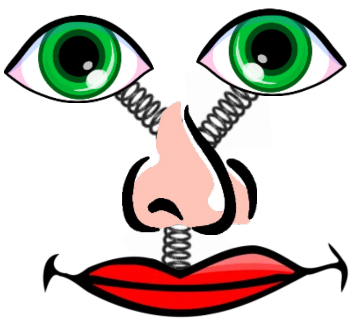
Review: Inkball Models



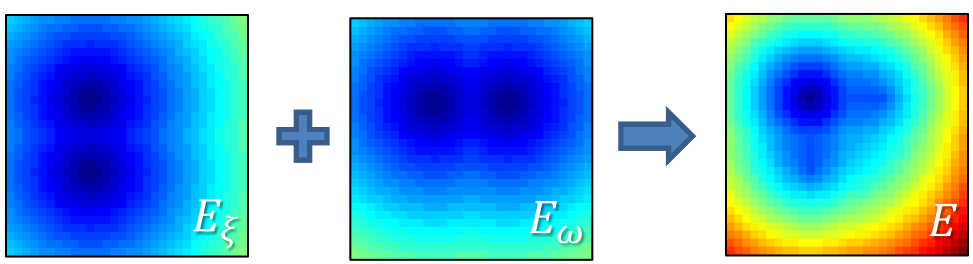
- Writing model = disks of ink in a particular configuration
- Any sample gives a model
- Flexible connections between adjacent disks
  - Gaussian distribution around offset point
  - Generative model: sampling gives new versions



Part-Structured Models



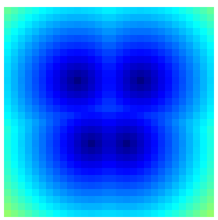
- Complex model is made of simple parts in a spatial relationship
- Proposed layout of parts is a ***configuration***
- Likelihood of configuration has two factors:
  - Do observations support layout of parts? E_ω
  - Does layout of parts match expected offsets? E_ξ



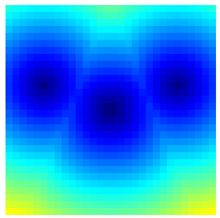
$$E = E_\xi + \lambda E_\omega$$

Efficient Inference

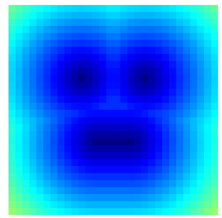
- Part detectors do some localization



Eyes

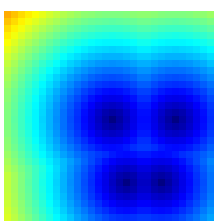


Nose

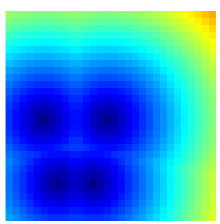


Mouth

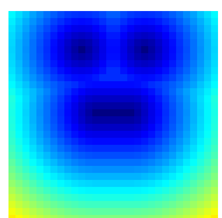
- Offset detections and combine



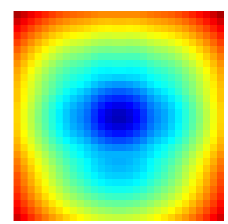
Left eye
to nose



Right eye
to nose



Mouth
to nose



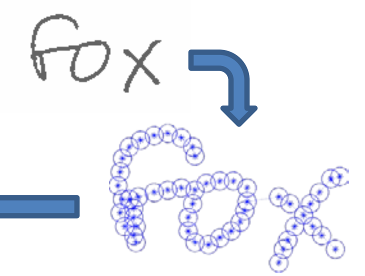
Combined
nose
likelihood

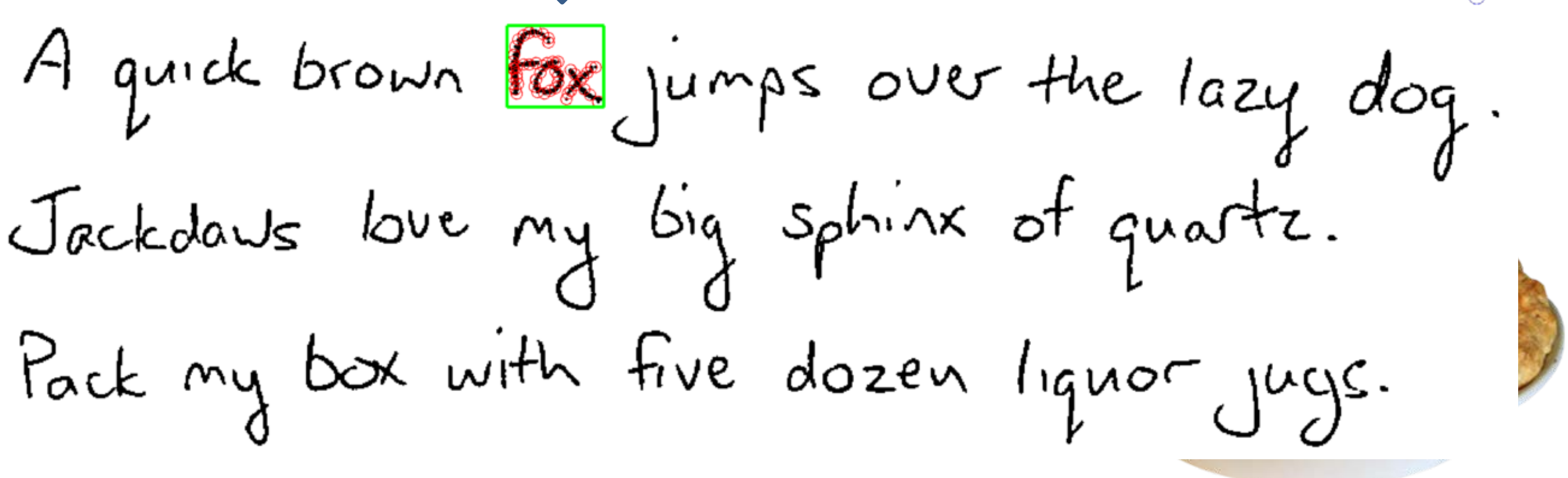
Accounting for subordinate parts clarifies nose position

Given nose position, can place subordinate parts

Prior Work: ICDAR 2013

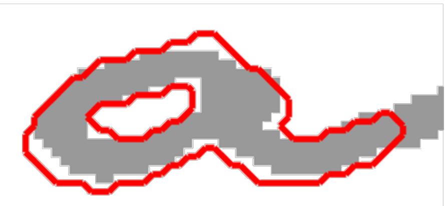
- Used inkball models for word spotting
- No training: each query word used as model
- Localizes target word on page of text





This Work: Two Goals

- Inkball models for character segmentation
  - Attribute individual pixels to characters
  - Known transcript only

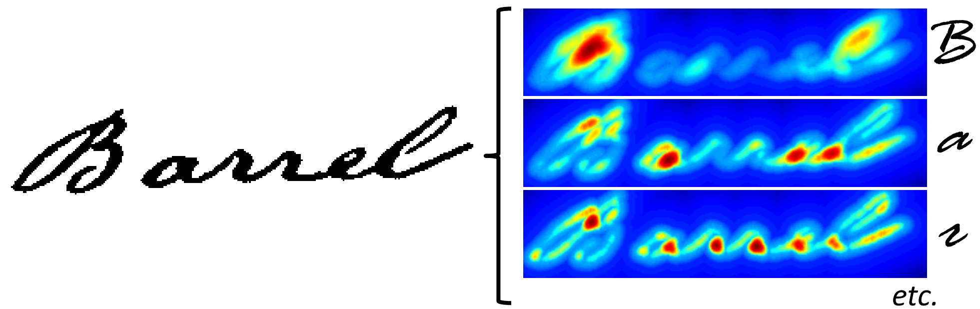


- Word spotting with text queries
  - Use synthetic word models
  - Relies on character models developed above

`Regiment`

Character Localization

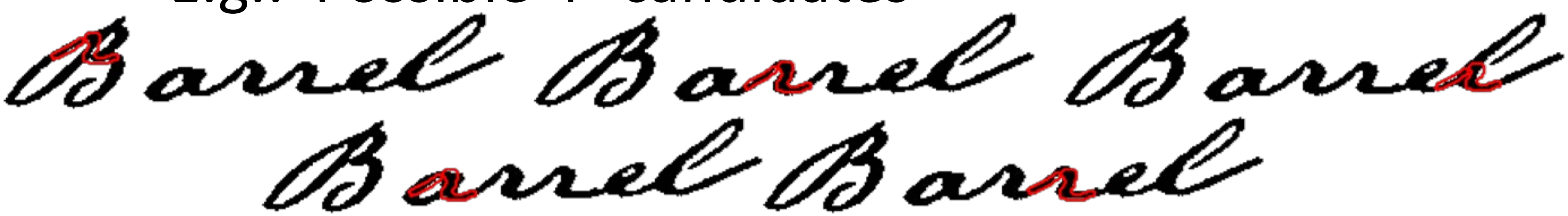
- Maximal points of character model fit
  - Multiple scales
  - Any location



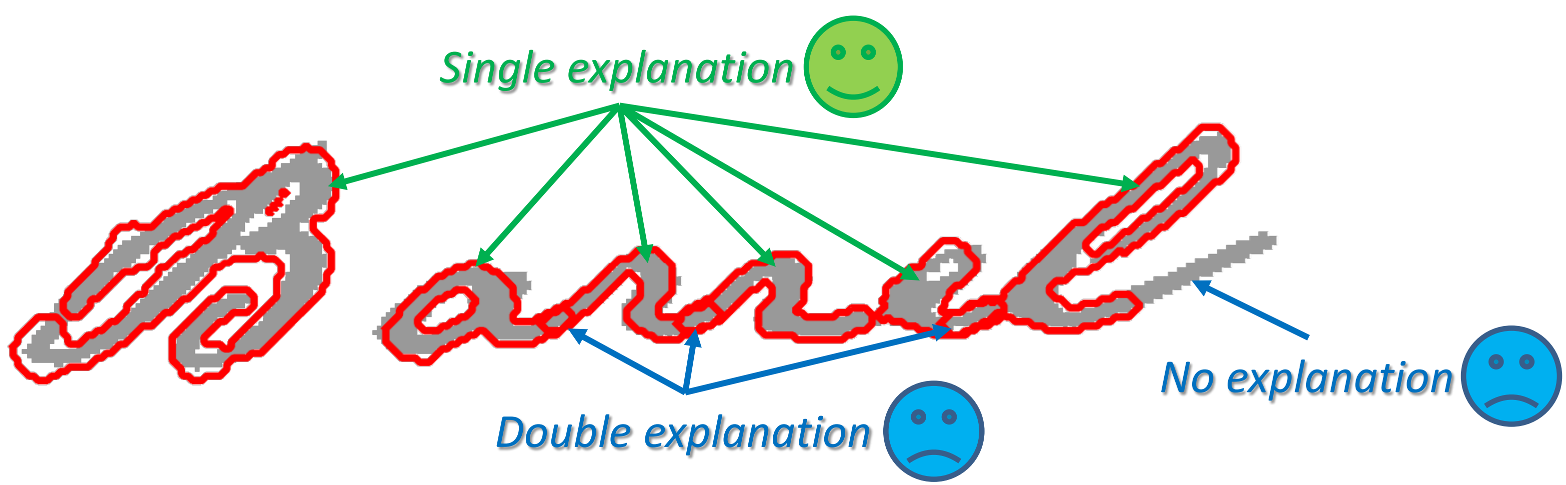
- Energy minimization chooses best sequence for entire word at once
  - Expected **(x,y)** displacement
  - Scale consistency
  - Explanation of all ink pixels

Every Pixel Wants to be Happy

- Render each candidate fit against image
  - E.g.: Possible ‘r’ candidates



- Pixels with exactly one explanation are best



Final Pixel Attribution

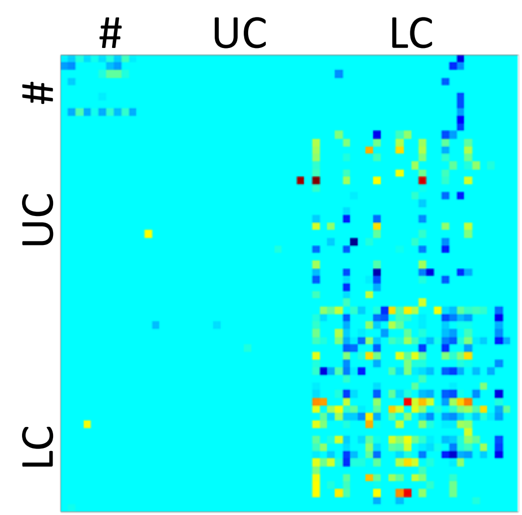
- Clean segmentation with some heuristics
  - Fix attribution problems using nearest neighbor

  - Untouched components are stray marks

Bulk Statistics

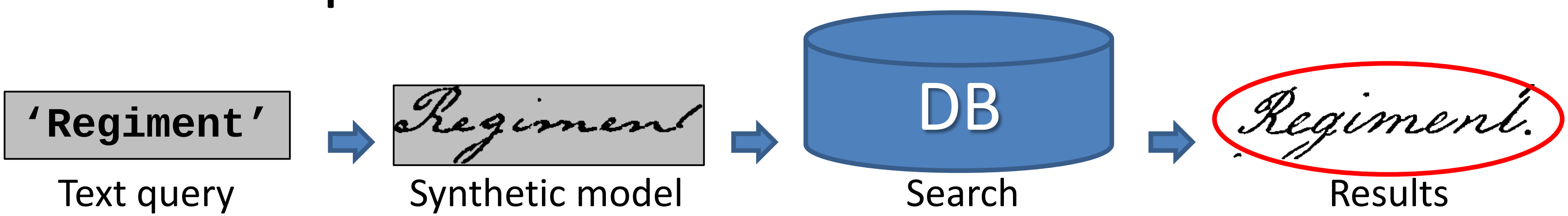
- Fit to large data set → Useful statistics
  - Character separation for bigrams
  - 1D or 2D offsets (e.g., superscripts)

- Problem of sparse data (tz rarer than th)
  - Bin samples by bigram
  - Add mean offset to every bin
  - Median bin value = offset estimate
  - Robust; conservative



Synthetic Queries

- What do you need to build synthetic words?
  - Model of each character
  - Displacement data for character bigrams
- Search process:



- Data sets:

Data Set Profile

GW20

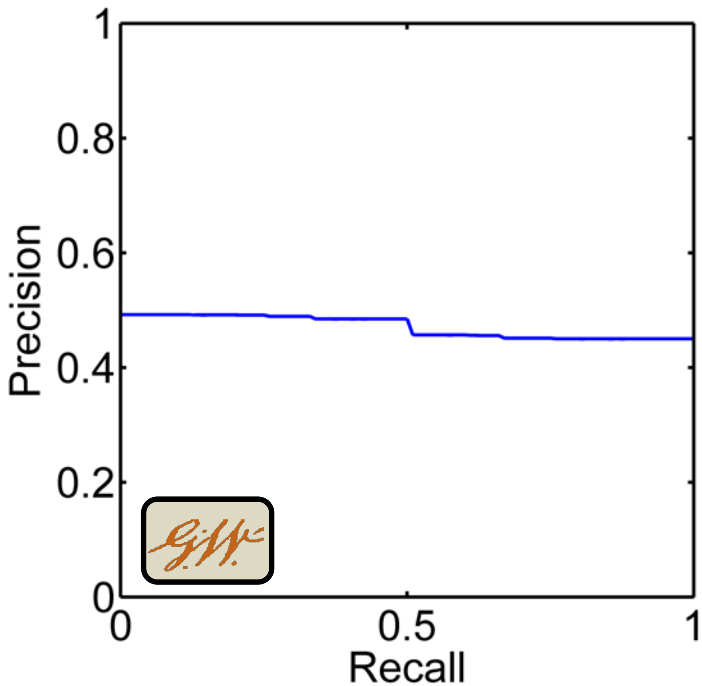
- Chancery script
- 60 characters
- Binarized
- Not deslanted
- 4857 words

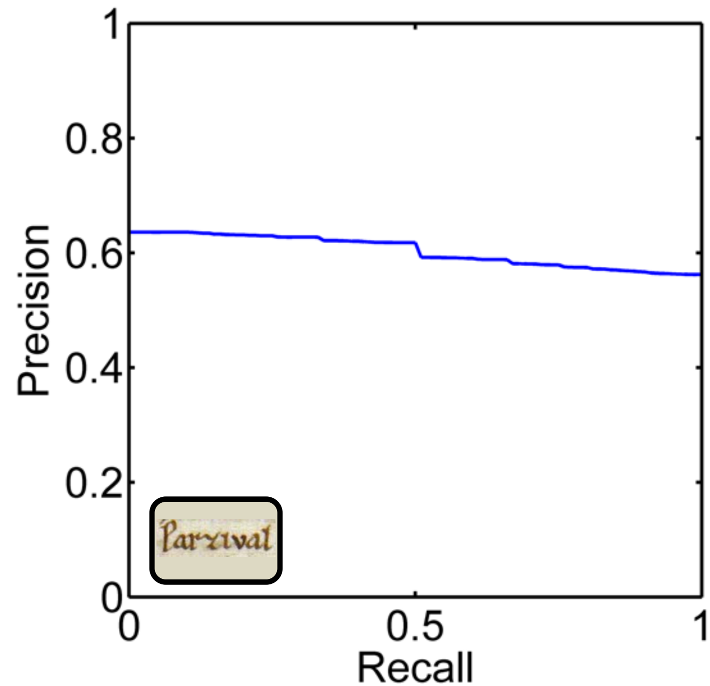
Parzival

- Medieval German
- 94 characters (includes accented)
- Low-quality binarized & deslanted
- 23485 words

OOV Performance

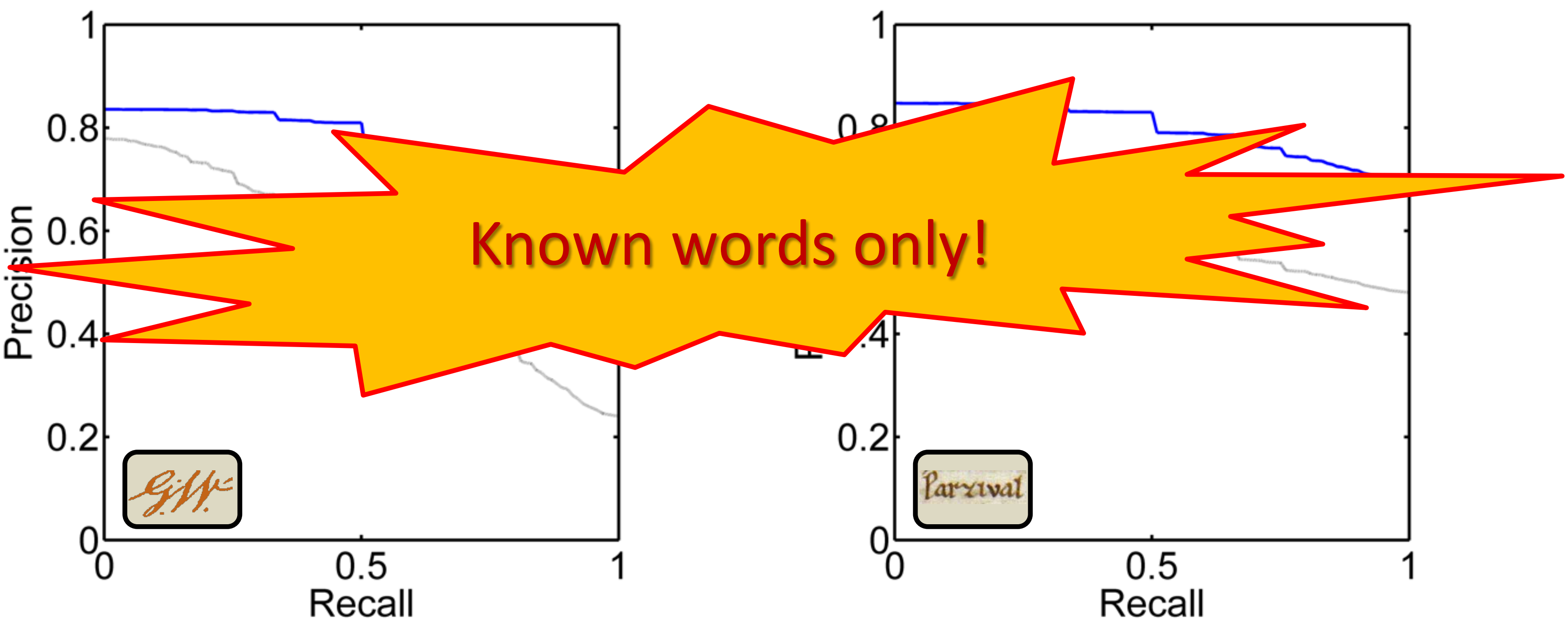
- Precision around 50% on rare (OOV) words





Real vs. Synthetic

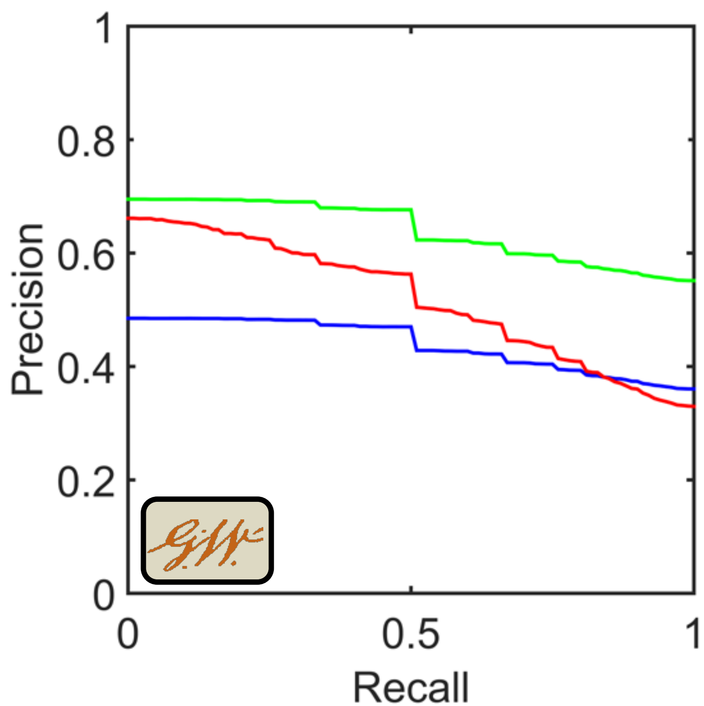
- Results for real query images vs. synthetic

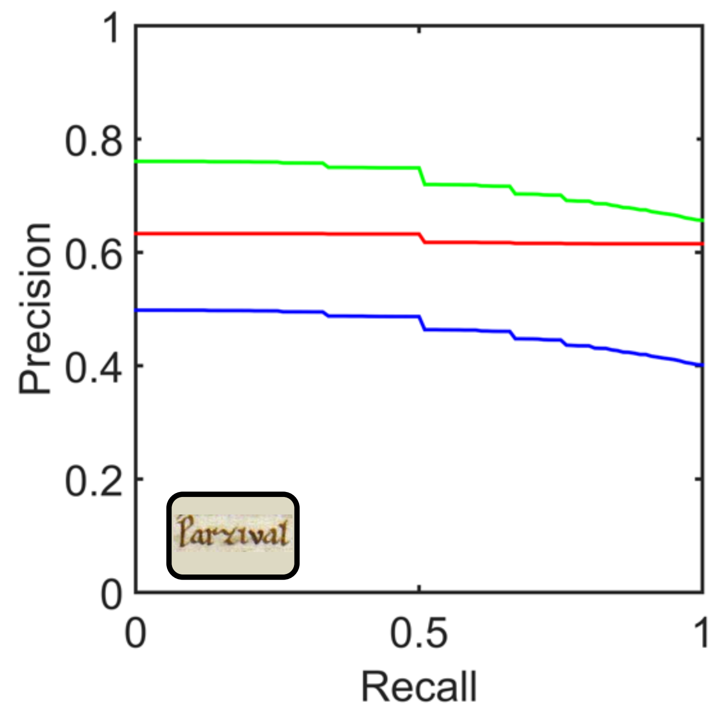


Full Vocabulary Results

- Results on all images (in & out of vocabulary)

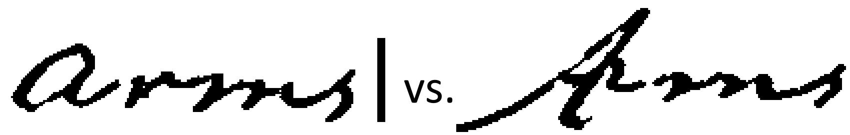
QBE only, Synthetic only, Hybrid QBE/Synthetic



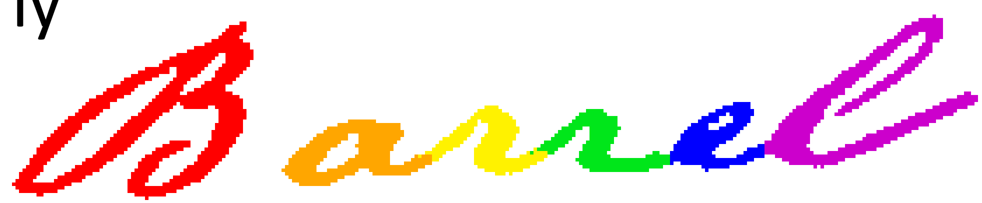


Conclusion

- Inkball models allow synthetic query images
- Improvement possible with future work
  - Letter variants
  - Better character joins

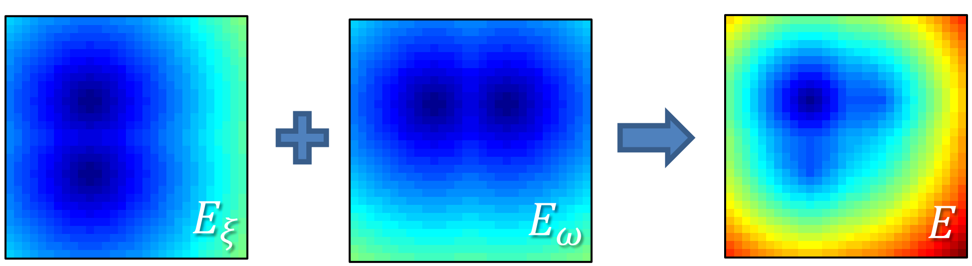


- Inkball models give algorithmic insight into handwritten forms
  - Locate letters and parts of letters
  - Attribute ink properly



Part-Structured Models

- Complex object made of simple parts in a spatial relationship
- Two factors give location likelihood:
  - Match of observations to part appearance E_ω
  - Proximity to offset locations of connected parts E_ξ

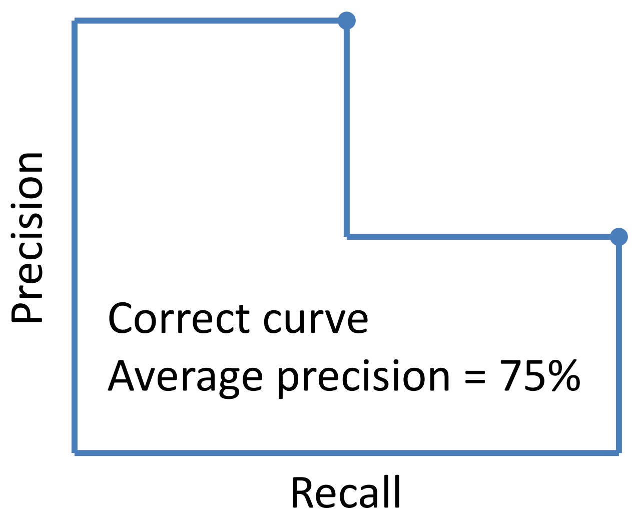


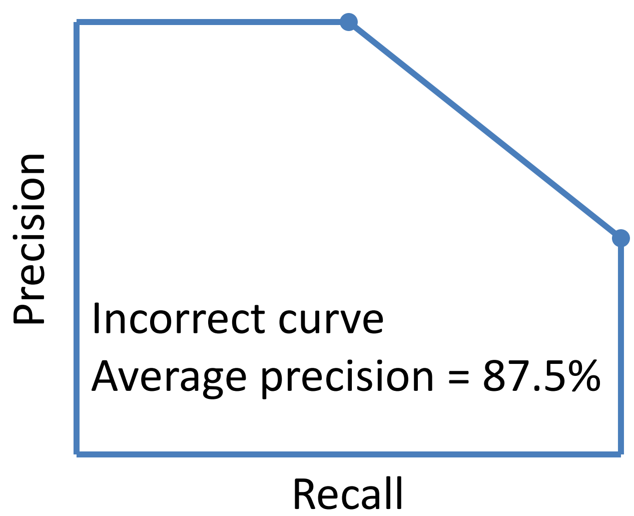
$$E = E_\xi + \lambda E_\omega$$

- Tree structure on parts ➔ efficient algorithm

What was the error?

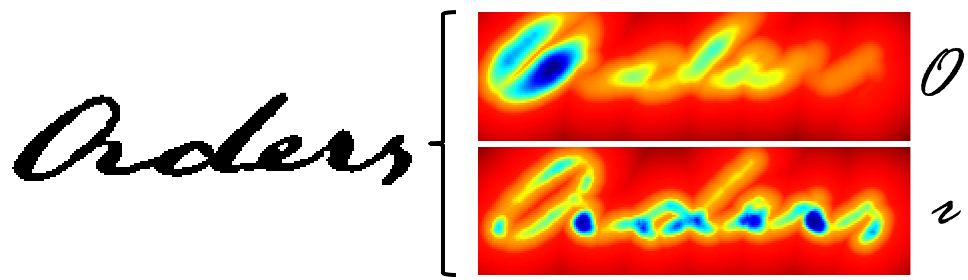
- Error in ICDAR 2013 paper: bad interpolation
- Significant when few exemplars
- Example: 2 relevant words, ranked 1 & 3





Character Localization

- Maximal points of character model fit
  - Multiple scales
  - Any location



- Energy minimization chooses best sequence for entire word at once
  - Expected **(x,y)** displacement
  - Scale consistency
  - Explanation of all ink pixels