Measuring Human Perception to Improve Handwritten Document Transcription

Walter J. Scheirer
Computer Vision Research Laboratory
Department of Computer Science and Engineering
University of Notre Dame
Has the AI renaissance solved handwritten document transcription?
Liber septimus regestorum domini Honorii pape III
odoribus pervenit, unguentum effusum nomen tuum.
Quomodo unguentum quod effusum est suavidadem longe latae.
Dispergit ille xvi nomen effusum eis. In muniera terra, xvi
nominatur in omnium mundo predicatur dvs.
Unguentum
enum effusum est nomen eius. Nunc morti nomen audetur,
quod prius audietur tum clam debatur angustia. Neque
enum grecorum quisquam meminit eius, nec ullam gentium
letterum historia deillo seu exeris scriptum aliquid inueni
mus; summum irradiavit in mundo eduxit secum legem & pro-
phoras. Absque cumplum est unguentum effusum nomen
tuum, propter eius unuenculae dilexerunt te. Quia perspex
sum caritati diuinitatis appetibilis.
Congrua nomen effusio
nus in septem unguentum effusum e nomen tuum.
haec de
sponsa adulescentularis consuetud, quando illa sponsa patre
rogavit. Xad ipsum sponsum comminus laquebatur, nec du
adulescentulae adecant.
In medius vero praeibus ingreditur
unuenculae chorus sponsae laudatur et quasi propter
d eius dilexerunt te, adulescentulae.
Objectives

- Segment the text
- Produce a faithful plaintext transcription of what appears on the page
- Perform machine translation
- Explore a text’s broader context via distant reading
script that hasn’t been used since the 13th century

*Liber septimus regestorum domini Honorii pape III*
Liber septimus regestorum domini Honorii pape III
Liber septimus regestorum domini Honorii pape III
Liber septimus regestorum domini Honorii pape III

“-or(um)”

cum ait. pratt. puludo. quop ac tahuit omhy a maru uscp ad nuovs vitce
cuattes. 7 a onhy, posessionhy, positer in monte se stehante plantys 7 aitre
Senagillie de nuee spiroo Senogalen. 7 Curte que nacte Trabelite. at castello
q7 nacetw. Origoo a omhy hominyi 7eo2 bomb. et luit puentes se Castre vaca
ry. Castre Ramustet e Castellare Aluy Leoni e Castellare Scorzalepons.
Liber septimus regestorum domini Honorii pape III
Machine Learning

- Learn a model for a task based on training data
- Success is often more related to data than the chosen learning algorithm
- Collecting data for problem domains where experts operate with specialized knowledge is difficult
- Special consideration for understanding what is easy and what is difficult
Perceptual Annotation: Measuring Human Vision to Improve Computer Vision

IEEE T-PAMI, August 2014
Supervised Learning

A “sink or swim” approach

No effort to tailor the learning to the human ability to learn from particular images.
Perceptual Annotation

- Much information about human capacities can be of direct value for machine learning:
  - Some images are learnable, and some are not.
  - Learnability varies with experience.
  - Some things are easily learned, other things take more time.

- Such detailed information reflecting human capacity is what we call a perceptual annotation.
Visual Psychophysics

Probe psychological and perceptual thresholds through controlled manipulation of stimuli.

Careful management of stimulus construction, ordering and presentation allows for precise determination of perceptual thresholds.
Perceptual Annotation

1. Citizen scientists participate in visual recognition tests on the web

2. Pattern of human performance is modeled

3. Human weighted loss function

4. Penalty for margins not consistent with human data

5. Resulting sparse solution makes predictions more consistent with human decisions

Queries

\[ \langle x_1, x_2, \ldots, x_n, y_1, y_2, \ldots, y_n \rangle \]
3 Alternative Forced Choice

Press the number (1, 2 or 3) corresponding to the image with the face.
Hinge loss:
\[ \phi_h(z) = \max(0, 1 - z) \]
where \[ z = yf(x) \]
Human-weighted loss:

\[ \phi_\psi(x, z) = \max(0, (1 - z) + M(x, z)) \]
FDDDB
Verba Volant, Scripta Manent: An Open Source Platform For Collecting Data To Train OCR Models For Manuscript Studies

Digital Humanities, July 2018
Measurements From Expert Annotators

- Bounding boxes at multiple levels
- Pixel-level annotation
- Reaction time
- Likert scale rating of difficulty
The Homer Multi-Text Project
http://www.homermultitext.org/
Image Citation Tool

Right-click here to copy URN

urn:cite:hmt:vaimg:VA012RN-0013@0.17.0.2515.0.325.0.0263

Focus on Selection | See Selection | Selection in Context

Christopher Blackwell and Neel Smith, https://github.com/cite-architecture/ict2
Software: https://github.com/grieggs/Psychometric-Annotator
Legenda aurea sive lombardica (http://www.e-codices.unifr.ch/en/list/one/sbe/0629) License: CC-BY-NC
Trace out each individual letter to segment them. To start on to the next letter click the "Add Annotation" button. Include accents with the characters they are most associated with. Press submit when finished.
Trace out each individual letter to segment them. To start on to the next letter click the "Add Annotation" button. Include accents with the characters they are most associated with. Press submit when finished.
Trace out each individual letter to segment them. To start on to the next letter click the "Add Annotation" button. Include accents with the characters they are most associated with. Press submit when finished.
Character Annotator v.0.0.1

Simply input the character shown into the text box and then rate on a scale of 1-10 how difficult it was to transcribe. Click The button below to start. Note that the button might take a moment to load.
Measuring Human Perception to Improve Handwritten Document Transcription

arXiv, April 2019
Objective: Psychophysical Loss for Neural Networks

- Incorporate measurements of human vision into the training process of a CNN
  - Original perceptual annotation approach only applied to classifiers

- Build a transcription pipeline that really works
  - Too much emphasis on beating datasets in this area
Perceptual Annotation for a Neural Network

Measured reaction times for different hand drawn ‘a’s

Psychophysical Loss

\[ L(x_i, l_i, \epsilon_i, z_i) \]

predicatur dnt.
nunc morti...
claudebatur ang...
Assigning Penalties for a Reading Task

- Easy
  - 2852 ms
  - $z = m - 2852 = 27148$
  - Larger Cost

- Hard
  - 24008 ms
  - $z = m - 24008 = 5992$
  - Smaller Cost

max. time = $m = 30000$
CTC Loss

\[ \mathcal{L}_{\text{CTC}} = - \sum \log p(l_i | y_i) \]

Sequence Produced by Network

Training Image

Ground-Truth Label Sequence

Training Data Set

Shi et al. T-PAMI 2017
Psychophysical Loss

Calculate a psychophysical penalty for images

Sorted reaction times: \[ R = \{r_1, \ldots, r_n\} \]

Maximum reaction time: \[ m = \max(R) \]

Psychophysical penalty: \[ z_i = m - r_i \]
Psychophysical Loss

\[ L_{\text{psych}} = \sum_{x_i, \epsilon_i, l_i \in \mathcal{X}} - \log p(l_i | y_i) + (\epsilon_i \times z_i) \]

Associated Character Error Rate
Another choice for transcription: CRNN

Shi et al. T-PAMI 2017
Results
odoribus perdundente, uinguentum effusum nomen tuum... quomodo uinguentum quod effusione sua ad orem longe latet: disperge tecum, nomen effusum e,... Nunquam terra, uinq... nominatur in omni mundo predicatur dixit: "Unguentum
enum effusum e nomen eius: Nunc morti nomen audite
quod prius iudicem tutum cladebatur angustius. Neque
enum greorum quisquam memint euis: neq; uella gentium
litterorum historia dedito seu eixi ecriptum aliquid inueni
nus: ut uir laudavit in mundo educit secum legem & pro-
phoras, iisque complices est uinguentum effusum: nomen
tuum... Propter ea uinculae, dilexerunt te... Quia perspici
sum caritus diuissa e incerta us. Congruo nomen effusio
nus in secur... uinguentum effusum e nomen tuum... Haviest
causa sponsa uincularum constat, quando illa sponsa uatu
rogavit: "Adipsum sponsum comminum laquebatur nec du
uincularum adeant, in medius vero praebitis ingredierunt
uincularum chorus; a sponsa laudatur et quaeis propter
uinculae dilexerunt te, uincularum..."
# Does it work?

<table>
<thead>
<tr>
<th>Input</th>
<th>URNN</th>
<th>LM</th>
<th>GT</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>comparing the failure of a subsidized duke</td>
<td>domins ad memoriem illius dedarauit</td>
<td>dominus ad memoriam illius declaravit</td>
</tr>
<tr>
<td></td>
<td>a grey superimposition of respectability over</td>
<td>et missis post virum dei legatis eium ad venire devuote rogaret</td>
<td>et missis post virum dei legatis eius advenire devote rogaret</td>
</tr>
<tr>
<td></td>
<td>the original color of his own natural vowels,</td>
<td>largiatur nos similitudinem auri faciemus non habemus aurum</td>
<td>largiatur nos similitudinem auri faciemus non habemus aurum</td>
</tr>
</tbody>
</table>
What impact does psychophysical loss have?

Practical impact: average of 598 character errors and 295 word errors were eliminated.
What impact does the language model have?

IAM

<table>
<thead>
<tr>
<th></th>
<th>CRNN</th>
<th>URNN</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>w/o Psych. Loss</td>
<td>w/ Language Model Correction</td>
</tr>
<tr>
<td></td>
<td>val1</td>
<td>val2</td>
</tr>
<tr>
<td>CER (% Improvement)</td>
<td>-0.67</td>
<td>-0.93</td>
</tr>
<tr>
<td>WER (% Improvement)</td>
<td>27.99</td>
<td>24.32</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>CER (% Improvement)</th>
<th>WER (% Improvement)</th>
</tr>
</thead>
<tbody>
<tr>
<td>val1</td>
<td>-6.92</td>
<td>-27.31</td>
</tr>
<tr>
<td>val2</td>
<td>-7.36</td>
<td>-25.44</td>
</tr>
<tr>
<td>test</td>
<td>-7.10</td>
<td>-23.68</td>
</tr>
<tr>
<td></td>
<td>-4.47</td>
<td>-4.77</td>
</tr>
<tr>
<td></td>
<td>-4.95</td>
<td></td>
</tr>
</tbody>
</table>
Off-the-Shelf tools popular in Digital Humanities

St. Gall

<table>
<thead>
<tr>
<th></th>
<th>Ocular</th>
<th>OCRopus</th>
<th>Tesseract</th>
<th>CRNN + LM</th>
<th>URNN + LM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CER</td>
<td>WER</td>
<td>CER</td>
<td>WER</td>
<td>CER</td>
</tr>
<tr>
<td>IAM [39]</td>
<td>92.83</td>
<td>122</td>
<td>33</td>
<td>43.1</td>
<td>55.71</td>
</tr>
<tr>
<td>St. Gall [26]</td>
<td>90.2</td>
<td>143.8</td>
<td>71</td>
<td>102.4</td>
<td>39.71</td>
</tr>
<tr>
<td>All Latin</td>
<td>90.9</td>
<td>138.2</td>
<td>62.9</td>
<td>94.2</td>
<td>51.69</td>
</tr>
</tbody>
</table>
Where do we go after we transcribe the registers?
Tesserae Project

https://tesserae.caset.buffalo.edu/
Supported by:

Notre Dame Research, Notre Dame College of Arts and Letters, the Medieval Institute, Notre Dame Office of Mission Engagement and Church Affairs, Notre Office of Digital Learning, and Adobe Systems