Data Mining and Machine Learning with Images

Benjamin Kiessling

September 22, 2023

Key Topics

- Principles of working with digital images
- Image processing with Python
- Working with IIIF and Python
- Computer Vision
- Principles of Neural Networks for images
- Feature Engineering for Neural Networks

Key Topics

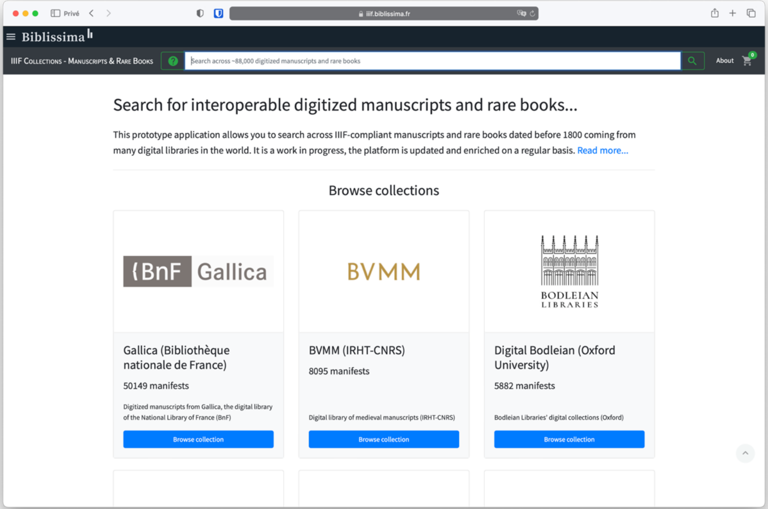
- Handwritten Text Recognition with Kraken
- Notebooks, academic code
- High Performance and Cloud computing
- Data archiving, publication, and reproducibility
- Group projects, including intermediate and final presentations

Introductions

Ben Kiessling - research engineer in the Digital Humanities

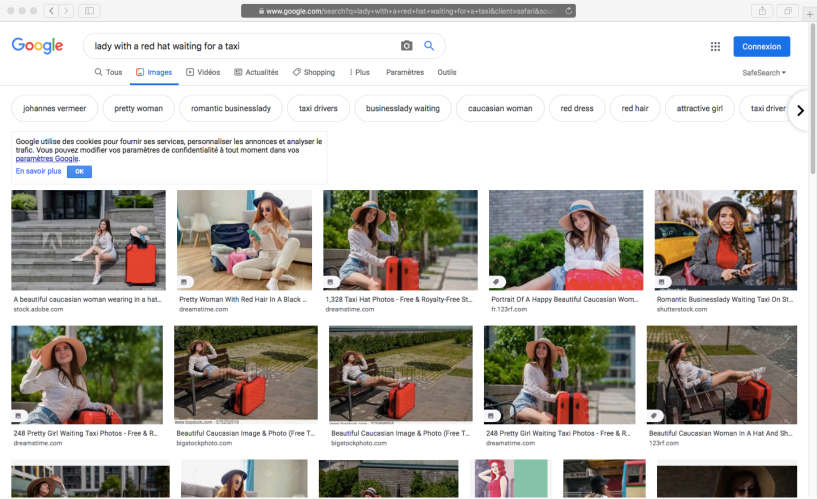
- Document analysis of historical material
- Focus on rare and esoteric documents
- Computer Scientist by training
- since 2018 at the EPHE

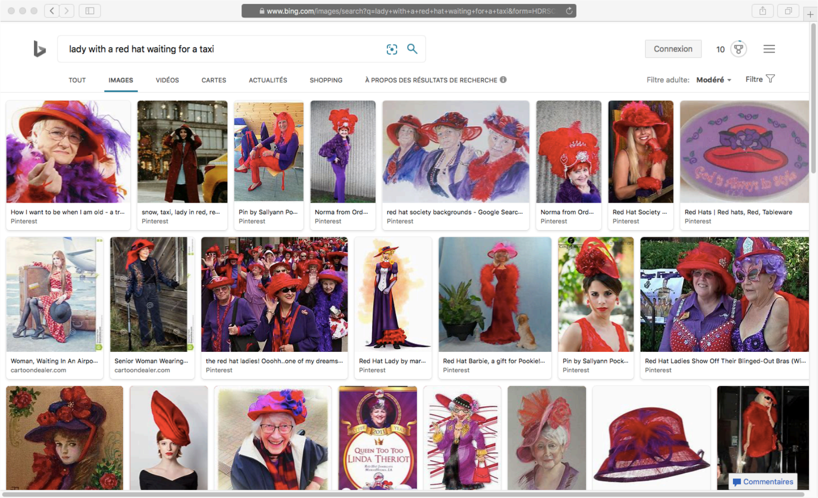
90000 digitised MSS via IIIF

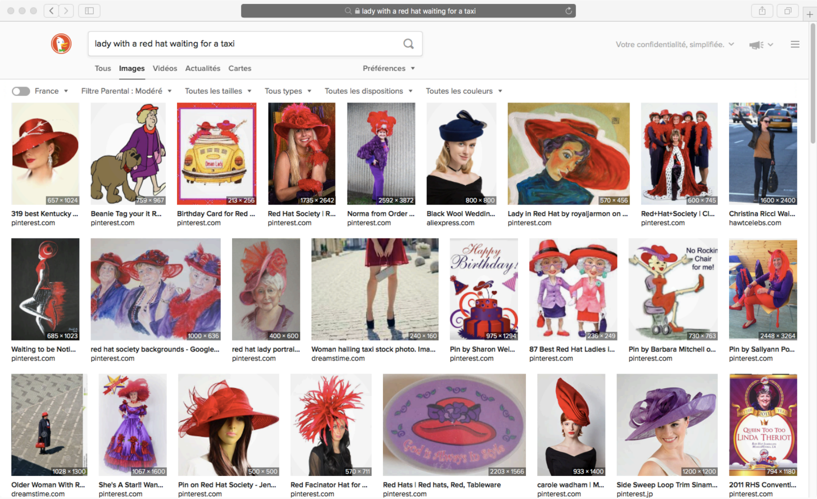


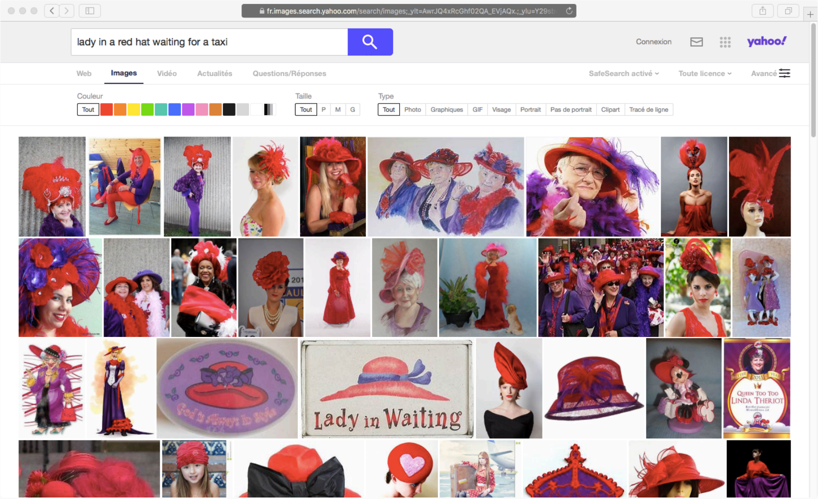
A Problem

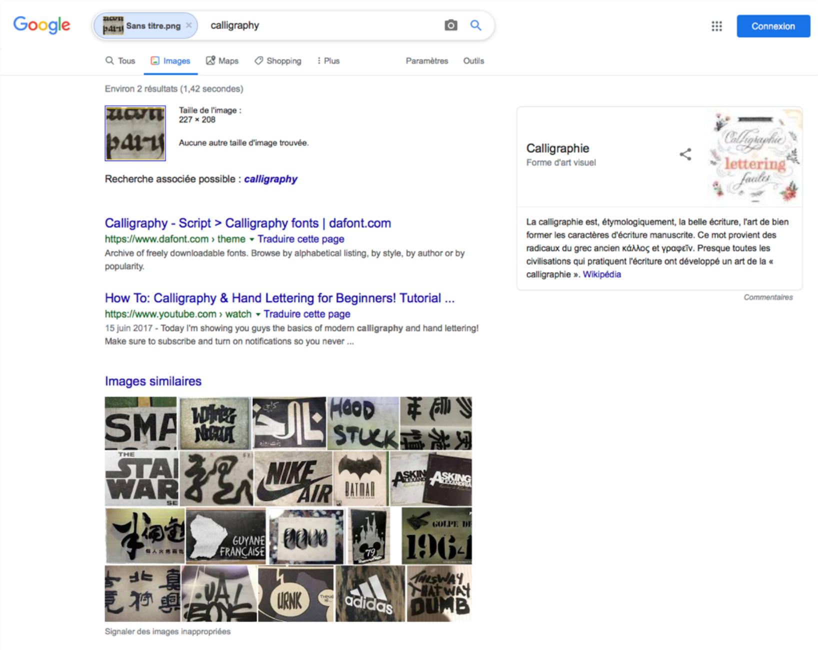
There is no efficient way of searching the www for, say, a picture of a "lady with a red hat waiting for a taxi", Mohammed Ghanbari, Video Coding: An Introduction to Standard Codecs (1999), §9.9


www.google.com/search?q=lady+with+a+red+hat+waiting+for+a+taxi&client=safari&sour
Google
lady with a red hat waiting for a taxi
Connexion
Tous
Images
Vidéos
Actualités
Shopping
Plus
Paramètres
Outils
SafeSearch
johannes vermeer
pretty woman
romantic businesslady
taxi drivers
businesslady waiting
caucasian woman
red dress
red hair
attractive girl
taxi driver
Google utilise des cookies pour fournir ses services, personnaliser les annonces et analyser le trafic. Vous pouvez modifier vos paramètres de confidentialité à tout moment dans vos paramètres Google.
En savoir plus
OK
A beautiful caucasian woman wearing in a hat...
stock.adobe.com
Pretty Woman With Red Hair In A Black ...
dreamstime.com
1,328 Taxi Hat Photos - Free & Royalty-Free St...
dreamstime.com
Portrait Of A Happy Beautiful Caucasian Wom...
fr.123rf.com
Romantic Businesslady Waiting Taxi On St...
shutterstock.com
248 Pretty Girl Waiting Taxi Photos - Free & R...
dreamstime.com
Beautiful Caucasian Image & Photo (Free T...
bigstockphoto.com
Beautiful Caucasian Image & Photo (Free T...
bigstockphoto.com
248 Pretty Girl Waiting Taxi Photos - Free & R...
dreamstime.com
Beautiful Caucasian Woman In A Hat And Sh...
123rf.com


www.bing.com/images/search?q=lady+with+a+red+hat+waiting+for+a+taxi&form=HDRSC
lady with a red hat waiting for a taxi
Connexion
10
TOUT
IMAGES
VIDÉOS
CARTES
ACTUALITÉS
SHOPPING
À PROPOS DES RÉSULTATS DE RECHERCHE
Filtre adulte: Modéré
Filtre
How I want to be when I am old - a tr...
Pinterest
snow, taxi, lady in red, re...
Pinterest
Pin by Sallyann Po...
Pinterest
Norma from Ord...
Pinterest
red hat society backgrounds - Google Searc...
Pinterest
Norma from Ord...
Pinterest
Red Hat Society ...
Pinterest
Red Hats | Red hats, Red, Tableware
Pinterest
Woman, Waiting In An Airpo...
cartoondealer.com
Senior Woman Wearing...
cartoondealer.com
the red hat ladies! Ooohh..one of my dreams...
Pinterest
Red Hat Lady by mar...
Pinterest
Red Hat Barbie, a gift for Pookie!...
Pinterest
Pin by Sallyann Pock...
Pinterest
Red Hat Ladies Show Off Their Blinged-Out Bras (Wi...
Pinterest
Queen Too Too
Linda Theriot
Commentaires

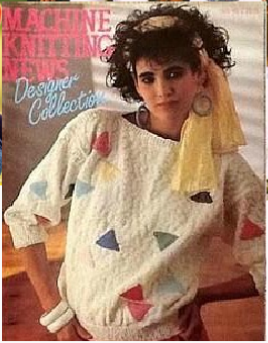

lady with a red hat waiting for a taxi
lady with a red hat waiting for a taxi
Votre confidentialité, simplifiée.
Tous
Images
Vidéos
Actualités
Cartes
Préférences
France
Filtre Parental : Modéré
Toutes les tailles
Tous types
Toutes les dispositions
Toutes les couleurs
657 × 1024
319 best Kentucky ...
pinterest.com
759 × 967
Beanie Tag your it R...
pinterest.com
213 × 256
Birthday Card for Red ...
pinterest.com
1735 × 2642
Red Hat Society | R...
pinterest.com
2592 × 3872
Norma from Order ...
pinterest.com
800 × 800
Black Wool Weddin...
aliexpress.com
570 × 456
Lady in Red Hat by royaljarmon on ...
pinterest.com
600 × 745
Red+Hat+Society | Cl...
pinterest.com
1600 × 2400
Christina Ricci Wai...
hawtcelebs.com
685 × 1023
Waiting to be Noti...
pinterest.com
1000 × 636
red hat society backgrounds - Google...
pinterest.com
400 × 600
red hat lady portrai...
pinterest.com
240 × 160
Woman hailing taxi stock photo. Ima...
dreamstime.com
Happy Birthday!
975 × 1294
Pin by Sharon Wei...
pinterest.com
236 × 249
87 Best Red Hat Ladies I...
pinterest.com
No Rockin Chair for me!
730 × 763
Pin by Barbara Mitchell o...
pinterest.com
2448 × 3264
Pin by Sallyann Po...
pinterest.com
1028 × 1300
Older Woman With R...
dreamstime.com
1067 × 1600
She's A Star!! Wan...
pinterest.com
500 × 500
Pin on Red Hat Society - Jen...
pinterest.com
570 × 711
Red Facinator Hat for ...
pinterest.com
2203 × 1566
Red Hats | Red hats, Red, Tableware
pinterest.com
933 × 1400
carole wadham | M...
pinterest.com
1200 × 1200
Side Sweep Loop Trim Sinam...
pinterest.jp
Queen Too Too
Linda Theriot
794 × 1180
2011 RHS Conventi...
pinterest.com


fr.images.search.yahoo.com/search/images;_ylt=AwrJQ4xRcGhf02QA_EVjAQx.;_ylu=Y29sb
lady in a red hat waiting for a taxi
Connexion
yahoo!
Web
Images
Vidéo
Actualités
Questions/Réponses
SafeSearch activé
Toute licence
Avancé
Couleur
Tout
Taille
Tout P M G
Type
Tout Photo Graphiques GIF Visage Portrait Pas de portrait Clipart Tracé de ligne
Lady in Waiting


Google
Sans titre.png
calligraphy
Connexion
Tous
Images
Maps
Shopping
Plus
Paramètres
Outils
Environ 2 résultats (1,42 secondes)
Taille de l'image :
227 × 208
Aucune autre taille d'image trouvée.
Recherche associée possible : calligraphy
Calligraphy - Script > Calligraphy fonts | dafont.com
https://www.dafont.com › theme ▾ Traduire cette page
Archive of freely downloadable fonts. Browse by alphabetical listing, by style, by author or by popularity.
How To: Calligraphy & Hand Lettering for Beginners! Tutorial ...
https://www.youtube.com › watch ▾ Traduire cette page
15 juin 2017 - Today I'm showing you guys the basics of modern calligraphy and hand lettering! Make sure to subscribe and turn on notifications so you never ...
Images similaires
Signaler des images inappropriées
Calligraphie
Forme d'art visuel
La calligraphie est, étymologiquement, la belle écriture, l'art de bien former les caractères d'écriture manuscrite. Ce mot provient des radicaux du grec ancien κάλλος et γραφεῖν. Presque toutes les civilisations qui pratiquent l'écriture ont développé un art de la « calligraphie ». Wikipédia
Commentaires

1

2



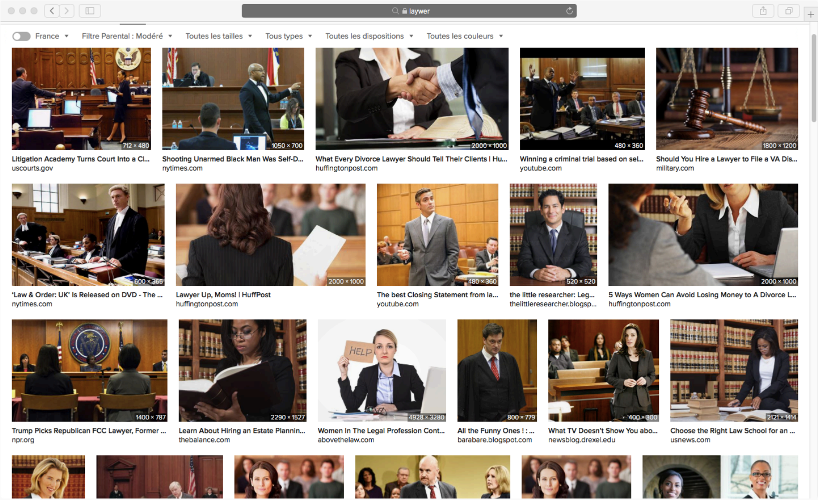
3

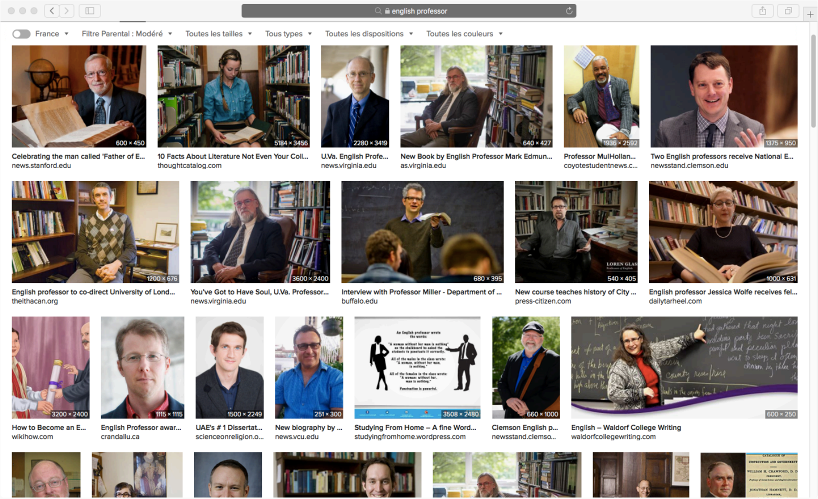
Searching for images

- Look at metadata
  - Title, data, author, ...
- Look at context
  - E.g. look at text around the image, captions, ...
- Content-based image retrieval (CBIR)
  - Try to analyse the image itself
  - Try to automatically identify what the image is
  - Try to find images that match an example (Query-by-example/reverse image search)

CBIR: Query by Example

- Calculate features of example and query set
- Measure distance between
- Which features? Color (histogram), textures (Fourier descriptors, neural features), scene identification (CNNs)

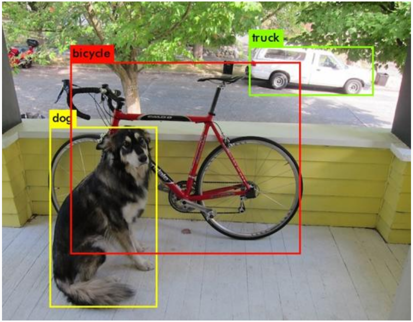

laywer
France
Filtre Parental : Modéré
Toutes les tailles
Tous types
Toutes les dispositions
Toutes les couleurs
712 × 480
Litigation Academy Turns Court Into a Cl...
uscourts.gov
1050 × 700
Shooting Unarmed Black Man Was Self-D...
nytimes.com
2000 × 1000
What Every Divorce Lawyer Should Tell Their Clients | Hu...
huffingtonpost.com
480 × 360
Winning a criminal trial based on sel...
youtube.com
1800 × 1200
Should You Hire a Lawyer to File a VA Dis...
military.com
600 × 365
'Law & Order: UK' Is Released on DVD - The ...
nytimes.com
2000 × 1000
Lawyer Up, Moms! | HuffPost
huffingtonpost.com
480 × 360
The best Closing Statement from la...
youtube.com
520 × 520
the little researcher: Leg...
thelittleresearcher.blogsp...
2000 × 1000
5 Ways Women Can Avoid Losing Money to A Divorce L...
huffingtonpost.com
1400 × 787
Trump Picks Republican FCC Lawyer, Former ...
npr.org
2290 × 1527
Learn About Hiring an Estate Plannin...
thebalance.com
HELP
4928 × 3280
Women In The Legal Profession Cont...
abovethelaw.com
800 × 779
All the Funny Ones ! : ...
barabare.blogspot.com
400 × 300
What TV Doesn't Show You abo...
newsblog.drexel.edu
2121 × 1414
Choose the Right Law School for an ...
usnews.com


english professor
France
Filtre Parental : Modéré
Toutes les tailles
Tous types
Toutes les dispositions
Toutes les couleurs
Celebrating the man called 'Father of E...
news.stanford.edu
10 Facts About Literature Not Even Your Coll...
thoughtcatalog.com
U.Va. English Profe...
news.virginia.edu
New Book by English Professor Mark Edmun...
as.virginia.edu
Professor MulHollan...
coyotestudentnews.c...
Two English professors receive National E...
newsstand.clemson.edu
English professor to co-direct University of Lond...
theithacan.org
You've Got to Have Soul, U.Va. Professor...
news.virginia.edu
Interview with Professor Miller - Department of ...
buffalo.edu
New course teaches history of City ...
press-citizen.com
English professor Jessica Wolfe receives fel...
dailytarheel.com
How to Become an E...
wikihow.com
English Professor awar...
crandallu.ca
UAE's # 1 Dissertat...
scienceonreligion.o...
New biography by ...
news.vcu.edu
Studying From Home – A fine Word...
studyingfromhome.wordpress.com
Clemson English p...
newsstand.clemso...
English – Waldorf College Writing
waldorfcollegewriting.com

Deep Learning and Object Recognition

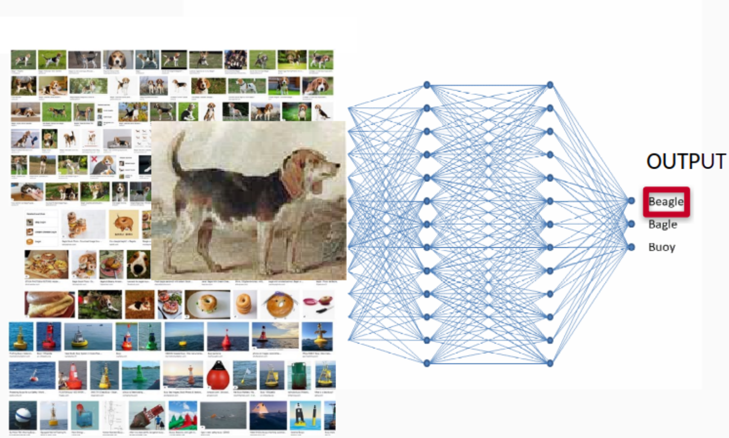
Convolutional Neural Networks (CNNs) have been very successful for object recognition

- Approx. 99.7% accurcay on MNIST database
- Over 97% accuracy on ImageNet database
- Fast enough for video (though lower accuracy, e.g. `https://youtu.be/yQwfDxBMtXg`)



How does it work?

Supervised Machine Learning ("AI") / "Deep Learning" for classification



Limitations of Deep Learning

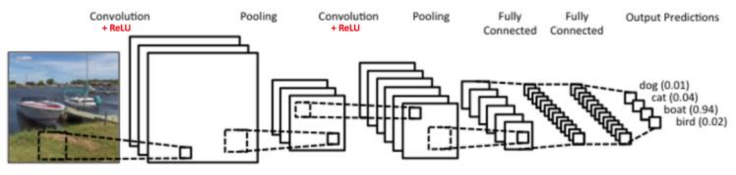
DL depends entirely on the data used for training. This has several important consequences:

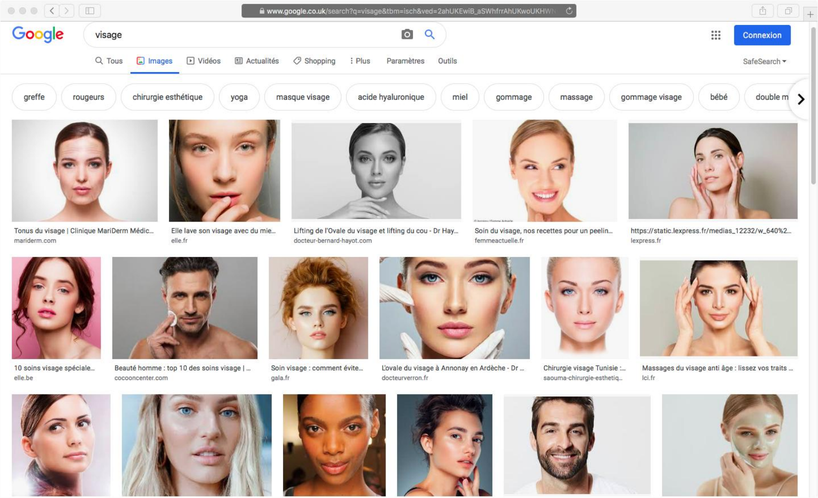
- The data is (probably) more important than the algorithm
- The is why Google etc. want our data so much:
  - Emails, web seraches, photos, SMS, movements, ...
  - Our data is worth may billions of dollars (but we usually give it away without thinking)
- (Supervised) DL only works if you know the answers in advance
- Biases in the data are transferred directly to the AI

Convolutional Neural Network

Applies filters across the input

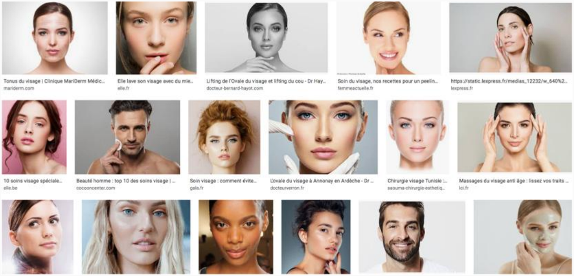
- e.g. edge detectors, line detection, ...for images
- Creates feature maps across the image
- In practice, filters are learned by the system, rarely intelligible to humans
- In general, layers build up more complex shapes



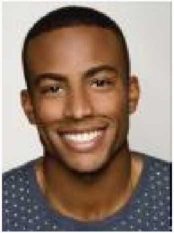
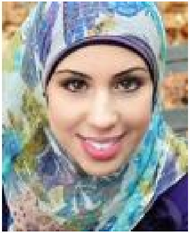

www.google.co.uk/search?q=visage&tbm=isch&ved=2ahUKEwiB_aSWhfrrAhUKwoUKHW
Google
visage
Connexion
Tous
Images
Vidéos
Actualités
Shopping
Plus
Paramètres
Outils
SafeSearch
greffe
rougeurs
chirurgie esthétique
yoga
masque visage
acide hyaluronique
miel
gommage
massage
gommage visage
bébé
double m
Tonus du visage | Clinique MariDerm Médic...
mariderm.com
Elle lave son visage avec du mie...
elle.fr
Lifting de l'Ovale du visage et lifting du cou - Dr Hay...
docteur-bernard-hayot.com
Soin du visage, nos recettes pour un peelin...
femmeactuelle.fr
https://static.lexpress.fr/medias_12232/w_640%2...
lexpress.fr
10 soins visage spéciale...
elle.be
Beauté homme : top 10 des soins visage | ...
cocooncenter.com
Soin visage : comment évite...
gala.fr
L'ovale du visage à Annonay en Ardèche - Dr ...
docteurverron.fr
Chirurgie visage Tunisie :...
saouma-chirurgie-esthetiq...
Massages du visage anti âge : lissez vos traits ...
lci.fr

Training Bias

If these are faces...



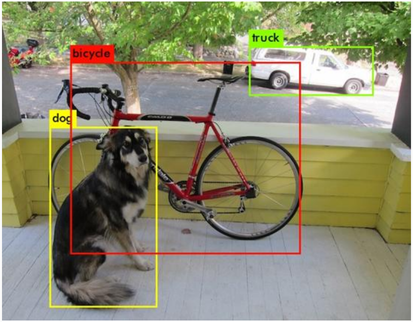
...then what are these?

Deep Learning and Object Recognition

Convolutional Neural Networks (CNNs) have been very successful for object recognition

- Approx. 99.7% accurcay on MNIST database
- Over 97% accuracy on ImageNet database
- Fast enough for video (though lower accuracy, e.g. `https://youtu.be/yQwfDxBMtXg`)



MNIST database

ImageNet database

Harvested $>$14M images from the Internet

Labelled by hand (Mechanical Turk)

Now used widely to train Computer Vision (for classification)

Questions for ImageNet

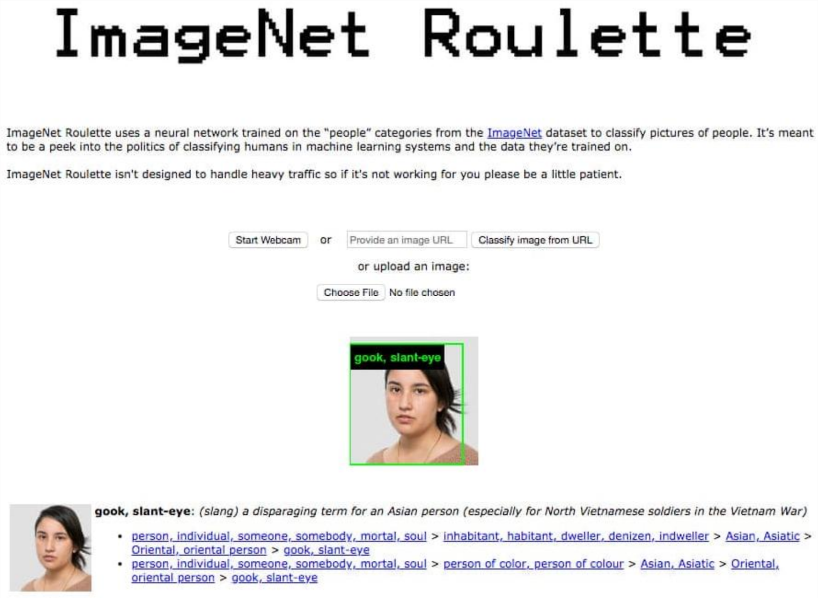
Who decides which labels to use?

Where did the images come from? Where they used with permission and respect data protection law?

What biases are in the image set and labels?

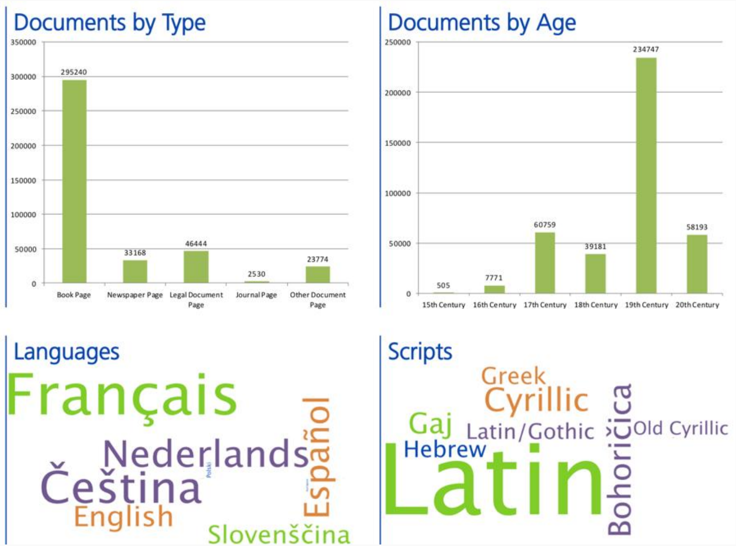
See, e.g., Kate Crawford and Trevor Paglen, "Excavating AI: The Politics of Training Sets for Machine Learning" (September 19, 2019) `https://excavating.ai`

ImageNet Roulette



IMPACT Dataset

More than half a million representative historical text-based images compiled from major European libraries



"Criminality" prediction

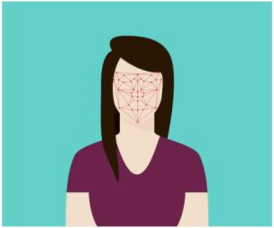
HARRISBURG UNIVERSITY OF SCIENCE AND TECHNOLOGY

About HU | Admissions | Degrees & Programs | The Campus | Esports

HU facial recognition software predicts criminality

HU facial recognition software predicts criminality

A group of Harrisburg University professors and a Ph.D. student have developed automated computer facial recognition software capable of predicting whether someone is likely going to be a criminal.



With 80 percent accuracy and with no racial bias, the software can predict if someone is a criminal based solely on a picture of their face. The software is intended to help law enforcement prevent crime.

Ph.D. student and NYPD veteran Jonathan W. Korn, Prof. Nathaniel J.S. Ashby, and Prof. Roozbeh Sadeghian titled their research "A Deep Neural Network Model to Predict Criminality Using Image Processing."

"We already know machine learning techniques can outperform humans on a variety of tasks related to facial recognition and emotion detection," Sadeghian said. "This research indicates just how powerful these tools are by showing they can extract minute features in an image that are highly predictive of criminality."

"Criminality" prediction

Automated Inference on Criminality using Face Images

Xiaolin Wu
McMaster University
Shanghai Jiao Tong University
xwu510@gmail.com

Xi Zhang
Shanghai Jiao Tong University
zhangxi_19930818@sjtu.edu.cn

Abstract

We study, for the first time, automated inference on criminality based solely on still face images, which is free of any biases of subjective judgments of human observers. Via supervised machine learning, we build four classifiers (logistic regression, KNN, SVM, CNN) using facial images of 1856 real persons controlled for race, gender, age and facial expressions, nearly half of whom were convicted criminals, for discriminating between criminals and non-criminals. All four classifiers perform consistently well and empirically establish the validity of automated face-induced inference on criminality, despite the historical controversy surrounding this line of enquiry. Also, some discriminating structural features for predicting criminality have been found by machine learning. Above all, the most important discovery of this research is that criminal and non-criminal face images populate two quite distinctive manifolds. The variation among criminal faces is significantly greater than that of the non-criminal faces. The two manifolds consisting of criminal and non-criminal faces appear to be con-

management science, criminology, etc.

In all cultures and all periods of recorded human history, people share the belief that the face alone suffices to reveal innate traits of a person. Aristotle in his famous work Prior Analytics asserted, "It is possible to infer character from features, if it is granted that the body and the soul are changed together by the natural affections". Psychologists have known, for as long as a millennium, the human tendency of inferring innate traits and social attributes (e.g., the trustworthiness, dominance) of a person from his/her facial appearance, and a robust consensus of individuals' inferences . These are the facts found through numerous studies [3, 39, 5, 6, 10, 26, 27, 34, 32].

Independent of the validity of pedestrian belief in the (pseudo)science of physiognomy, a tantalizing question naturally arises: what facial features influence average Joes' impulsive and yet consensual judgments on social attributes of a non-acquaintance member of their own specie? Attempting to answer the question, Todorov and Oosterhof proposed a data-driven statistical modeling method to find visual determinants of social attributes by asking human

"Criminality" prediction

Responses to Critiques on Machine Learning of Criminality Perceptions (Addendum of arXiv:1611.04135)

Xiaolin Wu
McMaster University
Shanghai Jiao Tong University
xwu510@gmail.com

Xi Zhang
Shanghai Jiao Tong University
zhangxi_19930818@sjtu.edu.cn

In November 2016 we submitted to arXiv our paper "Automated Inference on Criminality Using Face Images". It generated a great deal of discussions in the Internet and some media outlets. Our work is only intended for pure academic discussions; how it has become a media consumption is a total surprise to us.

Although in agreement with our critics on the need and importance of policing AI research for the general good of the society, we are deeply baffled by the ways some of them mispresented our work, in particular the motive and objective of our research.

1. Name calling

It should be abundantly clear, for anyone who reads our paper with a neutral mind setting, that our only motive is to know if machine learning has the potential of acquiring hu-

> *attributes and facial features are correlated, because being a criminal requires a host of abnormal (outlier) personal traits. If the classification rate turns out low, then the validity of face-induced social inference can be safely negated."*

By a magical stretch of imagination, few of our c intertwine the above passage into some of our hones servations and morph them into the following deductic they insist, ours:

> *"Those with more curved upper lips and eyes closer together are of a lower social order, prone to (as Wu and Zhang put it) "a host of abnormal (outlier) personal traits" ultimately leading to a legal diagnosis of "criminality" with high probability."*

Limitations of Machine Learning

The entire world of statistically based machine learning right now is based on learning from historical examples and from statistics. … By its nature, that means it will always be a reflection of the past. And if the past is the future you want, that's fine. I tend to think that it's not, so we need something else. *Kristian Hammond, quoted in 'Is AI finally closing in on human intelligence?' FT Magazine, 12 November 2020*

Association for Computing Machinery
US Public Policy Council (USACM)

usacm.acm.org
facebook.com/usacm
twitter.com/usacm

January 12, 2017

Statement on Algorithmic Transparency and Accountability

Computer algorithms are widely employed throughout our economy and society to make decisions that have far-reaching impacts, including their applications for education, access to credit, healthcare, and employment.[1] The ubiquity of algorithms in our everyday lives is an important reason to focus on addressing challenges associated with the design and technical aspects of algorithms and preventing bias from the onset.

An algorithm is a self-contained step-by-step set of operations that computers and other 'smart' devices carry out to perform calculation, data processing, and automated reasoning tasks. Increasingly, algorithms implement institutional decision-making based on analytics, which involves the discovery, interpretation, and communication of meaningful patterns in data. Especially valuable in areas rich with recorded information, analytics relies on the simultaneous application of statistics, computer programming, and operations research to quantify performance.

ACM Principles for Transparency

1. Awareness among designers and users
2. Access and redress of those affected
3. Accountability of institutions
4. Explanation encouraged from designers and users
5. Data Provenance of training data
6. Auditability of models, algorithms, data and decisions
7. Validation and testing should be routine, rigorous and documented


INDEPENDENT
HIGH-LEVEL EXPERT GROUP ON
ARTIFICIAL INTELLIGENCE
SET UP BY THE EUROPEAN COMMISSION
AI
THE ASSESSMENT LIST FOR
TRUSTWORTHY ARTIFICIAL
INTELLIGENCE (ALTAI)
for self assessment

AI reliability: An AI system is said to be reliable if it behaves as expected, even for novel inputs on which it has not been trained or tested earlier.

Explainability: Feature of an AI system that is intelligible to non-experts. An AI system is intelligible if its functionality and operations can be explained non technically to a person not skilled in the art.

Interpretability: Interpretability refers to the concept of comprehensibility, explainability, or understandability. When an element of an AI system is interpretable, this means that it is possible at least for an external observer to understand it and find its meaning.

Reproducibility: Reproducibility refers to the closeness between the results of two actions, such as two scientific experiments, that are given the same input and use the methodology, as described in a corresponding scientific evidence (such as a scientific publication). A related concept is *replication*, which is the ability to independently achieve non-identical conclusions that are at least similar, when differences in sampling, research procedures and data analysis methods may exist. Reproducibility and replicability together are among the main tools of the scientific method.

Traceability: Ability to track the journey of a data input through all stages of sampling, labelling, processing and decision making.

Human oversight, human-in-the-loop, human-on-the-loop, human-in-command: Human oversight helps ensure that an AI system does not undermine human autonomy or causes
other adverse effects. Oversight may be achieved through governance mechanisms such as a human-in-the-loop (HITL), human-on-the-loop (HOTL), or human-in-command (HIC) approach. Human-in-the-loop refers to the capability for human intervention in every decision cycle of the system, which in many cases is neither possible nor desirable. Human-on-the-loop refers to the capability for human intervention during the design cycle of the system and monitoring the system's operation. Human-in-command refers to the capability to oversee the overall activity of the AI system (including its broader economic, societal, legal and ethical impact) and the ability to decide when and how to use the system in any particular situation. This can include the decision not to use an AI system in a particular situation, to establish levels of human discretion during the use of the system, or to ensure the ability to override a decision made by a system. Moreover, it must be ensured that public enforcers have the ability to exercise oversight in line with their mandate. Oversight mechanisms can be required in varying degrees to support other safety and control measures, depending on the AI system's application area and potential risk. All other things being equal, the less oversight a human can exercise over an AI system, the more extensive testing and stricter governance is required.

Model Evasion: Evasion is one of the most common attacks on machine learning models (ML) performed during production. It refers to designing an input, which seems normal for a human but is wrongly classified by ML models. A typical example is to change some pixels in a picture before uploading, so that the image recognition system fails to classify the result.

Model Inversion: Model inversion refers to a kind of attack to AI models, in which the access to a model is abused to infer information about the training data. So, model inversion turns the usual path from training data into a machine-learned model from a one-way one to a two-way one, permitting the training data to be estimated from the model with varying degrees of accuracy. Such attacks raise serious concerns given that training data usually contain privacy-sensitive information.

Algorithmic Accountability

Computational approaches in the humanities often produce more questions than answers

A key principle of DH is that computers are not there to give definitive answers, but to help researchers pose questions and find data for study.

No software can be truly objective and unbiased.

Algorithmic Accountability

It's therefore imperative that researchers in the humanities can understand and use the results of software

- What do the results mean for questions in the humanities?
- What are the limits? The possibilities?
- What are the ideas behind the software?
- What biases, prejudices and assumptions underlie it?

A Problem

We have millions of images of handwriting

Palaeographers want to find examples of ‘similar’ writing

How can we do this?

- What does ‘similar’ mean?
- How similar is ‘similar’?
- How can we explain this to a computer?
- How can we describe handwriting?

(The same problem applies to much of the humanities, not just palaeography!)

A Problem

The central problem of AI is the question: What is the letter "a"? *Douglas Hofstadter (1985): Metamagical Themas, p. 633*

A Problem

Visualising Different Styles of Script

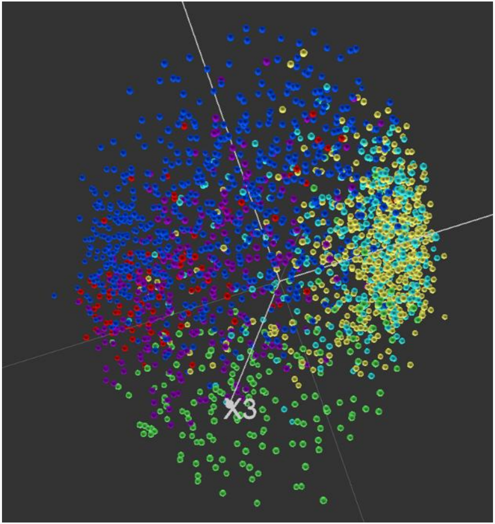


Figure 1: D. Stutzmann, Digital Medievalist 10 (2016)

Visualising Different Styles of Script

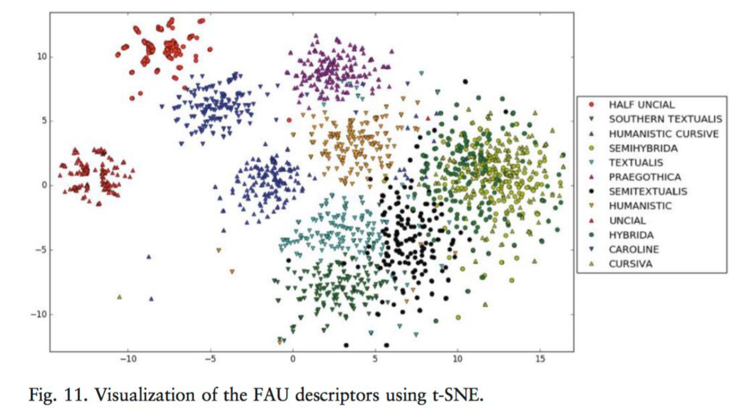


Figure 2: Kestemont, Christlein and Stutzmann, Speculum 92 (2017)

Visualising Different Styles of Script

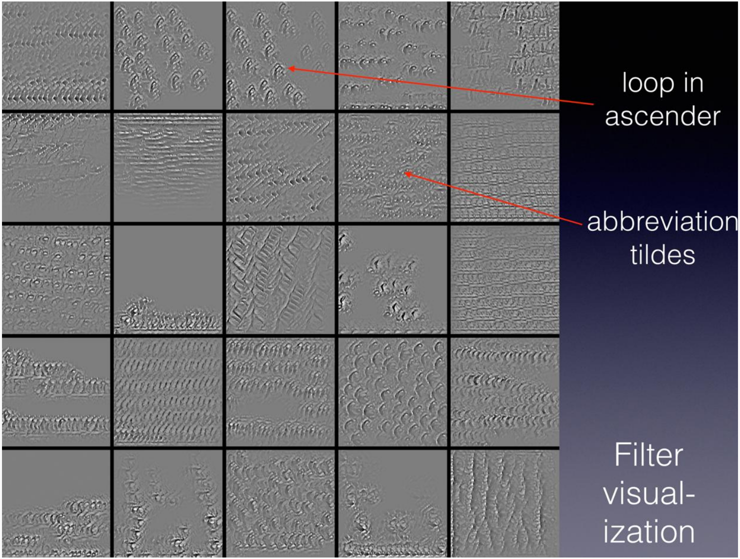


Figure 3: Kestemont (2016), DeepScript

A Need for Precision

Two things which are similar are always similar in certain respects. … Generally, similarity, and with it repetition, always presupposes the adoption of a point of view: some similarities or repetitions will strike us if we are interested in one problem or another. *Collette Sirat, Writing as Handwork (2006), p. 310, citing Popper 1968, pp. 420–22*

Applied Machine Learning: Handwritten Text Recognition

What is HTR?

'Hand-written text recognition'

Essentially OCR for manuscripts

NB that (usually) no longer uses character segmentation

- Treat text as sequence, not as independent images
- 'Long Short-Term Memory' (LSTM), not CNN
- Don't try to segment into characters, just lines
  - Character segmentation difficult for handwriting
  - Lines allow the system to use context

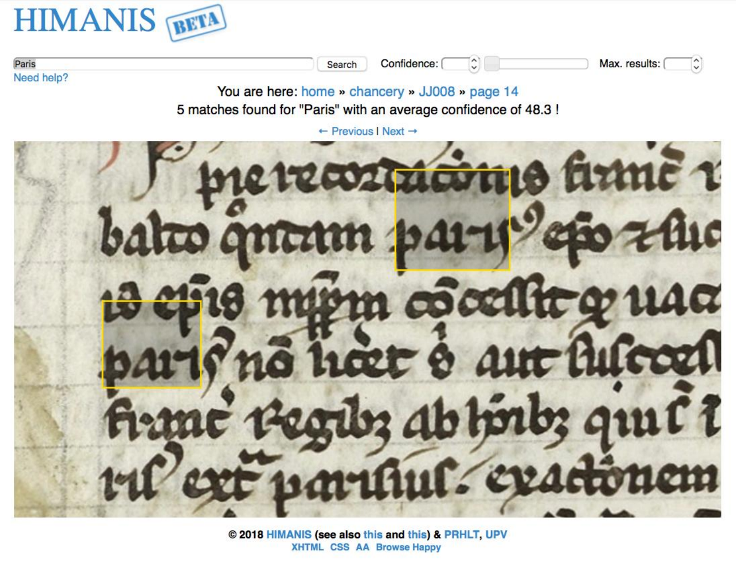
HTR/OCR vs Wordspotting

HTR/OCR results in a complete transcription

Input is an image, output is text

Wordspotting searches for images

- There is never a complete copy of the text
- Instead, the computer maps words (sequences of characters) to sections of image
- Input is text, output is an image


HIMANIS
BETA
Paris
Search
Confidence:
Max. results:
Need help?
You are here: home » chancery » JJ008 » page 14
5 matches found for "Paris" with an average confidence of 48.3 !
← Previous | Next →
© 2018 HIMANIS (see also this and this) & PRHLT, UPV
XHTML CSS AA Browse Happy

A Workflow for Automatic Transcription

1. Import images (via IIIF, PDF, zip, …)
2. Segment the images (via AI, via import, manually)
    - To find the regions and lines of text
    - Correct the results, retrain the AI, and repeat
3. Transcribe the text (via AI, via import, manually)
    - Correct the results, retrain the AI, and repeat
4. Export the text and/or the trained models

Also 2b: Change the sequence of lines via AI or by hand, correct, retrain, and repeat.

How does Segmentation work?

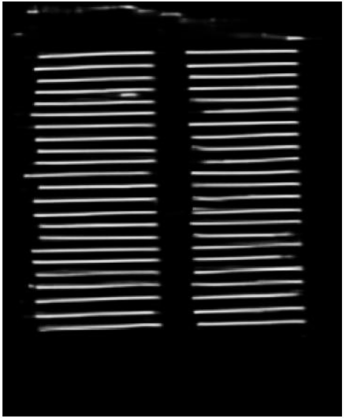
Segmentation is classification of pixels

- Input is an image (of a page), output is classification of pixels
- Writing, background, decoration, marginal addition, ...

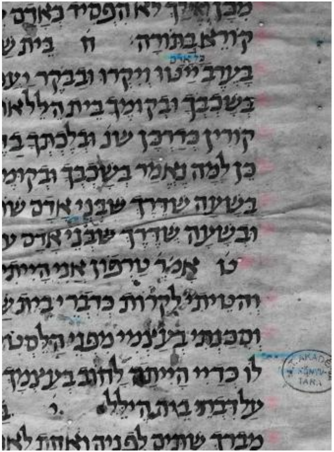
Distinction criteria: position or visual difference (different font, script, color)

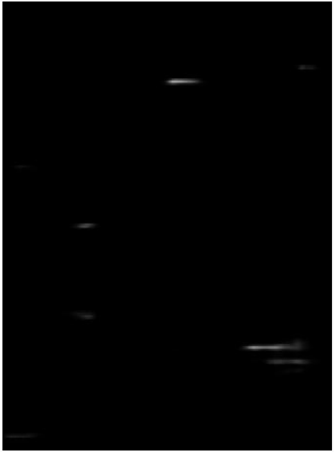
- Detection algorithm producing a series of 2D grayscale images, one for each type.
- The image is whiter where the system suspects the presence of an object of this type.
- Above a threshold: creation of a polygon for type X

Line Detection



Detection of Interlinear Additions





How does Line Transcription work?

Input is the image of a line of text

‘Answer’ is a line of text

NB that we use lines, not letters/characters

Can’t always divide handwriting into distinct characters

Can use context to resolve similar or ambiguous forms

But context makes it more complex than simple classification

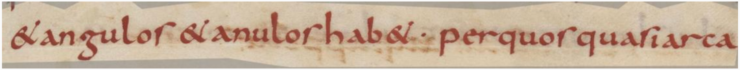


Figure 4: Image/query, label/solution: & angulos & anulos hab& . perquos quasi arca

Export and Interchange

Can import/export transcriptions

- (ALTO, PAGE, soon TEI export, probably Open Web Annotations export)

Can import/export trained models for layout and transcription

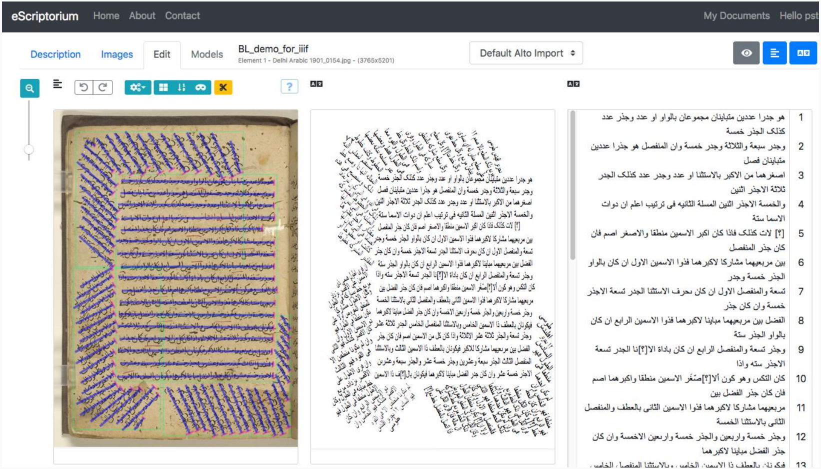
- Users not tied to eScriptorium
- Can share models, reducing significant load of training

But data sharing and interchange presupposes a common conceptual model and sufficiently precise standards

Languages in Scripta-PSL

Ancient Aramaic, Medieval Armenien, Pre-Imperian Chinese, Ptolomaic Egyptian, Elamite, Middle Iranian, Medieval Japanese, Old Javanese, Old Khmer, Meroitic, Pali, Soghdian, Sumerian, Classical Tamil, Tai-lue, Tokharian, Ugaritic, Umbrian, Old Vietnamese ...

eScriptorium



Challenges of Rare and Diverse Writing

Users working with dozens of different ‘rare’ and historical scripts

Different types of writing, fonts, Unicode blocks (if you’re lucky!)

- Alphabets, logographs, hieroglyphs...; ‘stroke’ scripts, ‘line’ scripts, ...

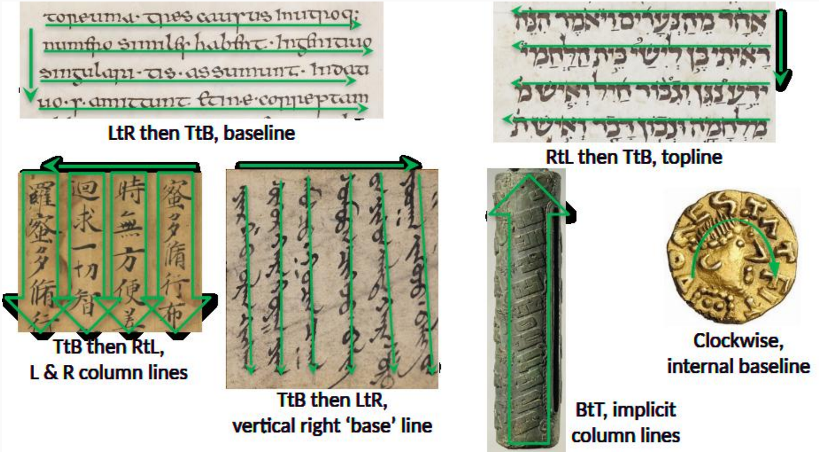
Different directions

- Left-right, right-left, top-bottom (then RtL or LtR), bottom-top, boustrophedon, diagonal, non-linear, mixed, ...
- Writing on baseline, from top-line, on vertical line, in grid, ...

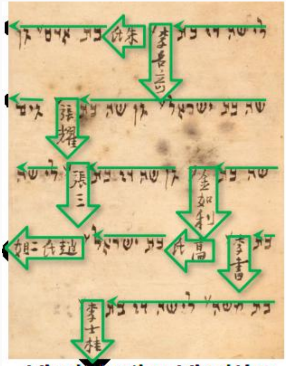
Different conventions for transcription and scholarly presentation Corpus often limited

- May be few exemplars
- May be no pre-existing models for language, layout etc.

A "simple" example: writing line & direction

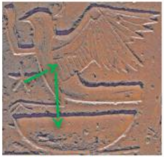


A "simple" example: writing line & direction



Mixed Direction, Mixed Line

Single Directionality, Single Line, Multiple Orientations



Non-linear

Some Challenges in Sharing Models and Data

Which transcription standards were used?

- Word spacing? Punctuation? Normalisation? Abbreviations? Illegible/unclear? Insertions? Allographs? Unicode normalisation?

Which segmentation standards were used?

- Baseline? Topline (e.g. for Hebrew)? Column lines (e.g. Chinese)? Nothing?
- Are there regions? Rectangular or polygonal? What ontology/typology (if any)?

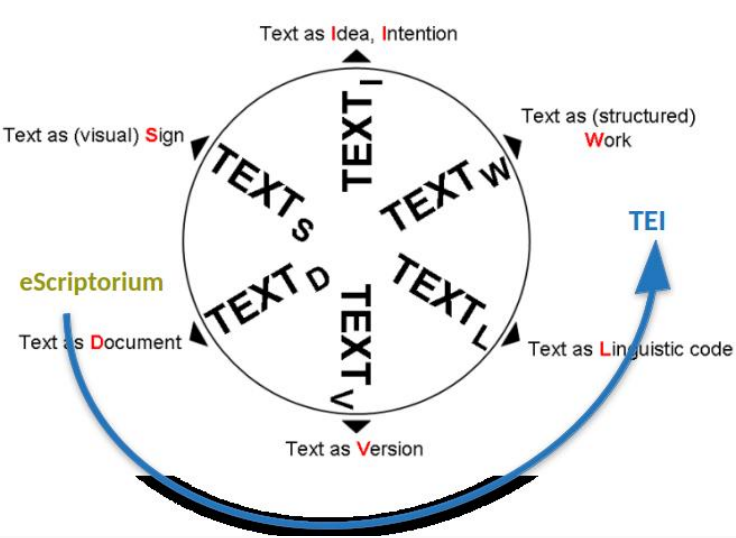
Which images were used for training?

- Is there a sufficient variety? Colour/grayscale? Can I reuse: PD, CC, restricted licence?

Which script-styles were used in training?

- What variety of scribal hands was there? Are they close enough to my scribe(s)? Which palaeographical terminology was used? Is it precise enough?

Patrick Sahle's "Text Wheel"

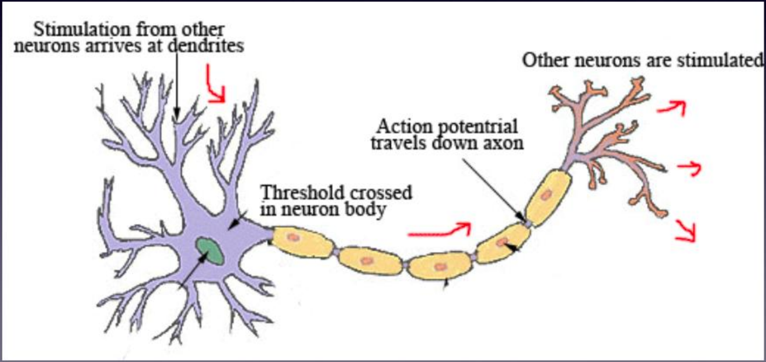


The best way to understand something is to know it so well that you can teach it to a computer … The process of seeking such explanations will surely be instructive for all concerned   Donald Knuth, ‘The Concept of a Metafont’ (1982), 5.

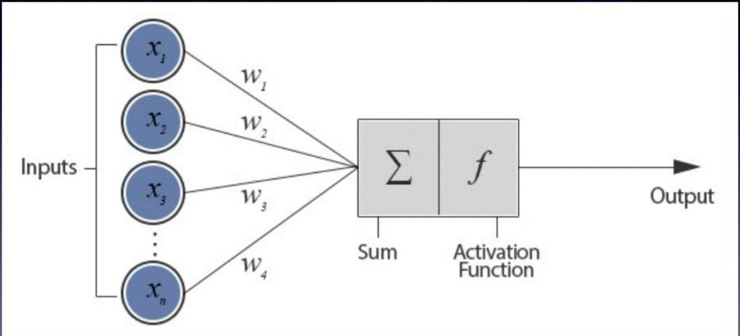
Artificial Neural Networks

Single neuron

Sum of incoming connections determines whether neuron will 'fire' (threshold)

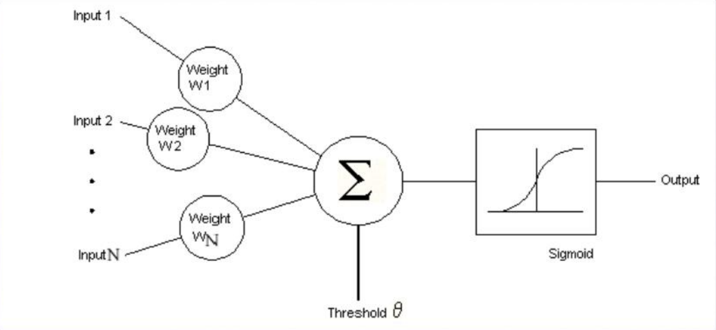


Mathematically



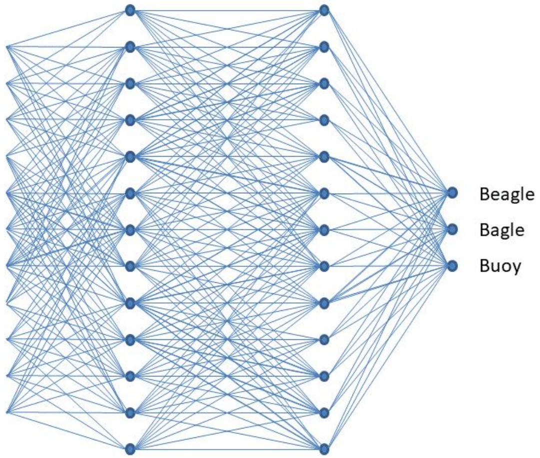
Weights control sensitivity of neuron to information

Perceptron



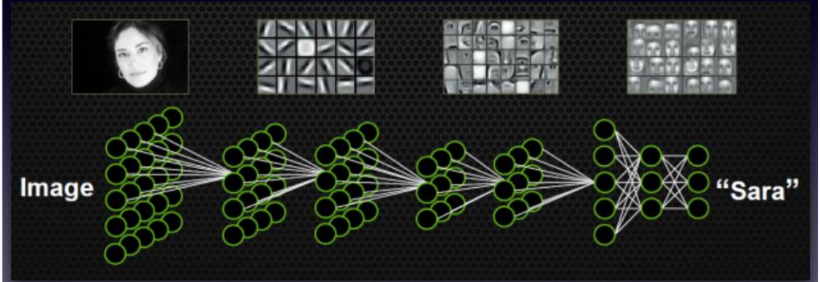
Already useful for regression (*single output*) in ML
E.g. predict house prices using location, bedrooms, …

Classification with an ANN



Computer Vision

Importance of layers



Training

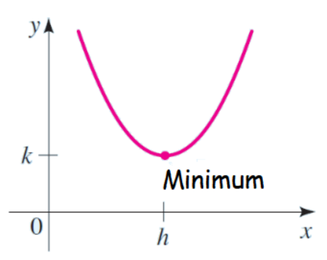
How do we avoid having to set weights ourselves?

Training

The key point is that we must find the values of all the weights (w) that produces the smallest possible error.

Optimization is a classical problem in calculus

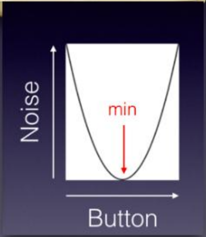
(Relatively) easy with one variable



But what about when we have millions or billions of variables??

Intuition?

- Left and right
- Movements get slower as you finetune: learning rate
- You don't know how the radio works internally: only knob and a loss estimate
- Naming conventions:
    - radio = system; knob = parameter
    - sound quality = loss function (which we want to minimize)



In neural networks?

System or function with many more knobs,
but exact same principle: one-by-one adjustments

Problem: local minima