

# Automatic Font Group Recognition in Early Printed Books



FRIEDRICH-ALEXANDER  
UNIVERSITÄT  
ERLANGEN-NÜRNBERG



Deutsche  
Forschungsgemeinschaft

JOHANNES GUTENBERG  
UNIVERSITÄT MAINZ



# Team

JGU Mainz, Buchwissenschaft  
Dr. Nikolaus Weichselbaumer  
Dr. Saskia Limbach

FAU Erlangen-Nürnberg, Pattern Recognition Lab  
Mathias Seuret M. Sc.  
Dipl.-Inf. Vincent Christlein  
Prof. Dr. Andreas Maier

# Ground Truth

Outstanding Ground Truth for books printed in the 15<sup>th</sup> century.

Two important sources:

- [Gesamtkatalog der Wiegendrucke](#) (GW)
- [Typenrepertorium der Wiegendrucke](#) (TW)

# Gesamtkatalog der Wiegendrucke

- Willkommen
- Ein Werk im Entstehen
- Anleitung
- Abkürzungen
- Hilfe
- Adressen/Kontakt
- Neuigkeiten
- Weitere Inkunabelprojekte im Internet
- ▼ Recherche
  - Allgemeine Recherche
  - Drucker: Name, Ort, Datum
  - Bibliographien
  - Zweite Lagen
  - Geographische

03100 Ave Maria, deutsch. Tübingen: [Johann Otmar, um 1498/1500]. 8°

8 Bl. [a⁸]. 26–28 Z. Typ. 7:170G, 13:80G. 1 Hlzs.

Reproduktionen: Stuttgart LB (Digitalisat).

Schr 3398a. ISTC ia01407600.

Stuttgart LB.

*Gesamtüberlieferung: 1 Exemplar/Fragment in öffentlichen Einrichtungen.*

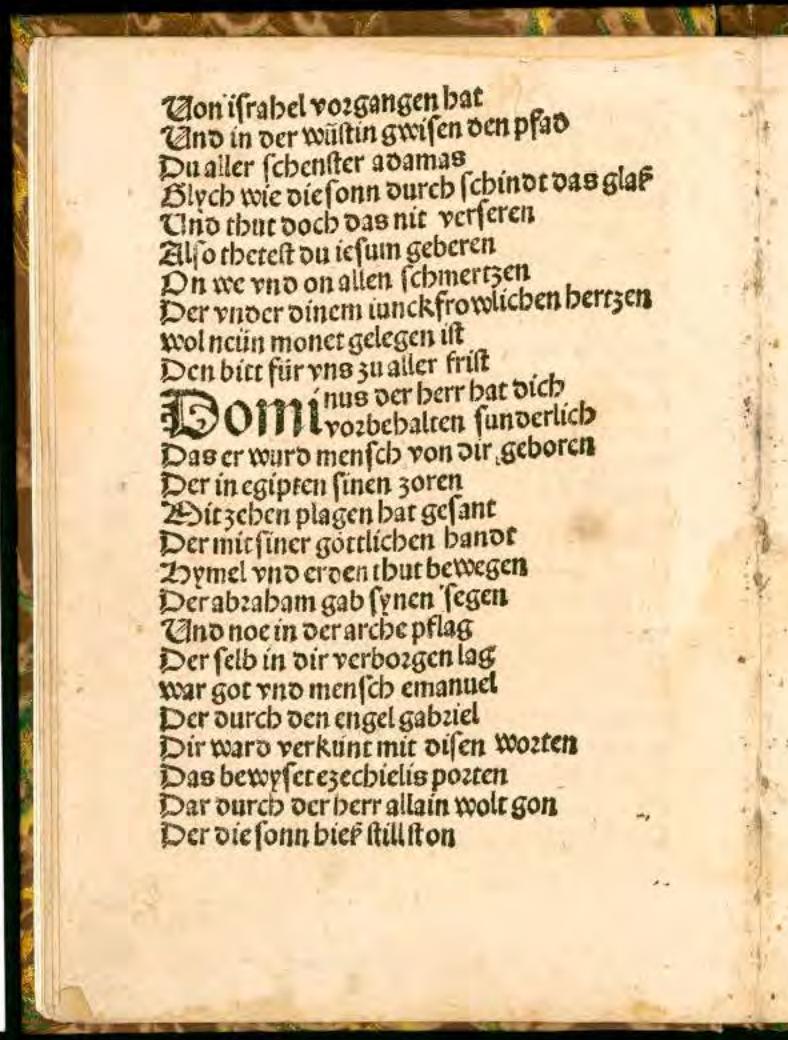
Gesamtkatalog der Wiegendrucke • <https://gesamtkatalogderwiegendrucke.de/docs/GW03100.htm> • Letzte Änderung: 2011-10-11



0  
1  
2  
3  
4  
5



Württembergische Landesbibliothek Stuttgart



<http://tw.staatsbibliothek-berlin.de>

The screenshot shows the homepage of the Typenrepertorium der Wiegendrucke (TW) website. The header features a logo of a stylized bird or letter 'W' inside a square, followed by the text "Typenrepertorium der Wiegendrucke". Below the header, there are four navigation links: "Recherche", "Register", "Dokumentation", and "Hilfe". On the right side of the header, there is a vertical sidebar with links to "Startseite der SBB", "Gesamtkatalog der Wiegendrucke", "Inkunabelreferat der SBB", and "Kontakt". The main content area contains several paragraphs of text. The first paragraph introduces the TW as a database of printing types from the 15th century, based on Haebler's work. It describes the classification system and the method used for identification. The second paragraph discusses the continuation of the work by Proctor, adding more information and material. The third paragraph provides details about the accompanying visual material. The fourth paragraph concludes with the current scope of the database.

**TW - Typenrepertorium der Wiegendrucke**

TW - Typenrepertorium der Wiegendrucke ist die durch das [Inkunabelreferat der Staatsbibliothek zu Berlin – Preußischer Kulturbesitz](#) erarbeitete Datenbank der im Buchdruck des 15. Jahrhunderts verwendeten Drucktypen.

Die Basis der Datenbank bildet das von Konrad Haebler herausgegebene "Typenrepertorium der Wiegendrucke", das zwischen 1905 und 1924 in fünf Bänden publiziert wurde. Haebler verzeichnete und beschrieb alle damals bekannten Drucktypen. Die methodische Neuerung Haeblers war eine Klassifizierung der Typen nach der Form der Majuskel M, für die er während seiner Arbeiten für das Typenrepertorium eine Übersichtstafel mit 101 verschiedenen Formen zusammenstellen konnte (vgl. die Tafel der M-Formen). Das zweite Klassifizierungsmerkmal war die auf die englischen Bibliographen Henry Bradshaw und Robert Proctor zurückgehende Bestimmung der Kegelhöhe der Typen, gemessen jeweils auf 20 Zeilen. Die Angabe von M-Form und Kegelhöhe ist bis heute das am weitesten verbreitete und bewährteste Klassifizierungsschema für Drucktypen der Inkunabelzeit. Durch die Typenbestimmung nach der Proctor-Haeblerschen Methode ist es möglich, auch für Inkunabeln ohne Impressum den Druckort, ihren Drucker oder das ungefähre Erscheinungsdatum zu bestimmen, indem die Typen mit jenen in lokalisierten bzw. firmierten und datierten Drucken verglichen werden.

Während der Arbeit am [Gesamtkatalog der Wiegendrucke](#), der bei seiner bibliographischen Arbeit die typographische Methode für die Bestimmung unfirmierter Inkunabeln anwendet, wurden und werden die Angaben aus Haeblers Typenrepertorium laufend ergänzt. Z.T. konnten weitere, Haebler noch unbekannte, Drucktypen verzeichnet werden, v.a. aber konnten die katalogisierten Typen mit genaueren Beschreibungen, Datierungen und Belegmaterial ergänzt werden.

Als ergänzendes Bildmaterial zu Haeblers Typenrepertorium wurden ab 1907 durch die "Gesellschaft für Typenkunde des Fünfzehnten Jahrhunderts" 2.460 großformatige Tafeln herausgegeben, auf denen Typenproben des 15. Jahrhunderts reproduziert und in vielen Fällen Typenalphabete aus den ausgewerteten Drucken zusammengestellt wurden.

Derzeit verzeichnet die Datenbank knapp 2000 verschiedene Druckwerkstätten (Offizinen) des 15. Jahrhunderts und das in ihnen verwendete typographische Material (Drucktypen, Initialen, Signete, Titelholzschnitte). In über 6.000 Datensätzen sind alle derzeit bekannten Drucktypen der Inkunabelzeit erfasst. Es ist hier nur mit geringfügigen Ergänzungen zu rechnen. Des Weiteren verzeichnet die Datenbank etwa 4.400 Initialen bzw. Initialalphabete, knapp 700 Drucker-



## Type 7:170G bei Johann Otmar (Tübingen, Offizin 1)

### Beschreibung

Bezeichnung (nach GW) Type 7:170G

Größe (20 Z.) 170mm

M-Form

TW-ID ma10705

Beschreibung: Andere Formen als Type 7:170G bei Johann Otmar (Reutlingen, Offizin 2)

C rund

E mit dünner Sehne außen

M mit Dorn im hinteren Teil

Divis

Nutzungszeit 1498-1500

Anmerkungen Abgrenzung zu Type 7\*:150G nicht klar! Die Formen scheinen in den Tübinger Drucken identisch zu sein. Nur über Kegelhöhe zu differenzieren, oder fehlerhafte Typenbestimmung bei GW12048?

A B C C D E F H J L M R  
D O P Q R S Z W X a a b c  
d o ð e ê f ff g h i i l m z n o ð p q  
q r s s f t u ü v x y z ð ñ



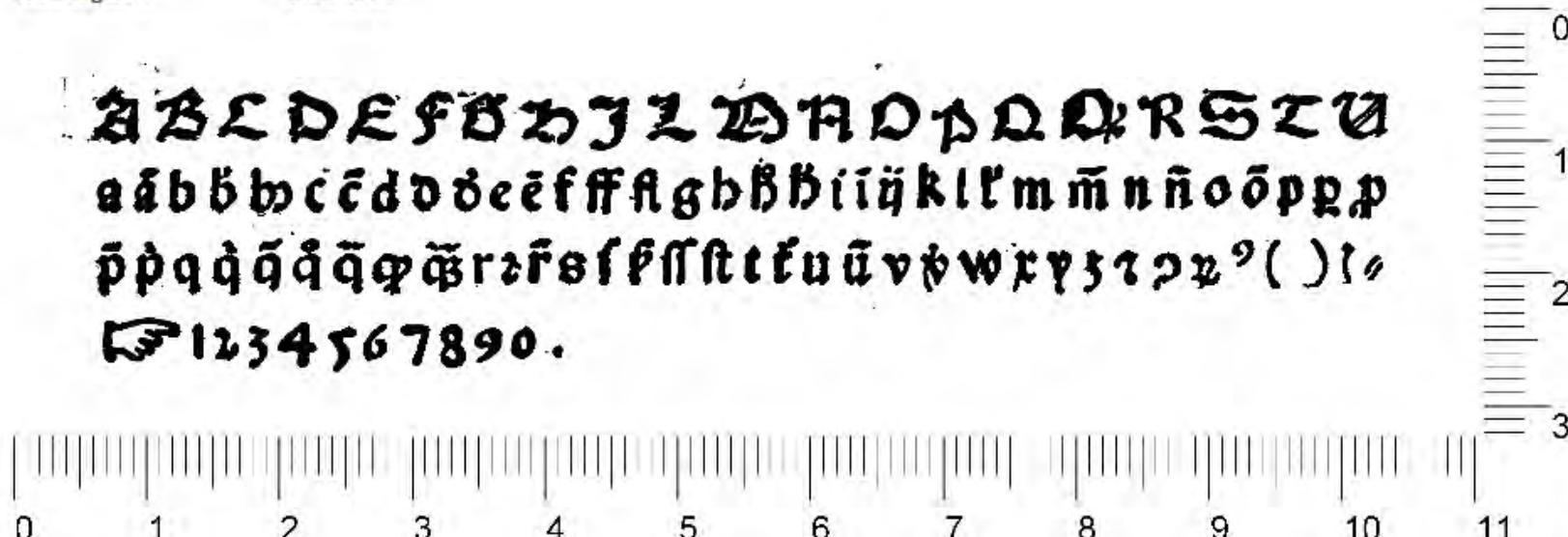
0  
1  
2  
3  
4  
5



## Type 13:80G bei Johann Otmar (Tübingen, Offizin 1)

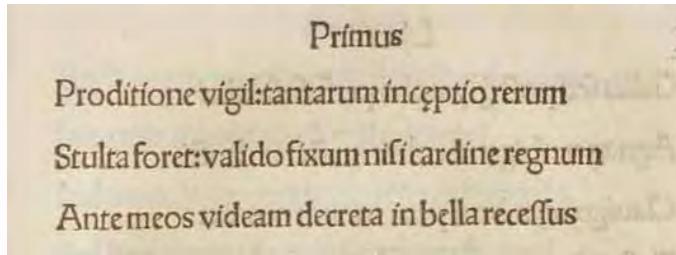
### Beschreibung

Bezeichnung (nach GW)	Type 13:80G
Größe (20 Z.)	80mm (Haebler: 79/80)
M-Form	ℳ M47
TW-ID	ma06965
Beschreibung	
Divis	
Nutzungszeit	1498-1500



# Classification

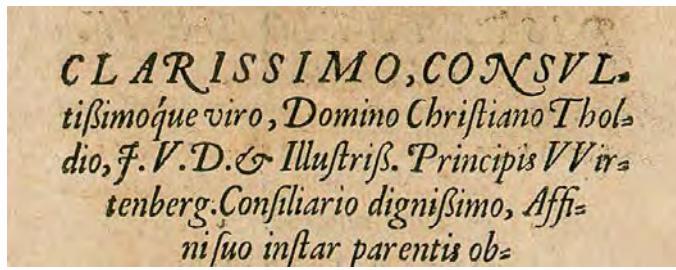
## Roman type



### Antiqua

Characteristics: very legible, letters are thin and not connected

Use: first in Italy in the 15th century, chiefly used for editions of classical authors, spread to Northern Europe, especially Basel and Strasbourg



### Italic

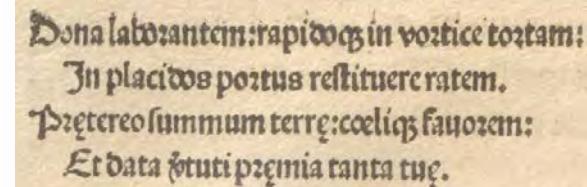
Characteristics: very legible, letters are thin and slope forward

Use: this type group spread from the 16th century onwards to many parts of Europe, often for Latin texts

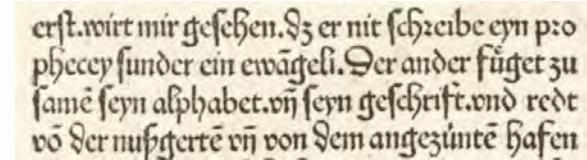
# Gothic type



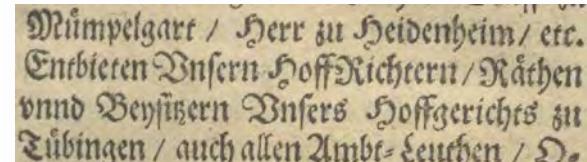
Chye endet sich die histori Eusebi von dem  
grossen künig Alexander. Als dy e der hpc  
gelert doctor Johann hartlieb zu München.  
durch liebe des durchleuchtigen fürsten Her



Dona laborantem: rapidoqz in vortice tortam:  
In placidus portus restituere ratem.  
Prætereosumnum terre: coeliqz favorem:  
Et data stuti præmia tanta tue.



erst. wirt mir geschen. Dz er nit schreibe eyn p:o  
phecey sunder ein ewägeli. Der ander füget zu  
samē seyn alphabet. vñ seyn geschrift. vnd redt  
vñ der nußgerte vñ von dem angezün̄te hafen



Mümpelgart / Herr zu Heidenheim / etc.  
Entbieten Unsern Hoff Richtern / Räthen  
vñnd Beyfzern Unser Hoffgerichts zu  
Tübingen / auch allen Ambs- Leuten / D.

## Textura

Characteristics: letters are thick, upright and narrow, strong vertical and thinner horizontal lines

Use: Northern Europe, especially the Holy Roman Empire, the Netherlands and England

## Rotunda

Characteristics: Thick curved letters. The letters b,h,l,p,q are short (no ascenders/decenders)

Use: in Southern Europe, especially Spain and Italy (where it was used for judicial texts, especially in the fifteenth century)

## Bastarda

Characteristics: strong regional variations, many of which have flourished ascenders for the letters b,d, h, k, l and v. The middle part of the letters f und s are thicker, the letter x looks like a p and the a is a single-decker.

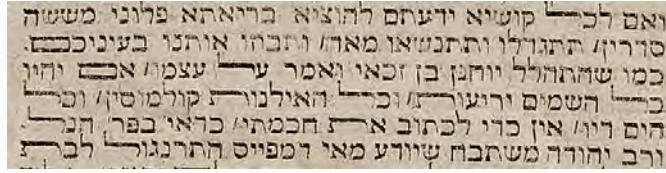
Use: France, Burgundian provinces of Netherlands, England and especially Germany

## Fraktur

Characteristics: many majuscles have a trunk, the letter a with only a single-decker

Use: since the early sixteenth century in Germany, becomes very popular until the twentieth century

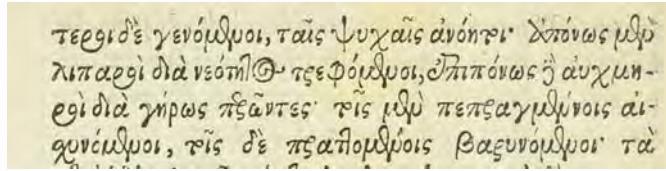
# Other type groups



## Hebrew

Characteristics: variations in square and semi-cursive letters

Use: at the beginning in Italy, Spain, Portugal and Turkey

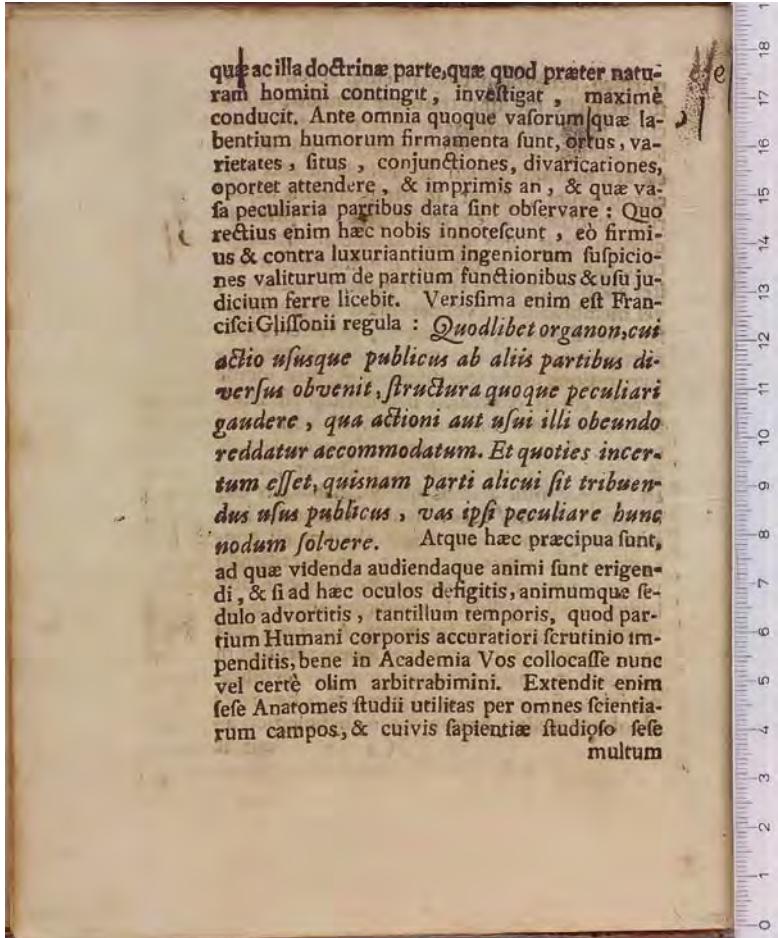


## Greek

Characteristics: at the beginning square, straight and with separate letters. During the early modern period letters more connected and slope forward

Use: in the 15th century mostly in Italy, from the 16th century all over Europe

# Classification



Input

{

"antiqua" : 0.7,  
"italic" : 0.4,  
"fraktur" : 0.1,  
"bastarda" : 0.2,

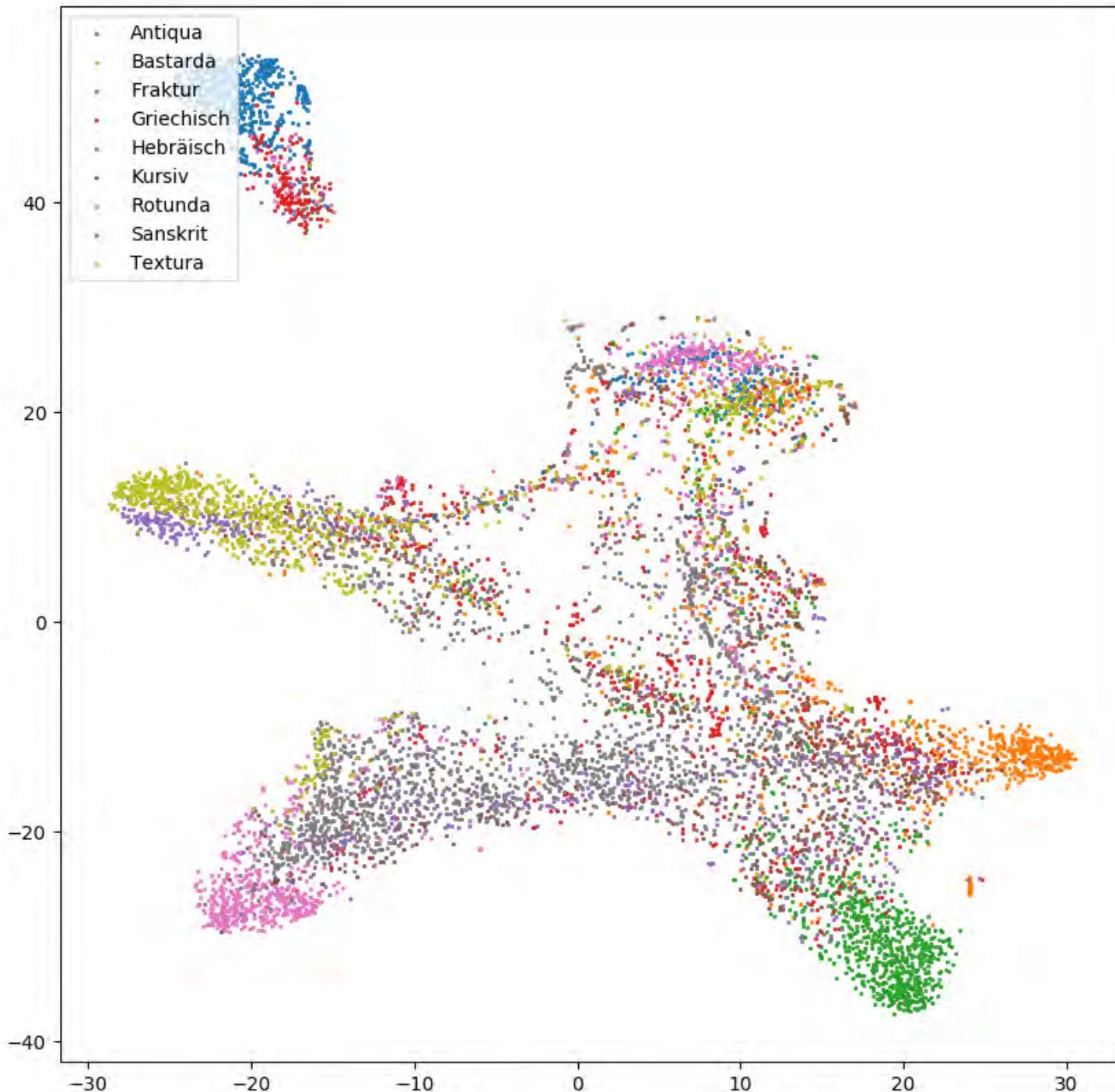
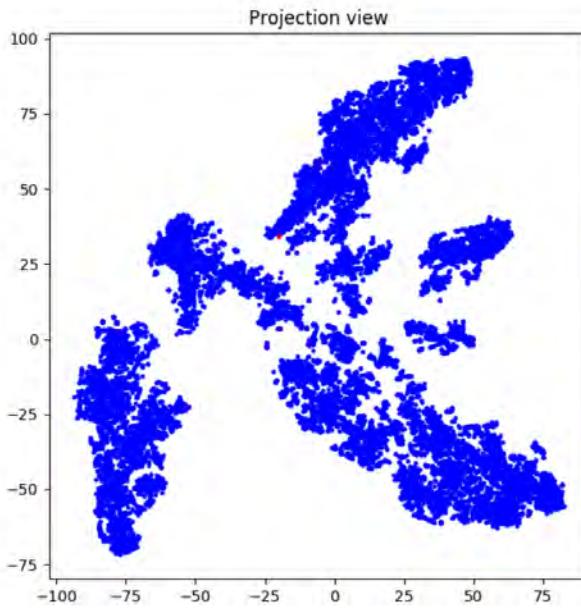
...

}

Output

# Clustering

- High-dimensional feature vector
- Goal : similar vectors for similar type groups

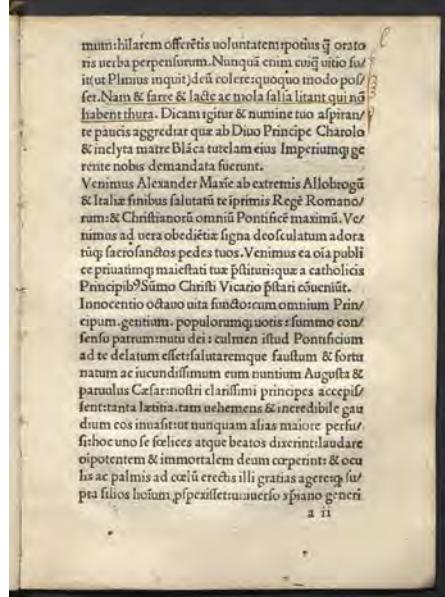


# Neural Network Approach

Neural networks return scores

Errors are computed – networks learn to decrease them

Tests : independent data (e.g., different books)

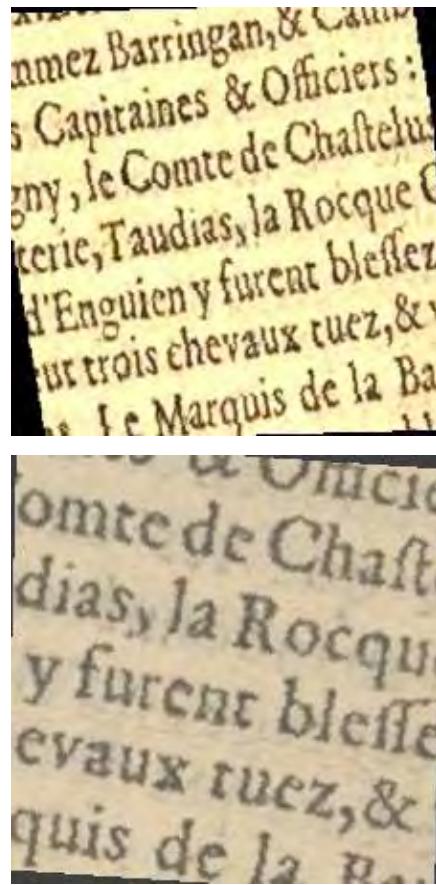


Result:	Wanted:	Error:
Antiqua 0.7	1	-0.3
Kursiva 0.4	0	+0.4
Fraktur 0.1	0	+0.1
...	...	...

# Training a Neural Network

Much data is needed – augmentations can help

bié son devoir: Le Marquis  
eux: Les Lieutenás Colonel  
ommez Barrinan, & Camb  
es Capitaines & Officiers :  
gny, le Comte de Chastelus  
terie, Taudias, la Rocque C  
d'Enguien y furent blessez;  
ut trois chevaux tuez, & v  
as. Le Marquis de la Bauli  
turent leurs chevaux blesse  
fauables sur eux, & la G



- Rotations
- Recaling
- Aspect ratio modification
- Shearing
- Color
- Luminosity
- Contrast
- JPEG compression

# Training a Neural Network

Much data is needed – augmentations can help

bié son devoir: Le Marquis  
eux: Les Lieutenās Colonel  
ommez Barrinan, & Camb  
es Capitaines & Officiers :  
gny, le Comte de Chastelus  
terie, Taudias, la Rocque C  
d'Enguien y furent blessez;  
ut trois chevaux tuez, & v  
as. Le Marquis de la Baul  
turent leurs chevaux blesse  
favorables sur eux. Le G

nnez Barrinan, & Camb  
es Capitaines & Officiers :  
gny, le Comte de Chastelus  
terie, Taudias, la Rocque C  
d'Enguien y furent blessez;  
ut trois chevaux tuez, & v  
as. Le Marquis de la Baul

ez Barrinan, & Camb  
es Capitaines & Officiers :  
gny, le Comte de Chastelus  
terie, Taudias, la Rocque C  
d'Enguien y furent blessez;  
ut trois chevaux tuez, & v  
as. Le Marquis de la Baul

tracvoir: Le Mar  
es Lieutenās Co  
ez Barrinan, & Camb  
es Capitaines & Offici  
e Comte de Chas  
Taudias, la Roc  
guien y furent ble  
ois chevaux tuez  
e Marquis de la  
ent leurs chevaux

pitaines & Offici  
e Comte de Chas  
Taudias, la Roc  
guien y furent ble  
ois chevaux tuez  
e Marquis de la  
ent leurs chevaux

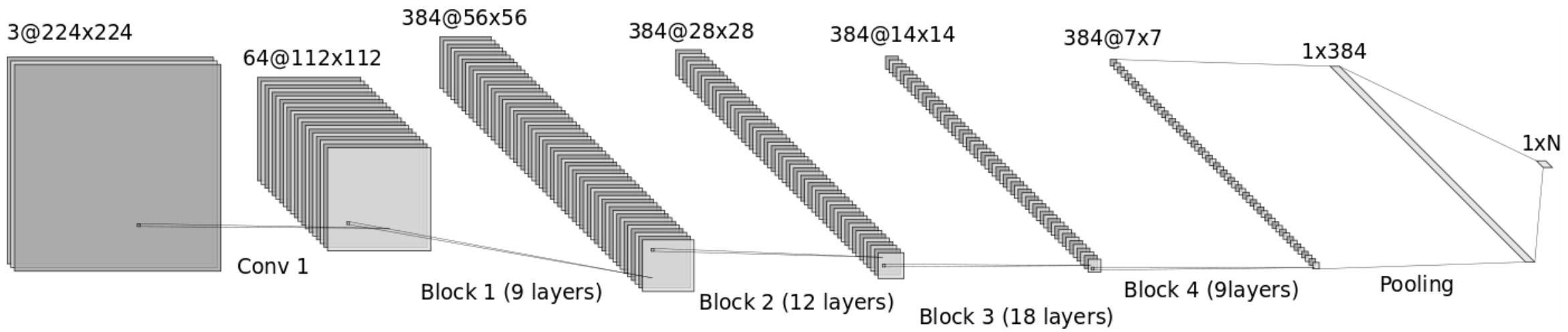
devoir: Le Marquis  
omte de Chast  
dias, la Rocqu  
y furent blesse  
chevaux tuez, &  
quis de la Baul

ez Barrinan, & Camb  
es Capitaines & Offici  
e Comte de Chas  
Taudias, la Roc  
guien y furent ble  
ois chevaux tuez  
e Marquis de la

ez Barrinan, & Camb  
es Capitaines & Offici  
e Comte de Chas  
Taudias, la Roc  
guien y furent ble  
ois chevaux tuez  
e Marquis de la

ez Barrinan, & Camb  
es Capitaines & Offici  
e Comte de Chas  
Taudias, la Roc  
guien y furent ble  
ois chevaux tuez  
e Marquis de la

# Classifier

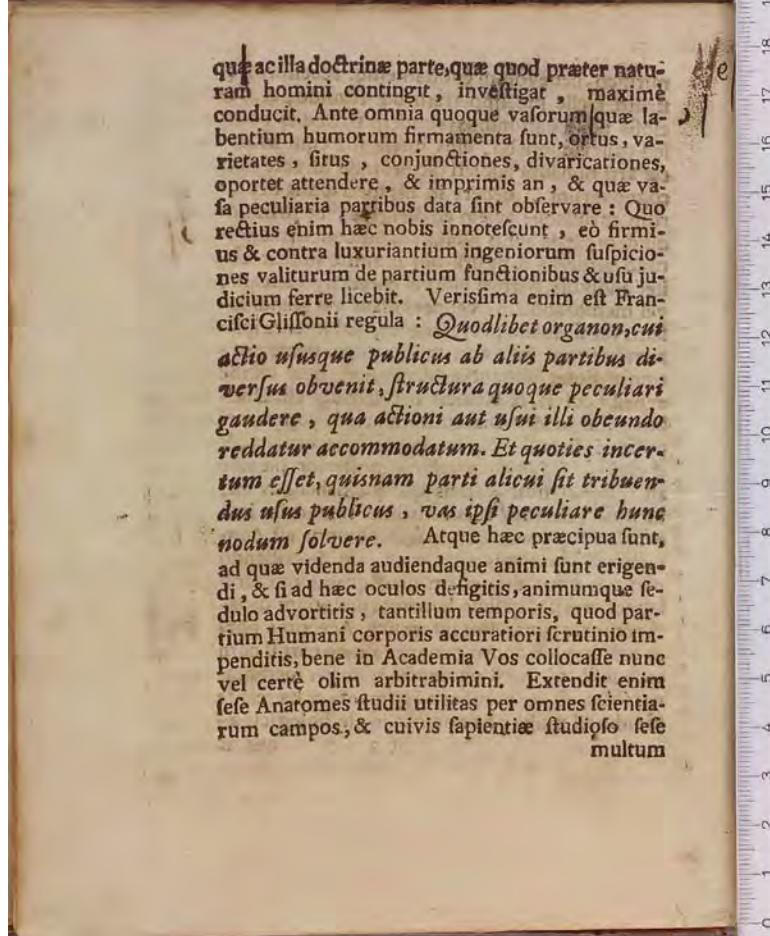


Adaptation of ResNet 50

Classifies image crops of 224x224 pixels

Average page result (random crops / grid-like pattern)

# Classification Granularity

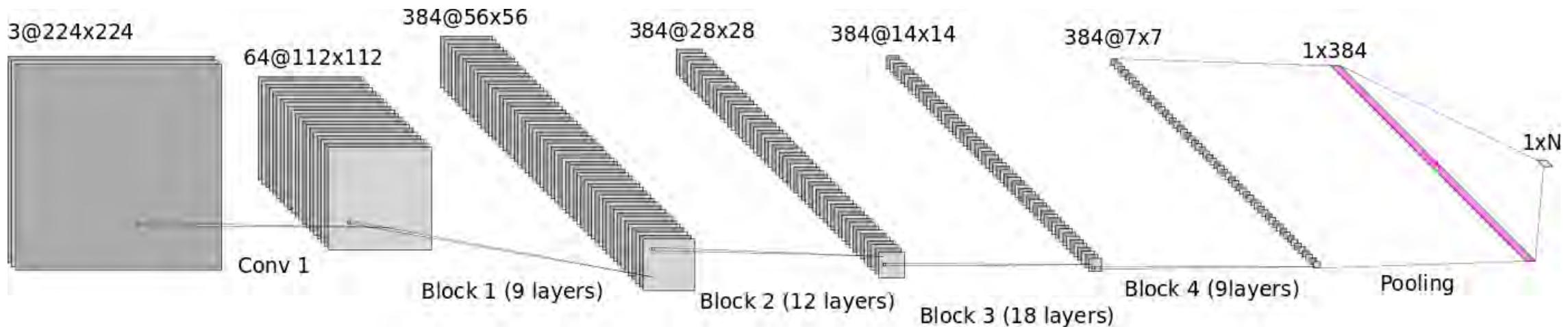


Input

Antiqua

Italic

# Feature Vector Extraction



Layers → levels of abstractions

Penultimate layer → high-level embedding

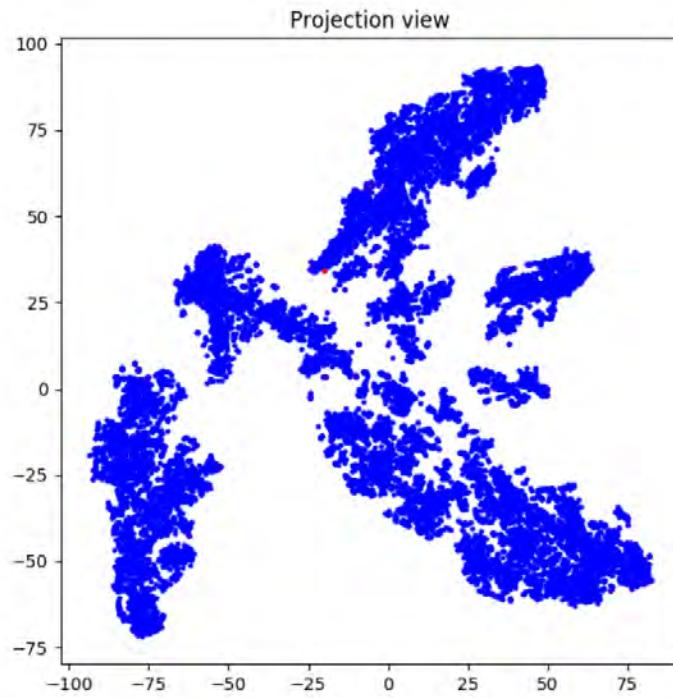
High-dimensional space → 2D / 3D projections

More information:

<https://github.com/seuretm/typegroups-classification-projection>

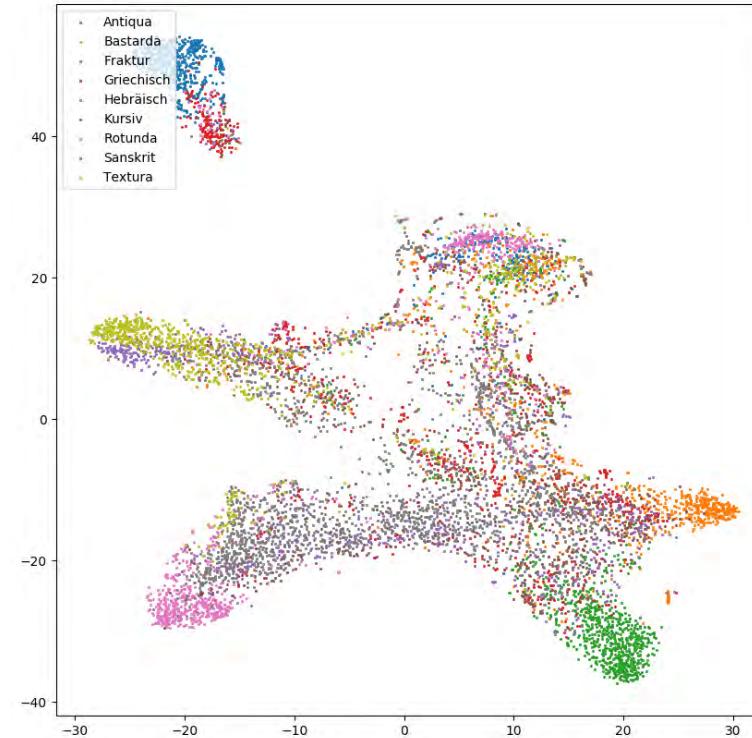
# Clustering

Similar data – similar location  
Projection : t-SNE, UMAP

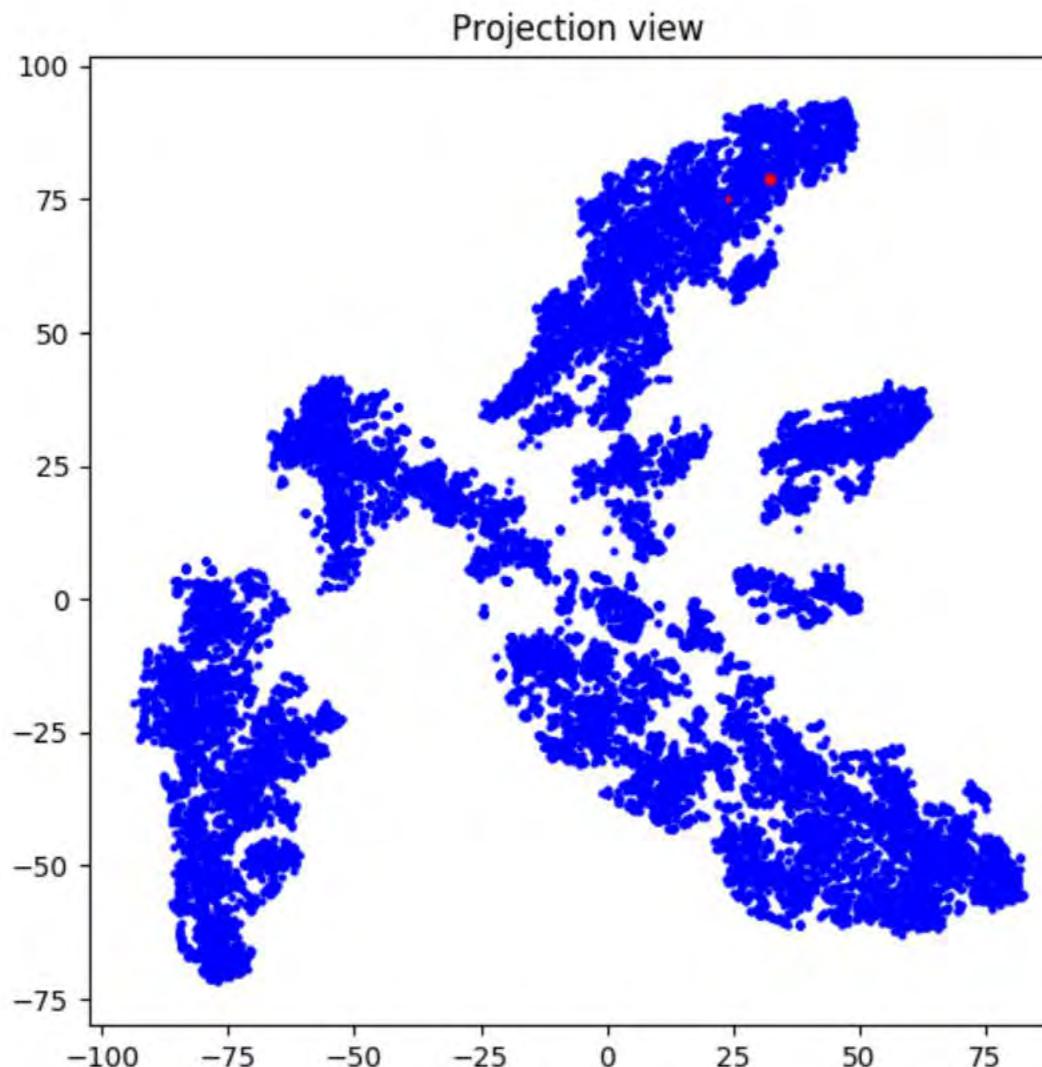


18k images  
Unknown type group

Data exploration  
Outliers spotting

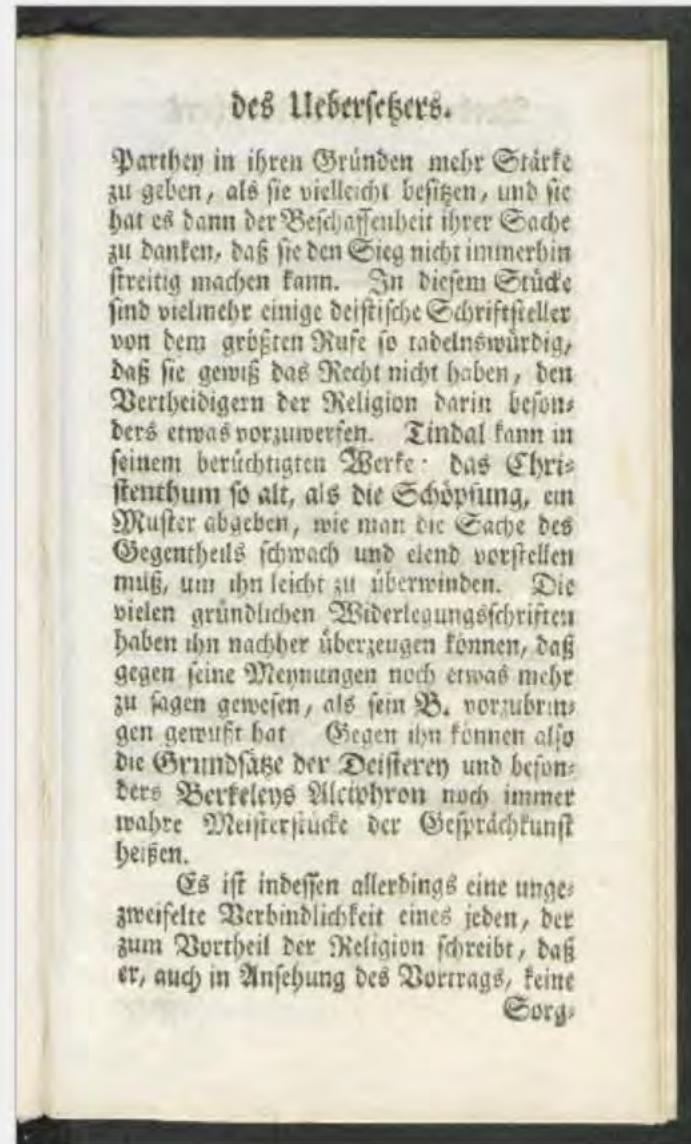


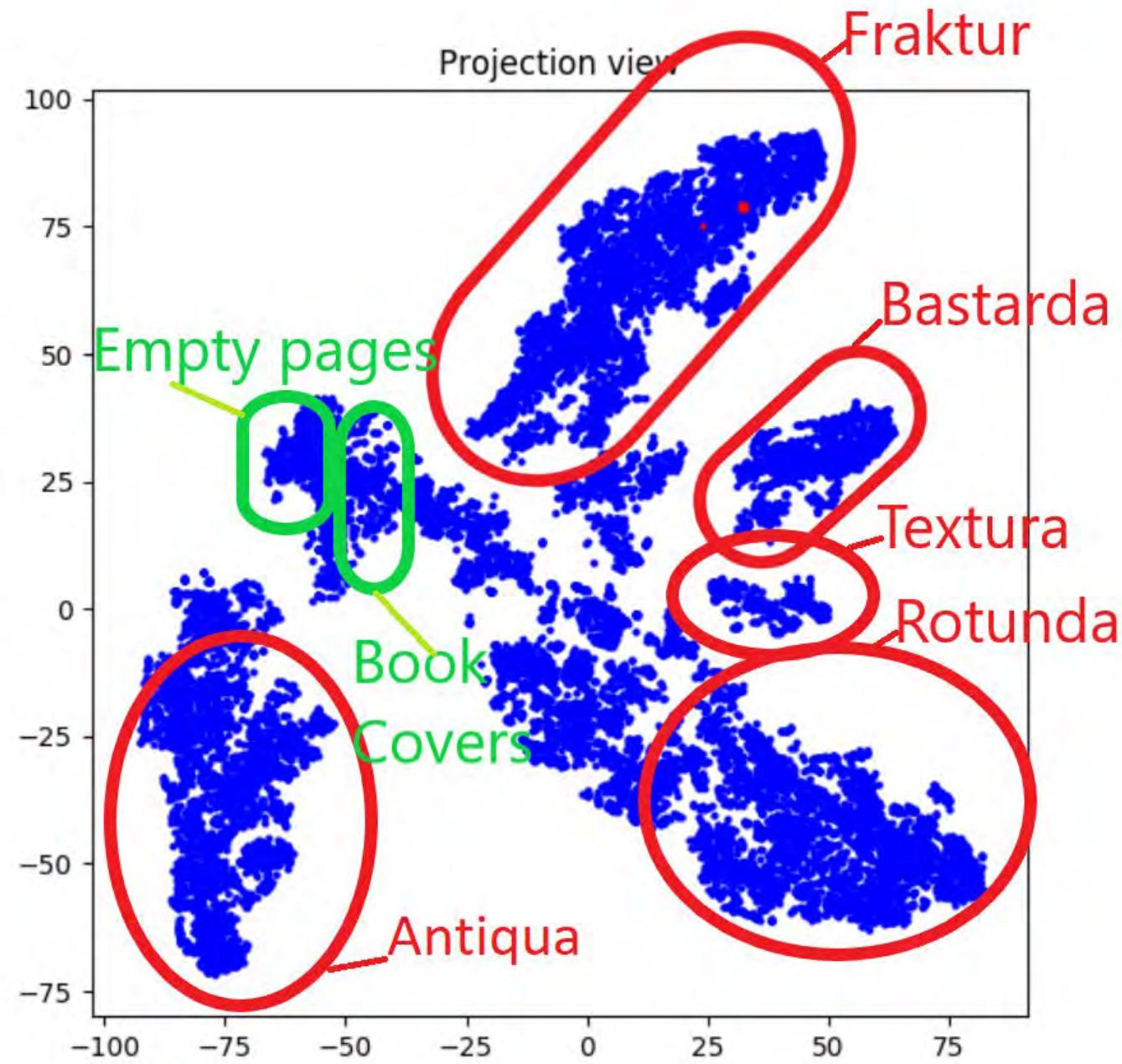
10k random crops  
Known type group

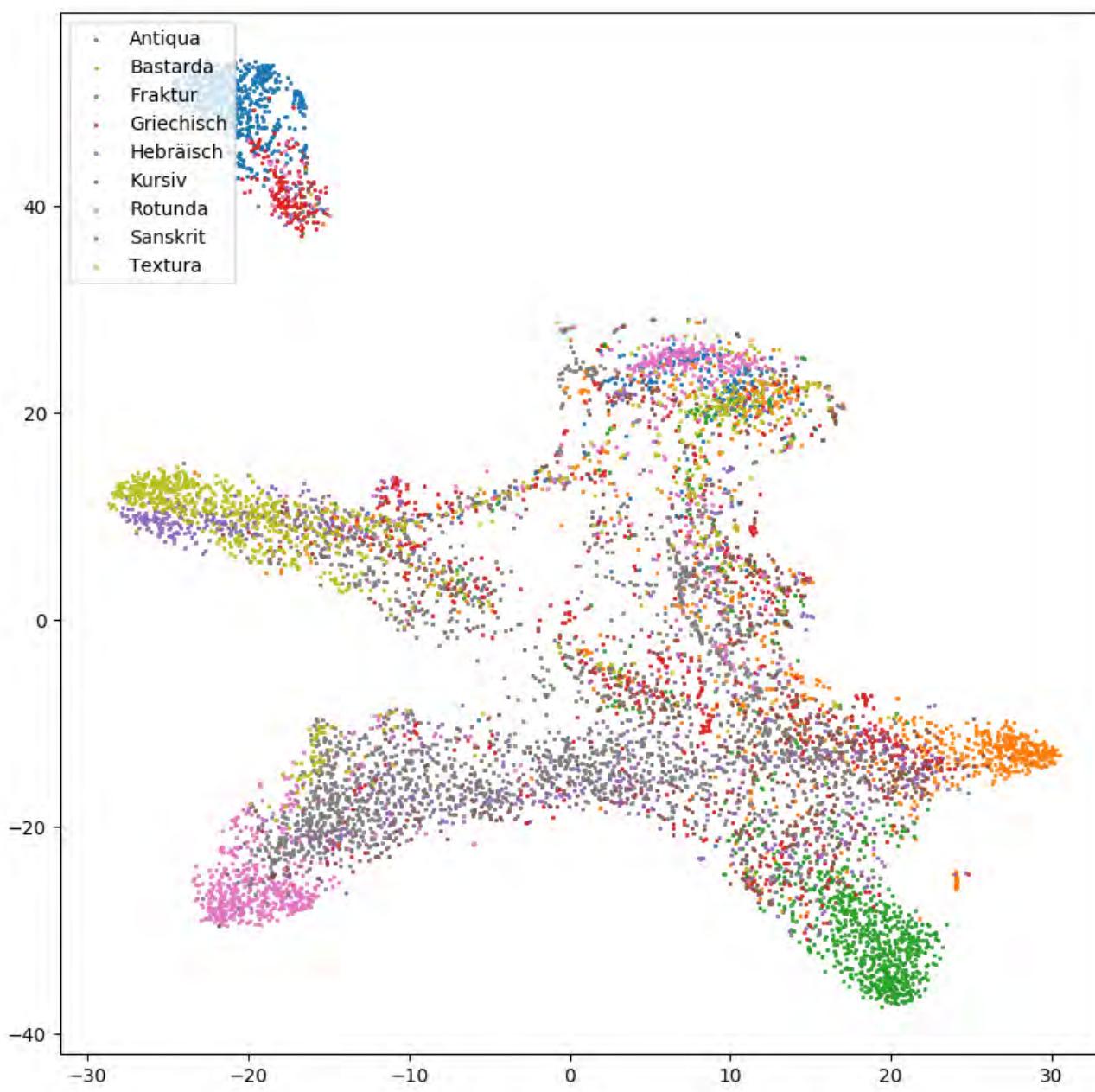


<https://content.staatsbibliothek-berlin.de/dc/672231506-0011/full/full/0/default.jpg>

Fit to screen







# Current & Future Work

Higher recognition accuracy

More information:

Higher granularity

<https://github.com/seuretm/typesgroups-classification-projection>

Creation & publication of multi-class dataset

Further development toward type identification

(of course: open source, non-contaminating license)