TINAJO 2022
27 de octubre
October 27

Organiza InterPARES Trust AI · Ayuntamiento de Tinajo

Colabora Cabildo de Lanzarote · Ajuntament de Girona

InterPARES TrustAI

AI in the Middle Age:
Parchment Study and Arrangement by Appearance-Based Recognition

ARCHIVIO DI STATO MILANO

unimc UNIVERSITÀ DI MACERATA
l'umanesimo che innova

Prof. Pierluigi Feliciati
Prof. Emanuele Frontoni
University of Macerata

TEAM

- **Archivio di stato di Milano**:

Benedetto Luigi Compagnoni - *Director*;

Carmela Santoro - *Archivist, Deputy director*;

Antonella Cesarini - *Archivist*;

Davide Bruno De Franco - *Archivist, Director Archivio di Stato di Novara*.

- **University of Macerata & Marche Polytechnic University:**

Emanuele Frontoni, *Professor - Computer Science and Computer Vision*

Pierluigi Feliciati - *Professor, Records and Information Science*

Marina Paolanti - *PhD, Assistant professor*

Andrea Felicetti - *PhD, Postdoc*

- **University of Pavia**:

Luca Fois - *PhD, Postdoc, Dipartimento di studi umanistici*

- **Washington University**:

Joseph T. Tennis - *Professor, Associate Dean for Faculty Affairs, Executive Director of Administrative Services*

InterPARES TrustAI

INTRODUCTION

- **Ancient written private documents on parchment** were an important communication mean to humankind and have, for that motive, an **invaluable historical value to our culture heritage** and at the same time they are **authentic deeds among people**, valid till nowadays.

- Therefore, it becomes imperative that they are **preserved** and perpetuated in order to extend their life span in the interests of population and **future generations**.

- Currently, **digitization of historical parchments is extraordinarily convenient as it allows easy access to the documents** from remote positions and removes the need for possible adverse physical management.

InterPARES TrustAI

The beginning of the most ancient original document preserved in Italian State Archives: the *mundium* of a Longobard woman called Anstruda dated 13 may 721
(ASMi, *Museo diplomatico,* b. 1, n.. 3.*)*

Two examples of papal bull and imperial charter

AN OUTSTANDING HERITAGE

- The **Archivio di Stato of Milan** (ASMi) preserves more than **130.000** documents on parchment, covering nearly a thousand years of history.
- The main collections of parchment documents are the **Museo diplomatico and the Pergamene per fondi**.
- The Museo diplomatico holds the most ancient Italian original private deed, **dating back to the year 721** and a large number of documents preceding the year 1000.
- The Pergamene fondi holds **imperial and royal charters, papal bulls, chancery documents, and private deeds** coming from a huge variety of ecclesiastical and secular institutions of northern Italy (12th-18th centuries).

InterPARES TrustAI

(S) Anno dominice incarnationis mille uno centesimo vigesimo tercio mense aprilis indictione prima. Nos in Dei nomine Ambrosius et Arialdus germani infantes filii quondam Petri et Gisla mater et filii atque Lanfrancus infans filius quondam Adami et Dorada mater eius et Bellunda avia eorum et relicta quondam Ambrosii omnes de loco Uduluco qui professi sumus lege vivere Longobardorum consentientibus nobis qui supra feminis predictis infantibus mundualdis nostris per consensum Marchisii et Obizonis germanorum tutorum eorum infantium atque per datam licentiam predictis infantibus ab Alberto iudice et misso domni regis [illegible] Dominus omnipotens ac redemptor noster animas quas condidit ad studium salutis semper invitat et ideo nos qui supra Ambrosius et Arialdus et Lanfrancus infantes et Gisla et Dorada atque Bellunda volumus et iudicamus seu per hoc nostrum inviolabile iudicatum confirmamus ut campus iuris nostri quem habere visi sumus in isto loco et fundo Uduluco ad locum ubi dicitur in Ronnano coheret ei a mane sancte Marie a meridie sancti Stefani a sero Iohannis a monte sancti Protasii et est ipse campus per mensura iusta pertice quinque omnis ipse campus cum finibus et accessionibus suis in integrum presenti die et hora deveniat in iure et proprietate monasterii sancti Ambrosii ubi eius sanctum conditum est corpus ita tamen ut nos qui supra infantes et nostri heredes teneamus ipsum campum ad laborandum et ad nostram utilitatem facienda ita tamen ut nos et nostri heredes solvamus atque deinde omni anno in festivitate sancti Martini fictum ad ipsum monasterium de formento modium unum ad mensuram ipsius loci et abbas ipsius monasterii et eius successores debent dare gustare omni qui ipsum fictum conduxerit panem de formento et carnem et vinum et hoc iudicamus ut si nos aut nostri heredes non dederimus ipsum fictum omni anno ad ipsum monasterium ut supra legitur tunc ipsum monasterium habeat et teneat ipsum campum quamdiu dederimus ipsum fictum ad ipsum monasterium ut supra legitur et si pars ipsius monasterii fecerit superimpositam de ipso campo nobis aut nostris heredibus tunc non persolvamus ipsum fictum donec pars ipsius monasterii non reliquerit nobis ipsam superimpositam et nostris heredibus et postquam dimiserit nobis ipsam superimpositam tunc persolvamus ipsum fictum ad ipsum monasterium ut supra legitur quia sic decrevit nostra bona voluntas. Unde due carte iudicati uno tenore scripte sunt. Actum in isto loco.

Signa manuum istorum Ambrosii et Arialdi germanorum et Lanfranci infantium et Gisle et Dorade atque Bellunde qui hoc iudicatum ut supra fieri rogaverunt et ipsi infantes predictis feminis consenserunt ut supra.

(S) Albertus iudex et missus domni regis predictis infantibus licentiam dedit ut supra et subscripsi.

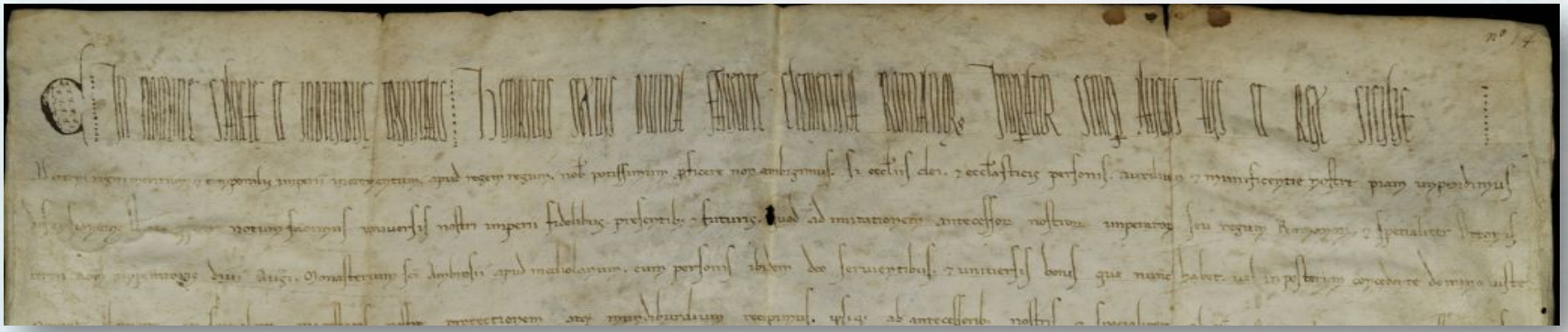
Signa manuum istorum Marchisii et Obizonis germanorum qui predictis infantibus ut tutores consenserunt ut supra et in hac carta ad confirmandum manus posuerunt.

Signa manuum Anselmi Arialdi Lanfranci Petri Arnaldi Iohannis Pagani Lanfranci testium.

(S) Ego Albertus notarius sacri palacii scripsi post traditam complevi et dedi.

NEW PATHS OF RESEARCH

- Though studied and edited since the eighteenth century, **this extensive material has not been investigated systematically in order to identify common features** recurrent in homogeneous groups of documents.
- **Processing automatically a large number of scanned documents** can lead to new understandings of general cross-cutting issues, not related to a single fund.



InterPARES TrustAI

AI & DEEP LEARNING FOR THE PARCHMENT DOCUMENTS

- One of these issues is **defining the number and the activity of the notaries in the city of Milan** and its contado.
- The AI deep learning approach can **automatically investigate a large number of private documents** and give a substantial contribution to the knowledge of **Milanese notaries in the Middle Ages**.
- The basic feature chosen to be identified for the research is **the signum, an authentication element** put by notaries.
- The **signum appears in about all private deeds and has specific and easily recognisable charateristics** (shape, position, etc.)

InterPARES TrustAI

THE SIGNUM: A USEFUL MARKER

- The **signum or Notarial sign** is a specific and personal drawn mark used by a single notary in the protocollo and before his signature.
- Identifying the signum means that every notary could be **recognised and tracked** in a virtually infinite series of documents.
- The AI will contribute either in create a **library of signa: a virtual matricula of notaries** and a basis to investigate their **less visible features.**




InterPARES TrustAI

PERGANET: A LIGHTWEIGHT DL-BASED SYSTEM

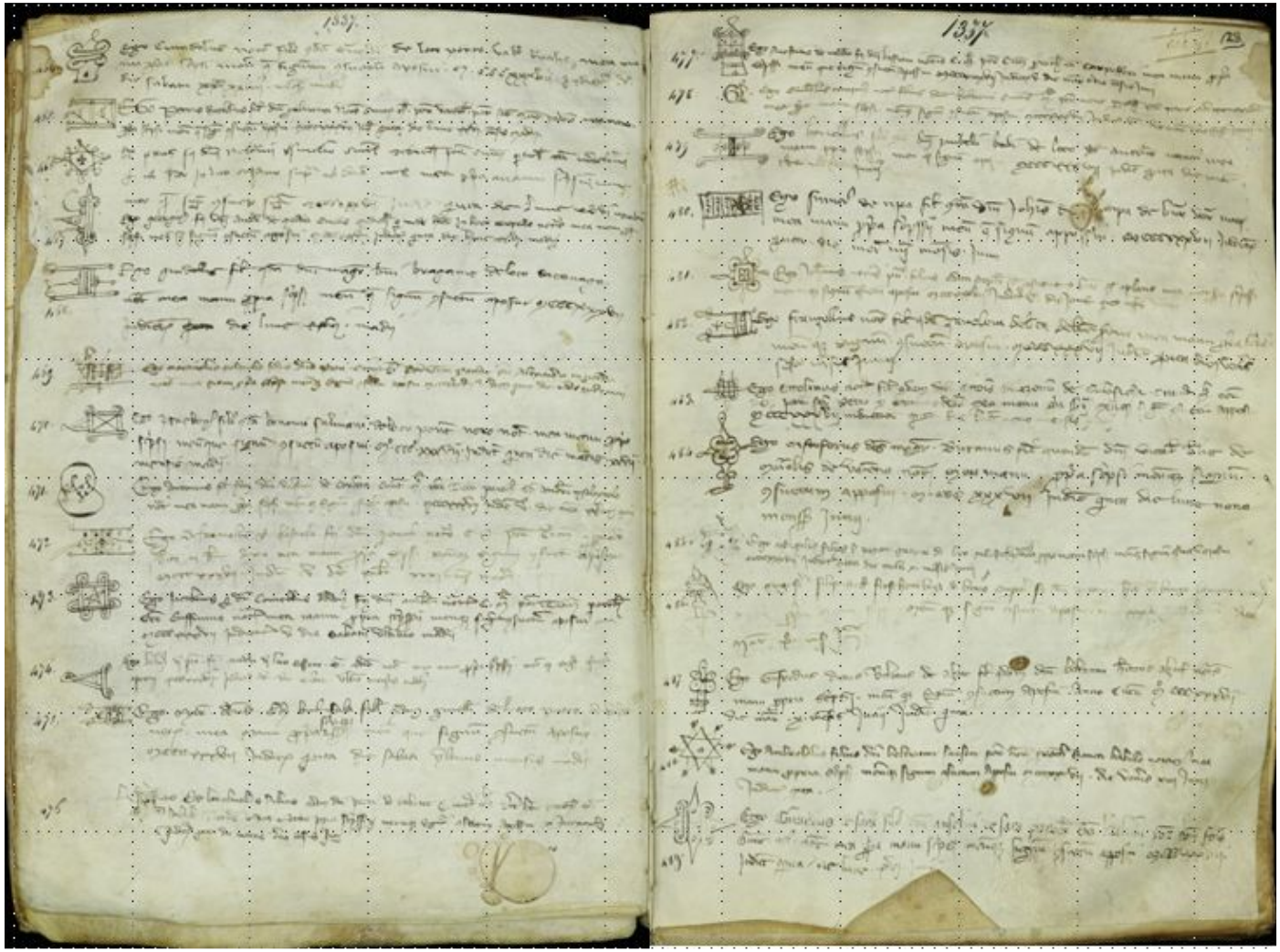
- The AI will be tested for an appearance-based recognition of ancient parchments and documents, implementing the use of AI in the archival science: a method that could be reproduced by many other Archives and for different types of documents.
- It is proposed PergaNet a lightweight DL-based system for historical reconstructions of ancient parchments.
- The aim of PergaNet is the automatic analysis and processing of huge amount of scanned parchments.

InterPARES TrustAI

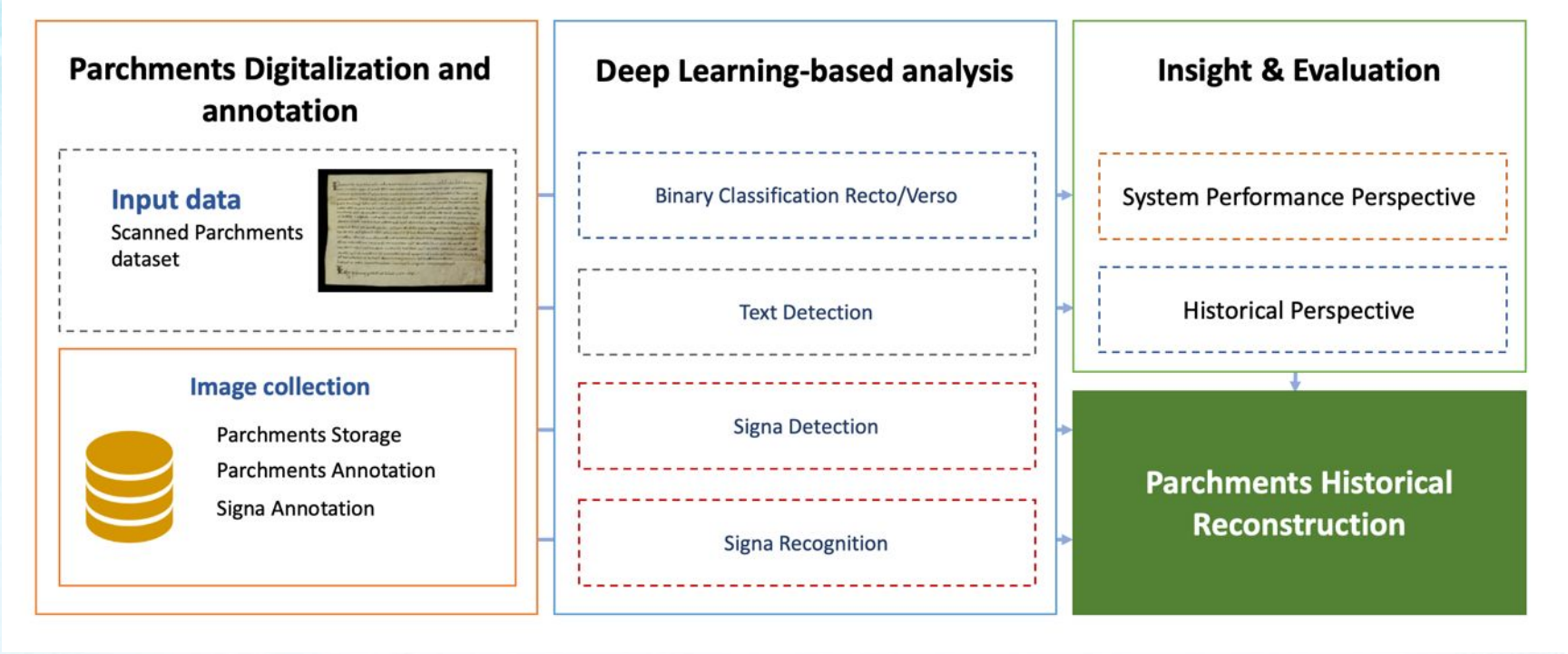
BUILDING UP A NOTARIAL MATRICULA BEFORE AD 1350

- One of the aims of the project is to use the **AI to build up a Matricula of Milanese notaries** of the twelfth and thirteenth centuries (in separate steps).
- The progressive investigation of all the parchment documents of the ASMi will lead to a **complete list of notaries and their identification**.
- The process will also generate **a list of all the surviving documents related to every single notary**.

BUILDING UP A NOTARIAL MATRICULA BEFORE AD 1350



PERGANET



InterPARES TrustAI

DATASET COLLECTION

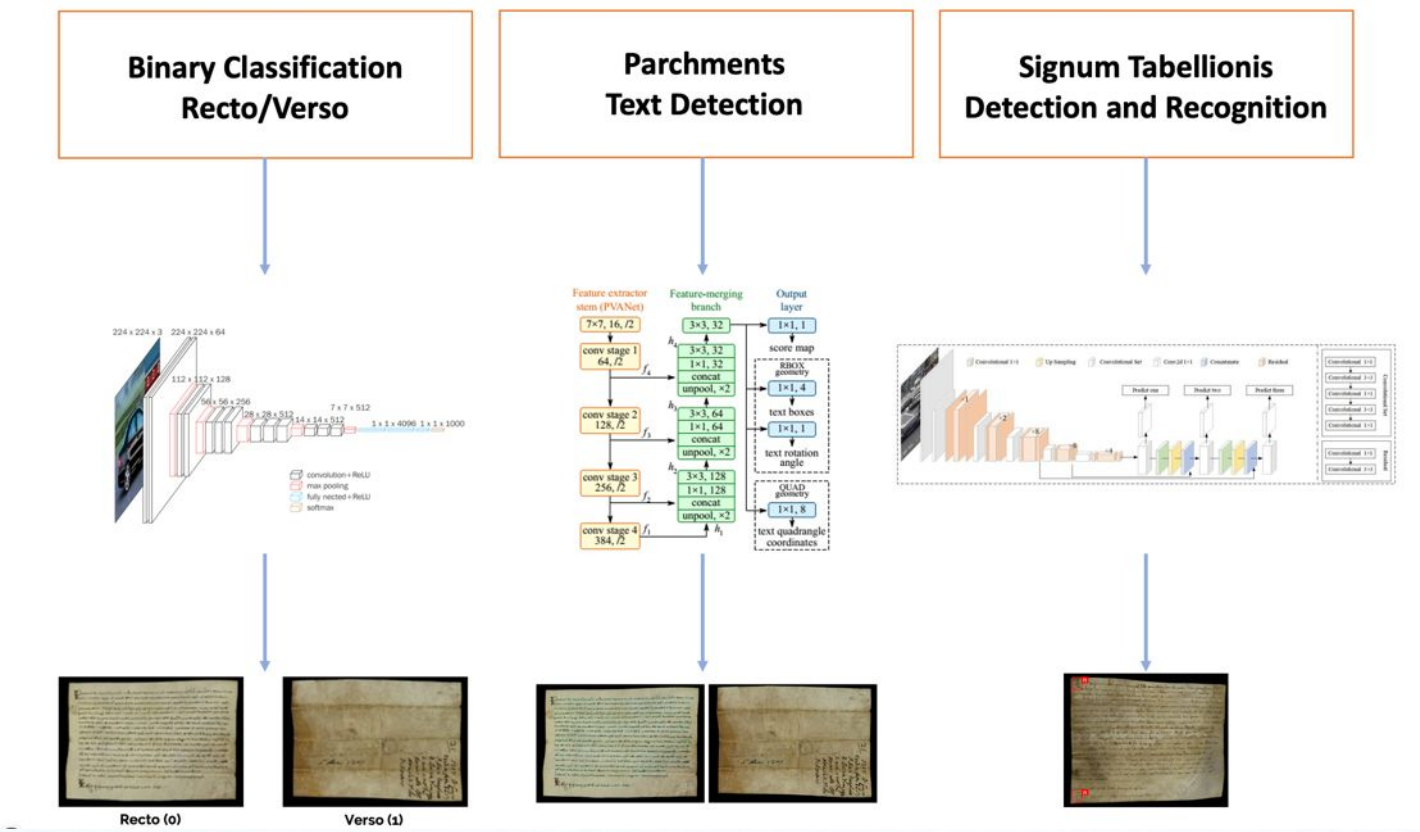
- Data collection is moving forward and actually we have the following **archives cooperation together and collecting data for the study**:
  - **Archivio di Stato di Milano**
    (220 documents already processed, nearly 1200 scanned and ready to be processed)
  - **Archivio di Stato di Novara**
    (91 documents ready to be processed)
  - **Archivio di Stato di Ascoli Piceno**
    (27 documents ready to be processed)

InterPARES TrustAI

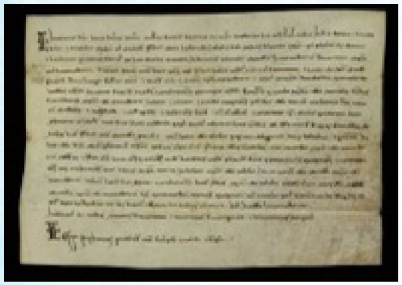
PARCHMENTS CLASSIFICATION & SEGMENTATION
https://drive.google.com/drive/folders/1lj8lj_VTr4wTtgyBHlGO9TmKK0qJ-xxV

InterPARES TrustAI

DEEP LEARNING PIPELINE



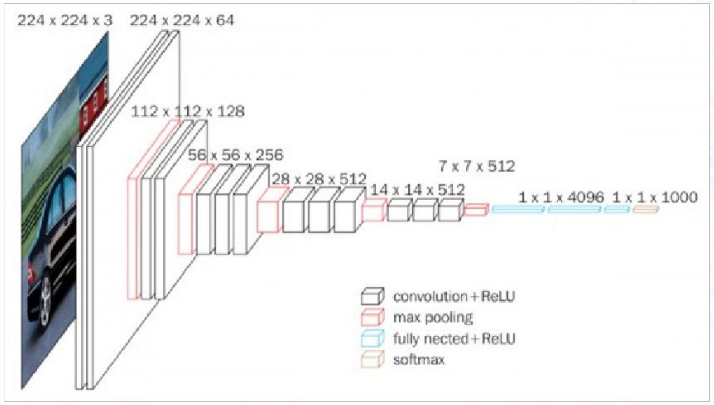
BINARY CLASSIFICATION RECTO/VERSO



Recto

Verso

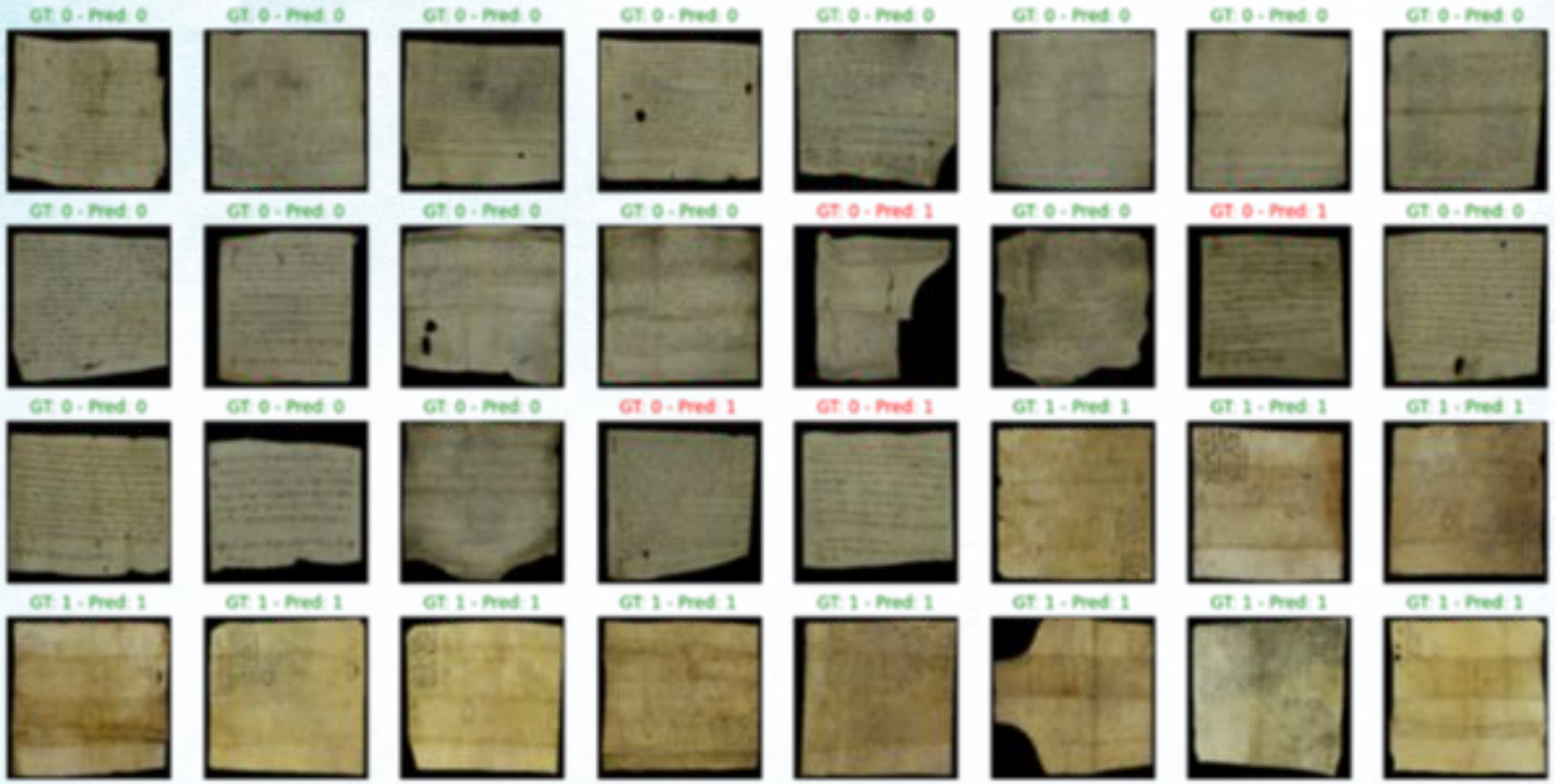
VGG16 DEEP NEURAL NETWORK



Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

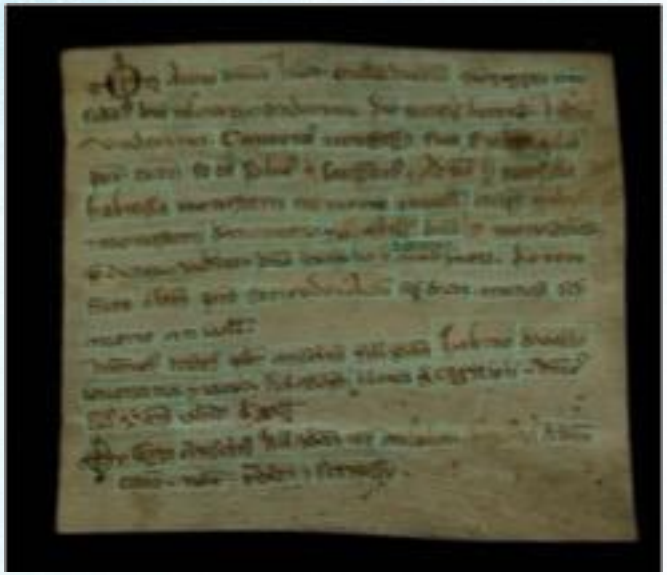
InterPARES TrustAI

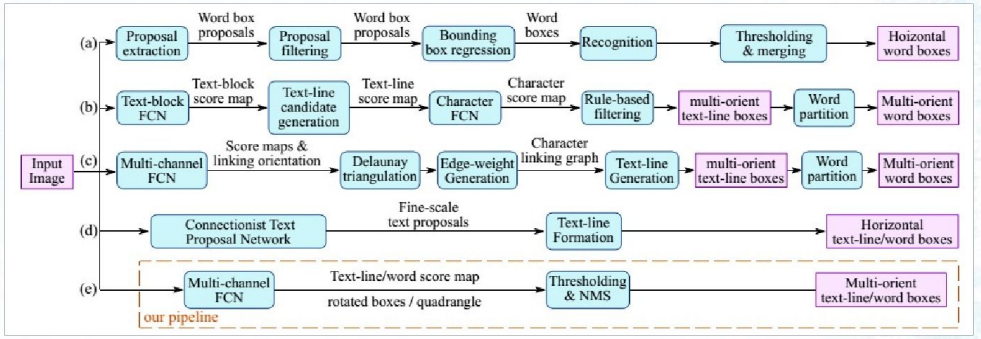
BINARY CLASSIFICATION RECTO/VERSO: RESULTS



InterPARES TrustAI

PARCHMENTS TEXT DETECTION





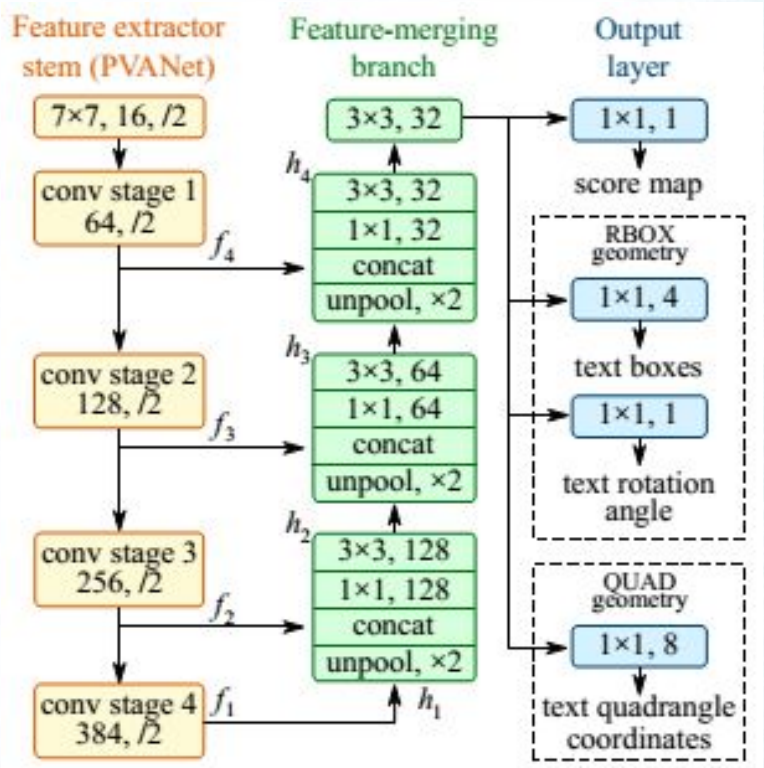
Xinyu Zhou., et all, A. (2017). EAST: An Efficient and Accurate Scene Text Detector. *arXiv:1704.03155v2*.

InterPARES TrustAI

PARCHMENTS TEXT DETECTION: EXPERIMENTS

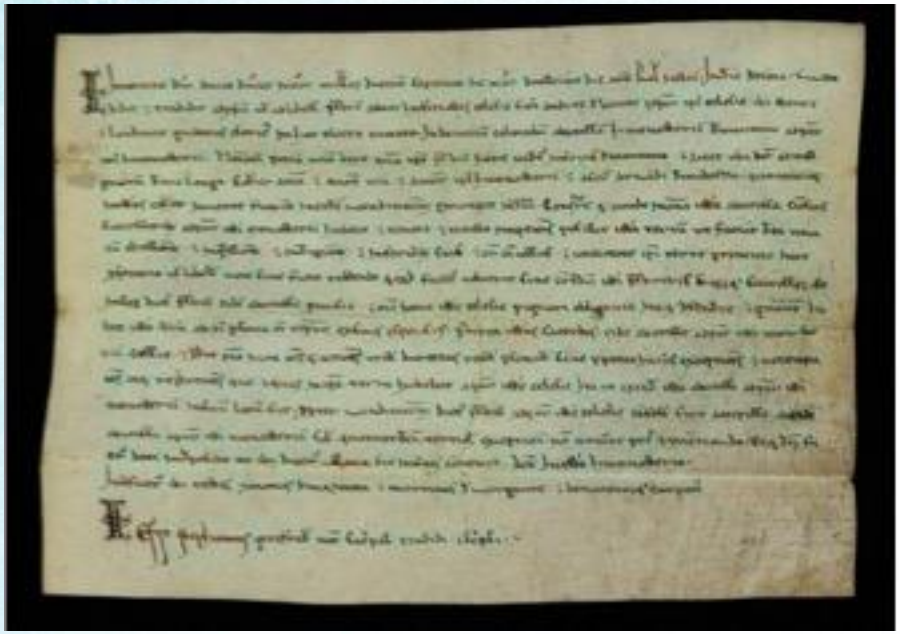
Word detection

- DNN Model: EAST
- Use of trained model (tested over datasets: ICDAR 2015, MSRA-TD500, COCO-Text)

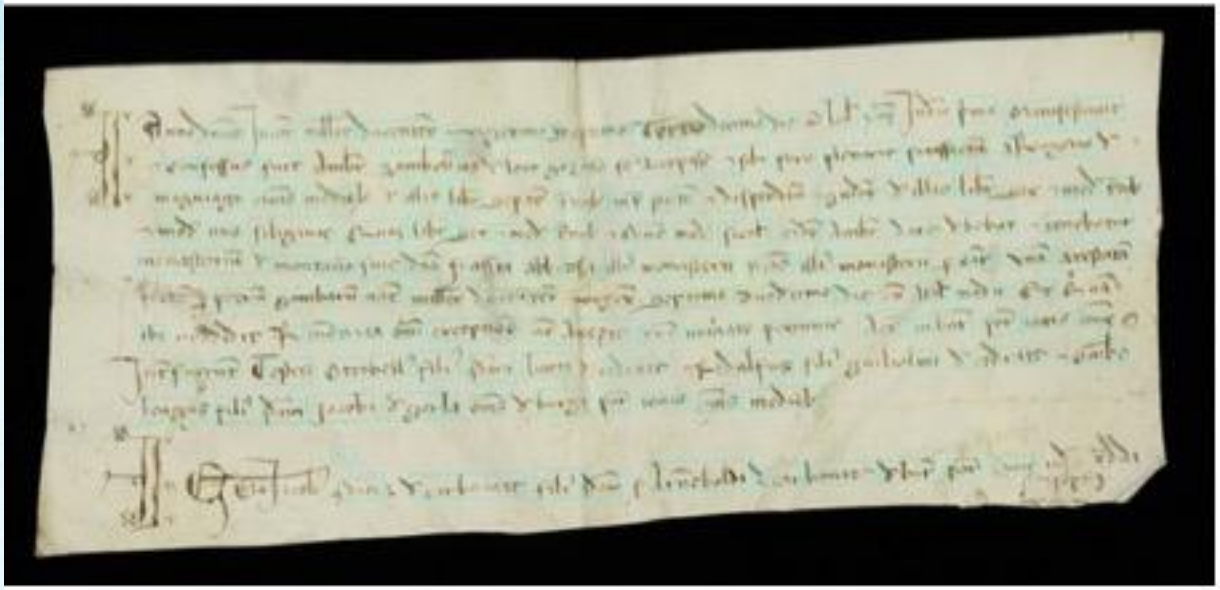


InterPARES TrustAI

PARCHMENTS TEXT DETECTION: RESULTS

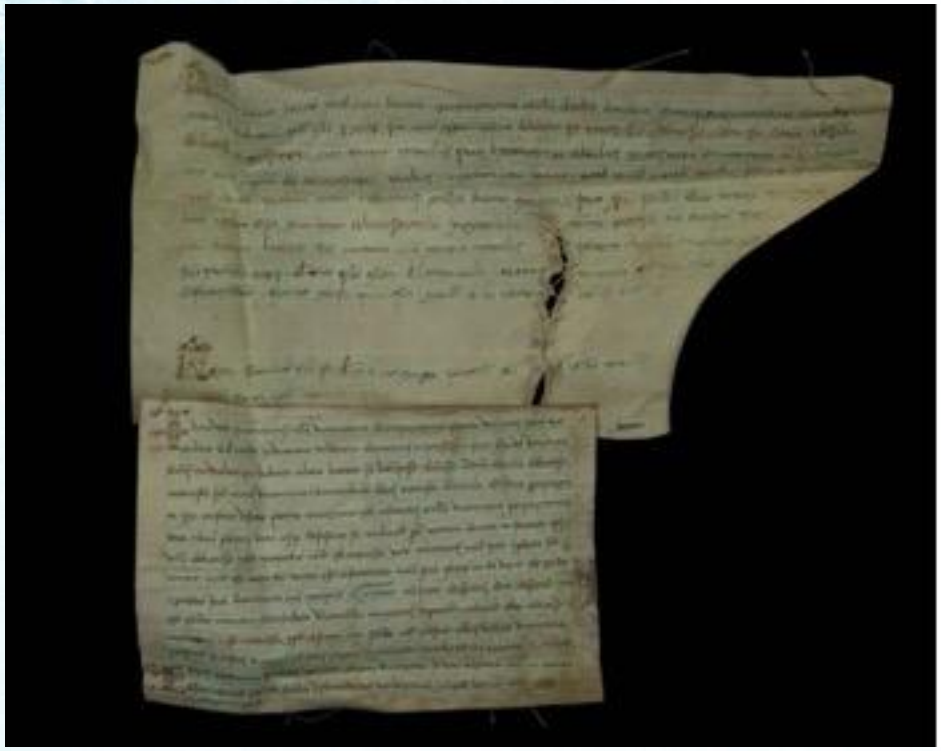


PARCHMENTS TEXT DETECTION:



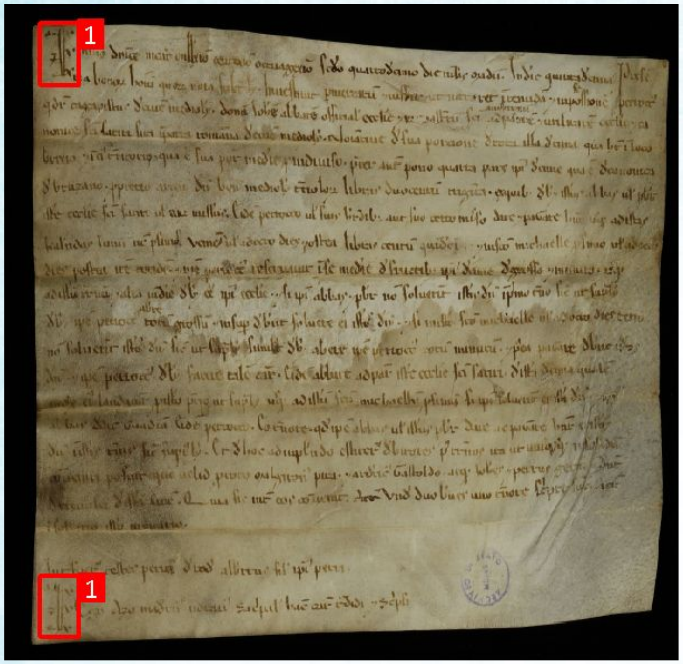
InterPARES TrustAI

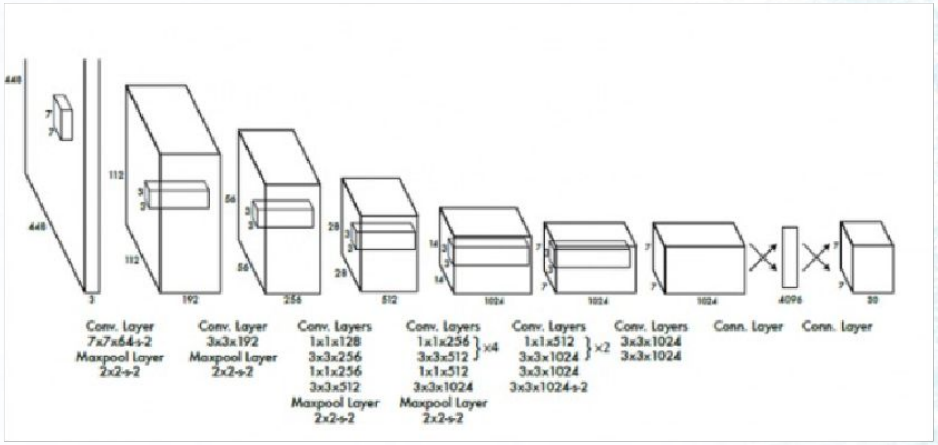
PARCHMENTS TEXT DETECTION: RESULTS



InterPARES TrustAI

SIGNA TABELLIONIS DETECTION AND RECOGNITION





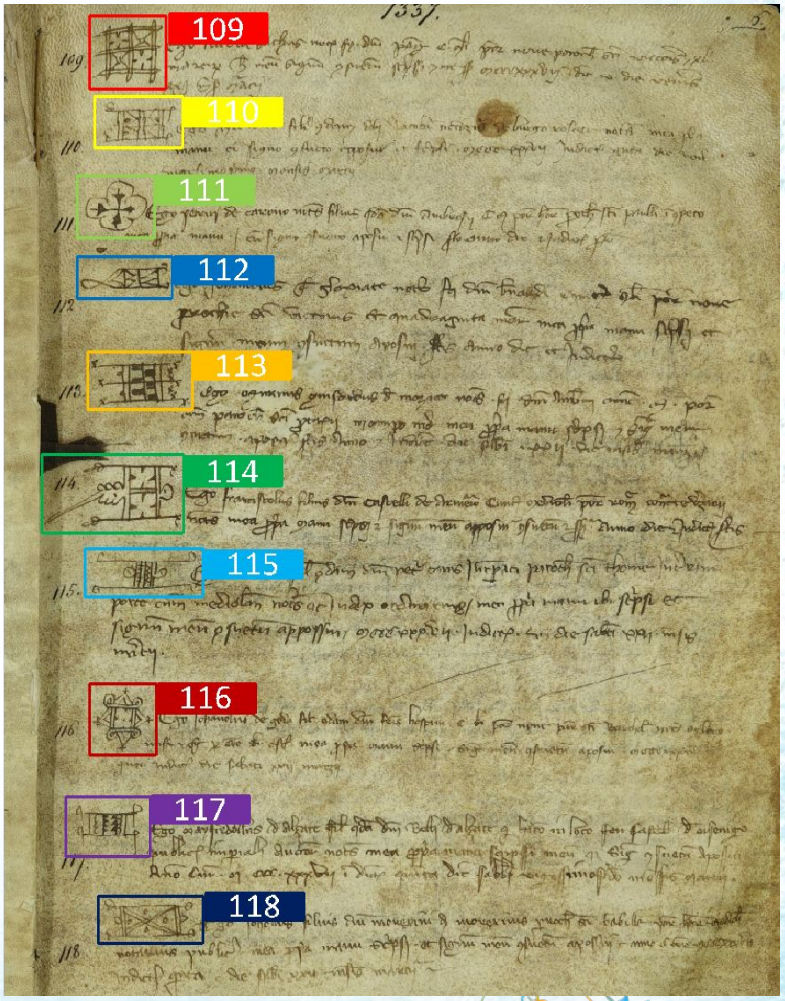
Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 779-788).

InterPARES TrustAI

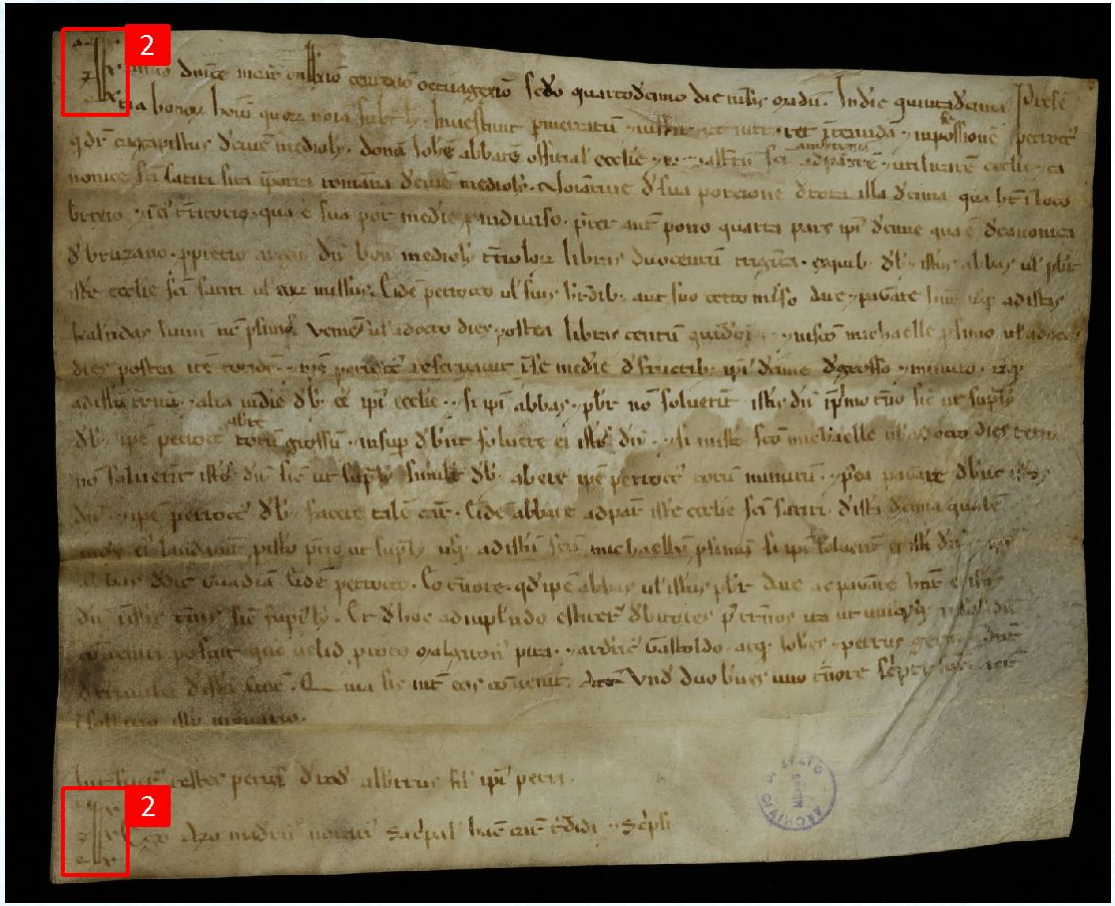
NOTARIAL MATRICULA ANNOTATION

2768 signa

ID	NAME	HISTORICAL INFORMATION			
…	…				
109	notary 109				
110	notary 110				
…	…				

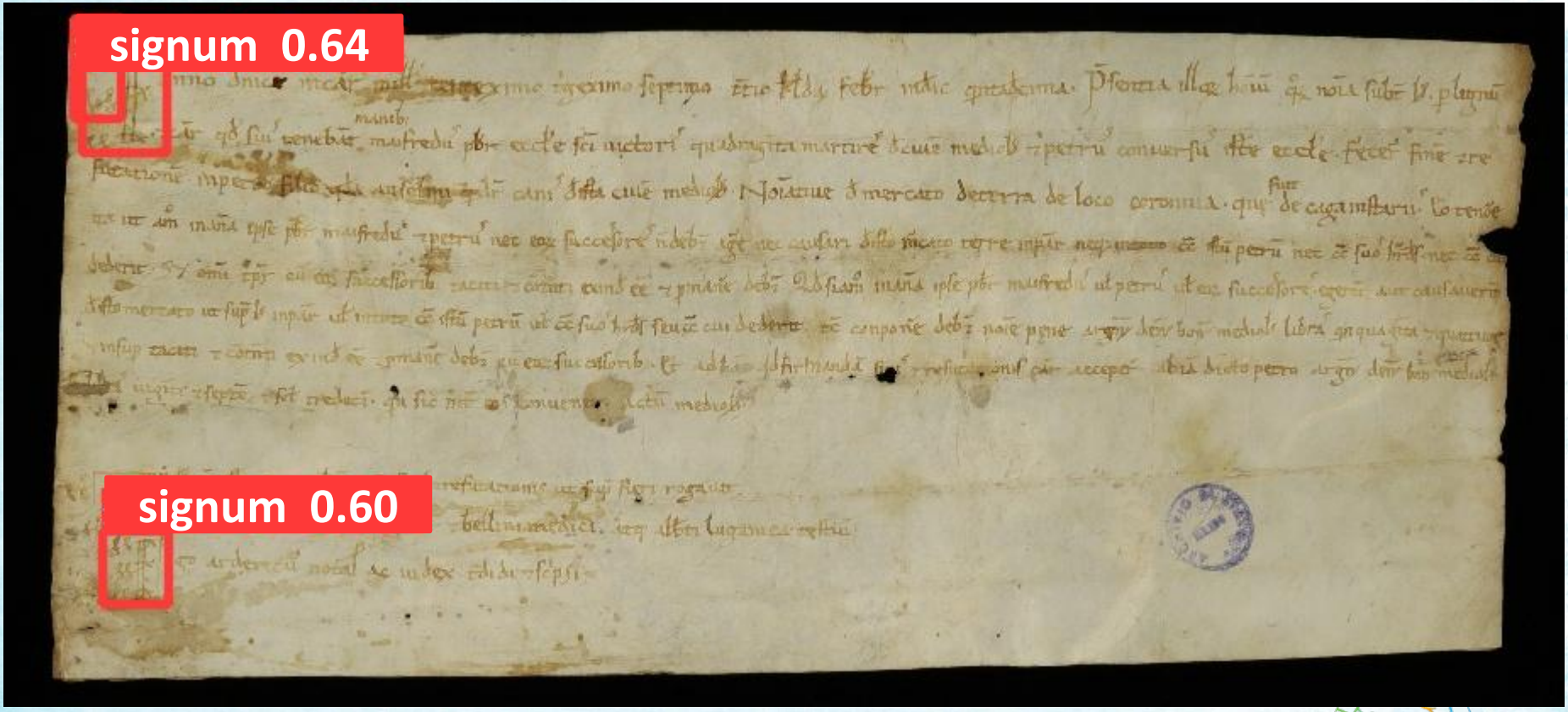


SIGNA TABELLIONIS DETECTION AND RECOGNITION: RESULTS

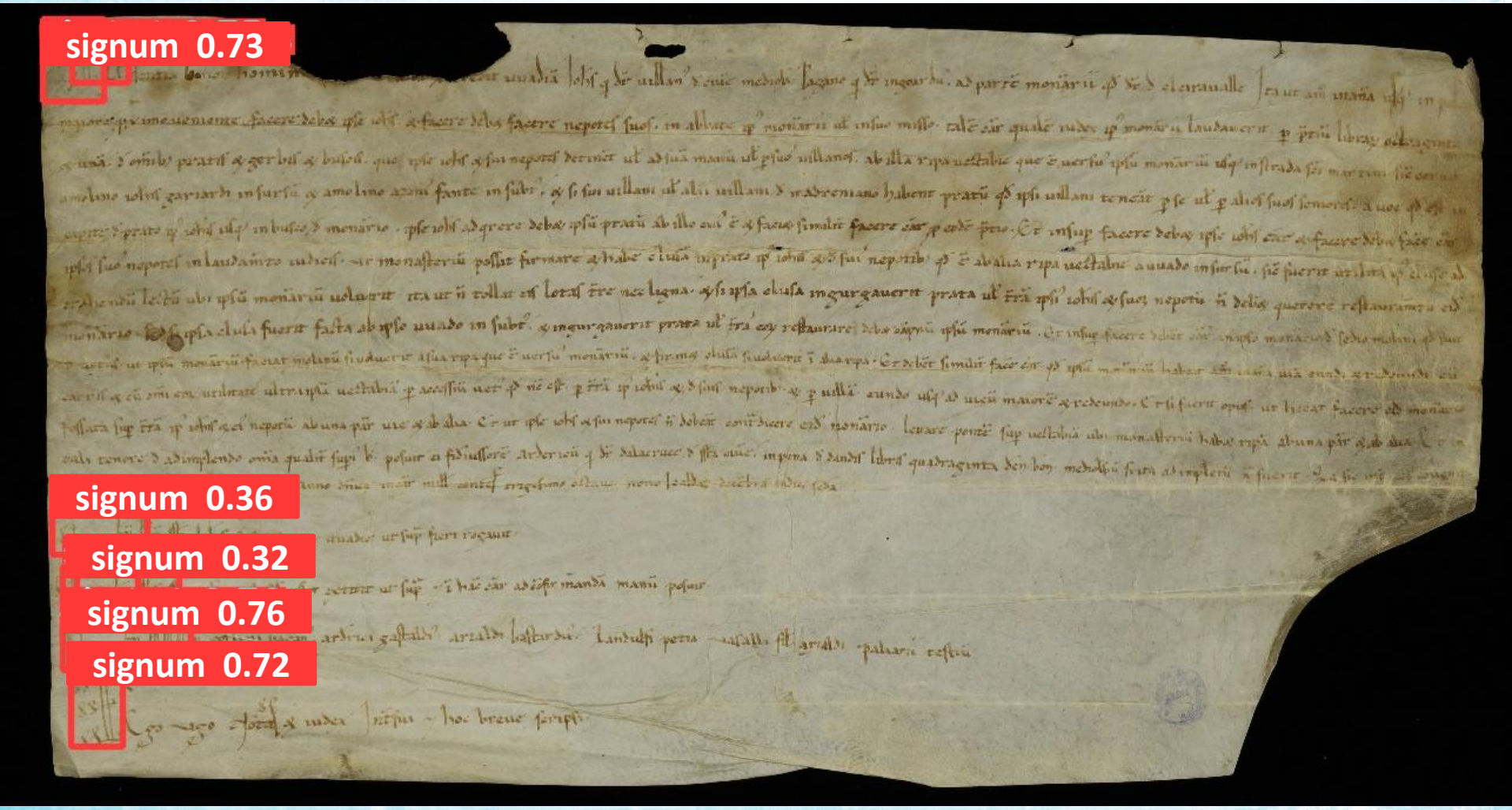


SIGNA TABELLIONIS DETECTION AND RECOGNITION: RESULTS

For each prediction is showed class (signum) and confidence

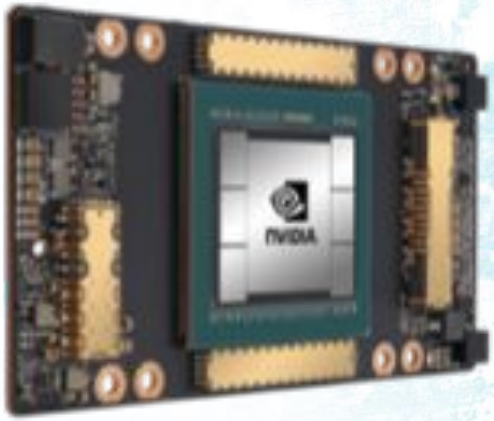


SIGNA TABELLIONIS DETECTION AND RECOGNITION: RESULTS



EXPERIMENT SETTINGS

- DNN Model: VGG16
- Image reshape to 224x224
- Data normalization: mean subtraction, std scaling
- Augmentation with image rotation and flip
- Batch size: 16
- Adam optimizer
- Learning Rate: 10e-4 (from 0th to 10th epoch)
- 10e-5 (from 10th to 20th epoch)
- Cross-Entropy loss
- Training from scratch
- 20 epochs



InterPARES TrustAI

OTHER ARCHIVAL APPLICATIONS

Once refined and fully developed, the AI could be used in a **wide range of applications**:

- Recognize the **peculiar system of writing** of every single notary**.**
- Analyze **archival notes on the back of the documents** and retrace previous archival arrangements or uses of groups of documents**.**
- Recognize **recurring images or other features** in huge series of documents**.**
- Identify **common patterns in manuscript maps or drawings**.
- Make publicly available **original & relevant AI Data Set.**
- And many others…

InterPARES TrustAI

AI in the Middle age
Arrangement of the documents via appearance-based recognition

Prof. Emanuele Frontoni
emanuele.frontoni@unimc.it

VRAI homepage: http://vrai.dii.univpm.it
Publications: http://vrai.dii.univpm.it/publications
Software & Datasets: http://vrai.dii.univpm.it/datasets
YouTube: shorturl.at/htvN2
LinkedIn: https://it.linkedin.com/in/emanuelefrontoni