L'Intelligenza Artificiale per il Medioevo: Ricognizione dei documenti antichi tramite riconoscimento appearance-based

A cura della SOPRINTENDENZA ARCHIVISTICA E BIBLIOGRAFICA DELLE MARCHE – SAB MAR

Benedetto Luigi Compagnoni, *Soprintendente SAB-MAR*
Emanuele Frontoni, *Professore ordinario Università di Macerata*
Stefano Leardi, *Direttore ASMI*

[Left document]

+ In dei noīe. Anno dnīce īca
nātio. M.C.L.xx.uiii. mēse ma
dii. indiē. i. Ego albtus stabit
ramoni. pa mea bona uo
lūtate. do. uendo. tibi atq
cedo. tibi pet guilen paga
ni. 7 tui hbs ad pptuāe pos
sidendū. rē ipsa res ī terri
torio esetano. ī plagis uocabu
lo ad seletra. tra. 7 silua. 7
pertī suū ibi habeo. 7 part
mea ī de ipsa silua. siū ibi co
par... i. do uobs sine ulla res
uatione. cū intritu. 7 ex trā
suo... q supse ut ī trase hab...
cui... res facapitē. 7 uio latē
tra. ī gti. a pede. tra silua
gard. 7 tri. a balio latē. tra giso
ni adūm. uñ recipiate ptiū
iiii. sol. pusui. 7 hpimntu. 7 obla
ne. 7 meos hedes. tibi st. pet. 7
tuis hbs ā remoūe. 7 auitstare
defende. sic au dupt 7 meliora
tū uot restui. 7 cartula ista firma
p maneat. q scrpi ego Barachā
noī. felicit.

+ Ego acto stabit ramonī. test
7 sp acea uice petrelli. test
7 Berad ramerī pagani. test

[Right document]

In dei nomine. Manifesta causa est qualit[er] nos dom[i]n[us]... bga...
uerabilis abbatissa de monesterio... sc̄i... cū conse[n]su...
tibus in t[er]ra mea uobis... cē scbla. hoc q[ui]a... bn̄
bona... uoluntate... nostra... cū consu annual. ui gē consensu... de... mo[n]e...
rego... uos... abbatissa. co[n]cedimus... ip[s]a... abbatissa... de bn̄... bga...
monesterio... nob[is]... ad nos... 7 tui... p[er] be... uidere... cec co... et... tī...
eīp... 7 p[er]... et ad fili ut... ano... nos... ip[s]a... uidere... 7 rem... uoia... sē... nra... ma[n]
una legra ma s[e]r ea... ide... ip[s]a ui... ra... cascatur... in loco q[ui] dicitur... qual duce
p[re]si... eq[ue]... e mea nra... cū corciu... adriaia... traberse... n̄i... da capriū... t[er]ra... legro... mōgie...
q[uo]remansi... de pede... neq[ue]... cē ra... aff... n̄ semp[er] mansi... e de uiuo loc[us]... nep... eq
pa... sto manere... p[ro]remansi... et abalio lat[er]e... sine rego de uere... c[er]... 7 q[uam]... mōt... et q[uo]reman
si... ip[s]a s[e]r t[er]... cū omnia q[ui] ip[s]a bene uiu... rep... ui... mini... ra ac ce desti... hoc est... ab tissa
no b[is]... uel... rec... ua... e... ac... ui... ab... ra... ab... ne... ip[s]a t[er]ra... nra... non bpi... des... dū... cedo
nomin... e cū ea bia...dh... ne... nullia q[ua] mea... eū... ra... mōgie... r[e]s[er]ua[n]di... et... ui... et... labora
no adh... re cū ea... bia... dh... ne... nullia q[ua] mea... e... ip[s]o mo nr... si... 7 o gū... bn̄... na[n]de... et de ip[s]o... usū... uel... tbi... ip[s]e
liū... ad ex... pugend... e... c[um] appmea... ie... de ip[s]o mo nr... si... faciam... ui... eq[ue]... ta... uē... ui... fine om̄ re calū... ua... lente
et dū... om̄ ni ea... nutri... t[er]... no bis descēde... eq[ue]s... faciam... ui... eq[ue]... ta... uē...... ui... fine om̄ re calū... ua... lente
su... p[er] sol bere... ut... la... e de beamus... nos... ip[s]a t[er]ra... no bia... ue... ra... tione... p ip[s]a t[er]ra... eq[ue]... p om̄... sī gu
le... for... nos... ī mēse madio... ī... seī... ra cū... arr... uei... gel... et uri ī... pi sage... tibi ui... de... nas... duodecim
le... for nos. ī mēse madio... ī... seī... ra... no... ue... re dū... e ne... ui... s[c]ī... p[ro]... nos... ut... p[ro]... mi... s[er]... nr... si... ab[er] op[er]a... tione
tū... c[er]a... et q p[re] gē... et... ra s[er]... ip[s]o... mo... ne... re... se... do ne ne... ui... s[er]q[ue]... nos... ut... p[ro]... mi s[er]... nr... si... ab[er] op[er]a tione
de dit... uel... ip[s]a abb[atiss]a... sap[er]... ra... ip[s]a... qua de bia... n[ost]ra de no stra... u[o]l[un]ta... bahe ri... sol qua
de dit... uel... ip[s]a abb[atiss]a... sap[er]... ra... ip[s]a... c[on]po... an su al de... rec[eptis]... et... ad us... de ip[s]o mōgie... et mu... s[er]et
d[e] gira... et... ip[s]a abb[atiss]a... c[on]po... an su al de... rec[eptis]... et... ad us... de ip[s]o mōgie... et mu... s[er]et
in p[er]... ra... ip[s]a... c[on]po an su al de... rec[eptis]... et... ad us... de ip[s]o mo[n]gie... et mu s[er]et... DE... rex... me... ui... mus... et obligauimus nos
in p[er]... ra... ne... q[ua] cales[er]... ip[s]a con le mē t[er]... et DE... rex me... ui mus et obligauimus nos
p t[er]ra... dam ne. et uoari... et lo... et clu bo... et bene dicere. eccl[esi]as et m[ar]... ti... et p t[er]ri... ut hī
rep[ar]a... re... ne... re ra tione... uob[is]... ip[s]a... abb[atiss]a. ut a p[ost]... nis... successores u[est]ri... et si nos... ficē... in arr... y[ri]... gū
et si non... st... habemus... q[uo] modo su p[er]... re... nostris... ut... tra... nostra... re ne p ra tione
q[uam] le de no[bis]... ut... pablis... ap[ar]e... est. co[n]ponam[us]... sot qua dra... m[e]a... st hanc escr... uob[is]...
ip[s]a abb[atiss]a ut a p[ost]... nis... successores u[est]ri... et q[ua]rcales... et si uer... nob[is]... con be me... ua...
ac dū... mascula... q[ui] be... na... n[ost]... scrī... tu... p[ra]ge... fui... s[er]... re... ego... rofredus iudex s[cri]psit et
r a yus... tero go de M[ar]ti... me scr[i]b[er]e... rogaberū... cū... ea... ra tione diū nos... q[ue]
cū... n[ost]... sua... ra in mille bicesimo... o... ce a bo ī mēse... ap[ri]lis... p indictione q[ua]rta decima

Signatures:

I signum manu... dui... mar... ca... ra nostra... et a fieri rogaberū... et si nostri u[er]se c[er]... ui...

+ I signum manu... marti... uel... mundi... regis... c[on]s[en]s[us]... ī... t[est]... et rog a tus... s...)))|||||||)

+ I signum ma... nū... le de... p[ro]... et test... e su... p[ro] ga tus... s)))|||||||)

+ I signum ma... nu q[ue]... dui... ui... et test... e... p[ro] ga tus... s)))|||||||)

A long story … with more than 80 partners worldwide

Archival science + Data Science & AI

InterPARES Trust AI ("I Trust AI") international research partnership, aims to (1) identify and develop specific AI technologies to address critical records and archives challenges; (2) determine the benefits and risks of employing AI technologies on records and archives; (3) ensure that archival concepts and principles inform the development of responsible AI; and (4) validate outcomes through a conglomerate of case studies and demonstrations.

https://interparestrustai.org

L'alleanza tra AI e Archivistica

Trusted Data Forever: Is AI the Answer?

Emanuele **Frontoni**[1], Marina **Paolanti**[1], Tracey P. **Lauriault**[2], Michael **Stiber**[3], Luciana **Duranti**[4] and Muhammad **Abdul-Mageed**[5]

[1] VRAI Vision Robotics and Artificial Intelligence Lab, University of Macerata, Italy
[2] Critical Media and Big Data Lab, Carleton University, Ottawa, ON K1S 5B6, Canada
[3] Intelligent Networks Lab, University of Washington Bothell, WA, USA
[4] InterPARES Lab, University of British Columbia, Vancouver, BC V6T 1Z4, Canada
[5] NLP and ML Lab, University of British Columbia, Vancouver, BC V6T 1Z4, Canada

Abstract

Archival institutions and programs worldwide work to ensure that the records of governments, organizations, communities, and individuals are preserved for future generations as cultural heritage, as sources of rights, and as vehicles for holding the past accountable and to inform the future. This commitment is guaranteed through the adoption of strategic and technical measures for the long-term preservation of digital assets in any medium and form — textual, visual, or aural. Public and private archives are the largest providers of data big and small in the world and collectively host yottabytes of trusted data, to be preserved forever. Several aspects of retention and preservation, arrangement and description, management and administrations, and access and use are still open to improvement. In particular, recent advances in Artificial Intelligence (AI) open the discussion as to whether AI can support the ongoing availability and accessibility of trustworthy public records. This paper presents preliminary results of the InterPARES Trust AI ("I Trust AI") international research partnership, which aims to (1) identify and develop specific AI technologies to address critical records and archives challenges; (2) determine the benefits and risks of employing AI technologies on records and archives; (3) ensure that archival concepts and principles inform the development of responsible AI; and (4) validate outcomes through a conglomerate of case studies and demonstrations.

L'alleanza tra AI e Archivistica

«Long before big data as an idea had been invented, archives already measured their collections in kilometers of files and folders»

Giovanni Colavizza, Tobias Blanke, Charles Jeurgens, and Julia Noordegraaf. 2021. Archives and AI: An Overview of Current Debates and Future Perspectives. J. Comput. Cult. Herit. 15, 1, Article 4 (December 2021)

Digitalised Heritage Data	Size
Fondo Ufficio italiano brevetti e marchi, Trademarks series: volumes with trademark registrations	30 TB
Official collection of laws and decrees	15 TB
Fund A5G (First World War): files with various documents (reports, reports, correspondence)	1 TB
Special collections (documents declassified under the Renzi and Prodi Directives): reports, reports, circulars	2 TB
Judgments of military courts	3 TB
Various photographic funds	2 TB
Digitised study room inventories	15 TB
National Archives of the US	1323 TB

Iterpares Survey 2022

Lucca
Beni
Culturali®
CANTIERE CULTURA

InterPARES
Trust AI

Ancient parchments

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*.)*
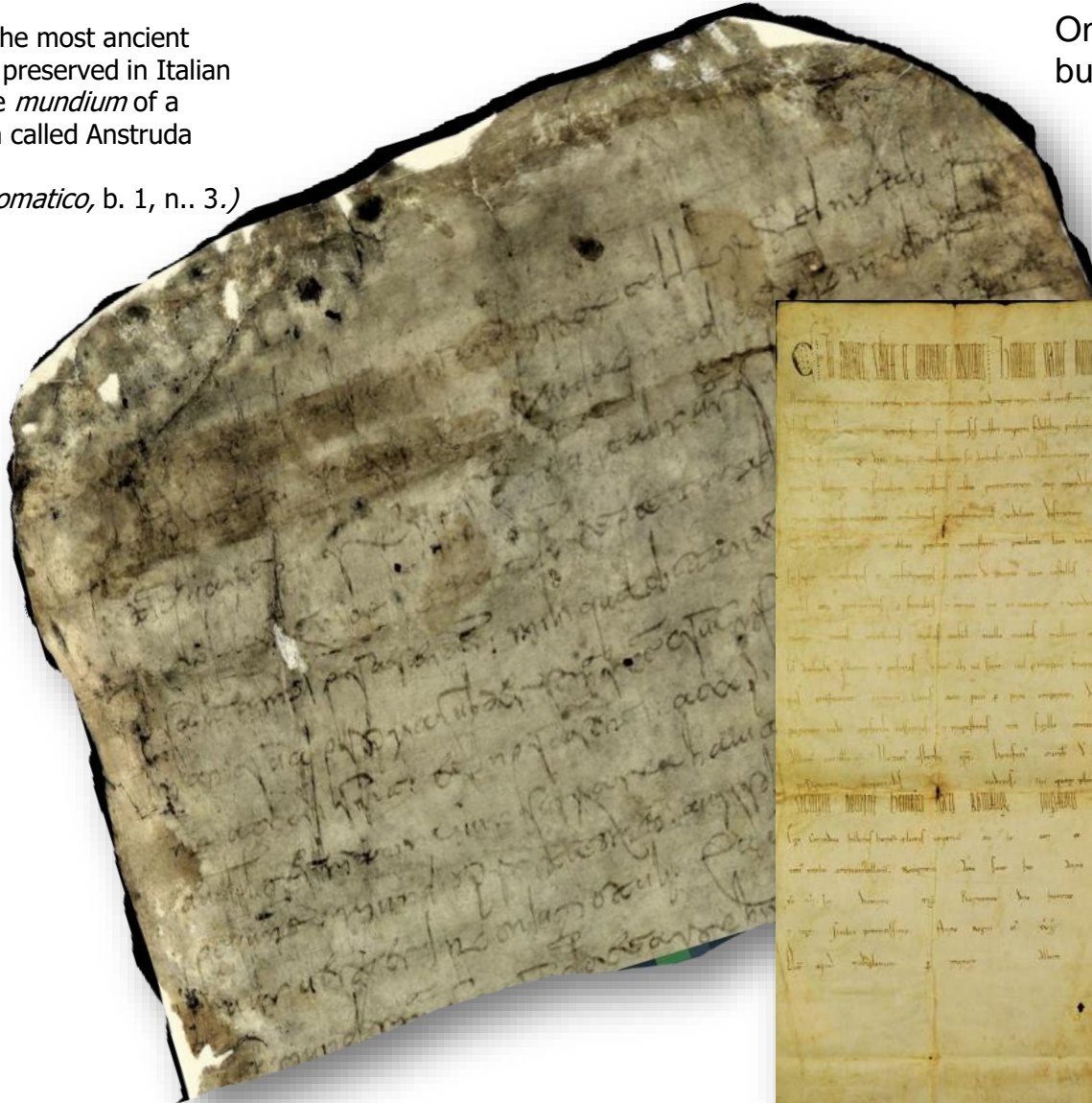
One example of a papal bull and an imperial charter

Anno dnice incar̄ millesimo centesimo quadragesimo tercio: sedo teal madii in die sexta. Placuit atq̄ convenit int̄ donu̅ guisredu̅ monachu̅ z abb monasterii sci ambrosii ubi e̅ sciom regelent

corp̄ cu̅ consilio monachu̅ suorum atq̄ landulfi qdr gratis avocati ipsius mo̅nari · guigerii z guico qdr stapa vasalli ist mo̅nari · ibi astanti̅b̄ z laudanti̅b̄ z subt̄ confirmanti̅b̄. Hec no̅ z int̄ ambrosiu̅ qde grega

nu̅ z meramu̅ qdr flouco z guilicione̅ qdr senart̄ atq̄ carmeriu̅ ad parte̅ utilitate̅ · ipsou̅ omiu̅ scolariu̅ igit̄ d scola eccle sci petri qd r̄ utali q̄ sc̄ z iu̅ria erit: Ut i̅di nobe o̅bett̄ due sic

ipseu̅ d̄ot̄ ipse abb. Eide silicet ambrosio · z meramu̅ seu guilicioni atq̄ carmerio ad habendu̅ z tenendu̅ seu censu̅ reddendu̅ libellario no̅e usq̄ ad annu̅ uiginti z none expletos z dino inuia usq̄ i̅ p̄peru̅

de modios tres dice̅t̄ t mesura mediol̄ sic̄ti omi anno dati z consignati in cameua ipsius mo̅nari sci ambrosii. Ea videlicet racione ut ad inuia usq̄ illa costitut̄ habet̄ z t̄e o̅bet̄ ipsi scolares

p̄dict̄ sic̄ti · sace ergo pd̄q̄ ipsis scolari̅b̄ silicet ipsis scole utile fuerit. Et psoluere erund d̄q̄ t̄ ipsi scolares eide mo̅nario si requisiti fuerit censu̅ singul̄ annis argenti · den boñ unu̅ z candellã unã illa stat̄

ipse abb sino fecert̄ eide scolari̅b̄ dr̄i dict̄ du̅ ist̄o · candellã. Et p̄misit̄ ipse abb silicet p̄u̅ti̅ ipsius mo̅nari sic hic int̄a salvabitur p̄dict̄ sic̄ti omi anno dare ipsis scolari̅b̄ sic suss̄ l̄ cu̅ suis sucessori

b̄. Alia sup iposita erund int̄ cal̄ no̅ sit̄ penã û not̄ posuer̄ ut q̄ eu ipsis ut eon h̄edi̅b̄ aut̄ sucessori̅b̄ se d̄ hac coue̅niencia libelli mouere p̄su̅fert̄ z inea omia ut suss̄ l̄ no̅ p̄mãsert̄ t̄e con

ponat̄ illa pars que hoc no̅ co̅suauerit̄ parti̅ side suamti̅ pene no̅e argenti du̅ denariis mediol̄. libris. dcc z octo z insup in eade coue̅niencia libelli p̄miett̄ und ipse abb accepit ab ipsis scolari̅b̄

p̄ ho̅c libello argenti du̅ bonou libra noue quos ipse abb d̄t̄ ad parte̅ ipsius mo̅nari ferreto filio qd̄a petri qd̄ icebat̄ rudeus d̄rie mediol̄ p̄manerts̄ qd̄ ipse abb fecer̄ cu̅ illo ad parte̅ ipsius

mo̅nani inte̅cis z finis do̅mergasco z cadello z mariano r̄q̄ r̄q̄ · z l̄go z iu̅dronio. hoc uel e̅ euacat̄ actii ist̄ mo̅nario:

⁜ Ego guisredus abb ut̄e facto h̄fi

Ego suzur t̄e h̄fi   ego romanus m̄ t̄b̄ ⁜ Ego rotst̄ m̄ z sac h̄fi

† Sigñu ⸭⸭⸭ manui ist̄o landulfi avocati q̄nt̄ fuit z firmauit:

† Sigñu ⸭⸭⸭ manui ist̄o guigerii z guigi vasallou q̄nt̄ fuer̄ z firmauer̄:

† Sigñu ⸭⸭⸭ manuu guarnerii cagatosico z onrici p̄ d̄anrolsi malmarini ast̄ t̄ratrum pedracu̅ d̄ fagnano scormoli atq̄ p ambrosii qd̄ d̄ lomaeu tellu̅:

⁜ ariald notar sacri palacii int̄ fui z roga h̄fi:

Ego petrus notar scpsi post t̄dot̄i co̅ptu̅ z d̄t̄i.

Metodi Apprarance-Based

Robot Localization in Partially Explored Environments

Appearance-Based Robotics

© DIGITAL STOCK

Robot localization has been recognized as one of the most fundamental problems in mobile robotics. Localization can be defined as the problem of determining the position of a robot. More precisely, the aim of localization is to estimate the position of a robot in its environment, given local sensorial data. This information is essential for a broad range of mobile robot tasks; in particular, the robot behavior may depend on its position.

This article presents a novel and efficient metric for appearance-based robot localization. This metric is integrated in a framework that uses a partially observable Markov decision process (POMDP) [8] as position evaluator, thus allowing good results even in partially explored environments and in highly perceptually aliased indoor scenarios.

Odometric sensors play a critical role to solve the localization problem in wheeled robots, as they provide information about robot movements. Unfortunately, these sensors are noisy and accumulate errors over time. Starting from a known position, they are accurate enough for local movements but are not suitable for long-term localization, and several techniques, using different sensor modalities, have

been proposed to permit robot self-localization. In this article, we shall only consider self-localization by means of vision [5], [6].

The extraction of visual features for positioning is not an easy task, due to the amount of information provided by visual sensing in the robot's environment. A preliminary task becomes the selection, or better, the learning of an appropriate set of visual features (often referred to as landmarks) to be used for the navigational trials [10]. Another difficulty of a vision-based self-localization process is to solve the matching between the observations provided by the robot's sensors and the landmarks themselves [12]. This problem could be considerably simplified if a robot's pose (localization prediction) is given a priori. On the contrary, without prediction, the absolute matching is quite difficult because the observations are not error-free.

Only recently, appearance-based (or view-based) approaches have been proposed [1], [9], [13], [18]. An appearance-based approach provides qualitative measurements of the position of the robot, thus monitoring the progress of the overall mission. Once certain relevant positions are attained, other navigation

BY PRIMO ZINGARETTI AND EMANUELE FRONTONI

Metodi Apprarance-Based

Pergamene antiche

Cosa vale la pena di «capire» meglio dei documenti antichi?

Forme
Signa
Posizione dei testi
Difetti e parti mancanti
Stile di scrittura / Abbreviazioni
Firme
Note (anche sul retro)

Come possiamo analizzare queste caratteristiche ?
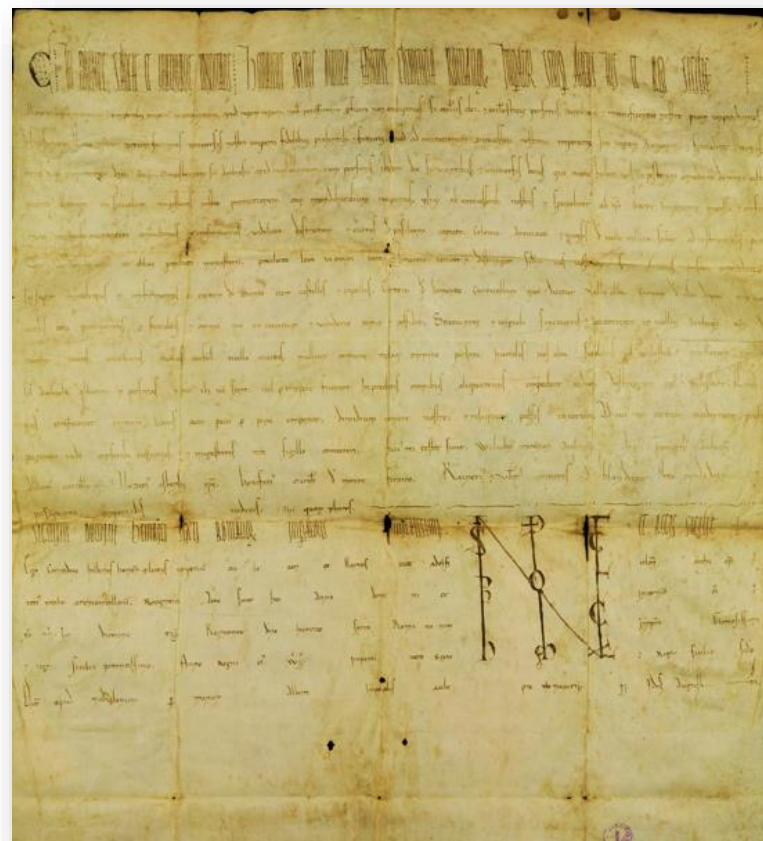
PERGANET Percorsi di Ricerca

- Benché studiati e redatti a partire dal XVIII secolo, questi ampi materiali non sono mai stati indagati in modo sistematico al fine di individuare caratteristiche comuni ricorrenti in gruppi omogenei di documenti.

-

- L'elaborazione automatica di un gran numero di documenti scansionati può portare a nuove comprensioni di questioni trasversali generali, non legate a un singolo fondo.
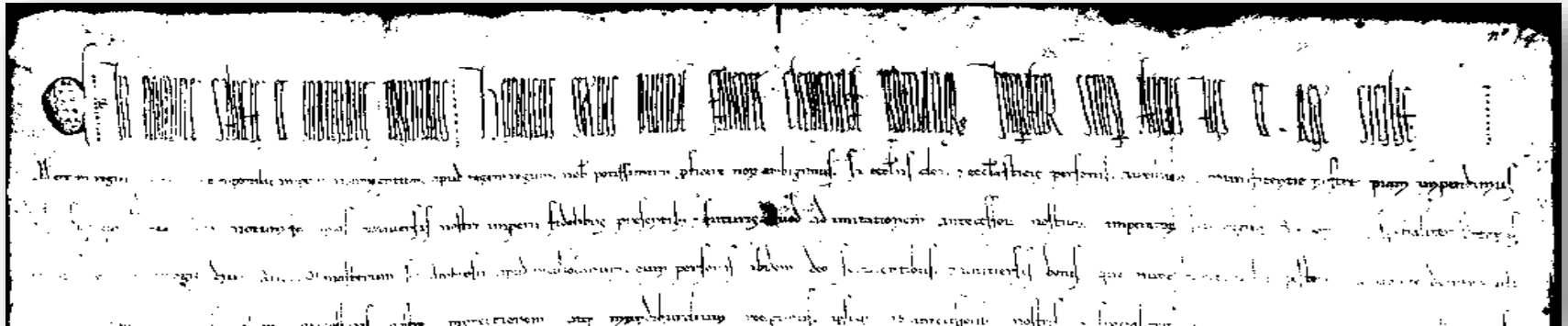
PERGANET Il Signum

- Il signum o segno notarile è un marchio disegnato specifico e personale utilizzato da un singolo notaio nel protocollo e prima della sua firma.

- Identificare il signum significa che ogni notaio può essere riconosciuto e rintracciato in una serie praticamente infinita di documenti.

- L'IA contribuirà sia a creare una biblioteca di signa: una matricola virtuale dei notai e una base per indagare le loro caratteristiche meno visibili.
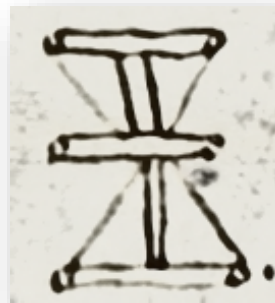
Una matricola notarile prima del 1350 d.C.

Una matricola notarile prima del 1350 d.C.

- Uno degli obiettivi del progetto è quello di utilizzare l'IA per costruire una Matricula di notai milanesi del XII e XIII secolo (in fasi separate).

- L'esame progressivo di tutti i documenti pergamenacei dell'ASMi porterà ad un elenco completo dei notai e alla loro identificazione.

- Il processo genererà anche un elenco di tutti i documenti sopravvissuti relativi a ogni singolo notaio.

InterPARES
Trust AI

PERGANET Dataset

- La raccolta dei dati sta andando avanti e attualmente abbiamo i seguenti archivi che collaborano e raccolgono dati per lo studio:

Archivio di Stato di Milano (220 documenti già elaborati, quasi 1200 scansionati e pronti per essere elaborati)

Archivio di Stato di Novara (91 documenti pronti per essere elaborati)

Archivio di Stato di Ascoli Piceno (27 documenti pronti per essere elaborati)

Perganet

Parchments Digitalization and annotation

Input data
Scanned Parchments dataset

Image collection
Parchments Storage
Parchments Annotation
Signa Annotation

Deep Learning-based analysis

Binary Classification Recto/Verso

Text Detection

Signa Detection

Signa Recognition

Insight & Evaluation

System Performance Perspective

Historical Perspective

Parchments Historical Reconstruction

PERGANET Datasets

PERGANET DL Pipeline

| Binary Classification Recto/Verso | Parchments Text Detection | Signum Tabellionis Detection and Recognition |
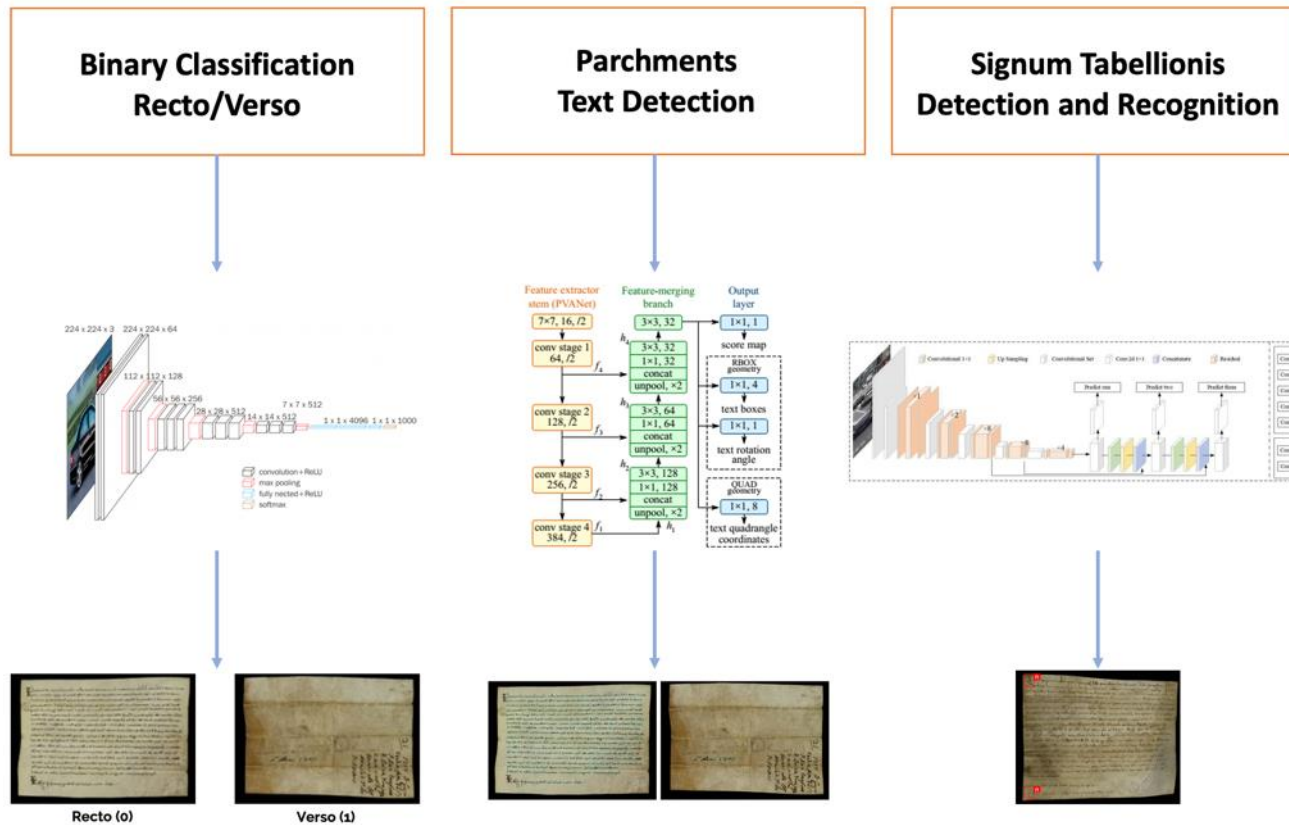
Binary Classification Recto/Verso

VGG16 DEEP NEURAL NETWORK

Recto

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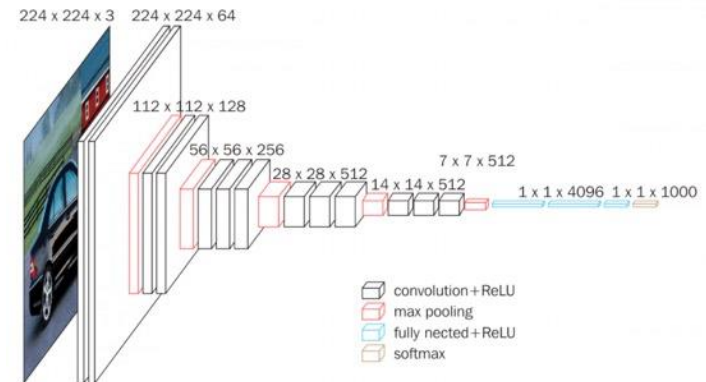
Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556.*

InterPARES
Trust AI

Binary Classification Recto/Verso

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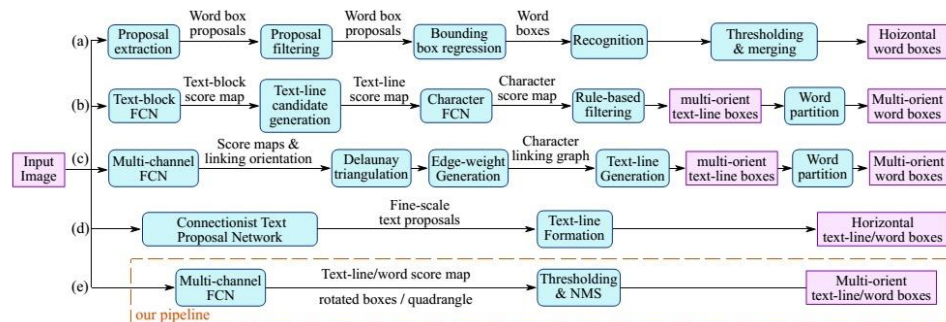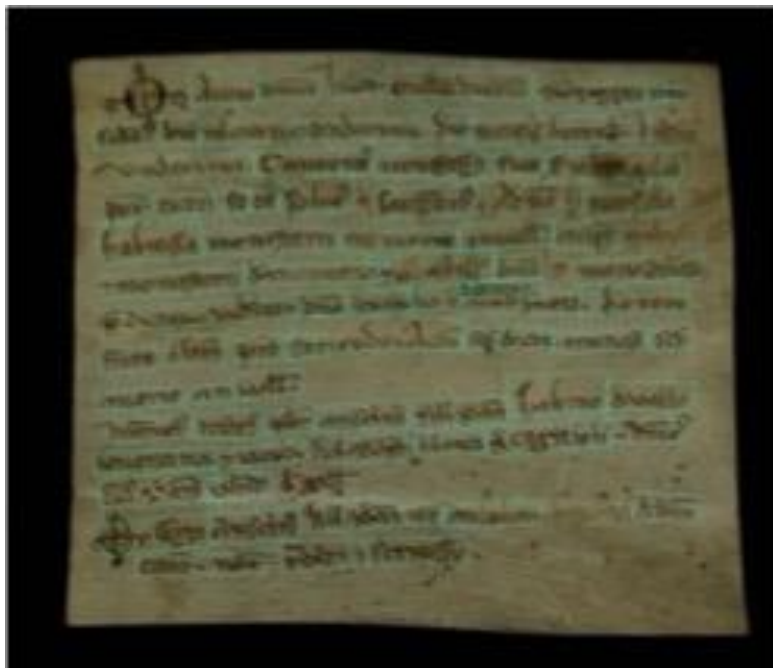
Xinyu Zhou., et all, A. (2017). EAST: An Efficient and Accurate Scene Text Detector. *arXiv:1704.03155v2*.

InterPARES
Trust AI

Parchments text detection

Word detection

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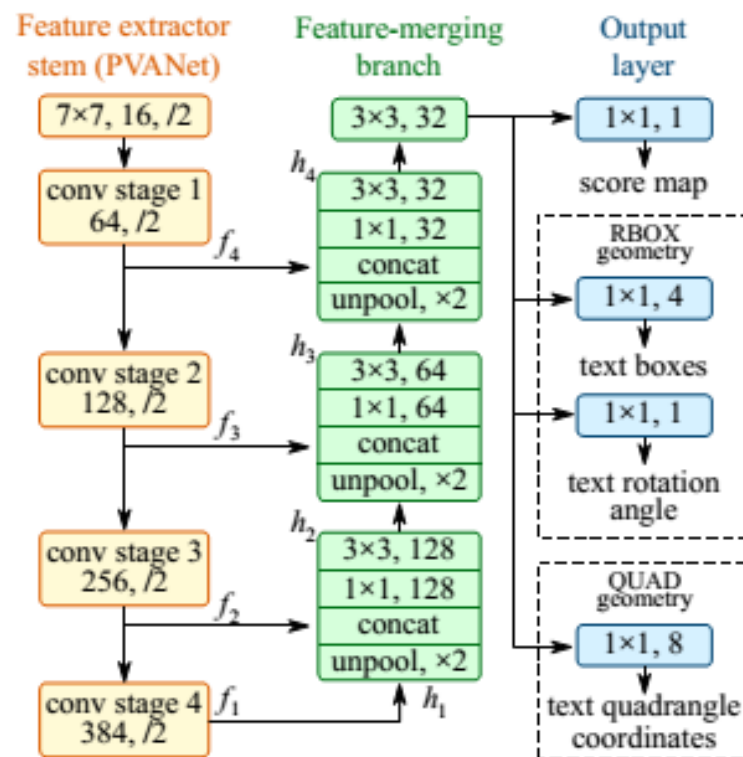
Parchments text detection

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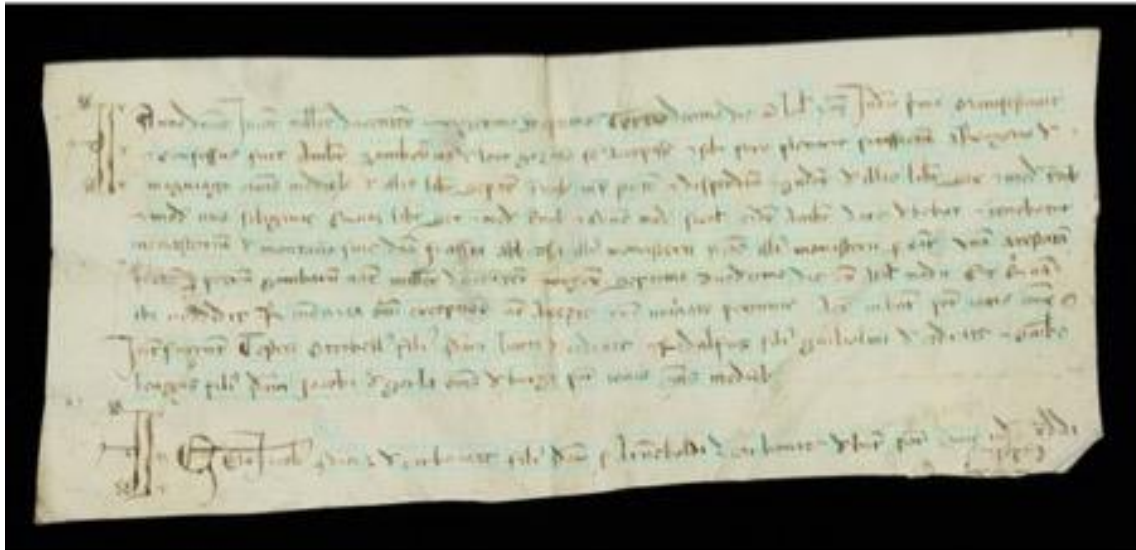
Parchments Text Detection
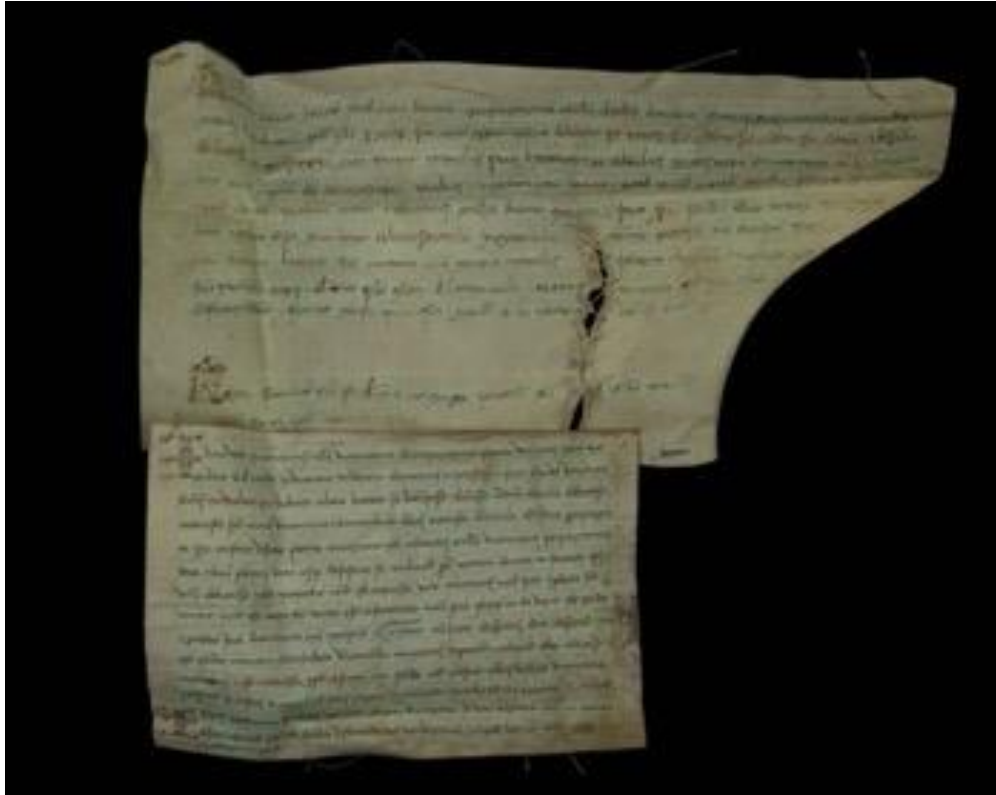
Signa Detection & Recognition
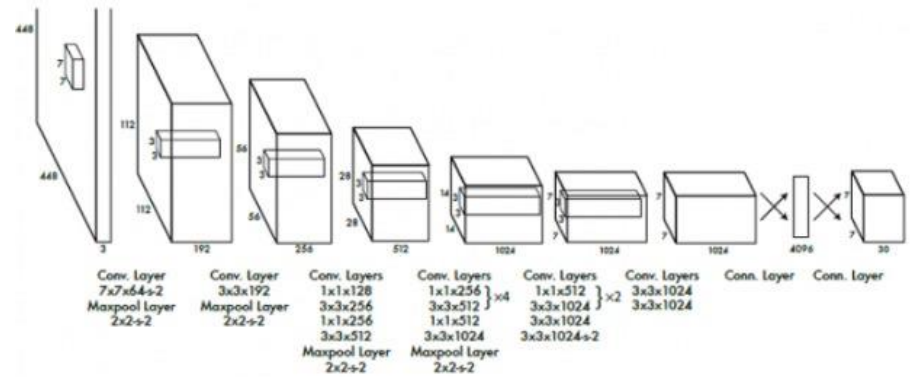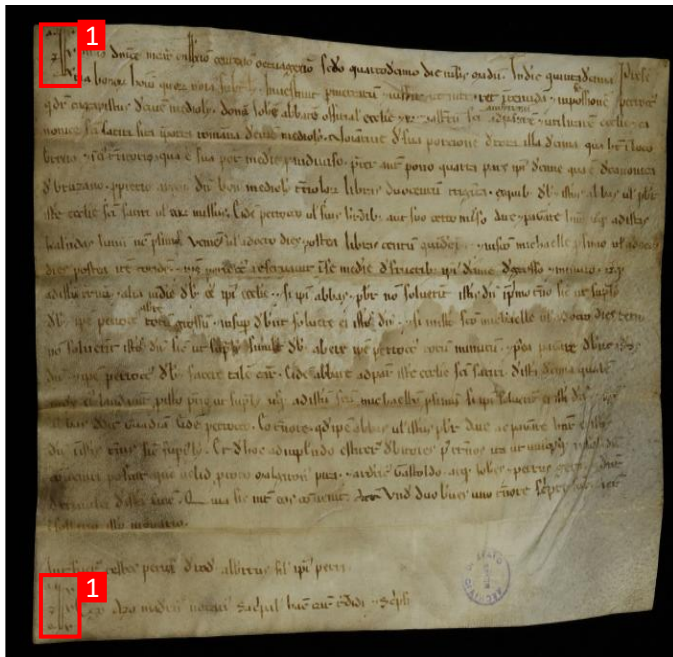
Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 779-788).

Annotazione matricole notarili

2768 signa

ID	NAME	HISTORICAL INFORMATION		
…	…			
109	notary 109			
110	notary 110			
…	…			
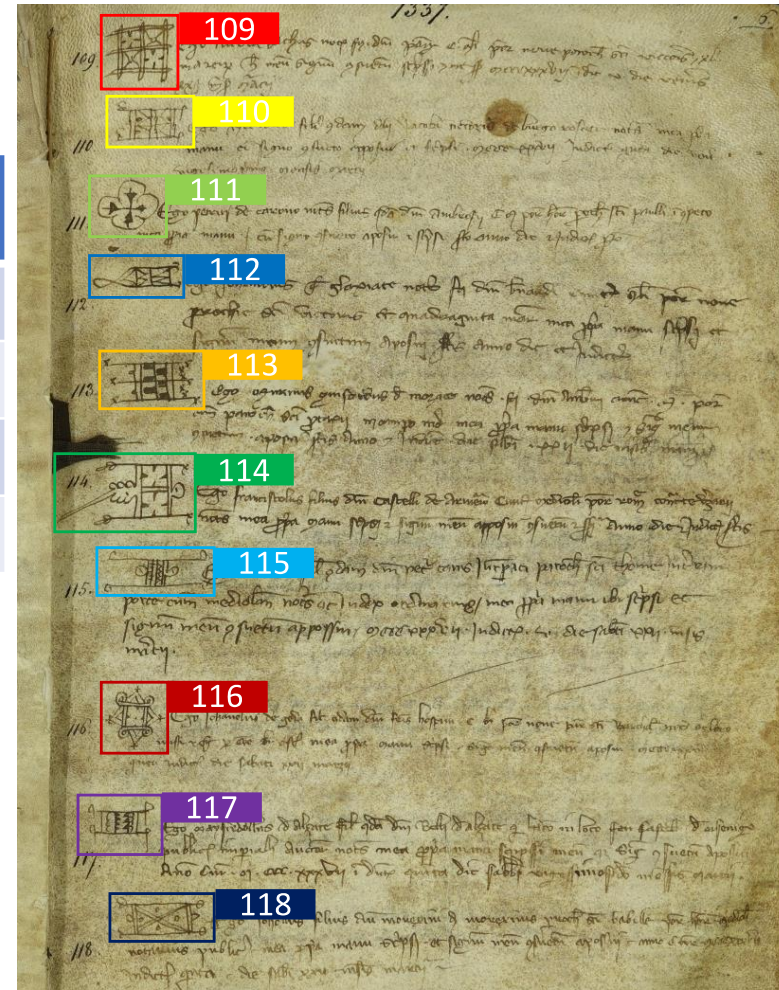
Signa Tabellionis Detection And Recognition: Results
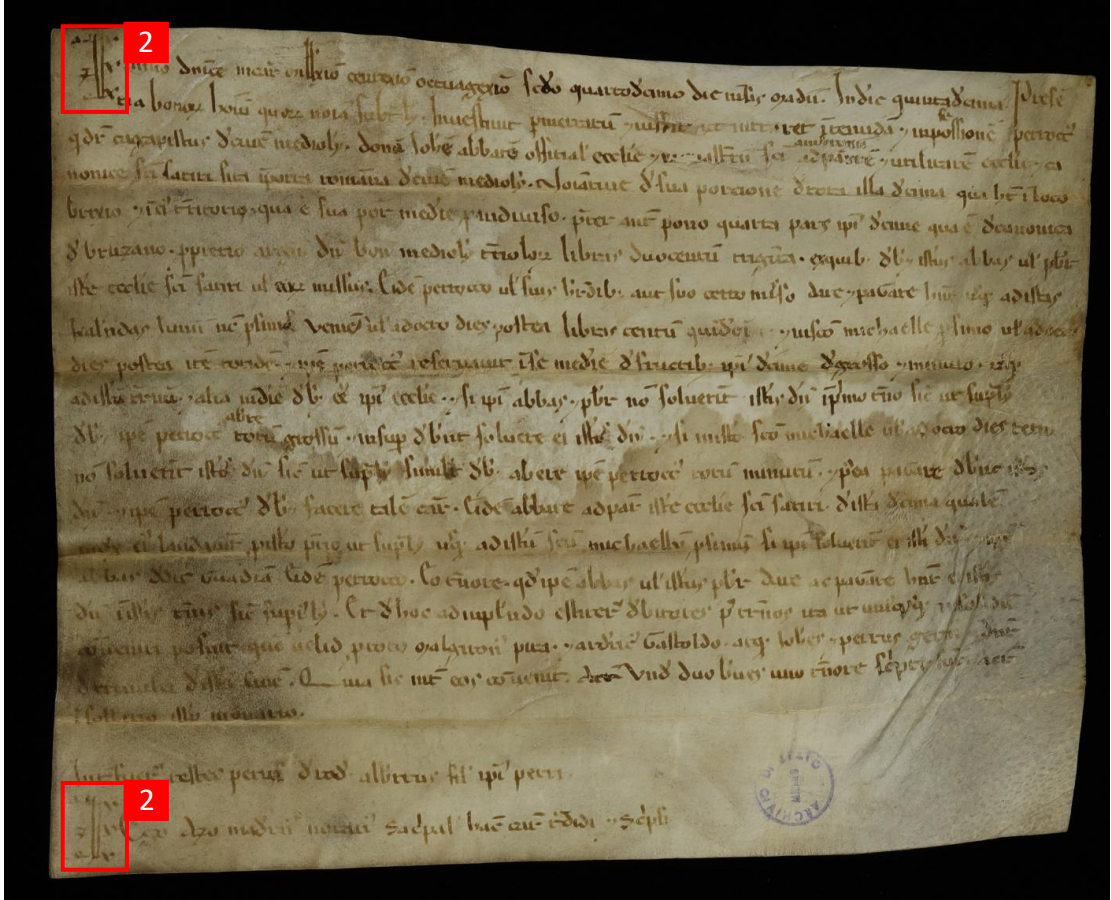
Rilevamento e Riconoscimento dei Signa Tabellionis: Risultati

Per ogni pronostico viene mostrata la classe (signum) e la confidenza
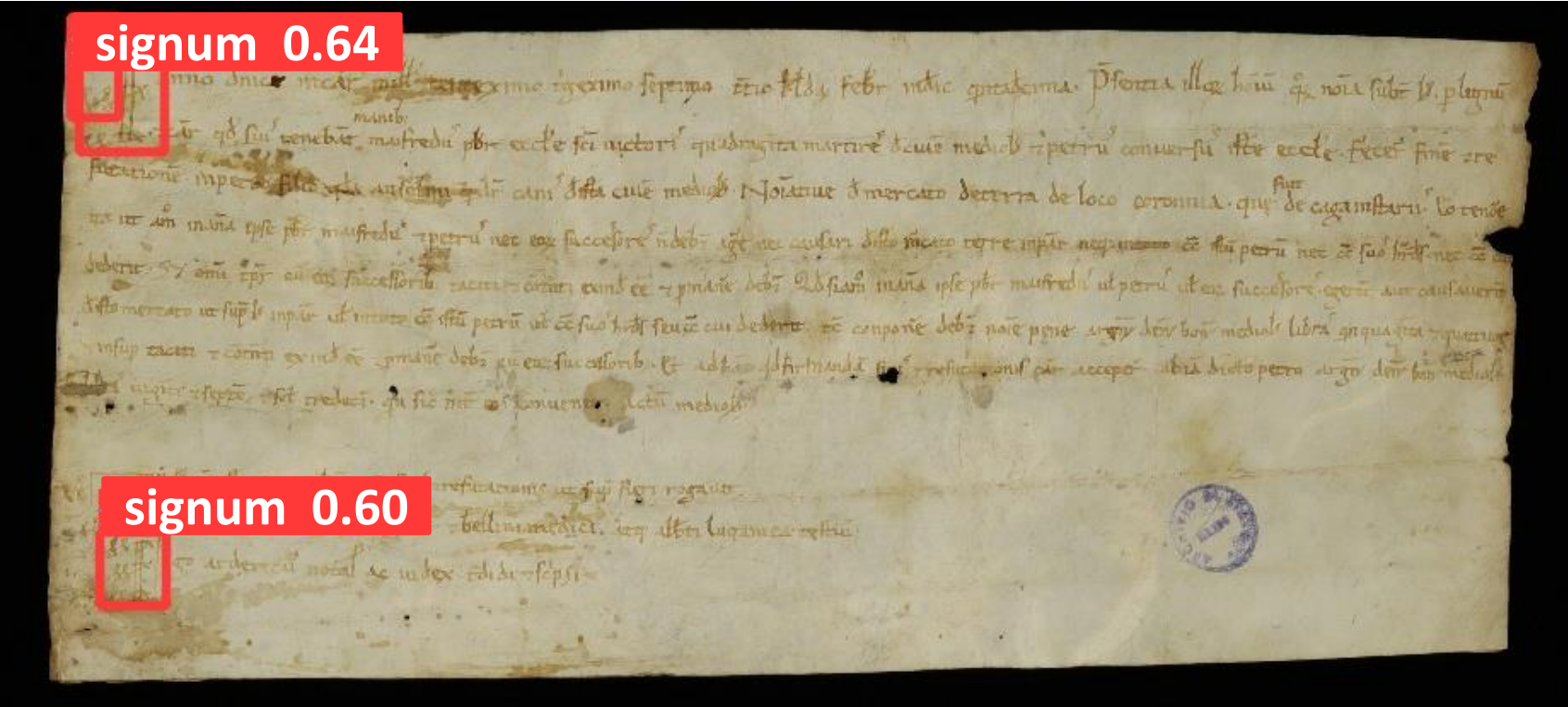
Rilevamento e Riconoscimento dei Signa Tabellionis: Risultati
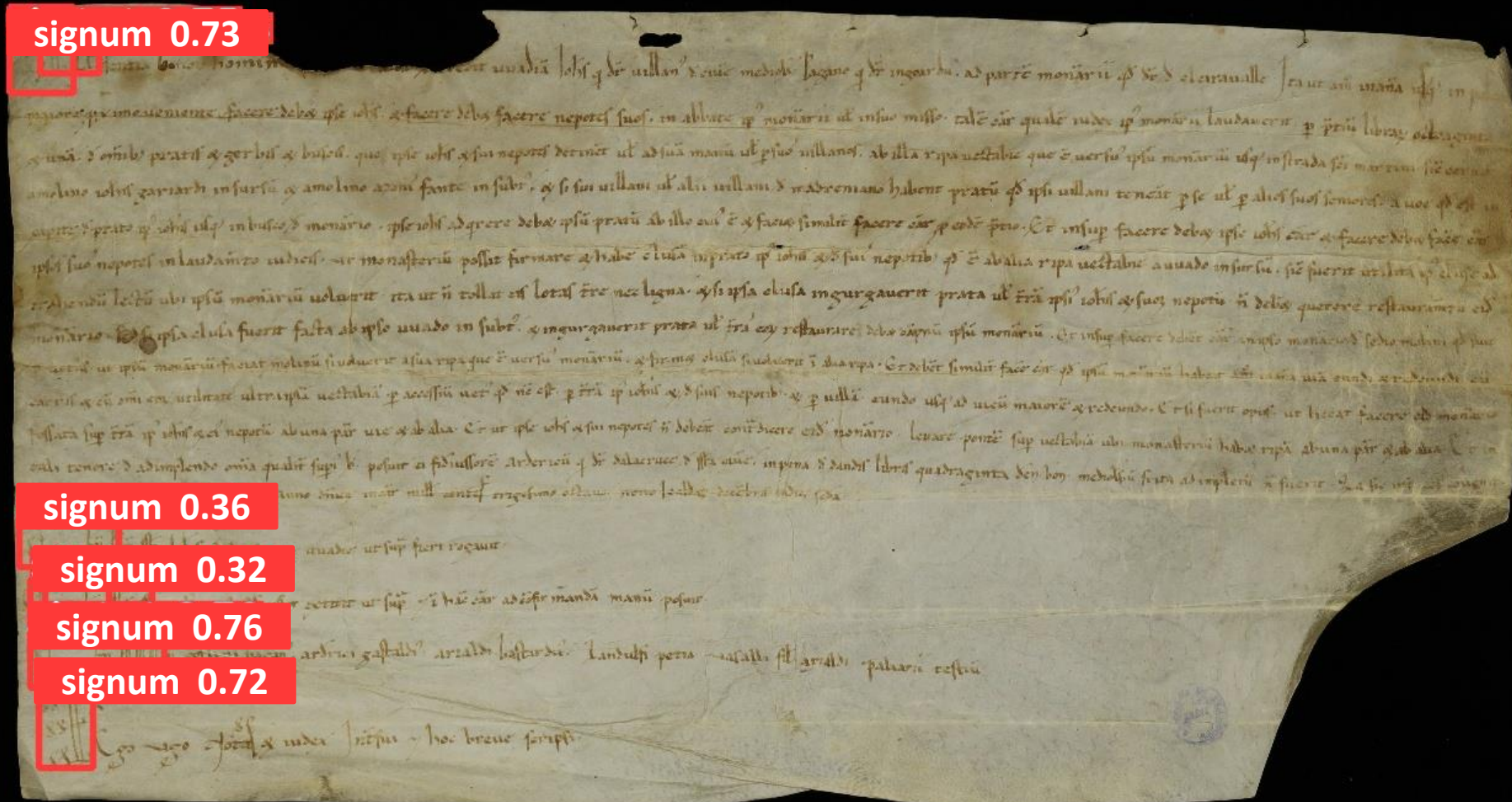
PERGANET2 LVLM

Possiamo progettare VLM specializzato per applicazioni Apperance-based in Archivistica ?
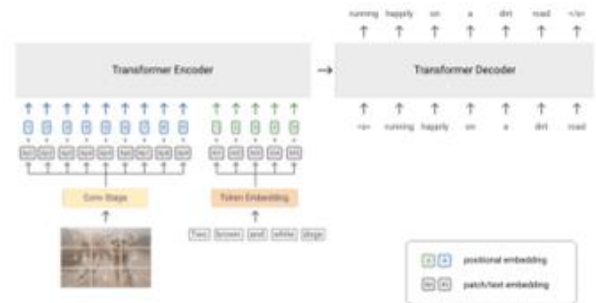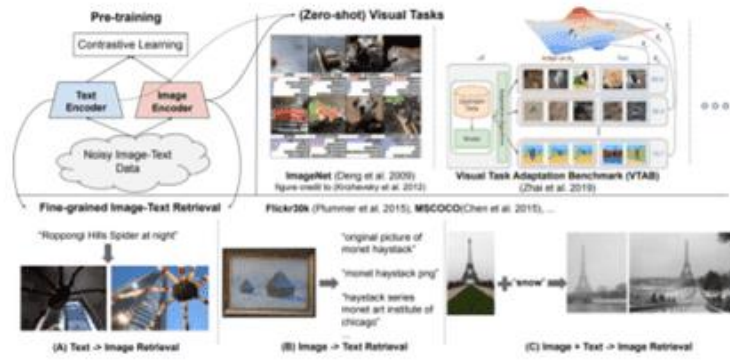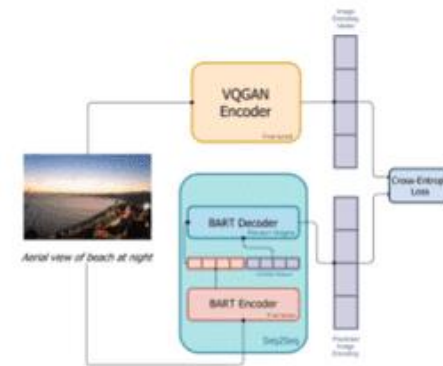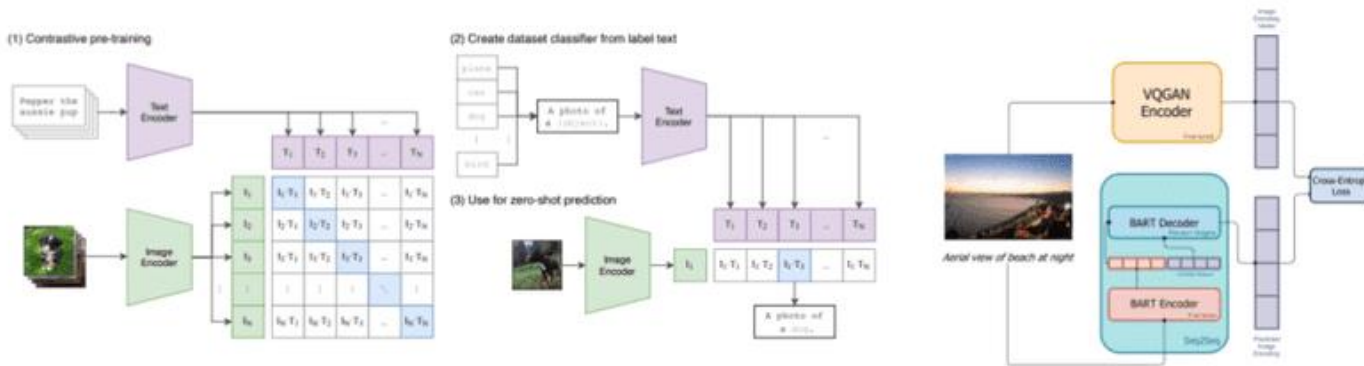
PERGANET2 LVLM

Prompt based approach:

«Detect Signa, percentage of text, and estimated datation on the following parchment»

«The picture reports the signa detection results. Each Notarium ID is reported in the attached table. Percentage of text is 73% Estimated datation is AD 1330»
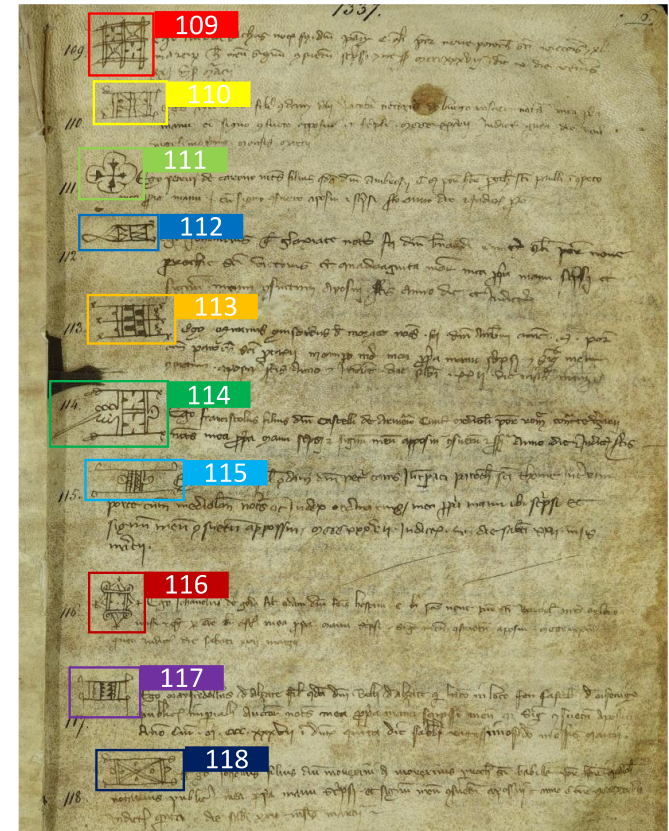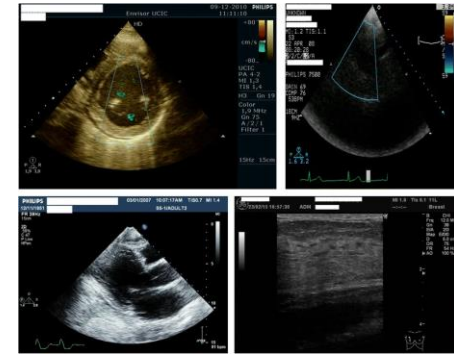
PERGANET2 Ethics by Design

Visione futura

Applicare l'approccio di PERGANET2 agli archivi di immagini Heath e alla segmentazione semantica delle nuvole di punti 3D
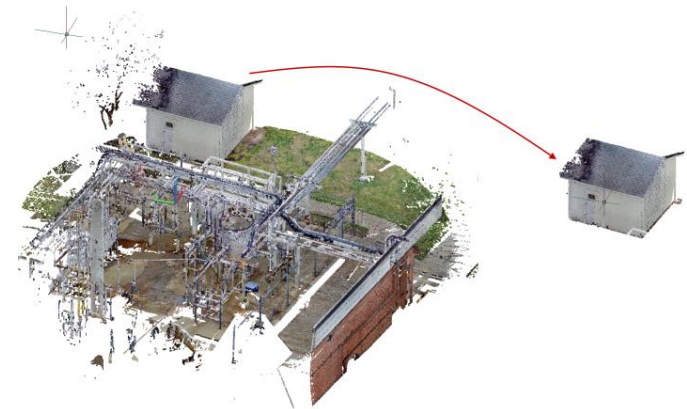
Raccoglere nuovi set di dati AS per approcci basati sull'Appearance

Mantenere la privacy e l'etica fin dalla progettazione allo stesso livello degli sviluppi tecnici.

Migliorare «l'alleanza» tra AI & AS !!

L'Intelligenza Artificiale per il Medioevo: Ricognizione dei documenti antichi tramite riconoscimento appearance-based

A cura della SOPRINTENDENZA ARCHIVISTICA E BIBLIOGRAFICA DELLE MARCHE – SAB MAR

Benedetto Luigi Compagnoni, *Soprintendente SAB-MAR*
Emanuele Frontoni, *Professore ordinario Università di Macerata*
Stefano Leardi, *Direttore ASMI*