











InterPARES Summer School Symposium – San Benedetto, Friday 7 July Artificial Intelligence to support the ongoing availability and accessibility of trustworthy public records - an overview of InterPARES Trust Al

Al for Multimedia understanding from computer vision to deep learning



Prof. Pierluigi Feliciati Prof. Emanuele Frontoni

University of Macerata

CONTENT

- Introduction
- From Machine Learning to Deep Learning
- Image understanding
- Deep Learning and Archival Science
- Managing Public Datasets
- The PergaNet study as a use case
- Other applications







INTRODUCTION

What is worth to better «understand» about ancient documents?

- Shapes
- Signa
- Position of texts
- Defects and missing parts
- Writing style / abbreviations
- Signatures
- notes (even on the back)



How we can study these features?



STUDY INTRODUCTION

- Ancient written private and public 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.

AN OUTSTANDING HERITAGE

- The **Archivio di Stato of Milan** (ASMi), among the others, 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 per 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).



In a demande les mis the que in the comme son falle of a mission in a posser and a solution and the first from one les entire longo by a form man of the second of the form one les entire longo by a form man of the second of the form one les entire longo by a form man of the second of the form of the second of the form of the second of the form of the second of the en ela relita parte abre e dalella como fecamer ser false probet ex amezone ne porche abre plante in porte false en marco plante en marco plante false en marco plante en ple amezone abestico a mundo alba mera de forma de colonte me albro infanto filos a mando de mando de monte de colon con de mando regi. Co note caballo reference de colonte a mando de monte de colonte con contra de principal de monte de colonte colonte colonte con contra de principal de monte de colonte colon the america a manication mere me se manication mere alle sole auguli a municalla mil Co un less that a motiva spongor to parente mind quo ner fute le Jugar plener a sellente ou una portane un una fair a una fine to fue le Jugar plener a sellente ou una fair a una fair a fair le Jugar plener a sellente ou una fair a de fair d comer sa a amizone a midule sume colerante for some and ab also hois milla par in understand no allena sportanea voluntate has our world face and for measure to for manufacture of the second form a milland form a migration of the second form a million of the second form a migration of the second form a million of the second form and the se can manufe traver and ab opti inquire of municipal of the control points provide gerbit a bufed our dies research in long about monier to que not sulf in habitum a consum as manu ut p mig will one quantities of the general and the general a points practice ger bit a bujet cura me record for another rotal garrand inful the a another fonce in fully a point all practique are plat rot as four inventor porture of the another or full practique are plat rot as four and made on de mayore & prese me us in busco & monare ou superior de sencional su inin le insuper estadon relacional monare de m of incourt prose of in the monator in full he form north of chie a trade in plu monafter in well at not light a monator in fall a light formare of haben at aliar of molar of the formare of haben at aliar of molar of the formare of the firmare of haben at aliar of molar of the formare of the firmare of the in in che monte de la la la monte de la monte de la me il chem querere estate and in inducert. Anar paque of und sonor in a firmage luta fruiducert malarrya Co entire monaria Co influencement malarrya Co entire monaria Co influencement malarrya Co entire monaria con of after monaria haben sent of ne of monagen a pulla engle a resound us abuci maver. Co upla monagen of apla monasterne trabent use por of part of monary that on the form and the flat one of part of monary that one of the court geredeunds ou corre & dist Francisco and the from fue que a zarefe en fabelle a amizon de beluse I hacin wend ut suf for requiert a uple will est matri sue a col suine com 1 1 1 1 mfanti Q ord zareje marre fue of the fecesmore Q ord stabelle congresse cofenier of an hac our advotor mand man popular or arrest of the fecesmore of ord stabelle congresse cofenier of an hac our advotor mand man popular ord arrest of the fecesmore of the feet o In the logian a unfred gering of sune a watall gerin of stabelle after the after of lovering court of behave a continuous of white car at cofe mand man popular bound Source marnery graffy got ling of pagan Lanfraer of to buzalem alber of the curator of each gararulla arnally & paule arrally baftard Johannon & pozel long 1 Se talliaferm andre tel tie Go to total a ruder corrept polt erratea coptur sededi og radeta beluita nepora mea cu polito lorerso parre fin introgani Man plant plit cefte dert unada poets with a arreto nepo ce of quanto magnified a albe gerne infante file of a ce alter habiter a feere habit of face alor plants file of a ce alter habiter file of a ce alter ha en just meder land amorit. ad parte in monara, d'es por post dan omini peri p fto pero recepto le si aplantante face habent face habent face heres of apla on ande posser or foustore this bennon. Super in ports & labert grapus ment doy bon. Ex iple solf deduc in sold in palle marrie facere habor face course lane on fill for car in landament indior, in monarm & fur por I delaw rote of police and fiduallore the bennone scure in pena labor cour. Configu that analis of roll or amero of perro of recipe habens our en laudaines under ab allis unland agh comere & upful practi of dare habens of car emperiori est monario. use inthe beats marcro ande poster of the sulling of the and inpena little ung

38 to Inno dince mear milletime centelime quadrigetime tercio: les teat madi indie sella: Clasier and consuent int doni quitredi monadi quitredi monadi quitredi monadi quitre monadi quitre de m 33 th corp cu contito monache fuou in landalf por graful aucent apfeul monare queco que topa unfalle off monare the atomorb fuot confermant let no ment ambrolie que green nu gumeranti qui stouco quillecone qui senare in commerci do parte quellerare i plou jomin scola este sei peri qui do la que sur jumin erie. It indi noe sbest due sie aplene som sple and the title fileer ambrolio gamenomo leu quilicioni any carimero adabenda y renenti leu centa redenda libellario noe uty asanto inquer y none expleto y sins mana uty imprerio De modol tre l'Streat Diffura mestol heti omi anno dati configuati monari lei ambroli: La modeler racione un addi mana usa inilio collecte ple telarel Broth heth face ever page upli feele upli feele unde fuert: &- plot und round ster tof feel with fuert confu fugue anul argu: Mi; bon und reandella und the har spie att fine feet Cità feelant: Sia siete su hete candella to punter spie att filicer purilire prul monnes sie hie selanteur pretti neta omi nune sare spil seelante sie sui fuellor by Alles superposited events interest in the pend is not posses in flew posses are successoris so that countries of the moner of summer or more primitive on the pend is not primitive or the pend is not posses are successoris. pomat illa part que los no confusiore parti fise fuanti pene nos argii di londi medol libert: Sec - octo - infup in cade consenienta libelli primer una uple abbaccepte abiptir scolaribi place libello mon du bonon liber noué quot iple abb Mor ad parte iplins monni servero file qua peri que bonon liber se medial promatoro que iple abb Mor ad parte iplins monari interi - funti domorgateo - cadello - marrano aquago intego fundiones: 1000 100 of cuma acti alla manario of ego wetredurator wire facto ff ego rome nui m is + Ego rotis in , sac fin a suni somis Fuzur 18 De Toma II II mani illi landulli auscari que fur y firmaure; 1 - Sun II man ich gungern gunger nafallou grit fier 7 homaner. Tout HI II manuit quarmeris caquelles sources qui d'antolus mal marini que s'envant per acti Stagnante scornoli arq: p ambrolis que sounce relivi; arrado nota factor satares me fue progue ffi. क perrul no lepli polt एकार्य coptin - अर्थ:

InterPARES Trust出版

The "Al in the Middle Age" study:

Parchments Study and Arrangement by Appearance-Based Recognition

TEAM

Archivio di stato di Milano:

Benedetto Luigi Compagnoni – *former Director*, Carmela Santoro - *Archivist, Deputy director*,

Archivio di Stato di Novara

Davide Bruno De Franco - Archivist, Director

Archivio di Stato di Ascoli Piceno

Emanuele Tedeschi, director

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:

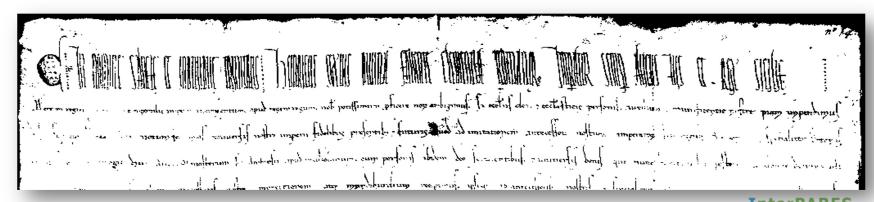
Joseph T. Tennis - *Professor, Associate Dean for Faculty Affairs, Executive Director of Administrative Services*





NEW PATHS OF RESEARCH

- Though studied and edited since the XVIII century, this extensive materials have never 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.







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 charatceristics (shape, position, etc.)





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.











FROM ML TO DL

MACHINE LEARNING -----CAR **NOT CAR INPUT FEATURE EXTRACTION** CLASSIFICATION **OUTPUT** DEEP LEARNING **NOT CAR**

FEATURE EXTRACTION + CLASSIFICATION

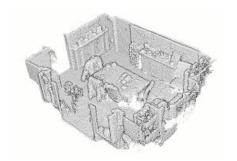
INPUT



OUTPUT

FROM ML TO DL





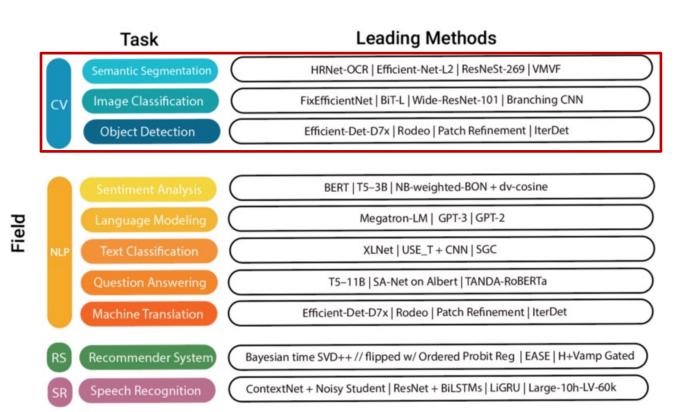






IMAGE UNDERSTANDING

Classification Instance **Object Detection** Classification + Localization Segmentation CAT, DOG, DUCK CAT, DOG, DUCK CAT CAT Multiple objects Single object

What about archives?



MANAGING PUBLIC DATASETS





BUILDING UP A NOTARIAL MATRICULA BEFORE AD 1350

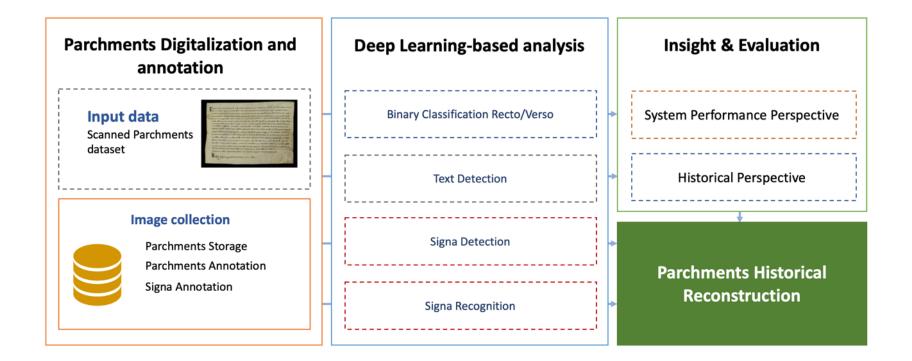
- One of the aims of the project is to use the AI to build up a Matricula of Milanese notaries of the twelfth and thirteenth centuries (in separate steps).
- The progressive investigation of all the parchment documents of the ASMi will lead to a complete list of notaries and their identification.
- The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



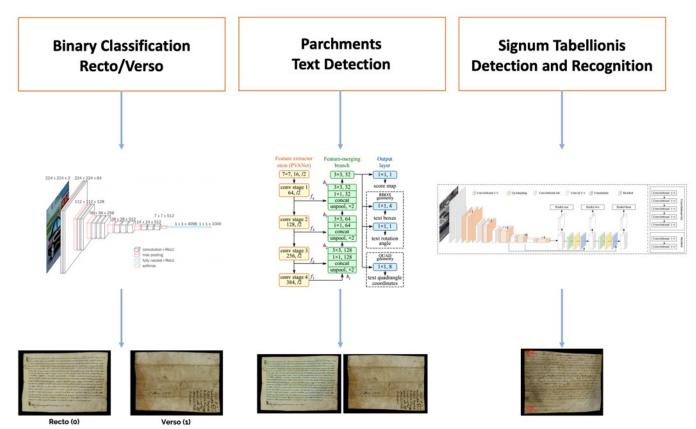


DATASET COLLECTION

- 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)



DEEP LEARNING PIPELINE







BINARY CLASSIFICATION RECTO/VERSO

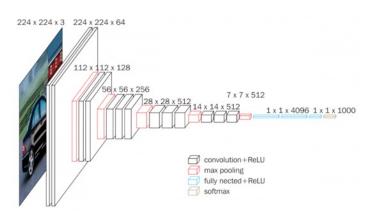
VGG16 DEEP NEURAL NETWORK



Recto



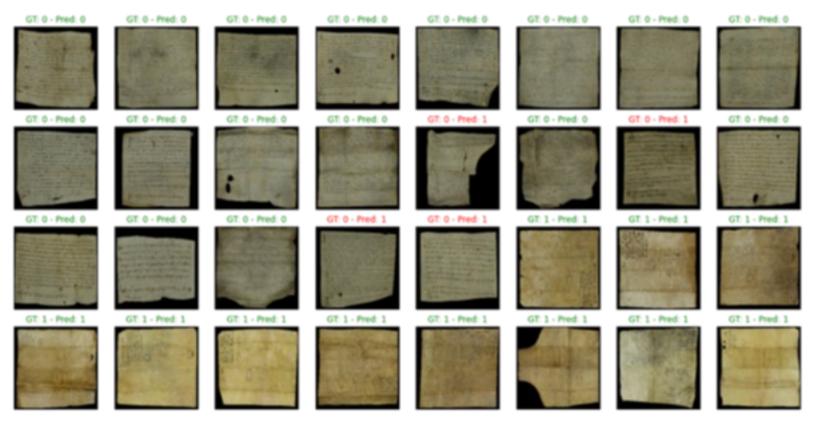
Verso



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



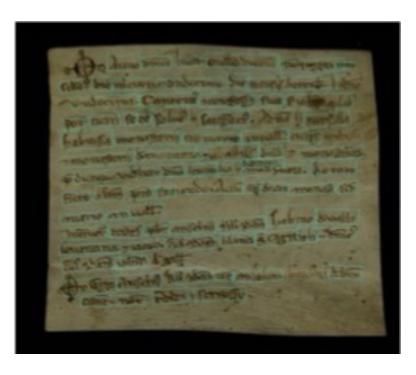
BINARY CLASSIFICATION RECTO/VERSO: RESULTS

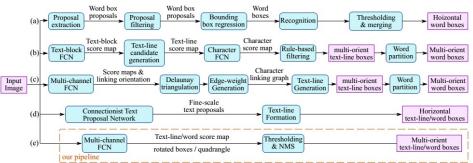






PARCHMENTS TEXT DETECTION





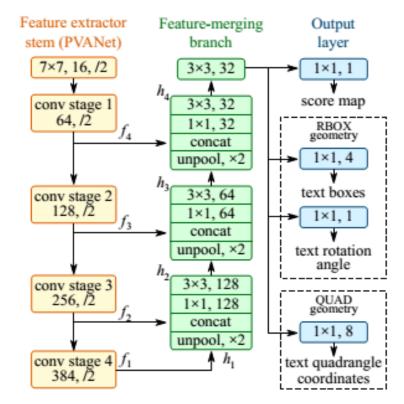
Xinyu Zhou., et all, A. (2017). EAST: An Efficient and Accurate Scene Text Detector. arXiv:1704.03155v2.



PARCHMENTS TEXT DETECTION: EXPERIMENTS

Word detection

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



InterPARES Trust⊞



PARCHMENTS TEXT DETECTION: RESULTS

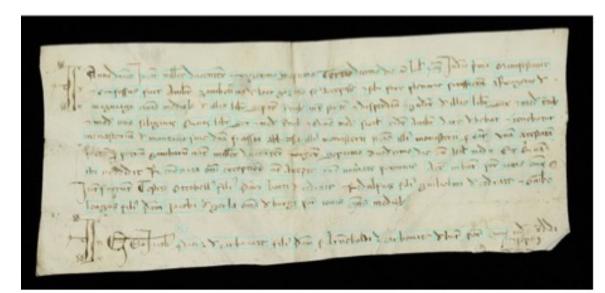
The board point of sold form under home depress to also participated but and pated of the tensor to the sound of the sound

of my grated in light which elight .





PARCHMENTS TEXT DETECTION

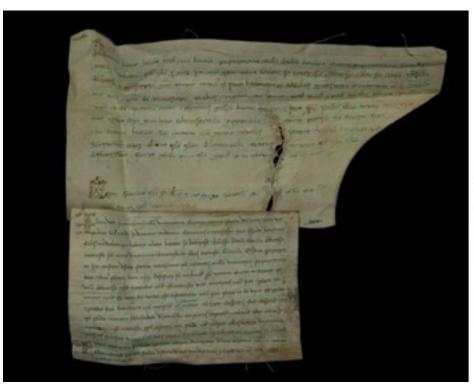


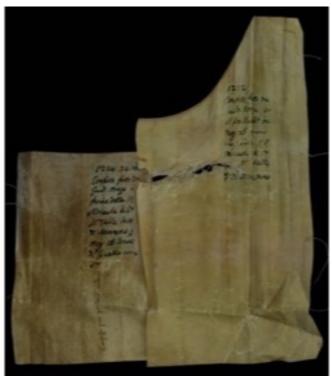


InterPARES Trust #188



PARCHMENTS TEXT DETECTION



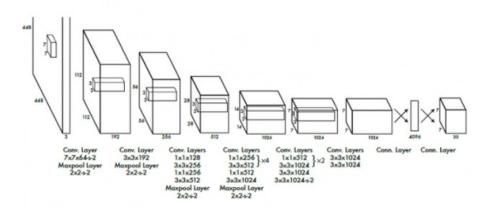






SIGNA TABELLIONIS DETECTION AND RECOGNITION

The period to the period occurs of the period of the period of the period occurs and the period occurs of the period occurs occurs of the period occurs of the period occurs occurs of the period occurs occurs occurs of the period occurs occurs occurs occurs of the period occurs occurs occurs occurs of the period occurs occurs occurs occurs occurs of the period occurs occurs occurs occurs of the period occurs occu



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).

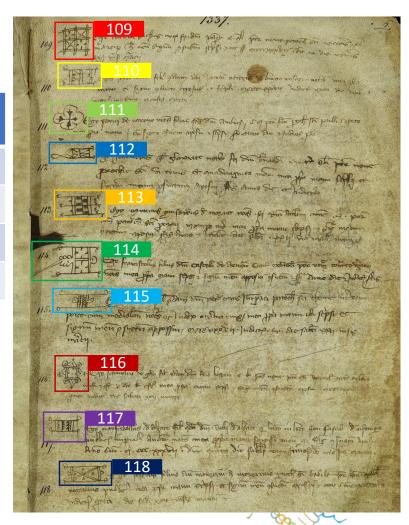




NOTARIAL MATRICULA ANNOTATION

2768 signa

ID	NAME	HISTORICAL INFORMATION
•••		
109	notary 109	
110	notary 110	



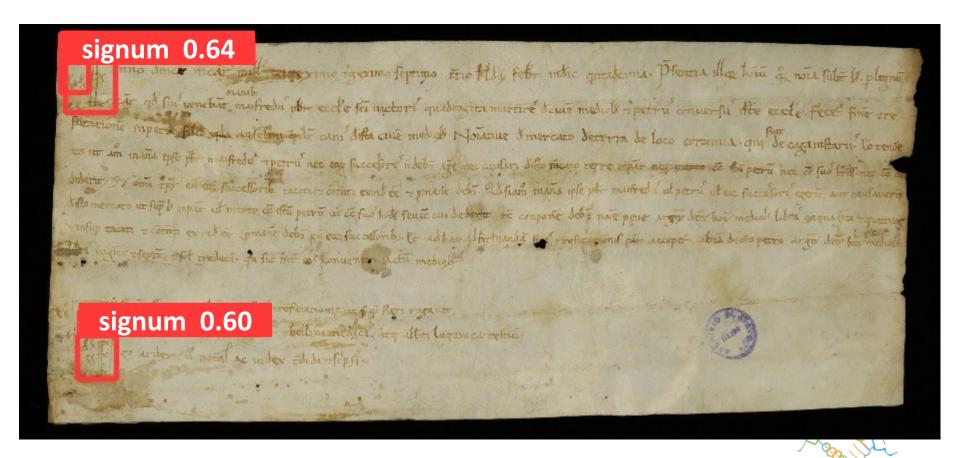
SIGNA TABELLIONIS DETECTION AND RECOGNITION: RESULTS

The man said man contrate contrate occurrent les quartes same security on sur. Indic quintes dense les In a bonos love ques note fate - hundhur puncture with the west tremes impollione person got augapellu. Saut meduly. Sona lobe albare official cretic ve salten for actione surlume cretic or nonce la latte for gotte comana Sent medioli Novature d'un porcione deven ille dans que le Flore berte vid' trevero qua e lus por medie sudinifo. pier sur pono quartes par ipi deme qua e desnouces & benzano perero ware du bon medioti contre liber duocenti curatir exquel Sti this al bar at por the code for farm of are mellion lose persons at fur liedels and two come mile due parties the sign adultion halinder lunn no plind vened it soure despoten liber course quider oufer michaelle pline utasses de polos ire ronde in mento i colorum ile medie d'inicibi un dime depento iminuto inque adelhe rous ales wide &b or up coche . Is up abbay plor no foluerer they do ipmo the he hape Il spe person tora grotte ruly & but folucie es the do . It with for muliaelle ut a var des term no Poluctie ille su he we light famile St abere we perwed wer minute. por private stine its In one person It facer tale car the about aspar the cutie for facer dithe some quite and of landrate with ping we hapt why with for michaelle plant to up tolure or the distinct at his Die ourder too perwe. To enous go up alder utilling the due reprine has chit In tally time he supilly by Shoe somplade other Studies perior us ar union's retained winter politic sac well proce outgrown pur varded talloldo any where got of the beauther Sille line. I me he me con whent her Vad due bust une inout leptoliste her Inches retter pare ded alberre fil un pers Azo moderi normi sadpel har aut ito "soph

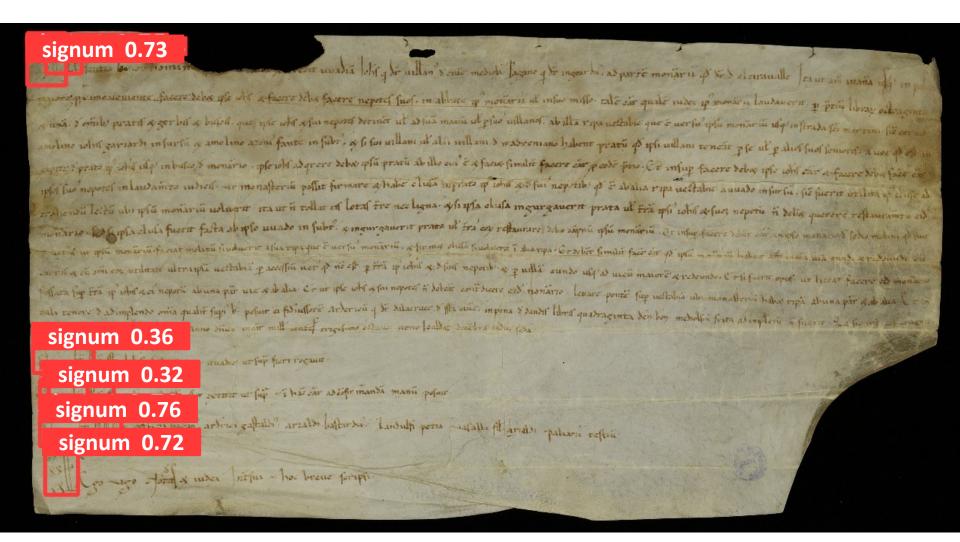


SIGNA TABELLIONIS DETECTION AND RECOGNITION: RESULTS

For each prediction is showed class (signum) and confidence



SIGNA TABELLIONIS DETECTION AND RECOGNITION: RESULTS



OTHER APPLICATIONS

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

- Recognizing the peculiar system of writing of every single notary.
- Training Handwritten Text Recognition tools to read text and abbreviations
- Analyzing archival notes on the back and retrace previous archival arrangements or uses of groups of documents.
- Recognizing recurring images or other features in huge series of documents.
- Identifying common patterns in manuscript maps or drawings.
- Opening original & relevant AI Data Sets.
- And many others...

















InterPARES Summer School Symposium – San Benedetto, Friday 7 July Artificial Intelligence to support the ongoing availability and accessibility of trustworthy public records - an overview of InterPARES Trust Al

Al for Multimedia understanding from computer vision to deep learning



Prof. Pierluigi Feliciati Prof. Emanuele Frontoni

University of Macerata