www.saintgeorgeonabike.eu

How Artificial Intelligence/Machine Learning methodologies are used in the “Saint George on a Bike” project: Describing cultural heritage imagery

Maria-Cristina Marinescu, Barcelona Supercomputing Center

Co-financed by the Connecting Europe Facility of the European Union

Table of Contents

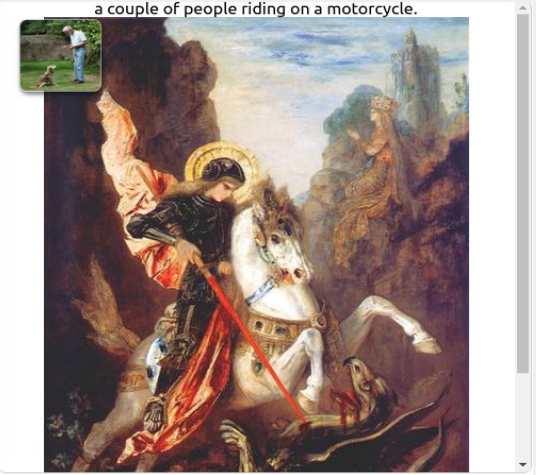
01 How Saint George on a Bike came to be

Motivation

- Focus on *cultural heritage* as a way to understand our past, approach the future, find inspiration, innovate
  - An area with a lot of (meta-)data quality issues
  - Rich, quality metadata enhance access to, and enable the reuse of cultural heritage digital content
  - Good descriptions enable research, education, cultural, social projects, and can improve web accessibility for the blind

Goal

- Contextualize the objects and image composition to ultimately endow AI with culture, symbols and tradition insight (and generate rich metadata)
  - Focus on (figurative) paintings of XII-XVIII centuries (especially iconography)



01 Basic approach

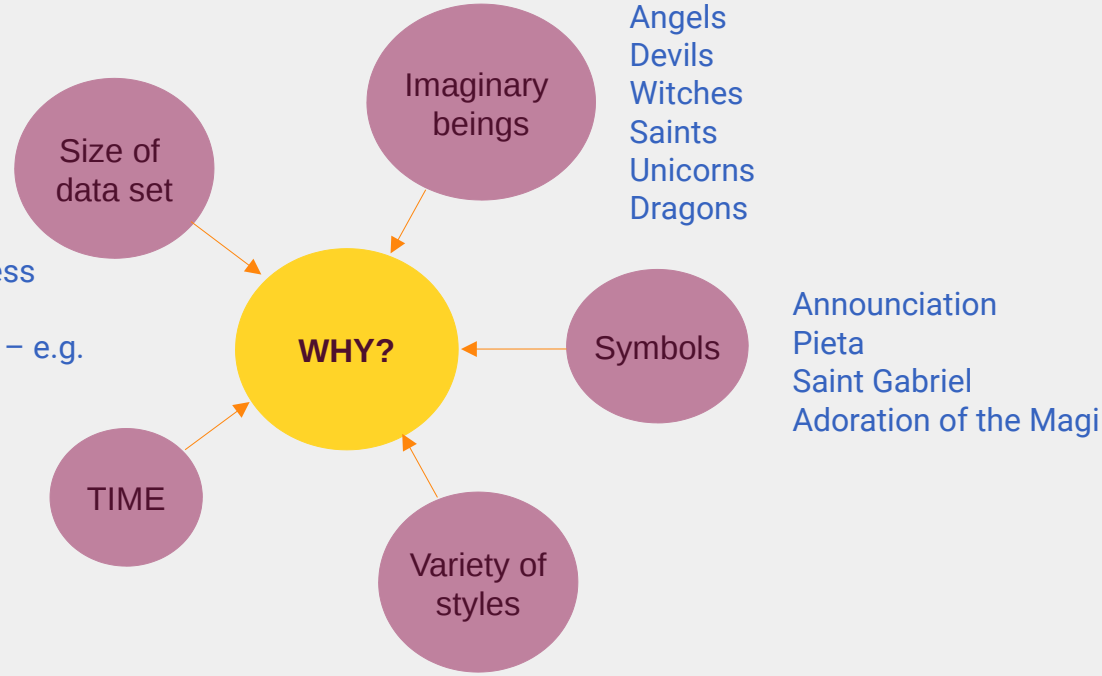
- Use jointly techniques from different (AI) fields to apply them to images or (image, text) pairs
  - Deep learning
  - Natural language-based models
  - Semantic metadata extraction and reasoning

01 The main challenge

Current approaches are very successful for everyday images, but fail for cultural heritage.
Work well for ***recent pictures***, given that the were trained on ***very large datasets*** with these characteristics.

And cultural Heritage?

Old objects not in use anymore – e.g. inkwell, printing press
Objects with different shapes in the past – e.g. plow
New objects, different but with similar shape as old ones – e.g. cell phone vs book
Unusual actions for everyday life, e.g. man killing a horse



01 Use cases

Relevant to our Europeana partner:

General service for enriching collections
Ingesting results from general enrichment service into Europeana
Search based on enrichment

Populate a crowdsourcing tool with candidate enrichments
Upload in data sharing platforms
Possibly: Browsing based on enrichment

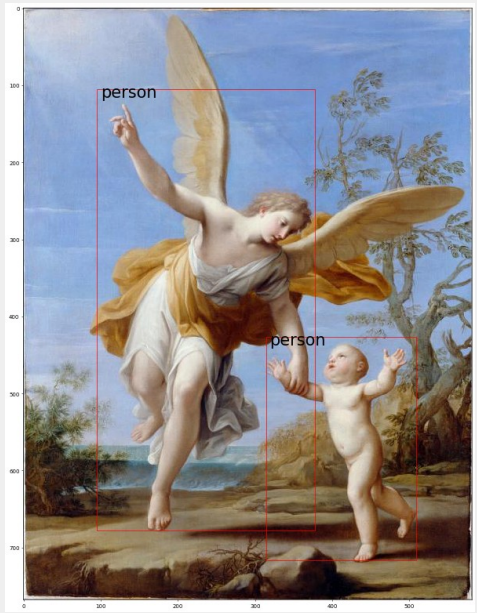
01 Knowledge sharing and capacity building

- Close collaboration with Europeana to develop capacity for digital transformation

- Interviewed for theTask Force on AI in relation to GLAMs

- Knowledge transfer via conferences, webinars, notebooks
    - Participation in EuropeanaTech x AI webinar series
    - Time Matrix seminar
    - PATC Big Data seminar series (organized by PRACE): Multidisciplinary research and data analytics: the cultural Heritage case
    - Publications

02 Deep learning for object detection



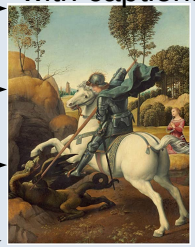
europeana

WIKIART
VISUAL ART ENCYCLOPEDIA

ARKYVES
ICONCLASS

WIKIMEDIA COMMONS

6K paintings with captions



TIME MATRIX

MS-COCO pre-trained model

Object detection / correction (Neural Network + Transfer learning)

Set of classes for transfer learning

Extract classes related to *Person*

E.g. **+** sword, dragon, mitre

Filter out anachronic COCO classes

E.g. **-** motorcycle, cell phone

Object detection

Generate candidate labels for each bounding box

02 Tackling object mis-identification by placing it in time

Assume we identified the following bounding boxes:
BB1: *teddy bear*
BB2: *bike*, horse, zebra
BB3: *baseball bat*, sword
BB4: dog
BB5: person

Anachronisms:
Teddy bear → woman
Bike → horse
Baseball bat → stick

Painting from 1506

TIME MATRIX

class label	first-time use

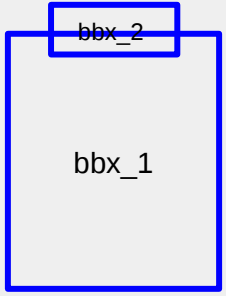
E.g. teddy bear appears in 1905

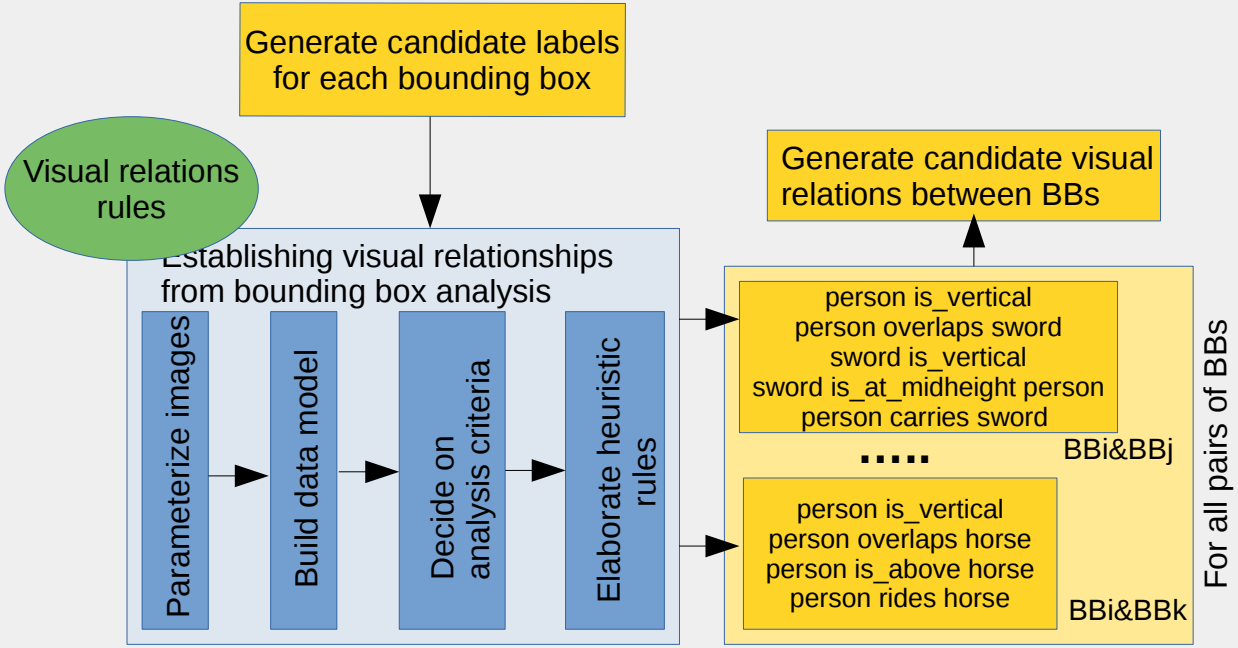
E.g. motorcycle appears in 1894

What we would like to get back?
BB1: princess
BB2: horse
BB3: sword
BB4: dragon
BB5: Knight (St. George)

02 Visual relations via bounding box analysis

Inferring visual relations
- Between detected objects
- In images w complex scenes
- Using bounding box analysis





if (bbx_1= '**crucifixion**' and bbx_1 is_vertical
and (bbx_2 = '**crown of thorns**'
and overlap with bbx_1
and is in upper region of bbx-1
and pairwise-proportions are respected))
then *'crucifixion'((person)) is ((jesus_christ))*

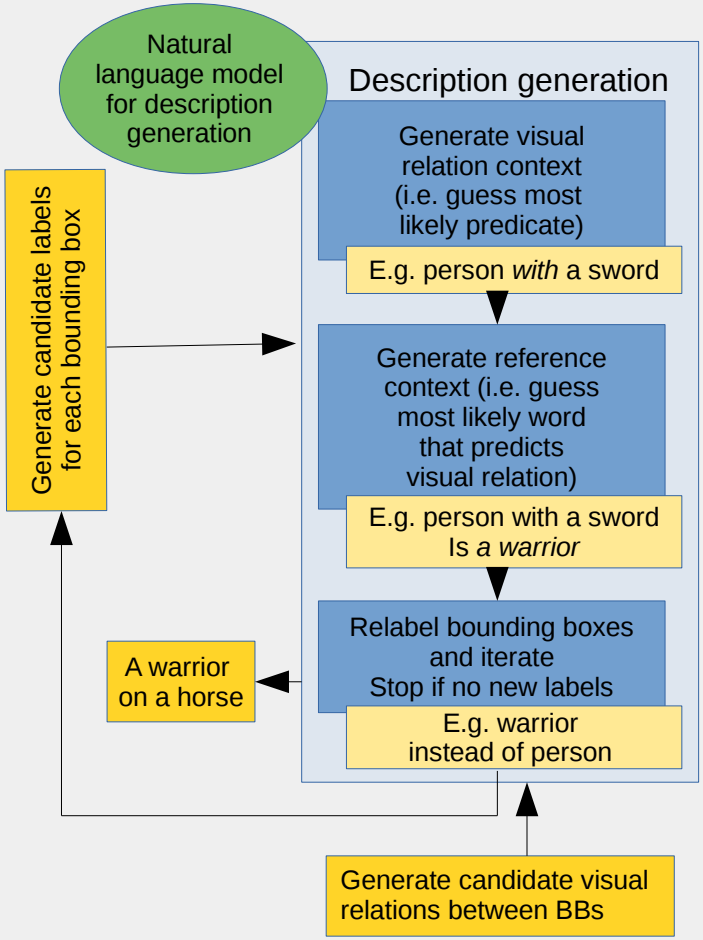
Heuristic rules based on different criteria: proportions, object location, overlap, orientation, etc

02 Refining object detection via a language model

A person wearing a crown is a _____ KING/QUEEN

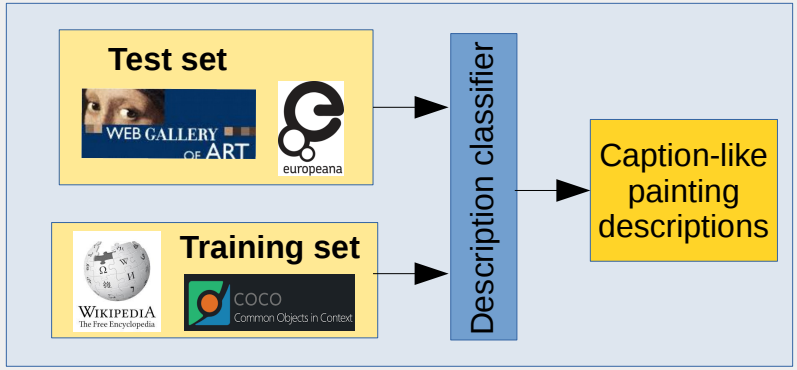
A person riding a horse is a _____ RIDER

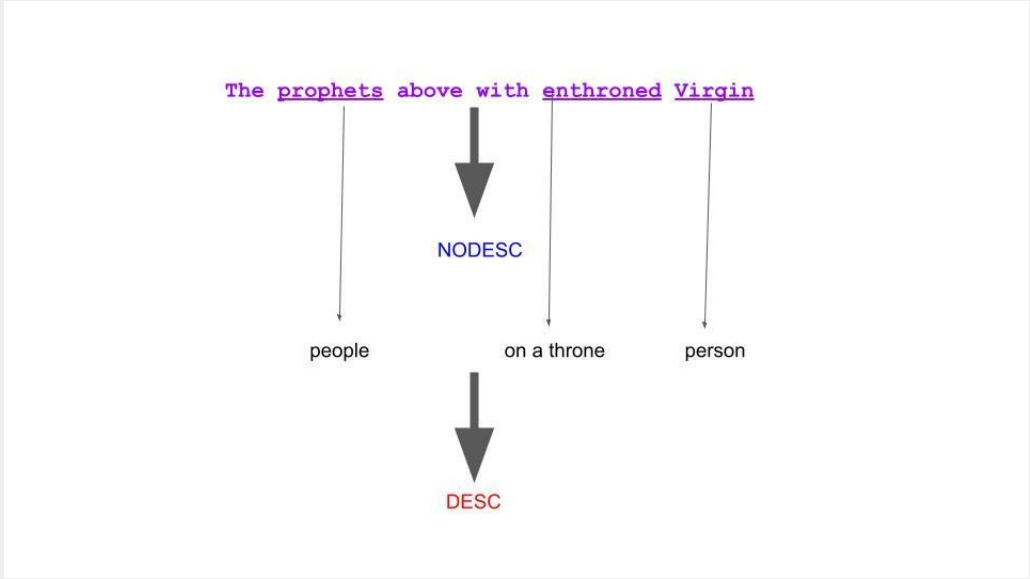
-Transformer-based language model
- Model attempts to predict the value of a masked word
- Prediction based on semantic context provided by the other, non-masked, words in the sequence



02 Caption classifier to extract descriptive content

- Descriptions are rarely about what can be seen
- Useful for caption generation via deep learning
    - with CNN
    - with LSTM
    - with transformers





03 Challenges so far

Small dataset of paintings by data mining standards

- Some classes are represented only in a few images; differences in style, medium, color
- Can't produce more paintings when needed!

...which requires complementary techniques to contextualize appropriately (e.g. detect anachronisms, imaginary objects, (unusual) actions

...top-down knowledge representation and reasoning can provide common sense

Poor metadata for training

- Labeled bounding boxes
- Descriptions of visual content
- Labeled visual relationships

→ Crowdsourcing can be fundamental!

Evaluation - quantifying enrichments quality and usefulness to the user

www.saintgeorgeonabike.eu

Thank you!

maria.marinescu@bsc.es

Co-financed by the Connecting Europe Facility of the European Union