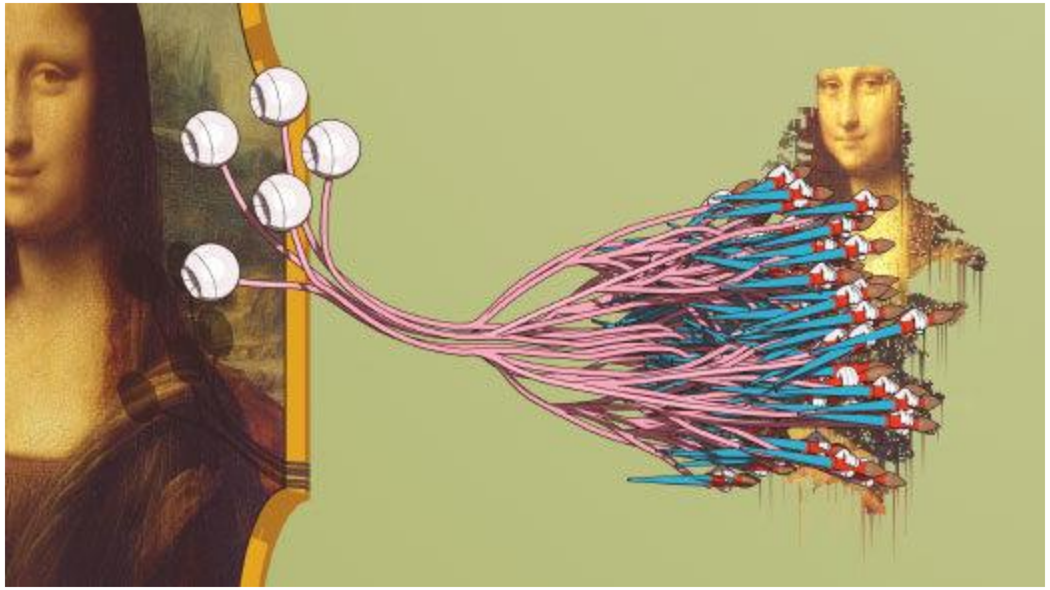
MonaLIA 2.0

Fusion of deep learning and semantic reasoning for image recognition and enrichment of Joconde database metadata



[DVDP](#) for Quanta Magazine

Motivation to Enrich Joconde Database

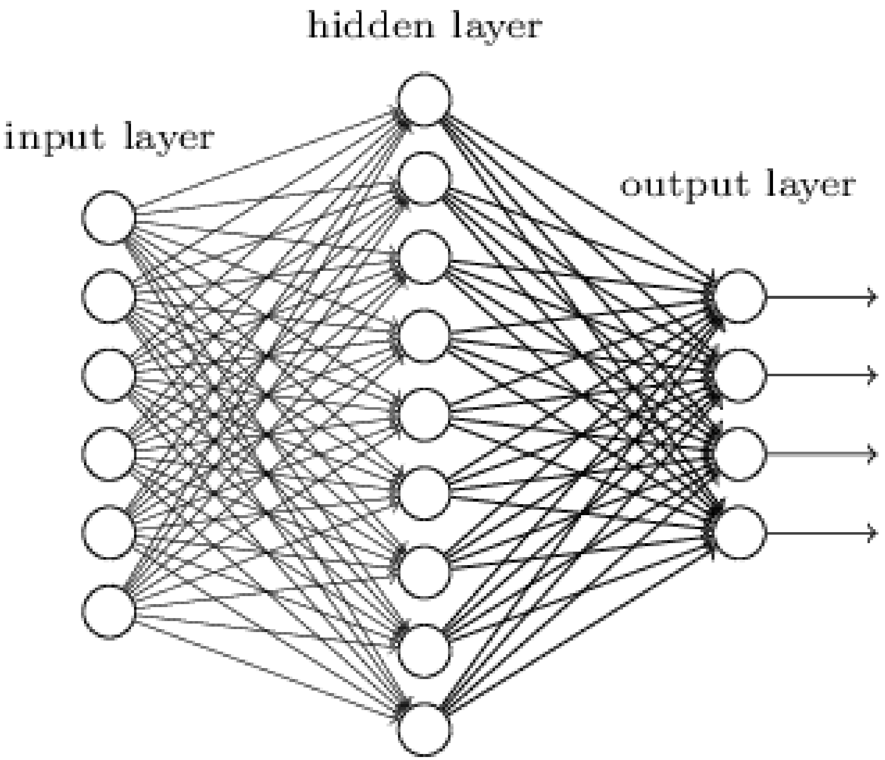
PORTRAIT DE MONA LISA (1479-1528) ; DITE LA JOCONDE by Leonardo Da Vinci

- ~300 000 illustrated artwork records from the French museums.
- RDF metadata about the artwork content and properties (media, author, museum, etc.).
- content metadata is organized according to Thesaurus Iconographique (Francois Garnier, 1981)

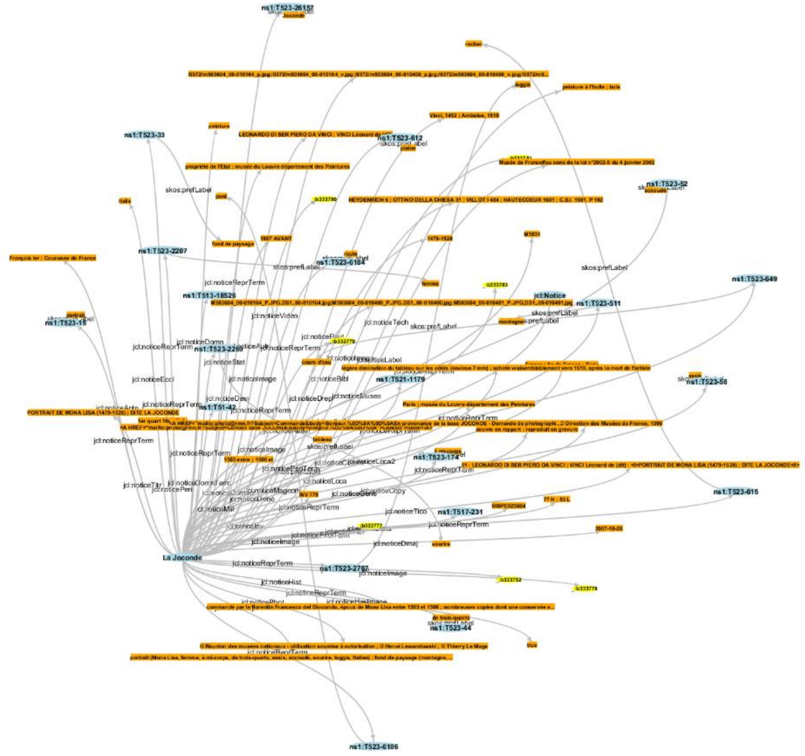
- **Can the digital artwork collections be automatically enhanced by combining Machine Learning and Knowledge Representation & Reasoning?**
- **Can annotation of the artworks be automated or semi-automated?**
- **Can the search results be ranked by the visual relevancy of a search criteria?**

Combining Strength of Two Methods

Deep Learning from unstructured data such as images



Semantic Reasoning and querying from semantic metadata

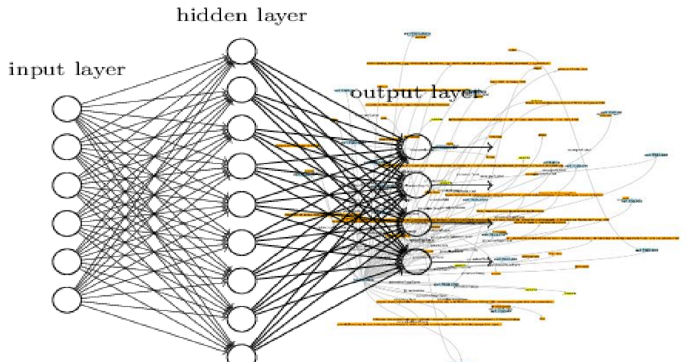


Combining Strength of Two Methods

Deep Learning from unstructured data such as images

Semantic Reasoning and querying from semantic metadata

- reason to **prepare training and validation sets** & labels



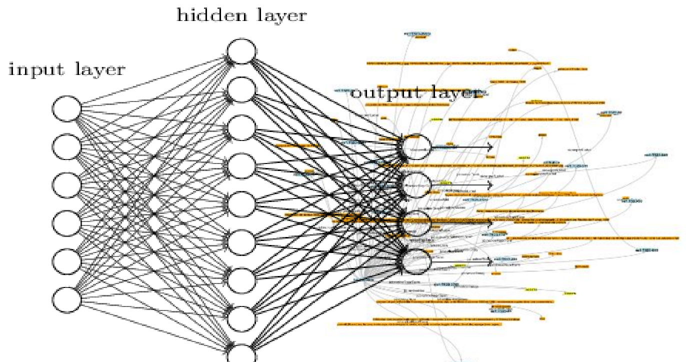
Combining Strength of Two Methods

Deep Learning from unstructured data such as images

- learn to classify image content to **annotate artworks** with efficiency

Semantic Reasoning and querying from semantic metadata

- reason to **prepare training and validation sets** & labels



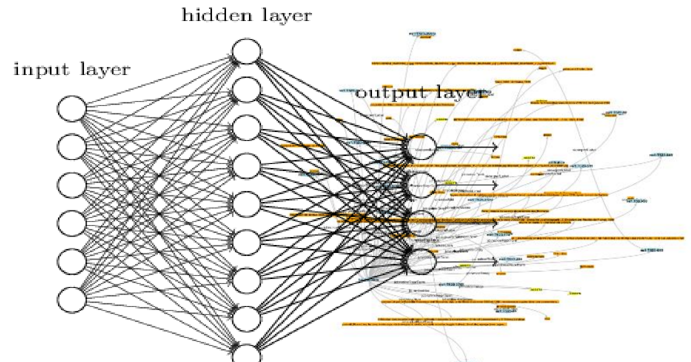
Combining Strength of Two Methods

Deep Learning from unstructured data such as images

- learn to classify image content to **annotate artworks** with efficiency

Semantic Reasoning and querying from semantic metadata

- reason to **prepare training and validation sets** & labels
- reason to **extend quality of the metadata** that is incomplete or noisy



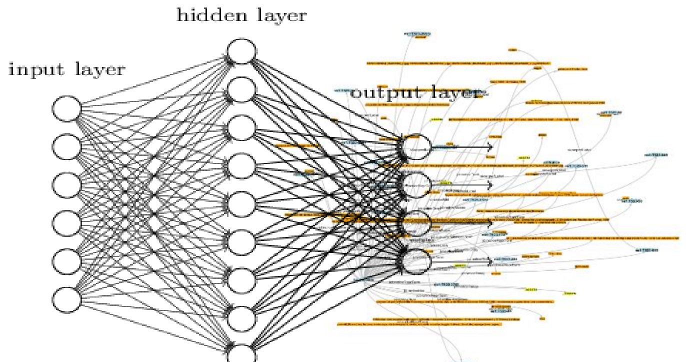
Combining Strength of Two Methods

Deep Learning from unstructured data such as images

- learn to classify image content to **annotate artworks** with efficiency
- learn to **extend the existing metadata** with quantitative measures of object relevance

Semantic Reasoning and querying from semantic metadata

- reason to **prepare training and validation sets** & labels
- reason to **extend quality of the metadata** that is incomplete or noisy



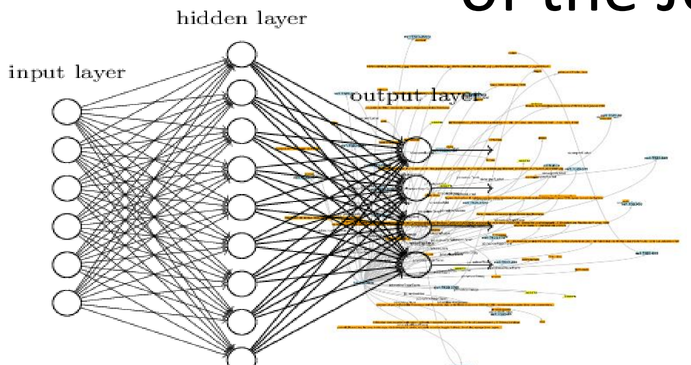
Combining Strength of Two Methods

Deep Learning from unstructured data such as images

- learn to classify image content to **annotate artworks** with efficiency
- learn to **extend the existing metadata** with quantitative measures of object relevance

Semantic Reasoning and querying from semantic metadata

- reason to **prepare training and validation sets** & labels
- reason to **extend quality of the metadata** that is incomplete or noisy
- reason to **improve searchability** of the Joconde database



Conclusions of MonaLIA 1.0

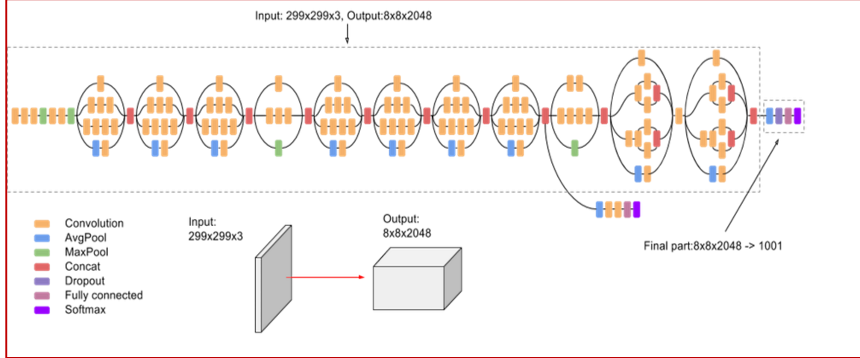
- Deep Learning Transfer works even from non cultural collections
  - The model gives good accuracy results even on the object categories that it hadn't been pre-trained on.
- The limitations are mostly from the labeled image availability.
  - Adding images from other museum collections and Wikipedia can increase the training set.
- Joconde database metadata can be improved by filling the missing values and structuring some fields
  - Missing values can be inferred by machine learning algorithms.
  - Introducing the ontologies on techniques, material, denomination, preservation state can help to deeper explore the relations between all possible structured representations

MonaLIA 2.0 Approach

- SPARQL+RDFS+SKOS on metadata to extract training and test subsets of images

```
prefix skos: <http://www.w3.org/2004/02/skos/core#>
prefix jcl: <http://jocondelab.iri-research.org/ns/jocondelab/>


select (replace(group_concat(?topCategory_label; separator="+" ), "\\+$", "")  as ?label)
       (sample(?noticeRepresentation) as ?repr)
       (sample(?imagePath) as ?image_path)
       (sample(?noticeReference) as ?ref)
where
{
       values (?topCategory_label) { ("chien"@fr) ("cheval"@fr) } .

       ?topCategory   a jcl:Term;
                        skos:prefLabel ?topCategory_label;
                        skos:narrower* | skos:related/skos:narrower* ?subCategory.

        ?subCategory  skos:prefLabel ?subCategory_label.

        ?notice jcl:noticeReprTerm ?subCategory;
                jcl:noticeHasImage true;
                jcl:noticeRef ?noticeReference.
}
group by ?noticeReference
```

- Train Multi-Label Deep Learning classifier



- Apply trained model and extend metadata

```
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix jcl: <http://jocondelab.iri-research.org/ns/jocondelab/> .
@prefix t: <http://data.culture.fr/thesaurus/resource/ark:/67717/> .
@prefix ml: <http://ns.inria.fr/monalia/> .

ml:classifierRepresentedSubject a rdfs:Class ;                    ml:vocabID "REPR" .
ml:classifier40_classes rdfs:label "40_classes" ;
                 rdfs:subClassOf ml:classifierRepresentedSubject .

<https://jocondelab.iri-research.org/data/notice/0000055013> ml:imageClassifier [ a
ml:classifier40_classes ;
      ml:detected [ a t:T523-6519 ; ml:score 0.8102 ],
                  [ a t:T523-6209 ; ml:score 0.0219 ],
                  [ a t:T523-175 ;  ml:score 0.3843 ],
                  . . .
                  [ a t:T523-2037 ; ml:score 0.0121 ].
```

- SPARQL on extended metadata to search the database

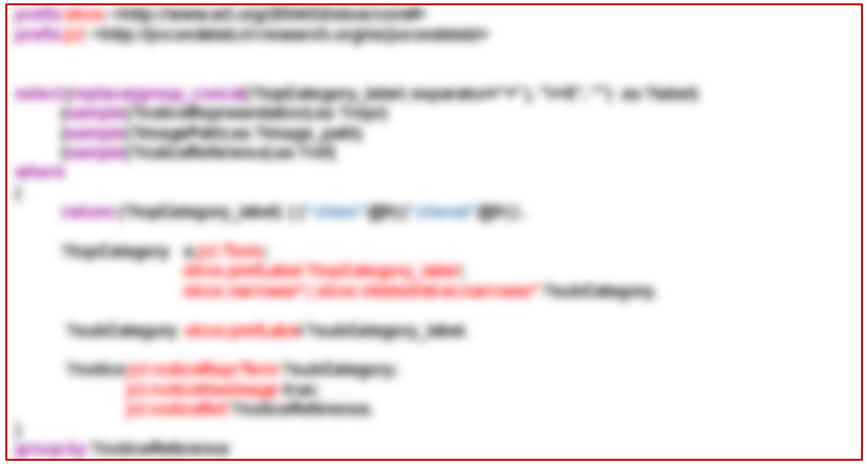
MonaLIA 2.0 Approach

- SPARQL+RDFS+SKOS on metadata to extract training and test subsets of images
  - label training and test sets including the “narrower” categories according to Garnier Thesaurus
  - create “missing” links between some categories
  - balance number of training images per class
  - filter out certain categories and images
- Train Multi-Label Deep Learning classifier
- Apply trained model and extend metadata
- SPARQL on extended metadata to search the database

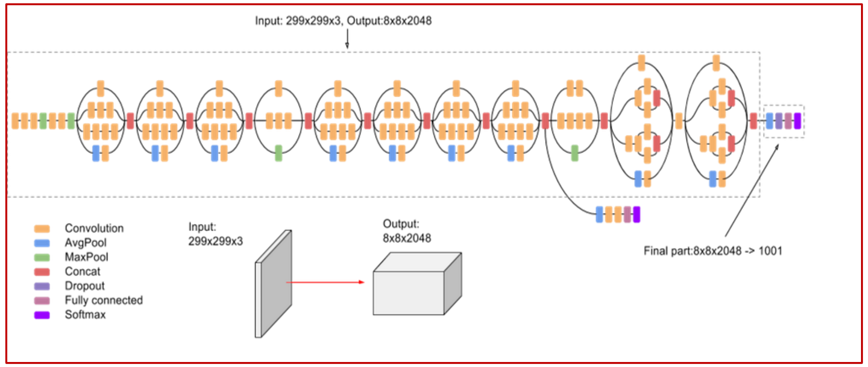
```
prefix skos: <http://www.w3.org/2004/02/skos/core#>
prefix jcl: <http://jocondelab.iri-research.org/ns/jocondelab/>


select (replace(group_concat(?topCategory_label; separator="+" ), "\\+$", "")  as ?label)
       (sample(?noticeRepresentation) as ?repr)
       (sample(?imagePath) as ?image_path)
       (sample(?noticeReference) as ?ref)
where
{
       values (?topCategory_label) { (“chien”@fr) (“cheval”@fr) } .

       ?topCategory   a jcl:Term;
                        skos:prefLabel ?topCategory_label;
                        skos:narrower* | skos:related/skos:narrower* ?subCategory.

        ?subCategory  skos:prefLabel ?subCategory_label.

        ?notice jcl:noticeReprTerm ?subCategory;
                jcl:noticeHasImage true;
                jcl:noticeRef ?noticeReference.
}
group by ?noticeReference
```

MonaLIA 2.0 Approach

- SPARQL+RDFS+SKOS on metadata to extract training and test subsets of images
  - create labeled training and test sets including the “narrower” categories according to Garnier Thesaurus
  - create “missing” links between some categories
  - balance number of training images per class
  - filter out certain categories and images



- Train Multi-Label Deep Learning classifier
  - select state-of-the-art pre-trained CNN model
  - adapt the model to multi-label classification
  - fine-tune model on labeled training sets
  - optimize model hyperparameters for best performance

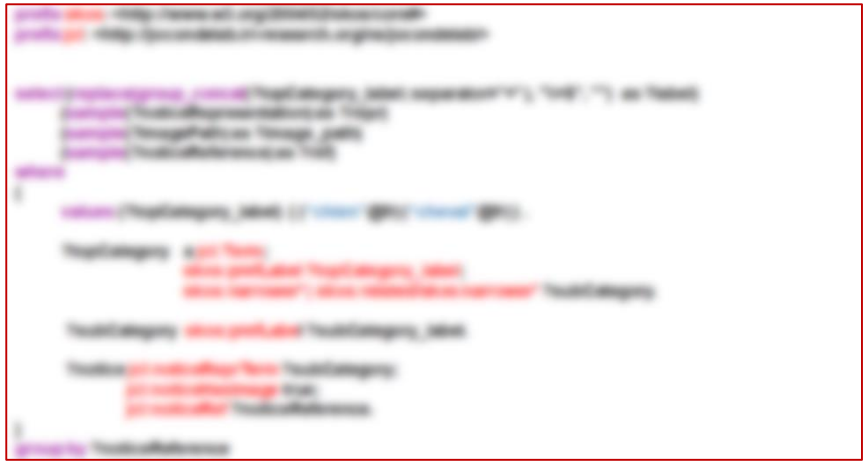


- Apply trained model and extend metadata
  - run all the images through the trained classifier
  - record the prediction score as RDF triples

- SPARQL on extended metadata to search the database

MonaLIA 2.0 Approach

- SPARQL+RDFS+SKOS on metadata to extract training and test subsets of images
  - create labeled training and test sets including the “narrower” categories according to Garnier Thesaurus
  - create “missing” links between some categories
  - balance number of training images per class
  - filter out certain categories and images
- Train Multi-Label Deep Learning classifier
  - select state-of-the-art pre-trained CNN model
  - adapt the model to multi-label classification
  - fine-tune model on artwork images
  - optimize model hyperparameters for best performance
- **Apply trained model and extend metadata**
  - **run all the images through the trained classifier**
  - **record the prediction score as RDF triples**
- SPARQL on extended metadata to search the database

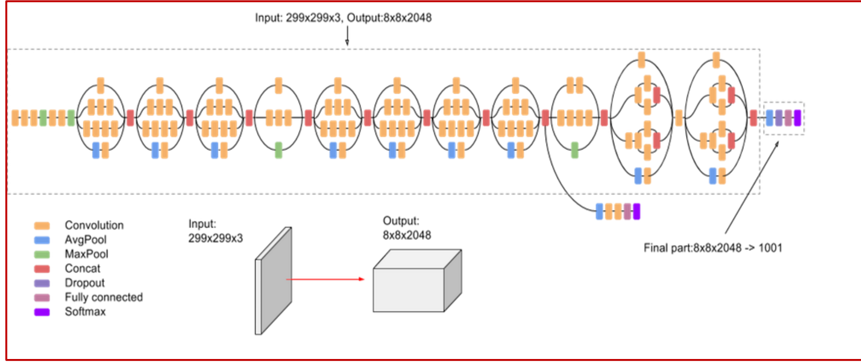


```
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix jcl: <http://jocondelab.iri-research.org/ns/jocondelab/> .
@prefix t: <http://data.culture.fr/thesaurus/resource/ark:/67717/> .
@prefix ml: <http://ns.inria.fr/monalia/> .

ml:classifierRepresentedSubject a rdfs:Class ;                ml:vocabID "REPR" .
ml:classifier40_classes rdfs:label "40_classes" ;
            rdfs:subClassOf ml:classifierRepresentedSubject .

<https://jocondelab.iri-research.org/data/notice/0000055013> ml:imageClassifier [ a
ml:classifier40_classes ;
      ml:detected [ a t:T523-6519 ; ml:score 0.8102 ],
                  [ a t:T523-6209 ; ml:score 0.0219 ],
                  [ a t:T523-175 ;  ml:score 0.3843 ],
                  . . .
                  [ a t:T523-2037 ; ml:score 0.0121 ].
```

MonaLIA 2.0 Approach

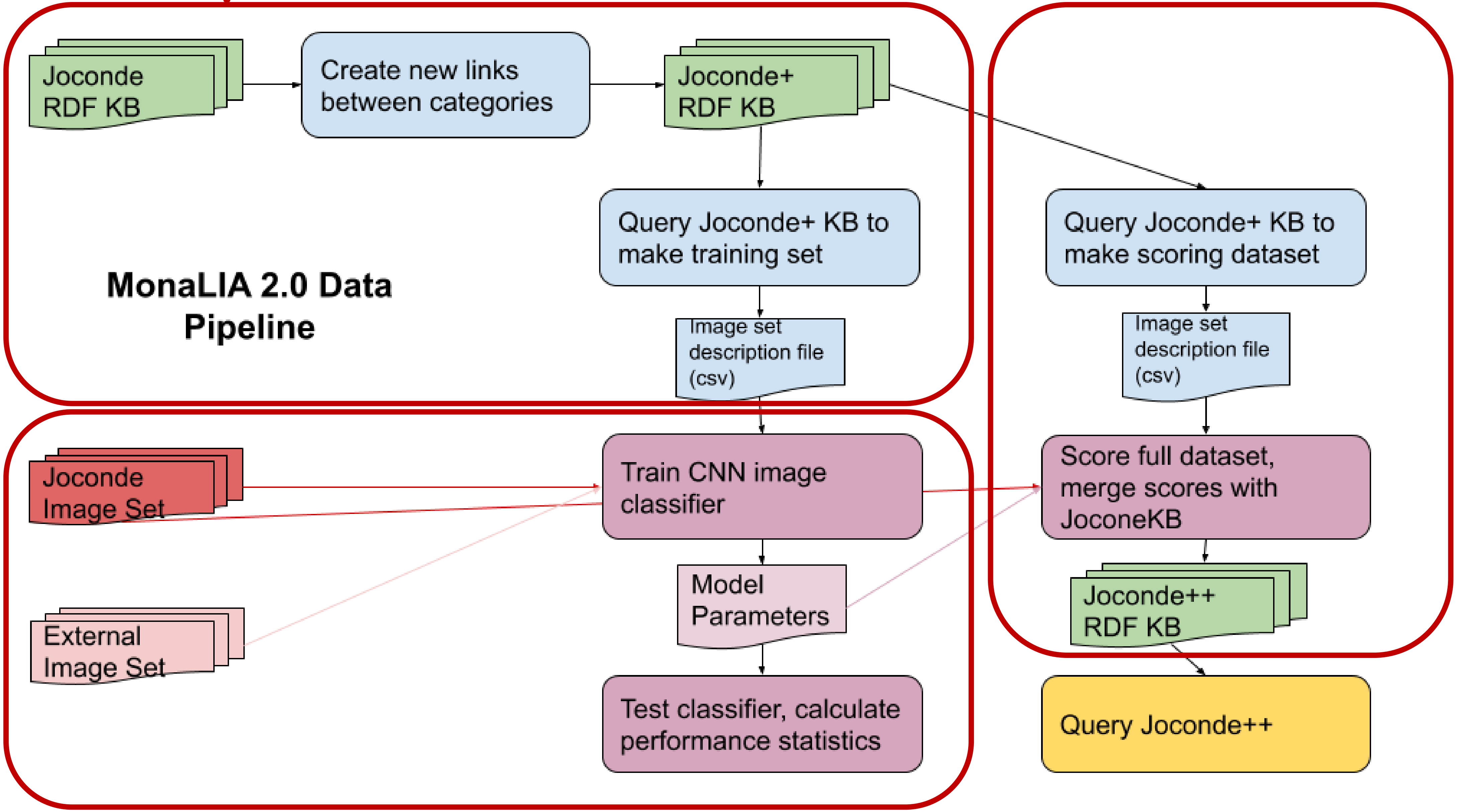
- SPARQL+RDFS+SKOS on metadata to extract training and test subsets of images
  - create labeled training and test sets including the “narrower” categories according to Garnier Thesaurus
  - create “missing” links between some categories
  - balance number of training images per class
  - filter out certain categories and images
- Train Multi-Label Deep Learning classifier
  - select state-of-the-art pre-trained CNN model
  - adapt the model to multi-label classification
  - fine-tune model on artwork images
  - optimize model hyperparameters for best performance
- Apply trained model and extend metadata
  - run all the images through the trained classifier
  - record the prediction score as RDF triples
- SPARQL on extended metadata to search the database

```
prefix skos: <http://www.w3.org/2004/02/skos/core#>
prefix jcl: <http://jocondelab.iri-research.org/ns/jocondelab/>


select (replace(group_concat(?topCategory_label; separator="+" ), "\\+$", "")  as ?label)
       (sample(?noticeRepresentation) as ?repr)
       (sample(?imagePath) as ?image_path)
       (sample(?noticeReference) as ?ref)
where
{
       values (?topCategory_label) { (“chien”@fr) (“cheval”@fr) } .

       ?topCategory   a jcl:Term;
                      skos:prefLabel ?topCategory_label;
                      skos:narrower* | skos:related/skos:narrower* ?subCategory.

        ?subCategory  skos:prefLabel ?subCategory_label.

        ?notice jcl:noticeReprTerm ?subCategory;
                jcl:noticeHasImage true;
                jcl:noticeRef ?noticeReference.
}
group by ?noticeReference
```




```
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix jcl: <http://jocondelab.iri-research.org/ns/jocondelab/> .
@prefix t: <http://data.culture.fr/thesaurus/resource/ark:/67717/> .
@prefix ml: <http://ns.inria.fr/monalia/> .

ml:classifierRepresentedSubject a rdfs:Class ;                    ml:vocabID "REPR" .
ml:classifier40_classes rdfs:label "40_classes" ;
              rdfs:subClassOf ml:classifierRepresentedSubject .

<https://jocondelab.iri-research.org/data/notice/0000055013> ml:imageClassifier [ a
ml:classifier40_classes ;
      ml:detected [ a t:T523-6519 ; ml:score 0.8102 ],
                  [ a t:T523-6209 ; ml:score 0.0219 ],
                  [ a t:T523-175 ;  ml:score 0.3843 ],
                  . . .
                  [ a t:T523-2037 ; ml:score 0.0121].
```

Data Pipeline



Linking Related Categories

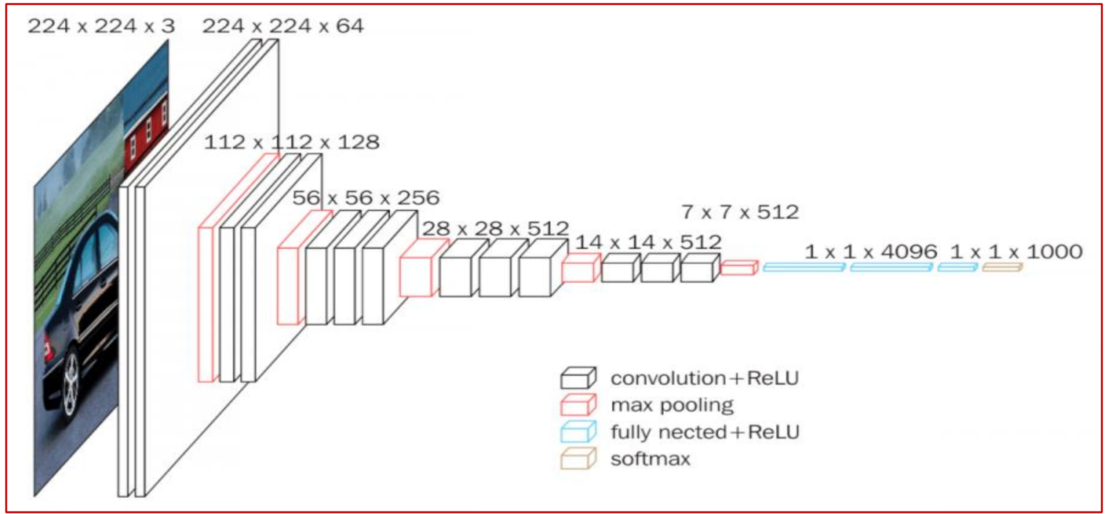
Lévrier tenant un lièvre dans sa gueule by Pierre Jules Mene

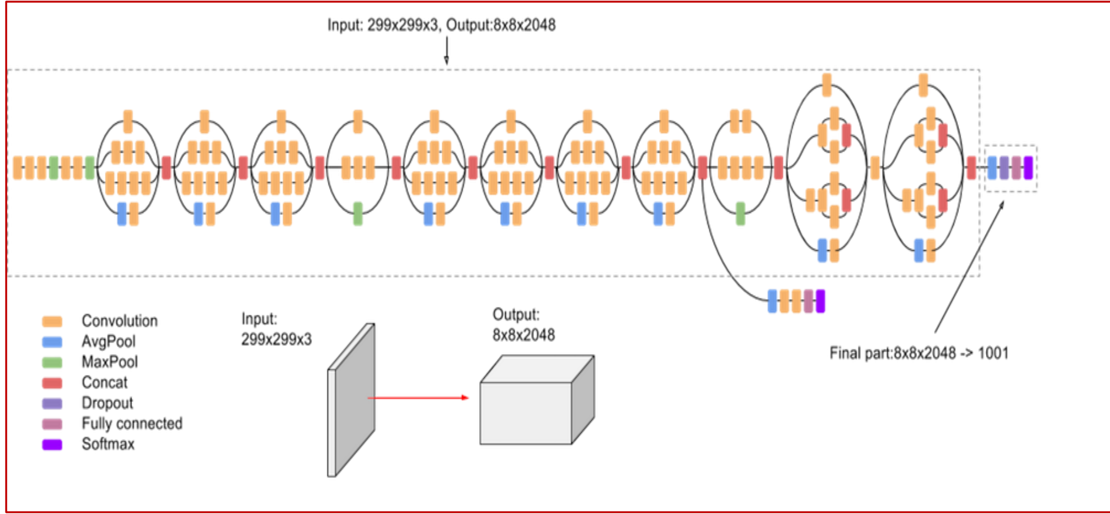
Sujet représenté	figure (Révolution française de 1848, soldat, cavalier, cheval, uniforme)

- Not all semantically similar categories are linked by Garnier Thesaurus (e.g. humans on the images not necessarily annotated by the terms in the “être humain”@fr hierarchy)
- Poor labeling -> poor model performance
- New RDF triples to link the categories that are not linked by the hierarchical thesaurus

```
@prefix skos: <http://www.w3.org/2004/02/skos/core#> .
insert { ?x skos:related ?y }
where {
        ?x skos:prefLabel "être humain"@fr.
        ?y skos:prefLabel "hiérarchie militaire“@fr.
}
```

Selecting Multi-label Classification Model



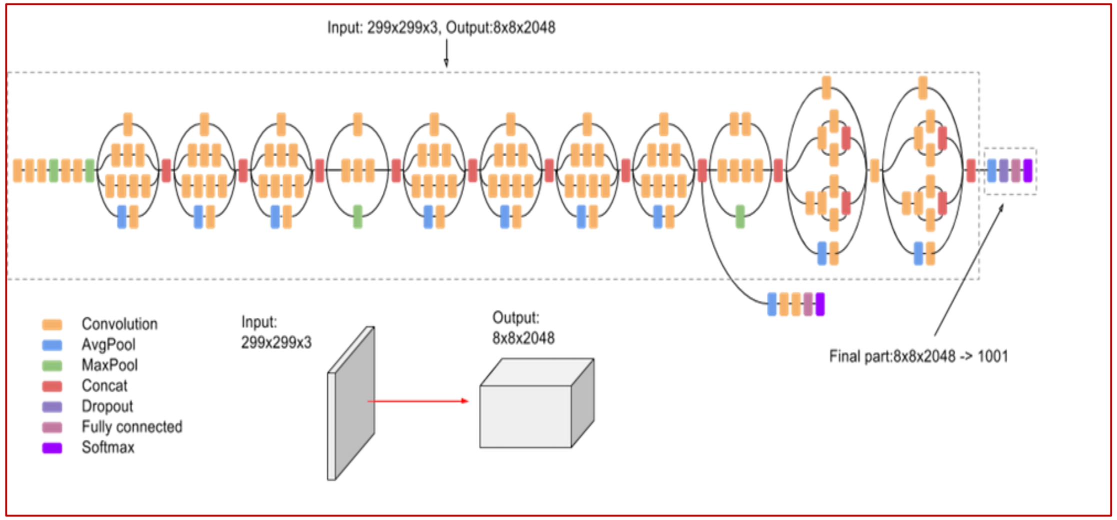


- VGG16 model
- 1st Runner-Up in ILSVRC2014
- 10 classes mAP = 0.78
- ~138 million parameters
- ~525 MB disk space
- 3h35m training time*

- Inception v3 model
- Inception v1 is a Winner in ILSVRC2014
- 10 classes mAP = 0.79
- ~4 million parameters
- ~96 MB disk space
- 3hr15m training time*

* Training time is benchmarked on 40 class multi-label classifier, 55 900 training samples on cluster node with 2x Xeon SP Gold 5115 @ 2.4 GHz CPU, 256 GB RAM with 2 GeForce GTX 1080 Ti GPUs cards connected with a PCIe gen3 16x interface, 11GB of RAM per card.

Tuning Classification Model for Best Performance



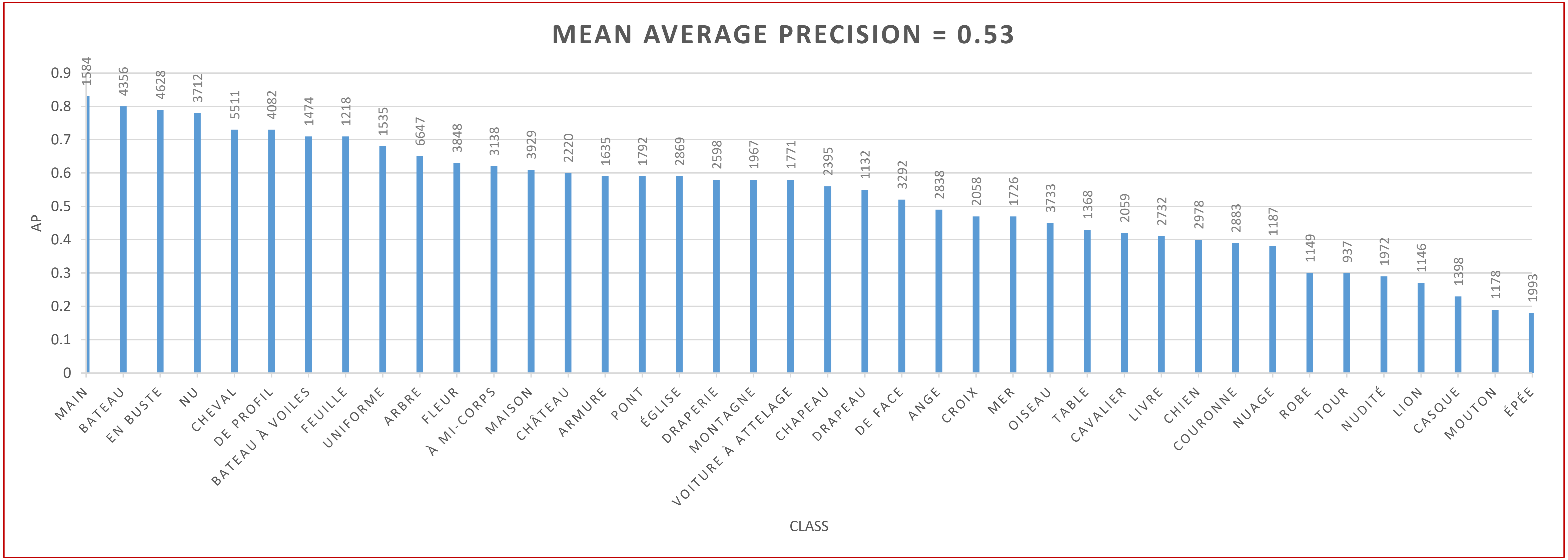
Hyperparameters

- Training mode: fine-tuning
- Dropout rate: 0.5
- Activation function: sigmoid
- Loss Function: Binary Cross Entropy
- Optimizer: Adam
- Initial learning rate: 0.001
- Training epochs: 20
- Learning rate decay schedule: reduce by 0.1 every 4 epochs
- Momentum : 0.9

- Inception v3 model
- Inception v1 in a Winner in ILSVRC2014
- ~4 million parameters
- ~95 MB disk space
- 3hr15m training time

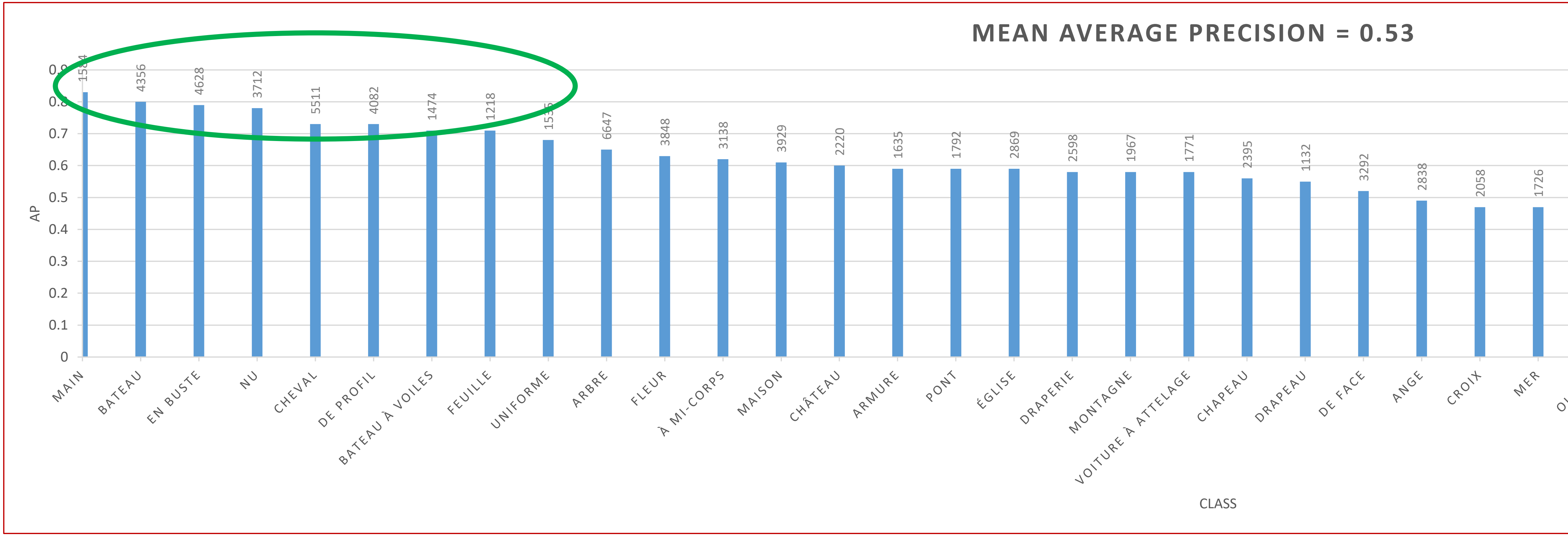
Multi-label Model Performance

- 40-class multi-label classifier
  - 40 of 100 categories from the Ministry of Culture list have adequate number of images for training



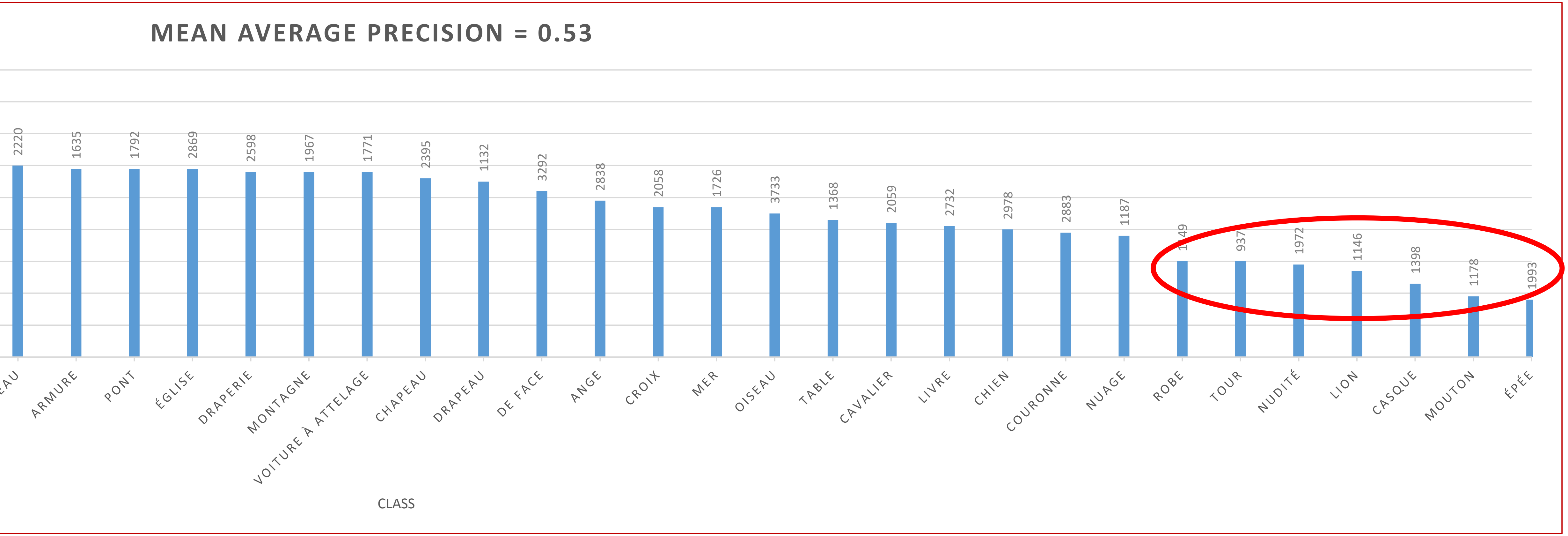
Multi-label Model Performance

- Some categories are easier to detect
  - general categories
  - large object



Multi-label Model Performance

- Some categories are harder to detect
  - relatively small objects
  - visually close to more prevalent categories



Extending Training Image Set

Study for Transportation Mural by William Sommer

- One of the limitations of object detection is labeled image availability
- 60 categories out of 100 from the Ministry of Culture list have insufficient number of images to train machine learning algorithms
- More contemporary objects are less represented
- Exploring other museum collections to increase the training set

Extending Training Image Set

Study for Transportation Mural by William Sommer

- One of the limitations of object detection is labeled image availability
- 60 categories out of 100 from the Ministry of Culture list have insufficient number of images to train machine learning algorithms
- More contemporary objects are less represented
- Exploring other museum collections to increase the training set

Collection	API	Number of available images	Number of labeled images
The Behance Artistic Media Dataset	SQLITE HTTP	78 687	8 338
Pinterest	HTTP		641
Cleveland Museum of Art	CSV HTTP	30 676	1 283
Smithsonian Museums: American Art Museum	SPARQL HTTP	3 900 000 (site) 47 133 (SPARQL)	In progress

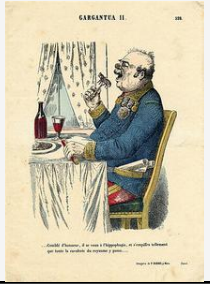
Generate & Query Extended Metadata

- Apply model to all images
  - classifier outputs prediction scores per category per image
- Extend metadata with prediction scores
  - store in RDF format
  - linked to artwork records
- SPARQL query on
  - object(s) +
  - classifier(s) +
  - prediction score

```
prefix skos: <http://www.w3.org/2004/02/skos/core#>
prefix jcl: <http://jocondelab.iri-research.org/ns/jocondelab/>
prefix ml: <http://ns.inria.fr/monalia/>

select     ?searchCategory_label  ?searchScore_value ?ref ?title ?repr
?score ?imagePath
where {
   ?searchCategory a jcl:Term;
                              skos:prefLabel "chien“@frr.

   ?classifier rdfs:subClassOf/ml:vocabID "REPR“;
                    rdfs:label ”40_classes".

   ?notice jcl:noticeRef  ?ref;
             jcl:noticeTitr  ?title;
               jcl:noticeRepr  ?repr;
               jcl:noticeImage [ jcl:noticeImageIsMain true ;
jcl:noticeImagePath ?imagePath];
               ml:imageClassifier [a ?classifier ; ml:detected [a
?searchCategory; ml:score ?score]] .

   filter ( ?score >= 0.90 )
}
order by desc(?score)
```

Detecting “noise”

- By querying the extended metadata for the objects with low scores we can detect the “noise” in the represented subject annotation

Image	Metadata	Score
	figure (saint Eloi de Noyon, évêque, en pied, bénédiction, vêtement liturgique, mitre, attribut, **cheval**, marteau, outil : ferronnerie) 000SC022652 C:/Joconde/joconde\0355/m079806_bsa0030101_p.jpg	cheval: 0.006
	figures bibliques (Vierge à l'Enfant, à mi-corps, assis, Enfant Jésus : nu, livre);fond de paysage (colline, cours d'eau, barque, **cavalier**) 000PE027041 C:/Joconde/joconde\0001/m503604_90ee1719_p.jpg	cheval: 0.009
	scène (satirique : Bismarck Otto von : Gargantua, repas, **cheval**, boisson : vin) 5002E006121 C:/Joconde/joconde\0074/m500202_atpico-g70128_p.jpg	cheval: 0.011

Detecting “silence”

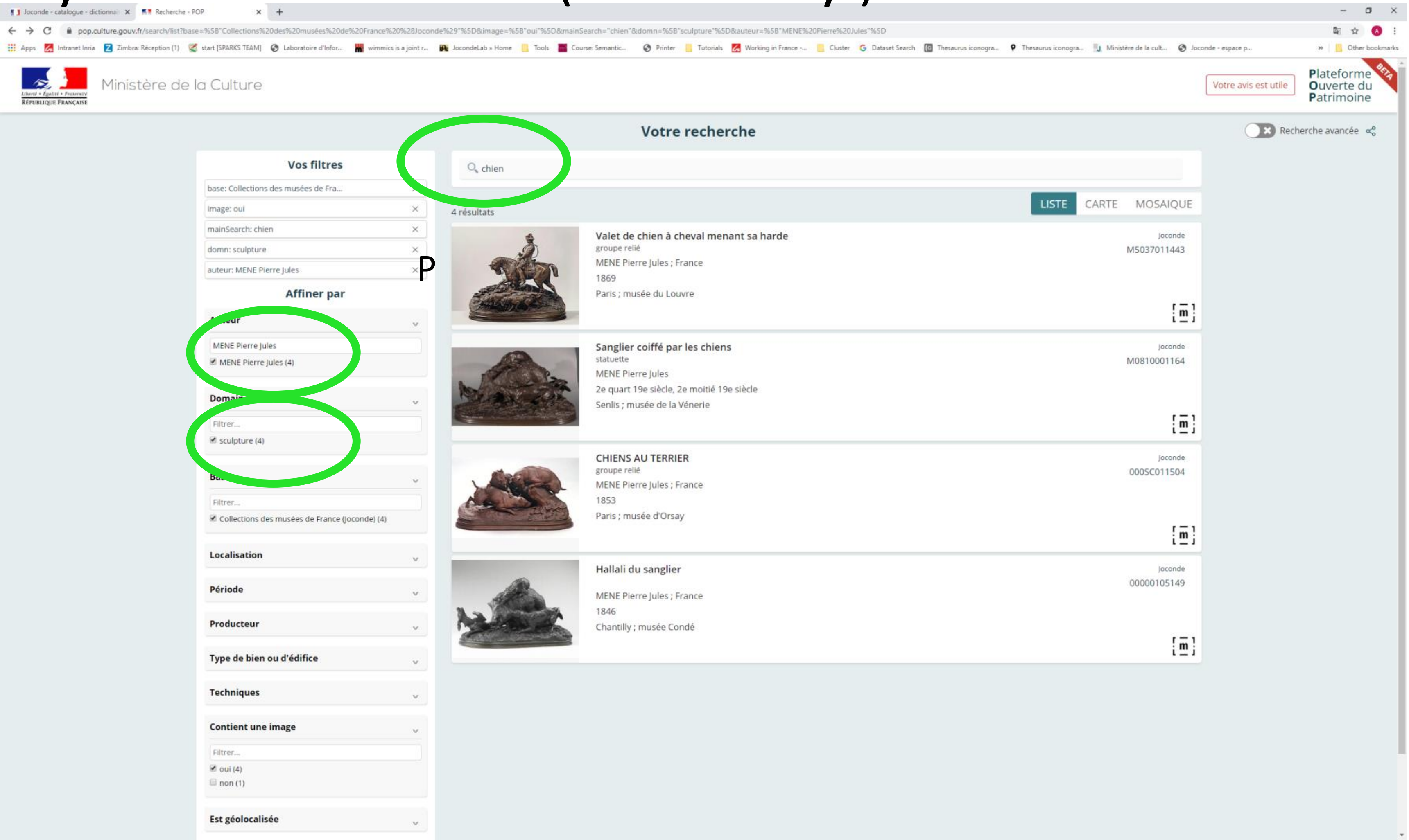
- By querying the extended metadata for the object with high scores and without object mentioned in annotation we can detect the “silence” in the annotation

Image	Metadata	Score
	portrait 50350012455 C:\Joconde\joconde\0138\m503501_d0012455-000_p.jpg	cheval: 0.999
	scène historique (guerre de siège : Lawfeld, Louis XV, Saxe maréchal de, bataille rangée) 000PE004371 C:\Joconde\joconde\0634\m507704_79ee519_p.jpg	cheval: 0.999
	figure (sainte Jeanne d'Arc, jeune fille, équestre passant, armure, asque, épée) M0301000355 C:\Joconde\joconde\0617\m030106_007305_p.jpg	cheval: 0.997

Improving search results

- This sculpture is not found as a result of the current search on the MiC portal for “chien” by Pierre-Jules Mêne (for brevity)

Lévrier tenant un lièvre dans sa gueule by Pierre Jules Mene
© musée de la Vénerie ; Senlis, © Direction des Musées de France, 2009



P

Sujet représenté	scène (chasse : lévrier, lièvre)

Ranking of the image content

- The thesaurus based labeling mechanism + classifier identified a dog on this image with highest degree of certainty

Image	**Labels & Metadata**	**Prediction**
	[0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0] chien scène (chasse : lévrier, lièvre) M0810001165 C:\Joconde\joconde\0466\m081003_028491_p.jpg	[0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0] chien : 0.993 cheval : 0.765 lion : 0.609 oiseau : 0.473 mouton : 0.422 de profil : 0.235 arbre : 0.124 cavalier : 0.112 fleur : 0.079 feuille : 0.036

Ranking of search results

- Running the same query on the Extended Joconde database and sorting by score gives a better result putting the image in the second place

Image	Metadata	Score
	représentation animalière (épagneul, debout) M0341003743 C:\Joconde\joconde\0534\m034186_006932_p.jpg	chien: 0.994
	scène (chasse : lévrier, lièvre) M0810001165 C:\Joconde\joconde\0466\m081003_028491_p.jpg	**chien: 0.993**
	représentation animalière (mise à mort, gros gibier : sanglier, chasse à courre, chien) 00000105149 C:\Joconde\joconde\0107\m505206_oa817_p.jpg	chien: 0.990

Conclusions

- **Organizing represented object categories only by Garnier Thesaurus is not sufficient for labeling images**
  - Creating new links between the categories that are semantically and/or visually related but do not exist in Grainier Thesaurus benefits the model training datasets thus creating more accurate classifiers.
- **Multi-label classifiers that recognize more than one object represented on the image can benefit the artwork annotation process**
  - Ranking the detected objects by prediction scores allows detection of “silence” and “noise” in image annotation.
- **Extending the Joconde metadata with DL model generated prediction scores can benefit information retrieval**
  - Using the scores to filter and order search query allows the retrieval of more relevant search results.

Next Steps: MonaLIA 3.0

Château de Bouillon : tête d'escalier by JEAN-HAFFEN Yvonne

- Develop semantic reasoning based on thesaurus and prediction scores
- Augment the training data by performing artistic style transformations of photos for underrepresented categories
- Scale up the multi-label classification to 300 classes depending on image availability
- Explore other state-of the-art object detection networks to improve on detection performance
- Explore the multi-task learning using the semantic hierarchy
- Explore semantic segmentation with state-of the-art networks

Thank you

- Acknowledgements:

Dr. Fabien Gandon

Dr. Frédéric Precioso

Francois Reygagne

Liberté • Égalité • Fraternité
RÉPUBLIQUE FRANÇAISE

Ministère de la Culture