



**HAL**  
open science

# Towards A Hybrid Approach for Medieval Illuminations Analysis

Cheikh Brahim El Vaigh, Martine Clouzot, Christophe Nicolle

► **To cite this version:**

Cheikh Brahim El Vaigh, Martine Clouzot, Christophe Nicolle. Towards A Hybrid Approach for Medieval Illuminations Analysis. 17th International Conference on SIGNAL IMAGE TECHNOLOGY & INTERNET BASED SYSTEMS, Nov 2023, BANGKOK, France. hal-04455810

**HAL Id: hal-04455810**

**<https://inria.hal.science/hal-04455810>**

Submitted on 13 Feb 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

# Towards A Hybrid Approach for Medieval Illuminations Analysis

Cheikh Brahim EL VAIGH  
CIAD UR 7533  
Université de Bourgogne, UB  
Dijon, France

Martine Clouzot  
ArTeHis UMR 6298  
Université de Bourgogne, UB  
Dijon, France

Christophe Nicolle  
CIAD UR 7533  
Université de Bourgogne, UB  
Dijon, France

**Abstract**—Medieval illuminations represent a system of graphic and visual information, based on metaphorical relationships whose meanings and messages are determined in a specific and evolving context. We consider an illumination as a knowledge graph used by certain elites during the Middle Ages to symbolize their social identity and convey events from their lives, along with their cultural values. This knowledge graph contains a set of entities connected through semantic relationships and was used to encode metaphors and impactful messages that can sometimes be challenging for non-experts to decipher. Our aim is to provide a straightforward approach to decoding those messages. Therefore, we first consider medieval illuminations as images and use YOLOv3 to identify the objects inside a given image using bounding boxes. Moreover, we retrieve the entities from the image, their relationships, and the implicit messages using logical inference. The ultimate aim is to simplify the interpretation of illuminations and provide in the future privacy tools for contemporary social media, such as a logical framework for encoding and transmitting information.

**Index Terms**—Hybrid AI, Object Detection, Medieval illumination, Semantic Web, Ontology

## I. INTRODUCTION

Social networks encompass online platforms and applications that enable individuals and entities to establish connections, engage in communication, exchange information, and cultivate relationships. In the Middle Ages, social networks existed also but in a different manner, using different tools and representations. In fact, Medieval illuminations are systems of graphic and visual information—images painted on parchment—based on metaphorical relationships, whose meanings and messages are determined within a particular context [1], [2]. They were used by certain elites to share encrypted information. In Figure 1, we can see a medieval illumination that encodes very a particular message that we will elucidate hereunder.

The Illumination Project <sup>1</sup> is a knowledge engineering tool that aims to logically formalize human knowledge so that it can be manipulated by a computerized system. Within the project, we specifically focus on characterizing and formalizing symbolic relationships between concepts inside medieval illuminations. The project is driven by the appreciation of the cultural heritage of libraries and museums in Burgundy, by the

semantic web as a tool for knowledge dissemination, and by the omnipresence of social networks in all areas of our lives.

In previous works [3]–[5] authors studied medieval illuminations and created the ontology **Medievelen** [5] which allows to encode experts knowledge. Moreover, they highlighted the strong structural correlations between medieval illuminations and social networks which may help elucidate the processes through which illuminations served as visual means of communication in medieval times. In this work, we study how we can annotate the images that represent medieval illuminations and then use these annotations along with the experts’ rules encoded with the ontology **Medievelen** to decode a given medieval illumination’s message.

In this paper, we build a pipeline to automatically decode medieval illuminations. Our goal is to combine machine learning techniques and semantics ontology to retrieve the message of medieval illumination. Specifically, we combine computer vision models to extract objects from a given image using object detection algorithms and the **Medievelen** [5] ontology defined in semantic web standard OWL. Thus, we first use version 3 of the real-time object detection model YOLO [6] to retrieve objects inside a given image. Then, we leverage SWLR rules [7] to define the entities inside the image from the association of different objects. Finally, we make use of experts’ SWLR rules to define the implicit message of medieval illuminations by looking at the set of entities found inside the images and their semantic relationships. These semantics rules are defined with the help of medievalist experts. The contribution of this paper is twofold. We built a pipeline combining machine learning for automatic image annotation and causal reasoning mechanisms to encode expert knowledge. Moreover, we conducted an experimental evaluation to show the effectiveness of our approach.

The rest of this paper presents our methodology for creating a pipeline to elucidate medieval illumination’s content using both machine learning and semantics ontology. We first introduce the basic concepts relative to first-order logic ontologies and the used object detection model in the preliminaries (Section II). Then we review similar work in the domain of medieval illumination analysis in Section III. Our proposed pipeline is shown in Section IV. Finally, we conduct experiments (Section V) to show the effectiveness of our pipeline in extracting implicit messages from medieval illumination and

Identify applicable funding agency here. If none, delete this.

<sup>1</sup><https://www.ciad-lab.fr/project/illumination-3-0/>

conclude in (Section VI).

## II. PRELIMINARIES

### A. Ontology and Inference

We define an ontology (a.k.a. semantic knowledge graph or KG) as a deductive incomplete database that contains two elements: an assertion box or ABox which contains the facts similarly to any classical database and a terminological box or TBox that represents the background knowledge of the application domain and provides a way to reason about the data beyond what is explicitly stored in the ABox.

Ontologies are useful in designing semantics knowledge graphs that have a data component and a reasoning component which allow inference mechanism based on first order logic.

These ontologies can be expressed using description logic, which serves as the basis for the Web Ontology Language (OWL2) [8] developed by the World Wide Web Consortium (W3C)<sup>2</sup>.

*Example 1 (Running example):* Let us consider the simple KG  $\mathcal{K} = (\mathcal{O}, \mathcal{D})$  where

$$\begin{aligned} \mathcal{O} = \{ & r_0 = \text{hasRightToKill}(X) \rightarrow \text{Knight}(X), \\ & r_1 = \text{hasRightToKill}(X) \rightarrow \text{Person}(X), \\ & \text{Weapon}(Y), \text{carry}(X, Y), \\ & r_2 = \text{Duke}(X) \rightarrow \text{Lord}(X), \\ & r_3 = \text{Duke}(X) \wedge \text{King}(X) \rightarrow \perp \}, \\ \mathcal{D} = \{ & \text{Duke}(\mathbf{a}), \text{Duke}(\mathbf{b}), \text{Duke}(\mathbf{g}), \text{Knight}(\mathbf{c}), \\ & \text{hasRightToKill}(\mathbf{g}) \}. \end{aligned}$$

$\mathcal{K}$  is expressed using the DL-lite $\mathcal{R}$  fragment of datalog $\pm$  and existential rules.  $\mathcal{O}$  states that a knight has the right to kill with ( $r_0$ ) someone carrying a weapon has the right to kill ( $r_1$ ), A Duke is also a Lord ( $r_2$ ), and a Duke can not be a King at the same time ( $r_3$ ).  $\mathcal{D}$  states that  $\mathbf{a}$ ,  $\mathbf{b}$ ,  $\mathbf{g}$ , are dukes (*Duke*),  $\mathbf{c}$  is a knight (*Knight*) and  $\mathbf{g}$  has the right to kill.  $\diamond$

**Consistency.** KB  $\mathcal{K}$  is qualified as *consistent* if and only if it does not entail  $\perp$ , i.e.,  $\mathcal{K} \not\models \perp$  where  $\models$  is the first order (FO) consequence relation. Inconsistency arises from the interaction between the negative rules in the ontology and the facts in the associated database to derive  $\perp$ .

*Example 2 (Cont.):*  $\mathcal{K}$  is consistent because  $\perp$  cannot be deduced from the unique negative rule  $r_3$  and the (entailed) facts for the *subClassOf* relation:  $\text{Duke} \subset \text{Lord}$  and  $\text{Duke}(\mathbf{a}), \text{Duke}(\mathbf{b}), \text{Duke}(\mathbf{g})$  in  $\mathcal{D}$  i.e., there is no Duke that is also a King.  $\diamond$

**Queries and answers.** For query answering, *FO queries* are defined as  $q(\bar{x}) = \phi$ , where  $\phi$  is an FO formula, and the set of free (non-quantified) variables in  $\phi$  matches exactly the tuple  $\bar{x}$  of answer variables. The arity of a query  $q(\bar{x})$  is the number of variables in  $\bar{x}$ . If  $\bar{x}$  is empty, the query  $q(\bar{x})$  is called Boolean. A *certain answer to a query*  $q(\bar{x})$  of arity  $n$  on a consistent KB  $\mathcal{K}$  is a tuple of  $n$  constants  $\bar{\mathbf{t}}$  such that  $\mathcal{K}$  satisfies the formula  $q(\bar{\mathbf{t}})$ , where  $q(\bar{\mathbf{t}})$  is the Boolean query obtained by

substituting  $\bar{x}$  with  $\bar{\mathbf{t}}$  in  $q$ . In the rest of this article,  $\text{ans}(q, \mathcal{K})$  denotes the set of answers of  $q$  in  $\mathcal{K}$ .

*Example 3 (Cont.):* Let us consider the query asking for the dukes that are also knights:  $q(x) = \text{Duke}(?x) \wedge \text{Knight}(?x)$ . Its answer set in  $\mathcal{K}$  is  $\text{ans}(q, \mathcal{K}) = \{\mathbf{g}\}$ :  $\mathbf{g}$  is obtained from  $\text{Duke}(\mathbf{g}) \in \mathcal{D}$ ,  $\text{hasRightToKill}(\mathbf{g}) \in \mathcal{D}$  and the fact  $\text{Knight}(\mathbf{g}) \in \mathcal{D}$  entailed from  $\text{hasRightToKill}(\mathbf{g}) \in \mathcal{D}$  and  $r_0$ .  $\diamond$

From Example 3, even though there is nothing in KB  $\mathcal{D}$  explicitly stating  $\text{Knight}(\mathbf{g})$  ( $\mathbf{g}$  is a knight), this information can be inferred based on rule  $r_0$ . In this case, the ontology is used to infer implicit facts using reasoning based on FO ontologies. In this work we use SWRL [7] to write logical rules, and as stated in [7] are equivalent to OWL assertions. The rule  $r_2$  is equivalent to (*Duke subClassOf Lord*) in OWL language [8]. The goal of this work is to bring high semantic reasoning capability to the field of medieval illumination analysis.

### B. Object Detection

Object detection methods are a crucial component of computer vision and artificial intelligence, enabling machines to identify and locate objects within images or videos. These techniques have wide-ranging applications, from autonomous vehicles to surveillance systems, robotics, and image analysis. Several object detection methods have emerged over the years, each with its strengths and weaknesses. Besides the traditional approaches that relied on handcrafted features [9]–[11], deep learning models such as R-CNN based models [12]–[14] first propose regions of interest and then classify and refine them. One-stage detectors perform detection in a single pass they are called two-stage detectors. Moreover, YOLO [6] and SDD [15] are one-stage detectors. SSD is known for its real-time object detection capabilities by combining multiple bounding box predictions at different scales. Similarly, YOLO is also a real-time object detection method that divides an image into a grid and predicts bounding boxes and class probabilities simultaneously. In this work, we use YOLO’s ability to detect objects in different regions of an image in order to retrieve entities from medieval illumination. As part of our pipeline, YOLO is used with the ontology for medieval illumination decoding.

## III. RELATED WORKS

Medieval Illumination Analysis is a multidisciplinary field that combines art history, computer science, and cultural studies to unlock the rich symbolism and cultural narratives embedded in illuminated manuscripts. The study of medieval illumination has a long history, but it has become increasingly sophisticated in recent years. Researchers use a variety of methods to analyze illuminated manuscripts, including:

- **Art historical analysis:** They consist in focusing on the formal elements of the illuminations, such as composition, style, and iconography [16].

<sup>2</sup><https://www.w3.org/>



Fig. 1: The Lady with the Unicorn at the Cluny Museum © Cluny Museum - National Museum of the Middle Ages.

- **Codicological analysis:** They are based on the physical characteristics of the manuscripts, such as the materials used, the binding, and the scribal script [17].
- **Paleographical analysis:** They devise analysis based on the handwriting of the manuscripts to identify different scribes and to date the manuscripts [18].
- **Machine learning based analysis:** This type of analysis examines the visual aspects of the digital images associated with the medieval illuminations in order to facilitate their classification [19], [20]. In [19], [20] authors devised a framework that facilitates the creation of a hierarchical classification system for medieval illuminations through visual analytics, enabling researchers to organize and categorize these artworks effectively.
- **Ontology-based analysis:** ontology-based approaches for medieval illumination analysis have gained prominence for their ability to provide structured and semantically rich insights into the symbolism, iconography, and cultural context of illuminated manuscripts and particularly medieval illumination [3]–[5], [21]. A medieval illumination can be considered as a knowledge graph which was used by some elites in the Middle Ages to represent themselves as a social group and exhibit the events in their lives and their cultural values. Authors in [3] used this idea to describe the meaning of the elements of

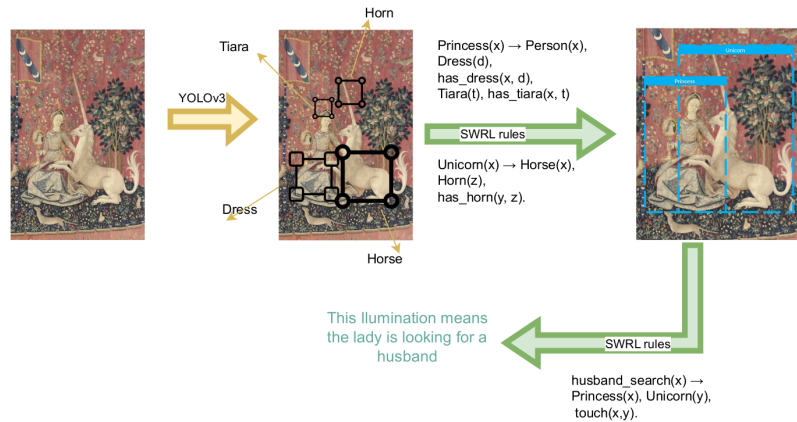


Fig. 2: Decoding the medieval illumination in Image 1 using our pipeline.

this knowledge graph through logical modeling using ontologies. Therefore, the authors built logical reasoning rules and used the inference mechanism to facilitate the interpretation of illuminations. Moreover, the **Medievelen** ontology was proposed in [5]. The goal of this work is to provide a formal ontology encoding the knowledge about medieval illuminations. An online annotation platform was proposed in [4] to allow illumination annotation by users which creates a collaborative tool for knowledge sharing. These annotations are very basic they can be the type of the objects inside an image e.g., Person, Weapon, Bell... ) or a semantic relation from the ontology **Medievelen** between two entities e.g., holding (between a Person and a Weapon). Our pipeline consists of using both causal reasoning (with an ontology such as **Medievelen**) and a computer vision model for object detection inside an image. An ontology makes it possible to define an entity from a set of different objects (see Example 4 for the definition of a Princess) using SWLR [7] rules. In addition with the help of medieval experts, it allows us to represent and store the rules to decode illuminations. Our work includes an automatic annotation tool, it is similar to [4], [19], [20], but we do not need human annotation. Finally, the main contribution of our work is its ability to automatically decode the message of an illumination.

*Example 4:* Let us consider the illumination in Figure 1, with Yolo [6], we can detect the objects Person, Dress and from these objects we can define a Princess as follow:  $Princess(x) \rightarrow Person(x), Dress(d), has\_dress(x, d), Tiara(t), has\_tiara(x, t)$

This SWLR rule means that a princess is a person who has a dress and a tiara.  $\diamond$

#### IV. MEDIEVAL ILLUMINATIONS DECODING PIPELINE

The objectives of the Illumination project are the characterization and formalization of symbolic relationships between concepts in a constantly evolving context for the case of

medieval illuminations. In this regard, research has been conducted to identify the knowledge inside an image e.g., object inside the image, concepts that can be built by combining these objects, and finally the semantics relationships between these concepts.

The proposed pipeline contains a first machine learning model to identify objects inside an image IV-A, then an annotation step to identify concepts associated with a set of objects and their relationships IV-B and the concepts and their relationships are used to elucidate the message of the illuminations.

#### A. Object Detection in the Illuminations

The advent of machine learning techniques, particularly those employing pre-trained convolutional neural networks (CNNs) like ResNet [22] or VGG [23], has rendered the need for manually crafted features obsolete. These pre-trained CNNs have demonstrated their ability to effectively capture meaningful representations of various entities, including text, natural images, and art pieces, making them a widely adopted, ready-to-use approach for classifying artworks [24], [25]. CNNs are also used within deep learning models such as R-CNN, Fast R-CNN, and Faster R-CNN, to identify objects inside an image or object detection. In our study, we relied on the model YOLO [6] (You Only Look Once). YOLO is an open-source object detection system that revolutionized the field of computer vision. It offers impressive speed and accuracy in identifying and locating objects in images or video frames. YOLO predicts bounding boxes around detected objects, providing precise location information for each object. Each bounding box is associated with a class label and a confidence score. The system has evolved over the years, introducing improvements in terms of accuracy, speed, and architecture. In our pipeline, we used YOLOv3 [26] which implemented several minor design modifications to enhance YOLO's performance.

The YOLO [6] detection network comprises 24 convolutional layers, followed by 2 fully connected layers. To reduce feature space, authors utilize alternating  $1 \times 1$  convolutional layers. Initially, they pre-train the convolutional layers on the ImageNet classification task with a lower resolution ( $224 \times 224$  input image) and subsequently double the resolution for the detection task. Both YOLO and YOLOv3 are trained to predict bounding boxes that are responsible for each object inside the image. The model is evaluated in terms of IoU (Intersection over Union) which is defined as the ratio between the overlap of the predicted instance with the real one. It is highly valued for image segmentation problems.

Models like YOLO were trained on datasets with modern images such as the COCO dataset<sup>3</sup>. Meanwhile, we have old images obtained from scanned illumination manuscripts. In fact, we used Mandragore an API provided by the BnF (the French National Library) which serves as the iconographic indexing database for illuminations at the BnF. Continuously

enriched since 1989, it currently describes 205,000 illuminations from nearly 7,000 manuscripts held in the Manuscripts Department and the Arsenal Library. Each identifiable element is indexed using a multilingual vocabulary of over 20,000 terms. The web application providing access to Mandragore since 2003 is now somewhat outdated and is the subject of a redesign project aimed at better showcasing the rich database content. As part of the Illumination project, more than 5,000 illumination images have been extracted from Mandragore, originating from various countries but primarily from France.

To conduct initial tests without annotating the entire dataset (which contains over 5000 complete illuminations and over 5000 illuminations manuscripts), we are limiting ourselves to detecting a minimum number of classes: Person, Animal, and Weapon. We have thus annotated approximately 300 images that may or may not contain these three classes. It is important to note that images without our classes are just as important as images containing our classes; this allows the neural network to identify what is not something. In addition, we used the Platform developed in [4] which allows any user to annotate the objects inside the images by surrounding the objects with bounding boxes which can speed up the annotation processing and improve the quality of the detection model.

We take several precautions to ensure that the model is satisfactory:

- Images are resized before annotation, which allows the network to maintain the overall proportions of the objects to be annotated.
- Images and annotations are placed in the same folder and saved in Yolo annotation format.
- The configuration files provided for training the model are modified to match our dataset, GPU configuration, and the number of classes. This number of classes is updated after each modification of the dataset or the addition of a class to be detected.

The off-the-shelf YOLOv3 model has poor accuracy on our images. In order to improve its accuracy, we fine-tuned YOLOv3 on our annotated images and tried two cases namely predicting only one class (Person) and predicting three classes (Person, Weapon, and Animals). The results for these experiments are shown in Section V.

#### B. Elucidating Medieval Illuminations

The automatic annotation of illumination has two parts: the object detection algorithm that yields the bounding boxes using YOLOv3 described in the previous section, and the reasoning mechanism that constructs a concept from a set of objects defined by their bounding boxes within the same image. In Example 4 the concept *Princess* is defined as a *Person*, with a *Dress* and a *Tiara*. We can define a set of SWRL rules defining all the concepts that are inside the ontology **Medievelen** [5]. We also make use of the inference mechanism to check the appropriate type of an object so the SWRL rules remain true. One can define the concept *hasRightToKill* as a *Person* carrying a *Weapon*:

$hasRightToKill(X) \rightarrow Person(X), Weapon(Y), carry(X, Y)$

<sup>3</sup><https://cocodataset.org/>



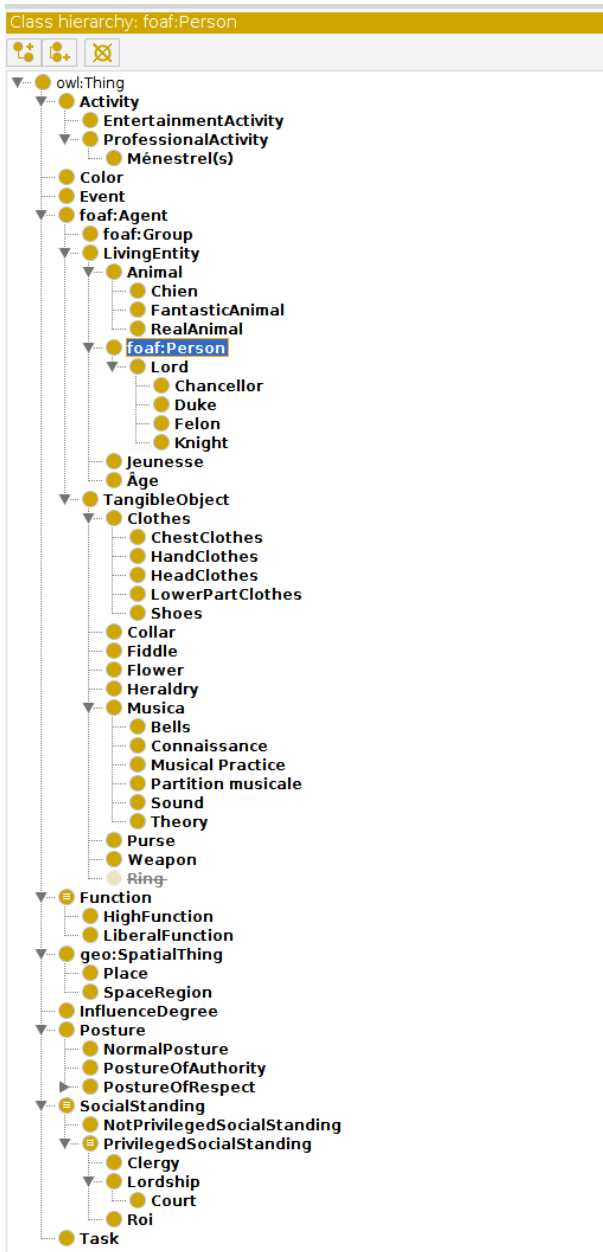


Fig. 3: The hierarchy of the classes inside the ontology **Medievalen** [5]

If we have in an image a Duke carrying a Sword ( $Duke(X), carry(X, Y), Sword(Y)$ ), the previous rule does not apply directly. But since in the **Medievalen** [5] and as shown in Image 3, we have  $Duke \subset Lord \subset Person$  meaning that  $\forall X Duke(X) \rightarrow Person(X)$  and similarly if we have  $Sword \subset Weapon$  we can write it as  $\forall X Sword(X) \rightarrow Weapon(X)$ . We can then use the inference mechanism to rewrite the previous facts ( $Duke(X), carry(X, Y), Sword(Y)$ ) as  $Person(X), carry(X, Y), Weapon(Y)$  which means a Duke carrying a Sword has the right to kill. We define the rules for the most general concept and using the inference mechanism

we can deduce implicit facts.

For example, there is no need to define the rule

$$hasRightToKill(X) \rightarrow Duke(X), Sword(Y), carry(X, Y)$$

In fact, as we show in the previous example, this rule can be deduced from the more general one:

$$hasRightToKill(X) \rightarrow Person(X), Weapon(Y), carry(X, Y)$$

For the relation between concepts we handle them partially by looking at a couple of concepts inside an image and we try to predict the most general relation from the ontology. Therefore relation detection inside the image is bound to the relationships defined inside the ontology **Medievalen** [5]

The inference mechanism is performed based on the reasoning engine Pellet [27]. Pellet is a popular reasoning engine used in the field of Semantic Web and ontology-based knowledge systems. It is known for its efficiency in performing Description Logic (DL) reasoning, which includes tasks such as ontology consistency checking, instance classification, and inferencing based on ontological knowledge. Pellet is often employed in applications that require automated reasoning to make logical inferences and deductions from complex knowledge structures, facilitating more intelligent and context-aware decision-making in various domains.



Fig. 4: A woman touching a unicorn, but she is not looking for a husband obtained from this online catalogue <https://www.bl.uk/catalogues/illuminatedmanuscripts/ILLUMIN.ASP?Size=mid&IllID=47419>

The process of elucidating medieval illuminations relies mainly on our annotation process, which enables concepts to be defined from the object detection stage. Once we define the concepts inside an image and their relationships, we define SWLR rules to elucidate the message of an illumination with the help of a medievalist expert.



(a) Examples of objects detection with one person



(b) Examples of objects detection with different faces positions

Fig. 5: Examples of objects detection using our finetuned YOLOv3 on the class Person.

Inside the Image 2 after identifying all the objects, we will be able to define the concepts *Princess* and *Unicorn* but we also have a special expert rule stating that if a princess is shown inside an image touching a unicorn it means she is looking for a husband. We then devised the following rule:

$$\text{husband\_search}(X) \rightarrow \text{Princess}(X), \\ \text{Unicorn}(Y), \text{touch}(X, Y)$$

These rules for elucidating medieval illumination are fixed by the expert and our contribution is to analyze the medieval images in order to find the best rule to use. We also make use of the inference mechanism with Pellet to reduce the quality of rules needed for illumination analysis.

For example, if we use our pipeline to understand the illumination in Image 4, we can see that we have a *Woman* touching the Unicorn. Can we deduce that the woman is looking for a husband? the answer is no. Actually, in Image 4 we have a *Person* who has the right to kill, and he is attacking the Unicorn a *Kinght* then.

We can write the facts inside the Image 4 as follow:

$$\text{hasRightToKill}(X), \text{Unicorn}(Y), \text{Attacking}(Y, X), \\ \text{Woman}(Z), \text{touch}(Z, Y)$$

This means the woman already found a husband and it is dangerous to come to make a proposal!



Fig. 6: Examples of objects detection using our finetuned YOLOv3 on the class Person with white skin and dark skin.

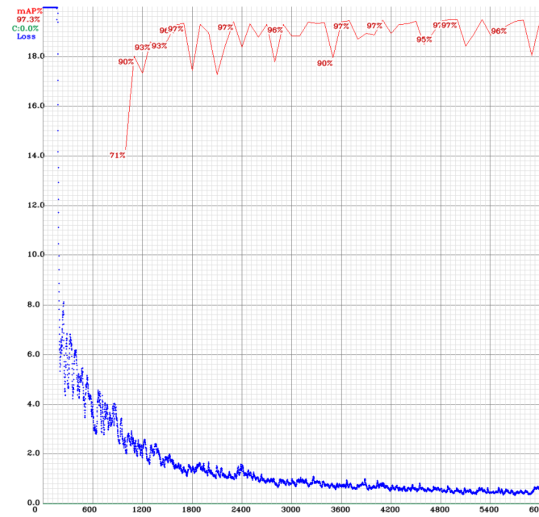


Fig. 7: Accuracy and loss function associated to the detection and bounding box for the class *Person*.

## V. EXPERIMENTAL EVALUATION

We conducted a comprehensive experimental evaluation of our pipeline using the benchmarks that we created from BnF illuminations. To demonstrate the effectiveness of our proposed approach we finetuned the object detection model YOLOv3 [26] to find objects in the illumination using bounding boxes. To evaluate our model we used the metric mAP, or mean Average Precision, which is a commonly used performance measure in object detection and information retrieval tasks. It evaluates the precision and recall of a model by calculating the average precision across various classes or queries. It provides a single number that quantifies the overall quality of the model's predictions, with higher mAP values indicating better performance in localizing and recognizing objects.

We first train the model to detect only the class *Person*. In any image, we will be looking for all the possible bounding boxes that contains a person. We observe a significant decrease in the loss function and a rapid increase in the mAP metric to

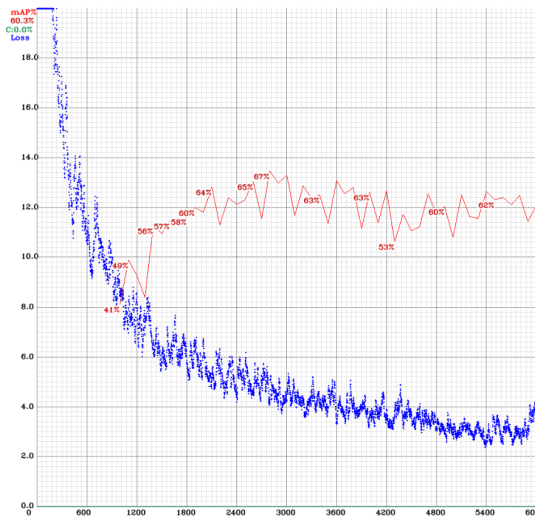


Fig. 8: Accuracy and loss function associated to the detection and bounding box for the the classes *Person*, *Weapon* and *Animal*.

around 97% (see Figure 7). It appears that object detection for the "Person" class is quite accurate. We can see that frontal faces are correctly detected, while profiles are detected less accurately as in Images 5b the model was not able to predict the last person in the image. Face detection via bounding boxes seems to work correctly with no false positives. Face detection appears to struggle or forget to detect individuals with darker skin tones. For example in Image 6, the model was not able to predict people with dark skin. We argue that the selection of illuminations plays a role; the choice of illuminations can influence the representation of people with different skin tones, and the dataset mainly consists of French illuminations where the majority of individuals are depicted as white. The number of annotations is also a major factor; it's easier for the model to develop biases when it has limited training data. It's important to note that only 300 images containing zero to multiple faces were annotated. The quality of illuminations varies; some illuminations are damaged, which may explain the model's difficulty in detecting certain faces affected by the aging of the document.

Impressed by the high accuracy of the model (97% in term of mAP), we increased the number of classes to 3 namely *Person*, *Weapon* and *Animal*.

In the same way as for the learning with one class, we observe a noticeable but slower decrease in the loss function (see Figure 8). However, it appears that the bounding box detection and the mAP calculation result in a lower percentage (67% in the best case). Nevertheless, certain results can still be discussed. We can see in Image 9 that faces still appear to be detected equally well. The detection of the *Animal* and *Weapon* classes is much less accurate. We can explain the decrease in accuracy with the following reasons:

- The choice of our other two classes, *Animal* and *Weapon*, encompasses a wide range of vastly different shapes,

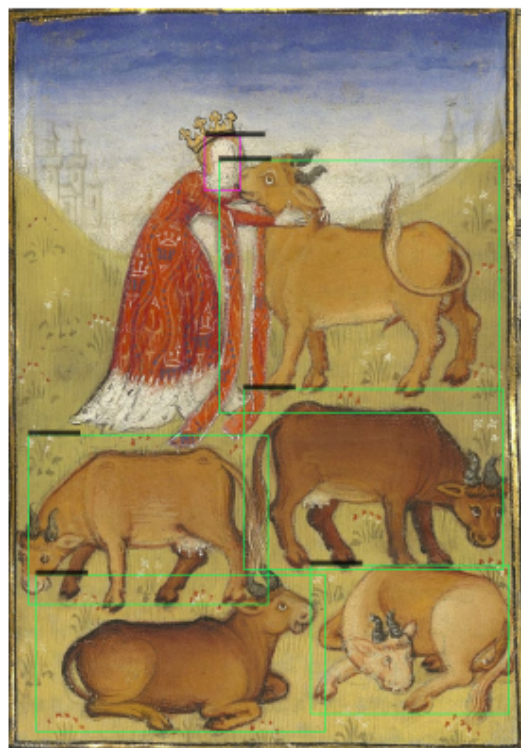


Fig. 9: Examples of objects detection using our finetuned YOLOv3 on the classes *Person*, *Animal*, *Weapon*.

reducing the model's precision.

- Objects of type *Animal* and *Weapon* often appear alongside humans and are frequently overshadowed by numerous other objects, diminishing the quality of our model.
- Annotations for the *Animal* and *Weapon* classes in a sample of 300 images are about 10 times less frequent than *Person*, significantly reducing the detection quality.

## VI. CONCLUSION

Medieval illuminations serve as intricate systems of graphic and visual information, characterized by metaphorical relationships whose interpretations evolve within a specific context. These illuminations functioned as knowledge graphs employed by medieval elites to symbolize their social identity, depict life events, and convey cultural values. However, decoding these metaphors and messages can prove challenging, particularly for non-experts.

To address this challenge, we propose a systematic approach that treats medieval illuminations as images and employs YOLOv3 for object identification through bounding boxes. Additionally, we utilize ontology to extract entities, relationships, and implicit messages from these images, facilitated by logical inference. Our ultimate goal is to simplify the interpretation of illuminations and potentially provide privacy tools for contemporary social media platforms by establishing a logical framework for encoding and transmitting information. This approach offers a bridge between historical and modern



communication methods, enabling a deeper understanding of the rich visual language of medieval illuminations.

In future work, we plan to improve the accuracy of our model by increasing the annotated images and the quality of the detection model. We also plan to devise logical communication tools based on the illumination idea where users can define a set of objects to encode a message. Finally, we plan to study how generative models such as Generative Adversarial Networks can be used to generate illumination from a plain text.

## VII. ACKNOWLEDGMENT

The authors would like to express their deep gratitude to Nicolas Gros and Jonathan Martin-Maestre, engineers at the CIAD Laboratory, for their contributions to the development of our final tool. Their expertise and dedication were instrumental in bringing this project to fruition.

## REFERENCES

- [1] R. G. Calkins, "Illuminated books of the middle ages," 1983.
- [2] L. F. Sandler, "A history of illuminated manuscripts," *The Art Bulletin*, vol. 70, no. 3, pp. 521–523, 1988.
- [3] D. Diarra, M. Clouzot, and C. Nicolle, "Causal reasoning and symbolic relationships in medieval illuminations," *Journal of Data Mining and Digital Humanities*, 2019.
- [4] D. Diarra, R. Belkaroui, M. Clouzot, and C. Nicolle, "Illumination3. 0: A semantic annotation platform based on ontology for medieval illuminations," in *30es Journées Francophones d'Ingénierie des Connaissances, IC 2019*, 2019, pp. 231–234.
- [5] D. Diarra, M. Clouzot, and C. Nicolle, "Mediev'en-la domain ontology for cultural heritage supports: Case of medieval illuminations of burgundy duke," *Available at SSRN 4386897*.
- [6] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 779–788.
- [7] I. Horrocks, P. F. Patel-Schneider, H. Boley, S. Tabet, B. Grosz, M. Dean *et al.*, "Swrl: A semantic web rule language combining owl and ruleml," *W3C Member submission*, vol. 21, no. 79, pp. 1–31, 2004.
- [8] "OWL 2 Web Ontology Language Profiles," <https://www.w3.org/TR/owl2-profiles/>, 2012.
- [9] S. Soo, "Object detection using haar-cascade classifier," *Institute of Computer Science, University of Tartu*, vol. 2, no. 3, pp. 1–12, 2014.
- [10] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, vol. 1. Ieee, 2005, pp. 886–893.
- [11] L. Cole, D. Austin, L. Cole *et al.*, "Visual object recognition using template matching," in *Australian conference on robotics and automation*, 2004.
- [12] X. Xie, G. Cheng, J. Wang, X. Yao, and J. Han, "Oriented r-cnn for object detection," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 3520–3529.
- [13] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," *Advances in neural information processing systems*, vol. 28, 2015.
- [14] R. Girshick, "Fast r-cnn," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1440–1448.
- [15] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "Ssd: Single shot multibox detector," in *Computer Vision—ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14*. Springer, 2016, pp. 21–37.
- [16] J. J. Alexander, "Art history, literary history, and the study of medieval illuminated manuscripts," *Studies in Iconography*, vol. 18, pp. 51–66, 1997.
- [17] Z. Gulácsi, "Mediaeval manichaean book art: A codicological study of iranian and turkic illuminated book fragments from 8th-11th century east central asia," in *Mediaeval Manichaean Book Art*. Brill, 2005.
- [18] R. Córdoba de la Llave, "Interdisciplinary exploration of medieval technical manuscripts from the iberian peninsula," *Journal of Medieval Iberian Studies*, vol. 14, no. 1, pp. 96–108, 2022.
- [19] C. Meinecke, "Labeling of cultural heritage collections on the intersection of visual analytics and digital humanities," in *2022 IEEE 7th Workshop on Visualization for the Digital Humanities (VIS4DH)*. IEEE, 2022, pp. 19–24.
- [20] C. Meinecke, E. Guéville, D. J. Wrisley, and S. Jänicke, "A visual analytics framework for composing a hierarchical classification for medieval illuminations," *arXiv preprint arXiv:2208.09657*, 2022.
- [21] D. Diarra, M. Clouzot, and C. Nicolle, "Hybrid ai for medieval illuminations analysis," in *workshop FAIR Heritage: Digital Methods, Scholarly Editing, and Tools for Cultural and Natural Heritage*, 2020.
- [22] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
- [23] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proceedings of International Conference on Learning Representations*, Y. Bengio and Y. LeCun, Eds., 2015.
- [24] C. B. El Vaigh, N. Garcia, B. Renoust, C. Chu, Y. Nakashima, and H. Nagahara, "Gcnboost: Artwork classification by label propagation through a knowledge graph," in *Proceedings of the 2021 International Conference on Multimedia Retrieval*, 2021, pp. 92–100.
- [25] Y. Wu, Y. Nakashima, and N. Garcia, "Not only generative art: Stable diffusion for content-style disentanglement in art analysis," in *Proceedings of the 2023 ACM International Conference on Multimedia Retrieval*, ser. ICMR '23. New York, NY, USA: Association for Computing Machinery, 2023, p. 199–208. [Online]. Available: <https://doi.org/10.1145/3591106.3592262>
- [26] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," *arXiv preprint arXiv:1804.02767*, 2018.
- [27] E. Sirin, B. Parsia, B. C. Grau, A. Kalyanpur, and Y. Katz, "Pellet: A practical owl-dl reasoner," *Journal of Web Semantics*, vol. 5, no. 2, pp. 51–53, 2007.