



Aouinti, Fouad, Victoria Eyharabide, Xavier Fresquet, and Frédéric Billiet. 2022. "Illumination Detection in IIF Medieval Manuscripts Using Deep Learning." *Digital Medievalist*, 15(1): pp. 1–18. DOI: <https://doi.org/10.16995/dm.8073>



Open Library of Humanities

Illumination Detection in IIF Medieval Manuscripts Using Deep Learning

Fouad Aouinti, STIH Laboratory, Sorbonne Université, FR, fouad.aouinti@sorbonne-universite.fr

Victoria Eyharabide, STIH Laboratory, Sorbonne Université, FR, maria-victoria.eyharabide@sorbonne-universite.fr

Xavier Fresquet, IReMus Laboratory, Sorbonne Université, FR, xavier.fresquet@sorbonne-universite.fr

Frédéric Billiet, IReMus Laboratory, Sorbonne Université, FR, frederic.billiet@sorbonne-universite.fr

Illuminated manuscripts are essential iconographic sources for medieval studies. With the massive adoption of IIF, old and new digital collections of manuscripts are accessible online and provide interoperable image data. However, finding illuminations within the manuscripts' pages is increasingly time consuming. This article proposes an approach based on machine learning and transfer learning that browses IIF manuscript pages and detects the illuminated ones. To evaluate our approach, a group of domain experts created a new dataset of manually annotated IIF manuscripts. The preliminary results show that our algorithm detects the main illuminated pages in a manuscript, thus reducing experts' search time.



1. Introduction

§1 In the Middle Ages, drawings and paintings in manuscripts often adorned the pages to illustrate or comment on the text. Illuminations are handmade decorations that adorn a manuscript, such as initials, borders, and illustrations. The history of illumination probably goes back to antiquity. Some documents have survived showing obvious decorative elements: for example, ancient Homeric poems (e.g., the Papyrus 114, Bankes Homer, created during the 2nd century and written in Greek [British Library 2021]) or biblical codices (e.g., the Codex Alexandrinus, a fifth-century Christian manuscript of the Greek Bible, containing the majority of the Greek Old Testament and the Greek New Testament: <https://www.bl.uk/collection-items/codex-alexandrinus>). During the Carolingian period, artistic centres with particular styles develop in different cities around workshops (Court School of Charlemagne, Rheimsian style, Touronian style, Drogo style, Court School of Charles the Bald). During the Merovingian period, symbolic ornaments appear in the manuscripts' pages and, in the 11th and 12th centuries, with the development of artistic centres in France (Benedictine monasteries, including the remarkable Abbey of Cluny in Burgundy), manuscripts sometimes reflect a local style. Illuminations thus became iconographic elements essential in medieval studies and sources of complementary historical knowledge in other fields of study (archaeology, art history, architecture, musicology, or literature). For example, in musicology, medieval iconography analysis brings useful information on the instruments' nature, instruments' organological characteristics, physical characteristics, or playing methods (the research team proposing this article focuses mainly on music in the Middle Ages; therefore, we have developed the first multi-object database in medieval musical iconography: Musiconis - <http://musiconis.huma-num.fr/fr/>), as shown in **Figure 1**: We observe a figure holding a key used to tune a string on a harp. Moreover, the medieval image is not always, as is sometimes thought, a "Bible for the illiterate, but rather a creation in which form and content are intrinsically linked, with a strong plastic thought and a visual language in which analogies, metaphors, and associations of meaning are numerous" (Baschet 2008). These images therefore also give us essential information about, for example, the role and place of sound in medieval society.

§2 For medievalists, art historians, and all scientists searching for these iconographic sources, one of the main problems lies in finding the right illuminations within the manuscripts' pages. The document analysis is often already partially or completely done within library catalogues, with precise references to the images located in the manuscripts' textual description: the position (folio number), techniques used, and subjects treated in each illumination are often already mentioned through



Figure 1: Illumination of a character “B.” Folio 7r, “St Omer” or “Rede” Psalter, MS 98, Christ Church, University of Oxford, 1201–1300. <https://digital.bodleian.ox.ac.uk/objects/1159e59e-b06e-44d9-bb0a-b5127a5b36fe/surfaces/9da1b4c2-2c04-40e9-a171-dc4b656af141>

associated metadata. However, few digital tools collect and classify these images. Where such tools exist – as in the CNRS *Initiale* database (Initiale 2021) – this meticulous work has been carried out manually. Unfortunately, it is limited to specific national, regional, and local collections. With the development of the IIF since 2011, researchers have been provided with massive digital iconographic data, which is a decisive factor in identifying medieval illustrations of various kinds (drawings, illuminations, diagrams). As of 2019, one billion images have become available, and about 300 million of them date from the medieval period. Some individual projects have emerged recently using deep learning and statistical learning techniques using IIF files (Moreux 2019; Nakamura 2019; Eyharabide et al. 2020). However, as far as we know, no team or project has yet undertaken a study to use state-of-the-art deep learning methods to find illuminations in medieval manuscripts across IIF libraries. When automatically searching for illuminations, the main difficulty lies in separating the text from the images (the background containing neither text nor images must also be considered). This article proposes an approach based on transfer learning that browses a collection of IIF manuscript pages and detects the illuminated ones. We used the HBA corpus (Mehri et al. 2017) for training, a pixel-based annotated dataset released at the ICDAR 2017 Competition on Historical Book Analysis. To evaluate our approach, a group of domain experts created a new dataset of manually annotated IIF manuscripts. The preliminary results show that our algorithm detects the main illuminated pages in a

given manuscript, thus reducing experts' search time and opening up new avenues for the massive exploration of digitized iconographic data.

§3 This article is organized as follows. First, after a brief review describing the context of the documents on which we are working - ornate early medieval illuminated manuscripts - we resume the state-of-the-art in the study of illustrations in historical documents using deep learning. Second, we present how we train our model on annotated manuscript images. Finally, we show our experimental results on IIIF images to extract the illuminations using the previously trained model, share feedback of the human expert's evaluations of our tool, and discuss the next steps for future research perspectives.

2. From manuscript to illumination

§4 For several centuries, almost all book production was concentrated in the most important monastic workshops: the *scriptoria*. The production of the manuscript is a complex process and teamwork. The parchment maker begins by choosing the skins to form the parchments after a treatment that makes them suitable to receive the ink. Once the notebooks have been formed, copyists can begin their work. Several copying techniques are used: direct copying from another manuscript or copying under dictation, allowing all copyists to write the same text simultaneously. Parallel lines are drawn on the parchment to obtain a carefully written text. First, the lines are marked with drypoint and then with a pencil lead. Some parchments still have "stitches," small regular holes along the margins' edge to trace this grid.

§5 Either at the beginning or at the end of each work, copyists frequently inserted a colophon (i.e., a small text giving the date and their name). The sheets could then be passed on to the rubricator, who wrote the titles (the rubrics), often in colour, and traced a few simple miniatures. Finally, the illuminator passed upon the richest manuscripts and sometimes created real masterpieces around important pages. Our work concentrates on decorations (i.e., illuminated illustrations in Western manuscripts). We aim at isolating this essential iconographic element using modern methods based on machine learning and transfer learning. Indeed, over the last ten years, the IIIF format development has given medievalists simultaneous access to manuscripts previously kept in separate libraries and museums. This exceptional opening of sources naturally goes hand in hand with an exponential multiplication of the pages to be analyzed. On this point, iconographic research can benefit from the contributions of computer science research results. To give a few examples, Western manuscripts dating from the 13th-15th centuries are generally works of several hundred

pages. As mentioned before, only some pages are decorated, and sometimes only one has a detailed illumination of large proportions.

§6 This general observation is the starting point for our proposal, which seeks to answer the following questions: To what extent can computer vision techniques isolate illuminations in digitized manuscripts as precisely as a human expert? If so, can this method be extended to a broad range of manuscripts held in different institutions around the world?

3. State of the art

§7 Nowadays, many large libraries around the world widely adopt the IIIF format for sharing images and documents. The new tool proposed by the Bibliissima consortium is an excellent example of this initiative. It provides access to more than 77,000 IIIF manifests of pre-1800 documents held in 24 of the largest digital collections. It offers a unified access to a set of digital data on ancient documents and allows one to view, consult, and query digitized manuscripts, catalogues, and specialized databases on various aspects of the study of the written heritage. Computer research teams have already analyzed several aspects of the automatic analysis of images in digitized documents. The first experiments in the use of AI and IIIF appeared as early as 2017, particularly in countries that adopted IIIF at an early stage (in France, the USA, or Japan, for example). However, our work shows some novelties since it is not limited to a collection or institution. The work proposed in Schlecht, Carqué, and Ommer (2011) presents a model-based detector for four types of gestures visualized in images of the Middle Ages. This is a method for learning a small set of templates representative of the gesture variability in medieval legal manuscripts through the application of an efficient version of normalized cross-correlation to vote for gesture position, scale, and orientation. In 2019, the work carried out around *dhSegment* (Oliveira, Seguin, and Kaplan 2018) implements an open-source CNN-based pixel-wise predictor coupled with task-dependent post-processing blocks. It is a very flexible tool that has analyzed illustrations (ornaments), measured using the standard Intersection over Union (IoU) metric. The next year, T. Monnier and M. Aubry (Monnier and Aubry 2020) presented *docExtractor*, a generic tool for extracting visual elements (text lines, illustrations) from historical documents without requiring any real data annotation. They demonstrated an excellent efficiency using an off-the-shelf solution across several datasets. They relied on a fast generator of rich synthetic documents and designed a fully convolutional neural network, which generalized better than a detection-based approach, using an encoder-decoder architecture combining ResNet with U-Net. Moreover, they

introduced the *IlluHisDoc* dataset to work on a satisfactory evaluation of illustration segmentation in historical documents. The work that we propose in this article is in line with these different teams' approaches. To further develop research on the subject, we propose an innovative concept described below.

4. Our proposal

§8 This paper proposes to use machine learning to solve an illumination detection issue in multiple IIF manuscript images using deep learning techniques. This classic machine learning problem has already been tested for object detection in modern images (post-1600) but never for ancient and rich IIF manuscript images.

§9 To detect illuminated pages in IIF medieval manuscripts, we implemented a neural network using the latest version of the state-of-the-art YOLO object detection system (version n°5) (Jocher et al. 2021). **Figure 2** shows an overview of our approach. Initially, as the HBA images are not annotated with bounding boxes surrounding the illuminations, for each page, we randomly generated 10000 rectangles and labelled them (as *graphic*, *text*, or *the_rest*) from their pixel-based annotations (Section 4.1). After preprocessing, we trained and validated YOLO on the HBA dataset (Section 4.2). Then, for each IIF manuscript manifest, we extracted the URLs of the page's images (Section 4.3). After evaluating each page's image on our model (Section 4.4), we merged all the bounding boxes obtained to discover the illuminations (Section 4.5). Later, for each illumination found on a page, we generated the corresponding IIF URL with only the

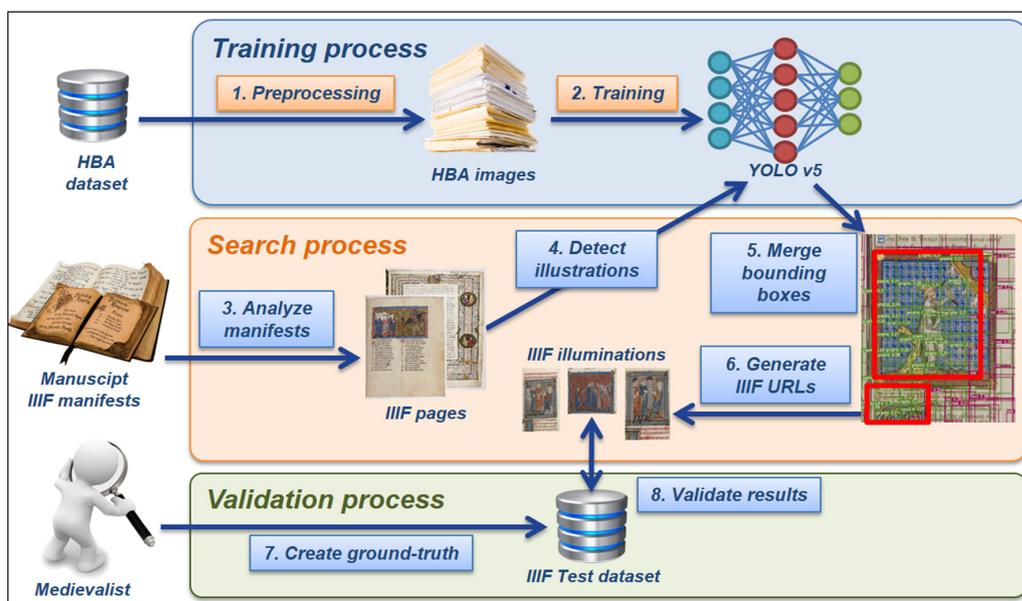


Figure 2: Overview of our proposal to detect illuminations in IIF medieval manuscripts.

page's area containing that image. For example, from the page https://www.e-codices.unifr.ch/loris/ubb/ubb-F-II-0020/ubb-F-II-0020_0054r.jp2/full/full/0/default.jpg, we obtained: https://www.e-codices.unifr.ch/loris/ubb/ubb-F-II-0020/ubb-F-II-0020_0054r.jp2/499,1790,702,661/full/0/default.jpg. Finally, we evaluated our model on a ground truth dataset of IIF images from different libraries (Valenciennes/BnF, Basel, Cambridge, Princeton) manually annotated by medievalist experts and discussed the results (Section 5).

4.1. Preparing the HBA dataset

§10 The HBA dataset (Mehri et al. 2017) comes from the French digital library Gallica. It contains 4436 real scanned ground-truthed images from 11 historical books published between the 13th and 19th centuries. These are grayscale or colour images that were digitized at 300 or 400 dpi and saved in the TIFF format. The HBA corpus also provides 1429 pixel-level annotated ground-truthed images. The annotations include six different content classes: *Graphics*, *Normal text*, *Capitalized text*, *Handwritten text*, *Italic text*, and *Footnote text*. Data preparation is the most crucial step in any deep learning task. Due to its nature, the HBA dataset contains some specificities that could significantly influence the learning process. The first challenge is the images' large and scattered dimensions, with a maximum of 5500 pixels in width and 8000 pixels in height. A second challenge is the HBA dataset's characteristics, including heterogeneity, illustration style differences, and complex layouts (irregular spacing, varying text column widths, margins). Finally, the HBA corpus can not be considered a purely random set of images because inherent links between images can exist since they belong to the same book. Before the images' feature extraction process, a preprocessing step was applied to the HBA images. First, we regroup all the different text styles (normal, capitalized, handwritten, italic, and footnote) in a single category to obtain three classes: *graphic*, *text*, and *the_rest*. Then, we eliminated 1063 images that do not contain graphics, were damaged, or with scanning defects. As a result, from the original 1429 pixel-annotated images using six classes, we obtained 366 pixel-annotated images using three classes after the preprocessing step.

4.2. Training YOLO on the HBA dataset

§11 For training YOLO, the 366 selected HBA images must be labelled and annotated with bounding boxes surrounding the illuminations. Therefore, for each image, we randomly generated 10000 bounding boxes of various sizes and assigned them, based on background colour threshold, the appropriate class (*graphic*, *text*, *the_rest*) to the corresponding rectangle in the ground-truthed image. This new dataset was split

into a training (80%) and a testing (20%) subset. The training subset contains 293 ground-truthed images (80% of images) and their corresponding annotation files. Each annotation is composed of: the object class (0: graphic, 1: text or 2: the_rest), the x and y centre coordinates, and the height and width of the bounding box (values in the range [0.0, 1.0]) as per YOLO specifications. As a result, the prepared HBA dataset contains 94,084 bounding boxes annotations classified into three classes: graphic (31,181 annotations), text (31,138 annotations), and the_rest (31,765 annotations). Thus, the annotations' distribution is strongly balanced and accurate for training purposes.

4.3. Retrieving pages' images from IIIF manifests

§12 According to the IIIF specification (<https://iiif.io/api/presentation/3.0/>), an IIIF manifest describes the structure and layout of a digitized object or other collection of images. Based on the *Shared Canvas Data Model* (<http://iiif-io.us-east-1.elasticbeanstalk.com/model/shared-canvas/>), a manifest has the following structure. Each IIIF manifest has one or more sequences. Each sequence (the order of the views) must have at least one canvas. Each canvas (a virtual container representing a page or view) should have one or more content resources (such as images or texts in a manuscript). The association of images with their respective canvases is done via annotations. Thanks to this well-organized IIIF specification, we read the JSON-LD manifest to extract the URLs' list to the images of each manuscript's page.

4.4. Extracting illustrated decorations

§13 Once the labelled dataset is in YOLO format, we describe our dataset parameters and modify the YOLOv5 configuration, which specifies the information on the names and number of classes. In our case, we chose the smallest and fastest base model of YOLOv5, which is yolov5s. During training, the YOLOv5 pipeline creates training data batches with three kinds of augmentations: scaling, colour space adjustments, and mosaic augmentation. We can visualize the training data ground truth as well as the augmented training data. Once the training is completed, we take our trained model and make inference on test images as illustrated in **Figure 3(a)**. For inference, we invoke the adapted weights with a specification of the model confidence. The higher confidence required, the fewer predictions made.

4.5. Merging bounding boxes

§14 As presented in **Figure 3(a)**, there is a strong overlap of rectangles/squares areas detected by the YOLO model. For this purpose, we developed an algorithm for combining a set of bounding boxes into bigger ones that would encompass areas of interest.

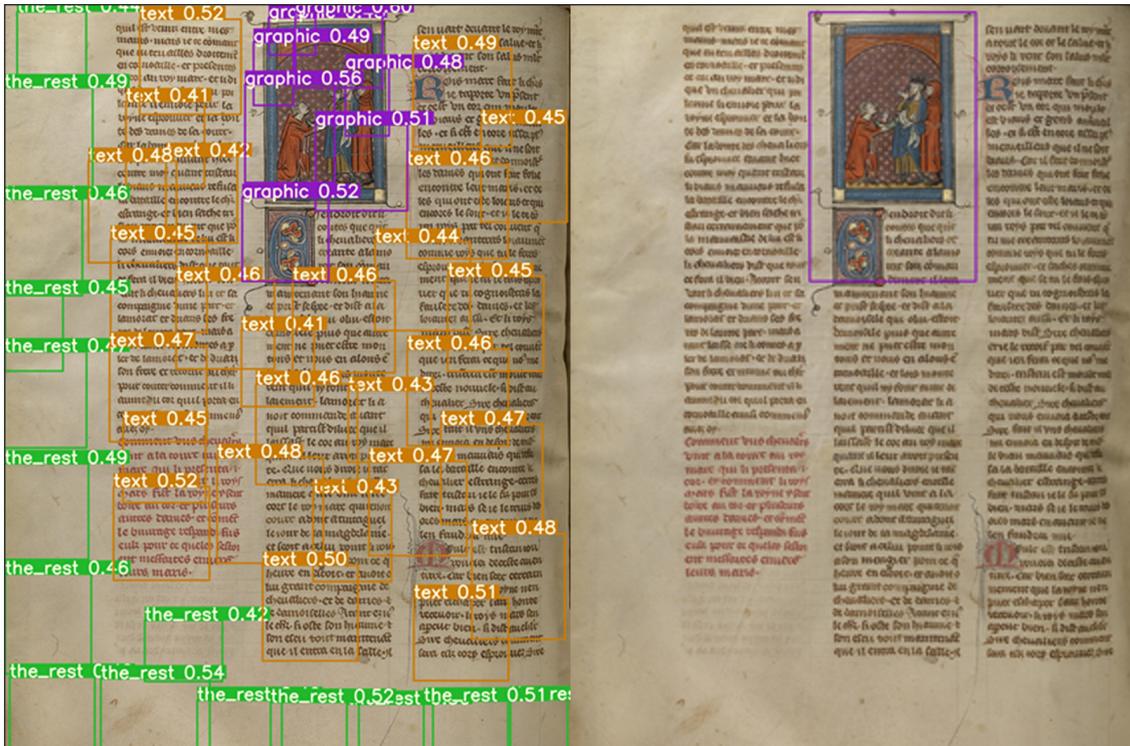


Figure 3: (a) Examples of the obtained bounding boxes, (b) Illumination detection after merging bounding boxes.

Algorithm 1 Merging overlapping bounding boxes.

```

bboxes ← bounding boxes
clusters ← merged illuminations
while not satisfaction of the stop criterion do
  clusters ← []
  for box in bboxes do
    matched ← false
    for cluster in clusters do
      if box and cluster overlap then
        matched ← true
        cluster.left ← min(cluster.left, box.left)
        cluster.right ← max(cluster.right, box.right)
        cluster.bottom ← min(cluster.bottom, box.bottom)
        cluster.top ← max(cluster.top, box.top)
      end
    end
    if not matched then
      append box to clusters
    end
  end
  bboxes ← clusters
end

```

§15 In this regard, we evaluated our proposed algorithm’s capabilities to merge overlapping graphics on different images, including the detected bounding boxes in **Figure 3(a)**. The obtained results in **Figure 3(b)** are satisfactory.

4.6. Generating IIIF URLs from bounding boxes

§16 IIIF Image API 3.0 (<https://iiif.io/api/image/3.0>) specifies a standardized URL syntax for transmitting an image through a standard HTTP request. This syntax applies to two different types of requests for the same image: the request for the image itself (image file) and the request for technical information on the image (JSON file). URL pattern for requesting an image must conform to the following URL template: `{scheme}://{server}/{prefix}/{identifier}/{region}/{size}/{rotation}/{quality}.{format}`

§17 We are interested in the region parameter, which defines the rectangular portion of the underlying image content to be returned. By default, the region is usually specified by the value *full*, which means that the full image should be returned. After applying the proposed algorithm for merging overlapping bounding boxes, we replace the default region value *full* by the coordinates of the illumination detected by our system following the syntax specified by the IIIF standard.

5. Evaluation and discussion

§18 We asked a group of expert medievalists to create a ground truth dataset of IIIF images to test our approach. As manual annotations are expensive in time and knowledge, they annotated only one manuscript from each library (BnF/Valenciennes, Basel, Cambridge, Princeton) for a total of 1835 annotated images. In this section, we present the testing dataset and discuss the results obtained.

5.1. Creating a testing IIIF dataset of medieval decorations and illustrations

§19 A group of medieval experts creates an IIIF dataset of medieval decorations and illustrations to test and validate our deep learning methods for identifying illustrations in medieval manuscripts. They selected four manuscripts with the particularity of having illustrations of various shapes, sizes, and characteristics. Their approximate production dates range from 800 to 1300. **Table 1** summarizes the manuscripts’ characteristics. Although this dataset applies only to a limited number of documents, we selected manuscripts from different libraries to demonstrate that the remote IIIF image analysis system could be applied to various collections. In our case, the documents are held at Gallica (Valenciennes/BnF), the Universitätsbibliothek Basel, the University of Cambridge, and the University of Princeton.

§20 The first manuscript (see **Figure 4(a)**), which is kept at the library of Valenciennes – part of the *Bibliothèque nationale de France* (BnF) collections – is a document mainly containing a book of Apocalypse published in the 9th century and is probably of Spanish origin. This Latin consists of 40 leaves (27×20 cm) bound with embossed calf binding. This document describes about forty paintings. To each painting corresponds a legend borrowed from the text of the Apocalypse. In addition to the IIF images, each illustration (i.e., illuminated initials, marginal illustrations) was also identified by human experts in the CNRS/IRHT Initiale database.

§21 The second manuscript is the *History of Animals* (see **Figure 4(b)**), a zoological work written in Greek around 343 BC by Aristotle. The document comprises 152 leaves (25x18 cm) and dates from the 13th century. The decoration of the initials in this manuscript is particularly rich in metaphors.

§22 The third manuscript (MS Nn.2.36) is currently being held by the University of Cambridge (**Figure 5(a)**). It is composed of 235 leaves (27x21 cm). It was probably produced in either 1120–1130 and is richly ornamented. Hollow rectangular headpieces precede Gospels with vegetative ornament (in red, blue, green, and white paint and gold). Narrower rectangular headpieces precede the beginning of other major sections.

Table 1: IIF testing dataset's characteristics.

Manuscript	Valenciennes BM, 0099 (092)	Garrett MS. 28.	MS Nn.2.36	F II 20
Title	Apocalypse	De animalibus	Gospel book	Bible
Date	800–900	1200–1300	1120–1130	1270
Location	Valenciennes/BnF	Universität Basel	University of Cambridge	Princeton University
Number of pages	100	316	489	930
Illustrated pages	40	19	10	72



Figure 4: (a) MS 99, f. 013, Valenciennes/BnF, (b) Aristoteles, *De animalibus* libri XIX, f. 31r. Universitätsbibliothek Basel.



Figure 5: (a) Gospel book (MS Nn.2.36), f. 179r, Cambridge University, (b) Garrett MS. 28. Princeton.

§23 The fourth manuscript (see **Figure 5(b)**), the Garrett MS. 28, is a Bible from circa the 1270s, probably produced in England. The document is a double psalter, consisting of two parallel versions in the left and right columns. It is composed of 458 leaves (24x16,8 cm). Decoration varies within the document from 1-line pen-work initials to 8-line historiated initials.

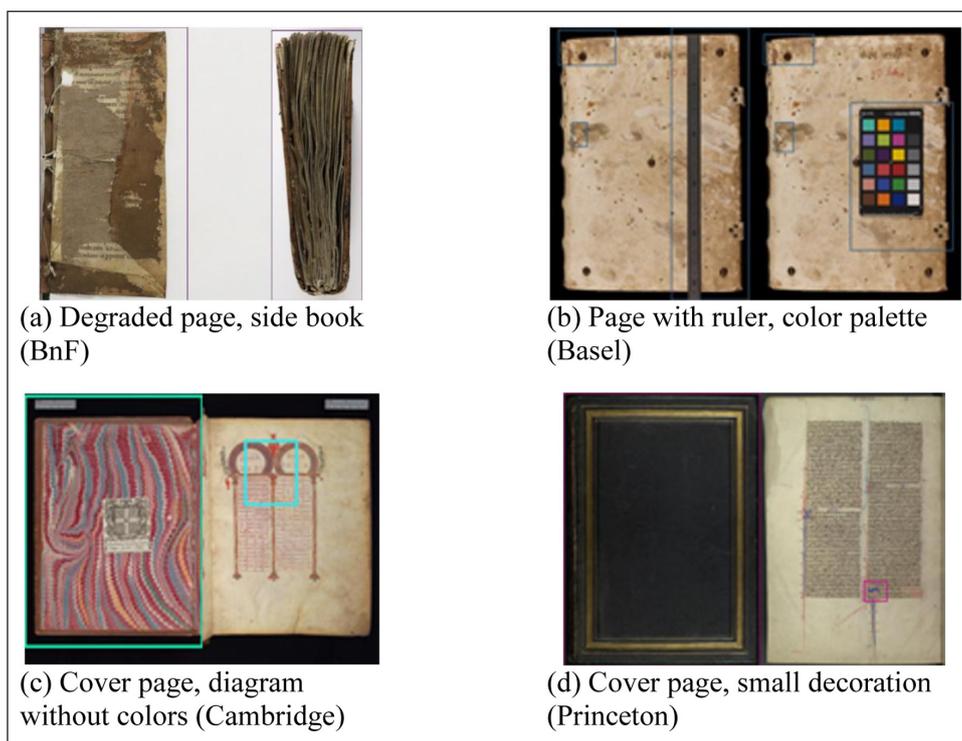
5.2. Experimental results

§24 The dimensions of the object(s) to detect can be very small compared to the whole page. Thus, downsizing the images could lead to completely erasing some of those objects, making them impossible to retrieve. Because of this aspect ratio challenge, multiple models are trained and tested with different input image sizes to select the most efficient. To evaluate the performance of our proposed approach, a confusion matrix is computed to get an overview of several classification metrics: precision, recall, and f-score. The obtained results of our system implementation, related to the training of the YOLO model from scratch, are presented in the following **Table 2**. At first glance, the results may seem slightly different, but this is more due to the nature of the test manuscripts, including degradation properties, scanning defects, and the method considered by domain experts to annotate the decorations.

§25 Starting with the Valenciennes/BnF manuscript, the 14 false positives (FP) predicted by our classification system are usually due to their shapes as shown in **Figure 6(a)**. These are the IIF images of cover pages, side book, plan, etc. Unfortunately, this kind of page images shares some visual similarities with HBA training images, which is why our deep neural network learns more easily some common features. Moreover, among the 40 illuminations annotated by domain experts, the YOLO-trained model can detect the 37 IIF targeted pages with a precision of 0.73. **Table 3** (row Valenciennes/BnF) shows us an example of IIF illumination URLs generated by our system from the whole page URL. In this example, the deep neural network predicts

Table 2: Experimental results of illumination detection.

Library	IIIF images	Annotated images	TP	TN	FP	FN	Precision	Recall	F-score
BnF	100	40	37	46	14	3	0.73	0.93	0.81
Basel	316	19	16	287	10	3	0.62	0.84	0.71
Cambridge	489	10	9	461	18	1	0.33	0.9	0.49
Princeton	930	72	70	824	34	2	0.67	0.97	0.8

**Figure 6:** Examples of false positive results.

all the graphics except the middle one (animal) because, in the learning phase, it is not trained on this kind of object to extract these characteristics. Thus, overall it can be noticed that with a small training dataset and with the issues affecting test image layout, our model pre-trained on the HBA dataset can detect the specified class in the IIIF manuscript of Valenciennes/BnF.

§26 A second implementation is carried out on the Basel manuscript to test our detection system's performance. In this case, we can see that our model can predict 16 of 19 true positives (TP) and detect illuminations with a precision of 0.62. As presented in **Table 3** (row Basel), our object detector generates the IIIF illumination correctly.

Table 3: Samples of true positive results with the generated IIIF URL.

IIIF Original image	IIIF generated URL
	<ul style="list-style-type: none"> - Valenciennes/BnF's page: https://gallica.bnf.fr/iiif/ark:/12148/btv1b84525958/f31/full/full/0/native.jpg - Obtained illuminations: https://gallica.bnf.fr/iiif/ark:/12148/btv1b84525958/f31/195,426,1210,2074/full/0/native.jpg - https://gallica.bnf.fr/iiif/ark:/12148/btv1b84525958/f31/1818,297,1420,2024/full/0/native.jpg - https://gallica.bnf.fr/iiif/ark:/12148/btv1b84525958/f31/466,2621,900,1201/full/0/native.jpg - https://gallica.bnf.fr/iiif/ark:/12148/btv1b84525958/f31/1981,2764,838,613/full/0/native.jpg
	<ul style="list-style-type: none"> - Original Basel's page: https://www.e-codices.unifr.ch/loris/ubb/ubb-F-II-0020/ubb-F-II-0020_0031r.jp2/full/full/0/default.jpg - Obtained illumination: https://www.e-codices.unifr.ch/loris/ubb/ubb-F-II-0020/ubb-F-II-0020_0031r.jp2/664,2657,430,774/full/0/default.jpg
	<ul style="list-style-type: none"> - Original Cambridge's page: https://images.lib.cam.ac.uk/iiif/MS-NN-00002-00036-000-00239.jp2/full/full/0/default.jpg - Obtained illumination: https://images.lib.cam.ac.uk/iiif/MS-NN-00002-00036-000-00239.jp2/477,1312,4120,2980/full/0/default.jpg
	<ul style="list-style-type: none"> - Original Princeton's page: https://iiif-cloud.princeton.edu/iiif/2/51%2F95%2F77%2F5195771668e94b969ea06a06a4ae1fd1%2Fintermediate_file/full/3000,/0/default.jpg - Obtained illumination: https://iiif-cloud.princeton.edu/iiif/2/51%2F95%2F77%2F5195771668e94b969ea06a06a4ae1fd1%2Fintermediate_file/75,225,1359,3519/3000,/0/default.jpg

However, as illustrated in **Figure 6(b)**, the ten false positives (FP), which were expected, are mainly caused by scanning defects, especially marginal edges, colour palette, and ruler. Taking the example of the IIF image containing a colour palette, our system considers it a set of graphics and thus returns the detected areas as an illumination after merging bounding boxes.

§27 On **Table 2** are also displayed the results of the third experimentation, which tested the model on the Cambridge manuscript. In fact, the obtained result does not reflect the truth for the simple reason that, in addition to the images with scanning defects, the YOLO-trained model detects other objects in IIF images, which are usually diagrams and decorations but are not annotated by domain experts. Eighteen false positives (FP) are therefore predicted, an example of which is depicted in **Figure 6 (c)**. Our system can obviously help experts in annotating images and detecting graphic objects, thus reducing their research time. On the other hand, in images where there are real illuminations, our model efficiently retrieves the IIF URL as shown in **Table 3** (row Cambridge). Indeed, the trained model still performs well enough, even though the specified class does not achieve sufficient scores.

§28 Regarding Princeton's latest test manuscript, the proposed neural network has considerable accuracy. In terms of precision, we obtained 0.67. In the case of the whole image containing a small decoration, the system considers it false positive (FP) because it is simply not annotated by the experts. 34 images are then predicted by our model as images where there are illuminations, including images with small decoration as shown in the **Figure 6(d)**. Because of its significant impact on the precision for detecting illuminations, a post-processing step can therefore be added to specify the minimum dimensions of illuminations. Additionally, as it can be seen in the example of **Table 3** (row Princeton), it should be noted that 70 images (at the exception of 2 among 72 images) are predicted true positives (TP).

§29 One of the major challenges for the neural network in this work is identifying illustrations on pages in IIF format, based on training with the HBA dataset, and transfer learning on reduced IIF data. This difficulty poses several detection problems because, as we were able to see quickly, the images concerned are of very different natures: Full-page painted images, richly ornamented initials of medium size, small illustrated initials, illustrations in the margins with different shapes and positions, extremely wide range of colours, use of the background for certain images (diagrams, schematics). All these types of images, easy to recognize for the medievalist's trained eye, constitute entirely different categories for the machine. Therefore, we shall see

how the system reacts to this diversity, and will naturally discuss the sometimes surprising results we have observed.

§30 We can notice the interest of the proposed approach in illuminations detection in IIF medieval manuscripts from these results. Nevertheless, it should be noted that the used dataset is relatively small with only 366 images (94084 annotations) and without using preloaded features. A larger annotated dataset would therefore be necessary to improve the performance and ability of our model to generalize.

6. Conclusions and perspectives

§31 In this work, we trained YOLOv5 on the HBA dataset to detect illustrations in IIF medieval manuscripts using transfer learning techniques. We have also shown that the proposed approach has positive results, which can be explained by the nature of the digitized documents themselves (front and back of the document, multicoloured sampling scale for shooting). This work is exploratory and deserves to be deepened. Therefore, we propose several future research perspectives that our team will pursue. First, to test the model on a much larger dataset, we propose to use the data proposed by Biblissima. In particular, that of the Initiale database to measure the accuracy of the model at the pixel level. Next, build a virtual library of illuminations through several libraries, and explore the tools available to annotate these images online. Finally, build a database of several thousand IIF images dedicated to music, dance, musical performance, and sound. Based on the Musiconis project, we will then explore IIF as a key element for advancing research in musicology using machine learning techniques.

Acknowledgements

We gratefully acknowledge support from the CNRS/IN2P3 Computing Center (Lyon - France) for providing computing and data-processing resources. Special thanks to Jean Philippe Moreux (BnF) for his expertise and advice. We would also like to thank Marcus Liwicki (LUT) for pointing out the potential interest of exploring YOLO for the analysis of images in IIIF format.

Competing interests

The authors have no competing interests to declare.

Contributions

Authorial contributions

Authorship is alphabetical after the drafting author and principal technical lead. Author contributions, described using the CASRAI CredIT typology, are as follows:

Fouad Aouinti : FA

Frédéric Billiet : FB

Victoria Eyharabide : VE

Xavier Fresquet : XF

Authors are listed in descending order by significance of contribution. The corresponding author is FA

Conceptualization: VE, XF, FA

Software: FA VE

Validation: XF

Supervision: FB

Editorial contributions

Recommending editor:

Franz Fischer, Università Ca' Foscari Venezia, Italy

Recommending referees:

Benjamin Albritton, Stanford University, US

Peter Bell, Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany

Section, copy, and layout editor:

Shahina Parvin, University of Lethbridge, Canada

References

Baschet, Jérôme. 2008. *L'icônographie médiévale*. Gallimard.

British Library. 2021. "Bankes Homer." Accessed November 15. <https://www.bl.uk/collection-items/the-bankes-homer>.

Eyharabide, Victoria, Xavier Fresquet, Fouad Aouinti, Marcus Liwicki, and Susan Boynton. 2020. "Towards the Identification of Medieval Musical Performance Using Convolutional Neural Networks and IIF." In IIF Annual Conference, Harvard University and MIT, USA.

Initiale. 2021. "Initiale: Catalogue de manuscrits enluminés." Accessed November 15. <http://initiale.irht.cnrs.fr>

Jocher, Glenn, Alex Stoken, Jirka Borovec, NanoCode012, Ayush Chaurasia; TaoXie, Liu Changyu, AbhiraV, Laughing, tkianai, yxNONG, AdamHogan, lorenzomamma, AlexWang1900, Jan Hajek, Laurentiu Diaconu, Marc, Yonghye Kwon, oleg, wanghaoyang0106, Yann Defretin, Aditya Lohia, ml5ah, Ben Milanko, Benjamin Fineran, Daniel Khromov, Ding Yiwei, Doug, Durgesh, Francisco Ingham. 2021. "Ultralytics/Yolov5: v5.0 - YOLOv5-P6 1280 Models, AWS, Supervise.ly and YouTube Integrations." Zenodo. Accessed November 21. <https://zenodo.org/record/4679653#.YZMUKS971QJ>

Mehri, Maroua, Pierre Héroux, Rémy Mullot, Jean-Philippe Moreux, Bertrand Couasnon, and Bill Barrett. 2017. "HBA 1.0: A Pixel-Based Annotated Dataset for Historical Book Analysis." In *Proceedings of the 4th International Workshop on Historical Document Imaging and Processing*, 107–112. DOI: <https://doi.org/10.1145/3151509.3151528>

Monnier, Tom, and Mathieu Aubry. 2020. "docExtractor: An Off-the-Shelf Historical Document Element Extraction." In *Proceedings of the 17th International Conference on Frontiers in Handwriting Recognition (ICFHR)*, 91–96. IEEE. DOI: <https://doi.org/10.1109/ICFHR2020.2020.00027>

Moreux, Jean-Philippe. 2019. "Using IIF for Image Retrieval in Digital Libraries: Experimentation of Deep Learning Techniques." In IIF Conference, Göttingen, Germany.

Nakamura, Satoru. 2019. "Development of Content Retrieval System of Scrapbook 'Kunshujo' Using IIF and Deep Learning." In IIF Conference, Göttingen, Germany. DOI: https://doi.org/10.1007/978-3-030-00066-0_39

Oliveira, Sofia Ares, Benoit Seguin, and Frederic Kaplan. 2018. "dhSegment: A Generic Deep-Learning Approach for Document Segmentation." In *Proceedings of the 16th International Conference on Frontiers in Handwriting Recognition (ICFHR)*, 7–12. IEEE.

Schlecht, Joseph, Bernd Carqué, and Björn Ommer. 2011. "Detecting Gestures in Medieval Images." In *Proceedings of the 18th IEEE International Conference on Image Processing*, 1285–1288. IEEE. DOI: <https://doi.org/10.1109/ICIP.2011.6115669>

