DIGITAL HERITAGE (2025)
S. Campana, D. Ferdani, H. Graf, G. Guidi, Z. Hegarty, S. Pescarin, and F. Remondino (Editors)

The IlluminAI project: a deep neural network and immersive visualization system to enhance illuminated manuscripts

V. Minisini[1,2], G. Gosti[2] B. Fanini[2]

[1] Sapienza University of Rome, Department of Classics, Italy
[2] National Research Council (CNR), Institute of Heritage Science, Italy

Abstract

In museum and archive digital catalogues, illuminated manuscript pages can often be found within heterogeneous groups of reproductions, coexisting with other types of artworks and objects. With them, figurative miniatures share the depicted subjects that are recognizable, regardless of the medium used, for their specific iconography. Thanks to the use of a visual vocabulary still common today, illuminations are also the element that mostly attracts the non-academic public, making the often-incomprehensible content partly accessible despite the language. The paper will present IlluminAI, a project still in progress, which aims at the enhancement of late medieval and Renaissance illuminated codices using artificial intelligence through an immersive visualization system capable of automatically recognizing manuscript sheets, analyzing their content, and relating specimens with similar illustrations or artworks from the same theme. After some brief references to contextualize the work, we will expose the first completed phase of the research focusing on the original dataset composition before outlining the chosen semi-automatic labeling strategy and the interactive machine learning approach. This was used to create with transfer learning a model able to recognize manuscript pages and identify inside of them five characteristic layout elements. We will then switch to the second ongoing part of the project with the design of the immersive Web3D system, based on the open-source ATON framework, that will give users the possibility to explore, inspect, compare and query large amounts of images in a three-dimensional space. The data aggregation criteria and the presentation modes will be described with particular attention to the spatial organization and novel 3D interfaces.

CCS Concepts

• ***Applied computing*** → ***Fine arts;*** • ***Computing methodologies*** → ***Activity recognition and understanding;*** • ***Human-centered computing*** → *Visual analytics;*

1. Introduction

Until the mid-15th century and the introduction in Europe of movable-type printing by Johannes Gutenberg, manuscript volumes produced in the scriptoria of monasteries and courts were the main vehicle for the preservation of knowledge. In the most valuable codices, the written word was accompanied by rich decorations, both abstract and figurative, which shared subjects and iconography with monumental arts. Especially in late medieval and Renaissance periods, the stylistic similarity between miniatures and other artworks is at its maximum with painters who also work as illustrators. The massive digitization of illuminated manuscripts places today's researchers in the unprecedented position to easily compare multiple volumes also if held by different institutions and studying the variations between them to better understanding the peculiarities of specific geographical areas or historical periods as well as following the cultural changes that are at the basis of the fortune of a particular depiction rather than another. The most widely used viewers today, like the open-source Mirador, are two-dimensional and allow to match only a limited number of images at a time, often implement annotations, but without being able to manipulate them or seek personalized paths and fruition strategies inside the codex. The paper presents IlluminAI, an ongoing project for the development of an innovative web3d visualization system that will permit users to move comfortably inside the digitized volumes, searching for miniatures, and capable of relating reproductions of different artworks with the same subject. Section 2 outlines the state of the art while Section 3 proceeds to present the purpose-built training dataset and the deep neural network created to automatically recognize manuscript sheets within heterogeneous groups of images as well as identify five layout elements: figurative miniatures, text, decorations, music, and ornamental initials. After exposing the interactive machine learning technique adopted and the dataset partition in three subsets of increasing size, in Section 4 we discuss the immersive interface under development to organize, compare and query large amounts of images in a three-dimensional space before the Conclusions in Section 5.

DOI: 10.2312/dh.20253119

2. Related Work

Historical document and manuscript research has benefited greatly from the use of deep neural network techniques for automatic content analysis [LM20]. Specific datasets concerning digitized handwritten and early printed texts [NSML22] are composed of materials merged according to a shared language, the creation period and resource-distributing institution [MHM*17] but their dimensions are always small if compared to those with natural images. In addition to character recognition [II23], attempts of illumination retrieval with different approaches [BGC14, AEFB22] have gradually shifted the attention towards the codex artistic aspects and specific tasks like the identification of figurative miniature subject. While few iconographic classification experiments have been conducted on the illuminated pages [Man23], considerable results have already been achieved with artworks [RML23] often taking advantage of the Iconclass classification system [MF21, SSR*24] used by different institutions around the world. Among the several neural network architectures ResNet [HZRS16] has often been chosen for its capacity to perform well on labeling and detection tasks with small datasets, and it can be easily trainable also with larger ones, given that its implementation of residual functions allows better error propagation. The annotation phase was implemented with Label Studio which gave a direct interface for the domain expert annotations [TMHL25]. To an interactive machine learning approach overview refer to [PTH13]. Regarding Immersive Analytics (IA), it's a rapidly expanding sector of information visualization due to the promising possibilities in 3D visual encoding and spatial interaction [SDBC*24]. The advantages of this approach over traditional non-immersive interfaces are now widely recognized [MCH*18] also in cultural heritage field [KAC20] where the opportunity to design complex but intuitive systems to re-contextualize artworks is opening up new strategies in user engagement.

3. The Deep Neural Network

The first step in creating the IlluminAI interface that will allow users to explore the history of illuminated manuscripts through their more intuitive figurative content and stylistic evolution was to train a dedicated model to automatically recognize pages from other digitized artwork reproductions and make it able also to distinguish five characteristic layout elements [MGF24]. The initial phase of the project was therefore dedicated to the creation of a dataset specifically built and its semi-automatic annotation through an iterative training cycle. In it a domain expert and an AI expert cooperate to create a model with domain-specific knowledge using transfer learning to rapidly compile three increasingly larger databases. Specifically, the domain expert starts by annotating or correcting the labels of a smaller dataset that the AI expert then uses to train the neural network using transfer learning and hyperparameter optimization. The resulting model is then used to generate labels for a larger dataset that the domain expert checks and manually modifies, if necessary, with Label Studio before the process starts again.

3.1. The Dataset

As far as we know, existing manuscript datasets don't include other artistic typologies so we merged an original one with reproductions of incunabula and later printed copies but also artworks and objects normally present in digital catalogs and often featuring figurative subjects as decoration. All files come from different institutional databases like the J. Paul Getty Collection of Los Angeles and Open Access images or released under creative common license CC-BY were preferred in anticipation of the dataset final publication. By using multiple sources we tried to obtain a certain variety of scanning and shooting conditions as well as reproduction quality, selecting both double and single pages, overall views and detail shots. In total the dataset, which cover the chronological span extends from the 9th to the 20th century, consists of 46.014 items divided into two macro-categories, "Pages" and "Art", and into ten subgroups distinguished by a specific identifier: manuscript sheets (P.01), printed pages (P.02), paintings (A.01), engravings (A.02), drawings (A.03), sculptures (A.04), stained glass windows (A.05), tapestries (A.06), art prints (A.07) and other objects (A.08). Since the project is focused on manuscripts, the number of images dedicated to these is higher than all other typologies and constitute 68% of the total, chosen to cover a wide variety of layouts and styles. Pages from printed books are 15% while the rest is made up of the "Art" category which is particularly rich in paintings (4%), other graphic works (4%) and sculptures (3%). Each element in our dataset has been classified using an alphanumeric code that allow to quickly trace back the original owner institution and provide essential information about the depicted subject, the century of production, and the specific author or, if unknown, at least the geographical area of provenance. To identify the represented theme, we also developed an intuitive system of abbreviations associated with the equivalent Iconclass codes necessary, in the immersive interface design, to group and order the images into thematic clusters. An example of the result thus obtained is "P.01_S(Mark)_14BMGetty" where "P.01" indicates the category and the subclass the image belongs to, in this case manuscript page, "S(Mark)" identifies the subject as Saint Mark while "14" is the century of creation, "BM" the author's initials and "Getty" is the provenance. The sheets without illustrations, instead, have been classified according to the presence of ornaments, text or musical scores considering the dominant component. For images belonging to the "Pages" category, the sequence following the subject specification may be common to multiple items as it uniquely associates all the sheets originally part of the same volume with the identification code assigned to it. This in turn is linked to a record containing the main bibliographic information, the original resource URL and the IIIF Manifest if available. Four partitions were thus obtained useful to balance first the composition of the dataset and then dividing it into three training groups of 400, 4.000 and 41.614 images.

3.2. The Training

To limit human effort in the annotation phase, the domain expert manually labeled in full only the first group of 400 images to report the presence or co-presence in pages of figurative miniatures (Fig), decorations (Deco), text (Text), music scores (Mus) and ornamental initials (Let) with a minimum of one and a maximum of four tags for a total of 818. The "Fig" label has also been given to all images in the "Art" category. Furthermore, the system automatically recovered from the classification described above the categories to which each item belongs, and the specific subclass as-

sociated with the type of digitized object. We have selected ResNet architecture pretrained on ImageNet dataset [RDS*15] and, since we had a relatively small dataset, we started with ResNet-18 because models with fewer parameters are less likely to be in an over-parametrization regime. ResNet-50 was also tested with no significant differences or improvements observable in the results. In transfer learning we tried both Fine-Tuning on all weights and Fixed Features as well as considering two gradient descent training methods: Stochastic Gradient Descent (SGD) and Adam. In the first group, for SGD we used an exponential learning rate scheduler with a starting learning rate 0.001, momentum 0.9, learning rate scheduler step size 7, and gamma 0.1. For Adam, we settled on a learning rate of 0.001. The set was divided, allocating 70% of the images for training and 30% to test and, in almost all cases, Fine Tuning and SGD gave the best performance even if the difference was not very large. The accuracy rate recorded with this combination is particularly high in distinguishing pages from art (97%) and manuscript material from printed one (95%) as well as in identifying music (97%), text (96%) and figurative miniatures (93%). The lowest values, however, were recorded for decorations (86%) and especially illuminated initials (82%) mainly due to the intrinsic ambiguity between simple ornamental letters and historiated initials. Considering these promising results, we proceeded to predict the labels of the next group using the models with the highest accuracy level for each target but, while in the first step we were able to keep the learning rate constant, in the second one for both SGD and Adam was necessary to vary the value between 0.00001 and 0.1. Of 4.000 images, only 1.529 (38%) required the domain expert manual intervention, saving several hours of work to reach a total of 7.204 annotations. Also with this set the combination of Fine Tuning and SGD gave the best accuracy results with 98% for "Mus", 95% for "Fig", 94% for "Text", 92% for "Deco" and 91% for "Let". Even in the recognition of pages and handwritten material there was a slight improvement, rising to 99% and 96% respectively. The models were then used one last time to predict the labels of the third and most consistent set of 41.614 items. This time, only 35% of the 41.614 images analyzed were modified, leading to a total of 63.677 labels of which 48.719 were assigned to the manuscript pages only with an average of two labels per copy. The precision rates in all tasks were very good with 99% for "Text", "Fig" and "Mus", 96% for "Deco" and 88% for "Let". The recall values were also evaluated as satisfactory with 99% for "Text", 98% for "Mus", 94% for "Fig" and "Deco", while the "Let" reached 91%. The accuracy instead was 99% for "Text" and "Mus", 98% for "Deco", 97% for "Fig" and 96% for "Let". Finally, for the distinction between "Pages" and "Art" as well as between "P.01" and "P.02", all the metrics reached a rate of 99%.

4. The interactive and immersive visualization system

We are now gradually moving from classification to object detection using the activation zones obtained so far to automatically locate and extract the miniatures, cropping their specific area before using image segmentation to get semantic masks of the figures represented inside of them. The ultimate purpose of the model is to make interactive an immersive web3D system under development, based on the open-source ATON framework [FFD*21] by CNR-ISPC and among the services of the H2IOSC Project - *Humanities and cultural Heritage Italian Open Science Cloud*. In its design we were inspired by Aby Warburg's Mnemosyne Bilderatlas, a series of large thematic panels that abandons the traditional sequential organization of information typical of the book format to leave the concepts free to branch out into two-dimensional space. The Atlas uses the images as nodes in a complex network of meaning and constitutes a perfect model also for immersive analytics [Hri16]. To enable the user to spatially inspect and interrogate large quantities of pages in a virtual 3D environment, juxtapose similar miniatures and compare them with other artworks of the same subject through the use of common web browsers, we proceeded by creating the structure underlying the webapp retrieval system based on an internal database automatically generated from several CSV files. The first and largest table derived from the model contains the fundamental information about the images, the paths to retrieve them from the cloud and the labels generated by the deep neural network. The second file, instead, consists of the Iconclass classification system whose codes have been associated with our abbreviations and the corresponding identifiers in the iconographic vocabulary of the Cultural Objects Name Authority (CONA) elaborated by the Getty Research Institute. A third document brings together all the bibliographic information on the more than two hundred volumes considered and a similar fourth one dedicated to the "Art" category is in preparation. In addition to the items used for the training phase, 13.845 images from the National Photographic Cabinet collection of the Central Institute for Cataloguing and Documentation of Italian cultural heritage were added to increase the items for comparison. The collected material is being spatially organized using the classification previously described and the hierarchical structure of Iconclass to divide the material into ten bigger thematic clusters. Each of them is divided into 6 subgroups related to specific narrative moments or characters except one which brings together the reproductions without illustrations based on the other layout elements detected. The subsets take spatially the form of segments in the bigger cluster as shown in Figure1 where the images are further divided chronologically in a radial direction, with the reproductions of the oldest artworks closer to the viewer and the most recent but less numerous ones on the outer circle. The user will be able to access IlluminAI either by using a virtual reality headset (HMD) to experiment the complete spatial interface, or through his computer or even personal smartphone. In addition to multi-level filtering contents, will be possible to highlight material of interest and relate images to each other as well as move them in space to confront their depictions and comparing the linked historical information. Once the development phase is completed, we plan to carry out user studies and usability tests to improve the human-interface interaction before the final release.

5. Conclusions

The IlluminAI project aims to demonstrate how immersive visualization interfaces can also be used for illuminated manuscripts, allowing the development of innovative enhancement strategies for the reuse of already acquired material even if not accompanied by rich metadata and specific historical information. The integrated deep neural network automatically extracts parameters from the image, organizing them into the tables needed by the webapp to function with minimal human effort. The system ability of relating

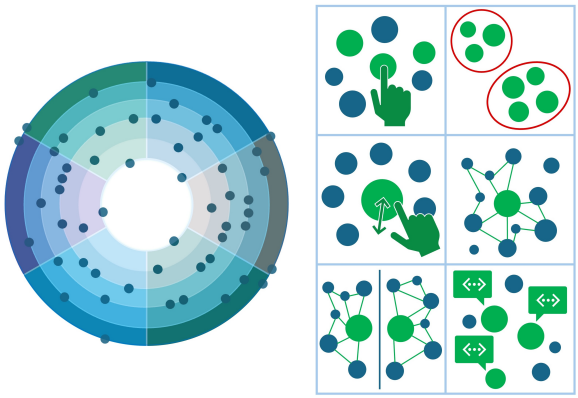

Figure 1: *On the left the organization scheme of images in clusters while next to six of the possible interactions: selection, grouping, zoom, relation, confront, information comparison.*

artwork reproductions that differ in terms of period and technique but share the same subject based on the user's interests is the constitutive feature of its flexibility and adaptability. When all phases of the project are completed, the hope is that it will help make objects, originally the prerogative of a few, accessible to all, freeing delicate cultural artifacts from their physical limitations.

Acknowledgments

This research is funded by H2IOSC Project - Humanities and cultural Heritage Italian Open Science Cloud (https://www.h2iosc.cnr.it/) funded by the European Union – NextGenerationEU – NRRP M4C2 - Project code IR0000029-CUP-B63C22000730005.

References

[AEFB22] AOUINTI F., EYHARABIDE V., FRESQUET X., BILLIET F.: Illumination detection in IIIF medieval manuscripts using deep learning. *Digital Medievalist 15*, 1 (2022), 1–18. doi:10.16995/dm.8073. 2

[BGC14] BORGHESANI D., GRANA C., CUCCHIARA R.: Miniature illustrations retrieval and innovative interaction for digital illuminated manuscripts. *Multimedia Systems 20* (2014), 65–79. doi:10.1007/s00530-013-0315-3. 2

[FFD*21] FANINI B., FERDANI D., DEMETRESCU E., BERTO S., D'ANNIBALE E.: ATON: an open-source framework for creating immersive, collaborative and liquid web-apps for cultural heritage. *Applied Sciences 11*, 22 (2021), 11062. doi:10.3390/app112211062. 3

[Hri16] HRISTOVA S.: Images as data: cultural analytics and Aby Warburg's Mnemosyne. *International Journal for Digital Art History*, 2 (Oct. 2016), 117–132. doi:10.11588/dah.2016.2.23489. 3

[HZRS16] HE K., ZHANG X., REN S., SUN J.: Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2016), Springer, pp. 770–778. doi:10.1109/CVPR.2016.90. 2

[II23] ISLAM M. A., IACOB I. E.: Manuscripts character recognition using machine learning and deep learning. *Modelling 4*, 2 (2023), 168–188. doi:10.3390/modelling4020010. 2

[KAC20] KOEBEL K., AGOTAI D., CÖLTEKIN A.: Exploring cultural heritage collections in immersive analytics: challenges, benefits, and a case study using virtual reality. In *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* (2020), vol. XLIII-B4-2020, p. 599–606. doi:10.5194/isprs-archives-XLIII-B4-2020-599-2020. 2

[LM20] LOMBARDI F., MARINAI S.: Deep learning for historical document analysis and recognition — a survey. *Journal of Imaging 6*, 10 (2020), 110. doi:10.3390/jimaging6100110. 2

[Man23] MANONI P.: AI4MSS : un esperimento di intelligenza artificiale alla Biblioteca Apostolica Vaticana. In *Guardando oltre i confini : partire dalla tradizione per costruire il futuro delle biblioteche : studi e testimonianze per i 70 anni di Mauro Guerrini*, Bergamin G., Possemato T., (Eds.). AIB - Associazione Italiana Biblioteche, 2023, pp. 231–244. doi:10.1400/294065. 2

[MCH*18] MARRIOTT K., CHEN J., HLAWATSCH M., ITOH T., NACENTA M. A., REINA G., STUERZLINGER W.: *Immersive Analytics*, vol. 11190 of *Lecture Notes in Computer Science*. Springer, 2018, ch. Just 5 questions: toward a design framework for immersive analytics, pp. 259–288. doi:10.1007/978-3-030-01388-2_9. 2

[MF21] MILANI F., FRATERNALI P.: A dataset and a convolutional model for iconography classification in paintings. *Journal on Computing and Cultural Heritage 14*, 4 (2021), 46. doi:10.1145/3458885. 2

[MGF24] MINISINI V., GOSTI G., FANINI B.: Transfer learning for Renaissance illuminated manuscripts: starting a journey from classification to interpretation. In *3rd Workshop on Artificial Intelligence for Cultural Heritage (IAI4CH 2024) co-located with the 23rd International Conference of the Italian Association for Artificial Intelligence (AIxIA 2024)* (2024), vol. 3865, CEUR Workshop Proceedings, pp. 30–35. URL: https://ceur-ws.org/Vol-3865/04_paper.pdf. 2

[MHM*17] MEHRI M., HÉROUX P., MULLOT R., MOREUX J. P., COÜASNON B., BARRETT B.: 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 (HIP '17)* (2017), Association for Computing Machinery, pp. 107–112. doi:10.1145/3151509.3151528. 2

[NSML22] NIKOLAIDOU K., SEURET M., MOKAYED H., LIWICKI M.: A survey of historical document image datasets. *International Journal on Document Analysis and Recognition (IJDAR) 25* (2022), 305–338. doi:10.1007/s10032-022-00405-8. 2

[PTH13] PORTER R., THEILER J., HUSH D.: Interactive machine learning in data exploitation. *Computing in Science e Engineering 15*, 5 (2013), 12–20. doi:10.1109/MCSE.2013.74. 2

[RDS*15] RUSSAKOVSKY O., DENG J., SU H., KRAUSE J., SATHEESH S., MA S., HUANG Z., KARPATHY A., KHOSLA A., BERNSTEIN M., BERG A. C., FEI-FEI L.: ImageNet large scale visual recognition challenge. *International Journal of Computer Vision 115* (2015), 211–252. doi:10.1007/s11263-015-0816-y. 3

[RML23] RESHETNIKOV A., MARINESCU M. C., LOPEZ J. M.: DEArt: dataset of european art. In *Computer Vision – ECCV 2022 Workshops* (2023), vol. 13801 of *Lecture Notes in Computer Science*, Springer, p. 218–233. doi:10.1007/978-3-031-25056-9_15. 2

[SDBC*24] SAFFO D., DI BARTOLOMEO S., CRNOVRSANIN T., SOUTH L., RAYNOR J., YILDIRIM C., DUNNE C.: Unraveling the design space of immersive analytics: a systematic review. *IEEE Transactions on Visualization and Computer Graphics 30*, 1 (Jan. 2024), 495–506. doi:10.1109/TVCG.2023.3327368. 2

[SSR*24] SPRINGSTEIN M., SCHNEIDER S., RAHNAMA J., STALTER J., KRISTEN M., MÜLLER-BUDACK E., EWERTH R.: Visual narratives: large-scale hierarchical classification of art-historical images. In *2024 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)* (2024), pp. 7195–7205. doi:10.1109/WACV57701.2024.00705. 2

[TMHL25] TKACHENKO M., MALYUK M., HOLMANYUK A., LIUBIMOV N.: Label Studio: data labeling software, 2020-2025. Open source software available from https://github.com/HumanSignal/label-studio. URL: https://github.com/HumanSignal/label-studio. 2