

ÉCOLE NATIONALE DES CHARTES
UNIVERSITÉ PARIS, SCIENCES & LETTRES

Pierre Husson

diplômé de master

INDEXATION ICONOGRAPHIQUE ET INTELLIGENCE ARTIFICIELLE

Expérimentations autour du Répertoire des
tableaux italiens dans les collections publiques
françaises (RETIF)

Mémoire pour le diplôme de master
« Technologies numériques appliquées à l'histoire »

2025

Résumé

Ce mémoire s'inscrit dans le cadre du projet IconographIA, mené par l’Institut national d’histoire de l’art en collaboration avec le consortium Huma-Num Pictoria. Ce projet a pour objectif d’indexer les œuvres recensées dans le RETIF (Répertoire des tableaux italiens dans les collections publiques françaises), en mobilisant les technologies relevant de l’intelligence artificielle et de la vision par ordinateur. L’indexation iconographique s’appuie sur le thésaurus Garnier et est destinée à être intégrée dans la plateforme AGORHA.

Au cours de cette étude, une réflexion est d’abord conduite sur les enjeux de cette pratique documentaire, en particulier pour la description des œuvres picturales, ainsi que sur ses instruments et son rôle dans l’exploration des bases de données. Les expérimentations menées pour automatiser cette tâche sont ensuite exposées, en détaillant les pistes envisagées, les défis rencontrés et les limites de cette méthode. Enfin, ce mémoire expose des approches alternatives à l’automatisation du processus, s’appuyant sur la vision par ordinateur pour faciliter le travail de l’indexeur ou pour repenser l’exploration au sein des bases de données.

Mots-clés : intelligence artificielle ; vision par ordinateur ; histoire de l’art, indexation iconographique ; thésaurus ; iconographie ; peinture italienne ; RETIF

Informations bibliographiques : Pierre Husson, *Indexation iconographique et intelligence artificielle. Expérimentations autour du Répertoire des tableaux italiens dans les collections publiques françaises (RETIF)*, mémoire de master « Technologies numériques appliquées à l’histoire », dir. Ségolène Albouy, École nationale des chartes, 2025.

Avant-Propos

Ce mémoire s'inscrit dans le cadre d'un stage mené conjointement à l'Institut national d'histoire de l'art et au sein du consortium Huma-Num pictorIA, au cours de ma seconde année de Master 2 « Technologies numériques appliquées à l'histoire » à l'École nationale des Chartes. L'objectif principal de ce stage était l'indexation iconographiques de peintures issues du RETIF (Répertoire des tableaux italiens dans les collections publiques françaises) à l'aide d'outils d'intelligence artificielle. Ce projet, intitulé *Iconographia*, s'est déroulé sur une période de quatre mois, du 1^{er} avril au 31 juillet 2025, sous la direction de Federico Nurra et de Julien Schuh. J'ai été chargé de cette mission en raison de ma double formation en histoire de l'art et en technologies numériques. Diplômé du deuxième cycle de l'École du Louvre en 2023, j'avais consacré mes recherches à la gravure d'interprétation en Italie à l'Époque moderne, ce qui a nourri mon intérêt pour l'histoire de la peinture italienne et enrichi mes connaissances dans ce domaine. Au cours de ce projet, j'ai abordé la question principalement sous l'angle de l'historien de l'art, étant la seule personne à travailler à temps plein pour le mener. Il convient toutefois de préciser que, malgré mon intérêt pour les technologies d'intelligence artificielle, ma maîtrise des modèles d'apprentissage profond et de leurs usages reste partielle. Cette limite a entraîné certaines difficultés à manipuler ces technologies, que je présenterai au fil des chapitres. Loin de constituer un simple frein, elles ont ouvert de nouvelles perspectives et permis d'apporter un regard neuf aux pratiques d'indexation iconographique.

Je tiens tout d'abord à exprimer ma gratitude à ma directrice de mémoire, Ségolène Albouy, qui m'a accompagné tout au long de sa rédaction, et à Gennaro Toscano de présider mon jury de soutenance. Je remercie chaleureusement Federico Nurra et Julien Schuh pour leur encadrement au cours de ce stage, ainsi que mes collègues à l'INHA, Pierre-Yves Laborde, Jean-Christophe Carius, Margaux Faure, Maria Francesca Bocchi et Natacha Grim pour leurs conseils et leur soutien. Ma reconnaissance va également à mes collègues stagiaires et amis Camille Samsa, Fantin Le Ber et Mathieu Taybi, ainsi qu'à l'ensemble de mes camarades de master, pour les échanges stimulants que nous avons partagés. Je remercie également Marion Charpier et Emmanuelle Bermès pour leur accompagnement. J'adresse un merci particulier à Teoman Akgönül et Agathe Ménétrier, pour leur disponibilité et leur écoute,

ainsi qu'à Johanna Daniel, dont les réflexions ont enrichi les miennes. Je voudrais témoigner ma gratitude envers Hans Brandhorst et Etienne Posthumus d'avoir accepté de s'entretenir avec moi au début de mes recherches. Enfin, je souhaite remercier mes amis et ma famille pour leur présence et leur soutien constant : Eléa Dargelos, Solveig d'Aboville, Grégoire Hignard, Thaïs Wsevoloski, Mathieu Darrietort, Théophile Piquemal-Tabou et Inès Hennequin. Pour leur aide dans la relecture de ce mémoire, j'adresse un merci particulier à ma maman et à Adèle Bugaut.

Bibliographie

- ADAMOVA (Alexandra) et WINGERT (Rosanne), *Le projet Gallica Images. L'intelligence artificielle pour l'accès aux grands corpus d'images*, 18 juin 2024, URL : <https://www.youtube.com/watch?v=IVMfpR96jrk> (visité le 04/09/2025).
- AMAR (Muriel), *Les fondements théoriques de l'indexation, une approche linguistique*, Paris, 2000.
- ASPERTI (Andrea), DESSÌ (Leonardo), TONETTI (Maria Chiara) et WU (Nico), *Does CLIP perceive art the same way we do ?*, version 1, 2025, DOI : 10.48550/ARXIV.2505.05229.
- Association française de normalisation (éd.), *Documentation : recueil de normes françaises 1993*, 5e éd, Paris-La Défense, 1993.
- AUBRY (Mathieu), COSTINER (Lisandra) et JAMES (Stuart), « Artificial Intelligence and Art History : A Necessary Debate ? », *Histoire de l'art*, 87–1 (2021), p. 55-60, DOI : 10.3406/hista.2021.3936.
- Augmented Artwork Analysis (AAA)*, URL : <https://icar.cnrs.fr/aaa/> (visité le 25/08/2025).
- BAROCCHI (Paola), BISOGNI (Fabio), CORTI (Laura) et AL., *First International Conference on Automatic Processing of Art History Data and Documents, Pisa, Scuola Normale Superiore, 4-7 September 1978 : conference transactions*, 2 t., [s.l.], 1978.
- BARRAL I ALTET (Xavier), « Introduction », dans *Dictionnaire critique d'iconographie occidentale*, Rennes, 2003, p. 866.
- BARRAL I ALTET (Xavier) et AL., *Dictionnaire critique d'iconographie occidentale*, Rennes, 2003.
- BEAUD (Mathieu), MARCHESIN (Isabelle) et BIAY (Sébastien), *Ontologie du christianisme médiéval en images*, 2023, URL : <https://omci.inha.fr/> (visité le 27/08/2025).
- BERGEON LANGLE (Ségolène) et CURIE (Pierre), *Peinture & dessin : vocabulaire typologique et technique*, 2 t., Paris, 2009 (Principes d'analyse scientifique).
- BERMÈS (Emmanuelle) et CHARPIER (Marion), « Repenser les collections patrimoniales par le prisme de l'IA 2025 », dans *Conférence nationale sur les applications de l'Intelligence artificielle*, Dijon, France, 2025, URL : <https://hal.science/hal-05138697>.

- BIBLIOTHÈQUE NATIONALE DE FRANCE, *Thésaurus iconographique de Mandragore*, 2024,
URL : <https://api.bnf.fr/fr/thesaurus-mandragore> (visité le 24/08/2025).
- *Les projets en intelligence artificielle à la BnF*, BnF - Site institutionnel, URL : <https://www.bnf.fr/fr/les-projets-en-intelligence-artificielle-la-bnf> (visité le 01/09/2025).
- BOUTÉ (Édouard), JULLIARD (Virginie), GÖDICKE (David), PAILLER (Fred) et ECREMENT (Victor), « PANOPTIC, un outil d'exploration par similarité de vastes corpus d'images », dans *Humanistica 2024, Association francophone des humanités numériques*, Meknès, Maroc (7-9 mai 2024), 2024, URL : <hal-04687627>.
- BRANDHORST (Hans), *A word is worth a thousand pictures - Why the use of Iconclass will make Artificial Intelligence smarter*, 2019, URL : https://iconclass.org/testset/ICONCLASS_and_AI.pdf (visité le 01/08/2025).
- *A virtual library for iconographic research*, ICONCLASS, 6 avr. 2023, URL : <https://forum.iconclass.org/t/a-virtual-library-for-iconographic-research/206> (visité le 03/09/2025).
- *ICONCLASSplus*, URL : <https://iconclass.org/help/plus> (visité le 04/09/2025).
- BRANDHORST (Hans) et POSTHUMUS (Etienne), « ICONCLASS, A key to collaboration in the digital humanities », dans *The Routledge Companion to Medieval Iconography*, dir. Colum Hourihane, Routledge, Londres et New York, 2017, p. 201-218.
- *Survey of versions of the Iconclass system*, ICONCLASS, URL : <https://iconclass.org/help/aboutb> (visité le 03/09/2025).
- CARON (Mathilde), TOUVRON (Hugo), MISRA (Ishan), JÉGOU (Hervé), MAIRAL (Julien), BOJANOWSKI (Piotr) et JOULIN (Armand), *Emerging Properties in Self-Supervised Vision Transformers*, version 2, 2021, DOI : [10.48550/ARXIV.2104.14294](https://arxiv.org/abs/2104.14294).
- CHAMPION (Erik Malcolm), « Digital humanities is text heavy, visualization light, and simulation poor », *Digital Scholarship in the Humanities* (, 11 nov. 2016), fqw053, DOI : [10.1093/llc/fqw053](https://doi.org/10.1093/llc/fqw053).
- COMAN (Florin), *L'histoire de l'art et l'informatique documentaire*, 2 t., Atelier national reproduction des thèses, Université Lille III ; Diffusion, Aux Amateurs de livres, Lille, Paris, 1988.
- COMMISSARIAT GÉNÉRAL AU DÉVELOPPEMENT DURABLE, *Consommation domestique en eau potable*, notre-environnement, 11 janv. 2023, URL : <https://www.notre-environnement.gouv.fr/themes/societe/le-mode-de-vie-des-menages-ressources/article/consommation-domestique-en-eau-potable> (visité le 31/08/2025).

- CORTI (Laura), SCHMITT (Marilyn) et AL., *Automatic processing of art history data and documents : [2d. international conference] Pisa, Scuola normale superiore, september 24-27, 1984*, 3 t., Pise, 1984.
- Crotos*, URL : <https://zone47.com/crotos/> (visité le 26/08/2025).
- DE SAINT-OURL Édouard et KERMORVANT (Christopher), « Le projet HikarIA : étude et mise en valeur du patrimoine photographique par l'intelligence artificielle », *Culture et Recherche*–147 (2024), p. 20-24.
- DE LUCA (Livio), « A digital ecosystem for the multidisciplinary study of Notre-Dame de Paris », *Journal of Cultural Heritage*, 65 (2024), p. 206-209, DOI : [10.1016/j.culher.2023.09.011](https://doi.org/10.1016/j.culher.2023.09.011).
- DE LUCA (Livio), ABERGEL (Violette), GUILLEM (Anaïs), MALAVERGNE (Olivier), MANUEL (Adeline), NÉROULIDIS (Ariane), ROUSSEL (Roxane), ROUSSET (Miled) et TOURNON (Sarah), *L'écosystème numérique n-dame pour l'analyse et la mémorisation multi-dimensionnelle du chantier scientifique Notre-Dame-de-Paris*, SCAN'22, 2022.
- DE LUCA (Livio), GUILLEM (Anaïs) et RÉBY (Kévin), « Intelligence humaine, collective et artificielle dans le chantier scientifique et de restauration de Notre-Dame de Paris », *Culture et Recherche*–147 (2024), p. 8-12.
- DEEPL, *Translate Text*, DeepL Documentation, URL : <https://developers.deepl.com/api-reference/translate> (visité le 02/09/2025).
- DHLAB (YALE), *PixPlot*, URL : <https://dhlab.yale.edu/projects/pixplot/> (visité le 30/08/2025).
- DI LENARDO Isabella, SÉGUIN (Benoît) et KAPLAN (Frédéric), « Visual Patterns Discovery in Large Databases of Paintings », dans *Digital Humanities 2016 : Conference Abstracts*, Cracovie, 2016, p. 169-172, URL : <https://dh2016.adho.org/abstracts/348> (visité le 05/08/2025).
- Documentation de FastAI*, URL : <https://docs.fast.ai/> (visité le 20/08/2025).
- Documentation de la librairie Python RDFLib*, URL : <https://rdflib.readthedocs.io/en/stable/> (visité le 01/09/2025).
- DUHAIME (Douglas) et LEONARD (Peter), *PixPlot* (Dépôt Github), Github, 2018-2023, URL : <https://github.com/pleonard212/pix-plot>.
- FÉLIBIEN (André), *Conferences de l'Academie royale de peinture et de sculpture. Pendant l'année 1667*, 1667.
- FONDATION INTERNATIONALE POUR LE LIMC, *Lexicon Iconographicum Mythologiae Classicae*, IX t., Zürich, 1981.
- (éd.), *Lexicon Iconographicum Mythologiae Classicae (LIMC) : Supplementum 2009*, Düsseldorf, 2009.

- FRANCESCHINI (Chiara), « Classifying image content in visual collections, a selective history », dans *The Routledge Companion to Medieval Iconography*, dir. Colum Hourihane, Routledge, Londres et New York, 2017, p. 184-191.
- GARNIER (François) et MINISTÈRE DE LA CULTURE, *Thésaurus iconographique, système descriptif des représentations*, Paris, 1984.
- GERVEREAU (Laurent), *Voir, comprendre, analyser les images : 2020* (Repères), DOI : 10.3917/dec.gerv.2020.01.
- GETTY RESEARCH INSTITUTE, *About CONA and IA*, révisé le 31 juillet 2024, URL : https://www.getty.edu/research/tools/vocabularies/cona/about.html#history_ia (visité le 03/09/2025).
- *About the AAT*, révisé le 26 juillet 2024, URL : <https://www.getty.edu/research/tools/vocabularies/aat/about.html> (visité le 03/09/2025).
 - *Getty Vocabularies*, URL : <https://www.getty.edu/research/tools/vocabularies/> (visité le 03/09/2025).
- GIANESELLI (Matteo), TCHERNIA (Marie) et VOLLE (Nathalie), *Interroger la base en ligne les peintures italiennes en France (RETIF) sur Agorha (INHA)*, Blog de l'APAHAU, 8 déc. 2011, URL : <http://blog.apahau.org/interroger-la-base-en-ligne-les-peintures-italiennes-en-france-retif-sur-agorha-inha/> (visité le 08/08/2025).
- GOOGLE, *Bienvenue dans Colab*, URL : <https://colab.research.google.com/?hl=fr> (visité le 01/09/2025).
- GOOGLE LABS, *Google NotebookLM*, URL : <https://notebooklm.google/> (visité le 02/09/2025).
- HAFFENDEN (Chris), REKATHATI (Faton) et RENDE (Emma), *Unearthing forgotten images with the help of AI*, The KBLab Blog, 20 oct. 2023, URL : <https://kb-labb.github.io/posts/2023-10-20-unearthing-forgotten-images-with-the-help-of-ai> (visité le 04/09/2025).
- HE (Kaiming), ZHANG (Xiangyu), REN (Shaoqing) et SUN (Jian), *Deep Residual Learning for Image Recognition*, version 1, 2015, DOI : 10.48550/ARXIV.1512.03385.
- HEITZINGER (Thomas) et STORK (David G.), « Improving semantic segmentation of fine art images using photographs rendered in a style learned from artworks », *IS&T International Symposium on Electronic Imaging 2022*, 34–13 (16 janv. 2022), p. 169-1-169-5, DOI : 10.2352/EI.2022.34.13.CVAA-169.
- HikarIA*, URL : <https://hikaria.org/> (visité le 18/08/2025).
- History of the Index*, The Index of Medieval Art, URL : <https://ima.princeton.edu/history/> (visité le 27/08/2025).

- HUMA-NUM, *Infrastructure multi-cibles pour l'usage de l'IA en sciences humaines et sociales*, Documentation de l'infrastructure Huma-Num, 27 août 2025, URL : <https://documentation.huma-num.fr/infrastructure-IA/> (visité le 01/09/2025).
- HUMM (Bernhard G.), *Fascinating with Open Data : openArtBrowser*, 2020, URL : https://ceur-ws.org/Vol-2535/paper_2.pdf (visité le 27/08/2025).
- *openArtBrowser*, URL : <https://openartbrowser.org/fr/> (visité le 03/09/2025).
- INSEE, *Émissions de gaz à effet de serre par habitant*, Tableau de bord de l'économie française, 6 janv. 2025, URL : https://www.insee.fr/fr/outil-interactif/5367857/details/90_DDE/92_DEV/92G_Figure7 (visité le 30/08/2025).
- INSTITUT NATIONAL D'HISTOIRE DE L'ART, AGORHA, URL : <https://agorha.inha.fr/> (visité le 03/09/2025).
- *Iconographie musicale : répertoire d'œuvres d'art à sujets musicaux publiées par Albert Pomme de Mirimonde*, AGORHA, URL : <https://agorha.inha.fr/database/17> (visité le 01/09/2025).
 - *Les Envois de Rome en peinture et sculpture, 1804-1914*, AGORHA, URL : <https://agorha.inha.fr/database/80> (visité le 01/09/2025).
 - *Recensement de la peinture produite en France au XVIe siècle*, AGORHA, URL : <https://agorha.inha.fr/database/50> (visité le 01/09/2025).
 - *Recensement des tableaux ibériques dans les collections publiques françaises (1300-1870)*, RETIB, AGORHA, URL : <https://agorha.inha.fr/database/49> (visité le 02/09/2025).
 - *Répertoire des tableaux français en Allemagne (XVIIe et XVIIIe siècles)*, REPFA, URL : <https://agorha.inha.fr/database/5> (visité le 01/09/2025).
 - *Répertoire des tableaux italiens dans les collections publiques françaises (XIIIe-XIXe siècles)*, RETIF - Agorha, AGORHA, URL : <https://agorha.inha.fr/database/4> (visité le 01/09/2025).
 - *Thésaurus AGORHA Garnier*, Thésaurus et vocabulaires contrôlés de l'Institut national d'histoire de l'art, 27 septembre 2021, dernière modification le 26 août 2025, URL : <https://thesaurus.inha.fr/thesaurus/page/ark:/54721/b50e6bb5-96f2-41ea-a42f-1155135b6295> (visité le 03/09/2025).
 - *Thésaurus et vocabulaires contrôlés de l'Institut national d'histoire de l'art*, URL : <http://thesaurus.inha.fr/thesaurus/> (visité le 26/08/2025).
- INSTITUT NATIONAL D'HISTOIRE DE L'ART et BIBLIOTHÈQUE NATIONALE DE FRANCE, *La fabrique de l'art. Couleurs et matériaux de l'enluminure*, INHA, URL : <https://agorha.inha.fr/database/89> (visité le 02/09/2025).
- KERMORVANT (Christopher) et DE SAINT-OURL Édouard, *Le projet HikarIA : analyse iconographique automatisée de photographies patrimoniales*, 16 juin 2025, URL : <https://>

- www.inha.fr/agenda/le-projet-hikaria-analyse-iconographique-automatisée-de-photographies-patrimoniales/.
- KIRILLOV (Alexander), MINTUN (Eric), RAVI (Nikhila), MAO (Hanzi), ROLLAND (Chloe), GUSTAFSON (Laura), XIAO (Tete), WHITEHEAD (Spencer), BERG (Alexander C.), LO (Wan-Yen), *et al.*, *Segment Anything*, version 1, 2023, DOI : 10.48550/ARXIV.2304.02643.
- LABORDE (Pierre-Yves), *Les bases de données de l'INHA, une longue histoire*, Numrha, 18 mars 2021, URL : <https://numrha.hypotheses.org/2164> (visité le 27/07/2025).
- LI (Junnan), LI (Dongxu), XIONG (Caiming) et HOI (Steven), *BLIP : Bootstrapping Language-Image Pre-training for Unified Vision-Language Understanding and Generation*, version 2, 2022, DOI : 10.48550/ARXIV.2201.12086.
- LIBRARY OF CONGRESS, *Thesaurus for Graphic Materials I (TGM) : An Introduction*, 2024, URL : <https://guides.loc.gov/tgm-i> (visité le 27/08/2025).
- *Planning Framework used by LC Labs for planning AI experiments towards responsible implementation*, Github, 25 juill. 2025, URL : <https://github.com/LibraryOfCongress/labs-ai-framework> (visité le 01/09/2025).
- LIU (Haotian), LI (Chunyuan), WU (Qingyang) et LEE (Yong Jae), *Visual Instruction Tuning*, version 2, 2023, DOI : 10.48550/ARXIV.2304.08485.
- MARTIN (Raiza) et JOHNSON (Steven), *Google NotebookLM*, Google, 12 juill. 2023, URL : <https://blog.google/technology/ai/notebooklm-google-ai/> (visité le 02/09/2025).
- METROPOLITAN MUSEUM OF ART, *Computers and their potential applications in museums. Conférences organisées les 15, 16 et 17 avril 1968 avec le soutien d'IBM*. New York, 1968.
- MILES (Alistair) et BECHHOFER (Sean), *SKOS (Simple Knowledge Organization System Reference)*, 18 août 2009, URL : <https://www.w3.org/TR/skos-reference/> (visité le 03/09/2025).
- MINISTÈRE DE LA CULTURE, *Thésaurus iconographique, système descriptif des représentations de François Garnier*, Musées, 4 juin 2020, URL : <https://www.culture.gouv.fr/thematiques/musees/pour-les-professionnels/conserver-et-gerer-les-collections/informatiser-les-collections-d-un-musée-de-france/vocabulaires-scientifiques-du-service-des-musées-de-france/thesaurus-iconographique-systeme-descriptif-des-représentations-de-francois-garnier> (visité le 26/08/2025).
- *compar :IA*, *Le comparateur d'IA conversationnelles*, oct. 2024, URL : <https://www.comparia.beta.gouv.fr/> (visité le 31/08/2025).
- «POP», *la plate-forme ouverte du patrimoine*, Ministère de la Culture, URL : <https://www.culture.gouv.fr/fr/espace-documentation/bases-de-donnees/Fiches->

- bases - de - donnees / pop - la - plate - forme - ouverte - du - patrimoine (visité le 26/08/2025).
- *Thésaurus iconographique Garnier*, Opentheso, URL : <https://opentheso.huma-num.fr/?idt=th285> (visité le 20/08/2025).
- MISTRAL AI, *Mistral 7B*, Mistral.ai, 27 sept. 2023, URL : <https://mistral.ai/fr/news/announcing-mistral-7b> (visité le 02/09/2025).
- *Our contribution to a global environmental standard for AI*, 22 juill. 2025, URL : <https://mistral.ai/news/our-contribution-to-a-global-environmental-standard-for-ai> (visité le 31/08/2025).
 - *Medium is the new large*, URL : <https://mistral.ai/fr/news/mistral-medium-3> (visité le 02/09/2025).
- MUSÉE D'ORSAY, *Les collections*, URL : <https://www.musee-orsay.fr/fr/collections> (visité le 03/09/2025).
- MUSÉE DE REIMS, *Les oeuvres en ligne*, Portail officiel des Musées de Reims, URL : <https://musees-reims.fr/fr/musee-numerique/oeuvres-en-ligne> (visité le 03/09/2025).
- Notes de version - Hikaria*, HikarIA, URL : <https://hikaria.org/about/release-notes> (visité le 01/09/2025).
- OKAYAMA (Yassu), *The Ripa index : personifications and their attributes in five editions of the Iconologia*, Doornspijk, 1992.
- OPENAI, *GPT-4*, OpenAI, URL : <https://openai.com/fr-FR/index/gpt-4/> (visité le 02/09/2025).
- OQUAB (Maxime), DAR CET (Timothée), MOUTAKANNI (Théo), VO (Huy), SZA FRANIEC (Marc), KHALIDOV (Vasil), FERNANDEZ (Pierre), HAZIZA (Daniel), MASSA (Francisco), EL-NOUBY (Alaaeldin), et al., *DINOv2 : Learning Robust Visual Features without Supervision*, version 2, 2023, DOI : 10.48550/ARXIV.2304.07193.
- P180 (*dépeint / depicts*), Wikidata, URL : <https://www.wikidata.org/wiki/Property:P180> (visité le 03/09/2025).
- PANOFSKY (Erwin), « Zum Problem der Beschreibung und Inhaltsdeutung von Werken der bildenden Kunst », *Logos*-21 (1932), p. 103-119.
- *Essais d'iconologie : thèmes humanistes dans l'art de la Renaissance*, trad. par Claude Herbette et Bernard Teyssèdre, Nouvelle éd., Paris, 2021 (Tel, 436).
- PEREZ (Sarah), *Wikipedia's Next Big Thing : Wikidata, A Machine-Readable, User-Editable Database Funded By Google, Paul Allen And Others*, TechCrunch, 30 mars 2012, URL : <https://techcrunch.com/2012/03/30/wikipedias-next-big-thing-wikidata-a-machine-readable-user-editable-database-funded-by-google-paul-allen-and-others/> (visité le 03/09/2025).

- Toni Petersen (éd.), *Art & Architecture Thesaurus*, New York, 1990, 2e édition 1994.
- Planification de projet IA, traduction de l'Artificial Intelligence Planning Framework de la Library of Congress*, Github, 23 mai 2025, URL : <https://github.com/altomator/Planification-de-projets-IA> (visité le 01/09/2025).
- POSTHUMUS (Etienne) et BRANDHORST (Hans), *Arkyves*, 1999, URL : <https://brill.com/display/db/arko> (visité le 03/09/2025).
- RADFORD (Alec), KIM (Jong Wook), HALLACY (Chris), RAMESH (Aditya), GOH (Gabriel), AGARWAL (Sandhini), SASTRY (Girish), ASKELL (Amanda), MISHKIN (Pamela), CLARK (Jack), et al., *Learning Transferable Visual Models From Natural Language Supervision*, version 1, 2021, DOI : 10.48550/ARXIV.2103.00020.
- RAMOS (Patrick), GONTIER (Nicolas), KHAN (Selina), NAKASHIMA (Yuta) et GARCIA (Noa), *No Annotations for Object Detection in Art through Stable Diffusion*, version 2, 2024, DOI : 10.48550/ARXIV.2412.06286.
- RDF WORKING GROUP, *Resource Description Framework (RDF)*, 25 févr. 2014, URL : <https://www.w3.org/RDF/> (visité le 03/09/2025).
- RESIG (John), *About The Site*, Ukiyo-e.org, 2012, URL : <https://ukiyo-e.org/about> (visité le 04/09/2025).
- *Ukiyo-e.org : Japanese Print Search and Database*, 2012, URL : <https://ukiyo-e.org/> (visité le 04/09/2025).
- RIPA (Cesare), *Iconologia*, Rome, 1593, 1603.
- ROUIT (Huguette), DUBOULZ (Jean-Pierre) et AL., *A l'écoute de l'oeil : les collections iconographiques et les bibliothèques. Actes du colloque organisé par la Section des Bibliothèques d'Art de l'IFLA, Genève, 13-15 mars 1985*, K.G. Saur, Munich, 1989 (IFLA Publications, 47).
- SANTINI (Cristian), POSTHUMUS (Etienne), TAN (Mary Ann), BRUNS (Oleksandra), TIETZ (Tabea) et SACK (Harald), *Multimodal Search on Iconclass using Vision-Language Pre-Trained Models*, version 1, 2023, DOI : 10.48550/ARXIV.2306.16529.
- SEGUIN (Benoit), STRILOLO (Carlotta), DI LENARDO Isabella et KAPLAN (Frederic), « Visual Link Retrieval in a Database of Paintings », dans *Computer Vision – ECCV 2016 Workshops*, Cham, 2016, t. 9913, p. 753-767, DOI : 10.1007/978-3-319-46604-0_52.
- SHATFORD (Sara), « Describing a Picture : A Thousand Words are Seldom Cost Effective », *Cataloging & Classification Quarterly*, 4-4 (1984), p. 13-30.
- « Analyzing the Subject of a Picture : A Theoretical Approach », *Cataloging & Classification Quarterly*, 6-3 (1986), p. 39-62.

- SHEN (Xi), EFROS (Alexei A.) et AUBRY (Mathieu), *Discovering Visual Patterns in Art Collections with Spatially-consistent Feature Learning*, version 2, 2019, DOI : 10.48550/ARXIV.1903.02678.
- SHEN (Xi), CHAMPENOIS (Robin), GINOSAR (Shiry), PASTROLIN (Ilaria), ROUSSELOT (Morgane), BOUNOU (Oumayma), MONNIER (Tom), GIDARIS (Spyros), BOUGARD (François), RAVERDY (Pierre-Guillaume), *et al.*, « Spatially-Consistent Feature Matching and Learning for Heritage Image Analysis », *International Journal of Computer Vision*, 130–5 (mai 2022), p. 1325-1339, DOI : 10.1007/s11263-022-01576-x.
- SINGH (Amanpreet), HU (Ronghang), GOSWAMI (Vedanuj), COUAIRON (Guillaume), GALUBA (Wojciech), ROHRBACH (Marcus) et KIELA (Douwe), *FLAVA : A Foundational Language And Vision Alignment Model*, version 3, 2021, DOI : 10.48550/ARXIV.2112.04482.
- SMITS (Thomas) et WEVERS (Melvin), « A multimodal turn in Digital Humanities. Using contrastive machine learning models to explore, enrich, and analyze digital visual historical collections », *Digital Scholarship in the Humanities*, 38–3 (31 août 2023), p. 1267-1280, DOI : 10.1093/lhc/fqad008.
- SOLOGUB (Polo) et FIEBRINK (Rebecca), *Sketchy Collections : Exploring Digital Museum Collections by Drawing via CLIP*, 2024, URL : https://ualresearchonline.arts.ac.uk/id/eprint/23541/1/NeurIPS_2024_Sketchy_Collections_camera-ready.pdf (visité le 04/09/2025).
- STORK (David G.), « Computer Vision, ML, and AI in the Study of Fine Art », *Communications of the ACM*, 67–5 (mai 2024), p. 68-75, DOI : 10.1145/3633454.
- *Pixels & paintings : foundations of computer-assisted connoisseurship*, Hoboken, New Jersey, 2024.
- TEKLIA, *Arkindex, the document processing platform*, TEKLIA, URL : <https://teklia.com/our-solutions/arkindex/> (visité le 02/09/2025).
- THE WARBURG INSTITUTE, *Summary guide to the photographic collection of the Warburg Institute*, Londres, 1988.
- THUILLIER (Jacques), « Image et informatique : de l'utilisation au système », dans *A l'écoute de l'oeil : les collections iconographiques et les bibliothèques. Actes du colloque organisé par la Section des Bibliothèques d'Art de l'IFLA, Genève, 13-15 mars 1985*, Munich, 1989, p. 255-261.
- « L'informatique en histoire de l'art : où en sommes-nous ? », *Revue de l'Art*, 97–3 (1992), p. 5-10, DOI : 10.3406/rvart.1992.347996.
- « L'informatique et l'histoire de l'art », *Revue de l'Art*, 117–3 (1997), p. 5-8, DOI : 10.3406/rvart.1997.348337.

- ULTRALYTICS, *YOLOv8 / YOLOv11*, 2025, URL : <https://github.com/ultralytics/ultralytics> (visité le 31/08/2025).
- UNIVERSITÉ PARIS CITÉ, *Plateforme de Calcul Haute-Performance NOVA*, Plateformes & plateaux - Université Paris Cité, URL : <https://plateformes.u-paris.fr/plateforme-de-calcul-haute-performance-nova/> (visité le 01/09/2025).
- VAN STRATEN Roelof, « Panofsky and ICONCLASS », *Artibus et Historiae*, 7–13 (1986), p. 165-181, DOI : 10.2307/1483254.
- *An Introduction to Iconography*, trad. par Patricia de Man, Yverdon (Suisse), 1994.
- *Iconography indexing iconclass : a handbook*, Leyde, 1994.
- VAN DE WAAL Henri, *Decimal Index of the Art of the Low Countries D.I.A.L., Abridged edition of the Iconclass System*, La Haye, 1968.
- « Système de classification iconographique “ICONCLASS” et la collection de reproductions D.I.A.L. (Decimal Index of the Art of the Low Countries) », dans *Colloques Internationaux du Centre National de la Recherche Scientifique*, Paris, 24, 25, 26 mars 1969 (Bibliographie d’Histoire de l’Art), p. 187-190.
- VAN DE WAAL Henri et COUPRIE (Leendert D.), *Iconclass an iconographic classification system*, 17 t., completed and edited by L.D. Couprie with R.H. Fuchs, E. Tholen & G. Vellekoop, North-Holland, Amsterdam, 1974.
- VOLLE (Nathalie), *Michel Laclotte et le Répertoire des Tableaux Italiens en France, XIIIe-XIXe siècles (RETIF)*, AGORHA, 5 févr. 2024, URL : <https://agorha.inha.fr/detail/940> (visité le 08/08/2025).
- Wikidata, URL : <https://www.wikidata.org/> (visité le 03/09/2025).

Introduction

En 1985, le Musée d'Art et d'Histoire de Genève accueillait de nombreux documentalistes, bibliothécaires et historiens de l'art réunis autour d'un même enjeu : réfléchir à l'avenir des collections iconographiques et à leur indexation au sein des institutions patrimoniales¹. Ce séminaire, intitulé "À l'Écoute de l'œil", plaçait l'informatisation au cœur des débats, à une époque où celle-ci commençait à transformer en profondeur les bibliothèques, les musées et les archives. Si les participants accueillaient ces innovations avec prudence, ils pressentaient déjà les perspectives considérables qu'elles offraient pour la découverbarilité des collections iconographiques, grâce à l'indexation et à leur mise à disposition du public. L'historien de l'art et professeur au Collège de France Jacques Thuillier soulignait même, de manière précoce, le rôle que pourrait jouer la recherche en intelligence artificielle pour interroger un jour de vastes banques d'images en langage naturel².

Par « intelligence artificielle », on désigne aujourd'hui l'ensemble des théories et techniques permettant de concevoir des programmes informatiques capables de simuler certains aspects de l'intelligence humaine, tels que le raisonnement ou l'apprentissage. Quarante ans plus tard, l'intelligence artificielle s'invite dans tous les secteurs d'activité et s'impose comme une nouvelle révolution technologique, porteuse de perspectives comparables à celles qu'avait ouvertes l'informatisation. Depuis le début des années 2020, le déploiement de modèles de vision par ordinateur auprès du grand public a favorisé l'émergence de nombreux projets explorant l'analyse automatisée d'images. Ces modèles reposent sur des algorithmes capables de détecter, analyser, reconnaître et interpréter des éléments visuels, de manière analogue à la perception humaine.

Dans les institutions patrimoniales, ils sont de plus en plus envisagés comme des outils susceptibles d'accélérer et d'enrichir les pratiques documentaires d'indexation iconographique, contribuant ainsi à améliorer la découverbarilité des collections. Selon l'AFNOR, l'in-

¹Huguette Rouit, Jean-Pierre Dubouloz et al., *A l'écoute de l'œil : les collections iconographiques et les bibliothèques. Actes du colloque organisé par la Section des Bibliothèques d'Art de l'IFLA, Genève, 13-15 mars 1985*, K.G. Saur, Munich, 1989 (IFLA Publications, 47).

²Jacques Thuillier, « Image et informatique : de l'utilisation au système », dans *A l'écoute de l'œil : les collections iconographiques et les bibliothèques. Actes du colloque organisé par la Section des Bibliothèques d'Art de l'IFLA, Genève, 13-15 mars 1985*, Munich, 1989, p. 255-261, p. 261.

dexion est la "représentation par les éléments d'un langage documentaire ou naturel, des notions résultant de l'analyse du contenu d'un document en vue d'en faciliter la recherche"³. Dans le cas de l'indexation iconographique des œuvres d'art, il s'agit donc de recourir à des descripteurs pour décrire les scènes et les éléments représentés dans une image.

C'est dans cette dynamique qu'a vu le jour le projet intitulé textitIconographIA, porté par l'Institut National d'Histoire de l'Art (INHA) et le consortium Huma-Num PictorIA. Son objectif est d'appliquer des modèles de vision par ordinateur aux peintures du Répertoire des tableaux italiens dans les collections françaises (RETIF), afin d'en produire une indexation à l'aide du thésaurus iconographique Garnier pour faciliter les recherches sur le portail AGORHA.

Ce projet s'inscrit dans une réflexion plus ancienne visant à utiliser les outils informatiques pour analyser et décrire les représentations au sein des tableaux. Dès leur apparition, les thésaurus iconographiques avaient été pensés pour un usage informatisé. Les premières réflexions commencent sans doute dans les années 1960, au moment où Henri van de Waal, créateur du système de représentation iconographique textitIconclass, réfléchissait déjà à l'utilisation de son système par ordinateur⁴. Le Metropolitan Museum of art de New York organisait alors, avec le soutien d'IBM, une conférence intitulée *Computers and their potential applications in museums*, où la question de l'informatisation et de l'indexation des catalogues muséaux étaient au centre du débat⁵. Dix ans plus tard, les réflexions se poursuivaient à Pise, d'abord en 1968 lors de la *First international conference on automatic processing of art history data and documents*⁶ puis lors de la seconde édition en 1984⁷. La même année, le Thésaurus iconographique Garnier était publié, avec la volonté d'uniformiser la pratique documentaire en vue d'une mise en commun des bases de données muséales en France. L'année suivante, Florin Coman soutenait sa thèse intitulée "Histoire de l'art et informatique documentaire", publiée plus tard en 1988⁸, et le séminaire "À l'Écoute de l'Oeil", que nous évoquons plus avant, discutait des enjeux de l'informatique pour les collections

³ Association française de normalisation (éd.), *Documentation : recueil de normes françaises 1993*, 5e éd, Paris-La Défense, 1993, vol. 1, p. 292.

⁴ Henri van de Waal, « Système de classification iconographique "ICONCLASS" et la collection de reproductions D.I.A.L. (Decimal Index of the Art of the Low Countries) », dans *Colloques Internationaux du Centre National de la Recherche Scientifique*, Paris, 24, 25, 26 mars 1969 (Bibliographie d'Histoire de l'Art), p. 187-190.

⁵ Metropolitan Museum of Art, *Computers and their potential applications in museums. Conférences organisées les 15, 16 et 17 avril 1968 avec le soutien d'IBM*. New York, 1968.

⁶ Paola Barocchi, Fabio Bisogni, Laura Corti et al., *First International Conference on Automatic Processing of Art History Data and Documents, Pisa, Scuola Normale Superiore, 4-7 September 1978 : conference transactions*, 2 t., [s.l.], 1978.

⁷ L. Corti, Marilyn Schmitt et al., *Automatic processing of art history data and documents : [2d. international conference] Pisa, Scuola normale superiore, september 24-27, 1984*, 3 t., Pise, 1984.

⁸ Florin Coman, *L'histoire de l'art et l'informatique documentaire*, 2 t., Atelier national reproduction des thèses, Université Lille III ; Diffusion, Aux Amateurs de livres, Lille, Paris, 1988.

iconographiques⁹. En dépit de ces initiatives, les technologies numériques ont tardé à être intégrées dans les pratiques d'histoire de l'art et à être acceptées par la communauté des chercheurs, comme le soulève Jacques Thuillier en 1992 : “un quart de siècle d'informatique n'a rien changé aux habitudes de l'histoire de l'art, tant internationale que française [...] l'échec n'est pas à mettre au compte de l'informatique, mais des historiens de l'art, qui n'ont pas su tirer parti d'elle.”¹⁰ Cinq ans plus tard, peu après l'apparition d'internet en France, le même auteur déplore la latence que prend la mise en ligne des bases de données liées au patrimoine, comme Joconde et Mérimée.¹¹ Les deux décennies suivantes voient l'accélération de la mise en ligne des contenus patrimoniaux, et la multiplication des projets numériques. Cependant, l'intelligence artificielle, telle qu'on la conçoit aujourd'hui, n'apparaît en histoire de l'art qu'au cours des années 2010, et surtout au début des années 2020. Dans le domaine de l'analyse des peintures, les projets et publications se sont multipliés ces dernières années. Les expérimentations menées par Benoît Seguin dans le cadre du projet *Replica*, en collaboration avec la fondation Cini à Venise et dont les premiers résultats ont été publiés en 2016¹², ainsi que des initiatives comme le projet international *Augmented Artwork Analysis* (AAA)¹³, fondé en 2021, mobilisent les outils de vision par ordinateur pour approfondir l'étude des peintures. Parallèlement, les nombreuses recherches conduites par David G. Stork sur l'analyse par ordinateur d'œuvres picturales ont abouti à une synthèse de référence sur le sujet¹⁴, soulignant le potentiel et les défis de ces approches dans le champ des études artistiques. Les revues spécialisées et les rencontres académiques ont contribué à alimenter la réflexion sur le sujet, comme en témoigne le numéro 87 de la revue *Histoire de l'art* paru en 2021, consacré aux Humanités numériques et incluant un échange modéré par Mathieu Aubry avec Lisandra Costiner et Stuart James sur les implications de l'intelligence artificielle dans la discipline¹⁵. L'indexation iconographique occupe une place croissante dans les réflexions sur l'usage de ces outils numériques. De nombreux projets visent à tirer parti de ces pratiques documentaires pour améliorer la gestion et la découvrabilité des collections. En 2019, Hans

⁹H. Rouit, J.P. Dubouloz et al., *À l'écoute de l'oeil...*

¹⁰J. Thuillier, « L'informatique en histoire de l'art : où en sommes-nous ? », *Revue de l'Art*, 97–3 (1992), p. 5-10, DOI : 10.3406/rvart.1992.347996, p. 5.

¹¹Id., « L'informatique et l'histoire de l'art », *Revue de l'Art*, 117–3 (1997), p. 5-8, DOI : 10.3406/rvart.1997.348337, p. 8.

¹²Benoit Seguin, Carlotta Striolo, Isabella di Lenardo et Frederic Kaplan, « Visual Link Retrieval in a Database of Paintings », dans *Computer Vision – ECCV 2016 Workshops*, Cham, 2016, t. 9913, p. 753-767, DOI : 10.1007/978-3-319-46604-0_52 ; I. di Lenardo, Benoît Séguin et Frédéric Kaplan, « Visual Patterns Discovery in Large Databases of Paintings », dans *Digital Humanities 2016 : Conference Abstracts*, Cracovie, 2016, p. 169-172, URL : <https://dh2016.adho.org/abstracts/348> (visité le 05/08/2025).

¹³*Augmented Artwork Analysis* (AAA), URL : <https://icar.cnrs.fr/aaa/> (visité le 25/08/2025).

¹⁴David G. Stork, *Pixels & paintings : foundations of computer-assisted connoisseurship*, Hoboken, New Jersey, 2024.

¹⁵Mathieu Aubry, Lisandra Costiner et Stuart James, « Artificial Intelligence and Art History : A Necessary Debate ? », *Histoire de l'art*, 87–1 (2021), p. 55-60, DOI : 10.3406/hista.2021.3936.

Brandhorst proposait une analyse des implications conjointes d'Iconclass et de l'intelligence artificielle¹⁶, tandis que des institutions telles que le Musée national des Arts Asiatiques – Guimet¹⁷ ou le Musée des Arts décoratifs¹⁸ ont développé des projets visant à enrichir la connaissance de leurs fonds grâce aux technologies de vision par ordinateur.

Le projet IconographIA s'inscrit dans ces réflexions en soulevant plusieurs questions liées à l'intégration des outils d'intelligence artificielle dans la pratique de l'indexation iconographique des œuvres picturales. Comment ces technologies peuvent-elles enrichir la connaissance et l'étude de ces collections ? Quels défis méthodologiques, sémantiques et techniques soulève l'intégration des systèmes d'apprentissage profond dans la description des représentations ? Quel rôle jouent les vocabulaires contrôlés dans ce processus, et comment assurer l'interopérabilité entre les algorithmes de vision par ordinateur et ces référentiels normés comme le thésaurus Garnier ? Dans un premier temps, il convient d'examiner les pratiques d'indexation et leurs enjeux, les vocabulaires employés ainsi que leur intégration dans les projets numériques et les bases de données. Cette analyse permet ensuite de réfléchir à la possibilité d'automatiser ce processus, en considérant les aspects techniques et méthodologiques et en identifiant les méthodes les plus pertinentes. Les avantages et limites d'une telle automatisation seront également discutés. Enfin, la réflexion s'élargit aux alternatives à l'automatisation, en explorant d'autres approches de vision par ordinateur susceptibles de faciliter le travail de l'indexeur, tout en permettant de nouvelles perspectives d'exploration des corpus d'images.

¹⁶Hans Brandhorst, *A word is worth a thousand pictures - Why the use of Iconclass will make Artificial Intelligence smarter*, 2019, URL : https://iconclass.org/testset/ICONCLASS_and_AI.pdf (visité le 01/08/2025).

¹⁷HikarIA, URL : <https://hikaria.org/> (visité le 18/08/2025).

¹⁸Emmanuelle Bermès et Marion Charpier, « Repenser les collections patrimoniales par le prisme de l'IA 2025 », dans *Conférence nationale sur les applications de l'Intelligence artificielle*, Dijon, France, 2025, URL : <https://hal.science/hal-05138697>.

Première partie

L'indexation iconographique des œuvres patrimoniales à l'ère du numérique

Chapitre 1

Principes et enjeux de l'indexation iconographique en histoire de l'art

I. L'indexation iconographique en histoire de l'art

Bien que la pratique de l'indexation remonte à plusieurs siècles, ce n'est qu'à partir du milieu du XX^e siècle qu'elle a fait l'objet d'une véritable théorisation, notamment pour la description du contenu intellectuel des livres et des documents conservés dans les bibliothèques¹. Lorsqu'elle s'applique aux objets muséaux, cette méthode reprend les principes développés pour les textes, tout en les enrichissant par les apports de l'archéologie et de l'histoire de l'art, indispensables à leur compréhension et à leur interprétation.

Pour définir et expliciter la pratique de l'indexation dans la pratique documentaire du catalogage, l'Association française de normalisation (AFNOR) la présente ainsi : l'indexation est « l'opération qui consiste à décrire et à caractériser un document à l'aide de représentations des concepts contenus dans ce document, c'est-à-dire à transcrire en langage documentaire les concepts après les avoir extraits du document par une analyse. La transcription en langage documentaire se fait grâce à des outils d'indexation tels que thésaurus, classifications, etc. [...] La finalité de l'indexation est de permettre une recherche efficace des informations contenues dans un fonds de documents et d'indiquer rapidement, sous une forme concise, la teneur d'un document² ». L'indexation iconographique, en histoire de l'art, constitue une application spécifique de ce principe : elle consiste à décrire le contenu des images à l'aide de descripteurs afin de faciliter l'accès aux œuvres, sans cependant dispenser sa consultation³.

¹Consulter à ce propos : Muriel Amar, *Les fondements théoriques de l'indexation, une approche linguistique*, Paris, 2000.

²*Documentation...*, norme NF Z47-102, vol. 1, p. 286.

³« Le rôle du système de recherche documentaire est de faciliter l'accès au document et non de dispenser sa consultation » François Garnier et Ministère de la Culture, *Thésaurus iconographique, système descriptif*

Avec le développement des bases de données, ces pratiques d'indexation reposent désormais presque systématiquement sur des vocabulaires contrôlés pour décrire les documents. Pour les œuvres d'art, où peuvent être présentés des objets, des personnages, mais aussi des concepts ou des scènes souvent codifiées, les thésaurus jouent un rôle central pour décrire ces éléments de manière normalisée. Ainsi, les méthodes et manuels d'indexation sont fréquemment associés aux thésaurus qu'ils accompagnent, afin de garantir une utilisation la plus uniforme possible du vocabulaire et d'assurer l'efficacité des recherches qui en découlent. Les principaux systèmes d'indexation iconographique, tels qu'*Iconclass* et le thésaurus Garnier, sont chacun accompagnés d'études et de guides destinés à encadrer leur utilisation, fournissant également des réflexions sur cette pratique documentaire⁴.

Cependant, les recommandations concernant l'usage des thésaurus ne fournissent pas une méthode claire et rigoureuse pour l'indexation des images, et ce pour plusieurs raisons. La première est le manque d'études consacrées à cette pratique en histoire de l'art. Certaines études comme celles de Sarah Shatford prennent en considération toutes les images et restent trop génériques pour des œuvres picturales⁵. Dans son manuel sur l'usage du système iconographique *Iconclass* publié en 1994, l'historien d'art Roelof van Straten remarque que la communauté scientifique s'est rarement penchée sur des questions pourtant essentielles liées à l'indexation, telles que la nature même de la tâche de l'indexeur ou la profondeur à laquelle l'indexation doit être effectuée.⁶ Une autre raison est celle de la grande variété des images à indexer présente dans les institutions patrimoniales ou les collections, rendant difficile l'écriture d'une méthode efficace. François Garnier, dans l'introduction de son thésaurus iconographique utilisé par les musées français, indique alors « qu'aucune grille d'analyse ne s'impose comme une nécessité »⁷.

S'il n'existe pas de consensus sur la manière de rédiger une indexation « parfaite » pour décrire une image, ces auteurs soulignent que l'indexeur doit avant tout se laisser orienter par la finalité de sa tâche : permettre un accès efficace aux images au sein d'une base de données et faciliter les recherches, présentes comme futures, des analystes d'images. Ils proposent à

des représentations, Paris, 1984, p. 15.

⁴Nous décrirons plus en détail l'histoire, l'utilisation et les particularités de ces thésaurus dans le chapitre 2.

⁵Sara Shatford, « Describing a Picture : A Thousand Words are Seldom Cost Effective », *Cataloging & Classification Quarterly*, 4–4 (1984), p. 13-30 ; Id., « Analyzing the Subject of a Picture : A Theoretical Approach », *Cataloging & Classification Quarterly*, 6–3 (1986), p. 39-62.

⁶Roelof van Straten, *Iconography indexing iconclass : a handbook*, Leyde, 1994, p. 37 : “It is astonishing that even the most basic questions regarding iconographic indexing, such as “What is the real task of the iconographic indexer ?” or “How far ‘in depth’ should we index ?” do not seem to have been discussed among art historians.”

⁷F. Garnier et Ministère de la Culture, *Thésaurus iconographique, système descriptif des représentations...*, p. 25.

cet effet une liste d’éléments essentiels à prendre en compte pour décrire une représentation⁸. Cette description doit d’abord rendre compte de la signification principale de l’image et de son contenu, en identifiant la scène, les personnages et les lieux représentés. Elle doit ensuite être enrichie par le relevé des motifs secondaires, liés ou non à l’action principale. Une attention particulière doit être portée aux éléments inhabituels, surtout lorsqu’ils présentent un intérêt historique, ainsi qu’aux détails que l’image met en avant. L’objectif est surtout de restituer autant de concepts du contenu visuel que possible, même s’il existe toujours une part inévitable de subjectivité dans le travail d’indexation. Sara Shatford insiste également sur le besoin d’être concis pour éviter de fausser les recherches dans les bases de données et d’abord privilégier les idées les plus importantes, en fonction du public visé⁹. La qualité de l’indexation dépend également de plusieurs paramètres : la qualité du système d’indexation utilisée et sa compréhension par l’indexeur, l’expérience et les connaissances de ce dernier, et la lisibilité de l’image¹⁰.

II. Sujets, genre, motifs et éléments

L’usage du vocabulaire générique que nous emploierons tout au long de ce mémoire pour décrire les images ne fait pas l’objet d’un accord unanime chez les historiens de l’art¹¹. Les termes « motifs », « thèmes », « sujet », « genre » et « éléments » ne sont pas toujours employés dans le même sens en fonction des publications. C’est pourquoi nous exposons ici la manière dont nous utiliserons ces termes.

Le terme « sujet » désigne le thème représenté dans une image ou une œuvre d’art, c’est-à-dire l’objet principal de la représentation. Il peut se référer au sujet général, correspondant à la catégorie de la représentation (portrait, paysage, scène religieuse ou mythologique), ou à une scène spécifique, c’est-à-dire une situation particulière représentée. Le sujet sert souvent à formuler le titre d’une œuvre lorsqu’aucun titre n’a été donné par l’artiste ou forgé par l’historiographie. Dans ce cas, nous préférons utiliser le terme « scène » pour plus de précision.

Dans l’histoire de la peinture, le sujet général, également appelé « genre », relève d’une

⁸R. van Straten, *Iconography indexing iconclass...*, p. 57-58, F. Garnier et Ministère de la Culture, *Thésaurus iconographique, système descriptif des représentations...*, p. 25

⁹S. Shatford, « Describing a Picture : A Thousand Words are Seldom Cost Effective »...

¹⁰F. Garnier et Ministère de la Culture, *Thésaurus iconographique, système descriptif des représentations...*, p. 24, R. van Straten, *Iconography indexing iconclass...*, p. 37

¹¹Nous avons consulté à ce propos différents ouvrages généralistes qui abordent le vocabulaire : Id., *An Introduction to Iconography*, trad. par Patricia de Man, Yverdon (Suisse), 1994, Xavier Barral i Altet et al., *Dictionnaire critique d’iconographie occidentale*, Rennes, 2003, Sérgolène Bergeon Langle et Pierre Curie, *Peinture & dessin : vocabulaire typologique et technique*, 2 t., Paris, 2009 (Principes d’analyse scientifique), Laurent Gervreau, *Voir, comprendre, analyser les images* : 2020 (Repères), DOI : 10.3917/dec.gerv.2020.01.

classification théorisée dès le XVII^e siècle. Dans son introduction aux conférences de l'Académie royale de peinture en 1667, André Félibien définit plusieurs genres iconographiques selon une hiérarchie¹² : la peinture d'histoire, qui occupe le premier rang, regroupe les œuvres représentant des événements historiques ou illustrant des textes, tels que la Bible ou la mythologie grecque et romaine. Le portrait constitue le deuxième genre pictural, suivi de la scène de genre, qui représente des scènes anecdotiques ou familières, souvent inspirées du quotidien. Viennent ensuite les genres qui mettent moins l'accent sur la figure humaine : le paysage, puis la nature morte, centrée sur la représentation d'éléments inanimés. Cette classification a profondément marqué l'histoire picturale, en particulier à l'Époque moderne. Si elle a désormais perdu son caractère hiérarchique, elle reste prégnante aujourd'hui dans les thésaurus comme les dictionnaires, avec quelques adaptations pour refléter la diversité de certaines catégories, en particulier la peinture d'histoire, et pour intégrer des genres apparus ultérieurement, comme la peinture abstraite.

Le « motif » désigne quant à lui un élément visuel particulier, un détail significatif ou un groupe d'éléments dans une œuvre, qui peut être isolé ou répété, et parfois associé à d'autres. Il peut être figuratif ou décoratif, notamment dans le cas de motifs ornementaux.

Enfin, un « élément » est la plus petite unité significative identifiable dans une image. Il peut s'agir d'une figure, d'un détail, d'un objet spécifique, d'un attribut iconographique ou d'une partie d'un motif paysager ou architectural.

Pour éclaircir ces notions, la description sommaire d'un tableau peut servir d'illustration. Prenons l'*Adoration des bergers* peinte à tempéra sur bois par le peintre florentin Taddeo Gaddi (vers 1300-1366) vers 1330 et conservé au Musée des Beaux-Arts de Dijon (inv. 1470) (figure A.1). Elle relève d'un *sujet* biblique, puisque la *scène* représentée provient du Nouveau Testament, plus particulièrement de l'Évangile selon saint Luc (chapitre 2, versets 8-20). Dans l'image, plusieurs *motifs* peuvent être identifiés, comme les bergers, la Vierge tenant l'Enfant Jésus dans ses bras ou encore saint Joseph se reposant en bas à droite de la composition. Le bâton du berger au centre du panneau, l'auréole de la vierge ou la fenêtre visible sur le mur du bâtiment peuvent être désignés comme des *éléments*.

III. La lecture de l'image : Iconographie et Iconologie

L'analyse des images en histoire de l'art, en particulier par l'identification de leurs thèmes et des scènes représentées, constitue un champ d'étude spécifique, appelé iconographie. La définition et la méthode proposées par l'historien de l'art Erwin Panofsky (1892-1968)

¹²André Félibien, *Conferences de l'Academie royale de peinture et de sculpture. Pendant l'année 1667*, 1667.

dès 1931 sont régulièrement citées pour définir cette discipline¹³. Il définit l'iconographie comme « cette branche de l'histoire de l'art qui se rapporte au sujet ou à la signification des œuvres d'art, par opposition à leur forme »¹⁴. Plus précisément, le terme iconographie revêt deux sens complémentaires : le premier désigne l'ensemble des représentations d'un même sujet ou thème dans les arts visuels ; le second correspond à la discipline d'histoire de l'art elle-même, centrée sur l'identification, la description et l'interprétation du contenu des images, en examinant les sujets représentés, les compositions, et tous les éléments distincts du style. Ce dernier terme, que Panofsky désigne comme “la forme”, recouvre les qualités plastiques de la représentation, notamment les lignes, les couleurs, la composition, les proportions, les volumes ou encore la lumière. L'iconographie s'intéresse ainsi essentiellement au contenu narratif ou symbolique des œuvres, plutôt qu'à leur aspect formel.

Le terme d'iconologie est généralement employé en complément de l'iconographie. Les historiens font remonter cette notion, dans son acceptation proche de son usage contemporain, à Cesare Ripa, érudit italien, qui publie en 1593 un manuel intitulé *Iconologia*¹⁵, largement diffusé au XVII siècle. Cet ouvrage servait de dictionnaire destiné à expliquer la manière de représenter et reconnaître les personnifications et allégories, en précisant leurs attributs¹⁶. Comme le souligne Xavier Barral i Altet, le sens du mot iconologie comprenait alors, au moment de la parution de cet ouvrage, ce que nous désignons aujourd'hui par « iconographie »¹⁷. Au XX siècle, une nouvelle approche de l'iconologie est développée à Hambourg en Allemagne autour de l'historien d'art Aby Warburg (1866-1929), puis surtout par Erwin Panofsky, qui entend l'iconologie comme une approche destinée à interpréter le sens profond des images et à dégager leur signification symbolique, en lien avec le contexte culturel, philosophique, religieux ou social de leur création.

Ainsi, Panofsky propose une méthode en trois niveaux pour interpréter le contenu d'une image. Bien que cette démarche ait suscité certaines critiques et ajustements au cours du temps, elle reste aujourd'hui largement respectée pour avoir théorisé la pratique iconographique et iconologique¹⁸. Le premier niveau de cette méthode, la description pré-

¹³Erwin Panofsky la définit pour la première fois en 1931 lors d'une conférence à Kiel auprès de la société d'études kantiennes : Erwin Panofsky, « Zum Problem der Beschreibung und Inhaltsdeutung von Werken der bildenden Kunst », *Logos*-21 (1932), p. 103-119. Il affine ensuite cette définition dans l'introduction de ses *Studies in Iconology* publiés en 1939 puis dans *Meaning in the visual arts* publié en 1955.

¹⁴Id., *Essais d'iconologie : thèmes humanistes dans l'art de la Renaissance*, trad. par Claude Herbette et Bernard Teyssèdre, Nouvelle éd., Paris, 2021 (Tel, 436), p. 19.

¹⁵Cesare Ripa, *Iconologia*, Rome, 1593, 1603.

¹⁶Voir à ce sujet Yassu Okayama, *The Ripa index : personifications and their attributes in five editions of the Iconologia*, Doornspijk, 1992

¹⁷X. Barral i Altet, « Introduction », dans *Dictionnaire critique d'iconographie occidentale*, Rennes, 2003, p. 866, p. 24.

¹⁸R. van Straten, *An Introduction to Iconography...*, en particulier le premier chapitre « What is Iconography ? », p. 3-25.

iconographique, consiste à identifier les éléments et les motifs visibles dans la représentation : figures humaines, animaux, objets et leurs relations mutuelles, y compris les événements représentés. Les qualités expressives, telles que la joie, la tristesse ou d'autres intentions, sont également relevées à cette étape. La connaissance de l'histoire du style permet de comprendre comment certains éléments sont représentés dans un contexte particulier : par exemple, un dauphin dans la tradition médiévale est figuré très différemment de la réalité. Le deuxième niveau, l'analyse iconographique, consiste à mettre en relation les motifs et compositions avec des sujets ou des thèmes précis, en identifiant les images, les scènes et les allégories représentées. Le troisième niveau, l'interprétation iconologique, vise à dégager le sens profond de l'image, souvent lié aux choix de l'artiste ou du commanditaire dans la représentation des motifs. Voici un exemple pour illustrer cette démarche : une description pré-iconographique pourrait consister à observer un vieillard levant un couteau et tenant un jeune garçon d'une main, avec un bœuf à ses côtés et un ange descendant du ciel. L'analyse iconographique identifierait Abraham sacrifiant son fils, comme raconté au chapitre 22 de la Genèse. L'interprétation iconologique, selon une lecture typologique, pourrait mettre en évidence une allusion à Dieu sacrifiant son fils, préfigurant la Passion du Christ dans le Nouveau Testament.

Comme le souligne Roelof van Straten, pour l'indexation iconographique, ce sont surtout les deux premières étapes qui importent, c'est-à-dire l'identification du sujet, des scènes, des motifs et des éléments représentés dans l'œuvre¹⁹. Les conclusions issues d'une interprétation iconologique peuvent également être intégrées à l'indexation, à condition que les concepts soient précis et validés par le contexte de création ou par des sources documentaires. C'est le cas, par exemple, des *Memento mori* présents dans les natures mortes, des représentations de vanités ou des paysages de l'Arcadie. Cela vaut aussi pour les préfigurations dans les associations typologiques entre l'Ancien et le Nouveau Testament.

¹⁹Id., « Panofsky and ICONCLASS », *Artibus et Historiae*, 7–13 (1986), p. 165–181, DOI : 10.2307/1483254.

Chapitre 2

Les instruments de l'indexation : vocabulaires contrôlés et systèmes d'indexation

Comme nous l'évoquions précédemment, les méthodes d'indexation sont intimement liées aux vocabulaires qu'elles utilisent. En histoire de l'art, la plupart des vocabulaires ont vu leur naissance dans la deuxième moitié du XX^e siècle, et surtout dans les années 1980, avec le développement des systèmes informatiques dans les institutions patrimoniales et l'amélioration des pratiques de catalogage. Des vocabulaires d'indexation iconographique génériques ont alors vu le jour, destinés à décrire l'ensemble des représentations artistiques, tandis que des vocabulaires spécialisés sont employés pour un domaine d'étude précis ou un projet particulier. La question de l'alignement de ces vocabulaires est un enjeu crucial pour l'interopérabilité des bases de données et la facilitation de la recherche iconographique.

I. Vocabulaires génériques et spécifiques

Différents types de vocabulaires peuvent être mobilisés pour décrire les représentations présentes au sein des objets patrimoniaux, selon l'ampleur de l'analyse envisagée et le degré de spécialisation des approches adoptées. Certains sont spécifiquement liés à des domaines d'étude particuliers, tandis que d'autres, plus généralistes, visent à proposer une description uniforme des représentations, quel que soit leur contexte de création. Ainsi, les vocabulaires spécialisés sont souvent associés à des projets de recherche ou à des disciplines précises. Pour les objets archéologiques, par exemple, le Lexicon Iconographicum Mythologiae Classicae (LIMC) constitue une référence. Publié en huit tomes doubles entre 1981 et 2009, il est consacré aux représentations figurées issues des mythologies grecque, étrusque et romaine,

appliquées à divers objets antiques (céramique, sculpture, orfèvrerie, glyptique, peinture, mosaïque, etc.) FONDATION INTERNATIONALE POUR LE LIMC, *Lexicon Iconographicum Mythologiae Classicae*, IX t., Zürich, 1981 ; Fondation Internationale pour le LIMC (éd.), *Lexicon Iconographicum Mythologiae Classicae (LIMC) : Supplementum 2009*, Düsseldorf, 2009.

Dans le domaine de l’art médiéval, plusieurs répertoires et dictionnaires iconographiques ont vu le jour, parmi lesquels l’Index of Medieval Art de l’université de Princeton. Fondé en 1917 par le professeur Charles Rufus Morey, il a d’abord été conçu au format papier, rassemblant un index des iconographies chrétiennes, puis juives, islamiques et non religieuses, accompagné de photographies. Informatisé en 1991, il est accessible en ligne depuis 2017¹. Plus récemment, des projets tels que l’Ontologie du christianisme médiéval en images, mise en ligne en 2023 par l’Institut national d’histoire de l’art, proposent également un lexique contrôlé pour les représentations figurées dans le contexte médiéval². Pour les œuvres d’art picturales, il existe certes des dictionnaires spécialisés, mais ce sont surtout des systèmes iconographiques généralistes qui sont employés pour leur description. Certains ont été développés dans le cadre d’institutions, comme le système d’indexation par mots-clés du Warburg Institute³, ou encore le Thesaurus for Graphic Materials, publié par la Library of Congress et principalement utilisé pour les estampes et les photographies⁴.

Aujourd’hui, les vocabulaires iconographiques les plus utilisés dans les institutions, notamment pour la description des œuvres d’époque moderne et contemporaine, sont *ICON-CLASS*, les thésaurus développés par le Getty Research Institute, et, en France, le thésaurus Garnier. Parallèlement, Wikidata prend une place croissante dans la description des représentations, grâce à l’implication grandissante de contributeurs et d’institutions qui y participent activement.

Le système iconographique Iconclass a été conçu à l’initiative de l’historien de l’art néerlandais Henri van de Waal. Publié une première fois en 1968 sous l’acronyme D.I.A.L. (textitDecimal Index of the Art of the Low Countries)⁵, il prend rapidement le nom d’Iconclass, contraction d’ICONographic CLASSification (parfois rendu ICONographic CLASsification System). Initialement prévu pour paraître en volumes imprimés, il fait l’objet d’une

¹ *History of the Index*, The Index of Medieval Art, URL : <https://ima.princeton.edu/history/> (visité le 27/08/2025).

² Mathieu Beaud, Isabelle Marchesin et Sébastien Biay, *Ontologie du christianisme médiéval en images*, 2023, URL : <https://omci.inha.fr/> (visité le 27/08/2025).

³ The Warburg Institute, *Summary guide to the photographic collection of the Warburg Institute*, Londres, 1988.

⁴ Aujourd’hui disponible en ligne : Library of Congress, *Thesaurus for Graphic Materials I (TGM) : An Introduction*, 2024, URL : <https://guides.loc.gov/tgm-i> (visité le 27/08/2025)

⁵ H. van de Waal, *Decimal Index of the Art of the Low Countries D.I.A.L., Abridged edition of the Iconclass System*, La Haye, 1968.

édition augmentée et révisée entre 1973 et 1985, comprenant sept volumes consacrés au système, sept volumes de bibliographie et trois volumes d’index⁶. Dès le début des années 1970, Van de Waal anticipe un usage informatisé du dispositif, ce qui conduit à une collaboration avec le Centraal Rekeninstituut de Leyde et au portage d’Iconclass sur un ordinateur IBM mainframe en 1978⁷. Après des versions diffusées sur CD-ROM dans les années 1990, le système est finalement mis en ligne entre 1999 et 2000⁸. Le principe d’Iconclass repose sur une classification hiérarchique des représentations, organisée en dix classes principales, chacune subdivisée en catégories plus fines, couvrant notamment les domaines de la Nature, de L’Être Humain, de l’Histoire, de la Bible ou des Mythologies. Chaque concept est identifié par un code alphanumérique, dont la longueur indique le degré de précision : plus le code est détaillé, plus la description est fine.

Parallèlement aux travaux de Van de Waal, des initiatives de normalisation des vocabulaires émergent aux États-Unis, portées par le Getty Trust, puis par le Getty Research Institute (GRI), fondé en 1985. Dès la fin des années 1970, est engagé le développement de l’Art & Architecture Thesaurus (AAT), destiné à répondre aux besoins des bibliothèques et des revues spécialisées en matière de catalogage et d’indexation⁹. Le projet est initié par les directeurs de bibliothèques et des historiens de l’art Toni Petersen, Dora Crouch et Pat Molholt, avec le soutien financier du Getty Trust. Structuré au cours des années 1980 par le Getty Vocabulary Program, l’AAT paraît en deux éditions imprimées, en 1990 et 1994¹⁰. De portée générale, l’AAT rassemble des termes multilingues pour désigner des agents, des types d’œuvres, des rôles, des matériaux, des styles, des cultures et des techniques de création. Si certains de ces termes concernent aussi les sujets iconographiques, c’est surtout la Getty Iconography Authority (IA), intégrée au programme des vocabulaires du GRI, qui s’attache directement aux représentations iconographiques. Plus spécialisé, ce vocabulaire recense des noms propres, des personnages religieux ou fictifs, des événements historiques, ainsi que des titres d’œuvres littéraires ou de spectacles, en intégrant des relations et des repères chronologiques. L’accent est particulièrement mis sur les cultures non occidentales, notamment asiatiques et américaines, tout en établissant des références croisées avec d’autres ressources,

⁶H. van de Waal et Leendert D. Couplie, *Iconclass an iconographic classification system*, 17 t., completed and edited by L.D. Couplie with R.H. Fuchs, E. Tholen & G. Vellekoop, North-Holland, Amsterdam, 1974.

⁷H. Brandhorst et Etienne Posthumus, « ICONCLASS, A key to collaboration in the digital humanities », dans *The Routledge Companion to Medieval Iconography*, dir. Colum Hourihane, Routledge, Londres et New York, 2017, p. 201-218, p. 202.

⁸Id., *Survey of versions of the Iconclass system*, ICONCLASS, URL : <https://iconclass.org/help/about> (visité le 03/09/2025).

⁹Getty Research Institute, *About the AAT*, révisé le 26 juillet 2024, URL : <https://www.getty.edu/research/tools/vocabularies/aat/about.html> (visité le 03/09/2025).

¹⁰Toni Petersen (éd.), *Art & Architecture Thesaurus*, New York, 1990, 2e édition 1994.

telles qu’Iconclass pour le contexte occidental¹¹. Afin de faciliter les mises à jour régulières, une version numérique de l’AAT est mise en ligne en 1997, accompagnée d’une interface de recherche. Depuis, l’ensemble des vocabulaires Getty, dont l’AAT et l’IA, est accessible via le site du Getty Research Institute¹².

En France, s’est développé à la même époque le *Thésaurus iconographique, système descriptif des représentations de François Garnier*, plus couramment appelé thésaurus Garnier. Initié en 1976 à la demande du Ministère de la Culture, il répondait aux besoins de l’Inventaire général des monuments et des richesses artistiques de la France ainsi qu’à ceux de la Direction des musées de France¹³. Son objectif était de fournir une liste de descripteurs destinés à l’indexation des documents iconographiques et à l’alimentation des bases de données créées par ces services. Conçu avec une vocation universelle, ce thésaurus visait à offrir un langage normalisé permettant de décrire toute représentation, indépendamment de sa date, de sa nature ou de son support. Le thésaurus Garnier demeure aujourd’hui « l’outil retenu par le ministère de la Culture pour l’analyse des représentations figurées dans les bases de données patrimoniales » en France¹⁴. La première version a été publiée sous forme d’ouvrage aux éditions Léopard d’or. Richement illustrée de reproductions de gravures sur bois accompagnées d’exemples d’indexation conforme aux règles du thésaurus, offrant ainsi aux utilisateurs des cas pratiques d’application¹⁵. Cette version initiale se composait de quatre rubriques : une liste hiérarchisée de termes pour décrire une représentation, une liste non exhaustive de noms propres ou de termes particuliers (par exemple des lieux), les sources textuelles dont les représentations sont issues et des termes permettant de dater le contenu représenté (et non l’exécution de l’œuvre). De nombreux musées français utilisent encore cette version imprimée, désormais mise à disposition en ligne grâce à une numérisation complète réalisée par le ministère de la Culture¹⁶. Celui-ci propose également une version hébergée sur Opentheso, structurée en 22 catégories et comprenant 32 895 concepts¹⁷. L’Institut national

¹¹Id., *About CONA and IA*, révisé le 31 juillet 2024, URL : https://www.getty.edu/research/tools/vocabularies/cona/about.html#history_ia (visité le 03/09/2025).

¹²Id., *Getty Vocabularies*, URL : <https://www.getty.edu/research/tools/vocabularies/> (visité le 03/09/2025).

¹³F. Garnier et Ministère de la Culture, *Thésaurus iconographique, système descriptif des représentations...*, p. 11.

¹⁴Ministère de la Culture, *Thésaurus iconographique, système descriptif des représentations de François Garnier*, Musées, 4 juin 2020, URL : <https://www.culture.gouv.fr/thematiques/musees/pour-les-professionnels/conserver-et-gerer-les-collections/informatiser-les-collections-d-un-musee-de-france/vocabulaires-scientifiques-du-service-des-musees-de-france/thesaurus-iconographique-systeme-descriptif-des-representations-de-francois-garnier> (visité le 26/08/2025)

¹⁵F. Garnier et Ministère de la Culture, *Thésaurus iconographique, système descriptif des représentations...*

¹⁶Ministère de la Culture, *Thésaurus iconographique, système descriptif des représentations de François Garnier...*

¹⁷Id., *Thésaurus iconographique Garnier*, Opentheso, URL : [https://openthoso.huma-num.fr/?idt=](https://opentheso.huma-num.fr/?idt=)

d’histoire de l’art (INHA) diffuse quant à lui une version distincte, totalisant 16 362 concepts répartis en 24 catégories¹⁸. Enfin, le thésaurus iconographique utilisé par Mandragore, la base de données recensant les enluminures des manuscrits de la Bibliothèque nationale de France, correspond à une version dérivée et simplifiée du thésaurus Garnier¹⁹.

Enfin, Wikidata constitue aujourd’hui une alternative pour décrire les représentations au moyen de ses propriétés²⁰. Créée en 2012 par la Wikimedia Foundation, cette base de connaissances libre et collaborative est accessible en ligne²¹. Elle fonctionne comme un référentiel sémantique où chaque entité est décrite par un élément (item), identifié par un identifiant unique. Wikidata est utilisée pour structurer et centraliser les données des projets Wikimedia (comme Wikimedia Commons ou Wikipedia), mais aussi par des institutions patrimoniales et scientifiques afin d’enrichir ou d’interconnecter leurs bases de données. Chaque élément y est défini par un label, c’est-dire le nom principal, complété d’une description et éventuellement d’alias correspondant à des variantes ou synonymes. Tous ces champs sont multilingues. Les images et les éléments peuvent être reliés entre eux grâce aux propriétés de Wikidata. Par exemple, dans le cas de la description iconographique, la propriété P180 (dépeint / depicts) permet d’associer une image à l’élément qu’elle représente²².

II. Systèmes descriptifs et systèmes d’indexation

En 1994, Roelof van Straten faisait la distinction entre des “systèmes descriptifs” (*descriptive systems*) qui utilisent uniquement du langage naturel et les “systèmes d’indexation” (*indexing systems*) dont le but n’est pas de décrire les images mais de les indexer en leur attribuant un ou plusieurs codes ou mot-clés²³. Cette distinction permet de comprendre l’utilisation des différents systèmes et vocabulaires utilisés dans la description iconographique, surtout dans le cadre de projets numériques. Avec l’apparition de l’informatisation, on ob-

¹⁸th285 (visité le 20/08/2025).

¹⁹Institut national d’histoire de l’art, *Thésaurus AGORHA Garnier*, Thésaurus et vocabulaires contrôlés de l’Institut national d’histoire de l’art, 27 septembre 2021, dernière modification le 26 août 2025, URL : <https://thesaurus.inha.fr/thesaurus/page/ark:/54721/b50e6bb5-96f2-41ea-a42f-1155135b6295> (visité le 03/09/2025).

²⁰Bibliothèque nationale de France, *Thésaurus iconographique de Mandragore*, 2024, URL : <https://api.bnf.fr/fr/thesaurus-mandragore> (visité le 24/08/2025).

²¹Wikidata, URL : <https://www.wikidata.org/> (visité le 03/09/2025).

²²Sarah Perez, *Wikipedia’s Next Big Thing : Wikidata, A Machine-Readable, User-Editable Database Funded By Google, Paul Allen And Others*, TechCrunch, 30 mars 2012, URL : <https://techcrunch.com/2012/03/30/wikipedias-next-big-thing-wikidata-a-machine-readable-user-editable-database-funded-by-google-paul-allen-and-others/> (visité le 03/09/2025).

²³P180 (dépeint / depicts), Wikidata, URL : <https://www.wikidata.org/wiki/Property:P180> (visité le 03/09/2025).

²⁴R. van Straten, *Iconography indexing iconclass...*, p. 101.

serve une mutation des systèmes descriptifs en systèmes d'indexation en s'appuyant sur les technologies fournies par les outils numériques.

Par système descriptif, on entend des systèmes reposant sur le langage naturel, utilisant des mots ou des expressions pour décrire une image. Les thésaurus entrent dans cette catégorie, car ils fournissent des termes contrôlés permettant de qualifier les représentations. Le thésaurus iconographique Garnier a été conçu dans cet esprit. Dans sa version papier, il propose ainsi une liste hiérarchisée de plus de 3 200 descripteurs, répartis en 23 classes. Parmi celles-ci, 14 classes sont considérées comme fixes, tandis que les autres peuvent être adaptées à des collections particulières²⁴. Ces classes se répartissent en deux catégories principales : thèmes et sujets. Les thèmes regroupent des motifs communs, classés en 14 catégories, telles que la nature, les transports et la communication, ou les ornements. Les sujets se réfèrent à des scènes ou des personnages identifiables, issus de textes bibliques, de la mythologie, de l'histoire ou de récits. Pour représenter les relations hiérarchiques entre les éléments et les catégories, des parenthèses sont ajoutées dans les descriptions. Par exemple, pour la peinture représentant Saint Apolline et Saint Sébastien du Pérugin, conservée au Musée de Grenoble (figure A.3), la description selon le thésaurus papier devait être formulée ainsi :

Groupe de figures (sainte Apolline, pince, dent, en pied, de face, Saint Sébastien, flèche).

Pour la peinture de *Persée délivrant Andromède* de Paul Véronèse (1528-1588) conservée au Musée des beaux-arts de Rennes (inv. 1801.1.1) (figure A.2), la description iconographique pourrait être formulée ainsi :

Scène mythologique (Persée, épée, Andromède, monstre : marin, bord de mer, délivrance, rocher, chaîne) - fonds de paysage (ville)

Ainsi, la description formulée s'apparente à une description formulée en langage naturel, en s'appuyant sur les termes fournis par le thésaurus.

Quant aux systèmes d'indexation, ils n'avaient pas pour objectif de fournir une description complète des images, mais plutôt d'attribuer des mots-clés ou des codes normalisés permettant de les classer et les retrouver dans une base de données. Ils reposent sur des concepts pouvant désigner une idée, un terme, une scène, un motif ou un élément de représentation, chacun étant identifié par un code ou un mot-clé. Iconclass en constitue l'exemple le plus abouti : il utilise des codes alphanumériques qui restituent la hiérarchie du système et précisent progressivement un concept. Le premier caractère du code correspond à la catégorie

²⁴F. Garnier et Ministère de la Culture, *Thésaurus iconographique, système descriptif des représentations...*, p. 15.

la plus générale, puis les suivants affinent la description. Ainsi, le *Persée délivrant Andromède* de Paul Véronèse est indexé par le code 94P211. Celui-ci indique successivement : la mythologie classique (9), les légendes héroïques grecques (94), l’histoire de Persée (P), ses amours (2) avec Andromède (1), et plus précisément l’épisode où Persée tue le monstre marin pour la délivrer (1). La notice correspondante décrit en français : « 94P211 Persée tue le monstre marin (ou le pétrifie en lui montrant la tête de Méduse) et délivre Andromède ; il se peut que les parents de celle-ci, Céphée et Cassiopée, observent la scène »²⁵. Dans d’autres cas, le code peut inclure un nom propre, comme pour *saint Sébastien et sainte Apolline* de Pérugin (figure A.3, qui serait décrit à l’aide de ces deux codes : (11H(SEBASTIAN))²⁶ et (11HH(APOLLONIA))²⁷. Des qualificatifs supplémentaires permettent également de préciser une action ou un contexte : par exemple, un (+1) à la suite d’un code désignant un saint signale la présence d’un ange à ses côtés, tandis que 34B12(+946) désigne un chat dormant (34B12 pour le chat, (+946) pour l’action de dormir)²⁸. L’usage des systèmes d’indexation comme Iconclass se distingue donc nettement d’un thésaurus. Ils sont plus complexes à utiliser, mais sont capables de cerner le sujet principal ou un motif précis tout en conservant la hiérarchie des idées associées.

III. Vers une interopérabilité des outils d’indexation

Avec la mise en ligne des thésaurus et systèmes de description iconographique, on observe une mutation des “systèmes descriptifs” à des “système d’indexation”. Dans ce cadre, chaque descripteur devient un concept doté d’un identifiant, permettant le plus souvent de former l’URL du concept. Ce concept contient toujours un libellé principal (label), parfois enrichi de variantes, de traductions, ou d’une description. C’est le cas des vocabulaires du Getty Research Institute, mais aussi du thésaurus Garnier dans ses instances gérées par le Ministère de la Culture et par l’Institut national d’histoire de l’art. La majorité de ces vocabulaires contrôlés reposent sur le format RDF (Resource Description Framework), standard du W3C pour représenter des données structurées et interopérables sur le web²⁹. Ce format exprime les informations sous forme de triplets : un sujet (par ex. un concept du thésaurus), un prédicat (par ex. « a pour libellé ») et un objet (par ex. « La Crucifixion »). Pour uniformiser ces repré-

²⁵ICONCLASS 94P211, <https://iconclass.org/94P211>

²⁶ICONCLASS 11H(SEBASTIAN), <https://iconclass.org/fr/11H%28SEBASTIAN%29>

²⁷ICONCLASS 11HH(APOLLONIA), <https://iconclass.org/11HH%28APOLLONIA%29>

²⁸R. van Straten, *Iconography indexing iconclass...*, p. 59.

²⁹RDF Working Group, *Resource Description Framework (RDF)*, 25 févr. 2014, URL : <https://www.w3.org/RDF/> (visité le 03/09/2025).

sentations, on utilise souvent le modèle SKOS (Simple Knowledge Organization System)³⁰, qui définit les unités du thésaurus comme des <skos :Concept>, leur libellé de référence avec <skos :prefLabel>, et les relations hiérarchiques avec <skos :broader> pour le concept plus large, soit le parent dans la hiérarchie, et <skos :narrower> pour le concept plus spécifique, soit l'enfant dans la hiérarchie. Cette approche favorise l'interopérabilité, le multilinguisme et la navigation hiérarchique dans les vocabulaires. Cependant, une véritable interconnexion entre vocabulaires reste rare. C'est l'ambition de Wikidata, qui relie ses entités à des référentiels externes via des propriétés dédiées. Ainsi, environ 1 109 concepts du Getty Iconography Authority³¹, 24 597 du Getty Art and Architecture Thesaurus³², 4 131 identifiants Iconclass³³ y sont déjà utilisés. Notons également que 257 entités Wikidata sont associées à la fois à un concept du Getty Iconography Authority et à un concept d'ICONCLASS³⁴. En revanche, aucune propriété n'existe encore pour le thésaurus Garnier, qu'il s'agisse de la version du ministère de la Culture ou de l'INHA. La création d'une telle propriété et l'alignement des descripteurs aux entités Wikidata pourraient permettre de désenclaver le thésaurus Garnier, d'assurer sa compatibilité multilingue et de relier les bases de données iconographiques françaises à celles internationales. Un obstacle majeur à prendre en compte pour l'interopérabilité des systèmes de description iconographique reste la différence de granularité entre eux. Iconclass décrit chaque concept par un identifiant unique, tel que 94P211 pour Persée délivrant Andromède, tandis que le thésaurus Garnier nécessiterait de combiner plusieurs termes (« Scène mythologique », « Persée », « Andromède », « délivrance ») pour exprimer la même idée. Malgré cela, Iconclass mène une politique active de liaisons vers d'autres bases et références bibliographiques, même si l'accès à ses services avancés (Iconclass+) reste payant³⁵.

La diversité des vocabulaires contrôlés et des systèmes d'indexation représente à la fois un potentiel majeur pour l'indexation iconographique et un défi pour son efficacité : il

³⁰Alistair Miles et Sean Bechhofer, *SKOS (Simple Knowledge Organization System Reference)*, 18 août 2009, URL : <https://www.w3.org/TR/skos-reference/> (visité le 03/09/2025).

³¹Calculé à partir de la requête SPARQL suivante : [https://query.wikidata.org/sparql?query=SELECT%20\(COUNT\(DISTINCT%20%3FId\)%20AS%20%3Fcount\)%20WHERE%20%7B%20%3Fitem%20wdt%3AP5986%20%3FId.%20%7D](https://query.wikidata.org/sparql?query=SELECT%20(COUNT(DISTINCT%20%3FId)%20AS%20%3Fcount)%20WHERE%20%7B%20%3Fitem%20wdt%3AP5986%20%3FId.%20%7D) (consultée le 3/09/2025)

³²Calculé à partir de la requête SPARQL suivante : [https://query.wikidata.org/sparql?query=SELECT%20\(COUNT\(DISTINCT%20%3FId\)%20AS%20%3Fcount\)%20WHERE%20%7B%20%3Fitem%20wdt%3AP1014%20%3FId.%20%7D](https://query.wikidata.org/sparql?query=SELECT%20(COUNT(DISTINCT%20%3FId)%20AS%20%3Fcount)%20WHERE%20%7B%20%3Fitem%20wdt%3AP1014%20%3FId.%20%7D) (consultée le 3/09/2025)

³³Calculé à partir de la requête SPARQL suivante : [https://query.wikidata.org/sparql?query=SELECT%20\(COUNT\(DISTINCT%20%3FId\)%20AS%20%3Fcount\)%20WHERE%20%7B%20%3Fitem%20wdt%3AP1256%20%3FId.%20%7D](https://query.wikidata.org/sparql?query=SELECT%20(COUNT(DISTINCT%20%3FId)%20AS%20%3Fcount)%20WHERE%20%7B%20%3Fitem%20wdt%3AP1256%20%3FId.%20%7D) (consultée le 3/09/2025)

³⁴Calculé à partir de la requête SPARQL suivante : [https://query.wikidata.org/sparql?query=SELECT%20\(COUNT\(DISTINCT%20%3Fitem\)%20AS%20%3Fcount\)%20WHERE%20%7B%0A%20%3Fitem%20wdt%3AP5986%20%3FiaId%3B%0A%20%20%20wdt%3AP1256%20%3FiconclassId.%0A%7D](https://query.wikidata.org/sparql?query=SELECT%20(COUNT(DISTINCT%20%3Fitem)%20AS%20%3Fcount)%20WHERE%20%7B%0A%20%3Fitem%20wdt%3AP5986%20%3FiaId%3B%0A%20%20%20wdt%3AP1256%20%3FiconclassId.%0A%7D) (consultée le 03/09/2025)

³⁵H. Brandhorst, *A virtual library for iconographic research*, ICONCLASS, 6 avr. 2023, URL : <https://forum.iconclass.org/t/a-virtual-library-for-iconographic-research/206> (visité le 03/09/2025).

reste nécessaire de mettre ces outils en correspondance afin de permettre le croisement des descriptions. Chacun d’eux contribue cependant à décrire les images avec précision, facilitant l’exploration des bases de données et la compréhension des représentations.

Chapitre 3

L'indexation iconographique à l'épreuve des technologies numériques

Comme nous l'avons souligné, la plupart des thésaurus iconographiques ont été conçus dès l'origine pour alimenter des bases de données informatisées, ou ont été adaptés par la suite pour pouvoir y être intégrés. Avec l'essor d'Internet et la mise en ligne massive des collections patrimoniales, ces thésaurus ont dû s'ajuster aux contraintes techniques et aux besoins spécifiques de leur environnement numérique. Aujourd'hui, les bases de données iconographiques sont très nombreuses, et la plupart intègrent un thésaurus, parfois spécialement élaboré pour elles. Dans certains cas plus rares, la base de données elle-même devient un véritable espace de valorisation du système d'indexation, où les images sont directement mises en relation avec les concepts qu'elles représentent. Enfin, l'essor des projets en intelligence artificielle suscite une réflexion sur l'intégration des données générées au sein des bases de données.

I. Quelques bases de données intégrant des systèmes d'indexation

Les bases de données intégrant des systèmes d'indexation et des vocabulaires contrôlés sont très nombreuses. Pour cette étude, nous nous concentrerons sur celles qui contiennent des œuvres d'art picturales et qui utilisent l'indexation iconographique pour faciliter la recherche.

Dans la plupart des cas, cette méthode de facilitation de l'exploration vient compléter une base de données existante, qu'il s'agisse d'une base interne à une institution, regroupant plusieurs collections, ou conçue pour un projet spécifique. En France, conformément aux recommandations du Ministère de la Culture, la majorité des bases de données utilisent

le thésaurus Garnier pour l’indexation iconographique. À l’origine, ce thésaurus avait été conçu pour les inventaires d’œuvres et de monuments sous la tutelle de ce ministère, qui sont aujourd’hui rassemblés dans la plateforme POP (Plateforme ouverte du patrimoine). Lancée en 2019, POP vise à centraliser et valoriser les données patrimoniales gérées par le ministère depuis la fin des années 1970, en regroupant plusieurs anciennes bases : Joconde, Mérimée, Palissy, Mémoire, MNR Rose Valland, Muséofile, Autor et Enluminures¹.

Certaines de ces bases, comme Joconde (catalogue collectif des collections des musées de France) ou Palissy (patrimoine mobilier protégé ou recensé dans le cadre de l’Inventaire général du patrimoine culturel), utilisent le thésaurus Garnier depuis les années 1980. Aujourd’hui, POP rassemble plus de 4,5 millions de notices, dont plus de 3,8 millions accompagnées d’images. La section « Sujet représenté » de chaque notice indique les descripteurs issus du thésaurus Garnier. Cependant, le moteur de recherche ne permet pas d’interroger directement la base à partir d’un descripteur provenant du thésaurus : l’utilisateur doit d’abord trouver une notice qui contient le terme et cliquer sur le lien associé. Ce lien est souvent défectueux, insérant des espaces supplémentaires codés en « %20 », ce qui empêche l’affichage des résultats. La seule solution consiste à corriger manuellement l’URL. Ainsi, malgré son ambition, POP ne facilite pas réellement la recherche à l’aide de l’indexation. Bien que la plateforme soit censée être connectée à l’instance Opentheso du thésaurus, de nombreux termes présents dans les notices de POP n’ont pas d’équivalent parmi les concepts de cet outil, limitant la cohérence et l’efficacité du dispositif². Certaines institutions, comme le Musée d’Orsay à Paris³ ou le Musée de Reims⁴, exposent également l’indexation grâce au thésaurus Garnier sur leurs propres bases.

Une autre plateforme utilisant le thésaurus Garnier est AGORHA (Accès global et organisé aux ressources en histoire de l’art)⁵. Développée par l’INHA, elle donne accès aux bases de données créées dans le cadre de projets de recherche en histoire de l’art et en archéologie menés par cette institution ou ses partenaires. Les données avaient d’abord été publiées sur le site institutionnel de l’INHA à partir de 2004, avant la mise en ligne d’AGORHA et l’ouverture au public en 2011. L’interface a été améliorée en 2018, avec une meilleure ergonomie, des recherches par auto-complétion ou par filtres et une présentation plus claire des données⁶.

¹Ministère de la Culture, « POP », *la plate-forme ouverte du patrimoine*, Ministère de la Culture, URL : <https://www.culture.gouv.fr/fr/espace-documentation/bases-de-donnees/Fiches-bases-de-donnees/pop-la-plate-forme-ouverte-du-patrimoine> (visité le 26/08/2025).

²Id., *Thésaurus iconographique Garnier...*

³Musée d’Orsay, *Les collections*, URL : <https://www.musee-orsay.fr/fr/collections> (visité le 03/09/2025).

⁴Musée de Reims, *Les œuvres en ligne*, Portail officiel des Musées de Reims, URL : <https://musees-reims.fr/fr/musee-numerique/oeuvres-en-ligne> (visité le 03/09/2025).

⁵Institut national d’histoire de l’art, *AGORHA*, URL : <https://agorha.inha.fr/> (visité le 03/09/2025).

⁶Pierre-Yves Laborde, *Les bases de données de l’INHA, une longue histoire*, Numrha, 18 mars 2021, URL :

La version actuelle, mise en ligne en 2022, est gérée par le Service numérique de la recherche à l'INHA et contient 54 bases de données, rassemblant plus de 250 000 notices. Celles-ci sont réparties en six catégories : « œuvres », « personnes », « événements », « références », « collections » et « articles & actualités ». L'indexation iconographique à l'aide du thésaurus Garnier intervient dans les notices de type « œuvre », dans la section « Description », puis « Représentations », puis dans l'onglet « Indexation Garnier-SMF ». Le thésaurus est hébergé sur la plateforme Thésaurus et vocabulaires contrôlés de l'INHA, grâce à l'outil GINCO développé par le ministère de la Culture⁷. Cette plateforme contient l'ensemble des vocabulaires contrôlés de l'INHA, au format SKOS, interrogables via une interface SPARQL et téléchargeables au format XML-RDF.

L'indexation iconographique des œuvres dans AGORHA reste toutefois limitée. Nous avons extrait les données de 12 bases contenant majoritairement des œuvres d'art pour calculer leurs statistiques d'indexation (figures B.1 et B.2)⁸. Certaines bases, comme celle liée à l'Iconographie musicale⁹ ou Envois de Rome en peinture et sculpture (1804-1914)¹⁰, ne contiennent aucune œuvre indexée. D'autres, comme la base Peinture produite en France au XVI^e siècle¹¹, affichent moins de 10 % des œuvres indexées. Le RETIF appartient à cette catégorie, avec moins de 1 % des notices indexées (48 sur 14 066), ce qui a motivé le projet d'indexation assistée par intelligence artificielle. À l'inverse, certaines bases connaissent une meilleure indexation. Toutes les notices du Répertoire des tableaux français en Allemagne (XVII^e-XVIII^e siècles, REPFall) contiennent au moins un descripteur, mais seulement 25 termes différents y sont utilisés, avec une moyenne d'un terme par notice, principalement des genres et sujets iconographiques (“portrait”, “paysage”, “scène de genre”, “Nouveau Testament”). Le RETIB, relatif aux tableaux ibériques dans les collections françaises, est très bien indexé, puisque 86 % des notices le sont, avec une moyenne de 2,35 termes par notice et 317 descripteurs différents. La base La fabrique de l'art, dédiée aux enluminures médiévales et développée par Charlotte Denoël et Teresa Knapowska, présente également une excellente indexation : trois quarts des notices contiennent au moins un label, avec une moyenne de 5,14 termes par notice parmi 1 028 concepts distincts.

<https://numrha.hypotheses.org/2164> (visité le 27/07/2025).

⁷Institut national d'histoire de l'art, *Thésaurus et vocabulaires contrôlés de l'Institut national d'histoire de l'art*, URL : <http://thesaurus.inha.fr/thesaurus/> (visité le 26/08/2025).

⁸Données calculées entre le 25 avril 2025 et le 14 mai 2025, voir les graphiques présentés sur Tableau-Public : <https://public.tableau.com/views/AGORHA/UtilisationdeGarnierdansAgorha>.

⁹Id., *Iconographie musicale : répertoire d'œuvres d'art à sujets musicaux publiées par Albert Pomme de Mirimonde*, AGORHA, URL : <https://agorha.inha.fr/database/17> (visité le 01/09/2025).

¹⁰Id., *Les Envois de Rome en peinture et sculpture, 1804-1914*, AGORHA, URL : <https://agorha.inha.fr/database/80> (visité le 01/09/2025).

¹¹Id., *Recensement de la peinture produite en France au XVIe siècle*, AGORHA, URL : <https://agorha.inha.fr/database/50> (visité le 01/09/2025).

AGORHA permet, via la recherche avancée, de limiter la recherche à la section de l’indexation iconographique à l’aide du champs ”Description > Représentations > Indexation Garnier-SMF”. Néanmoins, deux limites persistent et réduisent l’efficacité de la requête : premièrement, la recherche ne propose pas d’autocomplétion et fonctionne uniquement en texte intégral. Ainsi, il n’est pas possible de sélectionner un concept du thésaurus dans une liste déroulante et la requête se limite à une recherche traditionnelle du texte au sein des indexations des notices. Par exemple, une recherche sur ”Saint Jean” cherchera les deux mots parmi les indexations des œuvres et affichera également ”Saint Jean-Baptiste” ou ”Saint Jean Chrysostome”, sans possibilité de restreindre la chaîne exacte. Pour contourner ce problème, il faut procéder comme dans POP : repérer une notice contenant le terme souhaité, puis cliquer sur la fonction « rechercher à partir de ce terme », ce qui crée un critère de filtrage plus précis des résultats. Deuxièmement, la hiérarchie n’est pas prise en compte : rechercher le concept parent ”la nature” n’inclut pas automatiquement les concepts enfants issus de la faune et la flore, même en filtrant via une notice contenant le terme¹². Cela pose problème pour des concepts dont les concepts enfants sont très proches, comme ”chien” et les races liées, ou ”Vie de la Vierge” et les événements qui en font partie (Annonciation, la Visitation, le Couronnement, etc.). Ainsi, des améliorations restent à apporter pour améliorer l’efficacité de l’indexation par le thésaurus Garnier dans ces deux bases de données.

Afin de compléter ce tour d’horizon, il convient de mentionner plusieurs outils récents qui visent à valoriser les objets provenant de collections diverses indexées à l’aide d’un même système iconographique. C’est le cas de Crotos, un outil de visualisation qui recense actuellement plus de 443 000 œuvres issues de Wikidata, illustrées par des images provenant de Wikimedia Commons¹³. Grâce à son moteur de recherche, il permet d’interroger les artefacts patrimoniaux à partir d’entités Wikidata. Ses fonctionnalités restent toutefois limitées, puisqu’il n’est pas possible de combiner plusieurs filtres ou propriétés. L’outil openArtBrowser, développé par le Dr Bernhard G. Humm à Darmstadt, reprend le même principe tout en offrant une interface plus ergonomique et enrichie. Il propose notamment des visualisations chronologiques et la possibilité d’appliquer plusieurs filtres simultanément¹⁴. Enfin, citons Arkyves, qui agrège des objets patrimoniaux provenant de nombreuses collections indexées avec le système ICONCLASS. Il donne accès à plus de 900 000 images issues d’institutions patrimoniales du monde entier, principalement néerlandaises (comme le Rijksmuseum d’Amsterdam ou le Rijksbureau voor Kunsthistorische Documentatie à La Haye), mais aussi

¹²Lien de recherche pour ”la nature” : https://agorha.inha.fr/recherche?terms=concept_lie%3A%20%224f1f0a43-b194-4688-a19e-04bb95dd1f3a%22 (consulté le 3/09/2025)

¹³Crotos, URL : <https://zone47.com/crotos/> (visité le 26/08/2025).

¹⁴Bernhard G. Humm, *openArtBrowser*, URL : <https://openartbrowser.org/fr/> (visité le 03/09/2025) ; Id., *Fascinating with Open Data : openArtBrowser*, 2020, URL : https://ceur-ws.org/Vol-2535/paper_2.pdf (visité le 27/08/2025).

allemandes (Bayerische Staatsbibliothek de Munich, Herzog August Bibliothek de Wolfenbüttel), britanniques, italiennes ou américaines. La base, initialement baptisée Mnemosyne, a été enrichie dès ses débuts par Yassu Frossati, Peter van Huisstede et Hans Brandhorst, qui en demeure l'éditeur. L'interface logicielle a été développée par Etienne Posthumus et, depuis 1999, le service est diffusé par Brill¹⁵. Cet outil permet de filtrer les œuvres selon différents critères : le médium artistique, l'institution qui les conserve, et surtout grâce au système d'indexation ICONCLASS.

II. La place de l'indexation dans les projets d'intelligence artificielle

Aujourd'hui, peu de projets en intelligence artificielle ont abouti à une véritable mise en œuvre de l'indexation à l'aide de systèmes iconographiques. Certains, comme le projet HikarIA et Notre-Dame, ont toutefois documenté l'intégration de l'intelligence artificielle dans les pratiques d'indexation et la présentation des résultats.

Le projet Hikaria, porté depuis 2023 par le Musée national des Arts Asiatiques – Guimet à Paris et la société française Teklia, vise notamment la « mise en valeur du patrimoine photographique en améliorant l'indexation automatisée des photographies anciennes »¹⁶. Il porte sur un fonds de plus de 19 000 prototypes du Japon, datés des époques Bakumatsu et Meiji (1853-1912), issus en grande partie de la collection Joseph Dubois et conservés au Musée Guimet. Ces images ont été numérisées et mises en ligne selon le protocole IIIF¹⁷. Une première version de l'application a été rendue publique le 8 novembre 2024, et le site continue d'évoluer jusqu'à la fin prévue du projet en 2026¹⁸*NotesVersionSite*. Le traitement des images a d'abord consisté à numériser les reliures et pages d'albums, puis à détecter automatiquement les photographies. Un moteur de recherche permet ensuite des requêtes en langage naturel sur les contenus visuels¹⁹. Pour un corpus test de quelques centaines de photographies, le modèle GPT-4 d'OpenAI a été mobilisé afin de proposer de nouveaux termes d'indexation²⁰. L'usage de vocabulaires contrôlés a été écarté, car jugé trop contraignant dans l'interaction avec un

¹⁵E. Posthumus et H. Brandhorst, *Arkyves*, 1999, URL : <https://brill.com/display/db/arko> (visité le 03/09/2025).

¹⁶Édouard de Saint-Ours et Christopher Kermorvant, « Le projet HikarIA : étude et mise en valeur du patrimoine photographique par l'intelligence artificielle », *Culture et Recherche* 147 (2024), p. 20-24, p. 24.

¹⁷*Ibid.*, p. 20.

¹⁸*Ibid.*, p. 24.

¹⁹Nous abordons plus en détail ce moteur de recherche et la similarité d'images à l'aide du modèle CLIP dans le chapitre 8.

²⁰C. Kermorvant et É. de Saint-Ours, *Le projet HikarIA : analyse iconographique automatisée de photographies patrimoniales*, 16 juin 2025, URL : <https://www.inha.fr/agenda/le-projet-hikaria-analyse-iconographique-automatisee-de-photographies-patrimoniales/>.

grand modèle de langue²¹. Afin de distinguer les descripteurs générés automatiquement de ceux proposés par les indexeurs, les développeurs du site ont apposé une petite icône de robot pour les premiers et une coiffe académique pour les seconds. Par exemple, la dixième vue de l'album 47 (fonds Dubois) a été indexée à la fois par un documentaliste et par GPT-4²². La photographie représente deux femmes sous des ombrelles, devant un portail en bois menant à une maison. Le modèle a proposé certains termes pertinents non relevés par l'indexeur humain (par ex. « barrière », « portail », « chemin »), mais aussi de nombreux descripteurs erronés ou fantaisistes, tels que « bac à gaz », « punition » ou « consulat ». Teklia et le Musée Guimet ont néanmoins fait le choix de rendre visibles ces propositions automatiques avant d'en réaliser le contrôle²³.

Actuellement, l'onglet « Rechercher » ne permet d'interroger qu'à l'aide de la technologie d'encodage des images CLIP, sans prendre en compte les tags d'indexation. Pour accéder aux images liées à un tag, l'utilisateur doit procéder comme pour les plateformes POP et AGORHA, en cliquant sur un mot-clé déjà associé à une photographie.

Un autre projet, cette fois fondé sur l'usage systématique d'un vocabulaire contrôlé, est lié au chantier de restauration de la cathédrale Notre-Dame de Paris après l'incendie du 15 avril 2019. Dans ce cadre, le groupe Données numériques du Chantier scientifique Notre-Dame (CNRS/Ministère de la Culture), coordonné par Livio de Luca, a mis en place l'écosystème numérique « n-dame » pour rassembler les données du projet²⁴. Après la collecte des données de numérisation, une expérimentation a consisté à détecter et segmenter automatiquement les objets d'intérêt (colonnes, statues, vitraux, éléments décoratifs) dans les relevés 2D et 3D. Trois modèles ont été utilisés pour lier ces images au thésaurus : DINO, SAM et CLIP²⁵. Le modèle DINO a servi à la détection des éléments d'intérêt, le modèle SAM à la segmentation de ces objets pour extraire des masques, qui ont ensuite été analysés par CLIP, afin de leur associer une étiquette issue d'un thésaurus²⁶. Les données ne sont pas encore accessibles en

²¹É. de Saint-Ours et C. Kermorvant, « Le projet HikarIA : étude et mise en valeur du patrimoine photographique par l'intelligence artificielle »..., p. 23.

²²Voir : <https://hikaria.org/photograph/4119977d-eaef-4192-8d3c-16036bfa087e/> (consulté le 20/07/2025). Lors des vérifications effectuées le 1er septembre 2025, peu de temps avant la restitution de ce mémoire, toutes les indexations générées par le modèle GPT avaient été retirées du site, sans qu'aucune note de version ne précise les changements effectués *Notes de version - Hikaria*, HikarIA, URL : <https://hikaria.org/about/release-notes> (visité le 01/09/2025). Nous avons choisi de conserver nos remarques.

²³C. Kermorvant et É. de Saint-Ours, *Le projet HikarIA...*

²⁴Livio De Luca, Violette Abergel, Anaïs Guillem, Olivier Malavergne, Adeline Manuel, Ariane Néroulidis, Roxane Roussel, Miled Rousset et Sarah Tournon, *L'écosystème numérique n-dame pour l'analyse et la mémorisation multi-dimensionnelle du chantier scientifique Notre-Dame-de-Paris*, SCAN'22, 2022 ; L. De Luca, « A digital ecosystem for the multidisciplinary study of Notre-Dame de Paris », *Journal of Cultural Heritage*, 65 (2024), p. 206-209, DOI : 10.1016/j.culher.2023.09.011.

²⁵Nous développons davantage le rôle de ces modèles dans le chapitre 5.

²⁶L. De Luca, A. Guillem et Kévin Réby, « Intelligence humaine, collective et artificielle dans le chantier scientifique et de restauration de Notre-Dame de Paris », *Culture et Recherche*–147 (2024), p. 8-12, p. 10.

ligne.

Ces deux projets illustrent deux stratégies distinctes d'intégration de l'intelligence artificielle dans l'indexation : la première mise sur la génération libre de termes par un grand modèle de langue, avec un contrôle *a posteriori* et une différenciation claire entre les apports humains et automatiques ; la seconde s'appuie sur un vocabulaire contrôlé, intégré dans la chaîne opératoire, garantissant la qualité et la cohérence des descripteurs.

L'indexation iconographique est donc une pratique inscrite dans la discipline de l'histoire de l'art qui repose sur des normes définies par les systèmes et vocabulaires contrôlés mobilisés. Elle demeure un outil essentiel pour assurer l'accessibilité, la recherche et l'interconnexion des collections à l'ère numérique. Néanmoins, la complexité de sa mise en œuvre pose un véritable défi lorsqu'il s'agit de l'intégrer dans des projets recourant à l'intelligence artificielle. C'est pourquoi il importe désormais d'analyser les enjeux techniques et méthodologiques de l'automatisation de ce processus.

Deuxième partie

**Enjeux techniques et méthodologiques
de l'automatisation de la description
iconographique de peintures**

Chapitre 4

Comprendre et exploiter les données du projet

Dans l’« Artificial Intelligence Planning Framework » (Cadre de planification pour l’intelligence artificielle) proposé par le LC Labs de la Bibliothèque du Congrès en 2023¹, proposant un guide pour les projets menés par les institutions patrimoniales utilisant ces technologies, trois phases de planification sont mis en avant : Comprendre, expérimenter et mettre en œuvre, chacune soutenant l’évaluation de trois éléments nécessaires au projets : les données, les modèles et les personnes. La première phase consiste à comprendre le besoin et l’objectif de la tâche ainsi que le personnel mis en œuvre et les données qui seront utilisées pour la tâche.

Dans le cas du projet d’indexation du RETIF accompagnée par l’intelligence artificielle, le but de l’expérimentation est de tester son automatisation pour fournir une indexation grâce au thésaurus Garnier sur la plateforme AGORHA. Si au sein du consortium PictorIA et de l’INHA, à l’initiative du projet, des personnels avec des compétences pouvaient fournir une assistance ponctuelle, une seule personne se consacrait à mener le projet sur une période courte de quatre mois. Les données à prendre en compte sont le corpus d’images du RETIF et les métadonnées associées, ainsi que le thésaurus Garnier. Des étapes de préparation sont nécessaires pour préparer les données à leur intégration et envisager l’automatisation du processus.

¹Library of Congress, *Planning Framework used by LC Labs for planning AI experiments towards responsible implementation*, Github, 25 juill. 2025, URL : <https://github.com/LibraryOfCongress/labs-ai-framework> (visité le 01/09/2025). Nous utilisons la traduction de Jean-Philippe Moreux : *Planification de projet IA, traduction de l'Artificial Intelligence Planning Framework de la Library of Congress*, Github, 23 mai 2025, URL : <https://github.com/altomator/Planification-de-projets-IA> (visité le 01/09/2025).

I. Le corpus du RETIF

Le Répertoire des tableaux italiens dans les collections publiques françaises (XIII^e-XIX^e siècles) (RETIF) est issu d'un programme de recherche initié par Michel Laclotte et porté par l'Institut national d'histoire de l'art (INHA)². Lancé en 2001, le projet a d'abord été dirigé par Mickaël Szanto (2001–2004), puis successivement par Éric Pagliano (2004–2008), Nathalie Volle (2009–2015) et Servane Dargnies (2015–2019). Depuis 2019, la direction scientifique est assurée par Clémence Raccah, Michal Litwinowicz et Pierre-Yves Laborde. Le corpus recense exclusivement des peintures de format mobile : les pastels et peintures murales en sont exclus, à l'exception des fresques détachées et exposées dans les institutions patrimoniales. Son objectif est de signaler les œuvres présentes dans les collections françaises, dans les musées, édifices civils et religieux et monuments historiques, réalisées entre le XIII^e siècle et 1914 par des artistes italiens, ou étrangers ayant exercé en Italie³. Le répertoire inclut également certaines copies non italiennes d'œuvres de la péninsule, dès lors qu'elles présentent « un certain niveau de qualité »⁴. Aujourd'hui, RETIF est hébergé par l'INHA et consultable en ligne via le portail AGORHA (Accès global et organisé aux ressources en histoire de l'art).

Sur cette plateforme, le corpus RETIF rassemble 14 059 œuvres documentées, parmi lesquelles 11 974 sont accompagnées d'une image⁵. La chronologie du corpus s'étend du XIII^e siècle au début du XX^e siècle. Les œuvres médiévales restent marginales, avec seulement 21 pièces pour le XIII^e siècle et 347 pour le XIV^e. Le corpus devient plus abondant à l'époque moderne : autour de 1 000 œuvres pour le XV^e siècle, environ 2 600 pour le XVI^e, et plus de 5 600 pour le XVII^e, qui représente à lui seul plus du tiers du corpus. Le XVIII^e siècle compte près de 2 150 œuvres. En comparaison, l'époque contemporaine est peu représentée : environ 1 300 œuvres pour le XIX^e siècle et une cinquantaine pour le XX^e. Enfin, 948 œuvres ne comportent pas de datation, soit par absence d'information dans le champ prévu, soit parce que la date est inconnue, notamment dans le cas de copies. La plupart de ces œuvres non datées peuvent néanmoins être rattachées au XVII^e ou au XVIII^e siècle. L'analyse du corpus à l'aide des logiciels PixPlot et Panoptic permet de visualiser la répartition des su-

²Nathalie Volle, *Michel Laclotte et le Répertoire des Tableaux Italiens en France, XIII^e-XIX^e siècles (RETIF)*, AGORHA, 5 févr. 2024, URL : <https://agorha.inha.fr/detail/940> (visité le 08/08/2025).

³Matteo Ganeselli, Marie Tchernia et N. Volle, *Interroger la base en ligne les peintures italiennes en France (RETIF) sur Agorha (INHA)*, Blog de l'APAHAU, 8 déc. 2011, URL : <http://blog.apahau.org/interroger-la-base-en-ligne-les-peintures-italiennes-en-france-retif-sur-agorha-inha/> (visité le 08/08/2025).

⁴Institut national d'histoire de l'art, *Répertoire des tableaux italiens dans les collections publiques françaises (XIII^e-XIX^e siècles)*, RETIF - Agorha, AGORHA, URL : <https://agorha.inha.fr/database/4> (visité le 01/09/2025).

⁵*Ibid.*

jets iconographiques⁶. Une grande majorité des peintures représente des scènes religieuses, destinées au décor des lieux de culte ou à la dévotion, illustrant des personnages et épisodes de l'Ancien et du Nouveau Testament, ainsi que des représentations de saints, au cours de leur vie ou sous forme de portraits. Les paysages naturels et urbains constituent également une part importante du corpus, parmi lesquels on retrouve de nombreuses vedute de Venise et de ses environs, ou des caprices architecturaux peuplés de ruines peints en majorité au XVIII^e siècle. Quelques paysages historiques méritent d'être signalés, illustrant des scènes intégrées à des environnements naturels, un genre pictural apprécié au XVII^e siècle. Les scènes mythologiques, historiques ou allégoriques représentent une proportion plus modeste du corpus, à hauteur de quelques centaines d'images. Une fraction du corpus est composée de natures mortes, principalement des XVII^e et XVIII^e siècles, parfois associées à des personnages, comme l'étal d'un poissonnier ou d'un marchand de légumes. Le corpus inclut également quelques peintures animalières, des scènes de bataille et des peintures de genre.

La qualité et le format des images varient fortement selon les campagnes de prise de vue. Si la plupart des photographies provenant de collections muséales sont exploitables, nombreuses sont celles qui incluent le cadre de la peinture, peu pertinent pour la description iconographique. Une grande partie des images est en noir et blanc, ce qui peut compliquer la reconnaissance des objets. Certaines œuvres sont photographiées dans leur environnement d'origine, notamment pour les retables ou les tableaux d'autel dans les églises, ou lorsqu'elles sont décoratives dans une pièce d'habitation. D'autres reproductions sont difficilement analysables à l'œil nu, en raison d'une résolution insuffisante, d'un angle prononcé de la prise de vue ou d'un état dégradé de la peinture. Ces images posent un obstacle à un traitement automatisé, car elles risqueraient de produire des résultats peu fiables.

En revanche, le corpus bénéficie d'une grande richesse de métadonnées, produites par les équipes de recherche impliquées dans le programme RETIF : toutes les notices d'œuvres contiennent un titre descriptif indiquant le sujet ou la scène représentée. Par ailleurs, la grande majorité d'entre elles précise le nom de l'auteur de l'œuvre ou du modèle, associé au copiste dans le cas des copies, ainsi que la datation. En somme, le corpus d'images présente une grande variété et constitue un défi majeur pour l'analyse automatisée et la classification des œuvres.

II. Préparer le vocabulaire contrôlé

L'un des principaux enjeux du projet consiste à analyser et décrire les images en utilisant le système de description du thésaurus Garnier, afin de pouvoir intégrer de manière cohérente

⁶L'utilisation de ces logiciels pour comprendre le corpus est détaillée dans le chapitre 7.

ces données dans la plateforme AGORHA. Pour ce faire, il est nécessaire de préparer le vocabulaire contrôlé afin qu'il soit lisible par un utilisateur et exploitable par des langages de programmation, tels que Python. Le thésaurus Garnier de l'INHA est hébergé sur une instance Ginco au format SKOS (Simple Knowledge Organization System), un standard du web sémantique permettant de représenter des systèmes d'organisation de la connaissance, comme des thésauruses, vocabulaires contrôlés, des taxonomies ou des ontologies. Chaque terme ou notion y est défini comme un concept, contenant des informations telles que le terme du descripteur et les relations hiérarchiques parent/enfant entre eux, le tout codé en XML-RDF. L'instance du thésaurus Garnier de l'INHA contient 16 362 concepts.

Pour être exploité par la plupart des systèmes d'intelligence artificielle, le thésaurus devait être accessible depuis Python, langage utilisé pour l'ensemble de la chaîne de traitement. Cependant, le format RDF-SKOS présente certaines contraintes : il est relativement lourd et la hiérarchie des concepts y est difficile à reconstruire. Les concepts sont organisés par ordre chronologique de création, et la hiérarchie est explicitée dans chaque concept par les balises <skos :broader> et <skos :narrower>, indiquant respectivement l'identifiant du concept parent et les concepts enfants. Cette structure complexe rend l'exploitation directe en Python peu pratique et difficile à appréhender pour un utilisateur et il convient de créer une version allégée du vocabulaire contrôlé, dans laquelle la hiérarchie serait claire et les données réduites au strict nécessaire pour le projet. Le format JSON (JavaScript Object Notation) a été choisi pour cette conversion. Pour chaque concept, le fichier JSON restitue le label préféré, les autres labels alternatifs, l'URL vers le concept dans Ginco, ainsi que la liste des concepts enfants pour conserver la hiérarchie. Deux méthodes ont été envisagées pour passer du SKOS au JSON : la première consistait à interroger directement le XML-RDF avec XQuery, un langage adapté au traitement de documents XML. Cependant, cette approche s'est révélée très lourde en termes de puissance de calcul, car chaque requête nécessitait de parcourir l'ensemble du document. L'option la plus efficace a finalement été d'utiliser la librairie python RDF pour convertir directement le thésaurus en JSONII.⁷. Le document obtenu est ainsi légèrement structuré, facile à interroger et immédiatement exploitable par des algorithmes informatiques.

III. Exploiter les métadonnées existantes

Avant de recourir à des outils de vision par ordinateur pour analyser les images du corpus, il est pertinent d'essayer de tirer parti des métadonnées déjà accessibles dans AGORHA.

⁷Documentation de la librairie Python RDFLib, URL : <https://rdflib.readthedocs.io/en/stable/> (visité le 01/09/2025).

Comme nous l'avons évoqué au chapitre 4, toutes les notices du RETIF sont accompagnées d'un titre, qui décrit généralement la scène représentée ou le sujet de la peinture, et mentionne parfois certains éléments et motifs iconographiques. Il serait donc possible de réaliser une partie de l'indexation des peintures du RETIF en recherchant les termes du thésaurus au sein de leurs titres, en utilisant le thésaurus Garnier au format JSON que nous avions extrait (cf. chapitre 4). À cette fin, un script Python, sous la forme d'un Jupyter Notebook, a été conçu pour récupérer une grande partie des descripteurs du thésaurus présents dans les titres. La première fonction du script procède ainsi : lorsqu'un terme correspond exactement au titre, la recherche s'arrête immédiatement. Les déterminants sont supprimés dès le début du traitement et, à chaque étape, les mots sont convertis au singulier et en minuscules. Par exemple, pour une œuvre intitulée *Saint Pierre*, le script ne retiendra que le label « saint Pierre » et ne poursuivra pas la recherche. Si aucun terme n'est trouvé à cette étape, le script cherche si le titre correspond à une partie d'un descripteur du thésaurus. Ainsi, pour les saints, une peinture intitulée *Saint Augustin* pourra être associée au concept « saint Augustin d'Hippone ». Grâce à l'enchaînement de ces deux fonctions, un titre comme *Saint Thomas* sera correctement identifié, sans être confondu avec « saint Thomas d'Aquin ». Un problème subsiste toutefois dans le cas de titres ambigus : par exemple, pour un tableau ayant pour titre *Sainte Catherine*, le script pourrait suggérer à la fois « sainte Catherine d'Alexandrie » et « sainte Catherine de Sienne ». Des étapes de validation seraient nécessaires pour résoudre ce type de conflits.

Ensuite, le script recherche les termes exacts du thésaurus présents dans les titres. Par exemple, *Paysage avec une tour* renverra les termes « paysage » et « tour ». Cependant, cet outil demeure à l'état de prototype, car il reste à gérer de nombreuses exceptions liées à la formulation des titres ou aux spécificités du thésaurus. Au sein du corpus, certains titres sont imprécis ou ambigus : par exemple, *Sainte Famille avec Saint Jean* fait référence à saint Jean-Baptiste, cousin de Jésus, et non à saint Jean l'Évangéliste. Dans d'autres cas, le thésaurus complique la recherche : *L'extase de saint François* ne renverra que le mot « extase », car saint François correspond dans le thésaurus à « saint François d'Assise » ; de même, *La Mort de Cléopâtre* ne sera pas correctement indexée puisque Cléopâtre est référencée comme « Cléopâtre VII ». De plus, certains termes du thésaurus comportent des précisions entre parenthèses pour différencier les homonymes, comme par exemple pour différencier le héros « Castor (mythologie) » de l'animal. Enfin, les labels alternatifs du thésaurus ne sont pas encore exploités par le script, bien qu'ils pourraient permettre certaines identifications. Par exemple, le terme « tortue » est le label alternatif du concept ayant pour label principal « chélonien ».

Dans la configuration actuelle du script, son application aux 14 062 titres d'œuvres

du RETIF (y compris ceux des notices dépourvues d’images) a permis d’identifier 36 020 descripteurs, ce qui correspond à une moyenne de 2,56 termes par notice.

Le principal problème de cette méthode est qu’elle ne répond pas pleinement à l’objectif premier d’une indexation iconographique. En effet, elle n’améliore pas réellement la recherche dans la base de données, puisque les informations souhaitées sont déjà présentes dans les titres. Habituellement, lorsqu’un utilisateur recherche une représentation dans AGO-RHA, il utilise la recherche générale qui interroge les titres, plutôt que la recherche avancée reposant sur l’indexation Garnier, laquelle est plus difficile à exploiter. L’intérêt de ce script réside principalement dans le cadre de recherches quantitatives sur la fréquence de représentations de concepts issus du thésaurus ou pour constituer un ensemble de données utiles à des analyses statistiques. Faute de temps, le script est resté à un état inachevé afin de permettre l’exploration d’outils davantage centrés sur l’analyse des images.

Chapitre 5

Expérimenter les méthodes et proposer un protocole

En suivant les recommandations du cadre de planification des projets en intelligence artificielle de la Bibliothèque du Congrès, la deuxième étape correspond à l'expérimentation, qui succède à la phase de compréhension des données du projet¹. L'approche automatisée de l'indexation iconographique du RETIF n'a pas dépassé ce stade ni atteint la phase de planification, mais les tests effectués permettent néanmoins de dégager des pistes de pratiques à privilégier.

I. S'appuyer sur d'autres expériences

Afin de mieux comprendre quelles pratiques pourraient être pertinentes pour le corpus étudié, la première étape consiste à s'intéresser aux choix mis en œuvre dans d'autres projets similaires, en particulier les projets TORNE-H et Royère au Musée des Arts décoratifs, le projet Notre-Dame et le projet HikarIA. Parmi ces pistes figurait l'expérimentation du modèle YOLO (You Only Look Once), développé par Ultralytics, performant pour la détection et segmentation d'objets récurrents dans les images². Ce modèle est fréquemment mobilisé dans des projets de vision par ordinateur, comme le projet Torne-H, qui l'utilise pour analyser les œuvres graphiques du fonds Royère au Musée des Arts Décoratifs³. Dans notre cas, YOLO a été testé pour la reconnaissance d'objets dans les natures mortes. Plutôt que de mobiliser directement le corpus du RETIF, nous avons opté pour un corpus plus généraliste,

¹ Planification de projet IA, traduction de l'Artificial Intelligence Planning Framework de la Library of Congress...

² Ultralytics, YOLOv8 / YOLOv11, 2025, URL : <https://github.com/ultralytics/ultralytics> (visité le 31/08/2025).

³ E. Bermès et M. Charpier, « Repenser les collections patrimoniales par le prisme de l'IA 2025 »...

constitué d'images issues des collections du Musée du Louvre. Au total, 193 peintures ont été annotées à l'aide de 45 classes d'objets et d'animaux caractéristiques des natures mortes (par exemple : glass, fruits, musical instrument, dog, cat, knife, etc.). Le modèle retenu pour l'expérimentation était une version YOLOv11-small, entraînée sur 50 époques. Les résultats se sont toutefois révélés décevants : testé sur de nouvelles images, le modèle n'a pas été capable de reconnaître de manière fiable des objets similaires à ceux présents dans le jeu d'entraînement (Métriques : I. et I.). Plusieurs limites ont été identifiées : d'une part, le nombre de classes trop élevé compliquait significativement l'apprentissage et d'autre part, le volume d'images annotées était insuffisant pour permettre une répartition adéquate des objets et leur reconnaissance. Certaines classes, comme les instruments musicaux, présentaient une diversité de représentations trop importante. En définitive, l'approche d'utiliser YOLO s'est révélée inadaptée à notre problématique spécifique et a donc été abandonnée. Une autre approche pertinente est celle développée dans le cadre du projet Notre-Dame pour l'indexation de vues 2D et 3D. Cette méthodologie mobilise simultanément trois modèles d'intelligence artificielle, dont la manipulation demeure complexe et requiert à la fois des compétences d'ingénierie spécialisées et une infrastructure technique adaptée. Le processus s'appuie sur l'utilisation successive des modèles DINO, SAM et CLIP⁴. Le modèle DINO (Self-Distillation with No Labels), introduit en 2021 par Facebook AI Research⁵, constitue une approche d'apprentissage auto-supervisé appliquée aux images. Contrairement à des modèles supervisés tels que YOLO, qui nécessitent un jeu de données annoté par des experts humains, DINO apprend à identifier et à extraire des représentations visuelles sans recours à l'annotation manuelle. Il repose sur l'architecture des Vision Transformers (ViT), qui consistent à transformer les images en leur appliquant diverses modifications (recadrage, variations de couleurs, etc.), afin d'entraîner le modèle à reconnaître les parties significatives malgré leurs variations visuelles. Cette méthode permet à DINO d'isoler automatiquement des objets et des éléments au sein d'une image, sans entraînement spécifique à la segmentation. Une seconde version, publiée en 2023, a enrichi le modèle grâce à un apprentissage sur un corpus massif de 142 millions d'images et une amélioration de la technologie d'entraînement par distillation interne (*self-distillation*)⁶. Le modèle SAM (Segment Anything Model), développé par Meta en 2023, est conçu pour la segmentation d'images par génération de masques en réponse à

⁴L. De Luca, A. Guillem et K. Réby, « Intelligence humaine, collective et artificielle dans le chantier scientifique et de restauration de Notre-Dame de Paris »...

⁵Mathilde Caron, Hugo Touvron, Ishan Misra, Hervé Jégou, Julien Mairal, Piotr Bojanowski et Armand Joulin, *Emerging Properties in Self-Supervised Vision Transformers*, version 2, 2021, DOI : 10.48550/ARXIV.2104.14294.

⁶Maxime Oquab, Timothée Darcet, Théo Moutakanni, Huy Vo, Marc Szafraniec, Vasil Khalidov, Pierre Fernandez, Daniel Haziza, Francisco Massa, Alaaeldin El-Nouby, *et al.*, *DINOv2 : Learning Robust Visual Features without Supervision*, version 2, 2023, DOI : 10.48550/ARXIV.2304.07193.

des instructions sous forme de prompts, par exemple un clic, une boîte englobante ou une courte description textuelle⁷. Dans le cadre du projet Notre-Dame, DINO est utilisé pour vectoriser chaque élément d’une image de la cathédrale, par exemple une statue, un vitrail ou un détail architectural, afin d’identifier des similitudes visuelles, comme la représentation d’une même statue sous différents angles. SAM intervient alors en exploitant les régions d’attention détectées par DINO : celles-ci servent de prompt permettant à SAM de générer un masque précis. Ainsi, dans le cas d’une photographie de la Galerie des Rois, DINO identifie les silhouettes saillantes des statues par rapport à la niche dans laquelle elle se trouve, et SAM affine cette première analyse en produisant le contour exact de chaque figure. Enfin, une troisième étape mobilise le modèle CLIP, entraîné à associer représentations visuelles et langage naturel⁸. Cet outil multimodal permet notamment de relier les objets segmentés aux entrées d’un thésaurus, facilitant ainsi leur indexation dans une perspective documentaire. Dans d’une indexation automatisée du RETIF, l’association des modèles DINO, SAM et éventuellement CLIP pourrait constituer une solution prometteuse pour la détection automatique de motifs, de structures ou d’éléments visuels récurrents dans les peintures, et leur mise en correspondance avec un vocabulaire contrôlé restreint. Toutefois, cette approche n’a pas pu être expérimentée, en raison de la complexité de sa mise en œuvre, qui exige à la fois une expertise technique avancée et une capacité de calcul importante pour l’utilisation et l’articulation de ces modèles. Une troisième méthode, moins complexe à mettre en œuvre, consiste à exploiter les capacités des grands modèles de langage (LLM) pour générer des descriptions d’images. Cette approche a notamment été expérimentée dans le cadre du projet HikarIA. Pour décrire les photographies japonaises, trois modèles avaient alors été évalués : CLIP, Qwen2 et GPT-4, ce dernier s’étant distingué par la qualité de ses résultats. Grâce à l’application open source Arkindex, développée par la société TEKLIA, il a été possible d’accéder de manière simple à l’API de GPT-4 afin de traiter les images par l’intermédiaire du modèle de langage⁹. Près de deux cents images ont ainsi été décrites automatiquement, et les descripteurs générés ont été intégrés au site, avec une icône signalant qu’ils avaient été générés par un outil d’intelligence artificielle. Pour l’indexation iconographique du RETIF, le recours à des LLM avait d’ailleurs été envisagé dès les premières phases du projet. Toutefois, cette option suppose la capacité d’intégrer et de respecter un vocabulaire contrôlé.

⁷Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C. Berg, Wan-Yen Lo, *et al.*, *Segment Anything*, version 1, 2023, DOI : 10.48550/ARXIV.2304.02643.

⁸Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, *et al.*, *Learning Transferable Visual Models From Natural Language Supervision*, version 1, 2021, DOI : 10.48550/ARXIV.2103.00020.

⁹TEKLIA, *Arkindex, the document processing platform*, TEKLIA, URL : <https://teklia.com/our-solutions/arkindex/> (visité le 02/09/2025).

II. Intégrer le vocabulaire contrôlé au processus

L’intégration d’un vocabulaire contrôlé dans la chaîne de traitement constitue un défi majeur pour l’automatisation de l’indexation iconographique et conditionne le choix de la démarche adoptée. Les systèmes d’intelligences multimodaux génératifs offrent en effet la possibilité de produire des descriptions textuelles à partir d’images, lesquelles pourraient ensuite être exploitées par des scripts pour extraire les termes pertinents. Le script présenté au chapitre 3, conçu pour détecter les concepts du thésaurus Garnier dans les titres, pourrait ainsi être appliqué aux textes générés par les modèles. Toutefois, ce processus nécessite d’avoir des textes en français. Or, les modèles téléchargeables localement, tels que LLA-VA ou BLIP-3, bien qu’ils évitent le recours à des API coûteuses comme celles d’OpenAI pour les modèles GPT, demeurent nettement plus performants en anglais qu’en français. Plusieurs pistes peuvent être envisagées pour surmonter cette barrière linguistique. Une première consiste à recourir à un système de traduction performant, tel que DeepL, afin de convertir automatiquement les textes produits en français et d’y rechercher les termes du vocabulaire contrôlé¹⁰. Une deuxième option repose sur la traduction du thésaurus Garnier lui-même, afin de le rendre bilingue. Cette démarche pourrait être prolongée par une approche plus ambitieuse consistant à aligner les concepts du thésaurus avec ceux de Wikidata. Un tel alignement offrirait une interopérabilité accrue, en reliant chaque descripteur Garnier à un équivalent conceptuel multilingue plutôt qu’à une simple traduction littérale. Il permettrait également de croiser les données avec celles de Wikidata, notamment les peintures déjà labellisées, et d’ouvrir la voie à des rapprochements avec d’autres vocabulaires contrôlés tels qu’Iconclass, dont 4 215 concepts sont référencés dans Wikidata, ou le Getty Art and Architecture Thesaurus (AAT), dont 24 877 termes sont reliés¹¹. Une troisième piste réside dans le développement d’un outil de type RAG (Retrieval-Augmented Generation), reposant sur le thésaurus Garnier comme base de données, afin de traduire efficacement et de générer des résultats précis et contextualisés. Un tel modèle combine deux étapes : suite à un prompt, il interroge d’une base de données pour identifier les informations pertinentes, puis mobilise ces informations à l’aide d’un modèle génératif de langue qui produit une réponse en fonction des données des documents sources et du prompt. Dans ce cadre, un RAG pourrait analyser un texte en anglais, en identifier les correspondances pertinentes dans le thésaurus Garnier et restituer une liste des descripteurs. L’avantage de cette approche est que le modèle ne s’appuie pas uniquement sur ses paramètres internes liés à son entraînement, mais sur une base de données externe, ce qui garantit une contextualisation plus fine. Cependant, la mise

¹⁰DeepL, *Translate Text*, DeepL Documentation, URL : <https://developers.deepl.com/api-reference/translate> (visité le 02/09/2025).

¹¹Consulter le chapitre 2 à ce sujet.

en œuvre de ces solutions se heurte à plusieurs limites. L’alignement du thésaurus Garnier avec Wikidata, bien que porteur de perspectives d’interopérabilité, est un processus extrêmement long et exigeant : le thésaurus comporte plus de 16 000 concepts, et l’alignement, pour des raisons de qualité, doit être réalisé manuellement, ce qui excède le temps imparti au projet. De même, la création d’un RAG pleinement opérationnel dans une chaîne de traitement requiert à la fois des compétences d’ingénierie numérique élevées et des ressources de calcul importantes. Lors du projet, un test mené avec un modèle Mistral 7b s’est révélé insatisfaisant, sans doute en raison d’une méthodologie inadaptée ou de la puissance limitée du modèle¹². Pour pallier ces contraintes, nous avons eu recours à une solution alternative et gratuite proposée par Google Labs : NotebookLM¹³. Cet outil disponible en ligne depuis 2023 associe des documents textuels à un modèle Gemini et permet de les interroger aisément¹⁴. Il s’est montré particulièrement performant pour explorer le thésaurus et identifier des équivalences. Par exemple, à partir du terme “Cléopâtre”, il parvient à proposer l’entrée précise “Cléopâtre VII” du thésaurus Garnier. Toutefois, les résultats doivent être nuancés dans de très rares cas, puisque NotebookLM tend parfois à surinterpréter. Ainsi, face à la requête “une peinture d’une femme et son enfant”, il peut parfois proposer le terme “Vierge à l’Enfant”, en se fondant sur la fréquence de ce motif iconographique. Enfin, bien que très utile pour l’exploration et l’évaluation des modèles multimodaux dans la description des images, NotebookLM ne permet pas d’industrialiser la tâche : les requêtes doivent être effectuées une par une, ce qui limite son usage dans une chaîne automatisée de traitement.

III. Tester et comparer les performances des modèles

Afin d’évaluer les performances d’un ensemble de modèles, ceux-ci ont été associés à un RAG développé avec NotebookLM, intégrant la liste des termes du thésaurus Garnier, et testés sur un échantillon d’images provenant du RETIF. Cette expérimentation avait pour objectif principal de mesurer l’efficacité des modèles et d’envisager la conception d’un système RAG adapté au projet, ou, le cas échéant, de privilégier d’autres méthodes pour l’extraction des termes du vocabulaire contrôlé. L’échantillon utilisé comprenait trente images issues du corpus RETIF, couvrant l’ensemble de la période étudiée, du XIII^e siècle au début du XX^e siècle, et représentant une diversité de genres artistiques : paysages, portraits, iconographie

¹²Mistral AI, *Mistral 7B*, Mistral.ai, 27 sept. 2023, URL : <https://mistral.ai/fr/news/announcing-mistral-7b> (visité le 02/09/2025).

¹³Google Labs, *Google NotebookLM*, URL : <https://notebooklm.google/> (visité le 02/09/2025).

¹⁴Raiza Martin et Steven Johnson, *Google NotebookLM*, Google, 12 juill. 2023, URL : <https://blog.google/technology/ai/notebooklm-google-ai/> (visité le 02/09/2025).

religieuse, scènes bibliques ou mythologiques, ainsi que natures mortes¹⁵. Les œuvres ont été choisies de façon à représenter à la fois des œuvres facilement identifiables par les modèles, provenant de grandes collections telles que le Louvre, et des images plus difficiles à reconnaître, comme des copies, des reproductions photographiques en noir et blanc ou des compositions à l’iconographie complexe. Chaque œuvre a ensuite été indexée iconographiquement à l’aide du thésaurus Garnier de l’INHA¹⁶.

Quatre modèles de génération de langage naturel ont été retenus, en majorité des grands modèles de langue : GPT-4 (OpenAI)¹⁷, Mistral Medium 3¹⁸, LLaVA¹⁹ et BLIP²⁰. À chacun d’eux a été soumis le même prompt : “Identify the scene and write an iconographical description of this painting”. Celui-ci a été formulé en anglais, puisque certains modèles, comme LLaVA et BLIP, se révèlent moins performants lorsqu’ils sont sollicités en français.

L’outil NotebookLM était utilisé pour l’extraction des termes du thésaurus. Le prompt utilisé était : “Donne une liste séparée par des virgules des termes du thésaurus présents dans cette description”, suivi de la description en anglais générée par l’un des modèles. L’outil identifiait ainsi les correspondances entre les termes anglais de la description et les descripteurs du thésaurus, fournissant un liste de termes présents dans la description. La liste obtenue a ensuite été confrontée à la vérité terrain, afin d’évaluer la pertinence et la précision de l’analyse.

Modèle	TP moy	FP moy	FN moy	Précision moy	Rappel moy	F1 moy
GPT-4	5.267	23.7	6.8	0.192	0.469	0.261
Mistral M3	4.2	19.967	7.867	0.184	0.357	0.237
LLAVA	2.733	27.467	9.333	0.096	0.254	0.134
BLIP	0.867	3.2	11.2	0.229	0.085	0.116

TAB. 5.1 : Comparaison des modèles pour l’indexation iconographique en passant par NotebookLM selon les moyennes des TP, FP, FN, précision, rappel et F1.

Ce tableau montre les statistiques de chaque modèle pour décrire les trente images et obtenir les termes du thésaurus Garnier. Il présente une moyenne des vrais positifs, des faux positifs, des faux négatifs, de la précision, du rappel et du score F1 pour chaque modèle. Les

¹⁵La liste des œuvres est indiquées en annexes C.1

¹⁶Nous remercions Pierre-Yves Laborde pour avoir participé à la réalisation de cette vérité terrain.

¹⁷OpenAI, *GPT-4*, OpenAI, URL : <https://openai.com/fr-FR/index/gpt-4/> (visité le 02/09/2025).

¹⁸Mistral AI, *Medium is the new large*, URL : <https://mistral.ai/fr/news/mistral-medium-3> (visité le 02/09/2025).

¹⁹Haotian Liu, Chunyuan Li, Qingyang Wu et Yong Jae Lee, *Visual Instruction Tuning*, version 2, 2023, DOI : 10.48550/ARXIV.2304.08485.

²⁰Junnan Li, Dongxu Li, Caiming Xiong et Steven Hoi, *BLIP : Bootstrapping Language-Image Pre-training for Unified Vision-Language Understanding and Generation*, version 2, 2022, DOI : 10.48550/ARXIV.2201.12086.

vrais positifs correspondent au nombre de mots correctement prédits par image. Les faux positifs sont le nombre de mots prédits absents de la vérité terrain. Les faux négatifs sont les mots manquants, présents dans la vérité terrain mais pas prédits par un modèle. La précision est la proportion de mots prédits corrects parmi tous les mots prédits. Le rappel est la proportion de mots corrects retrouvés parmi les mots vrais, issus de la vérité terrain. Enfin le score F1 est une moyenne harmonique entre la précision et le rappel : elle permet de donner une mesure globale de la qualité de prédiction d’un modèle. Dans l’ensemble, ces combinaisons de modèles donnent de mauvais résultats. À première vue, les scores F1 sont très bas, puisque tous inférieurs à 0.7 qui serait un score F1 satisfaisant pour un modèle.

À part le modèle BLIP, l’ensemble des solutions donnent un nombre de faux positifs très importants, supérieurs à 20, soit presque le double de la moyenne de mots de la vérité terrain qui s’élève à 12,07. Cela impliquerait sûrement des corrections pour supprimer les mots supplémentaires créés par les modèles.

IV. Adapter l’automatisation aux spécificités de l’histoire de l’art

D’autres approches ont été explorées, ne cherchant pas à mobiliser l’ensemble des termes du corpus, mais à se concentrer sur une fraction spécifique pour accomplir une tâche précise. Comme le souligne David G. Stork, les modèles d’IA traditionnels, y compris les grands modèles de langues multimodaux, ont surtout été développés pour analyser des photographies, des vidéos ou des images spécialisées, notamment les radiographies médicales²¹. Dans le cas des peintures, ces modèles présentent deux principales difficultés : d’une part, les caractéristiques formelles des œuvres diffèrent de la réalité photographique, et d’autre part, les datasets d’entraînement sont souvent trop restreints, particulièrement pour des iconographies rares. Ainsi, l’entraînement de modèles pour réaliser des tâches de reconnaissance ciblées pourrait permettre de développer une approche mieux adaptée aux corpus d’histoire de l’art. Nous n’avons pas identifié de modèles spécialisés capables de reconnaître efficacement des motifs ou des scènes dans les peintures de l’Époque moderne pour notre corpus. L’outil NADA (No Annotations for Object Detection in Art), développé pour reconnaître des éléments iconographiques dans des tableaux, aurait pu être pertinent, mais sa complexité l’a rendu difficile à exploiter²². Comme mentionné précédemment, la technologie de segmentation YOLO a été

²¹D. G. Stork, « Computer Vision, ML, and AI in the Study of Fine Art », *Communications of the ACM*, 67–5 (mai 2024), p. 68–75, DOI : 10.1145/3633454, p. 72-73.

²²Patrick Ramos, Nicolas Gonthier, Selina Khan, Yuta Nakashima et Noa Garcia, *No Annotations for Object Detection in Art through Stable Diffusion*, version 2, 2024, DOI : 10.48550/ARXIV.2412.06286.

testée pour reconnaître des objets dans les natures mortes. Bien que cette expérience n’ait pas produit de résultats concluants, le modèle n’est pas à écarter complètement : d’autres applications restent envisageables, comme la détection de personnes ou l’extraction de la peinture de son cadre lorsque la reproduction inclut également celui-ci.

Une autre expérience utilisant un modèle plus simple de classification a donné de meilleurs résultats. L’objectif était d’identifier les genres picturaux au sein des images du corpus, afin de distinguer les natures mortes, paysages et portraits des scènes, elles-mêmes regroupant les scènes religieuses, de genre et mythologiques. Un corpus d’entraînement de 1 495 peintures du RETIF a été constitué, réparti en quatre classes : 258 natures mortes, 227 paysages, 384 portraits et 626 scènes. L’outil utilisé est la bibliothèque Python FastAI, construite sur PyTorch, conçue pour faciliter le développement et la manipulation de modèles d’apprentissage profond²³. La partie FastAI Vision de cette bibliothèque est dédiée au traitement d’images, notamment la classification, et sa simplicité d’utilisation permet des expérimentations rapides. Le modèle s’appuie sur le réseau de neurones convolutif ResNet, pré-entraîné sur le dataset ImageNet de 1,2 million d’images, ce qui lui permet de détecter des motifs visuels génériques tels que les contours, textures et formes²⁴. Lors du réentraînement du modèle de classification, des transformations sont appliquées automatiquement aux images (rotation, recadrage, inversion miroir, variation de luminosité) afin d’améliorer sa robustesse et la reconnaissance des classes qu’il doit identifier. Après seulement dix époques d’entraînement, le modèle a été évalué sur le corpus de validation, donnant des résultats très encourageants :

Exactitude	Précision	Rappel	F1-score
0.8428	0.8386	0.8432	0.8336

TAB. 5.2 : Performances du modèle de classification des genres picturaux

Ces statistiques montrent que le modèle est performant pour la reconnaissance des genres picturaux, avec un taux d’erreur relativement faible (Matrice de confusion : II.). Ce modèle peut ainsi constituer une première étape prometteuse pour l’indexation iconographique des œuvres. D’autres modèles avaient également été envisagés, notamment FLAVA (A Foundational Language And Vision Alignment Model), conçu pour établir des correspondances plus fines entre les images et leurs catégories de représentation. Développé par Facebook AI Research (FAIR) et présenté en 2021, ce modèle a été entraîné sur près de 70 millions de paires texte-image²⁵. Il est notamment capable de prédire si une image et un

²³ Documentation de FastAI, URL : <https://docs.fast.ai/> (visité le 20/08/2025).

²⁴Kaiming He, Xiangyu Zhang, Shaoqing Ren et Jian Sun, *Deep Residual Learning for Image Recognition*, version 1, 2015, DOI : 10.48550/ARXIV.1512.03385.

²⁵Amanpreet Singh, Ronghang Hu, Vedanuj Goswami, Guillaume Couairon, Wojciech Galuba, Marcus

texte correspondent en mobilisant trois encodeurs distincts, chacun dédié respectivement aux tâches de vision par ordinateur, de traitement du langage et aux tâches multimodales. Son utilisation avait été envisagée dans le cadre de notre étude pour exploiter les titres des images, en particulier pour les scènes à iconographie récurrente, telles que les scènes bibliques, les épisodes de l’Ancien Testament, de la vie de la Vierge ou de celles du Christ. La manipulation du modèle n’a pas pu être expérimentée, faute de temps et de compétences techniques.

Ces expériences, bien que parfois peu concluantes, ouvrent la voie à une première ébauche de protocole méthodologique adapté au corpus pour réaliser l’automatisation de son indexation iconographique. Une phase initiale pourrait consister à déterminer les genres iconographiques à l’aide du modèle de classification des images, avant d’affiner l’analyse par des outils plus performants en fonction de la complexité des représentations. Ainsi, les scènes, plus complexes à interpréter, pourraient faire l’objet d’un traitement spécifique, différent de celui appliqué aux natures mortes, où l’enjeu principal se limite à l’identification des objets. L’utilisation des grands modèles de langue pourrait alors être réservée aux cas les plus complexes, avec extraction des termes pertinents du thésaurus soit via un outil de recherche automatique dans le texte généré, soit au moyen d’une approche de génération augmentée par récupération (RAG).

Chapitre 6

Bénéfices et limites de l'automatisation dans un contexte institutionnel

L'automatisation constitue donc un véritable défi, car elle exige d'adapter les pratiques de l'histoire de l'art aux outils numériques. Si elle offre des perspectives prometteuses pour le traitement de vastes corpus, elle requiert néanmoins des ressources considérables, tant humaines que matérielles, à chaque étape du processus.

I. Des bénéfices non négligeables : traitement de masse, reproductibilité et transférabilité des méthodes

L'automatisation offre des avantages qui justifient son recours dans le cadre des collections patrimoniales. Elle permet avant tout de répondre à la contrainte de temps considérable nécessaire à l'étude et à l'inventaire des fonds. À titre d'exemple, dans le cadre des projets Royère et Torne-H, qui mobilisent des outils de vision par ordinateur pour analyser les collections du Musée des Arts Décoratifs, la réalisation manuelle de l'inventaire et de la description des près de 700 000 œuvres non encore répertoriées de la collection de dessins, papiers peints et photographies a été estimée à 643 ans¹. L'intelligence artificielle apparaît ainsi comme une solution pour accélérer le travail des conservateurs et documentalistes. Elle contribuerait à rendre les collections plus accessibles et à faciliter la consultation et l'exploration des bases de données. Un autre bénéfice réside dans la transférabilité des méthodes développées. Les techniques d'automatisation mises en œuvre dans un projet donné peuvent être adaptées à

¹E. Bermès et M. Charpier, « Repenser les collections patrimoniales par le prisme de l'IA 2025 »...

d’autres contextes institutionnels. C’est le cas du projet Torne-H, mais également du projet HikarIA, mené conjointement par le Musée national des Arts Asiatiques – Guimet et la société Teklia. L’objectif est d’étendre le traitement au-delà de la collection Dubois, afin d’inclure l’ensemble des albums de photographies japonaises anciennes conservées au musée, et avec la perspective de transposer la démarche à d’autres collections comparables dans les institutions patrimoniales au Japon².

Dans le cadre du projet de l’INHA consacré à l’indexation iconographique du RETIF, le recours à des méthodes automatisées offrirait la possibilité de transférer ce processus vers d’autres bases de données de peintures d’AGORHA insuffisamment indexées. Ainsi, la méthodologie développée dans le cadre de ce projet pourrait être appliquée à des corpus tels que l’Iconographie musicale : répertoire d’œuvres d’art à sujets musicaux publiées par Albert Pomme de Mirimonde³, le Répertoire des tableaux français en Allemagne (XVII^e et XVIII^e siècles) (REPFALL)⁴, le Recensement de la peinture produite en France au XVI^e siècle⁵, ou encore Les Envois de Rome en peinture et sculpture, 1804-1914⁶. L’automatisation offre ainsi les perspectives d’économie de temps et de transférabilité des méthodes, dont les apports seraient significatifs pour l’étude des collections patrimoniales.

II. Des limites institutionnelles et techniques

Cependant, l’automatisation d’un tel processus soulève des enjeux importants, tant institutionnels, organisationnels que infrastructurels, auxquels le projet de l’INHA a dû faire face. Contrairement à d’autres projets similaires, une équipe pluridisciplinaire dédiée n’avait pas été constituée pour travailler exclusivement sur ce projet. L’ensemble du travail reposait principalement sur une seule personne, soutenue ponctuellement par les ingénieurs du Service numérique de la recherche de l’INHA et par des compétences techniques issues de PictorIA. Un croisement plus important des compétences, combinant une expertise iconographique approfondie du corpus étudié et des savoir-faire en vision par ordinateur, aurait été nécessaire, tant pour la préparation du projet que pour le déroulement des expérimentations. Si quatre mois ont permis d’obtenir des premiers résultats, les expériences d’autres projets français ont révélé la nécessité de mettre en œuvre des périodes de préparation, de test et de manipulation des modèles beaucoup plus longues pour atteindre des résultats probants.

²C. Kermorvant et É. de Saint-Ours, *Le projet HikarIA...*

³Institut national d’histoire de l’art, *Iconographie musicale : répertoire d’œuvres d’art à sujets musicaux publiées par Albert Pomme de Mirimonde...*

⁴Id., *Répertoire des tableaux français en Allemagne (XVIIe et XVIIIe siècles)*, REPFALL, URL : <https://agorha.inha.fr/database/5> (visité le 01/09/2025).

⁵Id., *Recensement de la peinture produite en France au XVIe siècle...*

⁶Id., *Les Envois de Rome en peinture et sculpture, 1804-1914...*

À titre d'exemple, après un an d'expérimentations avec un modèle YOLO pour étudier le fonds de dessins du décorateur Jean Royère (1902-1981), Emmanuelle Bermès et Marion Charpier notent que « le degré de précision que nous recherchons pour répondre aux attentes du personnel du musée n'est pas encore atteint »⁷. Un autre facteur limitant concerne les infrastructures informatiques, en grande partie liées à la puissance de calcul requise pour entraîner et déployer des modèles de vision par ordinateur. Les processeurs graphiques, ou GPU (pour *Graphical processor unit*) initialement conçus pour le traitement massif d'images et de graphiques dans les ordinateurs et consoles de jeu, sont aujourd'hui indispensables pour l'intelligence artificielle, en particulier pour les réseaux de neurones profonds, qui impliquent des calculs intensifs de matrices et vecteurs. Ces architectures permettent des traitements parallèles massifs, nécessaires pour entraîner des modèles complexes. De plus en plus d'institutions menant des projets en sciences humaines se dotent de processeurs graphiques pour leurs travaux de recherche. Par exemple, l'École des Chartes et le MediaLab de SciencesPo⁸, ainsi que l'Université Paris-Cité⁹, disposent de ces ressources, tout comme l'infrastructure de recherche Huma-Num, qui met des infrastructures à disposition¹⁰. Ces équipements restent cependant coûteux, tant pour leur acquisition que pour leur consommation électrique. À l'Institut national d'histoire de l'art, à l'exception d'un ordinateur équipé d'une carte graphique Nvidia 4080 au sein du Service numérique de la recherche, aucune ressource comparable n'est disponible pour l'intelligence artificielle. Aucun projet de l'institution n'intègre actuellement de telles technologies, à l'exception du projet Gallica Images, mené en collaboration avec la Bibliothèque nationale de France et la Bibliothèque nationale universitaire de Strasbourg¹¹. Pour les expérimentations réalisées dans le cadre du projet d'indexation du RETIF, l'ordinateur disponible et les GPU gratuits proposés par Google Colab ont suffi, mais ils ne permettraient pas une mise à grande échelle d'un traitement automatisé¹². Pour des projets futurs nécessitant des infrastructures plus conséquentes, il apparaît donc préférable de privilégier des collaborations avec des institutions déjà équipées, plutôt que d'acquérir du matériel qui serait peu utilisé.

⁷E. Bermès et M. Charpier, « Repenser les collections patrimoniales par le prisme de l'IA 2025 »...

⁸Ces informations m'ont été communiquées par Mathieu Taybi et Lilla Conte.

⁹Université Paris Cité, *Plateforme de Calcul Haute-Performance NOVA*, Plateformes & plateaux - Université Paris Cité, URL : <https://plateformes.u-paris.fr/plateforme-de-calcul-haute-performance-nova/> (visité le 01/09/2025).

¹⁰Huma-Num, *Infrastructure multi-cibles pour l'usage de l'IA en sciences humaines et sociales*, Documentation de l'infrastructure Huma-Num, 27 août 2025, URL : <https://documentation.huma-num.fr/infrastructure-IA/> (visité le 01/09/2025).

¹¹Bibliothèque nationale de France, *Les projets en intelligence artificielle à la BnF*, BnF - Site institutionnel, URL : <https://www.bnf.fr/fr/les-projets-en-intelligence-artificielle-la-bnf> (visité le 01/09/2025).

¹²Google, *Bienvenue dans Colab*, URL : <https://colab.research.google.com/?hl=fr> (visité le 01/09/2025).

III. L'impact environnemental

Depuis le déploiement des grands modèles de langage auprès du grand public, leur impact environnemental a rapidement été souligné. Dans ce contexte, des outils sont apparus, tels que Compar :IA, conçu pour mesurer l'empreinte écologique des requêtes adressées aux agents conversationnels¹³. Parallèlement, les premières études ont commencé à être publiées afin de mieux documenter ces effets. Le 22 juillet 2025, la société française Mistral AI publiait une étude intitulée *textitOur contribution to a global environmental standard for AI*, menée en collaboration avec le cabinet de conseil Carbone 4 et l'Agence de l'environnement et de la maîtrise de l'énergie (ADEME)¹⁴. Cette recherche visait à quantifier l'empreinte environnementale des grands modèles de langage (LLM) selon trois critères principaux : les émissions de gaz à effet de serre, la consommation d'eau et l'épuisement des ressources non renouvelables. L'analyse s'appuie sur le suivi du modèle Mistral Large 2, utilisé pendant 18 mois jusqu'en janvier 2025. Les résultats font état de 20,4 kilotonnes de CO₂ consommés pour l'entraînement du modèle. Rapporté aux émissions annuelles moyennes par habitant en France en 2023 (9,4 tonnes), cela équivaut aux émissions de plus de 2 000 Français en un an¹⁵. Pour la consommation d'eau, l'étude chiffre à volume utilisé de 281 000 m³, soit l'équivalent de la consommation domestique annuelle de plus de 5 200 habitants (sur la base d'une moyenne de 54 m³ par personne en 2020)¹⁶. L'essentiel de cette empreinte provient de la phase d'entraînement des modèles, responsable de 85,5 % des émissions de CO₂ et de 91 % de la consommation d'eau potable. À l'inverse, l'impact d'une requête d'inférence reste marginal : environ 1,14 g de CO₂ et 45 mL d'eau par demande. Toutefois, l'accumulation de requêtes à des grands modèles de langue, par exemple appliquées aux 11 974 images du RETIF, peut représenter une dépense énergétique notable, d'autant plus lorsque les résultats peuvent nécessiter une curation ou une analyse complémentaire. L'impact environnemental de ces outils est donc non négligeable et doit être pris en compte lors de l'utilisation de ces outils.

¹³Ministère de la Culture, *compar :IA, Le comparateur d'IA conversationnelles*, oct. 2024, URL : <https://www.comparia.beta.gouv.fr/> (visité le 31/08/2025).

¹⁴Mistral AI, *Our contribution to a global environmental standard for AI*, 22 juill. 2025, URL : <https://mistral.ai/news/our-contribution-to-a-global-environmental-standard-for-ai> (visité le 31/08/2025).

¹⁵Insee, *Émissions de gaz à effet de serre par habitant*, Tableau de bord de l'économie française, 6 janv. 2025, URL : https://www.insee.fr/fr/outil-interactif/5367857/details/90_DDE/92_DEV/92G_Figure7 (visité le 30/08/2025).

¹⁶Commissariat général au développement durable, *Consommation domestique en eau potable*, notre-environnement, 11 janv. 2023, URL : <https://www.notre-environnement.gouv.fr/themes/societe/le-mode-de-vie-des-menages-ressources/article/consommation-domestique-en-eau-potable> (visité le 31/08/2025).

Si l’automatisation du processus peut être envisagée pour certaines tâches de reconnaissance, les tests menés ont mis en évidence de nombreuses difficultés, mobilisant un temps conséquent d’expérimentation ainsi que d’importants moyens techniques et humains afin d’établir un protocole. Si ces expériences ont permis de formuler des hypothèses pour l’analyse d’images intégrant en intégrant un vocabulaire contrôlé, d’autres approches peuvent être envisagées à l’automatisation afin de tirer pleinement parti des systèmes de vision par ordinateur dans l’indexation iconographique.

Troisième partie

Alternatives à l'automatisation : des outils d'intelligence artificielle au service des pratiques d'indexation iconographiques

Chapitre 7

Des solutions hybrides pour faciliter l'indexation iconographique plutôt que l'automatiser

I. Des logiciels pour explorer et annoter le corpus

Face aux limites évoquées tout au long de la partie précédente, il est nécessaire d'envisager des alternatives à l'automatisation de l'indexation iconographique des images. Plusieurs solutions logicielles, développées au cours des dernières années, intègrent désormais des algorithmes d'intelligence artificielle capables de repérer des similarités visuelles et de faciliter la navigation au sein de vastes corpus iconographiques. Parmi elles figurent notamment textit-PixPlot et textitPanoptic. L'un des outils les plus largement utilisés pour explorer de grandes bases de données d'images est textitPixPlot¹. Ce logiciel open source, développé en 2017 par le textYale Digital Humanities Lab de l'université de Yale, s'inspire des travaux de Bénoît Séguin, d'Isabella di Lenardo et de Frédéric Kaplan, qui ont appliqué des réseaux neuronaux convolutifs (textitConvolutional Neural Networks, CNN) à l'analyse d'images patrimoniales². PixPlot permet de visualiser et d'explorer un corpus iconographique en disposant les images dans un espace en deux dimensions, en fonction de leurs similarités visuelles détectées par les algorithmes. L'outil offre ainsi de nouvelles perspectives de consultation : il permet de repérer des textitclusters, c'est-à-dire des groupes d'images similaires, de mettre en évidence des relations et d'identifier des régularités au sein d'un ensemble d'images. En revanche, son utilisation suppose une configuration matérielle relativement puissante, si possible une carte graphique dédiée, et ses possibilités restent limitées en matière d'enrichissement des

¹DHLab (Yale), *PixPlot*, URL : <https://dhlab.yale.edu/projects/pixplot/> (visité le 30/08/2025).

²I. di Lenardo, B. Séguin et F. Kaplan, « Visual Patterns Discovery in Large Databases of Paintings »...

images par des métadonnées³. À part le nom du fichier de l'image numérique, aucune donnée supplémentaire ne peut être ajoutée pour décrire les images.

Le second outil, Panoptic, est une application Python développée par le Centre d'expérimentation en méthodes numériques pour les recherches en sciences humaines et sociales (Ceres), unité de service de la Faculté des Lettres de Sorbonne Université⁴. Contrairement à PixPlot, Panoptic met davantage l'accent sur l'association et la gestion des métadonnées, tout en simplifiant la visualisation des images. Le logiciel repose sur le modèle d'apprentissage automatique CLIP, entraîné à relier texte et image, et capable d'identifier des correspondances visuelles en fonction de leur contenu. Grâce à ce modèle, Panoptic peut former des clusters, détecter des images similaires et même permettre des recherches textuelles directement au sein du corpus iconographique. Lorsqu'on sélectionne une image, ses métadonnées apparaissent et l'outil présente les images proches, ordonnées selon leur degré de similarité. L'application propose en outre une interface pour enrichir les images par différents types de métadonnées (étiquettes, texte libre, nombres, dates, couleurs, liens ou cases à cocher). Enfin, à la différence de PixPlot, Panoptic ne requiert pas de puissance de calcul particulière et peut fonctionner sur des ordinateurs dépourvus de carte graphique.

II. Explorer le corpus du RETIF grâce à ces logiciels

Les 11 974 images du RETIF ont été intégrées dans les deux logiciels présentés précédemment, PixPlot et Panoptic, afin d'en évaluer l'intérêt pour l'étude du corpus.

Avec PixPlot, l'ensemble des peintures peut être visualisé dans un espace en deux dimensions, ce qui permet à la fois d'appréhender globalement la composition du corpus et de repérer des regroupements d'images similaires. La disposition spatiale obtenue dans le logiciel reflète les calculs de similarité du modèle utilisés : plus deux images sont rapprochées dans l'espace, plus le modèle les juge proches, et inversement pour les images éloignées (figure D.1). Des groupes sont ainsi identifiables aussi du corpus. Dans la partie inférieure gauche de la visualisation apparaissent les peintures représentant des figures isolées ou de petits groupes, en particulier les portraits et les images de saints (figure D.2). Les paysages, quant à eux, se répartissent dans la partie supérieure : les vues naturelles plutôt au centre, tandis que les vues urbaines – notamment les /textitvedute vénitiennes et les caprices architecturaux peuplés de ruines – se répartissent en direction de l'angle supérieur droit (figure D.3). Les

³Douglas Duhaime et Peter Leonard, *PixPlot* (Dépôt Github), Github, 2018-2023, URL : <https://github.com/pleonard212/pix-plot>.

⁴Édouard Bouté, Virginie Julliard, David Gödicke, Fred Pailler et Victor Ecrement, « PANOPTIC, un outil d'exploration par similarité de vastes corpus d'images », dans *Humanistica 2024, Association francophone des humanités numériques*, Meknès, Maroc (7-9 mai 2024), 2024, URL : hal-04687627.

scènes mythologiques et religieuses occupent principalement la zone centrale, surtout lorsqu’elles mettent en scène un grand nombre de personnages(figure D.4. Cependant, plusieurs limites de l’analyse des images apparaissent. Dans le coin supérieur gauche de la visualisation, l’algorithme regroupe toutes les photographies en noir et blanc, les isolant des peintures auxquelles elles pourraient être rapprochées visuellement (figure D.5). Sur le bord droit, les photographies incluant le cadre de la peinture brouillent également la classification. Toutefois, les panneaux de polyptyques et les retables photographiés dans leur contexte, églises ou chapelles par exemple, y sont aisément repérables (figure D.6). Enfin, un cluster isolé, situé dans l’angle inférieur droit, rassemble les œuvres de formats particuliers (circulaires, hexagonaux ou ovales), en particulier les tondi ou les *modelletti* de décors plafonnants (figure D.7). Ces résultats montrent que PixPlot est pertinent pour dégager des regroupements morphologiques au sein du corpus, mais que l’algorithme privilégie surtout des critères formels, notamment le format, les couleurs, et la qualité de la photographie, au détriment de la proximité iconographique. Ainsi, des peintures au contenu très proche peuvent se retrouver séparées en raison de différences liées à la prise de vue. Une harmonisation préalable des images, en convertisseur les images en noir et blanc, ou en suppression des cadres, aurait sans doute amélioré les résultats, mais au prix d’un traitement long et d’une lisibilité réduite. Enfin, l’absence de possibilité d’enrichir les images par des métadonnées limite fortement les perspectives d’analyse approfondie et d’indexation.

La même expérimentation a été réalisée avec le logiciel Panoptic. Grâce à son algorithme de clustering, l’outil parvient à identifier, au sein des 11 974 images du RETIF, des regroupements similaires à ceux obtenus avec PixPlot. Grâce à l’outil de génération de clusters, le logiciel répartit les peintures entre différents genres iconographiques avec un taux d’erreur relativement faible. Ainsi, certains regroupements se composent exclusivement de natures mortes, de portraits, de paysages naturels ou urbains, tandis que d’autres rassemblent des scènes plus complexes. Le premier cluster qui se détache est celui des Vierges à l’Enfant, reflet de l’importance majeure de cette iconographie dans le corpus(figure D.8). Lors de l’encodage des images par le modèle CLIP, Panoptic offre la possibilité de traiter les fichiers en couleur ou en noir et blanc. Cette flexibilité permet soit de tirer parti des propriétés chromatiques, utiles notamment pour le manteau bleu caractéristique de la Vierge ou l’habit pourpre de cardinal de saint Jérôme, soit de regrouper des images malgré des différences de reproduction photographique, en particulier pour les photographies en noir et blanc. De plus, l’intégration des métadonnées issues d’AGORHA, les titres des œuvres, dates de création, noms des artistes auteurs des modèles, constitue un atout majeur. Exploitables par le moteur de recherche ou comme critères de filtrage, cette fonctionnalité enrichit l’analyse en facilitant la navigation dans le corpus et en permettant de croiser les regroupements visuels avec les informations

documentaires préexistantes.

III. Test de *Panoptic* pour l’indexation et protocole d’utilisation

La possibilité d’associer de nouvelles métadonnées aux images dans l’application *Panoptic* a offert l’opportunité de réaliser un test en vue d’une future campagne d’indexation iconographique, s’appuyant sur un modèle d’intelligence artificielle tel que CLIP. Pendant environ cinquante heures de travail, nous avons exploré l’usage de cet outil pour cette tâche, afin d’évaluer son efficacité et de rédiger un protocole pour conduire une telle campagne. Dans un premier temps, une version facilement navigable du thésaurus a été préparée au format texte (txt), utilisant l’indentation pour restituer la hiérarchie des termes. Cette solution présentait moins de contraintes que l’importation des plus de 16 000 termes du thésaurus directement dans une propriété « étiquettes » (tags) de l’application, qui entraînait des problèmes de chargement. Les 11 974 images du corpus ont ensuite été importées dans l’application, avec les noms des artistes et les titres des œuvres. Dans *Panoptic*, plusieurs « propriétés » ont été créées pour permettre le remplissage d’étiquettes (« tags ») selon des catégories définies par l’indexeur : « scènes » pour les scènes représentées, une propriété pour le lieu lorsqu’il est identifié (notamment Rome, Venise ou Jérusalem), et d’autres pour différents types d’éléments présents, tels que « personnages », « animaux », « objets » ou « éléments de paysage ». À l’intérieur de ces différentes propriétés étaient ensuite ajoutés les termes du thésaurus sous forme d’étiquettes au fur et à mesure de l’indexation, par exemple les saints dans la propriété « personnages » ou les différents animaux représentés dans la propriété correspondante. Ces catégories facilitaient la localisation des termes du thésaurus associés à chaque image et permettaient un filtrage efficace des images. Grâce à la fonction de clustering, une grande partie des genres picturaux ou des « caractères généraux de la représentation », selon le thésaurus, a pu être identifiée, en particulier les portraits, les paysages et les natures mortes. La recherche par métadonnées a ensuite permis de compléter ces catégories, en utilisant les mots-clés présents dans les scènes.

Les algorithmes de similarité ont permis d’identifier des séquences entre les images ou de regrouper l’ensemble des copies d’une même œuvre. Par exemple, pour « La Sainte Famille » de François Ier, qui compte de nombreuses copies et dont les titres peuvent varier, l’algorithme les a regroupées, permettant leur indexation collective après vérification de leur similitude exacte (figure D.9). Le moteur de recherche CLIP a également simplifié l’indexation, en permettant de retrouver des éléments identifiables, tels que des animaux (chien, chat,

oiseau...), des éléments naturels (arbre, fleur, fruits...) des objets (livres, sabliers, crânes...) ou des lieux (Venise et Rome) (figure D.10). Comme CLIP est principalement entraîné sur des paires texte-image en anglais, les recherches étaient plus efficaces dans cette langue. Toutefois, l'utilisation d'autres langues, comme le français ou l'italien, produisait parfois des variations intéressantes dans les résultats. Tout au long du processus, lorsqu'une image était considérée comme entièrement indexée, elle pouvait être marquée comme telle, afin de l'exclure des étapes suivantes de l'indexation.

Après une cinquantaine d'heures consacrées à l'indexation iconographique des peintures du corpus RETIF grâce à Panoptic, un premier bilan peut être établi. Sur les 11 974 images qui le composent, 9 381 comportent au moins un descripteur issu du thésaurus Garnier, soit 78,34 % du total. Par ailleurs, 9 121 images ont pu être rattachées à une catégorie générale de représentation, c'est-à-dire le sujet iconographique. Parmi celles-ci, 4 523 relèvent de l'iconographie religieuse, une catégorie rarement mobilisée dans le thésaurus mais qui s'est révélée pertinente pour regrouper les peintures dévotionnelles, telles que les Vierges à l'Enfant ou les Saintes Familles, et les représentations de saints. L'examen plus détaillé des principales catégories montre que 1 759 peintures traitent de sujets bibliques issus de l'Ancien ou du Nouveau Testament, 1 049 relèvent du portrait, 477 du paysage, 421 de la mythologie, 333 de la nature morte, 133 de l'allégorie et 92 de sujets historiques. Les scènes de genre sont nettement plus rares, puisque seules 56 peintures ont été identifiées. D'autres thématiques apparaissent de manière ponctuelle, comme les représentations animalières, les vues architecturales, quelques sujets littéraires, principalement inspirés de La Jérusalem délivrée, ou encore quelques scènes de spectacle. En moyenne, chaque image est associée à 2,74 termes, et l'indexation a mobilisé 714 concepts distincts du thésaurus Garnier. Le processus a été mené à son terme pour 774 images, dont l'indexation a été précisée comme terminée. À titre de comparaison, le script que nous avions conçu pour extraire automatiquement dans les titres des œuvres les termes issus du thésaurus Garnier avait permis d'obtenir une moyenne de 2,56 termes par image (cf. chapitre 4). L'approche fondée sur Panoptic se distingue ainsi par sa capacité à enrichir de manière notable l'indexation iconographique dans un laps de temps relativement restreint.

Dans l'ensemble, les résultats statistiques obtenus se révèlent particulièrement satisfaisants si on les compare à ceux relevés en mai 2025 dans d'autres bases de données disponibles sur la plateforme AGORHA⁵. À l'exception de la base *Les Fabriques de l'art*, dont l'indexation résulte du travail des conservateurs du Département des Manuscrits de la Bibliothèque

⁵ Annexes figures D.11 et D.12 ; la comparaison de ces statistiques avec celles des autres bases de données est disponible au lien suivant : <https://public.tableau.com/app/profile/pierre.husson/viz/AGORHA-RETIF/UtilisationdeGarnierdansAgorha>

nationale de France dans la base Mandragore⁶, aucune autre base de données consacrée à la peinture ne dépasse la moyenne de termes issus du thésaurus par notice que nous avons atteinte, ni ne mobilise autant de descripteurs provenant du thésaurus. À titre d'exemple, *Le Recensement des tableaux ibériques dans les collections publiques françaises (1300-1870)* (*RETIB*), qui rassemble 461 notices, affiche des résultats sensiblement inférieurs aux nôtres : il présente une moyenne de 2,35 termes par notice et un total de 317 concepts distincts mobilisés, soit moins de la moitié de ceux exploités dans notre propre indexation⁷.

Ce travail a en outre mis en évidence certaines erreurs d'attribution dans les titres — par exemple, une scène représentant le Christ et la Samaritaine au puits identifiée comme un *Noli me Tangere*⁸, ou encore un saint Onufre pris pour saint Jean-Baptiste⁹ — ainsi que des lacunes dans le thésaurus, telles que l'absence du personnage de Lucrèce, figure pourtant centrale de l'histoire antique. Il convient également de souligner que ce processus présente un avantage majeur par rapport au processus automatisé décrit dans la section précédente : il ne nécessite pas de curation des résultats. Lors de l'étude des modèles présentée au chapitre 5, la plupart ont produit un nombre de faux positifs supérieur ou avoisinant les 20, ce qui impliquait de corriger une grande partie des métadonnées générées automatiquement.

Un autre aspect de ce mode de fonctionnement, qui pourrait entraîner un biais dans l'indexation, est la tendance à ne pas se concentrer suffisamment sur chaque image, en passant facilement d'une catégorie à l'autre. Si cette approche réduit fortement la redondance de la tâche en permettant de découvrir constamment de nouvelles images, elle peut également conduire à accorder moins de temps à l'analyse détaillée de certaines œuvres ou catégories de représentation. Quelques erreurs peuvent ainsi survenir au cours du processus. Toutefois, si l'indexeur travaille de manière rigoureuse, en examinant chaque image, en possédant une bonne connaissance de l'iconographie et du corpus étudié, et en n'indexant que les éléments qu'il peut observer, les erreurs ou confusions sont rares. Il s'agit alors plutôt d'omissions, compréhensibles dans le cadre d'un travail d'indexation de cet ampleur.

En renouvelant cette approche documentaire, la principale qualité de cette méthode réside dans la possibilité de consacrer moins de temps aux images facilement identifiables et récurrentes (telles que les très nombreuses Vierges à l'Enfant, les copies d'un même modèle et les variations sur un même sujet) pour accorder davantage de temps aux images complexes,

⁶Institut national d'histoire de l'art et Bibliothèque nationale de France, *La fabrique de l'art. Couleurs et matériaux de l'enluminure*, INHA, URL : <https://agorha.inha.fr/database/89> (visité le 02/09/2025).

⁷Institut national d'histoire de l'art, *Recensement des tableaux ibériques dans les collections publiques françaises (1300-1870)*, RETIB, AGORHA, URL : <https://agorha.inha.fr/database/49> (visité le 02/09/2025).

⁸<https://agorha.inha.fr/ark:/54721/faa8b19c-6436-472e-a6b1-9c5f5091ed48> (consulté le 2/09/2025)

⁹<https://agorha.inha.fr/ark:/54721/b7009df7-2168-4f46-8456-2105dcc3395a> (consulté le 2/09/2025)

qui nécessitent parfois la consultation de références bibliographiques afin d’identifier avec précision les scènes et les motifs iconographiques.

Chapitre 8

Des nouvelles méthodes d'exploration des corpus en histoire de l'art

L'indexation iconographique est une approche qui repose majoritairement sur le texte, en décrivant l'image à l'aide de concepts exprimés en langage naturel, ce qui peut en limiter la portée. L'intelligence artificielle ouvre de nouvelles perspectives sur cette pratique documentaire en établissant un lien multimodal entre texte et image. Elle permet également d'envisager des méthodes de recherche qui dépassent le langage naturel, en mobilisant directement des motifs visuels ou des images comme points d'accès à l'information.

I. La multimodalité texte-image offerte par l'intelligence artificielle

En intelligence artificielle, la multimodalité désigne la capacité d'un modèle à combiner et à raisonner à partir de types de données hétérogènes, tels que des textes, des images, des fichiers audio ou vidéo. En 2023, avec l'usage croissant de l'outil CLIP dans les projets liés aux images, Thomas Smits et Melvin Meyers ont évoqué « un tournant multimodal dans les humanités numériques »¹. CLIP (Contrastive Language-Image Pre-training) est une technologie conçue pour entraîner des modèles à associer textes et images². La version originale de ce modèle a été développée par la société américaine OpenAI et publiée en 2021. Elle a été entraînée sur près de 400 millions de paires image-texte, c'est-à-dire d'images accompagnées

¹Thomas Smits et Melvin Wevers, « A multimodal turn in Digital Humanities. Using contrastive machine learning models to explore, enrich, and analyze digital visual historical collections », *Digital Scholarship in the Humanities*, 38–3 (31 août 2023), p. 1267-1280, DOI : 10.1093/lhc/fqad008.

²A. Radford, J. W. Kim, C. Hallacy, *et al.*, *Learning Transferable Visual Models From Natural Language Supervision...*

de leurs légendes, afin de créer un espace de représentation commun dans lequel les images se rapprochent des descriptions textuelles qui leur correspondent et s'éloignent de celles qui ne correspondent pas. Pour atteindre cet objectif, CLIP repose sur deux encodeurs distincts. Le premier, basé sur des transformateurs, encode un texte en un vecteur reflétant son contenu sémantique. Le second, fondé sur des réseaux de vision, encode une image en un vecteur décrivant son contenu visuel. Chaque description textuelle et chaque image sont ainsi projetées dans un même espace de représentation vectorielle, ce qui permet de comparer directement le texte et l'image. Le modèle peut ainsi identifier les textes qui décrivent le mieux une image, ou les images correspondant à un mot ou une requête en langage naturel, sans nécessiter de nouvel entraînement spécifique pour chaque tâche. Par exemple, pour la requête « un chien », CLIP encode le texte en vecteur et le compare aux vecteurs correspondant à un ensemble d'images. Les images dont la représentation est la plus proche de celle du texte sont considérées comme représentant cet animal. Cette technologie permet donc, entre autres, de rechercher dans un corpus d'images des mots ou des phrases en langage naturel, même si ces images n'ont pas été préalablement annotées ou indexées. Les modèles entraînés à l'aide de cette technologie sont relativement légers et ne nécessitent pas de configuration informatique ou de processeurs graphiques puissants. Une fois téléchargé, il peut être utilisé localement pour encoder des images et effectuer des recherches dans un corpus, sans connexion à Internet.

Plusieurs projets à travers le monde ont mis en place des moteurs de recherche basés sur des modèles CLIP afin d'améliorer la découvrabilité de collections patrimoniales. Cette technologie est notamment utilisée pour explorer des ensembles de données variés, comme les collections de photographies japonaises du Musée Guimet dans le cadre du projet HikarIA³, ou les collections de cartes postales de la Bibliothèque nationale du Danemark⁴. La Bibliothèque nationale de France, dans le cadre de son projet Gallica Images, intègre CLIP en fin de chaîne de traitement. Soutenu par le programme France 2030 et réalisé en coopération avec l'INHA et la Bibliothèque nationale universitaire de Strasbourg, ce projet vise à rendre accessibles les fonds iconographiques numérisés dans Gallica, la bibliothèque numérique de la BnF et de ses partenaires⁵. Grâce aux outils d'intelligence artificielle, les images sont extraites de recueils, albums ou documents à l'aide d'algorithmes de segmentation. Un moteur de recherche intégrant CLIP facilite ensuite l'exploration de ces images, permettant de naviguer efficacement dans le vaste corpus d'images patrimoniales proposées sur Gallica⁶.

³ *HikarIA...*

⁴ Chris Haffenden, Faton Rekathati et Emma Rende, *Unearthing forgotten images with the help of AI*, The KBLab Blog, 20 oct. 2023, URL : <https://kb-labb.github.io/posts/2023-10-20-unearthing-forgotten-images-with-the-help-of-ai> (visité le 04/09/2025).

⁵ Bibliothèque nationale de France, *Les projets en intelligence artificielle à la BnF...*

⁶ Alexandra Adamova et Rosanne Wingert, *Le projet Gallica Images. L'intelligence artificielle pour l'accès aux grands corpus d'images*, 18 juin 2024, URL : <https://www.youtube.com/watch?v=IVMfpR96jrk> (visité

Dans le cadre du projet RETIF, nous avons développé une application web intégrant un moteur de recherche CLIP à l'aide de la bibliothèque open-source Python Streamlit, qui permet de créer rapidement une interface de présentation (annexe I.)⁷. Cette application reproduit les fonctionnalités offertes par l'outil Panoptic : elle récupère les encodages des images préalablement préparés et, grâce à un moteur de recherche textuel basé sur CLIP, propose des images proches correspondant à l'idée exprimée. Le modèle expérimenté est ViT-B/32⁸. Cette approche dépasse les limites de l'indexation iconographique traditionnelle, en permettant de détecter des concepts et des caractéristiques souvent négligés dans les pratiques documentaires classiques : la présence de couleurs, l'intensité lumineuse, la récurrence de motifs communs comme le ciel, les nuages ou l'herbe, ainsi que des dimensions affectives générales de la composition, telles que la tristesse ou la joie. Si les requêtes simples fournissent généralement des résultats pertinents et conformes à la demande (figures E.1 et E.2), les modèles CLIP présentent encore de nombreuses limites pour analyser et comprendre les peintures, notamment en ce qui concerne les styles picturaux ou les iconographies complexes⁹. Par ailleurs, le contenu exact de WebImageText (WIT), le jeu de données interne développé par OpenAI, n'étant pas publié pour des raisons de droits d'auteur et de filtrage, il reste impossible d'évaluer la proportion d'œuvres d'art picturales qu'il contient¹⁰.

II. Améliorer un modèle de CLIP pour répondre aux problématiques de la peinture ?

Ainsi, un test a été mené pour affiner un modèle CLIP préexistant afin d'améliorer la reconnaissance des motifs et des scènes dans les images du RETIF, en particulier avec des termes en français. Le jeu d'entraînement était constitué des 11 966 images du corpus accompagnées d'une courte légende mentionnant le nom de l'artiste, le titre de l'œuvre et les termes d'indexation associés, identifiés lors du travail sur l'application Panoptic. Après le fine-tuning et l'intégration du modèle dans l'application Streamlit développée précédemment, les résultats se sont révélés très décevants et nettement inférieurs à ceux du modèle VIT-B/32, utilisé précédemment. Le système ne présentait plus aucune cohérence, et même les recherches liées à des termes bien représentés et indexés dans le corpus (par exemple

le 04/09/2025). L'outil a été expérimenté dans les locaux de l'INHA entre avril et juillet 2025.

⁷streamlitDocumentation

⁸Le modèle Vit-L, plus puissant, demandait trop de ressources pour l'encodage des images par rapport à celles disponibles.

⁹Andrea Asperti, Leonardo Dessì, Maria Chiara Tonetti et Nico Wu, *Does CLIP perceive art the same way we do ?, version 1, 2025, doi : 10.48550/ARXIV.2505.05229.*

¹⁰A. Radford, J. W. Kim, C. Hallacy, *et al.*, *Learning Transferable Visual Models From Natural Language Supervision..., p. 4-5.*

« nature morte », « Vierge », « chien ») aboutissaient à des résultats aberrants. Nos compétences techniques limitées et le temps restreint alloué à l’expérience ne nous ont pas permis d’identifier précisément l’origine du problème. Toutefois, plusieurs hypothèses peuvent être formulées. Nos 12 000 paires images-textes constituaient peut-être un volume insuffisant pour atteindre un niveau de performance comparable à celui des modèles entraînés sur de plus grandes quantités de données. D’autres raisons techniques peuvent expliquer l’inefficacité de l’affinement de notre modèle : un entraînement limité à deux époques, l’utilisation d’un système d’exploitation Windows au lieu de Linux, qui est préconisé, et des ressources de calcul insuffisantes. Sur la machine de l’INHA, dotée d’une carte graphique NVIDIA GeForce RTX 4080, l’opération avait déjà mobilisé plus de quatre heures.

Un autre facteur de difficulté tient à la complexité visuelle des images du RETIF. Là où les jeux de données classiques présentent souvent une image contenant un seul objet accompagné d’une description simple (« un chien », « une pomme »), nos peintures réunissent fréquemment de nombreux éléments simultanés (une nature morte composée d’un globe terrestre, de cartes marines, d’instruments mathématiques sous une draperie de tissu, par exemple). L’absence d’annotations précises sur la localisation et la hiérarchie de ces objets peut limiter l’efficacité de l’apprentissage.

Une stratégie plus pertinente pour améliorer CLIP dans le contexte pictural aurait pu consister soit à constituer un corpus de motifs iconographiques isolés (par exemple, uniquement la figuration d’un chien ou d’un fruit dans une peinture) assortis de descriptions simples, soit à adapter des jeux de données photographiques existants pour leur donner une apparence picturale. Cette seconde approche, de transformer des photographies en images simulant les effets de texture de la peinture à l’huile, médium le plus présent dans le corpus du RETIF, est d’ailleurs recommandée par Thomas Heitzinger David G. Stork dans une étude où ils rapportent l’efficacité de la méthode pour des modèles de segmentation¹¹. Cette démarche permettrait au modèle de se concentrer moins sur la catégorisation des images du RETIF en tant que « peintures » et davantage sur l’identification des motifs iconographiques qu’elles représentent, indépendamment de leur caractère pictural.

III. Une approche image-à-image

En 2016, Erik Malcolm Champion mettait en évidence la prépondérance des études textuelles au sein des humanités numériques et plaideait pour une plus grande autonomie des

¹¹Thomas Heitzinger et D. G. Stork, « Improving semantic segmentation of fine art images using photographs rendered in a style learned from artworks », *IS&T International Symposium on Electronic Imaging 2022*, 34–13 (16 janv. 2022), p. 169-1-169-5, DOI : 10.2352/EI.2022.34.13.CVAA-169.

images et des médias par rapport au texte¹². L’année suivante, Chiara Franceschini appelait à repenser les méthodologies de l’iconographie, trop dépendantes des descriptions textuelles, et évoquait le « tournant visuel » en histoire de l’art rendu possible par les nouvelles technologies¹³. En effet, au sein des méthodes de cette discipline, la comparaison d’images est particulièrement pertinente pour identifier des copies, des variations ou des transferts de motifs entre œuvres, pour comparer différentes versions d’un même thème ou pour étudier les compositions dans le cadre de séries ou de multiples. Pour répondre à des problématiques nécessitant la recherche d’images comparatives à partir d’une image donnée, des outils numériques ont été développés afin de permettre une recherche visuelle en se détachant des contraintes textuelles.

C’est notamment le cas de la plateforme Ukiyo-e.org, créée en 2012 par l’ingénieur John Resig¹⁴. Elle contient plus de 200 000 estampes japonaises provenant de plus de 24 institutions dans le monde et propose deux méthodes de recherche : la première repose sur un moteur de recherche textuel classique exploitant les métadonnées des œuvres, tandis que la seconde permet de rechercher dans la base des images similaires à une photographie d’estampe téléchargée par l’utilisateur¹⁵. Les estampes issues des différentes collections sont regroupées par similarité pour mettre en valeur les différents états et impressions d’une même estampe répartis dans les diverses institutions.

Le projet Replica poursuivait un objectif similaire de restauration de l’autonomie des images et de facilitation des liens entre elles pour la recherche. Développé par Benoît Seguin à l’École Polytechnique Fédérale de Lausanne sous la supervision de Frédéric Kaplan, et en partenariat avec la Fondation Cini à Venise, les premiers résultats ont été publiés en 2016¹⁶. Le principe était de créer un moteur iconographique basé sur l’image, utilisant des réseaux de neurones convolutifs (CNN) pour identifier des motifs récurrents et des similarités compositionnelles entre les œuvres. Ces recherches s’appuyaient sur les grandes théories de l’histoire de l’art, notamment celles d’Aby Warburg et d’Ernst Gombrich, et permettaient d’exploiter l’immense fonds photographique de l’institution vénitienne. Xi Shen, Alexei A. Efros et Mathieu Aubry ont également exploré à plusieurs reprise la similarité de motifs dans

¹²Erik Malcolm Champion, « Digital humanities is text heavy, visualization light, and simulation poor », *Digital Scholarship in the Humanities* (, 11 nov. 2016), fqw053, DOI : 10.1093/11c/fqw053.

¹³Chiara Franceschini, « Classifying image content in visual collections, a selective history », dans *The Routledge Companion to Medieval Iconography*, dir. Colum Hourihane, Routledge, Londres et New York, 2017, p. 184-191, p. 190 : « The classic methodology of iconography, with its strong tendency to anchor image to text, if not to a source, and to a subject matter (described either with a text or a simple textual label), needs to be rethought. [...] Without sounding too paradoxical, it may be possible to say that we are today assisting in a visual turn in art history »

¹⁴John Resig, *Ukiyo-e.org : Japanese Print Search and Database*, 2012, URL : <https://ukiyo-e.org/> (visité le 04/09/2025).

¹⁵Id., *About The Site*, Ukiyo-e.org, 2012, URL : <https://ukiyo-e.org/about> (visité le 04/09/2025).

¹⁶B. Seguin, C. Striolo, I. di Lenardo, *et al.*, « Visual Link Retrieval in a Database of Paintings »...

les arts visuels, en particulier entre différents médiums comme les peintures et les dessins¹⁷.

Avec l’arrivée de CLIP et la mise à disposition de modèles entraînés avec cette technologie, de nouvelles perspectives se sont ouvertes pour la recherche par similarité d’images, ajoutant une dimension sémantique aux comparaisons visuelles. L’encodage et la création de vecteurs d’embeddings permettent de mesurer la proximité entre images selon leur contenu tant visuel que sémantique. Ainsi, de nombreux moteurs de recherche intègrent désormais cette fonctionnalité, permettant de trouver des images similaires non pas à partir de textes, mais directement à partir d’une image.

Par exemple, ICONCLASS propose à ses membres un outil multimodal permettant d’explorer un jeu de données entièrement indexé selon son système iconographique. L’outil facilite l’indexation en montrant des images similaires à une représentation que l’on souhaite décrire¹⁸.

Dans le cadre du projet RETIF, la recherche inversée à l’aide d’une image a été testée dans l’application développée avec Streamlit et CLIP que nous évoquions précédemment. Cette approche permet d’identifier non seulement des copies d’une même œuvre, mais aussi certains motifs iconographiques. Le modèle encode l’image ajoutée par l’utilisateur et présente les images similaires de la plus proche à la moins proche. En exploitant son apprentissage à partir de paires image-texte, il est capable d’identifier des représentations différentes d’un point de vue formel mais similaires en termes de contenu. Ainsi, une photographie d’un poisson peut permettre d’extraire les nature mortes du RETIF où cet animal est figuré (figure E.3). Selon Polo Sologub et Rebecca Fiebrink, CLIP offre même d’excellents résultats lorsqu’on lui soumet un croquis d’un objet à retrouver¹⁹. Ce type de recherche par l’image constitue un atout majeur pour les historiens de l’art et les utilisateurs de bases de données, en particulier pour identifier des éléments ou objets qui n’ont pas été décrits lors de l’indexation.

Ainsi, les outils d’intelligence artificielle fondés sur la similarité des images et la multimodalité, tels que CLIP, ouvrent de nouvelles perspectives pour l’indexation iconographique

¹⁷Xi Shen, Alexei A. Efros et M. Aubry, *Discovering Visual Patterns in Art Collections with Spatially-consistent Feature Learning*, version 2, 2019, DOI : 10.48550 / ARXIV.1903.02678 ; X. Shen, Robin Champenois, Shiry Ginosar, Ilaria Pastrolin, Morgane Rousset, Oumayma Bounou, Tom Monnier, Spyros Gidaris, François Bougard, Pierre-Guillaume Raverdy, *et al.*, « Spatially-Consistent Feature Matching and Learning for Heritage Image Analysis », *International Journal of Computer Vision*, 130–5 (mai 2022), p. 1325-1339, DOI : 10.1007/s11263-022-01576-x.

¹⁸Je tiens à remercier Hans Brandhorst et Etienne Posthumus de m’avoir permis de tester l’outil en mars 2025. Cristian Santini, E. Posthumus, Mary Ann Tan, Oleksandra Bruns, Tabea Tietz et Harald Sack, *Multimodal Search on Iconclass using Vision-Language Pre-Trained Models*, version 1, 2023, DOI : 10.48550 / ARXIV.2306.16529 ; H. Brandhorst, *ICONCLASSplus*, URL : <https://iconclass.org/help/plus> (visité le 04/09/2025), diapositive 14 de la présentation

¹⁹Polo Sologub et Rebecca Fiebrink, *Sketchy Collections : Exploring Digital Museum Collections by Drawing via CLIP*, 2024, URL : https://ualresearchonline.arts.ac.uk/id/eprint/23541/1/NeurIPS_2024_Sketchy_Collections_camera-ready.pdf (visité le 04/09/2025).

et proposent des approches complémentaires pour l’exploration des corpus visuels.

Conclusion

L'aboutissement de nos réflexions rejoint celle que Jacques Thuillier formulait déjà il y a quarante ans²⁰. Face à l'essor des premiers outils informatiques dans les institutions patrimoniales, il invitait alors les musées et les bibliothèques non pas à s'en servir pour reconduire des usages préexistants, mais à interroger en profondeur ces pratiques à la lumière des besoins réels des utilisateurs. Le même constat émerge aujourd'hui à travers les expérimentations menées sur l'indexation du RETIF à l'aide des technologies d'intelligence artificielle. Doit-on mobiliser ces outils pour reconduire des pratiques documentaires conçues dans un contexte matériel et scientifique antérieur, ou au contraire les envisager comme une occasion d'élaborer de nouveaux usages, en adéquation avec les possibilités qu'ils offrent et les besoins des utilisateurs ?

Ce projet a notamment mis en évidence plusieurs points d'attention concernant l'indexation iconographique et les instruments qu'elle emploie, en particulier les vocabulaires contrôlés comme le thésaurus Garnier. Son emploi constitue un véritable défi pour l'automatisation de l'indexation, mais il soulève également des interrogations plus larges sur sa pertinence dans les bases de données iconographiques. Le thésaurus Garnier reste en effet isolé des autres systèmes d'indexation, y compris entre ses propres déclinaisons (POP, AGORHA, Mandragore,...). Depuis sa création dans les années 1980, son usage a été multiple, ses adaptations et modifications institutionnelles compliquant l'interconnexion des bases. Or, depuis plusieurs années, ce vocabulaire semble en perte de vitesse et de plus en plus critiqué par la communauté des historiens de l'art pour son caractère monolingue et sa difficulté à s'adapter aux problématiques contemporaines. Dans un contexte où les bases de données se mondialisent et ne s'adressent plus uniquement à des chercheurs francophones, une enquête approfondie sur son utilisation paraît nécessaire, notamment pour envisager des passerelles avec d'autres systèmes d'indexation.

Un autre problème tient à la difficulté pour les bases de données liées à ce thésaurus de remplir la fonction première de l'indexation : favoriser la recherche documentaire. Les limites d'ergonomie de plateformes comme POP ou AGORHA réduisent fortement l'intérêt

²⁰J. Thuillier, « Image et informatique : de l'utilisation au système »...

de l’indexation iconographique à l’aide du thésaurus Garnier. Ces enjeux d’accessibilité et d’interface devraient être résolus en priorité, avant même d’envisager une automatisation à grande échelle de cette tâche complexe de description des représentations. Ainsi, si le projet d’indexation iconographique assistée par intelligence artificielle du RETIF a permis de révéler ces difficultés, il témoigne aussi d’une certaine tendance à vouloir exploiter des technologies en plein essor, démarche qui gagnerait à s’assortir d’une réflexion préalable quant au véritable intérêt pour les utilisateurs.

Un point crucial demeure donc celui de l’analyse préalable des pratiques de la communauté et de la définition précise des problèmes auxquels l’intelligence artificielle est censée répondre. Cette étape, indispensable pour encourager des méthodes adaptées et guider d’autres projets de ce type, n’a pas pu être menée en profondeur lors des expérimentations autour du RETIF. Une telle enquête aurait pourtant permis de mieux comprendre les attentes réelles de la communauté des usagers et d’orienter le choix de solutions adaptées et pertinentes. Des cadres méthodologiques, tels que ceux proposés par la Bibliothèque du Congrès²¹, ou encore des recommandations issues d’expériences similaires à la nôtre, mériteraient d’être valorisés et diffusés davantage, afin d’inciter les institutions à intégrer l’intelligence artificielle de façon réfléchie et efficace, lorsqu’une approche purement expérimentale n’est pas envisagée.

Un autre enjeu, encore trop peu mis en avant, concerne la qualité des données destinées à constituer des jeux d’entraînement exploitables par les systèmes d’intelligence artificielle. Cette réflexion est aujourd’hui portée par des institutions comme la Bibliothèque nationale de France ou encore ICONCLASS, qui travaillent à améliorer la valeur de leurs corpus et de leur indexation pour nourrir les algorithmes. Elle s’inscrit dans la continuité du rôle des institutions patrimoniales, en France comme à l’étranger, à la fois dans l’inventaire et le catalogage des collections et dans leur mise en libre accès.

Les technologies d’intelligence artificielle apparaissent et évoluent à grands pas, et les possibilités que pourront offrir les outils de vision par ordinateur dans les prochaines années surepasseront très certainement celles d’aujourd’hui. Il est donc préférable de privilégier des chaînes opératoires légères et adaptables, garantissant la pérennité des résultats face à l’émergence de solutions plus performantes. Dans ce contexte, de réelles perspectives se sont révélées grâce à l’usage d’outils comme Panoptic, permettant d’explorer et d’annoter les corpus plus efficacement. Sa capacité à permettre l’exploration de banques d’images et à y associer des métadonnées constitue un atout majeur pour l’indexation iconographique. Si une indexation complète du RETIF devait être entreprise en s’appuyant sur la vision par ordinateur, c’est cette solution que nous préconisons. Elle a l’avantage de prolonger des pratiques établies en

²¹Library of Congress, *Planning Framework used by LC Labs for planning AI experiments towards responsible implementation...*

histoire de l’art, sans exiger de lourdes manipulations techniques ni de ressources de calcul intensives, et sans nécessiter l’envoi de données vers des serveurs distants pour les effectuer. Elle évite également le risque de produire des informations de qualité hétérogène, ou d’imposer un travail de curation long et fastidieux *a posteriori*. Nous tenons à souligner ce point : les responsables de collections et les documentalistes n’ont pas la possibilité d’assumer cette charge et la politique de l’institution en matière de curation, notamment le temps consacré ou non à cette pratique, doit être définie dès la conception du projet.

Enfin, l’intelligence artificielle ouvre des perspectives inédites pour l’exploration des corpus iconographiques, notamment grâce aux modèles capables de relier texte et image. Aujourd’hui, la technologie CLIP s’impose comme une solution à la fois performante et efficace, largement adoptée dans de nombreux projets. Ces outils permettent d’interroger les banques d’images non seulement en langage naturel, mais également à partir d’éléments visuels, en exploitant les similarités entre représentations. Cette approche offre de nouvelles possibilités de navigation dans les corpus iconographiques et constitue un atout considérable pour la recherche en histoire de l’art. Elle répond en définitive à l’un des enjeux fondamentaux de l’indexation iconographique depuis ses origines : permettre l’exploration et la découverte des bases de données pour nourrir l’étude et l’investigation scientifiques.

Toutefois, loin d’être écartée au profit des technologies émergentes, l’indexation iconographique doit s’imposer comme un levier essentiel de production de données de qualité, au service à la fois d’analyses iconographiques précises, de méthodes de l’histoire de l’art computationnelle et d’études transdisciplinaires. Son avenir ne réside donc pas dans l’opposition avec les innovations numériques, mais dans leur complémentarité, afin de promouvoir des modèles interopérables et partagés pour la description des représentations.

Annexe A

Le corpus du RETIF

I. Exemples d'images provenant du RETIF



FIG. A.1 : Taddeo Gaddi, *L'Adoration des bergers*, vers 1330, tempéra sur bois, Dijon, Musée des Beaux-Arts, inv. 1470 © Musée des Beaux-Arts de Dijon



FIG. A.2 : Paul Véronèse, *Persée délivrant Andromède*, 1575-1580, peinture à l'huile sur toile, Rennes, Musée des Beaux-Arts de Rennes, inv. 801.1.1



FIG. A.3 : Pérugin, *Saint Sébastien et sainte Apolline*, 1513 -1523, peinture à l'huile sur toile, Grenoble, Musée de Grenoble, inv. MG 48



FIG. A.4 : Raphaël, *La Sainte Famille, dit La Grande Sainte Famille de François Ier*, 1518, peinture à l'huile sur toile, Paris, Musée du Louvre, INV 604 ; MR 432 © 2013 GrandPalaisRmn (musée du Louvre) / Adrien Didierjean

II. Exemples d'images inexploitables



FIG. A.5 : Giuseppe Recco *Nature morte : poisson de mer et chaudron*, Musée des Beaux-Arts (Orléans), inv. 1173, Lien AGORHA :
<https://agorha.inha.fr/ark:/54721/d5fc162e-48a3-4508-9a6c-562fb76c1a1a>



FIG. A.6 : Copie anonyme d'après Francesco Albani, *Le Baptême du Christ*, Eglise Saint-Jean-Baptiste (Esbly), Lien AGORHA :

<https://agorha.inha.fr/ark:/54721/d9cc8098-0959-422e-902d-d5a8a6d41c6d>

Annexe B

Le thésaurus Garnier dans AGORHA

I. L'indexation à l'aide du thésaurus Garnier dans les bases de données d'AGORHA

Ces graphiques sont disponibles au lien suivant : <https://public.tableau.com/app/profile/pierre.husson/viz/AGORHA/UtilisationdeGarnierdansAgorha>

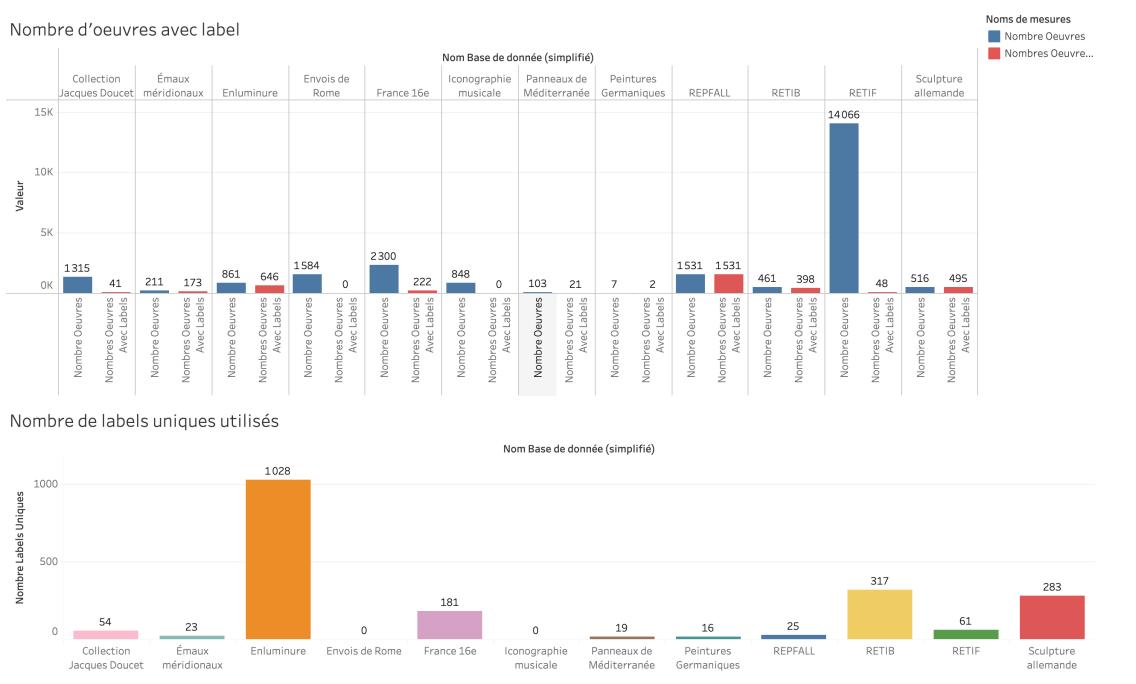


FIG. B.1 : Graphiques montrant le nombre de notices indexées avec au moins un label dans 12 bases d'AGORHA, et le nombre de labels distincts utilisés par bases.

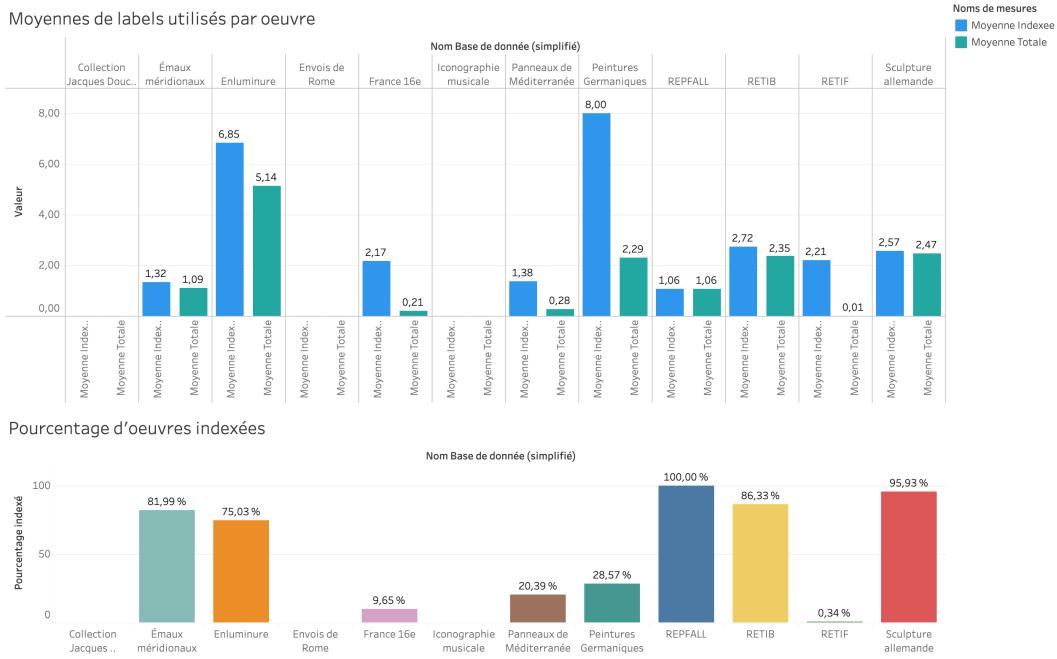


FIG. B.2 : Graphique montrant la moyenne de labels utilisés par notices dans 12 bases d'AGORHA, pour les œuvres indexées et toutes les œuvres ; et le pourcentage de œuvres qui possèdent au moins un label par base de données.

II. Code Python pour structurer le thésaurus Garnier en json

```
1 import json
2 from rdflib import Graph, Namespace
3 from collections import defaultdict
4
5 """
6 Ce script génère deux fichiers pour faciliter l'exploitation du thésaurus Garnier
7 → de l'INHA à partir d'un fichier XML-RDF :
8 - un fichier JSON qui représente la hiérarchie complète des concepts et qui
9 → inclut, pour chacun d'entre eux : l'URI (id), le label préféré (prefLabel),
10 → les labels alternatifs (altLabels), les concepts enfants (children). La
11 → hiérarchie est structurée sous forme de dictionnaires imbriqués, et tous les
12 → concepts sont triés par ordre alphabétique.
```

```

9  – un fichier TXT qui contient uniquement les labels principaux (prefLabel) des
   ↵ concepts. La hiérarchie est reproduite grâce à une indentation par niveau.
10 """
11
12 rdf_entree = "thesaurusGarnierINHA.rdf"
13 resultat_json = "JSONthesaurusGarnierINHA.json"
14 resultat_txt = "TXTthesaurusGarnierINHA.txt"
15
16 # Charger le fichier XML-RDF
17 g = Graph()
18 g.parse(rdf_entree, format="xml")
19
20 # définition du namespace SKOS
21 SKOS = Namespace("http://www.w3.org/2004/02/skos/core#")
22
23 # préparation des dictionnaires
24 labels = {}
25 alt_labels = defaultdict(list)
26 broader_relations = defaultdict(list)
27
28 # Extraction des "prefLabel"
29 for s, p, o in g.triples((None, SKOS.prefLabel, None)):
30     labels[str(s)] = str(o)
31
32 # Extraction des "altLabel"
33 for s, p, o in g.triples((None, SKOS.altLabel, None)):
34     alt_labels[str(s)].append(str(o))
35
36 # identification de la hiérarchie parents enfants des concepts
37 for s, p, o in g.triples((None, SKOS.broader, None)):
38     broader_relations[str(o)].append(str(s))
39
40 # Trouver les concepts racines (càd les concepts sans skos:broader)
41 all_concepts = set(labels.keys()) | set(alt_labels.keys())
42 children = set()
43 for children_list in broader_relations.values():
44     children.update(children_list)
45
46 roots = list(all_concepts - children)

```

```
47
48 # fonction récursive pour créer la hiérarchie JSON
49 def build_hierarchy(concept_uri):
50     children_sorted = sorted(
51         broader_relations.get(concept_uri, []),
52         key=lambda x: labels.get(x, "").lower()
53     )
54     return {
55         "id": concept_uri,
56         "label": labels.get(concept_uri, ""),
57         "altLabels": sorted(alt_labels.get(concept_uri, []), key=str.lower),
58         "children": [build_hierarchy(child) for child in children_sorted]
59     }
60
61 # Créer la hiérarchie complète (racines triées dans l'ordre alphabétique des
62 # labels)
63 roots_sorted = sorted(roots, key=lambda x: labels.get(x, "").lower())
64 hierarchy = [build_hierarchy(root) for root in roots_sorted]
65
66 # Sauvegarder en JSON
67 with open(resultat_json, "w", encoding="utf-8") as f:
68     json.dump(hierarchy, f, ensure_ascii=False, indent=2)
69     print(f"Le fichier {resultat_json} a bien été créé !")
70
71 # Sauvegarde d'un txt indenté contenant seulement les labels
72 def write_txt(node, f, indent=0):
73     f.write("\t" * indent + str(node["label"]) + "\n")
74     for child in node["children"]:
75         write_txt(child, f, indent + 1)
76
77 with open(resultat_txt, "w", encoding="utf-8") as f:
78     for root in hierarchy:
79         write_txt(root, f)
80         print(f"Le fichier {resultat_txt} a bien été créé !")
```

Annexe C

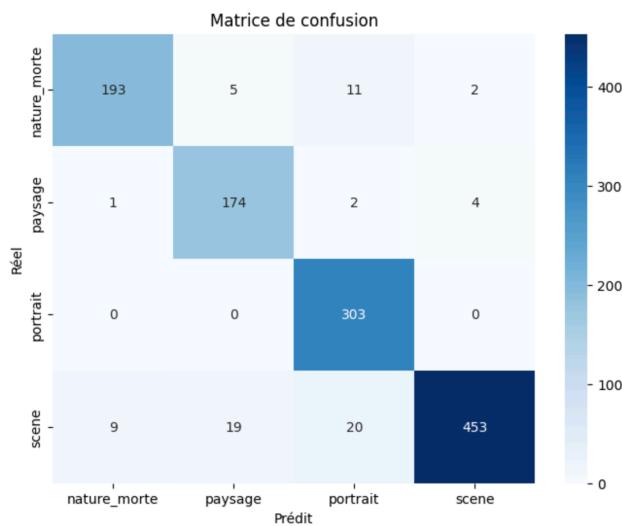
Les modèles d'intelligence artificielles

I. Métriques par classe pour le modèle YOLO entraîné sur des natures mortes

Classe	TP	FP	FN	Précision	Rappel	F1-score	Support
vase	0	0	0	0.0	0.0	0.0	0
dog	0	1	5	0.0	0.0	0.0	5
glass	0	1	0	0.0	0.0	0.0	0
flower	1	13	4	0.071	0.2	0.105	5
grapes	1	0	2	1.0	0.333	0.5	3
musical instrument	1	0	1	1.0	0.5	0.667	2
sheet music	0	2	0	0.0	0.0	0.0	0
skull	4	1	1	0.8	0.8	0.8	5
book	3	1	0	0.75	1.0	0.857	3
monkey	3	10	1	0.231	0.75	0.353	4
onion	0	5	0	0.0	0.0	0.0	0
cabbage	0	1	0	0.0	0.0	0.0	0
bunch of flowers	0	1	0	0.0	0.0	0.0	0
snail	0	0	0	0.0	0.0	0.0	0
fishes	1	2	1	0.333	0.5	0.4	2
caterpillar	1	6	1	0.143	0.5	0.222	2
butterfly	0	1	0	0.0	0.0	0.0	0
beetle	1	6	0	0.143	1.0	0.25	1
cat	0	0	0	0.0	0.0	0.0	0
rabbit	0	2	0	0.0	0.0	0.0	0
bottle	3	9	7	0.25	0.3	0.273	10
hourglass	3	2	3	0.6	0.5	0.545	6
bowl	0	0	0	0.0	0.0	0.0	0

Classe	TP	FP	FN	Précision	Rappel	F1-score	Support
bread	2	14	3	0.125	0.4	0.19	5
basket	0	1	1	0.0	0.0	0.0	1
shell	0	0	0	0.0	0.0	0.0	0
peacock	0	0	0	0.0	0.0	0.0	0
parrot	1	2	1	0.333	0.5	0.4	2
knife	0	1	0	0.0	0.0	0.0	0
cheese	0	0	1	0.0	0.0	0.0	1
violin	0	1	0	0.0	0.0	0.0	0
flute	0	3	0	0.0	0.0	0.0	0
bird	1	0	0	1.0	1.0	1.0	1
playing cards	0	0	0	0.0	0.0	0.0	0
insect	0	0	0	0.0	0.0	0.0	0
eggs	0	0	0	0.0	0.0	0.0	0
guitar	2	5	0	0.286	1.0	0.444	2
vegetable	0	6	1	0.0	0.0	0.0	1
fruit	6	0	1	1.0	0.857	0.923	7
apple	0	0	0	0.0	0.0	0.0	0
cherries	0	3	0	0.0	0.0	0.0	0
meat	0	0	2	0.0	0.0	0.0	2
globe	0	1	0	0.0	0.0	0.0	0
plate	0	2	0	0.0	0.0	0.0	0
squirrel	0	4	0	0.0	0.0	0.0	0

II. Matrice de confusion du modèle FastAI pour la reconnaissance du sujet iconographique



III. L'étude comparative des modèles GPT-4, Mistral M4, LLAVA et BLIP

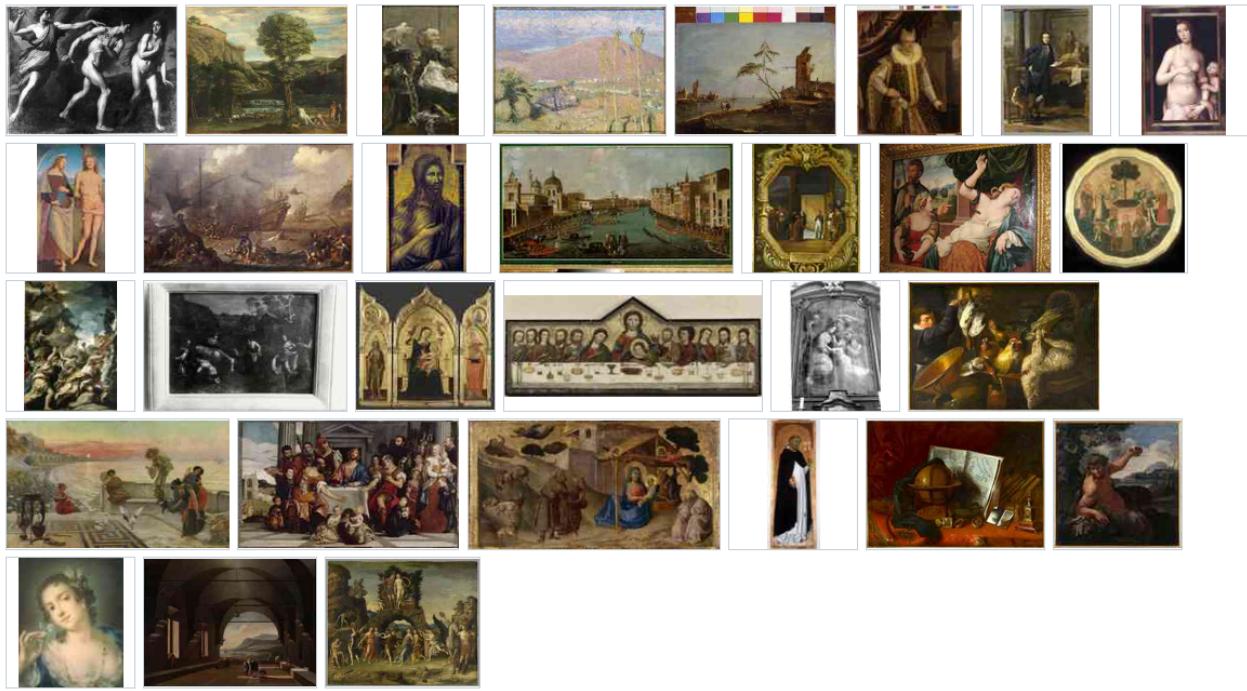


FIG. C.1 : Les trente tableaux choisis dans le RETIF pour constituer le benchmark des modèles de vision par ordinateur.

IV. Liste des peintures utilisées pour l'étude des modèles (Chapitre 6)

- Grifo di Tancredi, *Saint Jean-Baptiste*, 1e quart du XIV^e siècle, Musée des Beaux-Arts (Chambéry), inv. M 933. Lien AGORHA :
<https://agorha.inha.fr/ark:/54721/d72206a7-27c0-4260-8f16-dbf6fbcceab3>
- Maestro della Maddalena, *La Cène*, 3^e quart du XIII^e siècle, dépôt du musée du Louvre (Paris) au musée du Petit Palais (Avignon), inv. 1089.1. Lien AGORHA :
<https://agorha.inha.fr/ark:/54721/b5efecda-e688-41c0-a8cb-31fcde24e9b6>
- Andrea da Firenze, *La Fontaine d'amour*, XIV^e siècle, Douai, Musée de la Chartreuse, Inv. 1089. Lien AGORHA :
<https://agorha.inha.fr/ark:/54721/e5b631b2-a483-4171-bdf9-da6bfdbaa1115>

- Taddeo Gaddi, *L'Adoration des bergers*, vers 1330, Dijon, Musée des Beaux-Arts, inv. 1470.
Lien AGORHA :
<https://agorha.inha.fr/ark:/54721/f39e3826-f862-4489-b979-4c8063b9950b>
- Bartolo di Fredi, *Saint Dominique*, 1397, Musée des Beaux-Arts (Chambéry), inv. D.81.1.3.
Lien AGORHA :
<https://agorha.inha.fr/ark:/54721/7e0aada6-931f-4b4b-b9bc-ef0461daf983>
- Andrea Mantegna, *Mars et Vénus*, 1496-1497, Département des peintures du musée du Louvre (Paris), INV 370. Lien AGORHA :
<https://agorha.inha.fr/ark:/54721/687cce0f-9a9e-4e3d-a826-702cf81be157>
- Francesco Francia, *Vénus et Cupidon*, vers 1500, Musée des Beaux-Arts (Mulhouse), inv. 61.1.13. Lien AGORHA :
<https://agorha.inha.fr/ark:/54721/823a7d31-05ba-4545-81f1-b63a9b81f1b0>
- Le Pérugin, *Saint Sébastien et sainte Apolline*, 1er quart du XVI^e siècle, Musée de Grenoble, inv. MG.48. Lien AGORHA :
<https://agorha.inha.fr/ark:/54721/ec5f09f4-6f00-457b-ba37-feaf4d63acbe>
- Paolo Véronèse, *Les Pèlerins d'Emmaüs*, vers 1559, Musée du Louvre (Paris), inv. INV 146.
Lien AGORHA :
<https://agorha.inha.fr/ark:/54721/837ec2ef-7d88-45c5-9da1-32d9860c7a28>
- Attribué à Francesco Montemezzano, *Portrait d'une dame tenant un luth*, XVI^e siècle, Musée de Picardie (Amiens), inv. MP.Lav.1894.223. Lien AGORHA :
<https://agorha.inha.fr/ark:/54721/c306700a-1c71-424e-a136-32f7f7026739>
- Copie anonyme d'après Jacopo Bassano, *L'Automne*, XVII^e siècle, Musée Garinet (Châlons-en-Champagne), inv. 899.11.237 Lien AGORHA :
<https://agorha.inha.fr/ark:/54721/a793bae2-1995-487e-b06e-e081a94927fd>
- Girolamo Marchesi, *La Mort de Cléopâtre*, milieu du XVI^e siècle, Musée Baron Gérard (Bayeux), inv. P.0177. Lien AGORHA :
<https://agorha.inha.fr/ark:/54721/cc06cbe1-9bb2-4694-9084-4eb4ea861fe9>
- Ciro Ferri, *Faune au verre de vin. L'Automne (?)*, XVII^e siècle, Palais Fesch-musée des beaux-arts (Ajaccio), inv. MFA 852.1.81. Lien AGORHA :
<https://agorha.inha.fr/ark:/54721/dac0ec32-4208-41dc-a22d-4bed21591c55>
- Entourage de Viviano Codazzi, *Galerie voûtée avec un fond de marine*, milieu du XVII^e siècle, Musée des Beaux-Arts (Marseille), inv. 380. Lien AGORHA :
<https://agorha.inha.fr/ark:/54721/0dad77b7-bbee-4f1e-b33a-c96cb10203af>

- Giacomo Legi, *Jeune homme dans un garde manger (nature morte avec gibier, volaille et mouton)*, 1^e moitié du XVII^e siècle, Musée des Beaux-Arts (Bordeaux), inv. M.6243. Lien AGORHA :
<https://agorha.inha.fr/ark:/54721/50b22315-55a9-4a46-a5fc-051fd30be318>
- Attribué à Benedetto Fioravanti, *Nature morte avec globe terrestre et carte d'Italie*, XVII^e siècle, Palais Fesch-musée des beaux-arts (Ajaccio), inv. MFA 852.1.592. Lien AGORHA :
<https://agorha.inha.fr/ark:/54721/525f10ac-19b1-44ca-8f4c-12952f913b6b>
- Le Dominiquin, *Paysage avec Hercule combattant Acheloüs changé en taureau*, 1621-1622, Musée du Louvre (Paris), inv. INV 794. Lien AGORHA :
<https://agorha.inha.fr/ark:/54721/d629c1a1-4cd9-406c-834b-393f93d9bc24>
- Atelier de Luca Giordano, *La Mort d'Orphée*, XVII^e siècle, Musée Hyacinthe Rigaud (Perpignan), inv. 840.2.7. Lien AGORHA :
<https://agorha.inha.fr/ark:/54721/709a2a40-0b00-4d85-b1bd-4cac2d58d31f>
- Ottavio Vannini, *Adam et Eve chassés du Paradis terrestre*, XVII^e siècle, Musée des beaux-arts Jules Chéret (Nice), inv. 1884. Lien AGORHA :
<https://agorha.inha.fr/ark:/54721/de6f3938-90a3-4798-b0af-dcf9d2c85bf8>
- Cornelis de Wael, *Combat naval*, XVII^e siècle, Musée d'Art et d'Histoire (Cognac), inv. 892.1.24. Lien AGORHA :
<https://agorha.inha.fr/ark:/54721/9d64e054-0e1b-45c1-ba04-35765ea39e86>
- Attribué à Francesco Guardi, *Paysage avec marine et tours*, XVIII^e siècle, Musée de Picardie (Amiens), inv. MP.Lav.1894.217. Lien AGORHA :
<https://agorha.inha.fr/ark:/54721/e57b34cb-b772-4014-b639-ad661df64bb4>
- Attribué à Rosalba Carriera, *L'Été*, XVIII^e siècle, Fondation Bemberg (Toulouse), inv. 1064. Lien AGORHA :
<https://agorha.inha.fr/ark:/54721/88c8ac54-99e4-4d6a-98f9-bef4d2553784>
- Pompeo Batoni, *Portrait de Charles John Crowle (1738-1811)*, 1761-1762, Musée du Louvre (Paris), inv. RF 1981 37. Lien AGORHA :
<https://agorha.inha.fr/ark:/54721/652d90b7-46f0-406b-b34c-7c971ec44391>
- Luca Carlevarijs, *Venise, fête sur le grand canal*, 1^e moitié du XVIII^e siècle, Musée de la Castre (Cannes), inv. 35. Pe. Lien AGORHA :
<https://agorha.inha.fr/ark:/54721/cd495a47-ef36-4184-b91b-d56ddf610909>
- Antonio Mancini, *L'Homme au perroquet*, avant 1917, Musée d'Orsay (Paris), inv. RF 1980 137. Lien AGORHA :
<https://agorha.inha.fr/ark:/54721/26e71710-c577-473b-a0eb-88e1d9734462>

- Luigi Rubio, *Napoléon reçoit Mohammed Mirza Khan, ambassadeur de Perse, à Firkenstein, 27 avril 1807*, 1834, Musée national des châteaux de Versailles et de Trianon (Versailles), inv. 1850.1442. Lien AGORHA :
<https://agorha.inha.fr/ark:/54721/b1f66158-0117-4c85-a6be-da8b69d935c8>
- Emilio Vasarri, *Les Oies*, vers 1908, Musée d'arts de Nantes, inv. 1208. Lien AGORHA :
<https://agorha.inha.fr/ark:/54721/1f51a542-129b-45c6-897a-b94916541bca>
- Serafino Macchiati, *Heure automnale ou L'Automne*, 1914, Musée d'Orsay (Paris) inv. INV 20241. Lien AGORHA :
<https://agorha.inha.fr/ark:/54721/8229d01a-857d-4068-8e68-de929b02eff3>
- Attribué à Lorenzo Vecchietta, *Triptyque : La Vierge et l'Enfant couronnée par deux anges (partie centrale) ; Saint Jean-Baptiste (volet gauche) ; L'Ange de l'Annonciation (au-dessus) ; Saint Jérôme (volet droit) ; La Vierge de l'Annonciation (au-dessus)*, XVe siècle, Musée du Petit Palais (Avignon), inv.Cl. 7492. Lien AGORHA :
<https://agorha.inha.fr/ark:/54721/6730de2b-0efc-417f-8ece-7321f7ad4d28>
- Copie anonyme d'après Raphaël, *La Grande Sainte Famille de François Ier*, début 18e siècle, Eglise Saint-Pierre (Méhoncourt). Lien AGORHA :
<https://agorha.inha.fr/ark:/54721/cdeeb0cb-ac77-4012-80bf-128fe93d0765>

Annexe D

Pixplot et Panoptic

I. Visualisation du corpus dans PixPlot

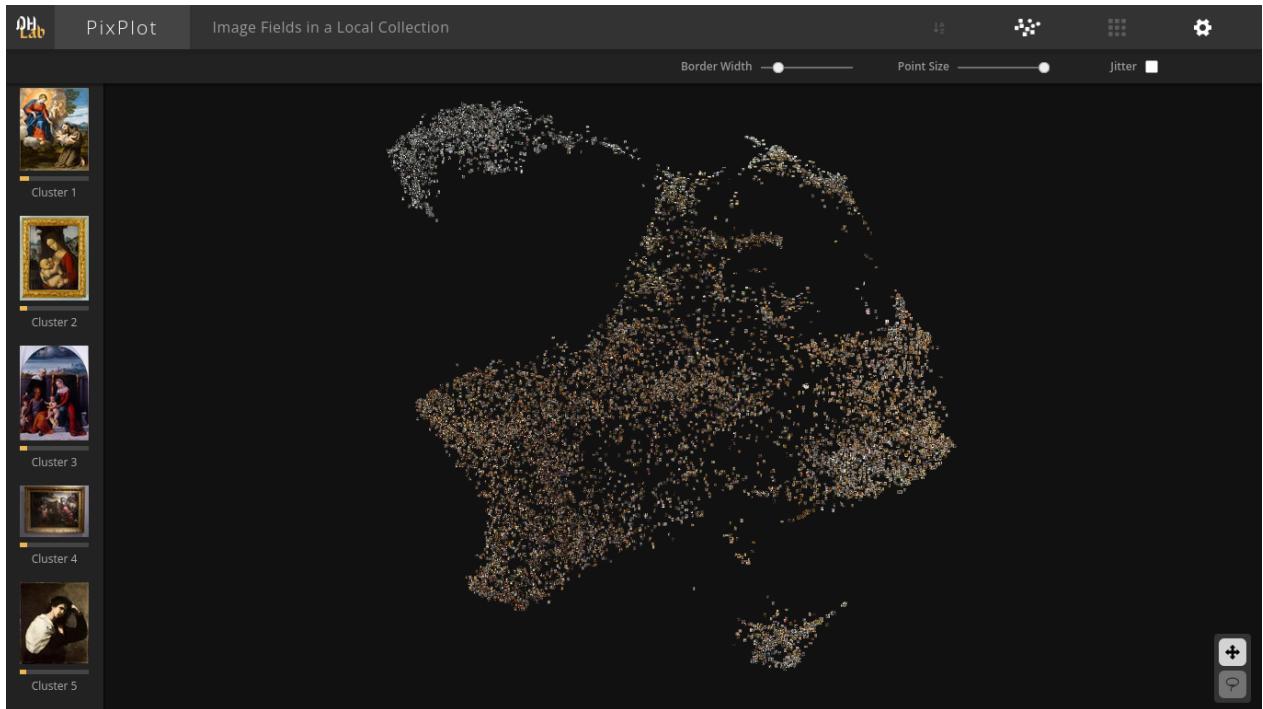


FIG. D.1 : Visualisation générale des tableaux du RETIF dans PixPlot.

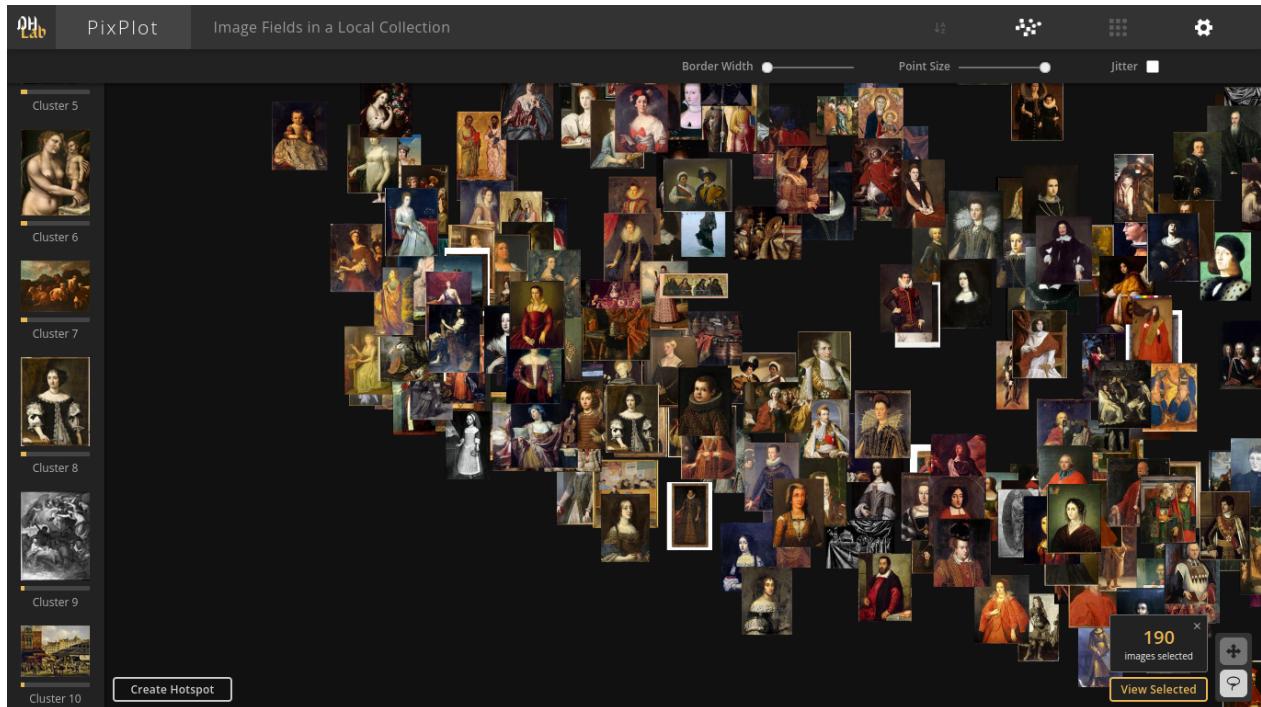


FIG. D.2 : Cluster de portraits identifié dans l'angle inférieur gauche de la visualisation du RETIF dans PixPlot.

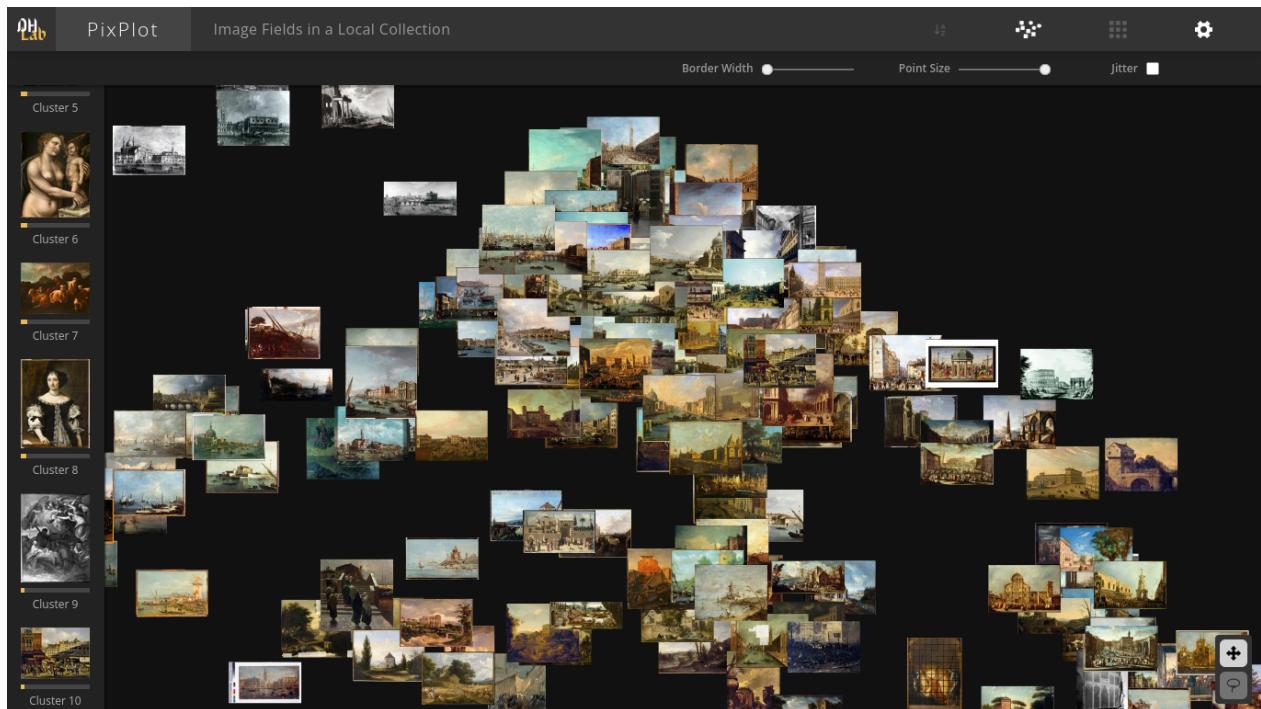


FIG. D.3 : Cluster de paysages urbains dans l'angle supérieur droit de la visualisation du RETIF dans PixPlot.

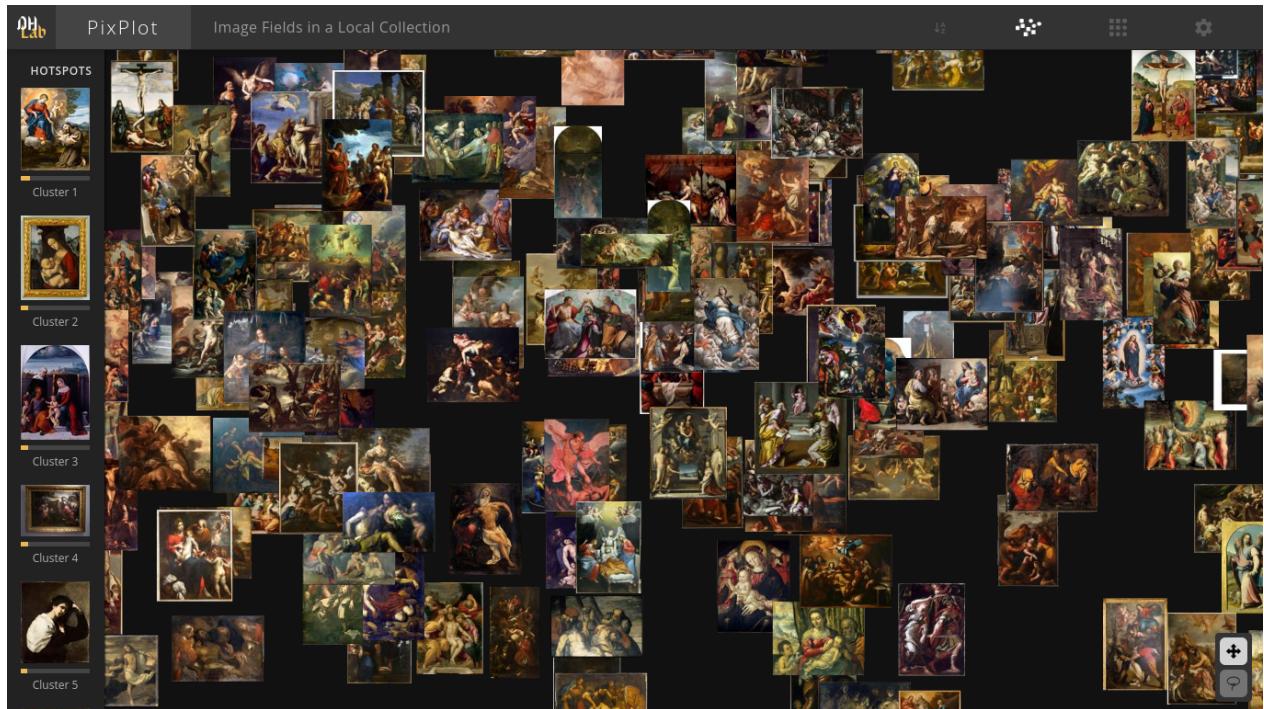


FIG. D.4 : Cluster des scènes religieuses et mythologiques au centre de la visualisation du RETIF dans PixPlot.

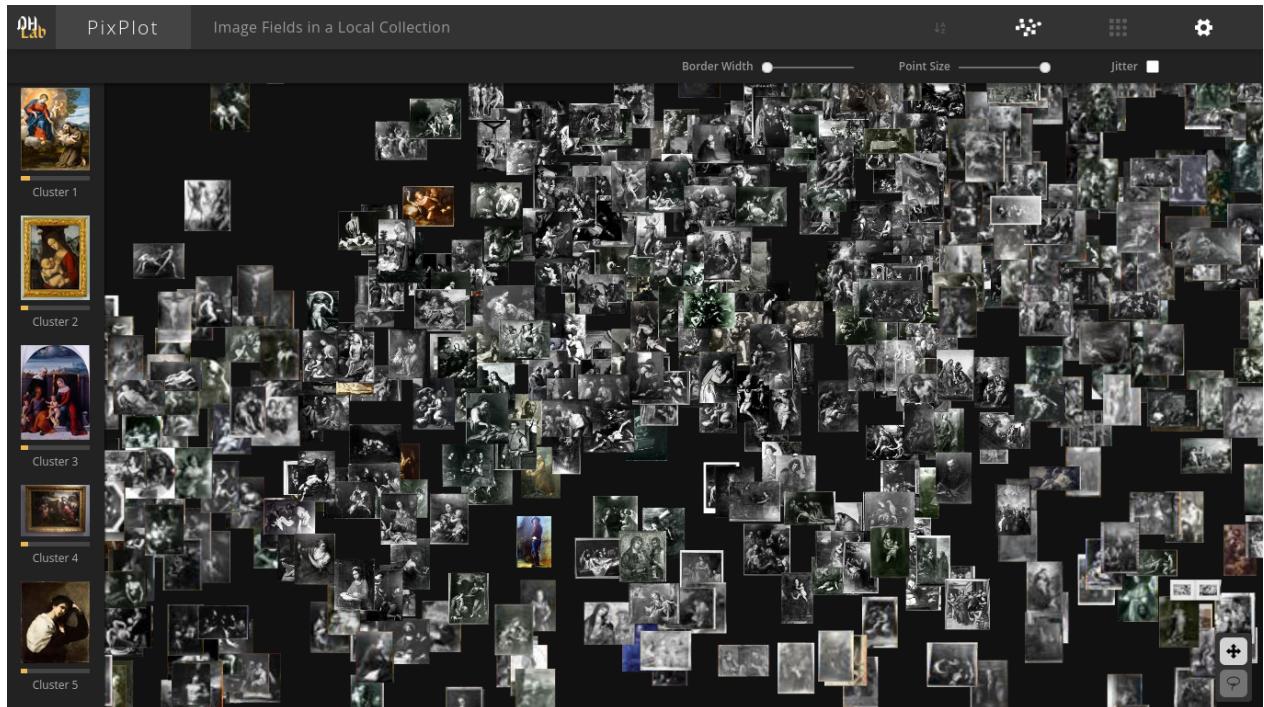


FIG. D.5 : Cluster des photographies en noir et blanc dans l'angle supérieur gauche de la visualisation du RETIF dans PixPlot.

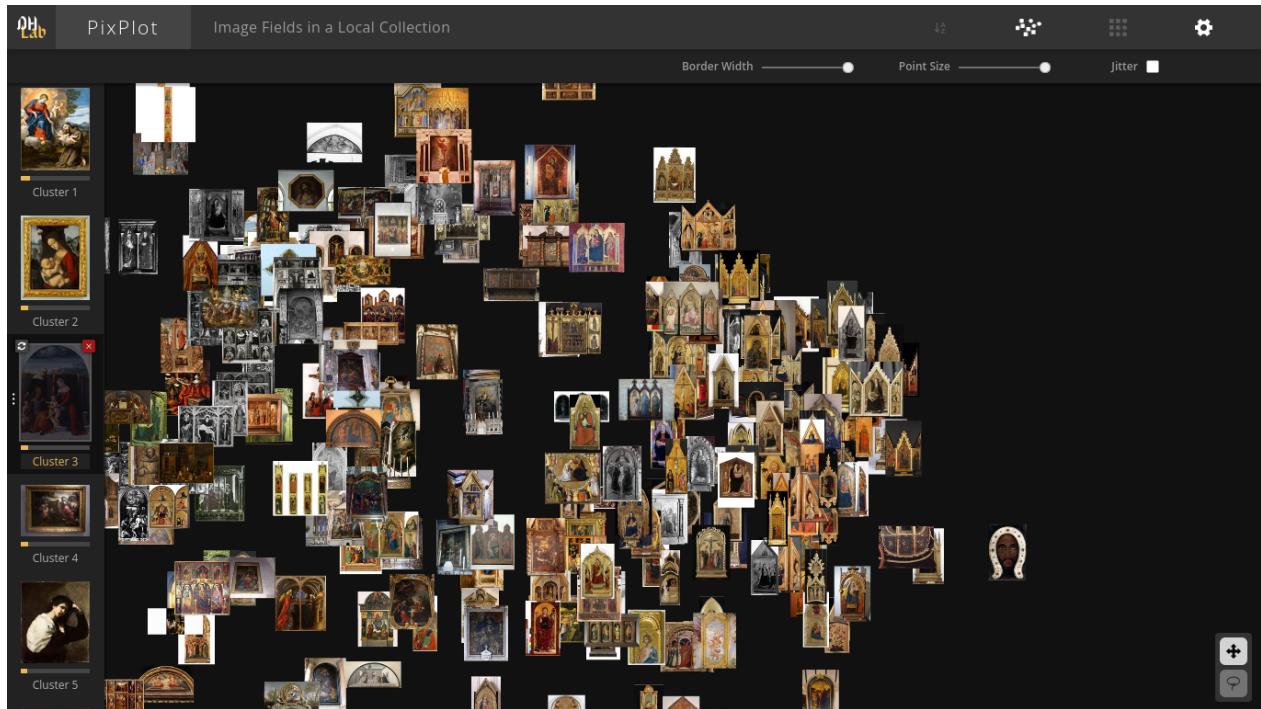


FIG. D.6 : Cluster des panneaux médiévaux et des tableaux dans leur contexte dans le bord droit de la visualisation du RETIF dans PixPlot.

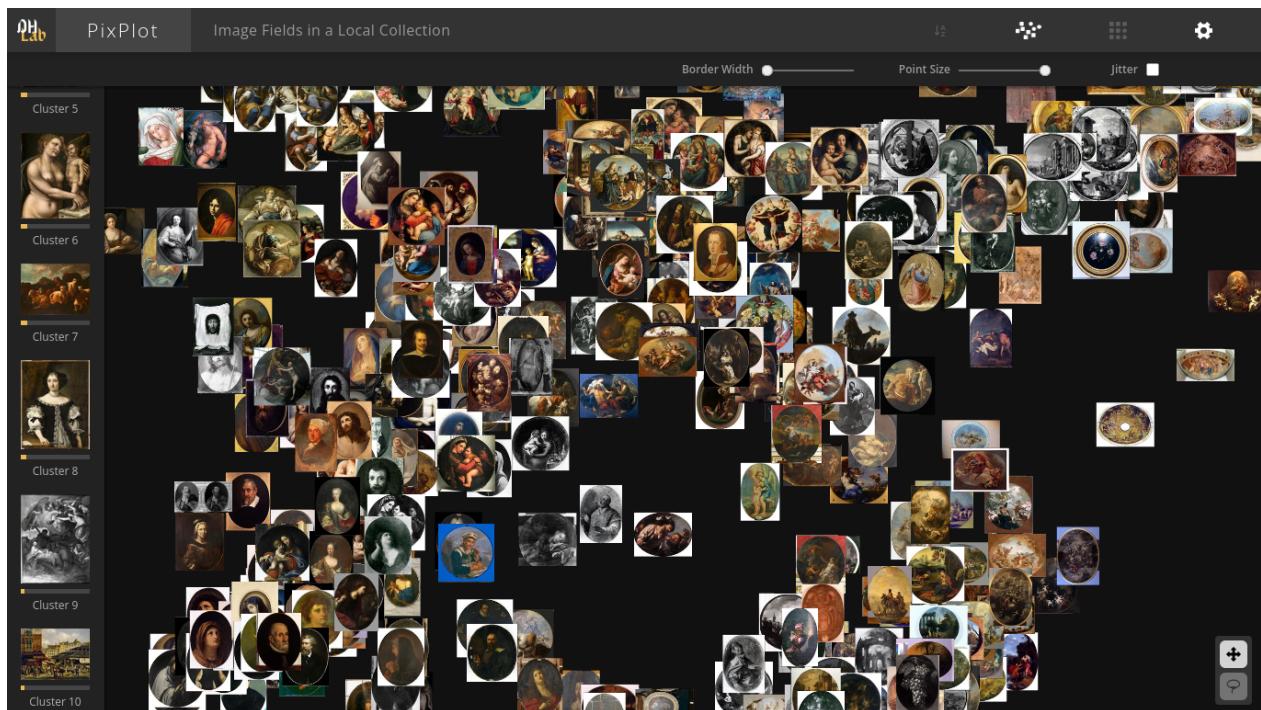


FIG. D.7 : Cluster des tableaux aux formats inhabituels (circulaire, ovale, hexagonal) dans l'angle inférieur droit de la visualisation du RETIF dans PixPlot.

II. Visualisation du corpus dans Panoptic

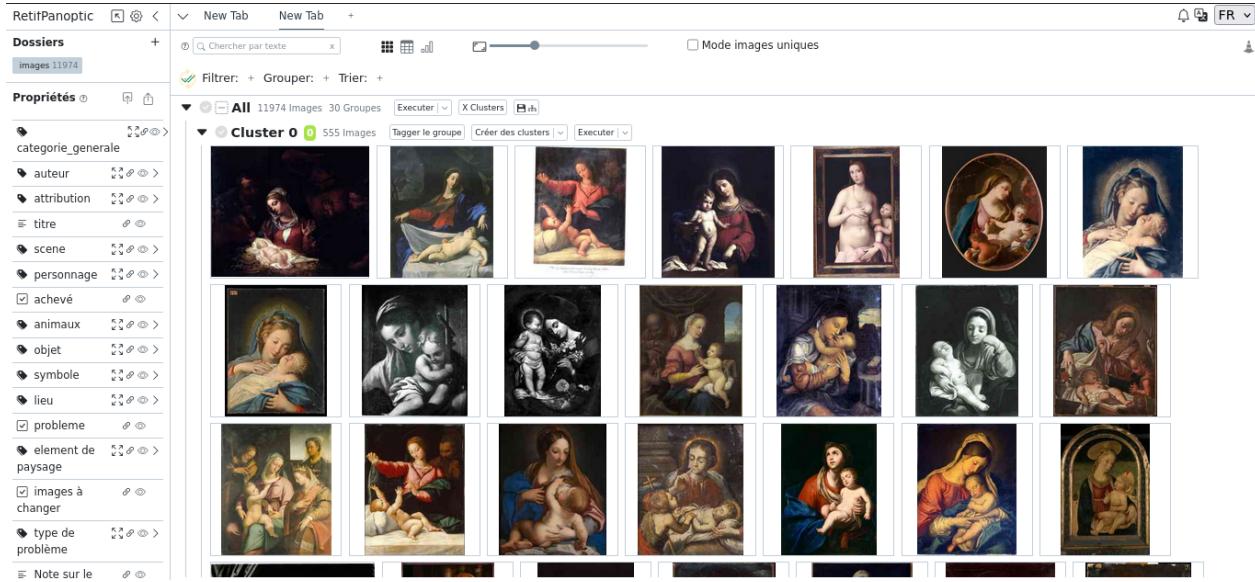


FIG. D.8 : Cluster des Vierge à l'Enfant dans Panoptic.

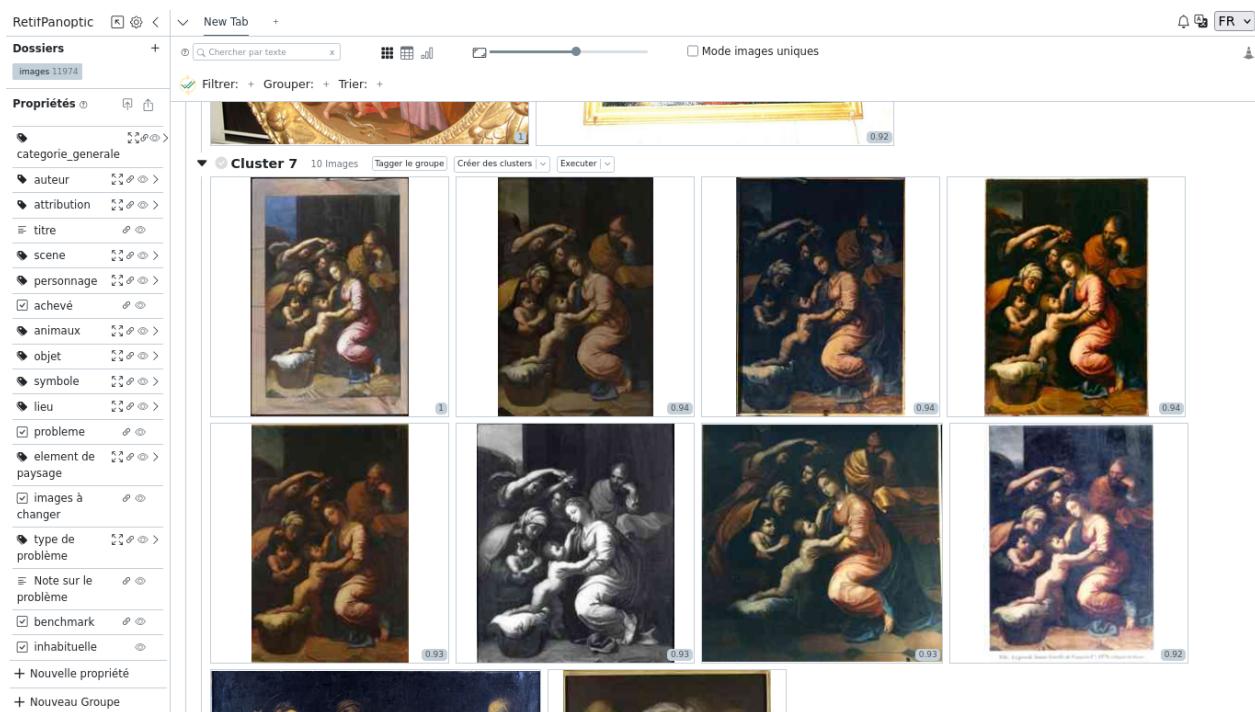


FIG. D.9 : Les copies de la Sainte Famille de François 1er de Raphaël rassemblées dans Panoptic

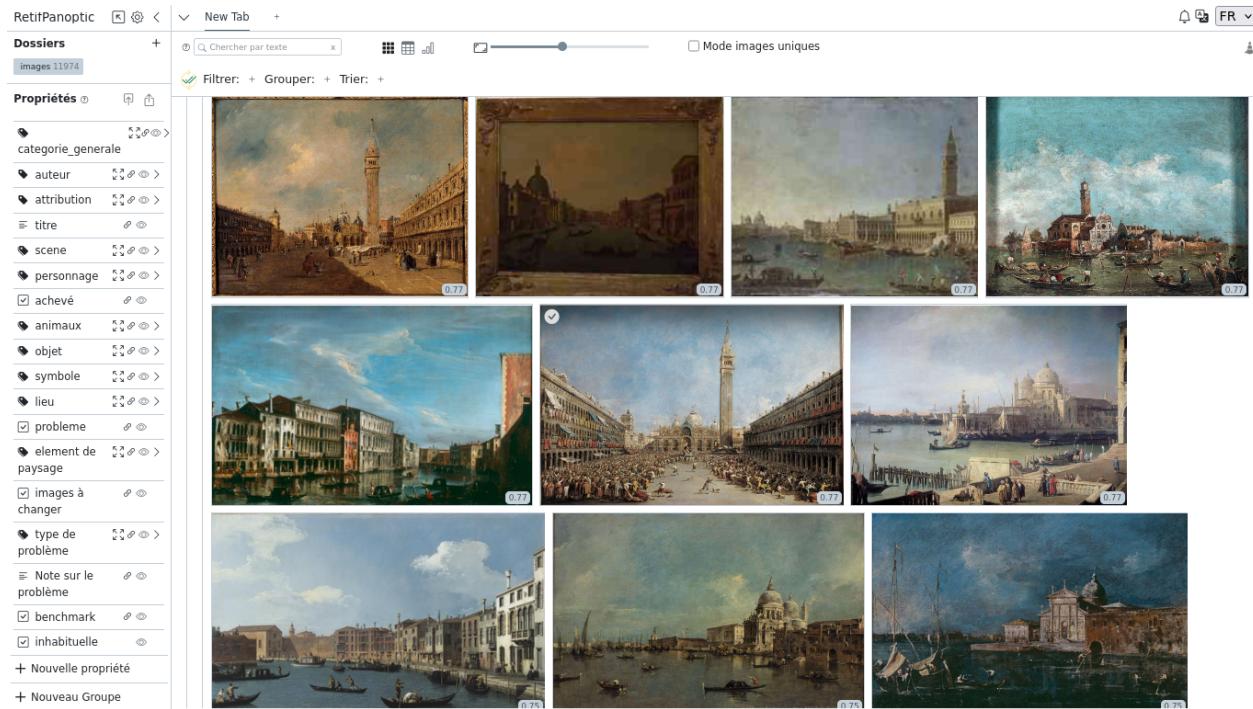


FIG. D.10 : Une recherche CLIP avec le terme "Venice" dans Panoptic

III. Statistiques après l'expérimentation d'indexation avec Panoptic

Ces graphiques sont disponibles au lien suivant : <https://public.tableau.com/app/profile/pierre.husson/viz/AGORHA-RETIF/UtilisationdeGarnierdansAgorha>

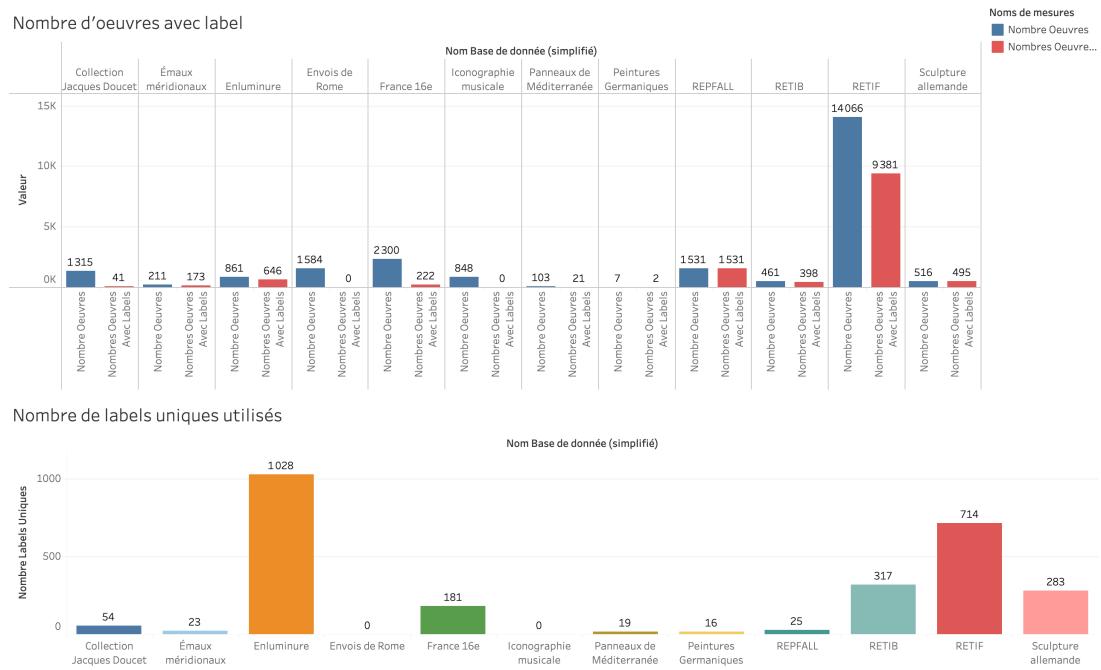


FIG. D.11 : Graphiques montrant le nombre de notices indexées avec au moins un label dans 12 bases d'AGORHA, et le nombre de labels distincts utilisés par bases.

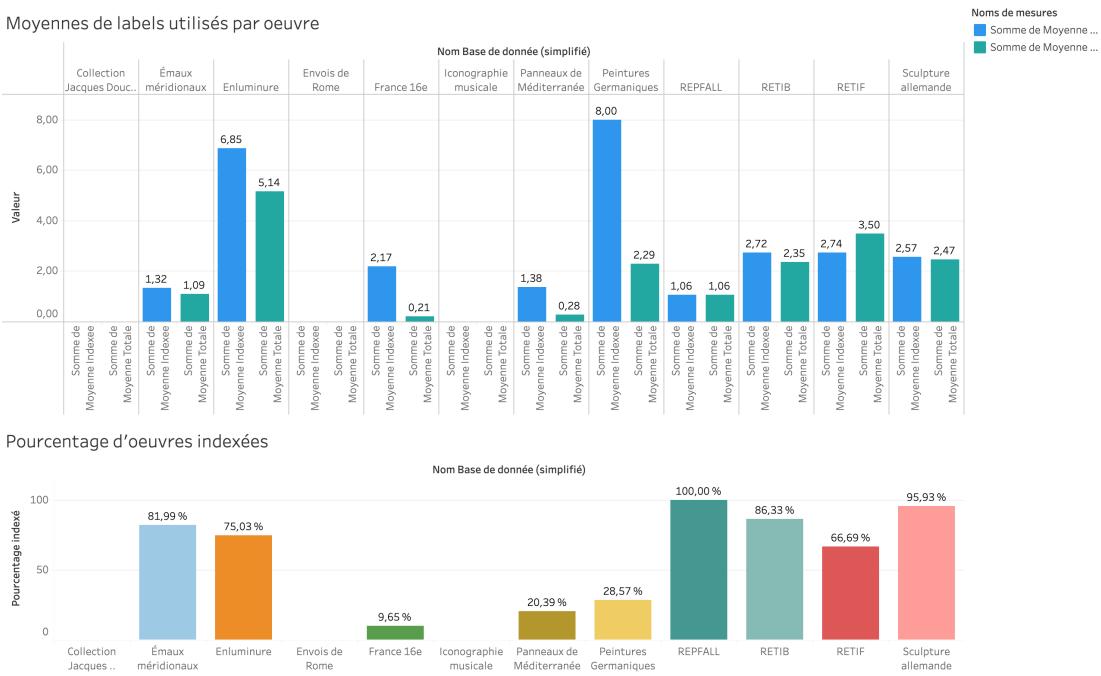


FIG. D.12 : Graphique montrant la moyenne de labels utilisés par notices dans 12 bases d'AGORHA, pour les oeuvres indexées et toutes les oeuvres ; et le pourcentage de d'oeuvres qui possèdent au moins un label par base de données.

Annexe E

Application Streamlit intégrant CLIP

I. Le code de l'application Streamlit

```
1 import streamlit as st
2 import os
3 import json
4 import clip
5 import torch
6 from PIL import Image
7 import numpy as np
8
9 # Configuration
10 st.set_page_config(layout="wide")
11 images_folder = "images/4" # à modifier si besoin
12 json_folder = "clip_json"
13
14 # Charger le modèle CLIP
15 @st.cache_resource
16 def load_clip_model():
17     device = "cuda" if torch.cuda.is_available() else "cpu"
18     model, preprocess = clip.load("ViT-B/32", device=device)
19     return model, preprocess, device
20
21 model, preprocess, device = load_clip_model()
22
23 # Charger les embeddings
24 @st.cache_data
```

```

25 def load_embeddings():
26     embeddings = []
27     for json_file in os.listdir(json_folder):
28         if json_file.endswith(".json"):
29             with open(os.path.join(json_folder, json_file), "r", encoding="utf-8") as f:
30                 data = json.load(f)
31                 embeddings.append({
32                     "filename": data["filename"],
33                     "embedding": np.array(data["embedding"]).astype("float32")
34                 })
35     return embeddings
36
37 data = load_embeddings()
38
39 # Interface
40 st.title("Recherche visuelle d'images du RETIF avec CLIP")
41 # Texte d'introduction
42 st.text(
43     "Bienvenue sur l'application Streamlit pour rechercher des images du RETIF
44     grâce à CLIP !\n"
45     "Assurez-vous d'avoir téléchargé les images et préparé les embeddings avec
46     CLIP."
47 )
48
49 # Recherche textuelle
50 with col1:
51     text_query = st.text_input("Recherche par mots-clés (ex : pyramid, dog...)")
52
53 # Recherche par image
54 with col2:
55     uploaded_image = st.file_uploader("Ou uploadez une image pour chercher par
56                                         contenu", type=["jpg", "jpeg", "png"])
57
58 # Résultats
59 if not text_query and not uploaded_image:
60     st.info("Entrez un mot-clé ou uploadez une image pour lancer une recherche.")

```

```

60 else:
61     if text_query:
62         with torch.no_grad():
63             tokens = clip.tokenize([text_query]).to(device)
64             text_features = model.encode_text(tokens)
65             text_features /= text_features.norm(dim=-1, keepdim=True)
66             query_vector = text_features.cpu().numpy()[0]
67             st.subheader(f"Résultats pour : *{text_query}*")
68
69 elif uploaded_image:
70     try:
71         pil_img = Image.open(uploaded_image).convert("RGB")
72         # Affichage dans la colonne 2
73         with col2:
74             st.image(pil_img, caption="Image envoyée", width='stretch')
75             image_input = preprocess(pil_img).unsqueeze(0).to(device)
76             with torch.no_grad():
77                 image_features = model.encode_image(image_input)
78                 image_features /= image_features.norm(dim=-1, keepdim=True)
79                 query_vector = image_features.cpu().numpy()[0]
80                 st.subheader("Résultats similaires à l'image")
81             except Exception as e:
82                 st.error(f"Erreur avec l'image envoyée : {e}")
83                 st.stop()
84
85     # Recherche par similarité
86     results = []
87     for item in data:
88         score = np.dot(query_vector, item["embedding"])
89         results.append((score, item["filename"]))
90
91     # Trier et afficher
92     results.sort(reverse=True)
93     cols = st.columns(5)
94     for idx, (score, filename) in enumerate(results[:50]):
95         with cols[idx % 5]:
96
97             st.image(os.path.join(images_folder, filename),
98                     caption=f"{filename}\nScore: {score:.2f}", width='stretch')

```

II. Visuels de l'application Streamlit

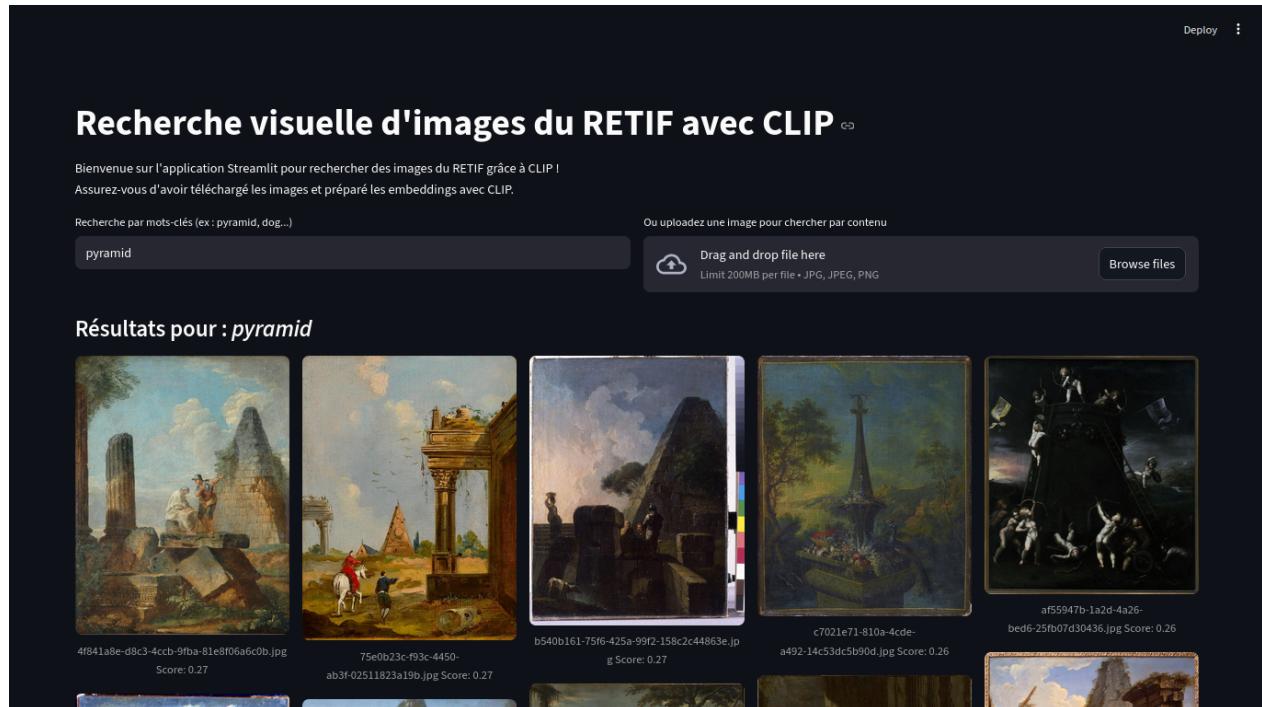


FIG. E.1 : Recherche du terme "pyramid" dans le moteur de recherche CLIP de l'application

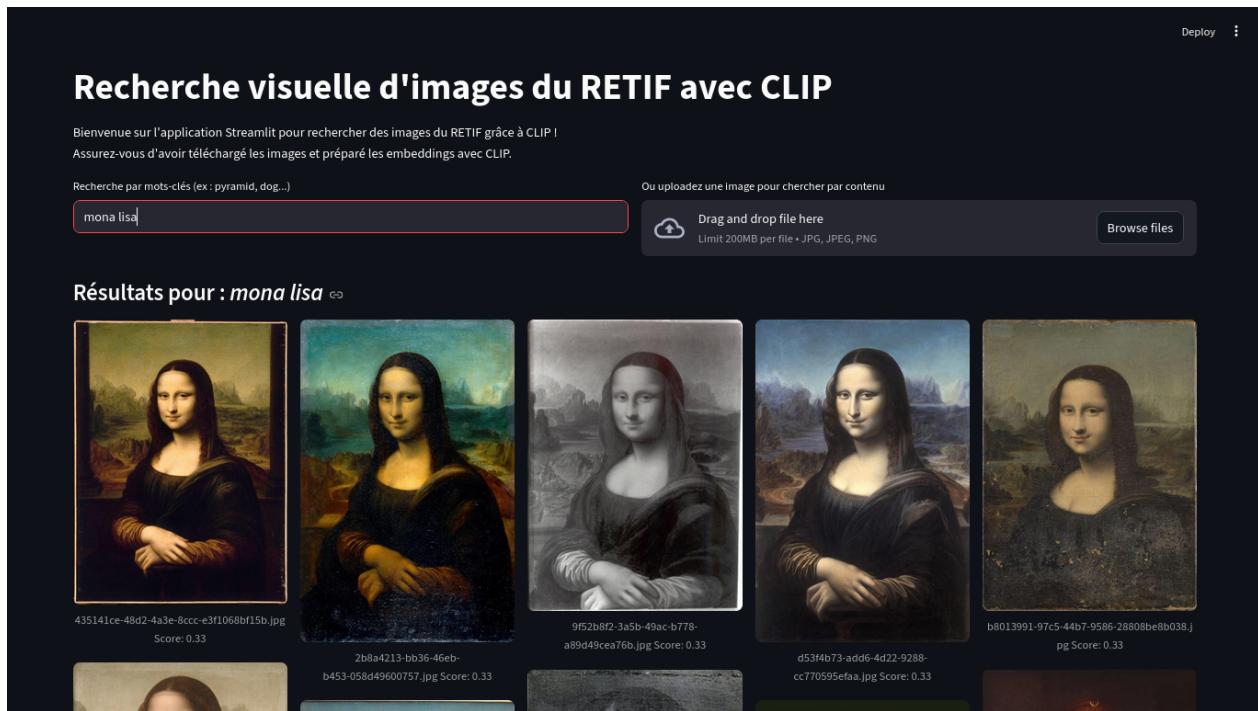


FIG. E.2 : Recherche des termes "mona lisa" dans le moteur de recherche CLIP de l'application

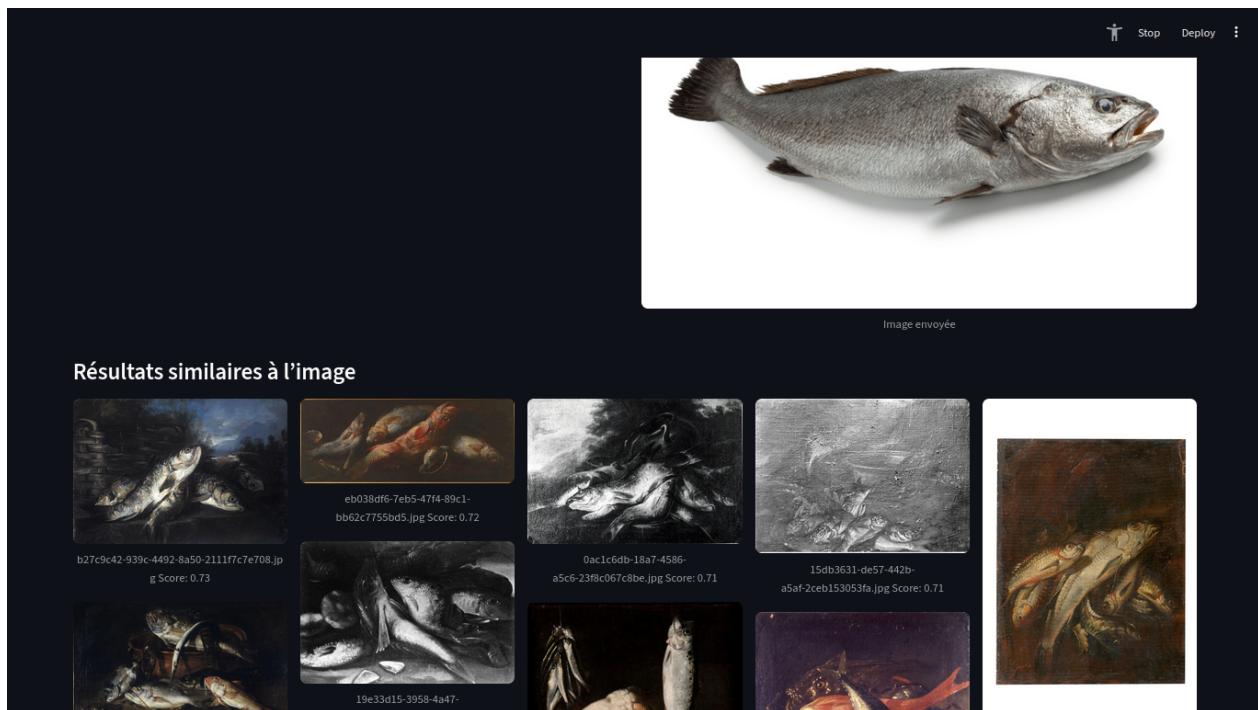


FIG. E.3 : Recherche par image à l'aide d'une photographie d'un poisson dans le moteur de recherche CLIP de l'application

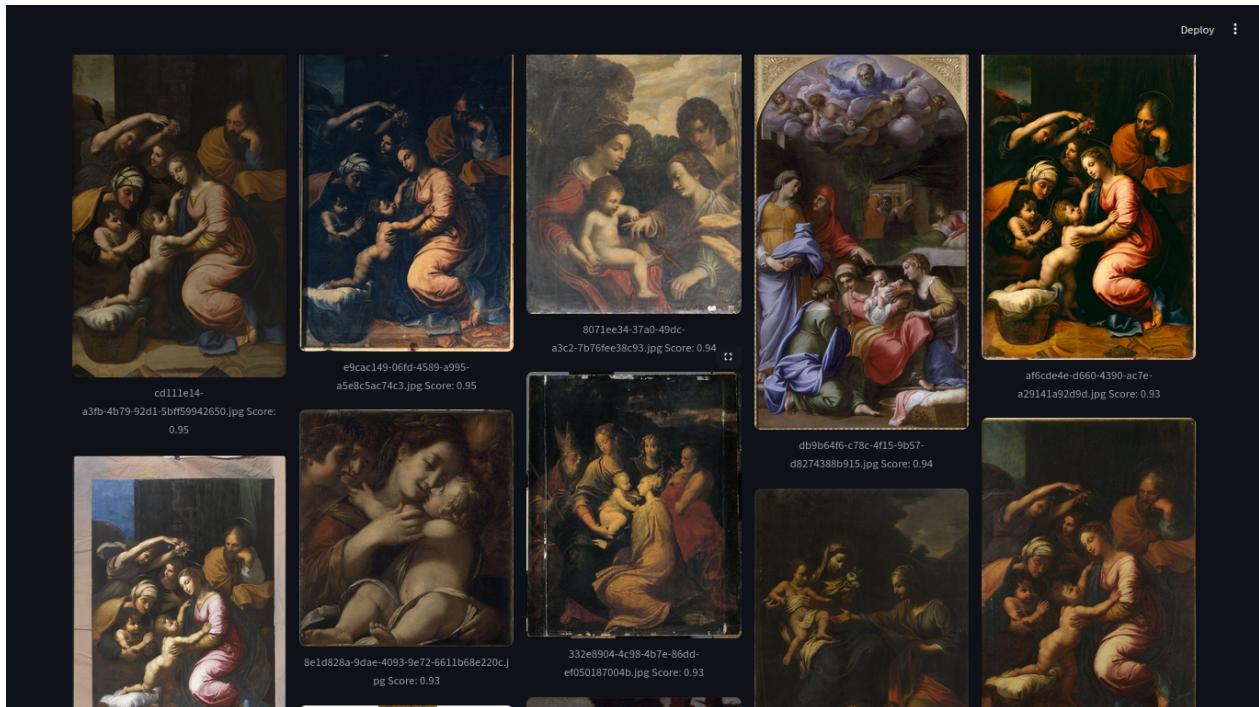


FIG. E.4 : Recherche par image à l'aide de la Sainte Famille de François 1er dans l'application (figure A.4).

Table des matières

Résumé	i
Avant-Propos	iii
Introduction	xv
I L'indexation iconographique des œuvres patrimoniales à l'ère du numérique	1
1 Indexation : principes et enjeux	3
I. L'indexation iconographique en histoire de l'art	3
II. Sujets, genre, motifs et éléments	5
III. Iconographie et Iconologie	6
2 Les instruments de l'indexation	9
I. Vocabulaires génériques et spécifiques	9
II. Systèmes descriptifs et systèmes d'indexation	13
III. Vers une interopérabilité des outils d'indexation	15
3 Indexation et numérique	19
I. Bases de données	19
II. Indexation et projets IA	23
II Enjeux techniques et méthodologiques de l'automatisation de la description iconographique de peintures	27
4 Les données du projet	29
I. Le corpus du RETIF	30

II.	Préparer le vocabulaire contrôlé	31
III.	Exploiter les métadonnées existantes	32
5	Expérimentations pour l'automatisation	35
I.	S'appuyer sur d'autres expériences	35
II.	Le vocabulaire contrôlé	38
III.	Tests de performances	39
IV.	Automatisation et histoire de l'art	41
6	Bénéfices et limites de l'automatisation	45
I.	Bénéfices	45
II.	Des limites	46
III.	L'impact environnemental	48
III	Alternatives à l'automatisation : des outils d'intelligence artificielle au service des pratiques d'indexation iconographiques	51
7	Faciliter l'indexation	53
I.	Des logiciels	53
II.	Exploration du RETIF	54
III.	Indexation avec Panoptic	56
8	Nouvelles méthodes d'exploration	61
I.	Multimodalité	61
II.	Amélioration de CLIP	63
III.	Une approche image-à-image	64
Conclusion		69
A	Le corpus du RETIF	73
I.	Exemples d'images provenant du RETIF	73
II.	Exemples d'images inexploitables	75
B	Le thésaurus Garnier dans AGORHA	77
I.	L'indexation dans AGORHA	77
II.	Code Python pour structurer le thésaurus Garnier en json	78

C Les modèles d'intelligence artificielles	81
I. Métriques par classe pour le modèle YOLO entraîné sur des natures mortes	81
II. Matrice de confusion du modèle FastAI pour la reconnaissance du sujet iconographique	82
III. Étude comparative des modèles	83
IV. Liste des peintures	83
D Pixplot et Panoptic	87
I. Visualisation du corpus dans PixPlot	87
II. Visualisation du corpus dans Panoptic	91
III. Statistiques d'indexation	92
E Application Streamlit intégrant CLIP	95
I. Le code de l'application Streamlit	95
II. Visuels de l'application Streamlit	98