



HAL
open science

L'image comme particule élémentaire, ou les prémisses d'un changement d'échelle à la BnF

Emmanuelle Bermès, Céline Leclaire, Jean-Philippe Moreux

► To cite this version:

Emmanuelle Bermès, Céline Leclaire, Jean-Philippe Moreux. L'image comme particule élémentaire, ou les prémisses d'un changement d'échelle à la BnF. *The Measurement of Images. Computational Approaches in the History and Theory of the Arts*, sous la direction de Clarisse Bardiot et Emmanuel Château-Dutier, Presses universitaires du Septentrion, A paraître, Humanités numériques et science ouverte. <hal-03991515>

HAL Id: hal-03991515

<https://hal.science/hal-03991515v1>

Submitted on 17 Mar 2023

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

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



Distributed under a Creative Commons CC BY 4.0 - Attribution - International License

L'image comme particule élémentaire, ou les prémisses d'un changement d'échelle à la BnF

Introduction

Le travail sur les images passionne les bibliothécaires et les chercheurs de la Bibliothèque nationale de France depuis des siècles. Il s'est étendu au fur et à mesure de l'intégration effective de nouveaux médias dans le périmètre du dépôt légal : les estampes, cartes et plans en 1648, les affiches en 1941, internet en 2006¹, et les productions numériques en 2021 (« signes, signaux, écrits, images, sons ou messages de toute nature mis à la disposition du public ou de catégories de public lorsqu'ils leur sont librement accessibles »²), pour n'évoquer que quelques étapes clés concernant les images fixes.

En 2021, la BnF s'est dotée d'une feuille de route pour le déploiement de l'intelligence artificielle³. Cette initiative correspond à un moment de bascule dans le domaine du traitement automatisé des images : après une dizaine d'années d'explorations techniques et d'expérimentations, l'établissement s'achemine vers une implantation généralisée de ce mode d'approche dans l'ensemble de Gallica, sa bibliothèque numérique et celle de ses partenaires. Ce premier tournant est soutenu par des développements similaires dans le monde de la recherche, en particulier au sein des sciences humaines et sociales, où les travaux d'exploration des images se multiplient : étude de la représentation des femmes sur les pages de couverture du magazine *Paris Match* (1949-1959)⁴, analyse de l'iconographie de la rubrique « cinéma » des quotidiens français des années 1920⁵, décryptage des motifs floraux des collections de papiers peints et de textiles... Cette aire de jeux très vaste, propice aux partenariats, contribue à un autre mouvement : celui du remplacement des logiques de séparation disciplinaires par des logiques de recomposition qui effacent – ou troublent – les frontières canoniques entre les disciplines, entre les supports, que ce soit à travers l'étude de documents complexes (tels que la presse, les documents fortement structurés comme les dictionnaires ou les catalogues de ventes, ou encore les ouvrages illustrés, auxquels plusieurs techniques d'analyse doivent pouvoir être appliquées), ou le regroupement de documents de types complètement différents. Ainsi, de nombreux projets de recherche pluridisciplinaires tentent d'identifier et de relier les fonds iconographiques avec d'autres sources documentaires (archives, annuaires et catalogues, bases muséales). Enfin – et ce troisième

¹ Voir <https://www.bnf.fr/fr/le-depot-legal-une-mission-cle-pour-la-bnf>

² LOI n° 2021-1901 du 30 décembre 2021 visant à conforter l'économie du livre et à renforcer l'équité et la confiance entre ses acteurs, <https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000044637892>

³ Voir <https://www.bnf.fr/fr/feuille-de-route-ia>

⁴ Projet ANR Numapress (2019)

⁵ Voir

https://docs.google.com/presentation/d/158Yw69D_kSmE5t5KEWvxaHal7u1asfOnsiKLr3EOzww/edit#39

tournant est lié aux précédents – si les corpus soumis aux traitements automatisés des images visaient à mettre la technologie à l'épreuve et respectaient une cohérence du point de vue de l'époque de production des images, de la technique de leur élaboration (dessin, gravure...), ou du support de leur diffusion, il est aujourd'hui permis d'espérer qu'une application moins (voire pas du tout) supervisée de la technologie à l'ensemble de la collection fasse émerger des corpus d'un nouveau type, *a posteriori*, c'est-à-dire de nouveaux rapprochements, iconoclastes, jusqu'à présent non envisagés.

Quels sont les défis à relever, d'un point de vue scientifique, technologique, matériel, organisationnel, pour aller jusqu'au bout de cette dynamique et faire advenir ce temps des corpus renouvelés ? En analysant la notion même d'image et les modalités de fouille qui lui sont attachées, en observant la manière dont l'image est devenue le limon qui fertilise plus que jamais l'ensemble de Gallica, nous verrons que bougent également les frontières de la bibliothèque tout entière, celles de la notion de collection, de métier, d'accès à la connaissance, tout comme l'insertion de cette action dans un écosystème qui la dépasse.

1. Extraire : la BnF, source d'images

Lorsqu'une image intègre les collections de la BnF, qu'elle ait été collectée dans le cadre du dépôt légal ou produite par la BnF et ses partenaires dans le cadre de la numérisation, elle devient le maillon d'une complexité que la définition informatique basique des images – une matrice en deux dimensions d'intensités RVB (rouge, vert, bleu), un fichier dont chaque pixel est défini par un nombre compris entre 0 et 255 – ne parvient pas totalement à réduire, pas plus qu'elle ne parvient, d'ailleurs, à rendre compte de certaines couleurs subtiles...

Un ensemble évolutif et polymorphe

À travers la numérisation, la Bibliothèque est source d'un nombre immense de nouvelles images, qui constituent autant de vues des objets matériels qu'elle conserve, quelle que soit leur nature. En effet, le choix effectué dès les années 1990 de la numérisation « en mode image » [Bermès, 2020] c'est-à-dire sous forme de fac-similés a conduit, y compris pour des documents considérés communément comme des textes, à produire des images pour chaque page, images qui doivent ensuite faire l'objet de traitements tels que l'OCR⁶ pour en extraire le contenu. L'image de la page, la « vue » numérisée, est ainsi l'unité minimale de consultation et de traitement des documents textuels, imprimés ou manuscrits, et implique un outillage adéquat. La détection de contenus iconographiques, dans un sens plus restreint, sur ces « vues » est, nous le verrons, l'un des enjeux majeurs des développements actuels.

L'ouverture qui préside au dépôt légal, l'absence de jugement de valeur (qu'il soit moral, esthétique ou social) quant aux documents appelés à rejoindre les collections [Bermès, 2018_2019], la diversité de l'accroissement des collections (dons, acquisitions) et l'évolution naturelle de la création éditoriale et artistique se reflètent dans la diversité des corpus numérisés accessibles dans Gallica : ouvrages scientifiques comprenant des diagrammes, des images, des formules, des tableaux ; titres de presse comprenant des photographies et des publicités ; manuscrits enluminés, cartes et affiches mêlant texte et illustrations, sceaux, médailles, vues d'objets en trois dimensions tels que les maquettes et les costumes de décors d'opéra.... Aujourd'hui, les frontières de cet ensemble s'ouvrent aux bannières de blogs, aux photos des banques d'images en ligne, aux fichiers correspondant aux affiches de très grand format des couloirs du métro, aux calques qui composent une œuvre créée avec le logiciel InDesign, et, le cas échéant, aux NFT⁷.

Aujourd'hui, la BnF offre même à ses publics le loisir de produire leurs propres vues grâce aux tentes de numérisation qu'elle met à leur disposition, contribuant, ainsi que le faisait remarquer Antoine Courtin lors du festival de l'histoire de l'art à Fontainebleau en 2022, à un vaste mouvement de production banalisée et démocratique des contenus iconographiques les plus

⁶ Génération automatique de contenu textuel à partir d'images numériques par reconnaissance optique de caractères (*Optical Character Recognition*)

⁷ Jetons non fongibles, titres de propriété associés à des œuvres d'art et fondés sur la technologie de la chaîne de blocs

divers, et ceci, sans qu'un support en remplace jamais un autre. Un tel étagement est nourri par la volonté d'élargir l'accès à ces contenus, dans l'immédiat en poursuivant activement la politique de collecte et de numérisation, et dans le futur, gageant que les technologies garantiront les modalités d'une lecture aujourd'hui impossible de certains contenus.

Cet article envisage donc la notion d'image dans un sens large, englobant les objets numériques définis ci-dessus.

L'enjeu de la patrimonialisation de contenus irréductibles

L'entreprise de développement de la collection d'images n'a d'intérêt que parce qu'elle accompagne son objet d'un ensemble d'enrichissements tels que le contexte de son élaboration et surtout les métadonnées, les liens vers des standards de description. C'est un ressort essentiel de l'opération de patrimonialisation [Bermès, 2020]. L'image n'est jamais envisagée seule, de manière absolue : la grande force de la BnF est qu'elle conserve des images environnées de texte. Plus qu'à la traduction informatique de l'image, la BnF s'intéresse à l'information qu'elle véhicule. C'est ce qui permet de rêver par exemple que la presse nous révèle des photographies d'auteurs et de personnes dont nous ne connaissons aucune représentation aujourd'hui... De ce point de vue, les différents documents de la collection se répondent, de même que les différentes strates historiques de son traitement.

Comme dans le cas des autres documents, il s'agit donc pour la Bibliothèque d'apprendre à traiter des volumes croissants de contenus et de faire en sorte qu'ils se répondent les uns les autres, mais le cas de l'image se pare d'une autre forme de complexité au moment de sa mise en données. Quelque chose résiste. En effet, comme l'a montré Jolan Wuyts [Wuyts, 2021], les campagnes d'annotation menées dans le cadre de projets de recherche visant à développer des systèmes de vision par ordinateur se heurtent à des formes de subjectivité, par exemple lorsqu'il s'agit de déterminer le genre architectural d'un monument construit selon plusieurs strates temporelles, d'apporter des nuances dans une classification, d'éviter la confusion sur des genres visuellement proches, d'identifier des visuels qui relèvent d'une notion aussi riche que la « publicité », etc. L'image est sujette à interprétation, particulièrement dans le cas d'une collection d'une telle diversité et profondeur chronologique. Les premiers projets conduits par la BnF sur l'indexation des collections iconographiques, comme le projet STITCH (2006-2007) qui portait sur l'alignement de vocabulaires de description d'enluminures [Angjeli et Isaac, 2008] ou le projet Classification d'images patrimoniales (Inria, BnF) qui portait sur l'analyse par apprentissage profond de manuscrits enluminés [CIP, 2019], ont montré que des interprétations relativement courantes (par exemple distinguer un lion d'une chimère, identifier un personnage ou une scène biblique) étaient déjà des défis pour des opérateurs humains : que penser de leur automatisation par un outil d'intelligence artificielle...

Du document « image » à l'unité iconographique

Si ce qui précède suggère que l'image ne peut être réduite à elle-même, il suggère aussi les limites des enrichissements classiques, textuels, en matière de métadonnées. Naît l'idée d'accorder à l'image une forme d'autonomie documentaire, de l'extraire du document dont elle fait partie, comme une particule qui constituerait un fondement élémentaire de la collection. Nous sommes ici face à ce double mouvement, quasiment perpétuel à la Bibliothèque, entre traitements par flux et traitements à l'unité, approches de masse et approches spécialisées, qui se nourrissent les uns les autres. Afin que la collection existe, il faut en identifier les unités au préalable. C'est le standard IIIF⁸ qui relèvera ce défi : il permet de modéliser les coordonnées et le contenu d'un élément iconographique sur une page. Si Gallica compte 250 millions de vues issues de la BnF et de 250 partenaires, 50 millions d'images actuellement enfouies dans les documents pourraient être mises au jour et rendues accessibles grâce à la conjonction de l'IA et de IIIF. Alors, un titre tel que le magazine *Vogue* pourra devenir une collection d'images à part entière⁹. L'avènement de l'unité iconographique fait sortir de l'ombre des images qui n'ont jamais circulé ni fait l'objet de traitements documentaires en tant qu'objets indépendants et autonomes. S'ouvre plus largement que jamais la possibilité d'analyser leurs différents statuts et usages.

⁸ <https://iiif.io>

⁹ Le prototype GallicaPix a permis d'identifier 42 000 images dans les fascicules de la version française de *Vogue*. Voir <https://gallicapix.bnf.fr/rest?run=findIllustrations-form.xq>

Un troisième sens du terme « corpus » peut alors apparaître : non uniquement « ensemble présélectionné » ou « ensemble nouveau qui surgit », le corpus est aussi un ensemble dans lequel les différentes unités se renforcent, où un élément peut conférer du sens à son voisin via une analyse comparative, la mobilisation de métadonnées... Cet effet puissant du rassemblement n'a jamais été aussi bien exploité qu'avec l'image. Et nous revenons au fondement même de la collection de bibliothèque : le tout représente davantage que la somme de ses parties.

2. La Bibliothèque comme réceptacle outillé, centre d'interprétation de l'image

Non contente de poser le concept d'unité iconographique, c'est-à-dire d'élément irréductible susceptible de faire l'objet d'une recherche à part entière, quelle que soit son origine (un document iconographique autonome ou un ensemble où il figure en tant qu'illustration), la Bibliothèque envisage aussi, pour reprendre les mots de Jean-Philippe Moreux [2019], d'entrer de plain-pied dans l'image, soit de se confronter à de nouveaux modes de « lecture » – ou, en l'occurrence, de « vision ». L'outillage nécessaire pour cela, et surtout l'évolution de cet outillage, suggèrent cette fois encore à quel point l'inscription des images dans une collection totale constitue une richesse, et la manière dont l'approche visuelle et l'approche textuelle d'un contenu nourrissent un dialogue continu. Opérons donc à présent, en quelque sorte, la traduction technologique du processus de patrimonialisation décrit en première partie.

La vision par ordinateur à la BnF : brève définition et principaux enjeux

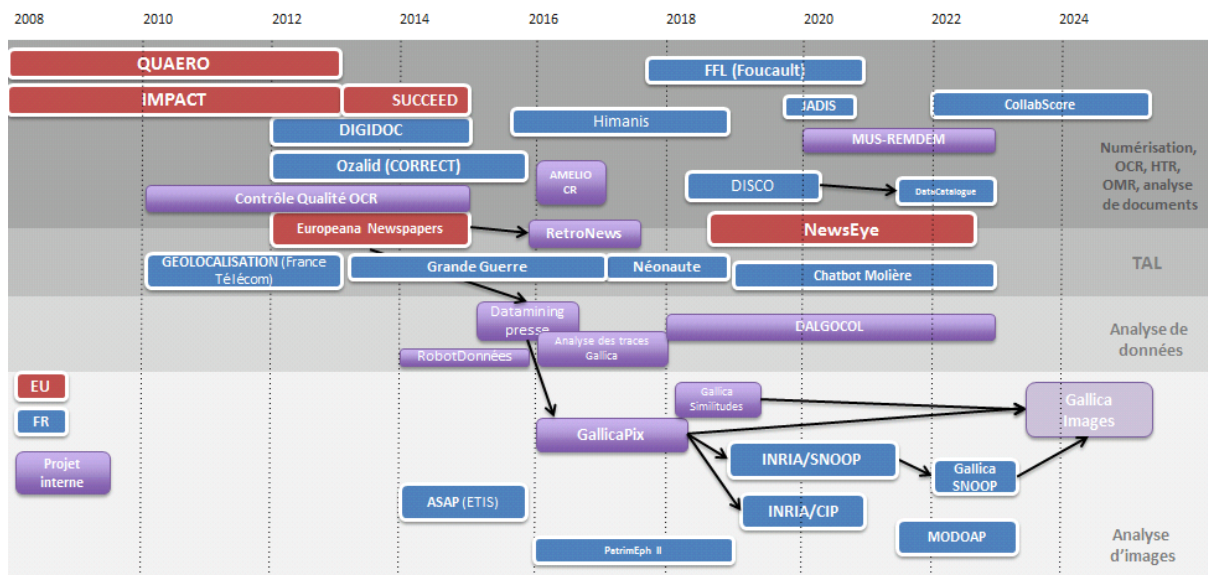
La lecture assistée par ordinateur d'une image donnée peut s'appuyer d'une part sur les caractéristiques visuelles de cette image (signature générée par extraction d'indices visuels) qui seront le fondement de la recherche d'images similaires à cette image, d'autre part sur des index sémantiques consistant en des annotations de cette image (ou « étiquettes », alimentées par les métadonnées de l'image ou générées automatiquement) et rendant possibles les recherches selon des concepts tels que « chat », « voiture »... [Moiraghi et Moreux, 2018].

L'usage de ces modes de vision est essentiel pour permettre à la BnF d'atteindre ses deux principaux objectifs dans le domaine de l'image : améliorer le traitement des collections et améliorer ses services aux publics (et aux chercheurs en particulier), c'est-à-dire leur offrir une lecture augmentée, un hyper-œil qui s'ajoute à l'œil organique et soutient leur faculté à extraire des connaissances [Bermès et Moiraghi, 2020]. Les travaux scientifiques de la Bibliothèque dans ce domaine s'orientent dans plusieurs directions, au-delà de l'extraction même des unités iconographiques : automatisation de certains traitements, application de technologies comme l'apprentissage machine (*machine learning*), développement d'interfaces, développement de fonctionnalités de recherche...

De la reconnaissance optique de caractères à la fouille d'images

Les avancées technologiques qui assoient aujourd'hui la fouille d'images à la BnF ont été précédées de nombreux développements, explorations et expérimentations, dont nous présentons ici les principales étapes.

Le schéma panoramique ci-contre dresse la liste des projets de recherche visant l'exploration des collections numériques depuis 2008, organisés selon les technologies mobilisées (de haut en bas : l'OCR et l'analyse de documents, le traitement automatisé de la langue ou TAL, l'analyse de données et l'analyse d'images).



Un tel schéma illustre tout d'abord le rôle essentiel des partenariats nationaux et internationaux (respectivement colorés en bleu et rouge), mais encore la manière dont divers projets ont précédé l'analyse outillée des images, sans substitution d'une technique par une autre : non seulement différentes approches sont nécessaires pour faire honneur à la richesse et à la complexité des documents numériques conservés, mais une technique donnée peut également devenir le socle de développements futurs dans un autre domaine. Ainsi, l'OCR appliqué à grande échelle par les bibliothèques depuis les années 2000 implique une étape de segmentation de la page et l'identification des éléments qui composent sa structure : titres, blocs de textes, éléments décoratifs, illustrations¹⁰. Il en résulte un gisement considérable de signalement de ressources iconographiques propice à nourrir des outils d'analyse et de fouille d'images [Moreux, 2019]. Enfin, il faut également du temps à une technologie pour s'intégrer naturellement aux processus de travail. Dans le cas de l'OCR – première technologie issue de l'intelligence artificielle à rejoindre, via une internalisation, les dispositifs informatiques régulièrement employés par la BnF [Bermès et Moiraghi, 2020] – ce temps a représenté une dizaine d'années, et il pourrait en aller de même pour les techniques de fouille d'image.

Pour autant, le traitement des collections d'images s'est lui aussi avéré moteur dans le cheminement général de la BnF vers l'intégration de l'intelligence artificielle dans ses processus et outils, tant la valeur ajoutée de la promesse de l'analyse sémantique des images était évidente. Ainsi, l'un des tout premiers projets, mené avec l'université de Cergy et le laboratoire ETIS¹¹ (ASAP, 2014-2016), concernait la banque d'images de la BnF et visait à en indexer les contenus. Il s'agissait de la première application d'un réseau de neurones à la BnF. Si ses résultats ont pu apparaître comme décevants d'un point de vue bibliothéconomique – puisqu'il a échoué à apporter des éléments sémantiques convaincants, susceptibles d'enrichir le catalogage – ils ont enthousiasmé les développeurs, qui ont pu améliorer leur connaissance du fonctionnement de leur algorithme [Picard et Gosselin, 2015]. Ce projet correspond à une première phase qui a consisté à prendre conscience de l'intérêt d'une approche par réseaux de neurones et apprentissage machine, même si, en 2015, le taux de réussite n'était que de 20%¹². Il a également montré l'importance d'expérimentations variées, visant à améliorer la technologie, avant de pouvoir envisager son industrialisation. Une autre étape clé a été franchie en 2017 grâce à l'élaboration de GallicaPix, prototype d'un moteur de recherche sémantique dans les images fonctionnant selon le principe de l'indexation (via l'analyse d'images) et dont l'enjeu était de satisfaire des cas d'usage classiques. D'autres prototypes ont suivi dans le même but –

¹⁰ On parle alors plus spécifiquement d'OLR – *Optical Layout Recognition* ou reconnaissance optique de mise en page

¹¹ Voir https://actions-recherche.bnf.fr/bnf/anirw3.nsf/IX01/E2015000032_equipements-traitement-de-l-information-et-systemes

¹² Ce qui signifie par exemple que la machine, une fois entraînée à classer des images dans une ontologie, ne parvenait à classer correctement qu'une seule image sur cinq

vérifier la capacité de modèles pré-entraînés à s'appliquer aux collections de la BnF – en particulier GallicaSimilitudes [Moiraghi et Moreux, 2018] puis GallicaSnoop, qui sont tous deux des moteurs de recherche par similarité visuelle. Cette démarche a pu s'imposer grâce aux progrès des techniques d'apprentissage profond supervisé au cours de la seconde moitié des années 2010 : ces prototypes s'appuyaient sur des modèles d'entraînement et un outillage technologique issus d'acteurs privés tels que Google et Amazon, suffisamment stables pour envisager un déploiement en production dans le système d'information.

Pour compléter cette équipée, l'année 2017 a vu la commande au laboratoire L3i de l'université de La Rochelle d'un état de l'art sur les outils de recherche d'image [Sidère, 2018]. Il aura montré que les approches neuronales étaient matures mais qu'il n'existait pas de progiciels sur étagère, *a fortiori* pour des contenus patrimoniaux.

Cinq approches fondamentales

Les différents projets évoqués ont permis d'identifier cinq situations ou approches susceptibles de guider la vision par ordinateur et l'analyse d'image. Même si c'est le croisement de ces approches qui favorise la qualité de l'interface homme-machine, la diversité des recherches et surtout leur inventivité, nous allons ici les envisager séparément afin de mieux mettre en évidence le rapport qui s'y joue entre le mot, le texte, et l'image.

La première approche est une approche par genre (reconnaître une carte, un dessin, une gravure, une photographie, etc.). Elle est par exemple indiquée pour analyser les cartographies réalisées durant la 1^{ère} Guerre Mondiale et destinées à expliquer, à donner à voir le conflit. Ici, l'apprentissage ne se fonde pas nécessairement sur des métadonnées textuelles : au sein d'un corpus, une analyse des caractéristiques stylistiques (répartition des couleurs, éléments récurrents...) et des rapprochements par similarité peuvent suffire. En quelque sorte, techniquement, l'image se suffit à elle-même. Cela permet de passer outre les barrières linguistiques. Dans une telle approche, un support textuel peut être interprété (et traité) comme une image (calligrammes et calligraphies, mais également texte manuscrit, etc.).

La deuxième approche s'intéresse au type de document (sélectionner un ensemble documentaire – imprimés, images, presse, etc. – pour restreindre une recherche, par exemple pour observer l'usage de la cartographie d'un conflit dans les seuls journaux). Ici, la solution vient surtout d'un apport de métadonnées, souvent préalablement renseignées dans le cadre du travail de signalement, même si la technologie fait se profiler d'autres modes d'identification de ces types de documents.

La troisième approche se tourne vers un nom propre ou « entité nommée » (trouver les représentations d'une personnalité – par exemple Georges Clemenceau – d'une ville, d'une œuvre d'art, d'un monument précis – par exemple la tour Eiffel). Ici, une connexion est étroitement établie entre un terme et sa représentation. L'image ne fonctionne pas seule, elle a besoin de s'appuyer sur un nom, que ce soit à travers des métadonnées associées ou, dans le cas d'une recherche par similarité visuelle, à partir de l'exemple d'une image préalablement nommée. Cette approche pourrait s'avérer particulièrement utile dans le cas du projet Histoire du quartier Richelieu¹³, le repérage d'un ou plusieurs bâtiments remarquables de ce quartier sur des gravures ou photographies anciennes (la façade de la Bourse par exemple) apportant de précieuses informations en complément des métadonnées ou légendes existantes. L'enjeu d'un tel mode de recherche peut également consister à mettre en cause des icônes communément partagées (pensons au portrait d'Arthur Rimbaud par Etienne Carjat en éternel adolescent...) en offrant d'autres vues, moins connues ou moins attendues, d'une entité. S'ajoute la question de l'identité, de l'authenticité, du lien entre l'original et la copie, en particulier dans le cas des sosies, lorsque l'image met le mot à l'épreuve.

La quatrième approche se fonde sur un motif ou un concept (reconnaître dans une image un visage, un être humain, un animal, un objet, un type de véhicule : des soldats, des avions, des chats...). Dans ce cas, le principe est le même que précédemment, à deux nuances près. Tout d'abord, l'image agit surtout comme une représentation générique, donc la question de l'identité se pose de manière moins aiguë (mais cela en soulève une autre, dans certains cas : celle de la subjectivité de la catégorisation [Wuyts, 2021]). En outre, la recherche par motif ou concept surmonte les barrières linguistiques et l'évolution (ou la synonymie) d'une langue, comme dans le

¹³ <https://quartier-richelieu.fr/>

cas de la première approche (par genre). Dans une recherche d'images par motif, un vélo reste un vélo, que le texte environnant mentionne un vélocipède ou une bicyclette...

Enfin, une approche selon un critère propre aux images permettra de sélectionner une couleur, une palette, une taille, un degré de luminosité, un format (paysage, portrait, carré, panoramique), etc., notamment pour observer les variations de couleur dans l'uniforme des soldats français lors de la 1^{ère} Guerre Mondiale. Ici, la recherche sort du champ sémantique de l'interprétation préalable.

Ces approches et les exemples qui les illustrent (essentiellement tirés de l'analyse de GallicaPix par Jean-Philippe Moreux [2019]) sont le reflet de développements en devenir et suggèrent la métamorphose possible des recherches classiques parmi les fonds documentaires des bibliothèques. Avant d'aborder cet aspect, il faut franchir une dernière étape : celle du passage à l'échelle.

La définition d'une stratégie ou l'enjeu de l'industrialisation

L'industrialisation de la fouille d'images, qui représente, à la BnF, l'enjeu majeur du début des années 2020 dans le domaine de l'intelligence artificielle, fait actuellement l'objet d'un projet d'envergure. Les défis à relever sont nombreux : financement, mise en cohérence d'ensembles jusqu'à présent silotés ou décrits selon des modèles de données variés, lacunes en matière de descripteurs indispensables à la recherche d'image, transfert d'apprentissage sur un ensemble de documents plus hétérogène, gestion du bruit occasionné par l'arrivée massive des unités iconographiques dans les résultats d'une recherche, etc.

Permis par des tests, par l'exploration des usages possibles, par des formations préalables, par le développement d'interfaces de recherche, ce projet s'inscrit dans une continuité scientifique. C'est également le cas du point de vue stratégique : le contrat d'objectifs et de performance (COP) conclu entre la BnF et sa tutelle (le ministère de la Culture) prévoyait dans son édition 2017-2021 de « développer des outils conviviaux d'exploration des corpus de Gallica (fouille d'images, reconnaissance d'articles, etc.) »¹⁴ afin, plus généralement, de faciliter l'usage de la cette bibliothèque numérique par le grand public, et de « mettre à disposition des outils de structuration, d'exploration et de reconnaissance des contenus, en particulier la reconnaissance de formes pour les images »¹⁵. Le COP 2022-2026 se fixe quant à lui pour objectif la mise en œuvre du projet de fouille d'images en 2024-2026, en écho à la feuille de route de la BnF sur l'intelligence artificielle déjà citée. Grâce au soutien de France Relance, obtenu en 2022, un premier projet en partenariat avec l'Institut national d'histoire de l'art et la bibliothèque nationale et universitaire de Strasbourg s'attellera à réaliser ce changement d'échelle sur des corpus relatifs à l'architecture et au patrimoine.

3. Ce que cette approche fait à la bibliothèque et les responsabilités qui en découlent

La vision par ordinateur, promesse, pour le bibliothécaire et l'utilisateur, d'être en capacité de jeter un regard nouveau sur les collections et de mettre au jour des « trouvailles » [Bermès, 2020], suppose donc de nombreux défis. Dans l'attente du lancement officiel du projet de fouille d'images industrialisée, les expérimentations se poursuivent à la BnF en s'appuyant sur l'écosystème « bibliothèque » tout entier, sur son art séculaire de la collecte et du signalement, sur sa capacité à faire converger les expertises... Tout au long de ce processus, l'intelligence artificielle interroge cet écosystème lui-même et rend d'autant plus saillants certains de ses constituants.

L'essence de la recherche

Avec la recherche d'image, la notion de « fouille » revêt le sens archéologique de « découverte », de « mise au jour », par exemple d'images de réalités jusqu'alors sans image et vice versa (recherches généalogiques portant sur des personnes dont aucun portrait n'a été conservé,

¹⁴ COP 2017-2021, p. 23, https://www.bnf.fr/sites/default/files/2019-08/BNF_CONTRAT%20PERF_P1_72_BAT.pdf

¹⁵ Ibidem, p. 35

personnes photographiées dont on ignore l'identité¹⁶...). De plus, il a été calculé que 50% des cinq cents requêtes les plus courantes de Gallica portent sur des entités de type personne, lieu ou événement historique, pour lesquelles des ressources iconographiques apporteraient une valeur ajoutée significative.

Enfin, par son universalité, l'image se montre capable de dépasser les frontières linguistiques. Pour l'instant, nous le constatons surtout à travers l'évolution du nombre d'abonnés au compte Instagram¹⁷ de Gallica, fondé sur l'image :

Nombre de fans ou d'abonnés des comptes de Gallica sur les réseaux sociaux (cumul)

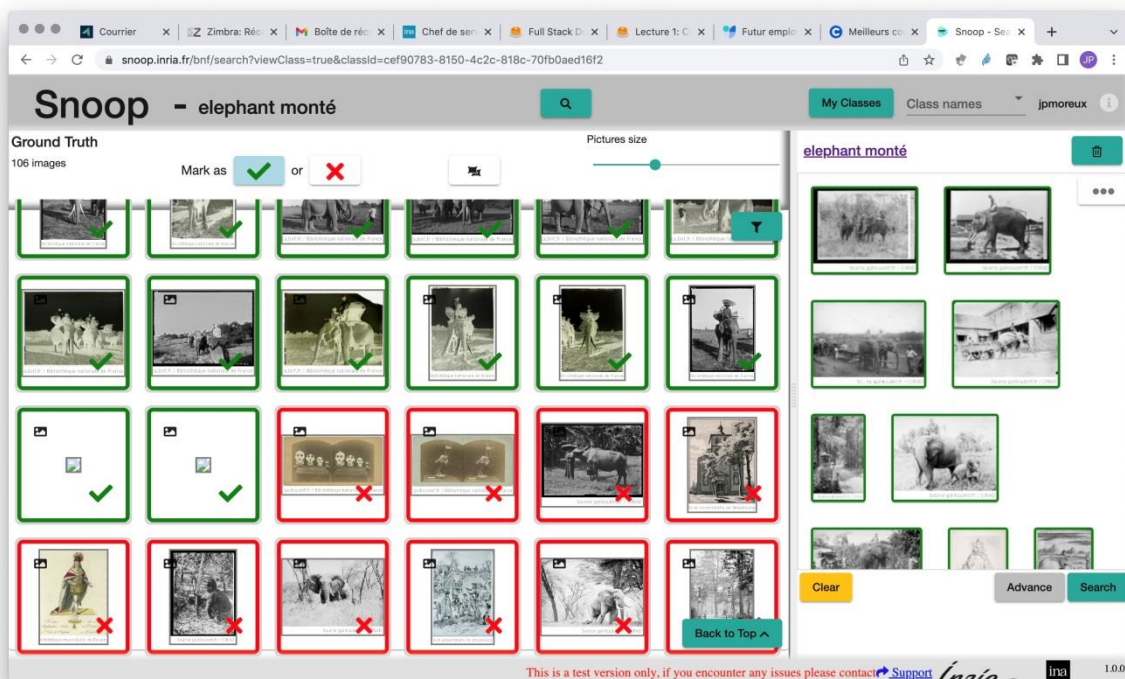
	2019	2020	2021
Facebook (compte ouvert en 2010)	146 540	161 983	170 085
Twitter (compte ouvert en 2010)	68 869	80 154	89 264
Pinterest (compte ouvert en 2012)	22 121	37 873	46 792
Instagram (compte ouvert en 2018)	73 857	138 706	162 651

L'enjeu est donc à la fois de poursuivre le recrutement d'un nouveau public permis par ce média et de faire profiter tous les publics des fonctionnalités de fouille, sans sous-estimer les virages cognitifs et l'information qu'impliquent ces nouveaux types de recherche : là où le texte est un agent efficace de pré-classement des résultats d'une recherche, qu'en sera-t-il de l'image ? Comment le moteur hiérarchisera-t-il la liste ? Quels sont les risques de biais ? Il semble qu'un nouveau type de médiation devienne nécessaire, non plus seulement sur les contenus, mais pour interroger et mettre en perspective les modalités de la recherche par algorithme et préserver l'art de chercher, c'est-à-dire la curiosité, le bon sens, et l'esprit critique. Parmi les réponses déjà apportées par la BnF à ce type d'interrogations, figure le choix d'une démarche qui implique l'utilisateur pour trouver avec lui la meilleure manière de composer sa relation à la machine. C'était justement l'objet du projet GallicaSnoop, co-financé par le ministère de la Culture dans le cadre de son partenariat avec Inria : une interface a été mise en place pour effectuer des recherches d'images dans un corpus donné en affinant progressivement le modèle, en « montrant » à la machine quels résultats sont satisfaisants et quels résultats ne correspondent pas à ce que l'on cherche.

Exemple GallicaSnoop : recherche d'images d'éléphants montés

¹⁶ Durant l'été 2022, la boîte arts graphiques du musée du Quai Branly-Jacques Chirac a présenté une exposition intitulée « Qui êtes-vous ? retrouver les identités perdues des personnes photographiées » suggérant la manière dont la recherche documentaire et le croisement des fonds permettent de retrouver des informations négligées voire effacées à l'époque de la première diffusion des contenus. Voir : <https://www.quaibrantly.fr/fr/collections/vie-des-collections/actualites/boite-arts-graphiques/qui-etes-vous-retrouver-les-identites-perdues-des-personnes-photographiees/>

¹⁷ <https://www.instagram.com/gallicabnf>



Figurent aussi une réflexion sur la notion de qualité (quel sera le niveau d'exigence retenu en ce qui concerne les résultats de l'algorithme et comment ce niveau sera-t-il porté à la connaissance des publics ?), et la complémentarité des modes de recherche possibles. Ainsi, de même que le développement de la numérisation n'a jamais, en soi, empêché un lecteur de consulter un manuscrit dans une salle de lecture, de même le développement de la fouille d'images s'inscrit dans le perfectionnement des outils de recherche en général et il restera possible – et conseillé – de scruter des images de la Bibliothèque à l'œil nu.

La valeur des ressources humaines et la plus-value du professionnel de l'information

Cette dernière idée invite à préciser les contours de la plus-value apportée par les professionnels de la bibliothèque : elle s'exprime en termes de responsabilité, elle-même inscrite dans la stratégie de la BnF¹⁸.

D'abord, l'entraînement d'un algorithme sera d'autant plus léger, moins énergétivore, qu'une analyse profonde des corpus et de leurs spécificités aura été menée auparavant : il peut s'avérer très utile, par exemple, de savoir, d'expérience, que les images contenues dans tel ou tel titre de presse sont en noir et blanc ou d'être capable d'identifier préalablement et de manière exhaustive les classes d'une typologie. Cette approche économe est essentielle pour gérer au mieux le travail humain que nécessitent les entraînements. Dans le cadre du projet GallicaCIP mentionné plus haut, financé par le ministère de la Culture et Inria, et consistant en une tentative de classification d'enluminures de la base Mandragore¹⁹, une étape a consisté à préparer un corpus en y entourant manuellement les images, et 8000 « boîtes » ont ainsi été produites. Plus globalement, le ressort technique de chacune des cinq approches détaillées précédemment peut commencer, au préalable ou dans le cas des corpus très restreints, par un traitement manuel des contenus. Une telle sobriété est également essentielle pour garantir un déploiement raisonné de la puissance machine, en évitant d'entraîner plusieurs fois de suite un modèle, en soupesant habilement l'intérêt de la surprise apportée par les résultats d'un apprentissage peu supervisé, ou encore en utilisant les performances des approches texte-image sans supervision (CLIP) [Moreux, 2022].

À l'autre bout de l'exercice, le professionnel pourra s'appuyer sur sa connaissance des collections, sur son expertise des données, pour évaluer les résultats obtenus et identifier les risques de biais.

¹⁸ Voir COP 2022-2026, en particulier l'axe 4.

¹⁹ <http://mandragore.bnf.fr>

Il aura à cœur d'injecter les enrichissements massifs produits par l'intelligence artificielle dans ses catalogues et outils de recherche, alimentant un cercle vertueux. Mais cette fois encore, circonspection et sobriété sont de mise, surtout si nous considérons le volume des données générées par le travail sur les images : « Comme le repérage d'entités visuelles demande énormément d'informations locales, les bases de données d'indexation constituées à partir de la reconnaissance des entités sont conséquentes (elles peuvent avoir une taille supérieure à celle des vidéos elles-mêmes) et font l'objet de recherches au sein de l'établissement pour en optimiser les temps de calcul. », explique Nicolas Hervé, chercheur en informatique à l'Ina (Institut national de l'audiovisuel), à propos du projet Diginpix, un système d'identification d'entités visuelles dans les images fixes et animées [Moiraghi et Moreux, 2018]. Le traitement d'un corpus représentant une infime fraction de la collection numérique peut générer plusieurs millions de métadonnées [Moreux, 2019], ce qui représente aussi un enjeu de conservation. Enfin, le professionnel veillera à la légalité du travail sur les collections qu'il met à la disposition des publics en cadrant l'usage des contenus protégés par le droit et en s'attachant au respect du règlement général sur la protection des données (RGPD²⁰).

De la nécessité et de l'exigence de ne jamais travailler seul

Pour continuer à apporter cette plus-value, le professionnel de l'information a de plus en plus besoin de s'entourer, de partager, de s'ouvrir, dans la dynamique de l'association, au sein du site Richelieu, des trois institutions que sont la BnF, l'École des Chartes et l'INHA (Institut national d'histoire de l'art), et dans la dynamique du DataLab²¹, inauguré à la BnF en 2021 pour favoriser les rencontres entre bibliothécaires, chercheurs et informaticiens, et améliorer ainsi, *in fine*, la qualité des outils de recherche proposés et de leurs résultats.

L'esprit d'ouverture n'est pas nouveau à la BnF et le prototype GallicaPix démontre que quiconque maîtrise suffisamment les API²² de récupération des contenus de Gallica, les données, et les logiciels d'intelligence artificielle peut créer une interface de recherche [Moreux, 2019]. Mais se dessinent de nouvelles formes de rencontre : il s'agit d'être sensible aux nouvelles communautés qui pourraient émerger grâce au développement de la fouille d'images à une échelle industrielle, en lien avec le DataLab, mais également avec le travail mené avec les industries culturelles et créatives. Il s'agit par ailleurs d'élargir le champ spatio-temporel de la collaboration, en défrichant des domaines peu développés. Ainsi, au moment d'industrialiser la fouille d'images dans Gallica, la BnF prévoit de se tourner vers l'ADEME (Agence de la transition écologique) pour étudier son impact environnemental et se doter de moyens d'action dans ce domaine.

Conclusion

L'histoire ici racontée est faite de choix stratégiques, de partenariats, de prototypes, de marchés, de sensibilisation, de technologies, de cadres de travail... Les questions éthiques la scandent, du début à la fin, aux côtés des enthousiasmes sans réserve qui font écho au caractère unique de la BnF. Ce caractère, qui la distingue d'autres immenses réservoirs d'images, s'exprime à travers l'hétérogénéité des supports susceptibles de contenir des images (lesquels se soutiennent mutuellement) et la dimension patrimoniale des contenus (qui exige un entraînement particulier des algorithmes), mais aussi à travers les attentes des publics et en particulier des chercheurs (qui sont aussi diverses que les interprétations et usages possibles d'une image), la logique de co-construction qui a présidé, depuis la fin des années 2000, à l'enrichissement de Gallica, ou encore l'attachement au respect du droit.

Comme dans un kaléidoscope, son enjeu est d'articuler plusieurs interfaces, réelles ou virtuelles : entre les différents professionnels, entre les machines existantes et les nouveautés technologiques, entre les différents projets, entre le passé et le futur.

²⁰ <https://www.cnil.fr/fr/reglement-europeen-protection-donnees>

²¹ <https://www.bnf.fr/fr/bnf-datalab>

²² API : interface de programmation d'application. Les API sont des outils informatiques qui permettent à des machines de dialoguer entre elles et d'échanger des données.

Bibliographie

- Angjeli, Anila et Isaac, Antoine, 2008, « Semantic web and vocabularies interoperability: an experiment with illuminations collections », World Library and Information Congress : 74th IFLA general conference and council, Montréal, Québec, Canada, https://www.researchgate.net/profile/Anila-Angjeli/publication/274375033_Semantic_web_and_vocabularies_interoperability_an_experiment_with_illuminations_collections/links/551c75250cf2fe6cbf791837/Semantic-web-and-vocabularies-interoperability-an-experiment-with-illuminations-collections.pdf
- Bermès, Emmanuelle, 2018-2019, « Quand le dépôt légal devient numérique : épistémologie d'un nouvel objet patrimonial », *Quaderni : communication, technologies, pouvoir*, n° 98, p. 73-86
- Bermès, Emmanuelle et Moiraghi, Eleonora, 2020, « Le patrimoine numérique national à l'heure de l'intelligence artificielle. Le programme de recherche Corpus comme espace d'expérimentation pour les humanités numériques », *Revue d'Intelligence Artificielle (RIA)*, p. 89-109, <https://roia.centre-mersenne.org/articles/10.5802/roia.5>
- Bermès Emmanuelle, 2020, *Le numérique en bibliothèque : naissance d'un patrimoine : l'exemple de la Bibliothèque nationale de France (1997-2019)*. Thèse de doctorat, <https://tel.archives-ouvertes.fr/tel-02475991>
- Bibliothèque nationale de France, « la BnF et l'intelligence artificielle », site web institutionnel, <https://www.bnf.fr/fr/feuille-de-route-ia>
- CIP (projet de Classification d'images patrimoniales), « Mandragore : jeu d'images annotées pour la détection d'objet », 2019, <https://api.bnf.fr/fr/node/212>
- Leroy-Terquem, Mélanie, et Leclaire, Céline, 2022, Dossier « L'intelligence artificielle à la BnF », *Chroniques*, n°93, http://chroniques.bnf.fr/pdf/Chroniques_93.pdf
- Moreux, Jean-Philippe, 2018, « Gallica.pix, une preuve de concept », Journée d'étude « Explorer des corpus d'images. L'IA au service du patrimoine ». Projet CORPUS, 28 mars 2018, <https://bnf.hypotheses.org/2809>
- Moreux, Jean-Philippe, 2019, « Recherche d'images dans les bibliothèques numériques patrimoniales et expérimentation de techniques d'apprentissage profond », *Documentation et Bibliothèques*, n°2, vol 65, avril-juin 2019, Montréal, <https://www.erudit.org/en/journals/documentation/2019-v65-n2-documentation04848/1063786ar>
- Moreux, Jean-Philippe, 2022, « Indexation d'images patrimoniales et intelligence artificielle : retour d'expérience de la BnF. 2012-2022, d'AlexNet au BnF Datalab », Commission Patrimoine Occitanie, https://docs.google.com/presentation/d/158Yw69D_kSmE5t5KEWvxaHal7u1asfOnsiKlr3EOzww/edit#slide=id.g123e1690ef5_0_83

Moreux, Jean-Philippe, 2022, « AI and heritage image indexing: feedbacks from the BnF. 2012-2022, from AlexNet to the BnF Datalab », IFLA World Library and Information Congress, Satellite Conference on Artificial Intelligence, Galway, Irlande : https://docs.google.com/presentation/d/1RbUvVw8mr3DVKfKncGBMeq9s0mbDT_LEmNfpgZ8Zuws/edit#slide=id.gc6f73a04f_0_0

Moiraghi, Eleonora et Moreux, Jean-Philippe, 2018, « Explorer des corpus d'images. L'IA au service du patrimoine », *Carnet de la recherche à la Bibliothèque nationale de France*, atelier dans le cadre du projet CORPUS, avril 2018, <https://bnf.hypotheses.org/2809>

Picard, David et Gosselin, Philippe-Henri, 2015, « Évaluation de descripteurs visuels pour l'annotation automatique d'images patrimoniales », Grets, Lyon, France, <https://hal.archives-ouvertes.fr/hal-01277486/document>

Sidère, Nicolas, 2018, *Etat de l'art sur des outils logiciels, actuels ou en développement et des labos et équipes de recherché* [Rapport Technique], BnF, <https://hal-bnf.archives-ouvertes.fr/hal-02432577/document>

Wuyts, Jolan, 2021, « Annotating datasets for computer vision to recognise architectural and artistic styles: lessons from the V4Design project », conference Futurs fantastiques, BnF, 10 décembre 2021, <https://www.bnf.fr/fr/mediatheque/presentation-projects-annotating-datasets-computer-vision-recognise-architectural-and>