Documentation et bibliothèques

DOCUMENTATION BIBLIOTHÈQUES

Recherche d'images dans les bibliothèques numériques patrimoniales et expérimentation de techniques d'apprentissage profond

Searching for Images in Heritage Digital Libraries and Experimenting With Deep Learning Technology

Jean-Philippe Moreux

Volume 65, numéro 2, avril–juin 2019

Techno, techno, techno...

URI : https://id.erudit.org/iderudit/1063786ar
DOI : https://doi.org/10.7202/1063786ar

Aller au sommaire du numéro

Éditeur(s)

Association pour l'avancement des sciences et des techniques de la documentation (ASTED)

ISSN

0315-2340 (imprimé)
2291-8949 (numérique)

Découvrir la revue

Citer cet article

Moreux, J.-P. (2019). Recherche d'images dans les bibliothèques numériques patrimoniales et expérimentation de techniques d'apprentissage profond. *Documentation et bibliothèques*, *65*(2), 5–27. https://doi.org/10.7202/1063786ar

Résumé de l'article

Si historiquement, les bibliothèques numériques patrimoniales furent d'abord alimentées par des images, elles profitèrent rapidement de la technologie OCR pour indexer les collections imprimées afin d'améliorer le service de recherche d'information offert aux utilisateurs. Mais l'accès aux ressources iconographiques n'a pas connu les mêmes progrès et ces dernières demeurent dans l'ombre : indexation manuelle lacunaire, hétérogène et impossible à généraliser ; silos par genre documentaire ; recherche dans le contenu des images encore peu opérationnelle sur les collections patrimoniales. Aujourd'hui, il serait pourtant possible de mieux valoriser ces ressources en exploitant les énormes volumes d'OCR produits durant les deux dernières décennies (tant comme descripteur textuel que pour l'identification automatique des illustrations des imprimés), en profitant de la maturité des techniques d'intelligence artificielle (en particulier l'apprentissage profond ou *deep learning*), pour mettre ainsi en valeur ces gravures, dessins, photographies, cartes, etc., pour leur valeur propre, mais aussi comme point d'entrée dans les collections, en favorisant découverte et rebond.

Cet article décrit une approche ETL (extract-transform-load) appliquée aux images d'une bibliothèque numérique à vocation encyclopédique : identifier et extraire l'iconographie partout où elle se trouve (dans les collections d'images, mais aussi dans les imprimés) ; transformer, harmoniser et enrichir ses métadonnées descriptives grâce à l'IA ; intégrer ces données dans une application web dédiée à la recherche iconographique. Cette approche est qualifiée de pragmatique à double titre, puisqu'il s'agit de valoriser des ressources numériques existantes tout en mettant à profit les acquis de l'IA.

RECHERCHE D'IMAGES DANS LES BIBLIOTHÈQUES NUMÉRIQUES PATRIMONIALES ET EXPÉRIMENTATION DE TECHNIQUES D'APPRENTISSAGE PROFOND

Jean-Philippe Moreux
Expert scientifique Gallica, département de la Coopération, Bibliothèque nationale de France, Paris
jean-philippe.moreux@bnf.fr

RÉSUMÉ | ABSTRACT

Si historiquement, les bibliothèques numériques patrimoniales furent d'abord alimentées par des images, elles profitèrent rapidement de la technologie OCR pour indexer les collections imprimées afin d'améliorer le service de recherche d'information offert aux utilisateurs. Mais l'accès aux ressources iconographiques n'a pas connu les mêmes progrès et ces dernières demeurent dans l'ombre : indexation manuelle lacunaire, hétérogène et impossible à généraliser ; silos par genre documentaire ; recherche dans le contenu des images encore peu opérationnelle sur les collections patrimoniales. Aujourd'hui, il serait pourtant possible de mieux valoriser ces ressources en exploitant les énormes volumes d'OCR produits durant les deux dernières décennies (tant comme descripteur textuel que pour l'identification automatique des illustrations des imprimés), en profitant de la maturité des techniques d'intelligence artificielle (en particulier l'apprentissage profond ou *deep learning*), pour mettre ainsi en valeur ces gravures, dessins, photographies, cartes, etc., pour leur valeur propre, mais aussi comme point d'entrée dans les collections, en favorisant découverte et rebond.

Cet article décrit une approche ETL (extract-transform-load) appliquée aux images d'une bibliothèque numérique à vocation encyclopédique : identifier et extraire l'iconographie partout où elle se trouve (dans les collections d'images, mais aussi dans les imprimés) ; transformer, harmoniser et enrichir ses métadonnées descriptives grâce à l'IA ; intégrer ces données dans une application web dédiée à la recherche iconographique. Cette approche est qualifiée de pragmatique à double titre, puisqu'il s'agit de valoriser des ressources numériques existantes tout en mettant à profit les acquis de l'IA.

Searching for Images in Heritage Digital Libraries and Experimenting With Deep Learning Technology

If historically, heritage digital libraries were initially made up of images, they rapidly benefited from the optical character recognition (OCR) technology to index print collections and improve reference services for users. However, access to iconographic resources has not experienced the same progression, remaining somewhat difficult to access. Manual indexation is not very efficient, it is varied and impossible to apply uniformly. Searching the content of an image is not as effective with heritage collections. Today, it is possible to improve the use of these resources by exploiting large volumes of OCR produced over the past two decades (both the textual descriptors as well as the automatic identification of the illustrations in the printed documents) and to take advantage of proven artificial intelligence techniques, especially deep learning. In doing so, it will showcase engravings, drawings, photographs, maps, etc. as such but also the point of entry to the collections by improving discovery and connections.

This article describes an ETL (extract-transform-load) approach as it applies to the images in a digital library with an encyclopedic vocation. There are three components: 1) identify and extract the iconography wherever it is found, either in images or in the printed documents, 2) transform, harmonise and enrich the descriptive metadata with the help of artificial intelligence, and 3) incorporate this data into a web application dedicated to iconographic research. This is a two-pronged approach because it highlights existing digital resources and takes advantage of the benefits of artificial intelligence.

Introduction

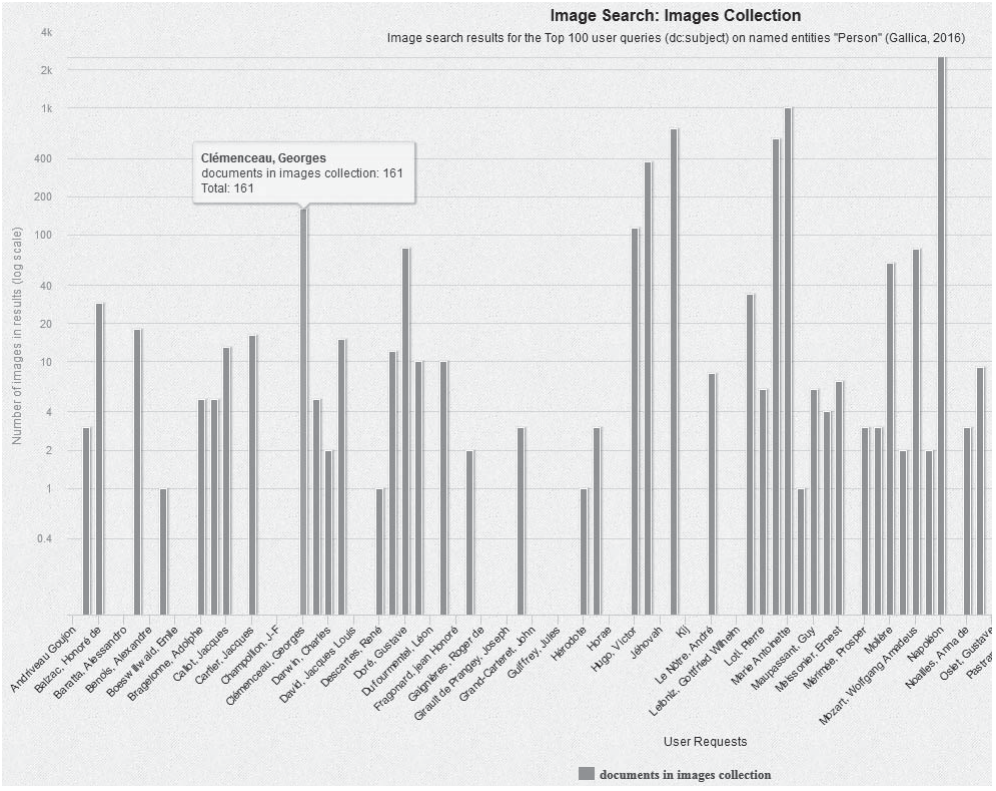
Alors même que la création des collections numériques patrimoniales a débuté par l'acquisition en mode image des fonds conservés, rechercher dans le contenu de ces images[1] plusieurs décennies après relève encore d'un futur plus ou moins éloigné (Gordea et Haskiya, 2017). Ce paradoxe apparent trouve son origine dans deux faits : les bibliothèques patrimoniales ont fait porter leur effort sur l'océrisation des imprimés, investissement qui en retour a rendu des services majeurs en termes de recherche d'information ; ensuite, interroger ou parcourir de grandes collections d'images demeure un défi, en dépit des efforts de la communauté scientifique comme des GAFA à les relever (Datta, Joshi, Li et Wang, 2008).

Les besoins sont pourtant bien réels, si l'on en croit tant les enquêtes conduites auprès des usagers de bibliothèques numériques patrimoniales (63 % des utilisateurs de Gallica, la bibliothèque numérique de la BnF, consultent des images[2], et 85 % connaissent l'existence d'une collection d'images (BnF, 2017)) que les études statistiques des comportements des usagers : parmi les cinq cents requêtes les plus courantes, la moitié portent sur des entités nommées de type Personne, Lieu ou Événement historique (Chiron, Doucet, Coustaty, Visani et Moreux, 2017). Or il est plausible que pour de telles requêtes, des ressources iconographiques puissent apporter une information complémentaire à celle présente dans les contenus textuels. Par ailleurs, les usagers des bibliothèques numériques sont familiers de ces fonctionnalités de recherche dans les images, mises en œuvre depuis de nombreuses années par plusieurs services grand public, comme Google Images depuis 2001, iPhoto depuis 2002 ou encore Flickr depuis 2006 (dans sa forme définitive).

Cherchant à satisfaire de telles recherches iconographiques encyclopédiques, les bibliothèques numériques ne sont pas sans ressource. La figure 1 ci-dessous (extrait) montre ainsi le nombre de documents de la collection iconographique de Gallica (photographies, gravures, dessins, affiches, cartes, etc.) pour le sous-échantillon des cent requêtes les plus fréquentes portant sur des personnes.

FIGURE 1

Nombre de documents de la collection Images de Gallica pour les 100 requêtes les plus fréquentes sur une entité nommée de type Personne (en interrogeant la métadonnée Sujet, dc:subject).



1. Pour paraphraser l'acronyme anglais consacré, CBIR, *content-based image retrieval*.

2. Par la suite, « image » désignera plutôt l'objet numérique résultant de la numérisation d'un document patrimonial (une affiche, une médaille, un fascicule de presse, etc.) et « illustration » une ressource iconographique numérique (le recto d'une affiche, l'avers et le revers d'une médaille, les illustrations contenues dans le fascicule, etc.). Le lecteur saura élucider de lui-même les cas ambigus.

Ce graphe met également en lumière des lacunes, liées à la taille de la collection Images (un million de documents environ), au regard du large spectre des domaines de connaissance et des périodes sondés par les usagers (de l'Antiquité au XXIe siècle). Ainsi, Champollion n'est pas représenté, et seules 161 illustrations concernent Clémenceau. Pour autant, les bibliothèques sont riches de nombreuses autres sources iconographiques, ainsi la presse, avec jusqu'à trois illustrations par page en moyenne pour les titres les plus illustrés de la première moitié du XXe siècle (Moreux, 2016). Mais ces sources sont parfois organisées en silos de données peu interopérables, manquant le plus souvent des descripteurs indispensables à la recherche d'image, qu'ils soient « physiques » (localisation de l'illustration dans la page, taille, couleur et forme) ou sémantiques (indexation thématique des documents dans les catalogues, campagne d'annotation participative[3], etc.). Et lorsque tous les genres documentaires sont réunis au sein d'un portail unique, ce qui est le cas de Gallica, les modes de recherche et de consultation auront le plus souvent été pensés selon un paradigme classique (métadonnées bibliographiques, indexation en texte intégral, feuilleteur de pages). Alors que l'interrogation de fonds iconographiques patrimoniaux pose des enjeux particuliers (Picard, Gosselin et Gaspard, 2015), répond à des usages variés, du butinage récréatif de photographies anciennes à l'étude scientifique des matériaux enluminés ou imprimés de manuscrits ou incunables (Coustaty, Pareti, Vincent et Ogier, 2011), cible des domaines de connaissance plus ou moins larges (encyclopédiques ou spécialisés) et appelle des interactions homme/machine spécifiques, de la saisie d'un mot-clé au tracé d'une ébauche de la forme de l'image recherchée (Wan et Liu, 2008 ; Breiteneder et Eidenberger, 2000 ; Datta, Joshi, Li et Wang, 2008). Enfin, la dynamique actuelle des humanités numériques incite les bibliothèques à investir ce champ, conjointement aux équipes de recherche travaillant sur des corpus illustrés (au sujet des *visual studies*, voir par exemple le fichier de données SIAMESET de Melvin Wevers et Juliette Lonij[4], les réalisations du Software Studies Initiative[5], ou la présentation Powerpoint de Paul Fyfe, Qian Ge et Joe Aguayo[6]). Dans ce contexte, les travaux des laboratoires numériques et autres pépinières d'innovation établis ces dernières années par les bibliothèques nationales témoignent de cet effort (voir à ce sujet KB Lab[7], British Library Labs[8], ou la demi-journée d'étude du 28 mars 2018 sur les différentes applications des méthodes d'analyse automatique d'images par le contenu, décrite par Eleonora Moiraghi (2018)).

Confrontées à ces défis, les bibliothèques numériques patrimoniales n'ont pas baissé les bras. On pourra citer les actions de médiation d'Europeana[9] et de Gallica[10] et l'indexation manuelle de la British Library[11], mais leur nature même les limite à des corpus restreints et généralement thématiques et/ou monogenres (photographie, enluminure, etc.). Plus récemment, la force brute de l'ordinateur aura été mise à profit pour extraire les illustrations de corpus numériques et les donner à voir, tout en sollicitant parfois une annotation participative des usagers. Le Mechanical Curator[12] de la British Library est représentatif de ce type d'expérimentation. Des réalisations tablant sur des techniques d'intelligence artificielle (*machine learning*, *deep learning*) ont également vu le jour : ainsi la Bayerische Staatsbibliothek[13] offre une fonctionnalité de recherche par similarité visuelle sur la totalité des illustrations de sa collection (voir aussi une expérimentation similaire de la BnF[14]). Le présent travail s'inscrit dans cette dernière catégorie, laquelle vise à faire profiter les collections patrimoniales numérisées de grande taille des avancées de l'intelligence artificielle.

Cet article se propose donc d'investiguer une double problématique, celle de la création automatisée d'une base d'images encyclopédique (dans une acception relative, c'est-à-dire couvrant toutes les collections d'une bibliothèque elle-même à vocation encyclopédique) et celle relative aux modalités d'indexation et de recherche dans les métadonnées ainsi créées. Une première section décrit la phase initiale de la nécessaire agrégation des données et métadonnées à disposition (le « e » d'une approche ETL, *extract-transform-load*). La deuxième section présente les transformations et enrichissements appliqués aux données collectées, en particulier l'application de traitements relevant des méthodes dites d'apprentissage profond (*deep learning*). L'interrogation multimodale des illustrations, sous la forme de l'application web GallicaPix, est ensuite décrite, et les cas d'usage rendus possible par cette preuve de concept (PoC) sont expérimentés et commentés. Enfin, on prendra pour prétexte cette expérimentation pour aborder la

3. Par ailleurs, l'annotation manuelle n'est pas un remède souverain (voir par exemple Nottamkandath, Oosterman, Ceolin et Fokkink, 2014 ; Welinder, Branson, Belongie et Perona, 2010).

4. Voir lab.kb.nl/dataset/siameset

5. Tools for the Analysis and Visualization of Large Image and Video Collections for the Humanities », Software Studies Initiative, University of California, http://lab.softwarestudies.com/2012/04/software-studies-initiative-awarded.html

6. Voir ncsu-las.org/wp-content/uploads/2016/05/wrm-presentation-slides-4-27-16-paul-fyfe.pdf

7. http://lab.kb.nl

8. http://labs.bl.uk

9. blog.europeana.eu/2017/04/galleries-a-new-way-to-explore-europeana-collections

10. gallica.bnf.fr/html/und/images/images

11. imagesonline.bl.uk

12. mechanicalcurator.tumblr.com et www.flickr.com/photos/britishlibrary

13. bildsuche.digitale-sammlungen.de

14. gallicastudio.bnf.fr/bo%C3%AEte-%C3%A0-outils/avec-gallica-similitudes-%C3%A0-vous-la-recherche-par-images-similaires

thématique des politiques de collaboration scientifique entre institutions patrimoniales et acteurs des humanités numériques, en particulier au prisme de la diffusion de l'IA dans les boîtes à outils des bibliothèques.

Extraire et agréger

Plusieurs décennies après la création des premières bibliothèques numériques (Gallica a fêté ses vingt ans en 2017), les ressources iconographiques conservées dans les magasins numériques sont à la fois conséquentes, en constante expansion et de nature variée (notamment en matière de techniques de production et de reproduction, voir figure 2). Un traitement massif et inclusif de ces ressources semble donc nécessaire. Et l'on peut penser qu'il devra débuter par une phase d'extraction et d'agrégation prenant en considération la variabilité des données à disposition, du fait tant de la nature des silos documentaires que de l'histoire des politiques de numérisation ayant présidé à leur constitution.

La base iconographique décrite dans cet article agrège environ 260 000 illustrations (collectées parmi un corpus de 475 000 pages) des collections d'images et d'imprimés de Gallica relatives à la Première Guerre mondiale et aux années la précédant (sur la période 1910-1920). Produite à l'aide des API Gallica[15] et des protocoles SRU (*Search Retrieval via URL*) et OAI-PMH (*Open Archives Initiative – Protocol for Metadata Harvesting*), elle suit un formalisme XML et a été stockée dans une base XML, BaseX[16]. Son modèle de données agrège les niveaux document, page et illustration et permet d'accueillir les informations disponibles dans les différents silos documentaires ciblés. L'accès aux illustrations elles-mêmes est réalisé à l'aide du protocole IIIF Image[17], appréciable standard qui autorise la manipulation homogène d'images indépendamment de leurs localisations physiques et des institutions qui les hébergent. Le retour d'expérience détaillé de cette première étape d'agrégation est présenté dans les sections suivantes.

Collections Images

Un *set* thématique préexistant de l'entrepôt numérique OAI-PMH de Gallica (« gallica :corpus :1418 ») est utilisé afin d'en extraire les métadonnées des documents images (œuvres graphiques, coupures de presse, médailles, cartes, partitions musicales, etc.). Ces documents présentent des défis particuliers : métadonnées souffrant de défauts d'incomplétude et d'inconsistance (du fait de la variabilité des pratiques d'indexation), et peu ou pas de métadonnées documentaires (genre – photo, gravure, dessin, etc. –, couleur, taille, etc.).

Les recueils reliés d'illustrations[18] (voir figure 3) posent des enjeux techniques de plus grande ampleur encore : les couvertures, pages de texte et pages vierges doivent être exclues de la base d'illustrations ; les pages collationnant plusieurs

FIGURE 2

Exemples de types documentaires porteurs d'iconographie : presse quotidienne, portfolio, affiche.

15. api.bnf.fr

16. basex.org

17. iiif.io/api/image/2.1

18. Voir gallica.bnf.fr/ark :/12148/btv1b8432784m

illustrations doivent être segmentées (voir figure 2, exemple central, et figure 3) ; enfin il convient d'associer une légende, quand elle existe, à chaque illustration.

Ce corpus a été complété par diverses requêtes SRU portant sur des métadonnées catalogue (« sujet = Guerre 14-18 », « source = agence photo Meurisse », « type = affiche », etc.).

Collections Imprimés

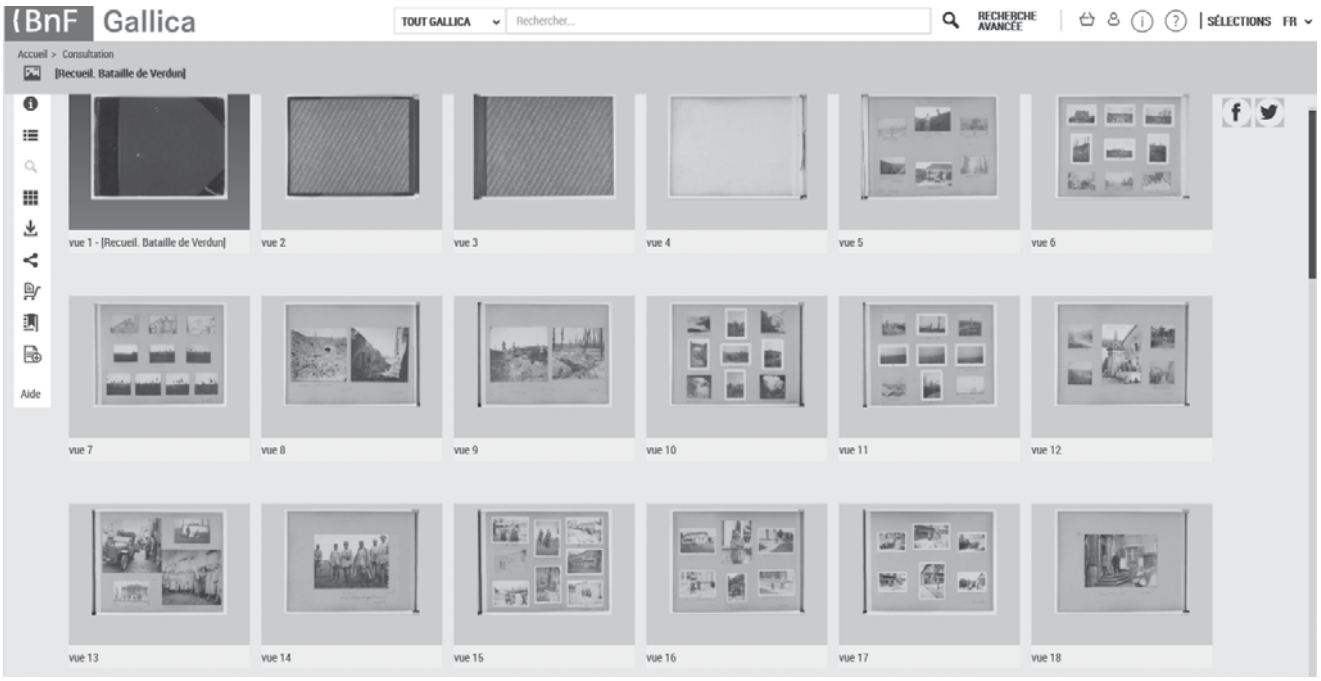
La base est alimentée par une sélection intellectuelle d'ouvrages et de revues ainsi que par un échantillonnage temporel de la collection de presse (une vingtaine de titres). Pour cette collection, l'OCR est une ressource doublement valorisée, puisqu'elle donne à voir l'emprise des illustrations au sein de la page ainsi que le texte environnant l'illustration, lequel est extrait et conservé pour sa qualité de descripteur de l'illustration. Comme on l'a précédemment évoqué, l'OCR étant appliqué à grande échelle par les bibliothèques, patrimoniales ou non, depuis les années 2000, il en résulte un gisement considérable de signalement de ressources iconographiques.

La collection des périodiques de la BnF se présente sous différents formalismes liés à l'histoire des projets de numérisation successifs. Dans tous les cas, il s'agit d'extraire des formats METS[19] et ALTO[20] les métadonnées descriptives des illustrations. Les projets de numérisation récents identifiant l'emprise de chaque article (OLR, *optical layout recognition*) facilitent cette tâche, du fait de leur structuration fine et contrôlée, alors que les anciens programmes offrent de l'OCR brut peu structuré. Des revues thématiques enrichissent la base : journaux de tranchées, revues scientifique et technique, revues de sciences militaires, journaux professionnels, etc. Dans le cas de la presse quotidienne, les illustrations récoltées se caractérisent par des singularités (taille variable des illustrations, de la vignette à la double page ; mauvaise qualité de reproduction, en particulier aux débuts de la photogravure), une grande diversité de genres (de la carte d'état-major à la bande dessinée) et un volume conséquent. Le bruit est également massif (blocs de texte que l'OCR reconnaît faussement comme illustration, filets et autres ornements, publicités illustrées et répétées au fil des publications, etc.). Il varie de 10 % (pour l'OLR) jusqu'à atteindre les proportions considérables de 80 % pour l'OCR brut (voir figure 4).

Diverses heuristiques sont appliquées afin de réduire ce bruit, en filtrant sur des critères physiques tels que taille, ratio largeur/hauteur (suppression des filets et autres culs-de-lampe), emplacement des illustrations (par exemple l'ours de la première page). Ces automatismes génèrent en retour des faux positifs qui seront traités ultérieurement (voir la section « Classification des genres documentaires »).

FIGURE 3

Exemple de recueil d'illustrations, avec plusieurs centaines d'illustrations placées sous une seule notice bibliographique chapeau.

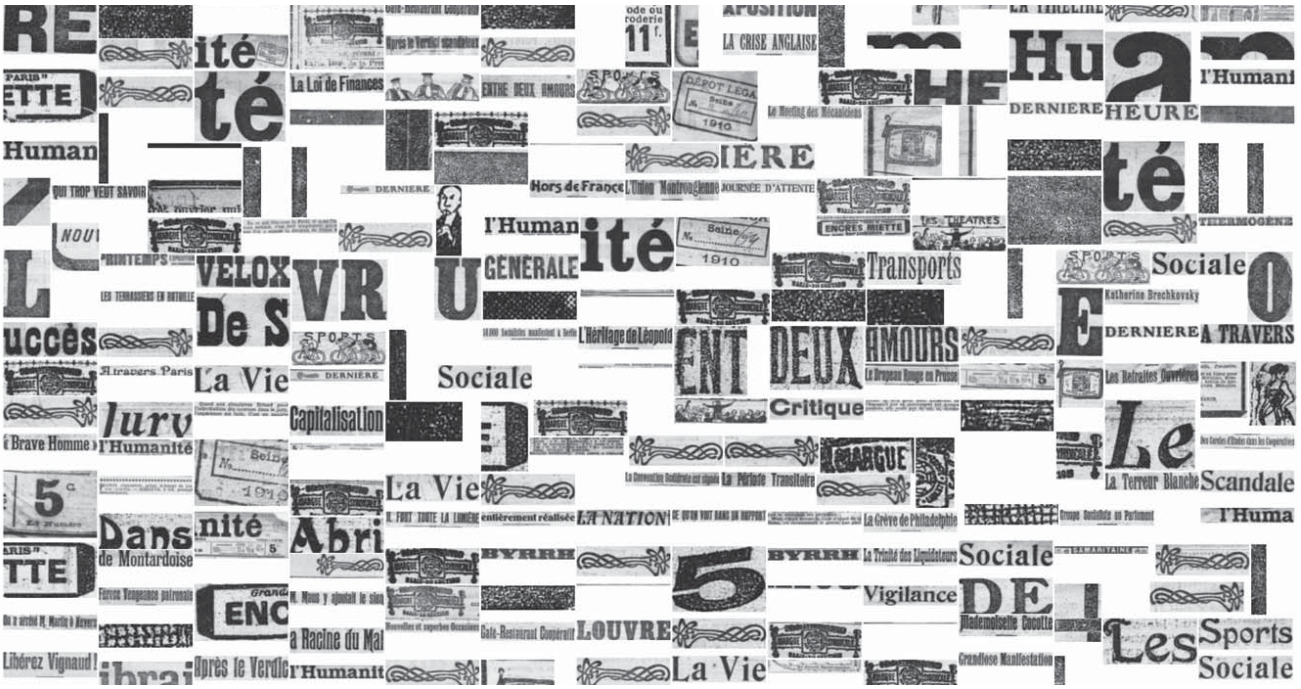


19. www.loc.gov/standards/mets

20. www.loc.gov/standards/alto

FIGURE 4

Exemples de blocs que l'OCR traite faussement comme des illustrations.



Dans le cas de la presse quotidienne, un deuxième enjeu concerne l'identification des illustrations de publicité et petites annonces, qui présentent un intérêt scientifique propre (voir le fichier de données SIAMESET), mais doivent être isolées du corpus des illustrations éditoriales. À ce stade du traitement, cela ne peut être fait que si la métadonnée est disponible (c'est le cas de l'OLR). Le même traitement est appliqué à l'OCR des monographies du corpus (ouvrages d'histoire et de sciences militaires, historiques de régiment, etc.).

Conclusion intermédiaire (1)

Cette seule étape justifie l'effort consenti dans la mesure où elle donne accès pour les utilisateurs à des illustrations jusqu'alors profondément cachées dans les collections numériques (en particulier celle des imprimés). De plus, elle construit les fondations requises pour l'élaboration des interfaces et des interactions homme/machine adaptées à la recherche iconographique (zoom sur l'illustration, mosaïque de résultats, etc.).

Elle met aussi en lumière les nombreux défis à relever pour la mener à bien :

- variabilité des formats, métadonnées et pratiques de numérisation des différentes collections ;
- identification des illustrations (on parle aussi de segmentation) sur des genres documentaires hétérogènes ;
- réduction du bruit résultant du processus d'identification, en particulier pour le cas difficile de la presse quotidienne (dans le cas de notre corpus, la moitié des illustrations extraites sont du bruit, et 99 % de ce bruit provient de la presse quotidienne) ;
- manque de métadonnées essentielles au grain de l'illustration (genre, couleur dominante, etc.).

Du point de vue de l'informaticien, le pipeline de traitement (voir figure 22) est linéaire, mais il présente des variations du fait des multiples sources de données analysées. Ses traitements sont simples (hormis pour la tâche de segmentation, pour laquelle des barrières scientifiques demeurent), mais intensifs (analyse massive de données catalogue et de contenus XML). Et il requiert un monitoring humain, en particulier pour la sélection documentaire, qui sera souvent l'occasion pour le curateur des fonds de faire part de connaissances a priori qui profiteront grandement au paramétrage du pipeline (« Ce titre de presse ne contient que des illustrations publicitaires », « Ce corpus est constitué de photographies noir et blanc », etc.).

Enfin, le bibliothéconomiste voit quant à lui poindre des questionnements singuliers liés aux caractéristiques des métadonnées descriptives générées :

- volume : le corpus considéré représente une infime fraction de la collection numérique, et pourtant son traitement conduit à la production de plus de 12 millions de métadonnées ;

- dispositif de stockage : où stocker ces métadonnées ? Comment articuler métadonnées « primaires » (à l'échelle des documents numériques) et « secondaires » (celles décrivant leurs contenus) ? Comment les rendre disponibles aux autres systèmes d'information de la bibliothèque qui pourraient y trouver de l'intérêt ?
- cycle de vie : il faudrait pouvoir alimenter le pipeline à mesure que la collection numérique croît. Par ailleurs, rejouer le pipeline suite à une amélioration technique d'une de ses composantes est également un cas d'usage à envisager.

Transformer et enrichir

Cette étape consiste à transformer, enrichir et aligner les métadonnées obtenues lors de la phase d'extraction et d'agrégation. En effet, les métadonnées descriptives des illustrations récoltées se caractérisent tant par leur hétérogénéité que par leur pauvreté au regard des fonctionnalités utilisateur attendues.

Descripteur textuel

Les illustrations des imprimées sans descripteur textuel (par exemple du fait d'une lacune de l'OCR originel) sont détectées et leur emprise élargie est traitée par le moteur OCR de l'API Google Cloud Vision, ce qui permet d'indexer textuellement ces illustrations muettes. Ce traitement peut également être appliqué à certains portfolios dotés de légendes n'ayant pas été extraites durant leur programme de numérisation.

Indexation thématique

Cas de la collection Images

Un alignement vers le standard d'indexation thématique de contenus de presse IPTC[21] (17 thèmes de premier niveau) est réalisé à l'aide d'une approche par réseau sémantique : les mots-clés des notices des documents de la collection Images (titre, sujet, description) et des légendes des illustrations (quand il y en a) sont lemmatisés puis alignés sur les thèmes IPTC. Une telle méthode n'est pas aisément généralisable (le réseau doit être affiné manuellement en fonction du corpus), mais sur un corpus réduit, elle permet d'offrir un classement thématique, certes rudimentaire, mais opératoire.

Cas de la collection Imprimés

A contrario, les imprimés se caractérisent par un riche appareil textuel (titraille et légende, texte précédant ou suivant l'illustration) qu'il est possible de thématiser. Les techniques de détection de thèmes (tâche non réalisée dans le cadre de cette expérimentation) seraient ici opérationnelles, par exemple la méthode de *topic modeling* LDA sans apprentissage supervisé (Underwood, 2012 ; Langlais, 2017 ; Velcin et al., 2017).

Dans le cas de la presse, médium polyphonique par essence, cette thématisation semble indispensable. Les corpus de presse numérisés avec une reconnaissance des articles (OLR) incluent parfois un rubriquage partiel, en général réalisé manuellement par les prestataires de numérisation (petites annonces, publicités, cours de la bourse, chroniques judiciaires, etc.). Notons que les illustrations de certaines revues thématiques (sciences, sports, arts militaires, etc.), dans une approche naïve, sont également assignables à un thème IPTC unique.

Indexation des images

La recherche dans des contenus image doit affronter une double non-coïncidence, d'une part la réalité du monde enregistrée dans une scène (dans notre contexte une « illustration ») et la description informationnelle de cette scène, et d'autre part l'interprétation d'une scène par différents utilisateurs ayant des objectifs de recherche possiblement différents. Réduire ou dépasser ces fossés sensoriel et sémantique implique notamment de fournir tant aux applications qu'à leurs utilisateurs des descripteurs opératoires (genres des illustrations, couleur, taille, texture, etc.), la recherche d'information opérant ensuite dans l'espace formé par ces descripteurs visuels.

La qualité est également un critère à prendre en compte, bien que par nature difficile à quantifier. Pour les photographies par exemple, un distinguo doit être établi entre tirage argentique et autres modes de reproduction dans les imprimés.

Taille, densité, localité, couleur

Quand elle n'est pas disponible dans les métadonnées de numérisation, le mode colorimétrique de chaque illustration est calculé. Un cas particulier concerne les documents originellement monochromatiques (noir et blanc, sépia, sélénium, etc.) numérisés en couleurs, cas qu'il conviendrait de traiter.

La localité, la taille et la densité des illustrations sont également extraites. Dans le cas de la presse quotidienne, interroger la seule page de « une » ou rechercher une illustration de grande taille sont représentatifs de cas d'usage courants et légitimes.

21. cv.iptc.org/newscodes/mediatopic

Classification des genres documentaires

Le genre des illustrations, considéré dans une acception large, de la technique de production (gravure, dessin, etc.) au type documentaire (cartes et plans, affiches, etc.), n'est pas toujours caractérisé dans les catalogues. Cette information n'est bien sûr pas plus disponible pour les illustrations présentes dans les imprimés.

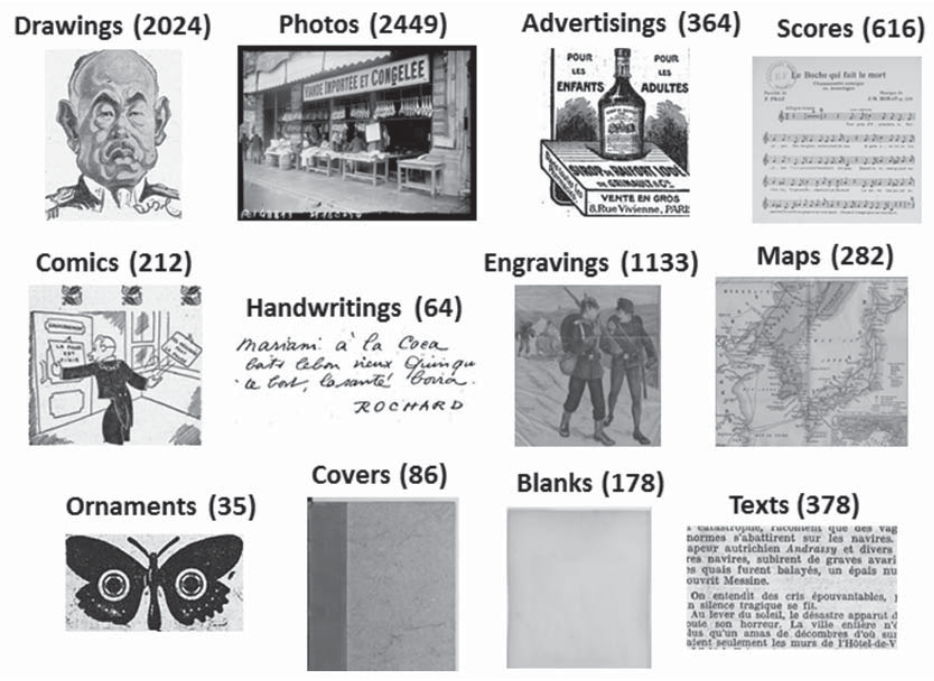
Afin de pallier ce manque, une méthode de classification automatique d'images par apprentissage profond est mise en œuvre[22]. Les réseaux de neurones modernes sont capables de reconnaître des milliers de types de concepts de la vie courante (bateau, table, chien, personne, etc.) et surpassent même les humains sur certains jeux de données. Le modèle Inception-v3 (Szegedy, Vanhoucke, Ioffe, Shlens et Wojna, 2016) est la troisième itération d'amélioration du modèle originel GoogLeNet (un réseau de neurones convolutif à 22 couches ayant gagné la compétition ILSVRC 2014). Les modèles de ce type sont généralement préentraînés sur des supercalculateurs et sont spécialement optimisés pour exceller sur des bases d'images telles que ImageNet, aujourd'hui une référence dans le domaine tant pour sa taille que pour sa représentativité (Deng, Dong, Socher, Li, Li et Fei-Fei, 2009). L'effort toujours plus important apporté à l'amélioration de ces modèles bénéficie à la communauté « vision par ordinateur » en général, mais pas seulement. En effet, il est aujourd'hui possible d'exploiter la puissance capitalisée par ces modèles sur des problématiques autres, ici la classification de documents patrimoniaux. Cela nécessite un réapprentissage léger d'une partie du modèle (en fait sa dernière couche, ce qui nécessite quelques heures de travail sur une machine classique, à comparer aux semaines qui seraient nécessaires pour réentraîner le modèle complet), en suivant une approche dite de *transfer learning* (Pan et Yang, 2010). Cette approche consiste à réutiliser les descripteurs visuels élémentaires trouvés durant la phase d'apprentissage originelle, dans la mesure où ils ont prouvé leur capacité à classifier un jeu de données, mais sur un nouveau jeu de données, avec l'espoir que ces descripteurs continueront à bien se comporter. De plus, en réduisant le nombre de classes, ce qui simplifie le problème en quelque sorte, il est tout à fait possible de conserver des scores honorables de classification, alors même que l'on utilise un modèle n'étant pas spécifiquement entraîné pour la tâche en question.

La figure 5 donne un aperçu des douze genres que l'on cherche à faire apprendre au modèle suivant l'approche décrite : bande dessinée, carte, dessin, gravure, écriture manuscrite, partition, photo, publicité, texte, page blanche, couverture et ornement.

Cet apprentissage nécessite de fournir un certain nombre d'illustrations étiquetées par leur classe. Dans notre cas, il s'agit d'une « vérité terrain[23] » (*ground truth*) d'environ 12 000 illustrations, en partie élaborée à l'aide de métadonnées catalogue. Une fois entraîné avec le framework

FIGURE 5

Les classes constituant le jeu de données d'apprentissage.



22. Cette tâche a été réalisée avec l'apport scientifique de Guillaume Chiron, laboratoire L3i, université de La Rochelle.

23. Information obtenue par observation directe de la réalité, par opposition à celle fournie par inférence.

Tensorflow[24], le modèle est ensuite évalué à l'aide d'un corpus de test. Le rappel (nombre de documents pertinents retrouvés par le classifieur au regard du nombre de documents pertinents que possède la base) et la précision (nombre de documents pertinents retrouvés au regard du nombre total de documents retrouvés par le classifieur) sont de 0,90 (moyenne pondérée des résultats de chaque classe). On peut considérer ces résultats comme appréciables : Viana, Nguyen, Smith et Gabrani (2017) rapportent une précision de 0,85 pour un scénario à 62 classes documentaires.

Notons que les performances sont meilleures avec un modèle moins générique (entrainé sur le seul corpus des monographies, avec six classes, le rappel s'établit à 0,95). Abandonner l'approche par *transfert learning* pour un modèle totalement entraîné (mais au prix d'un temps de calcul très supérieur) aurait également un effet bénéfique.

Ce modèle est également utilisé pour filtrer les illustrations indésirables qui auraient pu être manquées par le filtre heuristique de l'étape antérieure, en particulier les blocs de texte parasites de la presse et les couvertures et pages blanches des recueils d'images. Rappel et précision pour ces classes de bruit dépendent fortement de la difficulté de la tâche : de 98 % pour la collection Images à 85 % pour la presse. Enfin, le modèle est appliqué symétriquement pour tenter de « défiltrer » les faux positifs générés par le filtre heuristique.

Par contre, il ne permet pas (ou mal) d'isoler les illustrations publicitaires (qui représentent environ un tiers des illustrations de presse de notre corpus, volume qui appelle un traitement dédié). En effet, de nombreuses illustrations publicitaires ne sont pas distinguables visuellement du restant des contenus illustrés, puisqu'elles ressortent d'un type de communication et non d'une forme ou d'une technique graphique (voir figure 6).

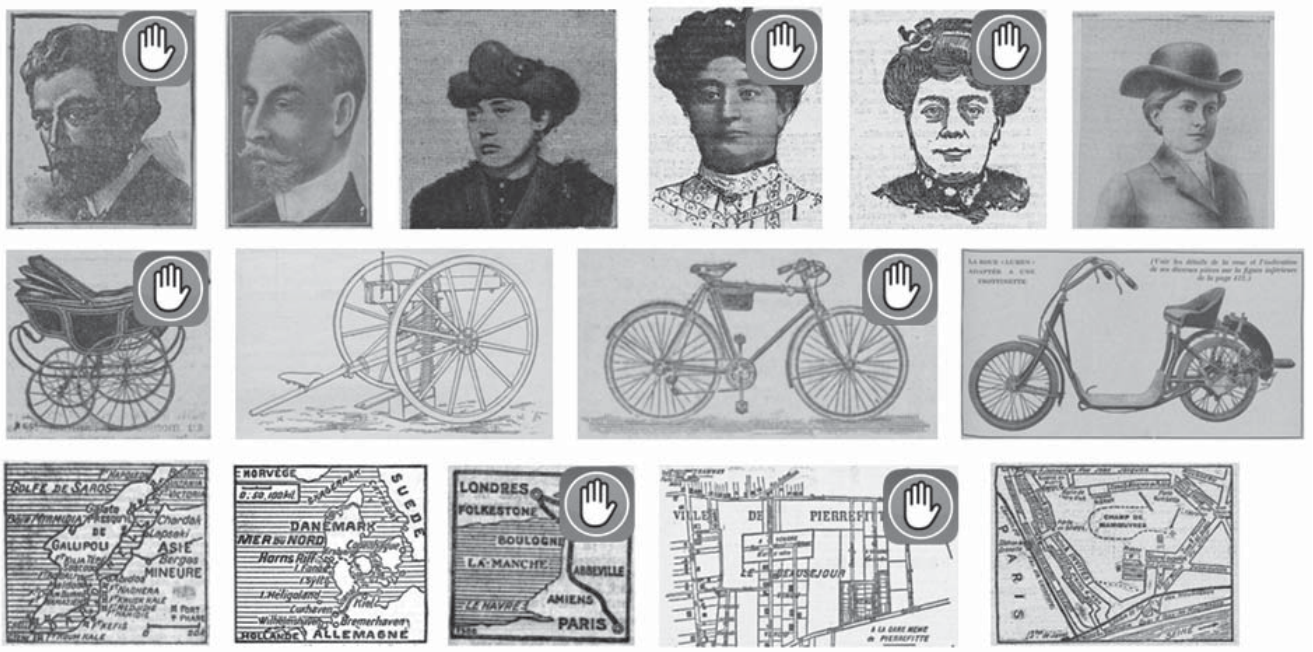
Il serait alors nécessaire d'envisager une recherche par similarité pour les publicités récurrentes, ou encore un modèle de classification bimodal opérant sur les contenus textuels (inclus et à proximité des illustrations) et visuels. Cette tâche n'a pas été réalisée dans le cadre de cette expérimentation.

L'utilisation d'un réseau de neurones convolutifs met en lumière des phénomènes intéressants. Ainsi, leur capacité à « généraliser » s'exprime par exemple dans les résultats de classification de documents hybrides, à la fois cartographiques et illustratifs, cas qui n'était pas présent dans le corpus d'apprentissage des véritables cartes (voir figures 7-1, 7-2 et 7-3).

Les réseaux de neurones sous-échantillonnent les images d'entrée. À une échelle de 299 × 299 pixels, cartes, partitions musicales et grandes pages de presse sont visuellement très proches.

FIGURE 6

Exemples de publicités illustrées de la presse quotidienne (repérées par une icône) et de contenus éditoriaux.



24. www.tensorflow.org/

FIGURE 7-1

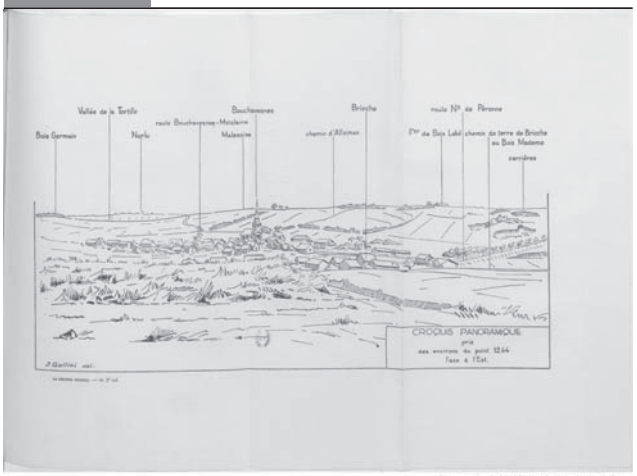
Exemples d'illustrations classées en tant que carte.

Source gallica.bnf.fr / Bibliothèque nationale de France

FIGURE 7-2

Source gallica.bnf.fr / Bibliothèque nationale de France

FIGURE 7-3



Source gallica.bnf.fr / Bibliothèque nationale de France

Enfin, en raison de la large couverture documentaire de la collection, même sur une courte période, des genres n'ayant pas été anticipés dans la base d'apprentissage peuvent apparaître (voir figure 8). De ce fait, une nouvelle classe d'apprentissage doit être construite et l'ensemble du modèle doit être réentraîné. La recherche vise aujourd'hui à éliminer cette lourdeur, avec des approches par apprentissage dynamique et incrémental, ou encore à l'aide de protocoles *low-shot learning* nécessitant peu de données d'apprentissage (Douze, Szlam, Hariharan et Jégou, 2018).

Classification des contenus image par reconnaissance visuelle

Historiquement, tout système de reconnaissance visuelle se devait d'extraire les descripteurs visuels d'une image, d'en déduire une signature numérique, puis d'opérer la recherche dans l'espace des signatures à l'aide d'une mesure de similarité, ce qui implique de fournir la requête sous la forme d'une signature (Datta, Joshi, Li et Wang, 2008). Cette dernière étant une image, cette contrainte a un impact négatif sur l'utilisabilité de ces systèmes (Breiteneder et Eidenberger, 2000 ; Wan et Liu, 2008). De plus, il est apparu qu'une mesure de similarité peinait à transcrire la richesse sémantique et la subjectivité d'interprétation des contenus image, en dépit des améliorations apportées (par exemple la prise en compte des sous-régions d'une image).

Plus récemment, les progrès des techniques d'apprentissage machine ont permis de dépasser ces limites, en particulier grâce à des approches dites de *clustering* et de classification (ou extraction de concepts), cette dernière ayant en outre l'avantage de laisser formuler la requête sous forme textuelle (Karpathy et Fei-Fei, 2017). Les API Visual Recognition d'IBM (qui s'inscrivent dans les services rendus par l'IA Watson[25]) et Cloud Vision[26] de Google sont représentatives de ces évolutions et les sections suivantes décrivent leur application à notre base d'images patrimoniales.

Ces API appliquent un modèle d'apprentissage profond à l'analyse d'images afin d'en extraire des concepts (objets, personnes, couleurs, etc.) identifiés au sein de la taxonomie des classes connues de l'API. Elles renvoient généralement des paires classe/estimation de confiance, et parfois l'emprise géométrique du concept au sein de l'image (on parle alors de « détection »). Une évaluation est menée sur la classification de personnes. Une vérité terrain est créée, couvrant la variété documentaire de la base. Une autre évaluation est menée sur la classe Soldat. Pour la classe Personne et l'API Watson[27], on constate un modeste taux de rappel de 55 %, mais une excellente précision de 98 %. Les

25. www.ibm.com/watson

26. cloud.google.com/vision

27. Appliquée sur la même vérité terrain, l'API Google Cloud Vision conduit à des performances proches (rappel : 48%, précision : 99%).

FIGURE 8

Nouveau genre dans l'illustration de presse : croquis, schéma technique, graphe.

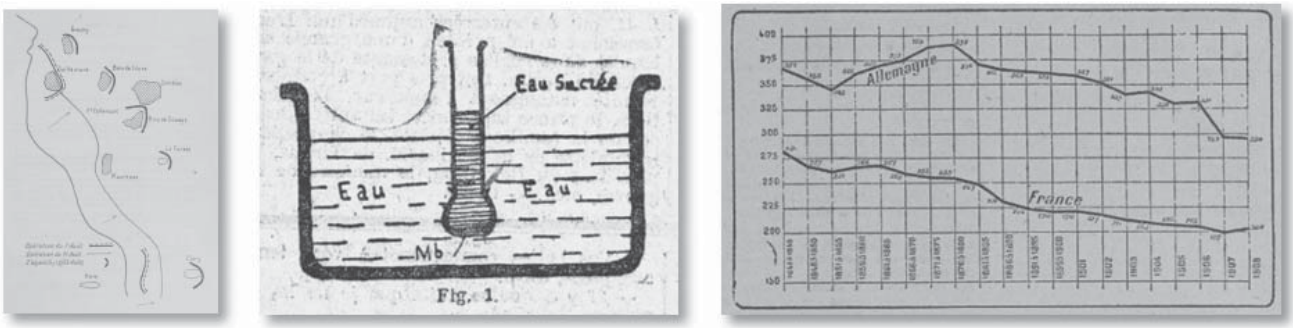
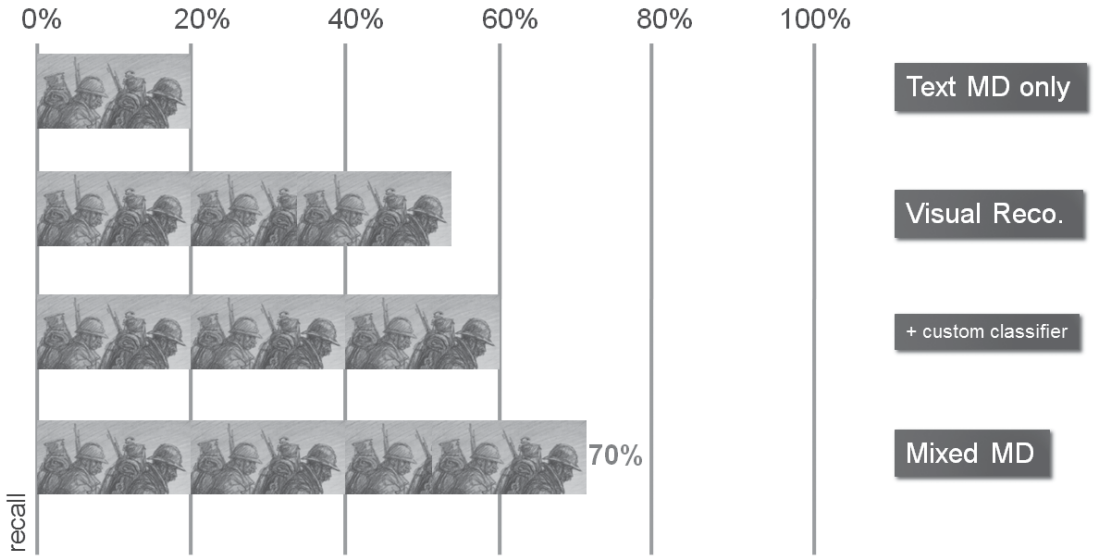


FIGURE 9

Taux de rappel pour la classe « Soldier » selon quatre modalités d'indexation.



taux sont moindres sur une classe plus spécialisée (Soldat, rappel : 50 %, précision : 80 %). Cependant, ces résultats sont à mettre en regard du relatif silence des approches classiques : le concept « soldat » n'existe pas dans les métadonnées bibliographiques (a fortiori pour des illustrations d'imprimés non cataloguées), et il serait nécessaire de composer une requête complexe du type « soldat OU officier OU militaire OU artilleur OU poilu… » pour aboutir à un taux de rappel de 20 %, à comparer aux 50 % obtenus avec la reconnaissance visuelle.

L'API Watson autorise également la création de classifieurs ad hoc (c'est-à-dire entraînés sur des données fournies par l'utilisateur). Une expérience sur la classe Soldat a apporté une amélioration significative des performances. La figure 9 résume les taux de rappel pour cette classe selon les quatre modalités analysées (descripteurs textuels, reconnaissance visuelle, reconnaissance visuelle avec classifieurs ad hoc, descripteurs textuels et visuels) et montre l'intérêt évident d'offrir aux utilisateurs un requêtage multimodal, qui aboutit ici à un rappel de 70 % (Wang, Yin, Wang, Wu et Wang, 2016).

Remarquons qu'un service générique tel que Watson se comporte convenablement sur des documents patrimoniaux de différents genres (voir figure 10), au-delà du seul genre de la photographie.

Cependant, certains types de document mettent en évidence les limites de ces approches par apprentissage. Le phénomène de généralisation conduit par exemple à des anachronismes ou des confusions (voir figure 11, respectivement à gauche et à droite). Gardons à l'esprit que les techniques d'apprentissage machine restent dépendantes des modalités selon lesquelles le corpus de d'entraînement a été créé (Ganascia, 2017). Les plus avancés de ces modèles, par exemple entraînés à reconnaître des chiens jouant au

FIGURE 10

Exemples de résultat pour la classe « Person ».

FIGURE 11

Une trottinette à moteur de 1917 classée en « Segway » et une locomotive à vapeur classée en « véhicule blindé ».

frisbee (Karpathy et Fei-Fei, 2017), ne seront bien sûr pas à leur avantage sur des documents du début du XX^{e} siècle.

Les scènes complexes (« multiobjets ») sont également des écueils que la recherche tente de dépasser (voir par exemple un modèle génératif de description en langage naturel d'images et de leurs zones d'intérêt (Karpathy et Fei-Fei, 2017)). Enfin, la qualité de la segmentation influe considérablement sur la reconnaissance visuelle : des documents numérisés avec leur cadre seront classés en tant que tels (et non d'après leur contenu, voir figure 12, à gauche) ; des imprimés à mise en page complexe seront associés à des classes génériques (« Affiche », « Document », « Presse ») de peu d'utilité (voir figure 12, au centre et à droite).

Les *gender studies* forment un champ de recherche à part entière et le réemploi de visuels numériques de visages humains pour des activités scientifiques (Ginosar et al., 2015) ou récréatives (Feaster, 2016) a ses praticiens. Il n'est donc pas anodin pour une bibliothèque numérique de prendre en compte de tels besoins, et ces API offrent en général un service de détection de visages (qui peut fournir en outre les âge et genre estimés des personnes).

Appliquer Watson[28] sur la vérité terrain conduit à un taux de rappel de 43 % pour les visages et une précision proche

28. L'API Google Cloud Vision se comporte de manière similaire (rappel : 40 %, précision : 100 %).

FIGURE 12

Exemple d'illustrations identifiées visuellement en classes génériques.

FIGURE 13

Exemples de résultats pour la détection des visages (IBM Watson).

de 100 %. Il s'agit d'un corpus difficile pour ce genre d'exercice, car incluant des dessins, gravures, défauts d'impression, visages de petite taille, etc. (voir figure 13).

Une dernière expérimentation de détection des visages est menée à l'aide du module dnn (*deep neural networks*) au sein de la bibliothèque OpenCV 3.3. Un réseau ResNet combiné à un détecteur SSD (*Single Shot MultiBox Detector*) est utilisé, produisant des taux de rappel supérieurs, quitte à sacrifier quelque peu la précision. Il apparaît qu'utiliser un framework de *deep learning* offre plus de souplesse qu'une API, ces dernières semblant paramétrées pour favoriser la précision dans le cas étudié. Là encore, on peut constater qu'un modèle même basique opère convenablement sur des données a priori difficiles (dessins de presse, caricatures, gravures ; voir figure 14).

FIGURE 14

Exemples de résultats pour la détection des visages (OpenCV/dnn).

FIGURE 15

Exemples de classification erronée : « lièvre » à gauche, « chat » à droite (au lieu de chien et lion, voir gallica.bnf.fr/ark:/12148/btv1b55008177v/f140, Liber quatuor tractatuum, Apulée, XIVe siècle).

Conclusion intermédiaire (2)

L'enrichissement consiste en un processus majoritairement linéaire, mais qui requiert de nombreux ajustements et une large panoplie d'outils (scripts, modèles entraînés, API, etc.) relevant du champ de l'informatique scientifique, domaine peu familier aux bibliothécaires. Là encore, les besoins en puissance informatique sont importants (qu'elle soit internalisée ou acquise dans le *cloud*), notamment lors de la phase d'apprentissage des modèles de classification.

À l'issue de cette phase d'enrichissement, les questionnements précédemment énoncés sont toujours valables, et certains se trouvent renforcés :

- volume : chaque illustration se voit enrichie de centaines de nouvelles métadonnées (étiquettes de classe sémantique, emprises géométriques, seuils de confiance, etc.) ;
- cycle de vie : on peut imaginer cette indexation visuelle comme étant hautement volatile, puisque produite par des techniques IA en permanente évolution.

Comme on l'a souligné, l'utilisation d'outils de reconnaissance visuelle soulève de nouvelles questions adressées tant aux ingénieurs qu'aux bibliothécaires, mais apportent aussi certains enseignements :

- les API commerciales rendent des services appréciables (en particulier sur des collections encyclopédiques des XIX^{e} et XX^{e} siècles), mais elles sont mises en difficulté sur des corpus spécialisés ou plus anciens. À titre d'exemple, les métaconcepts (animaux/mammifères) des enluminures présentées à la figure 15 sont convenablement identifiés, mais l'API Cloud Vision échoue à identifier l'espèce ;
- pour adapter ces techniques à de tels corpus, il faudra créer des modèles entraînés et/ou s'intéresser aux recherches en cours en matière d'apprentissage dynamique et incrémental et proposer aux équipes de recherche concernées des collaborations ;
- ces API et modèles de classification fournissent des étiquettes de classe pouvant aisément être ingérées par les moteurs d'indexation des portails documentaires[29] ;
- les taxonomies utilisées par les API commerciales ne sont pas interopérables, tant de par leur organisation interne (certaines sont plates, d'autres hiérarchiques) que du fait des choix de vocabulaire réalisés[30] ;
- une bonne segmentation des illustrations est un préalable majeur ; or il n'existe pas d'algorithme de segmentation universel (par exemple œuvrant indifféremment sur les enluminures de manuscrit et les illustrations de presse) ;
- le choix effectif des méthodes et outils peut être délicat : le domaine du *deep learning* est aujourd'hui foisonnant, facteur invitant et favorisant l'expérimentation, voire l'implémentation opérationnelle, mais dans le même temps complexe à appréhender (il existe de nombreuses offres commerciales, encore plus de frameworks, modèles préentraînés et autres architectures de modèle) ;
- enfin, l'application de ces techniques à grande échelle implique des infrastructures informatiques adaptées.

Charger et interroger : GallicaPix

GallicaPix est hébergé par l'espace contributif Gallica Studio[31]. Les métadonnées chargées dans BaseX sont interrogées en mode client/serveur REST à l'aide d'un formulaire HTML et d'expressions XQuery. La mosaïque d'images est créée avec la bibliothèque Javascript Mansory et peuplée par le serveur IIIF de Gallica. Un système rudimentaire de facettes (couleur, taille, type, date, etc.) permet d'imaginer ce que serait une interface utilisateur plus aboutie.

La complexité du formulaire de recherche et le très grand nombre de résultats qu'il fournit rappelle, s'il en était besoin, que chercher et naviguer dans des bases d'images de grande dimension pose des problèmes spécifiques d'usage et constitue un sujet de recherche à part entière (voir par exemple Lai, Visani, Boucher et Ogier, 2014). Ainsi les modalités opérationnelles de l'interrogation multimodale par les contenus image et par les descripteurs textuels (Wang, Yin, Wang, Wu et Wang, 2016) doivent être rendues intelligibles aux utilisateurs. De même, il faudra expliciter la présence inévitable d'erreurs de classification et de bruit (les fausses illustrations ayant échappé aux diverses stratégies de filtrage) dans les résultats affichés. Cela étant, ce mode de recherche contribue à réduire l'écart entre la formulation du besoin utilisateur (par exemple « une caricature de presse de George Clémenceau en 1914 ») et le modèle de données intelligible au système d'information, comme l'illustrent les exemples suivants, qui détaillent des cas d'usage représentatifs de l'interrogation d'une base d'images patrimoniales à vocation encyclopédique.

Requête sur un genre iconographique

La classification automatique des genres permet désormais de cibler un genre particulier tout en prenant en considération la diversité des collections. Le cas des cartes et plans est significatif à cet égard (voir figure 16), puisque pour la période et le sujet considérés, Gallica offre 150 documents et GallicaPix plus de 13 000, en grande majorité extraits de la presse quotidienne (cartes du front et d'état-major, plans de mouvement de troupe, etc.).

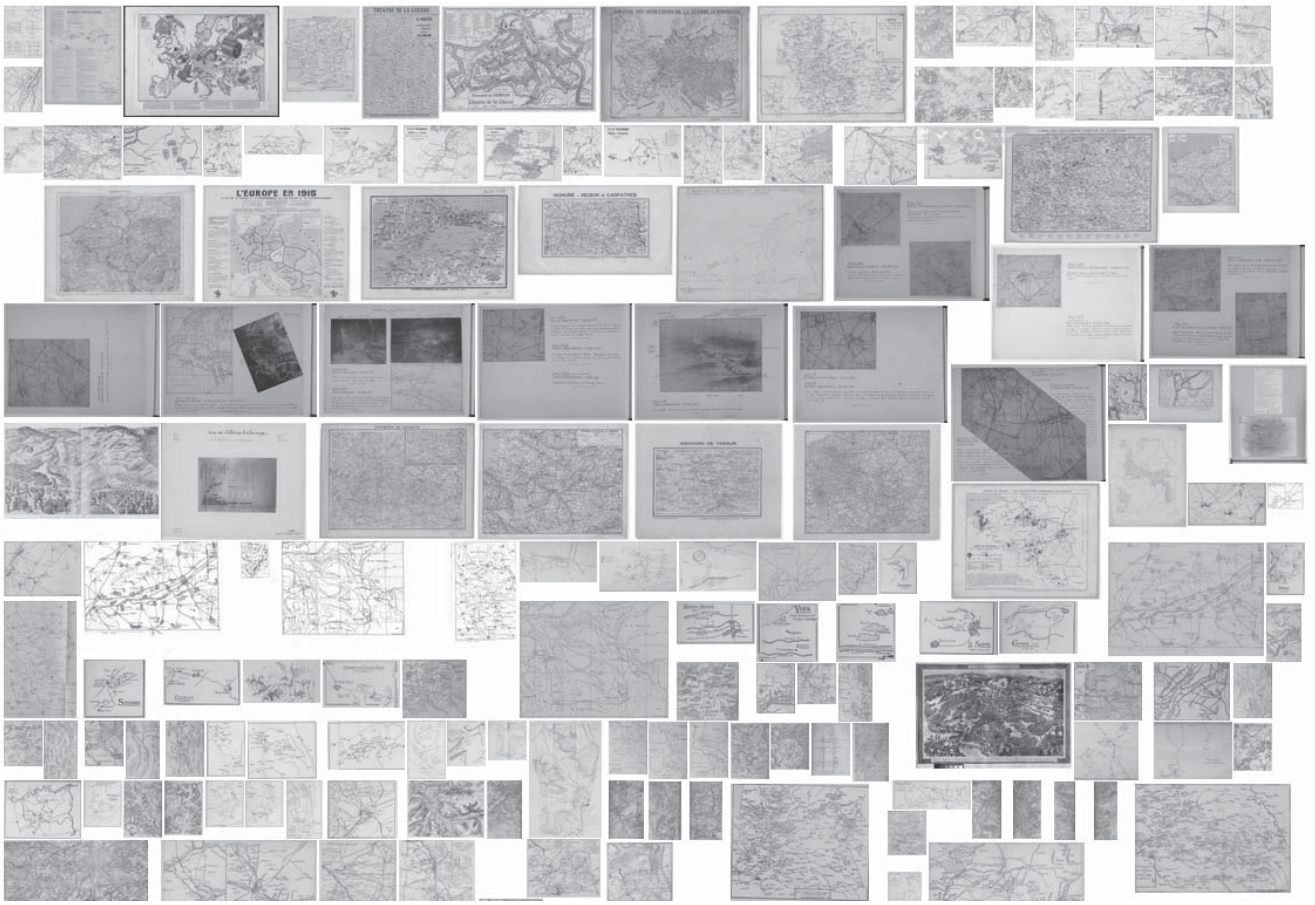
29. Ce qui n'est pas le cas des signatures numériques (voir le paragraphe sur l'indexation des images), qui appellent des algorithmes dédiés de recherche dans des espaces de grande dimension.

30. Ainsi, pour décrire la couleur orange, Watson utilisera l'étiquette « orange color » et Cloud Vision « orangered » ; pour un tank, respectivement « army tank » et « tank », etc. Et certains concepts n'existent que dans une API (seul Cloud Vision propose par exemple la classe « aerial photography »).

31. gallicastudio.bnf.fr/bo%C3%AEte-%C3%A0-outils/plongez-dans-les-images-de-14-18-en-testant-un-nouveau-moteur-de-recherche

FIGURE 16

Requête : genre = « carte » (extraits).



Requête sur une entité nommée

Les descripteurs textuels (métadonnées et OCR) sont mis à contribution. Une des cent requêtes les plus fréquentes sur une entité nommée de type Personne exprimées par les utilisateurs de Gallica porte sur Georges Clémenceau (avec 161 documents sur la période 1910-1920). La même requête dans GallicaPix renvoie désormais plus de 900 illustrations (voir figures 17-1 et 17-2), issues d'un large spectre de genres. Les facettes peuvent ensuite être mises en œuvre pour affiner la recherche (par exemple en filtrant les dessins et autres caricatures de presse).

Dans ce cas d'usage, précision et rappel sont liés à la qualité des descripteurs textuels, mais la recherche de caricatures a été rendue possible grâce à la classification des genres.

Requête sur un concept

Les classes conceptuelles extraites par les API d'indexation visuelle permettent de lever le silence des métadonnées bibliographiques et de l'OCR (ou au contraire leur trop grand bavardage), mais aussi de contourner les difficultés propres aux corpus multilingues (certains documents du corpus sont d'origine allemande) et à l'évolution des lexiques (par exemple le vieilli « casemate », aujourd'hui « bunker »).

Dans ce cas d'usage, l'utilisateur n'attend généralement pas exhaustivité et précision, mais plutôt des propositions. Dans le contexte de la Grande Guerre, on pense aux soldats, véhicules, armes, etc. Prenons l'argument d'une requête sur le mot-clé « avion », qui produit de nombreuses illustrations hors de propos (portraits d'aviateur, photographies aériennes, cartes d'état-major, etc.). Au contraire, la classe « Avion » fournit une palette d'illustrations plus restreinte, mais de meilleure précision, et des portraits d'aviateur à bord de leur machine volante pourront être isolés à l'aide de la facette « Personne » (voir figures 18-1 et 18-2).

FIGURE 17-1

Requête : mot-clé = « Clémenceau » (extraits) ; en dessous, avec la facette « Dessin ».

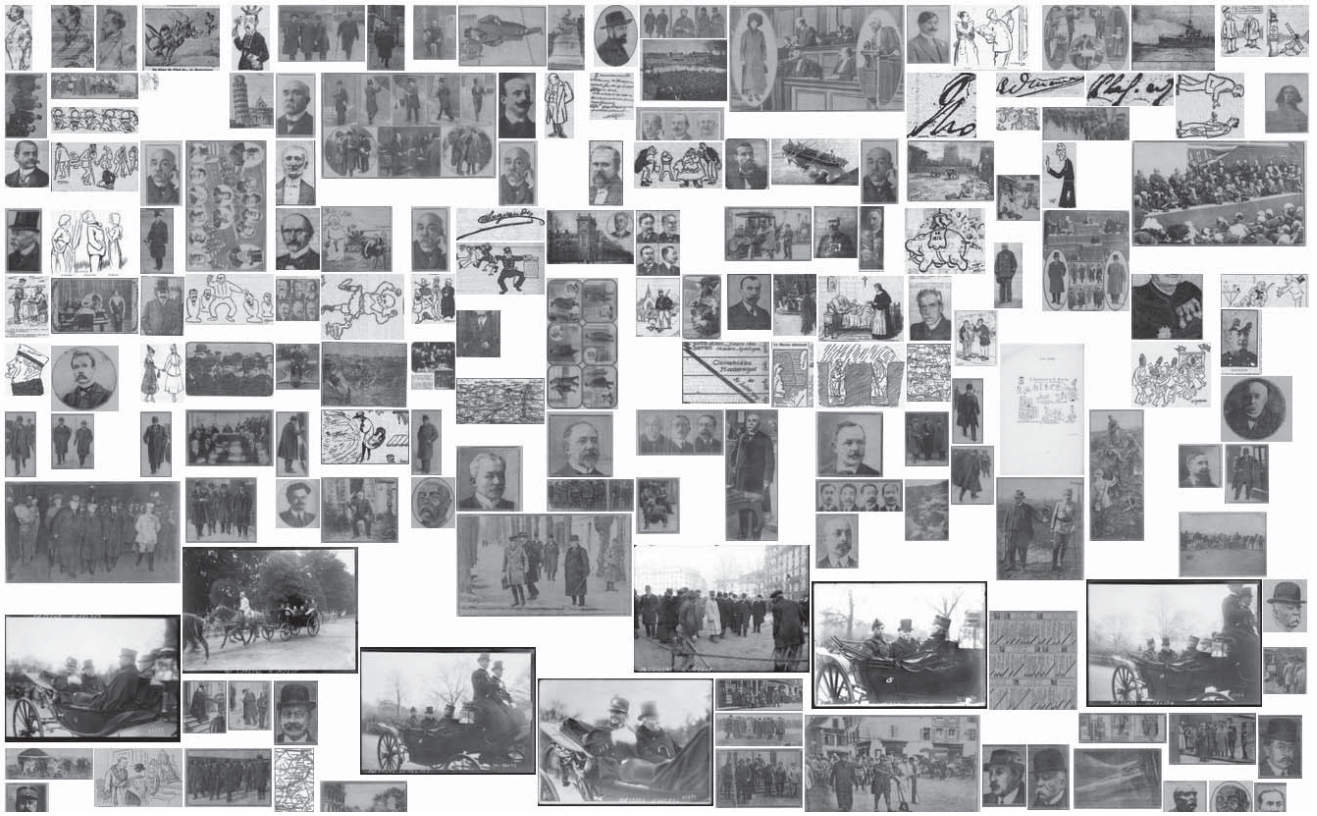


FIGURE 17-2

FIGURE 18-1

Requête : classe = « Avion » (extraits) ; en dessous, classes « Avion » et « Personne ».

FIGURE 18-2

FIGURE 19

Requête : classe = « Street » ET mot-clé = « Verdun » (extraits).

Requête multimodale

L'utilisation conjointe des métadonnées et des classes d'indexation visuelle autorise l'expression de requêtes avancées. La figure suivante montre l'exemple d'une recherche se rapportant aux destructions urbaines consécutives à la bataille de Verdun, exprimée à l'aide des classes « Rue », « Maison » et « Ruines », ainsi que du mot-clé « Verdun » (voir figure 19).

Autre exemple, une étude de l'évolution des uniformes des soldats français au cours du conflit pourra s'appuyer sur deux requêtes mettant en œuvre les classes conceptuelles (« Soldier », « Officer », etc.), une donnée bibliographique (date) et un critère image (mode colorimétrique = « couleur »), afin de mettre en évidence en quelques clics l'histoire du célèbre pantalon garance porté jusqu'au début de l'année 1915 (voir figures 20-1 et 20-2).

Perspectives

Diffusion des métadonnées visuelles

Le développement du PoC GallicaPix a été rendu possible du fait de la dynamique d'ouverture des données engagée par les bibliothèques, lesquelles proposent des API d'accès aux métadonnées et aux documents, avec en particulier le support du standard IIIF[32]. Symétriquement, les métadonnées que ce dernier a produites gagneraient à être pérennisées afin de favoriser leur réutilisation, tant par les systèmes d'information et applicatifs internes de la bibliothèque que par ses usagers, via les divers services d'accès aux données. À cet effet, l'API IIIF Presentation[33] offre un moyen élégant de décrire dans le manifeste IIIF les illustrations présentes dans un document, sous la forme d'une liste d'annotations (W3C Open Annotation) attachée à un calque (*canvas*).

La totalité des ressources iconographiques devient alors accessible à un visualiseur compatible IIIF, mais aussi actionnable par machine, pour des projets propres à la bibliothèque, pour le moissonnage de données par d'autres bibliothèques (Freire, Robson, Howard, Manguinhas et Isaac, 2017), ou encore à l'usage des communautés GLAM, hackers et autres *makers*.

On peut alors considérer que l'expansion en cours du protocole IIIF, avec pour corollaire la multiplication des portails documentaires dotés d'un jeu d'API et d'un serveur IIIF,

32 Notons qu'il aurait pu être développé par un tiers, de l'extérieur de la bibliothèque, et que ces API facilitent aussi les développements internes ou externalisés.

33 iiif.io/api/presentation/2.1

FIGURE 20-1

Requête : classe = « Soldier » ET mode = « couleur » ET date avant 31/12/1914.

FIGURE 20-2

Requête : classe = « Soldier » ET mode = « couleur » ET date après 01/01/1915.

contribuera à la généralisation de l'approche décentralisée suivie par ce PoC : moissonnage de ressources iconographiques auprès d'institutions patrimoniales (musées, bibliothèques, archives), traitements spécifiques des images, exposition des résultats dans des manifestes IIIF, lesquels pourraient en retour être consommés par les institutions détentrices des fonds (voir figure 21).

Passage à l'échelle

Le PoC GallicaPix a été mis en ligne pour test public et des améliorations lui sont apportées régulièrement (par exemple le moissonnage de la collection Welcome Library « WW1 » via Europeana, voir figure 21). Comme il a été démontré, il permet de satisfaire des cas d'usage classiques en matière de recherche iconographique, et Gallica pourra puiser dans les solutions esquissées lors de sa mise en œuvre de telles fonctionnalités, envisagée à l'horizon 2020.

L'industrialisation du processus d'extraction et d'enrichissement (voir figure 22) devra bien sûr prendre à bras-le-corps les enjeux techniques exposés par GallicaPix. De plus, au-delà de l'indispensable pipeline de prétraitement des données, il faudra développer le back-end et le front-end de l'application, le tout formant un système d'information complexe.

De futurs travaux devront aussi être consacrés aux défis d'utilisabilité posés par la navigation et la recherche dans un grand volume d'images : partitionnement visuel des

FIGURE 21

Bibliothèques ouvertes, ressources iconographiques et IIIF.

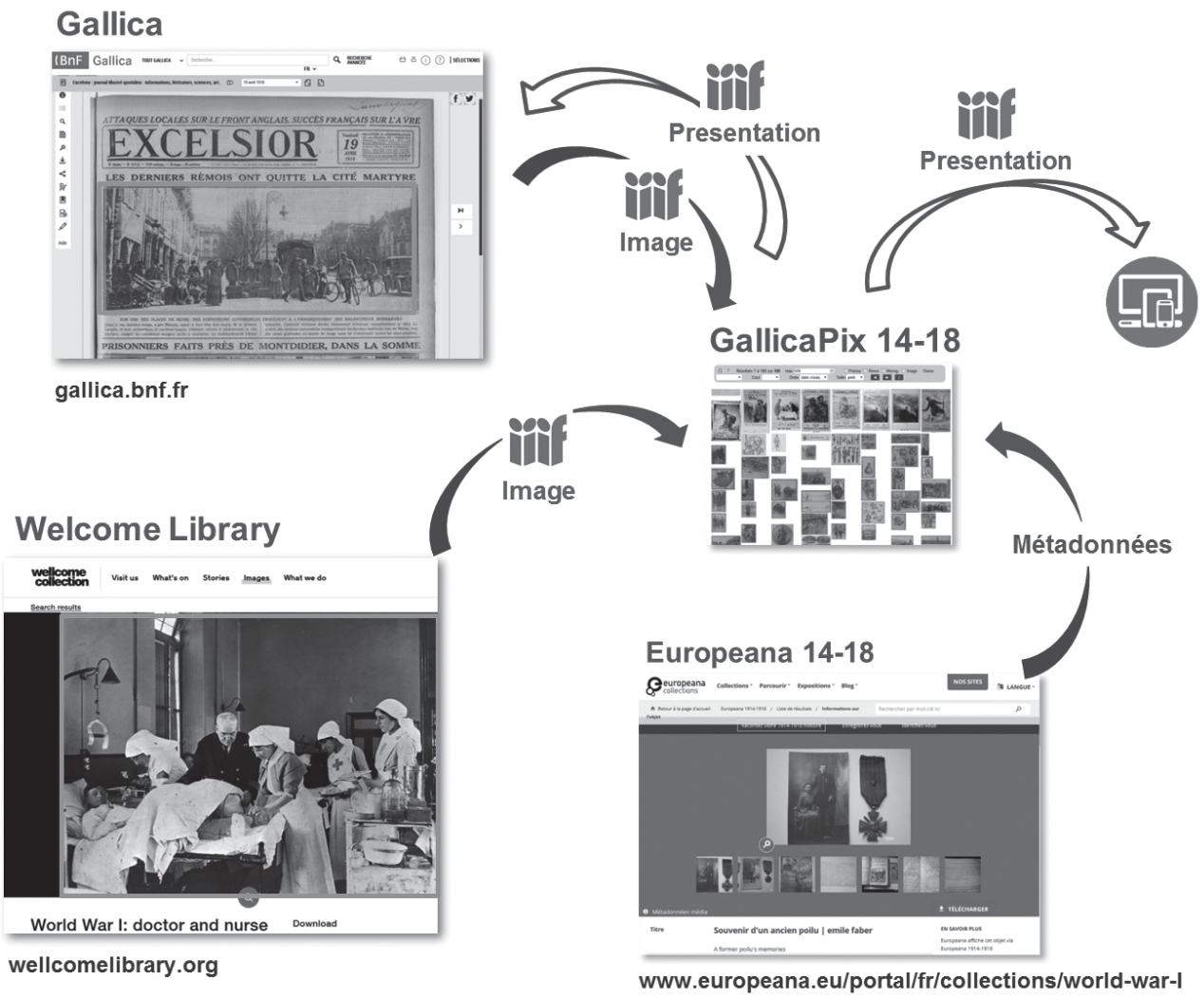
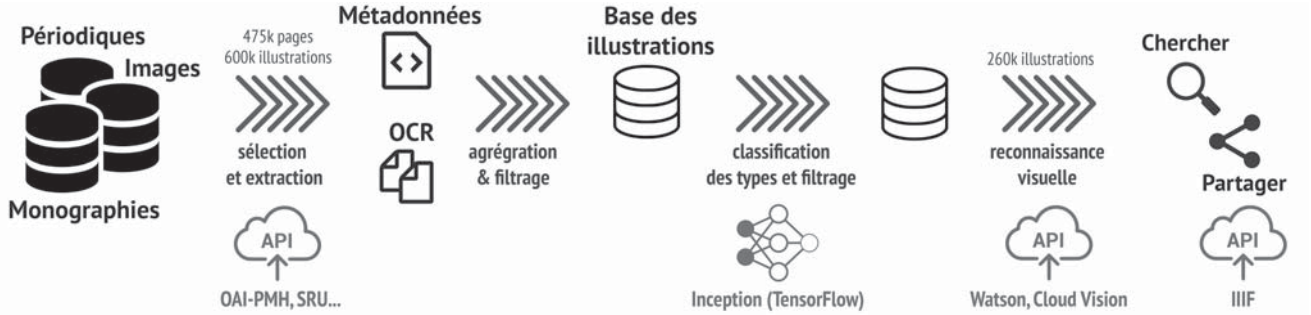


FIGURE 22

Schéma de principe du pipeline de traitement.



résultats, recherche itérative pilotée par les retours de l'utilisateur (Picard, Gosselin et Gaspard, 2015), ainsi qu'un mode de recherche complémentaire, celui de la recherche par similarité visuelle, qui n'a pas été abordé ici (voir par exemple l'exposé de Raphaël Baudiment dans Moiraghi, 2018, ainsi que la réalisation de la Bayerische Staatsbibliothek en note 13).

Humanités numériques et bibliothèques numériques

Les chercheurs en sciences humaines et sociales et en histoire de l'art expriment un intérêt encore limité (Gunthert, 2017), mais cependant croissant pour les études ayant pour terrain d'investigation des corpus visuels numériques (Ginosar et al., 2015 ; Moiraghi, 2018). Mais prendre en compte ces nouveaux publics et leurs usages offre l'occasion aux bibliothèques numériques de décliner le modèle de collaboration scientifique déjà en vigueur pour d'autres types de corpus, notamment textuels (Bermès, 2017) : en ouvrant ses collections numériques et en favorisant leur appropriation par les chercheurs, la bibliothèque devient partenaire à part entière de projets de recherche pluridisciplinaires, au cours desquels elle s'approprie techniques, outils et méthodes qui en retour lui permettront de mieux remplir ses missions (décrire un patrimoine, le rendre accessible, le faire connaître).

Illustrons ce propos avec la genèse de GallicaPix, qui croise le projet de recherche européen Europeana Newspapers et une collaboration avec l'université de La Rochelle. Au terme de l'expérimentation, la base iconographique GallicaPix, mise à disposition sur api.bnf.fr, devient la matrice de corpus visuels thématiques ou documentaires pour les humanités numériques, mais aussi pour la recherche fondamentale en informatique, ce type de bases annotées se révélant précieux en tant que données d'entrainement (pour les approches par *machine learning* et *deep learning*) et données de validation. Puis d'autres acteurs des deux champs disciplinaires évoqués, découvrant de telles ressources, sollicitent la bibliothèque afin d'engager de nouvelles collaborations. On voit poindre ici un modèle vertueux de co-construction d'une connaissance du patrimoine, chacun agissant selon des besoins propres mais tous mutualisant pratiques et outils.

Conclusion

L'accès unifié à toutes les illustrations d'une collection numérique encyclopédique est un service innovant répondant à un besoin attesté. Il participe de l'effort de valorisation des contenus patrimoniaux numériques à la granularité adéquate (ce qui implique d'abandonner le confortable modèle de l'image numérisée et d'entrer de plain-pied dans cette dernière), ainsi que de la politique d'ouverture des données (visant à favoriser leur réutilisation). Dans le contexte de ces deux enjeux, le protocole IIIF semble destiné à jouer un rôle majeur, en permettant d'exposer et de mutualiser les ressources iconographiques numérisées, toujours plus nombreuses à intégrer les entrepôts patrimoniaux.

Dans le même temps, la maturité des techniques d'IA en matière de traitement d'images encourage à les intégrer dans la panoplie technique des bibliothèques numériques. Leurs résultats, mêmes imparfaits, contribuent à rendre visibles les riches ressources iconographiques de nos collections, dont l'indexation manuelle est quoi qu'il en soit hors de portée. Mais il conviendra d'échapper à la tentation d'appliquer ces outils sans les comprendre ni les adapter au contexte particulier du patrimoine, et pour ce faire, adopter une stratégie proactive est indispensable.

On peut imaginer que la conjonction de cette abondance de contenus numériques, d'un contexte technique favorable et d'une politique de collaboration scientifique volontariste, permettra d'ouvrir à court terme un nouveau champ d'investigation pour les chercheurs et de nouveaux services de recherche iconographique pour tous les usagers[34].

34 Les jeux de données, modèles entraînés et scripts sont disponibles à github.com/altomator/Image_Retrieval

SOURCES CONSULTÉES

Bermès, E. (2017, août). *Text, Data and Link-Mining in Digital Libraries : Looking for the Heritage Gold*. Communication présentée à la conférence IFLA Satellite Meeting, Digital Humanities – Opportunities and Risks : Connecting Libraries and Research, Berlin, Allemagne. Repéré à www.ifla.org/files/assets/academic-and-research-libraries/conferences/emmanuelle_bermes_keynote.pdf

Bibliothèque nationale de France (BnF). (2017). *Enquête auprès des usagers de la bibliothèque numérique Gallica*. Repéré à www.bnf.fr/documents/mettre_en_ligne_patrimoine_enquete.pdf

Breiteneder, C. et Eidenberger, H. (2000, février). *Content-Based Image Retrieval in Digital Libraries*. Communication présentée à la Conférence internationale de Kyoto sur les bibliothèques numériques, Japon. doi.org/10.1109/DLRP.2000.942186

Chiron, G., Doucet, A., Coustaty, M., Visani, M. et Moreux, J.-P. (2017, juin). *Impact of OCR Errors on the Use of Digital Libraries*. Communication présentée à la 17e conférence commune ACM/IEEE sur les bibliothèques numériques, Toronto, Ontario. doi.ieeecomputersociety.org/10.1109/JCDL.2017.7991582

Coustaty, M., Pareti, R., Vincent, N. et Ogier, J.-M. (2011). Towards Historical Document Indexing : Extraction of Drop Cap Letters. *International Journal on Document Analysis and Recognition, 14*(3), 243-254. Repéré à hal.archives-ouvertes.fr/hal-00916007/document

Datta, R., Joshi, D., Li, J. et Wang, J. (2008). Image Retrieval : Ideas, Influences, and Trends of the New Age. *ACM Computing surveys, 40*(2), [5]. Repéré à infolab.stanford.edu/~wangz/project/imsearch/review/JOUR/datta.pdf

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K. et Fei-Fei, L. (2009, juin). *ImageNet : A Large-Scale Hierarchical Image Database*. Communication présentée à la conférence « IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2009 », Miami, Floride. doi.org/10.1109/CVPRW.2009.5206848

Douze, M., Szlam, A., Hariharan, B. et Jégou, H. (2018, juin). *Low-Shot Learning with Large-Scale Diffusion*. Communication présentée à la conférence « IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2018 », Salt Lake City, Utah. Repéré à openaccess.thecvf.com/content_cvpr_2018/papers/Douze_Low-Shot_Learning_With_CVPR_2018_paper.pdf

Feaster, P. (2016, 31 octobre). Time-Based Image Averaging [Billet de blogue]. Repéré à griffonagedotcom.wordpress.com/2016/10/31/time-based-image-averaging.

Freire, N., Robson, G., Howard, J. B., Manguinhas, H. et Isaac, A. (2017). Metadata Aggregation : Assessing the Application of IIIF and Sitemaps Within Cultural Heritage. Dans J. Kamps, G. Tsakonas, Y. Manolopoulos, L. Illiadis et I. Karydis (dir.), *Research and Advanced Technology for Digital Libraries. TPDL 2017*. doi.org/10.1007/978-3-319-67008-9_18

Ganascia, J.-G. (2017). *Le mythe de la Singularité. Faut-il craindre l'intelligence artificielle ?* Paris, France : Le Seuil.

Ginosar, S., Rakelly, K., Sachs, S., Yin, B., Lee, C., Krähenbühl, P. et Efros, A. A. (2015). A Century of Portraits : A Visual Historical Record of American High School Yearbooks. *IEEE Transactions on Computational Imaging, 3*(3), 421-431.

Gordea, S. et Haskiya, D. (2017). Europeana DSI 2 – Access to Digital Resources of European Heritage. MS6.1 : Advanced Image Discovery Development Plan. Repéré à https ://pro.europeana.eu/files/Europeana_Professional/Projects/Project_list/Europeana_DSI-2/Milestones/ms6.3-advanced-image-discovery-development-plan.pdf

Gunthert, A. (2017, 10 juin). Le « visual turn » n'a pas eu lieu [Billet de blogue]. Repéré à imagesociale.fr/4603

Karpathy, A. et Fei-Fei, L. (2017). Deep Visual-Semantic Alignments for Generating Image Descriptions. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 39*(4), 664-676. doi.org/10.1109/TPAMI.2016.2598339

Lai, H. P., Visani, M., Boucher, A. et Ogier, J.-M. (2014). A New Interactive Semi-Supervised Clustering Model for Large Image Database Indexing. *Pattern Recognition Letters, 37*, 94-106. doi.org/10.1016/j.patrec.2013.06.014

Langlais, P.-C. (2017). Identifier les rubriques de presse ancienne avec du *topic modeling*. Repéré à numapresse.hypotheses.org

Moiraghi, E. (2018). Explorer des corpus d'images. L'IA au service du patrimoine. Repéré à bnf.hypotheses.org/2809

Moreux, J.-P. (2016). Innovative Approaches of Historical Newspapers : Data Mining, Data Visualization, Semantic Enrichment. Facilitating Access for Various Profiles of Users. Repéré à http://library.ifla.org/2076/1/S21-2016-moreux-en.pdf

Nottamkandath, A., Oosterman, J., Ceolin, D. et Fokkink, W. (2014). Automated Evaluation of Crowdsourced Annotations in the Cultural Heritage Domain. *URSW'14 Proceedings of the 10th International Workshop on Uncertainty Reasoning for the Semantic Web, 1259*, 25-36. Repéré à http://ceur-ws.org/Vol-1259/ursw2014_submission_5.pdf

Pan, S. J. et Yang, Q. (2010). A Survey on Transfer Learning. *IEEE Transactions on Knowledge and Data Engineering, 22*(10), 1345-1359. doi.org/10.1109/TKDE.2009.191

Picard, D., Gosselin, P.-H. et Gaspard, M.-C. (2015). Challenges in Content-Based Image Indexing of Cultural Heritage Collections. *IEEE Signal Processing Magazine, 32*(4), 95-102. Repéré à hal.archives-ouvertes.fr/hal-01164409/document

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J. et Wojna, Z. (2016, juin). *Rethinking the Inception Architecture for Computer Vision*. Communication présentée à la conférence « 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) », Nevada, États-Unis. doi.org/10.1109/CVPR.2016.308

Underwood, T. (2012, 7 avril). Topic Modeling Made Just Simple Enough [Billet de blogue]. Repéré à tedunderwood.com/2012/04/07/topic-modeling-made-just-simple-enough

Velcin, J., Soulages, J.-C., Kurpiel, S., Dias, L., Del Vecchio, M. et Aubrun, F. (2017). Fouille de textes pour une analyse comparée de l'information diffusée par les médias en ligne : une étude sur trois éditions du Huffington Post. Repéré à hal.archives-ouvertes.fr/hal-01571265/document

Viana, M., Nguyen, Q.-B., Smith, J. et Gabrani, M. (2017, novembre). *Multimodal Classification of Document Embedded Images*. Communication présentée à la conférence « 12th IAPR International Workshop, GREC 2017 », Kyoto, Japon. http://doi.org/10.1007/978-3-030-02284-6_4

Wan, G. et Liu, Z. (2008). Content-Based Information Retrieval and Digital Libraries. *Information Technology and Librairies, 27*(1), 41-47. doi.org/10.6017/ital.v27i1.3262

Wang, K., Yin, Q., Wang, W., Wu, S. et Wang, L. (2016). A Comprehensive Survey on Cross-Modal Retrieval. Repéré à arxiv.org/pdf/1607.06215.pdf

Welinder, P., Branson, S., Belongie, S. et Perona, P. (2010). The Multidimensional Wisdom of Crowds. *NIPS'10 Proceedings of the 23rd International Conference on Neural Information Processing Systems, 2*, 2424-2432.