

L'émergence brutale de la discipline jusque la discrète qu'était « l'intelligence artificielle » dans la sphère publique depuis 2022 a grandement augmenté l'intérêt des équipes de recherche et des institutions patrimoniales pour ces technologies.

Si officiellement le terme Intelligence Artificielle recouvre le « Champ interdisciplinaire théorique et pratique qui a pour objet la compréhension de mécanismes de la cognition et de la réflexion, et leur imitation par un dispositif matériel et logiciel, à des fins d'assistance ou de substitution à des activités humaines »¹, dans ce texte, nous utiliserons ce terme comme une manière de désigner un ensemble de techniques issues de l'apprentissage-machine, ou *machine learning*. L'apprentissage machine est un ensemble de technologies faisant appel à des algorithmes statistiques pouvant apprendre à partir de données, et généraliser les résultats de cet apprentissage à des données inconnues. Si c'est un champ d'étude ancien, théorisé dès la fin des années 1950, il a connu des progrès spectaculaires depuis les années 2010, notamment lié à l'augmentation exponentielle des capacités de calcul disponibles.

Pour travailler sur des collections patrimoniales, nous allons nous intéresser à des technologies de vision par ordinateur (ou *computer vision*). Ces technologies ont comme point commun de partir d'images pour en extraire les informations que l'on recherche. Les principaux usages de ces technologies vont être par exemple les véhicules autonomes, ou le traitement d'imagerie médicale. Si elles sont très puissantes, très peu de ces technologies sont nativement adaptées au contexte patrimonial. Leur usage « commercial » premier leur permet grossièrement de ne considérer les images que sous l'angle de leurs caractéristiques visuelles directes.



Figure 1: Pourtant mon modèle est certain qu'il s'agit d'une pipe ! - *La Trahison des images*, René Magritte, 1928-29, LACMA.

Si l'exemple ci-dessus, avec la fameuse pipe de Magritte², est celui d'une œuvre dans lequel le paradoxe est pleinement assumé. Les mêmes difficultés se posent dans de nombreux autres

1 J.O. du 9 décembre 2018

2 Inspiré par l'article de E. Villaespesa et O. Murphy, « This is not an apple ! Benefits and challenges of applying computer vision to museum collections », Taylor & Francis, 2020

contextes, comme celui d'un personnage mythologique identifiable non pas à son apparence générale, mais à un attribut bien particulier qui le caractérise, dans le cadre d'objets anthropomorphes et zoomorphes ou encore dans le contexte d'œuvres abstraites. Le défi pour les institutions patrimoniales a été et est toujours d'adapter ces technologies aux questions spécifiques qui les concernent. En plus de cela, l'utilisation de petits modèles³ très spécialisées induit une économie d'énergie non négligeable comparativement aux gros modèles généralistes les plus connus.

Usages de l'IA pour l'étude des collections patrimoniales

Dans ce contexte, un certain nombre d'usages ont été identifiés pour la vision par ordinateur pour le traitement des collections. Ces usages sont principalement centrées autour des missions A et D des musées telles que définies par l'article L.441-2 du Code du patrimoine, c'est à dire : « Conserver, restaurer, étudier et enrichir leurs collections », et « Contribuer aux progrès de la connaissance et de la recherche ainsi qu'à leur diffusion ». Si l'on devait découper ces usages en 3 grands groupes, il s'agirait de la transcription de texte, de la description et classification d'images, et enfin de la navigation dans des grands corpus d'images.

La transcription de textes par ordinateur est une technologie ancienne. Dès 1965, la poste américaine utilise des machines capables de lire les codes postaux pour le tri du courrier. Ces technologies sont regroupées sous le nom d'OCR (*Optical Character Recognition*) pour le texte imprimé, et d'HTR (*Handwritten Character Recognition*) pour les textes manuscrits. Les technologies d'OCR sont aujourd'hui capables de transcrire sans les spécialiser la plupart des polices, et la principale difficulté reposera dans le fait de restituer l'agencement des différents textes d'une page. L'HTR, pour sa part, du fait de la plus grande variabilité des écritures manuscrites d'un individu à l'autre, et à travers le temps, est une technologie qui demande généralement un travail plus fin pour adapter les modèles IA aux besoins de chaque collection.

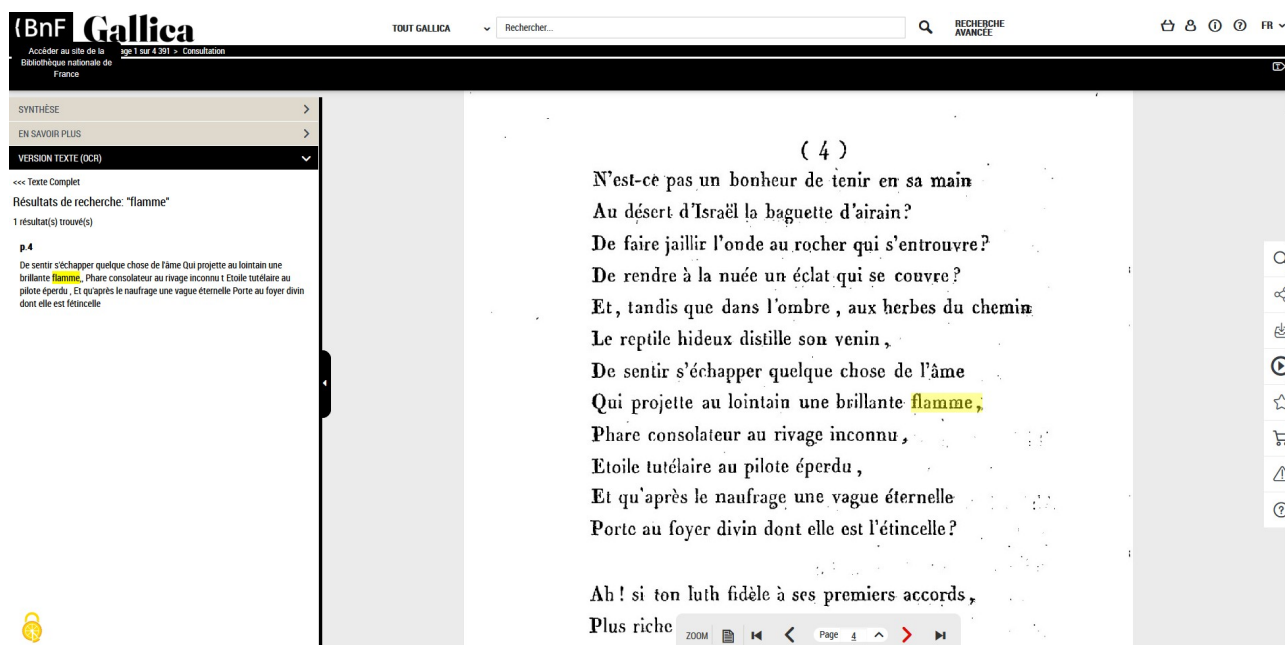


Figure 2: L'OCR pour rechercher en langage naturel dans des ouvrages numérisés (i.e Gallica)

En plus de rendre certaines collections numérisées plus facilement explorables, comme ci-dessus (figure 2), la transcription peut faciliter l'étude des parcours d'œuvres, comme ça a pu être fait sur les catalogues de ventes conservés par le département des arts de l'islam du Louvre, sur des albums de photos des Archives Nationales, ou encore sur des cartes postales de sources diverses.

3 Un « modèle » d'IA est concrètement le programme qu'on utilise pour effectuer une tâche.

Lot

107

A BRONZE STANDING FIGURE OF BUDDHA
CHOLA PERIOD, NAGAPATTINAM, TAMIL NADU, 12TH CENTURY

Numéro de lot

Date

Description

Dimensions

Estimation

Provenance

Contexte

Ouvrage de référence

Oeuvre comparable

standing sculpture on a round carved pedestal. The figure is shown in a full-length view on the left and in a smaller circular inset on the right. The figure is standing on a tiered base, wearing a robe with scalloped folds. The figure's right hand is in the Varada mudra (boon-giving gesture) and the left hand is in the Varada mudra (boon-giving gesture). The figure is shown in a full-length view on the left and in a smaller circular inset on the right.

21.5 cm high

€8,000 - 15,000
US\$3,000 - 20,000
£9,700 - 15,000

Private UK collection, acquired in the early 1980s.

While the artistic achievement of the Chola period is best known for its numerous Hindu figures, a thriving Buddhist community co-existed in Tamil Nadu and artisans produced figures of Buddhist and Jain deities in addition to Hindu gods. The most important Buddhist center in Tamil Nadu was the port city of Nagapattinam.

The Buddhist monastic complex in Nagapattinam, the largest in medieval south India, was built by a Javanese king under Chola patronage.

Buddhism came as early as the 3rd century BC to south India, as recorded by the inscriptions of the Emperor Ashoka. The region continued to be a centre of Buddhist learning in the 12th and 13th Centuries. Buddhist scholars and artisans received generous patronage for the creation of numerous Buddha images.

Buddhist art flourished and was sustained not only by the local community but also by transient communities of merchants and pilgrims from Sri Lanka, Southeast Asia, and China. Portable bronze figures of Buddha, such as the example seen here, would have been easy to carry around and served to propagate artistic and religious connections between India and South East Asia. The pointed flame atop the Buddha's umbilical is a late development in the depiction of the Buddha and is thought to have come to India from Southeast Asia. The robe with scalloped folds at the hem is also a later convention, seen from the 12th Century.

For further discussion see V. Deshpande, *The Serenous and the Sacred: Exhibition Catalogue*, New York, 2000, pp. 207-210.

There is a comparable example in the collections of the Art Institute of Chicago, museum no. 079.2038.

ART INSTITUTE CHICAGO

Figure 3: Transcription et extraction des informations des catalogues numérisés du département des arts de l'islam du Louvre (présenté sur le site de la société Teklia)

Pour ce qui est de la description ou classification d'image, si les LLMs ou VLMs les plus récents (comme les derniers modèles Gemini, Mistral, Qwen...) sont capables aujourd'hui de fournir des descriptions en plein texte plutôt satisfaisantes, il est beaucoup plus dur de leur faire rédiger des sorties structurées, par exemple limitées à un thesaurus, et donc de remplir des tâches traditionnelles d'inventaire et récolement. Et pour le format plein texte, il est aussi beaucoup plus dur de contrôler le niveau de fiabilité des sorties générées. Il faut néanmoins noter que cette affirmation est de moins en moins vraie, et qu'au rythme où vont les choses, des modèles génératifs de ce type seront probablement capables de produire des sorties structurées de qualité d'ici quelques mois ou années. En attendant, les technologies sur lesquelles on a du recul dans un contexte patrimonial pour ces usages sont principalement les techniques de segmentation et de classification par IA. Ces deux tâches ont pour avantage d'être nativement adaptées à un vocabulaire contrôlé, et de généralement proposer des seuils de certitude quand à leurs prédictions.

La segmentation est la tâche de vision par ordinateur qui consiste à repérer et extraire certains types d'objets d'une image. Elle peut par exemple être utilisée pour détecter les types d'objets présents dans les dessins préparatoires d'un designer (projet TORNE-H), les œuvres présentes dans des photographies d'exposition, ou encore extraire pour les mettre en valeurs les photographies contenues dans des carnets numérisés (projet Hikaria). La segmentation peut se faire de manière agnostique (extraction de formes) ou ciblée (extraction d'objets particuliers).

La classification, pour sa part, est la tâche qui consiste à associer à chaque image une classe parmi une liste de classes potentielles. Il est généralement difficile de classer à un niveau très avancé des numérisations de collections, mais pour des objets bien distincts (par exemple des collections archéologiques rassemblant vases, statues, intailles...) ces technologies peuvent être assez efficaces.

Le troisième grand usage, l'exploration de grands corpus, contrairement aux précédents usages, s'inscrit plutôt dans une optique de simplement aider le travail déjà réalisé aujourd'hui. A l'aide de technologies issues de l'apprentissage machine, on peut aujourd'hui transformer les images en *embeddings*, c'est à dire représentations mathématiques complexes. En comparant ensuite ces *embeddings* entre eux, on peut repérer les doublons, les images proches, et même, pour

certaines modèles, les images à partir de morceaux de texte. Un usage très facile de ces techniques dans le cadre d'un travail par exemple d'indexation, est donc de créer des groupes d'images suffisamment similaires (appelés *clusters*), pour ensuite annoter d'un seul coup l'intégralité de ces images. On peut également utiliser ces techniques pour rechercher des images sur la base de leurs caractéristiques visuelles dans des bases d'images sans aucune documentation (projet Torne-H).

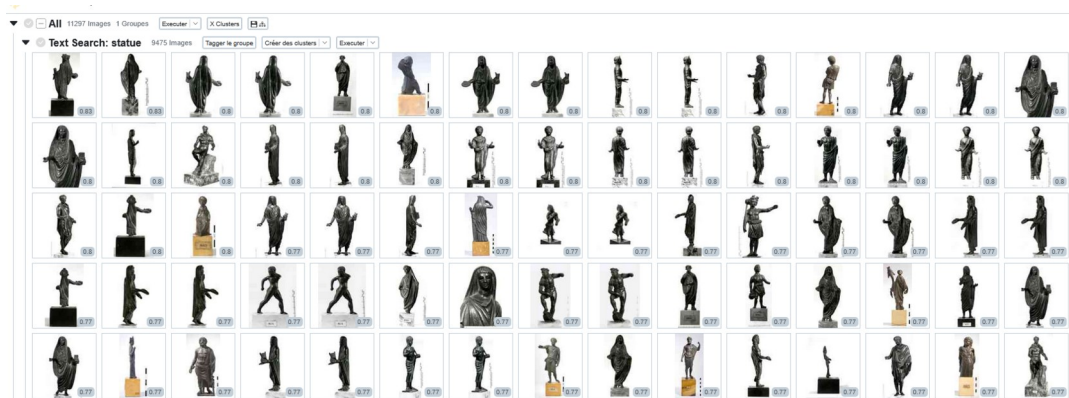


Figure 4: Recherche par texte dans une base d'images non indexées via l'outil Panoptic (CERES - Sorbonne Université)

Enfin, ces techniques peuvent aussi avoir un usage de « médiation », en permettant une navigation différente dans les collections numérisées, par exemple via des visualisations pixplot (BAOIA - affiches de la Contemporaine), ou simplement une proposition d'images similaires (projet Hikaria).

Conclusion

Comme on a pu le voir dans ce court texte, l'IA peut présenter un intérêt pour le traitement de collections patrimoniales, même si la temporalité de la popularisation de ces technologies, et la vitesse de leur évolution, induit un manque de maturité des pratiques à l'heure d'écriture de ces lignes. L'introduction d'IA dans les institutions représente un défi à la fois de mise à niveau technique, de gestion de données, et de réflexion méthodologique sur la manière dont on traite des collections numérisées. L'idée étant de toujours mesurer les besoins et coûts (financiers, sociétaux, environnementaux) réels de ces technologies, auxquelles prête parfois des propriétés presque magiques. Ces questions ont été à l'origine de beaucoup de réflexions, et pour conclure, nous ne pouvons qu'encourager à consulter des travaux comme ceux de la communauté AI4LAM, des documents comme le *toolkit* du réseau « The museums + AI », le *AI planning framework* de la librairie du congrès américaine, ou encore en France les travaux réalisés à la BNF ou dans le cadre du projet TORNE-H, qui ont tous servis à la préparation de cette intervention.