Visual Contagions: extraire et tracer la circulation d'images dans des imprimés illustrés

Robin Champenois, Béatrice Joyeux-Prunel

Visual Contagions: extraire et tracer la circulation d'images dans des imprimés illustrés

Robin Champenois
SACRe, EA7410, ENS, Université PSL, Paris, France
LIGM, Ecole des Ponts, Univ Gustave Eiffel, CNRS, Marne-la-Vallée, France
robin.champenois@ens.psl.eu

Béatrice Joyeux-Prunel
Université de Genève, Suisse
Beatrice.Joyeux-Prunel@unige.ch

Résumé

Le projet Visual Contagions vise à pister la circulation internationale des images, à partir notamment d'un corpus mondial d'imprimés illustrés numérisés. Cet article présente la chaîne de pré-traitement des sources du projet, par laquelle sont récupérés des lots d'images proches visuellement – reproductions exactes, images similaires. C'est à partir des résultats de cette chaîne, croisés avec des métadonnées de dates et lieux de publication, qu'une étude sur le temps long et d'échelle mondiale devient possible. La première partie de l'article détaille quels choix algorithmiques ont été faits pour regrouper des illustrations par similarité ; la seconde partie décrit les outils mis en place pour récupérer des données (au format IIIF), extraire les images et les traiter automatiquement, et enfin permettre un post-traitement humain des résultats du classement algorithmique.

Introduction

Depuis la diffusion des pratiques photographiques, de nombreux historiens de l'art ont tenté de comprendre la circulation des images dans l'espace et le temps en constituant d'importantes banques d'images. Mais pour comprendre ce qu'Aby Warburg appelait les véhicules des images *(Bilder Fahrzeuge)* (Warburg et Recht, 2012), c'est souvent l'étude de cas qui s'est imposée. Aujourd'hui, de nombreux fonds iconographiques ont été numérisés. Leur utilité pour étudier la mondialisation visuelle est limitée : ces corpus donnent un panorama essentiellement européen et artistique de la circulation des images. Mais la numérisation accélérée de périodiques illustrés du monde entier peut donner accès à d'autres sources, plus variées, jamais étudiées à grande échelle, et très souvent datées, localisées. La parution régulière des périodiques illustrés assure un corpus relativement homogène, sur lequel il est possible d'envisager des approches comparées.

Travailler sur la mondialisation à partir des illustrations dans la presse illustrée est l'objectif de Visual Contagions, un projet porté à l'université de Genève par la chaire des humanités numériques et financé par le FNS [1].

Le projet profite de la disponibilité numérique sans précédent, en Open Access, de revues illustrées numérisées en masse, dont les originaux ont été publiés de 1890 (développement des illustrations dans les périodiques) à 1990 (arrivée d'Internet) dans plus de 120 pays. Le corpus réuni à ce jour concerne plus de 4 000 périodiques [2]. Il doit être traité algorithmiquement pour en récupérer les illustrations, avant de repérer quelles images peuvent avoir circulé selon qu'il s'agit de reproductions d'un même original, d'images inspirées les unes des autres, ou d'images relevant de mêmes styles.

Cet article présente le processus de traitement mis en place pour le projet : d'abord, la description algorithmique des images, qui autorise leur comparaison ; ensuite, la chaîne de traitements automatisée et l'interface de contrôle et d'analyse, qui permet aux chercheurs d'exploiter et corriger les prédictions algorithmiques. Dans ce contexte, les choix effectués ont privilégié la robustesse et la fiabilité, ainsi que la facilité de mise en œuvre.

1. https://visualcontagions.unige.ch est financé pour 2021-2024 par le Fonds national suisse pour la Recherche. En 2019-2022, il a reçu le soutien de l'Europe dans le cadre du Centre d'excellence Jean Monnet IMAGO (Ecole normale supérieure, PSL, en partenariat avec l'université de Genève).

2. Voir sa présentation sur le site de l'exposition *Contagions visuelles*, dir. Béatrice Joyeux-Prunel et Nicola Carboni, Espace de création en ligne du Jeu de Paume, mai-décembre 2022, https://jdp.visualcontagions.net/

1 Détection d'images similaires

L'objectif général du projet est d'effectuer, à travers un corpus d'images conséquent, des regroupements d'image par similarité visuelle.

1.1 Recherches préliminaires

Quelle similarité chercher ? De multiples échelles sont possibles : celle du document (reproductions, duplicatas) ; celle de l'objet (reconnaissance d'un élément récurrent) ; celle du « style » ; ou même celle du motif (orientation d'un bras, position d'un corps, type de mise en page, …).

L'ambition initiale du projet était d'avoir une grande finesse : avec l'algorithme ArtMiner (Shen et al., 2022), un premier traitement du corpus d'environ 8 000 images représentant Vénus, constitué par K. Bender (2015) était prometteur pour repérer la circulation de motifs. Mais si cet algorithme était capable d'identifier les similarités avec une granularité élevée, il le faisait en étudiant individuellement chaque paire d'images : l'appliquer sur un ensemble de données de millions d'images aurait demandé plusieurs dizaines d'années de temps de calcul.

Il nous a donc fallu nous restreindre à une vision plus « distante », qui puisse mieux passer à l'échelle de nos ensembles de données. Il fallait toutefois une méthode qui reste plus tolérante à des variations visuelles que ne le serait une détection de duplicatas exacts : d'une part, parce que les procédés d'impression de l'époque (tampons de plus ou moins bonne qualité, clichés photographiques plus ou moins nets…) modifiaient leur apparence d'une version à l'autre ; d'autre part, pour parvenir à détecter des copies plus ou moins fidèles (gravures), des caricatures ou des variantes ; enfin, pour récupérer également des images de style ou de contenu proches.

1.2 Description des images

Le traitement choisi s'est donc limité à extraire, pour chaque image, un vecteur descripteur (*feature vector*), qui soit aisément comparable aux autres par un calcul de similarité cosinus (*cosine distance*). Le facteur décisif pour le regroupement des images est donc la méthode d'extraction des descripteurs, à l'aide de réseaux de neurones.

Prototype Dans un premier temps, ce traitement s'est limité à l'utilisation d'un réseau léger, `ResNet18` (11 millions de paramètres, He et al., 2016), pré-entraîné sur ImageNet, sans apprentissage supplémentaire (*off-the-shelf*). Les descripteurs extraits étaient ceux de la dernière couche convolutionnelle, moyennés linéairement : cela fournissait un vecteur de taille 512 pour chaque image. Si ces descripteurs donnèrent de premiers résultats encourageants, ils montraient un comportement plus *sémantique* que réellement *visuel* : deux images proches visuellement mais au contraste différent pouvaient être considérées comme très « éloignées » ; quand deux visages sans grande ressemblance pouvaient être très « proches ». En cherchant des images connues répandues dans le corpus, nous constations que le réseau ne rapportait qu'un petit nombre de duplicatas.

Améliorations Pour améliorer les résultats, nous avons comparé plusieurs architectures, et finalement opté pour un réseau `ViT` (Dosovitskiy et al., 2020) entraîné selon la méthode DINO (Caron et al., 2021) : les résultats nous convenaient mieux, avec un temps de calcul raisonnable.

Un réseau `ViT` (*Vision Transformer*) a un fonctionnement particulier : il découpe l'image en *tokens* localisés, sortes de « mots visuels », auquel il ajoute un *token* additionnel (dit « de classe »). Ces *tokens* sont ensuite transformés en utilisant des mécanismes dits « d'attention », plutôt que convolutionnels (CNN). Les Transformers affichent de nos jours des performances comparables à celles des CNNs, et leur usage est de plus en plus commun.

L'élément important est surtout l'utilisation de DINO (Caron et al., 2021), une méthode d'entraînement non supervisé qui se distingue de l'objectif de classification habituellement utilisé. Le principe de DINO est d'entraîner conjointement selon des modalités légèrement différentes un réseau « étudiant » et un réseau « enseignant », en cherchant à ce que les deux donnent des descriptions similaires d'une même image, altérée de diverses manières (recadrage, déformations, changement de couleurs…). Cette méthode rend le réseau particulièrement robuste aux perturbations typiques du corpus de Visual Contagions (perturbations de cadre, de contraste…). L'apprentissage non supervisé permet par ailleurs d'éviter que le réseau ne se spécialise trop sur la recherche de classes ou de formes spécifiques au détriment d'autres types d'images.

Post-traitement La détection d'images similaires est une des tâches étudiées par les auteurs

a.  b. 

FIGURE 1 – Deux images de test issues de WikiCommons (*Sir Willam Butts the Younger*, auteur inconnu ; *Ships Sailing and Beating up against the Wind in the Sound*, par Adolph Friedrich Vollmer). Ces images ne sont pas présentes dans l'ensemble de calcul de la PCA.

de DINO [3]. Dans leur évaluation, les descripteurs extraits sont issus de l'avant-dernière couche du Transformer : le descripteur issu du *token* de classe est gardé tel quel ; et on lui concatène la moyenne spatiale généralisée (puissance 4) des autres descripteurs – reprenant ainsi la procédure de (Radenović et al., 2019). Enfin, les auteurs appliquent une analyse en composantes principales (PCA) avec un facteur de blanchiment (*whitening*) de 0,5, qui d'après (Berman et al., 2019) améliore les résultats.

Leur évaluation utilise un réseau `ViT-B/8`, de 85 millions de paramètres, et dont la dimension des descripteurs est de 1 536 ; la PCA est calculée sur 20 000 images issues du dataset YFCC100M (Thomee et al., 2016). Pour le projet Visual Contagions, nous avons utilisé un réseau un peu moins précis mais plus léger (`ViT-S/16`, 4 fois plus léger et 20 fois plus rapide), aux descripteurs de taille 768, tel que fourni par les auteurs de DINO. La PCA a été calculée sur des images d'un domaine plus proche de celles utilisées par le projet (essentiellement des numérisations d'imprimés) : 20 000 images extraites de nos corpus ; 5 000 peintures issues de WikiData ; et 12 000 photographies venant de différents ensembles de données [4] (pour conserver une certaine diversité de contenus, et ne pas se spécialiser dans les imprimés). Dans une recherche d'optimisation du stockage, trois tailles de descripteurs ont été testées : 768, 384 et 256. La taille des images d'entrée a par ailleurs été fixée à 320×320 pixels.

Évaluation En l'absence de données annotées à utiliser comme référence, l'évaluation des méthodes n'a pas été faite de manière quantitative.

3. https://github.com/facebookresearch/dino/blob/main/eval_copy_detection.py

4. Notamment COCO (Lin et al., 2014), LAION400M (Schuhmann et al., 2021) et VOC2012 (Everingham et al., 2012)

Méthode	Plus proches voisins			
ResNet18	0.86 	0.86 	0.86 	0.86 
ResNet18 +PCA	0.32 	0.29 	0.29 	0.28 
DINO+PCA (768)	0.44 	0.23 	0.20 	0.19 
DINO+PCA (384)	0.55 	0.30 	0.29 	0.29 
DINO+PCA (256)	0.58 	0.38 	0.36 	0.35 

TABLEAU 1 – Comparaison de 5 recherches de plus proches voisins de l'image de référence (Figure 1a), selon différentes méthodes (voir section 1.2).
Au-dessus de chaque image : affichage du score de similarité rapporté par la chaîne de traitement.
Ici, les descripteurs ResNet ne parviennent pas à identifier la copie approchée présente dans l'ensemble de test, sans PCA ; les méthodes DINO+PCA trouvent systématiquement cette copie (avec un score autour de 0,5), et trouvent des images relativement semblables.

Une étude qualitative a permis de discriminer très clairement l'intérêt de certaines méthodes plutôt que d'autres. Trois approches ont été comparées, en appliquant la même PCA (avec blanchiment) aux descripteurs issus de chaque méthode :

— ResNet18 (méthode initale, taille 512)
— ResNet18+PCA (taille 512)
— DINO+PCA : descripteurs complets (taille 768), ou partiels (taille 384)

Les résultats ont été reportés dans des tableaux comparables aux Tableaux 1 et 2 qui ont été soumis à l'équipe, comparant les images rapportées par la recherche d'une image de départ (Figure 1) dans l'ensemble de données constitué pour le test (250 000 images extraites de notre corpus).

Interprétation L'application successive de la PCA et d'une étape de blanchiment augmente la qualité des résultats : avec elle, dans la plupart des cas, quand elle existe, la "bonne image" (c'est-à-dire, un duplicata effectif de l'image recherchée) est la plus proche trouvée. Les scores de similarité des images ainsi regroupées sont plus discriminants

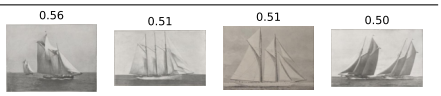
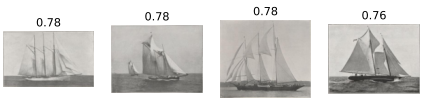
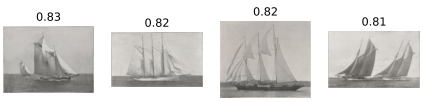
Méthode	Plus proches voisins			
ResNet18				
ResNet18 +PCA				
DINO+PCA (768)				
DINO+PCA (384)				
DINO+PCA (256)				

TABLEAU 2 – Voir Tableau 1 pour la légende.
Sur une image requête de bateau (voir Figure 1b), en l'absence de copie effectivement présente dans l'ensemble de recherche, les résultats font globalement apparaître d'autres bateaux, avec des scores potentiellement très élevés (au-dessus de $0{,}55$), d'autant plus élevés que les descripteurs sont petits.

avec la PCA : alors que pour les traitements sans PCA un score de $0{,}85$ peut être obtenu par des images quasiment identiques autant que par deux images très différentes, le passage par la PCA réserve les scores supérieurs à $0{,}5$ aux images très proches (le problème demeurant, cependant, pour les images très nombreuses de score entre $0{,}4$ et $0{,}5$).

Comme attendu, les descripteurs obtenus avec la méthode DINO+PCA sont plus robustes aux changements de contraste et de couleur – un élément décisif pour l'étude des publicités depuis le XIX^{e} siècle –, et la chaîne de traitement n'oblige pas à passer les images en noir et blanc [5]. Les descripteurs de dimensions 384 (et même 256) sont moins précis que ceux de dimension 768, mais leurs regroupements restent pertinents.

La question du « seuil » le plus efficace de similarité reste délicate. Dans le cas d'une vectorisation à 384 dimensions, une image et sa version en noir et blanc ont des scores de similarité entre $0{,}5$ et $0{,}7$. Mais, par exemple, deux peintures de bateaux similaires sans être identiques peuvent avoir un score de $0{,}8$, comme le montre le Tableau 2. Des scores de $0{,}75$ sont obtenus pour des images très proches. Au-delà de $0{,}85$, nous récupérons uniquement des copies exactes.

Nous avons finalement opté pour les descripteurs DINO+PCA, sur des vecteurs de 384 dimensions. Ce choix permet de récupérer et grouper davantage d'images similaires (meilleur rappel, ou *recall*), et les scores élevés sont réservés aux images vraiment similaires (plus de précision).

1.3 Regroupement en *clusters*

Une fois que nous avions une mesure de similarité entre deux images, notre objectif était de générer des groupes (*clusters*) d'images, utiles pour l'analyse ultérieure. De nombreux algorithmes de *clustering* existent, mais beaucoup sont trop gourmands en ressources pour fonctionner sur nos données. Ils peuvent être, par ailleurs, difficiles à interpréter.

Nous avons donc choisi un simple regroupement par composantes connexes dans le graphe formé par les images (comme sommets), reliées entre elles deux à deux dès que leur similarité est supérieure à un certain seuil (choisi par l'utilisateur). Autrement dit, deux images sont regroupées ensemble si et seulement s'il existe entre elles un chemin d'images suffisamment semblables deux-à-deux.

Cet algorithme a l'avantage d'être rapide ; il a l'inconvénient d'être très sensible à des similarités croisées, ou des mauvais recadrages. Par exemple, une illustration A et une publicité contenant beaucoup de texte peuvent être agglomérés dans le même cluster s'il existe dans le corpus une image contenant à la fois A et du texte. Par « contagion », l'algorithme de composantes connexes mettra ainsi toutes les occurrences de A et de texte dans le même (gros) groupe : c'est ainsi que le clustering aboutit souvent à un *cluster* contenant jusqu'à 20 % du corpus initial – ininterprétable. Il est possible toutefois de procéder itérativement à d'autres *clusterings* à l'intérieur même de ce groupe, avec des seuils de similarité plus élevés, pour contrebalancer cet effet (en sacrifiant un peu de rappel pour cette sous-partie du corpus).

2 Une plateforme pour effectuer les traitements de bout en bout : Explore

L'enjeu du projet n'est pas seulement de disposer d'algorithmes pour comparer les images, mais bien de permettre le traitement automatisé d'importants corpus par des chercheurs d'horizons divers. C'est

5. Les tests indiquent, de fait, une meilleure similarité entre une image en couleur passée en noir et blanc et une numérisation en noir et blanc ; mais l'amélioration est moins significative que pour ResNet.

ce que réalise la plateforme Explore [6].

2.1 Récupération des données IIIF

Pour mettre en place la chaîne de traitement des périodiques, le format interopérable d'images IIIF a été privilégié [7]. Depuis son introduction en 2012, ce standard permet la distribution, la description et la réutilisation simple d'images, accessible aux personnes comme aux machines. Il permet de réunir des corpus de sources très diverses, sans devoir les restocker ; quel que soit le serveur d'où est publié le document en question, celui-ci peut être visualisé et comparé avec un document servi par n'importe quelle autre institution publiant des numérisations IIIF ; ce standard permet aussi, à partir d'une simple URI, de récupérer toutes les pages d'un volume particulier dans le bon ordre, et les métadonnées qui le décrivent.

Dans la pratique, l'accès aux données relatives à un volume se fait d'abord par l'API de « présentation », qui consiste principalement en des fichiers au format JSON-LD, accessibles par une URI directe (utilisée comme identifiant unique du document). Ce fichier, appelé « manifeste », répertorie les métadonnées de publication du volume, et liste chaque page, dans l'ordre. Ces pages contiennent une référence à d'autres URIs correspondant à l'API « Image » IIIF, qui donnent accès aux numérisation des documents.

Un corpus Visual Contagions consiste donc en une liste d'URI de manifestes IIIF. À partir de ceux-ci, deux traitements parallèles sont menés : un premier télécharge toutes les pages, pour les transmettre à la segmentation (voir section 2.2) ; un second extrait les métadonnées disponibles dans le fichier JSON, et les normalise – souvent à l'aide d'une intervention humaine, les formats n'étant pas toujours unifiés.

L'API « Image » permet d'accéder directement, via de simples transformations d'URL, à des extraits de l'image d'origine, à la taille souhaitée : cette flexibilité nous permet de limiter la bande passante et le stockage que nous utilisons, en ne demandant aux serveurs IIIF que les images à la dimension dont nous avons besoin.

2.2 Traitement des données

Les données récoltées par IIIF sont souvent, dans nos corpus, des imprimés, contenant de multiples images par page, ainsi que du texte. Une première tâche consiste donc, si cela est nécessaire, à appliquer un algorithme de segmentation pour séparer les illustrations du texte. L'outil que nous utilisons est docExtractor (Monnier et Aubry, 2020), qui repose sur un réseau de neurones profond de type UNet (Ronneberger et al., 2015).

Les illustrations récupérées sont ensuite traitées selon les algorithmes décrit en section 1, afin d'en obtenir des vecteurs descripteurs, et de les regrouper. Afin d'accélérer nettement la recherche des plus proches voisins, sans devoir comparer chaque image avec des millions d'autres, une indexation reposant sur l'algorithme HNSW (Malkov et Yashunin, 2020) est effectuée.

2.3 Interface web et analyse ultérieure

La constitution des corpus IIIF, ainsi que le contrôle de la chaîne de traitement, est entièrement accessible à travers la plateforme Explore, développée à l'aide des bibliothèques `django` et `celery`.

La chaîne de traitement permet en fait de regrouper les images selon trois types de similarité très différentes pour un œil humain, sans vraiment faire la distinction :

— des images reproduisant un même original ;
— des images contenant un même objet décliné sous des angles et des formats divers ;
— des images de même style (essentiellement artistiques, certains styles étant mieux regroupés que les autres – en particulier l'abstraction géométrique et le cubisme).

Parmi ces regroupements, certains ne sont pas pertinents – par exemple les codes-barres apposés sur les volumes papier qui sont ensuite numérisés. D'où le besoin d'un outil pour retrier les clusters sortis de la chaîne de regroupement, avec un œil expert.

La plateforme Explore a donc été développée pour permettre :

1. De visualiser les groupements d'images, triés selon certains critères (le plus d'images ; ou concernant le plus de villes / de pays différents) ;
2. D'effectuer à partir de n'importe quelle image (locale ou avec URI) des recherches d'images similaires dans le corpus ;
3. D'observer chaque cluster d'images individuellement, le sélectionner, le nommer, le compléter ;

6. Accessible publiquement sur https://visualcontagions.unige.ch/explore/
7. https://iiif.io/

4. D'observer chaque image d'un cluster ; de connaître sa date, son lieu, son titre et son type de publication ; de la visualiser en contexte ;
5. De visualiser de manière spatio-temporelle un ou plusieurs clusters, et de comparer leur répartition dans l'espace et le temps ;
6. De récupérer les métadonnées complètes d'une série d'images proches au format CSV.

C'est à partir de ce nouveau corpus, proposé par le traitement algorithmique du premier corpus de revues, puis validé par les experts, que l'équipe du projet Visual Contagions peut mettre en route une analyse multi-scalaire de la circulation des images, avant d'approfondir les questions ouvertes par cette analyse avec d'autres méthodes et d'autres sources. 1 787 revues ont pour l'instant été analysées, dont ont été extraites à peu près 6,8 millions d'illustrations.

Bibliographie

K. Bender. 2015. Distant Viewing in Art History. A Case Study of Artistic Productivity. *International Journal for Digital Art History*, 1(1).

Maxim Berman, Hervé Jégou, Andrea Vedaldi, Iasonas Kokkinos, et Matthijs Douze. 2019. MultiGrain: A unified image embedding for classes and instances.

Mathilde Caron, Hugo Touvron, Ishan Misra, Hervé Jegou, Julien Mairal, Piotr Bojanowski, et Armand Joulin. 2021. Emerging Properties in Self-Supervised Vision Transformers. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 9630–9640.

Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, et Neil Houlsby. 2020. An Image is Worth 16x16 Words : Transformers for Image Recognition at Scale. In *International Conference on Learning Representations*.

M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, et A. Zisserman. 2012. The PASCAL Visual Object Classes Challenge 2012 (VOC2012) Results.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, et Jian Sun. 2016. Deep Residual Learning for Image Recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778.

Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, et C. Lawrence Zitnick. 2014. Microsoft COCO: Common Objects in Context. In *Computer Vision – ECCV 2014*, Lecture Notes in Computer Science, pages 740–755, Cham. Springer International Publishing.

Yu A. Malkov et D. A. Yashunin. 2020. Efficient and Robust Approximate Nearest Neighbor Search Using Hierarchical Navigable Small World Graphs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(4) :824–836.

Tom Monnier et Mathieu Aubry. 2020. docExtractor: An off-the-shelf historical document element extraction. In *2020 17th International Conference on Frontiers in Handwriting Recognition (ICFHR)*, pages 91–96.

Filip Radenović, Giorgos Tolias, et Ondřej Chum. 2019. Fine-Tuning CNN Image Retrieval with No Human Annotation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(7) :1655–1668.

Olaf Ronneberger, Philipp Fischer, et Thomas Brox. 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, Lecture Notes in Computer Science, pages 234–241, Cham. Springer International Publishing.

Christoph Schuhmann, Richard Vencu, Romain Beaumont, Robert Kaczmarczyk, Clayton Mullis, Aarush Katta, Theo Coombes, Jenia Jitsev, et Aran Komatsuzaki. 2021. LAION-400M: Open Dataset of CLIP-Filtered 400 Million Image-Text Pairs.

Xi Shen, Robin Champenois, Shiry Ginosar, Ilaria Pastrolin, Morgane Rousselot, Oumayma Bounou, Tom Monnier, Spyros Gidaris, François Bougard, Pierre-Guillaume Raverdy, Marie-Françoise Limon, Christine Bénévent, Marc Smith, Olivier Poncet, K. Bender, Béatrice Joyeux-Prunel, Elizabeth Honig, Alexei A. Efros, et Mathieu Aubry. 2022. Spatially-Consistent Feature Matching and Learning for Heritage Image Analysis. *International Journal of Computer Vision*, 130(5) :1325–1339.

Bart Thomee, David A. Shamma, Gerald Friedland, Benjamin Elizalde, Karl Ni, Douglas Poland, Damian Borth, et Li-Jia Li. 2016. YFCC100M: The new data in multimedia research. *Communications of the ACM*, 59(2) :64–73.

Aby Moritz Warburg et Roland Recht. 2012. *L'atlas Mnémosyne*. Numéro II in Écrits. l'Écarquillé Institut national d'histoire de l'art, INHA, Paris.