



PapierS

NOUVELLES DE L'ASSOCIATION FRANÇAISE POUR L'HISTOIRE ET L'ÉTUDE DU PAPIER & DES PAPETERIES. 33 AVENUE FAIDHERBE 93100 MONTREUIL
Numéro 18 Janvier 2024

Une fois de plus nous vous invitons à explorer des aspects parfois inattendus du papier :

Protecteur – Dans le monde arabe médiéval, face aux souffrances et lésions corporelles provoquées par le froid de l'hiver, certains médecins ont mis en avant les propriétés d'isolation thermique du papier. Ici, l'exploration des textes se double d'un volet terminologique susceptible de jeter un éclairage nouveau sur l'arrivée des techniques papetières dans le monde islamique.

Insolite – Nous découvrirons un acte notarié du XVIII^e siècle, dans lequel des papetiers expliquent la réglementation en vigueur relative aux marques du papier à un négociant de Cadix, incontournable port de commerce avec l'Amérique espagnole.

Clandestin – Le sulfureux manuscrit rédigé en secret par le marquis de Sade à la prison de la Bastille nous livrera quelques-uns des secrets de son étonnant support, un long rouleau de feuillets assemblés. Le décryptage, mené par l'équipe de restauration de la Bibliothèque de l'Arsenal, résulte d'une méticuleuse analyse matérielle et d'un travail de traitement d'image à l'aide de « Magic Eraser ».

Révélé – En complément à cette application pratique, Robert Erdmann nous explique la conception et les principes de l'outil d'intelligence artificielle très innovant qu'il a développé afin d'améliorer la lisibilité des filigranes. « Magic Eraser » permet aux chercheurs d'effacer virtuellement l'encre des manuscrits quand celle-ci s'interpose de façon trop prégnante et gêne la lecture des marques du papier.

Militant – Des ouvrières, mobilisées pour la préservation de leur emploi, ont collecté, consigné et conservé de nombreux documents retracant leurs activités au jour le jour, au sein de trois cahiers de bord qui livrent aujourd'hui un témoignage sensible et rare de cette lutte. Le décryptage, très détaillé des aspects matériels de ces cahiers, rend un bel hommage à ces militantes.

Tactile – Guidés par sa directrice, nous visiterons les coulisses d'une imprimerie braille et nous découvrirons les critères spécifiques de choix du papier et les techniques de production des livres en relief.

Bonne découverte !

MARYSE PIERRARD



SOMMAIRE

ÉDITORIAL

Maryse PIERRARD, Responsable de la publication

L'UTILISATION DU PAPIER DANS LA MEDECINE DES VOYAGES ARABES MEDIEVAUX : ENQUETE HISTORIQUE ET TERMINOLOGIQUE PRELIMINAIRE

Zsuzsanna CSORBA, Chercheuse,
Institut Avicenne d'études du Moyen-Orient
(Piliscsaba, Hongrie)

A PROPOS DES MARQUES DE PAPIERS

BEARNAIS NEGOCIES A CADIX (1781)

Jeanne VALOIS, Historienne

LES FEUILLETS DU MANUSCRIT DES « 120 JOURNEES DE SODOME » (1785) : PREMIERES OBSERVATIONS AVEC « MAGIC ERASER »

Marlène SMILAUER, Restauratrice,
Arts Graphiques-Livres, Bibliothèque de l'Arsenal,
Bibliothèque nationale de France

L'OUTIL « MAGIC ERASER » : QUAND L'INTELLIGENCE ARTIFICIELLE EFFACE L'ENCRE POUR REVELER LE FILIGRANE

Robert G. ERDMANN, Senior Scientist,
Rijksmuseum (Amsterdam) et Professeur titulaire,
Université d'Amsterdam

HISTOIRE D'OBSSESSION, CAHIERS DE GREVE

Asceline POURCELOT, Conservatrice-restauratrice,
Arts Graphiques-Livres

QUEL PAPIER POUR LA LECTURE TACTILE ?

Adeline COURSANT, Directrice du Centre de Transcription et d'Édition en Braille (Cteb)

LA PAPETERIE DE MANDEURE

Maryse PIERRARD

COMPTE RENDUS

MARLÈNE SMILAUER

LES FEUILLETS DU MANUSCRIT DES 120 JOURNÉES DE SODOME : PREMIÈRES OBSERVATIONS AVEC « MAGIC ERASER »

C'est à la prison de la Bastille, à l'automne 1785, que le marquis de Sade (1740-1814) recopie sur un rouleau de papier les brouillons des *120 journées de Sodome ou l'école du libertinage*. L'histoire rocambolesque de ce manuscrit, caché, volé, perdu, édité partiellement, censuré, exporté illégalement, devenu objet de spéculation, recèle encore des mystères¹. Déclaré Trésor national, le manuscrit a rejoint en juillet 2021, grâce à un mécénat, les collections publiques de la BnF, à la bibliothèque de l'Arsenal, parmi les archives de la Bastille où est conservé l'important dossier du prisonnier Sade². Les conservateurs et restaurateurs de la bibliothèque ont alors une occasion unique d'étudier le rouleau dans sa matérialité. Le manuscrit étant fort dégradé sur les feuillets des extrémités et sur ses bords, sa manipulation s'avère périlleuse.

Claire Lesage, conservatrice du fonds des manuscrits modernes, convient avec l'équipe de l'atelier d'intervenir rapidement sur le rouleau. Quatre restauratrices, Caroline Bertrand, Anne-Laure Fessart, Magali Dufour et Marlène Smilauer, commencent l'étude, coordonnent les actions avec différents services pour amorcer une étude matérielle poussée et mener à bien le projet de conservation et d'exposition. Elles réalisent au printemps 2022 la restauration qui permet une numérisation complète du manuscrit³.

Le rouleau est constitué d'une longue bande de feuillets collés bout à bout, écrits recto verso entre deux traits d'encadrement, à l'encre métallo-gallique. L'écriture du marquis, malgré son échelle microscopique, est clairement identifiable. Avant l'intervention de restauration, un constat d'état photographique intégral sur banc de reproduction en lumière réfléchie est réalisé avec un reflex numérique. En parallèle, chacun des feuillets est photographié sur table lumineuse avec un Smartphone afin de pouvoir effectuer dans un second temps la recherche systématique des marques papetières avec l'outil de traitement d'image numérique en phase de test « Magic Eraser⁴ ». Conjointement à ces prises de vue en lumière transmise, un relevé sur un polyester transparent, appelé relevé Mylar, va permettre d'avoir en échelle 1/1 un ensemble d'informations matérielles. Ce « patron » relève : les contours de chaque feuillet numéroté, les chevauchements, les dégradations et les taches. Ce deuxième substitut, épargnant la manipulation de l'original, permet d'effectuer des mesures précises.

La longueur du rouleau est vérifiée : 11,85 m. Les 33 feuillets de papier vergé, coupés en bandes dans le sens de la hauteur (parallèlement aux fils de chaînettes), ont été



Figure 1

Vue du rouleau lors de la restauration du feuillet 9. Passage des feuillets blancs aux feuillets bleus.

¹ En 2023, un article développant les péripéties du manuscrit, le constat d'état, l'étude technico-scientifique et la restauration est paru dans *Support tracé*, ainsi qu'un fascicule sur le rouleau des *Cent vingt journées de Sodome* écrit par Olivier Bosc et Claire LESAGE dans la collection « Cartel » de la BnF.

² Le dossier Sade correspond aux cotes MS-12455 et MS-12456, le rouleau est coté à son arrivée MS-15877.

³ La numérisation est disponible sur Gallica : <https://gallica.bnf.fr/ark:/12148/btv1b525185706/f1.planchecontact>

⁴ Nous tenons à remercier tout particulièrement Nadine Dumain pour sa communication sur l'outil, les essais très concluants qu'elle a réalisés pour nous, ainsi que la mise en relation avec Robert Erdmann qui nous a généreusement transmis le lien d'utilisation de « Magic Eraser » et avec qui nous avons communiqué à plusieurs reprises.

LES FEUILLETS DU MANUSCRIT DES 120 JOURNÉES DE SODOME : PREMIÈRES OBSERVATIONS AVEC « MAGIC ERASER »

MARLÈNE SMILAUER

assemblés par une jonction de 2 à 8 mm de collage. Leur largeur varie de 11,4 à 12 cm et leur hauteur de 36 à 37 cm. Deux types de papiers sont identifiables par leur couleur : les 8 premiers feuillets sont blancs, les 25 autres bleus (**figure 1**). À l'occasion du constat, nous observons en lumière transmise un filigrane, une demi-fleur de type marguerite décelée sur les marges non manuscrites de quatre feuillets qui se succèdent (du 21 au 24), elles sont relevées sur un Mylar où le motif coupé en son milieu peut être reconstitué sans perte. Quelques rares suspicions d'autres marques sont repérées, mais rien de clairement identifiable.

L'UTILISATION DE « MAGIC ERASER »

L'exploitation des 33 clichés en lumière transmise avec « Magic Eraser » nous réserve de nombreuses surprises et des moments d'enthousiasme collectif mémorables. En effet, sur ce manuscrit noirci par une écriture très serrée, nous n'imaginions pas trouver 24 marques sur les 33 feuillets. Cette découverte a nécessité un long et méthodique travail de traitement d'image.

Tout d'abord, l'image de base du feuillet en lumière transmise est traitée par « Magic Eraser » pour obtenir une seconde image (**figure 2**). Si une marque est perceptible à ce stade, une autre série de traitement d'image est effectuée à partir de la première image : rognage pour enlever le débord de table lumineuse qui génère du flou lors du passage dans l'outil, zoom sur la zone à observer, inversion miroir ou retournement à 180 degrés pour faciliter la lecture du motif et parfois inversion des couleurs. Pour l'ensemble, lorsque des prises de vues en détail n'ont pas été faites, car la marque n'était pas repérable à l'œil nu, nous refaisons une série de secondes prises de vue plus ciblées pour aller chercher une meilleure définition et repasser l'image dans l'outil de traitement. Un classement méticuleux est indispensable au fur et à mesure : pour chaque feuillet s'accumulent parfois jusqu'à quinze images ...

Une constante est observée : les motifs apparaissent en bord de feuille, au milieu de la hauteur, souvent coupés en leur centre et peuvent être rassemblés sans manques ou avec une légère perte d'un feuillet à l'autre. L'étude de possibles reconstitutions de feuillets complets est à mener, certainement avec une étude plus précise des vergeures et fils de chaînettes, de leurs écartements, etc.

Nous recherchons les motifs identifiés et des lettres repérées dans les catalogues de références classiques : Nicolaï⁵ et Gaudriault⁶ (**figures 3 et 4**). Nous nommons



Figure 2

Le feuillet 2 en lumière transmise et après passage par « Magic Eraser » : les lettres KOWEN sont lisibles en inversé, au centre du bord gauche.

⁵ Alexandre NICOLAÏ, *Histoire des moulins à papier du Sud-Ouest de la France (1300-1800)*: Périgord, Agenais, Angoumois, Soule, Béarn, Monein, Ed. Pyremonde, 2006-2009, p. 19, 31-35, Pl. XXXIV, XLIV-XLVI.

⁶ Raymond GAUDRIAUT, *Filigranes et autres caractéristiques des papiers fabriqués en France aux XVII^e et XVIII^e siècles*, Paris, Éd. du CNRS, J. Telford, 1995, p.102, 111, Pl. 51-54.

PAPIER ÉCRU	
F1 :	HOVEN
F2 :	KOUWEN
F3 :	Cornet
F4 :	Cornet
F5 :	Cornet
F6 :	KOUWEN
F7 :	Cornet
F8 :	HOVEN
PAPIER AZURÉ	
F9 :	Chapeau de Cardinal
F10 :	L
F11 :	Chapeau de Cardinal
F12 :	rien
F13 :	rien
F14 :	L
F15 :	rien
F16 :	L
F17 :	rien
F18 :	L
F19 :	L
F20 :	rien
F21 :	Chapeau de Cardinal
F22 :	Chapeau de Cardinal
F23 :	Chapeau de Cardinal
F24 :	Chapeau de Cardinal
F25 :	rien
F26 :	L
F27 :	rien
F28 :	L
F29 :	L
F30 :	rien
F31 :	L
F32 :	rien
F33 :	Chapeau de Cardinal

Figure 5

Liste des marques et contre-marques relevées par feuillets
(F1 est le feuillet numéro 1)

ainsi ces figures et résumons nos découvertes sur le tableau qui suit (**figure 5**). Le nom de KOUWENHOVEN n'est pas référencé dans ces catalogues, c'est un nom hollandais, la piste reste à suivre⁷.

Suite aux découvertes de filigranes, nous avons commencé à observer les deux volumes de la correspondance du marquis de Sade et de ses proches⁸. Si le motif du cornet est courant dans les correspondances du XVIII^e siècle, nous n'en avons trouvé aucun associé à la contremarque KOUWENHOVEN. Mais le chapeau de cardinal avec pot se retrouve souvent, associé à la contremarque « L ». Pour le moment, aucun KOUWENHOVEN n'a été observé. Les questions des lieux de production et de la manière de s'approvisionner en papier restent donc posées. La poursuite de ces recherches permettrait peut-être de trouver le format précis des feuillets *in plano* d'origine. Cela nous aiderait à comprendre la manière dont ils ont été découpés et assemblés, et aussi peut-être à reconstituer le nombre de feuillets *in plano*, partiels ou entiers utilisés pour réaliser le manuscrit en rouleau des *Cent vingt journées de Sodome*.

CONCLUSION

La restauration menée par l'équipe de la bibliothèque de l'Arsenal s'est voulue juste, décelable et humble. L'étude matérielle et technico-scientifique liée à la restauration a ouvert sur une transmission enrichie et éclairée. La présentation de cet objet patrimonial unique a fait l'objet de différentes manifestations au cours de l'année 2023. Une conférence a été consacrée à la restauration et à l'étude matérielle du rouleau dans le cadre du cycle « Éclats de papier » à la bibliothèque de l'Arsenal⁹, suivie d'une exposition au sein du nouveau musée de la BnF, site Richelieu, à partir du 16 septembre 2023. C'est une étape fondamentale, car cet objet est désormais rendu au regard du plus grand nombre, en tant que partie intégrante du patrimoine culturel public. Enfin, un colloque a rassemblé les meilleurs spécialistes du rouleau à la BnF le 22 septembre 2023. Ces événements clés ont permis d'ouvrir un temps nouveau, faisant la part belle à la connaissance, sous toutes ses formes.

⁷ Claire Bustarret nous signale que la base MUSE comporte plusieurs vergés portant le motif du chapeau avec pot étoilé associé à la contremarque « L » de la papeterie de Langlée, notamment ceux employés par Condorcet et conservés aux archives de l'Académie des sciences : AAS028, azuré, dimensions 372 x 488 mm (rogné) en 1774 et BNF568, écru, 372 x 498 mm (rogné) en 1776.

⁸ Il s'agit des lettres manuscrites d'origine, conservées à la Bibliothèque de l'Arsenal, Archives de la Bastille, sous les cotes MS-12455 et MS-12456.

⁹ Conférence du 13 février 2023 [[En ligne](https://www.bnf.fr/fr/mediatheque/papier-en-restauration-le-rouleau-des-120-journees-de-sodome)] <https://www.bnf.fr/fr/mediatheque/papier-en-restauration-le-rouleau-des-120-journees-de-sodome>=

LES FEUILLETS DU MANUSCRIT DES 120 JOURNÉES DE SODOME : PREMIÈRES OBSERVATIONS AVEC « MAGIC ERASER »

MARLÈNE SMILAUER

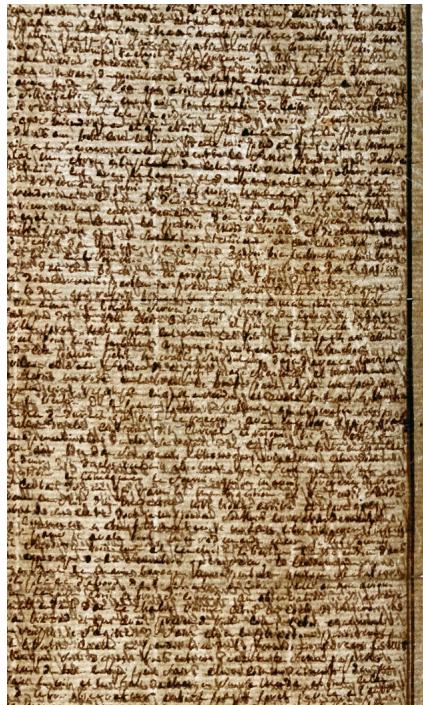


Figure 3

Le feuillet 5 en lumière transmise, puis après passage par « Magic Eraser ». Un motif coupé aux 2/3 est visible au centre du bord droit : un écu de couronné avec un cornet et une lettre R, appelé « Cornet » dans les répertoires de filigranes (A. NICOLAÏ ; R. GAUDRIAUT)

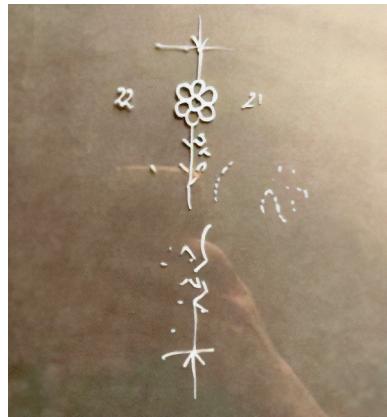


Figure 4a

Relevé Mylar sur le bord des feuillets 22 et 21.

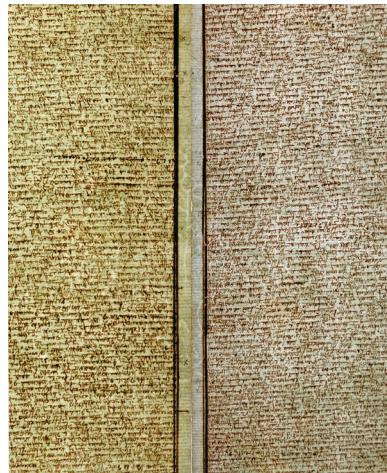


Figure 4b

Photographies en lumière transmise et mises côte à côte des feuillets 22 et 21

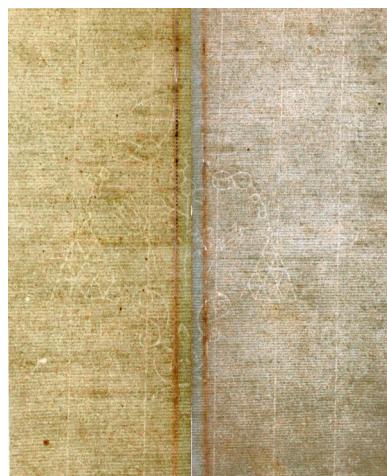


Figure 4c

Les mêmes photographies exploitées par « Magic Eraser »



Figure 4d

Chapeau de cardinal avec pot étoilé, A. NICOLAÏ, vol. 2, planche XXXIV

Un problème persistant dans l'identification des filigranes réside dans le fait que le papier est un support de texte, d'impression ou de dessin qui rendent le filigrane difficile à lire. L'élimination automatique de l'encre d'une image en lumière transmise d'un papier filigrané est un défi, car elle nécessite une connaissance de la structure des filigranes, du papier (avec ses chaînettes et vergeures), et de l'effet de l'encre de différentes couleurs et de différents degrés d'opacité sur les deux côtés du papier. Depuis 2018, nous nous sommes lancés dans la création d'un outil de reconnaissance automatique des filigranes dans le cadre de notre développement d'une boîte à outils logicielle pour le papier^{1,2}. Vu la difficulté de distinguer le filigrane sous l'encre, et inspirés par les progrès de l'intelligence artificielle, nous avons développé en octobre 2019 un réseau neuronal convolutif³ ainsi qu'un prototype d'application web, où l'utilisateur peut supprimer automatiquement l'encre des images. Via l'interface, il est possible de charger une photo en lumière transmise d'un papier encré contenant un filigrane, et l'application produit une version « désencree ». L'outil, baptisé « *Magic Eraser* » (« Effaceur Magique »), a été mis à la disposition de quelques chercheurs qui ont généreusement accepté de donner leur avis sur son utilité. L'outil a depuis été utilisé dans divers projets, notamment l'analyse d'un manuscrit de Sade (Paris, Bibliothèque de l'Arsenal ; voir l'article de Marlène Smilauer qui précède celui-ci). Nous allons décrire le développement et le fonctionnement de l'outil. Les fondations mathématiques du réseau neuronal dépassent le cadre de cet article, mais nous présentons les détails pertinents de l'outil, y compris un aperçu de l'apprentissage des réseaux neuronaux, la manière dont nous avons constitué nos données d'entraînement pour apprendre au réseau à supprimer l'encre, quelques résultats, et des réflexions sur les variations futures de l'outil.

LE TEMPS DE L'APPRENTISSAGE

Pour entraîner un réseau neuronal convolutif à effectuer une tâche, il faut disposer d'un ensemble de données d'entraînement composé de paires : données entrantes, et données sortantes désirées. Pour un réseau neuronal destiné à éliminer l'encre, les paires sont constituées d'une photo en lumière transmise avec de l'encre au recto et/ou au verso comme entrée, et d'une photo identique sans encre comme sortie souhaitée. Jusqu'à un milliard d'itérations sur un gigantesque corpus de ces paires sont requises pour que le réseau apprenne à réduire ses erreurs dans la transformation de ses entrées en sorties correspondantes à partir des données d'entraînement. Les

¹ Leila SAUVAGE , Marie-Noëlle GRISON, « Explorer les matérialités du papier grâce aux outils d'imagerie et d'intelligence artificielle au Rijksmuseum, méthodologie et résultats préliminaires » dans *Actes Des Journées d'étude : Les Filigranes, Une Marque à Explorer et Histoire du Papier et de La Papeterie - Actualités de La Recherche*, sous la dir de Claude LAROQUE, Valérie LEE, Maryse PIERRARD, Paris, [En ligne] HiCSA, 2019, p. 238-257 [En ligne] https://web.archive.org/web/20201212131632/https://hicsa.univ-paris1.fr/documents/pdf/PublicationsLigne/Actes%20Laroque%202020/12_SAUVAGE-GRISON.pdf

² Marie-Noëlle GRISON, Leila SAUVAGE, Robert G. ERDMANN, « Toward an Online Paper Study Toolkit: Developing Innovative Tools for Conservators and Researchers to Investigate the Materiality of Pre-Industrial

Paper », in *International Association of Paper Historians (IPH) 35th Congress*, 2021. [En ligne] <https://www.loc.gov/preservation/resources/iphpub/indexiph.html>

³ Les réseaux neuronaux convolutifs sont un type d'algorithme d'apprentissage automatique, inspiré à l'origine par les connexions entre les neurones du cortex visuel des mammifères, et utilisé pour le traitement de vastes corpus d'images. Composés de plusieurs couches, ils sont capables de s'adapter aux données entrantes pour apprendre à résoudre des problèmes. Le terme "convolutif" désigne une opération mathématique qui "filtre" et extrait certaines caractéristiques visuelles/sémantiques de l'image traitée entre une couche du réseau et la suivante.

L'OUTIL « MAGIC ERASER » : QUAND L'INTELLIGENCE ARTIFICIELLE EFFACE L'ENCRE POUR RÉVÉLER LE FILIGRANE

ROBERT G. ERDMANN

performances du réseau seront médiocres en l'absence d'un nombre suffisant de paires pour l'apprentissage ; une situation malheureusement courante dans le domaine du patrimoine.

Pour supprimer l'encre, on pourrait, en principe, créer manuellement un corpus de données d'entraînement en photographiant d'abord un ensemble de papiers filigranés non encrés, puis en appliquant de l'encre sur chacune d'eux des deux côtés, et en les photographiant à nouveau dans la même configuration. Cependant, la nécessité de disposer d'un vaste ensemble de données rend cette méthode inenvisageable. Notre principale contribution, et l'avancée majeure qui a permis au réseau d'atteindre des performances très élevées, a été la création d'un corpus numérique d'exemples d'entraînement, en utilisant un modèle basé sur la physique, pour simuler l'ajout d'encre au recto et au verso d'images de filigranes sans encre. Avec cet ensemble de données artificielles, l'entraînement du réseau neuronal s'est fait en échangeant les entrées et les sorties de la simulation d'ajout d'encre : les images numériquement encrées servent d'entrées au réseau, et les images originales sans encre servent de cible que le réseau doit reproduire.

Nous allons maintenant décrire le modèle physique utilisé pour simuler l'apparence de l'encre ajoutée au papier, puis nous expliquerons comment nous avons créé un vaste corpus d'apprentissage synthétique en utilisant ce modèle, ainsi que des images existantes de filigranes non encrés et de motifs d'encre.

UN PEU DE PHYSIQUE

Nous utilisons un modèle RVB (rouge, vert, bleu) de Kubelka-Munk pour simuler l'effet optique de l'application de l'encre sur le papier. Il s'agit d'une extension de la théorie de base de Kubelka-Munk⁴ qui est largement utilisée pour décrire l'interaction de la lumière avec des matériaux diffusants et absorbants tels que la peinture, le papier et les textiles⁵. Cette théorie fut originellement développée pour la lumière monochromatique, mais le modèle RVB l'étend pour traiter les systèmes de couleurs tristimulus. La théorie de base modélise le flux de lumière dans une dimension lorsqu'il interagit avec un support translucide tel que de la peinture ou de l'encre sur un substrat. Pour le modèle Kubelka-Munk monochromatique, les propriétés du support sont décrites par deux paramètres : le coefficient d'absorption (K) et le coefficient de diffusion (S). Nous ne nous attarderons pas ici sur les détails, mais une dérivation mathématique complète du modèle que nous utilisons peut être trouvée dans Nobbs⁶.

⁴ Paul KUBELKA, Franz MUNK, « An article on optics of paint layers », *Z. Tech. Phys.*, 1931, n° 12, p. 593-601.

⁵ Pour une excellente discussion sur la théorie de Kubelka-Munk, voir Roy S. BERNSTEIN, *Billmeyer and Saltzman's Principles of Color Technology*, New York, 2019.

⁶ James H. NOBBS, « Colour-match prediction for pigmented materials », in Roderick McDONALD (dir.), *Colour Physics for Industry*, vol 2, Bradford, 1997, p. 292-372.

Dans notre approche, ce modèle physique est appliqué pixel par pixel. Il faut donc trois données de base pour l'appliquer : (1) une image en couleur du support sur lequel l'encre sera appliquée ; (2) une image en niveaux de gris de l'encre à appliquer, dont l'intensité correspond à l'épaisseur ou à la densité souhaitée de l'encre ; et (3) les propriétés de l'encre, sous la forme de ses coefficients de diffusion et d'absorption dans les longueurs d'onde rouge, verte et bleue (K_R, S_R, K_G, S_G, K_B , et S_B). L'application du modèle permet alors d'obtenir une nouvelle image simulant l'ajout de l'encre donnée sur le support.

Pour appliquer l'encre au verso, le processus est légèrement modifié. En photographie en lumière transmise, la lumière frappe d'abord l'encre au verso du papier et se déplace ensuite à travers le papier, où elle se diffuse latéralement en se déplaçant vers le recto puis vers l'appareil photo. Le degré de diffusion latérale dépend des propriétés et de l'épaisseur du papier. Dans notre modèle, lorsque l'encre est appliquée au verso, l'image originale de la densité de l'encre est d'abord modifiée à l'aide d'un filtre gaussien. Cette opération floute les bords du trait d'encre, simulant la diffusion de la lumière. Pour ce projet, nous avons adopté une stratégie *ad hoc* en réduisant également l'image de la densité de l'encre d'un pourcentage aléatoire pour tenir compte de la prévalence accrue de la couleur du papier sur les zones encrées au verso par rapport à celles encrées au recto.

**Figure 1**

Image en lumière transmise sans encre (image 2-44) de l'ensemble de données « A » (photographies en lumière transmise) du projet Filigranes Pour Tous.

« FAKE IT ‘TIL YOU MAKE IT », OU COMMENT CRÉER UN ENSEMBLE DE DONNÉES SYNTHÉTIQUES ALÉATOIRES

Pour entraîner le réseau neuronal à supprimer l'encre, nous avons besoin d'un vaste ensemble de données contenant une grande variété de couleurs de papier, de tailles et de formes de filigrane, de couleurs et de densités d'encre, et de largeurs de trait, afin qu'il apprenne à traiter correctement une gamme variée de données entrantes. Le modèle physique que nous utilisons pour encrer nous a permis de synthétiser un grand nombre de paires d'images (non encrées/encrées) de filigranes, afin d'entraîner le réseau à inverser efficacement le processus par lequel l'encre a été ajoutée. Pour créer ce corpus d'entraînement, nous avions besoin des trois composantes susmentionnées pour alimenter notre modèle d'application de l'encre : des images en lumière transmise de filigranes sur papier sans encre, des images en niveaux de gris de motifs d'encre à appliquer au recto et au verso, et des coefficients de diffusion et d'absorption RVB pour l'encre à ajouter. En les combinant de manière aléatoire, nous pouvons générer une variation infinie de « jumeaux » encrés/non-encrés. Nous allons examiner chacune de ces composantes tour à tour.

L'OUTIL « MAGIC ERASER » : QUAND L'INTELLIGENCE ARTIFICIELLE EFFACE L'ENCRE POUR RÉVÉLER LE FILIGRANE

ROBERT G. ERDMANN

Images de filigranes sans encre

Pour trouver une source d'images de filigranes en lumière transmise, nous avons utilisé l'excellent jeu de données mis à disposition dans le cadre du projet Filigranes Pour Tous (FPT)⁷. Parmi les nombreuses ressources offertes par le projet, le jeu de données « A » fournit une collection de 5000 photos en lumière transmise d'archives du minutier central des notaires de Paris (xvi^e- xvii^e e siècles). La collection contient une gamme de couleurs d'encre et de papier, de tailles et de styles d'écriture et, surtout, de nombreuses images de feuilles non encrées. Chacune des images est carrée et d'une résolution suffisamment élevée (400×400 pixels) pour que l'on puisse distinguer la texture du papier, les chaînettes et vergeures, et le filigrane. Un autre avantage de cette collection est que, comme la prise de vue a été réalisée pour capturer correctement des filigranes de différentes tailles, les images couvrent également une large gamme de dimensions physiques. Cette variété fait la qualité des données d'apprentissage. Nous avons séparé manuellement 300 images en deux catégories : 200 images avec encre, et 100 sans encre, puis nous avons entraîné un réseau neuronal simple à départager ces deux catégories pour trier automatiquement la collection de 5000 images. Après vérification et élimination des classifications erronées, nous avons obtenu 1091 images à utiliser comme données entrantes de notre simulateur d'enrage. La **figure 1** montre un exemple sans encre. Lors de la production d'un encrage aléatoire, nous sélectionnons l'une de ces images avec une probabilité uniforme.

Motifs d'encre

Le deuxième élément nécessaire au modèle d'enrage est une collection d'images réalistes de motifs d'encre typiques contenant des lignes d'épaisseur, de densité et de courbure différentes, et réparties sur la page de façon variable. Nous avons utilisé comme matériau-source 545 images haute résolution (environ 3000×4000 pixels) sélectionnées aléatoirement parmi les décisions du banc des échevins de Louvain entre 1362 et 1795, générées automatiquement via la page « image aléatoire⁸ » du système Monk⁹. La **figure 2** montre une image typique d'un folio de manuscrit. Pour extraire un motif d'encre d'une zone aléatoire du folio pour l'appliquer à une image en filigrane, nous sélectionnons une région carrée dont la taille est choisie au hasard dans une fourchette prédéterminée. Le cadre bleu représente une de ces zones, illustrée en pleine résolution dans la **figure 3** et utilisée comme exemple dans les étapes suivantes. Pour obtenir une estimation de la densité de l'encre à partir d'une région carrée

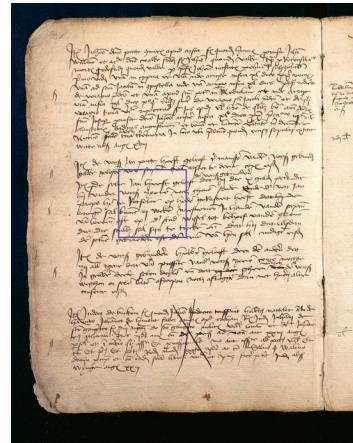


Figure 2

Un exemple d'image d'un folio aléatoire des archives de l'échevinat de Louvain. Le carré bleu représente une région typique à extraire et à réappliquer comme motif d'encre sur une image de filigrane.

[En ligne] <https://monk.hpc.rug.nl/cgi-bin/monkweb?cmd=randompage>

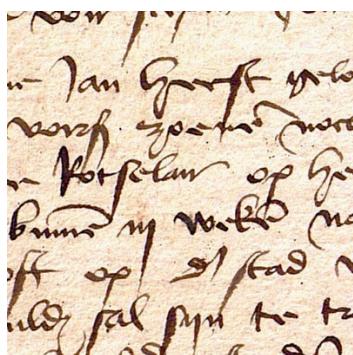


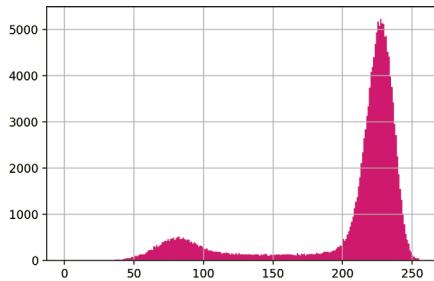
Figure 3

Le détail du carré bleu de la figure 2.

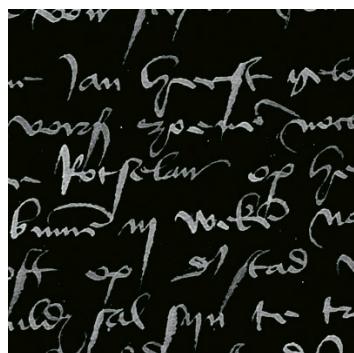
⁷ Xi SHEN, Ilaria PASTROLIN, Oumayma BOOUNOU, et al., « Large-Scale Historical Watermark Recognition: dataset and a new consistency-based approach », 2019 [En ligne] <https://arxiv.org/abs/1908.10254>. Voir aussi <https://filigranes.hypotheses.org/>

⁸ Monk. Monk - (Random page)-Pages. [En ligne] <https://monk.hpc.rug.nl/cgi-bin/monkweb?cmd=randompage> (consulté le 29 avril 2023).

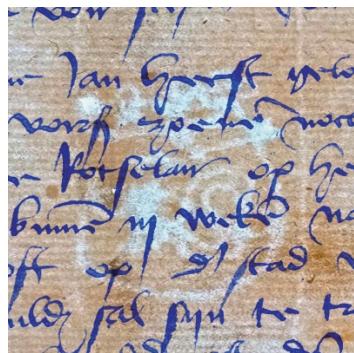
⁹ Lambert SCHOMAKER, Monk. Monk - Search & annotation tools for handwritten manuscripts. [En ligne] <https://monk.hpc.rug.nl/> (consulté le 27 avril 2023).

**Figure 4**

L'histogramme des valeurs du canal vert de la figure 3 montre deux pics (un bas et un haut) associés respectivement à l'encre et au papier. L'axe X indique la luminosité du pixel et l'axe Y le nombre de pixels de cette valeur.

**Figure 5**

Le motif de densité de l'encre extraite de la figure 3. Le blanc indique les zones où l'encre est la plus concentrée et le noir représente l'absence d'encre.

**Figure 6**

Résultat de l'application de la densité d'encre au recto de l'échantillon d'image de filigrane sous forme d'encre bleue translucide à l'aide de notre modèle Kubelka-Munk.

extraite, nous la convertissons dans un espace colorimétrique linéaire et extrayons l'histogramme des valeurs du canal vert (**figure 4**). Pour chaque pixel assigné par le modèle, nous modélisons le pic lumineux et l'utilisons pour estimer la probabilité qu'un pixel corresponde à du papier. Nous inversons et normalisons l'image obtenue pour produire une image de densité d'encre (**figure 5**). Un exemple d'application à un filigrane est présenté dans la **figure 6**. Lors de la création du jeu de données, une photo est choisie au hasard parmi les 545 folios, puis une région carrée est sélectionnée au hasard dans l'image, à l'exclusion des marges.

Propriétés de l'encre

La troisième donnée nécessaire au modèle d'encrage est un ensemble de paramètres de diffusion et d'absorption RVB pour l'encre. Ces paramètres contrôlent la couleur de l'encre, sa transparence pour une densité d'encre donnée, et sa sensibilité aux légères variations de densité de l'encre. Notre objectif étant de produire une grande variété d'échantillons synthétiques afin d'apprendre au réseau à traiter tout ce que les utilisateurs peuvent lui soumettre, il était important d'ajouter une part importante d'aléatoire dans les propriétés de l'encre pour l'ensemble de données d'entraînement. Ainsi, le réseau pourra reconnaître et supprimer l'encre indépendamment de sa couleur ou de son opacité. Nous avons donc choisi les paramètres RVB K et S de manière aléatoire dans une gamme uniforme de valeurs.

La **figure 7** présente plusieurs exemples de paires d'entraînement, où chacune montre une photo sans filigrane à l'encre à côté d'une version où le filigrane a été encré de manière aléatoire au recto et au verso. Notez la diversité de couleurs d'encre, d'opacité, d'orientation et d'échelle, ainsi que la variation de l'échelle et de la complexité du filigrane.

Le réseau neuronal

Pour l'architecture du réseau, nous avons utilisé une implémentation PyTorch¹⁰ d'un U-Net¹¹ avec une ossature ResNet34¹². Dans cette architecture de réseau neuronal convolutif, chaque couche du réseau est entraînée à reconnaître la coopération spatiale des données sortantes de la couche précédente. Au fur et à mesure qu'une image traverse la première moitié d'un réseau U-Net, l'ordinateur la déconstruit en caractéristiques de plus en plus abstraites : il reconnaît d'abord l'orientation des bords localement, puis les textures, les motifs, les éléments et les objets¹³. Dans la seconde

¹⁰ Adam PASZKE, Sam GROSS, Francisco MASSA, et al., « Pytorch: An imperative style, high-performance deep learning library », *Advances in Neural Information Processing Systems*, n° 32, 2019.

¹¹ Olaf RONNEBERGER, Philipp FISCHER, Thomas BROX, « U-Net: Convolutional Networks for Biomedical Image Segmentation », 2015 [En ligne] <https://arxiv.org/abs/1505.04597>

¹² Kaiming HE, Xiangyu ZHANG, Shaoqing REN, Jian SUN, « Deep Residual Learning for Image Recognition », 2015 [En ligne] <https://arxiv.org/abs/1512.03385>

¹³ Pour d'excellentes visualisations en ligne de la façon dont un réseau neuronal voit le monde, voir Chris OLAH, Alexander MORDVINTSEV, Ludwig SCHUBERT, « Feature Visualization », 2017 [En ligne] <https://distill.pub/2017-feature-visualization/>

L'OUTIL « MAGIC ERASER » : QUAND L'INTELLIGENCE ARTIFICIELLE EFFACE L'ENCRE POUR RÉVÉLER LE FILIGRANE

ROBERT G. ERDMANN

moitié du réseau, chaque couche transforme cette représentation abstraite en formes de moins en moins abstraites, jusqu'à la production d'une image RVB. Au sein du réseau, les représentations abstraites en entrée sont transformées en de nouvelles représentations abstraites dans lesquelles l'encre est supprimée tout en conservant les caractéristiques du support papier.

Un élément clé de l'apprentissage de tout réseau neuronal est la définition de la « fonction de perte », qui mesure le taux d'erreur du réseau dans la production du résultat souhaité (image-cible). Nous évaluons cette perte en faisant la somme des différences pixel par pixel entre l'image produite par le réseau et l'image-cible, et en calculant les différences dans les caractéristiques à chaque couche du réseau. Le réseau s'efforce d'égaler l'image-cible, depuis la représentation la plus élémentaire des pixels jusqu'aux représentations plus abstraites, en faisant correspondre les textures, les motifs, etc. Au cours de l'entraînement, il apprend ainsi les relations profondes entre les versions encrées et non encrées, de sorte qu'il peut même compléter des détails de façon plausible lorsqu'ils sont entièrement masqués par l'encre, comme le ferait une personne. Ainsi, lorsque l'encre est si épaisse qu'elle empêche toute transmission de la lumière dans certaines zones, le réseau arrive à prolonger les chaînettes, vergeures et filigranes à partir de zones voisines moins encrées¹⁴.

Au cours de la phase d'apprentissage, nous avons également utilisé une technique appelée « augmentation des données », dans laquelle les paires de données d'apprentissage sont modifiées de manière aléatoire pour en créer de nouvelles, par exemple en changeant leur échelle, en les faisant pivoter, en modifiant leur luminosité ou leur contraste, ou en appliquant des transformations de perspective pour simuler une photographie prise sous un certain angle. Au cours de l'apprentissage, le modèle est donc soumis à un nombre quasi-illimité de variations. Ceci développe la robustesse dont il aura besoin face aux variations rencontrées dans les photos que chargeront les utilisateurs.

L'apprentissage du modèle s'est déroulé par étapes, en commençant par des images du corpus d'apprentissage à résolution très réduite, allant progressivement jusqu'aux images à pleine résolution de 400×400 pixels. Lorsque l'ensemble du corpus d'images, dont chacune varie de manière aléatoire en fonction de la stratégie d'augmentation, est présenté au réseau pendant l'apprentissage, on parle d'« époque ». Après chaque époque, les performances du réseau sur un ensemble de paires d'entraînement qui ne sont pas utilisées pour entraîner le réseau (un « ensemble de test ») sont calculées

¹⁴ L'outil fournit un résultat statistique ; autrement dit, l'image qu'il produit reflète l'apparence la plus probable du filigrane caché sous l'encre. Il est techniquement impossible de mesurer une marge d'erreur dans ce cas précis, puisque nous ne disposons pas d'une image réelle du même filigrane, mais non encré, qui pourrait servir de point de référence. C'est toujours à l'utilisateur, qui a

probablement un savoir préalable au sujet des filigranes, de former un jugement sur la base de ce que soumet l'IA. Un outil de ce type ne prétend jamais remplacer l'intelligence humaine ni les connaissances des spécialistes, il est uniquement destiné à les augmenter et à assister dans la prise de décision.



Figure 7

Plusieurs exemples du processus d'encrage dans lequel de l'encre de différentes couleurs, opacités, largeurs de trait et orientations est ajoutée au verso et au recto de différentes images de filigrane sans encre à l'aide de notre modèle physique.



et visualisées (**figure 8**). L'entraînement s'arrête lorsque la fonction de perte sur l'ensemble de test ne diminue plus.

Une fois l'apprentissage terminé, le modèle est figé et prêt à être utilisé pour supprimer l'encre d'images qu'il n'a jamais vues. Le réseau peut traiter toute image de taille raisonnable et renverra une image de la même taille. Étant donné que la résolution du corpus d'apprentissage n'était que de 400 pixels de côté, le réseau n'a jamais vu de gros plans extrêmes, avec des traits d'encre larges de plusieurs centaines de pixels. Ses performances seront donc généralement médiocres sur de telles entrées, car il considérera ces marques d'encre comme des taches ou une coloration du papier à ne pas modifier.

Déploiement

Pour tester son utilité dans des scénarios réels, nous avons développé une interface web pour le réseau neuronal, baptisée « *Magic Eraser* ». L'utilisateur dispose d'une interface simple sur laquelle il peut glisser-déposer des images ou sélectionner des fichiers depuis son ordinateur. La version mobile permet de capturer directement une photo avec l'appareil photo d'un smartphone. L'interface affiche l'image originale et l'image traitée côté à côté pour la comparaison. D'un clic, l'utilisateur télécharge un fichier TIF de l'image désencrée, dont le nom de fichier dérive de celui du fichier original pour plus de commodité. La **figure 9** montre une capture d'écran de l'outil appliquée à deux images encrées du jeu de données FPT.

Figure 8
Visualisation des performances du modèle pendant son apprentissage. La colonne de gauche est une sélection d'images entrées du réseau, la colonne centrale est l'image sortante produite par le réseau et la colonne de droite est l'« image-cible » (la bonne réponse).



Figure 9
Une capture d'écran de l'outil « *Magic Eraser* », montrant deux entrées contenant de l'encre non synthétique de l'ensemble de données FPT (98_1-48 et 84_B-2 de l'ensemble de données « A ») dans la colonne de gauche et les versions « désencrées » dans la colonne de droite.

L'OUTIL « MAGIC ERASER » : QUAND L'INTELLIGENCE ARTIFICIELLE EFFACE L'ENCRE POUR RÉVÉLER LE FILIGRANE

ROBERT G. ERDMANN

L'outil a été partagé en privé pour être testé par un petit groupe de chercheurs, qui ont accepté de faire part de leurs commentaires sur ses performances, et de permettre que les images chargées soient stockées sur le serveur en vue d'une utilisation future afin d'améliorer l'outil.

Les premiers résultats sont prometteurs. Citons par exemple les travaux de Nadine Dumain (**figure 10**). Bien que l'outil n'ait été entraîné que sur des documents manuscrits, le large éventail de couleurs, d'orientations et d'échelles lui a permis d'apprendre à identifier les lignes d'encre de manière plus large, de sorte qu'il est également performant sur les estampes et les dessins.

CONCLUSION

Les progrès modernes de l'intelligence artificielle, associés à un modèle physique réaliste de l'effet de l'encre sur les images en lumière transmise, ont permis de mettre au point un outil gratuit, efficace et facile à utiliser par les chercheurs afin d'améliorer la lisibilité des filigranes. L'outil est encore en phase de test, mais nous serons heureux de fournir l'URL sur contact aux chercheurs intéressés, en échange d'un retour d'information sur les performances de l'outil et de suggestions pour le rendre plus utile (robert.g.erdmann@gmail.com). À l'avenir, nous prévoyons de l'intégrer à notre boîte à outils en ligne destinée à la communauté des spécialistes du papier.

Article traduit de l'anglais par Marie-Noëlle GRISON

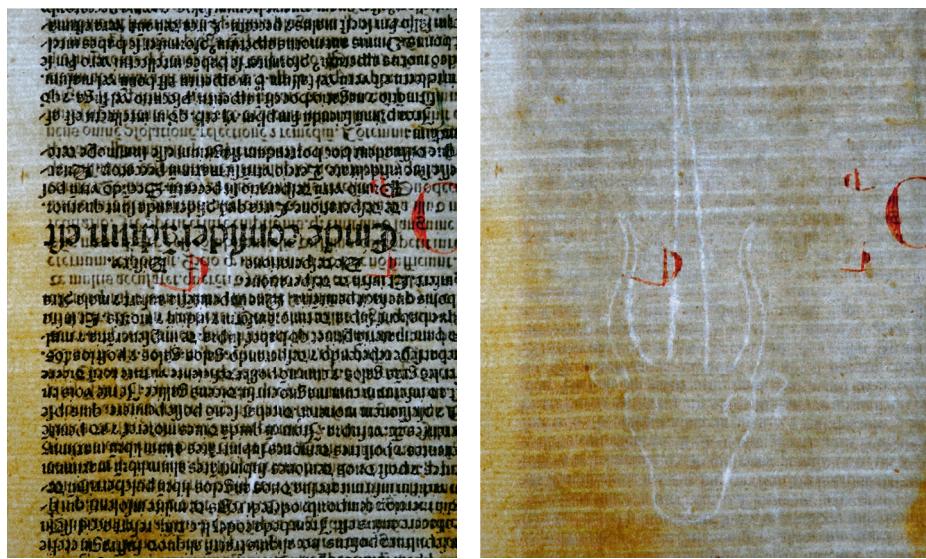


Figure 10

Un exemple de filigrane désencreé par « Magic Eraser », sur un folio du livre d'Anton Koberger, *Speculum Morale*, Nuremberg, 1485.
Photographie : Nadine Dumain.