

La machine voit-elle le kitsch ? Une analyse computationnelle de la culture visuelle des skyblogs

Emmanuelle Bermès¹ , and Marina Hervieu^{1,2}

¹ École nationale des chartes - PSL, Centre Jean Mabillon, Paris, France

² Bibliothèque nationale de France, Département du dépôt légal du web, Paris, France

Abstract

This presentation questions whether it is possible to use computational methods to identify and characterise the concept of kitsch, which is by its nature elusive. The work is based on images published and shared on the French blog platform Skyblog and draws on the collections of the BnF web archives. Using supervised and unsupervised visual analysis tools, the presentation offers a historical exploration of the digital aesthetics of the French-speaking web in the early 2000s.

Mots-clés: Archives du web, Vision par ordinateur, Kitsch, Skyblog, Culture numérique

Keywords: Web archive, Computer Vision, Kitsch, Skyblog, Visual and digital culture

1 Introduction

Une photo de chat encadrée de fleurs aux couleurs vives, un fond d'écran rose imitant une fourrure léopard, un dessin de dauphin dont le bleu scintille de paillettes au rythme saccadé d'un gif animé, un photomontage d'une cavalière galopant en bord de mer au coucher du soleil : le web des années 2000-2010 était saturé d'images numériques dont l'esthétique (motifs, textures, couleurs et effets visuels) nous fait aujourd'hui sourire. Au premier regard, on ne peut que s'exclamer avec spontanéité : « c'est kitsch ! » [23]. Pourtant l'omniprésence et la circulation de ces images d'un site à l'autre, dans une plateforme comme Skyblog (2002-2023), suggèrent qu'elles étaient à l'époque valorisées et appréciées. Bénéficiaient-elles de l'attrait de la nouveauté, dont on aurait ensuite saturé jusqu'au dégoût, ce qui en ferait des marqueurs temporels ? Étaient-elles réservées à certaines audiences, certains milieux socio-culturels, et déjà jugées kitsch à l'époque par d'autres qui tiendraient plutôt de la culture savante, seule détentrice du « bon goût » ?

Répondre à ces questions implique d'interroger ces sources nativement numériques et issues du web [4 ; 8 ; 25], afin de parvenir à caractériser et contextualiser ce que l'on entend par « kitsch ». Cette étude s'appuie sur l'archive Skyblog constituée par la Bibliothèque nationale de France, au titre du dépôt légal du web, lorsque la plateforme a été mise hors ligne à l'été 2023 [15]. Pour l'exploration de ce corpus massif de 12,6 millions de blogs archivés, composé de textes et d'images, le recours aux méthodes computationnelles s'avère indispensable. C'est dans ce cadre que voit le jour le projet de recherche SkyTaste : « Les skyblogs : un terrain d'analyse sensoriel, technologique et émotionnel » financé par l'Université PSL pour une durée de quatre ans dans le cadre du programme « Young Researcher Starting Grant ». L'un des axes du projet est consacré à l'analyse de l'atmosphère visuelle de la plateforme, et c'est dans ce contexte que s'inscrit cette étude sur le kitsch. Menée collectivement par plusieurs membres de l'équipe, en collaboration avec des partenaires institutionnels et des étudiants, cette étude participe plus largement à l'effort

Emmanuelle Bermès, and Marina Hervieu. "La machine voit-elle le kitsch ? Une analyse computationnelle de la culture visuelle des skyblogs." *Actes de la Conférence Humanistica*, éd. par Serena Crespi, Simon Gabay, Martin Grandjean, Ariane Pinche, Marie Puren et Léa Saint-Raymond. Vol. 4. Anthology of Computers et the Humanities. 2026, 92–98. <https://doi.org/10.63744/s8EJp9jEw9DM>.

© 2026 par les auteurs. Sous licence Creative Commons Attribution 4.0 International (CC BY 4.0).

commun de description et de compréhension de la collection Skyblog [6; 24] en tant que patrimoine numérique [2; 10; 14].

Si des tâches de *computer vision* telles que la détection d'objets [12], la segmentation ou encore l'analyse de similarité sont bien maîtrisées par les modèles actuels basés sur l'intelligence artificielle [5; 21], leur confrontation à une notion qui relève du jugement esthétique pose question [31]. Les machines peuvent-elles voir le kitsch? Telle est la problématique à laquelle nous nous proposons de répondre dans cette présentation.

À cette fin, nous verrons quelles sont les méthodes computationnelles à notre portée pour analyser un jeu massif d'images. Nous étudierons comment elles peuvent contribuer à définir le kitsch numérique [30] et à identifier sa place au sein de la collection Skyblog. Nous pourrions ainsi espérer saisir la place du kitsch dans la culture visuelle du web francophone diffusée via les blogs du début du XXI^e siècle et interroger, plus largement, les relations entre l'esthétique kitsch et la culture numérique au fil du temps [10; 13].

2 Objet de recherche et état de l'art

Apparu dans les textes entre 1850 et 1870, le terme « kitsch » désigne une forme esthétique qui, si elle a fait couler beaucoup d'encre depuis près de deux siècles dans le domaine de l'art, résiste toujours aux définitions en raison de sa nature mouvante, voire fuyante. Si, pour certains [20], un coup d'œil suffit à le reconnaître, établir des critères figés permettant de le caractériser à coup sûr est une gageure, en particulier parce qu'il est lié à une appréciation esthétique, culturelle et sociale qui est susceptible d'évoluer. L'histoire de l'art distingue trois périodes différentes où la définition du kitsch et les jugements de valeur dont il fait l'objet ont évolué. De la fin du XIX^e s. au début du XX^e, le kitsch est associé à la dévalorisation de l'art par sa reproductibilité dans le cadre d'une culture populaire de masse [26]. Au cours du XX^e siècle, des artistes du *ready-made*, Pop art ou encore du *Camp* réinvestissent les objets et les motifs industriels, revendiquant leur capacité à transformer ce qui était jugé laid, étrange, bizarre ou anecdotique, des objets ou des représentations relevant du kitsch, au rang d'art par le biais de la démarche artistique, de la performance ou de sa marchandisation. Enfin, au début du XXI^e siècle, le kitsch est revendiqué comme tel, notamment chez des artistes comme Jeff Koons qui le marchandisent. Dans ce troisième temps du kitsch, celui-ci est davantage assumé [29] et largement partagé via le web qui joue un rôle déterminant dans cette reconnaissance à travers la circulation des tropes visuels.

L'exemple de l'imprimé léopard illustre la difficulté à saisir le kitsch [9] : sa charge symbolique varie à travers le temps, alternant entre le luxueux et le vulgaire, le sauvage et l'érotique, le subversif et le convenu, le chic et le mauvais goût. Or, l'expression « mauvais goût » suppose en soi qu'il existe un « bon goût », c'est-à-dire une norme sociale et esthétique à laquelle il est nécessaire de se conformer pour acquérir une forme de reconnaissance. Le goût introduit une hiérarchisation des pratiques culturelles, qui se reflète dans des hiérarchies sociales et politiques.

Plus spécifiquement dans le domaine du numérique, le kitsch n'a pas fait en tant que tel l'objet d'une étude systématique. Si Alice Pfeiffer [28] préfère rattacher les skyblogs à une tendance qu'elle nomme le « néo-moche » plutôt qu'au kitsch, il nous semble que le besoin de caractériser l'esthétique de la culture numérique visuelle et de définir en quoi, pourquoi et comment le kitsch nous saute aux yeux dans le web des années 2000 reste aujourd'hui entier. Si le kitsch est si difficile à reconnaître, à caractériser et à décrire, même pour les historiens et historiennes de l'art, la question de son appréhension par les machines constitue un enjeu heuristique. L'objectif de notre étude n'est pas seulement d'outiller notre exploration du kitsch dans l'archive des skyblogs, mais de comprendre si la façon dont les machines analysent les images fait émerger des motifs, des critères, des récurrences qui permettraient d'alimenter une recherche encore inédite sur le kitsch en tant qu'esthétique numérique.

3 Sources : origine, statut juridique et modalité d'accès

Le corpus étudié provient de l'archive des skyblogs collectés en 2023 par la BnF au titre du dépôt légal du web [32]. Cette archive représente environ 40 To de données et se compose de 12,6 millions de blogs (ceux qui, parmi les 33 millions créés sur une période d'environ 20 ans, étaient encore en ligne en 2023), de 7,4 millions de pages de profil et de plusieurs jeux de données complémentaires (métadonnées statistiques, données éditoriales, données d'usage, documentation). L'ensemble représente près de 2 milliards d'URL, un milliard de fichiers multimédias (images, vidéos, sons) et plus de 500 millions d'articles. Bien que massive, cette archive ne constitue qu'environ 2% du volume global lorsqu'elle est rapportée à l'ensemble des archives du web conservées par la BnF, qui, depuis 1996, a cumulé jusqu'à 52 milliards d'URL en 2023.

Les archives du web sont protégées par les droits de propriété intellectuelle et ne sont pas accessibles en ligne. Leur consultation est restreinte aux emprises de la BnF, au sein de l'espace de recherche, en rez-de-jardin, ou dans certaines bibliothèques partenaires disposant d'un poste de consultation dédié. Dans ce contexte, l'accès se limite à une consultation individuelle des sites, blog par blog : une « lecture proche » qui réduit considérablement la recherche. Or, à l'échelle de la thématique étudiée, le recours à des méthodes computationnelles s'impose : celui-ci est rendu possible par l'offre de services du BnF DataLab, un service opéré conjointement par la BnF et l'IR* Huma-Num dans les espaces de recherche de la bibliothèque, et qui met à disposition des outils, des infrastructures numériques et un accompagnement pour les méthodes des humanités numériques.

Une convention de recherche a été établie entre la BnF et l'École nationale des chartes - PSL, dans le cadre des projets Skybox et SkyTaste financés conjointement [3]. Elle permet un accès élargi, non seulement pour la consultation, mais surtout pour le traitement et l'export des données dans un cadre sécurisé. Dans ce contexte, un membre de l'équipe bénéficie du statut de chercheur associé et dispose d'un poste informatique autorisant l'analyse et l'extraction des sources de façon autonome. Cette collaboration repose sur un travail en synergie avec les équipes du dépôt légal du web : les réflexions de la recherche alimentent l'évolution des procédures et contribuent au développement et à l'adaptation de l'offre de services destinée à la recherche.

4 Méthodologie et description de la chaîne de traitement

L'analyse porte exclusivement sur les images intégrées aux articles de blogs, ainsi sont exclues les fonds d'écran, avatars et galeries de profils. Plusieurs formats d'images sont traités : png, jpg et gif. L'étude se concentre sur un corpus de 50 000 images (soit 0.005% de la collection). Cet échantillonnage est nécessaire compte tenu de la nature surabondante du kitsch numérique et des contraintes actuelles d'export. La démarche de constitution de corpus est itérative, avec des tests successifs sur des ensembles de données de taille croissante. Les premières étapes de la chaîne de traitement s'effectuent à l'aide de la boîte à outils « Skybox » développée par la BnF, l'équipe de recherche ayant participé à sa définition fonctionnelle.

Trois scénarios de recherche, inspirés des méthodes computationnelles, sont mobilisés pour répondre à notre problématique. Le premier scénario repose sur une constitution ciblée, inductive et multimodale du corpus, suivie d'une analyse supervisée et déductive. Les deuxième et troisième scénarios s'appuient sur une sélection aléatoire de blogs, puis l'analyse reste supervisée dans le deuxième cas et non supervisée dans le troisième. Le premier scénario a été éprouvé par plusieurs projets de recherche et par des étudiants depuis plus d'un an, et a fait l'objet de plusieurs publications et conférences [17; 18; 19]. Le deuxième scénario est en cours d'expérimentation et le troisième est en perspective.

Dans le scénario 1, la première étape consiste à valider le premier périmètre du corpus en identifiant une liste de blogs pertinents à partir de mots-clés présents dans les pseudos. Une vérification

qualitative en « lecture proche » permet de valider ou d'exclure les résultats. Ces deux opérations sont menées de manière itérative car l'examen de certains résultats permet d'ajuster progressivement la liste. Les blogs retenus sont ensuite analysés avec l'outil Hyphe [27], un crawler web développé par le Medialab de SciencesPo, connecté aux archives du web de la BnF, pour lesquelles un critère de sélection spécifique aux skyblogs a été développé. Les réseaux de liens, entrants et sortants, sont étudiés afin de stabiliser ce second périmètre du corpus à exporter. Après une étape importante de traitement des données extraites qui consiste à générer les permaliens associés aux blogs puis à les trier (un blog peut se constituer de plusieurs milliers de pages mais toutes ne sont pas utiles pour notre étude), nous importons dans Panoptic [1 ; 7] le jeu de données constitué d'un corpus d'images associées à des fichiers de métadonnées descriptives.

Conçu pour la recherche en sciences humaines et sociales, l'outil permet d'explorer et d'annoter de larges corpus d'images, soit par similarité visuelle ou par description textuelle. En effet, Panoptic repose sur le modèle d'apprentissage CLIP (*Contrastive Language-Image Pretraining*) : il est pré-entraîné sur des images et sur des textes décrivant ces images. L'objectif du modèle est d'apprendre une correspondance entre le langage naturel et le contenu visuel. Dans le cadre de notre étude, cet outil est mobilisé pour faire émerger des catégories hypothétiques du kitsch. Celles-ci peuvent s'inspirer de la littérature de référence mais un travail de modélisation est nécessaire pour permettre une approche computationnelle :

- catégories structurelles : la composition des plans d'un photomontage, la forme d'un gif animé (carré, rond, en forme de dauphin), la surcharge visuelle (alternance de textes, d'images, de motifs décoratifs), l'encadrement ostentatoire (bordures épaisses, réalistes et/ou animées), etc. ;
- catégories chromatiques : une palette chaude, froide, saturée, brillante, dégradée, dorée, contrastée ;
- catégories d'usage : des expressions affectives (« Je t'aime », « Bonne journée »), des représentations stéréotypées du corps humain musculeux, lisse et brillant, un romantisme populaire (figures de couples, roses, animaux attendrissants) ou des décorations pieuses (anges et figures sulpiciennes, croix et symboles mystiques).

Ces catégories sont annotées, triées, filtrées et exportées individuellement et de façon semi-automatique afin d'être comparées entre les membres de l'équipe pour produire une grille consolidée, ensuite testée sur des corpus aléatoires dans le cadre du scénario 2. Ces deux approches d'analyse visuelle sont supervisées et elles permettent, dans un premier temps, d'établir de façon déductive des catégories de kitsch interprétées par l'œil humain, et, dans un second temps, de prédire les catégories. Enfin, l'approche non supervisée, qui est envisagée dans le troisième scénario, vise à identifier des regroupements visuels que l'œil humain ne saurait voir mais qui du point de vue algorithmique sont susceptibles de révéler des catégories inédites. Une telle perspective pourrait être complétée d'une phase de *fine-tuning* du modèle CLIP, afin d'adapter ses représentations aux résultats des deux premiers scénarios.

La comparaison des résultats issus des trois scénarios doit *in fine* permettre une cartographie du kitsch numérique et faciliter l'appropriation des techniques d'apprentissage automatique dans un cadre de recherche en sciences humaines.

5 Perspectives et discussions

Cette première étude, qui mobilise les méthodes d'analyse computationnelle au service d'un questionnement historique sur la culture visuelle du web, nous permet de faire émerger un tableau du kitsch dans les skyblogs. Cette esthétique est largement partagée par la reprise des images d'un site à l'autre et par leur déclinaison en de multiples variations qui évoquent un nouveau Pop art. Ces images kitsch qui saturent les skyblogs fascinent autant qu'elles rebutent, voire dérangent. Dans une forme de continuité avec le baroque du XVIIe siècle ou la peinture romantique du XIXe, le

kitsch numérique des années 2000 est une esthétique de l'excès et de la surenchère qui se décline en un vocabulaire visuel de référence dont l'usage semble parfois relever de l'invocation quasi mystique[11]. Cette dimension symbolique, voire religieuse de l'esthétique kitsch n'est pas sans implication politique : en contrepoint des images partagées et bricolées du web des années 2000 à 2010, nous sommes aujourd'hui confrontées à l'uniformisation esthétique des images générées par IA, qui nous acculent dans les stéréotypes [16 ; 22]. Une telle normalisation des représentations peut contribuer à éroder les frontières entre le réel et la fiction, ouvrant la porte à l'instrumentalisation par des régimes autoritaires. Notre étude contribue à inscrire ce constat dans un temps plus long, et à comprendre les mécanismes à l'œuvre dans le dialogue entre la machine et l'image, en contexte numérique.

Remerciements

Le projet de recherche SkyTaste est financé par l'Université PSL dans le cadre de son programme « Young Researcher Starting Grant » 2024 et par le plan quadriennal de la recherche de la Bibliothèque nationale de France. Ce travail a bénéficié du soutien du Grand Programme de Recherche « CultureLab » de l'Université PSL mis en œuvre par l'ANR sous la référence ANR-10-IDEX-0001.

Références

- [1] ALIÉ, Félix, BOUTÉ, Édouard et GÖDICKE, David. « Explore large datasets of images with Panoptic ». In : *FOSDEM*. Bruxelles : Université Libre de Bruxelles, 2025. URL : https://archive.fosdem.org/2025/events/attachments/fosdem-2025-4635-explore-large-image-datasets-with-panoptic/slides/238044/Presentat_ZPDqSu6.pdf.
- [2] BERMÈS, Emmanuelle. « De l'écran à l'émotion : quand le numérique devient patrimoine ». Paris : École nationale des chartes-PSL, 2024.
- [3] BERMÈS, Emmanuelle. « Skybox et SkyTaste : deux projets pour comprendre l'histoire des skyblogs ». *Web Corpora*. 24 nov. 2024. DOI : 10.58079/12wsu.
- [4] BERMÈS, Emmanuelle, GEBEIL, Sophie et MIURA, Grégory. « Le web comme source : En quête d'une science auxiliaire ? » In : *Les Cahiers du numérique 30*, no. 3 (2024), p. 195-215. DOI : 10.1684/lcn.2024.9.
- [5] BERMÈS, Emmanuelle, LECLAIRE, Céline et MOREUX, Jean-Philippe. « L'image comme particule élémentaire, ou les prémisses d'un changement d'échelle à la BnF ». In : *The Measurement of Images. Computational Approaches in the History and Theory of the Arts*, sous la direction de Clarisse Bardiot et Emmanuel Château-Dutier. Lille : Presses universitaires du Septentrion, 2023. URL : <https://hal.science/hal-03991515>.
- [6] BIGOT, Violaine et MAILLARD-DE LA CORTE GOMEZ, Nadja. « Skyblog – Focus sur l'émergence d'un genre littéraire numérique : les chroniques ». *Web Corpora*. 10 avr. 2025. DOI : 10.58079/13pyd.
- [7] BOUTÉ, Édouard, JULLIARD, Virginie, ALIÉ, Félix, GÖDICKE, David, PAILLER, Fred et ECREMENT, Victor. « PANOPTIC, un outil d'exploration par similarité de vastes corpus d'images ». In : *Humanistica 2024*. Meknès, Maroc : Association francophone des humanités numériques, 2024. URL : <https://hal.science/hal-04687627v1>.
- [8] BRÜGGER, Niels. *The Archived Web : Doing History in the Digital Age*. Cambridge : MIT Press, 2023. DOI : 10.7551/mitpress/10726.001.0001.
- [9] BUFFARD-MORET, Julie. « Le motif léopard dans le vestiaire féminin au XXe siècle ». In : *Apparence(s) 11* (2022). DOI : 10.4000/apparences.3773.

- [10] CARDON, Dominique. *Culture numérique*. Paris : Presses de Sciences Po, 19 mai 2019.
- [11] CAROBALANTE, Jean-Baptiste. *La Peinture de masse : art et société de consommation*. Strasbourg : L'atelier contemporain, Essais sur l'art, 2026.
- [12] CHARPIER, Marion, BERMÈS, Emmanuelle et STUTZMANN, Dominique. « TiamaT : un workflow de computer vision entre iconographie médiévale et design contemporain ». In : *Humanités numériques* (2026).
- [13] CUBITT, Sean. « Aesthetics of the Digital ». In : *A Companion to Digital Art*. John Wiley & Sons, Ltd, 2016, p. 265-280. DOI : 10.1002/9781118475249.ch10.
- [14] FAYE, Alexandre. « Aujourd'hui les skyblogs entrent dans l'Histoire ». Web Corpora. 4 sept. 2024. DOI : 10.58079/128yr.
- [15] FAYE, Alexandre. « Il était une fois Skyblog ». Web Corpora. 24 oct. 2024. DOI : 10.58079/12kh5.
- [16] GUILLAUD, Hubert. « Acculés dans les stéréotypes ». In : *Dans les algorithmes* (2024). URL : <https://danslesalgorithmes.net/2024/10/18/accules-dans-les-stereotypes/>.
- [17] HERVIEU Marina et Bermès, Emmanuelle. « Skyblog se met au vert, retour d'expérience du Hackathon 2025 Ecomedia – Pictoria x BNF Datalab ». Consortium-HN pictoria. 19 jan. 2026. URL : <https://pictoria.hypotheses.org/6332>.
- [18] HERVIEU, Marina. « Le projet Skybox à RESAW 2025 ». Web Corpora. 21 juill. 2025. DOI : 10.58079/14ece.
- [19] HERVIEU, Marina. « Skyblog : les étudiants lâchent leurs com's ». Web Corpora. 16 oct. 2025. DOI : 10.58079/14z04.
- [20] JOHANSSON, FRANZ et VALLESPER, Mathilde. *Le Kitsch ou le naufrage du sublime*. Strasbourg : Presses universitaires de Strasbourg, 2025.
- [21] KEE, Kevin et COMPEAU, Timothy J. *Seeing the Past with Computers : Experiments with Augmented Reality and Computer Vision for History*. Ann Arbor : University of Michigan Press, 2019. DOI : 10.3998/mpub.9964786.
- [22] KONRAD, Moritz, BRUHN, Matthias et WEINSTOCK, Katharina. « From Glitch to Kitsch ». In : *Begriffe des digitalen Bildes*, sous la dir. de Matthias BRUHN et Katharina WEINSTOCK. T. 6. München, Vienne : Ludwig-Maximilians-Universität München, 2025, p. 38-49. DOI : 10.5282/ubm/epub.126477.
- [23] LIPOVETSKY, Gilles et SERROY, Jean. *Le Nouvel âge du kitsch. Essai sur la civilisation du "trop"*. Paris : Gallimard, 2023.
- [24] LOBBÉ, Quentin. « Jarrête mon bloG ! JdéménaGe iiicii ! » Web Corpora. 5 nov. 2025. URL : <https://webcorpora.hypotheses.org/8791>.
- [25] MUSIANI, Francesca, PALOQUE-BERGÈS, Camille, SCHAFER, Valérie et G. THIERRY, Benjamin. *Qu'est-ce qu'une archive du web ?* Encyclopédie numérique. Marseille : OpenEdition Press, 2019.
- [26] OLALQUIAGA, Céleste. *Royaume de l'artifice : l'émergence du kitsch au XIXe siècle*. Lyon : Fage, 2008.
- [27] OOGHE-TABANOU, Benjamin, JACOMY, Mathieu, GIRARD, Paul et PLIQUE, Guillaume. « Hyperlink is not dead ! » In : *Proceedings of the 2nd International Conference on Web Studies*. New York : Association for Computing Machinery, 2018, p. 12-18. DOI : 10.1145/3240431.3240434.

- [28] PFEIFFER, Alice. *Le Goût du moche*. Paris : Flammarion, 2021. DOI : 10.3917/flam.pfeif.2021.01.
- [29] PIĄTEK, Maria. « Kitsch in Visual Arts : Deconstruction of Kitsch Through its Use in Elite and Popular Arts ». In : *Non-Artistic Kitsch*, sous la dir. de Michał SZOSTAK. Londres : Palgrave Macmillan, 2024, p. 45-64.
- [30] QUARANTA, Domenico. « Digital Kitsch : Art and Kitsch in the Informational Milieu ». In : *The Changing Meaning of Kitsch : From Rejection to Acceptance*, sous la dir. de Max RYNNÄNEN et Paco BARRAGÁN. Londres : Palgrave MacMillan, 2023, p. 205-228.
- [31] TAYLOR, Arnold et TILTON, Lauren. *Distant Viewing*. Cambridge, MA ; Londres : MIT Press, 2023. DOI : 10.7551/mitpress/14046.001.0001.
- [32] TYBIN, Vladimir. « Les skyblogs au service de la science ». In : *Chroniques, le magazine de la BnF* 99 (2024). URL : <https://www.bnf.fr/fr/les-skyblogs-au-service-de-la-science>.