

Classification of Abstract Painting

Raphaël Legrand, Enguerrand Le Niliot, Elliott Dionnet, Arthur Evain

Juin 2026

Encadrants : Henri Maître, Loïc le Folgoc

Résumé

Ce travail propose une classification non supervisée d'un corpus de peintures abstraites fondée exclusivement sur des propriétés visuelles objectives, sans recours à aucune donnée sur les artistes ou leurs écoles. Nous mobilisons des descripteurs bas-niveau (couleur, texture, composition) combinés aux espaces latents d'un réseau convolutif (CNN) et d'un Autoencodeur Variationnel (VAE), fusionnés par *alpha-blending* pondéré. Ces représentations multimodales ont été réduites par Analyse en Composantes Principales (PCA) et partitionnées par K-Means. La classification est affinée par un filtre de qualité basé sur le score de Silhouette. Le but du projet est d'évaluer dans quelle mesure les clusters émergents correspondent aux catégories produites par l'histoire et la critique de l'art.

1 Déclaration d'utilisation de l'Intelligence Artificielle

Conformément aux consignes du projet, cette section détaille de manière exhaustive l'utilisation des intelligences artificielles génératives tout au long de notre étude et lors de la rédaction de ce rapport. Les modèles sollicités pour ce projet sont Claude et Gemini.

De manière générale, l'intelligence artificielle a aidé au développement du code, exclusivement pour la visualisation avec Matplotlib. L'IA a aussi aidé à la mise en page de ce rapport en LaTeX.

Pour plus de clarté, nous détaillons ci-dessous nos requêtes spécifiques en séparant l'aide méthodologique de la génération de code pur.

1. Questions méthodologiques et pistes de réflexion

- **Extraction CNN** : Demande de renseignements et d'explications théoriques sur le processus d'extraction des caractéristiques (feature maps) issues des premières couches d'un réseau de neurones convolutif.
- **Espace siamois** : Demande de renseignements théoriques sur l'utilisation d'un espace siamois pour un VAE, et sa mise en pratique dans notre projet.

- **Stratégie de fusion** : Recherche de pistes et de conseils méthodologiques sur la manière la plus pertinente de fusionner nos différentes matrices de descripteurs (CNN, géométriques, etc.).
- **Visualisation en 2D** : Piste pour savoir quelle méthode utiliser pour projeter en 2D afin de s'en servir pour le logiciel (T-SNE/UMAP/PACMAP)
- **Aide pour l'état de l'art** : Recherche d'articles et aide à la compréhension de ces derniers sur des sujets dont ne nous sommes pas expert

2. Assistance au code pur et à la visualisation

- **Aide à la documentation et à l'implémentation des features** : L'IA nous a permis d'accéder rapidement à de la documentation pour traduire des idées abstraites de features en opérateurs machine concrets, et de savoir le nom de la librairie et de la fonction python qui l'implémente.
- **Génération de bilans (Fusion)** : Aide pour l'affichage de la répartition stylistique de nos clusters, ainsi que pour le tracé de la matrice des distances Inter-Artistes.
- **Inspection topologique (Fusion)** : Aide à la création d'un script permettant d'afficher visuellement 5 images situées au centre d'un cluster, 5 à sa frontière, et 5 tirées de manière aléatoire.
- **Logiciel de Visualisation** : Une aide au débogage ainsi qu'une aide à la mise en ligne du logiciel. L'IA a aussi été utile afin de générer des csv utilisables à la fin des codes, pour le logiciel.

2 Introduction

L'art abstrait constitue l'un des corpus les plus singuliers et les plus résistants que puisse rencontrer un système de classification automatique. Né au tournant du XX^e siècle du rejet délibéré des codes de la représentation figurative, il se définit précisément par l'absence des repères visuels (objets, personnages, perspectives)

sur lesquels s'appuient ordinairement les algorithmes de vision par ordinateur. Pourtant, depuis un siècle, historiens, critiques, philosophes, psychologues, sociologues et ingénieurs en apprentissage automatique ont tous tenté d'en ordonner le corpus. Ce projet s'inscrit dans cette tradition interdisciplinaire, avec une contrainte supplémentaire et délibérée : ignorer toute métadonnée sur les peintres, leurs écoles ou leurs biographies, pour proposer une **classification non supervisée** fondée exclusivement sur les propriétés visuelles objectives des œuvres elles-mêmes.

2.1 Enjeux philosophiques

Toute classification de l'art abstrait se heurte d'emblée à une objection philosophique. Weitz [Weitz, 1956] soutient que l'art est un *open concept* dont aucune condition nécessaire et suffisante ne peut délimiter les contours ; Danto radicalise ce point en montrant que deux tableaux visuellement identiques peuvent appartenir à des catégories esthétiques entièrement distinctes selon leur contexte [Danto, 1964]. Notre projet assume pleinement cette limite : ce que nous classifions sont des *signaux visuels*, non des œuvres au sens de Danto. Goodman offre néanmoins un fondement favorable [Goodman, 1968] : l'œuvre abstraite *exemplifie* des propriétés qu'elle met en avant (la lourdeur d'un Rothko, la régularité d'un Mondrian) propriétés accessibles à des descripteurs quantitatifs et qui justifient notre démarche.

2.2 Enjeux psychologiques et neuroesthétiques

La psychologie de la perception légitime directement nos choix techniques. Les variables « collatives » de Berlyne [Berlyne, 1971] (complexité, nouveauté, ambiguïté) se traduisent en mesures quantitatives comme l'entropie de texture ou la variance chromatique, ce qu'on calculera dans l'approche géométrique. Le modèle de Leder et al. [Leder et al., 2004] montre que face à l'abstraction, l'absence de reconnaissance d'objet force le système cognitif vers des traitements stylistiques de haut niveau, motivant le recours à des représentations profondes. Nous tenterons de les modéliser à l'aide de CNN et de VAE. Zeki fournit enfin une base neurobiologique à la pertinence de descripteurs différenciés, couleur, forme et mouvement relevant d'aires corticales distinctes [Zeki, 1999].

2.3 Perception et catégorisation humaine des œuvres abstraites

Face à une œuvre abstraite, l'observateur opère une catégorisation visuelle rapide fondée sur la détection de prototypes [Rosch, 1978] : régularité géométrique, palette chromatique, dynamisme gestuel sont extraits quasi-immédiatement, avant toute identification consciente du mouvement ou du style. Cette extraction repose sur des principes gestalistes de synthèse globale (tension,

équilibre, dynamisme) que l'œil traite en priorité sur les détails locaux [Arnheim, 1974], ce qui justifie l'inclusion dans notre pipeline de descripteurs compositionnels capturant la structure d'ensemble. Ces perceptions sont ensuite encodées en mémoire sous forme de *schémas stylistiques* implicites [Martindale, 1990] : un observateur regroupe intuitivement deux Rothko non parce qu'il verbalise le concept de *Color Field*, mais parce que la similarité de leur structure chromatique et de leur texture lumineuse active les mêmes représentations mnésiques. C'est précisément ce type de similarité (intersubjective, pré-verbale, fondée sur le seul signal visuel) que notre fusion multimodale (CNN, VAE, descripteurs classiques) cherche à capturer et à formaliser.

2.4 Motivation par l'exemple : des features visuelles discriminantes

Avant de décrire notre pipeline, nous illustrons par quelques exemples contrastés pourquoi les propriétés visuelles objectives que nous mesurons constituent des descripteurs pertinents pour organiser ce corpus.

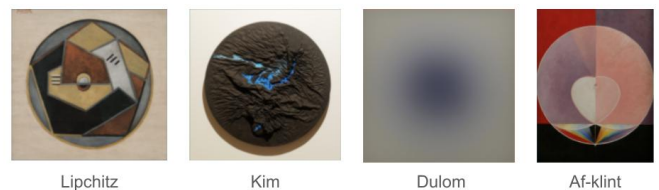


FIGURE 1 – Exemple d'œuvres de peintres différents présentant toutes un cercle centré : intuitivement, on aimerait toutes les classer ensemble.

Ces contrastes illustrent pourquoi une représentation multimodale est nécessaire : aucun descripteur isolé ne suffit à discriminer simultanément toutes ces dimensions. La taille du fichier JPEG distingue Soulages de Pollock mais confond Mondrian et Rothko ; les lignes de Hough séparent Mondrian de Kandinsky mais ignorent la gestualité de Hartung ; le score d'autocorrélation identifie Vasarely et Hantaï mais ne différencie pas leurs palettes respectives. C'est précisément la combinaison de ces descripteurs complémentaires (classiques, CNN, VAE) qui donne à notre pipeline sa capacité discriminante.

2.5 État de l'art computationnel

2.5.1 Descripteurs artisanaux

Les premières approches de classification computationnelle de l'art reposaient sur des descripteurs conçus manuellement. Les descripteurs exploités se regroupent en trois familles [Rathi et al., 2025] : *couleur* (histogrammes RGB/HSV, palettes dominantes par k -means), *texture* (filtres de Gabor, GLCM, LBP) et *composition* (bords Canny/Sobel, transformée de Fourier, symétrie) [Saleh and Elgammal, 2015]. Ces approches réussissent sur les styles les plus contrastés mais échouent sur les catégories finement définies, dont les distinctions reposent sur des critères contextuels inaccessibles au signal visuel brut.

2.5.2 Réseaux convolutifs

L'introduction des CNN profonds a considérablement fait progresser l'extraction automatique de représentations visuelles pour la classification de peintures. Lecoutre et al. [2017] montrent qu'un ResNet50 pré-entraîné sur ImageNet et ré-entraîné partiellement (environ 20 % des couches hautes) sur le corpus WikiArt produit des features nettement plus discriminantes pour le style pictural que les features génériques d'ImageNet : les couches hautes, trop spécialisées en reconnaissance d'objets naturalistes, doivent être adaptées au domaine artistique pour capturer les variations stylistiques pertinentes. Ce résultat motive directement notre choix d'utiliser les activations intermédiaires d'un CNN pré-entraîné comme source de features : elles encodent des structures visuelles de complexité croissante, des textures locales dans les premières couches aux relations compositionnelles d'ensemble dans les couches profondes.

Lelièvre and Neri [2021] apportent un éclairage complémentaire spécifique à l'art abstrait : dans une tâche d'orientation de peintures, les couches profondes du réseau sont bien plus sollicitées pour les œuvres abstraites que pour les œuvres figuratives, ce qui indique que la composition globale (et non les cues locaux) constitue le signal stylistique dominant dans ce corpus. Cette observation justifie notre choix de ne pas se restreindre aux activations de la dernière couche, mais d'en extraire à plusieurs profondeurs du réseau.

2.5.3 Autoencodeurs variationnels comme extracteurs de features

Parallèlement aux approches discriminatives, les autoencodeurs variationnels (VAE) ont émergé comme extracteurs de représentations latentes pour l'art. Contrairement aux CNN entraînés sur des labels, un VAE apprend une distribution probabiliste continue dans un espace latent compact, capturant des variations stylistiques globales sans supervision. Castellano and

Vessio [2020] montrent que de tels espaces latents encodent spontanément des attributs stylistiques à faible granularité, la frontière style/contenu émergeant du modèle et non d'un biais imposé. Notre pipeline exploite directement ces vecteurs latents, fusionnés avec des features CNN et des descripteurs classiques pour obtenir une représentation multimodale plus riche qu'aucune source isolée.

2.5.4 Classification non supervisée

La classification non supervisée de peintures reste nettement moins explorée que la classification supervisée. L'absence de *ground truth* incontestable impose de combiner des métriques internes de cohérence (score de silhouette, indice de Davies-Bouldin) et une validation humaine, deux sources de vérité de nature différente et parfois contradictoires.

Les approches les plus performantes combinent réduction dimensionnelle et algorithmes de clustering : la projection dans un espace de faible dimension préalablement au clustering améliore significativement la qualité des partitions, en éliminant le bruit propre aux espaces de haute dimension et en révélant la structure géométrique globale du corpus [Allaoui et al., 2023]. Dans notre pipeline, cette réduction est assurée par ACP séparée par famille de features, avant fusion pondérée et clustering par K -moyennes.

Un résultat récent [Castellano and Vessio, 2020] montre que des espaces latents appris sans supervision encodent spontanément des attributs stylistiques reconnaissables, confirmant que la structure visuelle de l'art contient suffisamment d'information pour permettre une organisation non triviale sans annotation.

2.5.5 Limites spécifiques à l'art abstrait et implications pour notre démarche

Le survey de Rathi et al. [2025] identifie quatre défis propres à ce corpus, qui constituent autant de contraintes concrètes sur notre pipeline.

Variance intra-classe élevée. Deux œuvres du même artiste peuvent être visuellement plus dissemblables que deux œuvres de styles distincts. Pour nous, cela signifie que la cohérence intra-peintre ne peut pas servir de critère de validation principal : un clustering visuellement pertinent pourrait séparer deux périodes d'un même artiste, et ce serait un résultat légitime. Nous utilisons néanmoins la cohérence intra-peintre comme indicateur secondaire, en gardant à l'esprit cette limite.

Ambiguïté des labels historiques. Les catégories de l'histoire de l'art ont des frontières floues, reconnues comme telles par les experts eux-mêmes. Cela implique qu'une partition nette pro-

duite par notre algorithme ne peut pas être jugée incorrecte au seul motif qu'elle ne coïncide pas avec les découpages historiographiques : notre classification vise des groupes visuellement homogènes, pas nécessairement des équivalents des mouvements historiques.

Problème de l'intention. Les modèles classifient sur des propriétés visuelles exhibées, en ignorant le contexte qui détermine pour une grande part la classification humaine experte. Nous assumons pleinement cette limite : notre objectif n'est pas de reproduire le jugement de l'historien de l'art, mais de produire des groupes cohérents au sens du signal visuel. Cette distinction est au cœur de notre démarche et de son évaluation.

Déséquilibre du corpus. L'art abstrait est sous-représenté dans les grands datasets de référence, ce qui implique que nos features CNN, pré-entraînées sur ImageNet ou WikiArt, ont été optimisées sur des distributions très différentes de notre corpus. C'est l'une des raisons pour lesquelles nous complétons les features CNN par un VAE entraîné directement sur notre corpus : il apprend une représentation spécifique à nos données, sans biais de distribution externe.

2.6 Positionnement et plan

Ce projet se situe à la convergence de toutes ces traditions. Il opère dans l'espace philosophique de Goodman (propriétés exhibées) plutôt que de Danto (contexte institutionnel) ; il reconnaît avec Bourdieu et Becker que les labels de référence sont des constructions sociales situées ; et il propose une ingénierie de *fusion multimodale* (descripteurs bas-niveau, features CNN et représentations probabilistes VAE, combinés par *alpha-blending* pondéré, réduits par PCA et partitionnés par K-Means, complété par un regroupement hiérarchique agglomératif et un filtre de qualité par score de Silhouette) pour proposer une partition émergente indépendante de ces constructions.

3 Dataset et Prétraitement

3.1 Description du corpus d'étude

La base de données exploitée est constituée de 1700 œuvres d'art abstrait. Si ce corpus offre une diversité précieuse pour une approche de classification non supervisée, son analyse quantitative préliminaire révèle plusieurs limites structurelles et biais méthodologiques importants. Il convient de les formaliser afin d'interpréter rigoureusement la topologie des clusters qui en découleront.

3.2 Biais de distribution et anomalies d'étiquetage

On observe en premier lieu un déséquilibre significatif dans la distribution des œuvres par artiste. Alors que certains peintres sont largement surreprésentés au sein du jeu de données, d'autres ne sont documentés que par une unique toile. Dans le cadre d'un *clustering* non supervisé, cette hétérogénéité numérique constitue un biais majeur : le modèle risque de regrouper les œuvres en fonction de la récurrence statistique des caractéristiques propres à un artiste surreprésenté (faible variance intra-artiste), plutôt que sur des critères stylistiques universels et transversaux.

De plus, le *dataset* présente des anomalies d'étiquetage critiques : plusieurs œuvres rigoureusement identiques apparaissent en double sous des attributions d'artistes différentes. Bien que notre démarche algorithmique s'affranchisse explicitement des métadonnées nominales lors de la phase d'apprentissage, ces collisions d'étiquettes compliquent la phase de validation et l'évaluation de la pertinence sémantique des classes obtenues.

3.3 Hétérogénéité des formats et impact géométrique des prétraitements

Une autre contrainte technique majeure réside dans la grande variabilité des formats géométriques d'origine, les œuvres présentant des ratios d'aspect (hauteur/largeur) très hétérogènes. Pour être encodées par des architectures de vision computationnelle standardisées (telles que les CNN), ces images doivent obligatoirement être ramenées à des dimensions matricielles uniformes.

Pour pallier cette divergence sans écraser ni anamorphoser les structures spatiales des œuvres, un *padding* par réflexion a dû être appliqué. D'un point de vue esthétique, ces manipulations altèrent l'intégrité de l'œuvre originale. D'un point de vue computationnel, elles introduisent des artefacts artificiels non négligeables :

- **Artefacts de symétrie :** Le dédoublement ou le miroitement des bordures crée des motifs de symétrie axiale totalement inexistant dans la composition initiale. Cela risque de fausser les descripteurs de texture et d'organisation spatiale (comme l'analyse des contours ou l'entropie locale).
- **Biais de distribution chromatique :** La réplique de zones périphériques surreprésente artificiellement certaines palettes de couleurs au détriment de l'équilibre chromatique initialement voulu par le peintre.

Par conséquent, l'impact de ces choix techniques de prétraitement devra être systématiquement pris en considération lors de l'analyse des critères discriminants du classifieur et de la confrontation des résultats avec notre propre observation humaine.

4 Extraction des features

Afin d'extraire des descripteurs pertinents, nous avons implémenté trois approches inspirées de l'état de l'art. Pour chaque ensemble de caractéristiques, une Analyse en Composantes Principales (ACP) a été appliquée afin de réduire la dimensionnalité.

4.1 Approche géométrique

Caractéristiques de Couleur

La dimension chromatique est essentielle dans l'art abstrait. Les caractéristiques suivantes visent à quantifier l'usage de la couleur, de la luminosité et de la composition chromatique globale :

- **Histogrammes HSV** : Analyse de la distribution des couleurs dans l'espace (Teinte, Saturation, Valeur) permettant de capturer la tonalité dominante et l'intensité des couleurs.
- **Statistiques Lab** : Calcul des moments statistiques (moyenne, écart-type, asymétrie, aplatissement) dans l'espace perceptuel $L^*a^*b^*$, modélisant la perception humaine des contrastes colorés.
- **Palette dominante** : Extraction des couleurs principales et de leurs proportions via l'algorithme K-Means, reflétant la signature chromatique de l'artiste.
- **Luminosité spatiale** : Distribution de la luminance sur une grille spatiale, capturant l'équilibre des clairs-obscur dans la composition.
- **Richesse chromatique** : Mesure de la diversité et de la variété des couleurs employées dans l'œuvre.
- **Métadonnées physiques** : Intégration de la taille du fichier et du ratio d'aspect (format portrait, paysage ou carré), offrant des informations sur le format et la complexité graphique de l'œuvre.

Caractéristiques de Texture

La texture traduit la matière, le coup de pinceau et la qualité de la surface picturale. Les descripteurs suivants ont été utilisés :

- **Matrices de co-occurrence** : Calcul de contraste, corrélation, énergie et homogénéité pour quantifier les relations spatiales entre les niveaux de gris voisins.
- **Motifs binaires locaux** : Analyse multi-échelle de la rugosité locale de la surface, efficace pour détecter les micro-textures.
- **Filtres de Gabor** : Utilisation d'une banque de filtres simulant la réponse du système visuel humain à des fréquences et orientations spatiales spécifiques.
- **Autocorrélation** : Mesure de la répétition et de la périodicité des motifs spatiaux dans l'image.

- **Analyse spectrale** : Identification des fréquences spatiales dominantes et des pics dans le domaine fréquentiel, révélant la rythmique de la texture.
- **Répétition** : Quantification explicite de la régularité et de la répétitivité des motifs picturaux.

Composition et Structure

Ce pilier analyse la composition spatiale, les formes, la complexité structurelle et l'organisation de l'œuvre :

- **Histogramme de Gradients Orientés (HOG)** : Capture de la distribution locale de l'intensité des gradients.
- **Analyse des contours et formes** : Détection des arêtes (filtre de Deriche), extraction des contours, analyse des formes géométriques et détection de lignes droites (transformée de Hough).
- **Analyse par grille** : Étude de la répartition de l'information (symétrie, centralité, gradients) selon une grille spatiale divisant l'image en sous-régions.
- **Symétrie globale** : Mesure du degré de symétrie axiale ou centrale de la composition.
- **Cohérence directionnelle** : Utilisation du tenseur de structure pour évaluer l'alignement et la force des structures locales (ex : un coup de pinceau fort et directionnel).
- **Complexité fractale** : Calcul de la dimension fractale pour estimer la complexité, l'irrégularité et l'auto-similarité des motifs et des contours.
- **Désordre visuel** : Mesure de l'entropie de Shannon pour quantifier l'incertitude et la densité d'information, complétée par l'analyse statistique de la magnitude des gradients.
- **Caractéristiques fréquentielles haut niveau** : Analyse de la pente du spectre de puissance (indiquant la décroissance de l'énergie à haute fréquence) et de l'énergie dans les hautes fréquences, reflétant le niveau de détail fin de la composition.

4.2 Approche par VAE

L'approche par un VAE (Variational Auto-Encoder) nous est vite apparue comme une solution idéale pour la classification d'art abstrait. Deux raisons principales motivent ce choix :

- sans vouloir générer de nouvelles images, un VAE peut nous offrir une visualisation directe des images dans l'espace latent.
- un VAE est une extension non linéaire d'une ACP, ce qui peut donc nous permettre de capter des features (autres que des features géométriques) dans nos images.
- c'est de l'apprentissage non-supervisé : sans label sur nos

images (hormis le nom de l'artiste), un VAE nous permet de regrouper certaines peintures ensemble de manière automatique.

Comme nous voulons que le rendu de notre espace latent nous donne quelque chose de compact et continu (c'est-à-dire sans vide entre nos vecteurs de l'espace latent), nous avons fait le choix d'utiliser un β -VAE afin de pénaliser davantage les erreurs de reconstruction pour que notre espace latent s'apparente au maximum à une loi normale standard.

De même, dans le cas des VAE, l'entraînement peut prendre un certain temps de calcul. C'est pourquoi nous avons fait le choix de redimensionner nos images en format 256×256 pixels.

4.2.1 Construction du β -VAE

Pour la première construction de notre VAE, nous choisissons d'appliquer 5 couches de convolutions à la suite, chacune réduisant de moitié la dimension de nos images. La dernière convolution nous donne en sortie : $8 \times 8 \times 512$, ce qui une fois mis à plat donne des vecteurs de taille 32768.

Comme nous voulons obtenir des vecteurs de taille LATENT-DIM = 128, nous passons une dernière fois ces vecteurs dans des couches linéaires : une couche qui renvoie la **moyenne** μ ainsi que le **log de la variance** $\log(\sigma^2)$.

On utilise l'astuce de reparamétrisation afin de générer un point issu de la loi normale de paramètres (μ, σ^2) .

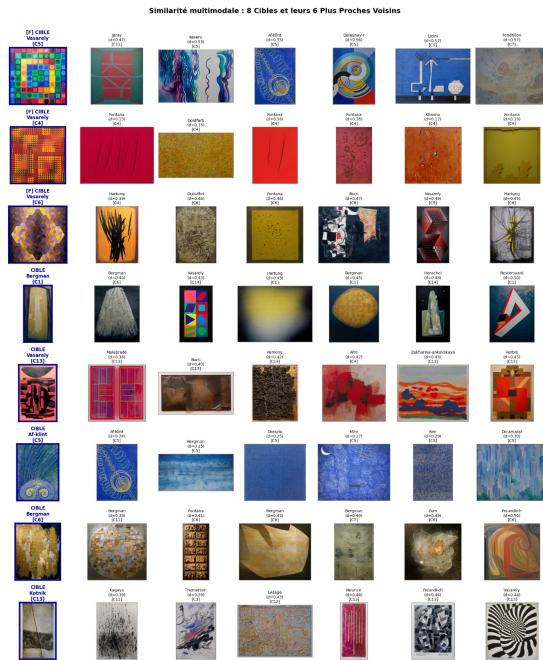


FIGURE 2 – Plus proches voisins de chacune des images. On voit des similarités dans les formes et les couleurs.

On peut ici facilement identifier le premier problème de notre

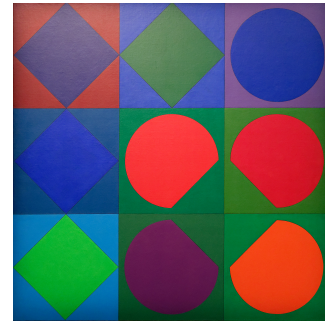
approche par les VAE : les structures géométriques ne sont pas préservées au sein des différents clusters. Notre implémentation cherche plutôt à regrouper les images d'un point de vue de la colorimétrie.

Après analyse de notre code et du fonctionnement d'un VAE, nous pouvons expliquer ce résultat à cause d'une seule chose : la fonction de perte. En effet, pour un VAE classique, la fonction de perte est Pixel-MSE, c'est-à-dire qu'elle va principalement pénaliser les écarts de couleur sur l'image. Si par exemple une tache de couleur se retrouve décalée de quelques pixels sur une autre image, la perte sera grande.

Or, comme nous l'avons dit dans l'introduction, quand on veut classer de l'art abstrait, il faut prendre principalement en compte une abstraction lyrique (axée sur la couleur) et une abstraction géométrique (axée sur les formes). Ainsi nous comprenons que toute une partie des features n'avait pas été prise en compte par le VAE.



(a) (Pollock) - Axé sur la colorimétrie



(b) (Vasarely) - Axé sur les formes

4.2.2 Une meilleure extraction des features

Dans notre VAE actuel, lors de la reconstruction, notre modèle va avoir tendance à générer des images floues. En effet, comme la pénalisation se fait surtout sur les différences de couleur, et non pas sur la structure globale, notre modèle va souvent générer des images floues, avec une couleur unie, afin de minimiser cette perte.

Comme nous voulons éviter cela, nous avons introduit **une nouvelle fonction de perte**, s'appuyant sur un réseau convolutif. Ici nous prenons le modèle VGG, pré-entraîné sur ImageNet, et qui a donc appris à détecter les structures présentes sur une image

(lignes, contours, contrastes, textures, ...). La fonction de perte fonctionne comme suit :

- on fait passer l'image originale dans le réseau VGG ;
- on fait passer l'image générée par le VAE dans le réseau VGG ;
- on calcule la différence entre les deux sorties du réseau VGG.

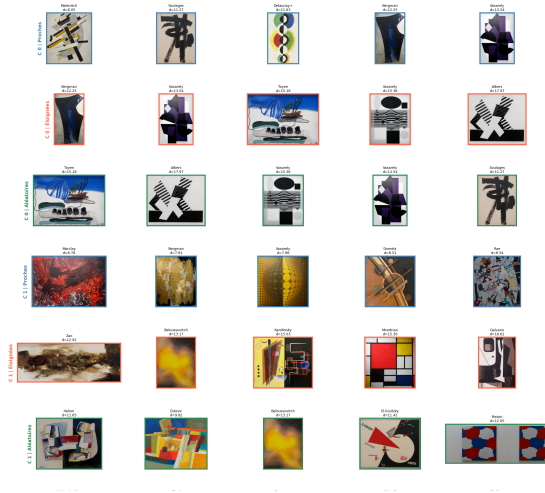


FIGURE 4 – Résultats VAE avec une perte VGG.

Erreur à la dernière epoch	VAE Classique	VAE + VGG
Train	12561.7	6823.2
Validation	12529.2	6551.1
Reconstruction	11540.4	6064.7
KL	255.3	189.6

TABLE 1 – Comparaison des erreurs entre le VAE classique et le VAE avec perte VGG.

4.2.3 Ouverture : utilisation d'un espace siamois

Bien que la fusion des espaces du β -VAE (axé sur les couleurs et le style global) fournisse une base de caractéristiques très riche, cet espace combiné souffre d'une limite majeure : il n'a aucune notion de « sens » artistique ou d'identité d'auteur. Les vecteurs y sont regroupés de manière purement statistique et mathématique.

C'est ici qu'intervient le réseau siamois, dont l'objectif est de métamorphoser cet espace brut en un espace métrique discriminant. Son utilité repose sur trois piliers fondamentaux :

- **Le rapprochement sémantique (raffinage par paires) :** En s'entraînant avec une perte contrastive (*Contrastive Loss*), le réseau siamois apprend explicitement à rapprocher les œuvres d'un même artiste et à repousser celles d'artistes

différents. Il force l'espace latent à s'organiser non plus seulement par « aspect visuel brut », mais par cohérence stylistique d'auteur.

- **La création de clusters ultra-nets :** Comme le met en évidence la visualisation t-SNE finale du projet, l'espace siamois transforme un nuage de points initialement diffus et mélangé (l'espace fusionné) en grappes (clusters) d'artistes extrêmement distinctes et séparées.
- **L'apprentissage *few-shot* (classification d'artistes rares) :** C'est le plus grand point fort de cette approche. Dans un dataset d'art abstrait où de nombreux artistes ne possèdent que deux ou trois œuvres disponibles, un classifieur classique (comme un gros réseau de neurones) échouerait par manque de données. L'espace siamois résout ce problème en permettant une classification par prototypes (méthode du plus proche voisin) : il suffit de calculer la moyenne des vecteurs siamois d'un artiste pour créer sa « signature », permettant d'identifier une nouvelle œuvre avec une précision remarquable, même si l'artiste est très rare.

En résumé, le réseau siamois agit comme le traducteur expert de notre modèle. Il prend des descripteurs visuels bruts pour en faire un espace de sens, indispensable pour l'authentification et la reconnaissance de style artistique. Nous n'avons cependant pas eu la possibilité de le mettre totalement en place avant le rendu du rapport, car c'est une méthode semi-supervisée et qui ne rentre donc pas totalement dans les critères du projet. C'est pourquoi nous n'avons pas de résultats ici.

4.3 Approche par un CNN

L'utilisation de réseaux de neurones convolutifs (CNN) pré-entraînés, bien que standardisée en vision par ordinateur, pose un défi majeur dans le cadre de l'art abstrait. Des architectures classiques entraînées sur ImageNet ont optimisé leurs filtres pour détecter des objets sémantiques. Or, l'art abstrait se définit précisément par l'absence d'objets identifiables. Extraire directement les vecteurs issus des dernières couches *Fully Connected* introduirait un biais sémantique fort : le réseau tenterait de regrouper les toiles en cherchant des illusions d'objets (paréidolie de la machine), plutôt qu'en analysant la technique picturale. Pour contourner ce biais, nous avons implémenté une approche inspirée des travaux de Neural Style Transfer, en nous concentrant exclusivement sur les corrélations de textures.

Le choix architectural : VGG16 plutôt que ResNet

Bien que des architectures plus récentes et profondes comme ResNet dominent aujourd'hui les tâches de classification, nous avons délibérément choisi de baser notre extracteur sur **VGG16**. L'architecture de ResNet repose sur des connexions résiduelles (*skip connections*) qui forcent l'addition des activations de couches peu profondes avec celles des couches profondes. Cette fusion mélange les informations de texture (bas niveau) et les informations sémantiques (haut niveau), rendant l'isolation du style pur mathématiquement instable. À l'inverse, l'architecture séquentielle et stricte de VGG16 permet une hiérarchisation nette des filtres : chaque bloc convolutif encode une échelle de texture précise, sans pollution sémantique résiduelle.

Extraction de la Signature Stylistique par Matrices de Gram

Plutôt que d'utiliser les cartes d'activation brutes, nous extrayons le style de chaque toile via la **Matrice de Gram** des cartes de caractéristiques à différentes profondeurs du réseau. Pour une couche donnée produisant un tenseur F de dimensions $C \times H \times W$ (C canaux, $H \times W$ dimensions spatiales), la matrice de Gram G de taille $C \times C$ est définie par :

$$G_{ij} = \frac{1}{C \cdot H \cdot W} \sum_{k=1}^{H \cdot W} F_{ik} F_{jk}$$

Ce produit scalaire entre les canaux annule l'information spatiale (la localisation du coup de pinceau) pour ne conserver que l'information de corrélation : la nature de la texture et l'interaction des couleurs.

Concrètement, notre classe `VGGStyleExtractor` intercepte les activations à la sortie de quatre blocs convolutifs successifs de VGG16, obtenus en découpant les `features` du réseau aux indices `[ :4 ]`, `[ 4:9 ]`, `[ 9:16 ]` et `[ 16:23 ]` dans l'implémentation PyTorch. Ces points d'extraction correspondent respectivement aux sorties de `conv1_2`, `conv2_2`, `conv3_3` et `conv4_3`. Ce choix couvre un spectre complet allant des micro-textures (grain de la toile, coups de pinceau fins) aux macro-structures (aplats de couleur, organisation chromatique globale). Pour optimiser l'espace mémoire, seul le triangle supérieur de chaque matrice symétrique est conservé via `triu_indices` et aplati en un vecteur unique avant concaténation.

Robustesse par Échantillonnage Spatial

L'art abstrait présente une forte variabilité intra-image : une toile de Pollock n'est pas stylistiquement uniforme d'un coin à l'autre. Pour obtenir une signature robuste, nous avons adopté une stratégie d'extraction par *patches* aléatoires. Pour chaque

œuvre, **10 régions aléatoires** (*random crops*) de taille 256×256 pixels sont extraites, chacune normalisée selon les statistiques ImageNet. Le vecteur de style (concaténation des matrices de Gram des 4 couches) est calculé indépendamment pour chaque patch via `torch.no_grad()`, garantissant une inférence sans mise à jour des poids. La signature finale de l'œuvre correspond à la **moyenne de ces 10 vecteurs**, produisant une représentation invariante à l'échelle et aux compositions décentrées.

Réduction de Dimension et Pipeline de Normalisation

Les vecteurs stylistiques bruts issus de la concaténation des quatre matrices de Gram présentent une dimensionnalité très élevée. Nous avons appliqué le pipeline de réduction suivant, dans cet ordre :

1. **Correction des valeurs manquantes** : remplacement des valeurs NaN éventuelles par la médiane de chaque dimension.
2. **Normalisation L_2** : harmonisation de l'énergie des vecteurs (`Normalizer(norm='l2')`) pour éviter que des contrastes locaux extrêmes ne dominent l'espace.
3. **Analyse en Composantes Principales (ACP)** : projection sur les **50 composantes principales**, retenant l'essentiel de la variance stylistique.
4. **Standardisation** : centrage et réduction (`StandardScaler`) pour garantir un poids équivalent à chaque dimension latente.

Sur cet espace latent de dimension 50, On a essayé d'étudier conjointement la méthode du coude (inertie intra-cluster) et le score de Silhouette afin de retenir un nombre optimal de clusters mais ce fut modifié lors de la fusion des approches.

5 Fusion des approches

Avant toute pondération, il est impératif d'harmoniser les échelles des trois sous-espaces pour éviter que l'un d'eux ne domine le calcul des distances par sa seule variance brute. Chaque matrice subit d'abord une normalisation L_2 (projetant les vecteurs sur une hypersphère unitaire), suivie d'une standardisation (centrage-réduction pour imposer une moyenne nulle et une variance unitaire). C'est uniquement sur ces espaces harmonisés que les poids sémantiques prennent leur sens.

Les trois espaces réduits sont ensuite concaténés après pondération (voir Équation 1) :

$$\mathbf{X}_{\text{final}} = \left[\tilde{w}_{\text{cl}} \cdot \mathbf{X}_{\text{cl}} \parallel \tilde{w}_{\text{CNN}} \cdot \mathbf{X}_{\text{CNN}} \parallel \tilde{w}_{\text{VAE}} \cdot \mathbf{X}_{\text{VAE}} \right] \in \mathbb{R}^{n \times 3d} \quad (1)$$

où les poids normalisés sont définis par :

$$\tilde{w}_\star = \frac{w_\star}{w_{\text{CNN}} + w_{\text{VAE}} + w_{\text{cl}}} \quad (2)$$

La multiplication par le poids avant concaténation est équivalente à une mise à l'échelle de chaque bloc de colonnes : une source avec un poids deux fois plus élevé contribue deux fois plus à la norme euclidienne de $\mathbf{X}_{\text{final}}$, et donc deux fois plus à l'inertie du K -moyennes.

Cette formulation présente deux avantages pratiques. D'une part, les poids sont des entiers relatifs ($w_{\text{CNN}}, w_{\text{VAE}}, w_{\text{cl}} \in \mathbb{N}^3$), ce qui rend la grille de recherche finie et interprétable. D'autre part, la dimension de $\mathbf{X}_{\text{final}}$ reste constante (3d) quelle que soit la configuration de poids, ce qui garantit la comparabilité des scores de silhouette entre configurations.

5.1 Validation topologique de l'espace unifié

Avant d'appliquer un algorithme de partitionnement discret tel que le K -moyennes, il est indispensable de vérifier si l'espace continu $\mathbf{X}_{\text{final}}$ code de manière convaincante une sémantique stylistique cohérente. En écho aux questionnements sur les « critères de décision de la machine », nous avons calculé la matrice de leurs distances cosinus inter-artistes.

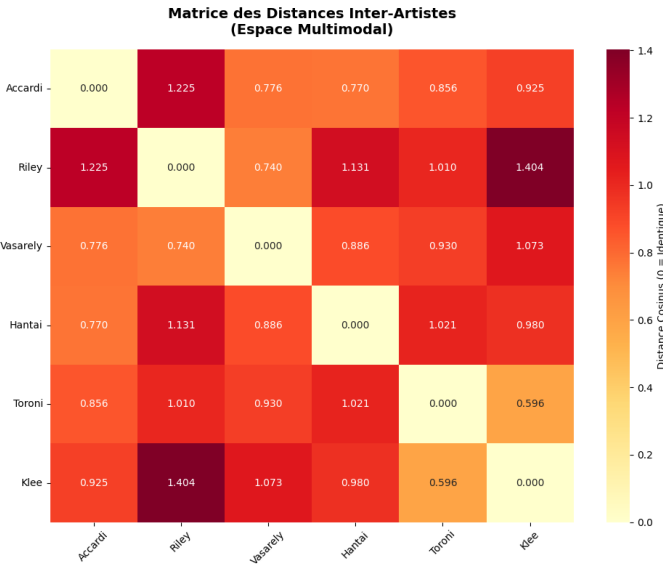


FIGURE 5 – Matrice des distances cosinus entre les centroïdes des artistes de référence dans l'espace fusionné $\mathbf{X}_{\text{final}}$. Une distance proche de 0 indique une forte similarité stylistique.

L'analyse de cette matrice (Figure 5) valide notre pipeline de fusion multimodale, car les rapprochements mathématiques épousent fidèlement l'histoire de l'art :

— **Le pôle de l'art optique et géométrique :** Bridget Riley et

Victor Vasarely présentent la distance cosinus la plus faible de la matrice (0.776). Notre espace a parfaitement capté l'essence de l'Op Art (rigueur géométrique, contraste des arêtes, récurrence des motifs) et les associe naturellement.

— **La rupture avec l'abstraction lyrique :** À l'inverse, ces mêmes artistes géométriques (Riley, Vasarely) affichent des distances extrêmes (allant jusqu'à 0.1.404) avec Klee.

Cette étape confirme que notre super-vecteur ne se contente pas d'empiler des caractéristiques, mais qu'il a organisé l'art abstrait en un *continuum* topologique valide. Le rôle du K -moyennes, détaillé dans la section suivante, sera de discrétiser cet espace continu en mouvements distincts.

5.2 Classification par K -moyennes

L'algorithme des K -moyennes minimise l'inertie intra-cluster :

$$\mathcal{J}(C_1, \dots, C_K) = \sum_{k=1}^K \sum_{\mathbf{x}_i \in C_k} \|\mathbf{x}_i - \boldsymbol{\mu}_k\|_2^2 \quad (3)$$

où $\boldsymbol{\mu}_k = \frac{1}{|C_k|} \sum_{\mathbf{x}_i \in C_k} \mathbf{x}_i$ est le centroïde du cluster k . Nous utilisons une initialisation k-means++ avec $n_{\text{init}} = 10$ répétitions pour limiter la sensibilité à l'initialisation aléatoire.

Les distances au centroïde fournies par la méthode transform de sklearn sont utilisées pour ordonner les images au sein de chaque cluster : une distance faible indique une image proche du prototype stylistique du cluster, une distance élevée indique un cas limite.

5.3 Recherche des hyperparamètres

Le pipeline comporte deux familles d'hyperparamètres : les poids de combinaison et le nombre de clusters K . Leur recherche suit un protocole en deux étapes séquentielles, puis une validation visuelle finale.

Étape 1 — Calibration des poids à K fixé

Les poids ($w_{\text{CNN}}, w_{\text{VAE}}, w_{\text{cl}}$) sont explorés empiriquement. Faute de vérité terrain absolue, l'optimisation des hyperparamètres a été conduite par une évaluation heuristique itérative. L'inspection de l'espace des plus proches voisins pour des requêtes complexes a permis d'ajuster les pondérations. L'équilibre optimal a été atteint avec la configuration ($w_{\text{CNN}} = 0.6, w_{\text{VAE}} = 0.2, w_{\text{cl}} = 0.2$).

Étape 2 — Sélection de K à poids fixés

Les poids optimaux étant fixés, le nombre de clusters est exploré sur la plage $K \in \{3, \dots, 50\}$. Pour chaque valeur de K , la même méthode que précédemment a été employée (en affichant en plus les classes). Nous avons essayé de la compléter par

l'**inertie** qui permet d'appliquer la méthode du coude.

Finalement, le résultat n'était pas convaincant (très variable selon la combinaison de poids précédente). Nous avons donc opté pour une méthode plus manuelle où nous observions, pour chaque cluster, les images les plus proches du centroïde, les images les plus éloignées et quelques images aléatoires. Nous avons trouvé que le résultat était convaincant à partir d'un nombre de clusters $K = 20$, et qu'il ne s'améliorait pas vraiment en augmentant K (cela créait des classes superflues et/ou hétérogènes au lieu de séparer des classes qui combinaient clairement deux classes séparables à l'œil nu).

Pour valider ce K , nous avons aussi inspecté les plus proches voisins en distance cosinus d'un ensemble d'images de référence correspondant aux artistes désignés par le jury : Accardi, Riley, Vasarely, Hantaï, Toroni et Klee.

$$d_{\cos}(\mathbf{u}, \mathbf{v}) = 1 - \frac{\mathbf{u}^\top \mathbf{v}}{\|\mathbf{u}\|_2 \|\mathbf{v}\|_2} \quad (4)$$

La distance cosinus est préférée à la distance euclidienne car elle est insensible à l'amplitude absolue des vecteurs : seule la direction dans l'espace des features importe, ce qui est cohérent avec une comparaison de profils stylistiques plutôt que d'intensités brutes. Le choix de la distance cosinus pour l'évaluation est par ailleurs mathématiquement cohérent avec notre partitionnement par K -moyennes : nos vecteurs ayant préalablement subi une normalisation L_2 , la minimisation de la distance euclidienne (au cœur du K -moyennes) est formellement équivalente à la maximisation de la similarité cosinus.

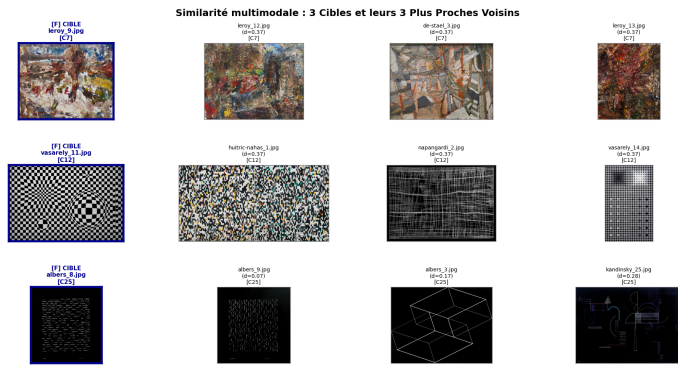


FIGURE 6 – Validation de l'espace de fusion via l'extraction des plus proches voisins. Les requêtes ciblent délibérément des extrêmes stylistiques (matérialité extrême, instabilité optique, minimalisme géométrique).

Comme l'illustre la Figure 6, la représentation finale parvient à capturer des nuances stylistiques complexes sans se laisser piéger par le bruit visuel. Ce panel d'artistes a été spécifiquement conçu

pour éprouver l'algorithme face à différents défis : la texture chaotique de Leroy, les hautes fréquences spatiales de Vasarely et les aplats stricts d'Albers. Enfin, l'alignement quasi systématique entre les plus proches voisins et les labels de clusters $[C_k]$ prouve que la continuité de notre espace de similarité se traduit fidèlement dans notre classification discrète.

5.4 Diagnostic de l'hétérogénéité des clusters

Malgré la pertinence globale de la partition obtenue, un audit visuel expert des clusters a mis en évidence une limite fondamentale de l'algorithme des K -moyennes appliqué à l'art abstrait. L'algorithme suppose implicitement que les clusters sont sphériques, isotropes et de densités comparables. Or, l'histoire de l'art révèle une topologie bien différente : certains mouvements (comme l'abstraction géométrique ou l'Op Art) sont extrêmement normés et forment des sous-espaces denses, tandis que d'autres (comme l'Action Painting ou l'Art Informel) présentent une variance intra-classe considérable.

Cette asymétrie topologique se traduit par la création de classes « fourre-tout ». Les œuvres situées aux frontières de ces mouvements à forte variance se retrouvent mathématiquement absorbées par des clusters dont elles ne partagent pas l'essence stylistique, dégradant ainsi la cohérence visuelle de la classe. Pour objectiver ce phénomène, nous nous sommes appuyés sur le score de silhouette individuel $s(\mathbf{x}_i) \in [-1, 1]$, qui évalue l'ancrage d'une œuvre dans son propre cluster par rapport au cluster voisin le plus proche.

5.5 Écrémage adaptatif par analyse du noyau dur

Pour purifier les clusters hétérogènes sans amputer les clusters naturellement denses, un filtrage par seuil global s'est révélé inefficace. Nous avons donc implémenté une méthode d'écrémage adaptatif fondée sur l'analyse du *noyau dur* (Core Silhouette) de chaque cluster.

Pour un cluster donné C_k , l'objectif est de définir un seuil d'exclusion dicté uniquement par ses représentants les plus caractéristiques, ignorant ainsi le bruit périphérique qui tirerait la moyenne vers le bas. Soit $S_k = \{s(\mathbf{x}_i) \mid \mathbf{x}_i \in C_k\}$ l'ensemble des scores de silhouette du cluster k . Nous définissons le noyau $\mathcal{N}_k \subset C_k$ comme la moitié supérieure de la distribution :

$$\mathcal{N}_k = \{\mathbf{x}_i \in C_k \mid s(\mathbf{x}_i) \geq \text{médiane}(S_k)\} \quad (5)$$

Nous calculons ensuite la moyenne $\mu_{\mathcal{N}_k}$ et l'écart-type $\sigma_{\mathcal{N}_k}$ des scores strictement sur ce noyau. Le seuil de tolérance τ_k pour le cluster k est alors défini par un couperet statistique, sécurisé par un plancher paramétrable $\tau_{\min}$:

$$\tau_k = \max(\tau_{\min}, \mu_{\mathcal{N}_k} - \lambda \cdot \sigma_{\mathcal{N}_k}) \quad (6)$$

où λ est un hyperparamètre de sévérité (fixé empiriquement à 2.5) et $\tau_{\min}$ est un paramètre ajustable définissant le seuil de silhouette minimal toléré. Toute œuvre $\mathbf{x}_i \in C_k$ vérifiant $s(\mathbf{x}_i) < \tau_k$ est rejetée de sa classe initiale et réaffectée à un ensemble global d'outliers, noté C_{-1} . Cette méthode a permis de restaurer une pureté visuelle exceptionnelle au cœur de nos K clusters principaux.

5.6 Classification récursive des marginaux

L'ensemble C_{-1} regroupe les œuvres rejetées lors de l'écrémage. Cependant, postuler que cet ensemble n'est constitué que de « bruit » absolu serait une erreur analytique. Au regard de la sociologie de l'art, ces marginaux peuvent représenter des sous-mouvements minoritaires, des écoles de transition, ou des fusions de styles qui ont été statistiquement écrasés par les mouvements majeurs lors du premier passage du K -moyennes.

Pour vérifier cette hypothèse, nous avons isolé la sous-matrice $\mathbf{X}_{\text{marginaux}} \subset \mathbf{X}_{\text{final}}$ correspondant aux œuvres de C_{-1} , et appliqué un sous-clustering indépendant afin d'y rechercher de nouvelles topologies. Cette approche récursive a permis de révéler K' nouvelles sous-classes (notées $A.0, A.1, \dots$). L'émergence de regroupements logiques parmi ces marginaux (par exemple, la réunion de toiles de Dubuffet et d'Alechinsky initialement dispersées) confirme que le rejet initial était bien le symptôme de ponts stylistiques inexplorés plutôt que d'une simple erreur du modèle.

5.7 Bilan : Matrice de répartition stylistique unifiée

L'intégration des clusters épurés et des nouvelles sous-classes aboutit à une partition finale. Afin d'évaluer dans quelle mesure cette taxonomie parvient à identifier les « signatures » des artistes, nous avons généré une matrice de répartition stylistique (Figure 7).

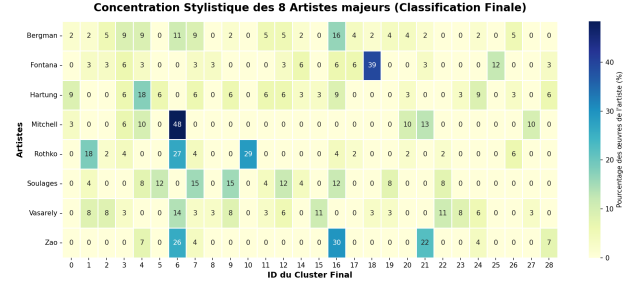


FIGURE 7 – Concentration stylistique des 8 artistes majeurs. Les fortes concentrations valident la capacité du modèle à isoler l'identité visuelle des auteurs.

L'analyse de ce support valide notre pipeline : les cellules présentant de fortes concentrations indiquent que le modèle a réussi à isoler l'identité visuelle d'un artiste de manière aveugle, s'affranchissant ainsi de toute métadonnée.

6 Visualisation des résultats

6.1 Projection dans un espace 2D par PaCMAP

Une fois la partition finale obtenue, la visualisation des K_{total} clusters dans un espace de faible dimension constitue une étape essentielle pour évaluer qualitativement la cohérence de notre pipeline. La projection directe de $\mathbf{X}_{\text{final}} \in \mathbb{R}^{n \times 3d}$ étant impossible à interpréter visuellement, nous recourons à une réduction dimensionnelle vers $\mathbb{R}^2$.

Nous avons retenu **PaCMAP** (*Pairwise Controlled Manifold Approximation* [Wang et al., 2021]) plutôt que le t-SNE classique. Le t-SNE optimise exclusivement la structure locale, les voisins proches dans l'espace d'origine restent proches dans la projection, mais déforme systématiquement les distances inter-clusters : deux groupes éloignés sur la carte t-SNE ne sont pas nécessairement éloignés dans l'espace réel. PaCMAP corrige cette limite en équilibrant explicitement trois types de paires lors de l'optimisation :

- **Paires proches** : maintien de la cohérence locale au sein de chaque cluster.
- **Paires à distance intermédiaire** : préservation de la structure mésoscopique entre clusters voisins.
- **Paires lointaines** : conservation de la topologie globale du corpus.

Cette propriété est particulièrement précieuse dans notre contexte : la position relative des clusters sur la carte reflète fidèlement les distances cosinus calculées dans $\mathbf{X}_{\text{final}}$, validées à la section 5.1. Un observateur peut ainsi interpréter la proximité spatiale de deux clusters non seulement comme une similarité locale, mais comme une vraie continuité stylistique dans l'espace des features.

Les coordonnées 2D produites par PaCMAP sont ensuite normalisées linéairement vers l'intervalle $[-60, 60]$ pour assurer la compatibilité avec le moteur de rendu interactif (voir section 6.2).

6.2 AbstractViz : plateforme de visualisation interactive

Les résultats de la classification sont exposés via **AbstractViz**¹, une application web interactive développée spécifiquement pour ce projet, reposant sur **deck.gl** (rendu WebGL via une vue orthographique) côté client et **Flask** côté serveur.

Architecture technique

Chaque œuvre est représentée par un point dans le plan PaCMAP, coloré selon son cluster ou selon son artiste au choix de l'utilisateur. Les 1 700 miniatures sont encodées en base64 côté serveur et transmises au client lors du chargement initial ; elles se substituent aux points à partir d'un seuil de zoom, offrant une transition fluide entre vue d'ensemble et inspection individuelle. Un mécanisme de *bloom* disperse localement les points superposés à fort grossissement, permettant la sélection précise d'œuvres dans les zones denses.

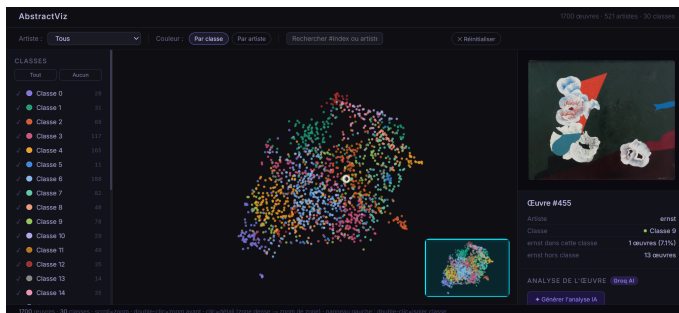


FIGURE 8 – Carte 2D de la visualisation AbstractViz.

Apport pour l'évaluation

AbstractViz joue un double rôle dans notre démarche. D'une part, il constitue l'outil de **validation visuelle** décrit à la section 5.3 : l'inspection des voisins et de la composition des clusters a directement guidé la sélection des hyperparamètres K et λ . D'autre part, il permet une **confrontation humaine** de la taxonomie obtenue : un observateur non expert peut naviguer librement dans le corpus, vérifier que les œuvres d'un même cluster partagent une cohérence perceptible, et identifier les cas limites révélateurs des frontières stylistiques.

7 Analyse de la classification finale

Vous pouvez trouver cette classification sur le site <https://perso.telecom-paristech.fr/edionnet-24/> via AbstractViz, ou regarder la classification complète ici : https://github.com/blastoncrush/projet-im06/tree/main/classification_finale.

Vous avez sur le lien ci-joint : <https://share.rezel.net/s/BP3zjVKp> une image qui présente, pour chaque cluster (identifié de C0 à C19 pour les clusters de base et de A.0 à A.9 pour ceux ajoutés pour les marginaux), en ligne : les 5 images les plus proches du centroïde, les 5 plus éloignées et 5 images aléatoires du cluster.

Voici l'identification cluster par cluster :

- **C0** · 28 œuvres : palette dominante bleue et jaune, mouvement horizontal. Des Dubuffet sont présents : bien qu'ils ont l'air à part, on peut comprendre leur présence de par leur mouvement et leur palette.
- **C1** · 31 œuvres : grosse forme centrale, rectangulaire pour la plupart.
- **C2** · 68 œuvres : rond centré. Le Lee et les Gorin en bordure sont compréhensibles (forme proche d'un rond).
- **C3** · 117 œuvres : palette bleue et jaune, mouvement présent mais contrôlé
- **C4** · 165 œuvres : mouvement et texture (incarné par Leroy), palette terne/sombre
- **C5** · 11 œuvres : peu d'œuvres, toutes majoritairement noires avec quelques formes linéaires blanches ou orangées.
- **C6** · 188 œuvres : palette de couleur riche et saturée, mouvement (incarné par Mitchell).
- **C7** · 82 œuvres : mouvement vertical, palette bleue et jaune.
- **C8** · 48 œuvres : petits carrés et rectangles (exclusivement) de couleurs variées.
- **C9** · 78 œuvres : carrés plus petits (comme Klee). On aurait aimé garder que des œuvres où les carrés sont bien ordonnés, mais le tableau Leger-f présent au centre introduit la présence de tableaux à motifs (Hartung par exemple).
- **C10** · 29 œuvres : empilement vertical de rectangles monochromes (incarné par Rothko).
- **C11** · 49 œuvres : sujet central composé de lignes sur fond blanc.
- **C12** · 35 œuvres : Effet d'éléments alignés horizontalement, souvent des lignes. Avec principalement des couleurs chaudes.
- **C13** · 14 œuvres : Une ligne verticale centrale souvent unique objet du tableau plongé dans un fond de couleur quasi uniforme.
- **C14** · 35 œuvres : sujet central et fond de vignettage
- **C15** · 84 œuvres : grande dynamique de mouvement (in-

1. <https://perso.telecom-paristech.fr/edionnet-24/>

carné par Pollock). Des tableaux chargés de petits motifs (comme certains Vasarely) sont aussi présents, ce qui est moins convaincant mais compréhensible (ils sont présents à la frontière du cluster). Une forme de symétrie grossière est aussi présente.

- **C16** · 121 œuvres : Motifs horizontaux avec une palette froide notamment bleue et des coups de pinceaux clairement visibles
- **C17** · 56 œuvres : 2 ou 3 formes géométriques avec un rendu très uniforme
- **C18** · 108 œuvres : Traits peu appuyés avec la présence de certaines formes géométriques (ce qui a attiré le vasarely à la frontière)
- **C19** · 29 œuvres : Polygones verticaux et fins
- **A0** · 52 œuvres : Oeuvre fractale avec de fortes formes géométriques
- **A1** · 47 œuvres : Couleurs froides pour faire des traits erratiques mais horizontaux avec des motifs carrés
- **A2** · 24 œuvres : Carrés et rectangles imbriqués sur toute l'oeuvre
- **A3** · 33 œuvres : COuleurs chaudes avec de petits motifs géométriques répétés et chargés
- **A4** · 44 œuvres : Motifs chaotiques avec une palette à majorité sombre et chargés en forme petite et complexe
- **A5** · 21 œuvres : Motifs verticaux avec des couleurs sombres
- **A6** · 24 œuvres : Demi-cercles et carrés avec un sujet au centre.
- **A7** · 38 œuvres : mouvement de pinceaux chaotiques avec la plupart des couleurs et des formes géométriques à très hautes fréquences.
- **A8** · 32 œuvres : Coups de pinceaux peu appuyés sur fond blanc avec des lignes bien marquées
- **A9** · 9 œuvres : Polygones rouges, noirs ou gris sur fond blanc

8 Conclusion

Ce projet s'est donné pour ambition de relever un défi à la frontière de la vision par ordinateur et de la philosophie de l'art : classifier un corpus complexe de 1700 peintures abstraites de manière entièrement non supervisée, en se fondant exclusivement sur le signal visuel brut et en faisant abstraction de toute méta-donnée contextuelle ou biographique Goodman [1968], Danto [1964].

Nos expérimentations ont mis en évidence l'insuffisance des

approches isolées pour capturer la richesse sémantique de l'abstraction. Si un VAE classique équipé d'une perte Pixel-MSE tend à privilégier une organisation purement colorimétrique au détriment des structures géométriques, l'introduction d'une perte basée sur le réseau VGG16 a permis de pallier ce biais en forçant le modèle à encoder des caractéristiques de haut niveau, comme les textures et les micro-structures. En harmonisant et en fusionnant ces représentations probabilistes avec les matrices de Gram de textures issues d'un CNN séquentiel et des descripteurs géométriques artisanaux, notre pipeline de fusion multimodale par *alpha-blending* pondéré a réussi à formaliser un espace continu doté d'une véritable cohérence stylistique.

Face aux problématiques de ce corpus (variance intra-classe, déséquilibre numérique, artefacts de prétraitement), les partitions obtenues démontrent la pertinence de notre approche. Si certains clusters émergents épousent fidèlement les catégories historiques validant ainsi l'existence de signatures visuelles propres à certains mouvements les écarts observés sont tout aussi instructifs. Ils démontrent que la machine s'affranchit des constructions sociales de l'histoire de l'art pour regrouper les œuvres selon des critères formels stricts, là où la classification humaine repose souvent sur l'intention invisible de l'artiste.

Perspectives

Bien que l'espace unifié fusionné offre une excellente base statistique, l'absence de labels lors du partitionnement limite sa capacité à regrouper finement les œuvres selon une signature d'auteur ou une parenté stylistique subtile. Une avancée majeure de ce travail réside dans la conception d'un prolongement méthodologique reposant sur un **réseau Siamois**.

En exploitant une perte contrastive (*Contrastive Loss*), ce réseau permet de métamorphoser notre espace brut en un espace métrique hautement discriminant. Bien que sa mise en œuvre complète dépasse le cadre des résultats quantitatifs immédiats de ce rapport, les fondements théoriques posés ouvrent trois perspectives capitalises pour l'analyse computationnelle de l'art abstrait :

- **Raffinage sémantique par paires** : Forcer l'espace latent à réorganiser ses distances pour rapprocher les œuvres de cohérence stylistique similaire et repousser les styles antagonistes.
- **Clustering ultra-net** : Transformer le nuage de points continu et parfois diffus de l'espace fusionné en grappes d'artistes nettement isolées, comme le suggèrent nos premières projections t-SNE.
- **Apprentissage Few-Shot par prototypes** : Résoudre le problème critique des artistes rares (ceux ne disposant que

d'une ou deux toiles dans le dataset) en calculant des vecteurs de signatures « prototypes » capables d'identifier le style d'un peintre à partir d'un nombre extrêmement restreint d'exemples.

En définitive, ce projet démontre que si la machine ne peut capturer l'« intention » invisible de l'artiste au sens de Danto, elle s'avère un outil puissant pour révéler la structure formelle de l'art abstrait, ouvrant la voie à des systèmes d'authentification et de recommandation affranchis des biais historiographiques traditionnels.

Références

- Mebarka Allaoui et al. Considerably improving clustering algorithms using UMAP. *Lecture Notes in Computer Science*, 2023.
- Rudolf Arnheim. *Art and Visual Perception : A Psychology of the Creative Eye*. University of California Press, 1974.
- Daniel E. Berlyne. *Aesthetics and Psychobiology*. Appleton-Century-Crofts, 1971.
- Giovanna Castellano and Gennaro Vessio. Deep convolutional embedding for digitized painting clustering. *arXiv :2003.08597*, 2020.
- Arthur Danto. The artworld. *Journal of Philosophy*, 61(19), 1964.
- Nelson Goodman. *Languages of Art*. Bobbs-Merrill, 1968.
- Adrian Lecoutre, Benjamin Negrevergne, and Florian Yger. Recognizing art style automatically in painting with deep learning. In *Asian Conference on Machine Learning*, pages 327–342, 2017.
- Helmut Leder, Benno Belke, Axel Oeberst, and Dorothee Augustin. A model of aesthetic appreciation and aesthetic judgments. *British Journal of Psychology*, 95 :489–508, 2004.
- Pierre Lelièvre and Peter Neri. A deep-learning framework for human perception of abstract art composition. *Journal of Vision*, 21(5) :9, 2021.
- Colin Martindale. *The Clockwork Muse : The Predictability of Artistic Change*. Basic Books, 1990.
- Vaibhav Rathi, Ankit Sharma, et al. A survey of computational techniques for fine art painting classification. *Image and Vision Computing*, 161 :105626, 2025.
- Eleanor Rosch. Principles of categorization. In Eleanor Rosch and Barbara B. Lloyd, editors, *Cognition and Categorization*, pages 27–48. Lawrence Erlbaum Associates, 1978.
- Babak Saleh and Ahmed Elgammal. A unified framework for painting classification. In *IEEE ICDMW*, pages 1254–1261, 2015.
- Yingfan Wang, Haiyang Huang, Cynthia Rudin, and Yaron Shaposhnik. Understanding how dimension reduction tools work. *Journal of Machine Learning Research*, 22(201) :1–73, 2021.
- Morris Weitz. The role of theory in aesthetics. *Journal of Aesthetics and Art Criticism*, 15(1) :27–35, 1956.
- Semir Zeki. *Inner Vision : An Exploration of Art and the Brain*. Oxford University Press, 1999.