

# Classification non supervisée d'œuvres d'art par style

## Rapport de projet d'IMA 206

Ethan Lazimy, Titouan Duhazé, Noé Mosca

27 juin 2025



## Table des matières

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Extraction des descripteurs visuels</b>	<b>2</b>
2.1	Features de couleur . . . . .	2
2.2	Features de texture . . . . .	2
2.3	Features thématiques . . . . .	4
<b>3</b>	<b>Représentations profondes</b>	<b>5</b>
3.1	Descripteurs visuels via ResNet50 . . . . .	5
3.2	Affinage des vecteurs CLIP . . . . .	5
3.3	Fusion des descripteurs . . . . .	5
<b>4</b>	<b>Réduction de dimension</b>	<b>5</b>
4.1	Application de l'ACP . . . . .	5
4.2	Justification du seuil de 1000 dimensions . . . . .	6
<b>5</b>	<b>Clustering non supervisé</b>	<b>7</b>
5.1	Méthode K-means . . . . .	7
5.2	Affinage des clusters et labellisation . . . . .	7
5.3	Optimisation des attributions . . . . .	8
<b>6</b>	<b>Analyse des résultats</b>	<b>10</b>
6.1	Clusters représentatifs . . . . .	10
6.2	Clusters problématiques . . . . .	12
6.3	Évaluation globale et interprétations . . . . .	12
<b>7</b>	<b>Conclusion</b>	<b>13</b>

# 1 Introduction

Dans le cadre du cours IMA206, nous avons choisi pour projet d'étudier l'application de méthodes d'apprentissage non supervisées afin de classifier des œuvres picturales selon leur style. L'objectif était de regrouper des peintures numériques sans étiquettes préalables, en exploitant uniquement leurs caractéristiques visuelles et sémantiques, regroupées par auteurs. Nous avions comme seule information la date de naissance de chaque auteur.

Nous avons d'abord extrait des descripteurs manuels (couleurs en HSV, textures via GLCM, thématiques avec CLIP) afin de caractériser chaque image selon des critères visuels simples. Cette approche a ensuite été enrichie par des représentations profondes extraites avec ResNet50 et CLIP, permettant de capturer l'apparence globale et le contenu sémantique des œuvres.

L'ensemble des descripteurs a été concaténé puis réduit à 1000 dimensions par ACP, en conservant plus de 90 % de la variance. Un clustering K-means a été appliqué dans cet espace pour regrouper les peintures, puis affiné à l'aide d'une labellisation par style à partir des artistes dominants. Des variantes probabilistes (softmax, pénalisation) ont permis d'améliorer les attributions dans les cas ambigus.

Ce rapport présente notre méthodologie mise en œuvre, ainsi que les résultats que nous obtenus, et les pistes explorées pour dépasser les limites d'une classification stricte en clusters disjoints.

## 2 Extraction des descripteurs visuels

### 2.1 Features de couleur

Dans un premier temps, nous avons adopté l'extraction de features de couleurs, qui sont des critères centraux dans la caractérisation stylistique d'une peinture. Pour extraire ces informations, nous avons utilisé deux types de descripteurs complémentaires sur chaque image, convertie préalablement dans l'espace de couleur HSV (Hue, Saturation, Valeur), plus adapté que le RGB pour capter la perception humaine.

**Statistiques globales.** Nous avons d'abord calculé, sur l'ensemble des pixels, la moyenne et l'écart-type des trois composantes HSV. Ces six valeurs permettent de quantifier la teinte dominante, la saturation globale et la luminosité moyenne d'une œuvre. Elles offrent une première mesure simple de l'ambiance colorée, utile pour distinguer des styles visuellement contrastés (ex. : couleurs pastel contre couleurs vives).

**Couleurs dominantes par K-means.** Nous avons ensuite extrait les trois principales teintes d'une œuvre à l'aide d'un clustering K-means appliqué aux pixels dans l'espace HSV. Chaque cluster correspond à une couleur dominante, représentée par son centre et sa proportion dans l'image. Ce procédé permet de résumer visuellement une peinture sous forme de palette colorée, avec pour chaque peinture, les K couleurs les plus importantes.

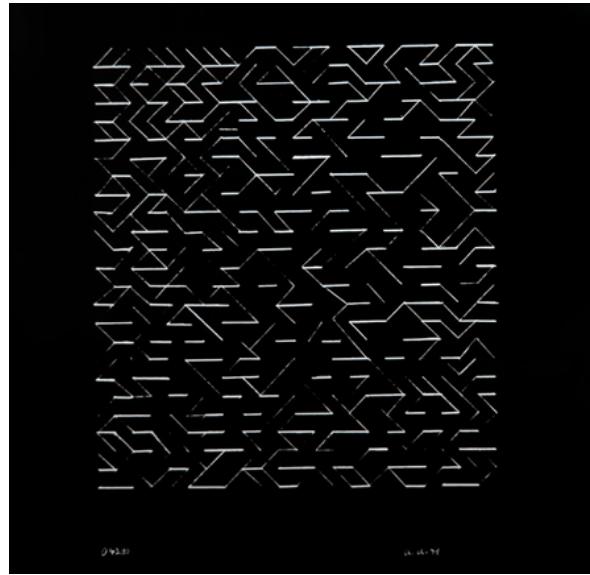
**Utilité.** Ces features sont peu coûteuses à calculer et facilement interprétables. Bien qu'elles ne capturent pas la disposition spatiale des couleurs, elles fournissent des indices précieux pour caractériser des styles riches en contrastes ou en harmonies chromatiques. Ces descripteurs seront combinés à d'autres types de features (textures, sémantiques, profonds) dans les étapes suivantes.

### 2.2 Features de texture

Au-delà des couleurs, la texture d'une peinture constitue un élément essentiel pour différencier les styles : par exemple, le pointillisme de Seurat ou la touche épaisse de Van Gogh. Pour extraire



(a) Delaunay



(b) Albers

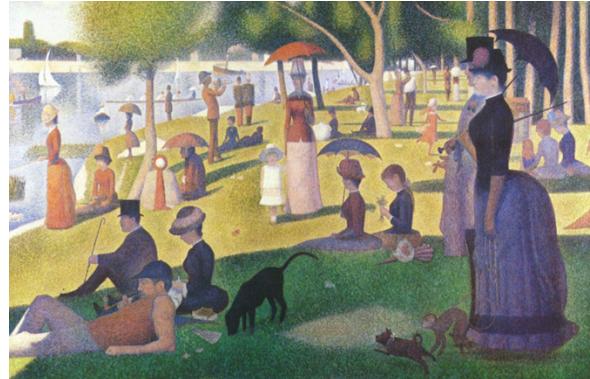
FIGURE 1 – Exemples de peintures avec des couleurs très différentes

des descripteurs texturaux robustes, nous avons utilisé la matrice de cooccurrence des niveaux de gris (GLCM — Gray Level Co-occurrence Matrix), une méthode classique mais efficace.

**Principe.** La GLCM mesure comment les niveaux de gris varient localement dans une image. Elle encode la fréquence des paires de pixels ayant des intensités spécifiques, séparées par une certaine distance et direction. Nous avons calculé cette matrice pour plusieurs angles ( $0^\circ$ ,  $45^\circ$ ,  $90^\circ$ ,  $135^\circ$ ) et une distance fixe.



(a) Van Gogh



(b) Seurat

FIGURE 2 – Exemples de peintures avec des textures très différentes

**Descripteurs extraits.** À partir de la GLCM, nous avons extrait six mesures standards :

- **Contraste** : intensité des variations locales (rugosité) ;
- **Dissimilitude** : différence moyenne entre pixels voisins ;
- **Homogénéité** : régularité des motifs ;
- **Énergie** : répétitivité de la texture ;

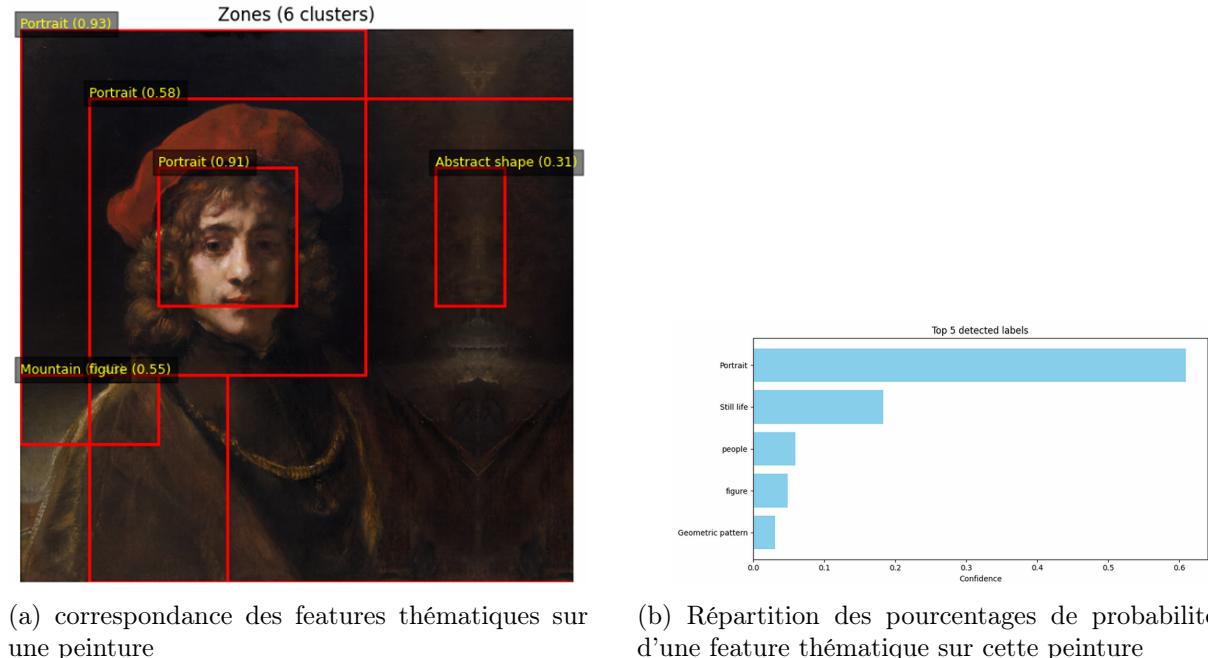
- **Corrélation** : dépendance linéaire entre intensités voisines ;
- **ASM (Angular Second Moment)** : stabilité de la distribution.

**Utilité.** Ces descripteurs permettent de quantifier le style pictural à un niveau plus fin que la couleur : ils captent la densité du trait, la granularité ou encore la répétitivité de motifs. Bien que sensibles au bruit et à la résolution, ils apportent une information complémentaire précieuse, notamment pour différencier les styles fondés sur la texture du coup de pinceau ou du fond.

### 2.3 Features thématiques

En complément des descripteurs visuels classiques, nous avons cherché à intégrer une représentation sémantique des œuvres, c'est-à-dire une information sur leur contenu ou leur signification. Pour cela, nous avons utilisé le modèle CLIP (Contrastive Language–Image Pretraining) développé par OpenAI, qui permet d'associer une image à un vecteur sémantique issu du langage.

**Principe.** CLIP est un modèle entraîné pour rapprocher dans un espace vectoriel les images et les textes qui se correspondent. Appliqué à notre contexte, il permet d'extraire pour chaque peinture un vecteur d'embedding de dimension 512 qui encode des informations thématiques globales : personnages, objets, ambiances, ou scènes représentées.



(a) correspondance des features thématiques sur une peinture

(b) Répartition des pourcentages de probabilité d'une feature thématique sur cette peinture

FIGURE 3 – Extraction de features thématiques

**Méthode utilisée.** Nous avons appliqué CLIP directement sur l'image entière, sans découpe préalable. Cela permet de simplifier le calcul et d'éviter certains effets de bord rencontrés lors d'une approche précédente par patchs. Le vecteur obtenu est ensuite normalisé pour être intégré aux autres features.

**Intérêt.** Cette représentation permet de capturer des différences sémantiques difficilement détectables par les seules couleurs ou textures. Elle est particulièrement utile pour distinguer des styles iconographiquement riches, comme le symbolisme ou le classicisme narratif. Ces features thématiques enrichissent donc significativement l'espace descripteur global.

## 3 Représentations profondes

Les descripteurs visuels précédents, bien que utiles, restent limités à des représentations manuellement conçues. Pour capturer des caractéristiques plus abstraites, liées à l'apparence globale des œuvres, nous avons eu recours à des représentations dites profondes, extraites automatiquement à l'aide de réseaux de neurones préentraînés.

### 3.1 Descripteurs visuels via ResNet50

Nous avons utilisé le réseau convolutionnel ResNet50, préentraîné sur ImageNet, pour extraire un vecteur de caractéristiques de bas niveau pour chaque peinture. Le réseau est appliqué à l'image entière, et nous récupérons le vecteur de dimension 2048 issu de la dernière couche avant classification, après un global average pooling.

Ce vecteur encode une représentation riche et compacte de l'apparence de l'œuvre : organisation spatiale, formes dominantes, contrastes, mais aussi certaines régularités visuelles apprises sur des millions d'images naturelles. Il est particulièrement adapté à la capture du "style visuel global" d'une peinture.

**Prétraitement.** Les images sont redimensionnées à  $224 \times 224$  pixels, comme attendu par ResNet50, puis normalisées. Le calcul de l'embedding est réalisé en batch sur GPU.

### 3.2 Affinage des vecteurs CLIP

Les vecteurs thématiques extraits précédemment via CLIP ont également été affinés. Contrairement à une approche antérieure où les images étaient découpées en sous-parties, nous avons choisi ici de traiter les œuvres sans découpe, pour trois raisons principales :

- éviter une perte d'informations liée au recadrage ou à une couverture incomplète ;
- accélérer le temps de calcul ;
- renforcer la cohérence sémantique globale des embeddings.

Le vecteur CLIP final, de dimension 512, est directement concaténé aux autres descripteurs.

### 3.3 Fusion des descripteurs

Les différents types de descripteurs — couleurs (6 valeurs), textures (6), thématiques (512), ResNet (2048) — sont concaténés pour former un vecteur unique par peinture. Après traitement des données manquantes (si nécessaire), chaque vecteur est centré et normalisé avant d'être utilisé pour la suite du traitement.

Cette concaténation permet de combiner des informations complémentaires : statistiques simples, structure locale, contenu sémantique et apparence globale. Elle aboutit à un espace de représentation complet, prêt à être réduit et analysé.

## 4 Réduction de dimension

Les vecteurs de descripteurs obtenus après concaténation sont de très grande dimension (plus de 2500 composantes par image), ce qui rend difficile une analyse directe. Pour simplifier l'espace de représentation tout en conservant l'essentiel de l'information, nous avons appliqué une réduction de dimension par ACP (Analyse en Composantes Principales, ou PCA en anglais).

### 4.1 Application de l'ACP

Nous avons appliqué l'ACP (ou PCA en anglais) sur l'ensemble des vecteurs concaténés décrivant les peintures. Avant cela, chaque composante a été centrée et normalisée, de façon à

ne pas favoriser arbitrairement un type de descripteur (par exemple ceux issus de ResNet, plus nombreux).

Le nombre de composantes principales conservées a été fixé à 1000. Ce choix s'appuie sur la courbe de variance cumulée, qui montre que plus de 90 % de la variance totale est capturée par ces 1000 composantes. Cette valeur constitue donc un compromis raisonnable entre richesse d'information et simplicité de représentation.

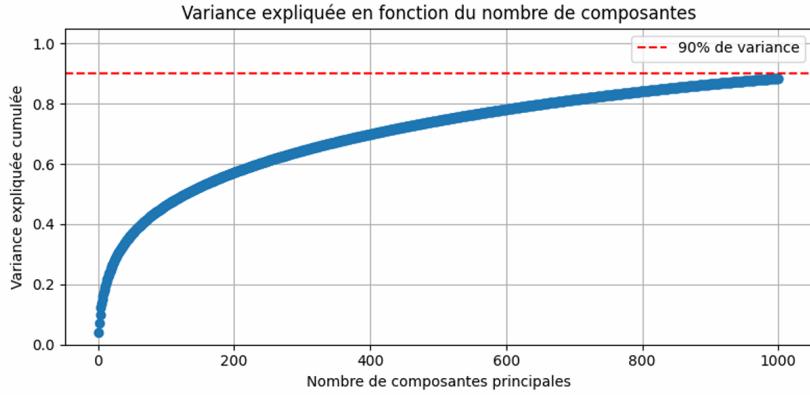


FIGURE 4 – Courbe de la variance cumulée en fonction du nombre de composantes principales retenues

**Résultat.** Chaque peinture est ainsi représentée, à ce stade, par un vecteur de dimension 1000. Ce nouvel espace sert ensuite de base pour l'algorithme de clustering non supervisé.

## 4.2 Justification du seuil de 1000 dimensions

Le choix de conserver 1000 composantes principales n'a pas été fixé arbitrairement. Il repose sur l'analyse de la variance expliquée par l'ACP. Lors de la projection des données, nous avons tracé la courbe de variance cumulée en fonction du nombre de composantes retenues.

Cette courbe montre que la barre des 90 % de variance totale est franchie autour de la 1000<sup>e</sup> composante. Ce seuil constitue donc un compromis judicieux : il permet de conserver la quasi-totalité de la structure informative des données tout en réduisant significativement leur dimension (de plus de 2500 à 1000). Une réduction plus agressive aurait entraîné une perte excessive d'information, tandis qu'un nombre plus élevé aurait alourdi inutilement les calculs de clustering.

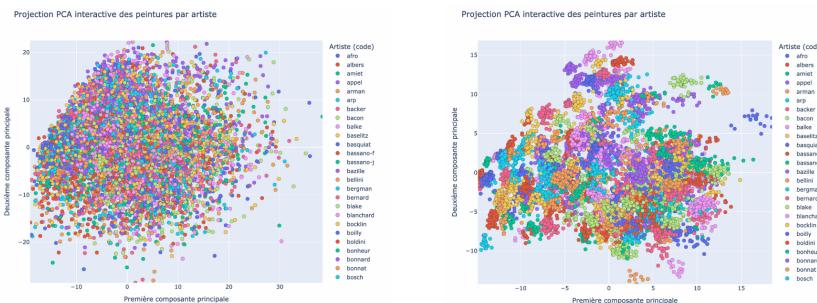


FIGURE 5 – Projection de la PCA et contraction par artiste

Par ailleurs, ce seuil s'est avéré empiriquement satisfaisant : les résultats du clustering obtenus sur cet espace réduit sont cohérents avec les styles artistiques attendus, comme on le verra dans la suite du rapport. Nous avons ensuite contracté les peintures autour du barycentre des artistes,

car les clusters étaient initialement trop éparpillés. Cela nous a donné un rendu plus satisfaisant et plus exploitable.

À la suite de la projection PCA, il s'est avéré que les données étaient trop dispersées, ce qui entraînait des résultats moins pertinents par la suite, notamment après l'application de l'algorithme K-means. Nous avons jugé judicieux de rapprocher chaque image, dans l'espace PCA, du barycentre des artistes auxquels elle appartenait.

## 5 Clustering non supervisé

Une fois les peintures représentées dans un espace réduit de 1000 dimensions par ACP, nous avons cherché à regrouper celles présentant des caractéristiques visuelles et sémantiques similaires. Pour cela, nous avons utilisé l'algorithme de clustering non supervisé K-means.

### 5.1 Méthode K-means

**Mise en œuvre.** Nous avons appliqué K-means sur l'espace réduit obtenu après ACP. Le nombre de clusters  $k$  a été déterminé empiriquement après plusieurs essais : il a été fixé à 50, en cherchant un compromis entre diversité stylistique et lisibilité des résultats. L'algorithme a été initialisé avec la méthode `k-means++`, pour une meilleure convergence, et a été exécuté plusieurs fois avec des seeds différentes pour vérifier la stabilité des regroupements.

**Distance utilisée.** La distance euclidienne a été utilisée pour mesurer la proximité entre vecteurs. Dans notre cas, chaque point correspond à une œuvre représentée dans un espace PCA de 1000 dimensions normalisées. Ce choix est cohérent avec les hypothèses de K-means (clusters de forme sphérique dans un espace métrique).

**Sortie.** Chaque peinture s'est vue attribuer un label de cluster. Ces regroupements ont ensuite été analysés visuellement et confrontés aux styles réels connus des artistes (via WikiArt) dans les parties suivantes. Le résultat brut de K-means a aussi été affiné, comme décrit dans la section suivante.

### 5.2 Affinage des clusters et labellisation

L'algorithme K-means nous a permis d'obtenir une partition automatique des œuvres en 50 groupes. Cependant, les clusters ainsi formés ne sont pas immédiatement interprétables d'un point de vue artistique. Un travail d'affinage a donc été mené afin de relier chaque cluster à un ou plusieurs styles picturaux.

**Labellisation manuelle.** Pour chaque cluster, nous avons identifié les artistes les plus représentés, en consultant les métadonnées associées aux peintures. Lorsqu'un ou plusieurs artistes étaient prédominants dans un groupe, nous avons attribué au cluster le ou les styles correspondant(s), d'après leur classification officielle sur WikiArt. Ce processus a permis d'orienter l'interprétation stylistique des clusters, en ancrant les regroupements obtenus dans des catégories artistiques reconnues.

**Analyse de cohérence.** Cette labellisation a révélé une bonne cohérence pour certains clusters. Par exemple, des groupes étaient presque exclusivement composés de peintures impressionnistes ou rococo, ce qui a facilité leur identification. À l'inverse, d'autres clusters étaient plus hétérogènes, reflétant soit une proximité visuelle entre styles historiquement distincts, soit un échec de la séparation par K-means.

**Limites du processus.** La principale difficulté de cette étape réside dans le caractère subjectif de la labellisation manuelle. Certains styles étant proches ou ambigus (réalisme vs romantisme, par exemple), la frontière entre clusters n'est pas toujours nette. De plus, le nombre inégal d'œuvres par style dans notre base a pu déséquilibrer certains groupes.

**Améliorations proposées.** Pour mieux refléter la pluralité des appartances stylistiques, nous avons envisagé une attribution probabiliste à l'aide d'une fonction softmax appliquée sur les distances aux centroïdes. Cette méthode, explorée dans une phase ultérieure (section 5.3), permet d'éviter une classification trop rigide et de rendre compte des incertitudes d'affectation.

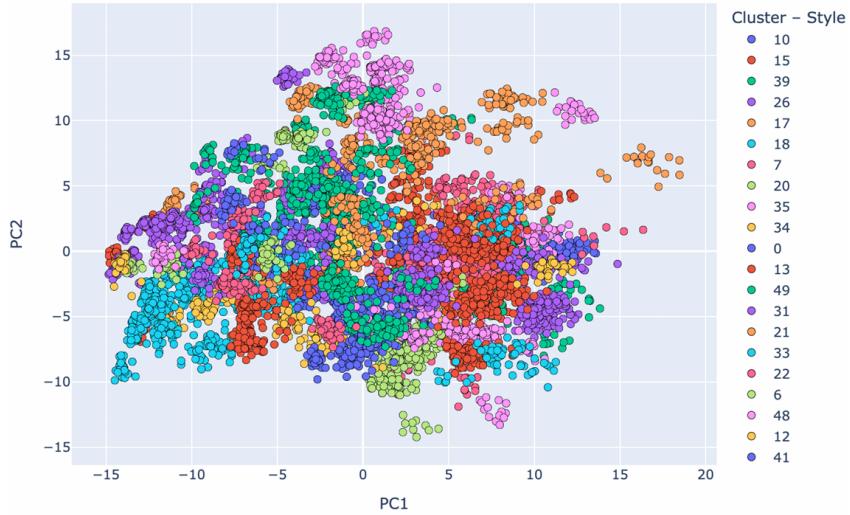


FIGURE 6 – Clustering des images dans l'espace PCA

### 5.3 Optimisation des attributions

Les résultats bruts de l'algorithme K-means sont fondés sur une affectation stricte de chaque peinture à un unique cluster, selon la distance euclidienne minimale. Cette méthode présente des limites : elle ne tient pas compte des cas ambigus, ni du fait qu'une œuvre peut partager des caractéristiques avec plusieurs styles. Pour affiner cette classification, nous avons exploré plusieurs approches permettant de mieux modéliser l'appartenance relative à chaque cluster.

**Utilisation d'un softmax sur les distances.** Plutôt que d'assigner une peinture à son centroïde le plus proche de manière binaire, nous avons appliqué une transformation softmax inversée sur les distances aux centroïdes.

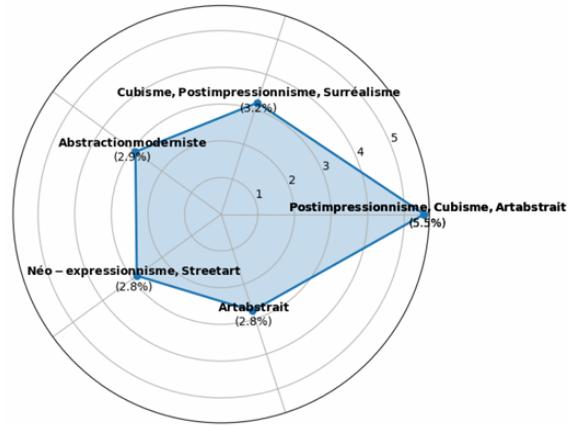
$$p_k = \frac{e^{-d_k/\tau}}{\sum_{j=1}^K e^{-d_j/\tau}}$$

Cette fonction permet de calculer une probabilité d'appartenance à chaque cluster. Plus une peinture est proche d'un centroïde, plus la probabilité associée est élevée. Cette méthode offre une lecture plus nuancée du résultat, en rendant visibles les cas d'hésitation ou de frontière entre styles.

**Introduction d'une pénalisation.** Pour limiter les effets des clusters très larges ou très denses, nous avons testé une variante pondérée du softmax, intégrant un terme de pénalité selon la densité ou la dispersion du cluster. Cette approche vise à éviter qu'un cluster dominant n'absorbe des œuvres simplement parce qu'il est statistiquement plus attractif.



(a) Léger

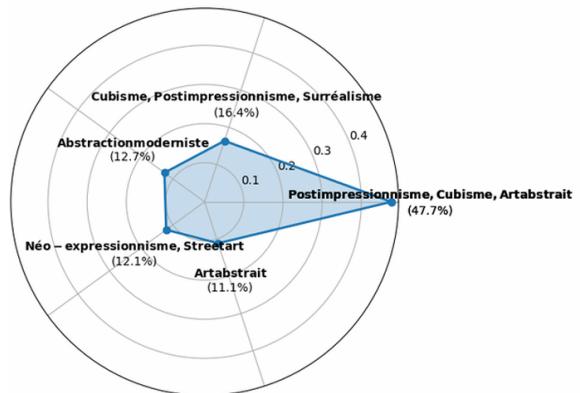


(b) Top 5 des clusters les plus proches

FIGURE 7 – Attribution des clusters avec la distance euclidienne



(a) Léger



(b) Top 5 des clusters les plus proches

FIGURE 8 – Attribution des clusters et introduction de la distance *softmax*

À partir de ces distances, nous avons défini une probabilité d'appartenance  $p_k$  via une fonction softmax, modulée par un paramètre de température  $\tau$  :

$$p_k = \frac{e^{-d_k/\tau}}{\sum_{j=1}^K e^{-d_j/\tau}}$$

où  $d_k = \text{dist}(\mathbf{x}, \mathbf{c}_k)$  représente la distance euclidienne entre l'image  $\mathbf{x}$  et le centre du cluster  $k$ .

Dans la variante pénalisée, nous avons modifié la distance utilisée en introduisant un facteur multiplicatif  $\lambda$  pour pénaliser certains clusters. Plus précisément, si l'on note  $C_{\text{true}}$  l'ensemble des clusters considérés comme pertinents pour une image donnée, la distance  $\tilde{d}_k$  est définie comme suit :

$$\text{dist}(\mathbf{x}, \mathbf{c}_k) = \|\mathbf{x} - \mathbf{c}_k\|^2$$

$$\tilde{d}_k = \begin{cases} d_k, & \text{si } k \in C_{\text{true}} \\ \lambda \cdot d_k, & \text{si } k \notin C_{\text{true}} \end{cases}$$

La probabilité d'appartenance devient alors :

$$p_k = \frac{e^{-\tilde{d}_k/\tau}}{\sum_{j=0}^{K-1} e^{-\tilde{d}_j/\tau}}$$

Cela permet d'atténuer l'influence des clusters statistiquement dominants, en tenant compte de leur structure interne.

**Effets observés.** Ces deux ajustements ont permis d'améliorer la lisibilité des regroupements dans les cas limites. Certaines œuvres initialement mal classées ont ainsi pu être réaffectées à un cluster plus cohérent d'un point de vue artistique. Néanmoins, ces raffinements n'ont pas toujours produit de gains visibles lorsque les distances inter-centroïdes étaient déjà bien séparées.

**Choix final.** Pour l'analyse finale des résultats, nous avons conservé les affectations issues de la méthode softmax, jugée plus représentative des continuités stylistiques réelles. Ce choix reflète notre volonté de dépasser une approche purement catégorielle au profit d'une interprétation plus souple, conforme à la réalité complexe de l'histoire de l'art.

## 6 Analyse des résultats

Après labellisation et affinage, les regroupements obtenus par clustering ont été confrontés à des styles artistiques reconnus. Nous avons évalué la pertinence des résultats de manière qualitative, en étudiant plusieurs clusters bien formés, d'autres plus ambigus, et enfin une estimation globale de la cohérence de la classification.

### 6.1 Clusters représentatifs

Certains clusters ont révélé une homogénéité particulièrement forte, à la fois du point de vue des artistes représentés et des caractéristiques stylistiques. Ces cas montrent que l'approche mise en œuvre permet effectivement de faire émerger des regroupements pertinents.

**Exemple : Cluster n°44 : Impressionnisme.** Ce groupe contient presque exclusivement des peintures aux couleurs claires, aux contours flous, et présentant des scènes de plein air. On y retrouve de nombreux tableaux d'artistes comme Claude Monet ou Camille Pissarro. Le vecteur moyen du cluster est très marqué par des teintes pastel, des textures douces et un embedding CLIP typique de paysages.



FIGURE 9 – Exemple d'une peinture classée dans le style impressionniste

**Exemple : Cluster n°28 : Rococo.** Les œuvres de ce cluster sont dominées par une palette lumineuse, des décors chargés et une iconographie légère ou mythologique. Les artistes prédominants incluent Fragonard, Watteau et Boucher. Ce résultat est en accord avec la caractérisation stylistique attendue du rococo.

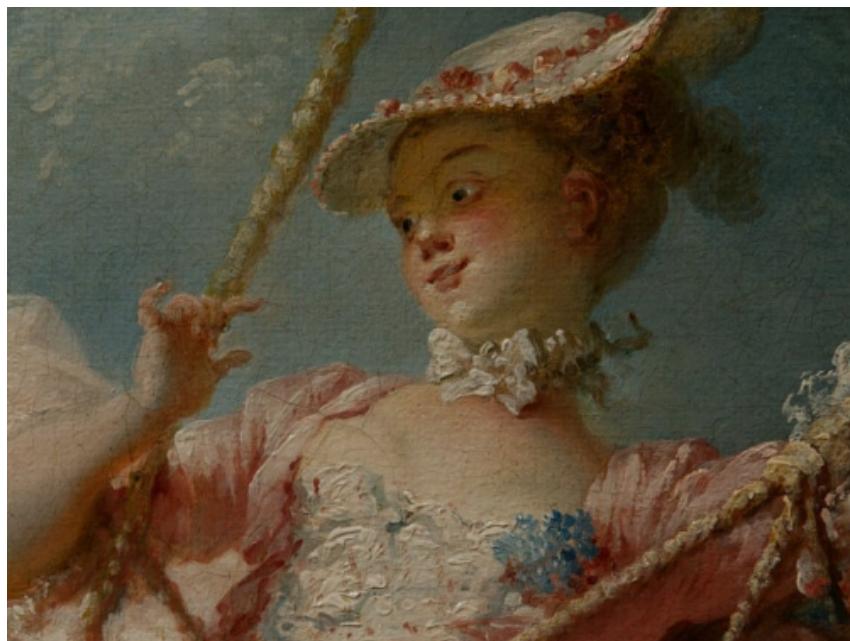


FIGURE 10 – Exemple d'une peinture classée dans le style Rococo

## 6.2 Clusters problématiques

D'autres clusters sont apparus moins cohérents ou plus difficiles à interpréter. Deux types de cas se dégagent : les regroupements confus et les clusters stylistiquement faibles.

**Cluster n°39 : Postimpressionnisme.** Ce cluster contient un mélange d'œuvres proches du style postimpressionniste, mais aussi des tableaux réalistes ou modernes. L'hétérogénéité visuelle est nette : certains tableaux rappellent Cézanne ou Gauguin, tandis que d'autres semblent sans lien apparent. Cette confusion s'explique sans doute par la complexité intrinsèque du style postimpressionniste, qui sert souvent de catégorie fourre-tout.



(a) Peinture postimpressionniste bien classée (subjectivement)



(b) Peinture mal classée dans le style postimpressionniste (subjectivement)

FIGURE 11 – Exemple de cluster contenant plusieurs styles différents

**Clusters ambigus ou hybrides.** Certains groupes mêlent des œuvres issues de styles différents mais visuellement proches. Par exemple, des peintures baroques et néoclassiques peuvent se retrouver dans un même cluster en raison de similarités dans les contrastes ou la composition. Cela montre les limites de la séparation automatique fondée uniquement sur des features visuelles.

## 6.3 Évaluation globale et interprétations

Pour estimer globalement la qualité du regroupement, nous avons passé en revue les 50 clusters manuellement et les avons classés selon trois niveaux :

- : correspondance forte avec un style identifiable — 26 clusters ;
- : regroupement partiellement cohérent — 13 clusters ;
- : cluster mal structuré ou incohérent — 11 clusters.

Ces résultats montrent que plus de la moitié des clusters révèlent une structure stylistique claire, ce qui confirme l'efficacité globale de la méthode. Toutefois, l'interprétation reste délicate pour plusieurs raisons :

- L'accès aux étiquettes reste limité aux artistes majoritaires, sans pourcentage d'appartenance formel.
- Certains styles sont peu représentés dans la base, ce qui empêche le modèle de bien apprendre leur structure.
- La réduction de dimension par ACP peut faire disparaître certaines distinctions fines mais pertinentes.

Malgré ces limites, l'approche permet de proposer un classement intéressant, structuré, et partiellement interprétable. Elle constitue une base solide pour des raffinements futurs.

## 7 Conclusion

Ce projet nous a permis d'explorer les capacités de l'apprentissage non supervisé à structurer une base d'œuvres d'art selon des critères visuels et stylistiques, sans connaissance préalable des styles. En combinant des descripteurs manuels (couleurs, textures, thématiques) à des représentations profondes issues de modèles comme ResNet50 et CLIP, nous avons construit un espace riche et multidimensionnel capable de capturer la diversité esthétique des peintures.

L'utilisation de l'ACP a permis de réduire efficacement cet espace tout en conservant l'essentiel de l'information, rendant possible une classification par K-means. L'analyse des clusters a montré que cette approche permet de faire émerger des regroupements stylistiquement cohérents dans plus de la moitié des cas, comme pour le style baroque.

Ce projet nous a offert une mise en pratique complète des méthodes étudiées dans le cours IMA206, à travers une problématique ouverte, exigeante, et ancrée dans un domaine culturellement riche. Il démontre l'intérêt des méthodes non supervisées pour explorer des structures latentes dans des données complexes et visuelles.