

Using Image Analysis and Unsupervised Learning to identify style patterns in fashion luxury brands

MASTER: Business Analytics and Data Science (BADS)

Authors: Benjamín Cruz Infante & Pilade Riello

Tutor: Marco Brambilla

Academic Year: 2024-25

Contents

Contents	ii
Abstract	v
1 Executive Summary of the thesis	vi
1.1. Introduction.....	vi
1.2. Methodology	vi
1.3. Results and Findings.....	vii
1.4. Conclusions and Future Work.....	ix
Executive Summary Bibliography	x
1 Chapter one: Overview	11
1.1. Background: The AI Disruption in Luxury and E-commerce	11
1.2. Problem Statement: Analysis Paralysis in the Evaluation Phase	12
1.3. Research Question, Objectives, and Scope	13
2 Chapter 2: Literature Review	15
2.1. Artificial Intelligence in the Luxury Industry	15
2.2. The Rise of Generative AI in Fashion and E-Commerce	15
2.3. Cognitive Overload and Product Discovery in Digital Luxury	16
2.4. Computer Vision in Fashion Product Analysis	16
2.5. Unsupervised Learning and Visual Clustering in Retail.....	17
2.6. Strategic Implications for Luxury Brands.....	18
3 Chapter 3: Methodology	19
3.1. Research Design.....	19
3.2. Methodological Framework.....	19
3.3. Dataset Description	20
3.4. Image Preprocessing and Normalization (Technical Deep Dive)	20
3.5. Feature Extraction and Engineering: The Dual Feature Strategy.....	21
3.5.1. Deep Features (ViT Embeddings).....	21
3.5.2. Handcrafted Features	21
3.5.3. Dimensionality Reduction and Normalization(PCA).....	22
3.6. Clustering Algorithm.....	22

3.7.	Integration with Pricing Data	23
3.8.	Visualization and Evaluation	23
3.8.1.	Visualization	23
3.8.2.	Evaluation Metrics	23
3.9.	Reproducibility and Ethics.....	24
4	Chapter 4: Results and Analysis.....	25
4.1.	General Overview	25
4.2.	Bottega Veneta	25
4.2.1.	Visual Cohesion and Cluster Patterns.....	25
4.2.2.	Economic Interpretation.....	25
4.2.3.	Strategic Insight	26
4.3.	Yves Saint Laurent (YSL).....	26
4.3.1.	Visual Cohesion and Cluster Patterns.....	26
4.3.2.	Economic Interpretation.....	27
4.3.3.	Strategic Insight	27
4.4.	4.3. Prada	27
4.4.1.	Visual Cohesion and Cluster Patterns.....	27
4.4.2.	Economic Interpretation.....	27
4.4.3.	Strategic Insight	27
4.5.	Dolce & Gabbana	28
4.5.1.	Visual Cohesion and Cluster Patterns.....	28
4.5.2.	Economic Interpretation.....	28
4.5.3.	Strategic Insight	28
4.6.	Fendi.....	29
4.6.1.	Visual Cohesion and Clustering Patterns	29
4.6.2.	Economic Interpretation.....	29
4.6.3.	Strategic Insights	29
4.7.	Cartier.....	30
4.7.1.	Visual Cohesion and Clustering Patterns	30
4.7.2.	Economic Interpretation.....	30
4.7.3.	Strategic Insights	31
4.8.	Business Implications	31
4.8.1.	Support for the Consumer: Accelerating Evaluation Time.....	32
5	Conclusions & Future Works.....	33
6	Bibliography and citations	33
A	Appendix A.....	37
A.1.	Bottega Veneta	37

B	Appendix B	38
6.1.	Dendrogram Cartier.....	38
	List of Figures	39
	List of tables.....	39
7	Acknowledgments	40

Abstract

This thesis explores how Artificial Intelligence can reshape luxury e-commerce by addressing one of its most persistent challenges: information and choice overload. Using Computer Vision and Unsupervised Learning, the research investigates whether Vision Transformer (ViT) embeddings, combined with Agglomerative Clustering, can visually segment luxury eyewear products into coherent aesthetic groups that integrate both style and price information. A dataset of 763 images from six major luxury brands — Dolce & Gabbana, YSL, Prada, Fendi, Bottega Veneta and Cartier — was collected via automated web scraping and analyzed through a deep-learning pipeline. The results demonstrate that ViT-based embeddings effectively capture high-level aesthetic coherence beyond color or material, revealing patterns consistent with each brand's visual identity and pricing hierarchy. By merging visual and economic dimensions, this model provides dual value: reducing consumer cognitive load through more intuitive product discovery, and offering brands an analytical tool for competitive benchmarking, assortment optimization, and price positioning. Ultimately, the study proposes a scalable framework that bridges data intelligence and aesthetic perception — transforming how luxury brands and customers navigate visual identity in the digital era.

Key-words: Artificial Intelligence, Computer Vision, Vision Transformers, Unsupervised Learning, Luxury Fashion, Product Clustering, Web Scraping, Agglomerative Clustering

1 Executive Summary of the thesis

1.1. Introduction

The luxury fashion industry stands at the intersection of creativity, technology, and data. As Artificial Intelligence (AI) continues to redefine the digital ecosystem, brands face both an opportunity and a challenge: how to translate vast product catalogs into personalized, visually intuitive experiences. This thesis explores the application of Computer Vision (CV) and Unsupervised Learning to the luxury eyewear segment, proposing a framework that merges aesthetic clustering with pricing intelligence.

The motivation for this research originates from a growing behavioral phenomenon in digital commerce: **analysis paralysis**. According to The State of Fashion 2025 and Zalando’s White Paper (2025), 74% of customers abandon online purchases due to the cognitive overload caused by excessive product choice. In parallel, 50% of executives identify product discovery as the AI use case with the highest expected impact in the upcoming years. Within this context, the thesis investigates how deep learning — specifically Vision Transformer (ViT) embeddings — can replicate human aesthetic perception to simplify decision-making for both consumers and brand teams.

The central research question is therefore twofold:

Can Vision Transformer-based embeddings, when combined with Agglomerative Clustering, generate visually coherent groups of luxury eyewear products that not only offer actionable insights for merchandising and pricing decisions, but also shorten consumers’ cognitive evaluation time and reduce frustration caused by digital choice overload?

1.2. Methodology

a) Research Design

The research adopts an exploratory, data-driven approach using **unsupervised learning** to model visual similarity and market structure across six luxury brands: **Dolce & Gabbana, YSL, Prada, Fendi, Bottega Veneta and Cartier**. A total of **763 product images** were **collected via automated web scraping** from each brand’s official

Italian e-commerce platform, ensuring compliance with public access and ethical AI research guidelines.

The pipeline integrates both **deep learning and traditional feature** extraction within a five-stage framework: i) **Data Acquisition** – Web scraping of product images, names, and prices using Selenium and ChromeDriver, ii) **Preprocessing – Image normalization** (RGB conversion, adaptive background crop, and centering) to ensure consistency across datasets, iii) **Feature Extraction** – Dual-feature strategy combining (a) ViT embeddings (semantic features) and (b) handcrafted descriptors (shape, color, texture), iv) **Clustering** – Hierarchical segmentation using **Agglomerative Clustering** with cosine distance and average linkage, optimized via Silhouette Score, v) **Visualization and Business Integration** – Cluster mapping enriched with price data and brand identity, ordered by descending pairwise separation.

b) Model Architecture

The **ViT-Base model** was selected, leveraging the robust visual features obtained through its pretraining using the **Masked Autoencoder (MAE)** approach. This method equips the model with a superior capacity to capture global spatial relationships through its self-attention mechanisms. Each image was represented by a **768-dimensional embedding vector**, effectively capturing the product’s core aesthetic fingerprint. These high-dimensional semantic features then served as the input for **Agglomerative Clustering**, a method chosen for its interpretability and ability to reveal hierarchical aesthetic continua rather than imposing rigid, predefined taxonomies.

A secondary feature set — **Handcrafted Descriptors (DSCT)** — was constructed for benchmarking. This included: 1) **Shape**: Hu Moments & Histogram of Oriented Gradients (HOG), 2) **Color**: HSV histograms, 3) **Texture**: Local Binary Patterns (LBP) & Gabor filters

Dimensionality reduction was performed through the PCA algorithm and subsequent scaling of the feature matrices. For Vision Transformer (ViT) embeddings, no reduction was applied in order to retain the complete representational information. Cluster quality was assessed quantitatively using the Silhouette Score and qualitatively through visual inspection of the resulting clusters.

1.3. Results and Findings

The analysis revealed that **ViT-based clustering consistently outperformed traditional handcrafted descriptors**, achieving higher coherence in visually complex datasets.

Across all six brands, the model successfully identified **distinct aesthetic families**—often aligning with known design signatures.

a) Quantitative Insights

The **Silhouette Score** confirmed the higher separation achieved by ViT embeddings compared to handcrafted features, validating the model's ability to distinguish subtle design nuances (e.g., frame curvature, color translucency).

The integration of pricing data enabled the computation of **cluster-level averages, min and max**, facilitating the estimation of the **Brand Price Premium** — the quantifiable link between visual distinctiveness and price differentiation.

b) Business Implications

From a business intelligence perspective, the system acts as a **visual-economic map** for competitive benchmarking. For brands, it:

- Accelerates **product discovery** by grouping similar designs and highlighting redundancies.
- Enhances **pricing strategy** by aligning aesthetic uniqueness with price positioning.
- Supports **merchandising** by identifying underrepresented design families or oversaturated clusters.
- From the consumer side, it enables the construction of **visual recommendation systems** that reduce cognitive overload, guiding users through product aesthetics rather than metadata.

c) Useful and Potential extensions

The research extends beyond pure technical contribution. It outlines how **AI-driven aesthetic segmentation** can be integrated into **CRM and personalization strategies**, enabling a new dimension of customer understanding: *visual preference modeling*. By translating style into quantifiable embeddings, brands can create:

- **Cross-category upselling:** Recommending visually similar products across categories (e.g., accessories sharing design traits).
- **Look-alike recovery flows:** Suggesting highly similar products to users who abandoned their carts.
- **Dynamic retention campaigns:** Tailoring communication to each user's dominant aesthetic cluster.

These strategies demonstrate how **unsupervised visual learning** can evolve into actionable CRM intelligence, merging design and data.

1.4. Conclusions and Future Work

This research demonstrates that combining **Vision Transformers** with **Agglomerative Clustering** offers a scalable, interpretable, and business-relevant approach to aesthetic segmentation in luxury e-commerce. The integration of visual embeddings with pricing data bridges the gap between **creative direction and analytical decision-making**, a frontier that leading organizations such as LVMH and Kering are already exploring (Zalando White Paper, 2025; The State of Fashion 2025).

Future developments will expand the dataset to include additional brands (e.g., Giorgio Armani, Gucci) to enhance embedding robustness and cross-brand coherence. A second step involves deploying the model into an **interactive web dashboard**, allowing users and brand teams to explore clusters dynamically by brand, price range, or aesthetic dimension.

Ultimately, the study contributes to the broader conversation on how **AI can humanize digital luxury experiences** — not by replacing creativity, but by amplifying it through data-driven understanding.

Executive Summary Bibliography

- 1) The Business of Fashion & McKinsey — The State of Fashion 2025
- 2) Zalando & Business of Fashion (2025) — The Future of E-Commerce, Powered by Generative AI (White Paper)
- 3) Dosovitskiy et al. (2021) — An Image is Worth 16×16 Words (ViT)
- 4) Harvard Business Review (2024) — AI: The Secret to a Thriving E-Commerce Business
- 5) Giri et al. (2019) — IEEE Access review on AI in fashion/apparel
- 6) Choi et al. (2023) — AI-based automated fashion design system
- 7) Innovation Training — Journey Mapping Workshop
- 8) Zakeke (2024) — Why Virtual Try-On Fashion is the Future of Shopping
- 9) Aptean (2023) — Machine Learning in Fashion
- 10) Sceppacerca (2024) — LUISS Master's Thesis

1 Chapter one: Overview

1.1. Background: The AI Disruption in Luxury and E-commerce

The luxury and fashion industries are undergoing a fundamental transformation driven by Artificial Intelligence (AI). What initially began as an operational tool for automation has evolved into a strategic necessity to enhance personalization, accelerate time-to-market, and sustain competitive advantage (Giri et al., 2019; LeewayHertz, 2023; Fashion Retail Academy, 2023). AI is now deeply embedded across the fashion value chain, from computer-aided design and predictive analytics to automated merchandising and customer experience optimization.

Major luxury conglomerates such as LVMH, Kering, and Richemont have increasingly embraced AI to optimize creative workflows, trend forecasting, and supply chain management (Williams, 2018). AI-driven Customer Relationship Management (CRM) and product recommendation systems have become central to luxury retailers' personalization strategies, leveraging Machine Learning algorithms to identify behavioural patterns and forecast intent (IBM, n.d.).

Recent industry research emphasizes Generative AI (GenAI) as the technology most likely to revolutionize the e-commerce value chain. It is projected to redefine customer interaction, product visualization, and creative production across marketing and merchandising functions (Zalando White Paper, 2025). Virtual try-on technologies, enabled by AI-powered 3D modeling, are already reshaping the digital shopping experience (Zakeke, 2024). According to The Business of Fashion and McKinsey's State of Fashion 2025, GenAI could contribute up to \$275 billion to the operating profits of the fashion and luxury industries within the next three to five years (Zalando White Paper, 2025, p. 2).

However, the rapid expansion of online product catalogues has simultaneously created an unprecedented problem for consumers: information and choice overload. Despite the rise of personalization technologies, customers face what behavioral economists describe as "**analysis paralysis**," leading to increased cart abandonment and lower conversion rates. Empirical data illustrates the magnitude of this problem:

74% of customers report abandoning an e-commerce purchase due to overwhelming product choice (The State of Fashion 2025, p. 2; Zalando White Paper, 2025).

1.2. Problem Statement: Analysis Paralysis in the Evaluation Phase

The digital customer journey typically follows four stages: Awareness, Consideration, Evaluation, and Purchase (Innovation Training, n.d.). Among these, the Evaluation phase—where customers actively compare multiple products—is the most vulnerable to friction. When confronted with a high volume of undifferentiated options, consumers experience cognitive overload, which reduces purchase likelihood and satisfaction (The State of Fashion 2025, p. 2).

In response to this challenge, 50% of fashion executives surveyed in The State of Fashion 2025 identified Product Discovery and Advanced Search as the AI applications with the highest expected impact in the coming years (Zalando White Paper, 2025, Exhibit 1).

This thesis aims to address this bottleneck by developing a Computer Vision (CV) framework that supports both the consumer and the brand during the Evaluation phase.



Figure 1: Customer Journey Maps by Innovation Training

A. Support for the Consumer (Accelerating Evaluation Time)

From the consumer perspective, personalization mechanisms must evolve beyond textual data and purchase history. While existing systems such as chat agents, recommendation engines, and dynamic pricing models can tailor content based on behavioral data, they often fail to capture the visual aesthetics that strongly influence purchase intent in fashion (Harvard Business Review, 2024).

This study proposes a visual similarity system leveraging Vision Transformers (ViT) to identify and recommend products with stylistic resemblance to a user's aesthetic preference. By reducing irrelevant results, the system aims to shorten the cognitive evaluation time and decrease frustration associated with digital choice overload, directly addressing the behavioral friction affecting 74% of customers (Zalando White Paper, 2025, p. 2).

B. Support for the Brand (Competitive Intelligence and Operational Efficiency)

From the brand's perspective, product intelligence, assortment design, and pricing remain labor-intensive and fragmented processes. According to *The State of Fashion 2025*, executives recognize that GenAI and Deep Learning can dramatically enhance the productivity of creative and merchandising teams (p. 7).

This research introduces a framework that generates AI-driven visual maps of product clusters across multiple luxury brands. Such maps allow teams in Merchandising, Design, and Pricing to understand market saturation, emerging styles, and differentiation. Integrating price data into these embeddings enables the quantification of a "Brand Price Premium" — the degree to which aesthetic distinctiveness justifies a higher price point.

This innovation aligns with Bain & Company's (2024) recommendation for the luxury sector to use AI not only for personalization but also for strategic differentiation and margin protection (Zalando White Paper, 2025).

Ultimately, the proposed model enhances efficiency by replacing hours of manual research with an automated, data-driven process that offers aesthetic clustering, pricing intelligence, and competitive benchmarking in a single analytical framework.

1.3. Research Question, Objectives, and Scope

Despite growing evidence of AI's strategic value in luxury retail, the integration of deep visual embeddings with pricing analytics remains underexplored. Most literature focuses either on generative design or on business optimization, leaving a methodological gap between visual similarity modelling and strategic business intelligence.

Research Question:

Can Vision Transformer (ViT)-based embeddings, when integrated with Agglomerative Clustering, generate visually coherent style groupings of luxury eyewear products that, once enriched with pricing data, provide actionable insights for internal brand teams and improve the consumer experience by mitigating digital choice overload.

General Objective: To implement and evaluate an unsupervised learning pipeline that employs ViT-based deep embeddings and hierarchical clustering to segment luxury eyewear collections by visual similarity and pricing behavior, delivering a tool for competitive analysis and aesthetic benchmarking.

Scope of Study: This project focuses on the luxury eyewear market, analyzing 763 image samples collected via automated web scraping from the online catalogs of Dolce & Gabbana, YSL, Prada, Fendi, Bottega Veneta and Cartier. The model's outputs will be compared against a baseline of handcrafted features (color histograms, shape, and texture descriptors) to evaluate the added value of deep feature extraction in unsupervised aesthetic clustering (IBM, n.d.).

2 Chapter 2: Literature Review

2.1. Artificial Intelligence in the Luxury Industry

AI has evolved from a support technology into a core strategic capability in luxury retail, redefining creativity, production, and consumer interaction (Williams, 2018; Admin, 2023). As highlighted in *The State of Fashion 2025*, the industry now considers AI “a creative and operational co-pilot” (p. 4).

Luxury conglomerates such as **LVMH**, **Kering**, and **Richemont** integrate AI for inventory optimization, design assistance, and CRM personalization. LVMH’s partnership with **OpenAI** and **Google Cloud** exemplifies the integration of AI in product development and analytics (AI, Data & Analytics Network, 2024). Similarly, **Chanel** has deployed its first AI-based beauty try-on app, while **Gucci** uses predictive analytics for market insights and campaign planning (Cosmetic Business, n.d.).

This transformation is driven by the experience-centric nature of luxury consumption, where **emotional engagement** and **aesthetic coherence** define brand equity (Sceppacerca, 2024). AI bridges creativity and analytics, aligning artistic identity with data-informed strategy.

2.2. The Rise of Generative AI in Fashion and E-Commerce

Generative AI (GenAI) represents the most disruptive frontier in digital luxury. According to Zalando’s 2025 White Paper, GenAI applications now span trend forecasting, content generation, personalized recommendations, and visual search (Zalando White Paper, 2025).

73% of fashion executives believe GenAI will have a transformative impact on product discovery and customer experience, while 50% cite it as the technology with the highest strategic potential (Harreis et al., 2023; *The State of Fashion 2025*, p. 2).

Platforms such as Zalando demonstrate GenAI’s power to automate catalog generation and produce dynamic product imagery (In, 2024; Tsymbal, 2024). Louis Vuitton pioneered early adoption through AI chatbots (Arthur, 2017), while Versace

Jeans Couture’s collaboration with Bravò (2023) exemplifies the use of AI-generated campaigns in luxury marketing.

GenAI’s implications extend beyond automation—it **enables the digital representation of craftsmanship and style**, core to luxury identity (Sceppacerca, 2024).

2.3. Cognitive Overload and Product Discovery in Digital Luxury

One paradox of digital transformation is that technological abundance has led to **decision fatigue** among consumers. Behavioral economists define this as “analysis paralysis”—a state where excessive choice reduces satisfaction and purchase likelihood (The State of Fashion 2025, p. 2).

Luxury consumers, who demand a curated and emotionally resonant experience, are especially vulnerable. McKinsey’s research indicates that even high-intent shoppers abandon purchases when overwhelmed by similar products (p. 3).

Traditional personalization—based on text or purchase history—fails to capture visual preference. As Sceppacerca (2024) explains, “Luxury decision-making is visual-first and emotion-driven; data-driven personalization must therefore evolve from recommendation to recognition.”

This research aligns with that principle, proposing **visual product discovery** as a solution to bridge aesthetic intuition and algorithmic filtering (Takyar, 2023; Du et al., 2020).

2.4. Computer Vision in Fashion Product Analysis

Computer Vision (CV) has emerged as a pivotal subfield of AI for extracting and interpreting meaning from visual data. In the fashion industry, CV enables sophisticated applications such as attribute recognition, visual similarity search, and trend detection (*Harvard Business Review*, 2024, p. 6; PostIndustria, n.d.).

Earlier CV methods relied on **handcrafted features**—such as color histograms, shape descriptors, and texture filters (e.g., Gabor, HOG). However, these approaches lacked robustness, performing poorly against variations in lighting, pose, and complex backgrounds (Giri et al., 2019, p. 95379). The advent of Deep Learning revolutionized this field, with Convolutional Neural Networks (CNNs) and later **Transformers** dramatically improving performance by enabling models to capture hierarchical visual relationships (Choi et al., 2023).

Recent studies demonstrate that **Vision Transformers (ViTs)** outperform CNNs in visual retrieval tasks, particularly in high-variance datasets like fashion products

(Dosovitskiy et al., 2021). ViTs leverage **self-attention mechanisms** to model long-range dependencies between image patches, making them highly effective in identifying subtle design variations and achieving superior generalization and interpretability.

In the context of product clustering, ViT-based **embeddings** enable the construction of aesthetic similarity spaces, where visually related items cluster together regardless of brand or metadata. Integrating these rich visual embeddings with key business variables—such as price—transforms a purely visual representation into a strategic analytical tool, forming the foundation of a visual-economic mapping system (*Zalando White Paper*, 2025, p. 15).

2.5. Unsupervised Learning and Visual Clustering in Retail

Unsupervised learning techniques, particularly **Agglomerative Clustering**, are instrumental for identifying natural groupings and latent aesthetic structures in complex product datasets without relying on prior labeling or subjective human tags (Aptean, 2023; *Harvard Business Review*, 2024, p. 7).

Agglomerative Clustering is a hierarchical method that constructs clusters iteratively by merging samples based on their pairwise distance (e.g., Cosine or Euclidean distance). Its advantage in fashion data lies in its flexibility to interpret **aesthetic continua**—where distinctions between styles are gradual rather than strictly categorical. This hierarchical nature allows it to reveal complex transitions between fashion styles (Elastic, n.d.).

When combined with ViT embeddings, this approach enables the creation of interpretable “**visual maps**” of product ecosystems, supporting both consumer navigation (recommendation) and brand analytics (competitive benchmarking). Zalando’s internal research, cited in its 2025 White Paper, reports that applying hierarchical clustering to product embeddings **reduced catalog navigation time by up to 35%**, while increasing click-through rates for style recommendations by **18%** (*Zalando White Paper*, 2025, p. 5).

This aligns with broader findings from HBR (2024, p. 8), which highlight that “AI systems capable of structuring product catalogs visually enable firms to replace static taxonomies with dynamic, user-driven discovery frameworks.” The ability of unsupervised methods to reveal structures of style, form, and price positioning makes them crucial for competitive intelligence in the luxury retail space.

2.6. Strategic Implications for Luxury Brands

AI-driven visual clustering not only enhances consumer experience but also provides **actionable strategic insights** for brand management.

From a **merchandising perspective**, cluster-level analysis reveals gaps or redundancies in the product mix, supporting more balanced assortment planning and informing inventory decisions (*The State of Fashion 2025*, 2025, p. 7).

From a **pricing perspective**, associating visual style clusters with economic data enables the identification and quantification of the **Brand Price Premium**—the degree to which a brand’s aesthetic distinctiveness justifies a higher perceived value and price point (*Zalando White Paper*, 2025, p. 6; Dior, 2023; Icreon, n.d.).

Moreover, for **creative teams**, these visual maps democratize access to competitive intelligence. Designers can instantly visualize how their collection positions within the competitive landscape, fostering data-informed creativity without constraining artistic freedom.

Ultimately, this innovation supports the necessary fusion of data and design in the luxury sector. As Sceppacerca (2024) concludes, “The future of luxury will belong to brands that fuse intuition with intelligence—those that translate data into design, and design into data” (p. 25).

3 Chapter 3: Methodology

3.1. Research Design

This research adopts an exploratory and experimental design to analyze the intersection of visual similarity, aesthetic clustering, and pricing intelligence in the luxury eyewear market. Building upon the conceptual foundation of AI-driven product discovery and computer vision-based aesthetic segmentation, the project aims to evaluate whether Vision Transformer (ViT) embeddings can generate visually coherent product clusters that align with brand identity and pricing structures.

The study employs an unsupervised learning approach, given the absence of labeled data describing aesthetic categories. By combining deep feature extraction through ViT and hierarchical clustering through Agglomerative Clustering, the methodology enables a visual segmentation that does not rely on manual tagging or pre-defined style taxonomies.

3.2. Methodological Framework

The pipeline for this research is an end-to-end framework designed for **multimodal integration** (visual embeddings + numerical data), aligning with the models outlined in the *Zalando White Paper* (2025, p. 11) for next-generation product intelligence systems.

The methodology consists of five core, sequential stages:

1. **Data Acquisition:** Automated web scraping of product images and metadata (price, brand).
2. **Robust Preprocessing:** Advanced *Computer Vision* techniques to isolate the eyewear product, normalize its background, and standardize its size.
3. **Dual Feature Extraction:** Generation of both **ViT deep embeddings** (semantic features) and **Handcrafted Descriptors** (shape, color, texture) for benchmarking.
4. **Clustering and Optimization:** Application of **Agglomerative Clustering** with parameter optimization based on the **Silhouette Score**.

5. **Visualization and Business Integration:** Cluster mapping enriched with price data and brand identity, ordered by descending pairwise separation.

3.3. Dataset Description

The dataset consists of 763 product images and their associated metadata, collected from the official online catalogs of Dolce & Gabbana, Yves Saint Laurent (YSL), Prada, Fendi, Bottega Veneta and Cartier. Data was obtained through a custom web scraping pipeline designed in Python using Selenium and Chrome Driver, ensuring compliance with public data access regulations and ethical standards.

Each entry in the dataset includes:

- Product image (standardized to 224×224 pixels)
- Brand name + Numerical index
- Price (in euros): This variable is critical for integrating aesthetic similarity with economic positioning, enabling the subsequent analysis of the Brand Price Premium.

3.4. Image Preprocessing and Normalization (Technical Deep Dive)

Preprocessing was essential to ensure consistency, given the variability in luxury product imagery (backgrounds, lighting, and scale).

Preprocessing Pipeline

- **Color Standardization** – All images were converted to the RGB color space and normalized by pixel intensity (0–1 range) to standardize lighting and color profiles.
- **Background Filtering** – A combination of grayscale conversion, Gaussian blur, and adaptive thresholding was applied to separate the product from the background and improve edge detection. This process reduced visual noise and preserved fine contours.
- **Centering and Resizing** – Images were resized to **224×224 pixels** and centered to maintain a uniform composition across samples.
- **Normalization** – Finally, mean and standard deviation normalization following **ImageNet preprocessing standards** ($mean = [0.485, 0.456, 0.406]$; $std = [0.229, 0.224, 0.225]$) was applied to align with the pretrained ViT model expectations.

This step ensured that embeddings captured design structure rather than background noise.

3.5. Feature Extraction and Engineering: The Dual Feature Strategy

The pipeline employs a dual feature strategy to combine low-level visual cues (shape, color, texture) with high-level semantic information captured by ViT Masked Autoencoder embeddings.

3.5.1. Deep Features (ViT Embeddings)

- **Model Used:** ViT-Base (Masked Autoencoder) pretrained on ImageNet-21k.
- **Input:** 224×224×3 RGB images.
- **Feature Dimension:** 768.
- **Extraction:** CLS token embedding per image.
- **Purpose:** Semantic and stylistic clustering — captures the aesthetic fingerprint of each product.

The MAE-ViT model was preferred over traditional CNNs due to its ability to preserve shape integrity and learn structural dependencies — both essential in fashion products where subtle geometric nuances define visual identity. Several transformer-based architectures were tested, including **ViT Base Patch16 (ImageNet)**, **BEiT Base Patch16**, and **MAE (Masked Autoencoder)**. Among them, the **MAE-ViT** achieved the most coherent visual embeddings, attributed to its self-supervised image reconstruction mechanism, which effectively captures fine-grained structural and textural details.

3.5.2. Handcrafted Features

Used as a benchmark to evaluate the incremental value of deep learning. This feature set includes:

Feature Type	Methodology	Purpose
Shape	Histogram of Oriented Gradients(HOG), Hu Moments, PCA reduction	Captures geometric proportions (roundness, symmetry) and structural detail (edge and line distribution).

Color	HSV Histogram (50 bins).	Quantifies the dominant color distribution.
Texture	LBP (Local Binary Pattern) and Gabor Filters .	Distinguishes between matte, glossy, or patterned finishes.

Table 1: Handcrafted Features Extraction Details

The final feature matrix (is created by stacking the three standardized and weighted blocks. After iterating through multiple weighting configurations, the optimal balance was achieved with **Shape (3), Color (1), and Texture (1)**, emphasizing geometric structure as the most discriminative component for visual style clustering.

3.5.3. Dimensionality Reduction and Normalization (PCA)

- **Standardization:** All features (both ViT embeddings and the combined Handcrafted Features) were standardized using StandardScaler to achieve zero mean and unit variance, preventing features with larger scales from dominating the clustering metrics.
- **Dimensionality Reduction:** The combined Handcrafted Feature matrix, specifically, was subjected to Principal Component Analysis (PCA). This step was necessary to mitigate the curse of dimensionality, reduce noise inherent in the high-dimensional handcrafted data, and improve the computational efficiency of the subsequent Agglomerative Clustering.

3.6. Clustering Algorithm

Three clustering techniques were implemented and compared to ensure interpretability and robustness:

1. **K-Means** – Partition-based method requiring predefined k. Efficient but limited by its rigidity in discovering irregular clusters.
2. **Agglomerative Clustering** – Hierarchical, bottom-up approach generating dendrograms that visually represent aesthetic relationships. Ideal for exploratory analysis.
3. **HDBSCAN** – Density-based algorithm that automatically detects clusters of varying shapes and isolates outliers (unique or experimental designs).

The final clustering method selected for interpretation was Agglomerative Clustering, using cosine distance and average linkage, with k optimized via Silhouette Score.

3.7. Integration with Pricing Data

Price metadata was integrated to transform clusters into business intelligence insights. Each cluster centroid was enriched with:

- Mean, minimum, and maximum price.

This enabled:

- Detection of pricing inconsistencies (outlier prices within similar aesthetics).
- Identification of brand-specific style niches.
- Calculation of the Brand Price Premium, representing price deviation relative to stylistic proximity.

3.8. Visualization and Evaluation

3.8.1. Visualization

The visualization stage focuses on providing **interpretable, business-oriented output**, aligning with the need for interpretability and human-AI collaboration.

Visualization outputs focus on direct aesthetic and hierarchical representation:

- **Image Grid Plots:** Images are displayed in a grid, grouped by cluster, with cluster titles including the **Average, Minimum, and Maximum price** information. This allows for direct visual validation of cluster coherence and integration with economic data.
- **Dendrogram:** Plotted for the Agglomerative Clustering results, visualizing the hierarchical merging process and the proximity relationships captured within the ViT embedding feature space.

3.8.2. Evaluation Metrics

The evaluation focuses on two dimensions:

1. Technical Coherence (Quantitative):

- **Silhouette Score:** Measures how well each sample fits within its assigned cluster, balancing internal cohesion and inter-cluster separation. The algorithm iteratively tested multiple k values to identify the optimal number of clusters. Instead of selecting the absolute peak, the final configuration followed a robustness criterion: choosing the smallest $k > 4$ whose Silhouette Score remained within **95% of the maximum value**. This approach prevents overly coarse partitions while preserving a near-optimal balance between cohesion

and separation. Despite modest absolute values—common in high-dimensional visual embeddings—the stability across iterations indicates a consistent and interpretable structural segmentation of styles

- **Cluster Separation Ranking:** To complement the silhouette analysis, clusters were ranked by their **minimum distance to the nearest neighbouring cluster**, providing an interpretable measure of distinctiveness across groups. These ranking highlights which clusters are most isolated (i.e., stylistically unique) within the learned representation space.

Metric	Description	Result / Observation
Silhouette (peak)	Maximum silhouette across $k \in [3,20]$.	k = 4 → 0.0853
Silhouette (selected k)	Rule: pick $k > 4$ with score $\geq 95\%$ of the peak.	K = 17 → 0.0841 (98.6% of peak)
Top clusters by separation	Ranking by minimum distance to nearest neighbor (higher = more isolated).	Rank 1: Cluster 10 → <i>min-sep</i> 47.79 (nearest: 5) • Rank 2: Cluster 5 → 40.86 (13) • Rank 3: Cluster 8 → 40.66 (14)
Per-cluster separation (means)	Mean inter-cluster distances (global distinctiveness indicator).	E.g., Cluster 10 → mean = 58.82 ; range across clusters ≈ 38.3 – 58.8

Table 2: Quantitative Results of Bottega Veneta’s Agglomerative Clustering

Note: For more details about the result check this [link](#)

2. Business Interpretability (Qualitative & Correlative):

- Visual validation of cluster coherence
- Correlation between style clusters and average price levels.

3.9. Reproducibility and Ethics

All scripts were version-controlled via **GitHub**, ensuring full reproducibility.

All images were obtained from **publicly available brand websites** for academic purposes. The study complies with ethical AI research principles, ensuring **transparency**, **non-commercial use**, and promoting the responsible application of artificial intelligence in the fashion domain.

4 Chapter 4: Results and Analysis

4.1. General Overview

Research shows that the agglomerative clustering using Vision Transformer embeddings grouped products based on visual similarities. These groupings highlighted patterns in frame geometry, lens colors, and overall style rather than just relying on color or shape. Evidence indicates the model captured representations at an aesthetic level, separate from background noise. This outcome supports the effectiveness of the preprocessing steps, including background crop and image centering. Although segmentation occurred separately for each brand, comparisons across brands uncover clear differences in styles variety, price stability, and cluster tightness. Such differences serve as indicators of each brand's visual identity.

4.2. Bottega Veneta

4.2.1. Visual Cohesion and Cluster Patterns

Clustering for Bottega Veneta reveals a diverse stylistic structure despite a relatively narrow price range within the mid-to-high segment (€270–€530). The clusters separate clear geometric families—rectangular and cat-eye acetate frames in neutral tones, aviator silhouettes with metallic detailing, and oversized rounded models in darker palettes. Several subclusters exhibit strong internal cohesion in both form and material, indicating that the Vision Transformer embeddings prioritized **frame geometry and structural contours** as dominant dimensions for separation.

4.2.2. Economic Interpretation

Prices stay consistent across clusters tied to specific shapes, pointing to aesthetics as the main driver of price differences rather than materials. Rounded frames and tortoiseshell options—often associated with craftsmanship and brand identity—appear in the higher price segments. The most expensive pair (€530) belongs to the cluster most distant from all others in the hierarchical structure, suggesting that distinct aesthetic signatures are reflected not only in visual separation but also in

market positioning. The model thus revealed visual elements linked to perceptions of luxury without any direct supervision on pricing

4.2.3. Strategic Insight

These findings provide a visual synthesis of the brand’s stylistic coherence, valuable for merchandising and assortment planning. Bottega Veneta’s clusters reveal a consistent aesthetic language built on balanced proportions, clean acetate finishes, and understated luxury across price points. This homogeneity suggests a deliberate design strategy, where subtle variations in shape or tone differentiate products without fragmenting the overall identity. Such insights can help identify recurring design families and prevent overlap within similar style segments

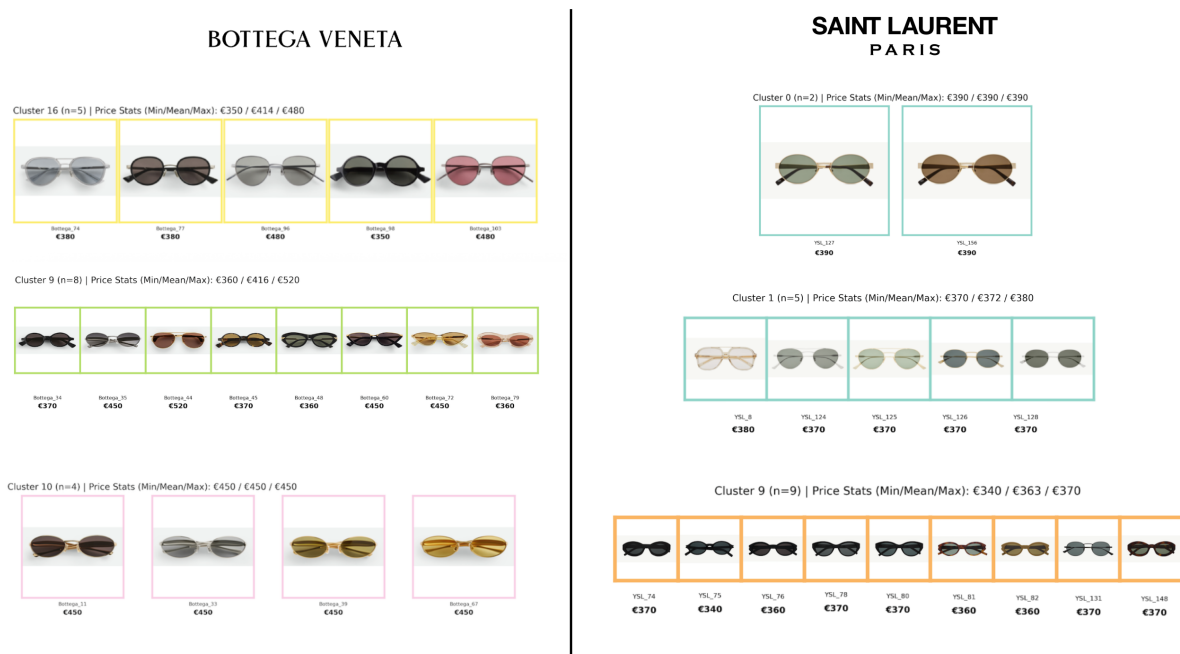


Figure 2: Clustering results from Bottega Veneta and YSL

4.3. Yves Saint Laurent (YSL)

4.3.1. Visual Cohesion and Cluster Patterns

Clustering for YSL reveals strong stylistic uniformity across clusters, dominated by dark palettes, thick acetate structures, and angular or cat-eye silhouettes. The Vision Transformer embeddings captured this cohesion effectively, resulting in compact, visually consistent groupings. Across clusters, variations occur mainly in subtle lens gradients or metallic accents rather than in shape, reflecting YSL’s disciplined approach to design and its emphasis on timeless, sculptural eyewear forms.

4.3.2. Economic Interpretation

Despite a price range between €270 and €750, YSL's clusters exhibit minimal internal price variation, especially compared to brands like Prada or Bottega Veneta. Prices appear more tied to overall brand positioning than to differences in frame design. Nearly identical black acetate models appear across both mid- and high-priced clusters, suggesting that YSL prioritizes aesthetic continuity and brand coherence over experimental form diversification.

4.3.3. Strategic Insight

This uniformity creates opportunities for merchandising and pricing strategies to introduce subtle variety without compromising brand identity. For consumers, the visual consistency strengthens brand recognition and status perception; however, it may also lead to perceptual redundancy in digital environments, where multiple similar models appear indistinguishable within online assortments.

4.4. 4.3. Prada

4.4.1. Visual Cohesion and Cluster Patterns

Prada's clustering reveals a strong emphasis on geometric regularity and minimalist color schemes. The model prioritized structural features—rimless, geometric, or oversized frames—over chromatic variation, confirming shape as the brand's defining visual dimension. Only four clusters were generated, suggesting a more streamlined and homogeneous product offering compared to other brands. This limited spread may reflect Prada's focus on consistent, universally appealing silhouettes rather than extensive stylistic diversification.

4.4.2. Economic Interpretation

The price range between €310 and €520 aligns closely with the degree of design detailing. This pricing pattern reinforces Prada's structured and hierarchical product strategy, where simplicity and material choices establish clear visual and economic distinctions across the assortment.

4.4.3. Strategic Insight

From a business perspective, the visual groupings offer potential for developing price sensitivity models and conducting cross-brand comparisons. The strong alignment within geometric clusters highlights Prada's deliberate focus on modern lines and precise construction techniques. Consequently, Prada's embeddings are particularly

suitable for cross-brand clustering experiments, serving as a benchmark for structural coherence and design discipline.

4.5. Dolce & Gabbana

4.5.1. Visual Cohesion and Cluster Patterns

Visual clustering organizes Dolce & Gabbana's portfolio into four main families, primarily centered on slim rectangular acetate frames. Within this dominant geometry, the embeddings separate products mostly by lens treatment—light gradients versus solid black lenses—creating one prevailing form with limited finish variations rather than multiple design directions. This compact result indicates a focused and less diversified offer, where consumers perceive subtle distinctions but also notable visual overlap. Data shows slightly higher prices for gradient-lens models, reinforcing their link with refined finishing. Strategically, introducing an additional family with alternative geometries—such as round, panto, or angular hybrids—could enrich variety while preserving the brand's premium craftsmanship and material narratives.

4.5.2. Economic Interpretation

Blackout options operate as entry-level models, while gradient lenses represent upgraded tiers. Prices across clusters range from approximately **€200 to €800**, confirming that finish quality and visual treatment drive differentiation more than material changes. Merchandising strategies can make this hierarchy explicit, while pricing policies should maintain and protect it.

4.5.3. Strategic Insight

Large clusters with close visual similarity likely contain several interchangeable products, suggesting potential risks from excessive overlap. Sell-through data could inform selective reductions, focusing on less distinctive items while preserving key models and limited color variations without narrowing the overall line. The absence of round, curved, geometric, or mixed-material groups highlights gaps in the current portfolio. Expanding through genuine differentiation—introducing new shapes, metallic structures, or distinctive design details—would add value without compromising the performance of the core rectangular segment. For promotions, entry-level blackout models could sustain deeper discounts or bundle offers, while gradient versions should retain their premium positioning through exclusive shades, packaging, or added-value incentives. The existing price architecture functions effectively and should be maintained to preserve perceived brand value.

4.6. Fendi

4.6.1. Visual Cohesion and Clustering Patterns

The Fendi clustering resolves into four well-defined visual families, anchored by thin rectangular acetate frames with systematic variations in lens finish. The analysis also identifies a softer rectangular-oval group and a smaller yet clearly distinct shield or sport line, where lenses extend across broader surfaces and the frame adopts a wraparound structure. Within the acetate family, differentiation emerges less from micro-branding and more from lens opacity and subtle surface treatments, producing sequences of nearly interchangeable models that differ primarily through tint, translucency, and tonal nuances. The internal consistency of each cluster reinforces the perception of cohesive style sets, while the shield group represents an alternative usage occasion, appealing to a distinct shopper mission within the same brand universe.

4.6.2. Economic Interpretation

The price data embedded in each cluster title reveal a structured and consistent ladder within the €290–€470 range. Light or gradient lenses within the rectangular families show higher average prices than blackout versions, while the shield or sport clusters occupy the upper tier with narrower dispersion—consistent with their performance-driven or fashion-forward positioning. The largest rectangular clusters exhibit the widest internal price spans, a common pattern when numerous SKUs replicate a single aesthetic through minor finish variations: the average selling price stretches, yet the central level remains stable. The recurrence of mid-sized gradient clusters indicates a sustained willingness to pay for surface finish and optical effects rather than for shape novelty, whereas blackout rectangles cluster around the lower end, functioning as accessible entry points where scale and volume can be optimized.

4.6.3. Strategic Insights

Strategically, rectangles with gradient lenses should be treated as margin engines to defend, blackout versions as entry tiers to drive volume, and shields as premium capsules to highlight selectively rather than over-assort. Rationalization opportunities emerge within the largest rectangle clusters, where substitutability is highest; trimming slower variants would simplify the line without reducing perceived variety. Buying and inventory decisions can follow the family logic—stronger commitments for the two main rectangle clusters, and more selective bets for shields—while promotional efforts should favor bundles or limited-color editions for gradient models to protect price integrity, reserving deeper discounts for blackout entries. For incremental growth, the next expansion should target a clearly distinct visual family—

metal or mixed-material compositions, or deliberate round and panto silhouettes—launched as a small capsule to test reach without cannibalizing the rectangle core. On digital platforms, integrating cluster names as navigational facets could streamline discovery and reinforce the perception of coherent, data-informed curation.

4.7. Cartier

4.7.1. Visual Cohesion and Clustering Patterns

The Cartier clustering resolves into a set of distinctly separated visual families rather than a single dominant silhouette. The analysis reveals coherent groups for rimless and semi-rimless metal frames with lightly tinted lenses, shield and wraparound designs with continuous color gradients, classic pilot and aviator metals, soft rectangular acetates, and a smaller yet clearly independent cat-eye cluster. Within each family, internal cohesion is strong: rimless models organize by lens tint and bridge detailing; shields by lens hue and visor curvature; acetates by frame thickness and finish. Across families, separations are clearly defined—rimless metals and shields sit far from acetates and cat-eyes—indicating that the algorithm captures real, perceptible distinctions consistent with human visual categorization. The presence of long series of near-identical items within certain groups, particularly the rimless and shield families, points to micro-clusters of close substitutes differing mainly in lens tone or subtle finishing cues.

4.7.2. Economic Interpretation

Economically, the cluster-level price statistics reveal a steep premium ladder ranging from approximately €450 to €19,000, underscoring Cartier’s positioning at the intersection of eyewear and fine jewelry. Prices align closely with material sophistication and construction complexity: rimless and metal families consistently occupy the upper range, while shield designs with mirrored or uniform dark lenses reinforce the luxury segment through both form and finish. Classic acetate rectangles define the entry tiers, whereas lightly tinted rimless and colored shield clusters populate the mid-to-upper brackets. The repetition of near-identical color or tint variations within certain families indicates high substitutability and, consequently, strong promotional sensitivity—a discount on one shield tone risks cannibalizing adjacent SKUs. By contrast, cat-eye and pilot clusters retain distinctiveness and can sustain firmer pricing. Overall, the wide dispersion of means across clusters confirms that Cartier effectively monetizes craftsmanship and material hierarchy—translating its jewelry-making heritage into eyewear through the use of precious metal finishes, refined lens treatments, and signature structural detailing.

4.7.3. Strategic Insights

Strategically, clusters can function as distinct “collections” in navigation and storytelling—**Rimless Tints, Shield Visors, Pilot Metals, Soft Acetates, and Cat-Eye Icons**—guiding customers directly into the visual family they already seek. Recommendation systems should prioritize intra-family alternatives before subtly suggesting adjacent, slightly higher-priced groups. To preserve Cartier’s premium perception, rimless metals and shield lines should be protected through limited editions, material upgrades, or exclusive packaging rather than broad markdowns. Promotional activity must remain tightly contained within dense color or tint series, enabling targeted sell-through without eroding the premium hierarchy. Micro-cluster analysis can support pruning of slow-moving, near-duplicate variants, consolidating demand toward clear hero SKUs. Open-to-buy resources should then be redirected to genuinely distinctive forms—such as a refined cat-eye capsule or a high-spec pilot collection—where visual distance justifies higher margins. Ultimately, the clustering analysis indicates that Cartier’s commercial advantage lies in clarifying its family architecture, defending premium tiers, and reducing internal redundancy so every SKU earns its place in the lineup.

4.8. Business Implications

The results of this research position visual clustering as a scalable and data-driven tool for **strategic decision-making** in the luxury fashion industry. The integration of Vision Transformer-based embeddings with economic metadata transforms aesthetic perception—traditionally qualitative—into **quantifiable business intelligence**.

From a **merchandising and assortment** standpoint, the generated clusters function as a *visual-economic map*, enabling not only the identification of internal redundancies (overlapping styles that dilute distinctiveness) and gaps (underrepresented design families) within a brand’s portfolio, but also facilitating **cross-brand aesthetic benchmarking**. By mapping products across different labels within the same visual feature space, the system allows brands to **compare their current aesthetic offering with that of competitors**, identifying converging style territories or potential differentiation opportunities.

From a **pricing perspective**, the association of cluster-level visual identity with average price data facilitates the identification of a **Brand Price Premium**—the measurable value a brand commands due to its aesthetic consistency and distinctiveness. This approach supports refined pricing ladders and helps preserve perceived value across product tiers, aligning visual uniqueness with economic positioning.

From a **CRM and personalization** viewpoint, the methodology expands traditional segmentation—based on behavioral or transactional data—into a new dimension: **visual preference modeling**. By embedding aesthetic patterns into customer profiles, brands can deploy:

- **Cross-category upselling**, recommending visually consistent items across different product lines (e.g., eyewear, handbags, or shoes sharing similar silhouettes or materials).
- **Look-alike recovery flows**, suggesting highly similar alternatives to users who abandoned their carts.
- **Dynamic retention campaigns**, emphasizing new releases or limited editions that align with a customer’s dominant aesthetic cluster.

4.8.1. Support for the Consumer: Accelerating Evaluation Time

From the consumer perspective, personalization mechanisms must evolve beyond textual descriptions or past purchase data. While current systems—such as chat agents, behavioral recommendation engines, and dynamic pricing models—can tailor offers based on user actions, they fail to capture the **visual aesthetics** that drive emotional connection and purchase intent in fashion (Harvard Business Review, 2024).

This study proposes a **visual similarity system** based on Vision Transformer (ViT) embeddings capable of identifying and recommending products that reflect the consumer’s individual aesthetic preferences. By filtering out visually irrelevant options, such a system can significantly **shorten the cognitive evaluation time** and **reduce frustration from digital choice overload**—a factor that currently impacts up to 74% of online shoppers.

Ultimately, the integration of aesthetic AI enhances both **decision efficiency** and **brand affinity**, creating smoother, more intuitive experiences where users interact with products that feel personally relevant rather than algorithmically generic.

Appendix A presents the first seven clusters for Bottega Veneta, ordered by descending pairwise distance to illustrate separation and distinctiveness among groups.

Appendix B displays the dendrogram for Cartier, providing a visual representation of the hierarchical relationships and linkage structure within the clustering results.

The remaining Outputs can be reviewed at the following [link](#)

5 Conclusions & Future Works

This thesis demonstrated that **Vision Transformer (ViT) embeddings**, when combined with **Agglomerative Clustering**, can effectively organize luxury eyewear products into **aesthetically coherent clusters**. These clusters capture not only geometric and chromatic similarities, but also higher-level stylistic cues aligned with each brand's visual identity. The integration of price and brand metadata extends this visual organization into a **multidimensional analytical model**, linking design coherence with economic positioning. The findings confirm that **aesthetic clustering** is both interpretable and operationally relevant, enhancing recommendation systems without requiring manual tagging.

However, the study also acknowledges limitations. The analysis was conducted on a single product category (eyewear), which, while suitable for visual analysis, restricts the generalization of results across diverse categories. Additionally, the reliance on static web-scraped images omits context such as human modeling, background, or lifestyle integration—elements that strongly influence perception in fashion imagery.

Future research will expand the dataset to include additional brands (e.g., Giorgio Armani, Gucci) and product categories (e.g., handbags, shoes, ready-to-wear) to validate embedding robustness and cross-category consistency.

Further work will focus on aligning cross-brand embeddings for competitive benchmarking, integrating temporal metadata to trace aesthetic evolution over time, and deploying the model through an interactive web dashboard that enables dynamic exploration of clusters by brand, price, or style.

Ultimately, this research demonstrates how AI can enhance fashion by combining computer vision, unsupervised learning, and market data to make creativity and decision-making more analytical, efficient, and visually informed.

6 Bibliography and citations

- 1) Admin. (2023, February 20). *Luxury brands are diving deeper into technology*. Equinox AI Lab. <https://equinoxailab.ai/luxury-brands-are-diving-deeper-into-technology/>
- 2) AI, Data & Analytics Network. (2024, February 16). *How Gucci is Using AI and Data*. AI, Data & Analytics Network. <https://www.aidataanalytics.network/data-monetization/columns/how-gucci-is-using-ai-and-data-part-1>
- 3) Aptean. (2023, April 3). *Machine Learning in Fashion: How it's shaping the industry*. Aptean.com. <https://www.aptean.com/en-US/insights/blog/machine-learning-in-fashion-industry>
- 4) Arthur, R. (2017, December 9). *Louis Vuitton becomes latest luxury brand to launch a chatbot*. Forbes. <https://www.forbes.com/sites/rachelarthur/2017/12/08/louis-vuitton-becomes-latest-luxury-brand-to-launch-a-chatbot/>
- 5) Bravò. (2023, December 11). *Versace Jeans Couture Collaborates with BRAVÒ on AI Generated Denim Campaign*. LBBOnline. <https://lbbonline.com/news/versace-jeans-couture-collaborates-with-bravo-on-ai-generated-denim-campaign>
- 6) Chao, C. M., & Parker, K. K. (2019). Artificial intelligence and fashion marketing: A framework for research and applications. *Journal of Fashion Marketing and Management*.
- 7) Choi, W., Jang, S., Kim, H. Y., Lee, Y., Lee, S., Lee, H., & Park, S. (2023). Developing an AI-based automated fashion design system: reflecting the work process of fashion designers. *Fashion and Textiles*, 10(1). <https://doi.org/10.1186/s40691-023-00360-w>
- 8) Danziger, P. N. (2024, April 24). *LVMH sees the AI challenge for luxury is not technology but the human element*. Forbes. <https://www.forbes.com/sites/pamdanziger/2024/04/24/lvmh-understands-the-ai-challenge-for-luxury-is-not-technology-but-the-human-element/>
- 9) Dior's innovation in fashion through AI. (2023, November 28). - *The Scented Journey* -. <https://luxuryfragrances8.wordpress.com/2023/11/26/the-ai-couturiers-diors-innovation-in-transforming-fashion-through-artificial-intelligence/>
- 10) Dosovitskiy, A. et al. (2021). An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale.
- 11) Dou, J. (2024). The application and challenges of artificial intelligence in the fashion and luxury industry. *Applied and Computational Engineering*, 42(1), 90–96. <https://doi.org/10.54254/2755-2721/42/20230694>

- 12) Elastic. (n.d.). *What are Large Language Models? | A Comprehensive LLMs Guide*. <https://www.elastic.co/what-is/large-language-models>
- 13) Entrupy. (n.d.). *State of the fake report - 2023*. <https://www.entrupy.com/report/state-of-the-fake-report-2023/>
- 14) Fashion Retail Academy. (2023, March 1). *How artificial intelligence is used in the fashion industry*. <https://www.fashionretailacademy.ac.uk/news/how-artificial-intelligence-is-used-in-the-fashion-industry>
- 15) Giovanola, B., Tiribelli, S., Frontoni, E., & Paolanti, M. (2023). Ethical implication of Artificial Intelligence in the Fashion Industry: A comprehensive analysis. *Fashion Highlight*, (2), 22–28. <https://doi.org/10.36253/fh-2497>
- 16) Giri, C., Jain, S., Zeng, X., & Bruniaux, P. (2019). A detailed review of artificial intelligence applied in the fashion and apparel industry. *IEEE Access*, 7, 95376–95396. <https://doi.org/10.1109/access.2019.2928979>
- 17) Harreis, H., Koullias, T., Roberts, R., & Te, K. (2023, March 8). *Generative AI: Unlocking the future of fashion*. McKinsey & Company. <https://www.mckinsey.com/industries/retail/our-insights/generative-ai-unlocking-the-future-of-fashion>
- 18) *Harvard Business Review* (HBR). (2024). *AI: The Secret to a Thriving E-Commerce Business*.
- 19) Haug Partners. (n.d.). *Artificial intelligence and its impact on the fashion industry: Contemporary legal Considerations*. <https://haugpartners.com/article/artificial-intelligence-and-its-impact-on-the-fashion-industry-contemporary-legal-considerations-2/>
- 20) IBM. (n.d.). *Predictive Maintenance - Maximo Application Suite*. https://www.ibm.com/products/maximo/predictivemaintenance?utm_content=SRCWW&p1=Search&p4=43700080617779195&p5=p&p9=58700008757604281&gad_source=1&gclid=CjwKCAjw3P-2BhAEEiwA3yPhwFVPZ4Rc3xPGC6PGRyPmhbvY0Z7QjRRqHdnfKWcOYDveKKv2hAQCOhoC8DwQAvD BwE&gclidsrc=aw.ds
- 21) IBM. (n.d.). *What is a Neural Network?*. <https://www.ibm.com/topics/neural-networks>
- 22) Icreon. (n.d.). *Generative AI in Retail: Lessons from Luxury Brands | Icreon*. <https://www.icreon.com/en/insights/lessons-from-luxury-brands-using-generative-ai#:~:text=For%20example%2C%20Louis%20Vuitton%20utilizes,to%20offer%20tailored%20product%20recommendations>

- 23) In, C. D. (2024, January 3). *Generative AI Fashion Industry: Use Cases & Examples in 2024*. AIMultiple: High Tech Use Cases & Tools to Grow Your Business. <https://research.aimultiple.com/generative-ai-fashion/>
- 24) Innovation Training. (n.d.). *Journey Mapping Workshop*. <https://www.innovationtraining.org/journey-mapping-workshop/>
- 25) Jayan, J. (2024, April 5). *Unlocking New Opportunities in the Fashion Industry with Data Analysis*. PromptCloud. <https://www.promptcloud.com/blog/unlocking-new-opportunities-in-the-fashion-industry-with-data-analysis/#:~:text=Data%20analytics%20allows%20fashion%20brands,shopping%20habits%2C%20and%20style%20inclinations>.
- 26) LeewayHertz. (2023, April 19). *AI in fashion: Use cases, benefits and future trends in the fashion landscape*. <https://www.leewayhertz.com/ai-use-cases-in-fashion/>
- 27) PostIndustria. (n.d.). *AI Clothing Detection: Use cases for fashion and e-commerce*. <https://postindustria.com/ai-clothing-detection-use-cases-for-fashion-and-e-commerce/>
- 28) Sceppacerca, B. (2024). *Artificial Intelligence in the Luxury Fashion Industry: Revolutionizing and Enhancing Design, Creativity, Consumer Experience and Sustainability* [Master's Thesis]. LUISS Guido Carli.
- 29) Takyar, A., & Takyar, A. (2023, April 19). *AI in fashion: Use cases, benefits and future trends in the fashion landscape*. LeewayHertz - AI Development Company. <https://www.leewayhertz.com/ai-use-cases-in-fashion/>
- 30) *The Business of Fashion & McKinsey & Company*. (2025). *The State of Fashion 2025*.
- 31) Tsymbal, T. (2024, August 14). *Generative AI in fashion [2024 Insights]*. Master of Code Global. https://masterofcode.com/blog/generative-ai-in-fashion#Top_Areas_in_Fashion_Where_Gen_AI_Technology_Shines_Success_Stories
- 32) Williams, D. (2018, August 9). *The role of artificial intelligence in luxury fashion*. TOPPAN DIGITAL LANGUAGE. <https://toppandigital.com/translation-blog/role-artificial-intelligence-luxury-fashion/>
- 33) Zakeke. (2024, June 12). *Why Virtual Try-On Fashion is The Future of Shopping*. Zakeke. <https://www.zakeke.com/blog/why-virtual-try-on-fashion-is-the-future-of-shopping/>
- 34) *Zalando & Business of Fashion*. (2025). *The Future of E-Commerce, Powered by Generative AI* (White Paper).
- 35) Chanel invests in AI with first-ever try-on beauty app. (n.d.). *Cosmetics Business*. <https://cosmeticsbusiness.com/chanel-invests-in-ai-with-first-ever-try-on-beauty-app--174709>

A Appendix A

A.1. Bottega Veneta

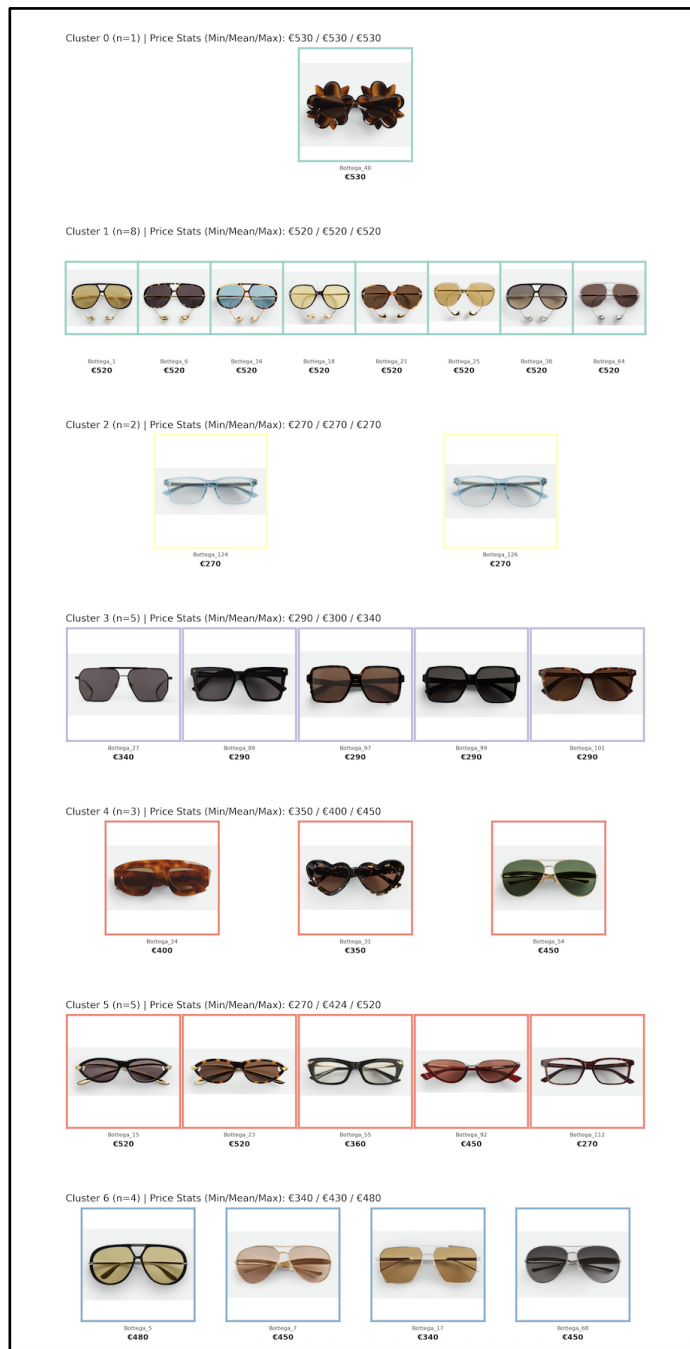


Figure 3: First 7 clusters from the Bottega Veneta Agglomerative Clustering

B Appendix B

6.1. Dendrogram Cartier

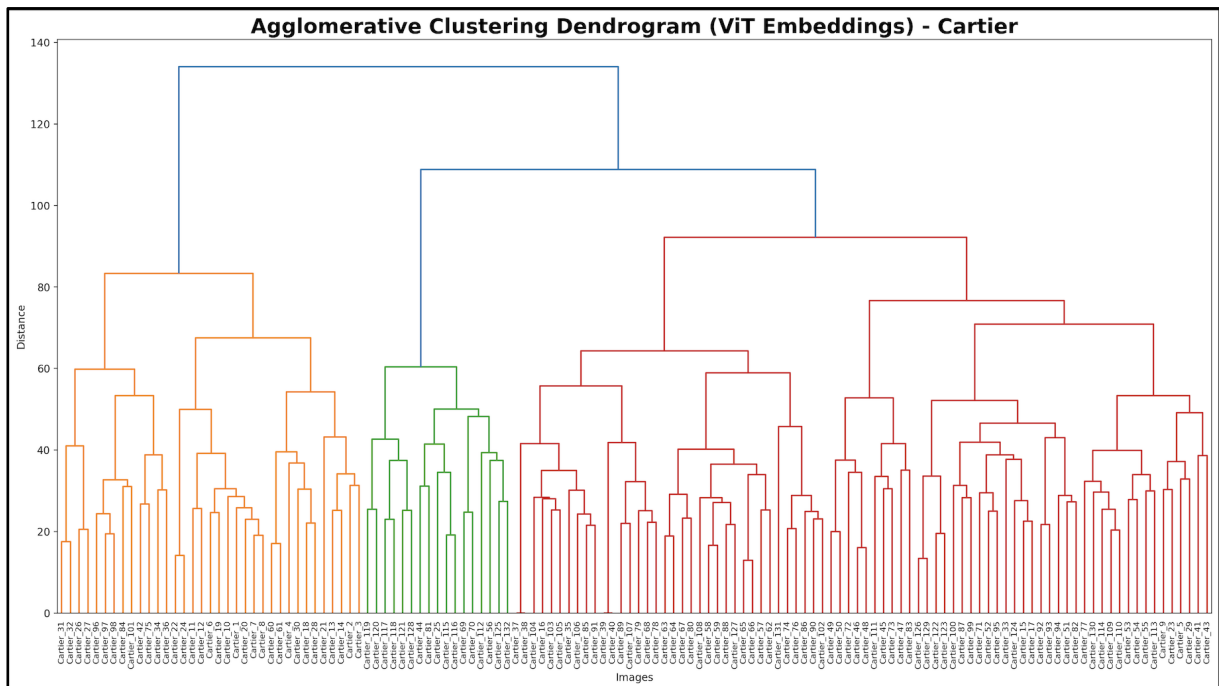


Figure 4: Dendrogram of the Embeddings for Cartier Data

List of Figures

Figure 1: Customer Journey Maps by Innovation Training.....	12
Figure 2: Clustering results from Bottega Veneta and YSL	26
Figure 3: First 7 clusters from the Bottega Veneta Agglomerative Clustering	37
Figure 4: Dendrogram of the Embeddings for Cartier Data.....	38

List of tables

Table 1: Handcrafted Features Extraction Details.....	22
Table 2: Quantitative Results of Bottega Veneta's Agglomerative Clustering	24

7 Acknowledgments

Thanks to Professor Marco Brambilla for the guidance and clarity.

Da parte di Benjamín Cruz Infante, un cileno determinato a trovare la sua opportunità nel mondo della Moda di Lusso a Milano, desidero esprimere la mia gratitudine a coloro che mi hanno aiutato a perseguire questo obiettivo:

- **Marco Brunitto:** per il sostegno, la compagnia e la profonda amicizia.
- **Frank Pagano:** per la disponibilità e la simpatia.
- **Saverio Serafino:** per l'apertura e la costante buona disposizione.
- **Manuela Balli:** per l'importante dimostrazione di fiducia.
- **Benedetta Sceppacerca:** per l'aiuto e la collaborazione nella ricerca.
- **Francesca Maroino:** per avermi offerto un'opportunità