Using Web Data to Reveal 22-Year History of Sneaker Designs

Sungkyu Park
Department of AI Convergence, Kangwon National University
Institute for Basic Science, Republic of Korea, shaun01.park@gmail.com

Hyeonho Song
School of Computing, KAIST
Institute for Basic Science
Daejeon, Republic of Korea
hyun78@kaist.ac.kr

Sungwon Han
School of Computing, KAIST
Institute for Basic Science
Daejeon, Republic of Korea
lion4151@kaist.ac.kr

Berhane Weldegebriel
School of Electrical Eng., KAIST
Institute for Basic Science
Daejeon, Republic of Korea
breteklu@kaist.ac.kr

Lev Manovich
City University of New York
New York City, USA
manovich.lev@gmail.com

Emanuele Arielli
Università Iuav di Venezia
Venezia, Italy
emanuele.arielli@gmail.com

Meeyoung Cha
Institute for Basic Science
School of Computing, KAIST
Daejeon, Republic of Korea
mcha@ibs.re.kr

ABSTRACT

Web data and computational models can play important roles in analyzing cultural trends. The current study presents an analysis of 23,492 sneaker images and metadata collected from a global reselling shop, StockX.com. Based on data encompassing 22 years from 1999 to 2020, we propose a *sneaker design index* that helps track changes in the design characteristics of sneakers using a contrastive learning method. Our data suggest that sneaker designs have been employing brighter colors and lower hue and saturation values over time. We also observe how popular brands have continued to build their unique identities in shape-related design space. The embedding analysis also predicts which sneakers will likely see a high premium in the reselling market, suggesting viable algorithm-driven investment and design strategies. The current work is one of the first publicly available studies to analyze product design evolution over a long historical period and has implications for the novel use of Web data to understand cultural patterns that are otherwise difficult to assess.

CCS CONCEPTS

• **Computing methodologies** → **Image representations**; • **Applied computing** → *Consumer products*; • **Information systems** → Web mining.

KEYWORDS

Sneaker design, neural-net embedding, transfer learning, contrastive learning, global fashion trends, cultural analytics.

ACM Reference Format:
Sungkyu Park, Hyeonho Song, Sungwon Han, Berhane Weldegebriel, Lev Manovich, Emanuele Arielli, and Meeyoung Cha. 2022. Using Web Data to Reveal 22-Year History of Sneaker Designs. In *Proceedings of the ACM Web Conference 2022 (WWW '22), April 25–29, 2022, Virtual Event, Lyon, France.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3485447.3512017

WWW '22, April 25–29, 2022, Virtual Event, Lyon, France.

ACM ISBN 978-1-4503-9096-5/22/04.
https://doi.org/10.1145/3485447.3512017

1 INTRODUCTION

The online reselling market is proliferating. One of the fastest-growing sectors is fashion[1], as noted by Anne-Marie Tomchak, a former Vogue digital director, "there are literally tens of millions of pounds worth of clothes sitting dormant in people's wardrobes." Many reselling markets, such as eBay, Depop, Vinted, and Etsy, have tens of millions of users, turning themselves into excellent platforms for trading fashion items. Data accumulated within these sites provide a unique opportunity to study how fashion trends have evolved (or sometimes revolved) over a longitudinal period.

StockX.com is one such reselling marketplace for sneakers. It facilitates auctions between sellers and buyers by verifying the seller's items and shipping them to international buyers. Now considered a multibillion-dollar business, the site offers data insights into street fashion trends via its stock market-like price history of transactions. To characterize long-term trends based on Web data, we collected information about approximately 23,492 sneakers from the site (Figure 1). The product images are utilized in a neural network model to extract a low-dimensional embedding that can explain characteristic design traits by jointly learning color and shape attributes. This embedding, as we demonstrate, excels in several practical tasks, such as the classification of sneaker designs by brand, consumer type, release year, and the resale premium.

Our embedding is generated by a combination of computer vision techniques, including pretrained ResNet-18 and fine-tuning with an unsupervised contrastive learning model. Unsupervised contrastive learning is a machine learning method to classify similar objects by bringing positive samples closer together while pushing negative samples farther away in the embedding space [8]. Despite the potential, conventional contrastive learning methods cannot be used directly in data with multiple attributes. This is because different visual features (such as color and shape) become entangled into a single embedding space [43], hindering further analysis of subjects with multiple visual aspects. We employ two innovative modules to overcome this challenge: multiaugmentation to accommodate subtle variations in image styles (such as flipped, cropped, or color-jittered images) and masking to create disentangled multiple projection heads to represent key visual attributes such as color and shape. As a result, our model can extract low-dimensional disentangled visual semantics, enabling broader applications, such as cultural analytics. Our technical novelty is to decompose the design

[1]Forbes Magazine, 27 June 2021, https://tinyurl.com/y2czc5re.

aspects by embedding and extracting the visual characteristics of sneaker images. We consider a unique method that can apply to other fashion items.

Based on Web data spanning 22 years, the current research makes the following contributions: 1) Feature engineering identifies an overall design change that sneakers have become more pastel-toned and brighter over time; 2) We present a novel neural-net-based embedding that jointly considers color and shape information from the given sneaker images; 3) The proposed embedding can infer key cultural trends of sneakers, including the product category, target consumer, and high reselling premium; 4) Based on the extracted latent representation, we learn that popular sneaker brands are becoming similar in color choice, but more distinct in shape-related design choice. To the best of our knowledge, these patterns observed via large-scale Web data have not yet been studied.

Our research has several implications. We can envision new AI-driven fashion consulting that can assist the industry in managing their data, predict trend trajectories, and propose designs that could be more sustainable (and hence produce less waste similar to AI-driven food design) [34]. The state-of-the-art embedding model can also be used to observe temporal patterns of other human artifacts and cultural products beyond fashion, as it does not require any other metadata besides images. This ensures domain-independent learning and helps analyze long-term historic data.

Implementation details of the model and code are made available via a repository at https://github.com/embSneakers/embSneakers. Please see the Appendix for a description of the data distributions, feature extraction methods, and comprehensive embedding results.

2 RELATED WORKS

2.1 Fashion Trend Analysis

Fashion trends are cultural and social phenomena that have undergone constant changes. In the past, social scientists have tried to study its underlying principles. Some trends are seasonal and short-term, while others are more long-term. It can also be cyclical with periodic revivals that give new life or can be repurposed as vintage. The mechanisms behind fashion trends are multifaceted: they can be both internal to the evolution of an object's shape [26, 29, 36] and the result of external social dynamics, such as people's drive toward differentiation and distinction within a group. Sociologist George Simmel observed that fashion is the result of tension between the individual's desire to conform to the dominant trends in his group and his desire to be unique and stand out from the crowd [37]. Later, Pierre Bourdieu defined the force of distinction as the process explaining how people from a social group conform to specific trends in taste, allowing them to separate themselves from the taste of a different social group from which they want to distance themselves [5].

These mechanisms partly explain how a few connoisseurs and trendsetters first define the so-called "early adopters" of a trend that becomes popular and mainstream over time. At the moment in which many persons follow the trend, this becomes increasingly fashionable for the trendsetters, who then abandon it to pass to something new by setting a new cycle in motion. Trendsetters might then revalue a trend that the majority had abandoned for a long time. This makes the reselling market a viable industry. The so-called hipster effect describes the search for exclusivity in trends, especially by those who like to revive styles of the past, and the fact that a global population of trend-conscious subjects, paradoxically, follows this quest, leading to a counterintuitive and industry-driven synchronization occurring among people who want to express uniqueness via fashion [40].

Many fashion items enjoy a global market, and consumers exchange information via the Web. Trend dynamics are very similar to financial markets, where few stockbrokers act upon exclusive information, followed by the mass of investors, generating a constant fluctuation of stock values. The world of finance has always used systematic and quantitative methods to capture these processes while realizing the difficulty of creating predictive models. In fashion, predictive models are even less systematic, as they are mostly based on the subjective intuition of experts in trends and styles. Fashion forecasting is an area of study and industry with a long history whose effectiveness has been the subject of debate [4]. Only in recent times can we witness the shift to quantitative analysis of data in fashion forecasting [14, 15].

Initially, these studies were based on small datasets, such as images from specific catwalks [21]. Today, the development of computational methods for analyzing large databases offers the possibility of conducting observations using Web-based data and machine learning. Researchers in data science, social science, and digital humanities have already published many studies that use computational methods to analyze changes in style, form, and content in the literature, visual arts, popular music, and mass media over long periods [1, 20]. In the fashion domain, images have been automatically analyzed, for instance, for automated parsing of clothing items [45], product identification, and style classification [22, 25]. Analysis on large databases allows us to discover broad general trends, such as seasonal changes in clothing colors [16, 42], and to detect and predict changes in fashion styles over space and time that escape direct human observation [28, 30].

2.2 Contrastive Self-Supervised Learning

Limited availability of ground-truth data or annotations led to advances in self-supervised learning. Contrastive learning is one such method to learn data distributions. Contrastive learning maximizes the agreement between similar instances (or so-called positive samples) while minimizing the agreement among dissimilar instances (negative samples). For example, SimCLR [8] defines an augmented version of an image as positive and the other images in the same batch as negatives. This work utilized image transformations such as color jitter and horizontal flipping, which do not deform the underlying characteristics so that the model can maintain the crucial information during training. MoCo [9] employs a similar framework but utilizes a momentum encoder and introduces a dynamic dictionary with a queue.

Contrastive learning methods adopt the InfoNCE loss [32]. Minimizing this loss maximizes the lower bound on the mutual information between positive pairs, enabling the model to learn invariant features against data augmentation. Due to the high representation power, this concept is being widely utilized in various inference tasks [17, 18]. However, the model becomes sensitive to the augmentations in downstream tasks [39]. Thus, a recent work, LooC [43],

proposed separating the embedding space and applying different data augmentation techniques on each head, which successfully improves the usability of the embeddings.

3 PROBLEM AND DATA

3.1 Research Questions

Our research questions are as follows. First, what can we learn from the Web data about how sneakers as a mass fashion item have evolved? (Section 3) Second, what models can explain latent dimensions of formal and aesthetic trends from a large dataset? (Section 4) Third, to what extent do embeddings of color and shape differ by product type and reselling premium? (Section 5) Fourth, what can we say about brand identity based on analysis? (Section 6) These questions are essential for understanding the evolution of street fashion and deciding future designs and investment strategies. We make use of the Web data to answer these questions.

3.2 Data Description

We collected sneaker images and their descriptive metadata from StockX.com. As shown in Figure 1, we crawled all features offered on the website, including the retail price, resale price history, show images, brand, and release date. We found information for approximately 23,492 sneakers from past years (e.g., the release year of some sneakers goes back to 1985). The transaction data of prices are available from 2012 since the platform's launch. The product images are in landscape resolution, so we added white padding to the upper and lower parts to make them square images. We then resized the images to 256×256 pixels to reduce the dimensions for feature engineering and for using the ImageNet pretrained network (i.e., ResNet with 18 layers). Pretraining with a large-scale dataset ensures that the model captures generalizable visual characteristics and helps create high-quality indices [10].

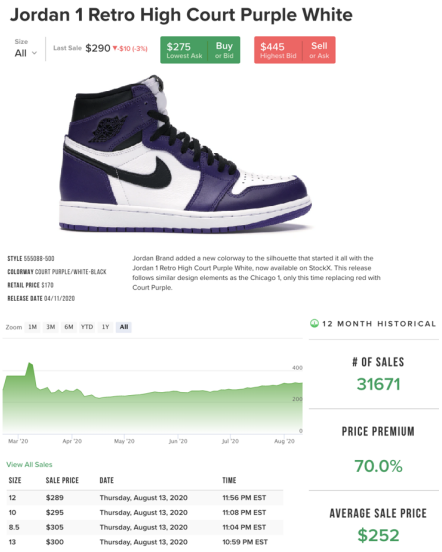


Figure 1: Snapshot of the information shown on StockX.com, indicating product details and transaction price history. ©Photo by StockX, taken on October 20, 2021.

Prior to analysis, we removed any products without proper shoe images, such as only showing a shoebox. The final data constitute 11.0 gigabytes, including both images and metadata for 22,331 sneakers with valid images. We list the frequency distributions of the retail price and reselling premium across brands and release times in Figure 7 in the Appendix.

3.3 Metadata Exploratory Analysis

We examined the scale and temporal patterns seen in the metadata of sneakers. Figure 2(a) shows that the product count and the brand count on the reselling market have increased rapidly over the past decade, as indicated by the exponential growth of the business. Since the spring of 2019, quarterly transactions have reached a million scale on the platform, and more than 10,000 unique items have been sold every quarter (see Figure 8(a) in the Appendix). We define profit or resale premium as $price_{resale} - price_{retail}$ per transaction where every price was adjusted by accounting for inflation of the USD (see Section B in the Appendix on how the adjustments are made).

As more sneakers are sold via the platform, the resale premium per transaction tends to decrease. This trend is shown in Figure 2(c), denoted by the median reselling premium of sneakers. A typical transaction leads to a USD \$60 ~ \$80 profit to the reseller, assuming the reseller bought the product at retail price. However, the top 10% of sneakers with the highest premium transactions led to a \$320 ~ \$400 premium per sale, pointing to a substantial profit considering the retail price of sneakers. The resale premium also tends to increase over the sneaker's age (i.e., release date), as shown in Figure 2(b).

3.4 Image Exploratory Analysis

We employed feature engineering to examine temporal changes in sneaker designs. We first segmented each image using a standard unsupervised image segmentation method by backpropagating partitions of an image into groups of pixels containing similar traits [24]. Segmentation can be a useful method to systematically remove the white background. Figure 3 shows one example with the identified segments. We then extracted the following features on the color attributes, distribution parameter, histogram, and entropy from the background-removed images. Details of this method are described in Section A in the Appendix.

Figure 3(b) shows changes in the HSV trends for sneakers released in the past five years. The plot shows quarterly averaged trends. The hue value H gradually decreases, indicating that larger proportions of sneakers come in colors in the yellow-orange wheel. The saturation value S shows an overall decrease with seasonal fluctuations, meaning that sneakers have become more pastel-toned. These patterns may be attributed to the fact that sneakers targeting female consumers have grown in proportion.[2] Alternatively, sneakers now adopt more novel colors than the blue-green wheel to appeal to more diverse customers. Meanwhile, for the brightness value V, more distinctive seasonal variations can be observed, with a gradual increase in value over time. Specifically, sneakers tend to become darker when targeted for the winter season in the Northern Hemisphere (Q4) and lighter when targeted for the summer season

[2]A report by ForwardPMX shows that the female sneaker market grew five times faster than the male market from 2016 to 2017 in https://bit.ly/3m1XoD7.

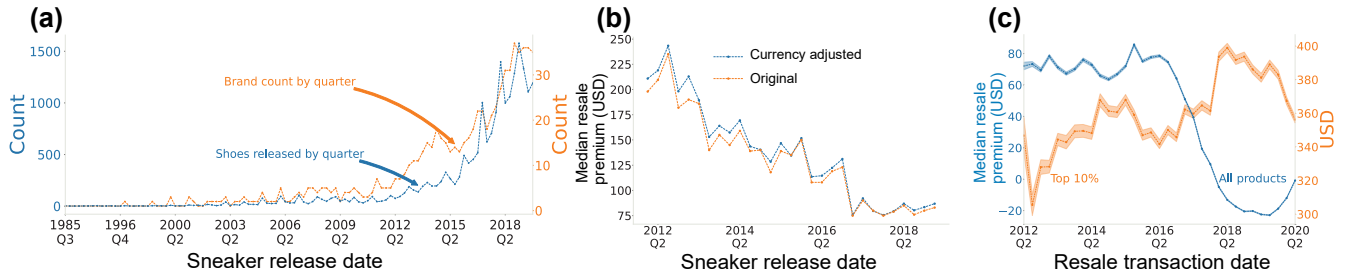
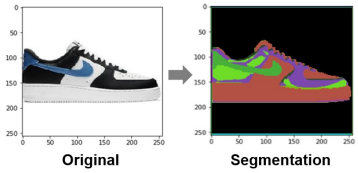
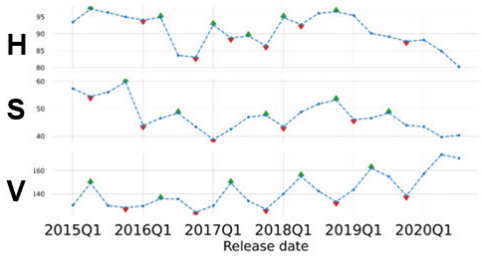


Figure 2: Data trends: (a) Quarterly item and brand counts based on the release date; (b) Quarterly median resale premium of transacted sneakers based on the release date; (c) Quarterly median resale premium for all products and top 10% based on the transaction date, with 1.0 standard error: prices were adjusted for inflation of US currency (USD). Note that the transaction logs are available only from the platform's launch year in 2012, whereas the sneakers' release year may go back further.



(a) Unsupervised image segmentation example. © Photo by StockX



(b) Quarterly color trends of released sneakers in 2015–2020.

Figure 3: Segmentation example and HSV color model constructed within the extracted segments

(Q2). Note that products released in a given quarter typically target the forthcoming season.

Our data analysis demonstrates that publicly available Web data gathered from a reselling market can be used to analyze the design evolution and those of high premium sneakers. However, one drawback of such an approach is that the derived features may not fully represent the data since each feature captures limited aspects of the products. We next investigate how to extract latent representations or embeddings from the given sneaker images by minimizing the loss of information while reducing the dimensions.

4 MODEL

The visual features of sneaker designs are learned via a neural network model, as illustrated in Figure 4. We developed an embedding model with two functions: (1) a multiaugmentation contrastive learning framework and (2) a masking module for disentangling representation. We explain conventional contrastive learning and introduce the two innovations added in this work.

Unsupervised contrastive learning is a self-supervised method that does not rely on ground-truth labels. Contrastive learning assumes a pretext task that gathers representations of similar instances (called positive samples) closer in the embedding space while pushing away representations of dissimilar instances (called negative samples). One may consider variations of the same sneaker image (e.g., enlarged, cropped, flipped) as positive samples and consider two separate sneaker images as negative samples. This method can hence efficiently learn invariant features against data augmentation. Formally, let $\mathcal{D} = \{\mathbf{x}_k\}_{k=1}^{N}$ denote a set of training images $\mathbf{x}_k$. Contrastive learning introduces the InfoNCE loss [8, 19], which is defined as follows:

$$L_{\mathrm{CL}}(\mathbf{z}_{(1)}, \mathbf{z}_{(2)}) = -\log \frac{\exp(\mathrm{sim}(\mathbf{z}_{(1)}, \mathbf{z}_{(2)})/\tau)}{\sum_{\mathbf{z}' \in \hat{\mathcal{B}}^{(-)}} \exp(\mathrm{sim}(\mathbf{z}_{(1)}, \mathbf{z}')/\tau)}, \quad (1)$$

where $\mathbf{z}_{(1)}$ and $\mathbf{z}_{(2)}$ are latent representations of two different views of the target image $\mathbf{x}_k$ and $\mathrm{sim}(\cdot)$ is the cosine similarity function. τ is a temperature parameter (τ=0.5 in our case) that controls the entropy of output: it should be positive, nonzero, and below 1.0 [8]. $\hat{\mathcal{B}}^{(-)}$ is the set of latent representations from batch images excluding $\mathbf{x}_k$. Optimizing this loss (i.e., pulling positive samples closer and pushing away negative samples) maximizes the lower bound on the mutual information between two different views [39].

A model must accurately learn the color choice's distinct effect and design sketches to handle fashion data. Therefore, the conventional method is not suitable, as the learned representations lead to "feature entanglement", where different visual features are entangled into the single embedding space, as identified by [43]. This inherent limitation led us to consider feature-specific heads. We added different nonlinear projection heads for important design aspects, focusing on the shape-invariance head and the color-invariance head, as illustrated in Figure 4. Each head enables the model to focus on a specific design aspect during training. Every training head takes a different type of augmented view pair as input. We generate multiple views for a given image $\mathbf{x}$, by adopting different augmentation strategies. For example, a predefined set of images T_{shape} consists of transformations that change only the shape information of the given image to train the shape-invariance embedding. Similarly, T_{color} includes a set of color-related transformations. We used random resized crops and random horizontal flips for T_{shape} and random grayscale and color jitter for T_{color}. T_{all} is the union of

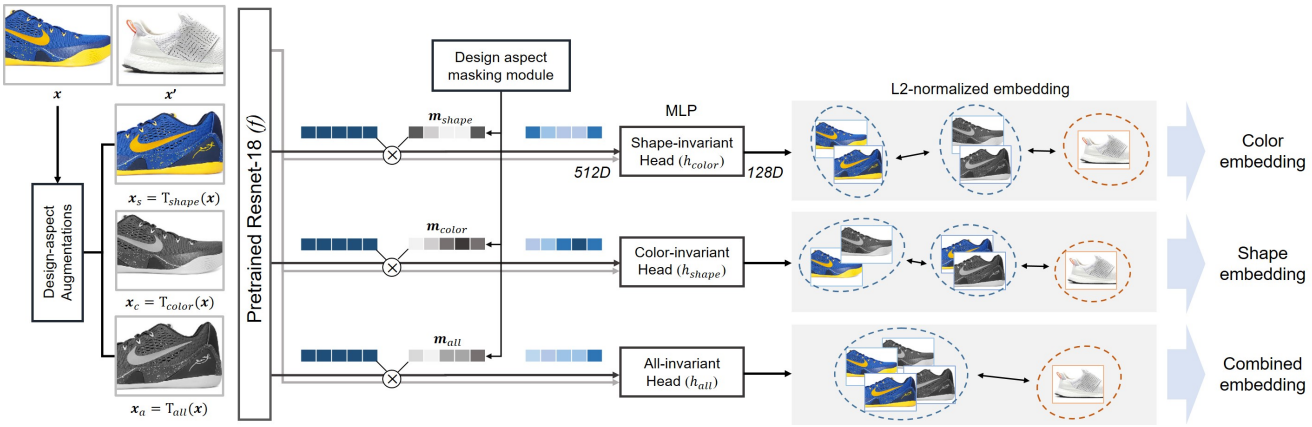


Figure 4: Illustration of the design embedding model. Shape, color, and combined attributes information become disentangled via the newly added masking module. The contrastive objectives for different heads (i.e., h_{shape}, h_{color}, and h_{all}) are jointly optimized in an end-to-end fashion. The shape-invariant head groups sneaker images by all features except for shape and, in turn, allows the embedding to contain (mostly) color-related information (i.e., color embedding shown at the top right corner). On the other hand, the color-invariance head groups images by all features except for color, which allows the embedding to now learn (mostly) shape-related information (i.e., shape embedding at the middle). The all-invariance head can accommodate augmentations in both shape and color (i.e., combined embedding at the bottom).

the shape and color transformations ($T_{all} = T_{shape} \cup T_{color}$) used to train both shape- and color-invariance embedding.

We also employ a masking module to enhance the feature disentanglement ability. The masking module has trainable parameters and provides a weight vector that helps the model concentrate on each design aspect. This is similar to the original attention technique [2, 41]. The masking module produces three different masking vectors by applying the sigmoid function and making the scale from 0 to 1 ($\mathbf{m}_{shape}$, $\mathbf{m}_{color}$ and $\mathbf{m}_{all} \in \Re^{512}$). Pointwise (or element-wise) multiplication with these masking vectors is performed by the encoder f so that the model can focus on a specific aspect of the input image for three different projection heads:

$$\mathbf{z}^i = h_i(\mathbf{m}_i \odot f(\mathbf{x})), \quad i \in \{\text{shape, color, all}\}, \tag{2}$$

where $\odot$ is the pointwise multiplication operator, and h_i is the projection head for each design aspect (i.e., shape, color, and all). The loss objective for the projection head is denoted as follows:

$$L = \sum_i L_{\text{CL}}(\mathbf{z}^i_{(1)}, \mathbf{z}^i_{(2)}), \quad i \in \{\text{shape, color, all}\}. \tag{3}$$

This model does not require any domain-specific features or labels. Hence, it can embed design patterns for other fashion items such as handbags or hats. The only requirement is an ample set of images for training. In addition, the proposed end-to-end method automatically extracts semantically meaningful features when discriminating images and is suitable for use as an innate design index. We next consider several practical inference tasks to test the model's efficacy in the following section. Hereinafter, we refer to the embeddings from each shape-invariance, color-invariance, and all-invariance head as the color embedding, shape embedding, and combined embedding, respectively.

5 EVALUATION

We evaluated the embedding model first with four inference tasks and then via data visualization.

5.1 Quantitative Evaluation

We considered three hypothetical classification tasks based on data attributes and transaction logs. In StockX.com, sneakers are labeled as belonging to one of 16 product categories, from which we choose the top 8 classes by frequency. These categories include Adidas (N=4,123), Air Jordan (3,675), Air Max (2,839), Nike Basketball (1,256), Air Force (1,213), Nike SB (816), LeBron (657), and Kobe (428). The platform also assigns seven types of target consumers for each sneaker, from which we chose the top 5 by frequency: men (3,000), women (2,344), children (1,460), preschool (420), and toddler (340). Given that classes are skewed, we chose a similar number of products per class. For the reselling premium, we limited the analysis to only sneakers with at least two transaction records (11,848) and considered a binary class of 'high (top-20 percentile of the premium)' and 'low (the remaining)' considering the heavy-tail distribution.[3] We identified 9,478 sneakers for the low premium group and 2,370 sneakers for the high premium group.

We assessed embeddings through three widely used off-the-shelf classifiers: multinomial logistic regression, XGBoost, and MLP (multilayer perceptron) neural-net model. XGBoost was trained for 100 epochs with early stopping and a learning rate of 5e-2. MLP was trained for 300 epochs with a learning rate of 1e-3 and a mini-batch size of 4. We adjusted the number of layers for each embedding in MLP according to its dimensions (e.g., 4 layers for 12 dimensions and 6 layers for 524 dimensions). The data were split into 60%, 20%,

[3]Premium refers to the difference in the maximum reselling price and the retail price, adjusted by the inflation rate. One may use the average or the median price instead.

and 20% for training, validation, and testing for all tasks. Training details will be released via the repository link later.

Table 1 shows the evaluation results. Due to the page limit, we present the results only for MLP; other results are provided in Section C of the Appendix. As baselines, we ran classifiers based on feature engineering of color, segmentation, and their concatenation. For the concatenation of color and segmentation information, we used the entropy of the color values that yielded the best performance. All features were normalized to z-scores before training and testing. As another baseline from the recent representation learning domain, we used LooC ,which has shown the best performance among contrastive learning models [43].

Our model yielded three representations after the masking module: color, shape, and combined embedding. Following the suggestion that the layer before the projection head has a reasonable quality for downstream tasks [8], we evaluated our model by freezing the backbone network and removing the projection heads. We then fit the classification and regression models on top of the learned representations (512 dimensions). The combined representation obtained from the all-invariance head showed the best results, outperforming all baselines. While LooC only separates the embedding space for each view (or the augmentation transformation) of the images in an intertwined manner, our model further guides the model to extract disentangled representations for each separated head via a masking module.

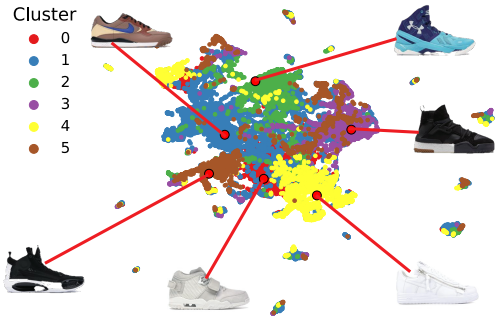
As a result, the proposed model's inference on the product category shows an exceptionally high performance of 0.926 in the F1 score. Inference on the consumer type appears to be a more challenging task, yet our model gives the best result, with an F1 score of 0.578. For predicting the reselling premium, the concatenation of feature engineering yields the best RMSE result of .087, and our model gives an RMSE of .096. However, our model shows the best result in an MAE of .044. These results consistently demonstrate that our embedding contains meaningful information for many downstream tasks. The ablation results shown at the bottom of Table 1 confirm that excluding any component from the model degrades the performance substantially for all three inference tasks. Furthermore, shape embedding was consistently better than color embedding at predicting product categories, target consumers, and high premium items.

5.2 Qualitative Evaluation

We also assessed the embedding by visualizing the clustering results. The strength of our embedding model comes from the use of separate heads. We obtained 128 dimensions of latent vectors on each image from the three heads, color-invariance, shape-invariance, and all-invariance,[4] by projecting each representation into L2-normalized space. We then ran k-means clustering for each embedding and reduced the dimensions by two via UMAP, a nonlinear dimensionality reduction technique that has shown novelty over existing visualization methods such as t-SNE [31]. Figure 5(a) shows the final visualization for the color embedding. The clustering result for the shape embedding is provided in Section D of the Appendix.

[4]As explained in Figure 4, the combined embedding is produced by the all-invariant head and groups sneaker images of the same design, ignoring subtle color- and shape-related augmentations.

To qualitatively examine the images within clusters, Figure 5(b) shows the nearest image samples for each centroid. We can observe distinctive color patterns, where sneakers of similar colors appear to be grouped within the same cluster. The number of clusters for this embedding was determined to be six by the elbow method. The first three authors qualitatively examined images within clusters for the other two embeddings and confirmed that the sneakers are well clustered by their color, shape, or combined attributes.



(a) K-means clustering result of the sneaker embedding for the color attribute.

(b) Examples of sneaker products by cluster for the color attribute.

Figure 5: Centroids within clusters and their 15-nearest neighbors based on color embedding. © Photo by StockX

6 EXPLORING TEMPORAL DESIGN PATTERNS

The extracted embedding allows us to explore temporal changes in designs over a particular subset of data, such as brands, product categories, or specific features. For visual comprehensiveness, we further used principal component analysis (PCA) to reduce the embedding size to a single dimension, which we call the **Sneaker Design Index**, thereby applying 1D PCA on top of the 2D UMAP coordinates previously compressed for clustering. This approach is taken to retain critical information, as reducing the dimension from 512 to 1 at once via PCA may break the local manifold structure (i.e., neighbor relationship) among data samples [3]. Eventually, the final 1D PCA coordinate on top of the 2D UMAP coordinates offers an intuitive comparison of any two groups.

Classifier & Regressor: MLP (Neural-net)		Primary Category			Consumer Type			Maximum Resale Premium				
		N = 15,007 sneakers			N = 7,564			N = 11,848			N = 11,848	
Attribute	Feature	Acc.	F1	κ	Acc.	F1	κ	Acc.	F1	κ	RMSE	MAE
Random	1 / n Classes	.167 (1/8)			.200 (1/5)			.500 (1/2)			N/A (regression)	
Feature Engineering:												
Color	Dist. parameter (12D)	.352	.291	.140	.440	.377	.112	.787	.693	.000	.165	.142
Color	Entropy (7D)	.460	.387	.292	.419	.387	.113	.788	.696	.009	.167	.142
Color	Histogram (128bin, 384D)	.401	.391	.250	.403	.402	.141	.727	.702	.057	.095	.045
Segmentation	Unsp. Image Seg. (5D) [24]	.384	.265	.167	.408	.329	.098	.787	.693	.000	.169	.143
Concatenation	Entropy + Segmentation (12D)	.476	.401	.308	.401	.387	.123	.787	.693	.000	**.087**	.046
Contrastive Learning:												
Color + Shape	LooC (384D) [43]	.882	.884	.854	.547	.548	.356	.767	.760	.238	.166	.142
Color + Shape	**Ours: All-inv. Rep. (512D)**	**.926**	**.926**	**.909**	**.579**	**.578**	**.392**	**.790**	**.785**	**.314**	.096	**.044**
Ablation Study:												
Color	Ours: Shape-inv. Rep. (512D)	.799	.801	.751	.569	.571	.384	.776	.772	.276	.165	.142
Shape	Ours: Color-inv. Rep. (512D)	.839	.836	.801	.623	.623	.458	.800	.793	.334	.168	.143
Color + Shape	Ours: 3 Rep. total (1536D)	.928	.928	.751	.602	.604	.431	.795	.785	.301	.095	.043

Table 1: Results of inferring various classification and regression tasks utilizing MLP: see Table 2 in the Appendix for other classifiers (Acc.: accuracy; κ: Cohen's kappa; Rep.: representation; RMSE: root mean square error; MAE: mean absolute error; the bold row demonstrates the most significant error among the seven compared embeddings; the bottom three rows demonstrate the ablation study within our model).

Figure 6 shows the index trajectory for seven popular brands known for their distinct styles. These brands had released over 500 sneaker kinds each, accounting for a total of 20,195 items in our dataset. The plot shows the trajectory separately for the color and shape embedding. We use data only from years with sufficient observations, limiting the analysis to 22 years (1999–2020). The plot shows a distinct trajectory per brand. For instance, Air Jordan shows a downward trajectory in the color embedding in Figure 6(a), whereas Nike shows an upward. In recent years, all brands' design indices have converged in terms of color embedding.

The shape embedding shown in Figure 3(b), in contrast, indicates a different design evolution. While color trends have become more similar across brands, the shape embedding continues to be similar for most brands, except for Nike, whose shape design trajectory continues to change, although in a consistent direction. The same mechanics can be observed by examining the pairwise similarity of brands. We computed the average pairwise similarity between every two brands over a fixed time span $\mathbf{t}$ as follows:

$$\text{Pairwise-Similarity}^{t}{}_{(i,j)} = \frac{\sum_{\mathbf{x}}^{\mathcal{D}_i^t} \sum_{\mathbf{x}'}^{\mathcal{D}_j^t} sim(x, x')}{|\mathcal{D}_i^t||\mathcal{D}_j^t|}. \quad (4)$$

where $\mathcal{D}_b^t = \{\mathbf{x}_k\}_{k=1}^N$ denotes a set of embeddings of sneaker images $\mathbf{x}_k$ within brand b. $sim(i, j)$ is the cosine similarity of two sneaker images i and j. These plots, binned by 3-months, are shown in Figures 6(c) and (d). The similarity plots show trends for all brands in the dataset, beyond the top seven brands.

The similarity of the color embedding continues to increase over time, reaching a high average similarity of 0.620. Investigating the HSV values of colors per brand, we consistently observe the trends of sneakers adopting brighter colors and pastel-toned designs, as shown in Figure 3(b). The shape embedding shows a lower correlation of 0.390–0.425, although there were some changes over time. The subtle increase in global similarity of the shape is likely caused by the shift in Nike's design, which reduced the distance across all other brands. Other factors may affect the increment in shape embedding distance; for example, some brands have gone through mergers during the studied period (e.g., Adidas acquired Reebok in 2006, which might have led to design changes).

7 DISCUSSION AND CONCLUSION

Summary. The current study presented an unsupervised neural embedding model of a mass fashion item mined from the Web. We jointly utilized color and shape information to embed sneaker designs from an extensive collection of Web images. This process required no label information, and the training was performed end to end. By further reducing the data dimensions, we proposed the **Sneaker Design Index**, which is an intuitive method to track design changes over time and across brands. Our data analysis revealed patterns of convergence and uniqueness in the design of major sneaker design houses over two decades.

Cultural analytics. The methodology and findings presented in this paper have implications for better predicting fashion trends. Our method can assist efforts in existing qualitative or nonneural methods to understand style evolution in mass fashion. Data science methods have practical uses in helping designers capture crucial traits of fashion trends from massive data that are otherwise challenging to find manually. While the use of AI in the fashion industry has focused on converting sales and automating processes, our research is an early effort toward using AI to capture design trends. We envision how the decades of sneaker trend trajectory identified by a neural net may assist the creative process of professional and amateur designers. The trajectory of high premium items may provide insight to designers in perceiving user perceptions of collectible goods. Furthermore, customers can use the information to anticipate new trends and plan investment strategies for collectors.

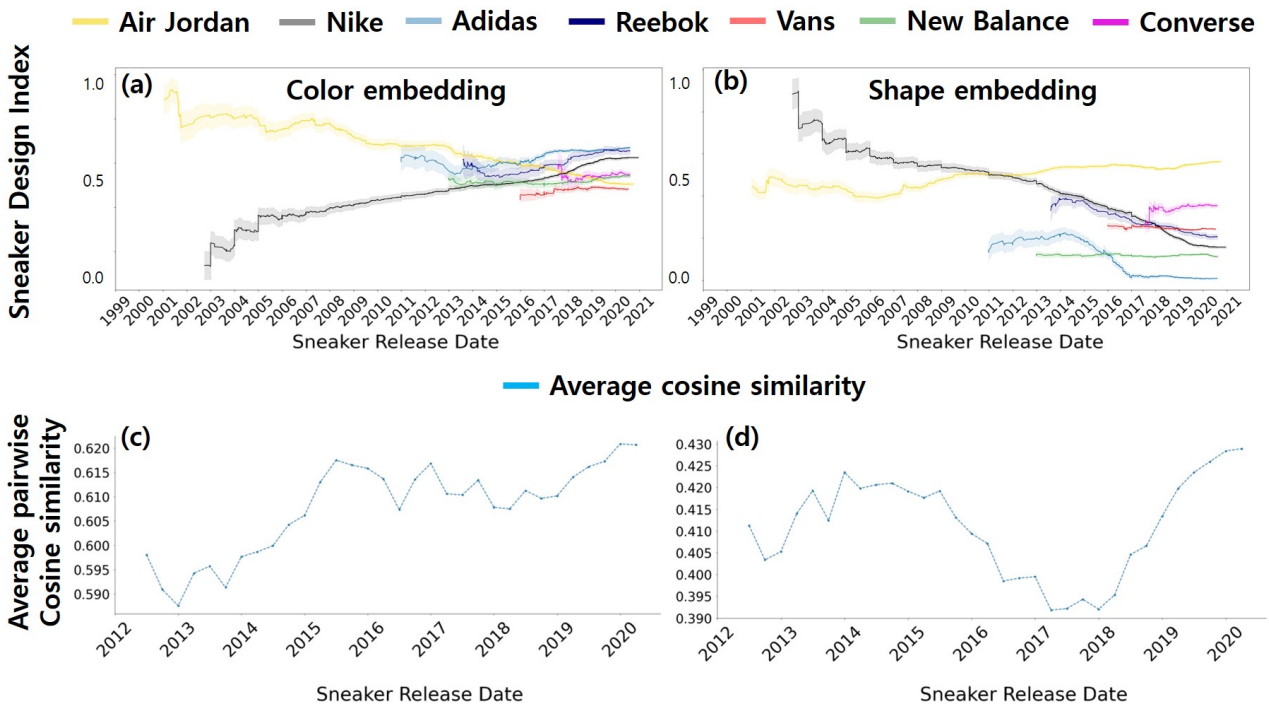


Figure 6: Temporal sneaker design patterns by the brand via embeddings. (a) Color embedding with one standard error shadow (the index values on the y-axis were normalized by min-max scaling); (b) Shape embedding; (c) Average pairwise cosine similarity between the top-12 brands (see Eq 4) via color embedding; (d) Average pairwise cosine similarity via shape embedding.

Model interpretation. Regarding the higher prediction power of the shape embedding compared to the color embedding in the ablation study (see Table 1), we speculate that this observation also aligns with the temporal patterns depicted by the **Sneaker Design Index**. Brands shared a similar trajectory in terms of color choice over the years yet sustained their particular shape-wise designs. This can represent a challenge in classifying product categories or reselling premiums. Nonetheless, our work shows potential for such tasks, and we plan to expand our work to other human artifacts (contents) and examine if this trait is coherent across various items. This attempt will reveal what a core (backbone) attribute would be depending on the content.

Business implications and future directions. There are many exciting future directions. One is to utilize additional information about sneakers, such as materials or user reviews. This information was not readily available for all sneaker items in our dataset, yet future research can combine such heterogeneous meta features for analysis. One can use sentence embedding from language models, such as BERT, to process text-based user reviews or product descriptions [7, 35, 44]. Another approach is to collaborate with the fashion industry of various other product types, extending the application of the proposed model. For example, in terms of shapes, we may also focus on dresses, jackets, or bags and examine their evolution over a longitudinal period. In general, this content-analysis-based approach could help analyze temporal patterns and design the evolution of any human artifacts for which there are large-scale data, contributing to cultural analytics.

Future directions for modeling. In terms of technique, our work can be improved in the following ways. We want to adopt advanced augmentation methods, such as CutMix [46], Autoaugment [12], and RandomAugment [13], used in recent contrastive learning studies to better represent latent information. In contrastive learning, CNN automatically learns the design characteristics that are helpful to discriminate between items. However, due to the black-box nature of deep learning algorithms, it is difficult to identify which particular feature contributes the most to the design index. For interpretability, we plan to apply post hoc methods such as LIME [33] and Saliency Maps [11] to visualize critical components. We could also use metadata such as retail and reselling prices. The reselling price can be considered a proxy of product popularity and adds contextual meaning to any design change. Expert evaluation can be used as feedback to improve machine learning predictions [23]. While StockX.com does not contain bidders' geographic information, other platforms (such as Flickr or Instagram, where hashtags such as #sneakerhead account for 23 million images) may offer metadata that would allow for analysis of the spatial location of trends [28]. Our model is unsupervised, which means it can be extended to additional features, such as materials and other fashion items to allow broader applications. Moreover, our model provides three projection heads that can focus on different attributes depending on a given task in a disentangled manner, providing some degree of interpretation of the model.

ACKNOWLEDGMENTS

M. Cha is the corresponding author of this work. This work was supported by the Institute for Basic Science (IBS-R029-C2) and the National Research Foundation of Korea (NRF-2017R1E1A1A01076400).

REFERENCES

[1] Ziad Al-Halah, Rainer Stiefelhagen, and Kristen Grauman. 2017. Fashion forward: Forecasting visual style in fashion. In *proc. of the ICCV*. 388–397.
[2] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2015. Neural machine translation by jointly learning to align and translate. In *proc. of the ICLR*.
[3] Etienne Becht, Leland McInnes, John Healy, Charles-Antoine Dutertre, Immanuel WH Kwok, Lai Guan Ng, Florent Ginhoux, and Evan W Newell. 2019. Dimensionality reduction for visualizing single-cell data using UMAP. *Nature Biotechnology* 37, 1 (2019), 38–44.
[4] Regina Lee Blaszczyk and Ben (eds.) Wubs. 2018. *The Fashion Forecasters. A Hidden History of Color and Trend Prediction.* Bloomsbury.
[5] Pierre Bourdieu. 1984. *Distinction.* Routledge.
[6] Rishav Chakravarti and Xiannong Meng. 2009. A study of color histogram based image retrieval. In *proc. of the ICITS*. 1323–1328.
[7] Chen-Hsi Chang, Hung-Ting Su, Jui-Heng Hsu, Yu-Siang Wang, Yu-Cheng Chang, Zhe Yu Liu, Ya-Liang Chang, Wen-Feng Cheng, Ke-Jyun Wang, and Winston H Hsu. 2021. Situation and Behavior Understanding by Trope Detection on Films. In *proc. of the Web Conference*. 3188–3198.
[8] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. 2020. A simple framework for contrastive learning of visual representations. In *proc. of the ICLR*. 1597–1607.
[9] Xinlei Chen, Haoqi Fan, Ross Girshick, and Kaiming He. 2020. Improved baselines with momentum contrastive learning. arXiv:2003.04297 [cs.CV]
[10] Brian Chu, Vashisht Madhavan, Oscar Beijbom, Judy Hoffman, and Trevor Darrell. 2016. Best practices for fine-tuning visual classifiers to new domains. In *proc. of the ECCV*. 435–442.
[11] Runmin Cong, Jianjun Lei, Huazhu Fu, Ming-Ming Cheng, Weisi Lin, and Qingming Huang. 2018. Review of visual saliency detection with comprehensive information. *IEEE Transactions on circuits and Systems for Video Technology* 29, 10 (2018), 2941–2959.
[12] Ekin D Cubuk, Barret Zoph, Dandelion Mane, Vijay Vasudevan, and Quoc V Le. 2019. Autoaugment: Learning augmentation policies from data. In *proc. of the CVPR*.
[13] Ekin D Cubuk, Barret Zoph, Jonathon Shlens, and Quoc V Le. 2020. RandAugment: Practical Automated Data Augmentation with a Reduced Search Space. *Advances in Neural Information Processing Systems* (2020).
[14] Mikayla DuBreuil and Sheng Lu. 2020. Traditional vs. big-data fashion trend forecasting: an examination using WGSN and EDITED. *International Journal of Fashion Design, Technology and Education* 13, 1 (2020), 68–77.
[15] Takao Furukawa, Chikako Miura, Mori Kaoru, Sou Uchida, and Makoto Hasegawa. 2019. Visualisation for analysing evolutionary dynamics of fashion trends. *International Journal of Fashion Design, Technology and Education* 12 (2019), 1–13.
[16] Ahyoung Han, Jihoon Kim, and Jaehong Ahn. 2021. Color Trend Analysis using Machine Learning with Fashion Collection Images. *Clothing and Textiles Research Journal* (2021).
[17] Sungwon Han, Sungwon Park, Sungkyu Park, Sundong Kim, and Meeyoung Cha. 2020. Mitigating embedding and class assignment mismatch in unsupervised image classification. In *proc. of the ECCV*. 768–784.
[18] Sungwon Han, Hyeonho Song, Seungeon Lee, Sungwon Park, and Meeyoung Cha. 2021. Elsa: Energy-based Learning for Semi-supervised Anomaly Detection. In *proc. of the BMVC*.
[19] Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. 2020. Momentum contrast for unsupervised visual representation learning. In *proc. of the CVPR*. 9729–9738.
[20] Ruining He and Julian McAuley. 2016. Ups and Downs: Modeling the Visual Evolution of Fashion Trends with One-Class Collaborative Filtering. In *proc. of the Web Conference*. 507–517.
[21] Shintami Chusnul Hidayati, Kai-lung Hua, Wen-Huang Cheng, and Shih-Wei Sun. 2014. What are the Fashion Trends in New York?. In *proc. of the ACM Multimedia*.
[22] Youngseung Jeon, Seungwan Jin, and Kyungsik Han. 2021. FANCY: Human-centered, Deep Learning-based Framework for Fashion Style Analysis. (2021), 2367–2378.
[23] Youngseung Jeon, Seungwan Jin, and Kyungsik Han. 2021. FANCY: Human-centered, Deep Learning-based Framework for Fashion Style Analysis. In *proc. of the Web Conference*. 2367–2378.
[24] Asako Kanezaki. 2018. Unsupervised image segmentation by backpropagation. In *proc. of the ICASSP*. 1543–1547.
[25] M. Hadi Kiapour, Kota Yamaguchi, Alexander C. Berg, and Tamara L. Berg. 2014. Hipster Wars: Discovering Elements of Fashion Styles. In *proc. of the ECCV*. 472–488.
[26] George Kubler. 1962. *The shape of time : remarks on the history of things.* Yale University Press.
[27] Er Manisha Lumb and Er Poonam Sethi. 2013. Texture Feature Extraction of RGB, HSV, YIQ and Dithered Images using Wavelet and DCT Decomposition Techniques. *International Journal of Computer Applications* 975 (2013), 8887.
[28] Utkarsh Mall Mall, Kevin Matzen, Bharath Hariharan, Noah Snavely, and Kavita Bala. 2019. GeoStyle: Discovering Fashion Trends and Events. arXiv:1908.11412 [cs.CV]
[29] Colin Martindale. 1990. *The clockwork muse: The predictability of artistic change.* Basic Books.
[30] Kevin Matzen, Kavita Bala, and Noah Snavely. 2017. StreetStyle: Exploring world-wide clothing styles from millions of photos. arXiv:1706.01869 [cs.CV]
[31] Leland McInnes, John Healy, and James Melville. 2018. Umap: Uniform manifold approximation and projection for dimension reduction. arXiv:1802.03426 [stat.ML]
[32] Aaron van den Oord, Yazhe Li, and Oriol Vinyals. 2018. Representation learning with contrastive predictive coding. arXiv:1807.03748 [cs.LG]
[33] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. 2016. " Why should i trust you?" Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*. 1135–1144.
[34] Amaia Salvador, Michal Drozdzal, Xavier Giro-i Nieto, and Adriana Romero. 2019. Inverse cooking: Recipe generation from food images. In *proc. of the CVPR*. 10453–10462.
[35] Mingi Shin, Sungwon Han, Sungkyu Park, and Meeyoung Cha. 2020. A risk communication event detection model via contrastive learning. In *proc. of the NLP4IF Workshop on NLP for Internet Freedom: Censorship, Disinformation, and Propaganda*. 39–43.
[36] Martin Siefkes and Emanuele Arielli. 2018. *The Aesthetics and Multimodality of Style.* Peter Lang.
[37] Georg Simmel. 1957. Fashion. *Amer. J. Sociology* 62, 6 (1957), 541–558.
[38] Junding Sun, Ximin Zhang, Jiangtao Cui, and Lihua Zhou. 2006. Image retrieval based on color distribution entropy. *Pattern Recognition Letters* 27, 10 (2006), 1122–1126.
[39] Yonglong Tian, Chen Sun, Ben Poole, Dilip Krishnan, Cordelia Schmid, and Phillip Isola. 2020. What makes for good views for contrastive learning? arXiv:2005.10243 [cs.LG]
[40] Jonathan D Touboul. 2019. The hipster effect: When anti-conformists all look the same. *Discrete & Continuous Dynamical Systems-B* 24, 8 (2019), 4379–4415.
[41] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in Neural Information Processing Systems* (2017).
[42] Sirion Vittayakorn, Kota Yamaguchi, Alexander C. Berg, and Tamara L. Berg. 2015. Runway to Realway: Visual Analysis of Fashion. In *proc. of the IEEE Winter Conference on Applications of Computer Vision*. 951–958.
[43] Tete Xiao, Xiaolong Wang, Alexei A Efros, and Trevor Darrell. 2021. What Should Not Be Contrastive in Contrastive Learning. In *proc. of the ICLR*.
[44] Hu Xu, Bing Liu, Lei Shu, and S Yu Philip. 2019. BERT Post-Training for Review Reading Comprehension and Aspect-based Sentiment Analysis. In *proc. of the NAACL-HLT*. 2324–2335.
[45] Wei Yang, Ping Luo, and Liang Lin. 2014. Clothing co-parsing by joint image segmentation and labeling. In *proc. of the CVPR*. 3182–3189.
[46] Sangdoo Yun, Dongyoon Han, Seong Joon Oh, Sanghyuk Chun, Junsuk Choe, and Youngjoon Yoo. 2019. Cutmix: Regularization strategy to train strong classifiers with localizable features. In *proc. of the CVPR*. 6023–6032.

Appendix

A FEATURE EXTRACTION METHODS

For the sneaker images, we constructed features from feature engineering to examine temporal changes in sneaker designs. We first constructed five segmentation-related features, number of segments, mean (μ) and standard deviation (std) for areas and perimeters of segments per image, i.e., a total of five dimensions (5D), based on unsupervised image segmentation [24], as presented with one example in Figure 3. Unsupervised image segmentation is a well-established method that learns segmentation information image by image via backpropagation that partitions an image into groups of pixels containing similar traits.

We then extracted three features from the color attributes: distribution parameter, histogram, and entropy. We extracted μ and std for the RGB and HSV color models for the distribution parameter (12D). R, G, and B in the RGB color model represent the red, green, and blue colors, respectively, while H, S, and V in the HSV model represent hue, saturation, and brightness (or value), respectively. RGB and HSV are the most widely used color models in image retrieval and computer vision [27]. For the histogram, we made 128 bins out of 256 values for each RGB channel, resulting in a total of 384D [6]. For color entropy, we computed the values based on the equation below for RGB, HSV, and grayscale (7D) to particularly conjecture brightness based on a given sneaker image [38].

$$Color\ Entropy(S) = -\Sigma_{i=0}^{m} p_i log_2(p_i),\ p_i = \frac{freq(C_i, S)}{|S|}, \quad (5)$$

where $S \in \{R, G, B, H, S, V, grayscale\}$, $m = 255$, C = set of values within a class, and $|S| = 65{,}536$ (i.e., 256×256). Specifically, all crawled images have white backgrounds, as shown on the left side of Figure 3(a). To minimize the unintended effect of the background when performing color feature engineering, we reconstructed target images by merging the extracted segmentation by image and used pixels only within the reconstructed areas.

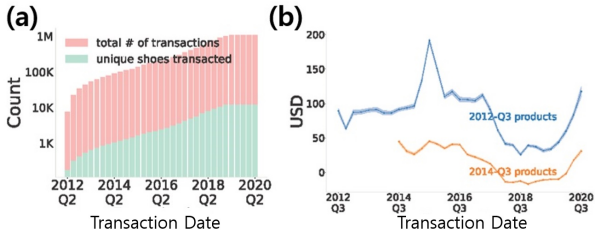


Figure 8: Additional data trends: (a) Released sneaker product and corresponding brand volume trends by quarter; (b) The median reselling premium trends for the sneakers released in 2012-Q3 and 2014-Q3, respectively, shown with one standard error: prices were adjusted based on the inflation of US currency (USD).

B DISTRIBUTIONS AND PRICE FEATURES

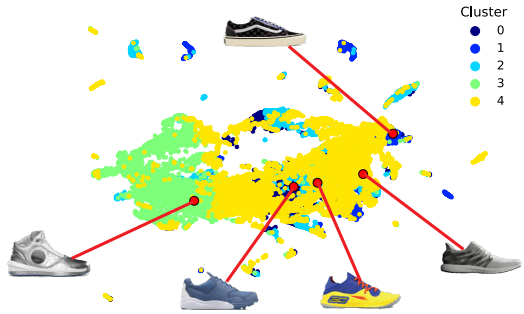
Considering the reselling market first, the platform has boomed, as shown in Figure 8(a). Based on the frequency distribution plots in Figure 7, we observe that retail prices have not changed much over 20 years. Concerning the characteristics of the reselling transactions, only a handful of brands, such as Adidas Yeezy and Air Jordan, obtain high resale premiums. Moreover, older sneaker products appear to obtain higher resale premiums; this pattern is also observed in Figure 8(b). We may interpret these findings as indicating that people are more zealous toward classical or original sneaker products. To compute the adjusted retail or resale price considering inflation of the US currency, we retrieved the annual purchasing power of the USD from the CPI library in Python, which uses data from the US Bureau of Labor Statistics, and computed the adjusted price as follows: $p'_t = \gamma_t p_t$, where p is the raw price, p' is the adjusted price, t is the target year, γ is the annual purchasing power, and the baseline year (i.e., where $\gamma_t = 1.0$) is set to the current year, 2021.

C CLASSIFICATION RESULTS

Linked to Section 5.1, we iterate running one classification task on Primary Category with three classifiers, as reported in Table 2: Multinomial Logistic Regression, XGBoost, and MLP. The results are consistent with those in Table 1: our embedding model outperformed other feature-engineered or SOTA models.

D VISUALIZATION OF SHAPE EMBEDDING

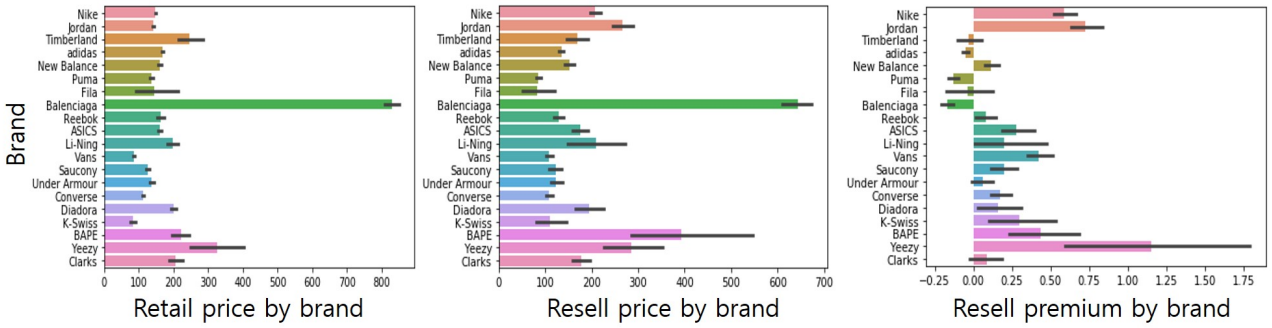
Expanding the discussion in Section 5.2, Figure 9(a) shows the derived clusters and centroids therein for the shape attribute. The optimal number of clusters is five, and we confirm that sneakers are surprisingly well matched within the group based on their shape, as presented by the centroids and their neighbors in Figure 9(b).



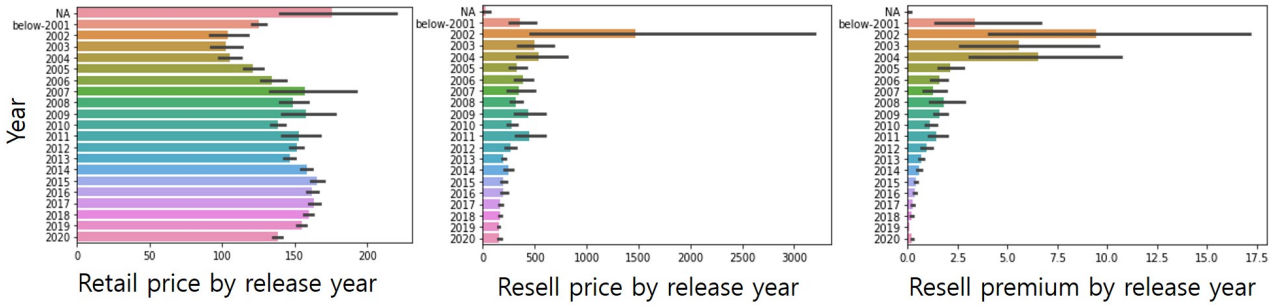
(a) K-means clustering of the embedding for the shape attribute.

(b) Sneaker examples by cluster for the shape attribute.

Figure 9: Centroids within clusters and their 15-nearest neighbors based on shape embedding.



(a) Frequency distributions of retail price and resale transaction price and corresponding premium across brands.



(b) Frequency distributions of retail price and resale transaction price and corresponding premium across release years.

Figure 7: Frequency distributions of price-related features by brand and release time, with bars showing one standard error.

N = 15,007 Sneaker Products		Multinomial Logistic Regression					XGBoost					MLP (Neural-net)				
Attribute	Feature	Acc.	Pre.	Rec.	F1	κ	Acc.	Pre.	Rec.	F1	κ	Acc.	Pre.	Rec.	F1	κ
Random	1 / 8 Classes	.167	—	—	—	—	.167	—	—	—	—	.167	—	—	—	—
Feature Engineering:																
Color	Dist. parameter (12D)	.209	.288	.209	.223	.084	.362	.250	.362	.289	.145	.352	.261	.352	.291	.140
Color	Entropy (7D)	.301	.392	.301	.323	.184	.462	.327	.462	.383	.289	.460	.337	.460	.387	.292
Color	Histogram (128bin, 384D)	.239	.355	.239	.274	.123	.430	.333	.430	.356	.242	.401	.388	.401	.391	.250
Segmentation	Unsp. Image Seg. (5D) [24]	.194	.275	.194	.188	.085	.359	.245	.359	.276	.137	.384	.203	.384	.265	.167
Concatenation	Entropy + Segmentation (12D)	.301	.411	.301	.327	.191	.469	.331	.469	.388	.297	.476	.389	.476	.401	.308
Contrastive Learning:																
Color+Shape	LooC (384D) [43]	.855	.860	.855	.857	.821	.856	.853	.856	.850	.820	.882	.887	.882	.884	.854
Color+Shape	**Ours: all-inv. Rep (512D)**	**.898**	**.899**	**.898**	**.898**	**.874**	**.905**	**.905**	**.905**	**.903**	**.882**	**.926**	**.927**	**.926**	**.926**	**.909**
Ablation Study:																
Color	Ours: color-inv. Rep (512D)	.901	.902	.901	.901	.877	.905	.905	.905	.903	.882	.839	.838	.839	.836	.801
Shape	Ours: shape-inv. Rep (512D)	.757	.767	.757	.761	.701	.819	.810	.819	.807	.772	.799	.804	.799	.801	.751
Color+Shape	Ours: three Reps total (1536D)	.893	.895	.893	.894	.868	.907	.907	.907	.905	.884	.928	.928	.928	.928	.751

Table 2: Results of predicting the PRIMARY CATEGORY utilizing 3 off-the-shelf classifiers (κ: Cohen's kappa; the bold row demonstrates the most significant of seven embeddings compared; the bottom three rows demonstrate the ablation study within our model).

E CLARIFICATION OF HOW TO USE OBJECT IMAGES TO BUILD DESIGN EMBEDDINGS

We used the sneaker images as an example cultural artifact. These images were well prepared on StockX.com with a white background. We chose images with the same camera angle to control exogenous factors, i.e., the current deep embedding model used one image per sneaker. Regarding other domains, the same treatment (i.e., controlling the background and camera angle) would make it easier to build the design embeddings.