Towards Style Classification for Fashion Recommender Systems

René Kessler
Department of Computing Science
University of Oldenburg
Ammerländer Heerstr. 114-118, 26129 Oldenburg
rene.kessler@uol.de

Jorge Marx Gómez
Department of Computing Science
University of Oldenburg
Ammerländer Heerstr. 114-118, 26129 Oldenburg
jorge.marx.gomez@uol.de

***Abstract*—In recommender systems, products can be recommended based on customer modeling. Especially in fashion e-commerce, recommendations can be a solution for many business and sustainability challenges. This paper presents how deep learning can be used for style classification in fashion e-commerce. The style of a person is classified based on a photo. Furthermore, this paper analyzes which data preprocessing and augmentation techniques can positively impact a style classification model. In order to be able to explain recommendations based on this, it was analyzed to what extent the resulting model provides accurate results and the predictions can be explained.**

***Keywords*—E-Commerce, Recommender Systems, Style Classification, Deep Learning, Explainable AI.**

I. Introduction

While direct consultation can occur in stationary retail, for example, by sales staff, this point is almost entirely omitted in online retail. Whenever a customer is not looking for a specific product they already know and has not yet made a clear product decision, this advice may still be necessary. For this reason, e-commerce retailers work with recommender systems, i.e., they try to present a specific product based on existing customer information in such a way that the user does not have to search through the entire product catalog but can reach the desired product quickly and effortlessly [1]. At the same time, targeted recommendations can also create incentives to buy. The recommendation quality always depends on the quality and quantity of the existing customer information. The more data available and the higher the information value of this data, the more accurate the customer profiling can be [2] [3]. The importance of recommendations becomes particularly clear regarding products that cannot be described exclusively textually by their properties. There is a high demand for individual advice, particularly in the case of stylistic features or visual characteristics, which can only be represented to a limited extent by textual descriptions or categorizations. For example, such products can be found in the fashion sector, which represents the highest share of all e-commerce sectors in terms of sales with 20.0 billion euros (Germany) in 2021 [4]. Compared to 2020, the online share in the fashion industry has risen by 3.21 %, from 39.8 % to 46.5 % - in absolute terms, this represents an increase of 3.21 billion euros, underscoring the importance and future potential of recommendations in online retail [3] [4].

The basis for recommender systems is the profiling of customers. The overall goal of customer profiling is to collect relevant data from the customer, extract information from this data, build a customer model, and use this model to predict the customer's future behavior to support purchase decisions [5]–[7]. Stereotyping or clustering of customers represents an important strategy here in order to be able to realize personalization for a group of customers [8]–[10].

Customer-relevant data sources (e.g., master data, transaction data, or data from marketing campaigns) can be used to gain information about the customer [11] [12]. In research and practice, mainly master data, personal data such as age, gender, or data related to origin or place of residence are used [13]. While these data provide essential information, transactional data, such as purchase history or other touchpoints, can be used to model and subsequently predict buying behavior in an online store or even user preference [13] [14]. In contrast, a more recent approach is the analysis of click paths in online platforms. This can be used to model the interaction behavior of customers without purchases having already taken place [15] [16]. Thus, the priority is to evaluate structured data in current approaches to model customer characteristics, behavior, and preference.

The paper is structured as follows: The introduction is followed by the introduction to the problem addressed. The third section describes related work, while the fourth section describes the research objectives and methodology. In the main body, the fifth section, the data basis is first described. Based on this, the experiments, their evaluation and the interpretability of the results are explained. Finally, the results are discussed and a short outlook for further research is given.

II. Problem Statement

To recommend a product to customers at all, the relationships between customers and products must be known. A distinction can be made between micro and macro behavior factors. The analysis of transaction data, e.g., purchases of products or product reviews, describes the macro behavior, i.e., the user's behavior that led to a purchase. The micro behavior is fundamentally more refined in its structure. Here, the user's interactions with products that are not directly linked to a purchase are also analyzed, i.e., in addition to

 ISBN: 978-1-61208-994-2

transaction data, behavioral patterns are also analyzed. The insights gained are incorporated into the recommendation process [1] [14]–[16]. In addition to transaction and interaction data, many approaches also process data about creating the product and demographic data (or master data, such as age or gender) [13]. The data sources used in recommender systems can be roughly divided into three types: user-related data, product-related data, and transaction- or behavior-related data. These data are primarily available in a structured form. Data of an unstructured nature has hardly played a role in existing approaches. When unstructured data is used, it is mainly product-related information, i.e., visual features of the products or textual descriptions, and product- and user-related information, i.e., textual product reviews [17]–[19]. The approaches considered mainly look at simple, individual products rather than product bundles (such as outfits) [20]–[22]. Therefore, these approaches can only be classified in the analysis of relationships between users and products.

Unstructured personal data (e.g., customer images) have been considered relatively rarely, although they can be crucial for modeling user preferences [18]. The subject of current research is processing customer images to extract new information that can be used to improve recommendations, as publications show [3] [23]. Although visual recommender systems exist, the focus is mainly on product images. Analyzing and using this data to generate recommendations has added value but cannot solve the cold start problem [24]. Various studies have shown that explainability can positively contribute to the quality of recommendations. Existing approaches indicate that explaining recommendations (e.g., via historical transactions, similarity metrics between products, or customer ratings and reviews) can increase customer satisfaction [24]–[27]. However, it is also a prerequisite in these approaches that historical transaction data of the customer is already available. In particular, accurately recognizing a clothing style can provide an essential building block for successful recommendations in fashion e-commerce. Nevertheless, clothing style can rarely be determined simply by utilizing textual information or purchase history; the customer's visual characteristics must be processed for this purpose.

III. Related Work

In order to make the best possible recommendations to a customer, the customer must be known as well as possible. In fashion e-commerce, it is essential to know the style in which the customer dresses to assist them in finding a product or to recommend products that they will probably like because they match their style. Especially in aesthetic use cases such as the determination of a clothing style, which cannot be described one hundred percent textually, deep learning or, in particular, computer vision offers a possibility to process these aesthetic features via images. There are already previous approaches in style classification, but they differ from the focus of this paper.

In a literature study, the literature databases *Scopus*, *ScienceDirect*, *Web of Science*, *arXiv*, *ACM*, *IEEE* were searched for relevant articles from 2013. The retrieved hits were then filtered by title and abstract screening.

Gu et al. (2017) combine traditional recommender system approaches with autoencoder-based processing of visual fashion features, so-called fashion coordinates. Sets of product images (three products each: outerwear, pants, and shoes) serve as input [28]. A similar approach is taken by McAuley et al. (2015). Here, the authors also consider product images and developed a prototype matching products to a product (input) [29]. Liu et al. (2017) in their approach try to classify fashion images into clusters representing a style via different clustering methods. In addition to individual product images, image details of people are also included [17].

An approach based on product data but which differs significantly is described by Guan and Qin (2019). In their work, product images are analyzed, and descriptive attributes are extracted from the images. These descriptive attributes (e.g., colors or contrast) are then associated with human concepts, feelings, or other customer characteristics to make product recommendations [30]. Properties and attributes of a style can also be modeled as a knowledge graph to represent styles from multiple properties and their interrelationships [31].

Ma et al. (2017) show an interesting approach in their paper that uses images to show a spectrum of different clothing styles to make them more understandable. In this approach, style features are first extracted from images, then processed by an autoencoder to cluster the autoencoder results. The resulting clusters were then classified in a Fashion Semantic Space [32]. Schindler et al. (2018) are similarly concerned with extracting features from images. Here, they use images of people and products crawled from online stores. These images are then classified using a pre-trained Convolutional Neural Network [33].

Existing work makes a valuable contribution to research in visual recommender systems but looks at the underlying problem of fashion style classification from a different angle. In the context of this work, an attempt will be made to assign a style to product images directly and dispense with the prior extraction of features. The resulting fashion style classification can then be used for further steps in the recommendation process.

IV. Research Objectives and Methodology

This paper presents a novel fashion style classification approach that analyzes users' image data (assuming consent of the user) in fashion e-commerce and extracts the fashion style. The detected fashion style can then be incorporated into the recommendation process, for example, garments that match the person's clothing style could be recommended. The focus is on the one hand on solving the cold start problem and expanding the database of recommendation systems, and on the other hand on improving the recommendations by making the generated recommendations explainable.

For this purpose, images first had to be collected and labeled. Based on the use case, three different datasets are created here. Based on this, various deep learning experiments

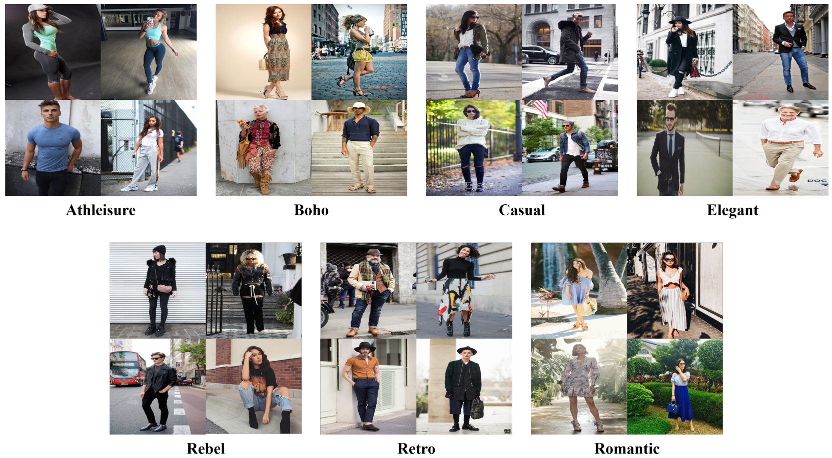

Fig. 1. Visualization of the different style classes in the dataset

were conducted. Deep Learning models were implemented and trained using different image preprocessing and augmentation techniques and critically compared. In addition to a qualitative evaluation of the models, a quantitative, technical evaluation of the deep learning experiments was also performed. To complement this, the best model was then selected, used, and the extent to which the model's results could be explained was examined.

V. Style Classification

This section describes the design and implementation of style classification. First, the creation of the dataset is discussed. This is followed by the description of the Deep Learning experiments and the evaluation of the resulting models. Finally, the procedure for evaluating the explainability of the models is described, and the results are presented.

A. Dataset

In the first step of creating the datasets for style classification, possible, suitable public data sources were researched, i.e., datasets containing images of people and, in the best case, labels about the person's clothing or style. This revealed that a large number of datasets are freely available (e.g., Figaro-1k, Appareal, CelebA, iMaterialist Fashion 2018, and 2019). However, a closer look at the datasets showed that while they can be used for some related tasks, they cannot be used as a basis for the intended classification of customer style, as none of the datasets focus on people in real scenes. There are no labels for the person's style. Therefore, the creation of a separate dataset was necessary.

Before data collection, a catalog of 7 style classes was created by researching fashion magazines/portals and interviewing three fashion experts. The three fashion experts are employees of a cooperating online fashion retailer. The seven different styles and their characteristics are described below:

- **Athleisure:** The style *Athleisure* describes mainly sporty styles, i.e. people who wear sportswear in everyday life. Examples may include plain, tighter-fitting, muscle-toned shirts or leggings.
- **Boho:** In the *Boho* style, wider-fitting, less muscle-emphasizing clothes are mainly worn. In addition, earth tones or eye-catching patterns, such as mandala patterns, are often features of the garments worn.
- **Casual:** The *Casual* style represents the typical everyday style, that is, the style that most people wear in everyday life. Examples include simple jeans, white shirts, or sweaters in shades of gray.
- **Elegant:** The *Elegant* style is mainly worn business clothes. Often suits, blouses or elegant hats can be found. Patterns are rather rare here (if, then, more inconspicuous check patterns) and colorwise, rather calm, neutral colors are to be found, such as black, gray or beige.
- **Rebel:** The *Rebel* look makes heavy use of punk and rock elements. Ripped jeans, studs, leather jackets or band shirts are often found.
- **Retro:** The *Retro* style describes clothing styles that seem to have fallen out of time, i.e., mainly classic clothing items, such as corduroy pants, hats or long coats.
- **Romantic:** The *Romantic* style describes playful looks and is mainly found on women. Core elements are, for example, dresses in pastel colors or floral patterns.

The defined styles were used in the next step to crawling public search engines and portals. In this course, the search engine *Bing* and the fashion social media portal *Chictopia* were used, and a total of 11200 images of people, as close to reality as possible, were collected. The images were manually sifted to clean them up (erroneous crawls or duplicates) and then annotated by a team of eleven, with the majority defined as label strategy. Each image was assigned exactly one style. In some cases, more than one person could be recognized in the image. If all persons to be seen could be assigned to one style, the image was kept in the dataset; otherwise, it was removed.

The resulting images were then used to form a data set (in

TABLE I
ACCURACY OF MODELS USING DIFFERENT DATASETS AND AUGMENTATION TECHNIQUES

	Accuracy		
Augmentation	**Original**	**Without Background**	**Cropped**
No Augmentation	0.698	0.698	0.755
Horizontal Flip	0.679	0.679	0.751
Blur	0.698	0.680	0.738
CLAHE	0.689	0.699	0.729
Coarse Dropout	0.698	0.699	0.738
Elastic Transform	0.699	0.704	0.739
Grid Distortion	0.707	0.694	0.729
Motion Blur	0.704	0.683	0.737
Optical Distortion	0.704	0.697	0.734
Random Resized Crop	0.692	0.683	0.739
Shift Scale Rotate	0.709	0.701	0.727
All Augmentations	**0.726**	0.706	0.749
AutoAugmentation	0.717	**0.733**	**0.825**

the further course: original) for the experiments. A 70/30 split was used, i.e., 70 % of the images were used for training, while 30 % of the images formed the validation dataset. This resulted in 1120 images per class for training and another 480 images per class to validate the models. Some examples of the resulting dataset and the annotated styles can be seen in Figure 1.

In a further step of data preprocessing, additional datasets were built. First, images with multiple people were split, i.e., object detection was used to detect people on images and extract the resulting bounding box. This resulted in smaller images but more images forming the dataset. The second dataset (Cropped) consisted of 13500 images (70/30 - Training/Validation-Split). On the other hand, we went one step further and completely removed the background of the person. For this purpose, we implemented a background removal service that detects the background of an image and removes all pixels except those belonging to the person to be seen. The resulting dataset (without background) also included a total of 13500 images. An example of the data preprocessing steps is shown in Figure 2.

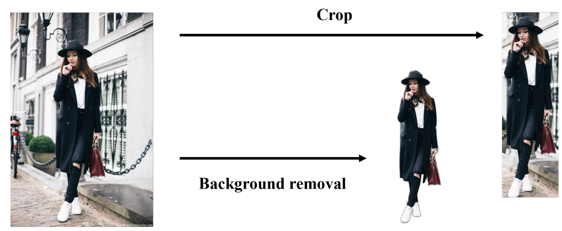

Fig. 2. Removing the background and cutting out persons to create the different datasets

B. Modeling Experiments

Numerous experiments were conducted to develop the style classification model. The EfficientNet architecture developed by *Google* was used as the model architecture. This architecture was chosen because it can be used and adapted flexibly and because it offers state-of-the-art results in the field of image classification [34]. The b0 variant of EfficientNet was tested for resource reasons. However, it can be assumed that the different architectures can show even better results due to the higher number of parameters and the possibility of processing larger images. For implementation, the *PyTorch*-implementation in the framework *timm* of EfficientNet was used. The following hyperparameters were used: Epochs 10, Batchsize 64, Learning rate 1e-2, Optimizer Adam, Finetuning (last 100 layers), Activation (Output-Layer) Softmax, Loss-Function Categorical Cross-Entropy.

Based on the architecture, further experiments were conducted using various image augmentation and preprocessing techniques. In total, ten established, different techniques for image enhancement (Horizontal Flip, Blur, CLAHE, Coarse Dropout, Elastic Transform, Grid Distortion, Motion Blur, Optical Distortion, Random Resized Crop and Shift Scale Rotate) were evaluated. For this purpose, the image augmentation framework *albumentations* was used, and a custom data loader was implemented to augment the input images with a defined probability of p=0.5 with the respective technique [35]. Models without augmentation were also implemented as baselines to ensure comparability. In addition, the ten different techniques were applied in combination (All Augmentations). As an alternative augmentation strategy, the AutoAugmentation approach from Google Brain was implemented and examined in more detail [36]. AutoAugmentation describes a procedure to learn an optimal data augmentation strategy based on the existing data material. The algorithm searches for the optimal strategy and adapts it individually in sub-strategies.

C. Evaluation of the Models

Looking at the results in Table I, it quickly becomes apparent that some augmentation techniques can positively affect the quality of the results (measured by the accuracy). The effects of the respective techniques also differ depending on the dataset used. For the *original* dataset, i.e., the dataset without further image preprocessing, all augmentation techniques perform relatively equally well. The manual combination of all augmentation techniques is the best, with an accuracy of 72.6 %. A similar picture emerges at first glance when looking at the individual augmentation techniques for the

 ISBN: 978-1-61208-994-2

dataset without *Without background* and *Cropped*. Thus, all individual techniques and their combinations within the dataset perform relatively similarly. However, it can be shown that the *Cropped* dataset performs much better than the other two datasets (about 5 % across all augmentation techniques). It is particularly striking, however, that for both datasets, the AutoAugmentation technique performs best. Overall, the best model in the experiments was the model trained on the *Cropped* dataset using AutoAugmentation. In addition, it is noticeable that only the AutoAugmentation approach shows an improvement in this dataset. This can be explained by the fact that the other augmentations may change the images too much. It is conceivable that the algorithmic search for the best strategy selects more realistic augmentation intervals.

Overall, the experiments show promising results with the best accuracy of 82.5 %. Manual evaluation of additional images that were not included in the training set shows that the models can solve the style classification problem (see Figure 3). In particular, the manual evaluation shows that outfits that can be localized (e.g., a complete sports outfit) are reliably classified. In contrast, mixed styles can be problematic for the model, e.g., when people deliberately combine elements of different styles. It can be stated from the experiments that a style classification benefits from a large amount of data, and the more data available, the better results can be obtained. Additionally, it shows that AutoAugmentation can be a promising strategy for data augmentation. The approach minimizes the manual effort, and at the same time, very good results are shown.

D. Explainability of the Models and Predictions

Now that an initial model for style classification has been developed, it will be examined to what extent the model's predictions can also be explained. In doing so, it will be investigated which areas of an image lead to the classification into a specific style. Furthermore, it is to be investigated whether the resulting model recognizes the correct image contents or, for example, learns styles from an image's context (e.g., the image background).

For this purpose, the widely used framework *lime* is used [37]. *lime* provides methods to provide so-called locally interpretable model-agnostic explanations. To this end, another experiment used the model previously identified as the best model and built an *lime* explanatory model based on it.

The results of this explanatory model can be seen in Figure 4. Clearly, the model learns the expected signal of an image, i.e., image areas where the person whose style can be classified can be seen. At the same time, it is noticeable that background areas are also perceived as a signal or partial areas of the person (see upper example on figure 4) are learned as a negative influence on the classification. This becomes especially clear if one looks closely at the Explanation Map.

In principle, explanatory models can be expected to produce better results if the model also produces more robust and better results since explainability depends on model performance. The results of the explanatory models can be used in the further course of development to enrich the generated recommendations with a visual explanation and thus create additional user acceptance.

VI. Conclusion and further Steps

In this state of research, it has already been shown that unstructured data can provide added information value for recommender systems and underlying customer profiling. A first approach to style classification is shown here. The results of the models so far are promising and already show practicality. Moreover, it becomes apparent that the manual search for an optimal augmentation strategy is not trivial. Especially in the case of complex elements in an image, as is the case with the detection and classification of fashion styles, changes to the image can lead to positive effects, but also to negative effects if the changes are incorrectly selected, as the results were able to show. The AutoAugmentation approach shows promise, where a very realistic augmentation strategy adapted to the data set can be found, which could lead to better results in our experiments.

In defining styles, it became clear that styles cannot always be clearly distinguished from one another. Often, fashion-conscious people deliberately combine different styles to make a fashion statement. On the other hand, certain clothing items are often worn in different styles. It can be assumed that styles are almost always associated with fashion elements but that the boundaries between styles can sometimes become blurred. This led to the fact that the developed models for style classification may have difficulties, although providing good results, especially in these mixed clothing styles. On the one hand, a more extensive and improved data set will be built for this purpose, as described earlier. On the other hand, further experiments will be conducted to investigate whether multi-label classification can solve this problem.

The technical evaluation of the different models by the developers and the manual visual inspection has shown that the results of the models can be considered promising and will therefore be followed up and extended. In the future, this very technical and subjective evaluation will be complemented by an empirical study to test the methods' suitability objectively. To this end, various test scenarios will be developed and conducted. This study will also measure whether the explainability of the approach can positively contribute to the perceived goodness of the models and whether they are advantageous compared to classical, non-explained recommendations.

References

[1] P. Kumar and R. Thakur, "Recommendation system techniques and related issues: a survey," *Int. j. inf. tecnol. (December 2018)*, 2018.

[2] Y.-J. Park and K.-N. Chang, "Individual and group behavior-based customer profile model for personalized product recommendation," *Expert Systems with Applications*, vol. 36, no. 2, Part 1, pp. 1932–1939, 2009.

[3] H. Zheng, K. Wu, J.-H. Park, W. Zhu, and J. Luo, "Personalized fashion recommendation from personal social media data: An item-to-set metric learning approach," *2021 IEEE International Conference on Big Data (Big Data)*, pp. 5014–5023, 2021.

[4] HDE Handelsverband Deutschland, "Online monitor 2022." Web, 2022. https://einzelhandel.de/index.php?option=com_attachments&task=download&id=10659, retrieved on 28.08.2022.

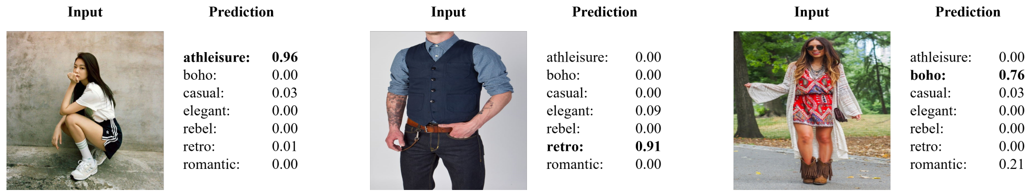

Fig. 3. Exemplary predictions of the style classification model

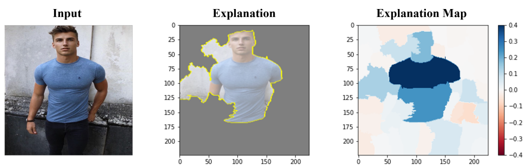

Fig. 4. Visualization to explain the predictions of the model

[5] S. Berkovsky, "Ubiquitous user modeling in recommender systems," in *User Modeling 2005* (L. Ardissono, P. Brna, and A. Mitrovic, eds.), (Berlin, Heidelberg), pp. 496–498, Springer Berlin Heidelberg, 2005.
[6] K. Lakiotaki, N. F. Matsatsinis, and A. Tsoukias, "Multicriteria user modeling in recommender systems," *IEEE Intelligent Systems*, vol. 26, no. 2, pp. 64–76, 2011.
[7] H.-N. Kim, A. Alkhaldi, A. E. Saddik, and G.-S. Jo, "Collaborative user modeling with user-generated tags for social recommender systems," *Expert Systems with Applications*, vol. 38, no. 7, pp. 8488–8496, 2011.
[8] E. Rich, "User modeling via stereotypes," *Cognitive Science*, vol. 3, no. 4, pp. 329–354, 1979.
[9] H. Rijn, A. Johnson, and N. Taatgen, *Cognitive user modeling.*, pp. 523–538. Handbook of human factors in web design, CRC Press, 2 ed., 2011.
[10] A. Johnson and N. Taatgen, *User modeling.*, pp. 424–438. Handbook of human factors in web design, Lawrence Erlbaum Associates Publishers, 2005.
[11] A. Kobsa *User Modeling and User-Adapted Interaction*, vol. 11, no. 1/2, pp. 49–63, 2001.
[12] A. Goy, L. Ardissono, and G. Petrone, *Personalization in E-Commerce Applications*, vol. 4321, pp. 485–520. Berlin, Heidelberg: Springer Berlin Heidelberg, the adaptive web. lecture notes in computer science ed., 2007.
[13] K. Wei, J. Huang, and S. Fu, "A survey of e-commerce recommender systems," in *2007 International Conference on Service Systems and Service Management*, pp. 1–5, IEEE, 2007.
[14] S. Sivapalan, A. Sadeghian, H. Rahnama, and A. M. Madni, "Recommender systems in e-commerce," in *2014 World Automation Congress (WAC)*, pp. 179–184, IEEE, 2014.
[15] Y. Gu, Z. Ding, S. Wang, and D. Yin, "Hierarchical user profiling for e-commerce recommender systems," in *Proceedings of the 13th International Conference on Web Search and Data Mining*, pp. 223–231, ACM, 2020.
[16] M. Zhou, Z. Ding, J. Tang, and D. Yin, "Micro behaviors," in *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*, pp. 727–735, ACM, 2018.
[17] Q. Liu, S. Wu, and L. Wang, "DeepStyle," in *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 353–361, ACM, 2017.
[18] P. Pérez-Núñez, "Taking advantage of images and texts in recommender systems: semantics and explainability," in *Fourteenth ACM Conference on Recommender Systems*, pp. 792–796, ACM, 2020.
[19] S. Wang and J. Qiu, "A deep neural network model for fashion collocation recommendation using side information in e-commerce," *Applied Soft Computing*, vol. 110, p. 107753, 2021.
[20] B. O. Viso, "Evolutionary approach in recommendation systems for complex structured objects," in *Fourteenth ACM Conference on Recommender Systems*, pp. 776–781, ACM, 2020.
[21] M. F. Dacrema, P. Cremonesi, and D. Jannach, "Are we really making much progress? a worrying analysis of recent neural recommendation approaches," in *Proceedings of the 13th ACM Conference on Recommender Systems*, pp. 101–109, ACM, 2019.
[22] K. Laenen and M.-F. Moens, "Attention-based fusion for outfit recommendation," in *Fashion Recommender Systems* (N. Dokoohaki, ed.), (Cham), pp. 69–86, Springer International Publishing, 2020.
[23] W.-C. Kang, E. Kim, J. Leskovec, C. Rosenberg, and J. McAuley, "Complete the look: Scene-based complementary product recommendation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 10524–10533, 2019.
[24] X. Chen, Y. Zhang, H. Xu, Y. Cao, Z. Qin, and H. Zha, "Visually explainable recommendation,"
[25] X. Chen, Y. Zhang, and Z. Qin, "Dynamic explainable recommendation based on neural attentive models," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, pp. 53–60, 2019.
[26] V. Dominguez, P. Messina, I. Donoso-Guzmán, and D. Parra, "The effect of explanations and algorithmic accuracy on visual recommender systems of artistic images," in *Proceedings of the 24th International Conference on Intelligent User Interfaces*, pp. 408–416, ACM, 2019.
[27] D. Pan, X. Li, X. Li, and D. Zhu, "Explainable recommendation via interpretable feature mapping and evaluation of explainability," in *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence*, IJCAI'20, p. 2690–2696, 2021.
[28] S. Gu, X. Liu, L. Cai, and J. Shen, "Fashion coordinates recommendation based on user behavior and visual clothing style," in *Proceedings of the 3rd International Conference on Communication and Information Processing*, ICCIP '17, (New York, NY, USA), p. 185–189, Association for Computing Machinery, 2017.
[29] J. McAuley, C. Targett, Q. Shi, and A. van den Hengel, "Image-based recommendations on styles and substitutes," in *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '15, (New York, NY, USA), p. 43–52, Association for Computing Machinery, 2015.
[30] C. Guan, S. Qin, and Y. Long, "Apparel-based deep learning system design for apparel style recommendation," *International Journal of Clothing Science and Technology*, vol. 31, no. 3, pp. 376–389, 2019.
[31] C. Zhang, X. Yue, W. Liu, and C. Gao, "Fashion style recognition with graph-based deep convolutional neural networks," in *Artificial Intelligence on Fashion and Textiles* (W. K. Wong, ed.), (Cham), pp. 269–275, Springer International Publishing, 2019.
[32] Y. Ma, J. Jia, S. Zhou, J. Fu, Y. Liu, and Z. Tong, "Towards better understanding the clothing fashion styles: A multimodal deep learning approach," in *AAAI*, Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI-17), pp. 38–44, 2017.
[33] A. Schindler, T. Lidy, S. Karner, and M. Hecker, "Fashion and apparel classification using convolutional neural networks," *CoRR - Forum Media Technology*, vol. abs/1811.04374, 2018.
[34] M. Tan and Q. V. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," 2020.
[35] A. Buslaev, V. I. Iglovikov, E. Khvedchenya, A. Parinov, M. Druzhinin, and A. A. Kalinin, "Albumentations: Fast and flexible image augmentations," *Information*, vol. 11, no. 2, 2020.
[36] E. D. Cubuk, B. Zoph, D. Mane, V. Vasudevan, and Q. V. Le, "Autoaugment: Learning augmentation policies from data," 2018.
[37] M. T. Ribeiro, S. Singh, and C. Guestrin, ""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, San Francisco, CA, USA, August 13-17, 2016*, pp. 1135–1144, 2016.