

Article

Development of Fashion Product Retrieval and Recommendations Model Based on Deep Learning

Jaechoon Jo ¹, Seolhwa Lee ², Chanhee Lee ², Dongyub Lee ³ and Heuseok Lim ^{2,*}

¹ Department of Smart Information and Communication Engineering, Sangmyung University, Seoul 31066, Korea; jae@smu.ac.kr

² Department of Computer Science and Engineering, Korea University; Seoul 02841, Korea; whiteldark@korea.ac.kr (S.L.); chanhee0222@korea.ac.kr (C.L.)

³ R&D Center, Kakao, Gyeonggi-do 13494, Korea; jude.lee@kakaocorp.com

* Correspondence: limhseok@korea.ac.kr; Tel.: (+82)02-3290-2396

Received: 8 February 2020; Accepted: 17 March 2020; Published: 19 March 2020



Abstract: The digitization of the fashion industry diversified consumer segments, and consumers now have broader choices with shorter production cycles; digital technology in the fashion industry is attracting the attention of consumers. Therefore, a system that efficiently supports the searching and recommendation of a product is becoming increasingly important. However, the text-based search method has limitations because of the nature of the fashion industry, in which design is a very important factor. Therefore, we developed an intelligent fashion technique based on deep learning for efficient fashion product searches and recommendations consisting of a Sketch-Product fashion retrieval model and vector-based user preference fashion recommendation model. It was found that the “Precision at 5” of the image-based similar product retrieval model was 0.774 and that of the sketch-based similar product retrieval model was 0.445. The vector-based preference fashion recommendation model also showed positive performance. This system is expected to enhance consumers’ satisfaction by supporting users in more effectively searching for fashion products or by recommending fashion products before they begin a search.

Keywords: deep learning; convolutional neural network (CNN); generative adversarial network (GAN); Image2Vec; fashion recommendation

1. Introduction

The size of the domestic and overseas fashion market is steadily increasing. It is expected that the size of the domestic fashion market will grow by 8.8% in 2024 compared to its size in 2020, resulting in a market size of USD (United States dollars) \$26.288 million [1,2]. In particular, the apparel and fashion online market is estimated to be USD \$113 million in 2018, and it is annually growing by 18%, of which mobile shopping accounts for approximately 40% [3].

The fashion industry encompasses not only apparel but also various products such as bags, shoes, and accessories. Through the automation, advancement, and digitization of production and distribution, the fashion industry has a new paradigm. With the digitization of the fashion industry, the system analyzes the fashion wants of consumers and attempts to quickly meet consumer needs. Therefore, the digital technology of the fashion industry is attracting the attention of consumers of many segments and who have a wider range of selections because of the shortened production cycle. In addition, as the number of consumers using smart devices exceeds 10 million, it plays an important role as a new e-commerce platform [4]. Reviewing the growth rate by distribution channel, the sales of offline store retailers, which were formerly traditional apparel sales channels, decreased, while sales of non-store retailers such as online or mobile retailers have the highest growth rate. In particular, the

online fashion market was expected to continue to grow at a rate of 16% per annum from 2011 to 2016 and to 20% from January to September 2017, compared to the same period of the previous year. It is expected to become a major channel in the fashion industry in the future [5].

When purchasing fashion apparel on an e-commerce platform, it is very different from purchasing in an offline retail store, as a wider range of fashion products are available. Therefore, a system that more efficiently supports users in searching for the product they want and recommending a desired product is increasingly important. In other words, in the field of fashion where visual factors and design characteristics are important, there are many limitations to expressing only using text. However, the existing fashion retrieval system uses a text-based retrieval method based on the attribute information of the product (product name, category, brand name, etc.). Because of the nature of the fashion industry, which has a strong design factor, a text-based search method has limitations in providing satisfactory search results. Recently, several systems such as “shopping how” and “smart lens” that search for products using fashion photographs were implemented, but they did not achieve satisfactory or meaningful performance. Additionally, when using these systems, consumers must search for the image of the product online or they must photograph products themselves.

Because of the rapid growth of the fashion market, consumers can obtain various desired fashion apparel. This provides an advantage to consumers by giving them more options, but it also results in a disadvantage in that it is difficult to find the desired style online. When looking for desired apparel, general consumers do not know which keywords they should use to find such a style, although they perceive the style they would like. Therefore, in online shopping malls, there is a need to understand the preferred styles to provide a customized recommendation service to overcome these limitations and to provide satisfactory services.

Accordingly, this study developed intelligent fashion techniques for Sketch-Product and personalized coordination. This system is composed of a Sketch-Product fashion retrieval model that can resolve the limitations of a picture image search, overcoming the restrictions of a traditional text-based search method, as well as a vector-based user preferred fashion recommendation model, applying an implicit user fashion profiling method that is different from the conventional profiling method to overcome the limitations of the preferred fashion profiling. The Sketch-Product fashion retrieval model works as follows: a user sketches a fashion product that he/she wants, which is up-sampled to the level of a product image using a generative adversarial network (GAN) technology; the GAN then extracts the properties of the up-sampled product image as vector values and searches the vector space for similar images. The vector-based user preferred fashion recommendation model works as follows: a profile acquired through professional filtering based on a deep neural network (DNN) is pre-trained and set as the basic weight value of the recommendation model, and customized coordination fashions are recommended as the weight values are learned for the individual as time passes based on preferred fashion profiling.

In Section 2, research related to the development of the deep-learning-based intelligent fashion technique for Sketch-Product and personalized coordination to search and recommend fashion products is outlined, and previously developed technologies are explained. In Section 3, the Sketch-Product fashion retrieval model and the vector-based user preferred fashion recommendation model are explained. In Section 4, experimental methods and results are described. In Section 5, the conclusion of this study is presented.

2. Related Works

2.1. User Fashion Profiling and Style Recommendation

User fashion profiling in a shopping mall can be divided into two types: static profiling through surveys and dynamic profiling through the analysis of the behavioral history of the user.

Static profiling follows the survey method, and Stitch Fix falls in this category. Stitch Fix collects information such as size, body type, preferred price, and preferred style (Bohemian or Edge) through a

five-step online survey for the profiling of a user's preferred fashion [6]. It proposes the final preferred fashion products using the collected information through machine learning such as logistic regression and support vector machine, as well as the advice of fashion professionals [7].

Dynamic profiling utilizes log information based on the users' clicks on the site. This profiling is based on the behavioral history such as previous purchase history [8]. This can be automatically associated with preferred fashion recommendations. Amazon uses this type, which is also the case with Amazon's recommendation system. Amazon has a separate category for apparel and offers recommendations with information such as previous purchases. Amazon provides a real-time recommendation service using item-to-item collaborative filtering. The item-to-item collaborative filtering method lists items that are rated as similar to those purchased by the user in the recommendation list, and then recommends items that other buyers tend to purchase together [9]. For the dynamic profiling method to make a recommendation, user data should be accumulated; otherwise, it may result in the issue of a cold start, a chronic issue in collaborative filtering [10]. The model proposed in this study is a new type of effective profiling method.

In the past, style recommendations for apparel were merely to identify and recommend the consumers' favorite fashion through surveys. However, in recent years, various attempts were made in recommending apparel because of the development of deep learning technology [11]. Charles et al. (2018) conducted a study to recommend products by providing users with interpretive feedback on their personal preferences using visual information regarding their favorite apparel product images to overcome the limitations of image feature extraction using convolutional neural network (CNN), where information is expressed too highly, making it difficult to interpret [12]. However, the study focused on a black-box design, which is among the essential problems of deep learning, using a different approach from the method proposed in this study.

2.2. Implicit User Fashion Profiling Method

The implicit user fashion profiling method, as proposed in this study, collects data needed for learning using the implicit profiling method in which the user selects their preferred style among the fashion styles that the system proposes. This method uses the term implicit because the system can collect the characteristics of the style inherent in the pictures, and users do not have to separately mention or record their favorite style by selecting a preferred style from the pictures.

The implicit user fashion profiling collection system provides two styles of top and bottom as shown in Figure 1 at random. The user can select their preferred style from the image pictures. If there is no preferred style in the pictures, then the system can provide the next set of style images. Through the implicit user fashion profiling collection system, each user has a preferred style image profile. Through the collected individual profile, the system learns a user's preferred coordination recommendation model that recommends top and bottom, as well as a matching recommendation model that recommends top or bottom. The implicit user fashion profiling collects learning data in two manners: filtering by professionals and data collection from general users.



Figure 1. Overview of the implicit fashion profiling system.

2.2.1. Filtering by Professionals

There are two purposes to performing “filtering by professionals”. Firstly, before providing general users with fashion styles, it filters the styles that users will not prefer according to the opinions of fashion professionals that are saved beforehand. Secondly, it is performed to reduce the recommendation weight of similar style images, when teaching the model to recommend top or bottom or top and bottom matching.

2.2.2. User Profiling

General users obtain style images through a profiling system which is already filtered by professionals. As the proposed images were filtered by professionals, general users can select their preferred style from the images that were already screened for uncomfortable or unattractive styles.

2.3. Image2Vec-Based Feature Extraction Model

The bag-of-features method and scale-invariant feature transform (SIFT) are two approaches to expressing image features [13–15]. Recently, deep learning technology such as CNN was used to extract image features. It was also utilized for image classification and pattern recognition [16–18]. Research reported that the performance of a CNN model can be improved when pre-trained and pre-weighted values are used with a large volume of image data such as that of ImageNet [19,20].

To classify the categories of fashion images before searching for similar images and extracting properties, the system collected fashion product images from Amazon (amazon.com) and learned 66,000 cases composed of 33 product categories using a deep residual network model that showed good performance in the vision field and was composed of 34 pre-trained layers [21]. Table 1 shows some categories configured using four depths.

Table 1. Amazon category samples.

Depth 1	Depth 2	Depth 3	Depth 4
Apparel, Shoes, and Jewelry	Boys	Apparel	Jackets and Coats
	Men	Apparel	Shirts Shorts
	Women	Apparel	Shirts Shorts

Afterward, it extracted fashion product image features using a learned image classification model using a residual network with 34 layers, and it analyzed the precision to check if these features were significant. The performance of the fashion image category classifier taught using Amazon fashion images showed a precision of 96%.

To extract product image features from the learned deep residual networks, a global average pooling layer was used. According to previous studies, the method using a fully connected layer for the extraction of image features has disadvantages in that it is easy to overfit and it is highly dependent on dropout regularization [22,23]. However, when a global average pooling layer is used, there is an advantage in that the models can be interpreted because it directly maps the spatial average value of the feature map with the category information [24].

Figure 2 shows the process of extracting product image features using the global average pooling layer of the learned model. The extracted image feature vector is composed of 64 dimensions, and they are embedded in three-dimensional space after the dimension of the extracted product image features are reduced using t-distributed stochastic neighbor embedding (t-SNE). Figure 3 shows the result of the embedding. It was found that the products with similar images were clustered in the embedding space, and the extracted features yielded a meaningful image representation.

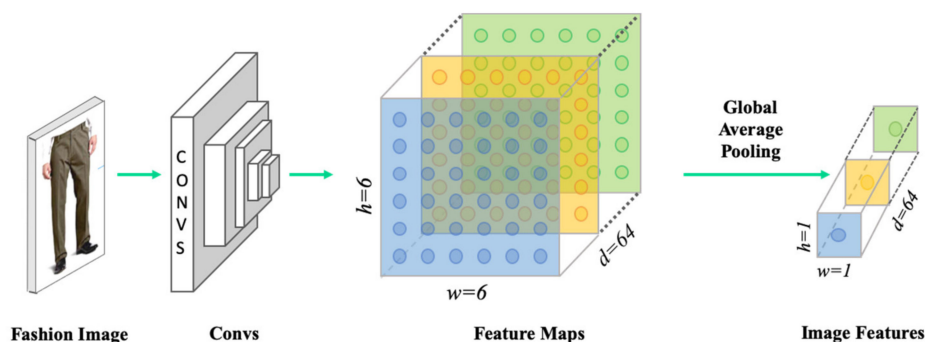


Figure 2. Extraction of features from a fashion image using a global average pooling layer.



Figure 3. Result of image feature embedding using t-distributed stochastic neighbor embedding (t-SNE).

3. Proposed Models

This study developed intelligent fashion techniques for Sketch-Product and personalized coordination composed of a Sketch-Product fashion retrieval model that overcomes the limitations of a text-based search method, as well as a vector-based user preferred fashion recommendation model applying an implicit user fashion profiling method to overcome the limitations of the preferred fashion profiling to search for and recommend fashion products.

3.1. Sketch-Product Fashion Retrieval Model

To develop the Sketch-Product fashion retrieval model, we firstly developed an image-based similar product retrieval model. Then, a sketch-based similar product retrieval model was additionally developed. It up-samples a sketch of the customer to the level of an image using deep learning technology and searches vector-based similar images with the feature vector value acquired from the up-sampled image. Figure 4 shows the structure of the Sketch-Product fashion retrieval model, including a sketch up-sampling module, an Image2Vec-based feature extraction model, the image-based similar product retrieval model, and the sketch-based similar product retrieval model.

3.1.1. Image-Based Similar Product Retrieval Model

Because of the inherent characteristics of the fashion domain, there are numerous types of fashion images in a category, and the fragmentation is severe. Therefore, it is very difficult for a user to find the product that they want using the text-based search. To resolve this problem, the image-based similar product retrieval model searches for fashion products that are similar to the fashion image that the user provides.

The vector space model represents the object of the search as a feature vector in space [25] and performs a search by quantifying the similarity using cosine similarity or Euclidean distance. If the image classifier using a CNN model is trained through a map learning method, feature vectors that can be generalized are acquired [26–29].

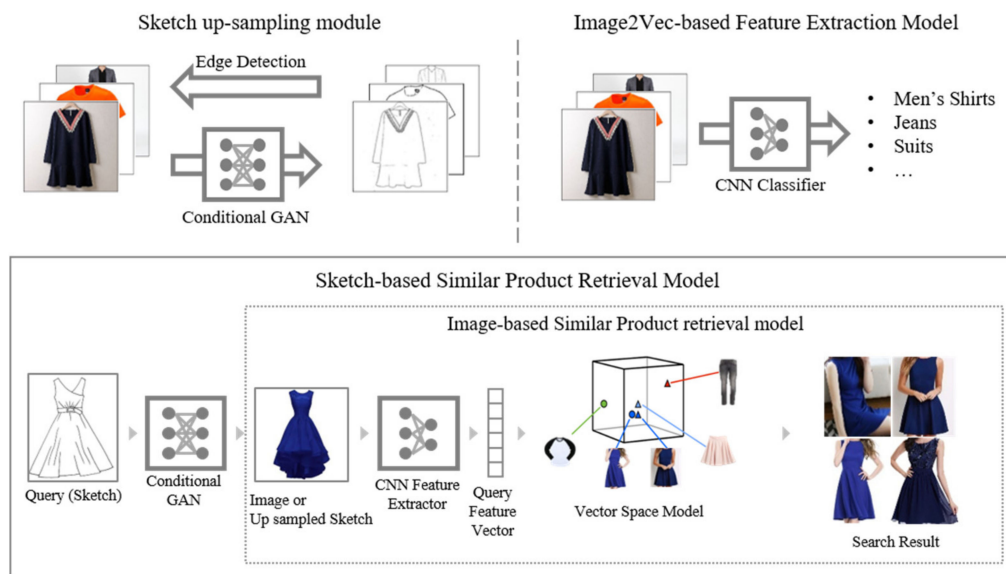


Figure 4. Structure of Sketch-Product fashion retrieval model.

If one uses the Image2Vec-based feature extraction model described in Section 2.3, the image can be converted into a real number vector in dimension n . If the real number that is obtained from the image is mapped in the vector space in dimension n , images with similar characteristics will be near the original image in the vector space. Vector-based image retrieval is performed by extracting image features of the image that the user searches, converting it to a real number vector, and arranging the images according to the vector value.

To reduce the time complexity, an approximate nearest neighbor search based on the Hddix library is used to index extracted image features and save them in the retrieval engine. This can reduce the search time by converting the image features on the high-dimensional space using the Hddix library into compact binary codes [30–32].

3.1.2. Sketch-Based Similar Product Retrieval Model

The sketch-based similar product retrieval model trains the Image2Vec-based feature extraction model using the map learning method and acquires an image feature vector. Using this image feature extraction model, it converts the fashion images for the search and query images created from the sketch up-sampling model into vector values, and then obtains similar product search results using the cosine similarity measurement method.

To perform a similar image search using the sketch provided by the user as a query, up-sampling of the sketch to an image level is required. The generative adversarial network (GAN) is a generation model used to train two contrasting neural networks [33], and it recently attracted attention in the image generation field as it provides better results than conventional methods [34,35]. The proposed model up-samples sketches to the level of fashion images using GAN.

To train the artificial neural network (ANN), a large volume of learning data is needed. For the training of the sketch up-sampling model, apparel image data tagged with category information, and the matching data between the apparel image and the sketch are needed. In the case of apparel image data tagged with category information, one can collect the crawler of the apparel e-commerce webpage. As most apparel sales webpages provide category information to which the corresponding apparel belongs, one can construct learning data using this information. When an apparel image is collected, one can convert it into a sketch using algorithms such as canny edge detection [36] or holistically nested edge detection (HED) [37]. These sketches can be used for training the sketch up-sampling module.

3.2. Vector-Based User Preferred Fashion Recommendation Model

The vector-based user preferred fashion recommendation model is a DNN-based recommendation model. The model learning process is divided into two stages. Firstly, it is pre-trained with profile data acquired from pre-filtering by professionals in the fashion industry. This was conducted to provide a default weight value to recommend fashion using the implicit user fashion profiling collection system. Secondly, the weight value of the model acquired from the filtering process by professionals is set as the default of the individualized model. During the learning process of the recommendation model, the weight value set as the default is updated in the user preference profile, and it is learned according to the user preferred styles. Using the profile data collected during the learning process, each user will have a separate learning model.

Figure 5 shows the model overview of the whole process. Each user has a profile collected through the system as described in Section 2.2, and the system converts the top and bottom image in the profile to vector values using the Image2vec-based feature extraction model specified in Section 2.3. Each image vector is used as Item1 and Item2, and they are merged into a vector through concatenation. During the learning process, general users have a weight value that is pre-trained through filtering by professionals as the default. When individual profiling data are added, they will be reflected in the model through weight training. The models learned by the system are (1) a vector-based user preferred fashion recommendation (style recommendation) and (2) a vector-based user preferred fashion matching recommendation (apparel matching recommendation). The former recommends a top and a bottom based on the user profile, and the latter recommends a top or a bottom matching the top or the bottom that the user presents. In other words, when a user presents a top, the system recommends a bottom matching the top, while, if a user presents a bottom, it recommends a top matching the bottom.

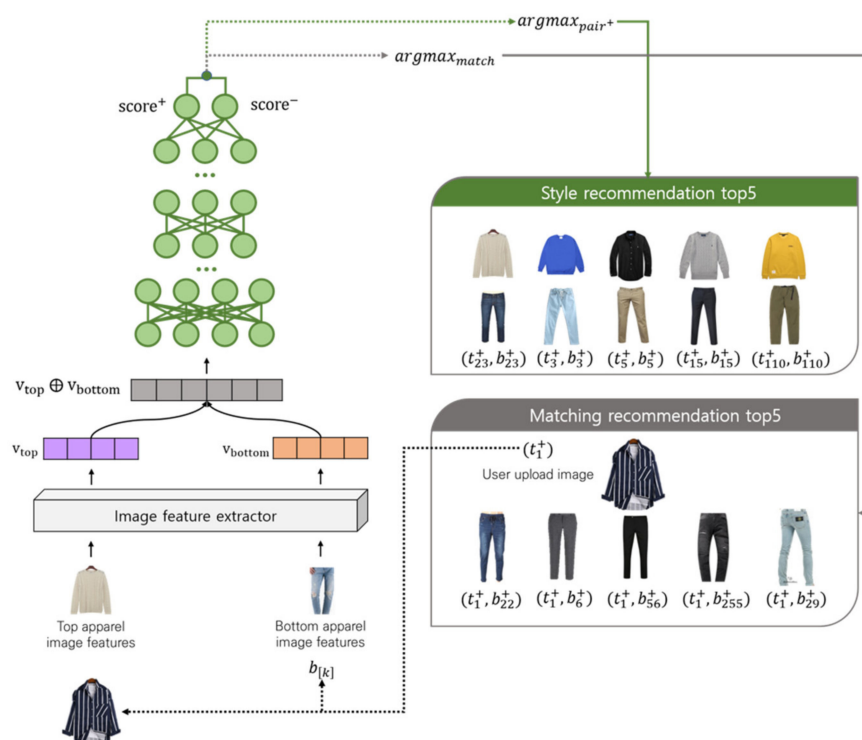


Figure 5. Overview of the vector-based user preferred fashion recommendation model.

3.2.1. Problem Formulation

The model uses the labeled data through the filtering process by professionals as learning data, and the format of the data used for learning is a pair composed of a top and a bottom.

The images of the top and the bottom are converted to a vector using the Image2Vec-based feature extraction model specified in Section 2.3. The images in the profiled data are $\{t_1, t_2, \dots, t_n\} \in top$ and $\{b_1, b_2, \dots, b_n\} \in bottom$, and they are labeled as $pair^-, pair^+ \in \{0, 1\}$ according to the preference. The images are labeled as $pair^+ = \{(t_1^+, b_1^+), (t_2^+, b_2^+), \dots, (t_n^+, b_n^+)\}$ when they are those that the user prefers and as $pair^- = \{(t_1^-, b_1^-), (t_2^-, b_2^-), \dots, (t_n^-, b_n^-)\}$ when they are those that the user does not prefer. Here, each image t and b can be converted to a vector as v_{top} and v_{bottom} , respectively, through the image feature extractor as follows:

$$v_{top} = ImageFeatureExtractor(t)v_{bottom} = ImageFeatureExtractor(b)x = v_{top} \oplus v_{bottom} \quad (1)$$

The vector value for the top and the bottom can be defined through concatenation (i.e., \oplus). The final model for the learning aims to maximize the distance between $pair^-$ and $pair^+$ and to find t_n^+ and b_n^+ that satisfy the goal as follows:

$$argmax_{(t_n^+, b_n^+) \in pair^+} \| score^+ - score^- \| \quad (2)$$

3.2.2. Model Learning

x , as defined in Section 3.2.1, is used for learning through the DNN. The score value on the preferred style is shown in Equation (3). $g^{[l]}$ and l^{th} are the layer's activation functions, while $W_{[l]}$ and $b_{[l]}$ are the weight and bias learned in the l^{th} layer, respectively. x^+ and x^- share the same weight value of $W_{[l]}$. The final score originates through the output layer. W_o and b_o are the weight and the bias from each output layer, respectively, and they are learned until the model is completed.

$$x^+ = g^{[l]}(W_{[l]}x + b_{[l]})x^- = g^{[l]}(W_{[l]}x + b_{[l]})score^+ = W_o x^+ + b score^- = W_o x^- + b \quad (3)$$

The loss function of the proposed model is learned to optimize the point-wise ranking loss function [38]. Point-wise ranking is used to prioritize pairs according to a relative ranking score compared to other pairs. Following this process, the pairs with higher ranking scores can be induced and the pair with the highest-ranking score is recommended to the user. The loss function is shown in Equation (4).

$$L = min[\uparrow_2(score^+) - \downarrow_2(score^-), \lambda] \quad (4)$$

where \uparrow_2 is the l2-regularization [39] and λ is the minimum threshold which is set as 10 through the heuristic experiments.

3.2.3. Fashion Matching Recommendation Model

The fashion matching recommendation model recommends a matching top or bottom when the user presents a bottom or top, respectively. For example, if the user presents a top, the system recommends a bottom that matches the presented top. The fashion matching recommendation model has the same structure of the model specified in Sections 3.2.1 and 3.2.2. If the user presents t_1^+ , it sets the input as $pair^+ = \{(t_1^+, b_1^+), (t_1^+, b_2^+), \dots, (t_1^+, b_n^+)\}$ to compare all bottoms that can match the top, induces a score value, and compares them. Equation (5) satisfies the following premise: if the user would like to obtain a recommendation for the bottom as $\{t_n^+, b_n^+ \in pair^+ | match = t_n^+ \text{ or } b_n^+\}$, $match$ becomes t_n^+ and the input to induce the score from the system is $(t_n^+, b_{[k]}^+)$. Here, $[k] = \{1, 2, 3, \dots, n\}$.

$$argmax_{match} score^+ \quad (5)$$

4. Experiments and Results

To assess the Sketch-Product fashion retrieval model, a search was performed for men's apparel. The performance was quantitatively measured, and a verification experiment of the implemented system

was performed. Picture and category information collected from a domestic apparel e-commerce service was used as experimental data. Using the method proposed by Isola (2017), the system learned the tops among the men's apparel, and the products to be searched were limited to five categories [40].

For the performance measurement, "Precision at 5" was used. When the search result product category matched the query product category, it was defined as relevant. Test data comprised 1000 queries. Apparel pictures that were not included in the learning data were used after processing using the HED method [41]. As the base model for the performance comparison, a method that randomly present search results was used. As the search category was five and the proper category was one out of five, the "Precision at 5" of the model was 0.200.

Using the image-based similar product retrieval model, we performed a similar product search on 1000 apparel product images and measured the performance. Out of the 5000 product images from the 1000 queries, the images that fell under the same category of the search target totaled 3871; thus, the Precision at 5 was 0.774.

In the performance verification of the sketch-based similar product retrieval model, we obtained 5000 search results from 1000 queries. Out of these, 2227 products fell under the same category of the search target; thus, the Precision at 5 was 0.445.

When the base model was a null hypothesis, the proposed model showed a statistically significant difference at $p < 0.01$. This showed that the proposed model dismissed the null hypothesis and showed significant performance enhancement. The performance measurement "Precision at k " was also used. Table 2 shows the performance experiment results of each model.

Table 2. Precision at k results.

Model	k	N of Data	N of Successes	Precision at k
Image-based similar product retrieval	1	1000	351	0.351
	3	3000	1838	0.613
	5	5000	3871	0.774
	10	10,000	8092	0.809
Sketch-based similar product retrieval	1	1000	183	0.183
	3	3000	1073	0.358
	5	5000	2227	0.445
	10	10,000	4925	0.493

The experiment of the vector-based user preferred fashion recommendation model was performed on 10,000 cases of menswear web shopping mall crawling data. Each image had tagging information such as season (spring/fall/summer/winter) and specification information such as top or bottom. In the tagging of each data point, there was additional tagging information from fashion professionals and information from shopping malls via crawling. Ten network layers were used, and the input vector was composed of 64 dimensions. Each layer consisted of 500 dimensions, the batch size was set as 10, and the L2-regularizer learning rate was set as 0.01. For optimization, the Adam optimizer [42] was used.

Figure 6a is the result of the preferred fashion recommendation of the top five tops and bottoms through the user profile, and Figure 6b,c are the matching recommendation results for a top and bottom that the user presented, respectively. When the user presents the clothing that he/she prefers, the system recommends the matching pair. As it is difficult to quantitatively evaluate the individualized fashion recommendation, it is qualitatively evaluated in this study. According to previous studies of the individualized recommendation model, Packer et al. (2018) used an affinity vector to represent the individualized preference and proposed a refined recommendation through a scaled affinity vector; however, the study did not perform a quantitative evaluation. It verified the proposed model by presenting samples of the recommendation results [12]. McAuley et al. (2015) also performed model verification by presenting recommendation result samples of the query [41].

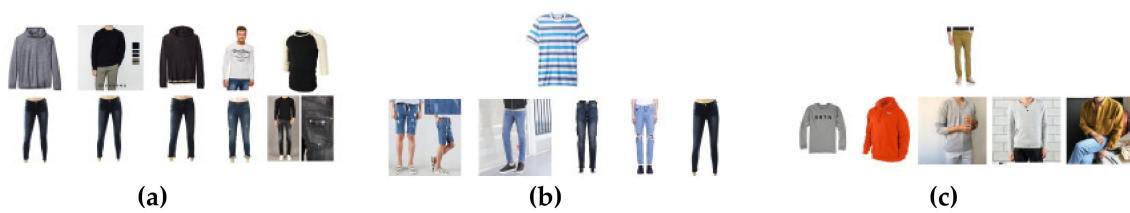


Figure 6. Recommendation results: (a) example of top and bottom apparel style recommendation (top five); (b) example of bottom matching recommendation for top apparel (top five); (c) example of top matching recommendation for bottom apparel (top five).

5. Conclusions

The size of the fashion market is continuously increasing, and the digital industry in the fashion market created a new paradigm. While customers can access various fashion apparel through the digitalized paradigm and they have more options to choose from, it is not easy for them to search for the style that they want online. General users do not know what keyword they should use to search for the style they want. Accordingly, shopping malls need to implement a more effective information retrieval system that should be able to recommend clothing by identifying the preference of the consumer to resolve limitations and provide optimal service. Therefore, it is increasingly important for shopping malls to implement a system that supports efficient searches and recommendations.

This research developed (1) a Sketch-Product fashion retrieval model that allows users to use sketch images instead of text for their search and to obtain fashion images similar to the sketch images as search results, and (2) a vector-based user preferred fashion recommendation model using an implicit profiling method. Intelligent fashion techniques for Sketch-Product and personalized coordination using the aforementioned models were developed.

To develop the proposed system, we firstly developed a feature extraction model and then an image-based similar product retrieval model. We applied technology that converts a sketch fashion drawing to an image to develop the sketch-based similar product retrieval model. To convert the fashion sketch to an image, we up-sampled the sketch that the user provided using GAN, converted it to an image feature vector using the CNN, and then obtained similar images from the vector space model. The Sketch-Product fashion retrieval model showed significantly improved performance. It is expected that this may support potential consumers with currently insufficient keyword search capabilities to conduct an effective fashion product search, as well as improve service satisfaction and sales profitability.

The vector-based user preferred fashion recommendation model is pre-trained with profile data obtained through pre-filtering by professionals and the weight of the model is set as the default. When the user preference profile is updated, the default weight of the model is learned according to the individual preference. The data format for the training is represented by pairs composed of a top and bottom. The images of the tops and the bottoms are converted to vectors through the Image2Vec-based feature extraction model. The recommended preferred fashion results showed positive performance.

If utilized in the fashion field, this model may improve customer satisfaction and increase sales revenue by recommending the preferred style of the user before they start the search process. At present, a system based on the proposed model is being implemented. Future studies will analyze the performance of the system and the purchase pattern of consumers via the introduction of the system to actual shopping mall sites.

Author Contributions: Conceptualization, H.L. and J.J.; methodology, C.L., S.L., and D.L.; software, J.J., C.L., S.L., and D.L.; validation, C.L., S.L., and D.L.; writing—original draft preparation, J.J.; writing—review and editing, J.J., C.L., S.L., and D.L.; visualization, J.J., C.L., S.L., and D.L.; project administration, H.L.; funding acquisition, H.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the MSIT (Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2018-0-01405) supervised by the IITP (Institute for Information and Communications Technology Promotion). This work was supported by the National Research Foundation of Korea (NRF) grant funded by the MSIT (No. NRF-2017M3C4A7068189).

Acknowledgments: The authors thank the MSIT, IITP, and NRF for supporting the research and project.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Korea Federation of Textile Industries. *Korea Fashion Market Trend 2019 Report*; Korea Federation of Textile Industries: Seoul, Korea, 2019.
2. Korea Fashion Association. *Global Fashion Industry Survey*; Korea Fashion Association: Seoul, Korea, 2019.
3. Online Marketing Trends -Clothing/Fashion. Available online: <http://www.econovill.com/news/articleView.html?idxno=376137> (accessed on 25 February 2020).
4. Joo, S.; Ha, J. Fashion industry system and fashion leaders in the digital era. *J. Korean Soc. Cloth. Text.* **2016**, *40*, 506–515. [[CrossRef](#)]
5. Kang, C.; Kim, H. *2018 Industry Credit Outlook*; Korea Ratings: Seoul, Korea, 2017.
6. What stitch fix figured out about mass customization. Available online: <https://hbr.org/2015/05/what-stitchfix-figured-out-about-mass-customization> (accessed on 25 February 2020).
7. Colson, E. Using human and machine processing in recommendation systems. In Proceedings of the First AAAI Conference on Human Computation and Crowdsourcing, Palm Springs, CA, USA, 6–9 November 2013; pp. 16–17.
8. Chen, X.; Zhang, Y.; Qin, Z. Dynamic Explainable Recommendation based on Neural Attentive Models. In Proceedings of the AAAI Conference on Artificial Intelligence, Honolulu, HI, USA, 27 January–1 February 2019; Volume 33, pp. 53–60.
9. Linden, G.; Smith, B.; York, J. Amazon. com recommendations: Item-to-item collaborative filtering. *IEEE Internet Comput.* **2003**, *7*, 76–80. [[CrossRef](#)]
10. Lam, X.N.; Vu, T.; Le, T.D.; Duong, A.D. Addressing cold-start problem in recommendation systems. In Proceedings of the 2nd International Conference on Ubiquitous Information Management and Communication ACM, Suwon, Korea, 31 January–1 February 2008; pp. 208–211. [[CrossRef](#)]
11. Eke, C.I.; Norman, A.A.; Shuib, L.; Nweke, H.F. A survey of user profiling: State-of-the-art, challenges, and solutions. *IEEE Access* **2019**, *7*, 144907–144924. [[CrossRef](#)]
12. Packer, C.; McAuley, J.; Ramisa, A. Visually-aware personalized recommendation using interpretable image representations. *arXiv*, 2018; arXiv:1806.09820.
13. Perronnin, F.; Dance, C. Fisher kernels on visual vocabularies for image categorization. In Proceedings of the 2007 IEEE Conference on Computer Vision and Pattern Recognition, Minneapolis, MN, USA, 17–22 June 2007; pp. 1–8.
14. Lowe, D.G. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* **2004**, *60*, 91–110. [[CrossRef](#)]
15. Wang, C.; Zhang, Y.; Zhou, X. Robust image watermarking algorithm based on ASIFT against geometric attacks. *Appl. Sci.* **2018**, *8*, 410. [[CrossRef](#)]
16. Kannan, A.; Talukdar, P.P.; Rasiwasia, N.; Ke, Q. Improving product classification using images. In Proceedings of the 2011 IEEE 11th International Conference on Data Mining, Vancouver, BC, Canada, 11 December 2011; pp. 310–319.
17. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. *arXiv*, 2015; arXiv:1512.03385.
18. Cui, X.; Qi, M.; Niu, Y.; Li, B. The Intra-Class and Inter-Class Relationships in Style Transfer. *Appl. Sci.* **2018**, *8*, 1681. [[CrossRef](#)]
19. Deng, J.; Dong, W.; Socher, R.; Li, L.J.; Li, K.; Fei-Fei, L. Imagenet: A large-scale hierarchical image database. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 248–255.
20. Wang, W.C.; Chen, L.B.; Chang, W.J. Development and experimental evaluation of machine-learning techniques for an intelligent hairy scalp detection system. *Appl. Sci.* **2018**, *8*, 853. [[CrossRef](#)]

21. Lee, D.; Lee, C.; Lee, S.; Jo, J.; Lim, H. Automatic classification and embedding of fashion products using deep residual network. *Korean Inst. Info. Scient. Engin.* **2017**, 975–977.
22. Krizhevsky, A.; Sutskever, I.; Hinton, G. Imagenet classification with deep convolutional neural networks. *Adv. Neural Info. Proc. Syst.* **2012**, 25, 1106–1114. [[CrossRef](#)]
23. Hinton, G.E.; Krizhevsky, N.S.A.; Sutskever, I.; Salakhutdinov, R. Improving neural networks by preventing co-adaptation of feature detectors. *arXiv*, 2012; arXiv:1207.0580.
24. Lin, M.; Chen, Q.; Yan, S. Network in network. In Proceedings of the International Conference on Learning Representations, Banff, AB, Canada, 14–16 April 2014.
25. Turney, P.D.; Pantel, P. From frequency to meaning: Vector space models of semantics. *J. Artif. Intell. Res.* **2010**, 37, 141–188. [[CrossRef](#)]
26. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, 521, 436–444. [[CrossRef](#)]
27. Sharif Razavian, A.; Azizpour, H.; Sullivan, J.; Carlsson, S. CNN features off-the-shelf: An astounding baseline for recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Columbus, OH, USA, 23–28 June 2014.
28. Donahue, J.; Jia, Y.; Vinyals, O.; Hoffman, J.; Zhang, N.; Tzeng, E.; Darrell, T. Decaf: A deep convolutional activation feature for generic visual recognition. In Proceedings of the International Conference on Machine Learning, Beijing, China, 22–24 June 2014; pp. 647–655.
29. Babenko, A.; Slesarev, A.; Chigorin, A.; Lempitsky, V. Neural codes for image retrieval. In Proceedings of the European Conference on Computer Vision, Zurich, Switzerland, 6–12 September 2014; pp. 584–599.
30. Jegou, H.; Matthijs, D.; Cordelia, S. Product quantization for nearest neighbor search. *Pattern Anal. Mach. Intell. IEEE Trans.* **2011**, 33, 117–128. [[CrossRef](#)]
31. Weiss, Y.; Antonio, T.; Rob, F. Spectral hashing. In *Advances in neural Information Processing Systems 21*; NeurIPS: San Diego, CA, USA, 2009; pp. 1753–1760.
32. Liu, C.; Lu, T.; Wang, X.; Cheng, Z.; Sun, J.; Hoi, S.C. Compositional coding for collaborative filtering. In Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval, Paris, France, 21–25 July 2019; pp. 145–154.
33. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Bengio, Y. Generative adversarial nets. In Proceedings of the Advances in Neural Information Processing Systems, Montreal, QC, Canada, 8–13 December 2014; pp. 2672–2680.
34. Ledig, C.; Theis, L.; Huszár, F.; Caballero, J.; Cunningham, A.; Acosta, A.; Shi, W. Photo-realistic single image super-resolution using a generative adversarial network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; Volume 2, p. 4.
35. Zhang, H.; Xu, T.; Li, H.; Zhang, S.; Huang, X.; Wang, X.; Metaxas, D. Stackgan: Text to photo-realistic image synthesis with stacked generative adversarial networks. *arXiv*, 2017; arXiv:1612.03242.
36. Canny, J. A computational approach to edge detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **1986**, 6, 679–698. [[CrossRef](#)]
37. Xie, S.; Tu, Z. Holistically-nested edge detection. In Proceedings of the IEEE International Conference on Computer Vision, Beijing, China, 21 October 2005; pp. 1395–1403.
38. Liu, T.Y. Learning to rank for information retrieval. *Found. Trends@Inf. Retr.* **2009**, 3, 225–331. [[CrossRef](#)]
39. Wager, S.; Wang, S.; Liang, P.S. Dropout training as adaptive regularization. *Adv. Neur. Inf. Proc. Syst.* **2013**, 26, 351–359.
40. Isola, P.; Zhu, J.Y.; Zhou, T.; Efros, A.A. Image-to-image translation with conditional adversarial networks. *arXiv*, 2017; arXiv:1611.07004v3.
41. McAuley, J.; Targett, C.; Shi, Q.; Van Den Hengel, A. Image-based recommendations on styles and substitutes. In Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval, Santiago, Chile, 9–13 August 2015; pp. 43–52.
42. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv*, 2014; arXiv:1412.6980.

