A CNN-Based Advertisement Recommendation through Real-Time User Face Recognition

Gihwi Kim¹, Ilyoung Choi¹, Qinglong Li¹ and Jaekyeong Kim¹,₂,*

¹Department of Big Data Analytics, Kyung Hee University, 26, Kyungheedae-ro, Dongdaemun-gu, Seoul 02447, Korea; kiwi@khu.ac.kr (G.K.); choice102@khu.ac.kr (I.C.); leecy@khu.ac.kr (Q.L.)
²School of Management, Kyung Hee University, 26, Kyungheedae-ro, Dongdaemun-gu, Seoul 02447, Korea
*Correspondence: jaek@khu.ac.kr; Tel.: +82-2-961-9355

Abstract: The advertising market’s use of smartphones and kiosks for non-face-to-face ordering is growing. An advertising video recommender system is needed that continuously shows advertising videos that match a user’s taste and displays other advertising videos quickly for unwanted advertisements. However, it is difficult to make a recommender system to identify users’ dynamic preferences in real time. In this study, we propose an advertising video recommendation procedure based on computer vision and deep learning, which uses changes in users’ facial expressions captured at every moment. Facial expressions represent a user’s emotions toward advertisements. We can utilize facial expressions to find a user’s dynamic preferences. For such a purpose, a CNN-based prediction model was developed to predict ratings, and a SIFT algorithm-based similarity model was developed to search for users with similar preferences in real time. To evaluate the proposed recommendation procedure, we experimented with food advertising videos. The experimental results show that the proposed procedure is superior to benchmark systems such as a random recommendation, an average rating approach, and a typical collaborative filtering approach in recommending advertising videos to both existing users and new users. From these results, we conclude that facial expressions are a critical factor for advertising video recommendations and are helpful in properly addressing the new user problem in existing recommender systems.

Keywords: recommender system; computer vision; scale-invariant feature transform; deep learning; convolutional neural network

1. Introduction

As the number of Internet users increases, users’ interactions with multimedia devices are also growing rapidly. Recently, in fast food restaurants and cafes, non-face-to-face services that take orders through a smartphone app or a kiosk are increasing rather than receiving orders through people. Moreover, because of COVID-19, non-face-to-face orders are increasing at an explosive rate. Advertisers have not missed this opportunity and want to place various advertisements while processing orders through kiosks or smart apps. In addition, online video platform providers such as YouTube and TikTok are working to increase advertising revenue by commercializing video products [1]. Particularly, video content providers have introduced a recommender system to attain more website traffic by making it easy for users to find their favorite videos.

However, online advertising video or video recommendations face a unique problem. Users often become bored with advertising videos while watching them, even if they run for less than 30 s [2]. It is challenging to capture changes in user interests from profiles using users’ historical data over a brief period. Recent advances in digital image processing technology have made it possible to track the facial expressions of online advertising video viewers in real-time. Some researchers have argued that facial expressions are important in predicting a user’s preference [3,4]. For example, a user frowns when watching brutal scenes. That is, users’ facial expressions make it possible to predict their...
emotions. Therefore, we can utilize facial expressions to advertise video recommendations to users. They allow us to discover a user's dynamic preferences while watching an advertising video.

In this study, we propose an advertising video recommendation procedure based on facial expression changes to cope with users' dynamic preferences within a short period. When a user is watching an advertisement, their emotion is captured by a webcam mounted on a kiosk, monitor, or smartphone. Then, their emotions are compared with those of other users to decide whether to keep the advertising video the user is watching or quickly replace it with another advertising video. To achieve the purpose of this study, we adopted a deep learning approach. Many studies have proposed a recommendation approach named collaborative filtering (CF), a traditional technique that recommends items suitable for users based on neighbors' preferences or purchasing history [5–7]. However, using only such a CF technique creates a “cold start” issue, whereby the recommendations for new users suffer from unpredictability because of a lack of historical data on their past purchases. Additionally, it has a “first start” issue that cannot offer recommendations until a user has reflected their preferences. In addition, there are also issues regarding the scalability of the model, arising from the continued growth of users' purchasing history or preference data [8–10]. In other words, it is challenging to offer recommendations in real time because it takes a long time to calculate data owing to model scalability problems. Many studies have been conducted to supplement issues such as data sparsity and scalability [11–13]. The recent deep learning approach has shown high performance in image processing or natural language processing areas and has received much attention [14,15]. Many studies that applied the deep learning approach have steadily been proposed in the study recommender systems [16–18]. However, recommender systems using deep learning techniques have focused on encoding various types of input data, which was not possible in traditional CF. Recommender systems using deep learning have been developed to improve the performance of recommender systems by reflecting vast amounts of internet review data or images in existing sales data or service use records. However, as is the case with the system we propose, a system that directly receives video information and recommends it without any prior information about its user has not yet been developed.

Therefore, this study applies a deep learning approach to offer personalized advertising videos by capturing new users’ facial expressions in real time. Though the proposed procedure follows the principle of CF, we create a dynamic profile of users based on the changes in their facial expressions every moment instead of using their historical records. To achieve this purpose, we developed a recommender system that applied a convolutional neural network (CNN) to predict how a user likes the currently watched advertising video by recognizing their facial expressions. Additionally, to recommend new advertising videos to users, we developed a scale-invariant feature transform (SIFT), an algorithm-based similarity model, to search users with similar preferences in real time. To evaluate the proposed procedure, we compare the performance of three different recommendation approaches, a random system, an average rating-based (best-selling) system, and a typical CF-based system as benchmark systems using eleven food advertising videos. The experiment results indicate that the performance of the proposed procedure is better than that of the benchmark systems in advertising video recommendations to both existing users and new users. The main contributions of this study are as follows:

- The proposed methodology in this study effectively captures user facial expressions in real-time through a deep learning approach to solve information overload and data sparsity problems.
- The proposed methodology can effectively recommend advertising videos without distinguishing between new users (with no history of watching or rating) and existing users (whose previous records or ratings are available). Therefore, the proposed methodology does not suffer from the new user problem.
- The study conducts several experiments that use real-world facial expressions data to demonstrate that the performance of the proposed methodology was excellent in
the existing recommendation approach. We also found that facial expressions are an important factor for advertising video recommendations.

The rest of this study is organized as follows. Section 2 discusses related work on deep learning-based recommender systems, SIFT, and CNN. Section 3 discusses the proposed methodology and its components in detail. Section 4 describes the experimental design and discusses the results of this study. Section 5 summarizes the study, describes its limitations, and presents ideas for future work.

2. Related Work

2.1. Deep Learning-Based Recommender Systems

Recommender systems are information-filtering systems that can solve the information overload problem by filtering important pieces of information among information generated according to a user’s interests, preferences, or observed behavior for a specific item [19,20]. The information overload problem is a phenomenon in which it becomes more difficult for individuals to make good decisions as the amount of data increases [21]. With the rapid spread of Internet use between the early and mid-1990s, recommender systems based on CF were developed to estimate which of the many items users would like; thus, the problem of information overload was solved [22,23]. The objective of the recommender system is to make meaningful recommendations from user data for items or products of interest [24]. While customers have more choice, Internet shopping malls and service providers face challenges in providing personalized product advertisements to customers. The recommender system collects user preference information for items such as movies, songs, books, travel destinations, and websites. CF is an algorithm that recommends items liked by other users with similar tastes, and word of mouth is implemented as a computer system. CF has been widely used as a recommendation methodology to this day, but it is also prone to the problem of performance degradation owing to data sparsity, as well as cold-start problems, long tail problems, and scalability problems [25,26].

Deep learning has advanced the structure of recommender systems and provides several ways to improve the performance of recommender systems. The development of deep learning-based recommender systems has received great attention because these systems can overcome the limitations of existing CF-models (e.g., data sparsity problem, cold-start problem, scalability problem, and long tail problem) and achieve high recommendation quality [7,8,27–30]. Li, et al. [31] extracted latent features of user preferences or ratings using restricted Boltzmann machines and an undirected two-layer graphic model as a kind of graphic probabilistic model. Hu, et al. [32] extracted high-level features from low-level features for user preferences through a deep belief network, which is a deep neural network composed of multiple layers of latent variables. Auto Encoder, a model of deep learning, is used to reduce the dimensionality of the user item matrix and extract more latent features using the encoder output [33,34]. Ko, et al. [35] analyzed user behavior changes over time using a recurrent neural network (RNN), a deep learning model specialized in processing sequence data, and combined analysis results from RNN and the latent factor regarding user preferences to offer more accurate recommendations. In addition, the CNN, which shows excellent performance in tasks such as image recognition and object classification, extracts latent factors and latent features from raw data such as audio data, text data, and image data, thereby providing good performance for recommender systems [36,37]. Zhang, et al. [38] proposed a recommender system to solve a problem that makes it difficult for viewers to find interest anchors. Therefore, the authors developed a multi-head connected device to capture the preferences between anchors and viewers and extract related functions for the expression. The results show that the proposed model outperforms state-of-the-art recommendation models.

In summary, deep learning-based recommender systems have been developed to predict a user’s preference or recommend items suitable for said user’s preference by extracting the hidden meaning between the user and the item from various types of input data such as images and unstructured data, which were previously impossible to
analyze. However, it is difficult to find a deep learning-based recommender system that can recommend in real time by recognizing only a user’s facial expression without using pre-registered data or information such as purchase records or images. In terms of existing related studies, as shown in Table 1, most researchers collect user facial expressions and extract features to build dynamic user profiles. However, most studies applied simple heuristic techniques instead of computer vision approaches to extract user facial features. Additionally, most studies applied the CF method in the recommendation phase. However, there are many limitations to this traditional approach because it is critical to capture a user’s facial expression and recommend advertisements quickly and accurately in real time. Users can quickly become bored, even with short videos. However, it is difficult to capture changes in user interest in a profile using the user’s historical data for a short period. Therefore, it has become necessary to capture a user’s facial expressions accurately and efficiently in real-time. In this study, we develop a recommender system through a comprehensive deep learning approach to recognize user expressions and predict how they enjoy the advertisement they are currently watching. This study elaborately extracts user facial expression characteristics in real time by applying SIFT, widely used in computer vision, to reduce the existing gap in such studies. We also applied CNN to accurately predict user preferences based on the extracted facial expression. The proposed methodology can effectively recommend advertising videos without distinguishing between new users, and it does not suffer from the new user problem.

Table 1. Summary of related work.

| Name of Work | Author | Year | Description |
|--------------|--------|------|-------------|
| A hybrid approach to building face shape classifier for hairstyle recommender system | Pasupa, et al. [39] | 2019 | This study proposed an automatic hairstyle recommender system according to face type. Feature extraction was performed from images through a deep learning model and classified based on the support vector machine. |
| Enhancing Retailer Marketing with a Facial Recognition Integrated Recommender System | Wu, et al. [40] | 2015 | This study proposed a facial recognition integrated recommender system to address profiling problems (user information recognition) in retail stores. Based on this, the user’s information (gender, age, and facial expression) is predicted by recognizing the face. |
| Enhancing recommender systems for TV by face recognition | De Pessemier, et al. [41] | 2016 | This study proposed a system that supplemented the TV content recommender system by detecting and recognizing the facial emotions of users watching TV. Through the proposed system, a task was conducted to predict age, gender, and emotion. |
| Music Recommendation System Based on Facial Emotion Recognition | Visnu Dharsini, et al. [42] | 2020 | This study used facial recognition technology to predict user emotions and developed a music recommender system. To this end, faces were recognized using AdaBoost, and facial emotions were classified into eight categories through an SVM. |

2.2. Scale-Invariant Feature Transform

Feature detection and image matching are the most critical tasks in the field of machine vision. Because they have different computational efficiency or accuracy depending on which feature detector and descriptor extraction algorithm are used, it is critical to choose a suitable algorithm according to feature matching tasks [43]. Algorithms such as SIFT, speeded up robust feature (SURF), and binary robust invariant scalable keypoint (BRISK) are mainly used, and each algorithm has differences in performance [44]. SIFT is known to be most robust in feature detection and matching. In particular, its robustness is apparent.
in terms of image scale, rotation, and affine transformation. SURF is based on SIFT and changes filter size instead of image pyramid. SURF is robust in relation to image scale and rotation, but weak for affine transformation; another advantage is its relatively fast calculation speed compared with SIFT [45]. BRISK is a binary corner detection algorithm robust to changes in image scale and rotation. It has the advantage of enabling a faster calculation than SIFT and SURF [46]. According to previous studies, SIFT was a little disadvantageous in terms of speed but had excellent accuracy [43,47]. Therefore, we perform feature extraction by applying the SIFT algorithm to accurately and effectively extract features included in user facial expressions.

The SIFT algorithm is an image descriptor developed for image-based matching and recognition [48,49]. The image descriptor is a means of expressing images such as keyframe and face, and it is used for image comparison when extracting scene transitions and searching for similar images. The SIFT algorithm extracts characteristic points unique to images, and it is also a strong algorithm in the detection of many environmental changes such as image size change, deformation, rotation, lighting change, and so on. The SIFT algorithm has proven to be useful for experimentally measuring the similarity between images, matching images, and object recognition [50]. The SIFT algorithm consists of four main processes: scale–space extrema detection, orientation assignment, keypoint localization, and keypoint descriptor. Scale–space extrema detection creates scale–space and detects extrema. Scale–space refers to images obtained by creating an image pyramid by resizing an original image with various scales and then using the image pyramid’s octave image as a growing Gaussian blur scale factor. After obtaining the scale–space, the difference of Gaussian (DoG) image is obtained. For DoG image generation, the difference in the gaussian function is used to perform subtraction operations with two different Gaussian-blurred images in octave, and potential interest points are identified. Finally, extrema detection is performed using DoG. If the coordinates are considered to be the local minimum point or the local maximum point, they are classified into a group of keypoint candidates. In DoG target, the target pixel (x-indicated) value is compared to a total of 26 surrounding pixels (gray points) to determine whether it is a vertex. In the keypoint localization phase, the candidate keystone is not located on the correct coordinate system and is unsuitable for size and location determination. Therefore, an unstable candidate keypoint is removed depending on stability measurements, and only a stable keypoint is selected. The orientation assignment phase determines the gradient direction for each keypoint based on the local image patch. The local image patch determines a 16 × 16 patch around the keypoint, and after Gaussian blurring the image in it, the orientation and size of the gradient are determined for each point. In the final phase, a keypoint descriptor is created to express the characteristics of the keypoint. The final phase uses the extracted keypoint and descriptors to match keypoint; this is called keypoint matching. Keypoint matching is an algorithm that calculates the Euclidean distance for each keypoint between two images and matches the closest keypoint. Image matching tasks such as object detection, recognition, image retrieval, and tracking are among the most challenging tasks in the field of computer vision. Keypoint matching has been reported to have good results in object recognition and object detection [49]. Therefore, this study developed a SIFT algorithm-based similarity model to search for users with similar preferences in real time by comparing facial expression changes between two users using keypoint matching.

2.3. Convolution Neural Networks

CNN is known to be the most active field of research in deep neural networks [51,52] and was first introduced in the study “Backpropagation applied to handwritten zip code recognition” published by LeCun in 1989 [53]. LeCun later proposed the first CNN, a network called LeNet, in 1998 [54]. CNN is mainly used to solve difficult pattern recognition tasks, mainly focusing on images, and consists of an accurate but simple architecture [52]. Recently, CNN has been recognized as a powerful tool that can be used
in various areas such as face, image, video, and voice analysis [51]. CNN is similar to traditional artificial neural network (ANN), which are composed of neurons that self-optimize through learning [52]. However, CNN is more proficient at reducing the number of ANN parameters. This allowed researchers and developers alike to solve tasks that could not be solved with a classic ANN, allowing them to access larger models. CNN assumes that features do not have spatial dependence in problem solving [51]. For example, in face recognition, a face is recognized even if that face exists in an arbitrary position in a given image. Another important aspect of CNN is that abstract features are extracted when an input propagates to a deeper layer; as layers are stacked, local features are advanced to global features, thus solving the problem of not reflecting the entire relationship of the image, which is a disadvantage of fully connected neural networks. As a result, it has robustness in the transformation of input data. NN creates a feature map from an input image through a convolution filter. If it extracts several different features, it can set the number of convolution kernels to extract different features. Sub-sampling reduces the size of the feature map and, through it, topology invariance can also be obtained. After several stages of convolution and sub-sampling, the size of the feature map decreases, leaving only the robust features that can represent the whole. The global feature obtained in this way is connected to the fully connected network (FCN) input. Like the feature of ANN, it is possible to produce an optimal recognition result through learning. In this study, the user’s face image is extracted and represented as a matrix in 3D. The difference between the current image matrix and the previous image matrix is obtained, and this is defined as a face change image. We developed a CNN-based rating prediction model trained on the obtained face change image over time and predicted a rating with FCN.

3. A Face Recognition-Based Recommender System

This study aims to develop a user-customized advertising video recommender system to search for similar users in real time and to recommend advertisements suitable for users’ preferences. A CNN-based prediction model was developed to predict a user’s rating while watching an advertising video in real time before watching the advertisement. It uses a CNN that learns the facial expression changes in people who have seen the advertising video in the past and the user’s real-time ratings. Additionally, a SIFT algorithm-based similarity model was developed to search for users with similar preferences in real time. Using the keypoint matching of the SIFT algorithm, the user’s facial expression changes over time are compared with other users to find neighbors similar to the user. To summarize the flow of our proposed recommender system, it is to continuously predict the rating of the advertising video using CNN while the target user is watching the advertisement. Additionally, when the predicted rating is below a certain threshold value, similar neighbors are found through the SIFT model, and advertising videos that have been evaluated by the neighbors in the past are recommended to the target user.

3.1. Overall Process

The overall process of our suggested methodology is shown in Figure 1. When a user watches an advertisement in front of the webcam, their appearance is photographed through the webcam. In addition, the user who watches the advertisement assigns his/her preference for each advertisement on a five-scale score. Then, the face image of the user in the video is extracted from the video at 0.5 s intervals. The rating prediction process is composed of independent CNN models every 0.5 s, and each CNN model learns to predict the rating from the face image of the corresponding time. The image of the first time is compared with the face image of each time, and the degree of change in the face image is measured with the SIFT algorithm. Thereafter, if users with similar facial image gradients are matched, then advertisements preferred by similar users are recommended.
3.2. Data Collection and Data Pre-Processing

A user who views an advertisement rates the advertisement in question. Therefore, user \( i \)'s rating for advertisement \( j \), \( r_{ij} \) can be expressed as Table 2, and Table 3 is an example of actual experimental data.

### Table 2. User rating on each item.

| User | Rating of Item 1 | Rating of Item 2 | ... | Rating of Item \( j \) |
|------|------------------|------------------|-----|---------------------|
| \( u_{i1} \) | \( r_{1,1} \) | \( r_{1,2} \) | ... | \( r_{1,j} \) |
| \( u_{i2} \) | \( r_{2,1} \) | \( r_{2,2} \) | ... | \( r_{2,j} \) |
| ... | ... | ... | ... | ... |
| \( u_{i,n-1} \) | \( r_{i-1,1} \) | \( r_{i-1,2} \) | ... | \( r_{i-1,j} \) |
| \( u_{i,n} \) | \( r_{i,1} \) | \( r_{i,2} \) | ... | \( r_{i,j} \) |

### Table 3. Actual user rating data.

| User  | Rating of Item 1 | Rating of Item 2 | ... | Rating of Item \( 11 \) |
|-------|------------------|------------------|-----|---------------------|
| \( u_{i1} \) | 4 | 3 | ... | 3 |
| \( u_{i2} \) | 1 | 3 | ... | 4 |
| ... | ... | ... | ... | ... |
| \( u_{i,76} \) | 1 | 1 | ... | 3 |
| \( u_{i,77} \) | 2 | 3 | ... | 5 |

The user’s face image is extracted each time from the real-time image of the user’s face who watches the advertisement. The face image of the user in the video uses a face detection model using CNN based on max-margin object detection (MMOD). Images are captured at 0.5 s intervals from the video, and the user’s face in the image is recognized and extracted, as shown in Figure 2. Our methodology proceeds to predict users’ ratings and search for similar users from the previously extracted images.
The data collection and preprocessing are shown in Figure 3. When user \( i \) watches an advertisement about item \( j \), user facial data and user rating data are accumulated and matched for later use.

The face data for each time extracted through the face detection model can be represented as shown in Table 4.

### Table 4. Face data about item \( j \).

|          | Face of Time\(_1\) for Item\(_j\) | Face of Time\(_2\) for Item\(_j\) | \ldots | Face of Time\(_T\) for Item\(_j\) |
|----------|------------------------------------|------------------------------------|--------|------------------------------------|
| user\(_1\) | \( F_{1,j,1} \)                  | \( F_{1,j,2} \)                  | \ldots | \( F_{1,j,T} \)                  |
| user\(_2\) | \( F_{2,j,1} \)                  | \( F_{2,j,2} \)                  | \ldots | \( F_{2,j,T} \)                  |
| \ldots   | \ldots                            | \ldots                            | \ldots | \ldots                            |
| user\(_{i-1}\) | \( F_{i-1,j,1} \)              | \( F_{i-1,j,2} \)              | \ldots | \( F_{i-1,j,T} \)              |
| user\(_i\) | \( F_{i,j,1} \)                  | \( F_{i,j,2} \)                  | \ldots | \( F_{i,j,T} \)                  |

In the similar user search algorithm, the face image is used as input data because similarity is measured by finding singularities in the face. However, as the rating prediction algorithm predicts the rating according to the degree of change in the image, the amount of change in the image is used as input data. The amount of change in the image is calculated as the difference in matrix (pixel) values between two adjacent images, as in Equation (1):

\[
\text{Equation (1)}
\]
change in the image is used as input data. The amount of change in the image is calculated as the difference in matrix (pixel) values between two adjacent images, as in Equation (1):

\[
\Delta F_{i,j,t} = F_{i,j,t} - F_{i,j,t-1}
\]

(1)

An example is shown in Figure 4.

3.3. CNN-Based Rating Prediction Model

In deep learning, the deeper the neural network (NN), the better the performance; however, it is difficult to train. In particular, a gradient vanishing problem and a gradient exploding problem cause a problem in which the training error increases as the network deepens. ResNet (residual neural network) is a CNN-based deep artificial neural network that makes training easy, even in deep neural networks using a residual learning framework, through the use of residuals [55].

In this study, the image in the video is analyzed with a deep convolutional neural network (DCNN); the larger and more complex the image, the deeper the neural network. Therefore, ResNet, which shows excellent performance even in deep neural networks, is used.

The face data \((F_{i,j,1}, F_{i,j,2}, \cdots, F_{i,j,T})\) for user \(i\) in Table 4, which is the amount of face change over time, correspond to the rating \(r_{i,j}\) in Table 2. ResNet and neural network models that predict ratings using face changes over time and rating data as inputs are trained as depicted in Figure 5. For the time \(t\) to be predicted, the average of the prediction results from \(time_1\) to \(time_t\) through ResNet and the neural network is defined as the prediction rating for \(time = t\), and is expressed as Equation (2):

\[
\text{Predicted rating } (time_t) = \frac{\sum_{i=1}^{t} \text{predicted rating } (time_i) \text{ through ResNet}}{t}
\]

(2)
3.4. Keypoint Score-Based Recommendation Model

Through the SIFT algorithm, keypoint of $F_{i,j,t-1}$ and $F_{i,j,t}$ are extracted, and keypoint matching is performed based on the extracted keypoint. Here, the number of matched keypoint is defined as keypoint score (KPS). Figure 6 illustrates the KPS calculation process from $F_{i,j,t-1}$ and $F_{i,j,t}$. Keypoint scores on each time for item $j$ are shown in Table 5.

**Figure 5.** Changes in image $t$ and image $t - 1$.

**Figure 6.** KPS score calculation from image changes.

| $\text{KPS of Time}_1$ for Item $j$ | $\text{KPS of Time}_2$ for Item $j$ | $\cdots$ | $\text{KPS of Time}_T$ for Item $j$ |
|-------------------------------------|-------------------------------------|---------|-------------------------------------|
| $u_{user_1}$ $K_{1,j,1}$ $K_{1,j,2}$ $\cdots$ $K_{1,j,t}$ |
| $u_{user_2}$ $K_{2,j,1}$ $K_{2,j,2}$ $\cdots$ $K_{2,j,t}$ |
| $u_{user_{\text{other}}_1}$ $K_{\text{other}_1,j,1}$ $K_{\text{other}_1,j,2}$ $\cdots$ $K_{\text{other}_1,j,t}$ |
| $u_{user_{\text{other}}_2}$ $K_{\text{other}_2,j,1}$ $K_{\text{other}_2,j,2}$ $\cdots$ $K_{\text{other}_2,j,t}$ |
This study searches for similar users by analyzing the similarity of KPS, and recommends items to users in real time based on similar users. Cosine similarity and Pearson similarity are usually used to obtain similarities between the vectors of two users \([5,11,56]\). This study uses cosine similarity. The cosine similarity between user \(m\) and user \(n\) is calculated in Equation (3):

\[
KPS \text{ similarity } (n,m,j) = \frac{K_{n,j} \cdot K_{m,j}}{\|K_{n,j}\| \cdot \|K_{m,j}\|}
\]

The similarity between users from time \(t_1\) to time \(t_7\) for item \(i\) is calculated using the cosine similarity of KPS data. Figure 7 illustrates the searching process of similar users, where user 3 is determined as similar to user 2. Therefore, in this study, the top N advertisements preferred by users with the highest similarity were recommended for the target user. If user 2 is a target user, then the advertisements preferred by user 3 (advertisement with max \(r_{3,j}\)) are recommended to them.

**Table 5.** Keypoint score on each time for item \(j\).

| user\(i\)    | KPS of Time\(1\) for Item\(j\) | KPS of Time\(2\) for Item\(j\) | \ldots | KPS of Time\(T\) for Item\(j\) |
|--------------|---------------------------------|---------------------------------|--------|---------------------------------|
| user\(1\)    | \(K_{1,i,1}\)                   | \(K_{1,i,2}\)                   | \ldots | \(K_{1,i,T}\)                   |
| user\(2\)    | \(K_{2,i,1}\)                   | \(K_{2,i,2}\)                   | \ldots | \(K_{2,i,T}\)                   |
| \ldots       | \ldots                          | \ldots                          | \ldots | \ldots                          |
| user\(i-1\)  | \(K_{i-1,i,1}\)                | \(K_{i-1,i,2}\)                | \ldots | \(K_{i-1,i,T}\)                |
| user\(i\)    | \(K_{i,i,1}\)                   | \(K_{i,i,2}\)                   | \ldots | \(K_{i,i,T}\)                   |

3.5. **Benchmark Model**

As benchmark models to compare the CNN-based model suggested in this study, \(1\) CF-based, \(2\) average rating (best-selling), and \(3\) random recommender system are used. As the objective of this study is to recommend advertising videos by comparing data between users, a typical CF is used, in which similarity is measured based on the ratings given by two users. A random recommender system and an average rating-based system (which uses the average of all other users’ ratings, excluding a target user’s rating for the recommendation) are used as benchmark systems. A random recommender system is used to show that the proposed recommendation approach is more profitable. If there is no difference between the accuracy of the random recommendation and that of the proposed recommendation, then the latter is not acceptable, whatever the measure of accuracy. An average rating-based system is used to show that the proposed approach outperforms a simple best-selling approach. Note that the average rating approach is essentially a form of best-selling recommendation algorithm adapted for the online video recommendation problem. In particular, a notable point is that the proposed approach can
recommend advertising videos without distinguishing between new and existing users; this is because the proposed method does not use a user’s past purchase history or rating records, but instead recommends advertising videos using only facial changes. However, as CF can only recommend advertising videos to users with existing data, it is only used to compare the performance of recommendations for existing users.

4. Experiments

4.1. Datasets

To evaluate the performance of the methodology proposed in this study, we collected user facial expression data from 1 May 2020 to 31 December 2020. We prepared 11 advertisement videos to collect the facial expressions of users watching the video. The advertising video used in the experiment was a food advertisement, and the average length of one advertising video was 20.1 s (minimum 11 s and maximum 30 s). The users watched each advertisement we prepared and gave ratings on a five-point Likert scale (1 = strongly dislike, 2 = dislike, 3 = neutral, 4 = like, and 5 = strongly like). While watching the advertisements, the users’ facial expressions were collected in real-time, and we performed face recognition and face image extraction through a deep learning approach. A total of 77 users participated in the experiment of this study. The demographic information of the users who participated in the experiment of this study is shown in Table 6.

Table 6. The demographic information of users.

| Characteristic | Subjects | Frequency | %  |
|---------------|----------|-----------|----|
| Sex           | Male     | 39        | 50.65 |
|               | Female   | 38        | 49.35 |
| Age           | <20      | 1         | 1.30 |
|               | 20–21    | 8         | 10.39 |
|               | 22–23    | 22        | 28.57 |
|               | 24–25    | 20        | 25.98 |
|               | 26–27    | 15        | 19.48 |
|               | 28–29    | 7         | 9.09 |
|               | >29      | 4         | 5.19 |
| Total         |          | 77        | 100.0 |

4.2. Experiment Design

For the evaluation of the CNN-based advertisement recommender system, two experiments were performed. In the first experiment, the accuracy of the CNN-based rating prediction model was measured. The extracted user’s face image was represented as a matrix in 3D, and the difference between the current image matrix and the previous image matrix was obtained; this is defined as a face change image. A CNN-based rating prediction model was trained on the obtained face change image over time to construct a rating prediction model, as explained in Section 3.3. As the dataset of this study is small, we use the LOOCV (leave-one-out cross validation) method, which showed good performance for learning fewer data. LOOCV is a method for evaluating the performance of a model using one sample as test data and the other \( n - 1 \) as training data from \( n \) data samples [57]. Therefore, there are \( n \) validations for one epoch, and the average of the validation accuracy is set as the accuracy of the model.

To measure the accuracy of the CNN-based rating prediction model, this study used mean absolute error (MAE). MAE has been used for the measurement and comparison
of average performance errors of a model, and has the advantage of being robust to outliers [6, 7, 57, 58]. The MAE is calculated as in Equation (4):

$$MAE = \frac{1}{N} \sum_{i} |y_i - \hat{y}_i|$$

(4)

In the second experiment, the accuracy of the advertisement recommendation model for new users was measured. The users’ keypoint score (KPS) by time was calculated through the proposed SIFT algorithm-based KPS calculation method using the extracted user’s face image. Similar users to a target user were searched by obtaining a cosine similarity of each user’s KPS for a specific time, as shown in Equation (3). The advertisement recommender system recommends Top K advertisements for the target user, which are selected based on similar users’ preferred advertisements. To measure the accuracy of the advertisement recommender system, this study used a recommendation hit ratio (RHR), defined as shown in Equation (5).

$$RHR@K = \frac{n(\text{User’s Top-K} \cap \text{Recommended Top-K})}{K}$$

where a user’s Top-K is defined as the top Top-K advertisements in which the advertisements preferred by the target user are arranged in descending order.

4.3. Experiment Result 1: CNN-Based Rating Prediction Model

To evaluate the performance of the CNN-based rating prediction model for new users, an average rating prediction system and a random rating prediction system were constructed as a benchmark system. As CF is a recommendation method based on historical data, it cannot be used as a benchmark system that evaluates new users in real time. When evaluating existing users, it is used as a benchmark system.

For the difference between the actual rating and the predicted rating by time, the evaluation index was used as MAE, and the average MAE over the entire time was considered as the performance of the model. The averages of the MAE for the CNN-based rating prediction system, the random rating system, and the average rating-based system were 0.756, 1.584, and 0.935, respectively. The CNN-based system was better than other systems in terms of preference rating predictions regarding advertising videos for the new users. More details are presented in Figure 8.

![Figure 8. Comparison of rating prediction models for new users.](image)

To compare the performances of existing users, a CF-based rating prediction system was added and analyzed. The analysis results are shown in Figure 9. The averages of the MAE values for the CNN-based rating prediction system, the random recommender system, the average rating-based system, and the CF-based system were 0.756, 1.584,
0.935, and 0.946, respectively. The CNN-based system was better than other systems in rating the prediction of advertising videos for existing users. However, in the case of video3, video6, and video9, the performance of the CF-based system was better than that of the CNN-based system. Furthermore, the CF-based system shows a more stable recommendation performance, regardless of the video type, than the suggested CNN-based system. Whereas the CF-based system used data evaluated and stored by existing users, the CNN-based system without existing user data predicted the degree of video preference by facial expressions only; thus, it was considered relatively unstable.

Figure 9. Comparison of rating prediction models for existing users.

For performance comparison, experiments were performed assuming there were existing data. However, as our proposed CNN-based system estimates a user’s rating value by detecting their face changes in real time, it is impossible to include the CF-based system in our analysis.

4.4. Experiment Result 2: Advertisement Recommendation Model

As with the rating prediction model, the performance evaluation for a real-time KPS-based advertisement recommender system is divided into two categories: situations for (1) new users and (2) existing users. For new users, best-selling and random recommender systems were used as benchmark systems, and a CF-based recommender system was added for existing users. As for the performance evaluation index, the \( RHR@K \) of Equation (5), as defined above, was used. Several experiments were performed regarding the various numbers of recommended (Top-K) videos, varying from 1 to 11. The performance from Top-1 to Top-11 was compared; the average result of \( RHR@K \) among all the users is shown in Figure 10.

As shown in Figure 10, it can be seen that the performance of the KPS-based recommender system is clearly higher than that of the best-selling or random recommender systems. Because it is a case of stopping an advertising video currently being viewed on a smartphone or kiosk and showing another advertising video, it is reasonable to see only the performance evaluation when actually recommending one advertising video. In this case, it can be seen that the performance of the KPS-based recommender system is overwhelmingly higher than the benchmark systems. Moreover, the KPS-based system showed improved robustness over the varying recommendation list sizes. Accordingly, we find that facial expressions are critical factors in recommendations to new users, and we claim that the KPS-based system addresses the new user problem. To compare the performance of existing users, a performance analysis was conducted by adding a CF-based recommender system. The analysis results are shown in Figure 11.
As a result of our performance analysis, the performance of the CF-based system was better than that of the KPS-based system when the recommendation list size was three or more. However, it is reasonable to view only the performance evaluation of Top-1, as the main problem in this study is the case of stopping the currently viewed advertising video on a smartphone or kiosk and showing another advertising video. In this case, it can be seen that our proposed system performs much better. The averages of the RHR@1 values for the KPS-based system, the random recommender system, the CF-based system, and the best-selling system were 0.348, 0.078, 0.195, and 0.221, respectively. The KPS-based system outperformed other systems in recommending videos for existing users.

5. Discussion and Conclusions

As computer vision and information technology advance, an environment has been established to collect various types of data such as facial expressions and purchase history. However, most online video recommender systems have built dynamic user profiles by capturing facial expressions through heuristic techniques. Such an approach is challenging when it comes to recommending advertisement videos in real time to new users without past video historical data. In this study, we propose a novel recommender system using computer vision and a deep learning approach to improve the limitations of existing video recommender systems. Thus, we developed a CNN-based prediction system to predict how a user enjoys an advertising video by recognizing their facial expressions. We also

Figure 10. The performance evaluation of recommender systems for the new users.

Figure 11. The performance evaluation of recommender systems for existing users.
applied SIFT to recommend advertisement videos to new users to search for users with similar preferences in real time.

The experiment results are as follows. First, the proposed CNN-based rating prediction system outperforms other systems in the preference rating predictions of advertising videos for new users and users with video records. The proposed CNN-based recommender systems predict user preferences based on real-time user facial expressions. Although the recommendation performance for new users varies depending on the video type, the overall predictive performance is excellent. However, the proposed method cannot offer a stable recommended performance considering the overall historical data. That is, when considering the overall historical data, the CF technique can be used more efficiently. Accordingly, we find that facial expressions are critical factors in recommendations to new users. Second, the proposed KPS-based advertisement recommender system indicates excellent performance when recommending large-size recommendation lists to both new and existing users. That is, when the recommended list size is small, the CF-based system shows a more stable recommendation performance. Therefore, we can determine that it is more suitable to use data such as click history than real-time user facial expression data when providing a small size recommendation list. Nevertheless, we found it more effective to recommend advertisements based on real-time user facial expressions when several advertisements were recommended.

The academic and practical implications of this study are as follows. First, we proposed a methodology for recommending online advertising videos through a deep learning approach. In previous studies, when measuring user facial expressions, these were collected through heuristic techniques, while we elaborately extracted user facial expressions by applying SIFT deep learning techniques. Similar users are calculated based on such facial expression characteristics, and customized advertisements are recommended in real time. We have contributed to the expansion of research areas related to recommender systems through this study. Second, existing recommender system research mainly uses memory-based CF techniques. However, the memory-based CF technique is a lazy learning technique that produces a result through a heuristic technique whenever a recommendation is required without building a model. As the types of online videos have become diverse, and the number of users has increased, the data size has increased, and the existing algorithms have consumed a lot of time and resources. In online advertising videos, in particular, it is necessary to quickly calculate and find recommendation items using real-time information and finally make personalized recommendations for each user. This study expands research on recommender systems by building a CNN-based model that recommends online advertising videos to users in real time after constructing a model using actual data. Third, the recommender system manager must accurately establish the user type and select a recommendation method suitable for a company’s strategy. Through experiments, we proved that the method proposed in this study is perfect for new users. However, our results showed that the CF-based method was more stable for users with historical data. Therefore, a manager should prepare a strategy for a recommendation method according to the user type and provide a personalized advertisement service. Fourth, when the proposed model provides more recommendation lists, managers need to prepare various advertisement videos because the recommendation performance is excellent. The proposed model can precisely grasp a user’s preference while capturing their facial expressions in real time. According to the experimental results of this study, a manager needs to set the recommendation list as extensively as possible to grasp users’ preferences.

However, there are several limitations and future research topics apparent in this study. The limitations and future research topics are listed as follows. First, the experiments were performed with a relatively small dataset, so the learning was insufficient. Owing to the nature of the image, the data should be significant, and learning should be repeated many times. However, this study did not obtain many data during data collection. Therefore, this represents an excellent future research topic regarding the continuous collection of data to
reveal the relationship between data and the according change in performance. Second, it takes a long time to process images, so the learning time was extensive. Improving learning time efficiency by optimizing the learning structure in the future also represents a suitable research topic. For example, SIFT used in this study as a feature detection and image matching algorithm can be compared with SURF, BRISK, and so on. Additionally, data augmentation optimization for efficient learning with few data will be a necessary research topic. Finally, in this study, facial expressions were used to evaluate users’ satisfaction with advertising videos; however, the same technology could be used for many other domains or purposes. For example, it can predict whether a customer searching for a particular product in a store is likely to purchase that product. Applying the idea of predicting customer purchase intentions in physical stores and metaverse platform situations will also be interesting for future research.

Author Contributions: Conceptualization, G.K. and J.K.; methodology, G.K. and J.K.; data curation, G.K.; writing—original draft preparation, G.K.; writing—review and editing, I.C., Q.L. and J.K.; visualization, I.C.; supervision, J.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the BK21 FOUR Program (5199990913932) funded by the Ministry of Education (MOE, Korea) and the National Research Foundation of Korea (NRF), and the Industrial Strategic Technology Development Program (20009050) funded by the Ministry of Trade, Industry and Energy (MOTIE, Korea).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

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

References
1. Krishnan, S.S.; Sitaraman, R.K. Understanding the effectiveness of video ads: A measurement study. In Proceedings of the 2013 Conference on Internet Measurement Conference, Barcelona, Spain, 23–25 October 2013; pp. 149–162.
2. Choi, I.Y.; Oh, M.G.; Kim, J.K.; Ryu, Y.U. Collaborative filtering with facial expressions for online video recommendation. Int. J. Inf. Manag. 2016, 36, 397–402. [CrossRef]
3. Ekman, P.; Oster, H. Facial expressions of emotion. Annu. Rev. Psychol. 1979, 30, 527–554. [CrossRef] [PubMed]
4. Somerville, L.H.; Fani, N.; McClure-Tone, E.B. Behavioral and neural representation of emotional facial expressions across the lifespan. Dev. Neuropsychol. 2011, 36, 408–428. [CrossRef] [PubMed]
5. Kim, H.K.; Ryu, Y.U.; Cho, Y.; Kim, J.K. Customer-driven content recommendation over a network of customers. IEEE Trans. Syst. Man Cybern. Part A Syst. Hum. 2011, 41, 42–56. [CrossRef]
6. Park, D.H.; Kim, H.K.; Choi, I.Y.; Kim, J.K. A literature review and classification of recommender systems research. Expert Syst. Appl. 2012, 39, 10059–10072. [CrossRef]
7. Kim, J.; Choi, I.; Li, Q. Customer satisfaction of recommender system: Examining accuracy and diversity in several types of recommendation approaches. Sustainability 2021, 13, 6165. [CrossRef]
8. Li, Q.; Li, X.; Lee, B.; Kim, J. A hybrid CNN-based review helpfulness filtering model for improving e-commerce recommendation Service. Appl. Sci. 2021, 11, 8613. [CrossRef]
9. Lu, J.; Wu, D.; Mao, M.; Wang, W.; Zhang, G. Recommender system application developments: A survey. Decis. Support Syst. 2015, 74, 12–32. [CrossRef]
10. Bobadilla, J.; Ortega, F.; Hernando, A.; Gutiérrez, A. Recommender systems survey. Knowl. Based Syst. 2013, 46, 109–132. [CrossRef]
11. Herlocker, J.L.; Konstan, J.A.; Terveen, L.G.; Riedl, J.T. Evaluating collaborative filtering recommender systems. ACM Trans. Inf. Syst. TOIS 2004, 22, 5–53. [CrossRef]
12. Sři, M.; Oussous, A.; Aīt Lahcen, A.; Mouline, S. Recommender systems based on collaborative filtering using review texts—A survey. Information 2020, 11, 317. [CrossRef]
13. Chen, R.; Hua, Q.; Chang, Y.-S.; Wang, B.; Zhang, L.; Kong, X. A survey of collaborative filtering-based recommender systems: From traditional methods to hybrid methods based on social networks. IEEE Access 2018, 6, 64301–64320. [CrossRef]
14. Schmidhuber, J. Deep learning in neural networks: An overview. Neural Netw. 2015, 61, 85–117. [CrossRef] [PubMed]
15. Abdi, A.; Shamsuddin, S.M.; Hasan, S.; Piran, J. Deep learning-based sentiment classification of evaluative text based on Multi-feature fusion. Inf. Process. Manag. 2019, 56, 1245–1259. [CrossRef]
16. Unger, M.; Tuzhilin, A.; Livne, A. Context-aware recommendations based on deep learning frameworks. ACM Trans. Manag. Inf. Syst. TMIS 2020, 11, 1–15. [CrossRef]
17. Zhang, S.; Yao, L.; Sun, A.; Tay, Y. Deep learning based recommender system: A survey and new perspectives. *ACM Comput. Surv.* **2019**, *52*, 1–38. [CrossRef]
18. Batmaz, Z.; Yurekli, A.; Bilge, A.; Kaleli, C. A review on deep learning for recommender systems: Challenges and remedies. *Artif. Intell. Rev.* **2019**, *52*, 1–37. [CrossRef]
19. Isinkaye, F.O.; Falajimi, Y.; Ojokoh, B.A. Recommendation systems: Principles, methods and evaluation. *Egypt. Inform. J.* **2015**, *16*, 261–273. [CrossRef]
20. Haruna, K.; Ismail, M.A.; Suhendroyono, S.; Damiasih, D.; Pierewan, A.C.; Chiroma, H.; Herawan, T. Context-aware recommender system: A review of recent developmental process and future research direction. *Appl. Sci.* **2017**, *7*, 1211. [CrossRef]
21. Gantz, J.; Reinsel, D. The digital universe in 2020: Big data, bigger digital shadows, and biggest growth in the far east. *IDC Iview IDC Anal. Future 2012*, 2007, 1–16.
22. Konstan, J.A.; Miller, B.N.; Maltz, D.; Herlocker, J.L.; Gordon, L.R.; Riedl, J. Grouplens: Applying collaborative filtering to usenet news. *Commun. ACM* **1997**, *40*, 77–87. [CrossRef]
23. Konstan, J.A.; Riedl, J. Recommender systems: From algorithms to user experience. *User Modeling User-Adapt. Interact.* **2012**, *22*, 101–123. [CrossRef]
24. Melville, P.; Sindhwani, V. Recommender systems. *Encycl. Mach. Learn.* **2010**, *1*, 829–838.
25. Herlocker, J.L.; Konstan, J.A.; Borchers, A.; Riedl, J. An algorithmic framework for performing collaborative filtering. *SIGIR Forum* **2017**, *51*, 227–234. [CrossRef]
26. Goldberg, D.; Nichols, D.; Oki, B.M.; Terry, D. Using collaborative filtering to weave an information tapestry. *Commun. ACM* **1992**, *35*, 61–70. [CrossRef]
27. Mu, R. A survey of recommender systems based on deep learning. *IEEE Access* **2018**, *6*, 69009–69022. [CrossRef]
28. Tian, Y.; Zheng, R.; Liang, Z.; Li, S.; Wu, F.-X.; Li, M. A data-driven clustering recommendation method for single-cell RNA-sequencing data. *Tsinghua Sci. Technol.* **2021**, *26*, 772–789. [CrossRef]
29. Shen, L.; Liu, Q.; Chen, G.; Ji, S. Text-based price recommendation system for online rental houses. *Big Data Min. Anal.* **2020**, *3*, 143–152. [CrossRef]
30. Nitu, P.; Coelho, J.; Madiraju, P. Improvising personalized travel recommendation system with recency effects. *Big Data Min. Anal.* **2021**, *4*, 139–154. [CrossRef]
31. Li, G.; Deng, L.; Xu, Y.; Wen, C.; Wang, W.; Pei, J.; Shi, L. Temperature based restricted Boltzmann machines. *Sci. Rep.* **2016**, *6*, 1–12. [CrossRef]
32. Hu, Z.; Hu, W.; Zhang, C. Training deep belief network with sparse hidden units. In *Proceedings of the Chinese Conference on Pattern Recognition*, Changsha, China, 17–19 November 2014; pp. 11–20.
33. Zuo, Y.; Zeng, J.; Gong, M.; Jiao, L. Tag-aware recommender systems based on deep neural networks. *Neurocomputing* **2016**, *204*, 51–60. [CrossRef]
34. Unger, M.; Bar, A.; Shapira, B.; Rokach, L. Towards latent context-aware recommendation systems. *Knowl. Based Syst.* **2016**, *104*, 165–178. [CrossRef]
35. Ko, Y.-J.; Maysstre, L.; Grossglauser, M. Collaborative recurrent neural networks for dynamic recommender systems. In *Proceedings of the Asian Conference on Machine Learning*, Hamilton, New Zealand, 16–19 November 2016; pp. 366–381.
36. Lei, C.; Liu, D.; Li, W.; Zha, Z.-J.; Li, H. Comparative deep learning of hybrid representations for image recommendations. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, 27–30 June 2016; pp. 2545–2553.
37. Shen, X.; Yi, B.; Zhang, Z.; Shu, J.; Liu, H. Automatic recommendation technology for learning resources with convolutional neural network. In *Proceedings of the 2016 International Symposium on Educational Technology (ISET)*, Beijing, China, 19–21 July 2016; pp. 30–34.
38. Zhang, S.; Liu, H.; He, J.; Han, S.; Du, X. Deep sequential model for anchor recommendation on live streaming platforms. *Big Data Min. Anal.* **2021**, *4*, 173–182. [CrossRef]
39. Pasupa, K.; Sunhem, W.; Loo, C.K. A hybrid approach to building face shape classifier for hairstyle recommender system. *Expert Syst. Appl.* **2019**, *120*, 14–32. [CrossRef]
40. Wu, C.-C.; Zeng, Y.-C.; Shih, M.-J. Enhancing retailer marketing with an facial recognition integrated recommender system. In *Proceedings of the 2015 IEEE International Conference on Consumer Electronics-Taiwan*, Taipei, Taiwan, 6–8 June 2015; pp. 25–26.
41. De Pessiemier, T.; Verlee, D.; Martens, L. Enhancing recommender systems for TV by face recognition. In *Proceedings of the 12th International Conference on Web Information Systems and Technologies*, Rome, Italy, 23–25 April 2016; pp. 243–250.
42. Visnu Dharsini, S.; Balaji, B.; Kirubha Hari, K. Music recommendation system based on facial emotion recognition. *J. Comput. Theor. Nanosci.* **2020**, *17*, 1662–1665. [CrossRef]
43. Karami, E.; Prasad, S.; Shehata, M. Image matching using SIFT, SURF, BRIEF and ORB: Performance comparison for distorted images. *arXiv 2017*, arXiv:1710.02726.
44. Bay, H.; Ess, A.; Tuytelaars, T.; Van Gool, L. Speeded-up robust features (SURF). *Comput. Vis. Image Underst.* **2008**, *110*, 346–359. [CrossRef]
45. Bay, H.; Tuytelaars, T.; Van Gool, L. Surf: Speeded up robust features. In *Proceedings of the European Conference on Computer Vision*, Graz, Austria, 7–13 May 2006; pp. 404–417.
46. Leutenegger, S.; Chli, M.; Siegwart, R.Y. BRISK: Binary robust invariant scalable keypoints. In Proceedings of the 2011 International Conference on Computer Vision, Barcelona, Spain, 6–13 November 2011; pp. 2548–2555.
47. Tareen, S.A.K.; Saleem, Z. A comparative analysis of sift, surf, kaze, akaze, orb, and brisk. In Proceedings of the 2018 International Conference on Computing, Mathematics and Engineering Technologies (iCoMET), Sukkur, Pakistan, 3–4 March 2018; pp. 1–10.
48. Lowe, D.G. Object recognition from local scale-invariant features. In Proceedings of the Seventh IEEE International Conference on Computer Vision, Kerkyra, Greece, 20–27 September 1999; pp. 1150–1157.
49. Lowe, D.G. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* **2004**, *60*, 91–110. [CrossRef]
50. Lindeberg, T. Scale Invariant Feature Transform. *Scholarpedia* **2012**, *7*, 10491. [CrossRef]
51. Albawi, S.; Mohammed, T.A.; Al-Zawi, S. Understanding of a convolutional neural network. In Proceedings of the 2017 International Conference on Engineering and Technology (ICET), Antalya, Turkey, 21–23 August 2017; pp. 1–6.
52. O’Shea, K.; Nash, R. An introduction to convolutional neural networks. *arXiv* 2015, arXiv:1511.08458.
53. LeCun, Y.; Boser, B.; Denker, J.S.; Howard, R.E.; Hubbard, W.; Jackel, L.D. Backpropagation applied to handwritten zip code recognition. *Neural Comput.* **1989**, *1*, 541–551. [CrossRef]
54. LeCun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **1998**, *86*, 2278–2324. [CrossRef]
55. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
56. Kim, H.K.; Kim, J.K.; Ryu, Y.U. Personalized recommendation over a customer network for ubiquitous shopping. *IEEE Trans. Serv. Comput.* **2009**, *2*, 140–151. [CrossRef]
57. Herlocker, J.L.; Konstan, J.A.; Riedl, J. Explaining collaborative filtering recommendations. In Proceedings of the 2000 ACM Conference on Computer Supported Cooperative Work, Philadelphia, PA, USA, 2–6 December 2000; pp. 241–250.
58. Liang, N.; Zheng, H.-T.; Chen, J.-Y.; Sangaiah, A.K.; Zhao, C.-Z. Tsrdl: Tag-aware recommender system based on deep learning–intelligent computing systems. *Appl. Sci.* **2018**, *8*, 799. [CrossRef]