

Research Article

Multimedia Recommendation System for Video Game Based on High-Level Visual Semantic Features

Fasiha Ikram ¹ and Humera Farooq ²

¹Department of Computer Science, Bahira University Karachi Campus, Karachi, Sindh, Pakistan

²Department of Computer Science, Bahira University Karachi Campus, Karachi, Sindh, Pakistan

Correspondence should be addressed to Fasiha Ikram; fasia.ikram@gmail.com

Received 16 November 2021; Revised 30 December 2021; Accepted 6 January 2022; Published 3 February 2022

Academic Editor: Waqas Nawaz

Copyright © 2022 Fasiha Ikram and Humera Farooq. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The increase of multimedia content in e-commerce and entertainment services creates a new research gap in the field of recommendation systems. The main emphasis of the presented work is on increasing the accuracy of multimedia recommendations using visual semantic content. Recent approaches have shown that the inclusion of visual information is helpful to understand the semantic features for a recommendation model. The researchers have contributed to the field of multimedia item recommendations using low-level visual semantic features. Here, we seek to extend this contribution by exploring the high-level visual semantic content using constant visual attributes for video game recommendation systems. With the exponential growth of multimedia content in the video game industry in the last decade, researchers investigate the importance of personalized video game recommendation techniques. Previous methods have not investigated the importance of visual semantic content for video game recommendations. A practical recommendation system for video games is challenging due to the data diversity, level of user interest, and semantic complexity of features involved. This study proposed a novel method named Deep Visual Semantic Multimedia Recommendation Systems (D_VSMR) to deal with high-level visual features for multimedia recommendation systems. A visual semantic-based video game recommendation system utilizing deep learning methods for visual content learning and user profile learning is introduced. The proposed approach employs content-based techniques to expand users' profiles. The user profile expansion is based on the visual content of games. The required datasets have been obtained from video game e-commerce platforms like Google Play Store and Amazon for evaluation purposes. The evaluation results have shown that the proposed approach's accuracy and effectiveness have been improved up to 95.87% compared to the other state-of-the-art methods.

1. Introduction

Recommendation systems are software applications that aim to support users in decision-making while interacting with large information spaces. Recommendation systems help overcome the information overload problem by exposing users to the most interesting items and offering novelty, surprise, and relevance [1]. Recommendation systems collect information about user preferences for a set of items (e.g., movies, songs, books, news, applications, websites, travel destinations, and e-learning materials) [2]. This information is then used to make personalized predictions on items or products, such as which video one may want to watch on

Netflix or which books to buy from Amazon [3]. Additionally, users can mark each item on these systems to express their preferences; these expressions are called ratings. The concept of recommendation systems was presented to decrease the challenges of sizeable digital trace data, inspect and retrieve the large data sets, and optimize relevant systems [4]. Recommendations can be about any product, for example, books, videos, music, TV programs, research resources, websites, and video games [5]. Traditionally, these systems use two recommendation techniques, that is, content-based filtering, in which recommendations are based on the content items targeted by the users and relevant features of the item [5], and collaborative filtering [6], which is based

on observed behaviors of the users while interacting with the systems and filters out those with similar behavior. In recent years, different features combination has been proposed to design these types of systems.

Two types of attributes are used for user profile expansion, that is, explicit attributes and implicit attributes. Explicit attributes help expand user profiles using demographic information such as gender, occupation, and city. In contrast, implicit information is obtained from multimedia features, that is, audio, text, and visual data. In the last two decades, recommendation systems have started getting feedback from users regarding ratings on their preferences for online purchasing, books, movies, and videos [7]. Recent research has investigated the importance of social, context-aware [8], and multimedia content recommendation [9]. In a social recommendation, the user's connectivity is collected and used for recommendations [10]. In the context-aware recommendation, the researchers investigate and evaluate the relation of contextual data and factors' impact on recommendation processes. The multimedia recommendations are content-based techniques that exploit the audio, textual, and visual features of the product and enhance the user profile, thus increasing the recommendation system's performance.

Recommendation systems leveraging multimedia datasets are rich in features that can help to expand and reveal user preferences. Multimedia feature-based recommendation systems play an important role in the decision-making process to cope with the information overload problem. Recently researchers have contributed to the field of multimedia recommendations using low-level and high-level visual semantic features. In the last decade, researchers investigated the importance of visual attributes in various fields. Most existing research on semantic high-level visual features is based on fashion, movies, food, and news. These features have been extracted using different techniques, that is, extracting low-level features and making a descriptor for movie recommendation systems [11]. Using high-level features, a CNN architecture recommends fashion items for the e-commerce domain [12]. In some of the applications of recommendation systems, high-level features are not constant; for example, in movies, actors have different guises in different characters.

The present work explores the importance of high-level visual content using constant attributes for multimedia recommendation systems for video games. Video game is a billion-dollar industry that aims to provide a better experience to users with the help of recommendation systems. However, video game recommendation systems suffer from a lack of innovation [13]. It is critical to study the video game genre compared to other fields due to its versatile and open nature [14]. Furthermore, the robust and varied attributes of the video game will be used in this research to investigate the relation of visual content in-game recommendation systems. To the best of our knowledge, very few research works have been conducted on the video game recommendation domain. Although many other recommendation techniques have been extrapolated for video game recommendation, such as rating and metadata, visual content is not used for

game recommendation. Video game recommendation systems play an important role in helping game users cope with information overload. Content-based video game recommendation systems use explicit structure (meta-data and description genre) information for product retrieval. It is necessary to require rich set side information for features about videos. Such information is human-generated and prone to error. In the last few years, researchers paid attention to visual and textual multimedia recommendation systems.

This work presents a multimedia recommendation system for video game recommendations based on the visual semantic features of games to extend user profiles. For that purpose, we obtain data from platforms like Play Store and Amazon to extract visual content, test validation, and evaluate. In addition, the user's demographical profile information is also used, such as rating, id, and product features that have been obtained from explicit and implicit sources such as tags, descriptions, and visual content. The following contributions are reported in the presented study:

- (a) A user preference profile is developed based on the high-level features of each product. These features are retrieved by semantic visual features. The retrieved features are then integrated with the user preference.
- (b) A recommendation model is proposed, namely, deep visual semantic-based multimedia recommendation systems (D-VSMR). A deep recommendation model is used to learn the user and product features. D-VSMR uses visual features to model user preferences and product attributes.
- (c) We have evaluated the recommendations based on novel visual features of video games data. Experiments are performed on the Google Play Store and recently released Amazon dataset [15].

The rest of the paper is organized as follows: in Section 2, related work based on the game recommendation techniques and visual semantic recommendation has been discussed. Section 3 will discuss the proposed methodology. In Sections 4 and 5, the experimental results and evaluation methods are detailed. The discussion is presented in Section 6.

2. Related Work

2.1. Game Recommendation Techniques. In recent decades, the video game market has flourished considerably, hindering decision-making while interacting with huge information. The existing studies can be categorized as game product recommendation [16], game development process recommendation [17], and video game recommendation to users [18]. The features used for game recommendation systems are varied; the existing work shows that the main features used for recommendations are metadata, profile expansion, textual and visual features, and ratings. A few state-of-the-art methods used for video game recommendation are shown in Table 1. Recently some researchers have paid importance to in-game product recommendations to

TABLE 1: State-of-the-art methods used for video game recommendation.

Study	Technique	Features				
		Metadata	Profile expansion	Textual	Visual	Rating
[18]	CB	✓	✗	✗	✗	✓
[19]	Regression	✓	✓	✗	✗	✓
[16]	Factorization algorithm	✗	✗	✗	✗	✓
[21]	Fm and Deepfm	✓	✓	✓	✗	✓
[22]	Hybrid	✓	✓	✗	✗	✓

users according to the need of the game. In [18], a game recommendation system has proposed a solution for cold-start problem using content-based behavioural telemetry features. Due to the lack of available datasets, they used a survey dataset for solving the cold-start problem. In addition, they have used only metadata and ratings for their study. In [19], they integrated the game-level activities and designed a neural network for the recommendation and collected data from 24 users. This approach was based on the player in-game profiling and taxonomy of game activities. However, they have not considered multimedia features for profile expansion. In another study, the Tensor Factorization algorithm was used to expand user profile and proposed a solution for in-game content recommendation using latent factor models [16]. Particularly, their focus was on role playing games only for in-game recommendations for users. Containerized Amazon recommendation system (CARS) [20] have been proposed based on textual features such as reviews for data validation and experiments that have used video game datasets. In [21], the author made a comparative study and tested deep learning recommendation architectures for game recommendation (FM, DeepNN, and Deepfm) and textual reviews of users. Their work concluded that sentiments of user reviews are not helpful for video game recommendation and deep learning architectures are efficient for video game recommendations. In [22], a hybrid video game recommendation system has been proposed based on a stem dataset, considering the hours of play to obtain implicit ratings for a particular user.

The above studies have done significant work in video game recommendations. The limitations exist in using multimedia features for video game recommendation systems. Despite this, the reported studies work on content-based recommendations [18] or collaborative recommendations [23, 24]. A systematic study was performed for a recommendation system for video game developers using the developer process and recommended a new project using previous game development experience [17]. In line with the survey reported in 2018, [25] described the importance of using multimedia features for multimedia recommendations. Few studies also work hybrid filtering-based algorithms for recommendations of video games.

2.2. Visual Semantic Recommendation. Generally, recommendation systems exploit visual content features by availing semantic interpretation [1]. Furthermore, multimedia content processing has the same steps, that is, segmentation, feature extractions, item representation, and

semantic orientations. These visually explainable recommendation systems mainly deal with machine learning techniques using high-level and low-level semantic visual features.

The initial work in visually explainable recommendation systems proposed using convolution neural network (CNN) and semantic regions [21–24]. Additionally, in [26], the author proposed a multimedia recommendation system, attentive collaborative filtering (ACF), for low-level semantic attributes using component and item-level hierarchical attention. Visual Part-based Object Representation (VPOR) [27], a personalized user preference recommendation system, was proposed based on user interest calculations, using low-level visual semantic features. In [28], the author investigates the importance of visual features for food recipe recommendation systems using low-level visual semantic features and metadata. The results have shown that recipe images are the key features to increase users’ attention and explore the feasibility of alternative meals with similar healthier dishes. Using their textual features, a dynamic RNN was proposed to capture user preference from the semantic video information [29]. Fashion recommendations are based on disjoint cloths parts and aesthetics representation, where the models learn from visual content to make personalized recommendations. Hierarchical Fashion Graph Network (HFGN) [30] attempted to aggregate the low-level visual information with user and item features to estimate the user preference using attention mechanisms based on outfits (fashion recommendation). Similar to (HFGN), [12] proposed a model (DVBPR) established on visual and personal preference-based recommendation, Siamese CNN, and conditional GANs used for visual semantic features extractions. Additionally, DVBPR was the extension of VBPR [9], where researchers investigate the importance of visual features on users’ opinions to make a personalized recommendation for multimedia products. The Semantic Attribute Explainable Recommender System (SAERS) [31] proposed a fashion recommendation based on user semantic attributes. The semantic user attributes have been extracted using a fine-grained interpretable semantic space based on high-level visual semantic features. In [32], the author proposed a dependent generative adversarial model to solve the pairing problem in fashion recommendations based on visual similarity. However, conventional visual features (low-level) are not enough to extract aesthetic information from images. Low-level visual features use MGPEC features and color histograms to portray user preference about a product. Researches also investigated the importance of aesthetic features in recommendations systems to model

user preferences [33, 34]. Aesthetic features contain more valuable information for modern recommendation systems having rich visual content such as video games, food recipes, and movie recommendation systems.

Tables 1 and 2 show existing studies in recommendation systems that provide a promising result using user preferences and visual features. Additionally, literature studies show that visual features for video games have not been considered for recommendation. However, the user's preference data is competence sparse, and the lack of textual information for new items does not help generate a user preference profile. Therefore, we proposed a model to predict user preference by combining visual features with a user profile. The presented model consists of two components: user profile expansion and designing of a deep recommendation model.

3. Deep Visual Semantic Multimedia Recommendation System

In this section, we introduced the detail of the proposed model. We first present the overall framework of the proposed methodology in Subsection 3.1. Next, we detail the profile expansion to extract visual semantic features and user profile design in Subsections 3.2 and 3.3, respectively.

3.1. Basic Framework D-VSMR. Considering the impact of multimedia features on user preference, we propose a recommendation model; namely, D-VSMR (deep learning-based visual semantic multimedia recommendation systems) is presented. The elements of the multimedia dataset are denoted as $M = [I_1, I_2, I_3, \dots, I_n]$, whereas the elements of items (I) consist of visual feature (V), user profile preference (UR), and product details (P), that is, $I \in [V \times UR \times P]$. We consider implicit product visual features, that is, V, and explicit user and product features, that is, UR and P of each item. The D-VSMR model consists of two main components, that is, user profile expansion and designing of deep recommendation model, as shown in Figure 1. For the user profile expansion component, first, we acquire a multimedia dataset and then retrieve the visual semantic features. The retrieved visual semantic features are integrated and extended to user profile preferences. The extended user profile features, including category, title, and description, are fed into the second component. In the second component, a deep recommendation model is created to learn the user preferences (user and product features). Overall, D-VSMR used visual semantic product attributes to learn the potential preference of users to predict the rating given by the users and estimate the ranking of the products using user preferences. The steps of D-VSMR are explained in Algorithm 1. Table 3 shows the notation used for developing the presented recommendation model. Figure 2 shows the detailed architecture of D-VSMR. In the following subsections, we will discuss the details of the profile expansion and deep recommendation model components.

3.2. User Profile Expansion. This component is designed to expand user profile preferences using visual semantic features. Traditionally recommendation algorithms utilize explicit details of product and user, that is, rating, occupation, country, and tags, to increase the recommendation accuracy. This work combines product features (category, genres, description, and title) and user preference (rating, user id, occupation, and gender) with visual semantic features. The profile expansion component is designed to extract visual semantic features, which is then used to expand user profile preference. In order to extract the visual semantic features, RCNN has been used [35]. RCNN is a region-based convolutional neural network. RCNN uses a selective search method and extracts 2000 regions from each image [36]. These regions are referred to as a region proposal. After extraction of proposed regions, rescale and transformation on each region are performed. Each region proposal is converted into a fixed image size. A pretrain convolutional neural network model AlexNet is applied for feature extraction [37]. In the next step, an SVM classifier is used to predict the object class. The task of the SVM classifier is to extract 2000 regions proposal using a selective search on a test image. Scoring for each extracted feature for each class is performed using the SVM. A greedy nonmaximum suppression is used to reject any region if it overlaps with higher scoring. Specifically, after the nonmaximum suppression, we add a new layer and remove all the features less than or equal to the defined threshold value, that is, $\text{score} \leq 0.5$. Finally, RCNN extracts implicit visual features, where the feature set is denoted by F and R is each set of regions. Hence, $F \in R^{f \times s}$, where f denotes the extracted proposed region in the image and s denotes the prediction probability of the feature, where $VSF = [F_1, F_2, F_3, \dots, F_n]$ and $F \in [[f_1 \times s, f_2 \times s, f_3 \times s, \dots, f_n \times s]]$.

User profile expansion builds a strong association between user profiles, textual product attributes, and each item's semantic visual attributes. While combining the product features, a dictionary of the bag of words is established using the term-frequency-inverse document (TF-ID) technique (eq1) and porter stemming algorithm. The TF-ID helps to identify the frequency of each word in the corpus and later on helps to identify the stop words (is, this, and products), stemming (games and game), and noise words (controller and price). The feature set has four (4) features users_id, rating, product id, implicit product features, and explicit product features.

$$tf(t, d) = \frac{f_{t,d}}{\sum_{t' \in d} f_{t',d}} \quad (1)$$

Based on the feature set mentioned in equation (1), we implement the deep recommendation system using the semantic feature vector to predict items to the targeted user where Tf is the term frequency and d reflects the dictionary.

3.3. Deep Recommendation. This component is designed to estimate the probability of user likeness using users and

TABLE 2: State-of-the-art methods used for visual semantic recommendation.

Methods	Approaches	Low level	High level
DVBPR [12]	CNN	✗	✓
VBPR [9]	BPR (Bayesian personalized algorithm)	✗	✓
SAERS [31]	PairWise learning	✗	✓
HybridFM [40]	Ann	✓	✗
ADCFA [33]	PairWise learning	✓	✗
TranSearch [39]	Searching algorithm	✗	✓

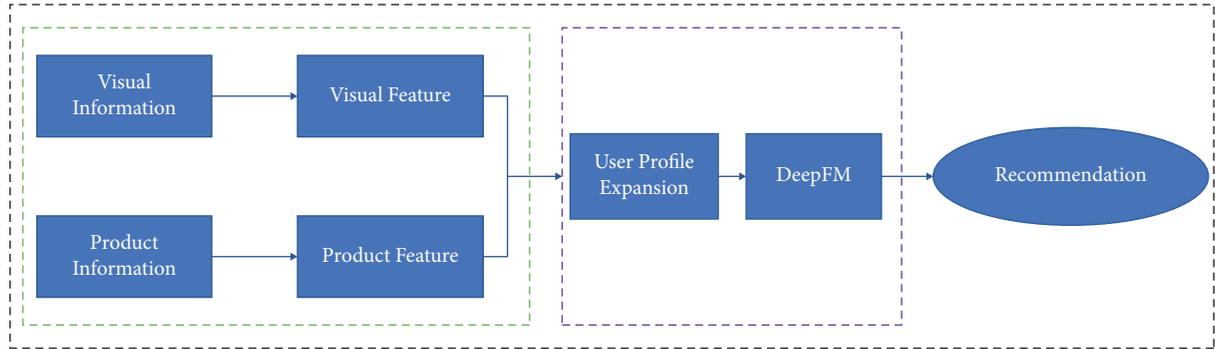


FIGURE 1: Component of proposed methodology D-VSMR.

TABLE 3: Notations used in this paper.

Notations	Description
VSMR	Visual semantic multimedia recommendation system
VSF	Visual semantic features
U	User set
I	Multimedia items
D	Document set
$F_{t,d}$	Frequency of each feature
$\sum_{t \in d}$	Total number of features
Y _{fm}	The output of the FM layer
Y _{DNN}	The output of the DNN layer
E	Embedding layers
A	Activation function
w_0	Global bias for FM
v_i, v_j	Product features
UPE	User profile expansion
R	Product rating
F	Feature extracted proposal region
S	Prediction probability/score
Tf	Term frequency

product features interaction. The goal of this component is to increase the recommendation accuracy. For this purpose, a deep neural network along with a factorization machine is implemented that learns user and product features interactions using the given feature vector or data. The given recorded data contains features consisting of categorical, highly sparse, and dense features. Therefore, an embedding layer transforms the input features to a low dimension feature vector input using FM. This layer helps to reduce the overwhelming of the training network. We used FM for elementwise multiplication to obtain a joint feature vector. The user preference is obtained from the visual preference of the product for the targeted product,

where (w, x) represents the user features and the inner product (v_i, v_j) represents the product features. A deep neural network component estimates the prediction and likeness score for item v from user u . Additionally, to reduce the need for an extensive pretraining process for a recommendation, DNN along with FM is used for learning [38]. These steps are shown in equations (2) to (6).

$$f(0) = [e1, e2, \dots, em], \quad (2)$$

$$a^{(l+1)} = \sigma(W^{(l)} a^{(l)} + b^{(l)}), \quad (3)$$

$$Y_{FM} = w_0 + \sum_{i=1}^d w_i x_i + \sum_{i=1}^d \sum_{j=i+1}^d (v_i, v_j) x_i x_j, \quad (4)$$

$$Y_{DNN} = \sigma(W^{|H|+1} \cdot a^H + b^{|H|+1}) \quad (5)$$

$$y_{Df m} = \sigma(Y_{FM} + Y_{DNN}). \quad (6)$$

The semantic information is data driven since the rules to extract visual semantic features are learned from the data. Compared with the traditional visual-aware recommendation system, our D-VSMR is more comprehensive and objective. Another advantage of the reported work is that it can be applied to recommend visually aware video games to reduce the lengthy process.

As mentioned in Algorithm 1, the proposed method is a combination of RCNN and Deepfm. However, the RCNN performs a selective search about 2000 times on each image. Therefore, it is the most time-consuming step in the proposed model. The time complexity of RCNN and Deepfm is all $O(N^2 + N)$, where N donates the number of epochs.

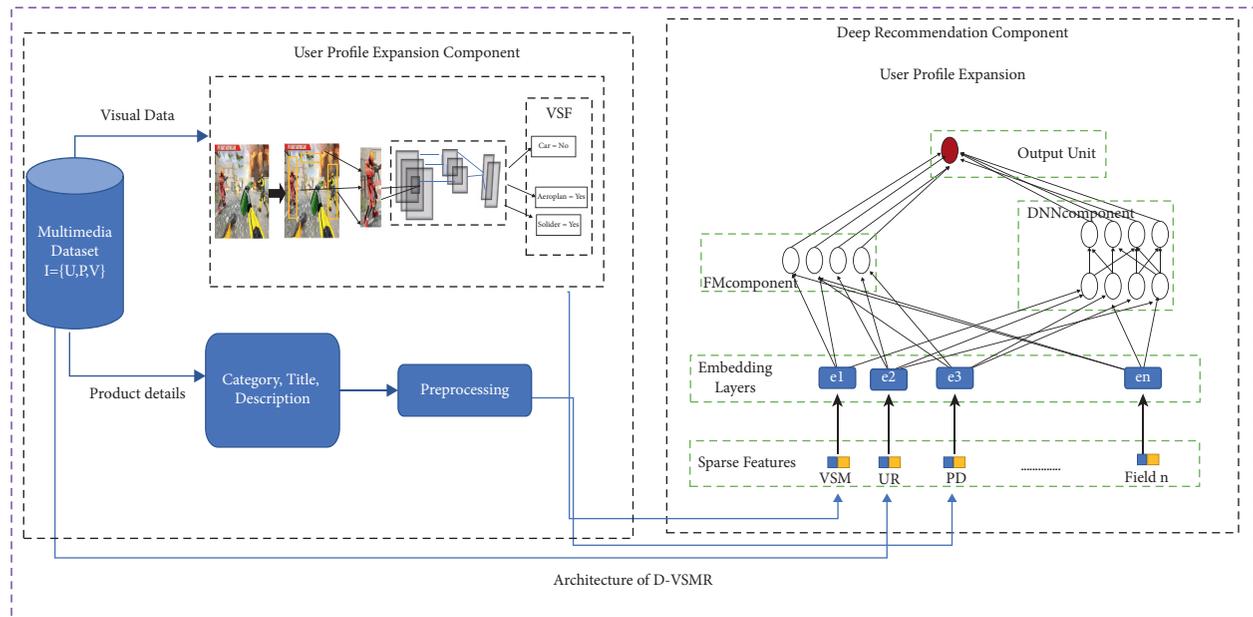


FIGURE 2: The detailed architecture of D-VSMR (<https://mizostudio.com/>).

Input: the user set U , the multimedia item set I , the item description document C , the one-hot coding user rating features, and product images URLs.

Epoch, the iterations for each training

- (1) Visual semantic feature learning (V (VSF) = visual semantic feature learning I (V , UR , and P))
- (2) Using the trained model, extract Visual Features of each image of items using the RCNN model
- (3) User profile expansion (user profile expansion I (VSF, UR , and P))
Extract the textual attributes of the product using description genre tags and product details
 - (i) Perform NLP tasks to remove noisy words
 - (ii) Stemming
 - (iii) TF-IDF: to remove high frequency and stop words
- (4) The given recorded data contain raw features, which contain categorical, highly sparse, and dense I (VSF, UR , and UPE)
- (5) Recommendation using Deepfm

Output: the recommendation for the targeted user

ALGORITHM 1: Deep-VSMR.

4. Evaluation Methodology

This section discusses the experiments, evaluation, and the existing datasets, their limitations, and the need for a new proposed dataset. In Subsection 4.1, we explain the available multimedia datasets, their limitations, and data collection for the experiments. To evaluate the performance of the D-VSMR, we perform our experiments on two datasets. In Subsection 4.2, the evaluation metrics will be discussed, whereas in Subsection 4.3, the experimental design and the comparison of the D-VSMR with the baseline methods are presented.

4.1. Datasets. Multimedia recommendation systems need large datasets for learning models and recommendations. To the best of our knowledge, very few multimedia datasets are available that can be used for multimedia recommendation

systems, for example, Netflix, MovieLens, and Amazon datasets. MovieLens and Netflix datasets contain movies and visual content that do not contain uniform semantic objects; for example, actors do not have the same guise and role in a different context. In addition, the Amazon dataset contains visual and textual content with minimal visual description. Table 4 shows the comprehensive statistics of the three datasets. These datasets lack uniform or semantic attributes. Netflix and MovieLens datasets are from different domains and contain different features, which are not suitable for verifying and validating the D-VSMR. However, the Amazon dataset contains very few visual features, that is, one image per item; such small data is not recommended for visual features learning. Therefore, we created a multimedia dataset (vision game) for learning visual semantic features and user preference for the proposed model; we used Amazon and our dataset (vision game) for testing and validation.

TABLE 4: Multimedia recommendation systems datasets.

Multimedia recommendation systems datasets					
Dataset	Multimedia	Users	Items	Ratings	Visual details (ground truth)
Amazon	✓	✓	✓	✓	✓
Netflix	✓	✓	✓	✓	✗
MovieLens	✓	✓	✓	✓	✗
Steam	✓	✓	✓	✓	✗

4.2. Data Collection and Preparation. A dataset is created using Google Scrapper (a library for Play Store data crawling). This library provides the information of apps such as users, rating, and visual content of apps. For data scrapping, games data is selected considering that it provides rich semantic visual features for model learning and attributes for data recommendation. Furthermore, games have been divided into genres, and each genre has semantic information associated with the genre. Table 5 shows the details of each genre’s data, including visual features. Therefore, semantic information of genres objects classification and retrieval is used for visual features collections. Additionally, metadata is also provided by the Play Store in which users, ratings, tags, and descriptions are included. The scrapper provides around 41 features like overall score, time, country, installs, versions, editors, and developer information. Features such as installed versions of applications that are not useful for prediction and recommendations have been removed.

4.3. Evaluation Metrics. In the literature of recommendation systems, the item-based recommendation is performed using item ranking accuracy models. The evaluation metric for ranking prediction is based on the performance of models, and we evaluate the performance using evaluation metrics AUC and NDCG according to equations (7) and (8). The AUC metrics measure the performance recommendation systems, whereas NDCG metrics rank the positive items in the list. Therefore, these metrics are used to measure the performance of the D-VSMR and compare results with baselines.

$$\text{AUC} = \frac{1}{|U|} \sum_u \frac{1}{|E(u)|} \sum_{(i,j) \in E(u)} \delta(x'_{u,i} > x'_{u,j}), \quad (7)$$

$$\text{NDCG} = \frac{1}{Z} \sum_{k=1}^N \frac{2^{rel_k} - 1}{\log_2(k+1)}. \quad (8)$$

For the sake of simplicity, all the reported results will be presented in a range of 0–100%. Some of the parameters remained constant throughout the experiments, that is, 30 epochs for training, dropout rate, 0.1, and batch size to 256.

4.4. Experimental Design and Baselines. We compared the performance of the proposed methodology with recommendations based on product implicit (semantic visual features) and explicit features (product details and metadata), verified the accuracy of the proposed model by comparing it with other recommendation models using two

datasets, that is, the existing datasets of Amazon, and created a dataset for the game on Google Play Store.

Three different evaluation strategies were used to verify the effectiveness of the D-VSMRS from different aspects:

- (i) The impact of the D-VSMR over other visual recommendation systems: to evaluate the performance of the D-VSMR in the recommendation, we compare it with the following state of the visual art features-based methods: DVBPR [12], VBPR [9], TranSearch mode (toys and game) [39], SAERS [31], ADCFA [33], and HybridFM [40].
- (ii) Performance comparison with state-of-the-art game recommendation methods: to evaluate the performance of the D-VSRM model, we compare the state-of-the-art game recommendation techniques to ensure the reliability of our results.
- (iii) The impact of the user profile expansion using semantic features: analyze the effectiveness of user profile expansion over traditional techniques to verify the importance and influence of visual features.
- (iv) The impact of Deepfm for visual semantic recommendation: to analyze the results of user profile expansion and the impact of the deep factorization machine on visual semantic attributes of user profile expansion.

5. Experimental Results

In this section, the experimental results of D-VSMR are discussed. Additionally, we also compare our results with the state-of-the-art game recommendation algorithms. We also compare the results of visual semantic features with traditional recommendation methods such as tags and product details.

Table 6 and Figure 3 show the results of the features during recommendation. These results have been produced using semantic visual features (VSF), tags (T), and product details (PD) on two different datasets, vision game and Amazon. The reported results have shown that the proposed approach provides better results in terms of accuracy and performance. We conclude that the recommendation and training performance of the D-VSMR is better with VSF on both datasets. As mentioned before, visual features of items contain plenty of user and item implicit and explicit preference information. This information helps the model to understand the relationship between users and items better. Encouragingly, the results show that our model showed better results than traditional metadata-based attributes and

TABLE 5: Visual features categories for semantic attributes.

Genre-class	Genre	Semantic attributes
Roleplaying	Adventure, first-person, third-person, and role playing	Unlock states and third-person games scene
Strategy	Turn-based, strategy, puzzle, and platformer	Blocks, words, robe, and rings
Sports	Sports, racing, and wrestling	Car, bat, ball tennis, and so on
Action	Action	Gun, soldiers, fire, and jeep

TABLE 6: Results Of D-VSMR where T, PD, and VSF denote tags, product details, and visual semantic features, respectively.

	T	Type PD	VSF	Accuracy	NDCG
Google	✓	✗	✗	81.34	0.0235
	✓	✓	✗	83.15	0.0209
	✓	✓	✓	95.84	0.0340
Amazon	✓	✗	✗	58.67	0.0342
	✓	✓	✗	66.25	0.0675
	✓	✓	✓	72.08	0.1384

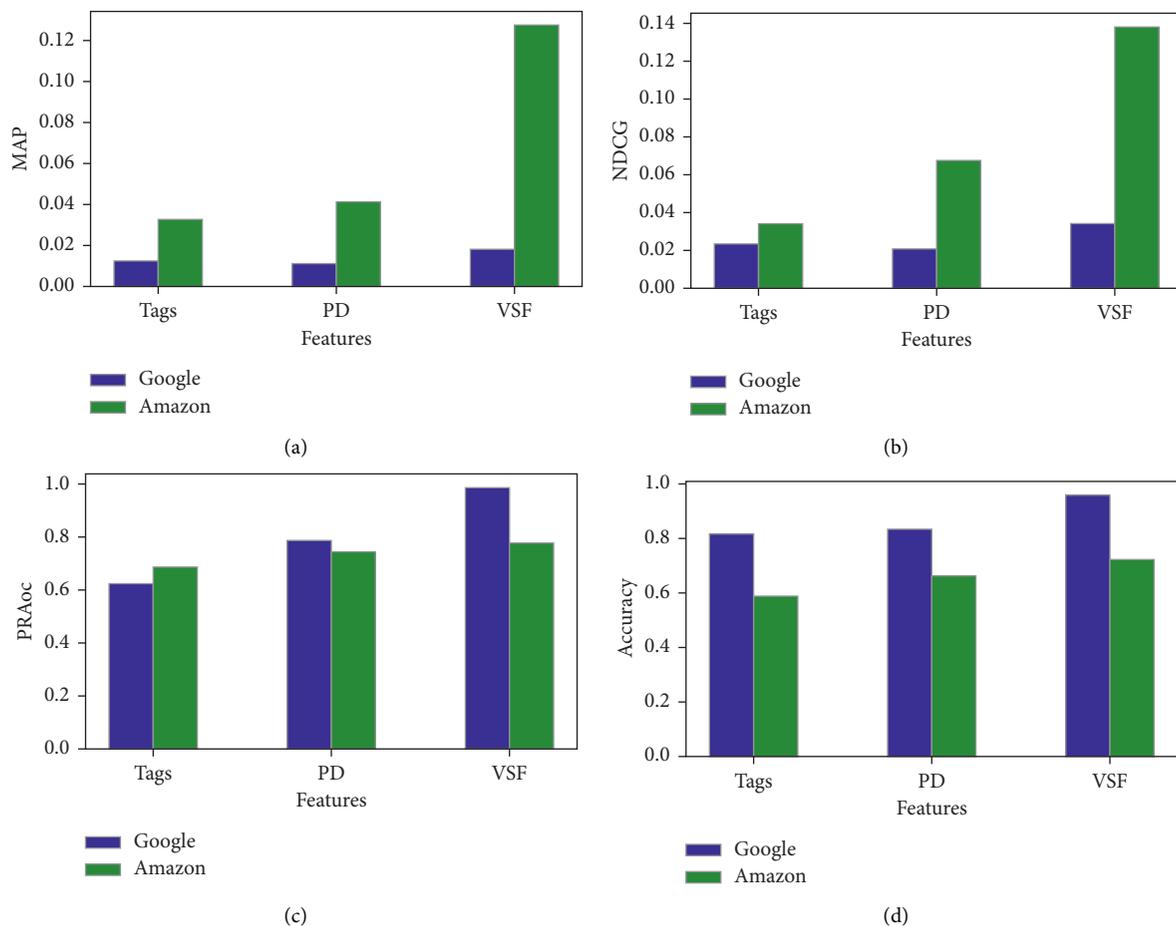


FIGURE 3: Performance compared with the product features on the Amazon and Google Play datasets. PD and VSM denote product features and visual semantic features, respectively.

product explicit information. The proposed approach provides the best results for multimedia game recommendation and outperforms other techniques.

Table 7 and Figure 4 show the comparison of the D-VSMR with other state-of-the-art techniques. Overall,

D-VSMR consistently and significantly outperforms all the baselines across two datasets in AUC and NDCG. The performance of vision game is better from a huge margin than the others, mainly because of the more numbers of visual features in vision game compared to Amazon.

TABLE 7: State-of-the-art visual features-based multimedia recommendation systems.

Methods	Evaluation methods	Results (%)
DVBPR [12]	AUC	79.64
VBPR [9]	AUC	78.34
SAERS [31]	AUC	80.17
HybridFM [40]	AUC	71.8
Proposed model (D-VSMR)	AUC	95.87
TranSearch mode (toys and game) [39]	NDCG	0.120
ADCFA [33]	NDCG	0.0200
Proposed model (D-VSMR)	NDCG	0.1384

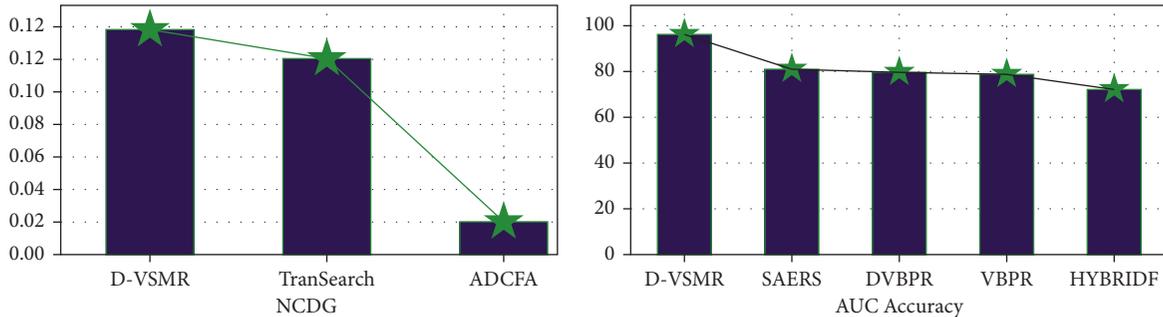


FIGURE 4: Baseline multimedia recommendation systems using visual features.

Another reason is that the vision game is purely based on games dataset; however, the Amazon dataset has some other products too. DVBPR is an extension of VBPR [9], tailored for a visually aware recommendation for fashion items. DVBPR achieved 79% accuracy using end-to-end learning and CNN model, and VBPR achieved 78% accuracy using learning of user preference from implicit feedback. At the same time, the D-VSMR outperformed over DVBPR and VBPR and achieved 95% accuracy, which shows that our model improved 16% compared to these existing multimedia recommendation models. TranSearch is a multimodal personalized product search method that relies on textual queries and gives preference to visual modalities. TranSearch shows overall 83% improved NDCG over clothing products and 7% improved results for office items, whereas, for toys and games, it shows 3% improved results. In contrast, the D-VSMR improved the NDCG to 0.034 and outperformed over Tran search.

The results show that VSF plays an important role in multimedia recommendation systems and outperforms other methods. The profile expansion using visual semantic features surpassed the other features for traditional tags and product details. This demonstrates the importance of profile expansion using visual features, which play an important role in multimedia product recommendations. Table 7 shows that the profile expansion using VSF shows 12% improved results for the vision game and 5% improved results for Amazon datasets. However, profile expansion using only tags shows the worst performance.

Due to a lack of studies, we could only compare our model with three state-of-the-art video game recommendation systems. The recommendation performance of STEAMer [41] and Implicit feedback reviews [21] performed

better than the D-VSMR model. This may prove that textual reviews contain more information regarding users' behaviors and preferences. However, both studies have used a steam dataset with many users, that is, 100 million and 70,912, respectively. Therefore, these systems outperformed our model. In the case of a large number of users, we believe that our system will also provide better recommendations. In another recent study, the authors utilized hybrid filtering that included users' account properties, reviews, achievements, and played hours [22]. D-VSMR model consistently achieves the best accuracy scores across the three baseline techniques, that is, tags, product attributes, and visual features. In contrast to the models that use user reviews and perform sentiment analysis on massive datasets, D-VSMR utilizes the visual features to decrease data sparsity, which leads to better recommendation results. Additionally, in the studies that utilized steam datasets and considered users' account properties achievements and played hours, D-VSMR utilized only implicit and explicit product features to better understand the user preferences.

5.1. Computational Complexity. Figure 5 shows the comparison among the time complexity of the profile expansion with tags, product details, and VSF. In the case of tags, the time needed for the training is less than the time needed for visual semantic features. This is attributed to the calculations and operations performed for visual features. However, such time delay is not observed for the achieved accuracy. The improvement in recommendation and user satisfaction is more important. The machine used for experiments has 6 GB-GPU, core i7, and Windows 10 installed.

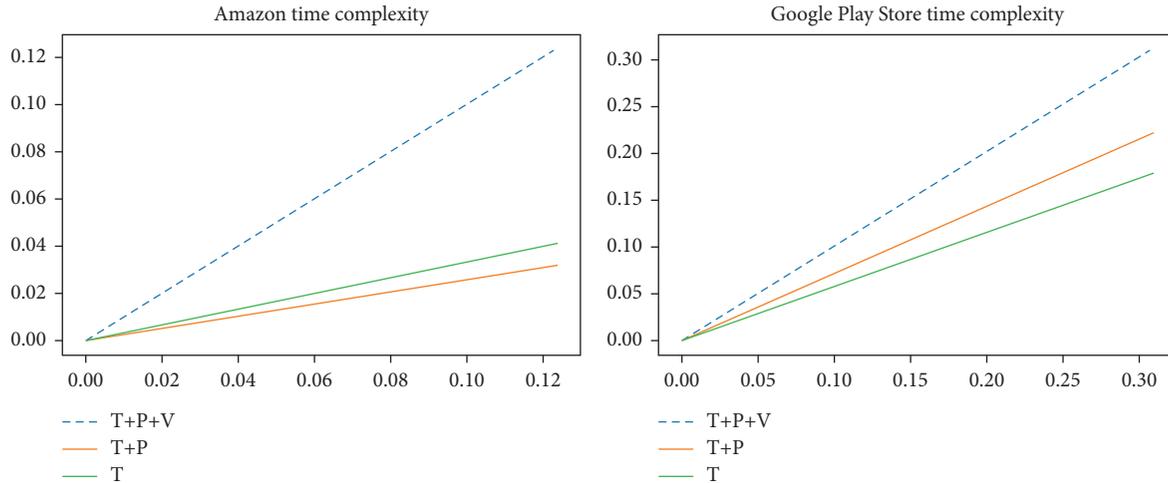


FIGURE 5: Time complexity of D-VSMR, where t tags, P is products details, and V is visual semantic features.

6. Discussion

The D-VSMR contributes to two research areas of modern recommendation systems, that is, improvement in multimedia recommendation systems and the role of VSF in the game recommendation systems. This section will discuss the importance of reported results according to the evaluation metrics for multimedia recommendations.

The basic idea of extracting semantic attributes using the deep learning model is to reduce the error of human-based tagging; therefore, deep learning models have been trained on semantic attributes. The training model extracts and provides semantic attributes without being human-biased. The personalized user profile, especially feeding all the recommendations' attributes, performs better than a nonpersonalized recommendation. Interestingly, the proposed model shows extraordinary results in terms of accuracy of recommendation. The presented work is one of its types that shows better performance in terms of items recommendation for video game systems. Additionally, the reported results show that the performance of the proposed model highly depends upon the visual features; that is, having better visual features results in better recommendation performance. In general, low-level visual features have been used for uniform objects such as food [28], games, and movies (very few) [11]. This research used high-level semantic visual features, and D-VSMR significantly outperformed over automatically extracted low-level visual features [11]. It is worth mentioning that our model outperformed the tag-based and user-based recommendation techniques, that is, MF and SDA. This is because we considered the interaction of multimedia and user features. This work is based on dual neural networks, one trained for feature extraction and the second for the recommendation of items. Remarkably, recommendation using visual features shows better results than traditional content-based recommendation techniques, that is, tags-based and manually extracted visual features [42–44].

We believe that such architecture will help form a richer recommendation, where content and users are more closely linked. To the best of our knowledge, this study is the first

and comprehensive attempt to consider the product's visual preferences and the details of the product to expand user profile. Deep factorization machine along with the visual features of product outperforms in most of the cases. The conducted experiments verify the effectiveness of the model. In the last decade, researchers paid attention to video games recommendation systems using textual attributes. However, the high-level visual semantic attributes exceed the baselines and show better results with a large margin. This indicates that visual preference contains more information. Integrating the high-level visual semantic features of products can boost the recommendation performance more. Furthermore, the robust and varied features of the video game used in this research to investigate the relation of visual content in-game recommendation systems have opened a new door for the video game recommendation industry.

7. Conclusion

The video game recommendation systems were presented in this research. This work is based on content-based recommendations along with visual semantic features. In addition, a deep neural network was presented to obtain better recommendations. The visual features have been obtained using an RCNN in which the model proposed regions of interest. These regions have been used to elevate the recommendation process and retrieve interesting features. The proposed system focused on the visual semantic features of video games that contain rich implicit information. A deep neural network with a factorization machine is used to improve the recommendation performance. The reported results show that the presented work outperforms traditional techniques and shows remarkable performance on a multimedia dataset with the visual modality. Finally, this work has investigated the importance of visual features. It indicates that incorporating the high-level semantic attributes of products can boost the recommendation performance for those systems where visual user preference contains more information. As a possible future line of research, it is possible to make a joint multimedia recommendation system based on visual and textual modalities.

Data Availability

The data that support the findings of this study (Amazon and Google Play Store) are openly available at <https://nijianmo.github.io/amazon/index.html>, reference number [15], and <https://play.google.com/store/search?q=games~amp;c=apps>. Google Play Store data were generated using `google_play_scraper` (<https://pypi.org/project/google-play-scraper/>). The raw data that support the findings are available on Google Drive at https://drive.google.com/drive/folders/1G_zaTX7Be8TgG0SYEpNc0ZGx0ytqV52N?usp=sharing, following a 6-month embargo from the date of publication to allow for the commercialization of research findings. Derived data supporting the findings of this study are available from the corresponding author (Fasiha Ikram) upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] Y. Deldjoo, M. Schedl, P. Cremonesi, and G. Pasi, "Recommender systems leveraging multimedia content," *ACM Computing Surveys*, vol. 53, no. 5, pp. 1–38, 2020.
- [2] F. Alyari and N. Jafari Navimipour, "Recommender systems," *Kybernetes*, vol. 47, no. 5, pp. 985–1017, 2018.
- [3] F. Ricci, L. Rokach, and B. Shapira, "Recommender systems: introduction and challenges," in *Recommender Systems Handbook*, pp. 1–34, Springer, Berlin, Germany, 2015.
- [4] P. Bedi and R. Sharma, "Trust based recommender system using ant colony for trust computation," *Expert Systems with Applications*, vol. 39, no. 1, pp. 1183–1190, 2012.
- [5] J. Gemmell, M. Ramezani, T. Schimoler, L. Christiansen, and B. Mobasher, "A fast effective multi-channelled tag recommender," *ECML PKDD Discovery Challenge*, vol. 497, pp. 59–70, 2009.
- [6] T. Rambharose and A. Nikov, "Computational intelligence-based personalization of interactive web systems," *WSEAS Transactions on Information Science and Applications*, vol. 7, no. 4, 2011.
- [7] M. C. Kim and C. Chen, "A scientometric review of emerging trends and new developments in recommendation systems," *Scientometrics*, vol. 104, no. 1, pp. 239–263, 2015.
- [8] G. Adomavicius and A. Tuzhilin, "Context-aware recommender systems," in *Recommender systems handbook*, pp. 217–253, Springer, Berlin, Germany, 2011.
- [9] R. He and J. McAuley, "VBPR: visual bayesian personalized ranking from implicit feedback," in *Proceedings of the AAAI Conference on Artificial Intelligence*, pp. 144–150, ACM, Washington, NY, USA, February 2016.
- [10] H. Ma, D. Zhou, C. Liu, M. R. Lyu, and I. King, "Recommender systems with social regularization," in *Proceedings of the fourth ACM international conference on Web search and data mining*, pp. 287–296, ACM, California, USA, February 2011.
- [11] Y. Deldjoo, M. Elahi, P. Cremonesi, F. Garzotto, P. Piazzolla, and M. Quadrana, "Content-based video recommendation system based on stylistic visual features," *Journal on Data Semantics*, vol. 5, no. 2, pp. 99–113, 2016.
- [12] W.-C. Kang, C. Fang, Z. Wang, and J. McAuley, "Visually-aware fashion recommendation and design with generative image models," in *Proceedings of the 2017 IEEE International Conference on Data Mining (ICDM)*, Nov. 2017.
- [13] M. A. Khenissi, F. Essalmi, M. Jemni, fnm Kinshuk, S. Graf, and N.-S. Chen, "Relationship between learning styles and genres of games," *Computers & Education*, vol. 101, pp. 1–14, 2016.
- [14] D. Clearwater, "What defines video game genre? Thinking about genre study after the great divide," *Loading*, vol. 5, no. 8, 2011.
- [15] J. Ni, J. Li, and J. McAuley, "Justifying recommendations using distantly-labeled reviews and fine-grained aspects," in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, University of California, San Diego, CA, USA, January 2019.
- [16] R. Sifa, R. Yawar, R. Ramamurthy, C. Bauckhage, and K. Kersting, "Matrix- and tensor factorization for game content recommendation," *KI - Künstliche Intelligenz*, vol. 34, no. 1, pp. 57–67, 2020.
- [17] C. Politowski, L. M. Fontoura, F. Petrillo, and Y.-G. Guéhéneuc, "Learning from the past: a process recommendation system for video game projects using post-mortems experiences," *Information and Software Technology*, vol. 100, pp. 103–118, 2018.
- [18] M. Viljanen, J. Vahlo, A. Koponen, and T. Pahikkala, "Content based player and game interaction model for game recommendation in the cold start setting," *Game recommendation is an important application of recommender systems*, vol. 107, pp. 2411–2502, 2020.
- [19] L. De Simone, D. Gadia, D. Maggiorini, and L. Anna Ripamonti, "Design of a recommender system for video games based on in-game player profiling and activities," in *Proceedings of the CHIItaly 2021: 14th Biannual Conference of the Italian SIGCHI Chapter*, pp. 1–8, ACM, Bolzano, Italy, July 2021.
- [20] A. Cassell, A. Muñoz, B. Blain-Castelli et al., "CARS: a containerized amazon recommender system," in *Proceedings of the ITNG 2021 18th International Conference on Information Technology-New Generations*, Springer, Berlin, Germany, February 2021.
- [21] G. Cheuque, J. Guzmán, and D. Parra, "Recommender systems for Online video game platforms: the case of STEAM," in *Proceedings of the Companion 2019 World Wide Web Conference*, ACM, New York, NY, USA, May 2019.
- [22] J. Pérez-Marcos, L. Martín-Gómez, D. M. Jiménez-Bravo, V. F. López, and M. N. Moreno-García, "Hybrid system for video game recommendation based on implicit ratings and social networks," *Journal of Ambient Intelligence and Humanized Computing*, vol. 11, no. 11, pp. 4525–4535, 2020.
- [23] R. Bunga, F. Batista, and R. Ribeiro, "From implicit preferences to ratings: video games recommendation based on collaborative filtering," in *Proceedings of the 13th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management*, pp. 209–216, ISCTE University Institute of Lisbon, Portugal, June 2021.
- [24] S. M. Anwar, T. Shahzad, Z. Sattar, R. Khan, and M. Majid, "A game recommender system using collaborative filtering (GAMBIT)," in *Proceedings of the 2017 14th International Bhurban Conference on Applied Sciences and Technology (IBCAST)*, IEEE, Islamabad, Pakistan, Jan. 2017.
- [25] R. Mu, "A survey of recommender systems based on deep learning," *Ieee Access*, vol. 6, pp. 69009–69022, 2018.

- [26] J. Chen, H. Zhang, X. He, L. Nie, W. Liu, and T.-S. Chua, "Attentive collaborative filtering: multimedia recommendation with item-and component-level attention," in *Proceedings of the 40th International ACM SIGIR conference on Research and Development in Information Retrieval*, ACM, New York, NY, USA, August 2017.
- [27] H.-Y. Chi, C.-C. Chen, W.-H. Cheng, and M.-S. Chen, "UbiShop: commercial item recommendation using visual part-based object representation," *Multimedia Tools and Applications*, vol. 75, no. 23, pp. 16093–16115, 2016.
- [28] D. Elsweiler, C. Trattner, and M. Harvey, "Exploiting food choice biases for healthier recipe recommendation," in *Proceedings of the 40th international acm sigir conference on research and development in information retrieval*, ACM, New York, NY, USA, August 2017.
- [29] J. Gao, T. Zhang, and C. Xu, "A unified personalized video recommendation via dynamic recurrent neural networks," in *Proceedings of the 25th ACM international conference on Multimedia*, ACM, New York, NY, USA, October 2017.
- [30] X. Li, X. Wang, X. He, L. Chen, J. Xiao, and T.-S. Chua, "Hierarchical fashion graph network for personalized outfit recommendation," in *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, ACM, New York, NY, USA, July 2020.
- [31] M. Hou, L. Wu, E. Chen, Z. Li, V. W. Zheng, and Q. Liu, "Explainable fashion recommendation: a semantic attribute region guided approach," in *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence {IJCAI-19}*, AAAI Press, Macao China, August 2019.
- [32] S. Kumar and M. D. Gupta, *C+GAN: Complementary Fashion Item Recommendation*, AAAI Press, Macao, China, 2019.
- [33] A. Paul, Z. Wu, K. Liu, and S. Gong, *Robust Multi-Objective Visual Bayesian Personalized Ranking for Multimedia Recommendation*, pp. 1–12, Springer Cham, New York, NY, USA, 2021.
- [34] W. Yu, X. He, J. Pei et al., "Visually aware recommendation with aesthetic features," *The VLDB Journal*, pp. 1–19, 2021.
- [35] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, IEEE, Columbus, OH, USA, June 2014.
- [36] J. R. R. Uijlings, K. E. A. van de Sande, T. Gevers, and A. W. M. Smeulders, "Selective search for object recognition," *International Journal of Computer Vision*, vol. 104, no. 2, pp. 154–171, 2013.
- [37] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems*, vol. 25, pp. 1097–1105, 2012.
- [38] H. Guo, R. Tang, Y. Ye, Z. Li, and X. He, "DeepFM: a factorization-machine based neural network for CTR prediction," in *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence*, AAAI press, Melbourne, Australia, August 2017.
- [39] Y. Guo, Z. Cheng, L. Nie, X.-S. Xu, and M. Kankanhalli, "Multi-modal preference modeling for product search," in *Proceedings of the 26th ACM international conference on Multimedia*, ACM, New York, NY, USA, October 2018.
- [40] N. Hazrati and M. Elahi, "Addressing the New Item problem in video recommender systems by incorporation of visual features with restricted Boltzmann machines," *Expert Systems*, vol. 38, no. 3, p. e12645, 2021.
- [41] D. Wang, M. Moh, and T.-S. Moh, "Using deep learning and steam user data for better video game recommendations," in *Proceedings of the 2020 ACM Southeast Conference*, ACM, New York, NY, USA, April 2020.
- [42] T. Bogers, "Tag-based recommendation," in *Social Information Access*, vol. 10100, pp. 441–479, Springer, 2018.
- [43] M. Ge, M. Elahi, I. Fernaández-Tobías, F. Ricci, and D. Massimo, "Using tags and latent factors in a food recommender system," in *Proceedings of the 5th International Conference on Digital Health 2015*, ACM, New York, NY, USA, October 2018.
- [44] A. K. Sahu, P. Dwivedi, and V. Kant, "Tags and item features as a bridge for cross-domain recommender systems," *Procedia Computer Science*, vol. 125, pp. 624–631, 2018.