Article

Deep Learning Model with Transfer Learning to Infer Personal Preferences in Images

Jaeho Oh ¹, Mincheol Kim ² and Sang-Woo Ban ^{1,2,*}

¹ Department of Information & Communication Engineering, Graduate School, Dongguk University, Gyeongju 38066, Korea; ojh2417@dongguk.ac.kr

² Department of Electronics Information & Communication Engineering, Dongguk University, Gyeongju 38066, Korea; kimmc0316@naver.com

* Correspondence: swban@dongguk.ac.kr; Tel.: +82-54-770-2270

Received: 30 September 2020; Accepted: 27 October 2020; Published: 29 October 2020

Featured Application: Image-based personal recommendation system superior to human experts on customized interiors.

Abstract: In this paper, we propose a deep convolutional neural network model with transfer learning that reflects personal preferences from inter-domain databases of images having atypical visual characteristics. The proposed model utilized three public image databases (Fashion-MNIST, Labeled Faces in the Wild [LFW], and Indoor Scene Recognition) that include images with atypical visual characteristics in order to train and infer personal visual preferences. The effectiveness of transfer learning for incremental preference learning was verified by experiments using inter-domain visual datasets with different visual characteristics. Moreover, a gradient class activation mapping (Grad-CAM) approach was applied to the proposed model, providing explanations about personal visual preference possibilities. Experiments showed that the proposed preference-learning model using transfer learning outperformed a preference model not using transfer learning. In terms of the accuracy of preference recognition, the proposed model showed a maximum of about 7.6% improvement for the LFW database and a maximum of about 9.4% improvement for the Indoor Scene Recognition database, compared to the model that did not reflect transfer learning.

Keywords: personal preference; deep learning; atypical features; Fashion-MNIST; LFW; Indoor Scene Recognition; convolutional neural networks; Grad-CAM

1. Introduction

Humans generally shape their individual preferences for complex and diverse visual information through various visual experiences in childhood. In the course of this process, based on the learning process from experiencing visual information in very different fields, personal preferences for common visual features of visual information from different domains are made. Therefore, in order to implement a preference model for individual visual information, it is necessary to form personal preferences through learning using image data with different visual characteristics. The user's preference classification problem is actively researched in the field of recommendation systems [1]. This helps the e-commerce market to continuously grow, and its influence is gradually expanding. From the standpoint of providing specific products or services, it is essential to predict and infer a customer's preferences. Thus, most recommendation systems are in the form of servicing the results from predicting the user's preferences [2–5]. In the case of visual-based preferences, it is difficult to predict them compared to text-based preferences, since visual-preference information is very sparse, based on each domain [1], and there are many difficulties in expressing them with formal information [5]. However, most of

these studies mainly used formal information and features about objects such as color, shape, and style as a way of inferring visual preference [6–12]. The problems with these studies are that it is difficult to transfer visual-preference features between domains [5,13], and it is difficult to express visual preference information when a preference has an atypical feature [5]. Furthermore, most human visual preferences have atypical features [6].

In this work, we propose a learning and classification method for these atypical visual preferences. In order to design a model for inferring visual preferences, the deep convolutional neural network (DCNN) [14] was applied, which has shown significant achievements in object-recognition/classification applications. In order to analyze the inference possibility and characteristics of inter-domain visual preferences, transfer learning was applied to the proposed model [6,15]. In addition, three datasets for verification of the proposed preference inference model were used: Fashion-MNIST [16], Labeled Faces in the Wild (LFW) [17], and Indoor Scene Recognition [18]. Images from the datasets were reclassified according to personal preferences. Additionally, in order to infer personal preferences from datasets with different morphological characteristics [6], we applied transfer learning to the proposed preference inference model, and analyzed the data characteristics according to cross-domain preferences [6]. Thus, we verified the possibility of classifying common atypical visual features of inter-domain data with completely different morphological and categorical characteristics.

In general, humans have personal specific preferences for visual information, but there is a tendency for the reason to not be explained accurately as to whether they prefer the corresponding visual information. In this work, we tried to partially explain the personal preferences for the visual information to which the gradient class activation mapping (Grad-CAM) method [19] was applied to explain why the proposed model shapes personal preferences for visual information. A Grad-CAM can be utilized to identify salient areas of an image that are important for determining user preferences in visual images. A partial explanation of the user's preferences in visual information might be possible from visual characteristics of a salient area, which leads to determining the preference. Moreover, the Grad-CAM approach was applied to graphically express the atypical visual features preferred by a specific user [19]. Through this method, the localization and patterns of atypical visual features were analyzed according to the preference classification result of the proposed model.

The rest of this paper is organized as follows. In Section 2, the proposed model is described in detail. For verification, Section 3 shows the experimental results from the proposed model's performance. Section 4 offers some conclusions and describes further work.

2. Proposed Model

2.1. Deep Convolutional Neural Networks

A DCNN model was utilized to implement the proposed personal visual-preference model, which is composed of three different types of layers: the convolution layer, the pooling layer, and the fully connected layer [6,20]. We mainly used 3×3 and 1×1 kernels, as used in VGGNet to construct a deeper neural network for improving classification performance [6,20]. Pooling layers are used as a means to reduce the computational cost by reducing the dimensions of the feature map and the number of parameters in the network [6,20]. Fully connected layers are used for the classification of feature maps generated in the final feature-extraction layer, which were composed of two hidden layers for sufficient classification performance [6,20]. In addition, we utilized a softmax function [6,21] as the activation function of the output layer, and the adaptive momentum (Adam) as an optimizer of the objective function to improve the classification performance [6,22]. Figure 1 shows the overall architecture of the proposed DCNN-based model for preference classification.

Figure 1. Architecture of the proposed convolutional neural network model.

2.2. Transfer Learning

In order to mimic a human-like incremental learning mechanism, we considered transfer learning for incrementally training and inferring visual preferences in inter-domain images with atypical visual characteristics. In general, transfer learning can improve the classification performance by reusing a part of the pre-trained network for another domain dataset similar to the domain of the original dataset when the deep learning model does not have significant data to increase the classification performance [6,15]. Thus, we reused the feature extraction layers of the pre-trained network, connecting the pre-learned feature extraction layer with the newly defined classification layer to fine-tune all the parameters of the network [6], as shown in Figure 2.

Figure 2. Transfer learning in the proposed approach.

2.3. Grad-CAM

Grad-CAM was utilized as a method to analyze and interpret the cause of classification results in a deep learning-based model [19]. In the class activation mapping (CAM) model, the attention map can be generated only when the global average pooling is applied (instead of the fully connected layer) to the feature map passing through the convolution layers [23]. However, Grad-CAM can be applied to a fully connected structure, not only a global average pooling layer [19]. We used Grad-CAM as a method for the analysis and interpretation of the results of personal preferences. We applied the Grad-CAM approach to implementing a personal preference inference model with plausible explanatory possibilities. Figure 3 shows the structure of Grad-CAM for generating a heat map reflecting the user's preference based on atypical visual features.

Figure 3. Structure of Grad-CAM for generating a heat map reflecting the user's preference based on atypical features.

2.4. Three Image Databases

In this paper, three well-known open image databases with atypical visual characteristics were considered for training in, and inferring of, personal visual preferences: Fashion-MNIST, LFW, and Indoor Scene Recognition [16–18]. Fashion-MNIST is an image dataset of fashion products, which is made up of 60,000 training images and 10,000 test samples including 10 categories of 28×28 grayscale images [6,16]. Figure 4 shows 10 class labels from Fashion-MNIST and some images sampled in Fashion-MNIST.

Label	Description	Examples
0	T-Shirt/Top	
1	Trouser	
2	Pullover	
3	Dress	
4	Coat	
5	Sandals	
6	Shirt	
7	Sneaker	
8	Bag	
9	Ankle boots	

Figure 4. Fashion-MNIST dataset images and labels [6,16].

LFW is a public benchmark dataset for face recognition and classification [6,17]. LFW is made of 13,233 facial images of 5749 people, and consists of 150×150 color (RGB) images [6,17]. Figure 5 shows some of the FL images sampled from LFW.

Figure 5. Labeled Faces in the Wild (LFW) dataset images and labels [6,17].

The Indoor Scene Recognition dataset is a group of images of various indoor scenes. The dataset contains 67 indoor categories and a total of 15,620 images. On average, there are more than 100 images per class. Resolutions of the images in the dataset vary, and they are also RGB [18]. Figure 6 shows several categories and examples of the data.

Figure 6. Indoor Scene Recognition dataset images and labels [18].

3. Experiment Results

For verification of the inference accuracy of personal visual preferences from the proposed model, we utilized the three public benchmark datasets Fashion-MNIST, LFW, and Indoor Scene Recognition [16–18]. Before training the proposed model, all images from these datasets were classified and labeled as preferred or non-preferred, based on the preferences of a specific user. Table 1 shows the ratios of the two classes (preferred and non-preferred) for the Fashion-MINIST, LFW, and Indoor Scene Recognition databases, according to the user's preferences. In the process of determining the subject's preference for each image, a subject was asked to determine a preference for each image within 1 s in order to maximally exclude interference from other factors including the subject's intrinsic episodic memory or other non-visual factors. In addition, in order to minimize the effect of the subject's fatigue,

the entire image database was divided into several subgroups. Additionally, a subject was asked to decide preference instantly and sensibly based on visual characteristics as much as possible.

Table 1. Data classification ratios of user preferences in each dataset.

	Fashion-MNIST	LFW	Indoor Scene Recognition
Preferred	57.92% (40,545/70,000)	55.93% (5034/9000)	54.77% (8246/15,000)
Non-preferred	42.08% (29,455/70,000)	44.07% (3966/1000)	45.03% (6754/15,000)

We considered two different models. Model 1 was trained and tested using only a training dataset and a test dataset. Model 2 utilized a validation dataset as well as training and test datasets. In the experiments for measuring classification performance of the proposed preference inference model using the Fashion-MNIST database, Model 1 used 60,000 training images and 10,000 test images. However, for Model 2, training was conducted with 50,000 training images, 10,000 validation images, and 10,000 test images, as shown in Figure 7. Table 2 shows the preference classification results for Model 1 and Model 2 with the Fashion-MNIST dataset.

Figure 7. Schematics for datasets used with Model 1 and Model 2.

Table 2. Preference classification results with the Fashion-MNIST dataset.

Data Set: Fashion-MNIST	Preference Classification Results			
	Model 1		Model 2	
Training dataset	TP (31,889/60,000)	FP (2876/60,000)	TP (25,512/50,000)	FP (20,859/50,000)
	FN (602/60,000)	TN (24,633/60,000)	FN (391/50,000)	TN (3238/50,000)
	Correct Accuracy (TP + TN)/(TP + FP + FN + TN) 94.20% (56,522/60,000)		Correct Accuracy (TP + TN)/(TP + FP + FN + TN) 92.74% (46,371/50,000)	
	TP (5229/10,000)	FP (137/10,000)	TP (5347/10,000)	FP (60/10,000)
Test dataset	FN (651/10,000)	TN (4083/10,000)	FN (433/10,000)	TN (4160/10,000)
	Correct Accuracy (TP + TN)/(TP + FP + FN + TN) 93.12% (9312/10,000)		Correct Accuracy (TP + TN)/(TP + FP + FN + TN) 95.07% (9507/10,000)	
			TP (5280/10,000)	FP (53/10,000)
			FN (755/10,000)	TN (3932/10,000)
Validation data-set		Correct Accuracy (TP + TN)/(TP + FP + FN + TN) 92.12% (9212/10,000)		

TP: True Positive, FP: False Positive, FN: False Negative, TN: True Negative in Tables 2–4.

Table 3. Preference classification results from Model 1 and Model 2 with the Labeled Faces in the Wild (LFW) and Indoor Scene Recognition datasets.

	LFW Dataset Classification Results		Indoor Scene Recognition Dataset Classification Results	
	TP	FP	TP	FP
Model 1	(3657/9000)	(780/9000)	(6935/15,000)	(1105/15,000)
	FN (1377/9000)	TN (3186/9000)	FN (1311/15,000)	TN (5649/15,000)
	Correct Accuracy (TP + TN)/(TP + FP + FN + TN) 76.03% (6843/9000)		Correct Accuracy (TP + TN)/(TP + FP + FN + TN) 83.89% (12,584/15,000)	
	TP (3643/9000)	FP (699/9000)	TP (6608/15,000)	FP (1332/15,000)
Model 2	FN (1397/9000)	TN (3267/9000)	FN (1638/15,000)	TN (5422/15,000)
	Correct Accuracy (TP + TN)/(TP + FP + FN + TN) 76.78% (6910/9000)		Correct Accuracy (TP + TN)/(TP + FP + FN + TN) 80.20% (12,030/15,000)	

Table 4. Preference classification results from Model 1 and Model 2 using transfer learning with the LFW and Indoor Scene Recognition databases.

	LFW Classification Results through Transfer Learning				Indoor Scene Recognition Classification Results through Transfer Learning			
	Model 1		Model 2		Model 1		Model 2	
Training data-set	TP (3823/8000)	FP (526/8000)	TP (3811/8000)	FP (522/8000)	TP (5988/12,000)	FP (575/12,000)	TP (5912/12,000)	FP (447/12,000)
	FN (652/8000)	TN (2999/8000)	FN (664/8000)	TN (3003/8000)	FN (499/12,000)	TN (4938/12,000)	FN (375/12,000)	TN (5066/12,000)
	Correct Accuracy (TP + TN)/(TP + FP + FN + TN) 85.28% (6822/8000)		Correct Accuracy (TP + TN)/(TP + FP + FN + TN) 85.18% (6814/8000)		Correct Accuracy (TP + TN)/(TP + FP + FN + TN) 91.05% (10,926/12,000)		Correct Accuracy (TP + TN)/(TP + FP + FN + TN) 91.48% (10,978/12,000)	
Test data-set	TP (465/1000)	FP (70/1000)	TP (451/1000)	FP (64/1000)	TP (1599/3000)	FP (219/3000)	TP (1585/3000)	FP (138/3000)
	FN (94/1000)	TN (371/1000)	FN (108/1000)	TN (377/1000)	FN (160/3000)	TN (1022/3000)	FN (174/3000)	TN (1103/3000)
	Correct Accuracy (TP + TN)/(TP + FP + FN + TN) 83.60% (836/1000)		Correct Accuracy (TP + TN)/(TP + FP + FN + TN) 82.80% (828/1000)		Correct Accuracy (TP + TN)/(TP + FP + FN + TN) 87.36% (2621/3000)		Correct Accuracy (TP + TN)/(TP + FP + FN + TN) 89.60% (2688/3000)	

From the experimental results, we were able to conclude that both Model 1 and Model 2 performed well in classifying preferences with the Fashion-MNIST dataset. The classification performance of Model 2 was 95.07%, which was superior to Model 1 for the test dataset. For the purpose of evaluating the performance of the proposed personal preference model, we applied the proposed model (pre-trained with the Fashion-MNIST database) for preference classification with the LFW database and the Indoor Scene Recognition database. In this experiment, Model 1 and Model 2 were applied to classify the LFW data and Indoor Scene Recognition data into only two classes (preferred and non-preferred) without any additional training process for fine-tuning. The preference classification experiment using the LFW dataset was conducted by randomly selecting 9000 images from 13,233 images in LFW. In addition, in the experiments using the Indoor Scene Recognition dataset [18], 15,000 images were randomly selected from 15,620 images. Table 3 shows the experimental results of preference classification with the LFW dataset and the Indoor Scene Recognition dataset by Model 1 and Model 2.

According to the experimental results in Table 3, the proposed model showed plausible preference classification performance for two databases with different characteristics from the prior-learning database.

Additionally, in order to verify the effectiveness of transfer learning between inter-domain datasets with different visual characteristics, transfer learning was applied to the learning process of the proposed preference inference model. Table 4 shows the experimental results of preference classification with the LFW and Indoor Scene Recognition datasets by the proposed model considering transfer learning. As shown in Table 4, the proposed preference classification model reflecting transfer learning showed better performance than the model that did not reflect transfer learning, as shown in Table 3.

From these experimental results, we can conclude that the proposed model was able to properly classify preferences for datasets with different visual characteristics through transfer learning.

This transfer learning process might be recognized as playing an important role in constructing the personality of each person, reflecting visual preferences by incrementally determining the individually preferred characteristics for visual information. Thus, we were able to conclude that, through numerous experiences with visual information that is tremendously diverse in characteristics, each human shapes their own preferences for visual data.

Finally, we applied a Grad-CAM model for the purpose of explaining the reasons for the proposed model's preference classifications [19]. Grad-CAM generated an attention heat map to provide explanations for the classification results of the specific user's preferences [19]. Figure 8 shows the visual attention map for the results of preferred and non-preferred classifications from the LFW dataset. Figure 8a shows some heat map results obtained by Grad-CAM for the correctly classified preferred images in the LFW dataset. Figure 8b shows some of the heat map results obtained by Grad-CAM for the correctly classified non-preferred images in the LFW dataset. According to the subject's opinions, non-facial visual features including a suit or tie had some influence on the preference of the face image when deciding the preference for each face image of LFW. In addition, it was judged that the racial issue or the presence of acquaintances was also partly affected, even though a subject was asked to decide preference momentarily and sensibly based on the visual face as much as possible. However, some of these influences might be acceptable because we aimed to develop a model that learns and infers individual preferences for general visual features.

Figure 8. Visual expressions from applying Grad-CAM to user preference-classification results with LFW: (a) sample heat maps for the results in preferred data, and (b) sample heat maps for the results in non-preferred data.

Figure 9 shows a visual attention heat map for the results of preferred and non-preferred classifications from the Indoor Scene Recognition dataset. Figure 9a shows some heat map results obtained by Grad-CAM for the correctly classified preferred images from the Indoor Scene Recognition dataset. Figure 9b shows the heat map results obtained by Grad-CAM for the correctly classified non-preferred images from the Indoor Scene Recognition dataset. As shown in the experimental results, we can see that the heat map provided by Grad-CAM is generally set in a meaningful feature area in the image where human preferences can be found. Therefore, we can conclude that it might be possible to explain a user's preferences using heat maps generated by Grad-CAM. In order to confirm the explanatory feasibility of the subject's preference using the results of Grad-CAM, we analyzed how well the area that influenced the subject's preference decision matched the attention area of Grad-CAM. As shown in Table 5, analysis was performed on the test datasets of LFW and Indoor Scene Recognition. In the case of the LFW dataset, it was analyzed that the Grad-CAM results properly reflected the subject's preference for 312 images out of 465 images correctly classified as preferred images among the 1000 test images. On the other hand, it was analyzed that the Grad-CAM results for 238 images out of 371 images correctly classified as non-preferred images properly reflected the subject's disfavor. In addition, the experimental results showed that the attention heat maps of the Grad-CAM results were matched to the subject's preference decision by 66.68% and 60.74% of the preferred images and

non-preferred images, respectively. Accordingly, the Grad-CAM results might be utilized to partially explain personal preference, even though it is insufficient to describe the subject's visual preference with only current Grad-CAM results.

Figure 9. Visual expressions applying Grad-CAM for user preference-classification results with LFW: (a) sample heat maps for the results in preferred data, and (b) sample heat maps for the results in non-preferred data.

Table 5. Consistency analysis of Grad-CAM results with the subject's preference decision with the LFW and Indoor Scene Recognition datasets.

	LFW (Test Set: 1000)		Indoor Scene Recognition (Test Set: 3000)	
	Preferred (465)	Non-Preferred (371)	Preferred (1585)	Non-Preferred (1103)
Grad-CAM results match with subject's preference decision	67.10% (312/465)	64.15% (238/371)	66.68% (1057/1585)	60.74% (581/1103)
Grad-CAM results do not match with subject's preference decision	32.90% (153/465)	35.85% (133/371)	33.32% (528/1585)	39.26% (522/1103)

4. Conclusions and Further Work

In this paper, we proposed a deep CNN-based transfer learning model for inferring personal visual preferences in visual images with atypical characteristics. For performance verification of the proposed model, we considered three public benchmark datasets (Fashion-MNIST, LFW, and Indoor Scene Recognition) that have atypical visual characteristics compared with other inter-domain datasets. Experimental results showed that the proposed model properly inferred personal preferences for visual inter-domain images. Moreover, the proposed model also used a transfer learning process that showed better performance for preference prediction with atypical datasets. In addition, the proposed model applied a Grad-CAM approach in order to try to explain personal preferences, even though it is partial to giving explanations about personal preferences, since personal preferences are typically unexplainable, even for humans.

As further work, we are considering many more experiments with multiple users in order to verify the generality of the proposed model. Moreover, we have to use different datasets with different atypical characteristics. Finally, we plan to enhance the proposed model in order to make a personal preference inference model with many more plausible explanatory possibilities. For a theoretical contribution on developing a personal visual preference model, we need to find more biological mechanisms related to generating attention or indirectly obtain insights from known biological mechanisms in further work since the attention mechanism of humans is very complex in nature. As long-term

further work, we are considering combining the personal visual preference model with the Grad-CAM based explanatory model to develop an automatically explainable personality model. We will apply the proposed model to a visual personal recommendation system that is superior to human experts. For example, the proposed personal preference inference model can be applied to image-based personal recommendation systems that might be superior to human experts on customized interiors.

Author Contributions: Conceptualization, J.O. and S.-W.B.; data curation, M.K.; implementation and experiments, J.O.; project administration, S.-W.B.; writing—original draft, J.O.; writing—review and editing, S.-W.B. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (2016-0-00564, Development of Intelligent Interaction Technology Based on Context Awareness and Human Intention Understanding).

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

References

- Shankar, D.; Narumanchi, S.; A Ananya, H.; Kompalli, P.; Chaudhury, K. Deep learning based large scale visual recommendation and search for e-commerce. *arXiv* **2017**, arXiv:1703.02344.
- Tallapally, D.; Sreepada, R.S.; Patra, B.K.; Babu, K.S. User preference learning in multi-criteria recommendations using stacked auto encoders. In Proceedings of the 12th ACM Conference on Recommender Systems, Association for Computing Machinery, New York, NY, USA, 2–7 October 2018; pp. 475–479.
- Yang, L.; Hsieh, C.-K.; Estrin, D. Beyond Classification: Latent User Interests Profiling from Visual Contents Analysis. In Proceedings of the 2015 IEEE International Conference on Data Mining Workshop, Atlantic City, NJ, USA, 14–17 November 2015; pp. 1410–1416.
- Subramaniaswamy, V.; Logesh, R. Adaptive KNN based Recommender System through Mining of User Preferences. *Wirel. Pers. Commun.* **2017**, *97*, 2229–2247. [[CrossRef](#)]
- Chu, W.-T.; Tsai, Y.-L. A hybrid recommendation system considering visual information for predicting favorite restaurants. *World Wide Web* **2017**, *20*, 1313–1331. [[CrossRef](#)]
- Oh, J.; Kim, M.; Ban, S. User preference Classification Model for Atypical Visual Feature. In Proceedings of the 8th International Conference on Green and Human Information Technology 2020, Hanoi, Vietnam, 5–7 February 2020; pp. 257–260.
- Chen, H.; Sun, M.; Tu, C.; Lin, Y.; Liu, Z. Neural Sentiment Classification with User and Product Attention. In Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, Austin, TX, USA, 1–5 November 2016; pp. 1650–1659.
- Gurský, P.; Horváth, T.; Novotný, R.; Vaneková, V.; Vojtáš, P. UPRE: User Preference Based Search System. In Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence, Hong Kong, China, 18–22 December 2006; pp. 841–844.
- McAuley, J.J.; Targett, C.; Shi, Q.; Hengel, A.V.D. Image-based recommendations on styles and substitutes. In Proceedings of the 38th international ACM SIGIR Conference on Research and Development in Information Retrieval, Santiago, Chile, 9–13 August 2015; pp. 43–52.
- Liu, Q.; Wu, S.; Wang, L. DeepStyle: Learning user preferences for visual recommendation. In Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval, Tokyo, Japan, 7–11 August 2017; pp. 841–844.
- Deldjoo, Y.; Elahi, M.; Cremonesi, P.; Garzotto, F.; Piazzolla, P.; Quadrana, M. Content-Based Video Recommendation System Based on Stylistic Visual Features. *J. Data Semant.* **2016**, *5*, 99–113.
- Savchenko, A.V.; Demochkin, K.V.; Grechikhin, I. User preference prediction in visual data on mobile devices. *arXiv* **2019**, arXiv:1907.04519.
- Farseev, A.; Samborski, I.; Filchenkov, A.; Chua, T.-S. Cross-Domain Recommendation via Clustering on Multi-Layer Graphs. In Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval, Tokyo, Japan, 7–11 August 2017; pp. 195–204.
- Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; MIT Press: Cambridge, MA, USA, 2016.

15. Yosinski, J.; Clune, J.; Bengio, Y.; Lipson, H. How transferable are features in deep neural networks? In *Advances in Neural Information Processing Systems*; Jordan, M.I., LeCun, Y., Solla, S.A., Eds.; MIT Press: Cambridge, MA, USA, 2014; pp. 3320–3328.
16. Xiao, H.; Rasul, K.; Vollgraf, R. Fashion-MNIST: A novel image dataset for benchmarking machine learning algorithms. *arXiv* **2017**, arXiv:1708.07747.
17. Huang, G.B.; Ramesh, M.; Berg, T.; Learned-Miller, E. Labeled Faces in the Wild: A Database For studying Face Recognition in Unconstrained Environments. Available online: <http://vis-www.cs.umass.edu/lfw> (accessed on 8 January 2020).
18. Quattoni, A.; Torralba, A. Recognizing indoor scenes. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 413–420.
19. Selvaraju, R.R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; Batra, D. Grad-cam: Visual explanations from deep networks via gradient-based localization. In Proceedings of the 2017 IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 618–626.
20. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
21. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. *Commun. ACM* **2017**, *60*, 84–90. [[CrossRef](#)]
22. Kingma, D.P.; Jimmy, B.A. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.
23. Zhou, B.; Khosla, A.; Lapedriza, A.; Oliva, A.; Torralba, A. Learning deep features for discriminative localization. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 2921–2929.

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).