ISSN 2436-5408 (Online)
ISSN 1346-2601 (Print)

Difference in Perceived Similarity Between Humans and Machines
——A Case Study on Ukiyo-e Artworks

Zhenao Wei (Doctoral student, Graduate School of Information Science and Engineering, Ritsumeikan University)
Shizhe Wang (Graduate, Graduate School of Information Science and Engineering, Ritsumeikan University)
Pujana Paliyawan (Senior researcher, Research Organization of Science and Technology, Ritsumeikan University)
Ruck Thawonmas (Professor, College of Information Science and Engineering, Ritsumeikan University)
E-mail ruck@is.ritsumei.ac.jp

要旨

本稿では、美術品画像間の類似度において、人間の知覚と機械の知覚との差異について述べる。ここで機械は深層学習モデルの特徴を用いた方法を指す。現在、深層学習モデルの特徴を用いて計算する類似度は、画像検索エンジンや画像推薦システムを構築するなど様々な領域に使用されている。しかし、人間が類似度を判断する際に、色、スタイル、明るさ、シーンなどの複数の側面を考慮にいれる。そこで ARC が所蔵している浮世絵の画像を対象に調査を行い、その結果を報告する。

abstract

In this paper, we describe difference between human perception and machine perception in the similarity of art images. Here the term "machine" is referred to methods using features extracted from a deep learning model. Currently, similarity calculated using such deep-learning features is used in various purposes such as for building image search engines and image recommendation systems. However, when humans judge similarity, they consider multiple aspects such as color, style, brightness, and scene. Therefore, we investigate such difference using ukiyo-e images in the possession of ARC and report our findings.

1. Introduction

The digitization of artworks is a recent trend, which has also driven the development of online databases and recommender systems (e.g. Li *et al.*, 2020; Messina *et al.*, 2019; He *et al.*, 2016, all the three being about online databases and the last two also about recommender systems). Ukiyo-e, a kind of Japanese woodblock print, has remained popular since the 17th century (Kikuchi *et al.*, 1969); it is an important part of art history (Li *et al.*, 2020). In the 1990s, Art Research Center (ARC) of Ritsumeikan University launched a project for digital archiving of cultural assets (Bincsik *et al.*, 2012). As one of the representative ukiyo-e databases, it has publicly released more than 19,000 ukiyo-e images online[1]. In 1996, the Web Gallery of Art[2], a database containing more than 51,400 art reproductions, was made public on the internet. More recently, online artwork databases Google Arts & Culture[3] and WikiArt[4] have flourished since their inception.

The online artwork databases mentioned above have a large amount of data. When the user browses their contents, this causes confusion and reduces the user experience. Recommender systems help users discover items of interest from a vast resource collection (Pradhan *et al.*, 2020b). Therefore, for later

developed online artwork databases (e.g., Google Arts & Culture and WikiArt), the introduction of a recommender system is the norm. Earlier online artwork databases, such as the ARC database and Web Gallery of Art, were built even before the introduction of classic content-based filtering (CBF) (Melville *et al.*, 2002) or collaborative filtering (CF) (Breese *et al.*, 1998) recommender systems, not to mention the more recent hybrid recommendation (HR) (He *et al.*, 2016; Pradhan *et al.*, 2020b) and network-based recommendation (Yu *et al.*, 2018). Despite the rapid changes in technology, the similarity between items still has a huge impact on CBF (Messina *et al.*, 2019), HR (Pradhan *et al.*, 2020b), and even network-based recommendation (Pradhan *et al.*, 2020a). To implement a recommender system for image databases, such as the ukiyo-e artwork database by ARC, it is needed to employ a method to measure the similarity in contents between images.

Similarity between images can be computationally measured using image-processing techniques. With the development of deep learning technology, methods that employ deep features (i.e., image features derived from a deep-learning network) have been proposed for solving image processing related tasks. For example, to represent the style of an image, the Gram matrix was used and applied to feature maps in a deep-learning network (Gatys *et al.*, 2016). In addition, Chu *et al.* (2016, pp. 402-406) and Matsuo *et al.* (2016, pp. 309-312) proposed several representations based on deep feature maps, respectively. Wei *et al.* (2017, pp. 60-62) explored whether those image representations could be used to calculate image-style similarity. Recently, Chu et al. (2018, pp. 2491–2502) and Zhang et al. (2021, pp. 3106–3114) used those image representations for image classification and recommendation, respectively. In this work, we aim at verifying the difference in image similarity between that perceived by humans and that calculated by machines, in particular, state-of-the-art methods using deep features. This study reveals that they are significantly different by comparing results in image selection (i.e., selecting two most similar images from a set of images) by 41 participants and three kinds of deep-feature-based methods. It is noted that this study also takes the difficulty of questions into consideration; and relevant results, to be discussed in Sections 4.2 and 4.3, reveal that the above three methods perceive the similarity of artworks differently from humans even on easy questions, in which most humans have the same answer for the most similar image pair of artworks.

2. Deep Feature Vectors

An image can be represented by a set of features. Such features include those derived from the image itself and those available in its keywords and description. We focus on the former type of features, in particular deep features which are shown promising for image processing related tasks in the literature (Gatys *et al.*, 2016; Chu *et al.*, 2016; Matsuo *et al.*, 2016; Wei *et al.*, 2017; Chu *et al.*, 2018; Zhang *et al.*, 2021).

This study continues the practice of the above work (Gatys et al., 2016; Chu et al., 2016; Matsuo et al., 2016; Wei et al., 2017; Chu et al., 2018), using the VGG-19 model trained on the ImageNet dataset. From our previous study (Wei *et al.*, 2017), image representations based on deep features using the Gram matrix, Cosine matrix, and Gram-dot-Cosine matrix are the best three, among five examined, in the similarity calculation of ukiyo-e images. Here, the ij^{th} element of the Gram matrix representation is

$$G_{ij}^{l} = \sum_k f_{ik}^{l} f_{jk}^{l}, \qquad (1)$$

where $\boldsymbol{f}_i$ and $\boldsymbol{f}_j$ are the i^{th} and j^{th} feature maps in the l^{th} layer of the neural network, and k is the k^{th} element of the feature map.

The ij^{th} element of the Cosine matrix and the Gram-dot-Cosine matrix, both proposed by Chu and Wu (2016, pp. 402-406), are given below, respectively.

$$C_{ij}^{l} = \sum_k \frac{f_{ik}^{l} f_{jk}^{l}}{||\boldsymbol{f}_i^{l}|| \, ||\boldsymbol{f}_j^{l}||} \qquad (2)$$

$$GdC_{ij}^{l} = G_{ij}^{l} C_{ij}^{l} \qquad (3)$$

Note that the Cosine matrix is the cosine correlation between the feature maps and the Gram-dot-Cosine is the matrix obtained by element-wise multiplying the values of both Gram and Cosine matrices. In addition, to use these matrices as deep features, they are reshaped to a one-dimensional vector, called feature vector. These three feature vectors are used in the experiment below.

3. Experiment

We conduct an experiment on a selected ukiyo-e image dataset, with the purpose of verifying whether image similarity calculated based on each of the above three deep features is consistent with human perception. All experimental data can be found on our Open Science Framework page[5].

3-1. Ukiyo-e Image Dataset

There are various genres of ukiyo-e images in the ARC ukiyo-e database. Only a few of them have more than one thousand images, and some of them have hundreds, but most of them have only a few dozen. Therefore, to obtain reliable experimental results, 100 samples were randomly selected from each of the top four genres that have the largest number of images. Those 100 samples per genre were then randomly divided into 20 image groups, in each of which participants will select the pair of the most similar images. It is noted that grouping is done by randomization to prevent bias in the data. Table 1 shows the four genres of ukiyo-e images that are used in our experiment.

3-2. Procedure

Our experimental procedure consists of three steps as follows:

(1) Employ the aforementioned deep features to calculate the similarity of a pair of images in each image group,

(2) Conduct a questionnaire to obtain the subjective perception of image similarity in each image group, and

(3) Compare and analyze the difference in similarity between that perceived by humans and that calculated by each deep-feature-based method.

3-2-1. Image Similarity Calculation

Each ukiyo-e image used in this experiment is input to the deep Convolutional Neural Network VGG-19 (Simonyan *et al.*, 2015), and

Table 1 Tested top four genres of ukiyo-e images

Example Image	Genre	Description
	Yakusha-e	Portraits of individual artists (sometimes in pairs).
	Bijin-ga	Images of beautiful women
	Meisho-e	Images of famous places
	Monogatari-e	Images of folktales and mythologies

the resulting values at the conv5_1 layer, containing 512 layers, of VGG-19 are extracted. This layer of VGG-19 was shown to have the best performance in previous work (Chu *et al.*, 2016; Matsuo *et al.*, 2016) for the task therein, and was also used in the work of Wei *et al.* (2017, pp. 60-62). Note that VGG was also recently used to extract the features and style information of images (Wu *et al.*, 2021), which shows that the model is still promising. Extracted values are then used to calculate any of the aforementioned three representations of the image, i.e., Gram matrix (*Gram*), Cosine matrix (*Cosine*), and Gram-dot-Cosine matrix (*Gram·Cosine*). Because all the three representation methods are symmetric, the redundant information in their 512 × 512 matrix is removed by only considering the elements of the lower triangle (including those on the diagonal) to form a 131,328 × 1 feature vector. In this work, the similarity of any two images is derived using the Euclidean distance between their feature vectors, those with the lower the distance the higher the similarity.

3-2-2. Image Similarity Perception

Participants: We initially recruited 54 multi-national students studying computer science in our university. Participants who self-reported of no visual impairment, through the questionnaire described below, and completed the entire questionnaire were included in our analysis. This results in the number of valid participants = 41.

Questionnaire: We conducted an online questionnaire. First, participants need to perform a self-assessment regarding ukiyo-e familiarity, their areas of study, and visual impairment. Then, each participant is asked 80 questions (4 genres with 20 image groups per genre), as shown in Figs. 2, 3, and 4. In each question, they were asked to select the most similar pair of images from each group of 5 images, where this number was empirically determined.

4. Results and Discussions

4-1. Overall Performance Comparison

Tables 2 and 3 show the overall experimental results. The notations in use are defined as follows:

- $Ps(x)$ is the ratio of selecting the image pair of type x in an image group by participants and is defined as

$$Ps(x) = \frac{Number\ of\ participants\ selecting\ x}{N} \quad (4)$$

  where N is the number of participants (in this study N = 41), and x represents *Majority*, *Gram*, *Cosine*, and *Gram·Cosine* whose image pair is most selected by the participants or is selected due to having the lowest Euclidian distance for *Gram*, *Cosine*, and *Gram·Cosine* feature vectors, respectively. In Table 2, *Avg Ps*(x) is the average of $Ps(x)$ for each genre. Please note that the sum of each column in Table 2 is not necessarily constrained by 100% because image pairs selected by these four methods do overlap.
- *Agree*(x) is the ratio that the image pair of type x, defined above, are the same as that of *Majority* and is defined as

$$Agree(x) = \frac{\sum_1^M (x == Majority\ ?\ 1:0)}{M} \quad (5)$$

  where M is the total number of questions per genre (in this study M = 20); and the "$x == Majority\ ?\ 1:0$" expression, using the ternary operator, means that when x is $Majority$, x is 1, otherwise 0.

As can be seen from Table 2, on average, 53.05% of the participants selected the same pair for each question in the questionnaire. For humans, Yakusha-e is the most difficult genre to select similar images, while Monogatari-e the least difficult one. The average ratios of participants selecting the same image pair as each of the three deep-feature- based methods are only 26.62%, 30.67%, and 26.13%, respectively. These results show that the deep-feature-based methods or machines select image pairs differently from participants or humans.

Table 3 shows that the agreement ratio between each deep-feature-based method and the majority of participants is low: 40%, 43.75% and 36.25% respectively. Among the three deep-feature-based methods, the one using the *Cosine* feature vector outperforms the others for

Table 2 Average ratio of participants selecting the same image pair as each method

	Yakusha-e (%)	Bijin-ga (%)	Meisho-e (%)	Monogatari-e (%)	All (%)
AvgPs(*Majority*)	47.68	50.49	53.66	60.37	53.05
AvgPs(*Gram*)	20.00	32.80	29.27	24.39	26.62
AvgPs(*Cosine*)	30.49	31.46	34.27	26.46	30.67
AvgPs(*Gram·Cosine*)	19.76	30.98	28.78	25.00	26.13

Table 3 Agreement ratio between the image selection by each deep-feature-based method and the majority of participants

	Yakusha-e (%)	Bijin-ga (%)	Meisho-e (%)	Monogatari-e (%)	All (%)
Agree(*Gram*)	35.00	**55.00**	35.00	35.00	40.00
Agree(*Cosine*)	**40.00**	50.00	**45.00**	**40.00**	**43.75**
Agree(*Gram·Cosine*)	35.00	45.00	30.00	35.00	36.25

Yakusha-e, Meisho-e, and Monogatari-e, except for Bijin-ga where the *Gram* feature vector has the highest performance. In fact, this trend can also be seen in Table 2. These results indicate that when building a recommender system with an expectation to recommend similar images to humans, it is important to keep in mind that there exists discrepancy between similarity perceived by machines and that by humans. As a result, results from such systems might need to be adjusted, which however is beyond the scope of this paper.

4-2. Performance and Question Difficulty

How much does the difficulty of questions influence the selection of similar images? Will machines think more like humans in easy questions in which people's judgement is less divided? In order to obtain some insights to these questions, we perform an analysis according to question difficulty. Here, we consider a question of interest an easy question when a large number of participants have the same answer for the most similar image pair. Similarly, we consider it a difficult question when participants' answers are split. In other words, our definition of "difficulty" is inversely related to "the agreement of similarity judgements among participants. We sort questions according to *Ps*(*Majority*) from high to low. Tables 4 and 5 show the results on the top-10 easiest and most difficult questions, respectively.

Table 4 Matching between ***Majority*** and each deep-feature-based method in the top-10 easiest questions, where "==" is the equality operator.

Rank	Genres	*Avg Ps* (*Majority*) (%)	*Majority* == *Gram*?	*Majority* == *Cosine*?	*Majority* == *Gram·Cosine*?
1	Meisho-e	95.12	**True**	**True**	**True**
2	Yakusha-e	92.68	False	**True**	False
3	Meisho-e	92.68	False	**True**	False
4	Monogatari-e	90.24	False	False	False
5	Monogatari-e	87.80	False	False	False
6	Meisho-e	85.36	False	False	False
7	Bijin-ga	82.92	**True**	**True**	**True**
8	Yakusha-e	80.48	False	**True**	False
9	Monogatari-e	78.04	False	False	False
10	Monogatari-e	75.60	False	False	False

Table 5 Matching between ***Majority*** and each deep-feature-based method in the top-10 most difficult questions, where "==" is the equality operator.

Rank	Genres	*Avg Ps*(*Majority*) (%)	*Majority* == *Gram*?	*Majority* == *Cosine*?	*Majority* == *Gram·Cosine*?
10	Bijin-ga	34.14	**True**	**True**	False
9	Yakusha-e	31.70	False	False	False
8	Bijin-ga	31.70	**True**	False	**True**
7	Meisho-e	29.26	False	False	False
6	Yakusha-e	24.39	False	False	False
5	Yakusha-e	24.39	**True**	False	**True**
4	Yakusha-e	24.39	False	False	False
3	Meisho-e	24.39	False	**True**	False
2	Bijin-ga	24.39	**True**	**True**	False
1	Bijin-ga	19.51	False	False	False

In the set of top-10 easiest questions, nearly half of them are Monogatari-e. However, there is no Monogatari-e question in the set of top-10 most difficult questions. These results indicate that, compared to the other genres used in this study, it is less difficult for humans to find the most similar Monogatari-e images. In addition, only one question using Bijin-ga images resides in the former set, but 40% of the questions in the latter set are from this genre, which indicates that Bijin-ga is more difficult than other genres for humans to find the most similar images.

The best deep-feature-based method, in terms of matching with *Majority*, is the one using the *Cosine* feature vector for Yakusha-e, Meisho-e, and Monogatari-e and the *Gram* feature vector for Bijin-ga. However, their average performance for all the genres is no more than 50%. This also shows that, as far as ukiyo-e images are concerned, there is a notably difference in the perceived similarity between humans and machines.

4–3. Cumulative Percentages for Each Similarity

Fig. 1 shows the cumulative moving average of $Ps(x)$ for each method x over the list of all questions sorted by $Ps(Majority)$ in descending order. All methods have a decrease trend because more and more difficult questions appear later in the list. The trend of *Gram* and that of *Gram·Cosine* are indistinguishable from each other. For *Cosine*, the method is remarkably more accurate for easy questions than both *Gram* and *Gram·Cosine*.

Let define easy questions as those with $Ps(Majority)$ above 80%, by which there are eight of such questions (questions 1 to 8), accounting for 10% of the questions in this study. For those questions, on average 88.41% of the participants form a consensus in image similarity. In contrast, the deep-feature-based methods could reach only 57.01% with *Cosine*, while much less 23.47% with both *Gram* and *Gram·Cosine*.

4–4. Treatment a Special Case

In all of our questions, there is only one tie result where the question has multiple *Majority* pairs (the '3' in Table 5). Fig. 2 shows the question. The three *Majority* pairs are (a) - (b), (c) - (d) and (c) - (e), all with 24.39% of votes. *Gram* and *Gram·Cosine* both take the (a) - (d) pair, but *Cosine* takes (a) - (b), which is one of the *Majority* pairs. As a result, for tie-breaking, *Majority* is deemed as *Cosine* in this case, as shown in Table 5.

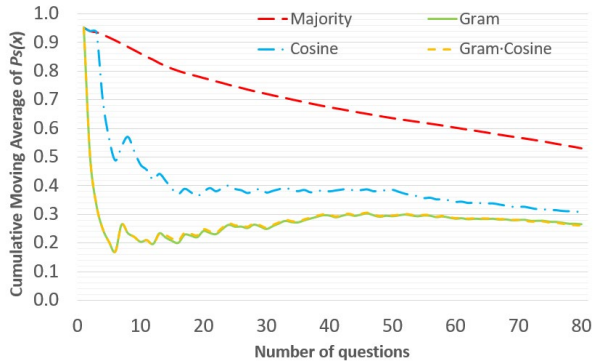

Fig. 1. The cumulative moving average of $Ps(Majority)$, $Ps(Gram)$, $Ps(Cosine)$, and $Ps(Gram·Cosine)$ in red, green, blue, and yellow, respectively, where the 80 questions are sorted in decreasing order of $Ps(Majority)$

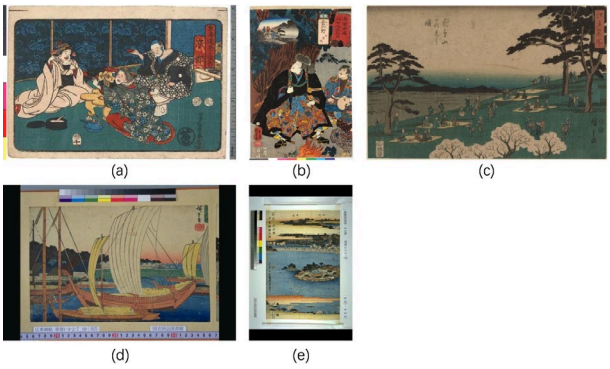

Fig. 2. A special case where multiple *Majority* pairs exist

Fig. 3. An example set of ukiyo-e images where the most similar image selection is the same between humans and the deep-feature-based methods

4–5. Discussions

4–5–1. Case Analysis

Fig. 3 shows an example where 95.12% of the participants and all the deep-feature-based methods chose (d) - (e) as the most similar images. Fig. 4 shows an example set of 5 ukiyo-e images where 90.24% of the participants chose (d) - (e) as the most similar images. It is evident that from the human perspective since this pair looks similar at first glance, both depicting a scene of war.

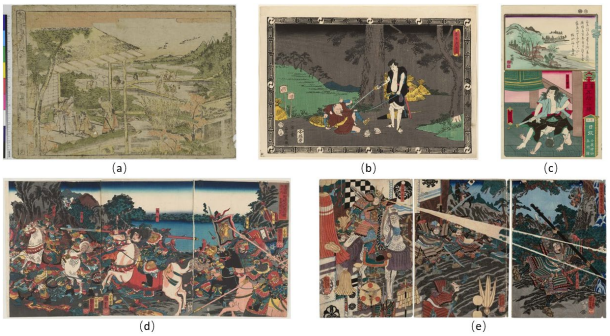

Fig. 4. An example set of ukiyo-e images where the most similar image selection is different between humans and the deep-feature-based methods

However, all the deep-feature-based methods took a different pair: (c) - (e).

In Fig. 3, (d) - (e) pair has three similarities: the size ratio (vertical print), the item (the person wearing the straw hat) and the artistic conception (the scene along the river). In Fig. 4, according to the results of *Majority*, the similarity of human perception is affected by the size ratio (horizontal prints of the triplet) and the artistic conception (the battle between the two armies). The deep-feature-based methods are more concerned about whether there is similar item (the victorious samurai in (c) and the rightmost soldier in (e)) in the images.

4–5–2. Discussion on the Difference of Human Perceived Similarity Measures

This paper focuses on the difference in perceived similarity between humans and machines. However, human-perceived similarity also has shortcomings; it is subjective as the understanding of similarity is different for each individual, and how to measure the human-perceived similarity is a difficult problem. Nevertheless, Díaz-Agudo *et al.* (2021, pp. 48–63) proposed a method of mixing different kinds of similarity--namely color similarity, content similarity, emotion similarity, and knowledge similarity--to measure the human-perceived similarity, and verified the effectiveness of the method through experiments. Their work provides us a direction for data collection in future work--we will collect human-perceived similarity in different kinds.

5. Conclusion

This paper investigated whether deep-feature-based methods perceive the similarity between ukiyo-e artworks similarly to humans or not. An online questionnaire was designed and answered by 41 participants, for verifying the performance of each deep-feature-based method used in the experimentation. We analyzed experimental results from the aspects of genre and difficulty, have revealed that there is a considerable difference between humans and the state-of- the-art deep-feature-based methods in the perception of ukiyo-e image similarity.

In the future, using a larger size of participant population with a higher variety of background, we plan to extend our method to examine the difference between humans and the deep-feature-based methods in the similarity perception of other kinds of artwork images. Moreover, we will develop a game with a purpose (GWAP) that allows players to label the image similarity for a given pair of images while playing the game. Based on the acquired similarity data through this game, we plan to develop deep-feature-based methods for calculation of image similarity whose results match more with human perception by taking into account aspect ratios and artist's conceptions.

[Acknowledgement]

This research was supported in part by Grant-in-Aid for Scientific Research (C), Number 19K12291, Japan Society for the Promotion of Science, Japan.

[Annotations]

1) http://www.dh-jac.net/db/nishikie-e/search.php?enter=default
2) https://www.wga.hu/index.html
3) https://artsandculture.google.com/
4) https://www.wikiart.org/
5) https://osf.io/rwmz5/

[References]

Belén Díaz-Agudo, Guillermo Jimenez-Diaz, Jose Luis Jorro-Aragoneses. User Evaluation to Measure the Perception of Similarity Measures in Artworks, in ICCBR 2021. Lecture Notes in Computer Science, vol 12877. 2021, pp. 48–63.

John S. Breese, David Heckerman, Carl Kadie. Empirical analysis of predictive algorithms for collaborative filtering, in Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence. 1998, pp. 24–26.

Kangying Li, Biligsaikhan Batjargal, Akira Maeda, Ryo Akama. Artwork information embedding framework for multi-source Ukiyo-e record retrieval, in Digital Libraries at Times of Massive Societal Transition. 2020, pp. 255–261.

Karen Simonyan, Andrew Zisserman. Very Deep Convolutional Networks for Large-Scale Image Recognition, in International Conference on Learning Representations 2015 (ICLR). 2015.

Leon A. Gatys, Alexander S. Ecker, Matthias Bethge. Image Style Transfer using Convolutional Neural Networks, in IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2016). 2016, pp. 2414-2423.

Monika Bincsik, Shinya Maezaki, Kenji Hattori. Digital archive project to catalogue exported Japanese decorative arts, in International Journal of Humanities and Arts Computing. 2012, pp. 42-56.

Pablo Messina, Vicente Dominguez, Denis Parra, Christoph Trattner, Alvaro Soto. Content-based artwork recommendation: integrating painting metadata with neural and manually-engineered visual features, in User Model. User-Adapt. Interact. 29. 2019, pp. 251–290.

Prem Melville, Raymond J. Mooney, Ramadass Nagarajan. Content-boosted collaborative filtering for improved recommendations, in Proceedings of Conference on Artificial Intelligence. 2002, pp. 187– 192.

Ruining He, Chen Fang, Zhaowen Wang, Julian McAuley. Vista: a visually, socially, and temporally-aware model for artistic recommendation, in Proceedings of the 10th ACM Conference on Recommender Systems (RecSys 2016). 2016, pp. 309–316.

Sadao Kikuchi. A Treasury of Japanese Wood Block Prints (Ukiyo-e), in New York: Crown Publishers. 1969, pp. 31.

Shin Matsuo, Keiji Yanai. CNN-based Style Vector for Style Image Retrieval, in Proceedings of ACM International Conference on Multimedia Retrieval (ICMR'16). 2016, pp. 309-312.

Shuo Yu, Jiaying Liu, Zhuo Yang, Zhen Chen, Huizhen Jiang, Amr Tolba, Feng Xia. Pave: Personalized academic venue recommendation exploiting co-publication networks, in Journal of Network and Computer Applications, vol 104. 2018, pp. 38–47.

Tribikram Pradhan, Sukomal Pal. CNAVER: A content and network-based academic Venue recommender system, in Knowledge-Based Systems, vol 189. 2020.

Tribikram Pradhan, Sukomal Pal. A hybrid personalized scholarly venue recommender system integrating social network analysis and contextual similarity, in Future Generation Computer Systems. 2020, pp. 1139-1166.

Wei-Ta Chu, Yi-Ling Wu. Deep correlation features for image style classification, in Proceedings of the 2016 ACM on multimedia conference (MM16). 2016, pp. 402-406.

Wei-Ta Chu, Yi-Ling Wu. Image style classification based on learnt deep correlation features, in IEEE Transactions on Multimedia, vol. 20, no. 9. 2018, pp. 2491–2502.

Xiaolei Wu, Zhihao Hu, Lu Sheng, Dong Xu. StyleFormer: Real-time arbitrary style transfer via parametric style composition, in Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). 2021, pp. 14618-14627.

Yiwei Zhang, Toshihiko Yamasaki. Style-Aware Image Recommendation for Social Media Marketing, in Proceedings of the 29th ACM International Conference on Multimedia. Association for Computing Machinery. 2021, pp. 3106–3114.

Zhenao Wei, Lilang Xiong, Kazuki Mori, Tung Duc Nguyen, Tomohiro Harada, Ruck Thawonmas, Keiko Suzuki, Masaaki Kidachi. Deep Features for Image Classification and Image Similarity Perception, in Japanese Association for Digital Humanities Conference 2017 (JADH2017). 2017, pp. 60-62.