System for Searching Illustrations of Anime Characters Focusing on Degrees of Character Attributes

Yuta Koyama[a], Tomohiro Fukuhara[ab], Koichi Yamada[a], Hironobu Abe[a], Hidetaka Masuda[a]
[a]Tokyo Denki University, 5 Senjuasahi-cho, Adachi-ku, Tokyo, JAPAN 120-8551
[b]MULTISOUP CO, LTD, 3-11 Kandanishiki-cho, Chiyoda-ku, Tokyo, JAPAN 101-0054

ABSTRACT

Keyword searches are generally used when searching for illustrations of anime characters. However, keyword searches require that the illustrations be tagged first. The illustration information that a tag can express is limited, and it is difficult to search for a specific illustration. We focus on character attributes that are difficult to express using tags. We propose a new search method using the vectorization degrees of character attributes. Accordingly, we first created a character illustration dataset limited to the hair length attribute and then trained a convolutional neural network (CNN) to extract the features. We obtained a vector representation of the character attributes using CNN and confirmed that they could be used for new searches.

Keywords: Illustration search, Anime characters, Vectorization, CNN

1. INTRODUCTION

On social networking sites (SNS) for posting illustrations of anime characters, illustration searches for anime characters use the metadata on the characters as tags for keyword searches. However, this method requires SNS users to manually add tags to illustrations in advance. Compared to specific information such as the character name, character attributes (e.g., hairstyle, clothes, facial expressions, and posture) tend not to be tagged. Therefore, even though users want to view illustrations with specific attributes, they cannot search for these attributes because the appropriate tags are not attached.

In addition, conventional methods use tag attributes as keywords; however, it is not possible to express details of degrees of character attributes such as the hair length. Therefore, it is difficult to perfectly express attributes that differ by small amounts depending on the anime character, such as hair color or hair length. Keyword searches using tags therefore cannot be used to elaborate on the search results.

We propose a new method that performs feature extraction, focusing on specific attributes of character illustrations using convolutional neural networks (CNNs), and creates vector representations of the illustrations based on degrees of character attributes such as hair length, hair color, and facial expression. Even though vector representations of illustrations using character attributes been generated in the past [1], the vector representation used here focuses only on the degrees of the character attributes. By extracting features for character attributes that have degrees, it is possible to represent illustrations as vectors. As a result, metadata on character attributes that are not tagged or are difficult to assign to illustrations of anime characters can be added to the illustrations. Using this vector representation, it is possible to better search for illustrations of anime characters.

2. RELATED WORK

In a study by Saito et al. [2], a CNN classifier was trained and tags were automatically assigned using the Danbooru[1] illustration dataset. In addition, vector representations of illustrations have been generated by extracting the output of the middle layer of a CNN and proposing “semantic morphing” to search for illustrations in the semantical middle of two illustrations.

---

1 Danbooru, https://danbooru.donmai.us/

International Workshop on Advanced Imaging Technology (IWAIT) 2020, edited by Phooi Yee Lau, Mohammad Shobri, Proc. of SPIE Vol. 11515, 115152B · © 2020 SPIE
CCC code: 0277-786X/20/$21 · doi: 10.1117/12.2566509

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 06 Jun 2020
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

We extended this method of extracting the output of the middle layer as a vector representation of an illustration to generate a vector representation of an illustration focusing on the degrees of character attributes. In this study, we focused on only one attribute, i.e., the hair length, and verified whether the classification of the character of this attribute is possible after vectorizing the concept of length.

3. PROPOSED METHOD

The goal of this study is to perform an illustration search focusing on character attributes that have degrees by extracting the features of the character attributes in illustrations and generating vector expressions. To automatically generate vector representations for use in an illustration search, we trained a CNN using the Danbooru illustration dataset, as in previous studies.

3.1 Dataset

We manually extracted character illustrations from the Danbooru illustration dataset under the following conditions: one female, upper body, white background, frontal composition, color illustration, and hair and face not hidden. A breakdown of the created dataset is shown in Table 1. As shown in Figure 1, the hair length was classified as follows: long hair extended lower than the shoulder blade, medium hair was roughly shoulder length, and short hair was above the chin. Figure 2 shows an example of each type of illustration.

Table 1. Number of training, validation, and testing images.

	Long hair	**Medium hair**	**Short hair**
Training	80	80	80
Validation	10	10	10
Testing	10	10	10
Total	100	100	100

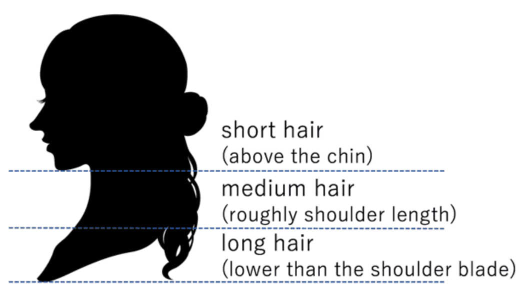

Figure 1. Classification of women's hair length.

Figure 2. Examples of character illustrations with different hair lengths (free illustrations).[2]

[2] Illust AC, https://www.ac-illust.com/

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 06 Jun 2020
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

3.2 CNN's configuration and training

The CNN learns based on the VGG16 [3] model. Because the number of training data in the prepared dataset is small, fine-tuning was performed using a model that had been trained on a large ImageNet image dataset [4]. Because the pre-trained VGG16 model is a CNN that classifies 1000 classes, we replaced fully connected layers with those that output the probability that each label (long hair, medium hair, or short hair) would be assigned according to the prepared dataset. During learning, the weights already learned with the pre-trained VGG16 were not updated up to the 14th layer and learning was performed only from the 15th layer onwards. In addition, random image processing was performed for horizontal flips, horizontal shifts, vertical shifts (only in the downward direction), scaling, and rotation. Other learning options included the following: the number of epochs was set to 100, the learning rate was 0.001, and the optimizer was stochastic gradient descent, which is the most basic algorithm in machine learning.

3.3 Vectorization

The probability of classification to each label was output by inputting the illustration into the trained CNN classifier. We extracted the 4096-dimensional feature vector of the middle layer of the CNN, which can be obtained in the feature extraction process, and treated that vector as a vector representation of the illustration. The value of each element of the vector of the middle layer does not have a clear meaning such as the predicted value; however, it is a high-dimensional vector just prior to the prediction. Therefore, it is considered that this vector expresses the features well, that is, if the character's hair is long, medium, or short in the illustration. We confirmed whether this vector could be used for illustration searches focusing on the degree of the character attributes.

4. RESULTS

Table 2 shows the precision, recall, and F-measure of the classification results for test data using the CNN. It was confirmed that the precision, recall, and F-measure were 0.90 on average and that the performance of the classifier was sufficient. However, the values for medium hair were below average.

Table 2. Validation result for test data.

	Long hair	**Medium hair**	**Short hair**	**Average**
Precision	0.91	0.89	0.90	0.90
Recall	1.00	0.80	0.90	0.90
F-measure	0.95	0.84	0.90	0.90

Figure 3 shows a two-dimensional scatter plot representing the results using principle component analysis (PCA) performed on the 4096-dimensional output of the CNN middle layer for each category. It can be seen that clusters exist for each hair length and that the medium hair cluster exists between long and short hair. In addition, an intermediate degree of illustration between the clusters (an illustration of the cluster representing medium hair with relatively long hair near the cluster representing long hair) is plotted near the boundaries of the clusters.

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 06 Jun 2020
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

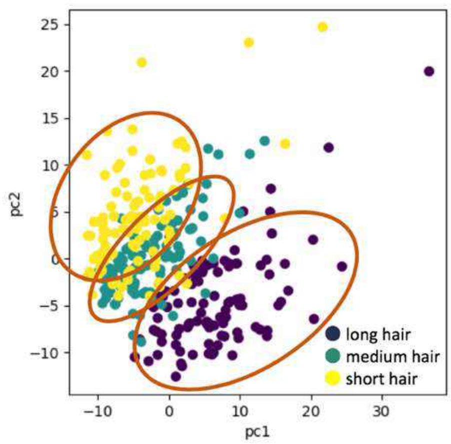

Figure 3. PCA result for the feature value of the output of the middle layer of the CNN.

5. DISCUSSION

We trained a CNN to extract the features of each illustration. It was found that feature extraction focusing on hair length was possible by narrowing the learning attributes to the hair length and learning the short state, intermediate state, and long state. However, for medium hair, the performance was below average. This is because medium hair is in the intermediate state; therefore, the chance of misclassification is higher compared to long hair and short hair whose features are easy to understand.

In addition, it can be seen from the scatter plot that the relationship in which the hair length changes from long hair to medium hair and from medium hair to short hair can be expressed in the vector space. Therefore, it may be possible to perform an illustration search focusing on hair length using this vector representation.

6. CONCLUSION

In this study, we focused on attributes with continuity of characters drawn in illustrations and extracted features to generate a vector representation that can represent the degrees of specific attributes. We focused on one attribute (the hair length) and confirmed that it is possible to extract and classify features focusing on hair length using CNN. The generated vector representation is correlated to the actual hair length.

In the future, we will apply the proposed method to attributes other than the hair length and improve the generalization performance. Then, we will develop a search method accounting for all of the generated vectors.

REFERENCES

[1] Yuta, K., Tomohiro, F., Koichi, Y., Hironobu, A., and Hidetaka, M., "Generation method of vector representation of illustration based on character attributes for illustration search," The 81st National Convention of IPSJ, 2019(1), 203-204 (2019) [In Japanese].

[2] Masaki, S. and Yusuke M., "Illustration2Vec: A Semantic Vector Representation of Illustrations," SIGGRAPH Asia 2015 Technical Briefs Article No. 5 (2015).

[3] Simonyan, K. and Zisserman, A., "Very deep convolutional networks for large-scale image recognition," ICLR (2015).

[4] Krizhevsky, A., Sutskever, I., and Hinton, G., "ImageNet classification with deep convolutional neural networks," NIPS'12 Proceedings of the 25th International Conference on Neural Information Processing Systems, 1097-1105 (2012).

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 06 Jun 2020
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use