International Journal of Internet, Broadcasting and Communication Vol.9 No.4 1-7 (2017)
http://dx.doi.org/10.7236/IJIBC.2017.9.4.1

IJIBC 17-4-1

Comparison of Fine-Tuned Convolutional Neural Networks for Clipart Style Classification

Seungbin Lee[1], Hyungon Kim[1], Hyekyoung Seok[1], Jongho Nang[1*]

[1]*Department of Computer Science and Engineering Sogang University, Seoul, Korea*

[mercileesb, hgk93, stonehye, jhnang]* *@ sogang.ac.kr*

Abstract

Clipart is artificial visual contents that are created using various tools such as Illustrator to highlight some information. Here, the style of the clipart plays a critical role in determining how it looks. However, previous studies on clipart are focused only on the object recognition [16], segmentation, and retrieval of clipart images using hand-craft image features. Recently, some clipart classification researches based on the style similarity using CNN have been proposed, however, they have used different CNN-models and experimented with different benchmark dataset so that it is very hard to compare their performances. This paper presents an experimental analysis of the clipart classification based on the style similarity with two well-known CNN-models (Inception Resnet V2 [13] and VGG-16 [14] and transfers learning with the same benchmark dataset (Microsoft Style Dataset 3.6K). From this experiment, we find out that the accuracy of Inception Resnet V2 is better than VGG for clipart style classification because of its deep nature and convolution map with various sizes in parallel. We also find out that the end-to-end training can improve the accuracy more than 20% in both CNN models.

Keywords: *Clipart Classification; Convolutional neural network; Computer vision; Clipart, style search; Fine tuning; Deep learning;*

1. Introduction

Today, we use a variety of visual content such as advertising images and flyers in the fields of advertising, marketing, and entertainment. In this content, clipart is mainly used to summarize the intentions and features of the content designers, and style is the main element that conveys the meaning of the image. Nevertheless, style is very subjective and it is difficult to create a label that fits different styles. However, the study of style has rarely been spotlighted. Despite Refs. [1-7], the main studies of style are rarely used and the majority of other studies calculate the distance between feature vectors by extracting the aggregate, and employ dimension reduction using feature extraction to search for similar style images. In these cases, performance varies depending on their hand-craft features. Recently, transfer learning was applied to solve clipart style classification problems using Convolutional Neural Network (CNN) models. CNN models, that have been proven to be superior in the existing Imagenet Challenge(ILSVRC), were used to classify the style of clipart and their performances were measured by fine-tuning with their own style dataset.

This paper presents some experimental results on clipart classification problem using two well-known CNN

Manuscript Received: Sep. 5, 2017 / Revised: Sep. 17, 2017 / Accepted: Sep. 30, 2017
Corresponding Author: jhnang@sogang.ac.kr
Tel: +82-2-705-8494, Fax: +82-2-704-8273
Department of Computer Science and Engineering, Sogang University, Korea

models via fine-tuning. The results show that the accuracy of Inception ResnetV2 is better than VGG for clipart style classification because of its deep nature and convolution map with various sizes in parallel. In addition, the end-to-end learning (i.e., an error backpropagation to feature extraction layer) could improve the accuracy more than 20%.

In Section 2, the search studies using existing styles will be reviewed. The data sets and CNN models used in the experiment will be discussed in Section 3, and the performance and characteristics of CNN models will be discussed in Section 4. Finally, conclusions and future research will be discussed in Section 5..

2. Related works

2.1 Similarity research using hand-craft visual features

There have been several studies focused on similar style searches. Table 1 shows the methods of extracting defined features and applying them to the retrieval, such as existing contents-based image retrieval. Refs. [1-3] simplify complex structures by using Query by Sketch. The simplified image is represented by a relationship structure between inclusion and adjacency through topological analysis to search for similar images. Ref. [4] extracts color, texture, and shape information from a raster image and uses the topology and geometry features extracted from a vector image as retrieval parameters. Ref. [5] extracts color, shading, texture, and stroke features from images, which are then metric-learned through information obtained from MTurk. Ref. [6] extracts various features for an infographic similarity search and those features are metric-learned using MTurk. Ref. [7] aggregates features extracted from various scales and performs dimension reduction to search for similar images.

Table 1. Previous Researches on Similar Style using Feature Descriptors

Work	Purpose	Method		Datatype	Benchmark dataset
		Used features	Metric		
Barroso [1]	retrieval	topological feature	graph matching	vector drawings	968 drawings
Fonseca [2]	retrieval	topological feature, geometrical feature	KNN search	CAD, clipart	968 drawings [1]
Sousa [3]	retrieval	topological feature, geometrical feature	graph matching	clipart	100(20*5) drawings
Martins [4]	retrieval	color, texture, shape, topological feature, geometrical feature,	L1, L2	clipart	100(10*10) drawings
Garces [5]	retrieval	color, shading, texture, stroke	metric learning [21]	clipart	MS 3.6k, commercial 200k
Saleh [6]	retrieval	GIST, HoG, LBP, histogram of color and luminance	metric learning [22]	infographic	2082 drawings
Furuya [7]	retrieval	color, texture	L2, cosine	clipart	MS 3.6k [5]

2.2 Similarity research using convolutional neural network

The results of the methods in Section 2.1 can vary depending on which features are selected and what processing is performed. In order to eliminate this uncertainty, CNN-based studies are suggested and these are shown in Table 2.

CNN abstracts images by layers and creates easily distinguishable features. Ref. [8] used various features such as L*a*b, GIST, and DeCAF5,6 [18]. Ref. [9] constructs a Siamese network that shares parameters of CNN obtained from actual and original photographs of different products and the system learns to embed them into a low-dimensional space to reduce the number of variables. Ref. [10] constructs a CNN model

with seven layers, allowing the system to learn how to emulate an auto encoder serving as an initial filter for any arbitrary image. Then, the distance is measured between the features extracted from fully connected layers using L1 distance, L2 distance and cosine similarity. Ref. [11] calculates the similarity in the OASIS method using features extracted from Alexnet's FC1, FC2, and FC3 [19].

However, those studies only focused on instance retrieval in images using CNN and did not conduct focused style-related research.

Table 2. Previous Researches on Similar Style using CNN

Work	Purpose	Method		Datatype	Benchmark dataset
		Used CNN model	Character		
Karayev [8]	classification	DeCAF [19][20]	learning classifier	Photograph	[Flickr style] : 80k images covering 20 styles, [Wikipaintings] : 85k images for 25 art genres
Bell [9]	retrieval	Alexnet [19], GoogLeNet [23]	siamese nets	Furniture	Millions of products and scenes from [Houzz.com]
Kuo [10]	retrieval	Self-designed CNN	auto encoder	Photograph	CIFAR-10, CIFAR-100
Wan [11]	retrieval	DeCAF [19][20]	similarity learning [20]	Photograph	ImageNet, Caltech256, Oxford, Pubfig83LFW Paris,

3. Clipart style classification

3.1 Benchmark Dataset

The Microsoft Style Dataset 3.6k, which was used in [5] and [7], was also used in our study as a benchmark dataset for learning and performance measurement of similar style classification of clipart. The dataset has a total of 3,591 clipart images and they are classified into 220 style classes. Figure 1 shows examples of categorized images.

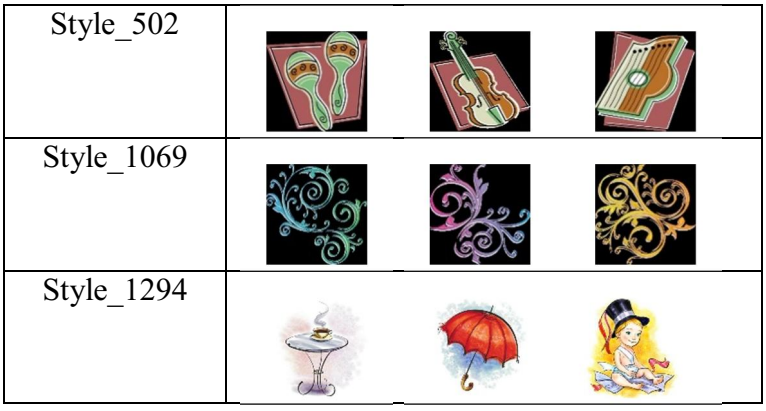


Figure 1. Examples of MS dataset clipart images

3.2 CNN for illustration style classification

In very few datasets, labels are generated according to style. Therefore, a SOTA model, which was proven to work well with images obtained from a broad domain, was used in our experiments rather than

other existing techniques where a CNN model must learn from scratch [13,14,15]. Eighty percent of the Microsoft Style Dataset's 3.6K images were used for learning and 20% of them were used as validation sets. First, learned classifiers were initialized to distinguish the existing ImageNet class. Then, a classifier for 220 classes, which is suitable for the MS dataset, was learned by using a relatively large learning rate (0.01) over 10K iterations. Finally, an end-to-end training session of 50K iterations was performed using a finer learning rate (0.001) in a feature extractor to fine-tune the CNN.

4. Experiments

For the transfer learning, classifier training (Task A) and end-to-end training (Task B) were conducted, as shown in Figure 2. After the pre-trained weight was initialized and trained classifier, its top-1 performance was measured in the range of 55-61% for two CNN models shown in Table 3 when the system was learning with a relatively large learning rate (0.01) for 220 classes. In order to improve performance, each network learned at a finer learning rate (0.001), and overall top-1 performance improved to 78-86%.

Table 3 shows the results after the transfer learning on the two CNN models. The Top-1 accuracy for the Inception Resnet V2 [13] averaged approximately 86%. On the other hand, the performance of the VGG-16 [14] averaged 78%, which is approximately 7% lower than the Inception. However, the Top-10 accuracy for the VGG-16 was 96%, which is comparable to that of the Inception model.

In order to analyze the effect of end-to-end training to two CNN models, the percent of the changed weights in both CNN models after fine-tuning are measured and their results are shown in Table 4. As shown in Table 4, after the end-to-end training, the VGG-16 model learned over 16 layers and showed a change of 2.78%. However, since the ResnetV2 model uses convolution maps of various sizes in parallel in the Inception module, it is very expressive and shows a much higher change. Moreover, although the ResnetV2 model has much deeper layers, it employs an effective backpropagation algorithm in combination with a residual structure, so the change of the filter is large in the domain and its performance is better, as shown in Table 3.

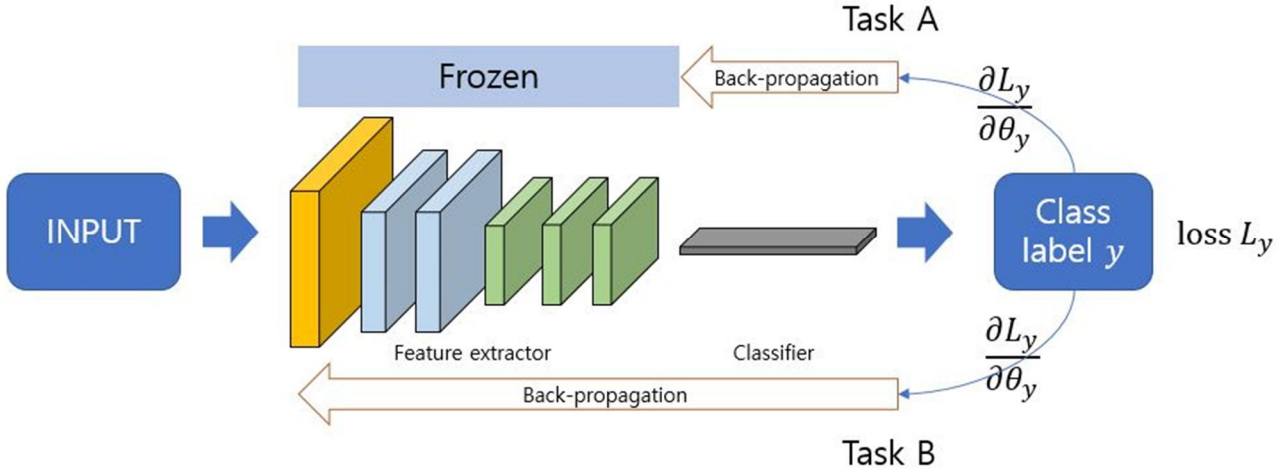


Figure 2. Comparison of classifier training (Task A) and end-to-end training (Task B)

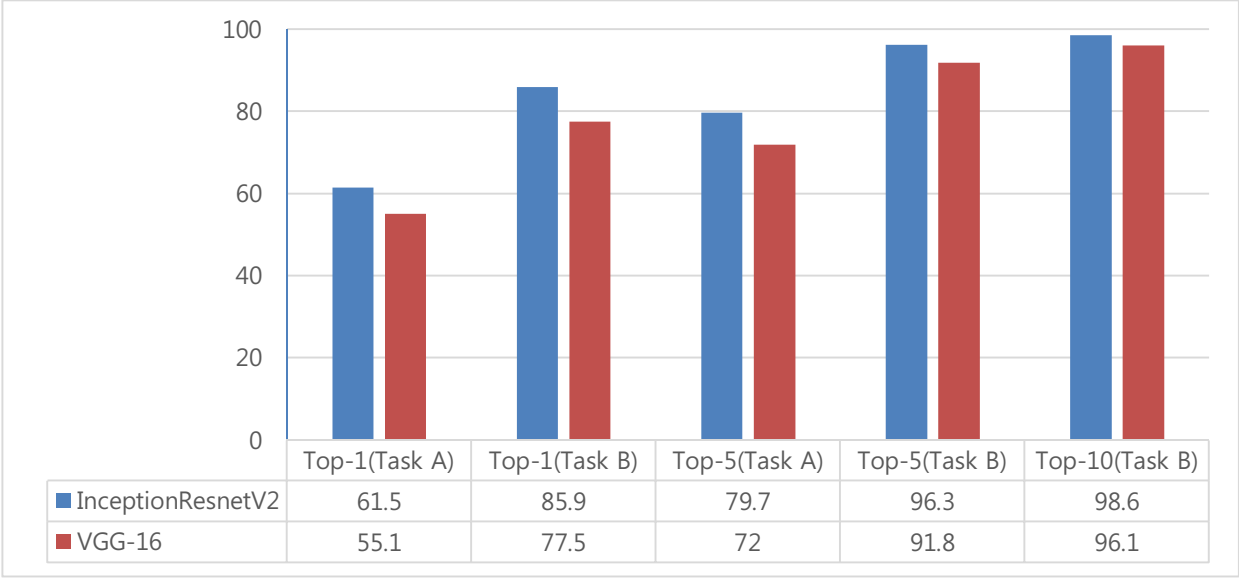
5. Conclusions

This paper conducted an experiment to classify styles based on the deep learning method used in the Computer Vision field. The majority of existing research uses methods for searching and sorting by objects, and it is necessary to select a feature descriptor even when attempting to classify styles, yielding relatively poor results. In order to compensate for this problem, two CNN models were trained using 220 predefined style classes and their performance was compared. Furthermore, in order to investigate learning differences among the models, the filters of a model pre-trained with the ImageNet dataset were compared with the filters trained with the style data, and the performance of the proposed method of employing CNN networks

was validated.

As a result, Inception model has a better performance than VGG model and that was satisfied enough for style classification. Moreover, End-to-end learning all layers of model was better than fine-tuning partially. However, there are not enough publicly-available labeled datasets dedicated to the clipart style. In order to fill this gap, further studies should be carried out with more extensive learning data, whether from additional available clipart datasets or perhaps by creating new datasets.

Table 3. Performance comparison for individual tasks



	Top-1(Task A)	Top-1(Task B)	Top-5(Task A)	Top-5(Task B)	Top-10(Task B)
InceptionResnetV2	61.5	85.9	79.7	96.3	98.6
VGG-16	55.1	77.5	72	91.8	96.1

Table 4. Effect of End-to-End training of CNN models

Model	# of Layers	% of changed weights (after end-to-end training)	Accuracy (after end-to-end training)
Inception Resnet V2	152	17.5%	85.9
VGG-16	16	2.7%	77.5

Acknowledgement

This work was supported by Institute for Information & communications Technology Promotion(IITP) grant funded by the Korea government(MSIT) (No.2017-0-00271, Development of Archive Solution and Content Management Platform)

References

[1] M. F. Barroso, M. J. Fonseca, B. Barroso, P. Ribeiro, and J. A. Jorge, "Retrieving ClipArt Images by C ontent." in *Proceedings of the 3rd International Conference on Image and Video Retrieval*, pp. 500-507, 2

004.

[2] M. J. Fonseca, A. Ferreira, and J. A. Jorge, "Sketch-Based Retrieval of Complex Drawings Using Hierarc hical Topology and Geometry," *Computer-Aided Design*, vol. 41, no. 12, pp. 1067-1081, 2009.

[3] P. Sousa and M. J. Fonseca, "Geometric Matching for Clip-Art Drawing Retrieval," *Journal of Visual Co mmunication and Image Representation,* vol. 20, no. 2, pp. 71-83, 2009.

[4] P. Martins, R. Jesus, M. J. Fonseca and N. Correia, "Clip Art Retrieval Combining Raster and Vector M ethods," in *Proceedings of 11th International Workshop on Content-Based Multimedia Indexing,* pp. 35-40, 2013.

[5] E. Garces, A. Agarwala, D. Gutierrez, and A. Hertzmann, "A Similarity Measure for Illustration Style," *A CM Transactions on Graphics,* vol. 33, no. 4, 2014.

[6] B. Saleh, M. Dontcheva, A. Hertzmann, and Z. Lui, "Learning Style Similarity for Searching Infographics, " in *Proceedings of the 41st Graphics Interface Conference,* pp. 59-64, 2015.

[7] T. Furuya, S. Kuriyama, and R. Ohbuchi, "An Unsupervised Approach for Comparing Styles of Illustratio ns," in *Proceedings of 13th International Workshop on Content-Based Multimedia Indexing,* pp. 35-40, 201 3.

[8] S. Karayev, M. Trentacoste, H. Han, A. Agarwala, T. Darrell, A. Hertzmann, and H. Winnemoeller, "Rec ognizing Image Style," in *Proceedings of British Machine Vision Conference*, no. 121, 2014.

[9] S. Bell and K. Bala, "Learning Visual Similarity for Product Design with Convolutional Neural Network s," *ACM Transactions on Graphics,* vol. 34, no. 4, 2015.

[10] C. Kuo, Y. Chou, and P. Chang, "Using Deep Convolutional Neural Networks for Image Retrieval," in *Proceedings of Visual Information Processing and Communication VII*, pp. 1-6, 2016.

[11] J. Wan, D. Wang, S. Hoi, P. Wu, J. Zhu, Y. Zhang, and J. Li, "Deep Learning for Content-based Imag e Retrieval : A Comprehensive Study," in *Proceedings of 22nd ACM Transactions on Multimedia*, pp. 15 7-166, 2014.

[12] E. Garces, A. Agarwala, A. Hertzmann, and D. Gutierrez, "Style-Based Exploration of Illustration Datase ts," *Multimedia Tools and Applications*, vol. 76, pp. 13067-13086, 2017.

[13] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning." in *Proceedings of the 31st Association for the Advancement of Artifici al Intelligence* , pp. 4278-4284, 2017.

[14] K. Simonyan, "Very Deep Convolutional Networks for Large-Scale Image Recognition," *arXiv preprint a rXiv: 1409.1556*, 2014.

[15] B. Chu, V. Madhavan, O, Beijbom, J. Hoffman, and T. Darrell, "Best Practices for Fine-tuning Visual Classifiers to New Domains," *in Proceedings of the European Conference on Computer Vision 2016*, pp. 435-442, 2016.

[16] A. Poernomo and D. Kang, "Content-Aware Convolutional Neural Network for Object Recognition Tast," *Internat ional Journal of Advanced Smart Convergence*, vol. 5, no. 3, pp. 1-7, 2016.

[17] OpenClipart Library. [online]. Available : https://openclipart.org

[18] J. Donahue, Y. Jia, O. Vinyals, J. Hoffman, N. Zhang, E. Tzeng, T. Darrell, and T. Eecs, "DeCAF : A Deep Convolutional Activation Feature for Generic Visual Recognition," in *Proceedings of the Internatio nal Conference on Machine Learning*, pp. 647-655, 2014.

[19] A. Krizhevsky, I, Sutskever, and G. E. Hinton, "Imagenet Classification with Deep Convolutional Neural Network," in *Proceedings of the 25th International Conference on Neural Information Processing Systems,* pp. 1097-1105, 2012.

[20] G. Chechik, V. Sharma, U. Shalit, and S. Bengio, "Large Scale Online Learning of Image Similarity Th rough Ranking," *Journal of Machine Learning Research*, vol. 11, pp. 1109-1135, 2010.

[21] P. Donova, J. Libeks, A. Agarwala, and A. Hertzmann, "Exploratory Font Selection Using Crowdsourced Attributes," *ACM Transactions on Graphics,* vol. 33, no. 4, 2014.

[22] B. Kulis, "Metric Learning: A Survey," *Fondations and Trends in Machine Learning*, vol. 5, no. 4, pp. 2 87-364, 2013.

[23] C. Szegedy, W. Lui, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabin ovich, "Going Deeper with Convolutions," in *Proceedings of the IEEE Conference on Computer Vision a nd Pattern Recognition* , pp. 1-9, 2015.

[24] B. Kim, "Combining Empirical Feature Map and Conjugate Least Squares Support Vector Machine for Real Time Image Recognition: Research with Jade Solution Company," *International Journal of Internet, Broadcasting and Communication*, vol. 9, no. 1, pp. 9-17, 2017.