Photo Selection for Family Album using Deep Neural Networks

Sijie Shen
The University of Tokyo
Tokyo, Japan
shensijie@hal.t.u-tokyo.ac.jp

Michi Sato
Chikaku Inc.
Tokyo, Japan
michisato@chikaku.co.jp

Toshihiko Yamasaki
The University of Tokyo
Tokyo, Japan
yamasaki@hal.t.u-tokyo.ac.jp

Kenji Kajiware
Chikaku Inc.
Tokyo, Japan
kajiware@chikaku.co.jp

ABSTRACT

The development of digital cameras and the web booming are the critical reasons for the increasing of digital portraits. However, such kind of daily photos are usually too many to select and organize, which leads to further requirement of better photo management services. In this paper, we are focusing on a significant part of daily photos – family photos. We collaborate with a family photo service provider, Chikaku¹ Inc., to create a family photo dataset. The dataset contains 12,140 images with corresponding rates (from 1 to 5) measuring if it is suitable to be selected for a family album. According to our experiments on classifying eligibility of the photos as a printed family photo album, the classification accuracy reaches 96.6% on the test set.

CCS CONCEPTS

• Computing methodologies → Visual content-based indexing and retrieval;

KEYWORDS

Image selection, family photo

ACM Reference Format:

Sijie Shen, Toshihiko Yamasaki, Michi Sato, and Kenji Kajiware. 2018. Photo Selection for Family Album using Deep Neural Networks. In *MMArt&ACM'18: International Joint Workshop on Multimedia Artworks Analysis and Attractiveness Computing in Multimedia, June 11, 2018, Yokohama, Japan*. ACM, Yokohama, Japan, 6 pages. <https://doi.org/10.1145/3209693.3209697>

1 INTRODUCTION

Since the first consumer digital cameras came into the market in late 1990s, digital images have become one of the most important media in our modern life. The digitized images are

¹<https://www.chikaku.co.jp/>

This work is licensed under a Creative Commons Attribution-NonCommercial International 4.0 License.

MMArt&ACM'18, June 11, 2018, Yokohama, Japan
© 2018 Copyright held by the owner/author(s).
ACM ISBN 978-1-4503-5798-2/18/06.
<https://doi.org/10.1145/3209693.3209697>

Figure 1: Samples selected from our created family photo dataset. The first row are three original images in dataset, belows are the rating distributions for each image.

widely accepted by its easiness to use and comparable image quality to traditional photographs. Even photography professionals gradually turned to digital because using such digital files fulfills the demands of employers or clients from photographic work. The first boom of digital portraits came with the growth of digital cameras as well as the development of digital devices (computers, printers, etc.). As the digital images increase, the traditional photo album management may not be suitable for collecting all taken images. Furthermore, because of the growth of web services and related techniques, digital album services are becoming even more popular for better security, organization, and ease of use.

The requirement of better photo management results in various photo services. One promising photo management service is the one shipped with a mobile phone. Such a service is becoming smarter by utilizing image geographical information as well as the techniques of face recognition. For example, the Apple Photo application² provides a bunch of

²<https://support.apple.com/en-us/HT206186>

Figure 2: The histogram of the number of rated people.

handy services for better photo management: the basic cross-device cloud sharing service, utilization of additional information (geographical and temporal information), face detection and recognition, and automatically selecting pictures for making summary videos, etc. Besides Apple's Photo application, there are also services from Google (google album³), Microsoft (Photo application⁴) as well as Facebook (a face recognition service to find link with other people⁵).

Thus there would be a large potential market of a smarter photo album management service benefited from the latest computer vision researches. However, in order to make use of deep neural networks (the most promising technology in the computer vision field), a good training dataset is required. As the photos are usually personal, there are few family photo datasets available. We collaborate with Chikaku Inc.⁶, focusing on the family photo in personal photos, and formed a dataset for family photos. Samples are shown in Figure 1. We rated photos from the view point of suitability for family photo albums. A convolutional neural network (CNN) model using this dataset achieved 96.6% accuracy for the task of selecting good photos for family photo albums.

The rest of this paper is organized as follows. We introduce related works on managing family photos in Section 2. The details of the collected dataset are shown in Section 3. Then a simple binary classification is performed on the collected dataset. The detailed process is introduced in Section 4. We perform classification on the collected dataset in Section 5. Finally concluding remarks are given in Section 6.

2 RELATED WORKS

There are researches trying to build a family photo classifier to select out family photos. For example, Chen et al. [2] tried to select photos based on their proposed Bag-of-Face

³<https://photos.google.com/albums>

⁴<https://support.microsoft.com/en-us/help/4026222/windows-how-to-use-albums-in-the-photos-app>

⁵<https://newsroom.fb.com/news/2017/12/managing-your-identity-on-facebook-with-face-recognition-technology/>

⁶<https://www.chikaku.co.jp/>

Figure 3: The averaged score distribution of all rated images.

subgraph (BoFG). The dataset they used was the photos collected from Flickr⁷ by Andrew et al. in [4]. The collected dataset contained 28,231 people together with human labeled face attributes (age and gender) and categorized to family images, group images and wedding images. The BoFG was a graph structure constructed from age, gender, and location information from images. The idea of BoFG is finding out if a photo contains representative subgraph: it would be more likely to be a family photo if it contains a couple and a child, and would be less likely to be a family photo if it contains a lot of adult pairs. Furthermore, Choia et al. [3] improved the framework of BoFG, by adding TF-IDF normalization part to the computed histogram of BoFG.

There are also researches focusing on generating good photo album. Myodo et al. [6] developed a system for better presenting wedding photos album. They first generated a lot of album layout templates, and used effective algorithms [7, 8] for mining generated layout templates. Then important regions were extracted using face detection techniques and saliency information for each images. Finally they computed penalty score for each template and chose the best template.

These works are all useful for creating a better album. But considering the variety of personal photos and the the number of them, it would be more wise to filter out non-target part of photos which usually consumes great part of personal images. We support this hypothesis during the process of collecting our dataset.

3 COLLECTED DATASET

We are collaborating with Chikaku Inc. – a company aiming to providing photo sharing services for a family between different generations. The collecting process is performed on the provided images from Chikaku. The usage of the photos for our research purpose was approved by the photo owners. The total number of collected images is 58,865, among which 12,140 images are rated by human for measuring how much the image is suitable for a printed family photo album. The

⁷<https://www.flickr.com/>

Table 1: User correlation of collected data rating. “Nan” means there were no overrapping photos between the two raters.

Rater ID	1	2	3	4	5	6	7	8	9	10	11	12	13	Avg.
1	1.00	0.80	0.82	0.79	0.84	0.82	nan	0.80	0.80	0.77	0.84	0.81	0.74	0.80
2	0.80	1.00	0.84	0.81	0.83	0.79	0.48	0.79	0.75	0.79	0.79	0.77	0.69	0.76
3	0.82	0.84	1.00	0.91	0.93	0.86	0.68	0.89	0.84	0.92	0.88	0.84	0.74	0.85
4	0.79	0.81	0.91	1.00	0.91	0.85	0.18	0.86	0.80	0.87	0.87	0.83	0.74	0.79
5	0.84	0.83	0.93	0.91	1.00	0.86	0.50	0.87	0.82	0.88	0.88	0.84	0.76	0.83
6	0.82	0.79	0.86	0.85	0.86	1.00	0.60	0.82	0.79	0.84	0.84	0.80	0.74	0.80
7	nan	0.48	0.68	0.18	0.50	0.60	1.00	0.83	0.56	nan	0.53	0.54	0.59	0.55
8	0.80	0.79	0.89	0.86	0.87	0.82	0.83	1.00	0.79	0.84	0.84	0.82	0.72	0.82
9	0.80	0.75	0.84	0.80	0.82	0.79	0.56	0.79	1.00	0.81	0.80	0.77	0.69	0.77
10	0.77	0.79	0.92	0.87	0.88	0.84	nan	0.84	0.81	1.00	0.85	0.86	0.74	0.83
11	0.84	0.79	0.88	0.87	0.88	0.84	0.53	0.84	0.80	0.85	1.00	0.83	0.75	0.81
12	0.81	0.77	0.84	0.83	0.84	0.80	0.54	0.82	0.77	0.86	0.83	1.00	0.71	0.79
13	0.74	0.69	0.74	0.74	0.76	0.74	0.59	0.72	0.69	0.74	0.75	0.71	1.00	0.72

Figure 4: The variance distribution of all rated images. **Figure 5: The variance distribution of all rated images after thresholding.**

rating is performed by 13 different people including 3 female and 10 male with the age ranged from 28 to 45. The rated score is integer and ranged from 1 to 5, with 1 for not suitable for family photo, 5 for perfectly suitable. Each image is rated by at least five different people, the histogram of the number of rated people is shown in Figure 2.

Since the dataset is collected from users' personal photo libraries, it is worth to find out how much part of the images in libraries is good for family photos. Thus we compute the average score for each images and order them by the average score. The score distribution of all rated images is shown in Figure 3. From the distribution figure, we can see that a large part of images are not appropriate for family photos at all (with score of 1).

Table 2: Confusion matrix of predicting suitability for family photo album on test set.

N=1025	Predicted:Yes	Predicted: No
Actual:Yes	236	19
Actual:No	16	754

Also since the images are rated by different people, there must be some disagreements. We plot the score variance distribution figure for better understanding. The variance distribution along with average score is shown in Figure 4. From the variance distribution figure, we could see that (1) there is only a small part of images that have large variance, and (2) the images with large variance are generally have higher average score.

4 SELECTING PHOTOS FOR PRINTED ALBUM

As shown in Figure 4, there are only a relatively small number of images that are rated with a relatively large variance. It would be a good idea to clean the dataset before using them. Here we performed two different cleaning strategies: filter out the large variance images and filter out outlier users. We firstly set the threshold the variance to 1.5 in our experiments (excluding images with variance score larger than 1.5). The filtered variance distribution is shown in Figure 5.

Secondly, we compute the rater correlation between all rater pairs from their rating scores. The rater correlation score of all rater pairs is shown in Table 1. We compute the average correlation score for each rater, and find that the rater 7 and rater 13 is a little bit out of the main raters. Our following experiments are performed on the dataset without rater 7 and 13's labels because it can be regarded that these raters did not conduct the rating task seriously.

After the dataset cleaning, there are 11,064 images remaining. The task is binary classification for selecting images suitable for printed family album. We divide the images with score equal to 1.5 as not suitable to family photo album class, and images with score greater or equal to 3.0 as suitable to family album class. The remaining part was eliminated from the experiments. The training/evaluation/test division is performed as 80%-10%-10%. We tried different neural networks models, and the ResNet-101 [5] works the best. We use a Stochastic Gradient Descent (SGD) optimization algorithm [1] with 0.001 learning rate and 0.0001 weight decay rate.

5 FAMILY ALBUM CLASSIFICATION

We test our trained neural network on the test set which contains 1025 images. The final classification accuracy is 96.6% and the prediction confusion matrix is shown in Table 2.

The accuracy shows the trained neural network works properly on the classifying images. Furthermore, to explore what the neural network learned to see where the neural network exactly looks at, we extract Class Activation Maps (CAMs) [9] for the classified images. The CAMs of correct classified images labeled as suitable for family album are shown in Figure 6, the CAMs of correct classified images labeled as unsuitable for family album are shown in Figure 7, the CAMs of miss classified images are shown in Figure 8.

From the CAMs listed above, we can see that the trained neural network is generally looking at child as well as the interaction with other people (dad or mom). Note that the network successfully classifies the images into positive class (suitable for family album) no matter how the baby and child's postures change (from side face, lying down, and image rotated). Also for the negative class images (unsuitable for family album), our CNN model is focusing on a particular position like foods on tables, long hair when a single female appears, etc. From the mis-classified CAMs, we could see the neural network makes mistakes when the appeared person is

Figure 6: Class Activation Maps (CAMs) of correctly classified images labeled as suitable for family album.

young without showing long hair (regarded as a mom), or close connection between two persons (regarded as a family).

6 CONCLUSION AND FUTURE WORK

In this paper, we focused on family photos. Due to the image privacy, there are few public datasets containing family photos. Thus we collaborate with Chikaku Inc. to create a family photo dataset containing 11,064 images with subjective scores.

We performed training on collected dataset for classifying photos suitable and unsuitable for family album, and got

Figure 7: CAMs of correctly classified images labeled as unsuitable for family album.

96.6% accuracy on the test set. Since the task was classification, it was possible to visualize the important area for classifying. We computed CAMs on the classified images shown in Figures 6, 7, and 8. The CAMs showed that the trained neural network actually learned efficient feature for family photos.

Since the network we trained generally filter out the part that apparently unsuitable for family album, the number of remaining images continues to be remarkably large. So there remains further room for selecting well-shot, attractive photos from the remaining images.

ACKNOWLEDGMENTS

This work was partially supported by the Grants-in-Aid for Scientific Research (no. 26700008) from JSPS, JST-CREST (JPMJCR1686), and Microsoft IJARC core13.

REFERENCES

- [1] Léon Bottou. 2010. Large-scale machine learning with stochastic gradient descent. In *Proceedings of COMPSTAT*. 177–186.
- [2] Yan-Ying Chen, Winston H Hsu, and Hong-Yuan Mark Liao. 2012. Discovering informative social subgraphs and predicting pairwise

Figure 8: CAMs of mis-classified images.

- relationships from group photos. In *Proceedings of the 20th ACM International Conference on Multimedia (ACMMM)*. 669–678.
- [3] Changmin Choi, YoonSeok Lee, and Sung-Eui Yoon. 2016. Discriminative subgraphs for discovering family photos. *Computational Visual Media* 2, 3 (2016), 257–266.
- [4] Andrew C Gallagher and Tsuhan Chen. 2009. Understanding images of groups of people. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 256–263.
- [5] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep Residual Learning for Image Recognition. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- [6] Emi Myodo, Satoshi Ueno, Koichi Takagi, and Shigeyuki Sakazawa. 2011. Automatic comic-like image layout system preserving image order and important regions. In *Proceedings of the 19th ACM International Conference on Multimedia*. 795–796.
- [7] Takamasa Tanaka, Kenji Shoji, Fubito Toyama, and Juichi Miyamichi. 2007. Layout analysis of tree-structured scene frames

- in comic images. In *Proceedings of the 20th International Joint Conference on Artificial Intelligence*. 2885–2890.
- [8] Mohammed J Zaki. 2002. Efficiently mining frequent trees in a forest. In *Proceedings of the eighth International Conference on Knowledge Discovery and Data mining*. 71–80.
- [9] B. Zhou, A. Khosla, Lapedriza. A., A. Oliva, and A. Torralba. 2016. Learning Deep Features for Discriminative Localization. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2016).