International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 22 (2018) pp. 15734-15739

Generative Adversarial Network-Based Face Recognition Dataset Generation

Sooyoung Cho[1], Daeyeol Kim[1], Sinwoo Yoo[1], and Chae-Bong Sohn [1]

[1]Department of Electronics and Communications Engineering, Kwangwoon University
20, Gwangun-ro, Nowon-gu, Seoul, 01897, Republic of Korea.

Abstract

Facial recognition technique has many advantages than other biometric recognition solutions and recent studies and outcomes of automation process level almost the same as a human does. Applying Deep-Learning solution in this region is very common these days, but there are many obstacles to put in. This paper deals one of them of which the preparation of a certain scale of a dataset by combining existing dataset and another dataset this paper suggests. Celeb A and 2nd version of VGG face dataset are the base dataset that the discriminator agent of Generative Adversarial Network can be trained, and the generator will refer the new dataset with thousands of western portraits we added. This suggested new dataset is tested with Deep-face network as the one of existing facial recognition solutions, and we confirmed that we can use this technique for other similar dataset preprocessing layers. There are some facts to need to consider when it applies to other targets, as analyzed differences between the real facial pictures and the ones was generated.

Keywords: Generative Adversarial Networks, Pattern recognition, Histogram, Deep learning, Deep face

INTRODUCTION

Facial recognition technology is one of decision-making solution that utilizes various numerical information that is obtained from a surface of a human face or other medium that has similar information. Previously applied methodologies had been behind of unresolved obstacles such as a low practical accuracy of the recognition, or the processing time that's longer than the expected. The improvement trend of the overall computing power and current neural network pushes the boundaries in recent. A powerful and efficient infrared sensor product being commercialized in the modern smart gadget industry gathers way much larger scale of facial information quantitively with better quality than any other days. Facebook, one of the biggest SNS service company is developing this technology that's able to recognize the each of individual whom a customer wants to search and sort of among the uploaded pictures, and its accuracy marked over 97% recently[1]. This technology is being developed and focused on among the other companies and industries as well.

Most popular algorithms for the recognition purpose of classifying given human faces as examples above are chosen with YOLO[2], and Deep-Face. They show considerable results through the supervised-learning process usually, and the better performance requires quantitatively bigger datasets in general. The preparation process for datasets needs lots of time, including many manpower that follows.

The supervised-learning process is affected by not only the structure of the network model but also the dataset being used for. The domain such as this, visual recognition, tends to derive a result which was processed by the specific feature of spatial information from the dataset. Hence, the composing dataset which consists of not biased images in order to expect a proper recognition processing.

Deep-face algorithm uses the color information from the dataset of prepared 3-dimensional modeled human faces those were captured before. And obtaining 3-dimensional information from actual human faces require a significant amount of time to prepare. This is where we suggest the way of generation of this kind of datasets.

This paper deals the way of generating a new dataset combining Celeb A[3], VGG face dataset ver. 2[4], and well-known western portraits from various artists. This paper also proceeds a proper analysis by using a specific facial analysis algorithm and a distribution of histogram among its dataset.

A. The goal and used methodology

This paper targets to generate a dataset which is able to be utilized for other Deep learning models besides the existing facial recognition purpose, e.g., Deep-face model. Overall process of this generation of the target dataset will on the Generative Adversarial Networks (GAN)[5] to produce all concluded variations of outcome, by which based on a generator and discriminator. We preprocessed the input sources of the dataset as categorized and scale-normalized since this network may lose its efficient processing its target dataset generation regarding the variance of image sources those be too large. We ignored the specific coordinates information of "face" among each of image sources as for the major element, they were fixed the coordinates of faces among the image sources. So, the learning process of GAN can be done with only the color information from each RGB-pixels as skipping the necessary process for convolution layers to signify its location among the image sources.

Generated image models by GAN processing is being done by two steps. First one that is called as "parametric" refers the outline information from each of the source images, and then filling it as the complete image as the second one that is called as "Non-parametric". The GAN image processing will iterate this process as many as we want in order to complete the image dataset generation.

International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 22 (2018) pp. 15734-15739

Before to put this interim image sources, we preprocessed to enhance the overall image quality by Laplacian parameter extending. Re-classification process comes after, via recurrent and de-convolution network process with these enhanced images after, in order to avoid increasing noise factor during GAN processing.

B. Foundation

Recent trend of utilizing this pattern recognition technique focuses on a machine-visioning business, scene-recognition automation, or material categorization and pattering in these days. The main target of this region can be named as maximizing a pattern classification accuracy.

Pattern recognition is one of the regions under the machine learning methodology, which can be achieved by a supervised that's having the certain label for each item from the dataset to classify them, or, a non-supervised one that categorizes all items from the given dataset without specific classification label.

Supervised learning process are up to specific characteristics the dataset has, each of item has its own classified label by which classified the operator.

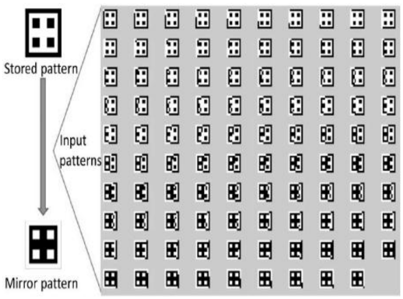

Figure 1: The Stored 10 by 10 pattern and an example of the input patter set

Pattern recognition would be considered as a set consisting of the specific characteristics of each item among the given dataset that induced as an input vector of the learning process, and it classifies them depending on spatial analysis, such as the distances, covariances, or other types of classification factors. The learning process regards the linearity and non-linearity depending on the type of dataset.

C. Collecting Data

As the mentioned explanation, the overall scale of a dataset is one of that determines the result of the learning process. Fortunately, the recent trend of rising the computing power and its cost are being down, which helps us to prepare the certain and required scale of dataset with practical manners.

Crawling and scraping the resources from internet is one of common method that are being used widely. Crawling includes the way of duplicating the content itself from web-

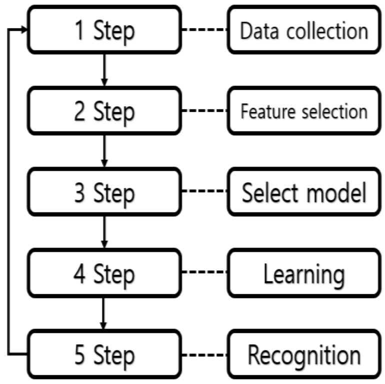

Figure 3: Experiment progress

pages, and scraping requires many ways of parsing the specific resources that you'd like to obtain. Two methods that mentioned above are usually being used as a combined manner.

D. Generative Adversarial Network

GAN image processing includes a supervised learning, by which GAN uses for the Generator network. Various input resources bring the result better and more accurate as long as each of labels are clearly indicates the answer that's expected.

Most of the published and open-sourced techniques of Deep Learning are being grounded in these days based on well-known parameters and environment which is used for the learning process. But as long as the entire classification is based on the answer tags from which human gave already, the result can be easily limited by given data itself. This makes some regions that we don't know exactly what we can expect

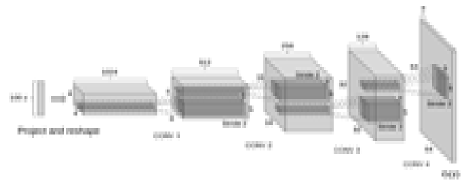

Figure 4: Deep Convolution Generative Adversarial Network

Figure 2: The process of character creation thought the learning of two models

by the result with use other way of it as the non-supervised learning. Combining the learning process with these both of ways lead GAN processing is unique. The Generator keeps suggesting its outcome to the discriminator which has the actual labelled dataset resources to help it to make better classification between the given answer and artificially generated outcome. In the case of this paper used and utilized, each of RGB pixels actually is composing something that discriminator can't define whether it's one of given image resources that's labeled at the first time by the learned and

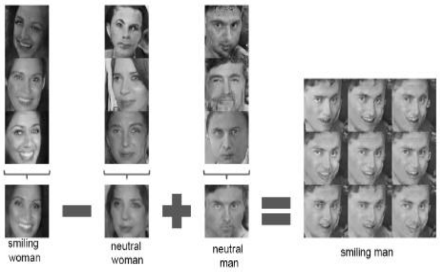

Figure 5: Vector arithmetic for visual concepts

shaped GAN[6]. The produced images via GAN processing are within the most characteristics from the original image dataset which was used for the resources[7].

This is one of the most unique features of GAN processing. Since each of the two agents keeps deliver their own outcomes of opposite trends, which keep overall weights and internal parameters of inside their networks away from mean-value merging or zero-value vanishing issues.

Generator model generates actual images based on RGB color information, and discriminator performs compare between images that injected as original image resources and images that was delivered from the generator model. The discriminator model learns the internal network itself via the

Figure 6: The process of character creation thought the learning of two models

supervised deep learning network, the determination process makes the generator producing its outcome of which having the maximum likelihood distribution that's from the original dataset by which the discriminator provided.

E. SIFT (Scale Invariant Feature Transform)

Scale Invariant Feature Transform algorithm which is known as SIFT is first presented by David G Lowe in 1999, which suggested a solution regarding the sensitivity issue that's caused by changing image scale on the previous algorithm that was suggested by Harris Corner before.

SIFT algorithm performs very robust regarding the image scale sensitivity issue in order to pick the "key-points" among various image resources which have many transforms within the certain covariance. Since of its unique methodology of finding and applying key-points which of scale-invariant into

the patch process, it's able to be utilized to recognize images regardless 4 factors that most of the image dataset has various size, brightness, rotation, and noise in general[8].

Instance Applications

Many deep-learning networks are being used for its own purpose, Deep-face is one of them for the facial recognition. It doesn't have the CNN preprocessing layer as like most other facial recognition networks. Deep-face has its own preprocessing that generalizes input images as the dataset in order to make the input vector can't be interfered by any other external environment factors[9].

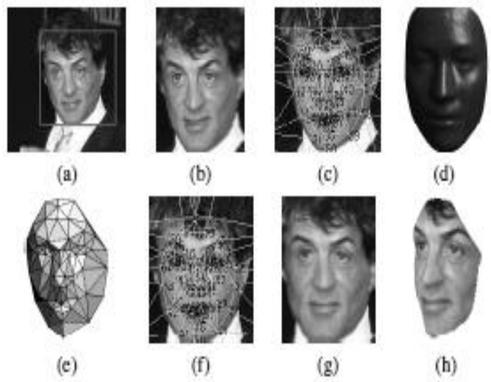

Figure 7: Result from feature points and 3D modeling process

Deep-face doesn't just use the raw image that has given as input vector, but yet it extracts major key-points from faces on the images and reconstruct its own way to obtain parametric values from them. Already the overall accuracy of the facial recognition of Deep-face achieved over 90% [10] [11], even it's very close to the same performance as human's.

Image dataset expansion was done by GAN processing which performed in an effective manner in this paper, massively produced and prepared facial image dataset could be very useful to other facial recognition algorithms for their performance benchmarks.

Experiment

A. A Data model

We set the original dataset with Celeb A, version 2 of VGG face dataset, and about thousands of portraits images those have been crawled and scrapped from internet of which produced from 13th century as western arts. We put the crawling margin as enough to human can recognize the face from it.

Figure 8: Asian face dataset

Figure 9: Guido Reni (1575 ~ 1642)

B. Training Resources and Environment

We didn't crop the faces from the raw images to give more variables in the test environment, but this was applied after enough amount of the learning process to recognize actual faces with raw dataset consist of Celeb A and VGG face dataset ver. 2 without western portrait arts additionally added.

C. Test Environment

Applied hyperparameters are indicated as the table below. Epoch size is 50k for the raw dataset, and 100k for entire dataset that was merged with additional western portraits arts. These parameters were fixed for all training processes in order to eliminate possible confounding variables. The size of input image set was fixed by 178 pixel by 218 pixel.

Table 1: Hyperparameters

parameter	value
learning rate	0.0002
batch size	64
beta1	0.5 (momentum term of adam)

D. N-step training

When the g_loss value is above number, we confirmed there's no facial image or possible object vector that human can recognize. Each step of the learning was shown in the following table 3. Loss values from each loss functions which are the discriminator and the generator. ***g_loss*** indicates the loss function value from the generator that has used, and ***d_loss*** does of the discriminator as well. GAN produced outcomes that we can possibly recognize as human facial images from them after 50k iterations, and further iterations gave us more sophisticated colors and saturation on each facial image.

Table 2: Learning to n-step

step	d_loss (average)	g_loss (average)
10K	0.8	7
20K	0.007	5.47
50K	0.0004	2.24
100K	0.00007	1.772

E. Key-points algorithm

We put a probe while the preprocessing layer extracts "key-points" from input vectors in order to make confirmation that the SIFT algorithm can extract them from generated images. Figure 10 represents that newly generated images also similarly are treated as a human face comparing the actual image example on the right.

International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 22 (2018) pp. 15734-15739

Figure 10: Apply SIFT algorithm using output image

F. Analysis of Information distribution

We also analyzed the distribution of information among input vector images by histograms. Analyzed distribution shows that there's a difference between original and generated facial images through the value of the derived probability density function for each class.

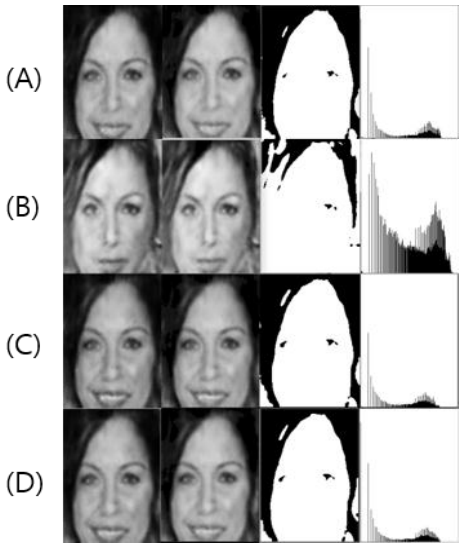

Figure 11: Analyze results using histogram

Table 3: Histogram value

Gray scale value	(A)	(B)	(C)	(D)
2	4337	505	4754	4673
23	3317	617	2976	3084
30	1645	748	1242	1262
41	606	587	638	595
45	466	534	525	575
61	236	299	258	256
78	111	281	164	123
87	92	244	105	100
97	91	218	117	106
114	77	174	109	102
122	111	160	121	149
139	119	164	163	121
148	106	148	184	198
154	166	145	251	181
166	408	253	486	500
179	298	161	238	260
200	118	210	60	135
210	68	153	20	5
212	37	331	3	0
233	0	100	0	0

Table 3 Indicates the each of cases of the histogram as above, 256- colors grayscale was used for this results.

CONCLUSIONS

This paper suggested a dataset consist of massive number of facial images that from actual raw resources and artificially generated one of which a facial recognition algorithm and neural network can utilized as an input dataset. GAN was used for actual generation process with all raw input dataset, Celeb A and VGG face dataset ver. 2.

Overall characteristics of generated images are within the range of the distribution of original image resources as the input vectors. By given images' distribution, outcomes are almost similar with the original but only additional colors and saturation variances and some images from GAN processing has additional objects around of it. Since this suggested dataset is based on the feature as a human facial image, extracted key-points tent to be derived from most of the color of the human face. But already we shown that there are clearly different distributions among the histograms between preprocessed raw image sources and GAN generated facial images.

We assume that the cause of the different information distribution comes from the difference between actual pictures those were taken by a camera, and western portrait arts which are mostly oil-paints done by human hands. Depicting a human face on the canvas is always depending on how a human can perceive them in their recognition process. Human doesn't accept and process entire information which they perceive, but only key-points that is able to be unmasked from most of common vectors among all other faces. Absolute amount of Information that could be included regardless of an actual picture from a camera and portrait paints done by human artist could be the same, but its information distribution and various key-points that could be obtained by machine learning process could be quite different, such as the overall shape of the skull or the jar, forehead, eyebrows, wrinkles, and others more. GAN preprocessing for diversifying the original dataset can help to solidify the variant distribution of input vectors, but we have to consider that there is non-uniformity of information expansion as well.

In the conclusion of this paper, VGG face dataset version 2 and Celeb A facial image dataset are appropriate to utilize together in other purposes in which the diversify the given dataset in this case. And also this result can be referred non-facial image reproduction purpose as building another proper dataset for it as well.

ACKNOWLEDGMENT

This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the ITRC(Information Technology Research Center) support program(IITP-2018-2016-0-00288) supervised by the IITP(Institute for Information & communications Technology Promotion).

REFERENCES

[1] Huang, G. B., Mattar, M., Berg, T., & Learned-Miller, E. (2008, October). Labeled faces in the wild: A database for studying face recognition in unconstrained environments. In Workshop on faces in 'Real-Life' Images: detection, alignment, and recognition.

[2] Redmon, J., & Farhadi, A. (2018). Yolov3: An incremental improvement. arXiv preprint arXiv:1804.02767.

[3] Liu, Z., Luo, P., Wang, X., & Tang, X. (2018). Large-scale CelebFaces Attributes (CelebA) Dataset. Retrieved August, 15, 2018. (http://mmlab.ie.cuhk.edu.hk/projects/CelebA.html)

[4] Parkhi, O. M., Vedaldi, A., & Zisserman, A. (2015, September). Deep face recognition. In BMVC (Vol. 1, No. 3, p. 6).

[5] Huang, G. B., Mattar, M., Berg, T., & Learned-Miller, E. (2008, October). Labeled faces in the wild: A database for studying face recognition in unconstrained environments. In Workshop on faces in 'Real-Life' Images: detection, alignment, and recognition.

[6] Radford, A., Metz, L., & Chintala, S. (2015). Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv preprint arXiv:1511.06434.

[7] Denton, E. L., Chintala, S., & Fergus, R. (2015). Deep generative image models using a Laplacian pyramid of adversarial networks. In Advances in neural information processing systems (pp. 1486-1494).

[8] Lindeberg, T. (2012). Scale invariant feature transform.

[9] Parkhi, O. M., Vedaldi, A., & Zisserman, A. (2015, September). Deep face recognition. In BMVC (Vol. 1, No. 3, p. 6).

[10] Huang, G. B., Mattar, M., Berg, T., & Learned-Miller, E. (2008, October). Labeled faces in the wild: A database for studying face recognition in unconstrained environments. In Workshop on faces in 'Real-Life' Images: detection, alignment, and recognition.

[11] Taigman, Y., Yang, M., Ranzato, M., & Wolf, L. (2014). Closing the gap to human-level performance in face verification. Deepface. In IEEE Computer Vision and Pattern Recognition (CVPR).