A Distinct Discriminant Canonical Correlation Analysis Network based Deep Information Quality Representation for Image Classification

Ryerson Multimedia Research Laboratory

Ryerson University, Toronto, Canada

Lei Gao, Zheng Guo, and Ling Guan

Outline

- Motivation
- The Proposed Method
- Experimental Results and Analysis
- Conclusions

Motivation

- With the advances of computing capability, deep neural network (DNN) based algorithms have greatly promoted state-of-the-art.

- Essentially, generation of the aforementioned deep-level feature representation is accomplished by extracting the abstract semantics of the input data sets with a deep cascade network structure.

- Nevertheless, since the deep-level features are learned from different layers, it is necessary to collect enormous data samples to guarantee that the parameters in the deep layers are able to be tuned successfully.

- Therefore, the learned deep-level feature representation is a data-driven solution, maybe leading to failures on the small scale data sets.

The proposed method-DDCCANet (1)

According to the aforementioned discussion, one potential solution to balancing the small scale and deep-level feature representation is to integrate the multi-view representation and the deep cascade structure effectively.

In this paper, a distinct discriminant canonical correlation analysis network (DDCCANet) based deep information quality representation is proposed for image classification.

DDCCANet possesses three different parts, including DDCCA filters, pooling operation, information quality representation.

The proposed method-DDCCANet (2)

❑ DDCCA filters

- Distinct Discriminant CCA (DDCCA)

DDCCA aims to find the discriminant information by the within-class and between-class correlation matrices across two data sets instead of the scatter matrix, it is able to explore more discriminant representation especially in multi-feature spaces.

$$\arg\max_{\omega_1,\omega_2} \rho = {\omega_1}^T \tilde{C}_{x_1 x_2} \omega_2,$$

subject to

$${\omega_1}^T C_{x_1 x_1} \omega_1 = {\omega_2}^T C_{x_2 x_2} \omega_2 = 1$$

Then Lagrange multiplier and GEV algorithms are utilized to find the solution.

The proposed method-DDCCANet (3)

1) DDCCA Filters: Given a set of images $[I_1, I_2, ...I_M]$, where M is the number of images. There are two-view data sets $[I_1^1, I_2^1, ...I_M^1]$ and $[I_1^2, I_2^2, ...I_M^2]$ with size $p \times q$. To extract the learned deep-level feature representation, first, the given two-view data sets are divided into different patches with the size being $l_1 \times l_2$. Then, we select an $l_1 \times l_2$ patch around each pixel and the vectorized operation is performed on the selected patches. Therefore, all patches collected from the kth sample of the dth ($d \in [1, 2]$) view set are written as the vectorial forms $I_{k,1}^d, I_{k,2}^d, ..., I_{k,pq}^d \in R^{l_1 l_2}$. The patches from all M samples of the dth view set are written as follows

$$I^d = [I_1^{d^*}, I_2^{d^*}, ..., I_M^{d^*}] \in R^{l_1 l_2 \times Mpq},$$

where

$$I_k^{d^*} = [I_{k,1}^d, I_{k,2}^d, ..., I_{k,pq}^d] \in R^{l_1 l_2 \times pq} (k \in [1, 2, ..., M]).$$

The proposed method-DDCCANet (4)

In DDCCANet, since DDCCA is utilized to learn the parameters of the deep network from the input samples to explore discriminant information, I^d is considered as the dth data set with size $l_1 l_2 \times Mpq$, where $l_1 l_2$ is the number of 'dimensions' and Mpq is the number of 'training samples', respectively. Afterwards, the optimization function of DDCCA on I^1 and I^2 is formulated as

$$\arg\max_{\omega_1,\omega_2} \rho = {\omega_1}^T \widetilde{C_{I^1 I^2}} \omega_2,$$

subject to

$${\omega_1}^T C_{I^1 I^1} \omega_1 = {\omega_2}^T C_{I^2 I^2} \omega_2 = 1.$$

The proposed method-DDCCANet (5)

Based on the Lagrange multiplier and GEV algorithms, the solutions ω_1 and ω_2 are with size $l_1 l_2 \times l_1 l_2$. Then, each column of ω_1 and ω_2 is chosen as the DDCCA filter for deep feature representation. Suppose we have L_i filters in the ith layer of the 1th view, the filters are formulated as follows

$$W_g^1 = res_{l_1,l_2}(\omega_{1,g}) \in R^{l_1 \times l_2}, g = 1, 2, .., L_i.$$

Similarly, the filters in the 2th view are formulated as follows

$$W_g^2 = res_{l_1,l_2}(\omega_{2,g}) \in R^{l_1 \times l_2}, g = 1, 2, .., L_i.$$

where $res_{l_1,l_2}(h)$ is a function to reshape $h \in R^{l_1 l_2}$ into a matrix with size $l_1 \times l_2$.

The proposed method-DDCCANet (6)

Therefore, in the ith layer of the dth view, it is capable of obtaining L_i outputs $I_{d,k,g}^{out} = res_{l_1,l_2}(I_k^{d*}) \otimes W_g^d$ from the kth sample, where '$\otimes$' is the 2D convolution operator. Essentially, the aforementioned analysis aims to accomplish the task of 2D convolution by vectors based product operator, which is drawn graphically in Figure. 1.

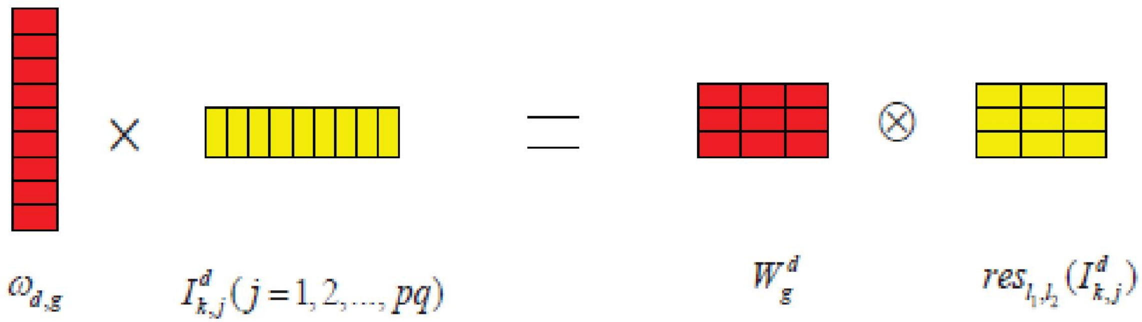


Figure 1. The representation of DDCCANet filters

The proposed method-DDCCANet (7)

2) *Pooling Operation:* Suppose there are $i+1$ layers for the proposed DDCCANet architecture, and the outputs are given as $I_{d,k,g}^{out} \otimes W_\ell^d$ for the kth sample of the dth view. Then outputs are binarized based on Hashing transform in following equation

$$S\{I_{d,k,g}^{out} \otimes W_\ell^d\}(\ell = 1, 2, .., L_{i+1}),$$

where

$$S(x) = \begin{cases} 1, & (x > 0) \\ 0, & others. \end{cases}$$

After that, the vector of ℓ binary bits is considered as a decimal number, resulting in a single integer-valued *"image"*

The proposed method-DDCCANet (8)

3) Information Quality Representation: Based on the previous subsection, each generated image $Q_{d,k,g}$ is partitioned into A blocks. Different from the existing CCANet and related algorithms, information quality (IQ) instead of histogram is employed to generate the deep-level feature representation.

$$H(p(t)) = -log(p(t)),$$

where $p(t)$ is the prior probability of t, the IQ of the decimal values in each block is calculated. Next, IQ values in all the A blocks are transformed into one vector as $H(Q_{d,k,g})$.

The proposed method-DDCCANet (9)

Therefore, in the DDCCANet, the deep level representation of the kth sample in the dth view is written as

$$o_{k,d} = [H(Q_{d,k,1}), \ldots, H(Q_{d,k,L_i})]^T \in R^{(2^{L_{(i+1)}} L_i A)}.$$

Finally, the final deep-level feature representation corresponding to the kth sample by integrating two different views is formulated as following

$$o_k = [o_{k,1}; o_{k,2}] \in R^{(2^{L_{(i+1)}+1} L_i A)}.$$

In summary, the proposed DDCCANet based deep-level feature representation architecture is depicted in Figure. 2.

The proposed method-DDCCANet (10)

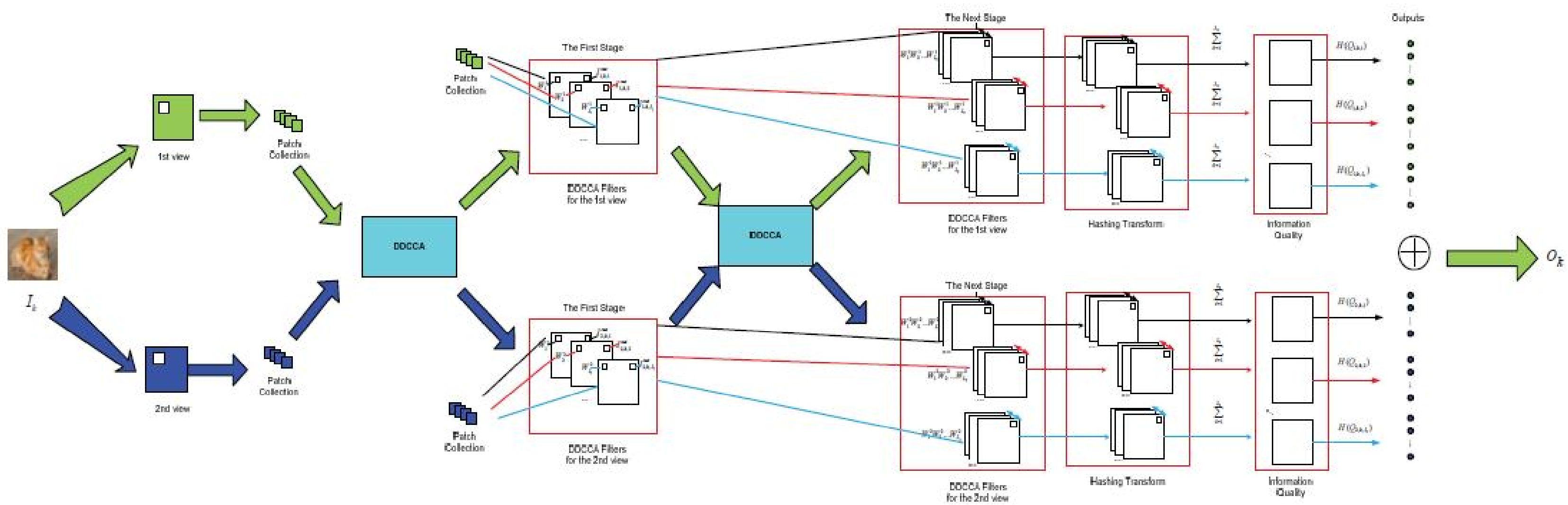


Figure. 2 The diagram of the proposed DDCCANet

Experimental results and analysis(1)

The ORL Database

During our experiments, the local binary patterns (LBP) operation is performed on each sample to generate the second view data set for CCANet and DDCCANet. As a result, the original images and LBP maps are utilized together to accomplish the task of deep-level feature representation for face recognition as shown in Figure. 3.

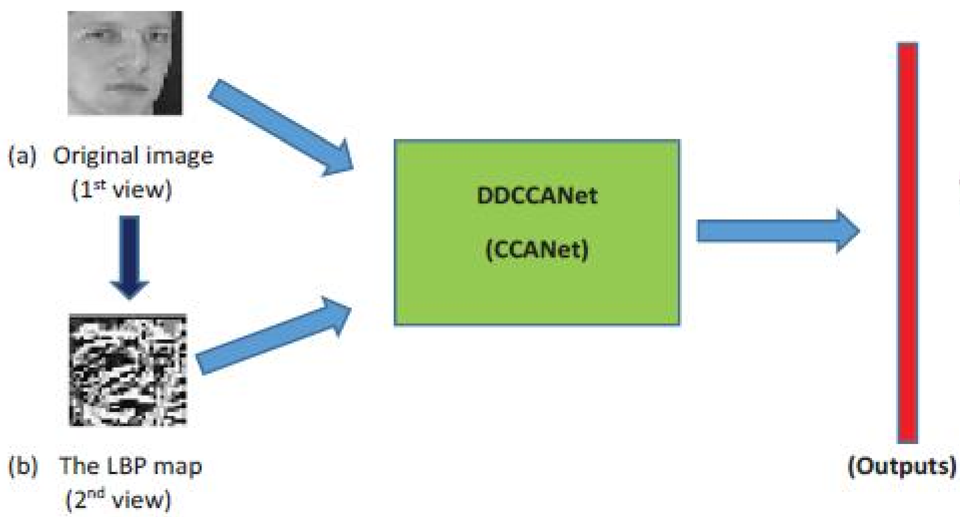


TABLE I
THE PERFORMANCE OF CCANET AND DDCCANET ON THE ORL DATABASE

Methods	Performance
CCANet [10]	96.90%
DDCCANet	**97.50%**

Figure 3. The diagram of face recognition

Experimental results and analysis(2)

TABLE II
THE PERFORMANCE WITH DIFFERENT ALGORITHMS ON THE ORL DATABASE

Methods	Performance
DDCCANet	**97.50%**
AOS+VGG [12]	93.62%
CDPL [13]	95.42%
ANFIS-ABC [14]	96.00%
SOLDE-TR [15]	95.03%
GDLMPP [16]	94.50%
CNN [6]	95.00%
PCANet [17]	96.50%
CS-SRC [18]	96.00%
ANFIS [11]	96.00%
LCCA [19]	95.50%

Experimental results and analysis(3)

ETH-80 Database

In the experiment, we choose 1000 images to construct the training subset and the remaining 2280 images are chosen as the testing samples. All samples are reduced as the size 64*64. The R and G sub-channel images are adopted as the two different views as shown in Figure. 4.

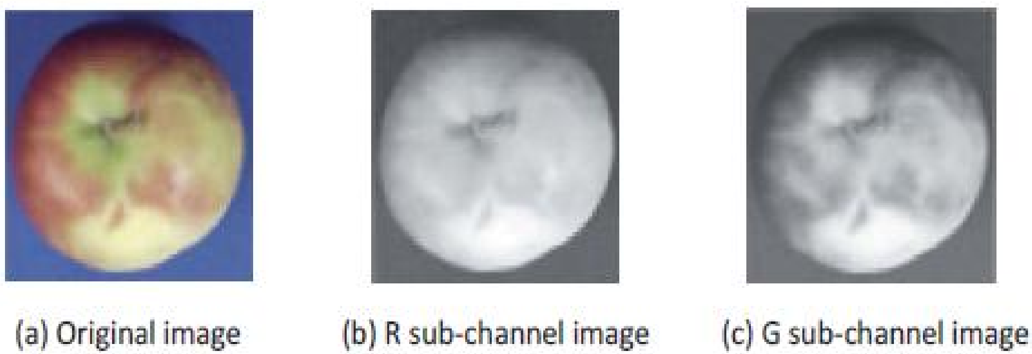


Figure. 4 The original image, R and G sub-channel images in the ETH-80

Experimental results and analysis(4)

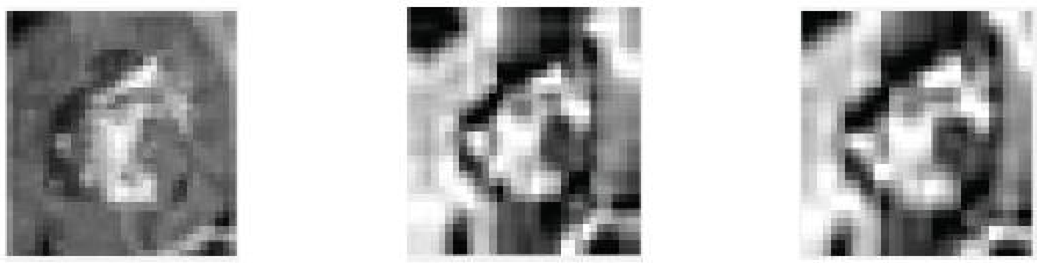
TABLE III
THE PERFORMANCE WITH DIFFERENT ALGORITHMS ON THE ETH80 DATABASE

Methods	Performance
DDCCANet	**91.67%**
CCANet [9]	91.45%
PCANet [17]	91.28%
RandNet-1 [20]	78.50%
RandNet-1 [20]	83.51%
DCC [21]	86.25%
LEML [22]	84.25%
PML [23]	89.00%
SDNN [26]	82.80%
MFD [27]	86.91%
ALP-TMR [28]	84.86%
CERML [29]	85.00%

Experimental results and analysis(5)

CIFAR10 Database

In this paper, we randomly select 10000 images for training and the average performance is reported. The Coiflets orthogonal wavelet transform and Daubechies orthogonal wavelet transform are performed on the original images to generate the two-view samples as drawn in Figure. 5.



(a) Original 'Frog' image (b) The Coiflets wavelet map (c) The Daubechies wavelet map

Figure. 5 The original image, Coiflets and Daubechies orthogonal wavelet maps in the CIFAR10

Experimental results and analysis(6)

TABLE IV
THE PERFORMANCE WITH DIFFERENT ALGORITHMS ON THE CIFAR10 DATABASE

Methods	Performance
DDCCANet	**62.05%**
DCCANet [10]	60.00%
CCANet[9]	53.50%
PCANet [17]	58.01%
DCTNet [24]	56.23%
RandNet [20]	45.11%
LDANet [25]	51.42%
Wide ResNet [30]	60.00%
VGG-16 [31]	56.00%

Conclusions

- This paper presents a DDCCANet with application to image classification.
- To extract more discriminant information between different data sets, the within-class and between-class correlation matrices are employed and optimized jointly.
- Benefiting from the strengths of DDCCANet, it is capable of improving the quality of feature representation from original images.
- Experimental results demonstrate its superior performance on image classification.

References(1)

[1] L. Wei, R. Su, B. Wang, X. Li, Q. Zou, and X. Gao. "Integration of deep feature representations and handcrafted features to improve the prediction of N6-methyladenosine sites." *Neurocomputing*, vol. 324, pp. 3–9, 2019.

[2] R. Venkatesh, N. Maheswari, and S. Jeyanthi. "Multiple criteria decision analysis based overlapped latent fingerprint recognition system using fuzzy sets." *International Journal of Fuzzy Systems*, vol. 20, no. 6, pp. 2016–2042, 2018.

[3] T. Lin, and H. Zha. "Riemannian manifold learning." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 5, pp. 796–809, 2008.

[4] S. Kim, and C.D. Yoo. "Underdetermined blind source separation based on subspace representation." *IEEE Transactions on Signal processing*, vol. 57, no. 7, pp. 2604–2614, 2009.

[5] I. Tosic, and P. Frossard. "Dictionary learning." *IEEE Signal Processing Magazine*, vol. 28, no. 2, pp. 27–38, 2011.

[6] A. Krizhevsky, I. Sutskever, and G. E. Hinton. "ImageNet classification with deep convolutional neural network."*In Advances in neural information processing systems*, pp. 1097–1105, 2012.

[7] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. "Going deeper with convolutions." *2015 IEEE conference on computer vision and pattern recognition*, pp. 1–9, 2015.

[8] X.T. Yuan, X. Liu, and S. Yan. "Visual classification with multitask joint sparse representation." *IEEE Trans. Image Process.*, vol. 21, no. 10, pp. 4349–4360, 2012.

[9] X. Yang, W. Liu, D. Tao, and J. Cheng. "Canonical correlation analysis networks for two-view image recognition." *Information Sciences*, vol. 385, pp. 338–352, 2017.

References(2)

[10] B.B. Gatto, and E.M. Dos Santos. "Discriminative canonical correlation analysis network for image classification." *2017 IEEE International Conference on Image Processing (ICIP)*, pp. 4487–4491, 2017.

[11] M.R. Rejeesh. "Interest point based face recognition using adaptive neuro fuzzy inference system." *Multimedia Tools and Applications*, vol. 78, no. 16, pp. 22691–22710, 2019.

[12] A. Singha, M.K. Bhowmik, and D. Bhattacherjee. "Akin-based Orthogonal Space (AOS): a subspace learning method for face recognition." *Multimedia Tools and Applications (Early Access)*, 2020.

[13] M. Meng, M. Lan, J. Yu, J. Wu, and D. Tao. "Constrained Discriminative Projection Learning for Image Classification." *IEEE Transactions on Image Processing*, vol. 29, no. 1, pp. 186–198, 2019.

[14] M.R. Rejeesh."Interest point based face recognition using adaptive neuro fuzzy inference system." *Multimedia Tools and Applications*, vol. 78, no. 16, pp. 22691–22710, 2019.

[15] X. Yang, G. Liu, Q. Yu, and R. Wang. "Stable and orthogonal local discriminant embedding using trace ratio criterion for dimensionality reduction." *Multimedia Tools and Applications*, vol. 77, no. 3, pp. 3071-3081, 2018.

[16] M. Wan, and Z. Lai. "Generalized Discriminant Local Median Preserving Projections (GDLMPP) for Face Recognition." *Neural Processing Letters*, vol. 49, no. 3, pp. 951–963, 2019.

[17] T. Chan, K. Jia, S. Gao, J. Lu, Z. Zeng, and Y. Ma. "PCANet: A simple deep learning baseline for image classification?." *IEEE transactions on image processing*, vol. 24, no. 12, pp. 5017–5032, 2015.

[18] G. Zhang, H. Sun, Z. Ji, Y. Yuan, and Q. Sun. "Cost-sensitive dictionary learning for face recognition." *Pattern Recognition*, vol. 60, pp. 613–629, 2016.

References(3)

[19] L. Gao, R. Zhang, L. Qi, E. Chen, and L. Guan. "The labeled multiple canonical correlation analysis for information fusion." *IEEE Transactions on Multimedia*, vol. 21, no. 2, pp. 375–387, 2019.

[20] X. Yang, and W. Liu. "Multiple Scale Canonical Correlation Analysis Networks for Two-View Object Recognition." *International Conference on Neural Information Processing*, pp. 325–334, 2017.

[21] T. Kim, J. Kittler, and R. Cipolla. "Discriminative learning and recognition of image set classes using canonical correlations." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 6, pp. 1005–1018, 2007.

[22] Z. Huang, R. Wang, S. Shan, X. Li, and X. Chen. "Log-euclidean metric learning on symmetric positive definite manifold with application to image set classification." *International conference on machine learning*, pp. 720–729, 2015.

[23] Z. Huang, R. Wang, S. Shan, and X. Chen. "Projection metric learning on Grassmann manifold with application to video based face recognition." *IEEE conference on computer vision and pattern recognition*, pp. 140–149, 2015.

[24] C. Ng, and A.B.J. Teoh. "DCTNet: A simple learning-free approach for face recognition." *2015 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA)*, pp. 761–768, 2015.

References(4)

[25] Y. Ge, J. Hu, and W. Deng. "PCA-LDANet: A simple feature learning method for image classification." *2017 4th IAPR Asian Conference on Pattern Recognition (ACPR)*, pp. 370–375, 2017.

[26] S.R. Kheradpisheh, M. Ganjtabesh, S.J. Thorpe, and T. Masquelier. "STDP-based spiking deep convolutional neural networks for object recognition." *Neural Networks*, vol. 99, pp. 56–67, 2018.

[27] C. Yang, and Q. Yu. "Multiscale Fourier descriptor based on triangular features for shape retrieval." *Signal Processing: Image Communication*, vol. 71, pp. 110–119, 2019.

[28] H. Zhang, Z. Zhang, M. Zhao, Q. Ye, M. Zhang, and M. Wang. "Robust triple-matrix-recovery-based auto-weighted label propagation for classification." *IEEE Transactions on Neural Networks and Learning Systems (Early Access)*, 2020.

[29] Z. Huang, R. Wang, S. Shan, L. Gool, and X. Chen. "Cross Euclidean-to-Riemannian metric learning with application to face recognition from video." *IEEE transactions on pattern analysis and machine intelligence*, vol. 40, no. 12, pp. 2827–2840, 2018.

[30] E. Oyallon, E. Belilovsky, and S. Zagoruyko. "Scaling the scattering transform: Deep hybrid networks." *2017 IEEE international conference on computer vision*, pp. 5618–5627, 2017.

[31] K. Simonyan, and A. Zisserman."Very Deep Convolutional Networks for Large-Scale Image Recognition." *2015 International Conference on Learning Representations*, pp. 1–14, 2015.

Thanks!