CLASS-CENTER INVOLVED TRIPLET LOSS FOR SKIN DISEASE CLASSIFICATION ON IMBALANCED DATA

Weixian Lei[1] *   *Rong Zhang*[1,2]*   *Yang Yang* [1,2]   *Ruixuan Wang* [1,2]   *Wei-Shi Zheng*[1,2]

[1] School of Data and Computer Science, Sun Yat-sen University, China
[2] Key Laboratory of Machine Intelligence and Advanced Computing, MOE, Guangzhou, China

ABSTRACT

It is ideal to develop intelligent systems to accurately diagnose diseases as human specialists do. However, due to the highly imbalanced data issue between common and rare diseases, it is still an open problem for the systems to effectively learn to recognize both common and rare diseases. In this paper, we novelly applied triplet modelling to overcome the data imbalance issue particularly for diagnosis of rare diseases. Moreover, we further applied a class-center based triplet loss in order to make the triplet-based learning more stable. Extensive evaluation on two skin image classification tasks shows that the triplet-based approach is very effective and outperforms the widely used methods for solving the imbalance problem.

***Index Terms*—** Data imbalance, triplet loss, medical image classification.

1. INTRODUCTION

Recent advances in deep learning has led to human-level performance on intelligent diagnosis based on medical images [1]. However, it often requires large account of training data to make deep neural networks work well. As a result, current intelligent systems are mainly for the diagnosis of commonly encountered diseases, leaving the intelligent diagnosis of rare disease as an open problem largely unsolved due to the limited available data. To deploy intelligent systems in real-world application, it would be important to make intelligent systems diagnose both common and rare disease as human specialists do. In this study, we aim to overcome the difficulty of training intelligent systems effectively under the condition of data imbalance, i.e., large samples for common diseases and small samples for rare diseases.

In order to solve the data imbalance issue, multiple approaches have been developed particularly to effectively handle the small-sample classes. For example, it has been widely used to oversample the limited data from small-sample classes or downsample the data from larger-sample classes to make training data balanced between classes [2]. Another approach is to set larger coefficient weights to the loss terms related to small-sample classes of training data, penalizing the mis-classification of each training data more heavily when the data is from a small-sample class [3]. Different from setting a single class-specific loss weight for all training data of the same class, another approach is to adaptively set a unique weight for each single training data, with higher weight for the data difficult to be correctly classified. Boosting [4, 5] and the recently proposed focal loss [6] belong to this approach. Besides these traditional approaches, transfer learning by fine-tuning a pre-trained convolutional neural network (CNN) classifier has been proven helpful to improve performance for small-sample classes [7, 8].

Different from the existing approaches, this paper novelly applied the triplet loss [9] to effectively overcome the data imbalance problem. In each triplet, two images are from the same class and the third image is from another class. In this way, the number of training data (i.e., triplets) relevant to small-sample classes can be increased exponentially compared to the number of individual images within the classes. While the number of triplet training data among large-sample classes would also be much more than individual images, the triplet data can be easily balanced between classes by sampling similar number of triplets for each class during model training (note that sampling some triplets rather than using all possible triplets for training is necessary due to the huge number of triplets). Although triplet-based model training has been employed in natural image analysis (e.g., FaceNet [9], ReID [10, 11], triplet-center loss [12]), to the best of our knowledge, no study has investigated the effect of triplet-based model training on data imbalance problems, particularly for the diagnosis of common and rare diseases together.

Another contribution of this study is the novel application of a class-center based triplet loss to further improve the performance on data-imbalanced classification tasks. Specifically, two data within each triplet are from the class centers rather than randomly sampled from individual images. Such

*These authors contributed equally to this work.

This paper was modified since acceptance to include a statement of data type and source (References [14, 15, 16]).

class-center involved triplet can help each image attracted to its class center, thus resulting in more compact distribution for each class of images.

Experimental evaluation on two skin image datasets, both of which are highly data-imbalanced between classes, shows that the triplet-based model training outperforms the widely used approaches such as oversampling, class weighting, and using focal loss, irrespective of model structures. The evaluation also shows that class-center involved triplets can further improves the classification performance.

2. METHODS

2.1. Triplets for data-imbalanced disease classification

In order to develop an intelligent medical image classification system to accurately diagnose both common and rare diseases, we need to effectively train the (e.g., CNN) classifier based on the limited training images especially for rare diseases. The highly imbalanced training data between common and rare diseases often cause biased prediction when applying the trained classifier to the diagnosis of new images. The biased prediction may still exist when applying existing strategies (e.g., over sampling, class weighting, etc.) to handle the data imbalance issue during classifier training, probably because it is difficult to learn to extract the appropriate visual features characterizing the corresponding rare diseases from the very limited data (e.g., 20 images).

Considering the learning process of medical students who learn to find and recognize unique features of one rare disease by comparing the disease with others, embedding such 'comparison' process into model training could help the model more effectively find discriminative features particularly for rare diseases. Let $\mathbf{x}_i$ and $\mathbf{y}_i$ denote two images randomly sampled from the same rare (or common) disease and $\mathbf{z}_i$ sampled from another (rare or common) disease. Here $\mathbf{x}_i$, $\mathbf{y}_i$, and $\mathbf{z}_i$ are often called *anchor*, *positive*, and *negative* respectively. Then, the 'comparison' process can be naturally embedded into model training by enforcing that the visual features of the two images $\mathbf{x}_i$ and $\mathbf{y}_i$ extracted from the model (here a CNN) are more similar than those between the two images $\mathbf{x}_i$ and $\mathbf{z}_i$ from different diseases, i.e.,

$$\|\mathbf{f}(\mathbf{x}_i;\boldsymbol{\theta}) - \mathbf{f}(\mathbf{y}_i;\boldsymbol{\theta})\| + \alpha < \|\mathbf{f}(\mathbf{x}_i;\boldsymbol{\theta}) - \mathbf{f}(\mathbf{z}_i;\boldsymbol{\theta})\| \,, \quad (1)$$

where $\boldsymbol{\theta}$ represents the CNN model parameters to be learned, and $\mathbf{f}(\cdot;\boldsymbol{\theta})$ represents the feature extraction process of the CNN model, with an image as input and a feature vector as output. $\|\cdot\|$ represents L_p norm (here $p = 2$), and α is a positive constant to further enforce such inequality constraint. Based on this inequality constraint, the loss function $l(\mathbf{x}_i, \mathbf{y}_i, \mathbf{z}_i; \boldsymbol{\theta})$ for the CNN model (i.e., feature extractor) can be defined as

$$\begin{aligned} l(\mathbf{x}_i, \mathbf{y}_i, \mathbf{z}_i; \boldsymbol{\theta}) = [\|\mathbf{f}(\mathbf{x}_i;\boldsymbol{\theta}) - \mathbf{f}(\mathbf{y}_i;\boldsymbol{\theta})\| + \alpha \\ - \|\mathbf{f}(\mathbf{x}_i;\boldsymbol{\theta}) - \mathbf{f}(\mathbf{z}_i;\boldsymbol{\theta})\|]_+ \,, \end{aligned} \quad (2)$$

where $[d]_+ = \max(0, d)$ is the hinge loss. In Equation (2), the three images form a triplet $(\mathbf{x}_i, \mathbf{y}_i, \mathbf{z}_i)$ which is used as a single training data for the loss function. Note that while the number of individual training images for one rare disease are often small, the number of triplets which includes at least one training image from the rare disease could be very large. That means, when considering a triplet as one training data, the number of training data relevant to each (particularly rare) disease becomes large enough, such that an equally large number of triplets relevant to each disease can be sampled during training. Suppose there are totally N triplets involved for model training, then the loss function over all the triplets becomes

$$L(\boldsymbol{\theta}) = \frac{1}{N}\sum_{i=1}^{N} l(\mathbf{x}_i, \mathbf{y}_i, \mathbf{z}_i; \boldsymbol{\theta}) \,. \quad (3)$$

We would like to emphasize that, while the triplet loss (Equation 3) has been proposed for other natural image analysis tasks (e.g., face identification [9]), we novelly apply the triplet loss to solve the data imbalance issue.

2.2. Class-center based triplet loss

While the trained CNN model based on the above triplet loss works well and already outperforms the existing approaches (as shown in the Experiment section below), there still exists one training issue due to the data imbalance between classes. For a large-sample class, different images within the class often demonstrate large visual differences, causing the distribution of this class of data spreading in a relatively large region in the original image feature space. Triplet-based model training ideally would help make the distribution of each class more compact in the learned feature space. However, when randomly sampling images partly from the large-sample class to form triplets, the images between two triplets but from the same large-sample classes would likely demonstrate large variations in visual contents. Such large variation may cause the triplet inequality constraints (Equation 1) satisfied for one triplet but not for another, which in turn causes instability during triple-based model training. This may prevent the distribution of the large-sample class in the learned feature space not as compact as expected. Occupying larger region by the large class in the feature space would likely cause more ambiguous regions where multiple classes of data appear, leading to worse performance than desired.

To further improve the performance of triplet-based model training, we propose to include global information on the distribution of each class in triplets, thus replacing the above triplet loss (Equation 2) by the class-center based triplet loss

$$\begin{aligned} l(\mathbf{x}_i, \mathbf{y}_i, \mathbf{z}_i; \boldsymbol{\theta}_t) = [\|\mathbf{f}(\mathbf{x}_i;\boldsymbol{\theta}_t) - \mathbf{c}(\mathbf{x}_i;\boldsymbol{\theta}_{t-1})\| + \alpha \\ - \|\mathbf{f}(\mathbf{x}_i;\boldsymbol{\theta}_t) - \mathbf{c}(\mathbf{z}_i;\boldsymbol{\theta}_{t-1})\|]_+ \,. \end{aligned} \quad (4)$$

Table 1. The details of the two skin datasets.

dataset	class number	image number in largest class	image number in smallest class
Skin7 [14, 15]	7	6705	115
Skin198 [16]	198	60	10

Here, the feature $\mathbf{f}(\mathbf{y}_i;\boldsymbol{\theta})$ of the *positive* image $\mathbf{y}_i$ in the previous loss function (Equation 2) was replaced by the *positive* center ($\mathbf{c}(\mathbf{x}_i;\boldsymbol{\theta}_{t-1})$) of the class containing both $\mathbf{x}_i$ and $\mathbf{y}_i$ in the feature space, where the center $\mathbf{c}(\mathbf{x}_i;\boldsymbol{\theta}_{t-1})$ was obtained by averaging over features of all training images of the same class based on the trained model (with parameters $\boldsymbol{\theta}_{t-1}$) from the previous $(t-1)^{th}$ epoch. Similarly, the feature $\mathbf{f}(\mathbf{z}_i;\boldsymbol{\theta})$ of the *negative* image $\mathbf{z}_i$ from another class was replaced by the *negative* center ($\mathbf{c}(\mathbf{z}_i;\boldsymbol{\theta}_{t-1})$) of the class containing $\mathbf{z}_i$.

With this new triplet loss, images of the same class would be attracted to its class center over training epochs, thus expecting to result in more compact distribution for each class. To initialize the model training from a good start, the pretrained model based on the triplet loss (Equation 2) can be adopted at the first training epoch. This class-center based triplet loss can be considered as the combination of the traditional triplet loss [9] with the single image-based class-center loss [13]. It is also similar to the recently proposed triplet-center loss for 3D object retreival [12]. Different from the triplet-center loss which consider only the nearest *negative* center for each *anchor*, the proposed method here considers all negative centers for each *anchor*. In this way, the global distribution of all classes in the image feature space is considered and could be updated more efficiently during training. Also differently, the proposed class-center triplet loss here is novelly applied to solve the data-imbalance issue.

3. EXPERIMENTAL EVALUATION

3.1. Experimental settings

Two image datasets involving human subjects were used to evaluate the proposed method (see Table 1 for details) [14, 15, 16]. Multiple baseline training strategies relevant to data imbalance were included for comparison, including the basic cross-entropy loss ('CE'), the cross entropy loss with class weights ('WCE'), cross-entropy loss with the strategy of oversampling ('OCE') and focal loss with class weights ('WFCE'). In training, batch size=32 for all baselines. For the proposed original triplet loss method ('TP'), random-hard triplet generation strategy (see reference [9]) was adopted, based on 70 randomly selected images (10 per class) on Skin7, 100 images (5 for each of 20 randomly chosen classes, due to limited GPU memory) on Skin198 at each training iteration. For the class-center involved triplet loss ('TPC'), 32 images were randomly selected at each training iteration. For each image, the feature representation of its class center was computed based the model parameters from last training epoch, and those (*negative*) classes centers which contribute a non-zero losses to Equation 4 were selected. Unless mentioned otherwise, ResNet-50 serves as the backbone for cross-entropy based classifiers, with the last original output layer removed and a new fully connected layer with the 128-dimensional feature vector as output for the proposed approach (including TP and TPC). kNN (k=3 by default) prediction and the nearest centroid based prediction were respectively used for TP and TPC. In all experiments, Adam optimizer was used with learning rate 0.001, $\beta_1 = 0.9$, and $\beta_2 = 0.99$. $\alpha = 0.5$ for triplet loss based training. Each classifier was evaluated with 5-fold cross validation, with average precision, average recall, and average F1 score ('MF1') over classes as the measurement for each validation. Each training lasts for 250 epochs with clear training convergence, and the best scores over 250 epochs were recorded on each validation set for each model. The mean and standard deviation of each measurement over the 5 validations were reported for each test.

3.2. Effectiveness of the triplet-based approach

This section evaluates the effectiveness of the proposed approach (TP and TPC). Table 2 shows that the proposed approach performs better than all the baseline methods on both Skin7 and Skin198 datasets. The proposed class-center involved triplet loss (TPC) overall performs slightly better than the proposed original triplet loss (TP), supporting the effect of class centers on performance improvement. To investigate more details about the effect of the proposed approach on small-sample classes, we also collected the performance of each method for the smallest class (only 115 images) on Skin7 dataset and average performance over 70 classes (less than 20 images for each class) on Skin198. Table 3 shows that on both datasets, the proposed approach (TP and TPC) improved the performance on the small-sample classes more than on the large-sample classes. For example, the improvement in mean F1 score from the strongest baseline OCE to the proposed TPC is 3% on the small class of Skin7 (see Table 3), while the improvement is only 1.1% when considering all classes (see Table 2). Similar larger improvement (from 3% to 4.6%) on smaller data was observed on Skin198. Wilcoxon Rank test showed that the performance of the proposed approach TPC is significantly better than the all the baseline methods (p-values < 0.05)

3.3. Generalizability of the triplet-based approach

To evaluate the generalization of the proposed approach, we evaluated the proposed approach on multiple model architectures and on different size of output dimension. Table 4 shows that on four different CNN models [17, 18, 19, 20], the proposed approach (TP and TPC) consistently outperforms all the baseline methods. Tests with varying output dimensions (not shown due to limited space) shows that the proposed

Table 2. Comparisons between the proposed approach with baseline methods on Skin7 and Skin198 datasets.

	Skin7						Skin198					
	BCE	WCE	OCE	WFCE	TP (Ours)	TPC (Ours)	BCE	WCE	OCE	WFCE	TP (Ours)	TPC (Ours)
MF1	83.65 (1.52)	82.45 (1.31)	83.53 (1.33)	83.52 (1.63)	84.31 (1.93)	**84.89** (0.91)	51.91 (1.10)	60.21 (1.36)	59.77 (1.89)	53.28 (2.65)	61.90 (1.80)	**63.21** (1.61)
Precision	86.96 (1.96)	83.35 (1.79)	87.26 (1.27)	86.43 (1.34)	88.31 (1.79)	**88.42** (0.62)	56.41 (1.27)	64.82 (1.34)	64.87 (2.06)	58.31 (2.77)	**66.11** (2.03)	65.55 (1.65)
Recall	81.15 (1.62)	82.06 (1.47)	80.81 (1.39)	81.25 (1.78)	81.11 (2.27)	**83.02** (0.71)	52.12 (1.14)	60.23 (1.12)	59.34 (1.87)	53.34 (2.58)	62.10 (1.81)	**64.68** (1.63)

Table 3. Performance of methods on small-samples classes of Skin7 and Skin198.

	Skin7						Skin198					
	BCE	WCE	OCE	WFCE	TP (Ours)	TPC (Ours)	BCE	WCE	OCE	WFCE	TP (Ours)	TPC (Ours)
MF1	73.67 (3.62)	77.96 (5.31)	74.05 (8.91)	76.21 (4.94)	75.58 (5.60)	**81.22** (5.07)	18.59 (2.43)	53.37 (1.99)	56.41 (3.55)	20.36 (2.08)	59.31 (3.36)	**61.03** (2.84)
Precision	79.03 (0.76)	87.18 (2.47)	84.93 (5.16)	84.96 (3.61)	86.89 (9.09)	**89.45** (6.97)	24.22 (3.00)	65.21 (2.52)	66.46 (4.25)	26.83 (2.74)	**66.68** (4.03)	63.69 (3.24)
Recall	69.39 (6.24)	70.83 (7.64)	66.17 (11.81)	69.35 (6.62)	67.78 (8.03)	**75.30** (8.39)	16.67 (2.78)	49.79 (2.68)	53.42 (3.17)	17.99 (2.21)	58.45 (3.14)	**63.87** (3.00)

Table 4. Classification performance with various model architectures on Skin198.

	ResNet-50			DenseNet-121			Inception-v4			VGG-19		
	MF1	Precision	Recall	MF1	Precision	Recall	MF1	Precision	Recall	MF1	Precision	Recall
BCE	51.91 (1.10)	56.41 (1.27)	52.12 (1.14)	41.60 (1.85)	44.57 (1.39)	42.85 (1.94)	50.22 (1.78)	53.73 (2.00)	50.83 (1.79)	33.75 (2.69)	35.69 (3.23)	35.52 (2.29)
WCE	60.21 (1.36)	64.82 (1.34)	60.23 (1.12)	55.04 (1.74)	61.42 (2.01)	54.62 (1.68)	57.92 (1.71)	62.30 (1.24)	58.08 (2.05)	47.09 (7.91)	50.43 (7.71)	48.08 (7.67)
OCE	59.77 (1.89)	64.87 (2.06)	59.34 (1.87)	57.72 (1.91)	63.86 (1.85)	56.96 (1.92)	56.79 (2.51)	61.83 (2.70)	56.49 (2.46)	50.85 (1.42)	53.99 (1.44)	52.09 (1.67)
WFCE	53.28 (2.65)	58.31 (2.77)	53.34 (2.58)	43.03 (1.28)	46.70 (1.08)	44.00 (1.37)	49.88 (2.65)	53.53 (2.33)	50.44 (2.86)	37.13 (1.98)	39.68 (2.41)	38.61 (1.73)
TP(Ours)	61.90 (1.80)	**66.11** (2.03)	62.10 (1.81)	60.04 (2.16)	64.03 (2.52)	60.48 (1.86)	57.24 (1.67)	61.09 (2.37)	57.59 (1.24)	50.29 (1.43)	53.94 (1.52)	50.69 (1.58)
TPC(Ours)	**63.21** (1.61)	65.55 (1.65)	**64.68** (1.63)	**62.62** (2.18)	**64.94** (2.41)	**64.26** (1.98)	**59.25** (1.18)	**62.43** (1.48)	**60.12** (1.12)	**52.49** (2.33)	**55.39** (2.84)	**53.49** (2.50)

method is still consistently better than corresponding baseline methods with similar model architectures. In addition, the effect of the hyperparameter α in the triplet loss was also evaluated, showing stable performance when varying in the range $[0.2, 0.8]$, e.g., the mean F1 score varied from 84.4% to 84.9% on Skin7 which all outperformed the strong baseline methods. These consistent results support that the proposed approach is stable with high generalizability.

4. CONCLUSION

This paper introduces a new way to handle data imbalance issues based on triplet loss. The improved class-center involved triplet loss, together with the original triplet loss, outperforms the widely used methods, which has been extensively verified on two skin image classification tasks. The stable and generalizable performance of the proposed approach on multiple model architectures, output dimensions, and hyperparameter settings further confirm its capability in solving the data imbalance issue. This new solution to data imbalance is especially helpful for the development of intelligent diagnosis systems for rare diseases. Future work includes the study of the proposed approach on more medical image datasets.

Acknowledgement. This work is supported in part by the National Key Research and Development Program (grant No. 2018YFC1315402), the Guangdong Key Research and Development Program (grant No. 2019B020228001), the National Natural Science Foundation of China (grant No. U1811461), and the Guangzhou Science and Technology Program (grant No. 201904010260).

5. REFERENCES

[1] Dinggang Shen, Guorong Wu, and Heung-Il Suk, "Deep learning in medical image analysis," *Annual review of biomedical engineering*, pp. 221–248, 2017.

[2] Nitesh V. Chawla, Kevin W. Bowyer, Lawrence O. Hall, and W. Philip Kegelmeyer, "Smote: synthetic minority over-sampling technique," *Journal of artificial intelligence research*, vol. 16, pp. 321–357, 2002.

[3] Yanmin Sun, Mohamed S. Kamel, Andrew KC Wong, and Yang Wang, "Cost-sensitive boosting for classification of imbalanced data," *Pattern Recognition*, pp. 3358–3378, 2007.

[4] Yoav Freund, Robert Schapire, and Naoki Abe, "A short introduction to boosting," *Journal-Japanese Society For Artificial Intelligence*, vol. 14, no. 1612, pp. 771–780, 1999.

[5] Paul Viola, Michael Jones, et al., "Rapid object detection using a boosted cascade of simple features," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. 511–518, 2001.

[6] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár, "Focal loss for dense object detection," in *Proceedings of the IEEE international conference on computer vision*, pp. 2980–2988, 2017.

[7] Hongyu Wang and Yong Xia, "Chestnet: A deep neural network for classification of thoracic diseases on chest radiography," *arXiv preprint arXiv:1807.03058*, 2018.

[8] Mateusz Buda, Atsuto Maki, and Maciej A. Mazurowski, "A systematic study of the class imbalance problem in convolutional neural networks," *Neural Networks*, vol. 106, pp. 249–259, 2018.

[9] Florian Schroff, Dmitry Kalenichenko, and James Philbin, "Facenet: A unified embedding for face recognition and clustering," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 815–823, 2015.

[10] Alexander Hermans, Lucas Beyer, and Bastian Leibe, "In defense of the triplet loss for person re-identification," *arXiv preprint arXiv:1703.07737*, 2017.

[11] Chunfeng Song, Yan Huang, Wanli Ouyang, and Liang Wang, "Mask-guided contrastive attention model for person re-identification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1179–1188, 2018.

[12] Xinwei He, Yang Zhou, Zhichao Zhou, Song Bai, and Xiang Bai, "Triplet-center loss for multi-view 3d object retrieval," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1945–1954, 2018.

[13] Yandong Wen, Kaipeng Zhang, Zhifeng Li, and Yu Qiao, "A discriminative feature learning approach for deep face recognition," in *Proceedings of the European conference on computer vision*, pp. 499–515, 2016.

[14] Philipp Tschandl, Cliff Rosendahl, and Harald Kittler, "The ham10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions," *Scientific data*, vol. 5, pp. 180161, 2018.

[15] Noel Codella, Veronica Rotemberg, Philipp Tschandl, M Emre Celebi, Stephen Dusza, David Gutman, Brian Helba, Aadi Kalloo, Konstantinos Liopyris, Michael Marchetti, et al., "Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging collaboration (isic)," *arXiv preprint arXiv:1902.03368*, 2019.

[16] Xiaoxiao Sun, Jufeng Yang, Ming Sun, and Kai Wang, "A benchmark for automatic visual classification of clinical skin disease images," in *Proceedings of the European conference on computer vision*, pp. 206–222, 2016.

[17] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

[18] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4700–4708, 2017.

[19] Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke, and Alexander A Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," in *Thirty-first AAAI conference on artificial intelligence*, 2017.

[20] Karen Simonyan and Andrew Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.