Comparative Study of Convolutional Neural Network Architecture in Lymphoma Detection

Michaella Yosephine[1(✉)], Rafita Erli Adhawiyah[1], Yasmin Salsabila Kurniawan[1], Isa Anshori[1], Ramadhita Umitaibatin[1], Vegi Faturrahman[1], Rey Ezra Langelo[1], Widyawardana Adiprawita[1], Hermin Aminah Usman[2], and Okky Husain[2]

[1] Bandung Institute of Technology, Bandung, Indonesia
michaella.yosephine@gmail.com
[2] Padjadjaran University, Bandung, Indonesia

Abstract. In this study, we propose an automatic classification of three common types of lymphoma: (1) lymphoma, (2) benign lesion, and (3) carcinoma using lymphoma cell images magnified by 100x and by 400x. A comparative study was performed to find the best architecture to classify lymphoma cell images using the Keras library in Tensorflow. The architectures used in this study are ResNet50, MobileNetV1, and VGG16. Based on the accuracy of lymphoma classification for each architecture, the MobileNet model had the highest accuracy in all three classes at both 100x and 400x magnification levels, which suggests that MobileNet is the best model for lymphoma cell classification. This study can be later used as the base argument in modifying the MobileNet architecture further to get more accurate results in future similar studies.

Keywords: Convolutional neural network · Lymphoma · MobileNet · ResNet50 · VGG16

1 Introduction

Lymphoma is a malignancy of inflammatory cells in the human body, usually involving lymphoid tissue, spinal cord, or both [1]. In general, lymphoma can be grouped into Hodgkin Lymphoma and Non-Hodgkin Lymphoma, differentiated by the presence of Reed-Sternberg cells on histopathological examination [2]. Non-Hodgkin Lymphoma is more common than other types, wherein 2012, the malignancy of this disease is one of the ten malignancies with the most frequent incidence in the world [3]. NHL is considered highly heterogeneous in histological type, symptoms, clinical response to treatment, and prognosis [4].

Lymphoma examination begins with checking the symptoms that appear in the patient and is followed by a biopsy, blood test, and bone marrow retrieval. From this series of tests, the patient is diagnosed with lymphoma if Reed-Sternberg cells are found in the tissue taken during a biopsy [5]. The results of this series of examinations could be misdiagnosed. Errors in the physical examination can lead to lymphoma being left undetected or misdiagnosed [6, 7].

A. Lelono et al. (Eds.): ICOLIB 2021, ABSR 27, pp. 193–202, 2023.
https://doi.org/10.2991/978-94-6463-062-6_19

Machine learning (ML) and deep learning (DL) approaches are often used in assisting medical analysis [8]. ML and DL technology development has assisted several lymphoma-related analyses using whole-slide image (WSI) data from hematoxylin and eosin (H&E) slides. Achi et al. [9] used Convolutional Neural Network (CNN) for the classification diagnosis of four lymphoma cases, namely (1) benign lymphoma, (2) diffuse large B cell lymphoma, (3) Burkitt lymphoma, and (4) small lymphocyte lymphoma. This method has a classification accuracy of 95% for the diagnosis of each image. However, the accuracy of this method reaches 100% when used to diagnose each set of images (one collection of images contains five images). This method can still be developed further in terms of the number of datasets or model improvisations. Syrykh et al. [10] also used CNN added with Bayesian Neural Network (BNN) for differentiation between follicular lymphoma and follicular hyperplasia that display similar features, resulting in classification accuracy varying between 92–99% depending on the image resolution used. The function of BNN in this method is to determine the level of certainty of the predictions generated by the model.

Gaidano et al. [11] also developed a classification of non-Hodgkin B-cell lymphocytes using a classification tree. From this classification tree, it is processed which features have the most significant influence on lymphocyte diagnosis. To perform this method, a large dataset is required. Gaidano et al. used lab data that has been collected from 2003 to 2019. In addition, the data available have specific markers of the 16-year study result. The advantage of using this method is being able to find features that are important and different from conventional (manual) diagnostic processes. These features also vary depending on the model formed in the classification tree process. However, the diagnosis using this method requires clean and large raw data.

In this study, ML is expected to reduce misdiagnosis by helping differentiate lymphoma cells and non-lymphoma cells, Hodgkins and non-Hodgkins lymphoma, lymphoma due to B cells and T cells, and indolent lymphoma and non-Hodgkins lymphoma using sample images.

2 Method

To find the best result, we performed a comparative study by using the architecture available from Keras in the Tensorflow library to diagnose H&E-stained images for three cases; (1) lymphoma, (2) benign lesion, and (3) carcinoma (Fig. 1). The architectures compared were MobileNet, VGG16, and ResNet50. The architectures were used as pre-trained models. Tuning was repeated until the optimum result was obtained.

Lymphoma detection in this study uses a dataset submitted by the Faculty of Medicine, University of Padjajaran. The lymphoma dataset is divided into three classes: lymphoma, benign lesion, and carcinoma. Each class consists of two sub-classes datasets, i.e., datasets with the magnification of 100x and 400x. The total images of each class and sub-class are shown in Table 1.

The quality and quantity of the dataset will be improved at the pre-processing stage. After pre-processing, we perform lymphoma detection using three architectures, namely ResNet50, MobileNet, and VGG16. The accuracy will be compared between the results of each architecture to find the best-performed architecture.

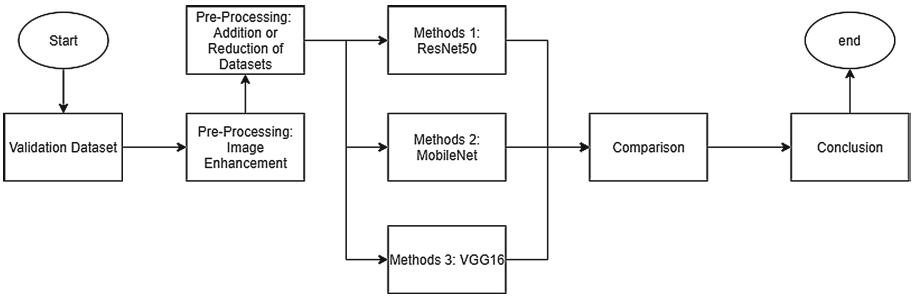


Fig. 1. Research workflow.

Table 1. Comparison of dataset total image for each class and subclass

Magnification	Lymphoma	Benign Lesion	Carcinoma
100x	69	37	47
400x	165	60	59

2.1 Pre-processing

Pre-processing was done in two main steps: image enhancement and dataset balancing (addition or reduction of the dataset), as displayed in Fig. 1. Dataset balancing was done to ensure all datasets had the same number of samples. Image enhancement was done to delete the resolution markers on the original dataset, increasing the contrast between the lighter and darker shades in the image.

Aside from enhancements made above, we also created a second version of our datasets – one containing only green channels extracted from the samples, aimed initially to help with noise reduction and better detection. There were two versions of samples: simply enhanced and enhanced with its green-channel extracted.

Pre-processing was done using photo-editing software Adobe Photoshop CC 2019. To ensure all samples receive the same treatment, we utilized Photoshop's "Actions" feature, which allows the user to record and save sets of the user's actions and choices. The "Actions" were performed on all of its image samples for every dataset. The comparison between "before" and "after" enhancement can be seen in Fig. 2.

The steps for the image enhancement part of the Action set are as follows:

1. Brightness: + 10 (first iteration), + 0 (second iteration)
2. Contrast: + 100 (first iteration), + 25 (second iteration)
3. Translation: (0, + 0.12) inches relative to the center (0, 0)
4. Transformation: 103.5% (both width and height)
5. Hide all channels but Green Channel (for green-channel version extracted only)
6. Copy Green Channel (for green-channel extracted version only).

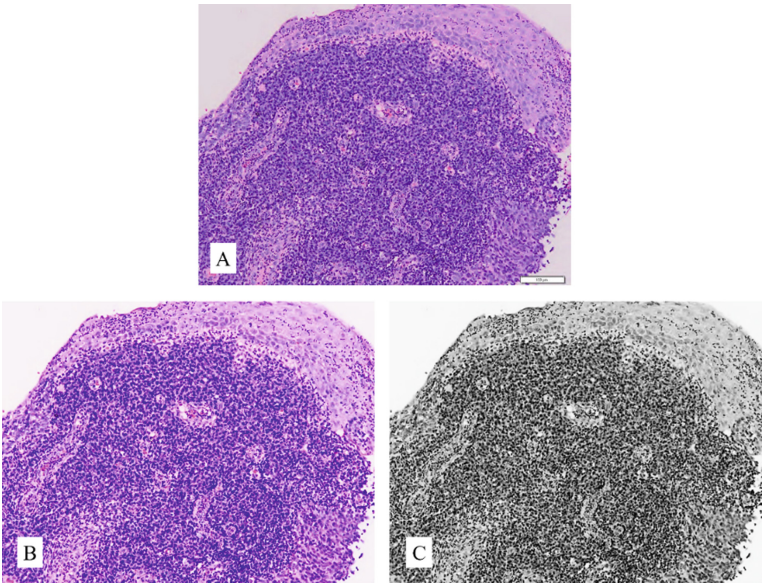


Fig. 2. A) Sample image before enhancement. B) Sample image after enhancement. C) Sample image after green channel extraction.

Table 2. The four steps and action added for each sample version (action set are marked by its letter code)

Steps	Actions		
	Translate (x,y) (inches), only simply enhanced	Translate (x,y) (inches), only simply enhanced	Rotate (degrees)
A	(+0.15, 0)	(+0.151, − 0.009)	0
B	(−0.16, 0)	(−0.16, 0)	0
C	(+0.129, − 0.042)	(+0.12, − 0.05)	+ 0.5
D	(−0.134, − 0.018)	(−0.123, − 0.015)	+ 0.1

Balancing was done by adding more steps to the existing Action set that previously only had enhancement steps, then performing the Action on samples of datasets lacking samples. Four Action sets were required to make sure all datasets had an equal number of samples: Shift Right (A), Shift Left (B), Rotate + 0.5 Degrees (C), and Rotate + 0.1 Degrees (D). More details are displayed in Table 2.

Dataset balancing resulted in 165 samples for each dataset, both for the simply enhanced and green-channel extracted versions. Depending on the original number of samples, datasets could have no replicated samples up to five replicated samples. Once finished, the pre-processed images were exported as.PNG image files sized 1376 × 1038 pixels. Details on the samples after dataset balancing is shown in Table 3.

Table 3. Three different samples with two level of magnification and five different actions

Type	Actions				
	Original	A	B	C	D
Carcinoma 100 ×	47	47	47	24	0
Carcinoma 400 ×	50	50	50	15	0
Benign Lesion 100 ×	37	37	37	37	17
Benign Lesion 400 ×	60	60	45	0	0
Lymphoma 100 ×	69	69	27	0	0
Lymphoma 400 ×	165	0	0	0	0

2.2 Method Implementation I: ResNet50

ResNet-50 is a convolutional neural network that is 50 layers deep. This training model also has other variants such as ResNet101 and ResNet152. Residual Networks (ResNet) are deep convolutional networks with the basic idea of skipping blocks of convolutional layers by using a shortcut connection [12].

The model tries to learn some residual blocks instead of features in residual learning [14]. ResNet can optimize trainable parameters in error backpropagation using these shortcut architectures and avoid the vanishing gradient problem. The learning process can help construct a deeper CNN structure and improve the final result.

2.3 Method Implementation II: MobileNet

MobileNet is a convolutional neural network feature extractor architecture developed by Google in 2018. It is 88 layers deep with 3,538,984 parameters for the feature extractor alone. MobileNetV1 uses depth-wise separable convolution, a strategy to reduce complexity while computing convolution. The classical MobileNetV1 architecture consists of blocks of layers organized as follows: (1) 3×3 depth-wise convolution with stride 1 with ReLU6 activation function; (2) 1×1 pointwise convolution with ReLU6 activation function [13].

2.4 Method Implementation III: VGG16

VGG16 is a convolutional neural network architecture that introduces the effect of depth (numberof layers stacked) in deep learning [16]. VGG16 is said to be an improved version of the older model AlexNet, mainly in its convolutional kernel size. VGG16 uses a smaller convolutional kernel and produces a more accurate result than AlexNet. Despite this, training VGG16 models usually takes a lot of time. One source reported it takes weeks to perform calibration and fine-tuning [15].

3 Result and Discussion

In summary, we trained all three models with all dataset sizes. As an alternative, we also provide green channels from the dataset at 100X and 400X magnifications. The data train used has a size of 1376 × 1037 pixels.

3.1 The 100x Magnification

With the ResNet50 model, we conducted two training sessions using the green channel dataset with a magnification of 100X. We use the pipeline to get the average accuracy of each class. We added four output layers in the form of the dense layer with ReLU activation, global average pooling layer, and three layers of batch normalization and dropout layer. We added a prediction layer with softmax activation. The highest mean accuracy was 0.56 in the lymphoma class from the first train trial. The mean accuracy of the other two classes was 0.383 in the carcinoma class and 0.04 in the benign lesion class. This trial is trained for up to 8 epochs. In the second train trial, the highest average accuracy was 0.925 in the carcinoma class. The mean accuracy of the other two classes was 0.259 in the benign lesion class and 0.022 in the lymphoma class. This trial was trained for up to 15 epochs. Using the VGG16 model, we trained the green channel dataset with 100X magnification. We also use Pipeline to get the average accuracy for each class. The layers used in this training are the same as the ResNet50 model. The highest average accuracy was obtained in the lymphoma class with a value of 0.386. The mean accuracy in the other two classes was 0.323 in the benign lesion class and 0.291 in the carcinoma class. This model is trained for up to 8 epochs.

Using the MobileNet model, we conducted training on our green channel dataset with a magnification of 100x. Before implementation, we divided the dataset into three sets: data train, test, and validation. For the model we created, we used a modified MobileNetV1 layer. We add three output layers in the form of the dense layer with ReLU activation and a prediction layer with softmax activation. We trained and validated the model with 25 epochs. The results of the trained model can be seen in Fig. 3.

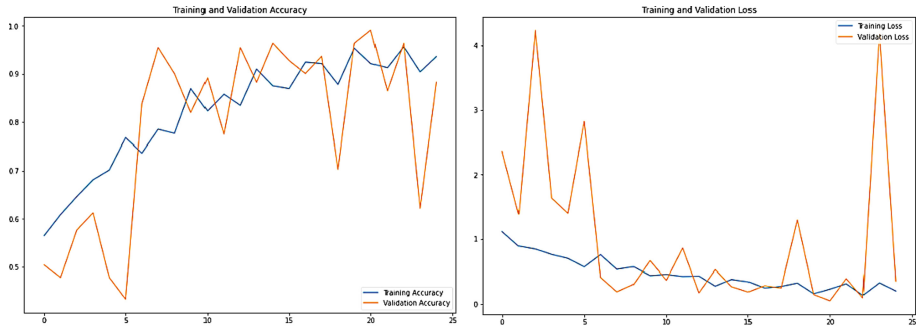
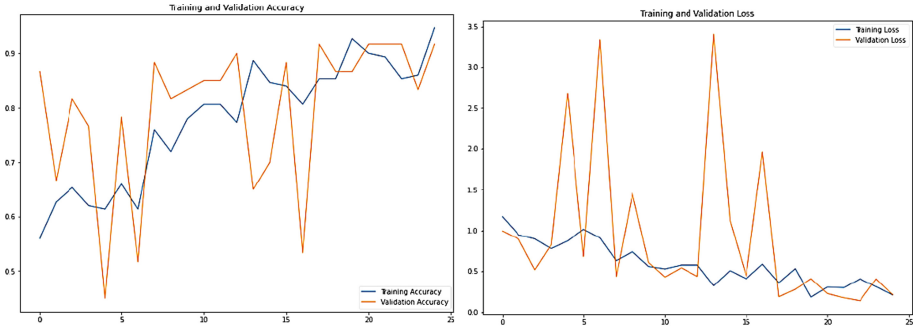


Fig. 3. A) Training and validation accuracy of MobileNet, B) Training and validation loss of MobileNet in 100 × magnified slide images.



A) B)

Fig. 4. A) Training and validation accuracy of MobileNet, B) Training and validation loss of MobileNet in 400 × magnified slide images.

Due to our device limitations, the iteration process can only be done 25 times. In the 25th iteration, it can be seen that the tendency of the accuracy value has stabilized, though it is not the case with the loss value. The accuracy value of the model ranges from 0.56 to 0.95, with the highest accuracy obtained at the 23rd epoch of 0.95. We also measured the pipeline accuracy for each class prediction. The highest accuracy was obtained for carcinoma with a value of 0.998. For the other two classes, it has an accuracy of 0.72 in the lymphoma and 0.991 in the benign lesion.

3.2 The 400x Magnification

We also conducted a training session using the green channel dataset with a magnification of 400X using the same pipeline to get the average accuracy of each class. We added four output layers in the form of the dense layer with ReLU activation, global average pooling layer, and three layers of batch normalization and dropout layer. We added a prediction layer with softmax activation. The highest mean accuracy was 0.839 in the carcinoma class. The mean accuracy of the other two classes was 0.873 in the benign lesion class and 0.161 in the lymphoma class. This trial is trained for up to 8 epochs.

Using the VGG16 model, we trained the green channel dataset with 400X magnification. We also use Pipeline to get the average accuracy for each class. The layers used in this training are the same as the ResNet50 model. The highest average accuracy was obtained in the lymphoma class with a value of 0.393. The mean accuracy in the other two classes was 0.317 in the benign lesion class and 0.288 in the carcinoma class. This model is trained for up to 8 epochs.

Using the MobileNet model, we also conduct training on our green channel dataset with a magnification of 400x and divide it into three classes (carcinoma, benign lesion, lymphoma). We use the same layer as the 100x magnification model in this trained model. The model is also trained with 25 epochs. The results of the trained model can be seen in Fig. 4.

The accuracy value of the model ranges from 0.56 to 0.95, with the highest accuracy obtained at the last epoch. As for the pipeline accuracy for each class, the highest accuracy

Table 4. Comparison of model accuracy for each class (lymphoma, carcinoma, benign lesion) and magnification level (100 × and 400 ×)

Model	Class					
	Lymphoma		Carcinoma		Benign Lesion	
	100 ×	400 ×	100 ×	400 ×	100 ×	400 ×
MobileNet	0.723	0.854	0.998	0.943	0.991	0.994
ResNet50	0.561	0.161	0.383	0.839	0.040	0.873
VGG16	0.386	0.393	0.291	0.288	0.323	0.317

was obtained in the class of benign lesions with a value of 0.99. For the other two classes, the model has an accuracy of 0.85 in the lymphoma class and 0.94 in the carcinoma class.

3.3 Comparison

The comparison of the model accuracy for each class between the three methods implemented is shown in Table 4.

From Table 4, the MobileNet model has the best accuracy for the carcinoma class detection at 100X magnification with an accuracy value of 0.998. From the 400X magnification dataset, the highest accuracy is in the benign lesion class, with an accuracy value of 0.994. As for the other models, the ResNet model accuracy is quite inconsistent for each case and dataset, while the VGG16 model accuracy is very low compared to MobileNet.

We see that there is room for improvement in this work by performing hyperparameter tuning and changing the hidden layer. In this work, the authors are constrained by the small number of datasets provided, thereby reducing the number of trains carried out in the experiment and constrained by the device's inadequate capabilities (requiring a long computation time for each iteration). Therefore, we hope to have larger datasets and computers with better specifications in the subsequent development.

The comparison of the models in the available architecture provided by the Keras library has shown that these architectures can classify H&E histopathology slide images into three classes: lymphoma, benign lesion, and carcinoma, with a significant accuracy score. Green channel extraction is often used to get a less noisy grayscale version of an image. However, it was proven that color profiles contributed a great deal to the algorithm's accuracy in this case. Furthermore, the datasets with a magnification of 100x and 400x also contributed greatly to the algorithm's accuracy, depending on the case predicted and the model architecture. The proposed transfer learning method using architecture models has its own set parameter, giving different results. Based on the accuracy, the best architecture to diagnose each class case is the MobileNet architecture. Despite the accuracy of MobileNet to predict lymphoma still less than 90%, further optimization of the model is still possible with fine-tuning. Changing the hidden layer and the weight of each layer should be considered to get the optimum result as it is meant to tune the architecture towards the dataset.

Acknowledgements. We thank Hermin Aminah Usman and Okky Husain as well as the Faculty of Medicine, University of Padjajaran for providing the dataset used in this study.

Author Contributions. All authors conceived and designed the study. OH and HEU provided the data as representatives from the Faculty of Medicine, University of Padjajaran. MY and REA conducted the data preparation process. REL, RU, VF, and YSK conducted the experiments using the architectures. IA and WA supervised the process. All author read and approved the final manuscript.

References

1. E. &. A. K. Verburgh, "Approach to lymphoma diagnosis and management in South Africa," South African Medical Journal, vol. 109, no. 10, p. 715, 2019. Available: https://doi.org/10.7196/samj.2019.v109i10.14360.
2. H. a, "essential haematology. 6th ed," in hodgkin lymphoma., westsussex, wiley, 2013, p. 246–52.
3. S. I. E. M. d. Ferlay J, "Cancer Incidence and Mortality Worldwide: IARC CancerBase No.11," in GLOBOCAN 2012 v1.1, vol. 1, International Agency for Research on Cancer, 2014. Available: https://doi.org/10.1002/ijc.29210
4. C. Reksodiputro AH, Limfoma Non-Hodgkin (LNH), S. S. H. R. K. J. T. D. Alwi I, Ed., Jakarta: Ilmu Penyakit Dalam, 2015, pp. 2975–86.
5. American Cancer Society, "Tests for Hodgkin Lymphoma," 1 May 2018. [Online]. Available: https://www.cancer.org/cancer/hodgkin-lymphoma/detection-diagnosis-staging/how-diagnosed.html
6. Paul and Perkins, "Lymphoma Misdiagnosis," 2013. [Online]. Available: https://paulandperkins.com/lymphoma/
7. A.M. Perry, R. A. Warnke, Q. Hu, P. Gaulard, C. Copie-Bergman, S. Alkan, H.-Y. Wang, J. X. Cheng, C. M. Bacon, J. Delabie, E. Ranheim, C. Kucuk, X. Hu, D. D. Weisenburger, E. S. Jaffe and W. C. Chan, "Indolent T-cell lymphoproliferative disease of the gastrointestinal tract," National Library of Medicine, 2013. Available: https://doi.org/10.1182/blood-2013-07-512830
8. K. Bera, K. A. Schalper, D. L. Rimm, V. Velcheti and A. Madabhushi, "Artificial intelligence in digital pathology—new tools for diagnosis and precision oncology," Nature reviews Clinical oncology, vol. 16, no. 11, pp. 703–715, 2019. Available: https://doi.org/10.1038/s41571-019-0252-y
9. e. a. Achi, "Automated Diagnosis of Lymphoma with Digital Pathology Images Using Deep Learning," Annals of Clinical and Laboratory Science, vol. 49, no. 2, pp. 153–160, 2019. Available: https://pubmed.ncbi.nlm.nih.gov/31028058/
10. e. a. Syrykh, "Accurate diagnosis of lymphoma on whole-slide histopathology images using deep learning," Digital Medicine, vol. 3, no. 63, 2020. Available: https://doi.org/10.1038/s41746-020-0272-0
11. V. T. N. S. S. V. A. C. G. P. G. S. G. D. R. a. M. G. V. Gaidano, "A Clinically Applicable Approach to the Classification of B-Cell Non-Hodgkin Lymphomas with Flow Cytometry and Machine Learning," cancers MDPI, vol. 12, no. 1684, 2020. Available: https://doi.org/10.3390/cancers12061684

12. E. Rezende, G. Ruppert, T. Carvalho, F. Ramos and P. d. Gues, "Malicious software classification using transfer learning of resnet-50 deep neural network," In 2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA), pp. 1011–1014, 2017. Available: https://doi.org/10.1109/ICMLA.2017.00-19
13. S. Ho-Tsang, "Review: MobileNetV2 — Light Weight Model (Image Classification) outperform MobileNetV1," towardsdatascience.com, 19 May 2019. [Online]. Available: https://towardsdatascience.com/review-mobilenetv2-light-weight-model-image-classification-8febb490e61c [Accessed December 2020].
14. S. B. a. D. S. J. Reddy, "Transfer Learning with ResNet-50 for Malaria Cell-Image Classification," 2019 International Conference on Communication and Signal Processing (ICCSP) , pp. 0945–0949, 2019. Available: https://doi.org/10.1109/ICCSP.2019.8697909
15. Z. Jiang, "A Novel Crop Weed Recognition Method Based on Transfer Learning from VGG16 Implemented by Keras," IOP Conference Series: Materials Science and Engineering , vol. 677, no. 3, 2019. Available: https://doi.org/10.1088/1757-899X/677/3/032073
16. K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," in International Conference on Learning Representations, California, 2015. Available: https://doi.org/10.48550/arXiv.1409.1556