2017 International Conference on Big Data, IoT and Data Science (BID)
Vishwakarma Institute of Technology, Pune, Dec 20-22, 2017

MobileNets for Flower Classification using TensorFlow

Nitin R. Gavai
SCOE, Pune
nrgavai.scoe@sinhgad.edu

Yashashree A. Jakhade
SCOE, Pune
yajakhade.scoe@sinhgad.edu

Seema A. Tribhuvan
SCOE, Pune
seematribhuvan11@gmail.com

Rashmi Bhattad
MMCOE, Pune
rashmibhattad@mmcoe.edu.in

***Abstract*— Classification of objects into their specific classes is always been significant tasks of machine learning. As the study of flower, categorizing specific class of flower is important subject in the field of Botany but the similarity between the diverse species of flowers, texture and color of flowers, and the dissimilarities amongst the same species of flowers, there still are some challenges in the recognition of flower images. Existing recent Google's inception-v3 model comparatively takes more time and space for classification with high accuracy. In this paper, we have shown experimental performance of MobileNets model on TensorFlow platform to retrain the flower category datasets, which can greatly minimize the time and space for flower classification compromising the accuracy slightly.**

Keywords—Classification, Inception-v3, Mobile Nets, TensorFlow

I. INTRODUCTION

Flower classification is a fundamental research work not just in the field of Botany but also the Image Processing through Machine Learning. Till now, it has been found that there are thousands of species of flowers, one of the most prosperous species on this planet. When people uses vision devices to shoot flowers, one may get confused if one doesn't know the species of flowers. Therefore, having a fast and accurate flower classifier will bring a lot of eagerness in peoples' lives. There are some challenges in flower classification like the similarity between the different species of flowers, the complex background of flower image. We cannot just rely on a single feature, such as color, texture or shape to distinguish them. The same species of flowers will be different because of the shades of colors, shape, scale, viewpoint etc.

The obvious conformist method of flower classification is to observe the growing habit and habitat, botanical structure and other features of flowers, compare with the distinguished flowers and ultimately determine their types. This classification method is entirely based on expertise of observer; the workload is large so need of professionally skilled person who is prominent in this domain has all the more increased.

In 2012 ImageNet Large Scale Visual Recognition Competition (ILSVRC) [24] Convolutional Neural Networks (CNN) have become more popular than AlexNet in computer vision after winning the competition[13].To achieve higher accuracy in image classification development and usage of deeper and complex CNN become trend in research [14][17][15][8]. However, the complex structure of models does not to not necessarily making networks more efficient with respect to speed and size to increase the accuracy. On computationally limited platform object detection and recognition has to be done with time critical manner, for example online games, self-driving car, robotics and automation, augmented reality etc.

Convolutional neural network is a competent image recognition method which has been established in latest years. It uses local receptive field as neurons in brain, weights sharing and linking information and greatly decreases the training constraints in comparison with the neural networks. It also performs image transformations with a certain degree of rotation, translation and distortion. This network avoids the complex preprocessing of the image, and people can input the original image directly. It has progressed greatly in the domain of image processing with classification.

In this paper we have shown a competent network design and a set of two hyper-parameters in order to construct very small, high correctness and speedier models that can be easily fit into the design necessities based on mobile and embedded vision applications. Section II describes existing work with Tensor Flow in constructing small models. Section III describes main theme, the MobileNet architecture and width multiplier and resolution multiplier as two hyper-parameters of the networks, to prove smaller and speedier MobileNets. Section IV gives building of classification model; Section V refer to experimentation on five different flowers' datasets. Section VI concludes with summary.

II. EXISTING WORK

In the traditional flower classification methods, convolutional neural network [7] uses multilayer convolution to extract features and combine them automatically. It also uses the pooling layer, fully connected layer and softmax. Google made TensorFlow [1]open source and used for arithmetic calculation,

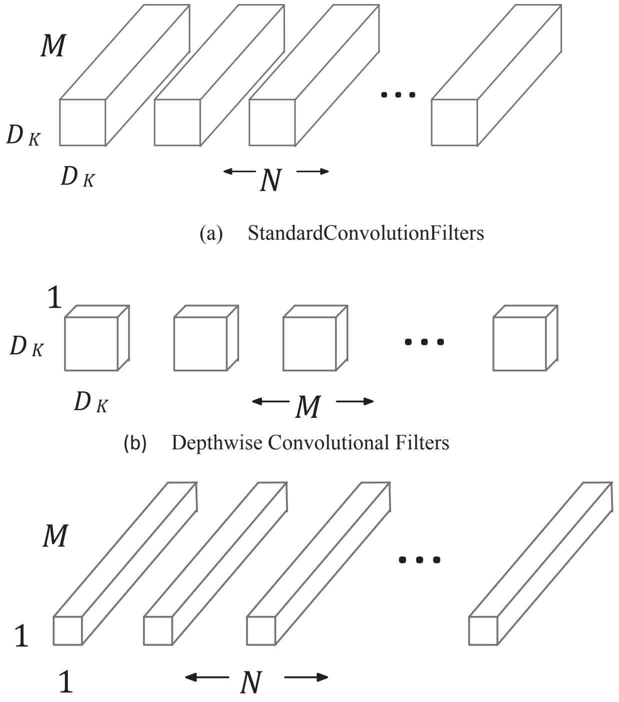
978-1-5090-6593-6/17/$31.00 ©2017 IEEE

specializing in machine learning applications. Second generation of Google artificial intelligence learning system got much attention and affirmation in the field of machine learning all over the world. TensorFlow has advantages of high accessibility, high flexibility, and high provision of TensorFlow researchers through github and online forum to progress its efficiency. Today, Google has unconfined number of pretrained models on the TensorFlow's official website, to expedite the use of researchers in different fields. MobileNets [3] is one of the pretrained models on the TensorFlow. It is an continuous improvement to the initial structure of computer vision after Inception-v1 [6], Inception-v2 [5], Inception-v3 [4] in 2015. The MobileNet [3] model is trained on the ImageNet datasets, comprising the facts that can identify 1000sof categories in ImageNet, the fault percentage of top-5 is upto 3.5%, the fault percentage of top-1 dropped to 17.3% [2] [12]. Transfer learning is a new machine learning method which can use the existing knowledge learned from one environment and solve the new problem which is different but has some relation with the old one. TensorFlow [1] provides comprehensive tutorials to reskill commencement's final layer for new classifications by means of transfer learning. For example, we can relate the information learned from the guitar to study the violin and any musical instruments. Compared with the traditional neural network, it only needs to use a small quantity of data to train the model, and attain high exactitude with a short training time.

Depth wise separable convolutions have been shown to be successful model used for image classification, in both the cases, obtaining better models than previously possible for an available parameter count [4] [5]and significantly dropping the number of parameters essential to perform at a given level[19] [3].

III. Architecture of MobileNet

In this section we described that the core architecture layer of MobileNets which is built on depth wise distinguishable filters [3]. In MobileNets the depth wise intricacy is used and that applies a distinct filter to each response channel. The MobileNet model is constructed on depth wise distinguishable intricacies which happens to be a form of factorized complexities that factorizes a standard complexity into a depth wise complexity [19] and a 1x1 complexity is termed as pointwise complexity. After that the point wise complexity applies a 1x1 complexity to add the outputs to the depth wise complexity. In a standard complexity both are filtered and add inputs to form a different output set in one step.



(a) StandardConvolutionFilters

(b) Depthwise Convolutional Filters

(c) 1×1 Complex Filters named Pointwise Complexity in the context of Depthwise Distinguishable Complexity

Fig 1. The regular complex filters in (a) are substituted by two layers: depthwise complexity in (b) and pointwise complexity (c) to build a depthwise distinguishable filter.

The depthwise distinguishable complexity separates this into two layers, one for filtering and the other for linking. By doing this separation it has a huge effect of reduction with computational time and size of the model. Figure 1indicationsdiverse scenes of how a regular complexity 1(a) is factorized into a depthwise complexity1(b) and a 1x1 pointwise complexity 1(c).

IV. Construction of Model

This section focuses on experimental setup for flower classification model using MobileNet on TensorFlow framework. Here classification model is separated into following four stages: image preprocessing, training, verification followed by testing phase.

A. Image Preprocessing

In the image preprocessing step we need to label the data since the learning method of convolution neural network fits into administered learning in machine learning.

`

B. Training Process

The main working flow of MobileNet [3] model is shown in figure 1.

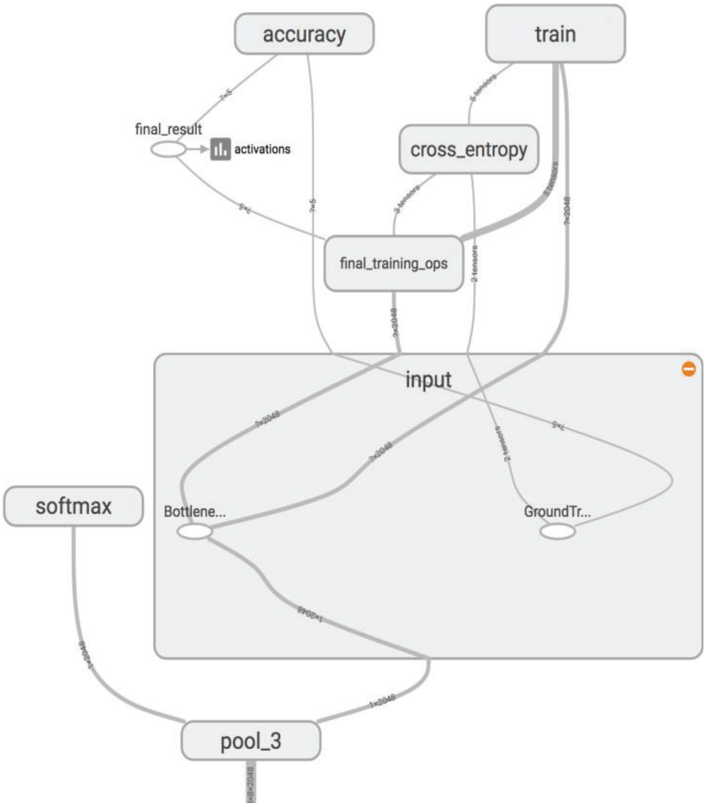


Fig 2. MobileNet model

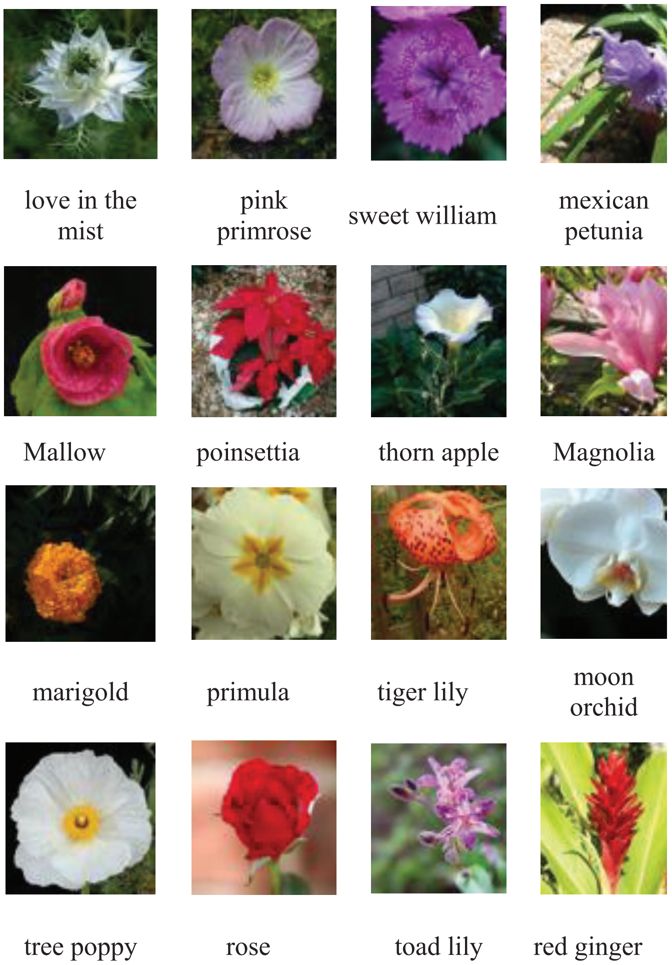
In the above given MobileNet model, on the left side the labeled node "softmax", is the output layer of the original model. On the other hand, all the nodes to the right side of "softmax" were further supplemented by the reskilling script.

C. Verification and testing process

For verification we use Oxford-102 flower dataset [9] [14]: this dataset contains 102 species of flowers, it was created by Maria-Elena Nilsback and Andrew Zisserman in 2008 [9] [11], each of species contains 40-258 images. Compared with the Oxford-17 flower dataset, the Oxford-102 flower dataset[14] contain more species of flowers and there are more similarities between the different types, so the flower grouping will be more intricate.

V. Experiment

This experiment is based on the MobileNet [3] model on python with TensorFlow [1] open source library, the hardware platform is Dell N-series laptop: processor 2.5GHz Intel i5, memory 12GB. The experimental datasetis the Oxford- 102 flower.



love in the mist | pink primrose | sweet william | mexican petunia

Mallow | poinsettia | thorn apple | Magnolia

marigold | primula | tiger lily | moon orchid

tree poppy | rose | toad lily | red ginger

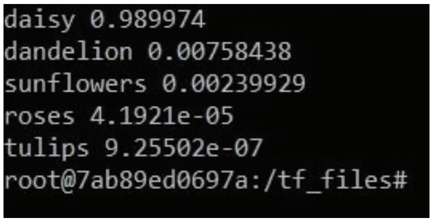
Fig 3. Few species of flowers in Oxford-102 flower dataset.

A. Training

While training the model we use Oxford-102 flower dataset [9] [14], every image is used multiple times through training process. Computing the layers behind the layer just before the final output layer whichperforms the grouping for each image takes a substantialtime. As the lower layers of the network are not being changed their outputs can be stored and used again.

B. Experiments' Result

We would get to see correctly classified flower with an accuracy value of between 85% and 99%, though the resulting value will change from case to case since there's arbitrariness in the process of training. On top of that, if we give training for only two flowers as two different categories, we should expect higher precision. This number value specifies the percentage of the images in the test set which are provided the exact label after training the model completely.



```
daisy 0.989974
dandelion 0.00758438
sunflowers 0.00239929
roses 4.1921e-05
tulips 9.25502e-07
root@7ab89ed0697a:/tf_files#
```

Fig 4. results of MobileNets

As it trains a series of output steps can be seen, each one showing training precision, validation accuracy, and the cross entropy:

Accuracy of training illustrates the percentage of the images used in the current training batch that were labeled with the precise class. *Validation accuracy* is the precision; percentage of correctly-labelled images, on a randomly selected images from a different set. *Cross entropy* is a loss function which gives a sight into how good the process of learning is moving ahead, lower numbers are better here [18] [3].

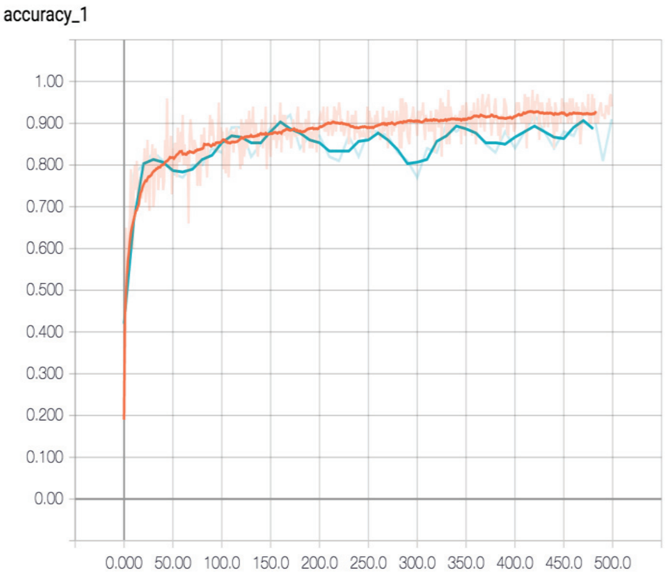


Fig 5. Accuracy with respect to training and validation

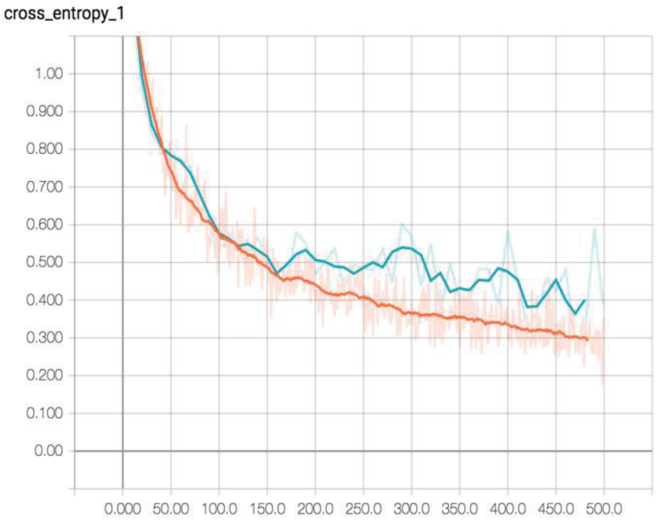


Fig 6. Cross entropy with respect to training and validation

If the precision of the training is high but the accuracy of validation remains low the meaning of which is that the network is over fitting, and that it is retaining the specific features in the training images which doesn't help it grouping images more mainly. An exact measure of the networks performance is to quantify its performance on a set of data that is not a subset of training data. This performance is quantified with the help of precision validation.

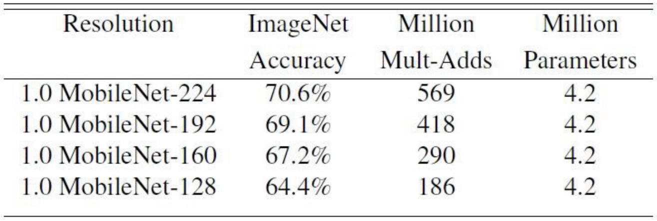
Table I indicates the precision, computation and size tradeoffs of shrinking the MobileNet [3] architecture with the width multiplier α. Accuracy reduces smoothly until the architecture is made too small at α = 0.25

Table II shows the correctness, computation and size tradeoffs for diverse resolution multipliers by training MobileNets with reduced input resolutions. The Precision reduces gradually across resolution.

TABLE I. MobileNet Width Multiplier [3]

Width Multiplier	ImageNet Accuracy	Million Mult-Adds	Million Parameters
1.0 MobileNet-224	70.6%	569	4.2
0.75 MobileNet-224	68.4%	325	2.6
0.5 MobileNet-224	63.7%	149	1.3
0.25 MobileNet-224	50.6%	41	0.5

TABLE II. MobileNet Resolution [3]

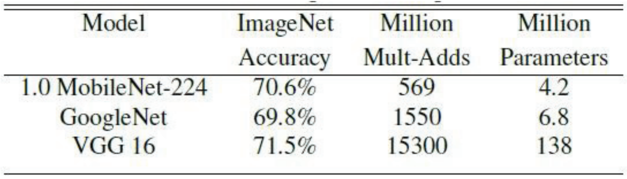


Resolution	ImageNet	Million	Million
	Accuracy	Mult-Adds	Parameters
1.0 MobileNet-224	70.6%	569	4.2
1.0 MobileNet-192	69.1%	418	4.2
1.0 MobileNet-160	67.2%	290	4.2
1.0 MobileNet-128	64.4%	186	4.2

Comparison with other models

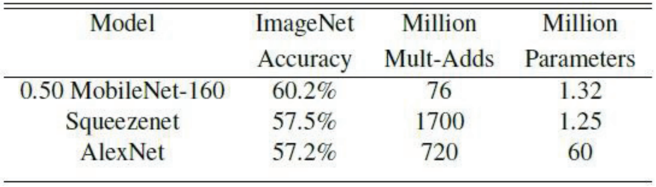
MobileNets is comparatively precise than GoogleNet being smaller and more than 2.5 times less computation. Table III compares full MobileNet to the original GoogleNet [16] and VGG16 [14] [18]. MobileNet is nearly as accurate as VGG16 despite being 32 times less and 27 times less compute intensive.

TABLE III. MobileNet Comparison to Popular Models [3]



Model	ImageNet	Million	Million
	Accuracy	Mult-Adds	Parameters
1.0 MobileNet-224	70.6%	569	4.2
GoogleNet	69.8%	1550	6.8
VGG 16	71.5%	15300	138

TABLE IV. Smaller MobileNet Comparison to Popular Models



Model	ImageNet	Million	Million
	Accuracy	Mult-Adds	Parameters
0.50 MobileNet-160	60.2%	76	1.32
Squeezenet	57.5%	1700	1.25
AlexNet	57.2%	720	60

Table IV compares a reduced MobileNet with width multiplier α = 0.5 and reduced resolution 160x160. Reduced MobileNet is 4% better than Squeezenet [10] at the same size and 22 times less calculation. It is also4% better than AlexNet [12] being 45 times less in size and 9.4 times less computation than AlexNet.

VI. CONCLUSION

We experimented flower classification with the Google's new model architecture, MobileNets found on depthwise distinguishable convolutional neural network. We proved how to build smaller and faster MobileNets application as flower classifier using width multiplier and resolution multiplier by trading off a rational amount of correctness to shrink size and latency. We concluded by validating MobileNet's efficiency when applied to an extensive range of image dataset. The MobileNets are optimized to become small and efficient by compromising on the accuracy aspect.

REFERENCES

[1] Peter Goldsborough Fakultät für Informatik, "A Tour of TensorFlowarXiv:1610.01178v1 [cs.LG] 1 Oct 2016.

[2] Abhineet Saxena, "Convolutional neural networks: an illustration in TensorFlow". ACM Crossroads 22(4): 56-58, 20 16.

[3] Andrew G. Howard Menglong Zhu Bo Chen Dmitry KalenichenkoWeijun Wang Tobias Weyand Marco Andreetto Hartwig Adam,"MobileNets Efficient Convolutional Neural Networks for Mobile Vision,"arXiv:1704.04861v1 [cs.CV] 17 Apr 2017.

[4] J. Krause, B. Sapp, A. Howard, H. Zhou, A. Toshev, T. Duerig, J. Philbin, and L. Fei-Fei. "The unreasonable effectiveness of noisy data for fine-grained recognition". arXiv preprint arXiv:1511.06789, 2015.

[5] Sergey loffe, Christian Szegedy, et al. "Batch Normalization: Accelerating Deep Network Training by Reducing Internal CovarIate Shift". ICML20 15: 448-456, 20 15.

[6] J. Christian Szegedy, Wei Liu, et aI, "Going Deeper with Convolutions". arXiv:1409.4842,2014.

[7] Angelova A, Zhu S, Efficient Object Detection and Segmentation for Fine-Grained Recognition[C]. Computer Vision and Pattern Recognition: 8 1 1-8 18, IEEE, 2013.

[8] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. arXiv preprint arXiv:1512.03385, 2015.

[9] Maria-Elena Nilsback, Andrew Zisserman: Automated Flower Classification over a Large Number of Classes. ICVGlP 2008: 722-729,2008.

[10] F. N. Iandola, M. W. Moskewicz, K. Ashraf, S. Han, W. J. Dally, and K. Keutzer. Squeezenet: Alexnet-level accuracy with 50x fewer parameters and¡ 1mb model size. arXiv preprint arXiv:1602.07360, 2016.

[11] Maria-Elena Nilsback, Andrew Zisserman: A Visual Vocabulary for Flower Classification. CVPR (2) 2006: 1447-1454,2006.

[12] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems, pages 1097–1105, 2012

[13] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, et al. Imagenet large scale visual recognition challenge. International Journal of Computer Vision, 115(3):211–252, 2015

[14] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556, 2014.

[15] C. Szegedy, S. Ioffe, and V. Vanhoucke. Inception-v4, inception-resnet and the impact of residual connections on learning. arXiv preprint arXiv:1602.07261, 2016

[16] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,pages 1–9, 2015.

[17] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna. "Rethinking the inception architecture for computer vision". arXiv preprint arXiv:1512.00567, 2015.

[18] K. He, X. Zhang, S. Ren, and J. Sun. "Deep residual learning for image recognition". arXiv preprint arXiv:1512.03385, 2015.

[19] Łukasz Kaiser, Aidan N. Gomez, François Chollet. "Depthwise Separable Convolutions for Neural Machine Translation". arXiv:1706.03059v2 [cs.CL] 16 Jun 2017