Image Classification

Kenneth Warmuth, Alexander Wo¨lfel, Abhinav Bhattacharya, Asad Ullah, Muhammad Faiz Zeeshan

*Dept. of Audio Visual Technology; Technische Universita¨t Ilmenau, Germany*

*Email:* [*kenneth.warmuth@tu-ilmenau.de,*](mailto:kenneth.warmuth@tu-ilmenau.de) [*alexander.woelfel@tu-ilmenau.de,*](mailto:alexander.woelfel@tu-ilmenau.de) [*abhinav.bhattacharya@tu-ilmenau.de,*](mailto:abhinav.bhattacharya@tu-ilmenau.de)[*asad.ullah@tu-ilmenau.de,*](mailto:asad.ullah@tu-ilmenau.de) [*muhammad-faiz.zeeshan@tu-ilmenau.de*](mailto:muhammad-faiz.zeeshan@tu-ilmenau.de)

***Abstract*—Recent work in the field of Machine- and Deep Learning has shown that Convolutional Neural Networks are interesting and promising to apply in different fields of use, and one such use case is the classification of images of a photo library. In this work, we compare the performance of three pre-trained CNNs on a dataset of 200 high-resolution images. For this purpose, EfficientNet B7, NasNetLarge, and DenseNet 201 are chosen, based on their state-of-the-art benchmarks of Top-1 and Top-5 accuracy for CNNs trained on the ImageNet database and implementability. The high-resolution images are processed with the Brute-Force approach to evaluate the accuracy of the Image Classification of each CNN. It is observed that the three CNNs perform poorly on the given dataset with a Top-1 and Top-5 accuracy of less than one-third of the state-of-the-art. The used method to evaluate this result by considering the easy-to-classify images shows, that these are often portraits. From this, it is concluded, that the on ImageNet trained CNNs perform worse on portraits in general.**

1. INTRODUCTION

Every day, generally every person visually recognizes a high number of things in his environment. They recognize for example a thing with two wheels, a handlebar, and a saddle and know it is a bike. People connect different information and combine them to a bigger result. For humans, this and also to establish a connection between a text and an image is an easier process, because they have learned it in early years. For computers, this is not an easy task, and image classification a big challenge. Computers need prior knowledge and have to learn to handle this process [[1].](#_bookmark0) Machine learning and deep learning are currently the most preferred ways to accomplish this task.

The analysis of unstructured data is widely researched in the field of Big Data and Deep Learning. The analysis and extraction regarding certain information in texts have made further progress than image analysis, respectively image classification. The management and feature-specific organization of photo libraries is a relevant use case due to the strong increase in digital photography

and the spread of smartphones since”1.2 trillion digital photos taken in 2017, 85% were captured via smartphones” [[6].](#_bookmark5) Besides, a wide range of applications in industry, medicine, and autonomous driving are already in use and more are conceivable in the future. This is to emphasize the relevance of the topic image classification. [[2].](#_bookmark1)

The goal of this paper is to analyze three image classification algorithms and rank them concerning their performance on a high-resolution image dataset.

1. STATE OF THE ART

As part of artificial intelligence, machine learning should enable computers to recognize patterns in given databases to find and create solutions for the given problem or to analyze the data independently. Different categories of machine learning algorithms exist supervised learning, unsupervised learning, and reinforcement learning. Supervised learning is divided into classification and regression. The classification supervised learning “[. . . ] involves predicting a class label. [[3]”](#_bookmark2) As a part of machine learning, Deep Learning is the next step in this topic. For information processing, Deep Learning uses neural networks and big datasets. The datasets are suitable for learning, especially to recognize samples. The base for this is the neural networks, which are getting trained during the learning process. The structure of the neural network is oriented to the structure of the human brain. They have a multi-layered structure and consist of neurons. The neurons are combined in a series of groups. In neural networks, it “is a multi-layered structure of algorithms. [[4]”](#_bookmark3) The “Neurons in each layer are connected to neurons of the next layer. [[5]”](#_bookmark4) As is the case with people, the more linkages there are, the more complex issues can be solved

and mapped.

The creation of a computer model based on the neural network of the human brain as a combination of algorithms and mathematics in 1943, by logician Walter Pitts and neurophysiologist Warren McCulloch, is described as the start point of Deep Learning. The first ideas to develop Deep Learning algorithms are made by mathematician Alexey Grigoryevich Ivakhnenko and author Valentin Grigorevich Lapa. The algorithm was able to choose the best statistical features to forward them to the next layer. The first CNN was developed by computer scientist Kunihiko Fukushima. He developed an artificial neural network, called Neocognitron, with a hierarchical and multilayered structure, in 1979. In 1999, computers become faster. This caused better results of neural networks using the same data. In 2009, ImageNet as a free database of more than 14 million labeled images was launched. The Cat Experiment was the next step in the development of Deep Learning. In 2012, it examines the complexity of unsupervised and supervised learning. This experiment was 70% more efficient than the forerunners by processing unlabeled images, but the accuracy of recognizing objects was worse [[16].](#_bookmark15)

1. EVALUATED NETWORKS

In the following subsections, three different Convolutional Neural Networks EfficientNetB7, Nas- NetLarge, and DenseNet201 are described in detail. All three are trained on different subsets of the ImageNet Database and were selected because of the integration in the TensorFlow Keras API, which provides the possibility for a target-oriented implementation, just like their performance in terms of accuracy on reference datasets also given in Table 1:

TABLE I

OVERVIEW OF ACCURACY OF THE CHOSEN CNNS

|  |  |  |  |
| --- | --- | --- | --- |
| **Database** | **CNN** | | |
| ***NASNetLarge*** | ***EfficientNet B7*** | ***DenseNet201*** |
| ImageNet | Top-1 accuracy:  82.7%  Top-5 accuracy 96.2% | Top-1 accuracy:  84.3% | Top-1 accuracy:  77.42%  Top-5 accuracy 93.66% |

*A. Fundamentals*

In this paper, a dataset of 200 high-resolution images is used. The paper aims to analyze and evaluate three image classification algorithms, which are based on objective reasons, like practicability of the implementation, computational power, and integrability of the given images. The algorithms can be classical- or machine-learning- based. All algorithms are applied to a given dataset. After that, the classified images of each classification algorithm should be analyzed and evaluated.

1. *NASNetLarge:* NAS is the abbreviation for Neural architecture search. It is a technique to automate the design of artificial neural networks (ANN). It is used to develop networks, that perform better than hand-created architectures [[7].](#_bookmark6) Corresponding methods can be distinguished between “[. . . ] search space, search strategy, and performance estimation strategy. [[8]”](#_bookmark7) The NASNet Search Space enables transferability. In the actual NASNet from 2018, there is a new regularization technique implemented, which is called Scheduled- Drop-Path and is a learning algorithm. This technique improves the generalization in the models. Generalization means, that the degree of details of information gets reduced. The current state of the art on ImageNet is a Top-1 accuracy of 82.7% and 96.2% Top-5 accuracy. Furthermore, the model is 28% less computational demanding compared to the previous model. Thereby the NasNet improves the processing time of learning on large datasets by applying a block for small datasets on larger datasets [[9].](#_bookmark8)
2. *EfficientNetB7:* EfficientNet is a CNN based on MobileNets and ResNet, which are scaled up from B0 to B7, each model includes more parameters and a higher accuracy [[10].](#_bookmark9) The result of EfficientNet B7 is much better accuracy and efficiency. This is done by inserting a new scaling method, that is applied to the dimensions depth, width, and resolution. The convolution is divided into depthwise and pointwise convolution. This is done to reduce the calculation time. EfficientNet B7 uses an Inverse ResNet. That means, that

the channels are extended and then compressed. Furthermore, it uses a linear bottleneck in the last layer in each block to prevent loss from the activation function by linear activation [[10].](#_bookmark9) The current state-of-the-art EffcientNet B7 has a Top-1 accuracy of 83.3% on ImageNet. Furthermore, it is 8.4 times smaller and 6.1 times faster “[. . . ] on interference than the best existing ConvNet. [[11]”](#_bookmark10)

1. *DenseNet201:* The Densely Connected Convolutional Network follows the principle of connecting each layer to every other layer in a feed-forward fashion and increasing “the depth of deep convolutional networks. [[12]”](#_bookmark11) The number behind DenseNet corresponds to the number of layers. The layers are more accurate due to shorter connections between the layers “[. . . ] close to the input and those close to the output. [[13]”](#_bookmark12) This is achieved by a direct connection between every layer with each other. The difference to other Convolutional Networks is, that the DenseNet201 has L(L+1)/2 direct connections instead of one between each layer and its subsequent layer. Due to these links, DenseNet needs fewer parameters compared to a traditional CNN. The current state of the art is a Top-1 accuracy of 77.42% and Top-5 accuracy of 93.66% [[14].](#_bookmark13)
2. EVALUATION

Besides the general network-specific pre-processing an image has to be resized to the required dimensions via Open CV. All scripts and results are available in the project GitHub: [link.](https://github.com/asad-62/IVP-DNN)

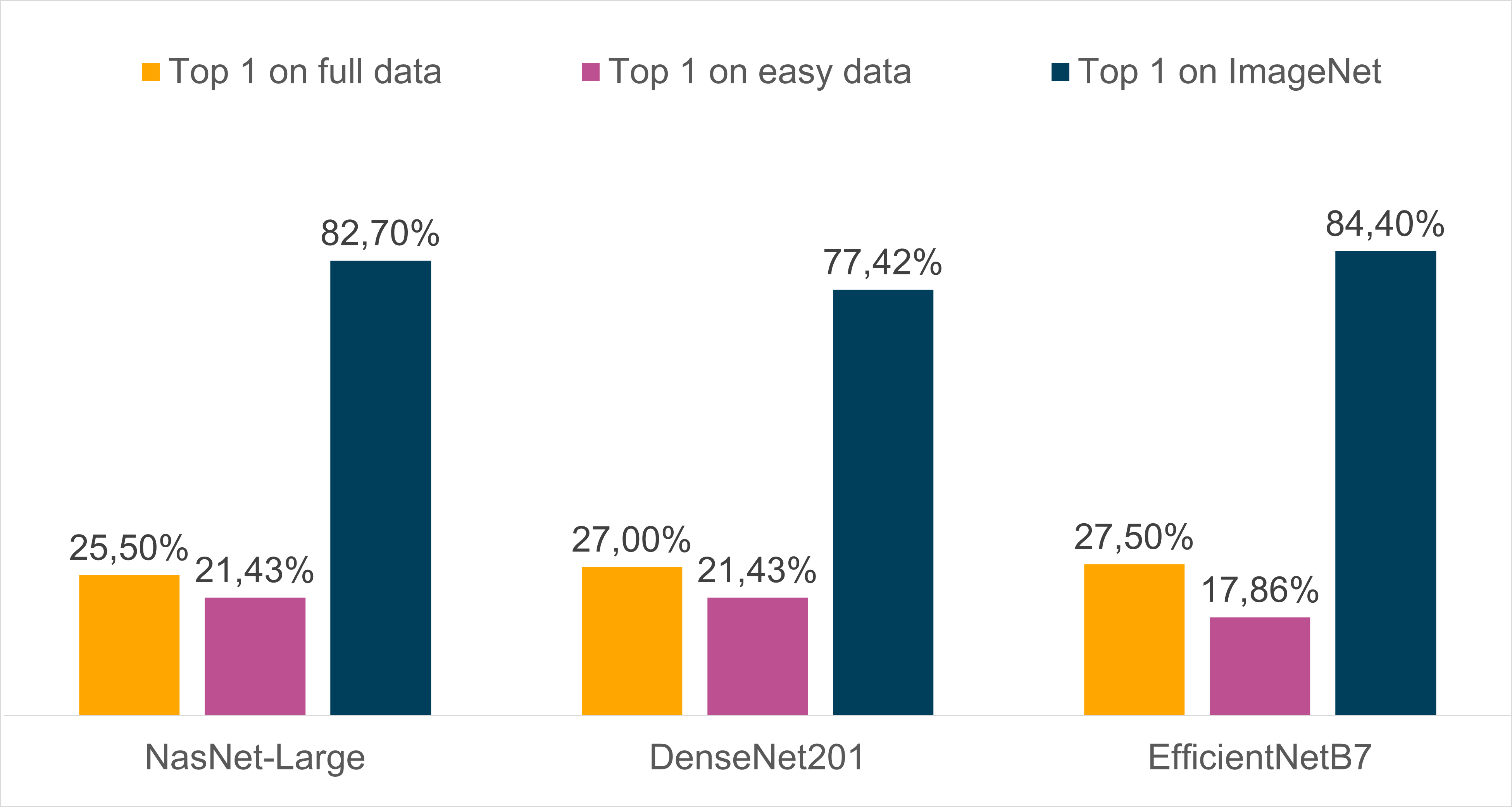
To compare the results with the ImageNet benchmark the TOP 1 and TOP 5 Accuracy is calculated by dividing the number of correct predictions by the total number of predictions [[17].](#_bookmark16)

*Number of correct predictions*

*Accuracy* =

*Total number of predictions*

1. RESULTS

In determining the accuracy of the CNNs, the observation was made that while the meaning of the ground truth labels often matched the content of the images, but the exact wording did not fit. Therefore, to avoid falsification of the results, the images were compared manually with the respective classification results.

1. *Ground Truth*

The ground truth is a dataset with 200 images with up to 20 labels per image. The image labels were generated by choosing up to five labels for each image by every group member. For each image, one label out of all labels was selected based on the frequency of the occurrence to the Top-1 accuracy.

1. *Pre-Processing*

All CNNs are applied to the images in full resolution. To work with high-resolution images, the Brute-force approach, which includes the resizing of the images to required dimensions in an Open CV within the pre-processing for the specific Convolutional Neural Network, is chosen.

1. *Test procedure*

To classify the images and run the chosen CNNs, the TensorFlow included Keras API is used.

Fig. 1. Top-1 accuracy in comparison on different datasets

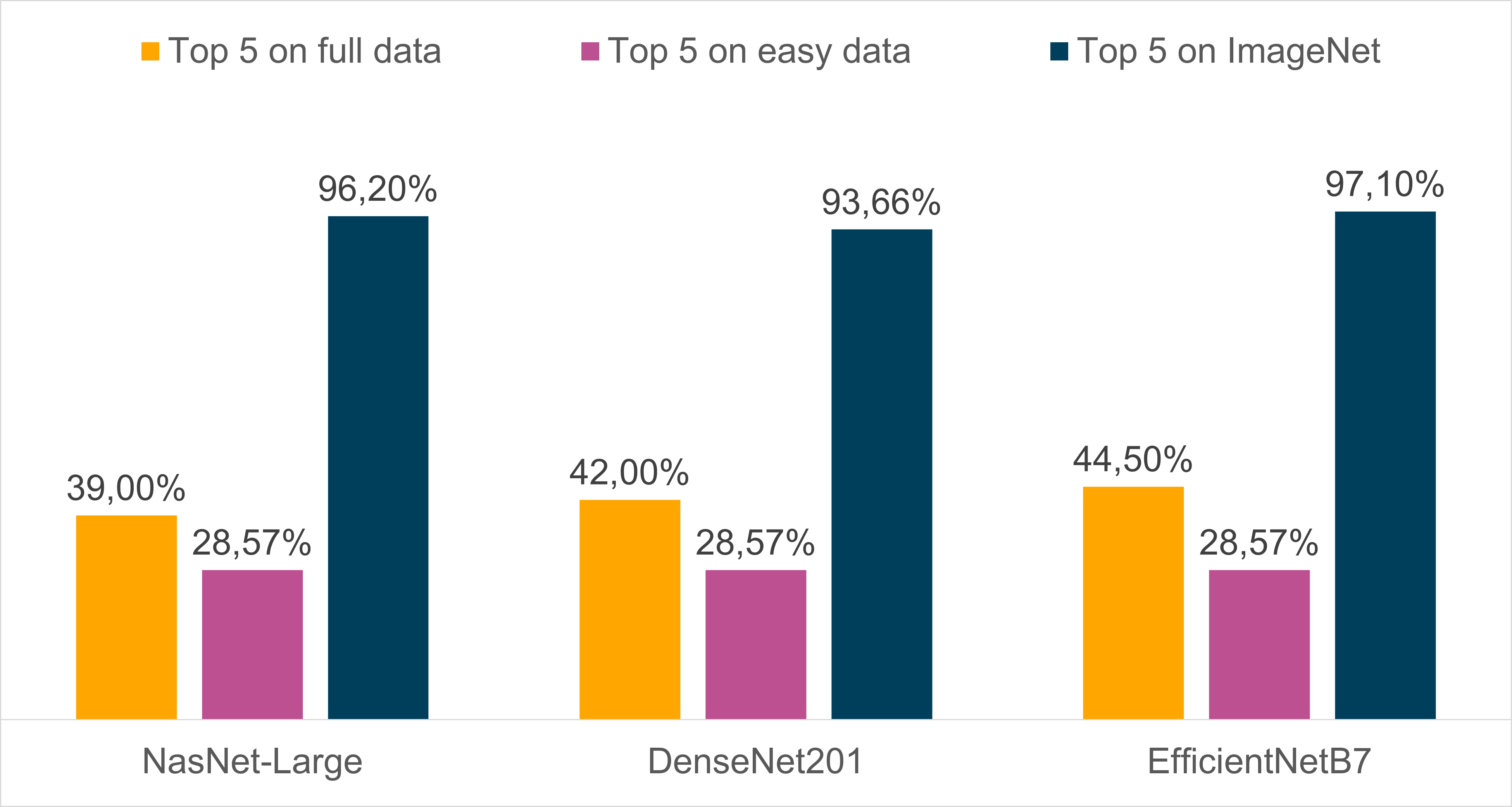


Fig. 2. Top-5 accuracy in comparison on different datasets

As can be seen from the figures [1][2], all the algorithms underperformed in terms of the reference

accuracy. Since all the used algorithms were trained on parts of the ImageNet dataset, the accuracy- results are only varying around 2% at all on the dataset. The best-performing algorithm Efficient- NetB7 has a Top1-accuracy of 27.50% for our dataset, which is less than one-third compared to the state-of-the-art accuracy of 84.3%.

The results for EfficientNet on the dataset are expected because it has shown the best accuracy in the state-of-the-art. Our results support the results of figure [1] and indicate, that even though, the EfficientNetB7 uses only 66 million parameters compared to the 89 millions of NasNet and 20 millions of DenseNet, the classification accuracy is 84.4% and thus 1.7% better than NasNet and 6.9% better than DenseNet, respectively when tested on ImageNet [[15].](#_bookmark14) It is surprising, that the differences between all networks are smaller than expected.

After the evaluation of the accuracy on the whole dataset, the images were segregated according to the ease of classification. For this purpose, each group member selected images from the dataset which, for instance, contain only one subject and should be easy to classify in his opinion. For the subset ”easy data” the 20 images with the most votes were selected.

Contrary to the expectations that a higher accuracy of the estimates can be achieved utilizing the subset, the accuracy decreased dramatically. Since most of the easy to classify pictures are portraits, the hypothesis was made that all CNNs perform especially badly on those pictures. To test this assumption, the dataset was divided into images of the class portraits and non-portraits.

The comparison of the results of all CNN's per category based on a normalized histogram, which can be seen in Figure 3, confirms this hypothesis. 0 indicates that no CNN could label the images correctly whereas 6 indicates all CNN would correctly label the images with their most likely prediction. It is recognizable, that the histogram bin for 0 has the highest percentage, especially for portraits, which means that the CNNs have performed over-proportional poorly for this.

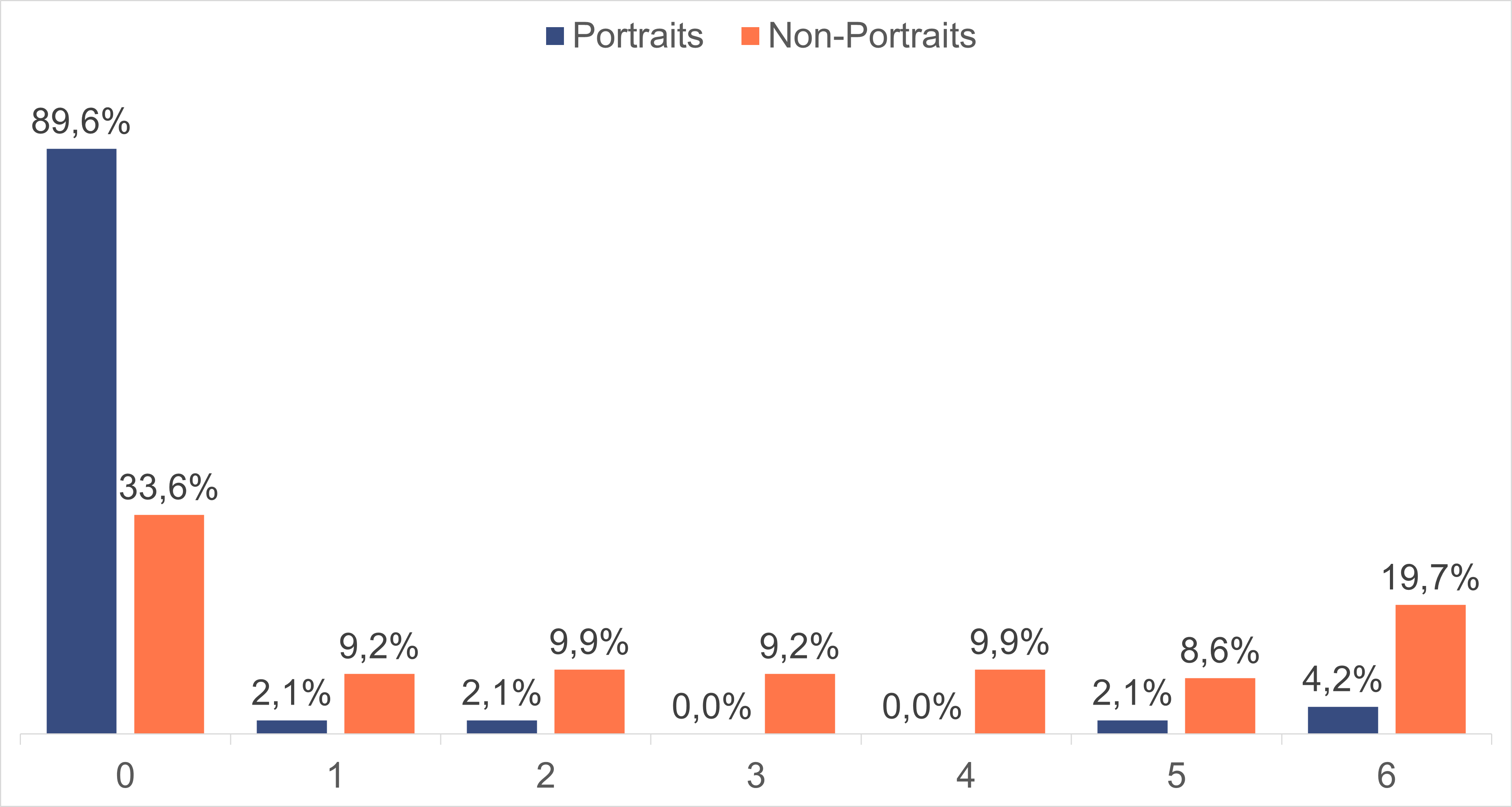


Fig. 3. Comparison by easy to classify images

1. CONCLUSION AND FUTURE WORK

Generally, it is concluded that the different Top-1 and Top-5 accuracy does not depend on the pure number of used parameters regarding their accuracy, rather than the architecture and implementation of neural networks.

The composition of the test dataset including

48 portrait photographs out of all 200 pictures explain the overall bad performance of all tested CNNs. Furthermore, all three CNNs are trained on different amounts of pictures but using the same labels and classes. This leads to the assumption, that portrait photographs are underrepresented in the used training parts of the ImageNet dataset. This statement could not be verified due to the fact, that there are over 14 million pictures labeled with

20.000 different classes.

For future work, it is interesting to evaluate the used CNNs on another less complicated dataset to get closer to the reference benchmark or find approaches that perform better on the used dataset. Based on the previous paragraph, the hypothesis according to the under-representation of portraits in the ImageNet database should be checked. Regarding this, a transfer learning with EfficientNet on portraits is a possible solution to achieve better accuracy on the used dataset. Furthermore, it would be interesting to have a look at the performance of a combined strategy including image classification and object detection algorithms.

REFERENCES

1. Jonathan Leban, “Image recognition with Machine Learning on Python, Image Processing” https://towardsdatascience.com/image-recognition- with-machine-learning-on-python-image-processing-3abe6b158e9a, May 22th, 2020.
2. ThinkAutomation, “ELI5: what is image classification in deep learning?” [https://www.thinkautomation.com/eli5/eli5-what-is-image-](http://www.thinkautomation.com/eli5/eli5-what-is-image-) classification-in-deep-learning/
3. Jason Brownlee, “14 Different Types of Learning in Machine Learning,”,https://machinelearningmastery.com/types-of-learning-in- machine-learning/, November 11th, 2019.
4. Artem Oppermann, “What is Deep Learning and How does it work?,”https://towardsdatascience.com/what-is-deep-learning-and-how- does-it-work-2ce44bb692ac, November 12th, 2019
5. Ksenia Sorokina, “Image Classification with Convolutional Neural Networks,” [https://medium.com/@ksusorokina/image-classification-](https://medium.com/%40ksusorokina/image-classification-) with-convolutional-neural-networks-496815db12a8, November 12th, 2017
6. Joseph Flynt, “The Growth and Evolution of Photography – Some Surprising Statistics,” https://3dinsider.com/photography-statistics/, Septem- ber 23rd, 2020
7. Barret Zoph, Quoc V., “Neural Architecture Search with Reinforcement Learning,”https://arxiv.org/abs/1611.01578, February 15th, 2017
8. Thomas Elsken, Jan Hendrik Metzen, Frank Hutter, “Neural Architecture Search: A Survey,”https://arxiv.org/abs/1808.05377, April 26th, 2019
9. Barret Zoph, Vijay Vasudevan, Jonathon Shlens, Quoc V. Le, “Learning Transferable Architectures for Scalable Image Recognition,”https://arxiv.org/abs/1707.07012v4, April 11th, 2018
10. Ayy¨ze Kizak, “Reviewing EfficientNet: Increasing the Accuracy and Robustness of CNNs,”https://heartbeat.fritz.ai/reviewing-efficientnet- increasing-the-accuracy-and-robustness-of-CNN's-6aaf411fc81d, January 17th, 2020
11. Mingxing Tan, Quoc V. Le, “EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks,”https://arxiv.org/abs/1905.11946,

September 11th, 2020

1. Pablo Ruiz, “Understanding and visualizing DenseNets,” https://towardsdatascience.com/understanding-and-visualizing- densenets-7f688092391a, October 10th, 2018
2. Gao Huang, Zhuang Liu, Laurens van der Maaten, Kilian

Q. Weinberger, “Densely Connected Convolutional Networks,” https://arxiv.org/abs/1608.06993, January 28th, 2018

1. Paperswithcode, “Image Classification on ImageNet,” https://paperswithcode.com/sota/image-classification-on-imagenet [15]Vardan Agarwal,“Complete Architectural Details of all EfficientNet

Models,” https://towardsdatascience.com/complete-architectural-details- of-all-efficientnet-models-5fd5b736142, May 24th, 2020

1. Keith D. Foote, “A Brief History of Deep Learning,” [https://www](http://www.dataversity.net/brief-history-deep-learning/).datav[ersity.net/brief-history-deep-learning/](http://www.dataversity.net/brief-history-deep-learning/) , February 7th, 2017
2. Google Developers, “Classification: Accuracy, ´´

https://developers.google.com/machine-learning/crash- course/classification/accuracy/, February 10th, 2020