Image Classification

The task of recognizing what an image means.

# Abstract

Recent work in the field of Machine Learning and Deep Learning has shown that Convolutional Neural Networks are interesting and promising to apply in different fields of use, for example to classify images of a photo library. In this research a test of three pre-trained CNNs on a dataset of 200 high-resolution images is explored, observed and compared. Therefore, 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. *Result…We observed that Conclusion… Our method…*

# Introduction

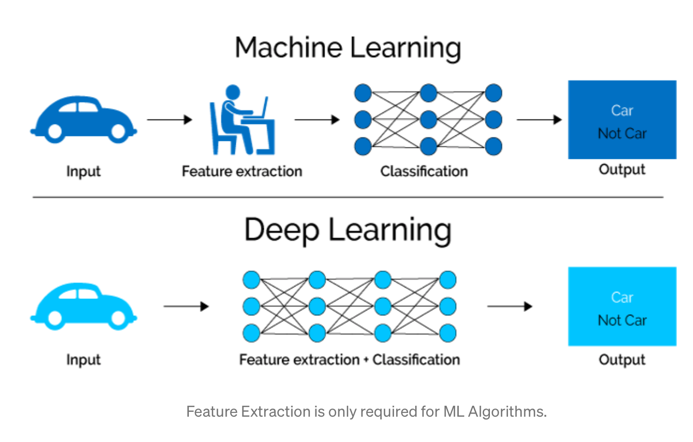
Every day, generally every person visually recognizes a high number of things in his environment. They recognize for example a thing with two wheels, a handlebar and a saddle and know it is a bike. People connect different information and combine them to a bigger result. For humans, this and also to establish a connection between a test and an image is an easier process, because they learn it in early years. For computers, this is not an easy task and furthermore image classification a big challenge. Computers need prior knowledge and have to learn to handle this process and therefore *Machine Learning* and *Deep Learning* is the chosen way.[[1]](#footnote-1)

The analysis of unstructured data is permanently a current topic and issue in the field of *Big Data* and *Deep Learning*. The analysis and extraction regarding to certain information in texts has made a further progress than the image analysis, respectively image classification. With Deep learning and image classification for example self-driving cars identify the things in their surroundings like other vehicles or people. Two other examples are the classification of symptoms of illness in the medical area or something easier like the image classification of photo libraries.[[2]](#footnote-2)

After a description of the current State-of-the-Art in the field of Deep Learning, especially Image Classification, the aim of the paper is described. As basis of the Evaluation chapter the Fundamentals of the three chosen Convolutional Neural Networks are presented. The Evaluation part explains the process of testing the CNNs on the given dataset and the test-result. The last part is drawn a conclusion and describes future work.

# State-of-the-Art

As part of artificial intelligence, *machine learning* in for example computers basically should enable to recognize patterns in given databases to find and create solutions for the given problem or to analyze the data independently.   
Machine learning is divided in different types of learning algorithms: supervised learning, unsupervised learning and reinforcement learning. Supervised learning is divided in classification and regression. The classification supervised learning “[…] involves predicting a class label.”[[3]](#footnote-3)  
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 and models. The base for this are the neural networks, which are getting trained during the learning process. The structure of the neural network is oriented on the structure of the human brain. They have a multi-layered structure and consisting of neurons. The neurons are combined in a series of groups. In neural networks, it “is a multi-layered structure of algorithms.”[[4]](#footnote-4) The “Neurons in each layer are connected to neurons of the next layer.”[[5]](#footnote-5) As is the case with people, the more linkages there are, the more and more complex issues can be solved and mapped.

******

[***https://towardsdatascience.com/what-is-deep-learning-and-how-does-it-work-2ce44bb692ac***](https://towardsdatascience.com/what-is-deep-learning-and-how-does-it-work-2ce44bb692ac)

In the neural network, the given Input is processed to different layers of nodes. The nodes “communicate their results to the next layer of nodes.”[[6]](#footnote-6) Deep Learning involves different types of neural networks. For image classification, Convolutional Neural Networks, CNNs, are often used.   
 Unlike the classical approach, the nodes do not always share their result to every node in the next layer. These layer is “known as convolutional layers.”[[7]](#footnote-7)   
The advantage of this Deep learning is, that algorithms get better with an increasing amount of data. Classical models of machine learning stop their classification after a saturation point. [[8]](#footnote-8)

With image classification, it is possible to capture an entire image. For humans, it is an image. For a computer, it is an array with a defined number of pixels with a height, width and number of color channels. The main goal is it to classify the image by allocate it to a set label. In contrast to “Object detection” only the object without his location is interesting for the classification. The State of the Art of CNNs is based on the accuracy, how “good” the object in an image, for example a woman, a cat or a plant, is classified and matching with the assigned label.[[9]](#footnote-9)

# Evaluated Networks

This section discusses the three different Convolutional Neural Networks that are considered for evaluation. Firstly, the fundamentals of the three considered networks is described in brief. Following that, we present the evaluation of these three networks which consists of detailing the groundtruth generation, pre-processing and final evaluation of the networks. The presented algorithms are used in order to achieve a good classification, comparable to the Top-1 accuracy and Top5-accuracy on ImageNet.

## Fundamentals

### NASNetLarge

NAS is the short version for Neural architecture search. It is basically a technique to automate the design of artificial neural networks, short version ANN. It is used to develop networks, that perform better than hand-created architectures.[[10]](#footnote-10) Corresponding methods can be distinguished between “[…] search space, search strategy, and performance estimation strategy.”[[11]](#footnote-11)  
The *NASNet Search Space* enables a transferability. With the actual NASNet from 2018, Barret Zoph, Vijay Vaseduan, Jonathon Shlens and Quoc V. implemented a new regularization technique, 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 includes 28% less computational demand compared to the previous model.[[12]](#footnote-12) Thereby the NasNet improves the process time of learning on large datasets by applying a block for small datasets on larger datasets.[[13]](#footnote-13)

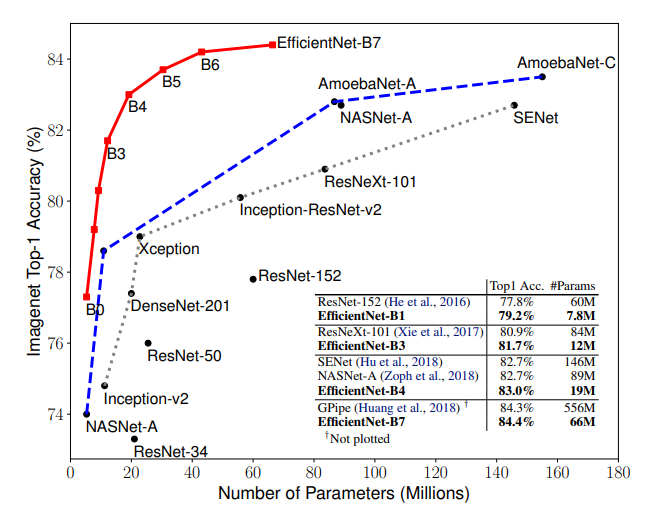
### EfficientNet B7

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.[[14]](#footnote-14) The result of EfficientNet B7 is a much better accuracy and efficiency. This is done by insert a new scaling method, that is applied on the dimensions depth, width and resolution.   
The convolution is divided into depthwise and pointwise convolution. This to stairs reducing 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.[[15]](#footnote-15)  
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 exisiting ConvNet.”[[16]](#footnote-16)

### DenseNet 201

|  |  |  |  |
| --- | --- | --- | --- |
|  | 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% |
| CIFAR | CIFAR-10: 2.4% Error rate | CIFAR-100:  91.7% accuracy | (DenseNet) CIFAR-100: 82.62% accuracy |

The Densly 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.”[[17]](#footnote-17) 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.”[[18]](#footnote-18) This is achieved by a directly 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 less 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%.[[19]](#footnote-19)



In this paper a dataset of 200 high-resolution images is used. The aim of the paper is to analyse and evaluate three image classification algorithms, which are based on objective reason, 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 on a given dataset.   
After that, the classified images of each classification algorithm should be analyzed and evaluated.

# Evaluation

## Ground Truth

The ground truth is a dataset with 200 images with up to 20 labels per image. (example image of table) For each image, one label is selected to test the Top 1 accuracy.

We segregated the pictures according to ease of classification. The decision was made in accordance to the simplicity of the image and the things represented in it. When comparing the accuracy for full data with the easily classified data, we see an unexpected deterioration in top 1 and top 5 accuracy of the easily classified data.  
We can conclude that the network has poor accuracy for portraits since 48 out of 200 images in the dataset are portraits and they were mostly sub grouped as easy to be classified. We proved this hypothesis by measuring the performance between the classes of portraits and non-portraits. In the graph, 0 indicates that no CNN could label the images correctly whereas 6 indicates all CNN would correctly label the images. We can see that the histogram bin for 0 has the highest percentage. Which means that the CNNs have performed poorly for portraits.

## Pre-Processing

The height and width of an image are based on number of pixels. One pixel corresponds to one point of the image and involves an information about the grayscale, in case of an black-white image, or RGB(a)-color with three integers between 0 and 255, where 0 is black and 255 white (and the opacity of the image).[[20]](#footnote-20)

All CNNs are applied on 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.

## Test procedure

### NasNet

The NasNet requires a minimum dimension of 32 x 32 pixels.

TensorFlow based

tf.keras.application

# Results

Table with original results from the paper and the new results

Further analysis by considering the “easy to classify” images

Table

# Conclusion and Future work

- What have we learned from our results?

In this paper, we classified a dataset containing of 200 high dimensional images. Three different classification algorithms were used and the corresponding results can be seen in the FIGURE [“insert the fig number of the results”]. The performance metric, that is being used is accuracy, namely, Top1-accuracy and Top5-accuracy. As can be seen from the fig, all the algorithms “underperformed” in terms of the performance metric. This is mainly attributed to the fact, that all the 200 images were used to test the algorithms. All the used algorithms were trained on other datasets for instance on the ImageNet dataset. The highest performing algorithm was EfficientNetB7 with a Top5-accuracy of 44.50% and Top1-accuracy of 27.50% while the least underperforming is NasNetLarge.

These results are expected, because EfficientNet has shown better accuracy in general with EfficientNetB7 as the best performing among the seven algorithms in the EfficientNet-family. EfficientNetB7 has a total number of 813 layers. Our results seconds the results of [“Reference I have added ”] figure [2] and indicate, that even though, the EfficientNetB7 uses only 66 million parameters compared to the 89 million of NasNet and 20 million 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.[[21]](#footnote-21)

Future work:

1. Using a less complicated dataset

# References

1. <https://towardsdatascience.com/image-recognition-with-machine-learning-on-python-image-processing-3abe6b158e9a> [↑](#footnote-ref-1)
2. <https://www.thinkautomation.com/eli5/eli5-what-is-image-classification-in-deep-learning/> [↑](#footnote-ref-2)
3. <https://machinelearningmastery.com/types-of-learning-in-machine-learning/> [↑](#footnote-ref-3)
4. <https://towardsdatascience.com/what-is-deep-learning-and-how-does-it-work-2ce44bb692ac> [↑](#footnote-ref-4)
5. <https://medium.com/@ksusorokina/image-classification-with-convolutional-neural-networks-496815db12a8> [↑](#footnote-ref-5)
6. <https://www.thinkautomation.com/eli5/eli5-what-is-image-classification-in-deep-learning/> [↑](#footnote-ref-6)
7. <https://www.thinkautomation.com/eli5/eli5-what-is-image-classification-in-deep-learning/> [↑](#footnote-ref-7)
8. <https://towardsdatascience.com/what-is-deep-learning-and-how-does-it-work-2ce44bb692ac> [↑](#footnote-ref-8)
9. <https://paperswithcode.com/task/image-classification> [↑](#footnote-ref-9)
10. <https://arxiv.org/abs/1611.01578> [↑](#footnote-ref-10)
11. <https://arxiv.org/abs/1808.05377> [↑](#footnote-ref-11)
12. <https://arxiv.org/abs/1707.07012> [↑](#footnote-ref-12)
13. <https://arxiv.org/abs/1707.07012v4> [↑](#footnote-ref-13)
14. <https://heartbeat.fritz.ai/reviewing-efficientnet-increasing-the-accuracy-and-robustness-of-cnns-6aaf411fc81d> [↑](#footnote-ref-14)
15. <https://heartbeat.fritz.ai/reviewing-efficientnet-increasing-the-accuracy-and-robustness-of-cnns-6aaf411fc81d> [↑](#footnote-ref-15)
16. <https://arxiv.org/abs/1905.11946> [↑](#footnote-ref-16)
17. <https://towardsdatascience.com/understanding-and-visualizing-densenets-7f688092391a> [↑](#footnote-ref-17)
18. <https://arxiv.org/abs/1608.06993> [↑](#footnote-ref-18)
19. <https://paperswithcode.com/sota/image-classification-on-imagenet> [↑](#footnote-ref-19)
20. <https://towardsdatascience.com/image-recognition-with-machine-learning-on-python-image-processing-3abe6b158e9a> [↑](#footnote-ref-20)
21. Agarwal, V. (2020, May 31). *Complete Architectural Details of all EfficientNet Models*. Medium. https://towardsdatascience.com/complete-architectural-details-of-all-efficientnet-models-5fd5b736142 [↑](#footnote-ref-21)