TU Chemnitz

Faculty of Natural Sciences
Institute of Physics

Bachelor Thesis

In the course of degree in computational science (B.Sc.)

To obtain the academic degree Bachelor of Science

Topic:	Investigation of strategies for image classification on small training data sets
Author:	Björn Hempel <bjoern@hempel.li> Matriculation number 025038
Version from:	February 1, 2020
First assessor:	Prof. Dr. Angela Thränhardt
Second assessor:	Dr. David Urbansky

Abstract

Neural networks are extraordinarily good at finding patterns in data. For this purpose, these networks must be trained with known data sets and adapted accordingly. Data sets are usually very expensive to obtain and should therefore be used with care and good quality. The training of the network takes place under many different parameters and process techniques. Care must be taken to use the best possible model with its best possible parameters. In this thesis, common modern methods of image classification will be presented and compared with each other. The main goal of the work is to find optimal parameters and techniques for the classification, which also allow to create an optimal model with little training data.

Contents

1 Introduction

In this thesis different techniques of image classification are compared. Variable parameters during training will have a decisive influence on the accuracy of the model and are compared here in detail. Not always only the accuracy is a decisive factor. Also the required computing time, which is necessary to determine the model, should not be disregarded and should be included in the evaluation. I assume that a small learning rate combined with many learning epochs and correspondingly more computing time required will achieve better results than a few learning epochs combined with a high learning rate (slow adaptation vs. fast adaptation). I also assume that a high quality and a larger amount of data will have a decidedly positive influence on the result. New and more complex convolutional neural networks are more successful in model accuracy than older and smaller models.

2 Background

2.1 Image Classification

Classifications are a process of identifying to which class an unobserved object belongs. A number of predefined classes can be specified and, based on their properties, an attempt can be made to classify unknown and previously unobserved objects. The procedure for image classification is similar. The previously mentioned objects are now simply images.

Figure 1: Is it a dog or a cat?[1]

For a long time, the automatic recognition of objects, people and scenes in images by computers was considered impossible. The complexity seemed too great to be programmatically taught to an algorithm. Until a few decades ago, attempts were made to achieve image classification by manually developed algorithms. Automated classification based on given and pre-classified images and the automated creation of models was a new step into a new approach. The neural networks developed in this process played a huge role and dramatically changed the way of approach! In the meantime, image recognition has become a widespread application area of machine learning. So-called "Convolutional Neural Networks[2]" or "ConvNets" are often used for images.

The image classification algorithm takes an image as input and classifies it into one of the output categories. Deep Learning has revolutionized the field of image classification and has achieved great results. Various Deep Learning networks, such as ResNet, DenseNet, Inception, etc. have been developed as high-precision networks for image classification. At the same time, image data sets were created to capture tagged image data. These are now primarily used to train existing networks and to organize annual challenges that compete with the model accuracies already known and developed. ImageNet is such a large data set with more than 11 million images and over 11,000 categories. Once a network has been trained with ImageNet data, it can be generalized with other data sets by simple re-compilation or optimization. In this transfer learning approach, a network is initialized with weights that come from a previously trained network. This previously initialized network is now simply adapted for a new image classification task.

The underlying work here is mainly concerned with supervised learning, in which a mathematical model is trained based on existing known data sets. The goal of the trained model is to make best

---

[1]Source: https://towardsdatascience.com/image-classifier-cats-vs-dogs-with-convolutional-neural-networks-cnns-and-google-colabs-4e9af21ae7a8

[2]Convolutional neural network, Wikipedia contributors, January 31, 2020, https://en.wikipedia.org/wiki/Convolutional_neural_network

possible predictions even for unknown images. These known data sets are usually created manually (ontologist), automatically determined based on known facts or determined in a semi-automatic process.

2.1.1 Deductive approach

Since the late 1960s, attempts have been made to classify images with self-written algorithms. This part of Computer Vision deals with techniques such as image creation, image processing and image segmentation. In the field of image processing, well-known techniques such as edge detection, feature detectors, edge linking, contrast enhancement, etc. are used[3]. Common to all techniques is the use of the deductive approach. With the deductive approach, one creates rules (feature detectors) which are supposed to predict the desired result. These rules are given and described and thus allow later classification of unknown objects. Since the model and its algorithm are sufficiently well known, this procedure is called white-box procedure.
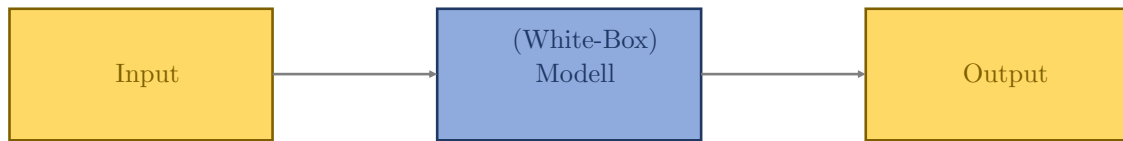
Figure 2: Deductive approach

2.1.2 Inductive approach

The inductive approach, on the other hand, takes a different approach to classifying images. The goal is not to specify a rule, but to learn a rule (model) automatically from already known individual objects. A model is usually a complex function and a mathematical representation of a space (VC dimension[4]), in which individual objects with their properties can be mapped and separated. The model is adapted piece by piece to the known objects in such a way that the input value corresponds to the output value or corresponds to a large extent (backpropagation). The goal is to create a function with this model, which is able to classify unknown objects in the best possible way. Because the space of this model is mostly far away from the imagination and the possibility of explanation, this procedure is also called black box procedure. The procedure described here is mostly used for any kind of supervised learning and is a part of machine learning.
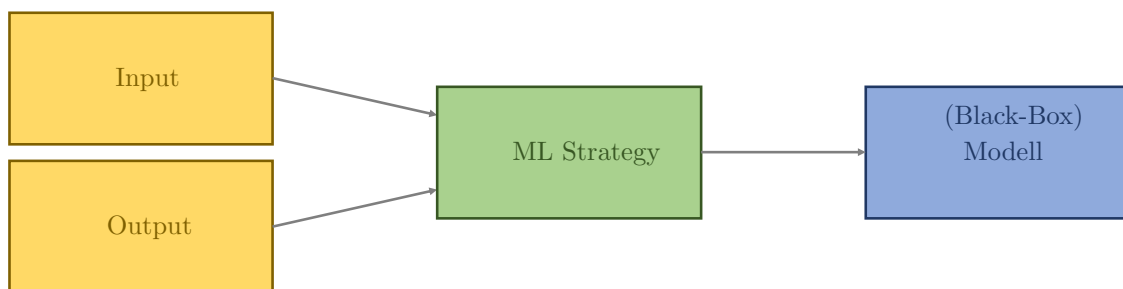
Figure 3: Inductive approach

2.1.3 Balanced training data set

Neural networks have made enormous progress in the field of pattern recognition in recent years. A decisive factor is that the data for learning must be of high quality and easy for the network to

---

[3]Szeliski, R.: Computer Vision: Algorithms and Applications, Springer Science Business Media, 10 (2010)

[4]Vapnik–Chervonenkis dimension, Wikipedia contributors, January 31, 2020, `https://en.wikipedia.org/wiki/Vapnik%E2%80%93Chervonenkis_dimension`

process. Wrongly classified or irrelevant data could cause the network to learn something wrong. This also applies to non-existent or unsuitable pre-processing.[5]

With the beginning of a classification project the question is what exactly you want to classify and how extensive the classification should be. Assuming you want to identify different classes of food, this could be classes like pizza, burgers, donuts and lasagna (etc.). For these classes you now need a large number of images. Ideally, this data should reflect reality as well as possible. A large variation is advantageous (balanced data set): different viewing angles, size, position, colour brightness, variations, number, etc. Images of e.g. only one colour brightness or only one viewing angle should be avoided. If the data are not balanced, they must be corrected accordingly: e.g. by adding further data, image processing or by removing data that causes an imbalance. Furthermore, the selected classes should be clearly optically separable from each other. If two classes are visually very similar and not really distinguishable even by a human, consideration should be given to combining them (e.g. "burger" and "veggie burger"):

Figure 4: Example pictures of a burger class

Figure 5: Example pictures of a donut class

Figure 6: Example pictures of a pizza class

Accessing data is often not that easy. Every data source has its own special features. One way to access data would be an automatic crawling of image databases, search engines or reviews in which images appear. A certain amount of creativity is advantageous:

- Google

- Bing

- Flickr

- TripAdvisor

- ...

Probably the most expensive way to obtain data is to search and classify them manually, e.g. by an ontologist. The ontologist evaluates and searches for different images and manually classifies them in the appropriate classes. A combined variant is also possible and probably preferable: automatic crawling and manual sorting out of incorrect, unfavorable or irrelevant images.

---

[5]Douwe Osinga. *Deep Learning Kochbuch: Praxisrezepte für einen schnellen Einstieg.* O'Reilly Verlag, 2019, pp. 22–26. ISBN: 9783960090977.

2.1.4 Training, test and evaluation data set

Before starting the training of balanced images, they must be divided into a training, a test and possibly a validation data set. This is necessary because neural networks will not generalize to some extent, but will learn by heart (overfitting[6]). The idea is to train with a training data set, while the validation data set is used to monitor the general validity of the network and its parameters. Based on the results, adjustments are made at runtime. Since the adjustment of the parameters is carried out using the test data, there is also an independent test data set, which carries out a renewed check of the model for previously uninvolved data. This ensures that hyperparameters are not inadvertently optimized for the validation data set only.[7] The use of the test data set is optional and simulates the model under real conditions. If the number of data is limited, this data record can also be added to the training data record, for example. In this thesis, the test data set is not used and all evaluations refer to the validation data set.

An optimal division of the training and validation data set depends on the existing classification problem and the amount of data available.

2.1.5 Unsupervised Classification

Unsupervised learning refers to machine learning without knowing in advance known target values. This paper will not go into this technique any further.[8]

...

2.1.6 Supervised Classification

Supervised learning refers to machine learning with known training data sets. Learning refers to the ability of an artificial intelligence to reproduce regularities and patterns. The results are known by laws of nature or expert knowledge and are used to teach the system. The learning algorithm tries to find a hypothesis that makes the most accurate predictions possible. A hypothesis is an image that assigns the assumed output value to each input value. This paper makes extensive use of supervised learning.[9]

...

2.1.7 Classification Metrics

...

2.1.7.1 Top-1 accuracy

The **top-1 accuracy** is the most important accuracy, which means that the model response (the one with the highest probability) must be exactly the expected response.

...

2.1.7.2 Top-5 accuracy

...

2.2 Machine Learning

Maschinelles Lernen ist ein Oberbegriff für die künstliche Generierung von Wissen aus Erfahrung. Es verfolgt den Ansatz des induktiven Lernens.

---

[6]Overfitting, Wikipedia contributors, January 31, 2020, https://en.wikipedia.org/wiki/Overfitting
[7]Osinga, *Deep Learning Kochbuch: Praxisrezepte für einen schnellen Einstieg.*
[8]Aurélien Géron. *Praxiseinstieg Machine Learning mit Scikit-Learn und TensorFlow: Konzepte, Tools und Techniken für intelligente Systeme.* O'Reilly Verlag, 2017, pp. 8–14. ISBN: 9783960090618.
[9]Ibid.

2.2.1 Neuronal Network

Explain the development of a neuronal network here...

In künstlichen neuronalen Netzen bezeichnet die Topologie die Struktur des Netzes. Damit ist im Allgemeinen gemeint, wie viele künstliche Neuronen sich auf wie vielen Schichten befinden, und wie diese miteinander verbunden sind. Künstliche Neuronen können auf vielfältige Weise zu einem künstlichen neuronalen Netz verbunden werden. Dabei werden Neuronen bei vielen Modellen in hintereinander liegenden Schichten (englisch layers) angeordnet; bei einem Netz mit nur einer trainierbaren Neuronenschicht spricht man von einem einschichtigen Netz.

Unter Verwendung eines Graphen können die Neuronen als Knoten und ihre Verbindungen als Kanten dargestellt werden. Die Eingaben werden gelegentlich auch als Knoten dargestellt.

Die hinterste Schicht des Netzes, deren Neuronenausgaben meist als einzige außerhalb des Netzes sichtbar sind, wird Ausgabeschicht (englisch output layer) genannt. Davorliegende Schichten werden entsprechend als verdeckte Schicht (englisch hidden layer) bezeichnet.

Let us imagine a simple classification function, where $\bar{x}$ represents the coordinates of the individual class points and $\bar{w}$ and b are learnable parameters:

$$f(\bar{x}, \bar{\omega}, b) = sgn(\bar{\omega}^{\mathsf{T}} \cdot \bar{x} + b) \tag{1}$$
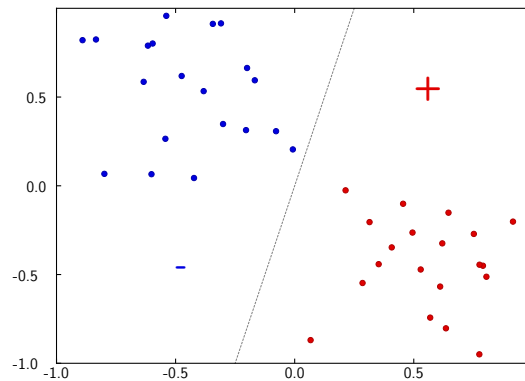

Figure 7: Simple class shattering

With this linear function it is easy to classify the above problem. The function corresponds to a neural network without hidden layer and contains only one input and one output layer. The dimension that this function can separate is 2 and is called VC dimension[10]. This function can separate exactly 2 classes. But what about nonlinear problems? In this case let's look at the following classification:

---

[10]Vapnik–Chervonenkis dimension: `https://en.wikipedia.org/wiki/Convolutional_neural_network`
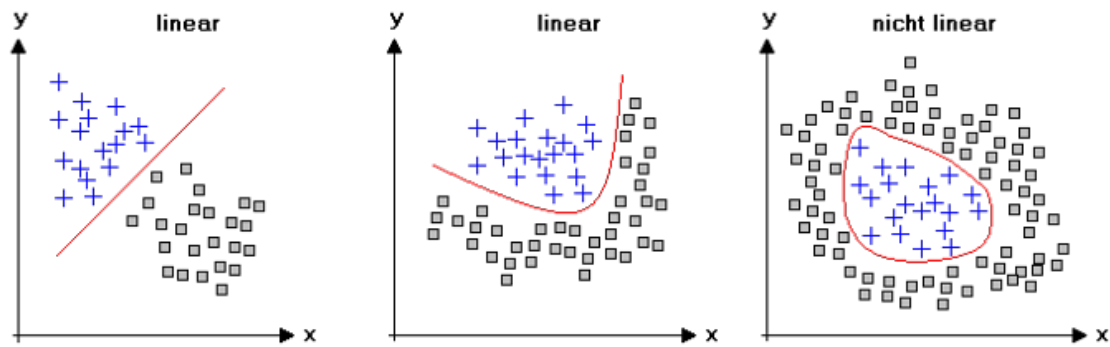
Figure 8: Linear vs. non-linear classificatioin[11]

For the second problem we can still adjust the function. For the third nonlinear problem, the classification space is no longer sufficient and requires a different algorithm.
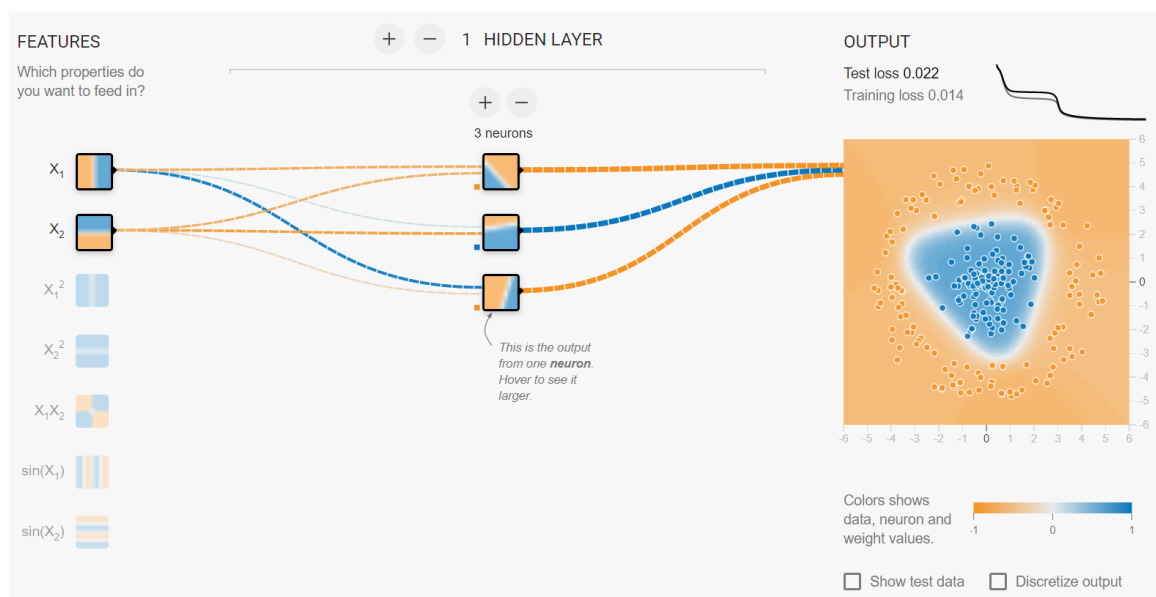

Figure 9: Simple neuronal network with one hidden layer[12]

2.2.2 Convolutional Neuronal Network

A neural network processes a vector. A convolutional neural network, on the other hand, can process a matrix. Describe more here.

2.2.3 Transfer Learning

Explain here the technique of Convolutional Neuronal Networks.

2.2.3.1 Transfer Learning Models

...

---

[11]Source: `http://www.statistics4u.info/fundstat_germ/cc_linvsnonlin.html`
[12]Source: `http://playground.tensorflow.org/`
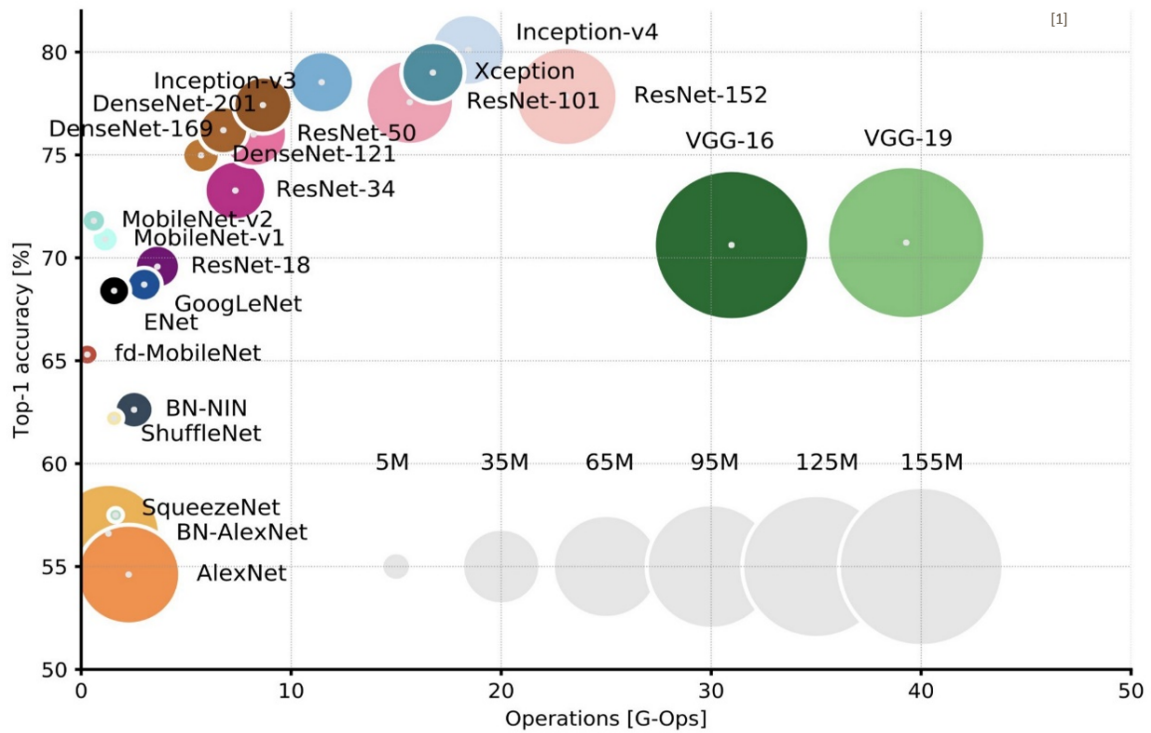
Figure 10: Overview of known transfer learning models.[13]

<mark>...</mark>

2.3 Related work

This is an example citation.[14]

3 Validation process

This is the part where I explain my approach.

3.1 Preamble

In the following, the best possible accuracy is to be achieved by testing various parameters. A learning set with the following properties was used:

- 14865 images

- classified within 50 classes

- different number of images per class (unbalanced)

With the exception of the model tests, all tests were based on the following parameters (whereby one value of the parameters varied depending on the chapter):

- model: resnet18

- learning rate: 0,001 (decreases every 7 epochs to 10% of the previous value)

---

[13]Source: https://towardsdatascience.com/neural-network-architectures-156e5bad51ba

[14]Jia Deng et al. "What does classifying more than 10,000 image categories tell us?" In: *European conference on computer vision*. Springer. 2010, pp. 71–84.

- batch size: 48

- epochs: 21 (learning rate from epoch 15 to 21: 0,00001)

- image size: 224x224 pixels

- the entire training and validation set (14865 images)

Different models were tried out in chapter ≪**??**[15]≫ with the same parameters as above:

- ResNet18

- ResNet50

- ResNet152

- AlexNet

- VGG

- SqueezeNet

- DenseNet

- Inception v3

3.2 Working environment

<mark>Explain in this part of the thesis the frameworks, environments and hardware used, etc.</mark>

3.3 Splitting and preparing the data

3.3.1 Situation

We have 14866 images differently distributed in 50 classes (unbalanced). We would like to divide these into 80% training and 20% validation images.

3.3.2 Unbalanced

The unbalanced dispersion data set is divided exactly in the same ratio:

- 2953 images for the training

- 11913 images for validation

For training with different training elements, the validation dataset of 2953 images is retained for a comparable result. The number of training elements deviating from the total data set results from this:

$$n_{train} = k \cdot 500; k \in 1 \ldots 26 \tag{2}$$

3.3.3 Balanced

...

3.4 Performance

...

---

[15]see on page **??** chapter **??** ≪**??**≫

3.5 Accuracy and evaluations

...

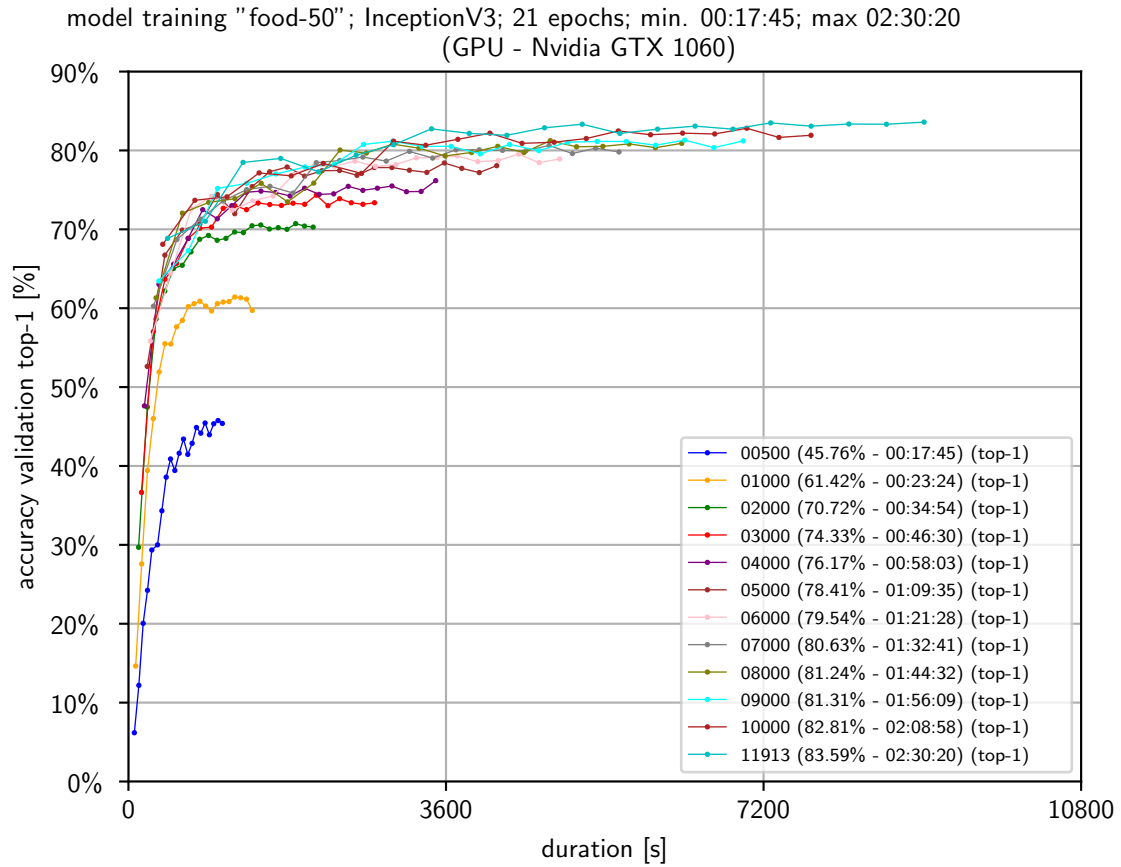### 3.5.1 Influence of number of trained images on accuracy

...

Figure 11: Overview of influence of number of trained images on accuracy

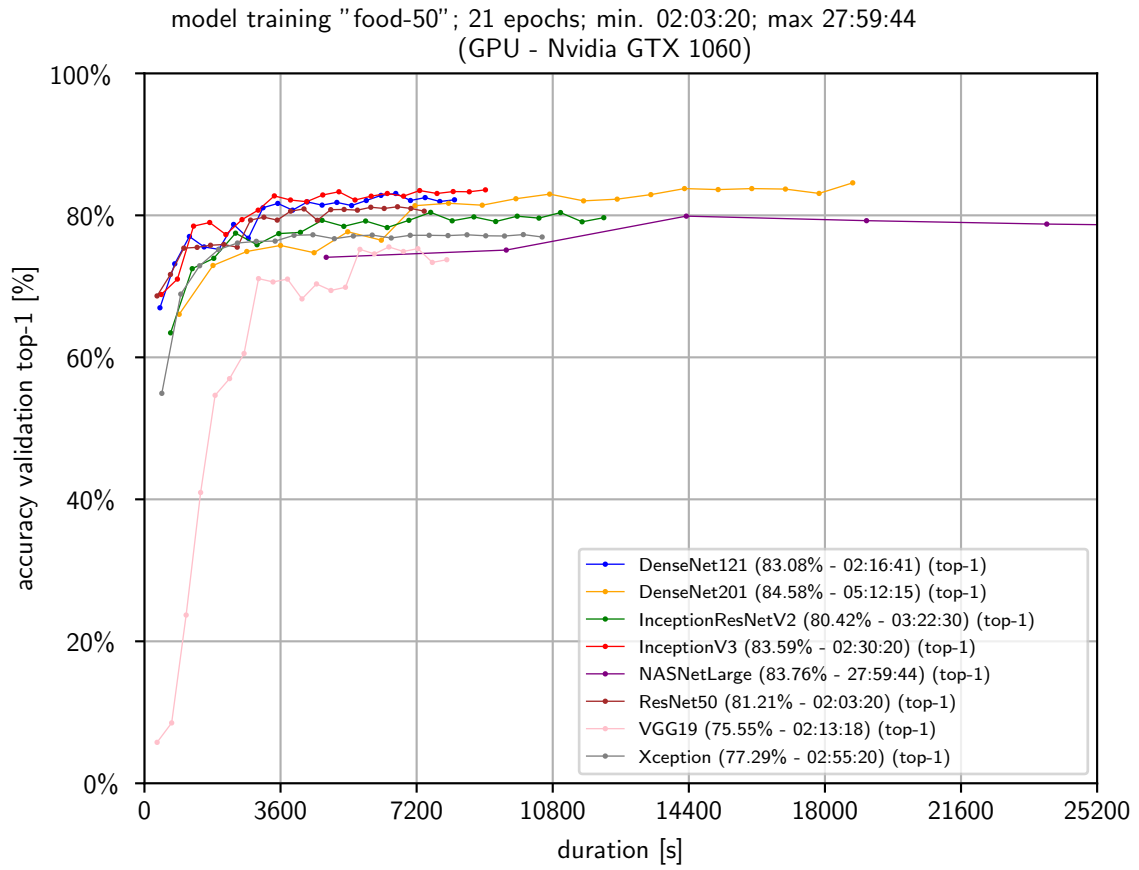...

3.5.2 Comparison of different CNN models

...

Figure 12: Overview of known transfer learning models

...

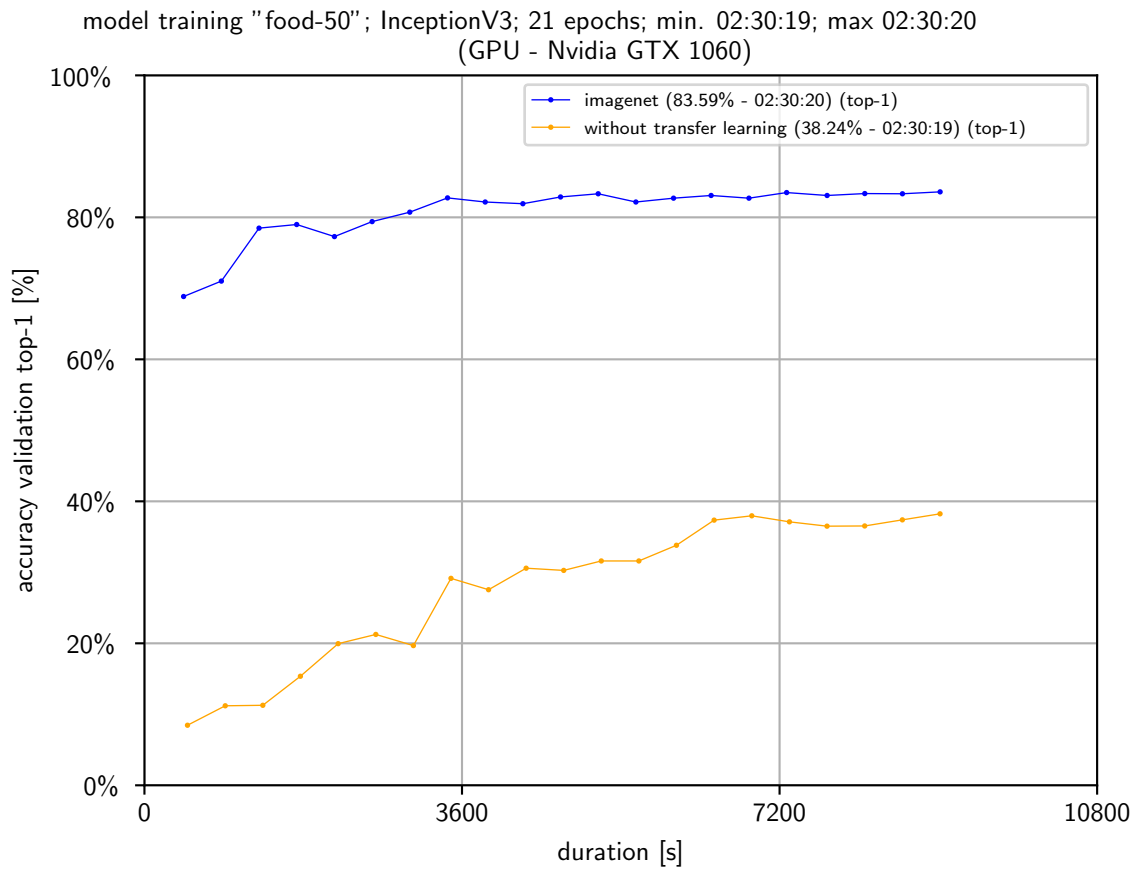### 3.5.3 Use of the transfer learning approach

...

Figure 13: Overview of use of the transfer learning approach

...

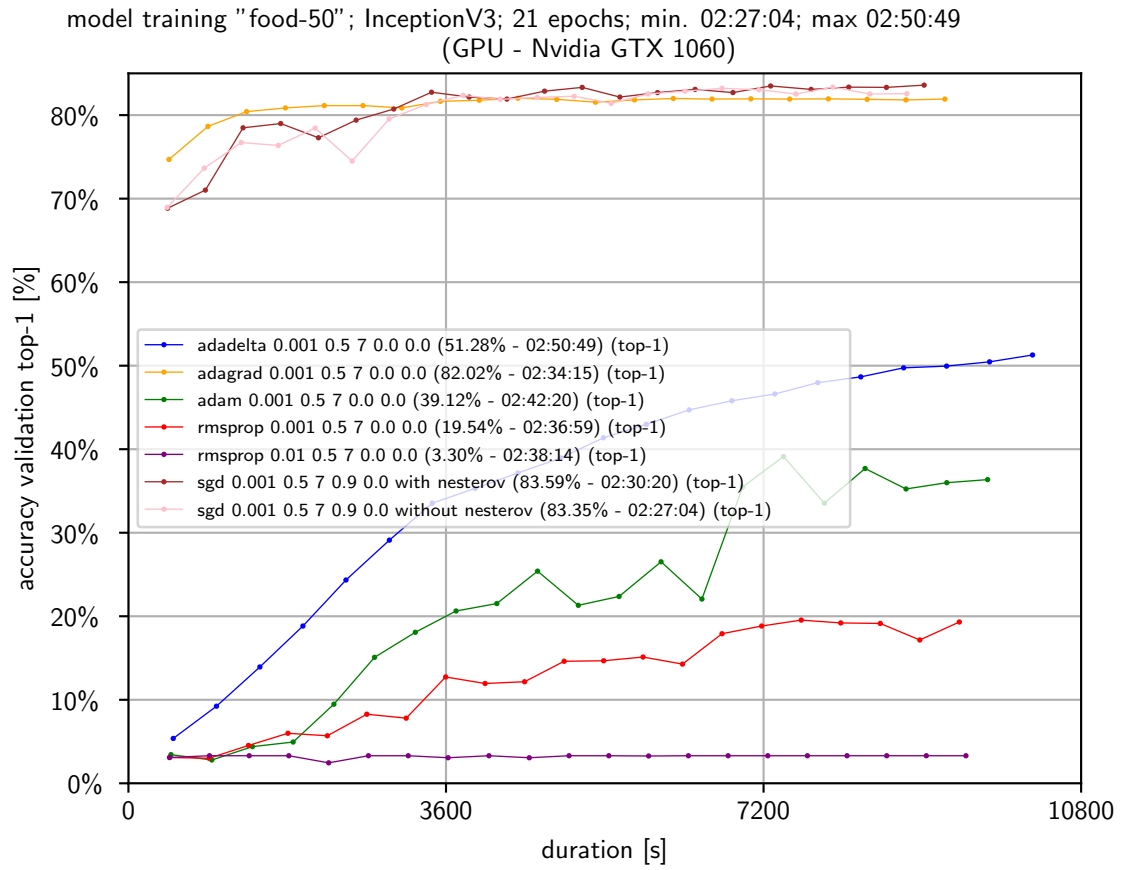### 3.5.4 Influence of different error optimizers

3.5.4.1 Comparison Optimizer

...

Figure 14: Overview of best optimizer

...

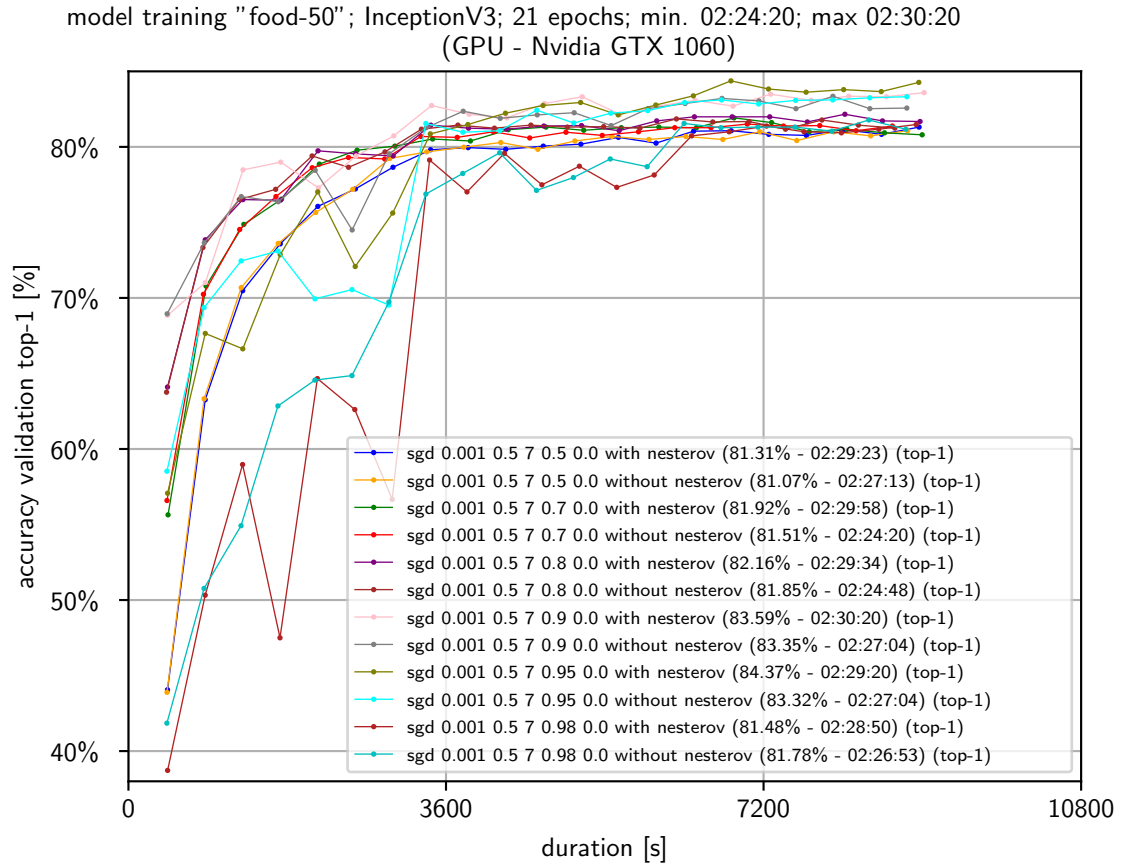### 3.5.4.2 Influence of the momentum and the Nesterov momentum

...

Figure 15: Overview momentum vs nesterov momentum

...

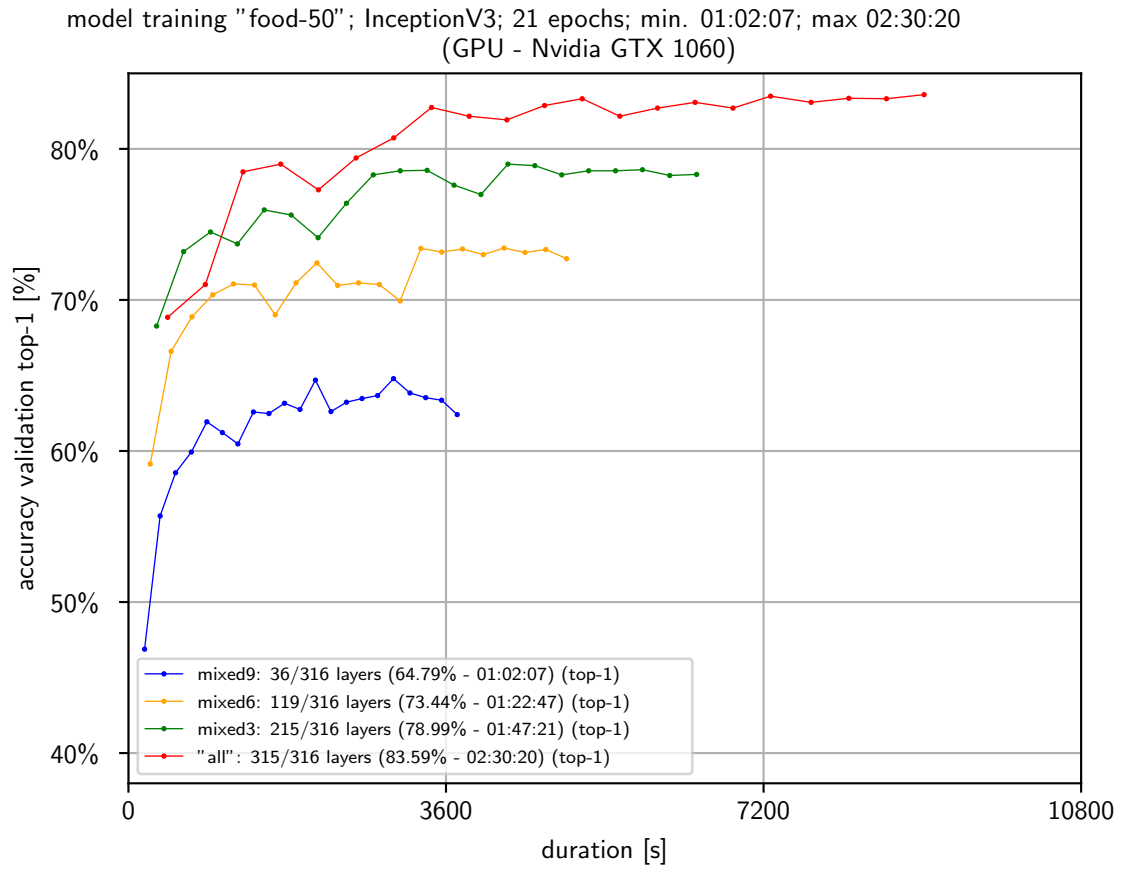### 3.5.5 Influence of the number of trained layers on the accuracy

...

Figure 16: Overview of influence of the number of trained layers

...

3.5.6 Influence of a dynamic learning rate on accuracy (scheduling)
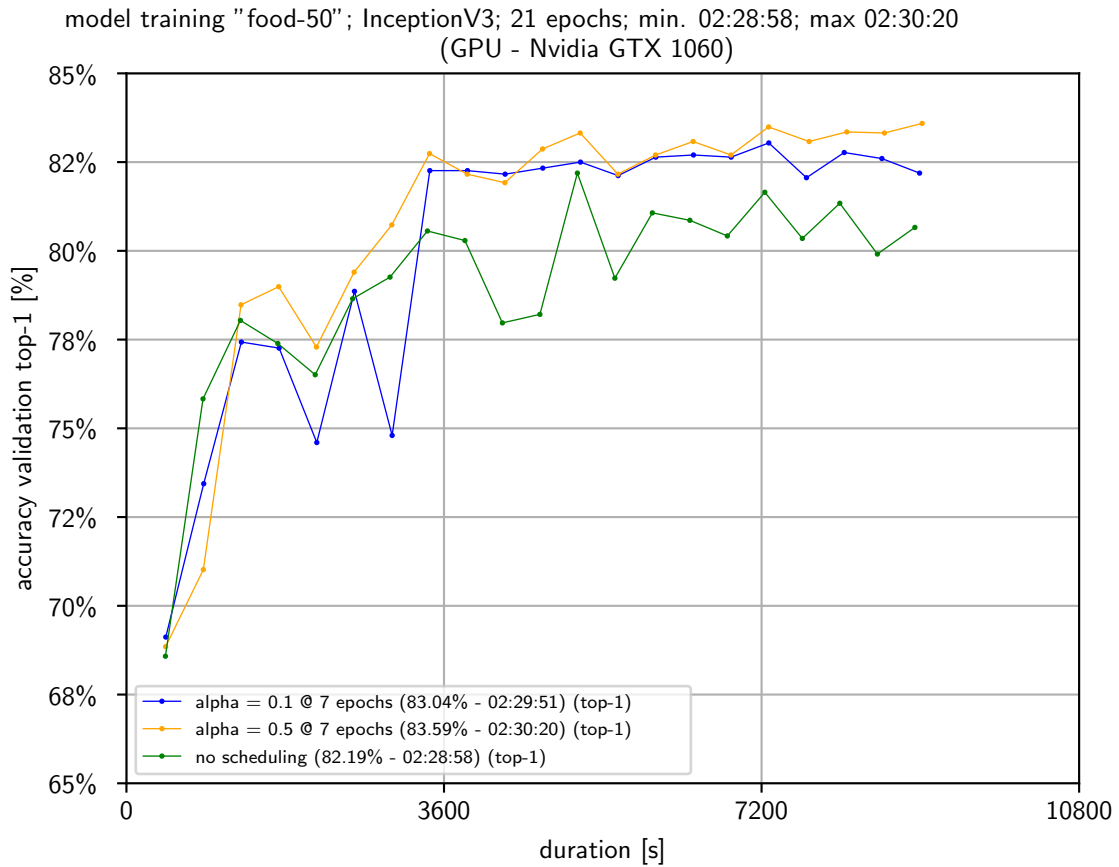
...

Figure 17: Overview of a dynamic learning rate on accuracy

...

3.5.7 Different batch sizes

...

3.5.8 Different image sizes

...

3.5.9 Different number of learned epochs

...

4 Optimization process

This chapter contains ideas, approaches and evaluations of more complex ideas, which do not fit into the range of simple parameter changes.

4.1 Preamble

...

4.2 Data augmentation

...

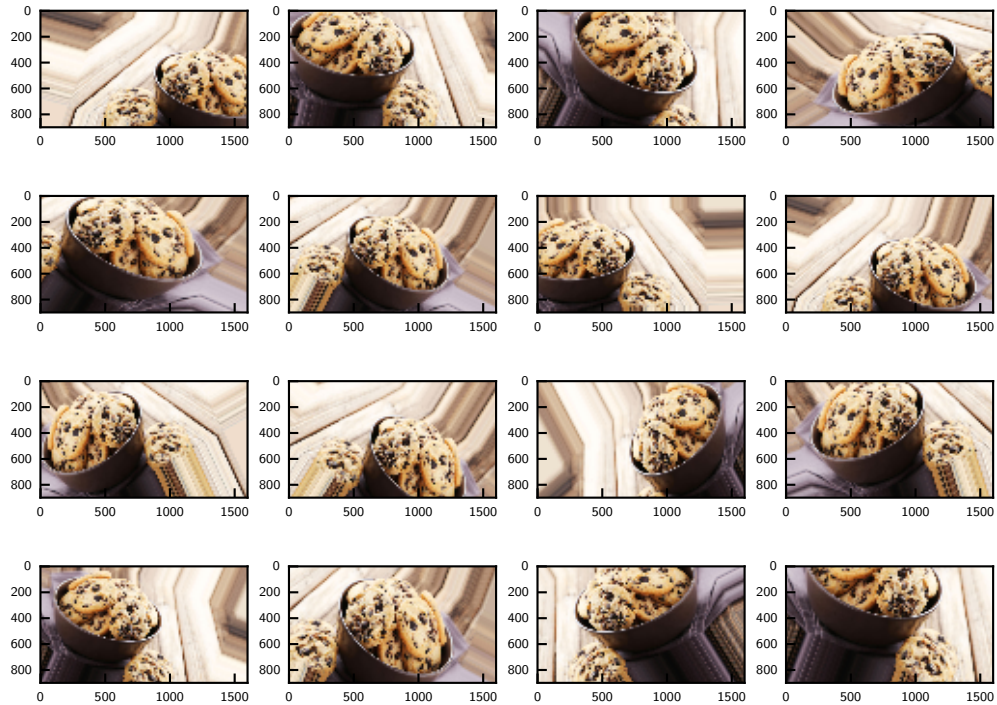F:/data/raw/food-50-augment/_other/chocolate-chips-cookies-american-cookies.jpg

Figure 18: Data Augmentation

4.3 Enrichment of the data set from other data sources

...

4.4 Analyses with multidimensional scaling

...

4.5 Hierarchical classification

By using a single model for all classes, previous classifiers have been trained to minimize the loss of the class output vector. Each class used so far has the same rank in both training and classification. The prediction of "Pizza" costs the same as the prediction of "Martini".

The human ability to classify objects does not only work on one level. Categories will naturally overlap and have a hierarchical structure. For example, a human will classify a picture under "pizza", "tuna pizza" or even "fast food", which is correct from this point of view. Depending on the classification, there will only be a "loss of information". However, a person will not mistake a "pizza" as a "Martini", which is more likely to be classified as a "drink" or "cocktail".[16]

---

[16]Eleanor Rosch et al. "Basic objects in natural categories". In: *Cognitive psychology: Key readings* 448 (2004).

4.6 Binary classifiers

...

4.7 Evaluation

...

4.8 Use of the model across programming languages

...

5   Summary and outlook

What's the outcome? What else is possible? How can this work be continued? In here!

List of figures

List of literature

Deng, Jia et al. "What does classifying more than 10,000 image categories tell us?" In: *European conference on computer vision*. Springer. 2010, pp. 71–84.

Géron, Aurélien. *Praxiseinstieg Machine Learning mit Scikit-Learn und TensorFlow: Konzepte, Tools und Techniken für intelligente Systeme*. O'Reilly Verlag, 2017, pp. 8–14. ISBN: 9783960090618.

Osinga, Douwe. *Deep Learning Kochbuch: Praxisrezepte für einen schnellen Einstieg*. O'Reilly Verlag, 2019, pp. 22–26. ISBN: 9783960090977.

Rosch, Eleanor et al. "Basic objects in natural categories". In: *Cognitive psychology: Key readings* 448 (2004).

List of links

- Deep learning unbalanced training data?

  - `https://towardsdatascience.com/deep-learning-unbalanced-training-data-solve-it-like-this-6c528e9efea6`

- Data Augmentation

  - `https://machinelearningmastery.com/how-to-configure-image-data-augmentation-when-training-deep-learning-neural-networks/`

- Stop Feeding Garbage To Your Model! — The 6 biggest mistakes with datasets and how to avoid them.

  - `https://hackernoon.com/stop-feeding-garbage-to-your-model-the-6-biggest-mistakes-with-datasets-and-how-to-avoid-them-3cb7532ad3b7`

Declaration

I hereby declare that the work presented in this thesis is solely my work and that to the best of my knowledge this work is original, except where indicated by references to other authors. No part of this work has been submitted for any other degree or diploma.

Signature :   *Place, Date* :