Preprocessing for Neural Network Image Classification

Martina Kuchtová

3Alb, 2022 - 2023

Abstract. Image classification represents one of the fundamental tasks in Computer Vision (CV), which is mainly solved by utilizing Neural Networks. The aim of this thesis is to investigate the effect of using image pre-processing on the results and the accuracy of the neural network. Based on the analysis of a particular dataset, we propose methods from the field of Computer Vision to be applied on the images and evaluate their impact on results.

Keywords: Neural Network, Deep Learning, Image Classification, Preprocessing

1 Introduction

With the arrival of new technologies and Neural Networks in particular, Image classification has reached a new dimension. Whereas in the past we had to use different bags of features to categorize an image, today a simple Deep Learning model is sufficient.

Image Classification is the task within the Computer Vision consisting of assigning labels and categorizing images into pre-defined classes. It can be also often referred to as an Image Recognition.

Images are complex objects, they are not just some random pixels a grid, they describe things. [1] Nevertheless, we are looking for ways to obtain the most information out of them. The more information we have, the more accurate our results will be.

The use of Neural Networks naturally arises the question of data pre-processing – in this case, image pre-processing which is aimed to improve the image data (enhance some relevant image features etc.). However, in this thesis we will not apply this pre-processing to the data on which the Neural Network learns, as is customary. Our focus is on the data that is fed into the Neural Network, which has already been trained on a dataset. We will investigate whether we can modify such images by applying different methods to improve the Image Classification results, even though the Neural Network was trained on raw data.

When classifying images, we distinguish between Single-label and Multi-label Classification. More common is the ***Single-label Classification***, which as its name suggests, each image is assigned just one label. Therefore, the model returns just one value, while the output layer contains a vector of length n (number of classes) and a value indicating the score that the image belongs to that class. It employs a *Softmax* activation function that ensures that the score is equal to one and selects the class with the highest score as output. ***Multi-label Classification***, on the other hand, is a task where each image is assigned multiple labels. Compared to Simple-label Classification, it is much more complex and has found its way into medical imaging. It can also be used, for example, to indicate the presence of objects in images.

Image Classification itself has a few uses and a great potential. And medical images, autonomous vehicles, security systems and Reverse Image Search are just some of the examples of real-life applications of image classification.

1.1 Deep Learning in Image Classification

We would venture to say that nowadays you will not come across a situation where the task of image classification is handled in any other way than by using Neural Networks. Applying them to the Image Classification tasks has greatly facilitated the work, allowing us to achieve much more accurate results, since thanks to the Deep Learning (DL) is Neural Network able to learn from the image data by analysing them, identifying, and extracting features from the images.

As human beings, we can distinguish individual objects, but we also had to learn this in our childhood - for example, to tell the difference between individual shapes such as a square, a circle, a triangle, and a star. As children we had sensory toys to help us. Based on the features and characteristics of the object, we could throw it through the correct hole in the box. And this is exactly what the Image Classification is all about.

Deep learning is a type of machine learning that is based on algorithms that are structured in a similar way to the human brain. The algorithms recognise patterns in the image dataset as well as features that may be unique to a particular label.

Image classification techniques can be divided into two main categories: Supervised and Unsupervised Image Classification approache.

Supervised Image Classification is a machine learning approach that is defined by the use of labelled (previously classified) datasets. So, a human analyst plays a crucial role here and the data are representative of a particular class. By means of labelled inputs and outputs, the model can measure its accuracy and can learn over time. Some common types of classification algorithms are linear classifiers, Support Vector Machines (SVM), decision trees and random forests. [2]

Unsupervised Image Classification uses machine learning algorithms to analyse and cluster data, looking for hidden patterns without the need for human intervention. The unlabelled data is grouped on the basis of similarities or differences. Among the popular algorithms used here are K-means [2] and Iterative Self-Organizing Data Analysis Technique (ISODATA). [3]

1.2 Models *(change this sub-heading)*

The most popular neural network model, not only in Image Classification, is the Convolutional Neural Network (CNN) that has proven to be highly effective because of its ability to automatically extract features from the input space. Although most image classification tasks involve Convolutional Neural Networks, it does not mean that this type of problem cannot be solved by other models.

In 2021 was introduced the new architecture of ResMLP that was built entirely upon Multi-layer Perceptrons and consists just of fully-connected layers without convolution. Despite this, the model achieves impressive results, especially on ImageNet classification benchmarks. [4]

1.2.1 Convolutional Neural Network architecture (CNN)

The main reason why is this type of Neural Network widely used for image classification and object detection tasks is its design and ability to process data with a spatial grid-like topology – such as images.

Let us outline the architecture of this type of Neural Network, as our work will revolve around it. Composition consists of multiple layers including convolutional, pooling and fully-connected layers. Each layer is a 3-dimensional grid structure, which has its height, width, and depth.

The convolutional layers are strong feature extractors composed of multiple filters, while each filter is responsible for detecting a specific feature such as texture or edges. These filters are learnable and used to scan the input image. [5] [6]

To reduce the spatial dimensions of the output from the convolutional layers, the pooling layers are used. [5] This involves applying a pooling operation, e.g., max-pooling or average-pooling, that works on small grid regions in each layer and creates another layer of the same depth (unlike filters). It results in reduced number of parameters and computation in the Neural Network. This makes it more robust to small translations in the input of the features detected by the convolutional layers. [7]

The fully-connected layers are used for the classification of the features extracted by the convolutional layers. Several convolutional and pooling layers are often stacked on top of each other in order to extract more abstract feature representations as the network traverses them. These feature representations are interpreted by the fully-connected layers and the function of high-level reasoning is performed. [5]

1.2.2 Convolutional Neural Network models (CNN)

The most popular Convolutional Neural Networks used for Image Classification are AlexNet, GoogLeNet, and ResNet50. When comparing these networks using the CIFAR-10 dataset (see 1.3 Datasets), RestNet performed the best (Table 1). [8] The VGG-16, which is cutting-edge and has already beaten the AlexNet, could also be included. All these models are already pre-trained.

Table 1. Performance of CNNs on the CIFAR10 test dataset [8]

CIFAR-10		AlexNet	GoogLeNet	ResNet50
Image Category	Airplane	41.80%	51.10%	90.80%
	Automobile	21.80%	62.10%	69.10%
	Bird	00.02%	56.70%	72.60%
	Cat	00.03%	78.80%	61.90%
	Deer	87.60%	49.50%	75.40%
	Dog	23.00%	57.50%	82.10%
	Frog	24.20%	90.20%	76.60%
	Horse	34.70%	78.20%	84.70%
	Ship	31.70%	95.50%	83.20%
	Truck	95.90%	97.10%	84.60%

1.3 Datasets

Although almost any dataset can be used for image classification, and we can even build our own, there are some that stand out due to their ubiquity and are very popular when it comes to solving Image Classification tasks:

- **MNIST** (Modified Institute for Advanced Research) – a dataset consisting of 60,000 training and 10,000 testing small square grayscale images representing hand-written single-digit numbers from 0 to 9. The images originated as scans and differ in the handwriting of different individuals.
- **CIFAR-10** (The Canadian Institute for Advanced Research) – a collection of training images used in Computer Vision projects. Most often used to train models to classify objects into one of the 10 classes. 60,000 images are evenly distributed across the classes, they are useful for quickly testing algorithms due to their low resolution. This dataset is often used to test on a small scale before more extensive training is performed using ImageNet. [7]
- **ImageNet** – a huge database is used in ImageNet Large Scale Visual Recognition Challenge (ILSVRC) and contains over 14 million images drawn from 1,000 different categories. Many state-of-the-art image recognition architectures have emerged from this competition (see 1.2.2). [7] There are provided only URLs and thumbnails of the images since the ImageNet does not own the image copyright.
- **ObjectNet** – a unique image dataset containing object captured in an unusual position within the complex scenes. This dataset serves only for testing the computer vision systems.

- **Intel Image Classification** – is focusing on different natural scenes such as forests, mountains, streets, building. It was originally created for the image classification competition purposes.

Some Deep Learning models may have a large data requirement - too little training data results in a poor accuracy. Dataset augmentation can help to increase the size of our dataset. By applying various transformations to existing image data, we are able to multiply the size of our dataset several times.

As a rule, each dataset should be divided into Test data and Training data. As the name suggests, training data is used to train the neural network - we feed it labelled examples and the network learns features that can distinguish one category from another. The test data is separate and not used for training – it allows us to assess the performance and the accuracy of the network on new unseen images.

2 Art Image Classification

Art Image Classification is the first task we have looked at. Given a Convolutional Neural Network model that classifies an image into five classes based on Art type: drawing, engraving, painting, iconography, sculpture.

The input dataset consisted of labelled data obtained from The Virtual Russian Museum. We did an analysis of the image data for each of the categories and a comparison between them. This enabled us to suggest some of the pre-processing methods to apply to the images.

There are our observations we made as humans, and based on what we ourselves would classify the image into one of the classes:

a) The drawings and the engravings are very similar in terms of the use of the colour palette and style – it is also difficult to distinguish them with a naked eye. Moreover, even pencil strokes resemble grooves.
b) The icons are characterised by high colouration and a large representation of yellow or gold, while the earthy colours are predominant in the paintings.
c) There must be a face on the iconography, but these are also dominant in other types of art, mostly painting.
d) Not only paintings are framed - but we also find them in drawings, engravings, and iconography. The icons contain distinctive circles that represent the halo.
e) Sculptures have the greatest depth of any of the above and the background is often monolithic or blurred, making the sculpture stand out in the foreground.

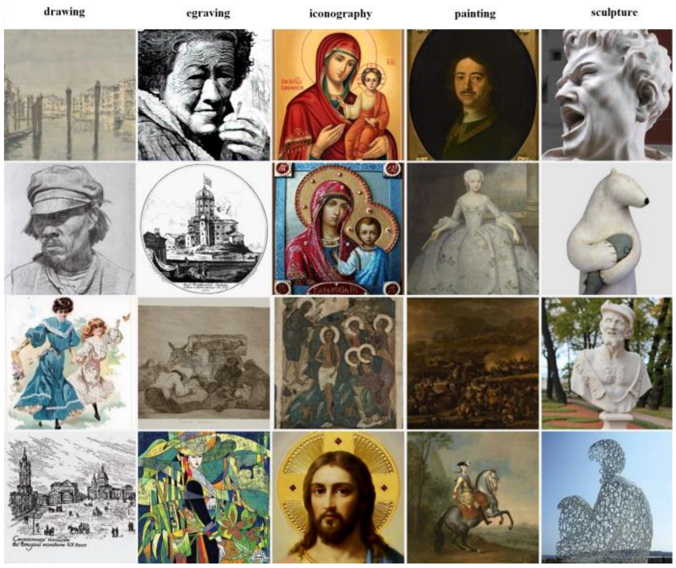

Figure 1. Sample representatives for each class from the dataset used to train the Art Image Classification Neural Network.

Given the diversity of the dataset, we assume that the Neural Network also takes colour into account when classifying and we were assured of this when we fed it greyscale images as input. The accuracy of the classification has dropped significantly by almost 40% (Table 2.).

Despite this finding, we also performed histogram equalization and adaptive histogram equalization on the grayscale images as well to see what effect it has on the results. We also used this procedure for colour images, where the results were compared to greyscale better but still negative – accuracy decreased.

The images in the dataset are of very good quality, we hypothesized that the increasing sharpness might also be one of the factors that would help the network to classify the drawings and engravings more accurately, and thus sharpening might help to highlight thin lines (pencil strokes). However, the accuracy of image classification on such pre-processed inputs was reduced by almost half.

In view of the recurring faces and point c), we have applied face recognition to each of the images using the Haar Cascade. For greater efficiency, we looked not only for faces that were perpendicular to the camera, but also those that were from profile. Despite the use of this feature extension, not all the faces that were present in the images were detected correctly. To draw attention to the face in the images, in case it was present, we tried two approaches. In the first we framed the face with a yellow square, in the second we decided to drop this drawing into the image and made an adjustment where we converted the image to greyscale, leaving the face area untouched and in its original colours. As the Neural Network had not been trained on this type of input images, there was no improvement in the classifying accuracy.

Based on the specific colour representation for some categories, we also tried to apply pixel clustering and therefor the K-Means algorithm - differentiating 3-colour, 9, 12 and 15-colour images, where the number of colours is represented by $K - number\ of\ clusters$. We can say that the use of this method is the closest to the original classification accuracy, i.e., it has the least degradation. The best results were obtained when $K = 12$, while the accuracy dropped again when $K = 15$. This means that when using 13 or 14 clusters, there is a possibility to increase the accuracy even more and this is something that we need to test further.

Table 2. Image Classification Accuracy on pre-processed Testing data

Method	Test Accuracy
None (original images)	77.45%
Adaptive Histogram Equalization (grey)	31.77%
Face recognition (framing)	73.83%
Face recognition (grayscale image, face in colour)	51.29%
Grayscale	41.47%
Histogram Equalization (colour)	52.80%
Histogram Equalization (grey)	42,40%
K-Means (K=3)	69.16%
K-Means (K=9)	75.93%
K-Means (K=12)	76.51%
K-Means (K=15)	75.23%
Sharpening	40.77%

3 Conclusions

We got an overview of how image classification can be approached, what the tasks are and what the trends are. We focused on working with Neural Network for Art Image Classification, where we applied some methods from the Computer Vision. In the near future, we want to take a closer look at the individual methods and their benefits. So far, our focus has only been on accuracy, but we will also look at precision, recall and F1 score.

To proceed further in this thesis, we will choose another Neural Network with a more complex model (perhaps one of those mentioned above) and possibly with a higher number of classes. Also, we would like to focus more on working with the already pre-trained Neural Networks and the cloud.

Literature

[1] Sharma N., Jain V., Mishra A.: Procedia Computer Science, (2018), 377-384
[2] Delua, J. (2021, March 12). *Supervised vs. Unsupervised Learning: What's the Difference?* IBM Blog. https://www.ibm.com/cloud/blog/supervised-vs-unsupervised-learning
[3] Boesch , G. (n.d.). *A Complete Guide to Image Classification in 2023*. Viso.ai. https://viso.ai/computer-vision/image-classification/
[4] Touvron H, Bojanowski P, Caron M et al. (2021). ResMLP: Feedforward networks for image classification with data-efficient training.
[5] RAWAT, W. AND WANG, Z.: Neural Computation, (2017), 2352-2449, 29(9)
[6] Aggarwal C.: Neural Networks and Deep Learning, (2018), 315-3687
[7] Xiang Z., Zhang R., Seeling P.: Computing in Communication Networks: From Theory to Practice, (2020), 325-338

[*] S. S. Haykin, S. S. Haykin, S. S. Haykin, and S. S. Haykin, Neural networks and learning machines, vol. 3. Pearson Upper Saddle River, NJ, USA:, 2009.