DRIVES

Development and Research on Innovative Vocational Education Skills

COMPUTER VISION FUNDAMENTALS

U2.E7. IMAGE CLASSIFICATION

Computer Vision Expert

May 2021, Version 1

Co-funded by the Erasmus+ Programme of the European Union

The Development and Research on Innovative Vocational Educational Skills project (DRIVES) is co-funded by the Erasmus+ Programme of the European Union under the agreement 591988-EPP-1-2017-1-CZ-EPPKA2-SSA-B. The European Commission support for the production of this publication does not constitute endorsement of the contents which reflects the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

LEARNING OBJECTIVES

The student is able to …

CVE.U2.E7.PC1	The student can define and understand image classification.
CVE.U2.E7.PC2	The student is able to analyse and explore practical applications of image classification.
CVE.U2.E7.PC3	The student knows different image classification models.
CVE.U2.E7.PC4	The student is able to select the image classification model that best fits a specific situation or problem.
CVE.U2.E7.PC5	The student is able to assess the performance of the classification model and proceed with the necessary measures.

IMAGE CLASSIFICATION

Image classification is the prediction of a specific class or label for something that is defined by a set of data points. Image classification is a subset of the classification problem, where an entire image is assigned a label. Perhaps a picture will be classified as a daytime or nighttime shot. Or, similarly, images of cars and motorcycles will be automatically placed into their own groups. There are numerous categories/classes, that a specific image can be classified.

IMAGE CLASSIFICATION

Image classification is a fundamental problem in computer vision and serves as the foundation of multiple tasks like object detection, image segmentation, object tracking, action recognition, and autonomous driving. There are some very popular datasets in Image Classification that are used across research, industry, and hackathons.

The most popular ones are:

- ImageNet
- CIFAR
- MNIST

APPLICATIONS OF IMAGE CLASSIFICATION

Some examples of image classification include:

- Labeling an x-ray as cancer or not (binary classification).
- Classifying a handwritten digit (multiclass classification).
- Assigning a name to a photograph of a face (multiclass classification).

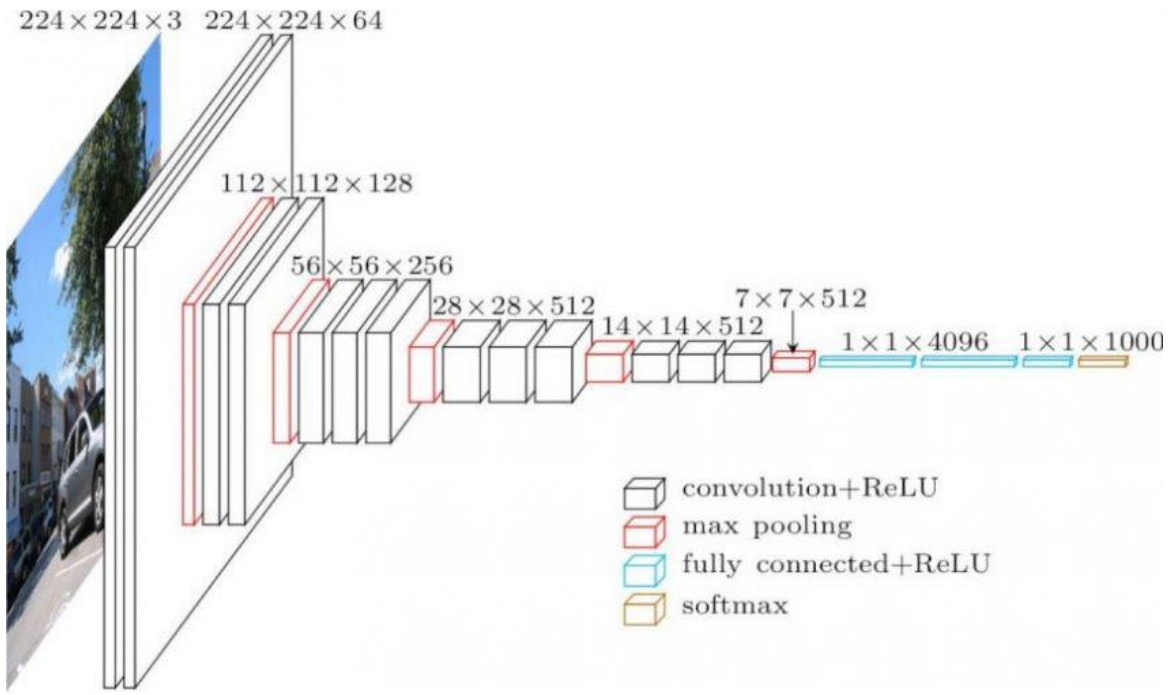
A popular example of image classification used as a benchmark problem is the MNIST dataset.

Deep neural networks

Generally, the deep learning architecture for image classification includes convolutional layers, making it a convolutional neural network (CNN). A typical use case for CNNs is when the network images are fed and it is the network that classifies the data. CNNs tend to start with an input “scanner” that is not intended to parse all the training data at once. For example, to input an image of 100 x 100 pixels, you wouldn’t want a layer with 10,000 nodes.

Very Deep Convolutional Networks for Large-Scale Image Recognition(VGG-16)

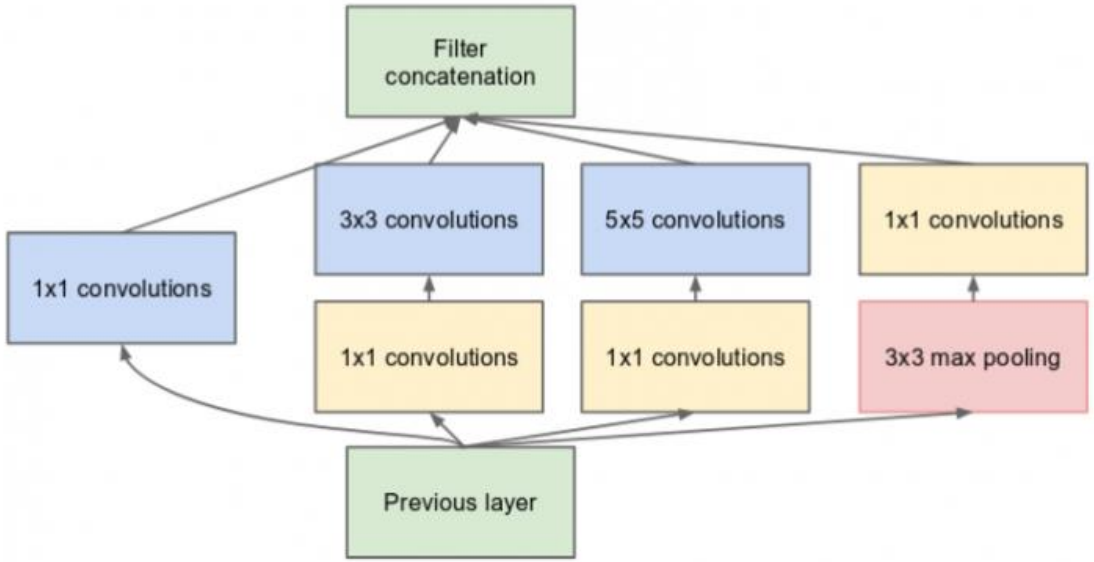
One of the most popular pre-trained models for image classification is called VGG-16. In the figure on the right, we can see its architecture.



The model is **sequential** in nature and uses lots of **filters**. At each stage, small 3 * 3 filters are used to **reduce** the number of **parameters**. All the hidden layers use the **ReLU** activation function. Nevertheless, the number of parameters is **138 Billion** – which makes it a slower and much larger model to train than others.

Inception

In simple terms, the Inception Module performs convolutions with different filter sizes on the input, performs Max Pooling, and concatenates the result for the next Inception module. The introduction of the 1 * 1 convolution operation drastically reduces the parameters.



Inception module with dimension reductions.

DRIVES
Development and Research on Innovative Vocational Education Skills

ResNet50

The principal motivation behind this model was to avoid poor accuracy as the model went on to become deeper.

When you work with **Gradient Descent**, you would have come across the *Vanishing Gradient* issue. The *ResNet* model aimed to tackle this issue.

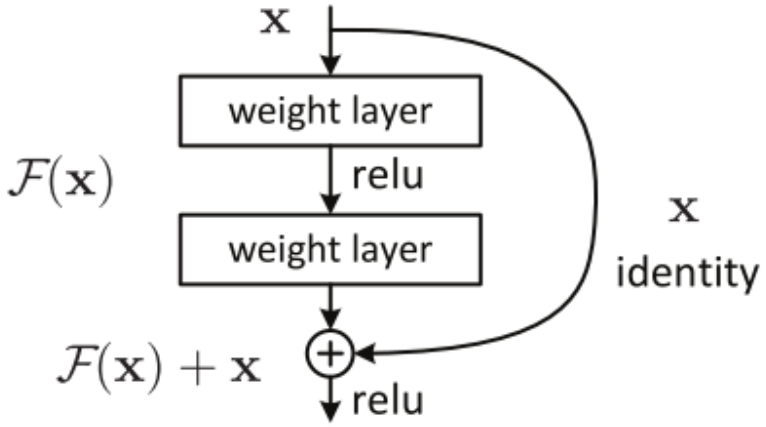
After started with a single Convolutional layer and Max Pooling, there are 4 similar layers with varying filter sizes (all using 3 * 3 convolution operation). After every 2 convolutions, it is bypassing/skipping the layer in-between. This is the main concept behind ResNet models. These skipped connections are called ‘identity shortcut connections” and use what is called **residual blocks**.

ResNet50

ResNet authors propose that fitting a residual mapping is much easier than fitting the actual mapping and thus apply it in all the layers.

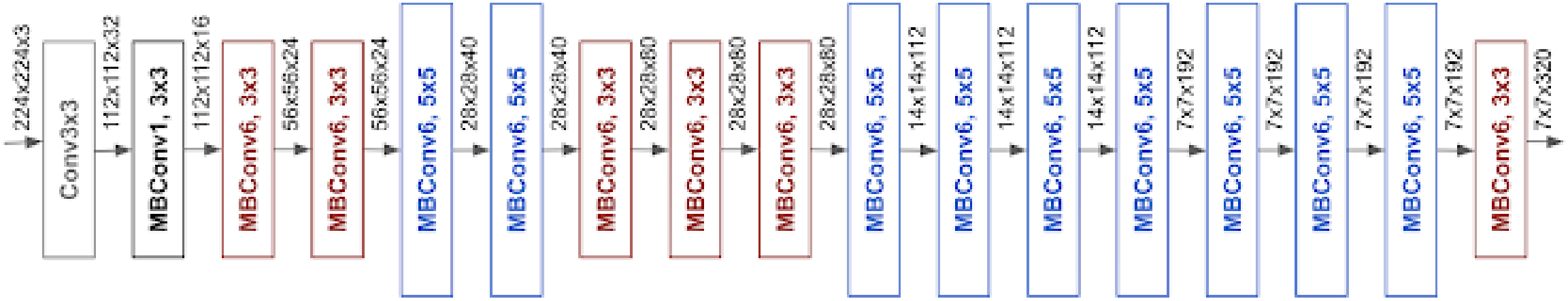
Also, they are of the opinion that the more layers it stacks, the model should not perform worse.

This is opposite to Inception and is almost similar to VGG16 in the sense that it is just stacking layers on top of the other. ResNet just changes the underlying mapping.



EfficientNet

It was developed from Google. *EfficientNet* authors propose a new Scaling method called **Compound Scaling.** The earlier models like *ResNet* follow the conventional approach of scaling the dimensions arbitrarily and by adding up more and more layers.



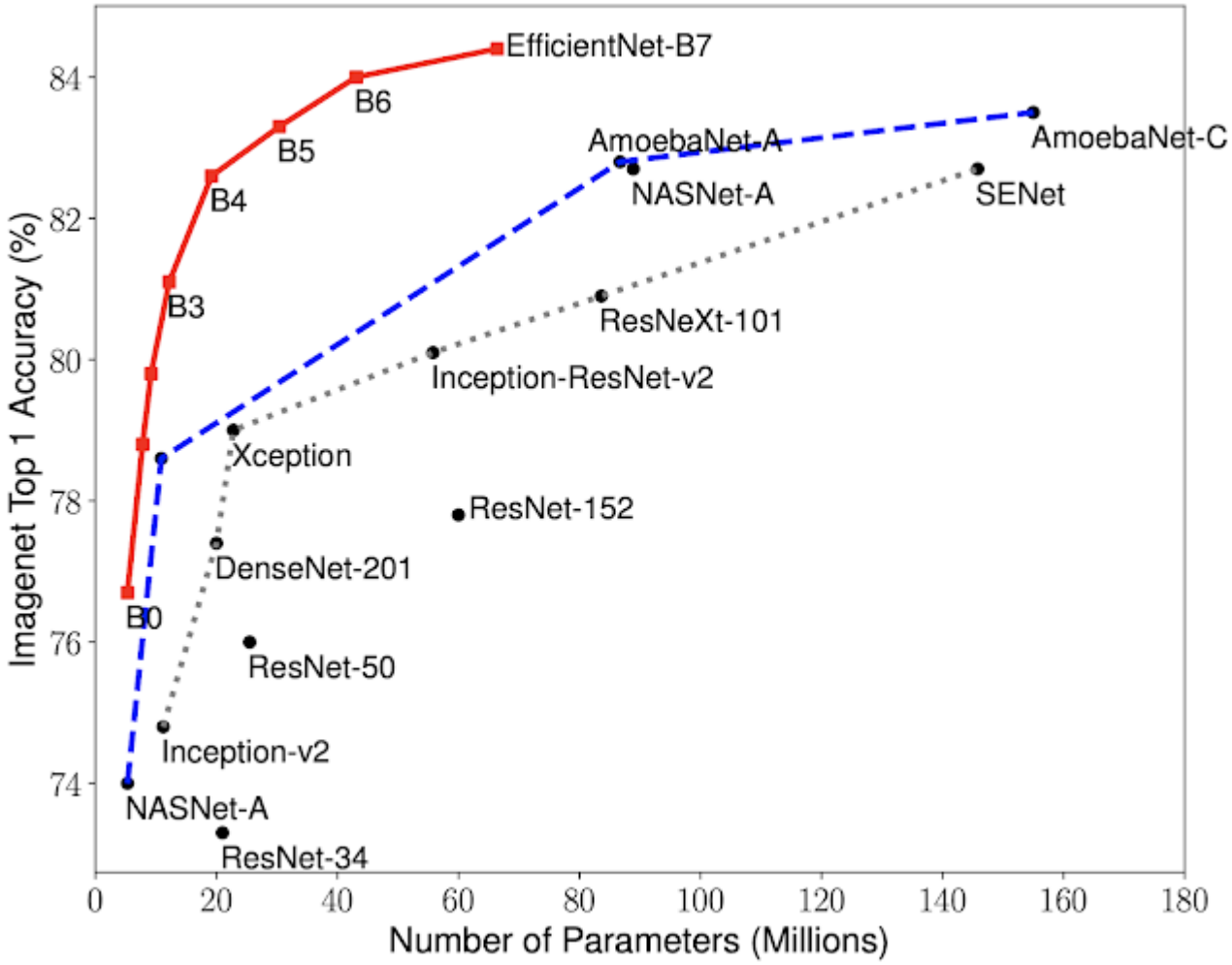
THE PERFORMANCE OF THE CLASSIFICATION MODEL

The robustness of the Deep Neural Network (DNN) models is evaluated using the four metrics:

Attack success rate, Distortion, CLEVER score, and **Transferability.**

Model	Year	Number of Parameters	Top 1 Accuracy
VGG-16	2014	138Million	74.5%
Res-Net-50	2015	25Million	77.15%
Inception V3	2015	24Million	78.8%
EfficientNetB0	2019	5.3Million	76.3%
EfficientNetB7	2019	66Million	84.4%

THE PERFORMANCE OF THE CLASSIFICATION MODEL



Graph showing the comparative performance of popular models

SUMMARY

- **Image classification** is the prediction of a specific class or label, for something that is defined by a set of data points. Image classification is a subset of the classification problem, where an entire image is assigneda label.
- Labeling an x-ray, classifying a handwritten digit and assigning a name to a photograph of a face are some examples that include **image classification**.
- Image Classification **Models**: VGG-16, Res-Net-50, Inception and EfficientNet.
- **EfficientNet** is the model with the highest accuracy.

REFERENCES

Computer Vision and Image Classification -A study. (n.d.). Retrieved from https://www.dexlabanalytics.com/blog/computer-vision-and-image-classification-a-study

Su, D., Zhang, H., Chen, H., Yi, J., Chen, P. Y., & Gao, Y. (2018). Is robustness the cost of accuracy? – A comprehensive study on the robustness of 18 deep image classification models. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *11216 LNCS*, 644–661. https://doi.org/10.1007/978-3-030-01258-8_39

Top 4 Pre-Trained Models for Image Classification | With Python Code. (n.d.). Retrieved from https://www.analyticsvidhya.com/blog/2020/08/top-4-pre-trained-models-for-image-classification-with-python-code/

REFERENCE TO AUTHORS

Ana Luísa Sousa

- PhD student in Information System and Tecnologies
- Research Collaborator of the Algoritmi Research Center

0000-0001-5731-3583

Regina Sousa

- PhD student in Biomedical Engineering
- Research Collaborator of the Algoritmi Research Center

0000-0002-2988-196X

Diana Ferreira

- PhD student in Biomedical Engineering
- Research Collaborator of the Algoritmi Research Center

0000-0003-2326-2153

REFERENCE TO AUTHORS

António Abelha

- Assistant Professor at the University of Minho
- Integrated Researcher of the Algoritmi Research Center

0000-0001-6457-0756

José Machado

- Associate Professor with Habilitation at the University of Minho
- Integrated Researcher of the Algoritmi Research Center

0000-0003-4121-6169

Victor Alves

- Assistant Professor at the University of Minho
- Integrated Researcher of the Algoritmi Research Center

0000-0003-1819-7051

REFERENCE TO AUTHORS

This Training Material has been certified according to the rules of **ECQA – European Certification and Qualification Association.**

The Training Material was developed within the international job role committee **“Computer Vision Expert”:**

UMINHO – University of Minho (https://www.uminho.pt/PT)

The development of the training material was partly funded by the EU under Blueprint Project DRIVES.

DRIVES

Development and Research on Innovative Vocational Education Skills

Thank you for your attention

DRIVES project is project under **The Blueprint for Sectoral Cooperation on Skills in Automotive Sector**, as part of New Skills Agenda.

The aim of the Blueprint is **to support an overall sectoral strategy and to develop concrete actions to address short and medium term skills needs.**

Follow DRIVES project at:

More information at:

www.project-drives.eu

Co-funded by the Erasmus+ Programme of the European Union

The Development and Research on Innovative Vocational Educational Skills project (DRIVES) is co-funded by the Erasmus+ Programme of the European Union under the agreement 591988-EPP-1-2017-1-CZ-EPPKA2-SSA-B. The European Commission support for the production of this publication does not constitute endorsement of the contents which reflects the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.