Computer Vision Project:
Prediction of Covid-19 infections with X-Ray images

David Budgenhagen, Leonidas Devetzidis, Alexandru Bârsan

March 2021

Motivation

The idea is to analyse medical images with deep learning. Deep Learning is used to aid doctors in the analysis of complex medical data and it aims to assign the patient to one of a small set of classes **[Amato et al., 2013] [3]**. It is not a new approach, medical image analysis exists since the 60s with the help of edge and line detector filters.
Some typical applications of Deep Learning in the medical image analysis are:

- the detection of body structures such as organs **[Lo et al., 1995] [8]**
- Segmentation for cardiac and brains analyses **[Ronneberger et al., 2015][9]**
- image classification, to determine if a disease is present or not by classifying X-Ray images **[Antony et al., 2016] [2]**

Why use Deep Learning for Medical Image Analysis?

Papers and statistics show that Deep Learning approaches exceed the accuracy of trained doctors in general, for example in skin cancer detection a CNN exceeds the accuracy of the one from dermatologists by 6% (66% - 72%)**[Esteva, 2017] [4]**. A benefit of Neural Networks is that once these are trained,debugged a fine-tuned, the classification of images is fast and an easy to implement process. During the COVID-19 outbreak a real-time polymerase chain reaction was the standard method to detect the disease, the problem is that such a large analysis for thousands of patients can't be carried out in a short time, especially in countries which don't have a good laboratory infrastructure.

The Project

The project is about the detection of COVID-19 in patients by classifying radiographic lung images. This is sometimes very challenging for doctors to interpret correctly because they have to also distinguish between different types of pneumonia **[Joarder, Crundwell, 2009] [6]**. There are approaches that have shown that the recognition of COVID-19 by X-ray images is possible and useful for the medical environment. **[Minaee, Zhgang et al., 2020] [10]**

Dataset

a **The initial dataset** was the 1st version of the COVID-19 radiography database with $\sim$ 200 COVID positive images and $\sim$ 1300 for normal and viral pneumonia images. The downside of this version of the dataset is that there is a low amount of COVID images which makes the dataset pretty unbalanced.

b **The current dataset** is the 3rd version of the COVID-19 Radiography Database in which 1000 more COVID images were added. To clean the dataset we removed the duplicates by an image-hash comparison and also removed side photos. Furthermore it got balanced, with 1028 images for each category. This was done due to the accuracy paradox, so that the model doesn't get biased by a class which comes up more frequently.

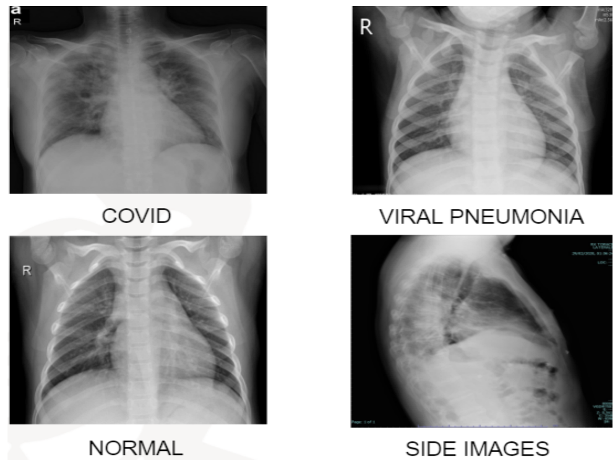


Figure 1: Image categories from dataset
[1]

Model Architecture

In this section two models are presented that were used for the experiments:

1. A naive CNN with 5 convolutional layers, a global-average-pooling layer, a dense layer and a softmax layer for classification. Figure 2 shows the architecture in more detail.

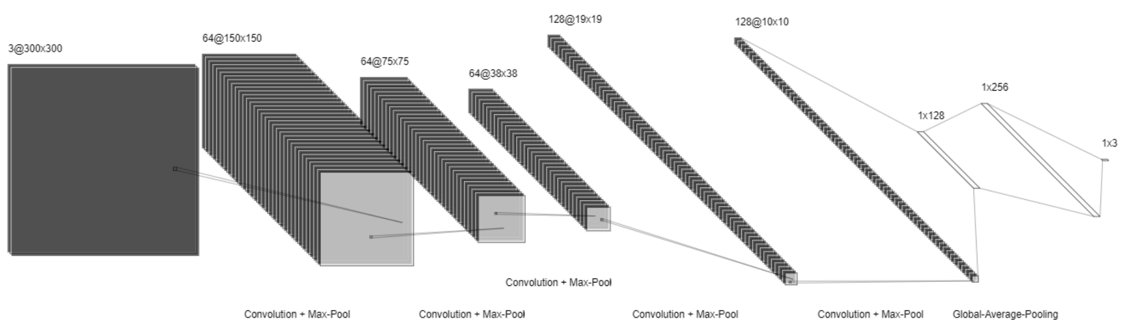


Figure 2: Naive CNN

2. Efficient Networks as presented in 2019 by Tan, Mingxing and Le, Quoc [11]

Efficient Networks

Tan, Mingxing and Le, Quoc presented a new *compound scaling method* that efficiently balances and scales network depth, -width and -resolution in order to achieve better performance. The scaling method was tested on their presented Efficient Network baseline model and resulted in state-of-the-art 84.4% top-1-accuracy ImageNet while being 8.4x smaller and 6x times faster on inference than the best existing convolutional model. Figure 3 shows the performance of EfficientNets compared to other model architectures.

The authors observed that in order to pursue better accuracy and efficiency, it is critical to balance all dimensions of network width, -depth and -resolution during model scaling. Their observation was based on the fact that scaling up any dimension of the network separately improved accuracy, but the accuracy-gain diminished as the model increased. Figure 4 illustrates that observation.

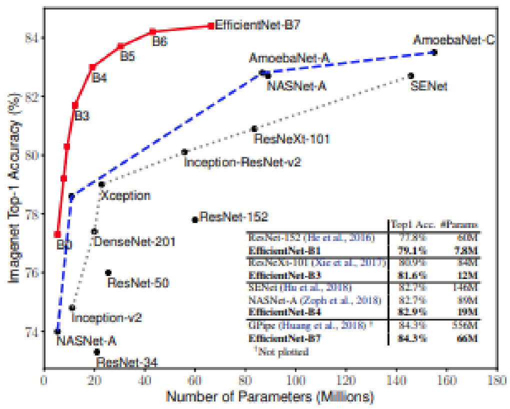


Figure 3: EfficientNet significantly outperforming other CNNs.

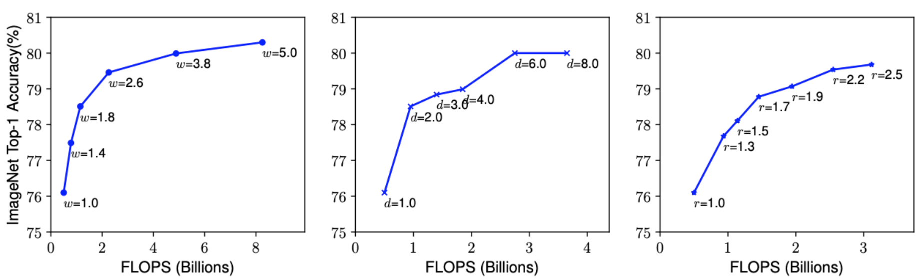


Figure 4: Single dimension scaling increases the accuracy, but with diminishing returns. In order to get the best performance, it is crucial to balance all dimensions of the network uniformly.

The observation of scaling all these dimensions uniformly lead them to present the **compound scaling method**, which uses a compound scaling coefficient ϕ to uniformly scale network width w, depth d and resolution r in a principled way:

$$
\begin{aligned}
& d=\alpha^{\phi}, w=\beta^{\phi}, r=\gamma^{\phi} \\
& \text{s.t. } \alpha \cdot \beta^{2} \cdot \gamma^{2} \approx 2 \\
& \alpha \geq 1, \beta \geq 1, \gamma \geq 1
\end{aligned} \tag{1}
$$

Through grid-search applied on the baseline model, the best values were found to be $\alpha = 1.2, \beta = 1.1$ and $\gamma = 1.15$. Afterwards, the authors fixed α, β, γ as constants and scaled up the baseline with different values of ϕ to achieve various Efficient Networks.
Because the compound scaling method does not change the layer operations, it it essential to find a good baseline model. For the experiments the baseline model from the paper was adopted. The baseline model consists of mobile inverted bottlenecks with additional squeeze-and-excitation block optimization. The building blocks and the baseline model are showing in Figure 5 and Table 1, respectively.

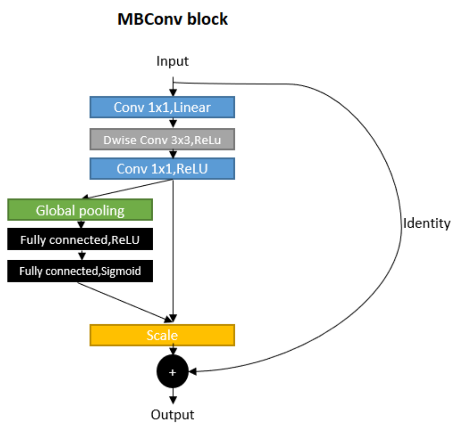


Figure 5: A mobile inverted bottleneck with squeeze-and-excitation optimization

Table 1: EfficientNet baseline model

Stage	**Operator**	**Exp. factor**	**Resolution**	**# Channels**	**# Layers**
1	Conv3x3	-	224x224	32	1
2	MBConv3x3	1	112x112	16	1
3	MBConv3x3	6	112x112	24	2
4	MBConv5x5	6	56x56	40	2
5	MBConv3x3	6	28x28	80	3
6	MBConv5x5	6	14x14	112	3
7	MBConv5x5	6	14x14	192	4
8	MBConv3x3	6	7x7	320	1
9	Conv1x1 &Pooling & FC	-	7x7	1280	1

Evaluation

This section evaluates the performance of the models for our experiments. A summary for the models are shown in Table 2. All models were trained for 100 epochs with the Adam [7] optimizer and an inital learning rate of 0.001. Because of memory limitations, a batch-size of 4 was the maximum that could fit on the provided GPU with an input resolution of 600x600. Because of this limitation, we trained several models with varying batch-sizes and input resolution. For the future, it would be interesting to train the 600x600 model, with a bigger batch-size, i.e 32 or 64.

Table 2: Model summary

Model	Input resolution	Batch size	Regularization
Naive CNN with 5 layers	300x300	64	$\times$
EfficientNet($\phi = 7$)	600x600	4	✓
EfficientNet($\phi = 7$)	360x360	16	✓
EfficientNet($\phi = 7$)	475x475	8	✓

Regularization

The following regularization methods were implemented:

- Dropout with p = 0.5
- Batch Normalization
- Early Stopping
- Data Augmentation with the following operations: Random Brightness, Flip Vertical, Random Rotation, Random Shear and Random Erasing. Each with p = 0.25.

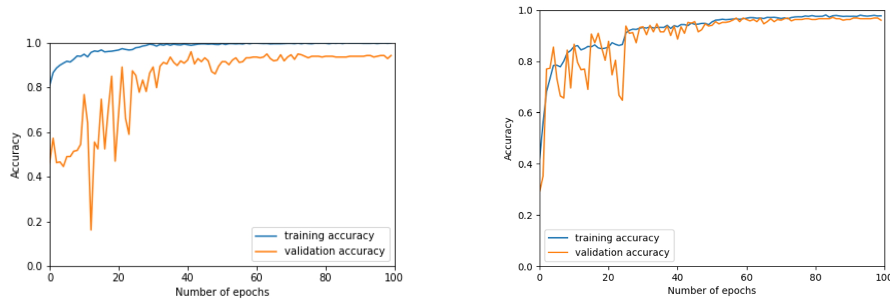
Validation accuracy

Table 3 shows the validation accuracy of the trained models. From the table it is apparent that the EfficientNet with an input resolution of 475x475 and a batch-size of 8 performs the best. It reaches a validation accuracy of 97.16% which is a substantial increase from the naive CNN.
Figure 6(a) also illustrates that using no regularization can lead to severe overfitting, especially between epochs 15-35. Using regularization leads to a much smoother curve, even when the batch-size is relatively low.

Because the accuracy itself is a somewhat misleading metric and can lead to problems such as the accuracy paradox [5], it is important to also measure different metrics such as recall and precision. Table 4 shows the different metrics measured for the different models. We also compared our model to the state-of-the-art COVIDNet [12]. As shown in the table, our model reaches comparable results to COVIDNet. What's important to note is that for the evaluation of the COVIDNet only 100 images were used, but for our experiments we used 359 images. We also plotted the confusion matrix to better interpret our results. Figure 7 shows the confusion matrix for our best performing EfficientNet model with an input resolution of 475x475. The confusion matrix shows that our model has some difficulties distinguishing patients infected with covid and patients infected with pneumonia. The model is more inclined to predict covid instead of pneuomonia when it is uncertain. That also explains the recall of 100% for covid classification, but only a precision of 92%.

Table 3: Model summary

Model	Input resolution	Batch size	Validation Accuracy
Naive CNN with 5 layers	300x300	64	91.43%
EfficientNet($\phi = 7$)	600x600	4	95.89%
EfficientNet($\phi = 7$)	360x360	16	96.88%
EfficientNet($\phi = 7$)	475x475	8	**97.16%**



(a) Naive CNN.

(b) EfficientNet($\phi = 7$) with 475x475 resolution

Figure 6: Validation accuracy of the naive CNN without any regularization (left) and the sophisticated EfficientNet($\phi = 7$) trained with regularization (right).

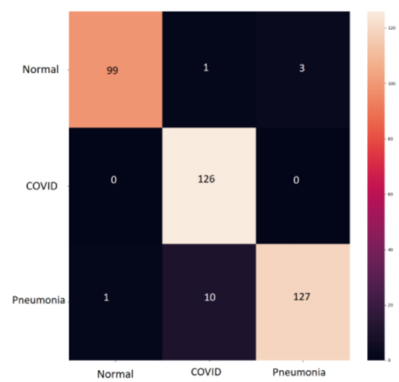


Figure 7: Confusion matrix of our best performing EfficientNet with input resolution 475x475 and batch-size of 8.

Table 4: Evaluation of different models

EfficientNet($\phi = 7$) 600x600	**Precision**	**Recall**	**F1-Score**	**Support**
Normal	0.99	0.98	0.99	103
Covid	0.86	1.00	0.92	126
Pneumonia	0.98	0.83	0.90	130
EfficientNet($\phi = 7$) 360x360	**Precision**	**Recall**	**F1-Score**	**Support**
Normal	0.99	0.99	0.99	103
Covid	0.91	1.00	0.95	126
Pneumonia	0.99	0.90	0.94	130
EfficientNet($\phi = 7$) 475x475	**Precision**	**Recall**	**F1-Score**	**Support**
Normal	0.99	0.96	0.98	103
Covid	0.92	**1.00**	0.96	126
Pneumonia	0.98	0.92	0.94	130
COVIDNet	**Precision**	**Recall**	**F1-Score**	**Support**
Normal	0.91	0.94	0.92	33
Covid	**0.99**	0.95	0.97	34
Pneumonia	0.93	0.94	0.93	33

Workload:

- **David Budgenhagen:** Removing duplicate images, implementing-, fine-tuning- and training state-of-the-art models. Implementing regularization methods. Implementing evaluation methods/confusion matrix
- **Alexandru Bârsan:** Searching for datasets, loading data, data preprocessing, merging data
- **Leonidas Devetzidis:** Data preprocessing, balancing datasets, implementing regularization methods

References

[1] Covid-19 radiography database; https://www.kaggle.com/tawsifurrahman/covid19-radiography-database.

[2] McGuinness K. O'Connor N. E. Moran K. , Antony J. *Quantifying radiographic knee osteoarthritis severity using deep convolutional neural networks. In 2016 23rd International Conference on Pattern Recognition (ICPR) (pp. 1195-1200).* 2016.

[3] Peña-Méndez E. M. Vaňhara P. Hampl A. Havel J. Amato F., López A. *Artificial neural networks in medical diagnosis.* 2013.

[4] Novoa R. A. Ko J. Swetter S. M. Blau H. M. Thrun S. Esteva A., Kuprel B. *Quantifying radiographic knee osteoarthritis severity using deep convolutional neural networks. In 2016 23rd International Conference on Pattern Recognition (ICPR) (pp. 1195-1200).* 2017.

[5] Francisco J. Valverde-Albacete and Carmen Peláez-Moreno. 100information transfer factor explains the accuracy paradox, Jan 2016.

[6] Crundwell N. Joarder R. *Chest X-ray in clinical practice.* 2009.

[7] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization, 2014. cite arxiv:1412.6980Comment: Published as a conference paper at the 3rd International Conference for Learning Representations, San Diego, 2015.

[8] Chan H. P. Lin J. S. Li H. Freedman M. T. Mun S. K. Lo, S. C. B. *Artificial convolution neural network for medical image pattern recognition. Neural networks, 8(7-8), 1201-1214.* 1995.

[9] Brox T. Ronneberger O., Fischer P. *U-net: convolutional networks for biomedical image segmentation. In: Proceedings of the Medical Image Comput- ing and Computer-Assisted Intervention. In: Lecture Notes in Computer Science, 9351, pp. 234–241.* 2015.

[10] Milan Sonka Shakib Yazdani Ghazaleh Jamalipour Soufi Shervin Minaee, Rahele Kafieh. *Chest X-ray in clinical practice.* 2020.

[11] Mingxing Tan and Quoc Le. EfficientNet: Rethinking model scaling for convolutional neural networks. In Kamalika Chaudhuri and Ruslan Salakhutdinov, editors, *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 6105–6114. PMLR, 09–15 Jun 2019.

[12] Linda Wang and Alexander Wong. Covid-net: A tailored deep convolutional neural network design for detection of covid-19 cases from chest x-ray images, 2020.