Custom Pretrainings and Adapted 3D-ConvNeXt Architecture for COVID Detection and Severity Prediction

Daniel Kienzle*, Julian Lorenz*, Robin Schön*, Katja Ludwig, Rainer Lienhart

Augsburg University, Augsburg 86159, Germany
{firstname.lastname}@uni-a.de

Abstract. Since COVID strongly affects the respiratory system, lung CT scans can be used for the analysis of a patient’s health. We introduce an neural network for the prediction of the severity of lung damage and the detection of infection using three-dimensional CT-scans. Therefore, we adapt the recent ConvNeXt model to process three-dimensional data. Furthermore, we introduce different pretraining methods specifically adjusted to improve the models ability to handle three-dimensional CT-data. In order to test the performance of our model, we participate in the 2nd COV19D Competition for severity prediction and infection detection.

Keywords: Machine Learning, COVID Detection, Severity Prediction, Medical Image Analysis, CT scans, 3D data

1 Introduction

The last few years have been strongly shaped by the COVID-19 pandemic, with a considerable amount of cases ending deadly. For the treatment of patients it is very important to diagnose the severity of lung damage caused by a SARS-CoV-2 infection accurately. The lung damage is visually detectable by visible ground-glass opacities and mucoid impactions on the slices of a three dimensional CT image ([17]). Thus, it might be beneficial to use CT-Scans for the diagnosis of the patients.

Goal of this paper is to develop an neural network that is able automatically predict four degrees of severity of lung damage from a lung CT-scan of the patient. Furthermore, we also adapt our architecture in order to predict infections with the SARS-CoV-2 virus using CT-scans. In order to improve these two tasks, our main contributions are:

1. We adapt the recent ConvNeXt architecture ([15]) to process three-dimensional data
2. We introduce different techniques for the pretraining of our architecture in order to increase the ability of our network to handle 3D CT-data
3. We introduce a robust estimation method of the performance of our model through the usage of 5-fold cross-validation.

* Authors contributed equally
2 Related Work

The idea of using neural networks for the prediction of certain properties visible in medical data has developed to increasing levels of importance in the last few years. The authors of [11], [10] and [12] for example have used CNNs and Recurrent Neural Networks (RNNs) for the prediction of Parkinson’s disease on brain MRI and DaT scans. Since its occurrence in late 2019, a considerable number of articles have concerned themselves with using neural networks for the purpose of predicting a potential SARS-CoV-2 infection from visual data. In [12] and [10] the authors make use of clustering in order to carry out a further analysis of the produced features vectors, and classify the CT scans according to their proximity to the cluster centers.

In the context of last years ICCV there has been a challenge with the aim of detecting COVID-19 from CT images ([9]). The winners of this contest ([7]) were using contrastive learning techniques in order to improve their networks performance. [16] and [14] use 3D-CNNs for the detection of the disease, whereas the teams in [22] and [20] happen to use transformer-based architectures.

Our architecture is, similarly to other previously existing approaches, designed for the processing of 3D-data, based on the idea, that 2D architectures can be directly extended to 3D architectures ([2], [19], [13]). In our case we modified the ConvNeXt architecture ([15]) to enable 3D processing. This particular approach is characterized by architectural similarities to Vision Transformers (ViT, [5]).

3 Methods

Goal of this work is to develop a neural network architecture capable of detecting SARS-CoV-2 and to develop a separate neural network for the prediction of the severity of the disease. We apply our models to the COV19-CT-DB database ([9]), which consists of CT-scans. Each scan consists of multiple two-dimensional image slices. These slices are concatenated into a three-dimensional tensor and cubic spline interpolation is applied to get tensors of the desired shape.

ConvNeXt 3D

The architecture we utilize is a three dimensional version of the recent ConvNeXt architecture ([15]) in order to be able to process the three-dimensional input tensors. This architecture is especially characterized by multiple alterations already proven to be useful in the context of Vision Transformers that are applied to the standard ResNet ([6]).

In order to make use of the ConvNeXt architecture, we simply replace all 2D components with corresponding 3D components.
Pretraining

In all cases we started from pretraining an architecture on a version of ImageNet (\cite{4}) that has been transformed to grayscale, since the CT only have a single color channel as well. In order to obtain weights for our 3D model, we inflate the weights of the pretrained 2D model in two different ways. The first way, called \textit{full inflation}, is the commonly used option of simply copying the weights along the new tensor axis. The second, that we will call \textit{2G inflation} is based on multiplying the 2D weights along 2 axes and can best be described mathematically. Let $K \in \mathbb{R}^{I \times O \times H \times W}$ and $K' \in \mathbb{R}^{I \times O \times H \times W \times D}$ be initial 2D and the resulting 3D tensors respectively, with $D$ being the newly added dimension.

$$
\forall i, o, h, w, d : K'_{i,o,h,w,d} = K_{i,o,h,w} \cdot \mathcal{N}(d, \frac{D}{2}, \frac{D}{8}) + K_{i,o,h,w} \cdot \mathcal{N}(w, \frac{W}{2}, \frac{W}{8})
$$

Since the images in the ImageNet database are very distinct from CT-images used in this paper, we introduce various further ways to adjust the model to three-dimensional CT-scans. Therefore, we use an additional dataset designed for lung lesion segmentation in CT scans (\cite{18}, \cite{1}, \cite{3}) and the STOIC dataset created for SARS-CoV-2 severity prediction (\cite{17}). In this work we apply 4 different pretraining methods:

1. We directly use the ImageNet-pretraining weights. This approach is referred to as \textit{ImageNet model}.
2. We use a second-stage net trained for the task of segmentation on the segmentation dataset. This approach is referred to as \textit{segmentation model}.
3. Pseudo-labels are generated with the segmentation model in order to get segmentation masks for the MIA COV19D dataset. We use a new second-stage network trained for segmentation on the MIA-COV19D dataset starting from ImageNet weights. This is referred to as \textit{segmia model}.
4. Pseudo-labels are generated with the segmentation model to get segmentation masks for the STOIC dataset. We train a new second-stage network to jointly perform severity classification and segmentation for the STOIC dataset starting from ImageNet weights. This is referred to as \textit{multitask model}.

In order to be capable of generating segmentation masks in addition to severity predictions, we extend the ConvNeXt model in similar way to \cite{21}. After each pretraining, we finetune our model for severity prediction or infection detection on the MIA COV19D dataset.

\textbf{Approaches for Increased Robustness}

In order to improve the robustness of our model, a set of data augmentations is applied: random flips and rotations, Gaussian noise and Gaussian blur. We also use a GPU implementation of elastic deformations. Furthermore, we take random crops of size $(224 \times 224 \times 224)$ out of the input images of size $(256 \times 256 \times 256)$. When inspecting the data, we discovered some tensors, where the slice resolution
happened to be internally inconsistent between different slices of the same CT. In those cases we discarded the inconsistent slices. Rescaling of the tensors is performed with cubic spline interpolation and the rescaled tensors are precomputed to decrease computation time.

In order to stabilize our performance we kept a second copy of our model, whose weights are computed as exponential moving average (EMA) of the trained models weights. This copy of the network is used for the evaluations and final predictions.

Since it is probably not possible to get a good estimate of the models performance using the validation set metrics as the official validation set might be too small to give accurate results, we used cross validation in order to avoid overfitting our training hyperparameters. Whenever we used cross validation, we evaluated on the resulting ensemble of networks instead of a single trained model.

4 Results

Challenge Submission

We participate in two challenges hosted in the context of the Medical Image Analysis (MIA) workshop on this years edition of the ECCV 2022 [8]. The two challenges consist of two different tasks: The detection of COVID infections and the prediction of the severity the patient will experience.

When inspecting the dataset, we found that few CT scans had different orientations. In order to gain robustness with respect to such outliers, an augmentation that randomizes the orientation of the CT tensor is implemented and we apply this augmentation with a certain probability.

In order to prevent overfitting our hyperparameters to the validation set, we implemented a 5-fold cross validation to improve our models’ robustness and gain better results on the test set.

We use the full initialization method for our ImageNet runs, balanced Cross Entropy as loss function and no learning rate scheduler.

The validation set results for our submission for the infection detection task can be found in Table 1, whereas the validation set results for the severity prediction task can be found in Table 2.

5 Conclusion

Our paper presented a deep learning based method for the purpose of detecting SARS-CoV-2 infections and predicting the severity of the patients symptoms in the case of an existing infection. In this paper we adapted the modern architecture of ConvNeXt to 3D data and increased the performance of our model using multiple pretraining methods. We furthermore paid special interest in increasing the robustness of our model.
Table 1. The performance of the models that we trained for the covid detection task. 

**CrossVal** indicates whether we used 5-fold cross validation for the training and **RandOrProb** indicates with what probability we randomized the orientation of the CT scan. F1 scores are macro F1 scores.

| CrossVal | RandOrProb | Pretraining | F1 Validation | F1 Cross Validation |
|----------|------------|-------------|---------------|---------------------|
| Yes      | 0.0        | ImageNet    | 0.8660        | 0.9173              |
| Yes      | 0.0        | Multitask   | 0.8780        | **0.9353**          |
| Yes      | 0.0        | Segmia      | 0.8831        | 0.9333              |
| No       | 0.25       | Segmia      | **0.9203**    | -                   |
| No       | 0.0        | Segmentation| 0.918         | -                   |

Table 2. The performance of the models that we trained for the severity prediction task. **CrossVal** indicates whether we used 5-fold cross validation for the training and **RandOrProb** indicates with what probability we randomized the orientation of the CT scan. F1 scores are macro F1 scores.

| CrossVal | RandOrProb | Pretraining | F1 Validation | F1 Cross Validation |
|----------|------------|-------------|---------------|---------------------|
| Yes      | 0.0        | ImageNet    | **0.6721**    | 0.6567              |
| Yes      | 0.0        | Segmentation| 0.6343        | 0.6725              |
| Yes      | 0.0        | Segmia      | 0.6089        | 0.6648              |
| Yes      | 0.0        | Multitask   | 0.5551        | 0.6818              |
| Yes      | 0.25       | Segmentation| 0.6002        | **0.7174**          |

References

1. An, P., Xu, S., Harmon, S.A., B, T.E., Sanford, T.H., Amalou, A., Kassin, M., Varble, N., Blain, M., Anderson, V., Patella, F., Carrafiello, G., Turkbey, B., Wood, B.J.: Ct images in covid-19 [data set]. the cancer imaging archive. (2020)
2. Carreira, J., Zisserman, A.: Quo vadis, action recognition? a new model and the kinetics dataset. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 4724–4733 (2017)
3. Clark, K., Vendt, B., Smith, K., Freymann, J., Kirby, J., Koppel, P., Moore, S., Phillips, S., Maffitt, D., Pringle, M., et al.: The cancer imaging archive (tcia): maintaining and operating a public information repository. Journal of digital imaging 26(6), 1045–1057 (2013)
4. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition. pp. 248–255 (2009)
5. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., Houlsby, N.: An image is worth 16x16 words: Transformers for image recognition at scale. In: International Conference on Learning Representations (2021)
6. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (June 2016)
7. Hou, J., Xu, J., Feng, R., Zhang, Y., Shan, F., Shi, W.: Cmc-cov19d: Contrastive mixup classification for covid-19 diagnosis. In: Proceedings of the IEEE/CVF In-
8. Kollias, D., Arsenos, A., Kollias, S.: Ai-mia: Covid-19 detection & severity analysis through medical imaging. arXiv preprint arXiv:2206.04732 (2022)
9. Kollias, D., Arsenos, A., Soukissian, L., Kollias, S.: Mia-cov19d: Covid-19 detection through 3-d chest ct image analysis. arXiv preprint arXiv:2106.07524 (2021)
10. Kollias, D., Bouas, N., Vlaxos, Y., Brillaikis, V., Seferis, M., Kollia, I., Sukissian, L., Wingate, J., Kollias, S.: Deep transparent prediction through latent representation analysis. arXiv preprint arXiv:2009.07044 (2020)
11. Kollias, D., Tagaris, A., Stafylopatis, A., Kollias, S., Tagaris, G.: Deep neural architectures for prediction in healthcare. Complex & Intelligent Systems 4(2), 119–131 (2018)
12. Kollias, D., Vlaxos, Y., Seferis, M., Kollia, I., Sukissian, L., Wingate, J., Kollias, S.D.: Transparent adaptation in deep medical image diagnosis. In: TAILOR. p. 251–267 (2020)
13. Kopuklu, O., Kose, N., Gunduz, A., Rigoll, G.: Resource efficient 3d convolutional neural networks. In: Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops. pp. 0–0 (2019)
14. Liang, S., Zhang, W., Gu, Y.: A hybrid and fast deep learning framework for covid-19 detection via 3d chest ct images. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops. pp. 508–512 (October 2021)
15. Liu, Z., Mao, H., Wu, C.Y., Feichtenhofer, C., Darrell, T., Xie, S.: A convnet for the 2020s. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2022)
16. Miron, R., Moisii, C., Dinu, S., Breban, M.E.: Evaluating volumetric and slice-based approaches for covid-19 detection in chest cts. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops. pp. 529–536 (October 2021)
17. Revel, M.P., Boussoar, S., de Marjorie-Mellon, C., Saab, I., Lapotre, T., Mompoint, D., Chassagnon, G., Milon, A., Lederlin, M., Bennani, S., et al.: Study of thoracic ct in covid-19: the stoic project. Radiology (2021)
18. Roth, H.R., Xu, Z., Diez, C.T., Jacob, R.S., Zember, J., Molto, J., Li, W., Xu, S., Turkbey, B., Turkbey, E., et al.: Rapid artificial intelligence solutions in a pandemic-the covid-19-20 lung ct lesion segmentation challenge. Research Square (2021)
19. Ruiz, J., Mahmud, M., Modashir, M., Shamim Kaiser, M., Alzheimer’s Disease Neuroimaging Initiative, f.t., et al.: 3d densenet ensemble in 4-way classification of alzheimer’s disease. In: International Conference on Brain Informatics. pp. 85–96. Springer (2020)
20. Tan, W., Liu, J.: A 3d cnn network with bert for automatic covid-19 diagnosis from ct-scan images. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops. pp. 439–445 (October 2021)
21. Xiao, T., Liu, Y., Zhou, B., Jiang, Y., Sun, J.: Unified perceptual parsing for scene understanding. In: European Conference on Computer Vision. Springer (2018)
22. Zhang, L., Wen, Y.: A transformer-based framework for automatic covid19 diagnosis in chest cts. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops. pp. 513–518 (October 2021)