PLOS ONE

RESEARCH ARTICLE

Efficient, high-performance semantic segmentation using multi-scale feature extraction

Moritz Knolle[1,2☯], **Georgios Kaissis**[1,2,3,4☯], **Friederike Jungmann**[1],
Sebastian Ziegelmayer[1], **Daniel Sasse**[1], **Marcus Makowski**[1], **Daniel Rueckert**[2,4],
Rickmer Braren[1] *

1 Department of Diagnostic and Interventional Radiology, Klinikum rechts der Isar, School of Medicine, Technical University of Munich, Munich, Germany, **2** Institute for Artificial Intelligence and Informatics in Medicine, Klinikum rechts der Isar, School of Medicine, Technical University of Munich, Munich, Germany, **3** OpenMined, **4** Department of Computing, Imperial College London, London, United Kingdom

☯ These authors contributed equally to this work.
* rbraren@tum.de

Abstract

The success of deep learning in recent years has arguably been driven by the availability of large datasets for training powerful predictive algorithms. In medical applications however, the sensitive nature of the data limits the collection and exchange of large-scale datasets. Privacy-preserving and collaborative learning systems can enable the successful application of machine learning in medicine. However, collaborative protocols such as federated learning require the frequent transfer of parameter updates over a network. To enable the deployment of such protocols to a wide range of systems with varying computational performance, efficient deep learning architectures for resource-constrained environments are required. Here we present *MoNet*, a small, highly optimized neural-network-based segmentation algorithm leveraging efficient multi-scale image features. *MoNet* is a shallow, *U-Net*-like architecture based on repeated, dilated convolutions with decreasing dilation rates. We apply and test our architecture on the challenging clinical tasks of pancreatic segmentation in computed tomography (CT) images as well as brain tumor segmentation in magnetic resonance imaging (MRI) data. We assess our model's segmentation performance and demonstrate that it provides performance on par with compared architectures while providing superior out-of-sample generalization performance, outperforming larger architectures on an independent validation set, while utilizing significantly fewer parameters. We furthermore confirm the suitability of our architecture for federated learning applications by demonstrating a substantial reduction in serialized model storage requirement as a surrogate for network data transfer. Finally, we evaluate *MoNet*'s inference latency on the central processing unit (CPU) to determine its utility in environments without access to graphics processing units. Our implementation is publicly available as free and open-source software.

Introduction

Access to large collections of data remains one of the key challenges in successfully applying machine learning to many problems in medicine. Common machine learning datasets, such as ImageNet [1] with >1 million images, are much larger than their counterparts used in medical studies. Even large recent studies [2, 3] use datasets significantly smaller than ImageNet and orders of magnitude smaller than the datasets used to train state-of-the-art language models [4]. Furthermore, current medical studies often source data from only few institutions, thus preventing the training of representative and unbiased models, suitable for application in a broad variety of patient collectives [5]. Algorithms trained on single-institutional data have recently been shown to cause generalization challenges to out-of-sample data [6]. One of the main hindrances to large-scale, multi-institutional medical data collection, which could address this challenge, is the strict regulation of patient data, preventing its exchange and mandating the development of decentralized learning systems [7].

Federated machine learning [8] allows for collaborative training of algorithms on data from different hospitals (*data silos*) or edge devices (such as wearable health sensors or mobile phones) without the need for central aggregation of said data. In federated learning, a model is trained in a distributed fashion. Individual models are trained locally on data which never leaves a participating site (*node*), and only parameter updates are sent via the network to be aggregated by the coordinating node (*hub-and-spoke topology*). Federated learning enhanced by privacy-preserving techniques [9] such as differential privacy [10] holds the promise of secure, large-scale machine learning on confidential, medical data.

The utilization of federated learning techniques on the largest possible number of institutions and patients from a diverse geographic, demographic and socio-economic background will require the development of systems suitable for execution on a broad range of hardware including mobile devices and systems without graphics processing units, which may be too expensive for deployment e.g. in the developing world. A further key component of this democratization is the improvement of system efficiency, as federated learning requires the frequent transfer of parameter updates over a network. Previous work [11] has mainly focused on improving communication efficiency in federated learning by compression of parameter updates or sophisticated update aggregation methods [12, 13]. Other works have focused on increasing the computational efficiency of model architectures: The MobileNet family of models [14] utilizes depth-wise separable convolutions and a reduced parameter count to achieve this goal and target deployment on edge computing/mobile devices. The EfficientNet model architectures [15] recently proposed attempts to achieve optimal trade-offs between input resolution, network depth and width for classification performance. However, the targeted design of small and efficient neural network architectures for the specific task of semantic segmentation has so far remained under-explored. To the contrary, deep learning-based segmentation has focused on expanding model size with large ensembles of neural networks [16], rendering them impractical for deployment in the federated setting.

Here, we introduce *MoNet*, a very small, shallow, *U-Net*-derived semantic segmentation architecture based on efficient multi-scale feature extraction using repeated decreasingly dilated convolution (RDDC) layers with two global down-sampling operations and a total of 403,556 parameters. We showcase our architecture's performance on the challenging task of pancreatic segmentation as well as brain tumor segmentation and demonstrate substantial efficiency gains and segmentation performance competitive with much larger models.

Methods

Training, validation and independent testing datasets

All neural network architectures presented in this work were trained two different datasets from the Medical Segmentation Decathlon (MSD) [17]: pancreas and brain tumor segmentation. A random, consistent 70%/30% training-validation split was employed for both datasets. For processing, images were bilinearly down-sampled to 256×256, and the segmentation labels were merged yielding a binary segmentation task. To assess out-of-sample generalization performance on the pancreas dataset, independent validation of the architectures was performed on an unseen, clinical PDAC dataset consisting of 85 abdominal CT scans in the portal-venous phase collected at our institution. For the brain tumor dataset, no in-house clinical dataset was available. All clinical data were collected according to Good Clinical Practice and in consent with the Declaration of Helsinki. The use of imaging data was approved by the institutional ethics committee (Ethikkommission der Fakultät für Medizin der Technischen Universität München, protocol number 180/17S, May 9th 2017) and the requirement for informed written consent was waived. The pancreas including the tumor was manually segmented by a third-year radiology resident, then checked and corrected as necessary by a sub-specialized abdominal radiologist. An exemplary ground truth label mask superimposed on a CT slice from the training set is shown in Fig 1.

Network architecture

The architecture of *MoNet* is depicted in Fig 2. In brief, 4-dimensional input tensors of shape $B \times 256 \times 256 \times 1$, with B denoting the batch size, are progressively down-sampled across the encoder branch of the network using convolutions with a stride length of 2, resulting in an $X \times Y$ resolution of 64×64 in the *bottleneck* segment of the network. The resulting feature maps are then progressively up-sampled by transposed convolution (*de-convolution*) in the decoder
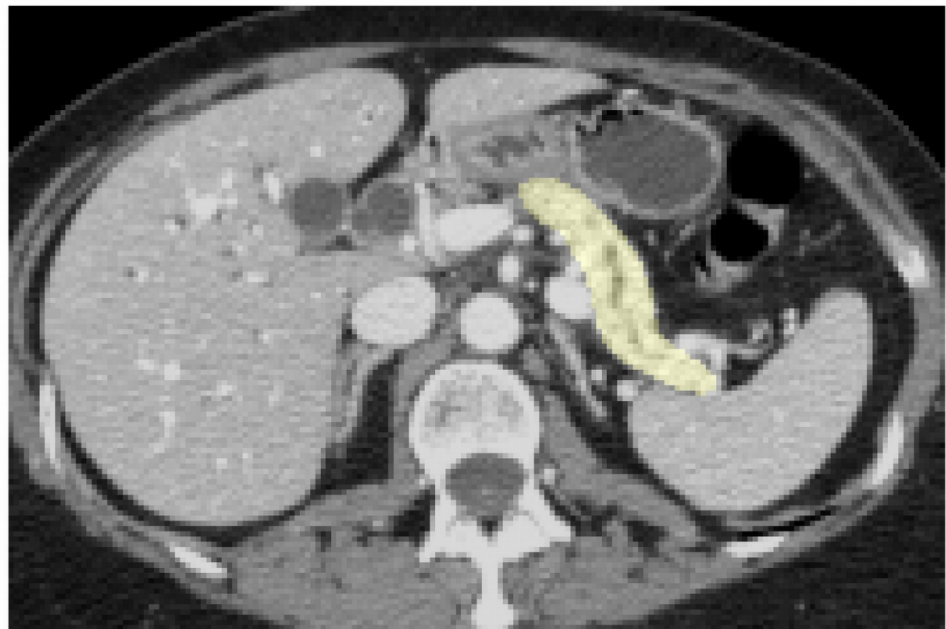

Fig 1. Axial slice of a ground truth pancreas segmentation in an abdominal CT scan (MSD), cropped to show detail of surrounding tissues.
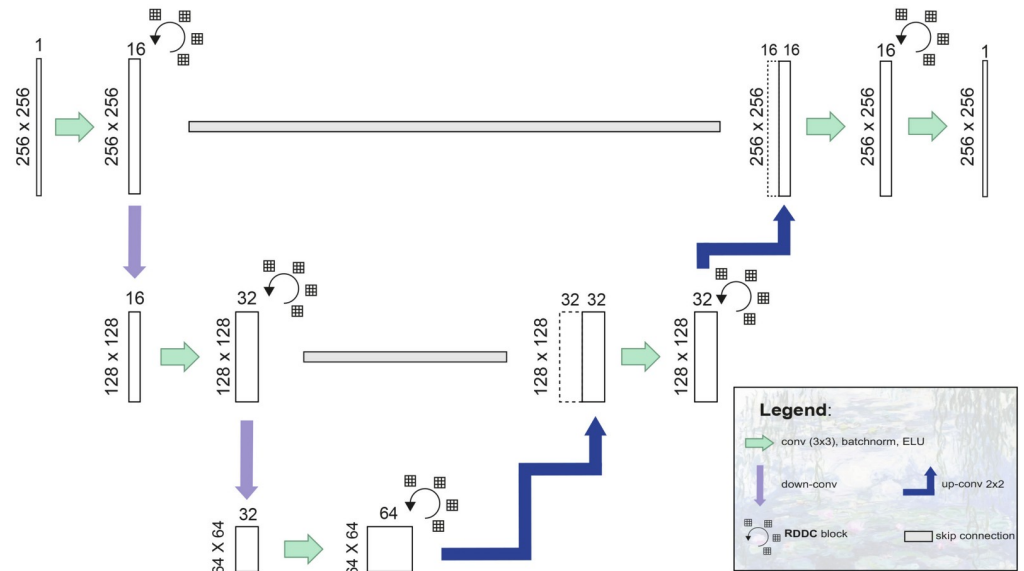
Fig 2. Schematic representation of the *MoNet* architecture.

branch resulting in output masks of identical dimensions as the input. Each (de-)convolution block consists of a 3x3 convolutional layer followed by batch normalization and an *exponential linear unit* (ELU) activation. At every stage in the *U-Net*-like architecture, the convolution blocks are followed by a *repeated decreasingly dilated convolution* (RDDC) block (Fig 3), consisting of four successive convolutional blocks as described above, but employing dilated convolutions [18] with a decreasing dilation rate (4, 3, 2, 1, respectively). This feature extraction strategy has been shown to perform well for small objects [19]. Each convolutional block within a *RDDC* block is followed by a spatial dropout layer [20]. Finally, residual-type longitudinal (short) connections are employed within each RDDC block and transverse (long) skip connections are employed between the encoder and the decoder branch to assist signal and gradient flow as originally described in [21, 22].

Model training

All architectures were trained to convergence using the *Nesterov-Adam* optimizer [23] with an initial learning rate of 5×10^{-4} and learning rate decay by a factor 10 upon validation loss
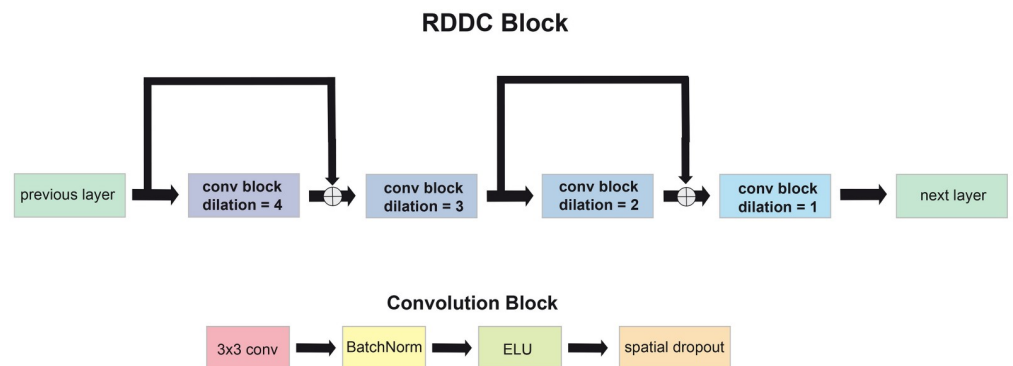

Fig 3. Schematic representation diagram of a *RDDC* block (top) and the constituent convolutional (bottom).

stagnation for ≥ 2 epochs. Weights were initialized using uniform He-initialization [22] and the Dice loss [24] was used to train all networks. Data augmentation was used in the form of random rotations up to 10 degrees, random zoom (±0.25) and random pixel shifts of a maximum magnitude of 0.2 of the image height/width. All architectures were trained to segment the entire pancreas including the tumor. This approach is owing to the fact that the exact delineation of the tumor border is often times infeasible and supported by literature findings noting the importance of the peritumoral tissue in PDAC [25–27] and in other tumor entities [28]. To maintain comparability, we also merged the labels in the brain tumor dataset to obtain a binary segmentation task.

Performance assessment

We compared *MoNet*'s performance to the following three *U-Net* baselines:

- original *U-Net* [21], 64 base filters (*U-Net-64*)

- original *U-Net* [21], 16 base filters (*U-Net-16*)

- Attention-gated *U-Net* [29], 2D, 64 base filters (*Attention U-Net*)

Results

Segmentation performance comparison

MoNet performed similarly or on par with other *U-Net* variants on the validation dataset (pancreas & brain tumor) while outperforming the other *U-Net* variants in out-of-sample generalization on the independent validation dataset (pancreas only). Results are summarized in Table 1 and visualized in Fig 4.

Training speed & inference time comparison

To compare the performance of *MoNet* in a typical inference setting on CPU, as well as when performing GPU training. We recorded the time required for doing inference with 150 256×256 images on CPU (2.4GHz 8-Core Intel Core i9) and the time per batch when training on GPU (batch size = 32). All experiments were performed with identical batch size and otherwise consistent environment for all architectures with $N = 5$ repetitions. *MoNet* significantly outperformed both *U-Net-64* and *Attention U-Net* with regards to inference and training time. Results are shown in Table 2.

Table 1. Comparison of *MoNet* with other *U-Net* variants in two different imaging modalities on the task of pancreas and brain lesion segmentation, CT and MRI respectively. We report performance on validation sets of the MSD datasets (brain tumor and pancreas) as well as out-of sample generalization performance on an independent validation set (IVD), collected and annotated in-house.

Task	Architecture	Dice Score, MSD	Hausdorff Distance, MSD	Dice Score, IVD
Pancreas	*U-Net-64*	0.76±0.11	1.86±0.66	0.50±0.2
Pancreas	*U-Net-16*	0.73±0.10	1.96±0.66	0.59±0.2
Pancreas	*Attention U-Net*	0.73±0.10	1.98±0.66	0.37±0.6
Pancreas	*MoNet (ours)*	**0.78±0.090**	**1.78±0.61**	**0.70±0.1**
Brain Tumor	*U-Net-64*	**0.78±0.12**	**2.01±0.38**	—
Brain Tumor	*U-Net-16*	0.77±0.12	2.03±0.38	—
Brain Tumor	*Attention U-Net*	0.74±0.15	2.09±0.38	—
Brain Tumor	*MoNet(ours)*	0.77±0.13	2.04±0.39	—

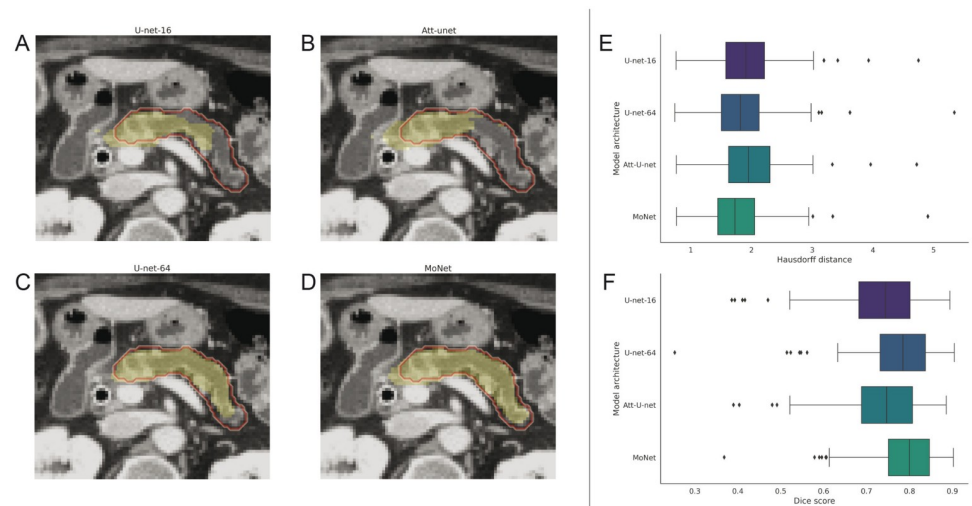https://doi.org/10.1371/journal.pone.0255397.t001

Fig 4. Exemplary segmentation results (yellow) of: *U-Net-16* (A), *Attention U-Net* (B), *U-Net-64* (C), *MoNet* (D), on the pancreas MSD validation set, Ground truth indicated by red outline. Box-plots of Hausdorff distances (E) and Dice scores (F) computed for the whole pancreas MSD validation set on a per-patient basis.

https://doi.org/10.1371/journal.pone.0255397.g004

Serialized model size as an indicator for network traffic in federated learning

We performed a comparison of the size taken up by the weights of *MoNet* and the other *U-Net* like architectures. Federated learning requires the frequent transfer of parameter updates over a network, hence the serialized model size of a given architecture can serve as an estimate of the amount of network traffic generated when deployed in a federated learning application. *MoNet* with its small number of parameters is significantly smaller in size than *U-Net-16* and an order of magnitude smaller than *U-Net-64* and *Attention U-Net*. Results are shown in Table 3.

Visualization of intermediate activations

To corroborate our hypothesis that *MoNet* achieves superior semantic segmentation performance due to an improved utilization of its convolutional filters, leading to more information-rich feature maps, we examined differences between the features extracted by *U-Net-64* and *MoNet* at early, intermediate and late convolutional layers. *MoNet* extracts feature maps with overall higher resolution. Moreover, we found the filter activations at all examined layers of *U-Net-64* to collapse to a region near zero. We thus assume that many of these filters remain essentially unutilised. On the contrary, *MoNet* produced non-zero activations at all examined layers. These results are presented in Fig 5.

Table 2. CPU inference time (sec) for a CT scan of 150 slices and timer per batch (sec) on GPU, both at a resolution of 256 × 256.

Architecture	Inference times(s) (CPU)	Time per batch(s) (GPU)
U-Net-64	45.34±1.77	0.747±0.30
U-Net-16	**7.03±0.21**	**0.152±0.11**
Attention U-Net	53.30±0.53	0.905±0.12
MoNet (ours)	14.88±0.32	0.462±0.16

https://doi.org/10.1371/journal.pone.0255397.t002

Table 3. Comparison of storage space occupied by _MoNet_ and other _U-Net_ variants.

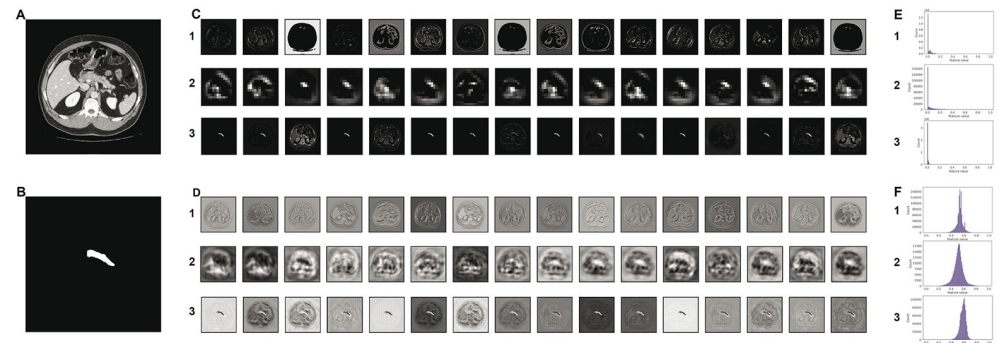| Architecture	Parameter count	Size in memory
U-Net-64	31, 054, 145	118.6 MB
U-Net-16	1, 946, 705	7.6 MB
MoNet (ours)	**403, 556**	**1.8 MB**

Fig 5. Input image (A) and target (B) alongside visualizations of the first 16 channels of intermediate activations for the given input image produced by early (1), middle (2) and late (3) convolutional layers in _U-net_ (C) and _MoNet_ (D). Histograms computed for all channels in the feature maps for early (1), middle (2), and late (3) convolution layers for _U-net_ (E) and _MoNet_ (F).

Discussion

We present an efficient, high-performance _U-Net_-like segmentation algorithm and show a substantial reduction in parameter count and expected network traffic in federated learning applications (indicated by serialized model size). Compared to _U-Net-64_ and _Attention U-Net_, our method achieves a substantial inference latency reduction on CPU hardware, enabling remote diagnosis applications in centers without GPUs. Furthermore we show reduced training time on GPU, which could benefit federated training as well as swift fine-tuning when model personalisation is required. Both are made possible by our method while exceeding or matching the segmentation performance of all other evaluated algorithms. We thus believe our architecture to be a promising candidate for utilization in large-scale collaborative medical imaging workflows and particularly in resource constrained environments.

We chose the tasks of pancreatic segmentation and brain tumor segmentation due to the poor prognosis and increasing incidence of PDAC [30, 31] and the typically dismal prognosis of brain tumors [32], both of which mandate the development of enhanced diagnosis and treatment strategies. Our recent findings suggest that quantitative image analysis can identify molecular subtypes related to different response to chemotherapeutic drugs [33] or predict patient survival [34] in PDAC. In all quantitative imaging workflows, automated region-of-interest definition increases the reliability and validity of such findings, and offers substantial time savings compared to manual expert-based segmentation. However, the success of automated segmentation algorithms is constrained by the findings' poor differentiability from adjacent structures of similar attenuation/signal, variability in position of the segmentation target and alterations due to pathology such as edema or other inflammatory changes. Existent work in deep learning-assisted semantic segmentation of medical images and the pancreas in particular has focused on expanding previously available architectures such as the _U-Net_ [21]

into the three-dimensional context [24] or on improving segmentation results by incorporating attention mechanisms into the architecture [29]. Other approaches have used complex ensembles of 2D and 3D models to extract the maximum amount of information in the CT images [16]. All these modifications however result in a further increase in the (already substantial) computational requirements of these architectures, rendering such *U-Net* derivatives impractical for the utilization in the above-mentioned decentralized learning applications. In other application domains, i.e. image classification, MobileNet and EfficientNet have demonstrated strong performance combined with high computational efficiency. MobileNet achieves this through the utilisation of depth-wise separable convolutions and EfficientNet through optimal trade-offs between model depth and width. In contrast, our method exploits properties of the feature space specific to semantic segmentation through the utilization of higher resolution feature maps in the *bottleneck* section of the network, and thus enables competitive segmentation performance with the state-of-the-art while offering substantial efficiency gains. Recent work on semantic segmentation provides evidence in favor of architectures performing image feature extraction at multiple scales by utilizing dilated convolutions instead of relying merely on the scale-decreasing backbones employed in traditional fully convolutional architectures [19, 35–37]. Our work corroborates this notion, since multi-scale feature extraction combined with larger receptive fields at the same hierarchical level seem to capture both more robust and higher quality features compared to the fixed kernel size design encountered in traditional *U-Net*-like architectures. Moreover, architectures with several down-sampling operations and/or many filters such as the *U-Net* (with 4 down-sampling stages) cannot leverage the large number of parameters sufficiently well to warrant their utilization at least in medical imaging tasks, typically characterized by small segmentation targets (such as the pancreas or small tumors). This is substantiated by our results from *U-Net-64*'s activation histograms, which were concentrated at a near-zero region.

Our results indicate that *MoNet* extracts more robust features that generalize better to out-of-sample data than the compared methods, as shown by *MoNet*'s performance on the independent validation set and the activation histograms. The poor performance of the 64 filter *U-Net* and *Attention U-Net* in the out-of-sample generalization challenge could potentially be caused by the overparameterization of these architectures, making them prone to over-fitting the data-generating distribution of the training data, while the two smaller models (*U-Net-16* and *MoNet*) tested seemed to generalize better to the out-of-sample data, supporting this hypothesis.

Our work is not without limitations. The generalizability of our findings should be confirmed using larger, multi-institutional training and validation sets. Furthermore, we only compared our algorithm against models based on the use of a single 2D *U-Net*-style network. Algorithms such as *nnU-Net* [16] based on *U-Net* ensembles offer superior performance, however at the expense of extremely high computational and post-processing requirements and thus much slower inference times (especially on CPU). Furthermore implementing and establishing a real-world federated learning application was out of scope for this study and will be addressed in future work.

Conclusion

In conclusion, we propose an optimized semantic segmentation algorithm with small size and low inference latency, particularly suited for decentralized applications such as federated learning. Our work can benefit both, radiological research and clinical translation of artificial intelligence workflows in medical imaging by providing consistent, high-quality segmentation for machine learning tasks.

Acknowledgments

Source code

Source code for *MoNet* based on TensorFlow is available at https://github.com/TUM-AIMED/MoNet.

Author Contributions

Conceptualization: Moritz Knolle, Georgios Kaissis, Rickmer Braren.

Data curation: Moritz Knolle, Georgios Kaissis, Friederike Jungmann, Sebastian Ziegelmayer, Daniel Sasse.

Formal analysis: Georgios Kaissis.

Funding acquisition: Georgios Kaissis, Rickmer Braren.

Investigation: Moritz Knolle, Georgios Kaissis, Friederike Jungmann.

Methodology: Moritz Knolle, Georgios Kaissis, Friederike Jungmann.

Project administration: Daniel Rueckert.

Resources: Daniel Rueckert.

Software: Moritz Knolle, Georgios Kaissis.

Supervision: Marcus Makowski, Daniel Rueckert, Rickmer Braren.

Validation: Moritz Knolle, Georgios Kaissis.

Visualization: Moritz Knolle, Georgios Kaissis.

Writing – original draft: Moritz Knolle, Georgios Kaissis, Friederike Jungmann, Rickmer Braren.

Writing – review & editing: Moritz Knolle, Georgios Kaissis, Friederike Jungmann, Sebastian Ziegelmayer, Daniel Sasse, Marcus Makowski, Daniel Rueckert, Rickmer Braren.

References

1. Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, et al. Imagenet large scale visual recognition challenge. International journal of computer vision. 2015; 115(3):211–252. https://doi.org/10.1007/s11263-015-0816-y

2. McKinney SM, Sieniek M, Godbole V, Godwin J, Antropova N, Ashrafian H, et al. International evaluation of an AI system for breast cancer screening. Nature. 2020; 577(7788):89–94. https://doi.org/10.1038/s41586-019-1799-6 PMID: 31894144

3. Bora A, Balasubramanian S, Babenko B, Virmani S, Venugopalan S, Mitani A, et al. Predicting the risk of developing diabetic retinopathy using deep learning. The Lancet Digital Health. 2020;. PMID: 33735063

4. Devlin J, Chang MW, Lee K, Toutanova K. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:181004805. 2018;.

5. Obermeyer Z, Mullainathan S. Dissecting racial bias in an algorithm that guides health decisions for 70 million people. In: Proceedings of the Conference on Fairness, Accountability, and Transparency; 2019. p. 89–89.

6.   Zhang L, Wang X, Yang D, Sanford T, Harmon S, Turkbey B, et al. Generalizing deep learning for medical image segmentation to unseen domains via deep stacked transformation. IEEE Transactions on Medical Imaging. 2020;.

7.   Winter JS, Davidson E. Big data governance of personal health information and challenges to contextual integrity. The Information Society. 2019; 35(1):36–51. https://doi.org/10.1080/01972243.2018.1542648

8.   Konečnỳ J, McMahan B, Ramage D. Federated optimization: Distributed optimization beyond the datacenter. arXiv preprint arXiv:151103575. 2015;.

9.   Kaissis GA, Makowski MR, Rückert D, Braren RF. Secure, privacy-preserving and federated machine learning in medical imaging. Nature Machine Intelligence. 2020; p. 1–7.

10.   Dwork C, Roth A, et al. The algorithmic foundations of differential privacy. Foundations and Trends in Theoretical Computer Science. 2014; 9(3-4):211–407. https://doi.org/10.1561/0400000042

11.   Li T, Sahu AK, Talwalkar A, Smith V. Federated learning: Challenges, methods, and future directions. IEEE Signal Processing Magazine. 2020; 37(3):50–60. https://doi.org/10.1109/MSP.2020.2975749

12.   Konečnỳ J, McMahan HB, Yu FX, Richtárik P, Suresh AT, Bacon D. Federated learning: Strategies for improving communication efficiency. arXiv preprint arXiv:161005492. 2016;.

13.   Bonawitz K, Eichner H, Grieskamp W, Huba D, Ingerman A, Ivanov V, et al. Towards federated learning at scale: System design. arXiv preprint arXiv:190201046. 2019;.

14.   Howard AG, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:170404861. 2017;.

15.   Tan M, Le Q. Efficientnet: Rethinking model scaling for convolutional neural networks. In: International Conference on Machine Learning. PMLR; 2019. p. 6105–6114.

16.   Isensee F, Petersen J, Klein A, Zimmerer D, Jaeger PF, Kohl S, et al. nnu-net: Self-adapting framework for u-net-based medical image segmentation. arXiv preprint arXiv:180910486. 2018;.

17.   Simpson AL, Antonelli M, Bakas S, Bilello M, Farahani K, Van Ginneken B, et al. A large annotated medical image dataset for the development and evaluation of segmentation algorithms. arXiv preprint arXiv:190209063. 2019;.

18.   Holschneider M, Kronland-Martinet R, Morlet J, Tchamitchian P. A real-time algorithm for signal analysis with the help of the wavelet transform. In: Wavelets. Springer; 1990. p. 286–297. https://doi.org/10.1007/978-3-642-75988-8_28

19.   Hamaguchi R, Fujita A, Nemoto K, Imaizumi T, Hikosaka S. Effective use of dilated convolutions for segmenting small object instances in remote sensing imagery. In: 2018 IEEE winter conference on applications of computer vision (WACV). IEEE; 2018. p. 1442–1450.

20.   Tompson J, Goroshin R, Jain A, LeCun Y, Bregler C. Efficient object localization using convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2015. p. 648–656.

21.   Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. Springer; 2015. p. 234–241.

22.   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; 2016. p. 770–778.

23.   Dozat T. Incorporating nesterov momentum into adam. ICLR 2016 Workshop. 2016;.

24.   Milletari F, Navab N, Ahmadi SA. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In: 2016 fourth international conference on 3D vision (3DV). IEEE; 2016. p. 565–571.

25.   Bauer AS, Nazarov PV, Giese NA, Beghelli S, Heller A, Greenhalf W, et al. Transcriptional variations in the wider peritumoral tissue environment of pancreatic cancer. International Journal of Cancer. 2017; 142(5):1010–1021. https://doi.org/10.1002/ijc.31087 PMID: 28983920

26.   Fukushima N, Koopmann J, Sato N, Prasad N, Carvalho R, Leach SD, et al. Gene expression alterations in the non-neoplastic parenchyma adjacent to infiltrating pancreatic ductal adenocarcinoma. Modern Pathology. 2005; 18(6):779–787. https://doi.org/10.1038/modpathol.3800337 PMID: 15791284

27.   Infante JR, Matsubayashi H, Sato N, Tonascia J, Klein AP, Riall TA, et al. Peritumoral Fibroblast SPARC Expression and Patient Outcome With Resectable Pancreatic Adenocarcinoma. Journal of Clinical Oncology. 2007; 25(3):319–325. https://doi.org/10.1200/JCO.2006.07.8824 PMID: 17235047

28.   Sun Q, Lin X, Zhao Y, Li L, Yan K, Liang D, et al. Deep Learning vs. Radiomics for Predicting Axillary Lymph Node Metastasis of Breast Cancer Using Ultrasound Images: Don't Forget the Peritumoral Region. Frontiers in Oncology. 2020; 10. https://doi.org/10.3389/fonc.2020.00053 PMID: 32083007

29. Oktay O, Schlemper J, Folgoc LL, Lee M, Heinrich M, Misawa K, et al. Attention u-net: Learning where to look for the pancreas. arXiv preprint arXiv:180403999. 2018;.

30. American Cancer Society. Cancer Facts & Figures 2020;. Available from: https://www.cancer.org/cancer/pancreatic-cancer/detection-diagnosis-staging/survival-rates.html#references.

31. Collisson EA, Bailey P, Chang DK, Biankin AV. Molecular subtypes of pancreatic cancer. Nature Reviews Gastroenterology & Hepatology. 2019; 16(4):207–220. https://doi.org/10.1038/s41575-019-0109-y PMID: 30718832

32. Patel AP, Fisher JL, Nichols E, Abd-Allah F, Abdela J, Abdelalim A, et al. Global, regional, and national burden of brain and other CNS cancer, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. The Lancet Neurology. 2019; 18(4):376–393. https://doi.org/10.1016/S1474-4422(18)30468-X

33. Kaissis GA, Ziegelmayer S, Lohöfer FK, Harder FN, Jungmann F, Sasse D, et al. Image-Based Molecular Phenotyping of Pancreatic Ductal Adenocarcinoma. Journal of Clinical Medicine. 2020; 9(3):724. https://doi.org/10.3390/jcm9030724 PMID: 32155990

34. Kaissis GA, Jungmann F, Ziegelmayer S, Lohöfer FK, Harder FN, Schlitter AM, et al. Multiparametric Modelling of Survival in Pancreatic Ductal Adenocarcinoma Using Clinical, Histomorphological, Genetic and Image-Derived Parameters. Journal of Clinical Medicine. 2020; 9(5):1250. https://doi.org/10.3390/jcm9051250 PMID: 32344944

35. Du X, Lin TY, Jin P, Ghiasi G, Tan M, Cui Y, et al. SpineNet: Learning scale-permuted backbone for recognition and localization. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2020. p. 11592–11601.

36. Chen LC, Papandreou G, Schroff F, Adam H. Rethinking atrous convolution for semantic image segmentation. arXiv preprint arXiv:170605587. 2017;.

37. Yu F, Koltun V. Multi-scale context aggregation by dilated convolutions. arXiv preprint arXiv:151107122. 2015;.