PROCEEDINGS OF SPIE

SPIEDigitalLibrary.org/conference-proceedings-of-spie

Lesion-wise evaluation for effective performance monitoring of small object segmentation

Groothuis, Irme, Sudre, Carole, Ingala, Silvia, Barnes, Jo, Gispert, Juan Domingo, et al.

Irme Groothuis, Carole H. Sudre, Silvia Ingala, Jo Barnes, Juan Domingo Gispert, Lauge Sørensen, Akshay Pai, Mads Nielsen, Sebastien Ourselin, M. Jorge Cardoso, Frederik Barkhof, Marc Modat, "Lesion-wise evaluation for effective performance monitoring of small object segmentation," Proc. SPIE 11596, Medical Imaging 2021: Image Processing, 1159608 (15 February 2021); doi: 10.1117/12.2580734



Event: SPIE Medical Imaging, 2021, Online Only

Lesion-wise evaluation for effective performance monitoring of small object segmentation

Irme Groothuis^a, Carole H. Sudre^{a,b}, Silvia Ingala^c, Jo Barnes^b, Juan Domingo Gispert^e, Lauge Sørensen^d, Akshay Pai^d, Mads Nielsen^d, Sebastien Ourselin^b, M. Jorge Cardoso^b, Frederik Barkhof^{c,b}, and Marc Modat^a

^aSchool of Biomedical Engineering & Imaging Sciences, King's College London

^bUniversity College London

^cVU Medical Centre Amsterdam

^dCerebriu A/S, Biomediq A/S

^eBarcelonabeta Brain Research Center (BBRC)

ABSTRACT

Object detection in medical images using deep learning is a challenging task, due to the imbalance often present in the data. Deep learning algorithms require large amount of balanced data to achieve optimal performance, as well as close monitoring and fine-tuning of hyper parameters. For most applications, such performance monitoring is done by simply feeding unseen data trough the network, and then using the loss function for evaluation. In the case of small or sparse objects, the loss function might not able to describe the features needed, but such features can be hard to capture in a loss function. In this paper we introduce a lesion-wise whole volume validation tool, which allows more a more accurate performance monitoring of segmentation of small and sparse objects. We showcase the efficacy of our tool by applying it to the task of microbleed segmentation, and compare the behaviour of lesionwise-whole volume validation compared to well known segmentation loss functions. Microbleeds are visible as small (less than 10 mm), ovoid hypo-intensities on T2*-weighted and susceptibility weighted magnetic resonance images. Detection of microbleeds is clinically relevant, as microbleeds can indicate the risk of recurrent stroke, and are used as imaging biomarker for various neurodegenerative diseases. Manual detection or segmentation is time consuming and error prone, and suffers from high inter- and intra-observer variability. Due to the sparsity and small size of the lesions, the data is severely imbalanced.

Keywords: Data imbalance, microbleed, evaluation

1. INTRODUCTION

Deep learning is a powerful tool to analyse medical images, but adequate performance is often dependent on the quality and amount of training data, as well hyper parameters chosen. Performance monitoring in deep learning applications is essential to fine-tune applications to reach optimal performance. Loss functions are used both for training and evaluation, and a network's performance is heavily dependent on how well the loss function describes the features of the ground truth. Depending on the task at hand, different functions can be used.

In the case of tasks in which the elements of interest are very sparse (subject-wise imbalance) and/or very small (voxel-wise imbalance), deep learning frameworks are often strongly challenged and specific solutions must be designed to enable the learning process despite those challenges. In particular, the Dice loss, based on the Dice measure of overlap, which is often successfully used for medical image segmentation tasks fails to cope with very imbalanced challenges. For instance, when training a network to segment objects of different sizes with the Dice loss, small objects, that do not contribute much to the overall amount of overlap voxel-wise, tend to be overlooked and are often ignored. As a result and due to the severe data imbalance associated with the task of segmenting small and sparse lesions, common approaches are rendered ineffective, where it is more loss effective for a network to always return empty segmentation rather than miss-segmenting lesions. Most loss functions

Further author information: irme.m.groothuis@kcl.ac.uk

Medical Imaging 2021: Image Processing, edited by Ivana Išgum, Bennett A. Landman, Proc. of SPIE Vol. 11596, 1159608 · © 2021 SPIE · CCC code: 1605-7422/21/\$21 · doi: 10.1117/12.2580734

will not properly reflect this behaviour, as the prediction for the majority of the image, the background, is still correct. We introduce a lesion-wise evaluation during training, which enables to monitor the true performance of the network, and make informed decisions regarding over-fitting, early stopping or convergence as well as for hyper-parameters tuning. This evaluation relies on measures that are otherwise only reported after training, as final measures of performance. These measures include for example the true positive rate, or the average amount of false positive lesions predicted. We argue that such measures are more informative of network performance for small objects segmentation than the loss function on its own. Additionally, we used a whole volume evaluation approach, which is also more informative than a patch-based evaluation even when only patches of the whole image are used at a time for training. We apply our framework to the task of microbleed segmentation.

From a technical standpoint, the main challenge of microbleed segmentation lies in the data imbalance due to the combination of subject- and voxel-wise imbalance. Not only the microbleeds are very small, but they also occur quite rarely; only 6% of individuals in a healthy ageing population will have microbleeds and this number only raises to 20% in subjects with Alzheimer's disease for instance. Furthermore microbleeds vary greatly within their limited size range, as can be seen in Figure 2. This double source of data imbalance results in the failure of classical segmentation approaches. Cerebral microbleeds (CMB)are small deposits of hemosiderin that are commonly observed as areas of hypo-intensity on T2*-weighted and susceptibility weighted imaging (SWI) MRI scans, 3 as shown in Figure 1, Ovoid in shape, CMBs are less than 10mm of diameter. 4

The clinical need for microbleed segmentation is based on their value as an imaging biomarker for a variety of pathologies. The number of microbleeds and their localisation are strongly related to the underlying etiology. Today, in both clinical and research settings, CMB are typically manually counted and/or segmented. Such tasks are very time-consuming and error-prone and require a high level of expertise due to the large number of mimics potentially present in images. Manual counting and segmentation thus suffer from high inter- and intra-observer variability.⁵

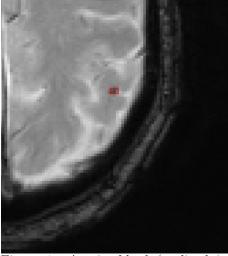


Figure 1: A microbleed (outlined in red) as seen in a T2* scan.

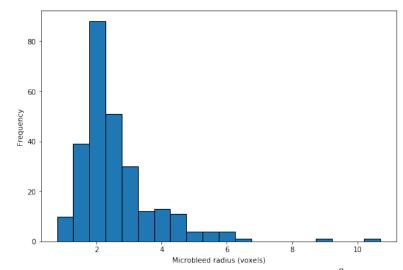


Figure 2: Distribution of microbleed radii in voxels, 1 mm³ isotropic.

Due to the small size of the CMBs, detection can be hindered by decreased contrast caused by partial volume effect occurring when multiple tissue types contribute simultaneously to the intensity of a given voxel. While CMBs are visible on both T2*-weighted and SWI as aforementioned, SWI are usually more sensitive to microbleeds and thus, when available, often the sequence of choice to detect them. It has been shown that the phase information from SWI could facilitate the distinction between microbleeds and possible mimics.⁶ In addition, some known mimics such as calcifications have a different intensity signature on SWI where they appear bright, whereas on T2* they appear dark, similar to microbleeds.

Other sources of mimics are commonly due to interference with the skull or patient motion.⁷ For these

reasons, most CMB segmentation approaches use SWI as input. T2*-weighted images are however more readily available as acquired as part of standard clinical protocols. A working tool for microbleed segmentation from T2*-weighted image could thus have a great impact.

2. RELATED WORK

Imbalanced tasks can be overcome in various ways, most commonly using over- & under-sampling, task specific loss functions and data augmentation. Object counting and object detection are similar tasks, but systems developed for dense object counting to not expand well to sparse data.^{8,9} Oversampling of the minority class is often used to overcome data imbalance, but this can lead to over-fitting or a high amount of false positive predictions at test time.¹⁰

In the field of deep learning, there are many networks that are able to detect various objects, and R-CNN is a well known two-phase approach.¹¹ R-CNN has been applied to medical imaging to the task of extremely small object counting. In this paper, Sudre *et al.* use the two-phase approach to detect small and sparse objects, whilst taking rater uncertainty into account.¹² Long nodule detection aims to to detect small objects, with varying frequency and size. As the imbalance present in the data impedes training, long nodule detection algorithms employ a two-phase approach, in which the first phase is used for candidate selection and the second phase for refinement.^{13,14}

Similarly, most deep learning microbleed segmentation approaches take a 2-phase approach, in which the first phase serves as candidate detection, and the second phase as candidate refinement. Dou $et\ al.$ introduced a deep learning based two-phase approach (detection and classification) for microbleed segmentation from SWI. ¹⁵ This work was later used by Chen $et\ al.$ They achieve a sensitivity of 92.31%, and a false positive rate of 2.90 per image.

3. METHODS

In order to demonstrate the advantages of lesion-wise evaluation, we trained multiple models using different losses. During validation iterations, performance was evaluated using the whole volume lesion wise approach, as well as standard validation approach.

3.1 Lesion-wise whole volume evaluation approach

To evaluate the performance of the network, we used an approach specifically tailored for our application where the loss function used for training provides little insight about the network performance. Due to the sparsity and small size of the microbleeds, patch-based training is employed. Patches are generated to contain a microbleed at a uniformly distributed location within the patch. The presence of microbleeds within the patch is required to ensure the network can learn despite the very imbalanced data. During training, every 1000 iterations validation iterations are performed. Both standard validation and lesions-wise whole volume validation are performed. In a standard validation iteration, a patch is sampled from the validation set in the same manner as from the training set, meaning that microbleed will be present within the patch. This patch is fed through the network, and the outcome is evaluated using the loss function. During lesion-wise whole volume validation, the patches are sampled using a sliding window approach, allowing the entire image to be fed trough the network in patches. After being processed by the network, the image is reconstructed into its original shape. A visual representation can be seen in figure 3. Now, the lesion-wise evaluation will be performed. In our case of microbleed segmentation, connected component analysis is done on both the ground truth and the proposed segmentation. Predicted lesions are classified as ground truth if there is an overlap of at least 10% between the ground truth and segmentation. If less, the lesion is classified as a false negative. If a predicted lesion covers two microbleeds in the ground truth, the lesion with the most overlap will be counted as correct and the other as a false negative. Whilst the network has correctly localised the lesion, the total count of lesions is incorrect. Each time an evaluation is done, we run a forward pass of the current version of the model on all available data used for training and validation.

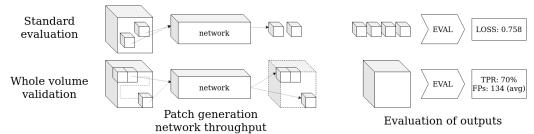


Figure 3: A visual representation of how lesion-wise whole volume evaluation compares to a standard validation approach, regarding network throughput and evaluation of results.

3.2 Network set-up

All models were trained for 100k iterations, using a 4 resolution-layer deep 3D U-net¹⁷ in PyTorch using an Ignite Engine. Apart from the loss functions used for training, identical hyper-parameters were used: the learning rate was set to 1e-4 with no decay, an L2 regularisation of 0.02 was applied, dropout was set to 30%, and random flipping of axes was applied with a 50% probability. Patches were generated according to a sampling strategy ensuring a microbleed would be present at a uniformly distributed location withing the patch.

By using a large patch size of $96 \times 96 \times 96$ voxels and ensuring a positive example in each patch, we are maintaining a data imbalance similar to the one at validation and test time.

3.3 Losses

The benefits of our proposed lesion-wise evaluation were highlighted by monitoring the performance of models trained with six different loss functions. The (i) Dice and (ii) cross entropy losses were chosen as they are commonly used loss functions for medical image segmentation tasks. Imbalance targeted loss functions were also included. The (iii) generalised Dice loss, which uses the class volume to re-weight the Dice, aims at preventing the minority class from being overwhelmed.² The Tversky similarity index at the basis of the (iv) Tversky loss proposes to cost differently the contribution of false positives and false negatives to the overall error.²⁰ It was implemented using the recommended values of $\alpha = 3$ and $\beta = 7$. A (v) hybrid loss combining Dice and cross entropy was introduced by Isensee *et al.* in,²¹ showing good performance on medical segmentation tasks. Extending it to (vi) combine generalised Dice and cross entropy seemed an intuitive adaptation to imbalanced data problems.

3.4 Data

Data from two different sources have been combined for our experiments, namely the ADNI database and the ALFA cohort. The Alzheimer's Disease Neuroimaging Initiative (ADNI) is a large multicentre study of elderly people (healthy controls, mild cognitive impairment and Alzheimer's disease), which includes a series of cognitive test outputs and MRI scans. Among the first-visit T2*-weighted MRI scans from the ADNI1 database, microbleeds were visually found in 132 cases. For each subject, they were manually identified and labelled as possible or definite lesions by a trained operator, according to the Microbleed Anatomical Rating Scale (MARS). The ALFA study (Alzheimer's and Family) is a single centre study, with one dedicated MR scanner, that involves Alzheimer's disease patients and descendants of people with Alzheimer's disease. Participants are submitted to a multitude of tests, including cognitive tests, DNA samples and MRI scans. Scans were annotated by a trained rater using the Brain Observer Microbleed Scale, generating a total of 88 scans. Images from both input datasets were re-sampled to be 1.0 mm isotropic, and normalised in the 0-1 range. Images were also skull stripped using FSL-bet. Scans with more than 10 microbleeds were excluded, as they are not representative of the elderly or typical Alzheimer's disease population. The average volume of all CMBs is 29 mm³.

Figure 2 presents the distribution of microbleed sizes in the dataset. All images used for this study contained at least one microbleed, which led to an average of 1.6 microbleeds per image. Data from different sources and labelled using different ratings were merged with the aim of increasing the robustness and generalisability of our networks.

4. RESULTS AND DISCUSSION

Figure 4 compiles the monitoring of all trained models. For each loss, the true positive rate (TP), average number of false positive per image (FP) and function loss value are shown across 500 000 iterations. Note that while the TP and FP scales are similar across losses, the loss value scales vary in range. In the remainder of this section, we will focus our analysis on three losses, which we believe are representative of the potential use of the proposed evaluation strategy.

The generalised Dice loss is bounded between 1.00 - 0.00, an average loss of 0.9 thus indicates a poor performance of the network. However, it performs relatively well on the training set, where it achieves about 80% TP rate. On the validation set, the performance drops significantly, to 50%. The difference in the generalised Dice loss between the training set and the validation set, is only of 0.07. A high false positive rate is a common problem in microbleed segmentation, but it is notably absent here. The exact segmentation provided by the network does not match the ground truth on a voxel-to-voxel basis, but for detection purposes the microbleeds are clearly marked. The voxelwise weighting of the Dice score causes the high loss score, giving a skewed view of the network's performance.

The cross-entropy loss is able to perform quite well, reaching a TPR of 90% on both the training and validation sets. The loss is showing a decline that is almost mirorring the the decline in FPs. The cross entropy is reporting a difference in training and validation performances that does not seem to be reflected in the lesion-wise analysis, suggesting that it is the voxel-wise comparison that is causing this difference.

The Dice and cross entropy loss performs well, also reaching a TPR of 90%, but a higher average of predicted false positives per image. The loss however, implies a stark difference in performance between training and validation sets, implying that the network is over-fitting. By using the lesion-wise analysis we can show that this is not the case, but that the model is able to generalise well.

As the quantity of correct lesions is the clinically relevant information, a straightforward approach would to be to design a loss function that takes this into account. Introducing a discrete number as the count into the loss functions can have adverse effect on training, as this can cause steep cliff in the gradient, impeding training even more. Regression based counting methods overcome this difficulty by approximating a density map, which allows for smoother gradients but also allows counting of partial objects. These methods are most effective when counting large items, and we have found they are unable to deal with the sparsity and small objects present in microbleed segmentation. Losses that are developed with data imbalance in mind, such as generalised Dice loss or Tversky loss, are able to increase the performance on data sets that have foreground vs. background imbalance, but are less powerful in tackling the imbalance of sizes within the classes itself. Both generalise Dice and Tversky losses are re-weightings of the Dice score, which is a measure of overlap. Smaller lesions contribute less to the overall overlap and are therefore more likely to be overlooked, but they should have an equal influence on performance. By using whole volume evaluation it is possible to closely monitor performance of the networks on a lesion-wise basis, giving a closer insight into the networks true performance than the loss function value itself.

5. CONCLUSION

Small object segmentation can be a challenging task due to the imbalance present in the data. Standard data imbalance approaches such as oversampling or specific loss functions alleviate the problem slightly, and can improve training performance. Measures like these have their own disadvantages, which are not always captured by the loss function.

We introduce lesion-wise whole volume validation, which allows for better performance monitoring of performance. We have highlighted the added value of a lesion-wise whole volume evaluation strategy, a tool that allows detailed performance measuring of tasks such as microbleed segmentation. It allows for a lesion-wise analysis of the networks performance on the entire dataset, and not only on a subset. The additional monitoring makes for better informed decisions regarding performance, convergence and over-fitting – all key information required to maximise the performance that a model can achieve.

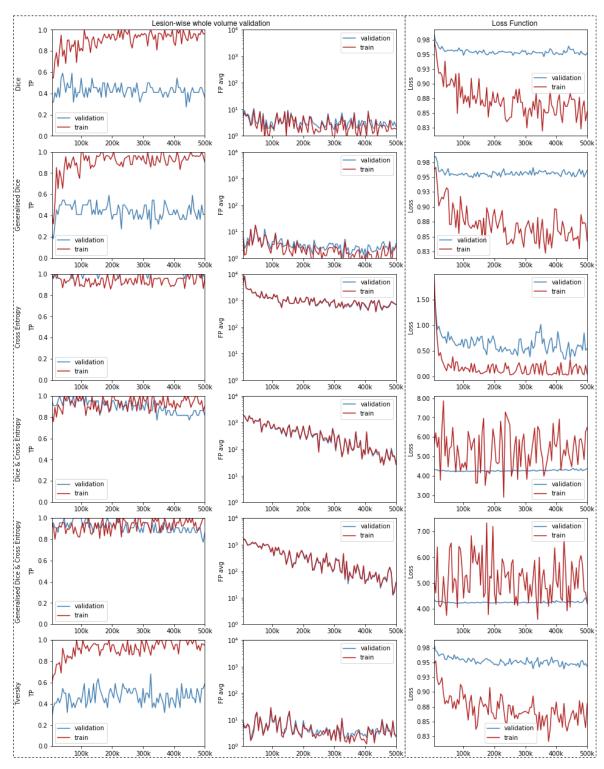


Figure 4: Evaluation measures monitored while training a network with seven different loss functions. From left to right, the true positive rate (TP), the average false positive number per image (FP) and the actual loss values are shown. The presented losses from top to bottom are Dice, generalised Dice, cross-entropy, combined Dice and cross-entropy, generalised Dice and cross-entropy and Tversky.

5.1 Limitations

Training a network using a loss function that does not adequately represent the end goal is not ideal, but using a lesion-wise whole volume evaluation strategy provides a better understanding of the network's performance. Unfortunately, whole volume lesion-wise measures cannot be used as a loss function. Using such ordinal measure will negatively impact the gradient used to optimise the network parameters. Using whole volume evaluation results in each iteration taking longer, as more data points need to be evaluated by reconstructing the original volume. Additionally, in the case of microbleed segmentation the lesion-wise evaluation is time consuming and computationally expensive. The microbleed segmentation network presented in this paper is in an early state of development, and is not yet at it is peak performance. The FPs rate per image is currently too high to allow its use in clinical studies or practices.

5.2 Future work

We have shown the effectiveness of whole volume evaluation purely as a monitoring tool, and we are eager to explore its application in a curriculum learning setting, where early stopping would be applied depending on performance A two-phase approach is seen in the state-of-the-art for object detection, where the first phase screens for potential candidates, and the second phase filters these candidates into true positives and false positives. For such a set up to work efficiently, a high sensitivity is required. Our current set-up yields a high number of false positives, but does also have a high true positive prediction rate. Therefore this set-up would be a good candidate for the initial candidate selection phase of such approaches. Using lesion-wise whole volume evaluation would clearly showcase the performance of both phases.

ACKNOWLEDGMENTS

The authors would like to thank the ADNI and ALFA consortia for providing access to their data. We gratefully acknowledge the support of NVIDIA Corporation with the donation of the P6000 GPU used for this research.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 721820.

REFERENCES

- [1] Sørensen, T., "A method of establishing groups of equal amplitude in plant sociology based on similarity of species and its application to analyses of the vegetation on Danish commons," *Biol. Skr.* 5, 1–34 (1948).
- [2] Sudre, C. H., Li, W., Vercauteren, T., Ourselin, S., and Jorge Cardoso, M., "Generalised Dice overlap as a deep learning loss function for highly unbalanced segmentations," in [arXiv], Cardoso, M. J., Arbel, T., Carneiro, G., Syeda-Mahmood, T., Tavares, J. M. R. S., Moradi, M., Bradley, A., Greenspan, H., Papa, J. P., Madabhushi, A., Nascimento, J. C., Cardoso, J. S., Belagiannis, V., and Lu, Z., eds., 240–248, Springer International Publishing, Cham (2017).
- [3] Greenberg, S. M., Eng, J. A., Ning, M. M., Smith, E. E., and Rosand, J., "Hemorrhage burden predicts recurrent intracerebral hemorrhage after lobar hemorrhage," *Stroke* **35**, 1415–1420 (jun 2004).
- [4] Wardlaw, J. M., Smith, E. E., Biessels, G. J., Cordonnier, C., Fazekas, F., Frayne, R., Lindley, R. I., T O'Brien, J., Barkhof, F., Benavente, O. R., and Others, "Neuroimaging standards for research into small vessel disease and its contribution to ageing and neurodegeneration," *The Lancet Neurology* 12(8), 822–838 (2013).
- [5] Cordonnier, C., Potter, G. M., Jackson, C. A., Doubal, F., Keir, S., Sudlow, C. L. M., Wardlaw, J. M., and Salman, R. A.-S., "Improving Interrater Agreement About Brain Microbleeds," *Stroke* **40**(1), 94–99 (2009).
- [6] Ayaz, M., Boikov, A. S., Haacke, E. M., Kido, D. K., and Kirsch, W. M., "Imaging cerebral microbleeds using susceptibility weighted imaging: One step toward detecting vascular dementia," *Journal of Magnetic Resonance Imaging* 31(1), 142–148 (2010).
- [7] Greenberg, S. M., Vernooij, M. W., Cordonnier, C., Viswanathan, A., Salman, R. A.-S., Warach, S., Launer, L. J., Buchem, M. A. V., and Breteler, M. M. B., "Cerebral microbleeds: a guide to detection and interpretation," *The Lancet Neurology* 8(2), 165–174 (2009).

- [8] Heinrich, K., Roth, A., and Zschech, P., "Everything counts: A taxonomy of deep learning approaches for object counting," in [27th European Conference on Information Systems Information Systems for a Sharing Society, ECIS 2019], (may 2020).
- [9] Sindagi, V. A. and Patel, V. M., "A Survey of Recent Advances in CNN-based Single Image Crowd Counting and Density Estimation," *Pattern Recognition Letters* **107**, 3–16 (jul 2017).
- [10] Dubey, R., Zhou, J., Wang, Y., Thompson, P. M., and Ye, J., "Analysis of sampling techniques for imbalanced data: An n=648 ADNI study," *NeuroImage* 87, 220–241 (2014).
- [11] Girshick, R., Donahue, J., Darrell, T., and Malik, J., "Rich feature hierarchies for accurate object detection and semantic segmentation," in [Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition], 580–587, IEEE Computer Society (sep 2014).
- [12] Sudre, C. H., Anson, B. G., Ingala, S., Lane, C. D., Jimenez, D., Haider, L., Varsavsky, T., Smith, L., Jäger, R. H., and Jorge Cardoso, M., "3D multirater RCNN for multimodal multiclass detection and characterisation of extremely small objects," arXiv (dec 2018).
- [13] Teramoto, A., Fujita, H., Yamamuro, O., and Tamaki, T., "Automated detection of pulmonary nodules in PET/CT images: Ensemble false-positive reduction using a convolutional neural network technique," *Medical Physics* 43, 2821–2827 (jun 2016).
- [14] Zhang, J., Xia, Y., Cui, H., and Zhang, Y., "Pulmonary nodule detection in medical images: A survey," (may 2018).
- [15] Dou, Q., Chen, H., Yu, L., Zhao, L., Qin, J., Wang, D., Mok, V. C., Shi, L., and Heng, P. A., "Automatic Detection of Cerebral Microbleeds from MR Images via 3D Convolutional Neural Networks," *IEEE Transactions on Medical Imaging* 35(5), 1182–1195 (2016).
- [16] Chen, G., Choi, W., Yu, X., Han, T., and Chandraker, M., "Learning efficient object detection models with knowledge distillation," (2017).
- [17] Ronneberger, O., Fischer, P., and Brox, T., "U-net: Convolutional networks for biomedical image segmentation," Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 9351, 234–241 (2015).
- [18] Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., and Others, "Pytorch: An imperative style, high-performance deep learning library," arXiv preprint arXiv:1912.01703 (2019).
- [19] Kerfoot, E., King, C. E., Ismail, T., Nordsletten, D., and Miller, R., "Estimation of Cardiac Valve Annuli Motion with Deep Learning," (oct 2020).
- [20] Mohseni Salehi, S. S., Erdogmus, D., and Gholipour, A., "Tversky loss function for image segmentation using 3D fully convolutional deep networks," in [arXiv], 379–387, Springer (2017).
- [21] Isensee, F., Petersen, J., Klein, A., Zimmerer, D., Jaeger, P. F., Kohl, S., Wasserthal, J., Koehler, G., Norajitra, T., Wirkert, S., and Maier-Hein, K. H., "nnU-Net: Self-adapting Framework for U-Net-Based Medical Image Segmentation," *Informatik aktuell*, 22 (sep 2018).
- [22] Gregoire, S. M., Chaudhary, U. J., Brown, M. M., Yousry, T. A., Kallis, C., Jäger, H. R., and Werring, D. J., "The Microbleed Anatomical Rating Scale (MARS): Reliability of a tool to map brain microbleeds," Neurology 73(21), 1759–1766 (2009).
- [23] Molinuevo, J. L., Gramunt, N., Gispert, J. D., Fauria, K., Esteller, M., Minguillon, C., Sánchez-Benavides, G., Huesa, G., Morán, S., Dal-Ré, R., and Camí, J., "The ALFA project: A research platform to identify early pathophysiological features of Alzheimer's disease," Alzheimer's and Dementia: Translational Research and Clinical Interventions 2, 82–92 (jun 2016).
- [24] Smith, S. M., "Fast robust automated brain extraction," Human Brain Mapping 17, 143–155 (nov 2002).