Weak labels for deep-learning-based detection of brain aneurysms from MR angiography scans

Tommaso Di Noto[1] tommaso.dinoto@chuv.ch
Guillaume Marie[1] guillaume.marie@chuv.ch
Sebastien Tourbier[1] sebastien.tourbier1@gmail.com
Yasser Alemán-Gómez[1] yasseraleman@gmail.com
Oscar Esteban[1] phd@oscaresteban.es
Guillaume Saliou[1] guillaume.saliou@chuv.ch
Meritxell Bach Cuadra[1,2] meritxell.bachcuadra@unil.ch
Patric Hagmann[1] patric.hagmann@chuv.ch
Jonas Richiardi[1] jonas.richiardi@chuv.ch
[1] *Radiology Department, Lausanne University Hospital and University of Lausanne*
[2] *CIBM: Center for Biomedical Imaging*

Editors: Under Review for MIDL 2022

Abstract

Unruptured Intracranial Aneurysms (UIAs) are focal dilatations in cerebral arteries. If overlooked, UIAs can rupture and lead to subarachnoid hemorrhages. Deep Learning (DL) models currently reach state-of-the-art performances for the automated detection of UIAs in Magnetic Resonance Angiography. However, there are still a few missing pieces to create robust DL models that can generalize across sites and be used during clinical practice. On one hand, the need for voxel-wise annotations from medical experts is hindering the creation of large datasets. On the other hand, multi-site validations are unfeasible since there exists to date only one open-access dataset. In this work, we summarize a full paper that we recently submitted to a journal and whose main contributions are the following: (a) a DL training approach that leverages fast-to-create weak labels and (b) the release of a second open-access dataset (the largest in the community) to foster model generalization.

Keywords: Weak Labels, Open Data, Aneurysm Detection, MR Angiography.

1. Introduction

Magnetic Resonance Angiography (MRA) is a routine imaging technique performed to detect UIAs. Since the visual inspection of MRA scans is time-consuming and small aneurysms can be non-trivial to spot, several automated DL methods (full list in Table 1, (Di Noto et al., 2021)) have been proposed to aid clinicians detecting UIAs in MRA scans. Although some of these models presented results on relatively large cohorts (115<N<1271), none of them made their dataset public. In addition, all these works rely on expensive, voxel-wise labels to drive the training process of their DL algorithms. Instead, in our work (Di Noto et al., 2021) we present a DL model that is trained using weak labels: these are oversized annotations that are less precise, but considerably faster to create for medical experts. Moreover, we openly release our in-house dataset on OpenNeuro. Together with the ADAM challenge dataset (Timmins et al., 2021), ours is the only open-access dataset.

2. Materials & Methods

In-House Dataset - We retrieved 284 MRA subjects: 156 had one (or more) UIAs, while 127 did not present any. A total of 198 aneurysms were annotated by one radiologist with 2 years of experience in neuroimaging, and double-checked by a senior neuroradiologist with over 15 years of experience. Two annotation schemes were followed: 1) **weak labels**: for most subjects (246/284), the radiologist used Mango to create the aforementioned weak labels which correspond to spheres that enclose the whole aneurysm (visual example in Figure 1). 2) **voxel-wise labels**: for the remaining subjects (38/284), the radiologist used ITK-SNAP to create voxel-wise labels drawn slice by slice scrolling in the axial plane.

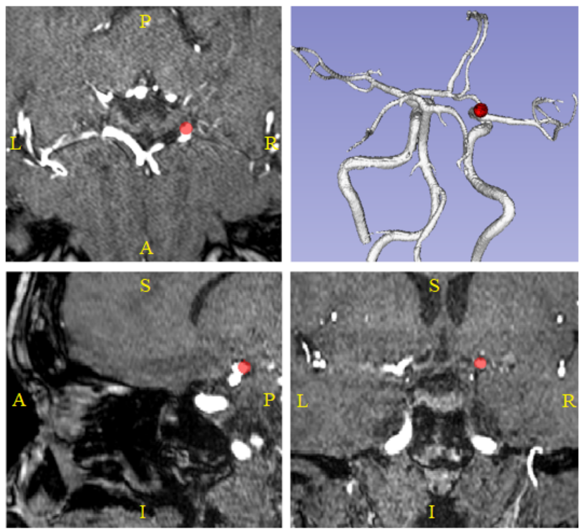

Figure 1: TOF-MRA orthogonal views of a 62-year-old female patient. Red areas correspond to our spherical weak labels.

Pre-processing - All volumes were skull-stripped. Second, we applied N4 bias-field correction. Third, we resampled all volumes to a median voxel spacing. Last, we registered a probabilistic vessel atlas to subject space to inject anatomical prior knowledge about the vessel locations (results related to the efficacy of this last step are not shown here).
Network - For detecting UIAs, we created a custom 3D UNET. We used 64^3-voxel MRA patches as input to our network. More technical details about the network can be found in (Di Noto et al., 2021). For patients, we extracted both negative (without aneurysm) and positive (with aneurysm) patches, while for controls we only extracted negative patches. At inference time, we adopted an anatomically-informed sliding-window approach (patch overlap = 50%). While not shown here, this network obtained competitive results in the ADAM challenge ranking 4th/18 in the open leaderboard.
Experiments - We wanted to answer two questions: (1) how much faster the creation of weak labels is with respect to voxel-wise labels? (2) what is the impact of using weak labels in terms of detection performances when comparing to voxel-wise labels? To answer (1), we selected 14 patients (mean aneurysm size = 5.2 mm), and we assessed the time difference between the two annotation schemes. To answer (2), we used the 38 subjects with voxel-

wise labels and for these we generated corresponding 'weakened' spherical labels (details in (Di Noto et al., 2021)). Then, to evaluate the influence of annotation quality in terms of detection performances, we conducted 3 experiments in which we used an increasing number of patients with voxel-wise labels: (i) all 38 patients with weakened labels, (ii) 19 patients with weakened labels and 19 with voxel-wise labels, and (iii) all 38 patients with voxel-wise labels. To compare the three configurations, we ran Wilcoxon signed-rank tests of the areas under the FROC curves (Chakraborty and Berbaum, 2004) across test subjects.

3. Results & Conclusion

Regarding the difference in annotation time, weak labels were four times faster to generate than their voxel-wise counterparts. Regarding the influence of label quality on detection results, we report in Table 1 the performances of our network when changing the ratio of voxel-wise/weakened labels for the 38 subjects described above. The configuration with all voxel-wise labels (Model 3) had higher sensitivity with respect to the other two (Model 1 and Model 2). However, this difference was not significant (W = 14.0, p = 0.054 when comparing to Model 1 and W = 685.5, p = 0.977 when comparing to Model 2).

In conclusion, we presented a DL framework which obtains competitive performances despite using weak labels. These labels are fast to create and thus facilitate the creation of large training datasets. In addition, we openly released our dataset to the community.

Table 1: Detection results on in-house dataset when changing the ratio of voxel-wise/weakened labels. Sensitivity reported as mean and 95% Wilson CI; FP = false positive.

Model Configuration	**Labels of 38 added subs**	**Avg. Sensitivity (CI)**	**Avg. FP rate**
Model 1	38 weakened	95/127 = 75% (65%, 80%)	1.3
Model 2	19 weakened, 19 voxel-wise	99/127 = 78% (68%, 82%)	0.9
Model 3	38 voxel-wise	101/127 = 80% (72%, 85%)	1.2

References

Dev P Chakraborty and Kevin S Berbaum. Observer studies involving detection and localization: modeling, analysis, and validation. *Medical physics*, 31(8):2313–2330, 2004.

Tommaso Di Noto, Guillaume Marie, Sebastien Tourbier, Yasser Alemán-Gómez, Oscar Esteban, Guillaume Saliou, Meritxell Bach Cuadra, Patric Hagmann, and Jonas Richiardi. Towards automated brain aneurysm detection in tof-mra: open data, weak labels, and anatomical knowledge. *arXiv preprint arXiv:2103.06168*, 2021.

Kimberley M Timmins, Irene C van der Schaaf, Edwin Bennink, Ynte M Ruigrok, Xingle An, Michael Baumgartner, Pascal Bourdon, Riccardo De Feo, Tommaso Di Noto, Florian Dubost, et al. Comparing methods of detecting and segmenting unruptured intracranial aneurysms on tof-mras: The adam challenge. *NeuroImage*, 238:118216, 2021.