Perimeter-based Loss Evaluation for Intracranial Hemorrhage Segmentation on Non-Contrast Head CT

Antoine Sanner[0000−0002−4917−9529] and Anirban Mukhopadhyay[0000−0003−0669−4018]

Technical University of Darmstadt, Karolinenpl. 5, 64289 Darmstadt, Germany
{firstname.lastname}@gris.tu-darmstadt.de

Abstract. Stroke remains one of the leading causes of permanent disability, with intracranial hemorrhage (ICH) having the highest mortality of all stroke types [1]. The volume of the hemorrhage is used in clinical diagnosis procedures, as it constitutes one of the most significant predictors [2]. As such, being able to accurately estimate this volume is crucial to allocate the level of care required by each patient. Currently, either a manual segmentation is done by a radiologist or the ABC/2 method [8] is used to solve this task. The former is highly time-consuming, while the latter is imprecise in its volume estimation. This calls for the development of a fully-automated solution, that is both fast and accurate. In this work, we show that the nnU-Net framework [4] can still reach near state-of-the-art results despite being not new. We also propose an evaluation of contour-based losses [6, 7].

Keywords: nnU-Net · Segmentation · Stroke

1 Introduction

Intracranial hemorrhage (ICH) has the highest mortality of all stroke subtypes while being common [1]. The volume of the hemorrhage is usually measured during the clinical routine, as it is one of the most significant predictors for the patient outcome [2] and allows allocating the level of care required by each patient.

Non-Contrast Computed Tomography (NCCT) is the most commonly used modality during the clinical routine for diagnosing ICH, as it is much faster than MR imaging. In general, the appearance of hemorrhage on NCCT is distinct because of the density of the blood [12]. Still, the segmentation of an ICH is challenging due to image noise and artifacts. The varying stages of clot formation also contribute to the heterogeneity of an ICH and make the boundaries more ambiguous.

A radiologist can delineate the ICH region by hand, but this task is time-consuming and prone to inter-rate variability. In contrast, the clinically-adopted ABC/2 method [8] is based on a simplification of the ellipsoid volume equation and computes the ICH volume from only a few measurements. While being faster, this method is still prone to inter-rater variability. It was also shown that it overestimates the ICH volume by circa 20% [14]. Regardless, the ABC/2 method, is still commonly used for its practicality. This calls for the development of an automated segmentation method that is both fast and accurate.

In this work, we show that the *nnU-Net* framework [4] can still reach near state-of-the-art results. We also propose an evaluation of contour-based losses [6, 7].

2 Related Work

In the literature, efforts to improve the segmentation results are either focused on adapting the network architecture to the specific task or on using specialized losses to capture some specific shape characteristics.

2.1 Network Architectures

Existing methods for segmenting ICH on NCCT imaging are mostly CNN-based. The first archetype uses 2D convolutions combined with another mechanism to capture the correlation between 2D slices. SLEX-Net [9] computes a segmentation for a 2D slice and its two neighboring slices. A Slice expansion Module is then used to combine these three segmentation masks to produce a final prediction for the central slice, as well as an uncertainty mask. ICHNet [5] uses a hypercolumn feature map for each 2D slice to produce a segmentation. The final 3D prediction is then post-processed with a 3D Conditional Random Field. In contrast, [12, 13] both simply use a 3D U-Net to capture the correlation between slices.

2.2 Segmentation Losses

While using a combination of a cross-entropy and Dice loss has become standard for segmentation [4], a variety of losses for segmentation exists, with new ones being developed each year [11]. Each of these new losses promise to improve the segmentation performance by capturing task-specific characteristics. For instance, the Hausdorff-distance [7] and contour [6] losses allow the network to also learn border irregularities within the target shapes and to not only focus on the overall volume.

3 Method

Our base architecture is the well-known nnU-Net as provided by [4]. We integrated both the Hausdorff-distance loss as proposed by [7] and the contour loss proposed by [6]. While the former estimates the Hausdorff distance, the latter extracts the contour of both the prediction, and the ground truth and minimizes the mean square error between them. The contour maps are produced by non-trainable max- and min-pooling layers.

In practice, we minimize the following loss:

$$\mathcal{L} = \lambda_{Dice}\mathcal{L}_{Dice} + \lambda_{CE}\mathcal{L}_{CE} + \lambda_{Surface}\mathcal{L}_{Surface} \tag{1}$$

where $\mathcal{L}_{Dice}$ is the Dice loss, $\mathcal{L}_{CE}$ is the cross-entropy loss and $\mathcal{L}_{Surface}$ is either the Hausdorff-distance loss or the contour loss depending on the experiment.

4 Experimental Setup

In this section, we briefly present the dataset as well as the implementation details.

4.1 Dataset

As part of the *INSTANCE2022* challenge [10], 200 non-contrast head CT volumes of clinically diagnosed patients with ICH. The data were labeled by 10 radiologists with more than 5 years clinical experience. The size of a CT volume is $512 \times 512 \times N$, with $N \in [20, 70]$. The pixel spacing of a CT volume is 0.42 mm $\times$ 0.42 mm $\times$ 5 mm. The labels are binary, with 0 for the background and 1 for ICH. The dataset was finally split in 100 training, 30 open validation and 70 test cases.

4.2 Implementation Details

We used nnU-Net as a framework[1] and we integrated the losses of [6] as provided[2]. In particular, the Hausdorff-distance loss is based on the euclidean distance transform.

Rather than using the standard z-normalization of nnU-Net for input images, we chose to clip the intensity values to $[0 - 100]$ as [3] has shown the importance of restricting the value range to a meaningful one. We empirically chose $\lambda_{Dice} = \lambda_{CE} = 1$ and $\lambda_{Surface} = 10^{-3}$.

Otherwise, the behavior of the nnU-Net framework has remained unchanged with regard to the training pipeline. For each method, a five-fold cross-validation was used to train five models for a fixed number of epochs. All models for each setup are then ensembled to make the final prediction. Finally, the standard nnU-Net data augmentation scheme was used for 2D models, while the "insane_DA" scheme was used for 3D models.

[1] https://github.com/MIC-DKFZ/nnU-Net
[2] https://github.com/rosanajurdi/Prior-based-Losses-for-Medical-Image-Segmentation

5 Results

Here, we first evaluate the benefit of ensembling 2D and 3D U-Nets trained with the traditional Dice + cross-entropy loss. Then we compare, the results for 3D U-Nets trained with the Hausdorff-distance [7] and contour [6] losses. We use the official metrics for the *INSTANCE2022* challenge, i.e. Dice, relative volume difference, Hausdorff distance and surface Dice. These results are computed for the 30 open validation cases.

5.1 Ensembling U-Nets

Method	Dice	Rel. Vol. Diff.	Hausdorff Dist.	Surf. Dice
2D U-Net	71.0 ± 30.0	29.3 ± 30.3	*null**	49.6 ± 20.9
3D U-Net	**72.8 ± 28.3**	**25.5 ± 27.4**	**2973.4 ± 3102.7**	**51.5 ± 19.2**
2D+3D Ensemble	71.6 ± 30.1	28.9 ± 30.1	*null**	50.6 ± 21.3

Table 1. Mean ± Std Dice, relative volume difference, Hausdorff distance and surface Dice for various U-Net models trained on the open validation data using the Dice + cross-entropy loss. A *null** Hausdorff distance means that the metric could not be computed for certain cases.

Here, we compare the performance of a 3D U-Net vs 2D U-Net vs ensembling 2D+3D. As can be seen in Table 5.1, the best results across metrics were obtained using the 3D U-Net. While the 2D U-Net is closely behind, ensembling the two architectures does not yield better results. The high variance in Dice across cases is caused by some cases having much smaller ICH volumes. For these, the models barely activate (3D U-Net) or even do not (2D U-Net), preventing the computation of the Hausdorff distance in the latter case.

The 3D U-Net outperforming the 2D one is likely due to the ambiguity of some annotations: some areas in a given slice may only be labeled as ICH because a hemorrhage was clearly found nearby in other slices. This context may help the model better distinguish between the two.

5.2 Perimeter-based Losses Evaluation

With these experiments, we evaluate the impact of using perimeter-based losses in combination with a 3-D U-Net. As can be seen in Table 5.2, the contour-loss does not yield improvements with regard to either volume-based or surface-based metrics. The results for the Hausdorff-distance loss were not included due to time constraints. Indeed during our testing, using a non-zero weight for this loss caused catastrophic forgetting, with all models only reaching $2-5\%$ Dice on training data. Tuning this weight is especially time-consuming as the implementation of this loss heavily relies on the CPU and multiplied by a factor $\times 4$ the training time.

Method	Dice	Rel. Vol. Diff.	Hausdorff Dist.	Surf. Dice
$\mathcal{L}_{Dice+CE}$	**72.8 ± 28.3**	**25.5 ± 27.4**	**2973.4 ± 3102.7**	**51.5 ± 19.2**
$\mathcal{L}_{Dice+CE+contour}$	69.1 ± 29.5	25.1 ± 23.8	4056.8 ± 3940.7	48.3 ± 19.9

Table 2. Mean ± Std Dice, relative volume difference, Hausdorff distance and surface Dice for 3D U-Nets on the open validation data trained using different losses.

6 Discussion

In this work, we show that despite not being new, the nnU-Net framework [4] can still reach near state-of-the-art results on the *INSTANCE2022* challenge [10]. We also propose an evaluation of contour-based losses [6, 7]. We find that the 3D U-Net outperforms the 2D variant, most likely due to having more context across slices to better label ambiguous areas. Both the Hausdorff-distance [7] and the contour loss [6] did not yield satisfying results and may require some further tuning. This is especially time-consuming for the former, as its implementation heavily relies on the CPU and slows down the training significantly. In the future, we would like to further investigate the weight scheduling for contour-based losses. We also want to work on a faster GPU-based implementation of the Hausdorff-distance loss for 3D segmentation.

References

1. van Asch, C.J., Luitse, M.J., Rinkel, G.J., van der Tweel, I., Algra, A., Klijn, C.J.: Incidence, case fatality, and functional outcome of intracerebral haemorrhage over time, according to age, sex, and ethnic origin: a systematic review and meta-analysis **9**(2), 167–176. https://doi.org/10.1016/S1474-4422(09)70340-0, https://linkinghub.elsevier.com/retrieve/pii/S1474442209703400
2. Broderick, J.P., Brott, T.G., Duldner, J.E., Tomsick, T., Huster, G.: Volume of intracerebral hemorrhage. a powerful and easy-to-use predictor of 30-day mortality. **24**(7), 987–993. https://doi.org/10.1161/01.STR.24.7.987, https://www.ahajournals.org/doi/10.1161/01.STR.24.7.987
3. Ertl, A., Franz, A., Schmitz, B., Braun, M.: 3d CNN-based identification of hyperdensities in cranial non-contrast CT after thrombectomy. In: Maier-Hein, K., Deserno, T.M., Handels, H., Maier, A., Palm, C., Tolxdorff, T. (eds.) Bildverarbeitung für die Medizin 2022, pp. 309–314. Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-36932-3_64, https://link.springer.com/10.1007/978-3-658-36932-3_64, series Title: Informatik aktuell
4. Isensee, F., Jaeger, P.F., Kohl, S.A.A., Petersen, J., Maier-Hein, K.H.: nnU-net: a self-configuring method for deep learning-based biomedical image segmentation **18**(2), 203–211. https://doi.org/10.1038/s41592-020-01008-z, http://www.nature.com/articles/s41592-020-01008-z
5. Islam, M., Sanghani, P., See, A.A.Q., James, M.L., King, N.K.K., Ren, H.: ICH-Net: Intracerebral hemorrhage (ICH) segmentation using deep learning. In: Crimi, A., Bakas, S., Kuijf, H., Keyvan, F., Reyes, M., van Walsum, T. (eds.) Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries, vol. 11383,

pp. 456–463. Springer International Publishing. https://doi.org/10.1007/978-3-030-11723-8_46, http://link.springer.com/10.1007/978-3-030-11723-8_46, series Title: Lecture Notes in Computer Science
6. Jurdi, R.E., Petitjean, C., Honeine, P., Cheplygina, V., Abdallah, F.: A surprisingly effective perimeter-based loss for medical image segmentation. In: Medical Imaging with Deep Learning (2021), https://openreview.net/forum?id=NDEmtyb4cXu
7. Karimi, D., Salcudean, S.E.: Reducing the hausdorff distance in medical image segmentation with convolutional neural networks. IEEE Transactions on Medical Imaging **39**(2), 499–513 (2020). https://doi.org/10.1109/TMI.2019.2930068
8. Kothari, R.U., Brott, T., Broderick, J.P., Barsan, W.G., Sauerbeck, L.R., Zuccarello, M., Khoury, J.: The ABCs of measuring intracerebral hemorrhage volumes **27**(8), 1304–1305. https://doi.org/10.1161/01.STR.27.8.1304, https://www.ahajournals.org/doi/10.1161/01.STR.27.8.1304
9. Li, X., Luo, G., Wang, W., Wang, K., Gao, Y., Li, S.: Hematoma expansion context guided intracranial hemorrhage segmentation and uncertainty estimation **26**(3), 1140–1151. https://doi.org/10.1109/JBHI.2021.3103850, https://ieeexplore.ieee.org/document/9511297/
10. Li, X., Wang, K., Liu, J., Wang, H., Xu, M., Liang, X.: The 2022 intracranial hemorrhage segmentation challenge on non-contrast head CT (NCCT) . https://doi.org/10.5281/ZENODO.6362221, https://zenodo.org/record/6362221, publisher: Zenodo
11. Ma, J., Chen, J., Ng, M., Huang, R., Li, Y., Li, C., Yang, X., Martel, A.L.: Loss odyssey in medical image segmentation. Medical Image Analysis **71**, 102035 (2021). https://doi.org/https://doi.org/10.1016/j.media.2021.102035, https://www.sciencedirect.com/science/article/pii/S1361841521000815
12. Patel, A., Schreuder, F.H.B.M., Klijn, C.J.M., Prokop, M., Ginneken, B.v., Marquering, H.A., Roos, Y.B.W.E.M., Baharoglu, M.I., Meijer, F.J.A., Manniesing, R.: Intracerebral haemorrhage segmentation in non-contrast CT **9**(1), 17858. https://doi.org/10.1038/s41598-019-54491-6, http://www.nature.com/articles/s41598-019-54491-6
13. Schmitt, N., Mokli, Y., Weyland, C.S., Gerry, S., Herweh, C., Ringleb, P.A., Nagel, S.: Automated detection and segmentation of intracranial hemorrhage suspect hyperdensities in non-contrast-enhanced CT scans of acute stroke patients **32**(4), 2246–2254. https://doi.org/10.1007/s00330-021-08352-4, https://doi.org/10.1007/s00330-021-08352-4
14. Webb, A.J., Ullman, N.L., Morgan, T.C., Muschelli, J., Kornbluth, J., Awad, I.A., Mayo, S., Rosenblum, M., Ziai, W., Zuccarrello, M., Aldrich, F., John, S., Harnof, S., Lopez, G., Broaddus, W.C., Wijman, C., Vespa, P., Bullock, R., Haines, S.J., Cruz-Flores, S., Tuhrim, S., Hill, M.D., Narayan, R., Hanley, D.F.: Accuracy of the ABC/2 score for intracerebral hemorrhage: Systematic review and analysis of MISTIE, CLEAR-IVH, CLEAR III **46**(9), 2470–2476. https://doi.org/10.1161/STROKEAHA.114.007343, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4550520/