Intracranial Hemorrhage Segmentation using E3-equivariant neural networks

Mario Geiger[1], Ivan Diaz[2], Sebastián Otalora Montenegro[2], and Richard McKinley[2]

[1] Department of Research Laboratory of Electronics, Massachusetts Institute of Technology, Boston MA
[2] Support Center for Advanced Neuroimaging (SCAN), University Institute of Diagnostic and Interventional Neuroradiology, University of Bern, Inselspital, Bern University Hospital, Bern, Switzerland

Abstract. This short paper describes the network architecture, preprocessing and data augmentation used by the SCAN team as part of the 2022 Intracranial Hemorrhage Segmentation Challenge on Non-Contrast head CT, to accompany the docker image submission.

1 Network Description

1.1 E3-Equivariant Neural Networks

The network makes use of the jax version of the e3nn library [2,3,1] which enables the creation neural networks equivariant to translations, rotations, and mirroring. This library implements a rotationally-equivariant voxel convolution described in [8]. This voxel convolution has the advantage of being definable independently of the underlying voxel spacing, as a product of spherical harmonics and a radial basis function in physical space. This allows the network to be trained and applied directly on different input voxel dimensions. Since the INSTANCE 2022 dataset [5,4] has a wide range of voxel spacings, we anticipate that this flexibility will be particularly valuable. The e3nn framework allows feature maps valued in a variety of irreducible representations (irreps): datatypes which transform linearly under rotation and parity shift. While the input and output to the network are respectively 3 even scalar volumes (Table 1) and one scalar volume (corresponding to ordinary scalar-valued data), the intermediate features take the form of vectors, pseudovectors, and higher-ranked tensors. The code is available on github.

1.2 UNet architecture

The networks follows the classic U-Net architecture [7], in which the convolution kernels in the original architecture are replaced by a 3D e3nn voxel convolution of diameter 5 mm. We used three 2x2x2 downsampling operations which halve the resolution in the encoding path (depending on the resolution of the image), and three corresponding trilinear upsampling operations on the decoding path (that maps back to the same resolution). The size of the voxel convolution diameter is doubled with every downsample and halved with every upsampling operation. A Gaussian error linear unit activation function and instance normalization was used after each convolution. The hidden irreducible representations (corresponding to features) consisted of 20 odd and 20 even scalars, 10 odd and 10 even vectors, and 5 odd and 5 even rank-2 tensor. These were doubled after each downsampling step and halved at each upsampling step.

2 Training

2.1 Pre-processing. Data Augmentation, and Training

For training, each CT volume was windowed to three different Hounsfield unit value ranges, scaled, and added to a separate channel which served as the model input. The values are in the following table:

Table 1. Channels used.

Channel	Range [HU]
0	0 - 1000
1	0 - 80
2	-50 - 220

Since all operations in the network are equivariant with respect to 3D rotations, we did not perform any rotation-based augmentation. To increase the variety in the data a random diffeomorphic deformation was performed on each training sample: the sampling of diffeomorphisms is detailed in [6]. We trained the networks for 200k steps of adam with a learning rate 10^{-3}. We used a batch size of 1 and no regularization. We performed weight averaging with $\epsilon = 5\ 10^{-5}$, $\tilde{w}_{t+1} = (1-\epsilon)\tilde{w}_t + \epsilon w_t$ and use $\tilde{w}_t$ for evaluation. Since a new model was computed for every training sampling, a batch size of 1 was used. We used the Adam optimizer with a learning rate of 1e-3. Eight models were trained, each on 80 randomly sampled subsets from the training dataset. The loss function employed was Cross Entropy Loss.

The final prediction was performed by applying each of the eight models to patches of size 144x144x13 with padding discarding 22x22x2 pixels on each side, a sliding window with an overlap of 26 pixels and gaussian weighing, and then averaging the model outputs.

For the final prediction we take the ensemble average of the eight models.

References

1. Diaz, I., Geiger, M., Richard, M.: e3nn_unet (2022)

2. Geiger, M., Smidt, T.: e3nn: Euclidean neural networks (2022). https://doi.org/10.48550/ARXIV.2207.09453, `https://arxiv.org/abs/2207.09453`
3. Geiger, M., Smidt, T., M., A., Miller, B.K., Boomsma, W., Dice, B., Lapchevskyi, K., Weiler, M., Tyszkiewicz, M., Batzner, S., Uhrin, M., Frellsen, J., Jung, N., Sanborn, S., Rackers, J., Bailey, M.: Euclidean neural networks: e3nn (2020). https://doi.org/10.5281/zenodo.5292912, `https://doi.org/10.5281/zenodo.5292912`
4. Li, X., Luo, G., Wang, W., Wang, K., Gao, Y., Li, S.: Hematoma expansion context guided intracranial hemorrhage segmentation and uncertainty estimation. IEEE Journal of Biomedical and Health Informatics **26**(3), 1140–1151 (2021)
5. Li, X., Wang, K., Liu, J., Wang, H., Xu, M., Liang, X.: The 2022 intracranial hemorrhage segmentation challenge on non-contrast head ct (ncct) (2022)
6. Petrini, L., Favero, A., Geiger, M., Wyart, M.: Relative stability toward diffeomorphisms indicates performance in deep nets. Advances in Neural Information Processing Systems **34**, 8727–8739 (2021)
7. 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. pp. 234–241. Springer (2015)
8. Weiler, M., Geiger, M., Welling, M., Boomsma, W., Cohen, T.S.: 3D Steerable CNNs: Learning Rotationally Equivariant Features in Volumetric Data. neurips 2018 p. 12 (2018)