Supplementary Materials
Self-PU: Self Boosted and Calibrated Positive-Unlabeled Training

Xuxi Chen* [1,2] **Wuyang Chen*** [1] **Tianlong Chen** [1] **Ye Yuan** [1] **Chen Gong** [3] **Kewei Chen** [4] **Zhangyang Wang** [1]

1. Network Structures

1.1. Network Structure for the MNIST dataset

For images from the MNIST dataset, we follow the same setting as in (Kiryo et al., 2017), where we first flatten the image into a one-dimensional vector and forward the vector through six fully-connected layers. The final output ($P(\boldsymbol{x}|Y = +1)$) is produced by a Sigmoid function.

1.2. Network Structure for the CIFAR10 dataset

For images from the CIFAR10 dataset, we also follow the same setting as in (Kiryo et al., 2017), where we first forward the image through 13 convolutional layers. After that, we flatten the feature map into a one-dimensional vector and forward the vector through three fully-connected layers. The final output ($P(\boldsymbol{x}|Y = +1)$) is produced by a Sigmoid function.

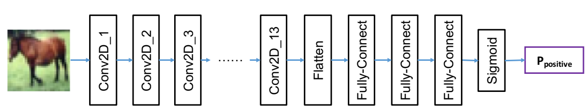

Figure 1. Network for images in CIFAR10 dataset. Each convolutional and fully-connected layer is followed by a batch normalization layer and a ReLU activation layer, which are omitted for simplicity.

1.3. Network Structure of Multi-scale Network for MRI Images in the ADNI dataset

To effectively learn from the MRI images, we design a special multi-scale network. As pointed out in (Khvostikov et al., 2018), the left and right hippocampus regions in the human brain are of high relevance to Alzheimer's Disease, and the remaining regions are of less relevance. Therefore in our multi-scale networks, we crop local regions of the left and right hippocampus as two $50 \times 50 \times 50$ inputs into two independent local branches. In addition, to reduce the noise that may reside in out-hippocampus regions, we downsample the global MRI image also to a size of $50 \times 50 \times 50$ and feed into the global branch. After two 3D convolutional layers, we adopt a "late fusion" strategy for three branches: three feature maps are flattened into one-dimensional vectors and are concatenated together with four demographic scalars (age, gender, education, and APOE gene type). Two fully-connected layers are then applied, and the final output ($P(\boldsymbol{x}|Y = +1)$) is produced by a Sigmoid function.

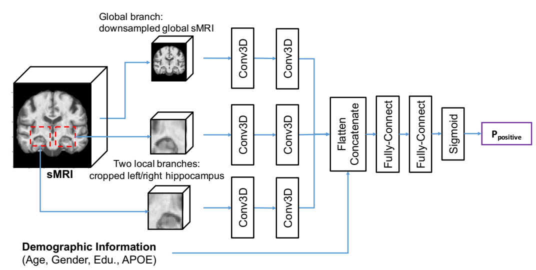

Figure 2. Multi-scale networks for MRI images in the ADNI dataset. Each convolutional and fully-connected layer is followed by a batch normalization layer and a ReLU activation layer, which are omitted for simplicity.

*Equal contribution [1]Texas A&M University [2]University of Science and Technology of China [3]Nanjing University of Science and Technology [4]Banner Alzheimer's Institute. Correspondence to: Zhangyang Wang <atlaswang@tamu.edu>.

Proceedings of the 37th International Conference on Machine Learning, Online, PMLR 119, 2020.

References

Khvostikov, A., Aderghal, K., Benois-Pineau, J., Krylov, A., and Catheline, G. 3d cnn-based classification using smri and md-dti images for alzheimer disease studies. *arXiv preprint arXiv:1801.05968*, 2018.

Kiryo, R., Niu, G., du Plessis, M. C., and Sugiyama, M. Positive-unlabeled learning with non-negative risk estimator. In *NeurIPS*, pp. 1675–1685, 2017.