HD-UNet: Hyper-dense UNet with asymmetric convolutions for multi-modal intervertebral disc segmentation

Jose Dolz, Christian Desrosiers, and Ismail Ben Ayed

École de technologie Superieure (ETS), Montreal, QC, Canada
jose.dolz@livia.etsmtl.ca

1 Methodology

Deep models have been largely employed recently in a broad span of medical applications such as detection, classification or segmentation [2]. Among all the proposed models, UNet [3] was a breakthrough in the medical field, leading to a huge amount of works based on this network. Recent research on multi-modal image segmentation has also demonstrated that the manner in which low-level features from multiple modalities are fused plays an important role on the performance of a learning model [1]. Indeed, in the context of multi-modal learning, it is difficult to discover highly non-linear relationships between the low-level features of different modalities, more so when such modalities have significantly different statistical properties. To deal with this, and unlike most approaches that adopt an early fusion strategy –i.e., multiple modalities are concatenated at the input of the network–, we followed the strategy presented in [1] and adapt it on a UNet-alike architecture. Further, we extended the Inception modules [4] by including two convolutional blocks with dilated convolutions of different scale, which help to capture different context. As in the inception module, all the convolutional blocks are replaced by asymmetric convolutions, which increase the representation ability of the model by adding more non-linearities without the need of increasing its complexity (i.e., number of learnable parameters) [4]. Figure 1 illustrates the proposed module.

2 Experiments

Out of the 16 multi-modal volumes provided, 13 were used for training and the remaining 3 for validating our model. Our network was trained on a NVidia TITAN XP GPU with 12 GBs RAM, using ADAM as optimizer and a learning rate of 1×10^{-5} during 200 epochs. Training took around 7 hours and segmentation of a whole volume took 3-4 seconds, as average.

The proposed architecture was compared to three baselines. First, an architecture similar to UNet, where all the convolutional blocks have been replaced by the module in 1,*letf*. In this model, all the image modalities have been fused at the input of the network (*Baseline_EarlyFusion*). The second architecture is

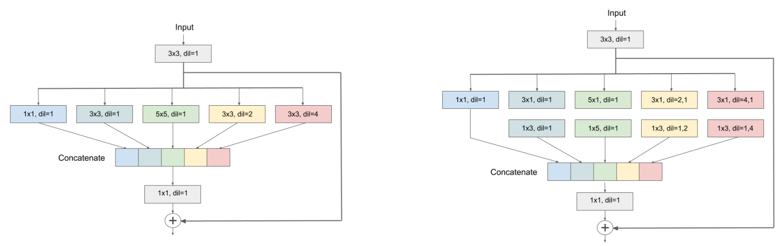

Fig. 1. Proposed module which extends inception modules. The module on the left employs standard convolutions while the module on the right adopts the idea of asymmetric convolutions [4].

similar to the first one, except by the strategy employed to fuse the different image modalities. Instead of being concatenated at the beginning each modality is processed independently in separated paths, and the features learned are fused before the decoding path (*Baseline_LateFusion*). Third, the multi-path architecture is extended by adding dense connections between the same and different paths, similar to [1] (*HD-UNet*). And last, the module including asymmetric convolutions (Fig. 1,*right*) is adopted in *HD-UNet*, obtaining the proposed model, referred to as *HD-UNet asym*. Results of the comparison are reported in Table 1.

Table 1. Results on the validation subjects for the different architectures.

Architecture	DSC
Baseline_EarlyFusion	0.8981
Baseline_LateFusion	0.9086
HD-UNet	0.9162
HD-UNet asym	**0.9191**

References

1. Dolz, J., Gopinath, K., Yuan, J., Lombaert, H., Desrosiers, C., Ben Ayed, I.: Hyperdense-Net: A hyper-densely connected CNN for multi-modal image segmentation. arXiv preprint arXiv:1804.02967 (2018)
2. Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., van der Laak, J.A., Van Ginneken, B., Sánchez, C.I.: A survey on deep learning in medical image analysis. Medical image analysis **42**, 60–88 (2017)
3. 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)
4. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z.: Rethinking the inception architecture for computer vision. In: CVPR. pp. 2818–2826 (2016)