

Dense Multi-path U-Net for Ischemic Stroke Lesion Segmentation in Multiple Image Modalities

Jose Dolz¹, Ismail Ben Ayed¹, Christian Desrosiers¹

Laboratory of Imaging, Vision and Artificial Intelligence
Ecole de technologie supérieure, Montreal, Canada
`jose.dolz@livia.etsmtl.ca`

Delineating infarcted tissue in ischemic stroke lesions is crucial to determine the extent of damage and optimal treatment for this life-threatening condition. However, this problem remains challenging due to high variability of ischemic strokes' location and shape. Recently, fully-convolutional neural networks (CNN), in particular those based on U-Net [27], have led to improved performances for this task [7]. In this work, we propose a novel architecture that improves standard U-Net based methods in three important ways. First, instead of combining the available image modalities at the input, each of them is processed in a different path to better exploit their unique information. Moreover, the network is densely-connected (i.e., each layer is connected to all following layers), both within each path and across different paths, similar to HyperDenseNet [11]. This gives our model the freedom to learn the scale at which modalities should be processed and combined. Finally, inspired by the Inception architecture [32], we improve standard U-Net modules by extending inception modules with two convolutional blocks with dilated convolutions of different scale. This helps handling the variability in lesion sizes. We split the 93 stroke datasets into training and validation sets containing 83 and 9 examples respectively. Our network was trained on a NVidia TITAN XP GPU with 16 GBs RAM, using ADAM as optimizer and a learning rate of 1×10^{-5} during 200 epochs. Training took around 5 hours and segmentation of a whole volume took between 0.2 and 2 seconds, as average. The performance on the test set obtained by our method is compared to several baselines, to demonstrate the effectiveness of our architecture, and to a state-of-art architecture that employs factorized dilated convolutions, i.e., ERFNet [26].

1 Introduction

Stroke is one the leading causes of global death, with an estimate of 6 million cases each year [19,28]. It is also a major cause of long-term disability, resulting in reduced motor control, sensory or emotional disturbances, difficulty understanding language, and memory deficit. Cerebral ischemia, which comes from the blockage of blood vessels in the brain, represents approximately 80% of all stroke cases [31,13]. Brain imaging methods based on Computed Tomography

(CT) and Magnetic Resonance Imaging (MRI) are typically employed to evaluate stroke patients [34]. Early-stage ischemic strokes appear as a hypodense regions in CT, making them hard to locate with this modality. MRI sequences, such as T1 weighted, T2 weighted, fluid-attenuated inversion recovery (FLAIR), and diffusion-weighted imaging (DWI), provide a clearer image of brain tissues than CT, and are preferred modalities to assess the location and evolution of ischemic stroke lesions [3,2,18].

The precise delineation of stroke lesions is critical to determine the extend of tissue damage and its impact on cognitive function. However, manual segmentation of lesions in multi-modal MRI data is time-consuming as well as prone to inter and intra-observer variability. Developing methods for the automatic segmentation can thus contribute to having more efficient and reliable tools to quantify stroke lesions over time [24]. Over the years, various semi-automated and automated techniques have been proposed for segmenting lesions [25,21]. Recently, deep convolutional neural networks (CNNs) have shown high performance for this task, outperforming standard segmentation approaches on benchmark datasets [20,35,14,4,17].

Multi-modal image segmentation based on CNNs is typically addressed with an *early fusion* strategy, where multiple modalities are merged from the original input space of low-level features [39,22,16,17,10,33]. This strategy assumes a simple relationship (e.g., linear) between different modalities, which may not correspond to reality [30]. For instance, the method in [39] learns complementary information from T1, T2 and FA images, however the relationship between these images may be more complex due to the different image acquisition processes. To better account for this complexity, Nie et al. [23] proposed a *late fusion* approach, where each modality is processed by an independent CNN whose outputs were fused in deep layers. The authors showed this strategy to outperform early fusion on the task of infant brain segmentation.

More recently, Aigün et al. explored different ways of combining multiple modalities [1]. In this work, all modalities are considered as separate inputs to different CNNs, which are later fused at an ‘early’, ‘middle’ or ‘late’ point. Although it was found that ‘late’ fusion provides better performance, as in [23], this method relies on a single-layer fusion to model the relation between all modalities. Nevertheless, as demonstrated in several works [30], relations between different modalities may be highly complex and they cannot easily be modeled by a single layer. To account for the non-linearity in multi-modal data modeling, we recently proposed a CNN that incorporates dense connections not only between the pairs of layers within the same path, but also between those across different paths [9,11]. This architecture, known as *HyperDenseNet*, obtained very competitive performance in the context of infant and adult brain tissue segmentation with multiple MRI data.

Despite the remarkable performance of existing methods, the combination of multi-modal data at various levels of abstraction has not been fully exploited for the segmentation of ischemic stroke lesions. In this paper, we adopt the strategy presented in [9,11] and propose a multi-path architecture, where each modality

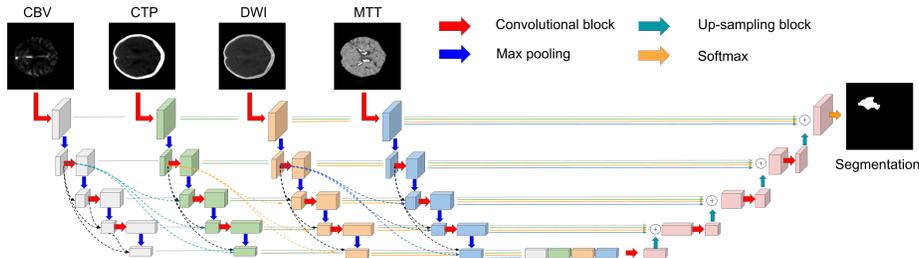


Fig. 1. Proposed architecture multi-path dense UNet. Dotted lines represent some of the dense connectivity patterns adopted in this extended version of UNet.

is employed as input of one stream and dense connectivity is used between layers in the same and different paths. Furthermore, we also extend the standard convolutional module of InceptionNet [32] by including two additional dilated convolutional blocks, which may help to learn larger context. Experiments on 103 ischemic stroke lesion multi-modal scans from the Ischemic Stroke Lesion Segmentation (ISLES) Challenge 2018 shows our model to outperform architectures based on early and late fusion, as well as state-of-art segmentation networks.

2 Methodology

The proposed models build upon the UNet architecture [27], which has shown outstanding performance in various medical segmentation tasks [8,12]. This network consists of a contracting and expanding path, the former collapsing an image down into a set of high level features and the latter using these features to construct a pixel-wise segmentation mask. Using skip connections, outputs from early layers are concatenated to the input of subsequent layers with the objective of transferring information that may be lost in the encoding path.

2.1 Proposed multi-modal UNet

Disentangling input data. Figure 1 depicts our proposed network for ischemic stroke lesion segmentation in multiple image modalities. Unlike most UNet-like architectures, the encoding path is split into N streams, which serve as input to each image modality. The main objective of processing each modality in separated streams is to disentangle information that otherwise would be fused from an early stage, with the drawbacks introduced before, i.e., limitation to capture complex relationships between modalities.

Hyper-Dense connectivity. Inspired by the recent success of densely and hyper-densely connected networks in medical image segmentation works [37,5,9,11], we propose to extend UNet to accommodate hyper-dense connections within the

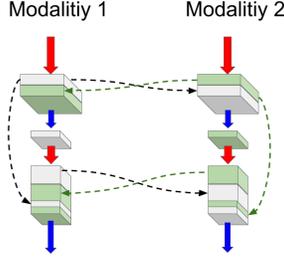


Fig. 2. Detailed version of a section of the proposed dense connectivity in multi-modal scenarios. For simplicity two image modalities are considered in this example.

same and between multiple paths. In addition to better modeling relationships between different modalities, employing dense connections also brings the three following benefits [15]. First, direct connections between all layers help improving the flow of information and gradients through the entire network, alleviating the problem of vanishing gradient. Second, short paths to all the feature maps in the architecture introduce an implicit deep supervision. And third, dense connections have a regularizing effect, which reduces the risk of over-fitting on tasks with smaller training sets.

In standard CNNs, the output of the l^{th} layer, denoted as \mathbf{x}_l , is typically obtained from the output of the previous layer \mathbf{x}_{l-1} by a mapping H_l :

$$\mathbf{x}_l = H_l(\mathbf{x}_{l-1}). \quad (1)$$

where H_l commonly integrates a convolution layers followed by a non-linear activation. In a densely-connected network, all feature outputs are concatenated in a feed-forward manner,

$$\mathbf{x}_l = H_l([\mathbf{x}_{l-1}, \mathbf{x}_{l-2}, \dots, \mathbf{x}_0]), \quad (2)$$

where $[\dots]$ denotes a concatenation operation.

As in HyperDenseNet [9,11], the outputs from layers in different streams are also linked. This connectivity yields a much more powerful feature representation than early or late fusion strategies in a multi-modal context, as the network learns the complex relationships between the modalities within and in-between all the levels of abstractions. Considering the case of only two modalities, let \mathbf{x}_l^1 and \mathbf{x}_l^2 denote the outputs of the l^{th} layer in streams 1 and 2, respectively. In general, the output of the l^{th} layer in a stream s can then be defined as follows:

$$\mathbf{x}_l^s = H_l^s([\mathbf{x}_{l-1}^1, \mathbf{x}_{l-1}^2, \mathbf{x}_{l-2}^1, \mathbf{x}_{l-2}^2, \dots, \mathbf{x}_0^1, \mathbf{x}_0^2]). \quad (3)$$

Inspired by the recent findings in [6,38,40], where shuffling and interleaving feature map elements in a CNN improved the efficiency and performance, while serving as a strong regularizer, we concatenate feature maps in a different order for each branch and layer:

$$\mathbf{x}_l^s = H_l^s(\pi_l^s([\mathbf{x}_{l-1}^1, \mathbf{x}_{l-1}^2, \mathbf{x}_{l-2}^1, \mathbf{x}_{l-2}^2, \dots, \mathbf{x}_0^1, \mathbf{x}_0^2])), \quad (4)$$

with π_l^s being a function that permutes the feature maps given as input. Thus, in the case of two image modalities, we have:

$$\begin{aligned} \mathbf{x}_l^1 &= H_l^1([\mathbf{x}_{l-1}^1, \mathbf{x}_{l-1}^2, \mathbf{x}_{l-2}^1, \mathbf{x}_{l-2}^2, \dots, \mathbf{x}_0^1, \mathbf{x}_0^2]) \\ \mathbf{x}_l^2 &= H_l^2([\mathbf{x}_{l-1}^2, \mathbf{x}_{l-1}^1, \mathbf{x}_{l-2}^2, \mathbf{x}_{l-2}^1, \dots, \mathbf{x}_0^2, \mathbf{x}_0^1]) \end{aligned}$$

A detailed example of hyper-dense connectivity for the case of two image modalities is depicted in Fig. 2.

2.2 Extended Inception module

Salient regions in a given image can have extremely large variation in size. For example, in ischemic stroke lesion segmentation, the area occupied by a lesion highly varies from one image to another. Therefore, choosing the appropriate kernel size is not trivial. While a smaller kernel is better for local information, a larger kernel is preferred to capture information that is distributed globally. InceptionNet [32] exploits this principle by including convolutions with multiple kernel sizes which operate on the same level. Furthermore, in versions 2 and 3, convolutions of the shape $n \times n$ are factorized to a combination of $1 \times n$ and $n \times 1$ convolutions, which have demonstrated to be more efficient. For example, a 3×3 convolution is equivalent to a 1×3 followed by a 3×1 convolution, which was found to be 33% cheaper.

We also extended the convolutional module of InceptionNet to facilitate the learning of multiple context. Particularly, we included two additional convolutional blocks, with different dilation rates, which help the module to learn from multiple receptive fields and to increase the context with respect to the original inception module. Since dilated convolutions were shown to be better alternative to max-pooling when capturing global context [36], we removed the latter operation in the proposed module. Our extended inception modules are depicted in Fig. 3.

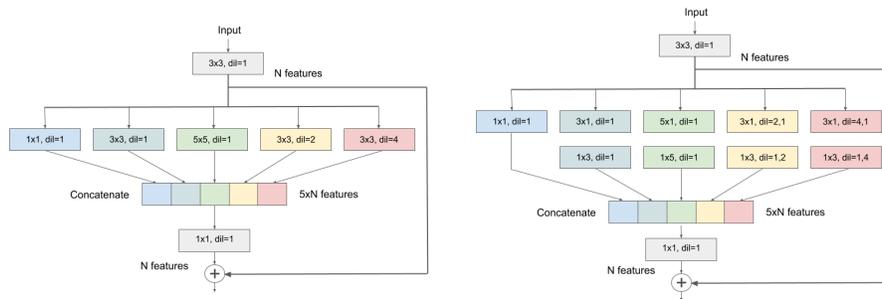


Fig. 3. Proposed extended inception modules. The module on the left employs standard convolutions while the module on the right adopts the idea of asymmetric convolutions [32].

3 Materials

3.1 Dataset

The training dataset, composed of 103 ischemic stroke lesion multi-modal scans, was provided by the ISLES organizers. We split the 94 stroke datasets into training and validation sets containing 83 and 11 examples, respectively. Each scan contains: Diffusion maps (DWI) and Perfusion maps (CBF,MTT,CBV,Tmax and CTP source data). In addition, the manual ground truth segmentation is provided only for the training samples. Detailed information about the dataset can be found in the ISLES website¹.

3.2 Evaluation metrics

Dice similarity coefficient (DSC). We first evaluate performance using Dice similarity coefficient (DSC), which compares volumes based on their overlap. Let V_{ref} and V_{auto} be, respectively, the reference and automatic segmentations of a given tissue class and for a given subject, the DSC for this subject is defined as:

$$\text{DSC}(V_{\text{ref}}, V_{\text{auto}}) = \frac{2 | V_{\text{ref}} \cap V_{\text{auto}} |}{| V_{\text{ref}} | + | V_{\text{auto}} |} \quad (5)$$

Modified Hausdorff distance (MHD). The second metric measures the accuracy of the segmentation boundary. Let P_{ref} and P_{auto} denote the sets of voxels within the reference and automatic segmentation boundary, respectively. MHD is given by

$$\text{MHD}(P_{\text{ref}}, P_{\text{auto}}) = \max \left\{ \max_{q \in P_{\text{ref}}} d(q, P_{\text{auto}}), \max_{q \in P_{\text{auto}}} d(q, P_{\text{ref}}) \right\}, \quad (6)$$

where $d(q, P)$ is the point-to-set distance defined by: $d(q, P) = \min_{p \in P} \|q - p\|$, with $\|\cdot\|$ denoting the Euclidean distance. In the MHD, the 95th percentile is used for the estimation of the maximum distance value. Low MHD values indicate high boundary similarity.

Volumetric similarity (VS). Volumetric similarity (VS) ignores the overlap between the predicted and reference segmentations, and simply compares the size of the predicted volume to that of the reference:

$$\text{VS}(V_{\text{ref}}, V_{\text{auto}}) = 1 - \frac{| |V_{\text{ref}}| - |V_{\text{auto}}| |}{|V_{\text{ref}}| + |V_{\text{auto}}|}. \quad (7)$$

where a VS equal to 1 reflects that the predicted segmentation is the same size as the reference volume.

¹ <http://www.isles-challenge.org>

3.3 Implementation details

Baselines. To demonstrate the effectiveness of hyper-dense connectivity in deep neural networks we compare the proposed architecture to the same network with early and late fusion strategies. For early fusion, all the MRI image modalities are merged into a single input, which is processed through a unique path, as many current works. On the other hand, each image modality is treated as an independent signal and processed by separate branches in the later fusion strategy, where features are fused at a higher level. The details of the late fusion architecture are depicted in Table 1. In both cases, i.e., early and late fusion, the left module depicted in Fig. 3 is employed. Furthermore, feature maps from the skip connections are summed before being fed into the convolutional modules of the decoding path, instead of concatenating them, as in standard UNet.

Proposed network. The proposed network is similar to the architecture with the late fusion strategy. Nevertheless, as introduced in section 2.1, feature maps from previous layers and different paths are concatenated and fed into the subsequent layers. The details of the resulted architecture are reported in Table 1, most-right columns. The first version of the proposed network employs the same convolutional module than the two baselines. The second version, however, adopts asymmetric convolutions instead (Fig. 3).

Table 1. Layers disposal of the architecture with late fusion and the proposed hyper dense connected UNet.

	Name	Late fusion		HyperDense connectivity	
		Feat maps (input)	Feat maps (output)	Feat maps (input)	Feat maps (output)
Encoding Path (each modality)	Conv Layer 1	$1 \times 256 \times 256$	$32 \times 256 \times 256$	$1 \times 256 \times 256$	$32 \times 256 \times 256$
	Max-pooling 1	$32 \times 256 \times 256$	$32 \times 128 \times 128$	$32 \times 256 \times 256$	$32 \times 128 \times 128$
	Layer 2	$32 \times 128 \times 128$	$64 \times 128 \times 128$	$128 \times 128 \times 128$	$64 \times 128 \times 128$
	Max-pooling 2	$64 \times 128 \times 128$	$64 \times 64 \times 64$	$64 \times 128 \times 128$	$64 \times 64 \times 64$
	Layer 3	$64 \times 64 \times 64$	$128 \times 64 \times 64$	$384 \times 64 \times 64$	$128 \times 64 \times 64$
	Max-pooling 3	$128 \times 64 \times 64$	$128 \times 32 \times 32$	$128 \times 64 \times 64$	$128 \times 32 \times 32$
	Layer 4	$128 \times 32 \times 32$	$256 \times 32 \times 32$	$896 \times 32 \times 32$	$256 \times 32 \times 32$
	Max-pooling 4	$256 \times 32 \times 32$	$256 \times 16 \times 16$	$256 \times 32 \times 32$	$256 \times 16 \times 16$
	Bridge	$1024 \times 16 \times 16$	$512 \times 16 \times 16$	$1920 \times 16 \times 16$	$512 \times 16 \times 16$
Decoding Path	Up-sample 1	$512 \times 16 \times 16$	$256 \times 32 \times 32$	$512 \times 16 \times 16$	$256 \times 32 \times 32$
	Layer 5	$256 \times 32 \times 32$	$256 \times 32 \times 32$	$256 \times 32 \times 32$	$256 \times 32 \times 32$
	Up-sample 2	$256 \times 32 \times 32$	$128 \times 64 \times 64$	$256 \times 32 \times 32$	$128 \times 64 \times 64$
	Layer 6	$128 \times 64 \times 64$	$128 \times 64 \times 64$	$128 \times 64 \times 64$	$128 \times 64 \times 64$
	Up-sample 3	$128 \times 64 \times 64$	$64 \times 128 \times 128$	$128 \times 64 \times 64$	$64 \times 128 \times 128$
	Layer 7	$64 \times 128 \times 128$	$64 \times 128 \times 128$	$64 \times 128 \times 128$	$64 \times 128 \times 128$
	Up-sample 4	$64 \times 128 \times 128$	$32 \times 256 \times 256$	$64 \times 128 \times 128$	$32 \times 256 \times 256$
	Layer 8	$32 \times 256 \times 256$	$32 \times 256 \times 256$	$32 \times 256 \times 256$	$32 \times 256 \times 256$
	Softmax layer	$32 \times 256 \times 256$	$2 \times 256 \times 256$	$32 \times 256 \times 256$	$2 \times 256 \times 256$

Training. Network parameters were optimized via Adam with β_1 and β_2 equal to 0.9 and 0.99, respectively and training is run during 200 epochs. Learning rate is initially set to 1×10^{-4} and reduced after 100 epochs. Batch size was equal to 4. For a fair comparison, the same hyper-parameters were employed

across all the architectures. The proposed architectures were implemented in pytorch. Experiments were performed on a NVidia TITAN XP GPU with 16 GBs RAM. While training took around 5 hours, inference on a single 2D image was done in 0.1 sec, as average. No data augmentation was employed. Images were normalized between 0 and 1 and no other pre- or post-processing steps were used. As input to the architectures we employed the following four image modalities in all the cases: CBV, CTP, DWI and MTT.

4 Results

Table 2 reports the results obtained by the different networks that we investigated in terms of mean DSC, MHD and VS values and their standard deviation. First, we compare the different multi-modal fusion strategies with the baseline UNet employed in this work. We can observe that fusing learning features in a higher level provides better results in all the metrics than early fusion strategies. Additionally, if hyper-dense connections are adopted in the late fusion architecture, i.e., interconnecting convolutional layers from the different image modalities, the segmentation performance is significantly improved, particularly in terms of DSC and VS. Specifically, while the proposed network outperforms the late fusion architecture by nearly 5% in both DSC and VS, the mean MHD is decreased by almost 1 mm, obtaining a mean MHD of 18.88 mm. On the other hand, replacing the standard convolutions of the proposed module (Fig 3, *left*) by asymmetric convolutions (Fig 3, *right*), brings another boost on performance on the proposed hyper-dense UNet. In this case, the mean DSC and MHD are the best ones among all the architectures, with mean values of 0.635% and 18.64 mm, respectively.

Table 2. Mean DSC, MHD and VS values, with their corresponding standard deviation, obtained by the evaluated methods on the independent validation group.

Architecture	Validation		
	DSC (%)	MHD (mm)	VS (%)
Early Fusion	0.497 ± 0.263	21.30 ± 13.25	0.654 ± 0.265
Late Fusion	0.571 ± 0.221	19.72 ± 12.29	0.718 ± 0.235
Proposed	0.622 ± 0.233	18.88 ± 14.87	0.764 ± 0.247
Proposed (Asymmetric conv)	0.635 ± 0.186	18.64 ± 14.26	0.796 ± 0.162
ERFNet [26]	0.540 ± 0.258	21.73 ± 11.46	0.823 ± 0.119

Then, we also compare the results obtained by the proposed network to another state-of-the-art network that includes factorized convolution modules, i.e., ERFNet. Even though its performance outperforms the baseline with early fusion, results are far from those obtained by the proposed network, except for the volume similarity, where both ERFNet and the proposed network with asymmetric convolutions obtain similar performances.

Qualitative evaluation of the proposed architecture is assessed in Fig. 4, where ground truth and automatic CNN contours are visualized on MTT images. We

can first observe that, by employing strategies where learned features are merged at higher levels, unlike *early fusion*, the region of the ischemic stroke lesion is generally better covered. Furthermore, by giving freedom to the architecture to learn the level of abstraction at which the different modalities should be combined segmentation results are visually improved, which is in line with the results reported in Table 2.

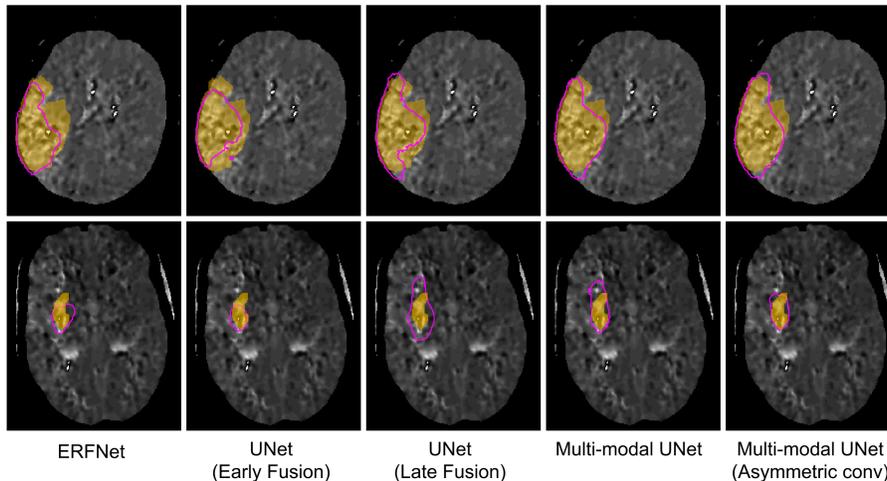


Fig. 4. Visual results for two subjects on the validation set. While the area in yellow represents the ground truth, purple contours depict the automatic contours for each of the architectures.

5 Discussion

In this work we extended the well-known UNet to leverage information in multi-modal data. Particularly, following recent work in multi-modal learning for medical image segmentation [9,11], we processed each image modality in different streams on the encoding path and densely connected all the convolutional layers from all the streams. Thus, each convolutional layer received as input the features maps from all the previous layers within the same stream, i.e., same modality, but also the learned features from previous layers in every different stream. In this way, the network has the freedom to learn any pattern at any level of abstraction of the network, which seems to improve its representation learning power.

Results obtained in this work demonstrate that better strategies to model multi-modal information can bring a boost on performance when compared to more naive fusion strategies. These results are in line with recent studies in multi-modal image segmentation on the medical field [23,9,11]. For instance, in [23], a *late fusion* strategy was proposed to combine high-level features to better

capture the complex relationships between different modalities. They used an independent convolutional network for each modality, and fused the outputs of the different networks in higher-level layers, showing better performance than early fusion in the context infant brain segmentation. More recently, we demonstrated that hyper dense connectivity can strength the representation power of deep CNNs in the context of multi-modal image infant and adult brain segmentation, surpassing the performance of several features fusion strategies [9,11].

One of the limitations of this work is that volumes were treated as a stack of 2D slides, where each slide was processed independently. Thus, 3D context was discarded, which might have improved the segmentation performance, as shown by recent works that employ 3D convolutions. One of the reasons for privileging 2D convolutions is that some of the volumes on the ISLES dataset contained a limited number of slides, i.e., 2 and 4 slides in many cases. One strategy to explore in the future could be to employ Long-Short Term Memory (LSTM) networks to propagate the spatial information extracted from the 2D CNN through the third dimension.

Acknowledgments This work is supported by the National Science and Engineering Research Council of Canada (NSERC), discovery grant program, and by the ETS Research Chair on Artificial Intelligence in Medical Imaging.

References

1. M. Aygün, Y. H. Şahin, and G. Ünal. Multi modal convolutional neural networks for brain tumor segmentation. *arXiv preprint arXiv:1809.06191*, 2018.
2. P. Barber, M. Hill, M. Eliasziw, A. Demchuk, J. Pexman, M. Hudon, A. Tomanek, R. Frayne, and A. Buchan. Imaging of the brain in acute ischaemic stroke: comparison of computed tomography and magnetic resonance diffusion-weighted imaging. *Journal of Neurology, Neurosurgery & Psychiatry*, 76(11):1528–1533, 2005.
3. J. A. Chalela, C. S. Kidwell, L. M. Nentwich, M. Luby, J. A. Butman, A. M. Demchuk, M. D. Hill, N. Patronas, L. Latour, and S. Warach. Magnetic resonance imaging and computed tomography in emergency assessment of patients with suspected acute stroke: a prospective comparison. *The Lancet*, 369(9558):293–298, 2007.
4. L. Chen, P. Bentley, and D. Rueckert. Fully automatic acute ischemic lesion segmentation in DWI using convolutional neural networks. *NeuroImage: Clinical*, 15:633–643, 2017.
5. L. Chen, Y. Wu, A. M. DSouza, A. Z. Abidin, A. Wismüller, and C. Xu. Mri tumor segmentation with densely connected 3D CNN. In *Medical Imaging 2018: Image Processing*. International Society for Optics and Photonics, 2018.
6. Y. Chen, H. Wang, and Y. Long. Regularization of convolutional neural networks using shufflenode. In *Multimedia and Expo (ICME), 2017 IEEE International Conference on*, pages 355–360. IEEE, 2017.
7. Y. Choi, Y. Kwon, H. Lee, B. J. Kim, M. C. Paik, and J.-H. Won. Ensemble of deep convolutional neural networks for prognosis of ischemic stroke. In *Workshop on Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, pages 231–243. Springer, 2016.

8. Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger. 3d u-net: learning dense volumetric segmentation from sparse annotation. In *MICCAI*, pages 424–432. Springer, 2016.
9. J. Dolz, I. Ben Ayed, J. Yuan, and C. Desrosiers. Isointense infant brain segmentation with a hyper-dense connected convolutional neural network. In *Biomedical Imaging (ISBI 2018), 2018 IEEE 15th International Symposium on*, pages 616–620. IEEE, 2018.
10. J. Dolz, C. Desrosiers, L. Wang, J. Yuan, D. Shen, and I. B. Ayed. Deep CNN ensembles and suggestive annotations for infant brain MRI segmentation. *arXiv preprint arXiv:1712.05319*, 2017.
11. J. Dolz, K. Gopinath, J. Yuan, H. Lombaert, C. Desrosiers, and I. B. Ayed. Hyperdense-net: A hyper-densely connected cnn for multi-modal image segmentation. *arXiv preprint arXiv:1804.02967*, 2018.
12. H. Dong, G. Yang, F. Liu, Y. Mo, and Y. Guo. Automatic brain tumor detection and segmentation using u-net based fully convolutional networks. In *MIUA*, pages 506–517. Springer, 2017.
13. V. L. Feigin, C. M. Lawes, D. A. Bennett, and C. S. Anderson. Stroke epidemiology: a review of population-based studies of incidence, prevalence, and case-fatality in the late 20th century. *The Lancet Neurology*, 2(1):43–53, 2003.
14. R. Guerrero, C. Qin, O. Oktay, C. Bowles, L. Chen, R. Joules, R. Wolz, M. Valdés-Hernández, D. Dickie, J. Wardlaw, et al. White matter hyperintensity and stroke lesion segmentation and differentiation using convolutional neural networks. *NeuroImage: Clinical*, 17:918–934, 2018.
15. G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger. Densely connected convolutional networks. In *CVPR*, volume 1, page 3, 2017.
16. K. Kamnitsas, L. Chen, C. Ledig, D. Rueckert, and B. Glocker. Multi-scale 3D convolutional neural networks for lesion segmentation in brain MRI. *Ischemic Stroke Lesion Segmentation*, 13, 2015.
17. K. Kamnitsas, C. Ledig, V. F. Newcombe, J. P. Simpson, A. D. Kane, D. K. Menon, D. Rueckert, and B. Glocker. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. *Medical image analysis*, 36:61–78, 2017.
18. M. G. Lansberg, G. W. Albers, C. Beaulieu, and M. P. Marks. Comparison of diffusion-weighted mri and ct in acute stroke. *Neurology*, 54(8):1557–1561, 2000.
19. A. D. Lopez, C. D. Mathers, M. Ezzati, D. T. Jamison, and C. J. Murray. Global and regional burden of disease and risk factors, 2001: systematic analysis of population health data. *The Lancet*, 367(9524):1747–1757, 2006.
20. O. Maier, B. H. Menze, J. von der Gablentz, L. Häni, M. P. Heinrich, M. Liebrand, S. Winzeck, A. Basit, P. Bentley, L. Chen, et al. Isles 2015-a public evaluation benchmark for ischemic stroke lesion segmentation from multispectral MRI. *Medical image analysis*, 35:250–269, 2017.
21. O. Maier, C. Schröder, N. D. Forkert, T. Martinetz, and H. Handels. Classifiers for ischemic stroke lesion segmentation: a comparison study. *PloS one*, 10(12):e0145118, 2015.
22. P. Moeskops, M. A. Viergever, A. M. Mendrik, L. S. de Vries, M. J. Benders, and I. Išgum. Automatic segmentation of MR brain images with a convolutional neural network. *IEEE Transactions on Medical Imaging*, 35(5):1252–1261, 2016.
23. D. Nie, L. Wang, Y. Gao, and D. Sken. Fully convolutional networks for multi-modality isointense infant brain image segmentation. In *13th International Symposium on Biomedical Imaging (ISBI), 2016*, pages 1342–1345. IEEE, 2016.

24. G. Praveen, A. Agrawal, P. Sundaram, and S. Sardesai. Ischemic stroke lesion segmentation using stacked sparse autoencoder. *Computers in biology and medicine*, 2018.
25. I. Rekik, S. Allasonnière, T. K. Carpenter, and J. M. Wardlaw. Medical image analysis methods in MR/CT-imaged acute-subacute ischemic stroke lesion: Segmentation, prediction and insights into dynamic evolution simulation models. A critical appraisal. *NeuroImage: Clinical*, 1(1):164–178, 2012.
26. E. Romera, J. M. Alvarez, L. M. Bergasa, and R. Arroyo. Erfnet: Efficient residual factorized convnet for real-time semantic segmentation. *IEEE Transactions on Intelligent Transportation Systems*, 19(1):263–272, 2018.
27. O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In *MICCAI*, pages 234–241. Springer, 2015.
28. S. Seshadri and P. A. Wolf. Lifetime risk of stroke and dementia: current concepts, and estimates from the framingham study. *The Lancet Neurology*, 6(12):1106–1114, 2007.
29. K. Sirinukunwattana, J. P. Pluim, H. Chen, X. Qi, P.-A. Heng, Y. B. Guo, L. Y. Wang, B. J. Matuszewski, E. Bruni, U. Sanchez, et al. Gland segmentation in colon histology images: The glas challenge contest. *Medical image analysis*, 35:489–502, 2017.
30. N. Srivastava and R. Salakhutdinov. Multimodal learning with deep boltzmann machines. *Journal of Machine Learning Research*, 15:2949–2980, 2014.
31. C. Sudlow and C. Warlow. Comparable studies of the incidence of stroke and its pathological types: results from an international collaboration. *Stroke*, 28(3):491–499, 1997.
32. C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna. Rethinking the inception architecture for computer vision. In *CVPR*, pages 2818–2826, 2016.
33. S. Valverde, M. Cabezas, E. Roura, S. González-Villà, D. Pareto, J. C. Vilanova, L. Ramió-Torrentà, À. Rovira, A. Oliver, and X. Lladó. Improving automated multiple sclerosis lesion segmentation with a cascaded 3D convolutional neural network approach. *NeuroImage*, 155:159–168, 2017.
34. H. B. Van der Worp and J. van Gijn. Acute ischemic stroke. *New England Journal of Medicine*, 357(6):572–579, 2007.
35. S. Winzeck, A. Hakim, R. McKinley, J. A. Pinto, V. Alves, C. Silva, M. Pisov, E. Krivov, M. Belyaev, M. Monteiro, et al. Isles 2016 and 2017-benchmarking ischemic stroke lesion outcome prediction based on multispectral mri. *Frontiers in Neurology*, 9, 2018.
36. F. Yu and V. Koltun. Multi-scale context aggregation by dilated convolutions. *arXiv preprint arXiv:1511.07122*, 2015.
37. L. Yu, J.-Z. Cheng, Q. Dou, X. Yang, H. Chen, J. Qin, and P.-A. Heng. Automatic 3D cardiovascular MR segmentation with densely-connected volumetric convnets. In *MICCAI*, pages 287–295. Springer, 2017.
38. T. Zhang, G.-J. Qi, B. Xiao, and J. Wang. Interleaved group convolutions. In *CVPR*, pages 4373–4382, 2017.
39. W. Zhang, R. Li, H. Deng, L. Wang, W. Lin, S. Ji, and D. Shen. Deep convolutional neural networks for multi-modality isointense infant brain image segmentation. *NeuroImage*, 108:214–224, 2015.
40. X. Zhang, X. Zhou, M. Lin, and J. Sun. Shufflenet: An extremely efficient convolutional neural network for mobile devices. *arXiv preprint arXiv:1707.01083*, 2017.