Medical Image Analysis 44 (2018) 1–13

Contents lists available at ScienceDirect

Medical Image Analysis

journal homepage: www.elsevier.com/locate/media

Learning normalized inputs for iterative estimation in medical image segmentation

Michal Drozdzal[a,b,c,*], Gabriel Chartrand[c], Eugene Vorontsov[a,b,c], Mahsa Shakeri[b], Lisa Di Jorio[c], An Tang[d], Adriana Romero[a], Yoshua Bengio[a], Chris Pal[a,b], Samuel Kadoury[b,d]

[a] Montreal Institute for Learning Algorithms, Montréal, Canada
[b] École Polytechnique de Montréal, Montréal, Canada
[c] Imagia Inc., Montréal, Canada
[d] CHUM Research Center, Montréal, Canada

ARTICLE INFO

Article history:
Received 4 May 2017
Revised 26 September 2017
Accepted 8 November 2017
Available online 14 November 2017

Keywords:
Image segmentation
Fully convolutionl networks
ResNets
Computed Tomography
Electron microscopy
Magnetic Resonance Imaging

ABSTRACT

In this paper, we introduce a simple, yet powerful pipeline for medical image segmentation that combines Fully Convolutional Networks (FCNs) with Fully Convolutional Residual Networks (FC-ResNets). We propose and examine a design that takes particular advantage of recent advances in the understanding of both Convolutional Neural Networks as well as ResNets. Our approach focuses upon the importance of a trainable pre-processing when using FC-ResNets and we show that a low-capacity FCN model can serve as a pre-processor to normalize medical input data. In our image segmentation pipeline, we use FCNs to obtain normalized images, which are then iteratively refined by means of a FC-ResNet to generate a segmentation prediction. As in other fully convolutional approaches, our pipeline can be used off-the-shelf on different image modalities. We show that using this pipeline, we exhibit state-of-the-art performance on the challenging Electron Microscopy benchmark, when compared to other 2D methods. We improve segmentation results on CT images of liver lesions, when contrasting with standard FCN methods. Moreover, when applying our 2D pipeline on a challenging 3D MRI prostate segmentation challenge we reach results that are competitive even when compared to 3D methods. The obtained results illustrate the strong potential and versatility of the pipeline by achieving accurate segmentations on a variety of image modalities and different anatomical regions.

1. Introduction

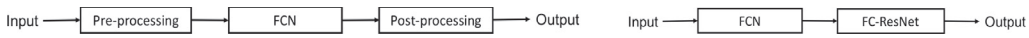
Segmentation is an active area of research in medical image analysis. With the introduction of Convolutional Neural Networks (CNNs), significant improvements in performance have been achieved in many standard datasets. For example, for the EM ISBI 2012 dataset (Arganda-Carreras et al., 2015), PROMISE12 challenge or MS lesions (Styner et al., 2008), the top entries are built on CNNs (Ronneberger et al., 2015; Chen et al., 2016b; Havaei et al., 2017; Yu et al., 2017).

The common view on CNN models is based on a *representation learning* perspective (Greff et al., 2017; Goodfellow et al., 2016). This view assumes that a CNN is built from layers (convolutional operations and non-linearities) that learn increasing levels of abstraction. In Zeiler and Fergus (2014), these different levels of abstraction were visualized, showing that in the first layer the network learns simple edge and blob detectors (e.g. Gabor-like filters), the second layer learns combination of these simple features, while the deeper layers learn to represent more complex object contours, such as faces or flowers. Moreover, in Veit et al. (2016), it was shown that removing any single layer of the network after training/finetuning can significantly harm the network's performance, implying that the transformations learnt by a CNN layer are very different from an identity mapping.

Recently, a new class of models called Residual Networks (ResNets) have been introduced (He et al., 2016a; 2016b). ResNets are built from hundreds of residual blocks. Each residual block is composed of two paths: the first one applies a series of nonlinear transformations (typically two or three transformations composed of Batch Normalization (Szegedy et al., 2015), convolution and a ReLu non-linearity), while the second one is an identity mapping. These two paths are summed up at the end of a residual block. This small architectural modification has three important implications. First, the gradient can flow uninterrupted allowing

* Corresponding author at : Montreal Institute for Learning Algorithms, Montréal, Canada.
E-mail address: michal.drozdzal@umontreal.ca (M. Drozdzal).

https://doi.org/10.1016/j.media.2017.11.005
1361-8415/



a) Traditional pipeline b) Our pipeline

Fig. 1. Traditional pipeline for medical image segmentation (a) and proposed pipeline for medical image segmentation (b).

parameters to be updated even in very deep networks. Second, ResNets are robust to layer removal at training time (Huang et al., 2016) and at inference time (Veit et al., 2016) implying that the operations applied by a single layer are only a small modification to identity operation. Third, ResNets are robust to layers permutation (Veit et al., 2016), suggesting that neighboring layers perform similar operations. These characteristics are not shared with traditional CNNs and researchers have attempted to propose possible explanations on the internal behaviors and mechanisms with ResNet-like models. Recently, two possible explanations of ResNets-like models have emerged. The first, explains the behavior of ResNets in terms of an *ensemble of relatively shallow networks* (Veit et al., 2016). The second, suggests that ResNets perform *iterative estimation*, where the input to the model is iteratively modified by small transformations (Greff et al., 2017; Liao and Poggio, 2016).

In recent years, state of the art segmentation methods for medical images have been based on Fully Convolutional Networks (FCNs) (Long et al., 2015; Ronneberger et al., 2015). While CNNs typically consist of a contracting path composed of convolutional, pooling and fully connected layers, FCNs add an expanding path composed of transposed convolutions or unpooling layers. The expanding path recovers spatial information by merging features skipped from the various resolution levels on the contracting path. Variants of these skip connections are proposed in the literature. In Long et al. (2015), upsampled feature maps are summed with feature maps skipped from the contractive path, while Ronneberger et al. (2015) concatenate them and add convolutions and non-linearities between each upsampling step. These skip connections have been shown to help recover the full spatial resolution at the network's output, making fully convolutional methods suitable for semantic segmentation. Since traditional FCNs are an extension of CNNs, they can be explained by the representation learning perspective on deep learning.

Although deep learning methods have proved their potential in medical image segmentation, their performance strongly depends on the quality of pre-processing and post-processing steps (Havaei et al., 2016). Thus, traditional image segmentation pipelines based on FCNs are often complemented by pre-processing and post-processing blocks (see Fig. 1(a)). Pre-processing methods vary among different imaging modalities and can include operations like standardization, histogram equalization, value clipping or range normalization (e.g. dividing by maximum intensity value). The tools of choice for post-processing are either based on Conditional Random Fields (Krähenbühl and Koltun, 2011) to account for spatial consistency of the output prediction or on morphological operations to clean the output prediction.

In Drozdzal et al. (2016), the Fully Convolutional Residual Networks (FC-ResNets) are introduced. FC-ResNets are a subclass of FCNs, which incorporate additional shortcut paths and, thus, increase the number of connections within a segmentation network. These additional shortcut paths have been shown not only to improve the segmentation accuracy but also help the network optimization process, resulting in faster convergence of the training. Since FC-ResNets are an extension of ResNets, their behavior should be interpreted in terms of iterative estimation or as an ensemble of relatively shallow networks. Moreover, the structure of FC-ResNets suggests that they might be more susceptible to image pre-processing than FCNs, making their performance potentially highly dependent on appropriate data preparation (e.g. data standardization or range normalization).

In this paper, we take advantage of recent advances in the understanding of both CNNs as well as ResNets and propose a new medical image segmentation pipeline. We use a FCN to obtain pre-normalized images, which are then iteratively refined by means of a FC-ResNet to generate a segmentation prediction (see Fig. 1(b)).

Thus, in our pipeline, the FCN can be thought of as a pre-processor that is learnt by means of back-propagation, and FC-ResNet can be thought as a powerful classifier that is an ensemble of exponential number of shallow models. This small modification to current segmentation pipelines allows to remove hand-crafted data pre-processing and to build end-to-end systems trained with back-propagation that achieve surprisingly good performance in segmentation tasks for bio-medical images. Our pipeline reaches state-of-the art for 2D methods on electron microscopy (EM) ISBI benchmark dataset (Arganda-Carreras et al., 2015) and outperforms both standard FCN (Long et al., 2015; Ronneberger et al., 2015) and FC-ResNet (Drozdzal et al., 2016) on in-house CT liver lesion segmentation dataset. Moreover, while applying our 2D pipeline off-the-shelf on a challenging 3D MRI prostate segmentation challenge, we reach results that are competitive even when compared to 3D methods. Similarly, we achieve highly competitive results on an MRI spine segmentation benchmark.

Thus, the contributions in this paper can be summarized as follows:

- We combine Fully Convolutional Residual Networks with Fully Convolutional Networks (Section 3).
- We show that a very deep network without any hand-designed pre- or post-processing achieves state-of-the-art performance on the challenging EM ISBI benchmark (Arganda-Carreras et al., 2015) (Section 4.1).
- We show that our pipeline outperforms other common FCN-based segmentation pipelines on our in-house CT liver lesion dataset (Section 4.2).
- We show that our 2D pipeline can be applied to untreated MR images reaching competitive results both on a challenging 3D MRI prostate segmentation task, outperforming many 3D based approaches (Section 4.3), and on a spine segmentation task (Section 4.4).
- We show that a FCN based pre-processor normalizes the data (Section 4.5).

2. Background

Recent advances in medical image segmentation often involve convolutional networks. Most of these state-of-the-art approaches are based on either variants of FCNs (FCN8 or UNet) or CNN architectures. FCN architectures process the input image end-to-end and provide a full resolution segmentation map, whereas CNN variants are applied to input patches and aim to solely classify the central pixel of each patch. In many cases, the application of an FCN/CNN is preceded by a pre-processing step and followed by a post-processing step. The former aims to account for the variability in the input images, whereas the latter helps refine the predictions made by the FCN/CNN.

In the remainder of this section, we review segmentation pipelines in medical imaging based on deep neural networks, for different imaging modalities and organs, with a particular focus on

pre-processing and post-processing steps proposed to normalize and regularize data, respectively.

Electron Microscopy (EM). EM is widely used to study synapses and other sub-cellular structures in the mammalian nervous system. EM data is the basis of two medical imaging segmentation challenges: 2D (Arganda-Carreras et al., 2015) and 3D (Arganda-Carreras et al., 2013). The core of the best performing methods is based on a patch-based CNN (Ciresan et al., 2012), FCN8 (Chen et al., 2016b), UNet (Ronneberger et al., 2015) and FC-ResNets (Drozdzal et al., 2016; Quan et al., 2016; Fakhry et al., 2016). The methods account for gray scale value variability by employing data augmentation (e.g. intensity shifts Ronneberger et al., 2015 and Quan et al., 2016) or data pre-processing (e.g standardization Drozdzal et al., 2016 or rescaling Quan et al., 2016). As for post-processing, a variety of FCN prediction refinement techniques have been proposed. In Beier et al. (2016), a minimum cost multicut approach is introduced, in Fakhry et al. (2016) and Chen et al. (2016b), watershed algorithm is used, while in Quan et al., (2016), median filtering is employed to improve EM segmentation results.

Magnetic Resonance Imaging (MRI). MRI has a broad range of diagnostic and interventional applications in the heart, brain, abdominal organs, including detection, lesion classification and tumor staging. As in many medical imaging applications, deep learning methods are nowadays playing an important role in the segmentation of lesions and organs from MRI scans, consistently improving state-of-the-art performance. Patch-based CNN architectures have been used to segment brain tumors (Kamnitsas et al., 2017; Pandian et al., 2016) and MS lesions (Birenbaum and Greenspan, 2016). Variants of FCN8 and UNets have also been applied to segment brain tumors (Chang, 2016; Lun and Hsu, 2016; McKinley et al., 2016; Casamitjana et al., 2016; Zhao et al., 2016), whereas FC-ResNets have been successfully employed to segment white and gray matter in the brain (Chen et al., 2016a) and prostate (Yu et al., 2017; Milletari et al., 2016). To account for 3D information, some of the proposed architectures apply convolutions in a 3D fashion (Chen et al., 2016a; Casamitjana et al., 2016; Pandian et al., 2016; Kamnitsas et al., 2017; Milletari et al., 2016) or make use of recurrent neural networks (LSTMs or GRUs) (Andermatt et al., 2016; Stollenga et al., 2015). As in the previous modalities, pre-processing techniques are applied to the input of the network. MRI pre-processing techniques include intensity normalization (Birenbaum and Greenspan, 2016; Chen et al., 2016a), rescaling (Chang, 2016), standardization (Kamnitsas et al., 2017; Pandian et al., 2016; Casamitjana et al., 2016; Zhao et al., 2016; Chen et al., 2016a; Andermatt et al., 2016; Stollenga et al., 2015; Yu et al., 2017), N4 bias correction (Pandian et al., 2016; Zhao et al., 2016; Birenbaum and Greenspan, 2016; Stollenga et al., 2015) and histogram equalization (Chang, 2016). Again, post-processing involving morphological operations (Pandian et al., 2016; Zhao et al., 2016), conditional random fields (CRFs) (Kamnitsas et al., 2017; Zhao et al., 2016) and interpolation (Andermatt et al., 2016) are often used to refine segmentation maps.

Computed Tomography (CT). Finally, CT scans are widely used by clinicians to detect a large variety of lesions in different organs. Patch-based CNNs have been used in the literature to segment pathological kidneys (Zheng et al., 2016; Thong et al., 2016), liver tumors (Vivanti et al., 2015), pancreas (Roth et al., 2015), liver (Hu et al., 2016; Dou et al., 2016) and urinary bladder (Cha et al., 2016). FCN8 and UNet variants have also been tried on CT scans, e.g. to segment liver tumors (Ben-Cohen et al., 2016; Christ et al., 2016). In many cases, CT scans are pre-processed by standardizing (Thong et al., 2016; Li et al., 2015; Cha et al., 2016), clipping (Christ et al., 2016), applying Gaussian smoothing (Li et al., 2015; Cha et al., 2016) or histogram equalization (Christ et al., 2016) to the input. Gaussian noise injection has also been explored as part of data augmentation to account for noise level variability in the CT scans (Christ et al., 2016). As in the other modalities, the most frequently used post-processing methods include morphological operations (Thong et al., 2016; Vivanti et al., 2015; Roth et al., 2015), CRFs (Christ et al., 2016) and level sets (Cha et al., 2016) to refine the segmentation proposals.

Table 1
Detailed FCN architecture used as a pre-processor in the experiments. Output resolution indicates the spatial resolution of feature maps for an input of size 512×512 and output width represents the feature map dimensionality. Repetition number indicates the number of times the layer is repeated.

Layer name	Block type	Output resolution	Output width	Repetition number
Input	–	512×512	1	–
Down 1	conv 3×3	512×512	16	2
Pooling 1	maxpooling	256×256	16	1
Down 2	conv 3×3	256×256	32	2
Pooling 2	maxpooling	128×128	32	1
Down 3	conv 3×3	128×128	64	2
Pooling 3	maxpooling	64×64	64	1
Down 4	conv 3×3	64×64	128	2
Pooling 4	maxpooling	32×32	128	1
Across	conv 3×3	32×32	256	2
Up 1	upsampling	64×64	256	1
Merge 1	concatenate	64×64	384	1
Up 2	conv 2×2	64×64	128	1
Up 3	conv 3×3	64×64	128	2
Up 4	upsampling	128×128	128	1
Merge 2	concatenate	128×128	192	1
Up 5	conv 2×2	128×128	64	1
Up 6	conv 3×3	128×128	64	2
Up 7	upsampling	256×256	64	1
Merge 3	concatenate	256×256	96	1
Up 8	conv 2×2	256×256	32	1
Up 9	conv 3×3	256×256	32	2
Up 10	upsampling	512×512	32	1
Merge 4	concatenate	512×512	48	1
Up 11	conv 2×2	512×512	16	1
Up 12	conv 3×3	512×512	16	2
Output	conv 3×3	512×512	1	1

3. Method

In this section, we present our segmentation pipeline. As highlighted in Section 1, our approach combines a FCN model with a FC-ResNet model (see Fig. 1(b)). The goal of FCN in our pipeline is to pre-process the image to a format that can be iteratively refined by FC-ResNet. In the remainder of this section, we describe the architecture of our FCN-based pre-processor (see Section 3.1), our FC-ResNet (see Section 3.2) as well as the loss function used to train our pipeline (see Section 3.3).

3.1. Fully convolutional pre-processor

Our FCN takes as input a raw image of size $N \times N \times 1$ (e.g. CT scan slice or EM image) without applying any pre-processing and outputs a processed $N \times N \times 1$ feature map. The FCN pre-processor architecture is described as a variation of the standard medical image segmentation pipeline: the UNet model from (Ronneberger et al., 2015) (see Table 1 for details). The contracting path is built by alternating convolutions and max pooling operations, whereas the expanding path is built by alternating convolutions and repeat operations. The expanding path recovers spatial information, lost in pooling operations, by concatenating the corresponding feature maps from the contracting path. In total, the model has 4 pooling operations and 4 repeat operations. All 3×3 convolutions are followed by ReLU non-linearity, while no non-linearity follows the 1×1 and 2×2 convolutions. In our experiments, we reduce the number of feature maps by a factor of 4 when compared to the original UNet (e. g. our first layer

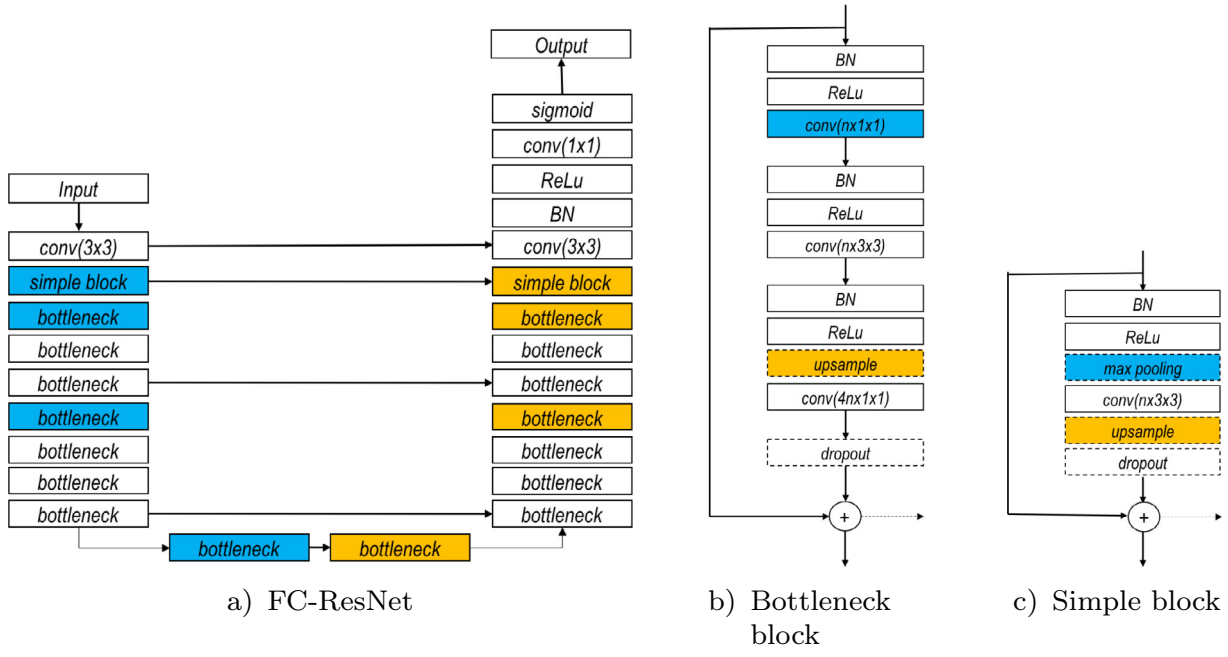


Fig. 2. An example of residual network for image segmentation. (a) Residual Network with long skip connections built from bottleneck blocks, (b) bottleneck block and (c) simple block. Blue color indicates the blocks where a downsampling is optionally performed, yellow color depicts the (optional) upsampling blocks, dashed arrow in Fig. (b) and (c) indicates possible long skip connections. Note that blocks (b) and (c) can have a dropout (Srivastava et al., 2014) layer (depicted with dashed line rectangle). The downsampling operation is performed with strided convolution in the bottleneck block and with maxpooling in the simple block. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

has 16 feature maps instead of 64 as in the original model). This step significantly reduces the memory foot-print of the FCN-based pre-processor. Thus, our UNet-like pre-processor has 1.8 million trainable parameters (vs 33 million in the original implementation of Ronneberger et al., 2015). Note that the UNet could potentially be replaced by other FCN models (e.g. FCN8 Long et al., 2015).

3.2. Iterative estimation with FC-ResNets

ResNets (He et al., 2016a) introduce a residual block that sums the identity mapping of the input to the output of a layer allowing for the reuse of features and permitting the gradient to flow directly to earlier layers. The resulting output x_ℓ of the ℓth block becomes

$$x_\ell = H_\ell(x_{\ell-1}) + x_{\ell-1}, \quad (1)$$

where H is defined as the repetition (2 or 3 times) of a block composed of Batch Normalization (BN) (Ioffe and Szegedy, 2015) followed by ReLU and a convolution.

The FC-ResNet (Drozdzal et al., 2016) model extends ResNets to be fully convolutional by adding an expanding (upsampling) path (Fig. 2(a)). Spatial reduction is performed along the contracting path (left) and expansion is performed along the expanding path (right). As in Long et al., (2015) and Ronneberger et al., (2015), spatial information lost along the contracting path is recovered in the expanding path by skipping equal resolution features from the former to the latter. Similarly to the identity connections in ResNets, the skipped features coming from the contracting path are summed with the ones in the expanding path.

Following the spirit of ResNets, FC-ResNets are composed of two different types of blocks: simple blocks and bottleneck blocks, each composed of at least one batch normalization followed by a non-linearity and one convolution (see Fig. 2(b) and (c)). These blocks can maintain the spatial resolution of their input (marked white in Fig. 2(b) and (c)), perform spatial downsampling (marked blue in Fig. 2(b) and (c)) or spatial upsampling (marked yellow in Fig. 2(a)-(c)). As in ResNets, bottleneck blocks are characterized by their 1×1 convolutions, which are responsible for reducing and restoring the number of feature maps and thus, aim to mitigate the number of parameters of the model.

Table 2
Detailed FC-ResNet architecture used in the experiments. Output resolution indicates the spatial resolution of feature maps for an input of size 512×512 and output width represents the feature map dimensionality. Repetition number indicates the number of times the block is repeated. For the definition of block types, please refer to Fig. 2(a) and (c).

Layer name	Block type	Output resolution	Output width	Repetition number
Down 1	conv 3×3	512×512	32	1
Down 2	simple block	256×256	32	1
Down 3	bottleneck	128×128	128	3
Down 4	bottleneck	64×64	256	8
Down 5	bottleneck	32×32	512	10
Across	bottleneck	32×32	1024	3
Up 1	bottleneck	64×64	512	10
Up 2	bottleneck	128×128	256	8
Up 3	bottleneck	256×256	128	3
Up 4	simple block	512×512	32	1
Up 5	conv 3×3	512×512	32	1
Classifier	conv 1×1	512×512	1	1

The detailed description of FC-ResNet architecture used in our pipeline is shown in Table 2. The contracting path contains 5 downsampling operations, one 3×3 convolution, one simple block and 21 bottleneck blocks. The contracting path is followed by 3 bottleneck blocks, which precede the expanding path. The expanding path contains 5 upsampling operations, 21 bottleneck blocks, one simple block and one last 3×3 convolution. This FC-ResNet has a total of 11 millions trainable parameters.

To sum up, our model is composed of a UNet-like model followed by a FC-ResNet. The UNet-like model is composed of 23 convolutional layers and 1.8 million trainable parameters. Our FC-ResNet has 140 convolutional layers and 11 millions of trainable

Table 3
Comparison to published FCN results for EM dataset. V_{rand} is the measure used to rank the submissions. For full ranking of all submitted methods please refer to the challenge website: http://brainiac2.mit.edu/isbi_challenge/leaders-board-new.

Method	V_{rand}	V_{info}	post-proc.	pre-proc.	average over	param.[M]
FusionNet (Quan et al., 2016)	0.978	**0.990**	YES	YES	8	31
CUMedVision (Chen et al., 2016b)	0.977	0.989	YES	NO	6	8
Unet (Ronneberger et al., 2015)	0.973	0.987	NO	YES	7	33
FC-ResNet (Drozdzal et al., 2016)	0.969	0.986	NO	YES	Dropout	11
Ours	**0.981**	0.988	NO	NO	10	13

parameters. Thus, our full segmentation pipeline has 12.8 millions of trainable parameters.

3.3. Dice loss

We train our model using the Dice loss (L_{Dice}) computed per batch:

$$L_{Dice} = -\frac{2\sum_i o_i y_i + 1}{\sum_i o_i + \sum_i y_i + 1}, \quad (2)$$

where $o_i \in [0, 1]$ represents the ith output of the last network layer (sigmoid output), $y_i \in \{0, 1\}$ represents the corresponding ground truth label and 1 is a smoothing factor. Note that the minimum value of the Dice loss is -1. This loss can be seen as an approximation (due to the usage of softmax) to the negative Dice coefficient of the foreground class.

The reason for using the Dice loss over traditional binary crossentropy is two-fold. First, the Dice coefficient is of the common metrics to assess medical image segmentation accuracy, thus, it is natural to optimize it during training. Second, as pointed out in Milletari et al. (2016), Dice loss is well adapted to the problems with high imbalance between foreground and background classes as it does not require any class frequency balancing.

4. Experiments

In this section, we present experimental results of the proposed pipeline for image segmentation. First, we show that our method achieves state-of-the-art results among all published 2D methods on challenging EM benchmark (Arganda-Carreras et al., 2015). Second, we compare our pipeline with standard FCNs (Long et al., 2015; Ronneberger et al., 2015) and with FC-ResNets (Drozdzal et al., 2016) on a dataset of CT scans of liver lesions, with 135 manually annotated scans. Third, we experiment with public MRI benchmarks of prostate and spine segmentations showing competitive results. Fourth, we show the normalization effect of the FCN-based pre-processing module on all three modalities. Finally, using the PROMISE12 MRI dataset we perform an analysis of the influence of having a learnable pre-processor on segmentation results, followed by an evaluation of potential pre-processor architecture designs.

In our experiments, we train with early stopping at the highest Dice value on the validation set with patience of 50 epochs. We implemented our model in the Keras framework (Chollet, 2015) using the Theano backend (Al-Rfou et al., 2016).

4.1. Electron microscopy dataset

The EM training dataset consists of 30 images (512×512 pixels) assembled from serial section transmission electron microscopy of the Drosophila first instar larva ventral nerve cord. The test set is a separate set of 30 images, for which labels are not provided.

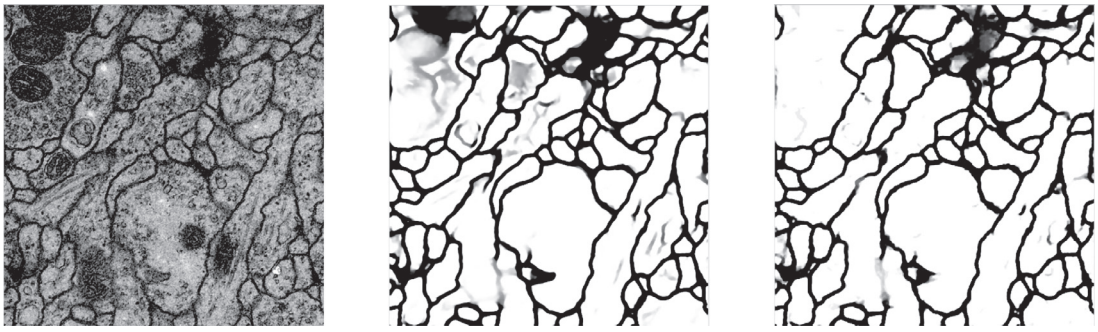
The official metrics used in this dataset are: Maximal foreground-restricted Rand score after thinning (V_{rand}) and maximal foreground-restricted information theoretic score after thinning (V_{info}) with (V_{rand}), being used to order the entries in the leader board. For a detailed description of the metrics, please refer to (Arganda-Carreras et al., 2015).

During training, we augmented the dataset using random flipping (horizontal and vertical), sheering (with maximal range of 0.41), rotations (with maximal range of 25), random cropping (256×256) and spline warping. We used the same spline warping strategy as Ronneberger et al. (2015). Thus, our data augmentation is similar to the one published in Drozdzal et al. (2016). We trained the model with RMSprop (Tieleman and Hinton, 2012) with an initial learning rate of 0.001, a learning rate decay of 0.001 and a batch size of 8. We used weight decay of 0.0001. For each training, the model with the best validation Dice was stored. In total, we trained 10 models and averaged their outputs at the test time. Each time the model was trained, we randomly split 30 images into 24 training images and 6 validation images.

The comparison of our method to other FCNs is shown in Table 3. Our pipeline outperforms all other fully convolutional approaches when looking at the primary metric (V_{rand} score), improving it by 0.003 over the second best fully convolutional approach (FusionNet (Quan et al., 2016)). When comparing our pipeline (FCN pre-processing followed by FC-ResNet) to the pipeline of (Drozdzal et al., 2016) (standardization pre-processing followed by FC-ResNet), we observe an increase in performance. The same happens when comparing our pipeline to FusionNet (Quan et al., 2016) (rescaling pre-processing followed by FC-ResNet). It is worth noting that FusionNet applies intensity shifts with Gaussian noise as data augmentation to account for input variability. For a fair comparison of the proposed method to (Drozdzal et al., 2016), we tested the latter model with the same ensembling strategy as proposed here. While we ensemble 10 models in this work, Drozdzal et al. (2016) ensembled multiple outputs sampled from a model using dropout at test time. We thus obtained a V_{rand} of 0.972, below the 0.981 V_{rand} obtained with the proposed approach (see Table 3). There are two additional fully convolutional approaches submitted for this dataset: UNet (Ronneberger et al., 2015) and CUMedVision (Chen et al., 2016b), the latter being based on FCN8 (Long et al., 2015). The pre-processing used on this dataset varies from range normalization to values from 0 to 1 (Quan et al., 2016) to data standardization (Drozdzal et al., 2016). For post-processing, median filter (Quan et al., 2016) and watershed algorithm (Chen et al., 2016b) are used. All methods use either prediction or model averaging at test time.

Table 4 compares our method to other published entries for EM dataset. As shown in the table, we report the highest V_{rand} score (at the moment of writing the paper), when compared to all published 2D methods. The second best entry is IAL (Beier et al., 2016), which introduces a graph-cut post-processing method to refine FCN predictions. In Beier et al. (2016), a 2D approach was tested resulting in 0.980 V_{rand} score and 0.988 V_{info} score. However, IAL leaderboard entrance incorporates 3D context information and achieves an improvement of 0.003 in V_{rand} score and 0.001 in V_{info} score[1] w.r.t. IAL 2D entry.

[1] Personal communication with the authors.



a) Input image

b) FC-ResNet with dropout at test time Drozdzal et al. (2016)

c) Segmentation result of our pipeline

Fig. 3. Qualitative segmentation results for EM data. (a) An image from the test set, (b) prediction of FC-ResNet with standardization as a pre-processor and dropout at test time and (c) prediction obtained by our pipeline with FCN-based processor followed by FC-ResNet.

Table 4
Comparison to published entries for EM dataset. V_{rand} is the measure used to order the submissions. For full ranking of all submitted methods please refer to the challenge website: http://brainiac2.mit.edu/isbi_challenge/leaders-board-new.

Method	V_{rand}	V_{info}	2D/3D
IAL (Beier et al., 2016)	0.980	0.988	2D
FusionNet (Quan et al., 2016)	0.978	**0.990**	2D
CUMedVision (Chen et al., 2016b)	0.977	0.989	2D
Unet (Ronneberger et al., 2015)	0.973	0.987	2D
IDSIA (Ciresan et al., 2012)	0.970	0.985	2D
motif (Wu, 2015)	0.972	0.985	2D
SCI (Liu et al., 2014)	0.971	0.982	3D
optree-idsia (Uzunbaş et al., 2014)	0.970	0.985	2D
FC-ResNet (Drozdzal et al., 2016)	0.969	0.986	2D
PyraMiD-LSTM (Stollenga et al., 2015)	0.968	0.983	3D
Ours	**0.981**	0.988	2D

Table 5
Results on the liver lesion dataset for both validation and test sets. Loss represents the value of the Dice loss and Dice represents the value of the Dice coefficient on validation and test sets. Note that the loss is computed as a mean of the Dice loss per batch, while the Dice coefficient is computed over the whole validation and test sets.

Method	param. [M]	Validation		Test	
		loss	$Dice_{lesion}$	loss	$Dice_{lesion}$
FCN8 (Long et al., 2015)	128	−0.419	0.589	−0.437	0.535
Unet (Ronneberger et al., 2015)	33	−0.451	0.553	−0.396	0.570
FC-ResNet (Drozdzal et al., 2016)	11	−0.223	0.551	−0.224	0.617
Ours	13	**−0.795**	**0.771**	**−0.796**	**0.711**

Finally, we show some qualitative analysis of our method in Fig. 3, where we display a prediction of a single test frame for two different pipelines: FC-ResNet from (Drozdzal et al., 2016) in Fig. 3(b) and of our pipeline in Fig. 3(c). The white color in the prediction images correspond to the cell class and the black color correspond to the cell membrane class. The different degrees of gray correspond to regions in which the model is (more or less) uncertain about the class assignment. The difference is especially visible when comparing our model to FC-ResNet, we can see that the predictions are sharper and clearer, suggesting that the FCN pre-processor has properly prepared the image for FC-ResNet.

4.2. Liver lesion dataset

Our in-house dataset consists of abdominal contrast-enhanced CT-scans from patients diagnosed with colorectal metastases (CRM). All images are 512×512 with a pixel size varying from 0.53 to 1.25 mm and a slice thickness varying from 0.5 to 5.01 mm. The pixel intensities vary between −3000 and 1500. For each volume, CRM were segmented manually using MITK Workbench (Nolden et al., 2015) by medical students and reviewed by professional image analysts, resulting in 135 CT scans with manually segmented CRM. In addition to that, manual liver segmentation was provided for 58 of the 135 CT scans. This data was collected with the specific goal of segmenting lesions *only* within the liver.

We split the dataset to have 77 training images with CRM segmented, 28 validation images with liver and CRM segmented and 30 test images with liver and CRM segmented. Note that for the training set we did not have liver segmentations available. The liver segmentations of both the validation and the test sets were used to limit the lesion segmentation evaluation only to the liver, treating the rest of the image as a void class.

During the training, we randomly cropped a 2D 128×128 pixel patch containing CRM from a CT scan. We trained the model with RMSprop (Tieleman and Hinton, 2012) with an initial learning rate of 0.001 and a learning rate decay of 0.001. We used weight decay of 0.0001 in the pre-processor and 0.0005 in FC-ResNet.

We follow the same training procedure for all FCN methods: FCN8 (Long et al., 2015), UNet (Ronneberger et al., 2015) and FC-ResNets (Drozdzal et al., 2016). Results are reported in Table 5. All models were trained with the Dice loss with batch size of 20 at training time and 1 at validation and test time. Our approach outperforms other methods on this challenging dataset, achieving the best validation loss of −0.795 and lesion validation Dice of 0.771. The second best validation loss was obtained by the UNet model (−0.451) and the second best lesion validation Dice was obtained by the FCN8 model (0.589). Moreover, our model generalizes well on test set reaching a loss of −0.796 and a Dice of 0.711.

Finally, we show some qualitative results for liver lesion segmentation in Fig. 4, obtained without any user interaction or manual initialization. Fig. 4(a) displays sample input CT images to segment, Fig. 4(b) shows ground truth annotations and Figs. 4(c)–(f) present predictions of FCN8, UNet, FC-ResNet and of our approach respectively. Our approach performs better for all types of lesions: small lesions (see third row in Fig. 4), medium size lesions (see first and second rows in Fig. 4) and large lesions (see forth row in Fig. 4). Furthermore, the lesion segmentation is better adjusted to ground truth annotation, has less false positives and does not have unsegmented holes or gaps within the lesions.

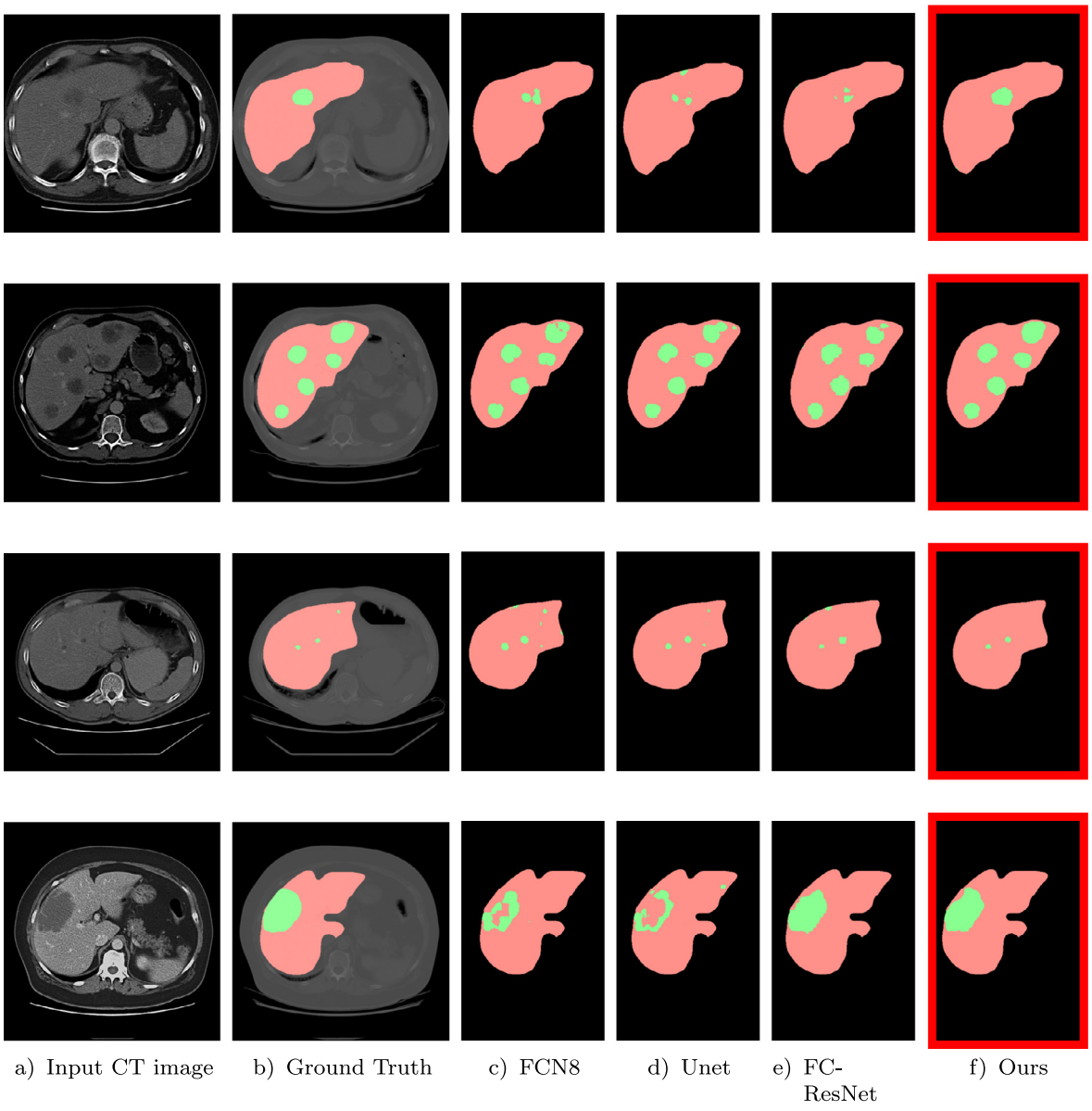


Fig. 4. Qualitative results on test set for the liver lesion dataset. Each line displays an example form the test set. From left to right: (a) represents an image, (b) displays the expert annotation of liver (red) and lesion (green), (c) displays a prediction for FCN8 model, (d) displays a prediction for UNet model, (e) displays a prediction for FC-ResNet model and (f) displays a prediction of our method. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

4.3. Prostate dataset

This experiment tested the segmentation framework on T2-w MR images of the prostate provided by the PROMISE12 challenge.[2] The training dataset contains 50 T2-w MR images of the prostate together with segmentation masks. The test set consists of 30 MR images for which the ground truth is held out by the organizer for independent evaluation. These MRIs are acquired in different hospitals, using different equipments, different acquisition protocols and include both patients with benign disease (e.g. benign prostatic hyperplasia) as well as with prostate cancer. Thus, the dataset variations include: voxel size, dynamic range, position, field of view and anatomic appearance. Contrary to all previously published methods we did not apply any pre-processing step nor volume resizing at training or testing time.

During training, we augmented the dataset using random sheering (with maximal range of 0.1), rotations (with maximal range of 10), random cropping (256×256) and spline warping. We trained the model with RMSprop (Tieleman and Hinton, 2012) with an initial learning rate of 0.0004, a learning rate decay of 0.001 and a batch size of 24. We used weight decay of 0.00001. For each training, the model with the best validation Dice was stored. In total, we trained 10 models and averaged their outputs at the test time. Each time the model was trained, we randomly split 50 training volumes into 40 training images and 10 validation images. Because the method is still based on 2D images, a connected component method was applied on the output to select the largest structure on each volume.

Overall, for our method the Dice coefficient is of 87.4 on the entire gland, with an average boundary distance of 2.17 mm and a volume difference of 12.37%. The comparison to other FCNs on the prostate data is shown in Table 6. In comparison, we use the score provided by the challenge organizer. As it can be seen,

[2] https://grand-challenge.org/site/promise12/home/.

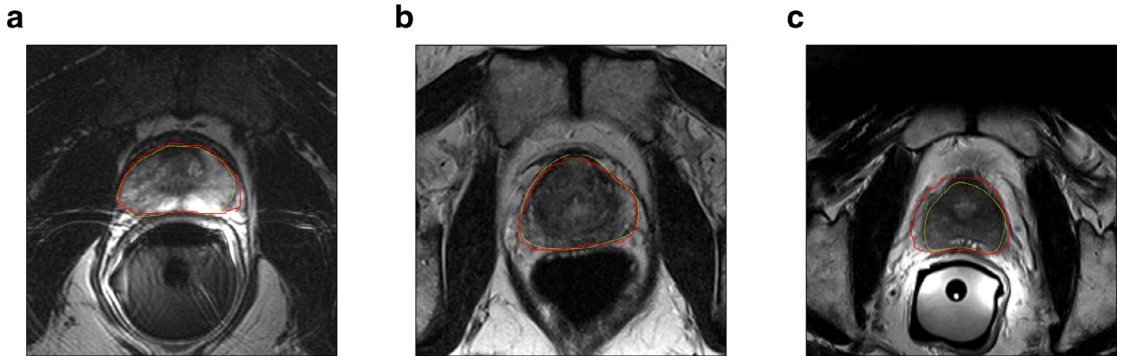


Fig. 5. Qualitative results of our pipeline on the test set. The yellow line shows ground truth annotations while the red line displays our prediction. For more examples of qualitative results please refer to https://grand-challenge.org/site/promise12/resultpro/?id=UdeM2D%folder=20170201000157_2703_UdeM2D_Result. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 6
Comparison to the automatic entries based on FCNs for the prostate dataset. For full ranking of all submitted methods please refer to the challenge website: https://grand-challenge.org/site/promise12/results/.

Method	Score [-]	Dice[%]	Avg. Dist.[mm]	Vol. Diff.[%]
2D FCNs				
Ours	**83.02**	87.4	2.17	12.37
SITUS	79.92	84.13	2.96	23.00
3D FCNs				
CUMED (Yu et al., 2017)	86.65	89.43	1.95	6.95
CAMP-TUM2 (Milletari et al., 2016)	82.39	86.91	2.23	14.98
SRIBHME	74.17	74.46	2.83	34.89

our pipeline outperforms other method based on 2D FCNs and is competitive with methods based on 3D FCNs.

Fig. 5 shows some qualitative results. The prostate segmentations are well adjusted to ground truth annotations and do not have unsegmented holes or gaps within the prostate. Due to the lack of 3D context information, our method, in some cases, is over-segmenting the base of the apex of the prostate (e.g. see Fig. 5(c)). These results rank amongst the best automated approaches for prostate segmentation and without tedious and application specific pre-processing steps.[3]

Finally, please note that PROMISE12 dataset is a multivendor and multisite dataset. The competitive results obtained by our method suggest that a single pre-processor based on a FCN can handle the multivendor and multisite variations in MRI data.

4.4. Spine dataset

The final experiment validated the proposed framework on the publicly available SpineWeb MRI segmentation challenge dataset.[4] This dataset consists of 23 subjects with 2D sagittal T2-weighted MRI sagittal acquisitions of the lower spine (total of 247 2D slices), acquired with a 1.5 Tesla Siemens scanner and resampled to a voxel size of $2 \times 1.25 \times 1.25$ mm^3. For each vertebral body of the spine from T11 to L5 (161 vertebrae in total), a reference manual segmentation was available. Results were evaluated with two-fold cross-validation and compared against two state-of-the-art methods on the same challenge dataset, achieving a Dice score of 92.1 ± 1.4 and an improvement of 3.4% and 1.4% over (Chu et al., 2015) and (Korez et al., 2016) (without the deformable model), respectively.

[3] For full ranking of all submitted methods, please refer to the challenge website: https://grand-challenge.org/site/promise12/results/.

[4] http://spineweb.digitalimaginggroup.ca.

Table 7
Jensen–Shannon divergence for different pre-processing methods over the validation set for EM, CT and MRI modalities. The lower the value, the more similar the distribution of pixels among validation samples.

Modality	Input data	Pre-processor output	Standardization
EM	2.99	2.48	2.98
CT	3.57	2.87	3.35
MRI	5.71	3.38	3.89

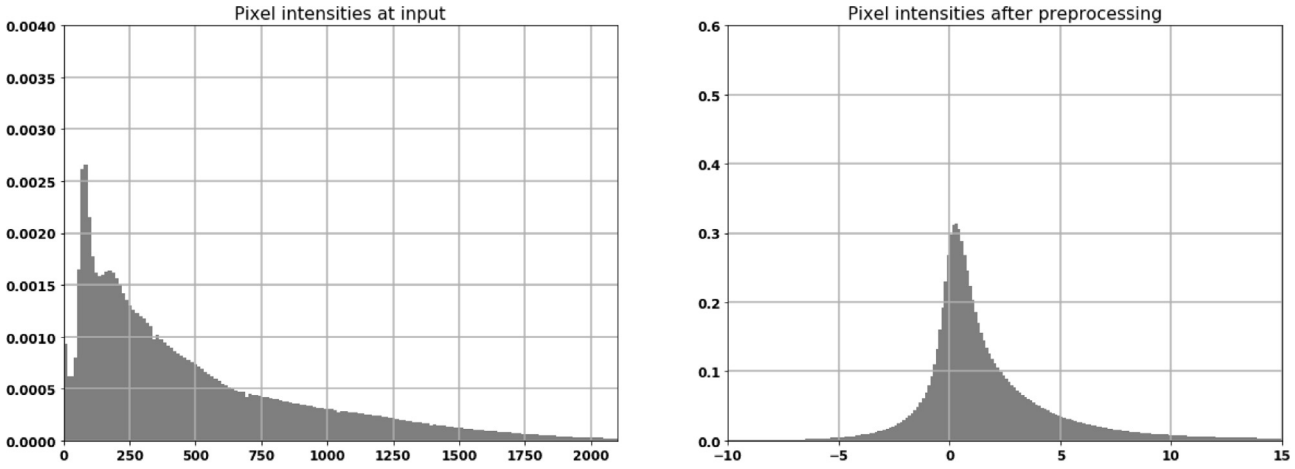
4.5. Data normalization

In this subsection, we provide a more detailed analysis on the trained models. In particular, we investigate the effect of a FCN-based pre-processor on the data distribution. The distribution of validation set pixel intensities at the input of our pipeline (input to FCN-based pre-processor) and at the input of the FC-ResNet are shown in Fig. 6.

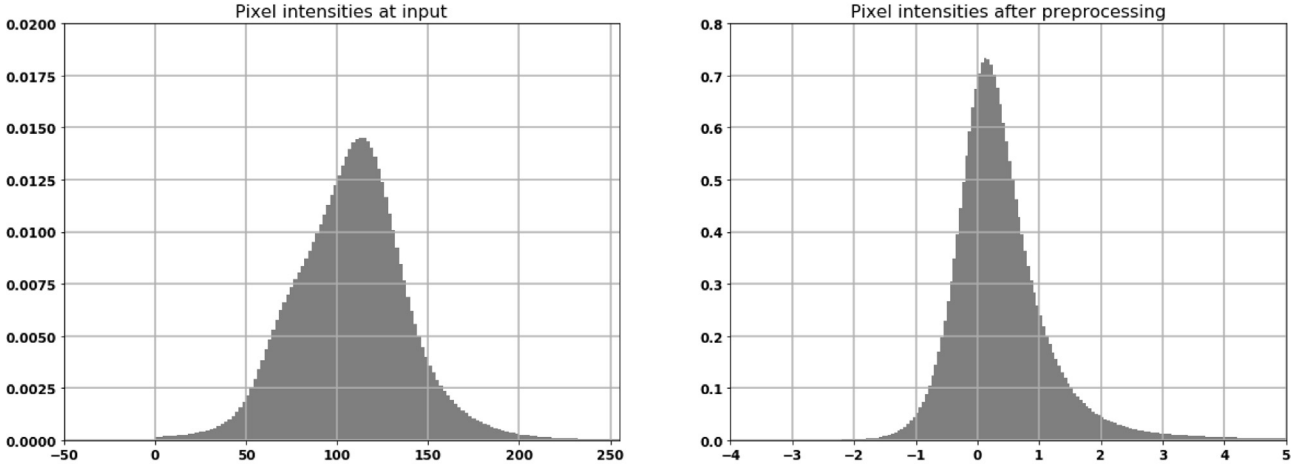
Fig. 6(a) shows the plots for prostate data. We observe that intensities are shifted from [0, 2000] at the FCN input to $[-5, 10]$ at the FC-ResNet input. For liver lesion dataset, we only plot the distributions for liver and lesion, ignoring the distribution of regions outside of the liver. The liver dataset distributions are shown in Fig. 6(b). We notice similar behavior as for the prostate data: intensities are shifted from [0, 200] at the FCN input to $[-2, 3]$ at the FC-ResNet input. Finally, EM data distributions are shown in Fig. 6(c). Please note that the distribution after FCN pre-processing does not resemble distributions that can be achieved by simple data standardization or scaling.

The qualitative evaluation of the FCN-based pre-processor is displayed in Figs. 7–9 for liver, EM data and prostate, respectively. Fig. 7 shows visualizations of how a liver image is transformed from the input of our pre-processor (Fig. 7(a)) to its output (Fig. 7(b)). Analogously, Fig. 7(c) and (d) emphasize how intensities within the liver change, by removing the void pixels of the image.

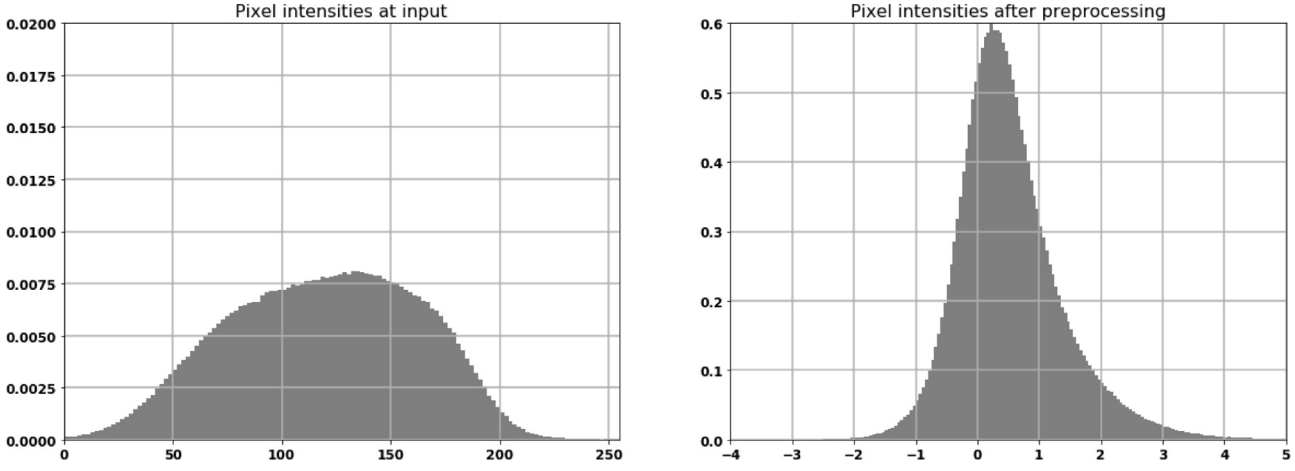
A good data pre-processor should reduce the variation across the intensity distributions of all input samples. For example, data standardization encourages the intensity distribution of each data point to represent standard normal distribution. One way to evaluate distance between two distribution is to compute the Jensen-Shannon (JS) distance between each pair of intensity histograms. Table 7 represents the mean JS distance across every pair of image intensity distributions in the validation set for all three modalities. The lower the mean divergence score, the more similar the intensity distributions among samples. In Table 7, we report the mean JS scores for the input data, for the feature maps at the output of the learnable pre-processor as well as for a standardized version of the input data. Our approach to data normalization improves



a) Prostate dataset

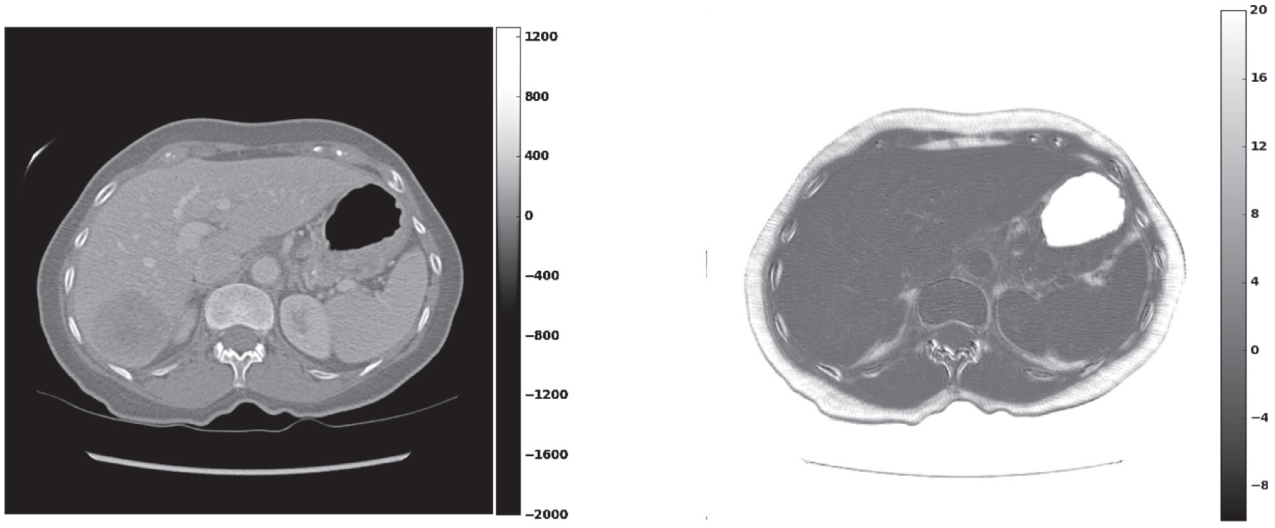


b) Liver lesion dataset



c) EM dataset

Fig. 6. Intensity distribution histograms for (a) Prostate dataset, (b) liver lesion dataset and (c) EM dataset. For each dataset, the plots represent the distribution of input pixels and the distribution after FCN pre-processing.



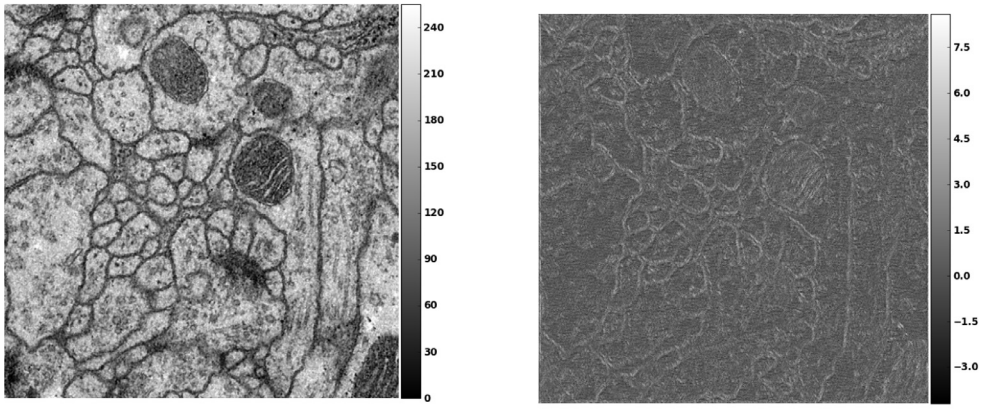
a) Input image

b) Pre-processor output

c) Image cropped to liver

d) Pre-processor output cropped to liver

Fig. 7. Visualization of the output of the pre-processor module for liver lesion dataset.



a) Input image

b) Pre-procesor output

Fig. 8. Visualization of the output of the pre-processor module for the EM dataset.

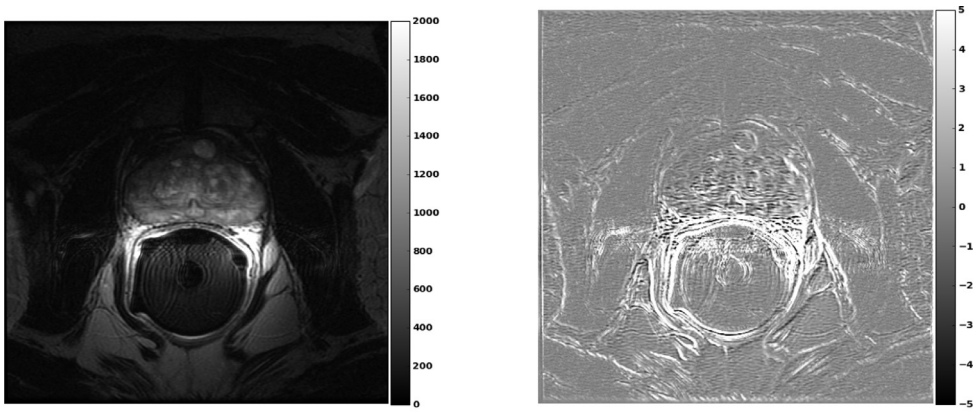


Fig. 9. Visualization of the output of the pre-processor module for the Promise12 dataset.

Table 8
Results of 5 fold cross validation on the PROMISE12 dataset using different segmentation pipelines. The numbers report the Dice coefficient for prostate class together with the standard deviation.

Model	Dice coefficient
Baselines	
FC-ResNetBig (no pre-processor)	56.7 ± 24.9
FC-ResNet (standardization)	78.8 ± 7.3
Learnable pre-processors	
FCN + FC-ResNet	82.4 ± 6.3
Shallow-FCN + FC-ResNet	80.2 ± 9.8
Context module + FC-ResNet	65 ± 25
BN-FCN + FC-ResNet	80.3 ± 13.4
FCN(2) + FC-ResNet	83.9 ± 11.1
FCN(16) + FC-ResNet	83.3 ± 6.5

the JS score for all modalities and outperforms standard data normalization techniques such as subtracting the mean and dividing by the standard deviation of the intensity on a per slice basis.

4.6. Preprocessor analysis

In this subsection, we provide an in depth analysis of the role of the learnable pre-processor compared to baselines. In addition to that, we assess the performance achieved by different pre-processor architectures. All experiments were performed using 5-fold cross-validation on the PROMISE12 dataset. Results are reported in Table 8.

We ran two baseline experiments. The first one (FC-ResNetBig) increases the capacity of the FC-ResNet to match the capacity of our pipeline (FCN + FC-ResNet). As reported in the table, the performance achieved by FC-ResNetBig is 56.7 ± 24.9. This result is significantly worse than the one obtained by FCN + FC-ResNet (82.4 ± 6.3). It is worth noting that FC-ResNetBig is not able to segment some of the PROMISE12 difficult cases, leading to poor results and high standard deviation. This suggests that the good performance of our pipeline is more related to the model design than the number of parameters. In the second experiment, we complement the FC-ResNet model with a per volume standardization in order to compare the performance achieved by different pre-processing techniques in conjunction with FC-ResNet. As reported in the table, FC-ResNet with standardization achieves a Dice coefficient of 78.8 ± 7.3. This result is lower than the one obtained with the learnable pre-processor, suggesting that a learnable pre-processor is better suited to work with FC-ResNet than per volume standardization.

A natural question arises regarding the pre-processor model: how important is the pre-processor architecture design? To address this question, we ran a set of experiments with varying pre-processor architectures. In the first experiment, we reduced the depth of the FCN pre-processor by removing one pooling and one upsampling operation together with the corresponding convolutions. We observe that reducing the depth of the pre-processor hurts the performance. In the second experiment, we used the context module of (Yu and Koltun, 2015) as a pre-processor. This module is built such that the resolution from the input to the output is maintained (no downsampling or upsampling operations) and the context is increased by means of repeated dilated convolutions. We found this architecture to obtain non-competitive results when used as a pre-processor on PROMISE12. Similarly to FC-ResNetBig, this architecture failed to segment some of the dataset's difficult cases, resulting in a high standard deviation. In the third experiment, we implemented a FCN model with batch normalization (BN) layers by following a pattern of convolution-BN-ReLU instead of convolution-ReLU (like in the proposed FCN pre-processor), reaching a Dice score of 80.3. Finally, in the last experiment, we increased the number of feature maps at the intersection between FCN and FC-ResNet. Note that when increasing the number of feature maps at the output of the FCN, the one to one mapping is lost, and thus, it should not be interpreted as a pre-processor. However, if the objective of the pipeline is to maximize the segmentation results, one can treat the number of feature maps as a hyperparameter and optimize it on the validation set. In our experiments, we observed a slight increase in the cross-validation Dice score when increasing the number of feature maps at the intersection between the models. We tried 3 different setups for the number of feature maps at the intersection: FCN with 1 feature map, FCN(2) with 2 feature maps and FCN(16) with 16 feature maps. The results are presented in Table 8. The Dice score is highest for 2 feature maps (83.9), while the model is most stable when having 1 feature map at the intersection (std of 6.3).

5. Discussion and conclusion

In this paper, we have introduced a simple, yet powerful segmentation pipeline for medical images that combines fully convolutional networks with fully convolutional ResNets. Our pipeline is built from a low-capacity FCN model followed by a very deep FC-ResNet (more than 100 layers). We have highlighted the importance of pre-processing when using FC-ResNets and shown that a low-capacity FCN model can serve as a pre-processor to normalize raw medical image data. We argued that FC-ResNets

are better complemented by the proposed FCN pre-processor than by traditional pre-processors, given the normalization they achieve. Finally, we have shown that using this pipeline we exhibit state-of-the-art performance on the challenging EM benchmark, improve segmentation results on our in-house liver lesion dataset, when compared to standard FCN methods and yield competitive results on challenging 3D MRI prostate segmentation task. These results illustrate the strong potential and versatility of the framework by achieving accurate segmentations on multi-modality images from different anatomical regions and organs.

The fact that FC-ResNets require some kind of data pre-processing is not very surprising, given the model construction. The identity path forwards the input data right to the output, while applying small transformations along the residual path. These small transformations are controlled by batch normalization. This input-forwarding especially affects medical imaging data, where the pixel intensities range is much broader than in standard RGB images. Therefore, adequate pre-processing becomes crucial for FC-ResNets to achieve significant improvement over standard FCN approaches, especially for some imaging modalities with greater variability in acquisition protocols. Standard FCNs do not contain the identity path and, thus, are more robust to the input data distribution. However, the lack of this identity path affects the optimization process and limits the depth of the tested models. Moreover, one cannot rule out the possibility that a FCN-based pre-processor complemented by an additional meta-pre-processor could perform even better.

In light of recent advancements in understanding both CNNs and ResNets, our pipeline can have two possible interpretations. On one hand, the pipeline could be explained as having the FCN model working as a pre-processor followed by the FC-ResNet model performing the task of an ensemble of relatively shallow models, leading to a robust classifier. On the other hand, the interpretation of our pipeline could revolve around the iterative inference point of view of ResNets (Greff et al., 2017). In this scenario, the role of the FCN would be to produce an input proposal that would be iteratively refined by the FC-ResNet to generate the proper segmentation map. It is worth noting that, in both interpretations, FC-ResNets should be relatively deep (hundreds of layers) in order to take full advantage of the ensemble of shallow networks or iterative refinement of the initial proposal. FC-ResNets might not have been deep enough in many medical image segmentation pipelines to achieve these effects.

Potential future direction might involve experimentation with different variants of architectures that could serve as a pre-processor. This architecture exploration should not be limited to FCN-like models. From medical image segmentation perspective, the model could potentially benefit by expanding it to 3D FCN.

Acknowledgments

We would like to thank all the developers of Theano and Keras. We gratefully acknowledge NVIDIA for GPU donation. The authors would like to thank Mohammad Havaei and Nicolas Chapados for insightful discussions. The authors would like to thank Drs Simon Turcotte, Réal Lapointe, Franck Vandenbroucke-Menu and Ms. Louise Rousseau from the CHUM Colorectal, Hepato-Pancreato-Biliary Cancer Biobank and Database, supported by the Université de Montréal Roger DesGroseillers Hepato-Pancreato-Biliary Surgical Oncology Research Chair, for enabling the selection of consenting patients with diagnostic imaging available for this study. This work was partially funded by Imagia Inc., MITACS (grant number IT05356) and MEDTEQ. An Tang was supported by a research scholarship from the Fonds de Recherche du Québec en Santé and Fondation de l'association des radiologistes du Québec (FRQS-ARQ #26993).

References

Al-Rfou, R., Alain, G., Almahairi, A., et al., 2016. Theano: a Python framework for fast computation of mathematical expressions. CoRR. arXiv:1605.02688.

Andermatt, S., Pezold, S., Cattin, P., 2016. Multi-dimensional gated recurrent units for the segmentation of biomedical 3d-data. In: Carneiro, G., Mateus, D., Peter, L., et al. (Eds.), Deep Learning and Data Labeling for Medical Applications, In Conjunction with MICCAI 2016, Athens, Greece, October 21, 2016, Proceedings. Springer International Publishing, Cham, pp. 142–151.

Arganda-Carreras, I., Seung, H. S., Vishwanathan, A., Berger, D. R., 2013. ISBI 2013 challenge: 3d segmentation of neurites in EM images. http://brainiac2.mit.edu/SNEMI3D/home.

Arganda-Carreras, I., Turaga, S.C., Berger, D.R., et al., 2015. Crowdsourcing the creation of image segmentation algorithms for connectomics. Front. Neuroanat. 9 (142). doi:10.3389/fnana.2015.00142.

Beier, T., Andres, B., Köthe, U., Hamprecht, F.A., 2016. An efficient fusion move algorithm for the minimum cost lifted multicut problem. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (Eds.), Computer Vision – ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part II. Springer International Publishing, Cham, pp. 715–730. doi:10.1007/978-3-319-46475-6_44.

Ben-Cohen, A., Diamant, I., Klang, E., et al., 2016. Fully convolutional network for liver segmentation and lesions detection. In: Deep Learning and Data Labeling for Medical Applications, In Conjunction with MICCAI 2016, Athens, Greece, October 21, 2016, Proceedings. Springer International Publishing, Cham, pp. 77–85.

Birenbaum, A., Greenspan, H., 2016. Longitudinal multiple sclerosis lesion segmentation using multi-view convolutional neural networks. In: Deep Learning and Data Labeling for Medical Applications, In Conjunction with MICCAI 2016, Athens, Greece, October 21, 2016, Proceedings. Springer International Publishing, Cham, pp. 58–67.

Casamitjana, A., Puch, S., Aduriz, A., Sayrol, E., Vilaplana, V., 2016. 3d convolutional networks for brain tumor segmentation. In: Proceedings of the MICCAI Challenge on Multimodal Brain Tumor Image Segmentation (BRATS) 2016, pp. 65–68.

Cha, K.H., Hadjiiski, L., Samala, R.K., et al., 2016. Urinary bladder segmentation in ct urography using deep-learning convolutional neural network and level sets. Med. Phys. 43 (4), 1882–1896.

Chang, P.D., 2016. Fully convolutional neural networks with hyperlocal features for brain tumor segmentation. In: Proceedings of the MICCAI Challenge on Multimodal Brain Tumor Image Segmentation (BRATS) 2016, pp. 4–9.

Chen, H., Dou, Q., Yu, L., Heng, P., 2016a. VoxResNet: deep voxelwise residual networks for volumetric brain segmentation. CoRR. arXiv:1608.05895.

Chen, H., Qi, X., Cheng, J., Heng, P.A., 2016b. Deep contextual networks for neuronal structure segmentation. In: Proceedings of the 13th AAAI Conference on Artificial Intelligence, February 12–17, 2016, Phoenix, Arizona, USA., pp. 1167–1173.

Chollet, F., 2015. Keras. https://github.com/fchollet/keras.

Christ, P.F., Elshaer, M.E.A., Ettlinger, F., et al., 2016. Automatic liver and lesion segmentation in ct using cascaded fully convolutional neural networks and 3d conditional random fields. In: Ourselin, S., Joskowicz, L., Sabuncu, M.R., Unal, G., Wells, W. (Eds.), MICCAI 2016: 19th International Conference, Athens, Greece, October 17–21, 2016, Proceedings, Part II. Springer International Publishing, Cham, pp. 415–423. doi:10.1007/978-3-319-46723-8_48.

Chu, C., Belav, D.L., Armbrecht, G., Bansmann, M., Felsenberg, D., Zhen, G., 2015. Fully automatic localization and segmentation of 3d vertebral bodies from ct/mr images via a learning-based method. PLOS One 10. e0143327.

Ciresan, D., Giusti, A., Gambardella, L.M., Schmidhuber, J., 2012. Deep neural networks segment neuronal membranes in electron microscopy images. In: NIPS 25. Curran Associates, Inc., pp. 2843–2851.

Dou, Q., Chen, H., Jin, Y., Yu, L., Qin, J., Heng, P.-A., 2016. 3d deeply supervised network for automatic liver segmentation from CT volumes. In: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016. Springer International Publishing, pp. 149–157. doi:10.1007/978-3-319-46723-8_18.

Drozdzal, M., Vorontsov, E., Chartrand, G., Kadoury, S., Pal, C., 2016. The importance of skip connections in biomedical image segmentation. In: Carneiro, G., Mateus, D., Peter, L., Bradley, A., Tavares, J.M.R.S., Belagiannis, V., Papa, J.P., Nascimento, J.C., Loog, M., Lu, Z., Cardoso, J.S., Cornebise, J. (Eds.), Deep Learning and Data Labeling for Medical Applications: First International Workshop, LABELS 2016, and Second International Workshop, DLMIA 2016, Held in Conjunction with MICCAI 2016, Athens, Greece, October 21, 2016, Proceedings. Springer International Publishing, Cham, pp. 179–187. doi:10.1007/978-3-319-46976-8_19.

Fakhry, A., Zeng, T., Ji, S., 2016. Residual deconvolutional networks for brain electron microscopy image segmentation. IEEE Trans. Med. Imaging PP (99). 1–1. doi:10.1109/TMI.2016.2613019.

Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep Learning. MIT Press. http://www.deeplearningbook.org.

Greff, K., Srivastava, R.K., Schmidhuber, J., 2017. Highway and residual networks learn unrolled iterative estimation. ICLR -.

Havaei, M., Davy, A., Warde-Farley, D., et al., 2017. Brain tumor segmentation with deep neural networks. Med. Image Anal. 35, 18–31. http://doi.org/10.1016/j.media.2016.05.004.

Havaei, M., Guizard, N., Larochelle, H., Jodoin, P., 2016. Deep learning trends for focal brain pathology segmentation in MRI. In: Machine Learning for Health Informatics - State-of-the-Art and Future Challenges, pp. 125–148.

He, K., Zhang, X., Ren, S., Sun, J., 2016a. Deep residual learning for image recognition. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

He, K., Zhang, X., Ren, S., Sun, J., 2016b. Identity mappings in deep residual networks. In: Computer Vision - ECCV 2016 - 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part IV, pp. 630–645. doi:10.1007/978-3-319-46493-0_38.

Hu, P., Wu, F., Peng, J., Liang, P., Kong, D., 2016. Automatic 3d liver segmentation based on deep learning and globally optimized surface evolution. Phys. Med. Biol. 61 (24), 8676–8698. doi:10.1088/1361-6560/61/24/8676.

Huang, G., Sun, Y., Liu, Z., et al., 2016. Deep networks with stochastic depth. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (Eds.), Computer Vision – ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part IV. Springer International Publishing, Cham, pp. 646–661. doi:10.1007/978-3-319-46493-0_39.

Ioffe, S., Szegedy, C., 2015. Batch normalization: accelerating deep network training by reducing internal covariate shift. In: Blei, D., Bach, F. (Eds.), Proceedings of the 32nd ICML (ICML-15). JMLR Workshop and Conference Proceedings, pp. 448–456.

Kamnitsas, K., Ledig, C., Newcombe, V.F., et al., 2017. Efficient multi-scale 3d {CNN} with fully connected {CRF} for accurate brain lesion segmentation. Med. Image Anal. 36, 61–78. http://doi.org/10.1016/j.media.2016.10.004.

Korez, R., Likar, B., Pernus, F., Vrtovec, T., 2016. Model-based segmentation of vertebral bodies from mr images with 3d CNNs. In: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016. Springer International Publishing, pp. 433–441. doi:10.1007/978-3-319-46723-8_50.

Krähenbühl, P., Koltun, V., 2011. Efficient inference in fully connected CRFs with Gaussian edge potentials. In: Shawe-Taylor, J., Zemel, R.S., Bartlett, P.L., Pereira, F., Weinberger, K.Q. (Eds.), Advances in NIPS 24. Curran Associates, Inc., pp. 109–117.

Li, W., Jia, F., Hu, Q., 2015. Automatic segmentation of liver tumor in ct images with deep convolutional neural networks. J. Comput. Commun. 3 (11), 146–151.

Liao, Q., Poggio, T.A., 2016. Bridging the gaps between residual learning, recurrent neural networks and visual cortex. CoRR. arXiv:1604.03640.

Liu, T., Jones, C., Seyedhosseini, M., Tasdizen, T., 2014. A modular hierarchical approach to 3d electron microscopy image segmentation. J. Neurosci. Methods 226, 88–102.

Long, J., Shelhamer, E., Darrell, T., 2015. Fully convolutional networks for semantic segmentation. CVPR.

Lun, T.K., Hsu, W., 2016. Brain tumor segmentation using deep convolutional neural network. In: Proceedings of the MICCAI Challenge on Multimodal Brain Tumor Image Segmentation (BRATS) 2016, pp. 26–29.

McKinley, R., Wiest, R., Reyes, M., 2016. Nabla-net: a deep dag-like convolutional architecture for biomedical image segmentation: application to high- and low–grade glioma segmentation. In: Proceedings of the MICCAI Challenge on Multimodal Brain Tumor Image Segmentation (BRATS) 2016, pp. 1–4.

Milletari, F., Navab, N., Ahmadi, S., 2016. V-net: fully convolutional neural networks for volumetric medical image segmentation. In: Fourth International Conference on 3D Vision, 3DV 2016, Stanford, CA, USA, October 25–28, 2016, pp. 565–571.

Nolden, M., Zelzer, S., Seitel, A., Wald, D., et al., 2015. The medical imaging interaction toolkit (MITK). http://mitk.org/wiki/The_Medical_Imaging_Interaction_Toolkit_(MITK).

Pandian, B., Boyle, J., Orringer, D.A., 2016. Multimodal tumor segmentation with 3d volumetric convolutional neural networks. In: Proceedings of the MICCAI Challenge on Multimodal Brain Tumor Image Segmentation (BRATS) 2016, pp. 1–4.

Quan, T. M., Hildebrand, D. G. C., Jeong, W.-K., 2016. FusionNet: a deep fully residual convolutional neural network for image segmentation in connectomics. ArXiv e-prints.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (Eds.), MICCAI 2015: 18th International Conference, Munich, Germany, October 5–9, 2015, Proceedings, Part III. Springer International Publishing, Cham, pp. 234–241. doi:10.1007/978-3-319-24574-4_28.

Roth, H.R., Lu, L., Farag, A., et al., 2015. Deeporgan: multi-level deep convolutional networks for automated pancreas segmentation. CoRR. arXiv:1506.06448.

Srivastava, N., Hinton, G., Krizhevsky, A., et al., 2014. Dropout: a simple way to prevent neural networks from overfitting. J. Mach. Learn. Res. 15 (1), 1929–1958.

Stollenga, M.F., Byeon, W., Liwicki, M., Schmidhuber, J., 2015. Parallel multi-dimensional LSTM, with application to fast biomedical volumetric image segmentation. In: Proceedings of the 28th NIPS. MIT Press, Cambridge, MA, USA, pp. 2998–3006.

Styner, M., Lee, J., Chin, B., et al., 2008. 3d segmentation in the clinic: a grand challenge II: MS lesion segmentation.

Szegedy, C., Liu, W., Jia, Y., et al., 2015. Going deeper with convolutions. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1–9. doi:10.1109/CVPR.2015.7298594.

Thong, W., Kadoury, S., Pich, N., Pal, C.J., 2016. Convolutional networks for kidney segmentation in contrast-enhanced CT scans. Comput Methods Biomech. Biomed. Eng. 1–6.

Tieleman, T., Hinton, G., 2012. Lecture 6.5—RmsProp: divide the gradient by a running average of its recent magnitude. COURSERA, Neural Networks for Machine Learning.

Uzunbaş, M.G., Chen, C., Metaxsas, D., 2014. Optree: a learning-based adaptive watershed algorithm for neuron segmentation. In: Golland, P., Hata, N., Barillot, C., Hornegger, J., Howe, R. (Eds.), MICCAI 2014: 17th International Conference, Boston, MA, USA, September 14–18, 2014, Proceedings, Part I. Springer International Publishing, Cham, pp. 97–105. doi:10.1007/978-3-319-10404-1_13.

Veit, A., Wilber, M.J., Belongie, S.J., 2016. Residual networks behave like ensembles of relatively shallow networks. In: Advances in NIPS 29: Annual Conference on NIPS 2016, December 5–10, 2016, Barcelona, Spain, pp. 550–558.

Vivanti, R., Ephrat, A., Joskowicz, L., et al., 2015. Automatic liver tumor segmentation in follow-up CT scans: preliminary method and results. In: Patch-MI 2015, Held in Conjunction with MICCAI 2015, Munich, Germany, October 9, 2015, pp. 54–61.

Wu, X., 2015. An iterative convolutional neural network algorithm improves electron microscopy image segmentation. CoRR. arXiv:1506.05849.

Yu, F., Koltun, V., 2015. Multi-scale context aggregation by dilated convolutions. CoRR. arXiv:1511.07122.

Yu, L., Yang, X., Chen, H., Qin, J., Heng, P.-A., 2017. Volumetric convnets with mixed residual connections for automated prostate segmentation from 3d mr images. In: Proceedings of AAAI.

Zeiler, M.D., Fergus, R., 2014. Visualizing and understanding convolutional networks. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (Eds.), Computer Vision – ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6–12, 2014, Proceedings, Part I. Springer International Publishing, Cham, pp. 818–833. doi:10.1007/978-3-319-10590-1_53.

Zhao, X., Wu, Y., Song, G., et al., 2016. Brain tumor segmentation using a fully convolutional neural network with conditional random elds. In: Proceedings of the MICCAI Challenge on Multimodal Brain Tumor Image Segmentation (BRATS) 2016, pp. 77–80.

Zheng, Y., Liu, D., Georgescu, B., Xu, D., Comaniciu, D., 2016. Deep learning based automatic segmentation of pathological kidney in CT: local vs. global image context. In: Lu, L., Zheng, Y., Carneiro, G., Yang, L. (Eds.), Deep Learning and Convolutional Neural Networks for Medical Image Computing. Springer.