Abdominal multi-organ segmentation with cascaded convolutional and adversarial deep networks

Pierre-Henri Conze, A. Emre Kavur, Emilie Cornec-Le Gall, N. Sinem Gezer, Yannick Le Meur, M. Alper Selver, François Rousseau

Version of Record: https://www.sciencedirect.com/science/article/pii/S0933365721001020
Manuscript_1ea8f32f5d80c0d35b8b9f8f99d8b0bd

Abdominal multi-organ segmentation with cascaded convolutional and adversarial deep networks

Pierre-Henri Conze[a,b,*], Ali Emre Kavur[c], Emilie Cornec-Le Gall[d,e], Naciye Sinem Gezer[c,f], Yannick Le Meur[d,g], M. Alper Selver[c], François Rousseau[a,b]

[a] *IMT Atlantique, Technopôle Brest-Iroise, 29238 Brest, France*
[b] *LaTIM UMR 1101, Inserm, 22 avenue Camille Desmoulins, 29238 Brest, France*
[c] *Dokuz Eylul University, Cumhuriyet Bulvarı, 35210 Izmir, Turkey*
[d] *Department of Nephrology, University Hospital, 2 avenue Foch, 29609 Brest, France*
[e] *UMR 1078, Inserm, 22 avenue Camille Desmoulins, 29238 Brest, France*
[f] *Department of Radiology, Faculty of Medicine, Cumhuriyet Bulvarı, 35210 Izmir, Turkey*
[g] *LBAI UMR 1227, Inserm, 5 avenue Foch, 29609 Brest, France*

Abstract

Abdominal anatomy segmentation is crucial for numerous applications from computer-assisted diagnosis to image-guided surgery. In this context, we address fully-automated multi-organ segmentation from abdominal CT and MR images using deep learning. The proposed model extends standard conditional generative adversarial networks. Additionally to the discriminator which enforces the model to create realistic organ delineations, it embeds cascaded partially pre-trained convolutional encoder-decoders as generator. Encoder fine-tuning from a large amount of non-medical images alleviates data scarcity limitations. The network is trained end-to-end to benefit from simultaneous multi-level segmentation refinements using auto-context. Employed for healthy liver, kidneys and spleen segmentation, our pipeline provides promising results by outperforming state-of-the-art encoder-decoder schemes. Followed for the Combined Healthy Abdominal Organ Segmentation (CHAOS) challenge organized in conjunction with the IEEE International Symposium on Biomedical Imaging 2019, it gave us the first rank for three competition categories: liver CT, liver MR and multi-organ MR segmentation. Combining cascaded convolutional and adversarial networks strengthens the ability of deep learning pipelines to automat-

*corresponding author: pierre-henri.conze@imt-atlantique.fr

ically delineate multiple abdominal organs, with good generalization capability. The comprehensive evaluation provided suggests that better guidance could be achieved to help clinicians in abdominal image interpretation and clinical decision making.

Keywords: multi-organ segmentation, convolutional encoder-decoders, adversarial learning, cascaded networks, abdominal images

1. Introduction

The development of non-invasive imaging technologies over the last decades has opened new horizons in studying abdominal anatomical structures. Segmentation has become a crucial task in abdominal image analysis with numerous applications including computer-assisted diagnosis, surgery planning (e.g. organ pre-evaluation for resection or transplantation), visual augmentation, extraction of quantitative indices or image-guided interventions [1]. In particular, the precise delineation of abdominal solid visceral organs including liver, kidneys and spleen for localization, volume assessment or follow-up purposes has critical importance. However, the analysis of Computed Tomography (CT) and Magnetic Resonance (MR) abdominal imaging datasets is challenging and time-consuming for clinicians since the abdomen is a complex body space. Robust automatic abdominal image segmentation is required to guide image interpretation, facilitate clinical decision making and improve patient care while avoiding traditional manual delineation efforts.

In this area, many interactive, semi- and fully-automated methods have been proposed with diverse methodologies including statistical shape models [3], multi-atlas segmentation [4] or machine learning [5] techniques. More recently, outstanding performance has been reached in almost all medical image analysis tasks using deep learning [6]. Despite the large variability in abdominal organ shape, size, location and texture, abdominal multi-organ segmentation has naturally benefited from this massive trend [7, 8, 9, 10]. Compared to conventional machine learning, the need for hand-crafted features no longer remains neces-

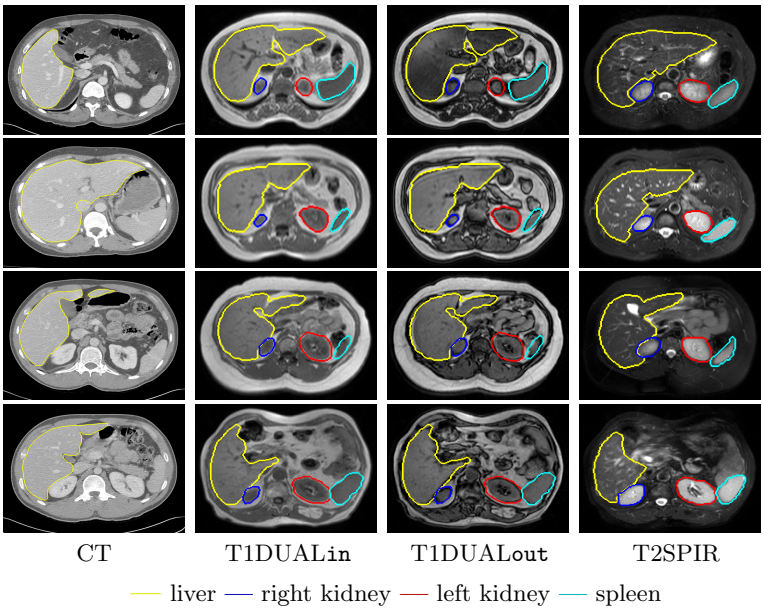

Figure 1: Samples of healthy abdominal CT and MR (T1-DUAL`in/out`, T2-SPIR) images arising from the CHAOS dataset [2], provided with ground truth organ delineations.

sary. In particular, huge efforts have been devoted to automatic segmentation based on variants of Fully Convolutional Networks (FCN) [11]. Recent architectures comprise a regular FCN to extract multi-scale features, followed by an up-sampling branch that enables to recover the input resolution through up-convolutions [6]. In the medical image processing community, UNet [12] is one of the most well-known approach among such Convolutional Encoder-Decoders (CED). Able to learn from relatively small datasets, CED architectures are the most likely to automatically infer high-level knowledge involved by radiologists when interpreting abdominal images.

Despite intensive developments in deep learning, it remains difficult to judge the effectiveness of deep networks for abdominal multi-organ segmentation since they are mainly assessed on one single organ only (liver most often), one single modality (usually CT) and/or relatively small and private datasets. Their robustness to delineate multiple abdominal structures from different modalities and to manage strong inter-subject variability is therefore under-investigated. Rather than organ or modality-specific strategies, the development of more com-

prehensive and generic computational models is needed [13]. Few challenges including the Combined Healthy Abdominal Organ Segmentation (CHAOS) challenge[1] [2], organized in conjunction with the IEEE International Symposium on Biomedical Imaging (ISBI) 2019, has been proposed to motivate further work on this perspective by making available a dataset (Fig.1) to segment multiple organs from two imaging modalities (CT, MR with T1-DUAL and T2-SPIR sequences) acquired for unpaired healthy subjects. Towards efficient combined segmentation and based on this unique dataset, we target robust and generic deep learning architectures for two main purposes: 1- segmentation of liver from CT scans and 2- segmentation of four abdominal organs (liver, right kidney, left kidney, spleen) from MR images.

The proposed healthy abdominal multi-organ segmentation methodology comprises three key aspects. First, deeper CED architectures using encoders pre-trained on non-medical data and extending the UNet [12] baseline are investigated. Second, we embed this architecture into a cascaded framework using auto-context and end-to-end training to benefit from simultaneous multi-level segmentation refinements. Third, such cascaded pipeline is used as generator within a conditional Generative Adversarial Network (cGAN). The resulting model thus includes a discriminator to strengthen the ability of the generative part to create delineations as realistic as possible. The step-by-step evaluation provided for each contribution in both CT and MR modalities highlights better performance than state-of-the-art encoder-decoder schemes. The pipeline also gave us the first rank for three CHAOS competition categories[2] (liver CT, liver MR and multi-organ MR segmentation) [2] which suggests that the proposed computational deep models can offer new insights for abdominal image interpretation and clinical decision making, in various computer-assisted tasks.

[1] https://chaos.grand-challenge.org
[2] https://chaos.grand-challenge.org/results_CHAOS/

2. Related works

Computational abdominal organ segmentation has attracted considerable attention over the last decades. This craze led to the development of a wide range of methodologies, from interactive to semi- and fully-automated [14]. Before the recent development of machine and deep learning, abdominal organ segmentation has often been carried out using statistical shape models [15, 3] to capture and then fit organ shapes through anatomical correspondences. Since deformations and limited datasets may prevent those models from managing the strong variability of abdominal organ shapes, aligning and merging manually segmented images could be followed as an alternative. Specifically, multi-atlas segmentation consists in leveraging label atlases through image registration and statistical fusion [16]. Applied to abdominal data, coarse-to-fine [17], region-wise local atlas selection [18], Selective and Iterative Method for Performance Level Estimation (SIMPLE) [4, 19] or dictionary learning and sparse coding [20] techniques can be employed to alleviate substantial registration errors. Nevertheless, robust inter-subject abdominal image registration is a challenging, computational intense and not yet solved issue [21] due to the diversity of organ shape, size, location and texture. This mainly explains the success of registration-free methods whose aim is to learn feature distributions that characterize abdominal anatomy from un-registered images.

Among registration-free methods, computational power and data availability have enabled the rise of machine learning techniques *via* voxel- [5, 22], patch- [23] or supervoxel-wise [24] classifiers. These methods require hand-crafted features and therefore, specialized knowledge to delineate structures from medical images. Conversely, deep Convolutional Neural Networks (CNN), data-driven learning models formed by multi-layer neural networks, automatically learn complex hierarchical features from data [25]. In this direction, huge efforts have been devoted to automatic segmentation based on variants of Fully Convolutional Networks (FCN) [11]. Further improvements are reached with architectures comprising a regular FCN to extract features, followed by an up-sampling

part which recovers the input resolution using up-convolutions [6]. UNet [12] and its 3D counterparts [26, 27] are among the most well-known Convolutional Encoder-Decoders (CED) in the medical community. They exploit long-range shortcuts to concatenate features between contracting and expanding paths for improving localization accuracy while allowing faster convergence.

CED networks have been widely adopted for automatic abdominal organ segmentation, as in [7] where 3D UNet [27] is exploited in a two-stage hierarchical fashion for multi-organ delineation purposes. Combining densely linked layers and shallow 3D UNet architecture [8] enables high-resolution activation maps through memory-efficient dropout and feature re-use. Some approaches consider post-processing steps for further contour refinement by exploiting organ probability maps arising from 3D CED as features for Conditional Random Field (CRF) [28], level-set [9] or graph-cut [29] models. Organ-attention networks with reverse connections followed by statistical fusion [10] tend to reduce uncertainties at weak boundaries and deal with relative organ size variations.

Feeding deep networks with volumetric images obviously faces memory and computational issues. Since increasing the network depth to extract discriminative features with a larger receptive field cannot be done *ad-infinitum*, many methods rely on small patches or downsampled images resulting in a significant loss of spatial context [8]. Reaching accurate organ delineations, however, requires to extract high-level contextual information, as do radiologists visually. Several key contributions in semantic segmentation arose to mimic visual medical image interpretations more closely. First, structure delineation can exploit transfer learning from large non-medical datasets [30, 31] to reduce the data scarcity issue while improving model generalizability [32]. Second, stacking multiple CEDs encourages the integration of more representative multi-level information [33, 34]. In particular, cascades of deep CEDs can embed auto-context [35] to fuse various amounts of spatial context using posterior probabilities resulting from one CED block to the subsequent. Third, conditional generative adversarial networks extends standard image-to-image translation [36] by including a discriminator whose role is to enforce the model to generate realis-

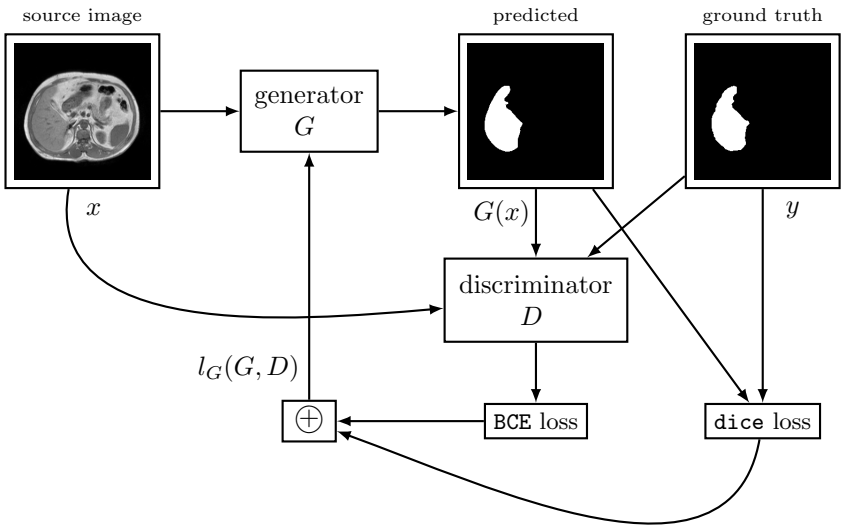

Figure 2: Conditional generative adversarial networks combining Dice and Binary Cross Entropy (BCE) losses for abdominal organ segmentation. The + symbol indicates that the loss fonction $l_G(G, D)$ for the generator G is a weighted sum of Dice and BCE metrics (Eq.1).

tic outputs. Successfully applied to medical images [37, 38, 39], introducing adversarial training to semantic segmentation not only leverages overall delineation performance but also alleviates high-order spatial incorrectness such as inaccurate boundaries or isolated false positives [40, 41]. In the context of abdominal image analysis, all these avenues represent promising methodological developments to achieve more generic computational models for CT and MRI multi-organ segmentation.

3. Methods

3.1. Conditional generative adversarial networks

Recent works including [37, 38, 39, 41] have demonstrated the feasibility of image-to-image translation [36] based on conditional Generative Adversarial Networks (cGAN) for medical image segmentation purposes. cGAN architectures (Fig.2) are made of a generator which provides segmentation masks through encoding and decoding layers and a discriminator (Fig.3) which assesses if a given segmentation mask is synthetic or real. Thus, the adversarial

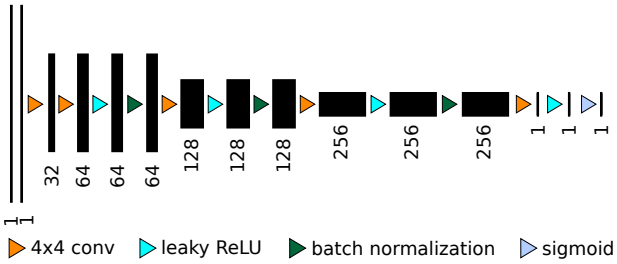

Figure 3: Discriminative part of conditional generative adversarial networks.

network learns to discriminate real (i.e. ground truth) from synthetic delineations (those arising from the generator) to enforce the generative part to create segmentation masks as plausible as possible. Contrary to standard iterative post-processing schemes such as Conditional Random Field (CRF) [28] or level-set [9], this refinement is performed in an end-to-end fashion [41].

cGAN pipelines usually use UNet [12] as generator G (Fig.4a). Its symmetrical architecture comprises an encoder which gradually reduces the spatial dimension using pooling layers, a decoder progressively recovering object details and initial resolution as well as long-range shortcuts which concatenate features between contracting and expanding paths. Specifically, UNet consists of sequential layers including 3×3 convolutional layers followed by Rectified Linear Unit (ReLU) activations. Spatial size is reduced using 2×2 max pooling layers. The first convolutional layer generates 32 channels [12]. This number doubles after each pooling as the network deepens. Following [37], the discriminator D consists of five 4×4 convolutional layers followed by leaky ReLU activation functions and batch normalization (Fig.3). The discriminator inputs are the concatenation of both source images and ground truth or predicted binary masks to be evaluated. The output is an array where each value is defined between 0 (fake) and 1 (plausible or real) and corresponds to the degree of segmentation likelihood for a given image crop and its associated segmentation mask. Let x and y be the source images and ground truth delineation masks, $\lambda = 150$ an empirically set weighting factor [37], $G(x)$ and $D(x, G(x))$ the outputs of G and D, $l_{\texttt{dice}}$ the Dice loss estimated by comparing predicted and ground truth masks.

As in [41, 37], the loss function $l_G(G, D)$ for the generator G is defined as the following combination:

$$l_G(G, D) = \mathop{\mathbb{E}}_{x,y}[-\texttt{log}(D(x, G(x)))] + \lambda \mathop{\mathbb{E}}_{x,y}[l_{\texttt{dice}}(G(x), y)] \quad (1)$$

Minimizing $l_{\texttt{dice}}$ tends to provide rough organ shape predictions whereas maximizing $\texttt{log}(D(x, G(x))$ aims at improving contour delineations. The loss function for the discriminator D is such that:

$$\begin{aligned} l_D(G, D) &= \mathop{\mathbb{E}}_{x,y}[-\texttt{log}(D(x, y))] \\ &+ \mathop{\mathbb{E}}_{x,y}[-\texttt{log}(1 - D(x, G(x)))] \end{aligned} \quad (2)$$

The optimizer fits D through Binary Cross Entropy (BCE) using estimated and ground truth masks. It maximizes loss values for ground truth ($\texttt{log}(D(x, y))$) and minimizes loss values for generated ($-\texttt{log}(1 - D(x, G(x)))$) masks. Optimization is performed sequentially by alternating at each batch gradient descents on G and D [42]. To further improve cGAN abilities to extract contours from the abdominal anatomy, investigations on more robust generators than traditional UNet are needed.

3.2. Partially pre-trained generator

CED architectures dedicated to medical image segmentation are typically trained from scratch, relying on randomly initialized weights. Since the amount of available images cannot be endlessly extended, reaching a generic model without over-fitting is therefore challenging. As deep classification networks which usually involve model pre-trained on large datasets, the encoder part of CEDs can be replaced by a classification network whose weights are previously trained on an initial classification task [30]. It exploits transfer learning and fine-tuning from large datasets like ImageNet [43] towards better semantic segmentation. In the literature, the encoder has been already replaced by pre-trained VGG-11 [30] or WideResnet-38 [44] networks. Our previous study [31] exploits pre-trained

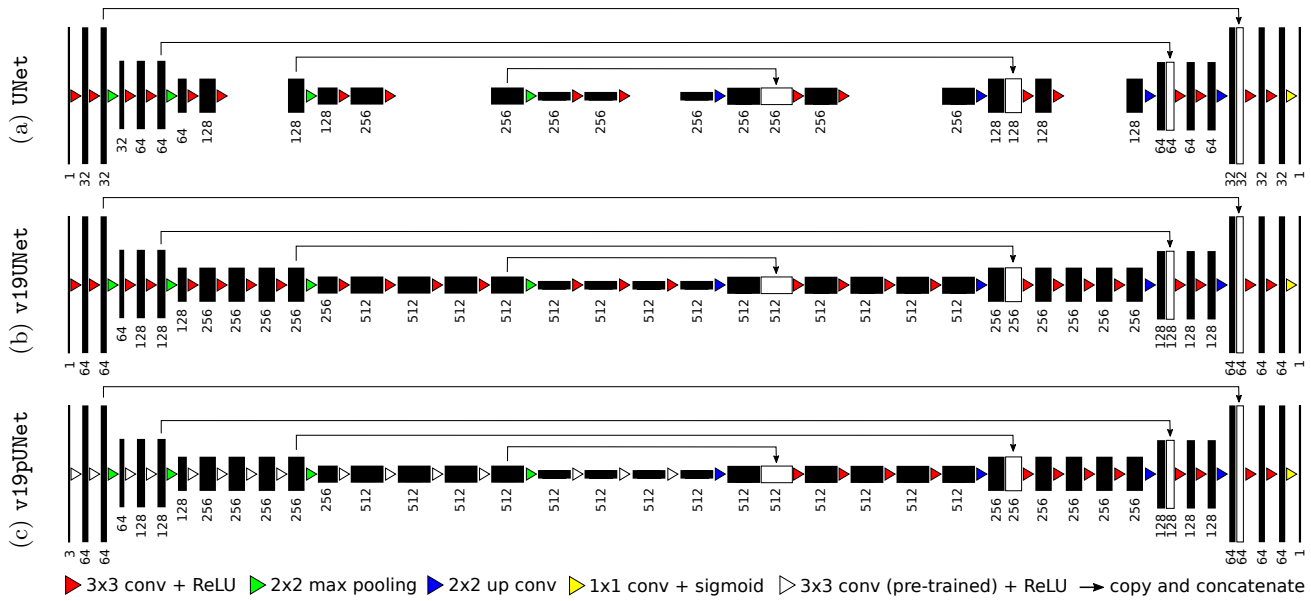

Figure 4: Extension of UNet [12] (a) by exploiting as encoder a slightly modified VGG-19 [45] without (b) and with (c) weights pre-trained on ImageNet [43]. The decoder is modified to get a symmetrical construction while keeping long-range shortcuts.

VGG-16 encoders and reveals significant improvements compared to their randomly weighted counterparts.

This approach can be further improved by extending standard UNet [12] by a deeper network from the VGG family [45] as encoder: the VGG-19 architecture. Compared to UNet (Fig.4*a*), the first convolutional layer of `v19UNet` (Fig.4*b*) generates 64 channels instead of 32. The number of channels doubles after each max pooling until it reaches 512 (256 for UNet). After the second max pooling, the number of convolutional layers differs from UNet with patterns of 4 consecutive layers instead of 2. Compared to VGG-19 [45], top layers including fully-connected layers and softmax are omitted. The three last convolutional VGG-19 layers serve as central part to separate the contracting and expanding paths. To improve performance, this encoder branch is pre-trained on ImageNet [43] to get the CED architecture referred to `v19pUNet` (Fig.4*c*). Pre-training this encoder using more than 1 million non-medical data collected for object recognition purposes improves predictive performance on abdominal data and

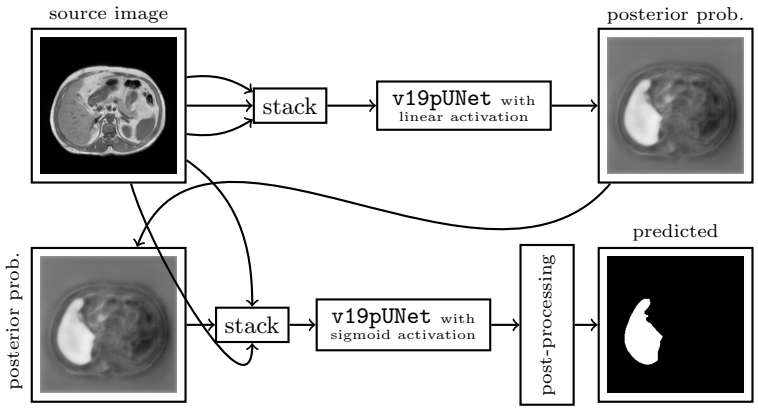

Figure 5: Cascaded convolutional encoder-decoders with auto-context to exploit multi-level contextual information. The `stack` blocks consist in concatenating multiple inputs.

requires less training time to reach convergence. In practice, axial slices are extended from single greyscale channel to 3 channels by repeating the same content to fit the RGB ImageNet image depth.

To get a symmetrical construction while keeping long-range shortcuts (Fig.4*c*), the decoder branch is extended in the same fashion by adding 4 convolutional layers and more features channels. Contrary to encoder weights which are initialized through ImageNet pre-training, decoder weights are set randomly. A final 1×1 convolutional layer with sigmoid activation function achieves pixel-wise segmentation at native resolution.

3.3. Cascaded generator with auto-context

Managing long-range spatial context is key to improve abdominal organ delineations [24]. However, increasing *ad-infinitum* the network depth to exploit larger receptive fields is not suitable for memory and computational issues. Alternatively, in the same spirit of [34], we propose to process abdominal images using a cascade of deep CEDs to exploit multi-level contextual information (Fig.5). Our strategy, referred to `v19pUNet`$_{1\text{-}1}$, consists in combining two partially pre-trained `v19pUNet` networks with auto-context [35], i.e. using posterior probabilities resulting from the first `v19pUNet` as features for the second one [46]. It extends with more complex architectures a proof-of concept given in [47] us-

ing standard UNet (UNet$_{1\text{-}1}$). The sigmoid activation of the first v19pUNet used in the last 1×1 convolution layer (Fig.4c) is replaced by a linear function to generate continuous output maps. These maps are normalized, concatenated to source images and given as inputs of the second v19pUNet which is trained to give final organ delineations. In practice, since the second CED requires 3-layers volumes as inputs due to its encodeur pre-trained on ImageNet, we concatenate the posterior probability map with the source image whose content is replicated twice (stack block in Fig.5). Instead of training both models separately [46], our pipeline is trained end-to-end to exploit simultaneous multi-level segmentation refinements. Making the first v19pUNet generating continuous instead of binary outputs propagates pixel-wise confidence information to the second v19pUNet and postpones the final segmentation decision to the pipeline ending part. Contrary to [34], both networks process source images at full-resolution. Moreover, we keep the largest connected segmented area as post-processing.

We propose to use this cascaded partially pre-trained v19pUNet$_{1\text{-}1}$ model as generator within the cGAN pipeline (cGv19pUNet$_{1\text{-}1}$). Robustness and generalization capabilities need to be assessed for abdominal multi-organ segmentation.

4. Results and discussion

4.1. Validation setup

Results are provided using the dataset[3] arising from the CHAOS challenge [2], collected by the Department of Radiology, Dokuz Eylul University Hospital, Izmir, Turkey and involving 80 patients. 40 abdominal CT scans acquired at portal venous phase after contrast agent injection are used with ground truth liver segmentation. The dataset also includes 40 T1-DUAL in phase (T1-DUALin), 40 T1-DUAL oppose phase (T1-DUALout) and 40 T2-SPIR abdominal MR images with ground truth delineations for liver, right kidney, left kidney and spleen. Three radiology experts (10, 12 and 28 years of experi-

[3]CHAOS data available at https://doi.org/10.5281/zenodo.3362844

ence) were involved for manual segmentation. Final ground truth masks were obtained through majority voting. T1-DUALin and T1-DUALout images are registered. Conversely, T1-DUAL and T2-SPIR sequences are not registered. Following the CHAOS challenge rules, CT and MR datasets are divided into training and test subsets, with a ratio of 50%.

Except for DeepMedic [48], VNet [26] and denseVNet [8] which process data in 3D, evaluated models independently process 2D axial slices and produce 2D segmentation masks which are then stacked together to recover 3D volumes. Image size for axial slices are 256×256 or 288×288 pixels for MR images, 512×512 for CT examinations. The number of axial slices varies from 26 to 50 (resp. 78 to 294) and slice thickness is between 4.4 and 8.0 (resp. 2.0 and 3.2) millimeters for MR (resp. CT) images.

Let S and G deal with segmentation results and ground truth. To assess standard CED (DeepMedic [48], VNet [26], denseVNet [8], `UNet` [12]), deeper CED without or with encoder pre-training (`v16UNet`, `v16pUNet` [31], `v19UNet`, `v19pUNet`), CED using nested and dense skip connections (`v19UNet+`, `v19pUNet+` [49]), cascaded CED ($\texttt{UNet}_{1\text{-}1}$ [47], $\texttt{v16pUNet}_{1\text{-}1}$, $\texttt{v19pUNet}_{1\text{-}1}$) and cGAN with partially pre-trained cascaded CED as generator ($\texttt{cGv16pUNet}_{1\text{-}1}$, $\texttt{cGv19pUNet}_{1\text{-}1}$), the accuracy of abdominal organ segmentation is quantified based on Dice coefficient (`dice`) estimated following $\frac{2|S \cap G|}{|S|+|G|}$ where $|.|$ denotes cardinality, relative absolute volume difference (`RAVD`) comparing S and G such as $\texttt{RAVD} = \frac{\texttt{abs}(|S|-|G|)}{|G|}$, average and maximum symmetric surface distances (`ASSD`, `MSSD`) which correspond to the average (resp. maximum) Hausdorff distance between border voxels in S and G. These metrics tend to provide an overall assessment of the involved networks. Following [2], we also provide model ranking scores by averaging all metrics after having transformed values to span the $[0, 100]$ interval so that higher values correspond to better segmentation. To discard unacceptable accuracy and increase metric sensitivity [2], thresholds are set up according to intra/inter-expert similarities: $\texttt{dice} > 80\%$, $\texttt{RAVD} < 5\%$, $\texttt{ASSD} < 15$mm and $\texttt{MSSD} < 60$mm. Metrics outside the threshold range get zero points. Scores reached for multi-organ segmentation are obtained by averag-

ing the scores obtained for each organ. Similarly, scores for MR images are the average between results arising from T1-DUAL`in/out` and T2-SPIR modalities.

In our experiments, a given model is dedicated to one single modality (T1-DUAL`in/out`, T2-SPIR, CT) and one single organ (liver, right kidney, left kidney, spleen). Each model thus performs binary instead of multi-class segmentation to extract robust organ-specific features. In addition, experiments on the T1-DUAL modality stack together T1-DUAL`in` and T1-DUAL`out` images as model inputs since both phases are registered. When 3 channels are required, as for `v16(p)UNet`, the third channel consists of the T1DUAL`in` image duplication. For CT and T2-SPIR, image content is replicated twice.

Deep CEDs are trained using data augmentation to teach networks efficient invariance and robustness properties [12]. Training axial slices undergo random scaling, rotation, shearing and shifting operations. 100 augmented images are produced for a single training slice. For CT (MR) images, models are trained with 6 (20) epochs, a batch size of 3 (5) images, an *Adam* optimizer with 10^{-5} as learning rate and a fuzzy Dice score as loss function. Models were implemented using Keras and trained using a single Nvidia GeForce GTX 1080 Ti GPU.

4.2. Evaluation on clinical data

CT and MR liver segmentation. Quantitative metric and score values are provided in Tab.1 for liver CT/MR delineations. For both modalities, standard architectures including DeepMedic, VNet, denseVNet and `UNet` are outperformed by deeper (`v16`/`v16pUNet`, `v19`/`v19pUNet`) and cascaded ($\texttt{UNet}_{1\text{-}1}$) networks which indicates that better predictive performance and generalizability are reached using more complex models. In one hand, comparisons between `v16`/`v19UNet` and their partially pre-trained extensions (`v16p`/`v19pUNet`) reveals that pre-training the encoder using non-medical ImageNet data makes the network converge towards a better solution. In particular, large gains in terms of `dice` (91.60 to 94.07%) and `ASSD` (2.96 to 1.70mm) are reported for MR images between `v16UNet` and `v16pUNet`. In the other hand, extending `UNet` into a cascaded pipeline ($\texttt{UNet}_{1\text{-}1}$) with auto-context and end-to-end-training al-

organ	model	CT					MRI				
		dice ↑	RAVD ↓	ASSD ↓	MSSD ↓	score	dice ↑	RAVD ↓	ASSD ↓	MSSD ↓	score
liver	DeepMedic [48]	96.70±1.36	3.18±3.42	1.24±0.48	27.90±10.0	73.32	89.74±7.54	6.52±8.27	4.74±4.83	122.5±53.2	47.64
	VNet [26]	89.58±8.54	6.78±12.6	4.87±8.80	42.52±48.6	60.01	74.55±6.23	42.5±26.2	9.21±3.63	75.59±31.2	16.81
	denseVNet [8]	95.26±1.14	2.89±2.53	1.57±0.48	23.89±9.19	73.78	76.75±6.86	17.4±13.1	8.27±3.12	54.98±28.4	28.91
	UNet [12]	97.35±0.50	1.80±1.35	1.09±0.46	22.72±10.6	79.07	90.68±5.30	7.89±8.69	3.29±2.39	44.49±24.0	58.02
	v16UNet	97.67±0.41	1.39±1.15	0.88±0.25	19.85±8.92	82.71	91.60±5.44	6.87±8.63	2.96±2.37	40.75±25.2	60.86
	v16pUNet [31]	97.86±0.32	1.29±1.01	0.80±0.24	19.09±8.84	83.71	94.07±2.32	4.25±3.46	1.70±0.94	29.54±12.2	67.99
	v19UNet	97.60±0.44	1.38±1.41	0.94±0.35	20.69±9.00	82.34	92.10±4.49	6.04±7.44	2.65±1.97	37.96±18.8	61.55
	v19pUNet	97.88±0.37	1.22±0.82	0.82±0.26	19.87±8.86	83.71	93.44±4.11	5.24±5.87	1.97±1.39	32.41±13.6	65.32
	v19UNet+ [49]	97.38±0.61	2.06±1.90	1.16±0.45	26.26±11.9	76.61	92.22±4.46	6.61±7.28	2.41±1.51	35.62±17.0	61.39
	v19pUNet+ [49]	97.80±0.42	1.49±1.45	0.85±0.26	18.97±6.71	82.69	92.83±6.92	5.91±8.73	2.12±2.04	31.54±18.3	66.14
	UNet$_{1\text{-}1}$ [47]	97.48±0.61	1.64±1.82	1.02±0.59	20.89±10.9	81.28	92.03±4.04	5.81±6.73	2.45±1.43	34.04±15.9	63.05
	v16pUNet$_{1\text{-}1}$	97.94±0.32	1.12±0.91	**0.76**±0.16	**17.08**±5.80	**85.53**	94.28±1.99	4.09±3.07	1.67±0.94	28.60±12.4	68.92
	v19pUNet$_{1\text{-}1}$	97.91±0.26	1.14±0.95	0.78±0.17	19.44±7.46	84.40	**94.52**±1.64	3.52±2.32	**1.62**±1.02	**27.02**±14.5	**70.05**
	cGv16pUNet$_{1\text{-}1}$	**97.95**±0.27	1.19±0.89	**0.76**±0.16	18.69±7.58	84.50	94.02±2.42	4.41±3.73	1.79±1.06	30.02±13.6	67.88
	cGv19pUNet$_{1\text{-}1}$	97.87±0.32	**1.09**±0.96	0.80±0.23	20.52±8.24	84.15	94.33±1.75	**3.49**±2.57	1.73±0.97	28.94±15.0	69.21
	MOvpUNet	97.94±0.32	1.12±0.91	0.76±0.16	17.08±5.80	85.53	94.45±1.74	3.45±2.45	1.67±1.01	27.46±14.5	70.17

Table 1: Quantitative assessment of DeepMedic [48], VNet [26], denseVNet [8], UNet [12], v16UNet, v16pUNet [31], v19UNet, v19pUNet, v19UNet+ [49], v19pUNet+ [49], UNet$_{1\text{-}1}$ [47] and proposed v16pUNet$_{1\text{-}1}$, v19pUNet$_{1\text{-}1}$, cGv16pUNet$_{1\text{-}1}$, cGv19pUNet$_{1\text{-}1}$ and MOvpUNet architectures for healthy liver segmentation in CT and MR images. Bold and underline results indicate first and second best scores.

lows to take advantage of multi-level context, with score improvements from 79.07 (58.02) to 81.28% (63.05%) in CT (MR). v19UNet and v19pUNet give better or slightly similar performance than their nested and dense counterparts (v19/v19pUNet+) which suggests that the great complexity brought by such heavy architectures [49] is not useful to provide relevant liver contours.

By combining these three contributions (deeper model, encoder pre-training, cascaded architecture), v16pUNet$_{1\text{-}1}$ (v19pUNet$_{1\text{-}1}$) discriminates more efficiently liver areas from surrounding structures by achieving the best score for CT (MR) scans with 85.53% (70.05%). Embedding v16pUNet$_{1\text{-}1}$ (v19pUNet$_{1\text{-}1}$) into a cGAN pipeline for CT (MR) liver segmentation gives broadly similar results but provides the second best scores. We note that cGv16pUNet$_{1\text{-}1}$ reaches the best dice (97.95%) and similar ASSD (0.76mm) in CT. In MR, the best RAVD (3.49%) is obtained using cGv19pUNet$_{1\text{-}1}$. $\lambda = 150$ was selected as in [37] to provide a good balance between $\underset{x,y}{\mathbb{E}}[-\texttt{log}(D(x, G(x)))]$ and $\underset{x,y}{\mathbb{E}}[l_{\texttt{dice}}(G(x), y)]$ in Eq.1.

Qualitative results for CT and MR liver segmentation are displayed in Fig. 6-7. Compared to standard networks as well as v16pUNet (v19pUNet/v19pUNet+) which are prone to under- or over-segmentation, sometimes combined with unrealistic shapes, better contour adherence and shape consistency are reached by

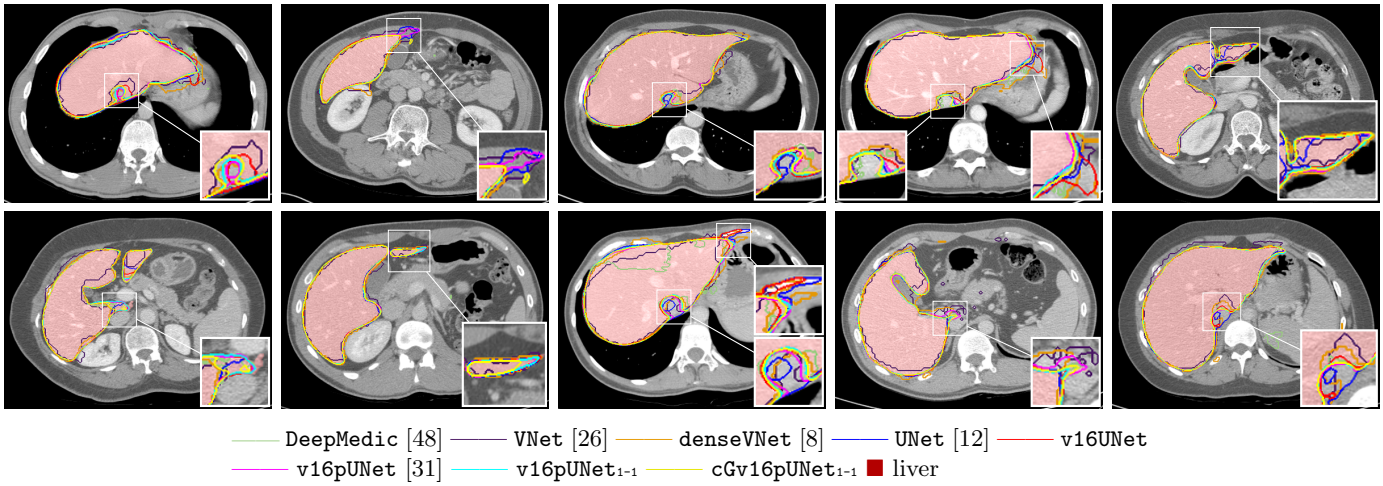

Figure 6: Liver CT segmentation using DeepMedic [48], VNet [26], denseVNet [8], UNet [12], v16UNet, v16pUNet [31], proposed v16pUNet$_{1\text{-}1}$ and cGv16pUNet$_{1\text{-}1}$. Ground truth is superimposed in red color.

v16pUNet$_{1\text{-}1}$ (v19pUNet$_{1\text{-}1}$) and cGv16pUNet$_{1\text{-}1}$ (cGv19pUNet$_{1\text{-}1}$) whose ability to mimic expert annotations is notable for CT (T1-DUAL and T2-SPIR). The diversity in terms of textures arising in MR images is accurately captured through cascaded partially pre-trained networks despite high similar visual properties with surrounding structures. Moreover, deep networks find it harder to segment abdominal MR compared to CT images due to lower contrast and resolution combined with higher spacing which makes the number of 2D axial slices substantially smaller.

Abdominal multi-organ MR segmentation. Tab.2 shows quantitative results for multi-organ MR segmentation. As for liver, the DeepMedic, VNet, denseVNet, UNet, v16UNet and v19UNet networks do not provide the required robustness for organ extraction. In T1-DUAL modality, significant improvements can be noticed using v16pUNet for left kidney, v19pUNet for right kidney and v19pUNet+ for spleen with the best reached scores (63.52, 68.07 and 69.04%) among all schemes. Except for spleen in T2-SPIR, the comparison UNet/UNet$_{1\text{-}1}$ indicates the appropriateness of exploiting networks in a cascaded manner, as proven for spleen (right kidney) in T1-DUAL (T2-SPIR) whose dice

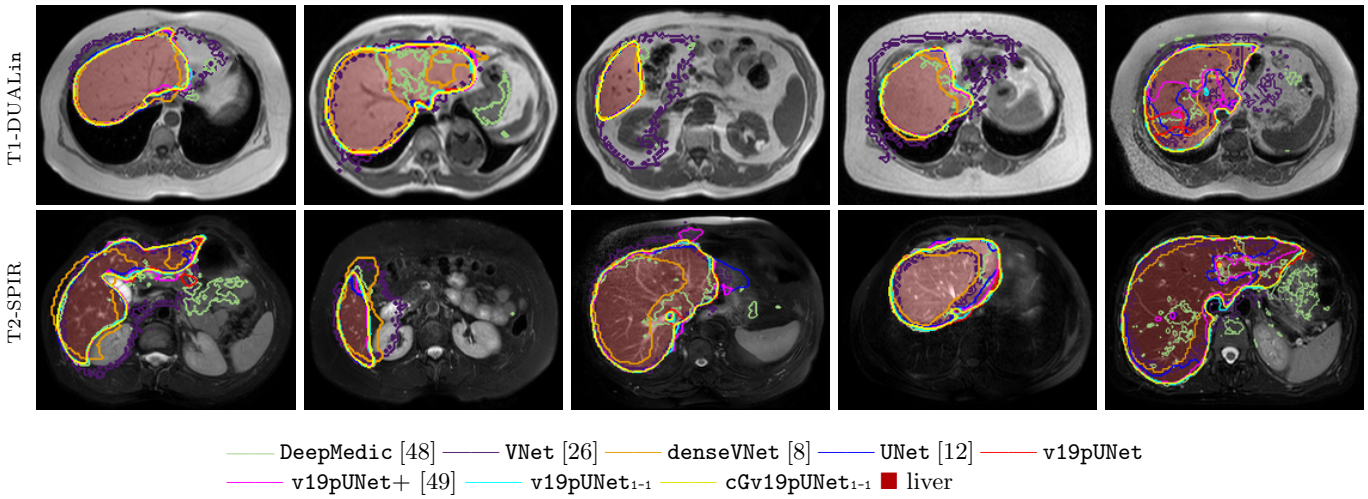

Figure 7: Liver MRI (T1-DUALin, T2-SPIR) segmentation using DeepMedic [48], VNet [26], denseVNet [8], UNet [12], v19pUNet, v19pUNet+ [49], proposed v19pUNet$_{1\text{-}1}$ and cGv19pUNet$_{1\text{-}1}$. Ground truth is superimposed in red color.

(RAVD) jumps from 81.56 (15.39) to 87.01% (9.04%). The same conclusion arises between v16p/v19pUNet and v16p/v19pUNet$_{1\text{-}1}$ with, for instance, a strong score improvement reached using v19pUNet$_{1\text{-}1}$ for right kidney in T2-SPIR (68.38 to 72.71%). Cascaded pre-trained cGAN strategies (cGv16p/cGv19pUNet$_{1\text{-}1}$) always belong to one of the two best methods in dice, except for left kidney in T1-DUAL. Gains for MSSD in T2-SPIR are highlighted with 11.23mm (10.96) obtained for right kidney (spleen) using cGv19pUNet$_{1\text{-}1}$ (cGv16pUNet$_{1\text{-}1}$). In T1-DUAL, cGv19pUNet$_{1\text{-}1}$ (cGv16pUNet$_{1\text{-}1}$) achieves the best RAVD scores for liver and right kidney (spleen). p-values obtained using Student's paired t tests for Dice, RAVD, ASSD and MSSD metrics between cGv19pUNet$_{1\text{-}1}$ and standard CED architectures such as DeepMedic or denseVNet globally confirmed that our contributions bring an added-value with statistical significance. Unsurprisingly, delineating small organs (kidneys) is more challenging than focusing on larger ones (liver, spleen). The vicinity between left kidney and spleen further complicates the contouring task. MR segmentation is easier with T2-SPIR than T1-DUAL since relative contrasts between structures is clearly enhanced.

As visually shown in Fig.8, many anomalies are present using standard networks with over (under-) detection issues for DeepMedic (denseVNet) in T2-

organ	model	T1-DUALin/out					T2-SPIR				
		dice ↑	RAVD ↓	ASSD ↓	MSSD ↓	**score**	dice ↑	RAVD ↓	ASSD ↓	MSSD ↓	**score**
liver	DeepMedic [48]	89.24±9.81	7.11±10.1	5.05±5.88	116.5±47.6	47.59	90.24±5.27	5.94±6.49	4.43±3.78	128.5±58.8	47.69
	denseVNet [8]	85.56±6.10	17.2±12.3	4.63±2.67	43.10±23.6	45.55	67.94±7.62	17.6±13.8	11.9±3.58	66.86±33.3	12.26
	UNet [12]	90.48±7.44	9.57±12.4	2.74±1.99	35.39±19.1	60.85	90.52±3.34	6.44±5.14	3.84±2.79	53.59±29.0	55.14
	v16UNet	91.32±7.66	8.38±12.3	2.36±1.94	29.62±18.0	63.43	91.75±3.19	5.54±4.94	3.61±2.80	52.42±32.6	57.63
	v16pUNet [31]	93.64±2.84	5.77±5.28	1.79±0.92	31.17±11.7	64.81	94.46±1.82	2.79±1.73	1.65±0.96	28.97±12.7	70.56
	v19UNet	91.83±6.28	7.38±10.3	2.32±1.75	32.04±16.7	62.29	92.40±2.70	4.82±4.49	2.94±2.21	43.86±20.9	60.21
	v19pUNet	92.39±6.30	7.71±9.96	2.31±1.81	35.70±14.3	59.77	94.48±1.92	2.84±1.80	1.63±0.96	29.17±12.8	70.56
	v19UNet+ [49]	91.64±6.01	7.94±10.2	2.44±1.58	36.62±19.8	60.24	92.68±2.89	5.25±4.51	2.39±1.44	34.61±14.1	62.91
	v19pUNet+ [49]	91.17±12.1	8.86±15.7	2.54±3.13	31.47±19.7	63.23	94.48±1.69	3.03±1.81	1.69±0.95	31.62±16.9	68.71
	UNet$_{1\text{-}1}$ [47]	92.06±5.34	6.35±9.10	2.30±1.61	31.45±16.4	65.22	91.97±2.74	5.41±4.43	2.56±1.24	36.42±14.9	60.74
	v16pUNet$_{1\text{-}1}$	93.86±2.28	5.49±4.72	1.77±0.88	29.87±10.6	65.93	93.69±1.68	**2.66**±1.41	**1.57**±0.99	27.31±13.8	72.04
	v19pUNet$_{1\text{-}1}$	**94.38**±1.33	4.25±3.17	1.68±0.97	**28.45**±14.9	67.70	**94.67**±1.92	2.80±1.49	1.57±1.06	**25.49**±14.1	**72.28**
	cGv16pUNet$_{1\text{-}1}$	93.45±2.96	6.05±5.77	1.96±1.05	32.30±13.4	64.57	94.60±1.89	2.78±1.77	1.63±1.07	27.74±13.9	71.26
	cGv19pUNet$_{1\text{-}1}$	94.23±1.54	**4.17**±3.42	**1.76**±0.96	29.42±15.0	**67.72**	94.44±1.95	2.86±1.71	1.71±1.00	28.46±15.0	70.49
right kidney	DeepMedic [48]	75.13±15.8	29.06±19.8	3.73±2.11	105.8±62.5	35.42	89.32±10.0	11.83±14.7	1.66±1.28	102.5±63.9	51.40
	denseVNet [8]	76.31±11.2	25.66±20.0	4.39±2.43	34.72±38.1	41.63	67.94±6.92	21.00±14.1	7.59±4.00	61.14±47.2	23.61
	UNet [12]	85.61±13.2	13.44±16.9	2.55±3.45	20.26±15.6	61.05	88.16±5.77	15.39±15.2	3.70±4.54	28.85±22.8	55.87
	v16UNet	87.19±6.11	11.52±10.5	2.61±2.82	23.83±16.3	57.85	91.68±3.72	9.85±6.54	1.42±1.14	18.00±10.4	64.19
	v16pUNet [31]	90.08±3.88	11.55±7.40	1.38±0.86	12.35±7.22	66.32	92.47±3.96	8.56±5.04	1.08±1.12	12.64±8.19	67.79
	v19UNet	87.36±7.39	13.67±14.7	2.12±2.03	20.27±14.6	61.69	92.12±3.91	8.89±6.16	1.28±1.19	15.30±8.80	66.59
	v19pUNet	90.26±4.28	10.96±8.39	1.29±0.81	**11.58**±6.35	**68.07**	92.66±4.08	7.85±5.11	1.07±1.14	12.65±8.77	68.38
	v19UNet+ [49]	86.49±9.47	14.40±13.1	2.09±1.31	19.32±9.03	59.67	86.43±20.8	10.18±7.28	6.61±22.8	24.02±29.6	61.91
	v19pUNet+ [49]	89.34±5.48	13.75±8.42	1.46±0.99	12.90±6.74	64.50	92.82±3.33	8.46±4.39	1.16±1.14	15.32±9.95	67.56
	UNet$_{1\text{-}1}$ [47]	86.32±9.30	14.82±14.9	2.11±1.71	17.15±10.3	60.61	91.17±4.46	9.04±8.30	1.77±1.90	18.07±13.6	65.75
	v16pUNet$_{1\text{-}1}$	90.27±3.19	11.92±6.32	**1.32**±0.72	11.95±7.25	66.22	92.78±4.19	8.71±5.32	1.02±1.14	12.46±7.87	68.39
	v19pUNet$_{1\text{-}1}$	90.30±3.73	11.66±7.28	1.47±1.16	13.96±10.5	66.19	93.21±2.84	**7.76**±6.26	**0.97**±1.05	12.63±8.12	**72.71**
	cGv16pUNet$_{1\text{-}1}$	90.29±3.91	11.38±8.50	1.38±1.02	13.60±8.18	67.26	**93.22**±3.45	8.06±7.89	1.01±1.06	15.87±10.2	71.34
	cGv19pUNet$_{1\text{-}1}$	**90.56**±4.28	**10.44**±8.92	1.37±0.99	13.39±9.21	66.67	93.02±3.74	7.94±6.77	0.99±1.09	**11.23**±7.80	71.21
left kidney	DeepMedic [48]	69.95±21.7	34.17±24.4	5.81±6.01	123.7±56.8	28.38	80.36±23.8	20.43±27.1	3.52±3.15	120.1±70.4	45.48
	denseVNet [8]	68.71±22.7	25.83±26.7	32.1±120.	61.22±119.	40.12	64.84±11.8	23.01±19.3	7.81±3.64	48.08±36.9	24.02
	UNet [12]	81.55±16.8	16.86±18.5	4.25±6.99	37.02±30.3	51.71	90.32±3.47	9.34±6.70	2.11±1.37	36.19±23.8	58.95
	v16UNet	83.33±15.3	16.94±17.6	3.04±3.65	30.92±31.5	53.49	91.52±2.52	9.32±5.06	1.74±1.41	26.91±24.5	62.75
	v16pUNet [31]	85.79±20.4	**10.13**±6.37	8.91±33.7	34.53±50.2	**63.52**	**92.83**±2.14	8.18±5.20	**1.32**±1.27	24.08±22.3	64.58
	v19UNet	82.06±20.3	16.45±16.4	7.97±23.3	44.09±55.9	52.73	90.64±4.01	9.90±5.45	1.77±1.55	26.94±22.9	61.76
	v19pUNet	85.58±20.5	14.38±16.5	8.68±31.5	34.84±51.1	59.64	92.60±2.39	9.00±6.11	1.57±2.08	23.62±22.2	63.52
	v19UNet+ [49]	82.69±20.4	16.49±19.8	7.72±23.9	39.19±57.6	56.03	90.97±4.05	11.59±10.1	2.26±2.59	31.25±24.9	58.61
	v19pUNet+ [49]	**88.76**±7.87	13.94±10.3	**1.62**±1.21	**27.17**±30.3	61.03	92.79±3.00	9.01±6.06	1.49±2.10	22.69±23.0	64.54
	UNet$_{1\text{-}1}$ [47]	83.88±12.2	16.72±16.2	2.91±2.33	33.55±29.7	51.35	89.91±4.65	10.89±7.62	1.67±1.03	23.50±20.4	64.30
	v16pUNet$_{1\text{-}1}$	85.56±20.5	11.40±9.22	8.58±31.7	37.23±48.2	61.66	92.78±2.97	8.76±7.59	1.46±2.11	**22.04**±23.2	64.57
	v19pUNet$_{1\text{-}1}$	84.01±20.5	14.18±10.1	9.01±32.2	35.68±49.2	56.64	92.10±3.03	9.45±8.11	1.87±2.79	24.31±23.0	62.63
	cGv16pUNet$_{1\text{-}1}$	84.70±20.4	12.10±9.23	8.70±30.3	38.39±48.6	56.13	92.83±2.27	**8.05**±5.40	1.35±1.41	23.90±22.3	**65.56**
	cGv19pUNet$_{1\text{-}1}$	85.31±20.4	13.17±15.1	7.04±23.8	36.08±49.1	59.89	92.67±3.30	8.88±8.80	1.67±2.91	23.89±23.6	64.64
spleen	DeepMedic [48]	75.33±24.2	23.23±27.4	5.59±6.74	155.3±66.2	35.16	88.24±5.57	13.15±8.59	5.07±4.75	165.8±87.9	40.50
	denseVNet [8]	69.38±17.3	31.26±19.8	6.18±3.98	61.09±71.7	31.06	48.56±19.3	25.11±17.0	16.1±9.94	90.46±67.5	9.33
	UNet [12]	81.56±19.8	20.68±25.0	3.38±4.08	26.41±18.9	53.99	89.33±5.83	8.56±7.73	2.02±1.86	23.46±16.2	59.85
	v16UNet	85.20±9.90	17.67±17.9	2.71±2.57	29.08±22.4	55.51	89.84±6.51	10.14±8.97	2.04±2.32	22.01±17.4	62.22
	v16pUNet [31]	89.66±3.75	11.79±7.30	1.41±0.77	15.24±11.5	66.35	92.60±3.29	7.92±5.43	1.03±0.63	15.04±9.68	68.93
	v19UNet	82.40±20.6	16.51±14.0	6.80±20.3	31.72±35.5	56.09	89.74±5.46	9.63±7.99	1.97±1.65	23.22±13.8	60.90
	v19pUNet	89.59±3.88	11.60±8.06	1.57±0.97	17.87±15.1	66.28	92.17±3.71	9.27±5.48	1.10±0.75	16.86±10.7	66.82
	v19UNet+ [49]	84.24±12.6	18.48±18.4	2.76±2.73	23.94±17.9	56.75	89.34±6.76	10.87±9.70	1.96±2.24	21.49±12.2	62.05
	v19pUNet+ [49]	89.54±4.10	11.81±8.89	1.43±0.88	**13.90**±9.25	**69.04**	92.14±4.48	7.82±6.43	1.29±1.51	16.40±13.5	68.56
	UNet$_{1\text{-}1}$ [47]	87.01±7.34	12.03±11.0	1.98±1.31	24.38±13.6	61.43	86.60±10.5	14.10±15.9	3.25±4.51	25.01±22.0	58.58
	v16pUNet$_{1\text{-}1}$	88.89±4.64	11.80±6.41	1.51±0.78	19.24±11.8	62.86	**93.16**±3.26	**7.57**±4.63	0.86±0.57	12.15±7.89	69.83
	v19pUNet$_{1\text{-}1}$	**89.93**±3.79	11.61±7.07	**1.34**±0.66	14.89±9.21	66.94	92.29±4.03	8.68±6.74	1.13±0.94	15.37±11.0	68.64
	cGv16pUNet$_{1\text{-}1}$	89.67±3.92	**10.90**±7.78	1.45±0.78	15.82±9.34	67.02	93.00±3.03	7.70±5.29	**0.84**±0.44	**10.96**±3.58	**70.79**
	cGv19pUNet$_{1\text{-}1}$	89.31±3.88	11.85±6.25	1.63±1.08	19.21±15.4	63.79	92.41±3.58	9.17±5.73	1.05±0.91	11.61±7.85	70.45

Table 2: Quantitative assessment of DeepMedic [48], denseVNet [8], UNet [12], v16UNet, v16pUNet [31], v19UNet, v19pUNet, v19UNet+ [49], v19pUNet+ [49], UNet$_{1\text{-}1}$ [47] and proposed v16pUNet$_{1\text{-}1}$, v19pUNet$_{1\text{-}1}$, cGv16pUNet$_{1\text{-}1}$ and cGv19pUNet$_{1\text{-}1}$ architectures for healthy abdominal multi-organ (liver, right kidney, left kidney, spleen) segmentation in T1-DUALin/out and T2-SPIR images. Bold and underline results indicate first and second best scores.

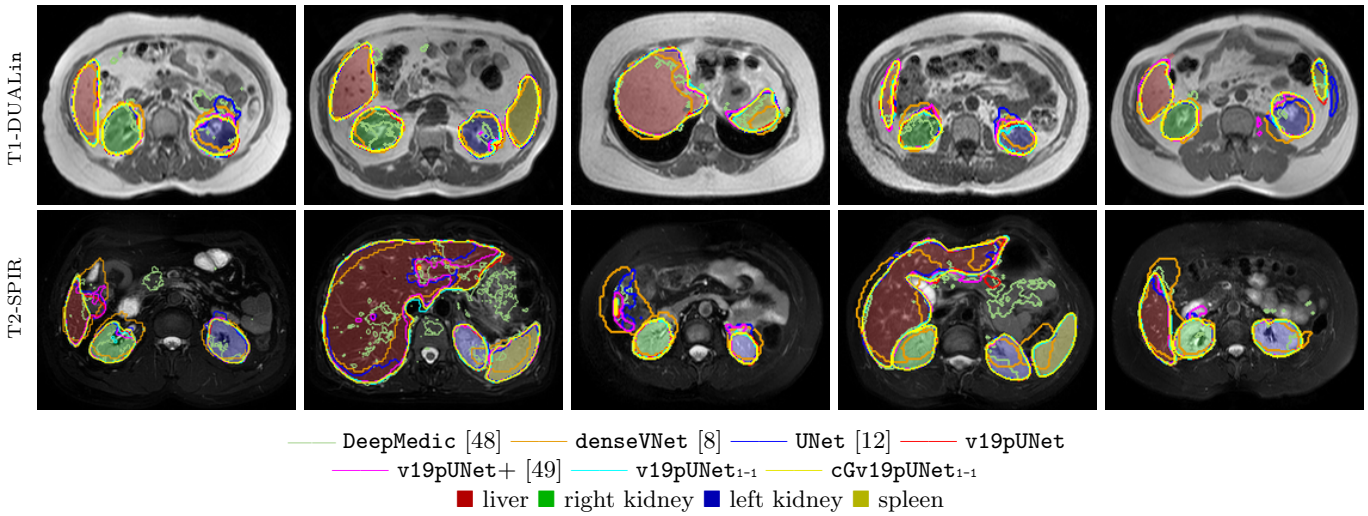

Figure 8: Abdominal multi-organ MRI (T1-DUALin, T2-SPIR) segmentation using DeepMedic [48], denseVNet [8], UNet [12], v19pUNet, v19pUNet+ [49], proposed v19pUNet$_{1\text{-}1}$ and cGv19pUNet$_{1\text{-}1}$. Liver, right kidney, left kidney and spleen ground truth delineations are respectively superimposed in red, green, blue and yelow colors.

SPIR. v19pUNet$_{1\text{-}1}$ and cGv19pUNet$_{1\text{-}1}$ show a better behavior than v19pUNet and v19pUNet+ in accurately fitting organ extents and offering plausible shape consistency, especially for bases and apexes where organs appear smaller. Despite visually similar performance compared to v19pUNet$_{1\text{-}1}$, cGv19pUNet$_{1\text{-}1}$ appears slightly better in providing realistic organ contours.

Towards better multi-organ segmentation. Under the team name PKDIA, the proposed cGv19pUNet$_{1\text{-}1}$ pipeline enabled us to win three CHAOS competition categories: liver CT, liver MR and multi-organ MR segmentation [2]. Nevertheless, since global findings are verified with varying degrees depending on the concerned modality or organ, we provide in Tab.3 an overall evaluation through scores and rankings for CT liver, MR liver, right kidney, left kidney, spleen as well as multi-organ segmentation. cGv19pUNet$_{1\text{-}1}$ indeed appears as the best strategy for MR multi-organ delineation purposes that reinforces the idea that combining deeper (v19) cascaded partially pre-trained convolutional and adversarial networks globally strengthens the generalization abilities of deep

model	CT		MRI									
	liver		liver		right kidney		left kidney		spleen		multi-organ	
	score	rank	**score**	rank	**score**	rank	**score**	rank	**score**	rank	**score**	rank
DeepMedic [48]	73.32	14	47.64	13	43.41	13	36.93	13	37.83	13	41.45	13
denseVNet [8]	73.78	13	28.91	14	32.62	14	32.07	14	20.20	14	28.45	14
UNet [12]	79.07	11	58.02	12	58.46	12	55.33	12	56.92	12	57.18	12
v16UNet	82.71	7	60.86	11	61.02	10	58.12	8	58.86	10	59.63	11
v16pUNet [31]	83.71	5/6	67.99	4	67.05	6	**64.05**	**1**	67.64	4	66.61	4
v19UNet	82.34	9	61.55	9	64.14	8	57.25	11	58.50	11	60.28	9
v19pUNet	83.71	5/6	65.32	7	68.23	4	61.58	5	66.55	6	65.38	7
v19UNet+ [49]	76.61	12	61.39	10	60.79	11	57.32	10	59.40	9	59.77	10
v19pUNet+ [49]	82.69	8	66.14	6	66.03	7	*62.79*	*3*	68.80	2	65.90	6
UNet$_{1\text{-}1}$ [47]	81.28	10	63.05	8	63.18	9	57.82	9	60.01	8	61.00	8
v16pUNet$_{1\text{-}1}$	**85.53**	**1**	*68.92*	*3*	67.31	5	63.11	2	66.35	7	66.44	5
v19pUNet$_{1\text{-}1}$	*84.40*	*3*	**70.05**	**1**	**69.45**	**1**	59.64	7	*67.79*	*3*	*66.72*	*3*
cGv16pUNet$_{1\text{-}1}$	84.50	2	67.88	5	69.30	2	60.85	6	**68.90**	**1**	66.74	2
cGv19pUNet$_{1\text{-}1}$	84.15	4	69.21	2	*68.94*	*3*	62.27	4	67.12	5	**66.86**	**1**
MOvpUNet	85.53	⋆	70.17	⋆	70.39	⋆	64.77	⋆	69.92	⋆	68.78	⋆

Table 3: Scoreboard and ranking of DeepMedic [48], denseVNet [8], UNet [12], v16UNet, v16pUNet [31], v19UNet, v19pUNet, v19UNet+ [49], v19pUNet+ [49], UNet$_{1\text{-}1}$ [47], proposed v16pUNet$_{1\text{-}1}$, v19pUNet$_{1\text{-}1}$, cGv16pUNet$_{1\text{-}1}$, cGv19pUNet$_{1\text{-}1}$ and MOvpUNet for healthy abdominal organ (liver, right kidney, left kidney, spleen) segmentation in CT and MR images. Bold, underline and italic results indicate first, second and third best scores.

learning pipelines. Except for left kidney where v16pUNet performs the best (64.05%), the first rank is always attributed to one of the proposed cascaded pre-trained scheme: v16pUNet$_{1\text{-}1}$ for CT liver, v19pUNet$_{1\text{-}1}$ for MR liver and right kidney (69.45%), cGv16pUNet$_{1\text{-}1}$ for MR spleen (68.9%) and cGv19pUNet$_{1\text{-}1}$ for MR multi-organ (66.86%) segmentation.

By combining the best sequence- and organ-specific networks towards better Multi-Organ (MO) segmentation, we obtain the so-called MOvpUnet computational model including v16pUNet$_{1\text{-}1}$ for liver in CT (Tab.1) as well as respectively for T1-DUAL and T2-SPIR cGv19pUNet$_{1\text{-}1}$ and v19pUNet$_{1\text{-}1}$ for liver, v19pUNet and v19pUNet$_{1\text{-}1}$ for right kidney, v16pUNet and cGv16pUNet$_{1\text{-}1}$ for left kidney, v19pUNet+ and cGv16pUNet$_{1\text{-}1}$ for spleen (Tab.2). The global ranking score reached by cGv19pUNet$_{1\text{-}1}$ for multi-organ MR segmentation is further improved about 2% with MOvpUnet, up to 68.78% (Tab.3). Visually comparing manual and MOvpUnet delineations in Fig.9 further supports the validity of our combined computational model. Outstanding performance is reached in terms of boundary adherence and shape consistency which suggests that integrating MOvpUnet as a guidance tool into clinical routine could greatly help clinicians for abdominal image interpretation.

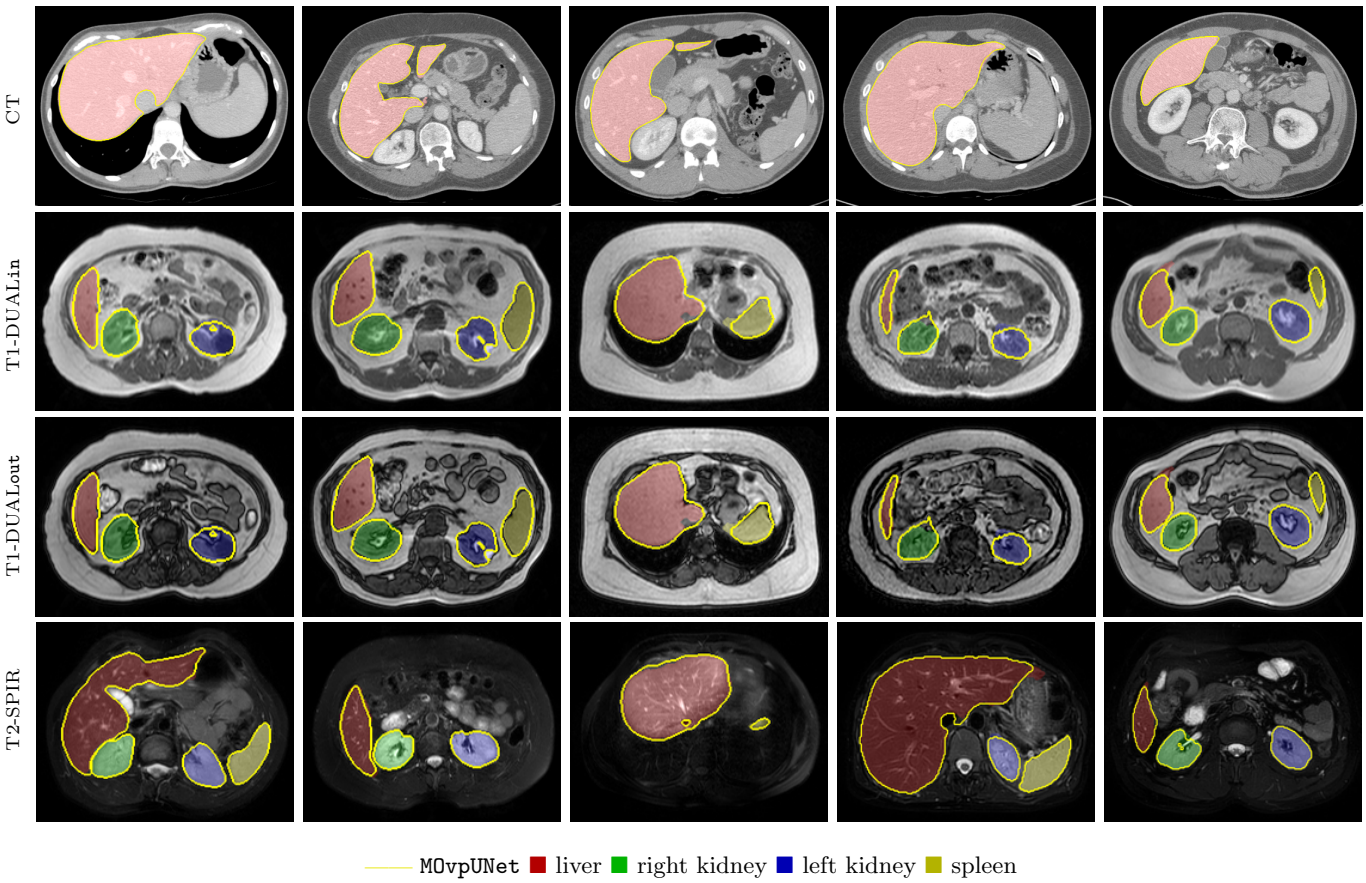

Figure 9: Liver CT and abdominal multi-organ MRI (T1-DUAL`in/out`, T2-SPIR) segmentation using the proposed `MOvpUNet`. Liver, right kidney, left kidney and spleen ground truth delineations are superimposed in red, green, blue and yelow colors.

5. Conclusion

This work tackles fully-automated abdominal organ CT and MR segmentation with deep learning. Standard segmentation networks are extended to cascades of partially pre-trained deep convolutional encoder-decoders. Encoder fine-tuning from a large amount of non-medical images improves predictive performance while alleviating data scarcity limitations. The cascaded architecture exploits multi-level contextual information through auto-context and end-to-end training. Such model is used as generator in a conditional generative adversarial network to further encourage the generative part to provide plausible organ delineations. Results highlight promising performance by outperforming state-of-the-art encoder-decoder schemes. Employed for the Combined Healthy Abdominal Organ Segmentation (CHAOS) challenge, our contributions reached the first rank for liver CT, liver MR and multi-organ MR segmentation competition categories. The proposed pipeline has the potential to support guidance for abdominal image interpretation, clinical decision making and patient care improvement while avoiding manual delineation efforts. Further work includes the evaluation of such deep models to other anatomical structures from the abdomen (pancreas, gallbladder) and the gastro-intestinal tract (esophagus, stomach, duodenum) arising from healthy and pathological subjects. More globally, our pipeline could be easily extended to other tissue types and imaging modalities to provide relevant clinical decision support. Methodological perspectives on unpaired cross-modality (CT, MR...) medical image segmentation with compact architectures could deserve further investigation to take advantage of multi-tasking properties of deep models as well as a larger amount of available data. Extending adversarial frameworks to incorporate anatomical priors using topological or shape constraints should also offer new insights to manage the strong diversity of abdominal organ appearance.

Conflicts of interest

None of the authors of this manuscript have any financial or personal relationships with other people or organizations that could inappropriately influence and bias this work.

References

[1] R. M. Summers, Progress in fully automated abdominal CT interpretation, American Journal of Roentgenology 207 (1) (2016) 67–79.

[2] A. E. Kavur, et al., CHAOS challenge : Combined (CT-MR) healthy abdominal organ segmentation, Medical Image Analysis 69 (2021).

[3] J. J. Cerrolaza, et al., Automatic multi-resolution shape modeling of multi-organ structures, Medical Image Analysis 25 (1) (2015) 11–21.

[4] Z. Xu, et al., Efficient multi-atlas abdominal segmentation on clinically acquired CT with SIMPLE context learning, Medical Image Analysis 24 (1) (2015) 18–27.

[5] R. Cuingnet, et al., Automatic detection and segmentation of kidneys in 3D CT images using random forests, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, 2012, pp. 66–74.

[6] G. Litjens, et al., A survey on deep learning in medical image analysis, Medical Image Analysis 42 (2017) 60–88.

[7] H. R. Roth, et al., Hierarchical 3D fully convolutional networks for multi-organ segmentation, arXiv preprint arXiv:1704.06382 (2017).

[8] E. Gibson, et al., Automatic multi-organ segmentation on abdominal CT with dense V-networks, IEEE Transactions on Medical Imaging 37 (8) (2018) 1822–1834.

[9] P. Hu, et al., Automatic abdominal multi-organ segmentation using deep convolutional neural network and time-implicit level sets, International Journal of Computer Assisted Radiology and Surgery 12 (3) (2017) 399–411.

[10] Y. Wang, et al., Abdominal multi-organ segmentation with organ-attention networks and statistical fusion, Medical Image Analysis 55 (2019) 88–102.

[11] J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, in: IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 3431–3440.

[12] O. Ronneberger, P. Fischer, T. Brox, U-Net: Convolutional networks for biomedical image segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, 2015, pp. 234–241.

[13] J. J. Cerrolaza, et al., Computational anatomy for multi-organ analysis in medical imaging: A review, Medical Image Analysis 56 (2019) 44–67.

[14] A. E. Kavur, et al., Comparison of semi-automatic and deep learning-based automatic methods for liver segmentation in living liver transplant donors, Diagnostic Interventional Radiology (2020) 11–21.

[15] X. Zhang, et al., Automatic liver segmentation from CT scans based on a statistical shape model, in: International Conference of the IEEE Engineering in Medicine and Biology, 2010, pp. 5351–5354.

[16] M. R. Sabuncu, et al., A generative model for image segmentation based on label fusion, IEEE Transactions on Medical Imaging 29 (10) (2010) 1714–1729.

[17] G. Yang, et al., Automatic kidney segmentation in CT images based on multi-atlas image registration, in: International Conference of the IEEE Engineering in Medicine and Biology Society, 2014, pp. 5538–5541.

[18] R. Wolz, et al., Automated abdominal multi-organ segmentation with subject-specific atlas generation, IEEE Transactions on Medical Imaging 32 (9) (2013) 1723–1730.

[19] Y. Huo, et al., Robust multi-contrast MRI spleen segmentation for splenomegaly using multi-atlas segmentation, IEEE Transactions on Biomedical Engineering 65 (2) (2017) 336–343.

[20] T. Tong, et al., Discriminative dictionary learning for abdominal multi-organ segmentation, Medical Image Analysis 23 (1) (2015) 92–104.

[21] Z. Xu, et al., Evaluation of six registration methods for the human abdomen on clinically acquired CT, IEEE Transactions on Biomedical Engineering 63 (8) (2016) 1563–1572.

[22] M. Bieth, et al., From large to small organ segmentation in CT using regional context, in: International Workshop on Machine Learning in Medical Imaging, 2017, pp. 1–9.

[23] R. Giraud, et al., An optimized patchmatch for multi-scale and multi-feature label fusion, NeuroImage 124 (2016) 770–782.

[24] P.-H. Conze, et al., Scale-adaptive supervoxel-based random forests for liver tumor segmentation in dynamic contrast-enhanced CT scans, International Journal of Computer Assisted Radiology and Surgery 12 (2) (2017) 223–233.

[25] Y. LeCun, et al., Gradient-based learning applied to document recognition, Proceedings of the IEEE 86 (11) (1998) 2278–2324.

[26] F. Milletari, N. Navab, S.-A. Ahmadi, V-net: Fully convolutional neural networks for volumetric medical image segmentation, in: International Conference on 3D Vision, 2016, pp. 565–571.

[27] Ö. Çiçek, et al., 3D U-Net: learning dense volumetric segmentation from sparse annotation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, 2016, pp. 424–432.

[28] Q. Dou, et al., 3D deeply supervised network for automatic liver segmentation from CT volumes, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, 2016, pp. 149–157.

[29] F. Lu, et al., Automatic 3D liver location and segmentation via convolutional neural network and graph cut, International Journal of Computer Assisted Radiology and Surgery 12 (2) (2017) 171–182.

[30] V. Iglovikov, A. Shvets, TernausNet: U-Net with VGG11 encoder pre-trained on imagenet for image segmentation, arXiv preprint arXiv:1801.05746 (2018).

[31] P.-H. Conze, S. Brochard, V. Burdin, F. T. Sheehan, C. Pons, Healthy versus pathological learning transferability in shoulder muscle MRI segmentation using deep convolutional encoder-decoders, Computerized Medical Imaging and Graphics 83 (2020).

[32] J. Yosinski, et al., How transferable are features in deep neural networks?, in: Advances in Neural Information Processing Systems, 2014, pp. 3320–3328.

[33] H. Choi, K. H. Jin, Fast and robust segmentation of the striatum using deep convolutional neural networks, Journal of Neuroscience Methods 274 (2016) 146–153.

[34] H. R. Roth, et al., A multi-scale pyramid of 3D fully convolutional networks for abdominal multi-organ segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, 2018, pp. 417–425.

[35] Z. Tu, X. Bai, Auto-context and its application to high-level vision tasks and 3D brain image segmentation, IEEE Transactions on Pattern Analysis and Machine Intelligence 32 (10) (2010) 1744–1757.

[36] P. Isola, et al., Image-to-image translation with conditional adversarial networks, in: IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 1125–1134.

[37] V. K. Singh, et al., Breast tumor segmentation and shape classification in mammograms using generative adversarial and convolutional neural network, Expert Systems with Applications 139 (2020).

[38] B. Lei, et al., Skin lesion segmentation via generative adversarial networks with dual discriminators, Medical Image Analysis 64 (2020).

[39] A. Boutillon, et al., Combining shape priors with conditional adversarial networks for improved scapula segmentation in MR images, in: IEEE International Symposium on Biomedical Imaging, 2020.

[40] P. Luc, C. Couprie, S. Chintala, J. Verbeek, Semantic segmentation using adversarial networks, arXiv preprint arXiv:1611.08408 (2016).

[41] Y. Huo, et al., Splenomegaly segmentation on multi-modal MRI using deep convolutional networks, IEEE Transactions on Medical Imaging 38 (5) (2019) 1185–1196.

[42] I. Goodfellow, et al., Generative adversarial nets, in: Advances in Neural Information Processing Systems, 2014, pp. 2672–2680.

[43] O. Russakovsky, et al., ImageNet large scale visual recognition challenge, International Journal of Computer Vision 115 (3) (2015) 211–252.

[44] V. Iglovikov, et al., TernausNetV2: Fully convolutional network for instance segmentation, arXiv preprint arXiv:1806.00844 (2018).

[45] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, arXiv preprint arXiv:1409.1556 (2014).

[46] S. S. M. Salehi, D. Erdogmus, A. Gholipour, Auto-context convolutional neural network (auto-net) for brain extraction in magnetic resonance imaging, IEEE Transactions on Medical Imaging 36 (11) (2017) 2319–2330.

[47] Y. Yan, et al., Cascaded multi-scale convolutional encoder-decoders for breast mass segmentation in high-resolution mammograms, in: IEEE International Engineering in Medicine and Biology Conference, 2019.

[48] K. Kamnitsas, et al., Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation, Medical Image Analysis 36 (2017) 61–78.

[49] Z. Zhou, et al., UNet++: A nested U-Net architecture for medical image segmentation, in: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, 2018, pp. 3–11.