Improving abdominal image segmentation with overcomplete shape priors

Amine Sadikine, Bogdan Badic, Jean-Pierre Tasu, Vincent Noblet, Pascal Ballet, Dimitris Visvikis, Pierre-Henri Conze

Computerized Medical Imaging and Graphics 113 (2024) 102356

Contents lists available at ScienceDirect

Computerized Medical Imaging and Graphics

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

Improving abdominal image segmentation with overcomplete shape priors

Amine Sadikine [a,b], Bogdan Badic [a,c], Jean-Pierre Tasu [a,d], Vincent Noblet [e], Pascal Ballet [a,b], Dimitris Visvikis [a], Pierre-Henri Conze [a,f,*]

[a] LaTIM UMR 1101, Inserm, Brest, 29200, France
[b] University of Western Brittany, Brest, 29200, France
[c] University Hospital of Brest, Brest, 29200, France
[d] University Hospital of Poitiers, Poitiers, 86000, France
[e] ICube UMR 7357, CNRS, Illkirch, 67412, France
[f] IMT Atlantique, Brest, 29200, France

ARTICLE INFO

Keywords:
Abdominal imaging
Semantic segmentation
Deep learning
Shape priors
Overcomplete representations

ABSTRACT

The extraction of abdominal structures using deep learning has recently experienced a widespread interest in medical image analysis. Automatic abdominal organ and vessel segmentation is highly desirable to guide clinicians in computer-assisted diagnosis, therapy, or surgical planning. Despite a good ability to extract large organs, the capacity of U-Net inspired architectures to automatically delineate smaller structures remains a major issue, especially given the increase in receptive field size as we go deeper into the network. To deal with various abdominal structure sizes while exploiting efficient geometric constraints, we present a novel approach that integrates into deep segmentation shape priors from a semi-overcomplete convolutional auto-encoder (S-OCAE) embedding. Compared to standard convolutional auto-encoders (CAE), it exploits an over-complete branch that projects data onto higher dimensions to better characterize anatomical structures with a small spatial extent. Experiments on abdominal organs and vessel delineation performed on various publicly available datasets highlight the effectiveness of our method compared to state-of-the-art, including U-Net trained without and with shape priors from a traditional CAE. Exploiting a semi-overcomplete convolutional auto-encoder embedding as shape priors improves the ability of deep segmentation models to provide realistic and accurate abdominal structure contours.

1. Introduction

Abdominal organ modeling from computed tomography (CT) or magnetic resonance (MR) imaging is of vital importance in computer-aided diagnosis (CAD), image guidance, surgery, or treatment planning purposes. However, semantic segmentation from abdominal images is a complex task owing to low contrast between target and surrounding soft tissues, morphological complexity of the abdominal anatomy (Fig. 1), target sparsity, and low signal-to-noise ratio (Boas and Fleischmann, 2012). Performing manual segmentation by clinicians is reliable but time-consuming, laborious, and prone to strong intra- and inter-expert labeling inconsistencies (Nelms et al., 2012). This explains the high interest towards robust fully-automated segmentation approaches.

Over the past few years, deep learning (DL) has reached many clinical applications (Litjens et al., 2017; Chen et al., 2022). This strongly impacted various medical image analysis tasks, including semantic segmentation, with remarkable performance (Conze et al., 2023). In particular, the U-Net convolutional architecture (Ronneberger et al., 2015) has gained tremendous popularity, with an excellent ability to extract large abdominal structures (e.g., liver) (Conze et al., 2021; Kavur et al., 2021). However, the capacity of U-Net inspired networks to segment smaller organs (e.g., kidneys, spleen, pancreas) or blood vessels remains challenging (Tekchandani et al., 2022a,b). Known as an ill-posed task, the solution remains highly undetermined or ill-conditioned. To alleviate this, dealing with additional priors formulated in a regularization fashion can restrict the space of feasible solutions to finally obtain more accurate and plausible delineation results.

Anatomical consistency can play an influential role in downstream segmentation applications. Since abdominal structures may vary in appearance, shape, topology, and spatial extent, incorporating prior knowledge into DL-based segmentation models is key (Nosrati and Hamarneh, 2016). Several works have thus exploited anatomical constraints such as topology specifications (Keshwani et al., 2020), edge polarity (Chen et al., 2017), or adjacency rules between

* Corresponding author at: IMT Atlantique, Brest, 29200, France.
E-mail address: pierre-henri.conze@imt-atlantique.fr (P.-H. Conze).

https://doi.org/10.1016/j.compmedimag.2024.102356
Received 25 May 2023; Received in revised form 11 December 2023; Accepted 6 February 2024
Available online 9 February 2024

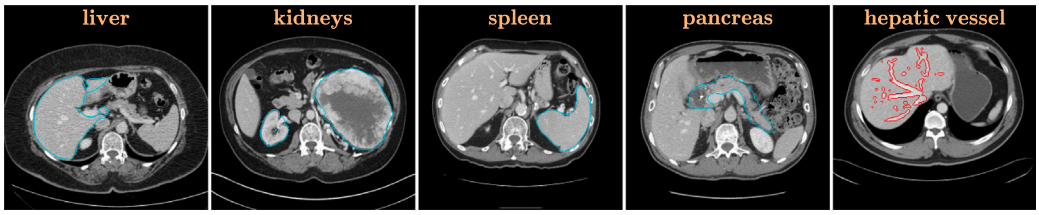

Fig. 1. Abdominal organ (liver, kidneys, spleen and pancreas) and (hepatic) blood vessel segmentation from CT scans with overlapping ground truth delineations, in blue and red respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

regions (Ganaye et al., 2019). Nevertheless, integrating shape constraints (Bohlender et al., 2021) remains one of the most commonly used and comprehensive strategies towards anatomically meaningful predictions. In this direction, conditional random fields (CRF) (Lafferty et al., 2001) were employed (Dou et al., 2016; Christ et al., 2016; Gao et al., 2016) as a post-processing step to incorporate high-level features over surrounding regions with the aim of refining segmentation masks forwarded by a convolutional neural network (CNN). In more recent works, promising performance was achieved by incorporating shape priors embedded at the loss function level, additionally to standard criteria such as cross-entropy or Dice. Thus, in cardiac image analysis, Oktay et al. proposed to add a penalty term to the global loss function dealing with the Euclidean distance between the projection of both predicted and ground truth segmentation masks in a convolutional auto-encoder (CAE) latent space (Oktay et al., 2017). This aimed at guiding the segmentation model to follow the global anatomy of the target. A similar approach was developed in Ravishankar et al. (2017) along with a cascade of segmentation U-Net models (Ronneberger et al., 2015) and a shape regularization network consisting of a denoising convolutional auto-encoder ensuring the projection of an arbitrary prediction onto a valid shape manifold. In an unsupervised learning context, a general and rapid approach for biomedical image segmentation suggested integrating rich probabilistic anatomical priors into a generative segmentation model towards brain cortical and subcortical tissue delineation from unpaired MR scans (Dalca et al., 2018). To go further, shape priors and adversarial learning were combined as regularizers to encourage the network to provide more realistic shapes for bone segmentation in MR modality (Boutillon et al., 2020, 2021, 2022). However, both segmentation and CAE models cannot fully deal with small structures, given the increase in receptive field size as we go deeper into the networks.

In the meantime, several approaches have recently focused on the exploration of more sophisticated deep architectures than the standard U-Net (Ronneberger et al., 2015) via multi-pathways (Kitrungrotsakul et al., 2019), residual connections (Yu et al., 2019; Oda et al., 2019), or overcomplete representations (Valanarasu et al., 2020). Overcomplete architectures (Valanarasu et al., 2020; Valanarasu and Patel, 2021) have appeared with the goal of projecting data onto higher dimensions in order to constrain the receptive field to be small and therefore improve the capacity of capturing finer low-level features details. The flexibility of overcomplete architectures to encode better anatomical structure shapes appears highly interesting for abdominal organ and vascular tree (Fig. 1) modeling purposes.

Following these directions, our contributions towards more efficient abdominal image segmentation are two-fold. First, our study evaluates the added value of standard CAE-based shape priors on various abdominal structures. Second, we propose to integrate into an end-to-end image segmentation framework a new semi-overcomplete convolutional auto-encoder (S-OCAE) with a multi-path encoder leveraging both meaningful, compact, and non-linear under and overcomplete shape representations. As an extension of preliminary results shown in Sadikine et al. (2022), the proposed fully automated method is evaluated and compared to state-of-the-art on abdominal organs and vessels using various publicly-available abdominal imaging datasets, covering both CT and MR modalities.

2. Proposed method

The central idea of our proposal is to improve the capability of a DL-based segmentation model to delineate abdominal structures (organs and vascular trees) by exploiting shape priors arising from a semi-overcomplete CAE, thus benefiting from more robust non-linear shape representations than those resulting from a standard CAE. Before reaching this, we gradually introduce all essential elements and concepts leading to the proposed segmentation pipeline (Fig. 2). Let us denote $\boldsymbol{x}$ as a discrete greyscale image and $\boldsymbol{y}$ as its corresponding ground truth segmentation mask. Supervised segmentation with DL consists of approximating a mapping function $\phi : \boldsymbol{x} \rightarrow \phi(\boldsymbol{x}) = \hat{\boldsymbol{y}}$ from q training samples $D = \{(\boldsymbol{x}_i, \boldsymbol{y}_i)\}_{i<q}$ with trainable weights $\boldsymbol{\Theta}_\phi$ by optimizing a loss function $\ell_\phi(\boldsymbol{y}, \hat{\boldsymbol{y}})$ through a stochastic optimizer (Kingma and Ba, 2014). We define ϕ as a U-Net shaped segmentation model, a convolutional encoder–decoder made of contracting and expansive paths with lateral skip-connections (Ronneberger et al., 2015).

2.1. Regularization and priors

Regularization plays a key role in machine learning, particularly in DL. A variety of techniques (Hanson and Pratt, 1988; Fan and Li, 2001; Yang and Hospedales, 2016; Sajjadi et al., 2016; Hernández-García and König, 2018) were developed to increase both the robustness and generalizability of a learned DL model applied to unseen data. Regularization methods are generally divided into several categories depending on whether they focus on data, network architecture, or loss function (i.e., by adding a regularization term). The regularization term $\mathcal{R}(\phi)$ consists of adding some prior knowledge to the model ϕ, and its regularization effect is achieved by adding the scaled regularizer $\lambda\mathcal{R}(\phi)$ to the loss function L to ensure further consistency between the prediction $\phi(\boldsymbol{x})$ and the target $\boldsymbol{y}$. The analytical expression of the resulting loss function is as follows:

$$\mathcal{L}(\phi|D) = \mathbb{E}_{(\boldsymbol{x},\boldsymbol{y})\sim P}\left[L(\boldsymbol{y}, \phi(\boldsymbol{x})) + \lambda\mathcal{R}(\phi)\right] \tag{1}$$

where P is the joint probability distribution over data from D and λ the weighting term which controls the importance of the regularization term $\mathcal{R}(\phi)$. However, in practice, the expected loss (Eq. (1)) cannot be directly minimized to estimate the model parameters $\hat{\boldsymbol{\Theta}}_\phi$ since the exact underlying probability distribution P is unknown which leads to approximate it empirically under the iid assumption of finite data samples D as:

$$\bar{\mathcal{L}}(\phi|D) = \frac{1}{|D|}\sum_{(\boldsymbol{x}_i,\boldsymbol{y}_i)\in D}\left[L(\boldsymbol{y}_i, \phi(\boldsymbol{x}_i)) + \lambda\mathcal{R}(\phi)\right] \tag{2}$$

The optimal parameters $\hat{\boldsymbol{\Theta}}_\phi$ are estimated through back-propagation by minimizing Eq. (2):

$$\hat{\boldsymbol{\Theta}}_\phi = \arg\min_{\boldsymbol{\Theta}_\phi} \bar{\mathcal{L}}(\phi|D) \tag{3}$$

Prior knowledge can be expressed via the regularization term $\mathcal{R}(\phi)$ according to some meaningful information relying on features extracted from $\boldsymbol{y}$ and $\hat{\boldsymbol{y}}$ in order to capture more accurate information of the underlying target, as outlined in the following.

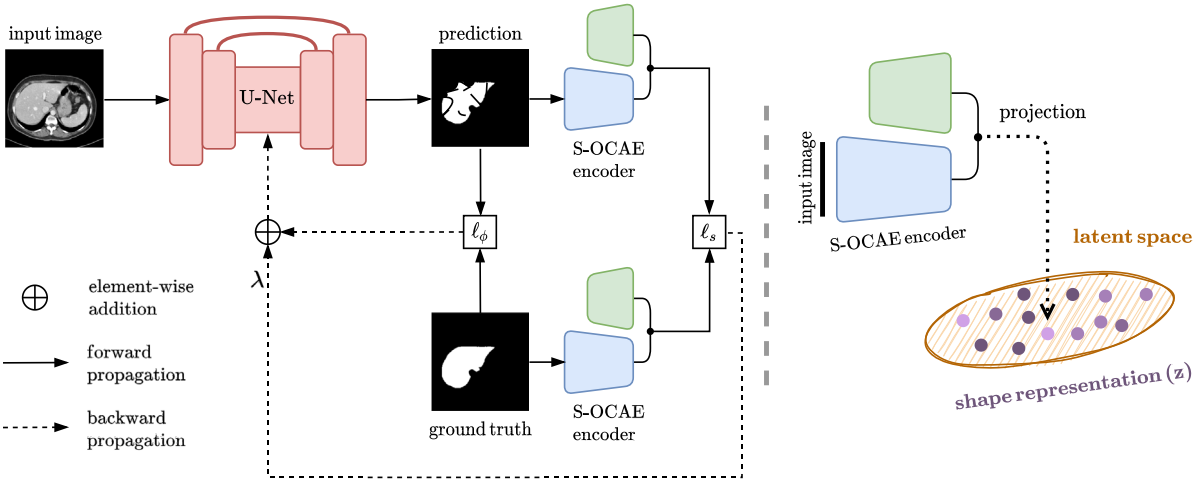

Fig. 2. Overview of the proposed pipeline. U-Net parameters are estimated by penalizing a segmentation loss ℓ_ϕ as well as a regularization term ℓ_s dealing with the similarity between projections of prediction and ground truth in a learned semi-overcomplete convolutional auto-encoder (S-OCAE) latent space.

2.2. Shape information

Commonly, segmentation networks ϕ are optimized by minimizing a loss function (e.g., cross-entropy), allowing to classify at the pixel-wise level without taking into account all contextual information regarding the object of the interest during back-propagation where shape can be unpredictable. In this context, incorporating high-level information constraints regarding the anatomy may overcome this limitation by providing a broader view of the problem.

The shape is considered one of the relevant sources of information in medical image analysis and plays a pivotal role among clinicians. Usually, shape representation for a given organ is carried out by an expert (e.g, radiologist) relying on spatial coordinates and prior knowledge of such anatomy and supplied as a ground truth segmentation mask. In a binary scenario, the resulting mask $\boldsymbol{y}_i$ is made of 0 (background) and 1 (target) as following:

$$\boldsymbol{y}_i(u,v) = \begin{cases} 1 & \text{if } (u,v) \in S \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

where S is the spatial domain of the structure of interest. Reaching shape descriptors from $\boldsymbol{y}_i$ deals with encoding features characterizing the shape into a compact representation. In this context, deep non-linear learned shape representations $\boldsymbol{z}$ are a powerful tool.

2.3. Standard undercomplete shape priors

To deal with shape priors, one can exploit a compact representation of the anatomy coming from ground truth segmentation masks using a CAE (Oktay et al., 2017; Boutillon et al., 2022) consisting of an encoder E (parameterized by $\boldsymbol{\Theta}_E$) and a decoder D. The encoder transforms a ground truth segmentation mask $\boldsymbol{y}$ through a cascade of convolutional, batch normalization (BN), non-linearity (e.g., ReLU), and max-pooling layers to a compressed representation $E(\boldsymbol{y};\boldsymbol{\Theta}_E) = \boldsymbol{z}$ referred to as latent code. Conversely, the decoder consists of decoding the information from the latent code $\boldsymbol{z}$ through a series of transposed convolutions, BN and non-linearity to reconstruct the input $\tilde{\boldsymbol{y}} = D(\boldsymbol{z}) = D \circ E(\boldsymbol{y})$. To train the CAE, we encourage it to accurately reconstruct the input segmentation masks by optimizing:

$$\ell_{CAE}(\boldsymbol{y},\tilde{\boldsymbol{y}}) \propto \sum_{i=0}^{q-1} \|\boldsymbol{y}_i - \tilde{\boldsymbol{y}}_i\|_2^2 \tag{5}$$

Let F_E^l and $F_D^{l'}$ be symmetrical hidden layer outputs at depths l and l' from E and D, defined respectively as $\boldsymbol{y}^l = F_E^l(\boldsymbol{y})$ and $\boldsymbol{z}^{l'} = F_D^{l'}(\boldsymbol{z})$, with $F_E^l = E^l \circ \cdots \circ E^1$ and $F_D^{l'} = D^{l'} \circ \cdots \circ D^1$. The spatial dimension of $\boldsymbol{y}^l$ and $\boldsymbol{z}^{l'}$ is far below the dimensionality of the input. For this reason, the CAE can be defined as an undercomplete auto-encoder. Furthermore, the first few layers F_E^l aim at capturing low-level features, and as l increases (i.e., as we go deeper in the network), the encoder extracts more high-level features and pays less attention to low-level ones due to down-sampling operations, leading to larger receptive fields (RF). Regarding where high-level and low-level features are located in E, the relationship between RF size and the level of features is not fixed. It can vary depending on the design of E and the specific task it is being used for. However, in general, it is common for high-level features to be generated by layers with larger RF, while low-level features are generated by layers with smaller RF. This is because larger RF enables the layer to see a broader context and extract more abstract features, while smaller RF allows the layer to focus on more specific and local details.

Once the standard CAE has been trained, shape priors can be incorporated into the segmentation pipeline by adding to the segmentation loss $L := \ell_\phi$ a shape regularization term $\mathcal{R}(\phi) := \ell_s$. A Euclidean distance between both latent shape representations (Oktay et al., 2017) is usually employed but other metrics can be exploited, such as:

$$\ell_s(\boldsymbol{y},\hat{\boldsymbol{y}}) \propto \sum_y (1 - \cos(E(\boldsymbol{y};\boldsymbol{\Theta}_E), E(\hat{\boldsymbol{y}};\boldsymbol{\Theta}_E))) \tag{6}$$

which measures the cosine distance between predicted outputs and ground truth masks in low-dimensional space. The total loss function is then expressed such that:

$$\bar{\mathcal{L}} = \ell_\phi(\boldsymbol{y}, \phi(\boldsymbol{x};\boldsymbol{\Theta}_\phi)) + \lambda \ell_s(\boldsymbol{y}, \phi(\boldsymbol{x};\boldsymbol{\Theta}_\phi)) \tag{7}$$

where $\boldsymbol{\Theta}_\phi$ denotes all trainable parameters of ϕ, $\lambda \in [0,\infty)$ is a empirically set hyper-parameter. Incorporating $\ell_s(.)$ into Eq. (7) constrains $\phi(\boldsymbol{x};\boldsymbol{\Theta}_\phi)$ to capture more anatomical shape features from targets. This strengthens plausible shape delineations in predicted masks and reduces false-positive detections.

2.4. Extension to semi-overcomplete shape priors

Compared to standard undercomplete CAE, an overcomplete CAE can be obtained by replacing max-pooling by up-sampling layers and vice-versa. In this scenario, intermediate layers $\boldsymbol{y}^l$ and $\boldsymbol{z}^{l'}$ are projected to a higher dimensionality than the input. This gives to the model a better flexibility to capture and encode low-level features since the RF is constricted. On the other hand, in the spatial sense, the overcomplete representation is too expensive to compute in memory. This limitation led us to design a new architecture that can be easily integrated into the segmentation pipeline (Fig. 2).

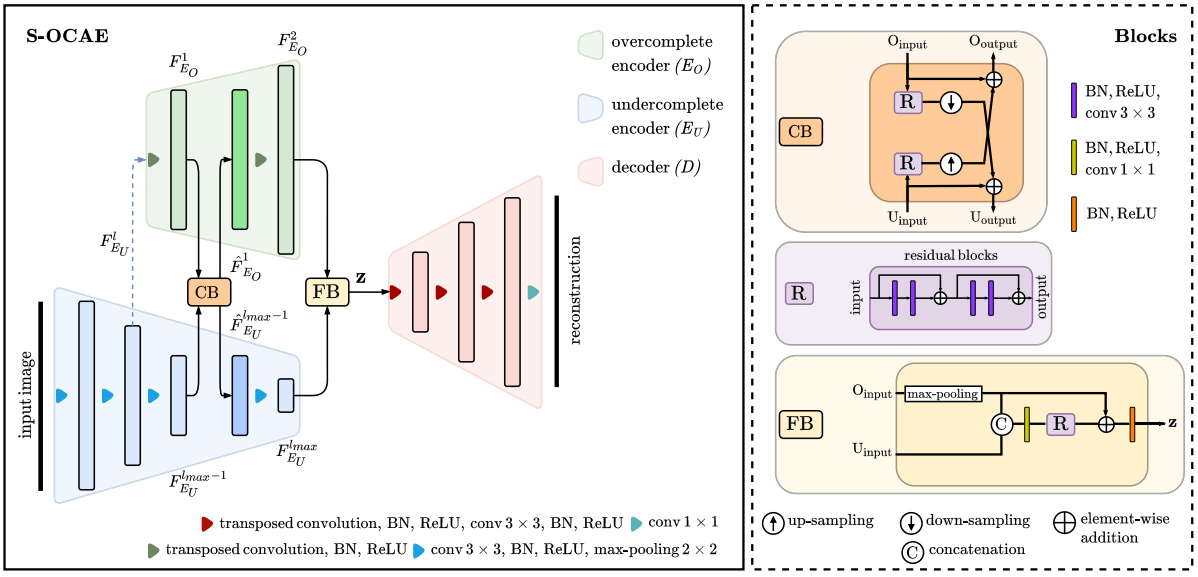

Fig. 3. The proposed S-OCAE network whose multi-path encoder is made of undercomplete (E_U) and overcomplete (E_O) branches includes communication (CB), fusion (FB) and residual blocks. The latent code z results from the fusion of the last layers of E_U (i.e., $F^{l_{max}}_{E_U}$) and E_O (i.e., $F^2_{E_O}$) using FB to combine the compact representation features $F^{l_{max}}_{E_U}$ with $F^2_{E_O}$. Then, z is decoded by the decoder (D) to reconstruct the input mask.

In the same spirit of Valanarasu and Patel (2021), we propose to extend traditional CAE by exploiting a multi-path encoder composed of an undercomplete branch E_U and an overcomplete branch E_O. Since an overcomplete encoder would be too expensive in terms of memory, we proposed to design a semi-overcomplete branch that takes as input hidden layer $F^l_{E_U}$ of E_U instead of the input image (Fig. 3). E_U allows to aggregate global information due to the expansion of RF in deeper $F^l_{E_U}$ layers but is not able to accurately encode small or tiny structures. To alleviate this issue, the branch E_O made of two layers $F^1_{E_O}$ and $F^2_{E_O}$ is added at $F^{l_{max}-1}_{E_U}$ and $F^{l_{max}}_{E_U}$ layer levels to restrict the RF. For a concrete example, please refer to Fig. 4 which allows to highlight the high-level details brought by the overcomplete layer $F^1_{E_O}$ compared to its under-complete counterpart layer $F^{l_{max}-1}_{E_U}$. Therefore, the resulting multi-path encoder manages both under and over-complete representations.

A specific block named communication block (CB) was designed to parallelly communicate residuals features between $F^{l_{max}-1}_{E_U}$ and $F^1_{E_O}$, with identity skip-connection, by adding to them residual features $[R_J]^{\downarrow\frac{1}{n}}$ and $[R_J]^{\uparrow n}$ respectively (Fig. 3). This results in new feature maps $\hat{F}^{l_{max}-1}_{E_U}$ and $\hat{F}^1_{E_O}$ to be forwarded in the network:

$$CB \triangleq \begin{cases} \hat{F}^{l_{max}-1}_{E_U} = F^{l_{max}-1}_{E_U} \oplus [R_J]^{\downarrow\frac{1}{n}} & \text{s.t. } R_0 = F^1_{E_O} \\ \hat{F}^1_{E_O} = F^1_{E_O} \oplus [R_J]^{\uparrow n} & \text{s.t. } R_0 = F^{l_{max}-1}_{E_U} \end{cases} \tag{8}$$

where $\oplus$ is the element-wise addition, $[\cdot]^{\downarrow\frac{1}{n}}$ and $[\cdot]^{\uparrow n}$ bilinear down-sampling and up-sampling operations with factors $\frac{1}{n}$ and n. $R_J = R_{J-1} \oplus \sum_{i=0}^{J} \mathcal{F}(R_i, \boldsymbol{w}_i)$ is a forward recursion to compute a serie of J full pre-activation (i.e., BN and ReLU come before convolution layers) residual units (He et al., 2016) where $\mathcal{F}$ is a residual function with trainable parameters $\boldsymbol{w}_i$ defined by a cascade of BN, ReLU, and convolution layers.

To combine the outputs from undercomplete and overcomplete encoding branches, respectively $F^{l_{max}}_{E_U}$ and $F^2_{E_O}$, a fusion block (FB) was introduced at the bottleneck of the network (Fig. 3) to compute a latent code z. FB projects $F^2_{E_O}$ to a lower dimension through max-pooling $\mathcal{M}^c_s(\cdot)$, with stride s=16 and pooling coefficient c=2, to enable the concatenation with $F^{l_{max}}_{E_U}$: $z' = \texttt{concat}[F^{l_{max}}_{E_U}, \mathcal{M}^c_s(F^2_{E_O})]$. Then, a 1×1 convolution operation is applied to z' with full pre-activation to get z'' with the same number of feature maps as $F^{l_{max}}_{E_U}$. R_J is finally applied (with $R_0 = z''$) to obtain the residual feature z'''. Finally, full post-activation was applied to $z''' \oplus \mathcal{M}^c_s(F^2_{E_O})$ to reach the final latent code z, considered as an undercomplete non-linear representation of the anatomical shape for a given input. In practice, CB and FB blocks employed $J = 2$ (Fig. 3), which was the optimal configuration during the network design.

3. Experiments

3.1. Imaging datasets

The proposed approach has been evaluated by relying on six miscellaneous and challenging public datasets: MSD (Antonelli et al., 2022), LiTS (Bilic et al., 2023), KiTS (Heller et al., 2021), CHAOS (Kavur et al., 2021), 3D-IRCADb (Soler et al., 2010) and DRIVE (Staal et al., 2004). Except for DRIVE, all these datasets were scanned from the abdomen with resolution 512 × 512 for CT scans (MSD, LiTS, KiTS, CHAOS) and 256 × 256 for MR images (CHAOS). As pre-processing, we sub-sampled all axial slices from CT scans by a factor of two. Diverse and numerous segmentation tasks were addressed in our experiments:

- **MSD**[1] (CT): liver (131 examinations), spleen (41), hepatic vessels (303), and pancreas (281) segmentation tasks were investigated. All volumes were acquired during the portal venous phase.
- **LiTS**[2] (CT): liver (131 examinations) manual segmentation was provided on data acquired from various clinical sites worldwide.
- **KiTS19**[3] (CT): 210 volumes with kidney ground truth contours were used.
- **CHAOS**[3] (CT/MR): contains 20 CT images acquired during portal venous phase with healthy annotated liver (including 15% of atypical liver shapes). 20 MR images from 2 different sequences (T1DUALin and T2SPIR) were also given with delineations for liver, left kidney, right kidney, and spleen.

[1] http://medicaldecathlon.com/
[2] https://competitions.codalab.org/competitions/17094
[3] https://kits19.grand-challenge.org/home/
[3] https://chaos.grand-challenge.org/

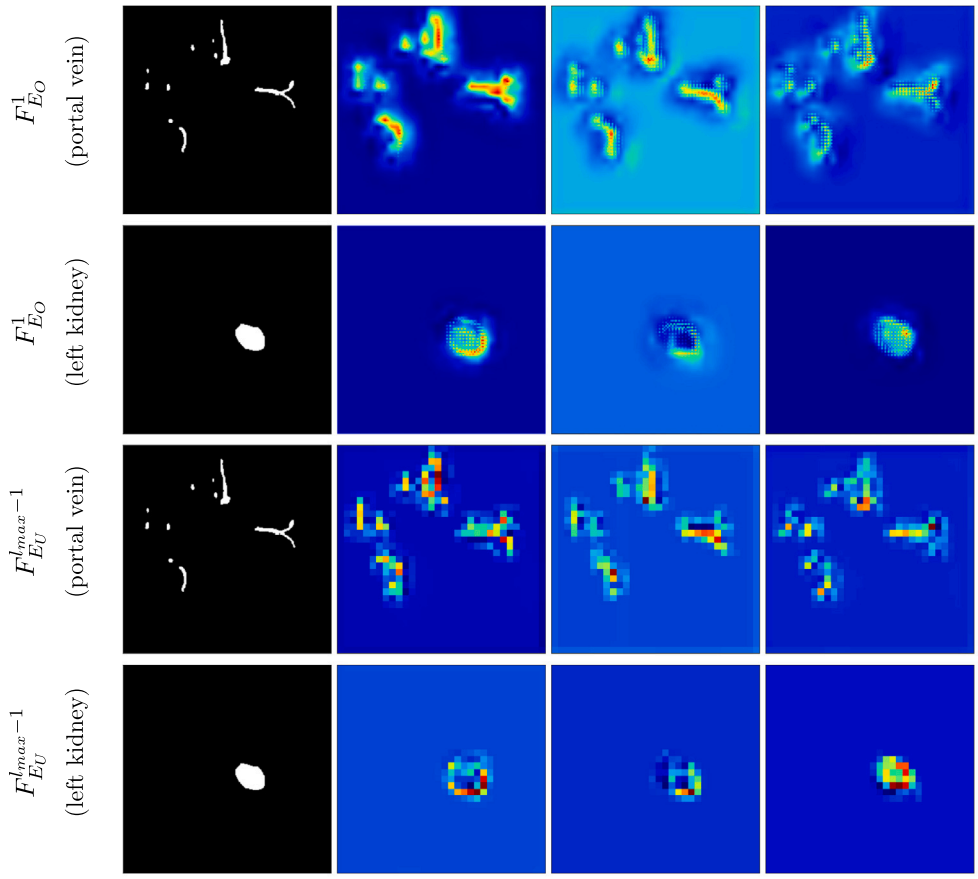

Fig. 4. Heatmap visualization of feature maps derived from overcomplete ($F^1_{E_O}$) and undercomplete ($F^{l_{max}-1}_{E_U}$) branches of S-OCAE with communication block (CB) extracted from two different anatomies: hepatic portal vein from 3D-IRCADb (Soler et al., 2010) and left kidney from CHAOS (Kavur et al., 2021).

- **3D-IRCADb**[4] (CT): contains CT scans with ground truth hepatic portal vein labels from 20 patients (10 women, 10 men) with liver tumors in 75% of cases. We resized axial slices from the liver bounding box in each CT scan.
- **DRIVE**[5]: retinal blood vessel dataset that contains 20 RGB images (including signs of mild early diabetic retinopathy) with associated ground truth. Color fundus photographs were resized from 584 × 565 to 512 × 512 pixels.

3.2. Comparison methods

In this work, U-Net (Ronneberger et al., 2015) using VGG13 with BN pre-trained on ImageNet as encoder (Conze et al., 2020, 2021) was compared with KiU-Net (Valanarasu et al., 2020) (for some datasets only due to the computational burden) which leverages overcomplete representation as well as the same U-Net architecture complemented with shape priors from either a standard CAE encoder (U-Net+CAE) (Oktay et al., 2017) or the proposed S-OCAE encoder with and without CB. In addition, it is worth noting that the CAE was implemented in a two-dimensional fashion, as for S-OCAE. This approach was followed to address memory limitations due to the overcomplete branch as well as to ensure comparability. All auto-encoders were trained from scratch. Then, their encoders were integrated into the segmentation pipelines for the regularization term (Eq. (7)) during training. We investigated the impact of the shape representation of our S-OCAE encoder with the CAE encoder on the segmentation model to learn and capture a more plausible anatomical shape of a specific target. Referring to Table 1, the S-OCAE encoder has a high demand in terms of memory (2.92 and 4.05MB for S-OCAE without and without CB respectively, compared to 0.38MB for standard CAE) owing to the projection in higher dimensions at the level of layers constituting the overcomplete branch. However, during inference, the segmentation model operates without any involvement of the encoder.

Table 1
Characteristics of CAE and S-OCAE without and with CB block encoders in terms of number of parameters in million (M) and megabytes (MB).

Encoders	#Params (M)	Params size (MB)
CAE	0.1	0.38
S-OCAE (w/o CB)	0.8	2.92
S-OCAE (w CB)	1.1	4.05

3.3. Implementation details

In the first stage, S-OCAE was trained using mean squared error as loss function (Eq. (5)), with Adam optimizer. The learning rate and batch size were set to 0.001 (0.0005) and 4 (32) for DRIVE (other datasets). The number of epochs for CHAOS, 3D-IRCADb, and DRIVE was set respectively to 150, 100 and 100. 50 epochs were considered for MSD, LiTS, and KiTS. In the second stage, the segmentation model was trained by optimizing Eq. (7) with weighted binary cross-entropy for ℓ_ϕ. The learning rate and batch size were set to 0.001 (0.0001) and 4 (16) for DRIVE (3D-IRCADb), and 0.0001 and 32 for all other datasets. Optimal λ factors (Eq. (7)) were empirically set for each dataset (values provided in Section 4). Data augmentation was applied during training

[4] https://www.ircad.fr/research/data-sets/
[5] https://drive.grand-challenge.org/

for both networks: rotation, shearing, and translation with additional flip, blurring and Gaussian noise for DRIVE. We used 5-fold cross-validation to validate the results. All networks were implemented in PyTorch.

3.4. Evaluation metrics

To evaluate the performance of our model (U-Net+S-OCAE) against existing approaches (U-Net (Ronneberger et al., 2015), KiU-Net (Valanarasu et al., 2020) and U-Net+CAE (Oktay et al., 2017)), we compared ground truth GT and prediction P, defined respectively by the surface S_{GT} and S_P, by means of the following metrics:

- **D**ice **S**imilarity **C**oefficient:

  $$\mathbf{DSC}(GT, P) = \frac{2|GT \cap P|}{|GT| + |P|}$$

- **A**bsolute **V**olume **D**ifference:

  $$\mathbf{AVD}(GT, P) = \frac{||GT| - |P||}{|GT|}$$

- **A**verage **S**ymmetric **S**urface **D**istance:

  $$\mathbf{ASSD}(GT, P) = \frac{1}{|S_{GT}| + |S_P|}\left(\sum_{s \in S_{GT}} d(s, S_P) + \sum_{s \in S_P} d(s, S_{GT})\right)$$

  where $d(s, S_k) = \min_{s_k \in S_k} \|s - s_k\|$ *and* $\|\cdot\|$ is the Euclidean distance.

- **H**ausdorff **D**istance:

  $$\mathbf{HD}(GT, P) = \max(h(GT, P), h(P, GT))$$

  where $h(A, B) = \max_{a \in A} \min_{b \in B} \|a - b\|$.

The predicted 2D images were stacked to convert them into the 3D domain. All evaluation metrics were therefore computed using 3D volumes.

4. Results and discussion

This section reports the results obtained for abdominal organ (Section 4.1) and vessel (Section 4.2) segmentation. Standard and over-complete architectures are compared through frequency analysis in Section 4.3. Section 4.4 provides an overall discussion regarding the integration of shape priors in DL-based segmentation pipelines, with a focus on abdominal imaging.

4.1. Abdominal organ segmentation

Abdominal organ segmentation results from CT and MR images are reported in Table 2 and 3, respectively. We focus on the automated delineation of liver from MSD (Antonelli et al., 2022), CHAOS (Kavur et al., 2021) and LiTS (Bilic et al., 2023) datasets, spleen and pancreas from MSD (Antonelli et al., 2022) as well as kidneys from the KiTS dataset (Heller et al., 2021). For computational reasons, comparisons with KiU-Net (Valanarasu et al., 2020) are employed for some organs only (Table 3). Fig. 5 provides a summary of qualitative results from the different targetted organs for visualization purposes.

Results in Table 2 show that our U-Net+S-OCAE model without (resp. with) CB performed well on CT datasets, achieving average DSC scores of 92.73% (resp. 92.51%) and outperforming both U-Net and U-Net+CAE. Furthermore, the proposed method without (resp. with) CB achieved the highest DSC score for liver segmentation with 95.80% (resp. 95.75%) on MSD, 97.20% (resp. 97.24%) on CHAOS and 95.90% (resp. 95.82%) on LiTS. The proposed approach without CB also obtained the highest DSC score for pancreas segmentation on MSD, with 76.53%. U-Net+S-OCAE with CB performed best for spleen segmentation on MSD with a DSC of 95.11% (+2% with respect to U-Net (Ronneberger et al., 2015)). Regarding AVD, the proposed model without (resp. with) CB had the lowest average score of 0.08 mm^3 (resp. 0.09 mm^3). Moreover, in terms of ASSD, the proposed model without CB got the lowest score with an average of 1.23 mm, while U-Net+S-OCAE with CB is the closest, with an average score of 1.27 mm. When focusing on the HD metric, it appears that the proposed model without CB achieved the lowest score with an average of 24.70 mm, while the same approach with CB is the closest, with an average of 24.73 mm. U-Net and U-Net+CAE (Oktay et al., 2017) were less efficient with HD about 25.8 mm and 43.7 mm, respectively. Overall, the proposed U-Net+S-OCAE model, without or with CB, performed well on the abdominal organ segmentation task from CT scans, achieving high scores across multiple metrics on multiple datasets. The models without/with CB revealed their effectiveness in liver, pancreas, and spleen segmentation. The proposed model without CB has the lowest scores in AVD, ASSD, and HD, suggesting that it provides more accurate delineations in terms of distance and volumetry with respect to ground truth delineations compared to U-Net and U-Net+CAE.

Regarding MR scans and as proven by the average metric values provided in Table 3, the proposed model, without and with CB, performed well across all abdominal organs on both T1DUALin and T2SPIR images. The proposed model with CB achieved on average (i.e., among the targeted organs) higher DSC scores than other models. In T2SPIR modality, our model with CB achieved the highest DSC score for the liver (90.67%) whereas its version without CB achieved a DSC score of 93.76% for liver segmentation in T1DUALin, outperforming U-Net+CAE (93.5%), U-Net (92.25%) and KiU-Net (90.77%). The proposed model (w/ CB) also achieved high DSC scores for left and right kidney segmentation, with scores of 88.05% and 90.07% in T1DUALin, 91.03% and 91.21% in T2SPIR. Results for AVD show that U-Net+S-OCAE with CB achieved the lowest scores for almost all organs: 0.04, 0.14, 0.12, and 0.17 mm^3 in T1DUALin for liver, left kidney, right kidney, and spleen respectively. The average AVD score across all organs was 0.11 mm^3, similarly to U-Net+CAE. The U-Net model performed relatively poorly, with an average AVD score of 0.16 mm^3. In terms of ASSD, the proposed U-Net+S-OCAE with CB achieved the best average ASSD (1.47 mm) against U-Net+CAE (1.59 mm) and U-Net (2.42 mm). Moreover, U-Net+S-OCAE without CB got an average HD of 22.79 mm which is lower than the HD obtained by U-Net (42.93 mm) and U-Net+CAE (28.50 mm). It is worth noting that KiU-Net performed worse in all organs and metrics in T1DUALin modality. Finally, concerning the communication block (CB), our results indicate that its integration into the S-OCAE architecture improves performance. This suggests that the CB plays an important role in the communication between both undercomplete and overcomplete pathways layers by leading to better feature aggregation. Qualitative results (Fig. 5) reveal that U-Net+S-OCAE globally reached better contour adherence and shape consistency than U-Net and U-Net+CAE. The ability of U-Net+S-OCAE to mimic expert annotations is all the more visible for the pancreas and kidneys, especially when CB is employed.

4.2. Abdominal vessel segmentation

Quantitative results (Table 4) indicate that adding shape priors to U-Net significantly improves delineation results in all assessment metrics despite the complex multi-scale geometry with decreasing diameters and contrast along tree-like networks, which is related to the intrinsic nature of blood vessels. Moreover, whatever the dataset, the proposed method outperforms U-Net+CAE in DSC, AVD, ASSD, and HD except for 3D-IRCADb where U-Net+CAE reaches a slightly better DSC. The main improvement from U-Net+CAE to U-Net+S-OCAE with CB block is observed at the level of the HD metric, which undergoes a rise of 3.08% for DRIVE and 4.03% for 3D-IRCADb. The attenuation of the largest segmentation errors suggests that vessel contours provided by U-Net+S-OCAE are more suitable with respect to clinical requirements. Qualitative illustrations provided in Fig. 6 further support these findings and highlight the ability of U-Net+S-OCAE to disintegrate thin vessels rather than erroneously agglomerate them. Better contour adherence can also be noticed when using our approach with CB, regardless of vessel size.

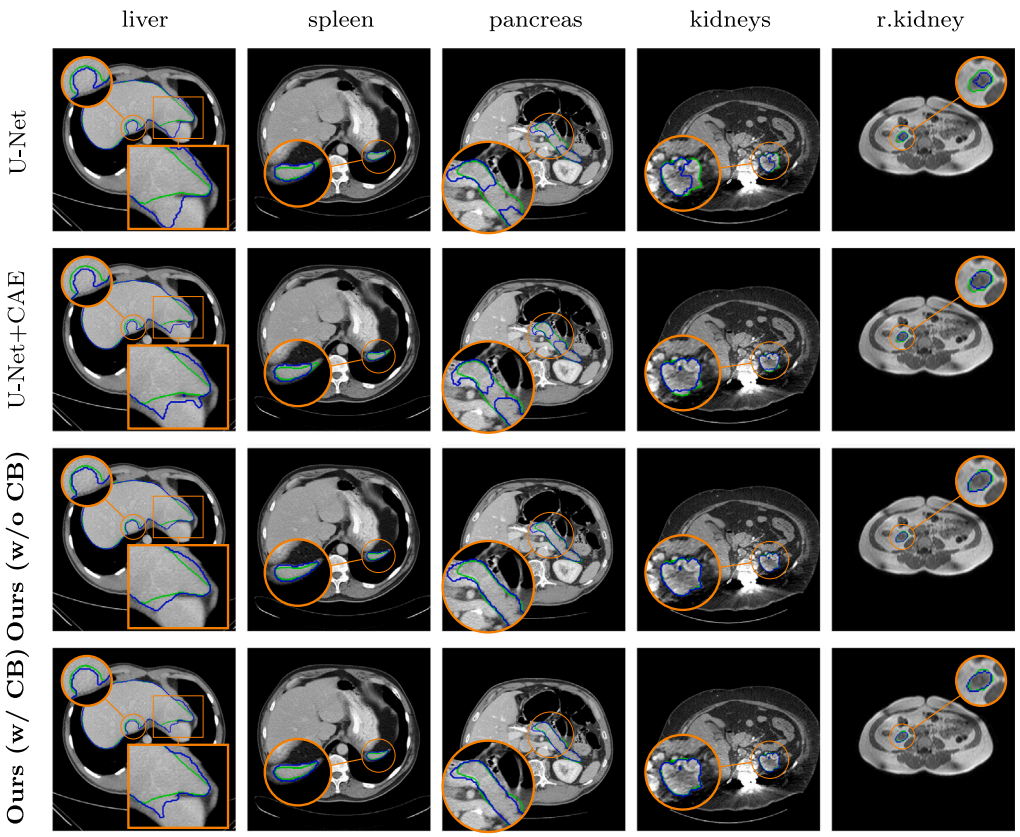

Fig. 5. Qualitative results of abdominal organ segmentation from CT and MR T1DUALin scans. The first to last columns correspond to the liver (CHAOS (Kavur et al., 2021)), spleen, pancreas (MSD (Antonelli et al., 2022)), kidneys (KiTS (Heller et al., 2021)), and right kidney (CHAOS). Ground truth and predicted contours are respectively in green and blue. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

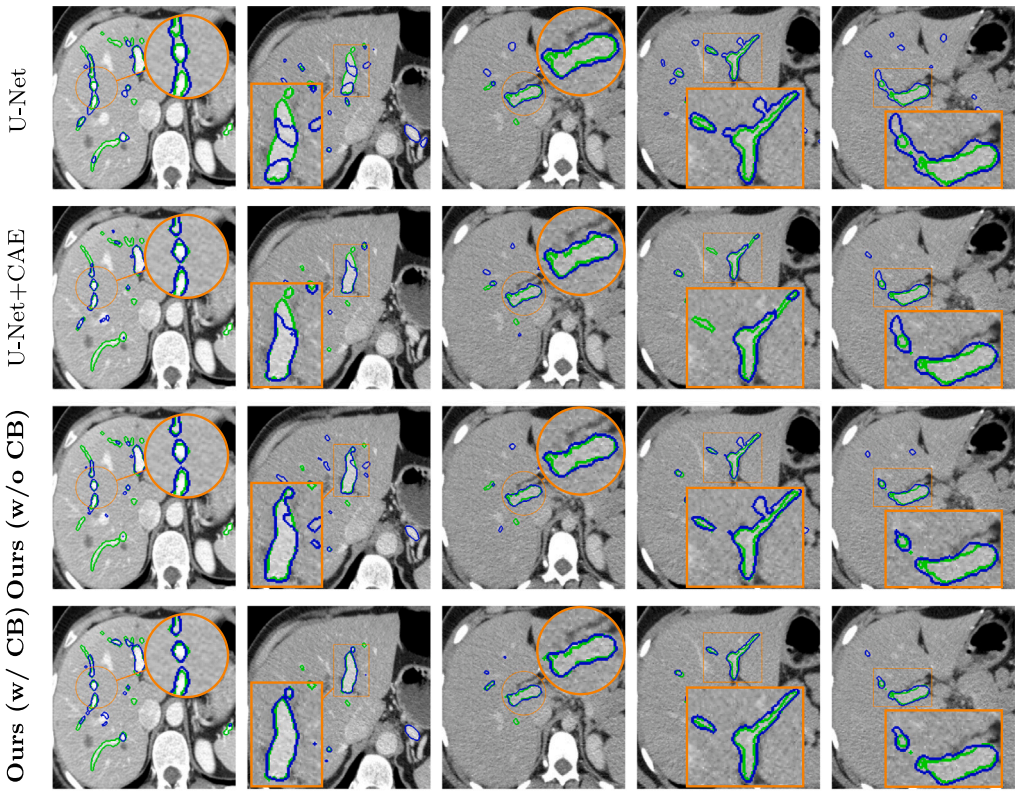

Fig. 6. Hepatic portal veins segmentation results (3D-IRCADb (Soler et al., 2010)). Ground truth and predicted contours are respectively in green and blue. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 2
Abdominal organ segmentation from CT scans using U-Net (Ronneberger et al., 2015), U-Net+CAE (Oktay et al., 2017) and the proposed U-Net+S-OCAE without (w/o) and with (w/) communication block (CB). Experiments focus on liver, spleen and pancreas from MSD (Antonelli et al., 2022), liver from CHAOS (Kavur et al., 2021) and LiTS (Bilic et al., 2023), kidney from KiTS (Heller et al., 2021). Best results in bold, second best results underlined.

Metrics	Models	MSD (Antonelli et al., 2022)			CHAOS (Kavur et al., 2021)	LiTS (Bilic et al., 2023)	KiTS (Heller et al., 2021)	avg
		liver ($\lambda = 40$)	spleen ($\lambda = 40$)	pancreas ($\lambda = 10$)	liver ($\lambda = 10$)	liver ($\lambda = 40$)	kidney ($\lambda = 50$)	
DSC↑ score (%)	U-Net (Ronneberger et al., 2015)	94.99 ± 0.74	93.11 ± 1.21	73.59 ± 4.84	96.96 ± 0.21	95.03±0.57	95.02 ± 0.65	91.45
	U-Net+CAE (Oktay et al., 2017)	95.68 ± 0.60	94.77 ± 0.56	75.00 ± 3.45	97.16 ± 0.26	95.77±0.64	95.54 ± 0.58	92.32
	Ours (w/o CB)	**95.80** ± 0.63	94.99 ± 0.73	**76.53** ± 2.55	97.20 ± 0.33	**95.90**±0.55	**95.97** ± 0.52	**92.73**
	Ours (w/ CB)	95.75 ± 0.63	**95.11** ± 0.24	75.42 ± 3.79	**97.24** ± 0.26	95.82±0.51	95.72 ± 0.46	92.51
AVD↓ dist. (mm^3)	U-Net (Ronneberger et al., 2015)	0.07 ± 0.01	0.13 ± 0.03	0.48 ± 0.24	0.03 ± 0.01	0.07 ± 0.02	0.07 ± 0.02	0.14
	U-Net+CAE (Oktay et al., 2017)	**0.04** ± 0.01	0.05 ± 0.02	0.36 ± 0.18	**0.02** ± 0.01	0.04 ± 0.01	0.04 ± 0.01	0.09
	Ours (w/o CB)	**0.04** ± 0.01	0.06 ± 0.02	**0.30** ± 0.10	**0.02** ± 0.01	**0.03** ± 0.01	**0.03** ± 0.01	**0.08**
	Ours (w/ CB)	0.04 ± 0.02	**0.05** ± 0.01	0.36 ± 0.20	**0.02** ± 0.01	0.04 ± 0.01	0.04 ± 0.01	0.09
ASSD↓ dist. (mm)	U-Net (Ronneberger et al., 2015)	2.01 ± 0.48	1.09 ± 0.46	2.94 ± 0.64	0.96 ± 0.18	1.98 ± 0.51	0.95 ± 0.33	1.66
	U-Net+CAE (Oktay et al., 2017)	1.50 ± 0.34	0.62 ± 0.12	2.67 ± 0.47	0.86 ± 0.13	1.41 ± 0.26	0.88 ± 0.22	1.32
	Ours (w/o CB)	**1.39** ± 0.24	0.53 ± 0.10	**2.45** ± 0.47	**0.79** ± 0.13	1.44 ± 0.31	**0.80** ± 0.23	**1.23**
	Ours (w/ CB)	1.40 ± 0.22	**0.51** ± 0.07	2.58 ± 0.50	0.80 ± 0.09	**1.41** ± 0.25	0.90 ± 0.20	1.27
HD↓ dist. (mm)	U-Net (Ronneberger et al., 2015)	59.89 ± 15.5	32.93 ± 23.2	30.92 ± 4.18	45.21 ± 19.8	60.59±15.1	32.82 ± 7.47	43.73
	U-Net+CAE (Oktay et al., 2017)	33.70 ± 7.43	11.69 ± 4.84	23.06 ± 2.61	26.65 ± 5.95	35.81±6.02	23.74 ± 2.48	25.78
	Ours (w/o CB)	34.83 ± 5.42	10.82 ± 4.36	**22.51** ± 2.25	**23.72** ± 5.77	34.39±5.55	21.92 ± 2.02	**24.70**
	Ours (w/ CB)	**32.36** ± 5.21	**9.25** ± 2.42	22.82 ± 1.93	29.43 ± 6.75	**34.10**±6.28	**20.40** ± 2.49	24.73

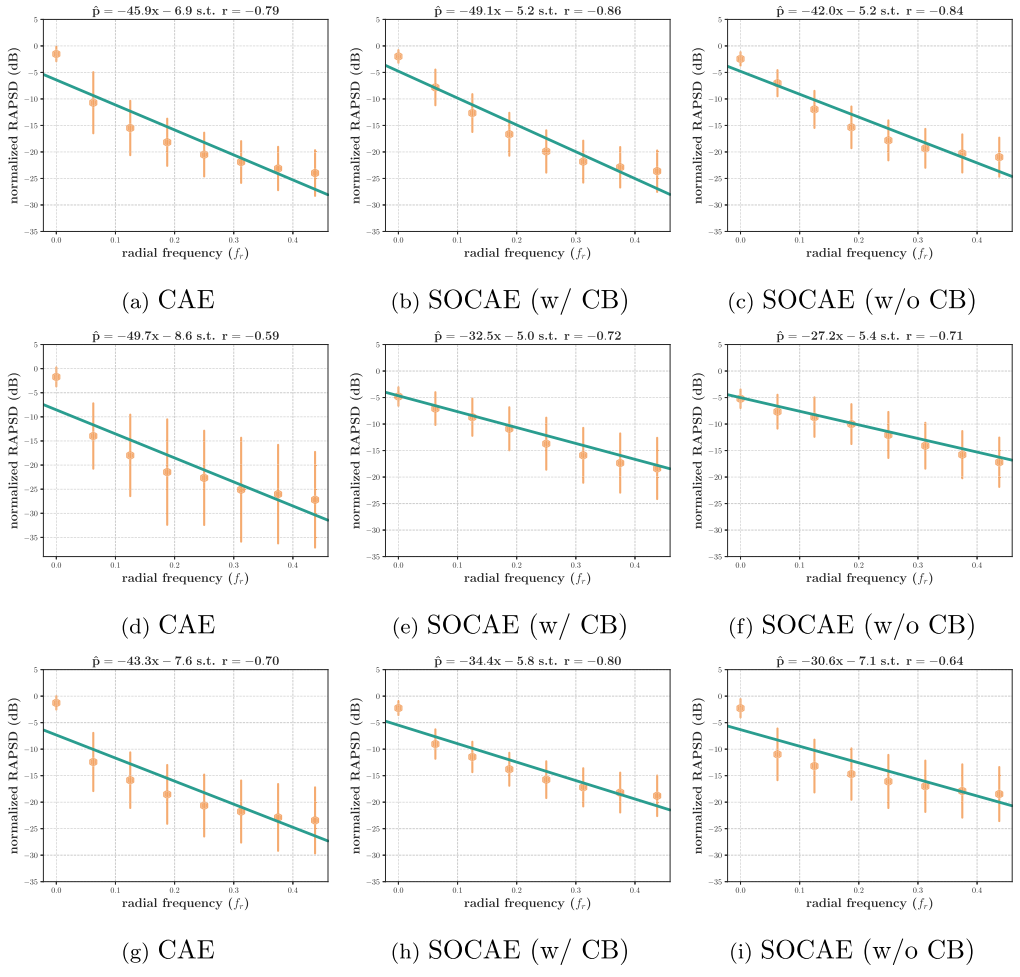

Fig. 7. Normalized RAPSD expressed in decibels (dB) for different organs at shape representation (latent code) level from CAE, S-OCAE with and without CB. The resulting normalized RAPSD data points at each radial frequency (f_r) are compressed by standard deviation error bars, symmetrically around the mean estimate, fitted to linear model $\hat{p}$ with r its Pearson correlation coefficient. (a), (b) and (c) summarize the results from liver shape latent code in CT from CHAOS (Kavur et al., 2021). The second row (d), (e), and (f) deals with the left kidney in MR from CHAOS. The last row (g), (h), and (i) focuses on the liver portal veins from 3D-IRCADb (Soler et al., 2010).

Table 3
Abdominal organ segmentation on MR T1DUALin and T2SPIR images from the CHAOS dataset (Kavur et al., 2021) using U-Net (Ronneberger et al., 2015), U-Net+CAE (Oktay et al., 2017) and the proposed U-Net+S-OCAE without (w/o) and with (w/) communication block (CB). Experiments focus on the liver, left kidney, right kidney and spleen. The best results are in bold, second best results are underlined.

Metrics	Models	CHAOS (Kavur et al., 2021)								avg
		T1DUALin				T2SPIR				
		liver ($\lambda = 50$)	l. kidney ($\lambda = 40$)	r. kidney ($\lambda = 10$)	spleen ($\lambda = 60$)	liver ($\lambda = 20$)	l. kidney ($\lambda = 20$)	r. kidney ($\lambda = 60$)	spleen ($\lambda = 50$)	
DSC↑ score (%)	U-Net (Ronneberger et al., 2015)	92.25 ± 2.07	87.30 ± 1.83	86.95 ± 2.64	83.81±1.97	90.61 ± 5.87	90.10 ± 1.44	91.34 ± 1.16	84.05 ± 10.1	88.30
	KiU-Net (Valanarasu et al., 2020)	90.77 ± 4.62	81.43 ± 8.68	82.53 ± 8.93	78.80±9.60	–	–	–	–	–
	U-Net+CAE (Oktay et al., 2017)	93.50 ± 1.31	87.95 ± 2.52	89.07 ± 3.01	85.71±3.95	89.87 ± 8.43	90.75 ± 0.57	91.10 ± 2.96	86.77 ± 9.00	89.34
	Ours (w/o CB)	**93.76** ± 0.88	87.18 ± 2.90	88.42 ± 3.76	86.30±2.79	89.58 ± 8.63	90.12 ± 1.28	**91.59** ± 2.70	**87.14** ± 8.74	89.26
	Ours (w/ CB)	93.49 ± 0.95	**88.05** ± 2.94	**90.07** ± 1.70	**86.44**±1.60	**90.67** ± 6.62	**91.03** ± 0.99	91.21 ± 2.68	86.80 ± 9.13	**89.72**
AVD↓ dist. (mm^3)	U-Net (Ronneberger et al., 2015)	0.09 ± 0.04	0.18 ± 0.07	0.15 ± 0.06	0.28 ± 0.05	0.10 ± 0.07	0.13 ± 0.02	0.12 ± 0.03	0.24 ± 0.13	0.16
	KiU-Net (Valanarasu et al., 2020)	0.11 ± 0.11	0.22 ± 0.11	0.20 ± 0.10	0.29 ± 0.09	–	–	–	–	–
	U-Net+CAE (Oktay et al., 2017)	0.04 ± 0.02	0.14 ± 0.05	**0.09** ± 0.03	0.19 ± 0.08	0.08 ± 0.09	0.10 ± 0.02	0.09 ± 0.03	0.16 ± 0.09	**0.11**
	Ours (w/o CB)	0.04 ± 0.02	0.16 ± 0.06	0.12 ± 0.06	0.17 ± 0.05	0.09 ± 0.09	0.10 ± 0.02	**0.08** ± 0.04	**0.14** ± 0.09	**0.11**
	Ours (w/ CB)	**0.04** ± 0.01	**0.14** ± 0.03	0.12 ± 0.04	**0.17** ± 0.04	**0.08** ± 0.08	**0.09** ± 0.04	0.10 ± 0.03	0.15 ± 0.09	**0.11**
ASSD↓ dist. (mm)	U-Net (Ronneberger et al., 2015)	2.72 ± 2.19	1.77 ± 0.31	1.97 ± 0.96	3.17 ± 1.37	2.66 ± 1.61	1.49 ± 0.61	1.41 ± 0.70	4.17 ± 4.75	2.42
	KiU-Net (Valanarasu et al., 2020)	4.19 ± 4.06	2.71 ± 1.33	3.16 ± 1.56	5.43 ± 6.39	–	–	–	–	–
	U-Net+CAE (Oktay et al., 2017)	1.58 ± 0.58	1.50 ± 0.51	1.59 ± 1.04	2.03 ± 1.15	2.53 ± 2.03	**1.13** ± 0.19	1.23 ± 0.93	1.15 ± 0.66	1.59
	Ours (w/o CB)	**1.42** ± 0.30	1.41 ± 0.39	**1.30** ± 0.66	1.88 ± 0.98	2.58 ± 2.01	1.21 ± 0.27	**0.97** ± 0.60	1.09 ± 0.70	1.48
	Ours (w/ CB)	1.61 ± 0.38	**1.37** ± 0.43	1.46 ± 0.93	**1.87** ± 0.69	**2.23** ± 1.35	1.15 ± 0.25	1.04 ± 0.59	**1.02** ± 0.57	**1.47**
HD↓ dist. (mm)	U-Net (Ronneberger et al., 2015)	55.56 ± 25.8	23.08 ± 8.99	35.26 ± 22.7	56.97±35.8	64.39 ± 31.1	30.90 ± 12.4	36.30 ± 11.6	40.94 ± 18.3	42.93
	KiU-Net (Valanarasu et al., 2020)	114.3 ± 38.2	96.93 ± 44.0	99.97 ± 34.5	116.4±38.3	–	–	–	–	–
	U-Net+CAE (Oktay et al., 2017)	35.73 ± 12.9	31.89 ± 27.1	24.82 ± 20.3	27.69±23.7	41.71 ± 19.1	27.72 ± 10.3	18.15 ± 11.7	20.27 ± 4.26	28.50
	Ours (w/o CB)	**27.41** ± 5.78	**15.17** ± 8.53	**11.87** ± 15.2	**19.60**±7.60	43.38 ± 18.0	27.22 ± 12.9	12.88 ± 5.16	24.80 ± 17.2	**22.79**
	Ours (w/ CB)	36.28 ± 6.29	15.50 ± 7.98	21.95 ± 15.2	21.13±8.30	**36.56** ± 6.63	**26.77** ± 12.3	**12.81** ± 5.27	**16.10** ± 5.66	23.39

4.3. Frequency analysis of shape representations

Frequency analysis aims at evaluating the frequencies of different features extracted from an image (e.g., edges, textures, shapes). This information may indirectly provide insight into how the RF of the different layers within a CNN architecture process the input image. In this study, in order to analyze the influence of the RF on both CAE and S-OCAE latent codes $\boldsymbol{z}$, we propose to exploit the radially averaged power spectral density (RAPSD). In this method, the power spectrum is calculated on the features learned by a CNN network and averaged over concentric circles of increasing radius, with the center of the circle corresponding to the center of the feature maps. It is also noted that low-level features tend to have higher frequency content, typically extracted in earlier layers of a CNN. Conversely, high-level features have lower frequency content generated in deeper CNN layers. In general, the frequency content of a feature depends on the scale at which it appears in the image. Low-level features are often associated with small scales and tend to contain higher frequencies. In contrast, high-level features, which are often associated with larger scales, tend to have lower frequency content. The resulting RAPSD plots can indirectly reveal how the RF is evolving through the CAE and the proposed S-OCAE shape representations.

In practice, we propose to fit a linear regression model to the RAPSD of CAE and S-OCAE with and without CB, trained on ground truth masks of different organs, in order to understand their general patterns regarding frequencies. This can provide insight into how such a dataset (i.e., organ shape, size) affects the encoding of shape representations. For this study, we considered different anatomical structures: two organs (liver and left kidney from CHAOS (Kavur et al., 2021)) and hepatic portal veins from 3D-IRCADb (Soler et al., 2010). Results are depicted in Fig. 7. We observed that the feature representations $\boldsymbol{z}$ of the S-OCAE with/without CB have a lower variability (i.e., standard deviation) of power values than the standard CAE, regardless of radial frequency (f_r), which implies compactness of the power spectrum in the proposed S-OCAE shape encoding. On the other, the slope of the resulting linear model indicates how power values change when

Table 4
Vessel segmentation using U-Net (Ronneberger et al., 2015), KiU-Net (Valanarasu et al., 2020), U-Net+CAE (Oktay et al., 2017) and the proposed U-Net+S-OCAE without (w/o) and with (w/) communication block (CB). Experiments focus on retinal vessels in fundus color images (DRIVE (Staal et al., 2004)), portal veins and hepatic vessels from CT scans (3D-IRCADb (Soler et al., 2010), MSD (Antonelli et al., 2022)). Best results in bold, second best results underlined.

Metrics	Models	DRIVE (Staal et al., 2004)	**3D-IRCADb** (Soler et al., 2010)	MSD (Antonelli et al., 2022)	avg
		retinal vasculitis ($\lambda = 40$)	portal veins($\lambda = 60$)	hepatic Vessel ($\lambda = 60$)	
DSC↑ score (%)	U-Net (Ronneberger et al., 2015)	75.14 ± 1.68	54.81 ± 1.87	52.52 ± 3.99	60.82
	KiU-Net (Valanarasu et al., 2020)	–	43.81 ± 1.70	–	–
	U-Net+CAE (Oktay et al., 2017)	79.25 ± 0.48	**59.26** ± 1.33	55.39 ± 2.33	64.63
	Ours (w/o CB)	77.75 ± 0.66	58.65 ± 1.09	**56.14** ± 2.43	64.18
	Ours (w/ CB)	**79.44** ± 0.65	59.14 ± 1.08	56.01 ± 2.89	**64.86**
AVD↓ dist. (mm^3)	U-Net (Ronneberger et al., 2015)	0.38 ± 0.13	0.62 ± 0.21	1.21 ± 0.64	0.74
	KiU-Net (Valanarasu et al., 2020)	–	1.09 ± 0.57	–	–
	U-Net+CAE (Oktay et al., 2017)	0.10 ± 0.05	0.33 ± 0.14	0.75 ± 0.39	0.39
	Ours (w/o CB)	0.13 ± 0.05	0.37 ± 0.12	**0.65** ± 0.29	0.38
	Ours (w/ CB)	**0.08** ± 0.02	**0.31** ± 0.15	0.71 ± 0.41	**0.37**
ASSD↓ dist. (mm)	U-Net (Ronneberger et al., 2015)	0.71 ± 0.24	4.25 ± 0.42	3.02 ± 0.46	2.66
	KiU-Net (Valanarasu et al., 2020)	–	5.63 ± 0.71	–	–
	U-Net+CAE (Oktay et al., 2017)	**0.54** ± 0.11	3.95 ± 0.21	2.82 ± 0.21	2.44
	Ours (w/o CB)	1.22 ± 0.19	**3.79** ± 0.31	**2.79** ± 0.28	2.60
	Ours (w/ CB)	0.54 ± 0.15	3.84 ± 0.30	2.80 ± 0.26	**2.39**
HD↓ dist. (mm)	U-Net (Ronneberger et al., 2015)	32.90 ± 1.85	67.60 ± 4.74	57.31 ± 12.2	52.60
	KiU-Net (Valanarasu et al., 2020)	–	86.32 ± 2.83	–	–
	U-Net+CAE (Oktay et al., 2017)	31.82 ± 2.78	54.48 ± 5.38	40.92 ± 2.57	42.41
	Ours (w/o CB)	30.08 ± 2.30	54.85 ± 8.86	**40.77** ± 3.72	41.90
	Ours (w/ CB)	**28.74** ± 0.58	**50.45** ± 8.99	41.88 ± 3.57	**40.36**

increasing f_r. In this context, the absolute value of the slopes for left kidney and liver portal veins indicates that the power value change is comparatively small for S-OCAE (with/without CB) compared to their counterparts CAE. This reflects the presence of more details in the latent code of the proposed networks. In contrast, the slopes for the liver are somewhat close between S-OCAE and CAE, which amounts to the size of the liver shape which is relatively large compared to the other organs. This leads us to conclude that S-OCAE is more desirable when trying to capture small anatomical structures.

4.4. Overall discussion

The results from CT and MR abdominal images show that the proposed U-Net+S-OCAE model performs well on both image modalities. However, the results for the two modalities differ slightly. The U-Net+S-OCAE model without CB performed better on CT datasets, achieving higher average scores and outperforming both U-Net and U-Net+CAE. On the other hand, the U-Net+S-OCAE model with CB performed better on MR scans, achieving higher average scores for both T1DUALin and T2SPIR images. The choice of imaging modality influences the appearance of abdominal organs and the contrast between tissues and hence impacts the accuracy of shape priors.

An overcomplete shape encoding results in a representation that captures more information than standard undercomplete encoders. When relying on overcomplete encoding to obtain shape representations, the encoder has the advantage of capturing rich complex shape information relationships and patterns from the input data. This overcomplete representation is handy for small objects as it captures more details (Fig. 4) and variations in the shape that may not be noticeable with an undercomplete representation. In contrast, overcomplete shape encoding can result in higher dimensionality, leading to increased computational and storage requirements. A trade-off should be considered when designing an overcomplete architecture for a given task. Overcomplete representations can provide more expressive features but at the cost of increased memory consumption. Therefore, it is important to balance the benefits of using overcomplete representations against the cost of increased memory usage when designing the S-OCAE architecture. In another side regarding data, the use of pathological data highlights the importance of further comparative studies between normal and abnormal anatomies. This would allow for a better understanding of the variations and abnormalities present in the pathological data and how they differ from healthy anatomies.

More generally, incorporating shape priors into a segmentation pipeline dedicated to abdominal organs and vessels can provide several benefits. First, the prior knowledge guiding the network towards the expected shape can help the algorithm better differentiating foreground from background pixels, thus leading to more accurate delineations results. This can be especially useful in medical imaging applications where accurate segmentation is critical for diagnosis and treatment planning. Second, using shape priors can also reduce the amount of annotated data needed to train the model by making the training process more efficient. This is of high interest for scenarios where annotated data is scarce and expensive to collect and annotate. Third, incorporating shape priors into the segmentation pipeline can improve the robustness of the algorithm to noise and variability in the input data. This last property is helpful since the quality of medical data can significantly vary due to factors such as patient motion or imaging equipment.

However, there are also limitations to the use of shape priors in DL-based segmentation pipelines. One significant limitation is that the performance highly depends on the quality and reliability of the shape priors being used. If the priors are incorrect or unreliable, they can harm the performance of the model and lead to inaccurate segmentation results. It is, therefore, important to carefully consider the sources and reliability of the shape priors being used and to validate the results of the model to ensure that the resulting contours remain accurate and reliable. Another potential limitation is that the approach may need to be more flexible to handle significant variations in the shape and appearance of the objects being segmented.

5. Conclusion

In conclusion, the proposed method using a semi-overcomplete convolutional auto-encoder (S-OCAE) to integrate shape priors into deep segmentation has been shown to be effective in extracting various abdominal structures, including smaller ones. It performs better than state-of-the-art techniques, including U-Net with traditional convolutional auto-encoders. Furthermore, using shape priors in deep

learning algorithms for abdominal image segmentation is a promising technique that can improve the accuracy and efficiency of the segmentation process. By leveraging the inherent structural information of the abdominal anatomy, these algorithms can more accurately identify and segment the various organs in medical images, providing valuable information for clinical diagnosis and treatment planning. As a perspective, incorporating both shape and topological priors into a segmentation network could deserve further investigation to take advantage of multiple prior knowledge embedding. By providing the network with information about the shape and topological properties of the targeted structures of interest, we expect the network to provide more plausible and realistic delineation predictions. Moreover, the proposed approach is generic enough to be easily extended to other structure types, anatomies, and imaging modalities, towards reliable decision support.

CRediT authorship contribution statement

Amine Sadikine: Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Bogdan Badic:** Conceptualization, Data curation, Project administration, Investigation. **Jean-Pierre Tasu:** Conceptualization, Investigation. **Vincent Noblet:** Conceptualization, Investigation. **Pascal Ballet:** Conceptualization, Project administration. **Dimitris Visvikis:** Conceptualization, Investigation, Methodology, Project administration, Validation. **Pierre-Henri Conze:** Conceptualization, Funding acquisition, Investigation, Methodology, Software, Supervision, Writing – original draft, Writing – review & editing.

Declaration of competing interest

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

Data availability

Only publicly-available abdominal imaging datasets are used.

Acknowledgments

This work was partially funded by LaBeX CAMI (grant ANR-11-LABX-0004) and France Life Imaging (grant ANR-11-INBS-0006).

References

Antonelli, M., Reinke, A., Bakas, S., Farahani, K., Kopp-Schneider, A., Landman, B.A., Litjens, G., Menze, B., Ronneberger, O., Summers, R.M., Van Ginneken, B., Bilello, M., Bilic, P., Christ, P.F., Do, R.K.G., Gollub, M.J., Heckers, S.H., Huisman, H., Jarnagin, W.R., McHugo, M.K., Napel, S., Pernicka, J.S.G., Rhode, K., Tobon-Gomez, C., Vorontsov, E., Meakin, J.A., Ourselin, S., Wiesenfarth, M., Arbeláez, P., Bae, B., Chen, S., Daza, L., Feng, J., He, B., Isensee, F., Ji, Y., Jia, F., Maier-Hein, K., Merhof, D., Pai, A., Park, B., Perslev, M., Rezaiifar, R., Rippel, O., Sarasua, I., Shen, W., Son, J., Wachinger, C., Wang, L., Wang, Y., Xia, Y., Xu, D., Xu, Z., Zheng, Y., Simpson, A.L., Maier-Hein, L., Cardoso, M.J., 2022. The medical segmentation decathlon. Nature Commun. 13 (1), 1–13.

Bilic, P., Christ, P., Li, H.B., Vorontsov, E., Ben-Cohen, A., Kaissis, G., Szeskin, A., Jacobs, C., Mamani, G.E.H., Chartrand, G., Lohöfer, F., Holch, J.W., Sommer, W., Hofmann, F., Hostettler, A., Lev-Cohain, N., Drozdzal, M., Amitai, M.M., Vivanti, R., Sosna, J., Ezhov, I., Sekuboyina, A., Navarro, F., Kofler, F., Paetzold, J.C., Shit, S., Hu, X., Lipková, J., Rempfler, M., Piraud, M., Kirschke, J., Wiestler, B., Zhang, Z., Hülsemeyer, C., Beetz, M., Ettlinger, F., Antonelli, M., Bae, W., Bellver, M., Bi, L., Chen, H., Chlebus, G., Dam, E.B., Dou, Q., Fu, C.-W., Georgescu, B., i Nieto, X.G., Gruen, F., Han, X., Heng, P.-A., Hesser, J., Moltz, J.H., Igel, C., Isensee, F., Jäger, P., Jia, F., Kaluva, K.C., Khened, M., Kim, I., Kim, J.-H., Kim, S., Kohl, S., Konopczynski, T., Kori, A., Krishnamurthi, G., Li, F., Li, H., Li, J., Li, X., Lowengrub, J., Ma, J., Maier-Hein, K., Maninis, K.-K., Meine, H., Merhof, D., Pai, A., Perslev, M., Petersen, J., Pont-Tuset, J., Qi, J., Qi, X., Rippel, O., Roth, K., Sarasua, I., Schenk, A., Shen, Z., Torres, J., Wachinger, C., Wang, C., Weninger, L., Wu, J., Xu, D., Yang, X., Yu, S.C.-H., Yuan, Y., Yue, M., Zhang, L., Cardoso, J., Bakas, S., Braren, R., Heinemann, V., Pal, C., Tang, A., Kadoury, S., Soler, L., Van Ginneken, B., Greenspan, H., Joskowicz, L., Menze, B., 2023. The liver tumor segmentation benchmark (liTS). Med. Image Anal. 84, 102680.

Boas, F.E., Fleischmann, D., 2012. CT artifacts: causes and reduction techniques. Imaging Med. 4 (2), 229–240.

Bohlender, S., Oksuz, I., Mukhopadhyay, A., 2021. A survey on shape-constraint deep learning for medical image segmentation. arXiv preprint arXiv:2101.07721.

Boutillon, A., Borotikar, B., Burdin, V., Conze, P.-H., 2020. Combining shape priors with conditional adversarial networks for improved scapula segmentation in MR images. In: IEEE International Symposium on Biomedical Imaging. pp. 1164–1167.

Boutillon, A., Borotikar, B., Burdin, V., Conze, P.-H., 2022. Multi-structure bone segmentation in pediatric MR images with combined regularization from shape priors and adversarial network. Artif. Intell. Med. 132, 102364.

Boutillon, A., Borotikar, B., Pons, C., Burdin, V., Conze, P.-H., 2021. Multi-structure deep segmentation with shape priors and latent adversarial regularization. In: IEEE International Symposium on Biomedical Imaging. pp. 999–1002.

Chen, H., Qi, X., Yu, L., Dou, Q., Qin, J., Heng, P.-A., 2017. DCAN: Deep contour-aware networks for object instance segmentation from histology images. Med. Image Anal. 36, 135–146.

Chen, X., Wang, X., Zhang, K., Fung, K.-M., Thai, T.C., Moore, K., Mannel, R.S., Liu, H., Zheng, B., Qiu, Y., 2022. Recent advances and clinical applications of deep learning in medical image analysis. Med. Image Anal. 102444.

Christ, P.F., Elshaer, M.E.A., Ettlinger, F., Tatavarty, S., Bickel, M., Bilic, P., Rempfler, M., Armbruster, M., Hofmann, F., D'Anastasi, M., Sommer, W.H., Ahmadi, S.-A., Menze, B.H., 2016. Automatic liver and lesion segmentation in CT using cascaded fully convolutional neural networks and 3D conditional random fields. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 415–423.

Conze, P.-H., Andrade-Miranda, G., Singh, V.K., Jaouen, V., Visvikis, D., 2023. Current and emerging trends in medical image segmentation with deep learning. IEEE Trans. Radiat. Plasma Med. Sci..

Conze, P.-H., Brochard, S., Burdin, V., Sheehan, F.T., Pons, C., 2020. Healthy versus pathological learning transferability in shoulder muscle MRI segmentation using deep convolutional encoder-decoders. Comput. Med. Imaging Graph. 83, 101733.

Conze, P.-H., Kavur, A.E., Cornec-Le Gall, E., Gezer, N.S., Le Meur, Y., Selver, M.A., Rousseau, F., 2021. Abdominal multi-organ segmentation with cascaded convolutional and adversarial deep networks. Artif. Intell. Med. 117, 102109.

Dalca, A.V., Guttag, J., Sabuncu, M.R., 2018. Anatomical priors in convolutional networks for unsupervised biomedical segmentation. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 9290–9299.

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: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 149–157.

Fan, J., Li, R., 2001. Variable selection via nonconcave penalized likelihood and its oracle properties. J. Amer. Statist. Assoc. 96 (456), 1348–1360.

Ganaye, P.-A., Sdika, M., Triggs, B., Benoit-Cattin, H., 2019. Removing segmentation inconsistencies with semi-supervised non-adjacency constraint. Med. Image Anal. 58, 101551.

Gao, M., Xu, Z., Lu, L., Wu, A., Nogues, I., Summers, R.M., Mollura, D.J., 2016. Segmentation label propagation using deep convolutional neural networks and dense conditional random field. In: IEEE International Symposium on Biomedical Imaging. pp. 1265–1268.

Hanson, S., Pratt, L., 1988. Comparing biases for minimal network construction with back-propagation. Adv. Neural Inf. Process. Syst..

He, K., Zhang, X., Ren, S., Sun, J., 2016. Identity mappings in deep residual networks. In: European Conference on Computer Vision.

Heller, N., Isensee, F., Maier-Hein, K.H., Hou, X., Xie, C., Li, F., Nan, Y., Mu, G., Lin, Z., Han, M., Yao, G., Gao, Y., Zhang, Y., Wang, Y., Hou, F., Yang, J., Xiong, G., Tian, J., Zhong, C., Ma, J., Rickman, J., Dean, J., Stai, B., Tejpaul, R., Oestreich, M., Blake, P., Kaluzniak, H., Raza, S., Rosenberg, J., Moore, K., Walczak, E., Rengel, Z., Edgerton, Z., Vasdev, R., Peterson, M., McSweeney, S., Peterson, S., Kalapara, A., Sathianathen, N., Papanikolopoulos, N., Weight, C., 2021. The state of the art in kidney and kidney tumor segmentation in contrast-enhanced CT imaging: Results of the KiTS19 challenge. Med. Image Anal. 67, 101821.

Hernández-García, A., König, P., 2018. Data augmentation instead of explicit regularization. arXiv preprint arXiv:1806.03852.

Kavur, A.E., Gezer, N.S., Baris, M., Aslan, S., Conze, P.-H., Groza, V., Pham, D.D., Chatterjee, S., Ernst, P., Ozkan, S., Baydar, B., Lachinov, D., Han, S., Pauli, J., Isensee, F., Perkonigg, M., Sathish, R., Rajan, R., Sheet, D., Dovletov, G., Speck, O., Nurnberger, A., Maier-Hein, K.H., Bozdagi Akar, G., Unal, G., Dicle, O., Selver, M.A., 2021. CHAOS challenge-combined (CT-MR) healthy abdominal organ segmentation. Med. Image Anal. 69, 101950.

Keshwani, D., Kitamura, Y., Ihara, S., Iizuka, S., Simo-Serra, E., 2020. TopNet: Topology preserving metric learning for vessel tree reconstruction and labelling. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 14–23.

Kingma, D.P., Ba, J., 2014. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

Kitrungrotsakul, T., Han, X.-H., Iwamoto, Y., Lin, L., Foruzan, A.H., Xiong, W., Chen, Y.-W., 2019. VesselNet: A deep convolutional neural network with multi pathways for robust hepatic vessel segmentation. Comput. Med. Imaging Graph. 75, 74–83.

Lafferty, J., McCallum, A., Pereira, F.C., 2001. Conditional random fields: Probabilistic models for segmenting and labeling sequence data.

Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., Van Der Laak, J.A., Van Ginneken, B., Sánchez, C.I., 2017. A survey on deep learning in medical image analysis. Med. Image Anal. 42, 60–88.

Nelms, B.E., Tomé, W.A., Robinson, G., Wheeler, J., 2012. Variations in the contouring of organs at risk: test case from a patient with oropharyngeal cancer. Int. J. Radiat. Oncol.*Biol.*Phys. 82 (1), 368–378.

Nosrati, M.S., Hamarneh, G., 2016. Incorporating prior knowledge in medical image segmentation: A survey. arXiv preprint arXiv:1607.01092.

Oda, M., Roth, H.R., Kitasaka, T., Misawa, K., Fujiwara, M., Mori, K., 2019. Abdominal artery segmentation method from CT volumes using fully convolutional neural network. Int. J. Comput. Assist. Radiol. Surg. 14 (12), 2069–2081.

Oktay, O., Ferrante, E., Kamnitsas, K., Heinrich, M., Bai, W., Caballero, J., Cook, S.A., De Marvao, A., Dawes, T., O'Regan, D.P., Kainz, B., Glocker, B., Rueckert, D., 2017. Anatomically constrained neural networks (ACNNs): application to cardiac image enhancement and segmentation. IEEE Trans. Med. Imaging 37 (2), 384–395.

Ravishankar, H., Venkataramani, R., Thiruvenkadam, S., Sudhakar, P., Vaidya, V., 2017. Learning and incorporating shape models for semantic segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 203–211.

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

Sadikine, A., Badic, B., Tasu, J.-P., Noblet, V., Visvikis, D., Conze, P.-H., 2022. Semi-overcomplete convolutional auto-encoder embedding as shape priors for deep vessel segmentation. In: IEEE International Conference on Image Processing. pp. 586–590.

Sajjadi, M., Javanmardi, M., Tasdizen, T., 2016. Regularization with stochastic transformations and perturbations for deep semi-supervised learning. Adv. Neural Inf. Process. Syst. 29.

Soler, L., Hostettler, A., Agnus, V., Charnoz, A., Fasquel, J., Moreau, J., Osswald, A., Bouhadjar, M., Marescaux, J., 2010. 3D Image Reconstruction for Comparison of Algorithm Database: A Patient Specific Anatomical and Medical Image Database. IRCAD Tech. Rep..

Staal, J., Abràmoff, M.D., Niemeijer, M., Viergever, M.A., Van Ginneken, B., 2004. Ridge-based vessel segmentation in color images of the retina. IEEE Trans. Med. Imaging 23 (4), 501–509.

Tekchandani, H., Verma, S., Londhe, N.D., 2022a. Improving the detection of abdominal and mediastinal lymph nodes in CT images using attention U-net based deep learning model. In: Artificial Intelligence Applications for Healthcare. pp. 181–202.

Tekchandani, H., Verma, S., Londhe, N.D., Jain, R.R., Tiwari, A., 2022b. Computer aided diagnosis system for cervical lymph nodes in CT images using deep learning. Biomed. Signal Process. Control 71, 103158.

Valanarasu, J.M.J., Patel, V.M., 2021. Overcomplete deep subspace clustering networks. In: IEEE Winter Conference on Applications of Computer Vision. pp. 746–755.

Valanarasu, J.M.J., Sindagi, V.A., Hacihaliloglu, I., Patel, V.M., 2020. Kiu-net towards accurate segmentation of biomedical images using over-complete representations. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 363–373.

Yang, Y., Hospedales, T.M., 2016. Trace norm regularised deep multi-task learning. arXiv preprint arXiv:1606.04038.

Yu, W., Fang, B., Liu, Y., Gao, M., Zheng, S., Wang, Y., 2019. Liver vessels segmentation based on 3D residual U-net. In: IEEE International Conference on Image Processing. pp. 250–254.