Hierarchical 3D Feature Learning for Pancreas Segmentation

Federica Proietto Salanitri[1(✉)], Giovanni Bellitto[1], Ismail Irmakci[2,3], Simone Palazzo[1], Ulas Bagci[3], and Concetto Spampinato[1]

[1] PeRCeiVe Lab, University of Catania, Catania, Italy
`federica.proiettosalanitri@phd.unict.it`
[2] CE, Ege University, Izmir, Turkey
[3] Department of Radiology and BME, Northwestern University, Chicago, IL, USA

Abstract. We propose a novel 3D fully convolutional deep network for automated pancreas segmentation from both MRI and CT scans. More specifically, the proposed model consists of a 3D encoder that learns to extract volume features at different scales; features taken at different points of the encoder hierarchy are then sent to multiple 3D decoders that individually predict intermediate segmentation maps. Finally, all segmentation maps are combined to obtain a unique detailed segmentation mask. We test our model on both CT and MRI imaging data: the publicly available NIH Pancreas-CT dataset (consisting of 82 contrast-enhanced CTs) and a private MRI dataset (consisting of 40 MRI scans). Experimental results show that our model outperforms existing methods on CT pancreas segmentation, obtaining an average Dice score of about 88%, and yields promising segmentation performance on a very challenging MRI data set (average Dice score is about 77%). Additional control experiments demonstrate that the achieved performance is due to the combination of our 3D fully-convolutional deep network and the hierarchical representation decoding, thus substantiating our architectural design.

Keywords: CT and MRI pancreas segmentation · Fully convolutional neural networks · Hierarchical encoder-decoder architecture

1 Introduction

Pancreatic cancer is a growing public health concern worldwide. In 2021, an estimated 60,430 new cases of pancreatic cancer will be diagnosed in the US and 48,220 people will die from this disease [19]. Early detection of pancreas cancer [14] is very hard and options in treatment are very limited. Radiology imaging and automated image analysis play key roles in diagnosis, prognosis, treatment, and intervention of pancreatic diseases; thus, there is a strong, unmet, need for computer aided analysis tools supporting these tasks. The first step in such analysis is to automate the medical image segmentation procedures, since manual segmentation (current standard) is tedious, prone to error, and it is not practical in routine clinical evaluation of the diseases [20]. Beyond the known challenges of medical image segmentation problems, pancreas is one of the most difficult organs to segment despite the recent advances in deep segmentation models.

Computed tomography (CT) and magnetic resonance imaging (MRI) are the two most common modalities for pancreas imaging. CT is the modality of choice for pancreatic cancer at the moment, while MRI is mostly used for finding other pancreatic diseases including cysts and diabetes. Compared to CT, MRI has advantages such as the lack of ionizing radiation, better resolution and soft tissue contrast. However, MRI has other unique difficulties, including field inhomogeneity, non-standard intensity distributions due to variations in scanners, patients, field strengths, and high similarity in pancreas and non-pancreas tissue densities.

Image-based pancreas analysis is by itself a challenging task. Shapes and sizes greatly vary across different patients, making it difficult to use robust priors for improving the delineation procedures. Intensity similarities to non-pancreatic tissues, and smooth or invisible boundaries (due to resolution limitations of medical scanners) are other challenges that need to be addressed in a successful segmentation method. Moreover, in presence of a cyst, tumor, or other abnormalities in pancreases, segmentation algorithms may easily fail to delineate correct boundaries.

To address these challenges, in this work we propose a novel 3D fully convolutional encoder-decoder network with hierarchical multi-scale feature learning, for general, fully-automated pancreas segmentation applicable to CT and MRI scans. Major contributions of this study are the following:

– Our segmentation network is unique in the sense that it is volumetric, learns to extract 3D volume features at different scales, and decodes features hierarchically, leading to improved segmentation results;
– We show the efficacy of our work both on CT and MRI scans. Our architecture successfully extracts pancreases from CT and MRI with high accuracy, obtaining new state-of-the-art results on a publicly-available CT benchmark and first-ever volumetric pancreas segmentation from MRI in the literature.
– Our work on MRI pancreas segmentation is an important application contribution, due to the very limited published research on this task using MRI data with deep learning. It is our belief that our method provides a significant state-of-the-art baseline to be compared with for further MRI pancreas research.

2   Related Work

Following the success of deep learning methods applied in medical image segmentation, researchers have recently shown an increasing interest in pancreas segmentation, in order to support physicians in early stage diagnosis for pancreas cancer. Although this application field is still in its infancy—also due to variabilities in texture, size and imaging contrast—a line of promising approaches has been proposed in the literature, mainly on CT scans [2,8,10–12,15–17,21,22,24,25]. We here describe the most significant ones which relate to our proposed model.

In [16], a two-stage cascaded approach for pancreas localization and pancreas segmentation is proposed. In the first stage, the method localizes the pancreas in the entire 3D CT scan, providing a reliable bounding box for a more refined segmentation step, based on an efficient application of holistically-nested convolutional networks (HNNs) on the three views of pancreas CT image. Per-pixel probability maps are then fused to

produce a 3D bounding box of the pancreas. Projective adversarial networks [8] incorporate high-level 3D information through 2D projections and introduce an attention module that supports a selective integration of global information from the segmentation module to an adversarial network. More recently, [22] proposes a dual-input v-mesh fully-convolutional network, which receives original CT scans and images processed by contrast-specific graph-based visual saliency, in order to enhance the soft tissue contrast and highlight differences among local regions in abdominal CT scans.

All of the above works tackle the problem of pancreas segmentation on CT scans. However, as already mentioned, MRI acquisitions have several advantages over CT—most importantly, fewer risks to the patients. On the other hand, MRI pancreas segmentation presents additional challenges to automated visual analysis. For this reason and others (e.g., the lack of public benchmarks), very few works have addressed pancreas segmentation on MRI data: to the best of our knowledge, the major attempts are [1–3]. In [3], two CNN models are combined to perform, respectively, tissue detection and boundary detection; the results are provided as input to a conditional random field (CRF) for final segmentation. In [1], an algorithmic approach based on hand-crafted features is proposed, employing an ad-hoc multi-stage pipeline: contrast enhancement within coarsely detected pancreas regions is applied to differentiate between pancreatic and surrounding tissue; 3D segmentation and edge detection through max-flow and min-cuts approach and structured forest are performed; finally, non-pancreatic contours are removed via morphological operations on area, structure and connectivity.

3   Method

Our 3D fully-convolutional pancreas segmentation model—*PankNet*—is based on an encoder-decoder architecture; however, unlike standard encoder-decoder schemes with a single decoding path (see Fig. 1a), we have parallel decoders at different abstraction levels, generating multiple intermediate segmentation maps (Fig. 1c). Hierarchical decoding is also fundamentally different from using skip connections (Fig. 1b), since these have the purpose to ease gradient flow and forward low-level features for output reconstruction, while our multiple decoders aim to extract local and global dependencies. The detailed architecture is shown in Fig. 2: the input data (either CT or MRI volume) is first processed by the encoder stream of the model which aggregates volumetric features at different abstraction levels. These features are then given as input to

(a) encoder-decoder   (b) skip-connection   (c) hierarchical decoding
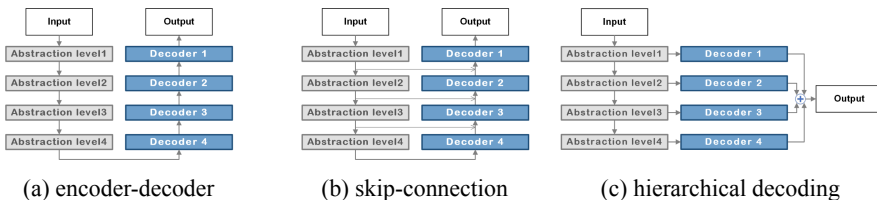
Fig. 1. A comparison between our proposed architecture and other types of networks used for segmentation: (a) standard encoder–decoder architecture; (b) encoder–decoder architecture with skip connections; (c) encoder–hierarchical decoder architecture (ours).
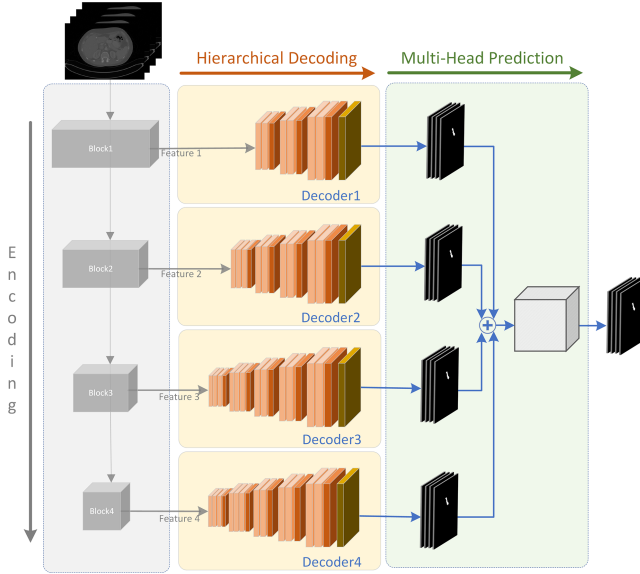
Fig. 2. *PanKNet* architecture: the encoding path extracts aggregated volumetric features, while the decoding path predicts four different intermediate segmentation masks (coarse to fine). Finally, intermediate segmentations are integrated into a detailed output mask. (Color figure online)

different decoder streams, each generating a segmentation mask volume. All intermediate masks are concatenated along the channel dimension and finally merged through a convolutional layer in order to predict the final segmentation mask for all input slices.

3.1   Volume Feature Encoding

The model's encoder performs aggregation of volumetric features from the input data. It is based on S3D [23], a network originally proposed for action recognition using 3D spatial and temporal separable 3D convolution layers, pretrained on the Kinetics Dataset [6]. We use the pretrained network, similarly to other works [3,8,9], to ease convergence given the limited training data we have from both CT and MRI datasets. Our encoder processes $D = 48$ slices from an input scan by progressively aggregating volumetric cues down to a more compact representation of size $1024 \times \frac{W}{8} \times \frac{H}{32} \times \frac{D}{32}$ (channels × width × height × depth). Features at the bottleneck and at the outputs of the second, third and fourth pooling layers are fed to separate decoders, described in the following section, to implement our hierarchical decoding strategy.

The proposed approach can be easily adapted to different encoder architectures. Thus, we additionally design a lightweight variant of our *PanKNet* network by replacing the S3D-based encoder with an encoder based on MobileNetV2 [18], where 2D convolutions are replaced with 3D ones through inflation. In particular, the 2D kernels are replicated along the third dimension, and the values of the weights are divided by the number of replications as proposed in [4]. In this case, as input to the decoders, we select

the output of the second, third, fourth and sixth *bottleneck* blocks of MobileNetV2, providing a more compact feature map of size $160 \times \frac{W}{16} \times \frac{H}{32} \times \frac{D}{32}$. This lightweight variant has 10 times fewer parameters (2.5 millions of parameters, 9.33 MB) and than the S3D counterpart (25.6 millions of parameters, 97.88 MB).

3.2   Hierarchical Decoding

Our hierarchical decoding strategy employs features at different points of the encoder stream to generate intermediate segmentation masks that aim to capture and combine fine segmentation (derived from decoders of deeper features) to coarse segmentation (derived from decoders of initial features). We include four decoders: each one processes a set of volumetric features taken from the corresponding level in the encoder stack and performs segmentation on the input volume (see Fig. 2, yellow blocks). Each decoder consists of a cascade of upsampling blocks, depending on the size of the input feature map: decoders operating on deeper features require less blocks to recover the original input size. Each upsampling block contains a 3D convolutional block (convolutional layer + batch normalization + ReLU), one or two 3D separable convolutional blocks, and a trilinear upsample layer. As last layer, a pointwise 3D convolution outputs a volume with size $2 \times W \times H \times D$, where W, H and D are the same as the input volume.

3.3   Pancreas Segmentation

Intermediate segmentation maps predicted by each of the model's decoders are combined into a global mask. In particular, the four intermediate maps are concatenated into a $8 \times W \times H \times D$ tensor, which then goes through a last layer performing a voxel-wise convolution to generate a single segmentation map of size $2 \times W \times H \times D$.

The whole model (encoder, hierarchical decoders and output layer) is trained end-to-end using a hierarchical Dice loss [13] between ground-truth mask, intermediate generated masks and the output segmentation mask. Formally, given the predicted output segmentation masks $\mathbf{S}_v$ for the input volume, the four maps $\hat{\mathbf{S}}_{v^i}$ estimated by the decoders, and the ground-truth segmentation maps $\mathbf{G}_v$ for the input data, the *segmentation loss* $\mathcal{L}_s$ is:

$$\mathcal{L}_s \left(\mathbf{S}_v, \hat{\mathbf{S}}_{v^i}, \mathbf{G}_v \right) = \sum_{i=1}^{4} \frac{2 \sum_j \hat{S}_{v^{i,j}} G_{vj}}{\sum_j \hat{S}_{v^{i,j}}^2 + \sum_j G_{vj}^2} + \frac{2 \sum_j S_{vj} G_{vj}}{\sum_j S_{vj}^2 + \sum_j G_{vj}^2} \qquad (1)$$

where index i iterates over the four intermediate maps and index j iterates over voxels.

4   Experiments

4.1   Dataset

We evaluate the accuracy of our proposed deep segmentation method in both CT and MRI modalities. For the former, we use the publicly available NIH Pancreas-CT dataset, which is the most used pancreas segmentation dataset for benchmarking [15]. This dataset includes 82 abdominal contrast-enhanced 3D CT scans. The resolution of the

CT scans is $512 \times 512 \times Z$, with Z (between 181 and 466) indicating the number of slices along the transverse axis. Voxel spacing ranges from 0.5 mm to 1 mm. More details on this dataset are available in [15].

In our experiments with MRI data, we use 40 in-house collected T2-weighted MRI scans from 40 patients, who have either IPMN (intraductal papillary mucinous neoplasm) cysts detected in their pancreases or invasive pancreatic ductal carcinoma. Two expert radiologists annotated pancreases manually and consensus segmentation masks were generated at the end of the ground-truth labeling procedure with agreement. MRI images were resized (in the transverse plane) to 256×256 pixels, with voxel spacing of varying from 0.468 mm to 1.406 mm. To minimize uncertainties in MRI scans, we applied a set of pre-processing steps: N4 bias field correction followed by an edge-preserving Gaussian smoothing, and intensity standardization procedure to standardize MRI scans across patients, scanners, and time.

4.2   Training and Evaluation Procedure

We apply the same training procedure for the two datasets, with the only difference regarding how model backbones are pre-trained. On the NIH Pancreas-CT dataset, we pre-train S3D on Kinetics [6] and MobileNetV2 on ImageNet [5] with weight inflation; on our MRI data, *Pancreas-MRI*, we employ the backbones pre-trained on the CT task.

Input CT and MRI scans are re-oriented using the RAS axes convention for consistency. We then perform voxel resampling through trilinear interpolation in order to have isotropic (1 mm) voxel spacing, and normalize the values of each scan between 0 and 1. During training, data augmentation is performed with random horizontal flipping, random 90° rotation and random crops of size $128 \times 128 \times 48$ (in RAS coordinates). We minimize our multi-part Dice loss with mini-batch gradient descent using the Adam optimizer (learning rate: 0.001) and batch size 8, for a total of 3000 epochs.

At inference time, we compute output segmentation masks by running a sliding window routine over an entire input scan, using $256 \times 256 \times 48$ windows overlapping by 25%. Voxel labels from overlapping segmentations are obtained by averaging the set of predictions. For evaluation, we carry out 4-fold cross-validation. At each iteration, the set of training folds is further split into the actual training set and a validation set, that is used to select the epoch at which Dice score on the test fold is reported. As metrics for quantitative evaluation, we employ: *Dice score coefficient* (DSC), *Positive Predictive Value* (PPV) and *Sensitivity*.

Experiments are performed on an NVIDIA Quadro P6000 GPU. The proposed approach was implemented in PyTorch and MONAI; all code will be publicly released.

4.3   Results

We first test our model (as well as its lightweight variant) on the NIH Pancreas-CT dataset and compare it to existing methods (which share our evaluation strategy with 4-fold cross-validation), namely, [2, 8, 10–12, 15–17, 21, 22, 24, 25]. Summarized in Table 1, our results indicate that *PanKNet* outperforms existing methods over different metrics. Note that PanKNet does not require any auxiliary regularization networks [8], nor additional inputs [22], nor upstream pancreas localization module [12]. Remarkably,

Table 1. Comparison of *PanKNet* against multiple state-of-the-art models for pancreas segmentation on NIH Pancreas-CT dataset using 4-fold cross-validation. **Best performance in bold**, *second best in italic*.

| Method	DSC			PPV	SENS
	Avg	Max	Min		
Roth et al. [15]	71.42 ± 10.11	86.29	23.99	–	–
Roth et al. [16]	78.01 ± 8.20	88.65	34.11	–	–
Roth et al. [17]	81.27 ± 6.27	88.96	50.69	–	–
Zhou et al. [25]	82.37 ± 5.68	90.85	62.43	–	–
Cai et al. [2]	82.40 ± 6.70	90.10	60.00	–	–
Li et al. (2019) [10]	83.50 ± 6.20	–	–	84.50 ± 6.90	83.70 ± 10.40
Liu et al. (2020) [11]	84.10 ± 4.90	–	–	83.60 ± 5.90	85.30 ± 08.20
You et al. [24]	84.50 ± 4.97	91.02	62.81	–	–
Khosravan et al. [8]	85.53 ± 1.23	88.71	83.20	–	–
Wang et al. (2020) [21]	85.90 ± 3.40	–	–	–	–
Man et al. (2019) [12]	86.90 ± 4.90	–	–	–	–
Wang et al. [22]	87.04 ± 6.80	–	–	$\mathbf{89.50 \pm 5.80}$	87.70 ± 7.90
PanKNet$_{Light}$	*87.13 ± 4.58*	93.49	72.77	86.85 ± 6.52	*88.48 ± 5.12*
PanKNet	$\mathbf{88.01 \pm 4.74}$	93.84	70.62	*88.25 ± 5.45*	$\mathbf{88.69 \pm 5.99}$

even the lightweight variant of PanKNet yields accuracy comparable to the full model, while outperforming existing models, showing that the choice of the backbone is not as important as the overall employed hierarchical architecture. The best trade-off between accuracy and computational resources for CT pancreas segmentation is represented by PanKNet$_{Light}$, whose memory occupation is about 10 MB compared to about 100 MB of PanKNet, but with very similar performance.

We then test our model on pancreas segmentation from MRI data. In this case, we compare the 3D-UNet, proposed in [7], pre-trained on the NIH Pancreas-CT dataset and fine-tuned on our MRI dataset. Furthermore, we add to this evaluation some control experiments to show the effectiveness of the designed architecture. Consequently, we define as baseline our encoder-decoder architecture without hierarchical decoding strategy, decoding only the features at the model's bottleneck. Results in Table 2 indicate that both PanKNet variants outperform the state-of-the-art 3D U-Net model [7]. The baseline (with either backbones) also performs better than 3D U-Net model [7] demonstrating that even our 3D fully convolutional network, ablated from the hierarchical decoding, is effective for MRI pancreas segmentation. Adding hierarchical decoding leads to enhanced segmentation performance, especially on DSC and PPV. Different from CT segmentation and from baseline models, PanKNet largely outperforms its lightweight counterpart, demonstrating that MRI pancreas segmentation is far more complex and challenging than CT segmentation and calls for high-capacity networks to be solved.

Example segmentation masks, corresponding to the highest and lowest Dice scores reported in Tables 1 and 2 for CT and MRI pancreas segmentation, are illustrated in Fig. 3.

Table 2. Segmentation performance on Pancreas-MRI dataset (4-fold CV).

Method	DSC			PPV	SENS
	Avg	Max	Min		
3D-UNet [7]	65.05 ± 9.17	84.58	49.80	61.55 ± 7.55	74.42 ± 13.99
Baseline$_{Light}$	69.17 ± 8.10	83.86	49.92	64.64 ± 7.49	84.19 ± 11.72
Baseline	65.16 ± 9.11	84.00	49.49	61.92 ± 8.22	75.22 ± 12.46
PanKNet$_{Light}$	72.96 ± 10.33	88.54	49.90	71.39 ± 11.21	79.76 ± 11.53
PanKNet	**77.46 ± 08.62**	**89.07**	**52.30**	**76.63 ± 08.66**	**80.91 ± 10.51**
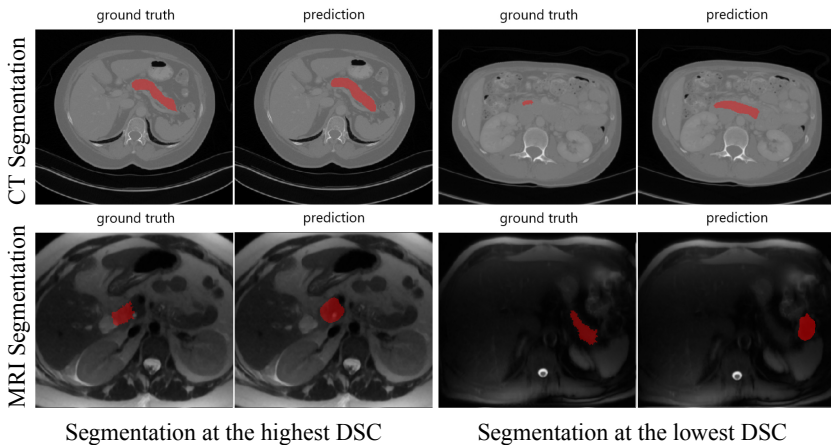
Fig. 3. Segmentation masks at the highest (left column) and lowest Dice score (right column) on NIH Pancreas-CT (first row) and Pancreas-MRI dataset (second row).

5 Conclusion

In this study, we propose a novel 3D fully-convolutional network for pancreas segmentation from MRI and CT scans. Our proposed deep network aims at learning and combining multi-scale features, namely a hierarchical decoding strategy, to generate intermediate segmentation masks for a coarse-to-fine segmentation process. The intermediate masks, capturing fine details, are derived from decoders of deeper features while coarse segmentation details are derived from decoders of initial features. We evaluated the efficacy of our method (a) on CT scans from the publicly available NIH CT-Pancreas benchmark, and obtained a new state of the art Dice score **88.01%**, outperforming all previous methods; and (b) on MRI scans, obtaining a Dice score of **77.46%**, which can be used as a baseline for future works on MRI pancreas segmentation. Noting that MRI pancreas segmentation methods are extremely limited due to the challenging nature of the problem, our study offers a fresh insight into MRI analysis of pancreas from a fully automated volumetric segmentation strategy. PanKNet is tested for pancreas segmentation, but its architecture is general and can be applied to any 3D object segmentation problem in medical domain.

References

1. Asaturyan, H., Gligorievski, A., Villarini, B.: Morphological and multi-level geometrical descriptor analysis in CT and MRI volumes for automatic pancreas segmentation. Comput. Med. Imaging Graph. **75**, 1–13 (2019)
2. Cai, J., Lu, L., Xie, Y., Xing, F., Yang, L.: Improving deep pancreas segmentation in CT and MRI images via recurrent neural contextual learning and direct loss function. arXiv preprint arXiv:1707.04912 (2017)
3. Cai, J., Lu, L., Xie, Y., Xing, F., Yang, L.: Pancreas segmentation in MRI using graph-based decision fusion on convolutional neural networks. In: Descoteaux, M., Maier-Hein, L., Franz, A., Jannin, P., Collins, D.L., Duchesne, S. (eds.) MICCAI 2017. LNCS, vol. 10435, pp. 674–682. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-66179-7_77
4. Carreira, J., Zisserman, A.: Quo vadis, action recognition? A new model and the kinetics dataset. In: CVPR, pp. 6299–6308 (2017)
5. Deng, J., Dong, W., Socher, R., Li, L., Li, K., Li, F.: ImageNet: a large-scale hierarchical image database. In: Computer Society Conference on Computer Vision and Pattern Recognition, pp. 248–255 (2009). https://doi.org/10.1109/CVPR.2009.5206848
6. Kay, W., et al.: The kinetics human action video dataset. arXiv preprint arXiv:1705.06950 (2017)
7. Kerfoot, E., Clough, J., Oksuz, I., Lee, J., King, A.P., Schnabel, J.A.: Left-ventricle quantification using residual U-Net. In: Pop, M., et al. (eds.) STACOM 2018. LNCS, vol. 11395, pp. 371–380. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-12029-0_40
8. Khosravan, N., Mortazi, A., Wallace, M., Bagci, U.: PAN: projective adversarial network for medical image segmentation. In: Shen, D., et al. (eds.) MICCAI 2019. LNCS, vol. 11769, pp. 68–76. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32226-7_8
9. LaLonde, R., et al.: INN: inflated neural networks for IPMN diagnosis. In: Shen, D., et al. (eds.) MICCAI 2019. LNCS, vol. 11768, pp. 101–109. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32254-0_12
10. Li, H., Lü, Q., Chen, G., Huang, T., Dong, Z.: Convergence of distributed accelerated algorithm over unbalanced directed networks. IEEE Trans. Syst. Man Cybern. Syst., 1–12 (2019). https://doi.org/10.1109/TSMC.2019.2946287
11. Liu, S., et al.: Automatic pancreas segmentation via coarse location and ensemble learning. IEEE Access **8**, 2906–2914 (2020). https://doi.org/10.1109/ACCESS.2019.2961125
12. Man, Y., Huang, Y., Feng, J., Li, X., Wu, F.: Deep Q learning driven CT pancreas segmentation with geometry-aware U-Net. IEEE Trans. Med. Imaging **38**(8), 1971–1980 (2019). https://doi.org/10.1109/TMI.2019.2911588
13. Milletari, F., Navab, N., Ahmadi, S.: V-Net: fully convolutional neural networks for volumetric medical image segmentation. In: 2016 Fourth International Conference on 3D Vision (3DV), pp. 565–571 (2016). https://doi.org/10.1109/3DV.2016.79
14. Oberstein, P.E., Olive, K.P.: Pancreatic cancer: why is it so hard to treat? Ther. Adv. Gastroenterol. **6**(4), 321–337 (2013)
15. Roth, H.R., et al.: DeepOrgan: multi-level deep convolutional networks for automated pancreas segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9349, pp. 556–564. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24553-9_68
16. Roth, H.R., Lu, L., Farag, A., Sohn, A., Summers, R.M.: Spatial aggregation of holistically-nested networks for automated pancreas segmentation. In: Ourselin, S., Joskowicz, L., Sabuncu, M.R., Unal, G., Wells, W. (eds.) MICCAI 2016. LNCS, vol. 9901, pp. 451–459. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46723-8_52

17. Roth, H.R., et al.: Spatial aggregation of holistically-nested convolutional neural networks for automated pancreas localization and segmentation. Med. Image Anal. **45**, 94–107 (2018)
18. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., Chen, L.C.: MobileNetV 2: inverted residuals and linear bottlenecks. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 4510–4520 (2018)
19. American Cancer Society: Cancer Facts & Figures. American Cancer Society (2021)
20. European Society of Radiology (ESR) communications@myesr.org Emanuele Neri Nandita de Souza Adrian Brady Angel Alberich Bayarri Christoph D. Becker Francesca Coppola Jacob Visser, E.S.: What the radiologist should know about artificial intelligence-an esr white paper. Insights into imaging **10**, 1–8 (2019)
21. Wang, W., et al.: A fully 3D cascaded framework for pancreas segmentation. In: 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI), pp. 207–211 (2020). https://doi.org/10.1109/ISBI45749.2020.9098473
22. Wang, Y., et al.: Pancreas segmentation using a dual-input V-Mesh network. Med. Image Anal. **69**, 101958 (2021)
23. Xie, S., Sun, C., Huang, J., Tu, Z., Murphy, K.: Rethinking spatiotemporal feature learning: speed-accuracy trade-offs in video classification. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds.) ECCV 2018. LNCS, vol. 11219, pp. 318–335. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-01267-0_19
24. Yu, Q., Xie, L., Wang, Y., Zhou, Y., Fishman, E.K., Yuille, A.L.: Recurrent saliency transformation network: incorporating multi-stage visual cues for small organ segmentation. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 8280–8289 (2018). https://doi.org/10.1109/CVPR.2018.00864
25. Zhou, Y., Xie, L., Shen, W., Wang, Y., Fishman, E.K., Yuille, A.L.: A fixed-point model for pancreas segmentation in abdominal CT scans. In: Descoteaux, M., Maier-Hein, L., Franz, A., Jannin, P., Collins, D.L., Duchesne, S. (eds.) MICCAI 2017. LNCS, vol. 10433, pp. 693–701. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-66182-7_79