Effective Pancreatic Cancer Screening on Non-contrast CT Scans via Anatomy-Aware Transformers

Yingda Xia[1,*], Jiawen Yao[2], Le Lu[2], Lingyun Huang[3], Guotong Xie[3], Jing Xiao[3], Alan Yuille[1], Kai Cao[4(✉)], Ling Zhang[2]

[1]Johns Hopkins University,[2]PAII Inc.,
[3]PingAn Technology,[4]Changhai Hospital

Abstract. Pancreatic cancer is a relatively uncommon but most deadly cancer. Screening the general asymptomatic population is not recommended due to the risk that a significant number of false positive individuals may undergo unnecessary imaging tests (e.g., multi-phase contrast-enhanced CT scans) and follow-ups, adding health care costs greatly and no clear patient benefits. In this work, we investigate the feasibility of using a single-phase non-contrast CT scan, a cheaper, simpler, and safer substituent, to detect resectable pancreatic mass and classify the detection as pancreatic ductal adenocarcinoma (PDAC) or other abnormalities (nonPDAC) or normal pancreas. This task is usually poorly performed by general radiologists or even pancreatic specialists. With pathology-confirmed mass types and knowledge transfer from contrast-enhanced CT to non-contrast CT scans as supervision, we propose a novel deep classification model with an anatomy-guided transformer. After training on a large-scale dataset including 1321 patients: 450 PDACs, 394 nonPDACs, and 477 normal, our model achieves a sensitivity of 95.2% and a specificity of 95.8% for the detection of abnormalities on the hold-out testing set with 306 patients. The mean sensitivity and specificity of 11 radiologists are 79.7% and 87.6%. For the 3-class classification task, our model outperforms the mean radiologists by absolute margins of 25%, 22%, and 8% for PDAC, nonPDAC, and normal, respectively. Our work sheds light on a potential new tool for large-scale (opportunistic or designed) pancreatic cancer screening, with significantly improved accuracy, lower test risk, and cost savings.

Keywords: Pancreatic Cancer · Large-scale Cancer Screening · Transformers · Non-contrast CT

1 Introduction

Pancreatic cancer is the third leading cause of death among all cancers in the United States, with a 5-year overall survival rate of ∼10% [14]. Surgical resection

* Work done during an internship at PAII Inc.
✉ Corresponding author (mdkaicao163@163.com).

by now remains the only treatment that offers curative potential [10], but more than 80% of patients with pancreatic cancer have already lost the opportunity of surgery at the first diagnosis. Thus, screening pancreatic cancer is very important to provide early diagnosis and patient risk monitoring. The most widely used imaging modality for the initial evaluation of suspected pancreatic cancer is the contrast-enhanced CT scan (CECT). The benefit of CECT for early pancreatic cancer detection includes high sensitivity and specificity, general standardization and availability and relatively easy interpretation [16]. However, CECT exposes patients to radiation and requires iodine contrast, which can cause reaction and potential risks in patients [16], making it hard to be recognized as a general protocol to screen for pancreatic cancer.

In this work, we investigate the possibility of using non-contrast CT scans (NCCT) to screen for pancreatic cancer with deep learning. Compared to CECT, NCCT is cheaper and safer, because it does not require iodine contrast and exposes patients to less radiation. NCCT has been generally applied in screening for lung nodules [11] which can possibly be reused for opportunistic pancreatic cancer screening as well. Nevertheless, due to the low contrast in NCCT pancreatic region, the difficulty of tumor detection rises significantly for radiologists without contrast enhancement. Deep networks, on the other hand, are particularly good at discovering local texture and shape geometry changes, which give us an opportunity to detect pancreatic cancer even without contrast enhancement on NCCT. Those miss detections by human eyes due to low visual contrast do not necessarily become the false negatives by deep learning (DL) detectors.

One major challenge of training deep learning models on NCCT is the difficulty of obtaining expert annotations. Even experienced radiologists could miss masses due to the low contrast on NCCT. This problem is tackled in the process of data collection from the following two aspects. (1) We obtain the pathology-confirmed mass type as classification ground-truth for patients with pancreatic ductal adenocarcinoma (PDAC) or non-PDAC. (2) For the pixel-level labeling of pancreatic tumor mass, the radiologist first annotates on the contrast-enhanced CT; we then transfer the segmentation mask from contrast-enhanced CT to non-contrast CT by performing volumetric registration on the same patient. The combined classification ground-truth labels and segmentation masks serve as the supervisions of our deep learning model with the input of non-contrast CT scans only. The pathology-confirmed mass type and knowledge transfer from CECT to NCCT are the two important pre-requisites for our model to surpass the human expert performance on detecting pancreatic cancer via NCCT.

In terms of the design of deep models, we extend the previous "Segmentation for Classification" [29] paradigm by building a deep classification on top of a segmentation model with transformers [18]. Given the fact that local texture could be insufficient to detect tumors in NCCT, we adopt Transformers to model the pancreas anatomy structure for better classification, which can capture the global context with multi-head attention. This is also in line with the practical diagnosis experience of the radiologists, where sometimes abnormality is discov-

ered by the secondary-sign, such as swelling pancreas head/tail or pancreatic duct dilation, without actually seeing the tumor.

To validate the feasibility of the proposed solution, we collect a large-scale dataset, which covers 1627 patients: 558 PDACs, 474 nonPDACs, and 595 normal. Our model achieves a sensitivity of 95.2% and a specificity of 95.8% on the holdout test set, in terms of abnormality detection. In contrast, the average performance of 11 expert radiologists is 79.7% and 87.6%. This result illustrates the superiority of our designed deep learning-based framework in this specialized task of detecting pancreatic cancer in NCCT. This work sheds light on a potential viable and safe protocol to screen pancreatic cancers on general population.

The main contributions of this paper are summarized as follows.

- For the first time, non-contrast CT (NCCT) is proposed and validated as an effective imaging modality for full-spectrum taxonomy of pancreatic mass/disease screening using deep learning. This sheds light on new computing tools for large-scale opportunistic or designated pancreatic cancer screening of improved accuracy, lower test risk, and cost savings.
- We utilize the pathology-confirmed mass labels, and transfer the imaging information and knowledge from CECT to NCCT as supervision, which is a prerequisite to surpass human expert performance in this task.
- We propose a new framework, named Anatomy-aware Hybrid Transformers, outperforming the mainstream "Segmentation for Classification" paradigm.
- We achieve a sensitivity of 95.2% and a specificity of 95.8% on a large-scale dataset with 1627 patients, demonstrating the good potential of using more convenient non-contrast CT scans for pancreatic cancer screening.

1.1 Related Work

Automated pancreatic tumor detection. Recent advances in deep learning have lead to tremendous improvement in pancreas segmentation [1, 9, 12, 13, 19, 23, 27, 28], an important pre-requisite step for pancreatic tumor detection. Researchers have started to explore the task of automated pancreatic tumor detection using contrast-enhanced CT scans with deep networks [4,20,24,25,29] and radiomics [3]; as well as the task of cancer prognosis prediction [22]. Different from previous work, our framework is designed for non-contrast CT scans, which is beneficial for general asymptomatic patients yet much more challenging.

Vision transformers. Transformer [18] utilizes attention mechanism originally designed for language tasks. It has recently been applied into vision task, *e.g.*, object detection [2], image recognition [18] and semantic segmentation [26], and achieved comparable or better performance than CNN based approaches.

2 Methodology

Problem statement. We formulate the task of pancreatic cancer detection in non-contrast CT scans as a three-class classification problem. We denote

$\mathcal{L} = \{0, 1, 2\}$ for the three patient classes, *i.e.*, normal, PDAC and non-PDAC. The reasons of having these three classes are: (1) PDAC is a unique group with the most dismal prognosis; (2) any pancreas CT findings with a influence on patient management options. Screening for pancreatic cancer is much more difficult than lung nodules or mammography screening due to the challenge and visual ambiguity of soft-tissue tumor detection without CT contrast enhancement. A key part in our processing pipeline is the availability of knowledge transfer from contrast-enhanced CT by incorporating pathology-confirmed mass type as classification labels and segmentation labels (tumor/pancreas) used for intermediate supervision, as shown in Fig.1. Denote the training set by $S = \{(\mathbf{X_i}, \mathbf{Y_i}, \mathbf{Z_i})|i = 1, 2, ..M\}$, where $\mathbf{X_i} \in \mathbb{R}^{H_i \times W_i \times D_i}$, is the 3D volume representing the non-contrast CT scans of the i-th patient. $\mathbf{Y_i}$ is the voxel-wise annotated label map with the same spatial size as $\mathbf{X_i}$. $\mathbf{Z_i} \in \mathcal{L}$ is the class label of the image, confirmed by pathology, radiology, or clinical records. In the testing phase, only $\mathbf{X_i}$ is given, and our goal is to predict a class label for $\mathbf{X_i}$.

Knowledge transfer from contrast-enhanced to non-contrast CT. Considering the difficulties of mass annotation on non-contrast CT scans (e.g., tumors are barely visible), radiologists first annotate the voxel-wise mass mask on the contrast-enhanced CT scan with the same patient. We then perform image registration using DEEDS [8] from CECT to NCCT and apply the registration field on the manually segmented mass mask. In this way, we can obtain a relatively coarse, but the most reliable mass mask $\mathbf{Y_i}$ on the NCCT image.

2.1 Anatomy-aware Classification with Transformers

Segmentation for classification is the most straightforward and adopted representation of the task of pancreatic tumor detection. We train a localization UNet [5] to segment pancreas and mass supervised by the transferred masks generated as above. This localization UNet is also used for cropping out the pancreas ROI region as shown in the test process in Fig 1.

Given the superiority of the attention mechanism in modelling the global context, we build a hybrid Vision Transformer [6] on top of the UNet segmentation model (see Fig 1). Since the transformer takes the input of the feature map of a segmentation network, we term it as Anatomy-ware Hybrid Transformer. We denote the pancreas ROI region by $\mathbf{X}$, and $\mathbf{X} \in \mathbb{R}^{H \times W \times D}$. We then forward the image $\mathbf{X}$ into a UNet, which consists of a feature extractor $\mathcal{F}$ and an output layer $\mathcal{G}$. This UNet has an intermediate supervision of the mask transferred from contrast-enhanced CT scan of the same patient where the human annotation is available. Therefore, the intermediate output segmentation $\mathbf{P}_s$ can be obtained by $\mathbf{P}_s = \mathcal{G}(\mathcal{F}(\mathbf{X}))$.

The input of the Vision Transformer $\mathcal{H}$ is the final feature map of the UNet right before the output layer, denoted as $\mathcal{F}(\mathbf{X})$, which has a spatial dimension of (C, W, H, D) and C is the number of channels of the feature map. We first use two consecutive 3D convolution layers with a kernel size of $k_1 \times k_2 \times 1$ and $1 \times 1 \times k_3$ to extract $\frac{W}{k_1} \times \frac{H}{k_2} \times \frac{D}{k_3}$ feature patches with C' dimensions each, where C' is also the dimension of the input sequences of the Transformer. Note that previous

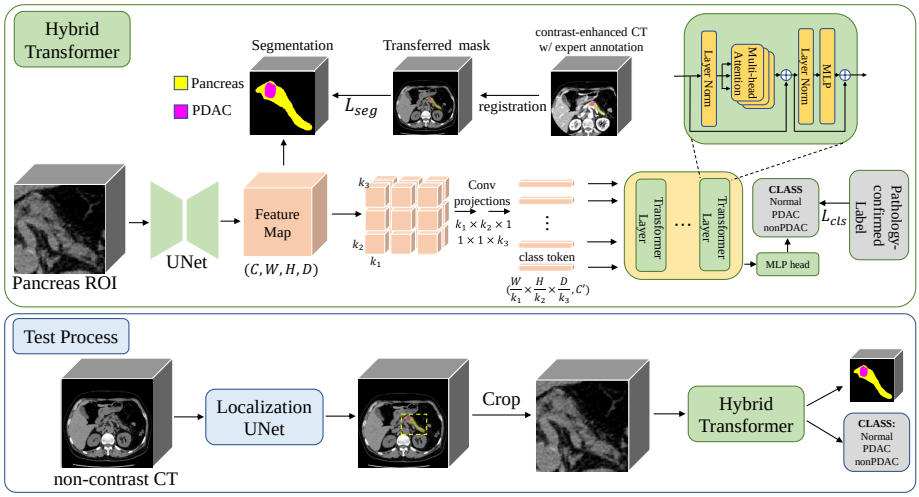


Fig. 1: A visual illustration of our whole framework. Top: we train our hybrid Vision Transformer on non-contrast CT via two supervisions: (i) class label of normal/PDAC/nonPDAC obtained by pathology-confirmed mass type, and (ii) coarse tumor segmentation label transferred from contrast-enhanced CT by registration. Bottom: in the testing phase, we first crop out the pancreas ROI with a localization UNet (separately trained) and output the class and segmentation prediction with the hybrid transformer given non-contrast CT scans.

work [6] directly use one single convolution layer to extract patch features, while we decompose it into two layers to reduce the number of parameters in our 3D settings. Learnable positional embeddings are then added to each patch. These patch features are forwarded through multiple transformer blocks with multi-head attention. Following ViT [6], we also use a class token for classification. The output embedding of the class token is used as the classification prediction after a MLP (multilayer perceptron). Our overall training objective is formulated as follows:

$$\mathcal{L} = L_{seg}(\mathbf{P}_s, \mathbf{Y}) + L_{cls}(\mathbf{P}_c, \mathbf{Z}), \tag{1}$$

where $\mathbf{P}_s = \mathcal{G}(\mathcal{F}(\mathbf{X}))$ and $\mathbf{P}_c = \mathcal{H}(\mathcal{F}(X))$ are the output segmentation of UNet and the final classification prediction of the Transformer, respectively. The loss function for classification is cross-entropy loss.

3 Experiments

Dataset and ground truth. Our dataset of CT scans of 1627 patients, is consecutively collected in the years of 2016∼2018 from a high-volume pancreatic cancer institution. PDAC is of the highest priority among all pancreatic

abnormalities with a 5-year survival rate of approximately 10% and is the most common type (about 90% of all pancreatic cancers). This is the main reason that we group all abnormalities into two classes of PDAC and nonPDAC (including nine subtypes [17,25]).The dataset is randomly split into a training and a testing dataset. The training set includes 450 PDACs, 394 nonPDACs, and 477 normal pancreases. The testing set includes 108 PDACs, 80 nonPDACs, 118 normal pancreases. Both PDAC and nonPDAC cases are confirmed by their pathology reports and normal cases by radiology reports and 2-year follow-up. Each patient has multi-phase CT scans. The median imaging spacing is 0.68×0.68×3.0mm in [X,Y,Z]. The manual annotations of masses are performed by an experienced radiologist (with 14 years of specialized experience in pancreatic imaging) on either arterial/pancreatic phase or venous phase with better mass visibility. The annotations of the pancreas are performed automatically by a segmentation model, which is trained on three datasets, including the single-phase pancreas CTs [15] and abdominal CTs [7] as well as our multi-phase CT dataset, by following a self-training strategy [21,24].

Reader study. Eleven radiologists (four are board-certified pancreatic imaging specialists) from four high-volume pancreatic cancer institutions read the 306 non-contrast CTs in the testing dataset without time constraint (WOTC), with a three-class decision by each reader: PDAC, nonPDAC, or normal.

Implementation details. Each CT volume is firstly resampled into 0.68 × 0.68 × 3.0 mm spacing and normalized into zero mean and unit variance. In the training phase, we crop the foreground 3D bounding box of the pancreas region, randomly pad a small margin on each dimension, and resize the bounding box into a volume of shape (256, 256, 64). The input of the Vision Transformer is the final feature map of the UNet right before the output layer, which has a shape of (32, 256, 256, 64). The two consecutive 3D convolution layers have the kernel size of 32×32×1 and 1×1×8 which leads to 512 feature patches with 256 dimensions each. The transformer contains 12 consecutive 8-head attention blocks. We train our hybrid model in an end-to-end fashion with SGD optimizer. The initial learning rate is set to 1×10^{-3} and decays with a cosine learning rate schedule. In addition to the hybrid UNet-Transformer model, we also trained a standard UNet on the whole image for the localization of pancreas. In the inference phase, we first localize the bounding box of the pancreas region with aforementioned UNet, resize the pancreas region into (256, 256, 64) volume and then classify the pancreas region with our hybrid UNet-Transformer model.

Evaluation methods and metrics. We randomly split the training dataset into a training set (80% data) and a validation set (20% data). Since the primary goal of non-contrast CT screening is to distinguish between abnormal (PDAC+nonPDAC) and normal, a cutoff point (i.e., threshold) is used to dichotomize model's output probabilities into binary predictions. The cutoff point is predefined on the validation set by maximizing the value of (sensitivity + specificity) before model evaluation on the testing set. To further classify the abnormal as PDAC or nonPDAC, the one with a larger output probability is selected as the prediction. We first report the result of the 2-class classification

	2-class			3-class		
Method	AUC	Sensitivity	Specificity	PDAC	nonPDAC	Normal
S4C with UNet [29]	95.98	91.48	95.76	75.00	73.75	95.76
Hybrid CNN	98.25	94.68	94.91	76.85	80.00	94.91
Hybrid Transformer	**98.37**	**95.21**	**95.76**	**78.70**	**80.00**	**95.76**
Mean radiologists WOTC	-	79.66	87.58	53.63	57.96	87.58

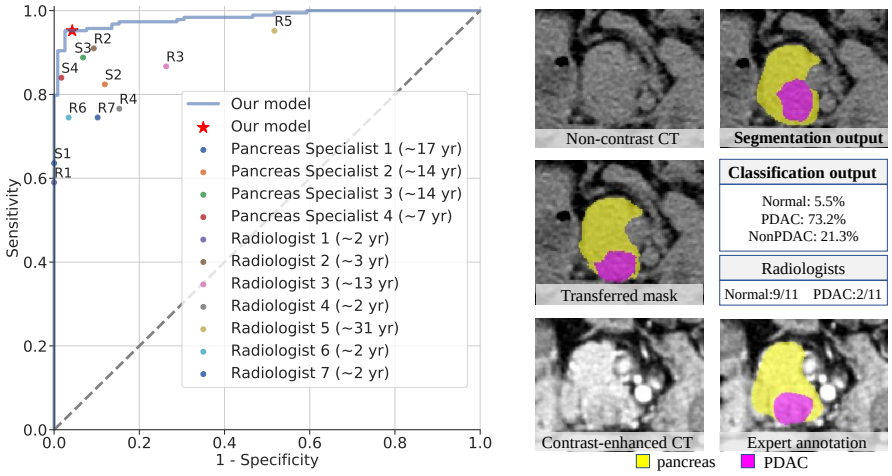
Table 1: Results on two-class classification (PDAC+nonPDAC vs. normal) and three-class classification (PDAC vs. nonPDAC vs. normal). WOTC: without time constraint.

(PDAC + nonPDAC vs. normal). The evaluation metrics include AUC (area under the ROC curve), sensitivity ($\frac{TP}{TP+FN}$) and specificity ($\frac{TN}{TN+FP}$). We also report the result of the 3-class classification (PDAC vs. nonPDAC vs. normal), measured by class accuracy. In addition, the mass detection rate by our model is assessed. A detection is considered successful if the intersection (between the ground truth and segmentation mask) over the ground truth is $> 0\%$ – a coarse localization of mass would be useful in this application scenario.

Compared methods. We compare our method to two baseline approaches. One is "segmentation for classification" [29] full-filled by a standard UNet where we classify a case as positive if the detected PDAC or nonPDAC tumor volume is larger than a certain threshold, which maximize the value of (sensitivity+specificity) on the validation set. The other is a hybrid CNN classifier built on the UNet feature map trained in an end-to-end fashion. Specifically, we integrate a classification head into the segmentation model. We extract multiple level of the UNet feature map, apply global max pooling on each feature map, concatenate them and forward into a single-layer perceptron for classification. Quantitative results are shown in Table 1. The 2-class ROC curve and a case study are shown in Fig 2.

Anatomy-aware transformer outperforms baselines. Compared to two baselines, *i.e.*, segmentation for classification (S4C) and hybrid CNN classifier, our hybrid transformer shows the best performance in all metrics (Table 1), with a relative low STD (about 1%) on sensitivity and specificity. Most medical segmentation models focus on local texture changes and lacks the ability to model the global context. In contrast, our anatomy-aware Hybrid Transformer is built on the locally discriminative features of the UNet, and captures the structural relationship over the whole pancreas region with multi-head attention. Our model is capable of improving the global decision process.

AI models outperforms expert radiologists on non-contrast CT scan. The performance of all 11 radiologists (WOTC) is below our model's ROC curve (Fig 2). Our model has a mass detection rate of 87.76%, and its sensitivity in abnormality prediction (95%) outperforms the mean human performance (80%) by a large margin and also surpasses the best performing radiologist (R2: 91%) and specialist (S3: 89%), which is the main goal of pancreatic cancer screening using non-contrast CTs. More surprisingly, for the 3-class classification task



(a) ROC curve (b) A case study

Fig. 2: (a) ROC diagram for our model result versus all other experts' referrals on the test set of n=306 patients for 2-class classification. The asterisk denotes the performance of our model. Filled markers denote 11 experts' performances using the same non-contrast CT only. S1: Pancreas Specialist 1, R1: Radiologist 1. (b) A case study in the test set. This PDAC case is extremely challenging for radiologists (only 2/11 are correct) given the limited intensity contrast in non-contrast CT scans whereas our model can successfully locate the mass and predicts the class label.

(Table 1), our model outperforms the mean radiologists (WOTC) by absolute margins of 25%, 22%, and 8% for PDAC, nonPDAC, and normal, respectively.

Human vision system requires adequate visual intensity contrast to distinguish mass from pancreas tissue, which is why contrast-enhanced CT scans are necessary for the diagnosis purpose. Given the surprising performance of DL models on non-contrast CT, we empirically hypothesize that machine vision is better at magnifying the local contrast changes to locate masses. Another crucial reason why computerized model substantially outperforming human performance on non-contrast CT is that we transfer the expert findings from contrast-enhanced CT. Most models are restricted to the performance upper bound of the human annotators. With annotations transfer from CECT (a more "doctor-friendly" modality), and pathology-confirmed labels, DL models are equipped with the essential information/knowledge to break the limit of human observers.

Impact and future work. From the reference of computerized performance using contrast-enhanced CT, (sensitivity, specificity) of PDAC vs. Normal is recorded as (92.7%, 99%) [29] and (97.1%, 96%) [20]. This work involves dealing with nonPDAC masses and is generally harder for deep learning [25]. Our performance on non-contrast CT scans (95.2%, 95.8%) is approaching those methods

using contrast-enhanced CT. This finding sheds light on the opportunity to use automated methods to screen pancreatic cancers via non-contrast CT imaging. This may be very beneficial for patients, because non-contrast CT is much cheaper, simpler, and safer than its contrast-enhanced counterpart. We plan to conduct multi-institutional studies to validate the generalizability of our system.

4 Conclusion

In this paper, we explore detecting pancreatic cancer from non-contrast CT scans, as a relatively cheap, convenient, simple and safe imaging modality. We propose a hybrid transformer model which is trained by the supervision of pathology-confirmed mass types and the segmentation masks transferred from contrast-enhanced CT scans. We achieve high sensitivity and specificity on a large-scale dataset and outperform the mean radiologists by large margins. Our work suggests the good feasibility of using non-contrast CT scans as a promising clinical tool for large-scale pancreatic cancer screening.

References

1. Cai, J., Lu, L., Zhang, Z., Xing, F., Yang, L., Yin, Q.: Pancreas segmentation in mri using graph-based decision fusion on convolutional neural networks. In: MICCAI. pp. 442–450. Springer (2016)
2. Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., Zagoruyko, S.: End-to-end object detection with transformers. In: European Conference on Computer Vision. pp. 213–229. Springer (2020)
3. Chu, L.C., Park, S., Kawamoto, S., Fouladi, D.F., Shayesteh, S., Zinreich, E.S., Graves, J.S., Horton, K.M., Hruban, R.H., Yuille, A.L., et al.: Utility of ct radiomics features in differentiation of pancreatic ductal adenocarcinoma from normal pancreatic tissue. American Journal of Roentgenology **213**(2), 349–357 (2019)
4. Chu, L.C., Park, S., Kawamoto, S., Wang, Y., Zhou, Y., Shen, W., Zhu, Z., Xia, Y., Xie, L., Liu, F., et al.: Application of deep learning to pancreatic cancer detection: Lessons learned from our initial experience. Journal of the American College of Radiology **16**(9), 1338–1342 (2019)
5. Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 3d u-net: learning dense volumetric segmentation from sparse annotation. In: MICCAI. pp. 424–432. Springer (2016)
6. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al.: An image is worth 16x16 words: Transformers for image recognition at scale. ICLR (2021)
7. Gibson, E., Giganti, F., Hu, Y., Bonmati, E., Bandula, S., Gurusamy, K., Davidson, B., Pereira, S.P., Clarkson, M.J., Barratt, D.C.: Automatic multi-organ segmentation on abdominal CT with dense V-networks. IEEE Transactions on Medical Imaging **37**(8), 1822–1834 (2018)
8. Heinrich, M.P., Jenkinson, M., Brady, M., Schnabel, J.A.: Mrf-based deformable registration and ventilation estimation of lung ct. IEEE transactions on medical imaging **32**(7), 1239–1248 (2013)

9. Man, Y., Huang, Y., Feng, J., Li, X., Wu, F.: Deep q learning driven ct pancreas segmentation with geometry-aware u-net. IEEE transactions on medical imaging **38**(8), 1971–1980 (2019)
10. Mizrahi, J.D., Surana, R., Valle, J.W., Shroff, R.T.: Pancreatic cancer. The Lancet **395**(10242), 2008–2020 (2020)
11. Oudkerk, M., Liu, S., Heuvelmans, M.A., Walter, J.E., Field, J.K.: Lung cancer ldct screening and mortality reduction—evidence, pitfalls and future perspectives. Nature Reviews Clinical Oncology pp. 1–17 (2020)
12. Roth, H.R., Lu, L., Farag, A., Shin, H.C., Liu, J., Turkbey, E.B., Summers, R.M.: Deeporgan: Multi-level deep convolutional networks for automated pancreas segmentation. In: MICCAI. pp. 556–564. Springer (2015)
13. Roth, H.R., Lu, L., Farag, A., Sohn, A., Summers, R.M.: Spatial aggregation of holistically-nested networks for automated pancreas segmentation. In: MICCAI. pp. 451–459. Springer (2016)
14. Siegel, R.L., Miller, K.D., Jemal, A.: Cancer statistics, 2020. CA: A Cancer Journal for Clinicians **70**(1), 7–30 (2020). https://doi.org/10.3322/caac.21590
15. Simpson, A.L., Antonelli, M., Bakas, S., Bilello, M., Farahani, K., Van Ginneken, B., Kopp-Schneider, A., Landman, B.A., Litjens, G., Menze, B., et al.: A large annotated medical image dataset for the development and evaluation of segmentation algorithms. arXiv preprint arXiv:1902.09063 (2019)
16. Singhi, A.D., Koay, E.J., Chari, S.T., Maitra, A.: Early detection of pancreatic cancer: opportunities and challenges. Gastroenterology **156**(7), 2024–2040 (2019)
17. Springer, S., Masica, D.L., Dal Molin, M., Douville, C., Thoburn, C.J., Afsari, B., Li, L., Cohen, J.D., Thompson, E., Allen, P.J., et al.: A multimodality test to guide the management of patients with a pancreatic cyst. Science Translational Medicine **11**(501) (2019)
18. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I.: Attention is all you need. arXiv preprint arXiv:1706.03762 (2017)
19. Xia, Y., Xie, L., Liu, F., Zhu, Z., Fishman, E.K., Yuille, A.L.: Bridging the gap between 2d and 3d organ segmentation with volumetric fusion net. In: MICCAI. pp. 445–453. Springer (2018)
20. Xia, Y., Yu, Q., Shen, W., Zhou, Y., Fishman, E.K., Yuille, A.L.: Detecting pancreatic ductal adenocarcinoma in multi-phase ct scans via alignment ensemble. In: MICCAI. pp. 285–295. Springer (2020)
21. Xie, Q., Luong, M.T., Hovy, E., Le, Q.V.: Self-training with noisy student improves imagenet classification. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 10687–10698 (2020)
22. Yao, J., Shi, Y., Lu, L., Xiao, J., Zhang, L.: Deepprognosis: Preoperative prediction of pancreatic cancer survival and surgical margin via contrast-enhanced ct imaging. In: MICCAI. pp. 272–282. Springer (2020)
23. 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: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 8280–8289 (2018)
24. Zhang, L., Shi, Y., Yao, J., Bian, Y., Cao, K., Jin, D., Xiao, J., Lu, L.: Robust pancreatic ductal adenocarcinoma segmentation with multi-institutional multi-phase partially-annotated ct scans. In: MICCAI. pp. 491–500. Springer (2020)
25. Zhao, T., Cao, K., Yao, J., Nogues, I., Lu, L., Huang, L., Xiao, J., Yin, Z., Zhang, L.: 3D graph anatomy geometry-integrated network for pancreatic mass segmentation, diagnosis, and quantitative patient management. arXiv preprint arXiv:2012.04701 (2020)

26. Zheng, S., Lu, J., Zhao, H., Zhu, X., Luo, Z., Wang, Y., Fu, Y., Feng, J., Xiang, T., Torr, P.H., et al.: Rethinking semantic segmentation from a sequence-to-sequence perspective with transformers. arXiv preprint arXiv:2012.15840 (2020)
27. 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: MICCAI. pp. 693–701. Springer (2017)
28. Zhu, Z., Xia, Y., Shen, W., Fishman, E., Yuille, A.: A 3d coarse-to-fine framework for volumetric medical image segmentation. In: 2018 International Conference on 3D Vision (3DV). pp. 682–690. IEEE (2018)
29. Zhu, Z., Xia, Y., Xie, L., Fishman, E.K., Yuille, A.L.: Multi-scale coarse-to-fine segmentation for screening pancreatic ductal adenocarcinoma. In: MICCAI. pp. 3–12. Springer (2019)