Holistic Analysis of Abdominal CT for Predicting the Grade of Dysplasia of Pancreatic Lesions

Konstantin Dmitriev and Arie E. Kaufman

Department of Computer Science, Stony Brook University,
Stony Brook NY, USA
{kdmitriev,ari}@cs.stonybrook.edu

Abstract. Diagnosis of various pancreatic lesions in CT images is a challenging task owing to a significant overlap in their imaging appearance. An accurate diagnosis of pancreatic lesions and the assessment of their malignant progression, or the grade of dysplasia, is crucial for optimal patient management. Typically, the grade of dysplasia is confirmed histologically via biopsy, yet certain radiological findings, including extrapancreatic, can serve as diagnostic clues of the disease progression. This work introduces a novel method of transforming intermediate activations for processing intact imaging data of varying sizes with convnets with linear layers. Our method allows to efficiently leverage the 3D information of the entire abdominal CT scan to acquire a holistic picture of all radiological findings for an improved and more precise classification of pancreatic lesions. Our model outperforms current state-of-the-art methods in classifying four most common lesion types (by 2.92%), while additionally diagnosing the grade of dysplasia. We conduct a set of experiments to illustrate the effects of a holistic CT analysis and the auxiliary diagnostic data on the accuracy of the final diagnosis.

1 Introduction

Pancreatic cancer (PC) is one of the most aggressive types of cancer and is currently the fourth most common cause of cancer-related deaths in the United States [1]. PC is often asymptomatic but is associated with distinct precursor cystic lesions, such as intraductal papillary mucinous neoplasms (IPMNs) and mucinous cystic neoplasms (MCNs). Early detection and diagnosis of such lesions offer an opportunity to prevent the progression of the disease. However, cystic pancreatic lesions are a heterogeneous group of lesions, which also include serous cystadenomas (SCAs) and solid-pseudopapillary neoplasms (SPNs), which are considered to have a low malignant potential [23]. Despite the widespread use of high-resolution imaging, such as computed tomography (CT), the non-invasive diagnosis and characterization of pancreatic lesions is still a challenge even for an experienced radiologist due to the overlapping demographic and appearance characteristics (Fig. 1) of lesions [10]. The potential complications and risks

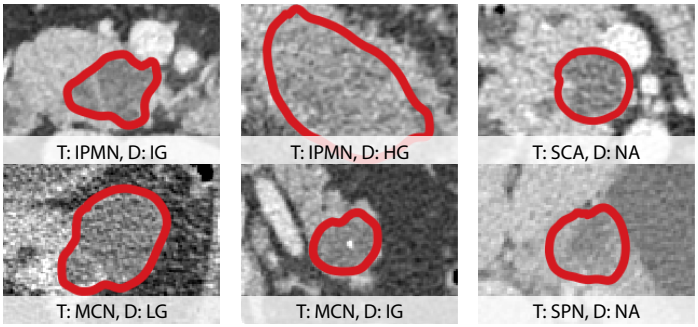

Fig. 1. Types (T) of different pancreatic lesions in CT and associated grades of dysplasia (D), namely, low- (LG), intermediate- (IG), or high-grade (HG) dysplasia.

associated with surgical resection make an accurate non-invasive diagnostic assessment critically important.

Optimal management of pancreatic lesions also relies on the malignant progression of a lesion to estimate which patients would benefit the most from a surgical resection as only a small portion of lesions progress to PC during the lifetime of a patient. The malignant progression of a lesion is associated with the grade of dysplasia, namely, low- (LG), intermediate- (IG), or high-grade (HG) dysplasia. Current guidelines recommend to observe LG and IG precursor lesions and to resect HG lesions to decrease the risk of invasive PC [2]. Traditionally, the grade of dysplasia is determined pathologically via a biopsy procedure. However, some studies show that certain radiological features indicate an association with the HG dysplasia in pancreatic lesions. These features include the dilation of the main pancreatic duct [6] or the common bile duct [19], extrapancreatic malignant neoplasms [4], and lesion size and location [20].

Previous work. Significant progress has been made in the development of computer-aided diagnosis (CAD) systems to aid the clinicians in the process of differentiation of various abnormalities in radiological images [5, 14, 17, 18, 27, 29, 30, 33], including for classification of pancreatic lesions [9, 13, 15, 24, 25, 32]. Latter methods mainly focus on the classification of the pre-segmented lesions into the four most common lesion types, or their sub-groups (benign vs. malignant, SCN vs. non-SCN, etc.), ignoring the grades of dysplasia for lesions with malignant potential. In addition, due to the significant variations in the size of the original 3D regions with lesions, previous methods often divide them into smaller 3D or 2D subvolumes or patches of fixed size. Each patch is then analyzed individually, and the final probabilities are generated as the average between all patches. In such an approach, the spatial relationship between the patches is neglected, leading to the potential inconsistencies between individual classifications and an overall limited diagnostic performance. A similar in spirit, yet different, issue is common in the analysis of gigapixel pathology images [22, 26]. Finally, another common limitation of these works is the diagnosis of pancreatic lesions in isolation from the surrounding vascular structures and organs,

including the pancreas, which leads to the loss of additional potential diagnostic clues.

In this paper, we attempt to approach the diagnosis of pancreatic lesions along with the associated grade of dysplasia in a holistic, rather than an isolated way. We hypothesize that while the radiological appearance of lesions provides some diagnostic clues, a more accurate diagnosis is possible by analyzing the entire peritoneal region in a CT scan, and obtaining a complete picture of all clinical findings, including extrapancreatic. The **contributions** of our paper are the following: a method for efficient transformation of intermediate activations in convolutional networks (convnets) with linear layers to process CT volumes of varying sizes; improved pancreatic lesion classification results (by 2.92%), along with the prediction of the associates grade of dysplasia; and, to the best of our knowledge, this work is the first to describe a lesion classification method based on the holistic analysis of the entire intact abdominal CT scan. The importance of the holistic approach is illustrated by a comprehensive set of experiments.

2 Method

Let $\mathbf{X}^i$ be the available diagnostic records of a patient i with a pathologically confirmed pancreatic lesion of type t^i with the associated grade of dysplasia d^i. These records $\mathbf{X}^i$ include an abdominal CT scan I^i, and the optional binary segmentation masks of the pancreas M_p^i and the lesion M_l^i, and the patient's age a_i at the acquisition time and their gender g_i. The goal is to model a function $\mathcal{F}_\theta(\mathbf{X}^i) = (\hat{\mathbf{t}}^i, \hat{\mathbf{d}}^i)$ using a convnet, where $\hat{\mathbf{t}}^i \in T = \{\text{IPMN, MCN, SCA, SPN}\}$ and $\hat{\mathbf{d}}^i \in D = \{\text{NA, LG, IG, HG}\}$ are the estimated probabilities of the lesion type and dysplasia. Each abdominal scan I^i of size $Z_i \times H_i \times W_i$ is a stack of Z_i axial 2D slices of size $H_i \times W_i$. While the dimensions of each 2D slice are typically unvarying and are equal to 512×512 regardless of the scanning equipment, the number of slices Z_i in each scan varies between patients depending on patient's size and positioning inside the scanner. Unlike the traditional solutions, our proposed model allows us to avoid dividing the input scan I_i into subvolumes of predefined size or individual 2D slices, but keep it intact to preserve the spatial information and to obtain a holistic picture of all pathological findings which may be present in the CT scans along with the pancreatic lesions.

Base Model: The overview of our model is illustrated in Fig. 2. It consists of the convolutional backbone $\mathcal{F}_\theta^1(\mathbf{X}^i) = \mathbf{A}^i$ for feature extraction (Fig. 2(a)), and the classification head $\mathcal{F}_\theta^2(\mathbf{A}^i) = (\hat{\mathbf{t}}^i, \hat{\mathbf{d}}^i)$ (Fig. 2(c)). In other words, our target function can be expressed as $\mathcal{F}_\theta(\mathbf{X}^i) = \mathcal{F}_\theta^2(\mathcal{F}_\theta^1(\mathbf{X}^i)) = (\hat{\mathbf{t}}^i, \hat{\mathbf{d}}^i)$. The backbone $\mathcal{F}_\theta^1$ consists of three convolutional layers of various kernel sizes and strides, each followed by a Leaky ReLU activation function, proceeded by nine ResNet bottleneck blocks [11]. The classification head $\mathcal{F}_\theta^2$ includes an additional $1 \times 1 \times 1$ convolutional layer to reduce the size of the activation maps generated by the backbone and two linear layers which perform the final classification.

The convolutional backbone $\mathcal{F}_\theta^1$ is agnostic to the size $Z_i \times H_i \times W_i$ of the input volumes in $\mathbf{X}^i$, and generates the activation maps $\mathbf{A}^i$ of size $\frac{Z_i}{16} \times \frac{H_i}{16} \times \frac{W_i}{16}$,

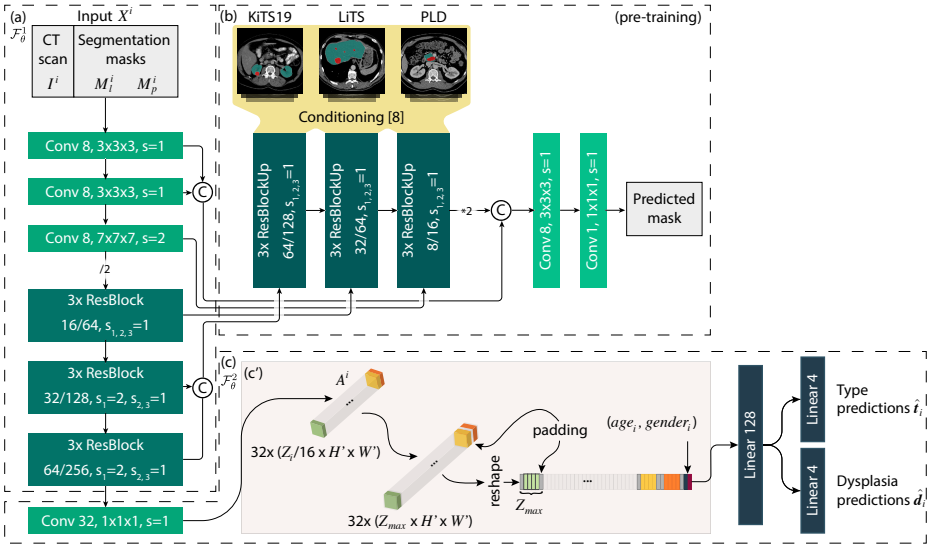

Fig. 2. Schematic view of our classification framework. The main classification model consists of (a) a feature extractor $\mathcal{F}_\theta^1$, or backbone, and (c) a classification head $\mathcal{F}_\theta^2$. The backbone processes a full intact 3D CT scan of arbitrary size $Z_i \times H \times W$ with optional segmentation masks of the pancreas M_p^i and lesion(s) M_l^i provided as a second channel, and generates a tensor of activations maps of size $Z_i/16 \times H' \times W'$. These activation maps are (c') adaptively padded to a predefined size $Z_{max} \times H' \times W'$, flattened, and passed to the classification head which generated the final output. During training, padded activation maps can be organized into batches for faster processing (see Sec. 2). To improve the classification performance, the backbone was augmented with a (b) conditioned decoder and pre-trained for segmentation.

or more specifically, $\frac{Z_i}{16} \times H' \times W'$, as the dimensions $H_i \times W_i$ of 2D slices are constant, but the number of slices Z_i varies between each volume. As a result, the size of the input $\mathbf{A}^i$ to the classification head $\mathcal{F}_\theta^2$ constantly varies. However, traditionally organized networks with linear layers require the input to be of fixed size. Common workarounds include zero-padding of the original 3D volumes in $\mathbf{X}^i$ and adaptive pooling of the final activations to a fixed size. These approaches come with limitations. The former significantly increases memory requirements and computational complexity due to a larger size of the input and the intermediate activation maps. On the other hand, adaptive pooling can potentially lead to a severe loss of spatial information encoded in the activation maps. Instead, we propose a simple, yet effective way of transforming the activation maps to a fixed size, which can be processed by the linear layers. Particularly, we propose to adaptively pad the final activations $\mathbf{A}^i$ generated by the convolutional backbone $\mathcal{F}_\theta^1$ to a pre-defined size Z_{max} before reshaping, or flattening, them to a 1D tensor and passing them to the classification head $\mathcal{F}_\theta^2$. Our proposed approach

allows one to process input data $\mathbf{X}^i$ of varying size, while avoiding the increase of computational complexity and the loss of the important spatial information.

However, as the original data samples $\mathbf{X}^i$ are left unpadded and can significantly vary in size, they cannot be easily collated into batch tensors for more efficient processing that takes advantage of the parallelism. To accelerate the training process, we propose the following: during the forward step, each sample $\mathbf{X}^i$ from a desired batch of size N is first processed individually by the backbone $\mathcal{F}_\theta^1$. Then, the padded activations maps $\mathbf{A}^i$ are combined into a batch tensor of size N and are processed jointly by the classification head $\mathcal{F}_\theta^2$. Consequently, during the backward step, gradients can be accumulated across all N samples in the batch, and the gradient updates can be applied to both $\mathcal{F}_\theta^2$ and $\mathcal{F}_\theta^1$. The efficiency of the proposed framework is evaluated and compared against the traditional input padding in Sec. 3.

Due to many medical datasets being often rather small, training a classification model from scratch might lead to overfitting. Moreover, it is unlikely that the model will learn to reliably detect and use extrapancreatic abnormalities, such as renal lesions, for the diagnosis of pancreatic lesions. Instead, the model will likely use irrelevant features. To alleviate this issue, we propose to pretrain our feature extractor $\mathcal{F}_\theta^1$ for a segmentation task using an additional decoder branch. The architecture of the decoder (Fig. 2(b)) is similar to the backbone, yet significantly narrower. Such an architecture puts more load on the backbone and forces it to learn robust feature representations, potentially at the cost of a limited segmentation performance. Additionally, we condition the decoder [8] to relax the dataset requirements, and to allow us to train a single multi-class segmentation model on multiple binary or ternary, rather than multi-class, segmentation datasets.

3 Experiments

Datasets: To train and evaluate the proposed model, we utilize three datasets of CT scans with various abdominal pathologies. Specifically, during the initial segmentation pre-training step, we use the publicly available Liver Tumor Segmentation (LiTS) dataset [3], the Kidney Tumor Segmentation (KiTS19) dataset [12], and a dataset with pancreatic lesions (PLD). The latter is also used during the final classification training and evaluation steps. The CT images in each dataset were minimally preprocessed by being downsampled from $Z_i \times 512 \times 512$ to $Z_i \times 128 \times 128$ ($Z_i \in [76, 190]$). The information loss due to this step is minimal. Furthermore, this is optional and is not a principal part of our framework. Additionally, we normalize the CT intensities to $[-1, 1]$ range after clipping the original values around $[-300, 300]$ range. This procedure imitates the standard radiologists reading protocol for pancreatic lesion evaluation (similar to window/level adjustments), while also preserving intensity consistency between scans.

LiTS dataset consists of 131 manually annotated contrast-enhanced abdominal CT volumes with various hepatic tumors collected from several clinical sites.

Table 1. Comparison of the segmentation accuracy (mean DSC, %) of our *single* pre-trained conditioned model against previous works evaluated on the same or similar datasets. (Note that outperforming SOTA results was not the goal of this paper.)

Method	LiTS [3]		KiTS19 [12]		Misc.	
	Liver	Tumor	Kidneys	Tumor	Pancreas	Lesion
Ours	83.53	65.45	81.48	75.69	66.05	85.93
Myronenko et al. [31]	-	-	97.42	81.03	-	-
Isensee et al. [16]	95.8	72.5	-	-	-	-
Zhou et al. [34]	-	-	-	-	86.65	63.94

The gender and age distribution of the subjects in the dataset, as well as the used imaging equipment, are unknown.

KiTS19 dataset consists of 210 manually annotated contrast-enhanced CT scans (60% males and 40% females; mean age, 60) with renal lesions collected from several clinical sites. The imaging equipment is unknown.

PLD dataset contains 141 contrast-enhanced abdominal CT scans (44 males and 97 females; mean age, 59 ± 17) collected with Siemens SOMATOM scanner (Siemens Medical Solutions, Malvern, PA), using venous phase protocol. The dataset exhibits examples of the four most common pancreatic lesions: 74 cases of IPMNs, 17 cases of MCNs, 33 cases of SCAs, and 17 cases of SPNs. Each scan was accompanied with the segmentation outlines of the pancreas and the lesion(s) generated by a semi-automatic framework [7]. The histopathological diagnosis for each subject was confirmed by a pancreatic pathologist based on the subsequent resection. The grade of dysplasia on IMPNs and MCNs was noted on a three-grade scale: LG (IPMNs: 18, MCNs: 11), IG (IPMNs: 29, MCNs: 5), HG (IPMNs: 27, MCNs: 1). Examples of various lesions from this dataset are illustrated in Fig. 1.

Data augmentation: The data augmentation routine plays an important role in our method. Each image was augmented with random rotations within $\pm 25^\circ$ range, random vertical and horizontal translations within ± 15 pixels range, and random scaling within $\pm 5\%$ range, during both segmentation pre-training and the final classification training stages. More importantly, as padding the activations extracted by the backbone might introduce an additional source for data overfitting (the model might get biased to the original size of the tensor), we perform random volume clippings (± 15 voxels) along Z direction to alleviate such detrimental effects.

Pre-training: During the initial segmentation pre-training step, our backbone network $\mathcal{F}_\theta^1$ and the additional decoder were trained to minimize a loss function, based on the common Dice Similarity Coefficient (DSC) metric, $\mathcal{L}(M, \hat{M}) = 1 - \frac{2 \sum M \bigodot \hat{M}}{\sum M + \sum \hat{M}}$. We use AdaBound optimizer [28] with the initial rate of $1e{-}4$, and the final rate of 0.01. Prior to training, we defined a lookup table of conditional values for the decoder for each of the six segmentation classes (i.e., liver, liver tumors, kidneys, etc.) with random values sampled from $[-20, 20]$, as suggested by Dmitriev et al. [8]. A mix of all three datasets was randomly split into training,

Table 2. Classification performance comparison across different experiments and competitive models. Different inputs X^i to our model are in parenthesis: raw CT scan; binary lesion mask, M_p; patient's age and gender, AG.

	Type	Dysplasia (4-point scale)	Both	Dysplasia (3-point scale)	Both
Li et al. [25]	72.80%	-	-	-	-
Dmitriev et al. [9]	83.60%	-	-	-	-
Ours (CT)	48.23%	30.49%	19.15%	47.52%	27.66%
Ours (CT + M_p)	85.81%	71.63%	61.70%	72.34%	62.41%
Ours (CT + M_p + AG)	**86.52%**	**75.17%**	**67.37%**	**75.88%**	**68.08%**
Ours (CT + M_p + AG) without pre-training	73.75%	59.57%	35.46%	62.41%	38.29%

validation, and testing sets using a 70/10/20 ratio. The results were evaluated using the DSC metric and are presented in Table 1. We additionally compare our results against previous works evaluated on the same or similar datasets. It is important to emphasize that updating the state-of-the-art (SOTA) segmentation results was not our goal during this pre-training step, especially given the decoder architecture deliberately bounded in performance.

Classification Results and Discussion: Our experiments are conducted on the PLD dataset using stratified 5-fold cross-validation with similar type and dysplasia distributions in training and testing folds. Each training fold was further split into training and validation sets using a 90/10 ratio for early stopping to prevent overfitting. Each experiment considered different key parameters. Particularly, we studied the effects of various combinations of input data X_i on the final performance, namely, the raw CT scan I^i, binary masks of the pancreas M_p^i and the lesion(s) M_l^i, and the age and gender of the patient. Additionally, we compared the impact of segmentation pre-training on classification performance. In each experiment, the model was trained to minimize the joint class-balanced cross-entropy loss (for lesion type and dysplasia) using Adam optimizer [21] with the initial learning rate of $1e-4$. We recorded the overall accuracy of predicting lesion type, the associated grade of dysplasia on 4-point (LG, IG, HG, NA) and 3-point (LG or IG, HG, NA) scales, and both (lesion type and dysplasia together). The final classification results are reported in Table 2.

The model trained only on the raw CT scans performed poorly and did not surpass the 50% accuracy mark for predicting lesions type, and performed even worse on predicting the grade of dysplasia. A likely explanation of such performance is the models inability to deduct the target structure for classification, namely, pancreatic lesions. To bring attention of the model to the pancreas and its lesions, we experimented with augmenting the raw CT scan with binary masks of the pancreas M_p^i, lesion(s) M_l^i, and both, as a second channel. This simply helps to bring the attention of the model to the pancreas and pancreatic lesions within the CT scan rather than to mask out areas outside of the mask. We observed significant, yet compatible, performance improvements in each experiment, and the best results were achieved using only binary masks of the

lesions. Finally, we examined if these results could be improved by feeding the age and gender of the patient to the classification head, as these demographical features are considered important diagnostic clues [10]. In this experiment, we slightly increased the number of units in the first linear layer of the classification head from 128 to 130, to accommodate the extra input, and combined the activations generated by the backbone with patient's age and gender, which were normalized and binarized, respectively. We observed an additional improvement in the performance, outperforming previous solutions [9] by 2.92% in predicting the type, while also predicting the grade of dysplasia with >67% accuracy.

Additionally, we examined the importance of our holistic diagnostic approach, namely, analysis of the entire CT volume without splitting it into smaller pieces and identification of all clinical findings, such as extrapancreatic abnormalities. In our approach, the latter was encouraged through pre-training the backbone $\mathcal{F}_\theta^1$ to detect and segment various lesions, such as renal and hepatic tumors. To evaluate the importance of this step, we compared the performance of the pre-trained model and the model trained from scratch, using the raw CT scan, a binary mask of the lesion, and the age and gender of the patient as the input data. The results are reported in Table 2. Notably, model ability to diagnose the lesion type was affected significantly less than its ability to diagnose the grade of dysplasia. We believe this can be attributed to the model being able to learn to utilize demographical and radiological features of the lesions to correctly predict their type, but not being able to reliably identify additional diagnostic clues outside of the pancreas to diagnose the grade of dysplasia. The results of the experiment support the importance of our holistic diagnostic approach.

Finally, we conducted another experiment to assess and compare the benefits of our proposed adaptive padding of the activation maps against the traditional padding of input data. Particularly, we measured the differences in the memory requirements and computation time required to complete one forward and backward pass between these two approaches. For this experiment, we set $Z_{max} = 12$, which equals to $Z_i = 16 \times 12 = 192$ for traditional padding, and we fixed the number of samples per batch $N = 5$ (batching of the padded activation maps passed to the classification head vs. batching of the padded original input volumes). All experiments were performed on NVidia RTX6000. The proposed method required 2.3 GB of GPU memory as opposed to 6.1 GB when padding the original volumes. While the proposed method was 56.8% slower in this experiment, one can speed up the processing with larger N. The same approach might be impossible when using the traditional padding technique as the memory requirements will increase significantly, especially for very large volumes (e.g., $192 \times 512 \times 512$).

4 Conclusion

We presented a simple, yet effective, method of adaptive padding of intermediate activations for processing intact imaging data of varying size with convnets with linear layers, and an efficient training procedure for such a setup. Our method applied for a holistic diagnosis of the pancreatic lesion showed improved perfor-

mance, while also predicting the associated grades of dysplasia. Despite extensive evaluation and experiments with various sources of input data, our study has several limitations. Specifically, a loss function which explicitly addresses the fact that benign lesions do not have associated grade of dysplasia could be an interesting direction for future work. Furthermore, potential performance gains could be achieved by using or combining images of different modalities. Additionally, a large, potentially multi-center, clinical study is needed to verify the robustness of our system.

Acknowledgments

This research was supported in part by NSF grants NRT1633299, CNS1650499, OAC1919752, and ICER1940302.

References

1. Cancer Facts & Figures. American Cancer Society (2020)
2. Basturk, O., Hong, S.M., Wood, L.D., Adsay, N.V., Albores-Saavedra, J., Biankin, A.V., Brosens, L.A., Fukushima, N., Goggins, M., Hruban, R.H., et al.: A revised classification system and recommendations from the Baltimore consensus meeting for neoplastic precursor lesions in the pancreas. Am J Surg Pathol **39**(12) (2015)
3. Bilic, P., Christ, P.F., Vorontsov, E., Chlebus, G., Chen, H., Dou, Q., Fu, C.W., Han, X., Heng, P.A., Hesser, J., et al.: The liver tumor segmentation benchmark (LiTS). arXiv preprint arXiv:1901.04056 (2019)
4. Buerke, B., Domagk, D., Heindel, W., Wessling, J.: Diagnostic and radiological management of cystic pancreatic lesions: important features for radiologists. Clinical Radiology **67**(8), 727–737 (2012)
5. Chen, T., Ma, X., Liu, X., Wang, W., Feng, R., Chen, J., Yuan, C., Lu, W., Chen, D.Z., Wu, J.: Multi-view learning with feature level fusion for cervical dysplasia diagnosis. Med Image Comput Comput Assist Interv pp. 329–338 (2019)
6. Del Chiaro, M., Beckman, R., Ateeb, Z., Orsini, N., Rezaee, N., Manos, L., Valente, R., Yuan, C., Ding, D., Margonis, G.A., et al.: Main duct dilatation is the best predictor of high-grade dysplasia or invasion in intraductal papillary mucinous neoplasms of the pancreas. Annals of Surgery (2019)
7. Dmitriev, K., Gutenko, I., Nadeem, S., Kaufman, A.: Pancreas and cyst segmentation. Medical Imaging 2016: Image Processing **9784**, 97842C (2016)
8. Dmitriev, K., Kaufman, A.E.: Learning multi-class segmentations from single-class datasets. Proc IEEE Conf Comput Vis Pattern Recognit (June 2019)
9. Dmitriev, K., Kaufman, A.E., Javed, A.A., Hruban, R.H., Fishman, E.K., Lennon, A.M., Saltz, J.H.: Classification of pancreatic cysts in computed tomography images using a random forest and convolutional neural network ensemble. Med Image Comput Comput Assist Interv pp. 150–158 (2017)
10. Farrell, J.J., Fernández-del Castillo, C.: Pancreatic cystic neoplasms: management and unanswered questions. Gastroenterology **144**(6), 1303–1315 (2013)
11. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. Proc IEEE Conf Comput Vis Pattern Recognit pp. 770–778 (2016)

12. Heller, N., Sathianathen, N., Kalapara, A., Walczak, E., Moore, K., Kaluzniak, H., Rosenberg, J., Blake, P., Rengel, Z., Oestreich, M., et al.: The KiTS19 challenge data: 300 kidney tumor cases with clinical context, CT semantic segmentations, and surgical outcomes. arXiv preprint arXiv:1904.00445 (2019)
13. Hu, H., Li, K., Guan, Q., Chen, F., Chen, S., Ni, Y.: A multi-channel multi-classifier method for classifying pancreatic cystic neoplasms based on ResNet. International Conference on Artificial Neural Networks pp. 101–108 (2018)
14. Hussain, M.A., Hamarneh, G., Garbi, R.: ImHistNet: Learnable image histogram based DNN with application to noninvasive determination of carcinoma grades in CT scans. Med Image Comput Comput Assist Interv pp. 130–138 (2019)
15. Hussein, S., Kandel, P., Bolan, C.W., Wallace, M.B., Bagci, U.: Lung and pancreatic tumor characterization in the deep learning era: novel supervised and unsupervised learning approaches. IEEE Trans Med Imaging **38**(8), 1777–1787 (2019)
16. Isensee, F., Petersen, J., Kohl, S.A., Jäger, P.F., Maier-Hein, K.H.: nnu-net: Breaking the spell on successful medical image segmentation. arXiv preprint arXiv:1904.08128 (2019)
17. Jiménez-Sánchez, A., Mateus, D., Kirchhoff, S., Kirchhoff, C., Biberthaler, P., Navab, N., Ballester, M.A.G., Piella, G.: Medical-based deep curriculum learning for improved fracture classification. Med Image Comput Comput Assist Interv pp. 694–702 (2019)
18. Kanayama, T., Kurose, Y., Tanaka, K., Aida, K., Satoh, S., Kitsuregawa, M., Harada, T.: Gastric cancer detection from endoscopic images using synthesis by GAN. Med Image Comput Comput Assist Interv pp. 530–538 (2019)
19. Kawamoto, S., Horton, K.M., Lawler, L.P., Hruban, R.H., Fishman, E.K.: Intraductal papillary mucinous neoplasm of the pancreas: Can benign lesions be differentiated from malignant lesions with multidetector CT? RadioGraphics **25**(6), 1451–1468 (2005)
20. Khashab, M.A., Shin, E.J., Amateau, S., Canto, M.I., Hruban, R.H., Fishman, E.K., Cameron, J.L., Edil, B.H., Wolfgang, C.L., Schulick, R.D., et al.: Tumor size and location correlate with behavior of pancreatic serous cystic neoplasms. American Journal of Gastroenterology **106**(8), 1521–1526 (2011)
21. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
22. Kong, B., Wang, X., Li, Z., Song, Q., Zhang, S.: Cancer metastasis detection via spatially structured deep network. Inf Process Med Imaging pp. 236–248 (2017)
23. Kowalski, T., Ali Siddiqui, D.L., Mertz, H.R., Mallat, D., Haddad, N., Malhotra, N., Sadowski, B., Lybik, M.J., Patel, S.N., Okoh, E., et al.: Management of patients with pancreatic cysts: analysis of possible false-negative cases of malignancy. Journal of Clinical Gastroenterology **50**(8), 649 (2016)
24. LaLonde, R., Tanner, I., Nikiforaki, K., Papadakis, G.Z., Kandel, P., Bolan, C.W., Wallace, M.B., Bagci, U.: INN: Inflated neural networks for IPMN diagnosis. Med Image Comput Comput Assist Interv pp. 101–109 (2019)
25. Li, H., Lin, K., Reichert, M., Xu, L., Braren, R., Fu, D., Schmid, R., Li, J., Menze, B., Shi, K.: Differential diagnosis for pancreatic cysts in CT scans using densely-connected convolutional networks. arXiv preprint arXiv:1806.01023 (2018)
26. Li, Y., Ping, W.: Cancer metastasis detection with neural conditional random field. arXiv preprint arXiv:1806.07064 (2018)
27. Liang, D., Lin, L., Hu, H., Zhang, Q., Chen, Q., Han, X., Chen, Y.W., et al.: Combining convolutional and recurrent neural networks for classification of focal liver lesions in multi-phase CT images. Med Image Comput Comput Assist Interv pp. 666–675 (2018)

28. Luo, L., Xiong, Y., Liu, Y., Sun, X.: Adaptive gradient methods with dynamic bound of learning rate. Proc. of ICLR (May 2019)
29. Luo, L., Chen, H., Wang, X., Dou, Q., Lin, H., Zhou, J., Li, G., Heng, P.A.: Deep angular embedding and feature correlation attention for breast MRI cancer analysis. Med Image Comput Comput Assist Interv pp. 504–512 (2019)
30. Manvel, A., Vladimir, K., Alexander, T., Dmitry, U.: Radiologist-level stroke classification on non-contrast CT scans with deep U-Net. Med Image Comput Comput Assist Interv pp. 820–828 (2019)
31. Myronenko, A., Hatamizadeh, A.: 3d kidneys and kidney tumor semantic segmentation using boundary-aware networks. arXiv preprint arXiv:1909.06684 (2019)
32. Wei, R., Lin, K., Yan, W., Guo, Y., Wang, Y., Li, J., Zhu, J.: Computer-aided diagnosis of pancreas serous cystic neoplasms: a radiomics method on preoperative MDCT images. Technology in Cancer Research & Treatment **18** (2019)
33. Zhao, Z., Lin, H., Chen, H., Heng, P.A.: PFA-ScanNet: Pyramidal feature aggregation with synergistic learning for breast cancer metastasis analysis. Med Image Comput Comput Assist Interv pp. 586–594 (2019)
34. Zhou, Y., Li, Y., Zhang, Z., Wang, Y., Wang, A., Fishman, E.K., Yuille, A.L., Park, S.: Hyper-pairing network for multi-phase pancreatic ductal adenocarcinoma segmentation. Med Image Comput Comput Assist Interv pp. 155–163 (2019)