CURRICULUM KNOWLEDGE SWITCHING FOR PANCREAS SEGMENTATION

Yumou Tang, Kun Zhan⋆, Zhibo Tian, Mingxuan Zhang, Saisai Wang, Xueming Wen

School of Information Science and Engineering, Lanzhou University
`https://github.com/kunzhan/CKS_Pancreas`

ABSTRACT

Pancreas segmentation is challenging due to the small proportion and highly changeable anatomical structure. It motivates us to propose a novel segmentation framework, namely Curriculum Knowledge Switching (CKS) framework, which decomposes detecting pancreas into three phases with different difficulty extent: straightforward, difficult, and challenging. The framework switches from straightforward to challenging phases and thereby gradually learns to detect pancreas. In addition, we adopt the momentum update parameter updating mechanism during switching, ensuring the loss converges gradually when the input dataset changes. Experimental results show that different neural network backbones with the CKS framework achieved state-of-the-art performance on the NIH dataset as measured by the DSC metric.

Index Terms— Pancreas Segmentation, Curriculum Learning, Knowledge Switching

1. INTRODUCTION

This paper focuses on pancreas segmentation from CT-scanned images, which is a crucial step in diagnosing and supporting pancreatic cancer surgery. The segmentation of the pancreas is challenging for two basic reasons: (1) The pancreas only takes up about 1.5% of a typical 2D abdominal multi-organ CT image, which means that both size and resolution have posed significant challenges to the segmentation network; (2) In addition, unlike the heart, kidneys, and other organs, whose geometric shapes are essentially fixed, the pancreatic organs of different people differ greatly.

Most contemporary approaches employ the coarse-to-fine framework [1, 2, 3, 4], in which the coarse network uses the original image as input to determine the approximate pancreatic position and produces a bounding box. The created bounding box is then used to locate the pancreas and crop the original image to obtain a smaller cropped image with a larger pancreas-to-background ratio, followed by the fine network, which is used to segment the cropped image with good precision. The two-stage technique trains two distinct networks separately on large and small images, resolving the issue that the pancreatic content of the original image is too small to segment directly. However, the method is effective but heavily

Fig. 1. Overview of our proposed framework. Several Pancreatic Pixel Discriminators (PPD) are built to progressively switch from straightforward, difficult, and challenging datasets, $\mathcal{D}_1$, $\mathcal{D}_2$, and $\mathcal{D}_3$. Learned knowledge is stored in the two cache models by the momentum update mechanism and is switched to the subsequent PPDs.

dependent on the detection results of the coarse model. Recent efforts tend to directly apply the raw dataset to train the coarse network, and we argue that it is challenging for the model to learn representation on the raw dataset properly because the number of pancreatic pixels in a CT picture is too few to be resolved by the coarse network alone.

According to this observation, we propose a novel Curriculum Knowledge Switching framework inspired by curriculum learning [5]. Instead of directly mapping the input domain to the label domain, we employ four specialized training phases with distinctive datasets $\mathcal{D}_1$, $\mathcal{D}_2$, and $\mathcal{D}_3$ and neural networks to decompose the learning of pancreatic patterns into smaller components. The framework progressively switches from $\mathcal{D}_1$ to $\mathcal{D}_3$ in order to learn from detecting the pancreatic region properties to extracting the pancreatic-specific pixels. In order to effectively inherit knowledge learned in previous stages, we employ the momentum update mechanism to incrementally switch the learned knowledge to two cache models, as shown in Fig. 1.

Experimental results on the NIH dataset reveal that our framework with multiple backbone types outperforms state-of-the-art approaches based on the DSC metric. The CKS

framework incrementally enhances segmentation performance and identifies the global optimum of precise pancreas segmentation.

The main contributions of this paper are listed as follows:

- We propose a Curriculum Knowledge Switching (CKS) framework to teach neural networks constructing pancreas discriminate ability in a consecutive manner, which starts from straightforward to challenging problems.

- The momentum update method is used for knowledge switching, ensuring the learned knowledge from previous stages inherits to the following tasks.

- We employ a data augmentation strategy to refine the learning of the segmentation stage. Experimental results have shown the superiority of the proposed frameworks across different backbones.

2. METHODOLOGY

2.1. The Curriculum Knowledge Switching Framework

We observe that semantic segmentation models for the pancreas could be regarded as Pancreatic Pixel Discriminators (PPDs), which are utilized to estimate pancreatic pixel locations while filtering out background pixels. Due to the varying shape of the pancreas organ, accurately detecting the pancreas poses great challenges. We construct units with suitable levels of complexity to overcome this issue, enabling the PPDs to gradually learn the location and specific pixels of the pancreas.

In this work, we set three phases with varying difficulty levels, straightforward, difficult, and challenging, to progressively guide the framework to identify the pancreas from the complicated CT scan background, as shown in Fig. 2. These three steps correspond to three training sets: the ground-truth bounding box cropped from the label set ($\mathcal{D}_1$) dominated by pancreatic pixels; the cropped bounding box predicted by the coarse model ($\mathcal{D}_2$), which has more background than $\mathcal{D}_1$ and contains boundary ambiguity pixels; and the raw image ($\mathcal{D}_3$), which is the raw CT slices and is mainly occupied by the background pixels. At each level, we assign a distinctive PPD to map the corresponding domains.

Since the model of the detection stage learns the requisite pancreatic characteristics, we wish to switch the learned knowledge from the detection stage to the segmentation stage during the succeeding segmentation step.

2.1.1. Momentum Update for Knowledge Transferring

As we utilize multiple PPDs to locate and learn precise features of the pancreas, retaining the performance of the trained model when switching between phases is necessary. We implement two methods for effective knowledge switching. One is to directly copy the trained model to the second one, and the other

Fig. 2. The detection stage of the proposed framework. First, f_1 is trained by the image cropped by its label in $\mathcal{D}_1$. Second, the detection cache model f_{mu} is utilized to produce the training set $\mathcal{D}_2$ and the f_2 is further trained by datasets in $\mathcal{D}_2$. Lastly, the PPD f_3 is trained by the image in $\mathcal{D}_3$.

is to incrementally collect the acquired information, introduced as follows:

A deep neural network model is denoted by $f(X|\Theta)$, where X is the input image and Θ is the model weight parameters, updated by the back-propagation algorithm. The momentum update approach gradually updates the model parameters Θ_{mu} from Θ by:

$$\Theta_{\mathrm{mu}} = m\Theta_{\mathrm{mu}} + (1-m)\Theta \tag{1}$$

where $m \in [0,1]$ is a momentum coefficient that defines to what extent the historical models should influence the current model. Accordingly, the momentum update method is able to update Θ_{mu} smoothly [6].

2.1.2. Detection Stage

From Fig. 2, the detection stage of our CKS system employs three PPDs $\{f_1, f_2, f_3\}$ and three training sets $\{\mathcal{D}_1, \mathcal{D}_2, \mathcal{D}_3\}$. The PPDs switch from $\mathcal{D}_1$ to $\mathcal{D}_3$ and acquire the ability to detect the pancreas. In addition, we build an extra cache model, namely Detection Cache, to accumulate the knowledge learned in the three phases with the momentum update mechanism, making it more robust and stable than all three PPDs.

Phase I In $\mathcal{D}_1$, the proportion of the pancreas is relatively large, and the positive and negative examples are well balanced. We employ a PPD here, namely f_1 to learn detailed patterns of the pancreas. In this phase, the input $X \in \mathcal{D}_1$ is cropped by utilizing the ground-truth T to generate image X' and we feed X' to f_1,

$$\begin{cases} X' = \mathrm{crop}(X, T) \in \mathcal{D}_1 \\ P_1 = f_1(X'|\Theta_1) \end{cases} \tag{2}$$

where the first PPD is denoted by $f_1(X'|\Theta_1)$, P_1 is the prediction of f_1, and Θ_1 is its parameters, updated by the back-propagation algorithm. The parameters of the detection cache,

Θ_{mu}, are updated by the momentum update:

$$\Theta_{\mathrm{mu}} = \alpha\Theta_{\mathrm{mu}} + (1-\alpha)\Theta_1 \qquad (3)$$

After the first phase, $f_{\mathrm{mu}}(X|\Theta_{\mathrm{mu}})$ is able to identify the overall shape of the pancreas and yet failed to extract pancreas features from datasets with low pancreas pixels ratio. This issue is further addressed in the second phase.

Phase II   In this phase, the detection cache model produces candidate frames with more boundary ambiguity pixels and a larger percentage of non-pancreatic pixels. The produced training set is $\mathcal{D}_2$. Using the pancreatic pixel discrimination capabilities of f_{mu}, the raw picture $X \in \mathcal{D}_3$ is cropped to generate X''.

$$\begin{cases} P_{\mathrm{mu}} = f_{\mathrm{mu}}(X|\Theta_{\mathrm{mu}}), \\ X'' = \mathrm{crop}(X, P_{\mathrm{mu}} > 0.5) \in \mathcal{D}_2 \\ P_2 = f_2(X''|\Theta_2) \\ \Theta_{\mathrm{mu}} = \alpha\Theta_{\mathrm{mu}} + (1-\alpha)\Theta_2 \end{cases} \qquad (4)$$

With $\mathcal{D}_2$, we employ a PPD, f_2, with weights initialized by Θ_1 to guide the framework switching to the difficult phase of learning. We argue that switching from $\mathcal{D}_1$ to $\mathcal{D}_2$ bridges a critical step to guarantee the final detection accuracy.

Phase III   In the third phase, $i.e.$, the challenging phase, we intend to guide the framework to extract pancreatic features directly from the raw CT scan input slices $\mathcal{D}_3$. We employ a PPD, f_3, with weights initialized by Θ_2. This step renders the f_3 consistent with the real situation and obtains high pancreatic pixel discriminative ability. The learned knowledge is switched and stored in the Detection Cache model with the momentum update mechanism. This phase is formulated as:

$$\begin{cases} P_3 = f_3(X|\Theta_3), \\ \Theta_{\mathrm{mu}} = \alpha\Theta_{\mathrm{mu}} + (1-\alpha)\Theta_3 \end{cases} \qquad (5)$$

Notably, our strategy is heavily based on the momentum update mechanism, which guarantees effective knowledge switching. The detailed pancreatic properties learned during the first phase of training are switched to f_2 in the second phase, passing the first phase categorization capabilities to the second, and the momentum update mechanism provides additional information regarding the peripheral characteristics of the pancreas to f_2 during the second phase. In the third phase, dataset $\mathcal{D}_3$ enhances the model PPD ability to generalize from real-world scenarios. The entire training process is dynamic, establishing not only a balance between the detection and segmentation stages but also high performance.

2.1.3. Segmentation Stage

This stage aims to precisely segment the pancreas area according to the detection of the previous stage. The pixel discrimination abilities vary at various stages of the detection training. However, following the completion of the three-step coarse

training, the Detection Cache model f_{mu} is fixed, and its prediction behaves identically to the actual test circumstance. Therefore, we combine $\mathcal{D}_2$ into $\mathcal{D}_1$ to train the segmentation model f^{seg}, allowing the framework to switch from detecting the pancreas to extracting the detailed features of the pancreas. The segmentation model parameters are also stored in the Segmentation Cache Model $f_{\mathrm{mu}}^{\mathrm{seg}}$ with the momentum update, which is typically more reliable and smoother than f^{seg}.

2.2. Prediction

The training stage yields two networks: the Cache Detection Model f_{mu} and the Cache Segmentation Model $f_{\mathrm{mu}}^{\mathrm{seg}}$. The prediction workflow is as follows: First, we feed the test image X forward: $P_{\mathrm{det}} = f_{\mathrm{mu}}(X|\Theta_{\mathrm{mu}})$. Second, we crop image X to produce $X'' \in \mathcal{D}_2$ by X, $X'' = \mathrm{crop}(X, P_{\mathrm{det}} > 0.5)$. Third, we feed X'' forward: $P_{\mathrm{seg}} = f_{\mathrm{mu}}^{\mathrm{seg}}(X''|\Theta_{\mathrm{mu}}^{\mathrm{seg}})$. Fourth, we do thresholding on P_{seg} to obtain the segmentation result.

2.3. Supervision

In this paper, we adopt the intersection-over-union loss for model learning supervision:

$$\mathcal{L}_{\mathrm{iou}} = 1 - \frac{\sum_{ij} t_{ij} p_{ij}}{\sum_{ij} t_{ij} + \sum_{ij} p_{ij} - \sum_{ij} t_{ij} p_{ij}}. \qquad (6)$$

where $T = [t_{ij}]$ denotes the ground-truth binary label matrix and $P = [p_{ij}] = f(X|\Theta)$ is the prediction probability of pancreatic pixels.

Since the Pancreatic Pixel Discriminator is regarded as the binary classifier, we add the binary cross-entropy loss for better supervision:

$$\mathcal{L}_{\mathrm{bce}} = -\sum_{ij} \big(t_{ij} \ln p_{ij} + (1-t_{ij}) \ln(1-p_{ij}) \big). \qquad (7)$$

As the adjacent pancreatic pixels have a high probability of being pancreatic area [7], we use Gaussian kernel smoothed $\hat{T} = [\hat{t}_{ij}]$ and $\hat{P} = [\hat{p}_{ij}]$ to define a new loss, and it is formulated as:

$$\mathcal{L}_{\mathrm{s}} = 1 - \frac{1}{N} \sum_{ij} \frac{2\hat{t}_{ij}\hat{p}_{ij}}{\hat{t}_{ij}^2 + \hat{p}_{ij}^2} \qquad (8)$$

where N is the number of pixels of an image.

The overall loss is given by:

$$\mathcal{L}(P, T) = \mathcal{L}_{\mathrm{iou}} + \mathcal{L}_{\mathrm{bce}} + \mathcal{L}_{\mathrm{s}}. \qquad (9)$$

3. EXPERIMENTAL RESULTS

3.1. Experimental Setting

We validate the CKS framework using the NIH pancreas dataset [1] consisting of 82 contrast-enhanced abdominal CT volumes. Each CT volume is cut into a set of 2D slices and

Fig. 3. Results of the detection stage of our proposed **CKS** framework comparing to FPM [3] and RSTN [4] with two backbones (FCN-8s [8] and U-Net [9]).

Fig. 4. The detection stage of **CKS** adapts real domain $\mathcal{D}_3$ gradually. Note that the prediction error (highlighted in green color) decreases as the framework switches to the challenging phase, along with the improvement of the DSC metric.

Table 1. DSC performance of the proposed **CKS** algorithm with thresholds(%) following [3].

Method	Mean	Max	Min
Detection	83.94	91.01	58.45
$d_t > 0.95$	**85.42 ± 4.39**	91.57	67.59
$d_t > 0.98$	85.41 ± 4.44	91.58	67.38
$d_t > 0.99$	85.37 ± 4.43	91.61	66.50

Table 2. DSC performance of **CKS** and SOTA methods (%). Methods with FCN-8s are above the double lines and the ones with U-Net are below the double lines.

Method	Mean	Max	Min
DeepOrgan [1]	71.42±10.11	86.29	23.99
Roth *et al.* [10]	78.01±8.20	88.65	34.44
FPM [3]	82.37±5.68	90.85	62.43
Roth *et al.* [11]	81.27±6.27	88.96	50.69
Cai *et al.* [12]	82.40±6.70	90.10	60.00
RSTN [4]	84.50±4.97	91.02	62.81
Zhu *et al.* [13]	84.59 ± 4.86	91.45	69.92
Zhang *et al.* [14]	84.47 ± 4.36	91.54	**70.61**
STFFM [15]	84.90	91.46	61.82
U-Net [9]	79.70 ± 7.60	89.30	43.40
Attention U-Net [16]	83.10 ± 3.80	89.38	66.77
CKS	**85.42** ± 4.39	**91.57**	67.59

3.3. Overall Performance

This section reports the prediction results. We utilize U-net as our network backbone in line with methods listed in Table 2.

Following [3], DSC metrics of the detection model and the overall model under different thresholds are summarized in Table 1. The comparison results are provided in Table 2. For the sake of a fair comparison, the metrics are from their papers or reproducing the publicly available code. The outcome demonstrates improvement over the state-of-the-art algorithms.

4. CONCLUSION

We present a novel Curriculum Knowledge Switching (CKS) framework that directs the Pancreas Pixel Discriminators with designed adaptive phases to incrementally switch from easy to difficult training sets and acquire the capacity to localize and segment the pancreas gradually. The augmented dataset $\mathcal{D}_2$ is employed to train the second phase of the detection stage, bridging the direct switching from $\mathcal{D}_1$ to $\mathcal{D}_3$. Moreover, we utilize the momentum update to facilitate progressive knowledge switching, inheriting the learned knowledge to the subsequent ones. Experimental results demonstrate that CKS surpasses the state-of-the-art accuracy and is helpful for computer-assisted clinical diagnosis and detection of small objects.

each slice is cut along three axes, *i.e.*, the coronal, sagittal, and axial views. Following [3], experiments are conducted in a cross-validation manner on four-folds. The test set contains 20 images and each image has at least 5,000 valid slices in both directions, and thereby the reliability of our results is guaranteed. Finally, the Dice-Sørensen coefficient (DSC) is exploited to measure performance.

3.2. Detection Stage Performance

According to Fig. 3, the detection stage results of CKS outperform the performance of FPM [3] and RSTN [4] under the same baseline, which is beneficial to the subsequent segmentation stage.

Fig. 4 depicts the process of the proposed CKS framework progressively switching to the final detection result across three difficulty levels in detail. As the figure shows, the CKS framework guides the PPDs switching from $\mathcal{D}_1$ to $\mathcal{D}_3$ with the DSC metrics increasing steadily. Experiments reveal that training set $\mathcal{D}_2$ contains more non-pancreatic pixels than $\mathcal{D}_1$ and numerous pixels with border ambiguity data. The second training phase is closer to the real-world situation than $\mathcal{D}_2$ and gives more attention to ambiguous locations.

The detection result without the constructed first and the second phases show inferior performance, which are 81.33% and 81.59% in terms of the DSC metric, respectively, as compared to the result of 85.42% with three full training phases.

5. REFERENCES

[1] Holger R Roth, Le Lu, Amal Farag, Hoo Chang Shin, Jiamin Liu, Evrim Turkbey, and Ronald M Summers, "DeepOrgan: Multi-level deep convolutional networks for automated pancreas segmentation," in *MICCAI*, 2015, pp. 556–564.

[2] Amal Farag, Le Lu, Holger R Roth, Jiamin Liu, Evrim Turkbey, and Ronald M Summers, "A bottom-up approach for pancreas segmentation using cascaded superpixels and (deep) image patch labeling," *TIP*, vol. 26, no. 1, pp. 386–399, 2017.

[3] Yuyin Zhou, Lingxi Xie, Wei Shen, Yan Wang, Elliot K Fishman, and Alan L Yuille, "A fixed-point model for pancreas segmentation in abdominal CT scans," in *MICCAI*, 2017, pp. 693–701.

[4] Qihang Yu, Lingxi Xie, Yan Wang, Yuyin Zhou, Elliot K Fishman, and Alan L Yuille, "Recurrent saliency transformation network: Incorporating multi-stage visual cues for small organ segmentation," in *CVPR*, 2018, pp. 8280–8289.

[5] Yoshua Bengio, Jérôme Louradour, Ronan Collobert, and Jason Weston, "Curriculum learning," in *ICML*, 2009, pp. 41–48.

[6] Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick, "Momentum contrast for unsupervised visual representation learning," in *CVPR*, 2020, pp. 9729–9738.

[7] Rafael Müller, Simon Kornblith, and Geoffrey E Hinton, "When does label smoothing help?," *NeurIPS*, vol. 32, 2019.

[8] Evan Shelhamer, Jonathan Long, and Trevor Darrell, "Fully convolutional networks for semantic segmentation," *TPAMI*, vol. 39, no. 4, pp. 640–651, 2017.

[9] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-net: Convolutional networks for biomedical image segmentation," in *MICCAI*, 2015, pp. 234–241.

[10] HR Roth, L Lu, A Farag, A Sohn, and RM Summers, "Spatial aggregation of holistically-nested networks for automated pancreas segmentation," in *MICCAI*, 2016, pp. 451–459.

[11] Holger R. Roth, Le Lu, Nathan Lay, Adam P. Harrison, Amal Farag, Andrew Sohn, and Ronald M. Summers, "Spatial aggregation of holistically-nested convolutional neural networks for automated pancreas localization and segmentation," *Medical Image Analysis*, pp. 94–107, 2018.

[12] Jinzheng Cai, Le Lu, Yuanpu Xie, Fuyong Xing, and Lin Yang, "Improving deep pancreas segmentation in CT and MRI images via recurrent neural contextual learning and direct loss function," in *MICCAI*, 2017.

[13] Zhuotun Zhu, Yingda Xia, Wei Shen, Elliot Fishman, and Alan Yuille, "A 3D coarse-to-fine framework for volumetric medical image segmentation," in *3DV*. IEEE, 2018, pp. 682–690.

[14] Yue Zhang, Jiong Wu, Yilong Liu, Yifan Chen, Wei Chen, Ed X Wu, Chunming Li, and Xiaoying Tang, "A deep learning framework for pancreas segmentation with multi-atlas registration and 3D level-set," *Medical Image Analysis*, vol. 68, pp. 101884, 2021.

[15] Dingwen Zhang, Jiajia Zhang, Qiang Zhang, Jungong Han, Shu Zhang, and Junwei Han, "Automatic pancreas segmentation based on lightweight DCNN modules and spatial prior propagation," *Pattern Recognition*, vol. 114, pp. 107762, 2021.

[16] Ozan Oktay, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Y Hammerla, Bernhard Kainz, et al., "Attention U-net: Learning where to look for the pancreas," in *MIDL*, 2018.