HACL-Net: Hierarchical Attention and Contrastive Learning Network for MRI-Based Placenta Accreta Spectrum Diagnosis

Mingxuan Lu[1], Tianyu Wang[2(✉)], Hao Zhu[3], and Mian Li[2]

[1] Columbia University in the City of New York, New York, USA
[2] Shanghai Jiao Tong University, Shanghai, China
gunnerwang27@sjtu.edu.cn
[3] Obstetrics and Gynecology Hospital affiliated to Fudan University, Shanghai, China

Abstract. Placenta Accreta Spectrum (PAS) can lead to high risks like excessive blood loss at the delivery. Therefore, prenatal screening with MRI is essential for delivery planning that ensures better clinical outcomes. For computer-aided PAS diagnosis, existing work mostly extracts radiomics features directly from ROI while ignoring the context information, or learns global semantic features with limited awareness of the focal area. Moreover, they usually select single or few MRI slices to represent the whole sequences, which can result in biased decisions. To deal with these issues, a novel end-to-end Hierarchical Attention and Contrastive Learning Network (HACL-Net) is proposed under the formulation of a multi-instance problem. Slice-level attention module is first designed to extract context-aware deep semantic features. These slice-wise features are then aggregated via the patient-level attention module into task-specific patient-wise representation for PAS prediction. A plug-and-play contrastive learning module is introduced to further improve the discriminating power of extracted features. Extensive experiments with ablation studies on a real clinical dataset show that HACL-Net can achieve state-of-the-art prediction performance with the effectiveness of each module.

Keywords: Placenta Accreta Spectrum · MRI · Multi-Instance Learning · Hierarchical Attention · Contrastive Learning

1 Introduction

Placenta Accreta Spectrum (PAS) occurs when the placenta becomes abnormally adherent to the myometrium rather than the uterine decidua [15]. The primary risk of PAS is hemorrhage and even life-threatening associated complications [15]. Particularly, it can lead to excessive blood loss and transfusions of blood products at the delivery [9]. Prenatal screening helps identify women

with potential PAS and allows for appropriate delivery planning, which is critical for better clinical outcomes. Meanwhile, Magnetic Resonance Imaging (MRI) features high resolution and sensitivity, playing an increasingly important role in prenatal screening [12]. MRI is often used after an inconclusive ultrasonic examination to assess the invasive condition of PAS [9]. A detailed background illustration is included in Section A of supplementary material.

In the field of computer-aided PAS diagnosis for prenatal screening, some work based on traditional machine learning has been proposed [7,11–13] and achieved promising progress. In these works, radiomics features were first extracted from the Region of Interest (ROI, placenta) of raw images and a classifier was then fitted between the features and clinical outcomes. In [12], Random Forests (RF), K Nearest Neighbor (KNN), Naive Bayes (NB), and Multi-Layer Perception (MLP) were compared. Statistical analysis and multivariate logistic regression were employed in [13]. The features are mostly extracted only on placenta ROI with Pyradiomics [17] or predefined rules, ignoring the context information such as the location of the placenta and its interaction with nearby organs that are important in PAS diagnosis [10].

On the other hand, deep learning stands out in terms of automatic deep feature extraction [24] and end-to-end learning fashion [3]. Deep learning has been recently explored for image-based PAS diagnosis [20,23]. A deep dynamic convolution neural network with autoencoder training manner was proposed in [20] to extract deep features. In [23] the encoder of a placenta segmentation network was employed to capture the semantic features. To enhance such deep features with the awareness of the focal area, attention mechanisms [25] are considered in this paper. The ROI relevant to the prediction task can then be automatically localized, leading to more comprehensive features that capture both semantic and context information. Moreover, prior knowledge about the focal area can be incorporated by annotating a limited number of images with ROIs. That is more efficient than directly including the localization task in addition to the prediction task, which requires extra pixel-level annotation of each image.

Another gap to fill is the inherent diversity of MRI sequences. There are T1-weighted and T2-weighted MRI scans, upon which three planes with multiple slices are further present. Therefore, each patient corresponds to dozens of slice images. The physicians usually need to inspect a large portion of these images in order to identify the focus of suspected PAS and reach the patient-level diagnosis. For the existing work, a single (or few) slice image is selected to represent the information of a patient, which can lead to potentially biased results. Multi-Instance Learning (MIL), as a potential solution, has been widely adopted in lesion classification and localization [19,21,22]. Some recent works, such as slice attention transformer and convolutional transformer-based MIL, have studied the problem of patient as bag and images as instances [6,8,16]. It remains to explore how MIL can benefit the computer-aided PAS diagnosis.

To deal with the issues above, a novel end-to-end Hierarchical Attention and Contrastive Learning Network (HACL-Net) is proposed under the formulation of a multi-instance problem. Slice-level spatial residual attention with prior knowledge is designed to extract context-aware deep semantic features from individual
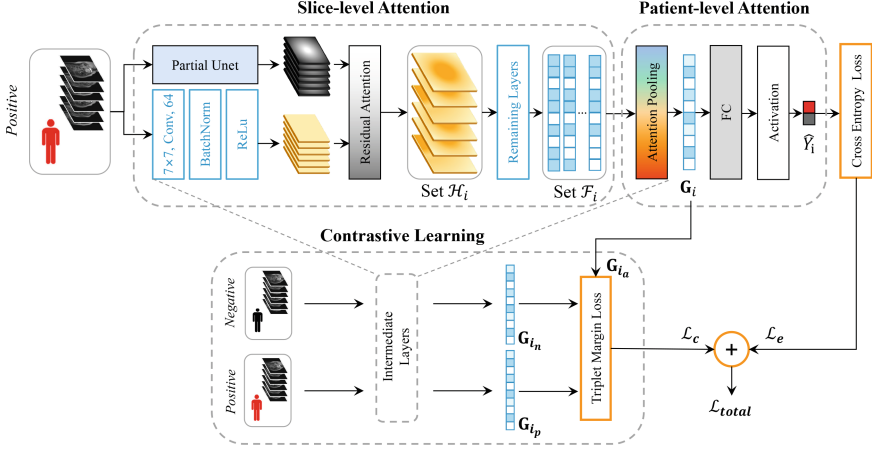
Fig. 1. An architecture overview of the proposed HACL-Net.

MRI slices of a patient. Patient-level attention-based pooling is then applied to aggregate these features into patient representation for PAS prediction. To make such features more discriminative, a plug-and-play contrastive loss is further included. Extensive experiments with ablation study show the proposed network can achieve state-of-the-art performance on a real clinical dataset involving 359 distinct patients.

2 Methodology

Consider each patient bag $\mathcal{P}_i$ includes K_i valid 2D MRI slices from axial, sagittal, and coronal views (possibly both T1-weighted and T2-weighted scans exist). In the training dataset, there is an associated binary label Y_i indicating whether PAS is reported at delivery. Denote $\mathcal{P}_i = \{\mathcal{X}_i, Y_i\}$, $\mathcal{X}_i = \{\mathbf{X}_i^{(1)}, \cdots, \mathbf{X}_i^{(K_i)}\}$, where $\mathbf{X}_i^{(j)}$ is the j-th MRI slice of i-th patient. Note that the number K_i can vary with different patients. The proposed HACL-Net aims to learn the relationship between $\mathcal{X}_i$ and Y_i in a weekly-supervised and end-to-end manner (Fig. 1). The slice-level attention module is designed to extract context-aware deep semantic features from each MRI slice. Then the slice-wise features are fused in a task-specific way via the patient-level attention module, after which being connected to FC layer for PAS prediction. A plug-and-play contrastive learning module is presented to facilitate discriminative patient-wise features between classes.

2.1 Slice-Level Spatial Residual Attention

Trunk Branch. The trunk branch aims to extract global semantic information from each MRI slice that is important for the PAS prediction task. It is composed of the top convolution layer, batch normalization layer, and ReLU layer from
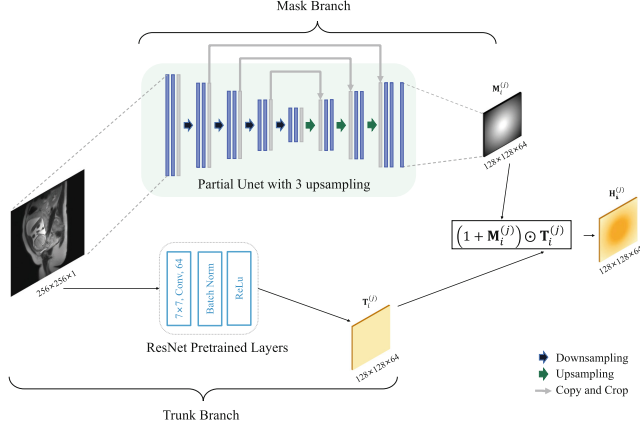
Fig. 2. The details of slice-level attention module with a two-branch structure.

ResNet18. For each grayscale MRI slice $\mathbf{X}_i^{(j)} \in \mathbb{R}^{256 \times 256 \times 1}$, a preliminary feature map $\mathbf{T}_i^{(j)} \in \mathbb{R}^{128 \times 128 \times 64}$ is generated. Note that to realize the spatial residual attention with the mask branch, shallow features are first derived since they can be aligned with the context features. Deep features will be extracted later after the spatial residual attention layer.

Mask Branch. The mask branch aims to extract the context information that indicates the location of ROI. Unlike the traditional implementation where a general bottom-up top-down network is employed, here the U-Net dedicated for medical image segmentation [14] is considered because the placenta area accounts for most of the ROI. Automatic segmentation of placenta with U-Net has proved effective in recent years [4]. The original network consists of 4 downsampling and 4 upsampling modules. To preserve the context information surrounding the placenta, a partial U-Net is constructed, as shown in Fig. 2. The last upsampling module to generate the single-channel mask ($\mathbb{R}^{256 \times 256 \times 1}$) is removed and the intermediate feature map $\mathbf{M}_i^{(j)} \in \mathbb{R}^{128 \times 128 \times 64}$ is obtained as the output of mask branch. Note that the entire U-Net is first pre-trained with a small number of placenta annotated MRI slices. Then the partial U-Net gets fine-tuned with a small learning rate in the end-to-end prediction task.

Spatial Residual Attention and Deep Feature Extraction. The mask branch output $\mathbf{M}_i^{(j)}$ has been aligned with the trunk branch output $\mathbf{T}_i^{(j)}$ in terms of the field of view. Therefore, $\mathbf{M}_i^{(j)}$ is multiplied in an element-wise manner with $\mathbf{T}_i^{(j)}$ as the attention mechanism. Besides, to avoid potential performance drop with attention, residual learning is further considered [18]. The shallow features after spatial residual attention become $\mathbf{H}_i^{(j)} = (1 + \mathbf{M}_i^{(j)}) \odot \mathbf{T}_i^{(j)}$. Note that the
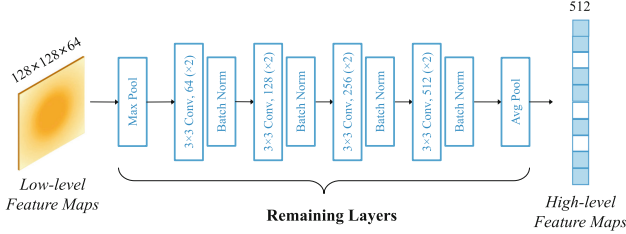
Fig. 3. Remaining layers of pre-trained ResNet18 except for the top three layers.

attention is not performed directly on deep features because the ROI considered has explicit physical meanings instead of being "hidden".

The remaining layers of pre-trained ResNet18 (excluding the top three layers used previously) are employed to further extract deep features, as shown in Fig. 3. These layers are fine-tuned in end-to-end learning so that task-specific deep features can be obtained. For each patient bag, a set of slice-wise deep features $\mathcal{F}_i = \{\mathbf{F}_i^{(1)}, \cdots, \mathbf{F}_i^{(K_i)}\}$ are eventually generated, where $\mathbf{F}_i^{(j)} \in \mathbb{R}^{512}$. These features capture the context-aware semantic information of MRI slices.

2.2 Patient-Level Attention-Based Pooling

Due to the MIL nature of the studied problem, MRI slices in a patient bag contribute unevenly to the PAS diagnosis. There are no detailed annotations to reflect the role of each slice. Therefore, patient-level attention mechanism is further proposed to aggregate the deep features from each slice in a task-specific manner, *i.e.*, deriving the patient-wise features $\mathbf{G}_i$ based on a set of slice-wise features $\mathcal{F}_i$. Some recent work [5,22] has proved that attention-based pooling is useful for feature aggregation in MIL problems. It shows certain similarities with channel attention which has been widely applied in image classification. Besides, such aggregators can be easily embedded in the network as a trainable layer. Concretely, the attention-based pooling layer works as follows:

$$\mathbf{G}_i = \sum_{j=1}^{K_i} a_j \mathbf{F}_i^{(j)}, \quad a_j = \frac{\exp\{\mathbf{w}^\top \tanh(\mathbf{V}\mathbf{F}_i^{(j)})\}}{\sum_{m=1}^{K_i} \exp\{\mathbf{w}^\top \tanh(\mathbf{V}\mathbf{F}_i^{(m)})\}}, \tag{1}$$

where the calculation of attention weight a_j involves trainable parameters $\mathbf{w} \in \mathbb{R}^{64 \times 1}$ and $\mathbf{V} \in \mathbb{R}^{64 \times 512}$. Such weight quantifies the degrees of "activation" toward the final prediction for each MRI slice. Note that the weight calculation resembles the softmax operation, and can deal with the varying number of slices for different patients unless the variation is not significant. The aggregated patient-wise features $\mathbf{G}_i$ are then mapped to an output vector $\hat{Y}_i$ with an FC layer. Binary Cross Entropy (BCE) is used as the classification loss function,

$$\mathcal{L}_e = -\frac{1}{N_b} \sum_{i=1}^{N_b} \left(Y_i \log(\hat{Y}_i) + (1 - Y_i) \log(1 - \hat{Y}_i) \right). \tag{2}$$

2.3   Contrastive Learning for Discriminative Deep Features

For the studied problem, the MRI slices of a positive patient (with PAS) that reflect the clues about PAS can account for only a small portion of all slices, while most ones look normal. Meanwhile, some slices of a negative patient (without PAS) can manifest suspected patterns. Therefore, the extracted patient-wise deep features should be more discriminative under the existence of inherent noises. In addition to aggregating the slice-wise features with attention, contrastive learning is further considered to realize that [2]. Specifically, the representation $\mathbf{G}_i$ needs to keep certain distances between positive and negative patients while getting relatively closer for patients in the same class. Since the proposed network does not require self-supervised learning, a plug-and-play contrastive loss is employed to avoid additional capacity.

Concretely, to make up each batch, a positive patient and a negative patient are first sampled from the dataset. Then another patient is randomly picked up. It is thus guaranteed that a matching pair and a non-matching pair exist, respectively. Assume the third patient is positive, the patient-wise features are denoted as $\mathbf{G}_{i_a}, \mathbf{G}_{i_n}, \mathbf{G}_{i_p}$. The triplet loss [1] is calculated between these feature maps ($\mathbf{G}_{i_a}$ is the anchor, $\mathbf{G}_{i_n}$ is the non-matching sample and $\mathbf{G}_{i_p}$ is the matching sample). The triplet loss is used as the contrastive loss:

$$\mathcal{L}_c = \max\left\{ \|\mathbf{G}_{i_a} - \mathbf{G}_{i_p}\|_2 - \|\mathbf{G}_{i_a} - \mathbf{G}_{i_n}\|_2 + c, 0 \right\} \tag{3}$$

where c is the positive margin constant that quantifies the distance. It is added to the BCE losses of the samples in a batch: $\mathcal{L}_{\text{total}} = \mathcal{L}_e + \mathcal{L}_c$.

3   Experiment

3.1   Dataset and Implementation Details

The experiments are conducted on a real clinical dataset collected from a large obstetrics and gynecology hospital. The dataset involves 359 distinct patient subjects with a total of 15,186 MRI slices. A 5-fold cross-validation is performed and the average metrics values with standard deviations over the folds are reported. To pre-train the U-Net in the mask branch, 185 MRI slices are annotated with placenta by the experts. The ResNet18 is pre-trained on ImageNet. HACL-Net is trained end to end with Adam optimizer and batch size $N_b = 3$ for 100 epochs. The learning rate of the slice-level attention module and final FC layer is set as 10^{-6} and 10^{-4} respectively, while that of the rest parts is set as 10^{-5}. PyTorch 1.10.1, and an NVIDIA GeForce RTX 2080 Ti GPU with CUDA 11.3 are used to train our method. More details can be found in Section B of supplementary. The code is publicly available on https://github.com/LouieBHLu/HACL-Net.

3.2   Performance Comparison of PAS Diagnosis

Section C of supplementary material illustrates the effect of different selections of trunk branch. Table 1 shows the performance of the pre-trained baseline U-Net

Table 1. Performance of pre-trained ROI networks. PA is pixel accuracy.

Method	Train			Test		
	DICE	IOU	PA	DICE	IOU	PA
U-Net	80.6 ± 0.5	89.2 ± 0.3	78.3 ± 0.9	77.1 ± 0.9	87.0 ± 0.5	72.4 ± 1.2
MaNet	82.2 ± 0.6	90.2 ± 0.3	79.8 ± 0.7	80.4 ± 1.0	89.1 ± 0.6	76.2 ± 1.0
FPN	85.4 ± 0.5	92.1 ± 0.4	81.7 ± 0.7	80.7 ± 0.9	89.3 ± 0.4	72.8 ± 0.9
LinkNet	79.5 ± 0.3	88.5 ± 0.2	77.0 ± 0.7	76.8 ± 0.6	86.9 ± 0.4	72.8 ± 1.1

Table 2. Performance comparison of HACL-Net and existing methods for PAS diagnosis. The results in bold denote the best performance.

Method	Accuracy	AUC	Sensitivity	Specificity
Radio+RF	70.1 ± 5.4	61.8 ± 4.3	15.4 ± 4.2	93.9 ± 4.2
Radio+NB	69.9 ± 4.0	48.1 ± 3.9	0.0 ± 0.0	$\mathbf{100.0 \pm 0.0}$
Radio+KNN	68.2 ± 5.7	60.2 ± 3.4	24.9 ± 8.3	87.0 ± 3.2
Radio+MLP	70.4 ± 6.4	61.7 ± 6.3	26.7 ± 8.8	89.4 ± 4.5
Xuan [20]	65.7 ± 1.2	51.4 ± 1.0	1.7 ± 2.4	91.3 ± 2.5
Ye [23]	55.7 ± 6.5	46.7 ± 1.4	35.0 ± 8.2	64.0 ± 12.3
HACL-Net	$\mathbf{84.2 \pm 0.9}$	$\mathbf{84.9 \pm 0.6}$	$\mathbf{79.2 \pm 5.9}$	86.9 ± 2.8

compared with other three common networks. It turns out such network choice does not have a significant impact. The placenta area can be accurately captured by U-Net, which justifies the usage of Partial U-Net in the mask branch. Visual ROI prediction can be found in Section D of supplementary material.

To evaluate the performance of machine learning methods with Pyradiomics features and deep learning methods [12,20,23], the slice-wise features are aggregated with simple mean to derive the patient-wise features. The accuracy, area under the ROC curve (AUC), sensitivity, and specificity are employed as the metrics. As shown in Table 2, HACL-Net achieves the best 84.9% AUC and 79.2% sensitivity with 23.1% and 44.2% improvement respectively over the second-best approach. The outcomes are reflected in Fig. 4, where the compared methods suffer from very low sensitivity and high specificity. Since only a small portion of MRI slices have clues about a patient's being positive, traditional aggregation of slices tends to dilute the effective information. As the features are extracted from a segmentation network in [23], the sensitivity is relatively improved but with highly degraded specificity. In general deep features even perform worse than ROI-based radiomics features, partially due to the increased complexity of input is not properly explained by the clinical outcome of PAS.

3.3   Ablation Study

Table 3 and Fig. 5 illustrate the effect of each module in HACL-Net. For experiments without patient-level attention, slice-wise features are also aggregated
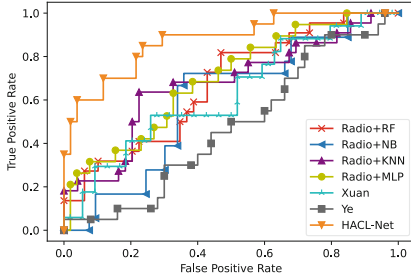
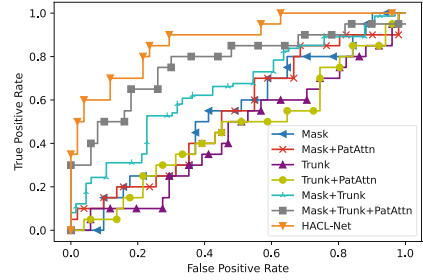**Fig. 4.** ROC of HACL-Net and existing methods for PAS diagnosis.

Fig. 5. ROC of ablation study of each component of HACL-Net. PatAttn is patient-level attention module.

Table 3. Results of ablation study for each module in the proposed HACL-Net.

Mask Branch	Trunk Branch	Patient Attention	CL	Accuracy	AUC	Sensitivity	Specificity
✓				60.4 ± 13.5	50.1 ± 1.3	26.0 ± 32.2	74.1 ± 31.7
✓		✓		71.6 ± 0.2	50.0 ± 1.5	7.5 ± 2.5	94.1 ± 2.0
	✓			71.6 ± 1.4	51.5 ± 1.5	5.0 ± 3.5	98.0 ± 2.4
	✓	✓		70.7 ± 2.7	53.9 ± 5.2	12.5 ± 10.9	95.2 ± 4.4
✓	✓			77.7 ± 0.6	66.9 ± 1.5	35.0 ± 4.1	94.7 ± 1.0
✓	✓	✓		80.7 ± 4.1	79.2 ± 5.2	62.9 ± 1.9	$\mathbf{98.7} \pm 1.3$
✓	✓	✓	✓	$\mathbf{84.2} \pm 0.9$	$\mathbf{84.9} \pm 0.6$	79.2 ± 5.9	86.9 ± 2.8
Sigmoid w/ 0.50 threshold				66.0 ± 9.6	60.9 ± 8.2	38.7 ± 4.1	84.8 ± 12.8
Sigmoid w/ 0.45 threshold				68.3 ± 9.8	68.1 ± 8.4	69.2 ± 6.6	67.0 ± 13.3
Sigmoid w/ 0.40 threshold				78.0 ± 8.0	82.1 ± 4.4	77.0 ± 10.1	81.3 ± 14.0
Sigmoid w/ 0.35 threshold				77.8 ± 8.4	81.6 ± 8.2	75.6 ± 0.9	78.5 ± 15.2
Sigmoid w/ 0.30 threshold				70.6 ± 3.1	66.3 ± 7.0	$\mathbf{87.8} \pm 8.7$	49.7 ± 13.8

with simple mean to derive patient-level features. The combination of mask branch and trunk branch achieves 15.4% AUC improvement over a single trunk branch, indicating the gain of additional context information. A 12.3% AUC improvement is witnessed when patient-level attention is used as an aggregator instead of mean. Contrastive learning manifests 5.7% AUC and 16.3% sensitivity improvement with 11.8% specificity drop as a trade-off. It is believed that sensitivity can be more important from a clinical perspective, and the trade-off can be tuned according to specific demands. Moreover, the softmax activation is replaced with sigmoid, and different confidence thresholds are used. As the threshold gets lower, a non-monotonic trend of sens-spec trade-off is observed, while the optimal balance lies around 0.4. That reveals the potential effect of class imbalance, *i.e.,* fewer positive cases. To this end, contrastive learning can help separate the decision boundary between two classes. Sigmoid results differ from softmax due to the explicitly set decision threshold. The predicted scores between positive and negative patients can be close for hard samples, leading to large deviations in sigmoid results. Figure 6 shows feature maps of a positive
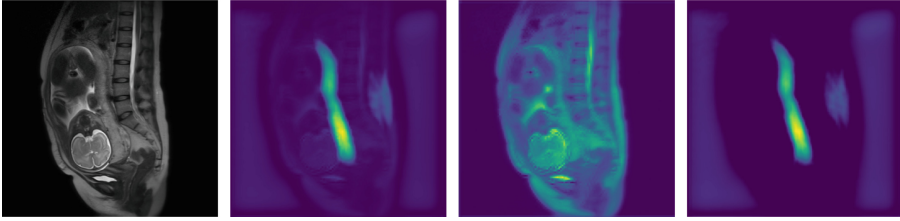
Fig. 6. Feature maps generated by HACL-Net of a positive patient. Figures from left to right are the MRI slice, the spatial residual attention feature map, the trunk branch feature map, and the mask branch feature map.

patient generated by HACL-Net. Spatial residual attention features capture both local focal area information without omitting global semantic context features compared to features of mask branch and trunk branch.

4   Conclusion

This paper presents a novel end-to-end HACL-Net for MRI-based computer-aided PAS diagnosis that facilitates more efficient prenatal screening. The proposed network utilizes slice-level spatial residual attention to extract context-aware deep semantic features, and aggregate them as a comprehensive representation of a patient with patient-level attention-based pooling. Moreover, a contrastive loss is added to improve the discriminating power of learned patient-wise features. Extensive experiments on a real clinical dataset show that HACL-Net achieves superior performance, with great potential for clinical applications.

References

1. Balntas, V., Riba, E., Ponsa, D., Mikolajczyk, K.: Learning local feature descriptors with triplets and shallow convolutional neural networks. In: Proceedings of the British Machine Vision Conference (BMVC), pp. 119.1-119.11 (2016)
2. Chopra, S., Hadsell, R., LeCun, Y.: Learning a similarity metric discriminatively, with application to face verification. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2005), vol. 1, pp. 539–546 (2005)
3. Cummins, C., Petoumenos, P., Wang, Z., Leather, H.: End-to-end deep learning of optimization heuristics. In: 2017 26th International Conference on Parallel Architectures and Compilation Techniques (PACT), pp. 219–232. IEEE (2017)
4. Han, M., et al.: Automatic segmentation of human placenta images with u-net. IEEE Access **7**, 180083–180092 (2019)
5. Ilse, M., Tomczak, J., Welling, M.: Attention-based deep multiple instance learning. In: International Conference on Machine Learning, pp. 2127–2136. PMLR (2018)

6. Jiao, M., Liu, H., Liu, J., Ouyang, H., Wang, X., Jiang, L., Yuan, H., Qian, Y.: Mal: Multi-modal attention learning for tumor diagnosis based on bipartite graph and multiple branches. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., Li, S. (eds.) MICCAI 2022. LNCS, vol. 13433, pp. 175–185. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-16437-8_17

7. Kohli, M., Prevedello, L.M., Filice, R.W., Geis, J.R.: Implementing machine learning in radiology practice and research. Am. J. Roentgenol. **208**(4), 754–760 (2017)

8. Li, H., Chen, L., Han, H., Kevin Zhou, S.: Satr: slice attention with transformer for universal lesion detection. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., Li, S. (eds.) MICCAI 2022. LNCS, pp. 163–174. Springer Nature Switzerland, Cham (2022)

9. Liu, X., et al.: What we know about placenta accreta spectrum (PAS). Eur. J. Obstet. Gynecol. Reprod. Biol. **259**, 81–89 (2021)

10. Oyelese, Y., Smulian, J.C.: Placenta previa, placenta accreta, and vasa previa. Obstet. Gynecol. **107**(4), 927–941 (2006)

11. Ren, H., et al.: Prediction of placenta accreta spectrum using texture analysis on coronal and sagittal T2-weighted imaging. Abdom. Radiol. **46**, 5344–5352 (2021)

12. Romeo, V., et al.: Machine learning analysis of MRI-derived texture features to predict placenta accreta spectrum in patients with placenta previa. Magn. Reson. Imaging **64**, 71–76 (2019)

13. Romeo, V., et al.: Prediction of placenta accreta spectrum in patients with placenta previa using clinical risk factors, ultrasound and magnetic resonance imaging findings. Radiol. Med. (Torino) **126**(9), 1216–1225 (2021). https://doi.org/10.1007/s11547-021-01348-6

14. Ronneberger, O., Fischer, P., Brox, T.: U-net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28

15. Silver, R.M., Lyell, D.J.: Placenta accreta spectrum. Protocols for High-Risk Pregnancies: an evidence-based approach, pp. 571–580 (2020)

16. Tian, Y., et al.: Contrastive transformer-based multiple instance learning for weakly supervised polyp frame detection. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., Li, S. (eds.) MICCAI 2022. LNCS, pp. 88–98. Springer Nature Switzerland, Cham (2022). https://doi.org/10.1007/978-3-031-16437-8_9

17. Van Griethuysen, J.J., et al.: Computational radiomics system to decode the radiographic phenotype. Can. Res. **77**(21), e104–e107 (2017)

18. Wang, F., et al.: Residual attention network for image classification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3156–3164 (2017)

19. Wang, S., et al.: RMDL: recalibrated multi-instance deep learning for whole slide gastric image classification. Med. Image Anal. **58**, 101549 (2019)

20. Xuan, R., Li, T., Wang, Y., Xu, J., Jin, W.: Prenatal prediction and typing of placental invasion using MRI deep and radiomic features. Biomed. Eng. Online **20**(1), 56 (2021)

21. Yao, J., Zhu, X., Huang, J.: Deep multi-instance learning for survival prediction from whole slide images. In: Shen, D., et al. (eds.) MICCAI 2019. LNCS, vol. 11764, pp. 496–504. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32239-7_55

22. Yao, J., Zhu, X., Jonnagaddala, J., Hawkins, N., Huang, J.: Whole slide images based cancer survival prediction using attention guided deep multiple instance learning networks. Med. Image Anal. **65**, 101789 (2020)

23. Ye, Z., Xuan, R., Ouyang, M., Wang, Y., Xu, J., Jin, W.: Prediction of placenta accreta spectrum by combining deep learning and radiomics using t2wi: a multi-center study. Abdominal Radiol. **47**(12), 4205–4218 (2022)
24. Zhang, D., Zou, L., Zhou, X., He, F.: Integrating feature selection and feature extraction methods with deep learning to predict clinical outcome of breast cancer. IEEE Access **6**, 28936–28944 (2018)
25. Zhang, Y., et al.: Spatiotemporal attention for early prediction of hepatocellular carcinoma based on longitudinal ultrasound images. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., Li, S. (eds.) MICCAI 2022. LNCS, vol. 13433, pp. 534–543. Springer, Cham (2022)