Self-Knowledge Distillation for First Trimester Ultrasound Saliency Prediction

Mourad Gridach[1], Elizaveta Savochkina[1], Lior Drukker[2,3], Aris T. Papageorghiou[2], and J. Alison Noble[1]

[1] Institute of Biomedical Engineering, University of Oxford, Oxford, UK
mourad.gridach@eng.ox.ac.uk
[2] Nuffield Department of Women's Reproductive Health, University of Oxford, Oxford, UK
[3] Rabin Medical Center, Sackler Faculty of Medicine, Tel-Aviv University, Israel

Abstract. Self-knowledge distillation (SKD) is a recent and promising machine learning approach where a shallow student network is trained to distill its own knowledge. By contrast, in traditional knowledge distillation a student model distills its knowledge from a large teacher network model, which involves vast computational complexity and a large storage size. Consequently, SKD is a useful approach to model medical imaging problems with scarce data. We propose an original SKD framework to predict where a sonographer should look next using a multi-modal ultrasound and gaze dataset. We design a novel Wide Feature Distillation module, which is applied to intermediate feature maps in the form of transformations. The module applies a more refined feature map filtering which is important when predicting gaze for the fetal anatomy variable in size. Our architecture design includes ReSL loss that enables a student network to learn useful information whilst discarding the rest. The proposed network is validated on a large multi-modal ultrasound dataset, which is acquired during routine first trimester fetal ultrasound scanning. Experimental results show the novel SKD approach outperforms alternative state-of-the-art architectures on all saliency metrics.

Keywords: Self-knowledge distillation · Saliency map prediction · Fetal ultrasound · Wide Feature Distillation · ReSL loss

1 Introduction

Ultrasound screening [1, 2] in the first trimester of pregnancy is important for early-stage abnormality detection and health assessment of mother and baby. This requires skill to perform and is not available worldwide. High skill required to scan contributes to this, and automated guidance would contribute to changing current practice and democratising access to ultrasound. Our interest is automatic prediction of saliency maps to assist an operator in guidance to imaging planes, to help a non-expert with plane finding and abnormality detection.

Knowledge distillation (KD) [3] consists of transferring knowledge from a cumbersome network (teacher) to a shallow network (student). In medical image analysis, there is a growing interest in KD [4–11]. It has its limitation and is

only suitable when using a large well-validated model. First, the exploitation of all knowledge from the teacher network by the student can be difficult. This can result in the student performance being lower than the teacher's. Second, pre-training a teacher model requires heavy computational resources. SKD bypasses the training of a large teacher network, only training a network with the same architecture as the student [12, 25, 26]. In computer vision, KD is widely used with applications in classification, object detection, instance and semantic segmentation [12–14]. KD for medical imaging has been reported in [4–11]. Qin et al. [4] proposed a model where a teacher transfers knowledge to a student in the form of semantic region information. Li et al. [7] developed a framework using mutual KD applied to cross-modality medical image segmentation. However, to the best of our knowledge, it is the first time SKD approach is used in medical image analysis and first trimester fetal ultrasound.

Motivated by the importance of the ultrasound screening in early stages of pregnancy and little research produced in this clinical domain, we consider SKD approach which solves the lack of large well-validated teacher models. Assistive technologies which help guide a user during ultrasound scanning can making fetal ultrasonography more accessible in medically underserved regions. We focus on the task of predicting where a sonographer should look next to guide to the anatomical structures. Sonographer's visual attention can be quantified via the distribution of gaze points, hereafter referred to as a *saliency map*. We propose a novel SKD approach which extracts and transfers useful information to allow a student network to learn from itself with no help from a large teacher model. First, the SKD architecture filters information from a student network and transfers it to a second network (identical to the structure of a student model), a *Peer-student* network. The process is called *peer-learning*. Second, we propose a Wide Feature Distillation (WFD) module as a transformation function applied to the intermediate student feature maps. The module plays an important role allowing the student model to effectively distill its own knowledge. Furthermore, we propose a ReLU Squared Loss (ReSL) to filter important information while blocking the rest of the information unnecessary for the learning process. The architecture has several advantages over the basic KD and SKD approaches. Our model has fewer connecting paths between the intermediate feature map levels, this minimizes information redundancy caused in the previous methods [4, 12]. The inclusion of the WFD module enables the model to learn the location of key anatomical structures (palate, nasal bone, rump and nuchal translucency) which are examined by a sonographer during a first trimester scan. The method helps locate small, medium and large structures. The main contributions are:

- To the best of our knowledge, we make the first attempt at using SKD approach for a real-time first trimester ultrasound saliency map prediction.
- We propose a WFD module with ReLU squared loss to enable the student to distill its own knowledge, transfer the useful information from the peer-student to the student while blocking the unnecessary information.
- Our network is computationally inexpensive which is important for both real-time saliency prediction and research replication.

2 Method

2.1 Problem Formulation

Assume a list of feature maps $F_{l_i}^{in}$ with different spatial resolutions, where F represents a feature map at level l_i. Our goal is to find a transformation f that can effectively filter information and output a list of new features: $F^{out} = f(F^{in})$. The transformation function f varies from a basic convolution operation (Figure 1a) to complex transformations (Figure 1b, 1c). Fig. 1b shows a top-down and bottom-up approach [12]. It takes $1-4$ levels of input features, $F^{in} = (F_1^{in}, F_2^{in}, F_3^{in}, F_4^{in})$, where F_k^{in} is a feature level with a pre-defined resolution of the input image. The final output is $F^{out} = (F_1^{out}, F_2^{out}, F_3^{out}, F_4^{out})$ after applying a series of transformations:

$$F^{out} = f(g(F^{in})) \tag{1}$$

$$g(x) = (BiFPN)^{(n)}(x) \tag{2}$$

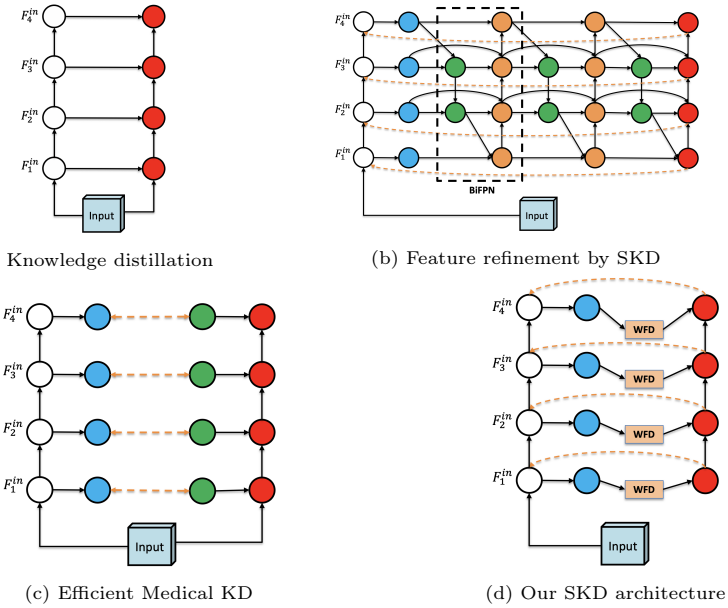
where n represents the number of times the Bi-directional Feature Pyramid Network (BiFPN) [18] module is applied to the input x, here $n = 3$. The same process is followed in Efficient Medical Knowledge Distillation (EMKD) (Fig. 1c) apart from a different transformation f used. Following Occam's razor principle, our goal is to design a novel transformation f that is simple and efficient in distilling information from intermediate feature maps at different levels.

2.2 Wide Feature Distillation Module

The KD and SKD approaches handle natural and medical images. Both approaches follow the student/teacher framework by applying feature transformation function to differentiate intermediate feature maps at different levels. We found that none of the approaches include modules which can distill the features efficiently [4, 6, 12]. Moreover, unlike the fetal ultrasound images, natural images are relatively easy to predict the saliency for. A significant size variation in the fetal anatomy makes it challenging to acquire an accurate prediction.

To this end, we propose a WFD module (Figure 2) to efficiently train the student network to distill its own knowledge without the help of the teacher network. Given an input feature map, we distill features using parallel branches leveraging diverse convolution filter sizes. In the SKD context, the WFD architecture creates a more refined feature map filtering mechanism that maximizes feature distillation. A complex mechanism enhances gaze prediction of fetal anatomy variable in size, enabling the student network to learn from itself on a more profound level.

Within the WFD module, convolution layers (sub-regions) are stacked in parallel where each layer starts by a 3×3 convolution followed by one or more dilated convolutions with different dilation rates (r = 1, 3, 5) where the two are added in a skip connection manner. The sub-regions compose the final feature map. By relying on dilated convolution with different rates, the input feature map is divided into multiple sub-regions from finer to coarser levels and combines



(a) Knowledge distillation

(b) Feature refinement by SKD

(c) Efficient Medical KD

(d) Our SKD architecture

Fig. 1: Comparison between different KD and SKD architectures. The relevant learning approaches include the student model (white), teacher model and peer-student model (red), blue circles were obtained by applying a convolution to the feature maps, feature transformation (green), forward path (solid black) and distillation path (orange).

local features within each. The sub-regions obtained from different parallel layers are added to the original feature map to form the final global distilled feature map. The knowledge from the whole, half, or small sections of an ultrasound image are extracted by the parallel layers, to fuse and construct the final gaze saliency prediction.

2.3 ReLU Squared Loss

The distillation loss plays an important role in a knowledge distillation framework [3, 14, 15]. We consider that useful information is represented by positive values, while negative values represent the unnecessary information for the student network. The ReLU function is widely used for this purpose and eliminates the negative values by setting them to zero. That way, the unnecessary information will not go back into the model, while positive values (useful information) will pass through. We design a distillation loss to benefit from the power of ReLU with the mean squared loss (SL) applied to the intermediate feature maps of the student ($ReLU(F^{in})$) and those of the peer-student ($ReLU(F^{out})$). The representation of ReSL loss:

$$\mathcal{L}_{ReSL} = d(\mathbf{ReLU}(\mathbf{F^{in}}), \mathbf{ReLU}(\mathbf{F^{out}})) \tag{3}$$

where d is a L_2 distance. ReSL allows the student model to mimic feature maps distilled from the peer-student.

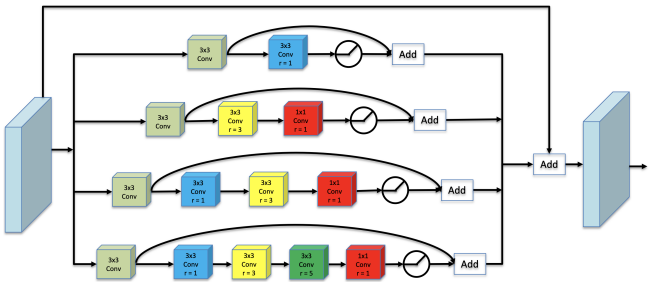


Fig. 2: Wide Feature Distillation Module. It consists of four parallel branches: first, we compute a 3×3 convolution, then followed by convolutions with different dilation rates and ReLU non-linearity to distill features at different scales (skip connections is used).

2.4 Proposed Self-Distillation Network

The KD and SKD related approaches [4, 12, 18, 19] suffer from a number of issues. First, for an efficient multi-scale feature fusion, these approaches use lateral layers to extract information causing the appearance of redundant information. Second, the approaches allow the student network to learn irrelevant information causing the model performance to drop. Furthermore, all the previous approaches were developed to classify and segment either natural, or MRI and CT images. However, no ultrasound work has been done using fetal ultrasound data.

To overcome these limitations, we design a novel SKD architecture for saliency map prediction using first trimester fetal ultrasound images. Figure 3 shows the proposed model architecture (best to view in color). Given an input ultrasound image I, ResNet generates a set of feature maps $F_{in} = (F_{in}^{(1)}, F_{in}^{(2)}, F_{in}^{(3)}, F_{in}^{(4)})$. Next, we compute the intermediate feature maps (green) at different levels using: $P^{(l)} = Conv(F_{in}^{(l)})$, where $P^{(l)}$ represents the intermediate feature map at a level l. To generate the distilled feature map $S^{(l)}$, every feature map $P^{(l)}$ is fed to the WFD module:

$$S^{(l)} = Conv(w_1^{(l)}.WFD(P^{(l)}) + w_2^{(l)}.MaxPool(P^{(l-1)})) \tag{4}$$

To increase model's efficiency, we utilize a depth-wise convolution [17] and max-pooling for down-sampling. We apply a fast normalized fusion [18] with parameters, $w_1^{(l)}$ and $w_2^{(l)}$, which outperform softmax-based and unbounded fusions.

In the first level, Equation 4 is reduced to $S^{(1)} = Conv(WFD(P^{(1)}))$. Using the knowledge distillation frameworks' loss [12, 15, 16], we filter knowledge through the soft labels as follows:

$$\mathcal{L}_{SKD} = KL(p^s || p^{ps}) \tag{5}$$

where $KL(.)$ is the Kullback-Leibler divergence loss, p^s and p^{ps} represent the probabilities of a given pixel in the saliency prediction map extracted from the student and the peer-student networks, respectively. Finally, the total loss function is given as:

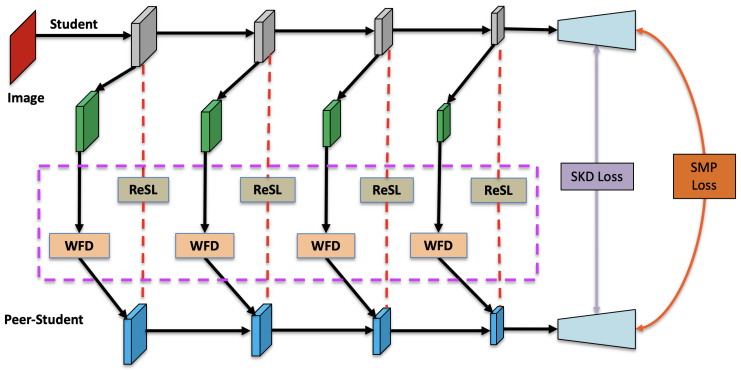


Fig. 3: Proposed Architecture. The network contains a student (grey) and a peer-student (blue) which distills its knowledge through multiple WFDs applied to different intermediate feature maps (green) and learn useful information using ReSL loss.

$$\mathcal{L}_{total} = \alpha_1 \mathcal{L}^{s}_{smp} + \alpha_2 \mathcal{L}^{ps}_{smp} + \beta \mathcal{L}_{SKD} + \gamma \mathcal{L}_{ReSL} \tag{6}$$

where $\mathcal{L}^{s}_{smp}$ and $\mathcal{L}^{ps}_{smp}$ are two cross-entropy losses. The hyper-parameters α_1 and α_2 are set to 1, β and γ are set to 0.9 based on the validation set results.

3 Experimental Results

Data and Experimental Setup. Routine clinical fetal ultrasound exams with real time sonographer gaze tracking data is used in this work [4]. The exams were performed on a GE Voluson E8 scanner (GE, USA) while the video signal of the machine's monitor is recorded at 30 Hz. Gaze is simultaneously recorded at 90 Hz with a Tobii Eye Tracker 4C (Tobii, Sweden). We inspected 150 fetal US videos, manually deleted corrupt and double-image frames which left us with 45,630 US frames and their corresponding saliency maps. Saliency maps are computed by smoothing the binary gaze maps with a Gaussian kernel (st.d 13.5, and window 81 chosen experimentally). The data are spilt into 70 videos (29,250 frames) for training, 17 videos (7,290 frames) for validation and 29 videos (9,126 frames) for testing. We implement our models using PyTorch1.6. During training, we use SGD with an initial learning rate of 10^{-3} and weight decay of 5.10^{-4}. The models were trained for 200 epochs with a batch size of 16, and the image size is 288 by 224. Random rotation and flipping were used as part of data augmentation. In all the experiments, DeepLab [22] with ResNet18 was used as the base model. Models are trained on a 12 GB TitanX GPU. We use Kullback-Leibler divergence (KL), Normalised Scanpath Saliency (NSS), Correlation Coefficient (CC) and Similarity (SIM) as the evaluation metrics [24].

Quantitative Results. To demonstrate the effectiveness of the proposed framework, we compare it to four state-of-the-art models: (1) U-Net architecture [23],

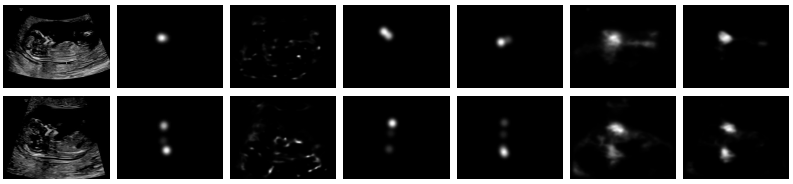
[4] This study was approved by the UK Research Ethics Committee (Reference 18/WS/0051) and the ERC ethics committee

Table 1: Performance of the our network against the state-of-the-art models. Best performance is marked in **bold**.

Model	KLD ↓	SIM ↑	NSS ↑	CC ↑
U-Net [23]	3.18 (0.10)	0.14 (0.09)	2.86 (0.08)	0.16 (0.09)
RA [24]	2.16 (0.07)	0.27 (0.08)	4.19 (0.07)	0.39 (0.07)
Mix. RA [24]	2.28 (0.08)	0.26 (0.08)	4.34 (0.06)	0.38 (0.06)
FRSKD [12]	2.08 (0.08)	0.29 (0.09)	4.37 (0.06)	0.42 (0.06)
Ours	**1.72 (0.06)**	**0.38 (0.05)**	**4.63 (0.04)**	**0.53 (0.05)**

Table 2: Ablation study of the proposed approach. The best performing model is marked in **bold**.

Model	KLD ↓	SIM ↑	NSS ↑	CC ↑
SKD	2.11 (0.10)	0.31 (0.09)	4.15 (0.10)	0.41 (0.09)
+ReSL	2.07 (0.07)	0.32 (0.08)	4.21 (0.08)	0.43 (0.09)
+WFD	1.73 (0.07)	0.36 (0.08)	4.59 (0.09)	0.52 (0.08)
Ours	**1.72 (0.06)**	**0.38 (0.05)**	**4.63 (0.04)**	**0.53 (0.05)**



(a) US (b) GT (c) UNet (d) RA (e) Mix.RA (f) FRSKD (g) Ours

Fig. 4: Qualitative performance. The columns show the ultrasound (US) input frames, the ground truth (GT) saliency annotations, saliency predictions of UNet, RA, Mix. RA and FRSKD against our proposed method, respectively.

(2) Random Augmentation (RA) and (3) Mixed Random Augmentation (Mix. RA) methods [24] (both used for ultrasound saliency prediction), and (4) FRSKD (Feature Refinement by SKD) [12], a recent approach applied to classification and semantic segmentation. The results are shown in Table 1. Our SKD network significantly outperforms U-Net architecture. Moreover, the proposed model outperforms the RA and Mix. RA methods [24] specifically developed to predict saliency maps in ultrasound. Our method outperforms FRSKD [12] which is a recent SKD approach.

Qualitative Results. To measure the effectiveness of gaze prediction, we qualitatively assess the image results. Figure 4 shows exemplary test results of the comparative state-of-the-art models against our proposed method. The rows show the raw ultrasound data, the ground truth (GT) saliency annotations, UNet, RA, Mix. RA, FRSKD and our model, respectively. In the bottom row, RA assigns high saliency values to fetus head whereas ground truth fixations are mainly at the bottom. Mix. RA under-estimates the gaze of the sonographer looking at the head. Our method correctly predicts both locations with slight misalignment at the bottom and assigns equal probability densities to all anatomical structures. The predictions are less spread out compared to the FRSKD method.

Computational Complexity. Our SKD model aims to make a network shallow while maximizing its performance. Consequently, we calculate the computational complexity in terms of number of parameters and FLOPs (number of operations required to run a single instance) in each model. However, due to no literature found on SKD method used for saliency prediction, medical segmentation ap-

proaches mentioned in Section 1 were used for the comparison instead. Table 3 compares our model against the previous state-of-the-art methods.

Table 3: Computational complexity.

Model	Backbone	#Param (M)	FLOPs (G)
UNet++	ResNet	20.6	312.14
UNet	ResNet	34.5	220.19
PSPNet	ResNet	46.7	293.83
MobileNetv2	ResNet	2.11	19.14
FRSKD	ResNet	12.11	24.81
Ours	ResNet	6.24	4.38

Ablation Study. We performed an ablation study with results reported in Table 2. The original SKD baseline model lacks WFD module and ReSL loss. For comparison, we replaced the WFD module with a 3×3 conv with dilation rate of 3, followed by 1×1 conv with dilation rate of 1 and ReLU non-linearity. A simple L2 loss replaced the ReSL. The addition of each component, including WFD module and ReSL loss, each contributes a considerable performance increase over the baseline SKD on all saliency metrics. The WFD module has the highest positive effect on performance metrics. Finally, a combination of the two components in a model shows the best performance and adds the most value to the fetal ultrasound saliency prediction.

4 Discussion and Conclusion

To the best of our knowledge, this is the first SKD framework for ultrasound saliency prediction as well as the first reported application of SKD in medical image analysis. Our model performs best on all saliency metrics. In comparison to four state-of-the-art models, our SKD framework led to a decrease in KLD score of 0.36. The KLD metric is highly penalized if any GT fixation locations are missed. The NSS metric is sensitive to false positives which is seen in Fig. 3 showing little to no false saliency prediction. SIM and CC metrics improved by 0.09 and 0.11 compared to the second best FRSKD model, respectively. CC penalizes false negatives and SIM penalizes predictions that fail to account for all the GT density. Finally, our network is compared against the previous state-of-the-art methods in its computational complexity. Our SKD method comes second in parameter complexity after the non-saliency method MobileNetv2 and first in terms of FLOPs.

In conclusion, we have presented a novel SKD architecture tailored to predict visual saliency for first trimester ultrasound images. Our SKD framework differs from the previous methods [4, 12, 18, 19] due to its architecture design and structure which combines modules in a novel way leading to feature distillation and filter false saliency predictions. The model's additional improvements are vital for guidance to fetal anatomical structures. Our architecture is computationally inexpensive which is of the essence for the replication purposes of other researchers. The experiments demonstrate the potential of the proposed method to track changes in sonographer gaze and automatically guide to important structures during real-time first trimester US scanning.

References

1. Vakanski, A., Xian, M. and Freer, P.E., 2020. Attention-enriched deep learning model for breast tumor segmentation in ultrasound images. Ultrasound in Medicine Biology, 46(10), pp.2819-2833.
2. Ning, Z., Zhong, S., Feng, Q., Chen, W. and Zhang, Y., 2021. SMU-Net: Saliency-Guided Morphology-Aware U-Net for Breast Lesion Segmentation in Ultrasound Image. IEEE Transactions on Medical Imaging, 41(2), pp.476-490.
3. Hinton, G., Vinyals, O. and Dean, J., 2015. Distilling the knowledge in a neural network. arXiv preprint arXiv:1503.02531, 2(7).
4. Qin, D., Bu, J.J., Liu, Z., Shen, X., Zhou, S., Gu, J.J., Wang, Z.H., Wu, L. and Dai, H.F., 2021. Efficient medical image segmentation based on knowledge distillation. IEEE Transactions on Medical Imaging, 40(12), pp.3820-3831.
5. Dou, Q., Liu, Q., Heng, P.A. and Glocker, B., 2020. Unpaired multi-modal segmentation via knowledge distillation. IEEE transactions on medical imaging, 39(7), pp.2415-2425.
6. Wang, H., Zhang, D., Song, Y., Liu, S., Wang, Y., Feng, D., Peng, H. and Cai, W., 2019, April. Segmenting neuronal structure in 3D optical microscope images via knowledge distillation with teacher-student network. In 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019) (pp. 228-231). IEEE.
7. Li, K., Yu, L., Wang, S. and Heng, P.A., 2020, April. Towards cross-modality medical image segmentation with online mutual knowledge distillation. In Proceedings of the AAAI conference on artificial intelligence (Vol. 34, No. 01, pp. 775-783).
8. Zhou, Y., Chen, H., Lin, H. and Heng, P.A., 2020, October. Deep semi-supervised knowledge distillation for overlapping cervical cell instance segmentation. In International Conference on Medical Image Computing and Computer-Assisted Intervention (pp. 521-531). Springer, Cham.
9. Hu, M., Maillard, M., Zhang, Y., Ciceri, T., La Barbera, G., Bloch, I. and Gori, P., 2020, October. Knowledge distillation from multi-modal to mono-modal segmentation networks. In International Conference on Medical Image Computing and Computer-Assisted Intervention (pp. 772-781). Springer, Cham.
10. Ju, L., Wang, X., Wang, L., Liu, T., Zhao, X., Drummond, T., Mahapatra, D. and Ge, Z., 2021, September. Relational subsets knowledge distillation for long-tailed retinal diseases recognition. In International Conference on Medical Image Computing and Computer-Assisted Intervention (pp. 3-12). Springer, Cham.
11. Xing, X., Hou, Y., Li, H., Yuan, Y., Li, H. and Meng, M.Q.H., 2021, September. Categorical Relation-Preserving Contrastive Knowledge Distillation for Medical Image Classification. In International Conference on Medical Image Computing and Computer-Assisted Intervention (pp. 163-173). Springer, Cham.
12. Ji, M., Shin, S., Hwang, S., Park, G. and Moon, I.C., 2021. Refine myself by teaching myself: Feature refinement via self-knowledge distillation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 10664-10673).
13. Zhang, L., Song, J., Gao, A., Chen, J., Bao, C. and Ma, K., 2019. Be your own teacher: Improve the performance of convolutional neural networks via self distillation. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 3713-3722).
14. Heo, B., Kim, J., Yun, S., Park, H., Kwak, N. and Choi, J.Y., 2019. A comprehensive overhaul of feature distillation. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 1921-1930).

15. He, T., Shen, C., Tian, Z., Gong, D., Sun, C. and Yan, Y., 2019. Knowledge adaptation for efficient semantic segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 578-587).
16. Liu, Y., Chen, K., Liu, C., Qin, Z., Luo, Z. and Wang, J., 2019. Structured knowledge distillation for semantic segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 2604-2613).
17. Howard, A.G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M. and Adam, H., 2017. Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861.
18. Tan, M., Pang, R. and Le, Q.V., 2020. Efficientdet: Scalable and efficient object detection. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 10781-10790).
19. Zhou, Y., Chen, H., Lin, H. and Heng, P.A., 2020, October. Deep semi-supervised knowledge distillation for overlapping cervical cell instance segmentation. In International Conference on Medical Image Computing and Computer-Assisted Intervention (pp. 521-531). Springer, Cham.
20. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V. and Rabinovich, A., 2015. Going deeper with convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1-9).
21. He, K., Zhang, X., Ren, S. and Sun, J., 2016. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).
22. Chen, L.C., Zhu, Y., Papandreou, G., Schroff, F. and Adam, H., 2018. Encoder-decoder with atrous separable convolution for semantic image segmentation. In Proceedings of the European conference on computer vision (ECCV) (pp. 801-818).
23. Ronneberger, O., Fischer, P. and Brox, T., 2015, October. U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention (pp. 234-241). Springer, Cham.
24. Savochkina, E., Lee, L.H., Drukker, L., Papageorghiou, A.T. and Noble, J.A., 2021, July. First Trimester Gaze Pattern Estimation Using Stochastic Augmentation Policy Search for Single Frame Saliency Prediction. In Annual Conference on Medical Image Understanding and Analysis (pp. 361-374). Springer, Cham.
25. Xu, T.B. and Liu, C.L., 2019, July. Data-distortion guided self-distillation for deep neural networks. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 33, No. 01, pp. 5565-5572).
26. Yun, S., Park, J., Lee, K. and Shin, J., 2020. Regularizing class-wise predictions via self-knowledge distillation. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 13876-13885).