ESORDA: Enhance Small Object Representation and Domain Adaption for Unsupervised Mitosis Detection

This is a paper pre-demo. Once our paper ongoing the review procedures and receive positive response, we will upload to arxiv for manuscript display and make code publicly available. Hope you can understand :-)

1. Contributions

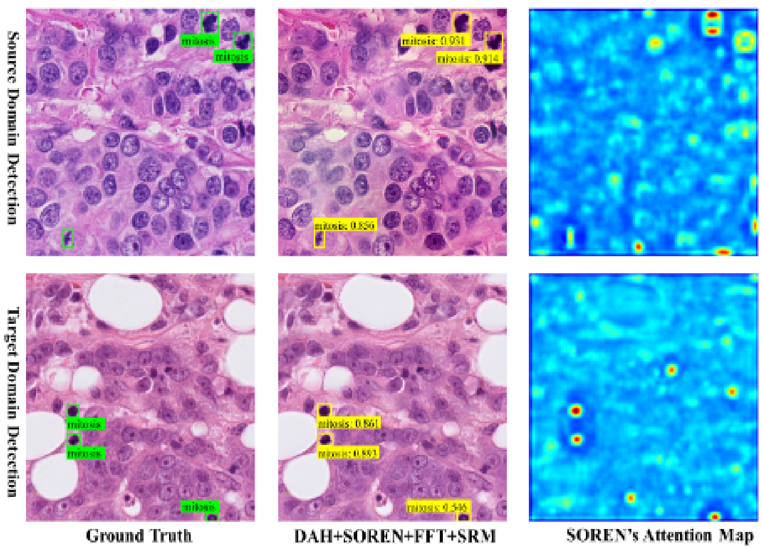

Fig. 1. There is a obvious domain gap between the source and target domain in the ground truth. With our network, the source pathology image becomes target-like and much more clear by implementing the FFT and SRM in image pre-processing. A domain adaption method DAH is used to extract the similar mitosis features in source and target domain. Via SOREN to augment the small object representation, our network pay more attention to the small mitosis cells.

In specific the contributions of this paper are summarized as follow:

(1) We propose a novel neck network named Small Object Representation Enhancement Network (SOREN) to specialize on enhancing the representation of small mitosis cells in feature-level. Via the SOREN, the augmented feature maps help our detector emphasis on the positive small samples and yield state-of-the-art results. To our best known, we are the first proposing a frame that specialize on small mitosis cells detection.

(2) We develop a domain style transfer method Fast Fourier Transform (FFT) to reduce the domain gap in image-level. The FFT maps the style of source domain into that of target domain in an efficient way which almost costs no additional computational resource.

(3) We create a Domain Adaption Head (DAH) to learn the domain-invariant representation in feature-level, where the feature extractor benefits from the global alignment and center-aware alignment to obtain
address the domain-shift problem.

(4) We propose a Super-Resolution Module (SRM) to improve the resolution and quality of images to facilitate the small object detection in image-level, which provides a feasible way to achieve better results.

TABLE I

OUR CONTRIBUTIONS AND THEIR RELATIONSHIPS.

Operation / Issue	Image-level	Feature-level
Small Object	SRM	SOREN
Domain Adaption	FFT	DAH

We propose four novel components. SRM and SOREN are specialized for the issue of small object detection. DAH and FFT based style transfer component focus on the issue of domain adaption. SRM and FFT work in image-level, while SOREN and DAH enhance representation ability in feature-level.

2. Method

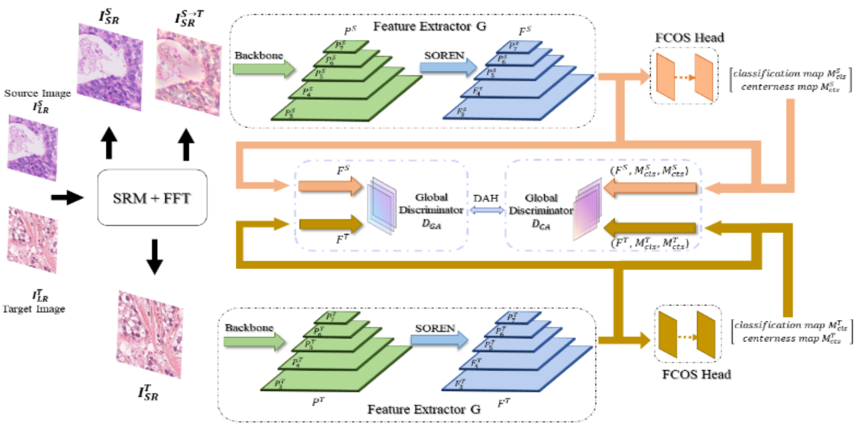

Fig. 2. Pipeline of our proposed model: Enhanced Small Object Representation and Domain Adaption (ESORDA) network. ESORDA has five components: (1) Super-Resolution Module (SRM) for improving the low-resolution images; (2) FFT for domain style transferring to reduce the domain gap in image-level; (3) feature extractor G with Feature Pyramid Network (FPN) and Small Object Representation Enhanced Network (SOREN) for obtaining feature maps in two domains, where the SOREN was proposed to enhance the representation of small objecT; (4) FCOS head for object detection and producing prediction maps to feed into Domain Adaption Head (DAH); (5) the DAH is used for mitosis-related feature alignment such that the domain-shift problem can be reduced in feature-level. The I^S_{LR} means images in low-resolution from source domain, and the subscript "SR" indicates super-resolution. Similar definition about target domain. $I^{S\rightarrow T}_{SR}$ denotes the source images mapped the texture style to target domain.

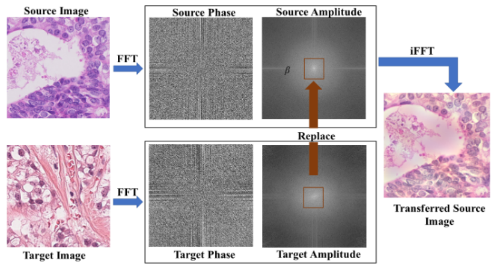

Fig. 3. Illustration of domain style transfer via FFT. We randomly sample the images from source and target domain. The low-frequency component of the target amplitude is used to replace the source amplitude, such that the source image can maintain the image content but map its style to target domain.

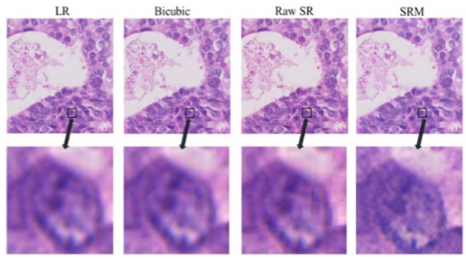

Fig. 4. Visulization of different methods for improving image resolution. The LR denotes the original image in low-resolution. The results of bicubic interpolation (Bicubic) and ESRGAN without fine-tune (Raw SR) improve the resolution accompanied with grid masks, while our SRM addresses it in detail. Zoom in for best view.

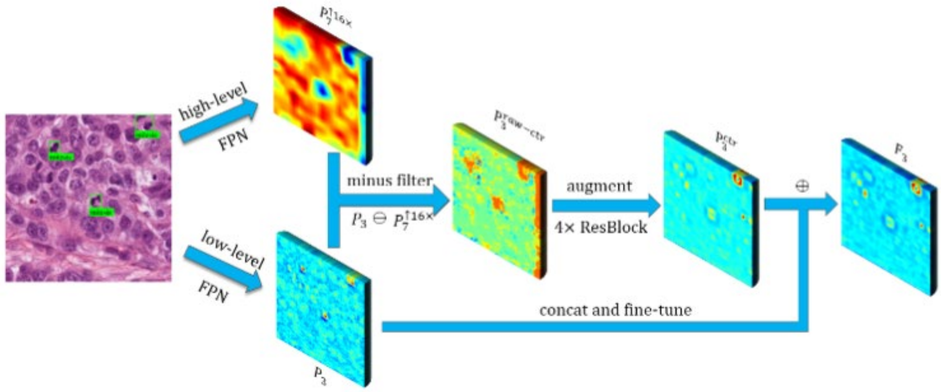

Fig. 5. Demonstration of our proposed Small Object Representation Enhancement Network(SOREN).Given feature maps produced by FPN, the high-level features P_7 firstly are upsampled as $P_7^{\uparrow\ 16\times}$ to match the shape of low-level features (e.g., P_3). The attention map of $P_7^{\uparrow\ 16\times}$ mainly focus on the irrelevant larger things (e.g., texture, larger normal cells) where the color becomes deeper. This condition still exist in P_3 where despite the representation notes the correct small mitosis cells, there are noises indicating the larger things semantics disturb P_3. A minus operation is implemented for filtering the irrelevant semantics such that the small mitosis cells are centralized and emphasized. We also deploy a plain-net with 4 residual blocks to successively augment it and obtain P_3^{ctr}. After concatenate the vanilla feature P_3 with P_3^{ctr}, the enhanced representation specialized for small object F_3 is formed. The SOREN outputs the same features compared to FPN, but augments the low-level features P_3 and P_4 to produce F_3 and F_4 respectively.

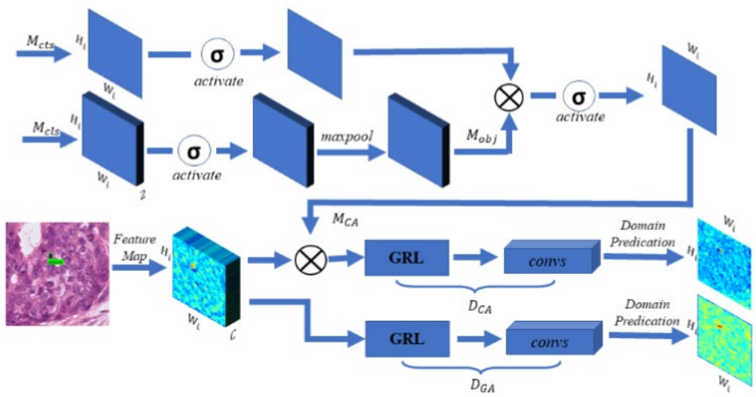

Fig. 6. The workflow of Domain Adaption Head (DAH). There are two blocks for domain adaption trained with adversarial manner. The global alignment discriminator D_{GA} backwards the global features of source and target domain, while the center-aware alignment classifier D_{CA} delivers the foreground information. Together these two components reduce the domain gap in feature-level. A gradient inversion layer (GRL) is also deployed to ensure the extracted features become similar in distribution.

3. Results

We test our model on the MIDOG 2021 dataset for mitosis detection. Our method achieves *state-of-the-art* performance with the powerful and novel component Small Object Representation Enhancement Network (SOREN).

TABLE III
COMPARISON ON MIDOG 2021 DATASET

Team	*Precision*	*Recall*	F_1
AI Medical	**0.643**	**0.780***	**0.728**
IAMLAB	0.642	0.757	0.721
No.0	0.562	0.718	0.698
Jdex	0.575	0.751	0.682
TIA Centre	0.625	0.759	0.677
XidianU-OUC	0.528	0.696	0.655
SK	0.563	0.730	0.644
CGV	0.547	0.728	0.643
Tribvn	0.608	0.768	0.631
PixelPath-AI	0.542	0.721	0.610
Ours(DAH)	0.542	0.746	0.628
Ours(DAH+SOREN)	0.713	0.711	0.712
Ours(DAH+SOREN+FFT)	0.649	**0.778**	0.708
Ours(Full Components)	**0.756***	0.706	**0.730***

Comparison on MIDOG 2021 dataset. The existing methods adapt full labeled images (S_1, S_2, S_3) to train and test on unlabeled space (S_4), while our methods are trained only with S_1 and testing in unsupervised way.

TABLE V
PERFORMANCE COMPARISON EVALUATED IN SUPERVISED WAY

Team	*Precision*	*Recall*	F_1
Tribvn	0.731	**0.875**	**0.848**
TIA Centre	0.692	0.857	0.837
CGV	**0.750**	0.867	0.829
No.0	0.652	0.837	0.826
XidianU-OUC	0.703	0.863	0.800
AI Medical	0.729	0.830	0.793
jdex	0.737	0.820	0.782
Leeds	0.624	0.795	0.774
ML	0.586	0.821	0.738
Ours(DAH)	0.618	0.869	0.723
Ours(DAH+SOREN)	0.794	0.932	0.858
Ours(DAH+SOREN+FFT)	0.760	0.955	0.850
Ours(Full components)	**0.844**	**0.969**	**0.902**

Comparison on MIDOG 2021 dataset in supervised. The existing methods adapt full labeled images (S_1, S_2, S_3) to train and test on unlabeled space (S_1), while our methods are trained only with S_1 and test on S_1.

TABLE VI
TESTING RESULTS OF PROPOSED METHODS IN SUPERVISED AND UNSUPERVISED WAY

Components	Supervised	F_1	AP	AP_S
DAH	✓	0.723	0.729	0.597
DAH+SOREN	✓	0.858	0.741	0.681
DAH+SOREN+FFT	✓	0.850	0.786	0.777
Full Components	✓	**0.902**	**0.841**	**0.813**
DAH	×	0.628	0.383	0.331
DAH+SOREN	×	0.712	0.417	0.359
DAH+SOREN+FFT	×	0.708	0.450	0.414
Full Components	×	**0.730**	**0.481**	**0.449**

Testing results of proposed methods in supervised and unsupervised way. "Supervised" means we test on the source domain that contains labels during training, or we test on target domain that is involved in training without labels (i.e., in unsupervised way). AP_S means the area of objects is lower than 32×32 in COCO style.

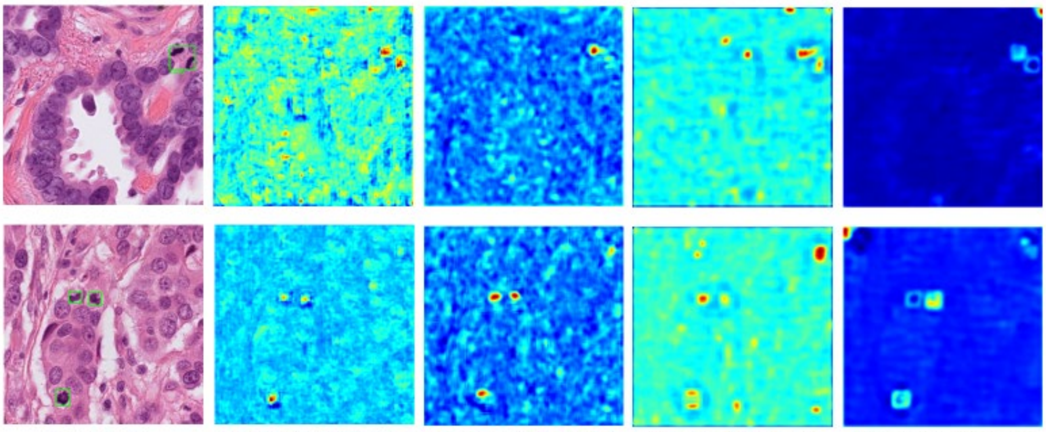

Fig. 7. Comparison of the representation ability. When we do not deploy SOREN, the backbone (ResNet+FPN) produces low-quality of features where the attention maps contain background noises and do not notice the the potential mitosis proposals, and even in the CA domain adaption this condition relieves litter. When we implement SOREN, it can obviously observe that there are much more prospectively small mitosis candidates in SOREN's attention map. Besides, the CA adaption of SOREN becomes more precisely. The comparison indicates our proposed SOREN has powerful ability to enhance the small object representation.