2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) | 979-8-3503-2447-1/23/$31.00 ©2023 IEEE | DOI: 10.1109/EMBC40787.2023.10340280

SSPT-bpMRI: A Self-supervised Pre-training Scheme for Improving Prostate Cancer Detection and Diagnosis in Bi-parametric MRI*

Yuan Yuan, Euijoon Ahn, Dagan Feng, Fellow, IEEE, Mohamad Khadra and Jinman Kim, Member, IEEE

***Abstract*— Prostate cancer (PCa) is one of the most prevalent cancers in men. Early diagnosis plays a pivotal role in reducing the mortality rate from clinically significant PCa (csPCa). In recent years, bi-parametric magnetic resonance imaging (bpMRI) has attracted great attention for the detection and diagnosis of csPCa. bpMRI is able to overcome some limitations of multi-parametric MRI (mpMRI) such as the use of contrast agents, the time-consuming for imaging and the costs, and achieve detection performance comparable to mpMRI. However, inter-reader agreements are currently low for prostate MRI. Advancements in artificial intelligence (AI) have propelled the development of deep learning (DL)-based computer-aided detection and diagnosis system (CAD). However, most of the existing DL models developed for csPCa identification are restricted by the scale of data and the scarcity in labels. In this paper, we propose a self-supervised pre-training scheme named SSPT-bpMRI with an image restoration pretext task integrating four different image transformations to improve the performance of DL algorithms. Specially, we explored the potential value of the self-supervised pre-training in fully supervised and weakly supervised situations. Experiments on the publicly available PI-CAI dataset demonstrate that our model outperforms the fully supervised or weakly supervised model alone.**

I. INTRODUCTION

Prostate cancer (PCa) is the second most common cancer in men. The latest statistics of cancers states that more than 370,000 people worldwide died from PCa in 2020 alone [1]. Early diagnosis and treatment of clinically significant PCa (csPCa) are crucial in improving the cure rate of PCa [2]. Prostate-specific antigen (PSA) testing and digital rectal examination (DRE) used in routine screening of PCa may under-diagnose of csPCa and over-diagnose of benign tissues or indolent lesions, which can miss malignant lesions or lead to unnecessary biopsies [3], [4]. Since 2010, the application of multi-parametric magnetic resonance imaging (mpMRI) in detecting csPCa and guided targeted biopsy has become a rapidly developing clinical research topic in the field of urological oncology [5]. In recent years, bi-parametric MRI (bpMRI) has gained popularity owing to its applicability in high-volume and population-based screening with enhanced efficiency when compared to mpMRI. This technique offers reduced scan time, decreased utilization of contrast agent and lower costs [6]. However, detection and diagnosis of csPCa lesions in bpMRI is challenging due to the difficulty of distinguishing csPCa from non-progressive clinical symptoms such as benign prostatic hyperplasia, prostatitis and cysts in prostate MR images.

Recently, advancement in artificial intelligence (AI) has spurred the use of automated computer-aided detection and diagnosis system (CADs) to help radiologists detect csPCa lesions from mpMRI/bpMRI. There has been development of customized deep neural network (DNN) models for csPCa detection and diagnosis [7], [8]. DNN models rely heavily on large-scale, high-quality labeled datasets that are difficult to obtain due to data silos, privacy protection, tedious labeling, etc. Prostatectomy specimen confirmed labels are more accurate but less common. Most existing models trained on small datasets in a fully supervised manner have problems such as poor generalization ability and model over-fitting [7], [9]. Recently, Joeran et al. [10] proposed a semi-supervised model to automatically generate labels using image diagnostic reports and partially labeled image data. Their aim was to reduce labeling costs and improve labeling efficiency, paving the way for the use of larger datasets. However, the labels generated by the model were often inaccurate and the performance of the deep learning (DL) models was heavily affected by noisy labels. Alternatively, self-supervised learning (SSL), a form of unsupervised learning, where the data themselves can generate supervisory signals for the model training can be used. It has shown great success in many computer vision [11] and medical image segmentation tasks [12].

In this paper, we propose a new 3D self-supervised pre-training framework on the bpMRI of prostate (SSPT-bpMRI), which can be applied to various image analysis tasks, such as prostate segmentation and PCa grading. Our model makes up for the shortcomings of the existing SSL models that are pre-trained on single-modal or specific multi-modal images with difficulty to generalize to bpMRI. Furthermore, we design an image restoration pretext task for pre-training of the model which captures distinctive image characteristics of bpMRI, focusing on their appearance, content and texture. Subsequently, we fine-tuned the model on a small but fully labeled dataset and also on a larger dataset with noisy labels added, respectively. We analyzed the effect of self-supervised pre-training on fully supervised learning and weakly supervised learning. The experimental results on the PI-CAI dataset [13] demonstrate that our self-supervised pre-training improves

*This work was supported by Australian Research Council (ARC) grants.

Yuan Yuan, Dagan Feng and Jinman Kim are with School of Computer Science, Faculty of Engineering, The University of Sydney, Sydney, Australia (corresponding author: jinman.kim@sydney.edu.au)

Euijoon Ahn is with College of Science & Engineering, James Cook University, Cairns, Australia

Dagan Feng is also with the Med-X Research Institute, Shanghai Jiao Tong University, Shanghai, China

Mohamad Khadra is with Department of Urology, Nepean Hospital, Kingswood, Australia

This work is licensed under a Creative Commons Attribution 3.0 License.
For more information, see http://creativecommons.org/licenses/by/3.0/

the generalization performance of the model trained from scratch on a small dataset, and enhances the stability of the model in the presence of noisy labels.

II. Materials and Methods

A. Materials

The PI-CAI dataset contains images from three Dutch centers (Radboud University Medical Center (RUMC), Ziekenhuis Groep Twente (ZGT), University Medical Center Groningen (UMCG)) and one Norwegian center (Norwegian University of Science and Technology (NTNU)). Two sub-datasets of this dataset are used. We use the Public Training and Development Dataset for model training and validation and the Hidden Validation and Tuning Cohort for model testing. All patient exams are of men suspected of harboring csPCa (e.g., due to elevated levels of PSA, abnormal DRE findings). Patients are included only if they do not have a history of treatment or prior International Society of Urological Pathology (ISUP) grading ≥ 2 findings. For all the patients, bpMRI scans (including axial T2-weighted imaging (T2W), axial high b-value ($\geq$ 1000 s/mm^2) diffusion-weighted imaging (DWI), axial apparent diffusion coefficient maps (ADC)) acquired using Siemens Healthineers or Philips Medical Systems-based scanners with surface coils are offered.

There are 1500 cases in the Public Training and Development Dataset. Among them, 1075 cases have benign tissues or indolent PCa, and 425 cases have malignant lesions. The benign cases are labeled as zero. 220 malignant cases are manually labeled. Labels of the rest 205 malignant cases are AI-derived in a semi-supervised learning strategy proposed by Joeran et al. [10]. This sub-dataset contains MRI scans acquired using seven different scanners, from two vendors, at three centers (RUMC, ZGT, UMCG). Thus, the spatial resolutions of its images vary across different patient exams. For instance, in the case of the axial T2W scans, the most common voxel spacing (in mm/voxel) observed is 3.0×0.5×0.5 (43%), followed by 3.6×0.3×0.3 (25%), 3.0×0.342×0.342 (15%) and others (17%). There are 100 cases in the Hidden Validation and Tuning Cohort. This sub-dataset contains MRI scans acquired using nine different scanners, from two vendors, at three centers (RUMC, ZGT, UMCG). All the cases are manually labeled by experts. Details of the images are not publicly available.

B. Methods

The overview of our framework is shown in Fig. 1. It consists of self-supervised pre-training (SSPT-bpMRI) and fully supervised or weakly supervised fine-tuning. In the self-supervised pre-training stage, the 3D U-Net [14] is trained based on an image restoration task that unifies four customized image transformations. In the fine-tuning stage, we perform csPCa detection based on the nnU-Net framework [15], and the network is initialized by the pre-trained 3D U-Net with network topology given by nnU-Net framework.

1) Data Pre-processing: All the data were transferred to NifTI format and then re-sampled to the same dimensions and spatial resolution as their corresponding original T2W images. For SSPT-bpMRI, all the data are re-sampled to 3.0×0.5×0.5 (mm/voxel), and the central area mainly including prostate sized 256×256×20 are extracted as the region of interest (ROI). T2W and DWI images are normalized to [0, 1] by the min-max normalization, and the intensity values in each ADC map are clipped within the range of [0, 3000] and then normalized to [0, 1]. For fine-tuning on the downstream csPCa detection and diagnosis task, the data pre-processing is automatically completed by following the nnU-Net framework settings.

2) Self-supervised Pre-training: To pre-train the 3D U-Net architecture, four specially tailored data augmentation methods are used in a unified image restoration task.

- Non-linear transformation. By recovering the intensity values of the images that have undergone a set of monotonically nonlinear transformations, the model can learn the appearance of the anatomic structures present in the images.
- Local shuffling guides the model to learn rich local texture and boundary information of objects while keeping its global structure understandable.
- Inner-cutout guides the model to learn the local continuous of organs.
- Outer-cutout guides the model to learn the spatial layout and global geometry of organs.

More detailed descriptions of data augmentation methods are also available in [16].

The network is trained on the sub-volumes sized 64×64×16 which were randomly extracted from the images to scale up the amount of data and support 3D SSL training on limited computing resources. The model converges towards the object of decreasing the mean squared error (MSE) loss [17]. Given N samples, we select S sub-volumes from each sample to form the new dataset $O = \{s_1, s_2, s_i, \ldots, s_{S\times N}\}$, then the MSE loss can be expressed as

$$L_{mse} = \frac{1}{S \times N} \sum_{i=1}^{i=S\times N} (g(f(s_i)) - s_i)^2 \qquad (1)$$

where $f(\cdot)$ denotes the image transformation function and $g(\cdot)$ on behalf of the mapping of the network from the generated image to the original image. Well, there are three channels in s_i including T2W, ADC and DWI images.

3) Fine-tuning for csPCa Detection and Diagnosis: Based on the nnU-Net framework, all the layers in the pre-trained U-Net are fine-tuned. The last layer of the pre-trained network is not retained in order to adapt to the downstream detection task. The loss function is a combination of focal loss [18] and cross-entropy loss [19]. The output of the softmax activation generated by nnU-Net is transformed to the detection map where each lesion region on the map corresponds to a unique probability value through a candidate lesion extraction method [10]. The candidate lesions are only extracted from the centrally cropped region of physical size

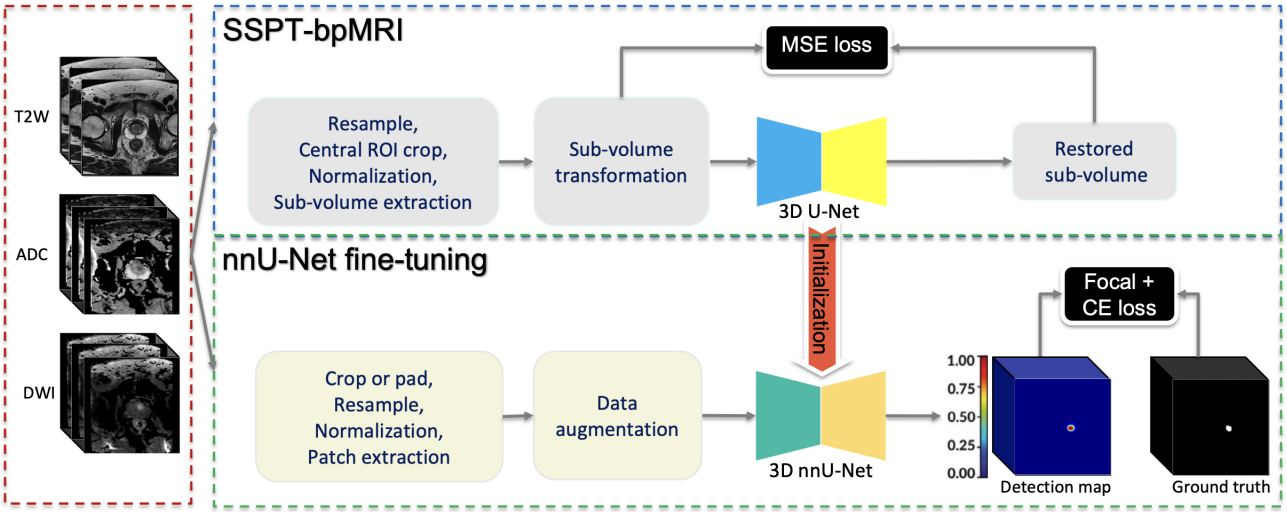

Fig. 1. The overview of our framework for csPCa detection and diagnosis.

$192 \times 192 \times 60\ mm^3$. The maximum probability value of the predicted lesions in each image is regarded as the result of the patient-level diagnosis.

4) Implementation Details: For the self-supervised pre-training, we extracted 16 3D sub-volumes from each case and the ratio of the training set to the test set is 4:1. The stochastic gradient descent (SGD) with a momentum of 0.9 is selected as the optimizer. The initial learning rate (0.1) is gradually reduced according to the step learning rate policy. The maximum of epoch is set to 1000 and the early stopping strategy is adopted with 20 epochs of patience. For the csPCa detection and diagnosis, the initial learning rate (0.01) is gradually reduced according to the "poly" learning rate policy and the maximum of epoch is set to 1000.

III. Results and Discussions

A. Experiments Setup

Our experiments include: 1) self-supervised pre-training, 2) fine-tuning on the supervised PI-CAI Public Training and Development Dataset with 1295 cases and testing the model on the PI-CAI Hidden Validation and Tuning Cohort; 3) fine-tuning on the weakly supervised PI-CAI Public Training and Development Dataset with 1500 cases and testing the model on the PI-CAI Hidden Validation and Tuning Cohort. To pre-train the network, 80% of all the 1500 cases are used for training and the rest 20% are used for validation. Five-fold cross-validation with the splits provided by the PI-CAI challenge organizers was used for model fine-tuning.

B. Evaluation

Lesion-level detection performance is evaluated using the Average Precision (AP) metric. Patient-level diagnosis performance is evaluated using the Area Under Receiver Operating Characteristic (AUROC) metric. Overall score is the average of both task-specific metrics:

$$Overall\ Score = (AP + AUROC)/2$$

C. Results and Discussions

We compare the cross-validation and test results of our model with the results of the baselines of the PI-CAI challenge including U-Net [14], nnDetection [20] and nnU-Net [15]. Performance of the models on the fully supervised dataset and the weakly supervised dataset was compared. Table I presents the cross-validation results and Table II shows the test results of the models trained on fully supervised dataset. In Table III cross-validation results of weakly supervised dataset are presented with testing results in Table IV. The models were evaluated based on the manually labeled data. For cross-validation results, the mean and standard deviation are calculated.

On the fully supervised dataset, our model achieved the best score with 0.6636 on cross-validation and 0.600 on test, best AP with 0.4647 on cross-validation and 0.479 on test. In AUROC, our model scored 0.8624 on cross-validation and 0.721 on test. This shows that self-supervised pre-training can improve the generalization performance of the fully supervised model. The results generated by nnU-Net trained from scratch was better than both the U-Net and the nnDetection (see Table I and Table II). On the weakly supervised dataset, our model obtained better results comparing to the nnU-Net trained from scratch, with AUROC of 0.8358 and AP of 0.4352 on cross-validation and AUROC of 0.823 and AP of 0.613 on test (see Table III and Table IV). This is attributed to self-supervised pre-training that has a certain anti-noise ability which helps in improving model's learning ability from noisy labels. On both cross-validation and test, nnDetection had the best AUROC (cross-validation: AUROC of 0.8405, test: AUROC of 0.885) but lower AP (cross-validation: AP of 0.4236, test: AP of 0.582). This may be because nnDetection pursues object-level detection rather than pixel-level segmentation. In addition, our results show that increasing the data size improves the generalization performance of the model, even in the presence of noisy labels.

TABLE I
CROSS-VALIDATION RESULTS ON THE SUPERVISED DATASET

Models/Metrics	Score	AUROC	AP
UNet	0.5981±0.0634	0.8162±0.0462	0.3800±0.0808
nnDetection	0.6253±0.0280	0.8450±0.0188	0.4057±0.0452
nnUNet	0.6609±0.0183	**0.8661**±0.0070	0.4556±0.0390
SSPT-bpMRI (ours)	**0.6636**±0.0330	0.8624±0.0163	**0.4647**±0.0570

TABLE II
TEST RESULTS OF MODELS TRAINED ON THE SUPERVISED DATASET

Models/Metrcis	Score	AUROC	AP
UNet	0.576	0.689	0.463
nnDetection	0.502	0.735	0.269
nnUNet	0.597	**0.737**	0.457
SSPT-bpMRI (ours)	**0.600**	0.721	**0.479**

TABLE III
CROSS-VALIDATION RESULTS ON THE WEAKLY SUPERVISED DATASET

Models/Metrics	Score	AUROC	AP
UNet	0.6012±0.0690	0.7910±0.0466	0.4115±0.0934
nnDetection	0.6321±0.0535	**0.8405**±0.0334	0.4236±0.0760
nnUNet	0.6334±0.0530	0.8348±0.0253	0.4320±0.0828
SSPT-bpMRI (ours)	**0.6355**±0.0541	0.8358±0.0281	**0.4352**±0.0851

TABLE IV
TEST RESULTS OF MODELS TRAINED ON THE WEAKLY SUPERVISED DATASET

Models/Metrcis	Score	AUROC	AP
UNet	0.731	0.829	**0.633**
nnDetection	**0.734**	**0.885**	0.582
nnUNet	0.714	0.818	0.610
SSPT-bpMRI (ours)	0.718	0.823	0.613

IV. CONCLUSIONS

In this paper, we propose a self-supervised pre-training scheme named SSPT-bpMRI with a unified image restoration task that incorporates multiple image transformation strategies to pre-train a model using large-scale bpMRI datasets. It learns general feature representations of image appearance, content, and texture of bpMRI. We conducted csPCa detection and diagnosis based on the nnU-Net framework, and the pre-trained model was fine-tuned and tested on fully supervised and weakly supervised datasets. Our experimental results demonstrate that our self-supervised pre-training can effectively improve the generalization performance and provide anti-noise learning ability of models trained on weakly supervised datasets.

REFERENCES

[1] H. Sung, J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, and F. Bray, "Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: a cancer journal for clinicians*, vol. 71, no. 3, pp. 209–249, 2021.

[2] P. R. Carroll, J. K. Parsons, G. Andriole, R. R. Bahnson, D. A. Barocas, E. P. Castle, W. J. Catalona, D. M. Dahl, J. W. Davis, J. I. Epstein, *et al.*, "Prostate cancer early detection, version 2.2015," *Journal of the National Comprehensive Cancer Network*, vol. 13, no. 12, pp. 1534–1561, 2015.

[3] D. C. Grossman, S. J. Curry, D. K. Owens, K. Bibbins-Domingo, A. B. Caughey, K. W. Davidson, C. A. Doubeni, M. Ebell, J. W. Epling, A. R. Kemper, *et al.*, "Screening for prostate cancer: Us preventive services task force recommendation statement," *Jama*, vol. 319, no. 18, pp. 1901–1913, 2018.

[4] L. Naji, H. Randhawa, Z. Sohani, B. Dennis, D. Lautenbach, O. Kavanagh, M. Bawor, L. Banfield, and J. Profetto, "Digital rectal examination for prostate cancer screening in primary care: a systematic review and meta-analysis," *The Annals of Family Medicine*, vol. 16, no. 2, pp. 149–154, 2018.

[5] R. T. Gupta, K. A. Mehta, B. Turkbey, and S. Verma, "Pi-rads: Past, present, and future," *Journal of Magnetic Resonance Imaging*, vol. 52, no. 1, pp. 33–53, 2020.

[6] L. Boesen, N. Nørgaard, V. Løgager, I. Balslev, R. Bisbjerg, K.-C. Thestrup, M. D. Winther, H. Jakobsen, and H. S. Thomsen, "Assessment of the diagnostic accuracy of biparametric magnetic resonance imaging for prostate cancer in biopsy-naive men: the biparametric mri for detection of prostate cancer (bidoc) study," *JAMA network open*, vol. 1, no. 2, pp. e180219–e180219, 2018.

[7] Y. Song, Y.-D. Zhang, X. Yan, H. Liu, M. Zhou, B. Hu, and G. Yang, "Computer-aided diagnosis of prostate cancer using a deep convolutional neural network from multiparametric mri," *Journal of Magnetic Resonance Imaging*, vol. 48, no. 6, pp. 1570–1577, 2018.

[8] A. Saha, M. Hosseinzadeh, and H. Huisman, "End-to-end prostate cancer detection in bpmri via 3d cnns: effects of attention mechanisms, clinical priori and decoupled false positive reduction," *Medical image analysis*, vol. 73, p. 102155, 2021.

[9] Y. K. Tsehay, N. S. Lay, H. R. Roth, X. Wang, J. T. Kwak, B. I. Turkbey, P. A. Pinto, B. J. Wood, and R. M. Summers, "Convolutional neural network based deep-learning architecture for prostate cancer detection on multiparametric magnetic resonance images," in *Medical imaging 2017: Computer-aided diagnosis*, vol. 10134, pp. 20–30, SPIE, 2017.

[10] J. S. Bosma, A. Saha, M. Hosseinzadeh, I. Slootweg, M. de Rooij, and H. Huisman, "Annotation-efficient cancer detection with report-guided lesion annotation for deep learning-based prostate cancer detection in bpmri," *arXiv e-prints*, pp. arXiv–2112, 2021.

[11] X. Liu, F. Zhang, Z. Hou, L. Mian, Z. Wang, J. Zhang, and J. Tang, "Self-supervised learning: Generative or contrastive," *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 1, pp. 857–876, 2021.

[12] A. Taleb, W. Loetzsch, N. Danz, J. Severin, T. Gaertner, B. Bergner, and C. Lippert, "3d self-supervised methods for medical imaging," *Advances in neural information processing systems*, vol. 33, pp. 18158–18172, 2020.

[13] A. Saha, J. J. Twilt, J. S. Bosma, B. van Ginneken, D. Yakar, M. Elschot, J. Veltman, J. Fütterer, M. de Rooij, and H. Huisman, "Artificial Intelligence and Radiologists at Prostate Cancer Detection in MRI: The PI-CAI Challenge (Study Protocol)," May 2022.

[14] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3d u-net: learning dense volumetric segmentation from sparse annotation," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II 19*, pp. 424–432, Springer, 2016.

[15] F. Isensee, P. F. Jaeger, S. A. Kohl, J. Petersen, and K. H. Maier-Hein, "nnu-net: a self-configuring method for deep learning-based biomedical image segmentation," *Nature methods*, vol. 18, no. 2, pp. 203–211, 2021.

[16] Z. Zhou, V. Sodha, J. Pang, M. B. Gotway, and J. Liang, "Models genesis," *Medical image analysis*, vol. 67, p. 101840, 2021.

[17] P. Christoffersen and K. Jacobs, "The importance of the loss function in option valuation," *Journal of Financial Economics*, vol. 72, no. 2, pp. 291–318, 2004.

[18] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proceedings of the IEEE international conference on computer vision*, pp. 2980–2988, 2017.

[19] Z. Zhang and M. Sabuncu, "Generalized cross entropy loss for training deep neural networks with noisy labels," *Advances in neural information processing systems*, vol. 31, 2018.

[20] M. Baumgartner, P. F. Jäger, F. Isensee, and K. H. Maier-Hein, "nndetection: a self-configuring method for medical object detection," in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part V 24*, pp. 530–539, Springer, 2021.