Computers in Biology and Medicine 150 (2022) 106124

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

Knowledge distillation driven instance segmentation for grading prostate cancer

Taimur Hassan [a,b,*], Muhammad Shafay [a], Bilal Hassan [a,c], Muhammad Usman Akram [b], Ayman ElBaz [d], Naoufel Werghi [a]

[a] KUCARS and C2PS, Department of Electrical Engineering and Computer Science, Khalifa University, Abu Dhabi, 127788, United Arab Emirates
[b] Department of Computer and Software Engineering, National University of Sciences and Technology, Islamabad, 44000, Pakistan
[c] School of Automation Science and Electrical Engineering, Beihang University (BUAA), Beijing, 100191, China
[d] Department of Bioengineering, University of Louisville, Louisville, KY 40292, USA

ARTICLE INFO

Keywords:
Histopathology
Deep learning
Prostate cancer
Prostate tissues
Incremental learning
Instance segmentation

ABSTRACT

Prostate cancer (PCa) is one of the deadliest cancers in men, and identifying cancerous tissue patterns at an early stage can assist clinicians in timely treating the PCa spread. Many researchers have developed deep learning systems for mass-screening PCa. These systems, however, are commonly trained with well-annotated datasets in order to produce accurate results. Obtaining such data for training is often time and resource-demanding in clinical settings and can result in compromised screening performance. To address these limitations, we present a novel knowledge distillation-based instance segmentation scheme that allows conventional semantic segmentation models to perform instance-aware segmentation to extract stroma, benign, and the cancerous prostate tissues from the whole slide images (WSI) with incremental few-shot training. The extracted tissues are then used to compute majority and minority Gleason scores, which, afterward, are used in grading the PCa as per the clinical standards. The proposed scheme has been thoroughly tested on two datasets, containing around 10,516 and 11,000 WSI scans, respectively. Across both datasets, the proposed scheme outperforms state-of-the-art methods by 2.01% and 4.45%, respectively, in terms of the mean IoU score for identifying prostate tissues, and 10.73% and 11.42% in terms of F1 score for grading PCa according to the clinical standards. Furthermore, the applicability of the proposed scheme is tested under a blind experiment with a panel of expert pathologists, where it achieved a statistically significant Pearson correlation of 0.9192 and 0.8984 with the clinicians' grading.

1. Introduction

Prostate cancer (PCa) is the second deadliest form of cancer in men [1,2], causing a high mortality rate if not treated timely. PCa can be accurately identified through biopsy tests [3,4], where its severity can be objectively graded by measuring the Gleason scores from the majority, and minority distribution of the Gleason patterns as per the International Society of Urological Pathologists (ISUP) standards [5,6]. The ISUP severity grades were developed in 2014 and range between 1 to 5 [7]. As evident from Table 1, ISUP-1 represents the individual, discrete well-formed glands, whereas ISUP-5 represents glands that lack the gland formation or have necrosis with or without poorly formed, fused, and cribriform patterns [7,8]. The structural deformities within the prostate tissues (due to PCa) can be analyzed through Gleason patterns (GPs), which typically range from 3 to 5 (i.e., GP-3, GP-4, GP-5), showcasing unique appearances within the whole slide images (WSI) [9]. Moreover, semantic segmentation, in the context of WSI segmentation, involves the extraction of foreground and background information, where the foreground information is related to the tissue regions, and the background information is associated with the non-tissue areas. In addition to this, the semantic segmentation networks can also solve the multi-class segmentation problems, e.g., extracting multiple GPs along with the healthy stroma or benign tissues from the candidate WSI patches [9,10]. However, the semantic segmentation models produce more false positives than the instance segmentation models in the context of WSI segmentation [11]. This elevated rate of false positives (with the semantic segmentation models) is because of the high similarity within the textural and semantic features of the WSIs containing the non-cancerous and cancerous regions [11,12] (see

* Corresponding author at: KUCARS and C2PS, Department of Electrical Engineering and Computer Science, Khalifa University, Abu Dhabi, 127788, United Arab Emirates.
E-mail address: taimur.hassan@ku.ac.ae (T. Hassan).

https://doi.org/10.1016/j.compbiomed.2022.106124
Received 18 January 2022; Received in revised form 29 August 2022; Accepted 17 September 2022
Available online 28 September 2022

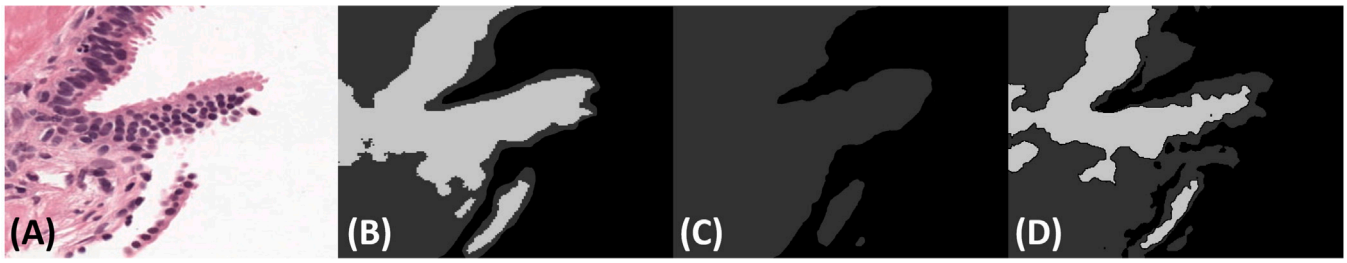

Fig. 1. (A): A WSI patch that depicts prostate tissues, (B): ground truth of the WSI patch in (A), (C): prostate tissues obtained through semantic segmentation, (D): prostate tissues segmented through incremental instance segmentation.

Table 1
Definition of PCa grading as per ISUP-2014 standards [7].

Gleason score	Gleason pattern	ISUP grade	Description
≤6	≤6	1	Contains only individual, discrete well-formed glands.
7	3 + 4	2	Contains predominantly well-formed glands with lesser component of poorly formed, fused, and/or cribriform glands.
7	4 + 3	3	Contains predominantly poorly formed, fused, and/or cribriform glands with lesser component of well-formed glands.
8	4 + 4, 3 + 5, 5 + 3	4	1. Contains only poorly formed, fused, and/or cribriform glands, 2. Contains predominantly well-formed glands with lesser component of lacking gland formation, 3. Contains predominantly lacking gland formation and lesser component of well-formed glands. It should be noted here that poorly formed, fused, and/or cribriform glands can be a minor component here.
9 or 10	4 + 5, 5 + 4, or 5 + 5	5	Contains gland that lack formation, or have necrosis, with or without poorly formed, fused, and/or cribriform glands.

Fig. 1-A). Furthermore, learning together all the pixel-level categories (via semantic segmentation) requires large-scale and well-annotated datasets, which is quite costly to procure in the real world [13,14], especially for the medical imaging problems [15,16]. One approach to overcome these issues is to develop an incremental instance segmentation approach that, with few-shot learning, can incrementally train the underlying segmentation model to recognize different instances (categories) of the tissue regions, especially the GPs.

The recent advances in instance segmentation models have led many researchers to detect highly correlated and cluttered object instances [17,18] within the conventional photographic [19] and X-ray [20] imagery. Similarly, researchers have also adopted these models in histopathological image analysis to identify and extract cellular communities [11], and tissue nucleuses [21] to grade breast [22], thyroid [23], and colorectal cancer [24]. However, such instance segmentation models require extensive training on the large-scale data [25–27] which makes them an infeasible choice for the spontaneous nucleus classification and tissue analysis tasks within the clinical settings. Furthermore, thoroughly annotating the histology images (having up to 40x magnification level) to recognize different tissue communities is an impractical and tedious job for the clinicians and often requires a team of experts to ensure validity. Here, developing a sophisticated instance segmentation approach that can provide competitive performance towards recognizing different morphological structures of cellular tissues (after being trained with a few training examples) is an enticing and worthwhile alternative.

1.1. Related work

Over the past, many researchers have worked on screening PCa from histopathology [9] and mp-MRI [28] imagery. The initial methods were based on classical machine learning schemes [29–31] for extracting the cancerous tissues. However, the recent frameworks employ deep learning to accurately segment cancerous tissues to screen PCa [32]. In this context, Wang et al. [33] provided a noticeable comparison of the deep learning systems with classical methods. However, the current trend in grading PCa (especially clinically significant PCa [5]) is moving towards the identification of prostate tissues from the WSI imagery [34]. Towards this end, Arvaniti et al. [10] extracted the class activation maps from the pre-trained MobileNet [35] to recognize Gleason PCa tissues. Hassan et al. [9] recently developed a dilated residual and hierarchically fashioned segmentation network, dubbed DRN, that is trained via a custom loss function (L_h) to extract the Gleason patterns from the whole slide images. Otalora et al. [36] developed a teacher–student learning scheme to classify prostate histopathology scans in which the teacher network (dubbed ResNext50-32 × 4d [37]), having 22M parameters, is trained in a semi-supervised or semi-weakly-supervised manner using strongly annotated and weakly annotated labels, respectively. Moreover, the student network (dubbed DenseNet-121 [38]), having 7M parameters, is trained with four custom strategies, namely, variant-I, II, III, and a fully supervised baseline. In variant-I, the student network is tuned with the pseudo-labeled data only. In variant-II, the student network is pre-trained first on the pseudo-labeled data, and then it is refined on the strongly labeled data. In variant III, the student network is trained on the combined pseudo-labeled and strongly-annotated data, and in the fully supervised baseline variant, the student network is trained only on the strongly-annotated data only [36]. Apart from this, Bulten et al. [39] developed a two-staged approach, where in the first stage, they employed U-Net [40] to segment healthy and tumorous glands. Afterward, they formulated a rule-based scheme to classify the biopsy as malignant if at least 10% of the epithelial tissue is identified as cancerous [39]. Strom et al. [41] developed a two-staged ensemble of deep classification models, where in the first stage, they classified the WSI patches into benign or malignant, and in the second stage, they recognized each malignant patch into GP-3,

GP-4, and GP-5 to grade PCa as per the ISUP standards. Apart from this, the recent works in the hispathology also leveraged few-shot learning paradigm to overcome the data limitation [42,43].

Although many frameworks have been proposed in the past that can extract the cancerous tissues from the WSIs to grade PCa [5,44]. But, to the best of our knowledge, these frameworks cannot be scaled in clinical settings for PCa grading because either their performance is low [41], or they require an extensive amount of well-annotated data for training [10]. Also, as mentioned above, the extraction of prostate tissues (which are essential for PCa grading) is an instance segmentation problem that the semantic segmentation networks cannot solve accurately. Apart from this, we also have a lot of sophisticated instance segmentation frameworks that rely on region-based [19,45,46], prototype-based [47], and contour-based [20,48] proposals. However, such models also require large-scale data for training and are not very scalable for learning new categories in the future with limited training samples [49].

1.2. Contributions

To overcome the limitations mentioned above, we present a novel knowledge distillation-based instance segmentation approach in which a conventional semantic segmentation network is iteratively tuned to perform instance segmentation for extracting prostate tissues in order to grade PCa. Unlike its competitors [9,11], the proposed scheme is adaptable and does not need thorough fine-tuning or re-training efforts. Furthermore, due to its incremental nature, it can accurately extract the stroma, benign, and PCa Gleason patterns from the WSI patches (in any order) when trained with very few training examples. In addition to this, to the best of our knowledge, this paper presents a first attempt at developing a knowledge distillation-based instance segmentation scheme for incrementally learning the stroma, benign, GP-3, GP-4, and GP-5 tissues extraction tasks (from the WSI patches) to calculate Gleason scores in order to grade the PCa progression as per the ISUP standards.

The rest of the paper is organized as follows: Section 2 discusses the proposed scheme, Section 3 enlists the experimental protocols and the details about the datasets, Section 4 shows the evaluations of the proposed scheme and its detailed comparison with the state-of-the-art methods, and Section 5 presents a detailed discussion on the extent of the proposed scheme along with its future prospects.

2. Methods

Our approach fuses knowledge distillation-driven incremental learning with few-shot learning to mold a conventional encoder–decoder, scene parsing, or fully convolutional network to perform instance segmentation. The purpose of performing instance segmentation is to accurately extract the stroma, benign, GP-3, GP-4, and GP-5 tissues from the input WSI samples. Moreover, after obtaining these tissue regions, we compute the majority and minority distribution of Gleason patterns within the candidate WSI scans in order to calculate the Gleason scores. We then use the calculated Gleason scores to grade the severity of PCa as per the ISUP grades. The block diagram of the proposed scheme is depicted in Fig. 2. Here, although the segmentation model is illustrated as an encoder–decoder, but the proposed scheme can also work with a scene-parsing and fully convolutional network (as explained in Section 4.1.1). Apart from this, before we present the detailed description of the proposed approach, we give a brief overview of knowledge distillation, incremental learning, few-shot learning, and incremental few-shot learning within the subsequent sections

2.1. Overview of knowledge distillation

In the context of deep learning, knowledge refers to the model's weights and biasing factors, and knowledge distillation is the process of transferring the knowledge of one model to another during the training phase to accelerate network convergence [50]. The concept of knowledge distillation was initially introduced in model compression [50]. Model compression is a paradigm in which the knowledge of a dense network, referred to as a teacher, is transferred to small student models, allowing them to be deployed on the resource-constrained devices while retaining the teacher network's performance without a significant information loss [51]. Similarly, knowledge distillation has also been introduced within the context of incremental learning, where it allows the model to distill its prior learned experiences (in the current increment) while learning new task representations so that it avoids the catastrophic forgetting phenomena [14,15]. Knowledge distillation, in incremental learning, is typically accomplished by constraining the candidate model through multi-objective functions so that it learns new tasks (in each increment) from the provided examples while retaining its prior knowledge either through episodic memories [52], the subset of previously used training samples [51,53], or through previously trained model instances [14,54].

2.2. Overview of incremental learning

Incremental learning is a paradigm that enables the deep neural networks to continuously learn new classification (or segmentation) tasks o_n while retaining its existing knowledge about the previously learned tasks o_p such that $o = [o_p, o_n]$. To learn these tasks, the network is fed, in each incremental increment i, with I training samples such that $I < \infty$ and $I = [I_p, I_n]$. I_p refers to the training examples related to the previous classes o_p (learned in increment 1 to $i-1$) and I_n represents the training examples used for learning newly stacked classes o_n (in ith increment). Apart from this, the network's prediction error in each increment is minimized through the ground truths labels G corresponding to I in each ith increment. Here, $G = [G_p, G_n]$, where G_p denotes the ground truths of I_p, whereas G_n indicates the ground truths of I_n. Furthermore, in each ith increment, the model produces the output logits $\mathfrak{L}$ such that $\mathfrak{L} = \mathcal{W}\mathcal{F} + \alpha$, where $\mathcal{F}$ defines the feature representations (feature vector), $\mathcal{W}$ signifies the learnable weights, and α represents the biasing factor. Note here that $\mathfrak{L}$ is the aggregation of two groups of logits, $\mathfrak{L}_p$ and $\mathfrak{L}_n$, where $\mathfrak{L}_p$ denotes the logits generated through I_p and $\mathfrak{L}_n$ represents the logits produced through I_n. These logits ($\mathfrak{L} = [\mathfrak{L}_p, \mathfrak{L}_n]$) are subsequently processed by the final activation functions (such as sigmoid, softmax, etc.) to produce final output probabilities $\rho(\mathfrak{L})$. For instance, employing softmax activation function, $\rho(\mathfrak{L}_{x,y})$, for xth training sample belonging to yth class is computed as:

$$\rho(\mathfrak{L}_{x,y}) = \frac{\exp(\mathfrak{L}_{x,y})}{\sum_{t=1}^{o_i-1} \exp(\mathfrak{L}_{x,t})}, \tag{1}$$

where o_i represents the total number of classes being learned in the ith increment. It is emphasized here that $\rho(\mathfrak{L})$ reflects the hard probability generated by $\mathfrak{L}$. Hard probabilities are preferred for traditional classification (or segmentation) problems because they unambiguously identify the most desirable result for the input sequence (at the inference stage). Conversely, in incremental learning, the output logits $\mathfrak{L}$ are usually scaled by the constant factor $\mathcal{T}$ (called temperature) to generate soft probabilities, as expressed below:

$$\rho(\mathfrak{L}^{\mathcal{T}}_{x,y}) = \frac{\exp(\mathfrak{L}^{\mathcal{T}}_{x,y})}{\sum_{t=1}^{o_i-1} \exp(\mathfrak{L}^{\mathcal{T}}_{x,t})}, \tag{2}$$

where $\mathfrak{L}^{\mathcal{T}} = \mathfrak{L}/\mathcal{T}$. These soft labels let the network acquire new information successfully over time without relying on previously acquired knowledge [13,14].

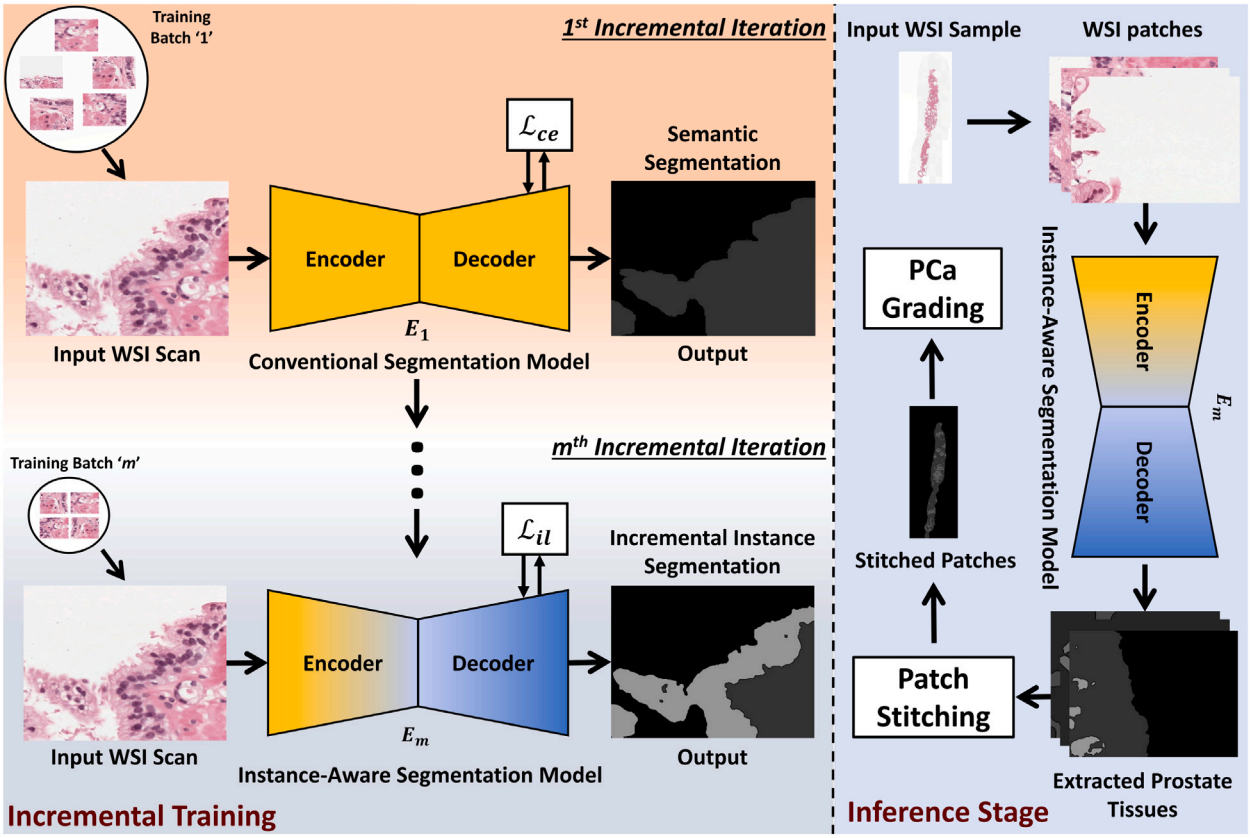

Fig. 2. Block diagram of the proposed scheme. During incremental training, we tune the conventional model E_1 as E_i (in each ith increment) by stacking new prostate tissue classes and constraining it via $\mathcal{L}_{il}$ so that it learns these classes accurately while retaining its previous knowledge. The incremental training is performed in m increments, where the model E_m is updated to perform $m+1$ segmentation tasks related to extracting the background region and the prostate tissues, such as stroma, benign, GP-3, 4, and 5 from the WSI patches. Moreover, at the inference stage, the input scan is first divided into non-overlapping patches that are given to the E_m model for segmenting the prostate tissues. Afterward, the segmented tissue patches are stitched together to generate the segmented representation of the input WSI sample, which is passed to the PCa grader to compute the Gleason scores by analyzing the majority and minority distribution of segmented tissues. Moreover, the computed Gleason scores are then utilized for grading the PCa as per the ISUP standards.

2.3. Overview of few-shot learning

Few-shot learning is a paradigm in which the network is trained for O classification (or segmentation) tasks using I training samples in a supervised $O-$way-$I-$shot approach [55]. Few-shot learning has the benefit of generalizing the network to successfully learn the underlying task(s) given a small amount of training data. Few-shot learning is usually accomplished through either data or parametric optimizations. In data optimizations, the model has the ability to supplement (or reconstruct) their synthesized depictions (typically via adversarial learning) to generalize well (for the underlying task) in a higher dimensional feature space using small training data for each category. In parametric optimizations, the model uses regularized loss functions to learn from a pool of experiences across the related (and comparable) tasks over time. In this case, another model (teacher) may be utilized, which is trained first to encapsulate a large collection of experiences (acquired from larger datasets) to low parametric space, around which the student model optimizes its learning capacity using a small number of training instances. The parametric optimization approach is also similar to meta-learning, more precisely model-agnostic meta-learning [16], where the instructor model is also known as a base-learner, and the student model is alluded to as a meta-learner.

2.4. Incremental few-shot learning

Incremental few-shot learning is a recently introduced paradigm that allows the network to learn a particular set of classification tasks incrementally by using a small number of training instances [56]. Regularization methods (and loss functions) are often used to achieve incremental few-shot learning by allowing networks to learn new tasks (in a gradual manner) without abruptly losing the previously acquired information. Furthermore, incremental learning, especially the one which is driven via knowledge distillation, is different than traditional transfer learning or fine-tuning schemes [14,56], i.e., in a traditional learning schemes, the model is constrained so that its pre-trained weights allow it to converge faster for the newly fed tasks. However, the model (after training) loses its capacity to retain its previously learned knowledge due to the alternation in the network weights as well as by the removal of prior classes from the last network layers. However, in an incremental learning scheme, the model is constrained in a way that it learns new tasks periodically while retaining its knowledge about the previously learned tasks [14].

2.5. Proposed incremental few-shot learning scheme

Unlike the conventional incremental few-shot learning approaches, primarily designed for the classification tasks [14], the proposed incremental few-shot learning scheme, as depicted in Fig. 2, modifies the conventional semantic segmentation network E_1 to perform instance-aware segmentation by incrementally learning different categories of prostate tissues, such as stroma, benign and GP-3, GP-4, and GP-5 from the WSI patches while retaining its previously learned knowledge. Moreover, the network E_i (in ith increment) is constrained through our proposed knowledge distillation-driven loss function $\mathcal{L}_{il}$, which enables it to learn these newly added categories (using their corresponding training examples) while distilling its prior knowledge about prostate

Table 2
Tissue categories which are learned and distilled in each increment.

Increment	New categories	Distilled categories	Description of categories	Loss function
1	0, 1	None	0: Background (BG), 1: Stroma	$\mathcal{L}_{ce}$
2	2	0, 1	0: BG, 1: Stroma, 2: Benign	$\mathcal{L}_{il}$
3	3	0, 1, 2	0: BG, 1: Stroma, 2: Benign, 3: GP-3	$\mathcal{L}_{il}$
4	4	0, 1, 2, 3	0: BG, 1: Stroma, 2: Benign, 3: GP-3, 4: GP-4	$\mathcal{L}_{il}$
5	5	0, 1, 2, 3, 4	0: BG, 1: Stroma, 2: Benign, 3: GP-3, 4: GP-4, 5: GP-5	$\mathcal{L}_{il}$

tissues (learned in the increment 1 to $i-1$) through a small subset of the previously used examples. At the inference stage, the input WSI scan is first divided into a set of non-overlapping patches. These patches are then passed to the E_m model for extracting the stroma, benign, and the GP-3, 4, and 5 regions, where m denotes the total training increments. Afterward, the resultant patches are stitched together to generate the segmented tissue representation of the input WSI sample, which is passed to the PCa grader block, that first computes the Gleason scores by analyzing the majority and minority distributions of the prostate tissues to grade the severity of PCa as per the ISUP standards.

It should be noted here that architecturally, each model instance $E = \{E_1, E_2, \ldots, E_m\}$ is the same (except for the last layer, which is updated periodically to learn new categories of prostate tissues in each increment as defined in Table 2). Apart from this, in the ith increment, the weights of E_{i-1} are first transferred to E_i, which are then refined to recognize the newly added categories of the prostate tissues while retaining the prior knowledge. The total number of segmentation tasks that the network learns during incremental training is $m + 1$. The extra class here is added to extract the background region from the patched WSI scans. More details about the proposed incremental training scheme are presented in Section 2.7.

2.6. Comparison of proposed scheme with existing approaches

Many researchers have worked on developing incremental few-shot learning schemes over the past to tackle classification and segmentation problems. Rebuffi et al. [15] developed a class-incremental representational learning scheme that constrains the classification model via classification and distillation loss functions to learn the generic image classification tasks progressively. Tian et al. [57] developed contrastive representational distillation via Bayesian inference to perform model compression and classification tasks using photographic imagery. Arslan et al. [52] proposed an averaged gradient episodic memories (A-GEM) pipeline to perform class-incremental learning tasks using small chunks of episodic memories. A-GEM [52] allowed the underlying model to retain its previous knowledge during the incremental training. Zhang et al. [54] developed a deep model consolidation loss function that distills the prior knowledge of the classification model (during training) by minimizing the differences between the latent features of the previously trained and the currently tuned model instance. Apart from this, recently, we also developed an incremental few-shot learning scheme that employs mutual distillation objective function to constrain the underlying model towards learning the contextual and semantic similarities within the X-ray imagery in order to perform the desired classification [58] or segmentation [59] tasks at the inference stage. Although, there are many similarities between the proposed scheme and these approaches, including the fact that, during incremental training, the proposed scheme also distills the prior knowledge using a subset of previously used examples. However, to avoid catastrophic forgetting phenomena, the proposed scheme, unlike these approaches, introduces a β hyperparameter that controls the remembrance capacity of the underlying model to retain previously learned knowledge representations while allowing it to learn new segmentation tasks (related to prostate tissues extraction). Furthermore, the proposed scheme does not incur additional computational costs to store the ground truth supervisions [52] or previously trained model instances [54] to distill prior knowledge, unlike these state-of-the-art approaches.

2.7. Incremental training

The proposed scheme, unlike other instance segmentation approaches, is based on incremental learning. Incremental learning allows the deep learning models to periodically learn new categories (classes) while retaining their prior learned experiences. However, traditional incremental learning approaches suffer from catastrophic forgetting. The inability of incremental learning systems to forget their previously knowledge while learning new information. This phenomenon occurs due to the alternation in the networks' weights to accurately learn the newly added classes (from their provided examples), which makes the network forget its past knowledge. However, many researchers have recently proposed contrastive learning [57] and knowledge distillation [53] based strategies to overcome catastrophic forgetting. These approaches allow the model to learn new classes while remembering their past knowledge through multi-objective loss functions [13,56].

In this work, we also propose a novel knowledge distillation approach towards incrementally training the conventional semantic segmentation network to perform instance segmentation. Such instance-aware segmentation allows the conventional segmentation models to distinguish the severity of cancerous prostate tissue within WSI patches to grade the PCa as per the clinical standards. Furthermore, the proposed scheme does not incur any additional computational or training cost (like its competitors), and with few-shot incremental learning, it can effectively extract the prostate tissues. As reported in Table 2, in the first training increment, the model E_1 is trained to recognize two classes, representing the background region and stroma tissues, from the WSI patches. E_1 learns this task using the categorical cross-entropy loss function ($\mathcal{L}_{ce}$). In the second increment, the model E_2 learns to distinguish between the background, stroma, and benign tissue regions depicted within the WSI patches. Consecutively, in the ith increment, the model instance E_i learns $i + 1$ tasks using a small training set that contains examples to learn new prostate tissue categories while distilling the prior representations learned in the increment 1 to $i - 1$. The training process continues for m increments, where the network E_m learns to perform the $m + 1$ segmentation tasks. In each increment i where $i > 1$, the model E_i is trained using the proposed loss function $\mathcal{L}_{il}$, as expressed below:

$$\mathcal{L}_{il} = \mathcal{L}_o + \mathcal{L}_n, \tag{3}$$

where

$$\mathcal{L}_o = -\frac{\beta}{b_s} \sum_{j=0}^{b_s-1} \sum_{k=0}^{c_o-1} t^o_{j,k} \log(p(l^o_{j,k})), \tag{4}$$

and

$$\mathcal{L}_n = \frac{1}{b_s} \sum_{j=0}^{b_s-1} \sum_{k=0}^{c_n-1} q(t^n_{j,k}) \log(q(t^n_{j,k}) / p(l^n_{j,k})). \tag{5}$$

b_s denotes the batch size, c_o and c_n denotes the number of old and new classes, respectively, $p(l^o_{j,k})$ represents the probability of the logit ($l^o_{j,k}$) obtained from the jth training example for the kth class (learned in increment 1 to $i - 1$). From Eq. (4), we can see that $\mathcal{L}_o$ constrains the model (in ith increment) to its retain prior knowledge, and β is a hyperparameter that controls its remembrance capacity, i.e., a large value of β enables the model to focus more on retaining the existing

Table 3
The number of WSI patches used in the proposed incremental training and conventional fine-tuning schemes.

Increment	Incremental training		Fine-tuning	
	Louisville	PANDA	Louisville	PANDA
1	12.90M	12.70M	64.5M	63.5M
2	2.58M	3.17M		
3	2.58M	3.17M		
4	2.58M	3.17M		
5	2.58M	3.17M		

knowledge, while decreasing the β value ensures that the network pays more attention towards learning newly added categories. The optimal value of β which gives the best trade-off for learning new classes while retaining the prior knowledge is determined through extensive ablative experiments (reported in Section 4.1.3).

Moreover, to effectively learn new prostate tissue categories in increment i, the E_i model is also constrained via $\mathcal{L}_n$, as shown in Eq. (5), where $q(t^n_{j,k})$ is the probability of the true labels ($t^n_{j,k}$) belonging to jth example of the kth class, and $p(l^n_{j,k})$ denotes the probability of the logit ($l^n_{j,k}$) generated from training example j for the kth class (learned in the current ith increment). Apart from this, it should be noted that 'm' in the proposed study is 5, i.e., $m = 5$, and it is chosen from the pixel-level labels of prostate tissues (that range from 0 to 5), as reported in Table 2. Also, unlike the state-of-the-art methods, the proposed scheme, due to its incremental nature, can learn the prostate tissue extraction tasks independent of their order. For example, the model E_1 can be trained to identify the background and the GP-5 regions, and E_2 can be modified to extract GP-4 tissues along with the background and GP-5 regions. Similarly, E_3, E_4, and E_5 can be tweaked to recognize GP-3, benign, and stroma regions, respectively, along with retaining the knowledge about segmenting the GP-4, GP-5, and the background regions. Reversing the order does not affect much the overall performance of the employed segmentation model (as it will be evidenced in Section 4.2.5). But the ability of the proposed scheme to perform out-of-order prostate tissue extraction is clinically significant because it allows the clinicians to focus more on screening the early PCa cases or the severe PCa cases, as per the trend of PCa cancer in their clinics and hospitals. More details are presented in Section 4.2.5.

Apart from this, we also trained the segmentation model in a conventional fine-tuning manner to provide the baseline comparison for the proposed incremental instance segmentation approach. For this baseline comparison, we first appended six classes in the final layer of model to extract the background, stroma, benign, GP-3, 4, and 5 regions from the WSI patches in one go. Moreover, after appending the classes, we trained the model using $\mathcal{L}_{ce}$ loss function for 100 epochs (having 512 iterations in each epoch). The comparison between the incremental instance segmentation and the conventional fine-tuning segmentation is presented in Section 4.2.3.

2.8. Inference stage

After training the segmentation model till m incremental increments, we apply it to the test WSI scans, where the scans are first decomposed into fixed-size non-overlapping patches. These patches are then passed to the E_m, which extracts and identifies the underlying prostate tissues from it as per the clinical standards defined in Table 1. Moreover, the segmented patches are then stitched together and are passed to the PCa grading block to compute Gleason scores in order to grade the input WSI sample.

2.8.1. PCa grading

We developed a simple yet effective rule to grade the PCa (within the candidate WSI sample) as per the ISUP standards defined in Table 1 [7]. When the stitched scan containing the segmented prostate tissues is obtained, we search for the presence of majority and minority distribution of the prostate tissues (excluding the background region) to compute Gleason scores, where the majority distribution is the one that is most frequently observed within the input WSI sample, and the minority distribution is the second-most frequently observed prostate tissue representation within the WSI sample. For example, if the majority distribution is GP-4 (having the label '4'), and the minority distribution has GP-3 (having label '3'), then the Gleason score is computed as 4+3, where the candidate WSI sample will be graded as ISUP-3 [7]. Also, for the minority distribution to be considered for the Gleason score calculation, its total area should be more than 3% of the overall tumor tissue area (represented by GP-3, 4, and 5 within the WSI sample). Otherwise, the Gleason score is computed by multiplying the majority distribution label by 2. For example, if the majority distribution label is GP-4 and the minority is GP-3, but the GP-3 area is less than 3% of the total PCa tumoral area, then the Gleason score would be 4+4, grading the WSI sample as having ISUP-4 level PCa [7].

3. Experimental setup

3.1. Datasets

We evaluated the proposed scheme on two PCa datasets. The first one is our local dataset, dubbed the Louisville dataset, and the second dataset is the popular and publicly available Prostate Cancer Grading Assessment (PANDA) dataset [60]. The detailed description of each dataset is presented below:

3.1.1. Louisville dataset

The first dataset on which we tested the proposed scheme is the Louisville dataset that contains 10,516 digitized H&E stained prostate cancer histology samples, which are obtained after biopsy examination. The Louisville dataset has been acquired from the Louisville Hospital, and it has been extensively marked (both pixel-wise and biopsy-wise) by the expert pathologists from the University of Louisville School of Medicine, USA. In addition to this, each WSI has been divided into non-overlapping patches of $350 \times 350 \times 3$ (yielding around 71.7M patches). As reported in Table 3, for the conventional fine-tuning (discussed in Section 4.2.3), we allocated 90% of the data (within the Louisville dataset) for training and the rest of 10% for testing purposes. Moreover, for the proposed incremental instance segmentation, we used only 40% of that 90% training data (that was used in the conventional fine-tuning), where 50% of that 40% data is used in the first incremental (to differentiate background and stroma regions), that the rest of the 50% is used in the subsequent training increments to learn the remaining prostate tissue categories.

3.1.2. PANDA dataset

The second dataset on which we trained the proposed scheme and compared it with the state-of-the-art works is the PANDA dataset [60]. PANDA dataset was recently introduced in one of the MICCAI-2020 challenges that have been organized by Radboud University Medical Center and Karolinska Institute in collaboration with Tampere University. Overall, the dataset contains 11,000 H&E-stained WSI samples that have been acquired after a thorough biopsy examination. Moreover, for these samples, detailed mask-level annotations were provided in order to train and validate the computer-aided diagnostic systems for extracting the stroma, benign, GP-3, 4, and 5 regions from the WSI samples. Also, the dataset contains biopsy-level annotations to train and validate these systems for computing Gleason scores and the ISUP grades. For extracting the prostate tissues, we divided the dataset into non-overlapping patches of $350 \times 350 \times 3$ size, which results in around 79.4M patches. Apart from this, for the baseline experiments, we used 80% of the WSI samples for training purposes and the rest of 20% for testing purposes. However, for incremental instance segmentation,

we used only 40% of that 80% training data (used in the baseline experiments). From that 40% data, we used 50% in the first iteration and the rest of 50% data in the subsequent training increments as reported in Table 3.

In addition to this, we only used the labels provided with the PANDA dataset in order to maintain a fair comparison with the state-of-the-art. Similarly, we discarded those cases for which the pixel-level annotations were not provided by the dataset authors. The labels marked in the PANDA dataset range from 0 to 5, where '0' represents the background region, '1' represents the stroma region, '2' represents the benign region, '3' illustrates the GP-3 region, '4' represents GP-4, and '5' represents the GP-5. Furthermore, since the annotations of the Karolinska institute scans do not differentiate between different Gleason pattern labels, therefore, in order to train the proposed framework, we used a single tag '3' to denote the cancerous tissues for the Karolinska institute scans. Similarly, we used '0' to represent the background region and '1' to denote stroma/benign tissues. Nonetheless, we believe that combining the samples of Karolinska and Radboud institutes is not a good approach, and such an experimental plan can also undermine the performance of the underlying screening system towards accurately extracting the prostate tissues. So, in the future, we envisage using only the Radboud institute scans to grade prostate cancer tissues as per the ISUP standards.

Moreover, as reported in Table 3, the proposed incremental instance segmentation scheme requires 23.22M patches from the proposed (Louisville) dataset and 25.38M patches from the PANDA dataset in order to train the model. The number of these patches is around 60% less (or 'fewer') than the number of patches required in the conventional fine-tuning process (as evident from Table 3). Hence, relative to the traditional training approaches, we dubbed the proposed scheme as 'few-shot incremental learning'.

3.2. Implementation details

The proposed scheme is developed using Python 3.7.8, with Keras 2.3.0 and TensorFlow 2.1.0. The machine on which the experiments were conducted has Intel(R) Core(TM) i9-10940X@3.30 GHz CPU, 160 GB RAM, NVIDIA Quadro RTX 6000 with CUDA v11.0.221, and cuDNN v7.5. Apart from this, ADADELTA [61] was used as an optimizer during the training with a default learning and decay rate of 1 and 0.95, respectively. Also, the training was conducted using a batch size of 16.

3.3. Evaluation metrics

The prostate tissues extraction performance of the proposed scheme is measured using intersection-over-union (IoU), as expressed below:

$$IoU = \frac{T_p}{T_p + F_p + F_n}, \tag{6}$$

where T_p denotes the pixel-wise true positives indicating the correct extraction of prostate tissue instances as compared to the ground truths, F_p denotes the false positives showcasing the incorrect recognition of background pixels as prostate tissues, and F_n represents the false negatives indicating the incorrect extraction of prostate tissues as background. Moreover, we also computed the mean IoU (μIoU) scores by averaging the IoU scores of all the prostate tissue categories, respectively. Apart from this, we used the area under the curve (AUC) and the mean average precision (mAP) scores obtained from the ROC and PR curves, respectively, to measure the tissue extraction performance. Also, to measure the PCa grading performance of the proposed scheme, we used biopsy-level true positive rate (T_{PR}), positive predicted value (P_{PV}), the F1 scores, and the quadratic weighted kappa, as expressed below:

$$k = 1 - \frac{\sum_{i=0}^{u-1}\sum_{j=0}^{u-1} w_{i,j}x_{i,j}}{\sum_{i=0}^{u-1}\sum_{j=0}^{u-1} w_{i,j}e_{i,j}}, \tag{7}$$

where u represents the number of codes, w represents the quadratic weights, x represents the predictions, and e represents the ground truths. We also want to highlight here that we computed the quadratic weighted kappa score at the ISUP (class) level in this research, where to calculate kappa for each ISUP grade (class), we decomposed that respective class as positive '1' and the rest of the categories as negative '0'. For example, to compute kappa for ISUP-1, we marked the prediction of the proposed scheme for ISUP-1 as '1' and the rest as '0'. Similarly, to compute kappa for ISUP-2, we marked the predictions for ISUP-2 as '1' and the rest as '0'. To calculate kappa for ISUP-3 grade, we kept the predictions of the proposed scheme as '1' and the rest as '0'. To compute kappa for ISUP-4, we marked its predictions as '1', and the rest as '0'. Moreover, to calculate kappa for ISUP-5, we kept its predictions as '1' and the rest as '0'.

We also want to highlight here that in order to maintain a fair comparison of the proposed framework with state-of-the-art, we used the same experimental protocols and the evaluation metrics (including the quadratic weighted kappa) for all the methods to compute their quantitative scores.

4. Results

To evaluate the proposed scheme, we first conducted a series of ablative experiments through which we determined the segmentation network (to be used within the proposed scheme), the optimal backbone network to be coupled with the chosen segmentation model, and the optimal value of the β that produces maximum segmentation performance at the inference stage. After discussing the ablative experiments, we present a comparative study in which we compared the performance of the proposed scheme with state-of-the-art methods through different metrics on both datasets. We also want to highlight here that the scores for all of the experimentation (reported in this section) are obtained by taking the mean of five consecutive runs. This ensures that these obtained scores are not random but are statistically significant and repeatable.

4.1. Ablation studies

The ablation studies which we report for the proposed scheme relate to (1) determining the best segmentation network for extracting prostate tissues from the WSI patches, (2) determining the best backbone network to give the optimal tissue extraction performance when coupled with the best segmentation model, and (3) determining the optimal remembrance factor β within $\mathcal{L}_o$ objective function to achieve the best incremental instance segmentation performance.

4.1.1. Choice of segmentation network

The proposed incremental instance segmentation scheme is independent of the segmentation model, and thus it can be coupled with any pre-trained network. To validate this, in the first ablation study, we plugged different encoder–decoder, scene parsing, and fully convolutional networks, such as DRN [9], RAGNet [62], PSPNet [63], UNet [40], and FCN-8 [64] within the proposed scheme, and determined the best segmentation network which can accurately extract prostate tissues from the WSI patches. Moreover, for fairness, all of the networks are backboned through ResNet-50 [65]. Apart from this, we incrementally trained these models to extract different categories of the prostate tissues, such as stroma, benign, GP-3, 4, and 5, and measured their performance through mean IoU (μIoU) scores, as shown in Table 4. We can observe here that, on both the Louisville and PANDA datasets, the overall best performance is achieved by the DRN [9] which is leading the second-best RAGNet [62] by 4.01% and 5.78%, respectively. Although DRN [9] achieved second-best performance for extracting GP-5 on the Louisville dataset, and it is lagging behind PSPNet [63] by 3.11%. But since it is outperforming the PSPNet [63] for extracting other prostate tissues on both datasets, we chose it as a candidate segmentation model within the proposed scheme to perform the rest of the experimentation.

Table 4
Performance comparison of different segmentation networks for extracting the prostate tissues as per the ISUP standards [7]. Bold indicates the best performance, while the second-best performance is underlined. For fairness, all the segmentation networks are backboned through ResNet-50 [65] with $\beta = 10$, and are trained incrementally using the same experiment protocols.

Dataset	Tissues	DRN [9]	RAGNet [62]	PSPNet [63]	UNet [40]	FCN-8 [64]
Louisville	Stroma	**0.7803**	0.7524	0.7618	0.7296	0.7149
	Benign	**0.5226**	0.4816	0.4652	0.4293	0.3724
	GP-3	**0.4198**	0.4034	0.3759	0.3481	0.3156
	GP-4	**0.3887**	0.3724	0.3497	0.3156	0.2803
	GP-5	0.2951	0.2728	**0.3043**	0.2657	0.2518
	μIoU	**0.4782**	0.4598	0.4513	0.4176	0.3870
PANDA [60]	Stroma	**0.8143**	0.7952	0.7748	0.7425	0.7294
	Benign	**0.5891**	0.5176	0.4939	0.4402	0.4106
	GP-3	**0.4684**	0.4482	0.4075	0.3518	0.3268
	GP-4	**0.4057**	0.3940	0.3652	0.3264	0.3142
	GP-5	**0.3482**	0.3271	0.3158	0.2953	0.2853
	μIoU	**0.5251**	0.4964	0.4714	0.4312	0.4132

Table 5
Comparison of different backbone models, in terms of μIoU scores, when plugged within DRN [9] to extract the prostate tissues. Bold indicates the best performance, while the second-best performance is underlined. The β value is chosen to be 10.

Dataset	MobileNet [35]	VGG-16 [66]	ResNet-50 [65]
Louisville	0.4419	0.4051	**0.4782**
PANDA [60]	0.4805	0.4638	**0.5251**

Table 6
Performance of the proposed scheme (in terms of μIoU score) for extracting prostate tissues after being incrementally trained with different values of β. Here, the proposed scheme is driven via DRN [9] with ResNet-50 [65] backbone. Bold indicates optimal performance.

Dataset	β			
	5	10	15	20
Louisville	0.3581	**0.4782**	0.3653	0.3128
PANDA [60]	0.4786	**0.5251**	0.4839	0.4452

4.1.2. Choice of backbone

The second series of ablation experiments is related to determining the best backbone network that can produce distinct feature representations to optimally drive the DRN model (selected in the previous ablative study) towards extracting the prostate tissues. For this purpose, when installed different pre-trained networks, such as, MobileNet [35], VGG-16 [66], and the ResNet-50 [65] within the encoder end of the DRN [9], and trained it incrementally to extract the stroma, benign, GP-3, 4, and 5 tissues. From Table 5, we can observe that, on both the Louisville and PANDA [60] dataset, the best tissue extraction performance, in terms of μIoU, is achieved when the latent features are generated through ResNet-50 [65], and it is leading the second-best MobileNet [35] by 8.21% on the Louisville dataset and 9.28% on the PANDA [60] dataset. The performance improvements with ResNet-50 emanate from the fact that it uses more residual operators to produce richer feature distribution while leading to the best tissue extraction performance. Although, it uses more parameters than the MobileNet [35]. But since our focus is on achieving more accurate results, therefore, we chose ResNet-50 [65] as a candidate backbone network for the rest of the experimentation.

4.1.3. Optimal remembrance factor

During the incremental training of the proposed scheme, β in $\mathcal{L}_o$ constrains the underlying segmentation network to retain its prior knowledge while learning new segmentation categories. In this experiment, we incrementally trained the selected DRN [9] model by varying the β value from 5 to 20 in the step of 5 and then evaluated the trained network on the test WSI scans in order to determine which β values gives the best prostate tissues extraction performance. From Table 6, we can see that, on both datasets, increasing β value makes the model retain its prior knowledge by showing good resistance to catastrophic forgetting (see the performance gain from $\beta = 5$ to $\beta = 10$). However, increasing the value of β too much constrains it towards properly learning the newly added categories, which results in the performance degradation (please see the outcome of $\beta > 10$ in Table 6 for both datasets). Therefore, we chose $\beta = 10$ as the optimal remembrance factor for incrementally training the proposed scheme for extracting the prostate tissues.

4.2. Comparative study

In this series of experiments, we compared the performance of the proposed scheme with state-of-the-art methods for extracting the prostate tissues and grading PCa as per the ISUP standards [7]. Furthermore, we also compared the performance of the proposed scheme with its baseline variant that is trained using the fine-tuning approach with the dataset reported in Table 3. The purpose of comparing the incrementally trained proposed scheme with the fine-tuning variant is to see the extent of performance improvement which we get towards grading PCa based on our incremental instance segmentation scheme as compared to conventional semantic segmentation approach. More details are presented in 4.2.3.

4.2.1. Comparison of prostate tissues extraction

In this experiment, we compared the performance of the proposed scheme with the state-of-the-art methods on both datasets to extract the prostate tissue from the WSI patches [6,10]. The comparison is reported in Tables 7, and 8 where we can see that the proposed scheme outperforms the other frameworks by 2.01% on the Louisville dataset, and by 4.45% on the PANDA dataset [60] in terms of μIoU score. Moreover, on both datasets, the second, and third-best performance is achieved by the Mask Scoring (MS) R-CNN and the HoVer-Net [11] with the μIoU score of 0.4686 and 0.4667, respectively, and 0.5027, and 0.4960, respectively. Furthermore, its worth mentioning that the proposed scheme, although trained incrementally, produces competitive performance with the conventional instance segmentation models such as MS R-CNN [19], HoVer-Net [11], Mask R-CNN [45], and YOLACT [47], as it can be seen in Tables 7, and 8. It is due to the fact that these models localize the area of interest either through region-based [11,19,45], or prototype-based proposals [47], which makes them vulnerable to the scans having highly correlated structural and textural regions (such as the prostate tissues within the WSI patches). On the contrary, the proposed scheme uses DRN [9] for tissue extraction that generates latent space representations by decomposing the scans across various scales and then fuses these representations together to preserve the contextual properties of the prostate tissues [9]. Nevertheless, for extracting the large regions of prostate tissues (especially GP-4 and GP-5),

Table 7
Prostate tissues extraction comparison in terms of (μIoU) on the Louisville dataset. For fairness, the backbone of all the models is comprised of residual blocks [65]. Moreover, 'DSRL' stands for Dual Super-Resolution Learning [67]. Also, bold indicates the best score, while the second-best performance is underlined.

Tissues	Proposed	MS R-CNN [19]	Mask R-CNN [45]	YOLACT [47]	DSRL [67]	HoVer-Net [11]
Stroma	**0.7803**	0.7696	0.7619	0.7483	0.6853	0.7641
Benign	**0.5226**	0.5013	0.4972	0.4628	0.3495	0.5083
GP-3	**0.4198**	0.3974	0.3906	0.3591	0.2847	0.4016
GP-4	0.3732	**0.3746**	0.3631	0.3066	0.2416	0.3652
GP-5	0.2951	**0.3005**	0.2960	0.2423	0.2245	0.2946
μIoU	**0.4782**	0.4686	0.4617	0.4238	0.3571	0.4667

Table 8
Prostate tissues extraction comparison in terms of (μIoU) on the PANDA dataset [60]. For fairness, the backbone of all the models is comprised of residual blocks [65]. Moreover, 'DSRL' stands for Dual Super-Resolution Learning [67]. Also, bold indicates the best score, while the second-best performance is underlined.

Tissues	Proposed	MS R-CNN [19]	Mask R-CNN [45]	YOLACT [47]	DSRL [67]	HoVer-Net [11]
Stroma	**0.8143**	0.7841	0.7984	0.7558	0.7154	0.7796
Benign	**0.5891**	0.5317	0.5243	0.4981	0.3781	0.5217
GP-3	**0.4684**	0.4253	0.4386	0.3824	0.3126	0.4452
GP-4	**0.4057**	0.3984	0.3969	0.3549	0.2843	0.3861
GP-5	0.3482	0.3280	**0.3557**	0.2756	0.2905	0.3475
μIoU	**0.5251**	0.4935	0.5027	0.4533	0.3961	0.4960

the DRN [9] within the proposed scheme sometimes produces eroded results, which affects a bit on the PCa grading performance, as shown in Fig. 5. However, this limitation can be easily addressed through additional post-processing schemes, as discussed within Section 5.

Apart from this, on the Louisville dataset, we also measured the performance of the proposed scheme through AUC and mAP scores generated through ROC and PR curves. From Fig. 4(A and B), we can observe the proposed scheme achieved the AUC and mAP scores of 0.9658 and 0.9527, respectively, for extracting stroma regions as compared to the rest of the tissues. Similarly, it achieved the AUC and mAP scores of 0.9631, and 0.9514, for extracting benign regions, 0.9235, and 0.9134, for extracting GP-3 regions, 0.9044, and 0.8964, for extracting GP-4 regions, and 0.9007, and 0.8852, for extracting GP-5 regions, respectively. Here, we can also notice that there is a significant gap between prostate tissues extraction performance in terms of μIoU (as evident from Table 9) and AUC (or mAP) scores (as evident from Fig. 4). This performance difference emanates from the fact that the AUC and mAP scores are computed through pixel-wise binary classification, where the region of respective prostate tissue was assigned one while the rest of the pixels were assigned zero, which resulted in the lower number of F_p and higher number of T_n. On the contrary, the μIoU scores (in Table 9) are computed through pixel-wise multi-class prostate tissue segmentation which produces more F_p. However, considering the fact that the proposed scheme outperforms the state-of-the-art despite being trained incrementally, the performance of the proposed scheme is appreciable.

4.2.2. Comparison of PCa grading

In the next experiment, we compare the performance of the proposed scheme with state-of-the-art schemes for grading PCa. Here, at the inference stage, we first divide the candidate WSI sample into non-overlapping patches and extract the prostate tissues from these patches. Then we stitch these patches together and search for the presence of majority and minority distribution of prostate tissues to compute Gleason scores, where the majority distribution is the one which is most frequently observed within the stitched WSI landscape, and the minority distribution is the second-most frequently observed tissue within the stitched WSI sample. After computing the Gleason scores, the biopsy-level PCa grading is computed as per the ISUP standards reported in Table 1. Then, the graded scans (from both datasets) are evaluated using the T_{PR}, P_{PV}, F_1, and the quadratic weighted kappa scores, as reported in Tables 9, and 10. To ensure fairness, we trained all the competitive methods using the same number of scans and the experimental protocols which we followed to train the proposed scheme. From Table 9, we can observe that, overall, the best performance for grading ISUP-2, ISUP-3, ISUP-4, and ISUP-5 on the Louisville dataset is achieved by the proposed scheme, where it leads the other frameworks by 1.24%, 10.73%, 0.458%, and 4.25%, respectively, in terms of F_1 score. Similarly, on the PANDA dataset [60], the proposed scheme achieved the performance improvement of 2.27%, 1.77%, 11.42%, 1.38%, and 4.29% in terms of F_1 score towards extracting the ISUP-1, ISUP-2, ISUP-3, ISUP-4, and ISUP-5 PCa, respectively. Although, it should be noted that on the Louisville dataset, for ISUP-1 grade, the performance of the proposed scheme lags behind the training variant-III teacher–student topology proposed by Marini et al. [36]. Nevertheless, considering the fact that the proposed scheme has outperformed all the competitive frameworks for extracting the rest of the ISUP grades on the Louisville dataset, which is more critical, and also outperformed the state-of-the-art towards extracting all the ISUP grades on the PANDA dataset [60], the performance of the proposed scheme is appreciable.

4.2.3. Comparison with baseline variant

In these series of experiments, we compared the proposed scheme's performance with the baseline variant that is trained in a conventional fine-tuning manner. Here, for each dataset, the number of scans that we used for training is shown in Table 3. For fine-tuning variant, we feed the data for all the five prostate tissue categories (plus the background) at once, where using $\mathcal{L}_{ce}$ loss function, the selected DRN model is trained for 100 epochs (containing 512 iterations in each epoch) with a batch size of 16. Moreover, after training the model, we applied it to the patched test WSI scan to extract the prostate tissues. After extracting these tissues, each patch is stitched together, and this stitched representation is passed to the PCa grading block to assess the severity of PCa. Apart from this, the incremental instance segmentation variant is trained using the same protocols as described in Section 2.7. From Table 11, we can observe that, on the Louisville dataset, the incrementally trained variant achieved 15.21%, 20.46%, 58.16%, 15.87%, and 38.51% performance improvements over the fine-tuning variant in terms of F1 score towards grading ISUP-1, ISUP-2, ISUP-3, ISUP-4, and ISUP-5 PCa, respectively. Similarly, on the PANDA [60] dataset, the incrementally trained variant achieved 18.42%, 21.70%, 58.46%, 17.42%, and 41.37% improvements in term of F1 score over the baseline variants for grading ISUP-1, ISUP-2, ISUP-3, ISUP-4, and ISUP-5 PCa, respectively. These improvements stem from the fact that the

Table 9
PCa grading comparison with state-of-the-art methods on the Louisville dataset. For fairness, all the models are trained incrementally using the same scans and experimental protocols.

Method	Metric	ISUP-1	ISUP-2	ISUP-3	ISUP-4	ISUP-5
Proposed	T_{PR}	0.602	0.607	0.568	0.742	0.567
	P_{PV}	0.327	0.539	0.118	0.311	0.125
	F_1	0.424	0.571	0.196	0.438	0.245
	k	0.238	0.311	0.102	0.257	0.147
MS R-CNN [19]	T_{PR}	0.565	0.598	0.508	0.753	0.684
	P_{PV}	0.316	0.534	0.107	0.309	0.141
	F_1	0.405	0.564	0.177	0.436	0.235
	k	0.221	0.303	0.085	0.255	0.131
HoVer-Net [11]	T_{PR}	0.561	0.575	0.496	0.749	0.673
	P_{PV}	0.305	0.520	0.101	0.301	0.135
	F_1	0.395	0.546	0.168	0.429	0.226
	k	0.206	0.280	0.076	0.241	0.121
Nagpal et al. [34]	T_{PR}	0.549	0.568	0.484	0.736	0.684
	P_{PV}	0.297	0.511	0.097	0.293	0.134
	F_1	0.385	0.538	0.162	0.420	0.225
	k	0.193	0.267	0.069	0.229	0.119
Arvaniti et al. [10]	T_{PR}	0.669	0.550	0.365	0.733	0.561
	P_{PV}	0.320	0.511	0.077	0.293	0.116
	F_1	0.433	0.530	0.127	0.419	0.192
	k	0.230	0.263	0.039	0.228	0.091
Marini et al. [36]	T_{PR}	0.678	0.539	0.377	0.755	0.572
	P_{PV}	0.324	0.515	0.080	0.298	0.118
	F_1	0.438	0.527	0.132	0.428	0.196
	k	0.236	0.266	0.043	0.237	0.096
Bulten et al. [39]	T_{PR}	0.557	0.593	0.496	0.747	0.572
	P_{PV}	0.303	0.517	0.101	0.299	0.119
	F_1	0.392	0.552	0.167	0.427	0.197
	k	0.202	0.280	0.075	0.238	0.097
Strom et al. [41]	T_{PR}	0.478	0.519	0.334	0.709	0.421
	P_{PV}	0.234	0.433	0.059	0.249	0.077
	F_1	0.314	0.472	0.101	0.369	0.130
	k	0.100	0.155	0.009	0.155	0.029

Table 10
PCa grading comparison with state-of-the-art methods on the PANDA dataset [60]. For fairness, all the models are trained incrementally using the same scans and experimental protocols.

Method	Metric	ISUP-1	ISUP-2	ISUP-3	ISUP-4	ISUP-5
Proposed	T_{PR}	0.669	0.608	0.568	0.741	0.567
	P_{PV}	0.393	0.540	0.118	0.310	0.125
	F_1	0.495	0.572	0.195	0.438	0.243
	k	0.296	0.330	0.106	0.267	0.142
MS R-CNN [19]	T_{PR}	0.635	0.598	0.508	0.741	0.684
	P_{PV}	0.382	0.530	0.106	0.305	0.140
	F_1	0.477	0.562	0.175	0.432	0.233
	k	0.277	0.318	0.088	0.259	0.135
HoVer-Net [11]	T_{PR}	0.628	0.575	0.496	0.749	0.673
	P_{PV}	0.371	0.518	0.101	0.299	0.135
	F_1	0.467	0.545	0.168	0.428	0.225
	k	0.262	0.297	0.079	0.250	0.126
Nagpal et al. [34]	T_{PR}	0.597	0.568	0.484	0.746	0.684
	P_{PV}	0.357	0.504	0.096	0.291	0.133
	F_1	0.446	0.534	0.160	0.419	0.222
	k	0.239	0.280	0.071	0.238	0.122
Arvaniti et al. [10]	T_{PR}	0.694	0.554	0.365	0.731	0.561
	P_{PV}	0.370	0.502	0.075	0.288	0.114
	F_1	0.483	0.527	0.125	0.413	0.189
	k	0.264	0.274	0.041	0.232	0.093
Marini et al. [36]	T_{PR}	0.697	0.541	0.377	0.741	0.572
	P_{PV}	0.370	0.502	0.077	0.289	0.115
	F_1	0.484	0.521	0.129	0.416	0.192
	k	0.264	0.271	0.044	0.234	0.095
Bulten et al. [39]	T_{PR}	0.607	0.593	0.496	0.739	0.572
	P_{PV}	0.361	0.508	0.099	0.293	0.117
	F_1	0.453	0.547	0.165	0.419	0.194
	k	0.246	0.290	0.076	0.239	0.098
Strom et al. [41]	T_{PR}	0.513	0.519	0.334	0.692	0.421
	P_{PV}	0.281	0.418	0.057	0.239	0.074
	F_1	0.363	0.463	0.098	0.355	0.126
	k	0.127	0.158	0.009	0.151	0.029

prostate tissues have similar textural representations within the WSI patches and also have imbalanced data distribution between classes (e.g., GP-3, 4, and 5 have a low number of samples as compared to stroma and benign). Therefore, when the model is trained to recognize all of the prostate tissue categories at once, its performance deteriorates (especially towards extracting GP-3, 4, and 5). On the other hand, when the model is incrementally trained with few-shot data distribution, where it is constrained to equally emphasize on learning the new categories while retaining the prior knowledge (through β factor), the proposed scheme produces more robust results towards extracting the prostate tissues, leading to a better PCa grading performance. Also, due to the incremental nature of the proposed scheme, it can be deployed in the clinical settings for grading PCa, where, if needed, it can be easily adapted to learn new categories of cancerous tissues with few-shot training, unlike the conventional fine-tuning system, which requires a thorough training on large-scale and well-annotated data in order to produce decent screening performance [16].

4.2.4. Qualitative evaluations

Fig. 3 reports the qualitative evaluations in which we can see that the proposed scheme has decently extracted the prostate tissues from the WSI patches. For example, see the pairs (A, B, C), (G, H, I), (J, K, L), (P, Q, R), (S, T, U), and (AB, AC, AD) in Fig. 3. Although, we can notice that the background region and the dark artifacts in many of the WSI patches within Fig. 3 are classified as stroma and GP-5, respectively, by the proposed framework. Nevertheless, such results are obtained because of the discrepancies within the ground truth annotations, which the clinicians marked, and since the proposed system is trained on such data (that has irregular ground truth supervisions), it also classifies the test WSI patches in a similar way. Apart from this, we also have some pixel-level misclassifications for extracting GP-4, and GP-5 in Fig. 3(F) compared to Fig. 3(E) but noticing the extreme overlap between stroma, GP-4, and GP-5 tissues in Fig. 3(D), the extraction of GP-5 in Fig. 3(E) is quite appreciable.

Table 11
Comparison between the incrementally trained and fine-tuned version of the proposed scheme towards grading PCa as per the ISUP standards.

Dataset	Variant	Metric	ISUP-1	ISUP-2	ISUP-3	ISUP-4	ISUP-5
Louisville	Increment	T_{PR}	0.602	0.607	0.568	0.742	0.567
		P_{PV}	0.327	0.539	0.118	0.311	0.125
		F_1	0.424	0.571	0.196	0.438	0.205
		k	0.238	0.311	0.102	0.257	0.107
	Fine-tuning	T_{PR}	0.586	0.504	0.258	0.711	0.471
		P_{PV}	0.268	0.447	0.049	0.258	0.088
		F_1	0.368	0.474	0.082	0.378	0.148
		k	0.149	0.175	0.004	0.170	0.046
PANDA [60]	Increment	T_{PR}	0.669	0.608	0.568	0.741	0.567
		P_{PV}	0.393	0.540	0.118	0.310	0.125
		F_1	0.495	0.572	0.195	0.438	0.205
		k	0.296	0.330	0.106	0.267	0.112
	Fine-tuning	T_{PR}	0.614	0.507	0.258	0.709	0.471
		P_{PV}	0.317	0.437	0.048	0.252	0.086
		F_1	0.418	0.470	0.081	0.373	0.145
		k	0.180	0.185	0.002	0.174	0.048

4.2.5. Reversing the prostate tissue learning order during training

In this experiment, we evaluated the proposed scheme's capacity to recognize the prostate tissues when their learning order is reversed. For example, we trained E_1 to segment GP-5 and the background regions in the first increment. Afterward, we modified E_2 to extract GP-4, background, and GP-5 tissues. In the next increment, we updated

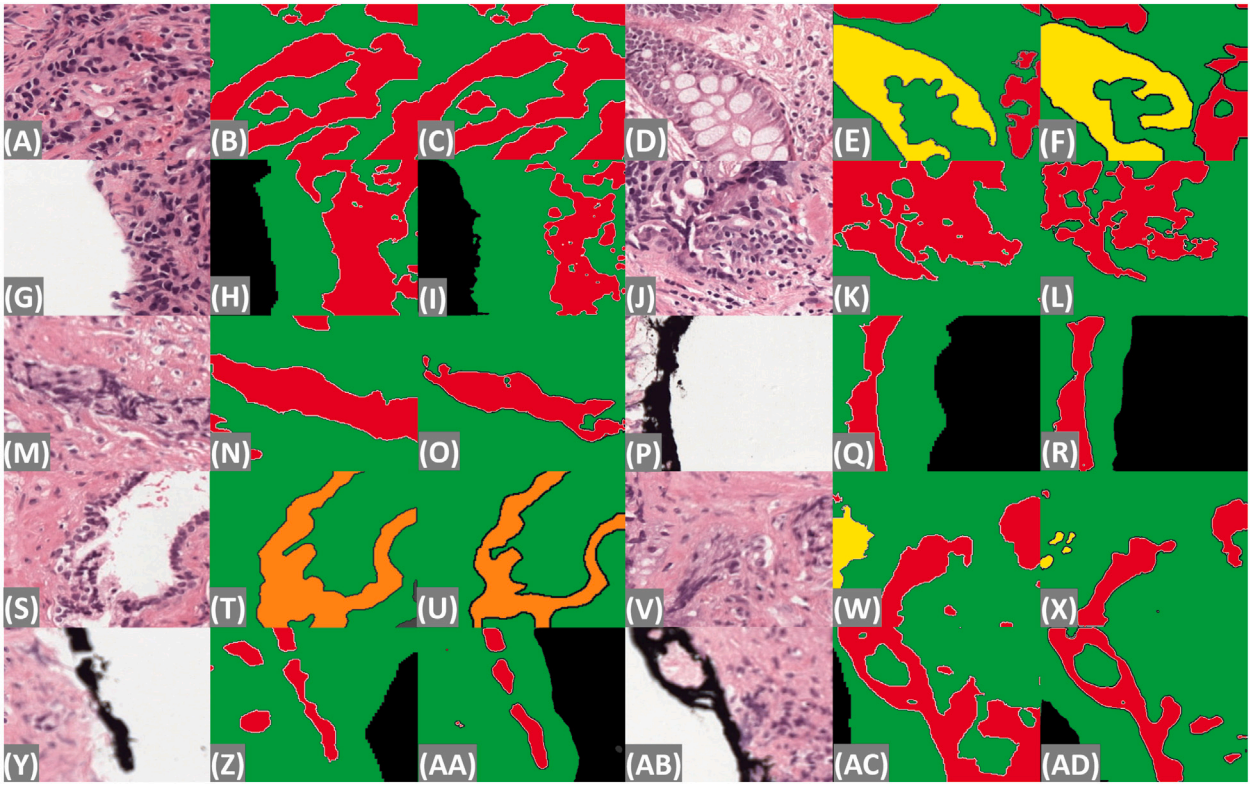

Fig. 3. Qualitative evaluations of the proposed scheme. The first and fourth column shows the original WSI patches, the second and fifth column shows the ground truths, and the third and sixth column shows the extracted results. Moreover, for better visualization, we have color-coded the scans here, where the black color in the second, third, fifth, and sixth columns represents the background region, the green color represents the stroma regions, orange color represents GP-3 regions, yellow color represents GP-4 regions, and the red color represents GP-5 tissues.

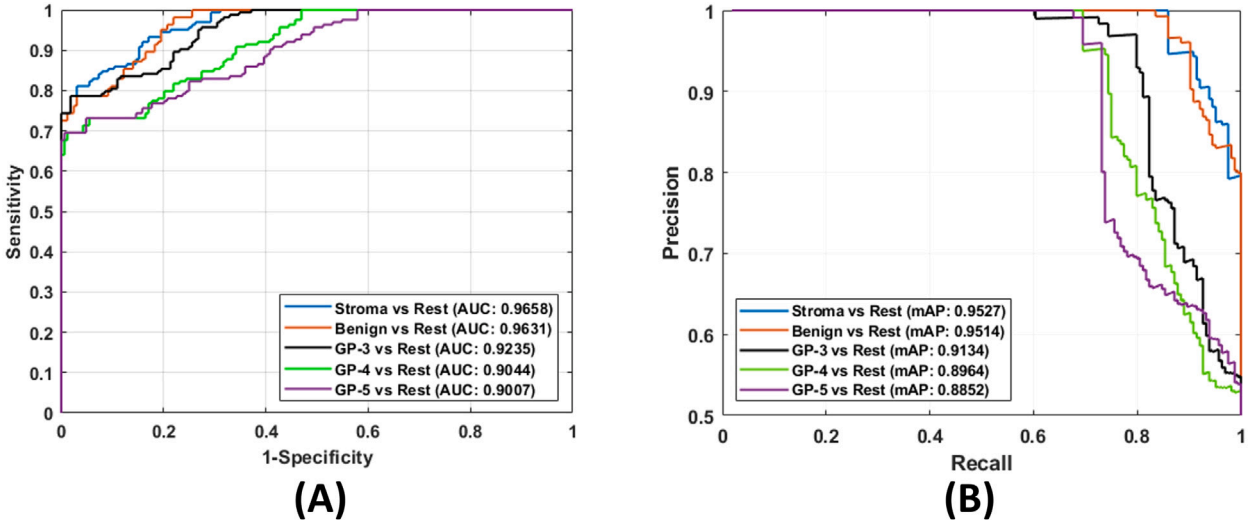

Fig. 4. Performance evaluation of the proposed scheme for extracting prostate tissue patterns through (A) ROC and (B) PR curves.

E_3 to segment GP-3 regions while retaining its previous knowledge about the background, GP-5, and GP-4. In the fourth increment, E_4 was modified to extract benign tissues along with previously learned tissue categories, and in the fifth increment, E_5 was updated to extract the background, GP-5, GP-4, GP-3, benign, and stroma tissue regions. From Table 12, we can observe that although there is not much difference in the overall performance of the proposed scheme (in terms of μIoU) when the order of prostate tissue learning is reversed across both datasets. But reversing the prostate learning order gave a considerable performance boost of 31.40% and 56.96% on the Louisville dataset, in terms of μIoU score, towards recognizing the severe PCa grades (i.e., GP-4 and GP-5) as compared to the conventional order. Similarly, on the PANDA dataset [60], reversing the order gave 27.31% and 50.89% performance improvements over conventional order towards extracting GP-4 and GP-5 regions, respectively. However, it should also be noted that, at the same time, the overall performance of the

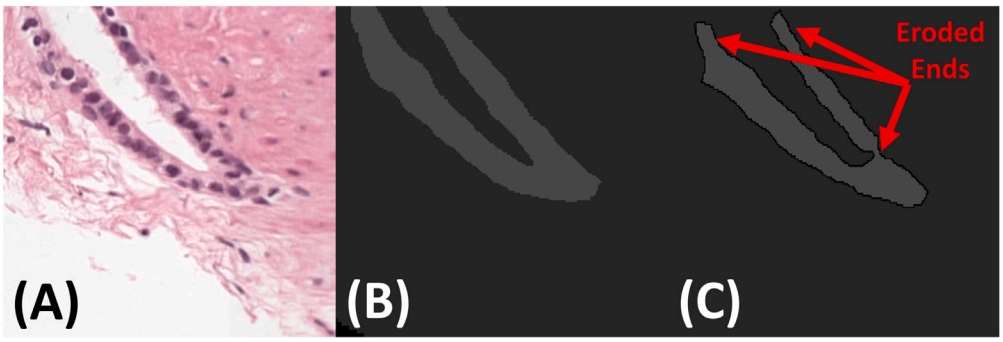

Fig. 5. Limitation of the proposed scheme to produce eroded results. (A) original patch, (B) ground truth, and (C) the segmented tissue regions. These eroded results can be corrected by employing additional morphological post-processing steps.

Table 12
Performance evaluation of the proposed scheme (in terms of μIoU) when the order of learning the prostate tissue extraction tasks is changed.

Dataset	GP-5	GP-4	GP-3	Benign	Stroma	Mean
Louisville	0.6858	0.4904	0.4151	0.3511	0.3626	0.4610
PANDA [60]	0.7091	0.5165	0.4364	0.3702	0.3943	0.4853

Table 13
Clinical verification of the proposed scheme through the statistically significant r_c coefficient (having p-value < 0.05).

Metric	Clinician-1	Clinician-2
r_c	0.9192	0.8984
p-value	3.71×10^{-25}	2.12×10^{-22}

proposed scheme drops significantly for extracting the stroma, benign, and GP-3 cases when the prostate tissue learning order is reversed (see Tables 7, 8, and 12).

Within the clinical settings, it is sometimes more critical to yield detection alarms for the early cancerous cases so that proper and timely treatment can be given to the subject to diminish the PCa spread. Similarly, it is equally important to accurately detect the severe PCa (especially the GP-4 and GP-5 level tissues) to prevent the loss of the subject's life. Therefore, deciding the prostate tissue learning order (on which the proposed scheme must be trained) depends on the clinical needs. But it is worth noting that the proposed approach, to the best of our knowledge, is the first PCa grading scheme (to date) that provides this trade-off and the flexibility to extract prostate tissues independent of their order due to its incremental nature, which can be tamed for any clinical needs to timely and effectively treat PCa cases.

4.2.6. Blind testing

We also conducted another series of blind testing experiments through which we validated the clinical applicability of the proposed scheme by establishing its correlation with the diagnostics of the expert pathologists in the clinical settings. Within these experiments, we randomly passed 60 test WSI samples (from the Louisville and PANDA [60] datasets) to the proposed scheme and to the pathologists for PCa grading and measured the relationship between their prognosis via Pearson correlation coefficient (r_c). From Table 13, we can observe that with both clinicians, the proposed scheme achieved the r_c of 0.9192, and 0.8984, which is also statistically significant as p-value $<$ 0.05. This indicates that the proposed scheme possesses the potential to be deployed within the clinical settings for real-world screening and grading of PCa progression via WSI biopsies.

4.3. Failure cases

Although, the proposed scheme (driven via DRN [9]) is quite robust in extracting the prostate tissues to grade the PCa. Nevertheless, there are some cases in which the performance of the proposed scheme is found to be bounded. The first case is related to the limitation of the proposed scheme to produce eroded results, e.g., see pairs (D, E, F), (M, N, O), (V, W, X), and (Y, Z, AA) in Fig. 3, and also (C) in Fig. 5, which affects the overall performance of the proposed scheme towards accurately grading the progression of PCa. However, this issue can be easily addressed through morphological dilation or an opening-based post-processing mechanism which can be appended at the top of the employed segmentation network (DRN [9]) during the inference stage. Moreover, as the frequency of the GP-4 and GP-5 samples are very less than stroma, benign, and GP-3 regions within both datasets (which is natural), the segmentation model within the proposed scheme gets more biased towards correctly predicting the healthy or early cancerous cases (due to the imbalanced data), producing the second limitation of the proposed scheme. However, this limitation can remedied in the future by incorporating some imbalanced learning strategies such as modulating factors [68], max-margins [69], and gaussian affinity [70, 71], within the proposed $\mathcal{L}_{il}$ loss function.

5. Discussion and conclusion

This paper presents a novel knowledge distillation-based instance segmentation scheme for the extraction of highly correlated prostate tissues from the WSI biopsies in order to grade the PCa progression as per the ISUP standards. To the best of our knowledge, this is the first attempt to blend knowledge distillation and incremental learning with instance segmentation in order to extract prostate tissues which are scarcely distinguishable even by expert pathologists. Moreover, thanks to its incremental nature, the proposed scheme can robustly recognize similarly structured prostate tissues (in any desired order) with few-shot training.

We rigorously tested the proposed scheme under different experimental settings where it outperforms its competitors for both prostate tissue segmentation and PCa severity grading tasks. For example, from Tables 7 and 8, we can see that, in terms of μIoU score, the proposed scheme outperformed its competitors by 2.01% on the Louisville dataset, and 4.45% on the PANDA dataset [60] for extracting the prostate tissues, such as stroma, benign, GP-3, 4, and 5. Similarly, from Table 9, we can observe that the proposed scheme outperformed the state-of-the-art by 1.24%, 10.73%, 0.458%, and 4.25% in terms of F_1 scores on the Louisville dataset towards grading WSI biopsies as ISUP-2, ISUP-3, ISUP-4, and ISUP-5, respectively. Also, from Table 10, we can observe that on the PANDA dataset [60], the proposed scheme produced 2.27%, 1.77%, 11.42%, 1.38%, and 4.29% performance improvements over state-of-the-art in terms of F_1 score for extracting ISUP-1, ISUP-2, ISUP-3, ISUP-4, and ISUP-5 PCa, respectively. Apart from this, under blind testing experiments, the proposed scheme achieved statistical significance correlation scores with the expert pathologists towards grading PCa progression, which signifies

its capacity to be deployed in clinical settings for screening PCa. Furthermore, unlike the competitive methods, the proposed scheme, due to its incremental nature, can learn the prostate tissue extraction tasks periodically. For example, in clinical practice, the proposed scheme can be initially deployed to differentiate between benign and GP-3 tissue regions. Afterward, with the availability of annotated GP-4 and GP-5 prostate samples (even in a smaller quantity), the proposed scheme can learn the new GP-4 and GP-5 extraction tasks while retaining the previously learned knowledge. To perform this action, the proposed scheme would only require a small subset of examples (belonging to both new and old categories) on which it will be constrained via $\mathcal{L}_{il}$ in a few training increments. On the contrary, the same training behavior cannot be achieved for the competitive methods as they are built upon conventional transfer learning and fine-tuning approaches, which require large-scale and well-annotated data at once to learn the GP extraction tasks from the WSI patches. Here, if we mold these models to learn GP extraction tasks in multiple passes. They would still be re-learning the previously learned tasks using the same amount of data in order to maintain their performance [53]. For example, if in the first pass, they use x1 samples to learn t1 tasks. Then in the second pass, they would need [x1, x2] samples to learn [t1, t2] tasks. However, the proposed scheme, in the second pass, would only require around [0.1*x1, 0.1*x2] samples to accurately learn [t1, t2] tasks. Hence, the proposed scheme is practically more adaptable and scalable to be deployed in the clinical practice for prostate tissue extraction and PCa grading tasks as compared to the competitive methods. We also want to highlight here that the proposed scheme has a trade-off between retaining its old knowledge or retaining the new one during incremental training. But this can be tuned through the β hyperparameter that controls the proposed scheme's ability to pay more attention to learning the newly added classes or retaining its prior knowledge. By adjusting the β parameter accurately, we can improve the trade-off of screening the severe cancerous cases or the early ones, which can lead to their timely treatment within clinical settings. This feature also reflects the adaptability of the proposed scheme for real-world deployment, as the proposed scheme can be adjusted easily as per the trend of patients that are more frequently come to the hospital.

We also want to highlight here that, although, the μIoU score within Tables 4, 7, and 8 is relatively lower than 0.5. The main reason for these lower values of μIoU is not the random classification. Instead, it is the extreme class imbalance between stroma, benign, GP-3, 4, and 5 tissues in both datasets. For example, Tables 4, 7, and 8 show that all the models can decently extract the stroma and benign tissues after the first two incremental iterations. However, when they are incrementally trained in the third, fourth, and fifth iterations to recognize GP-3, 4, and 5, respectively, their performance deteriorates at a much higher rate because of the imbalanced sample ratios between stroma, benign, and GP-3, 4, and 5 tissues. Similarly, sometimes, the proposed scheme produces the eroded results, especially on the imbalanced data, as discussed in Section 4.3. Nevertheless, these limitations can be easily addressed in the future by incorporating post-processing measures and imbalanced resistant objective functions within the $\mathcal{L}_{il}$ loss function. Also, in the future, we envisage validating the proposed scheme across other types of cancerous pathologies.

CRediT authorship contribution statement

Taimur Hassan: Conceptualization, Methodology, Software, Data curation, Investigation, Writing – original draft, Writing – review & editing, Visualization. **Muhammad Shafay:** Conceptualization, Methodology, Software, Data curation, Investigation, Writing – original draft, Writing – review & editing, Visualization. **Bilal Hassan:** Conceptualization, Methodology, Software, Data curation, Investigation, Writing – original draft, Writing – review & editing, Visualization. **Muhammad Usman Akram:** Supervision, Writing – review & editing, Project administration. **Ayman ElBaz:** Supervision, Writing – review & editing, Project administration. **Naoufel Werghi:** Supervision, Writing – review & editing, Project administration.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work is supported by research funds from the Terry Fox Foundation, United States of America Ref: I1037, Khalifa University, Ref: CIRA-2019-047, and the Advanced Technology Research Center Program (ASPIRE), United Arab Emirates, Ref: AARE20-279. Apart from this, we would also like to thank the two expert pathologists from Watim Medical & Dental College, Rawalpindi, Pakistan, for thoroughly screening the WSI biopsies under blind testing experiments.

References

[1] American Cancer Society, Key Statistics for Prostate Cancer, American Cancer Society Facts & Figures, 2022.
[2] P. Rawla, Epidemiology of prostate cancer, World J. Oncol. (2019).
[3] M.N. Gurcan, L.E. Boucheron, A. Can, A. Madabhushi, N.M. Rajpoot, B. Yener, Histopathological image analysis: A review, IEEE Rev. Biomed. Eng. 2 (2009) 147–171.
[4] J. Silva-Rodriguez, A. Colomer, M.A. Sales, R. Molina, V. Naranjo, Going deeper through the Gleason scoring scale: An automatic end-to-end system for histology prostate grading and cribriform pattern detection, Comput. Methods Programs Biomed. (2020).
[5] A. Matoso, J.I. Epstein, Defining clinically significant prostate cancer on the basis of pathological findings, Histopathology 74 (1) (2019) 135–145.
[6] J. Silva-Rodriguez, A. Colomer, V. Naranjo, WeGleNet: A weakly-supervised convolutional neural network for the semantic segmentation of Gleason grades in prostate histology images, Comput. Med. Imag. Graph. (2021).
[7] L. Egevad, B. Delahunt, J.R. Srigley, H. Samaratunga, International society of urological pathology (ISUP) grading of prostate cancer–An ISUP consensus on contemporary grading, APMIS (2016).
[8] S. Otálora, N. Marini, H. Müller, M. Atzori, Semi-weakly supervised learning for prostate cancer image classification with teacher-student deep convolutional networks, in: IMIMIC 2020, MIL3ID 2020, LABELS 2020: Interpretable and Annotation-Efficient Learning for Medical Image Computing, 2020.
[9] T. Hassan, B. Hassan, A. El-Baz, N. Werghi, A dilated residual hierarchically fashioned segmentation framework for extracting Gleason tissues and grading prostate cancer from whole slide images, in: IEEE Sensors Applications Symposium (SAS), 2021.
[10] E. Arvaniti, K.S. Fricker, M. Moret, N. Rupp, T. Hermanns, C. Fankhauser, N. Wey, P.J. Wild, J.H. Rueschoff, M. Claassen, Automated Gleason grading of prostate cancer tissue microarrays via deep learning, Sci. Rep. 8 (1) (2018) 1–11.
[11] S. Graham, Q.D. Vu, S.E.A. Raza, A. Azam, Y.W. Tsang, J.T. Kwak, N. Rajpoot, HoVer-Net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images, Med. Image Anal. (2019) 101563.
[12] T. Hassan, S. Javed, A. Mahmood, T. Qaiser, N. Werghi, N. Rajpoot, Nucleus classification in histology images using message passing network, Med. Image Anal. (2022) 102480.
[13] T. Hassan, B. Hassan, M.U. Akram, S. Hashmi, A.H. Taguri, N. Werghi, Incremental cross-domain adaptation for robust retinopathy screening via Bayesian deep learning, IEEE Trans. Instrum. Meas. (2021).
[14] T. Hospedales, A. Antoniou, P. Micaelli, A. Storkey, Meta-learning in neural networks: A survey, IEEE Trans. Pattern Anal. Mach. Intell. (TPAMI) (2021).
[15] S.-A. Rebuffi, A. Kolesnikov, G. Sperl, C.H. Lampert, ICaRL: Incremental classifier and representation learning, in: IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), 2017.
[16] C. Finn, P. Abbeel, S. Levine, Model-agnostic meta-learning for fast adaptation of deep networks, in: International Conference on Learning Representations (ICLR), 2017.
[17] T. Hassan, S. Akçay, M. Bennamoun, S. Khan, N. Werghi, Tensor pooling driven instance segmentation framework for baggage threat recognition, Neural Comput. Appl. (2021).
[18] T. Hassan, M. Shafay, S. Akçay, S. Khan, M. Bennamoun, E. Damiani, N. Werghi, Meta-transfer learning driven tensor-shot detector for the autonomous localization and recognition of concealed baggage threats, Sensors (2020).
[19] Z. Huang, L. Huang, Y. Gong, C. Huang, X. Wang, Mask scoring R-CNN, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 6409–6418.
[20] T. Hassan, N. Werghi, Trainable structure tensors for autonomous baggage threat detection under extreme occlusion, in: Proceedings of the Asian Conference on Computer Vision, 2020.

[21] N.A. Koohbanani, M. Jahanifar, N.Z. Tajadin, N. Rajpoot, NuClick: A deep learning framework for interactive segmentation of microscopic images, in: Medical Image Analysis, 2020.

[22] S. Graham, M. Jahanifar, A. Azam, M. Nimir, Y.-W. Tsang, K. Dodd, E. Hero, H. Sahota, A. Tank, K. Benes, et al., Lizard: A large-scale dataset for colonic nuclear instance segmentation and classification, in: Proceedings of the IEEE/CVF International Conference on Computer Vision Workshop (ICCVW), 2021, pp. 684–693.

[23] J. Gamper, N.A. Koohbanani, K. Benes, A. Khuram, N. Rajpoot, PanNuke: An open pan-cancer histology dataset for nuclei instance segmentation and classification, in: European Congress on Digital Pathology (ECDP), Springer, 2019, pp. 11–19.

[24] J. Gamper, N.A. Koohbanani, K. Benes, S. Graham, M. Jahanifar, S.A. Khurram, A. Azam, K. Hewitt, N. Rajpoot, PanNuke dataset extension, insights and baselines, 2020, arXiv preprint arXiv:2003.10778.

[25] T. Hassan, A. Ahmed, B. Hassan, M. Shafay, A. ElBaz, J. Dias, N. Werghi, Incremental instance segmentation for the Gleason tissues driven prostate cancer prognosis, in: 2nd IEEE International Conference on Digital Futures and Transformative Technologies (ICoDT2), 2022.

[26] S. Akbar, T. Hassan, M.U. Akram, U.U. Yasin, I. Basit, AVRDB: Annotated dataset for vessel segmentation and calculation of arteriovenous ratio, in: International Conference on Image Processing, Computer Vision, and Pattern Recognition (IPCV), 2017.

[27] Y.F.A. Gaus, N. Bhowmik, S. Akcay, T.P. Breckon, Evaluating the transferability and adversarial discrimination of convolutional neural networks for threat object detection and classification within X-Ray security imagery, in: 2019 18th IEEE International Conference on Machine Learning and Applications (ICMLA), 2019.

[28] R. Cao, A.M. Bajgiran, S.A. Mirak, S. Shakeri, X. Zhong, D. Enzmann, S. Raman, K. Sung, Joint prostate cancer detection and gleason score prediction in mp-MRI via FocalNet, IEEE Trans. Med. Imaging 38 (11) (2019) 2496–2506.

[29] A. Nasim, T. Hassan, M.U. Akram, B. Hassan, M.A. Shami, Automated identification of colorectal glands morphology from benign images, in: Proceedings of the International Conference on Image Processing, Computer Vision, and Pattern Recognition (IPCV), 2017, pp. 147–152.

[30] T. Hassan, M.U. Akram, M.F. Masood, U. Yasin, BIOMISA retinal image database for macular and ocular syndromes, in: International Conference Image Analysis and Recognition, 2018.

[31] T. Hassan, A. Usman, M.U. Akram, M.F. Masood, U. Yasin, Deep learning based automated extraction of intra-retinal layers for analyzing retinal abnormalities, in: IEEE 20th International Conference on E-Health Networking, Applications and Services, 2018.

[32] Z. Wang, C. Liu, D. Cheng, L. Wang, X. Yang, K.-T. Cheng, Automated detection of clinically significant prostate cancer in mp-MRI images based on an end-to-end deep neural network, IEEE Trans. Med. Imaging 37 (5) (2018) 1127–1139.

[33] X. Wang, W. Yang, J. Weinreb, J. Han, Q. Li, X. Kong, Y. Yan, Z. Ke, B. Luo, T. Liu, L. Wang, Searching for prostate cancer by fully automated magnetic resonance imaging classification: deep learning versus non-deep learning, Sci. Rep. 7 (1) (2017) 1–8.

[34] K. Nagpal, D. Foote, Y. Liu, P.-H.C. Chen, E. Wulczyn, F. Tan, N. Olson, J.L. Smith, A. Mohtashamian, J.H. Wren, G.S. Corrado, R. MacDonald, L.H. Peng, M.B. Amin, A.J. Evans, A.R. Sangoi, C.H. Mermel, J.D. Hipp, M.C. Stumpe, Development and validation of a deep learning algorithm for improving Gleason scoring of prostate cancer, NPJ Digital Med. 2 (1) (2019) 1–10.

[35] A.G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, H. Adam, MobileNets: Efficient convolutional neural networks for mobile vision applications, 2017, arXiv preprint arXiv:1704.04861.

[36] N. Marini, S. Otalora, H. Muller, M. Atzori, Semi-supervised training of deep convolutional neural networks with heterogeneous data and few local annotations: An experiment on prostate histopathology image classification, Med. Image Anal. (2021).

[37] S. Xie, R. Girshick, P. Dollar, Z. Tu, K. He, Aggregated residual transformations for deep neural networks, in: IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), 2017.

[38] G. Huang, Z. Liu, L. van der Maaten, K.Q. Weinberger, Densely connected convolutional networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 4700–4708.

[39] W. Bulten, et al., Automated deep-learning system for Gleason grading of prostate cancer using biopsies: a diagnostic study, Lancet Oncol. (2020).

[40] O. Ronneberger, P. Fischer, T. Brox, U-Net: Convolutional networks for biomedical image segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), Springer, 2015, pp. 234–241.

[41] P. Strom, et al., Artificial intelligence for diagnosis and grading of prostate cancer in biopsies: a population-based, diagnostic study, Lancet Oncol. (2020).

[42] W. Shuai, J. Li, Few-shot learning with collateral location coding and single-key global spatial attention for medical image classification, Electronics (2022).

[43] J. Deuschel, D. Firmbach, C.I. Geppert, M. Eckstein, A. Hartmann, V. Bruns, P. Kuritcyn, J. Dexl, D. Hartmann, D. Perrin, T. Wittenberg, M. Benz, Multi-prototype few-shot learning in histopathology, in: IEEE International Conference on Computer Vision Workshop (ICCVW), 2021.

[44] S.F.H. Naqvi, S. Ayubi, A. Nasim, Z. Zafar, Automated gland segmentation leading to cancer detection for colorectal biopsy images, in: Future of Information and Communication Conference, Springer, 2019, pp. 75–83.

[45] K. He, G. Gkioxari, P. Dollár, R. Girshick, Mask R-CNN, in: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2017, pp. 2961–2969.

[46] B. Hassan, S. Qin, T. Hassan, R. Ahmed, N. Werghi, Joint segmentation and quantification of chorioretinal biomarkers in optical coherence tomography scans: A deep learning approach, IEEE Trans. Instrum. Meas. (2021).

[47] D. Bolya, C. Zhou, F. Xiao, Y.J. Lee, YOLACT: Real-time instance segmentation, in: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2019, pp. 9157–9166.

[48] T. Hassan, M. Bettayeb, S. Akçay, S. Khan, M. Bennamoun, N. Werghi, Detecting prohibited items in X-ray images: A contour proposal learning approach, in: 27th IEEE International Conference on Image Processing (ICIP), 2020.

[49] Q. Meng, S. Shin'ichi, ADINet: Attribute driven incremental network for retinal image classification, in: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020.

[50] C. Bucilua, R. Caruana, A. Niculescu-Mizil, Model compression, in: Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2006.

[51] G. Hinton, O. Vinyals, J. Dean, Distilling the knowledge in a neural network, in: NIPS Deep Learning Workshop, 2014.

[52] A. Chaudhry, M. Ranzato, M. Rohrbach, M. Elhoseiny, Efficient lifelong learning with A-GEM, in: International Conference on Learning Representations (ICLR), 2019.

[53] M. Sirshar, T. Hassan, M.U. Akram, S.A. Khan, An incremental learning approach to automatically recognize pulmonary diseases from the multi-vendor chest radiographs, Comput. Biol. Med. 134 (2021) 104435.

[54] J. Zhang, J. Zhang, S. Ghosh, D. Li, S. Tasci, L. Heck, H. Zhang, C.-C.J. Kuo, Class-incremental learning via deep model consolidation, in: IEEE Winter Conference on Applications of Computer Vision (WACV), 2020.

[55] Z. Chen, J. Ge, H. Zhan, S. Huang, D. Wang, Pareto self-supervised training for few-shot learning, in: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021.

[56] Z. Chen, B. Liu, Continual learning and catastrophic forgetting, in: Lifelong Machine Learning, Morgan & Claypool Publishers, 2018.

[57] Y. Tian, D. Krishnan, P. Isola, Contrastive representation distillation, in: International Conference on Learning Representations (ICLR), 2020.

[58] A.M. Khan, T. Hassan, M.U. Akram, N.S. Alghamdi, N. Werghi, Continual learning objective for analyzing complex knowledge representations, Sensors (2022).

[59] T. Hassan, S. Akcay, M. Bennamoun, S. Khan, N. Werghi, A novel incremental learning driven instance segmentation framework to recognize highly cluttered instances of the contraband items, IEEE Trans. Syst. Man Cybern. Syst. (2021).

[60] W. Bulten, et al., Artificial intelligence for diagnosis and Gleason grading of prostate cancer: the PANDA challenge, Nature Med. (2022).

[61] M.D. Zeiler, ADADELTA: An adaptive learning rate method, 2012, arXiv preprint arXiv:1212.5701.

[62] T. Hassan, M.U. Akram, N. Werghi, M.N. Nazir, RAG-FW: A hybrid convolutional framework for the automated extraction of retinal lesions and lesion-influenced grading of human retinal pathology, IEEE J. Biomed. Health Inf. 25 (1) (2021) 108–120.

[63] H. Zhao, J. Shi, X. Qi, X. Wang, J. Jia, Pyramid scene parsing network, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 2881–2890.

[64] J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 3431–3440.

[65] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770–778.

[66] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, 2014, arXiv preprint arXiv:1409.1556.

[67] L. Wang, D. Li, Y. Zhu, L. Tian, Y. Shan, Dual super-resolution learning for semantic segmentation, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 3774–3783.

[68] T.Y. Lin, P. Goyal, R. Girshick, K. He, P. Dollar, Focal loss for dense object detection, in: IEEE/CVF International Conference on Computer Vision (ICCV), 2017.

[69] A. Ahmed, D. Velayudhan, T. Hassan, B. Hassan, J. Dias, N. Werghi, Baggage threat detection under extreme class imbalance, in: 2nd IEEE International Conference on Digital Futures and Transformative Technologies (ICoDT2), 2022.

[70] M. Hayat, S. Khan, S.W. Zamir, J. Shen, L. Shao, Gaussian affinity for max-margin class imbalanced learning, in: IEEE/CVF International Conference on Computer Vision (ICCV), 2019.

[71] A. Ahmed, A. Obeid, D. Velayudhan, T. Hassan, E. Damiani, N. Werghi, Balanced affinity loss for highly imbalanced baggage threat contour-driven instance segmentation, in: IEEE International Conference on Image Processing (ICIP), October, 2022.