

Simultaneous Nuclear Instance and Layer Segmentation in Oral Epithelial Dysplasia

Adam J. Shephard^{1*}, Simon Graham¹, R.M. Saad Bashir¹, Mostafa Jahanifar¹,
 Hanya Mahmood², Syed Ali Khurram², Nasir M. Rajpoot¹
¹Department of Computer Science, University of Warwick, Coventry, UK
²School of Clinical Dentistry, University of Sheffield, Sheffield, UK

adam.shephard@warwick.ac.uk

Abstract

Oral epithelial dysplasia (OED) is a pre-malignant histopathological diagnosis given to lesions of the oral cavity. Predicting OED grade or whether a case will transition to malignancy is critical for early detection and appropriate treatment. OED typically begins in the lower third of the epithelium before progressing upwards with grade severity, thus we have suggested that segmenting intra-epithelial layers, in addition to individual nuclei, may enable researchers to evaluate important layer-specific morphological features for grade/malignancy prediction. We present HoVer-Net+, a deep learning framework to simultaneously segment (and classify) nuclei and (intra-)epithelial layers in H&E stained slides from OED cases. The proposed architecture consists of an encoder branch and four decoder branches for simultaneous instance segmentation of nuclei and semantic segmentation of the epithelial layers. We show that the proposed model achieves the state-of-the-art (SOTA) performance in both tasks, with no additional costs when compared to previous SOTA methods for each task. To the best of our knowledge, ours is the first method for simultaneous nuclear instance segmentation and semantic tissue segmentation, with potential for use in computational pathology for other similar simultaneous tasks and for future studies into malignancy prediction.

1. Introduction

Head and neck cancer is among the top 10 most common cancers worldwide, with an estimated incidence of 150,000 new cases every year in Europe alone [6]. These cancers are usually detected late, with 60% of new cases presenting with an advanced stage at diagnosis. Prognosis for advanced stages is poor, having a five-year survival rate of just 40%. This increases drastically with early diagnosis to 80-90% [6]. Head and neck cancers encompass a large group of

cancers; however, oral squamous cell carcinomas (OSCCs) of the oral cavity are some of the most common.

Clinically, oral cancers often present as white or red lesions of the oral mucosa. Such lesions are typically preceded by a pre-cancerous state, a group of lesions termed oral potentially malignant disorders (OPMDs) including homogeneous/non-homogeneous leucoplakia (white/white-red speckled lesions) or erythroplakia (red lesions) [26]. Biopsies can be taken of these lesions allowing detailed microscopic examination by pathologists to determine the presence or absence of pre-cancer (oral epithelial dysplasia/OED) or cancer. Lesions of the oral mucosa exhibiting dysplasia are statistically more likely to transition into OSCC than non-dysplastic lesions [16]. Other risk factors of malignancy transition include sex, site/type of lesion, and habits (e.g. tobacco and alcohol consumption) [26].

OED lesions can be diagnosed and graded based on a variety of different classification systems, with the criteria set out by the World Health Organisation (WHO) in 2017 [27] being considered the gold standard, which grades OED lesions as mild, moderate or severely dysplastic based on cytological and architectural features [27]. Although severely dysplastic cases are most likely to become cancerous, progression has also been reported in OED with lower grades (i.e. mild and moderate). This could be attributed to the fact that OED grading is subjective with evidence of both inter- and intra-rater variability [26, 13, 23]. This lack of reproducibility may lead to over-treatment or in most cases insufficient treatment. This is critical as appropriate treatment and transformation prediction (based on correct grading) can stop a cancer from developing or allow treatment at an early stage conferring better prognosis.

Recent advances in digital pathology have allowed the digitisation of histology slides as whole slide images (WSIs) using high-resolution digital scanners. This has also led to a significant growth in the field of computational pathology [3]. The parallel advances of new deep learn-

ing techniques have complimented both pathology and radiology, allowing the automation of pipelines and displaying the promise of deep learning for predicting patient outcomes [3, 28]. In pathology, deep learning has been used to automatically segment the epithelium within a range of histology images (*e.g.* oral, cervical, prostate) [25, 4, 2] and to further segment and classify individual nuclei within WSIs [9, 20]. Within OED, it has been used to segment the epithelium into sub-regions: the lower basal layer, the middle epithelial layer and the superior keratin layer [2]. Initial OED changes typically manifest in the lower third of the epithelium – encompassing the basal layer – progressing upwards into the middle and upper thirds of the epithelium as the grade worsens. Thus, it has been proposed that segmentation of the different epithelial layers may allow the generation of layer-specific features (*e.g.* layer widths) that may aid in OED grading and in predicting malignant progression [2]. Deep learning tools therefore provide a potential avenue for reducing grading variability, whilst ensuring cross-site consistency in informing treatment decisions [18].

In this work, we suggest that the segmentation of both the individual nuclei and the three layers of the epithelium are important steps in any downstream analyses of dysplasia in the epithelium, not necessarily confined to oral epithelium. We hypothesise that these tasks are complementary and may aid each individual segmentation/classification task. However, manual segmentation of individual nuclei in WSIs is time-consuming, thus collecting such annotations at a large-scale provides a unique challenge. A range of automatic deep learning techniques have been proposed for this task [9, 20, 21]; however, most methods struggle to differentiate nuclei in crowded scenes or in new images. Semi-automated deep learning techniques [1] have been proposed to combat this problem, by manually providing a seed for each nucleus. Due to the complexity in generating ground truth (GT) nuclear annotations, we propose using a semi-automated multi-stage deep learning framework with manual refinement to generate our GT nuclei annotations. Finally, we present a deep learning framework, *HoVer-Net+*, for simultaneous nuclear instance segmentation (and classification) and semantic segmentation of the layers of the epithelium based on Haematoxylin and Eosin (H&E) stained histopathology slides of the oral mucosa in OED. We believe this to be the first approach to combine both nuclear and layer segmentation in a single pipeline.

2. Related Work

2.1. Nuclear Instance Segmentation/Classification

Multiple approaches have been proposed for the segmentation of nuclear instances on WSIs, using traditional machine learning techniques such as watershed-based techniques, active contours and pixel classification [8, 29, 19].

In recent years, deep learning methods have proven to be most common, due to their popularity and superior performance achieved throughout the computer vision literature for segmentation [15, 22, 17]. Deep learning models extract a hierarchy of features that are used to inform the final segmentation. U-Net is a popular approach consisting of an encoder and a decoder with multiple skip connections between layers [22]. It has been used for semantic segmentation in a plethora of different tasks, and has since been adapted to segment nuclear instances by the incorporation of a weighted loss function. However, a criticism of U-Net is its inability to efficiently disentangle touching or overlapping nuclei. Micro-Net is another deep learning network that has been used for the segmentation of nuclei in histology images [21]. This model combines multiple image resolutions to achieve superior results.

More recent methods have attempted to improve nuclear instance segmentation performance by treating the task as a regression problem. DIST formulates the problem as a regression task of the distance map [20], where the distance map gives the distance of a nuclear pixel to the centre of mass of the nuclear instance. In [9], the authors incorporate a novel HoVer branch in their encoder-decoder framework that estimates the horizontal and vertical distances of each nuclear pixel from their respective nucleus' centre of mass. This results in large gradients at the shared boundaries of touching instances, thus allowing more efficient nuclear edge discrimination. HoVer-Net comprises of a single encoder branch and three decoder branches for simultaneous nuclear instance segmentation and classification and has achieved state-of-the-art (SOTA) performance on multiple H&E stained histology image datasets (*e.g.* Kumar [14] and CoNSeP [9] *etc.*). This demonstrates its transferability to different histology images/pathologies. In this work, HoVer-Net is employed for nuclear segmentation and classification. We further hypothesise that the incorporation of an additional decoder branch for layer segmentation may aid in improving both nuclear and layer classification.

In [1], the authors proposed a supervised framework NuClick that can interactively segment nuclear instances provided only with a single point (click) inside each nucleus. The encoder-decoder design of the NuClick architecture, benefiting from multi-scale convolutional blocks and incorporating a weighted loss function, ensures high-quality segmentation of nuclei at various scales even in the cases of touching instances/crowded scenes. The generalisability of this method, due to the provided guiding signals (clicks), enables the users to collect nuclear instance segmentations for a large set of data rapidly and efficiently [1]; however, we suggest that this may be an impractical way of generating instance segmentations for large cohorts of WSIs. In our work, we have therefore resorted to first generating a reliable dataset of patches in which nuclear instances are an-

notated with NuClick, and then fine-tuning HoVer-Net on that dataset to make it ready for inference at the WSI level.

2.2. Layer Segmentation

Deep learning approaches have been used extensively for semantic segmentation, with models such as the fully convolutional neural network [17], U-Net [22] and DeepLabv3 [5] being considered the SOTA. Multiple works have used these methods to automatically segment the epithelium in histology images. In [4], the authors used U-Net to successfully segment the epithelium in IHC and H&E stained images. In [25], the authors presented EpithNet, a variant of VGG [24], to segment epithelium in cervical histology images. This method incorporated a novel step approximating edge contours with Bezier curves to outperform U-Net. In [2], DeepLabv3 was used to further segment oral epithelium into its constituent layers: basal layer, (middle) epithelial layer, (superficial) keratin layer, in order to generate layer-specific morphological features for predicting OED grade.

3. Methodology

3.1. Network Architecture

The network proposed in this study is based on HoVer-Net [9], which performs nuclear instance segmentation and classification. HoVer-Net is a deep learning framework that consists of an encoder branch, and three decoder branches. The encoder branch is based on the pre-activated residual network with 50 layers [10], consisting of four consecutive pre-activated residual units. The decoder branches of the network perform nearest-neighbour up-sampling, through up-sampling operations and densely connected units [11], in conjunction with skip connections from the encoder [22].

HoVer-Net consists of three decoder branches: the nuclear pixel (NP) branch, the horizontal and vertical distances (HoVer) branch, and the nuclear classification (NC) branch. The NP branch performs binary classification, thus predicting whether a pixel is foreground (nucleus) or background. The HoVer branch estimates the horizontal and vertical distances of each pixel in a nucleus from the respective nucleus' centre of mass, resulting in large gradients at shared nuclear boundaries. Finally, the NC branch predicts the type of nucleus for each pixel. Together, the NP and HoVer branches perform nuclear instance segmentation, whilst the NC branch performs nuclear classification. In this work, we have adapted HoVer-Net to include a fourth decoder branch that acts as the layer segmentation (LS) branch, in an adaption we have named HoVer-Net+. Within this work, we consider a larger input size (256×256) than was used in the original HoVer-Net to speed up inference. Figure 1 shows an overview of the model architecture.

3.2. Loss Function

The same overall loss function as utilised in [9] was used, with the addition of two extra terms relating to the LS branch. The proposed method consists of four decoder branches, each containing weights that are jointly optimised with the following overall loss \mathcal{L} :

$$\begin{aligned} \mathcal{L} = & \mathcal{L}_a \lambda_a + \mathcal{L}_b \lambda_b + \mathcal{L}_c \lambda_c + \mathcal{L}_d \lambda_d \\ & + \mathcal{L}_e \lambda_e + \mathcal{L}_f \lambda_f + \mathcal{L}_g \lambda_g + \mathcal{L}_h \lambda_h, \end{aligned} \quad (1)$$

where \mathcal{L}_a and \mathcal{L}_b are regression losses with respect to the output of the HoVer branch, \mathcal{L}_c and \mathcal{L}_d are losses with respect to the output at the NP branch, \mathcal{L}_e and \mathcal{L}_f are losses with respect to the output at the NC branch, and \mathcal{L}_g and \mathcal{L}_h are losses with respect to the output of the LS branch, and λ_a to λ_h are weights assigned to each loss function. We assigned λ_b a value of 2 in line with [9], whilst we found values of 1 best for the remaining weights.

At the output of the HoVer branch we calculated the mean squared error between the predicted horizontal and vertical distances and the GT (\mathcal{L}_a). We also calculated the mean squared error between the horizontal and vertical gradients of the horizontal and vertical maps respectively, and the corresponding gradients of the GT (\mathcal{L}_b). At the output of the NP, NC and LS branches, we calculated the cross-entropy loss (\mathcal{L}_c , \mathcal{L}_e , \mathcal{L}_g) and the dice loss (\mathcal{L}_d , \mathcal{L}_f , \mathcal{L}_h). Consequently, each branch consists of two loss terms summed together. For the full formulation of the loss terms used in this work, see [9].

3.3. Post-Processing

The three original branches of HoVer-Net were used to segment and classify nuclei as either epithelial or other (*e.g.* connective/immune cells) nuclei. The same post-processing steps for nuclear instance segmentation and classification were used as in [9]. This generated instance segmentations using the HoVer and NP branches, combined with a marker-based watershed. Majority voting was used to determine the class of each nucleus via the NC branch. The output layer segmentation of the LS branch was post-processed via the morphological opening and closing of layers, along with the removal of small objects and holes. Any nuclei that were classified as epithelial via the original HoVer-Net architecture were then further differentiated by the associated predicted layer that their centroid was located in. Thus, if a nucleus was predicted to be an epithelial nucleus, then that nucleus would be further sub-classified as a basal, epithelial or keratin nucleus depending on the associated predicted layer of the LS branch. This additionally allows for other nucleus types, such as lymphocytes that are known to penetrate the epithelial layer in OED, to be found in the epithelial layer. Note, the LS branch was used to assign layers to each epithelial nuclei over the NC branch, to ensure a smooth

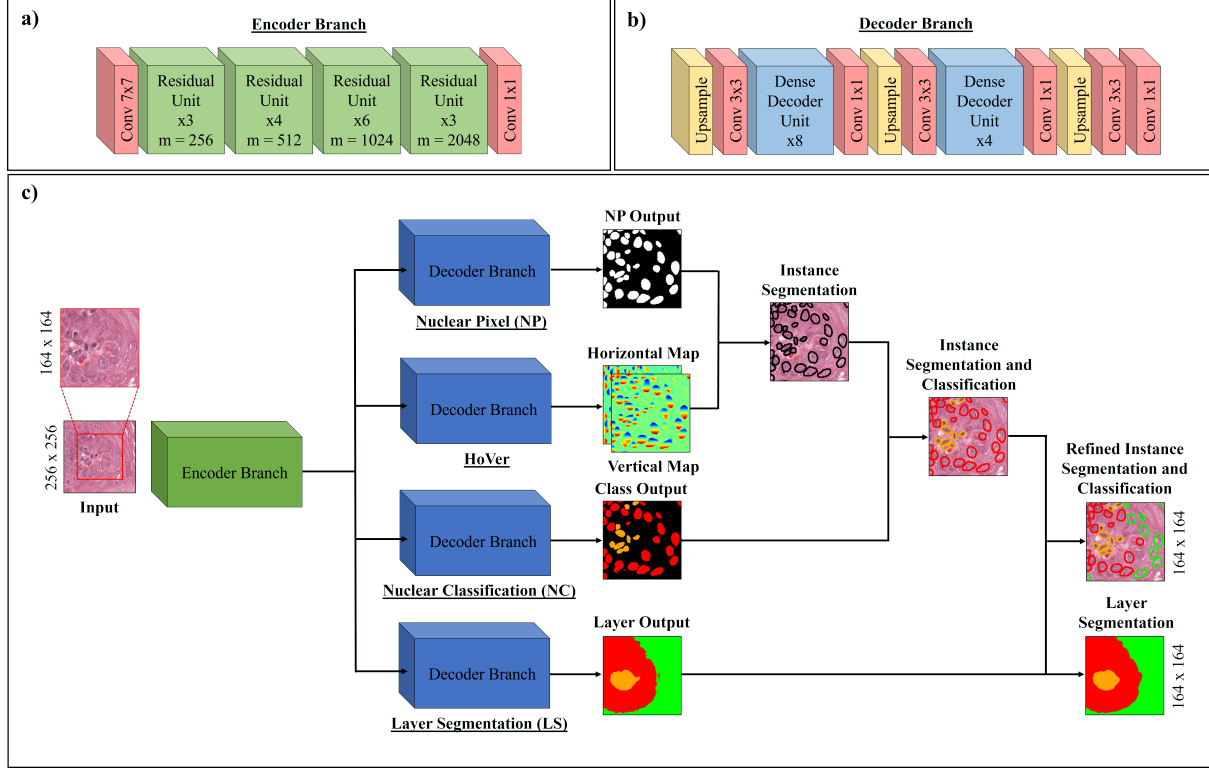


Figure 1. The architecture of HoVer-Net+, including the post-processing steps taken to achieve the final output, showing a) the encoder branch, b) the decoder branch and c) the overall architecture, where m denotes the number of feature maps within the residual units.

boundary between nuclei of different layers. HoVer-Net+ produces two sets of outputs: nuclear classification maps and layer segmentation maps. Each of these maps consists of 5 classes: background, other, basal, epithelium, keratin. Figure 1 displays example outputs showing other tissue as orange, basal as red, and epithelium as green. No background or keratin classes are present in the example images.

3.4. Evaluation Metrics

For the evaluation of nuclear instance segmentation and classification we have used the same metrics as used in [9]. Therefore, nuclear instance segmentation performance was assessed based on Panoptic Quality (PQ), a score originally proposed by [12], unifying the results of instance segmentation and semantic segmentation. Although PQ provides a good overall metric for comparing models, it is beneficial to interpret this in conjunction with its constituent elements: detection quality (DQ; or F1-score) and segmentation quality (SQ). We further report a Dice score comparing all segmented nuclei against the background, and an aggregated Jaccard index (AJI) [14]. PQ and AJI compare each pair of predicted and GT nuclei, classifying a matched pair (true positive) as when their intersection over union is greater than 0.5. For the evaluation of nuclear classification per-

formance, we report the average value over all images for: the F1-score for detection F_d (*i.e.* all nuclear types); and F1-score for classification F_c for each of the individual nuclei types (*e.g.* other: F_c^o , basal: F_c^b , epithelium: F_c^e). We refer the reader to [9] for an in-depth description of these metrics. Finally, for layer segmentation, we report aggregated precision, recall, accuracy and F1-score. We further provide a breakdown of the F1-score by each class studied (*e.g.* F1-score for basal *vs.* rest).

4. Preliminary Experiment

4.1. Data

The study data consisted of 43 WSIs from H&E stained sections of the oral mucosa attained at the University of Sheffield, UK, and scanned at $20\times$ (0.4952 microns-per-pixel) using an Aperio CS2 scanner. Of these cases, 38 had OED and 5 were healthy controls. An expert oral and maxillofacial pathologist (SAK) graded the OED cases as either mild, moderate, or severe according to the WHO 2017 classification system [27]. Of the 38 cases, 14 were rated as mild, 13 as moderate, and 11 as severe. Analysis of the data was conducted under ethical approval by the NHS Health Research Authority West Midlands titled “Artificial Intelligence to grade oral epithelial dysplasia and

predict malignant transformation” (Ref: 18/WM/0335). The pathologist further annotated the epithelium into three layers (basal layer, epithelium, superficial (keratin) layer) using the Automated Slide Analysis Platform software (<https://computationalpathologygroup.github.io/ASAP/>). Data were split into training/validation/testing sets with a ratio of 29/7/7, stratified by OED grade (including normal cases). Tissue masks were generated for each WSI via Otsu thresholding and the further removal of small holes and objects. Layer masks were generated for each WSI by combining the layer segmentations with the tissue mask.

4.2. Nuclear Ground Truth Generation

4.2.1 NuClick

The manual segmentation of individual nuclei in WSIs is time-consuming and subject to inter/intra-rater variability, thus nuclear instance maps were instead generated for a subset of cases ($n = 20$) using NuClick [1]. NuClick takes nuclear instance point annotations as inputs and outputs nuclear instance segmentations that are superior to those produced by fully automated approaches (specifically with touching nuclei) [1]. Thus, NuClick is suitable for generating GT annotations. The expert pathologist drew large regions of interest (ROIs) for these 20 cases and assigned a single point to each nucleus. Each instance was assigned a class as either epithelial or other nuclei. Following processing, all NuClick annotations were visually inspected and edited using the GIMP software (<https://www.gimp.org/>) where necessary. This was generally only necessary in the granular layer of the epithelium. This resulted in 1,139 patches (256×256 at $20\times$) containing nearly 50,000 nuclei.

4.2.2 HoVer-Net

The refined NuClick annotations were further used to fine-tune the original HoVer-Net (pre-trained on the PanNuke dataset [7]) to segment and classify nuclei in OED H&E images as either epithelial or other nuclei. This model was trained/validated on the manually-refined NuClick patches from the training and validation sets alone (12 and 4 ROIs respectively), to ensure an unbiased result on the test data in the proceeding experiments. The NuClick test set (4 ROIs from 4 cases) was held back for final evaluation in the main experiments. HoVer-Net was implemented with the PyTorch library on a workstation equipped with two NVIDIA RTX 5000 GPUs. It was trained in two phases, with only the decoder branches trained for the first 50 epochs, and with all branches in the next 50 epochs. A batch size of 8 and 4 on each GPU were used across these two phases respectively. The Adam optimizer was used with a learning rate that decayed from 10^{-4} to 10^{-5} after 25 epochs in each phase. The following quality metrics were attained for validation on the

patches obtained from the NuClick ROIs for instance segmentation: Dice = 0.715, AJI = 0.669, DQ = 0.798, SQ = 0.710, PQ = 0.562. For classification we obtained: $F_d = 0.838$, $F_c^o = 0.821$, $F_c^e = 0.781$. These results demonstrate the success of our NuClick/HoVer-Net pipeline for gaining OED specific GT segmentations.

Following fine-tuning, the trained HoVer-Net model was used for inference on all 43 WSIs in order to produce nuclear instance segmentations for each nucleus in every WSI. We further implemented a quality control step here, checking and manually refining each nuclear class where necessary, to ensure that the classes assigned to each nucleus in all 43 WSIs were correct. The nuclear masks were superimposed on the layer masks for each WSI, to further differentiate epithelial nuclei by intra-epithelial layer (basal, epithelial, keratin). This resulted in nuclear classes corresponding to basal, epithelial, keratin and other nuclei. The other nuclei mainly included nuclei of connective tissue cells, endothelial cells and immune cells. All WSIs were then divided into non-overlapping 256×256 images (252×252 and 268×268 patches for comparative experiments). A balanced dataset was ensured by following the below criteria on patching: a) patch contains some epithelial tissue (any of the three layers); b) patch is greater than 10% tissue (*i.e.* mean intensity of binary tissue map is ≥ 0.1)

5. Experiments and Results

5.1. Data

The preliminary experiment provided GT nuclear and layer segmentations (and classes) for the 43 WSIs used in this study. This resulted in 17,086 patches for training, 3,358 for validation and 4,188 for testing. The efficacy of the proposed HoVer-Net+ model at simultaneous nuclear segmentation (and classification) and layer segmentation, was tested using the data described in the preliminary experiment in a series of further experiments.

5.2. Implementation

HoVer-Net+ was tested and compared to other models for nuclear instance segmentation and classification, and layer segmentation. On testing layer segmentation, HoVer-Net+ was further adapted for comparison, to perform layer segmentation only. This trial is labelled HoVer-Net+ (LS). This architecture used a single decoder branch (LS), and aimed to establish whether simultaneous nuclear and layer segmentation improves performance by the incorporation of other histological information. All HoVer-Net(+) models were trained in two phases, with only the decoder branches trained for the first 20 epochs, and with all branches in the next 20 epochs. A batch size of 8 and 4 on each GPU were used across these two phases respectively. The Adam optimizer was used with a learning rate that decayed initially

Model	Dice	AJI	DQ	SQ	PQ	F _d	F _c ^o	F _c ^b	F _c ^e
DIST	0.788	0.666	0.745	0.735	0.564	0.868	0.747	0.583	0.593
U-Net	0.752	0.627	0.746	0.679	0.519	0.881	0.761	0.599	0.598
Micro-Net	0.855	0.787	0.838	0.841	0.715	0.892	0.777	0.627	0.582
HoVer-Net	0.848	0.775	0.823	0.866	0.725	0.890	0.761	0.600	0.580
HoVer-Net+	0.849	0.778	0.825	0.867	0.727	0.890	0.783	0.590	0.597

Table 1. Comparative experiments for nuclear instance segmentation and classification on the OED dataset.

Model	Dice	AJI	DQ	SQ	PQ
DIST	0.801	0.647	0.775	0.735	0.570
U-Net	0.769	0.612	0.755	0.693	0.525
Micro-Net	0.829	0.684	0.840	0.793	0.666
HoVer-Net	0.836	0.691	0.860	0.801	0.689
HoVer-Net+	0.839	0.694	0.861	0.801	0.690

Table 2. Comparative experiments for nuclear instance segmentation on the hold-out manually-refined NuClick OED test set.

from 10^{-4} to 10^{-5} after 10 epochs in each phase. Model hyper-parameters were found through optimisation on the training/validation set. Further hold-out testing took place after retraining the models on the combined training and validation set. Comparative experiments were performed to test HoVer-Net+’s performance against other benchmark networks for nuclear classification and semantic layer segmentation. For nuclear classification, HoVer-Net+ was tested against U-Net [22], DIST [20] and Micro-Net [21], where models were trained/tested based on their default parameters, however, where necessary they were adapted to perform classification as described in [9]. HoVer-Net+ was compared against DeepLabv3 [5] and U-Net [22] for the segmentation of layers, owing to the successes of these models in achieving SOTA results for semantic segmentation. During the training of all models, patches were augmented using flip, rotation, Gaussian blur and median blur. All models were confirmed for convergence after 40 epochs.

5.3. Nuclear Instance Segmentation

We evaluated HoVer-Net+ for nuclear instance segmentation on the entire OED cohort testing set, by comparing it to the other SOTA approaches. The results for instance segmentation are included in Table 1; however, these results will be biased towards the HoVer-Net architectures as HoVer-Net was used to generate the GT instance segmentations, with only the classes manually-refined. Therefore, we further display the results of all models on the hold-out manually-refined NuClick testing set, consisting of four large ROIs from four subjects (see Table 2). This test set was not evaluated in the preliminary experiment. Both HoVer-Net and HoVer-Net+ achieved the best performance across both test sets, particularly in relation to PQ. How-

ever, Micro-Net was only marginally worse in both scenarios, dissimilar to U-Net and DIST that have a PQ between 0.16 and 0.21 lower than the HoVer-Net models.

5.4. Nuclear Classification

Our proposed model classified nuclei as background, other, basal epithelium, epithelium or keratin layer epithelium. Since few nuclei are generally found in the keratin layer, a metric has not been included for this class. The HoVer-Net(+) models and Micro-Net, have again achieved the highest performance for the segmentation of all nuclei (see Table 1), with only minimal differences. However, for classification, HoVer-Net+ has attained the best performance for other nuclei, whilst Micro-Net was best for basal epithelial nuclei. U-Net achieved the highest score for the classification of epithelial nuclei, having a score marginally better than HoVer-Net+. Overall, there was little difference in the performances of the HoVer-Net(+) architectures and Micro-Net for the classification of nuclei, but all performed better than U-Net and DIST. Therefore, simultaneous nuclear and layer segmentation has resulted in a comparative performance with other SOTA approaches, demonstrating the efficiency of our method.

Examination of the quality of each model’s segmentation performance compared to the GT in Figure 2 shows how most models have effectively classified each type of nuclei, including other nuclei infiltrating the epithelium. However, it can be seen that U-Net and DIST have regularly segmented small areas of false positives in cells in the granular layer of the epithelium, where no nuclei are visible (see blue arrows). The HoVer-Net(+) and Micro-Net architectures have avoided this. All models have areas of false negatives for some faint nuclei in the epithelium, highlighted by the yellow boxes in Figure 2.

5.5. Layer Segmentation

We further compared HoVer-Net+ to other SOTA semantic segmentation methods: DeepLabv3 and U-Net, for the segmentation of the epithelium in OED (Table 3). HoVer-Net+ achieved superior performance for layer segmentation when performing simultaneous layer and nuclear segmentation (LS & NC) and when performing layer segmentation only (LS) than when compared to both other models. This difference in F1-score was particularly apparent in seg-

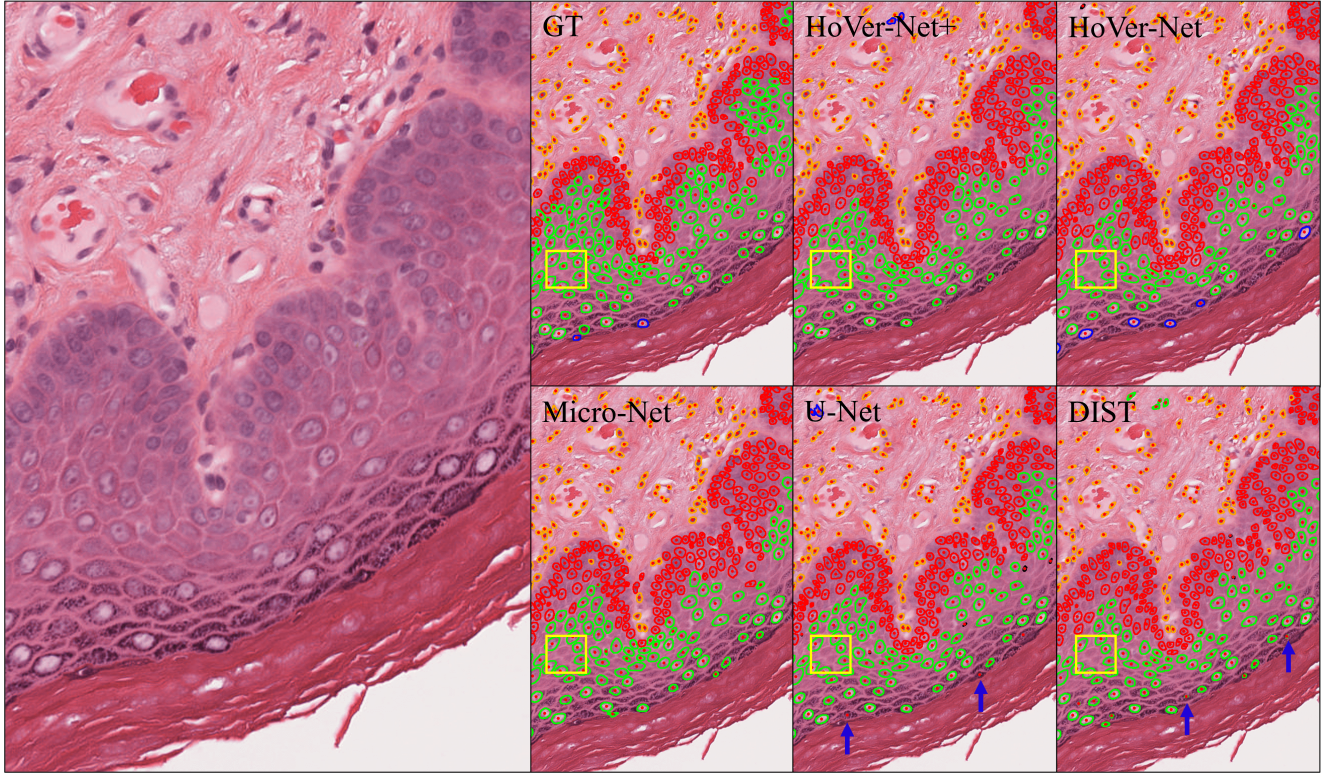


Figure 2. The nuclear segmentations from each of the models compared to the GT. Here, red represents basal, green represents epithelial, blue keratin, and orange other nuclei. The yellow boxes show areas of false negatives when compared to the GT, whilst the blue arrows show false positives.

menting the keratin and other tissue; however, was somewhat counteracted in the basal layer, where DeepLabv3 performed the best. Overall, these results, show the superiority of the HoVer-Net+ architecture for semantic segmentation of the epithelium when compared to other SOTA approaches. The marginal difference between HoVer-Net+ for simultaneous layer and nuclear segmentation *vs.* for layer segmentation alone, suggest that simultaneous segmentation provides a more efficient approach to these tasks.

A comparison of the layer segmentations of HoVer-Net+ (NC & LS) *vs.* the GT on a ROI from the test set can be seen in Figure 3. The red boxes show areas where HoVer-Net+ has segmented regions that the GT segmentation has misclassified, whilst the blue boxes show regions where HoVer-Net+ has misclassified some basal layer epithelium and other tissue. There are also areas of connective tissue that have been misclassified as keratin. Overall, the predicted segmentations appear to be smoother than the GT.

6. Discussion and Conclusion

In this work we have proposed HoVer-Net+, a new model for simultaneous nuclear instance segmentation (and classification) and semantic segmentation of three layers within

the epithelium of OED WSIs. The accurate segmentation of both nuclei and layers within OED cases is a vital step in down-stream analysis, due to the importance of finding layer-specific morphological features for the prediction of grade and outcome/malignancy. Indeed, previous work by [2] used such features to predict OED grade; however, was limited by a small cohort. The proposed method provides an avenue to perform rapid segmentation of nuclei/layers, allowing studies to scale-up approaches for outcome prediction with large cohorts. However, we emphasise that the potential of the proposed method is not confined to oral epithelial layers, or indeed the epithelium.

This study has shown HoVer-Net+ to perform in line with the current SOTA approaches for both nuclear segmentation/classification and layer segmentation. For layer segmentation, we found HoVer-Net+ to achieve better results than both U-Net and DeepLabv3. We suggest that this may be due to our experiments taking place at 20 \times , a necessity for ensuring accurate nuclei segmentation, as opposed to 10 \times , which has been used in previous work [25, 2]. We have shown that simultaneous segmentation of both nuclei and layers can yield SOTA performance by combining both tasks without any loss in performance. This further suggests

Model	F1-score						Prec.	Recall	Acc.
	Bkgd.	Other	Basal	Epith.	Keratin	Mean			
U-Net	0.874	0.666	0.719	0.740	0.494	0.699	0.698	0.720	0.721
DeepLabv3	0.863	0.751	0.748	0.832	0.601	0.759	0.755	0.768	0.786
HoVer-Net+ _{LS}	0.900	0.876	0.718	0.868	0.760	0.824	0.834	0.819	0.841
HoVer-Net+ _{LS,NC}	0.898	0.871	0.720	0.860	0.752	0.820	0.824	0.821	0.835

Table 3. Comparative experiments for layer segmentation on the OED dataset.

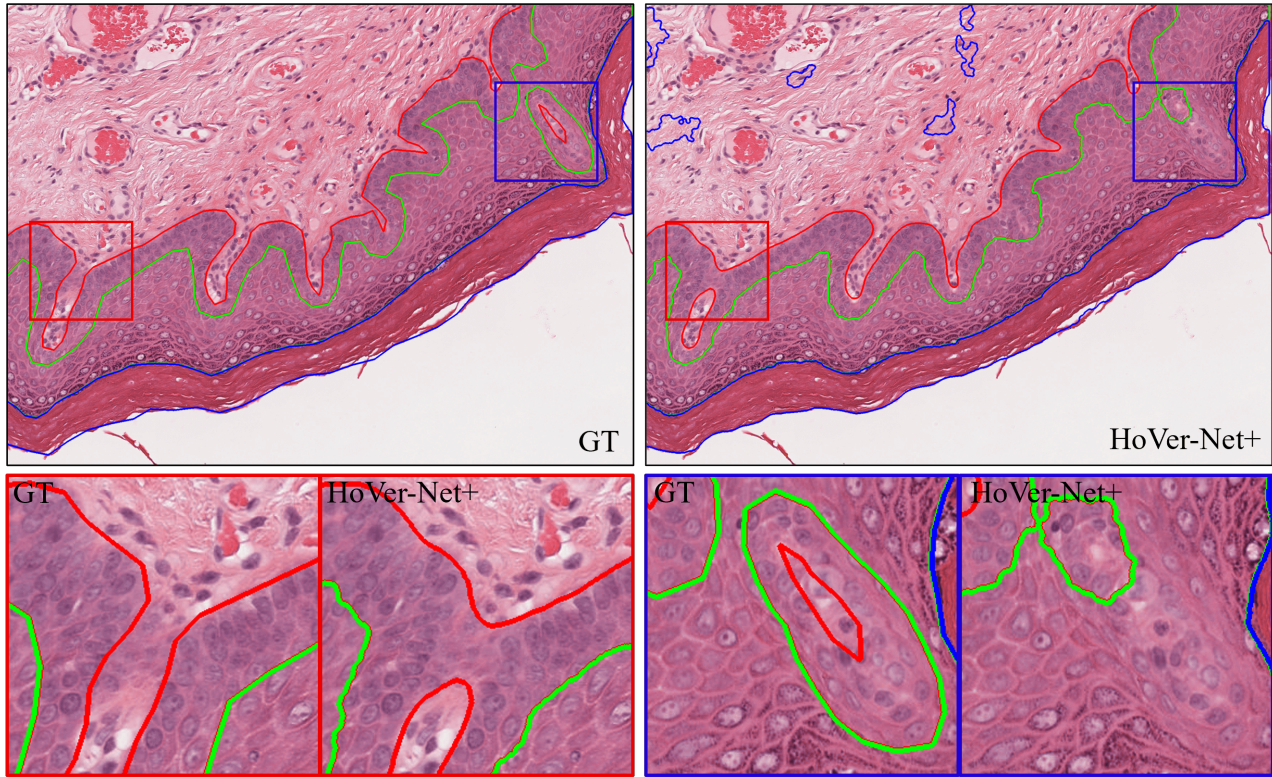


Figure 3. A comparison of the GT vs. the HoVer-Net+ layer segmentation. In the images, the red line represents the basal layer, the green epithelium, and blue keratin. For clarity, other tissue regions have been excluded from these results.

that important domain-specific features are learnt by the encoder that are useful for both nuclear and layer segmentation. It also demonstrates the adaptability and efficiency of the proposed network for multi-task learning.

This study was limited in its ability to assess nuclear instance segmentation based on the whole cohort, due to the GT nuclear instances being produced by a combination of NuClick and HoVer-Net. However, this method was chosen since the manual segmentation of each nucleus in all 43 WSIs would have been too time consuming. We did, however, provide an unbiased assessment of nuclear segmentation performance by testing each model on the hold-out NuClick ROIs, where HoVer-Net+ still obtained the best performance. Future work should further separate nuclei into more classes (*e.g.* immune and connective cells).

We believe this work advances the field of computational pathology by presenting an efficient model for the simultaneous segmentation of nuclear instances and epithelial layers. We have demonstrated its success in a complex cohort of OED cases, and emphasize its potential in other cohorts for a multitude of tasks. HoVer-Net+ could be used as an essential step to scale-up future studies for the rapid generation of morphological features for use in down-stream analyses to better understand the tumour microenvironment and ultimately improve outcome prediction. Future work could instead use the LS branch to simultaneously segment nuclei and tumour/stroma regions, or other similar simultaneous tasks, more efficiently by incorporating knowledge of underlying cytological changes between the different tissue constituents as captured by the proposed model.

References

- [1] Navid Alemi Koohbanani, Mostafa Jahanifar, Neda Zamani Tajadin, and Nasir Rajpoot. NuClick: A deep learning framework for interactive segmentation of microscopic images. *Medical Image Analysis*, 65, 2020.
- [2] Raja Muhammad Saad Bashir, Hanya Mahmood, Muhammad Shaban, Shan E. Ahmed Raza, Muhammad Moazam Fraz, Syed Ali Khurram, and Nasir Rajpoot. Automated grade classification of oral epithelial dysplasia using morphometric analysis of histology images. In *Medical Imaging 2020: Digital Pathology*, volume 11320. International Society for Optics and Photonics, 2020.
- [3] Kaustav Bera, Kurt A Schalper, David L Rimm, Vamsidhar Velcheti, and Anant Madabhushi. Artificial intelligence in digital pathology - new tools for diagnosis and precision oncology. *Nature Review Clinical Oncology*, 16(11):703–715, 2019.
- [4] Wouter Bulten, Péter Bándi, Jeffrey Hoven, Rob van de Loo, Johannes Lotz, Nick Weiss, Jeroen van der Laak, Bram van Ginneken, Christina Hulsbergen-van de Kaa, and Geert Litjens. Epithelium segmentation using deep learning in H&E-stained prostate specimens with immunohistochemistry as reference standard. *Scientific Reports*, 9(1):1–10, 2019.
- [5] Liang-Chieh Chen, George Papandreou, Florian Schroff, and Hartwig Adam. Rethinking Atrous Convolution for Semantic Image Segmentation. *arXiv*, 2017.
- [6] European Cancer Patient Coalition. European Cancer Patient Coalition: Head & Neck Cancers.
- [7] Jevgenij Gamper, Navid Alemi Koohbanani, Ksenija Benes, Simon Graham, Mostafa Jahanifar, Syed Ali Khurram, Ayesha Azam, Katherine Hewitt, and Nasir Rajpoot. Pan-Nuke Dataset Extension, Insights and Baselines. pages 1–12, 2020.
- [8] Yi Gao, Vadim Ratner, Liangjia Zhu, Tammy Diprima, Tahsin Kurc, Allen Tannenbaum, and Joel Saltz. Hierarchical nucleus segmentation in digital pathology images. *Medical Imaging 2016: Digital Pathology*, 9791(March 2016):979117, 2016.
- [9] Simon Graham, Quoc Dang Vu, Shan E.Ahmed Raza, Ayesha Azam, Yee Wah Tsang, Jin Tae Kwak, and Nasir Rajpoot. Hover-Net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images. *Medical Image Analysis*, 58:101563, 2019.
- [10] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Identity mappings in deep residual networks. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 9908 LNCS:630–645, 2016.
- [11] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q. Weinberger. Densely connected convolutional networks. *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, 2017-Janua:2261–2269, 2017.
- [12] Alexander Kirillov, Kaiming He, Ross Girshick, Carsten Rother, and Piotr Dollar. Panoptic segmentation. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2019-June:9396–9405, 2019.
- [13] Omar Kujan, Ammar Khattab, Richard J. Oliver, Stephen A. Roberts, Nalin Thakker, and Philip Sloan. Why oral histopathology suffers inter-observer variability on grading oral epithelial dysplasia: An attempt to understand the sources of variation. *Oral Oncology*, 43(3):224–231, 2007.
- [14] Neeraj Kumar, Ruchika Verma, Sanuj Sharma, Surabhi Bhargava, Abhishek Vahadane, and Amit Sethi. A Dataset and a Technique for Generalized Nuclear Segmentation for Computational Pathology. *IEEE Transactions on Medical Imaging*, 36(7):1550–1560, 2017.
- [15] Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen A.W.M. van der Laak, Bram van Ginneken, and Clara I. Sánchez. A survey on deep learning in medical image analysis. *Medical Image Analysis*, 42:60–88, 2017.
- [16] Wei Liu, Zhe Xuan Bao, Lin Jun Shi, Guo Yao Tang, and Zeng Tong Zhou. Malignant transformation of oral epithelial dysplasia: Clinicopathological risk factors and outcome analysis in a retrospective cohort of 138 cases. *Histopathology*, 59(4):733–740, 2011.
- [17] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 07-12-June:3431–3440, 2015.
- [18] H. Mahmood, M. Shaban, B. I. Indave, A. R. Santos-Silva, N. Rajpoot, and S. A. Khurram. Use of artificial intelligence in diagnosis of head and neck precancerous and cancerous lesions: A systematic review. *Oral Oncology*, 110(April):104885, 2020.
- [19] Aymen Mouelhi, Mounir Sayadi, Farhat Fnaiech, Karima Mrad, and Khaled Ben Romdhane. Automatic image segmentation of nuclear stained breast tissue sections using color active contour model and an improved watershed method. *Biomedical Signal Processing and Control*, 8(5):421–436, 2013.
- [20] Peter Naylor, Marick Laé, Fabien Rey, and Thomas Walter. Segmentation of Nuclei in Histopathology Images by Deep Regression of the Distance Map. *IEEE Transactions on Medical Imaging*, 38(2):448–459, 2019.
- [21] Shan E.Ahmed Raza, Linda Cheung, Muhammad Shaban, Simon Graham, David Epstein, Stella Pelengaris, Michael Khan, and Nasir M. Rajpoot. Micro-Net: A unified model for segmentation of various objects in microscopy images. *Medical Image Analysis*, 52:160–173, 2019.
- [22] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 9351:234–241, 2015.
- [23] A. R. Shubhasini, B. N. Praveen, Usha Hegde, K. Uma, G. Shubha, G. Keerthi, and Shiladitya Sil. Inter- and intra-observer variability in diagnosis of oral dysplasia. *Asian Pacific Journal of Cancer Prevention*, 18(12):3251–3254, 2017.

- [24] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, pages 1–14, 2015.
- [25] Sudhir Sornapudi, Jason Hagerty, R. Joe Stanley, William V. Stoecker, Rodney Long, Sameer Antani, George Thoma, Rosemary Zuna, and Shellaine R. Frazier. EpithNet: Deep Regression for Epithelium Segmentation in Cervical Histology Images. *Journal of Pathology Informatics*, 11:1–10, 2020.
- [26] Paul M. Speight, Syed Ali Khurram, and Omar Kujan. Oral potentially malignant disorders: risk of progression to malignancy. *Oral Surgery, Oral Medicine, Oral Pathology and Oral Radiology*, 125(6):612–627, 2018.
- [27] T Takata and P Slootweg. Tumours of the oral cavity and mobile tongue: epithelial precursor lesions. In AK El-Naggar, JKC Chan, JR Grandis, T Takata, and PJ Slootweg, editors, *WHO Classification of Head and Neck Tumours*. 4th edition, 2017.
- [28] Jeroen van der Laak, Geert Litjens, and Francesco Ciompi. Deep learning in histopathology: the path to the clinic. *Nature Medicine*, 27(5):775–784, 2021.
- [29] Mitko Veta, Paul J. van Diest, Robert Kornegoor, André Huisman, Max A. Viergever, and Josien P.W. Pluim. Automatic Nuclei Segmentation in H&E Stained Breast Cancer Histopathology Images. *PLoS ONE*, 8(7):1–12, 2013.