An Algorithm for Microscopic Specimen Delineation and Focus Candidate Selection

Yilun Fan, Yaniv Gal and Andrew P. Bradley
The University of Queensland, School of Information Technology and Electrical Engineering, St Lucia, QLD 4072, Australia.
Email: y.fan3@uq.edu.au

Abstract
In this paper, we compare four field-of-view (FOV) metrics that, when applied to a low-resolution image of a microscope slide, are capable of both accurately delineating the specimen and selecting a subset of focus candidate FOVs required for construction of a high-resolution focus map. The metrics evaluated are: threshold index (TI) that measures image intensity; normalised auto-correlation index (NACI) that measures spatial image similarity; auto-phase correlation index (APCI) that measures image phase diversity; and entropy index (EI) that measures the predictability of image intensities. Experiments are undertaken on a data set of forty slides including PAP stained Thin-prep cervical cytology and breast fine-needle aspiration slides and Hematoxylin and eosin (HE) stained histology slides. These slides were scanned on an automated bright-field microscope and chosen to be indicative of variety pathology specimens, containing artefacts such as excess coverslip glue and ink markers. Results are presented on the performance of each metric for correct ranking/segmentation of foreground (specimen) from background, and subsequently selecting focus candidate FOVs characteristic of the specimen's focal plane(s). The experimental results demonstrate that while NACI, APCI and EI are all effective at specimen delineation, only APCI is capable of effectively selecting superior focus candidates and ignoring artefacts.

Keywords: Specimen delineation, focus map, focus candidates, auto-phase correlation, tile-based processing.
1. Introduction
The acquisition time of high-resolution microscope slide images is normally prohibitively long. This is because of the large number (many thousands) of fields-of-view (FOVs) and the requirement to maintain focus over the whole specimen so that the final digital slide image is clear and sharp [4, 17]. To reduce acquisition time, it is typical to initially delineate the specimen from a low-resolution scan to minimise acquisition of blank FOVs, and then to construct a focus map, using a small subset of FOVs, that the system can follow rather than focusing on each high-resolution image [1, 8, 17]. The most straightforward approach to find the region of interest (specimen) from the background (blank areas) is to use an intensity thresholding followed by spatial smoothing [1]. This scan map can then be used to guide the acquisition of high-resolution FOVs over the specimen [1, 17]. Once the scan map is constructed, a focus map is formed by interpolating the focal positions measured from a limited number of FOVs within the specimen's scan map. Typically, these focus candidates are either randomly selected or follow a predefined pattern [10].

In our previous work we have shown that conventional methods for producing scan and focus maps are not robust to slide artefacts such as dust, coverslip glue and ink markers [8]. Specifically, as the thresholding method selects an image intensity value that distinguishes specimen from background, it is likely that artefacts, having a similar or darker opacity to the specimen, will also be included in the scan map. As the artefacts typically lie at a different focal plane to the specimen, the quality of the focus map could be severely affected.

This has led to the development of robust methods that reject outliers when interpolating the focus map constructed from a random or fixed pattern subset of possible focus candidates [10]. In the absence of such techniques the interpolated focus map has the potential to be biased towards the artefacts (say) on top of the slide and hence the images acquired become de-focused and appear blurred. This problem becomes worse as the magnification increases, as the depth of field (DOF) of the objective lens reduces. For example, at X40 magnification the DOF is around one micron [5, x15.5]), while dust artefacts are several hundreds of microns away from the specimen on the surface of the coverslip. Another potential solution to this problem is to acquire multiple high-resolution images from multiple focal planes and to then construct an extended depth of field (EDF) image of the specimen [4]. However, this further increases scan time and produces images that lack the familiar depth cue provided by the limited depth of field. Therefore, in this paper we propose a combined scan and focus map algorithm that aims to be more effective and efficient by not including these outliers in either the scan or focus map.

The algorithm proposed in this paper analyses a low-resolution image of the entire microscope slide to produce both a specimen scan map and an ordered list of focus candidates from which to build the high-resolution focus map. It utilises a tile-based
approach, where each tile in the low-resolution image relates directly to a potential (i.e., yet to be acquired) high-resolution FOV image of the specimen. This tile-based approach is inspired by the popular FOV digital slide acquisition technique, which uses an area scan image sensor to capture overlapping, high-resolution, images of the specimen that are then stitched together to produce an image (mosaic) of the entire specimen [8]. However, in our approach all of the low-resolution tiles are quantitatively evaluated by a metric to determine the amount of specimen presents. Although studies have been done to examine the high-resolution scanned images to determine whether they are artefacts or in focus, our approach aims to uncover this information from the low-resolution scanned images. Therefore, it can be seen as a pre-processing step to these methods [12, 15]. Ideally, the tile-evaluation metric should indicate not only how many objects, or amount of specimen, the tile contains, but also the degree to which these objects are in focus. In this way, the metric has the potential to distinguish specimen (in focus) from artefact (out of focus) utilising information acquired from the low-resolution image alone. Based on the value of the metric, tiles can then be classified as foreground (specimen to be scanned at high-resolution) and background (blank areas not to be scanned). More importantly, the tiles can also be treated as potential focus candidates, and hence can be ranked based on the value of the metric. More specifically, the tiles ranked by the metric can be used to guide the selection of focus candidates for the focus map (because the more specimens a tile contains the better it should be as a focus candidate). Therefore, for a tile-based method to be robust to artefacts, the tile evaluation metric should also be robust to artefacts.

In this study, we investigate and compare four different metrics suitable for evaluating tile-based specimen delineation. These metrics are: threshold index (TI) that measures the percentage of foreground to background pixels in each tile; normalised auto-correlation index (NACI) that measures spatial image similarity within each tile; auto-phase correlation index (APCI) that measures the phase diversity of each tile; and entropy index (EI) that measures the predictability of image intensities within each tile. The four metrics are evaluated on forty pathology slides including: sixteen Thin-prep cervical cytology and nine breast fine-needle aspiration (FNA) specimens, both PAP stained, as well as fifteen Hema-toxylin and eosin (HE) stained histology neuro-anatomy slides. The work presented in this paper is a significant extension and improvement of that previously published in [8], which only compared APCI and TI on nine PAP stained Thin-prep slides.

This paper is structured as follows: Section 2 details the tile-based specimen delineation method and the four tile evaluation metrics; Section 3 describes the slides analysed and the experimental methodology; Section 4 illustrates the results for both slide delineation and the evaluation of focus candidates; Section 5 discusses the results and compares the advantages and disadvantages of each metric.
2. Methodology
The flow chart of the tile-based specimen delineation method is shown in Fig. 1. The tile-based specimen delineation method includes five steps, namely, low-resolution acquisition of the whole slide image, coverslip detection and removal, tile evaluation, scan map segmentation and focus candidate selection.

![Flow chart of the tile-based specimen delineation method](image)

Figure 1: Flow chart (left) and example images of the tile-based specimen delineation method. A: Low-resolution image of the slide; B: Detected lines (green); C: Cropped slide image without the coverslip; D: Tile allocation on the low-resolution slide (each blue box represents a high-resolution tile); E: Tile matrix (showing the APCI metric); F: Tile matrix after morphological filtering; G: Scan map (red is foreground); H: Ranked tiles across the scan map; I: Positions of 30 selected focus candidates (blue points).

2.1. Low-resolution Scan
The algorithm starts with image acquisition of the whole slide under low magnification (say x10). The initial low-resolution scan is relatively fast either from a stand-alone macro camera, used in most commercially available whole slide scanners,
or from a low magnification objective of motorised microscope [17]. Here, we assume the system has an area scan sensor that takes overlapping FOV images sequentially over the whole slide. Then the low-resolution FOV images are stitched together to form a single image of the whole slide [1].

2.2. Coverslip Detection
Normally, a coverslip is glued on the top of the glass microscope slide to protect the specimen and an identification label is put on the side. On the low-resolution slide image, the edges of the coverslip and the glass slide normally have clear and strong boundaries allowing relatively easy detection. Therefore, the coverslip is detected and only the image area inside the coverslip is processed at the next step. This not only reduces the effect of artefacts such as coverslip edges, glue and labels, but also reduces the number of tiles to be analysed. First, we apply a Sobel edge detector [9] to the low-resolution image. Next, we apply a Hough transform [9] to the edge image and select the strongest peaks from the Hough matrix. The locations of lines corresponding to these peaks are found which, because these are the strongest lines in the image, relate to the locations of the edges of the slide and coverslip, e.g., as shown in Fig. 1.B. Finally, we remove the image areas outside of these lines and process the image representing the enclosed area (under the coverslip).

2.3. Tile Evaluation
We first locate the tiles to be processed in the low-resolution image produced in Step 2.2. The size of these tiles depends on the difference in magnification between the initially acquired low-resolution image and the high-resolution image yet to be acquired. The tiles are positioned adjacent to each other so that they form a mosaic of the whole low-resolution image, such as shown in Fig. 1.D. The positions of each tile are determinate based on the initial position of the first tile and any overlap (for stitching) between tiles [1].

After tile allocation, all of the tiles are evaluated individually by the proposed metrics, and their values are placed into a tile matrix based on their position in the image, an example is shown in Fig. 1.F. The tile matrix does not contain the actual positions of the tiles, rather they can be found based on the geometry and physical position of the slide. The following sub-sections describe the four tile evaluation metrics studied in this paper, namely: thresholding index (TI), normalised auto-correlation index (NACI), auto-phase correlation index (APCI) and entropy index (EI). Among the equations below, we let I represent a low-resolution image tile, which has dimensions of M by N, where normally M = N, I(m; n) represents a pixel intensity at position (m; n) of the image I, where 0 <= m <= M and 0 <= n <= N. A number of cytology images, their corresponding values given by the metrics and processing time are shown in Fig. 2 and Table 1. In Table 1, the threshold index of images B, C and D is calculated using the threshold value obtained over the whole low-resolution image of the specimen, e.g., A in Fig. 2.
Figure 2: Sample cytology images at magnification of 40X.

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>Average Time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thresholding Index</td>
<td>0.1385</td>
<td>0.0737</td>
<td>0.0289</td>
<td>0</td>
<td>0.0025</td>
</tr>
<tr>
<td>NACI</td>
<td>0.9991</td>
<td>0.9984</td>
<td>0.9964</td>
<td>0.2082</td>
<td>0.4328</td>
</tr>
<tr>
<td>APCI</td>
<td>0.7569</td>
<td>0.6792</td>
<td>0.5300</td>
<td>0.1028</td>
<td>0.0362</td>
</tr>
<tr>
<td>Entropy Index</td>
<td>0.8190</td>
<td>0.8020</td>
<td>0.7450</td>
<td>0.6543</td>
<td>0.0031</td>
</tr>
</tbody>
</table>

Table 1: Evaluation of images in Fig. 2 and average computation time over 1000 512x512 images.

2.3.1. Thresholding Index (TI)
The most straightforward way to tell whether a tile is empty or contains cellular objects is by measuring the image intensity. This is effectively the conventional approach to specimen delineation, which assumes that darker pixels relate to specimen and brighter pixels to background. Therefore, we initially apply Otsu's classical image thresholding method [16] to the low-resolution image obtained from the previous step of the algorithm 2.2. This produces a binary image that, to a first approximation, separates foreground from background. Next we calculate the threshold index of each tile as the percentage of foreground pixels in the tile,

\[ TI = \frac{1}{MN} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} (I(m, n) > t) \]  

(1)

Where \( t \) is a global threshold value found by Otsu's method in order to distinguish specimen from background. Due to its connection with the traditional specimen delineation methods, we treat the TI as the baseline for comparison.

2.3.2. Normalised Auto-Correlation Index (NACI)
It is well known that cross correlation is able to find the similarity between two signals and to co-register two images that have some degree of overlap [13]. However, the method fails when the two images contain objects with strong intensities or an uneven illumination gradient, because these high intensity areas result in high correlation measurements as their signal power is strong [13]. Therefore, normalised
cross-correlation (NCC) is typically implemented as it is more robust [13]. We also know that the auto-correlation is able to highlight repeated or periodic signals. Thus, if we correlate two images of the same scene but with different background noise, the auto-correlation index will estimate the power of the signal with respect to the power of the (random) underlying noise [9]. Here, we acquire two low-resolution images $I_1$ and $I_2$ of each slide with the same exposure and illumination settings, but with different realisations of the background noise. Therefore, the normalised auto-correlation index (NACI) can be used to directly compare and measure the amount of specimen in each tile. As we only require the complete correlation of two tile images, NACI is as follows:

$$NACI = \frac{1}{MN} \left( \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} (I_1(m, n) - \mu_1)(I_2(m, n) - \mu_2) \right) \frac{1}{\sigma_1 \sigma_2 + \epsilon}$$ (2)

Where $\mu_1$, $\mu_2$ are the mean value and $\sigma_1$, $\sigma_2$ are the standard deviation of the image $I_1$, $I_2$, and $\epsilon$ is a small constant to prevent division by zero.

As shown in Fig. 2 and Table 1, the more cellular objects in an image, the higher NACI. In contrast, if the tile is empty, such as image D in Fig. 2, a low NACI value is obtained.

2.3.3. Auto-Phase Correlation Index (APCI)

As an alternative to NCC, phase correlation can be used for robust registration of images [11]. Unlike NCC, phase correlation considers the phase of the images only, effectively equalising image spectral magnitudes [11]. The principle of APCI can be explained in the same way as NACI. The two realisations of each image are correlated in the phase domain and APCI reflects the degree of coherent phase of the specimen in the tile compared to the incoherent phase of the background noise. For example, the cytology images in Fig. 2 contain more cells and have more edges (of nuclei and cytoplasm) and hence are assigned higher APCI values.

Unlike NACI, APCI is conveniently calculated in the phase domain by normalising the power spectrum of the two images [8]:

$$APCI = \frac{1}{MN} \left| \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \left( \frac{\hat{I}_1(m, n) \hat{I}_2(m, n)^*}{|\hat{I}_1(m, n) \hat{I}_2(m, n)^*|} \right) \right|$$ (3)

Where $\hat{I}_1$ and $\hat{I}_2$ are Fourier transforms of the image $I_1$ and $I_2$, and $*$ indicates complex conjugate. To minimise boundary effects we apply a Tukey window [9] with
a radius of 0.95 of M and N.

2.3.4. Entropy Index (EI)
The entropy of an image reflects how unpredictable or random an image is by estimating the probability of individual intensity values. In this way, an empty tile has low entropy as it has predictable values, while a tile containing cellular objects has larger entropy as it is more unpredictable. All three metrics proposed above have an operating range of \{0, 1\}, and so we normalise entropy (4) in the same way, as shown in (5). As shown in Fig. 2 and Table 1, EI successfully ranks the four images based on the number of cells, although the empty tile is given a value close to those with specimen.

\[
Entropy = -\sum_i (p_i \log(p_i)) \quad (4)
\]

\[
EI = 1 - \frac{1}{Entropy + 1} \quad (5)
\]

Where \(p_i\) is the probability of the intensity value \(i\) in image \(I\). Here, image \(I\) has eight bits of precision and so has 256 distinct grey levels.

2.4. Scan Map Segmentation
The tile matrix from the tile evaluation process above is processed with morphological filters and then the filtered tile matrix is segmented to produce the final scan map.

2.4.1. Morphological Filtering
The morphological filters are used to increase the sensitivity of specimen delineation at the cost of reduced specificity. In practice, an ideal scan map should consist of tiles that encompass the complete specimen, with the minimum amount of background and other outliers. In this study, we assume the slides are prepared so that the specimens appear in one or two main areas, either as a tissue slice or a contiguous deposition (smear) of cells. Therefore, the morphological filter may remove isolated tiles containing isolated cells or small tissue fragments. However, we also believe that the scan map should be efficient for the high-resolution scan. For example, for a system using a line scan sensor, it is faster and easier for the sensor to scan over a small area of background, which means the scan map should not contain small holes. To meet these goals, we apply a number of grey scale morphological operators [9] to the tile matrix. The combination of grey scale morphology followed by a binary threshold has been shown to be preferable to the initial application of a threshold followed by binary morphology [3].

2.4.2. Segmentation
After the application of the morphological filters, a threshold value needs to be found to distinguish the tiles that contain foreground (specimen) from tiles that contain primarily background. The classical Otsu thresholding method is an effective way of automatically selecting a threshold value by analysing the distribution of tile index values. The method assumes that each object distribution follows a Gaussian (normal) distribution. In this way, if the intensities of foreground and background distributions overlap, Otsu's method picks a threshold value around the centre of the overlap [9]. However, in our application, the foreground (specimen) distribution is both more important and more variable than the background, as a user would prefer to acquire the entire specimen and tolerate a degree of background. Therefore, we implemented a modified version of the Otsu method to bias the threshold away from the specimen distribution in order to increase sensitivity.

First, the classical Otsu method is applied to the tile matrix and an initial threshold ($k_1$) that separates the two intensity classes (foreground and background) is found. We assume a bimodal distribution because we want the metric to detect (foreground) specimen and additional classes may represent artefacts. Next, we find the mean ($\mu$) and standard deviation ($\sigma$) of the foreground class, which is defined to have higher index. We place a biased threshold ($k_2$) two standard deviations below the mean of the foreground distribution ($k_2 = \mu - 2\sigma$). Assuming a Normal distribution, this assigns 97.5% of the foreground distribution to the foreground class. The two standard deviation distance is obtained heuristically using a number of different slide types and metrics with the aim of maximising sensitivity. The final threshold ($k$) is then taken as the minimum of $k_1$ and $k_2$. Finally, the scan map is 10 obtained using the threshold $k$ as shown in Fig. 1.G.

2.5. Focus candidate selection
The tile matrix produced from Section 2.3 also provides a ranking of all tile indexes with regard to the extent of objects in the tile. Obviously, in the Fig. 2, the image tile A is a good focus candidate and should be selected with priority as a focus candidate, while the tile D is a poor focus candidate and should not be selected.

Here we utilise a simple greedy search method to select a subset of focus candidates based on their tile index value and spatial location. First, we rank all tiles in the tile matrix in descending order. Next, we select the highest valued tile as the first focus candidate, and iterate down through the ranking. We select a tile as a focus candidate if the tile is contained in the scan map (obtained from Section 2.4) and if the position of the tile is a sufficient distance away from a previously selected focus candidate. This process is repeated until the required number of focus candidate is found or until the list of tiles is exhausted.

3. Experimental Methodology
3.1. Experimental Slides
The experimental slides include sixteen cervical cytology slides, prepared using the Thin-prep monolayer slide preparation technique with PAP stain, nine slides consisting of PAP stained smears from fine needle aspirates (FNA) from breast cancer diagnosis, and fifteen histology slides of Hematoxylin and eosin (HE) stained neuroanatomy specimens. To examine the robustness of the proposed method, we purposely selected slides with a range of conditions that are seen in routine clinical practice, including a varying degree of stain intensity, age and cell density. Some slides are contaminated by dust and glue artefacts, while some are post-diagnosis and so contain ink markers from previous manual diagnosis (as may be scanned for research or quality assurance purposes). All slides are standard microscope slides with dimensions of 75 by 25mm, thickness of around 1mm and with coverslips around 40 by 20mm. The sample slides are summarised in Table. 2.

<table>
<thead>
<tr>
<th></th>
<th>Thin-prep</th>
<th>FNA</th>
<th>Histology</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>With Artefacts</td>
<td>10</td>
<td>1</td>
<td>5</td>
<td>16</td>
</tr>
<tr>
<td>Without Artefacts</td>
<td>6</td>
<td>8</td>
<td>10</td>
<td>24</td>
</tr>
<tr>
<td>Total</td>
<td>16</td>
<td>9</td>
<td>15</td>
<td>40</td>
</tr>
</tbody>
</table>

Table 2: The number of slides in each specimen type and condition.

3.2. Digital Slide Acquisition System
A fully functional digital slide acquisition platform is used in this study. The system consisted of a motorised stage, an Olympus BX40 microscope with 4X and 40X objective lens, and a four mega-pixel SPOT Insight camera with a pixel size of 7.4 µm and 100% fill factor. The maximum resolution produced by the system is 0.185 µm/pixel under the 40X objective.

3.3. Scan Map Ground truth

Figure 3: Two low-resolution images of a Thin-prep slide (top left), a FNA slide (top middle) and a Histology slide (top right) and their ground truth (bottom).

On the Thin-prep slides, the cellular objects are approximately evenly spaced inside a
circular area, which has a measured diameter of around 20 mm. Therefore, a circle of diameter of 5400 pixels ( \approx 20,000 / 3.7 \times 3:70 \mu m/pixel image resolution) was manually placed to represent the specimen ground truth. The FNA and Histology slides are arbitrarily shaped and so a simple linear polygon was constructed around the cellular objects of interest as the ground truth. This was done using the Photoshop software (Adobe Systems Inc., San Jose, CA, USA). Examples of the scan map ground truth can be seen in Fig. 3.

3.4. Experimental Settings
This section gives details of specific parameter settings for the proposed method.

3.4.1. Low-resolution Scan
All glass slides were imaged using the 4ifi objective so that the low-resolution FOV images had a spatial resolution of 3:70 \mu m/pixel (as 2x pixel binning was used on the camera which effectively gives a pixel size of 14:8 \mu m). Before acquisition, each slide was manually focused and the focal axis of the stage was fixed during the scan. This is a plausible approach at low-resolution as the focal depth under the 4ifi objective is larger than the thickness of our specimens (around 20 \mu m) and smaller than the thickness of a slide.

In addition, for the calculations of NACI and APCI we require two images for each tile. Therefore, during acquisition of low-resolution images, we acquired two images at each FOV with the same exposure time and the illumination. As the number of images and the image exposure time are much smaller than for high-resolution scan, the increase in scan time due to the acquisition of the second image is insignificant. For example, with our system, each whole slide needs around 200 FOVs to be acquired and each exposure takes around 0.2 seconds, so the total time due to the second acquisition is around 40 seconds, which is acceptable and could be further reduced. Two images of the slide were then produced: both of them were used for calculating the NACI and APCI, and the average image was used for calculating TI and EI.

3.4.2. Coverslip Detection
To reduce processing time, we searched for lines only in the horizontal and vertical directions with a five degree tolerance and a step size of 0.2 degree. We down-sampled the original image by a factor of eight to further reduce processing time without sacrificing accuracy. The number of peaks examined in Hough matrix was 50, while the minimum line length was 225 pixels, which is around one third of the length of the short edges of the coverslip in the down-sampled image. After all lines meeting these criteria were found, we extended them to intersect with the edge of the image. The four most inward lines closest to the centre of the image then define the coverslip. With the above parameters, the coverslip of all experimental slides were
successfully identified. The above parameters were selected heuristically for the image resolution and the coverslip sizes used in this study.

3.4.3. Tile Allocation
At high magnification, each high-resolution FOV captures a spatial area of 379 µm by 379 µm, so the size of the tile projected on to the low-resolution image is 103 by 103 pixels (~379/3.7). For simplicity, we assume that no overlap is required between two high-resolution FOVs and hence the tiles. We put the first tile at the top left corner of the low-resolution image. For calculation of the APCI and NACI metrics, we tile the second slide image in the same way as the first image.

3.4.4. Morphological Filtering
The sizes of the structuring elements are first guided by the composition of the specimen such as cell density and gaps between tissue parts. This information is then transferred to the final sizes based on the image resolution, size of the FOV and overlap between tiles. Finally these parameters are fine-tuned with experiments of sample slides to maximise the segmentation performance.

First, we smooth the tile matrix via a circular averaging filter with size of three tiles because the values of tiles vary greatly even they all represent foreground, some may form distinct groups, which makes classification challenging. Next, we applied an image opening on the tile matrix with a disk structuring element (assuming the specimen is isotropic) of size three (tiles) in order to remove isolated and small-valued tiles that are less likely to be specimen. Then, we applied an image closing on the tile matrix with a disk structuring element of size five (tiles) to expand the potential specimen area and fill small holes. The same set of morphological filters was applied to all four tile evaluation metrics. Though the above filters were initially tuned to the Thin-prep slides, they all showed good performance on processing the FNA and histology slides.

3.5. Scan Map Segmentation Evaluation
We measure the Area Under of each Receiver operating characteristic (ROC) curve (AUC) to estimate the ranking performance of the four metrics [2]. As AUC measures the probability of correct ranking a foreground pixel higher than a background pixel it is independent of the threshold chosen to delineate (segment) the specimen. For all generated scan maps, we calculate the dice similarity index (DSI) [9] which compares the obtained scan maps with the ground truth described in 3.3.

3.6. Focus Map Evaluation
In order to evaluate the quality of the selected focus candidates, we attempt to construct the focus map in the same way as in real digital slide acquisition. That is by interpolating the focus map across the specimen area using a subset of selected focus candidates. Then, the focus maps are quantitatively compared with the ground truth as the number of focus candidates used for focus map interpolation is increased. This
approach highlights the effectiveness of each metric at selecting focus candidates, i.e. whether it produce a better ranking, and so fewer candidates are required to build an accurate focal map.

3.6.1. Focus Candidate Sampling and Ground-truth Generation
First, we need to sample the raw focal position data of the slide. To do this, we first locate each tile (from the tile allocation) at high magnification (40X) and run a focus searching algorithm to find the best focal plane of the high-resolution FOV. Due to poor stage positional accuracy of our system, FOV registration (using the normalised cross-correlation) of high-resolution FOVs with the low-resolution image was implemented to enable precise tile localisation. However, it is believed that the stage in modern scan systems have sufficient positional accuracy to relocate the tiles at high magnification. The focus metric used was normalised variance [18] and for each tile, we sample a range of focal depths with progressively smaller focal steps till a peak focus position is found. Next, we manually identify outlier FOVs, such as those focused on top of the coverslip, glue or ink marker artefacts. The ground truth of the focus map excludes all these outlier tiles and includes all valid tiles containing specimen.

3.6.2. Focus Map Construction
After we obtain the raw focal positions and focus map ground truth for each slide, we interpolate the focus map using the selected focus candidates and its focus information, and compare it with the ground truth using Mean Square Error (MSE). Specifically, we used a third order polynomial surface fitting in both directions for generating the focus map, while minimising the effect of any remaining outliers with the robust regression option enabled.

The optimal distance between two focus candidates depends on the topographic variation of the specimen's focal plane. For instance, in general, the histology slides are of a more uniform depth than the cytology slides, hence it requires a smaller number of data points for accurate interpolation. In this study, we set the minimum distance to be three FOVs, which may be a sub-optimal distance, but is sufficiently small for all three slide-types. A smaller distance will result in an increase in the number of focus candidates sampled, which may not be desirable in practice. We also implement a benchmark focus candidate selection algorithm based on a random selection of tiles (i.e. ignoring the tile evaluation metric) for comparison. Specifically, we randomly select a tile as a candidate and feed it to the greedy search process rather than using an ordered tile list. The ground truth scan map of the slide is used as the constrained map for the focus candidates in this case. We run the random selection of focus candidates 200 times and present the averaged results. Since acquisition of the full focus map and focus map interpolation are time consuming (may take more than 30 hours), we selected eight slides and sampled a rectangle area of the slide containing a combination of the specimen and artefacts.
4. Results
4.1. Scan Map Segmentation
The AUC of the four metrics are shown in Table 3. Averaged over all slide types, NACI achieved the highest AUC both before and after post-processing. Despite TI showing good performance when processing histology slides, the metric has the lowest AUC on average. Overall, all four of metrics achieved good overall performance with an average AUC of 0.97 after post-processing. When looking at the AUC based on the existence of artefacts, NACI again achieves the best results, both with/without artefacts and before/after post-processing. The performance of all four metrics is affected by the presence of artefacts. In particular, there was a 3-4% improvement in AUC when processing slides without artefacts.

The evaluation results of the scan map segmentation obtained after the tile matrices were thresholded are illustrated in Fig. 4 and Fig. 5. The TI metric achieved the poorest segmentation results for Thin-prep and FNA specimens, as shown with the lowest median DSI and largest interquartile range (IQR). However, the same metric achieved a competitive DSI (0.9502) when processing the Histology specimens, which is considerably better than that of Thin-prep (0.7782) and FNA (0.7854) specimens. In contrast, the scan maps generated using the other three metrics showed consistent high quality regardless of specimen type. Although the three metrics (APCI, NACI and EI) achieved similar DSI scores, APCI received the lowest False Positive Rate (FPR): 3.94% (APCI), 6.51% (NACI) and 6.02% (EI) for the Thin-prep slides; 0.97% (APCI), 3.21% (NACI) and 1.17% (EI) for the FNA slides; and 1.40% (APCI), 3.58% (NACI) and 2.16% (EI) for the Histology slides.

<table>
<thead>
<tr>
<th>Post-processing</th>
<th>Slide Type</th>
<th>Area Under ROC Curve</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TI</td>
<td>NACI</td>
</tr>
<tr>
<td>Before</td>
<td>Thin-prep</td>
<td>0.818</td>
</tr>
<tr>
<td></td>
<td>FNA</td>
<td>0.875</td>
</tr>
<tr>
<td></td>
<td>Histology</td>
<td>0.993</td>
</tr>
<tr>
<td>Average</td>
<td></td>
<td>0.895</td>
</tr>
<tr>
<td>After</td>
<td>Thin-prep</td>
<td>0.958</td>
</tr>
<tr>
<td></td>
<td>FNA</td>
<td>0.955</td>
</tr>
<tr>
<td></td>
<td>Histology</td>
<td>0.997</td>
</tr>
<tr>
<td>Average</td>
<td></td>
<td>0.970</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Post-processing</th>
<th>Slide Type</th>
<th>Area Under ROC Curve</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Artefacts Before</td>
<td>0.849</td>
</tr>
<tr>
<td></td>
<td>No Artefacts</td>
<td>0.904</td>
</tr>
<tr>
<td></td>
<td>Artefacts After</td>
<td>0.942</td>
</tr>
<tr>
<td></td>
<td>No Artefacts</td>
<td>0.970</td>
</tr>
</tbody>
</table>

Table 3: Area under the ROC curve before and after post-processing for the four tile valuation.
metrics. The highest value among the four is shown in bold.
When all slides are grouped as to the presence of slide artefacts, TI produces the worst performance among the four metrics. Specifically, its performance is affected significantly by the existence of artefacts, producing a decline in median DSI (from 0.9224 and an IQR 0.0043 without artefacts to a DSI of 0.7854 and IQR 0.0904 with artefacts). However, the other three metrics are less affected by the presence of artefacts, with their corresponding DSI values being maintained above 0.90.
The modified Otsu method also proved effective. Averaged across all experimental slides, the FNR decreased consistently across all metrics while DSI also increased, this is shown in Table 4. Regarding the computational complexity of the four metrics, we estimated average computation time of the tile evaluation process for all forty slides. All processing was performed in the MATLAB environment using a single Intel i7-3770 3.4GHz processor with 8 Gigabytes of RAM. The NACI takes considerably longer than the of other three metrics, at 95.86 seconds compared with 14.35 seconds for APCI, 0.98 seconds for EI and 0.43 seconds for TI. This result is consistent with the complexity of each metric as defined in Section 2.3.

4.2. Focus Map Evaluation
Figure 6: Average ranking of MSE from the focus map evaluation of eight slides. Note: ranking is relative to random selection (rank - 0) and so a negative rank means the MSE is better than that of random selection.

In order to summarise the focus map evaluation results, we rank the mean square error (MSE) of the focus maps generated from the same number of focus candidates for each slide against random selection as the baseline. As shown in Fig. 11 - 18, the MSE varies greatly between slides and metric and so averaging the MSE may bias the results between different slides/metrics. Therefore, we present a ranking which more directly compares the relative performance of selected focus candidates across each of the slides. The averaged ranks from the eight slides are shown in Fig. 6. Here, a negative rank means that the MSE of the focus map produced by a metric is smaller than that produced by random selection. Remarkably, only APCI achieves average rankings that are consistently better than random selection. Furthermore, according to the Sign Test [6], APCI is better than the three other metrics (NACI, EI and TI) at a very high level of significance (p<0.005). The other three metrics cross the random selection benchmark (rank 0) initially when the number of focus candidate being used is small (<15), but the rankings of all four metrics reduce as more focus candidates are used for the interpolation. In particular, TI requires more than 50 focus candidates before its performance is better than random selection.
Table 5: Sufficient number of focus candidates (SNFC) for each metric and slide. Note: a `-' means that the focus map did not achieve a sufficient MSE, the smallest number of sufficient focus candidates is shown in bold.

The ranking shown in Fig. 6, summaries the overall performance of the four metrics when the same number of focus candidates is used for focus map interpolation. Table 5 shows the 'sufficient' number of focus candidate (SNFC) and the corresponding sufficient MSE. Here, SNFC is broadly equivalent to settling or convergence time, where SNFC is defined to be the number of focus candidates required to achieve at least 50% of the globally minimum MSE observed from all four metrics, plus random selection, on each slide. Therefore, SNFC shows information of how quickly a metric converges to with 50% of the minimum MSE that can be expected on each slide.

Table 5 shows that overall APCI achieves the lowest SNFC on six out of the eight slides, which means that on average the APCI focal map MSE converges most rapidly to the sufficient MSE. Moreover, among four focus-candidate selection metrics only APCI converge to the sufficient MSE on all eight slides. In contrast, random selection achieved the sufficient MSE on only or four out of the eight slides. The intensity based TI metric performed slightly better than random selection, but still failed to converge on three of the slides.

Moreover, APCI also performed better in the presence of artefacts, with fastest convergence in all five artefact-affected slides. For the three slides without artefacts, NACI produced the minimum SNFC on two slides, while APCI produced one minimum. These three slides (Fig. 11 and Fig. 15-16) showed that the MSE curves of four metrics were close to each other. In terms of the artefact type, the two Thin-prep slides contain glue and dirt artefacts, and APCI showed a clear dominance. Particularly for the first Thin-prep slide in Table 5 (also illustrated in Fig. 7), the MSE curve of APCI is considerably below (smaller) all of the other metrics at all points (number of focus candidates) and the focus maps from all of the other metrics never reach an acceptable quality (see Fig. 17). The Random selection and TI were the
worst performers in processing these two slides, both of them failed to converge. On the other hand, the FNA slides contain ink marker artefacts. Here APCI also showed superior effectiveness, producing the smallest SNFC for all three of these slides.

Low-resolution images of the focus scan area for each slide and the resulting MSE for each tile metric as the number of focus candidates is increased from 10 to 100 are shown in Figs. 11 to 18 in the Supplementary Material.

5. Discussion
5.1. Scan Map Segmentation
The results in Fig. 4 and 5 demonstrate that APCI, NACI and EI are all capable of correctly discriminating specimen from background. Despite the experiment evaluating a modest number of slides (40), these were chosen to be representative of three common specimen types, containing commonly observed artefacts. Further, the scan maps obtained using APCI achieved the lowest average FPR, which implies that these scan maps contain less artefacts. Overall, the results show that the tile based delineation method is effective and that APCI is capable of distinguishing between (glue and ink related) artefacts and specimen.

<table>
<thead>
<tr>
<th>Ranks</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>APCI</td>
<td>0.823</td>
<td>0.818</td>
<td>0.818</td>
<td>0.817</td>
<td>0.815</td>
<td>0.806</td>
<td>0.801</td>
<td>0.794</td>
<td>0.792</td>
</tr>
<tr>
<td>NACI</td>
<td>0.818</td>
<td>0.219</td>
<td>0.183</td>
<td>0.219</td>
<td>0.758</td>
<td>0.634</td>
<td>0.710</td>
<td>0.815</td>
<td>0.204</td>
</tr>
<tr>
<td>EI</td>
<td>0.219</td>
<td>0.351</td>
<td>0.190</td>
<td>0.256</td>
<td>0.183</td>
<td>0.217</td>
<td>0.270</td>
<td>0.787</td>
<td>0.239</td>
</tr>
<tr>
<td>TI</td>
<td>0.190</td>
<td>0.201</td>
<td>0.219</td>
<td>0.351</td>
<td>0.142</td>
<td>0.363</td>
<td>0.256</td>
<td>0.183</td>
<td>0.175</td>
</tr>
</tbody>
</table>

Figure 7: Top: Tile matrix of the four metrics for a Thin-prep slide (top row) and the corresponding scan maps (bottom row). From left to right: TI, NACI, APCI and EI. Each pixel in the maps represents a single tile, red colour indicates high values (close to 1), while the dark blue indicate small values (close to 0). Middle and Bottom: The nine top ranked tiles by each of the four metrics, and in the table the values of these tiles evaluated by APCI. The APCI values of (blurred) outlier tiles are shown in bold.

To further interpret these results, we first examine a Thin-prep slide that contains glue artefacts, the tile matrix and scan map of which are shown in Fig. 7. Compared with the low-resolution image of the slide shown in Fig. 3(left), it can be seen that the tiles contaminated by glue artefacts (mainly upper centre part of the slide) have a lower APCI value than those containing specimen (position of the specimen refers to the ground truth of the slide in Fig. 3). In contrast, the other three metrics, NACI, EI and TI all had some contaminated tiles highlighted, and hence these outlier tiles are included in the final scan map. To verify this observation, the nine top ranked tiles from each of the four metrics were found, as shown in Fig. 7. It is remarkable that the top ranked tiles by APCI all contain real specimen, while the top ranked tiles by TI all contain (out of focus) glue and dirt artefacts. By this measure EI is the worst performer as it assigned eight out of nine outlier tiles the highest index, while the
NACI miss-identified four tiles. However, when all outlier tiles were evaluated by the APCI metrics, they all obtained small (\(<0.351\) compared to \(0.7-0.8\)) values. Clearly, APCI is robust to glue artefacts because these objects are on the surface of the coverslip and so are out of focus in the low-resolution scan. The glue on top of the coverslip is at least 0.4mm (around the thickness of the coverslip) away from the specimen, which is larger than the focal depth of the objective lens (4x, around 100 \(\mu m\)) used for low-resolution image acquisition. Thus, the glue and other objects around it appear defocused, and hence produce reduced phase diversity and a lower APCI value. Meanwhile, specimen is correctly focused at image acquisition: the nuclei and cytoplasm of individual cells are clear and sharp, and so tiles containing these objects have a high APCI as shown in Fig. 7. Similarly, the objects in the specimen focal plane, but directly below the glue also appear defocused, because the transparent glue causes light to be diffracted and alters the pathway of the illumination. The light of objects under the glue will focus on different points and so the objects appear defocused and blurred, consequently, a smaller APCI value is obtained. However, the defocused dirt glue still has low opacities and these tiles give high TI values as seen in Fig. 7. The EI also assigned these tiles high values, because the light of the object will be unevenly distributed to more pixels when the object is defocused, and so the randomness of the pixel intensities increases. The NACI, on the other hand, is sensitive to objects with either strong edges or high intensity, which often yield tiles with high contrast.

<table>
<thead>
<tr>
<th>Ranks</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>APCI</td>
<td>0.847</td>
<td>0.847</td>
<td>0.847</td>
<td>0.846</td>
<td>0.843</td>
<td>0.842</td>
<td>0.841</td>
<td>0.840</td>
<td>0.840</td>
</tr>
<tr>
<td>NACI</td>
<td>0.414</td>
<td>0.358</td>
<td>0.557</td>
<td>0.488</td>
<td>0.509</td>
<td>0.423</td>
<td>0.290</td>
<td>0.637</td>
<td>0.502</td>
</tr>
<tr>
<td>EI</td>
<td>0.842</td>
<td>0.840</td>
<td>0.766</td>
<td>0.785</td>
<td>0.818</td>
<td>0.770</td>
<td>0.803</td>
<td>0.755</td>
<td>0.820</td>
</tr>
<tr>
<td>TI</td>
<td>0.526</td>
<td>0.385</td>
<td>0.476</td>
<td>0.285</td>
<td>0.319</td>
<td>0.271</td>
<td>0.419</td>
<td>0.362</td>
<td>0.257</td>
</tr>
</tbody>
</table>

Figure 8: Top: Tile matrix of the four metrics on an FNA slide (top row) and the corresponding scan maps (bottom row). From left to right: TI, NACI, APCI and EI. Each pixel in the map represents a tile, red colour indicates high values (close to 1), while the dark blue indicate small values (close to 0). Middle and Bottom: The nine top ranked tiles by each of the four metrics, and in the table the values of these tiles evaluated by APCI. The APCI values of (blurred) outlier tiles are shown in bold.

We next examine an FNA slide with ink marker artefacts. The tile matrix, scan map and nine top ranked tiles by the four metrics are shown in Fig. 8, the low-resolution image and ground truth can be found in Fig. 3 (centre). Again, it can be seen that all of the top ranked tiles by the APCI metric contain specimen, while all tiles with high values of TI are fully occupied by the ink markers, which have a TI index of 1. Therefore, in the TI tile matrix, the tiles associated with markers are clearly visible and they have a strong influence on the scan map which has the largest FNR (34.62\% compared with 4.67\% (APCI), 1.22\% (NACI) and 2.15\% (EI)). The top ranked tiles
by NACI contain edges from the ink markers, as these have the strongest contrast and so the highest correlation. The tiles assigned high values by APCI and EI all have significant edges, but EI favours edges with high contrast, which miss-identifies four tiles containing part of an ink marker (tiles ranked 3, 4, 7 and 8). It worth noting that some outlier tiles are also assigned high APCI values, i.e. the tile ranked 8 by NACI (0.637) and tiles ranked 3, 4, 7, 8 by EI (0.766, 0.785, 0.803 and 0.755 respectively), however these are not among the top ranked tiles by APCI. Interestingly, the top nine ranked tiles by TI have APCI values not equal to zero. This is because the intensities in these tiles are not uniformly zero and thus contain small phase variations. The segmentation results in Fig. 5 show that NACI is able to distinguish tiles containing both artefacts and specimens from background, but the above two examples illustrate that it cannot reliably distinguish specimen from artefact. We can also observe from the example tile matrices above, that NACI highlights tiles that contain dirt, glue, isolated cells and dust. This implies that the correlation due to either specimen or artefact is both much stronger than the background noise and residual illumination gradient. Figures 7 and 8 show that TI is not robust to glue and ink-marker artefacts, particularly in segmentation of cytology specimens. However, Fig. 4 shows that TI is much more effective at segmenting histology slides, producing results that are more competitive with other metrics. The first reason for this is that the histology specimens contain tissue samples that have higher and more uniform cell density than that of the cytology specimens. Secondly, the example slides contain only minor glue artefacts without the significant artefacts seen in Thin-prep and FNA slides. Therefore, TI is only suitable for tile-based delineation of Histology slides without any artefacts.

5.2. Focus Map Evaluation
The ranking results presented in Fig. 6 show that the MSE of focus maps produced using candidates selected via the four metrics eventually surpass that of random selection. This demonstrates that the selection of focus candidates based on information extracted from each individual tile is an effective approach. However, TI is clearly the least effective as it produced results worse than random selection even when a considerable numbers of focus candidate were used (around 50). As Fig. 7 and 8 illustrate, the tiles most highly ranked by TI are primarily outliers, which must subsequently be detected and handled via a robust interpolation scheme (as was done here [10]). However, when the number of outliers is large, they dominate and so no longer appear to be outliers and so will have a significant effect on the focal depth estimated from the focal map. When this is the case even robust regression methods are likely to fail. The other three metrics exhibit performance that is superior to random selection once a reasonable number of focus candidates have been selected (around 15), but only APCI was significantly better than random selection over the complete range of focus candidates used (10 to 100).
Table 5 and Fig. 6 indicate that not only are the focus maps constructed from the focus candidates selected by APCI consistently better (lower MSE) than the other three metrics (TI, NACI and EI), but that APCI requires a smaller number of focus candidates to construct an acceptable focus map. The reasons for this are twofold: Firstly, the scan maps generated using APCI contain less artefacts. For instance, the APCI scan map of the Thin-prep example, shown in Fig. 7, contains the minimum number of artefacts and consequently far fewer "bad" focus candidates, as shown in Fig. 9. In contrast, the scan maps produced by NACI and EI cover the whole artefact area, and so contain more outlier focus candidates contaminated by glue artefacts. Secondly, the artefact affected tiles included in the scan maps, such as the ink markers in the FNA slide in Fig. 8, are given much lower rankings by APCI and so are far less likely to be selected as focus candidates. In fact, as Fig. 9 shows, the highest ranking tiles from APCI are almost exclusively "good" focus candidates that contain primarily specimen. Therefore, based on these results we can conclude that APCI is the best metric considered here for the selection of focus candidates due to its ability to rank in-focus specimen ahead of (out of focus) artefacts.

Figure 9: The APCI has the minimum number of focus candidates that locate on top of the glue artefacts. Notice: The minimum distance between two focus candidates was set to five tiles here for better illustration.
When there are few artefacts, as is the case for the histology specimens, all four metrics produce similar performance as their scan maps are all close to the ground truth and almost all tiles contain clean specimen. For example from Fig. 10, which shows the top ranked tiles from a histology slide, it can be seen that all tiles from all metrics contain a variable degree of specimen that are likely to give valid focal positions. In this situation, the MSE of the focus maps consistently reduces as more focal candidate safe selected and all metrics are equally good. However, Fig. 10 also shows that NACI selects tiles with high contrast, i.e., on the edge of specimen and background, while APCI and EI select tiles containing almost entirely specimen.

From Table 5 it can be seen that APCI produces the minimum SNFC on the majority of specimens. This means that APCI requires a smaller number of focus candidates to construct an accurate focus map. In practice, this means that APCI will minimise the amount of time spent on focusing on high-resolution FOVs, thus minimising scan times. Furthermore, Table 5 shows that on the two Thin-prep slides APCI is the only metric that reliably reaches the sufficient MSE criterion.

Obviously, the higher the resolution of the low-resolution image, the more information can be extracted for tile analysis. However, increasing the resolution, increases both the number of FOVs acquired and the data to be processed. There is clearly a trade-off between the low-resolution scan times and scan map and focal map accuracy. Here we have chosen to compromise on a 4X objective and demonstrated that accurate scan and focus maps can be produced. This not only saves considerable amount of high-resolution scan time, but also avoids wasting time on a re-scan if the high resolution images are found to be unsatisfactory. Particularly for APCI, which measures the phase diversity of the tile, the size of cellular material will be smaller at lower magnifications and so the accuracy of the method may be affected. Therefore, standalone macro camera with a unity (or lower) magnification may not be suitable as the depth of field may be larger than the depth of the specimen.

Future work will include the development of improve threshold criteria and
post-processing algorithms for delineation of specific slide types, which could lead to improved performance on these types of slides. Furthermore, the proposed modification of Otsu's method assumes that there are only two classes (foreground and background). This could be further improved by using a Gaussian Mixture Model that explicitly models multiple classes consisting of different types of cellular material (such as cytoplasm and nucleus) and artefacts (such as ink markers) [7, 14]. The method to date, has only been tested on conventionally stained (Pap and HE) bright-field microscopy images. However, in principle the technique would appear to be suitable for other types of specimens such as those acquired in fluorescence mode.

6. Conclusions
In this paper, we have proposed a novel tile-based algorithm capable of both delineating a microscopic specimen and selecting a subset of FOVs that are good candidates from which to construct a high-resolution focus map. In particular, we investigated four tile evaluation metrics for their suitability for both detecting specimen and highlighting good focus candidates. The experimental results on a set of 40 real-world pathology specimens demonstrated that while the NACI, APCI and EI were superior to TI at specimen delineation, only APCI was capable of selecting good focus candidates while ignoring artefacts.
7. Acknowledgements
The authors would like to thank the anonymous reviewer for their constructive comments on earlier drafts of this paper. Professor Bradley is supported by an Australian Research Council Future Fellowship (FT110100623).
References


Figure 11: Top: low-resolution of a sample slide (Row 1 in Table 5) and the blue box shows the part of the slide used for focus map evaluation. Bottom: The mean square error of the interpolated focus map compared with the ground truth as the number of the sampled focus candidate increase.
Figure 12: Top: low-resolution of a sample slide (Row 2 in Table. 5) and the blue box shows the part of the slide used for focus map evaluation. Bottom; The mean square error of the interpolated focus map compared with the ground truth as the number of the sampled focus candidate increase.
Figure 13: Top: low-resolution of a sample slide (Row 3 in Table 5) and the blue box shows the part of the slide used for focus map evaluation. Bottom: The mean square error of the interpolated focus map compared with the ground truth as the number of the sampled focus candidate increase.
Figure 14: Top: low-resolution of a sample slide (Row 4 in Table. 5) and the blue box shows the part of the slide used for focus map evaluation. Bottom; The mean square error of the interpolated focus map compared with the ground truth as the number of the sampled focus candidate increase.
Figure 15: Top: low-resolution of a sample slide (Row 5 in Table 5) and the blue box shows the part of the slide used for focus map evaluation. Bottom: The mean square error of the interpolated focus map compared with the ground truth as the number of the sampled focus candidate increase.
Figure 16: Top: low-resolution of a sample slide (Row 6 in Table. 5) and the blue box shows the part of the slide used for focus map evaluation. Bottom; The mean square error of the interpolated focus map compared with the ground truth as the number of the sampled focus candidate increase.
Figure 17: Top: low-resolution of a sample slide (Row 7 in Table 5) and the blue box shows the part of the slide used for focus map evaluation. Bottom: The mean square error of the interpolated focus map compared with the ground truth as the number of the sampled focus candidate increase.
Figure 18: Top: low-resolution of a sample slide (Row 8 in Table. 5) and the blue box shows the part of the slide used for focus map evaluation. Bottom: The mean square error of the interpolated focus map compared with the ground truth as the number of the sampled focus candidate increase.