Virtual Double Staining Based Learning for Segmentation of Tumors in Primary Immunohistochemical Biomarkers

Christian Hansen
1 Technical University of Denmark.

Jeppe Thagaard1,2
2 Visiopharm A/S, Hoersholm, Denmark.

Anders Bjorholm Dahl1

Editors: Under Review for MIDL 2019

Abstract
We propose a deep learning-based segmentation method of tumor for WSI biomarker quantification of Ki67. The model is evaluated by comparing with the clinical score and other statistical measures of an existing method based on Virtual Double Staining (VDS). The results show a 98.85% agreement between the two methods with no significant difference.

Keywords: Semantic Segmentation, Whole Slide Images (WSI), Virtual Double Staining (VDS), Ki67, Immunohistochemistry (IHC), Proliferation Index (PI), U-Net & DeepLabv3.

1. Introduction
Breast cancer are categorized by four different molecular subtypes, each with different optimal strategies for treatment. It is therefore crucial to determine the molecular subtype precisely due to the decisive effect on cancer treatment plans (Stålhammar, 2017). Current practice rely on histological evaluation, where multiple immunohistochemical markers are quantified to determine proliferation (Ki67), hormone receptors status (ER, PR) and HER2-enrichment (Duffy et al., 2017). Due to the possible expression of the nuclear markers in normal cells, it becomes imperative that the quantification is only carried out in tumor regions to obtain precise subtype evaluation. Clinical image analysis methods therefore involve either pathologist selecting regions for quantification manually, which is time consuming and has been shown to suffer from great inter-observer variability (Stålhammar et al., 2016) or an automated method called Virtual Double Staining (Visiopharm A/S, c) which requires additional histological processing in the lab and whole-slide registration of serial sections. In this study, we aim to develop a method for automated segmentation of tumor regions in order to ease biomarker quantification while also exploring the possibility of using VDS to overcome one of the major obstacles in deep learning for histopathology; insufficient ground truth (Komura and Ishikawa, 2018).

2. Method & Materials
VDS requires the use of an adjacent tissue sample stained with pancytokeratin (PCK), a protein which is expressed by epithelial breast cancer cells and therefore enables separation of tumor and stroma. We do not consider the discrimination between invasive and...
We automatically detect the positive PCK regions (DAB staining) with a simple image analysis algorithm (Visiopharm A/S, d). After WSI registration in Tissuealign™ (Visiopharm A/S, b), we transfer the PCK positive regions to the adjacent Ki67 stained WSI thus limiting the biomarker quantification to the detected tumor regions. Quantification of Ki67 stains in tumor regions is then done automatically (Visiopharm A/S, a) by assessing the amount of positive and negative stained nuclei and calculating the proliferation index according to Equation (1):

$$\text{Proliferation Index (PI)} \ [%] = \frac{100 \cdot \text{Pos Nuclei [\#]}}{\text{Pos Nuclei [\#]} + \text{Neg Nuclei [\#]}}$$

(1)

To obtain a method that requires less manual labour and is less sensitive to alignment issues, we propose the approach shown in Figure 1 using VDS to create ground truth annotations of WSI as training input for the two tested architectures; U-Net (Ronneberger et al.) and DeepLabv3+ (Chen et al.). Since no manual ground truth exist, the performance of the segmentation by the network is evaluated by calculating the biomarker quantification (PI) and comparing with the PI of the approved VDS approach described above. We included 46 WSI stained with Ki67 together with the adjacent PCK sections obtained from three different hospitals. We manually assessed the alignment results and large artifact regions such as holes of tissue samples, blurs, markers etc. were withheld from the analysis. 80% of the data (37 WSI) were used for training and the remaining 20% (9 WSI) for testing. However to obtain a larger test set, we split the 9 test WSI into 10 field-of-views (FOVs) of 512x512 pixels at a 5X magnification, thus obtaining a total of 90 test images.

Figure 1: The overall setup highlighting the VDS principal where a biopsy sample is stained with Ki67 and the adjacent tissue slide with PCK. The WSIs are afterwards digitized and registered so tumor outlines detected in the PCK marker can be precisely transferred to the Ki67 stained WSI to achieve fast and precise tumor and stroma training labels.

3. Results and discussion

We compare PI obtained by best neural network with the usual VDS method using Passing-Bablock (PB) regression, Bland-Altman analysis, the Wilcoxon signed-rank test and the clinical outcome to assess any significant differences. We exclude three outliers from the
statistical analysis because of registration misalignment or the segmentation of very small areas with very few nuclei leading to a very large PI. The resulting plots in Figure 2 shows the regression line follows the optimal line of equality. The Wilcoxon signed rank test yielded a p-value of 0.213 further indicating no significant differences in the two methods. Furthermore, the Bland-Altman plot indicates no systematic differences, however we observe five points outside the lines of agreement also indicated by deviations in the PB plot. Two of these are false positive segmentation by the proposed method, one due to slight misalignment of VDS and the remaining because our method struggles to detect smaller areas of tumor in adipose tissue. In Table 1, the agreement between the two methods is found using a cutoff value for PI of 14%. 1 of the 87 cases is affected in clinical outcome due to over-segmentation, yielding a high kappa value indicating very good agreement between the methods.

![Figure 2: Comparison of the method usually used by Visiopharm A/S based on VDS and the new method based on VDS training with data augmentation of a U-Net.](image)

<table>
<thead>
<tr>
<th></th>
<th>VDS</th>
<th>DL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>87</td>
<td>-</td>
</tr>
<tr>
<td>VDS</td>
<td>63</td>
<td>1</td>
</tr>
<tr>
<td>+</td>
<td>0</td>
<td>23</td>
</tr>
<tr>
<td>Agreement (%)</td>
<td>98.85</td>
<td></td>
</tr>
<tr>
<td>Cohen’s Kappa</td>
<td>0.971</td>
<td></td>
</tr>
</tbody>
</table>

Table 1: Comparison of clinical outcome between VDS and the Deep Learning based method, notice one test image is classified differently.

4. Conclusion

The results evidently demonstrate good agreement between the two methods thus successfully creating a less demanding approach for biomarker quantification in breast cancer. We also show that VDS-based ground truth annotations demonstrates to be an alternative to obtain large scale annotations for training deep learning networks in histopathology. For future work, we aim to lower false positives with more extensive evaluation of segmentation performance, improve VDS labeling by automatically detecting misalignment and expand the method to ER and PR while also incorporating discrimination between invasive and non-invasive tumor components.
Conflict of Interest

J. Thagaard is an employee of Visiopharm A/S, Hoersholm, Denmark.

References


