Article

ICOSeg: Real-Time ICOS Protein Expression Segmentation from Immunohistochemistry Slides Using a Lightweight Conv-Transformer Network

Vivek Kumar Singh [1,*], Md. Mostafa Kamal Sarker [2], Yasmine Makhlouf [1], Stephanie G. Craig [1], Matthew P. Humphries [1,3,4], Maurice B. Loughrey [1,5], Jacqueline A. James [1,6,7], Manuel Salto-Tellez [1,6,8], Paul O'Reilly [1,9] and Perry Maxwell [1]

1   Precision Medicine Centre of Excellence, The Patrick G Johnston Centre for Cancer Research, Queen's University Belfast, Belfast BT9 7AE, UK
2   National Subsea Centre, Robert Gordon University, Aberdeen AB21 0BH, UK
3   Leeds Teaching Hospitals NHS Trust, Leeds LS1 3EX, UK
4   Leeds Institute of Cancer and Pathology, University of Leeds, St. James' University Hospital, Leeds LS9 7TF, UK
5   Cellular Pathology, Belfast Health and Social Care Trust, Belfast City Hospital, Lisburn Road, Belfast BT9 7AB, UK
6   Regional Molecular Diagnostic Service, Belfast Health and Social Care Trust, Health Sciences Building, Queen's University Belfast, Belfast BT9 7AE, UK
7   Northern Ireland Biobank, The Patrick G Johnston Centre for Cancer Research, Queen's University Belfast, Belfast BT9 7AE, UK
8   Division of Molecular Pathology, The Institute of Cancer Research, Sutton SM2 5NG, UK
9   Sonrai Analytics Ltd., Lisburn Road, Belfast BT9 7BL, UK
*   Correspondence: v.singh@qub.ac.uk

Simple Summary: Inducible T-cell COStimulator (ICOS) is a biomarker of interest in checkpoint inhibitor therapy, and as a means of assessing T-cell regulation as part of a complex process of adaptive immunity. The aim of our study is to segment the ICOS positive cells using a lightweight deep-learning segmentation network. We aim to assess the potential of a convolutional neural network and transformer together that permits the capture of relevant features from immunohistochemistry images. The proposed study achieved remarkable results compared to the existing biomedical segmentation methods on our in-house dataset and surpassed our previous analysis by only utilizing the Efficient-UNet network.

Abstract: In this article, we propose ICOSeg, a lightweight deep learning model that accurately segments the immune-checkpoint biomarker, Inducible T-cell COStimulator (ICOS) protein in colon cancer from immunohistochemistry (IHC) slide patches. The proposed model relies on the Mobile-ViT network that includes two main components: convolutional neural network (CNN) layers for extracting spatial features; and a transformer block for capturing a global feature representation from IHC patch images. The ICOSeg uses an encoder and decoder sub-network. The encoder extracts the positive cell's salient features (i.e., shape, texture, intensity, and margin), and the decoder reconstructs important features into segmentation maps. To improve the model generalization capabilities, we adopted a channel attention mechanism that added to the bottleneck of the encoder layer. This approach highlighted the most relevant cell structures by discriminating between the targeted cell and background tissues. We performed extensive experiments on our in-house dataset. The experimental results confirm that the proposed model achieves more significant results against state-of-the-art methods, together with an $8\times$ reduction in parameters.

Keywords: colon cancer; immunohistochemistry; ICOS; deep learning; channel attention

1. Introduction

According to the World Health Organization (WHO), colorectal cancer (CRC) is the third most fatal cancer, causing approximately 0.9 million deaths each year. By the year 2030, the number of new cases is anticipated to rise to 2.2 million per year, resulting in 1.1 million fatalities worldwide [1]. An early diagnosis with proper screening could help to prevent colorectal cancer. Additionally, early identification may provide a better treatment response to prevent cancer cells spread and save lives.

The number of molecular stratified therapy opportunities targeting metastatic CRC is rising with the application of molecular biomarkers to benefit prognosis and treatment decision-making [2]. For instance, patients with metastatic microsatellite instability (MSI-H)/mismatch repair deficient (dMMR) tumors can benefit from immunotherapy, and MSI/MMR testing can be employed as a marker of genetic instability with therapeutic implications [3]. Other innovative therapeutic medicines are approved based on the results of a companion biomarker, such as PD-L1 immunohistochemistry (IHC) [4]. Translational oncology has been improved significantly by tissue-based biomarker analysis using IHC, which maintains spatial and cell-specific information, and allows for reliable biomarker expression analysis inside the tumor micro-environment [5]. A pathologist can assess the quantification of biomarker-expressing cells and their location to provide significant prognostic patient information [6,7]. As a means of evaluating biomarkers in a high volume/high throughput manner, Tissue microarrays (TMA) can be incorporated into the development pipeline of biomarker research [8,9]. Ultimately, manual TMA examination with IHC is a long and laborious procedure in evaluating biomarker-expressing cells, making it unsuitable for effective TMA investigation [10,11]. As a result, a Computer-Aided Diagnosis (CAD) system is necessary for examining the TMA and helping to train pathologists in their investigation workflows.

With today's CAD image analysis technology, large-scale quantitative evaluation of IHC on TMAs is feasible. Recently, there has been a lot of research conducted on the advantages of computer-assisted quantitative cell count studies [12], cell segmentation tasks [13], and the effect of the automatic versus manual estimation [14]. Current artificial intelligence (AI) guided CAD systems show promising outcomes using deep learning algorithms in digital pathology, for example in predicting colon cancer metastases [15], nuclei segmentation [16], lymphocyte identification [17], in assessing tumor grade [18], and predicting molecular signatures [19]. Most of the developed CAD systems focus on identifying tiles in whole-slide images. These CAD systems have some limitations for successfully segmenting biomarker-expressing cells on IHC-stained tissue images. In our previous work [20], the evaluation of immune and immune-checkpoint biomarkers proposed utilizing deep learning techniques by automated identification of cells positive for the ICOS. The results were encouraging, but the Efficient-UNet model utilized a huge number of parameters (i.e., 66 million), making them difficult to implement in real-world clinical situations.

There is always a trade-off between the model complexity in terms of parameters and its performance. Usually, an attempt to reduce the number of trainable parameters through a pruning strategy or adopting a lightweight model may achieve the lower results. Therefore, it is important to design a model with lower computational complexity while keeping a similar performance or achieving better results. The lightweight model deployment creates many opportunities in various applications in pathology. It offers the use of a digital image without requiring additional processing resources, and allows to perform image analysis. Moreover, the application of this work will help to deploy on a cloud server where a pathologist can easily access to segment the positive cells from IHC slides.

We propose a lightweight model viz ICOSeg in segmenting ICOS positive cells. ICOS is a biomarker of interest in checkpoint inhibitor therapy and as a means of assessing T-cell regulation as part of a complex process of adaptive immunity. We merge the convolutional neural network (CNN) and transformer to capture the local and global feature representation with lower computational complexity. Figure 1 shows examples of patches,

which highlight the ICOS positive cells in brown extracted from IHC slides. This study is an extension of our previous work [20] and shows higher computational complexity for similar tasks.
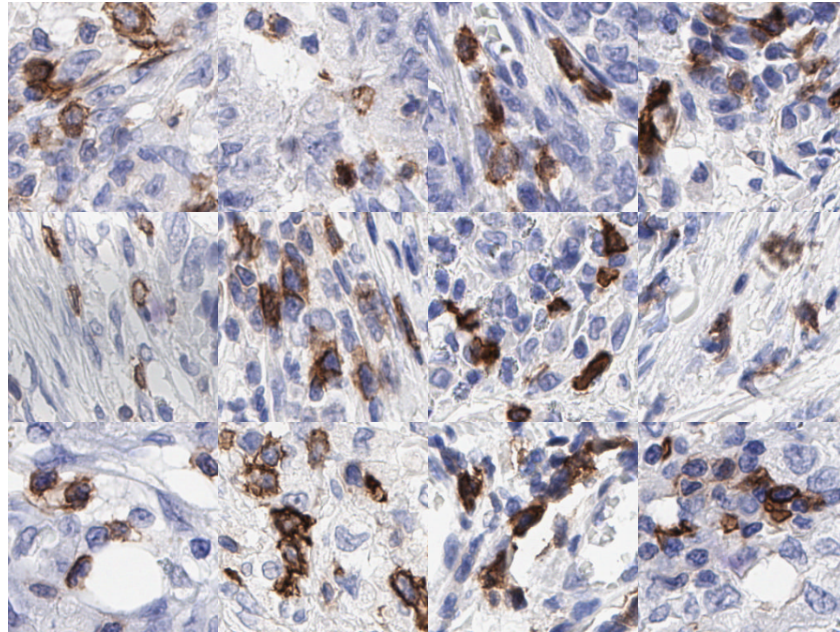

Figure 1. Examples of IHC patches extracted at $40\times$ magnification containing ICOS positive (i.e., brown cytoplasmic and blue nuclear staining) and ICOS negative (blue nuclear stain only) cells in colon cancer.

The main contributions of this paper are listed as follows:

- We propose an efficient, lightweight segmentation method called ICOSeg in segmenting ICOS positive cells that combines the CNN and transformer as a feature extractor with a lower latency rate. The proposed model extracts the local spatial features through CNN, and global representation learned using transformer block;
- We use the channel attention mechanism in the final encoder layer that enables the network to extract rich features and discriminate between the targeted cell's structure and background pixels;
- We perform extensive experiments incorporating ablation studies and employ four pixel-wise evaluation metrics (i.e., Dice coefficient, aggregated Jaccard index, sensitivity, and specificity) that confirm the effectiveness of the proposed method;
- Experimental results confirm that ICOSeg efficiently outperforms several state-of-the-art methods (i.e., U-Net, Attention U-Net, FCN, DeepLabv3+, U-Net++, and Efficient U-Net) with lower trainable parameters.

The remainder of this paper is structured as follows. Section 2 discusses the proposed ICOSeg method in detail. The dataset description with experimental plans and limitations of our work is explained in Section 3. We conclude by summarizing our findings and suggest some future lines of research in Section 4.

2. Methods

In this section, we present the architectural details of the proposed ICOSeg segmentation model motivated by the existing work [21].

2.1. ICOSeg Architecture

The proposed model consists of an encoder and decoder network. The IHC slide patch containing cells feeds to encoding layers, extracting the relevant spatial and global features. However, decoding layers allows reconstructing the feature representation into a

segmentation map. We explore the encoder structure by [21], that merges the potential of learning local and global features of convolutions and transformers, respectively. Figure 2 shows the general overview of the proposed ICOSeg positive cell segmentation model.
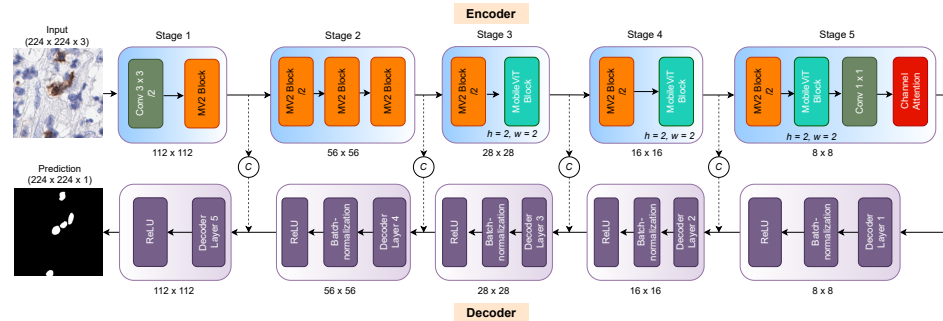

Figure 2. Overview of proposed segmentation model. The input IHC patch image feed to the encoder layer that extracts the spatial and global feature information. The decoder reconstructs the features into the segmentation map. Where *C* refers to the feature concatenation. The dash (- -) line corresponds to the skip connection between each encoder and decoder layers.

Encoder: This consisted of five key stages using convolutional and MobileVit blocks. The first stage combined the two layers with a strided 3×3 convolutional layer that helped to extract the spatial features from the input patch images. The extracted feature maps are down-sampled by the factor of 2. MobileNetv2 (i.e., MV2) block [22] helped encode the cell's feature maps in low-dimensional spaces and advanced feature representation power. Note that MV2 blocks in the encoder were mainly responsible for the down-sampling of the resolution of the intermediate feature maps. Stage two incorporated three MV2 blocks that allowed for enriching the spatial features and applied feature down-sampling. Stage three integrated the MV2 block and MobileViT block which, together, captured the global representation of the features. The MobileViT block learned contextual information with less computational complexity. In addition, it merged the advantage of Transformer and convolution. Stage four follows the stage three combination and featured maps downsampled by a factor of 2. The final stage combined the MV2 block, MobileViT block, and point-wise convolutional layers, which provided the bottleneck features that fed to the decoder layers. All the encoding layers helped to extract salient features, such as shape, texture, intensity, and margin. The last encoder layer used the channel attention [23] mechanism that highlights the targeted cell-related features and distinguished between relevant and the background pixels. It dynamically recalibrated channel-wise feature maps feedback through modeling precisely the interdependencies between channels.

The main functionality of a MobileViT block required an input tensor $I \in \mathbb{R}^{h \times w \times}$ and involved an $n \times n$ convolutional layer followed by a point-wise convolution to create the map of $I_L \in \mathbb{R}^{h \times w \times d}$. This convolutional layer enabled learning feature representations to capture local information, such as shape or intensity. To learn the global representation, unfolded I_L into another N non-overlapped vector (i.e, patch) with a shape $I_U \in \mathbb{R}^{p \times N_{patch} \times d}$. Where p is the area of a small patch and $N_{patch} = \frac{hw}{p}$ in which h and w refer to the height and width of that specific patch. To overcome the issue of not losing the spatial order of pixels, inter-patch connections are encoded using Transformers.

$$\mathbf{I}_G(p) = \text{Transformer}\left(\mathbf{I}_U(p)\right), 1 \leq p \leq P \tag{1}$$

where $p \in \{1...P\}$. Further, the output vector I_g was again folded into a vector of shape to receive $I_F \in \mathbb{R}^{h \times w \times d}$. This output feature vector used the point-wise (1×1) convolutional layer to project into a low dimensional space. This resultant feature was concatenated with the input feature map of I. To merge these concatenated features, an $n \times n$ convolutional layer was applied.

Decoder: We employed five decoding layers with a kernel size of 4×4 and stride of 2×2 that downsampled the features by two. Each decoder layer was followed by batch normalization and the non-linear ReLU activation function helped to confound the overfitting. The skip connection was used in between the encoding and decoding layers to narrow down the semantic gaps during feature reconstruction at the decoding stage. We used a threshold value of 0.5 to generate the binary segmentation map.

2.2. Loss Function

In this study, we used the combination of Binary cross-entropy (BCE) and Dice loss functions. The BCE loss can be formulated as follows:

$$\mathcal{L}^{\text{BCE}}(gt_{mask}, p_{mask}) = -(gt_{mask} \cdot log(p_{mask}) + (1 - gt_{mask}) \cdot log(1 - p_{mask})), \qquad (2)$$

Additionally, the Dice loss can be explained as follows:

$$\mathcal{L}^{\text{Dice}}(gt_{mask}, p_{mask}) = 1 - Dice(gt_{mask}, p_{mask}) = 1 - \frac{2|gt_{mask}| \cdot |p_{mask}|}{|gt_{mask}|^2 + |p_{mask}|^2}, \qquad (3)$$

Therefore, the total loss is the weighted summation of $\mathcal{L}^{\text{BCE}}$ and $\mathcal{L}^{\text{Dice}}$ described below:

$$\mathcal{L}^{\text{Total}}(gt_{mask}, p_{mask}) = \gamma \mathcal{L}^{\text{BCE}}(gt_{mask}, p_{mask}) + (1 - \gamma)\mathcal{L}^{\text{Dice}}(gt_{mask}, p_{mask}) \qquad (4)$$

where gt_{mask} is the ground truth mask, p_{mask} corresponds to predicted mask generated by the model, and γ is weighting factor. Here, we used the γ equal to 0.4.

3. Experimental Results and Discussion

3.1. Dataset

In this article, we used the ICOS IHC data from our previous research [20]. ICOS-stained IHC slides were scanned with 40× obj. magnification. These are relatively large in pixels fed into the input of deep models. Therefore, we generated definitive input patches that were employed to train models. The patches were 256×256 pixels in size. The entire dataset was split into training, validation, and test sets with the ratio of 60%, 10%, and 30%, respectively, which is the same split, that was used in the previous work [20].

3.2. Training Details

We used the input patch and rescaled with the size of 224×224 pixels. The extracted patches were normalized from 0–255 pixel values to the range 0–1. An ADAM optimizer with an initial learning rate of 0.0001 was employed to optimize the proposed model reasonably. During the training, we applied the step decay if the dice coefficient on a validation set did not increase two consecutive epochs . We used a batch size of 8 patch images and trained the model from scratch for 50 epochs. To increase the samples size, we performed the data augmentation approach consisting of rotation at 90 degrees, horizontal flipping, and scaling with a probability of 0.5. The selected strategy was used to mitigate overfitting and provide more variability of features in the training network.

Model development platform: We developed the model using the PyTorch neural network library on the National High-Performance Computing (HPC) Kelvin-2 managed by Queen's University Belfast (QUB). The model was trained on the 32 GB GPU memory with CUDA version 11.2. We evaluated the model on our local workstation, including GPU RTX2080 Ti of 11 GB video RAM. It is worthwhile to note that all the computational measurement and frame per second (FPS) speed was computed on a local workstation.

3.3. Evaluation Metrics

We pixel-wise assessed each method's segmentation results using four evaluation metrics. It includes the dice coefficient score (Dice), Aggregated Jaccard Index (AJI), sensitivity, and specificity. These metrics were measured using true positive (TP), false

positive (FP), false-negative (FN) rates, and true negative (TN). We assumed gt_{mask} and p_{mask} are the ground truth mask and predicted mask generated by the model for the same input patch images, respectively. The common area of segmented cells in gt_{mask} and p_{mask}, referred to $TP = gt_{mask} \cap p_{mask}$. The segmented region which is out of the common region to gt_{mask} corresponds to the FP that is defined as $\overline{gt}_{mask} \cap p_{mask}$. Further, FP is the region defined by the ground truth that is not segmented through the model formulated as $gt_{mask} \cap \overline{p}_{mask}$. Lastly, TN enfolds the background pixels.

The employed metrics can be formulated as follows:

$$Dice = \frac{2TP}{(2TP + FP + FN)}, \tag{5}$$

$$Sensitivity = \frac{TP}{(TP + FN)}, \tag{6}$$

$$Specificity = \frac{TN}{(TN + FP)}, \tag{7}$$

The aggregated Jaccard index (AJI) can be defined as:

$$\text{AJI} = \frac{\sum_{i=1}^{M} \left| gt_{maski} \cap p_{maskj}^{*}(i) \right|}{\sum_{i=1}^{S} \left| gt_{maski} \cup p_{maskj}^{*}(i) \right| + \sum_{S \in \text{Ind}} p_{maskS}} \tag{8}$$

where, gt_{mask_t} is the ith ground-truth mask of ICOS cells pixels, p_{maskS} is the predicted cells segmentation mask generated through the model, $p_{maskj}^{*}(i)$ is the connected component from the predicted binary mask that advances the Jaccard index, and *Ind* is the list of indices of pixels that do not belong to any element in the ground-truth pixels.

3.4. Results

In this subsection, we present and discuss the segmentation results of the proposed model with several state-of-the-art methods. We conducted an ablation study including the effect of attention mechanism and the effect of the loss functions.

3.4.1. Effect of Attention Block

Table 1 shows the performance of the proposed model with and without the attention block employed for ICOS positive cells segmentation. We first built the segmentation model without adding the attention block called Baseline, which only combines the plain encoder and decoder structure. The developed baseline model obtained the Dice score of 75.09% and AJI of 59.67%. On seeing the great success of the attention technique on multiple datasets however, we also merged it into our encoder layers. We achieved a better segmentation result than Baseline since the attention block highlighted the most relevant features and ignores the irrelevant ones. Defined kernels must focus on the target nuclei region. Some artifacts however, due to the dye and blurring in certain areas could infer the results. The attention approach enables the filters to focus on capturing the appropriate features and learn the pattern to discriminate between artifacts and nuclei. The proposed model (i.e., baseline + attention block) improved 1% in all the computed metrics. The attention block did not add any extra burden on computational complexity. The proposed model allowed segmentation of the patches with 154 frames per second (FPS) on the current GPU.

Table 1. Examining the effect of adopted attention block on segmentation result (mean ± standard deviation) incorporated with Baseline model. The best results are highlighted in bold.

Model	Metrics				Para (M)	FPS
	Dice	AJI	Sensitivity	Specificity		
Baseline	75.09 ± 13.52	59.67 ± 12.90	81.34 ± 15.99	99.54 ± 0.43	8.1	161
Proposed	**76.01 ± 13.35**	**60.88 ± 12.85**	**82.45 ± 15.96**	**99.61 ± 0.40**	**8.1**	**154**

3.4.2. Effect of the Loss Function

Table 2 demonstrates the ablation study conducted to measure the effect of a different combination of loss functions on the proposed model. The weighted sum of $\mathcal{L}^{\text{BCE}}$ + $\mathcal{L}^{\text{Dice}}$ outperformed the segmentation results in all metrics than other individual loss function. This combined loss function achieved the 2–3% better performance in Dice and AJI scores. Additionally, advance the model sensitiveness by 4% compared to only $\mathcal{L}^{\text{BCE}}$. The $\mathcal{L}^{\text{BCE}}$ obtained lower scores than the proposed combination and generated more false-positive with higher standard deviation. In turn, $\mathcal{L}^{\text{Dice}}$ reduces the pixel-wise false positives leading to better segmentation results. Overall, the combined loss functions provide more generalizability and help converge the model faster with minimal error.

Table 2. Ablation study of the loss functions. The best results are highlighted in bold.

Model	Metrics				Para (M)	FPS
	Dice	AJI	Sensitivity	Specificity		
$\mathcal{L}^{\text{BCE}}$	73.28 ± 15.77	57.41 ± 14.55	78.82 ± 16.61	99.23 ± 0.70	8.1	154
$\mathcal{L}^{\text{Dice}}$	74.97 ± 13.83	59.30 ± 13.72	80.65 ± 16.08	99.58 ± 0.41	8.1	154
$\mathcal{L}^{\text{BCE}}$ + $\mathcal{L}^{\text{Dice}}$ **(Proposed)**	$\mathbf{76.01 \pm 13.35}$	$\mathbf{60.88 \pm 12.85}$	$\mathbf{82.45 \pm 15.96}$	$\mathbf{99.61 \pm 0.40}$	**8.1**	**154**

3.4.3. State-of-the-Art Result Comparison

Table 3 presents the result of the proposed model and compared with six state-of-the-art methods. Specifically, we employed the U-Net [24], FCN [25], attention U-Net [26], DeepLabv3+ [27], U-Net++ [28], and Efficient U-Net [20]. The proposed model outperformed the compared methods in segmenting the ICOS positive cells with AJI, sensitivity, and Dice coefficient scores of more than 76%. The second highest U-Net++ achieved the Dice score of 74.93% and AJI with 55.20%, which is 1–5% lower than our proposal. However, U-Net, attention U-Net, and FCN secured similar results in segmenting the positive cells. The proposed model surpasses the DeepLabv3+ with 4.6% and 9.5% on Dice and AJI scores, respectively. We compared the proposed model with previous work performed by [20]. Our model results significantly exceed the earlier work with a margin of 2.5% and 2% in Dice score and AJI. Additionally, we computed each method of computational complexity with corresponding FPS at the inference level. We observed the proposed model shows the lowest trainable parameters of 8.1 M among all and attained the second-best 154 FPS on current GPU. Conclusively, the significant benefit of incorporating the CNN in local level features extraction and Transformer block with attention mechanism allows capturing a global representation of cells structure. The other methods only considered the standard CNN to work well but exhibit some limitations to segmenting the challenging cell structures in the presence of undefined boundaries or artifacts. These issues, at some level addressed by the proposed model, lead to better segmentation results than the rest.

Table 3. Segmentation results (mean ± standard deviation) of the proposed model compared with six state-of-the-art methods. The best results are highlighted in bold.

Model	Metrics				Params (M)	FPS
	Dice	AJI	Sensitivity	Specificity		
U-Net [24]	73.66 ± 12.61	53.49 ± 12.75	70.69 ± 15.46	99.71 ± 0.29	34.5	116
FCN [25]	73.79 ± 12.59	53.98 ± 12.57	72.01 ± 15.7	99.68 ± 0.35	135.53	57
Attention U-Net [26]	74.91 ± 13.58	55.66 ± 13.07	74.14 ± 16.05	99.65 ± 0.36	34.87	106
DeepLabv3+ [27]	71.38 ± 14.63	51.21 ± 14.19	66.68 ± 17.78	99.74 ± 0.27	59.33	103
U-Net++ [28]	74.93 ± 12.31	55.20 ± 12.57	72.74 ± 15.40	99.69 ± 0.31	9.61	168
Efficient-UNet [20]	73.44	58.98	81.81	99.30	66.0	83
Proposed	$\mathbf{76.01 \pm 13.35}$	$\mathbf{60.88 \pm 12.85}$	$\mathbf{82.45 \pm 15.96}$	$\mathbf{99.61 \pm 0.40}$	**8.1**	**154**

We measured descriptive statistics of each method's results using box-plot analysis. Figure 3 shows the box-plots generated with the help of the Dice coefficient and AJI scores obtained by all the evaluated methods. The multiple color boxes highlight the score range of different segmentation methods; the black line inside each box shows the median value, box limits incorporate interquartile spans Q2 and Q3 (from 25% to 75% of samples), upper and lower whiskers are measured as 1.5× the distance of upper and lower limits of the box, and all values outside the whiskers are assumed outliers. We noticed that the proposed model achieved better results than the rest. However, compared methods generated multiple outliers with the lowest Dice and AJI scores.
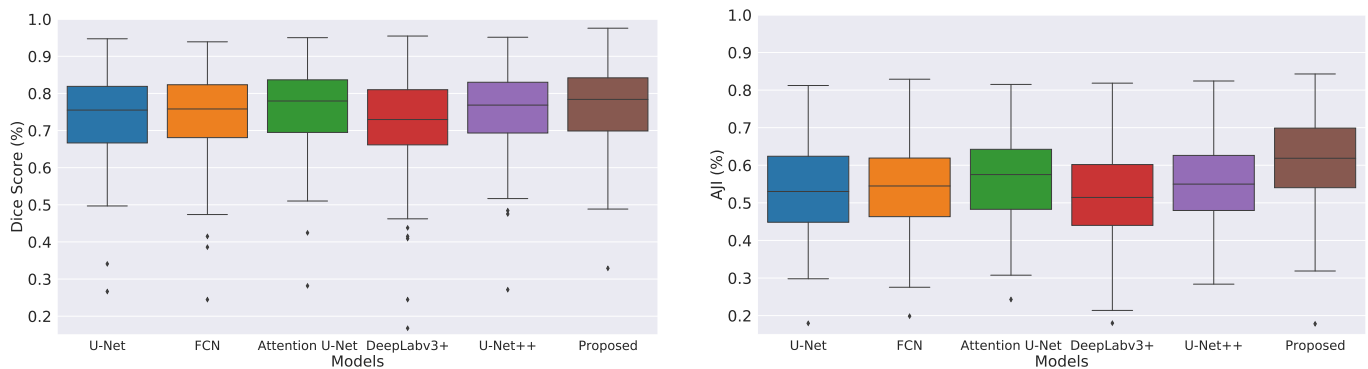
Figure 3. Boxplots of Dice and AJI scores on test dataset. Multiple color boxes highlight the score range of different segmentation methods; the black line inside each box shows the median value, box limits incorporate interquartile ranges Q2 and Q3 (from 25% to 75% of samples), upper and lower whiskers are measured as 1.5× the distance of upper and lower limits of the box, and all values outside the whiskers are assumed outliers.

3.4.4. Qualitative Evaluation

The eight qualitative examples obtained by the proposed ICOSeg for ICOS positive cell segmentation are depicted in Figure 4. We used color maps to represent the following description: orange/yellow denotes true positive, green denotes false positive, and red indicate false negative. All the examples presented yielded better results. Visual inspection confirms that ICOSeg generates very few false positives. For any deep neural network, splitting the two connected cells boundaries containing similar dye artifact in certain area is challenging task. The proposed method attempted to capture the majority of the positive cells in patch images.
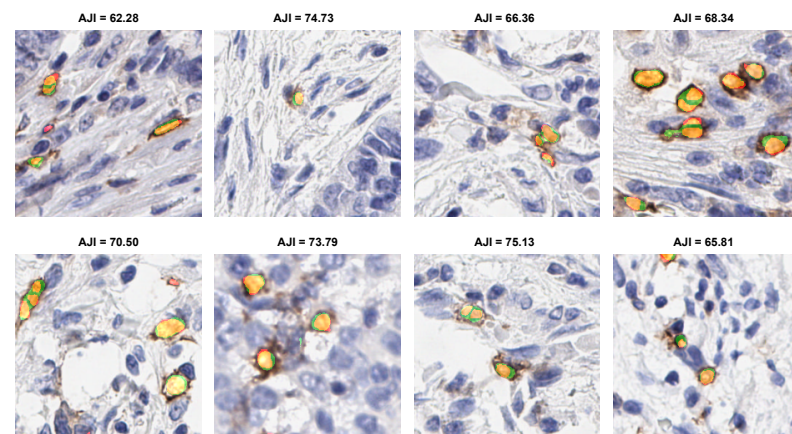
Figure 4. Illustration of proposed model segmentation result. Color maps shows the following description: orange/yellow (true positive), green (false positive), and red (false negative).

In addition, we provided a comprehensive comparison of ICOSeg with six state-of-the-art methods. Images with the associated mask shown to visualize the cell's textural patterns and boundaries. Figure 5 represents the two qualitative examples for ICOS positive cell segmentation. This visual finding, combined with quantitative evidence, confirms that the proposed model outperforms the other methods. Figure 5 (i.e., example 1) attempted to segment the small cell boundaries correctly; however, when compared methods performed well but produced more false positives. Similarly, example 2 depicts undefined cell boundaries. The ambiguous cell boundaries were not segmented by any of the six methods, which resulted in it being identified as a single structure. With the combined effect of CNN and transformer, the proposed model allows better feature representation in situations where other models do not show significant improvement.
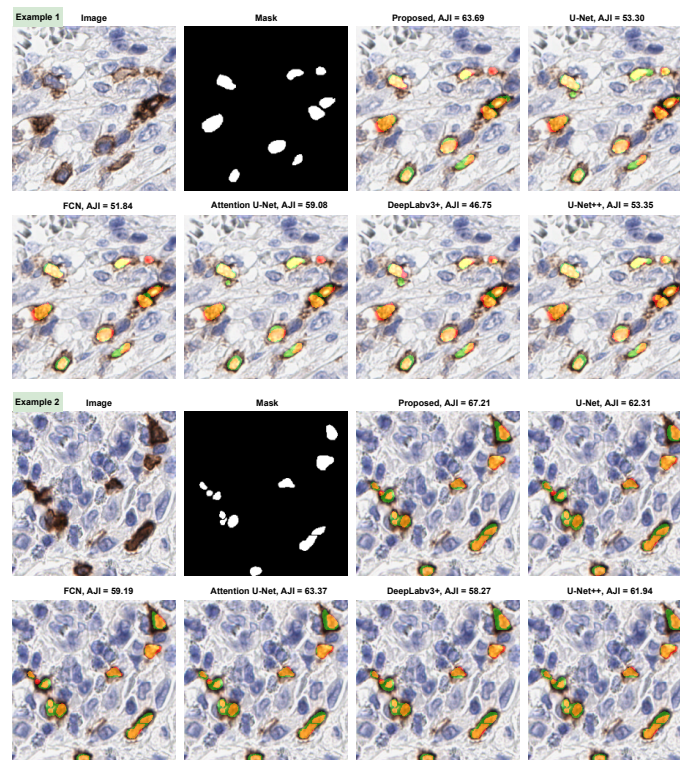
Figure 5. Illustration of proposed model segmentation results compare with five state-of-the-art methods. Color maps shows the following description: orange/yellow (true positive), green (false positive), and red (false negative).

Figure 6 provides descriptive statistics of the total cell count identified by ICOSeg compared to actual cells of the ground-truth. We count the total number of cells in each patches by finding its contour from image processing operations. There are 642 cells marked by the pathologist in a test set. However, ICOSeg correctly identified the 495 cells with the same set of patches. This plot shows the majority of the patch contains an average of five to seven patches. However, some patches have more than 10 cells represented as outliers in a distribution plot.
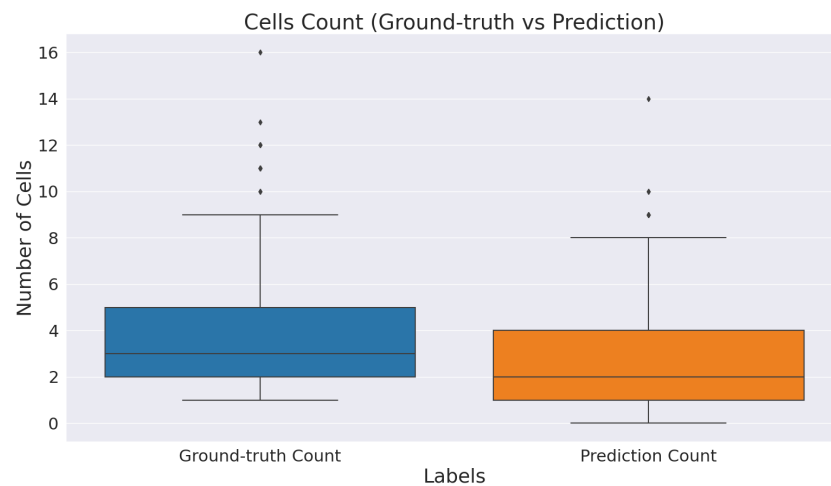
Figure 6. Boxplot of total cell identified by ICOSeg compared to actual cells of the ground-truth.

3.5. Limitation

Figure 7 shows four qualitative examples that illustrate the limitations of the proposed model on IHC patch images. Example one exhibits an area representing the structure of the normal cell. However, it was marked by a pathologist but failed to recognize by ICOSeg. Next, the second example (see in Figure 7b) has two cells marked in which only one of them correctly identified. The top-left cell presents the structure that more correlates to the normal cells, the proposed model failed to recognize it which was clinically found to be an outlier. In addition, the third and fourth examples (see in Figure 7c,d) show similar characteristics of some positive cells' ambiguous boundaries that the model failed to provide precise segmentation.
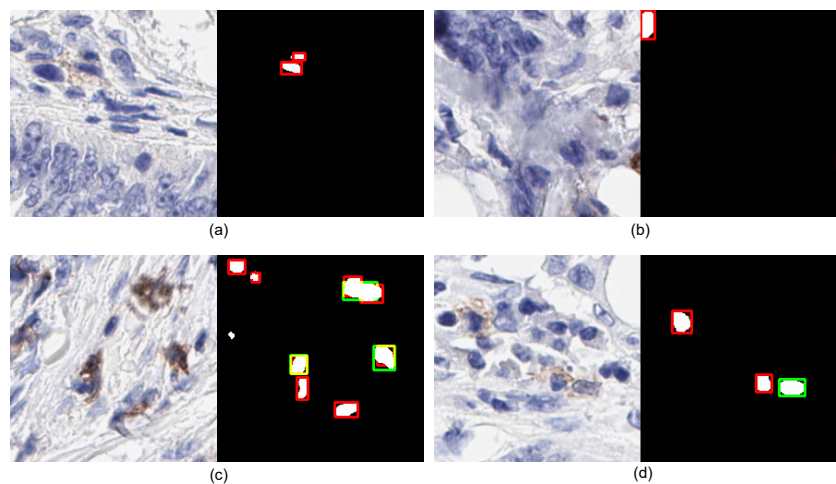
Figure 7. Illustration of four examples shows the limitation of ICOSeg failed to accurately identify and segment the cells boundaries. The red refers to the ground-truth, green corresponds to the ICOSeg prediction and orange/yellow represents the overlap of cell boundary both are common in ground-truth and prediction. The four different examples (**a–d**) refers to the positive cells in brown.

4. Conclusions

In this article, we presented an automated lightweight ICOSeg model for ICOS positive cell segmentation. This study is an extension of our previous work that demonstrates higher computational complexity for similar tasks. Our proposed method differs from earlier research findings in several ways. We included the use of Transformers in the encoder for global feature representation with lower computational complexity and attained better segmentation results compared to CNN alone. We used the feature extractor motivated

by the MobileViT by adding a channel attention mechanism at the last encoding layer and built the decoder sub-network to reconstruct the features into a binary segmentation map. The proposed model outperformed multiple state-of-the-art segmentation methods with less trainable parameters with higher FPS speed. The lower computational complexity allowed the feasibility of translating the model into a limited resource system. The future work would include extending the potential of our proposed model to multiple protein cell datasets for other cancer types.

References

1. Sung, H.; Ferlay, J.; Siegel, R.L.; Laversanne, M.; Soerjomataram, I.; Jemal, A.; Bray, F. Global cancer statistics 2020: GLOBO-CAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **2021**, *71*, 209–249. [CrossRef] [PubMed]
2. Sveen, A.; Kopetz, S.; Lothe, R.A. Biomarker-guided therapy for colorectal cancer: Strength in complexity. *Nat. Rev. Clin. Oncol.* **2020**, *17*, 11–32. [CrossRef] [PubMed]
3. Li, K.; Luo, H.; Huang, L.; Luo, H.; Zhu, X. Microsatellite instability: A review of what the oncologist should know. *Cancer Cell Int.* **2020**, *20*, 1–13. [CrossRef] [PubMed]

4.   Humphries, M.P.; McQuaid, S.; Craig, S.G.; Bingham, V.; Maxwell, P.; Maurya, M.; McLean, F.; Sampson, J.; Higgins, P.; Greene, C.; et al.   Critical appraisal of programmed death ligand 1 reflex diagnostic testing: Current standards and future opportunities. *J. Thorac. Oncol.* **2019**, *14*, 45–53. [CrossRef] [PubMed]

5.   Tan, W.C.C.; Nerurkar, S.N.; Cai, H.Y.; Ng, H.H.M.; Wu, D.; Wee, Y.T.F.; Lim, J.C.T.; Yeong, J.; Lim, T.K.H. Overview of multiplex immunohistochemistry/immunofluorescence techniques in the era of cancer immunotherapy. *Cancer Commun.* **2020**, *40*, 135–153. [CrossRef] [PubMed]

6.   Wilkinson, K.; Ng, W.; Roberts, T.L.; Becker, T.M.; Lim, S.H.S.; Chua, W.; Lee, C.S. Tumour immune microenvironment biomarkers predicting cytotoxic chemotherapy efficacy in colorectal cancer. *J. Clin. Pathol.* **2021**, *74*, 625–634. [CrossRef] [PubMed]

7.   Dagenborg, V.J.; Marshall, S.E.; Grzyb, K.; Fretland, Å.A.; Lund-Iversen, M.; Mælandsmo, G.M.; Ree, A.H.; Edwin, B.; Yaqub, S.; Flatmark, K. Low Concordance Between T-Cell Densities in Matched Primary Tumors and Liver Metastases in Microsatellite Stable Colorectal Cancer. *Front. Oncol.* **2021**, *11*, 1629. [CrossRef] [PubMed]

8.   Behling, F.; Schittenhelm, J. Tissue microarrays–translational biomarker research in the fast lane. *Expert Rev. Mol. Diagn.* **2018**, *18*, 833–835. [CrossRef] [PubMed]

9.   Storz, M.; Moch, H. Tissue microarrays and biomarker validation in molecular diagnostics. In *Molecular Genetic Pathology*; Springer: New York, NY, USA, 2013; pp. 455–463.

10.   Russo, J.; Sheriff, F.; Pogash, T.J.; Nguyen, T.; Santucci-Pereira, J.; Russo, I.H. Methodological Approach to Tissue Microarray for Studying the Normal and Cancerous Human Breast. In *Techniques and Methodological Approaches in Breast Cancer Research*; Springer: New York, NY, USA, 2014; pp. 103–118.

11.   Thallinger, G.G.; Baumgartner, K.; Pirklbauer, M.; Uray, M.; Pauritsch, E.; Mehes, G.; Buck, C.R.; Zatloukal, K.; Trajanoski, Z. TAMEE: Data management and analysis for tissue microarrays. *BMC Bioinform.* **2007**, *8*, 1–12. [CrossRef] [PubMed]

12.   Frøssing, L.; Hartvig Lindkaer Jensen, T.; Østrup Nielsen, J.; Hvidtfeldt, M.; Silberbrandt, A.; Parker, D.; Porsbjerg, C.; Backer, V. Automated cell differential count in sputum is feasible and comparable to manual cell count in identifying eosinophilia. *J. Asthma* **2022**, *59*, 552–560. [CrossRef] [PubMed]

13.   Dvanesh, V.D.; Lakshmi, P.S.; Reddy, K.; Vasavi, A.S. Blood cell count using digital image processing. In Proceedings of the 2018 International Conference on Current Trends towards Converging Technologies (ICCTCT), Tamil Nadu, India, 1–3 March 2018; pp. 1–7.

14.   Jagomast, T.; Idel, C.; Klapper, L.; Kuppler, P.; Proppe, L.; Beume, S.; Falougy, M.; Steller, D.; Hakim, S.; Offermann, A.; et al. Comparison of manual and automated digital image analysis systems for quantification of cellular protein expression. *Histol. Histopathol.* **2022**, *37*, 18434.

15.   Schiele, S.; Arndt, T.T.; Martin, B.; Miller, S.; Bauer, S.; Banner, B.M.; Brendel, E.M.; Schenkirsch, G.; Anthuber, M.; Huss, R.; et al. Deep learning prediction of metastasis in locally advanced colon cancer using binary histologic tumor images. *Cancers* **2021**, *13*, 2074. [CrossRef] [PubMed]

16.   Graham, S.; Jahanifar, M.; Vu, Q.D.; Hadjigeorghiou, G.; Leech, T.; Snead, D.; Raza, S.E.A.; Minhas, F.; Rajpoot, N. CoNIC: Colon Nuclei Identification and Counting Challenge 2022. *arXiv* **2021**, arXiv:2111.14485.

17.   Budginaitė, E.; Morkūnas, M.; Laurinavičius, A.; Treigys, P. Deep learning model for cell nuclei segmentation and lymphocyte identification in whole slide histology images. *Informatica* **2021**, *32*, 23–40. [CrossRef]

18.   Vuong, T.L.T.; Lee, D.; Kwak, J.T.; Kim, K. Multi-task deep learning for colon cancer grading. In Proceedings of the IEEE 2020 International Conference on Electronics, Information, and Communication (ICEIC), Barcelona, Spain, 19–22 January 2020; pp. 1–2.

19.   Bilal, M.; Raza, S.E.A.; Azam, A.; Graham, S.; Ilyas, M.; Cree, I.A.; Snead, D.; Minhas, F.; Rajpoot, N.M. Novel deep learning algorithm predicts the status of molecular pathways and key mutations in colorectal cancer from routine histology images. *medRxiv* **2021**. [CrossRef]

20.   Sarker, M.M.K.; Makhlouf, Y.; Craig, S.G.; Humphries, M.P.; Loughrey, M.; James, J.A.; Salto-Tellez, M.; O'Reilly, P.; Maxwell, P. A Means of Assessing Deep Learning-Based Detection of ICOS Protein Expression in Colon Cancer. *Cancers* **2021**, *13*, 3825. [CrossRef] [PubMed]

21.   Mehta, S.; Rastegari, M. Mobilevit: Light-weight, general-purpose, and mobile-friendly vision transformer. *arXiv* **2021**, arXiv:2110.02178.

22.   Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.C. Mobilenetv2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 4510–4520.

23.   Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 7132–7141.

24.   Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional networks for biomedical image segmentation. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 5–9 October 2015; pp. 234–241.

25.   Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 3431–3440.

26.   Oktay, O.; Schlemper, J.; Folgoc, L.L.; Lee, M.; Heinrich, M.; Misawa, K.; Mori, K.; McDonagh, S.; Hammerla, N.Y.; Kainz, B.; et al. Attention U-Net: Learning where to look for the pancreas. *arXiv* **2018**, arXiv:1804.03999.

27. Chen, L.C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-decoder with atrous separable convolution for semantic image segmentation. In Proceedings of the European Conference on Computer Vision, Munich, Germany, 17–24 May 2018; pp. 801–818.

28. Zhou, Z.; Siddiquee, M.M.R.; Tajbakhsh, N.; Liang, J. Unet++: A nested u-net architecture for medical image segmentation. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*; Springer: Berlin/Heidelberg, Germany, 2018; pp. 3–11.