Journal of Medical and Biological Engineering
https://doi.org/10.1007/s40846-020-00545-4

ORIGINAL ARTICLE

Automatic Detection and Counting of Lymphocytes from Immunohistochemistry Cancer Images Using Deep Learning

I. Keren Evangeline[1] · J. Glory Precious[2] · N. Pazhanivel[3] · S. P. Angeline Kirubha[1]

Received: 10 March 2020 / Accepted: 16 June 2020

Abstract

Purpose Cancer is one of the most life-threatening and devastating diseases in the world. The generally recognized standard for cancer staging is the TNM staging system. In addition, a new parameter called the Immunoscore has been developed recently to assess the survival rate of patients. The Immunoscore is based on counts of CD3+ and CD8+ lymphocytes in the tumour core and margin. Counting the number of lymphocytes is a tedious process for pathologists. This paper examines the use of deep learning techniques for automatic detection and counting of lymphocytes from immunohistochemistry images of breast, colon, and prostate cancers.

Methods We used an object detector called Faster R-CNN with four feature extractors: Resnet-50, VGG-16, Inception-V2, and Resnet-101 for automatic detection and counting of lymphocytes. A total of 11,136 lymphocytes were annotated after performing data augmentation on 1228 images. The test images are separated into three regions of interest (ROI): scattered lymphocytes, groups of lymphocytes, and artefacts. In each ROI, the performance of the object detector was checked by evaluation metrics.

Results On comparing the F1-score for all three ROIs, we found that Resnet-101 provides better performance than the other feature extractors when using Faster R-CNN. The mean error in lymphocyte count for all ROIs appears to be insignificant. The detection time for a single image is less for VGG-16 than for other feature extractors.

Conclusion This study presents a fine-tuned Faster R-CNN object detector for automatic detection and counting of lymphocytes in three different cancer tissues for scoring lymphocytes. Our results suggest that the Faster R-CNN method is efficient and yields good results. Thus, the proposed method can assist pathologists in providing a cancer prognosis.

Keywords Immunohistochemistry · Faster R-CNN · Deep learning · Cancer · Object detection

1 Introduction

Cancer is a deadly disease characterised by uncontrollable multiplication of cells, which may invade other parts of the body and destroy tissues. In India, the current population is 1.34 billion, with nearly 50% of the population below the age of 25. Nonetheless, India reportedly has the youngest cancer population in the world and 550,000 cancer deaths per year [1]. The TNM system is used for staging and classifying cancer. Treatment is given to improve the prognosis based on the stage and grade of the cancer. Recently, a parameter called the Immunoscore has been considered as a new approach in cancer classification. Because it accounts for the immune system, it may have significant prognostic value compared to TNM classification [2–5]. The Immunoscore is a diagnostic test performed to identify the response of the immune cells at the tumour area by measuring the density

✉ S. P. Angeline Kirubha
angeline.sp@ktr.srmuniv.ac.in4

I. Keren Evangeline
kerenevangeline@gmail.com

J. Glory Precious
glorypreciously88@gmail.com

N. Pazhanivel
drnpvel@yahoo.co.in3

[1] Department of Biomedical Engineering, SRM Institute of Science and Technology, Chennai, India

[2] Department of Electronics and Communication Engineering, SRM Institute of Science and Technology, Chennai, India

[3] Department of Veterinary Pathology, Madras Veterinary College, Vepery, Chennai, India

Published online: 04 July 2020

of T lymphocytes identified by CD3+ and CD8+ markers in both the core and the invasive margin of the tumour [6, 7]. The density of tumour-infiltrating lymphocytes varies by tumour subtype and grade [8]. The Immunoscore facilitates the administration of immunotherapy, which is effective in patients resistant to chemotherapy or radiation; its side effects are comparatively less severe [9–11]. Deep learning is a state-of-the-art method for classification and object detection in which features are detected automatically with the help of filters [12, 13]. Some of its applications include detecting lymph node metastases [14], counting ovarian follicles [15], detecting microscopic cells [16–20] and intervertebral discs [21], and classifying and detecting cancer [22–26].

Recently, deep learning has been used in the field of immunohistochemistry for object detection and classification [27]. Swiderska-Chadaj et al. investigated the effectiveness of four deep learning algorithms for the detection of lymphocytes in breast, colon, and prostate cancer tissues. U-net seemed to perform better than other approaches [28]. Rijthoven et al. presented a YOLO version 2 model for the detection of lymphocytes in whole slide immunohistochemistry cancer images of breast, colon, and prostate tissues. The modifications made in their work increased the speed up to 4.3 × and there was also a 3% increase in detection performance [29]. Garcia et al. created an approach to automatically detect and count lymphocytes from gastric immunohistochemistry [30]. Chen and Chefd'Hotel introduced an algorithm which automatically detects immune cells in immunohistochemistry images to assist in immune profile studies [31].

To assign an Immunoscore, a pathologist must evaluate the density of the lymphocytes in the core and margin of the tumour. This is a time-consuming and difficult process. To assist pathologists in counting, we aim to automatically detect and count lymphocytes using a deep learning algorithm called Faster R-CNN (FR-CNN). Faster R-CNN is one of the best object-detection networks. It uses region proposals to determine objects' locations, improving the overall accuracy of the detection [32]. In this study, we augmented 1228 images approximately 10 times to obtain 11,136 images; these were manually annotated with the help of a pathologist. We conducted data augmentation to increase the performance of the feature extractor during training. We fine-tuned the hyperparameters to prevent overfitting. We used 1500 images for testing, with 500 images in each ROI. Training was done using Tensorflow and Keras [33, 34]. The training images, along with their annotations of the lymphocytes, were given to four different feature extractors (VGG-16, Resnet-50, Inception V2, and Resnet-101) to create the feature map. The possible candidate regions from the feature map can be identified by the region proposal network (RPN) after applying non-maximum suppression (NMS). These regions are pooled and given to the detector network to predict the class of the detected objects and the location of the bounding boxes, which gives the object's position. This method yields reliable and accurate results.

2 Methodology

2.1 Data Set

The data set used in our study was obtained from the Lymphocyte Assessment Hackathon (LYSTO) [35]. The data set has image patches of size 299 × 299 × 3 obtained at a magnification of × 40 from breast, prostate, and colon cancer whole slide images stained with CD3 and CD8 markers. The data set also provides the total number of lymphocytes in each image patch. Figure 1 shows the distribution of the immunohistochemistry image data set after augmentation.

2.2 Faster R-CNN for Detecting and Counting Lymphocytes

The Faster Region-Based Convolutional Neural Network (FR-CNN) is one of the fastest and most accurate object detectors available. It can be divided into three parts: the feature extractor, the region-proposal network, and the detection network. The feature extractor consists primarily of convolutional nets used to extract features from the input image and form feature maps. From these maps, the region-proposal network gives the possible region proposals that may contain the object of interest. These regions of interest are then pooled to bring all the proposals to a similar size. Non-maximum suppression is used to remove anchors, which overlap with the same object to reduce error. The final region proposals are given to a fully connected layer. A classifier is used to predict if the region proposal has object or background and also to give the class of the object. A regressor is used to give the position of the predicted object by fine-tuning the coordinates of the bounding box. The feature extractor can be any network.

Faster R-CNN was initially trained using the ImageNet data set, which has millions of images. It is difficult to obtain huge numbers of medical images. Training a feature extractor from scratch with insufficient data leads to severe overfitting and the model cannot generalise. Moreover, the accuracy and performance of the model are very poor and the computation cost is high. Therefore, the process of transfer learning for the feature extractor is highly preferable in this scenario. In this process, pre-trained model weights are fine-tuned to achieve good results. Because pre-trained models are trained on data sets with diverse images, they can capture common features such as edges and curves in their initial

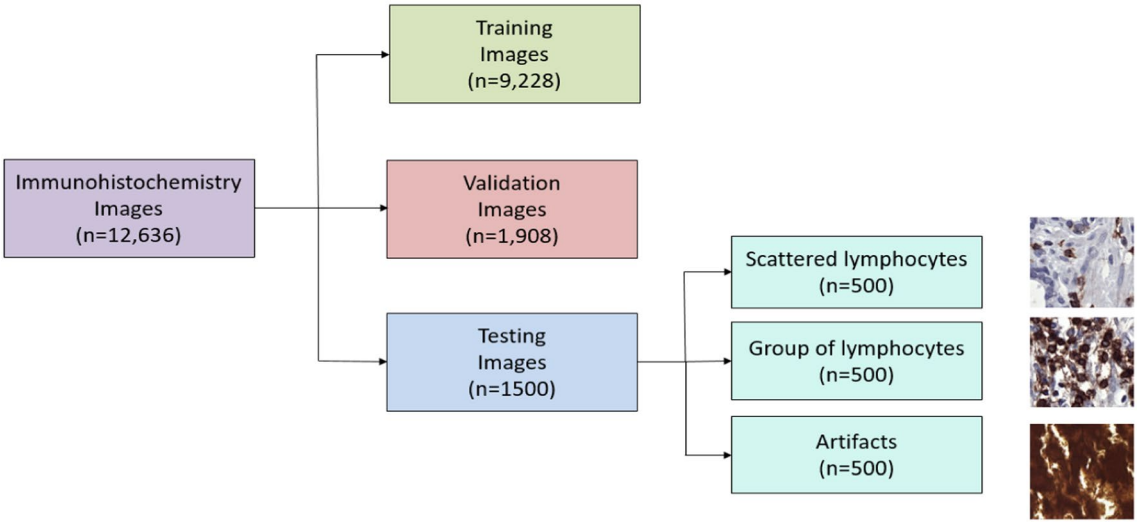

Fig. 1 Distribution of the immunohistochemistry image dataset after data augmentation

layers. These common features are useful even for medical images.

Faster R-CNN initially used ZF net and VGG-16 as its feature extractors to obtain feature maps. VGG-16 performs better than ZF net due to its deeper architecture and higher number of parameters. In this study, we fine-tuned the pre-trained weights with the lymphocyte data set so that the CNN could learn both common and data set-specific features by convolution and pooling on many levels. Here, the last layer that classifies the image is removed; hence, the CNN acts as a feature extractor. Shallu and Rajesh Mehra (2018) found that limited medical histopathological images perform well by the transfer-learning process of VGG-16 and Resnet-50 when pre-trained on an ImageNet data set, rather than using fully trained networks [36]. Hence, models pre-trained with an ImageNet data set learn medical image-specific features during the transfer-learning process, thereby yielding good performance and generalisation. Since the pre-trained network is already well-trained, the weights should not be modified immediately or excessively. Rather, by using a small learning rate, it should be fine-tuned to achieve good results.

We used VGG-16, Resnet-50, Inception-V2, and Resnet-101 as feature extractors, comparing them to determine which feature extractor performed best for this data set. VGG-16 has 16 layers and yields high accuracy due to its depth, as it learns complex features at a lesser cost. Resnet-50 is a 50-layer residual network and a deep network. It overcomes the problem of degrading accuracy, seen in all deep neural networks, by residual learning; thus, it is easier to train Resnet-50 and it gives good accuracy. Resnet-101 is similar to Resnet-50 but is much deeper, having 101 layers. Inception-V2 aims to reduce computational time by making the network wider instead of deeper. It has filters of various sizes. In Inception-V2, the 5×5 convolutional filter is factorised into two 3×3 filters to reduce computation, which boosts performance.

In addition, our study implements a counter in the detection network that automatically counts and displays the number of lymphocytes present in the image after detection. This can aid pathologists while also calculating the Immunoscore—the pathologist does not have to manually count the lymphocytes detected by Faster R-CNN. We aim to determine which feature extractor performs well in both detection and counting. Figure 2 shows the block diagram for the Faster R-CNN object detector.

The block diagram shown in Fig. 3 outlines the steps for training and validation, using four different networks in the Faster R-CNN object detector. The training data set has an equal number of images from each region of interest. The lymphocytes present in each ROI are manually annotated with the help of a pathologist and given as input for training and validation. Training is done on four different networks in Faster R-CNN; the weights are obtained for each network after training. Several hyperparameters are fine-tuned to improve the performance of the object detector. By trial and error, the hyperparameter values that yielded the best results were identified. The same parameters were set for all feature-extractor models. Table 1 shows the list of adjusted hyperparameters.

The block diagram in Fig. 4 illustrates the steps for the testing phase, which uses four networks in Faster R-CNN. In the testing phase, the weights that are obtained after training are applied; the detector network is used to detect the location of the bounding boxes and the class. Faster R-CNN with Resnet-50, VGG-16, Inception-V2, and Resnet-101

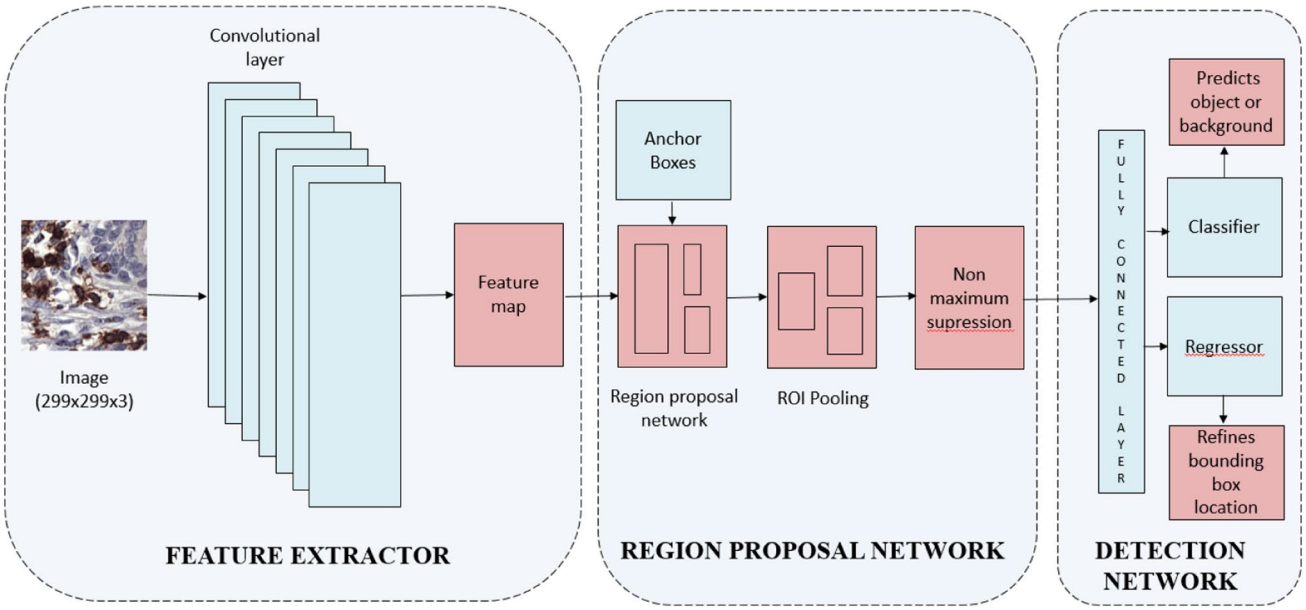

Fig. 2 Block diagram of Faster RCNN

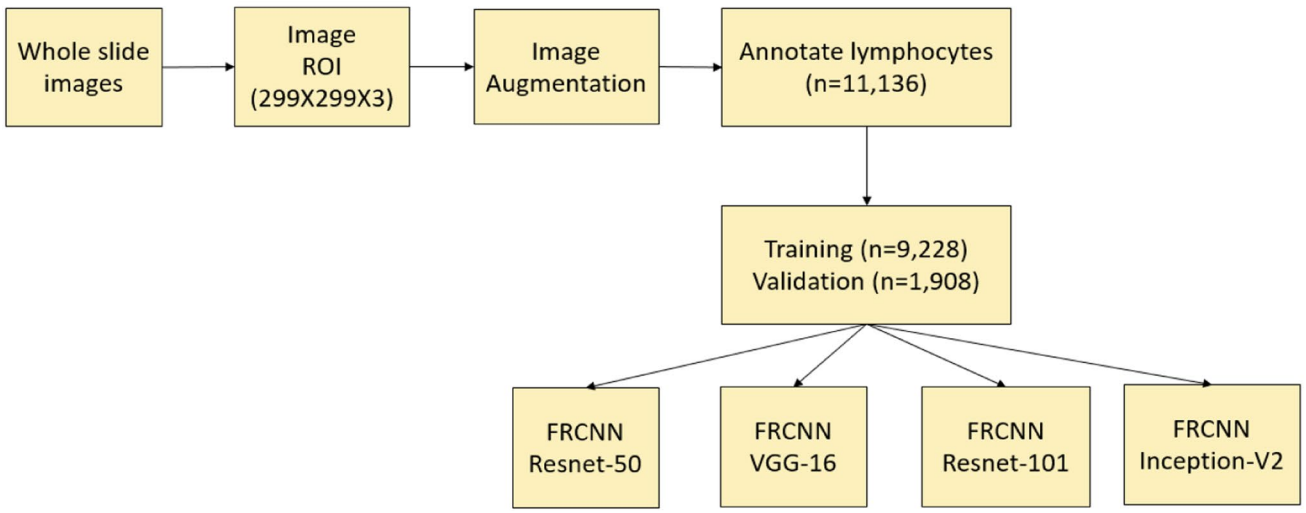

Fig. 3 Block diagram of the training phase

Table 1 List of hyperparameters fine-tuned

	Faster RCNN Resnet-50, VGG-16, inception V2, resnet-101
Optimizer	Adam
Learning rate	0.00001
No. of epochs	100
Epoch length	100
Loss function	Mean absolute error
Confidence score threshold	0.8
Overlap threshold for NMS	0.7

are used to detect and count lymphocytes from three sets of test images, including images with scattered lymphocytes, groups of lymphocytes, and artefacts.

2.3 Performance Analysis of FR-CNN

The performance of the object detector can be assessed using several evaluation metrics. When a lymphocyte is correctly detected, this is a true positive. False positives occur when the object detector incorrectly detects a lymphocyte in the absence of any lymphocytes. When a lymphocyte is not detected, this is a false negative. These parameters are

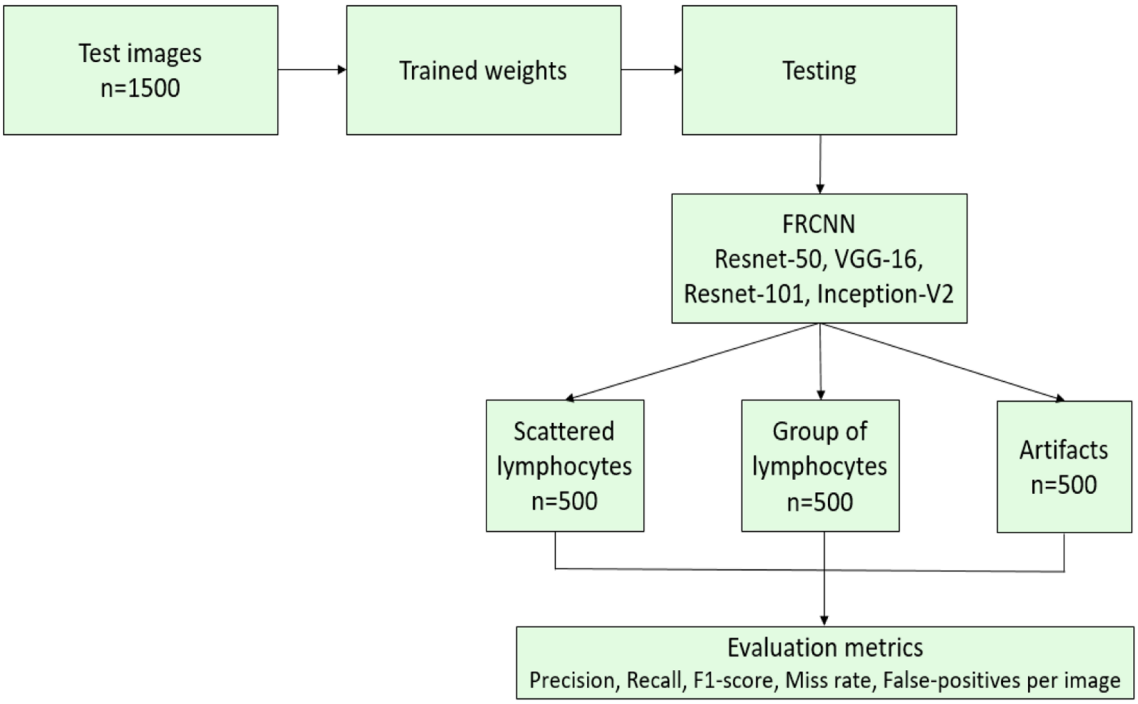

Fig. 4 Block diagram of the testing phase

used to calculate the precision, recall, F1 score, miss rate, and false positives per image. Precision is the ratio of true positives to all detections. A value of one indicates that there are no false positive detections. Recall is the number of true positives divided by all possible detections. A value of one means that all visible lymphocytes have been detected and there are no false negatives. The F1 score is the harmonic mean of recall and precision. The miss rate can be defined as the ratio of false negatives to all objects. The miss rate is zero when all possible lymphocytes have been detected. False positives per image can be stated as the number of false positives divided by the number of all lymphocytes detected. A value of zero indicates that the detected objects are correctly classified.

3 Results and Discussion

We used faster R-CNN with four feature extractors (VGG-16, Resnet-50, Inception-V2, and Resnet-101). We implemented the program using Tensorflow for GPU version 1.10 and Keras version 2.2.0. We used Nvidia Titan XP GPU for training. The input image annotations were given as input to the convolution net or feature extractors, yielding the feature map; the region-proposal network was used to get the region proposals from the feature map after applying non-maximum suppression. We used the detection network to classify the object and refine the coordinates of the bounding box. The hyperparameters were fine-tuned to improve detection performance. The lymphocytes were counted automatically based on the number of bounding boxes.

3.1 Training and Testing of the Data Set

Approximately 70% of the data were given for training, with 20% for testing and 10% for validation; 9228 images with their annotations were given for training and 1908 images were given for validation. For testing, we used 1500 images subdivided into three sets of 500 images. Each set depicts different regions of the slide, such as areas of clustered lymphocytes, areas having fewer lymphocytes, and areas with artefacts such as a background ink stain, in which there is non-specific staining or the tissues are stained in the absence of lymphocytes.

3.2 Lymphocyte Detection and Count

Faster R-CNN yields rapid detections. It is the fastest region-based convolutional neural network. Figure 5 shows that VGG-16 is much faster than the other models. This is because the depth of the other models is greater than that of VGG-16; it is also based on the number of parameters that each network produces. In images with artefacts, Resnet-101 takes the longest amount of time (about 1 s) to detect the lymphocytes in a single image. For images containing a group of lymphocytes, Resnet-50, Resnet-101, and

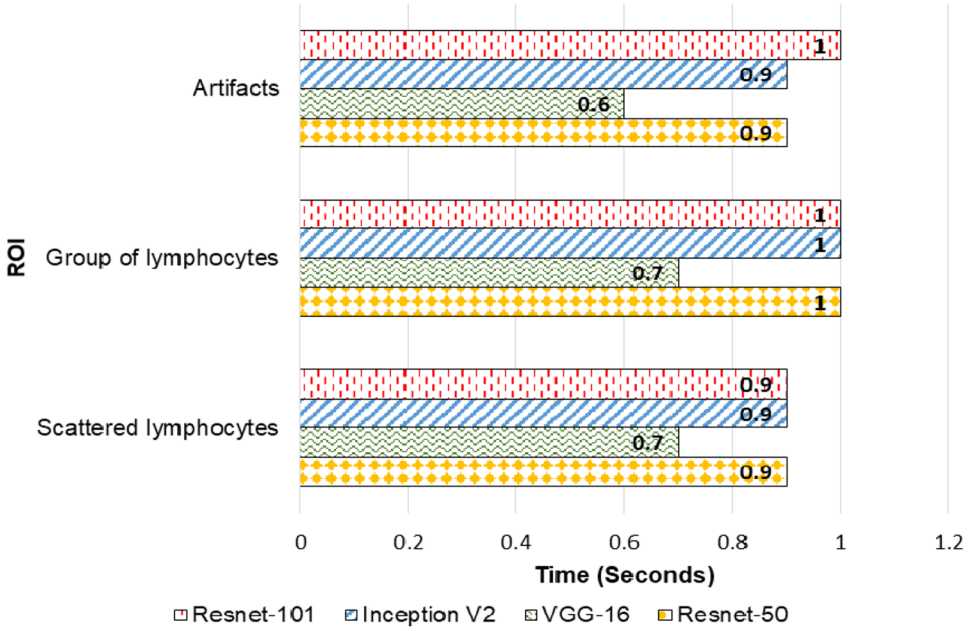

Fig. 5 Lymphocyte detection time per image

Inception-V2 take approximately the same length of time to detect lymphocytes from a single image. This may be because these images contain more objects and, therefore, determining the region proposals for these objects takes time. Similarly, for images having scattered or fewer lymphocytes, the time taken for detection in a single image is 0.9 s for all models except for VGG-16. However, the detection time is approximately equal for all ROIs.

Figure 6 shows the image patches having scattered lymphocytes in colon, breast, and prostate cancer, as well as their detection and count in Faster R-CNN, using various feature extractors such as Resnet-50, VGG-16, Inception-V2, and Resnet-101, respectively.

Figure 7 shows the image patches having groups of lymphocytes in colon, breast, and prostate cancer. Faster R-CNN can accurately detect and count the lymphocytes even though the lymphocytes are in clusters.

Figure 8 shows image patches that have artefacts. Notably, faster R-CNN correctly predicts the absence of lymphocytes in artefacts and gives the count as zero.

Figure 9 shows the number of test images with the count correctly detected. Out of 500 test images with scattered lymphocytes, VGG-16 gives the correct count for 496 images, Resnet-101 gives the correct count for 495 images, Inception-V2 gives the correct count for 494 images and Resnet-50 gives the correct count for 485 images. VGG-16 gives the most accurate count in the images with scattered lymphocytes. For groups of lymphocytes, Resnet-101 gives the correct count for 489 images, Resnet-50 for 472 images, Inception-V2 for 465 images and VGG-16 for 463 images. Resnet-101 performed well on this ROI. For images with artefacts, both Resnet-50 and Resnet-101 give the correct count for 488 images, VGG-16 gives the correct count for 487 images, and Inception-V2 gives the correct count for 483 images. Both Resnet-50 and Resnet-101 perform well on this ROI.

The average count error for images with scattered lymphocytes is 1.9 ± 1.65 for Resnet-50, 1.5 ± 0.89 for VGG-16, 1.6 ± 0.86 for Inception-V2, and 2.3 ± 0.77 for Resnet-101. For images with artefacts, the average count error is 2.2 ± 1.79 for VGG-16, 2.5 ± 1.92 for Resnet-50, 1.8 ± 1.02 for Inception-V2, and 1.2 ± 0.85 for Resnet-101. For the images with groups of lymphocytes, the average error is 3.2 ± 1.8 for Resnet-50, 3.8 ± 2.48 for VGG-16, 2.6 ± 2.02 for Inception-V2, and 2.8 ± 1.66 for Resnet-101. Figure 10 shows the mean error in lymphocyte count at various regions of interest for different feature extractors.

3.3 Accuracy and Loss

Table 2 shows the accuracy and loss of the object detector. The mean number of bounding boxes from RPN overlapping the ground truth boxes is greater in Resnet-101 than in other models. The accuracy of the classifier for bounding boxes from RPN is 98% for Resnet-101, the highest for any model. The loss of both the RPN classifier and RPN regression is minimal. Similarly, the value of the loss in the detector classifier and regression is also minimal for all feature extractors.

3.4 Performance Evaluation Metrics

Table 3 shows that the precision of VGG-16 in Faster R-CNN for scattered lymphocytes images is 92% which is

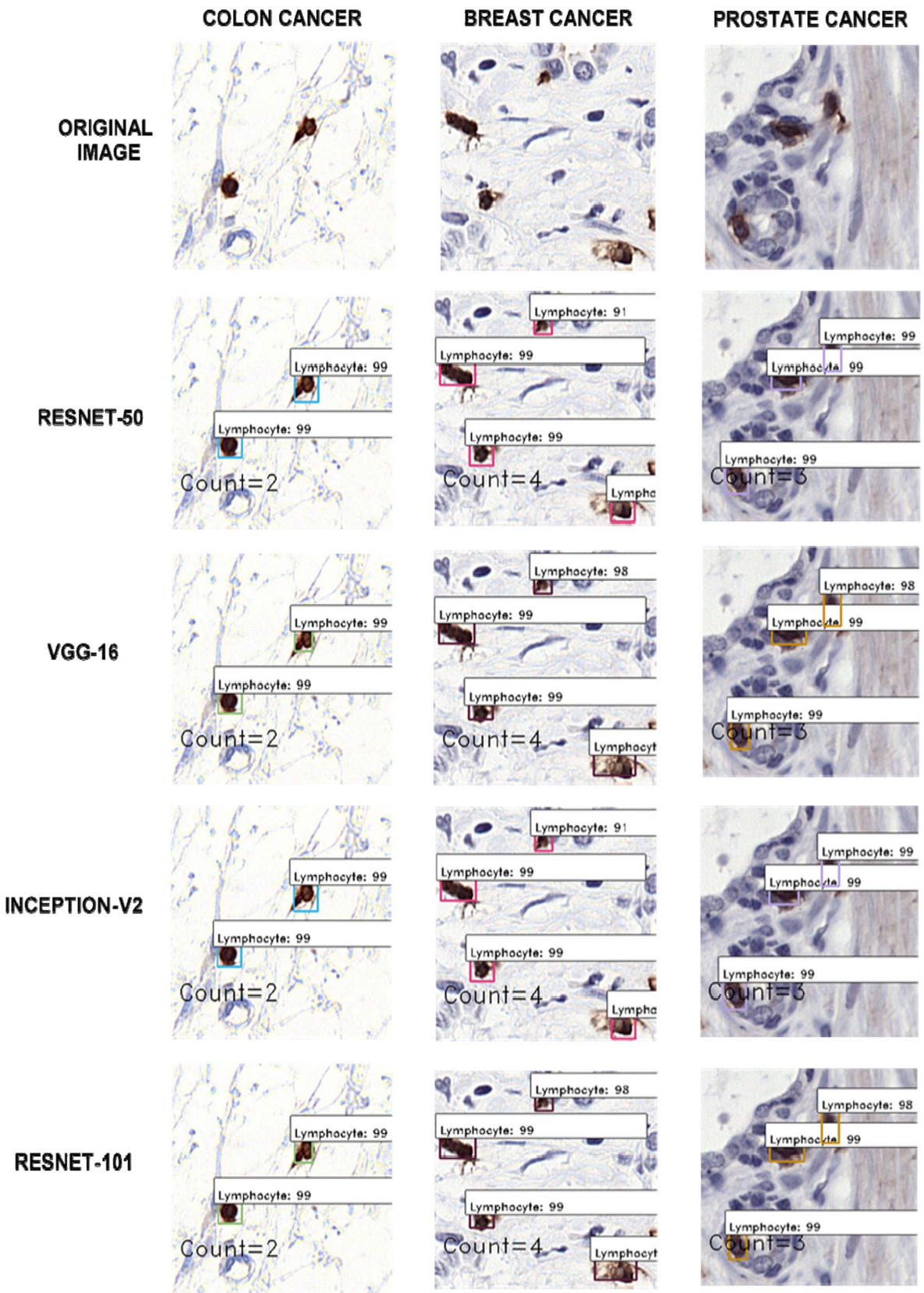

Fig. 6 Lymphocyte detection and count in image patches with scattered lymphocytes

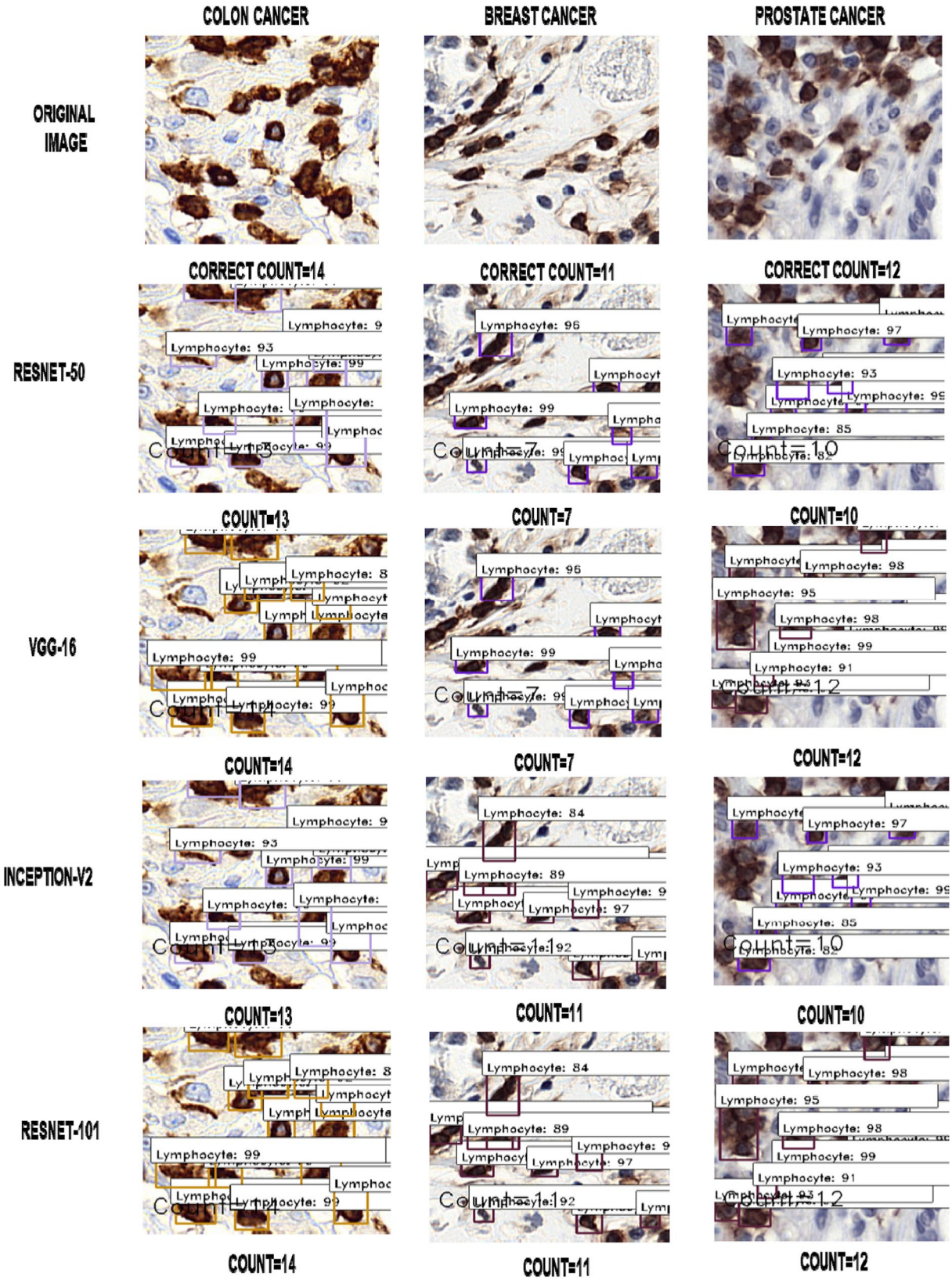

Fig. 7 Lymphocyte detection and count in image patches with group of lymphocytes

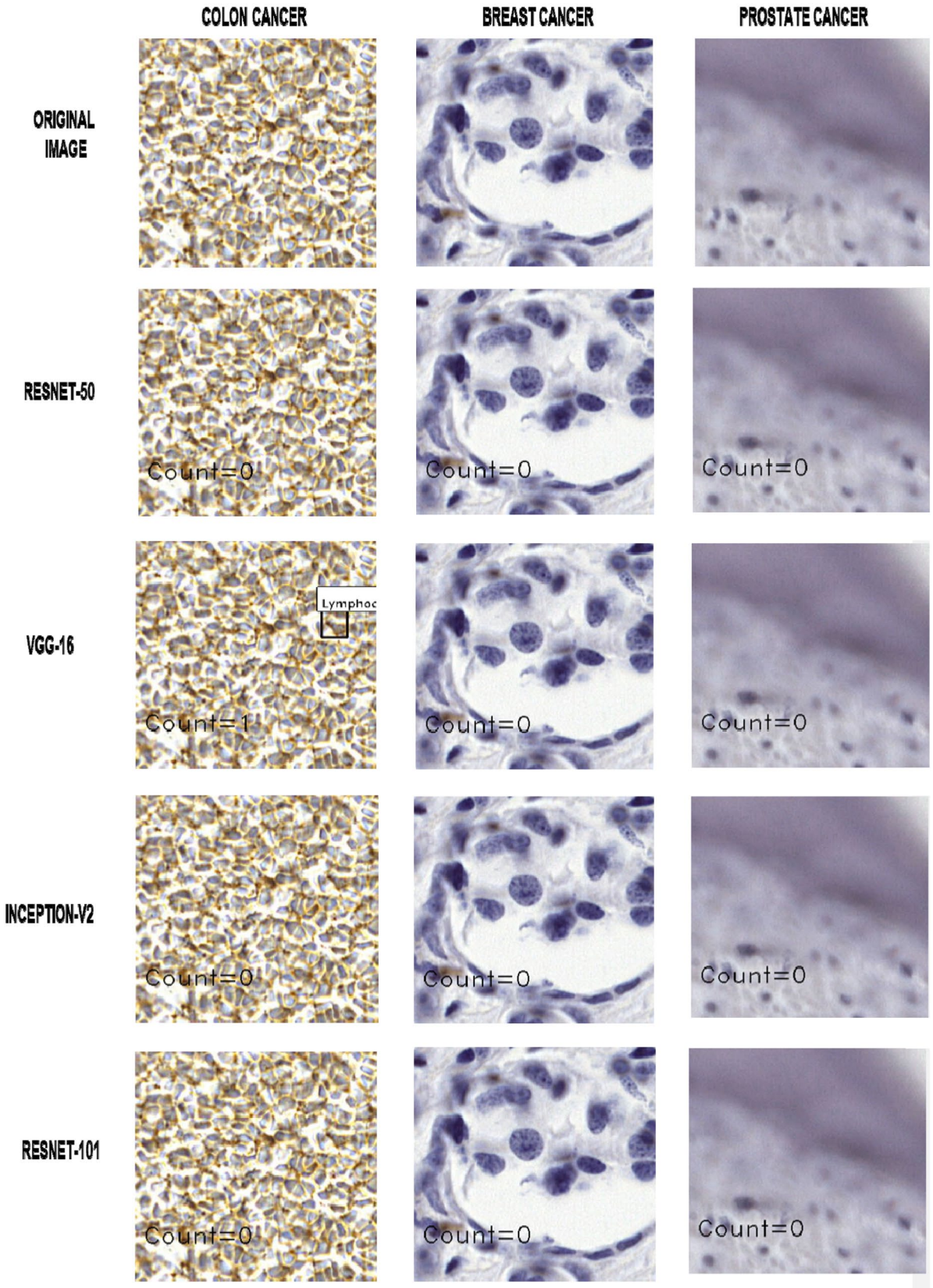

Fig. 8 Lymphocyte detection and count in image patches with artifacts

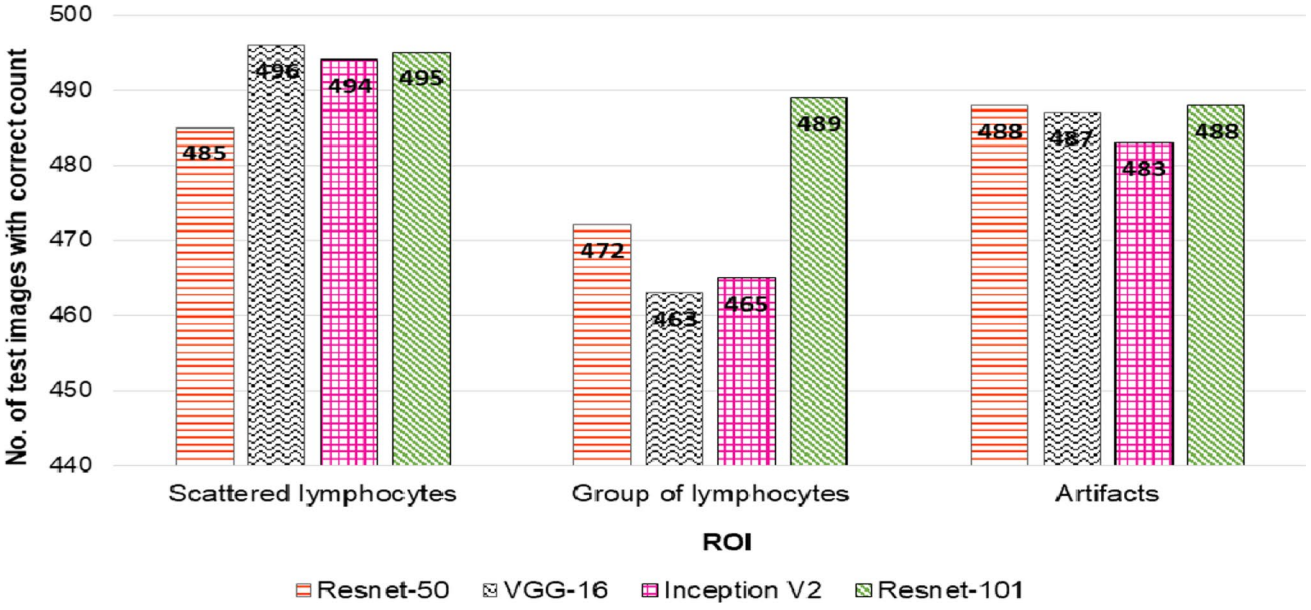

Fig. 9 No. of test images with correct count

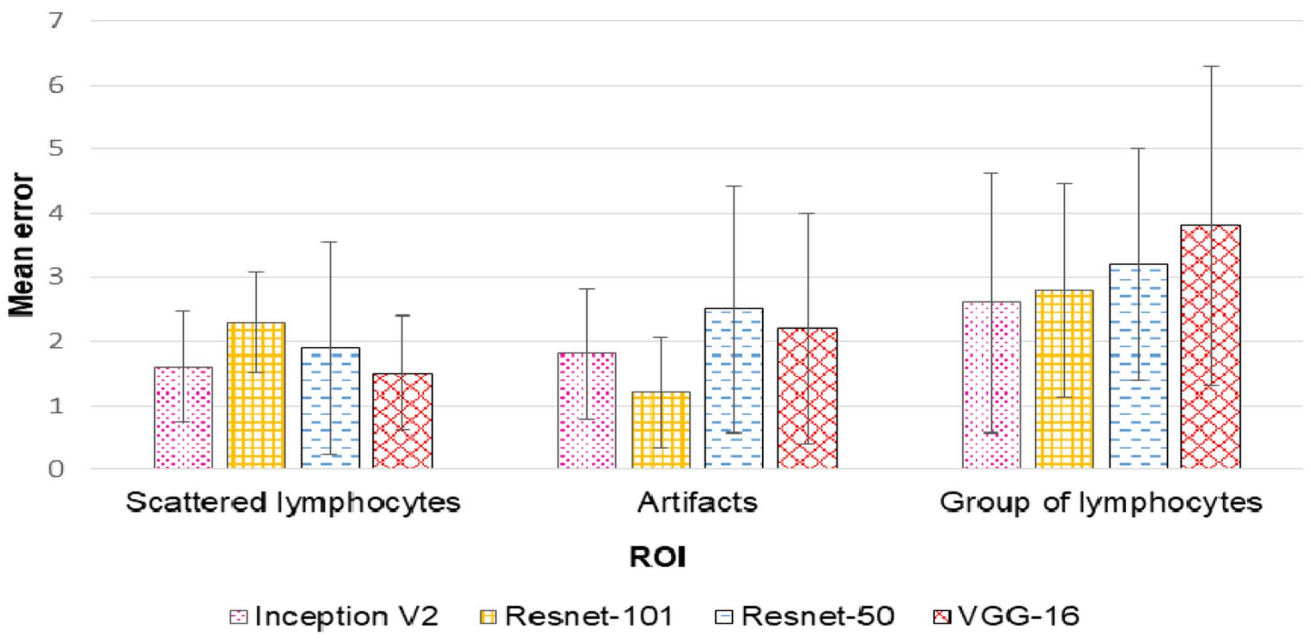

Fig. 10 Mean error in lymphocyte count

higher when compared to other networks. Recall is higher for both Inception-V2 and Resnet-101 in scattered lymphocytes.

The F1 score for Resnet-101 is 0.96—the highest among the feature extractors—in images with scattered lymphocytes. In the case of image patches with groups of lymphocytes, the performance of Resnet-101 is higher than the other feature extractors because its precision, recall, miss-rate, and false positive per image values are higher than the other models; Resnet-101 has an F1-score of 0.86. Similarly, for image patches with artefacts, Resnet-101 has a higher precision value (0.79) than the other models, although VGG-16 has the highest recall value. Resnet-101 performed better on this ROI than the other models, with an F1-score of 0.81. Table 3 shows the performance evaluation metrics of Faster R-CNN Resnet-50, VGG-16, Inception-V2, and Resnet-101.

Table 2 Accuracy and loss

	Faster RCNN Resnet-50	Faster RCNN VGG-16	Faster RCNN Inception-V2	Faster RCNN Resnet-101
Mean number of bounding boxes from RPN overlapping ground truth boxes	6.87	6.58	6.37	7.03
Classifier accuracy for bounding boxes from RPN	0.9746	0.9518	0.9427	0.9852
Loss RPN classifier	0.1454	0.1422	0.1401	0.1389
Loss RPN regression	0.0194	0.0337	0.0635	0.0125
Loss detector classifier	0.0767	0.1079	0.1143	0.0587
Loss detector regression	0.0265	0.0494	0.0879	0.0211

Table 3 Performance evaluation metrics

ROI	FRCNN Feature extractor	Precision	Recall	F1 score	Miss Rate	False-positives per image
Scattered lymphocytes	Resnet-50	0.8793	0.9684	0.9217	0.0316	0.1207
	VGG-16	**0.9293**	0.9598	0.9442	0.0402	**0.0707**
	Inception-V2	0.8369	**0.9898**	0.9069	**0.0102**	0.1631
	Resnet-101	0.9131	**0.9898**	**0.9615**	**0.0102**	0.0869
Group of lymphocytes	Resnet-50	0.9585	0.7569	0.8458	0.2431	0.0415
	VGG-16	0.9158	0.7253	0.8094	0.2747	0.0842
	Inception-V2	0.9231	0.7403	0.8216	0.2597	0.0769
	Resnet-101	**0.9651**	**0.7829**	**0.8645**	**0.2171**	**0.0349**
Artifacts	Resnet-50	0.7411	0.8614	0.7967	0.1386	0.2589
	VGG-16	0.7065	**0.8812**	0.7842	**0.1188**	0.2935
	Inception-V2	0.7410	0.83	0.7829	0.17	0.259
	Resnet-101	**0.7982**	0.8425	**0.8197**	0.1575	**0.2018**

Bold indicate highest values

3.5 Discussion

We utilised a deep learning-based approach to automatically detect and count lymphocytes in immunohistochemistry images. Our proposed method achieved satisfactory results after data augmentation. Data augmentation reduces overfitting and achieves good performance during training and testing on pre-trained models with transfer learning. We used four feature extractors (VGG-16, Inception-V2, Resnet-50, and Resnet-101) and compared their performance. Each yielded agreeable results as their F1-scores were high. Resnet-101 yielded the highest F1-score (0.96) in images with scattered lymphocytes. For images with groups of lymphocytes, Resnet-101 also performed well, with an F1-score of 0.86; its F1 score for images with artefacts was 0.81. The F1-scores for groups of lymphocytes can be further improved by annotating additional clustered images and giving them for training. Resnet-101 performed well in areas with scattered lymphocytes when the lymphocytes were present without overlapping and had a clear background. On the other hand, the F1-score for areas with artefacts was lower because the object detector was unable to effectively distinguish between background ink stains and lymphocytes. This could be improved by training the detector with additional images with artefacts, followed by pre-processing. The F1 score values achieved in our study are slightly higher than the values (0.71–0.82) obtained in previous papers that used copious annotation data for training (Swiderska-Chadaj et al. 2019 and Rijthoven et al. 2018). The classifier accuracy for bounding boxes from RPN in this paper is 98% for Resnet-101, only slightly higher than 96%, found in a study by Garcia et al. (2017). We have introduced automatic counting of lymphocytes based on the number of bounding boxes predicted, in addition to automatic detection of lymphocytes. We also calculated the error in count for all three regions; the deviation in error appears acceptable. The overall performance of Resnet-101 was good in both detection and counting. However, VGG-16 detects lymphocytes in images more quickly than the other feature extractors. The performance of Resnet-50 is much better than Inception-V2 and VGG-16, which have similar performance. This suggests that faster R-CNN as an object detector is efficient and

performs well after data augmentation. Thus, our proposed method can reduce the time and effort taken by pathologists to give an Immunoscore.

4 Conclusion

This paper has presented a comparison of various fine-tuned models for a Faster R-CNN object detector to automatically detect and count tumour-infiltrating lymphocytes in three different cancer tissues for scoring lymphocytes. We achieved a classifier accuracy for bounding boxes of 98% with Resnet-101. Our findings suggest that faster R-CNN with the Resnet-101 method is quite accurate and produces good results. Faster R-CNN detects objects in a single image much more quickly than the R-CNN and Fast R-CNN techniques. By reducing the number of region proposals in Faster R-CNN, the detection speed can be increased effectively. Testing the performance of the object detector with evaluation metrics like precision, recall, the F1-score, miss rate, and false-positives per image shows that the overall performance of Resnet-101 as a feature extractor is good for all three ROIs, although the differences in performance among the other feature extractors are negligible. VGG-16 provides faster detection times in single images in Faster R-CNN. Since the object detector is trained with three cancer tissues from breast, prostate, and colon cancer, it can detect lymphocytes from any image taken from these organs. Our proposed model includes a counter in the detection network that displays the number of lymphocytes detected. This reduces the workload on pathologists, who would normally have to count the detected lymphocytes manually. Thus, the method we propose could aid pathologists in providing a prognosis for cancer patients.

4.1 Future Scope

Real-time clinical immunohistochemistry data from various organs could be used for training and testing and determining the Immunoscore. The lymphocyte count can be obtained from both the core and invasive margin of the tumour; the count can be classified as low- or high-density and the Immunoscore can automatically be given accordingly. Other feature extractors and object-detection models can be used to check performance.

Funding This study is not funded.

Compliance with Ethical Standards

Conflict of interest The authors declare that they have no conflict of interest to this work.

Ethical Approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

1. Munshi, A., Ganesh, T., & Mohanti, B. K. (2019). Radiotherapy in India: History, current scenario and proposed solutions. *Indian Journal of Cancer, 56*, 359. https://doi.org/10.4103/ijc.IJC_82_19.
2. Pagès, F., Mlecnik, B., Marliot, F., Bindea, G., Ou, F. S., Bifulco, C., et al. (2018). International validation of the consensus Immunoscore for the classification of colon cancer: A prognostic and accuracy study. *The Lancet, 391*, 2128–2139. https://doi.org/10.1016/S0140-6736(18)30789-X.
3. Mlecnik, B., Tosolini, M., Kirilovsky, A., Berger, A., Bindea, G., Meatchi, T., et al. (2011). Histopathologic-based prognostic factors of colorectal cancers are associated with the state of the local immune reaction. *Journal of Clinical Oncology, 29*, 610–618. https://doi.org/10.1200/JCO.2010.30.5425.
4. Galon, J., Costes, A., Sanchez-Cabo, F., Kirilovsky, A., Mlecnik, B., Lagorce-Pagès, C., et al. (2006). Type, density, and location of immune cells within human colorectal tumors predict clinical outcome. *Science, 313*, 1960–1964. https://doi.org/10.1126/science.1129139.
5. Mlecnik, B., Bindea, G., Angell, H. K., Maby, P., Angelova, M., Tougeron, D., et al. (2016). Integrative analyses of colorectal cancer show immunoscore is a stronger predictor of patient survival than microsatellite instability. *Immunity, 44*, 698–711. https://doi.org/10.1016/j.immuni.2016.02.025.
6. Trabelsi, M., Farah, F., Zouari, B., Jaafoura, M. H., & Kharrat, M. (2019). An immunoscore system based on CD3+ and CD8+ infiltrating lymphocytes densities to predict the outcome of patients with colorectal adenocarcinoma. *OncoTargets and Therapy, 12*, 8663. https://doi.org/10.2147/OTT.S211048.
7. Galon, J., Pagès, F., Marincola, F. M., Angell, H. K., Thurin, M., Lugli, A., et al. (2012). Cancer classification using the Immunoscore: A worldwide task force. *Journal of Translational Medicine, 10*, 205. https://doi.org/10.1186/1479-5876-10-205.
8. Lee, K. H., Kim, E. Y., Yun, J. S., Park, Y. L., Do, S. I., Chae, S. W., et al. (2018). The prognostic and predictive value of tumor-infiltrating lymphocytes and hematologic parameters in patients with breast cancer. *BMC Cancer, 18*, 938. https://doi.org/10.1186/s12885-018-4832-5.
9. Zgura, A., Galesa, L., Bratila, E., & Anghel, R. (2018). Relationship between tumor infiltrating lymphocytes and progression in breast cancer. *Maedica, 13*, 317. https://doi.org/10.26574/maedica.2018.13.4.317.
10. Rosenberg, S. A., & Restifo, N. P. (2015). Adoptive cell transfer as personalized immunotherapy for human cancer. *Science, 348*, 62–68. https://doi.org/10.1126/science.aaa4967.
11. Varn, F. S., Wang, Y., Mullins, D. W., Fiering, S., & Cheng, C. (2017). Systematic pan-cancer analysis reveals immune cell interactions in the tumor microenvironment. *Cancer research, 77*, 1271–1282. https://doi.org/10.1158/0008-5472.CAN-16-2490.
12. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*, 436–444. https://doi.org/10.1038/nature14539.
13. Gautam, R., & Sharma, M. (2020). Prevalence and diagnosis of neurological disorders using different deep learning techniques: A meta-analysis. *Journal of Medical Systems, 44*, 49. https://doi.org/10.1007/s10916-019-1519-7.
14. Bejnordi, B. E., Veta, M., Johannes van Diest, P., van Ginneken, B., Karssemeijer, N., Litjens, G., et al. (2017). Diagnostic assessment of deep learning algorithms for detection of lymph node

metastases in women with breast cancer. *JAMA, 318*, 2199–2210. https://doi.org/10.1001/jama.2017.14585.
15. İnik, O., Ceyhan, A., Balcıoğlu, E., & Ülker, E. (2019). A new method for automatic counting of ovarian follicles on whole slide histological images based on convolutional neural network. *Computers in Biology and Medicine, 112*, 103350. https://doi.org/10.1016/j.compbiomed.2019.103350.
16. Yang, S., Fang, B., Tang, W., Wu, X., Qian, J., & Yang, W. (2017). Faster R-CNN based microscopic cell detection. *International Conference on Security, Pattern Analysis, and Cybernetics (SPAC)*. https://doi.org/10.1109/SPAC.2017.8304302.
17. Xing, F., & Yang, L. (2016). Robust nucleus/cell detection and segmentation in digital pathology and microscopy images: A comprehensive review. *IEEE Reviews in Biomedical Engineering, 9*, 234–263. https://doi.org/10.1109/RBME.2016.2515127.
18. Falk, T., Mai, D., Bensch, R., Çiçek, Ö., Abdulkadir, A., Marrakchi, Y., et al. (2019). U-Net: Deep learning for cell counting, detection, and morphometry. *Nature methods, 16*, 67. https://doi.org/10.1038/s41592-018-0261-2.
19. Xie, Y., Xing, F., Kong, X., Su, H., & Yang, L. (2015). Beyond classification: Structured regression for robust cell detection using convolutional neural network. *Medical Image Computing and Computer-Assisted Intervention, 9351*, 358–365. https://doi.org/10.1007/978-3-319-24574-4_43.
20. Bayramoglu, N., & Heikkilä, J. (2016). Transfer learning for cell nuclei classification in histopathology images. In G. Hua & H. Jégou (Eds.), *Computer vision – ECCV 2016 workshops. ECCV 2016. Lecture notes in computer science* (pp. 532–539). Cham: Springer.
21. Ruhan, S., William, O., Raymond, W., Mark, S., Donald, C., Kenneth, B., Alexander, G., Robert, R., Adam, H., John, C., & Vipin, C. (2020). Intervertebral disc detection in X-ray images using faster RCNN: A deep learning approach. In: *39th annual international conference of the IEEE engineering in medicine and biology society (EMBC)* (pp. 564–567). https://doi.org/10.1109/EMBC.2017.8036887.
22. Agnes, S. A., Anitha, J., Pandian, S. I. A., & Peter, J. D. (2020). Classification of mammogram images using multiscale all convolutional neural network (MA-CNN). *Journal of Medical Systems, 44*, 30. https://doi.org/10.1007/s10916-019-1494-z.
23. Hamid, M. A. A., & Khan, N. A. (2020). Investigation and classification of MRI brain tumors using feature extraction technique. *Journal of Medical and Biological Engineering*. https://doi.org/10.1007/s40846-020-00510-1.
24. Ronoud, S., & Asadi, S. (2019). An evolutionary deep belief network extreme learning-based for breast cancer diagnosis. *Soft Computing, 23*, 13139–13159. https://doi.org/10.1007/s00500-019-03856-0.
25. Dai, X., Huang, L., Qian, Y., Xia, S., Chong, W., Liu, J., et al. (2020). Deep learning for automated cerebral aneurysm detection on computed tomography images. *International Journal of Computer Assisted Radiology and Surgery*. https://doi.org/10.1007/s11548-020-02121-2.
26. Sirinukunwattana, K., Ahmed Raza, S., Tsang, Y. W., Snead, D. R. J., Cree, I. A., & Rajpoot, N. M. (2016). Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images. *IEEE Transactions on Medical Imaging, 35*, 1196–1206. https://doi.org/10.1109/TMI.2016.2525803.
27. Klauschen, F., Müller, K. R., Binder, A., Bockmayr, M., Hägele, M., Seegerer, P., et al. (2018). Scoring of tumor-infiltrating lymphocytes: From visual estimation to machine learning. *Seminars in Cancer Biology, 52*, 151–157. https://doi.org/10.1016/j.semcancer.2018.07.001.
28. Swiderska-Chadaj, Z., Pinckaers, H., Rijthoven, M., Balkenhol, M., Melnikova, M., Geessink, O., et al. (2019). Learning to detect lymphocytes in immunohistochemistry with deep learning. *Medical Image Analysis, 58*, 101547. https://doi.org/10.1016/j.media.2019.101547.
29. Rijthoven, M., Swiderska-Chadaj, Z., Seeliger, K., Laak, J., & Ciompi, F. (2018). You only look on lymphocytes once. In: *1st conference on medical imaging with deep learning (MIDL)*.
30. Garcia, E., Hermoza, R., Castanon, C. B., Cano, L., Castillo, M., & Castanneda, C. (2017). Automatic lymphocyte detection on gastric cancer IHC images using deep learning. *IEEE 30th international symposium on computer-based medical systems (CBMS)*, (pp. 200–204). https://doi.org/10.1109/CBMS.2017.94.
31. Chen, T., & Chefd'Hotel, C. (2014). Deep learning based automatic immune cell detection for immunohistochemistry images. In G. Wu, D. Zhang, & L. Zhou (Eds.), *Machine learning in medical imaging. MLMI 2014. Lecture notes in computer science* (pp. 17–24). Cham: Springer.
32. Ren, S., He, K., Girshick, R., Sun, J. (2015). Faster R-CNN: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems,* (pp. 91–99).
33. Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., & Ghemawat, S. (2015). Tensorflow: Large-scale machine learning on heterogeneous distributed systems. http://arxiv.org/abs/1603.04467.
34. Chollet, F. (2015). Keras. *GitHub repository*. https://github.com/fchollet/keras
35. Ciompi, F., Jiao, Y., & Laak, J. (2019). *Lymphocyte assessment hackathon (LYSTO)*. Zenodo.
36. Mehra, R. (2018). Breast cancer histology images classification: Training from scratch or transfer learning? *ICT Express, 4*(4), 247–254. https://doi.org/10.1016/j.icte.2018.10.007.