2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Multi-Scale Aggregated-Dilation Network for *ex-vivo* Lung Cancer Detection with Fluorescence Lifetime Imaging Endomicroscopy

Qiang Wang[1] and James R. Hopgood[2] and Marta Vallejo[3]

Abstract— **Multi-scale architectures at a granular level are characterised by separating input features into groups and applying multi-scale feature extractions to the split input features, and thus the correlations among the input features as global information are no longer retained. Moreover, they usually require more input features due to the separation, and therefore, more complexity is introduced. To retain the global information while utilising the advantages of feature-level hierarchical multi-scale architectures, we propose a multi-scale aggregated-dilation architecture (MSAD) to perform hierarchical fusion of features at a layer level, with the integration of dilated convolutions to overcome these issues. To evaluate the model, we integrate it into ResNet, and apply it to a unique dataset, containing over 60,000 fluorescence lifetime endomicroscopic images (FLIM) collected on *ex-vivo* lung normal/cancerous tissues from 14 patients, by a custom fibre-based FLIM system. To evaluate the performance of our proposal, we use accuracy, precision, recall, and AUC. We first compare our MSAD model with eight networks achieving a superiority over 6%. To illustrate the advantages and disadvantages of multi-scale architectures at layer and feature-level, we thoroughly compare our MSAD model with the state-of-the-art feature-level multi-scale network, namely Res2Net, in terms of parameters, scales, and effective convolutions.**

I. Introduction

Fluorescence lifetime is a unique characteristic of fluorophore, which is independent of its intensity, but sensitive to various internal and external factors, such as fluorophore structure and its biological environment. Due to its high sensitivity and diversity, tissue lifetime contrast has been utilised to differentiate human diseases [1]. Various FLIM systems have been applied to detect cancer and other conditions [2]. Conventionally, statistical methods dominated the discrimination of cancerous tissue, with the assistance of auxiliary information, e.g. histopathological images. Surprisingly, little attention has been paid to FLIM-based cancer differentiation using ML technologies. In Chen *et al.* [3], artificial features extracted from lifetime reconstruction were used for the automatic detection of skin lesions by a support vector machine (SVM). Jo *et al.* [4] used quadratic discriminant analysis to classify malignant and benign oral cancer lesions with six FLIM-based features. Authors also made some effort in applying ML methods. In [5], four ML algorithms, namely K-nearest neighbour, SVM, neural network, and random forest, were applied to FLIM images directly, yielded AUC scores below 0.78. Later, several classic convolutional neural networks (CNNs) were applied to evaluate the performance of deep learning on a FLIM dataset [6]. This research demonstrated, without surprise, that deep learning is superior to conventional ML methods. It also proved that combining intensity and lifetime together as the input, improves the prediction.

In contemporary CNNs, multi-scale architectures have been broadly applied [7], [8], [9], due to its capability of extracting and fusing features rich in spatial and contextual information. Multi-scale styles can be roughly categorised into three groups: architecture-level, layer-level, and feature-level aggregation, which are different in complexity, flexibility, and reusability. Architecture-level aggregation usually deals with multiple inputs [10], and thus, it is relatively rigid and complex. Layer-level aggregation parallels multiple convolutions in a block to retrieve features at different scales, like in Inception [7], which usually requires more computational time and resources due to extra convolutions appended. Feature-level aggregation usually concerns the subsets of input features, e.g. group convolution and its variants [11], [12], but correlations among features are ignored.

Here, we propose a novel CNN architecture, namely MSAD, that integrates the advantages of layer-level feature aggregation, feature-level multi-scale feature extraction, and dilated convolutions. We parallel a number of 3×3 convolutions with distinctive dilation rates at a layer-level, and introduce an extra aggregation before the convolutions to fuse the features extracted from previous dilated convolutions with the global features at layer-level. In addition, an identity shortcut is also presented to improve the flow of information and gradient. By incorporating the proposed block into ResNet [13], we apply the model, combining different types of filters with various numbers of dilated convolutions, to over 60,000 FLIM images collected *ex-vivo* on 14 pairs of normal/cancerous tissues from 14 patients by a fibre-based custom FLIM system. Accuracy, precision, recall, and AUC are utilised as metrics. To fully evaluate the performance of the model, we compare it with eight state-of-the-art CNNs: ResNet, ResNeXt [14], DenseNet [15], Inception-v3 [16], Xception [12], SENet [17], Res2Net [8], and Res2NeXt [8].

To emphasise the novelty of the proposed model, we list the primary differences between ours and feature-level multi-scale models, particularly Res2Net [8] and FPENet [9]:

- In feature-level multi-scale models, such as ResNeXt and Res2Net, input features are split into several groups and multi-scale feature extraction is performed on the individual groups. In contrast, our model processes

[1]Qiang Wang is with the Centre for Inflammation Research, University of Edinburgh, Edinburgh, UK q.wang@ed.ac.uk

[2]James R. Hopgood is with the School of Engineering, University of Edinburgh, Edinburgh, UK james.hopgood@ed.ac.uk

[3]Marta Vallejo is with the School of Engineering and Physical Science, Heriot-Watt University, Edinburgh, UK m.vallejo@hw.ac.uk

978-1-7281-1178-0/21/$31.00 ©2021 IEEE

input features altogether to retain the correlations;
- Feature level models, particularly Res2Net, eliminate feature redundancy of ResNet and create new features by combining separation and hierarchical aggregation, whereas our model achieves the elimination via much narrower backbones and the creation by the aggregation;
- Due to separation, feature-level multi-scale architectures require more input features than our model to maintain the width and scale. Consequently, they are usually more complex than ours. For example, our model has up to 20% fewer parameters than Res2Net, given the same settings, but with comparable results; and
- Res2Net employed standard 3×3 convolutions for multi-scale features extraction, and FPENet used depth-wise dilated convolutions, whereas ours utilised 3×3 dilated convolutions.

The rest of the paper is organised as follows. Section II introduces the technical details of our method. The results are presented in Section III, followed by the conclusions and future work in Section IV.

II. Methodology

Fig. 1 depicts the overall procedure of our approach. Raw data was collected *ex-vivo* using a custom fibre-based FLIM system on pairs of normal/cancerous tissues from individual patients (Step 1). Afterwards, the collected images are pre-processed for quality enhancement. Later, intensity and life-time images are stacked as the input to the MSAD network (Step 2). Eventually, the stacked images are classified by the proposed model (Step 3).

A. Image Collection

A custom fibre-based FLIM system was deployed to collect data (Fig. 1, step 1). Settings included two exposure times (6 and 20 μs) and two spectral bands (498-570 and 594-764 nm). Lifetime values were calculated using rapid lifetime determination (RLD) [18]. For each *ex-vivo* experiment, a pair of normal/cancerous tissues from a patient were scanned, and multiple measurements were collected at different physical points by direct contact of the fibre with the tissue.

Fig. 1, step 2 shows the pre-processing steps. To derive plausible lifetime from intensity using the RLD method, optimal signal-noise ratio (SNR) of the intensity is needed. In this study, we used a threshold value $\sqrt{\hat{I}}$ related to SNR, where $\hat{I}$ is the mean of the measured intensity. A fluorescence intensity greater than $\sqrt{\hat{I}}$ is required to perform a lifetime calculation of acceptable accuracy. Afterwards, the thresholded intensity images are normalised with dark background and lightfield images. The normalisation is adapted from [19]. Later, a histogram-based contrast enhancement [20] is applied to the normalised images to further improve their quality. Then, the contrast-enhanced intensity image is used as a mask on the thresholded lifetime image to yield the pre-processed intensity and corresponding lifetime images. Eventually, an intensity and its lifetime image are stacked into two channels of an RGB image, keeping the remaining channel blank, which are the input to the MSAD model.

B. Multi-Scale Aggregated-Dilation Architecture

As shown in Fig. 1 (right side of Step 3), the proposed MSAD architecture parallels several 3×3 convolutions with different dilation rates in order for multi-scale contextual features to be retrieved simultaneously. Similar to [8] and [9], aggregation is introduced to fuse the information from the previous branch, so that, both layer-level global features and branch-wise local features can be considered together. Following the ideas of [13], an identity shortcut is also employed, together with the aggregation operator to improve the flow of information and gradient throughout the block.

Let $d_i(r_i)$ denote the i^{th} branch in the MSAD block with dilation rate r_i, and I and O are the input and output of the block. Therefore, O can be defined as:

$$O = B([I, d_0(r_0), ..., d_n(r_n)]) \tag{1}$$

where $[I, d_0(r_0), ..., d_n(r_n)]$ is the concatenation of all branch-wise outputs, and B is a composite operation containing a 3×3 convolution, batch normalisation [21], followed by a rectified linear unit [22]. Suppose the receptive field of $d_{i-1}(r_{i-1})$ is f_{i-1}, the output from $d_i(r_i)$, therefore, reflects the aggregated receptive field of $(r_i - 1) * 2 + 3$ from I and $((r_i - 1) * 2 + 3) * f_{i-1}$ from $d_{i-1}(r_{i-1})$. Taking the model in Fig. 1 as an example, the concatenated features include four parts from I and the parallel dilated convolutions, and thus are rich in scale and contextual information.

In this study, we use ResNet50 as the backbone network, by only replacing the original bottleneck block with the MSAD block. Following [13], [14], and [8], we use w as the width in the MSAD block, and s as the scale for the number of parallel dilated convolutions. For example, MSAD-ResNet50-w40-s2 represents the MSAD based on ResNet50, with width 40 and scale 2. Considering that a shortcut connection is included in the module, the dilation rate of each branch in a MSAD module is in the subset of the first *s-1* elements of (1, 2, 3, 5, 7, 9, 11, 13, 15). For example, given s equal to 4, there have three dilated convolutions in the module, and the dilation rate for each convolution is 1, 3, and 5, respectively. All MSAD variations and the classic models are implemented using *PyTorch*. In addition, we also evaluate eight CNNs for comparison purposes: ResNet50, ResNeXt50, DenseNet121, Inception-v3, Xception, SENet50, Res2Net50, and Res2NeXt50.

C. Training and Testing

13 patients' images were utilised for training and the remaining one patient dataset for testing. 10% of training data was split out for validation. All models were trained with stochastic gradient descent for 200 epochs of batch size 64. The learning rate was set to 0.1, and divided by 10 at epoch 100 and 150. In addition, we also employed weight decay 1e-4. For data augmentation, we utilised a strategy reported in [13], [15], and [8], except that vertical flipping was also applied. It is worth noting that all models, including the classic CNNs, were trained from scratch for fair comparison.

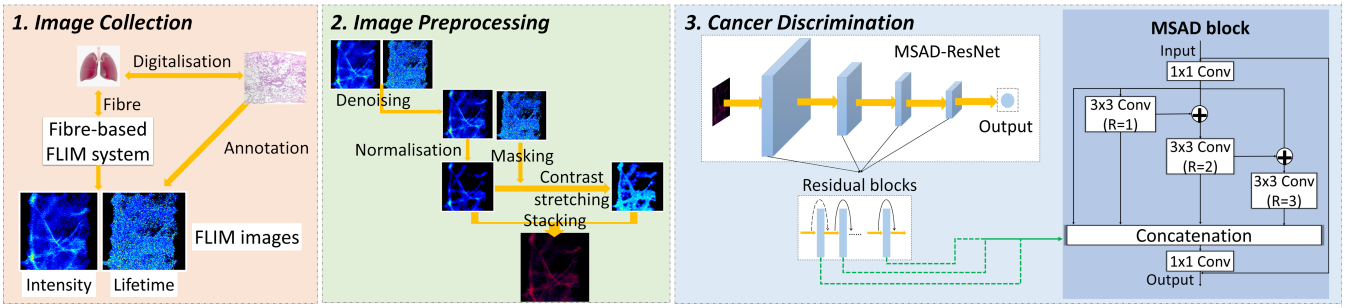

Fig. 1. Schematic diagram of the methodology.

TABLE I

PERFORMANCE COMPARISON. OVERALL BEST RESULTS ARE IN **BOLD**.

	Accuracy	Precision	Recall	AUC	Params
ResNet50	83.14	95.40	80.77	85.12	23.5
ResNeXt50	84.66	95.58	80.75	87.92	23.0
DenseNet121	80.43	95.22	76.96	83.33	7.0
Inception	83.61	96.99	78.04	88.26	21.8
Xception	83.24	96.67	78.29	87.37	20.8
SENet50	84.07	94.32	83.10	84.88	26.0
Res2Net50	84.35	93.89	80.89	88.26	23.7
Res2NeXt50	84.18	95.36	80.32	87.40	22.6
MSAD-ResNet50-w40-s2	84.81	**98.83**	83.06	88.76	18.3
MSAD-ResNet50-w24-s4	85.91	97.97	82.34	88.89	18.8
MSAD-ResNet50-w24-s6	**86.88**	98.01	**83.61**	**89.61**	25.8

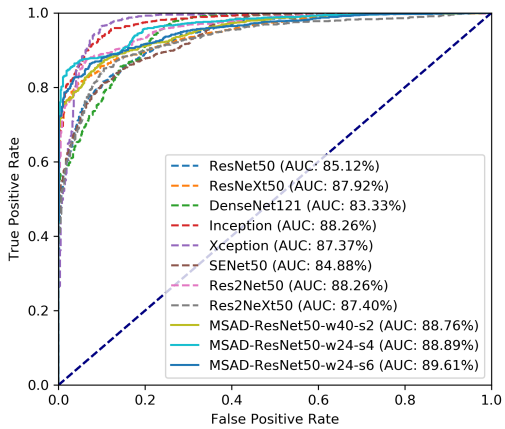

Fig. 2. ROC curves generated by the networks listed in Table I.

III. RESULTS

A. Overall performance

The resultant scores, including accuracy, precision, recall, and AUC, along with the complexity of the models, are listed in Table I, and ROC curves are plotted in Fig. 2. In general, all classic CNNs have a comparable performance in all the metrics. Overall, our MSAD model is superior to all the classic CNNs evaluated. For accuracy, all three MSAD models outperform the classic CNNs, with a gap of up to 6.45%. In particular, MSAD-ResNet50-w40-s2 surpasses the backbone ResNet50 for 1.65%, but only with 18.3M parameters, over 22% parameters less than the backbone. Similar results can also be found on precision, where all three models are better than the classic CNNs. In this case, MSAD-ResNet50-w40-s2 achieves the best precision, with up to 5% superiority. It is worth mentioning that all networks perform very well on precision, where the majority of them reach over 95%. This means that there is only a small number of normal-tissue images that are incorrectly classified as cancerous ones. As far as recall is concerned, our model still obtains remarkable outcomes. MSAD-ResNet50-w24-s6 is the best one on recall, and the rest two models are superior to all the CNNs, except for SENet50. As a result, MSAD is better than the evaluated CNNs with less number of cancerous images wrongly predicted to normal ones. When it comes to AUC, our proposed model has similar performance on accuracy and precision, where all three variations outperform the classic ones, and the discrepancy is up to 6.28%.

B. MSAD-ResNet vs Res2Net

Since MSAD is mainly inspired by Res2Net, we include a comparison of both models. We adapted the original Res2Net with dilated convolutions, so that the comparison is fair. We only replace the dilation rates in Res2Net with the identical ones used in our model. To fully evaluate the advantages and disadvantages of the hierarchical multi-scale architectures at layer and feature levels, we conducted experiments on parameter, scale, and convolution efficiency.

Parameter efficiency. To evaluate the parameter efficiency of our model and Res2Net, we conduct the experiments on five pairs of models, where each pair is identical in settings. The results are depicted in Fig. 3, where the first impression is that, given the same settings, Res2Net requires more parameters than MSAD-ResNet. The gap becomes larger with wider and deeper backbone ResNet, which can be up to 24%. For accuracy (first plot in Fig. 3), the MSAD model is more parameter efficient than Res2Net, although the best accuracy is achieved by Res2Net. As far as precision is concerned (second plot in Fig. 3), Res2Net performs better than the MSAD, where the networks are relatively simple with less than 5M parameters. With the increase of complexity, MSAD-ResNet becomes superior to Res2Net, and it yields the best precision overall, meaning it is better at providing correct prediction of cancer tissue. When it comes to recall (third plot in Fig. 3), MSAD-ResNet is remarkably better than Res2Net, particularly when they are relatively simple, and the gap can be over 8%. This indicates that the

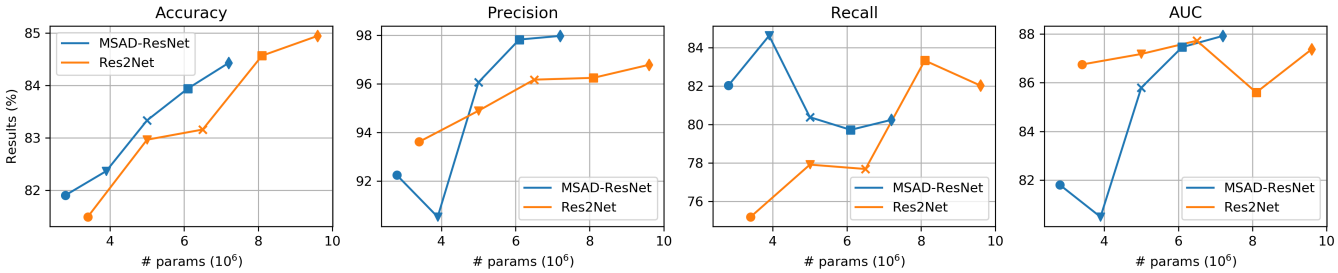

Fig. 3. Parameter efficiency of MSAD-ResNet and Res2Net.

MSAD architecture makes less error on predicting unhealthy to healthy tissue. The situation changes on AUC (fourth plot in Fig. 3), where Res2Net outperforms MSAD-ResNet, again in less complex configurations, although MSAD-ResNet is able to achieve the overall best AUC.

Scale efficiency. To compare the scale efficiency of our model and Res2Net, we use ResNet38-s32 as the backbone with identical settings. Five different scales, 2, 3, 4, 5, and 6, are evaluated. Fig. 4 depicts the impact of scale on the metrics, where ResNet38-w32 (red diamond in Fig. 4) is utilised as the baseline. As far as accuracy is concerned (first plot in Fig. 4), MSAD-ResNet is superior to Res2Net on scale 2 and 4, but inferior to Res2Net on scale 3, 5, and 6. For precision (second plot in Fig. 4), the situation changes. Our proposal surpasses Res2Net at scale 5 and 6, but Res2Net surpasses our proposal at scale 2, 3, and 4. This implies that Res2Net is more efficient on precision at a small scale, whereas MSAD-ResNet is better at larger scale. When it comes to recall (third plot in Fig. 4), the performance of Res2Net deteriorates dramatically, and with scales 2, 3, and 4, Res2Net is even worse than the baseline. With scales smaller than 5, MSAD-ResNet is significantly better than Res2Net, with a gap that can be over 6%. For scales 5 and 6, Res2Net outperforms MSAD-ResNet. The results on AUC (fourth plot in Fig. 4), are similar to those on precision. Res2Net achieves better outcomes with smaller scales, whereas MSAD-ResNet achieves better outcome with larger scales. In summary, on accuracy and recall, our MSAD model obtains higher scores with small scales. In contrast, Res2Net with small scales reaches higher scores on precision and AUC.

Convolution efficiency. Due to the hierarchical multi-scale architecture, both MSAD-ResNet and Res2Net introduce more convolution operations than the backbone, and this extra number becomes dramatic at a large scale. Therefore, it is important to evaluate the influence of the number of effective convolutions on the metrics. The results are illustrated in Fig. 5, where ResNet-w32 with different widths is used as the baseline. Overall, both networks outperform the backbone ResNet. On accuracy (first plot in Fig. 4), MSAD-ResNet outperforms Res2Net, except for the first variation. Note that Res2Net is even worse than the baseline when the effective number of convolutions is over 100. When it comes to precision (second plot in Fig. 4), MSAD-ResNet is comparable to Res2Net, although Res2Net is inferior to the baseline ResNet with over 100 effective convolutions. The results on recall are similar to those on accuracy, where MSAD-ResNet is over Res2Net, except for the case with over 100 effective convolutions, which is even worse than the baseline. When it comes to AUC (fourth plot in Fig. 4), MSAD-ResNet still surpasses Res2Net, except for the case with 42 convolutions. It is worth noting that Res2Net struggles outperforming the baseline in most cases on AUC. In conclusion, the MSAD model is superior to Res2Net in the number of effective convolutional operations for all metrics.

IV. Conclusions

In this study, we proposed a multi-scale architecture called MSAD. With ResNet as the backbone, we applied the proposed model to *ex-vivo* cancer discrimination using FLIM endomicroscopic images. The empirical results demonstrated the superiority of the proposed network over eight state-of-the-art CNNs for lung cancer classification with FLIM images. Since our model is inspired by feature-level multi-scale architectures, particularly Res2Net, we thoroughly compared our MSAD model with Res2Net, adapted with dilated convolutions in terms of parameter, scale, and convolution efficiency. Through the results, we can conclude that the MSAD model is more parameter efficient than Res2Net on accuracy, precision, and recall, but less on AUC. Moreover, MSAD-ResNet performs better with small scales on accuracy and recall, whereas Res2Net performs better with small scales on precision and AUC. In addition, the proposed model is overall superior to Res2Net with respect to effective convolutions, although there are few cases where our proposal is inferior to Res2Net.

It is worth noting that the proposed MSAD architecture is not designed specifically for FLIM-based lung cancer classification. Instead, it is expected to be also applicable to general image classification problems. As a result, future work will be conducted on migrating the model to general visual recognition tasks. In addition, the concepts within MSAD could also be applied to more complex computer vision problems, such as semantic segmentation. For this reason, we will also extend our research on this direction.

Acknowledgement

This work is supported by the Engineering and Physical Sciences Research Council (EPSRC, UK) Interdisciplinary Research Collaboration (grant number EP/K03197X/1 and

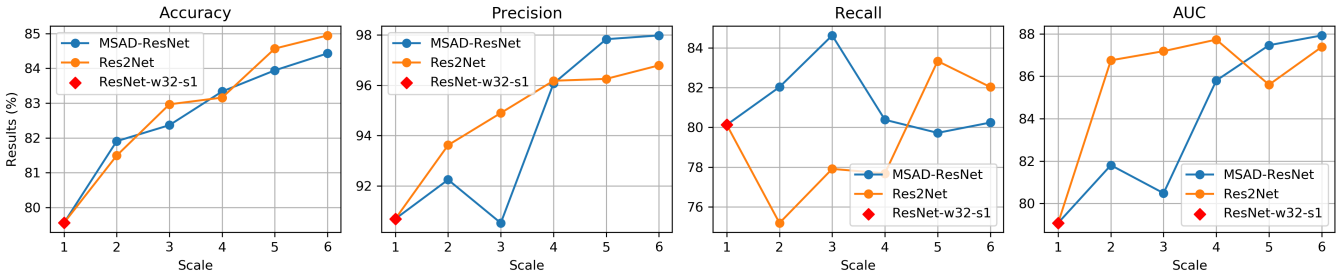

Fig. 4. Scale efficiency of MSAD-ResNet and Res2Net with different scales on the metrics, where ResNet38 is used as the backbone.

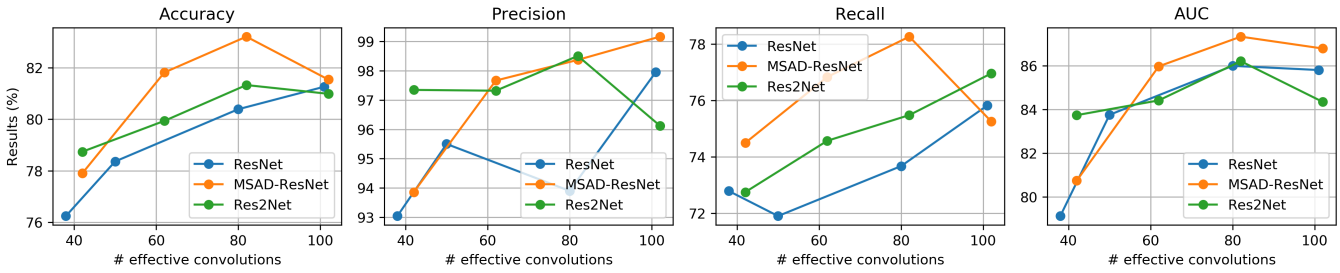

Fig. 5. Convolution efficiency of MSAD-ResNet and Res2Net, with ResNet of different depths as the baseline.

EP/R005257/1). We appreciate Dr Catharine Ann Dhaliwal for annotating the histological images. This work used the Cirrus UK National Tier-2 HPC Service at EPCC funded by the University of Edinburgh and EPSRC (EP/P020267/1). This project also made use of time on Tier 2 HPC facility JADE, funded by EPSRC (EP/P020275/1).

REFERENCES

[1] L. Marcu, "Fluorescence lifetime techniques in medical applications," *Annals of biomedical engineering*, vol. 40, no. 2, pp. 304–331, 2012.

[2] K. Suhling, L. M. Hirvonen, J. A. Levitt, P. H. Chung, C. Tregidgo, A. Le Marois, D. A. Rusakov, K. Zheng, S. Ameer-Beg, S. Poland, S. Coelho, R. Henderson, and N. Krstajic, "Fluorescence lifetime imaging (FLIM): Basic concepts and some recent developments," *Medical Photonics*, vol. 27, pp. 3–40, 2015.

[3] B. Chen, Y. Lu, W. Pan, J. Xiong, Z. Yang, W. Yan, L. Liu, and J. Qu, "Support Vector Machine Classification of Nonmelanoma Skin Lesions Based on Fluorescence Lifetime Imaging Microscopy," *Analytical Chemistry*, vol. 91, no. 20, pp. 10 640–10 647, 2019.

[4] J. A. Jo, S. Cheng, R. Cuenca-Martinez, E. Duran-Sierra, B. Malik, B. Ahmed, K. Maitland, Y. L. Cheng, J. Wright, and T. Reese, "Endogenous fluorescence lifetime imaging (flim) endoscopy for early detection of oral cancer and dysplasia*," in *40th International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2018, pp. 3009–3012.

[5] Q. Wang, M. Vallejo, and J. Hopgood, "Fluorescence Lifetime Endomicroscopic Image-based ex-vivo Human Lung Cancer Differentiation Using Machine Learning," *TechRxiv Preprint*, Jan. 2020.

[6] Q. Wang, J. R. Hopgood, N. Finlayson, G. O. Williams, S. Fernandes, E. Williams, A. R. Akram, K. Dhaliwal, and M. Vallejo, "Deep Learning in ex-vivo Lung Cancer Discrimination using Fluorescence Lifetime Endomicroscopic Images," in *42nd International Conference of IEEE Engineering in Medicine and Biology Society (EMBC)*, 2020.

[7] C. Szegedy, Wei Liu, Yangqing Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015, pp. 1–9.

[8] S. Gao, M. Cheng, K. Zhao, X. Zhang, M. Yang, and P. H. S. Torr, "Res2net: A new multi-scale backbone architecture," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–1, 2019.

[9] M. Liu and H. Yin, "Feature pyramid encoding network for real-time semantic segmentation," in *British Machine Vision Conference*, 7 2019.

[10] N. Alemi Koohbanani, M. Jahanifar, A. Gooya, and N. Rajpoot, "Nuclear instance segmentation using a proposal-free spatially aware deep learning framework," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*, D. Shen, T. Liu, T. M. Peters, L. H. Staib, C. Essert, S. Zhou, P.-T. Yap, and A. Khan, Eds. Cham: Springer International Publishing, 2019, pp. 622–630.

[11] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[12] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *IEEE conference on computer vision and pattern recognition*, 2017, pp. 1251–1258.

[13] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016, pp. 770–778.

[14] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in *IEEE conference on computer vision and pattern recognition*, 2017, pp. 1492–1500.

[15] G. Huang, Z. Liu, L. v. d. Maaten, and K. Q. Weinberger, "Densely Connected Convolutional Networks," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 2261–2269.

[16] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *IEEE conference on computer vision and pattern recognition*, 2016, pp. 2818–2826.

[17] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *IEEE conference on computer vision and pattern recognition*, 2018, pp. 7132–7141.

[18] R. M. Ballew and J. Demas, "An error analysis of the rapid lifetime determination method for the evaluation of single exponential decays," *Analytical Chemistry*, vol. 61, no. 1, pp. 30–33, 1989.

[19] T. N. Ford, D. Lim, and J. Mertz, "Fast optically sectioned fluorescence HiLo endomicroscopy," *Journal of Biomedical Optics*, vol. 17, no. 2, p. 021105, 2012.

[20] M. Sonka, V. Hlavac, and R. Boyle, *Image processing, analysis, and machine vision*. Cengage Learning, 2014.

[21] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *International conference on machine learning*. PMLR, 2015, pp. 448–456.

[22] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," in *14th international conference on artificial intelligence and statistics*, 2011, pp. 315–323.