International Journal of
Molecular Sciences

MDPI

Article

Multiple Parallel Fusion Network for Predicting Protein Subcellular Localization from Stimulated Raman Scattering (SRS) Microscopy Images in Living Cells

Zhihao Wei [1], Wu Liu [1], Weiyong Yu [1], Xi Liu [1], Ruiqing Yan [1], Qiang Liu [1] and Qianjin Guo [1,2,*]

[1] Academy of Artificial Intelligence, Beijing Institute of Petrochemical Technology, Beijing 102617, China
[2] School of Mechanical Engineering & Hydrogen Energy Research Centre, Beijing Institute of Petrochemical Technology, Beijing 102617, China
* Correspondence: guoqianjin@bipt.edu.cn or guoqj@iccas.ac.cn

Abstract: Stimulated Raman Scattering Microscopy (SRS) is a powerful tool for label-free detailed recognition and investigation of the cellular and subcellular structures of living cells. Determining subcellular protein localization from the cell level of SRS images is one of the basic goals of cell biology, which can not only provide useful clues for their functions and biological processes but also help to determine the priority and select the appropriate target for drug development. However, the bottleneck in predicting subcellular protein locations of SRS cell imaging lies in modeling complicated relationships concealed beneath the original cell imaging data owing to the spectral overlap information from different protein molecules. In this work, a multiple parallel fusion network, MPFnetwork, is proposed to study the subcellular locations from SRS images. This model used a multiple parallel fusion model to construct feature representations and combined multiple nonlinear decomposing algorithms as the automated subcellular detection method. Our experimental results showed that the MPFnetwork could achieve over 0.93 dice correlation between estimated and true fractions on SRS lung cancer cell datasets. In addition, we applied the MPFnetwork method to cell images for label-free prediction of several different subcellular components simultaneously, rather than using several fluorescent labels. These results open up a new method for the time-resolved study of subcellular components in different cells, especially cancer cells.

Keywords: label-free live cell imaging; protein subcellular localization; nonlinear optical microscopy; multiple parallel fusion network; deep learning

check for updates

Citation: Wei, Z.; Liu, W.; Yu, W.; Liu, X.; Yan, R.; Liu, Q.; Guo, Q. Multiple Parallel Fusion Network for Predicting Protein Subcellular Localization from Stimulated Raman Scattering (SRS) Microscopy Images in Living Cells. *Int. J. Mol. Sci.* **2022**, *23*, 10827. https://doi.org/10.3390/ijms231810827

Academic Editor: Ancha Baranova

Received: 17 August 2022
Accepted: 13 September 2022
Published: 16 September 2022

1. Introduction

Cells can be divided into different organelles for metabolic processes and complicated intra-cellular organizations. They are the basic biological, structural, and functional unit of all living organisms, and the dysfunction of organelles is often closely linked to the occurrence, development, and metastasis of tumors [1–6]. Accordingly, the detection and digging deeper into the structure, function, and micro-environment of cell organelles will help us further learn about the important role of organelles in life activities and provide effective suggestions for the diagnosis and treatment of cell organelle-related diseases [5–10]. However, the resolution of organelle structure in the natural tissue environment and its functional consequences are still not clear [6–12]. Accordingly, the capability of imaging, extracting, and exploring cells and their subcellular compartments is essential in various research fields, such as cell physiology and pathology, and is closely related to a variety of diseases.

Based on the above reasons, many related technologies have been developed and applied to cells and their subcellular compartment detection and research [13–16]. These detection methods mainly include traditional biological methods such as gel electrophoresis, protein immunoblotting, and mass spectrometry, while emerging microscopy technologies include electron microscopy, atomic force microscopy, and different type of optical imaging technologies such as fluorescence imaging technology, confocal microscopy imaging, phase contrast imaging, Raman imaging technology, and super-resolution fluorescence microscopy. These methods are extensively applied in single-cell investigations and offer an important way to study different angles of cell information [13–22].

Compared with other cell imaging methods, the optical method of single-cell imaging has certain improved properties, such as high detection sensitivity, high quality, and low cost, which tremendously boost the proceedings of non-destructive cell research [18–22]. Recently, large amounts of unlabeled optical imaging instruments have been utilized for cell surveys [18–23]. Compared with pathological images, which need staining, and fluorescent images, which need labeling, label-free optical imaging overcome the unfavorable influence of staining reagents on cytoactive and cell signal transduction and can be used for long time detection in tissues and living cells [20–23]. Therefore, there is increasing demand to develop advanced label-free cell optical imaging analysis methods to mine the rich information contained in the optical cell images [24–26].

Due to the successive enhancement and usability of sophisticated computing power and analytical methods in recent years, the deep neural network learning method has been prevalent in the field of label-free cell optical imaging techniques for deeply exploring cellular structure and morphological information [25–28]. In comparison with the conventional intelligence method, deep learning is able to automatically perform a series of target recognition, feature extraction, and analysis, which makes it possible to automatically discover image target features, and automatically explore feature levels and interactions [27–30]. The learning-enhanced cell optical image analysis model is capable of acquiring texture details from low-level source images and achieving higher resolution for label-free cell optical imaging techniques [29–32]. The deep learning pipeline of cell optical microscopy imaging can extract complex data representations in a hierarchical way, which is helpful for finding hidden cell structures from microscope images, such as the size of a single cell, the number of cells in a given area, the thickness of the cell wall, the spatial distribution between cells, subcellular components and their density, etc. [32–46]. The U-net model has been proven effective in semantic segmentation and label-free prediction for cell optical microscopy imaging [33]. However, for the multiple spectral channels, the spatial and spectral information are mixed during the feature encoding in a problematic fashion for image reconstruction in traditional Unet architecture. Although other recently reported modifications of Unet, such as UwU-Net architecture, are good at extracting local feature regions and can dedicate tunable free parameters to both spectral and spatial information, they still experience difficulty in capturing global representations [36].

Motivated by the above analysis, the goal of this work was to employ a multiple parallel fusion Deep Networks framework that boosts the performance of label-free cell optical imaging techniques when solving sub-cellular organelle location problems. In our work, we find that the multiple parallel fusion method (MPF)—which incorporates all the merit features of both high-resolution local detailed spatial information from CNN features and the global context information from transformers—presents a better way to predict the location of cellular organelles from label-free cell optical images compared with previous CNN-based self-attention methods. Moreover, we demonstrate that subcellular structures could be more precisely reconstructed with the combination of transformer and Unet than both methods working individually. The model also has a strong generalization ability and can be extended to the new cell imaging investigation.

2. Results

2.1. Experimental Settings

To compare the performance of different models, the setting of experimental parameters should be as consistent as possible. Firstly, the development, training, prediction, and image processing of all models were calculated by Pytorch platforms, and the graphics card of the server adopts Geforce RTX 3080. Secondly, during model training, the value of momentum was set at 0.9, the value of the batch size was set to 8, and the weight attenuation for the training neural network was set to 1×10^{-4}. At the same time, the maximum number of epochs for the contrasting models was set at 200. In order to select the initial learning rate, a series of values was computed to test its training effect in the model. According to the experimental comparison, it was proved that 0.001 is the best choice to set as the initial learning rate.

The neural network training curves for three different prediction methods are shown in Figure 1. For the better-performing multiple parallel branch fusion model, as Figure 1 depicts, the training process only took about 120 steps until the training accuracy increased by over 96%. The error decay in Figure 1 demonstrates that the method with the MPFnet mechanism gained better performance on different training samples in comparison with the classical Unet and UwUnet model, where our strategy avoids overfitting because the error does not increase with the change of training mode, and the error attenuation remains stable.

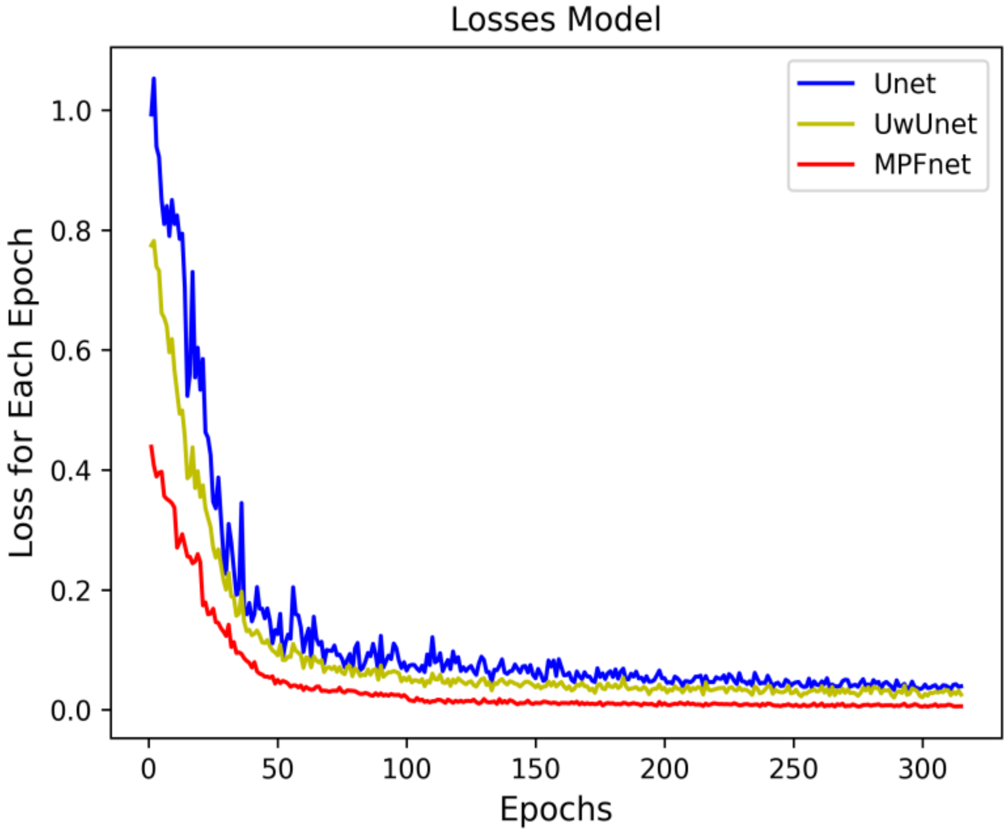


Figure 1. The neural network training curves for three prediction models include Unet, UwUnet, and MPFnet.

2.2. Metrics for Performance Evaluation

In order to verify the credibility of predictions, five quantified metrics were applied to measure the performance of different prediction algorithms. All the evaluation metrics mentioned above can be consecutively calculated as follows.

The accuracy (AY) and overall accuracy (OA) are common standard metrics for predicting subcellular locations, which can be calculated as follows:

$$AY(i) = \frac{R(i)}{S(i)} \tag{1}$$

$$OA = \frac{\sum_{i=1}^{10} R(i)}{\sum_{i=1}^{10} S(i)} \tag{2}$$

where $R(i)$ is the correctly predicted values in the ith subcellular locations, and $S(i)$ represents the total values in the ith subcellular locations.

Mean Intersection over Union (*MIoU*) is another standard metric for segmentation purposes [47]. Intersection over Union (*IoU*) is a ratio computed on a per-class basis between the ground truth and the protein subcellular location prediction. Mean Intersection over Union is the average IoU ratio, which can be calculated as follows:

$$IoU = \frac{T \bigcap P}{T \bigcup P} \tag{3}$$

$$MIoU = \frac{1}{k+1} \sum_{i=0}^{k} \frac{p_{ii}}{\sum_{j=0}^{k} p_{ij} + \sum_{j=0}^{k} p_{ji} - p_{ii}} \tag{4}$$

where it is assumed that the total of classes is (k + 1) and p_{ij} is the number of pixels of class i inferred to class j. p_{ii} represents the number of true positives, while p_{ij} and p_{ji} are usually interpreted as false positives and false negatives, respectively.

Pearson's correlation coefficient (PCC) ($r_{py} \in [-1, 1]$) is another metric that gives the relationships between the feature values and the predicted values by measuring the correlation between the pixels of the true and predicted images [48]. Given N sample pairs $\{(p_1, y_1), \ldots, (p_N, y_N)\}$, we obtain:

$$r_{py} = \frac{\sum_{i=1}^{N}(p_i - \overline{p})(y_i - \overline{y})}{\sqrt{\left(\left(\sum_{i=1}^{N}(p_i - \overline{p})^2\right)\left(\sum_{i=1}^{N}(y_i - \overline{y})^2\right)\right)}} \tag{5}$$

where $\overline{p}$ and $\overline{y}$ are the sample means. Note that when p_i and y_i are binary, r_{py} becomes the Matthews correlation coefficient, which is known to be more informative than the F_1 score (Dice coefficient) on imbalanced datasets.

MSE (Mean Square Error) is a function that is used to evaluate the difference between the targeted values and the predicted values [49]. *RMSE* (Root Mean Square Error) further evaluates the spatial detail information between images, while *NRMSE* (Normalized Root Mean Square Error) normalizes *RMSE* for easier observation and comparison. For the image prediction work, the *NRMS* can be applied in computing the accuracy between a pixel in the predicted image and the same pixel in the true image, which is obtained by:

$$MSE = \frac{1}{M \times N} \sum_{i=1}^{M} \sum_{j=1}^{N} \left(u'(i,j) - u(i,j)\right)^2 \tag{6}$$

$$RMSE = \sqrt{MSE(u', u)} \tag{7}$$

$$NRMSE = \frac{\sqrt{\frac{1}{M \times N} \sum_{i=1}^{M} \sum_{j=1}^{N} \left(u'(i,j) - u(i,j)\right)^2}}{u'(i,j)_{max} - u'(i,j)_{min}} \tag{8}$$

where $u'(i,j)$, $u(i,j)$ represent the image to be evaluated and the original image, respectively. N represents the length and width of the image.

Peak Signal to Noise Ratio (*PSNR*) is the most commonly used metric in image quality assessment, which can be obtained by:

$$PSNR = 10\, log_{10} \left(\frac{m_x \times m_y \times V_{max}^2}{\sum_{r,t}[t(x,y) - d(x,y)]^2} \right) \tag{9}$$

where V_{max} denotes the maximum predicted value of the source image. *t(x,y)* is the matrix of the raw source image, *d(x,y)* is the matrix of the noise-removed image. (*x*,*y*) denotes the pixel coordinate in a given $m_x \times m_y$ image.

Structural similarity index (SSIM) can be used as a quality evaluation index for similarity comparison among image prediction results, which can be obtained by:

$$s(x,y) = \frac{\sigma_{xy} + c_3}{\sigma_x \sigma_y + c_3} \tag{10}$$

$$l(x,y) = \frac{2\mu_x\mu_y + c_1}{\mu_x^2 + \mu_y^2 + c_1} \tag{11}$$

$$c(x,y) = \frac{2\sigma_x\sigma_y + c_2}{\sigma_x^2 + \sigma_y^2 + c_2} \tag{12}$$

The SSIM value is calculated for signals and images after combining Equations (10)–(12) as:

$$\begin{aligned} SSIM(x,y) &= [l(x,y)]^m[c(x,y)]^n[s(x,y)]^p \\ &= \frac{(2\mu_x\mu_y+C_1)(2\sigma_{xy}+C_2)}{(\mu_x^2+\mu_y^2+C_1)(\sigma_x^2+\sigma_y^2+C_2)} \end{aligned} \tag{13}$$

where m, n, and p denote the magnitude values of the structure component *s(x,y)*, the luminance component *l(x,y)*, and the contrast component *c(x,y)*, respectively. μ_x, μ_y is the average of x_i, y_i. σ_x, σ_y is the variance of x_i, y_i.

The Dice Similarity Coefficient (DSCs), also named Sørensen–Dice similarity, is another one of the frequently used metrics in medical image competitions. It is a collective similarity metric, which is usually used to calculate the similarity of two segmented images. The calculation formula of the Dice coefficient is as follows:

$$Dice = \frac{2 \times (T \bigcap P)}{T \bigcup P} \tag{14}$$

where the *Dice* similarity coefficient value threshold is [0, 1]. The best result of prediction or segmentation in a medical image is 1, and the worst result is 0.

2.3. The Analysis of the Performance of Multiple Parallel Fusion Deep Networks

The experiment of this section mainly discusses the results of the labeling-free method based on deep learning for protein subcellular localization from femtosecond stimulated Raman spectroscopic microscope images. Compared with other classical optical imaging methods, Stimulated Raman spectroscopy imaging has the advantages of not requiring fluorescent molecular markers and obtaining more information. However, the rich and overlapped information in the same collected image also causes difficulties for image analysis and feature extraction. Although some of the label-free staining methods based on Raman imaging have shown promising results in some organelles, there is still a lack of rich and effective means to predict the subtle changes of Raman spectra for single organelles.

The results for subcellular localizations can be seen in Figure 2A,B. The output results for different organelle locations, including nuclei (second column), mitochondria (third column), and endoplasmic reticulum (right) from the single raw SRS imaging cell (left), are shown in Figure 2A. One SRS raw image for A549 lung cancer cells from ATCC was output at the different positions of the multiple fusion model at the same time to determine the accuracy of subcellular localization predictions, including nuclei (top row), mitochondria (third column), and endoplasmic reticulum (bottom row). Figure 2B shows the output of nuclei (top row), mitochondria (third column), and endoplasmic reticulum (bottom row) from the transformer branch in the first column. The output from the bifusion branch and multi-model are shown in the second and third columns, respectively.

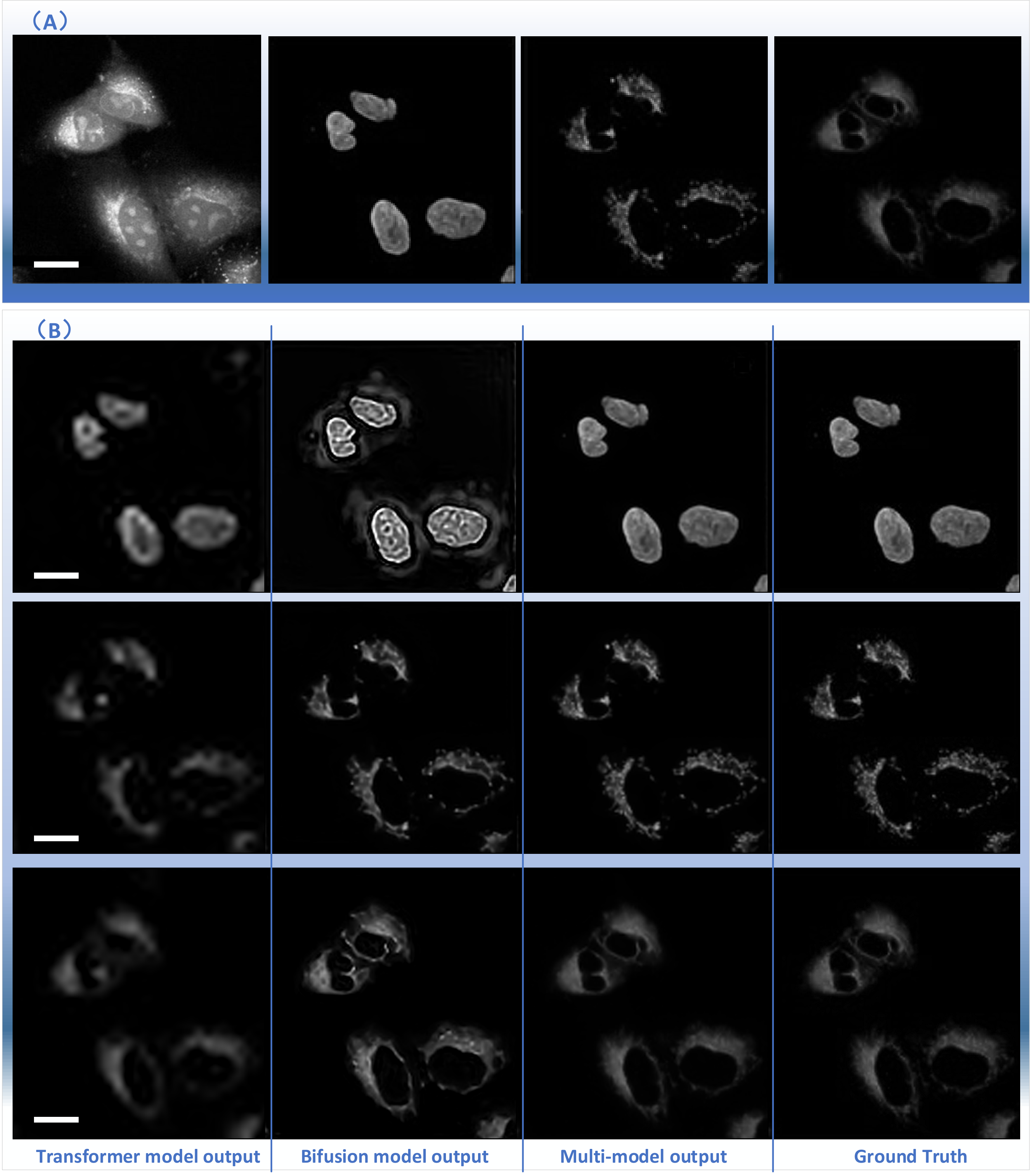


Figure 2. Subcellular localization prediction results of differentiation model from label-free cell imaging experiments. The results for different organelle locations, including nuclei (second column), mitochondria (third column), and endoplasmic reticulum (right) from the single raw SRS imaging cell (left), are shown in (**A**). The prediction results for organelle locations, including nuclei (top row), mitochondria (third column), and endoplasmic reticulum (bottom row), from the raw SRS imaging cells (left) are shown in (**B**). Scale bar, 25 μm.

2.4. Comparison of the Performance with Various Prediction Models

In this section, predicting experiments of subcellular organelle localization results are investigated and compared among different deep learning models from label-free SRS microscopy images. Even though the traditional imaging-based pipeline has cells stained, SRS imaging can give more information on cell shape and subcellular structure without using molecular probes. On the other hand, it also produces some disadvantages, such as low-contrast and complex images, which create obstacles to clearly indicating the biochemical features of cells. Therefore, for SRS cell imaging, some challenges still exist in using these deep learning-based methods to clearly identify, segment, and quantify each subcellular structure in cell optical images. As a result, some advanced analysis methods are needed to explore the rich information hidden in cell images. Based on the above reasons, the new MPFnet method is proposed in this work, which bridges the transform model and convolutional neural networks to automatically segment organelles.

To demonstrate the application of deep learning models in label-free organelle prediction, we used fluorescence imaging of the fixed lung cancer cells as a ground-truth model and SRS microscopy images as the source image model. In Figure 3, the first column shows a live-cell Raman optical image, the second column is ground-truth fluorescence images taken after the cells are stained, and the following three columns are predicted fluorescence cell images with the UwUnet method, Unet method, and MPFnet method, respectively. From the experimental analysis results, we can see that the MPFnet method can accurately predict the location of each organelle from cell optical imaging data at the same time.

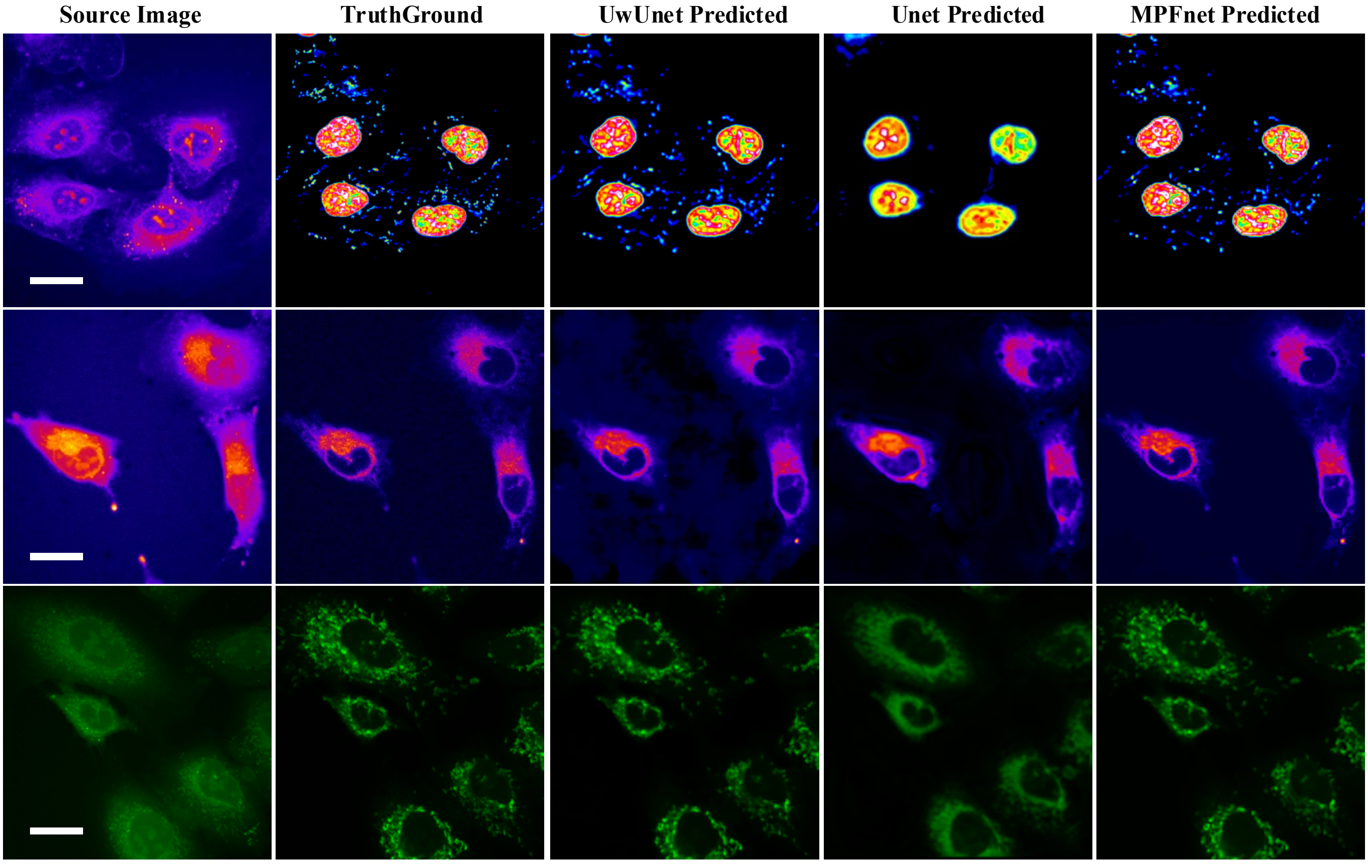


Figure 3. Predicted organelle fluorescence from hyperspectral SRS microscopy images by using UwUNet, U-Net, and MPFNet methods. The first column shows the Input SRS image, the second column shows the ground-truth fluorescence image, and the following three columns display the predicted fluorescence results by UwUNet, U-Net, and MPFNet, respectively, for nuclei (**top**), mitochondria (**middle**), and endoplasmic reticulum (**bottom**). Scale bar, 25 µm.

In order to quantitatively compare and analyze the effects of different prediction methods, we calculated the quantitative metrics: NRMSE, SSIM, PCC, DICE, and mean IoU to explore the differences between predicted and expected results of different methods, to compare the prediction performance between the methods proposed in this work with other classical methods. We firstly give a comparison of the prediction performance with the mean pixel accuracy curves and radar chart for five quantitative metrics, including NRMSE, SSIM, PCC, DICE, and mean IoU among Unet, UwUnet, and MPFnet models, as shown in Figure 4. One can observe from Figure 4 that the MPFnet ensemble method achieved the highest mean pixel accuracy of 0.92.

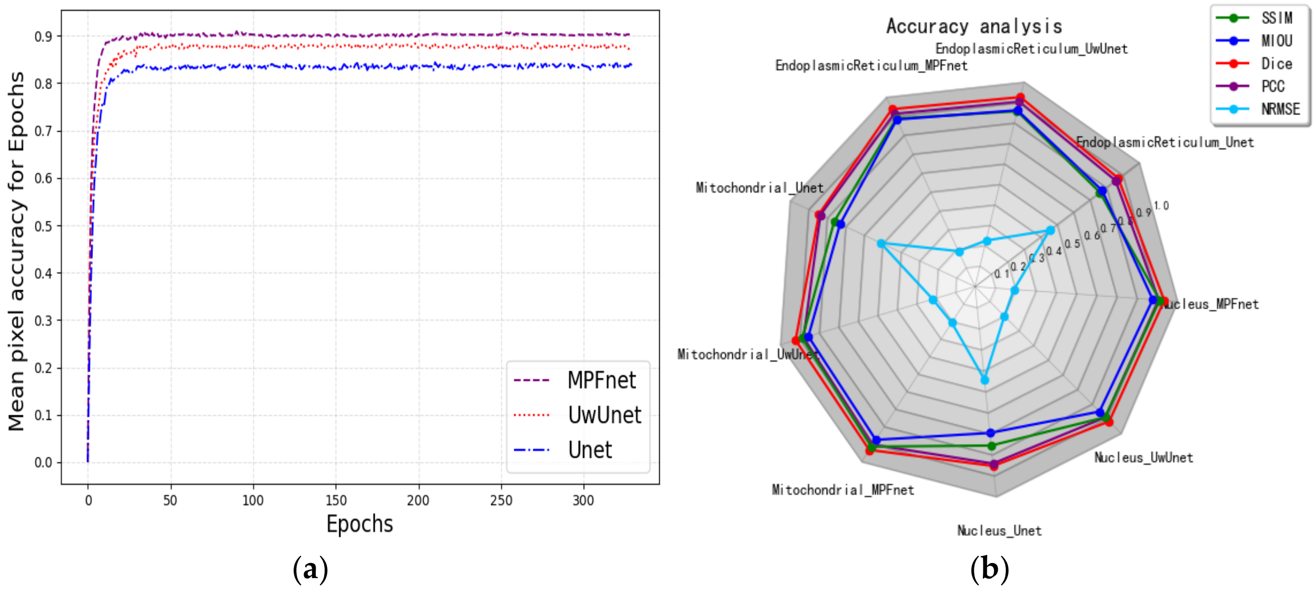


Figure 4. (**a**) Mean pixel accuracy comparison among different methods. Purple dash line, red dot line, and blue dash-dot line represent MPFnet, UwUnet, and Unet models, respectively. (**b**) Radar chart for quantitative comparisons of different predictive models.

To verify the valuable quantitative evaluation parameter in subcellular organelle localization, the segmentation performance of all trained networks was also assessed by the Dice similarity coefficients to evaluate the similarity between the ground truth fluorescence labels and model-generated images (Figure 5). Compared to the observed datasets, the Dice similarity of the prediction for the nuclei, endoplasmic reticulum, and mitochondria task set was the highest for MPFnet and the lowest for the Unet approach.

The lowest Dice similarity coefficient of the Unet method for the nuclei prediction task set was attributed to the edges of the bony structures not being included in the gold standard as it lacked the long-distance segmenting in these whole regions. The lowest Dice similarity coefficients of the Unet method for the mitochondria and endoplasmic reticulum prediction task sets were due to the larger excluded region of the outer contour area in its gold standard, while this excluded region was segmented by the Unet method.

In addition, we also measured the accuracy of label-free prediction algorithms using the mean Intersection Over Union (IOU) evaluation metric. Here, the mIOU for each epoch comparison among different methods is shown in Figure 6a. Then, we used the box plot graph to give a more visual and intuitive representation of the quantitative evaluation of mIOU parameters on the prediction results of different algorithms (Figure 6). The box plot in Figure 6 shows five statistics in the data: minimum, first quartile, median, third quartile, and maximum. In Figure 6, the minimum value is represented by the extension of the red lines at the bottom, while the maximum value is represented by the extension of the red line at the top. The range of these two black lines refers to the mIOU accuracy range, and the top and bottom of the box refer to the accuracy of the upper quartile (=0.75) and lower quartile (=0.25), respectively. The blue solid line in the box indicates the median accuracy.

It can be seen from Figure 6b–e that compared with other methods, the MPFnet method achieves the best performance on all the nuclei, mitochondria, and endoplasmic reticulum datasets. Compared with the observed datasets, MPFnet performed favorably by metric mIOU with 0.879, 0.876, and 0.881 for the nuclei, mitochondria, and endoplasmic reticulum task set, respectively, against alternative UwUnet approaches with mIOU mean values of 0.852, 0.861, 0.854. The classical Unet approach performed significantly worse with mIOU mean values of 0.716, 0.771, and 0.731.

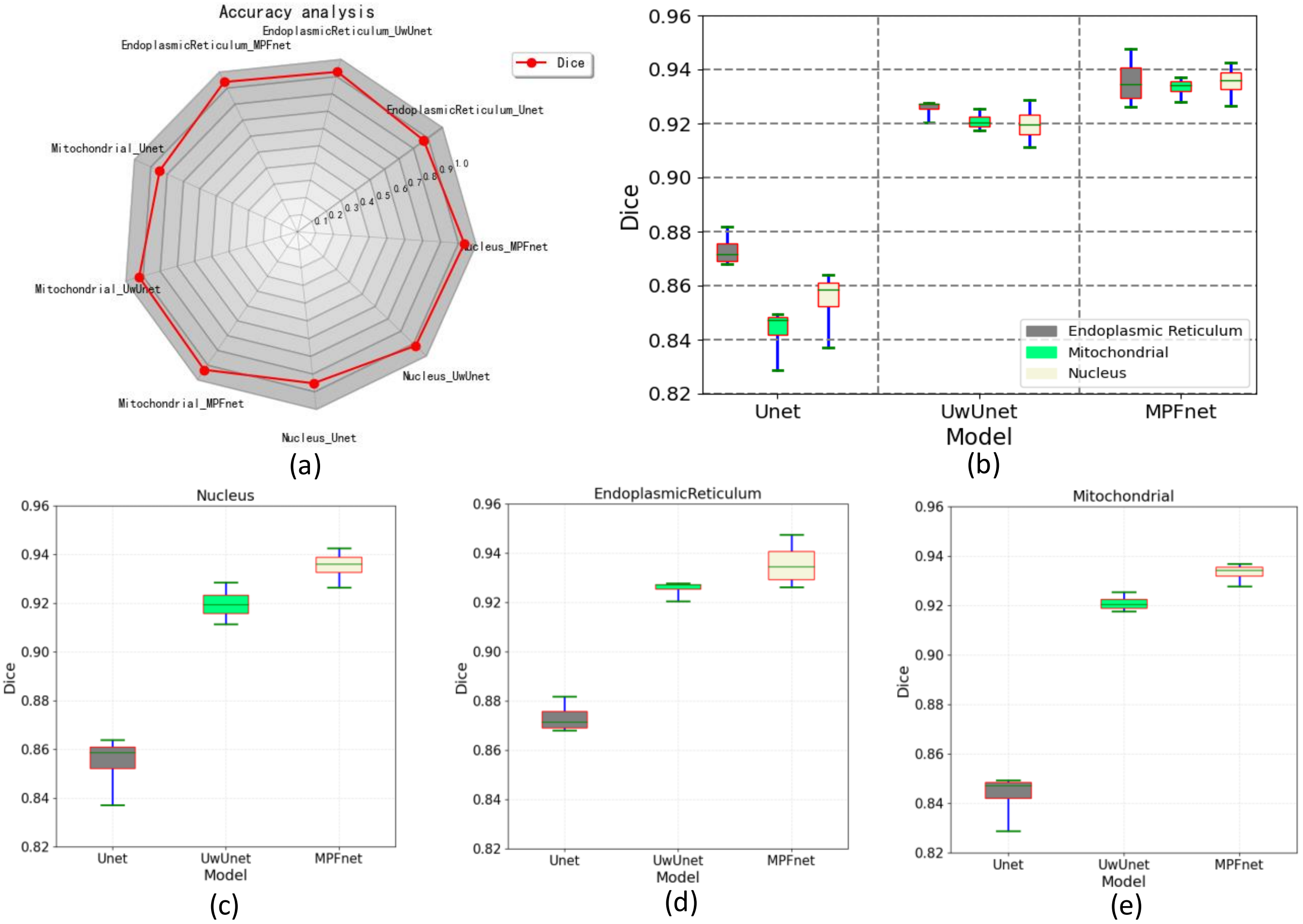


Figure 5. Comparing dice performance among various prediction algorithms. (**a**) Dice similarity coefficient radar chart for quantitative comparisons of different predictive models. (**b**) Box plot of dice values for all organelles (nuclei, mitochondria, and endoplasmic reticulum) prediction task set with MPFnet-based learning model and compared with that of varied deep neural network-based prediction models, such as Unet and UwUnet learning model on all. (**c**) Box plot of dice coefficient value on nuclei prediction task set with MPFnet-based learning model and compared with that of varied deep neural network-based prediction models, such as Unet and UwUnet learning model. (**d**) Box plot of dice coefficient value on endoplasmic reticulum prediction task set with MPFnet-based learning model and compared with that of varied deep neural network-based prediction models, such as Unet and UwUnet learning model. (**e**) Box plot of dice values on mitochondria prediction task set with MPFnet-based learning model and compared with that of varied deep neural network-based prediction models, such as Unet and UwUnet learning model.

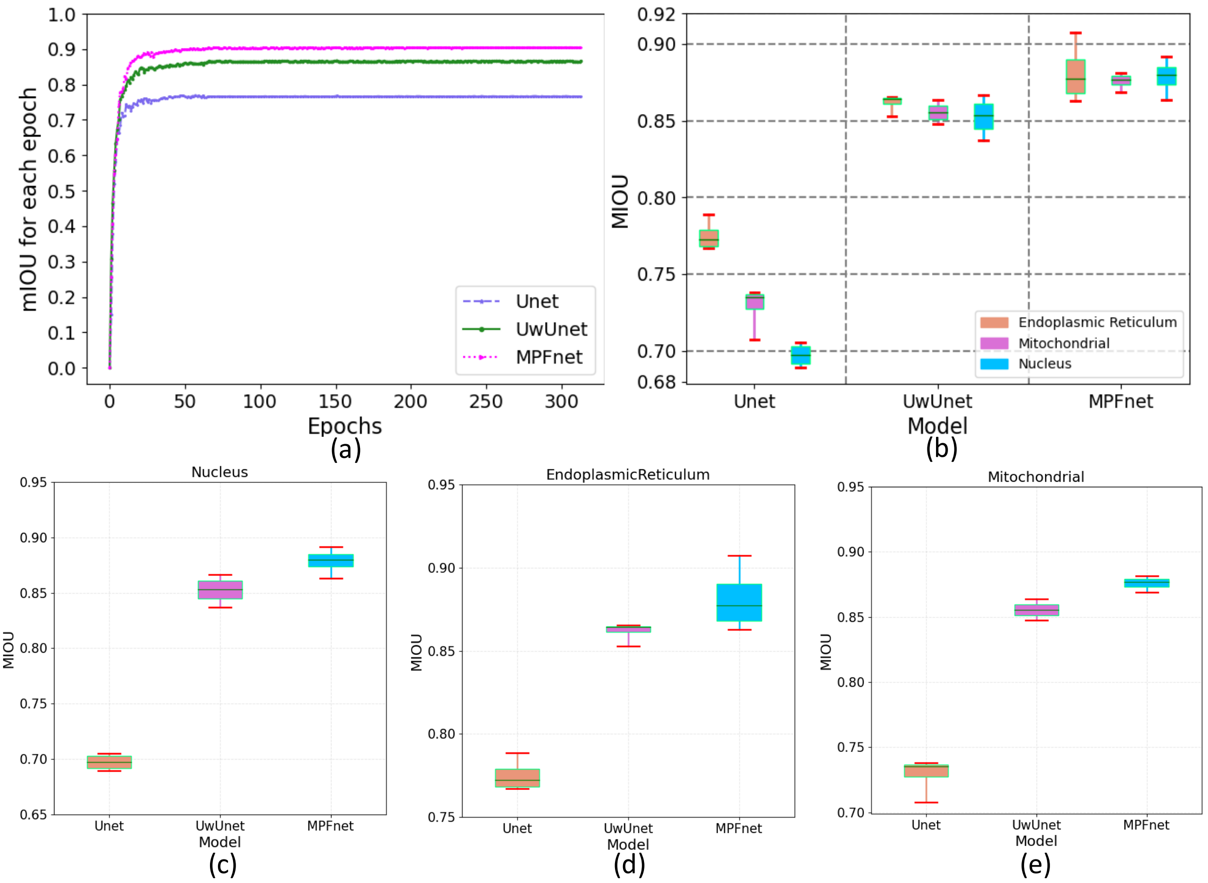
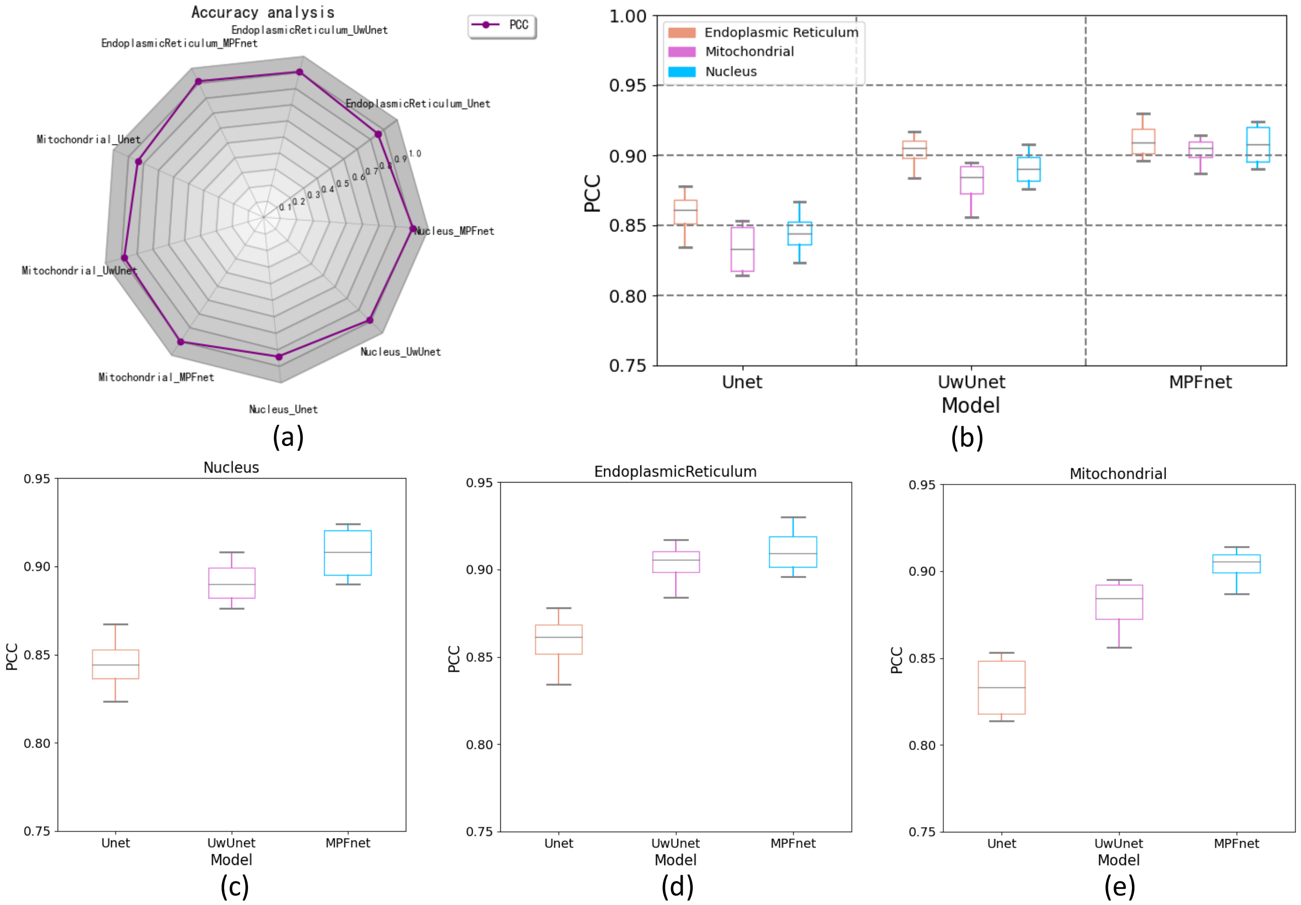


Figure 6. The mean intersection over union (mIOU) of different deep learning models over all SRS microscopy images in fixed lung cancer cells (A549, from ATCC) detection dataset. (**a**) mIOU for each epoch comparison among different methods. Unet (Purple triangle dash line), UwUnet (Green circle solid line), MPFnet (Pink arrow dotted line) represent Unet, UwUnet, MPFnet, respectively. (**b**) Box plot of mIOU accuracy of all organelles (nuclei, mitochondria, and endoplasmic reticulum) prediction task set with MPFnet-based learning model and compared with that of varied deep neural network-based prediction models, such as Unet and UwUnet learning models. (**c**) Box plot of mIOU accuracy on nuclei prediction task set with MPFnet-based learning model and compared with that of varied deep neural network-based prediction models such as Unet and UwUnet learning model. (**d**) Box plot of mIOU accuracy on endoplasmic reticulum prediction task set with MPFnet-based learning model and compared with that of varied deep neural network-based prediction models such as Unet and UwUnet learning model on all. (**e**) Box plot of mIOU accuracy on mitochondria prediction task set with MPFnet-based learning model and compared with that of varied deep neural network-based prediction models such as Unet and UwUnet learning models.

To further characterize the predictive performance of three variants of the deep learning-based predictor on organelle (nuclei, mitochondria, and endoplasmic reticulum) segmentation tasks and to give comparable measures, we also provide Pearson's correlation coefficient (PCC) to quantify the accuracy of the predictions. The PCC similarity value is used here to determine the similarity between the pixels of the truth and predicted images with various deep learning methods.

The top left radar chart of the PCC in Figure 7a shows the PCC metric value with three different test models for the cell SRS images. The box plot graph in Figure 7b–e provides a more visual and intuitive representation of the quantitative evaluation of PCC parameters on the prediction results of different algorithms. Compared to the observed datasets, MPFnet performed favorably in PCC mean value for the nuclei, endoplasmic reticulum, and mitochondria task sets, respectively, against the alternative UwUnet and classical Unet approaches. Comparing the PCC value with other methods, MPFnet results had good advantages over other methods. Therefore, the PCC similarity value in Figure 7a–e shows that the MPFnet method achieved top performance on all the nuclei, mitochondria, and endoplasmic reticulum datasets. Based on the PCC coefficients, our selected MPFnet method scheme achieved the best performance in all data sets.



(a) (b)

(c) (d) (e)

Figure 7. Comparing PCC performance among various prediction algorithms. (**a**) PCC radar chart for quantitative comparisons of different predictive models. (**b**) Box plot of PCC value over all organelle (nuclei, mitochondria, and endoplasmic reticulum) prediction task sets with MPFnet-based learning model and compared with that of varied deep neural network-based prediction models, such as Unet and UwUnet learning model. (**c**) Box plot of PCC value on nuclei prediction task set with MPFnet-based learning model and compared with that of varied deep neural network-based prediction models such as Unet and UwUnet learning model. (**d**) Box plot of PCC values on endoplasmic reticulum prediction task set with MPFnet-based learning model and compared with that of varied deep neural network-based prediction models such as Unet and UwUnet learning models. (**e**) Box plot of PCC values on mitochondria prediction task set with MPFnet-based learning model and compared with that of varied deep neural network-based prediction models such as Unet and UwUnet learning model.

Moreover, the NRMSE metric value was also employed in this section to assess the effectiveness of the different models by comparing the predicted images and the actual images. The top-right side of Figure 8a shows the radar chart of the NRMSE with three different test models for the cell images. Compared with other methods, it can be seen in Figure 8a–e that the MPFnet method achieved the top performance on all nuclei, endoplasmic reticulum, and mitochondria datasets. Compared to the observed datasets, MPFnet performed favorably in NRMSE mean values of the nuclei, endoplasmic reticulum, and mitochondria task sets, respectively, against the alternative UwUnet and classical Unet approaches.

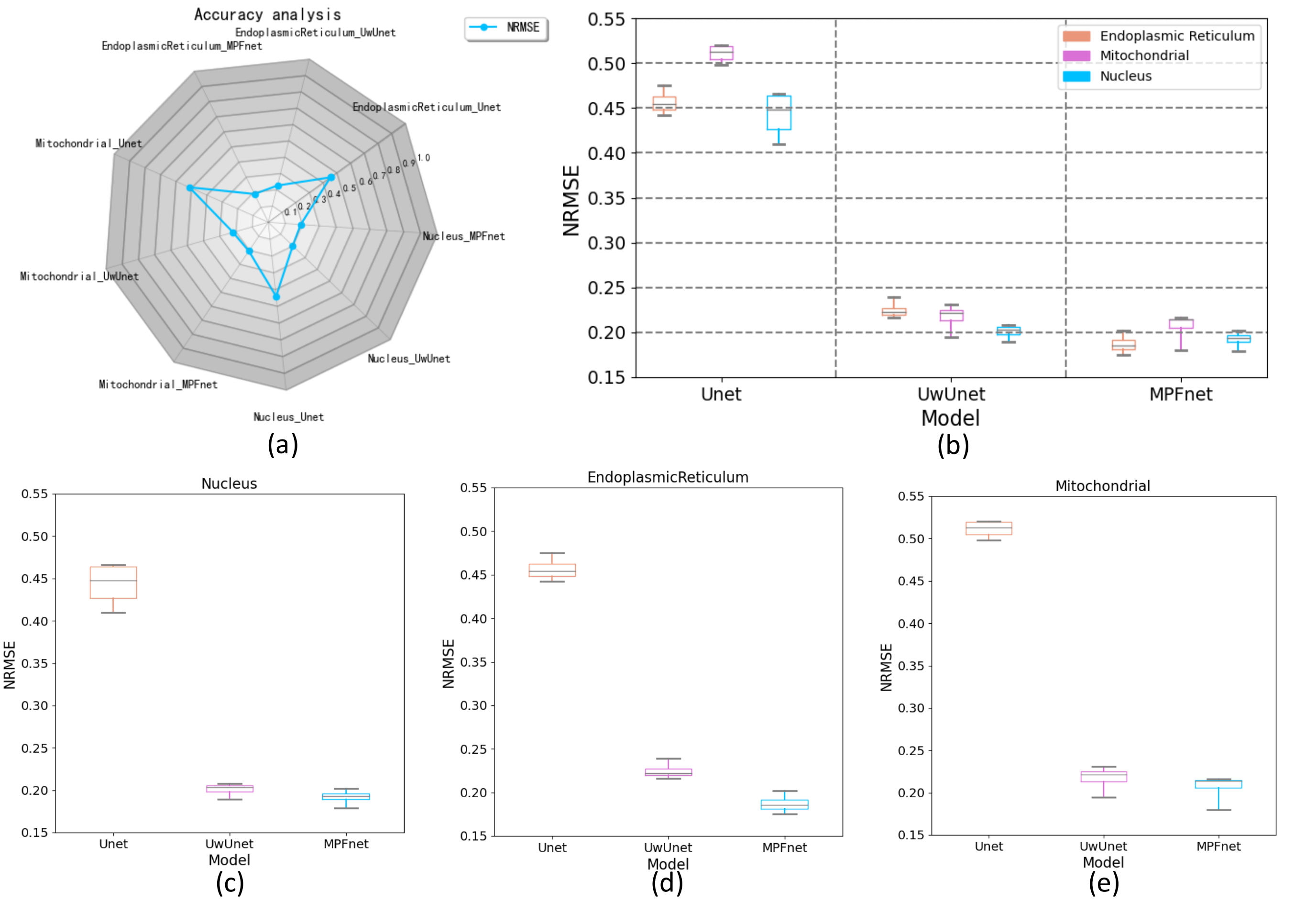


Figure 8. Comparing NRMSE performance among various prediction algorithms. (**a**) NRMSE radar chart for quantitative comparisons of different predictive models. (**b**) Box plot of NRMSE value over all organelles (nuclei, mitochondria, and endoplasmic reticulum) prediction task set with MPFnet-based learning model and compared with that of varied deep neural network-based prediction models such as Unet and UwUnet learning models. (**c**) Box plot of NRMSE value on nuclei prediction task set with MPFnet-based learning model and compared with that of varied deep neural network-based prediction models such as Unet and UwUnet learning model. (**d**) Box plot of NRMSE values on endoplasmic reticulum prediction task set with MPFnet-based learning model and compared with that of varied deep neural network-based prediction models such as Unet and UwUnet learning model. (**e**) Box plot of NRMSE values on mitochondria prediction task set with MPFnet-based learning model and compared with that of varied deep neural network-based prediction models such as Unet and UwUnet learning model.

At last, the SSIM index, which is based on structure, luminance, and contrast comparison, was applied for quality assessment in various different prediction methods. The radar chart of the SSIM quality assurance with three different test models for the cell images is shown in Figure 9a. A small SSIM index appeared in all data sets in the Unet method corresponding to the target area of prediction, having a relatively large difference between the ground truth fluorescence labels and model-generated images. Moreover, from the box plot of SSIM values in the all-organelles (nuclei, mitochondria, and endoplasmic reticulum) prediction task set with three different methods, we can obtain more definite quantitative comparison information. As shown in Figure 9b–e, MPFnet performed favorably in SSIM mean value for the nuclei, endoplasmic reticulum, and mitochondria task sets against alternative UwUnet and classical Unet approaches. The MPFnet method achieved the top SSIM performance on all nuclei, endoplasmic reticulum, and mitochondria datasets.

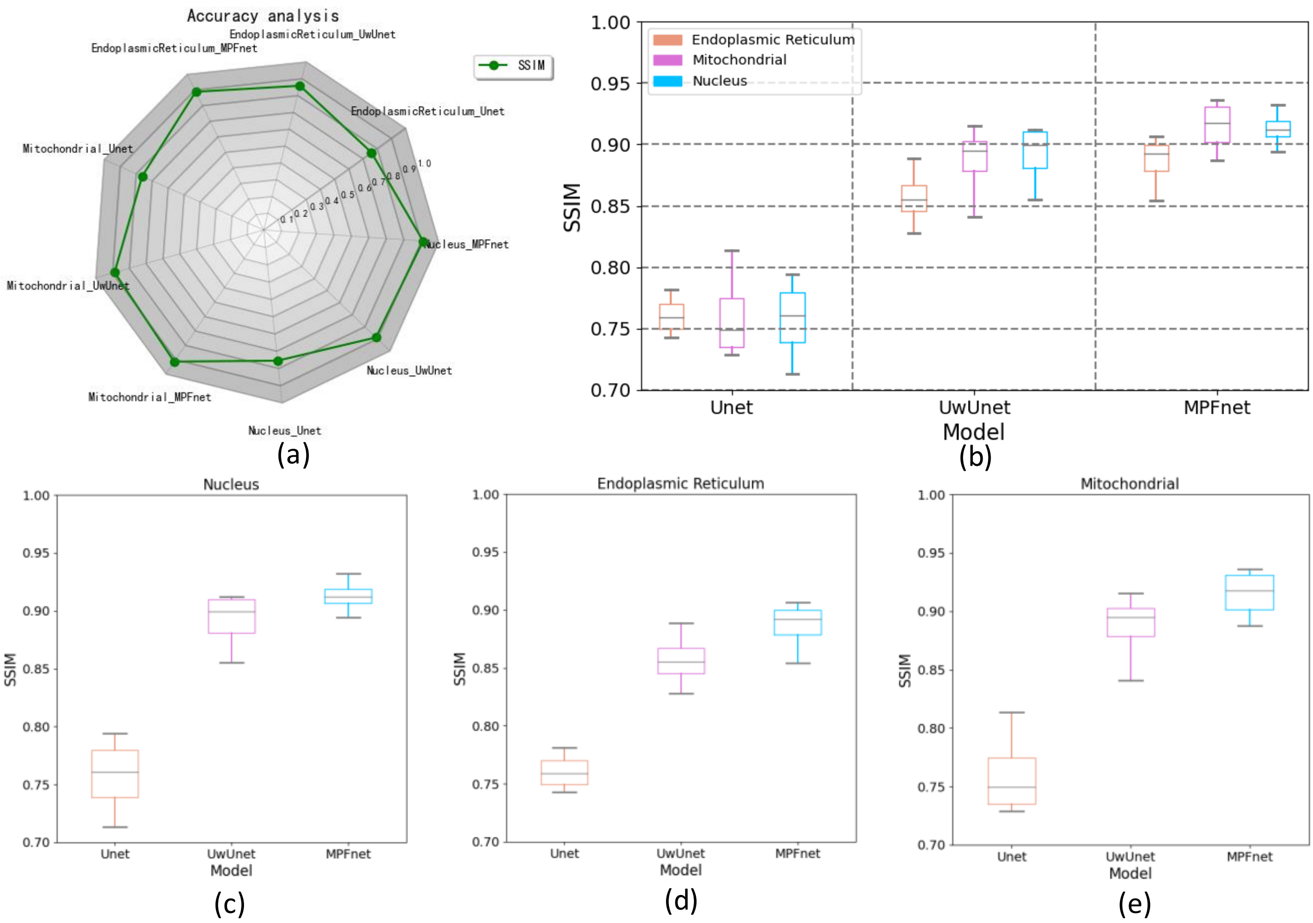


Figure 9. Comparing SSIM performance among various prediction algorithms. (**a**) The SSIM performance measures typical of predictive tasks. (**b**) Box plot of SSIM values in all organelle (nuclei, mitochondria, and endoplasmic reticulum) prediction task sets with the MPFnet-based learning model and compared with that of varied deep neural network-based prediction models such as Unet and UwUnet. Box plot of SSIM values on (**c**) nuclei, (**d**) endoplasmic reticulum, and (**e**) mitochondria prediction task sets with MPFnet-based learning model and compared with that of varied deep neural network-based prediction models such as Unet and UwUnet learning model.

In terms of explicit and quantitative analysis, the detailed evaluation results and calculations are shown in Tables 1–3. For different prediction models, five different quantitative parameters (NRMSE, SSIM, PCC, Dice, and mIOU) were computed to compare and analyze the accuracy of protein subcellular localization from label-free live cell imaging. Tables 1–3 present the label-free prediction results of three variants of deep learning-based predictors on organelle (nuclei, mitochondria, and endoplasmic reticulum) segmentation tasks in terms of quality metric values with NRMSE, SSIM, PCC, DICE, and mIOU. Compared with the Unet and UwUnet prediction methods, our proposed MPFnet method outperformed all quality metric values with NRMSE, SSIM, PCC, DICE, and mean IOU. Especially for the nuclei prediction task, MPFnet achieved a 3.2% improvement over the UwUnet method and a 22.7% enhancement over the Unet method in terms of mIOU. For the mitochondria prediction task, MPFnet achieved a 2.6% improvement over the UwUnet method and a 19.8% enhancement over the Unet method in terms of mIOU. To sum up, through the comprehensive analysis of mIOU quantitative indicators corresponding to different methods (Tables 1–3), we can draw a more accurate conclusion from the quantitative standard that our method is the best of all tested methods.

Table 1. Comparison of quality measures for label-free prediction results with MPFnet model. Here, ↓ indicates that the lower the index value, the better the performance; ↑ indicates that the higher the index value, the better the performance of the model.

Organelle/ Model	**MFPnet Model**				
	NRMSE↓	**SSIM↑**	**PCC↑**	**Dice↑**	**mIOU↑**
Nucleus	0.192 ± 0.013	0.913 ± 0.019	0.908 ± 0.018	0.935 ± 0.009	0.879 ± 0.014
Endoplasmic Reticulum	0.187 ± 0.015	0.886 ± 0.032	0.911 ± 0.019	0.936 ± 0.011	0.881 ± 0.016
Mitochondrial	0.206 ± 0.028	0.915 ± 0.028	0.903 ± 0.016	0.933 ± 0.005	0.876 ± 0.007

Table 2. The prediction result measures of protein subcellular localization using the UwUnet model. Here, ↓ indicates that the lower the index value, the better the performance; ↑ indicates that the higher the index value, the better the performance of the model.

Organelle\Model	**UwUnet Model**				
	NRMSE↓	**SSIM↑**	**PCC↑**	**Dice↑**	**mIOU↑**
Nucleus	0.201 ± 0.012	0.892 ± 0.037	0.892 ± 0.016	0.919 ± 0.009	0.852 ± 0.015
Endoplasmic Reticulum	0.225 ± 0.014	0.857 ± 0.030	0.903 ± 0.019	0.925 ± 0.005	0.861 ± 0.009
Mitochondrial	0.217 ± 0.017	0.886 ± 0.046	0.880 ± 0.024	0.921 ± 0.004	0.854 ± 0.008

Table 3. Comparison of quality measures for label-free prediction results with the Unet model. Here, ↓ indicates that the lower the index value, the better the performance; ↑ indicates that the higher the index value, the better the performance of the model.

Model/Organelle	**Unet Model**				
	NRMSE↓	**SSIM↑**	**PCC↑**	**Dice↑**	**mIOU↑**
Nucleus	0.442 ± 0.024	0.757 ± 0.044	0.843 ± 0.024	0.855 ± 0.018	0.716 ± 0.008
Endoplasmic Reticulum	0.454 ± 0.022	0.761 ± 0.021	0.856 ± 0.022	0.873 ± 0.008	0.771 ± 0.013
Mitochondrial	0.511 ± 0.013	0.760 ± 0.053	0.835 ± 0.021	0.843 ± 0.015	0.731 ± 0.021

Furthermore, not only was the mIOU metric used as an evaluation index, but the NMSE quantitative indicators were also utilized to compare and analyze the performance of different prediction models in this section. For the nuclei prediction test set shown in Tables 1–3, the obtained NMSE of the MPFnet model was 0.192, which has less than half that of the classical UNet model, and a 4.48% reduction compared to the UwUnet model. On the mitochondria prediction test set, the lowest NMSE value was acquired by the MPFnet model as 0.206, achieving a 59.69% depression compared to the classical UNet model and a 5.07% decrement compared to the UwUnet model. As for the endoplasmic reticulum prediction test set, the obtained NMSE of the MPFnet model was also lowest at 0.187, which has less than half that of the classical UNet model, and 16.89% diminution compared to the UwUnet model.

In order to further explore the prediction performance of different models, we give more calculations of the correlated pixels for the obtained organelle fluorescence images and the predicted organelle fluorescence from SRS microscopy images with three variants of deep learning-based predictors, respectively. Another quantitative parameter, PCC, was also applied to detect the consistency between the prediction results and the target values to further study the variability. From Tables 1–3, it can be observed that the MPFnet model shows top performance in terms of PCC coefficient. The predicted PCC value of the nuclear validation set was as high as 0.908 for our proposed MPFnet model. Similar results were also observed in mitochondrial samples and endoplasmic reticulum test sets (Pearson's r = 0.903, 0.911, respectively). In terms of nuclei, mitochondria, and endoplasmic reticulum test sets, the PCC similarity coefficient results from MPFnet were all higher than the classical UNet performance as follows: 7.71%, 8.14%, and 6.43%, and the PCC similarity coefficient results from MPFnet were all higher than the UwUNet performance as follows: 1.79%, 2.61%, and 0.89%.

In addition, the SSIM metric was also used as the evaluation index for similarity comparisons among image prediction results. From the nuclei prediction test set shown in Tables 1–3, the obtained SSIM of the MPFnet model was 0.913, which achieved a 20.61% improvement over the classical UNet model with 0.757, and a 2.35% increment compared to the UwUnet model. Meanwhile, on the mitochondria prediction test set, we observed that the highest SSIM value was acquired by the MPFnet model as 0.915, a 20.39% improvement compared to the classical UNet model and a 3.27% increase compared to the UwUnet model. As for the endoplasmic reticulum prediction test set, the obtained NMSE of the MPFnet model was also best at 0.886, which improved 16.43% over the classical UNet model and 3.38% compared to the UwUnet model.

At the same time, another quantitative parameter, Dice, was also applied to detect the consistency between the prediction results and the target values to further study the variability. Tables 1–3 show that the MPFnet model had top performance in terms of Dice coefficients. The predicted Dice value of the nuclear validation set was as high as 0.935 for our proposed MPFnet model. Similar results were also observed in mitochondrial samples and endoplasmic reticulum test sets (Dice r = 0.933, 0.936, respectively). In terms of nuclei, mitochondria, and endoplasmic reticulum test sets, the Dice coefficient results from MPFnet were all higher than classical UNet performance as follows: 9.36%, 10.68%, and 9.35%, and the Dice metric coefficient results from MPFnet were all higher than the UwUNet performance as follows: 1.74%, 1.30%, and 1.19%.

3. Discussion

From the comparison of protein subcellular localization results among various prediction algorithms shown in Figure 10, it can be clearly observed that the Unet model was good at extracting local feature regions (red box on the second column), but experienced difficulty in capturing global representations (region of green box on the second column). Compared with CNN, the global cues were still blurred (green box on the third column), even though the local details were better for the UwUnet (red box on the third column). In contrast, the global cues were significantly enhanced (green box on the third column), and the local details were retained for the MPFnet model (red box on the third column).

To sum up, through the comprehensive analysis of all three quantitative indicators in Tables 1–3, we can draw a more accurate conclusion from the quantitative standard that our method is the best among all modules in Tables 1–3. In conclusion, our results show that deep learning creates some new opportunities for accurately predicting the location of cellular organelles from label-free cell optical images. Existing U-net-based medical image prediction methods are insufficient for catching long-range dependencies in tested images. The multiple parallel fusion predictor combines the merits of transform and UNet methods. The new multiple parallel fusion method can intelligently reveal and extract the nonlinear correlation between features to improve prediction performance. Additionally, as illustrated in Section 4.2, our deep learning approach also improves the image SNR, which offers a solution to highly suppress image artifacts and solve the distortion problems for high-speed SRS cell imaging.

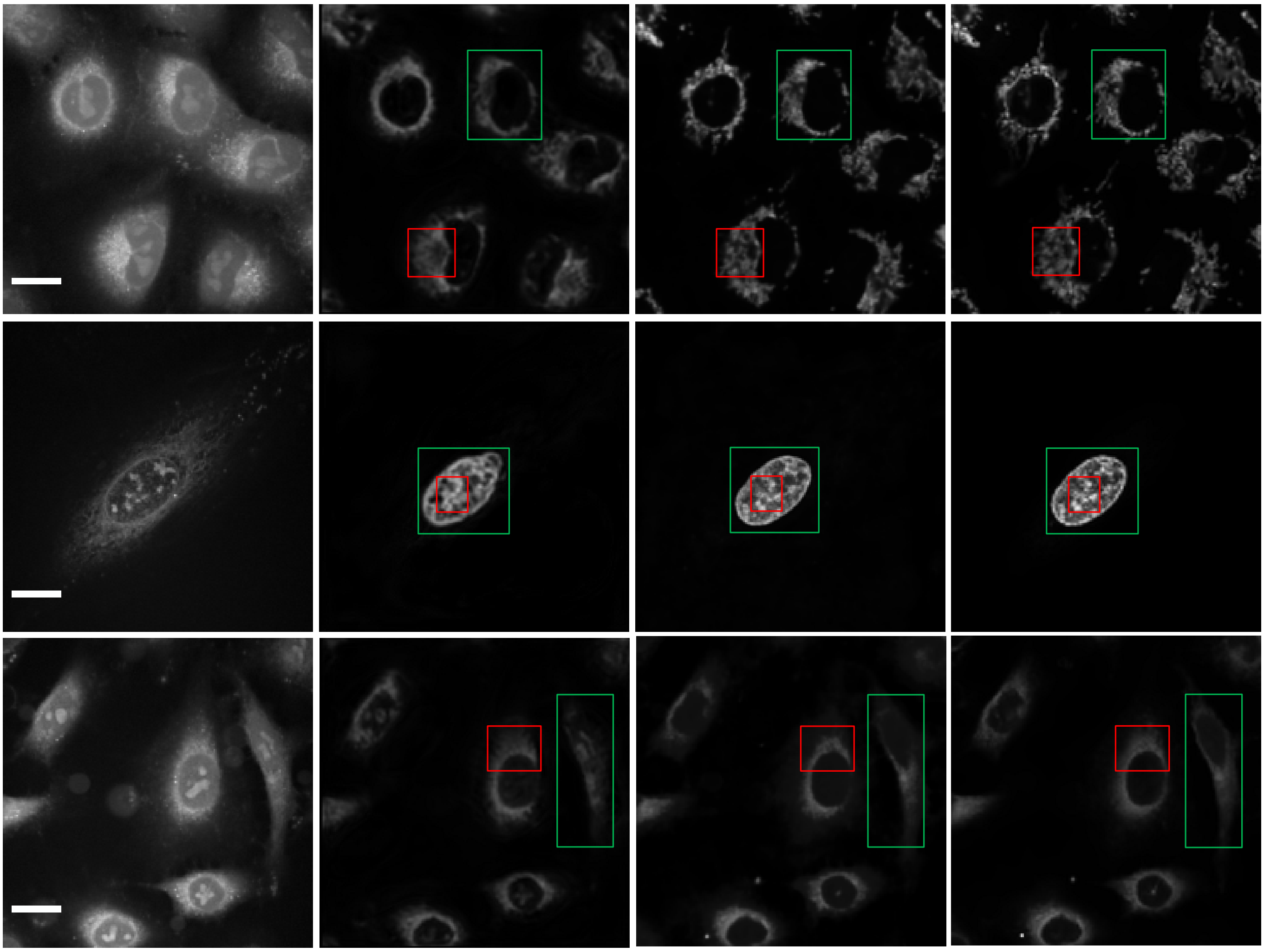


Figure 10. Comparison of the protein subcellular localization results from raw cell images (**first column**) among various prediction algorithms, including Unet (**second column**), UwUnet (**third column**), and MPFnent (**fourth column**). Local feature regions (red box), global representations (green box). Scale bar, 25 µm.

4. Methods and Materials

4.1. Experiment of the Simultaneous SRS and Fluorescence Microscopy

The complete experiment and process of predicting the subcellular protein localization based on the deep learning network are shown in Figure 11. A dual-output, 80 MHz femtosecond pulsed laser (InSight, Spectra-Physics, USA) provides the pump beam (tunable from 680 to 1300 nm) and the Stokes beam (fixed at 1040 nm) for the SRS system. Both the synchronized pump beam with 798 nm and Stokes beam with 1040 nm through the time delay stage were combined on a dichroic mirror before being directed through the microscope by the two-dimensional galvo scanning. The SRS loss signal from the lock-in amplifier and fluorescence signal from the photomultiplier tube were collected simultaneously. Images were acquired with 512 × 512 pixels, and a pixel dwell time of 8 µs at each of the ten vibrational transitions.

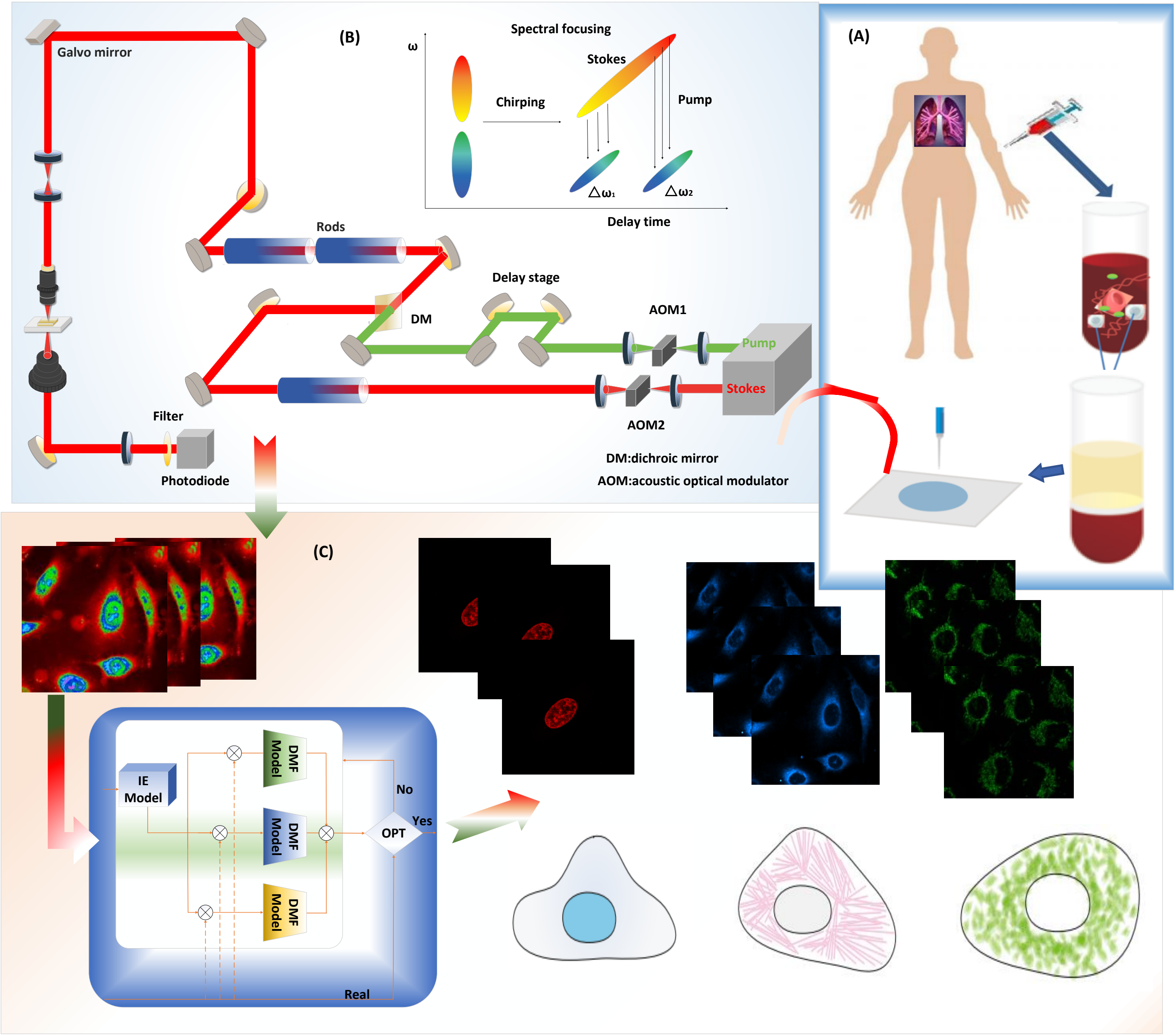


Figure 11. Workflow of the Single-Cell experiment by Stimulated Raman Scattering Imaging and deep learning model prediction process. (**A**) The process of prepearing the lung cancer cell sample. (**B**) Stimulated Raman Scattering microscopy setup for collecting SRS signal of the lung cancer cell samples. (**C**) Different machine learning techniques for the subcellular protein localization of lung cancer cells.

The deep learning-based computer-aided method for detecting subcellular protein localization using a Stimulated Raman Scattering microscopic images framework consists of the following stages: Firstly, the cell sample is prepared. Later, The SRS signal and fluorescence signal of different lung cancer cell samples were collected simultaneously using Stimulated Raman Scattering microscopy. Finally, the subcellular protein localization of lung cancer cells is performed using different machine learning techniques.

Specifically, lung cancer cells (A549, from ATCC, USA) were first cultured in ATCC F-12K medium (ATCC, USA). Then, the cells were fixed using 2% paraformaldehyde after being dyed. For the prepared live cells, after installing the living cell samples, the prepared cells were imaged with Stimulated Raman Scattering microscopy. After that, fluorescence images of nuclei, mitochondria, and endoplasmic reticulum were detected with fluorescent dyes, including Hoescht 33342, MitoTracker Red CMXRos, and ER-Tracker Green of different colors. In this work, a Unet algorithm was utilized to significantly improve the signal-to-noise ratio of nonlinear optical images. The Unet architecture consists mainly of convolutional layers that take an input feature map and convolves it with a set of filters to produce an output feature map. Based on the minimum loss from the optimizer on the training set, we chose the best hyperparameters and denoised the cell imaging test set. After denoising and enhancing the collected image, the 1200 processed cell sampling images were divided into two subsets, 80% of which was used for training, and the deep learning algorithm based on different algorithms was used to train the model. The remaining 20% of the subset was used as a test set to validate the model.

4.2. Protein Subcellular Localization Based on Multiple Parallel Fusion Deep Networks

The bottleneck in predicting subcellular protein locations of SRS cell imaging lies in modeling complicated relationships concealed beneath the original cell imaging data owing to the spectral overlap information from different protein molecules. Concerning the above issue, a multiple parallel fusion (MPF) Deep Network for Protein Subcellular Localization from Label-free live cell imaging is proposed to overcome the crowded and highly convoluted information, as shown in Figure 12. The main processes are as follows:

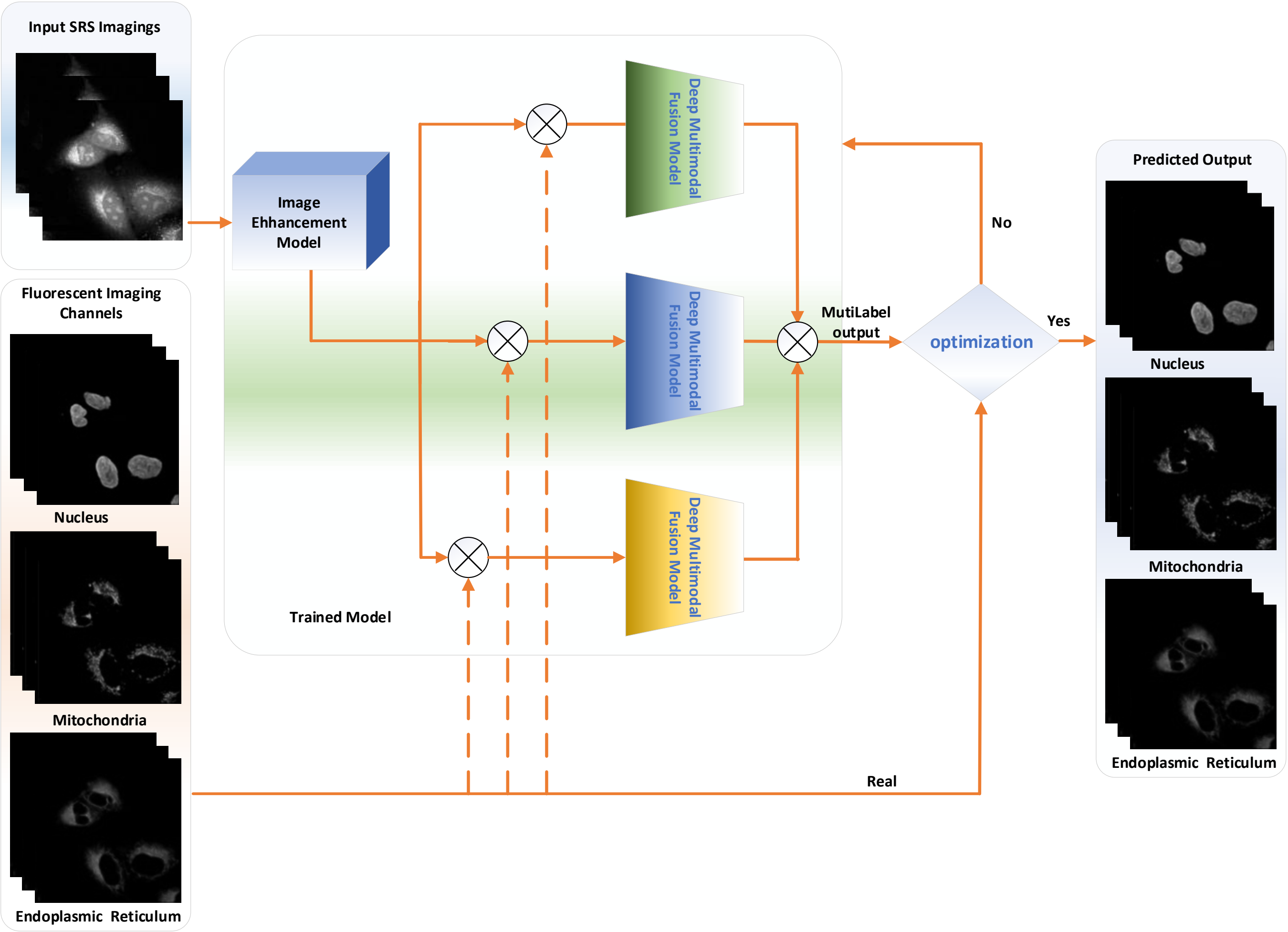


Figure 12. Graphical overview of our proposed deep MPF architecture. It includes five Modules: Input SRS imagings, Fluorescent Imaging Channels, Trained Model, Optimization and Predicted Output.

Step 1. According to the lung cancer cell imaging experiment, establish and store a lung cancer cells SRS imaging data set.

Step 2. Preprocess lung cancer cell SRS raw data sequences for the deep-learning-enabled image denoising and restoration [50–52].

Step 3. Build an integrated multiple parallel fusion deep networks framework.

Step 3.1. Construct independent transformer and CNN fusion models corresponding to different protein subcellular sites and fluorescence imaging labels.

Step 3.2. Train multiple parallel fusion models. Evaluate the protein subcellular location prediction performance according to the quantified metrics.

Step 4. Apply the different cell data sets to optimize the model parameters and find the optimal model combination.

Step 5. Locate the subcellular protein sites by using new cell data.

4.3. Multiple Parallel Fusion Neural Network Architecture and Implementation

As shown in Figure 13, the multiple parallel fusion network consists of mainly four components, which include two parallel branches and one multiple parallel fusion model in order to process input cell imaging information differently: (1) CNN branch, which gradually increases the receptive field and encodes the feature from local to global; (2) transformer branch, which starts with global self-attention and recovers the local details at the end; (3) multiple parallel fusion module, where fused features of the same resolution are extracted from each branch; and (4) gated skip-connection, where it combined the multi-level fused feature maps and generates prediction results. Each of the components are introduced in the following sections.

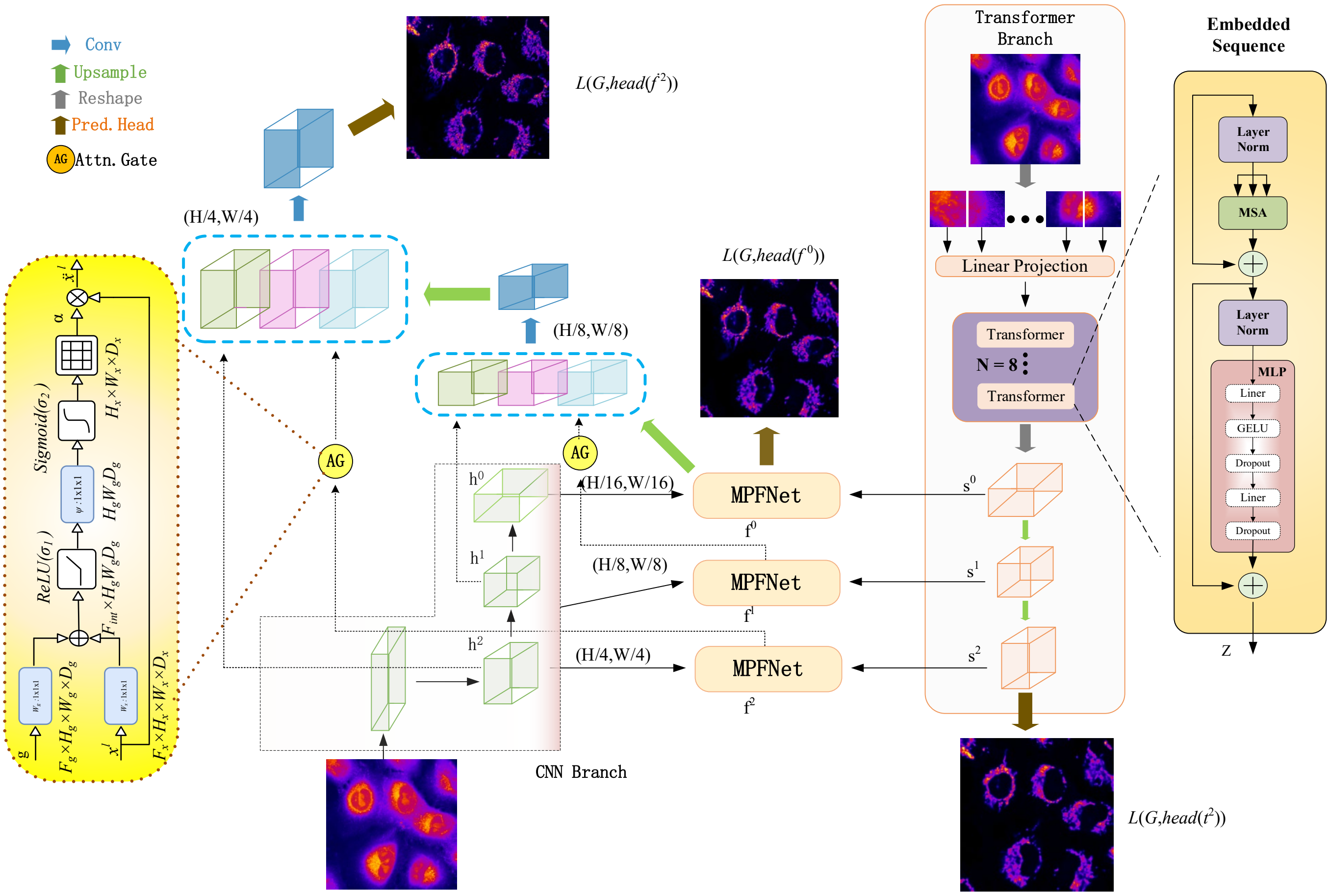


Figure 13. Multiple Parallel Fusion Deep Networks for the Protein Subcellular Localization from Label-free live cell imaging.

4.3.1. CNN Branch

As shown in Figure 13, the CNN branch adopts hierarchical feature extraction, in which the resolution of the feature map decreases with the deepening of the network while the channel number increases. We divided the whole CNN branch into four stages. Each stage is composed of multiple residual convolution blocks. Each residual convolution block contains multiple convolution layers, specification layers, and activation layers. In the first stage, we increased the number of channels of the image without reducing the resolution of the image. The next three stages are feature extraction, resolution reduction to one-half of the previous layer, and increasing the number of channels to enrich the feature information.

In experiments, CNN branches were used to obtain shallow features and retain richer original information by increasing the number of channels. In order to further obtain the global context information, the transformer branch was also employed for further accurate detection and location prediction while retaining the original information. Further, the characteristic maps of each level of the CNN branch and the characteristic map of the characteristic pyramid structure generated by the transformer are simultaneously input into the MPFnet model, which includes a CBAM block, channel attention block, spatial attention block, and convolution block for multi-feature information fusion.

It is worth noting that the CNN branch is used in the process of down sampling and combined into the MPFnet model for multi-feature information fusion with transformer output to enrich the original details. Moreover, in the up-sampling process, the high-resolution characteristic images at all levels of the encoder CNN are jump connected and added to the decoder path to obtain more detailed information and more accurate location.

4.3.2. Transformer Branch

Because of CNN's lack of translation invariance and global representation capture, and transformers good long-distance dependence capturing between sequences, we added transformer branches to the encoder. As shown in Figure 13, the transformer branch also follows the encoder-decoder structure in general. In the encoder part, the input image x $\in \mathrm{R}^{\mathrm{H}\times\mathrm{W}\times\mathrm{C}}$ with C channels is directly divided into a grid of $N = \frac{H}{M} \times \frac{W}{M}$ patches. Then, each patch is mapped into a one-dimensional vector through linear mapping $f:\ p \to e \in R^{C}$, and the H × W dimensions are flattened to form the input feature map to a linear embedding layer with output dimension S_0, obtaining the raw embedding sequence $e \in R^{N\times S_0}$.

$$X_0 = [E_{x_1}, E_{x_1}, \ldots, E_{x_N}] \tag{15}$$

To capture positional information, learnable position coding pos = $[\mathrm{ps}_1, \ldots, \mathrm{ps}_\mathrm{N}] \in \mathrm{R}^{\mathrm{N}\times\mathrm{S}_0}$ is added at the same time to the sequence of patches to obtain the resulting input sequence of tokens.

$$z_0 = x_0 + pos \tag{16}$$

The encoder in the transformer module with L = 8 transformer layer modules is applied to the sequence of tokens z_0 to generate a sequence of contextualized encoding $z_L \in R^{N\times S_0}$, and each transformer layer module includes multi-head self-attention (MSA), multi-layer perceptron (MLP) block, layer Norm block, Gelu layer, and dropout layer. Among them, the self-attention layer is the core module, which converts picture pixels into timing sequences in order to capture long-term dependence and retain global information. In the self-attention block, it first maps each input sequence $z_L \in R^{N\times S_0}$ to Q, K, and V, and then calculates the correlation between Q and K as the weight of the subsequent corresponding V.

$$q = z_l Wq,\ k = z_l Wk, v = z_l Wv\ \ W_q, W_k,\ W_v \in R^{S\times d} \tag{17}$$

where W_Q, W_K, $\mathrm{W}_\mathrm{V} \in R^{S\times d}$ are the learnable parameters of three linear projection layers, and d is the dimension of (query, key, value).

Self-attention (*SA*) is then formulated as:

$$SA(z_i) = softmax\left(\frac{q_i k^T}{\sqrt{D_h}}\right)v \tag{18}$$

$$SA(Z_{l-1}) = Z_{l-1} + \mathit{softmax}\left(\frac{Z_{l-1}W_Q(ZW_K)^\top}{\sqrt{d}}\right)(Z_{l-1}W_V). \tag{19}$$

MSA is an extension with *m* independent *SA* operations and projects their concatenated outputs:

$$MSA(Z_{l-1}) = [SA_1(Z_{l-1}); SA_2(Z_{l-1}); \cdots; SA_\mathrm{m}(Z_{l-1})]W_\mathrm{O} \tag{20}$$

where $\mathrm{W_O} \in \mathrm{R}^{\mathrm{md}\times\mathrm{C}}$. l is typically set to C/m.

The output of *MSA* is then transformed by an *MLP* block with a residual skip as the layer output as:

$$Z_l = MSA(LN(Z_{l-1})) + MLP(LN(MSA(Z_{l-1}))) \in \mathbb{R}^{L\times C} \tag{21}$$

where $\{Z_1, Z_2, \ldots, Z_L\}$ is the features of transformer layers, *LN*(.) is the layer norm applied before *MSA* and *MLP* blocks.

The contextualized encoding sequence Z_l contains rich semantic information used by the Decoder. For the decoder part of this work, the three steps of the progressive upsampling method are used. In the first step, the encoder feature Z is reshaped back from a 2D shape of $\frac{H\times W}{16\times 16} \times C_0$ to a 2D feature map of $s^0 \in R^{\frac{H}{8}\times\frac{W}{8}\times C_0}$. In the second step, the spatial resolution is more recovered from the 2D feature map of $s^0 \in R^{\frac{H}{8}\times\frac{W}{8}\times C_0}$ to $s^1 \in R^{\frac{H}{8}\times\frac{W}{8}\times D_1}$. In the third step, the upsampling-convolution layer is consecutively utilized to reshaped back from $s^1 \in R^{\frac{H}{8}\times\frac{W}{8}\times D_1}$ to $s^2 \in R^{\frac{H}{8}\times\frac{W}{8}\times D_2}$. At last, all feature maps with different scales will be sent to multiple parallel fusion modules for coupling with the output feature of the CNN branch, respectively.

4.3.3. Multiple Parallel Fusion Module

Considering the feature misalignment between CNN and transformer features, the multiple parallel fusion model (MPF) is designed as the bridge, which is applied to effectively combine the encoded features from different branches (Figure 14). Since CNN and transformer branches tend to capture features of different levels ranging from the local to global scale, MPF modules are inserted into every block to consecutively eliminate the semantic divergence between them in an interactive fashion. We take the outputs from the fourth ($\mathrm{h}^0 \in \mathrm{R}^{\frac{\mathrm{H}}{16}\times\frac{\mathrm{W}}{16}\times\mathrm{C}_0}$), third ($\mathrm{h}^1 \in \mathrm{R}^{\frac{\mathrm{H}}{8}\times\frac{\mathrm{W}}{8}\times\mathrm{C}_1}$), and second ($\mathrm{h}^2 \in \mathrm{R}^{\frac{\mathrm{H}}{4}\times\frac{\mathrm{W}}{4}\times\mathrm{C}_1}$) blocks to fuse with the results from transformer. In sum, the fused feature representation $\boldsymbol{f}^i$, i = 0, 1, 2 can be obtained by:

$$\widehat{\boldsymbol{s}}^i = ChannelAttn\left(\boldsymbol{s}^i\right) \tag{22}$$

$$\widehat{\boldsymbol{h}}^i = SpatialAttn\left(\boldsymbol{h}^i\right) \tag{23}$$

$$\widehat{\boldsymbol{b}}^i = Conv\left(\boldsymbol{h}^i\boldsymbol{W}_1^i \odot \boldsymbol{h}^i\boldsymbol{W}_2^i\right) \tag{24}$$

$$\boldsymbol{f}^i = Residual\left(\left[\widehat{\boldsymbol{b}}^i, \widehat{\boldsymbol{s}}^i, \widehat{\boldsymbol{h}}^i\right]\right) \tag{25}$$

$$\widehat{\boldsymbol{f}}^{i+1} = Conv\left(\left[Up\left(\boldsymbol{f}^i\right), AG\left(\boldsymbol{f}^{i+1}\right)\right]\right)\ for\ i = 0, 1, \ldots \tag{26}$$

where $W_1^i \in R^{D_i\times L_i}$, $W_2^i \in R^{C_i\times L_i}$, $| \odot |$ is the element-wise dot product and *Conv* is a 3 × 3 convolution layer.

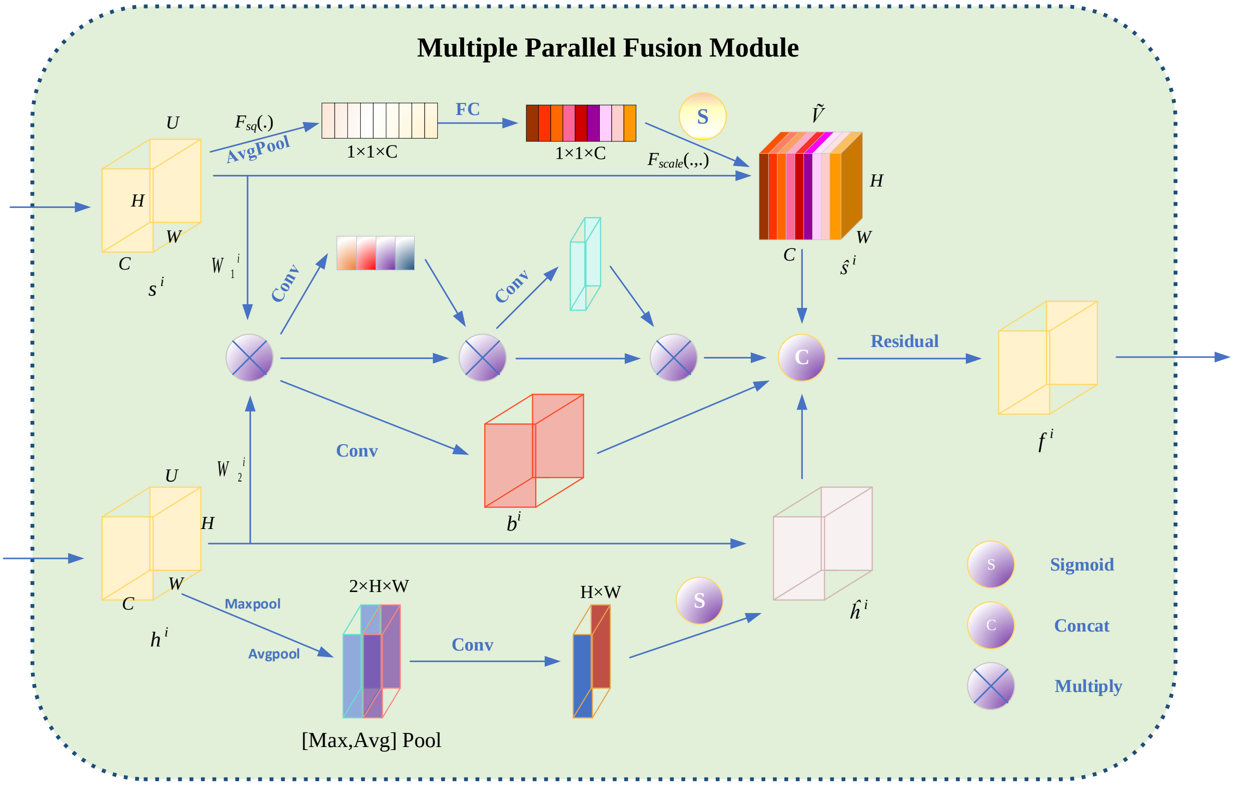


Figure 14. Multiple Parallel Fusion Model for the Protein Subcellular Localization from Label-free live cell imaging.

In the Multiple Parallel Fusion Model, the high-level CNN features and low-level transformer features were further fused with different attention mechanisms, including spatial attention block, channel-wise attention block, and residual block, as discussed in detail below.

4.3.4. Spatial Attention Block

In cell images, the information importance of different positions of the subcellular image is also different. For example, the edge position information of subcellular is generally more important than that from other positions. Consequently, spatial attention is imperative to strengthen such important information. Compared with channel-wise attention, spatial attention pays more attention to the content information in the spatial position. Therefore, the goals of spatial attention lie in uncovering lateral areas of cell images, which provide the greatest contribution to the final high accuracy subcellular prediction and assign these areas higher weights. By distributing the weight in each spatial position, we find which spatial position information is most important and consequently enhance the characteristics of that part of the position, meanwhile inhibiting the extraction of noise features.

The scheme of the spatial attention module is shown in Figure 15, which is applied to identify valuable cell regions for subcellular recognition. Firstly, the original feature maps of $F' \in R^{C\times H\times W}$ are aggregated by a pair of average pooling and maximum pooling operations, which are formulated as:

$$F_{avg}(h,w) = \frac{1}{c_v}\sum_{c=1}^{c_v} F(c,h,w), h = 1,2,\ldots,h_v; w = 1,2,\ldots,w_v \tag{27}$$

$$F_{max}(h,w) = max(c,h,w), h = 1,2,\ldots,h_v; w = 1,2,\ldots,w_v \tag{28}$$

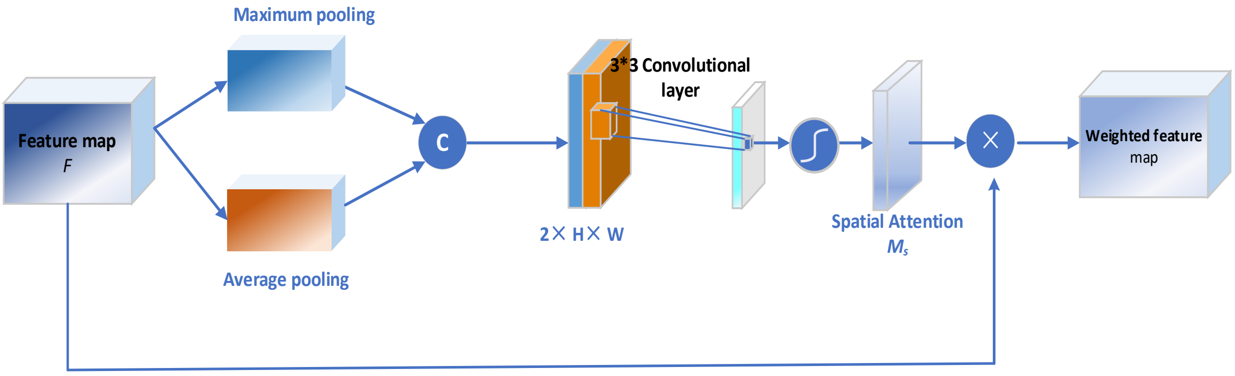


Figure 15. Diagram of spatial attention sub-module. As illustrated, the spatial sub-module utilizes max-pooling outputs and average-pooling outputs that are pooled along the channel axis and forwarded to a convolution layer.

Each outputs a weight matrix of all spatial positions. Then, $F_{avg}(h,w)$ and $F_{max}(h,w)$ are concatenated and convoluted by the pyramid kernels with size 3 × 3, 3 × 3, 3 × 3, respectively, to generate the 2D spatial attention map after the sigmoid active function. The process can be defined as:

$$F_{cat} = f^{(2\times k-1)\times(2\times k-1)}\left(\left[F^s_{avg}; F^s_{max}\right]\right) \tag{29}$$

$$F_{spa} = \sum_{k=1}^{3} \sigma\left(f^{(2\times k-1)\times(2\times k-1)}\left(\left[F^s_{avg}; F^s_{max}\right]\right)\right) \tag{30}$$

where $f^{(2\times k-1)\times(2\times k-1)}(.)$ is a convolution operation with the filter size of $(2\times k-1)\times(2\times k-1)$.

The elements of F_{spa} represent the importance of the corresponding regions of the spatial domain. Subsequently, the weighted feature map F'' is obtained by multiplying the feature map F' with the spatial attention map:

$$F'' = F_{spa} \otimes F' \tag{31}$$

where F'' denotes the final output 3D feature tensor of the convolutional attention module.

4.3.5. Channel-Wise Attention

Note that the variety of texture information in the feature map of cell imaging still requires preliminarily removing useless redundant information to keep important texture features before weight calculation. Hence, a channel-wise attention squeeze-excitation module is introduced in this section to amend the texture and global context features from the input feature maps and give them higher weights. A diagram illustrating the structure of a squeeze-excitation block is shown in Figure 16.

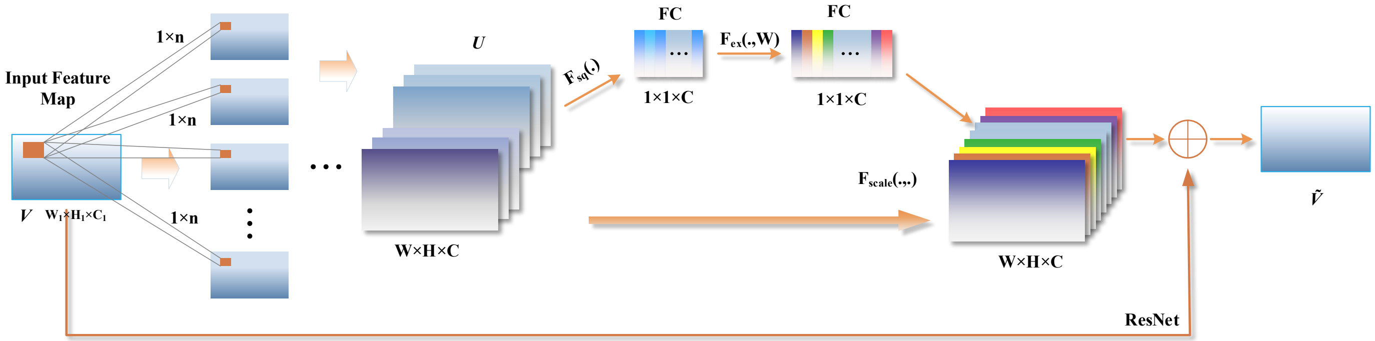


Figure 16. A structure scheme of the Channel-wise Attention based on squeeze-excitation block.

Given input sample V to a channel-wise attention block, a set of learned convolutional filters are applied on V to produce corresponding feature matrix responses $U \in R^{H \times W \times C}$, where $H \times W$ is the spatial dimension, and C is the channel dimension. We take F_{tr}: $V \rightarrow U$, $V \in R^{H' \times W' \times C'}$, $U \in R^{H \times W \times C}$ to be a convolutional operator.

$$U = F_{tr}(V) \tag{32}$$

$$u_c = k_c * V = \sum_{l=1}^{C'} K_C^l * V^l \tag{33}$$

where $*$ denotes convolution, $k_c = \left[k_c^1, k_c^2, \ldots, k_c^{C'}\right]$ and $X = \left[x^1, x^2, \ldots, x^{C'}\right]$ and $u_c \in R^{H \times W}$, while k_c^l is a 2D spatial kernel and therefore represents a single channel of k_c, which acts on the corresponding channel of V.

Then, a squeeze operation and an excitation operation are applied on U sequentially to re-weight channel-wise feature responses. In order to squeeze the global spatial information into a channel descriptor, channel-wise global average pooling was utilized to squeeze global spatial information into a channel descriptor. Formally, a channel-wise statistic $z \in R^C$ is generated by shrinking U through spatial dimensions $H \times W$, where the c-th element of z is calculated by:

$$Z_C = F_{sq}(u_c) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} u_c(i, j) \tag{34}$$

The channel-wise output Z_C of the information aggregated in the squeeze operation is next used to modulate nonlinear inter-dependencies of all channels through an excitation operation. Here, a gating mechanism with a sigmoid activation is employed as follows:

$$s = F_{ex}(z, W) = \sigma(g(z, w)) = \sigma(W_2 \delta(W_1 z)) \tag{35}$$

where σ is the sigmoid function, m is a scaling parameter. δ refers to the ReLU function, $W_1 \in R^{\frac{C}{m} \times C}$ and $W_2 \in R^{C \times \frac{C}{m}}$.

The final output of the block is obtained by rescaling the transformation output U with the activations:

$$\widetilde{V_c} = F_{scale}(u_c, s_c) = s_c \cdot u_c \tag{36}$$

where $\check{V}_c = \left[v_1, v_2, \ldots, v_C\right]$ and $F_{scale}(u_c, s_c)$ refers to channel-wise multiplication between the feature map $u_c \in R^{H \times W}$ and the scalar s_c.

4.3.6. Attention Gate Block

In the semantic segmentation for cell imaging based on transformation with CNN networks, the depth of the convolution layer is usually increased to expand the acceptance domain and capture more semantic context information. However, at the same time, it is still difficult to reduce false positive predictions, especially for small objects with large shape changes. In order to avoid this dilemma, the attention gate (AG) block is incorporated in the MPFnet, as shown in Figure 12, to highlight the significant information of the down-sampling output. The AG block uses contextual information from the gating signal $g \in R^{F_g}$ to prune the skip connection s_i^l, highlighting ROIs and therefore reducing false positive predictions. Through this module, we retain the spatial positioning information of some important objects as much as possible while maintaining a large receptive field.

The structure of an attention gate is shown in Figure 17. This AG receives two inputs, the gating signal $g \in R^{F_g}$ and the associated skip connection s_i^l generated at that level. The gating vector signal g_i is used for each pixel i to determine the regions of focus that originate from the deepest layer of the neural network, where feature representation is the greatest at the cost of significant down-sampling. The gating vector contains contextual information to reduce lower-level feature responses.

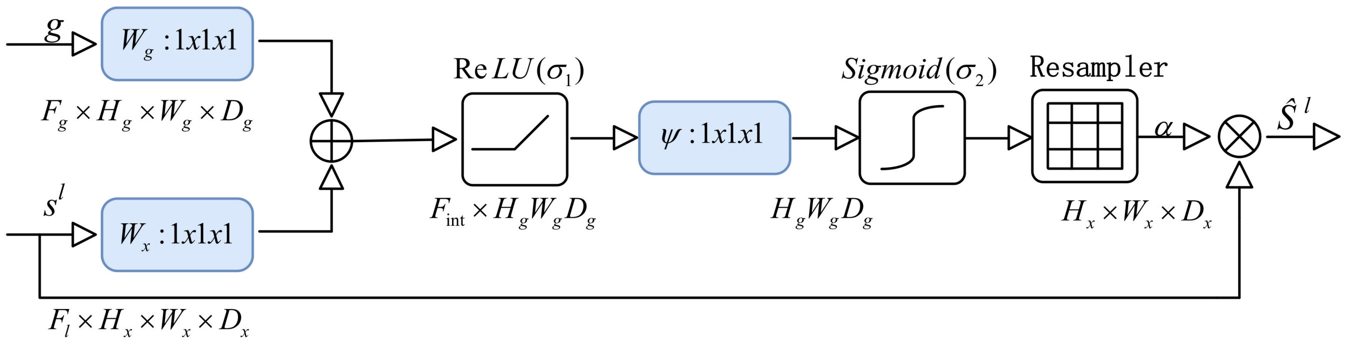

g
$W_g : 1x1x1$
$F_g \times H_g \times W_g \times D_g$
s^l
$W_x : 1x1x1$
$F_l \times H_x \times W_x \times D_x$
$\mathrm{Re}LU(\sigma_1)$
$F_{\text{int}} \times H_g W_g D_g$
$\psi : 1x1x1$
$Sigmoid(\sigma_2)$
$H_g W_g D_g$
Resampler
α
$H_x \times W_x \times D_x$
$\hat{S}^l$

Figure 17. The structure of the attention gate block to highlight the significant information of the down-sampling output.

For the AG gate block in Figure 17, the gating signal and skip connection s_i^l are first resized and then combined to form attention coefficients α^l calculated by:

$$q^l_{\text{att},\,i} = \psi^T\left(\sigma_1\left(W_x^T s_i^l + W_g^T g + b_{\text{xg}}\right)\right) + b_\psi \tag{37}$$

$$\alpha^l = \sigma_2\left(q^l_{\text{att}}\left(s^l, g; \Theta_{att}\right)\right) \tag{38}$$

Then, the output $\widehat{s}^l$ acquired with the element-wise multiplication between the original skip connection s^l and the attention coefficients α^l provides spatial context highlighting ROIs, whose specific formula is as follows:

$$\widehat{s}^l = s^l \cdot \alpha^l \tag{39}$$

Attention coefficient, $\alpha_i \in [0,\ 1]$ emphasizes salient image regions and significant features to preserve only relevant activations specific to the actual task.

4.4. Dataset

We employed a subset of SRS images in the fixed lung cancer cells (A549, from ATCC) dataset as one of our pre-trained data sources. These data sets were acquired simultaneously using image software by collecting the SRS signals from lock-in amplifiers and fluorescence signals from photomultiplier tubes [36]. For the fluorescence signals, all dyeing schemes are based on the standards provided that three different color fluorescent dyes were used to label and track the nucleus, mitochondria, and endoplasmic reticulum, respectively. The optical cell images with 512 × 512 pixels were obtained at a dwell time of 4 µs.

Another trained source of data we employed is the dataset cell images, which are acquired using GE's IN Cell Analyzer systems [53]. These datasets were applied to test different deep learning methods in the work and evaluate their performance.

5. Conclusions

In this work, we have presented a methodology based on the MPF network, a novel multiple parallel fusion deep learning model for protein subcellular location prediction from SRS images. The MPFnet includes four modules: CNN branch, receptive field block, Multiple Parallel Fusion Module, transformer branch and Attention gate module. The multiple parallel fusion module is composed of three core components: spatial attention block, channel-wise attention block, and residual block. The fused multi-level encoded features from the CNN branch focus on local information, and the Transformer branch focuses on global information to eliminate the semantic divergence between them. The performance of the proposed multiple parallel fusion method was estimated and compared with other deep learning models such as UwUnet and Unet. It is shown from the experimental results that the new multiple parallel fusion deep network achieves top prediction performance and is faster compared with other deep learning methods while reducing the number of parameters, suggesting that it has great potential in the subcellular prediction of label-free cell optical images. In future work, we will further develop more advanced fusion methods to hybridize the Transform and Unet methods to further improve the performance of protein subcellular location in cell optical imaging.

Author Contributions: Q.G. conceived the prediction method and wrote the paper. Z.W. developed the computer programs. Q.G. analyzed the results. Q.G., W.L., R.Y., X.L., W.Y. and Q.L. revised the paper. All authors have read and agreed to the published version of the manuscript.

Funding: The work is supported by the fund of the Beijing Municipal Education Commission, China, under grant number 22019821001, and the fund of the Climbing Program Foundation from the Beijing Institute of Petrochemical Technology (Project No. BIPTAAI-2021007).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data that support the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Parlakgül, G.; Arruda, A.P.; Pang, S.; Erika, C.; Nina, M.; Ekin, G.; Grace, Y.L.; Karen, I.; Hess, H.F.; Shan, X.C.; et al. Regulation of liver subcellular architecture controls metabolic homeostasis. *Nature* **2022**, *603*, 736–742. [CrossRef] [PubMed]
2. Mottis, A.; Herzig, S.; Auwerx, J. Mitocellular communication: Shaping health and disease. *Science* **2019**, *366*, 827–832. [CrossRef] [PubMed]
3. Yuan, H.; Cai, L.; Wang, Z.Y.; Hu, X.; Zhang, S.T.; Ji, S.W. Computational Modeling of Cellular Structures Using Conditional Deep Generative Networks. *Bioinformatics* **2019**, *35*, 2141–2149. [CrossRef] [PubMed]
4. Koenig, F.; Knittel, J.; Stepp, H. Diagnosing cancer in vivo. *Science* **2001**, *292*, 401–1403. [CrossRef]
5. Szabo, V.; Ventalon, C.; De Sars, V.; Bradley, J.; Emiliani, V. Spatially selective holographic photoactivation and functional fluorescence imaging in freely behaving mice with a fiberscope. *Neuron* **2014**, *84*, 1157–1169. [CrossRef]
6. Chou, K.C.; Shen, H.B. Recent progress in protein subcellular location prediction. *Anal. Biochem.* **2007**, *370*, 1–16. [CrossRef]
7. Xu, Y.Y.; Yang, F.; Zhang, Y.; Shen, H.B. An image-based multi-label human protein subcellular localization predictor (iLocator) reveals protein mislocalizations in cancer tissues. *Bioinformatics* **2013**, *29*, 2032–2040. [CrossRef]
8. Hung, M.C.; Link, W. Protein localization in disease and therapy. *J. Cell Sci.* **2011**, *124*, 3381–3392. [CrossRef]
9. Guo, X.; Liu, F.; Ju, Y.; Wang, Z.; Wang, C. Human Protein Subcellular Localization with Integrated Source and Multi-label Ensemble Classifier. *Sci. Rep.* **2016**, *6*, 28087. [CrossRef]
10. Vicar, T.; Balvan, J.; Jaros, J.; Jug, F.; Kolar, R.; Masarik, M.; Gumulec, J. Cell segmentation methods for label-free contrast microscopy: Review and comprehensive comparison. *BMC Bioinform.* **2019**, *20*, 360. [CrossRef]
11. Zhang, L.; Wu, Y.; Zheng, B.; Su, L.; Chen, Y.; Ma, S.; Hu, Q.; Zou, X.; Yao, L.; Yang, Y.; et al. Rapid histology of laryngeal squamous cell carcinoma with deep-learning based stimulated Raman scattering microscopy. *Theranostics* **2019**, *9*, 2541–2554. [CrossRef] [PubMed]
12. Ounkomol, C.; Seshamani, S.; Maleckar, M.M.; Collman, F.; Johnson, G.R. Label-free prediction of three- dimensional fluorescence images from transmitted-light microscopy. *Nat. Methods* **2018**, *15*, 917–920. [CrossRef] [PubMed]
13. Chou, K.C.; Shen, H.B. Cell-PLoc 2.0: An improved package of web-servers for predicting subcellular localization of proteins in various organisms. *Nat. Sci.* **2010**, *2*, 1090–1103. [CrossRef]
14. Chou, K.C.; Cai, Y.D. Prediction of protein subcellular locations by GO-FunD-PseAA predicor. *Biochem. Biophys. Res. Commun.* **2004**, *320*, 1236–1239. [CrossRef] [PubMed]
15. Falk, T.; Mai, D.; Bensch, R.; Çiçek, Ö.; Abdulkadir, A.; Marrakchi, Y.; Böhm, A.; Deubner, J.; Jäckel, Z.; Seiwald, K.; et al. U-Net: Deep learning for cell counting, detection, and morphometry. *Nat. Methods* **2019**, *16*, 67–70. [CrossRef]
16. Wang, A.; Zhang, Q.; Han, Y.; Megason, S.; Hormoz, S.; Mosaliganti, K.R.; Lam, J.C.K.; Li, V.O.K. A novel deep learning-based 3D cell segmentation framework for future image-based disease detection. *Sci. Rep.* **2022**, *12*, 342. [CrossRef]
17. Wei, L.; Ding, Y.; Ran, S.; Tang, J.; Zou, Q. Prediction of human protein subcellular localization using deep learning. *J. Parallel Distrib. Comput.* **2018**, *117*, 212–217. [CrossRef]
18. Jing, S.; Attila, T.; Su, X.T. Deep Learning-Based Single-Cell Optical Image Studies. *Cytom. A* **2020**, *97*, 226–240.
19. Buggenthin, F.; Buettner, F.; Hoppe, F.; Endele, P.S.; Kroiss, M.; Strasser, M.; Schwarzfischer, M.; Loeffler, M.; Kokkaliaris, D.; Hilsenbeck, K.D.; et al. Prospective identification of hematopoietic lineage choice by deep learning. *Nat. Methods* **2017**, *14*, 403–406. [CrossRef]
20. Hasan, M.; Ahmad, S.; Molla, M. Protein Subcellular Localization Prediction Using Multiple Kernel Learning Based Support Vector Machine. *Mol. BioSyst.* **2017**, *13*, 785–795. [CrossRef]

21. Juan, J.; Armenteros, A.; Sønderby, C.K.; Sønderby, S.K.; Nielsen, H.; Winther, O. DeepLoc: Prediction of protein subcellular localization using deep learning. *Bioinformatics* **2017**, *33*, 3387–3395.
22. Wang, Y.; Xu, Y.; Zang, Z.; Wu, L.; Li, Z. Panoramic Manifold Projection (Panoramap) for Single-Cell Data Dimensionality Reduction and Visualization. *Int. J. Mol. Sci.* **2022**, *23*, 7775. [CrossRef] [PubMed]
23. Zhang, X.; Roeffaers, M.; Basu, S.; Daniele, J.; Fu, D.; Freudiger, C.; Holtom, G.; Xie, X. Label-Free Live-Cell Imaging of Nucleic Acids Using Stimulated Raman Scattering Microscopy. *ChemPhysChem* **2012**, *13*, 1054–1059. [CrossRef] [PubMed]
24. Jiang, Y.; Lei, C.; Yasumoto, A.; Kobayashi, H.; Aisaka, Y.; Ito, T.; Guo, B.; Nitta, N.; Kutsuna, N.; Ozeki, Y.; et al. Label-free detection of aggregated platelets in blood by machine-learning-aided optofluidic time-stretch microscopy. *Lab Chip* **2017**, *17*, 2426–2434. [CrossRef]
25. Kobayashi, H.; Lei, C.; Wu, Y.; Mao, A.; Jiang, Y.; Guo, B.; Ozeki, Y.; Goda, K. Label-free detection of cellular drug responses by high-throughput bright-field imaging and machine learning. *Sci. Rep.* **2017**, *7*, 12454. [CrossRef]
26. Wei, H.; Xie, L.; Liu, Q.; Shao, C.; Wang, X.; Su, X. Automatic Classification of Label-Free Cells from Small Cell Lung Cancer and Poorly Differentiated Lung Adenocarcinoma with 2D Light Scattering Static Cytometry and Machine Learning. *Cytom. Part A* **2019**, *95A*, 302–308. [CrossRef]
27. Li, J.T.; Chen, J.; Bai, H.; Wang, H.W.; Hao, S.P.; Ding, Y.; Peng, B.; Zhang, J.; Li, L.; Huang, W. An Overview of Organs-on-Chips Based on Deep Learning. *Research* **2022**, *2022*, 9869518. [CrossRef]
28. Siu, D.M.D.; Lee, K.C.M.; Lo, M.C.K.; Stassen, S.V.; Wang, M.L.; Zhang, I.Z.Q.; So, H.K.H.; Chan, G.C.F.; Cheah, K.S.E.; Wong, K.K.Y.; et al. Deep-learning-assisted biophysical imaging cytometry at massive throughput delineates cell population heterogeneity. *Lab Chip* **2020**, *20*, 3696–3708. [CrossRef]
29. Yao, K.; Rochman, N.D.; Sun, S.X. Cell Type Classification and Unsupervised Morphological Phenotyping From Low-Resolution Images Using Deep Learning. *Sci. Rep.* **2019**, *9*, 13467. [CrossRef]
30. Chen, C.; Mahjoubfar, A.; Tai, L.C.; Blaby, I.K.; Huang, A.; Niazi, K.R.; Jalali, B. Deep Learning in Label-free Cell Classification. *Sci. Rep.* **2016**, *6*, 21471. [CrossRef]
31. Wang, F.S.; Wei, L.Y. Multi-scale deep learning for the imbalanced multi-label protein subcellular localization prediction based on immunohistochemistry images. *Bioinformatics* **2022**, *38*, 2602–2611. [CrossRef] [PubMed]
32. Zhang, J.; Zhao, J.; Lin, H.; Tan, Y.; Cheng, J.X. High-Speed Chemical Imaging by Dense-Net Learning of Femtosecond Stimulated Raman Scattering. *J. Phys. Chem. Lett.* **2020**, *11*, 8573–8578. [CrossRef] [PubMed]
33. Imboden, S.; Liu, X.; Lee, B.S.; Payne, M.C.; Hsieh, C.J.; Lin, N.Y. Investigating heterogeneities of live mesenchymal stromal cells using AI-based label-free imaging. *Sci. Rep.* **2021**, *11*, 6728. [CrossRef] [PubMed]
34. Lynch, A.W.; Theodoris, C.V.; Long, H.W.; Brown, M.; Liu, X.S.; Meyer, C.A. MIRA: Joint regulatory modeling of multimodal expression and chromatin accessibility in single cells. *Nat Methods* **2022**, *19*, 1097–1108. [CrossRef] [PubMed]
35. Melanthota, S.K.; Gopal, D.; Chakrabarti, S.; Kashyap, A.A.; Radhakrishnan, R.; Mazumder, N. Deep learning-based image processing in optical microscopy. *Biophys. Rev.* **2022**, *14*, 463–481. [CrossRef] [PubMed]
36. Manifold, B.; Men, S.; Hu, R.; Fu, D. A versatile deep learning architecture for classification and label-free prediction of hyperspectral images. *Nat. Mach. Intell.* **2021**, *3*, 306–315. [CrossRef]
37. Kobayashi, H.; Cheveralls, K.C.; Leonetti, M.D.; Royer, L.A. Self-supervised deep learning encodes high-resolution features of protein subcellular localization. *Nat. Methods* **2022**, *19*, 995–1003. [CrossRef]
38. Donovan-Maiye, R.M.; Brown, J.M.; Chan, C.K.; Ding, L.; Yan, C.; Gaudreault, N.; Theriot, J.A.; Maleckar, M.M.; Knijnenburg, T.A.; Johnson, G.R. A deep generative model of 3D single-cell organization. *PLoS Comput. Biol.* **2018**, *18*, e1009155. [CrossRef]
39. Gomariz, A.; Portenier, T.; Helbling, P.M.; Isringhausen, S.; Suessbier, U.; Nombela-Arrieta, C.; Goksel, O. Modality attention and sampling enables deep learning with heterogeneous marker combinations in fluorescence microscopy. *Nat. Mach. Intell.* **2021**, *3*, 799–811. [CrossRef]
40. Chen, X.; Li, Y.; Wyman, N.; Zhang, Z.; Fan, H.; Le, M.; Gannon, S.; Rose, C.; Zhang, Z.; Mercuri, J.; et al. Deep Learning Provides High Accuracy in Automated Chondrocyte Viability Assessment in Articular Cartilage Using Nonlinear Optical Microscopy. *Biomed. Opt. Express* **2021**, *12*, 2759–2772. [CrossRef]
41. Lu, A.X.; Kraus, O.Z.; Cooper, S.; Moses, A.M. Learning unsupervised feature representations for single cell microscopy images with paired cell inpainting. *PLoS Comput. Biol.* **2019**, *15*, e1007348. [CrossRef] [PubMed]
42. Wang, X.; Liu, J.; Zhang, C.; Wang, S. SSGraphCPI: A Novel Model for Predicting Compound-Protein Interactions Based on Deep Learning. *Int. J. Mol. Sci.* **2022**, *23*, 3780. [CrossRef]
43. Voronin, D.V.; Kozlova, A.A.; Verkhovskii, R.A.; Ermakov, A.V.; Makarkin, M.A.; Inozemtseva, O.A.; Bratashov, D.N. Detection of Rare Objects by Flow Cytometry: Imaging, Cell Sorting, and Deep Learning Approaches. *Int. J. Mol. Sci.* **2020**, *21*, 2323. [CrossRef] [PubMed]
44. Calizo, R.C.; Bell, M.K.; Ron, A.; Hu, M.; Bhattacharya, S.; Wong, N.J.; Janssen, W.G.M.; Perumal, G.; Pederson, P.; Scarlata, S.; et al. Cell shape regulates subcellular organelle location to control early Ca^{2+} signal dynamics in vascular smooth muscle cells. *Sci. Rep.* **2020**, *10*, 17866. [CrossRef] [PubMed]
45. Chen, J.; Lu, Y.; Yu, Q.; Luo, X.; Adeli, E.; Wang, Y.; Lu, L.; Yuille, A.L.; Zhou, Y. MPFnet: Transformers Make Strong Encoders for Medical Image Segmentation. *arXiv* **2021**, arXiv:2102.04306.
46. Fang, J.; Yang, C.; Shi, Y.T.; Wang, N.; Zhao, Y. External Attention Based MPFnet and Label Expansion Strategy for Crack Detection. *IEEE Trans. Intell. Transp. Syst.* **2022**, *17*, 1–10.

47. Rezatofighi, H.; Tsoi, N.; Gwak, J.Y.; Sadeghian, A.; Reid, I.; Savarese, S. Generalized Intersection over Union: A Metric and A Loss for Bounding Box Regression. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2019, Long Beach, CA, USA, 15–20 June 2019; pp. 658–666.
48. Boslaugh, S.; Watters, P.A. *Statistics in a Nutshell: A Desktop Quick Reference*; O'Reilly Media: Sebastopol, CA, USA, 2008; ISBN 9780596510497.
49. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [CrossRef] [PubMed]
50. Esakkirajan, S.; Veerakumar, T.; Subramanyam, A.N.; PremChand, C.H. Removal of High Density Salt and Pepper Noise Through Modified Decision Based Unsymmetric Trimmed Median Filter. *IEEE Signal Process. Lett.* **2011**, *18*, 287–290. [CrossRef]
51. Hsieh, M.h.; Cheng, F.C.; Shie, M.C.; Ruan, S.J. Fast and efficient median filter for removing 1–99% levels of salt-and-pepper noise in images. *Eng. Appl. Artif. Intell.* **2013**, *26*, 1333–1338. [CrossRef]
52. Manifold, B.; Thomas, E.; Francis, A.T.; Hill, A.H.; Fu, D. Denoising of stimulated Raman scattering microscopy images via deep learning. *Biomed. Opt. Express* **2019**, *10*, 3860–3874. [CrossRef]
53. Al-Kofahi, Y.; Zaltsman, A.; Graves, R.; Marshall, W.; Rusu, M. A deep learning-based algorithm for 2-D cell segmentation in microscopy images. *BMC Bioinform.* **2018**, *19*, 365. [CrossRef] [PubMed]