Article

Symmetry Breaking in the U-Net: Hybrid Deep-Learning Multi-Class Segmentation of HeLa Cells in Reflected Light Microscopy Images

Ali Ghaznavi [1,*] **, Renata Rychtáriková** [1] **, Petr Císař** [1] **, Mohammad Mehdi Ziaei** [1,2] **and Dalibor Štys** [1]

1. South Bohemian Research Center of Aquaculture and Biodiversity of Hydrocenoses, Faculty of Fisheries and Protection of Waters, Institute of Complex Systems, University of South Bohemia in České Budějovice, Zámek 136, 37333 Nové Hrady, Czech Republic; rrychtarikova@frov.jcu.cz (R.R.); cisar@frov.jcu.cz (P.C.); ziaei@frov.jcu.cz (M.M.Z.); stys@frov.jcu.cz (D.Š.)
2. Faculty of Science, University of South Bohemia, Branišovská 1760, 37005 České Budějovice, Czech Republic
* Correspondence: ghaznavi@frov.jcu.cz

Abstract: Multi-class segmentation of unlabelled living cells in time-lapse light microscopy images is challenging due to the temporal behaviour and changes in cell life cycles and the complexity of these images. The deep-learning-based methods achieved promising outcomes and remarkable success in single- and multi-class medical and microscopy image segmentation. The main objective of this study is to develop a hybrid deep-learning-based categorical segmentation and classification method for living HeLa cells in reflected light microscopy images. A symmetric simple U-Net and three asymmetric hybrid convolution neural networks—VGG19-U-Net, Inception-U-Net, and ResNet34-U-Net—were proposed and mutually compared to find the most suitable architecture for multi-class segmentation of our datasets. The inception module in the Inception-U-Net contained kernels with different sizes within the same layer to extract all feature descriptors. The series of residual blocks with the skip connections in each ResNet34-U-Net's level alleviated the gradient vanishing problem and improved the generalisation ability. The m-IoU scores of multi-class segmentation for our datasets reached 0.7062, 0.7178, 0.7907, and 0.8067 for the simple U-Net, VGG19-U-Net, Inception-U-Net, and ResNet34-U-Net, respectively. For each class and the mean value across all classes, the most accurate multi-class semantic segmentation was achieved using the ResNet34-U-Net architecture (evaluated as the m-IoU and Dice metrics).

Keywords: categorical segmentation; neural network; cell detection; microscopy image segmentation; U-Net; tissue segmentation; semantic segmentation; bright-field microscopy cell segmentation; cell analysis

1. Introduction

Cell detection and segmentation are fundamental processes in microscopy cell image analysis. These are challenging tasks due to the complexity of these images. On the other hand, the information from the segmented living cells can play an essential role in further analysis, such as observing and estimating cell behaviour, their number, and their dimensions. Recently developed artificial intelligence (AI) methods have achieved promising outcomes in this field. The machine learning (ML) segmentation methods for cell analysis can be categorised as traditional machine learning or recently developed deep learning (DL) methods.

1.1. Cell Culture Segmentation with Traditional Machine Learning Methods

The number of traditional cell detection–segmentation ML methods has grown rapidly because of the low performance of simple techniques, such as threshold-based [1], region-based [2], or morphological approaches [3,4] when processing such complex images. The traditional ML methods can be further classified as supervised or unsupervised.

The supervised methods use training data to generate a mathematical function or a model to map a new data sample [5]. Trained and optimised parameters using the graph-based Supervised Normalized Cut Segmentation (SNCS) with loosely annotated images separate overlapping and curved cells better than the traditional image processing methods [6]. Mah et al. [7] proposed a classification method using Fast Random Forest (FRF) and Trainable WEKA Segmentation for extracting the Interstitial cells of Cajal networks in 3D confocal microscopy images. The proposed method represents better performance than the Decision Table and Naïve Bayes classification methods in terms of accuracy and F-measure metric. However, the method showed higher computational costs due to the FRF's structure. A method combining the Support Vector Machine (SVM) and the Histogram of Oriented Gradients extracted and classified the feature descriptors as cells or non-cells in bright-field microscopy data. The method was sensitive to the training iterations, which is a crucial step in eliminating false positive detections [8]. A Logistic Regression classification with intensity values of 25 focal planes as features, followed by the binary erosion with a large circular structuring element, counted the cells in bright-field microscopy images. However, the method showed mis-segmentation and a low recall rate [9].

The training data for the unsupervised ML algorithms need not be labelled or scored a priori [10]. Unsupervised segmentation using the Markov Random Field considered an image as a series of planes based on Bit Plane Slicing. The planes were used as initial labelling for an ensemble of segmentations. The robust cell segmentation was achieved with pixel-wise voting. However, this method was too sensitive to the confidence threshold [11]. A combination of a Scale-Invariant Feature Transform, a self-labelling, and two clustering methods segmented unstained cells in bright-field micrographs. The method was fast and accurate but sensitive to the feature selection to avoid overfitting [12]. A self-supervised (i.e., a kind of unsupervised) learning approach combined unsupervised initial coarse segmentation (K-means clustering) followed by supervised segmentation refinement (SVM pixel classifier) to separate white blood cells. However, the unsupervised part of the method generates a rough segmentation result. In the case of complex datasets, the supervised part of the method cannot work efficiently due to fuzzy boundaries [13].

1.2. Cell Culture Segmentation with Deep Learning Methods

In recent years, a subset of new machine learning techniques—deep learning (DL) methods—has been developed to solve cell segmentation problems with higher accuracy and performance. The deep neural networks have integrated low-/medium-/high-level features and classifiers into a comprehensive multi-layer structure. The depth of the network, or the number of layers stacked, determines the "levels" of features [14].

Mask RCNN with a Shape-Aware Loss generated the HeLa cell's segmentation masks with a good performance [15]. A Convolutional Blur Attention (CBA) network for nuclei segmentation in standard datasets [16,17] with an acceptable aggregated Jaccard index consisted of down- and up-sampling procedures. The reduced number of trainable parameters reasonably decreased the computational cost [18]. The input images of a convolutional network can be of different custom sizes so that they can be trained end-to-end and pixel-to-pixel to produce an output of the appropriate size. Effective inference and learning can achieve successful semantic segmentation in complex microscopic and medical images [19,20].

A U-Net architecture containing a contracting path to obtain context and a symmetric expanding path for precise localisation showed strong data augmentation in the training process. It was optimised when applied to small datasets and performed efficiently in semantic segmentation of photon microscopy (phase contrast and DIC) images [21]. A Feedback U-Net with the convolutional Long Short-Term Memory network, working on the *Drosophila* cell image dataset and mouse cell image dataset, generally showed a low level of accuracy, depending on the segmented class (cytoplasm, cell membrane, mitochondria, synapses) [22]. A Residual Attention U-Net-based method segmented living HeLa cells in bright-field light microscopy data with a high IoU metric. The method combined the self-attention mechanism (to highlight the remarkable features and suppress activations

in the irrelevant image regions) and the residual mechanism (to overcome the vanishing gradient problem [23]. Multi-class cell segmentation in fluorescence images combining U-Net (a deeper network) with ResNet34 (a residual mechanism) achieved a good value of IoU score [24]. A two-step U-Net method segmented HeLa cells in microscopy images. The first U-Net localised the position of each cell. The second U-Net was trained with the first U-Net to determine the cell boundaries [25]. A fully automated U-Net-based algorithm recognised different classes (colonies, single, differentiated, and dead) of human pluripotent stem cells from each other with a satisfying m-IoU value in phase contrast images [26].

1.3. Our Motivation for a New Image Segmentation Method

In segmentation, especially of tiny cells, the traditional ML methods struggle with microscopy images with complex backgrounds [7,8]. The traditional ML methods have also not been very efficient in training the multi-class segmentation models in large time-lapse image series. Compared with the traditional ML methods, some Convolution Neural Networks (CNNs) architectures require many manually labelled training datasets and higher computational costs [19]. Deep learning methods have shown better results in segmentation tasks than other methods.

The main goal of our research is to develop and compare variants of a fully convolutional network as the encoder part of the original U-Net architecture and find the most accurate categorical segmentation algorithm. The U-Net was chosen since it is one of the most promising methods for semantic segmentation [21]. Later, the encoder part of the U-Net architecture was modified and replaced with a VGG-19, Inception, and ResNet34 encoder architecture and was examined to find the most suitable architecture for multi-class segmentation. We used unique telecentric bright-field reflected light microscopy multi-class labelled images of the cells to be automatically classified according to their morphological shapes to predict their cell cycle phases.

We captured image series of HeLa cells to test the algorithms. The HeLa is a cell line of human Negroid cervical epithelioid carcinoma that is used in tissue culture laboratories as the gold standard. Each image contains HeLa cells in different cell cycle states. The raw microscopy data are specific for their high pixel resolution in rgb mode and require pre-processing steps to reduce optical vignetting and camera noise. The data show unlabelled in-focus and out-of-focus living cells in their physiological state.

2. Materials and Methods

2.1. Cell Preparation and Microscope Specification

The cells were prepared as written in [23], Section 2.1. The European Collection of Cell Cultures with Cat. No. 93021013 of the human HeLa cell line was selected and prepared for time-lapse experiments. The cells were cultivated overnight with low optical density conditions at 37 °C, 5% CO_2, and 90% relative humidity. The nutrient solution includes Dulbecco's modified Eagle medium (87.7%) with high glucose (>1 g L^{-1}), fetal bovine serum (10%), antibiotics and antimycotics (1%), L-glutamine (1%), and gentamicin (0.3%; all provided by Biowest, Nuaille, France). The HeLa cells were maintained in a Petri dish with a cover glass bottom and lid at a room temperature of 37 °C.

Several time-lapse image series experiments on living HeLa cells growing on a glass Petri dish were collected using a high-resolved reflected light microscope with the light source and the microscope objective located on the same side when the light refracted or emitted from the specimen is analysed, giving the bright image with a dark background. This microscope was designed by the Institute of Complex Systems (ICS, Nové Hrady, Czech Republic) and was built by Optax (Prague, Czech Republic) and ImageCode (Brloh, Czech Republic) in 2021. The microscope has a simple construction of the optical path. The sample is illuminated by a Schott VisiLED S80-25 LED Brightfield Ringlight. The light reflected from a sample goes through a telecentric measurement objective TO4.5/43.4-48-F-WN (Vision & Control GmbH, Shul, Germany) to an Arducam AR1820HS 1/2.3-inch 10-bit RGB camera with a chip of 4912 × 3684 pixel resolution. The software (developed by the

ICS) controls the capture of the primary signal (raw image with a theoretical pixel size of 113 nm) with a camera exposure of 998 ms.

2.2. Data Preparation and Pre-Processing

Several time-lapse experiments were completed with HeLa cells using a reflected bright-field microscope (Section 2.1). The microscope control software calibrated the microscope optical path and corrected all image series using the algorithm proposed in [27] to avoid image background inhomogeneities and noise.

The calibration step was followed by converting the raw image representations to 8-bit colour (rgb) images of a quarter number of pixels [28] in order to preserve the information maximally and ensure mutual comparability of the images through the time-lapse series. The green channel on a typical camera sensor has a larger transparency, and its intensities dominate the signal (Figure 1). The background noise in converted 8-bit rgb images was minimised at preserving the texture details [29]. Afterwards, different time-lapse series were cropped to the 1024×1024 pixel size, giving the main dataset with 650 images (accessible at [30]).
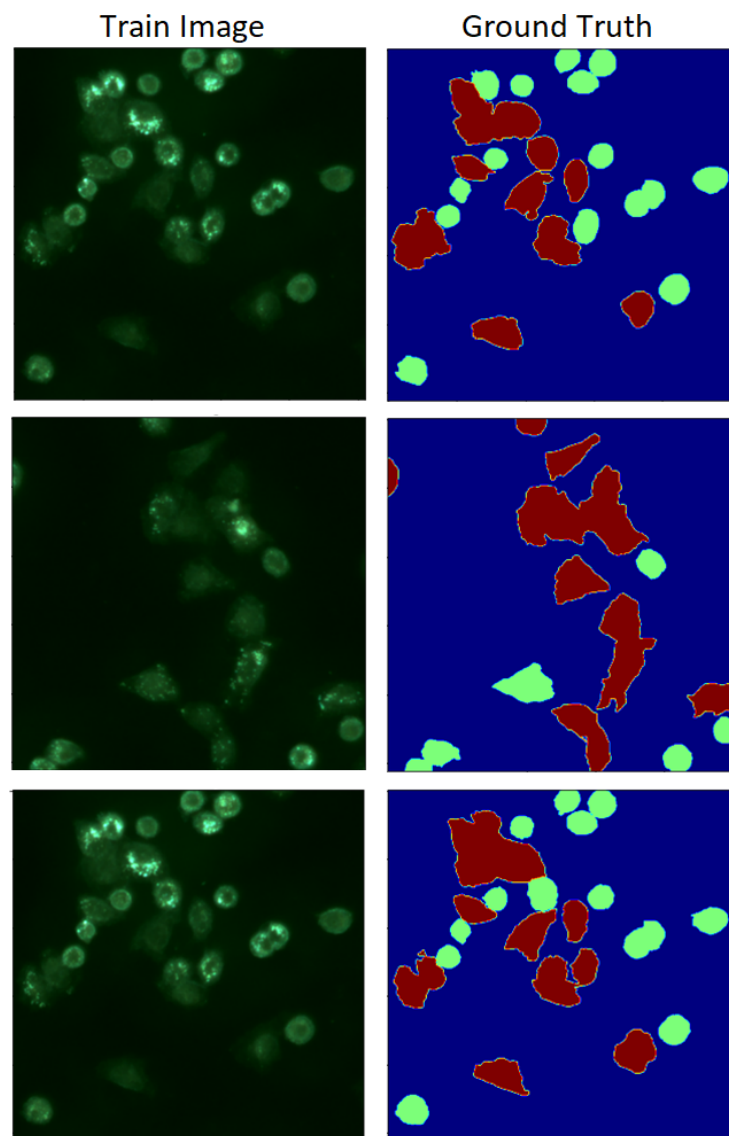
Figure 1. Examples of the train sets and corresponding ground truths. The image size is 512×512. The images in the left column visualise primary data from the camera sensor where, without any white balancing, the green intensity channel dominates (see Section 2.2). The green and red classes in the right column represent the roundish sharp cells and the migrating unclear cells, respectively.

For multi-class segmentation, one of three cell states was assigned to each cell manually using Apeer platform [31]: (1) a background class containing no cells, (2) a cell class containing larger dilated adhered or migrating cells with unclear borders by which we anticipate they are growing, and (3) a cell class including roundish cells with sharper borders when the cells are assumed in their early stage of the life cycle, having no division state yet, or at the beginning of the division. Identifying the proportion of cells in mitosis holds significance across various biomedical endeavours, including biological research and medical diagnosis [32]. Figure 1 depicts a sample of the resized dataset and relevant generated mask classes as ground truth of the size of 512×512 pixels. The manually segmented images were part of training (80%), testing (20%), and evaluation (20% of the training set) sets in the proposed neural network architectures.

2.3. The Neural Network Model Architectures

2.3.1. U-Net

The U-Net [21] is well-known as a deep neural network for semantic image segmentation. The U-Net architecture is based on encoder–decoder layers. The U-Net combines many shallow and deep feature channels. In this research, a five-"level" simple U-Net was implemented as the first method for multi-class segmentation purposes. The extracted deep features served for object localisation, whereas the shallow features were used for precise segmentation.

The first input layer accepts rgb 512×512-sized training set images. Each level of the proposed U-Net contains two 3×3 convolutions. Batch normalisation follows each convolution, and "ReLU" is used as an activation function. Each encoder "level" in the down-sampling (encoder) part (Figure 2A) consists of a 2×2 max-pooling operation with a stride of two. The max-pooling process obtains the highest value within the 2×2 region. The convolutions lead to double the number of feature channels by completing the down-sampling in each level of the encoder section.
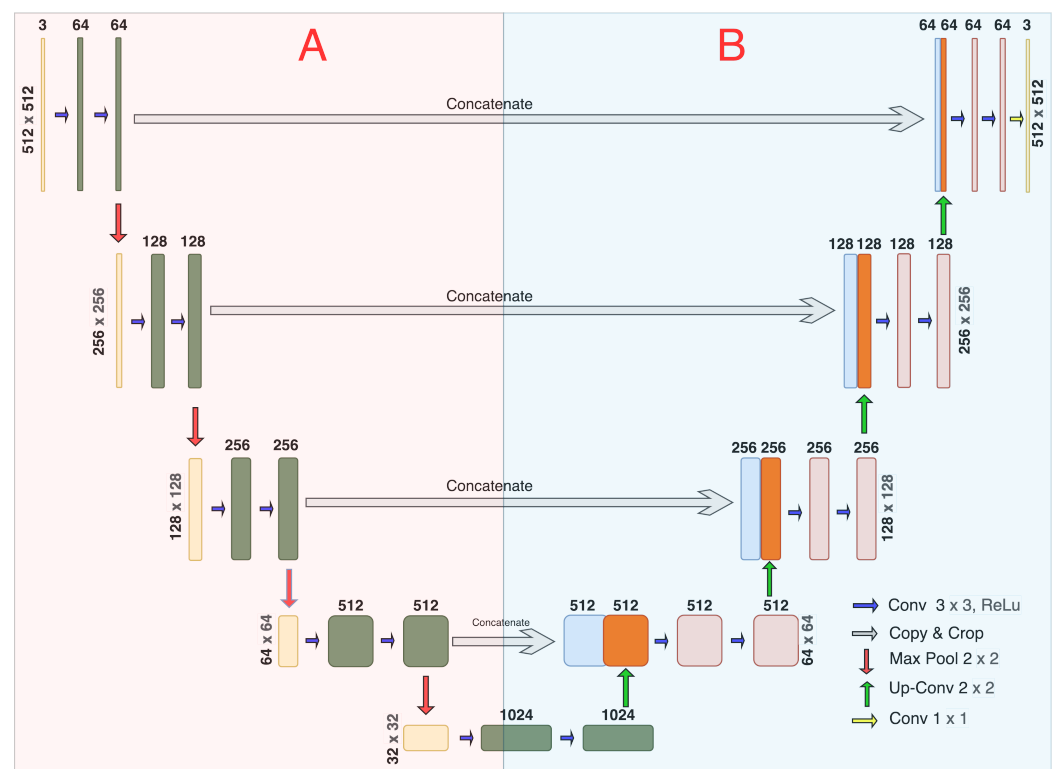
Figure 2. The simple U-Net model architecture. (**A**) The encoder section. (**B**) The decoder section.

In each level (from bottom to top) within the up-sampling (decoder) part (Figure 2B), the dimensions of the feature maps were multiplied by two in both height and width.

In the concatenation step, the encoder section's feature maps were integrated with the high-resolved shallow and deep semantic features. After concatenation, the channel sizes of the output feature maps are double the dimensions of the input feature maps. The "softmax" activation function in the top, 1×1 convolution-sized output decoder layer predicts the occurrence of each pixel in each of the three classes. We obtained the same input and output layer sizes by utilising padding in the convolution process. Each of those classes, achieved by the softmax activation, represents the probability of belonging each pixel into each class. In the final step, the "argmax" operation assigned each pixel to the class, where the highest probability value was achieved. This computational result, combined with the Categorical Focal Loss function, generated the energy function of the proposed U-Net architecture.

2.3.2. The VGG19-U-Net

Many modified artificial neural networks, such as AlexNet [33], ZFNet [14], and VGG [34], have been developed as hybrids with the U-Net to simplify U-Net. In this study, a VGG-Net architecture replaced the U-Net encoder path. In this way, we combined two powerful architectures to improve the categorical segmentation of our unique microscopy dataset. The VGG-Net was proposed by Simonyan and Zisserman [34] from Oxford's Visual Geometry Group (VGG). A VGG16 proved to be one of the most efficient classification networks. However, a VGG19 performed even more effectively than VGG16 [35]. The VGG19 comprises a network with a deeper topology and smaller convolution kernels to simulate a perceptual field of view. This architecture is designed to reduce the number of trainable parameters and decrease computational costs compared with the simple U-Net. Figure 3 represents the VGG19-U-Net proposed in this study. The left side of the network (Figure 3A) shows the architecture of the VGG19 encoder section with 16 convolution layers, 3 fully connected layers, and 5 MaxPool layers in 5 blocks. The convolution blocks at each level are followed by a 2×2 max-pooling operation with a stride of two to extract the maximal value in the 2×2 area. The proposed VGG19 network initiates with 64 channels in its first layer, and the channel numbers were doubled in each subsequent layer up to 512 channels. The right side of the network (Figure 3B) is a schema of the decoder part with five blocks. A concatenation step between each VGG19 encoder layer and each U-Net decoder layer (Figure 3) combines the feature maps from the encoder part with the high-resolution deep semantic and shallow features from the decoder part. The last decoder layer has a convolution size of 1×1 and predicts the probability values for each pixel and each of the three classes using the "softmax" activation function.
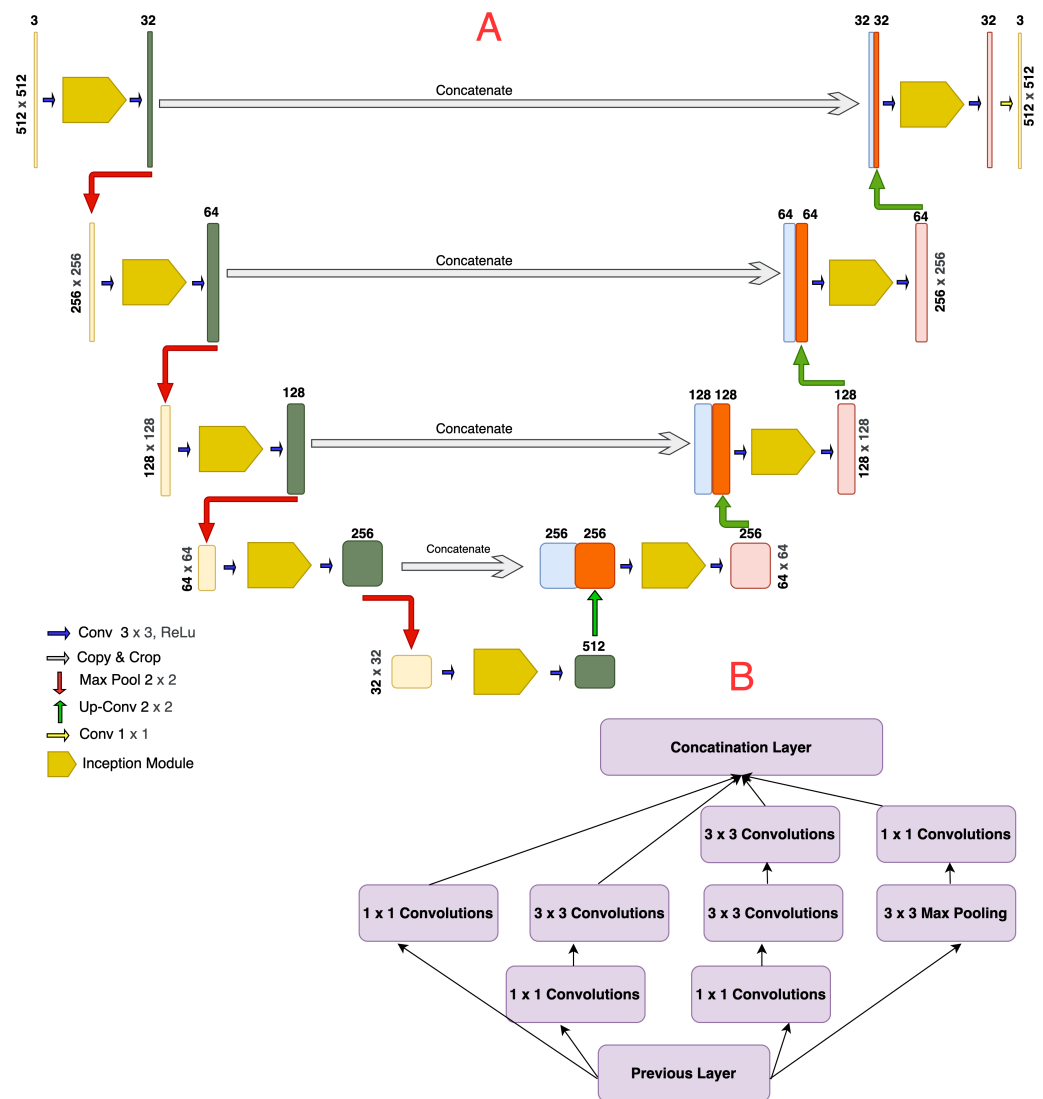
2.3.3. The Inception-U-Net

The complexity of the U-Net network about the number of trainable parameters leads to higher runtime and computational costs (Table 1). On the other hand, in image analysis, applying fixed kernel size in all convolution layers can make it difficult to extract all feature descriptors of different sizes. For example, in microscopy image analysis, some (tiny) features are at the local level, and some (larger) are at the global level. The network cannot extract the representative features for big objects when the small kernel is selected in convolution operations. If the kernel size is big, the network will miss extracting the features representative at the pixel level. In other words, the larger kernel can extract a global feature representation over a large image area, and the smaller kernel has been considered for detecting area-specific features. Google's inception deep learning method [36], known as the Inception architecture, was selected to build a hybrid Inception-U-Net architecture (Figure 4) to improve segmentation results in our datasets further.
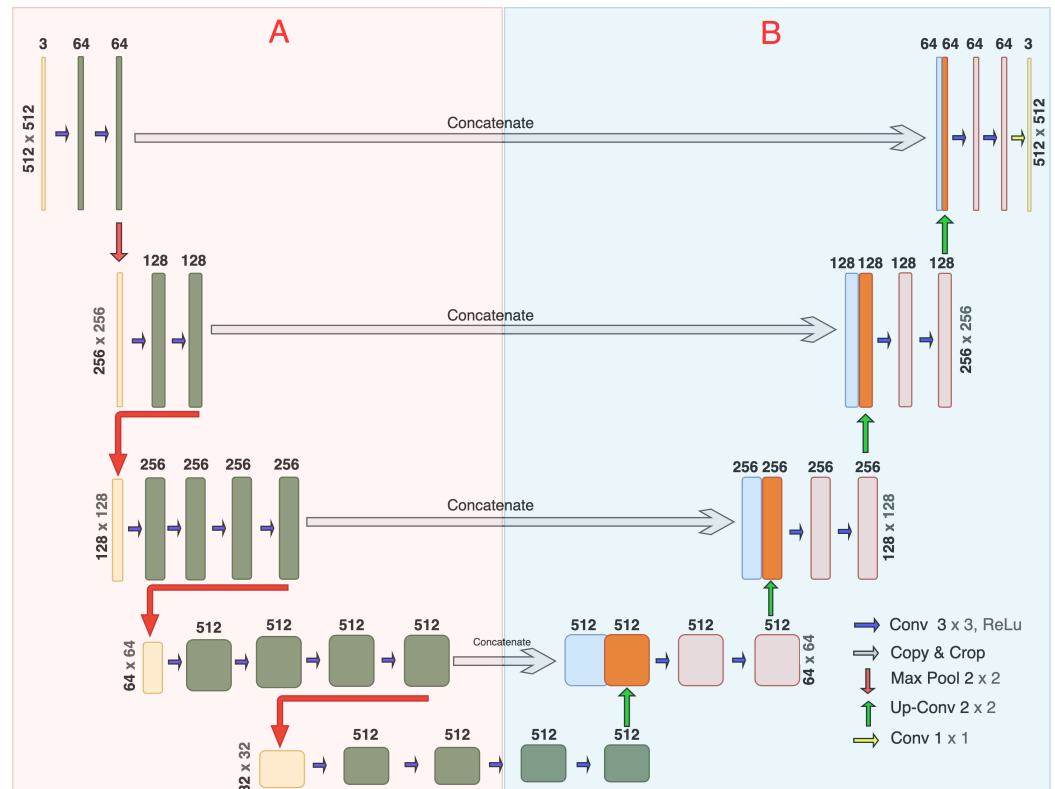
Figure 3. The hybrid VGG19-U-Net architecture. (**A**) The VGG19 encoder part. (**B**) The U-Net decoder part.

The inception module is well known for its computational efficiency by integrating different sizes of convolutions. The inception module applies kernels of different sizes within the same architecture layer and becomes wider (instead of deeper) with the layers (Figure 4B). The convolution layers were replaced with an inception module (Figure 4A) in all five levels of the encoder and decoder sections of the original U-Net structure. The inception module consists of different sizes of 3×3 convolutions, 1×1 convolutions, 3×3 max-pooling, and cascaded 3×3 convolutions. The number of filters at each convolution layer was doubled within the encoder side. The output feature map size (height and width) was reduced by half on the last encoder layer.

The up-sampling (decoder) architecture section (Figure 4A, left side) was also equipped with an inception module at each level. The skip connection linked the encoder and decoder section to enhance the performance of the prediction. The encoder spatial feature maps are concatenated with the decoder feature maps. The rectified linear unit (ReLU) was selected as an activation function for each layer to perform batch normalisation in each inception module. At the last layer, a 1×1 convolution layer together with the "softmax" activation function generated three segmentation classes of the feature maps for the given input image. Each pixel was assigned to one class according to the highest probability value achieved among the classes. The Categorical Focal Loss function has been considered an energy function for this Inception-U-Net.

Figure 4. (**A**) The Inception-U-Net architecture. (**B**) The internal architecture of one inception module.

2.3.4. The ResNet34-U-Net

To further improve the categorical segmentation of our datasets, the Residual Convolutional Neural Network (ResNet) [37] was joined to the U-net. Neural networks with deeper architecture are more effective for complex classification and segmentation tasks. However, during the training process, the vanishing gradient problem appears in the very deep CNN. Moreover, a high number of CNN layers makes the training process slower, and the calculated value of the backpropagation derivative becomes increasingly insignificant. Thus, the model's accuracy gets saturated and rapidly declines instead of improving. The series of residual blocks with the skip connections were implemented into the CNN to alleviate the gradient vanishing and improve the network's generalisation ability during the training process. The skip connections were added to the deep neural networks to bypass one or more layers and update the gradient values from one or more previous layers into the following layers.

The ResNet34-U-Net architecture used in our study (Figure 5) has 34 layers and 4 residual convolution steps with a total of 16 residual blocks (red and purple arrows). The first convolution layer has 64 filters with a kernel size of 7 × 7, followed by a max-pooling layer. Each residual block consists of two 3 × 3 convolution layers followed by the ReLU activation function and batch normalisation with the identity shortcut connection.
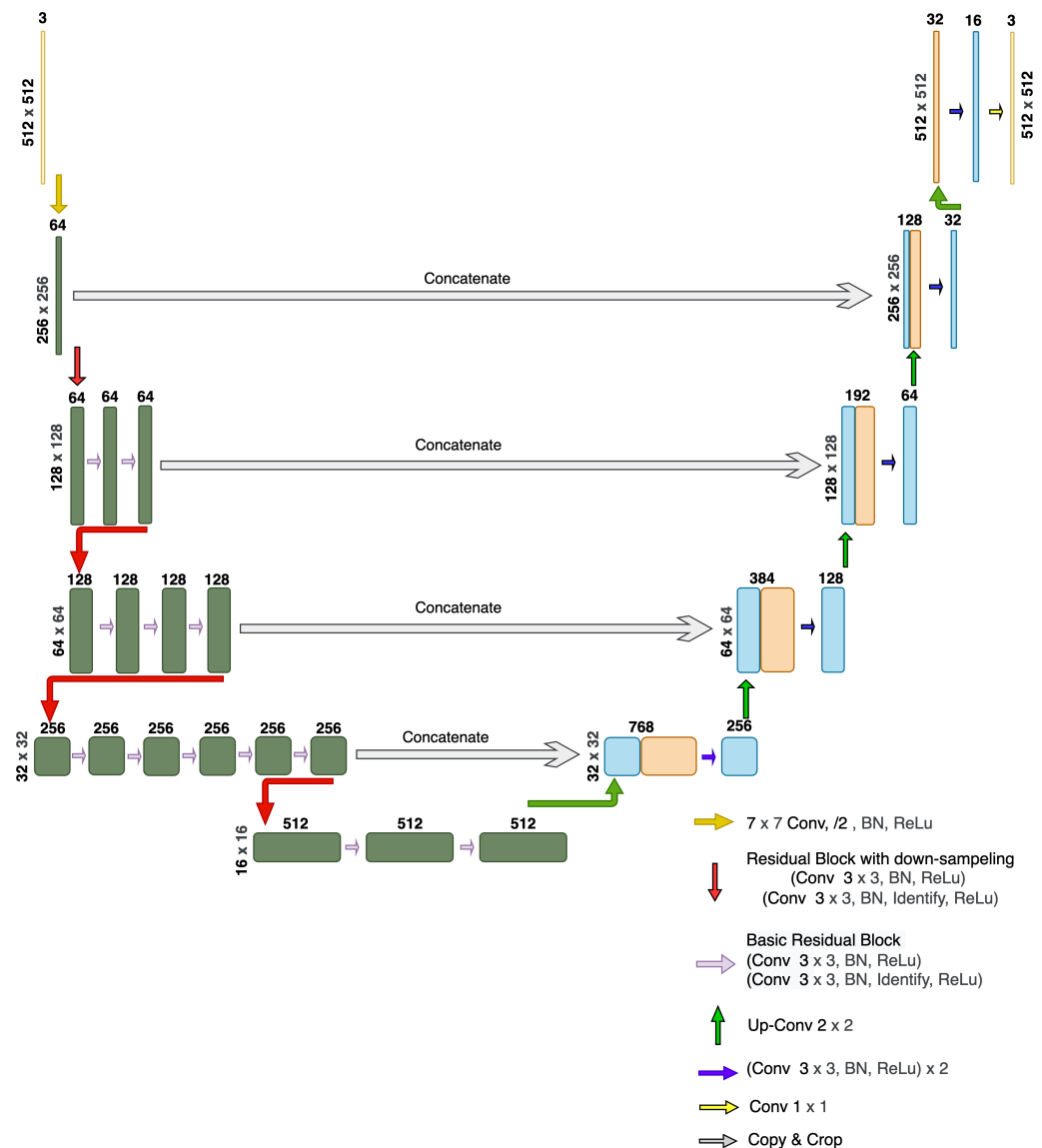
Figure 5. The hybrid ResNet34-U-Net architecture.

After the first 7×7 convolution layer, the feature map size halved to 256×256. At the first residual level, three residual convolution blocks were applied to the achieved feature maps, and the output size of the feature maps was halved to 128×128. Four residual convolution blocks in the second residual step decreased the size of the output feature maps to 64×64. Six residual convolution blocks in the third residual step gave a feature map size of 32×32. The last residual step consists of three residual convolution blocks to achieve a feature map with a size of 16×16.

The up-sampling section of the network (Figure 5) gets the input with the feature map size of 16×16 with 512 channels and a 2×2 up-convolution step with a stride of two. The decoder section has the same structure as the simple U-Net architecture. After passing the U-Net decoder part, the "softmax" activation function was employed to achieve the probability map across three different classes for each pixel of the input images. Afterwards, each pixel was assigned to a certain class according to the highest probability value selected by the "argmax" function.

With the usage of the ResNet34, the number of trainable parameters decreased significantly compared with the VGG19-Net and the simple U-Net. Thus, the runtime for training the model was shortened.

2.4. Training Models

The implementation platform for this research was based on Python 3.9. The deep learning framework was Keras with the Tensorflow backend [38]. All CNN architectures were first developed and completed on a personal computer and then transferred to the Google Colab Pro+ premium cluster account to train the most stable models. The Google Colab Pro+ cluster is equipped with an NVIDIA Tesla T4 or the NVIDIA Tesla P100 GPU with 16 GB of GPU VRAM, 52 GB of RAM, and two vCPUs [39].

The basic dataset included under-focused, over-focused, and focused images (650 images total) from various time-lapse series. Portions of the basic dataset were randomly selected to train the model (416 images, 64%) and validate the process (104 images, 16%) to avoid over-fitting. The rest 130 images (20%) were used to test and evaluate the model after training.

All images were normalised (see the pre-processing step in Section 2.2) and resized to 512×512 pixels suitable for inputting the designed neural networks. The optimised hyperparameter values (Table 2) correspond to training the most stable CNN models. The ReLU was selected as the activation function for all architecture. The early stopping hyperparameter was used to prevent overfitting during model training. A patient value was set at 30. The batch size was set to the maximal value of eight due to the complexity of the CNN structures and GPU-VRAM limitation. The Adam algorithm was chosen to optimise the neural networks. The learning rate was set to 10^{-3} for all proposed CNN models. The suitable number of object classes was set as 3 (Section 2.2). The best number-of-steps-per-epoch value equals 52 (achieved after dividing the length of the trainset of value 416 by the batch size of value 8). The number of epochs when all CNN models converged and were well-trained was 200.

Table 1. Number of the trainable parameters and the computational time for the U-Net models.

Network	Run Time	# Training Parameters
U-Net	$3:33':29''$	31,402,639
VGG19-U-Net	$1:44':38''$	31,172,163
Inception-U-Net	$1:05':47''$	18,083,535
ResNet34-U-Net	$0:56':22''$	24,456,444

Table 2. Hyperparameter settings for training all proposed models.

Hyperparameters Name	Value
Activation function	ReLU
Learning rate	10^{-3}
Number of classes	3
Batch size	8
Epochs number	200
Early stop	30
Step per epoch	52
γ for loss function	2

Categorical image segmentation entails classifying pixels into either cell classes or the background class. During training progress, all segmented cell images were compared to the GT to minimise the difference between these two as much as possible by using the Dice loss. One of the well-known loss functions used for categorical segmentation which is an extension of the cross entropy loss is the Categorical Focal Loss [40].

The Categorical Focal Loss is more efficient for the multi-class classification of imbalanced datasets, when some classes are determined easily, whereas others are not. During training progress, the loss function down-weights easy classes and focuses training on hard-to-classify classes. Thus, the focal loss reduces the loss value for "well-classified" examples (e.g., roundish sharp cells) and increases the loss for hard-to-classify objects (e.g., migrated

vanish cells) by tuning the right value of the focusing parameter γ in the categorical focal loss function. In summary, the categorical focal loss turns the model's attention towards the difficult-to-classify pixels to achieve more precise classification results.

2.5. Evaluation Metrics

The common evaluation metrics were used to assess all categorical semantic segmentation models (Equations (1)–(5)). The TP, FP, FN, and TN correspond to the true positive, false positive, false negative, and true negative metric, respectively, [41]. The metrics were calculated across all test sets within each class and reported as mean values across all classes (Tables 3 and 4).

Table 3. m-IoU values for the classes. C1—background, C2—divided and unclear cells, C3—roundish and sharp cells, green—the highest m-IoU value for the relevant class.

Network	m-IoU C1	m-IoU C2	m-IoU C3	m-IoU
U-Net	0.9894	0.4839	0.6452	0.7062
VGG19-Net	0.9885	0.5489	0.6160	0.7178
Inception-Net	0.9915	0.6614	0.7194	0.7907
ResNet 34-Net	0.9911	0.6911	0.7378	0.8067

Table 4. The metric results evaluating the U-Net models. The green values display the highest accuracy in segmentation for the corresponding metric.

Network	Accuracy	Precision	Recall	m-IoU	m-Dice
U-Net	0.9869	0.7897	0.8833	0.7062	0.8104
VGG19-U-Net	0.9865	0.8051	0.8614	0.7178	0.8218
Inception-U-Net	0.9904	0.8684	0.8905	0.7907	0.8762
ResNet 34-U-Net	0.9909	0.8795	0.8975	0.8067	0.8873

The overall pixel accuracy (Acc) indicates the percentage of image pixels correctly assigned to segmented cells:

$$\text{Acc} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{FN} + \text{TN}} \tag{1}$$

Precision (Pre) measures the ratio of correctly segmented cell pixels in the results that match the Ground Truth (GT). This metric is identified as a positive predictive value and holds significance in segmentation performance as it is sensitive to over-segmentation:

$$\text{Pre} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{2}$$

Recall (Recl) denotes the percentage of cell pixels in the GT identified correctly during the segmentation process. This metric represents the percentage of annotated objects in the GT that were identified as positive predictions:

$$\text{Recl} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{3}$$

The combination of Pre and Recl provides another crucial metric known as the F1 score, used to assess the segmentation outcome. The F1-score or Dice similarity coefficient evaluates the alignment and level of detail between the predicted segmented area and the GT and considers the false alarms and missed values for each class. The accuracy of the segmentation boundaries was evaluated by this metric [42] and takes precedence over the Acc metric:

$$\text{Dice} = \frac{2 \times \text{Pre} \times \text{Recl}}{\text{Pre} + \text{Recl}} \tag{4}$$

The Jaccard similarity index, or Intersection over Union (IoU), says what the correlation between the prediction and GT is [19,43] and represents the overlap and union area ratio for the predicted and GT segmentation:

$$IoU = \frac{TP}{TP + FP + FN} \tag{5}$$

3. Results

The models were trained for 200 epochs with assessing the training/validation loss and the Jaccard criterion (Figure 6). The values of the hyperparameters provided in Table 2 were utilised to obtain optimal training performance and stability. Then, the performances of the trained models were assessed and evaluated using the test datasets and the metrics in Equations (1)–(5) (Table 4).
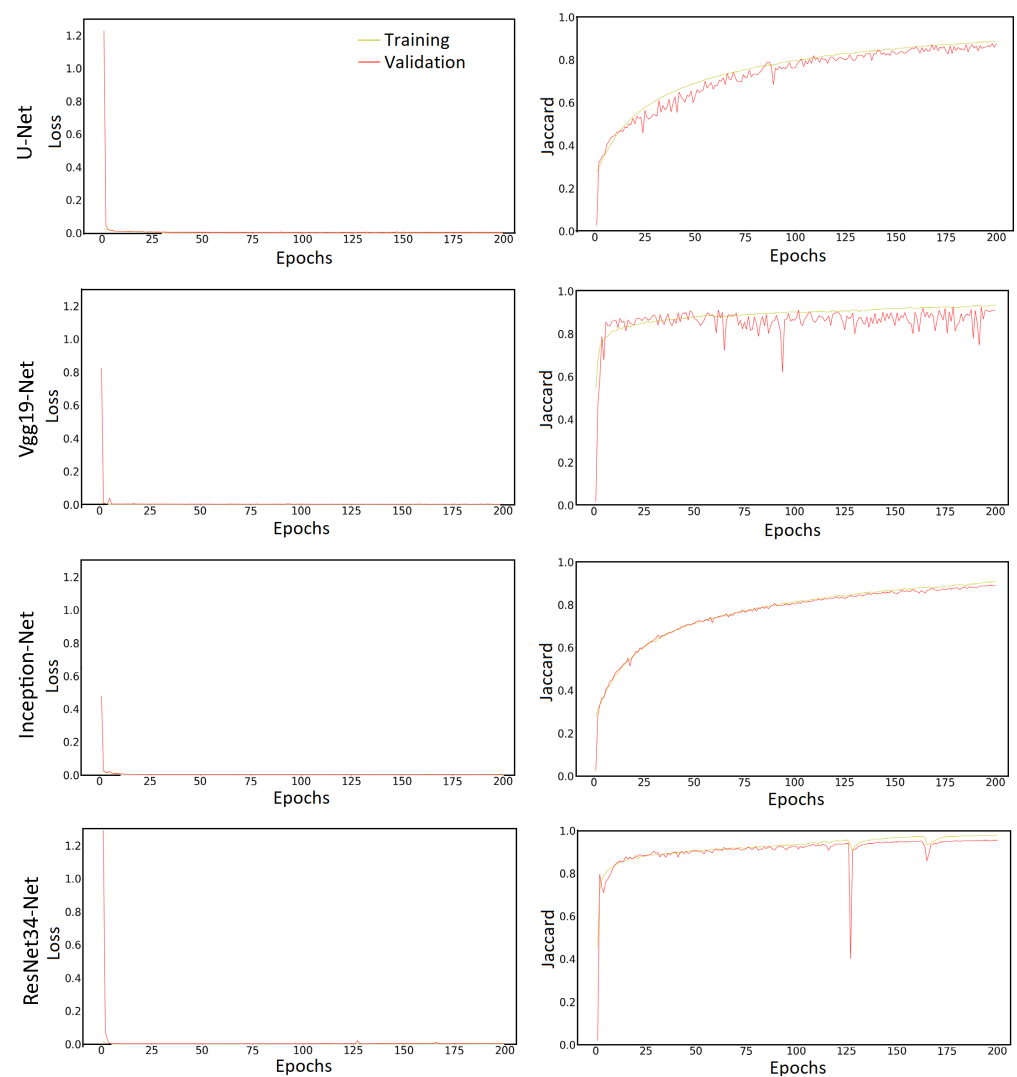

Figure 6. Training/validation plots for the loss criterion (**left**) and the Jaccard criterion (**right**) for the simple U-Net (1st row), VGG19-U-Net (2nd row), Inception-U-Net (3rd row), and ResNet34-U-Net (4th row).

The computational cost is one of the critical factors in training high-performance models based on the lowest computational resources. The four described methods differ significantly in runtime, the number of trainable parameters, and network structures (Table 1). Training the simple U-Net took the longest runtime with the highest number of training parameters. The VGG19-U-Net was trained well in a significantly shorter time

due to the network structure; the number of training parameters was slightly lower than in the simple U-Net. The Inception-U-Net runtime was even faster than the previous two methods. This runtime reduction was followed by a further significant decrease in the number of trainable parameters and higher segmentation performance. The last—ResNet34-U-Net method—achieved the shortest computational cost with the best segmentation performance.

Figure 7 presents the segmentation results for the U-Net-based models proposed in this paper. At the same conditions, the simple U-Net achieved a lower categorical segmentation performance than the other models (when the evaluation metrics are compared). The simple U-Net was inefficient in classifying the cell pixels into the suitable classes and suffered from wrongly segmented cells into the wrong classes (Figure 7, yellow circle). Applying the VGG19-U-Net improved the categorical segmentation performance in terms of the evaluation metrics (Tables 3 and 4). The cells segmented wrongly by the simple U-Net were improved slightly, but wrong classifications still occurred (Figure 7, purple circle). The Inception-U-Net was applied to our datasets as the third hybrid CNN method. This significantly improved the multi-class segmentation results in terms of evaluation metrics (Tables 3 and 4). However, this method suffers from over-segmentation in all classes (Figure 7, black circle). The hybrid ResNet34-U-Net was employed to further improve the object segmentation and classification (Tables 3 and 4). This method achieved mean class accuracies (MCA) of 0.9916 (for the background), 0.9915 (for the divided and unclear cells), and 0.9895 (for the roundish and sharp cells). The confusion matrix (Figure 8) illustrates the related true and predicted classes for the segmentation results.
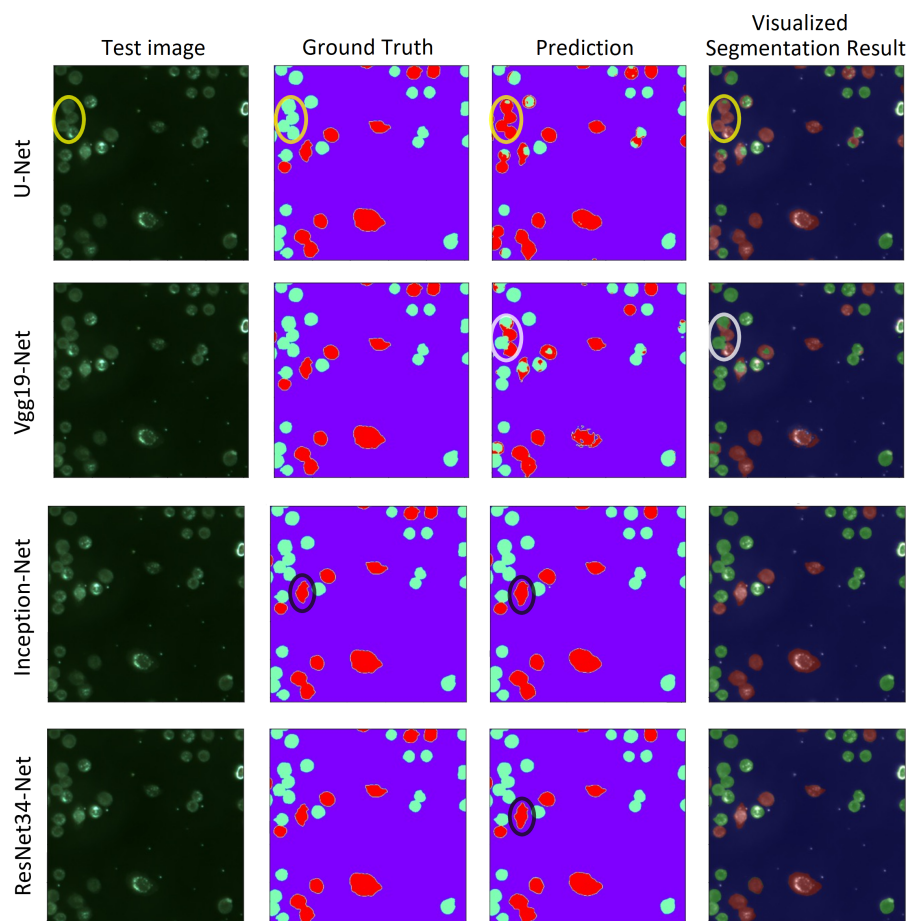
Figure 7. Test image, ground truth, prediction, and 8-bit visualisation of the segmentation results for the U-Net, VGG19-U-Net, Inception-U-Net, and ResNet34-U-Net. The yellow and white circles highlight the wrongly classified and segmented cells. The black circle highlights a different, smoother segmentation result achieved by the ResNet34-U-Net. The image size is 512 × 512.
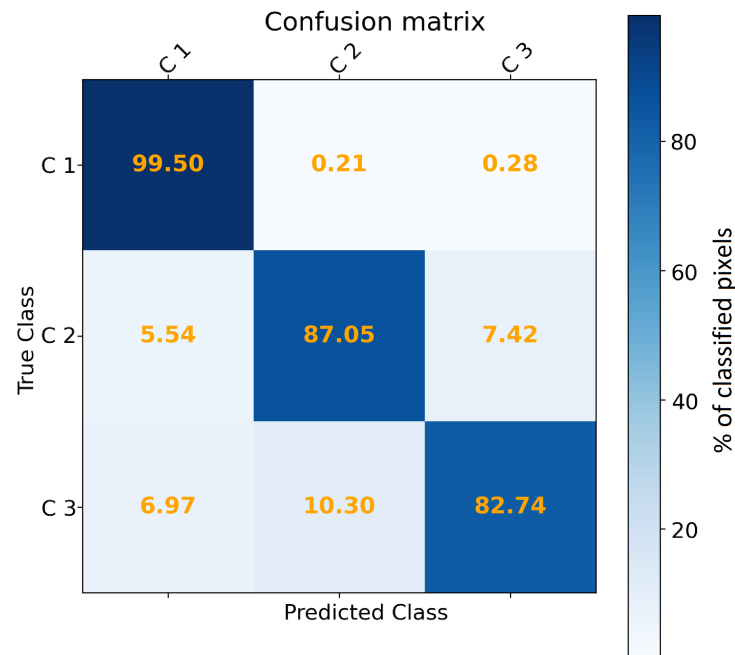
Figure 8. The confusion matrix for the ResNet34-U-Net. Classes: C1—background, C2—divided and unclear cells, and C3—roundish and sharp cells. The columns represent the predicted classes, the rows represent the true classes. Data are presented in % of classified pixels.

Table 3 shows the mean value of the IoU metric for all combinations of class and method. Achieving a higher IoU value for the class of divided unclear cells (C2) was challenging for all methods. The ResNet34-U-Net achieved the highest m-IoU value in all classes.

4. Discussion

The light microscope enables observing living cells in their most natural possible states. However, analysing live cell behaviour in an ordinary light transmission (bright-field) microscope over time is difficult for these technical and biological reasons: (1) The cell morphology and position change significantly depending on the life cycle. (2) Illumination conditions are unstable over image and time. (3) The field of view is small to ensure sufficient statistics on cell behaviour. (4) The images of observed cells are insufficiently spatially resolved and distorted by microscope optics. (5) The traditional image processing methods, including machine learning approaches, have shown sensitivity to the number of training iterations, mis-segmentation, and low computational and runtime performance and recall rate.

Therefore, we enhanced the method described in [23] and developed a microscopic technique with a connecting deep-learning multi-class image segmentation to obviate these complications: (1) Locating the object-sided telecentric objective on the side of the light source (reflection mode) enables us to capture "simple", high-resolved, and low-distorted images on a black background (similar to fluorescence images). (2) Calibrating the microscope optical path balanced the intensities in the whole images for following processing by the CNNs. (3) The larger field of view provides a satisfactory number of cells per snapshot to evaluate cell behaviour. (4) The images of individual cells were segmented and categorised according to their current physiological state.

In the studied neural networks, the symmetric element is the U-Net, composed of two mutually, more or less, symmetric parts: a contracting path to capture the image context vs. an expanding path for precise localisation [21].This symmetry is suitable for image segmentation [44]. The encoder part of the U-Net was replaced with another, more effective, asymmetrical architecture—VGG19, Inception, or ResNet-34—originally designed

for image classification. Both image classification and image segmentation require feature extraction [45].

In the symmetric architecture (U-Net), neurons perform similarly during forward and backward propagation in the convolutional blocks at each level of the network. However, the demand is a network to learn various and more representative features. This requires asymmetric behaviour by enhancing the encoder section performance using the VGG19, ResNet34, or Inception. In this way, we can extract more representative area-specific features together with global features. The hybrid architecture of the U-Net allows more exacting categorical segmentation over microscopy images.

The microscope and relevant image data used in this study are unique. No similar research on categorical segmentation of light reflection microscopy data has ever been performed before. Thus, comparing the results achieved in this study with the literature is hard. The results for the proposed hybrid U-Net-based models could only be compared with similar described methods (Table 5).

Table 5. Values of the evaluation metrics of the CNNs designed for microscopy and medical applications. Comparison with the literature. Green highlights the highest segmentation accuracy value for each metric.

Models	IoU	Dice	Acc
prop. U-Net	0.7062	0.8104	0.9869
prop. VGG19-U-Net	0.7178	0.8218	0.9865
prop. Inception-U-Net	0.7907	0.8762	0.9904
prop. ResNet34-U-Net	0.8067	0.8873	0.9909
Self-Attention U-Net [46]	-	0.799	-
U-Net [26]	0.777	0.753	-
U-Net [47]	-	0.618	-
U-Net+ [48]	0.567	-	-
VGG16-U-Net [49]	-	-	0.961
VGG19-U-Net [50]	-	0.8715	0.8764
Inception-U-Net [51]	-	0.887	-
Inception-U-Net [24]	-	0.95	-
ResNet34-U-Net [52]	0.6915	-	-
SMANet [53]	0.665	0.769	-
DMMN-M3 [54]	0.706 - 0.870	-	-

A simple U-Net structure was the first proposed model. Its final m-IoU score (mean value of all categorical segmentation classes) was 0.7062. The hyperparameter optimisation is expected to lead to a better value of the m-IoU (Table 2).

Sugimoto et al. [46] reached a m-Dice score of 0.799 for multi-class segmentation of cancer and non-cancer cells over the medical PD-L1 dataset. Nishimura et al. [47] applied a U-Net-based weakly supervised method on various microscopy datasets and reached an average m-Dice segmentation score of 0.618. Piotrowski et al. [26] applied a U-Net-based multi-class segmentation method over human-induced pluripotent stem cell images and achieved segmentation IoU and Dice accuracy scores of 0.777 and 0.753, respectively. Long [48] achieved the m-IoU score of 0.567 in single-class semantic segmentation of bright-field, dark-field, and fluorescence images using the enhanced U-Net (U-Net+).

The U-Net encoder part was replaced with the VGG19 architecture to enhance the multi-class segmentation result. The final VGG19-U-Net was optimised for our dataset to decrease trainable parameters in the convolution layers and improve the computational costs and segmentation performance using a dipper network topology and a smaller convolution kernel. In this way, the categorical segmentation accuracy increased to 0.7178 for the m-IoU score in the testing phase. Pravitasari et al. [49] applied a VGG16-U-Net with transfer learning to single-class semantic segmentation of brain tumours in magnetic resonance images and achieved an accuracy of 0.961. Nillmani et al. [50] applied a VGG19-

U-Net to X-ray images for single-class segmentation of COVID-19 infections and achieved accuracy and Dice scores of 0.8764 and 0.8715, respectively.

In the next step, we replaced Google's inception architecture for the U-Net encoder and made a hybrid Inception-U-Net network. The inception module contained kernels of various sizes in the same layer to make the network topology wider instead of deeper and extract more representative features. The m-IoU metric for categorical segmentation increased significantly to 0.7907. The number of trainable parameters was reduced. The computational costs were improved efficiently. Haichun et al. [51] proposed an Inception-U-Net for single-class segmentation of brain tumours and achieved the m-Dice score of 0.887 in the testing phase. Sunny et al. [24] applied an Inception-U-Net to categorical segmentation of fluorescence microscopy datasets and achieved an average Dice metric over all segmentation classes of 0.95.

The model performance was further improved using a hybrid ResNet34-U-Net architecture. The series of residual blocks with the skip connection was implemented into the CNN architecture during the training process to overcome the vanishing gradient and generalisation ability in very deep neural networks. It increased the m-IoU to 0.8067 after the multi-class segmentation. Sunny et al. [24] built up a ResNet34-U-Net, which showed the m-IoU of 0.6915 in the cross-validation phase of fluorescence microscopy multi-class image segmentation. Gao et al. [53] applied a selected Multi-Scale Attention Network (SMANet) for multi-class segmentation in pancreatic pathological images and achieved m-Dice and m-IoU scores of 0.769 and 0.665. Ho et al. [54] proposed Multi-Encoder Multi-Decoder Multi-Concatenation (DMMN-M3) deep CNN for multi-class segmentation in two different image sets of breast cancer and reached an m-IoU score of 0.870 and 0.706.

5. Conclusions

The main objective of this research was to develop an efficient algorithm to segment living HeLa cells and classify them according to their shapes and life cycle stages. We selected the HeLa aggressive cancer cells because they can proliferate rapidly with a replication rate of up to two times in 24 h [55]. Its replication rate and ubiquity in cell culture laboratories make HeLa an efficient and appropriate living cell line for research, industrial, and medical applications. However, the methods described in this study can be employed to analyse other tissue cell lines. Deep learning approaches to reflected light microscopy data analysis delivered efficient and promising outcomes. This research involved variants of hybrid U-Net-based CNN architecture: a simple U-Net, VGG19-U-Net, Inception-U-Net, and ResNet34-U-Net.

The longest training time, the highest number of trainable parameters, and the lowest categorical segmentation performance were observed for the simple U-Net (Table 1). On the contrary, the hybrid ResNet34-U-Net showed the best run time and categorical segmentation performance (Table 4). The computational cost and the number of trainable parameters of the inception network are lower than in the U-Net. Thus, the inception networks are better utilisable for bigger datasets. However, running the inception network requires a higher computational GPU memory.

The Residual Convolutional Neural Network (ResNet) was applied as a hybrid with the U-Net to overcome the gradient vanishing and improve the generalisation ability during training. Using a series of residual blocks with skip connection in each level of the ResNet34-U-Net network resulted in better categorical segmentation. The skip connections in each level of the deep neural networks bypass one or more layers and continuously update the gradient values from one or more previous layers into the layers ahead.

The categorical segmentation gradually improves from simple U-Net to ResNet34-U-Net (as evaluated using performance metrics, Table 4). The ResNet34 encoder network achieved the best categorical segmentation by integrating the residual learning structure to overcome the gradient vanishing with the U-Net as a hybrid ResNet34-U-Net method. However, weakly supervised multi-class semantic segmentation methods need to be further studied to be able to generate the ground truth for any huge datasets. Ensemble-learning ap-

proaches applied in the prediction step could also help achieve more accurate segmentation results using hybrid CNN architectures.

These segmentation methods are potentially applicable to observing and predicting cell behaviour in time-lapse experiments during their life cycles and 3D visualisation of the cell.

References

1. Tang, J.R.; Mat Isa, N.A.; Ch'ng, E.S. A Fuzzy-c-Means-Clustering Approach: Quantifying Chromatin Pattern of Non-neoplastic Cervical Squamous Cells. *PLoS ONE* **2015**, *10*, e0142830. [CrossRef]
2. Rojas-Moraleda, R.; Xiong, W.; Halama, N.; Breitkopf-Heinlein, K.; Dooley, S.; Salinas, L.; Heermann, D.W.; Valous, N.A. Robust Detection and Segmentation of Cell Nuclei in Biomedical Images Based on a Computational Topology Framework. *Med. Image Anal.* **2017**, *38*, 90–103. [CrossRef]
3. Wang, Z. A Semi-Automatic Method for Robust and Efficient Identification of Neighboring Muscle Cells. *Pattern Recogn.* **2016**, *53*, 300–312. [CrossRef]
4. Buggenthin, F.; Marr, C.; Schwarzfischer, M.; Hoppe, P.S.; Hilsenbeck, O.; Schroeder, T.; Theis, F.J. An Automatic Method for Robust and Fast Cell Detection in Bright Field Images from High-Throughput Microscopy. *BMC Bioinform.* **2013**, *14*, 297. [CrossRef]
5. Russell, S.J. *Artificial Intelligence: A Modern Approach*, 3rd ed.; Prentice Hall: Hoboken, NJ, USA, 2010.
6. Huang, X.; Li, C.; Shen, M.; Shirahama, K.; Nyffeler, J.; Leist, M.; Grzegorzek, M.; Deussen, O. Stem Cell Microscopic Image Segmentation Using Supervised Normalized Cuts. In Proceedings of the 2016 IEEE International Conference on Image Processing (ICIP), Phoenix, AZ, USA, 25–28 September 2016; pp. 4140–4144. [CrossRef]
7. Mah, S.A.; Avci, R.; Du, P.; Vanderwinden, J.M.; Cheng, L.K. Supervised Machine Learning Segmentation and Quantification of Gastric Pacemaker Cells. In Proceedings of the 42nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Montreal, QC, Canada, 20–24 July 2020; pp. 1408–1411. [CrossRef]
8. Tikkanen, T.; Ruusuvuori, P.; Latonen, L.; Huttunen, H. Training Based Cell Detection from Bright-Field Microscope Images. In Proceedings of the 9th International Symposium on Image and Signal Processing and Analysis (ISPA), Zagreb, Croatia, 7–9 September 2015; pp. 160–164. [CrossRef]
9. Liimatainen, K.; Ruusuvuori, P.; Latonen, L.; Huttunen, H. Supervised Method for Cell Counting from Bright Field Focus Stacks. In Proceedings of the 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI), Prague, Czech Republic, 13–16 April 2016; pp. 391–394. [CrossRef]
10. Hinton, G.; Sejnowski, T. *Unsupervised Learning: Foundations of Neural Computation*; MIT Press: Cambridge, MA, USA, 1999.
11. Antal, B.; Remenyik, B.; Hajdu, A. An Unsupervised Ensemble-Based Markov Random Field Approach to Microscope Cell Image Segmentation. In Proceedings of the 2013 International Conference on Signal Processing and Multimedia Applications (SIGMAP), Reykjavik, Iceland, 29–31 July 2013; pp. 94–99. [CrossRef]
12. Mualla, F.; Schöll, S.; Sommerfeldt, B.; Maier, A.; Steidl, S.; Buchholz, R.; Hornegger, J. Unsupervised Unstained Cell Detection by SIFT Keypoint Clustering and Self-labeling Algorithm. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI 2014): Proceedings of the 17th International Conference, Boston, MA, USA, 14–18 September 2014*; Lecture Notes in Computer Science; Golland, P., Hata, N., Barillot, C., Hornegger, J., Howe, R., Eds.; Springer: Cham, Switzerland, 2014; Volume 8675, pp. 377–384. [CrossRef]
13. Zheng, X.; Wang, Y.; Wang, G.; Liu, J. Fast and Robust Segmentation of White Blood Cell Images by Self-supervised Learning. *Micron* **2018**, *107*, 55–71. [CrossRef]

14. Zeiler, M.D.; Fergus, R. Visualizing and Understanding Convolutional Neural Networks. In *Computer Vision (ECCV 2014): 13th European Conference, Zurich, Switzerland, 6–12 September 2014*; Lecture Notes in Computer Science; Springer: Berlin/Heidelberg, Germany, 2014; Volume 8689; pp. 818–833. [CrossRef]

15. Lin, S.; Norouzi, N. An Effective Deep Learning Framework for Cell Segmentation in Microscopy Images. In Proceedings of the 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Online, 1–5 November 2021; pp. 3201–3204. [CrossRef]

16. Kumar, N.; Verma, R.; Anand, D.; Sethi, A. Multi-Organ Nuclei Segmentation Challenge. Available online: https://monuseg.grandchallenge.org/ (accessed on 5 May 2021).

17. Caicedo, J.C.; Goodman, A.; Karhohs, K.W.; Cimini, B.A.; Ackerman, J.; Haghighi, M.; Heng, C.; Becker, T.; Doan, M.; McQuin, C.; et al. Broad Bioimage Benchmark Collection. Available online: https://bbbc.broadinstitute.org/BBBC038 (accessed on 5 May 2021).

18. Thi Le, P.; Pham, T.; Hsu, Y.C.; Wang, J.C. Convolutional Blur Attention Network for Cell Nuclei Segmentation. *Sensors* **2022**, *22*, 1586. [CrossRef] [PubMed]

19. Long, J.; Shelhamer, E.; Darrell, T. Fully Convolutional Networks for Semantic Segmentation. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 3431–3440. [CrossRef]

20. Ben-Cohen, A.; Diamant, I.; Klang, E.; Amitai, M.; Greenspan, H. Fully Convolutional Network for Liver Segmentation and Lesions Detection in Deep Learning and Data Labeling for Medical Applications. In *Proceedings of the Deep Learning and Data Labeling for Medical Applications: 1st International Workshop (LABELS 2016), the 2nd International Workshop (DLMIA 2016), Held in Conjunction with MICCAI 2016, Athens, Greece, 21 October 2016*; Lecture Notes in Computer Science; Carneiro, G., Mateus, D., Peter, L., Eds.; Springer: Cham, Switzerland, 2016; Volume 10008, pp. 77–85. [CrossRef]

21. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, 5–9 October 2015*; Lecture Notes in Computer Science; Navab, N., Hornegger, J., Wells, W., Frangi, A., Eds.; Springer: Cham, Switzerland, 2015; Volume 9321, pp. 234–241. [CrossRef]

22. Shibuya, E.; Hotta, K. Cell Image Segmentation by Using Feedback and Convolutional LSTM. *Vis. Comput.* **2021**, *38*, 3791–3801. [CrossRef]

23. Ghaznavi, A.; Rychtáriková, R.; Saberioon, M.; Štys, D. Cell Segmentation from Telecentric Bright-Field Transmitted Light Microscopy Images Using a Residual Attention U-Net: A Case Study on HeLa line. *Comp. Biol. Med.* **2022**, *147*, 105805. [CrossRef] [PubMed]

24. Sunny, S.P.; Khan, A.I.; Rangarajan, M.; Hariharan, A.; Birur N, P.; Pandya, H.J.; Shah, N.; Kuriakose, M.A.; Suresh, A. Oral Epithelial Cell Segmentation from Fluorescent Multichannel Cytology Images Using Deep Learning. *Comput. Methods Programs Biomed.* **2022**, *227*, 107205. [CrossRef]

25. Bakir, M.E.; Yalim Keles, v.H. Deep Learning Based Cell Segmentation Using Cascaded U-Net Models. In Proceedings of the 2021 29th Signal Processing and Communications Applications Conference (SIU), Istanbul, Turkey, 9–11 June 2021; pp. 1–4. [CrossRef]

26. Piotrowski, T.; Rippel, O.; Elanzew, A.; Nießing, B.; Stucken, S.; Jung, S.; König, N.; Haupt, S.; Stappert, L.; Brüstle, O.; et al. Deep-Learning-Based Multi-Class Segmentation for Automated, Non-invasive Routine Assessment of Human Pluripotent Stem Cell Culture Status. *Comp. Biol. Med.* **2021**, *129*, 104172. [CrossRef] [PubMed]

27. Platonova, G.; Štys, D.; Souček, P.; Lonhus, K.; Valenta, J.; Rychtáriková, R. Spectroscopic Approach to Correction and Visualization of Bright-Field Light Transmission Microscopy Biological Data. *Photonics* **2021**, *8*, 333. [CrossRef]

28. Štys, D.; Náhlík, T.; Macháček, P.; Rychtáriková, R.; Saberioon, M. Least Information Loss (LIL) Conversion of Digital Images and Lessons Learned for Scientific Image Inspection. In *Bioinformatics and Biomedical Engineering: 4th International Conference (IWBBIO 2016), Granada, Spain, 20–22 April 2016*; Lecture Notes in Computer Science; Ortuno, F., Rojas, I., Eds.; Springer: Cham, Switzerland, 2016; Volume 9656, pp. 527–536. [CrossRef]

29. Buades, A.; Coll, B.; Morel, J.M. A Non-local Algorithm for Image Denoising. In Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, USA, 20–25 June 2005; Volume 2, pp. 60–65. [CrossRef]

30. Ghaznavi, A.; Rychtáriková, R.; Císař, P.; Ziaei, M.; Štys, D. Telecentric Bright-Field Reflected Light Microscopic Dataset. Available online: https://doi.org/10.5061/dryad.6q573n637 (accessed on 1 January 2024).

31. Zeiss. APPEAR—Automated Image Analysis. Available online: https://www.apeer.com/ (accessed on 12 December 2021).

32. Lu, Y.; Liu, A.A.; Su, Y.T. Chapter 6 - Mitosis detection in biomedical images. In *Computer Vision for Microscopy Image Analysis*; Computer Vision and Pattern Recognition; Chen, M., Ed.; Academic Press: Cambridge, MA, USA, 2021; pp. 131–157. [CrossRef]

33. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet Classification with Deep Convolutional Neural Networks. *Commun. ACM* **2017**, *60*, 84–90. [CrossRef]

34. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. In Proceedings of the 3rd International Conference on Learning Representations (ICLR2015), San Diego, CA, USA, 7–9 May 2015; pp. 1–14. [CrossRef]

35. Hamwi, W.A.; Almustafa, M.M. Development and Integration of VGG and Dense Transfer-Learning Systems Supported with Diverse Lung Images for Discovery of the Coronavirus Identity. *Inform. Med. Unlocked* **2022**, *32*, 101004. [CrossRef] [PubMed]

36. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going Deeper with Convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2015), Boston, MA, USA, 7–12 June 2015; pp. 1–9. [CrossRef]

37. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778. [CrossRef]

38. Abadi, M.; Agarwal, A.; Barham, P.; Brevdo, E.; Chen, Z.; Citro, C. TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. In Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation, Savannah, GA, USA, 2–4 November 2016; pp. 265–283.

39. Google Research Colaboratory. Available online: https://colab.research.google.com/?utm_source=scs-index (accessed on 12 December 2021).

40. Lin, T.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal Loss for Dense Object Detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *42*, 318–327. [CrossRef]

41. Pan, X.; Li, L.; Yang, H.; Liu, Z.; Yang, J.; Fan, Y. Accurate Segmentation of Nuclei in Pathological Images via Sparse Reconstruction and Deep Convolutional Networks. *Neurocomputing* **2017**, *229*, 88–99. [CrossRef]

42. Csurka, G.; Larlus, D.; Perronnin, F. What Is a Good Evaluation Measure for Semantic Segmentation? In Proceedings of the British Machine Vision Conference (BMVC 2013), Bristol, UK, 9–13. September 2013; BMVA Press: Durham, UK, 2013; pp. 32.1–32.11. [CrossRef]

43. Vijay, B.; Kendall, A.; Cipolla, R. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *39*, 228–233. [CrossRef]

44. Sankesara, H. UNet. Introducing Symmetry in Segmentation. *Towards Data Science*, 23 January 2019. Available online: https://towardsdatascience.com/u-net-b229b32b4a71 (accessed on 1 January 2024).

45. Gao, Y.; Che, X.; Xu, H.; Bie, M. An enhanced feature extraction network for medical image segmentation. *Appl. Sci.* **2023**, *13*, 6977. [CrossRef]

46. Sugimoto, T.; Ito, H.; Teramoto, Y.; Yoshizawa, A.; Bise, R. Multi-Class Cell Detection Using Modified Self-Attention. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), New Orleans, LA, USA, 21–24 June 2022; pp. 1854–1862. [CrossRef]

47. Nishimura, K.; Wang, C.; Watanabe, K.; Ker, D.F.E.; Bise, R. Weakly Supervised Cell Instance Segmentation under Various Conditions. *Med. Image Anal.* **2021**, *73*, 102182. [CrossRef] [PubMed]

48. Long, F. Microscopy cell nuclei segmentation with enhanced U-Net. *BMC Bioinform.* **2020**, *21*, 8. [CrossRef]

49. Pravitasari, A.A.; Iriawan, N.; Almuhayar, M.; Azmi, T.; Irhamah, I.; Fithriasari, K.; Purnami, S.W.; Ferriastuti, W. UNet-VGG16 with Transfer Learning for MRI-Based Brain Tumor Segmentation. *TELKOMNIKA* **2020**, *18*, 1310–1318. [CrossRef]

50. Nillmani.; Sharma, N.; Saba, L.; Khanna, N.N.; Kalra, M.K.; Fouda, M.M.; Suri, J.S. Segmentation-Based Classification Deep Learning Model Embedded with Explainable AI for COVID-19 Detection in Chest X-ray Scans. *Diagnostics* **2022**, *12*, 2132. [CrossRef] [PubMed]

51. Li, H.; Li, A.; Wang, M. A Novel End-to-End Brain Tumor Segmentation Method Using Improved Fully Convolutional Networks. *Comp. Biol. Med.* **2019**, *108*, 150–160. [CrossRef] [PubMed]

52. Patel, G.; Tekchandani, H.; Verma, S. Cellular Segmentation of Bright-field Absorbance Images Using Residual U-Net. In Proceedings of the 2019 International Conference on Advances in Computing, Communication and Control (ICAC3), Mumbai, India, 20–21 December 2019; pp. 1–5. [CrossRef]

53. Gao, E.; Jiang, H.; Zhou, Z.; Yang, C.; Chen, M.; Zhu, W.; Shi, F.; Chen, X.; Zheng, J.; Bian, Y.; et al. Automatic Multi-Tissue Segmentation in Pancreatic Pathological Images with Selected Multi-Scale Attention Network. *Comp. Biol. Med.* **2022**, *151*, 106228. [CrossRef] [PubMed]

54. Ho, D.J.; Yarlagadda, D.V.; D'Alfonso, T.M.; Hanna, M.G.; Grabenstetter, A.; Ntiamoah, P.; Brogi, E.; Tan, L.K.; Fuchs, T.J. Deep Multi-Magnification Networks for Multi-Class Breast Cancer Image Segmentation. *Comput. Med. Imaging Graph.* **2021**, *88*, 101866. [CrossRef]

55. Rahbari, R.; Sheahan, T.; Modes, V.; Collier, P.; Macfarlane, C.; Badge, R.M. A novel L1 retrotransposon marker for HeLa cell line identification. *Biotechniques* **2009**, *46*, 277–284. [CrossRef]

56. Ghaznavi, A. GitHub Repository. Available online: https://github.com/AliGhaznavi1986/Hybrid-CNNs-for-multi-class-segmentation.