

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/373166030>

Hybrid Deep-Learning Multi-Class Segmentation of Hela Cells in Reflected Light Microscopy Images

Preprint · August 2023

DOI: 10.2139/ssrn.4543079

CITATIONS

0

READS

55

5 authors, including:



Ali Ghaznavi

Federal Institute For Materials Research and Testing

13 PUBLICATIONS 50 CITATIONS

[SEE PROFILE](#)



Renata Rychtarikova

University of South Bohemia in České Budějovice

64 PUBLICATIONS 247 CITATIONS

[SEE PROFILE](#)



Petr Cisar

University of South Bohemia in České Budějovice

116 PUBLICATIONS 1,339 CITATIONS

[SEE PROFILE](#)

Hybrid deep-learning multi-class segmentation of HeLa cells in reflected light microscopy images

Ali Ghaznavi^a, Renata Rychtáriková^a, Petr Císař^a, Mohammadmehdi Ziaeī^b, Dalibor Štys^a

^a*Faculty of Fisheries and Protection of Waters, South Bohemian Research Center of Aquaculture and Biodiversity of Hydrocenoses, Institute of Complex Systems, University of South Bohemia in České Budějovice, Zámek 136, 373 33 Nové Hrady, Czech Republic*

^b*Faculty of Science, University of South Bohemia, Branišovská 1760, 37005 České Budějovice, Czech Republic*

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 images of this kind. 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. Different hybrid convolution neural networks – a simple U-Net, VGG19-U-Net, Inception-U-Net, and ResNet34-U-Net architectures – 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-

*Corresponding author: Ali Ghaznavi
Email address: ghaznavi@frov.jcu.cz (Ali Ghaznavi)

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 1. Introduction

² Cell detection and segmentation is a fundamental process in microscopy cell
³ image analysis. This is also a challenging task due to the complexity of these
⁴ images. On the other hand, the information from the segmented living cells
⁵ can play an essential role in further analysis, such as observing and estimat-
⁶ ing cell behaviour, their number and dimensions. Recently developed artificial
⁷ intelligence (AI) methods have achieved promising outcomes in this field. The
⁸ segmentation methods for analysing cell cultures can be categorised as machine
⁹ learning (ML) or deep learning (DL).

¹⁰ 1.1. *Cell culture segmentation with machine learning methods*

¹¹ The number of cell detection-segmentation ML methods has grown rapidly
¹² as a result of the low performance of simple techniques such as threshold-based
¹³ [1], region-based [2], or morphological approaches [3, 4] when processing such
¹⁴ complex images. The ML methods can be further classified as supervised or
¹⁵ unsupervised.

¹⁶ The supervised methods generate a mathematical function or a model from
¹⁷ the training data to map a new data sample [5]. Trained and optimised param-
¹⁸ eters using the graph-based Supervised Normalized Cut Segmentation (SNCS)
¹⁹ with loosely annotated images separate overlapping and curved cells better than
²⁰ the traditional image processing methods [6]. The Fast Random Forest (FRF)

21 classification using Trainable WEKA Segmentation outperformed the Decision
22 Table and Naïve Bayes classification methods in sensitivity, accuracy, and F-
23 measure when extracting the Interstitial cells of Cajal networks from 3D con-
24 focal microscopy images. However, the method showed higher computational
25 costs due to the FRF’s structure [7]. A method combining the Histogram of
26 Oriented Gradients and the Support Vector Machine (SVM) extracted and clas-
27 sified the feature descriptors as cells or non-cells in bright-field microscopy data.
28 The method was susceptible to the number of iterations in the training process,
29 which is a crucial step to eliminate false positive detections [8]. A Logistic
30 Regression classification with intensity values of 25 focal planes as features, fol-
31 lowed by the binary erosion with a large circular structuring element, counted
32 the cells in bright-field microscopy images. However, the method showed miss-
33 segmentation and a low recall rate [9].

34 The unsupervised ML algorithms require no pre-assigned labels or scores for
35 the training data [10]. Unsupervised segmentation using the Markov Random
36 Field considered an image as a series of planes based on Bit Plane Slicing. The
37 planes were used as initial labelling for an ensemble of segmentations. The
38 robust cell segmentation was achieved with pixel-wise voting. However, this
39 method was too sensitive to the confidence threshold [11]. A combination of a
40 Scale-Invariant Feature Transform, a self-labelling, and two clustering methods
41 segmented unstained cells in bright-field micrographs. The method was fast and
42 accurate but sensitive to the feature selection to avoid overfitting [12]. A self-
43 supervised (i.e., a kind of unsupervised) learning approach combined unsuper-
44 vised initial coarse segmentation (K-means clustering) followed by supervised
45 segmentation refinement (SVM pixel classifier) to separate white blood cells.
46 However, the unsupervised part of the method generates a rough segmentation
47 result. In the case of complex datasets, the supervised part of the method
48 cannot work efficiently due to fuzzy boundaries [13].

49 1.2. Cell culture segmentation with deep learning methods

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

56 Mask RCNN with a Shape-Aware Loss generated the HeLa cell's segmen-
57 tation masks with a good performance [15]. A Convolutional Blur Attention
58 (CBA) network consisted of down- and up-sampling procedures for nuclei seg-
59 mentation in standard challenge datasets [16, 17], with a good value of the
60 aggregated Jaccard index. The reduced number of trainable parameters led to
61 a reasonable decrease in the computational cost [18]. The size of input images of
62 a convolutional network can be of different custom sizes so that it can be trained
63 end-to-end, pixel-to-pixel, and produce an output of the appropriate size. Ef-
64 fective inference and learning can achieve successful semantic segmentation in
65 complex microscopic and medical images [19, 20].

66 A U-Net architecture containing a contracting path to capture context and a
67 symmetric expanding path for precise localisation showed strong data augmen-
68 tation in the training process. It was optimised when applied to small datasets
69 and performed efficiently in semantic segmentation of photon microscopy (phase
70 contrast and DIC) images [21]. A Feedback U-Net with the convolutional Long
71 Short-Term Memory network, working on *Drosophila* cell image dataset and
72 mouse cell image dataset, generally showed a low level of accuracy, depend-
73 ing on the segmented class (cytoplasm, cell membrane, mitochondria, synapses)
74 [22]. A Residual Attention U-Net-based method segmented living HeLa cells in
75 bright-field light microscopy data with a high IoU metric. The method combined
76 the self-attention mechanism to highlight the remarkable features and suppress
77 activations in the irrelevant image regions, and the residual mechanism to over-
78 come with vanishing gradient problem [23]. Multi-class cell segmentation in
79 fluorescence images combining U-Net (a deeper network) with ResNet-34 (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. [8, 7]. The ML methods were also not very efficient in training the multi-class segmentation models in large time-lapse image series. Compared with the 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 is specific for its high pixel resolution in rgb mode and requires pre-processing steps to suppress optical

110 vignetting and camera noise. The data shows unlabelled in-focused and out-of-
111 focus living cells in their physiological state. Thus, the obtained segmentation
112 method is applicable to observing and predicting cell behaviour in time-lapse
113 experiments during their life cycles and 3D visualisation of the cell.

114 **2. Materials and methods**

115 *2.1. Cell preparation and microscope specification*

116 The cells were prepared as written in [23], Section 2.1. Human HeLa cell line
117 (European Collection of Cell Cultures, Cat. No. 93021013) was prepared and
118 cultivated to low optical density overnight at 37°C, 5% CO₂, and 90% relative
119 humidity. The nutrient solution consisted of Dulbecco's modified Eagle medium
120 (87.7%) with high glucose (>1 g L⁻¹), fetal bovine serum (10%), antibiotics and
121 antimycotics (1%), L-glutamine (1%), and gentamicin (0.3%; all purchased from
122 Biowest, Nuaille, France). The HeLa cells were maintained in a Petri dish with
123 a cover glass bottom and lid at room temperature of 37°C.

124 The data was collected by running several time-lapse image series experi-
125 ments of living human HeLa cells on a glass Petri dish using a high-resolved
126 reflected light microscope to observe the microscopic objects and cells. This mi-
127 croscope was designed by the Institute of Complex System (ICS, Nové Hrady,
128 Czech Republic) and built by Optax (Prague, Czech Republic) and ImageCode
129 (Brloh, Czech Republic) in 2021. The microscope has a simple construction
130 of the optical path. The light from a Schott VisiLED S80-25 LED Brightfield
131 Ringlight was reflected from a sample to reach a telecentric measurement ob-
132 jective TO4.5/43.4-48-F-WN (Vision & Control GmbH, Shul, Germany) and an
133 Arducam AR1820HS 1/2.3-inch 10-bit RGB camera with a chip of 4912×3684
134 pixel resolution. The images were captured as a primary (raw) signal with a
135 theoretical pixel size (size of the object projected onto the camera pixel) of 113
136 nm. The software (developed by the ICS) controls the capture of the primary
137 signal with a camera exposure of 998 ms. All these experiments were performed
138 in time-lapse to observe cells' behaviour over time.

139 *2.2. Data preparation and pre-processing*

140 Several time-lapse experiments were completed with HeLa cells using a re-
141 flected bright-field microscope (Sect. 2.1). The microscope control software cal-
142 ibrated the microscope optical path and corrected all image series using the al-
143 gorithm proposed in [27] to avoid image background inhomogeneities and noise.

144 After the calibration step, the raw image representations were converted to
145 8-bit colour (rgb) images of resolution (number of pixels) quarter of the original
146 raw images. The Bayer mask pixels quadruplets [28] were merged as follows:
147 each pair of green camera filter pixels' intensities were averaged as the green
148 image channel. The red and blue camera filter pixels were adopted into the
149 relevant image channel. Then, images were rescaled to 8 bits after creating
150 the image series intensity histogram and omitting unoccupied intensity levels.
151 This bit reduction ensured the maximal information preservation and mutual
152 comparability of the images through the time-lapse series.

153 After generating 8-bit images, the denoising method [29] was applied to
154 minimise the background noise in the constructed rgb images at preserving the
155 texture details. Afterwards, the image series from different time-lapse experi-
156 ments were cropped into the 1024×1024 pixel size to achieve 650 images as
157 the main dataset. The image dataset is accessible at the Dryad data publishing
158 platform [30].

159 For multi-class segmentation, one of three cell states was assigned to each
160 cell manually using Apeer platform [31]: (1) a background class containing
161 no cells, (2) a cell class containing larger dilated adhered or migrating cells
162 with unclear borders by which we anticipate they are growing, and (3) a cell
163 class including roundish cells with sharper borders when the cells are assumed
164 in their early stage of the life cycle, having no division state yet, or at the
165 beginning of the division. The detection of the ratio of cells in mitosis plays
166 an important role in many biomedical activities, such as biological research and
167 medical diagnosis [32]. Figure 1 depicts a sample of the resized dataset and
168 relevant generated mask classes as ground truth of the size of 512×512 pixels.
169 The labelled images were used as training (80%), testing (20%), and evaluation

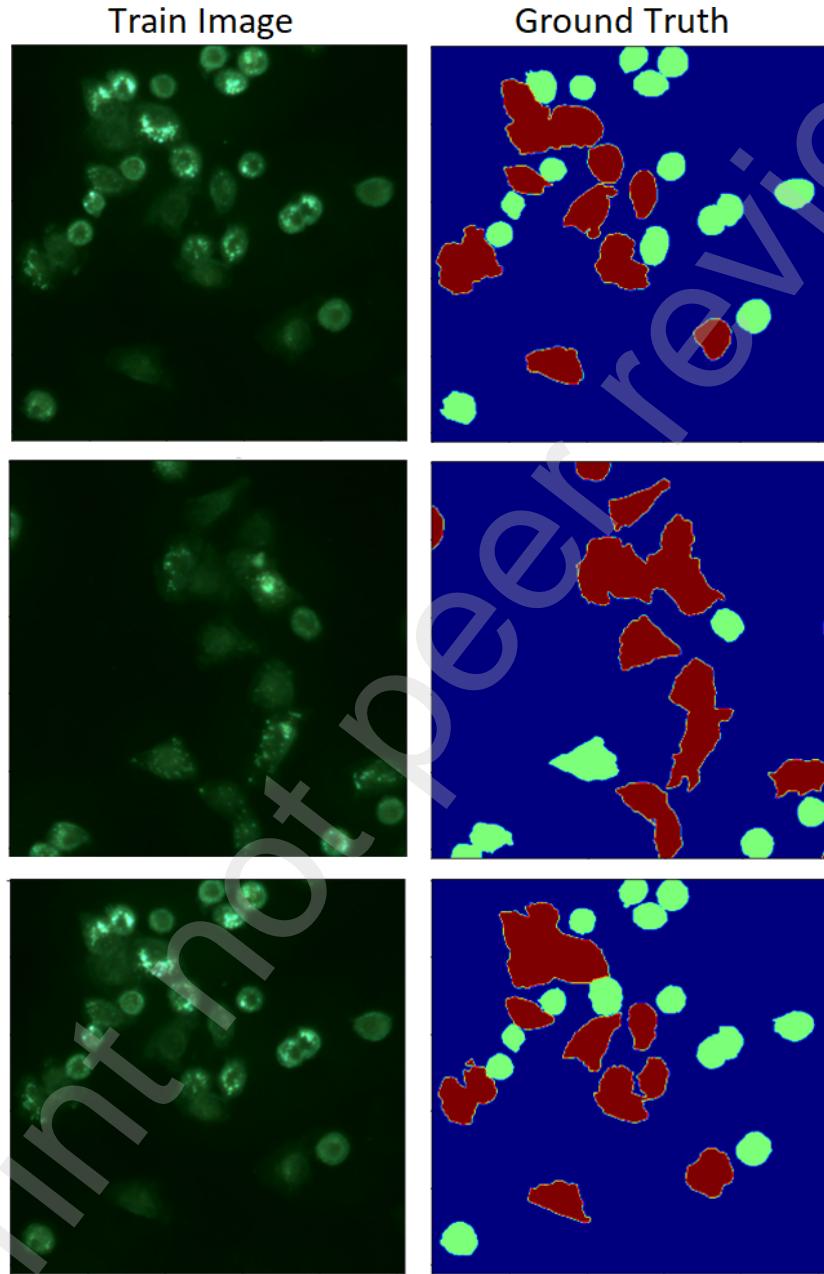


Figure 1: Examples of the train sets and their ground truths. The image size is 512×512 . The green and red class represents the roundish sharp cells and the migrating unclear cells, respectively.

170 (20% of the training set) sets in the proposed neural network architectures.

171 *2.3. The Neural Network Model Architectures*

172 *2.3.1. U-Net*

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

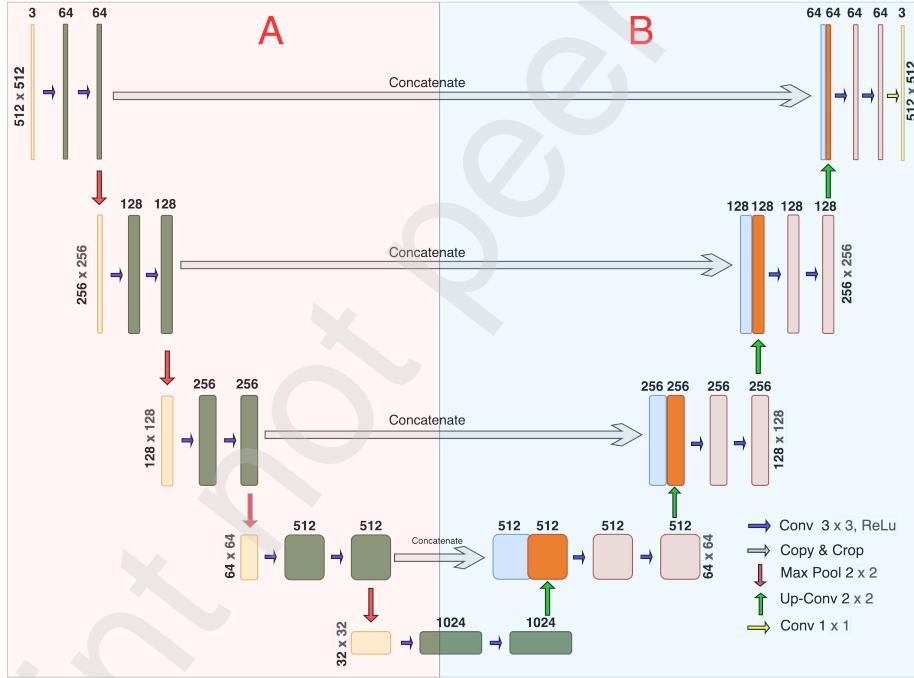


Figure 2: The simple U-Net model architecture. A) The encoder section. B) The decoder section.

179 The first input layer accepts rgb 512×512-sized training set images. Each
180 level of the proposed U-Net includes two 3×3 convolutions. Batch normalisation
181 follows each convolution, and ”ReLU” is used as an activation function. In

182 the down-sampling (encoder) part (Fig. 2A), each encoder "level" consists of
183 a 2×2 max-pooling operation with a stride of two. The max-pooling process
184 extracts the maximal value in the 2×2 area. By completing the down-sampling
185 in each level of the encoder part, convolutions will double the number of feature
186 channels.

187 In each level (from bottom to top) of the up-sampling (decoder) section
188 (Fig. 2B), the height and width of the existing feature maps are doubled. In the
189 concatenation step, the high-resolution deep semantic and shallow features were
190 combined with the feature maps from the encoder section. After concatenation,
191 the output feature maps have channels twice the size of the input feature maps.
192 The "softmax" activation function in the top, 1×1 convolution-sized, output
193 decoder layer predicts the occurrence of each pixel in each of the three classes.
194 Padding in the convolution process allowed us to achieve the same input and
195 output layers size. Each of those classes, achieved by the softmax activation,
196 represents the probability of belonging each pixel into each class. In the final
197 step, the "argmax" operation assigned each pixel to the class, where the highest
198 probability value was achieved. This computational result, combined with the
199 Categorical Focal Loss function, becomes the energy function of the U-Net.

200 2.3.2. The VGG19-U-Net

201 Many modified artificial neural networks, such as AlexNet [33], ZFNet [14],
202 and VGG [34], have been developed as hybrids with the U-Net to simplify U-
203 Net. In this study, a VGG-Net architecture replaced the U-Net encoder path.
204 In this way, we combined two powerful architectures to improve the categorical
205 segmentation of our unique microscopy dataset. The VGG-Net was proposed by
206 Simonyan and Zisserman [34] from Oxford's Visual Geometry Group (VGG). A
207 VGG-16 proved to be one of the most efficient classification networks. However,
208 a VGG-19 performed even more effectively than VGG-16 [35]. The VGG-19
209 comprises a network with a deeper topology and smaller convolution kernels
210 to simulate a perceptual field of view. This architecture is designed to reduce
211 the number of trainable parameters and decrease computational costs compared

212 with the simple U-Net. Figure 3 represents the VGG19-U-Net proposed in this
 213 study. The left side of the network (Fig. 3A) shows the architecture of the VGG-
 214 19 encoder section with 16 convolution layers, three fully connected layers, and 5
 215 MaxPool layers in 5 blocks. The convolution blocks at each level are followed by
 216 a 2×2 max-pooling operation with the stride of two to extract the maximal value
 217 in the 2×2 area. The first layer of the VGG network has 64 channels, and each
 218 subsequent layer is doubled up to 512 channels. The right side of the network
 219 (Fig. 3B) is a schema of the decoder part with five blocks. A concatenation
 220 step between each VGG-19 encoder layer and each U-Net decoder layer (Fig. 3)
 221 combines the feature maps from the encoder part with the high-resolution deep
 222 semantic and shallow features from the decoder part. The last decoder layer
 223 has a convolution size of 1×1 and predicts the probability values for each pixel
 224 and each of the three classes using the "softmax" activation function.

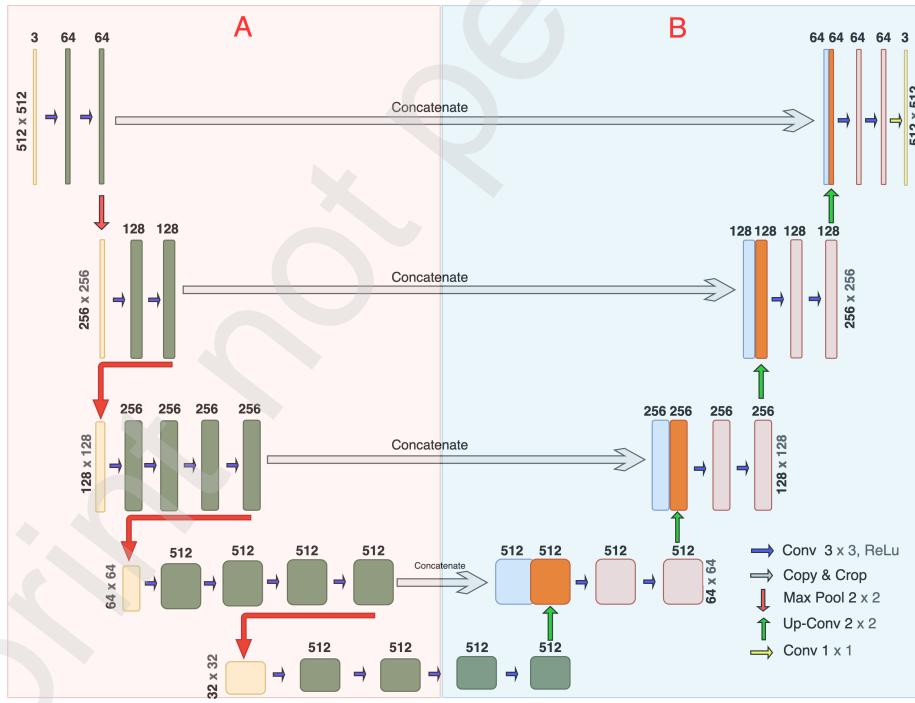


Figure 3: The hybrid VGG19-U-Net architecture. A) The VGG-19 encoder part. B) The U-Net decoder part.

225 *2.3.3. The Inception-U-Net*

226 The complexity of the U-Net network about the number of trainable param-
227 eters leads to higher runtime and computational costs (Tab. 4). On the other
228 hand, in image analysis, applying fixed kernel size in all convolution layers can
229 make it difficult to extract all feature descriptors of different sizes. For example,
230 in microscopy image analysis, some (tiny) features are at the local level, and
231 some (larger) are at the global level. The network cannot extract the represen-
232 tative features for big objects when the small kernel is selected in convolution
233 operations. If the kernel size is big, the network will miss extracting the features
234 representative at the pixel level. In other words, the larger kernel can extract
235 a global feature representation over a large image area, and the smaller kernel
236 has been considered for detecting area-specific features. Google’s inception deep
237 learning method [36], known as the Inception architecture, was selected to build
238 a hybrid Inception-U-Net architecture (Fig. 4) to improve segmentation results
239 in our datasets further.

240 The inception module is well known for its computational efficiency by inte-
241 grating different sizes of convolutions. The inception module applies kernels of
242 different sizes within the same architecture layer and becomes wider (instead of
243 deeper) with the layers (Fig. 4B). The convolution layers were replaced with an
244 inception module (Fig. 4A) in all five levels of the encoder and decoder sections
245 of the original U-Net structure. The inception module consists of multiple sets
246 of 3×3 convolutions, 1×1 convolutions, 3×3 max-pooling, and cascaded 3×3
247 convolutions. The number of filters at each convolution layer was doubled on
248 the encoder side. The size of the output feature map (height and width) was
249 halved on the last encoder layer.

250 The up-sampling (decoder) architecture section (Fig. 4A, left side) was also
251 equipped with an inception module at each level. The skip connection connected
252 the encoder and decoder parts to produce a finer prediction. The spatial feature
253 maps from the encoder are concatenated with the decoder feature maps. The
254 rectified linear unit (ReLU) was selected as an activation function for each

255 layer and performed batch normalisation in each inception module. At the last
 256 layer, a 1×1 convolution layer together with the "softmax" activation function
 257 generated three segmentation classes of the feature maps for the given input
 258 image. Each pixel was assigned to one class according to the highest probability
 259 value achieved among the classes. The Categorical Focal Loss function has been
 260 considered an energy function for this Inception-U-Net.

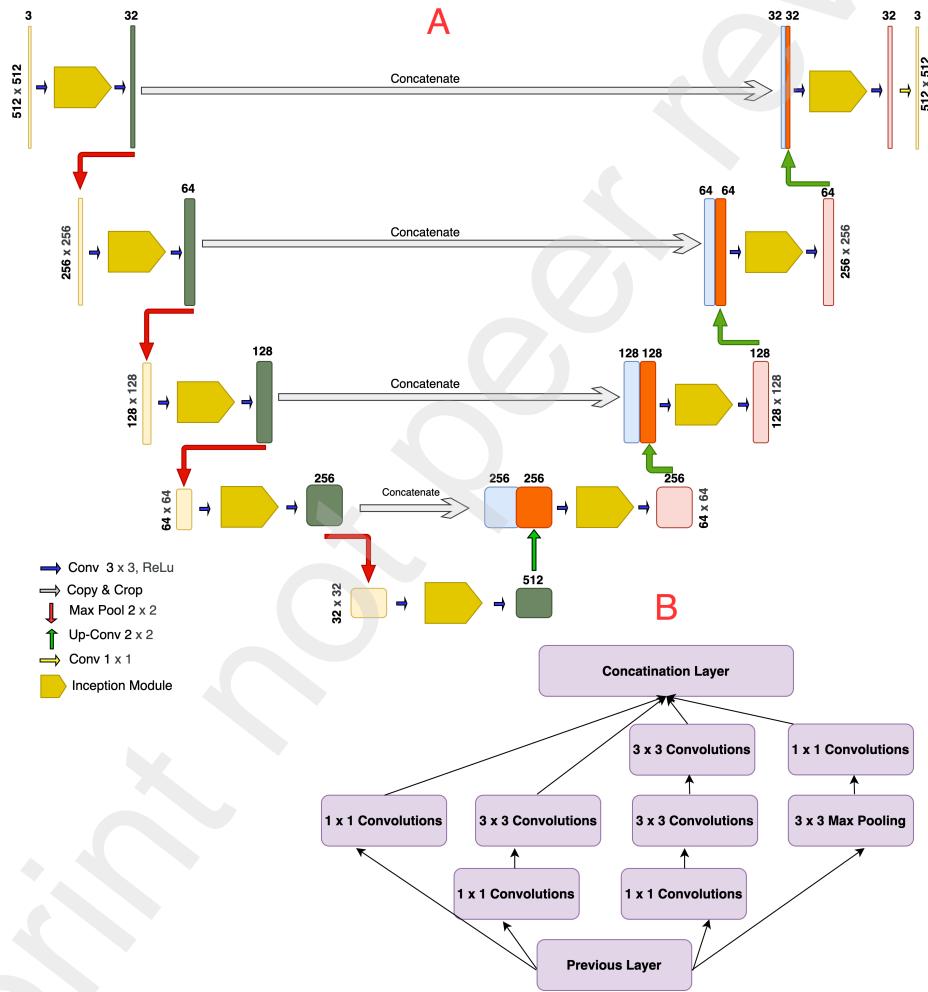


Figure 4: A) The Inception-U-Net architecture. B) The internal architecture of one inception module.

261 2.3.4. *The ResNet34-U-Net*

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

275 The ResNet-34-U-Net architecture used in our study (Fig. 5) has 34 layers
276 and four residual convolution steps with a total of 16 residual blocks (red and
277 purple arrows). The first convolution layer has 64 filters with a kernel size
278 of 7×7 , followed by a max-pooling layer. Each residual block consists of two
279 3×3 convolution layers followed by the ReLU activation function and batch
280 normalisation with the identity shortcut connection.

281 After the first 7×7 convolution layer, the feature map size halved to 256×256 .
282 At the first residual level, three residual convolution blocks were applied to the
283 achieved feature maps, and the output size of the feature maps was halved to
284 128×128 . Four residual convolution blocks in the second residual step decreased
285 the size of the output feature maps to 64×64 . Six residual convolution blocks
286 in the third residual step gave a feature map size of 32×32 . The last residual
287 step consists of three residual convolution blocks to achieve a feature map with
288 a size of 16×16 .

289 The up-sampling section of the network (Fig. 5B) gets the input with the
290 feature map size of 16×16 with 512 channels and a 2×2 up-convolution step with

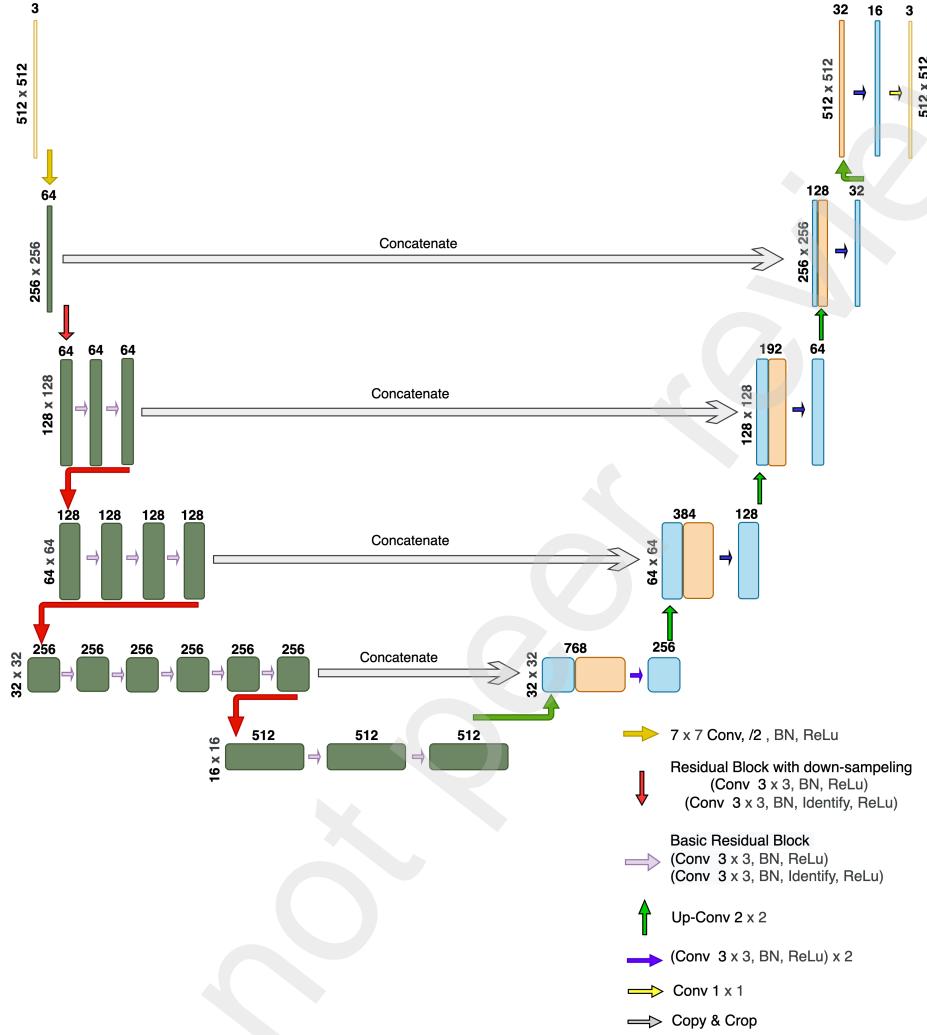


Figure 5: The hybrid ResNet-34-U-Net architecture.

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 ResNet-34, the number of trainable parameters de-

298 creased significantly compared with the VGG-Net and the simple U-Net. Thus,
299 the runtime for training the model was shortened.

300 *2.4. Training Models*

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

308 The basic dataset included 650 images from different time-lapse experiments
309 and consisted of under-, over-, and focused images. As a trainset, 416 images
310 (64%) were randomly selected to train the model, and 104 images (16%) were
311 chosen randomly to validate the process to avoid over-fitting. The rest of the
312 130 dataset images (20%) were considered for testing and evaluating the model
313 after training.

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

Network	Run time	Training parameter
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

314 All images were normalised (see the pre-processing step in Sect. 2.2) and
315 resized to 512×512 pixels suitable for inputting the designed neural networks.
316 The optimised hyperparameter values (Tab. 2) correspond to training the most
317 stable CNN models. The ReLU was selected as the activation function for
318 all architecture. The early stopping hyperparameter was employed to avoid
319 over-fitting during the model training. The patient value was considered 30.
320 The batch size was set to the maximal value of eight due to the complexity
321 of the CNN structures and GPU-VRAM limitation. The Adam algorithm was

322 chosen to optimise the neural networks. The learning rate was set to 10^{-3} for
 323 all proposed CNN models. The suitable number of object classes was set as 3
 324 (Sect. 2.2). The best number-of-steps-per-epoch value equals 52 (achieved after
 325 dividing the length of the trainset of value 416 by the batch size of value 8).
 326 The number of epochs when all CNN models converged and were well-trained
 327 was 200.

Table 2: Hyperparameters setting 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

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

334 The Categorical Focal Loss is more efficient for the multi-class classification
 335 of imbalanced datasets, when some classes are classified easily and others are
 336 not. During training progress, the loss function down-weights easy classes and
 337 focuses training on hard-to-classify classes. Thus, the focal loss reduces the loss
 338 value for "well-classified" examples (e.g., roundish sharp cells) and increases
 339 the loss for hard-to-classify objects (e.g., migrated vanish cells) by tuning the
 340 right value of the focusing parameter γ in the categorical focal loss function.
 341 In summary, the categorical focal loss turns the model's attention towards the
 342 difficult-to-classify pixels to achieve more precise classification results.

343 *2.5. Evaluation metrics*

344 All categorical semantic segmentation models were evaluated using the com-
345 mon metrics (Eqs. 1–5). The TP, FP, FN, and TN correspond to the true
346 positive, false positive, false negative, and true negative metric, respectively
347 [41]. The metrics were computed for all test sets in each class and explained as
348 mean values for all classes (Tab. 4).

349 Overall pixel accuracy (Acc) represents a per cent of image pixels belonging
350 to the correctly segmented cells.

$$\text{Acc} = \frac{\text{Pixels Predicted Correctly}}{\text{Total Number of Image Pixels}} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{FN} + \text{TN}} \quad (1)$$

351 Precision (Pre) is a proportion of the cell pixels in the segmentation results
352 that match the GT. The Pre, known as a positive predictive value, is a valuable
353 segmentation performance metric due to its sensitivity to over-segmentation.

$$\text{Pre} = \frac{\text{Correctly Predicted Cell Pixels}}{\text{Total Number of Predicted Cell Pixels}} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (2)$$

354 The Recall (Recl) represents the proportion of cell pixels in the GT correctly
355 identified through the segmentation process. This metric says what proportion
356 of the objects annotated in the GT was captured as a positive prediction.

$$\text{Recl} = \frac{\text{Correctly Predicted Cell Pixels}}{\text{Total Number of Actual Cell Pixels}} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (3)$$

357 The Pre and Recl together give another important metric—F1 score—to eval-
358 uate the segmentation result. The F1-score or Dice similarity coefficient states
359 how the predicted segmented region matches the GT in location and level of
360 details and considers each class's false alarm and missed value. This metric
361 determines the accuracy of the segmentation boundaries [42] and has a higher
362 priority than the Acc.

$$\text{Dice} = \frac{2 \times \text{Pre} \times \text{Recl}}{\text{Pre} + \text{Recl}} = \frac{2 \times \text{TP}}{2 \times \text{TP} + \text{FP} + \text{FN}} \quad (4)$$

363 Another essential evaluation metric for semantic image segmentation is the
364 Jaccard similarity index, known as Intersection over Union (IoU). This metric is
365 a correlation among the prediction and GT [19, 43], and represents the overlap
366 and union area ratio for the predicted and GT segmentation.

$$\text{IoU} = \frac{|y_t \cap y_p|}{|y_t| + |y_p| - |y_t \cap y_p|} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}} \quad (5)$$

367 **3. Results**

368 The models were trained well and converged after running 200 epochs (eval-
369 uated as training/validation loss and Jaccard criterion vs epochs, Fig. 6). The
370 hyperparameter values listed in Table 2 were used to achieve the best train-
371 ing performance and stability. Then, the performances of the trained models
372 were assessed and evaluated using the test datasets and the metrics in Eqs. 1–5
373 (Tab. 4).

374 The computational cost is one of the critical factors in training high-performance
375 models based on the lowest computational resources. The four described meth-
376 ods differ significantly in runtime, the number of trainable parameters, and
377 network structures (Tab. 1). Training the simple U-Net took the longest run-
378 time with the highest number of training parameters. The VGG19-U-Net was
379 trained well in a significantly shorter time due to the network structure; the
380 number of training parameters was slightly lower than in the simple U-Net.
381 The Inception-U-Net runtime was even faster than the previous two methods.
382 This runtime reduction was followed by a further significant decrease in the
383 number of trainable parameters and higher segmentation performance. The
384 last – ResNet34-U-Net method – achieved the shortest computational cost with
385 the best segmentation performance.

386 Figure 7 presents the segmentation results for the U-Net-based models pro-
387 posed in this paper. At the same conditions, the simple U-Net achieved a lower
388 categorical segmentation performance than the other models (when the eval-
389 uation metrics are compared). The simple U-Net was inefficient in classifying

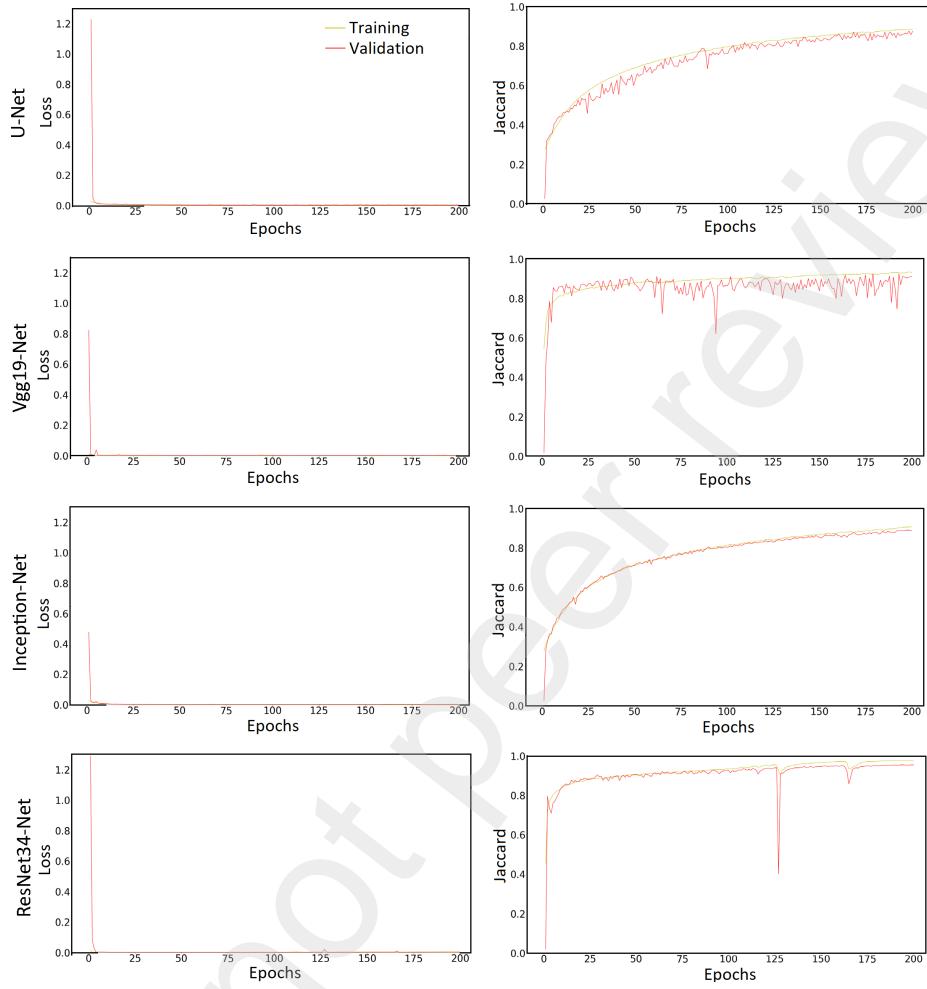


Figure 6: Training/validation plots for the loss criterion (left) and the Jaccard criterion (right) for the simple U-Net (1st row), Vgg19-U-Net (2nd row), Inception-U-Net (3rd row), and ResNet34-U-Net (4th row).

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

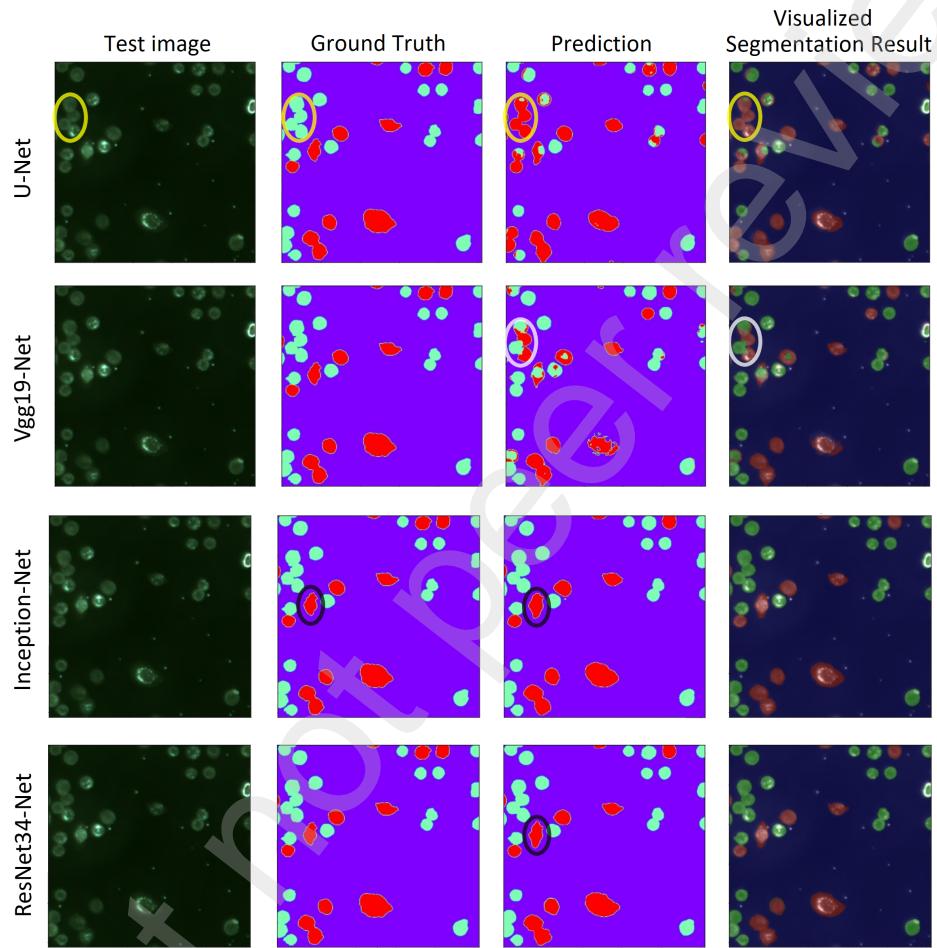


Figure 7: Test image, ground truth, prediction, and 8-bit visualisation of the segmentation results for the U-Net, VGG19-U-Net, Inception-U-Net, and ResNet34-U-Net. The yellow and white circles highlight the wrongly classified and segmented cells. The black circle highlights a different, smoother segmentation result achieved by the ResNet34-U-Net. The image size is 512×512 .

390 the cell pixels into the suitable classes and suffered from wrongly segmented
391 cells into the wrong classes (Fig. 7, yellow circle). Applying the VGG19-U-Net
392 improved the categorical segmentation performance in terms of the evaluation
393 metrics (Tab. 3–4). The cells segmented wrongly by the simple U-Net were
394 improved slightly, but wrong classifications still occurred (Fig. 7, purple cir-
395 cle). The Inception-U-Net was applied to our datasets as the third hybrid CNN
396 method. It leads to significant improvement of the multi-class segmentation
397 results in terms of evaluation metrics (Tab. 3–4). However, this method suf-
398 fers from over-segmentation in all classes (Fig. 7, black circle). The hybrid
399 ResNet34-U-Net was employed to improve further the object segmentation and
400 classification (Tab. 3–4).

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

Table 4: Results for metrics evaluating the U-Net models. Green values represent the
highest segmentation accuracy for the related metric.

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

405 **4. Discussion**

406 The light microscope enables observing living cells in their most natural pos-
407 sible states. However, analysing live cell behaviour in an ordinary light trans-
408 mission (bright-field) microscope over time is difficult for these technical and
409 biological reasons: (1) The cell morphology and position change significantly
410 depending on the life cycle. (2) Illumination conditions are unstable over image
411 and time. (3) The field of view is small to ensure sufficient statistics on cell

412 behaviour. (4) The images of observed cells are insufficiently spatially resolved
413 and distorted by microscope optics. (5) The traditional image processing meth-
414 ods, including machine learning approaches, were sensitive to the number of
415 iterations in the training process, showed mis-segmentation, low computational
416 and runtime performance and recall rate.

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

427 One of the most well-known efficient semantic segmentation methods for mi-
428 croscopy and biomedical images is U-Net [21]. The U-Net consists of encoder
429 and decoder parts with many convolution layers. The encoder part of the net-
430 work was replaced with other different and more effective architecture as the
431 hybrid architecture of the U-Net for more challenging segmentation purposes
432 like categorical segmentation over microscopy images.

433 The microscope and relevant image data used in this study are unique. No
434 similar research on categorical segmentation of light reflection microscopy data
435 has ever been performed before. Thus, comparing the results achieved in this
436 study with the literature is hard. Despite this, the performances of the proposed
437 hybrid U-Net-based models were compared with similar microscopy and medical
438 works (Tab. 5). The first proposed model was based on a simple U-Net structure
439 and achieved the m-IoU score of 0.7062 as the mean value of all classes for
440 categorical segmentation purposes. We assume that a better value of the m-IoU
441 will be achieved after the hyperparameter optimization (Tab. 2).

442 Sugimoto et al. [44] achieved a m-Dice score of 0.799 for multi-class segmen-

443 tation of cancer and non-cancer cells over the medical PD-L1 dataset. Nishimura
 444 et al. [45] applied a U-Net-based weakly supervised method on various mi-
 445 croscopy datasets and reached a m-Dice segmentation score of 0.618 as an av-
 446 erage over all datasets. Piotrowski et al. [26] applied a U-Net-based multi-
 447 class segmentation method over human induced pluripotent stem cell images
 448 and achieved segmentation IoU and Dice accuracy scores of 0.777 and 0.753,
 449 respectively. Long [46] applied the enhanced U-Net (U-Net+) to bright-field,
 450 dark-field, and fluorescence microscopy images and achieved the m-IoU score of
 451 0.567 for single class semantic segmentation.

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 [44]	-	0.799	-
U-Net [26]	0.777	0.753	-
U-Net [45]	-	0.618	-
U-Net+ [46]	0.567	-	-
VGG16-U-Net [47]	-	-	0.961
VGG19-U-Net [48]	-	0.8715	0.8764
Inception-U-Net [49]	-	0.887	-
Inception-U-Net [24]	-	0.95	-
ResNet34-U-Net [50]	0.6915	-	-
SMANet [51]	0.665	0.769	-
DMMN-M3 [52]	0.706 - 0.870	-	-

452 The U-Net encoder part was replaced with the VGG-19 architecture to im-
 453 prove the multi-class segmentation result. The final VGG19-U-Net was op-
 454 timized for our dataset to reduce the number of trainable parameters in the
 455 convolution layers and improve the computational costs and segmentation per-
 456 formance using a dipper network topology and a smaller convolution kernel. In
 457 this way, the categorical segmentation accuracy increased to 0.7178 for the m-
 458 IoU score in the testing phase. Pravitasari et al. [47] applied a VGG16-U-Net

459 with transfer learning to single-class semantic segmentation of brain tumours in
460 magnetic resonance images and achieved an accuracy of 0.961. Nillmani et al.
461 [48] applied a VGG19-U-Net to X-ray images for single-class segmentation of
462 Covid-19 infections and achieved accuracy and Dice scores of 0.8764 and 0.8715,
463 respectively.

464 In the next step, we replaced Google’s inception architecture for the U-Net
465 encoder and made a hybrid Inception-U-Net network. The inception module
466 contained kernels of various sizes in the same layer to make the network topol-
467 ogy wider instead of deeper and extract more representative features. The m-
468 IoU metric for categorical segmentation increased significantly to 0.7907. The
469 number of trainable parameters was reduced. The computational costs were
470 improved efficiently. Haichun et al. [49] proposed an Inception-U-Net for single-
471 class segmentation of brain tumours and achieved the m-Dice score of 0.887 in
472 the testing phase. Sunny et al. [24] applied an Inception-U-Net to categorical
473 segmentation of fluorescence microscopy datasets and achieved the average Dice
474 metric over all segmentation classes of 0.95.

475 The model performance was further improved using a hybrid ResNet34-U-
476 Net architecture. The series of residual blocks with the skip connection was
477 implemented into the CNN architecture during the training process to over-
478 come the vanishing gradient and generalisation ability in very deep neural net-
479 works. It increased the m-IoU to 0.8067 after the multi-class segmentation.
480 Sunny et al. [24] built up a ResNet34-U-Net which showed the m-IoU of 0.6915
481 in the cross-validation phase of fluorescence microscopy multi-class image seg-
482 mentation. Gao et al. [51] applied a selected Multi-Scale Attention Network
483 (SMANet) for multi-class segmentation in pancreatic pathological images and
484 achieved m-Dice and m-IoU scores of 0.769 and 0.665. Ho et al. [52] proposed
485 Multi-Encoder Multi-Decoder Multi-Concatenation (DMMN-M3) deep CNN for
486 multi-class segmentation in two different image sets of breast cancer and reached
487 m-IoU of 0.870 and 0.706.

488 **5. Conclusion**

489 The main objective of this research was to develop an efficient algorithm
490 to detect and segment living human HeLa cells and classify them according
491 to their shapes and life cycles stages. Deep learning approaches to reflected
492 light microscopy data analysis delivered efficient and promising outcomes. This
493 research involved variants of hybrid U-Net-based CNN architecture: a simple
494 U-Net, VGG19-U-Net, Inception-U-Net, and ResNet34-U-net.

495 The simple U-Net (Tab. 1) has the longest training time, the biggest number
496 of trainable parameters, and the lowest categorical segmentation performance.
497 On the other hand, the hybrid ResNet34-U-Net achieved the best categorical
498 segmentation performance (Tab. 4) with a run time significantly lower than the
499 other proposed models. The computational cost and the number of trainable
500 parameters of the inception network are lower than in the U-Net. Thus, the
501 inception networks are better utilisable for bigger datasets. However, running
502 the inception network requires a higher computational GPU memory.

503 The Residual Convolutional Neural Network (ResNet) was applied as a hy-
504 brid with the U-Net to overcome the gradient vanishing and improve the gen-
505 eralisation ability during training. Using a series of residual blocks with skip
506 connection in each level of the ResNet34-U-Net network resulted in better cat-
507 egorical segmentation. The skip connections in each level of the deep neural
508 networks bypass one or more layers and continuously update the gradient val-
509 ues from one or more previous layers into the layers ahead.

510 The categorical segmentation gradually improves from simple U-Net to ResNet34-
511 U-Net (as evaluated using performance metrics, Tab. 4). The ResNet34 encoder
512 network achieved the best categorical segmentation by integrating the residual
513 learning structure to overcome the gradient vanishing with the U-Net as a hy-
514 brid ResNet34-U-Net method. Nevertheless, future works are still essential to
515 expand the knowledge on multi-class semantic segmentation using the weakly
516 supervised method to generate the ground truth for huge datasets independently
517 and apply ensemble learning steps to combine different and efficient CNN ar-

518 chitectures in prediction to achieve the most accurate segmentation result.

519 **FUNDING**

520 This work was supported by the Ministry of Education, Youth and Sports of
521 the Czech Republic (project CENAKVA, LM2018099), from the European Re-
522 gional Development Fund in the frame of the project ImageHeadstart (ATCZ215)
523 in the Interreg V-A Austria–Czech Republic programme, and the project GAJU
524 114/2022/Z.

525 **DECLARATION OF COMPETITITNG INTEREST**

526 The authors declare no conflict of interest, or known competing financial
527 interests, or personal relationships that could have appeared to influence the
528 work reported in this paper.

529 **ACKNOWLEDGEMENT**

530 The authors would like to thank our lab colleagues Šárka Beranová and
531 Pavlína Tláskalová (both from the ICS USB), Jan Procházka (from the USB),
532 and Guillaume Dillenseger (from the FS USB) for their support.

533 **DATA AND CODE AVAILABILITY**

534 The implemented methods and trained models are hosted on the GitHub [53]
535 and other data on the Dryad [30].

536 **References**

- 537 [1] J. R. Tang, N. A. Mat Isa, E. S. Ch'ng, A fuzzy-c-means-clustering ap-
538 proach: Quantifying chromatin pattern of non-neoplastic cervical squamous
539 cells, PLoS ONE 10 (11) (2015) e0142830. doi:10.1371/journal.pone.
540 0142830.

- 541 [2] R. Rojas-Moraleda, W. Xiong, N. Halama, K. Breitkopf-Heinlein, S. Doo-
542 ley, L. Salinas, D. W. Heermann, N. A. Valous, Robust detection and
543 segmentation of cell nuclei in biomedical images based on a computa-
544 tional topology framework, *Med. Image Anal.* 38 (2017) 90–103. doi:
545 10.1016/j.media.2017.02.009.
- 546 [3] Z. Wang, A semi-automatic method for robust and efficient identification
547 of neighboring muscle cells, *Pattern Recogn.* 53 (2016) 300–312. doi:10.
548 1016/j.patcog.2015.12.009.
- 549 [4] F. Buggenthin, C. Marr, M. Schwarzfischer, P. S. Hoppe, O. Hilsenbeck,
550 T. Schroeder, F. J. Theis, An automatic method for robust and fast cell
551 detection in bright field images from high-throughput microscopy, *BMC
552 Bioinform.* 14 (2013) 297. doi:10.1186/1471-2105-14-297.
- 553 [5] S. J. Russell, *Artificial Intelligence: A Modern Approach*, Third Edition,
554 Prentice Hall, 2010.
- 555 [6] X. Huang, C. Li, M. Shen, K. Shirahama, J. Nyffeler, M. Leist, M. Grze-
556 gorzek, O. Deussen, Stem cell microscopic image segmentation using super-
557 vised normalized cuts, in: *IEEE Int. Conf. Image Processing (ICIP)*, 2016,
558 pp. 4140–4144. doi:10.1109/ICIP.2016.7533139.
- 559 [7] S. A. Mah, R. Avci, P. Du, J.-M. Vanderwinden, L. K. Cheng, Super-
560 vised machine learning segmentation and quantification of gastric pace-
561 maker cells, in: *42nd Ann. Int. Conf Proc. IEEE Eng. Med. Biol. Soc.*,
562 2020, pp. 1408–1411. doi:10.1109/EMBC44109.2020.9176445.
- 563 [8] T. Tikkanen, P. Ruusuvuori, L. Latonen, H. Huttunen, Training based
564 cell detection from bright-field microscope images, in: *9th International
565 Symposium on Image and Signal Processing and Analysis (ISPA)*, 2015,
566 pp. 160–164. doi:10.1109/ISPA.2015.7306051.
- 567 [9] K. Liimatainen, P. Ruusuvuori, L. Latonen, H. Huttunen, Supervised
568 method for cell counting from bright field focus stacks, in: *Proc. IEEE*

- 569 13th Int. Symp. Biom. Imaging, 2016, pp. 391–394. doi:10.1109/ISBI.
570 2016.7493290.
- 571 [10] G. Hinton, T. Sejnowski, Unsupervised Learning: Foundations of Neural
572 Computation, MIT Press, MIT Press, 1999.
- 573 [11] B. Antal, B. Remenyik, A. Hajdu, An unsupervised ensemble-based markov
574 random field approach to microscope cell image segmentation, in: 2013 Interna-
575 tional Conference on Signal Processing and Multimedia Applications
576 (SIGMAP), 2013, pp. 94–99. doi:10.5220/0004612900940099.
- 577 [12] F. Mualla, S. Schöll, B. Sommerfeldt, A. Maier, S. Steidl, R. Buchholz,
578 J. Hornegger, Unsupervised unstained cell detection by SIFT keypoint clus-
579 tering and self-labeling algorithm, in: P. Golland, N. Hata, C. Barillot,
580 J. Hornegger, R. Howe (Eds.), Med. Image Comput. Comput. Assist. In-
581 terv. 2014. Lect. Notes Comput. Sci., Vol. 8675, Springer International Pub-
582 lishing, Cham, 2014, pp. 377–384. doi:10.1007/978-3-319-10443-0_48.
- 583 [13] X. Zheng, Y. Wang, G. Wang, J. Liu, Fast and robust segmentation of
584 white blood cell images by self-supervised learning, Micron 107 (2018) 55–
585 71. doi:10.1016/j.micron.2018.01.010.
- 586 [14] M. D. Zeiler, R. Fergus, Visualizing and understanding convolutional neural
587 networks, ECCV: Computer Vision – ECCV 2014, Lect. Notes Comp. Sci.
588 8689 (2014) 818–833. doi:10.1007/978-3-319-10590-1_53.
- 589 [15] S. Lin, N. Norouzi, An effective deep learning framework for cell segmenta-
590 tion in microscopy images, in: Ann. Int. Conf. Proc. IEEE Eng. Med. Biol.
591 Soc., 2021, pp. 3201–3204. doi:10.1109/EMBC46164.2021.9629863.
- 592 [16] N. Kumar, R. Verma, D. Anand, A. Sethi, Multi-organ nuclei segmentation
593 challenge, Retrieved on 05/05/2021 (2018).
594 URL <https://monuseg.grandchallenge.org/>
- 595 [17] J. C. Caicedo, A. Goodman, K. W. Karhohs, B. A. Cimini, J. Ackerman,
596 M. Haghghi, C. Heng, T. Becker, M. Doan, C. McQuin, M. Rohban,

- 597 S. Singh, A. E. Carpenter, Broad bioimage benchmark collection, Retrieved
598 on 05/05/2021 (2018).
- 599 URL <https://bbbc.broadinstitute.org/BBBC038>
- 600 [18] P. Thi Le, T. Pham, Y.-C. Hsu, J.-C. Wang, Convolutional blur attention
601 network for cell nuclei segmentation, Sensors 22 (4) (2022) 1586. doi:
602 [10.3390/s22041586](https://doi.org/10.3390/s22041586).
- 603 [19] J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for seman-
604 tic segmentation, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognit.,
605 2015, pp. 3431–3440. doi:[10.1109/CVPR.2015.7298965](https://doi.org/10.1109/CVPR.2015.7298965).
- 606 [20] A. Ben-Cohen, I. Diamant, E. Klang, M. Amitai, H. Greenspan, Fully
607 convolutional network for liver segmentation and lesions detection in deep
608 learning and data labeling for medical applications, in: G. Carneiro, D. Ma-
609 teus, L. Peter (Eds.), Deep Learning and Data Labeling for Medical Appli-
610 cations DLMIA 2016, LABELS 2016. Lect. Notes Comp. Sci., Vol. 10008,
611 Springer, Cham, 2016, pp. 77–85. doi:[10.1007/978-3-319-46976-8_9](https://doi.org/10.1007/978-3-319-46976-8_9).
- 612 [21] O. Ronneberger, P. Fischer, T. Brox, U-Net: Convolutional networks for
613 biomedical image segmentation, in: N. Navab, J. Hornegger, W. Wells,
614 A. Frangi (Eds.), Med. Image Comput. Comput. Assist. Interv. Lect. Notes
615 Comp. Sci., Vol. 9321, Springer, Cham, 2015, pp. 234–241. doi:[10.1007/978-3-319-24574-4_28](https://doi.org/10.1007/978-3-319-24574-4_28).
- 617 [22] E. Shibuya, K. Hotta, Cell image segmentation by using feedback and
618 convolutional LSTM, Vis. Comput. 38 (2021) 3791–3801. doi:[10.1007/s00371-021-02221-3](https://doi.org/10.1007/s00371-021-02221-3).
- 620 [23] A. Ghaznavi, R. Rychtáriková, M. Saberioon, D. Štys, Cell segmentation
621 from telecentric bright-field transmitted light microscopy images using a
622 residual attention U-net: A case study on HeLa line, Comp. Biol. Med.
623 147 (147) (2022) 105805. doi:[10.1016/j.combiomed.2022.105805](https://doi.org/10.1016/j.combiomed.2022.105805).

- 624 [24] S. P. Sunny, A. I. Khan, M. Rangarajan, A. Hariharan, P. Birur N,
625 H. J. Pandya, N. Shah, M. A. Kuriakose, A. Suresh, Oral epithelial
626 cell segmentation from fluorescent multichannel cytology images using
627 deep learning, Comput. Methods Programs Biomed. 227 (2022) 107205.
628 doi:10.1016/j.cmpb.2022.107205.
- 629 [25] M. E. Bakir, v. H. Yalim Keles, Deep learning based cell segmenta-
630 tion using cascaded U-Net models, in: 2021 29th Signal Processing and
631 Communications Applications Conference (SIU), 2021, pp. 1–4. doi:
632 10.1109/SIU53274.2021.9477937.
- 633 [26] T. Piotrowski, O. Rippel, A. Elanzew, B. Nießing, S. Stucken, S. Jung,
634 N. König, S. Haupt, L. Stappert, O. Brüstle, R. Schmitt, S. Jonas, Deep-
635 learning-based multi-class segmentation for automated, non-invasive rou-
636 tine assessment of human pluripotent stem cell culture status, Comp. Biol.
637 Med. 129 (2021) 104172. doi:10.1016/j.compbio-med.2020.104172.
- 638 [27] G. Platonova, D. Štys, P. Souček, K. Lonhus, J. Valenta, R. Rychtáriková,
639 Spectroscopic approach to correction and visualization of bright-field light
640 transmission microscopy biological data, Photonics 8 (8) (2021) 333. doi:
641 10.3390/photonics8080333.
- 642 [28] D. Štys, T. Náhlík, P. Macháček, R. Rychtáriková, M. Saberioon, Least in-
643 formation loss (LIL) conversion of digital images and lessons learned for sci-
644 entific image inspection, in: F. Ortuno, I. Rojas (Eds.), Bioinformatics and
645 Biomedical Engineering. IWBBIO 2016. Lect. Notes Comp. Sci., Vol. 9656,
646 Springer, Cham, 2016, pp. 527–536. doi:10.1007/978-3-319-31744-1_
647 47.
- 648 [29] A. Buades, B. Coll, J.-M. Morel, A non-local algorithm for image denoising,
649 in: Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., Vol. 2,
650 2005, pp. 60–65. doi:10.1109/CVPR.2005.38.
- 651 [30] A. Ghaznavi, R. Rychtáriková, P. Císař, M. Ziae, D. Štys, Telecentric
652 bright-field reflected light microscopic dataset (2023).

- 653 URL [https://datadryad.org/stash/share/
654 D19v1HCHpsvAos7DPFZaJ0AvyMP80ZjiskRruodzwKs](https://datadryad.org/stash/share/D19v1HCHpsvAos7DPFZaJ0AvyMP80ZjiskRruodzwKs)
- 655 [31] Zeiss, APPEAR – automated image analysis, Retrieved on 12/12/2021.
- 656 URL <https://www.apeer.com/>
- 657 [32] Y. Lu, A.-A. Liu, Y.-T. Su, Chapter 6 - mitosis detection in biomedical
658 images, in: M. Chen (Ed.), Computer Vision for Microscopy Image Analy-
659 sis, Computer Vision and Pattern Recognition, Academic Press, 2021, pp.
660 131–157. doi:10.1016/B978-0-12-814972-0.00006-0.
- 661 [33] A. Krizhevsky, I. Sutskever, G. E. Hinton, Imagenet classification with
662 deep convolutional neural networks, Commun. ACM 60 (6) (2017) 84–90.
663 doi:10.1145/3065386.
- 664 [34] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-
665 scale image recognition, in: The 3rd International Conference on Learning
666 Representations (ICLR2015), 2015, pp. 1–14. doi:10.48550/arXiv.1409.
667 1556.
- 668 [35] W. A. Hamwi, M. M. Almustafa, Development and integration of VGG
669 and dense transfer-learning systems supported with diverse lung images
670 for discovery of the Coronavirus identity, Inform. Med. Unlocked 32 (2022)
671 101004. doi:10.1016/j imu.2022.101004.
- 672 [36] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan,
673 V. Vanhoucke, A. Rabinovich, Going deeper with convolutions, in: Proc.
674 IEEE Conf. Comput. Vis. Pattern Recognit., 2015, pp. 1–9. doi:10.1109/
675 CVPR.2015.7298594.
- 676 [37] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recog-
677 nition, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2016, pp.
678 770–778. doi:10.1109/CVPR.2016.90.

- 679 [38] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, Tensor-
680 Flow: large-scale machine learning on heterogeneous distributed systems,
681 in: Proc. USENIX Symp. Oper. Syst. Implement., 2016, pp. 265–283.
- 682 [39] Colab, Google Research, Colaboratory, Retrieved on 12/12/2021.
683 URL https://colab.research.google.com/?utm_source=scs-index
- 684 [40] T. Lin, P. Goyal, R. Girshick, K. He, P. Dollár, Focal loss for dense object
685 detection, IEEE Trans. Pattern Anal. Mach. Intell. 42 (2) (2020) 318–327.
686 doi:[10.1109/TPAMI.2018.2858826](https://doi.org/10.1109/TPAMI.2018.2858826).
- 687 [41] X. Pan, L. Li, H. Yang, Z. Liu, J. Yang, Y. Fan, Accurate segmentation
688 of nuclei in pathological images via sparse reconstruction and deep con-
689 volutional networks, Neurocomputing 229 (2017) 88–99. doi:[10.1016/j.neucom.2016.08.103](https://doi.org/10.1016/j.neucom.2016.08.103).
- 690 [42] G. Csurka, D. Larlus, F. Perronnin, What is a good evaluation measure
691 for semantic segmentation?, in: Proceedings of the British Machine Vision
692 Conference, BMVA Press, 2013, pp. 32.1–32.11. doi:[10.5244/C.27.32](https://doi.org/10.5244/C.27.32).
- 693 [43] B. Vijay, A. Kendall, R. Cipolla, SegNet: A deep convolutional encoder-
694 decoder architecture for image segmentation, IEEE Trans. Pattern Anal.
695 Mach. Intell. 39 (12) (2015) 228–233. doi:[10.1109/TPAMI.2016.2644615](https://doi.org/10.1109/TPAMI.2016.2644615).
- 696 [44] T. Sugimoto, H. Ito, Y. Teramoto, A. Yoshizawa, R. Bise, Multi-class cell
697 detection using modified self-attention, in: IEEE/CVF Conference on Com-
698 puter Vision and Pattern Recognition Workshops (CVPRW), 2022, pp.
699 1854–1862. doi:[10.1109/CVPRW56347.2022.00202](https://doi.org/10.1109/CVPRW56347.2022.00202).
- 700 [45] K. Nishimura, C. Wang, K. Watanabe, D. F. E. Ker, R. Bise, Weakly
701 supervised cell instance segmentation under various conditions, Med. Image
702 Anal. 73 (2021) 102182. doi:[10.1016/j.media.2021.102182](https://doi.org/10.1016/j.media.2021.102182).
- 703 [46] F. Long, Microscopy cell nuclei segmentation with enhanced U-Net, BMC
704 Bioinform. 21 (2020) 8. doi:[10.1186/s12859-019-3332-1](https://doi.org/10.1186/s12859-019-3332-1).

- 706 [47] A. A. Pravitasari, N. Iriawan, M. Almuhayar, T. Azmi, I. Irhamah,
707 K. Fithriasari, S. W. Purnami, W. Ferriastuti, UNet-VGG16 with transfer
708 learning for mri-based brain tumor segmentation, TELKOMNIKA 18 (3)
709 (2020) 1310–1318. doi:10.12928/telkomnika.v18i3.14753.
- 710 [48] Nillmani, N. Sharma, L. Saba, N. N. Khanna, M. K. Kalra, M. M. Fouda,
711 J. S. Suri, Segmentation-based classification deep learning model embedded
712 with explainable AI for COVID-19 detection in chest X-ray scans, Diag-
713 nóstics 12 (9) (2022) 2132. doi:10.3390/diagnostics12092132.
- 714 [49] H. Li, A. Li, M. Wang, A novel end-to-end brain tumor segmentation
715 method using improved fully convolutional networks, Comp. Biol. Med.
716 108 (2019) 150–160. doi:10.1016/j.combiomed.2019.03.014.
- 717 [50] G. Patel, H. Tekchandani, S. Verma, Cellular segmentation of bright-field
718 absorbance images using residual U-Net, in: 2019 International Conference
719 on Advances in Computing, Communication and Control (ICAC3), 2019,
720 pp. 1–5. doi:10.1109/ICAC347590.2019.9036737.
- 721 [51] E. Gao, H. Jiang, Z. Zhou, C. Yang, M. Chen, W. Zhu, F. Shi, X. Chen,
722 J. Zheng, Y. Bian, D. Xiang, Automatic multi-tissue segmentation in
723 pancreatic pathological images with selected multi-scale attention net-
724 work, Comp. Biol. Medicine 151 (Pt A) (2022) 106228. doi:10.1016/
725 j.combiomed.2022.106228.
- 726 [52] D. J. Ho, D. V. Yarlagadda, T. M. D’Alfonso, M. G. Hanna, A. Graben-
727 stetter, P. Ntiamoah, E. Brogi, L. K. Tan, T. J. Fuchs, Deep Multi-
728 Magnification Networks for multi-class breast cancer image segmenta-
729 tion, Comput. Med. Imaging Graph. 88 (2021) 101866. doi:10.1016/j.
730 compmedimag.2021.101866.
- 731 [53] A. Ghaznavi, Github repository (2023).
- 732 URL <https://github.com/AliGhaznavi1986/>
- 733 Hybrid-CNNs-for-multi-class-segmentation