Panoptic Segmentation with Multi-Modal Dataset Using an Improved Network Model

Koppagiri Jyothsna Devi[1], Gouranga Mandal[2,*]

[1,2]School of Computer Science and Engineering, VIT-AP University, Andhra Pradesh, India.

Email id: Jyothsnakoppagiri1302@gmail.com; gourangamandal@yahoo.com

Abstract

For biomedical image analysis, instance segmentation is crucial. It is still difficult because of the intricate backdrop elements, the significant variation in object appearances, the large number of overlapping items, and the hazy object borders. Deep learning-based techniques, which may be separated into proposal-free and proposal-based approaches, have been frequently employed recently to overcome these challenges. The existing approaches experience information loss due to their concentration on either local-level instance features or global-level semantics. To solve this problem, this work proposes an improved dense Net ($ID-Net$) that mixes instance and semantic data. The suggested $ID-Net$ promotes the acquisition of semantic contextual information by the instance branch by linking instance prediction and semantic features via a residual attention feature integration strategy. The confidence score of each item is then matched with the accuracy of the prediction using a dense quality sub-branch that is created. A consistency regularisation technique is also proposed for the robust learning of segmentation for instance branches and the semantic segments tasks. By proving its utility, the proposed $ID-Net$ outperforms prevailing approaches on various biomedical datasets.

Keywords: panoptic segmentation; multi-modal; prediction: semantic features; instance

1. Introduction

Medical image processing requires the stage of instance segmentation, which splits each item within the same class while simultaneously assigning each pixel a class label [1]. The form, geographical placements, and distribution of each thing may be further investigated to study the prediction activities from the provided photographs by giving each object a distinct ID. The positioning of cancerous nuclei in space aids in comprehending cancer prognostic predictions in digital pathology; the tumour and cancer grading ranges from [2] depending on the size and form of the nucleus. The secret to comprehending how plants work and how they grow in the field of agriculture and horticulture is the ability of experts to identify each image and learn the image; this details the disease, maturity stage and associated cultivars [3]. Due to its labor-intensive nature and length, traditional manual segmentation evaluation for biomedical image examples, the existing practice needs to be improved [4]. Furthermore, because of the intra- and inter-observer heterogeneity, limits in objectivity and repeatability are inherent [5]. Consequently, segmentation in biology photos is a highly desired and required automated and precise procedure [6].

Instance segmentation tasks for biomedical images still present considerable difficulties [7]. First, specific background structures, such as cytoplasm or stroma in histopathology photos, resemble the foreground item in appearance. Methods that rely on thresholding are, therefore, unsuccessful. Second, there is a lot of variation in the objects' size, shape, texture, and intensity among the photos in the same dataset [8]. The varied structures and activities bring it on when obtaining distinct images. Third, there are groups of things that are overlapping one another. Due to uneven dye absorption and comparable item intensities, the lines separating these contacting objects need to be clarified [9]. It might lead to the segmentation of many things into a single one. Deep learning-based approaches are common and successfully address these problems by learning from feature representations [10].

Doi : https://doi.org/10.54216/JCIM.140207
Received: January 17, 2024 Revised: March 03, 2024 Accepted: July 03, 2024

Approaches for segmenting instances using CNN may be divided into proposals-free and proposals-based methods into two categories. In the strategies for proposal-free instance segmentation, a class label is initially given to each pixel using a semantic segmentation model [11]. The next stage segregates each foreground item within the same category using post-processing techniques based on morphological characteristics, architectural design, and spatial planning. Even after processing can separate the relevant components, during overlapping object segmentation, these approaches still encounter erroneous boundaries [12]. Even though concentration to identify the touching objects based on global contextual data is still based on the semantic segmentation step's limits lacking, mainly when their borders are unclear. The segmentation approaches combine the detection and segmentation [13]. A bounding box is first used to determine the spatial position of each object. After that, each item inside each expected bounding box is segmented using a generator [14]. The proposal-based methods can distinguish between the contacting items by individually detecting and segmenting each object. They have certain limitations, though, because it is difficult to discern between the foreground and background due to an absence of global semantic data [15].

The global semantic information, as well as the local instance information, is crucial for the segmenting of instance duties. The global semantic information shows the scene context's contextual hints, showing how the foreground and background are related and how the foreground items are distributed spatially [16]. Despite this, each item's precise location and contour are described by local-level instance information. Segmenting instances using semantics have been combined to form panoptic segmentation to combine the advantages of global and local characteristics. In [17], the panoptic level segmentation is examined using predictions from two semantic and instance segmentation branches that were separately trained [18]. A substantial computational cost is associated with training if components are not shared across the two branches. Additionally, introducing a network for both tasks is possible, according to research in [19] concurrently is superior to training it separately. It is suggested to prepare the instance segmentation and semantic branching simultaneously using the same U-Net backbone to achieve this. It has maintained memory efficiency and achieved cutting-edge performance on panoptic segmentation [20]. When segmenting the nuclei in histopathological images, we previously presented. To evaluate the worldwide and regional data included in histopathological images, this work developed the pre-trained network with improved dense layers, which was inspired by our analysis of the tasks for segmentation by semantics and instance.

According to our proposed model, the two branches might be trained simultaneously using the same backbone model, in contrast to [21] two independently optimized branches. Some existing work suggested pushing to facilitate contextual learning at the semantic level in the model for segmenting scenarios; the instance branch allows for direct learning about semantic-level properties. In the following sections, the author in [22], a work by U-Net that is currently referred to as U-Net. First, we present a fresh prediction for semantic segmentation using the U-Net example branch. The quality maps are then combined using a feature integration technique and prompt the instance segmentation branch's decoder to learn semantic features by fusing the predictions, for instance, and semantic branches were hidden. For example, segmentation can benefit from using a dual-model generator to reduce information loss. By directly integrating the features from the two components, the Panoptic FPN could not further encourage learning semantic features in the instance branch since it could only use a single backbone. The semantic and instance segmentation branches will be jointly optimized as our U-Net was able to achieve.

In this paper, a Panoptic Feature Integration Net ($ID-Net$) is introduced to address several unresolved problems from our first; as a result, the U-Net is enhanced. First, the trait integration process in [23] immediately swaps out the features from the generator's output for those in the feature map's semantic segmentation. The global signals from the segmentation prediction are still significant even if the instances branch's predictions interpret instance-level data than semantic [24]. This study suggests replacing the current or residual attention feature integration technique. The incident branch's local characteristics are combined with the global semantic features in our recently proposed without any semantic-level features being deprecated [25]. In the general design, the collaborative optimization of two semantic segmentation problems using the same real-world data to ensure semantic coherence, we add a regularization step to the two segmentation tasks to impose as many similarities as possible to make two semantic forecasts originating from two distinct branches, making it easier to learn the two segmentation tasks well. Furthermore, as indicated, specific low-quality predictions still have excellent classification scores when using the conventional network. The segmentation accuracy will suffer by considering these incompetently created outcomes as the most confident ones. To achieve this, in this work, we offer an exceptional quality sub-branch that develops a secondary quality score relies on Intersection-over-Union (IoU) score and Dice score for each forecast. The matching grade for quality is multiplied by the classification score for each network to make adjustments during inference. The $ID-Net$ suggested in this work might be referred to as improved dense because it is a development of U-Net. We are the first to use

Doi : https://doi.org/10.54216/JCIM.140207
Received: January 17, 2024 Revised: March 03, 2024 Accepted: July 03, 2024

panoptic segmentation utilizing our prior U-Net models to evaluate photos from the biomedical fields. Following is a summary of this work's overall contributions concerning $ID - Net$:

- To combine the attributes of each item that are known at the instance-level semantics, this work develops a method for fusing features with an improved dense Net ($ID - Net$).
- To regularize robust semantic segmentation issues training, this work provides a consistent segmentation technique.
- Verify that the anticipated segmentation quality of each image matches its performance score; we included a new sub-branch for quality.
- Using a variety of biomedical datasets such as images from histopathology and fluorescence microscopy on instance segmentation tasks, our proposed $ID - Net$ is tested. When compared to cutting-edge methods, our findings consistently outperform them for all measures.

The work is structured as follows: Section 2 provides a broader analysis of various prevailing approaches. The proposed methodology is discussed in section 3, with numerical outcomes in section 4. The summary is provided in section 5.

2. Related works

Panoptic segmentation aims to distinguish between the things and stuff in a scene by combining instance and semantic segmentation. Panoptic segmentation divides objects into junk and things [26]. The uncountable areas, including the sky, the pavements, and the grounds, are referred to as stuff. All countable items, such as automobiles, people, and groups of three, are considered things. In instance and semantic segmentation, we can identify overlap among objects of similar kind, the panoptic method segments things and stuff by assigning each of them a unique hue that sets it apart from the rest [27]. Panoptic segmentation further enables effective visualization of diverse scene components and may be described as a general approach that includes detecting, localizing, and classifying distinct scene components. This results in a clear and valuable comprehension of the scenario [28].

The capacity of panoptic segmentation approaches to define an image's scene content and allow for its thorough comprehension dramatically facilitates the analysis, enhances performance, and offers answers to many computer vision issues [29]. These include video surveillance, self-driving cars and medical image analysis. Panoptic segmentation enables the study of specific targets without looking at all of the image's regions. It shortens computation time, reduces the likelihood that some objects will be mistakenly identified or missed, and establishes the edge saliency of various areas of an image. An illustration of the chronological segmentation initiating with object segmentation and finishing with panoptic segmentation [30], to help study the development of panoptic segmentation about the associated activities carried out on things and stuff. Typically, the well-known networks utilized to complete each training have also been emphasized.

The combinatorial perspective of "things" and "stuff" is made possible by panoptic segmentation, a breakthrough in computer vision. As a result, it indicates a fresh approach to image segmentation. The literature has suggested several panoptic segmentation investigations, presented and in-depth explored in this part to enlighten the latest technology. Segmentation by instance and semantics are used independently by specific panoptic segmentation algorithms before the results are combined or aggregated to produce the panoptic segmentation. To exploit the shared backbone, other network elements use the features created by the backbone. Different frameworks have applied the same strategies while explicitly connecting instance and semantic networks. Most panoptic segmentation frameworks described worked with RGB photos, but some applied their techniques to LIDAR data and medical images. The author discusses the frameworks based on data used. Many frameworks have been developed to segment an image using the panoptic technique utilizing instance and semantic segmentation before concatenating component's outcomes to produce the panoptic segmentation findings [30]. The author initially carried out model and semantic segmentation separately to obtain panoptic segmentation. The panoptic quality (PQ) metric is then developed using the non-maximum suppression (NMS) method. As a result, the semantic segmentation of the image and the NMS-like technique are used to create the non-overlapping instance segments.

3. Methodology

This section provides a wider analysis on the anticipated model with multi-modal dataset. The evaluation is done with diverse performance metrics and evaluated with other approaches.

Doi : https://doi.org/10.54216/JCIM.140207
Received: January 17, 2024 Revised: March 03, 2024 Accepted: July 03, 2024

3.1. Dataset

TCGA: This dataset, collected from TCGA, the Cancer Genome Atlas, magnified of 40, has 30 histopathological images with a size of $1000 * 1000$. Each image represents the seven organs: the liver, stomach, kidney, colon, prostate, and breast. We have the same data split so that we may compare it to cutting-edge methods. Three photos from each organ combine to make up 12 images used for training. 20 patches with $256 * 256$ are chosen during training from images. Then, basic augmentation techniques are applied, such as rotations of 90, 180, and 270 degrees in both the directions. Then, complex augmentation is used to guarantee resilience to noise and colour fluctuation in histology images, Gaussian noise, and Gaussian noise with median blur. The validation set consists of four images of the prostate, kidneys, liver, and breasts. The following 14 images are the unseen testing set comprised of 6 photographs from the remaining 3 organs that weren't included. In contrast, the visible testing set consists of 8 photos from the identical four organs similar to those in the practice set. Each 1000x1000 image used in testing is used directly for segmenting nucleus instances.

TNBC: Histopathology dataset is for Triple Negative Breast Cancer (TNBC), which was published. The TNBC collection includes 11 distinct Curie Institute patients' 30 512×512 histopathology photos at a magnification 40. On this dataset, we do 3-fold cross-validation for each experiment. Following data augmentation using features such as flipping the data horizontally and vertically, applying blurring using Gaussian, median, and noise, and rotating it by 90, 180, and 270 degrees, 5 256 x 256 patches are clipped from each 512 x 512 image during training. Each 512512 image is used directly during testing.

Fluorescence Microscopy Images: We validate our $ID - Net$ by analyzing the photographs taken using a fluorescent microscope and the histology images. We use the BBBC039V1 dataset, which consists of 200 520 × 696 fluorescent microscopy images. Cell shape and density variations are observed utilizing the DNA channel's single field of vision. Each image concentrates on the U2OS cells. According to the data split, 50 photos are used for validation, 50 for testing, and 100 for training (available at https://data.broadinstitute.org/bbbc/BBBC039). The initial step in preparing training data is randomly selecting ten 256 ×256 patches from each image. Only minimal data augmentation is used in this dataset, including 90°, 180°, and 270° rotations concerning both the horizontal and vertical axes. This is because this dataset's background components are less complicated than those in the others. Each 520 x 696 image is utilized immediately during inference.

3.2. Prediction

Many modifications is experimented over traditional U-Net design to enhance promote segmentations were swapped out with bigger 5×5 convolution operators. In our experience, utilizing 5×5 convolutions when using U-Net for image segmentation tasks led to superior results. A 10% dropout layer was added after each convolution. Dropout layers aid in network regularization and prevent over-fitting. Although padding was not necessary for the original U-Net design while employing convolution operators, this work uses zero padding to fulfil the input feature map's size was appropriate because the final feature map was the same. The conventional U-Net's last layer contraction path was a 2D layer with 1024 feature maps and a 32 by 32 dimension. For the initial contraction route layer, this work used 512×512 input size with 32 filters and 64×64 2D size at each maximum pool and doubled the feature maps until extracting 256 feature maps. This work refers explicitly to "max pool" as the traditional convolutional neural network (CNN) operation is employed to input down-sampling and "feature maps" as a convolution layer output to emphasize how input are transformed to outcomes with hidden features. Typically, the input's rectangular non-overlapping sub-regions are only used up to their maximum value. For instance, the output dimensions are lowered by the factor of two in height and breadth if we choose the maximum from each of the 22 input sub-regions to draw conclusions and achieve high precision in driving quickly and augmented reality scenarios; $ID - Net$ was developed. The $ID - Net$ design used three convolutional layers of leftover network construction materials. These included a typical convolutional layer, 1×1 expansion, batch normalization and 1×1 projection that decreased dimensionality. To create an image segmentation network in the encoder/decoder paradigm, $ID - Net$ used a variety of convolutions. $ID - Net$ contained asymmetric convolutions in some layers, distinguished by 5×1 and 1×5 convolution sequences that can be split apart. The comparable asymmetric convolution, used to minimize the network size, contained 10 parameters instead of 25 in the 5×5 convolution. The $ID - Net$ have multiple bottleneck layer versions and a single starting block. A layer that has been continually down-sampled to make a "bottleneck" is the point at which a network learns the most important input data. By doing this, the network can weed out pointless data from the incoming stream.

Doi : https://doi.org/10.54216/JCIM.140207
Received: January 17, 2024 Revised: March 03, 2024 Accepted: July 03, 2024

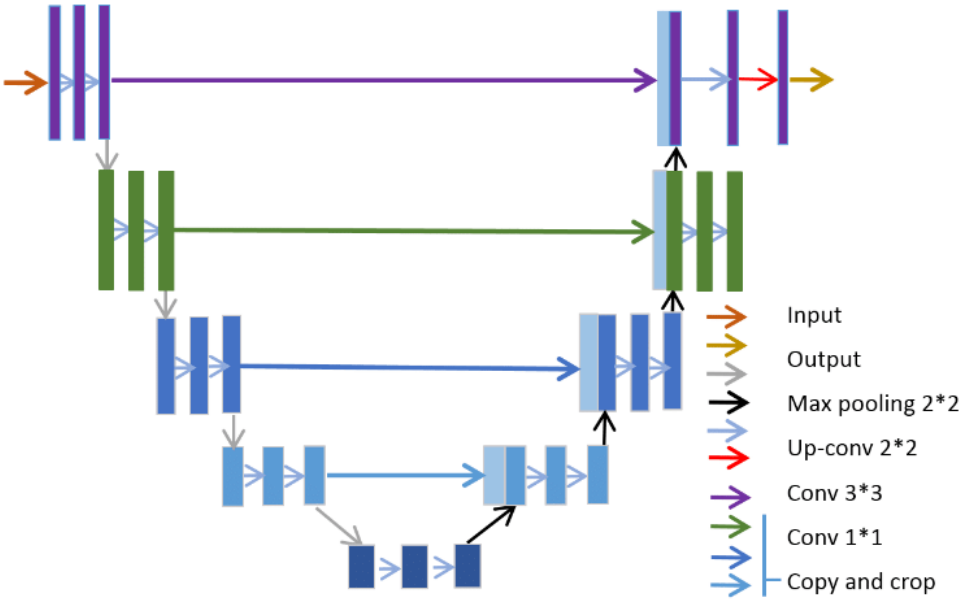


Figure 1: Improved UNet architecture

3.2. Loss function analysis

Diverse loss functions are discussed in the literature. The overlap between anticipated and actual segmentation is measured by DSC which is frequently employed as a loss function specifically for medical image segmentation. It calculates and equalizes harmonic means of FNs and FPs to put it another way. The Tversky loss was proposed to modify FPs and FNs weight-based on Tversky index:

$$TI(\propto, \beta, P, G) = \frac{|P \cap G|}{\alpha \left|\frac{P}{G}\right| + \beta \left|\frac{G}{P}\right| + |P \cap G|} \quad (1)$$

When P and G represent the sets of expected and actual labels, α and β determine the size of the penalties for false positives and false negatives, respectively. G on P has a relative complement known as $\frac{P}{G}$. This is how the Tversky index is used to define the Tversky loss:

$$TI(\propto * \beta) = \frac{\sum_{i=1}^{N} p_{0i} g_{oi}}{\sum_{i=1}^{N} p_{0i} g_{oi} + \propto \sum_{i=1}^{N} p_{0i} g_{1oi} + \beta \sum_{i=1}^{N} p_{10i} g_{oi}} \quad (2)$$

Where p_{0i} represents likelihood that voxel i is a component of the cancer image and p_{10i} is the likelihood, i.e. background component of the soft-max layer, which creates the network's top layer. The designation g_{oi} for ground truth training also stands for cancer image is 1, and the label g_{1oi} (for everything else) is 0, respectively. The trade-off between FPs and FNs can be managed by altering the parameters α and β. Setting $\alpha = \beta = 0.5$ results in the well-known DSC; however, setting $\alpha + \beta = 1$ results in a set of F_β scores; $\beta's$ more significant than 0.5 weight recall more than accuracy by emphasizing slices with small foreground areas have FNs.

3.3. Training process

Training, validation, and testing sets are usually separated into three dataset regions in machine learning and deep learning methodologies. The holding-out set sometimes called the testing set, is kept separate throughout training for use exclusively in reporting final findings, also known as the hold-out set. Using $k-$fold cross-validation may also be employed if there is just a little similar to how medical image processing works and large data accessible for the training phase. $K-$folds have been used to split the available data. The remaining folds are then joined to form the validation set, one of the folds, with the training set being the other. The verification group is each fold for further iterations of this procedure, while the different sets serve as the training set. Results might be more robust if done in this way. This work used 5-CV technique by partitioning the entire dataset to 5

Doi : https://doi.org/10.54216/JCIM.140207
Received: January 17, 2024 Revised: March 03, 2024 Accepted: July 03, 2024

folds randomly of 8 patients due to the low available availability for training stage (32 patients). By integrating 4 and 5-CV to training set and maintains patients as the validation fold, we could train five models for all network. Individual slices were used in the training fold from each research and fed into our models because all network models under consideration are 2D models. Results from the five validation folds were averaged to see how well the k-fold CV performed. For 10 samples (the testing set) that weren't utilized for k −fold CV procedure were then employed to create the performance metrics.

Data enhancement is a popular method for reducing over-fitting when training neural network models, especially when training data is lacking. The input training slices were also subjected to random rotation, translation in the shearing, horizontal flipping, zooming, and x and y dimensions. This work employed a total of 6 distinct data augmentation methods. Each study's whole set of image slices was utilized, just like in the training phase. Data normalization or standardization is frequently carried out as a first stage of machine learning. This minimizes numerical instability and speeds up model convergence by preventing the weights from growing too big. Using all the training data, we determine the average and range of each fold in 2D pixels. This work divided and subtracted the mean by the previously determined standard deviation before being sent into the training pipeline.

We used a first set of 16 patient investigations to find optimal learning rates empirically. With the Adam optimizer, 0.0001 learning rates were used for $ID - Net$. The batch size used for all tests was eight slices, and the Tversky loss functions were set to $\alpha = 0.3$ and $\beta = 0.7$. We trained network for 100 epochs during segmentation. Our decision to stop was based on the training lost. We halted after 10 epochs if the training loss had not decreased steadily. Lastly, we instructed the networks and performed inference on an 8 GB Intel processor. Finally, the testing phase involved 10 instances. To gauge how well-automated segmentation is, we created a set of performance criteria for each clinical case often utilized in the literature to compare shapes. The mean, standard deviation, and confidence interval (CI) were used to compute where several measurements are considered for analysis. An analysis of variance (ANOVA) was carried out to evaluate statistical conflicts amongst networks on the DSC. For p-values below 0.05, statistical significance was taken into account.

4. Numerical results and discussion

This work used the Aggregated Jaccard Index (AJI), Panoptic Quality (PQ), F1-score (F1), and DS to assess the efficacy of the anticipated model. AJI is used to evaluate object-level segmentation.

$$AJI = \frac{\sum_{i=1}^{N} |G_i \cap P_M^i|}{\sum_{i=1}^{N} |G_i \cap P_M^i| + \sum_{F \in U} |P_F|} \tag{3}$$

Where G_i is the i^{th} nuclear element out of N nuclei altogether in the ground truth. It is the collection of falsely positive forecasts that lack the associated ground truth. M denotes the prediction's index that overlaps each ground truth item G_i the most, and each M may only be utilized as follows:

$$M = argmax \frac{P_M^i \cap G_i}{P_M^i \cup G_i} \tag{4}$$

The metric for detection performance is the F1-score for objects, which is determined by the proportion of correct and false detections:

$$F1 = \frac{2TP}{FN + 2TP + FP} \tag{5}$$

Where the percentages of true positives, false negatives, and positives are denoted by TP, FN, and FP, respectively (things that have been discovered and corrected), false negative (items that have been missed), and false positive (items that have been detected but have no associated ground truth) detections. It should be noted that a true positive object should cross more than half of the relevant ground truth to receive an object-level F1 score. Panoptic Quality (PQ), the outcome of object segmentation and object detection quality (DQ and SQ,

Doi : https://doi.org/10.54216/JCIM.140207
Received: January 17, 2024 Revised: March 03, 2024 Accepted: July 03, 2024

respectively), has previously been used to assess the effectiveness of panoptic segmentation tasks. PQ is expressed as in Eq. (6):

$$PQ = \frac{2|TP|}{2|TP| + |FP| + |FN|} * \frac{\sum_{(p,g)\in TP} IoU(p,g)}{|TP|} \quad (6)$$

$|TP|, |FN|, and$ | FP | specifies percentages of false negative, true positive and false positive detections. Every set of predictions (p, g) is generated using real positive detections and the ground truth. If $I\ (p,q) > 0.5$, a prediction is only considered the real positive. The PQ metric represents how well object detection and segmentation are performed, as shown in Eq. (7). Pixel-level Dice score is used to compare the ground truth with the binarized prediction to assess the accuracy of the foreground and background segmentation:

$$Dice\ = \frac{2|P \cap G|}{|P| + |G|} \quad (7)$$

Here, P and G specifies ground truth and binarization prediction, respectively. Then,$|.|$ denotes all pixels in the foreground image. This work immediately uses Best Dice to determine the metrics for the leaf segmentation challenge:

$$SBD\ (P,T) = \min(BD\ (P,T), BD(T,P)) \quad (8)$$

Where P and T stand for forecasts and reality, respectively. $BD(P,T)$ is the ideal die between $P_i(i = 1,\ldots,M)\ and\ T_j(j\ =\ 1,\ldots,N)$:

$$BP\ (P,T) = \frac{1}{M}\sum_{i=1}^{M} \max_{j=1,\ldots,N} \frac{2|P_i \cap T_j|}{|P_i| + |T_j|} \quad (9)$$

Where the entire number of pixels in the foreground is represented by $|.|$.

4.1. Execution details

The following are the pre-trained network backbone's weights initialized using initialization, whereas the other layers' weights undergo pre-training for the classification job for ImageNet. The $ID - Net$ is trained and optimized using SGD of 0.9 velocity and 0.0001 weight decay. A little batch size, or batch size, is regarded as one. Rather than using the standard batch normalization layers, we used group normalizing layers with a group size of 32. With 500 iterations, starting learning rate is set at 0.003. When 3/4 of the total training repetitions have been completed, the learning rate is diminished to 0.0003. Using MATLAB 2020a, we researched two GeForce 1080Ti GPUs from Nvidia.

4.2. Evaluation with other approaches

TCGA: Several cutting-edge nucleus instance segmentation techniques are compared with our findings, including CNN, R-CNN, D-CNN, ResNet, LSTM and U-net. This work directly contrasts the results indicated using the same data split. The prediction is used for a fair comparison with group normalization and is all re-implemented using the same parameters as our suggested $ID - Net$. As a result, they are superior to existing approaches. Tables 1 to 6 respectively display the outcomes of our quantitative and qualitative comparison analyses. Table 1 demonstrates that our proposed $ID - Net$ beats all other approaches in the four criteria on visible and unseen testing sets. Testing using samples from the invisible organs suggests that $ID - Net$ has a great generalization capacity. To get the p-value, we performed a paired t-test with one tail to statistically compare the findings of $ID - Net$ with other techniques. Table 2 demonstrates that, except for the F1 of the proposed model, our increases for all four measures are 0.05 p-value required for statistical significance.

Doi : https://doi.org/10.54216/JCIM.140207
Received: January 17, 2024 Revised: March 03, 2024 Accepted: July 03, 2024

Regardless of the level of segmentation for each identified item, F1 considers the total number of corrected detected objects. Our $ID-Net$ outperforms CNN3 on nucleus segmentation tasks while surpassing it by roughly 10% on AJI and 4% on Dice, the other two segmentation metrics) by a substantial margin. All approaches using the four measures, demonstrating that our suggested $ID-Net$ performs better than other methods but is more reliable and stable.

TNBC: On the second histopathology dataset, we ran comparison tests using the outcomes of a threefold cross-validation are displayed in Table 3. $ID-Net$ performs better than earlier iterations in all three criteria, according to Table 4. The $ID-Net$ performs significantly better than the R-CNN in terms of efficacy. The TNBC dataset's backdrop elements are intricate, and some background textures resemble the foreground. As a result, analyzing the semantic-level data improves segmentation and detection precision. The ID-Net improvement is less noticeable when compared to the TCGA dataset, notably the F1 score for objects. Despite this, $ID-Net$ uses a feature integration method that makes it easier to learn semantic features in the instance branch; there aren't enough contextual features to surround each object since a portion of the semantic feature has depreciated. As an outcome, when the borders of two touching things are clear, the detection accuracy could improve. Results from the TCGA dataset show our proposed $ID-Net$ model is accurate and effective and performs significantly better than similar methods.

BBBC039V1: Our suggested $ID-Net$ is effective for segmentation of medical and histopathology images. Table 5 shows how effectively $ID-Net$ perform functions better than any of the compared approaches. The background elements in the fluorescence microscopy photos are less complex than in the histopathological images. Accuracy cannot be improved because R-CNN cannot handle the background objects' contextual information by training the backbone encoder's semantic features. Creating a dual-modal generator and making it available, with instructions to understand global semantic-level properties, boosts the segmentation accuracy in the instance branch. The $ID-Net$ pixel-level Dice score has improved, but only a little. Our suggested $ID-Net$ has a pre-trained network at its core, producing significant gains in all four measures. Using the 33 test photographs for segmentation, we also carried out a comparative experiment with other prior studies to emphasize the effectiveness of $ID-Net$ for image analysis. The comparison between our work and cutting-edge techniques is shown in Table 6, which also outperforms all previously published studies on this dataset regarding segmentation accuracy.

Table 1: Comparison with TCGA Dataset

Methods		**AJI**			**Dice**			**F1**			**PQ**		
		K*	**UK****	**All**	**K***	**UK****	**All**	**K***	**UK****	**All**	**K***	**UK****	**All**
CNN	Avg	0.51	0.49	0.50	0.73	0.80	0.76	0.82	0.83	0.82	0.47	0.48	0.48
	Std	0.08	0.08	0.06	0.05	0.10	0.09	0.09	0.07	0.09	0.07	0.06	0.06
R-CNN	Avg	0.55	0.56	0.55	0.77	0.80	0.78	0.83	0.84	0.83	0.48	0.49	0.49
	Std	0.05	0.06	0.07	0.04	0.05	0.05	0.09	0.07	0.09	0.07	0.06	0.06
D-CNN	Avg	0.54	0.53	0.53	0.76	0.76	0.76	0.69	0.64	0.67	0.48	0.47	0.47
	Std	0.06	0.12	0.09	0.04	0.04	0.06	0.13	0.19	0.15	0.08	0.17	0.12
LSTM	Avg	0.55	0.56	0.55	0.78	0.77	0.77	0.75	0.74	0.75	0.50	0.50	0.50
	Std	0.05	0.11	0.08	0.04	0.04	0.05	0.09	0.14	0.11	0.08	0.13	0.10
ResNet	Avg	0.57	0.59	0.58	0.78	0.78	0.80	0.80	0.80	0.80	0.55	0.55	0.55
	Std	0.05	0.11	0.08	0.04	0.04	0.06	0.07	0.10	0.08	0.07	0.13	0.10
U-Net	Avg	0.05	0.62	0.61	0.79	0.79	0.82	0.80	0.83	0.83	0.58	0.59	0.58
	Std	0.60	0.09	0.07	0.04	0.04	0.05	0.06	0.05	0.06	0.07	0.10	0.08
$ID-Net$	Avg	**0.61**	**0.63**	**0.62**	**0.80**	**0.80**	**0.83**	**0.81**	**0.84**	**0.84**	**0.59**	**0.60**	**0.60**
	Std	0.06	0.09	0.07	0.04	0.04	0.05	0.06	0.05	0.06	0.07	0.11	0.11

K- known and UK- Unknown

Table 2: p-value and mean comparison of various methods with the TCGA dataset

Methods	**AJI**		**Dice**		**F1**		**PQ**	
	p-value	**Mean**	**p-value**	**Mean**	**p-value**	**Mean**	**p-value**	**Mean**
CNN	0.0026	0.53	0.0016	0.73	0.34	0.75	0.0020	0.51
R-CNN	0.0023	0.57	0.0017	0.76	0.45	0.81	0.0025	0.56

Doi : https://doi.org/10.54216/JCIM.140207
Received: January 17, 2024 Revised: March 03, 2024 Accepted: July 03, 2024

D-CNN	0.0064	0.59	0.0014	0.77	0.55	0.81	0.0014	0.58
LSTM	0.0039	0.63	0.0023	0.80	0.25	0.86	0.0061	0.62
ResNet	0.0071	0.65	0.0010	0.84	0.13	0.88	0.0010	0.65
U-Net	0.0082	0.71	0.0015	0.88	0.14	0.90	0.0015	0.71
$\boldsymbol{ID-Net}$	0.0085	0.73	0.0010	0.90	0.15	0.94	0.0016	0.73

Table 3: Comparison with the TNBC Dataset

Methods		**AJI**			**Dice**			**F1**			**PQ**		
		K*	**UK****	**All**	**K***	**UK****	**All**	**K***	**UK****	**All**	**K***	**UK****	**All**
CNN	Avg	0.53	0.51	0.51	0.74	0.81	0.77	0.83	0.84	0.83	0.48	0.49	0.49
	Std	0.08	0.09	0.06	0.05	0.10	0.09	0.09	0.07	0.09	0.07	0.06	0.06
R-CNN	Avg	0.57	0.57	0.56	0.78	0.81	0.79	0.84	0.85	0.84	0.49	0.50	0.50
	Std	0.06	0.07	0.07	0.04	0.05	0.05	0.09	0.07	0.09	0.07	0.06	0.06
D-CNN	Avg	0.56	0.55	0.54	0.77	0.77	0.77	0.70	0.66	0.68	0.49	0.48	0.48
	Std	0.07	0.13	0.09	0.04	0.04	0.06	0.13	0.19	0.15	0.08	0.17	0.12
LSTM	Avg	0.57	0.56	0.56	0.79	0.78	0.78	0.76	0.75	0.76	0.51	0.51	0.51
	Std	0.06	0.11	0.08	0.04	0.04	0.05	0.09	0.14	0.11	0.08	0.13	0.10
ResNet	Avg	0.58	0.60	0.59	0.79	0.79	0.81	0.81	0.81	0.81	0.56	0.56	0.56
	Std	0.05	0.11	0.08	0.04	0.04	0.06	0.07	0.10	0.08	0.07	0.13	0.10
U-Net	Avg	0.59	0.63	0.62	0.80	0.80	0.83	0.81	0.84	0.84	0.59	0.60	0.59
	Std	0.60	0.09	0.07	0.04	0.04	0.05	0.06	0.05	0.06	0.07	0.10	0.08
$\boldsymbol{ID-Net}$	Avg	**0.62**	**0.64**	**0.63**	**0.81**	**0.81**	**0.84**	**0.82**	**0.85**	**0.85**	**0.60**	**0.61**	**0.61**
	Std	0.06	0.09	0.07	0.04	0.04	0.05	0.06	0.05	0.06	0.07	0.11	0.11

Table 4: p-value and mean comparison of various methods with the TNBC dataset

Methods	**AJI**		**Dice**		**F1**		**PQ**	
	p-value	**Mean**	**p-value**	**Mean**	**p-value**	**Mean**	**p-value**	**Mean**
CNN	0.0028	0.54	0.0018	0.74	0.36	0.76	0.0021	0.52
R-CNN	0.0025	0.58	0.0019	0.77	0.46	0.82	0.0026	0.57
D-CNN	0.0066	0.60	0.0016	0.78	0.54	0.82	0.0015	0.59
LSTM	0.0041	0.64	0.0025	0.81	0.26	0.87	0.0062	0.63
ResNet	0.0073	0.66	0.0012	0.85	0.14	0.89	0.0011	0.66
U-Net	0.0084	0.72	0.0017	0.89	0.15	0.91	0.0016	0.72
$\boldsymbol{ID-Net}$	0.0087	0.74	0.0012	0.91	0.16	0.95	0.0017	0.74

Table 5: Comparison with BBBC039V1 dataset

Methods		**AJI**			**Dice**			**F1**			**PQ**		
		K*	**UK****	**All**	**K***	**UK****	**All**	**K***	**UK****	**All**	**K***	**UK****	**All**
CNN	Avg	0.54	0.52	0.53	0.76	0.83	0.79	0.85	0.86	0.85	0.50	0.51	0.51
	Std	0.08	0.08	0.06	0.05	0.10	0.09	0.09	0.07	0.09	0.07	0.06	0.06
R-CNN	Avg	0.58	0.59	0.58	0.79	0.83	0.81	0.86	0.87	0.86	0.51	0.51	0.52
	Std	0.05	0.06	0.07	0.04	0.05	0.05	0.09	0.07	0.09	0.07	0.06	0.06
D-CNN	Avg	0.57	0.56	0.55	0.79	0.79	0.79	0.72	0.67	0.70	0.51	0.50	0.55
	Std	0.06	0.12	0.09	0.04	0.04	0.06	0.13	0.19	0.15	0.08	0.17	0.12

Doi : https://doi.org/10.54216/JCIM.140207
Received: January 17, 2024 Revised: March 03, 2024 Accepted: July 03, 2024

LSTM	Avg	0.58	0.59	0.57	0.81	0.80	0.81	0.78	0.77	0.78	0.53	0.53	0.53
	Std	0.05	0.11	0.08	0.04	0.04	0.05	0.09	0.14	0.11	0.08	0.13	0.10
ResNet	Avg	0.60	0.62	0.61	0.82	0.81	0.83	0.83	0.83	0.83	0.58	0.58	0.58
	Std	0.05	0.11	0.08	0.04	0.04	0.06	0.07	0.10	0.08	0.07	0.13	0.10
U-Net	Avg	0.63	0.65	0.63	0.82	0.81	0.85	0.83	0.86	0.86	0.61	0.62	0.61
	Std	0.60	0.09	0.07	0.04	0.04	0.05	0.06	0.05	0.06	0.07	0.10	0.08
ID − Net	Avg	**0.64**	**0.66**	**0.65**	**0.83**	**0.83**	**0.86**	**0.84**	**0.87**	**0.87**	**0.62**	**0.63**	**0.63**
	Std	0.06	0.09	0.07	0.04	0.04	0.05	0.06	0.05	0.06	0.07	0.11	0.11

Table 6: p-value and mean comparison of various methods with the BBBC039V1 dataset

Methods	**AJI**		**Dice**		**F1**		**PQ**	
	p-value	**Mean**	**p-value**	**Mean**	**p-value**	**Mean**	**p-value**	**Mean**
CNN	0.0029	0.56	0.0019	0.76	0.37	0.78	0.0023	0.54
R-CNN	0.0026	0.60	0.0020	0.79	0.48	0.84	0.0028	0.59
D-CNN	0.0067	0.62	0.0017	0.80	0.57	0.84	0.0017	0.61
LSTM	0.0042	0.65	0.0026	0.83	0.28	0.89	0.0064	0.65
ResNet	0.0074	0.68	0.0013	0.87	0.16	0.91	0.0013	0.68
U-Net	0.0085	0.74	0.0018	0.94	0.17	0.93	0.0018	0.74
ID − Net	0.0088	0.76	0.0013	0.96	0.18	0.97	0.0019	0.77

Table 7: IoU comparison

Methods	**IoU**
CNN	0.80 ± 0.06
R-CNN	0.81 ± 0.07
D-CNN	0.82 ± 0.05
LSTM	0.80 ± 0.06
ResNet	0.92 ± 0.07
U-Net	0.93 ± 0.07
ID − Net	0.95 ± 0.07

Doi : https://doi.org/10.54216/JCIM.140207
Received: January 17, 2024 Revised: March 03, 2024 Accepted: July 03, 2024

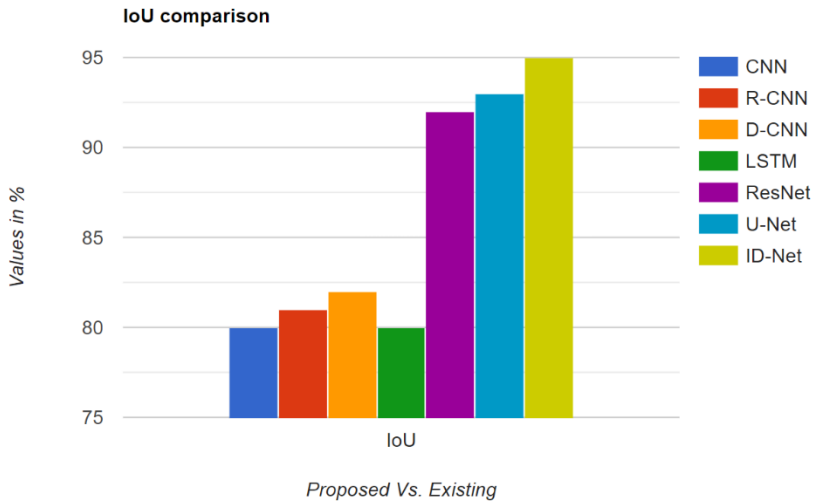
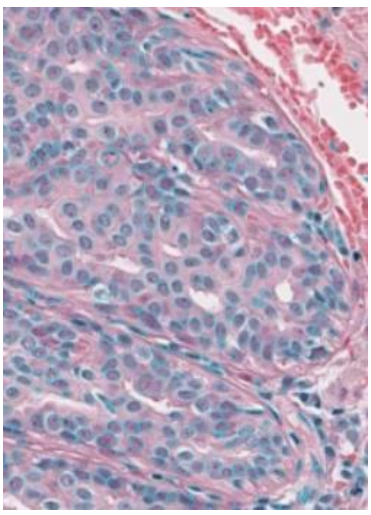
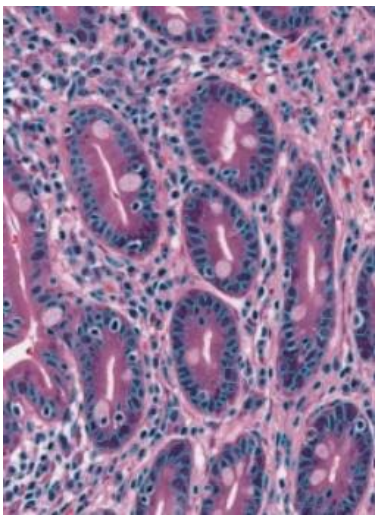


Figure 2: IoU comparison

Recurrent with attention, RNN and RIS processed one event at a time, using the temporal chain from either short-term memory (LSTM). In addition, the existing model outperformed in terms of performance thanks to the proposal-based design and the attention module. Other approaches focus on effectual spatial link and suited for the issue because the leaf instance segmentation task lacks time information. Techniques, for instance, segmentation without proposals, separate each set of cases un-touching into several groups before processing each group independently. The projections of the acquired high-dimensional embedding maps for each leaf image rather than learning the instance prediction directly. Then, during inference, each instance was divided using clustering techniques. Their performance is still constrained even without focusing on every item since there is insufficient local-level data. The CNN is an offer-based R-CNN architecture that is similar to our method. Most of the most recent cutting-edge methods are outperformed by these two technologies with the aid of the auxiliary synthetic images. In the given phenotypic images, the image creation techniques described, on the other hand, are entirely based on the characteristics such as texture and orientation. Therefore, applying the methodologies to other datasets with task-specific attributes takes time and effort. Without any task-specific design, our suggested-based $ID - Net$ performs better than all other techniques on the TCGA dataset owing to panoptic-level properties from both a local and global viewpoint. Further proof of the generalizability of $ID - Net$ comes from additional instance segmentation tasks; it performs at a competitive level.

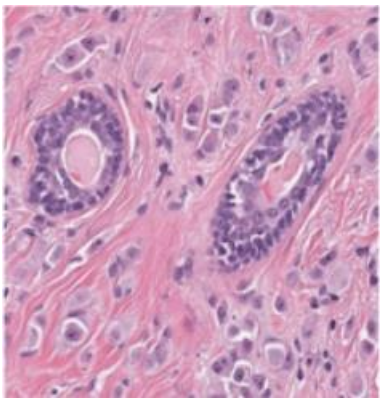


Input with TCGA dataset

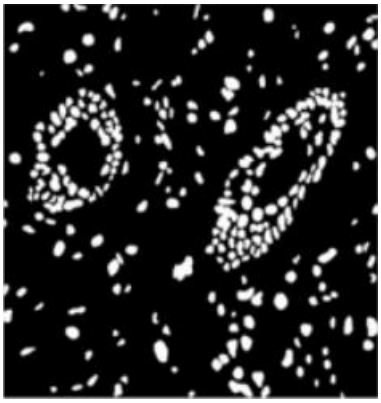


Segmented outcome of samples from TCGA dataset

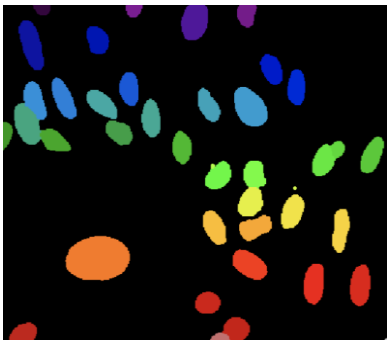
Doi : https://doi.org/10.54216/JCIM.140207
Received: January 17, 2024 Revised: March 03, 2024 Accepted: July 03, 2024



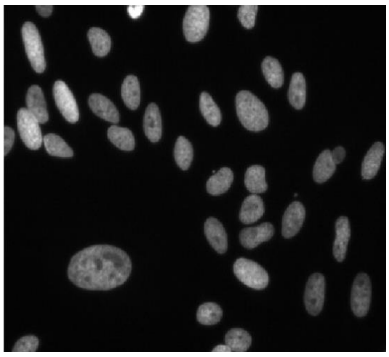
Input with TNBC dataset



Segmented outcome of samples from TNBC dataset



Input with BBBC039V1 dataset



Segmented outcome of samples from BBBC039V1 dataset

Figure 3: Segmented outcomes

4.4. Generalization

This work conducts generalization research by training the model and validating it that cannot be seen to show the generalization capability of d $ID-Net$. The dataset were trained using training set and TCGA testing set to verify it; we adhered to the experimental setup. Before training, each TCGA training image was randomly cropped into 10 $256*256$ patches. The patches are then enhanced using In addition to flipping them horizontally and vertically, rotating them 90, 180, and 270 degrees, we also consider Gaussian turbulence, median blur, and Gaussian noise. The TCGA testing set's fourteen test images validate the well-trained model for testing. Table 7 is a complete illustration of the experimental findings. Table 7 demonstrates that our proposed $ID-Net$ outperforms our old R-CNN on all criteria. A further instance is evidence of the effectiveness of our recommended method compared to other regularization modules for semantic task consistency and the residual attention feature integration method by improving the models' capacity for generalization. The model utilizes the nucleus instance that employs point annotations for training and inference, in addition to what we have already said. The existing model outperforms our $ID-N$et in performance with access to the point annotations for the test images. This work observes that U-Net performs similarly to our $ID-Net$ under instance-level segmentation; the IoU score is used, while the Dice score is used for pixel-level segmentation (See Fig 3). However, there are still expenses associated with getting the nucleus point annotations for the datasets. We prefer low-cost human annotations instead of any annotations from the test datasets used in secret if a less competitive accuracy is acceptable.

5. Conclusion

This research merges instance-level and semantic data to create an improved dense Net ($ID-Net$) for segmenting instances in biomedical images in this study. Our recently suggested $ID-Net$ is enhanced with feature integration approach and consistency regularization by expanding conventional U-Net. The segmentation provides extra global contextual data that generators may access; the feature integration process preserves the quality of foreground items. By learning the quality scores for each prediction during training using quality score

Doi : https://doi.org/10.54216/JCIM.140207
Received: January 17, 2024 Revised: March 03, 2024 Accepted: July 03, 2024

to classification instance where the anticipated model addresses the issue of misalignment among prediction quality and classification score. The two semantic segmentation tasks should be simultaneously regularized using the semantic consistency technique to make it possible to accurately and consistently determine the semantic properties. Our $ID-Net$ considerably outperforms numerous methods using biomedical datasets. The proposed $ID-Net$ has shown effectiveness on various biomedical datasets by meeting the requirements for the $ID-Net$ future development. This work would improve $ID-Net$ in subsequent experiments to better suit requirements for general image processing. As a result of our $ID-Net$ success with 2D image analysis, this work may extend it to segment 3D microscope image instances, another significant and engaging issue linked to this study.

References

[1] Xing and L. Yang, "Robust nucleus/cell detection and segmentation in digital pathology and microscopy images: A comprehensive review," IEEE Rev. Biomed. Eng., vol. 9, pp. 234–263, 2016.

[2] Song, L. Xiao, and Z. Lian, "Contour-seed pairs learning-based framework for simultaneously detecting and segmenting various overlapping cells/nuclei in microscopy images," IEEE Trans. Image Process., vol. 27, no. 12, pp. 5759–5774, Dec. 2018.

[3] Payer, D. Štern, M. Feiner, H. Bischof, and M. Urschler, "Segmenting and tracking cell instances with cosine embeddings and recurrent hourglass networks," Med. Image Anal., vol. 57, pp. 106–119, Oct. 2019

[4] De Brabandere, D. Neven, and L. Van Gool, "Semantic instance segmentation with a discriminative loss function," 2017, arXiv:1708.02551. [Online]. Available: http://arxiv.org/abs/1708.02551

[5] Ambeth Kumar, V.D. (2017). Automation of Image Categorization with Most Relevant Negatives. Pattern Recognition and Image Analysis, 27(3), 371–379.

[6] Kumar, I., Kumar, A., Kumar, V.D.A. et al. (2022) Dense Tissue Pattern Characterization Using Deep Neural Network. Cogn Comput 14, 1728–1751.

[7] Liu et al., "Nuclei segmentation via a deep panoptic model with semantic feature integration," in Proc. 28th Int. Joint Conf. Artif. Intell., AAAI Press, Aug. 2019, pp. 861–868.

[8] Chen, A. Hermans, G. Papandreou, F. Schroff, P. Wang, and H. Adam, "MaskLab: Instance segmentation by refining object detection with semantic and direction features," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., Jun. 2018, pp. 4013–4022

[9] Ambeth Kumar, V.D. Vaishali,S. Shweta, B. (2015). Basic Study of the Human Foot. Biomedical and Pharmacology, 8(1), 435-444.

[10] Y. Li et al., "Attention-guided unified network for panoptic segmentation," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2019, pp. 7026–7035.

[11] Chen et al., "Hybrid task cascade for instance segmentation," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2019, pp. 4974–4983.

[12] He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2016, pp. 770–778.

[13] Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jul. 2017, pp. 2117–2125.

[14] Ambeth Kumar, V.D. Ramakrishnan,M. (2013). Temple and Maternity Ward Security using FPRS. Journal of Electrical Engineering & Technology, 8(3), 633-637.

[15] Salvador et al., "Recurrent neural networks for semantic instance segmentation," 2017, arXiv:1712.00617. [Online]. Available: http://arxiv.org/abs/1712.00617

[16] Cuocolo, R.; Stanzione, A.; Ponsiglione, A.; Romeo, V.; Verde, F.; Creta, M.; La Rocca, R.; Longo, N.; Pace, L.; Imbriaco, M. Clinically significant prostate cancer detection on MRI: A radio mic shape features study. Eur. J. Radiol. 2019, 116, 144–149.

[17] Comelli, A.; Bignardi, S.; Stefano, A.; Russo, G.; Sabini, MG; Ippolito, M.; Yezzi, A. Development of a new fully three-dimensional methodology for tumour delineation in functional images. Comput. Biol. Med. 2020, 120, 103701.

[18] Christe, A.; Peters, A.A.; Drakopoulos, D.; Heverhagen, J.T.; Geiser, T.; Stathopoulou, T.; Christodoulidis, S.; Anthimopoulos, M.; Mougiakakou, S.G.; Ebner, L. Computer-Aided Diagnosis of Pulmonary Fibrosis Using Deep Learning and CT Images. Invest. Radiol. 2019, 54, 627–632.

[19] Kumar, V.D.A., Sharmila, S., Kumar, A. et al. (2023). A novel solution for finding postpartum haemorrhage using fuzzy neural techniques. Neural Comput & Applic. 35(33), 23683–23696

[20] Torrisi, S.E.; Palmucci, S.; Stefano, A.; Russo, G.; Torcitto, A.G.; Falsaperla, D.; Gioè, M.; Pavone, M.; Vancheri, A.; Sambataro, G.; et al. Assessment of survival in patients with idiopathic pulmonary fibrosis using quantitative HRCT indexes. Multidiscip. Respir. Med. 2018, 13, 1–8.

Doi : https://doi.org/10.54216/JCIM.140207
Received: January 17, 2024 Revised: March 03, 2024 Accepted: July 03, 2024

[21] Gerard, S.E.; Herrmann, J.; Kaczka, D.W.; Musch, G.; Fernandez-Bustamante, A.; Reinhardt, J.M. Multi-resolution convolutional neural networks for fully automated segmentation of acutely injured lungs in multiple species. Med. Image Anal. 2020, 60, 101592.

[22] Sun, K.; Xiao, B.; Liu, D.; Wang, J. Deep, high-resolution representation learning for human pose estimation. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–20 June 2019; pp. 5686–5696.

[23] Cuocolo, R.; Cipullo, MB; Stanzione, A.; Ugga, L.; Romeo, V.; Radice, L.; Brunetti, A.; Imbriaco, M. Machine learning applications in prostate cancer magnetic resonance imaging. Eur. Radiol. Exp. 2019, 3, 35.

[24] Sathya Preiya, V., and V. D. Ambeth Kumar. (2023). Deep Learning-Based Classification and Feature Extraction for Predicting Pathogenesis of Foot Ulcers in Patients with Diabetes. Diagnostics 13(12), 1983.

[25] Park, B.; Park, H.; Lee, S.M.; Seo, J.B.; Kim, N. Lung Segmentation on HRCT and Volumetric CT for Diffuse Interstitial Lung Disease Using Deep Convolutional Neural Networks. J. Digit. Imaging 2019, 32, 1019–1026

[26] Hemamalini, Selvamani, and Visvam Devadoss Ambeth Kumar. (2022). Outlier Based Skimpy Regularization Fuzzy Clustering Algorithm for Diabetic Retinopathy Image Segmentation. Symmetry, 14(12), 2512

[27] Piyush K. Pareek, Pixel Level Image Fusion in Moving objection Detection and Tracking with Machine Learning ",Fusion: Practice and Applications, Volume 2 , Issue 1 , PP: 42-60, 2020

[28] Shivam Grover, Kshitij Sidana, Vanita Jain, "Egocentric Performance Capture: A Review", Fusion: Practice and Applications, Volume 2, Issue 2 , PP: 64-73, 2020.

[29] Abdel Nasser H. Zaied, Mahmoud Ismail and Salwa El- Sayed, A Survey on Meta-heuristic Algorithms for Global Optimization Problems, Journal of Intelligent Systems and Internet of Things,Volume 1 , Issue 1 , PP: 48-60, 2020

[30] Mahmoud H.Alnamoly, Ahmed M. Alzohairy, Ibrahim M. El-Henawy, "A survey on gel images analysis software tools, Journal of Intelligent Systems and Internet of Things,Volume 1 , Issue 1 , PP: 40-47, 2021.

Doi : https://doi.org/10.54216/JCIM.140207
Received: January 17, 2024 Revised: March 03, 2024 Accepted: July 03, 2024