

## RESEARCH ARTICLE

# DRD-UNet, a UNet-Like Architecture for Multi-Class Breast Cancer Semantic Segmentation

MAURICIO ALBERTO ORTEGA-RUIZ<sup>1,2</sup>, CEFA KARABAĞ<sup>3</sup>, EDGAR ROMAN-RANGEL<sup>4</sup>,  
AND CONSTANTINO CARLOS REYES-ALDASORO<sup>2</sup>, (Senior Member, IEEE)

<sup>1</sup>Department of Engineering, CIIDETEC-Coyoacan, Universidad del Valle de México, Mexico City 53220, Mexico

<sup>2</sup>School of Science and Technology, City, University of London, EC1V 0HB London, U.K.

<sup>3</sup>School of Computing and Digital Media, London Metropolitan University, N7 8DB London, U.K.

<sup>4</sup>Department of Computer Science, Instituto Tecnológico Autónomo de México, Mexico City 01080, Mexico

Corresponding author: Constantino Carlos Reyes-Aldasoro (reyes@city.ac.uk)

The work of Edgar Roman-Rangel was supported by Asociación Mexicana de Cultura A.C.

**ABSTRACT** Staining of histological slides with Hematoxylin and Eosin is widely used in clinical and laboratory settings as these dyes reveal nuclear structures as well as cytoplasm and collagen. For cancer diagnosis, these slides are used to recognize tissues and morphological changes. Tissue semantic segmentation is therefore important and at the same time a challenging and time-consuming task. This paper describes a UNet-like deep learning architecture called *DRD-UNet*, which adds a novel processing block called *DRD* (Dilation, Residual, and Dense block) to a UNet architecture. *DRD* is formed by the combination of dilated convolutions (D), residual connections (R), and dense layers (D). *DRD-UNet* was applied to the multi-class (tumor, stroma, inflammatory, necrosis, and other) semantic segmentation of histological images from breast cancer samples stained with Hematoxylin and Eosin. The histological images were released through the Breast Cancer Semantic Segmentation (BCSS) Challenge. *DRD-UNet* outperformed the original UNet architecture and 15 other UNet-based architectures on the segmentation of 12, 930 image patches extracted from regions of interest that ranged in size between  $1036 \times 1222$  to  $6813 \times 7360$  pixels. *DRD-UNet* obtained the best performance as measured with Jaccard similarity index, Dice coefficient, in a per-class comparison and accuracy for overall segmentation.

**INDEX TERMS** UNet, deep learning architectures, histopathology, breast cancer, segmentation.

## I. INTRODUCTION

Image segmentation has been one of the most important tasks in computer vision, pattern recognition, and image analysis. The essence of image segmentation consists of assigning a label or class to each element of the image. Labels can correspond to organs in medical imaging [1], or human bodies in videos of surveillance [2]. A plethora of segmentation approaches have been proposed through the years; from traditional image processing techniques [3] and more recently deep learning [4]. The fact that segmentation algorithms continue to be published every year, illustrates

The associate editor coordinating the review of this manuscript and approving it for publication was Essam A. Rashed<sup>5</sup>.

that it is a difficult problem, with many challenges yet to be addressed [5], [6]. The use of deep learning architectures has grown significantly in recent years [7], in part due to the superior results that have been achieved in many areas, medical image analysis being just one of them [8], [9]. Another reason for the popularity of deep learning is that hand-crafted features and analysis are replaced by a learning process that results from training with a large amount of labeled data [10]. Deep learning-based semantic segmentation of biomedical images has grown dramatically since the first publication of UNet by Ronneberger in 2015 [11]. Whilst the standard UNet architecture, sometimes called *vanilla UNet* [12], has been used in numerous applications, modifications to the architecture have improved

results [13]. However, it is considered that there is still room for improvement [14], [15], and any improvement would be welcome in medical environments where decisions like diagnosis or treatment could be influenced by the outcome of previous segmentation or classification steps. Indeed, it has been observed that the survival in cases of colon cancer can be predicted from histology slides [16].

The probability of developing breast cancer in females is the highest compared to that in any other organ at approximately 12.4% [17]. Early detection through screening programs is the best way to improve patient outcome and reduce mortality [18]. As such, mortality has decreased slightly since the 1970s [17], but breast cancer remains a significant concern for females worldwide [19]; thus, any improvement in treatment and diagnosis is to be pursued. Mammography is the first tool for diagnosis of breast cancer, and screening programs have shown significant reductions in breast cancer mortality [20]. However, mammography has limitations and harms, besides the radiation inherent in the procedure, sensitivity and specificity are limited as compared with magnetic resonance imaging [21], and most importantly, its use can lead to over diagnosis, which is considered the greatest harm of the technique [22]. Histopathology, or the examination of tissues under a microscope, can provide sufficient information to avoid over diagnosis and over treatment of breast cancer [23]. Histopathology provides a huge amount of information [16], and advances in molecular techniques, bioinformatics, and image analysis can complementary information to reveal diagnosis, treatment, and survival [24], [25], [26]. The development of whole slide imaging [27], which scan slides and capture images at different resolutions, has grown in popularity and provides huge datasets with a wealth of information, and besides the disadvantage of capital investment [28], there is the technical challenge of analysis images that can comprise billions of pixels. To analyze these large images, which can be part of even larger datasets, computational approaches are thus required and can provide new perspectives in clinical practice [29], [30], [31], [32], [33], [34], [35], [36], [37], [38], like grading tissue [39], [40], [41], [42], or proliferation of tumors [43].

This work presents a UNet-like architecture, which provides a multi-class semantic segmentation breast cancer images stained with H&E. The main contributions of this paper are the following:

- 1) A novel processing block called DRD, which is a pyramid block with dilated convolutions, a residual connection, and a dense layer is proposed. The addition of this block improved the segmentation performance of histological images.
- 2) Sixteen UNet-like architectures were objectively compared with the original UNet used as a baseline. The architecture here proposed, DRD-UNet outperformed all other architectures/

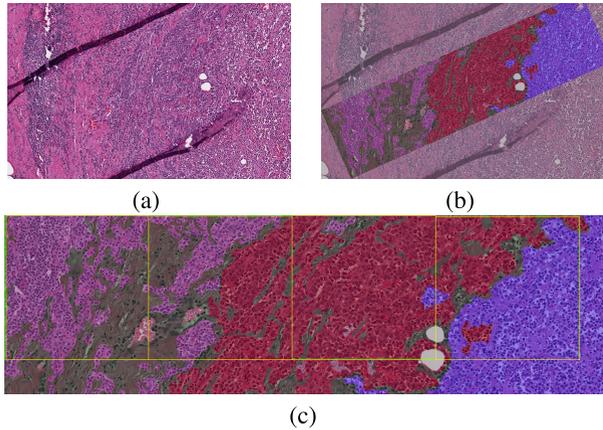
## II. RELATED WORK

Pyramids [44], [45] and quad trees [46] have been used for decades as multiresolution tools that allow the analysis of images at distinct levels of detail. The underlying concept of these techniques is to combine the values of neighboring pixels and subsampling to create a new image, or *level* of the pyramid, which is smaller in size than the previous one and whose values correspond to a neighborhood in a lower level and the filter or kernel applied. In this way, decisions that are taken at higher levels of the pyramid will have information of larger and larger neighborhoods of pixels. One of the first attempts to use these schemes for visual pattern recognition was the Neocognitron proposed by Fukushima [47]. A few years later Le Cun and Bengio extended the approach by extending the features through a series of convolutions, thus creating the widely popular convolutional networks [48].

Deconvolutional Neural Networks (DCNN) [49] perform an inverse convolutional model by an upsampling process thus increasing the resolution of the output, extensively used in segmentation tasks; the precursor of DCNN is the Neocognitron proposed by Fukushima. A fully Convolutional Network (FCN) is the first example of a deconvolutional model, for example, [50] incorporated a dense prediction layer, and a fully connected Conditional Random Field.

Convolution with an extended kernel dilation is also known as Atrous convolution [51] and it has been proved for the segmentation of objects at multiple scales, as well as for weakly supervised semantic segmentation [52].

UNet has been widely used for segmentation in medical imaging, and several modifications and improvements have been proposed [53], [54]. One of the first modifications of UNet was the addition of residual blocks [55] for liver lesion segmentation. ResUNet uses a UNet with residual connections at encoders and decoders and achieved first place in the liver tumor segmentation challenge in 2017. RMS UNet added a residual block with dilated section convolutions [55], [56] and demonstrated high accuracy over different publicly available datasets. Even a recent generative adversarial network (GAN) with a dense UNet-based segmentor and a radiomics discriminator for liver lesion segmentation [57] have been proposed. In brain segmentation, the stack Multi-Connection Simple Reducing Net (SMCSRNET) [58] formed with stacked coder-encoder blocks improved segmentation with less training time, using a reduced parameters model. MH-UNet [59], a multi-scaled network is formed with dense and residual blocks and as a result, it is used for multi-organ segmentation. MI-UNet [60] uses a procedure called brain parcellation to generate an input to UNet for brain stroke segmentation under MRI. Spatial weighted UNet is used for 3D CT brain images [61] with residual-inception blocks densely connected that reduces trainable parameters over the MRI dataset. A Purified and Residual UNet or P-ResUNet [62] is based on a Dilated Pyramid Block (DPB) and was used for brain tissue segmentation. This block consists of dilations of distances



**FIGURE 1. Illustration of the histological images and annotations.** (a) One representative histological image stained with Hematoxylin and Eosin (H&E) from the Breast Cancer Semantic Segmentation (BCSS) Challenge. (b) A rectangular region of interest is denoted by overlaid colors that correspond to labels of interest for this work: tumor (red), stroma (green), inflammatory (purple), necrosis (blue), and other (gray). (c) A zoom of the of the region of interest with additional yellow lines that indicate patches of size  $256 \times 256$ , which will be extracted for the training and validation of the proposed DRD-UNet and other architectures.

$D = 1, 2, 3$  in parallel. For retinal vessel segmentation, the images require enhanced contrast for accurate vessel detection and approaches based on Residual attention and supervised UNet have been proposed [63], [64], [65]. The network MI-UNet [66], which consists of two MI-UNets connected into one S-UNet has been used for Brain Stroke Lesions. For nuclei and cell segmentation, RIC-UNet [67], a network with the mechanism residual blocks, multi-scale, and channel attention mechanism demonstrated superior performance over traditional methods. Residual-dense blocks have been explored also at the bottleneck connection to form D-UNet for lung vessel segmentation [68]. Segmentation of breast tumor cancer has been also explored by different UNet architectures, the ResUNet [69] estimates volumetric measurement of breast cancer on magnetic resonance imaging. An architecture called Connected-UNets using two UNets and additional modified skip connections showed better visual results in segmenting the mass lesions over mammograms [70]. Also, generative networks have been explored, RDA-UNET-WGAN [71] employs a Residual-Dilated-Attention-Gate-UNet as the generator network. The UNet3+, a full-scale skip connections and deep supervision framework, has also been employed for 6 different architecture segmentation models [72].

### III. MATERIALS AND METHODS

#### A. MATERIALS

The datasets used in this paper were obtained from the Breast Cancer Semantic Segmentation (BCSS) Challenge [73], which is publicly available through the Grand Challenges website (<https://bcsegmentation.grand-challenge.org/>).

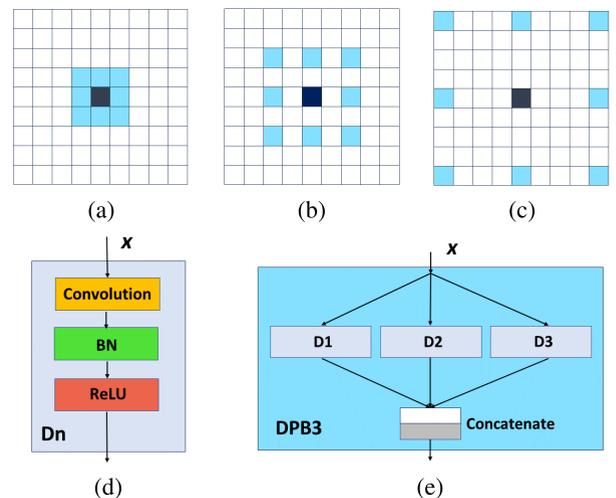
The dataset was extracted from 151 formalin-fixed paraffin-embedded tissues stained with hematoxylin and eosin (H&E), all with histologically-confirmed breast cancer

**TABLE 1. Principal characteristics of the images from the breast cancer semantic segmentation challenge (BCSS) dataset.**

Description	Parameter
Number of images	151
Magnification	$40\times, 20\times$
Maximum image size	$6813 \times 7360$ pixels
Minimum image size	$1036 \times 1222$ pixels
Mean ROI	$1.18 \text{ mm}^2$ $39\text{e}6 \text{ pixel}^2$
Tissue Classes	22

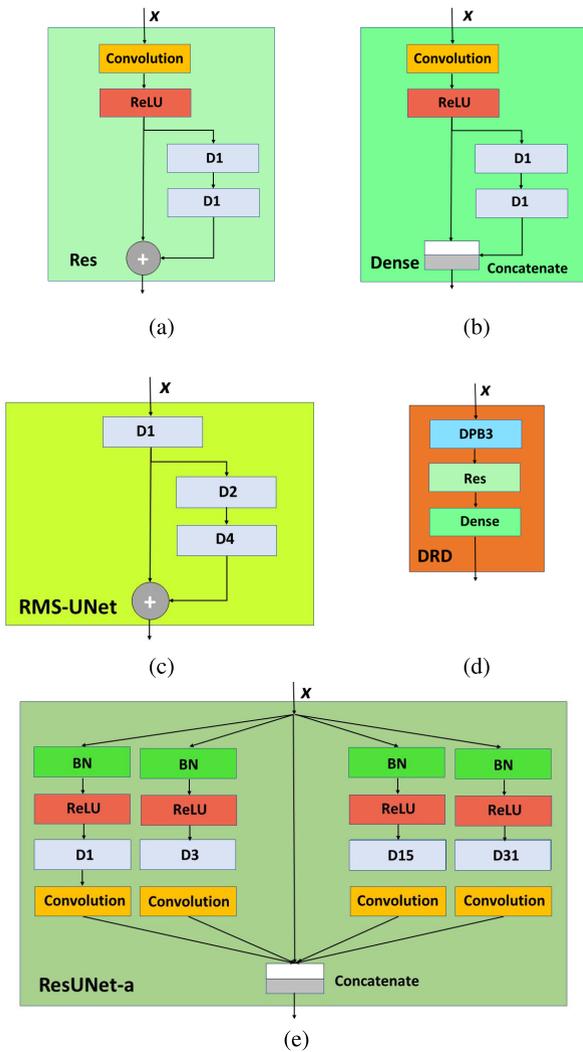
**TABLE 2. Principal characteristics of the patches that were extracted from the images.**

Description	Parameter
Patch size	$256 \times 256$ pixels
Image resize	1/4
Data Augmentation	Rotation $+90, +180$ , flip up/down, flip left/right
Total pairs of image/label patches	12,930
Training set	10,736
Validation set	2,194
Test set	15 ROI images
Tissue classes	5



**FIGURE 2. Illustration of dilated convolutions (Atrous Convolution) of different sizes and representation as blocks.** (a) Dilation  $D = 1$ . (b) Dilation  $D = 2$ . (c) Dilation  $D = 4$ . (d) The operation  $Dn = \text{CONV} + \text{BN} + \text{ReLU}$  with a Dilation  $D = n$ . (e) A dilated pyramid block (DPB) with 3 parallel convolutions with dilations of  $D = 1, 2, 3$  [62].

and are part of The Cancer Genome Atlas Program (<https://portal.gdc.cancer.gov>). Whole-slide images (WSIs) were acquired at a magnifications of  $40\times$  ( $n = 138$ ) and  $20\times$  ( $n = 13$ ). Then, regions of interest (ROI) were selected from the WSIs and manually annotated by pathologists, pathology residents, and medical students through a crowd-sourcing process as described in [73]. Figure 1.a,b illustrates the selection of an ROI from a WSI, and annotations associated

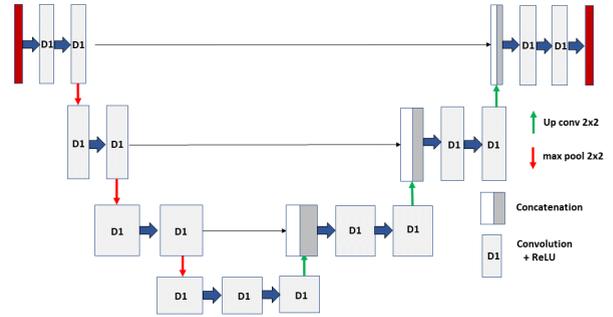


**FIGURE 3. Basic blocks inserted into the original UNet. (a) Residual block adds input with convolution blocks. (b) Dense block, which combines both channels with a concatenation operation. (c) RMS-UNet, a Residual block with convolution dilations of size  $D = 2, 4$ . (d) Our proposal, a block integrated with a DPB3 + Residual + Dense (DRD). (e) ResUNet-a, which includes several parallel convolutions with dilations  $D = 1, 3, 15, 31$  as proposed by [75].**

to the ROI. The maximum and minimum image sizes are  $6813 \times 7360$  and  $1036 \times 1222$  pixels, respectively. The mean ROI size was  $39e6$  pixels, which correspond to  $1.18 \text{ mm}^2$ . A summary of the features of this data set is presented in Table 1.

From the ROI images, 136 were selected to perform the training of the architectures and fifteen were used as a separate test set. Then, from the 136 ROIs 12,930 non-overlapping pairs of data and label patches of dimensions  $256 \times 256$  pixels were generated as illustrated in Figure 1.c. The main characteristics of the patches are shown in Table 2.

The manual annotation of the tissue contained 22 different tissue classes (tumor, stroma, inflammatory, necrosis, glandular secretions, blood, fat, plasma cells, other immune infiltrate, mucoid material, normal acinus or duct, lymphatics, undetermined, nerve, skin, blood vessels, etc.), of which



**FIGURE 4. The UNet architecture as proposed by Ronneberger [11]. It consists of encoder-decoder sections and contains a series of downsampling steps (encoder) obtained by convolutions and downsampling operations, and then upsampling steps (decoder) formed by upsampling plus convolution operations.**

the first four classes (tumor, stroma, inflammatory, necrosis) contained the majority of pixels. All other classes were merged in a single category called *other* following the instructions of the BCSS Challenge. After the merger, the proportions of the five classes were tumor 0.3964, stroma 0.3598, inflammatory 0.1031, necrosis 0.0660, other 0.0747. These proportions suggest a moderate class imbalance and as such, no measures to compensate the class imbalance were taken.

A 3D tensor structure (a cube) was used to represent the ground-truth (GT), i.e., the expected output of a segmented image. This structure has the same size, in terms of pixels, as its corresponding input image, and consists of five channels in depth, which mark the presence of a pathology each pixel.

## B. METHODS

Experimental modifications to UNet architecture were systematically evaluated on segmentation of the BCSS dataset. The modifications focused on two operations: dilated convolutions (Figure 2.a) and residual connections (Figure 3.a). Convolution is a fundamental operation that traverses an image and extracts the values of a neighborhood of pixels, normally contiguous, but it is possible to dilate by a factor of  $n$  and thus cover a larger area as illustrated in Figures 2.b and 2.c. These dilated convolutions are sometimes called Atrous convolutions [56] and produce an increase in the receptive field. The essence of residual connections, also known as skip connections, is to add the output of a processing block with the input of that block. In terms of branches, these would go in parallel as opposed to a single sequential process as illustrated in Figure Figure 3.a. Residual connections were introduced at the computer vision and pattern recognition (CVPR) 2016 conference [74], and have been demonstrated to minimize training errors, overfitting, and vanishing gradient effects [55].

### 1) UNET ARCHITECTURE

The UNet [11] is a convolutional neural network, which was proposed for semantic segmentation. It is a purely

**TABLE 3. Details of the experimental design. A traditional UNet was used as a baseline model, and a series of modifications of the encoder and decoder were applied (ablation studies). All models were trained in the same way: mini-batch of size 8, 10 epochs with a random shuffle of images at every epoch, initial learning rate of 1e-3, adaptive moment estimation optimization algorithm. Columns 2-6 indicate: total number of trainable parameters in millions, number of layers, brief description, figure if applicable and reference if applicable respectively.**

Model	Param.	Layers	Description of the modifications over UNet	Fig.	Ref.
UNet-3 (baseline)	7.69	46	Original UNet architecture of depth 3.	4	[11]
ResUNet-i	7.84	62	Residual block inserted at the input of UNet.	3.a	[55]
ResUNet-e	8.65	66	Residual block inserted at all layers of the encoder, shown in light green	5.a	
ResUNet	15.28	75	Residual UNet with residual blocks at encoder and decoder, shown in green	5.a	
ResUNet-BN	15.28	89	Residual UNet with extra batch normalization (BN) layer, original UNet by Ronneberger does not include BN operation		
P-ResUNet	15.36	96	Purified ResUNet with dilation blocks at the encoder	2.e	[62]
DPB3-i	7.99	62	DPB3 block at the input of the encoder	2.e	[62]
DPB3-e	10.09	74	DPB3 block at all layers of the encoder	2.e	
DPB3-a	21.41	107	DPB3 block at encoder and decoder		
DPB5-i	8.29	74	DPB5 block is similar as DPB3 with dilations of $D = 1, 3, 5$ at the input of the encoder	2.e	
DPB3-i+Res	11.82	80	DPB3 block at the first layer of the encoder and Residual block at all layers of the encoder and decoder		
DPB4-i+Res	11.94	86	DPB4 block is similar to DPB3 with dilations $D = 1, 2, 4$ at the first layer of the encoder and Residual Block at all layers of the encoder and decoder		
ResUNet-a	13.88	119	Residual block with dilated convolution up to $D = 31$ inserted at all layers of the encoder	3.e	[75]
Series-i	7.84	60	Dilations $D = 1, 2, 3$ ; in series at the first layer of the encoder	5.b	[56]
Series-e/d	10.93	75	Dilations with $D = 1, 2, 3$ ; in series at all layers of the encoder	5.c	
RMS-UNet	14.03	85	Residual UNet block with dilations $D = 2, 4$ at all layers of the encoder	3.c	
DRD-UNet	15.40	130	A DPB3 block + Residual block + Dense block at all layers of the encoder	3.d	

**TABLE 4. Performance of sixteen deep learning architectures and the proposed DRD-UNet described in Table 3. Per-class performance is measured with Jaccard Similarity Index and Dice coefficient, Per-image performance is measured in accuracy. Best results are highlighted in bold.**

architectures	Jaccard Similarity Index					Dice coefficient					Accuracy
	Other	Tumor	Stroma	Inflammatory	Necrosis	Other	Tumor	Stroma	Inflammatory	Necrosis	
UNet 3	0.33±0.29	0.73±0.10	0.60±0.17	0.61±0.21	0.15±0.23	0.43±0.33	0.84±0.08	0.73±0.16	0.73±0.22	0.20±0.31	0.80±0.05
ResUNet-i	0.29±0.27	0.70±0.12	0.51±0.20	0.48±0.22	0.14±0.22	0.39±0.32	0.82±0.09	0.65±0.20	0.62±0.23	0.20±0.29	0.73±0.01
ResUNet-e	0.28±0.23	0.70±0.13	0.61±0.16	<b>0.68±0.10</b>	0.19±0.28	0.40±0.29	0.82±0.09	0.74±0.16	<b>0.80±0.07</b>	0.24±0.35	0.78±0.06
ResUNet	0.31±0.27	0.73±0.12	0.59±0.17	0.62±0.18	0.16±0.25	0.41±0.32	0.84±0.09	0.72±0.16	0.75±0.16	0.22±0.33	0.79±0.06
ResUNet-BN	0.27±0.23	0.72±0.13	0.59±0.17	0.61±0.19	0.16±0.27	0.38±0.29	0.83±0.09	0.73±0.16	0.73±0.21	0.20±0.34	0.79±0.05
P-ResUNet	0.30±0.25	0.67±0.18	0.60±0.17	0.62±0.19	0.17±0.26	0.41±0.30	0.79±0.15	0.74±0.16	0.74±0.22	0.22±0.33	0.78±0.07
DPB3-i	0.31±0.25	0.71±0.14	0.60±0.16	0.62±0.19	0.19±0.28	0.42±0.30	0.82±0.11	0.74±0.16	0.74±0.22	0.25±0.35	0.79±0.06
DPB3-e	0.30±0.28	0.73±0.14	0.59±0.18	0.61±0.15	0.15±0.25	0.40±0.33	0.83±0.10	0.73±0.17	0.75±0.12	0.20±0.31	0.78±0.07
DPB3-a	0.31±0.29	0.73±0.15	<b>0.62±0.19</b>	0.62±0.20	<b>0.23±0.31</b>	0.41±0.32	0.84±0.12	0.74±0.19	0.74±0.22	<b>0.29±0.38</b>	<b>0.81±0.06</b>
DPB5-i	0.31±0.25	<b>0.75±0.10</b>	0.59±0.17	0.65±0.12	0.20±0.30	0.42±0.31	<b>0.85±0.07</b>	0.72±0.16	0.78±0.10	0.25±0.36	0.80±0.05
DPB3-i+Res	0.32±0.28	0.66±0.19	0.60±0.16	0.59±0.20	0.21±0.30	0.43±0.33	0.78±0.16	0.73±0.17	0.72±0.21	0.26±0.37	0.78±0.06
DPB4-i+Res	0.26±0.25	0.72±0.13	0.48±0.19	0.53±0.17	0.11±0.19	0.36±0.30	0.83±0.10	0.62±0.20	0.68±0.15	0.16±0.25	0.72±0.11
ResUNet-a	0.31±0.26	0.73±0.14	0.60±0.17	0.60±0.23	0.16±0.25	0.42±0.31	0.84±0.11	0.74±0.17	0.71±0.25	0.21±0.32	0.80±0.07
Series-i	0.28±0.27	0.73±0.13	0.48±0.19	0.57±0.20	0.10±0.17	0.37±0.32	0.84±0.09	0.63±0.20	0.70±0.19	0.15±0.24	0.72±0.11
Series-e/d	0.32±0.28	0.70±0.16	0.61±0.16	0.62±0.20	0.20±0.29	0.42±0.33	0.81±0.13	0.74±0.16	0.74±0.22	0.25±0.36	0.79±0.07
RMS-UNet	0.30±0.25	0.74±0.12	0.60±0.17	0.63±0.20	0.20±0.28	0.40±0.30	<b>0.85±0.08</b>	0.74±0.16	0.75±0.22	0.25±0.35	0.80±0.05
DRD-UNet	<b>0.34±0.28</b>	<b>0.75±0.11</b>	<b>0.62±0.17</b>	0.64±0.20	0.18±0.27	<b>0.44±0.33</b>	<b>0.85±0.08</b>	<b>0.75±0.16</b>	0.75±0.22	0.24±0.35	<b>0.81±0.05</b>

convolutional model, with neither perceptrons nor other type of trainable units, and it is an autoencoder-like architecture (AE). Following the traditional AE model, the UNet compresses its input image with a cascade of encoding blocks composed of convolutional and maxpooling layers. Later, using a cascade of decoding blocks, composed of upsampling and convolutional layers, it decodes the

compressed representation of the input and performs the prediction of its output. As required in deep learning, all convolutions are followed by non-linear ReLU activation functions and a pixel-wise softmax activation function at the output of the network. As shown in Figure 4, by using as many decoding blocks as there are encoding blocks, the UNet model describes a ‘U’ shape, hence its name. The

**TABLE 5. Performance of sixteen deep learning architectures and the proposed DRD-UNet described in Table 3. Per-class performance is measured with Sensitivity and Specificity. Accuracy is repeated from Table 4 for convenience. Best results are highlighted in bold.**

architectures	Sensitivity					Specificity					Accuracy
	Other	Tumor	Stroma	Inflammatory	Necrosis	Other	Tumor	Stroma	Inflammatory	Necrosis	
UNet 3	0.42±0.31	<b>0.94±0.06</b>	0.62±0.20	0.51±0.25	0.55±0.14	0.98±0.02	0.72±0.18	0.88±0.11	<b>0.99±0.00</b>	0.97±0.04	0.80±0.05
ResUNet-i	0.44±0.31	0.88±0.12	0.73±0.14	0.70±0.18	0.71±0.12	0.98±0.01	0.83±0.13	0.86±0.10	0.98±0.01	0.98±0.02	0.73±0.01
ResUNet-e	0.44±0.31	0.88±0.12	0.73±0.14	0.70±0.18	0.71±0.12	0.98±0.01	0.83±0.13	0.86±0.10	0.98±0.01	<b>0.98±0.02</b>	0.78±0.06
ResUNet	0.45±0.32	0.88±0.12	0.73±0.14	0.73±0.22	0.56±0.17	0.98±0.01	0.83±0.14	0.87±0.08	0.97±0.01	<b>0.99±0.00</b>	0.79±0.06
ResUNet-BN	<b>0.56±0.33</b>	0.82±0.14	0.71±0.14	0.87±0.08	0.54±0.32	0.96±0.03	0.89±0.10	<b>0.88±0.06</b>	0.95±0.04	<b>0.99±0.00</b>	0.79±0.05
P-ResUNet	0.48±0.32	0.76±0.21	0.77±0.13	0.83±0.09	0.69±0.15	0.97±0.01	0.91±0.10	0.83±0.10	0.96±0.03	0.97±0.02	0.78±0.07
DPB3-i	0.42±0.30	0.80±0.17	<b>0.78±0.12</b>	0.81±0.11	0.65±0.12	0.98±0.00	0.90±0.09	0.83±0.09	0.96±0.02	0.98±0.01	0.79±0.06
DPB3-e	0.42±0.31	0.86±0.14	0.74±0.15	0.70±0.18	0.74±0.12	0.98±0.01	0.86±0.11	0.86±0.09	0.98±0.01	0.96±0.04	0.78±0.07
DPB3-a	0.36±0.30	0.84±0.17	0.75±0.16	0.86±0.10	0.71±0.15	<b>0.99±0.00</b>	0.87±0.17	0.87±0.07	0.95±0.04	0.98±0.01	<b>0.81±0.06</b>
DPB5-i	0.44±0.33	0.87±0.10	0.71±0.15	0.80±0.17	0.78±0.19	0.98±0.01	0.85±0.14	<b>0.88±0.06</b>	0.96±0.03	0.97±0.02	0.80±0.05
DPB3-i+Res	0.43±0.32	0.72±0.21	0.76±0.13	<b>0.91±0.06</b>	0.69±0.10	0.98±0.01	<b>0.93±0.07</b>	0.82±0.10	0.93±0.06	0.98±0.01	0.78±0.06
DPB4-i+Res	0.39±0.32	0.90±0.12	0.60±0.21	0.55±0.18	0.71±0.13	0.98±0.01	0.80±0.15	0.86±0.13	<b>0.99±0.00</b>	0.93±0.06	0.72±0.11
ResUNet-a	0.44±0.33	0.84±0.15	0.76±0.13	0.81±0.21	0.59±0.18	0.42±0.01	0.88±0.12	0.85±0.10	0.96±0.03	<b>0.98±0.01</b>	0.80±0.07
Series-i	0.32±0.30	0.91±0.11	0.59±0.19	0.61±0.23	0.77±0.09	0.99±0.00	0.81±0.13	0.88±0.11	<b>0.99±0.01</b>	0.91±0.06	0.72±0.11
Series-e/d	0.45±0.31	0.77±0.19	<b>0.78±0.12</b>	0.85±0.11	0.71±0.17	0.98±0.01	0.92±0.07	0.83±0.10	0.96±0.03	0.97±0.02	0.79±0.07
RMS-UNet	0.48±0.32	0.84±0.12	0.72±0.13	0.87±0.08	<b>0.79±0.15</b>	0.97±0.02	0.90±0.08	<b>0.88±0.06</b>	0.96±0.03	0.97±0.01	0.80±0.05
<b>DRD-UNet</b>	0.43±0.32	0.87±0.11	0.76±0.13	0.83±0.11	0.71±0.14	0.98±0.00	0.88±0.10	<b>0.88±0.06</b>	0.96±0.03	0.98±0.01	<b>0.81±0.05</b>

key feature of the UNet is that it makes use of residual connections that concatenate the output of the  $i$ -th encoding block with its corresponding decoding counterpart. These residual connections serve two purposes, they mitigate the possibility of encountering vanishing gradient effects, and they increase the chance of exploiting visual patterns that might prove relevant for prediction but that could have been overlooked by the encoding process [11].

## 2) UNET MODIFICATIONS

The UNet architecture was systematically modified through different architectures, illustrated in Figure 5 and Table 3. The first five architectures implemented in this work were done with the inclusion of the residual connections [55]. The residual connection was evaluated at the input layer of the encoder (ResUNet-i), at all layers of the encoder (ResUNet-e), and at all layers of the encoder and decoder (ResUNet). A modified ResUNet with added batch normalization after the convolutional layer was evaluated in ResUNet-BN. Notice that the original UNet does not include the BN layer.

Inception blocks with standard convolutions of size  $3 \times 3$  were modified to include dilated convolutions of size 2 and 4 as illustrated in Figure 2.a -2.c. Next, different convolution dilations were evaluated by using the DPB block at the input layer of the encoder (DPB3-i), at all layers of the encoder (DPB3-e), at all layers of the encoder and decoder (DPB3-a), and with dilation of  $D = 5$  at the input layer of the encoder (DPB5-i). Combinations of residual connections and Atrous convolutions were also evaluated. Dilations of size  $D = 3$  at the input layer of the encoder with residual connection over all the layers of both the encoder and decoder (DPB3-i+Res), and a combination with an input dilation  $D = 4$  (DPB4i+Res). A ResUNet-a block as presented by [75], used for the segmentation of remote sense data with dilations up to size  $D = 31$ , was evaluated at the encoder. Dilation in series was also evaluated at the input (Series-i) and at the encoder and decoder blocks (Series-e/d). RMS-UNet [56] combines residual connections and Atrous convolution with

$D = 2, 4$  inserted in the residual connection (Figure 3.c). This configuration achieves high performance with minimal loss and low computational cost and was also evaluated in RMS-UNet. Finally, a Dilation DPB3 block, residual and dense block (DRD) as seen in Figure 3.d, is proposed, and evaluated in this work.

## 3) HYPER-PARAMETER EXPLORATION

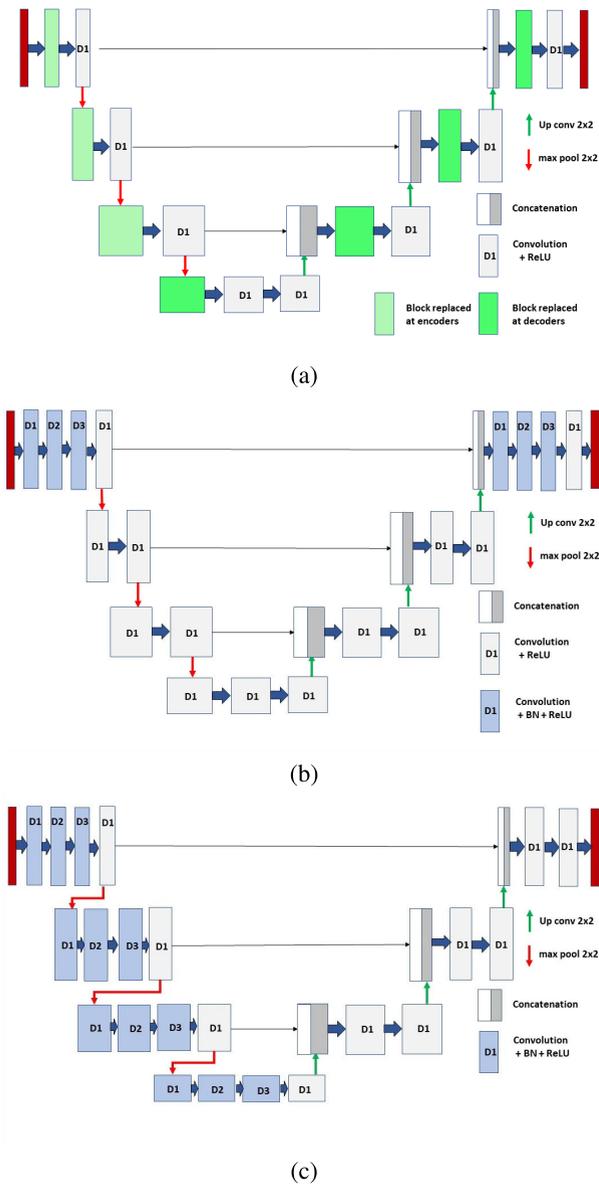
Hyper-parameters were explored on the baseline UNet and then were maintained for all the other architectures. Batch size was evaluated between 1 and 16 and the best results were provided by a mini-batch of size 8. Three optimizers were evaluated: stochastic gradient descent with momentum (sgdm), root mean square propagation (rmsprop), and adaptive moment estimation (adam). Adam provided the best results. Training scenarios of 5, 10 and 30 epochs were evaluated, and 10 epochs were chosen as it provided good results at a fraction of the time. The loss function employed was the Cross Entropy for  $k$  Mutually Exclusive Classes [76]. The same hyper-parameters were used for all subsequent variations of the UNet architecture.

## 4) IMPLEMENTATION DETAILS

All the programming was performed in a computer with processor Intel core i7-7700k, 16 GB RAM, and CPU at 4.20 GHz with an Intel graphics P4000 GPU. The platform used was Matlab<sup>®</sup> version 2023a (The Mathworks<sup>™</sup>, Natick, MA, USA) with the deep learning toolbox. The code is publicly available in the GitHub repository: <https://github.com/mauOrtRuiz/DRD-UNet>.

## 5) PERFORMANCE ESTIMATION

To assess the performance of the architectures proposed in this work, pixels were classified into four categories: True Positive ( $TP$ ), True Negative ( $TN$ ), False Positive ( $FP$ ) and False Negative ( $FN$ ) when compared against the ground truth as illustrated in Figure 6. Where these have been calculated in a *per-class* basis, the sub-index  $i$  has been used, i.e.,  $TP_i$ .



**FIGURE 5. Illustration of the modifications over the original UNet architecture. (a) Input encoders (blocks shaded in light green) and decoders (blocks shaded in darker green) were replaced by the corresponding block under study. (b) Atrous Convolution at input encoder up to  $D = 3$  in series were inserted. (c) Atrous Convolution in series was inserted at every encoder.**

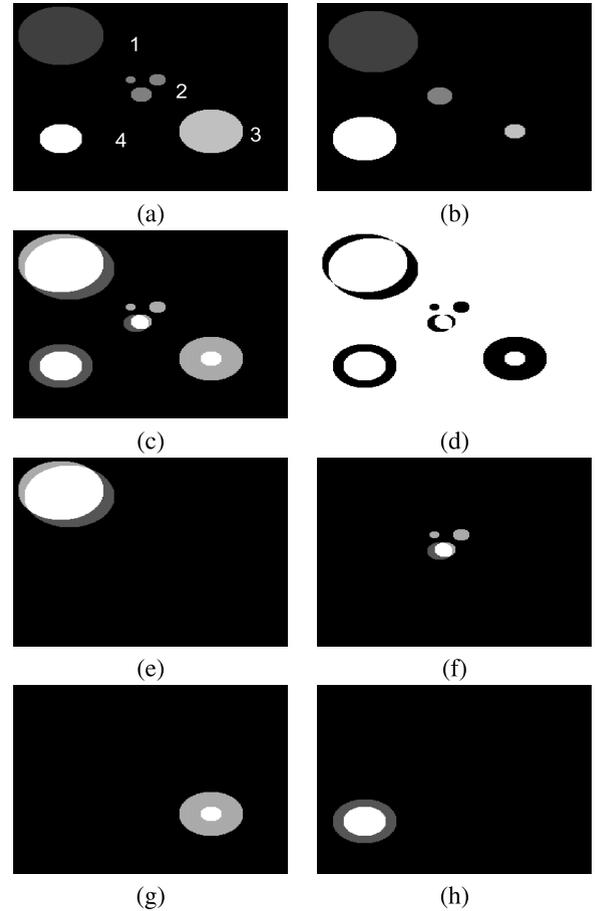
From these, the following pixel-wise, class-wise metrics were calculated: Dice coefficient, Jaccard similarity index (also known as intersection over union), specificity, and sensitivity as follows.

$$\text{Dice coefficient} = \frac{2TP_i}{2TP_i + FP_i + FN_i}, \quad (1)$$

$$\text{Jaccard Similarity Index} = \frac{TP_i}{TP_i + FP_i + FN_i}, \quad (2)$$

$$\text{Sensitivity} = \frac{TP_i}{TP_i + FN_i}, \quad (3)$$

$$\text{Specificity} = \frac{TN_i}{TN_i + FP_i}. \quad (4)$$



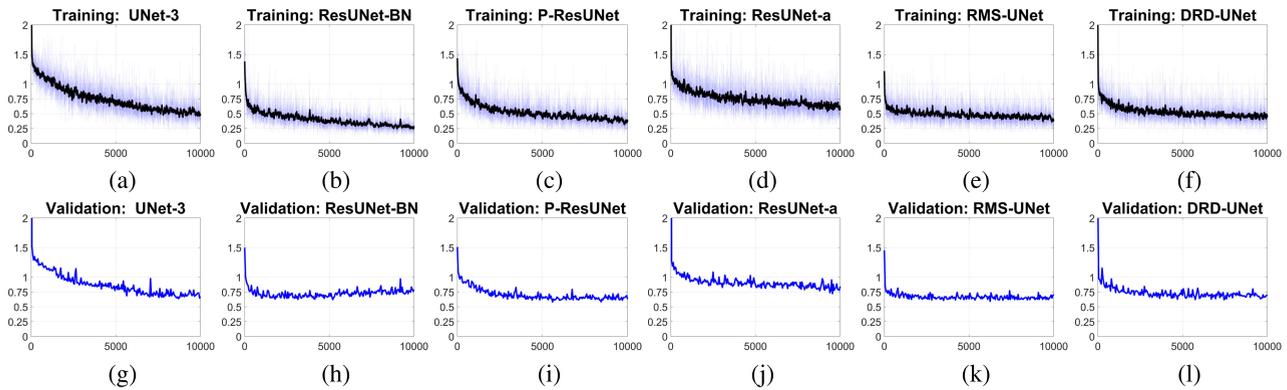
**FIGURE 6. Illustration of calculation of performance metrics with a synthetic image with 4 classes. (a) Ground truth. The classes are denoted by different shades and numbers next to them (the numbers are not part of the image). (b) Estimated image. (c) Ground truth and estimated images overlaid. True Positives (TP) are white, True Negatives (TN) are black, False Positives (FP) are dark gray, False Negatives (FN) are bright gray. (d) Illustration of overall accuracy; pixels correctly estimated are shown in white and incorrect are shown in black. (e-h) Illustration of per-class TP, TN, FP, FN. Notice the FN in class 3 and the FP in class 4.**

In addition, overall accuracy (i.e., not on a per-class basis) was calculated as follows:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}. \quad (5)$$

#### IV. RESULTS AND ABLATION ANALYSIS

From the 12,930 pairs of image and label patches, 10,736 were used to train the architectures previously described and 2,194 were used for validation. Figure 7 shows the training and validation loss curves of the architectures that have been published and DRD-UNet. Then, the trained models were used to perform semantic segmentation of fifteen ROIs, which had been selected as the test set. The performance of each architecture was measured as described previously. The results for Dice, Jaccard and Accuracy are presented in Table 4, the results for Sensitivity and Specificity are presented in Table 5. Results of the semantic segmentation of selected regions of the images are presented in Figures 8 and 9 for all architectures.



**FIGURE 7. Comparison of training loss and validation loss curves for the architectures that have been published UNet-3 [11], ResUNet-BN [55], P-ResUNet [62], ResUNet-a [75], RMS-UNet [56], and DRD-UNet (ours). (a-f) correspond to training loss curves, (g-l) correspond to validation loss curves. For all cases, the training loss is shown in blue with transparency and a weighted average thicker black line is overlaid. DRD-UNet shows a rapid decrease in training loss, similar to RMS-UNet and there is no indication of overfitting in the validation curves. The case with slowest decrease is ResUNet-a, which ends above UNet-3. ResUNet-BN shows a fast decrease in the training, with probably the lowest values, but on the validation, there is an increase, which suggest some overfitting.**

The sixteen models presented in this work are the result of a series of investigations that correspond to a study of ablation by themselves, as they incrementally incorporate variants into the base model. Therefore, we analyze our results from such an ablation perspective.

First, we can notice that adding residual connections (ResUNet) increases the number of parameters but seems to reduce the overall performance, except for class “Inflammatory” where the performance increases. This behavior holds regardless of the section of the UNet where the residual connections are placed. Additionally, notice that Batch Normalization (ResUNet-BN) seems to have random impact, sometimes increasing and sometimes lowering the performance, although just marginally.

The performance rises for classes “stroma” and “necrosis” when the dilated convolution operation is added to the architecture (P-ResUNet). This shows that residual connections with dilated convolutions are a suitable combination. Moreover, when the dilated convolution is used as part of the DPB module, it has a null contribution, except for class “necrosis”. Such results show that the dilated convolution by itself brings little contribution, as opposed to its combination with the residual connections.

Regarding the architectures where the dilation of the convolutions is increased gradually (Series and RMS), it seems that also these architectures improve and lower the performance depending on particular classes but fail to remain consistently better for all cases. It is only the integration of all variants that make a robust architecture (DRD-UNet) that achieves the highest performance for three classes, as well as on average for all five classes.

In terms of Sensitivity and Specificity, the best results were distributed among the architectures, and in several cases, there were draws between several architectures (DRD-UNet, ResUNet-BN, DPB5-i, Series-i) like the specificity for stroma. For necrosis, the specificity results were remarkably close between the best with the best (ResUNet, ResUNet-BN)

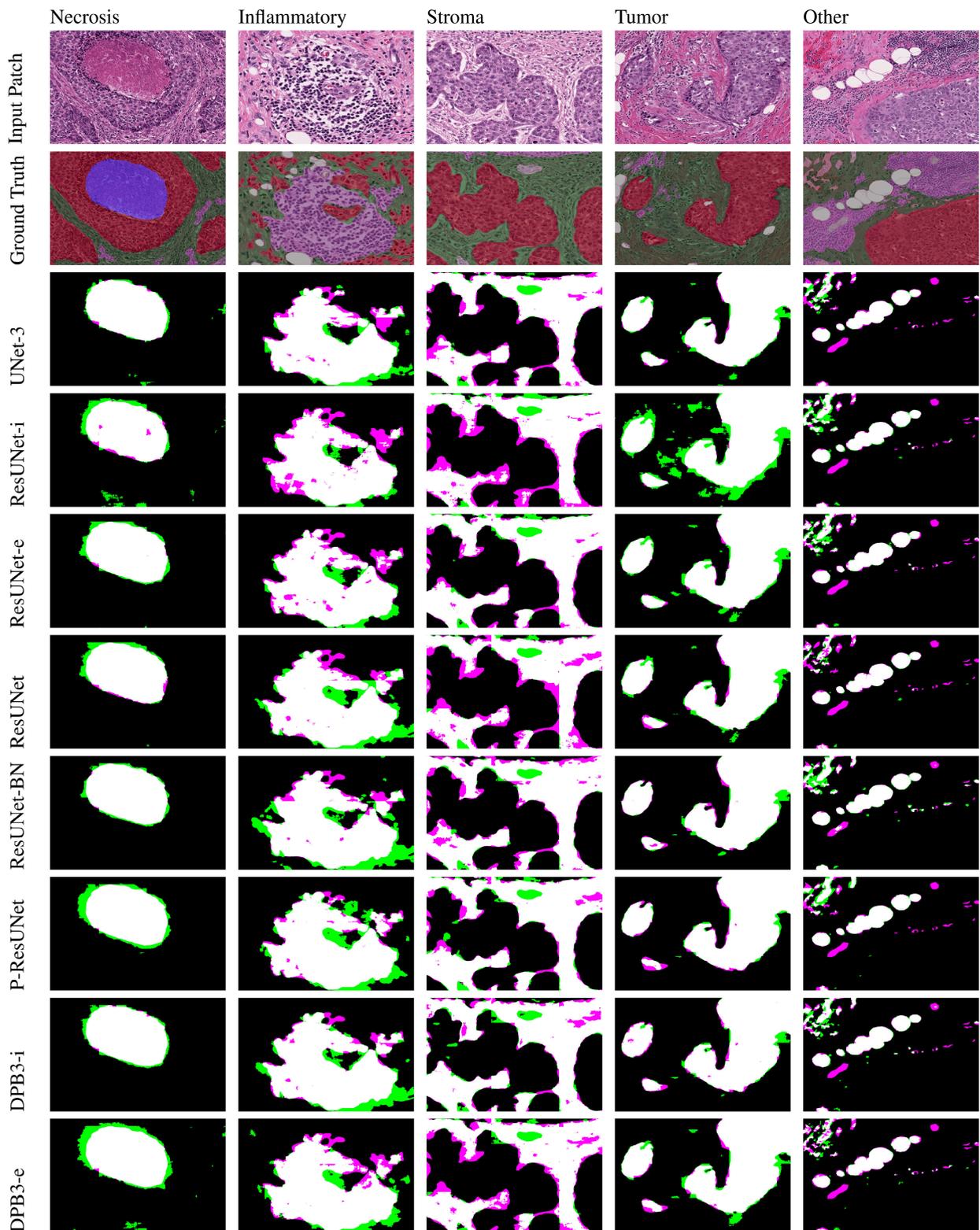
at 0.99 and DRD-UNet 0.98. Similarly, the specificity for Other of DRD-UNet was 0.98 while DPB3-a was 0.99.

Figures 8 and 9 illustrate the semantic segmentation results with selected sections of the images that contain a representative region of each of the five classes in each column and rows correspond to the architectures. TP are labeled in white; TN are labeled in black; FP are labeled in green, and FN are labeled in pink. The best results of DRD-UNet are visible in the smaller green and pink regions for tumor, stroma and other, which correspond to the values of Jaccard and Dice shown in Table 4. For inflammatory and necrosis, the areas are slightly higher than other architectures, but since tumor and inflammatory are the most common classes, these are the ones that have a greater impact on the overall accuracy giving DRD-UNet an advantage.

## V. DISCUSSION

The addition of the proposed DRD block into DRD-UNet provided better results of Jaccard, Dice, and Accuracy in the fifteen images of the BCSS challenge. Whilst the set is relatively small, the results are encouraging and are worth considering in future experimentation. It can be noted the importance of dilated convolutions to increase the receptive field and we proposed the purified DPB3 [62] block with an Atrous convolution of dilation  $D = 1, 2, 3$ . Experimentation revealed there was no significant improvement after adding several residual blocks in different UNet levels. Under the assumption that we can extend performance by adding a deeper layer, a dense block was proposed after the residual. The reduction of training loss indicates a rapid computational convergence in terms of training time. Further work can be done to evaluate this operation over different datasets and other types of cancers and stainings.

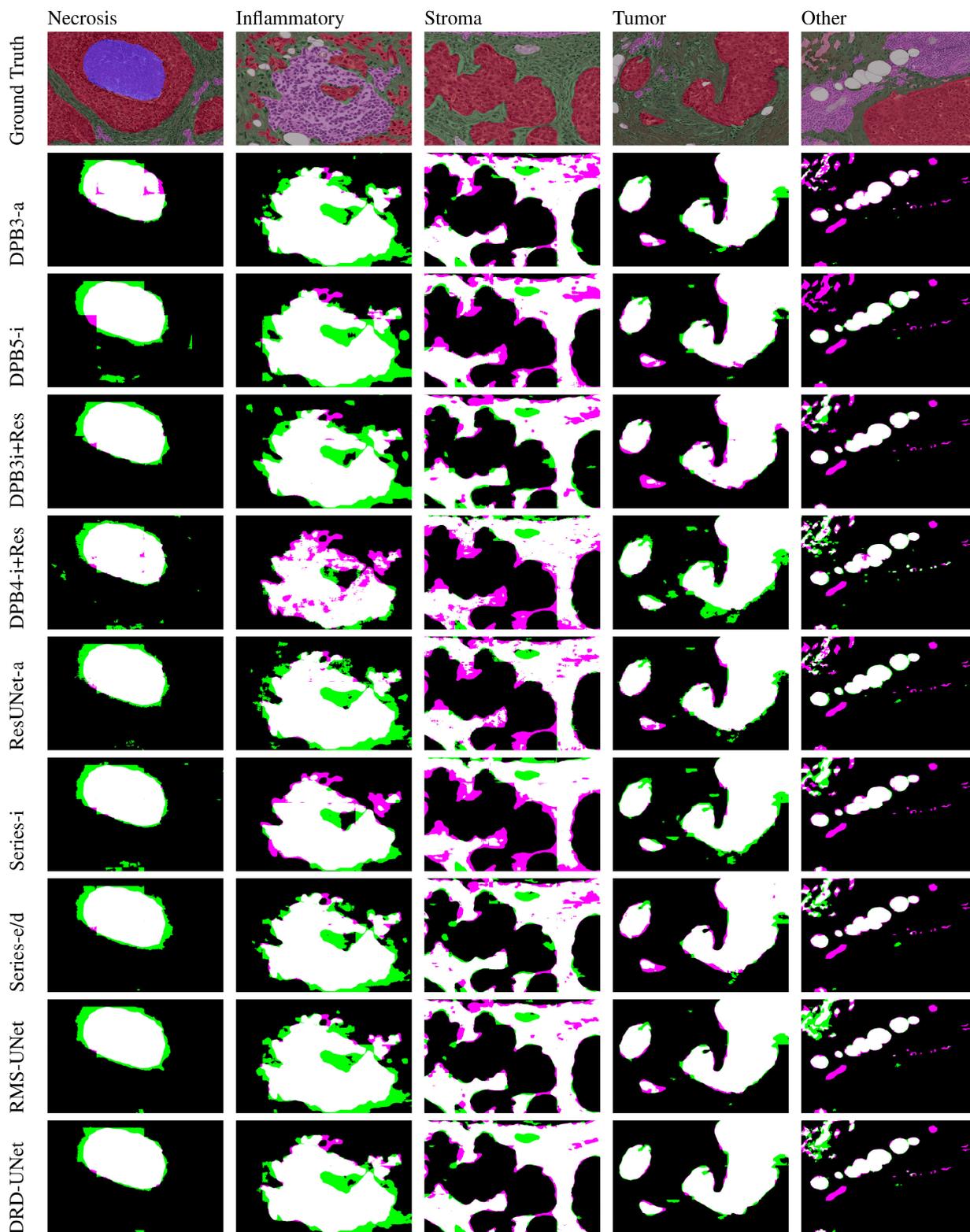
Our analysis was performed on a single dataset that contained only triple-negative breast cancer cases. Whilst this is a limitation in the diversity of cancer cases, we consider that for the purposes of comparing a series of architectures,



**FIGURE 8.** Semantic segmentation results of five different classes with the first 8 models evaluated. The first row shows representative input patches that focus on a region where each class is present. Second row shows the classes of the ground truth overlaid on the input patch with colors: necrosis (dark purple), inflammatory (light purple), stroma (green), tumor (red), and other (gray label). Rows 3 to 10 are the results of the first 8 models. TP are labeled in white; TN are labeled in black; FP are labeled in green, and FN are labeled in pink.

the conclusions in terms of performance should still be valid. However, further studies to confirm the performance

of DRD-UNet with larger datasets and with other cancers would be useful to demonstrate the capabilities of the



**FIGURE 9.** Semantic segmentation results of five different classes with models 9 to 17. The first row shows representative input patches that focus on a region where each class is present. Second row shows the classes of the ground truth overlaid on the input patch with colors: necrosis (dark purple), inflammatory (light purple), stroma (green), tumor (red), and other (gray label). Rows 2 to 10 are the results of the evaluated models. TP are labeled in white; TN are labeled in black; FP are labeled in green, and FN are labeled in pink.

architecture. A closely related observation is that all the images were stained with H&E. Further experimentation with

other stainings should be done to explore the capabilities of the architectures here compared. Second, in some cases

the differences in performance are quite small. Indeed, for specificity several implementations provided exactly the same metrics or varied only by 0.01. Third, in this work we focused on the accuracy of the semantic segmentation provided by sixteen architectures derived from the original UNet. Further work could consider the segmentation and classification, not of tissue regions, but of specific cells (e.g., Tumor Infiltrating Lymphocytes [77], [78]), or nuclei (e.g., [79], [80], [81], [82]), which can provide further information to analyze cancer datasets. Further, by using the nuclei as a starting point, it is possible then determine the extent of cell and from there isolate and analyze other structures like extracellular matrix [83].

## VI. CONCLUSION

In this work, we presented a UNet-like architecture, called DRD-UNet (Dilation, Residual, and Dense block). DRD-UNet was used to perform semantic segmentation of a multi-class breast cancer image dataset provided by the Breast Cancer Semantic Segmentation Challenge. DRD-UNet was compared systematically against the basic UNet architecture and fifteen variations and provided the best results in terms of overall accuracy and Dice coefficient and Jaccard similarity index for three (tumor, stroma, other) of the five classes.

## REFERENCES

- [1] R. Wolz, C. Chu, K. Misawa, M. Fujiwara, K. Mori, and D. Rueckert, "Automated abdominal multi-organ segmentation with subject-specific atlas generation," *IEEE Trans. Med. Imag.*, vol. 32, no. 9, pp. 1723–1730, Sep. 2013.
- [2] M. Gruosso, N. Capece, and U. Erra, "Human segmentation in surveillance video with deep learning," *Multimedia Tools Appl.*, vol. 80, no. 1, pp. 1175–1199, Jan. 2021, doi: 10.1007/s11042-020-09425-0.
- [3] N. R. Pal and S. R. Pal, "A review on image segmentation techniques," *Pattern Recognit.*, vol. 26, no. 9, pp. 1277–1293, 1993.
- [4] S. Minaee, Y. Boykov, F. Porikli, A. Plaza, N. Kehtarnavaz, and D. Terzopoulos, "Image segmentation using deep learning: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 7, pp. 3523–3542, Jul. 2022.
- [5] Y. Yu, C. Wang, Q. Fu, R. Kou, F. Huang, B. Yang, T. Yang, and M. Gao, "Techniques and challenges of image segmentation: A review," *Electronics*, vol. 12, no. 5, p. 1199, Mar. 2023. [Online]. Available: <https://www.mdpi.com/2079-9292/12/5/1199>
- [6] A. Foucart, O. Debeir, and C. Decaestecker, "Shortcomings and areas for improvement in digital pathology image segmentation challenges," *Computerized Med. Imag. Graph.*, vol. 103, Jan. 2023, Art. no. 102155.
- [7] Y. Arafat and C. C. Reyes-Aldasoro, "Computational image analysis techniques, programming languages and software platforms used in cancer research: A scoping review," in *Medical Image Understanding and Analysis (Lecture Notes in Computer Science)*, G. Yang, A. Aviles-Rivero, M. Roberts, and C.-B. Schönlieb, Eds. Cham, Switzerland: Springer, 2022, pp. 833–847.
- [8] M. H. Hesamian, W. Jia, X. He, and P. Kennedy, "Deep learning techniques for medical image segmentation: Achievements and challenges," *J. Digit. Imag.*, vol. 32, no. 4, pp. 582–596, Aug. 2019, doi: 10.1007/s10278-019-00227-x.
- [9] R. Wang, T. Lei, R. Cui, B. Zhang, H. Meng, and A. K. Nandi, "Medical image segmentation using deep learning: A survey," *IET Image Process.*, vol. 16, no. 5, pp. 1243–1267, Apr. 2022, doi: 10.1049/ipr2.12419.
- [10] R. Jahandideh, A. T. Targhi, and M. Tahmasbi, "Physical attribute prediction using deep residual neural networks," 2018, *arXiv:1812.07857*.
- [11] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, in Lecture Notes in Computer Science, vol. 9351, 2015, pp. 234–241.
- [12] J. Ye, H. Wang, Z. Huang, Z. Deng, Y. Su, C. Tu, Q. Wu, Y. Yang, M. Wei, J. Niu, and J. He, "Exploring vanilla U-Net for lesion segmentation from whole-body FDG-PET/CT scans," 2022, *arXiv:2210.07490*.
- [13] V. Ummadi, "U-Net and its variants for medical image segmentation: A short review," 2022, *arXiv:2204.08470*.
- [14] O. Ali, H. Ali, S. A. A. Shah, and A. Shahzad, "Implementation of a modified U-Net for medical image segmentation on edge devices," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 69, no. 11, pp. 4593–4597, Nov. 2022.
- [15] H. Huang, L. Lin, R. Tong, H. Hu, Q. Zhang, Y. Iwamoto, X. Han, Y.-W. Chen, and J. Wu, "UNet 3+: A full-scale connected UNet for medical image segmentation," 2020, *arXiv:2004.08790*.
- [16] J. N. Kather, J. Krisam, P. Charoentong, T. Luedde, E. Herpel, C.-A. Weis, T. Gaiser, A. Marx, N. A. Valous, D. Ferber, L. Jansen, C. C. Reyes-Aldasoro, I. Zörnig, D. Jäger, H. Brenner, J. Chang-Claude, M. Hoffmeister, and N. Halama, "Predicting survival from colorectal cancer histology slides using deep learning: A retrospective multicenter study," *PLOS Med.*, vol. 16, no. 1, Jan. 2019, Art. no. e1002730.
- [17] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2019," *CA, Cancer J. Clinicians*, vol. 69, no. 1, pp. 7–34, Jan. 2019.
- [18] C. Coleman, "Early detection and screening for breast cancer," *Seminars Oncol. Nursing*, vol. 33, no. 2, pp. 141–155, May 2017.
- [19] R. Lozano, R. Lozano, M. Naghavi, K. Foreman, S. Lim, K. Shibuya, V. Aboyans, J. Abraham, T. Adair, R. Aggarwal, and S. Y. Ahn, "Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: A systematic analysis for the global burden of disease study 2010," *Lancet*, vol. 380, no. 9859, pp. 2095–2128, 2012.
- [20] A. Dibden, J. Offman, S. W. Duffy, and R. Gabe, "Worldwide review and meta-analysis of cohort studies measuring the effect of mammography screening programmes on incidence-based breast cancer mortality," *Cancers*, vol. 12, no. 4, p. 976, Apr. 2020.
- [21] A. Conti, A. Duggento, I. Indovina, M. Guerrisi, and N. Toschi, "Radiomics in breast cancer classification and prediction," *Seminars Cancer Biol.*, vol. 72, pp. 238–250, Jul. 2021.
- [22] M. Lberg, M. L. Lousdal, M. Brethauer, and M. Kalager, "Benefits and harms of mammography screening," *Breast Cancer Res.*, vol. 17, no. 1, p. 63, May 2015.
- [23] B. C. Calhoun and C. A. Livasy, "Mitigating overdiagnosis and overtreatment in breast cancer: What is the role of the pathologist?" *Arch. Pathol. Lab. Med.*, vol. 138, no. 11, pp. 1428–1431, Nov. 2014.
- [24] E. A. Rakha and F. G. Pareja, "New advances in molecular breast cancer pathology," *Seminars Cancer Biol.*, vol. 72, pp. 102–113, Jul. 2021.
- [25] S. Jenkins, M. E. Kachur, K. Rechache, J. M. Wells, and S. Lipkowitz, "Rare breast cancer subtypes," *Current Oncol. Rep.*, vol. 23, no. 5, p. 54, Mar. 2021.
- [26] M. Veta, J. P. W. Pluim, P. J. van Diest, and M. A. Viergever, "Breast cancer histopathology image analysis: A review," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 5, pp. 1400–1411, May 2014.
- [27] N. Kumar, R. Gupta, and S. Gupta, "Whole slide imaging (WSI) in pathology: Current perspectives and future directions," *J. Digit. Imag.*, vol. 33, no. 4, pp. 1034–1040, Aug. 2020.
- [28] J. N. Nyengar, "Whole slide imaging: The futurescape of histopathology," *Indian J. Pathol. Microbiol.*, vol. 64, no. 1, pp. 8–13, 2021.
- [29] W. R. Drioua, N. Benamrane, and L. Sais, "Breast cancer histopathological images segmentation using deep learning," *Sensors*, vol. 23, no. 17, p. 7318, Aug. 2023. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10490494/>
- [30] A. Madabhushi, "Digital pathology image analysis: Opportunities and challenges," *Imag. Med.*, vol. 1, no. 1, pp. 7–10, Oct. 2009.
- [31] J. Słodkowska and M. García-Rojo, "Digital pathology in personalized cancer therapy," *Studies Health Technol. Inform.*, vol. 179, pp. 143–154, Jan. 2012.
- [32] P. Siregar, N. Julen, P. Hufnagl, and G. L. Mutter, "Computational morphogenesis—Embryogenesis, cancer research and digital pathology," *Biosystems*, vols. 169–170, pp. 40–54, Jul. 2018.
- [33] B. J. Williams, J. Lee, K. A. Oien, and D. Treanor, "Digital pathology access and usage in the U.K.: Results from a national survey on behalf of the national cancer research Institute's CM-path initiative," *J. Clin. Pathol.*, vol. 71, no. 5, pp. 463–466, May 2018.
- [34] A. V. Parwani, "Digital pathology enhances cancer diagnostics," *Med. Lab. Observer*, vol. 49, no. 2, p. 25, Feb. 2017.

- [35] J. T. Kwak and S. M. Hewitt, "Multiview boosting digital pathology analysis of prostate cancer," *Comput. Methods Programs Biomed.*, vol. 142, pp. 91–99, Apr. 2017.
- [36] A. Heindl, S. Nawaz, and Y. Yuan, "Mapping spatial heterogeneity in the tumor microenvironment: A new era for digital pathology," *Lab. Invest.*, vol. 95, no. 4, pp. 377–384, Apr. 2015.
- [37] R. Pell, K. Oien, M. Robinson, H. Pitman, N. Rajpoot, J. Rittscher, D. Snead, and C. Verrill, "The use of digital pathology and image analysis in clinical trials," *J. Pathol., Clin. Res.*, vol. 5, no. 2, pp. 81–90, Apr. 2019.
- [38] K. Bera, K. A. Schalper, D. L. Rimm, V. Velcheti, and A. Madabhushi, "Artificial intelligence in digital pathology—New tools for diagnosis and precision oncology," *Nature Rev. Clin. Oncol.*, vol. 16, no. 11, pp. 703–715, Nov. 2019.
- [39] R. Sparks and A. Madabhushi, "Statistical shape model for manifold regularization: Gleason grading of prostate histology," *Comput. Vis. Image Understand.*, vol. 117, no. 9, pp. 1138–1146, Sep. 2013.
- [40] K. Dimitropoulos, P. Barmopoulos, C. Zioga, A. Kamas, K. Patsiaoura, and N. Grammalidis, "Grading of invasive breast carcinoma through grassmannian VLAD encoding," *PLoS ONE*, vol. 12, no. 9, Sep. 2017, Art. no. e0185110.
- [41] J. Jørgensen, J. Emborg, R. Røge, and L. Østergaard, "Exploiting multiple color representations to improve colon cancer detection in whole slide H&E stains," in *Proc. 1st Int. Workshop Comput. Pathol. (COMPAY)*, vol. 11039. Granada, Spain: Springer, Sep. 2018, pp. 61–68.
- [42] S. Saxena and M. Gyanchandani, "Machine learning methods for computer-aided breast cancer diagnosis using histopathology: A narrative review," *J. Med. Imag. Radiat. Sci.*, vol. 51, no. 1, pp. 182–193, Mar. 2020.
- [43] F. Martino, S. Varricchio, D. Russo, F. Merolla, G. Iardi, M. Mascolo, G. O. dell'Aversana, L. Califano, G. Toscano, G. De Pietro, M. Frucci, N. Brancati, F. Frassetto, and S. Staibano, "A machine-learning approach for the assessment of the proliferative compartment of solid tumors on hematoxylin-eosin-stained sections," *Cancers*, vol. 12, no. 5, p. 1344, May 2020.
- [44] E. H. Adelson, C. H. Anderson, J. R. Bergen, P. J. Burt, and J. M. Ogden, "Pyramid methods in image processing," *RCA Eng.*, vol. 29, no. 6, pp. 33–41, 1984.
- [45] P. Burt and E. Adelson, "The Laplacian pyramid as a compact image code," *IEEE Trans. Commun.*, vol. COM-31, no. 4, pp. 532–540, Apr. 1983. [Online]. Available: <https://ieeexplore.ieee.org/document/1095851>
- [46] M. Spann and R. Wilson, "A quad-tree approach to image segmentation which combines statistical and spatial information," *Pattern Recognit.*, vol. 18, nos. 3–4, pp. 257–269, Jan. 1985. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/0031320385900512>
- [47] K. Fukushima, "Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position," *Biol. Cybern.*, vol. 36, no. 4, pp. 193–202, Apr. 1980, doi: [10.1007/bf00344251](https://doi.org/10.1007/bf00344251).
- [48] Y. Lecun and Y. Bengio, *Convolutional Networks for Images, Speech and Time Series*. Cambridge, MA, USA: MIT Press, Oct. 1995, pp. 255–258.
- [49] S. Guan, N. Kamona, and M. Loew, "Segmentation of thermal breast images using convolutional and deconvolutional neural networks," in *Proc. IEEE Appl. Imag. Pattern Recognit. Workshop (AIPR)*, Oct. 2018, pp. 1–7. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/8707379>
- [50] E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 640–651, Apr. 2017. [Online]. Available: <https://ieeexplore.ieee.org/document/7478072>
- [51] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," 2017, *arXiv:1706.05587*.
- [52] L. Xu, H. Xue, M. Bennamoun, F. Boussaid, and F. Sohel, "Atrous convolutional feature network for weakly supervised semantic segmentation," *Neurocomputing*, vol. 421, pp. 115–126, Jan. 2021. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0925231220314533>
- [53] M. K. A. AnbuDevi and K. Suganthi, "Review of semantic segmentation of medical images using modified architectures of UNET," *Diagnostics*, vol. 12, no. 12, p. 3064, Dec. 2022. [Online]. Available: <https://www.mdpi.com/2075-4418/12/12/3064>
- [54] V. Ummadi, "U-Net and its variants for medical image segmentation : A short review," 2022, *arXiv:2204.08470*.
- [55] X. Han, "Automatic liver lesion segmentation using a deep convolutional neural network method," *Med. Phys.*, vol. 44, no. 4, pp. 1408–1419, Apr. 2017.
- [56] R. A. Khan, Y. Luo, and F.-X. Wu, "RMS-UNet: Residual multi-scale UNet for liver and lesion segmentation," *Artif. Intell. Med.*, vol. 124, Feb. 2022, Art. no. 102231. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0933365721002244>
- [57] X. Xiao, Y. Qiang, J. Zhao, X. Yang, and X. Yang, "Segmentation of liver lesions without contrast agents with radiomics-guided densely UNet-nested GAN," *IEEE Access*, vol. 9, pp. 2864–2878, 2021. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/9309035>
- [58] Y. Ding, J. Sohn, M. Kawczynski, H. Trivedi, R. Harnish, N. Jenkins, D. Lituiev, T. Copeland, M. Aboian, C. Aparici, S. Behr, R. Flavell, S.-Y. Huang, K. Zalocusky, L. Nardo, Y. Seo, R. Hawkins, M. Pampaloni, D. Hadley, and B. Franc, "A deep learning model to predict a diagnosis of Alzheimer disease by using <sup>18</sup>F-FDG PET of the brain," *Radiology*, vol. 290, no. 3, pp. 456–464, Mar. 2019.
- [59] P. Ahmad, H. Jin, R. Alroobaea, S. Qamar, R. Zheng, F. Alnajjar, and F. Aboudi, "MH UNet: A multi-scale hierarchical based architecture for medical image segmentation," *IEEE Access*, vol. 9, pp. 148384–148408, 2021.
- [60] Y. Zhang, J. Wu, Y. Liu, Y. Chen, E. X. Wu, and X. Tang, "MI-UNet: Multi-inputs UNet incorporating brain parcellation for stroke lesion segmentation from T1-weighted magnetic resonance images," *IEEE J. Biomed. Health Inform.*, vol. 25, no. 2, pp. 526–535, Feb. 2021. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/9099078>
- [61] R. Sun, E. J. Limkin, M. Vakalopoulou, L. Dercle, S. Champiat, S. R. Han, L. Verlingue, D. Brandao, A. Lancia, S. Ammari, A. Hollebecque, J.-Y. Scoazec, A. Marabelle, C. Massard, J.-C. Soria, C. Robert, N. Paragios, E. Deutsch, and C. Ferte, "A radiomics approach to assess tumour-infiltrating CD8 cells and response to anti-PD-1 or anti-PD-L1 immunotherapy: An imaging biomarker, retrospective multicohort study," *Lancet Oncol.*, vol. 19, no. 9, pp. 1180–1191, Sep. 2018.
- [62] K. Niu, Z. Guo, X. Peng, and S. Pei, "P-ResUnet: Segmentation of brain tissue with purified residual unet," *Comput. Biol. Med.*, vol. 151, Dec. 2022, Art. no. 106294. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0010482522010022>
- [63] G. Lian, "Rotation invariant color texture classification using multiple sub-DLBPs," *J. Vis. Commun. Image Represent.*, vol. 31, pp. 1–13, Aug. 2015.
- [64] Y. Yang, W. Wan, S. Huang, X. Zhong, and X. Kong, "RADCU-Net: Residual attention and dual-supervision cascaded U-Net for retinal blood vessel segmentation," *Int. J. Mach. Learn. Cybern.*, vol. 14, no. 5, pp. 1605–1620, May 2023, doi: [10.1007/s13042-022-01715-3](https://doi.org/10.1007/s13042-022-01715-3).
- [65] J. Wang, X. Li, P. Lv, and C. Shi, "SERR-U-Net: Squeeze-and-excitation residual and recurrent block-based U-Net for automatic vessel segmentation in retinal image," *Comput. Math. Methods Med.*, vol. 2021, pp. 1–16, Aug. 2021. [Online]. Available: <https://www.hindawi.com/journals/cmmm/2021/5976097/>
- [66] J. Hu, H. Wang, S. Gao, M. Bao, T. Liu, Y. Wang, and J. Zhang, "S-UNet: A bridge-style U-Net framework with a saliency mechanism for retinal vessel segmentation," *IEEE Access*, vol. 7, pp. 174167–174177, 2019. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/8842560>
- [67] Z. Zeng, W. Xie, Y. Zhang, and Y. Lu, "RIC-UNet: An improved neural network based on UNet for nuclei segmentation in histology images," *IEEE Access*, vol. 7, pp. 21420–21428, 2019. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/8632907v>
- [68] H. Yuan, Z. Liu, Y. Shao, and M. Liu, "ResD-UNet research and application for pulmonary artery segmentation," *IEEE Access*, vol. 9, pp. 67504–67511, 2021. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/9408588>
- [69] W. Yue, H. Zhang, J. Zhou, G. Li, Z. Tang, Z. Sun, J. Cai, N. Tian, S. Gao, J. Dong, Y. Liu, X. Bai, and F. Sheng, "Deep learning-based automatic segmentation for size and volumetric measurement of breast cancer on magnetic resonance imaging," *Frontiers Oncol.*, vol. 12, Aug. 2022, Art. no. 984626, doi: [10.3389/fonc.2022.984626](https://doi.org/10.3389/fonc.2022.984626).
- [70] A. Baccouche, B. Garcia-Zapirain, C. Castillo Olea, and A. S. Elmaghraby, "Connected-UNets: A deep learning architecture for breast mass segmentation," *NPJ Breast Cancer*, vol. 7, no. 1, pp. 1–12, Dec. 2021. [Online]. Available: <https://www.nature.com/articles/s41523-021-00358-x>

- [71] A. Negi, A. N. J. Raj, R. Nersisson, Z. Zhuang, and M. Murugappan, "RDA-UNET-WGAN: An accurate breast ultrasound lesion segmentation using wasserstein generative adversarial networks," *Arabian J. Sci. Eng.*, vol. 45, no. 8, pp. 6399–6410, Aug. 2020, doi: 10.1007/s13369-020-04480-z.
- [72] T. Alam, W.-C. Shia, F.-R. Hsu, and T. Hassan, "Improving breast cancer detection and diagnosis through semantic segmentation using the UNet3+ deep learning framework," *Biomedicines*, vol. 11, no. 6, p. 1536, May 2023. [Online]. Available: <https://www.mdpi.com/2227-9059/11/6/1536>
- [73] M. Amgad, "Structured crowdsourcing enables convolutional segmentation of histology images," *Bioinformatics*, vol. 35, no. 18, pp. 3461–3467, Sep. 2019.
- [74] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," 2015, *arXiv:1512.03385*.
- [75] F. I. Diakogiannis, F. Waldner, P. Caccetta, and C. Wu, "ResUNet-A: A deep learning framework for semantic segmentation of remotely sensed data," *ISPRS J. Photogramm. Remote Sens.*, vol. 162, pp. 94–114, Apr. 2020. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0924271620300149>
- [76] S. J. Prince, *Understanding Deep Learning*. Cambridge, MA, USA: MIT Press, 2023. [Online]. Available: <http://udlbook.com>
- [77] F. Klauschen, "Scoring of tumor-infiltrating lymphocytes: From visual estimation to machine learning," *Seminars Cancer Biol.*, vol. 52, pp. 151–157, Oct. 2018.
- [78] X. Zhang, X. Zhu, K. Tang, Y. Zhao, Z. Lu, and Q. Feng, "DDTNet: A dense dual-task network for tumor-infiltrating lymphocyte detection and segmentation in histopathological images of breast cancer," *Med. Image Anal.*, vol. 78, May 2022, Art. no. 102415.
- [79] S. Lal, D. Das, K. Alabhya, A. Kanfode, A. Kumar, and J. Kini, "NucleiSegNet: Robust deep learning architecture for the nuclei segmentation of liver cancer histopathology images," *Comput. Biol. Med.*, vol. 128, Jan. 2021, Art. no. 104075.
- [80] I. Ahmad, S. M. Israr, and Z. Ul Islam, "A three in one bottom-up framework for simultaneous semantic segmentation, instance segmentation and classification of multi-organ nuclei in digital cancer histology," 2023, *arXiv:2308.11179*.
- [81] I. Ahmad, Y. Xia, H. Cui, and Z. U. Islam, "DAN-NucNet: A dual attention based framework for nuclei segmentation in cancer histology images under wild clinical conditions," *Expert Syst. Appl.*, vol. 213, Mar. 2023, Art. no. 118945.
- [82] S. Graham, Q. D. Vu, S. E. A. Raza, A. Azam, Y. W. Tsang, J. T. Kwak, and N. Rajpoot, "Hover-Net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images," *Med. Image Anal.*, vol. 58, Dec. 2019, Art. no. 101563.
- [83] A. Aggarwal, S. Khalighi, D. Babu, H. Li, S. Azarianpour-Esfahani, G. Corredor, P. Fu, M. Mokhtari, T. Pathak, E. Thayer, S. Modesitt, H. Mahdi, S. Avril, and A. Madabhushi, "Computational pathology identifies immune-mediated collagen disruption to predict clinical outcomes in gynecologic malignancies," *Commun. Med.*, vol. 4, no. 1, pp. 1–12, Jan. 2024.



**CEFA KARABAĞ** was born in Arpaçay, Türkiye, in 1978. He received the B.Sc. degree in physics from the University of İstanbul, in 1998, the M.Sc. (German Diplom) degree in physics from Leibniz Universität Hannover, in 2002, and the Ph.D. degree in electrical and electronic engineering from the City, University of London, in 2020. He was a Lecturer of mathematics with the Centre for Applied Sciences, from 2013 to 2021. He was a Lecturer of computer science with the Department of Computer Science, City, University of London, from 2021 to 2023. He is currently a Senior Lecturer of computer science and applied computing with London Metropolitan University. His research interests include medical image analysis and computing, image processing, and data visualization. He is a member of The British Machine Vision Association, The British Association for Cancer Research, and The European Association for Cancer Research. He is also a Reviewer of *PLOS One* and IEEE ACCESS.



**EDGAR ROMAN-RANGEL** received the Master Diploma degree from Instituto Tecnológico y de Estudios Superiores de Monterrey (ITESM), Mexico, in 2006, and the Ph.D. degree from École Polytechnique Fédérale de Lausanne (EPFL), Switzerland, in 2013. He is currently a Professor with the Academic Department of Computer Science, Instituto Tecnológico Autónomo de México (ITAM), where he conducts research in computer vision, representation learning, and cross-modal signal processing, applied to cultural and societal challenges. Previous to his current appointment, he was a Postdoctoral Researcher with the CVM-Laboratory, University of Geneva, Switzerland, from 2013 to 2017, with a break from 2014 to 2015, in which he was a Visiting Researcher with the National Institute of Anthropology and History (INAH), Mexico.



**CONSTANTINO CARLOS REYES-ALDASORO** (Senior Member, IEEE) received the B.Sc. degree in electrical engineering from Universidad Nacional Autónoma de México, in 1993, the M.Sc. degree in communications and signal processing from the Imperial College of Science Technology and Medicine, in 1994, and the Ph.D. degree in computer science from Warwick University, in 2005. He was a Postdoctoral Research Associate with the Tumor Microcirculation Group, Medical School, University of Sheffield, from 2005 to 2011. He was with Sussex University, from 2011 to 2013, and joined the City, University of London, in 2013. His research interests include image analysis of biomedical images with special interest in cancer, inflammation, and microcirculation. He has worked with data from x-rays, computed tomography, magnetic resonance imaging, light, fluorescent, confocal, multiphoton, and electron microscopy. He served as the Chair for the Vision and Imaging Technical Network of the Institute of Engineering and Technology and a member for the Executive Committee of the British Association for Cancer Research and the Data Analysis in Imaging Section of the Royal Microscopical Society. He was the Chair of the Medical Image Understanding and Analysis (MIUA) Conference in 2014 and The British Machine Vision Conference (BMVC) in 2022.



**MAURICIO ALBERTO ORTEGA-RUIZ** was born in Mexico City. He received the B.Sc. degree in electrical engineering from Universidad Nacional Autónoma de México, in 1994, and the M.Sc. degree in communications and signal processing from the Imperial College of Science Technology and Medicine, in 1995. He was responsible for the seismic instrumentation network with CENAPRED, Mexico City, and trained in specialized instrumentation systems in Japan. As a Service Engineer and a Service Coordinator with Gretag Imaging Company, he was responsible for technical service. He is currently a Senior Lecturer of electronics and signal processing with UVM University, Mexico, a Researcher with CIDETEC, UVM, and has been the Ph.D. Researcher with the City, University of London, since 2018. He has been the Co-Founder of the Research Group DigPatho, for the Americas pathology laboratories region. His research interest includes deep learning methods for breast cancer imaging analysis. He has been a member of the Medical Image Computing and Computer Assisted Intervention (MICCAI).