

Review

# Review of Semantic Segmentation of Medical Images Using Modified Architectures of UNET

M. Krithika alias AnbuDevi and K. Suganthi \* 

Vellore Institute of Technology, Chennai 600127, India

\* Correspondence: suganthi.k@vit.ac.in

**Abstract:** In biomedical image analysis, information about the location and appearance of tumors and lesions is indispensable to aid doctors in treating and identifying the severity of diseases. Therefore, it is essential to segment the tumors and lesions. MRI, CT, PET, ultrasound, and X-ray are the different imaging systems to obtain this information. The well-known semantic segmentation technique is used in medical image analysis to identify and label regions of images. The semantic segmentation aims to divide the images into regions with comparable characteristics, including intensity, homogeneity, and texture. UNET is the deep learning network that segments the critical features. However, UNETs basic architecture cannot accurately segment complex MRI images. This review introduces the modified and improved models of UNET suitable for increasing segmentation accuracy.

**Keywords:** UNET; semantic segmentation; dice similarity coefficient; CNN; MRI



**Citation:** Krithika alias AnbuDevi, M.; Suganthi, K. Review of Semantic Segmentation of Medical Images Using Modified Architectures of UNET. *Diagnostics* **2022**, *12*, 3064. <https://doi.org/10.3390/diagnostics12123064>

Academic Editor: Costin Teodor Streba

Received: 8 October 2022

Accepted: 22 November 2022

Published: 6 December 2022

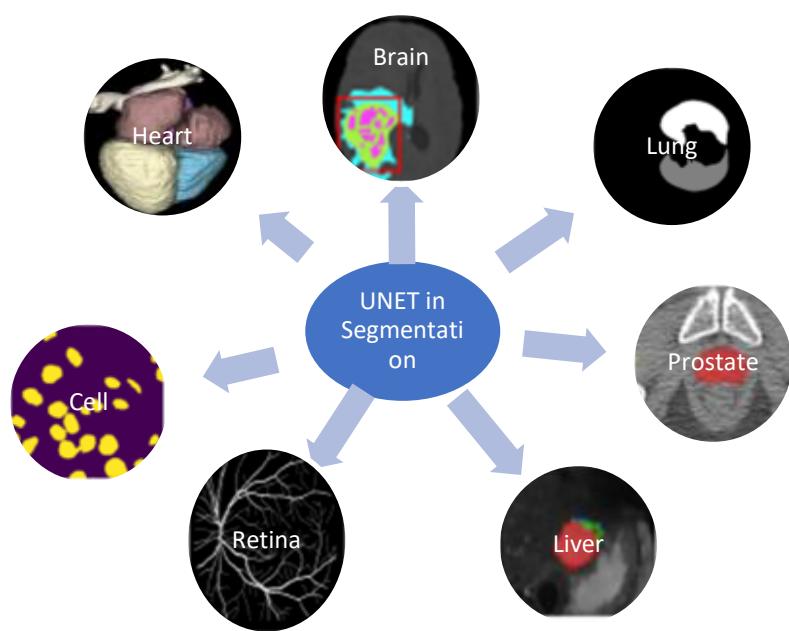
**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

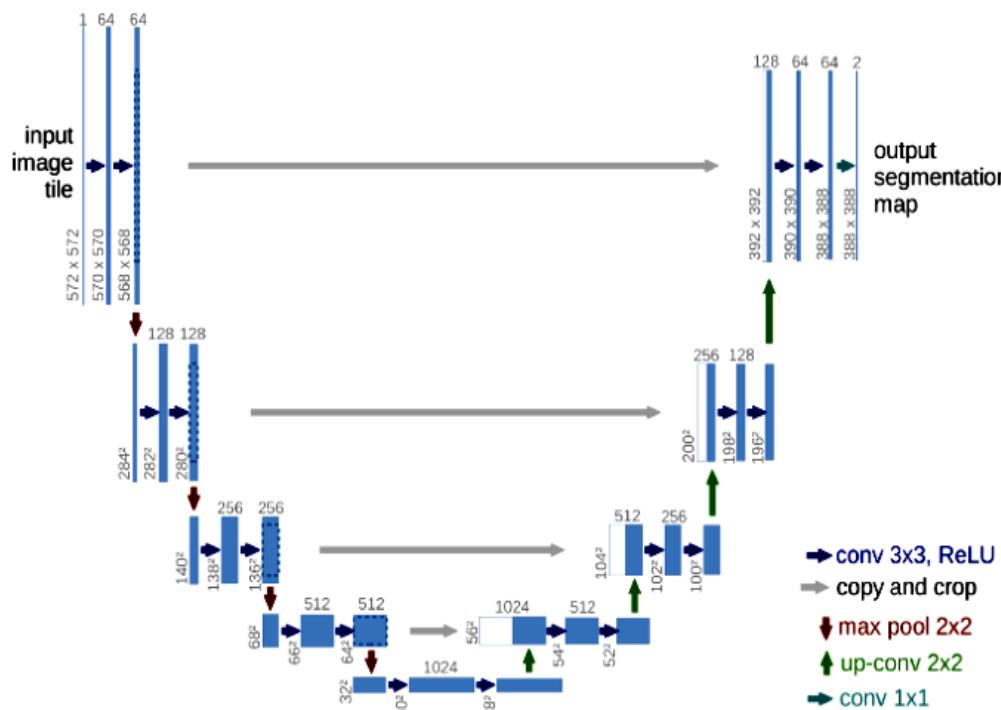
## 1. Introduction

Principal component analysis [1], fuzzy c-means Hsieh [2], Gabor filter [3], and multi-level fuzzy c-means [4] are examples of traditional machine learning techniques. However, the performance of these algorithms in the field of computer vision is not sufficient. Therefore, deep learning is now widely employed in various industries [5–13], for example, to tackle problems in computer vision and succeed in image recognition. Deep learning techniques are used to assess complex and diverse pathological images. Deep learning techniques can learn coarse and fine representations in all layers and perform end-to-end learning. There are the following two basic frameworks: CNN and the FCN for segmentation. Convolutional neural networks (CNN) perform well in classifying images and significantly improve segmentation. Initially, the categorization of image patches was a widely used deep learning approach where each pixel was sorted into matching categories separately by employing image blocks around each pixel. On the other hand, the FCN framework expands the fundamental CNN structure without a fully connected layer to enable intensive prediction in medical image processing. The problem of pixel location is solved using the shallower high-resolution layer, while the issue of pixel categorization is solved using the deeper layer. This structure is used in almost all current medical image semantic segmentation research. The internal structure of the human body is extremely complex. Hence, it is difficult for doctors to determine the disease's severity and location. Many approaches have been developed to overcome this challenge, and new research is constantly developing more novel and innovative methods. With the widespread adoption of image-aided medical diagnosis, segmentation is the desired process in medical image analysis. This is supported by the large number of papers explicitly published for the segmentation process, in which U-net survive prominent method [14,15]. UNET can improve the efficiency of segmenting disease-affected regions of the brain, lung, retina, liver, etc., as depicted in Figure 1.



**Figure 1.** Application of UNET in medical image segmentation.

Semantic segmentation is the classification of features in images based on pixels. Due to the lack of image detail, it is impossible to derive precise boundaries using image semantic feature information. The UNET model [16] designed by Olaf Ronneberger, Philipp Fischer, and Thomas Brox is shown in Figure 2, an ideal solution for medical image segmentation tasks, it efficiently uses the skip connection to merge feature maps of low-resolution and high-resolution images [17]. UNET is the CNN framework; it has a simple encoder and decoder network shaped like a U. This model can be well-trained with fewer samples. Despite the small training dataset, it provides precise segmentation results. The features were learned optimally using a UNET-based model.



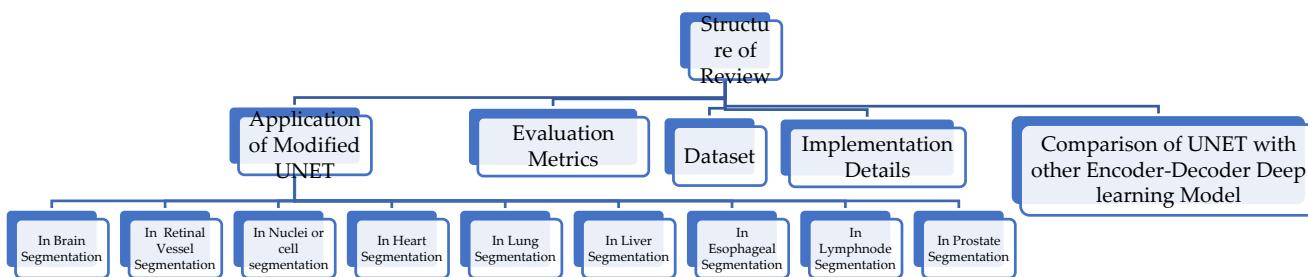
**Figure 2.** UNET model [16].

The survey articles [18,19] are related review works in which the application of UNET in various imaging modalities and UNET variants used in medical image segmentation are discussed. Our survey provides an

- In-depth review of UNET-modified architectures;
- Benchmark datasets and semantic architectures specifically designed for medical image segmentation;
- Presents the application of modified architectures of UNET in the segmentation of anatomical structures and a lesion in different organs to diagnose diseases;
- An updated survey of the improvement mechanisms, latest techniques, evaluation metrics, and challenges.

## 2. Study Method

The references are taken between the time frame of from 2015 to 2022. This survey is confined to the application of modified architectures of UNET in biomedical image segmentation. To determine the relevant quality of the paper, the references are taken from peer-reviewed journals. All architectures are thoughtfully collected from the original paper with a unique model focusing on enhancing accuracy and reducing complexity. Managing and comprehending the database format is a difficult task for researchers. Hence, this survey includes a separate section describing the medical image analysis database. It explains the benefits of adding the networks to the UNET in segmenting the lesion and tumor from different organs using images from imaging modalities. The structure of this review is given in Figure 3.



**Figure 3.** Structure of the review.

## 3. Application of Modified UNET

This section highlights the modified architecture of UNET for segmenting the region of interest from different imaging modalities to identify the severity of diseases.

### 3.1. InBrain Segmentation

#### 3.1.1. UNET with Generalized Pooling

This model modifies the pooling operation to enhance segmentation [20]. In the CNN and FCN models, the dimension is reduced to address the overfitting issue via max pooling or average pooling. Features are not precisely defined for variable data in down-sampling. A brain tumor's characteristics are very minute, so it is vital to minimize feature loss. A new generalized pooling (GP) method was developed to extract more prominent features from downsampling and improve segmentation performance. This approach adapts a pooling kernel's weights based on the input MRI images or feature maps. The initial average weight  $\alpha_0$  of each element is assigned as in Equation (1). The mean is given in Equation (2) as follows:

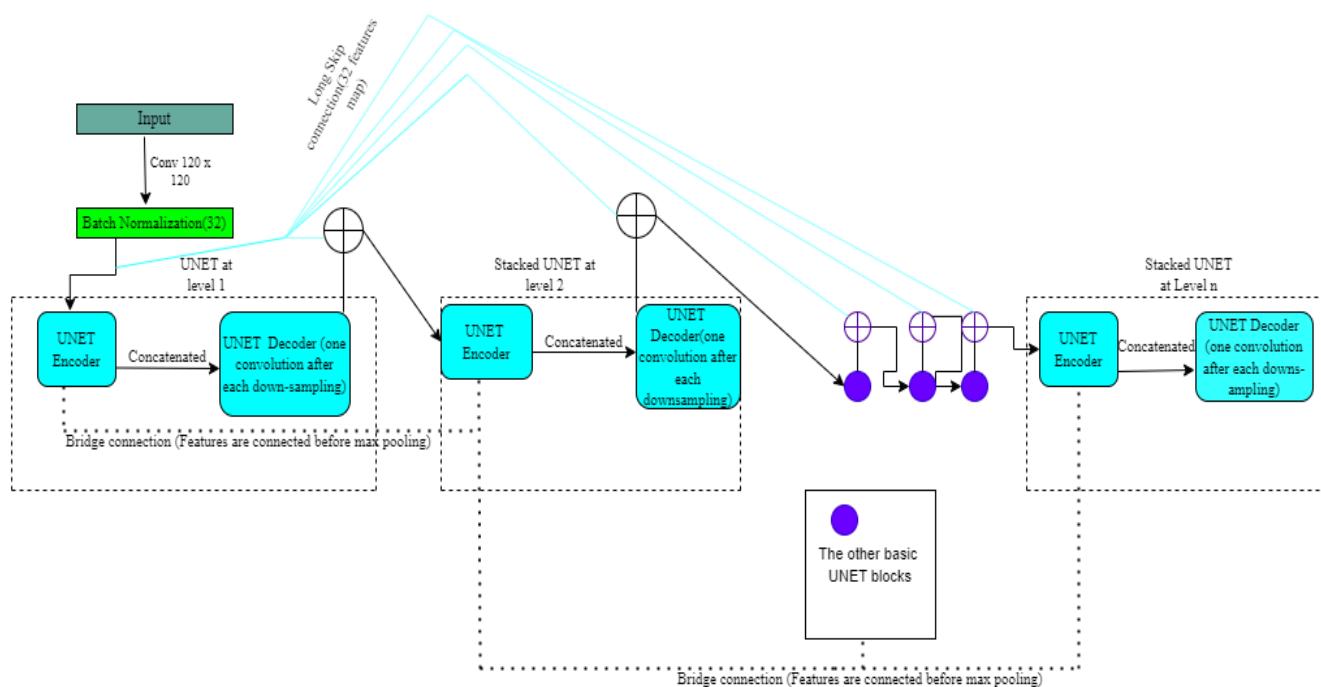
$$\alpha_0 = \frac{1}{p \times q} \quad (1)$$

where  $p$  is the length and  $q$  is the width of the pooling kernel.

$$\hat{z} = \frac{\sum_{r=1}^p \sum_{s=1}^q z_{rs}}{p \times q} \quad (2)$$

### 3.1.2. Stack Multi-Connection Simple Reducing Net (SMCSRNet)

Multi-connection stack, a novel framework known as simple reducing net (SMCSRNet) [21], is constructed using certain fundamental building elements (SRNet). Four down-sampling/up-sampling procedures were carried out throughout the encoding/decoding phases. UNET was further improved to better suit stacking to segment brain tumors. There is only one convolution process before each down-sampling. The processes of cropping and copying are maintained between decoding and encoding. This design aims to reduce parameters and simplify the network structure. It is important to note that the SMCSRNet model requires significantly less training time than the stacked UNET. In addition, the precision of this model has increased. The final block contains 32 feature maps stacked to the input image using the long skip connection depicted in Figure 4.



**Figure 4.** The architecture of SMCSRNet.

### 3.1.3. 3D Spatial Weighted UNET

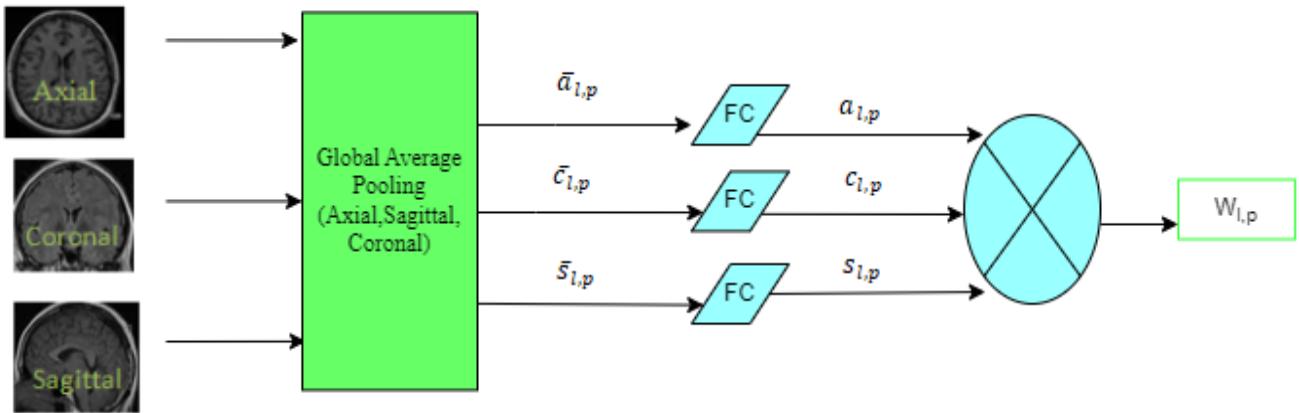
To properly utilize spatial contextual data at the intra-level plane and apply it to volumetric spatial weighting at the inter-level plane, the volumetric feature recalibration layer (VFR) is added to a 3D spatially weighted UNET [22]. It extracts geographic statistical information. The spatial information is compressed using global average pooling. The VFR is incorporated in this model before the de-convolutional layer and the max pooling layer in the encoder and decoder, respectively. Prior to resizing, it can be used to improve the features to prevent the loss of spatial information. Spatial statistical information is obtained by applying the global average pooling operation in each plane in Equation (3). The entire plane's spatial information is multiplied by the tensor product term to form the lower-weight tensor and change the weights of the volumetric input information. The workflow of VFR is shown in Figure 5.

$$\begin{aligned}\bar{a}_{lp} &= GAP_a(f_l, p) = \frac{1}{IJ} \sum_{ij} f_{l,p}(i, j, k), \\ \bar{c}_{lp} &= GAP_c(f_l, p) = \frac{1}{IK} \sum_{ik} f_{l,p}(i, j, k), \\ \bar{s}_{lp} &= GAP_s(f_l, p) = \frac{1}{JK} \sum_{jk} f_{l,p}(i, j, k).\end{aligned}\quad (3)$$

where  $f_l$  is the volumetric feature tensor input to the first VFR layer,  $i$  is the length,  $j$  is the width,  $k$  is the height, and  $p$  channels. The statistical information in three planes (axial,

coronal, and sagittal) are  $\bar{a}_{l,p}$ ,  $\bar{c}_{l,p}$  and  $\bar{s}_{l,p}$ . The weighted feature tensor is mathematically given in Equation (4) as follows:

$$w_{l,p} = a_{l,p} \otimes c_{l,p} \otimes s_{l,p} \quad (4)$$



**Figure 5.** WorkFlow of VFR.

This model is extended to the multimodality images with feature tensor values three times higher than for a single modality.

### 3.1.4. Anatomical Guided UNET

The segmentation and the anatomical attention sub-networks are the two sub-networks used in this model [23]. The segmentation network provides the local contextual information and learns the feature map from the image intensity. The anatomical images in the atlases train the anatomical networks. This anatomical gated network guides the segmentation network to segment the appropriate region of interest. The proposed anatomical guided architecture UNET is laid out in Figure 6. This work uses an anatomical gate to combine the features created by two sub-networks.

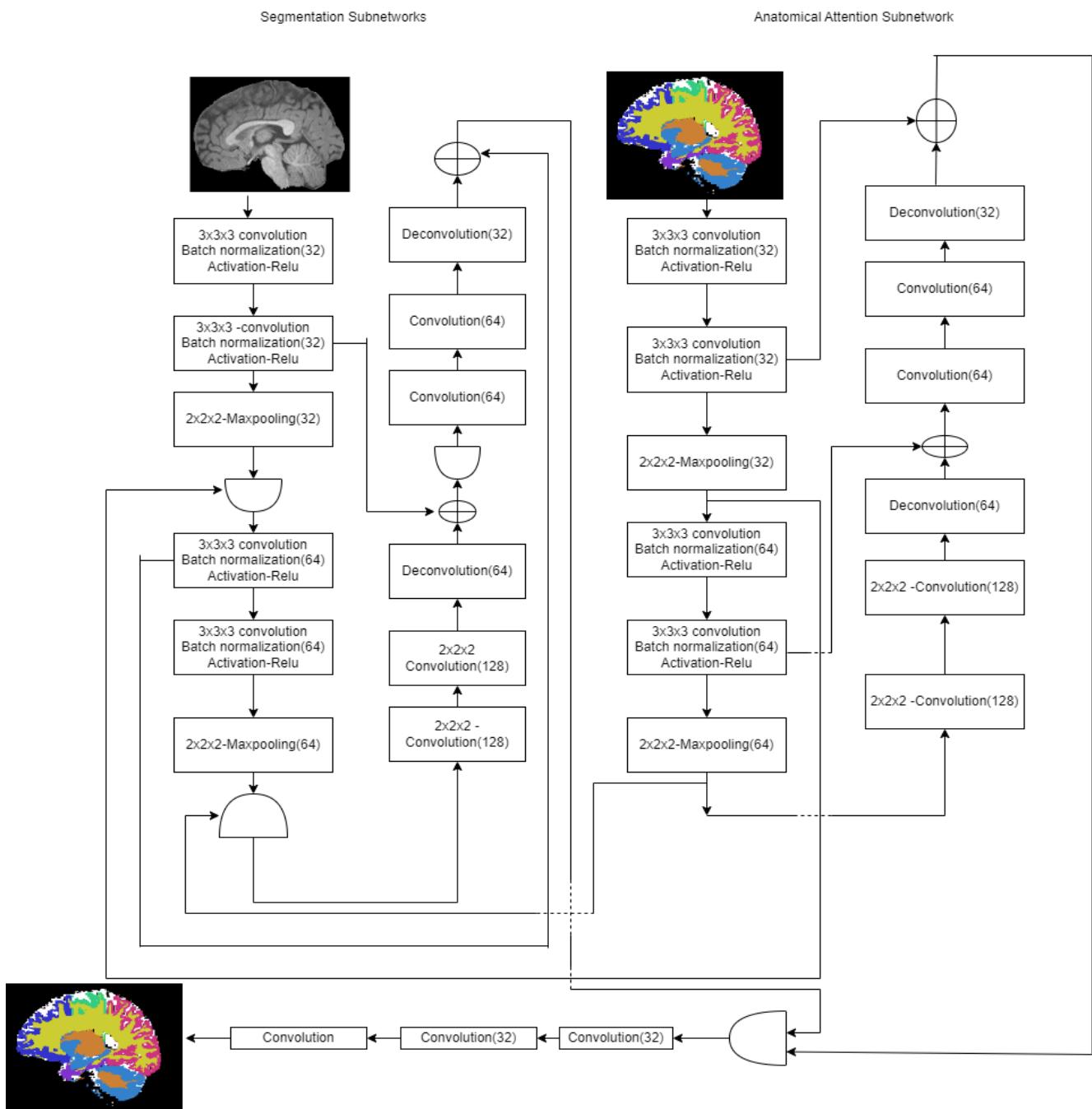
The feature maps [ $f_i^s$  (feature map from segmentation network in the  $s$ th network),  $f_a^s$  (feature map from anatomical attention subnetwork)] are concatenated channel-wise. It is fed into two convolutional layers (size:  $1 \times 1 \times 1$ ), and a non-linear sigmoid unit follows each convolutional layer to learn the weight tensor. (e.g.,  $o_i^s$ ) for each input feature map. The learning mechanism of weighted tensor is given in Equation (5) as follows:

$$\begin{aligned} o_i^s &= \sigma(W_i^s[f_i^s, f_a^s] + b), \\ o_a^s &= \sigma(W_a^s[f_i^s, f_a^s] + b). \end{aligned} \quad (5)$$

The anatomical gate, feature map output ( $f_o^s$ ) is given by the following:

$$f_o^s = o_i^s \cdot f_i^s + o_a^s \cdot f_a^s \quad (6)$$

The anatomical attention gate contains brain structure information provided by multiple atlases at different scales. This model automatically learns the optimal weights generated by the two subnetworks and efficiently fuses the two subnetworks for accurate ROI segmentation.



**Figure 6.** Anatomical attention guided model.

### 3.1.5. MH-UNET

In multi-scale UNET [24], several dense blocks, residual inception blocks, and hierarchical blocks are included in the decoder and encoder, which reduce the trainable parameters. Residual inception blocks (in Figure 7) extract valuable features. It learns much global and local information from a large receptive field. Residual inception block output is given in Equation (7).

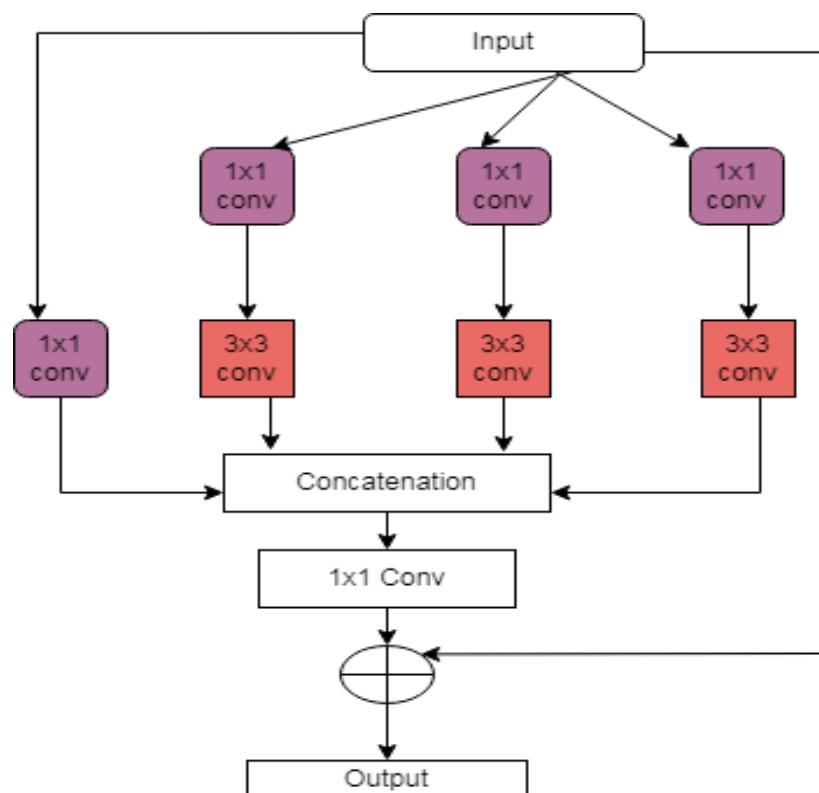
$$y_{l+1} = ((f_{one}(f_d(y_l) \odot y_l)) \oplus f_{one}(y_l)) \quad (7)$$

where  $y_l$  is the output of current layer,  $f_d(\cdot)$  is for Dilated Conv-IN-LeakyReLU,  $f_{one}$  is for  $1 \times 1 \times 1$  Conv-IN-LeakyReLU. The hierarchical block extracts multi-scale information

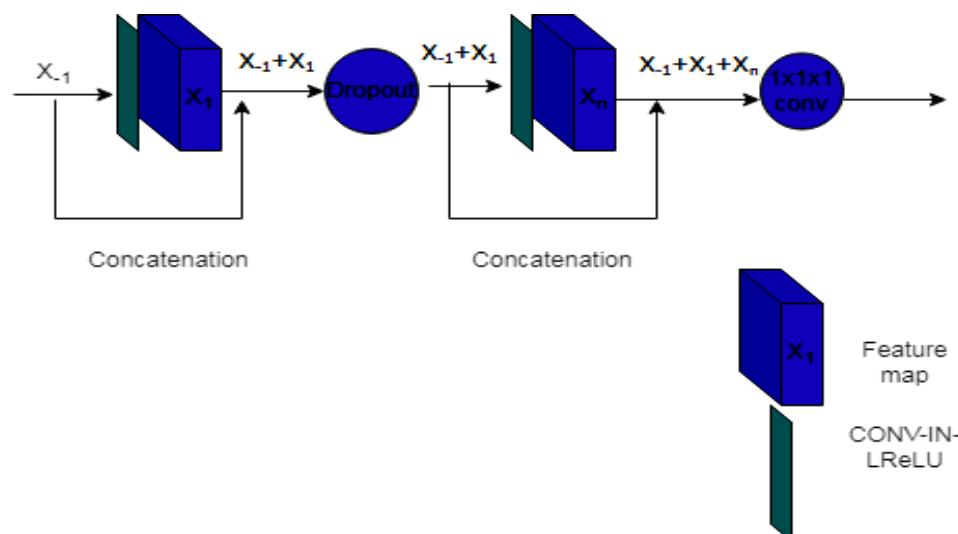
features. In the hierarchical block, dilated convolutional layers increase the receptive field without increasing the dimensions. On the other hand, a dense network (in Figure 8) decreases the trainable parameter and redundant feature for 3D convolution. The working condition of a dense block is described in Equation (8).

$$x_{l+1} = g(x_l) \circ x_l \quad (8)$$

where  $x$  is the output of the current layer and  $g$  represents the flow of Conv-IN-LeakyReLU and  $\circ$  is the concatenation function. Deep supervision is also proposed for superior segmentation accuracy and faster convergence.



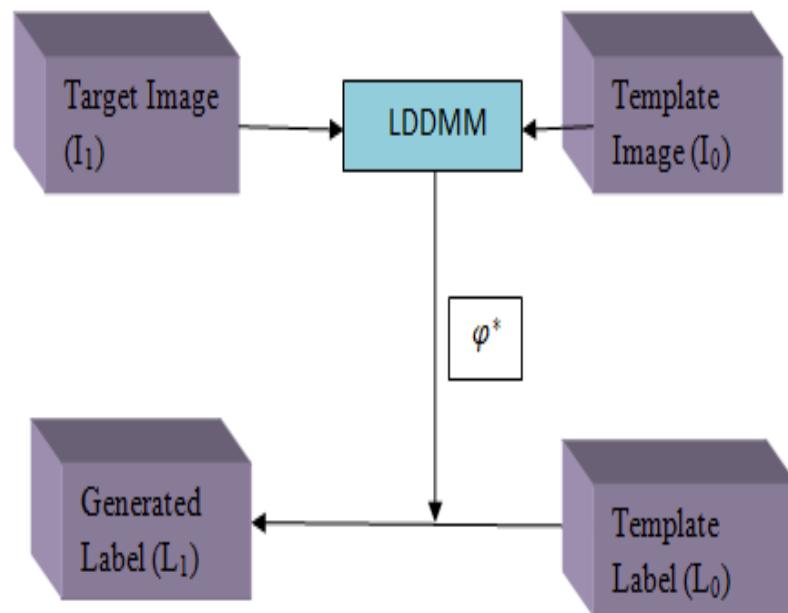
**Figure 7.** Residual inception block.



**Figure 8.** Dense block.

### 3.1.6. MI-UNET

In MI-UNET [25], brain parcellation information is obtained for the input MRI, and this information is additionally given as the input to the UNET (shown in Figure 9). LDMM [26] image registration algorithm is used for extracting the segmentation details from the atlas-based registration, and the MRI image is segmented into GM, WM, and LV.



**Figure 9.** MI-UNET architecture.

The brain parcellation is obtained as follows:

$$L_1 = L_0 \odot \Phi_a^* \quad (9)$$

In Equation (9),  $L_1$  is the brain parcellation,  $L_0$  is the template label and  $\Phi_a^*$  is the transformation. The GM, WM, and LV parcellation are obtained using atlas-based segmentation, which is independent of the subsequent deep learning-based stroke lesion segmentation.

### 3.1.7. Multi-Res Attention UNET

In multi-res attention gate UNET [27], Multi resnet [28] block reduces the filter dimension by splitting the  $5 \times 5$  and  $7 \times 7$  into the series of  $3 \times 3$ . In addition, two-layer filters ( $L1, L2$ ) are implemented to reduce the requirement of high memory.  $L1$  and  $L2$  filter parameters are given in Equations (10) and (11), respectively.

$$\text{No of the filters parameter in } L1 = k^2 \times n \times l \quad (10)$$

$$\text{No of the filter parameter in } L2 = (k')^2 \times l^2 \quad (11)$$

A residual path is added to overcome the semantic gap problem between the encoder and decoder.

$$\text{Resx} = \theta_{X_{3 \times 3}} \cdot \mu_i + w_{X_{1 \times 1}}(\mu_i) + b_x \quad (12)$$

$$\text{Resy} = \theta_{Y_{3 \times 3}} \cdot \mu_i + w_{Y_{1 \times 1}}(\mu_i) + b_y \quad (13)$$

In Equations (12) and (13), variable x represents the first layer, and variable y represents the second layer. Whereas  $\theta$  is the filter term,  $\mu_i$  is the feature map, w is the convolution, and b is for bias. The attention-gating block has the GS(gating signal). This signal guides the attention block to choose the exact features. Extracted spatial information is passed

through a  $1 \times 1$  ( $w_{GS}$ ) convolution operation. Finally, a ReLU activation function is applied to the output. As shown in Equation (14), the resulting signal is the attention-gating signal.

$$GS = \text{ReLU}(w_{GS}(s) + b_{GS}) \quad (14)$$

### 3.2. In Retinal Vessel Segmentation

#### 3.2.1. GLUE [29]

A weighted U-Net (WUN) and a weighted residual U-Net(WRUN) form this model. The WUN first creates a coarse segmentation map using patches that have been globally improved. The WRUN then enhances the locally upgraded patches, whose parameters are automatically updated rather than adjusted. Discriminative features are obtained by adding residual connections to the second half of the model (WRUN). Additionally, it uses the cascaded U-Net structure, which stands to gain improvements in retinal imaging both locally and globally. On retinal images, the contrast-limited adaptive histogram equalization (CLAHE) operation [30] is used to increase contrast. A circular template mask for the region of interest is created to obtain the location of the fundus. This mask can be used as the weighted attention mask to segment only the fundus and leave the irrelevant area. The weighted attention mask is multiplied by the feature map of the last WRUN layer, and the skip connection improves the depth and accuracy of UNET. It is implemented as in Equation (15).

$$y = F(x, \{w_i\}) + H(x) \quad (15)$$

where  $x$  represents the input,  $H$  represents the identity mapping function and  $w_i$  represents the weight.

#### 3.2.2. S-UNET

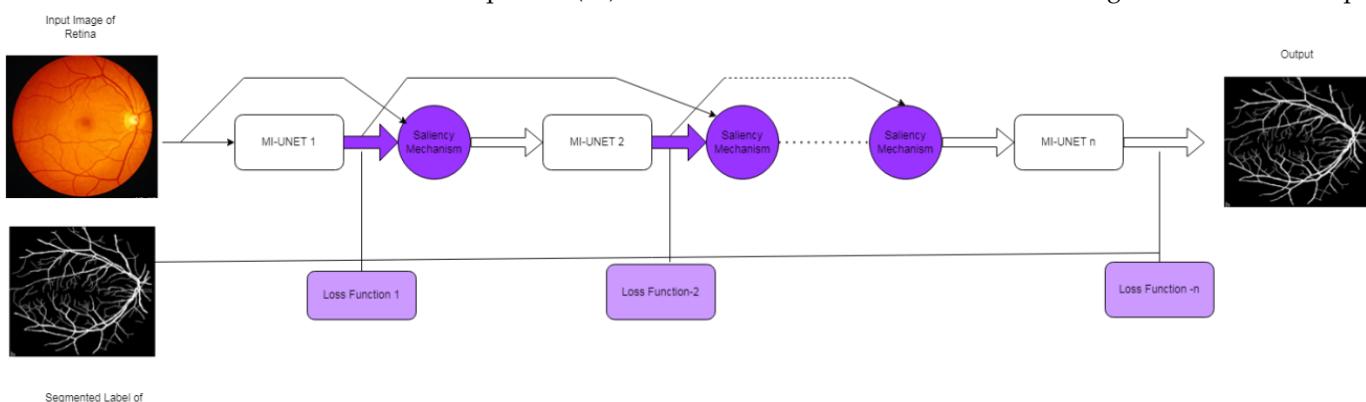
The minimum UNET is the foundation of the salient UNET [31] architecture. The network parameter can be decreased from 31.03 M to 0.07 M with minimal UNET. The bridge-style architecture, with two Mi-UNETs cascading, provides a prominent mechanism. Some features were taken from the first MI-UNET and provided as foreground attention directions for the next MI-UNET (shown in Figure 10). Features from all the output units are concatenated with the input block. It is given in Equation (16).

$$O_1 = W_1 \times 1 \quad (16)$$

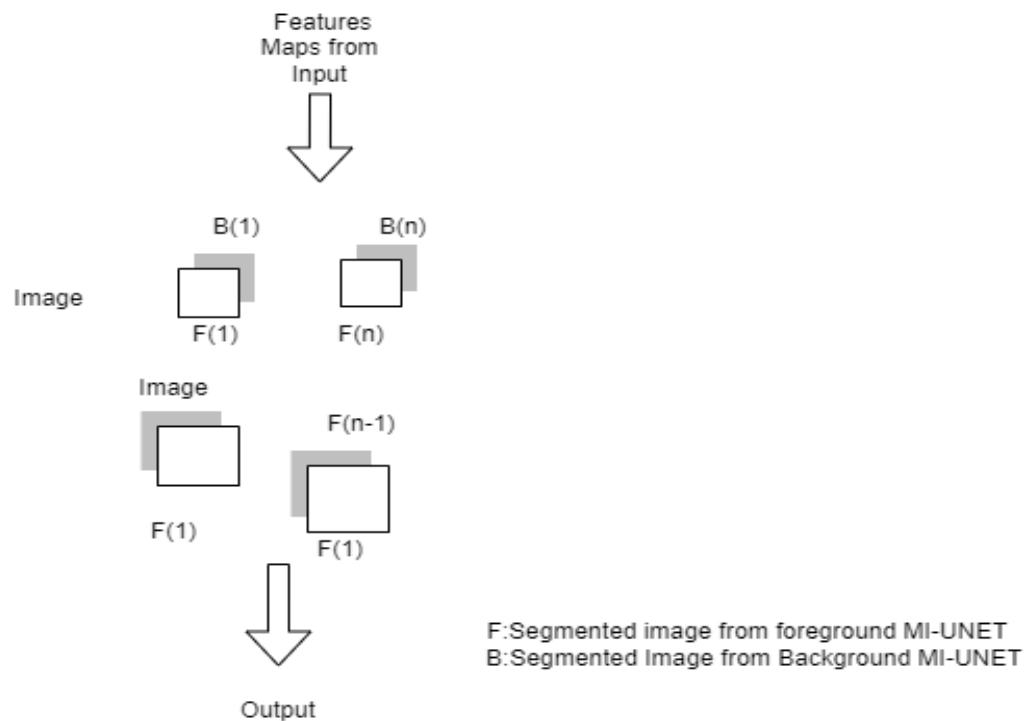
The saliency mechanism is shown in Figure 11 and defined in Equation (17).

$$sO_1 = (W_1 X)_f \oplus X_1 \quad (17)$$

From Equation (17), it is clear that the second minimal UNET gets the enhanced input.



**Figure 10.** S-UNET.



**Figure 11.** Saliency mechanism.

### 3.3. In Nuclei or Cell Segmentation

#### 3.3.1. As-Unet

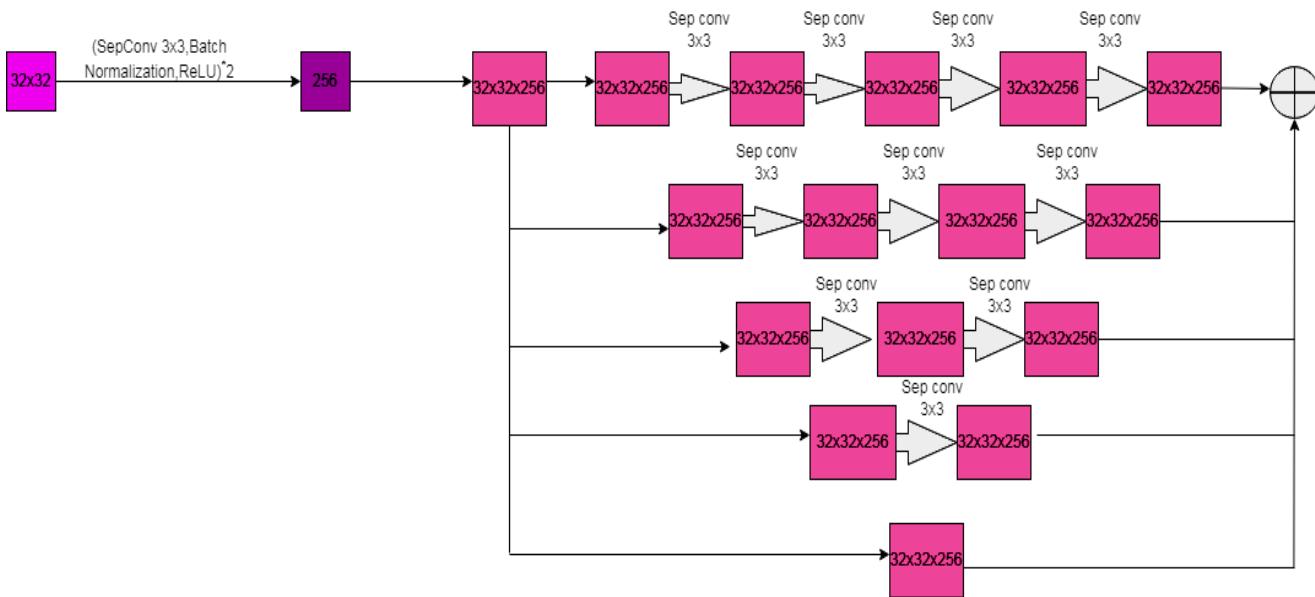
The atrous convolution is added between the encoder–decoder to increase the network’s receptive field without affecting the image resolution. Atrous convolution can change the convolution step for multi-scale information. The  $3 \times 3$  Separable convolutional is added with the ReLu activation function. There are 4 dilation rates, and 5 parallel and cascade atrous separable convolutions are added, and it is shown in Figure 12. The size of the AS-UNET [32] model, the number of trainable parameters, and the evolution time decreases using separable convolution. In AS-UNET, log-Dice loss and the focal loss are added to calculate the loss function as in Equation (18).

$$\text{Loss} = \lambda * \text{logDL} + (1 - \lambda) * \text{FL} \quad (18)$$

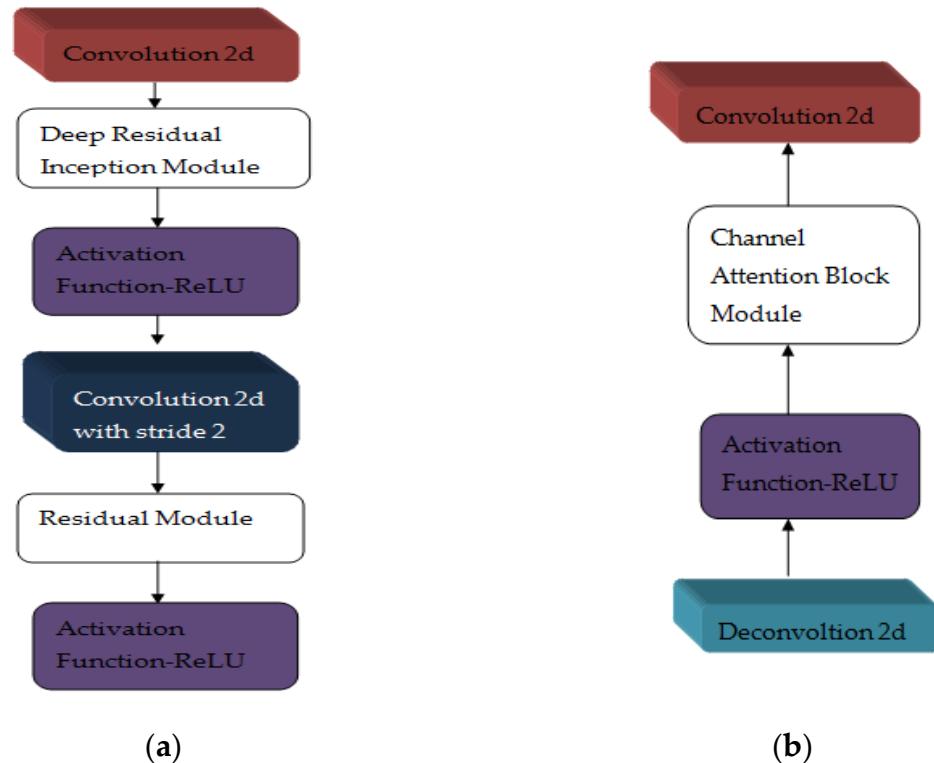
In Equation (18),  $\text{LogDL} = -\log(2*(y_t \cap y_p)) / (|y_t| + |y_p|)$  is the logDice loss and  $\text{FL} = y_t * \log(y_p) * (1 - y_p)^\gamma$  is the focal loss,  $y_t$  is the GT value,  $y_p$  is the predicted value, and  $\lambda$  is the training parameter.

#### 3.3.2. RIC-UNET

The multi-scale residual inception block and channel gate are applied in RIC UNET [33]. The residual inception block extracts the multi-scale feature information. Cell contour obtained from this network is used to segment the dense cell and reduce the cell level error. The channel attention block selects the high-resolution features with the low-resolution information taken from the up-sampling process. The structure of the RI block and DC block is laid out in Figure 13.



**Figure 12.** Part of AS-UNET.

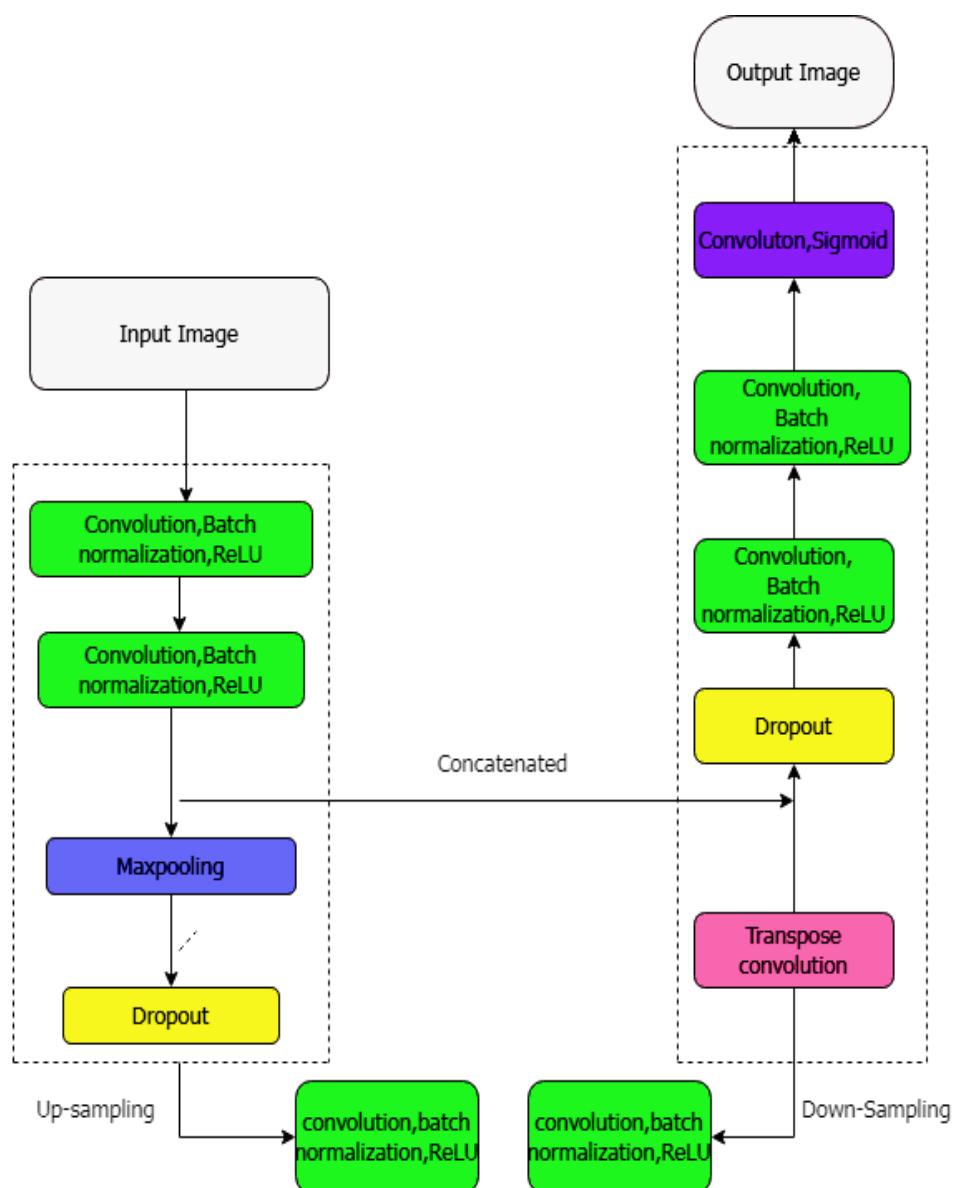


**Figure 13.** (a) RI Block (b) DC Block.

### 3.4. UNET in Heart CT Segmentation

#### 3.4.1. Modified 2D UNET

A modified 2D UNET model [34] is the next-level model of the fundamental 2D UNET model. It adds a dropout and batch normalization before each convolution block (depicted in Figure 14) to segment the aorta and coronary artery. The internal covariate shift affects the training process. The batch normalization stabilizes the training by normalizing the inputs for each mini-batch, which was achieved by ciphering the standard deviation and mean of each input variable for the layer of a single mini-batch. By randomly setting the weights to zero, the over-fitting was reduced using the dropout layer.



**Figure 14.** The architecture of adding batch-normalization and dropout layer in up-sampling and down-sampling processes.

### 3.4.2. UCNNet with Attention Mechanism

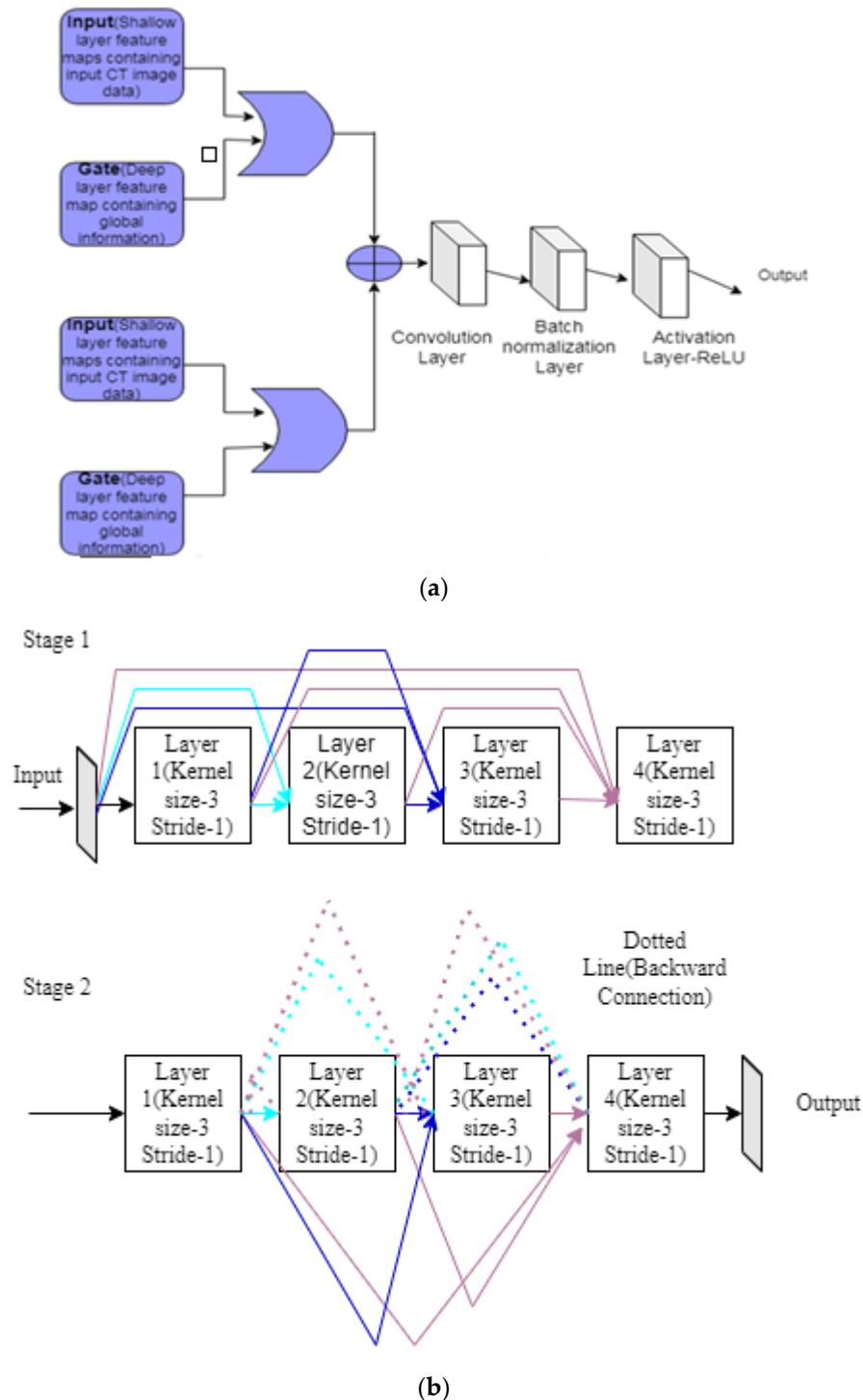
A negative mining technique is used in this model [35] to suppress the uninterested area. First, the number of negative sample examples  $N_s$  for each training sample was estimated using Equation (19).

$$N_s = N_n \max(2N_p, \frac{N_n}{8}) \quad (19)$$

In Equation (19),  $N_s$  is the number of negative samples, and  $N_p$  is the number of positive samples.

The attention mechanism and U-clique net focus only on the vital region. In the attention mechanism, input is in the shallow layer, and the gate uses the deep layer. Both are added to generate the attention map (Figure 15a) and are given to convolutional block, batch normalization, and RELU. U-clique UNET is laid out in Figure 15b. In stage 1, each layer is connected with the previous layer to update the next layer. In the next stage, layer 2 is concatenated to layer 1 in a forward direction, and the third and fourth layers in the

feedback directly to stage 1. This process will improve communication between the layers. Finally, heart regions are divided into segments, and the Jaccard score is calculated.



**Figure 15.** (a) Attention mechanism (b) U-clique NET.

### 3.5. UNET in Lung Segmentation

#### 3.5.1. Cascaded UNET [36]

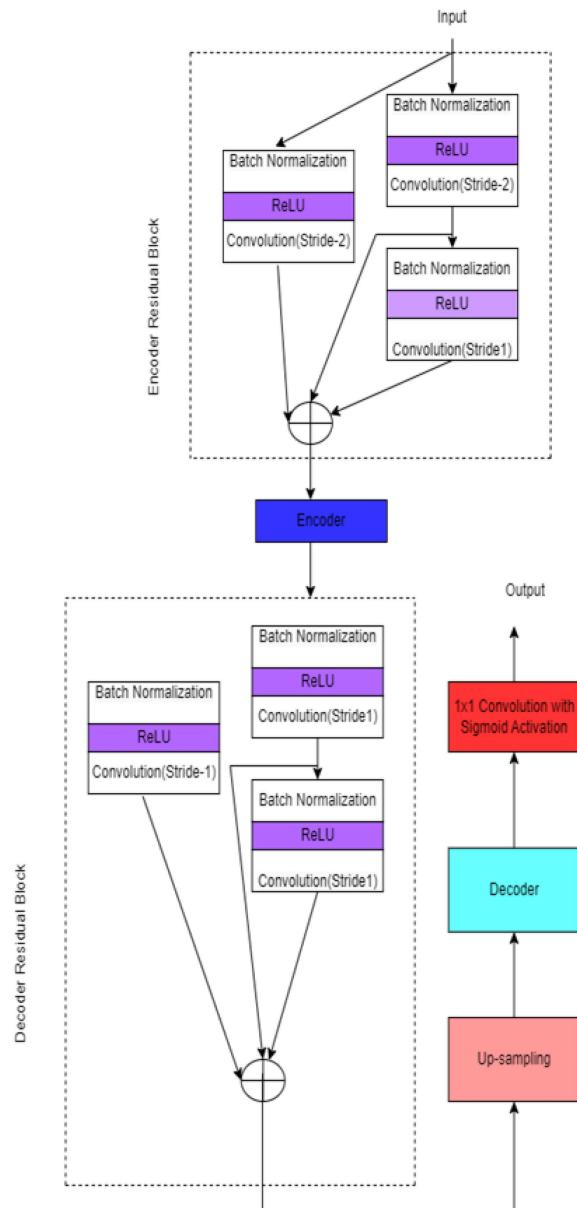
The network includes the EM (expectation maximization) framework [37] to account for the prior function of the disease-affected area. UNET is initially fine-tuned to discover the consolidated region from the labels at the patient level by applying the EM algorithm

after being trained with labeled, segmented image of the region of interest. Then, the latent variable  $y$  is solved pixel-wise with the EM algorithm given in Equation (20).

$$y_{ij} = \begin{cases} 1, & |f_j(x_i; \theta' + \varphi(z_i, x_{ij}))| > 1 \\ 0, & \text{Otherwise} \end{cases} \quad (20)$$

### 3.5.2. Res-D-UNET

Res-D-UNET [38] extracts all the high-level features from the intra-slice plane. An overview of a residual dense block is shown in Figure 16. The exclusive feature from the top layer to the bottom layer gets utilized; hence, vanishing gradient problem is reduced during the training period of the network. Binary cross entropy, similarity index, and dice loss are the loss functions calculated in this model.



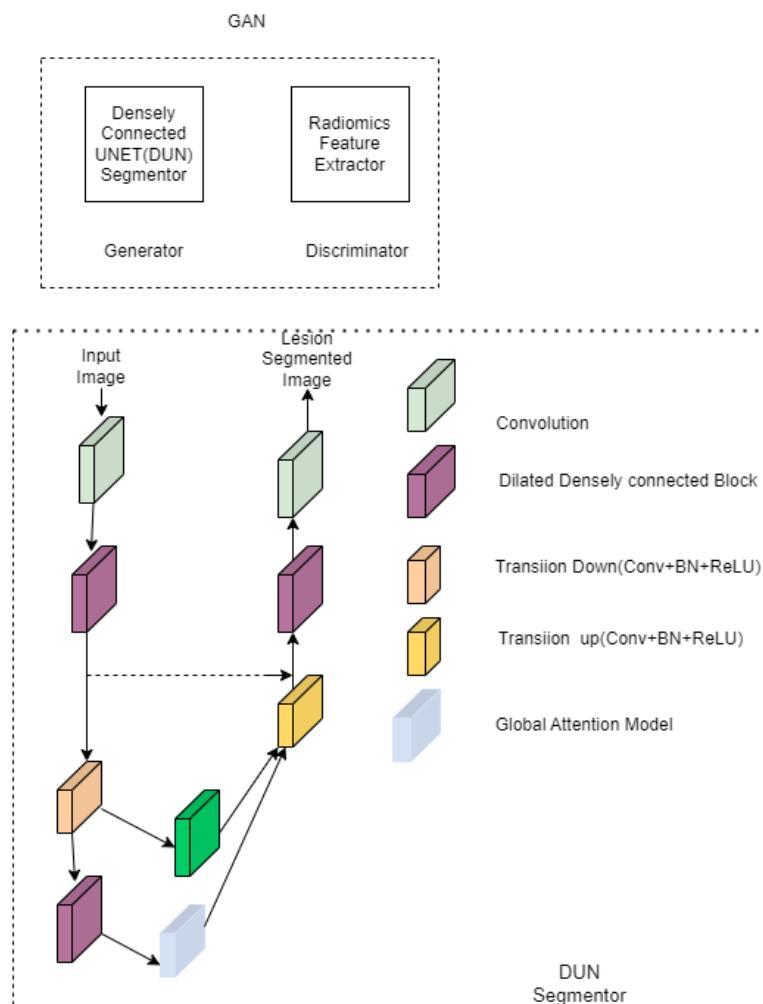
**Figure 16.** Res-D-UNET.

A ReLU activation layer, a batch normalization layer, and two convolution layers with strides of 2 and 1 are included in each convolution block. In addition, a convolutional

layer connects encoder input and output with a stride of 2, and a BN layer is used in identity mapping.

### 3.6. UNET in Liver Segmentation

Multi-phase dynamic contrast enhancement MRI radiomics features [39] insist on extracting the ICR characteristics from non-contrast images. Therefore, it is carried out without the use of contrast chemicals. In this work [40], the radiomics features guide UNET and generational adversarial network. Radiomics features are used at the discriminator, and the DUN (shown in Figure 17) is used as the segmenter at the generator network. UNET disseminates the directed knowledge. The gradient disappearance is reduced by combining a dilated and densely convolutional network. A global attention model extracts the desired characteristics from the pixels in low-contrast images. The discriminator of the GAN receives the MCRF (multi-phase radiomics feature) as input, which easily separates lesions from non-contrast images. Radiomics and semantic feature extraction models are connected with radiomic-guided layer connections at the discriminator. Semantic features are extracted using VGG 16 [41]. PyRadiomics [42] is an open-source tool to extract the features from the MRI.

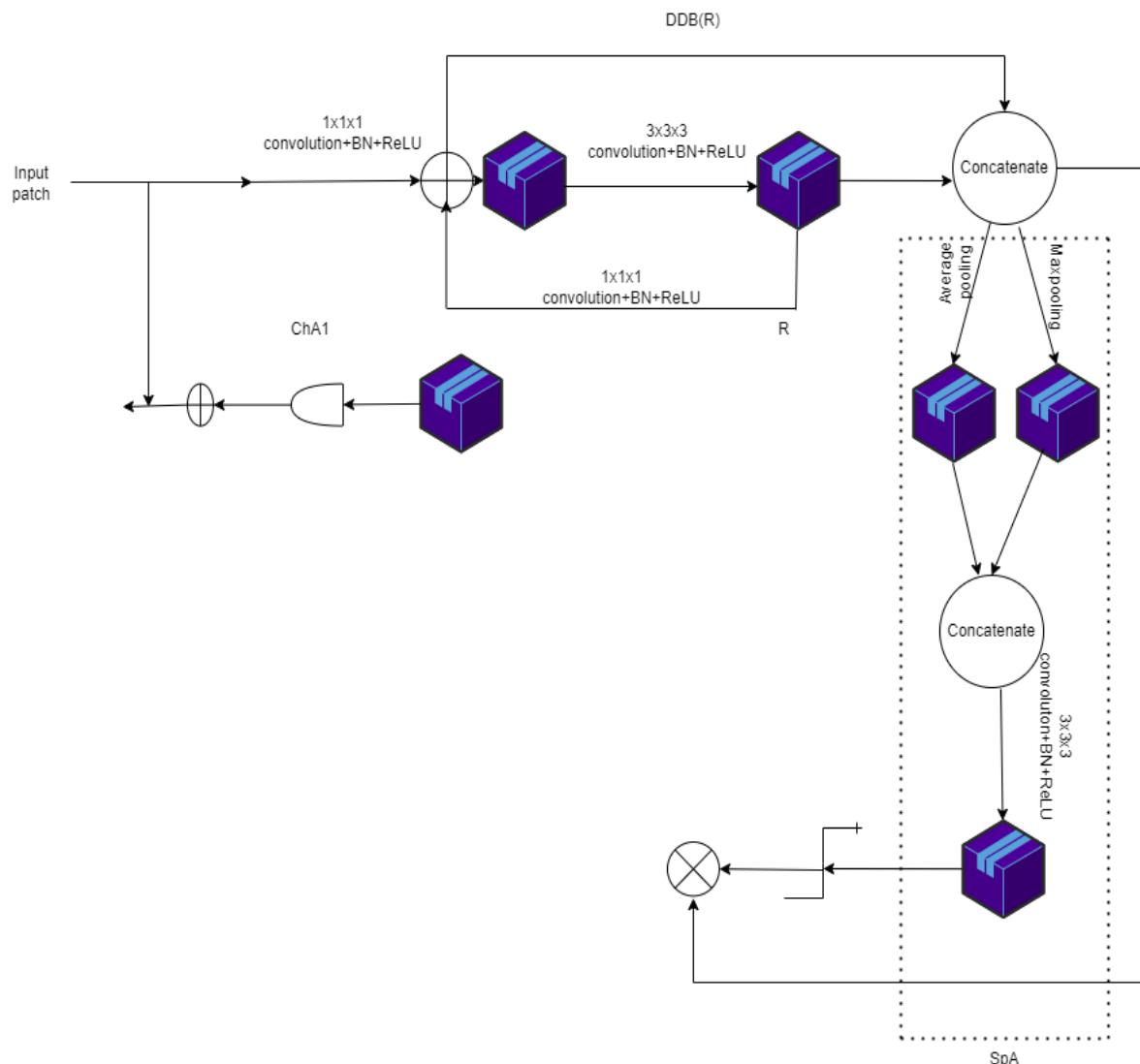


**Figure 17.** Densely connected UNET.

### 3.7. UNET in Esophageal Segmentation

In this model of a dilated dense block, channel attention (CHA1) and spatial attention (SPA) gates are used. The spatial gate retrieved tumor features in the main block were retrieved by the spatial gate. In the space between the paths of extracting and contracting,

the channel gate filtered out the unimportant features. Dubbed dilated dense attention UNET model [43] (DDAUNET), it segments the esophageal GTV (gross tumor volume). Its architecture is shown in Figure 18.

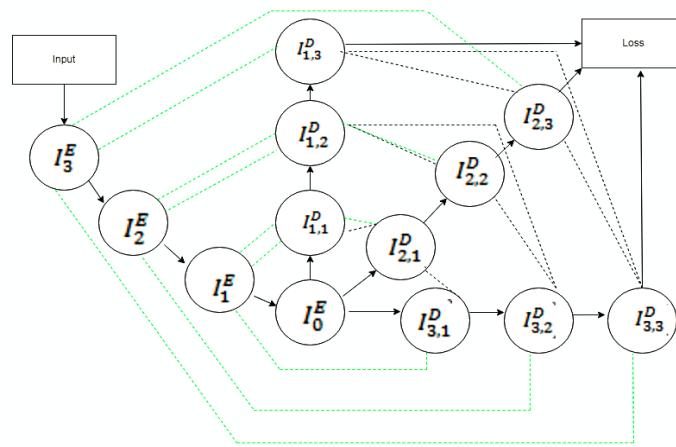


**Figure 18.** Architecture of DDAUNET.

Figure 18 denotes DDSCAB (dilated dense spatial and channel attention block) and DDB (dilated dense block). R represents the number of sub-DDBs. For example, chA1 is a skip connection channel attention gate, ChA2 is a DDSCAB block channel attention gate, and SpA is a DDSCAB block spatial attention gate. Although ChA1 is not included in the final network (DDAUnet), it is used in some of the experiments.

### 3.8. UNET in Lymphnodes Segmentation

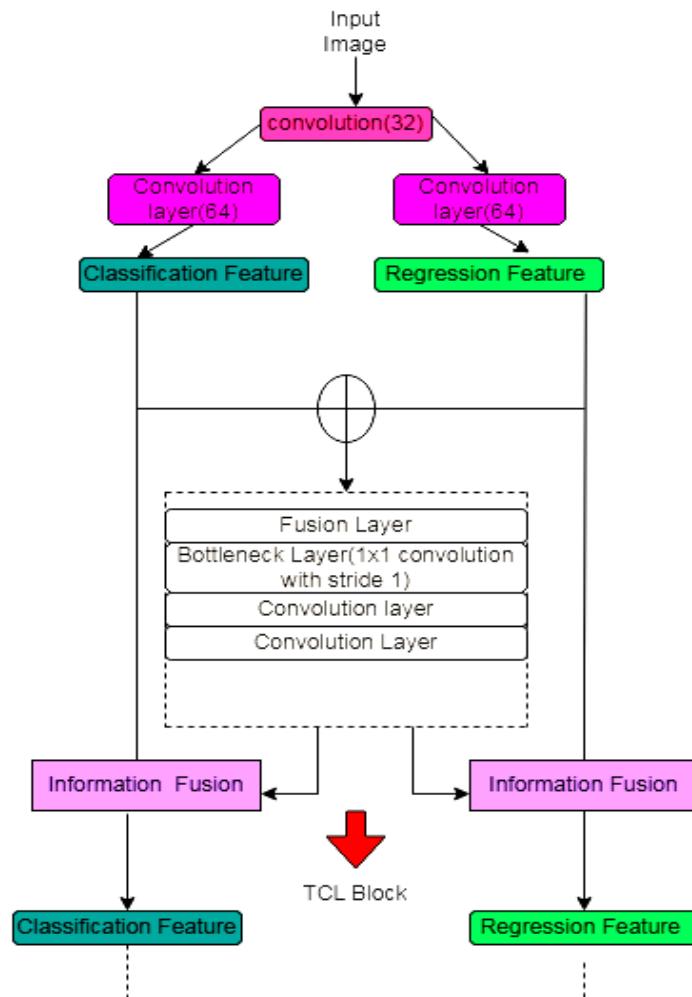
The body has lymph nodes and lymphoid tissues in all parts, making it challenging to distinguish lymphoma on a full-body CT scan. Hyperdense encoding using UNET architecture and recurrent dense Siamese decoding is employed in this model [44] at the encoder and decoder, respectively. The segmentation accuracy is increased using bootstrapping in re-sampling and a stable-gradient adaptive similarity dice loss function. The recurrent dense Siamese UNET in Figure 19 enables the spatial and temporal correlation. The Siamese decoder has two similar subnetworks for generating the feature vector for the input and eradicating the duplicate features.



**Figure 19.** RDS UNET.

### 3.9. UNET in Prostate Segmentation

A challenging task in prostate segmentation is (1) fast localization of the prostate boundary and (2) accurate segmentation. Hierarchically fused UNET is the multitask FCN. Adding an attention-based task consistency learning (TCL) module allows the encoder and decoder to share task-related knowledge. This research [45] implements a channel-based and a position-based attention network to learn the best information (shown in Figure 20).



**Figure 20.** Task consistency block (TCL).

#### 4. Evaluation Metrics

- DSC

The dice similarity coefficient (DSC) was first proposed by Dice [46]. It uses a reproducibility validation metric and an index of spatial overlap. Fleiss also referred to it as the percentage of explicit agreement [47]. DSCs values range from 0 to 1, which denotes the entire spatial similarity between two data sets from binary segmentation, indicating total spatial overlap. It predicts the similarity index between the ground truth and the predicted image by comparing the pixel-wise agreement between the two images.

$$\text{DSC} = \frac{2*|X \cap Y|}{|X|+|Y|} \quad (21)$$

In Equation (21), DSC is the dice similarity coefficient, X is the ground truth image pixels, and Y is the predicted image pixels. It should be higher.

- PPV—positive predictive value or precision

It measures the precision of prediction [48–52] by counting the number of actual samples. It is formulated in Equation (22).

$$\text{PPV} = \frac{TP}{TP + FP} \quad (22)$$

- Accuracy

Accuracy calculates the correctly classified pixels in the images. The formula for the accuracy is given in Equation (23).

$$\text{Accuracy} = \frac{TP + TN}{TP + FN + TN + FP} \quad (23)$$

- Sensitivity or recall

It measures [53,54] the number of false and true images. It is otherwise known as the positive rate. The calculation of recall is given in Equation (24).

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad (24)$$

- F1 score

This metric [55] gives the balance value in-between precision and recall. The result of 1 represents the best prediction. F1 score is formulated in Equation (25).

$$\text{F1} = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \quad (25)$$

- AUC (area under curve) [56]

It is the plot of the receiver under the operation curve according to the true positive rate(TPR) at the vertical axis and false positive rate(FPR) at the horizontal axis. TPR and FPR are given in Equations (26) and (27), respectively.

$$\text{TPR} = \frac{TN}{TN + FP} \quad (26)$$

$$\text{FPR} = \frac{FP}{FP + TN} \quad (27)$$

- The 95th percentile Hausdroff distance

Hausdroff distance [57] is the prediction of the distance between prediction and ground truth images. Small value of HD represents the high segmentation accuracy.

$$\text{HD}(S, L) = \max \left\{ k^{\text{th}}_{s \in S} \min_{g \in G ||S-L||}, \left\{ k^{\text{th}}_{g \in G} \min_{s \in S ||L-S||} \right\} \right\} \quad (28)$$

In Equation (28),  $S$  is the segmented image, and  $G$  is the ground truth image.

- Absolute volume difference

It predicts the difference between segmentation and label in terms of volume. A smaller range of AVD [58] gives better segmentation.

$$\text{AVD}(S, L) = \frac{V_s - V_L}{V_L} \times 100\% \quad (29)$$

In formula (29),  $V_s$  is the volume of the segmented image, and  $V_L$  is the volume of the labeled image.

- Jaccard score or IOU [59]

$$\text{Jaccard}(A, B) = \frac{|A \cap B|}{|A| + |B| - |A \cap B|} \quad (30)$$

In Equation (30),  $A$  is the ground truth, and  $B$  is the segmented image.

- Matthews correlation coefficients (MCC) [60]

It is a statistical tool to identify the difference between predicted and actual images, which Brain Matthew formulated.

$$\text{MCC} = \frac{TN \times TP - FN \times FP}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad (31)$$

## 5. Datasets

### 5.1. MRBrainS18 [61,62]

The image data for this challenge were collected at the UMC Utrecht using a 3T scanner (The Netherlands). T1-weighted, T1-weighted inversion recovery, and T2-FLAIR scans of 30 subjects have been fully annotated. Alzheimer's patients, patients with dementia, Diabetes, and, as well as matched controls (with increased cardiovascular risk) with varying degrees of atrophy and white matter lesions (age > 50), were included in the study. The voxel sizes for all scans are 0.958 mm, 0.958 mm, and 3.0 mm. The N4ITK algorithm is used to correct the bias fields in the scans.

### 5.2. IBRS

The Internet Brain Segmentation Repository (IBSR) [63] encourages the advancement of segmentation methods and the evaluation of MRI brain images. There are eighteen subjects ranging in age from 7 to 71. It is also worth noting that these data were subjected to the CMA'autoseg'bias field correction routines.

### 5.3. BRATS

A trained human expert manually annotated multi-contrast MRI scans of ten patients with low-grade glioma and twenty patients with high-grade glioma with two tumor labels [64,65]. Furthermore, the training data consist of simulated images of 25 high-grade and 25 low-grade glioma patients with the same 2 "ground truth" labels. The test images included 11 high-quality and 4 low-quality real-world cases and 10 high-quality and 5 low-quality simulated images.

#### 5.4. ADNI

Alzheimer's MRI images were taken from the ADNI (Alzheimer's disease Neuroimaging Initiative) database [66,67]. The primary purpose of ADNI is to track the progress of the disease and study the variation in brain function and structure during the four stages of the disease. ADNI has a clinical record of patients between 55 and 90, including males and females. Patients have undergone all the tests at subsequent intervals. This project is for collecting the anatomic, diffusion, perfusion, and resting-state MRI images.

#### 5.5. ATLAS

A 955 T1-weighted MRI scans are available in the Anatomical Tracing of Lesions after Stroke (ATLAS) dataset [68]. These scans are divided into training ( $n = 655$  T1w MRIs with manually segmented lesion masks) and testing ( $n = 300$  T1w MRIs only; lesion masks are not released). T1-weighted average structural template images from MNI152 standard space are used. The database contains lesion and scanner metadata in two.csv files. The LONI Probabilistic Brain Atlas (LPBA40) is a collection of anatomical maps of the brain that can be found in Atlas. These maps were created using data from 40 human volunteers' whole-head MRIs. Each MRI was manually delineated to identify 56 brain structures, most of which are located in the cortex.

#### 5.6. CHASE\_DB1

A child heart and health study in England (CHASE\_DB1) [69] contains 28 color retina images with a resolution of  $999 \times 960$  pixels taken from the left and right eyes of 14 school children for segmenting retinal vessels.

#### 5.7. DRIVE

The fundus images in the Digital Retinal Images for Vessel Extraction (DRIVE) [70] dataset include 7 abnormal pathology instances. It contains 40 images in JPEG format. The dataset is equally split for training and testing. The images are taken from a Netherlands screening program for diabetic retinopathy.

#### 5.8. STARE [71]

The dataset contains 20 eye fundus images with a resolution of  $700 \times 605$ . In addition, two sets of ground-truth vessel annotations are available. Six images in this dataset are normal, and 11 indicate ophthalmological disease.

#### 5.9. RITE [72,73]

Based on the publicly accessible DRIVE database, the RITE (Retinal Images Vessel Tree Extraction) database was created to enable comparative investigations on the segmentation or categorization of arteries and veins using retinal fundus images. Like DRIVE, RITE has 40 images evenly divided into training and test subsets. A fundus image, a vascular reference standard, and an arteries/veins (A/V) reference standard are included for each set. Four different types of vessels are identified for the A/V reference standard based on the vessel reference standard using four different colors. The image of the fundus is in tif format. The A/V and vessel reference standards are also in the png file format.

#### 5.10. CCAP IEEE Data Port [74]

It is obtained from the IEEE Data Port and consists of the following five distinct sets of lung CT images: Viral Pneumonia, COVID-19, Bacterial, Pneumonia, Normal lung, and Mycoplasma Pneumonia (MP).

#### 5.11. SARS-CoV-2 CT-Scan Dataset [75]

It included 1252 CT scans from patients infected with the disease and 1230 CT scans from patients not infected, for a total of 2482 CT scans.

### 5.12. CHAOS [76]

CHAOS provides CT and MRI data from healthy subjects for single and multiple abdominal organ segmentation.

### 5.13. ISLES [77]

In ISLES 201,863 patients' information was included for training, while 40 patients' information was added for testing. Furthermore, the developed methods are tested on a 40-stroke research dataset.

### 5.14. TCGA [78]

The TCGA project produced a massive amount of genomic, epigenomic, transcriptomic, and proteomic data. Transcriptomics technologies are methods for studying an organism's transcriptome, the sum of its RNA transcripts. A proteome is a collection of proteins made by an organism. This information has improved our ability to diagnose, treat, and prevent cancer.

### 5.15. MOD [79]

It is a data set of pathological images with 30 images from the following 7 organs: colon, stomach, prostate, liver, breast, kidney, and bladder. The images in the dataset have a resolution of  $1000 \times 1000$ , with a total of about 21,000 nuclei. Professional pathologists label the boundaries.

### 5.16. BNS [80]

BNS is a  $512 \times 512$ -byte breast cancer image data set with 33 HE-stained pathological images. There are also manually labelled nuclei (2754) with tissue data from seven TNBC patients.

### 5.17. Medical Segmentation Decathlon (MSD) [81]

This repository includes segmented images and masks of liver, pancreas, spleen, colon, lungs, brain, hippocampus, prostate, heart, and hepatic vessels.

## 6. Implementation Details

NVIDIA Deep Learning GPUs offer high processing power for deep learning model training. A software development kit (SDK) called NVIDIA CUDA-X AI [82] is intended for researchers and developers creating deep learning models. It utilizes powerful GPUs and satisfies several industrial benchmarks, including MLPerf. Computer vision tasks, recommendation systems, and conversational AI are all developed for NVIDIA CUDA-X AI. The following functionalities are supported by libraries in the NVIDIA Deep Learning SDK:

- Deep learning primitives are pre-built building blocks that can be used to define training elements such as tensor transformations, activation functions, and convolutions;
- Deep learning inference engine, a runtime you may use to deploy models in real-world settings;
- GPU-accelerated transcoding and inference are made possible by deep learning for video analytics, which also offers a high-level C++ runtime and API;
- Linear algebra—uses GPU acceleration to provide functionality for BLAS (basic linear algebra subprograms). Compared to the CPU this is 6–17 times faster;
- Sparse matrix operations let to use of GPU-accelerated BLAS with sparse matrices, such as those required for natural language processing (NLP);
- Multi-GPU communication—allows for group communications over up to eight GPUs, including broadcast, reduction, and all-gather.

Tensor flow [83,84] is a free and open-source end-to-end platform for performing machine learning tasks, and Keras [85,86] is a tensor flow-based neural network library at a high level.

## 7. Comparison of UNET with Other Encoder–Decoder Deep Learning Model

The encoder–decoder deep learning model to segment the medical images alternate to UNET are FCN, FPN, Segnet, and Deeplab. FCN is the first encoder-decoder model. The convolution layer in the FCN [87] is the  $1 \times 1$  convolution, which classifies and creates the mask at the pixel level by upsampling the last convolution layer through the deconvolution layer. However, the global contextual information is not obtained in the FCN, which reduces its segmentation performance and does not tune the parameters according to the image's content. FPN (feature pyramid network) transmits the feature's gradient information from the encoder to the decoder through the skip connection [88]. The depth of the model and separate encoder in the FPN increase the computational complexity [89]. UNET outperforms the segnet by producing higher accuracy in the multi-class classification of the COVID-19 dataset [90]. In addition, the segmentation accuracy for segnet can be improved with UNET. For example, a patch-wise residual-based squeeze U-SegNet model can increase the segmentation accuracy of the brain MRI to segment the GM, WM, and CSF [91]. In Deeplab [92], spatial pyramid pooling is used to adapt the pooling operation according to the different input images. Dilated or atrous convolution and depth separable convolution are other building blocks in the deeplab model applied to consider the spacing between the pixels and reduce the convolutional operation for RGB input.

## 8. Discussion

There are many medical image processing performed using the deep learning technique. However, segmentation is of great interest in diagnosing diseases. UNET can be fine-tuned according to the application and still has significant advancement potential in application range, training speed optimization, feature enhancement and fusion, a small sample training set, and training accuracy. Modified architectures of U-Net have recently been used to achieve precise segmentation of different lesions by embedding attention mechanisms, dense modules, residual structures, and other modules. Choosing an efficient UNET model is challenging; hence, it is implemented for different datasets. Evaluation metrics and limitations of different models are discussed in Table 1. The computational time, learning rate, and contribution of each model are summarized in Table 2.

**Table 1.** Evaluation metrics and limitations of different UNET models.

Model	Type of Disease Diagnosed	Evaluation Metrics	Limitations
UNET with generalized pooling [17]	Tumor	<p>For the BRATS 18 dataset</p> <ul style="list-style-type: none"> <li>• DSC WT-0.839 TC-0.6594 ET-0.7341</li> <li>• PPV WT-0.9175 TC-0.6564 ET-0.8175</li> <li>• Sensitivity WT-0.7879 TC-0.7169 ET-0.7367</li> </ul> <p>For the BRATS19 dataset</p> <ul style="list-style-type: none"> <li>• DSC WT-0.8764 TC-0.7465 ET-0.7926</li> <li>• PPV WT-0.9079 TC-0.7667 ET-0.8801</li> <li>• Sensitivity WT-0.8697 TC-0.8568 ET-0.8167</li> </ul>	Assigning the average initial weight to each element complicates the model.

**Table 1.** Cont.

Model	Type of Disease Diagnosed	Evaluation Metrics	Limitations
Stack Multi-Connection Simple Reducing Net (SMCSRNet) [18]	Tumor	Dice score-0.831 PPV-0.73 Sensitivity-0.87	When stacking more basic blocks (after 10), the performance decreases, and the number of parameters continuously increases. Therefore, it does not perform well for enhanced tumors. However, it is the end-to-end model which predicts the entire image.
3D spatial weighted UNET [19]	Psychological changes in the brain with age.	<ul style="list-style-type: none"> <li>• DSC GM-86.58 ± 1.76% WM-89.87 ± 1.43% CSF-84.81 ± 2.33%</li> <li>• HD GM-1.29 ± 0.25 WM-1.73 ± 0.50 CSF-1.84 ± 0.31</li> <li>• AVD(Absolute volume difference) GM-5.75 ± 3.58 WM-5.47 ± 5.19 CSF-6.84 ± 4.14</li> </ul>	It can be implemented only in the 3D input.
AnatomicallygatedUNET [20]	Alzheimer's disease	ADNI DC-0.8864 ± 0.0212 ASD-0.386 ± 0.058	Two sub-networks increase the segmentation's memory burden. The similarity between an atlas and a segmented MRI is not considered. Image intensity data is not included
MH-UNET [21]	Tumor, stroke	<p>Tumor-</p> <ul style="list-style-type: none"> <li>• DSC WT-90% TC-83% ET-78% HD WT-4.164 TC-9.809 ET-32.200 DC-56.72% HD-23.94</li> <li>Stroke ASSD-7 Precision-65.45 Recall-59.38</li> </ul>	During the segmentation of whole tumor, dice score will become zero.
MI-UNET [22]	Stroke	DC-76.62% Precision-87.97% Recall-67.09%	The registration step occupies computational time. Difficult to segment the small lesions. Attention gating signal should be optimally chosen to increase the recall rate. The model's first part (WUN) has 23.49 M parameters, and the second part (WRUN) has 32.43 M parameters. Therefore, it has to be separately trained.
Multi-Res Attention UNET [24]	Epilepsy	For DRIVE Dataset Accuracy-0.9692 Sensitivity-0.8278 Specificity-0.9861 Precision-0.8637	
GLUE [26]	Ophthalmic diseases	For STARE Dataset Accuracy-0.9740 Sensitivity-0.8342 Specificity-0.9916 Precision-0.8823	

**Table 1.** Cont.

Model	Type of Disease Diagnosed	Evaluation Metrics	Limitations
S-UNET [28]		For CHASE-DB1 dataset MCC-0.8065 SE-0.8044 SP-0.9841 Accuracy-0.9.58 AUC-0.9867 F1 score-0.8242 For MOD dataset Accuracy- $92.82 \pm 0.43$ Precision- $88.54 \pm 0.58$ Recall- $86.46 \pm 0.84$ F1 score- $87.35 \pm 0.75$ IoU- $77.72 \pm 1.15$	For TONGREEN Dataset MCC-0.7806 SE-0.7822 SP-0.9830 Accuracy-0.9652 AUC-0.9824 F1 score-0.7994 For DRIVE dataset MCC-0.8055 SE-0.8312 SP-0.9751 Accuracy-0.9567 AUC-0.9821 F1 score-0.8303
UNET with atrous Separable [29]	Cancer	Accuracy- $92.82 \pm 0.43$ Precision- $88.54 \pm 0.58$ Recall- $86.46 \pm 0.84$ F1 score- $87.35 \pm 0.75$ IoU- $77.72 \pm 1.15$	3.96 million parameters for sepconvolution with atrous and 1.01 million parameters without atrous.
RIC UNET [30]	Cancer	Aggregated Jaccard index-0.5635 Dice-0.8008 F1 score-0.8278	It has a more substantial discrimination effect on some deeper backgrounds. Small regions of the proximal coronary artery are occasionally missed while using this model. Cannot produce high accuracy for segmenting aorta with coronary artery. Attention mechanisms must be carefully selected for each task based on their characteristics
Modified 2D UNET [31]	Coronary artery disease	Only aorta- DC-91.20% IoU-83.82%	Aorta with coronary artery DC-88.80% IoU-79.85%
UCNET with attention Mechanism [32]	Cardiac arrhythmia and Congenital cardiac diseases	Single modality DSC-0.9112 Jaccard-0.8420	Multimodality DSC-0.91112
Cascaded UNET [33]	COVID-19	DSC-62.8%	The tradeoff between TPR and FPR rate.
Res-D-UNET [35]	Pulmonary embolism	For CT lung dataset DSC-0.982 Precision-0.985 Recall-0.980 SSIM-0.961 DSC- $93.47 \pm 0.83$ Accuracy-96.23 Recall-91.79	For CHAOS dataset DSC-0.969 Precision-0.966 Recall-0.968 SSIM-0.951
Radiomics guided -DUN GAN [37]	Liver lesions		Segmentor and discriminator have to be trained separately. Performance is worse for Smaller tumor cells (30cc), while patients with a disturbance in esophageal, hiatal hernia, proximal tumor had no discernible network strength. Only 60% of the training volumes are used in model selection, reducing the trained models' generalization ability and validation performance.
Dilated Dense attention UNET [40]	Esophageal tumor segmentation	DSC- $0.79 \pm 0.20$ , Mean surface distance- $5.4 \pm 20.2$ mm 95% Hausdorff distance- $14.7 \pm 25.0$ mm	Choosing the information weight as 0 and 1 will degrade the late and dual branch network.
HDRDS UNET [41]	Lymph node cancer	DSC-0.7811 SEN-0.9357 HMSD-0.8514	
HF-UNET [42]	Prostate cancer	DC-0.88 ASD-1.31 SEN-0.88 PPV-0.89	

**Table 2.** Summary of Model.

References	Modification in UNET	Dataset		Area of Segmentation	Contributions	Computational Time
		Clinically Available Dataset	Publically Available Dataset			
[17]	Generalized pooling and adaptive weight.		BRAT 2018 and BRAT 2019		Extract valuable features during down-sampling. Generalized pooling is applied to varying data.	Learning rate is 0.0001.
[18]	Stacking three SRUNET. In total, 32 feature maps are added in the last UNET, stacked by a long skip connection to the input image.		BRAT2015		Reduces 4/5 parameters compared to the original UNET. Additionally, it reduces multi-scale feature fusion.	Learning rate- $4 \times 10^{-5}$ . Epoch-12. This model takes 9.6 s to segment the tumor, and training time is 4 h 29 min (two stack level). Therefore, the learning rate is $4 \times 10^{-5}$ . Batch size is 10. Reduces the computational time.
[19]	A volumetric feature recalibration layer is included.		Multi-atlas Labeling (MIAL) MICCAI 2012 Grand Challenge		Spatial information loss can be avoided, and the power of the features can be enhanced.	This model is trained for 20,000 iteration with initial learning rate is 0.001. After that, the learning rate becomes half every 5000 iterations. It takes 1 day to train the model.
[20]	The anatomical gate learns the anatomical features from the brain atlases and guides the segmentation network for segmenting the correct region of interest.		ADNI and LONI-LPBA40	Brain	The feature map learned from the input image fuses with the multi-label atlases to increase segmentation performance.	It takes approximately one day to train the model. Learning rate-0.001, number of epoch is 1000, minibatch size-1.
[21]	Dense block, residual inception block, and hierarchical blocks are included		MICCAI BraTS and ISLES		Gradient vanishing and exploitation gets reduced. Less learnable parameter	For MICCAI BrasChallengedatabase, The learning rate is $4 \times 10^{-5}$ . Batch size is 1. Epochs-300. For ISLES dataset, Initial learning rate $5 \times 10^{-4}$ , Epochs-300, batch size-4
[22]	The LDDMM algorithm performs brain parcellation.		ATLAS		It can be applied to all types of input regardless of the dimensions.	Learning rate 0.001. It takes 140 s to segment strokes. Batch size-32.
[24]	The chain of the $3 \times 3$ kernel is connected in series.	SCTIMST, Trivandrum, India.			Consider the large semantic gap feature map between encoder and decoder. It suppresses redundant features. It reduces higher memory requirements.	The learning rate is 0.0001

**Table 2.** Cont.

References	Modification in UNET	Dataset		Area of Segmentation	Contributions	Computational Time
		Clinically Available Dataset	Publically Available Dataset			
[26]	Weighted attention mechanism, and skip connection are added.		DRIVE and STARE dataset		Data imbalance reduced.	Learning rate is $5 \times 10^{-5}$ . (batch size 128). A number of epochs is 60. DRIVE dataset takes 91 min to train and STARE dataset takes 65 min to train the model. Segments the 20 retinal images within 6.2 s.
				Eye		
[28]	Two MI-UNET with saliency mechanism is included.	TONGREN	DRIVE, HASE_DB1		Data imbalance reduced.	DRIVE dataset-It takes 3 h to train the model and segment the vessel within 33 ms. TONGREN dataset-9 h for training and 0.49 s to segment. CHASE-DB1 dataset-5 hours for training and 91 ms to segment
[29]	Convolutional operation is changed into sep convolution.		MOD and BNS		Size, trainable parameter, and evolution time reduced.	The learning rate is $1 \times 10^{-3}$ . Epochs-50
[30]	Residual block, channel gate, and multi-scale are applied in UNET.		The Cancer Genomic Atlas	Cell or nuclei	Extract the different cell shapes from the dense cell.	The learning rate is 0.0001, which is reduced by ten percent per 1000 iterations. Batch size is 2. Epoch-100
[31]	Batch normalization and dropout layer are added.	University College Hospital London and Barts Health NHS Trust.			Reduced overfitting and stabilized the training process.	The learning rate is $1 \times 10^{-5}$ . Epochs-200. Segmenting time is 40–41 s.
[32]	SNEM, attention mechanism, and clique UNET are included.	Cardiac CT angiography at Shuguang Hospital, Shanghai, China.		Heart	More salient features can focus.	Learning rate 0.001, drop out rate is 0.8. Epochs-80,000
[33]	Expectation maximization algorithm.	CT datasets from Iran, Italy, South Korea, and the United States from multiple institutions			Semantic label not required.	The learning rate is 0.0005
[35]	Residual and dense networks are embedded in UNET.	China-Japan Friendship Hospital,	CHAOS CT images	Lung	Attenuate the problem of degradation and vanishing gradient. Overfitting gets reduced.	Learning Rate $2 \times 10^{-4}$ (batch size is 4). Running time 1096.7 s. Number of epochs is 100.

**Table 2.** Cont.

References	Modification in UNET	Dataset	Area of Segmentation	Contributions	Computational Time
		Clinically Available Dataset	Publically Available Dataset		
[37]	Radiomics features, dense layer, and GAN are added.	McGill University Health Centre	Liver	Network converges faster and smoother.	The learning rate is $1 \times 10^{-6}$ for segmentor and discriminator. Batch size is 2 for segmentor and 64 for discriminator.
[40]	Dilated dense spatial attention gate and channel attention gate are included.	Dataset approved by Leiden University Medical Center's Medical Ethics Review Committee in The Netherlands	Esophageal	Receptive field increases without increasing the network size	Training time-6 days. Batch size 7
[41]	Hyper dense encoder and recurrent dense siamese decoder are added.	General Hospital of Shenyang Military Area Command(F-FDG PET/CT Scan)	Lymphoma	Stable gradient, explore spatial-temporal correlation.	The initial learning rate is 0.001, and it will be halved after each 10,000 iterations. Validation of model is performed after each 200 iterations.
[42]	The contour extracts the prostate region. Attention-based task consistency learning block learns the data from segmentation and regression.	National Cancer Institute—International Symposium on Biomedical Imaging (NCI-ISBI) 2013 Automated Segmentation of Prostate Structures Challenge dataset.	Prostate	Accurate contours are created to segment the prostate.	A number of epochs 60. The learning rate is decreased from 0.01 to 0.0001 by a step size of $2 \times 10^{-5}$

## 9. Conclusions and Future Work

Clinical applications and academic research are significantly influenced by the analysis and processing of medical data. Deep learning can generate novel concepts for medical image techniques that enable texture morphology detection purely from data. It has emerged as the primary component in numerous medical image research. The outcomes demonstrate that the DL approach on CNN has received widespread acclaim for its medical image segmentation, classification, and other areas. This article examines the evolution of UNET architecture for segmenting the region of interest from different internal organs. This review also specified the evaluation metrics and segmentation regions obtained from the UNET models according to the diseases. In future work, segmentation accuracy can be improved by increasing the segmentation validation metrics. UNET can be cascaded with GAN for synthesizing the medical images and can be utilized for efficiently segmenting, classifying, and synthesizing the images. The architecture of UNET can be modified to predict the statistical information from the segmented region.

**Author Contributions:** Conceptualization, formal analysis, investigation and writing draft, M.K.a.A.; Supervision, K.S. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** No new data were created or analyzed in this study. Data sharing is not applicable to this article.

**Conflicts of Interest:** There are no conflicts of interest that could be perceived as prejudicing the impartiality of the research reported.

## References

- Li, B.N.; Wang, X.; Wang, R.; Zhou, T.; Gao, R.; Caccio, E.J.; Green, P.H. Celiac Disease Detection from Videocapsule Endoscopy Images Using Strip Principal Component Analysis. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2021**, *18*, 1396–1404. [CrossRef] [PubMed]
- Chang, H.-H.; Hsieh, C.-C. Brain segmentation in MR images using a texture-based classifier associated with mathematical morphology. In Proceedings of the 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Jeju Island, Republic of Korea, 11–15 July 2017; pp. 3421–3424. [CrossRef]
- Venkatachalam, K.; Siuly, S.; Bacanin, N.; Hubalovsky, S.; Trojovsky, P. An Efficient Gabor Walsh-Hadamard Transform Based Approach for Retrieving Brain Tumor Images from MRI. *IEEE Access* **2021**, *9*, 119078–119089. [CrossRef]
- Haghghi, S.J.; Komeili, M.; Hatzinakos, D.; El Beheiry, H. 40-Hz ASSR for Measuring Depth of Anaesthesia During Induction Phase. *IEEE J. Biomed. Health Inform.* **2018**, *22*, 1871–1882. [CrossRef] [PubMed]
- Tang, C.; Yu, C.; Gao, Y.; Chen, J.; Yang, J.; Lang, J.; Liu, C.; Zhong, L.; He, Z.; Lv, J. Deep learning in the nuclear industry: A survey. *Big Data Min. Anal.* **2022**, *5*, 140–160. [CrossRef]
- Jalali, S.M.J.; Osorio, G.J.; Ahmadian, S.; Lotfi, M.; Campos, V.M.A.; Shafie-Khah, M.; Khosravi, A.; Catalao, J.P.S. New Hybrid Deep Neural Architectural Search-Based Ensemble Reinforcement Learning Strategy for Wind Power Forecasting. *IEEE Trans. Ind. Appl.* **2022**, *58*, 15–27. [CrossRef]
- Tran, M.-Q.; Elsisi, M.; Liu, M.-K.; Vu, V.Q.; Mahmoud, K.; Darwish, M.M.F.; Abdelaziz, A.Y.; Lehtonen, M. Reliable Deep Learning and IoT-Based Monitoring System for Secure Computer Numerical Control Machines Against Cyber-Attacks with Experimental Verification. *IEEE Access* **2022**, *10*, 23186–23197. [CrossRef]
- Cao, Q.; Zhang, W.; Zhu, Y. Deep learning-based classification of the polar emotions of “moe”-style cartoon pictures. *Tsinghua Sci. Technol.* **2021**, *26*, 275–286. [CrossRef]
- Liu, S.; Xia, Y.; Shi, Z.; Yu, H.; Li, Z.; Lin, J. Deep Learning in Sheet Metal Bending with a Novel Theory-Guided Deep Neural Network. *IEEE/CAA J. Autom. Sin.* **2021**, *8*, 565–581. [CrossRef]
- Monteiro, N.R.C.; Ribeiro, B.; Arrais, J.P. Drug-Target Interaction Prediction: End-to-End Deep Learning Approach. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2020**, *18*, 2364–2374. [CrossRef]
- Mohsen, S.; Elkaseer, A.; Scholz, S.G. Industry 4.0-Oriented Deep Learning Models for Human Activity Recognition. *IEEE Access* **2021**, *9*, 150508–150521. [CrossRef]
- Lee, S.Y.; Tama, B.A.; Choi, C.; Hwang, J.-Y.; Bang, J.; Lee, S. Spatial and Sequential Deep Learning Approach for Predicting Temperature Distribution in a Steel-Making Continuous Casting Process. *IEEE Access* **2020**, *8*, 21953–21965. [CrossRef]
- Usamentiaga, R.; Lema, D.G.; Pedrayes, O.D.; Garcia, D.F. Automated Surface Defect Detection in Metals: A Comparative Review of Object Detection and Semantic Segmentation Using Deep Learning. *IEEE Trans. Ind. Appl.* **2022**, *58*, 4203–4213. [CrossRef]
- Minaee, S.; Boykov, Y.Y.; Porikli, F.; Plaza, A.J.; Kehtarnavaz, N.; Terzopoulos, D. Image segmentation using deep learning: A survey. *arXiv* **2020**, arXiv:2001.05566. [CrossRef] [PubMed]
- Liu, L.; Cheng, J.; Quan, Q.; Wu, F.-X.; Wang, Y.-P.; Wang, J. A survey on U-shaped networks in medical image segmentations. *Neurocomputing* **2020**, *409*, 244–258. [CrossRef]
- Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015*; Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F., Eds.; Springer: Cham, Switzerland, 2015; pp. 234–241. [CrossRef]
- Xu, Y.; Zhou, Z.; Li, X.; Zhang, N.; Zhang, M.; Wei, P. FFU-Net: Feature Fusion U-Net for Lesion Segmentation of Diabetic Retinopathy. *BioMed Res. Int.* **2021**, *2021*, 6644071. [CrossRef]
- Du, G.; Cao, X.; Liang, J.; Chen, X.; Zhan, Y. Medical Image Segmentation based on U-Net: A Review. *J. Imaging Sci. Technol.* **2020**, *64*, 20508. [CrossRef]
- Siddique, N.; Paheding, S.; Elkin, C.P.; Devabhaktuni, V. U-Net and its variants for medical image segmentation: A review of theory and applications. *IEEE Access* **2021**, *9*, 82031–82057. [CrossRef]
- Hao, K.; Lin, S.; Qiao, J.; Tu, Y. A Generalized Pooling for Brain Tumor Segmentation. *IEEE Access* **2021**, *9*, 159283–159290. [CrossRef]
- Ding, Y.; Chen, F.; Zhao, Y.; Wu, Z.; Zhang, C.; Wu, D. A Stacked Multi-Connection Simple Reducing Net for Brain Tumor Segmentation. *IEEE Access* **2019**, *7*, 104011–104024. [CrossRef]
- Sun, L.; Ma, W.; Ding, X.; Huang, Y.; Liang, D.; Paisley, J. A 3D Spatially Weighted Network for Segmentation of Brain Tissue From MRI. *IEEE Trans. Med. Imaging* **2020**, *39*, 898–909. [CrossRef]
- Sun, L.; Shao, W.; Zhang, D.; Liu, M. Anatomical Attention Guided Deep Networks for ROI Segmentation of Brain MR Images. *IEEE Trans. Med. Imaging* **2020**, *39*, 2000–2012. [CrossRef] [PubMed]
- Ahmad, P.; Jin, H.; Alroobaee, R.; Qamar, S.; Zheng, R.; Alnajjar, F.; Aboudi, F. MH UNet: A Multi-Scale Hierarchical Based Architecture for Medical Image Segmentation. *IEEE Access* **2021**, *9*, 148384–148408. [CrossRef]

25. Zhang, Y.; Wu, J.; Liu, Y.; Chen, Y.; Wu, E.X.; Tang, X. MI-UNet: Multi-Inputs UNet Incorporating Brain Parcellation for Stroke Lesion Segmentation from T1-Weighted Magnetic Resonance Images. *IEEE J. Biomed. Health Inform.* **2020**, *25*, 526–535. [CrossRef] [PubMed]
26. Wu, J.; Tang, X. A Large Deformation Diffeomorphic Framework for Fast Brain Image Registration via Parallel Computing and Optimization. *Neuroinformatics* **2020**, *18*, 251–266. [CrossRef]
27. Thomas, E.; Pawan, S.J.; Kumar, S.; Horo, A.; Niyas, S.; Vinayagamani, S.; Kesavadas, C.; Rajan, J. Multi-Res-Attention UNet: A CNN Model for the Segmentation of Focal Cortical Dysplasia Lesions from Magnetic Resonance Images. *IEEE J. Biomed. Health Inform.* **2021**, *25*, 1724–1734. [CrossRef]
28. Ibtehaz, N.; Rahman, M.S. MultiResUNet: Rethinking the U-Net architecture for multimodal biomedical image segmentation. *Neural Netw.* **2020**, *121*, 74–87. [CrossRef]
29. Lian, S.; Li, L.; Lian, G.; Xiao, X.; Luo, Z.; Li, S. A Global and Local Enhanced Residual U-Net for Accurate Retinal Vessel Segmentation. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2019**, *18*, 852–862. [CrossRef]
30. Pour, A.M.; Seyedarabi, H.; Jahromi, S.H.A.; Javadzadeh, A. Automatic Detection and Monitoring of Diabetic Retinopathy Using Efficient Convolutional Neural Networks and Contrast Limited Adaptive Histogram Equalization. *IEEE Access* **2020**, *8*, 136668–136673. [CrossRef]
31. Hu, J.; Wang, H.; Gao, S.; Bao, M.; Liu, T.; Wang, Y.; Zhang, J. S-UNet: A Bridge-Style U-Net Framework with a Saliency Mechanism for Retinal Vessel Segmentation. *IEEE Access* **2019**, *7*, 174167–174177. [CrossRef]
32. Pan, X.; Li, L.; Yang, D.; He, Y.; Liu, Z.; Yang, H. An Accurate Nuclei Segmentation Algorithm in Pathological Image Based on Deep Semantic Network. *IEEE Access* **2019**, *7*, 110674–110686. [CrossRef]
33. Zeng, Z.; Xie, W.; Zhang, Y.; Lu, Y. RIC-Unet: An Improved Neural Network Based on Unet for Nuclei Segmentation in Histology Images. *IEEE Access* **2019**, *7*, 21420–21428. [CrossRef]
34. Cheung, W.K.; Bell, R.; Nair, A.; Menezes, L.J.; Patel, R.; Wan, S.; Chou, K.; Chen, J.; Torii, R.; Davies, R.H.; et al. A Computationally Efficient Approach to Segmentation of the Aorta and Coronary Arteries Using Deep Learning. *IEEE Access* **2021**, *9*, 108873–108888. [CrossRef]
35. Wang, W.; Ye, C.; Zhang, S.; Xu, Y.; Wang, K. Improving Whole-Heart CT Image Segmentation by Attention Mechanism. *IEEE Access* **2020**, *8*, 14579–14587. [CrossRef]
36. Wu, D.; Gong, K.; Arru, C.D.; Homayounieh, F.; Bizzo, B.; Buch, V.; Ren, H.; Kim, K.; Neumark, N.; Xu, P.; et al. Severity and Consolidation Quantification of COVID-19 From CT Images Using Deep Learning Based on Hybrid Weak Labels. *IEEE J. Biomed. Health Inform.* **2020**, *24*, 3529–3538. [CrossRef]
37. Zhu, W.; Vang, Y.S.; Huang, Y.; Xie, X. Deepem: Deep 3d convnets with em for weakly supervised pulmonary nodule detection. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Granada, Spain, 16–20 September 2018; pp. 812–820. [CrossRef]
38. Yuan, H.; Liu, Z.; Shao, Y.; Liu, M. ResD-Unet Research and Application for Pulmonary Artery Segmentation. *IEEE Access* **2021**, *9*, 67504–67511. [CrossRef]
39. Shiradkar, R.; Ghose, S.; Jambor, I.; Taimen, P.; Ettala, O.; Purysko, A.S.; Madabhushi, A. Radiomic features from pretreatment biparametric MRI predict prostate cancer biochemical recurrence: Preliminary findings. *J. Magn. Reson. Imaging* **2018**, *48*, 1626–1636. [CrossRef]
40. Xiao, X.; Qiang, Y.; Zhao, J.; Yang, X.; Yang, X. Segmentation of Liver Lesions without Contrast Agents with Radiomics-Guided Densely UNet-Nested GAN. *IEEE Access* **2020**, *9*, 2864–2878. [CrossRef]
41. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. *NIPS* **2017**, *60*, 84–90. [CrossRef]
42. van Griethuysen, J.J.M.; Fedorov, A.; Parmar, C.; Hosny, A.; Aucoin, N.; Narayan, V.; Beets-Tan, R.G.H.; Fillion-Robin, J.-C.; Pieper, S.; Aerts, H.J.W.L. Computational radiomics system to decode the radiographic phenotype. *Cancer Res.* **2017**, *77*, e104–e107. [CrossRef]
43. Yousefi, S.; Sokooti, H.; Elmahdy, M.S.; Lips, I.M.; Shalmani, M.T.M.; Zinkstok, R.T.; Dankers, F.J.W.M.; Staring, M. Esophageal Tumor Segmentation in CT Images Using a Dilated Dense Attention Unet (DDAUnet). *IEEE Access* **2021**, *9*, 99235–99248. [CrossRef]
44. Wang, M.; Jiang, H.; Shi, T.; Yao, Y.-D. HD-RDS-UNet: Leveraging Spatial-Temporal Correlation Between the Decoder Feature Maps for Lymphoma Segmentation. *IEEE J. Biomed. Health Inform.* **2022**, *26*, 1116–1127. [CrossRef] [PubMed]
45. He, K.; Lian, C.; Zhang, B.; Zhang, X.; Cao, X.; Nie, D.; Gao, Y.; Zhang, J.; Shen, D. HF-UNet: Learning Hierarchically Inter-Task Relevance in Multi-Task U-Net for Accurate Prostate Segmentation in CT Images. *IEEE Trans. Med. Imaging* **2021**, *40*, 2118–2128. [CrossRef]
46. Dice, L.R. Measures of the amount of ecologic association between species. *Ecology* **1945**, *26*, 297–302. [CrossRef]
47. Fleiss, J.L. The measurement of interrater agreement. In *Statistical Methods for Rates and Proportions*, 2nd ed.; John Wiley & Sons: New York, NY, USA, 1981; pp. 212–236.
48. Oktay, O.; Ferrante, E.; Kamnitsas, K.; Heinrich, M.; Bai, W.; Caballero, J.; Cook, S.A.; De Marvao, A.; Dawes, T.; O'Regan, D.P.; et al. Anatomically constrained neural networks (ACNNs): Application to cardiac image enhancement and segmentation. *IEEE Trans. Med. Imaging* **2018**, *37*, 384–395. [CrossRef]
49. Dalca, A.V.; Guttag, J.; Sabuncu, M.R. Anatomical priors in convolutional networks for unsupervised biomedical segmentation. In Proceedings of the IEEE/CVF Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 9290–9299. [CrossRef]

50. Larrazabal, A.J.; Martinez, C.; Ferrante, E. Anatomical priors for image segmentation via post-processing with denoising autoencoders. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Shenzhen, China, 13–17 October 2019; Springer: Cham, Switzerland, 2021; Volume 9, pp. 585–593. [CrossRef]
51. Ito, R.; Nakae, K.; Hata, J.; Okano, H.; Ishii, S. Semi-supervised deep learning of brain tissue segmentation. *Neural Netw.* **2019**, *116*, 25–34. [CrossRef] [PubMed]
52. de Vos, B.D.; Berendsen, F.F.; Viergever, M.A.; Sokooti, H.; Staring, M.; Işgum, İ. A deep learning framework for unsupervised affine and deformable image registration. *Med. Image Anal.* **2019**, *52*, 128–143. [CrossRef]
53. Chi, W.; Ma, L.; Wu, J.; Chen, M.; Lu, W.; Gu, X. Deep learning-based medical image segmentation with limited labels. *Phys. Med. Biol.* **2020**, *65*, 235001. [CrossRef]
54. He, Y.; Yang, G.; Chen, Y.; Kong, Y.; Wu, J.; Tang, L.; Zhu, X.; Dillenseger, J.-L.; Shao, P.; Zhang, S.; et al. DPA-DenseBiasNet: Semi-supervised 3D fine renal artery segmentation with dense biased network and deep prior anatomy. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Shenzhen, China, 13–17 October 2019; Springer: Cham, Switzerland, 2019; pp. 139–147.
55. Dong, S.; Luo, G.; Tam, C.; Wang, W.; Wang, K.; Cao, S.; Chen, B.; Zhang, H.; Li, S. Deep atlas network for efficient 3D left ventricle segmentation on echocardiography. *Med. Image Anal.* **2020**, *61*, 101638. [CrossRef]
56. Zheng, H.; Lin, L.; Hu, H.; Zhang, Q.; Chen, Q.; Iwamoto, Y.; Han, X.; Chen, Y.-W.; Tong, R.; Wu, J. Semi-supervised segmentation of liver using adversarial learning with deep atlas prior. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Shenzhen, China, 13–17 October 2019; Springer: Cham, Switzerland, 2019; pp. 148–156. [CrossRef]
57. Imran, A.; Li, J.; Pei, Y.; Yang, J.-J.; Wang, Q. Comparative Analysis of Vessel Segmentation Techniques in Retinal Images. *IEEE Access* **2019**, *7*, 114862–114887. [CrossRef]
58. García, V.; Dominguez, H.D.J.O.; Mederos, B. Analysis of Discrepancy Metrics Used in Medical Image Segmentation. *IEEE Lat. Am. Trans.* **2015**, *13*, 235–240. [CrossRef]
59. Eelbode, T.; Bertels, J.; Berman, M.; Vandermeulen, D.; Maes, F.; Bisschops, R.; Blaschko, M.B. Optimization for Medical Image Segmentation: Theory and Practice When Evaluating with Dice Score or Jaccard Index. *IEEE Trans. Med. Imaging* **2020**, *39*, 3679–3690. [CrossRef]
60. Khan, M.Z.; Gajendran, M.K.; Lee, Y.; Khan, M.A. Deep Neural Architectures for Medical Image Semantic Segmentation: Review. *IEEE Access* **2021**, *9*, 83002–83024. [CrossRef]
61. Landman, B.A.; Warfield, S. MICCAI 2012: Grand challenge and workshop on multi-atlas labeling. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Nice, France, 1–5 October 2012; Volume 2012.
62. Mendrik, A.M.; Vincken, K.L.; Kuijf, H.J.; Breeuwer, M.; Bouvy, W.H.; de Bresser, J.; Alansary, A.; de Bruijne, M.; Carass, A.; El-Baz, A.; et al. MRBrains challenge: Online evaluation framework for brain image segmentation in 3T MRI scans. *Comput. Intell. Neurosci.* **2015**, *2015*, 813696. [CrossRef]
63. Valverde, S.; Oliver, A.; Cabezas, M.; Roura, E.; Lladó, X. Comparison of 10 brain tissue segmentation methods using revisited IBSR annotations. *J. Magn. Reson. Imaging* **2015**, *41*, 93–101. [CrossRef] [PubMed]
64. Menze, B.H.; Jakab, A.; Bauer, S.; Kalpathy-Cramer, J.; Farahani, K.; Kirby, J.; Burren, Y.; Porz, N.; Slotboom, J.; Wiest, R.; et al. The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE Trans. Med. Imaging* **2015**, *34*, 1993–2024. [CrossRef]
65. Available online: <https://www.med.upenn.edu/sbia/brats2018/registration.html> (accessed on 22 April 2022).
66. Jack, C.R.; Bernstein, M.A.; Fox, N.C.; Thompson, P.; Alexander, G.; Harvey, D.; Borowski, B.; Britson, P.J.; Whitwell, J.L.; Ward, C.; et al. The Alzheimer’s disease neuroimaging initiative (ADNI): MRI methods. *J. Magn. Reson. Image.* **2008**, *27*, 685–691. [CrossRef]
67. Available online: <http://adni.loni.usc.edu/ADNI> (accessed on 15 December 2020).
68. Shattuck, D.W.; Mirza, M.; Adisetiyo, V.; Hojatkashani, C.; Salamon, G.; Narr, K.L.; Poldrack, R.A.; Bilder, R.M.; Toga, A.W. Construction of a 3D probabilistic atlas of human cortical structures. *NeuroImage* **2008**, *39*, 1064–1080. [CrossRef]
69. Owen, C.G.; Rudnicka, A.; Mullen, R.; Barman, S.; Monekosso, D.; Whincup, P.; Ng, J.; Paterson, C. Measuring retinal vessel tortuosity in 10-year-old children: Validation of the computer-assisted image analysis of the retina (CAIAR) program. *Investig. Ophthalmol. Vis. Sci.* **2009**, *50*, 2004–2010. [CrossRef] [PubMed]
70. Available online: <https://drive.grand-challenge.org/> (accessed on 23 January 2022).
71. Available online: <https://cecas.clemson.edu/ahoover/stare/> (accessed on 4 March 2022).
72. Hu, Q.; Abràmoff, M.D.; Garvin, M.K. Automated separation of binary overlapping trees in low-contrast color retinal images. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Nagoya, Japan, 22–26 September 2013; Springer: Berlin/Heidelberg, Germany, 2013; pp. 436–443. [CrossRef]
73. Hoover, A.; Kouznetsova, V.; Goldbaum, M. Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response. *IEEE Trans. Med. Imaging* **2000**, *19*, 203–210. [CrossRef] [PubMed]
74. Yan, T. CCAP, IEEE Dataport, 2020. 2020. Available online: <https://doi.org/10.21227/ccgv-5329> (accessed on 4 March 2022).
75. Soares, E.; Angelov, P.; Biaso, S.; Froes, M.H.; Abe, D.K. SARS-CoV-2 CT-scan dataset: A large dataset of real patients CT scans for SARS-CoV-2 identification. *MedRxiv* **2020**. [CrossRef]
76. CHAOS-Combined (CT-MR) Healthy Abdominal Organ Segmentation. Available online: [https://chaos.grand-challenge.org/Combined\\_Healthy\\_Abdominal\\_Organ\\_Segmentation/](https://chaos.grand-challenge.org/Combined_Healthy_Abdominal_Organ_Segmentation/) (accessed on 6 May 2022).

77. The ISLES Challenge 2018 Website. Available online: <https://www.smir.ch/ISLES/Start2018> (accessed on 5 November 2021).
78. The Cancer Genome Atlas (TCGA). Available online: <http://cancergenome.nih.gov/> (accessed on 14 May 2016).
79. Kumar, N.; Verma, R.; Sharma, S.; Bhargava, S.; Vahadane, A.; Sethi, A. A dataset and a technique for generalized nuclear segmentation for computational pathology. *IEEE Trans. Med. Imaging* **2017**, *36*, 1550–1560. [CrossRef]
80. Naylor, P.; Lae, M.; Reyal, F.; Walter, T. Nuclei segmentation in histopathology images using deep neural networks. In Proceedings of the 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI), Melbourne, VIC, Australia, 18–21 April 2017; pp. 933–936. [CrossRef]
81. Available online: <http://medicaldecathlon.com/index.html> (accessed on 19 September 2022).
82. Available online: <https://developer.nvidia.com/deep-learning-software> (accessed on 7 June 2022).
83. Available online: <https://www.tensorflow.org/> (accessed on 9 February 2022).
84. Abadi, M.; Barham, P.; Chen, J.; Chen, Z.; Davis, A.; Dean, J.; Devin, M.; Ghemawat, S.; Irving, G.; Isard, M.; et al. Tensorflow: A system for large-scale machine learning. In Proceedings of the 12th USENIX conference on Operating Systems Design and Implementation, Savannah, GA, USA, 2–4 November 2016; pp. 28–265.
85. Available online: <https://keras.io> (accessed on 10 August 2022).
86. Li, A.; Li, Y.-X.; Li, X.-H. Tensor flow and Keras-based convolutional neural network in CAT image recognition. In Proceedings of the 2nd International Conference Computational Modeling, Simulation Applied Mathematics (CMSAM), Beijing, China, 22 October 2017; p. 5.
87. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE/CVF Computer Vision and Pattern Recognition Conference (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 3431–3440.
88. Lin, T.-Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature pyramid networks for object detection. In Proceedings of the IEEE/CVF Computer Vision and Pattern Recognition Conference (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 2117–2125.
89. Syazwany, N.S.; Nam, J.-H.; Lee, S.-C. MM-BiFPN: Multi-Modality Fusion Network with Bi-FPN for MRI Brain Tumor Segmentation. *IEEE Access* **2021**, *9*, 160708–160720. [CrossRef]
90. Saood, A.; Hatem, I. COVID-19 lung CT image segmentation using deep learning methods: U-Net versus SegNet. *BMC Med. Imaging* **2021**, *21*, 19. [CrossRef]
91. Dayananda, C.; Choi, J.Y.; Lee, B. A Squeeze U-SegNet Architecture Based on Residual Convolution for Brain MRI Segmentation. *IEEE Access* **2022**, *10*, 52804–52817. [CrossRef]
92. Chen, L.C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *40*, 834–848. [CrossRef]