Evolutionary Attention Network for Medical Image Segmentation

Tahereh Hassanzadeh
School of Engineering and Information Technology (SEIT)
University of New South Wales
Canberra, Australia
t.hassanzadehkoohi@student.unsw.edu.au

Daryl Essam
SEIT
University of New South Wales
Canberra, Australia
d.essam@adfa.edu.au

Ruhul Sarker
SEIT
University of New South Wales
Canberra, Australia
r.sarker@adfa.edu.au

***Abstract*—Medical image segmentation is an active research topic to analyse medical images to find an organ or possible abnormalities in an image. Using a Convolutional Neural Network (CNN) is a successful technique for medical image segmentation. However, developing a CNN is a difficult task, especially when it includes complex structures, such as an attention mechanism. A CNN equipped with an attention mechanism is able to focus on a specific part of an image to extract a Region Of Interest (ROI), that can play a significant role to increase the accuracy of an image segmentation. Due to the difficulty of developing an attention network, in this paper, we introduce a new evolutionary technique to generate an attention network automatically for medical image segmentation. To the best of our knowledge, this is the first attempt to create an attention network using an evolutionary technique. To do this, a new encoding model is introduced to create a network topology, along with its training parameters, to ease the complexity of developing a CNN. Also, a Genetic Algorithm (GA) is applied to evolve the networks. To show the capability of the proposed technique, we used three publicly available medical segmentation datasets. The obtained results show that the proposed model can generate networks corresponding to each dataset, such that the developed networks have high performance for medical image segmentation.**

***Index Terms*—Medical Image segmentation, Convolutional Neural Network, Evolutionary Attention Network, Genetic Algorithm.**

I. Introduction

Image segmentation is the process of partitioning an image into multiple sections. By doing this, the aim is to find a Region Of Interest (ROI), that can be one or multiple organs or abnormalities, such as a tumour, cancer, or lesion. Medical image segmentation is a challenging task because the image can be noisy and contains artefacts that can be the cause of the utilised imaging technology and/or patients movement during the imaging. Furthermore, in most of the cases, the ROI is a tiny part of an image and/or the ROI colour, and texture is similar to the surrounding organs that makes image segmentation even more difficult.

A Convolutional Neural Network (CNN) is a kind of Artificial Neural Network (ANN), which is mostly utilised for image analysis [1]. CNN has been applied in various applications, including image segmentation, image classification, and object detection [2]–[5]. The success of CNN to analyse various types of images has encouraged researchers to apply CNN for medical image analysis [6]–[8]. To improve the accuracy of CNN, several techniques have been introduced in the literature. For example, Dense network, proposed by Huang et al. [9], applied dense connections to improve the quality of extracted features. Also, Residual network [10], utilising residual connections is another method to enhance the accuracy of a network.

Attention mechanism is another technique to increase the quality of extracted features. Attention networks were first introduced by Bahdanau et al. [11] for machine learning translation, which then became a predominant concept in the area of Artificial Intelligence (AI) and neural networks [12]. Attention is inspired by the cognitive process, where, selectively, the focus is on some specific part of the information. Particularly, in image analysis, some part of the input image can have more valuable information compared to others. Therefore, applying the attention mechanism along with a neural network can help the network to concentrate on a specific part of the image. Several papers addressed the application of attention mechanism and neural network to solve various problems in machine vision [13]. Generally, attention mechanism can be created using various types of connections and operations.

Creating a CNN for a specific application, notably when using the more complex structures such as using Dense [9], or Residual [10] connections, and particularly attention mechanis,m is a difficult task. The use of a trial-and-error approach (also known as manual approach) for designing such an intricate network requires a considerable amount of effort, time and computation. To ease the process of developing a neural network, an evolutionary technique can be applied. However, as of the literature, only a few evolutionary techniques are reported that help to improve attention networks [14], [15]. However, the existing evolutionary approaches were only used to simply evolve a specific part of a network.

In this paper, leveraging the capability of an attention neural network to segment images and the difficulty of generating an attention neural network, we propose an automatic evolutionary approach to create an attention network. To do this we develop a new encoding model to generate a network topology equipped with attention section. Furthermore, all the network training parameters are determined using the evolutionary model. A Genetic Algorithm (GA) [16], is applied to evolve the network structures along with their parameters. In the proposed model, the final network is a U–Net [17] based network, such that network structure is equipped with an attention mechanism. Three publicly available datasets are utilised for evaluating the proposed model . The obtained results show that the proposed model is general enough to apply for segmentation of various types of medical images. Besides, the automatically designed model outperform previous manual–designed and automated–designed networks. Therefore, our research contribution is defined as follows:

A. Research Contribution

- In this paper, for the first time, we propose an evolutionary technique using GA to develop an evolutionary U–Net–based attention network automatically.
- We propose four different attention gates.
- The proposed encoding model can develop networks with various sizes; however, it is compatible with standard crossover and mutation methods.
- The proposed attention network can segment various medical images precisely.
- Users without having in–depth knowledge about deep learning and attention network will be able to develop networks for various other segmentation applications.

The rest of this paper is organised as follows. Section 2 provides a literature review concerning the attention network and evolutionary networks. Section 3 demonstrates the proposed model. The dataset and experimental results are discussed in section 4. Section 5 provides the discussion and conclusion.

II. Literature Review

Since this paper is the combination of an evolutionary network and an attention network, the literature review is divided into two sections. The first part provides a review on evolutionary networks, and the second contains a summary on attention networks.

A. Evolutionary Network

In regard to the difficulty of developing a neural network, there are techniques to generate networks automatically, including using evolutionary techniques (Neuroevolution) [18] and also Reinforcement Learning (RL) [19]. Neuroevolution means using an evolutionary computation technique to develop a neural network. Montana et al. [20] introduced this technique in 1989. The very early works in this area mainly used an evolutionary model for the network's weights initialisation [20]–[22]. Gradually more parameters have been included in the search space and evolutionary models applied to develop networks for various applications, such as image classification, image segmentation, and speech recognition. [23]–[26].

EvoDeep [27] is an evolutionary technique to develop the topology and parameters of a neural network for image classification applications. EvoDeep applied a state machine to find a valid sequence of layers. Besides, the unique encoding that they utilised, a new crossover and mutation model was also introduced. Moreover, another evolutionary model to generate a network for image classification proposed by Wang et al. [28]. In their proposed model, a Differential Evolution (DE) algorithm, along with a new crossover and mutation method, was used to create variable length network structures. Also, to create the genotype of the possible solutions an IP (Internet Protocol) based encoding model proposed. Sun et al. [29] also offers another evolutionary network for image classification. In the proposed model, a GA [30] as well as the combination of Dense blocks [9] and Residual blocks [10] are utilised to create a block–based evolutionary network. To do this, three types of units, including Dense Unit (DN) (a combination of the multiple Dense blocks), Res Unit (RU) (a combination of multiple Residual blocks) and Pooling Layer Unit (PLU) are used to create networks.

Evolutionary techniques are also applied to fine–tune or create networks for image segmentation applications. For example, a classification model for the segmentation of brain MRI using an evolutionary Artificial Neural Network (ANN) is proposed by Kumari et al. [31]. In this model, using feature detection techniques, several features were extracted and then an evolutionary ANN is used for classification of the features. Besides, an evolutionary U–Net [17] based model was introduced to create a deep CNN for medical image segmentation [32]. The U–Net was created from down–sampling (for feature extraction), up–sampling (for segmentation reconstruction) and a bridging block to connect those sections. Baldeon et al. [32], used a fixed U–Net–based network structure, however, a multi–objective evolutionary algorithm was applied to find a part of the hyperparameters, to create a network with high accuracy and fewer parameters. Another U–Net–based evolutionary model for medical image segmentation was introduced by Hassanzadeh et al. [33], such that the whole network structure along with its hyperparameters were found using a GA. In their paper, a new encoding model to create a variable length network was introduced.

Based on the literature, most of the previous proposed evolutionary models have been applied to create networks for classification applications. At the same time, only limited papers are available in the area of segmentation. Furthermore, most of them addressed the issue of finding the best values for only a part of the hyperparameters used to create whole network structures.

B. Attention Network

Attention mechanism helps the network to focus and concentrate on more important information. This technique is mostly applied in the area of Natural Language Processing (NLP) [34]–[36] and computer vision for image or video analysis [37]–[39]. In the following, a review on using attention networks for medical image processing is provided.

Zhang et al. [37] introduced a deep CNN equipped with an attention mechanism for skin lesion classification. In the proposed model, to force the network to focus on the semantically meaningful lesion part, a combination of a Residual connection [10], and an attention connection was utilised. Leveraging the proposed model the number of learning layers and consequently, the number of parameters is also reduced. To create the attention mechanism, the element–wise sum of the block's input, block's output, and the multiplication of the input and weighted output (with softmax activation function) is used. Another attention CNN for skin lesion recognition was developed by Yan et al. [40]. Two attention modules are used after the last two blocks of the VGG16 [41] network structures to refine the extracted features. Each attention module is able to improve features with element–wise sum and pixel–wise multiplication on the intermediate features (which are extracted using convolution layers).

Another deep attention CNN was proposed by Wang et al. [42] for prostrate ultrasound segmentation. In the proposed model, deep attention features, which were obtained from different layers, are used to refine the extracted features of each layer. Therefore, it can suppress the non-prostate pixels at shallow layers and add increasingly more prostate details into features at deep layers. Moreover, they developed an attention network for 3D ultrasound prostate segmentation [43]. In this paper, they utilised single and multiple layer features along with the combination of the above features that are obtained after applying several operations as an attention mechanism, to get more informative features.

Besides, attention gate has been utilised to improve the accuracy of a U–Net–based network [44]. In the proposed model, an attention gate that is constructed from two convolution layers followed by activation functions and up–pooling layers, are used. Such that the attention applied to extracted features before concatenation of the down–sampling features with up–sampling features uses long connections. Another U–Net–based attention network was developed by Oktay et al. [45] for Pancreas segmentation. In this paper, again a new attention gate, which is slightly different from the previous one, is applied to improve the quality of the extracted features from the down–sampling section before concatenating with up–sampling features. Also, Kaul et al. [46] proposed an attention model to improve the accuracy of skin lesion segmentation. In this model, two branches of networks, such that each branch is an encoder–decoder model was used. To improved the quality of features, an attention gate is applied on the extracted features of the decoder section of the first branch before concatenating with extracted features of the encoder section of the second branch.

Based on the literature, since medical images are highly affected by noise, and also finding an ROI is difficult in most cases, using an attention mechanism along with CNN is a useful model for medical image analysis. As shown, in all of the above papers, researchers only used an attention gate several times in a CNN to improve the quality of features. However, the effect of using some combination of the attention gates in a network was never addressed. Also, all the attention networks are manual–designed. Consequently, in most cases, significant effort is needed to find an appropriate network structure, along with an attention gate, for a specific application.

TABLE I
THE HYPER–PARAMETERS AND THEIR CORRESPONDING RANGE TO CREATE A NETWORK.

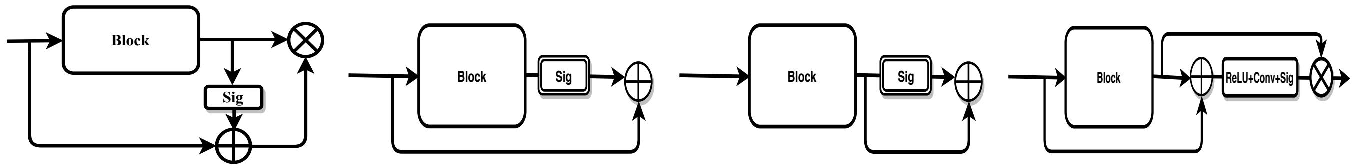
Hyper–parameters	Range
Number of blocks	7
Number of convolution layers	$[1-3]$
Filter size	$[1\times1, 3\times3, 5\times5, 7\times7]$
Number of filters	$[8, 16, 32, 64]$
Dropout	$[0-0.6]$
Pooling	[Averagepooling (0), Maxpooling (1)]
Activation function	[Sigmoid, Relu, tanh, elu]
Type of Attention	$[0-4]$
Long connection	$[0, 1]$
Batch normalisation	$[0, 1]$
Optimiser	[adam, rmsprop, adagrad, adadelta]
Learning rate	$[0.1, 0.01, 0.001, 0.0001]$
Batch size	$[4, 8, 16]$
Augmentation size	$[8000, 16000, 32000, 64000]$
Initialisation	[RandomNormal, RandomUniform, TruncatedNormal VarianceScaling, glorot-normal, glorot-uniform, he-normal, he-uniform]

III. Proposed Model

Since generating a deep attention network is a difficult task, in this section, we propose a new evolutionary model to create an attention network along with its training parameters. To create an evolutionary network, at first, we need an encoding model to create the network's genotypes, and then we need an evolutionary model to evolve the initial networks. Since our proposed model can generate U–Net–based networks, we propose a block–based encoding model, such that each genotype corresponds to down–sampling and possibly a bridging block of a network. Therefore, to create a viable solution, we need to know the maximum number of blocks and the hyper–parameters' ranges to develop networks—the list of parameters and their corresponding ranges are provided in Table I. As can be seen from Table I, the number of blocks is seven, this means the maximum number of blocks in each part of the network can be six, plus the bridging block. Because the size of the input image is 128×128, and in the U–Net–based structure after each block a pooling layer half the size of the feature maps is used; therefore, seven is the maximum number that we can use as the number of blocks. Another important parameter in Table I, is the type of attention. In the proposed model, we are using four types of attention blocks (see Figure 1). If the type of attention is zero, it means the block is not using an attention mechanism; otherwise, one of the attention models will be applied to the block.

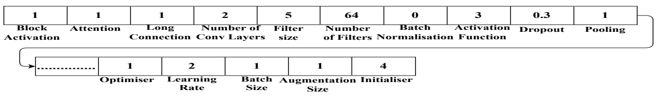
As shown in Figure 1, four types of attention methods are proposed to increase the quality of feature maps. To do this, various types of connections and operations (element–wise sum ($\bigoplus$) and element–wise multiplication ($\bigotimes$)) along with Sigmoid activation function to weight the features are applied to improve the quality of the feature maps.

In the proposed model, ten parameters are needed to create a block, plus four parameters to train a network, including optimiser, learning rate, batch size, and augmentation size. Also, the last parameter indicates the type of weight initialisation model. In order to create networks with various depths, we are using an activation flag for each block that shows if a block is active (1) or not (0). For an example, the genotype of a block and its corresponding phenotype are provided in Figure 2. Figure 2a, shows a part of a network's genotype. As can be seen from Figure 2a, the first block of the network is active; therefore, this block is included in the network's

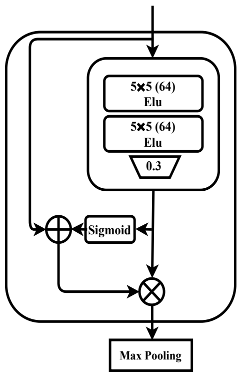


(a) Attention1 (b) Attention2 (c) Attention3 (d) Attention4

Fig. 1. Four attention models to develop attention networks.



(a) Genotype



(b) Phenotype

Fig. 2. An example of the Genotype and Phenotype of a block.

phenotype. Also, the first type of attention is selected for this block, and there is no long connection between this block and its corresponding block in the up–sampling section (long connections can be used to concatenate the extracted feature maps from a down–sampling section to an up–sampling section). Besides this, there are two convolution layers in this block with a filer size of 5×5. Also, each convolution layer has 64 output feature maps. Batch normalisation and an activation function are block–based, this means that if batch normalisation is active, after each convolution layer there is a batch normalisation, and the same activation function will be applied to all convolution layers. In the end, there is dropout with the probability of 0.3 (it may be zero in some cases). In the end, the selected pooling will be applied to the output of the block. In this way the network structure is created block by block based on the above explanation and again the up–sampling section will be a copy of the down–sampling section. However, instead of a pooling layer, a deconvolution layer will be utilised. Then the network will be trained using the training parameters.

Based on the population size, a number of networks will be generated and trained based on the training parameters up to specific epochs. The segmentation accuracy shows the fitness of each network. Then a roulette wheel section model is used to select two eligible networks. Also, to crossover selected genotypes, single point recombination is utilised. To increase the quality of the solutions, zero to three mutations can happen to each network, which is a random change to a random gene based on the specified range. This process will be continued to the stated generation. In the end, the ten best networks will be selected, and training will be continued to find the final segmentation.

TABLE II
THE LIST OF DATASETS AND THE NUMBER OF IMAGE SLICES IN TRAIN, TEST, AND VALIDATION SETS.

Dataset	Train	Test	Validation
	Volumes/Slices	Volumes/Slices	Volumes/Slices
Prostate	22/408	5/94	5/100
SLIVER07	12/2712	4/538	4/909
Spleen	29/2203	6/619	6/491

IV. Experiments

A. Dataset

For the evaluation of the proposed model, three publicly available datasets are used. They are, Prostate–Decathlon [47] for prostate MRI segmentation, SLIVER07 [48] for liver CT segmentation, and Spleen–Decathlon [47] for spleen CT segmentation. The detailed information of the datasets are provided in Table II. Image slices are resized to 128×128, and all slices were used for training. Also, because of the limited number of training images, we need to increase the number of images using augmentation models, including: rotation, zooming, vertical and horizontal flips, and elastic transformation [49]. However, the number of augmentations are chosen based on the evolution process.

B. Implementation

We implemented the proposed model using the Keras python package [50]. All experiments were carried out on one Nvidia GPU. To find the best network, we trained the proposed model using each dataset separately, based on the parameters mentioned in Table III. The initial population was 50 and then in the second generation was decreased to 30, and training continued to the stated generation. To save time during the training process each network was trained up to five epochs, and each experiment was repeated three times. For evaluating the evolved networks, Dice Similarity Coefficient (DSC) [51] is employed (see equation 1). Where, $Y^{'}$ shows the label image, Y represents predicted segmented image, and $\mid Y^{'} \mid$ and $\mid Y \mid$ indicates the cardinality of $Y^{'}$ and Y. Moreover, in the proposed model, DSC was exploited as the loss function for training the networks.

$$DSC = \frac{2 \mid Y^{'} \cap Y \mid}{\mid Y^{'} \mid + \mid Y \mid} \quad (1)$$

TABLE III
THE PARAMETERS AND THEIR CORRESPONDING VALUES FOR TRAINING THE MODELS.

Training parameters	Range
Number of generations	12
Number of epochs	5
Number of runs	3
Early stopping	3
Initial population size	50
Population size	30

C. Network's Evolution

As mentioned above, to find an appropriate network in regards to each dataset, we trained the proposed model for each dataset three times. Figure 3, shows the evolution of the ten best networks of the best run for each dataset. For the Prostate dataset, in the first generation, the ten best networks average's DSC was 0.755; however, after 12 generations, the average DSC increased to 0.865. Also, as can be seen from the second diagram, the average DSC of the ten best networks in the first generation is 0.831 for the Sliver dataset. Then with increasing the number of generations to 12, the DSC increased to 0.893. Finally, for the Spleen dataset, the average DSC in the first generation is 0.571 and at the end of evolution it had increased to 0.889. These diagrams show the efficiency of the proposed model to develop more accurate networks during evolution.

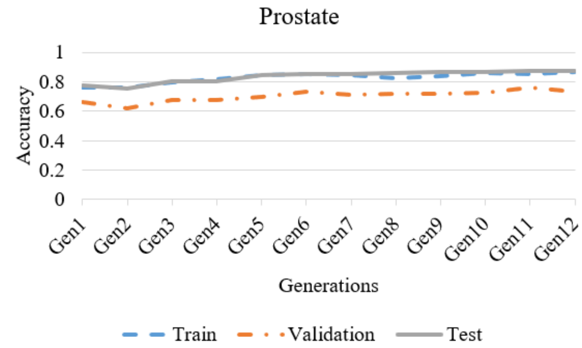
D. Training Time

As mentioned above, one of the significant issues that make developing a DNN a complicated task is required time. Therefore, in addition to the performance of the proposed model, time must also be considered. Since the proposed model starts training with a set of random networks and then continues training to find a set of more accurate networks, training time can be different in various datasets and runs. This is because, in some cases in the evolution process, smaller networks obtain better results and the training time will be decreased and vice versa. To show the required time of the proposed model, the average training time for each dataset in three runs is demonstrated in Figure 4. It needs to be noted that in the proposed model, the initial population was 50, and in the second generation the size of the population was decreased to 30 and training continued to 12 generations. Given that, as we just utilised one GPU to do the experiments; the required training time is accessible and cheap.

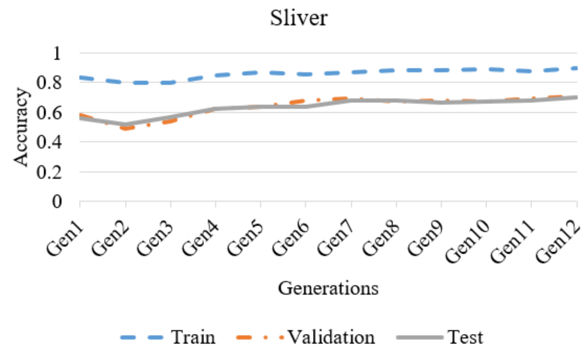
E. Experimental results

At the end of the evolution process, the ten best networks were selected as the best networks from the last generation for each dataset. Then, training continued for 45 more epochs with the early stopping of 20 to find the final segmentation. The DSC of the ten best networks in regards to each dataset can be seen in Table IV. As can be seen from Table IV, the final ten best networks are able to segment images with high accuracy.

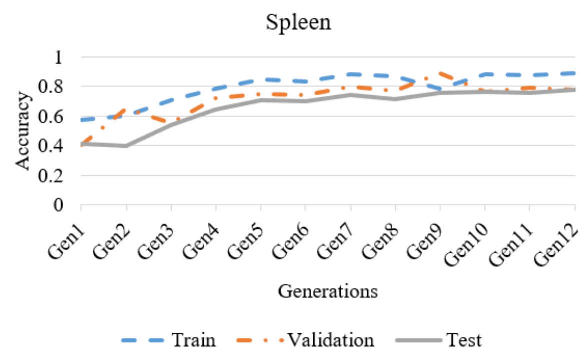
The comparison of the proposed model with previous works is provided in Table V. U–Net [17], Dense U–Net [52], and Res U–Net [53] are manually designed networks. NAS U–Net [54] is an automatic designed network using Reinforcement Learning (RL) such that the network structure is kept fixed, and RL is applied to find the block's topology. Also, we compared the proposed networks versus AdaResU–Net [32]



(a) Prostate Dataset



(b) SLIVER07 dataset



(c) Spleen dataset

Fig. 3. The evolution of the networks in each dataset.

TABLE IV
THE DSC OF THE TEN BEST NETWORKS IN REGARDS TO EACH DATASET AFTER 50 EPOCHS.

Networks	Prostate	SLIVER07	Spleen
Network 1	**0.90**	**0.952**	**0.948**
Network 2	0.889	0.951	0.936
Network 3	0.889	0.948	0.933
Network 4	0.887	0.948	0.933
Network 5	0.887	0.947	0.93
Network 6	0.886	0.926	0.927
Network 7	0.879	0.92	0.921
Network 8	0.876	0.914	0.918
Network 9	0.871	0.894	0.918
Network 10	0.868	0.891	0.905

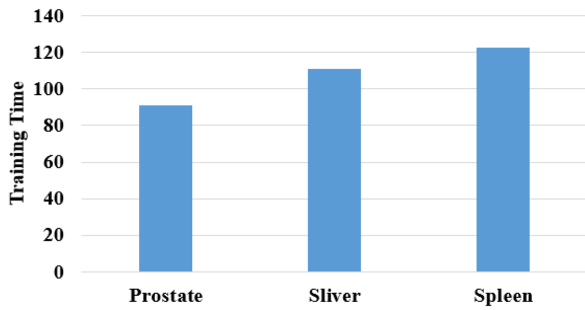


Fig. 4. proposed model training time to find the best networks for each dataset.

TABLE V
COMPARISON OF THE OBTAINED DSC OF THE PROPOSED MODEL VERSUS PREVIOUS WORK.

Models	Prostate	SLIVER07	Spleen	Trainable Parameters
U–Net [17]	0.73	0.70	0.74	31.0×10^6
Dense U–Net [52]	0.837	0.85	0.82	15.48×10^6
Res U–Net [53]	0.82	0.84	0.78	4.04×10^6
NAS U–Net [54]	0.821	0.93	0.89	30.0×10^6
AdaResU–Net [32]	0.846	0.893	0.90	4.1×10^6, 4.9×10^6, 3.0×10^6
EvoU–Net [33]	0.854	0.90	0.90	1.6×10^6, 2.1×10^6, 1.9×10^6
Proposed model	**0.90**	**0.952**	**0.948**	1.4×10^6, 1.1×10^6, 3.9×10^5

and EvoU–Net [33] that are evolutionary neural networks. In the AdaResU–Net [32], the network structure was kept fixed again, and an evolutionary technique was applied to find a part of the parameters. Finally, EvoU–Net [33] has also been used to generate a network for medical image segmentation.

As can be seen from Table V, the proposed model outperformed the previous works for the segmentation of three different medical datasets. Compared to manually designed networks, our proposed model is able to find a network for each dataset separately, that can increase the chance of finding a better network. Also, compared to automatically designed networks, in the proposed model, we utilised the attention mechanism that can help the network to focus more on ROI.

Besides, we compared the developed evolutionary attention networks with previous works regarding the size of the network (see Table V). As shown, the established networks use fewer parameters that can decrease the training time and required computation, while obtaining better segmentation accuracy.

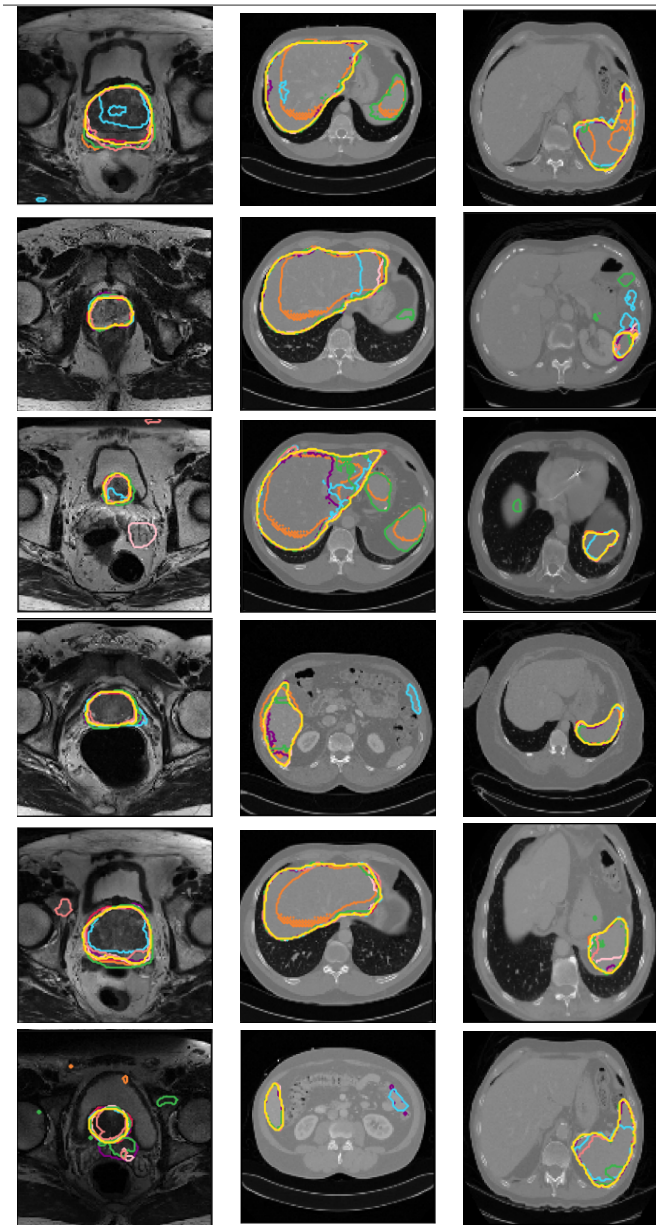
Also, the obtained results have been compared to previous works subjectively (see Table VI). As shown in Table VI, the proposed model (yellow border), segments images precisely. However, in most of the cases, the previous works were not precise enough for segmentation. These results show the effect of designing a network regarding a particular dataset, also the efficacy of using an attention mechanism for segmentation.

F. The Best Network

In this section, the genotype (see Table VII) and corresponding phenotype (see Figure 5) of the best network that was obtained for the Prostate dataset is provided as an example. As shown in Table VII, the network has four active blocks in

TABLE VI
SIX SAMPLE SEGMENTED IMAGES FROM EACH DATASET. THE RED CONTOUR IS THE GROUND TRUTH, ORANGE IS U–NET, PURPLE IS ADARES U–NET, PINK IS EVOU–NET, CYAN IS RES U–NET, GREEN IS DENSE U–NET, LIGHT CORAL IS NAS U–NET, AND YELLOW IS THE PROPOSED ATTENTION MODEL.

Prostate	SLIVER07	Spleen



the down–sampling section that is copied to the up–sampling section, also a bridging block. The training parameters were also found using the proposed evolutionary attention model. Besides, Figure 5 shows the phenotype of the best network. As shown the down–sampling blocks and a bridging block is constructed based on the obtained genotype. The down–sampling blocks need to be copied to the up–sampling section instead of the dashed lines. Also, as shown in Figure 5, the pooling layers are replaced with deconvolution layers to increase the size of the feature maps. Besides, two long

TABLE VII
THE GENOTYPE OF THE BEST NETWORK STRUCTURE CORRESPONDING TO THE PROSTATE DATASET. THE ABBREVIATIONS IN THE NETWORK'S GENOTYPE ARE: BA (BLOCK ACTIVATION), AB (ATTENTION BLOCK), LC (LONG CONNECTION), NOC (NUMBER OF CONVOLUTION), CS (CONVOLUTION SIZE), NOF (NUMBER OF FILTERS), BN (BATCH NORMALISATION), AF (ACTIVATION FUNCTION), D (DROPOUT), AND P (POOLING).

Block	BA	AB	LC	NOC	CS	NOF	BN	AF	D	P	Training Parameters
Block 1	1	4	0	2	5	16	1	2	0.4	0	
Block 2	0	0	1	1	5	8	1	1	0.3	1	Optimiser = Adam
Block 3	1	2	1	1	5	64	0	2	0.4	1	Learning Rate = 0.0001
Block 4	1	0	1	3	5	16	1	1	0.0	1	Batch Size = 32
Block 5	0	0	1	1	7	8	0	2	0.0	0	Augmentation = 16000
Block 6	1	2	0	3	7	64	0	3	0.0	0	Initialisation = Random Uniform
Bridge	1	3	0	2	3	16	1	3	0.1	0	

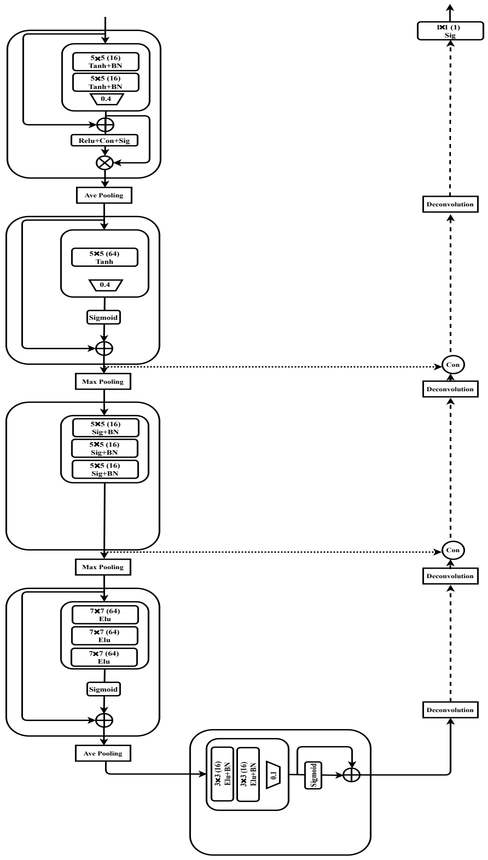


Fig. 5. The Phenotype of the best network for the Prostate dataset.

connections exist to copy and concatenate the feature maps from the down–sampling section to the up–sampling section.

V. DISCUSSION AND CONCLUSION

In this paper, we proposed an evolutionary attention model to create U–Net–based deep convolutional neural networks for medical image segmentation. To do this, we proposed a new block–based encoding model to decode the genotypes and convert them to their phenotypes. The main contribution of this work is utilising an attention mechanism, which has never been previously considered in the literature to create evolutionary networks. Firstly, we proposed four different attention mechanisms based on the literature. Then using the genetic algorithm, the best network structure potentially equipped with an attention mechanism and training parameters was found for each dataset. The experimental results, using three different datasets, show that the proposed evolutionary model is able to find a set of high–performance networks for each dataset. Also, our experiments show that not only that our proposed model can find high–performance networks, but also does not need huge computation to find the networks.

One on the problems of developing a neural network manually is lack of clear rules to find an appropriate network structure quickly, therefore the trial and error and using rules of thumb are the only available methods. However, in our proposed model, we include almost all parameters in the search space; therefore, the end–user can, without in–depth knowledge, develop a network for image segmentation. Also, because each network is built regarding each particular dataset, the accuracy of the found networks is high.

However, we need to do more experiments to improve the proposed model. The first plan is developing more attention mechanisms, instead of just using four attention models. The second plan is using more datasets to evaluate the proposed model. Also, in the future, we intend to consider this model for 3D image segmentation. Finally, changing the proposed encoding model to develop networks for other applications, such as image classification is another future goal.

ACKNOWLEDGEMENT

This research was undertaken with the assistance of resources and services from the National Computational Infrastructure (NCI), which is supported by the Australian Government.

REFERENCES

[1] M. Valueva, N. Nagornov, P. Lyakhov, G. Valuev, and N. Chervyakov, "Application of the residue number system to reduce hardware costs of the convolutional neural network implementation," *Mathematics and Computers in Simulation*, 2020.

[2] Y. Li, Z. Hao, and H. Lei, "Survey of convolutional neural network," *Journal of Computer Applications*, vol. 36, no. 9, pp. 2508–2515, 2016.

[3] Z. Qin, F. Yu, C. Liu, and X. Chen, "How convolutional neural network see the world-a survey of convolutional neural network visualization methods," *arXiv preprint arXiv:1804.11191*, 2018.

[4] F. Altenberger and C. Lenz, "A non-technical survey on deep convolutional neural network architectures," *arXiv preprint arXiv:1803.02129*, 2018.

[5] W. Liu, Z. Wang, X. Liu, N. Zeng, Y. Liu, and F. E. Alsaadi, "A survey of deep neural network architectures and their applications," *Neurocomputing*, vol. 234, pp. 11–26, 2017.

[6] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. Van Der Laak, B. Van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Medical image analysis*, vol. 42, pp. 60–88, 2017.

[7] G. Haskins, U. Kruger, and P. Yan, "Deep learning in medical image registration: A survey," *Machine Vision and Applications*, vol. 31, no. 1, p. 8, 2020.

[8] E. Miranda, M. Aryuni, and E. Irwansyah, "A survey of medical image classification techniques," in *2016 International Conference on Information Management and Technology (ICIMTech)*. IEEE, 2016, pp. 56–61.

[9] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700–4708.

[10] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[11] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," *arXiv preprint arXiv:1409.0473*, 2014.
[12] S. Chaudhari, G. Polatkan, R. Ramanath, and V. Mithal, "An attentive survey of attention models," *arXiv preprint arXiv:1904.02874*, 2019.
[13] F. Wang and D. M. Tax, "Survey on the attention based rnn model and its applications in computer vision," *arXiv preprint arXiv:1601.06823*, 2016.
[14] Y. Li, Z. Zhu, D. Kong, H. Han, and Y. Zhao, "Ea-lstm: Evolutionary attention-based lstm for time series prediction," *Knowledge-Based Systems*, vol. 181, p. 104785, 2019.
[15] R. A. Vázquez, H. Sossa, and B. A. Garro, "An evolutionary feature-based visual attention model applied to face recognition," in *International Conference on Hybrid Artificial Intelligence Systems*. Springer, 2010, pp. 376–384.
[16] J. H. Holland *et al.*, *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence*. MIT press, 1992.
[17] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.
[18] T. Back, *Evolutionary algorithms in theory and practice: evolution strategies, evolutionary programming, genetic algorithms*. Oxford university press, 1996.
[19] R. S. Sutton, A. G. Barto *et al.*, *Introduction to reinforcement learning*. MIT press Cambridge, 1998, vol. 2, no. 4.
[20] D. J. Montana and L. Davis, "Training feedforward neural networks using genetic algorithms." in *IJCAI*, vol. 89, 1989, pp. 762–767.
[21] S. Ding, H. Li, C. Su, J. Yu, and F. Jin, "Evolutionary artificial neural networks: a review," *Artificial Intelligence Review*, vol. 39, no. 3, pp. 251–260, 2013.
[22] A. Baldominos, Y. Saez, and P. Isasi, "On the automated, evolutionary design of neural networks: past, present, and future," *Neural Computing and Applications*, pp. 1–27.
[23] Q.-H. Ling, Y.-Q. Song, F. Han, and H. Lu, "An improved evolutionary random neural networks based on particle swarm optimization and input-to-output sensitivity," in *International Conference on Intelligent Computing*. Springer, 2017, pp. 121–127.
[24] T. Hassanzadeh, K. Faez, and G. Seyfi, "A speech recognition system based on structure equivalent fuzzy neural network trained by firefly algorithm," in *2012 International Conference on Biomedical Engineering (ICoBE)*. IEEE, 2012, pp. 63–67.
[25] V. K. Ojha, A. Abraham, and V. Snášel, "Metaheuristic design of feedforward neural networks: A review of two decades of research," *Engineering Applications of Artificial Intelligence*, vol. 60, pp. 97–116, 2017.
[26] K. O. Stanley, J. Clune, J. Lehman, and R. Miikkulainen, "Designing neural networks through neuroevolution," *Nature Machine Intelligence*, vol. 1, no. 1, pp. 24–35, 2019.
[27] A. Martín, R. Lara-Cabrera, F. Fuentes-Hurtado, V. Naranjo, and D. Camacho, "Evodeep: a new evolutionary approach for automatic deep neural networks parametrisation," *Journal of Parallel and Distributed Computing*, vol. 117, pp. 180–191, 2018.
[28] B. Wang, Y. Sun, B. Xue, and M. Zhang, "A hybrid differential evolution approach to designing deep convolutional neural networks for image classification," in *Australasian Joint Conference on Artificial Intelligence*. Springer, 2018, pp. 237–250.
[29] Y. Sun, B. Xue, M. Zhang, and G. G. Yen, "Automatically evolving cnn architectures based on blocks," *arXiv preprint arXiv:1810.11875*, 2018.
[30] J. H. Holland, "Adaptation in natural and artificial systems: An introductory analysis with applications to biology, control, and artificial intelligence." 1975.
[31] R. S. Kumari, T. Varghese, C. Kesavadas, N. A. Singh, and P. Mathuranath, "A genetic algorithm optimized artificial neural network for the segmentation of mr images in frontotemporal dementia," in *International Conference on Swarm, Evolutionary, and Memetic Computing*. Springer, 2013, pp. 268–276.
[32] M. Baldeon-Calisto and S. K. Lai-Yuen, "Adaresu-net: Multiobjective adaptive convolutional neural network for medical image segmentation," *Neurocomputing*, 2019.
[33] T. Hassanzadeh, D. Essam, and R. Sarker, "Evou-net: an evolutionary deep fully convolutional neural network for medical image segmentation," in *Proceedings of the 35th Annual ACM Symposium on Applied Computing*, 2020, pp. 181–189.
[34] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in neural information processing systems*, 2017, pp. 5998–6008.
[35] Z. Yang, D. Yang, C. Dyer, X. He, A. Smola, and E. Hovy, "Hierarchical attention networks for document classification," in *Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: human language technologies*, 2016, pp. 1480–1489.
[36] C. Liu, X. Chen, E. C. Shin, M. Chen, and D. Song, "Latent attention for if-then program synthesis," in *Advances in Neural Information Processing Systems*, 2016, pp. 4574–4582.
[37] J. Zhang, Y. Xie, Y. Xia, and C. Shen, "Attention residual learning for skin lesion classification," *IEEE transactions on medical imaging*, vol. 38, no. 9, pp. 2092–2103, 2019.
[38] G.-S. Xie, L. Liu, X. Jin, F. Zhu, Z. Zhang, J. Qin, Y. Yao, and L. Shao, "Attentive region embedding network for zero-shot learning," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 9384–9393.
[39] B. Chen and W. Deng, "Hybrid-attention based decoupled metric learning for zero-shot image retrieval," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 2750–2759.
[40] Y. Yan, J. Kawahara, and G. Hamarneh, "Melanoma recognition via visual attention," in *International Conference on Information Processing in Medical Imaging*. Springer, 2019, pp. 793–804.
[41] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.
[42] Y. Wang, Z. Deng, X. Hu, L. Zhu, X. Yang, X. Xu, P.-A. Heng, and D. Ni, "Deep attentional features for prostate segmentation in ultrasound," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2018, pp. 523–530.
[43] Y. Wang, H. Dou, X. Hu, L. Zhu, X. Yang, M. Xu, J. Qin, P.-A. Heng, T. Wang, and D. Ni, "Deep attentive features for prostate segmentation in 3d transrectal ultrasound," *IEEE transactions on medical imaging*, vol. 38, no. 12, pp. 2768–2778, 2019.
[44] N. Abraham and N. M. Khan, "A novel focal tversky loss function with improved attention u-net for lesion segmentation," in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*. IEEE, 2019, pp. 683–687.
[45] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz *et al.*, "Attention u-net: Learning where to look for the pancreas," *arXiv preprint arXiv:1804.03999*, 2018.
[46] C. Kaul, S. Manandhar, and N. Pears, "Focusnet: an attention-based fully convolutional network for medical image segmentation," in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*. IEEE, 2019, pp. 455–458.
[47] Medical segmentation decathlon. https://decathlon-10.grand-challenge.org/.
[48] T. Heimann, B. Van Ginneken, M. A. Styner, Y. Arzhaeva, V. Aurich, C. Bauer, A. Beck, C. Becker, R. Beichel, G. Bekes *et al.*, "Comparison and evaluation of methods for liver segmentation from ct datasets," *IEEE transactions on medical imaging*, vol. 28, no. 8, pp. 1251–1265, 2009.
[49] P. Y. Simard, D. Steinkraus, J. C. Platt *et al.*, "Best practices for convolutional neural networks applied to visual document analysis." in *Icdar*, vol. 3, no. 2003, 2003.
[50] F. Chollet *et al.*, "Keras," 2015.
[51] L. R. Dice, "Measures of the amount of ecologic association between species," *Ecology*, vol. 26, no. 3, pp. 297–302, 1945.
[52] M. Kolařík, R. Burget, V. Uher, K. Říha, and M. K. Dutta, "Optimized high resolution 3d dense-u-net network for brain and spine segmentation," *Applied Sciences*, vol. 9, no. 3, p. 404, 2019.
[53] Z. Zhang, Q. Liu, and Y. Wang, "Road extraction by deep residual u-net," *IEEE Geoscience and Remote Sensing Letters*, vol. 15, no. 5, pp. 749–753, 2018.
[54] Y. Weng, T. Zhou, Y. Li, and X. Qiu, "Nas-unet: Neural architecture search for medical image segmentation," *IEEE Access*, vol. 7, pp. 44 247–44 257, 2019.