Deep Neural Architectures for Medical Image Semantic Segmentation: Review

MUHAMMAD ZUBAIR KHAN[1], **(Graduate Student Member, IEEE),**
MOHAN KUMAR GAJENDRAN[1], **(Graduate Student Member, IEEE),**
YUGYUNG LEE[1], **(Senior Member, IEEE),**
AND MUAZZAM A. KHAN[2], **(Senior Member, IEEE)**

[1]School of Computing and Engineering, University of Missouri—Kansas City, Kansas City, MO 64110, USA
[2]Department of Computer Science, Quaid-i-Azam University, Islamabad 44000, Pakistan

Corresponding authors: Muhammad Zubair Khan (mkzb3@mail.umkc.edu) and Muazzam A. Khan (muazzam.khattak@qau.edu.pk)

ABSTRACT Deep learning has an enormous impact on medical image analysis. Many computer-aided diagnostic systems equipped with deep networks are rapidly reducing human intervention in healthcare. Among several applications, medical image semantic segmentation is one of the core areas of active research to delineate the anatomical structures and other regions of interest. It has a significant contribution to healthcare and provides guided interventions, radiotherapy, and improved radiological diagnostics. The underlying article provides a brief overview of deep convolutional neural architecture, the platforms and applications of deep neural networks, metrics used for empirical evaluation, state-of-the-art semantic segmentation architectures based on a foundational convolution concept, and a review of publicly available medical image datasets highlighting four distinct regions of interest. The article also analyzes the existing work and provides open-ended potential research directions in deep medical image semantic segmentation.

INDEX TERMS Deep learning, convolution neural network, medical image analysis, semantic segmentation, skip-connections, encoder-decoder, computer-aided diagnostics, healthcare.

I. INTRODUCTION

The convolutional neural techniques have an enormous impact on various areas of medical science. The underlying study is performed to overview the impact of deep learning techniques on medical image segmentation. It is an important step towards image-guided interventions, radiotherapy, and improved radiological diagnostics. A wide range of deep network architectures for medical image segmentation has been proposed for various modalities. Architectural compression and network optimization are the primary focus areas of AI researchers to build small yet precise models. The applications of deep learning to medical image analysis grew rapidly since 2015 [1]–[8]. The topic becomes dominant at major conferences and well-reputed journals. Initially, Fully Convolutional Network (FCN) [9], and U-Net [4] architectures have revolutionized medical image segmentation and provided a pathway for later introduced inspired models. The structure

The associate editor coordinating the review of this manuscript and approving it for publication was Chao Tong.

of the neural network contributes to the success of deep learning over conventional machine learning models. It follows an incremental feature learning approach that eliminates the need for hard-coded features and domain expertise [10]–[15]. Besides, publically available datasets and model optimization techniques privileged researchers to validate new variants of segmentation models and reach optimal performance. Additionally, the appearance of graphical processing units (GPUs) on-premises and over the cloud made the training process 30 times faster than conventional processing units (CPUs) with open-source GPU-compatible software libraries and packages [16]–[18]. Deep learning's success motivated scientists and researchers to apply these algorithms in medical image segmentation for diagnosis and treatment. It is observed that these deep network based-methods have achieved superior performance in segmenting region-of-interest than conventional methods when applied for brain, lung, pancreas, and retinal pathologies [19], [20].

The prominent deep learning applications for medical image analysis includes image construction [21],

enhancement [22], [23], segmentation [24], registration [25], computer-aided detection [26], view detection [27], landmark detection [28], and automated report generation [29]. A novel recurrent generative adversarial model is produced in [30] to reduce the pixel imbalance problem from medical images. The network includes bi-directional Long short-term memory (LSTM) to obtain inter-slice and intra-slice representation of the features. The authors presented a collection of techniques used in multi-modal image fusion for medical image segmentation. The method in [31] developed to support coarse localization with Retina U-Net. The technique combines RetinaNet's one-stage detector with U-Net architecture for image segmentation. It restores lossy signals with object detection without additional complexity found in a two-stage variant. An asymmetric U-Net-based convolutional block is used in [32] to define multi-scale architecture with skip-connections. It fuses low and high-end feature maps with different scales and strengthens the representational capacity of convolutional blocks. In [33], a self-supervised framework is designed to lessen the requirement for annotated samples. The network includes an adaptive prototyping module to resolve the class-imbalance problem. The study conducted in [34] designed a network for joint object detection. It encompasses the multi-vision task and a novel loss function using a transfer learning approach.

In article [35], a review of deep semantic segmentation is provided. The work is divided into three segments, where each core uncovers certain aspects of semantic segmentation. It highlights fully convolutional network (FCN)-based, region-based, and weakly supervised segmentation methods mainly. The authors of [36] presented a survey on deep learning models, commonly used datasets, and prominent evaluation metrics for natural image deep semantic segmentation. Similarly, the authors of [37] provided a ceiling review of medical image analysis and its application using deep learning. Hesamian *et al.* [38] also presented an overview of the latest deep network method and model training techniques for medical image segmentation. Karimi *et al.* [39] reviewed the techniques used for handling label noise and benchmarked existing segmentation and classification for medical image analysis. The authors in [40] discussed deep architectures with weakly supervised, fully supervised, and transfer learning techniques. It also uncovers the related data-scarcity and class-imbalance problems. The authors reviewed the techniques to target small data sample problems common in medical imaging. It also covered the literature related to transfer learning and active transfer learning methods. Most of the surveys conducted were categorized into multiple sub-groups based on their emphases like [41]–[44]. This mainly featured deep convolutional networks or extensions like recurrent neural networks, generative adversarial networks for image restoration, classification, segmentation, compression, and registration. Similarly, [11], [12], [14], [18], [45] focused on specialized areas such as domain adaptation, brain tissue, COVID-19, cardiac, and digital pathology. In contrast, other reviews discussed deep segmentation architectures for

specific medical application. The surveys [42], [46], [47] highlighted major deep networks and training strategies for medical image analysis.

The authors in [47] have provided a comprehensive review of deep learning-based image segmentation architectures used for general computer vision tasks. This survey includes medical image segmentation architectures; however, it primarily focused on object detection and segmentation of general image datasets. The authors in [48] have focused on medical image segmentation with limited supervision. This review is limited to efficient models capable of dealing with partially annotated datasets. The authors in [42], and [49] have reviewed most of the medical image segmentation architectures. They have provided the state-of-the-art performance of various medical modalities; nonetheless, they have not compared various architectures' performance and limitations for a given dataset. One dedicated review in line with our focused objectives is presented in [41] however, this article was published in 2017 and therefore does not include the latest literature. Our survey presents an in-depth, comprehensive review of different aspects, including the benchmark dataset, semantic architectures designed explicitly for medical image segmentation, an updated survey of the latest designed techniques, improvement mechanisms developed over time, evaluation metrics, challenges, and potential recommendations to fix highlighted challenges.

A. STUDY METHOD

A systematic review is conducted through an academic search engine, google scholar. The time frame considered for article selection mainly ranges between 2015 to 2021, except for papers reflecting core domain concepts with high citations. The keywords used for article selection include; deep neural architectures, semantic segmentation, medical imaging, image analysis, deep learning applications, computer-aided diagnosis, health AI, multi-modal medical system, and benchmark medical datasets. A total of 115 relevant articles were thoroughly studied and selected for this review. The literature is reviewed in a confined mode to keep the focus on deep learning applications in medicine. The database used to query relevant contributions includes PubMed, MedRxiv, and ArXiv. We also searched through top-tier conference portals such as MICCAI, ISBI, SPIE, ICHI, and EMBC. The articles that do not address medical image analysis and/or discuss conventional neural applications with handcrafted and hard-coded features are eliminated from the review process. All architectures are thoughtfully selected and placed concerning their contribution of building new methods, enhancing existing ones, improving accuracy, and reducing computational complexity. All the data collected for this article were taken from the original papers that contain state-of-the-art work performed in medical image semantic segmentation. The review process observed distinct aspects of proposed methods, including the choice of the network model, modality and data examined, the strategy adopted for training and inference, and vital contributions. It will help
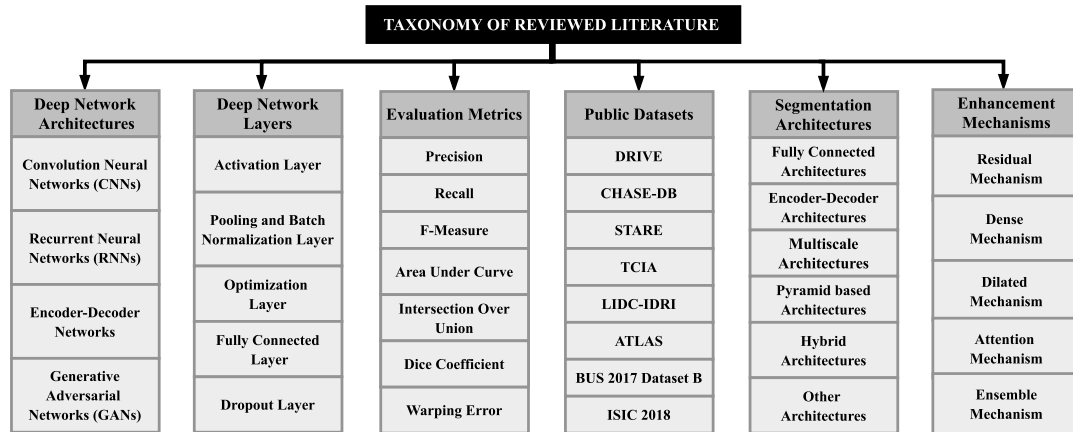
TAXONOMY OF REVIEWED LITERATURE					
Deep Network Architectures	**Deep Network Layers**	**Evaluation Metrics**	**Public Datasets**	**Segmentation Architectures**	**Enhancement Mechanisms**
Convolution Neural Networks (CNNs)	Activation Layer	Precision	DRIVE	Fully Connected Architectures	Residual Mechanism
Recurrent Neural Networks (RNNs)	Pooling and Batch Normalization Layer	Recall	CHASE-DB	Encoder-Decoder Architectures	Dense Mechanism
		F-Measure	STARE		
Encoder-Decoder Networks	Optimization Layer	Area Under Curve	TCIA	Multiscale Architectures	Dilated Mechanism
	Fully Connected Layer	Intersection Over Union	LIDC-IDRI	Pyramid based Architectures	Attention Mechanism
Generative Adversarial Networks (GANs)	Dropout Layer	Dice Coefficient	ATLAS	Hybrid Architectures	
		Warping Error	BUS 2017 Dataset B	Other Architectures	Ensemble Mechanism
			ISIC 2018		

FIGURE 1. Hierarchically-structured taxonomy of underlying survey.

the researchers to refine their baseline concepts for medical image analysis. The taxonomy of this literature is shown in the Fig. 1, and the reviewed architectures are organized into the following:

1) Fully Convolutional Neural Architectures
2) Encoder-Decoder Based Network
3) Multi-Scale Neural Architectures
4) Pyramid-based Network Architectures
5) Hybrid Architectures
6) Other Architectures
7) Enhancement Mechanisms

The main contributions of this article are provided below:

- The article highlights the deep convolutional neural network, its layered structure with applications, and standard performance evaluation metrics.
- The detail of benchmark datasets commonly used for model evaluation and a comprehensive review of deep semantic segmentation architectures for medical imaging is provided.
- The review emphasizes how different segmentation techniques have achieved state-of-the-art performance by analyzing model structural schema and assessing the behavior of the different medical datasets.
- The set of challenges and potential research directions with possible hypothetical solutions are provided for medical image analysis using deep learning methods.

II. OVERVIEW

This section highlights an overview and operational details of convolution neural networks. It encompasses the evaluation metrics used to assess the performance of a neural model.

A. DEEP NEURAL NETWORK ARCHITECTURES
1) CONVOLUTION NEURAL NETWORKS (CNNs)

A convolutional neural network (CNN) is an advanced neural network architecture developed for analyzing two-dimensional images [50]. Nevertheless, it can be applied to one-dimensional and three-dimensional data.

The convolutional layer is the fundamental block of CNN. Convolution is a linear operation performed by multiplying weights with the input, similar to a traditional neural network. Since this technique was primarily intended for two-dimensional input (images), this linear operation is performed between the input data and a two-dimensional array of weights, called a mask or a filter. This filter size is smaller than the input data. An element-wise multiplication (dot product) is performed between the filter and filter-sized patch of the input resulting in a single value as shown in Fig. 2. This operation is repeated systematically across the entire input image resulting in a two-dimensional output array called a feature map. Systematic application of the identical filter across an entire image allows the algorithm to discover a feature hidden anywhere in the image. This important ability is known as translation invariance. The major advantage of CNN is that the filters need not be handcrafted; instead, filters can be determined automatically by training with back-propagation, where the backward pass uses convolution operation but with spatially flipped filters [41], [45], [51]–[53]. Stacking the convolutional layers in CNN is a very effective approach.

The layers near the input learn low-level features like lines and edges, and layers deeper in the model learn higher-order features like shapes. CNN is a regularized variant of multilayer perceptrons. In multilayer perceptrons (FCN), each neuron of one layer is linked to all neurons in the following layers. The fully connectedness of these networks makes them susceptible to overfitting, which can be avoided through regularization. Conventional modes of regularization include adding some weights to the loss function. However, CNN uses a different regularization strategy; they utilize the hierarchical pattern in data and assemble more complex patterns using smaller and simpler patterns, reducing the connectedness and complexity [49], [54]–[56]. CNN learns the filters which were hand-engineered in traditional algorithms. This self-sufficiency from prior knowledge in feature engineering is its significant advantage.
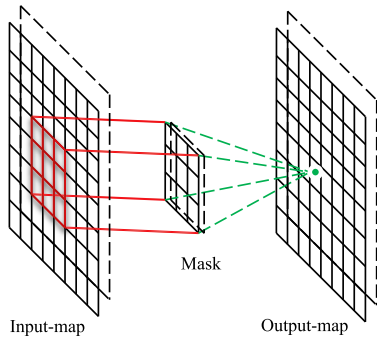
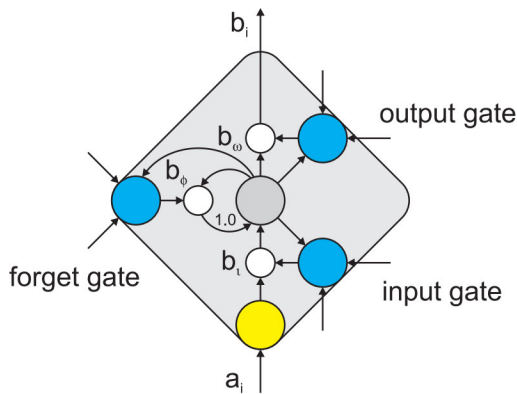**FIGURE 2. Convolution operation with *m* × *n* mask.**

FIGURE 3. LSTM memory cell with gating units [60].

2) RECURRENT NEURAL NETWORKS (RNNs)

Recurrent neural networks permit previous outputs to be used as inputs while having hidden states. RNNs are widely used in natural language processing, speech recognition, machine translation, music generation, and sentiment classification. Major advantages of RNN include their ability to process inputs of any length and compact model size despite larger inputs; computation accounts historical information, weights are shared across time. However, it suffers from higher computational costs, difficulty accessing information from the very past, and gradient vanishing or exploding problems. Modern variants of RNN such as Gated Recurrent Unit (GRU) [57], and LSTM [58] were designed to address the vanishing gradient problem faced by conventional RNNs. Bidirectional RNN (BRNN) [59], and Deep RNN (DRNN) are other advanced varients of RNN. An LSTM memory cell, as shown in Fig. 3 contains an input gate, output gate, and forget gate to regulate information flow.

3) ENCODER-DECODER NETWORK MODELS

Encoder-decoders [61], [62] are a class of models with a two-stage network as shown in Fig. 4. These models are widely used in image segmentation, image enhancement alongside machine translation, and Natural Language Processing (NLP) due to their capability of handling variable-length input and output sequences. These architectures have two major components encoders and decoders.
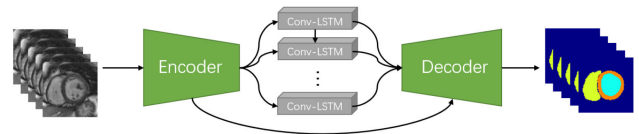

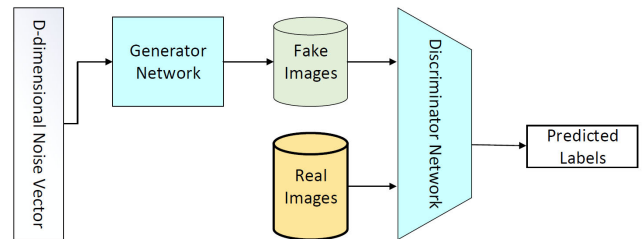**FIGURE 4. Encoder-Decoder Model [68].**

FIGURE 5. Generative Adversarial Network Model [47].

Typically, the encoder takes an input sequence and transforms it into a fixed-shaped state (hidden representation), and the decoder maps the encoded state to an output sequence. Auto-encoders are a specific case of encoder-decoder models in which both input and output are identical [44], [63]–[67].

4) GENERATIVE ADVERSARIAL NETWORKS (GANs)

The generative adversarial network (GAN) are recent techniques employed in unsupervised, and semisupervised learning [24], [39], [47], [69], [70]. GANs were inspired by noise-contrastive estimation, and they utilize implicit modeling of high-dimensional distributions in the data. GANs consist of two networks as shown in Fig. 5; a generative network generates plausible new data by learning the database's patterns, and a discriminative network classifies samples as either original data from domain or generated data. These two neural networks contest against each other where one contestant's loss is the other contestant's gain. GAN's fundamental idea is that the generator is inspired to generate a distribution of data to match that of real data, and an indirect training of generative networks is performed through a dynamically updated discriminative network. Early generative adversarial networks (GANs) take noise as input; however, recent GANs take images instead of noise. Various GAN architecture has been developed, such as Fully connected GANs, Convolutional GANs [71], Conditional GANs [72], Wasserstein-GAN [73], GANs with inference models, and Adversarial autoencoders.

B. DEEP NETWORK LAYERS
1) ACTIVATION LAYER

An activation layer decides what weights are to be fired to the layer such that the network learns the complex pattern in the data. In CNN, a convolution layer is typically followed by an activation layer containing activation functions. The feature maps from the convolution layer are operated using these functions to determine whether a neurons' input should
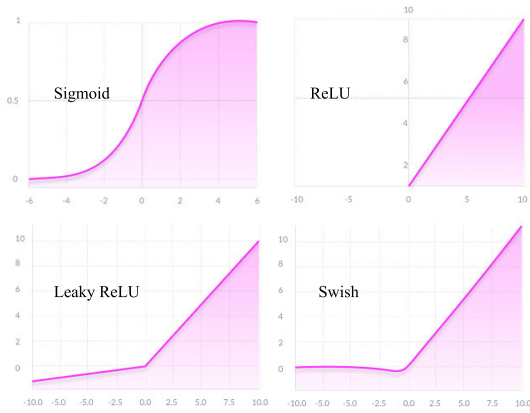
FIGURE 6. Plot of activation functions.

FIGURE 7. Maxpooling operation with mask and stride = 2.

be activated. It is based on whether each neuron input is important for the models' prediction. These functions have to be computationally efficient since they are computed across millions of neurons [53], [74]–[76]. Especially in CNN, backpropagation is employed, requiring evaluation of these functions' derivatives; This mandates these functions to be non-linear, further increasing the computational strain. Commonly used activation functions are plotted in Fig. 6.

The sigmoid activation function ranges from 0 to 1. Therefore, it is suitable for models that predict the probability as an output [77]. The softmax function is a more generalized sigmoid function, widely used for multi-class classification [78]–[81]. However, these functions are computationally expensive and suffer from vanishing gradient problem for very high or low input values, making the network very less sensitive to inputs in this range. Rectified Linear Units (ReLU) overcomes this problem [82]–[86]. It is also a widely accepted activation function ranging from 0 to ∞. This activation function is very easy to compute and is free from the vanishing gradient problem. However, it suffers from the dying-ReLU problem. It gives 0 as output for all negative inputs, thereby preventing backpropagation for negative inputs. Leaky ReLU is an extension of ReLU that solves the dying ReLU problem for negative inputs, but it suffers from inconsistency in negative input predictions. It should be noted that leaky ReLU and ReLU perform identically for positive inputs; therefore, the exploding problem still exists for higher input values as both functions are not bounded. Swish, a self-gated activation function, is another extension of ReLU recently discovered by Google researchers. It is identical in computational efficiency to ReLU and can outperform it by up to 0.9% on the ImageNet database [87]. Leaky ReLU does solve the dying ReLU problem for negative inputs, but for positive inputs, it performs the same as ReLU, which means the exploding problem cannot be solved.

2) POOLING AND BATCH NORMALIZATION
Pooling is a simple mathematical operation where a small grid region of the feature map is reduced to a single value,
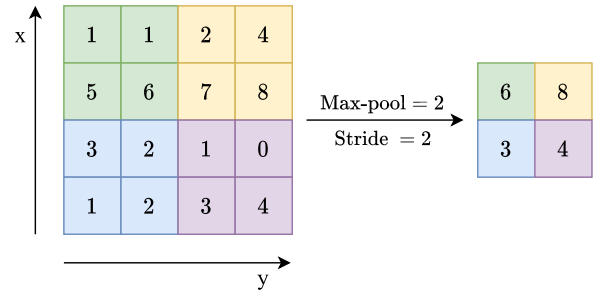
thereby producing a downscaled feature map. Numerous pooling techniques have been developed such as max pooling, average pooling, stochastic pooling [88], spatial pyramid pooling [89], multiscale orderless pooling [90], and spectral pooling [91]. Max-pooling is the most commonly employed technique that returns the maximum of the grid value, as shown in Fig. 7. Pooling operations on a feature map results create a shift in the activation map alongside downsampling of feature maps. Therefore pooling layers add some translational invariance to the network. It is worth noticing that downsampling can also be performed without pooling layers. One such technique is increasing the stride length resulting in a simpler network architecture without necessarily sacrificing the performance [92]–[95].

Similarly, the batch-normalization is another mathematical operation performed on a feature map resulting from pooling layers. It is performed by subtracting the mean and dividing it by the standard deviation for each training batch. This process forces the network to change its activations to zero mean and unit standard deviation periodically. Therefore, this process acts as a regularizer for the network and significantly enhances the stability, pace of the training process and eliminates a need for careful initialization of parameters [96].

3) OPTIMIZERS
The optimizers are the algorithms or methods used to change the attributes of a neural network, such as weights and learning rate to reduce the losses [97]–[99]. A generic form of an optimizer is depicted in (1).

$$\theta_{t+1} = \theta_t + \Delta\theta_t \tag{1}$$

The gradient descent is the most fundamental optimization algorithm as shown in (2) where, θ is the weight parameter, η is the learning rate and $\nabla J(\theta; x, y)$ is the gradient of weight parameter θ. It is a first-order optimization algorithm that is dependent on the first-order derivative of a loss function. There exist variants of gradient descent such as stochastic gradient descent and mini-batch gradient descent. The major drawback of gradient descent-based algorithms includes their inability to detect global minima, difficulty in determining optimal learning rate, and inability to have different and variable learning rates for different parameters [100].

$$\theta = \theta - \eta\nabla J(\theta; x, y) \tag{2}$$

The adaptive gradient algorithm (Adagrad) is shown in (3), where g_t is the gradient, and G_t is the sum of the squares of the past gradients with respect to all parameters θ, and ϵ is a smoothing term of order (10^{-8}) which prevents division by zero. This algorithm enables having different learning rates for different parameters alongside automatic tuning of learning rate. However, it is expensive computationally and has a decreasing learning rate resulting in slower training [101].

$$\Delta\theta = -\frac{\eta}{\sqrt{G_t + \varepsilon}} \cdot g_t \qquad (3)$$

Similarly, the other most commonly optimizer is Adadelta. It is an extension to Adagrad, shown in (4), where γ is the decay term that ranges from 0 to 1. Although it is computationally expensive, it resolves the monotonically reducing learning rate problem [102].

$$\Delta\theta = -\frac{\eta}{\sqrt{(1-\gamma)g^2 t - 1 + \gamma g_t + \varepsilon}} \cdot g_t \qquad (4)$$

The Adaptive Moment Estimation refered as Adam optimizer is shown in equation (5). It is by far the most used optimization algorithm after gradient decent. $\hat{m}_t$ and $\hat{v}_t$ are the bias corrected estimates of first and second moment and η is the learning rate respectively, given by $\hat{m}_t = m_t/(1-\beta_1^t)$ and $\hat{v}_t = v_t/(1-\beta_2^t)$ where $\beta_1 = 0.9$ and $\beta_2 = 0.999$. This optimizer does careful search of global minima without skipping over it. This optimizer rectifies most of the above stated problems such as vanishing learning rate, high variance, and local minima but is computationally expensive [103].

$$\Delta\theta = -\frac{\eta\hat{m}_t}{\sqrt{\hat{v}_t + \varepsilon}} \qquad (5)$$

4) FULLY CONNECTED LAYER

The final process before an output can predict results in flattened nodes is known as a fully connected layer. Since this layer is a single vector, no more operations are typically performed after this layer. However, several fully connected layers can be stacked like a traditional network before the final prediction is made [104].

5) DROPOUT FUNCTION

Deep neural networks can experience over-fitting problems. Over-fitting is a phenomenon when a model learns the statistical noise in the training data, leading to inadequate response when the model is tested on unseen data. Therefore, a deep learning model is expected to generalize its performance and prevent over-fitting. The dropout is a function developed to prevent over-fitting and improve model performance [105]. In this technique, neurons are randomly removed during training, leading to different connectivity with the previous layer. Consequently, the dropout breaks up the connections where network layers co-adapt to rectify mistakes from prior layers, making the model more generalized. One popular generalization descended from dropout is called

drop-connect [106], where weights are randomly dropped instead of activations. No-drop, dropout, and drop-connect variants are illustrated in Fig. 8.

C. PLATFORMS AND APPLICATIONS

This section encompasses common deep learning platforms available publically for research and development. It also highlights the generic categorization of deep convolutional networks in terms of their application. The most prominent and highly used platforms include the Google released TensorFlow library for the deep learning community. It used C++, Java, Python, and Go APIs and supports a multi-grid environment with CPUs and GPUs with CUDA and SYCL compatibility. It also comes with a lite version to aid mobile computing [107]. AI researchers and industry groups also acknowledge Keras. Keras is a high-level deep learning API, built on the top of the TensorFlow framework. It is highly modular, user-friendly, and supports rapid prototyping with an extendable interface. It can easily be integrated with other data, visualization, and machine learning packages build in python, such as NumPy, Matplotlib, seaborn, pandas, and scikit-learn, and is compatible with both CPU and GPU platforms [108]. Besides, PyTorch is one of the fascinating platforms built by a Facebook team to facilitate deep learning researchers. It is built on top of Torch using python as the primary scripting language. This platform supports highly scalable engines as well as mobile and embedded devices. It is more dynamic, memory efficient, and allows working with multiple GPUs in parallel without extensive effort [109]–[112]. The prevalence of convolution function for medical image analysis results in highly acknowledged application models in the AI community; these are classification and segmentation models. Image classification was the first area that was explored. In medical image classification, the networks are trained with multiple images as input with a single diagnostic variable as the output. The classification can be binary (diseased or healthy) or multi-class (different diseases/ stages of a disease) depending upon the nature of the dataset. [113]. Similarly, medical image segmentation is the process of separating different parts of an image into a set of groups such that all parts in a group have identical properties. Medical images often contain additional components other than tissues or organs of interest; segmentation (separation) of these anatomical structures facilitates better visualization of regions of interest. Therefore, segmentation plays a vital role in medical diagnosis [114].

D. PERFORMANCE EVALUATION METRICS

Verifying a deep semantic segmentation model is an essential task that tells how architecture performs on provided data. There is a wide range of evaluation metrics used to analyze model behavior in semantic segmentation [115]–[119]. This article discusses those metrics that are reviewed. The detail and mathematical depiction of each are enlisted below;
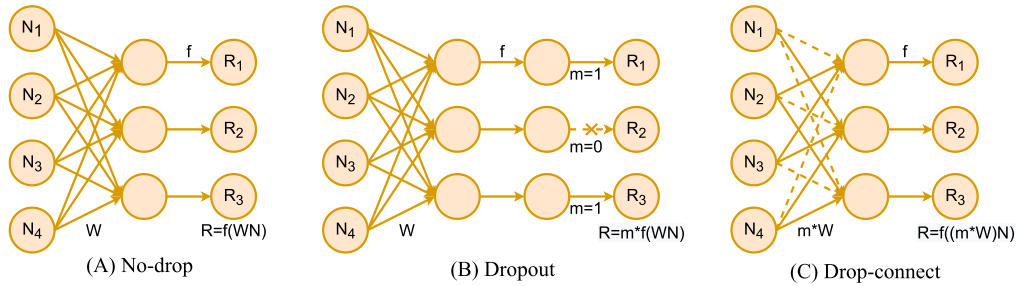
FIGURE 8. Visual depiction of dropout function to avoid overfitting and eliminate vanishing gradient problem.

1) PRECISION

The precision, also called a true predictive assessment, is a ratio between the number of true positives to the sum of the count of true positives and false positives. The true positive indicates the positive class's correct occurrence, whereas the false positive depicts its incorrect occurrence. It is a useful metric to measure prediction success. More specifically, it indicates what fraction of positive findings is true. Higher values of this metric dictate that the model returns precise and accurate outcomes. Hence, the higher the value of precision, the better the architecture is trained on provided data. The mathematical form of precision can be depicted as shown in (6). In the equation, TP is for true positive, and FP refers to false positive [120]–[124].

$$\text{Precision (PR)} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (6)$$

2) RECALL

The recall, also known as sensitivity, is the ratio between the number of true positives to the sum of the count of true positives and false negatives. Here, the true positive indicates the correct occurrence of the positive class, whereas the false negative depicts the negative class's incorrect prediction. More explicitly, it indicates what fraction of actual positives are found correctly. Like precision, recall is also a useful metric to measure prediction success, especially when the classes are not balanced. Higher values of this metric dictate that the model returns mainly all positive outcomes that are labeled correctly. Hence, the higher the value of recall, the better the architecture is trained [125]–[129]. The mathematical form of recall is provided in (7); here, TP is true positive, and FN refers to a false negative.

$$\text{Recall (RE)} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (7)$$

3) F-MEASURE

It is one of the metrics required to verify the model's accuracy. It unites precision and recall together in the form of harmonic mean. The purpose of using a harmonic mean rather than a simple average is to penalize extreme values. This metric is useful to bring a balance between precision and recall. The score obtained reaches its best value at 1 and worst at 0. Therefore, models that result in an f-measure close to

1 are considered optimal, which means that there are low false positives and false negatives. This metric applies to binary as well as multi-class classification and segmentation problems [127]. Mathematically it is represented in (8);

$$\text{F-Measure} = 2 \times \frac{\text{PR} \times \text{RE}}{\text{PR} + \text{RE}} \quad (8)$$

4) AREA UNDER THE CURVE

An area-under-the-curve (AUC) is the aggregated performance measuring metric used for classification and segmentation problems. It is computed using receiver operating characteristic (ROC) plots. It measures the 2D space beneath the ROC curve from coordinate [0, 0] to [1, 1], the degree to which the curve is in the northwest direction. It ranges from 0 to 1. The higher the AUC value interprets, the better result. Moreover, the AUC is scale and classification-threshold-invariant. This metric is evaluated using the ROC curve, which depends on the false positive rate (FPR) and true positive rate (TPR). The FPR is the horizontal axis, while TPR is the vertical axis of the ROC graph [128]. The mathematical form of TPR and FPR is shown in (9) and (10).

$$\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (9)$$

$$\text{FPR} = \frac{\text{FP}}{\text{FP} + \text{TN}} \quad (10)$$

5) INTERSECTION OVER UNION

The intersection-over-union (IoU) is a statistical validation tool, otherwise called the Jaccard index. It is one of the most regularly utilized measurements in segmentation. The IoU is a straightforward measurement that is very successful and commonly used for evaluating segmentation architectures. The intersection-over-union is defined as the overlapped area (AoO) ratio between the ground truth and prediction by the union of the area (AoU) between ground truth and prediction [129]. This evaluation metric range is between 0 and 1, where 0 means no overlapping and 1 refers to perfect overlap. Mathematically, it can be viewed in (11).

$$\text{IoU} = \frac{\text{AoO}}{\text{AoU}} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}} \quad (11)$$

TABLE 1. Metadata of datasets used in our survey article.

Dataset	Organ	Samples	Search Keywords
DRIVE [131]	Eye	40	Retina, Vessels, Segmentation, Fundus Image.
STARE [133]	Eye	20	Retina, Vessels, Segmentation, Fundus Image.
CHASE_DB1 [132]	Eye	28	Retina, Vessels, Segmentation, Fundus Image.
TCIA [134]	Pancreas	82	CT Scans, Pancreatic Lesions, Abdominal Pathology.
LIDC-IDRI [135]	Lung	1018	Lung Nodules, CT Scan, Lesions Segmentation.
ATLAS [136]	Brain	304	MRI Scans, Lesions Segmentation, Stroke.
BUS 2017 B [137]	Breast	562	Breast Tumor, Ultrasound Image, Cancer Dataset.
ISIC 2018 [138]	Skin	13,000	Skin Cancer, Lesion Extraction, Melanoma Detection.

6) DICE COEFFICIENT

The dice coefficient (DC) is a commonly used measure for segmentation applications and fundamentally the same as the IoU. It is defined as twice the overlapped area over the total pixels count in both images. Similar to IoU, it also ranges from 0 to 1. Value 1 depicts the highest similarity between the predicted value and the ground truth [130]. Hence, it finds the similarity between the two data samples. Mathematically the Dice coefficient can be expressed as shown in (12).

$$DC = \frac{2 \times TP}{(TP + FP) + (TP + FN)} \quad (12)$$

7) WARPING ERROR

The warping error is another measure used for boundary label comparison. It has tolerance to the variance for boundary location and can directly be used for learning boundaries. This metric focuses on objects and calculates the topological error amongst them rather than focusing on pixel variance. It is essentially the least mean square error between the pixels of the objective segmentation and the pixels of a topology-preserving distorted source, segmentation [130]. Mathematically, it is expressed in (13) using euclidean distance, or equivalently, hamming distance. In the provided equation, D represents distance, T is the candidate label, L* represents reference labeling, and L shows the pixel error.

$$D\left(T\|L^*\right) = \min_{L \lhd L^*} \|T - L\|^2 \quad (13)$$

III. DATASETS FOR MEDICAL IMAGE SEGMENTATION

In this section, various publicly available datasets for medical image segmentation are thoroughly discussed. These datasets are created and annotated under the supervision of domain experts and target multiple organs, depicted in Fig. 9. The details of each of these datasets is provided in the below subsections and a metadata is given in Table 1.

A. DRIVE

Digital Retinal Images for Vessel Extraction (DRIVE) dataset is built up to empower comparative investigations on blood vessel segmentation in retinal fundus images. The width and length of blood vessels, the tortuosity, branch structure, and angular directions are used to screen, diagnose, treat, and assess different ophthalmologic and cardiac problems, such as hypertension neovascularization, diabetes, and arteriosclerosis. Automatic localization and
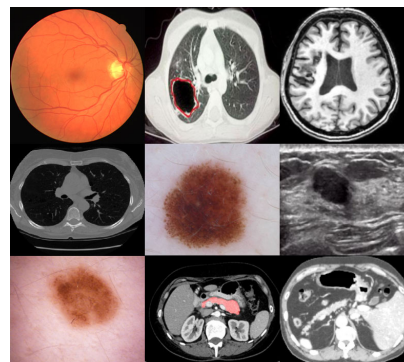

FIGURE 9. Medical Images acquired from various semantic segmentation challenges. The data is captured with specialized sensors at a particular angle and distinct image format.

investigation of the vessel structure can aid the execution of screening programs of aforementioned challenges. Additionally, the vascular structure is unique for each individual and can be utilized as a biometric ID. Data is collected from 400 diabetic patients with age ranges from 25 to 90 years. Out of 40 randomly selected color fundus images with 768 × 584, 33 images are normal, while 7 show signs of diabetic retinopathy. Images are in JPEG format and captured using a 45° field-of-view CR5 non-mydriatic 3CCD sensor. The dataset is partitioned into 20 training and 20 testing images. Single, manually-segmented vasculature is available for training, while testing contains two manuals. Moreover, for interpreting ROI, the mask of each fundus image is provided [131].

B. CHASE_DB1

This repository is a subset of the Child Heart and Health Study in England (CHASE) dataset, which contains retinal fundus images of multi-ethnic patients. It is the main subset in CHASE that has ground truth images of retinal vessels. Retinal fundus images of both eyes of patients were captured at a 35° field of view with an NM-200-D fundus image sensor. Data attained is in the TIF file format having 1280 × 960 image resolution. The dataset contains 28 images captured from both (left and right) eyes of the patients. The ground truths are manually segmented by two observers [132].

C. STARE

The Structured Analysis of the Retina (STARE) contains 400 retinal color fundus images. The images are captured with a 35° field of view using a TRV-50 fundus image sensor

having a resolution of 605×700 pixels. The dataset contains 20 fundus images suitable for segmentation challenge as it comes with ground truth images. Out of these 20 fundus images, 9 belongs to a normal class, while 11 shows signs of ophthalmological disease. These images are segmented manually by two domain experts [133].

D. TCIA

The National Institutes of Health Center produced 82 abdominal contrast-enhanced 3D CT scans from 53 male and 27 female subjects. Seventeen subjects are healthy kidney donors, scanned before nephrectomy. A radiologist selected the remaining 65 patients from patients without any significant abdominal pathologies or pancreatic lesions. The subjects' age ranges from 18 years to 76 years, with a mean age of 46.8 ± 16.7. Scans have 512×512 pixel resolution with varying pixel sizes and slice thickness between 1.5 - 2.5 mm, acquired on Philips and Siemens MDCT scanners (120 kVp tube voltage). A medical student manually performed slice-by-slice segmentation of the pancreas as ground-truth, and these were verified by an experienced radiologist [134].

E. LIDC-IDRI

The Lung Image Database Consortium image collection contains lung cancer screening thoracic computed tomography (CT) scans with marked-up annotated lesions. It is a web-accessible international resource for developing, training, and evaluating computer-assisted diagnostic methods for lung cancer diagnosis. It is initiated by the National Cancer Institute, further modified by the Foundation for the National Institutes of Health, and accompanied by the Food and Drug Administration; this public-private partnership depicts a consortium's success. Fifteen companies collaborated to create this data set containing 1018 case studies. Each subject includes images from a clinical thoracic CT scan and an XML file that records the results of a two-phase image annotation performed by four radiologists. In the initial blinded-read phase, every radiologist independently reviewed each CT scan and marked lesions belonging to any of three categories ("nodule > or = 3 mm," "nodule < 3 mm," and "non-nodule > or = 3 mm"). In the unblinded-read phase, each radiologist independently reviewed their marks along with the anonymized marks of the three other radiologists to render a final opinion. This dataset aims to identify all lung nodules in each CT scan as completely as possible without requiring forced consensus [135].

F. ATLAS

Anatomical Tracings of Lesions After Stroke, abbreviated as ATLAS, is an open-source dataset. This repository's main objective is to facilitate researchers with a benchmark dataset to evaluate their algorithms on T1-weighted MRIs for lesion segmentation challenge. There are 304 T1-weighted MRIs gathered from the ENIGMA Stroke Recovery Working Group consortium. An open-source tool, MRIcron, points brain lesions and draws manual masks for every individual brain image. At least one mask is recognized for every individual MRI. Likewise, an expert investigated all lesions to give extra subjective depictions of the type of stroke, the intensity of white matter disease, and the avascular area. In short, the repository includes; CSV file with metadata, 229 T1-weighted MRI scans, and MNI152 standard-space T1-weighted average image. Hence, it is an asset to evaluate and improve the precision of current lesion segmentation strategies [136].

G. BUS 2017 DATASET B

Cancer caused by Breast tumors is one of the dominant causes of deaths in women. From the statistical analysis, it is found that one in eight women in the United States of America have breast cancer in their life. Precise breast tumor segmentation is crucial and challenging for further diagnosis and treatment. Various approaches are used for Breast Ultrasound Segmentation, but most of these methods are evaluated using comparatively tiny private datasets, which causes inconsistency for performance comparison. Therefore, to fill that gap of having a large-scale public dataset, a benchmark Breast Ultrasound Segmentation Mode B is published. It contains 562 images gathered from multiple resources using various ultrasound sensors, including Philips IU22, Siemens ACUSON S2000, Hitachi EUB-6500, GE VIVID 7, and LOGIQ E9. Moreover, four domain experts are involved in generating ground truths for this dataset [137].

H. ISIC 2018

The input sub-data contains dermoscopic lesion images in JPEG format. All lesion images are provided in the format of ISIC_<image_id>.jpg, where image id is a 7-digit identifier. The lesion images were acquired with various dermatoscopy types, from all anatomic sites (excluding mucosa and nails), from a historical sample of patients presented for skin cancer screening from several different institutions. Every lesion image contains exactly one primary lesion; other fiducial markers, smaller secondary lesions, or other pigmented regions may be neglected. The disease states' distribution represents a modified real-world setting whereby there are more benign lesions than malignant lesions but an over-representation of malignancies. The response data contains binary mask images in PNG format, indicating the primary skin lesion's location within each input lesion image. Mask images are encoded as grayscale 8-bit PNGs images. Each pixel represents areas outside the primary lesion as a background region or the foreground of the image or areas inside the primary lesion. Mask image ground truth (provided for training and used internally for scoring validation and test phases) data were generated using several techniques, but all data were reviewed and curated by practicing dermatologists with expertise in dermoscopy [138].

IV. MEDICAL IMAGE SEGMENTATION ARCHITECTURES

Medical image semantic segmentation is performed to delineate the anatomical structures and other regions of interest
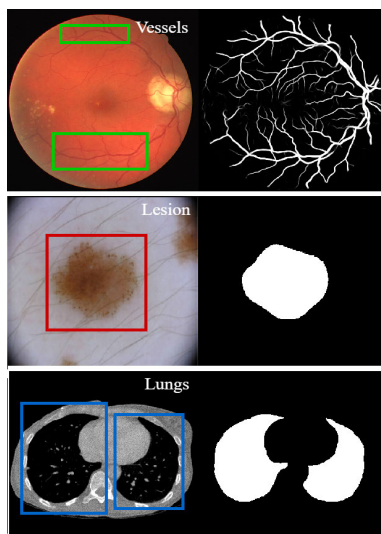
FIGURE 10. The figure depicts original medical images with semantically segmented outcomes. The regions of interest are highlighted in color boxes.
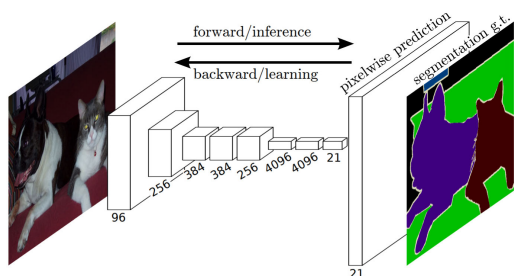

FIGURE 11. FCN for pixel-wise semantic segmentation [9].

where each pixel in an image is classified into a predefined set of classes as shown in Fig. 10.

A. FCN ARCHITECTURES

A fully convolutional network (FCN) is an advanced, deep learning-based semantic segmentation architecture developed in 2015 [9]. FCN has been widely used in medical image segmentation for various applications [139]–[142]. An FCN utilizes a CNN to convert image pixels to pixel categories such that the predictions and input image pixels have a one-to-one correspondence as shown in Fig. 11. This architecture transforms the intermediate layer feature map's dimensions back to the input image's size using transposed convolution layer. The major benefit of FCN is that it provides an end-to-end solution for semantic segmentation, even on variable-sized images. The shortcomings of FCN include its high computational cost and difficulty in adapting to three-dimensional images.

B. ENCODER-DECODER BASED ARCHITECTURES

SegNet introduced semantic pixel-wise image labeling [62]. It used forward evaluation of a fully learned function to get a
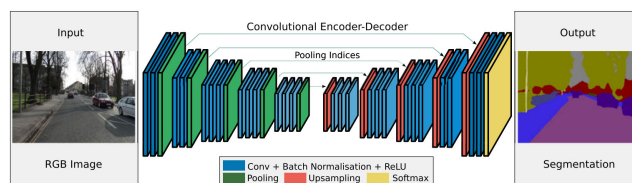

FIGURE 12. The SegNet architecture [62].

smooth prediction of the label. With depth increase, a broader context is considered for pixel labeling, which improves accuracy. It makes it easy to visualize the impact of feature activation in the pixel label space at any depth. It comprised a stack of encoders followed by a corresponding decoder stack, which feeds into a soft-max classification layer as shown in Fig. 12. The decoders help map low-resolution feature maps at the encoder stack's output to full-size feature maps identical to the input size. It addressed an essential drawback of recent deep learning approaches, which have adopted networks designed for object categorization for pixel-wise labeling. It also resorts to ad hoc methods to upsample features and results in noisy predictions, restricts the number of pooling layers to avoid too much upsampling, and reduces spatial context. SegNet resolved these problems by learning to map encoder outputs to image pixel labels.

The U-Net architecture is the most prominent medical image segmentation model applied to various medical problems, published in 2015. It is a common belief that successful architecture training needs a massive amount of data [4]. U-Net presents a strategy that strongly depends on the data augmentation technique to use limited available data more effectively. It is built on fully convolutional layers arranged to produce better segmentation results explicitly for medical images. The architecture consists of encoder/decoder paths, depicted in Fig. 13. The encoding path is a stack of unpadded convolutions followed by a max-pooling operation for down-sampling. It helps to find insight by exploring advanced features but, at the same time, causes a reduction in the size of the feature map. The symmetric decoding path uses transpose convolutions to perform precise localization. Hence, the concatenation of feature maps associated with encoding-decoding units at the same level is performed. It supports localization from an encoder to a decoder path. Since annotating a large number of training samples in medicine is expensive and time-consuming, U-Net architecture helps and supports the segmentation process by lowering the cost and time it takes to annotate. For architecture evaluation, PhC-U373 and DIC-HeLa Datasets have been used to segment Electron Microscopic (EM) images. The main contribution of this architecture includes data augmentation, separation of joint objects, and overlapping-tile approach.

V-Net, a volumetric convolutional neural network, which is FCN for Volumetric Medical Image Segmentation, was proposed in [143]. It is a modified version of U-Net with a foundational encoder-decoder structure as depicted in Fig. 14.
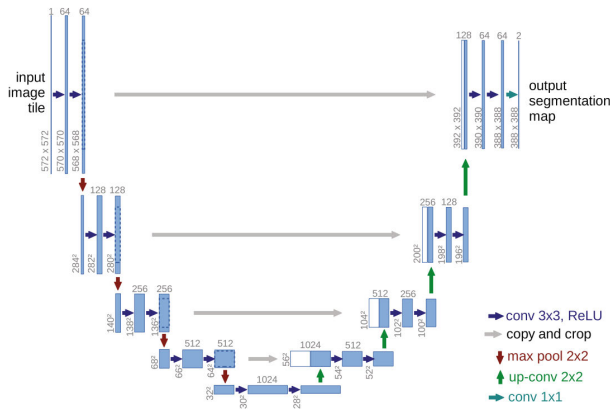
FIGURE 13. The U-Net architecture [4].

FIGURE 14. The VNet architecture [143].

Medical data comes in different dimensions, so it is always crucial to handle such details flawlessly. Many existing SOTA approaches only entertain 2D information, whereas various clinical information comes in 3-Dimensional volumes. The article demonstrated volumetric FCN for 3-Dimensional image data manipulation (segmentation). The proposed network is simulated over 3D MRI volumes for prostate segmentation; a critical diagnostic task for assessing the prostate condition. Moreover, a unique Dice coefficient-based objective function is introduced, which deals with an imbalance between foreground and background voxels. The model made use of histogram matching and random non-linear transformations for data augmentation. Experimental evaluation showed that the underlying approach is time efficient and needs fractional processing time to execute.

In 2016, 3D U-Net, a Dense Volumetric Segmentation network, was proposed by the University of Freiburg, University Hospital Freiburg, University Medical Center Freiburg, BIOSS Centre for Biological Signalling Studies, and Google DeepMind [144]. Proposed architecture replicated the U-Net as a baseline with analysis and synthesis paths. The analysis path contains 3-Dimensional convolutions, followed by ReLU activation and max-pooling layers. Similarly, the synthesis path contains 3-Dimensional up-convolution leading to convolutions and ReLU activation. Finally, the output channels are reduced to meet the number of labels with $1 \times 1 \times 1$ convolution. It should be noted that Batch-Normalization is performed before each activation. The framework operates in Bi-mode to semi-automatically and fully automatically segment 3-Dimensional volume from limited annotations. Elastic deformation is used for data augmentation during network training. It is observed that the proposed architecture responded well when tested on the Xenopus kidney.

C. MULTISCALE ARCHITECTURES

Multiscale segmentation architectures were developed to address the existence of scale variance in the regions of interest. Multiscale segmentation architectures can reveal the information at different scales by splitting a given image into
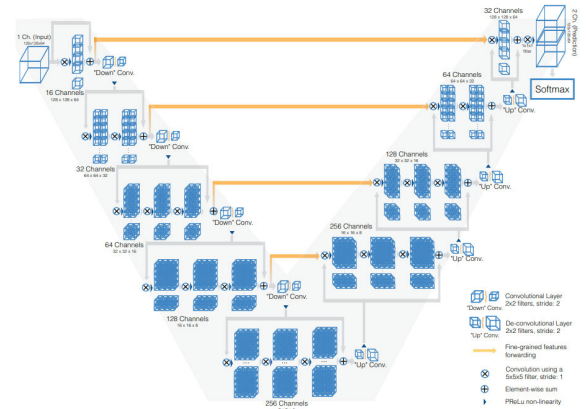
several homogeneous regions at various scales. These architectures used a combination of different scales of features (local and global) to predict the classification of each pixel, thereby improving high-level segmentation operation.

In 2018, U-Net++ was proposed by Arizona State University, USA to overcome U-Net's limitation of utilizing same-scale feature maps alone. The architecture used the concept of Dense-Block to improve original U-Net performance as shown in Fig. 15. Unlike the foundation model (U-Net), it included convolutions and dense skip connections on skip-pathway to fill the gap between feature maps across modules and to improve gradient flow [145]. It also incorporated the concept of deep supervision, which performs architecture pruning. The proposed architecture is evaluated in a multi-modal environment by considering four different medical image repositories, including; cell nuclei, colon polyp, liver, and lung nodule. The proposed model is compared with two foundational models (U-Net and Wide U-Net). The results have shown that U-Net++, even without the concept of deep supervision, achieved phenomenal results in terms of Intersection-over-Union (IoU) metric for both baseline U-Net and Wide U-Net models. DMCNN is a deep multiscale convolutional neural network developed in 2019 to tackle image quality issues such as low gray contrast and blurred tissue boundaries in medical images [146].

D. PYRAMID BASED ARCHITECTURES

The PSPNet [147] has proposed a pyramid pooling module to aggregate the context and used auxiliary loss as shown in Fig. 16. Global scene categories matter because it gives a clue on the distribution of the segmentation classes. The pyramid pooling module captures this information by applying large kernel pooling layers. Dilated convolutions are used to modify Resnet, and a pyramid pooling module is added to it. This module concatenates ResNet's feature maps with an upsampled output of parallel pooling layers with kernels covering the whole, half, and small portions of an image. An auxiliary loss, additional to the main branch loss, is applied after the fourth stage of ResNet.
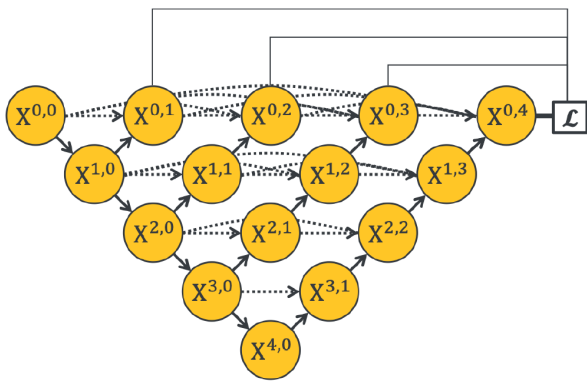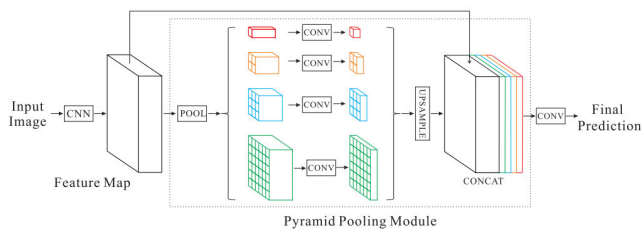
FIGURE 15. The U-Net++ architecture [145].

FIGURE 16. Overview of PSPNet Architecture from [147].

E. HYBRID ARCHITECTURES

Hybrid densely connected U-Net, abbreviated as H-DenseUNet, was proposed in 2018 by the Chinese University of Hong Kong [13]. In the Proposed Model, for precise extraction of intra-slice features, 2D DenseUNet is used whereas, for a hierarchical collection of volumetric frames of reference, the 3D counterpart is utilized. Moreover, intra-slice and inter-slice representations are collectively optimized via a hybrid feature fusion layer. The hybrid feature is the sum of intra-slice and inter-slice representations from 2-Dimensional and 3-Dimensional networks. The model is evaluated using 3DIRCADb Dataset for liver and tumor segmentation. It is found that H-DenseUNet outstripped DeepX for lesion and liver segmentation. It has also shown an enormous response compared to U-Net, with a 14.5 percent improvement in segmenting tumors' Dice score.

In an article published in 2018, Cascaded 3D U-Net is proposed by Nagoya University, Nagoya University Graduate School of Medicine, and Aichi Cancer Center [148]. The proposed architecture is a two-stage 3-dimensional U-Net model. The first stage is a fully convolutional neural network, trained to approximately trace organs of interest. The second stage executes another fully convolutional Network architecture to get a more detailed view of organ segmentation. It ensures a coarse-to-grain smooth transition. Candidate Regions C1 and C2 are specified for stage 1 and stage 2, respectively, to reduce the search space. Consequently, for reducing the output channel to make it equivalent to the number of output classes, the final layer contains $1 \times 1 \times 1$ convolution. Moreover, the size of each specified channel is $44 \times 44 \times 28$. The

loss function used for the proposed architecture is weighted voxel-wise cross-entropy loss. The proposed approach is operated on 331 contrast-enhanced abdominal clinical CT images with random rotations and Smooth B-spline deformations techniques to augment available data. The designed paradigm outperformed previous multi-organ segmentation approaches with an improved dice score of 0.792 from 0.717.

Cascaded Deep Convolutional Neural Network is proposed in [149]. The proposed architecture is considered as an ancient and conceptually simple alternative to know existing structured methods well. It follows a two-pathway approach that learns local information of the brain as well as the global context. The model also incorporates a bi-phase training process, which is significant in dealing with imbalanced tumor labels. The model is evaluated using the 2013 BRATS test dataset. Results proclaim that architecture is 30 times faster than existing state-of-the-art methods as it takes 25 seconds to 3 minutes for segmenting brain images.

Models Genesis, an approach to segment 3-Dimensional medical image data using transfer learning strategy was presented by Zhou *et al.* [85] in 2019. Applying transfer learning to medical data is crucial; often, the clinical data available is in a 3-D format such as MRI and CT. By taking the benefit of pre-trained models without training medical images from scratch, there exists a need to manipulate 3D data and transform it into 2D images. This manipulation consequently causes the loss of valuable spatial information, which ultimately impacts overall application performance. To resolve this problem, the authors of this paper presented a set of models named Models Genesis that are generic, self-taught, and capable enough to operate with no manual labeling. The authors of the article aimed to produce application-specific target architecture using a transfer learning approach. They evaluated the paradigm on five 3-Dimensional applications targeting both classification and segmentation challenges. It is significantly observed in experimentation that models trained from scratch on 3-Dimensional medical data may not essentially perform well compared to transfer learning from the 2D-ImageNet dataset. However, the proposed framework uniformly showed SOTA performance to any 2D strategy, which confirms its significance. Moreover, the authors have shown their intention to extend the approach from application-specific to domain/modality-specific.

In 2018 Google DeepLab had come up with DeepLabv3+ after series of DeepLab versions. Along with the foundational Contraction-Expansion pattern that can randomly manage the resolution of extracted encoder representations using Atrous convolution to compensate runtime and precision, the architecture also incorporates the Xception model with Separable Convolution (Depth-wise Convolution combined with point-wise convolution) for making architecture more robust and computationally efficient. The Architecture obtained SOTA performance on publicly available datasets [150]. Moreover, multiple variants of the proposed model are analyzed in the reviewed article. It is observed that the resultant model performs better than its previous version, DeepLabV3.

F. OTHER SEGMENTATION ARCHITECTURES

The Regional CNN (R-CNN) architectures were developed for object detection [151]. R-CNN extensions such as Mask R-CNN solves the instance segmentation problem, which requires simultaneous object detection and segmentation [152]. CenterMask is a recent Mask R-CNN extension that provides real-time anchor-free instance segmentation [153]. The authors of [154] proposed M-GAN, a conditional generative adversarial network for retinal vessel segmentation. M-GAN uses stacked deep FCN to balance the losses. It consists of M-generator with deep residual blocks and M-discriminator with a deep network to train the adversarial model. Besides, it includes a multi-kernel pooling block in between the stacks to achieve scale-invariance. DU-Net was published in 2019 by Tianjin University and Linkoping University. It utilized a U-Shape model like U-Net architecture for exploiting local features of retinal blood vessels with transpose convolutions for fetching context information [136]. It concatenated high- and low-level feature maps for accurate localization. Moreover, the model autonomously adapts, adjusting local, dense, and receptive fields concerning vessel shape and scale by incorporating Deformable Convolutional blocks. Each block contains offset, Convolution, batch-norm, and activation layers. Articulated architecture is evaluated using three benchmark retinal fundus image datasets: DRIVE, STARE, and Chase_DB1. Experimental results have shown that DU-Net performed better than other competing baseline models (DCN and U-Net) with fewer parameters.

V. ENHANCEMENT MECHANISMS

The classical models based on a U-Shaped structure had certain pitfalls. To resolve these gaps, researchers produced various enhancement mechanisms. In this section, we briefly reviewed the literature build on these mechanisms.

A. RESIDUAL MECHANISM

In [155], the authors proposed ResNet architecture. It performed enormously for image classification and segmentation. The ResNet used residual block to enhance feature representation and boost network performance. It supports deep net-works with increased depth without falling into the problem of vanishing gradient. It was initially developed for image classification; however, scientists utilized its significant impact by joining residual blocks with U-Shaped models. In [156] the authors proposed U-Net with Residual block variant. The work demonstrated the impact of long and short skip connections with FCN for segmenting medical images. Unlike FCN, which only uses long skip connections to skip representations from an encoder to decoder path, underlying architecture enhances FCN with short skip connections used in a residual network for deeper architecture. Full two-resolution Electron Microscopy (EM) Image dataset without any post-processing is used to evaluate the proposed model with three variants. It is noticed that high accuracy with small loss is obtained with an architectural variant containing both long and short skip connections. The authors in [157] proposed RDN, a Residual de-ConvNet inspired by U-Net architecture. It captures image features and contextual details. The residual block not only imparts depth support but also aid for multiscale features. Similarly, in [158], the authors integrated residual function with a U-Shaped structure. The model contained multi-scale kernels to aggregate the mappings from different size kernels for capturing more context information. In R2U-Net, the author proposed two deep semantic segmentation networks called Recurrent U-Net (RUNet) and Recurrent Residual U-Net (R2U-Net) for medical image segmentation [159]. The proposed model utilizes the foundational structure of U-Net along with residual units, which helps in training deep architectures along with recurrent convolutions, which assure improved feature representation for segmentation. The comparison of proposed models with other methods are made by keeping the number of parameters constant. The VoxResNet is proposed in [160]. The residual block is used to mitigate the performance degradation with increased model depth. In [161], the authors used residual structure to improve the convergence rate of back-propagation for lesion segmentation. The authors of [162]–[164] embedded residual units to enhance model feedback.

B. DENSE MECHANISM

The authors of [165] proposed DenseNet, which takes away the conventional concept of ResNet and improved network response by advancing a more comprehensive approach. It is convolution neural network with dense connections. The dense mechanism used in a network maximizes the information and gradient flow. It restrains the vanishing gradient problem and depreciates the number of training parameters. In dense blocks, the two layers are well-connected so that the input of each layer is the sum of the output of all previously processed layers, and features are learned incrementally by the network. Hybrid densely connected U-Net abbreviated as H-DenseUNet was proposed in [13]. In the proposed model, for precise extraction of intra-slice features, 2D DenseUNet is used, whereas, for a hierarchical collection of volumetric frames of reference, the 3D counterpart is utilized. Moreover, intra-slice and inter-slice representations are collectively optimized via a hybrid feature fusion layer. The hybrid feature is the sum of intra-slice and inter-slice representations from 2-Dimensional and 3-Dimensional networks. The model is evaluated using 3DIRCADb Dataset for liver and tumor segmentation. It is found that H-DenseUNet outstripped DeepX for lesion and liver segmentation. The dense unit is integrated with two symmetrical U-shaped models in [166]. The authors used multiple kernels to obtain useful features from sparse pixels. The performance of the developed method is analyzed with two publicly available medical image datasets. In [167], the authors developed a network model for gland segmentation. The network employed three multiscale dense units connected with a U-shaped structure. The authors of [168] have proposed DRINet architecture,

an inspiration of the U-Net for semantic image segmentation. It encompasses analysis and synthesis paths. The analysis path contains dense units; however, the synthesis path comprises residual units.

C. DILATED MECHANISM

In [169], the authors introduced dilated convolution concept in image segmentation. The conventional convolution networks used downsampling to reduce provided image and upsampling to bring the compressed image back to its original shape. The process causes essential information loss. However, the dilated convolution operates without a pooling function and mitigates the loss of essential pixel information. It enlarges the size of the receptive field to promote an extensive range of feature details. The dilated convolutions include dilation rate to reflect expansion size and share convolution core size. The networks that operate with dilated convolution also have an edge with minimalistic training parameters. The dilated convolution has the simplest structure to improve segmentation accuracy. In [170], the authors applied dilated convolution in STRAINet architecture to localize organs from MRI data. It helped the model to get the expended receptive field cost-effectively. The authors in [171] used multi-scaled dilated units in dense model to segment stroke lesion. The multi-scaled units efficiently handled the lesion of variable size. Similarly, [172] used dilated convolution in the 3D segmentation model that has shown promising impact compared with a model proposed in [173]. Also, [174] and [175] have also employed dilated convolution for segmenting atrial and brain lesions.

D. ATTENTION MECHANISM

The inspiration of the attention mechanism is taken from the human brain. The human brain does not understand all the information transmitted to the visual and auditory cortex; however, it keeps track of essential details extracted from the bulk of input data. The process improves the processing time and efficiency [176]. The attention mechanism was originally used in the recurrent neural network; later, it was widely adopted by other AI areas such as image localization, speech recognition, and machine translation [177]–[180]. Google DeepMind team proposed a recurrent visual attention model, which has received great appreciation from the research community [176]. Similarly, the authors in [177] used attention mechanisms for solving natural language processing problems. The RNN's extensions employed attention blocks for solving NLP problems in [181]. Researchers found that the attention mechanism has its application in medical image segmentation and became the core research topic. In [136], the authors used an attention mechanism in medical image semantic segmentation. The RA-UNet used this mechanism to extract context information by joining low and high-end feature maps. It helped the model to restore missing details. Besides, [155] has embedded attention gates with a deep network to segment abdominal organs from 2D CT scans. The attention mechanism aids the model to learn organs

of variable size and shape. The authors in [182] proposed CSAU-Net for vessel extraction. Before producing output, the model embedded connection loss and attention weight to the features. In [183], the authors proposed an attention-based U-Net model for segmenting CT-150 and CT-82 datasets with various settings. It is observed that without much increase in the number of parameters, without applying post-processing by just using Attention Gates, Attention U-Net outperformed the U-Net model with a higher Dice value.

E. ENSEMBLE MECHANISM

The ensemble mechanism is efficient in training for reducing the overfitting problem [184]. The ensemble unit contains a group of networks that execute in parallel, and their output finally joins together to complete the segmentation process [184]. It comprises a model with image pre-processing and post-processing units. Each phase contributes to the final prediction so that pre-processing unit enhances the input image provided to the network model to segment the region of interest, followed by a post-processing step to refine the outcome. Often it is found by the deep community that ensemble mechanism yields better response than conventional singular sub-network. Each sub-network in the ensemble module contributes to learning different features during the training process [185]. In [186], the authors proposed an ensemble network of networks for tumor localization and won the BraTs 2016 challenge. The authors in [185] defined multiple U-Shaped models to target multi-modalities using MRI data [187]. Similarly, [188] proposed two different U-Net-inspired models for medical image semantic segmentation. One of them captures residual signals from a low-resolution path; however, the other encompasses orthogonal-wavelet frames. The technique has better results than the conventional U-Net model. The authors in [136] came up with DUNet for vessel extraction. The model fairly captures the local features at various scales and shapes by adaptively managing receptive fields. Similarly, [3] presented a novel system for optic cup segmentation to detect chronic ocular glaucoma. They have used the polar conversion method to design their system with multiscale U-shaped network architectures.

VI. DISCUSSION

This section demonstrates the comparative impact of deep semantic network architectures in medical imaging. It is observed that searching an efficient deep network strategy is still a critical task. It is hard to obtain a huge amount of medical data records for training segmentation model. The limited availability of data can easily lead to an overfitting problem. Increasing the number of data instances and reducing the complexity of the network can be an optimal solution to overcome this problem. Moreover, imbalanced data training can result in network instability, especially in segmenting small structures. Resampling the data space, using a two-phase training process along with careful path sampling and selection of appropriate loss function can

TABLE 2. Comparison of deep neural techniques evaluated over retinal image datasets for detecting chronic eye disease.

Dataset	Method	Evaluation Metrics	
		F1-Score	AUC
DRIVE	U-Net [4]	0.8142	0.9755
	Residual U-Net [189]	0.8149	0.9779
	R2U-Net [190]	0.8171	0.9784
	LadderNet [191]	0.8202	0.9793
	DUNet [136]	0.8237	0.9802
	VGN [133]	0.8263	0.9802
	IterNet [192]	0.8205	0.9816
	CE-Net [193]	—	0.9779
	SA-UNet [194]	0.8263	0.9864
	SGL [195]	0.8316	0.9886
	RV-GAN [196]	0.8690	0.9887
CHASE_DB1	U-Net [4]	0.7783	0.9772
	Residual U-Net [189]	0.7800	0.9779
	DUNet [136]	0.7883	0.9804
	VGN [133]	0.8034	0.9830
	R2U-Net [190]	0.7928	0.9815
	LadderNet [191]	0.8031	0.9839
	IterNet [192]	0.8073	0.9851
	SA-UNet [194]	0.8153	0.9905
	SGL [195]	0.8271	0.9920
	RV-GAN [196]	0.8957	0.9914
STARE	U-Net [4]	0.8373	0.9898
	Residual U-Net [189]	0.8388	—
	DUNet [136]	0.8143	0.9832
	VGN [133]	0.8429	0.9877
	R2U-Net [190]	0.8475	0.9914
	RV-GAN [196]	0.8323	0.9887

eliminate this problem. It is becoming a global practice in the deep learning community to out-source code and data to the public. It helps expedite the research and open the door to design robust segmentation models. The segmentation architectures developed for medical image analysis have a profound impact. This claim is supported by empirical evidence provided in Table 2, 3, 4 and 5. It is found that U-Net architecture has clear dominancy over other architectures in terms of medical applicability for the multi-modal environment. This architecture is observed to be the universal semantic segmentation approach with a promising response for multi-organ segmentation challenge, depicted in Table 2-5. The R2U-Net is dominating on STARE and LUNA datasets when compared with other methods interpreted in Table 2 and Table 5. Residual U-Net supports the concept of having a better solution with more deep networks by showing a better response to the conventional U-Net model mentioned in Table 2, 3 and 5. The deformable U-Net has a low tolerance for entertaining data variation. It is observed through a systematic review that it has a fluctuating response for retinal vessel segmentation shown in Table 2. For DRIVE, it supersedes U-Net, Residual U-Net, and R2U-Net; however, it exhibits low response for the STARE dataset. The IterNet model is applied and more suitable to retina pathologies, as can be seen in Table 2. The Attention U-Net is used for lesion and pancreas segmentation. It has achieved a comparable response to other techniques when applied to BUS 2017, ISIC 2018, TCIA, and CT-150

datasets as shown in Table 3 and 4. It is found that ensemble networks with multiple techniques together are producing better performance than unified architectures, demonstrated through the results given in Tables 3, 4 and 5 respectively.

A. MOTIVATION

The motivation behind our effort is to provide a comprehensive application and methodology-driven overview that reflects key medical imaging modalities. The study reviews the state-of-the-art deep learning techniques designed to perform medical image analysis; it also assesses critical challenges associated with medical diagnosis and provides future directions with recommendations. The concept emerged from the fact that medical applications have a direct, immediate impact on human life and safety. Any intelligent solution proposed is acceptable if it operates with a high certainty of diagnostic decisions. The more deep networks operate with a massive number of layers. The increase in network depth positively impacts model performance; however, it negatively impacts architectural complexity with increased parameters and training cost. The models with fewer layers and parameters have to compromise on their efficiency. This gap needs to be filled with optimal techniques, specifically designed for healthcare, which reduces overall system cost without degraded system response applicable to the multi-modal environment.

B. CHALLENGES

The critical challenges that exist in medical image analysis observed while conducting this review are listed below:

- The skip-connections equip the network with rich feature representations at additional memory and computational cost. Often, it results in transferring non-discriminative features. There is a need to optimize the amount of data being transferred via skip-connections.
- The intensity-based region of interest segmentation using prior knowledge, is one of the active research paths.
- Medical images are larger than natural images, which inhibits the ability to load them entirely onto the memory for processing. They need to be preprocessed with patch-based or sub-volume-based techniques. It makes it difficult for segmentation models to capture spatial relationships. It leads to a potential need for building architectures capable of incorporating spatial relationships with high-resolution medical data.
- A universal solution that yields adequate response across multiple medical imaging modalities is needed.
- Data scarcity is a major obstacle to the healthcare automation process, where medical image datasets typically are smaller in size. The large 2D and 3D datasets creation allow researchers to accurately benchmark segmentation methods and make incremental enhancements to the problem's solution.

TABLE 3. Comparison of deep neural techniques experimented for lesion segmentation.

Dataset	Method	Evaluation Metrics			
		Precision	Recall	Dice	IOU
ATLAS	SegNet [62]	0.3938	0.2532	0.2767	0.1911
	PSPNet [147]	0.4769	0.3335	0.3571	0.2540
	U-Net [4]	0.5994	0.4449	0.4606	0.3447
	DeepLab v3+ [197]	0.5831	0.4491	0.4609	0.3458
	ResUNet [189]	0.5941	0.4537	0.4702	0.3549
	2D Dense-UNet [13]	0.5613	0.4875	0.4741	0.3559
	X-Net [198]	0.6000	0.4752	0.4867	0.3723
	D-UNet [136]	0.6331	0.5243	0.5349	—
BUS 2017 Dataset B	Attn U-Net + DL [199]	—	—	0.615	—
	U-Net + FTL [199]	—	—	0.669	—
	Attn U-Net + Multi-Input + FTL [199]	—	—	0.804	—
ISIC 2018	Attn U-Net + DL [199]	—	—	0.806	—
	U-Net + FTL [199]	—	—	0.829	—
	Attn U-Net + Multi-Input + FTL [199]	—	—	0.856	—
	MCGU-Net [200]	—	—	0.895	—
	DoubleU-Net [138]	—	—	0.896	—
ISIC 2017	Automatic skin lesion segmentation with fully convolutional-deconvolutional networks [141]	—	—	—	0.765
ISLES-2015	3D CNN + CRF [201]	—	—	0.590	—

TABLE 4. Comparison of deep neural architectures and techniques proposed for pancreas segmentation.

Dataset	Method	Evaluation Metrics			
		Precision	Recall	Dice	JI
TCIA Pancreas-CT Dataset	T. Okada et al. [202]	—	—	0.733	0.600
	U-Net [4]	—	—	0.820	—
	A. Saito et al. [203]	—	—	0.744	0.623
	Recurrent Saliency Transformation Network [134]	—	—	0.876	—
	S. Li et al. [204]	—	—	0.789	0.654
	Att U-Net [183]	—	—	0.831	—
	J. Cai et al. [205]	—	—	0.833	0.718
CT-150	T. Tong et al. [206]	—	—	0.655	—
	U-Net [4]	0.848	0.806	0.814	—
	K. Karasawa et al. [207]	—	—	0.763	0.639
	Att U-Net [183]	0.849	0.841	0.840	—

TABLE 5. Comparison of deep semantic techniques applied for lung nodule segmentation.

Dataset	Method	Evaluation Metrics					
		F1-Score	AUC	Dice	IOU	Sensitivity	Specificity
LUNA	U-Net [4]	0.9658	0.9784	—	—	—	—
	Residual U-Net [189]	0.9690	0.9849	—	—	—	—
	R2U-Net [190]	0.9823	0.9889	—	—	—	—
	Gong et al. [208]	—	—	—	—	0.7930	—
	BCDU-Net [209]	0.9904	0.9946	—	—	—	—
	ET-Net [209]	—	—	—	0.9623	—	—
LIDC-IDRI	Shaukat et al. [210]	—	—	—	—	0.9199	0.9648
	Zhang et al. [211]	—	—	—	—	0.7930	—
	Aoc et al. [212]	—	—	—	—	0.8560	—
	Naqi et al. [213]	—	—	—	—	0.8930	—
	Gupta et al. [214]	—	—	—	—	0.9220	0.9864
	Silva et al. [215]	—	—	—	—	0.9860	0.9820
	Model Genesis [85]	—	—	0.7586	0.7762	—	—
	Semantic Genesis [135]	—	—	—	77.24	—	—

- Methods are needed to find the root cause of false-positive and negative predictions in segmentation models and datasets.
- Need for reinforcement learning techniques for semantic (medical) image segmentation to mimic the way humans delineate objects of interest. Deep CNNs successfully extract features of different classes of objects, but they lose the local spatial information of where an object's borders should be. Some researchers resort to traditional computer vision methods such as conditional random fields (CRFs) to overcome this problem, making the model computationally extensive.
- Deep segmentation architectures designed for medical image analysis rely mostly on data obtained from a clinic. There often exists valuable information containing subject data in other imaging modalities not used by segmentation models. Hence, the deep community is searching for intelligent models capable of leveraging multi-modal patient data to improve prediction results.
- Deep neural networks mainly perform backpropagation using gradient descent method to alter network weights. Exploring methods that do not rely on backpropagation would be another significant research direction.
- Another potential need is to develop a network for searching optimal segmentation networks and design new layers that could capture novel data aspects other than convolutions and transform convolutions.
- The 3D networks defined for medical image segmentation need a massive amount of training parameters. The volume depth ranges from 20 to 400 slice/scan, where each scan contains vital medical detail of a subject. These scans are downsampled with different interpolation techniques. This reduction in volume-size causes information loss. Also, there exist constraints on resizing medical image volumes before being supplied to the network model.

C. RECOMMENDATIONS

The heterogeneous manifestation of the organs being examined and the extraction of such organs is crucial as the body parts appear in different sizes and shapes. Increasing the architecture depth is observed as an effective solution. The overfitting problem can be avoided with random weight initialization, dropout function, Ensemble, and Transfer Learning. Different additional datasets containing variation in image resolution and number of objects targeting diverse modalities need to be created. It would help models to handle dense object environments better. Additionally, higher-dimensional datasets would support boosting the efficiency of medical diagnosis. The plateau response of deep learning algorithms for medical image analysis can be enhanced by integrating with classical model-based image segmentation techniques like graph cuts, active contours, and other correlated strategies. A comprehensive study to analyze the dynamics of deep neural architectures by emphasizing both theoretical and technical aspects would enable progress

in semantic image segmentation. It is found that pixel annotation is critical and human-intensive in medical image analysis. The unsupervised and self-supervised learning received great appreciation in capturing minor details even with limited data samples. Besides, reinforcement learning is also an active research direction in health-AI. In medical applications, the accuracy of each segmented pixel is vital. Most of the work reviewed in this article emphasized increasing the accuracy of a model. In some applications, it is crucial to have higher precision and real-time execution with a frame-rate close to at least a typical camera sensors (25 frames/sec). The design of networks with modules like spatial-channel attention and spatial pyramid provides task-specific feature extraction more objectively. Similarly, the atrous convolution uses fewer parameters to capture wide receptive fields (WRF). Methods with dilated convolutions are found with better response. Knowledge distillation and model compression techniques are required to develop memory-efficient models for medical devices with improved performance. Another way to better design model architecture is to automate the loss function search using neural architecture search methods and apply domain-specific knowledge to develop task-specific loss functions. The existing methods are limited to operate with inadequate labeled data for reliable hyperparameter tuning, causing higher variation in performance. The meta-learning attempts to advance an algorithm itself with prior knowledge for magnifying computational response. Efficient pre-processing and post-processing techniques can help enhance data visibility to support networks to learn better representations and improve segmented output maps.

VII. CONCLUSION

This article provides a brief overview of deep convolution neural networks, its application, and metrics commonly used for model evaluation. Besides, this work also discussed state-of-the-art work performed in medical image semantic segmentation. The publically available benchmark datasets are also highlighted. After reviewing the massive amount of domain-centric work, it is found that the image segmentation has a significant contribution to healthcare in recent years. Researchers have produced efficient frameworks with groundbreaking improvements in the performance of the segmentation models. Additionally, the accuracy of segmentation results mainly depends on the structure of a network and the attributes of data. Several challenges directly related to data and model were also pointed out in this article with guided reference points to resolve them.

REFERENCES

[1] J. Enguehard, P. O'Halloran, and A. Gholipour, "Semi-supervised learning with deep embedded clustering for image classification and segmentation," *IEEE Access*, vol. 7, pp. 11093–11104, 2019.

[2] V. Cherukuri, P. Ssenyonga, B. C. Warf, A. V. Kulkarni, V. Monga, and S. J. Schiff, "Learning based segmentation of CT brain images: Application to postoperative hydrocephalic scans," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 8, pp. 1871–1884, Aug. 2018.

[3] H. Fu, J. Cheng, Y. Xu, D. W. K. Wong, J. Liu, and X. Cao, "Joint optic disc and cup segmentation based on multi-label deep network and polar transformation," *IEEE Trans. Med. Imag.*, vol. 37, no. 7, pp. 1597–1605, Jul. 2018.

[4] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2015, pp. 234–241.

[5] T.-H. Song, V. Sanchez, H. EIDaly, and N. M. Rajpoot, "Dual-channel active contour model for megakaryocytic cell segmentation in bone marrow trephine histology images," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 12, pp. 2913–2923, Dec. 2017.

[6] S. Wang, M. Zhou, Z. Liu, Z. Liu, D. Gu, Y. Zang, D. Dong, O. Gevaert, and J. Tian, "Central focused convolutional neural networks: Developing a data-driven model for lung nodule segmentation," *Med. Image Anal.*, vol. 40, pp. 172–183, Aug. 2017.

[7] Y. Onishi, A. Teramoto, M. Tsujimoto, T. Tsukamoto, K. Saito, H. Toyama, K. Imaizumi, and H. Fujita, "Multiplanar analysis for pulmonary nodule classification in CT images using deep convolutional neural network and generative adversarial networks," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 15, no. 1, pp. 173–178, Jan. 2020.

[8] J. Zhang, Y. Shi, J. Sun, L. Wang, L. Zhou, Y. Gao, and D. Shen, "Interactive medical image segmentation via a point-based interaction," *Artif. Intell. Med.*, vol. 111, Jan. 2021, Art. no. 101998.

[9] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 3431–3440.

[10] X.-W. Chen and X. Lin, "Big data deep learning: Challenges and perspectives," *IEEE Access*, vol. 2, pp. 514–525, 2014.

[11] L. Dora, S. Agrawal, R. Panda, and A. Abraham, "State-of-the-art methods for brain tissue segmentation: A review," *IEEE Rev. Biomed. Eng.*, vol. 10, pp. 235–249, 2017.

[12] C. Chen, C. Qin, H. Qiu, G. Tarroni, J. Duan, W. Bai, and D. Rueckert, "Deep learning for cardiac image segmentation: A review," *Frontiers Cardiovascular Med.*, vol. 7, p. 25, Mar. 2020.

[13] X. Li, H. Chen, X. Qi, Q. Dou, C.-W. Fu, and P.-A. Heng, "H-DenseUNet: Hybrid densely connected UNet for liver and tumor segmentation from CT volumes," *IEEE Trans. Med. Imag.*, vol. 37, no. 12, pp. 2663–2674, Dec. 2018.

[14] S. Deng, X. Zhang, W. Yan, I. Eric, C. Chang, Y. Fan, M. Lai, and Y. Xu, "Deep learning in digital pathology image analysis: A survey," *Frontiers Med.*, vol. 14, pp. 1–18, Jul. 2020.

[15] J. Peng and Z. Yuan, "Mitochondria segmentation from EM images via hierarchical structured contextual forest," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 8, pp. 2251–2259, Aug. 2020.

[16] M. Z. Khan and U. Qamar, "Towards service evaluation and ranking model for cloud infrastructure selection," in *Proc. IEEE 12th Int. Conf. Ubiquitous Intell. Comput., IEEE 12th Int. Conf. Auton, Trusted Comput., IEEE 15th Int. Conf. Scalable Comput. Commun. Associated Workshops (UIC-ATC-ScalCom)*, Aug. 2015, pp. 1282–1287.

[17] J. Peng, J. Yi, and Z. Yuan, "Unsupervised mitochondria segmentation in EM images via domain adaptive multi-task learning," *IEEE J. Sel. Topics Signal Process.*, vol. 14, no. 6, pp. 1199–1209, Oct. 2020.

[18] J. J. Cerrolaza, M. L. Picazo, L. Humbert, Y. Sato, D. Rueckert, M. Á. G. Ballester, and M. G. Linguraru, "Computational anatomy for multi-organ analysis in medical imaging: A review," *Med. Image Anal.*, vol. 56, pp. 44–67, Aug. 2019.

[19] C. Han, Y. Duan, X. Tao, and J. Lu, "Dense convolutional networks for semantic segmentation," *IEEE Access*, vol. 7, pp. 43369–43382, 2019.

[20] C. Ye, W. Wang, S. Zhang, and K. Wang, "Multi-depth fusion network for whole-heart CT image segmentation," *IEEE Access*, vol. 7, pp. 23421–23429, 2019.

[21] G. Wang, Y. Zhang, X. Ye, and X. Mou, *Machine Learning for Tomographic Imaging*. Bristol, U.K.: IOP Publishing, 2019.

[22] M. Gaillochet, K. C. Tezcan, and E. Konukoglu, "Joint reconstruction and bias field correction for undersampled MR imaging," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2020, pp. 44–52.

[23] B. E. Dewey, C. Zhao, J. C. Reinhold, A. Carass, K. C. Fitzgerald, E. S. Sotirchos, S. Saidha, J. Oh, D. L. Pham, P. A. Calabresi, P. C. M. van Zijl, and J. L. Prince, "DeepHarmony: A deep learning approach to contrast harmonization across scanner changes," *Magn. Reson. Imag.*, vol. 64, pp. 160–170, Dec. 2019.

[24] N. Tajbakhsh, L. Jeyaseelan, Q. Li, J. N. Chiang, Z. Wu, and X. Ding, "Embracing imperfect datasets: A review of deep learning solutions for medical image segmentation," *Med. Image Anal.*, vol. 63, Jul. 2020, Art. no. 101693.

[25] Y. Fu, Y. Lei, T. Wang, W. J. Curran, T. Liu, and X. Yang, "Deep learning in medical image registration: A review," *Phys. Med. Biol.*, vol. 65, no. 20, 2020, Art. no. 20TR01.

[26] H.-P. Chan, L. M. Hadjiiski, and R. K. Samala, "Computer-aided diagnosis in the era of deep learning," *Med. Phys.*, vol. 47, no. 5, pp. e218–e227, 2020.

[27] D. Liu, K. S. Zhou, D. Bernhardt, and D. Comaniciu, "Search strategies for multiple landmark detection by submodular maximization," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 2831–2838.

[28] Z. Xu, Y. Huo, J. Park, B. Landman, A. Milkowski, S. Grbic, and S. Zhou, "Less is more: Simultaneous view classification and landmark detection for abdominal ultrasound images," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2018, pp. 711–719.

[29] B. Jing, P. Xie, and E. Xing, "On the automatic generation of medical imaging reports," in *Proc. 56th Annu. Meeting Assoc. Comput. Linguistics*, 2018, pp. 2577–2586.

[30] M. Rezaei, H. Yang, and C. Meinel, "Recurrent generative adversarial network for learning imbalanced medical image semantic segmentation," *Multimedia Tools Appl.*, vol. 79, nos. 21–22, pp. 15329–15348, Jun. 2020.

[31] P. F. Jaeger, S. A. Kohl, S. Bickelhaupt, F. Isensee, T. A. Kuder, H.-P. Schlemmer, and K. H. Maier-Hein, "Retina U-Net: Embarrassingly simple exploitation of segmentation supervision for medical object detection," in *Proc. Mach. Learn. Health Workshop*, 2020, pp. 171–183.

[32] R. Li, C. Duan, S. Zheng, C. Zhang, and P. M. Atkinson, "MACU-Net for semantic segmentation of fine-resolution remotely sensed images," 2020, *arXiv:2007.13083*. [Online]. Available: http://arxiv.org/abs/2007.13083

[33] C. Ouyang, C. Biffi, C. Chen, T. Kart, H. Qiu, and D. Rueckert, "Self-supervision with superpixels: Training few-shot medical image segmentation without annotation," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2020, pp. 762–780.

[34] L. Huang, M. He, C. Tan, D. Jiang, G. Li, and H. Yu, "Jointly network image processing: Multi-task image semantic segmentation of indoor scene based on CNN," *IET Image Process.*, vol. 14, no. 15, pp. 3689–3697, Dec. 2020.

[35] Y. Guo, Y. Liu, T. Georgiou, and M. S. Lew, "A review of semantic segmentation using deep neural networks," *Int. J. Multimedia Inf. Retr.*, vol. 7, no. 2, pp. 87–93, Jun. 2018.

[36] Q. Geng, Z. Zhou, and X. Cao, "Survey of recent progress in semantic image segmentation with CNNs," *Sci. China Inf. Sci.*, vol. 61, no. 5, May 2018, Art. no. 051101.

[37] A. Garcia-Garcia, S. Orts-Escolano, S. Oprea, V. Villena-Martinez, P. Martinez-Gonzalez, and J. Garcia-Rodriguez, "A survey on deep learning techniques for image and video semantic segmentation," *Appl. Soft Comput.*, vol. 70, pp. 41–65, Sep. 2018.

[38] M. H. Hesamian, W. Jia, X. He, and P. Kennedy, "Deep learning techniques for medical image segmentation: Achievements and challenges," *J. Digit. Imag.*, vol. 32, no. 4, pp. 582–596, Aug. 2019.

[39] D. Karimi, H. Dou, S. K. Warfield, and A. Gholipour, "Deep learning with noisy labels: Exploring techniques and remedies in medical image analysis," *Med. Image Anal.*, vol. 65, Oct. 2020, Art. no. 101759.

[40] X. Li, W. Zhang, Q. Ding, and X. Li, "Diagnosing rotating machines with weakly supervised data using deep transfer learning," *IEEE Trans. Ind. Informat.*, vol. 16, no. 3, pp. 1688–1697, Mar. 2020.

[41] 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," *Med. image Anal.*, vol. 42, pp. 60–88, Dec. 2017.

[42] S. A. Taghanaki, K. Abhishek, J. P. Cohen, J. Cohen-Adad, and G. Hamarneh, "Deep semantic segmentation of natural and medical images: A review," *Artif. Intell. Rev.*, vol. 54, no. 1, pp. 137–178, Jan. 2021.

[43] S. K. Zhou, H. Greenspan, C. Davatzikos, J. S. Duncan, B. Van Ginneken, A. Madabhushi, J. L. Prince, D. Rueckert, and R. M. Summers, "A review of deep learning in medical imaging: Imaging traits, technology trends, case studies with progress highlights, and future promises," *Proc. IEEE*, vol. 109, no. 5, pp. 820–838, May 2021.

[44] X. Yi, E. Walia, and P. Babyn, "Generative adversarial network in medical imaging: A review," *Med. Image Anal.*, vol. 58, Dec. 2019, Art. no. 101552.

[45] F. Shi, J. Wang, J. Shi, Z. Wu, Q. Wang, Z. Tang, K. He, Y. Shi, and D. Shen, "Review of artificial intelligence techniques in imaging data acquisition, segmentation, and diagnosis for COVID-19," *IEEE Rev. Biomed. Eng.*, vol. 14, pp. 4–15, 2021.

[46] T. Heimann *et al.*, "Comparison and evaluation of methods for liver segmentation from CT datasets," *IEEE Trans. Med. Imag.*, vol. 28, no. 8, pp. 1251–1265, Aug. 2009.

[47] S. Minaee, Y. Y. Boykov, F. Porikli, A. J. Plaza, N. Kehtarnavaz, and D. Terzopoulos, "Image segmentation using deep learning: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, early access, Feb. 17, 2021, doi: 10.1109/TPAMI.2021.3059968.

[48] J. Peng and Y. Wang, "Medical image segmentation with limited supervision: A review of deep network models," *IEEE Access*, vol. 9, pp. 36827–36851, 2021.

[49] S. K. Zhou, H. Greenspan, C. Davatzikos, J. S. Duncan, B. Van Ginneken, A. Madabhushi, J. L. Prince, D. Rueckert, and R. M. Summers, "A review of deep learning in medical imaging: Imaging traits, technology trends, case studies with progress highlights, and future promises," *Proc. IEEE*, vol. 109, no. 5, pp. 820–838, May 2021.

[50] K. Fukushima and S. Miyake, "Neocognitron: A self-organizing neural network model for a mechanism of visual pattern recognition," in *Competition and Cooperation in Neural Nets*. Berlin, Germany: Springer, 1982, pp. 267–285.

[51] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017.

[52] S. Moccia, E. De Momi, S. El Hadji, and L. S. Mattos, "Blood vessel segmentation algorithms—Review of methods, datasets and evaluation metrics," *Comput. Methods Programs Biomed.*, vol. 158, pp. 71–91, May 2018.

[53] S. A. Taghanaki, K. Abhishek, J. P. Cohen, J. Cohen-Adad, and G. Hamarneh, "Deep semantic segmentation of natural and medical images: A review," *Artif. Intell. Rev.*, vol. 54, pp. 1–42, Jun. 2020.

[54] N. Kalchbrenner, E. Grefenstette, P. Blunsom, D. Kartsaklis, N. Kalchbrenner, M. Sadrzadeh, N. Kalchbrenner, P. Blunsom, N. Kalchbrenner, and P. Blunsom, "A convolutional neural network for modelling sentences," in *Proc. 52nd Annu. Meeting Assoc. Comput. Linguistics*. Stroudsburg, PA, USA: Association for Computational Linguistics, 2014, pp. 212–217.

[55] S. Mohapatra, T. Swarnkar, and J. Das, "Deep convolutional neural network in medical image processing," in *Handbook of Deep Learning in Biomedical Engineering*. Amsterdam, The Netherlands: Elsevier, 2021, pp. 25–60.

[56] M. Umer, I. Ashraf, S. Ullah, A. Mehmood, and G. S. Choi, "COVINet: A convolutional neural network approach for predicting COVID-19 from chest X-ray images," *J. Ambient Intell. Humanized Comput.*, pp. 1–13, Jan. 2021.

[57] K. Cho, B. van Merrienboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using RNN encoder–decoder for statistical machine translation," in *Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP)*, 2014, pp. 1724–1734.

[58] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[59] A. Graves and J. Schmidhuber, "Framewise phoneme classification with bidirectional LSTM and other neural network architectures," *Neural Netw.*, vol. 18, nos. 5–6, pp. 602–610, Jul. 2005.

[60] M. Sundermeyer, R. Schlüter, and H. Ney, "LSTM neural networks for language modeling," in *Proc. 13th Annu. Conf. Int. Speech Commun. Assoc.*, 2012, pp. 194–197.

[61] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, *Deep Learning*, vol. 1, no. 2. Cambridge, MA, USA: MIT Press, 2016.

[62] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.

[63] S. K. Zhou, H. Greenspan, C. Davatzikos, J. S. Duncan, B. van Ginneken, A. Madabhushi, J. L. Prince, D. Rueckert, and R. M. Summers, "A review of deep learning in medical imaging: Imaging traits, technology trends, case studies with progress highlights, and future promises," 2020, *arXiv:2008.09104*. [Online]. Available: http://arxiv.org/abs/2008.09104

[64] F. Altaf, S. M. S. Islam, N. Akhtar, and N. K. Janjua, "Going deep in medical image analysis: Concepts, methods, challenges, and future directions," *IEEE Access*, vol. 7, pp. 99540–99572, 2019.

[65] C. L. Srinidhi, O. Ciga, and A. L. Martel, "Deep neural network models for computational histopathology: A survey," *Med. Image Anal.*, vol. 67, Jan. 2021, Art. no. 101813.

[66] X. Xie, J. Niu, X. Liu, Z. Chen, S. Tang, and S. Yu, "A survey on incorporating domain knowledge into deep learning for medical image analysis," *Med. Image Anal.*, vol. 69, Apr. 2021, Art. no. 101985.

[67] W. Anjali, B. Anuj, and V. S. Verma, "A review on brain tumor segmentation of MRI images," *Magn. Reson. Imag.*, vol. 61, pp. 247–259, Sep. 2019.

[68] X. Du, S. Yin, R. Tang, Y. Zhang, and S. Li, "Cardiac-DeepIED: Automatic pixel-level deep segmentation for cardiac bi-ventricle using improved end-to-end encoder-decoder network," *IEEE J. Transl. Eng. Health Med.*, vol. 7, pp. 1–10, 2019.

[69] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in Neural Information Processing Systems*, vol. 27. Red Hook, NY, USA: Curran Associates, 2014, pp. 2672–2680.

[70] P. Zhang, Y. Zhong, Y. Deng, X. Tang, and X. Li, "A survey on deep learning of small sample in biomedical image analysis," 2019, *arXiv:1908.00473*. [Online]. Available: http://arxiv.org/abs/1908.00473

[71] S. Vallecorsa, F. Carminati, and G. Khattak, "3D convolutional GAN for fast simulation," in *Proc. EPJ Web Conf.*, vol. 214, 2019, p. 02010.

[72] B. Dai, S. Fidler, R. Urtasun, and D. Lin, "Towards diverse and natural image descriptions via a conditional GAN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2970–2979.

[73] Y. Cao, B. Liu, M. Long, and J. Wang, "HashGAN: Deep learning to hash with pair conditional wasserstein GAN," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 1287–1296.

[74] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, Apr. 2018.

[75] S. Hao, Y. Zhou, and Y. Guo, "A brief survey on semantic segmentation with deep learning," *Neurocomputing*, vol. 406, pp. 302–321, Sep. 2020.

[76] S. Sedai, B. Antony, R. Rai, K. Jones, H. Ishikawa, J. Schuman, W. Gadi, and R. Garnavi, "Uncertainty guided semi-supervised segmentation of retinal layers in OCT images," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2019, pp. 282–290.

[77] J. Han and C. Moraga, "The influence of the sigmoid function parameters on the speed of backpropagation learning," in *Proc. Int. Workshop Artif. Neural Netw.* Berlin, Germany: Springer, 1995, pp. 195–201.

[78] L. Yu, S. Wang, X. Li, C.-W. Fu, and P.-A. Heng, "Uncertainty-aware self-ensembling model for semi-supervised 3D left atrium segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2019, pp. 605–613.

[79] Y. Li, J. Chen, X. Xie, K. Ma, and Y. Zheng, "Self-loop uncertainty: A novel pseudo-label for semi-supervised medical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2020, pp. 614–623.

[80] G. Bortsova, F. Dubost, L. Hogeweg, I. Katramados, and M. de Bruijne, "Semi-supervised medical image segmentation via learning consistency under transformations," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2019, pp. 810–818.

[81] H. Shan, Y. Zhang, Q. Yang, U. Kruger, M. K. Kalra, L. Sun, W. Cong, and G. Wang, "3-D convolutional encoder-decoder network for low-dose CT via transfer learning from a 2-D trained network," *IEEE Trans. Med. Imag.*, vol. 37, no. 6, pp. 1522–1534, Jun. 2018.

[82] S. Liu, D. Xu, S. K. Zhou, O. Pauly, S. Grbic, T. Mertelmeier, J. Wicklein, A. Jerebko, W. Cai, and D. Comaniciu, "3D anisotropic hybrid network: Transferring convolutional features from 2D images to 3D anisotropic volumes," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2018, pp. 851–858.

[83] M. Raghu, C. Zhang, J. Kleinberg, and S. Bengio, "Transfusion: Understanding transfer learning for medical imaging," 2019, *arXiv:1902.07208*. [Online]. Available: http://arxiv.org/abs/1902.07208

[84] B. Zoph, G. Ghiasi, T.-Y. Lin, Y. Cui, H. Liu, E. D. Cubuk, and Q. Le, "Rethinking pre-training and self-training," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 33, 2020, pp. 1–20.

[85] Z. Zhou, V. Sodha, M. M. R. Siddiquee, R. Feng, N. Tajbakhsh, M. B. Gotway, and J. Liang, "Models genesis: Generic autodidactic models for 3D medical image analysis," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2019, pp. 384–393.

[86] R. Volpi, H. Namkoong, O. Sener, J. Duchi, V. Murino, and S. Savarese, "Generalizing to unseen domains via adversarial data augmentation," 2018, *arXiv:1805.12018*. [Online]. Available: http://arxiv.org/abs/1805.12018

[87] P. Ramachandran, B. Zoph, and Q. V. Le, "Swish: A self-gated activation function," 2017, *arXiv:1710.05941*. [Online]. Available: https://arxiv.org/abs/1710.05941

[88] M. D. Zeiler and R. Fergus, "Stochastic pooling for regularization of deep convolutional neural networks: 1st international conference on learning representations," in *Proc. 1st Int. Conf. Learn. Represent. (ICLR)*, 2013, pp. 1–9.

[89] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 9, pp. 1904–1916, Sep. 2015.

[90] Y. Gong, L. Wang, R. Guo, and S. Lazebnik, "Multi-scale orderless pooling of deep convolutional activation features," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2014, pp. 392–407.

[91] O. Rippel, J. Snoek, and R. P. Adams, "Spectral representations for convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 2449–2457.

[92] J. T. Springenberg, A. Dosovitskiy, T. Brox, and M. Riedmiller, "Striving for simplicity: The all convolutional net," 2014, *arXiv:1412.6806*. [Online]. Available: http://arxiv.org/abs/1412.6806

[93] H. Guo, Y. Mao, and R. Zhang, "Mixup as locally linear out-of-manifold regularization," in *Proc. AAAI Conf. Artif. Intell.*, 2019, vol. 33, no. 1, pp. 3714–3722.

[94] E. Panfilov, A. Tiulpin, S. Klein, M. T. Nieminen, and S. Saarakkala, "Improving robustness of deep learning based knee MRI segmentation: Mixup and adversarial domain adaptation," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshop (ICCVW)*, Oct. 2019, pp. 450–459.

[95] J. M. J. Valanarasu, R. Yasarla, P. Wang, I. Hacihaliloglu, and V. M. Patel, "Learning to segment brain anatomy from 2D ultrasound with less data," *IEEE J. Sel. Topics Signal Process.*, vol. 14, no. 6, pp. 1221–1234, Oct. 2020.

[96] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 448–456.

[97] Z. Yuan, J. Yi, Z. Luo, Z. Jia, and J. Peng, "EM-Net: Centerline-aware mitochondria segmentation in EM images via hierarchical view-ensemble convolutional network," in *Proc. IEEE 17th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2020, pp. 1219–1222.

[98] A. Zhao, G. Balakrishnan, F. Durand, J. V. Guttag, and A. V. Dalca, "Data augmentation using learned transformations for one-shot medical image segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 8543–8553.

[99] C. Chen, C. Qin, H. Qiu, C. Ouyang, S. Wang, L. Chen, G. Tarroni, W. Bai, and D. Rueckert, "Realistic adversarial data augmentation for MR image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2020, pp. 667–677.

[100] I. Oksuz, B. Ruijsink, E. Puyol-Antón, J. R. Clough, G. Cruz, A. Bustin, C. Prieto, R. Botnar, D. Rueckert, J. A. Schnabel, and A. P. King, "Automatic CNN-based detection of cardiac MR motion artefacts using k-space data augmentation and curriculum learning," *Med. Image Anal.*, vol. 55, pp. 136–147, Jul. 2019.

[101] D. E. Schlesinger and C. M. Stultz, "Deep learning for cardiovascular risk stratification," *Current Treatment Options Cardiovascular Med.*, vol. 22, no. 8, pp. 1–14, Aug. 2020.

[102] V. Verma, A. Lamb, C. Beckham, A. Najafi, I. Mitliagkas, D. Lopez-Paz, and Y. Bengio, "Manifold mixup: Better representations by interpolating hidden states," in *Proc. Int. Conf. Mach. Learn.*, 2019, pp. 6438–6447.

[103] S. Olut, Z. Shen, Z. Xu, S. Gerber, and M. Niethammer, "Adversarial data augmentation via deformation statistics," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2020, pp. 643–659.

[104] Z. Shen, Z. Xu, S. Olut, and M. Niethammer, "Anatomical data augmentation via fluid-based image registration," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2020, pp. 318–328.

[105] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov, "Improving neural networks by preventing co-adaptation of feature detectors," 2012, *arXiv:1207.0580*. [Online]. Available: http://arxiv.org/abs/1207.0580

[106] L. Wan, M. Zeiler, S. Zhang, Y. Le Cun, and R. Fergus, "Regularization of neural networks using dropconnect," in *Proc. Int. Conf. Mach. Learn.*, 2013, pp. 1058–1066.

[107] M. Abadi *et al.*, "TensorFlow: Large-scale machine learning on heterogeneous distributed systems," 2016, *arXiv:1603.04467*. [Online]. Available: http://arxiv.org/abs/1603.04467

[108] A. Gulli and S. Pal, *Deep Learning With Keras*. Birmingham, U.K.: Packt, 2017.

[109] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer, "Automatic differentiation in PyTorch," in *Proc. NIPS*, 2017, pp. 1–4.

[110] P. Costa, A. Galdran, M. I. Meyer, M. Niemeijer, M. Abramoff, A. M. Mendonca, and A. Campilho, "End-to-end adversarial retinal image synthesis," *IEEE Trans. Med. Imag.*, vol. 37, no. 3, pp. 781–791, Mar. 2018.

[111] K. Chaitanya, N. Karani, C. F. Baumgartner, A. Becker, O. Donati, and E. Konukoglu, "Semi-supervised and task-driven data augmentation," in *Proc. Int. Conf. Inf. Process. Med. Imag.* Cham, Switzerland: Springer, 2019, pp. 29–41.

[112] J. Cai, Z. Zhang, L. Cui, Y. Zheng, and L. Yang, "Towards cross-modal organ translation and segmentation: A cycle- and shape-consistent generative adversarial network," *Med. Image Anal.*, vol. 52, pp. 174–184, Feb. 2019.

[113] M. Ghafoorian, A. Mehrtash, T. Kapur, N. Karssemeijer, E. Marchiori, M. Pesteie, C. R. G. Guttmann, F.-E. de Leeuw, C. M. Tempany, B. van Ginneken, A. Fedorov, P. Abolmaesumi, B. Platel, and W. M. Wells, III, "Transfer learning for domain adaptation in MRI: Application in brain lesion segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2017, pp. 516–524.

[114] E. Smistad, T. L. Falch, M. Bozorgi, A. C. Elster, and F. Lindseth, "Medical image segmentation on GPUs—A comprehensive review," *Med. Image Anal.*, vol. 20, no. 1, pp. 1–18, Feb. 2015.

[115] B. Yu, L. Zhou, L. Wang, Y. Shi, J. Fripp, and P. Bourgeat, "Ea-GANs: Edge-aware generative adversarial networks for cross-modality MR image synthesis," *IEEE Trans. Med. Imag.*, vol. 38, no. 7, pp. 1750–1762, Jul. 2019.

[116] Y. Qin, H. Zheng, X. Huang, J. Yang, and Y.-M. Zhu, "Pulmonary nodule segmentation with CT sample synthesis using adversarial networks," *Med. Phys.*, vol. 46, no. 3, pp. 1218–1229, Mar. 2019.

[117] S. Wang, S. Cao, D. Wei, R. Wang, K. Ma, L. Wang, D. Meng, and Y. Zheng, "LT-Net: Label transfer by learning reversible voxel-wise correspondence for one-shot medical image segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 9162–9171.

[118] Z. Xu and M. Niethammer, "DeepAtlas: Joint semi-supervised learning of image registration and segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2019, pp. 420–429.

[119] E. J. Rosana, C. Petitjean, P. Honeine, and F. Abdallah, "BB-UNet: U-Net with bounding box prior," *IEEE J. Sel. Topics Signal Process.*, vol. 14, no. 6, pp. 1189–1198, Oct. 2020.

[120] O. Oktay, E. Ferrante, K. Kamnitsas, M. Heinrich, W. Bai, J. Caballero, S. A. Cook, A. de Marvao, T. Dawes, D. P. O'Regan, B. Kainz, B. Glocker, and D. Rueckert, "Anatomically constrained neural networks (ACNNs): Application to cardiac image enhancement and segmentation," *IEEE Trans. Med. Imag.*, vol. 37, no. 2, pp. 384–395, Feb. 2018.

[121] A. V. Dalca, J. Guttag, and M. R. Sabuncu, "Anatomical priors in convolutional networks for unsupervised biomedical segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 9290–9299.

[122] A. J. Larrazabal, C. Martinez, and E. Ferrante, "Anatomical priors for image segmentation via post-processing with denoising autoencoders," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2019, pp. 585–593.

[123] R. Ito, K. Nakae, J. Hata, H. Okano, and S. Ishii, "Semi-supervised deep learning of brain tissue segmentation," *Neural Netw.*, vol. 116, pp. 25–34, Aug. 2019.

[124] B. D. de Vos, F. F. Berendsen, M. A. Viergever, H. Sokooti, M. Staring, and I. Išgum, "A deep learning framework for unsupervised affine and deformable image registration," *Med. Image Anal.*, vol. 52, pp. 128–143, Feb. 2019.

[125] W. Chi, L. Ma, J. Wu, M. Chen, W. Lu, and X. Gu, "Deep learning-based medical image segmentation with limited labels," *Phys. Med. Biol.*, vol. 65, no. 23, Dec. 2020, Art. no. 235001.

[126] Y. He, G. Yang, Y. Chen, Y. Kong, J. Wu, L. Tang, X. Zhu, J.-L. Dillenseger, P. Shao, S. Zhang, H. Shu, J.-L. Coatrieux, and S. Li, "DPA-DenseBiasNet: Semi-supervised 3D fine renal artery segmentation with dense biased network and deep priori anatomy," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2019, pp. 139–147.

[127] S. Dong, G. Luo, C. Tam, W. Wang, K. Wang, S. Cao, B. Chen, H. Zhang, and S. Li, "Deep atlas network for efficient 3D left ventricle segmentation on echocardiography," *Med. Image Anal.*, vol. 61, Apr. 2020, Art. no. 101634.

[128] H. Zheng, L. Lin, H. Hu, Q. Zhang, Q. Chen, Y. Iwamoto, X. Han, Y.-W. Chen, R. Tong, and J. Wu, "Semi-supervised segmentation of liver using adversarial learning with deep atlas prior," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2019, pp. 148–156.

[129] M. Vakalopoulou, G. Chassagnon, N. Bus, R. Marini, E. I. Zacharaki, M.-P. Revel, and N. Paragios, "AtlasNet: Multi-atlas non-linear deep networks for medical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2018, pp. 658–666.

[130] A. J. Larrazabal, C. Martínez, B. Glocker, and E. Ferrante, "Post-DAE: Anatomically plausible segmentation via post-processing with denoising autoencoders," *IEEE Trans. Med. Imag.*, vol. 39, no. 12, pp. 3813–3820, Dec. 2020.

[131] A. D. Hoover, V. Kouznetsova, and M. Goldbaum, "Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response," *IEEE Trans. Med. Imag.*, vol. 19, no. 3, pp. 203–210, Mar. 2000.

[132] P. Khojasteh, B. Aliahmad, and D. K. Kumar, "Fundus images analysis using deep features for detection of exudates, hemorrhages and microaneurysms," *BMC Ophthalmol.*, vol. 18, no. 1, pp. 1–13, Dec. 2018.

[133] S. Y. Shin, S. Lee, I. D. Yun, and K. M. Lee, "Deep vessel segmentation by learning graphical connectivity," *Med. Image Anal.*, vol. 58, Dec. 2019, Art. no. 101556.

[134] Q. Yu, L. Xie, Y. Wang, Y. Zhou, E. K. Fishman, and A. L. Yuille, "Recurrent saliency transformation network: Incorporating multi-stage visual cues for small organ segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 8280–8289.

[135] F. Haghighi, M. R. H. Taher, Z. Zhou, M. B. Gotway, and J. Liang, "Learning semantics-enriched representation via self-discovery, self-classification, and self-restoration," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2020, pp. 137–147.

[136] Q. Jin, Z. Meng, T. D. Pham, Q. Chen, L. Wei, and R. Su, "DUNet: A deformable network for retinal vessel segmentation," *Knowl.-Based Syst.*, vol. 178, pp. 149–162, Aug. 2019.

[137] A. Vakanski, M. Xian, and P. E. Freer, "Attention-enriched deep learning model for breast tumor segmentation in ultrasound images," *Ultrasound Med. Biol.*, vol. 46, no. 10, pp. 2819–2833, Oct. 2020.

[138] D. Jha, M. A. Riegler, D. Johansen, P. Halvorsen, and H. D. Johansen, "DoubleU-Net: A deep convolutional neural network for medical image segmentation," 2020, *arXiv:2006.04868*. [Online]. Available: http://arxiv.org/abs/2006.04868

[139] G. Wang, W. Li, S. Ourselin, and T. Vercauteren, "Automatic brain tumor segmentation using cascaded anisotropic convolutional neural networks," in *Proc. Int. MICCAI Brainlesion Workshop.* Cham, Switzerland: Springer, 2017, pp. 178–190.

[140] Y. Li, H. Qi, J. Dai, X. Ji, and Y. Wei, "Fully convolutional instance-aware semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2359–2367.

[141] Y. Yuan, "Automatic skin lesion segmentation with fully convolutional-deconvolutional networks," 2017, *arXiv:1703.05165*. [Online]. Available: http://arxiv.org/abs/1703.05165

[142] N. Liu, H. Li, M. Zhang, J. Liu, Z. Sun, and T. Tan, "Accurate iris segmentation in non-cooperative environments using fully convolutional networks," in *Proc. Int. Conf. Biometrics (ICB)*, Jun. 2016, pp. 1–8.

[143] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-Net: Fully convolutional neural networks for volumetric medical image segmentation," in *Proc. 4th Int. Conf. 3D Vis. (3DV)*, Oct. 2016, pp. 565–571.

[144] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-Net: Learning dense volumetric segmentation from sparse annotation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2016, pp. 424–432.

[145] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: Redesigning skip connections to exploit multiscale features in image segmentation," *IEEE Trans. Med. Imag.*, vol. 39, no. 6, pp. 1856–1867, Jun. 2020.

[146] L. Teng, H. Li, and S. Karim, "DMCNN: A deep multiscale convolutional neural network model for medical image segmentation," *J. Healthcare Eng.*, vol. 2019, pp. 1–10, Dec. 2019.

[147] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2881–2890.

[148] H. R. Roth, H. Oda, X. Zhou, N. Shimizu, Y. Yang, Y. Hayashi, M. Oda, M. Fujiwara, K. Misawa, and K. Mori, "An application of cascaded 3D fully convolutional networks for medical image segmentation," *Comput. Med. Imag. Graph.*, vol. 66, pp. 90–99, Jun. 2018.

[149] S. Hussain, S. M. Anwar, and M. Majid, "Brain tumor segmentation using cascaded deep convolutional neural network," in *Proc. 39th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2017, pp. 1998–2001.

[150] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," 2017, *arXiv:1706.05587*. [Online]. Available: http://arxiv.org/abs/1706.05587

[151] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.

[152] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2017, pp. 2961–2969.

[153] Y. Lee and J. Park, "CenterMask: Real-time anchor-free instance segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 13906–13915.

[154] K.-B. Park, S. H. Choi, and J. Y. Lee, "M-GAN: Retinal blood vessel segmentation by balancing losses through stacked deep fully convolutional networks," *IEEE Access*, vol. 8, pp. 146308–146322, 2020.

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

[156] A. Fakhry, T. Zeng, and S. Ji, "Residual deconvolutional networks for brain electron microscopy image segmentation," *IEEE Trans. Med. Imag.*, vol. 36, no. 2, pp. 447–456, Feb. 2017.

[157] L. Chen, P. Bentley, K. Mori, K. Misawa, M. Fujiwara, and D. Rueckert, "DRINet for medical image segmentation," *IEEE Trans. Med. Imag.*, vol. 37, no. 11, pp. 2453–2462, Nov. 2018.

[158] M. Drozdzal, G. Chartrand, E. Vorontsov, M. Shakeri, L. Di Jorio, A. Tang, A. Romero, Y. Bengio, C. Pal, and S. Kadoury, "Learning normalized inputs for iterative estimation in medical image segmentation," *Med. Image Anal.*, vol. 44, pp. 1–13, Feb. 2018.

[159] M. Z. Alom, M. Hasan, C. Yakopcic, T. M. Taha, and V. K. Asari, "Recurrent residual convolutional neural network based on U-Net (R2U-Net) for medical image segmentation," 2018, *arXiv:1802.06955*. [Online]. Available: http://arxiv.org/abs/1802.06955

[160] H. Chen, Q. Dou, L. Yu, J. Qin, and P.-A. Heng, "VoxResNet: Deep voxelwise residual networks for brain segmentation from 3D MR images," *NeuroImage*, vol. 170, pp. 446–455, Apr. 2018.

[161] L. Liu, S. Chen, F. Zhang, F.-X. Wu, Y. Pan, and J. Wang, "Deep convolutional neural network for automatically segmenting acute ischemic stroke lesion in multi-modality MRI," *Neural Comput. Appl.*, vol. 32, pp. 1–14, Feb. 2019.

[162] N. Ibtehaz and M. S. Rahman, "MultiResUNet: Rethinking the U-Net architecture for multimodal biomedical image segmentation," *Neural Netw.*, vol. 121, pp. 74–87, Jan. 2020.

[163] A. Clèrigues, S. Valverde, J. Bernal, J. Freixenet, A. Oliver, and X. Lladó, "SuNet: A deep learning architecture for acute stroke lesion segmentation and outcome prediction in multimodal MRI," 2018, *arXiv:1810.13304*. [Online]. Available: https://arxiv.org/abs/1810.13304

[164] R. Guerrero, C. Qin, O. Oktay, C. Bowles, L. Chen, R. Joules, R. Wolz, M. C. Valdés-Hernández, D. A. Dickie, J. Wardlaw, and D. Rueckert, "White matter hyperintensity and stroke lesion segmentation and differentiation using convolutional neural networks," *NeuroImage, Clin.*, vol. 17, pp. 918–934, 2018.

[165] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4700–4708.

[166] L. Liu, F.-X. Wu, and J. Wang, "Efficient multi-kernel DCNN with pixel dropout for stroke MRI segmentation," *Neurocomputing*, vol. 350, pp. 117–127, Jul. 2019.

[167] J. Zhang, Y. Jin, J. Xu, X. Xu, and Y. Zhang, "MDU-Net: Multi-scale densely connected U-Net for biomedical image segmentation," 2018, *arXiv:1812.00352*. [Online]. Available: http://arxiv.org/abs/1812.00352

[168] M. Shaikh, G. Anand, G. Acharya, A. Amrutkar, V. Alex, and G. Krishnamurthi, "Brain tumor segmentation using dense fully convolutional neural network," in *Proc. Int. MICCAI Brainlesion Workshop*. Cham, Switzerland: Springer, 2017, pp. 309–319.

[169] A. Tureckova and A. J. Rodríguez-Sánchez, "ISLES challenge: U-shaped convolution neural network with dilated convolution for 3D stroke lesion segmentation," in *Proc. Int. MICCAI Brainlesion Workshop*. Cham, Switzerland: Springer, 2018, pp. 319–327.

[170] S. Vesal, N. Ravikumar, and A. Maier, "Dilated convolutions in neural networks for left atrial segmentation in 3D gadolinium enhanced-MRI," in *Proc. Int. Workshop Stat. Atlases Comput. Models Heart*. Cham, Switzerland: Springer, 2018, pp. 319–328.

[171] D. Nie, L. Wang, Y. Gao, J. Lian, and D. Shen, "STRAINet: Spatially varying sTochastic residual AdversarIal networks for MRI pelvic organ segmentation," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 5, pp. 1552–1564, May 2019.

[172] H. Li, A. Zhygallo, and B. Menze, "Automatic brain structures segmentation using deep residual dilated U-Net," in *Proc. Int. MICCAI Brainlesion Workshop*. Cham, Switzerland: Springer, 2018, pp. 385–393.

[173] F. Isensee, P. Kickingereder, W. Wick, M. Bendszus, and K. H. Maier-Hein, "Brain tumor segmentation and radiomics survival prediction: Contribution to the brats 2017 challenge," in *Proc. Int. MICCAI Brainlesion Workshop*. Cham, Switzerland: Springer, 2017, pp. 287–297.

[174] H. Chen, X. Qi, L. Yu, Q. Dou, J. Qin, and P.-A. Heng, "DCAN: Deep contour-aware networks for object instance segmentation from histology images," *Med. Image Anal.*, vol. 36, pp. 135–146, Feb. 2017.

[175] J. Dolz, C. Desrosiers, and I. B. Ayed, "3D fully convolutional networks for subcortical segmentation in MRI: A large-scale study," *NeuroImage*, vol. 170, pp. 456–470, Apr. 2018.

[176] V. Mnih, N. Heess, A. Graves, and K. Kavukcuoglu, "Recurrent models of visual attention," 2014, *arXiv:1406.6247*. [Online]. Available: http://arxiv.org/abs/1406.6247

[177] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," 2014, *arXiv:1409.0473*. [Online]. Available: http://arxiv.org/abs/1409.0473

[178] K. M. Hermann, T. Kočiský, E. Grefenstette, L. Espeholt, W. Kay, M. Suleyman, and P. Blunsom, "Teaching machines to read and comprehend," 2015, *arXiv:1506.03340*. [Online]. Available: http://arxiv.org/abs/1506.03340

[179] J. Chorowski, D. Bahdanau, D. Serdyuk, K. Cho, and Y. Bengio, "Attention-based models for speech recognition," 2015, *arXiv:1506.07503*. [Online]. Available: http://arxiv.org/abs/1506.07503

[180] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," 2017, *arXiv:1706.03762*. [Online]. Available: http://arxiv.org/abs/1706.03762

[181] H. Salehinejad, S. Sankar, J. Barfett, E. Colak, and S. Valaee, "Recent advances in recurrent neural networks," 2017, *arXiv:1801.01078*. [Online]. Available: http://arxiv.org/abs/1801.01078

[182] R. Li, M. Li, J. Li, and Y. Zhou, "Connection sensitive attention U-Net for accurate retinal vessel segmentation," 2019, *arXiv:1903.05558*. [Online]. Available: http://arxiv.org/abs/1903.05558

[183] O. Oktay, J. Schlemper, L. Le Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz, B. Glocker, and D. Rueckert, "Attention U-Net: Learning where to look for the pancreas," 2018, *arXiv:1804.03999*. [Online]. Available: http://arxiv.org/abs/1804.03999

[184] D. Opitz and R. Maclin, "Popular ensemble methods: An empirical study," *J. Artif. Intell. Res.*, vol. 11, pp. 169–198, Aug. 1999.

[185] H. Li, G. Jiang, J. Zhang, R. Wang, Z. Wang, W.-S. Zheng, and B. Menze, "Fully convolutional network ensembles for white matter hyperintensities segmentation in MR images," *NeuroImage*, vol. 183, pp. 650–665, Dec. 2018.

[186] K. Kamnitsas, W. Bai, E. Ferrante, S. McDonagh, M. Sinclair, N. Pawlowski, M. Rajchl, M. Lee, B. Kainz, D. Rueckert, and B. Glocker, "Ensembles of multiple models and architectures for robust brain tumour segmentation," in *Proc. Int. MICCAI Brainlesion Workshop*. Cham, Switzerland: Springer, 2017, pp. 450–462.

[187] Y. Han and J. C. Ye, "Framing U-Net via deep convolutional framelets: Application to sparse-view CT," *IEEE Trans. Med. Imag.*, vol. 37, no. 6, pp. 1418–1429, Jun. 2018.

[188] X. Feng, N. J. Tustison, S. H. Patel, and C. H. Meyer, "Brain tumor segmentation using an ensemble of 3D U-Nets and overall survival prediction using radiomic features," *Frontiers Comput. Neurosci.*, vol. 14, p. 25, Apr. 2020.

[189] D. Li, D. A. Dharmawan, B. P. Ng, and S. Rahardja, "Residual U-Net for retinal vessel segmentation," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2019, pp. 1425–1429.

[190] M. Z. Alom, C. Yakopcic, M. Hasan, T. M. Taha, and V. K. Asari, "Recurrent residual U-Net for medical image segmentation," *J. Med. Imag.*, vol. 6, no. 1, 2019, Art. no. 014006.

[191] J. Zhuang, "LadderNet: Multi-path networks based on U-Net for medical image segmentation," 2018, *arXiv:1810.07810*. [Online]. Available: http://arxiv.org/abs/1810.07810

[192] L. Li, M. Verma, Y. Nakashima, H. Nagahara, and R. Kawasaki, "IterNet: Retinal image segmentation utilizing structural redundancy in vessel networks," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Mar. 2020, pp. 3656–3665.

[193] Z. Gu, J. Cheng, H. Fu, K. Zhou, H. Hao, Y. Zhao, T. Zhang, S. Gao, and J. Liu, "CE-Net: Context encoder network for 2D medical image segmentation," *IEEE Trans. Med. Imag.*, vol. 38, no. 10, pp. 2281–2292, Oct. 2019.

[194] C. Guo, M. Szemenyei, Y. Yi, W. Wang, B. Chen, and C. Fan, "SA-UNet: Spatial attention U-Net for retinal vessel segmentation," 2018, *arXiv:1810.07810*. [Online]. Available: https://arxiv.org/abs/1810.07810

[195] Y. Zhou, H. Yu, and H. Shi, "Study group learning: Improving retinal vessel segmentation trained with noisy labels," 2021, *arXiv:2103.03451*. [Online]. Available: http://arxiv.org/abs/2103.03451

[196] S. A. Kamran, K. F. Hossain, A. Tavakkoli, S. L. Zuckerbrod, K. M. Sanders, and S. A. Baker, "RV-GAN: Segmenting retinal vascular structure in fundus photographs using a novel multi-scale generative adversarial network," 2021, *arXiv:2101.00535*. [Online]. Available: http://arxiv.org/abs/2101.00535

[197] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 801–818.

[198] K. Qi, H. Yang, C. Li, Z. Liu, M. Wang, Q. Liu, and S. Wang, "X-Net: Brain stroke lesion segmentation based on depthwise separable convolution and long-range dependencies," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2019, pp. 247–255.

[199] N. Abraham and N. M. Khan, "A novel focal tversky loss function with improved attention U-Net for lesion segmentation," in *Proc. IEEE 16th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2019, pp. 683–687.

[200] M. Asadi-Aghbolaghi, R. Azad, M. Fathy, and S. Escalera, "Multi-level context gating of embedded collective knowledge for medical image segmentation," 2020, *arXiv:2003.05056*. [Online]. Available: http://arxiv.org/abs/2003.05056

[201] K. Kamnitsas, C. Ledig, V. F. J. Newcombe, J. P. Simpson, A. D. Kane, D. K. Menon, D. Rueckert, and B. Glocker, "Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation," *Med. Image Anal.*, vol. 36, pp. 61–78, Feb. 2017.

[202] T. Okada, M. G. Linguraru, M. Hori, R. M. Summers, N. Tomiyama, and Y. Sato, "Abdominal multi-organ segmentation from CT images using conditional shape–location and unsupervised intensity priors," *Med. Image Anal.*, vol. 26, no. 1, pp. 1–18, Dec. 2015.

[203] A. Saito, S. Nawano, and A. Shimizu, "Joint optimization of segmentation and shape prior from level-set-based statistical shape model, and its application to the automated segmentation of abdominal organs," *Med. Image Anal.*, vol. 28, pp. 46–65, Feb. 2016.

[204] S. Li, H. Jiang, Z. Wang, G. Zhang, and Y.-D. Yao, "An effective computer aided diagnosis model for pancreas cancer on PET/CT images," *Comput. Methods Programs Biomed.*, vol. 165, pp. 205–214, Oct. 2018.

[205] J. Cai, L. Lu, F. Xing, and L. Yang, "Pancreas segmentation in CT and MRI images via domain specific network designing and recurrent neural contextual learning," 2018, *arXiv:1803.11303*. [Online]. Available: http://arxiv.org/abs/1803.11303

[206] T. Tong, R. Wolz, Z. Wang, Q. Gao, K. Misawa, M. Fujiwara, K. Mori, J. V. Hajnal, and D. Rueckert, "Discriminative dictionary learning for abdominal multi-organ segmentation," *Med. Image Anal.*, vol. 23, no. 1, pp. 92–104, Jul. 2015.

[207] K. Karasawa, M. Oda, T. Kitasaka, K. Misawa, M. Fujiwara, C. Chu, G. Zheng, D. Rueckert, and K. Mori, "Multi-atlas pancreas segmentation: Atlas selection based on vessel structure," *Med. Image Anal.*, vol. 39, pp. 18–28, Jul. 2017.

[208] J. Gong, J.-Y. Liu, L.-J. Wang, X.-W. Sun, B. Zheng, and S.-D. Nie, "Automatic detection of pulmonary nodules in CT images by incorporating 3D tensor filtering with local image feature analysis," *Phys. Medica*, vol. 46, pp. 124–133, Feb. 2018.

[209] R. Azad, M. Asadi-Aghbolaghi, M. Fathy, and S. Escalera, "Bi-directional ConvLSTM U-Net with densley connected convolutions," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshop (ICCVW)*, Oct. 2019, pp. 1–10.

[210] F. Shaukat, G. Raja, A. Gooya, and A. F. Frangi, "Fully automatic detection of lung nodules in CT images using a hybrid feature set," *Med. Phys.*, vol. 44, no. 7, pp. 3615–3629, Jul. 2017.

[211] W. Zhang, X. Wang, X. Li, and J. Chen, "3D skeletonization feature based computer-aided detection system for pulmonary nodules in CT datasets," *Comput. Biol. Med.*, vol. 92, pp. 64–72, Jan. 2018.

[212] A. O. D. C. Filho, A. C. Silva, A. C. de Paiva, R. A. Nunes, and M. Gattass, "3D shape analysis to reduce false positives for lung nodule detection systems," *Med. Biol. Eng. Comput.*, vol. 55, no. 8, pp. 1199–1213, Aug. 2017.

[213] S. M. Naqi, M. Sharif, and M. Yasmin, "Multistage segmentation model and SVM-ensemble for precise lung nodule detection," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 13, no. 7, pp. 1083–1095, Jul. 2018.

[214] A. Gupta, T. Saar, O. Martens, and Y. L. Moullec, "Automatic detection of multisize pulmonary nodules in CT images: Large-scale validation of the false-positive reduction step," *Med. Phys.*, vol. 45, no. 3, pp. 1135–1149, Mar. 2018.

[215] G. L. F. da Silva, T. L. A. Valente, A. C. Silva, A. C. de Paiva, and M. Gattass, "Convolutional neural network-based PSO for lung nodule false positive reduction on CT images," *Comput. Methods Programs Biomed.*, vol. 162, pp. 109–118, Aug. 2018.

MOHAN KUMAR GAJENDRAN (Graduate Student Member, IEEE) received the bachelor's degree from Hindustan University, in 2014, India, and the M.Sc. degree from the University of Missouri—Kansas City (UMKC), USA, in 2019, where he is currently pursuing the Ph.D. degree. He was a Guest Researcher with the Delft University of Technology, The Netherlands, in 2017. He served as an Instructor for UMKC, where he is currently a Research Assistant. His research interests include computational fluid dynamics, signal processing, turbulence modeling, machine learning, and deep learning.

YUGYUNG LEE (Senior Member, IEEE) received the Ph.D. degree in computer and information science from the New Jersey Institute of Technology, Newark, NJ, USA, in 1997. Since 2016, she has been working as a Professor with the School of Computing and Engineering, University of Missouri—Kansas City, MO 64110, USA. She is the author and a reviewer of high ranked research articles. Her research interests include machine learning, deep learning, medical informatics, cloud computing, distributed computing, software engineering, semantic web, and big data analytics and applications. She received prestigious awards for her quality work.

MUHAMMAD ZUBAIR KHAN (Graduate Student Member, IEEE) received the M.S. degree in software engineering from the National University of Sciences and Technology (NUST), Islamabad, Pakistan, in 2015. He is currently pursuing the Ph.D. degree in computer science from the University of Missouri—Kansas City, MO 64110, USA. Since 2018, he has been with the School of Computing and Engineering, University of Missouri—Kansas City, as a Research Fellow and a Graduate Teaching Assistant. His research interests include image processing, deep learning, medical image analysis, computer-aided medical diagnostics, cloud and parallel computing, and software analytics. He is awarded with the higher education commission scholarship for faculty development to complete his doctoral studies.

MUAZZAM A. KHAN (Senior Member, IEEE) received the Ph.D. degree in computer sciences as a Sandwich Program from IIUI and UMKC, Kansas City, MO, USA, in 2011. He was a Postdoctoral Researcher with the University of Missouri, in 2016. He is currently working as a Tenured Associate Professor with the Department of Computer Science, Quaid-i-Azam University, Islamabad, Pakistan. His research interests include wireless networks sensor, body area networks and security, image processing, compression, and encryption.

• • •