Artificial Intelligence Review (2022) 55:5845–5889
https://doi.org/10.1007/s10462-022-10152-1

Modality specific U-Net variants for biomedical image segmentation: a survey

Narinder Singh Punn[1] · **Sonali Agarwal**[1]

Accepted: 9 February 2022 / Published online: 1 March 2022

Abstract
With the advent of advancements in deep learning approaches, such as deep convolution neural network, residual neural network, adversarial network; U-Net architectures are most widely utilized in biomedical image segmentation to address the automation in identification and detection of the target regions or sub-regions. In recent studies, U-Net based approaches have illustrated state-of-the-art performance in different applications for the development of computer-aided diagnosis systems for early diagnosis and treatment of diseases such as brain tumor, lung cancer, alzheimer, breast cancer, etc., using various modalities. This article contributes in presenting the success of these approaches by describing the U-Net framework, followed by the comprehensive analysis of the U-Net variants by performing (1) inter-modality, and (2) intra-modality categorization to establish better insights into the associated challenges and solutions. Besides, this article also highlights the contribution of U-Net based frameworks in the ongoing pandemic, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) also known as COVID-19. Finally, the strengths and similarities of these U-Net variants are analysed along with the challenges involved in biomedical image segmentation to uncover promising future research directions in this area.

Keywords Biomedical image segmentation · Deep learning · U-Net

1 Introduction

Biomedical information technologies have a great importance, particularly in medi-cine to solve different problems (Goceri and Songul 2018; Kaya et al. 2017; Goceri 2016; Göçeri et al. 2015; Göçeri 2013), and deep learning-based approaches have been widely used recently (Goceri 2021; Göçeri 2020). The evolving medical imaging acquisition system (Alexander et al. 2019) has brought the consideration of the research community towards the non-invasive practice of disease diagnosis. Every diagnostic procedure involves the careful and critical examination of medical scans which represents the complex interior structure within the body, illustrating the functioning of various organs. With a wide

✉ Narinder Singh Punn
pse2017002@iiita.ac.in

[1] IIIT Allahabad, Prayagraj 211015, India

variety of medical imaging such as magnetic resonance imaging (MRI), X-ray, computerized tomography/computerized axial tomography (CT/ CAT), ultrasound (US), positron emission tomography (PET), etc., the medical domain has experienced exponential growth in the diagnosis practices. Each of these scans varies in the imaging procedure, usecases and its average diagnosis duration (Hughes 2019; TMI 2019), as shown in Table 1. For any radiologist, analyzing such complex scans is tedious and time consuming, thereby to fill this void of complexity, deep learning approaches are well explored to address the automated assistance in diagnosis procedure, resulting in faster and better practices for monitor, cure and treatment of the diseases (Elnakib et al. 2011; Masood et al. 2015; Deepa et al. 2011; Maintz and Viergever 1998).

Segmentation (Minaee et al. 2020) is one such automation task that helps to identify and detect the desired regions or objects of interest for the concerned issue. Depending on the depth of identifying the classes of objects, segmentation is divided into two levels as semantic and instance. The semantic segmentation (Liu et al. 2019b) segregates the objects belonging to different classes, whereas instance segmentation (Chen et al. 2019a) goes deeper to also segregate the objects within the common class. With the exhaustive analysis (Minaee et al. 2020; Haque and Neubert 2020), it is observed that among the latest advancements to perform segmentation, mostly U-Net (Ronneberger et al. 2015) based frameworks are adopted to achieve state-of-the-art segmentation performance which follows from its symmetrical encoder-decoder structure to extract and reconstruct the feature maps.

1.1 Motivation and contribution

With recent developments in deep learning technologies, there are a lot of review articles on biomedical image segmentation (BIS) using deep learning. The understanding of the available methods is critical for developing computer-aided diagnosis systems; however, to contribute to this domain as a researcher, one needs to understand the underlying mechanics of the methods that make those systems achieve promising results. For instance, the work of Zhou et al. (2019b) explored the comprehensive analysis focused on multi-modality fusion approaches, whereas Haque and Neubert (2020) reviewed the standard deep learning approaches for BIS using different modalities. Table 2 shows the overview of some of the survey articles proposed for biomedical image analysis using deep learning approaches. In contrast to the existing review articles, the present article is intended to contribute for an exhaustive analysis of the state-of-the-art modality specific U-Net based approaches by performing inter-modality and intra-modality categorization, and establishing better insights of technological solutions for each modality. Furthermore, the present article also uncovers general and modality specific challenges to perform biomedical image segmentation and make the researchers or readers reap the most benefits from the current advancements in U-Net and aid in further contributions towards the research in this area.

1.2 Review process

The basis of including a research article in this survey is that the article describes the research on U-Net based biomedical image segmentation. The articles confirming vivid architectures or frameworks are only included if the authors claimed certain advancements or novel contributions, whereas articles with pure discussions are excluded; fortunately, such articles are limited and hence will not affect the outcome of this survey.

Table 1 Medical imaging approaches for diagnosis

Imaging type	Approach	Usecase	Duration (in min.)
MRI	Magnetic fields and radio waves	Multiple sclerosis, stroke, tumors, spinal cord disorders, etc.	45–60
X-ray	Ionizing radiation	Fractures, arthritis, osteoporosis, breast cancer, etc.	10–15
CT/CAT	Ionizing radiation	Trauma injuries, tumors and cancers, vascular and heart diseases, etc.	10–15
US	Sound waves	Gallbladder illness, breast lumps, genital disorder, joint problems, etc.	30–60
PET	Radioactive tracer	Alzheimer, epilepsy, seizures, parkinsons' disease, etc.	90–120

The search for the articles is performed on Google Scholar, which is one of the best academic search engine (SearchEngines 2020), where relevant articles are identified using the search string, SS1 as shown in Table 3. Among the acquired papers, the high quality journals or conferences are confirmed by analyzing its impact factor (high), h-index (high), peer-review process (transparent), indexing (MEDLINE, Elsevier Scopus and EMBASE, Clarivate Analytics Web of Science, Science Citation Index, etc.) and scientific rigor. These reputed journals are identified from the ranked list, CORE (CORE 2020). However, some articles are also included from popular preprint servers such as arXiv. With such a huge pool of acquired articles, the most relevant articles are filtered with a thorough examination (journal or conference quality, cite score and contribution) to include in this survey. These articles are analysed by categorizing it in one of the proposed classes and highlighting the architectural design of U-Net variant along with the achieved results and further possible improvements to address modality specific segmentation challenges.

1.3 Research trend in BIS

A comparative literature exploration is conducted on the Google Scholar search engine using the search strings, SS2 and SS3, as shown in Table 3. The number of BIS approaches without U-Net are acquired by subtracting the number of BIS U-Net articles from the pool of BIS articles, to understand the latest trend of research. Fig. 1 illustrates that the latest approaches are developed by employing the U-Net framework while experiencing exponential growth every year. In order to analyse such trend, this article aims to provide an exhaustive review of the variants of U-Net architectural design developed for segmentation. It is evident that the U-Net model incorporates the huge potential for further advancements due to its mutable and modular structure that would result in a state-of-the-art diagnosis system.

1.4 Article structure

The remaining portion of the article is divided into several sections, where Sect. 2 presents the overview of biomedical image analysis and in Sects. 3, 4 and 5 the comprehensive analysis of U-Net variants is presented that covers implementation strategies and advancements. Later, Sect. 6 presents the observations concerned with the current advancements in

Table 2 Summary of existing review articles for biomedical image analysis

Author	Contribution
Havaei et al. (2016)	Reviews CNN based approaches along with the key challenges for brain pathology segmentation using MRI such as tumor, lesion, etc.
Razzak et al. (2018)	Explores the potential of deep learning based approaches for various medical imaging applications across different modalities.
Hesamian et al. (2019)	Investigates state-of-the-art deep learning techniques for medical image segmentation along with the network training techniques and state-of-the-art solutions to the challenges.
Taghanaki et al. (2021)	Covers the comprehensive analysis of deep learning approaches in the segmentation of natural and medical images with different categories and applications.
Zhou et al. (2019b)	Explores multi-modality fusion based approaches for medical image segmentation.
Chen et al. (2020)	Reviews deep learning approaches for cardiac image segmentation using ultrasound, CT and MRI imaging along with the various techniques to address the challenges.
Haque and Neubert (2020)	Presents literature survey of deep learning technologies for biomedical image segmentation with different modalities.
Lei et al. (2020)	Reviews various deep learning models for medical image segmentation with supervised and weakly supervised learning aspects.
Renard et al. (2020)	Emphasizes the variability and reproducibility of the deep learning approaches for medical image segmentation.

U-Net based approaches, followed by the scope and challenges in Sect. 7 and concluding remarks in Sect. 8.

2 Biomedical image analysis

The success of deep learning in image analysis has encouraged biomedical imaging researchers to investigate its potential in analyzing various medical modalities to aid clinicians in faster diagnosis and treatment of diseases or infections like the ongoing pandemic of SARS-CoV-2 (COVID-19). Following the deep learning usecases, the implication of classification can ascertain the presence or absence of disease in some modality e.g. the ground glass opacification (GGO) in the lungs via CT imaging. Furthermore, in localization, normal anatomy can be identified e.g. lungs in the CT or X-ray imaging, and later segmentation can generate refined boundaries around the GGOs to understand its impact on the anatomical structures for further analysis. Since, segmentation is an extension to

Table 3 Search strings to acquire research papers and analyse research trend using GoogleScholar

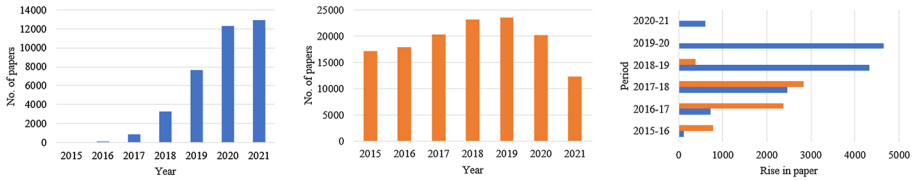
No.	Search string	Queried date	Year	No. of papers
SS1	(U-Net segmentation CT OR X-ray OR PET OR US OR MRI)	November 10, 2021	2015-21	32,530
SS2	(biomedical image segmentation)	November 10, 2021	2015-21	172,190
SS3	(biomedical image segmentation "U-Net")	November 10, 2021	2015-21	38,900

classification, localization or detection, it offers very rich information about the disease and infected regions. With this interest, many architectures have been proposed for the segmentation of the targeted regions from vivid modalities (Haque and Neubert 2020). In addition, segmentation is the most widely researched application of deep learning in biomedical image analysis (Litjens et al. 2017), where U-Net based segmentation architectures have gained significant popularity to develop computer-aided diagnosis (CAD) systems.

2.1 Rise of segmentation architectures

Despite the advancements in deep learning, segmentation is still one of the challenging tasks due to the varying dimensions, shape and locale of the target tissues. Traditionally, the segmentation process was carried manually by expert clinicians to illuminate the regions of interest in the whole volume of samples, thereby it is ideal to automate this process for faster diagnosis and treatment. In recent years, various deep learning models are developed for BIS that are categorized into manual, semi-automatic and fully automatic approaches (Haque and Neubert 2020). Fig. 2 presents the schematic representation of the pipeline of the recent deep learning based segmentation frameworks for biomedical images, which is divided into data preprocessing (Bhattacharyya 2011), deep learning model (Minaee et al. 2020), and post-processing (Zhou et al. 2019a; Christ et al. 2016). In the data preprocessing stage, the data undergoes a certain set of operations like resize and normalization to reduce the intensity variation in the image samples, augmentation to generate more training samples for avoiding the class biasness and overfitting problem, removal of irrelevant artefacts or noise from the data samples, etc. The pre-processed data is then fed to train the deep neural segmentation network, where mostly U-Net based architectures are deployed. The output of the network undergoes post-processing with techniques such as morphological and conditional random field based feature extraction to refine the final segmentation results.

Initiated from the sliding window approach by Ciresan et al. (2012) to classify each pixel while also localizing the regions using patch based input, the model outperformed in the ISBI 2012 challenge; however, the training was slow because of a large number of overlapping patches and also lacked the balance of context and localization accuracy. Long et al. (2015) proposed fully convolutional neural network (FCN) for semantic segmentation, defined on the state-of-the-art classification networks like Alex-Net, VGG-Net and Google-Net. This model can process images of arbitrary size and produce the segmentation mask of same size by using deconvolution; however, it does not utilize global information context and hence generates fuzzy segmentation masks. Later, the U-Net model proposed by Ronneberger et al. (2015), consists of FCN along with the contraction-expansion paths and skip connections to gradually adapt the long-range affinities. The contraction phase



(a) U-Net based approaches for BIS. **(b)** Other than U-Net based approaches for BIS. **(c)** Growth in approaches for BIS.

Fig. 1 Research trend in biomedical image segmentation per year

tends to extract high and low level features, whereas the expansion phase follows from the features learned in the corresponding contraction phase (skip connections) to reconstruct the image into the desired dimensions with the help of transposed convolutions or upsampling operations. The U-Net model won the ISBI 2015 challenge and outperformed its predecessors. Later, a similar approach is proposed by Çiçek et al. (2016) in the three dimensional feature space to perform volumetric segmentation of Xenopus kidney and achieved promising results. Following from the state-of-the-art potential of the U-Net model, many variants have been proposed based on the variation in the convolution and pooling operations, skip connections, the arrangement of the components in each layer and hybrid approaches that make use of the state-of-the-art deep learning models, to tackle the challenges associated with different applications.

2.1.1 U-Net

With the sense of segmentation being a classification task where every pixel is classified as being part of the target region or background, Ronneberger et al. (2015) proposed a U-Net model to distinguish every pixel, where input is encoded and decoded to produce output with the same resolution as input. As shown in Fig. 3, the symmetrical arrangement of encoder-decoder blocks efficiently extracts and concatenates multi-scale feature maps, where encoded features are propagated to decoder blocks via skip connections and a bottleneck layer.

The encoder block (contraction path) consists of a series of operations involving valid 3×3 convolution followed by a ReLU activation function (as shown in Fig. 4(a)), where a 1-pixel border is lost to enable processing of the large images in individual tiles. The obtained feature maps from the combination of convolution and ReLU are downsampled with the help of max pooling operation, as illustrated in Fig. 4(b). Later, the number of feature channels are increased by a factor of 2, following each layer of convolution, activation and max pooling, while resulting in spatial contraction of the feature maps. The extracted feature maps are propagated to decoder block via bottleneck layer that uses cascaded convolution layers. The decoder block (expansion path) consists of sequences of up-convolutions (as shown in Fig. 4(c)) and concatenation with high-resolution features from the corresponding encoded layer. The up-convolution operation uses the kernel to map each feature vector to the 2×2 pixel output window followed by a ReLU activation function. Finally, the output layer generates a segmentation mask with two channels comprising background and foreground classes. In addition, the authors addressed the challenge to segregate the touching or overlapping regions by inserting the background pixels between

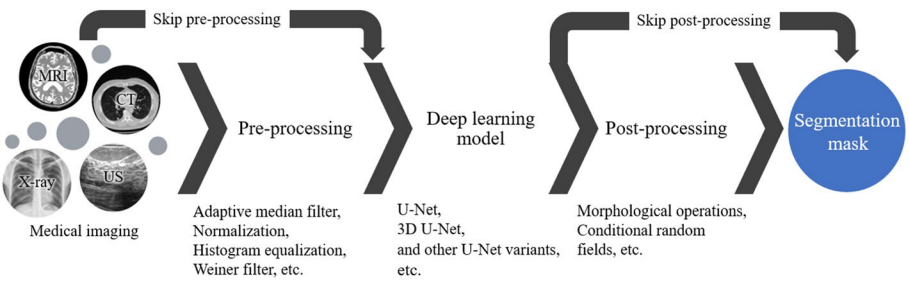


Fig. 2 Schematic representation of deep learning based segmentation architectures

the objects and assigning an individual loss weight to each pixel. This energy function is represented as a pixel-wise weighted cross entropy function as shown in Eq. 1. The authors established the state-of-the-art results by winning the ISBI 2015 challenge.

$$E = \sum_{x \in \Omega} \left(w_c(x) + w_0 \cdot \exp\left(-\frac{(d_1(x) + d_2(x))^2}{2\sigma^2} \right) \right) \log(p_{\ell(x)}(x)) \tag{1}$$

where softmax, $p_k(x) = \exp(a_k(x)) / \left(\sum_{k'=1}^{K} \exp(a_{k'}(x)) \right)$ with activation, $a_k(x)$ for channel k and pixel $x \in \Omega$ with $\Omega \in \mathbb{Z}^2$, w_c is the weight map, d_1 and d_2 are the distances to the nearest and the second nearest boundary pixels, and w_o and σ are constants.

2.1.2 Other than U-Net

U-Net is the most suitable segmentation model in the area of biomedical image analysis because of its ability to simultaneously combine high and low level information which helps to extract complex features and improve accuracy, respectively. However, there are various other deep learning based models that are utilized for segmentation such as FCN (Long et al. 2015), DeepLab (Chen et al. 2017a), SegNet (Badrinarayanan et al. 2017), mask R-CNN (He et al. 2017), etc.

Long et al. (2015) introduced FCN that has set the foundation of segmentation architectures across various domains. In contrast to classical CNN models (VGG, ResNet, etc.) where fully connected layers are employed to categories an entire image, FCN uses 1×1 convolution layers to perform pixel level classification and generate segmentation mask by upsampling the feature maps of the last convolution layer via deconvolution layer. However, with this arrangement of operations the generated masks are relatively fuzzy and insensitive to the global context information (Minaee et al. 2020). Unlike FCN which uses deconvolution to upsample the feature maps, SegNet (Badrinarayanan et al. 2017) is designed as a symmetric encoder-decoder structure, where encoder block uses VGG16

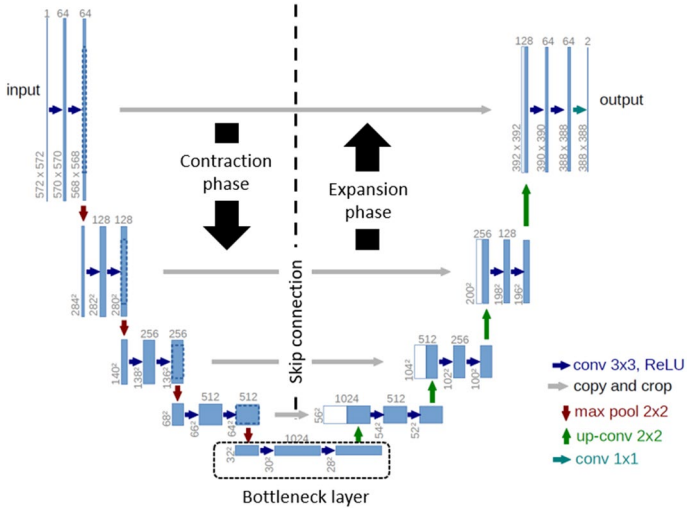


Fig. 3 U-Net architecture

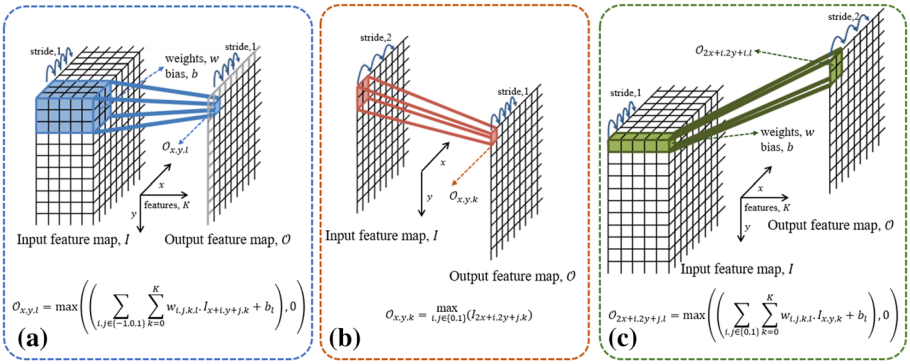


Fig. 4 Summary of operations in U-Net. (a) 3 × 3 convolution + ReLU, (b) 2 × 2 max-pooling and (c) 2 × 2 up-convolution operation

network topology for feature extraction and a corresponding decoder block uses max pooling indices that are transferred from encoder to decoder blocks, to generate sparse upsampled feature map without using any training parameters. However, this arrangement of operations ignores the pixel adjacent information especially during upsampling of low dimensional feature maps. U-Net addresses this issue by transferring the entire feature map from encoder to decoder during upsampling, but at the cost of more memory requirement; however, it can be neglected due to the significant improvements in the segmentation results.

Inspired from the potential of faster R-CNN model (Ren et al. 2015) to perform object detection, He et al. (2017) proposed mask R-CNN model to further refine the object boundaries for segmentation by first computing the object detection with bounding boxes, predicting the associated classes and finally computing the binary mask to segment objects. Vuola et al. (2019) analysed mask R-CNN model for nuclei segmentation, where the network accurately detected nuclei with bounding boxes but struggles to generate a better segmentation mask. Following this, the authors integrated mask R-CNN with U-Net to improve the overall segmentation performance.

DeepLab is another family of segmentation models that have improved over the years, where each phase of enhancement is named as DeepLabv1 (Chen et al. 2014), DeepLabv2 (Chen et al. 2017a), DeepLabv3 (Chen et al. 2017b) and DeepLabv3+ (Chen et al. 2018a). DeepLabv1 model uses VGG16 model, where fully connected layers are removed and pooling layers are replaced with atrous convolution. DeepLabv2 model address the difficulty of the DeepLabv1 model to segment the same objects with different sizes in an image by using ResNet101 as the backbone model and atrous spatial pyramid pooling (ASPP) to capture the multi-scale context of the objects in an image. To further refine the results, in DeepLabv3 parallel or cascaded atrous convolution block is designed with multiple dilation rates to better capture multi-scale context. DeepLabv3+ further extends the DeepLabv3 with a decoder block to improve the segmentation results. It uses feature maps from the middle layer and the Xception model for segmentation. Moreover, it also uses depthwise separable convolutions with ASPP to reduce the training parameters. In most of the U-Net variants, these modules are integrated with the network to achieve better segmentation results.

2.1.3 Implementation strategies

The implementation strategies of segmentation architectures can be divided into two categories: 1) training from scratch and 2) training using a pre-trained model (also known as transfer learning). In first approach (as shown in Fig. 5(a)), an entire model is trained in which training parameters are initialized with Xavier initialization (Glorot and Bengio 2010) or Kaming initialization (He et al. 2015a). Due to which this approach requires a large number of labelled data samples to optimize the training parameters and learn the desired task. Hence, this approach requires intensive time and effort to develop and train the model. In the transfer learning paradigm, as simulated in Fig. 6, a pre-trained model (models trained on benchmark datasets such as ImageNet) is utilized as a backbone model to train on different data involving similar or different tasks such as object detection and image segmentation. As shown in Fig. 5(b) and Fig. 5(c), the transfer learning or domain adaptation can be applied in two schemes, either freezing the base model and training the later layers for prediction, or semi-freezing the base model, where few high level layers are retrained along with the prediction layers. The transfer learning approach typically produces better results than the random initialization of the training parameters (Garcia-Garcia et al. 2018).

2.1.4 Performance metrics

The performance metrics are the key factors to evaluate and compare the segmentation performance of the models. Due to the unavailability of the standard metrics, each system requires an appropriate and different selection of metrics that can quantify time, computational and memory space requirements and overall performance (Fenster and Chiu 2006). Table 4 presents the most popular evaluation metrics that are utilized to analyse the performance in BIS models. In BIS, mostly the datasets are imbalanced i.e. the number of pixels/voxels concerning the target region (region of interest) are relatively less than the dark pixels/voxels (background region), due to which the metrics such as accuracy, which are best suited for a balanced distribution of data samples, are not recommended for BIS evaluation of the models. Among the discussed metrics intersection-over-union (IoU or Jaccard index) and dice similarity coefficient are the most widely used evaluation metrics in BIS for various modalities. More details can be found in the recent review articles (Haque and Neubert 2020; Minaee et al. 2020).

2.1.5 Loss functions

The loss functions or objective functions drive the training procedure of the deep learning models. For the BIS task, loss functions are tuned to alleviate the above discussed class imbalance problem by refining the distributions of the training data. With each dataset introducing its complexities and challenges, the loss functions are grouped into four categories based on the distribution, region, boundary and hybrid (Ma 2020), as shown in Table 5, along with their respective usecases. For ease in representation, the loss functions are summarized for the semantic segmentation scenario, where the number of classes is limited to two (background and target region). The effect of these loss functions for biomedical image segmentation using various modalities over nnU-Net model (Isensee et al. 2021) is explored by Nasalwai et al. (2021), and also proposed an accelerated tversky loss (ATL) function to achieve faster model training or convergence.

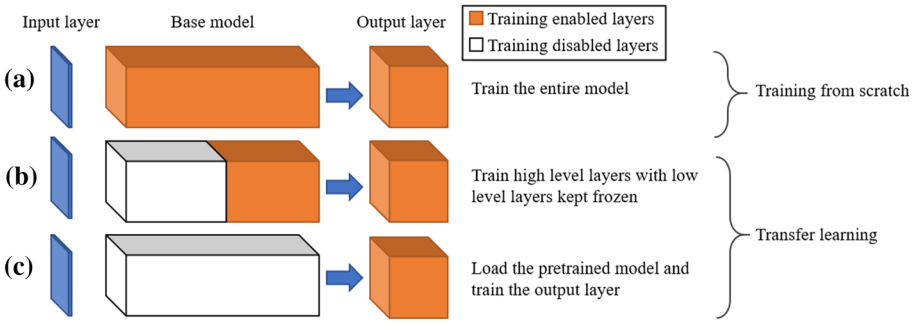


Fig. 5 Typical approaches for model training

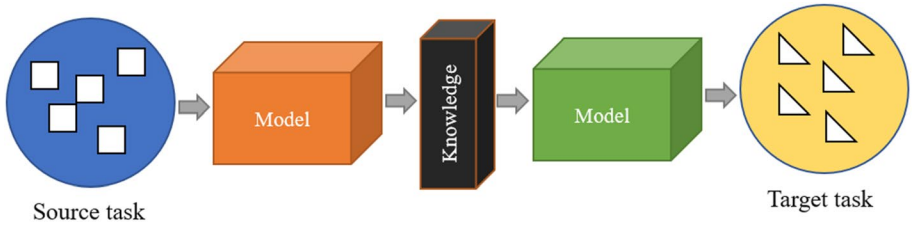


Fig. 6 Illustration of transfer learning approach to adapt to new task

3 U-Net variants for medical imaging

The numerous development in medical imaging acquisition systems and deep learning technologies have resulted in the rise of usage frequency of modalities for computer-aided diagnosis. Despite vanilla U-Net being super-efficient in the ISBI cell tracking challenge, there is still a void to fill with improvements in certain aspects. The most apparent problem in the vanilla U-Net is that the learning may slow down in deeper layers of the U-Net model which increases the possibility of the network ignoring the layers representing abstract features of the target structure. This slack in the learning process is due to the generation of diluted gradients in the deeper layers. Another major issue is concerned with the localized convolutions which tend to limit the capability of the model to efficiently capture global and long-range dependencies. Furthermore, the distinct challenges (discussed in a later section) introduced in performing segmentation using different modalities needs to be addressed; however, it is not optimal with vanilla U-Net. Following this context, various U-Net variants are proposed to improve the segmentation performance across vivid modalities. To establish a better understanding of these variants, the present review performs: 1) inter-modality categorization - to show variation in the segmentation approaches across the different modalities (X-ray, CT, MRI, PET and ultrasound), and 2) intra-modality categorization - to group each U-Net variant within the same modality based on its most profound technical contribution (better U-Nets, attention U-Nets, inception U-Nets and ensemble U-Nets), as shown in Fig. 7. Within each modality, a similar type of categorization is performed to better distinguish the type of approaches introduced for each modality. In the case where a U-Net variant uses multiple modifications, then based on its most profound

Table 4 Summary of performance metrics for BIS in terms of number of true positive (*TP*), true negative (*TN*), false positive (*FP*) and false negative (*FN*), predicted mask ($\mathcal{P}$) and ground truth ($\mathcal{G}$), $\mathcal{H}(X, Y)$ is the directed *AHD* from *X* to *Y* with *d* as euclidean distance, $\mathcal{V}_p$ and $\mathcal{V}_g$ refer to the volumes of generated and reference segmentation

Metric	Expression
Accuracy	$A = \frac{(TP+TN)}{(TP+TN+FP+FN)}$
Precision	$P = \frac{TP}{(TP+FP)}$
Recall	$R = \frac{(TP)}{(TP+FN)}$
F1-score	$F1 = 2 \times \frac{(P \times R)}{(P+R)}$
Specificity	$S = \frac{TN}{(TN+FP)}$
Dice similarity coefficient	$DSC = \frac{2\times\lvert\mathcal{P}\cap\mathcal{G}\rvert}{\lvert\mathcal{P}\rvert+\lvert\mathcal{G}\rvert} = \frac{2TP}{2TP+FP+FN}$
Intersection-over-union	$IoU = \frac{\mathcal{P}\cap\mathcal{G}}{\mathcal{P}\cup\mathcal{G}} = \frac{TP}{TP+FP+FN}$
Average Hausdo-rff distance	$AHD = \frac{1}{2}\left(\frac{\mathcal{H}(\mathcal{P},\mathcal{G})}{\mathcal{P}} + \frac{\mathcal{H}(\mathcal{G},\mathcal{P})}{\mathcal{G}}\right)$ $= \frac{1}{2}\left(\frac{1}{\mathcal{P}}\sum_{p\in\mathcal{P}}\min_{g\in\mathcal{G}} d(p,g) + \frac{1}{\mathcal{G}}\sum_{g\in\mathcal{G}}\min_{p\in\mathcal{P}} d(p,g)\right)$
Absolute Volume Difference	$AVD = \frac{\lvert\mathcal{V}_p - \mathcal{V}_g\rvert}{\mathcal{V}_g} \times 100$

Table 5 Summary of the most widely used loss functions for biomedical image segmentation with respect to the predicted mask ($\mathcal{P}$) and ground truth mask ($\mathcal{G}$), α and γ as constants, *h* is Hausdorff distance and *d* is the operator for Euclidean distance

Type	Objective functions	Usecase
Distribution	$\mathcal{L}_{BCE} = -(glog(p) + (1-g)log(1-p))$	Balanced distribution of data
	$\mathcal{L}_{WCE} = -(\alpha.glog(p) + (1-g)log(1-p))$	Skewed dataset
	$\mathcal{L}_{BaCE} = -(\alpha glog(p) + (1-\alpha)(1-g)log(1-p))$	Skewed dataset
	$\mathcal{L}_{Focal} = \begin{cases} -\alpha(1-p)^{\gamma} log(p), & \text{if } g = 1 \\ -(1-\alpha)(p)^{\gamma} log(1-p), & \text{otherwise} \end{cases}$	Focuses on hard samples
Region	$\mathcal{L}_{DSC} = 1 - \frac{2gp+1}{g+p+1}$	Widely used for segmentation
	$\mathcal{L}_{IoU} = 1 - \frac{gp}{g+p-gp}$	Widely used for segmentation
	$\mathcal{L}_{SS} = \alpha * \text{sensitivity} + (1-\alpha) * \text{specificity}$	Focuses to improve true positive rate
	$\mathcal{L}_{Tversky} = 1 - \frac{1+gp}{1+gp+\alpha(1-g)p+(1-\alpha)g(1-p)}$	Introduces weights for false predictions
Boundary	$\mathcal{L}_{HD} = \frac{1}{N}\sum_{i=0}^{N}\left((p_i - g_i)^2.(h_{pi}^2 + h_{gi}^2)\right)$	Widely used for segmentation
	$\mathcal{L}_{SA} = -\sum_i CE(p_i, g_i) - \sum_i \alpha_i d(\mathcal{P}, \mathcal{G}) CE(p_i, g_i)$	Focuses to segment boundaries of the regions
Compound	$\mathcal{L}_{Combo} = \alpha\mathcal{L}_{BaCE}(g,p) - (1-\alpha)\mathcal{L}_{DSC}(g,p)$	Leverages features of *BaCE* and *DSC* for skewed data
	$\mathcal{L}_{EL} = \alpha_{DSC} e^{(-ln(\mathcal{L}_{DSC})^{\gamma})} + \alpha_{CE}\, e^{(-ln(\mathcal{L}_{CE})^{\gamma})}$	Focuses on less accurate predictions

BCE binary cross-entropy, *WCE* weighted cross-entropy, *BaCE* balanced cross-entropy, *DSC* dice similarity coefficient, *IoU* intersection-over-union, *SS* sensitivity-specificity, *HD* Hausdorff distance, *SA* shape-aware, *EL* exponential-logarithmic

enhancement or technical contribution it is added in that category, e.g. RCA-IUnet model (Punn and Agarwal 2021b) for breast cancer segmentation using ultrasound imaging, uses residual cross-spatial attention and inception convolutions, is categorized as an attention U-Net variant under ultrasound modality. Each of the intra-modality categories are described as follows:

- Better U-Nets (BU): This category consists of U-Net models that are better than raw U-Net model by slight modifications such as integrating with FCN or SegNet models, transfer learning, dense or residual blocks, multi-stage training, multi-tasking, etc.
- Attention U-Nets (AU): These are the approaches with variation in the attention mechanism of the feature maps to filter the relevant features such as spatial and channel attention, mixed attention, non-local attention, etc.
- Inception U-Nets (IU): These are the approaches that use multi-scale feature fusion strategies to effectively learn the feature representations.
- Ensemble U-Nets (EU): These are the approaches that use multiple models or sub-models with or without the other enhancements to improve the segmentation performance.

Hence, for faster and efficient computer-aided diagnosis practices, the following sections present wide varieties of U-Net based approaches for biomedical image segmentation using various modalities. Table 6 summarizes the various U-Net variants reviewed in the following sections.

3.1 X-ray

In radiology, X-ray imaging is utilized as a diagnostic procedure of human bones and tissues. X-ray possesses the properties of penetrability, photographic effect and fluorescence effect. Human body tissues vary in density and thickness due to which X-rays are absorbed with different degrees, resulting in black and white contrast images (Bercovich and Javitt

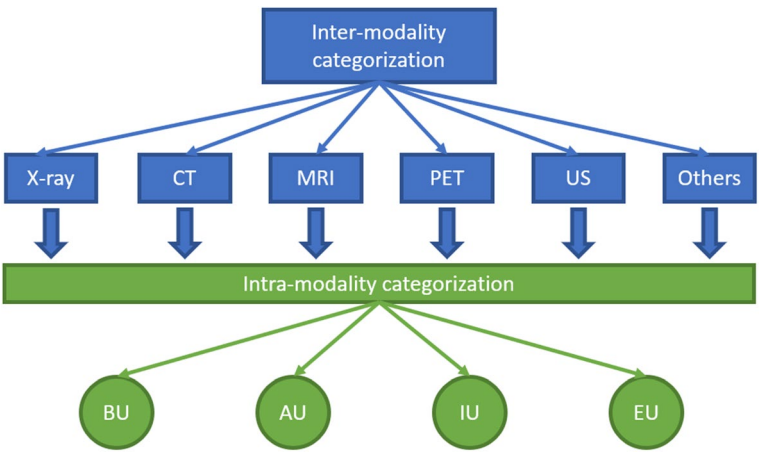


Fig. 7 Categorization scheme for U-Net variants

2018). The wide and easy availability of X-ray imaging has encouraged the research community to contribute towards smart diagnosis systems.

3.1.1 Better U-Nets

The segmentation of lungs from chest X-ray (CXR) imaging is a crucial step for any CAD system. Following this, Rashid et al. (2018) exploits the potential of U-Net model to generate the segmentation masks of the lungs from CXR images, where the produced masks are iteratively refined with post-processing techniques such as flood fill algorithm and morphological operations. Significant improvement is observed as compared to traditional segmentation approaches such as adaptive region growing, edge detection, statistical shape models, etc., over multiple datasets. To further improve segmentation performance, Frid-Adar et al. (2018a) employed a pre-trained VGG-16 model in the encoder phase, where the decoder or the expansion phase uses upsampling and standard convolution operations sequentially for multi-class segmentation involving anatomical structures like lungs field, heart and clavicles in chest X-ray samples. While training, the pre-trained weights are fine-tuned to better extract or encode the desired features of the target classes. Unlike the approach by Rashid et al. (2018), this model with transfer learning achieved promising results without any post-processing overhead. Besides, the authors also analysed the proposed model with multiple loss functions like *DSC*, *IoU*, Tversky and *BCE*, where the use of *DSC* produced the best results.

In another work, Abedalla et al. (2020) proposed a deep learning framework 2STU-Net to perform segmentation of pneumothorax (collapsed lung) in the CXR samples. It comprises a state-of-the-art residual network (ResNet-34) that is pre-trained on the ImageNet dataset and arranged in the U-Net topology. Similar to the work by Frid-Adar et al. (2018a), the encoder is built with ResNet-34 (He et al. 2016a) by removing the last layers, whereas the decoder follows standard blocks of CNN with upsampling. Initially, the data is pre-processed to produce images of dimensions 256×256 and 512×512 for 2 stage training scheme and multi-scale feature learning. The ResNet34U-Net is first trained with lower resolution images and later the same model is fine-tuned (keeping previously learned weights as initial weights) to adapt high resolution images. The authors also utilized stochastic weight averaging (SWA) and test-time augmentation (TTA) techniques to improve the test results. The significance of 2 stage training is justified with the faster convergence of the second training stage and better segmentation results, thereby highlighting the effectiveness of multi-scale feature representation learning. However, the overall training overhead is increased due to two stage training. Wang et al. (2020a) synthesized a CXR dataset annotated with clavicles, anterior ribs, posterior ribs and bones, on which a multi-task dense connection U-Net (MDU-Net) is trained for multi-class segmentation. A feature separation network is introduced for multi-label segmentation where a pixel value is associated with more than one label e.g. the pixels in the overlapped regions of anterior and posterior ribs have multiple tags. For every CXR image, multiple masks are generated concerning different annotations, thereby multiple networks are trained to generate the corresponding mask. The implication of increased training time is addressed with the help of transfer learning, where the network uses a pre-trained DenseNet201 (Huang et al. 2017) model for feature extraction. However, due to the 2D projection of X-ray imaging, each annotated mask also covers features representing other masks categories which may deviate the network from learning the class specific feature representations.

Table 6 Summary of popular U-Net variants for BIS

Author	U-Net variant	Modality	TL	SL	Pr	Po	Category	Description
Dong et al. (2017)	Modified U-Net	MRI	–	✓	✓	–	BU	FCN based U-Net
Rashid et al. (2018)	Modified U-Net	X-ray	–	✓	✓	✓	BU	FCN based U-Net
Frid-Adar et al. (2018a)	Modified U-Net	X-ray	✓	–	✓	–	BU	U-Net with pre-trained VGG-16 encoder
Que et al. (2018)	CardioXNet framework	X-ray	–	✓	✓	✓	EU	Two parallel U-Net models with binary contours
Oktay et al. (2018)	Attention U-Net	CT	–	✓	✓	–	AU	Attention skip-connections
Kohl et al. (2018)	Probabilistic U-Net	CT	–	✓	–	–	EU	U-Net with conditional variational autoencoder
Tong et al. (2018)	Improved U-Net	CT	–	✓	✓	–	BU	Mini-residual connections within encoder-decoder phases
Janssens et al. (2018)	Two stage U-Net model	CT	–	✓	✓	–	EU	3D FCN LocalizationNet followed by SegmentationNet
Kumar et al. (2018)	U-SegNet	MRI	✓	–	✓	–	BU	Integration of skip connections with SegNet
Kermi et al. (2018)	Residual U-Net	MRI	–	✓	✓	–	BU	Residual blocks between two convolution layers
Chen et al. (2018b)	S3DU-Net	MRI	–	✓	✓	–	IU	U-Net with spatiotemporal separable convolution
Blanc-Durand et al. (2018)	Vanilla 3D U-Net	PET	–	✓	✓	✓	BU	CNN based 3D U-Net
Zhao et al. (2018)	3D FCN	PET	–	✓	✓	–	IU	3D FCN multi-modal fusion network
Almajalid et al. (2018)	U-Net + SRAD	US	–	✓	✓	✓	BU	Base U-Net with speckle reducing anisotropic diffusion
Wang et al. (2018b)	cU-Net	US	–	✓	✓	–	EU	Classification and segmentation U-Net
Alom et al. (2018)	R2U-Net	Multi-modality	–	✓	✓	✓	BU	Recurrent Residual convolutional neural network based on U-Net (R2U-Net)
Zhou et al. (2018a)	UNet++	Multi-modality	–	✓	✓	–	BU	Nested U-Net model
Subramanian et al. (2019)	CVC framework	X-ray	✓	–	–	–	EU	Two parallel U-Net models with spatial priors and pre-trained NN-RF
Li et al. (2019a)	U-Net based framework	X-ray	✓	–	✓	✓	AU	SE and residual based attention CNN
Dong et al. (2019b)	U-Net-GAN	CT	–	✓	✓	–	EU	U-Net act as a generator and FCN as discriminator network
Liu et al. (2019c)	GIU-Net	CT	–	✓	✓	✓	EU	Deeper U-Net model with graph cut algorithm
Man et al. (2019)	GAU-Net	CT	✓	–	✓	–	AU	Deformable geometry-aware U-Net with deep Q learning
Seo et al. (2019)	mU-Net	CT	–	✓	–	–	AU	Object dependent filters in skip connections
Hiasa et al. (2019)	Bayesian U-Net	CT	–	✓	✓	✓	EU	Cascaded U-Net and Bayesian U-Net models
Song et al. (2019)	U-NeXt	CT	–	✓	✓	–	AU	U-Net model with attention blocks, SkipSPP and dense convolutions

Table 6 (continued)

Author	U-Net variant	Modality	TL	SL	Pr	Po	Category	Description
Rundo et al. (2019)	USE-Net	MRI	–	✓	✓	✓	AU	U-Net model with the squeeze-and-excitation blocks
Wang et al. (2019b)	MSU-Net	MRI	–	✓	✓	–	IU	Multiscale statistical U-Net
Dong et al. (2019a)	DAU-Net	MRI	–	✓	-	-	AU	Deep attention U-Net with deep supervision
Wang et al. (2019a)	3D DSD-FCN	MRI	–	✓	✓	✓	EU	3D FCN with deep supervision and group dilation
Guo et al. (2019)	3D Dense U-Net	PET	–	✓	✓	–	IU	3D U-Net with dense convolution fusion blocks
Yang et al. (2019)	DPU-Net	US	–	✓	✓	–	IU	Dual path U-Net with parallel multi-branch encoding and decoding
Li et al. (2019b)	DU-Net	US	–	✓	✓	✓	BU	Dense convolution U-Net
Lin et al. (2019)	SSU-Net	US	–	✓	✓	–	AU	Semantic-embedding and shape-aware U-net
Azad et al. (2019)	BCDU-Net	Multi-modality	–	✓	✓	–	BU	Bi-directional ConvLSTM U-Net with densley connected convolutions
Gu et al. (2019)	CE-Net	Multi-modality	✓	-	✓	–	IU	U-Net based context encoder network
Abedalla et al. (2020)	2STU-Net	X-ray	✓	-	✓	✓	BU	Two stage U-Net with pre-trained ResNet-34 model
Zhang et al. (2020a)	DEFU-Net	X-ray	-	✓	✓	-	IU	U-Net with encoder fusion of dense and inception CNN
Wang et al. (2020a)	MDU-Net	X-ray	✓	-	✓	-	BU	Multi-task dense connection U-Net
Park et al. (2020)	3D U-Net	CT	-	✓	✓	✓	BU	3D U-Net with segmentation error correction
Fan et al. (2020b)	MA-Net	CT	-	✓	✓	-	AU	U-Net based multi-scale attention model
Dong et al. (2020)	DeU-Net	MRI	-	✓	✓	-	AU	3D deformable attention U-Net
Punn and Agarwal (2020d)	3D inception U-Net	MRI	-	✓	✓	-	EU	3D inception U-Net with modality fusion
Lu et al. (2020)	Modified U-Net	PET	✓	-	-	✓	BU	U-Net with pre-trained VGG-19 encoder
Leung et al. (2020)	Modified U-Net	PET	✓	-	✓	-	EU	Physics guided minimal U-Net with dropout regularization
Dunnhofer et al. (2020)	Siam-U-Net	US	-	✓	✓	-	EU	U-Net with siamese tracking framework
Zhang et al. (2020b)	AU-Net	US	-	✓	✓	-	AU	Attention guided U-Net with total variation regularization
Byra et al. (2020a)	SKU-Net	US	-	✓	✓	-	AU	Attention based selective kernel U-Net
Punn and Agarwal (2020c)	IU-Net	Histopathol-ogical	-	✓	✓	-	IU	Inception U-Net model with hybrid spectral pooling
Ibtehaz and Rahman (2020)	MR-UNet	Multi-modality	-	✓	✓	-	IU	MultiResUNet with multiple inception based skip connections
Wang et al. (2020b)	NL-Unet	Multi-modality	-	✓	-	-	AU	Non-local Unet with global context aggregation

Table 6 (continued)

Author	U-Net variant	Modality	TL	SL	Pr	Po	Category	Description
Xia et al. (2021)	MC-Net	CT	-	✓	✓	-	IU	Multi-scale context extraction with residual attention
Li et al. (2021)	MSA-Unet	MRI	-	✓	✓	-	AU	U-Net with dual branch multi-scale attention
Fu et al. (2021)	MSAM-Net	PET	-	✓	✓	-	AU	Multi-modal spatial attention network
Punn and Agarwal (2021b)	RCA-IUnet	US	-	✓	-	-	AU	Residual cross-spatial attention guided inception U-Net model
Cao et al. (2021)	Swin-Unet	Multi-modality	✓	-	-	-	EU	Unet-like pure transformer network
Wang et al. (2021)	MTM-Unet	Multi-modality	-	✓	-	-	EU	U-Net with mixed transformer module
Isensee et al. (2021)	nnU-Net	Multi-modality	-	✓	✓	✓	EU	Self-adapting no-newU-Net Framework

TL transfer learning, *SL* scratch learning, *Pr* pre-processing, *Po* post-processing

3.1.2 Attention U-Nets

Motivated by the success of squeeze-and-excitation network (SENet) (Hu et al. 2018) to suppress the irrelevant features, Li et al. (2019a) proposed an attention guided deep learning framework divided into three components: preprocessing, region of interest (RoI) segmentation with transfer learning followed by pneumonia detection model. In the preprocessing stage, apart from the trivial processes like resizing, the authors synthesized the adversarial samples to gain attention of the model towards pneumonia. The pneumonia infected area is erased by replacing it with an average pixel value of the image and then labelled as non-pneumonia, which helped to distinguish between noise and relevant data. To further suppress the background interference, authors adopted the approach proposed by Rashid et al. (2018) to perform the lungs segmentation followed by post-processing with conditional random fields. The segmented, original and synthesized images together form the training and validation set for the pneumonia segmentation network. The network follows SENet design in which SE-ResNet34 is utilized as a backbone architecture. The proposed framework tends to learn the pneumonia features effectively and achieves a significant reduction in the false positive predictions with an FPR value of 0.19, in contrast to mask R-CNN (He et al. 2017) and RetinaNet (Lin et al. 2017) on RSNA challenge; however, the overall framework relies heavily on the pre-processing and post-processing of data, thereby limiting its usability across multiple datasets.

3.1.3 Inception U-Nets

In another U-Net variant, Zhang et al. (2020a) proposed a DEFU-Net model that uses the fusion of dual encoder models to better extract the spatial features and a standard decoder network with upsampling. The dual encoder network is equipped with a densely connected recurrent convolutional (DCRC) neural network (inspired from DenseNet (Huang et al. 2017) and R2U-Net (Alom et al. 2018)) and dilated inception convolution neural network (inspired from GoogleNet (Szegedy et al. 2015)), where the output from each layer is merged by addition operation which is later concatenated with the corresponding decoder layer. The DCRC aids in extracting high level features, whereas the inception block facilitates to increase the network width and improve the horizontal feature representation using various receptive fields with dilated convolutions. The advantage of using dilated convolutions is that it tends to increase the receptive field without changing the number of training parameters (Yu and Koltun 2016). The significance of each module of the network is established by achieving considerable improvements over several U-Net variants such as residual U-Net (He et al. 2016b), BCDU-Net (Azad et al. 2019), R2U-Net and attention R2U-Net (Alom et al. 2018), etc. with dice score of 0.97 on the chest X-ray dataset.

3.1.4 Ensemble U-Nets

With cardiomegaly being one of the most common inherited cardiovascular diseases, Que et al. (2018) proposed a CardioXNet framework to identify and localize the cardiomegaly present in the chest X-ray images. CardioXNet is equipped with two parallel U-Net models to generate the segmentation masks for cardiac and thorax respectively, that follows typical CNN architecture in contraction and expansion paths. To address

the limited availability of the data samples, authors utilized data augmentation strategies such as rotation, zooming, shearing, etc. Due to the possibility of the presence of noise in the output masks, post-processing is applied to keep the binary contours that represent the largest area. Later, the processed output mask is utilized to compute the cardiothoracic ratio defined as $CTR = (L + R)/(T)$, where L and R indicates the maximum distances from the center to the left and right farthest boundaries of the heart region, and T is the maximum horizontal distance between the lungs boundaries. The CTR value is then utilized to determine the cardiomegaly from the generated masks. In another approach, Subramanian et al. (2019) proposed an automated system involving two U-Net models, where the output features are exploited to identify the type of central venous catheters (CVC) as peripherally inserted central catheters (PICC), internal jugular (IJ), subclavian and Swan-Ganz catheters. The first U-Net model is utilized for CVC segmentation by using the exponential logarithmic loss to address the class imbalance problem, whereas the other U-Net model tends to extract the anatomical structures to distinguish the ambiguous classes such as PICC and subclavian lines. Clinicians manually annotated the CVCs to obtain the signature spatial priors which undergo pixel-wise multiplication with the segmentation output. Later, the produced output is fed to the pre-trained neural network random forest (NN-RF) classifier to distinguish the type of CVC. This hybrid combination of segmentation and classification achieved promising results on the NIH database.

3.2 Computed tomography

Computed tomography imaging is based on the principle of utilizing the series of the system of rotating X-rays to develop cross-sectional images or series of slices of bones, blood vessels and soft tissues of the body (Bercovich and Javitt 2018). In contrast to plain X-ray imaging, CT scans provide rich information with high quality images. This is generally utilized to examine people with serious injuries or diseases like trauma, tumors, pneumonia, etc., and also to plan medical, surgical or radiation treatment. Hence, various deep learning based approaches are developed for faster diagnosis and treatment using CT imaging.

3.2.1 Better U-Nets

Tong et al. (2018) proposed a U-Net framework for lung nodule segmentation, where mini residual connections are introduced within the encoder and decoder phases to address the vanishing gradient problem. The algorithm initiates with the process of generating the segmentation of lung parenchyma with morphological operations and removal of irrelevant features. The segmented lung parenchyma images are divided into 64×64 slices along with the input images. Finally, the improved U-Net model is trained and validated against the preprocessed dataset for segmenting the pulmonary nodules. The authors evaluated the approach on LUNA2016 dataset against various models and achieved promising results with a dice score of 0.74; however, the samples of pulmonary nodules were very limited and the approach also lacked the 3D volumetric analysis. Recently, Park et al. (2020) utilized a 3D U-Net model to segment the lung lobe regions while also addressing the missdetection of the lobar fissure. Initially, the volumetric CT scans are preprocessed with thresholding to identify lungs parenchyma, and region growing techniques (Leader et al. 2003) to separate overlapping left and right lung regions. Later, these lobe segmentations

are generated with the help of the 3D U-Net model, where the segmentation results are further refined with the upsampling and segmentation error correction. The authors utilized CT volumes from multiple centres (hospitals) to evaluate the model performance while achieving significant improvements.

3.2.2 Attention U-Nets

When the target is the segmentation of the internal organs, then models adopting the attention mechanism help to focus the network on regions of interest. Oktay et al. (2018) proposed a novel attention gate based U-Net framework to focus on pancreas regions and generate the corresponding segmentation masks. The attention approach tends to suppress irrelevant features and highlight the prominent features corresponding to the target regions. The authors utilized the FCN with U-Net connectivity, where the skip connections are loaded with these attention filters. Inspired from the work of Shen et al. (2017), each pixel is associated with a gating vector to determine the regions to focus. The incorporation of this attention mechanism allowed the authors to achieve significant improvements in the segmentation results over other approaches. The performance of this model could easily be improved by incorporating transfer learning, multi-stage training, etc. To exploit the potential of attention mechanism, Seo et al. (2019) proposed a modified U-Net (mU-Net) framework that addressed the classical problems associated with the standard U-Net model concerning skip connection (Han and Ye 2018) and pooling operation (loss of spatial information). In the mU-Net, the standard skip connections are replaced by object-dependent filters to dynamically filter the feature maps based on the object size, where features concerning the small objects are preserved by blocking the deconvolution path and in the case of large objects, feature maps indicating boundary information is propagated to avoid duplication. The authors verified the effectiveness of adaptive filters to preserve the features using the permeation rate while achieving the *DSC* values of 0.98 and 0.89 on the liver and liver-tumor segmentation respectively. The approach could be extended for 3D volumetric analysis, where computation cost can be addressed by modifying operation schemes such as depthwise separable convolution instead of standard convolution. These object-dependent filters could easily be integrated with other networks and modalities for segmentation.

Song et al. (2019) proposed a U-NeXt model to segment CT images of gallstones, which is one of the common and frequently occurring diseases worldwide. The U-NeXt model is equipped with the attention up-sampling blocks, spatial pyramid pooling (He et al. 2015b) of skip connections (SkipSPP), and multi-scale feature extraction with the series of convolution layers along with the dense connections. The overall architecture design is similar to U-Net++ model (Zhou et al. 2018b) with slight variation in connections, convolution and pooling operations. The authors trained and evaluated the model on the proposed dataset with 5,350 images using deep supervision and reported improvement in IoU by 7% over baseline biomedical image segmentation models. However, for complex target structures, the network produces soft edges in the mask. To address this issue, a deep Q network (DQN) (Mnih et al. 2015) driven approach is proposed by Man et al. (2019) that uses deformable U-Net to efficiently generate the segmentation mask of the pancreas from CT scans with the extraction of its contextual information and anisotropic features. Initially, the 3D volumes are split into axial, coronal and sagittal 2D slices for each of which, DQN-based deep reinforcement learning (DRL) agents tend to localize the pancreas to form RoI slices. These slices are fed to the deformable U-Net models and finally, based on the

majority voting scheme 3D segmentation mask is generated. The deformable U-Net (Dai et al. 2017) follows standard encoder-decoder architecture, where convolution operations are replaced with deformable convolutions (DC). In DC, the regular convolution operation is accompanied by another convolution layer to learn 2D offset for each pixel. It leverages the deep network's ability to learn the required receptive field rather than being fixed for segmenting the regions having varying geometrical structures. This can also be understood as a learnable dilated convolution.

Unlike other U-Net variants that applies multi-scale feature fusion, Fan et al. (2020b) recently proposed a multi-scale attention U-Net model that uses a self attention scheme for adaptive feature extraction. The self attention design comprises position-wise attention block (PAB - installed on bottleneck layer) and multi-scale fusion attention block (MFAB - installed on every stage of encoder path), where PAB captures feature interdependencies in spatial dimension and MFAB captures the channel dependencies for any feature map. The MA-Net is trained and evaluated on the 2017 LiTS challenge and achieved a *DSC* score of 0.96 and 0.75 for liver and liver-tumor segmentation respectively. However, the results are not as promising as achieved using mU-Net model (Seo et al. 2019).

3.2.3 Inception U-Nets

Recently, Xia et al. (2021) proposed MC-Net that uses a multi-scale context extraction module with a context residual attention approach to model the local and global semantic information of the target regions using CT imaging and alleviate the problem of not capturing the long-range dependencies by most of the U-Net models. Overall, MC-Net is built with a multi-feature extraction module (MIE) to obtain multi-scale feature maps similar to inception network (Szegedy et al. 2017), context information extraction (CIE) with parallel dilated convolutions and residual attention enabled skip connections. Though the model achieved promising results over multiple CT datasets; however, the diversity of multi-scale feature extraction is limited to a feature map at a single level, which could be improved by considering feature maps across different encoding layers.

3.2.4 Ensemble U-Nets

Janssens et al. (2018) proposed a cascaded 3D FCN based deep learning model consisting of "LocalizationNet" and "SegmentationNet" to estimate the bounding box (RoI) and generate volumetric segmentation masks of lumbar vertebrae respectively. The LocalizationNet comprises a 3D FCN regression model which is trained to regress the displacement vectors associated with a voxel, representing diagonal corners of the rectangular box. The localized information is fed to SegmentationNet comprising an FCN 3D U-Net model to produce a segmentation mask for lumbar vertebrae. This two stage approach exhibited significant improvement over the existing approaches but with the overhead computations of two dedicated models. In the real world scenario, modalities may suffer from inherent ambiguities that coagulate the actual nature of the disease. Following this, Kohl et al. (2018) introduced a probabilistic U-Net framework that combines the standard U-Net model with conditional variational autoencoder (CVAE). For a sample image, CVAE generates diverse plausible hypotheses from a low-dimensional latent space which are fed to U-Net to generate the corresponding segmentation mask. It is shown that the model can generate diverse segmentation samples, given the ground-truth delineation from multiple

experts. The trained model is evaluated on LIDC-IDRI and Cityscapes datasets which outperformed other approaches in reproducing the segmentation probabilities and masks. Inspired by this work many other variants have been developed to capture the uncertainties, e.g. (Raghu et al. 2019; Baumgartner et al. 2019; Tanno et al. 2019).

Motivated by the success of adversarial techniques, Dong et al. (2019b) proposed a U-Net-GAN framework in which a set of U-Nets is trained as a generator to produce organs-at-risk (OARs) segmentation and FCN as a discriminator to distinguish segmented masks from the ground-truth masks. The generator and discriminator networks followed adversarial training, where each network competes to achieve optimal segmentation masks of OARs. The model achieved satisfactory improvements at the cost of heavy computations and resource requirements, moreover, the model struggles in the presence of complex structures. In another work, Liu et al. (2019c) proposed a liver CT image segmentation framework named GIU-Net, inspired by the supervised interactive segmentation approach named, graph cut (Boykov and Funka-Lea 2006). The improved U-Net model is designed with increased depth to better extract semantic features that are trained to generate the segmentation mask of the liver regions. Later, to further refine the segmentation results, a slice covering the maximum liver region is used as an initial slice to generate graph cut energy function followed by the maximum flow minimum cut algorithm. The process is then repeated for all the slices to generate a complete sequence of precise and stable segmentation masks with smother boundaries.

In another application, a Bayesian CNN with U-Net model and Monte Carlo (MC) dropout is introduced by Hiasa et al. (2019) for automated muscle segmentation from CT imaging for musculoskeletal modelling. The design comprises two cascaded U-Net models, where first is standard U-Net that localizes the skin surface and later individual muscles (21 muscles) are segmented with Bayesian U-Net (Kendall et al. 2015) that uses MC dropout based on the structure-wise uncertainty, predictive structure-wise variance (PSV) and predictive dice coefficient (PDC). Besides, the authors employed an active learning method to produce segmentation and uncertainty from the unlabeled data, where the high uncertain data are relabeled manually by experts while other data is directly used as training data. The authors achieved promising results; however, the data samples were very limited which limits the diversity of the model.

3.3 Magnetic resonance imaging

Magnetic resonance imaging is synthesized by using the principles of nuclear magnetic resonance (NMR) (Morris and Slesnick 2018). It is utilized in radiology to visualize the anatomy and physiological process of the body organs. It uses a large magnetic field and radio waves to create detailed images of organs and tissues within the body. Based on the different attenuation values of the tissues e.g. T1-weighted (T1), fluid attenuation inversion recovery (FLAIR), Dixon, etc., the electromagnetic waves emitted from the gradient magnetic field is detected using the applied strong magnetic field by which the position and type of the nucleus can be drawn inside the object. Unlike the X-rays, CT scans and PET scans; MRI scans do not involve the usage of ionizing radiations.

3.3.1 Better U-Nets

MRI is mostly utilized in computer-aided diagnosis systems involving brain tumor segmentation. Inspired from the BraTS 2015 challenge, Dong et al. (2017) analysed the potential

of FCN based U-Net model for brain tumor segmentation via MRI sequences, where the authors achieved significant improvement over the traditional segmentation approaches. SegNet is another model that is most widely used for semantic segmentation (Badrinarayanan et al. 2017). Following this, Kumar et al. (2018) proposed a hybrid approach, U-SegNet, by integrating skip connections into the base SegNet model. This enabled the model to efficiently identify the tissue boundaries concerning the white matter (WM), gray matter (GM), and cerebro-spinal fluid (CSF). The authors achieved significant improvement over the base SegNet and U-Net models with *DSC* value of 0.90 on IBSR-18 dataset.

In recent years, to improve the biomedical image segmentation results, multi-modality fusion (MMF) (James and Dasarathy 2014) approaches are utilized. The fused scans are rich in information and offer multi-dimensional features. In this context, Kermi et al. (2018) proposed a modified U-Net model to segment the whole tumor and intra tumor regions like enhancing tumor, edema and necrosis affected with high grade glioma (HGG) and lower grade glioma (LGG) following from the BraTS 2018 challenge. The authors fused the T1, T2, T1c and FLAIR modalities and resized them to form the input feature map with rich tumor information. In the modified model, residual blocks (He et al. 2016b) are added between two convolution blocks and the max-pooling operation is replaced with the strided convolutions (Ayachi et al. 2018). The model is trained and evaluated with fused modalities to obtain the multi-class segmentation masks. Though the authors achieved good results but lacked the 3D volumetric analysis. In another application, skull stripping is an essential step to study brain imaging, where Hwang et al. (2019) proposed to utilize a standard 3D U-Net model to automate the process of skull stripping (brain extraction) from T1 MRI scans for faster diagnosis and treatment. The training is carried with dice loss and adam optimizer on neurofeedback skull-stripped (NFBS) dataset. The authors achieved a dice value of 0.99; however, the comparative study is limited to brain surface extractor (BSE) and robust brain extraction (ROBEX) algorithms.

3.3.2 Attention U-Nets

With the introduction of modality transformations, Dong et al. (2019a) proposed a deep attention U-Net (DAU-Net) model to automate the process of multi-organ segmentation for prostate cancer diagnosis via synthetic MRI, that is generated by processing the computed tomography scans using a cyclic generative adversarial network (CycleGAN) (Zhu et al. 2017). Initially, the CycleGAN model is trained to learn CT to MRI transformation which tends to add additional soft-tissue information without additional data acquisition techniques to produce sMRI data. Later, the sMRI data is used to train 3D DAU-Net model which incorporates conventional attention scheme (Oktay et al. 2018) and deep supervision (Wang et al. 2019a) with the U-Net model. The approach is trained and evaluated with 140 datasets from prostate patients to achieve *DSC* value of 0.95, 0.87 and 0.89 for segmentation of bladder, prostate and rectum respectively, while also showing improvement over using raw CT images.

The prostate cancer diagnosis is another challenging task for which Rundo et al. (2019) proposed an automated approach, USE-Net, that uses the U-Net model by incorporating the squeeze-and-excitation (SE) blocks (Hu et al. 2018) in skip connections to perform multi-class segmentation. Similar to the attention scheme (Oktay et al. 2018), the SE blocks tend to calibrate the channel-wise correlation while improving the generalization capability of the model across multi-institutional datasets. The USE-Net model outperformed its competitors when trained and evaluated on all datasets combined, where for other scenarios

(individual dataset and mixed datasets), USE-Net struggled to achieve better results. Dong et al. (2020) integrated 3D U-Net model with deformable convolutions (Dai et al. 2017) for cardiac MRI segmentation. The deformable U-Net (DeU-Net) includes a temporal deformable aggregation module (TDAM) to generate fused feature maps using an offset prediction network. The fused feature maps are then fed to deformable global position attention (DGPA) network to map the multi-dimensional contextual information into generalized and localized features. The proposed approach outperformed other models to generate efficient segmentation masks involving subtle structures. Recently, Li et al. (2021) proposed multi-scale attention enabled U-Net model (MA-Unet) to segment lumbar spinar using MRI. The dual branch multi-scale attention block represents the most relevant feature maps at different target scales by using aggregating features obtained from two branches, where the first branch works as a multi-scale feature extraction block by using dense connections and the second branch uses the channel and spatial attention network to suppress irrelevant information. The authors achieved promising results; however, the introduction of MSA block across every stage in the entire network results in expensive computation and resources which could be reduced by using depthwise separable convolutions.

3.3.3 Inception U-Nets

Chen et al. (2018b) improved the performance of the vanilla 3D U-Net model by adding spatiotemporal-separable 3D convolutions (Xie et al. 2018) to form S3DU-Net model. The S3D convolution involves two convolution layers i.e. 2D convolution operation to extract spatial features and then additional 1D convolution to learn temporal features, furthermore, inception (Szegedy et al. 2015) and residual connections (He et al. 2016a) are added to better learn the complex patterns. The S3DU-Net model is trained with dice loss and evaluated on dice coefficient and Hausdorff distance metrics. The authors achieved average dice scores of 0.69, 0.84 and 0.78, for enhancing tumor, whole tumor and tumor core respectively on BraTS 2018 challenge. For real-time applications, Wang et al. (2019b) proposed a multiscale statistical U-Net (MSU-Net) to segment cardiac regions in MRI. The MSU-Net incorporates statistical CNN (SCNN) (Wang et al. 2019c) to fully exploit the temporal and contextual information present in various channels of an input image or feature map along with the multiscale parallelized data sampling approach. For multi-scale data sampling, independent component analysis (ICA) (Wang et al. 2019c) is applied over the patches of data to form clusters of canonical form distributions which represent spatio-temporal correlations at coarser scales. This data sampling parallelization tends to speed up the performance significantly by 26.8% as compared to the standard U-Net model and achieved an increased dice score by 1.6% on ACDC MICCAI 2017 challenge, while also improving significantly over state-of-the-art GridNet (Zotti et al. 2018) model.

3.3.4 Ensemble U-Nets

Wang et al. (2019a) proposed a 3D FCN model with deep supervision and group dilation (DSD-FCN model) to address various challenges concerning the automated MRI prostate segmentation like inhomogeneous intensity distribution, varying prostate anatomy, etc., which makes it hard for manual intervention. The proposed architecture follows vanilla U-Net topology in which deep supervision is adopted to learn discriminative features, whereas group dilated convolutions tend to acquire multi-scale contextual information. The model is trained with the objective function defined as the weighted average of

cosine similarity and cross entropy using the manually annotated institutional dataset and MICCAI PROMISE12 dataset, where authors achieved the *DSC* values of 0.86 and 0.88 respectively. However, this achievement comes at the cost of complex computations due to group dilated convolutions (Wang et al. 2018a). Recently, Punn and Agarwal (2020d) proposed a 3D U-Net based framework for volumetric brain tumor segmentation. The proposed architecture is divided into three components: 1) multi-modalities fusion - to merge the MRI sequences with hierarchical inception convolution blocks, 2) tumor extractor - to learn the tumor patterns with 3D inception U-Net model using fused modalities, and 3) tumor segmenter - to decode the multi-scale extracted features into multi-class tumor regions. To achieve a better understanding of the input feature maps, each inception block aggregates multi-scale feature representation by using multiple filters equipped with short skip connections. With such dedicated components trained using a weighted average of dice and IoU loss functions, the authors achieved significant improvement over the existing approaches for BraTS 2017 and BraTS 2018 datasets. The increased computations due to inception convolution in each module could be reduced by using depthwise separable convolutions (Chollet 2017).

3.4 Positron emission tomography

The positron emission tomography (Ollinger and Fessler 1997) is a widely used imaging in various clinical applications like oncology, brain, heart, etc., that helps in visualizing the biochemical and physiological reaction processes within the human body. The PET images are obtained by injecting a full dose of radioactive tracer or inhalation of gas to meet the clinical requirements. However, for minimal harm to human health, low-dose PET imaging is adopted to produce high quality imaging (Wang et al. 2018c).

3.4.1 Better U-Nets

With the huge success of U-Net in biomedical image segmentation, Blanc-Durand et al. (2018) demonstrated the potential of 3D U-Net model in ^{18}F-fluoro-ethyl-tyrosine (^{18}F-FET) PET lesion detection and segmentation. F-FET PET/CT scans were acquired using a dynamic protocol from 37 patients, where the ground-truth segmentation masks were generated using manual delineation and binary thresholding. The 3D U-Net model comprises three stages of encoder and decoder paths with standard convolutions and pooling operations. The authors achieved a *DSC* value of 0.79 on training and validation sets. However, the results could further be improved by increasing the data size with GAN based data augmentation techniques (Frid-Adar et al. 2018b) and other U-Net based approaches. Recently, Lu et al. (2020) proposed U-Net based automatic tumor segmentation approach in PET scans. The authors employed a transfer learning approach, where pre-trained VGG-19 blocks are added in the encoder phase to address the challenge of limited data availability. The authors adopted the DropBlock as a replacement for dropout to effectively regularize the convolution blocks. The model is fine-tuned using the Jaccard distance (IoU) as the loss function and the performance is validated with 1,309 PET images provided by the Shanghai Xinhua hospital (XH), which displayed improvements over the vanilla U-Net model.

3.4.2 Attention U-Nets

The integration of PET and CT modalities offer metabolic and anatomical information simultaneously. High contrast in PET scan enables the network to effectively segregate soft tissues around the tumor boundaries that are identified using CT imaging with high spatial resolution. In this context, Fu et al. (2021) proposed multi-modal spatial attention module (MSAM) to segment tumor using PET-CT modalities. The MSAM block can be integrated with any U-Net model to leverage the multi-modalities features. Two individual encoder-decoder models are used, where one model is trained with PET imaging to generate spatial attention maps and another model utilizes this attention map at different scales by aggregating with each decoder layer to perform tumor segmentation using CT imaging. These PET based multi-scale attention maps guide the CT model towards the areas with high tumor likelihood. This approach of cross-modality multi-scale attention achieves promising results; however, the performance could further be improved by performing 3D volumetric analysis.

3.4.3 Inception U-Nets

Zhao et al. (2018) proposed to utilize the multi-modalities (PET and CT) for computer-aided cancer diagnosis and treatment with the help of 3D FCN based V-Net (Milletari et al. 2016) model, which is an extension of the U-Net model for volumetric segmentation. A feature or intermediate level fusion approach is adopted, where two independent sub-segmentation networks are constructed to extract dedicated feature maps from each modality and are later fused with the cascaded convolution blocks that follow the V-Net model scheme to finally compute the tumor segmentation mask. The proposed framework is trained and validated on a limited clinical dataset of 84 patients suffering from lung cancer that consists of PET and CT imaging, where a dice value of 0.85 is achieved while outperforming other traditional models that use unary modality. In a similar approach, Guo et al. (2019) adopted the fusion of PET and CT modalities to segment head and neck cancer (HNC) labelled as gross tumor volume (GTV). Due to resources limitations, the images are cropped during pre-processing. The authors utilized the modified 3D U-Net model in which the convolution blocks in encoder and decoder paths are replaced by dense convolution blocks (Huang et al. 2017) that aggregates the multi-scale feature maps across various levels. The authors trained and evaluated the model on TCIA-HNC dataset, while achieving the *DSC* value of 0.73 on the dedicated test set. The performance could be improved by analyzing complete 3D volume and obtain better affinities.

3.4.4 Ensemble U-Nets

To address the need for a reliable and robust PET based tumor segmentation model, Leung et al. (2020) proposed a novel physics guided deep learning based framework comprising three dedicated modules that segment each slice of PET volume to generate a complete mask. The first module tends to extract the realistic tumors with the available ground-truth boundaries via stochastic kernel-density estimation and physics based approach to generate simulated images. These images are fed to the improved U-Net model in the second module, which has minimal convolution and pooling blocks accompanied by dropout layers to aid in learning the complex features and generate efficient masks. Later, in the third module, the network is fine-tuned with delineation provided by the radiologist as surrogate

masks to improve the learned features. The proposed framework achieved dice scores of 0.87 and 0.73 to segment primary tumors on simulated and patient images and outperformed several semi-automated approaches. This approach could perform 3D segmentation by generating a mask for each slice, but avoids the 3D correlations of the voxels which is crucial for real-time applications.

3.5 Ultrasound

Ultrasound is acoustic energy in the form of waves having a frequency beyond the human hearing range. These are generated with the help of piezoelectric crystals which deform under the influence of electric field and generate compression waves when an alternating voltage is applied. Ultrasonography (Moore and Copel 2011) is an ultrasound based diagnostic imaging technique used for visualizing the internal body organs by processing the reflected signals. The deep learning technologies aid in diagnosing US imaging to segments regions of interest like breast mass, pelvic floor levator muscle discontinuity, etc.

3.5.1 Better U-Nets

In consideration of breast cancer being the deadliest cancer among women, Almajalid et al. (2018) proposed an automatic breast ultrasound (BUS) image segmentation system to aid in its diagnosis and treatment. The authors utilized the vanilla U-Net model on the preprocessed BUS images. The images are preprocessed using the contrast enhancement with histogram equalization and noise reduction with speckle reducing anisotropic diffusion (SRAD) (Yu and Acton 2002) techniques to improve the image quality. Finally, with the assumption of the presence of a single tumor region the authors filtered the false positive regions to remove the noisy regions. This assumptions limits the capability of the model to generate masks for multiple tumor regions. In this regard, Li et al. (2019b) incorporated dense connections in the U-Net model (DenseU-Net) to efficiently segment levator hiatus from ultrasound images. The implication of dense connections enabled feature reuse and reduction in the trainable parameters. The DenseU-Net model is trained to generate the binary segmentation mask which is post-processed with binary thresholding to generate mask sharp boundaries, and localized regions are generated with active contour model (Li et al. 2016) without compromising the performance of the model.

3.5.2 Attention U-Nets

In another application of eyeball segmentation, Lin et al. (2019) proposed a semantic embedding and shape-aware U-Net model (SSU-Net), where the authors employed a signed distance field (SDF) instead of a binary mask as the label to learn the shape information. In addition, the model is equipped with a semantic embedding module (SEM) to fuse the semantic information at coarser levels of the SSU-Net model. The SEM block draws features from two low-level stages and one corresponding stage, where lower level features are convolved and bilinear interpolation is applied to restore the resolution at the same scale. This enabled the network to efficiently identify the ambiguous and discontinuous boundaries and achieved better segmentation performance. Due to the low signal to noise ratio (SNR) in US imaging, real-time analysis is still a challenging task. Recently, Zhang et al. (2020b) proposed a U-Net based deep learning approach to realize the multi-needle

segmentation in the 3D transrectal US (TRUS) images of high dose rate (HDR) prostate brachytherapy. The U-Net model is loaded with the attention scheme in the skip connections to address the challenge of identifying the smaller needles, while spatial continuity of the needles is maintained with total variation regularization. The model is trained with a deep supervision approach, where patches of needle masks are generated to compute the cross entropy loss and accordingly optimize the training weights. With the proposed framework, the authors achieved adequate performance gain on multi-needle segmentation for prostate brachytherapy.

Byra et al. (2020a) proposed a selective kernel U-Net (SKU-Net) model for breast mass segmentation in US imaging while also addressing the challenge of variable breast mass size and image properties. In SKU-Net, each convolution layer of the U-Net model is replaced by an SK block, that tends to dynamically adapt the receptive field. Similar to the concept of dual path U-Net (Yang et al. 2019), the SK module (Li et al. 2019c) is designed using two branches, where one uses dilated convolutions and other is without dilation to generate feature maps. Later, these features are merged and global average pooling, followed by FC layer and sigmoid activation is applied to construct attention coefficients for each channel in the feature map. With this approach, authors achieved significant improvement over the vanilla U-Net model across multiple datasets. In another work, Punn and Agarwal (2021b) proposed residual cross-spatial attention (CSA) block in the skip connections of inception U-Net model (Punn and Agarwal 2020c) to further improve the segmentation performance. The authors validated the performance of the model with breast cancer segmentation using ultrasound imaging. In contrast to standard attention (Oktay et al. 2018), the CSA block uses multi-level encoded feature maps to obtain better spatial correlation and develop spatial attention feature maps that are concatenated with corresponding decoder block to reconstruct the multi-scale spatial information. To address computational and memory challenges model uses depth-wise separable convolutions. The model achieved promising results on multiple US datasets and establishes scope for further utilization of this model across different modalities.

3.5.3 Inception U-Nets

In another approach, Yang et al. (2019) proposed a dual path U-Net model for segmentation of lumen and media-adventitia from the IntraVascular UltraSound (IVUS) scans to aid in cardiovascular diseases diagnosis. Due to the limited availability of the data samples, the DPU-Net is trained with the real-time augmentor that generates and integrates three types of artefacts: bifurcation, side vessel, and shadow, and other common augmentation operations with training images. In contrast to vanilla U-Net, DPU-Net involves multi-branch parallel encoding and decoding operations, where feature maps are extracted and reconstructed with different kernel sizes at the same hierarchical level to address the challenge of a large variation in shape and size of lumen or media region. With this network-in-network architecture and real-time augmentation approach, the authors achieved Jaccard measure (IoU) of 0.87 and 0.90 on 40 MHz and 20 MHz frames respectively from IVUS dataset.

3.5.4 Ensemble U-Nets

Wang et al. (2018b) proposed a multi-feature guided CNN model for classification and segmentation of the bone surfaces in the US scans. The US images are initially processed with a pre-enhancing (PE) net to synthesize a US scan that highlights the bone surface, by using

a B-mode US scan and three filtered image features, including local phase tensor image (LPT), local phase bone image (LB) and bone shadow enhanced image (BPE). The feature enriched images are then used by a classification embedded U-Net model (cU-Net) to produce the segmentation mask and identify the type of the bone surface. This multi-task deep learning framework achieved promising segmentation and classification results with F1-score of 0.96 and 0.90 on SonixTouch and Clarius C3 datasets respectively. In another application area, Dunnhofer et al. (2020) emphasized on the tracking of knee femoral condyle cartilage during ultrasound guided invasive procedures. The Siam-U-Net model combines the potential of the U-Net model with siamese framework (Gomariz et al. 2019) for tracking the cartilage in the real-time ultrasound sequences. In Siam-U-Net two encoder blocks are adopted which are fed with resized-cropped US sequences named as, searching area and target cartilage. After five blocks of encoding layers, the acquired feature maps of two inputs are cross-correlated using convolution operation applied to searching area feature maps with target embedding as a filter, which results in localizing the implicit position of the cartilage in the searching area slice. Later, the slice is reconstructed in the decoder phase to generate the segmentation mask of the cartilage. The Siam-U-Net model achieved an average dice score of 0.70 with significant improvement over other approaches. However, the results could further be improved by expanding the dimension of the model into 3D space for considering the neighbouring voxels correlation.

4 Other U-Net variants and imaging

In this section, various U-Net variants are presented that are introduced as the biomedical image segmentation networks, where each model acts as a generic architecture that is trained and evaluated on multiple/different modalities.

4.1 Better U-Nets

In the growing phase of biomedical image segmentation, Alom et al. (2018) integrated the potential of multiple state-of-the-art deep learning models such as recurrent CNN (Mikolov et al. 2011), residual CNN (He et al. 2016a) and U-Net to form RU-Net and R2U-Net for BIS. In the RU-Net model, the standard convolution and up-convolution units are improved by incorporating recurrent convolutional layers (RCL), whereas in R2U-Net both RCL and residual units are added. These models are trained and evaluated on three different modalities such as retina blood vessel segmentation (DRIVE, STARE, and CHASH-DB1 datasets), skin cancer segmentation (ISIC 2017 Challenge), and lung segmentation (KDSB 2017 challenge). Zhou et al. (2018a) proposed a nested U-Net architecture, U-Net++, to narrow down the gap between the encoded and decoded feature maps. In contrast to the U-Net model, U-Net++ model follows convolutions on dense and nested skip connections to effectively capture the coarser details. Furthermore, a deep supervision approach is adopted to prune the model based on the loss (combined binary cross entropy and dice coefficient) estimated at different semantic levels. The performance of the model is validated with multiple datasets involving KDSB18, ASU-Mayo, MICCAI 2018 LiTS Challenge and LIDC-IDRI, while outperforming other models. Azad et al. (2019) proposed another extension of U-Net, where bi-directional ConvLSTM (BConvLSTM) with densely connected convolutions (BCDU-Net) is introduced for BIS. The skip connections are equipped with BConvLSTM (Song et al. 2018) to concatenate the feature maps between the encoded layer

and the corresponding decoded layer. Furthermore, the dense connections are added at the bottleneck layer to extract and propagate features with minimal parameters. The authors achieved promising results across DRIVE, ISIC 2018 and LUNA datasets.

4.2 Inception U-Nets

Gu et al. (2019) addressed the loss of spatial information while using the strided convolutions and pooling in U-Net with context-encoder network (CE-Net) to capture and preserve the information flow for BIS. In CE-Net the encoder unit is loaded with pre-trained ResNet blocks, the bottleneck layer (context extractor) includes dense atrous convolution (DAC) and residual multi-kernel pooling (RMP) blocks, and decoder block follows consecutive convolution and deconvolution blocks. The DAC module combines the design of Inception-ResNet-V2 model and atrous or dilated convolution, whereas RMP generates stacked feature maps followed from the pooling operations with varying window sizes to effectively model the target feature representations. This arrangement of operations achieved promising results on multiple modalities.

For histopathological image segmentation, Punn and Agarwal (2020c) proposed an inception U-Net model where standard convolution layers are replaced by inception blocks that consist of parallel convolutions of varying filter sizes and a hybrid pooling operation. The hybrid pooling operation draws the potential feature maps from the spectral domain via Hartley transform (Zhang and Ma 2018) to preserve more spatial information and spatial domain with the help of max pooling to aim for sharp features, by using the 1×1 convolution. The authors achieved significant improvement over other models using KDSB18 dataset with less number of parameters. Ibtehaz and Rahman (2020) proposed another extension of the U-Net model as MultiResU-Net, where the convolution operations are replaced with MultiRes blocks in encoder-decoder paths, and Res path is added in the bottleneck layer. Inspired from the inception and residual model, the MultiRes blocks are built using stacked convolutions with a succession of 3×3 filters, and a residual 1×1 convolution connection is added. The Res path tends to propagate the feature maps from the encoder phase to decoder phase with the series of residual convolution blocks. The model is evaluated on different datasets covering fluorescence images, ISBI-2012, ISIC-2017, CVC-ClinicDB and BraTS17.

4.3 Attention U-Nets

In most of the U-Net models, the long-range dependencies are gradually acquired with local convolutions which limit the overall efficiency and effectiveness of the model. Inspired from the transformer models (Vaswani et al. 2017), Wang et al. (2020b) proposed non-local Unet (NL-Unet) that comprises of self attention based global context aggregation module to extract full context information which can be easily integrated with feature extraction and reconstruction operation in any U-Net model. The traditional spatial (Oktay et al. 2018) and channel attention (Hu et al. 2018) mechanisms lack to establish correlations with different targets and features, due to which non-local attention based models exhibit potential to perform better in biomedical image segmentation.

4.4 Ensemble U-Nets

With the immense application of U-Net model in the medical domain, Isensee et al. (2021) proposed a self-adapting framework, no-newU-Net (nnU-Net) to establish the generalized architecture and training mechanism for vivid modalities, inspired by the medical segmentation decathlon (MSD) challenge. The nnU-Net framework comprises an ensemble of 2D U-Net, 3D U-Net and 3D U-Net cascade, along with an automated pipeline to adapt the requirements of the dataset such as preprocessing, data augmentation and post-processing. The model achieved state-of-the-art segmentation results without manual intervention for different modalities in the medical segmentation decathlon challenge.

With the recent success of transformer models in sequence-to-sequence modelling (Vaswani et al. 2017), it has also been integrated with U-Net for medical image segmentation and has achieved satisfactory results. In vanilla U-Net, there is limited learning of global context information due to the local convolutions and hence, it cannot capture long-range dependencies. To address this challenge, Cao et al. (2021) proposed Swin-Unet which is a Unet-like pure transformer architecture. This model uses Swin transformer (Liu et al. 2021) with shifted windows as the encoder for feature extraction and patch-expanding Swin transformer for restoration of image resolution. In similar context, to further improve the performance, Wang et al. (2021) introduced a mixed transformer module (MTM) in U-Net that refined the self attention mechanism by simultaneously obtaining intra- and inter-correlations while using local-global Gaussian-weighted self attention (LGG-SA) and external attention, respectively. These MTM blocks are arranged in U-Net topology for medical image segmentation. Though these models achieved better results; however, rely on large-scale pre-training and have high computational complexity.

5 U-Net in COVID-19 diagnosis

The ongoing pandemic of the severe acute respiratory syndrome - coronavirus (SARS-CoV-2) also known as COVID-19 has brought the worldwide crisis along with the rampant loss of lives. This contagious virus initiated from Wuhan, the People's Republic of China in December 2019 and till November 17, 2021, have caused 254,174,536 infections and 5,112,325 deaths worldwide (Hopkins 2020). Currently, the most reliable COVID-19 diagnosis approach follows the reverse-transcriptase polymerase chain reaction (RT-PCR) testing; however, it is time consuming and less sensitive to identify the virus at the early stages.

With the advancements in the technology and data acquisition systems (Agarwal et al. 2020; Shi et al. 2020), deep learning based approaches are developed to assist in the COVID-19 diagnosis with the help of CT and X-ray modalities (Huazhu et al. 2020) to control the exponential growing trend (Punn et al. 2020a) of the spread. Wu et al. (2020) proposed a JCS framework (similar to cU-Net) for joint classification and segmentation of COVID-19 from chest CT scans using the U-Net model. In another U-Net based implementation, a feature variation block is introduced in the COVID-SegNet model (Yan et al. 2020) to better segment the COVID-19 infected regions by highlighting the boundaries and diverse infected regions. The lung infection segmentation deep network (Inf-Net) (Fan et al. 2020a) followed U-Net topology with diverse modifications including reverse attention and parallel partial decoder. The authors validated the performance in the supervised and semi-supervised modes to address the challenge of limited availability of the

labelled data. Recently, Punn and Agarwal (2020b) introduced a hierarchical segmentation approach, CHS-Net that involves two cascaded residual attention inception U-Net (RAIU-Net) models, where first generates lungs contour, which is fed to the second model to identify COVID-19 infected regions using CT images. The RAIU-Net model is designed with a residual inception U-Net model and spectral-spatial-depth attention blocks. The authors achieved promising results in generating the infected segmentation masks.

Furthermore, similar approaches are also developed for X-ray imaging for the screening of COVID-19 (Punn and Agarwal 2020a). Zahangir Alom et al. (2020) proposed a robust classification and segmentation framework of coronavirus infected X-ray and CT images, where classification is performed using inception residual recurrent convolutional neural network (IRRCNN) with transfer learning and NABLA-N model is used for localizing the infected regions. In addition, other deep learning based application areas are also explored to control the spread of the virus such as automated social distancing monitoring (Punn et al. 2020b), mask detection (Chowdary et al. 2020), etc. Furthermore, the survey of deep learning based approaches for COVID-19 diagnosis (Shi et al. 2020) reveals the significant impact of U-Net for CAD systems. Following these developments, it is believed that these artificial intelligent approaches will continue to evolve and contribute towards the faster and efficient diagnosis of the coronavirus.

6 Analysis

Over the years, the advancements in deep learning and computer vision techniques have attracted many researchers to contribute to the healthcare domain with a variety of tasks e.g. classification, detection, segmentation, etc. With segmentation being a critical task that drives the diagnosis process (Hesamian et al. 2019), researchers have developed a keen interest to develop a computer-aided diagnosis system to speed up the treatment process.

Among the published approaches or frameworks, U-Net appears to be the prominent choice (Minaee et al. 2020) to develop novel architectures to adapt multiple modalities with optimal segmentation performance. Following such high utility of the model, this article presented the recent developments in U-Net based approaches for biomedical image segmentation. Due to the high mutability and modularity design, U-Net topology can easily be integrated with other state-of-the-art deep learning models such as AlexNet (Krizhevsky et al. 2012), VGGNet (Simonyan and Zisserman 2014), ResNet (He et al. 2016a), GoogLeNet (Szegedy et al. 2015), MobileNet (Howard et al. 2017), DenseNet (Huang et al. 2017), etc., to produce the desired results depending on the application. This ease of integration opens a wide spectrum of applications for U-Net with endless possibilities of novel architecture designs. In the most recent developments of U-Net based biomedical image segmentation models following observations are drawn:

- More emphasis is given to multi-scale feature extraction and fusion to explicitly model global and long-range feature dependencies.
- Inspired by the state-of-the-art performance of self attention mechanism in transformer models many transformer based U-Net variants are utilized to enhance its capability to capture global contexts.

- For the training phase, most models employed a hybrid loss function that combines the binary cross entropy loss with dice similarity coefficient loss or with Jaccard loss, which tends to better penalize the false positive and false negative predictions
- Considering the implementation strategies mostly authors applied an end-to-end training-from-scratch approach with minimal pre-processing i.e. resizing and normalization and without any post-processing.
- Mostly depthwise separable convolutions are employed to reduce the overall number of computations and training parameters of the model.
- Multi-modality fusion based approaches are also developed for better feature representation learning concerning target regions.

From the reviewed articles it is observed that some of the segmentation approaches utilize the local dataset (datasets that are not publicly accessible), which tend to limit their reusability and reachability. In order to develop a widely acceptable solution, the summary of most widely utilized publicly available datasets for BIS is provided in Table 7. These benchmark datasets aid the research community to validate the existing performance and propose further improvements. Among the reviewed articles, CT and MRI modalities cover a wide range of U-Net variants for biomedical image segmentation. Moreover, for PET scan and ultrasound imaging most of the proposed approaches are validated on the local dataset, whereas for X-rays the approaches aim to localize the target structure with the bounding boxes. Despite such variants, it is difficult to conduct an effective comparative analysis of the results because each approach is evaluated with different evaluation metrics such as accuracy, F1-score, Jaccard index, etc. However, among these metrics, dice similarity coefficient is most widely utilized to quantize segmentation performance.

Considering the present survey it is also observed that each modality requires a different approach to address the corresponding challenges. Though there are segmentation approaches that are validated on multiple modalities to form generic architectures like nnUNet, U-Net++, MR-Unet, etc. but it is difficult to achieve optimal performance in all segmentation tasks. The main reason is due to the diverse variation in the features corresponding to the target structures involving lungs nodule, brain tumor, skin lesions, retina blood vessels, nuclei cells, etc. and hence require different mechanisms (dense, residual, inception, attention, fusion, etc.) to integrate with U-Net model to effectively learn the complex target patterns. Moreover, the presence of noise or artefacts in different modalities adds another factor to propose different segmentation methods.

7 Scope and challenges

Deep learning technologies have played a vital role in advancements towards medical diagnosis and applications. Generally, deep learning based technologies such as U-Net aims to develop CAD systems to achieve the desired results with minimal error. Despite U-Net based models being superefficient for biomedical image segmentation, there are various challenges involved, as shown in Fig. 8 with general and modality specific challenges, for developing the real-world implications of the deep learning models. With regular advancements in deep learning, these challenges are tackled with hardware and software oriented approaches which consequently attracts researchers to develop novel architectures and frameworks for biomedical image segmentation.

Table 7 Summary of popular BIS datasets

Dataset	Description	Availability
ISBI 2012	Electron microscopy cell slides for cell segmentation	http://brainiac2.mit.edu/isbi_challenge/
CVC-ClinicDB	Endoscopic colonoscopy frames for polyp detection	https://www.kaggle.com/balraj98/cvcclinicdb
ISBI	2D and 3D videos of moving cells for cell tracking	http://celltrackingchallenge.net/
KDSB 2018	Histopathological cell images for nuclei segmentation	https://www.kaggle.com/c/data-science-bowl-2018
PanNuke	Histopathological slides for nuclei segmentation	https://jgamper.github.io/PanNukeDataset/
DRIVE	Retinal fundus images for vessel extraction	https://drive.grand-challenge.org/
STARE	Retinal fundus imaging for blood vessel segmentation	http://cecas.clemson.edu/%7Eahoover/stare/
CHASE_DB1	Retinal fundus imaging for blood vessel segmentation	https://blogs.kingston.ac.uk/retinal/chasedb1/
LiTS	Liver CT scans for tumor segmentation	https://competitions.codalab.org/competitions/17094
LIDC-IDRI	Lung CT scans for cancer segmentation	https://wiki.cancerimagingarchive.net/display/Public/LIDC-IDRI
LUNA 2016	CT scans for lung nodule segmentation	https://luna16.grand-challenge.org/
xVertSeg	CT spine images for vertebra segmentation	http://lit.fe.uni-lj.si/xVertSeg/
SIIM-ACR	Chest X-rays for pneumothorax segmentation	https://www.kaggle.com/c/siim-acr-pneumothorax-segmentation/data
ISIC	Dermoscopy images for skin lesion segmentation	https://www.isic-archive.com/
BraTS 2012 - 2020	MRI modalities (T1, T2, FLAIR) for brain tumor segmentation.	http://braintumorsegmentation.org/
ISLES	MRI scans for stroke lesion segmentation	http://www.isles-challenge.org/
ICCVB	Prostate MRI and retinal fundus imaging	http://i2cvb.github.io/
IBSR	Repository of MRI imaging	https://www.nitrc.org/projects/ibsr
ACDC 2017	MRI imaging for cardiac diagnosis and segmentation	https://www.creatis.insa-lyon.fr/Challenge/acdc/index.html
PROMIS 2012	Prostate MRI image segmentation	https://promise12.grand-challenge.org/
Med. Seg. Decathlon	MRI and CT modalities for tumor segmentation in various organs like liver, brain, lung, etc.	http://medicaldecathlon.com/
OASIS	MRI and PET images for aging analysis and segmentation	https://www.oasis-brains.org/
Head-Neck-PET-CT	PET and CT imaging for tumor segmentation	https://wiki.cancerimagingarchive.net/display/Public/Head-Neck-PET-CT
BUSIS	Ultrasound imaging for breast tumor segmentation	http://cvprip.cs.usu.edu/busbench/
BUSI	Breast ultrasound scans for tumor segmentation	https://scholar.cu.edu.eg/?q=afahmy/pages/dataset

7.1 General challenges

One major challenge is concerned with the computational power requirement which tends to limit the feasibility of the approach. Following this many cloud based high performance computing environments are developed for mobile, efficient and faster computations. Although progress is also made towards the model compression and acceleration techniques (Cheng et al. 2017) with great achievements; however, it is still required to establish the concrete benchmark results for real-time applications. Recently, Tan and Le (2019) proposed an EfficientNet framework that uses compound coefficients for uniform scaling in all dimensions. This could make the U-Net design streamlined for complex segmentation tasks with minimal change in the parameters. Besides several attempts are also made towards automation of model architecture design (Ren et al. 2021) to develop optimal model for different applications; however, there is still long way to go.

Furthermore, these powerful deep learning approaches are data-hungry i.e. the amount of data available directly affects the model performance towards achieving robust results. However, the expense of data acquisition and delineation, and data security, results in the limited availability of the data which bottlenecks the development of real-world systems. In this context, various data augmentation strategies (Shorten and Khoshgoftaar 2019) are proposed that tend to alleviate the performance of the model while drawing the advantages of big data. Generally, the image augmentation strategies involve geometric transformations, color space augmentations, kernel filters, mixing images, random erasing, feature space augmentation, adversarial training, generative adversarial networks, neural style transfer, and meta-learning. However, the diversity of augmented data is limited by the available data which could result in overfitting. In another approach, U-Net models utilize transfer learning approaches (Byra et al. 2020b) to optimize the pre-trained model to adapt to the targeted task while having insufficient training data. These deep transfer learning techniques are categories under four broad areas: instances based, mapping based, network based and adversarial based (Tan et al. 2018). The self-supervised learning (SSL) (Jing and Tian 2020) is an emerging technology that also addresses this challenge. In SSL strategies initially, pre-training is performed with un-annotated samples for some pretext task to learn feature representations such as predicting rotations, identifying the image patch, solving jigsaw puzzles, etc. and later model is fine-tuned to perform actual segmentation. The potential of this approach attracts many researchers to advance the U-Net based BIS approaches. Furthermore, with the fusion of different modalities, rich information can be extracted concerning desired features for training the model. However, developing an appropriate fusion approach representing vivid modalities is still a challenging task.

The performance of models are also affected by the low imaging quality caused by the noise and artefacts, where noise may obscure features of an image, while artefact adds irrelevant features following some pattern. For instance in CT imaging, noise can make images grainy with small variations in contrast, whereas a streak artefact appears to make the region of low density. There are several pre-processing strategies proposed to remove or minimize the presence of noise and artefacts from data. For denoising the most common approaches that are followed are wavelets thresholding, partial differential equations (PDEs) (minimization problem of total variation method), NL-means and fast NL-means algorithms, anisotropic diffusion, etc. (Oulhaj et al. 2012; Ravishankar et al. 2017). The artefacts can be reduced by using newer reconstruction or metal artefact reduction (MAR) techniques (Chen et al. 2019b; Triche et al. 2019).

In general, the decision made in the rule-based applications can be traced back to its origin; however, deep CNN models lack transparency in the decision making process, where the input and output are well-presented but the processing in the hidden layers is difficult to interpret and understand, and hence these are also termed as black-box models. To better interpret these models various visualization based approaches are proposed such as local interpretable model-agnostic explanations (LIME) (Mishra et al. 2017), shapley additive explanation (SHAP) (Lundberg and Lee 2017), partial dependence plots (PDP) (Friedman 2001), anchor (Ribeiro et al. 2018), etc. Currently, these approaches are applied to explain and interpret the obtained results from deep learning models, but still, a concrete benchmark scheme is required to be established.

7.2 Specific challenges

In addition to the general challenges each modality also exhibit unique challenges. Some of these major challenges that are profound in X-ray, CT, MRI, PET and US imaging are shown in Fig. 8. In X-ray imaging because of 2D projection of the 3D human body, features representing physiological structures overlap each other which may result in variation in the anatomy representation. For instance, in chest X-rays, due to the presence of scarring the lung contours are substantially blurred and hence segmentation models must learn the global concept for resolving the ambiguities and producing the correct mask. This challenge is mostly addressed by using U-Net based models that use long skip connection variants to transfer the knowledge from extraction to reconstruction phase (Rashid et al. 2018; Frid-Adar et al. 2018a), and adversarial learning strategies (Gaál et al. 2020). However, such model requires a large amount of pixel-level annotated training data which can be addressed by using data augmentation, self-supervised learning (Punn and Agarwal 2021a)

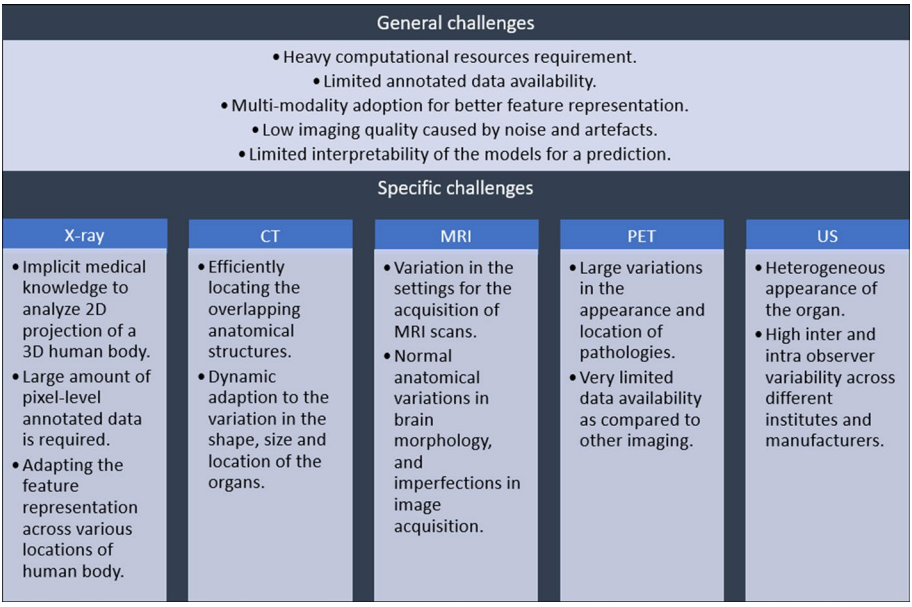


Fig. 8 Challenges involved in biomedical image segmentation

or semi-supervised learning (Yu et al. 2018) strategies. In another challenge, a model needs to be designed with dynamic feature adaption to generate segmentation masks at different locations in the human body. It can be considered as one of the important aspects for bone segmentation (Wang et al. 2020a) while using X-ray imaging.

With CT imaging most of the challenges arise due to overlapping anatomical structures and large variations in the shape, size and location of the organs from person to person. For example, in the case of an abnormal lung CT segmentation, lung parenchyma (Skourt et al. 2018) needs to be segregated from the bronchus regions that represent similar features as lung tissue, along with the segmentation of nodules and blood vessels. Moreover, pulmonary inflation with an elastic chest wall can result in large variation in volumes and margins (Mansoor et al. 2015). To address these challenges mostly attention based U-Net models are employed in CT image segmentation (Fan et al. 2020b; Seo et al. 2019; Song et al. 2019), while the performance could further be improved by incorporating other networks or operational designs such as object dependent filters, residual blocks, etc.

Unlike other modalities, MRI is most widely used for segmentation (brain tissue, tumor, skull, prostate, etc.) due to the ample availability of datasets. However, automated analysis using MRI is challenging due to intensity inhomogeneity, changes in settings for the acquisition of MRI scans, fluctuations in the appearance of pathology, anatomical variations in brain morphology, and imperfections in image acquisition. For instance, the performance of brain tumor segmentation models is affected by large variations in brain tumors location, size, shape and heterogeneity (image uniformity, contrast uptake and texture) (Akkus et al. 2017; Wadhwa et al. 2019). To address these challenges multi-modality fusion based approaches are most widely studied to effectively learn the inconsistent tumor features (Zhou et al. 2019b). Depending on the segmentation problem under consideration these challenges can be addressed by integrating several state-of-the-art architectural designs (inception, cascaded, attention, dense, etc.) and operational designs (atrous, spectral, hybrid, etc.) resulting in ample possibilities of approaches (Dong et al. 2020; Punn and Agarwal 2020d; Zhang et al. 2019). Similar to MRI, PET imaging analysis is mostly utilized by oncologists to diagnose and analyse severe problems such as gliomas, perfusion evaluation, cerebrovascular accidents, parkinson's disease, etc. This brings similar challenges of large variations in the appearance and location of pathologies (Weller et al. 2013) in addition to very limited data availability due to privacy and security concerns which can be addressed by using U-Net based models assisted with data augmentation, self-supervised learning or transfer learning strategies (Lu et al. 2020).

There are various clinical applications of ultrasound imaging including cardiology, breast cancer, prostate, and other diseases (Noble and Boukerroui 2006) which can be assisted with automated segmentation. However, the heterogeneous appearance of the organ due to variations in depth, neighbouring tissues and location is one of the major challenges in ultrasound image segmentation (Zhou et al. 2020). In this regard, most of U-Net based approaches uses inception based attention with dense encoder modules to develop multi-scale feature representation (Yang et al. 2019; Li et al. 2019b; Wang et al. 2018b). Another major challenge is concerned with the high variability in the inter and intra-observer among physicians and sonographers, which depends on the acquisition protocols and observer preference, thereby a larger training dataset is required to alleviate the variation (Liu et al. 2019a).

8 Conclusion

The deep learning approaches especially U-Net has great potential to influence the clinical applications involving automated biomedical imaging segmentation. With U-Net being a breakthrough development, it sets up the foundation for the development of novel architectures concerning the identification and localization of the target regions or sub-regions. Following from this context, in present article, various U-Net variants are explored, covering current advancements and developments in the area of biomedical image segmentation serving different modalities. Each U-Net variant features unique developments over the challenges incurred due to different modalities. With such high utility and potential of the U-Net models, it is believed that U-Net based models would be widely applied to address various challenging problems experienced in the biomedical image segmentation for developing real world computer-aided diagnosis systems.

Acknowledgements We thank our institute, Indian Institute of Information Technology Allahabad (IIITA), India and Big Data Analytics (BDA) lab for allocating the necessary resources to perform this research. We extend our thanks to our colleagues for their valuable guidance and suggestions.

References

Abedalla A, Abdullah M, Al-Ayyoub M, Benkhelifa E (2020) 2st-unet: 2-stage training model using u-net for pneumothorax segmentation in chest x-rays. In: 2020 International Joint Conference on Neural Networks (IJCNN), IEEE, pp 1–6

Agarwal S, Punn NS, Sonbhadra SK, Nagabhushan P, Pandian K, Saxena P (2020) Unleashing the power of disruptive and emerging technologies amid covid 2019: A detailed review. arXiv preprint arXiv:2005.11507

Akkus Z, Galimzianova A, Hoogi A, Rubin DL, Erickson BJ (2017) Deep learning for brain mri segmentation: state of the art and future directions. J Digital Imaging 30(4):449–459

Alexander A, McGill M, Tarasova A, Ferreira C, Zurkiya D (2019) Scanning the future of medical imaging. J Am College Radiol 16(4):501–507

Almajalid R, Shan J, Du Y, Zhang M (2018) Development of a deep-learning-based method for breast ultrasound image segmentation. In: 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), IEEE, pp 1103–1108

Alom MZ, Hasan M, Yakopcic C, Taha TM, Asari VK (2018) Recurrent residual convolutional neural network based on u-net (r2u-net) for medical image segmentation. arXiv preprint arXiv:1802.06955

Ayachi R, Afif M, Said Y, Atri M (2018) Strided convolution instead of max pooling for memory efficiency of convolutional neural networks. International conference on the Sciences of Electronics. Springer, Technologies of Information and Telecommunications, pp 234–243

Azad R, Asadi-Aghbolaghi M, Fathy M, Escalera S (2019) Bi-directional convlstm u-net with densely connected convolutions. In: Proceedings of the IEEE International Conference on Computer Vision Workshops, pp 0–0

Badrinarayanan V, Kendall A, Cipolla R (2017) Segnet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE transactions on pattern analysis and machine intelligence 39(12):2481–2495

Baumgartner CF, Tezcan KC, Chaitanya K, Hötker AM, Muehlematter UJ, Schawkat K, Becker AS, Donati O, Konukoglu E (2019) Phiseg: Capturing uncertainty in medical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, pp 119–127

Bercovich E, Javitt MC (2018) Medical imaging: from roentgen to the digital revolution, and beyond. Rambam Maimonides medical journal 9(4)

Bhattacharyya S (2011) A brief survey of color image preprocessing and segmentation techniques. J Pattern Recogn Res 1(1):120–129

Blanc-Durand P, Van Der Gucht A, Schaefer N, Itti E, Prior JO (2018) Automatic lesion detection and segmentation of 18f-fet pet in gliomas: a full 3d u-net convolutional neural network study. PLoS One 13(4):e0195798

Boykov Y, Funka-Lea G (2006) Graph cuts and efficient nd image segmentation. Int J Computer Vision 70(2):109–131
Byra M, Jarosik P, Szubert A, Galperin M, Ojeda-Fournier H, Olson L, O'Boyle M, Comstock C, Andre M (2020) Breast mass segmentation in ultrasound with selective kernel u-net convolutional neural network. Biomed Signal Proc Control 61:102027
Byra M, Wu M, Zhang X, Jang H, Ma YJ, Chang EY, Shah S, Du J (2020) Knee menisci segmentation and relaxometry of 3d ultrashort echo time cones mr imaging using attention u-net with transfer learning. Mag Res Med 83(3):1109–1122
Cao H, Wang Y, Chen J, Jiang D, Zhang X, Tian Q, Wang M (2021) Swin-unet: Unet-like pure transformer for medical image segmentation. arXiv preprint arXiv:2105.05537
Chen C, Qin C, Qiu H, Tarroni G, Duan J, Bai W, Rueckert D (2020) Deep learning for cardiac image segmentation: a review. Fronti Cardiovas Med 7:25
Chen L, Strauch M, Merhof D (2019a) Instance segmentation of biomedical images with an object-aware embedding learned with local constraints. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, pp 451–459
Chen LC, Papandreou G, Kokkinos I, Murphy K, Yuille AL (2014) Semantic image segmentation with deep convolutional nets and fully connected crfs. arXiv preprint arXiv:1412.7062
Chen LC, Papandreou G, Kokkinos I, Murphy K, Yuille AL (2017) Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE transactions on pattern analysis and machine intelligence 40(4):834–848
Chen LC, Papandreou G, Schroff F, Adam H (2017b) Rethinking atrous convolution for semantic image segmentation. arXiv preprint arXiv:1706.05587
Chen LC, Zhu Y, Papandreou G, Schroff F, Adam H (2018a) Encoder-decoder with atrous separable convolution for semantic image segmentation. In: Proceedings of the European conference on computer vision (ECCV), pp 801–818
Chen M, Xia D, Wang D, Han J, Liu Z (2019b) An analytical method for reducing metal artifacts in x-ray ct images. Mathematical Problems in Engineering 2019
Chen W, Liu B, Peng S, Sun J, Qiao X (2018b) S3d-unet: separable 3d u-net for brain tumor segmentation. In: International MICCAI Brainlesion Workshop, Springer, pp 358–368
Cheng Y, Wang D, Zhou P, Zhang T (2017) A survey of model compression and acceleration for deep neural networks. arXiv preprint arXiv:1710.09282
Chollet F (2017) Xception: Deep learning with depthwise separable convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 1251–1258
Chowdary GJ, Punn NS, Sonbhadra SK, Agarwal S (2020) Face mask detection using transfer learning of inceptionv3. arXiv preprint arXiv:2009.08369
Christ PF, Elshaer MEA, Ettlinger F, Tatavarty S, Bickel M, Bilic P, Rempfler M, Armbruster M, Hofmann F, D'Anastasi M, et al. (2016) Automatic liver and lesion segmentation in ct using cascaded fully convolutional neural networks and 3d conditional random fields. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, pp 415–423
Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O (2016) 3d u-net: learning dense volumetric segmentation from sparse annotation. In: International conference on medical image computing and computer-assisted intervention, Springer, pp 424–432
Ciresan D, Giusti A, Gambardella LM, Schmidhuber J (2012) Deep neural networks segment neuronal membranes in electron microscopy images. In: Advances in neural information processing systems, pp 2843–2851
CORE (2020) Computing research and education association of australasia. https://www.core.edu.au/, [Online; accessed December 06, 2020]
Dai J, Qi H, Xiong Y, Li Y, Zhang G, Hu H, Wei Y (2017) Deformable convolutional networks. In: Proceedings of the IEEE international conference on computer vision, pp 764–773
Deepa S, Devi BA et al (2011) A survey on artificial intelligence approaches for medical image classification. Indian J Sci Technol 4(11):1583–1595
Dong H, Yang G, Liu F, Mo Y, Guo Y (2017) Automatic brain tumor detection and segmentation using u-net based fully convolutional networks. In: annual conference on medical image understanding and analysis, Springer, pp 506–517
Dong S, Zhao J, Zhang M, Shi Z, Deng J, Shi Y, Tian M, Zhuo C (2020) Deu-net: Deformable u-net for 3d cardiac mri video segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, pp 98–107
Dong X, Lei Y, Tian S, Wang T, Patel P, Curran WJ, Jani AB, Liu T, Yang X (2019) Synthetic mri-aided multi-organ segmentation on male pelvic ct using cycle consistent deep attention network. Radio Oncol 141:192–199

Dong X, Lei Y, Wang T, Thomas M, Tang L, Curran WJ, Liu T, Yang X (2019) Automatic multiorgan segmentation in thorax ct images using u-net-gan. Med Phys 46(5):2157–2168
Dunnhofer M, Antico M, Sasazawa F, Takeda Y, Camps S, Martinel N, Micheloni C, Carneiro G, Fontanarosa D (2020) Siam-u-net: encoder-decoder siamese network for knee cartilage tracking in ultrasound images. Med Image Analy 60:101631
Elnakib A, Gimel'farb G, Suri JS, El-Baz A (2011) Medical image segmentation: a brief survey. In: Multi Modality State-of-the-Art Medical Image Segmentation and Registration Methodologies, Springer, pp 1–39
Fan DP, Zhou T, Ji GP, Zhou Y, Chen G, Fu H, Shen J, Shao L (2020a) Inf-net: Automatic covid-19 lung infection segmentation from ct images. IEEE Transactions on Medical Imaging
Fan T, Wang G, Li Y, Wang H (2020) Ma-net: A multi-scale attention network for liver and tumor segmentation. IEEE Access 8:179656–179665
Fenster A, Chiu B (2006) Evaluation of segmentation algorithms for medical imaging. In: 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference, IEEE, pp 7186–7189
Frid-Adar M, Ben-Cohen A, Amer R, Greenspan H (2018a) Improving the segmentation of anatomical structures in chest radiographs using u-net with an imagenet pre-trained encoder. In: Image Analysis for Moving Organ, Breast, and Thoracic Images, Springer, pp 159–168
Frid-Adar M, Diamant I, Klang E, Amitai M, Goldberger J, Greenspan H (2018) Gan-based synthetic medical image augmentation for increased cnn performance in liver lesion classification. Neurocomputing 321:321–331
Friedman JH (2001) Greedy function approximation: a gradient boosting machine. Annals of statistics pp 1189–1232
Fu X, Bi L, Kumar A, Fulham M, Kim J (2021) Multimodal spatial attention module for targeting multimodal pet-ct lung tumor segmentation. IEEE Journal of Biomedical and Health Informatics
Gaál G, Maga B, Lukács A (2020) Attention u-net based adversarial architectures for chest x-ray lung segmentation. arXiv preprint arXiv:2003.10304
Garcia-Garcia A, Orts-Escolano S, Oprea S, Villena-Martinez V, Martinez-Gonzalez P, Garcia-Rodriguez J (2018) A survey on deep learning techniques for image and video semantic segmentation. Appl Soft Comput 70:41–65
Glorot X, Bengio Y (2010) Understanding the difficulty of training deep feedforward neural networks. In: Proceedings of the thirteenth international conference on artificial intelligence and statistics, pp 249–256
Göçeri E (2013) A comparative evaluation for liver segmentation from spir images and a novel level set method using signed pressure force function. PhD thesis, İzmir Institute of Technology, İzmir
Goceri E (2016) Automatic labeling of portal and hepatic veins from mr images prior to liver transplantation. Int J Comput Ass Radiol Surg 11(12):2153–2161
Göçeri E (2020) Impact of deep learning and smartphone technologies in dermatology: Automated diagnosis. 2020 Tenth International Conference on Image Processing Theory. Tools and Applications (IPTA), IEEE, pp 1–6
Goceri E (2021) Diagnosis of skin diseases in the era of deep learning and mobile technology. Comput Biol Med 134:104458
Goceri E, Songul C (2018) Biomedical information technology: image based computer aided diagnosis systems. In: International Conference on Advanced Technologies, Antalaya, Turkey
Göçeri E, Ünlü MZ, Dicle O (2015) A comparative performance evaluation of various approaches for liver segmentation from spir images. Turkish Journal of Electrical Engineering & Computer Sciences 23(3):741–768
Gomariz A, Li W, Ozkan E, Tanner C, Goksel O (2019) Siamese networks with location prior for landmark tracking in liver ultrasound sequences. In: 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), IEEE, pp 1757–1760
Gu Z, Cheng J, Fu H, Zhou K, Hao H, Zhao Y, Zhang T, Gao S, Liu J (2019) Ce-net: Context encoder network for 2d medical image segmentation. IEEE Trans Med Imaging 38(10):2281–2292
Guo Z, Guo N, Gong K, Li Q et al (2019) Gross tumor volume segmentation for head and neck cancer radiotherapy using deep dense multi-modality network. Phys Med Biol 64(20):205015
Han Y, Ye JC (2018) Framing u-net via deep convolutional framelets: Application to sparse-view ct. IEEE Trans Med Imaging 37(6):1418–1429
Haque IRI, Neubert J (2020) Deep learning approaches to biomedical image segmentation. Informat Med Unlocked 18:100297
Havaei M, Guizard N, Larochelle H, Jodoin PM (2016) Deep learning trends for focal brain pathology segmentation in mri. In: Machine learning for health informatics, Springer, pp 125–148

He K, Zhang X, Ren S, Sun J (2015a) Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In: Proceedings of the IEEE international conference on computer vision, pp 1026–1034
He K, Zhang X, Ren S, Sun J (2015) Spatial pyramid pooling in deep convolutional networks for visual recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence 37(9):1904–1916
He K, Zhang X, Ren S, Sun J (2016a) Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 770–778
He K, Zhang X, Ren S, Sun J (2016b) Identity mappings in deep residual networks. In: European conference on computer vision, Springer, pp 630–645
He K, Gkioxari G, Dollár P, Girshick R (2017) Mask r-cnn. In: Proceedings of the IEEE international conference on computer vision, pp 2961–2969
Hesamian MH, Jia W, He X, Kennedy P (2019) Deep learning techniques for medical image segmentation: achievements and challenges. J Digital Imaging 32(4):582–596
Hiasa Y, Otake Y, Takao M, Ogawa T, Sugano N, Sato Y (2019) Automated muscle segmentation from clinical ct using bayesian u-net for personalized musculoskeletal modeling. IEEE Transact Med Imaging 39(4):1030–1040
Hopkins J (2020) 2019 novel coronavirus covid-19 (2019-ncov) data repository by johns hopkins csse. https://github.com/CSSEGISandData/COVID-19, [Online; accessed November 17, 2021]
Howard AG, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, Andreetto M, Adam H (2017) Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861
Hu J, Shen L, Sun G (2018) Squeeze-and-excitation networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 7132–7141
Huang G, Liu Z, Van Der Maaten L, Weinberger KQ (2017) Densely connected convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 4700–4708
Huazhu F, Deng-Ping F, Geng C, Tao Z (2020) Covid-19 imaging-based ai research collection. https://git.io/JYAtL, [Online; accessed January 11, 2021]
Hughes Z (2019) Medical imaging types and modalities. https://www.ausmed.com/cpd/articles/medical-imaging-types-and-modalities, [Online; accessed November 25, 2020]
Hwang H, Rehman HZU, Lee S (2019) 3d u-net for skull stripping in brain mri. Appl Sci 9(3):569
Ibtehaz N, Rahman MS (2020) Multiresunet: Rethinking the u-net architecture for multimodal biomedical image segmentation. Neural Networks 121:74–87
Isensee F, Jaeger PF, Kohl SA, Petersen J, Maier-Hein KH (2021) nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. Nature methods 18(2):203–211
James AP, Dasarathy BV (2014) Medical image fusion: A survey of the state of the art. Information Fusion 19:4–19
Janssens R, Zeng G, Zheng G (2018) Fully automatic segmentation of lumbar vertebrae from ct images using cascaded 3d fully convolutional networks. In: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), IEEE, pp 893–897
Jing L, Tian Y (2020) Self-supervised visual feature learning with deep neural networks: A survey. IEEE transactions on pattern analysis and machine intelligence
Kaya B, Goceri E, Becker A, Elder B, Puduvalli V, Winter J, Gurcan M, Otero JJ (2017) Automated fluorescent miscroscopic image analysis of ptbp1 expression in glioma. Plos One 12(3):e0170991
Kendall A, Badrinarayanan V, Cipolla R (2015) Bayesian segnet: Model uncertainty in deep convolutional encoder-decoder architectures for scene understanding. arXiv preprint arXiv:1511.02680
Kermi A, Mahmoudi I, Khadir MT (2018) Deep convolutional neural networks using u-net for automatic brain tumor segmentation in multimodal mri volumes. In: International MICCAI Brainlesion Workshop, Springer, pp 37–48
Kohl S, Romera-Paredes B, Meyer C, De Fauw J, Ledsam JR, Maier-Hein K, Eslami SA, Rezende DJ, Ronneberger O (2018) A probabilistic u-net for segmentation of ambiguous images. In: Advances in Neural Information Processing Systems, pp 6965–6975
Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems, pp 1097–1105
Kumar P, Nagar P, Arora C, Gupta A (2018) U-segnet: fully convolutional neural network based automated brain tissue segmentation tool. In: 2018 25th IEEE International Conference on Image Processing (ICIP), IEEE, pp 3503–3507
Leader JK, Zheng B, Rogers RM, Sciurba FC, Perez A, Chapman BE, Patel S, Fuhrman CR, Gur D (2003) Automated lung segmentation in x-ray computed tomography: development and evaluation of a heuristic threshold-based scheme. Academic Radiol 10(11):1224–1236

Lei T, Wang R, Wan Y, Zhang B, Meng H, Nandi AK (2020) Medical image segmentation using deep learning: a survey. arXiv preprint arXiv:2009.13120
Leung KH, Marashdeh W, Wray R, Ashrafinia S, Pomper MG, Rahmim A, Jha AK (2020) A physics-guided modular deep-learning based automated framework for tumor segmentation in pet. Physics in Medicine & Biology
Li B, Kang G, Cheng K, Zhang N (2019a) Attention-guided convolutional neural network for detecting pneumonia on chest x-rays. In: 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE, pp 4851–4854
Li H, Luo H, Huan W, Shi Z, Yan C, Wang L, Mu Y, Liu Y (2021) Automatic lumbar spinal mri image segmentation with a multi-scale attention network. Neural Computing and Applications pp 1–14
Li X, Li C, Fedorov A, Kapur T, Yang X (2016) Segmentation of prostate from ultrasound images using level sets on active band and intensity variation across edges. Med Phys 43(6):3090–3103
Li X, Hong Y, Kong D, Zhang X (2019) Automatic segmentation of levator hiatus from ultrasound images using u-net with dense connections. Phys Med Biol 64(7):075015
Li X, Wang W, Hu X, Yang J (2019c) Selective kernel networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)
Lin F, Liu C, Xie H, Zha ZJ, Zhang Y (2019) Semantic-embedding and shape-aware u-net for ultrasound eyeball segmentation. In: 2019 IEEE International Conference on Multimedia and Expo (ICME), IEEE, pp 892–897
Lin TY, Goyal P, Girshick R, He K, Dollár P (2017) Focal loss for dense object detection. In: Proceedings of the IEEE international conference on computer vision, pp 2980–2988
Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, Van Der Laak JA, Van Ginneken B, Sánchez CI (2017) A survey on deep learning in medical image analysis. Med Image Analys 42:60–88
Liu S, Wang Y, Yang X, Lei B, Liu L, Li SX, Ni D, Wang T (2019) Deep learning in medical ultrasound analysis: a review. Engineering 5(2):261–275
Liu X, Deng Z, Yang Y (2019) Recent progress in semantic image segmentation. Artificial Intelligence Review 52(2):1089–1106
Liu Z, Song YQ, Sheng VS, Wang L, Jiang R, Zhang X, Yuan D (2019) Liver ct sequence segmentation based with improved u-net and graph cut. Expert Systems with Applications 126:54–63
Liu Z, Lin Y, Cao Y, Hu H, Wei Y, Zhang Z, Lin S, Guo B (2021) Swin transformer: Hierarchical vision transformer using shifted windows. arXiv preprint arXiv:2103.14030
Long J, Shelhamer E, Darrell T (2015) Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 3431–3440
Lu Y, Lin J, Chen S, He H, Cai Y (2020) Automatic tumor segmentation by means of deep convolutional u-net with pre-trained encoder in pet images. IEEE Access 8:113636–113648
Lundberg SM, Lee SI (2017) A unified approach to interpreting model predictions. In: Advances in neural information processing systems, pp 4765–4774
Ma J (2020) Segmentation loss odyssey. arXiv preprint arXiv:2005.13449
Maintz JA, Viergever MA (1998) A survey of medical image registration. Med Image Analy 2(1):1–36
Man Y, Huang Y, Feng J, Li X, Wu F (2019) Deep q learning driven ct pancreas segmentation with geometry-aware u-net. IEEE Trans Med Imaging 38(8):1971–1980
Mansoor A, Bagci U, Foster B, Xu Z, Papadakis GZ, Folio LR, Udupa JK, Mollura DJ (2015) Segmentation and image analysis of abnormal lungs at ct: current approaches, challenges, and future trends. Radiographics 35(4):1056–1076
Masood S, Sharif M, Masood A, Yasmin M, Raza M (2015) A survey on medical image segmentation. Curr Med Imaging 11(1):3–14
Mikolov T, Kombrink S, Burget L, Černockỳ J, Khudanpur S (2011) Extensions of recurrent neural network language model. In: 2011 IEEE international conference on acoustics, speech and signal processing (ICASSP), IEEE, pp 5528–5531
Milletari F, Navab N, Ahmadi SA (2016) V-net: Fully convolutional neural networks for volumetric medical image segmentation. In: 2016 fourth international conference on 3D vision (3DV), IEEE, pp 565–571
Minaee S, Boykov Y, Porikli F, Plaza A, Kehtarnavaz N, Terzopoulos D (2020) Image segmentation using deep learning: A survey. arXiv preprint arXiv:2001.05566
Mishra S, Sturm BL, Dixon S (2017) Local interpretable model-agnostic explanations for music content analysis. In: ISMIR, pp 537–543
Mnih V, Kavukcuoglu K, Silver D, Rusu AA, Veness J, Bellemare MG, Graves A, Riedmiller M, Fidjeland AK, Ostrovski G et al (2015) Human-level control through deep reinforcement learning. Nature 518(7540):529–533

Moore CL, Copel JA (2011) Point-of-care ultrasonography. New England J Med 364(8):749–757
Morris SA, Slesnick TC (2018) Magnetic resonance imaging. Visual Guide to Neonatal Cardiology pp 104–108
Nasalwai N, Punn NS, Sonbhadra SK, Agarwal S (2021) Addressing the class imbalance problem in medical image segmentation via accelerated tversky loss function. In: Pacific-Asia Conference on Knowledge Discovery and Data Mining, Springer, pp 390–402
Noble JA, Boukerroui D (2006) Ultrasound image segmentation: a survey. IEEE Transactions on Medical Imaging 25(8):987–1010
Oktay O, Schlemper J, Folgoc LL, Lee M, Heinrich M, Misawa K, Mori K, McDonagh S, Hammerla NY, Kainz B, Glocker B, Rueckert D (2018) Attention u-net: Learning where to look for the pancreas. arXiv: 1804.03999
Ollinger JM, Fessler JA (1997) Positron-emission tomography. IEEE Signal Processing Magazine 14(1):43–55
Oulhaj H, Amine A, Rziza M, Aboutajdine D (2012) Noise reduction in medical images - comparison of noise removal algorithms -. 2012 International Conference on Multimedia Computing and Systems pp 344–349
Park J, Yun J, Kim N, Park B, Cho Y, Park HJ, Song M, Lee M, Seo JB (2020) Fully automated lung lobe segmentation in volumetric chest ct with 3d u-net: validation with intra-and extra-datasets. J Digital Imaging 33(1):221–230
Punn N, Agarwal S (2020a) Automated diagnosis of covid-19 with limited posteroanterior chest x-ray images using fine-tuned deep neural networks. Applied Intelligence
Punn NS, Agarwal S (2020b) Chs-net: A deep learning approach for hierarchical segmentation of covid-19 infected ct images. arXiv preprint arXiv:2012.07079
Punn NS, Agarwal S (2020) Inception u-net architecture for semantic segmentation to identify nuclei in microscopy cell images. ACM Transactions on Multimedia Computing, Communications, and Applications TOMM 16(1):1–15
Punn NS, Agarwal S (2020d) Multi-modality encoded fusion with 3d inception u-net and decoder model for brain tumor segmentation. Multimedia Tools and Applications pp 1–16
Punn NS, Agarwal S (2021a) Bt-unet: A self-supervised learning framework for biomedical image segmentation using barlow twins with u-net models. arXiv preprint arXiv:2112.03916
Punn NS, Agarwal S (2021b) Rca-iunet: A residual cross-spatial attention guided inception u-net model for tumor segmentation in breast ultrasound imaging. arXiv preprint arXiv:2108.02508
Punn NS, Sonbhadra SK, Agarwal S (2020a) Covid-19 epidemic analysis using machine learning and deep learning algorithms. medRxiv
Punn NS, Sonbhadra SK, Agarwal S (2020b) Monitoring covid-19 social distancing with person detection and tracking via fine-tuned yolo v3 and deepsort techniques. arXiv preprint arXiv:2005.01385
Que Q, Tang Z, Wang R, Zeng Z, Wang J, Chua M, Gee TS, Yang X, Veeravalli B (2018) Cardioxnet: Automated detection for cardiomegaly based on deep learning. In: 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE, pp 612–615
Raghu M, Blumer K, Sayres R, Obermeyer Z, Kleinberg B, Mullainathan S, Kleinberg J (2019) Direct uncertainty prediction for medical second opinions. In: International Conference on Machine Learning, pp 5281–5290
Rashid R, Akram MU, Hassan T (2018) Fully convolutional neural network for lungs segmentation from chest x-rays. In: International Conference Image Analysis and Recognition, Springer, pp 71–80
Ravishankar A, Anusha S, Akshatha H, Raj A, Jahnavi S, Madhura J (2017) A survey on noise reduction techniques in medical images. In: 2017 International conference of Electronics, Communication and Aerospace Technology (ICECA), IEEE, vol 1, pp 385–389
Razzak MI, Naz S, Zaib A (2018) Deep learning for medical image processing: Overview, challenges and the future. Classification in BioApps pp 323–350
Ren P, Xiao Y, Chang X, Huang PY, Li Z, Chen X, Wang X (2021) A comprehensive survey of neural architecture search: challenges and solutions. ACM Computing Surveys (CSUR) 54(4):1–34
Ren S, He K, Girshick R, Sun J (2015) Faster r-cnn: Towards real-time object detection with region proposal networks. Adv Neural Informat Proces Syst 28:91–99
Renard F, Guedria S, De Palma N, Vuillerme N (2020) Variability and reproducibility in deep learning for medical image segmentation. Sci Rep 10(1):1–16
Ribeiro MT, Singh S, Guestrin C (2018) Anchors: High-precision model-agnostic explanations. AAAI 18:1527–1535

Ronneberger O, Fischer P, Brox T (2015) U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention, Springer, pp 234–241

Rundo L, Han C, Nagano Y, Zhang J, Hataya R, Militello C, Tangherloni A, Nobile MS, Ferretti C, Besozzi D et al (2019) Use-net: Incorporating squeeze-and-excitation blocks into u-net for prostate zonal segmentation of multi-institutional mri datasets. Neurocomputing 365:31–43

SearchEngines (2020) The top list of academic search engines. https://paperpile.com/g/academic-search-engines/, [Online; accessed December 06, 2020]

Seo H, Huang C, Bassenne M, Xiao R, Xing L (2019) Modified u-net (mu-net) with incorporation of object-dependent high level features for improved liver and liver-tumor segmentation in ct images. IEEE Trans Med Imaging 39(5):1316–1325

Shen T, Zhou T, Long G, Jiang J, Pan S, Zhang C (2017) Disan: Directional self-attention network for rnn/cnn-free language understanding. arXiv preprint arXiv:1709.04696

Shi F, Wang J, Shi J, Wu Z, Wang Q, Tang Z, He K, Shi Y, Shen D (2020) Review of artificial intelligence techniques in imaging data acquisition, segmentation and diagnosis for covid-19. IEEE reviews in biomedical engineering

Shorten C, Khoshgoftaar TM (2019) A survey on image data augmentation for deep learning. J Big Data 6(1):60

Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556

Skourt BA, El Hassani A, Majda A (2018) Lung ct image segmentation using deep neural networks. Procedia Comput Sci 127:109–113

Song H, Wang W, Zhao S, Shen J, Lam KM (2018) Pyramid dilated deeper convlstm for video salient object detection. In: Proceedings of the European conference on computer vision (ECCV), pp 715–731

Song T, Meng F, Rodriguez-Paton A, Li P, Zheng P, Wang X (2019) U-next: A novel convolution neural network with an aggregation u-net architecture for gallstone segmentation in ct images. IEEE Access 7:166823–166832

Subramanian V, Wang H, Wu JT, Wong KC, Sharma A, Syeda-Mahmood T (2019) Automated detection and type classification of central venous catheters in chest x-rays. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, pp 522–530

Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A (2015) Going deeper with convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 1–9

Szegedy C, Ioffe S, Vanhoucke V, Alemi AA (2017) Inception-v4, inception-resnet and the impact of residual connections on learning. In: Thirty-first AAAI conference on artificial intelligence

Taghanaki SA, Abhishek K, Cohen JP, Cohen-Adad J, Hamarneh G (2021) Deep semantic segmentation of natural and medical images: a review. Artificial Intelligence Review 54(1):137–178

Tan C, Sun F, Kong T, Zhang W, Yang C, Liu C (2018) A survey on deep transfer learning. In: International conference on artificial neural networks, Springer, pp 270–279

Tan M, Le QV (2019) Efficientnet: Rethinking model scaling for convolutional neural networks. arXiv preprint arXiv:1905.11946

Tanno R, Saeedi A, Sankaranarayanan S, Alexander DC, Silberman N (2019) Learning from noisy labels by regularized estimation of annotator confusion. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp 11244–11253

TMI (2019) Types of medical imaging. https://www.doc.ic.ac.uk/~jce317/types-medical-imaging.html, [Online; accessed November 25, 2020]

Tong G, Li Y, Chen H, Zhang Q, Jiang H (2018) Improved u-net network for pulmonary nodules segmentation. Optik 174:460–469

Triche BL, Nelson JT Jr, McGill NS, Porter KK, Sanyal R, Tessler FN, McConathy JE, Gauntt DM, Yester MV, Singh SP (2019) Recognizing and minimizing artifacts at ct, mri, us, and molecular imaging. RadioGraphics 39(4):1017–1018

Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I (2017) Attention is all you need. In: Advances in neural information processing systems, pp 5998–6008

Vuola AO, Akram SU, Kannala J (2019) Mask-rcnn and u-net ensembled for nuclei segmentation. In: 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), IEEE, pp 208–212

Wadhwa A, Bhardwaj A, Verma VS (2019) A review on brain tumor segmentation of mri images. Magnetic Reson Imaging 61:247–259

Wang B, Lei Y, Tian S, Wang T, Liu Y, Patel P, Jani AB, Mao H, Curran WJ, Liu T et al (2019) Deeply supervised 3d fully convolutional networks with group dilated convolution for automatic mri prostate segmentation. Med Phys 46(4):1707–1718

Wang H, Xie S, Lin L, Iwamoto Y, Han XH, Chen YW, Tong R (2021) Mixed transformer u-net for medical image segmentation. arXiv preprint arXiv:2111.04734
Wang P, Chen P, Yuan Y, Liu D, Huang Z, Hou X, Cottrell G (2018a) Understanding convolution for semantic segmentation. In: 2018 IEEE winter conference on applications of computer vision (WACV), IEEE, pp 1451–1460
Wang P, Patel VM, Hacihaliloglu I (2018b) Simultaneous segmentation and classification of bone surfaces from ultrasound using a multi-feature guided cnn. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, pp 134–142
Wang T, Xiong J, Xu X, Jiang M, Yuan H, Huang M, Zhuang J, Shi Y (2019b) Msu-net: Multiscale statistical u-net for real-time 3d cardiac mri video segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, pp 614–622
Wang T, Xiong J, Xu X, Shi Y (2019) Scnn: A general distribution based statistical convolutional neural network with application to video object detection. Proceedings of the AAAI Conference on Artificial Intelligence 33:5321–5328
Wang W, Feng H, Bu Q, Cui L, Xie Y, Zhang A, Feng J, Zhu Z, Chen Z (2020a) Mdu-net: A convolutional network for clavicle and rib segmentation from a chest radiograph. Journal of Healthcare Engineering 2020
Wang Y, Yu B, Wang L, Zu C, Lalush DS, Lin W, Wu X, Zhou J, Shen D, Zhou L (2018) 3d conditional generative adversarial networks for high-quality pet image estimation at low dose. Neuroimage 174:550–562
Wang Z, Zou N, Shen D, Ji S (2020) Non-local u-nets for biomedical image segmentation. Proceedings of the AAAI Conference on Artificial Intelligence 34:6315–6322
Weller M, Pfister SM, Wick W, Hegi ME, Reifenberger G, Stupp R (2013) Molecular neuro-oncology in clinical practice: a new horizon. Lancet Oncol 14(9):e370–e379
Wu YH, Gao SH, Mei J, Xu J, Fan DP, Zhao CW, Cheng MM (2020) Jcs: An explainable covid-19 diagnosis system by joint classification and segmentation. arXiv preprint arXiv:2004.07054
Xia H, Ma M, Li H, Song S (2021) Mc-net: multi-scale context-attention network for medical ct image segmentation. Applied Intelligence pp 1–12
Xie S, Sun C, Huang J, Tu Z, Murphy K (2018) Rethinking spatiotemporal feature learning: Speed-accuracy trade-offs in video classification. In: Proceedings of the European Conference on Computer Vision (ECCV), pp 305–321
Yan Q, Wang B, Gong D, Luo C, Zhao W, Shen J, Shi Q, Jin S, Zhang L, You Z (2020) Covid-19 chest ct image segmentation–a deep convolutional neural network solution. arXiv preprint arXiv:2004.10987
Yang J, Faraji M, Basu A (2019) Robust segmentation of arterial walls in intravascular ultrasound images using dual path u-net. Ultrasonics 96:24–33
Yu E, Sun J, Li J, Chang X, Han XH, Hauptmann AG (2018) Adaptive semi-supervised feature selection for cross-modal retrieval. IEEE Transactions on Multimedia 21(5):1276–1288
Yu F, Koltun V (2016) Multi-scale context aggregation by dilated convolutions. arXiv: 1511.07122
Yu Y, Acton ST (2002) Speckle reducing anisotropic diffusion. IEEE Transactions on Image Processing 11(11):1260–1270
Zahangir Alom M, Shaifur Rahman M, Shamima Nasrin M, Taha TM, Asari VK (2020) Covid_mtnet: Covid-19 detection with multi-task deep learning approaches. arXiv pp arXiv–2004
Zhang H, Ma J (2018) Hartley spectral pooling for deep learning. arXiv preprint arXiv:1810.04028
Zhang L, Liu A, Xiao J, Taylor P (2020a) Dual encoder fusion u-net (defu-net) for cross-manufacturer chest x-ray segmentation. arXiv: 2009.10608
Zhang Y, Chen JH, Chang KT, Park VY, Kim MJ, Chan S, Chang P, Chow D, Luk A, Kwong T et al (2019) Automatic breast and fibroglandular tissue segmentation in breast mri using deep learning by a fully-convolutional residual neural network u-net. Academic Radiol 26(11):1526–1535
Zhang Y, Lei Y, Qiu RL, Wang T, Wang H, Jani AB, Curran WJ, Patel P, Liu T, Yang X (2020b) Multi-needle localization with attention u-net in us-guided hdr prostate brachytherapy. Medical Physics
Zhao X, Li L, Lu W, Tan S (2018) Tumor co-segmentation in pet/ct using multi-modality fully convolutional neural network. Phys Med Biol 64(1):015011
Zhou B, Yang X, Liu T (2020) Artificial intelligence in quantitative ultrasound imaging: A review. arXiv preprint arXiv:2003.11658
Zhou J, Zhang Q, Zhang B, Chen X (2019) Tonguenet: A precise and fast tongue segmentation system using u-net with a morphological processing layer. Appl Sci 9(15):3128
Zhou T, Ruan S, Canu S (2019) A review: Deep learning for medical image segmentation using multi-modality fusion. Array 3:100004

Zhou Z, Siddiquee MMR, Tajbakhsh N, Liang J (2018a) Unet++: A nested u-net architecture for medical image segmentation. In: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, Springer, pp 3–11

Zhou Z, Siddiquee MMR, Tajbakhsh N, Liang J (2018b) Unet++: A nested u-net architecture for medical image segmentation. In: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, Springer, pp 3–11

Zhu JY, Park T, Isola P, Efros AA (2017) Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE international conference on computer vision, pp 2223–2232

Zotti C, Luo Z, Lalande A, Jodoin PM (2018) Convolutional neural network with shape prior applied to cardiac mri segmentation. IEEE J Biomed Health Informatics 23(3):1119–1128