Artificial Intelligence Review
https://doi.org/10.1007/s10462-020-09854-1

Deep semantic segmentation of natural and medical images: a review

Saeid Asgari Taghanaki[1] · **Kumar Abhishek**[1] · **Joseph Paul Cohen**[2] · **Julien Cohen-Adad**[3] · **Ghassan Hamarneh**[1]

Abstract

The semantic image segmentation task consists of classifying each pixel of an image into an instance, where each instance corresponds to a class. This task is a part of the concept of scene understanding or better explaining the global context of an image. In the medical image analysis domain, image segmentation can be used for image-guided interventions, radiotherapy, or improved radiological diagnostics. In this review, we categorize the leading deep learning-based medical and non-medical image segmentation solutions into six main groups of deep architectural, data synthesis-based, loss function-based, sequenced models, weakly supervised, and multi-task methods and provide a comprehensive review of the contributions in each of these groups. Further, for each group, we analyze each variant of these groups and discuss the limitations of the current approaches and present potential future research directions for semantic image segmentation.

Keywords Semantic image segmentation · Deep learning

Saeid Asgari Taghanaki and Kumar Abhishek Joint first authors.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s10462-020-09854-1) contains supplementary material, which is available to authorized users.

✉ Saeid Asgari Taghanaki
sasgarit@sfu.ca

Kumar Abhishek
kabhishe@sfu.ca

Joseph Paul Cohen
joseph@josephpcohen.com

Julien Cohen-Adad
jcohen@polymtl.ca

Ghassan Hamarneh
hamarneh@sfu.ca

[1] School of Computing Science, Simon Fraser University, Burnaby, Canada

[2] Mila, Université de Montréal, Montreal, Canada

[3] NeuroPoly Lab, Institute of Biomedical Engineering, Polytechnique Montréal, Montreal, Canada

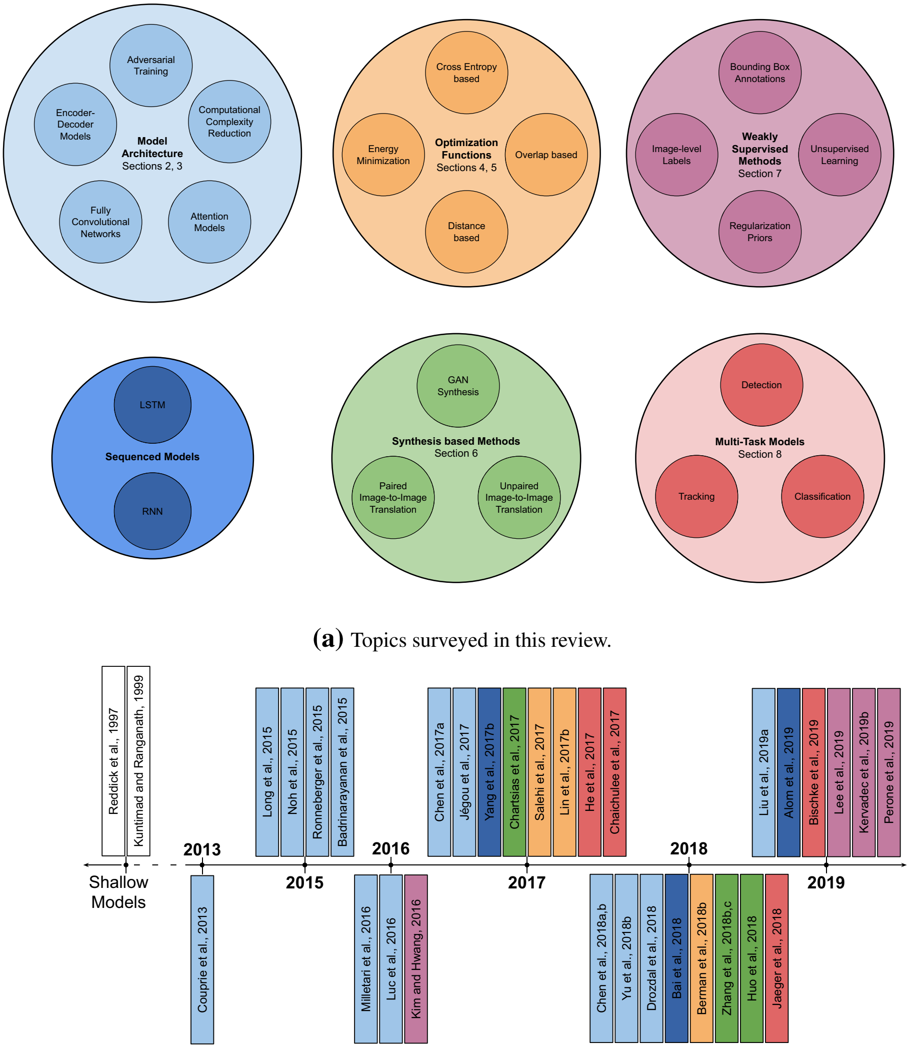
Published online: 13 June 2020

1 Introduction

Deep learning has had a tremendous impact on various fields in science. The focus of the current study is on one of the most critical areas of computer vision: medical image analysis (or medical computer vision), particularly deep learning-based approaches for medical image segmentation. Segmentation is an important processing step in natural images for scene understanding and medical image analysis, for image-guided interventions, radiotherapy, or improved radiological diagnostics, etc. Image segmentation is formally defined as "the partition of an image into a set of nonoverlapping regions whose union is the entire image" (Haralick and Shapiro 1992). A plethora of deep learning approaches for medical image segmentation have been introduced in the literature for different medical imaging modalities, including X-ray, visible-light imaging (e.g. colour dermoscopic images), magnetic resonance imaging (MRI), positron emission tomography (PET), computerized tomography (CT), and ultrasound (e.g. echocardiographic scans). Deep architectural improvement has been a focus of many researchers for different purposes, e.g., tackling gradient vanishing and exploding of deep models, model compression for efficient small yet accurate models, while other works have tried to improve the performance of deep networks by introducing new optimization functions.

Guo et al. (2018) provided a review of deep learning based semantic segmentation of images, and divided the literature into three categories: region-based, fully convolutional network (FCN)-based, and weakly supervised segmentation methods. Hu et al. (2018b) summarized the most commonly used RGB-D datasets for semantic segmentation as well as traditional machine learning based methods and deep learning-based network architectures for RGB-D segmentation. Lateef and Ruichek (2019) presented an extensive survey of deep learning architectures, datasets, and evaluation methods for the semantic segmentation of natural images using deep neural networks. Similarly, for medical imaging, Goceri and Goceri (2017) presented an high-level overview of deep learning-based medical image analysis techniques and application areas. Hesamian et al. (2019) presented an overview of the state-of-the-art methods in medical image segmentation using deep learning by covering the literature related to network structures and model training techniques. Karimi et al. (2019) reviewed the literature on techniques to handle label noise in deep learning based medical image analysis and evaluated existing approaches on three medical imaging datasets for segmentation and classification tasks. Zhou et al. (2019b) presented a review of techniques proposed for fusion of medical images from multiple modalities for medical image segmentation. Goceri (2019a) discussed the fully supervised, weakly supervised and transfer learning techniques for training deep neural networks for segmentation of medical images, and also discussed the existing methods for addressing the problems of lack of data and class imbalance. Zhang et al. (2019) presented a review of the approaches to address the problem of small sample sizes in medical image analysis, and divided the literature into five categories including explanation, weakly supervised, transfer learning, and active learning techniques. Tajbakhsh et al. (2020) presented a review of the literature for addressing the challenges of scarce annotations as well as weak annotations (e.g., noisy annotations, image-level labels, sparse annotations, etc.) in medical image segmentation. Similarly, there are several surveys covering the literature on the task of object detection (Wang et al. 2019c; Zou et al. 2019; Borji et al. 2019; Liu et al. 2019b; Zhao et al. 2019), which can also be used to obtain what can be termed as rough localizations of the object(s) of interest. In contrast to the existing surveys, we make the following contributions in this review:

- We provide comprehensive coverage of research contributions in the field of semantic segmentation of natural and medical images. In terms of medical imaging modalities, we cover the literature pertaining to both 2D (RGB and grayscale) as well as volumetric medical images.
- We group the semantic image segmentation literature into six different categories based on the nature of their contributions: architectural improvements, optimization function based improvements, data synthesis based improvements, weakly supervised models, sequenced models, and multi-task models. Figure 1 indicates the categories we cover



(a) Topics surveyed in this review.

(b) A timeline of the various contributions in deep learning based semantic segmentation of natural and medical images. The contributions are colored according to their topics in (a) above.

Fig. 1 An overview of the deep learning based segmentation methods covered in this review

in this review, along with a timeline of the most influential papers in the respective categories. Moreover, Fig. 2 shows a high-level overview of the deep semantic segmentation pipeline, and where each of the categories mentioned in Fig. 1 belong in the pipeline.

- We study the behaviour of many popular loss functions used to train segmentation models on handling scenarios with varying levels of false positive and negative predictions.
- Followed by the comprehensive review, we recognize and suggest the important research directions for each of the categories.

In the following sections, we discuss deep semantic image segmentation improvements under different categories visualized in Fig. 1. For each category, we first review the improvements on non-medical datasets, and in a subsequent section, we survey the improvements for medical images.

2 Network architectural improvements

This section discusses the advancements in semantic image segmentation using convolutional neural networks (CNNs), which have been applied to interpretation tasks on both natural and medical images (Garcia-Garcia et al. 2018; Litjens et al. 2017). Although artificial neural network-based image segmentation approaches have been explored in the past using shallow networks (Reddick et al. 1997; Kuntimad and Ranganath 1999) as well as works which relied on superpixel segmentation maps to generate pixelwise predictions (Couprie et al. 2013), in this work, we focus on deep neural network based image segmentation models which are end-to-end trainable. The improvements are mostly attributed to exploring new neural architectures (with varying depths, widths, and connectivity or topology) or designing new types of components or layers.

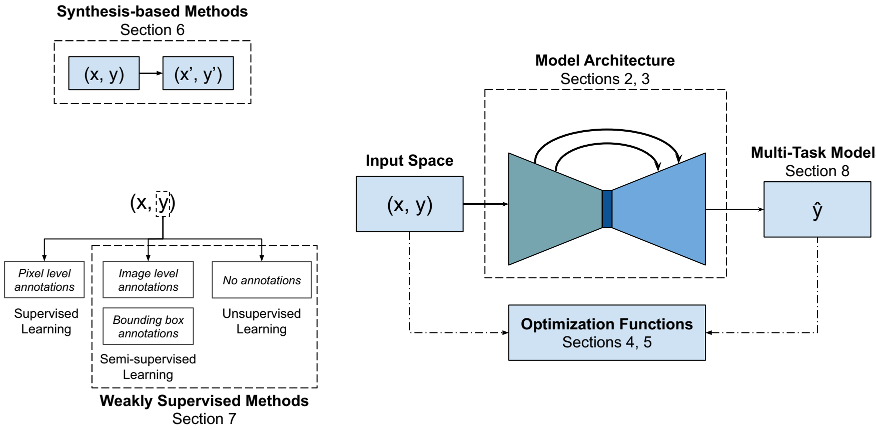


Fig. 2 A typical deep neural network based semantic segmentation pipeline. Each component in the pipeline indicates the section of this paper that covers the corresponding contributions

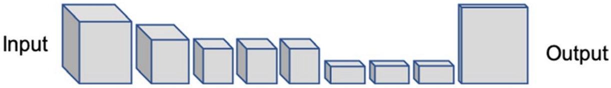


Fig. 3 Fully convolutional networks can efficiently learn to make dense predictions for per-pixel tasks like semantic segmentation (Long et al. 2015)

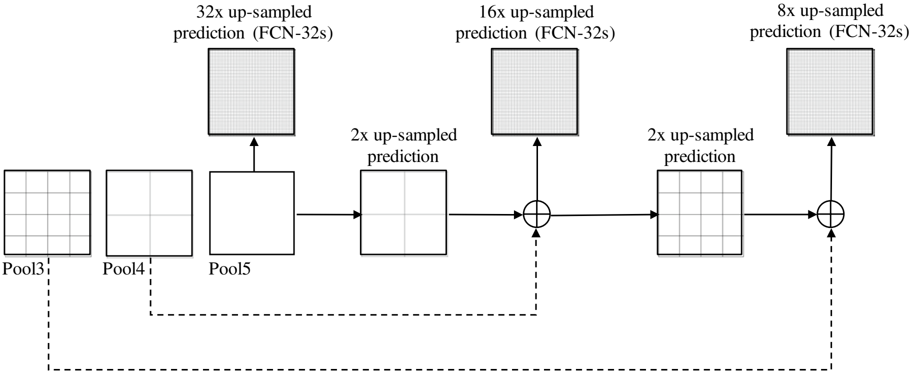

32x up-sampled prediction (FCN-32s)
16x up-sampled prediction (FCN-32s)
8x up-sampled prediction (FCN-32s)
2x up-sampled prediction
2x up-sampled prediction
Pool3
Pool4
Pool5

Fig. 4 Upsampling and fusion step of the fully convolution networks (Long et al. 2015)

2.1 Fully convolutional neural networks for semantic segmentation

As one of the first high impact CNN-based segmentation models, Long et al. (2015) proposed fully convolutional networks for pixel-wise labeling. They proposed up-sampling (deconvolving) the output activation maps from which the pixel-wise output can be calculated. The overall architecture of the network is visualized in Fig. 3.

In order to preserve the contextual spatial information within an image as the filtered input data progresses deeper into the network, Long et al. (2015) proposed to fuse the output with shallower layers' output. The fusion step is visualized in Fig. 4.

2.2 Encoder-decoder semantic image segmentation networks

Next, encoder-decoder segmentation networks (Noh et al. 2015) such as SegNet, were introduced (Badrinarayanan et al. 2015). The role of the decoder network is to map the low-resolution encoder feature to full input resolution feature maps for pixel-wise classification. The novelty of SegNet lies in the manner in which the decoder upsamples the lower resolution input feature maps. Specifically, the decoder uses pooling indices (Fig. 5) computed in the max-pooling step of the corresponding encoder to perform non-linear upsampling. The architecture (Fig. 5) consists of a sequence of non-linear processing layers (encoder) and a corresponding set of decoder layers followed by a pixel-wise classifier. Typically, each encoder consists of one or more convolutional layers with batch normalization and a ReLU non-linearity, followed by non-overlapping max-pooling and sub-sampling. The sparse encoding due to the pooling process is upsampled in the decoder using the max-pooling indices in the encoding sequence.

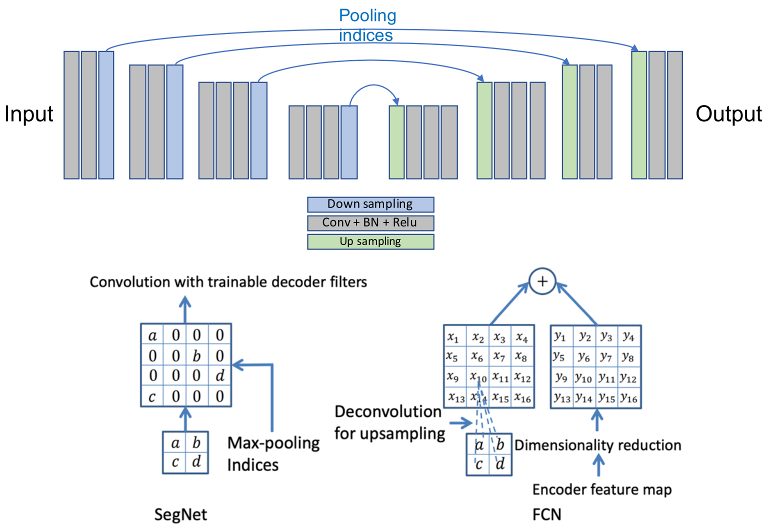


Fig. 5 **Top** An illustration of the SegNet architecture. There are no fully connected layers, and hence it is only convolutional. **Bottom** An illustration of SegNet and FCN (Long et al. 2015) decoders. *a*, *b*, *c*, *d* correspond to values in a feature map. SegNet uses the max-pooling indices to upsample (without learning) the feature map(s) and convolves with a trainable decoder filter bank. FCN upsamples by learning to deconvolve the input feature map and adds the corresponding encoder feature map to produce the decoder output. This feature map is the output of the max-pooling layer (includes sub-sampling) in the corresponding encoder. Note that there are no trainable decoder filters in FCN (Badrinarayanan et al. (2015))

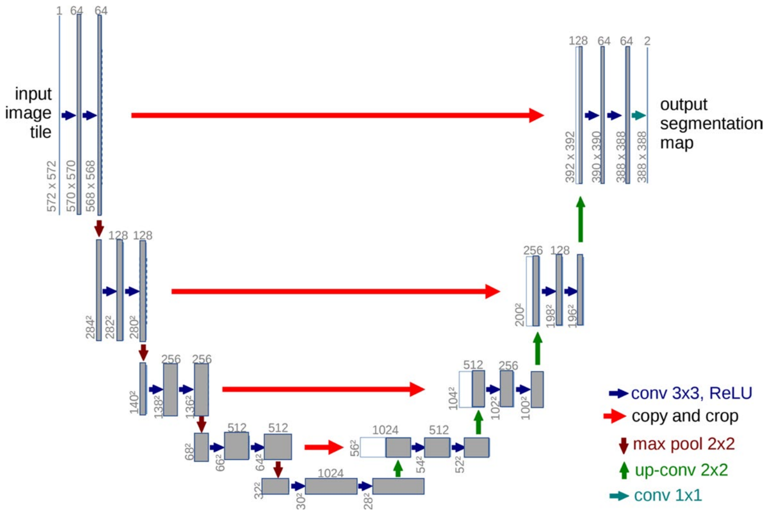


Fig. 6 An illustration of the U-Net (Ronneberger et al. 2015) architecture

Ronneberger et al. (2015) proposed an architecture (U-Net; Fig. 6) consisting of a contracting path to capture context and a symmetric expanding path that enables precise localization. Similar to the image recognition (He et al. 2016) and keypoint detection (Honari et al. 2016), Ronneberger et al. (2015) added *skip connections* to the encoder-decoder image segmentation networks, e.g., SegNet, which improved the model's accuracy and addressed the problem of vanishing gradients.

Milletari et al. (2016) proposed a similar architecture (V-Net; Fig. 7) which added residual connections and replaced 2D operations with their 3D counterparts in order to process volumetric images. Milletari et al. also proposed optimizing for a widely used segmentation metric, i.e., Dice, which will be discussed in more detail in the Sect. 4.

Jégou et al. (2017) developed a segmentation version of the densely connected networks architecture (DenseNet (Huang et al. (2017)) by adapting the U-Net like encoder-decoder skeleton. In Fig. 8, the detailed architecture of the network is visualized.

In Fig. 9, we visualize the *simplified* architectural modifications applied to the first image segmentation network i.e. FCN.

Several modified versions (e.g. deeper/shallower, adding extra attention blocks) of encoder-decoder networks have been applied to semantic segmentation (Amirul Islam et al. 2017; Fu et al. 2019b; Lin et al. 2017a; Peng et al. 2017; Pohlen et al. 2017; Wojna et al. 2017; Zhang et al. 2018d). Recently in 2018, DeepLabV3+ (Chen et al. 2018b) has outperformed many state-of-the-art segmentation networks on PASCAL VOC 2012 (Everingham et al. 2015) and Cityscapes (Cordts et al. 2016) datasets. Zhao et al. (2017b) modified the feature fusing operation proposed by Long et al. (2015) using a spatial pyramid pooling module or encode-decoder structure (Fig. 10) are used in deep neural networks for semantic segmentation tasks. The spatial pyramid networks are able to encode multi-scale contextual information by probing the incoming features with filters or pooling operations at multiple rates and multiple effective fields-of-view, while the latter networks can capture sharper object boundaries by gradually recovering the spatial information.

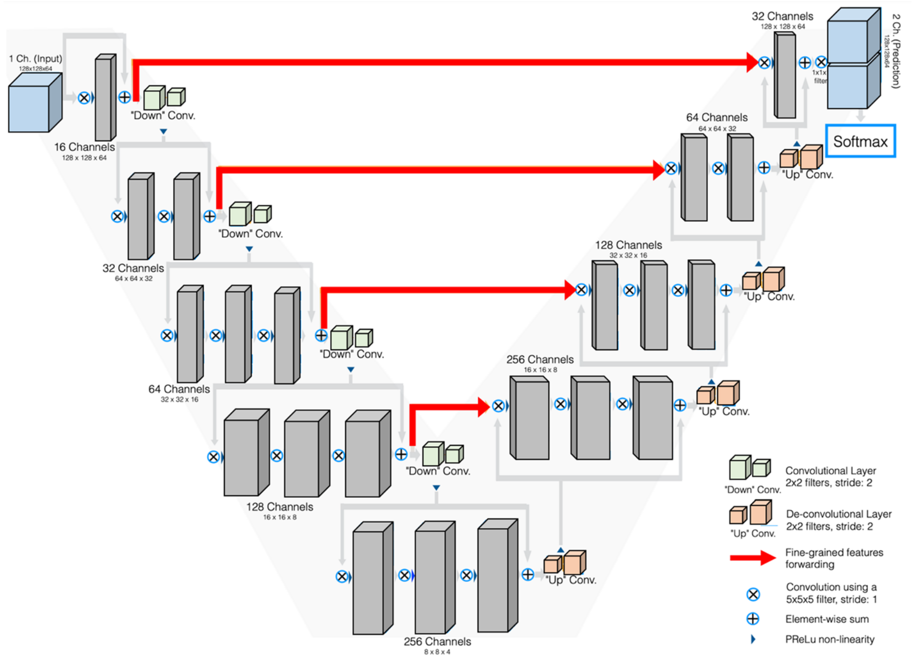


Fig. 7 An illustration of the V-Net (Milletari et al. 2016) architecture

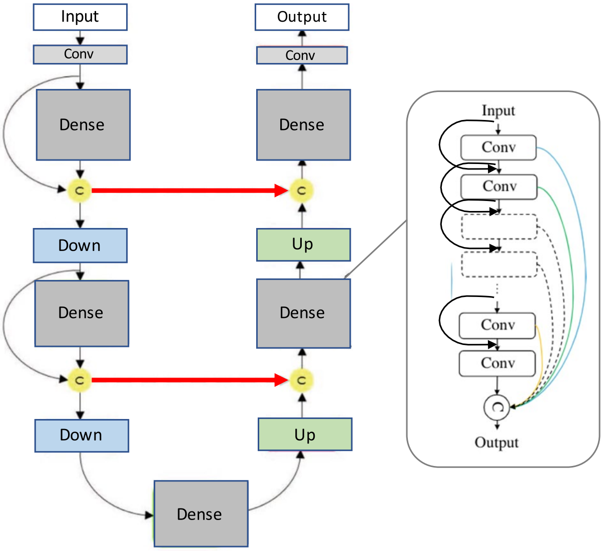


Fig. 8 Diagram of the one-hundred layers Tiramisu network architecture (Jégou et al. 2017). The architecture is built from dense blocks. The architecture is composed of a downsampling path with two transitions down and an upsampling path with two transitions up. A circle represents concatenation, and the arrows represent connectivity patterns in the network. Gray horizontal arrows represent skip connections, where the feature maps from the downsampling path are concatenated with the corresponding feature maps in the upsampling path. Note that the connectivity pattern in the upsampling and the downsampling paths are different. In the downsampling path, the input to a dense block is concatenated with its output, leading to linear growth of the number of feature maps, whereas in the upsampling path, it is not the case

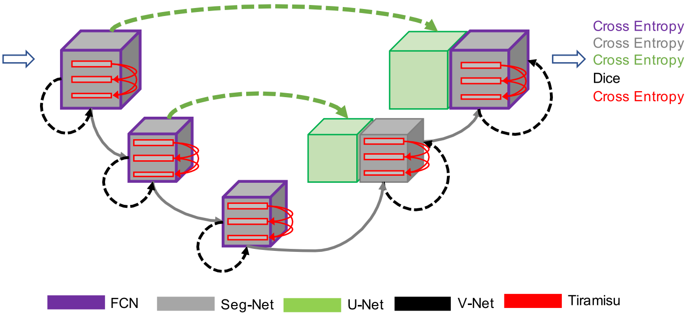


Fig. 9 Gradual architectural improvements applied to FCN (Long et al. 2015) over time

Chen et al. (2018b) proposed to combine the advantages from both dilated convolutions and feature pyramid pooling. Specifically, DeepLabv3+, extends DeepLabv3 (Chen et al. 2017b) by adding a simple yet effective decoder module (Fig. 11) to refine the segmentation results, especially along object boundaries using dilated convolutions and pyramid features.

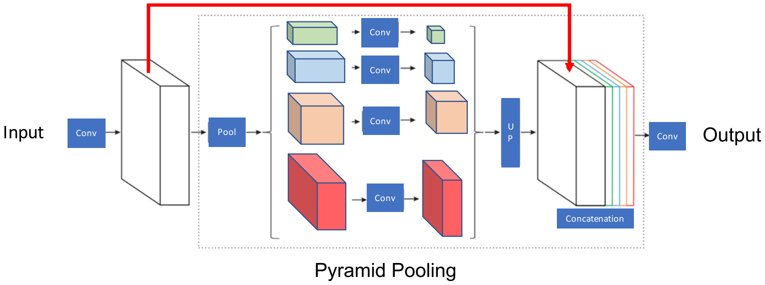


Fig. 10 Overview of the pyramid scene parsing networks. Given an input image (**a**), feature maps from last convolution layer are pulled (**b**), then a pyramid parsing module is applied to harvest different sub-region representations, followed by upsampling and concatenation layers to form the final feature representation, which carries both local and global context information in **c**. Finally, the representation is fed into a convolution layer to get the final per-pixel prediction (**d**) Zhao et al. (2017b)

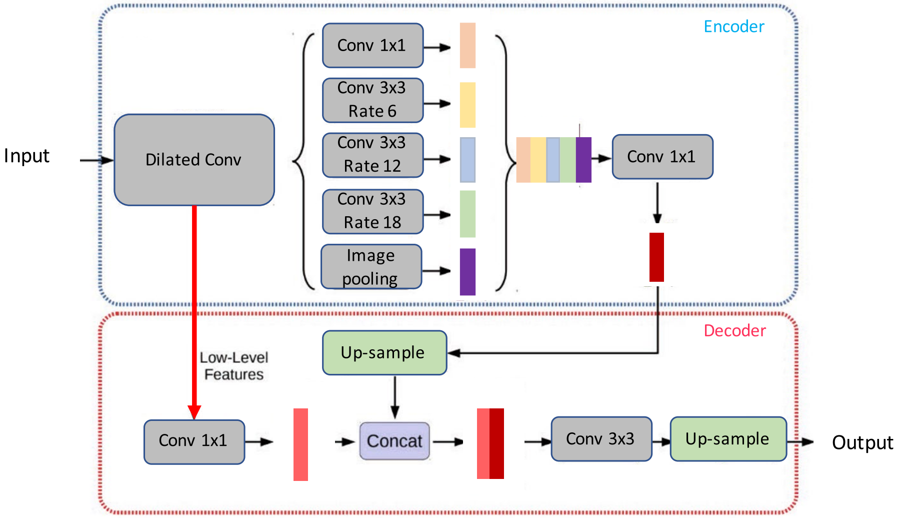


Fig. 11 An illustration of the DeepLabV3+; The encoder module encodes multi-scale contextual information by applying atrous (dilated) convolution at multiple scales, while the simple yet effective decoder module refines the segmentation results along object boundaries (Chen et al. 2018b)

2.3 Computational complexity reduction for image segmentation networks

Several works have been done on reducing the time and the computational complexity of deep classification networks (Howard et al. 2017; Leroux et al. 2018). A few other works have attempted to simplify the structure of deep networks, e.g., by tensor factorization (Kim et al. 2015), channel/network pruning (Wen et al. 2016), or applying sparsity to connections (Han et al. 2016). Similarly, Yu et al. (2018b) addressed the high computational cost associated with high resolution feature maps in U-shaped architectures by proposing spatial and context paths to preserve the rich spatial information and obtain a large receptive field. A few methods have focused on the complexity optimization of deep image segmentation networks. Similar to the work of Saxena and Verbeek (2016), Liu et al. (2019a) proposed a hierarchical neural architecture search for semantic image segmentation by performing both cell and network-level search and achieved comparable results to the state-of-the-art results on the PASCAL VOC 2012 (Everingham et al. 2015) and Cityscapes (Cordts et al. 2016) datasets. In contrast, Chen et al. (2018a) focused on searching the much smaller atrous spatial pyramid pooling module using random search. Depth-wise separable convolutions (Sifre 2014; Chollet 2017) offer computational complexity reductions since they have fewer parameters and have therefore also been used in deep segmentation models (Chen et al. 2018b; Sandler et al. 2018).

Besides network architecture search, Srivastava et al. (2015) modified ResNet in a way to control the flow of information through a connection. Lin et al. (2017a) adopted one step fusion without filtering the channels.

2.4 Attention-based semantic image segmentation

Attention can be viewed as using information transferred from several subsequent layers/ feature maps to select and localize the most discriminative (or salient) part of the input signal. Wang et al. (2017a) added an attention module to the deep residual network (ResNet) for image classification. Their proposed attention module consists of several encoding-decoding layers. Hu et al. (2018a) proposed a selection mechanism where feature maps are first aggregated using global average pooling and reduced to a single channel descriptor. Then an activation gate is used to highlight the most discriminative features. Wang et al. (2018b) proposed non-local operation blocks for encoding long range spatio-temporal dependencies with deep neural networks that can be plugged into existing architectures. Fu et al. (2019a) proposed dual attention networks that apply both spatial and channel-based attention operations.

Li et al. (2018) proposed a pyramid attention based network, for semantic segmentation. They combined an attention mechanism and a spatial pyramid to extract precise dense features for pixel labeling instead of complicated dilated convolution and artificially designed decoder networks. Chen et al. (2016) applied attention to DeepLab (Chen et al. 2017a) which takes multi-scale inputs.

2.5 Adversarial semantic image segmentation

Goodfellow et al. (2014) proposed an adversarial approach to learn deep generative models. Their generative adversarial networks (GANs) take samples z from a fixed (e.g., standard Gaussian) distribution $p_z(z)$, and transform them using a deterministic differentiable deep network $p(.)$ to approximate the distribution of training samples x. Inspired by adversarial learning, Luc et al. (2016) trained a convolutional semantic segmentation network along with an adversarial network that discriminates segmentation maps coming either from the ground truth or from the segmentation network. Their loss function is defined as

$$\begin{aligned}\ell(\boldsymbol{\theta}_s, \boldsymbol{\theta}_a) = \sum_{n=1}^{N} &\ell_{\text{mce}}(s(\boldsymbol{x}_n), \boldsymbol{y}_n) \\ &- \lambda\left[\ell_{\text{bce}}(a(\boldsymbol{x}_n, \boldsymbol{y}_n), 1) + \ell_{\text{bce}}(a(\boldsymbol{x}_n, s(\boldsymbol{x}_n)), 0)\right],\end{aligned} \tag{1}$$

where $\boldsymbol{\theta}_s$ and $\boldsymbol{\theta}_a$ denote the parameters of the segmentation and adversarial model, respectively. l_{bce} and l_{mce} are binary and multi-class cross-entropy losses, respectively. In this setup, the segmentor tries to produce segmentation maps that are close to the ground truth, i.e., which look more realistic.

The main models being used for image segmentation mostly follow encoder-decoder architectures as U-Net. Recent approaches have shown that dilated convolutions and feature pyramid pooling can improve the U-Net style networks. In Sect. 3, we summarize how these methods and their modified counterparts have been applied to medical images.

3 Architectural improvements applied to medical images

In this section, we review the different architectural based improvements for deep learning-based 2D and volumetric medical image segmentation.

3.1 Model compression based image segmentation

To perform image segmentation in real-time and be able to process larger images or (sub) volumes in case of processing volumetric and high-resolution 2D images such as CT, MRI, and histopathology images, several methods have attempted to compress deep models. Weng et al. (2019a) applied a neural architecture search method to U-Net to obtain a smaller network with a better organ/tumor segmentation performance on CT, MR, and ultrasound images. Brügger et al. (2019) by leveraging group normalization (Wu and He 2018) and leaky ReLU function, redesigned the U-Net architecture in order to make the network more memory efficient for 3D medical image segmentation. Perone et al. (2018) and Bonta and Kiran (2019) designed a dilated convolution neural network with fewer parameters as compared to the original convolution-based one. Some other works (Xu et al. 2018; Paschali et al. 2019) have focused on weight quantization of deep networks for making segmentation networks smaller.

3.2 Encoder decoder based image segmentation

Drozdzal et al. (2018) proposed to normalize input images before segmentation by applying a simple CNN prior to pushing the images to the main segmentation network. They showed improved results on electron microscopy segmentation, liver segmentation from CT, and prostate segmentation from MRI scans. Gu et al. (2019) proposed using a dilated convolution block close to the network's bottleneck to preserve contextual information.

Vorontsov et al. (2019), using a dataset defined in Cohen et al. (2018), proposed an image-to-image based framework to transform an input image with object of interest (presence domain) like a tumor to an image without the tumor (absence domain) i.e. translate diseased image to healthy; next, their model learns to add the removed tumor to the new healthy image. This results in capturing detailed structure from the object, which improves the segmentation of the object. Zhou et al. (2018) proposed a rewiring method for the long skip connections used in U-Net and tested their method on nodule segmentation in the low-dose CT scans of the chest, nuclei segmentation in the microscopy images, liver segmentation in abdominal CT scans, and polyp segmentation in colonoscopy videos.

3.3 Attention based image segmentation

Nie et al. (2018) designed an attention model to segment prostate from MRI images with higher accuracy compared to baseline models, e.g., V-Net (Milletari et al. 2016) and FCN (Long et al. 2015). Sinha and Dolz (2019) proposed a multi-level attention based architecture for abdominal organ segmentation from MRI images. Qin et al. (2018) proposed a dilated convolution base block to preserve more detailed attention in 3D medical image segmentation. Similarly, other papers (Lian et al. 2018; Isensee et al. 2019; Li et al. 2019b; Ni et al. 2019; Oktay et al. 2018; Schlemper et al. 2019) have leveraged the attention concept into medical image segmentation as well.

3.4 Adversarial training based image segmentation

Khosravan et al. (2019) proposed an adversarial training framework for pancreas segmentation from CT scans. Son et al. (2017) applied GANs for retinal image segmentation. Xue et al. (2018) used a fully convolutional network as a segmenter in the generative adversarial framework to segment brain tumors from MRI images. Other papers (Costa et al. 2017; Dai et al. 2018; Jin et al. 2018; Moeskops et al. 2017; Neff et al. 2017; Rezaei et al. 2017; Yang et al. 2017a; Zhang et al. 2017) have also successfully applied adversarial learning to medical image segmentation.

3.5 Sequenced models

The Recurrent Neural Network (RNN) was designed for handling sequences. The long short-term memory (LSTM) network is a type of RNN that introduces self-loops to enable the gradient flow for long duration (Hochreiter and Schmidhuber 1997). In the medical image analysis domain, RNNs have been used to model the temporal dependency in image sequences. Bai et al. (2018) proposed an image sequence segmentation algorithm by combining a fully convolutional network with a recurrent neural network, which incorporates both spatial and temporal information into the segmentation task. Similarly, Gao et al. (2018) applied LSTM and CNN to model temporal relationship in brian MRI slices to improve segmentation performance in 4D volumes. Li et al. (2019a) applied U-Net to obtain initial segmentation probability maps and further improve them using LSTM for pancreas segmentation from 3D CT scans. Similarly, other works have also applied RNNs (LSTMs) (Alom et al. 2019; Chakravarty and Sivaswamy 2018; Yang et al. 2017b; Zhao and Hamarneh 2019a, b) to medical image segmentation.

4 Optimization function based improvements

In addition to improved segmentation speed and accuracy using architectural modifications as mentioned in Sect. 2, designing new loss functions has also resulted in improvements in subsequent inference-time segmentation accuracy.

4.1 Cross entropy

The most commonly used loss function for the task of image segmentation is a pixel-wise cross entropy loss (Eq. 2). This loss examines each pixel individually, comparing the class predictions vector to the one-hot encoded target (or ground truth) vector. For the case of binary segmentation, let $P(Y=0)=p$ and $P(Y=1)=1-p$. The predictions are given by the logistic/sigmoid function $P(\hat{Y}=0)=\frac{1}{1+e^{-x}}=\hat{p}$ and $P(\hat{Y}=1)=1-\frac{1}{1+e^{-x}}=1-\hat{p}$, where x is output of network. Then cross entropy (CE) can be defined as:

$$\mathrm{CE}(p,\hat{p})=-(p\log(\hat{p})+(1-p)\log(1-\hat{p})). \tag{2}$$

The general form of the equation for multi-region (or multi-class) segmentation can be written as:

$$\mathrm{CE}=-\sum_{classes} p\log\hat{p} \tag{3}$$

4.2 Weighted cross entropy

The cross-entropy loss evaluates the class predictions for each pixel vector individually and then averages over all pixels, which implies equal learning to each pixel in the image. This can be problematic if the various classes have unbalanced representation in the image, as the most prevalent class can dominate training. Long et al. (2015) discussed weighting the cross-entropy loss (WCE) for each class in order to counteract a class imbalance present in the dataset. WCE was defined as:

$$\mathrm{WCE}(p,\hat{p})=-(\beta p\log(\hat{p})+(1-p)\log(1-\hat{p})). \tag{4}$$

To decrease the number of false negatives, β is set to a value larger than 1, and to decrease the number of false positives β is set to a value smaller than 1. To weight the negative pixels as well, the following balanced cross-entropy (BCE) can be used (Xie and Tu 2015).

$$\mathrm{BCE}(p,\hat{p})=-(\beta p\log(\hat{p})+(1-\beta)(1-p)\log(1-\hat{p})). \tag{5}$$

Ronneberger et al. (2015) added a distance function to the cross-entropy function to enforce learning distance between the components to enforce better segmentation in case of having very close objects to each other as follows:

$$\mathrm{BCE}(p,\hat{p})+w_0\cdot\exp\left(-\frac{\left(d_1(x)+d_2(x)\right)^2}{2\sigma^2}\right) \tag{6}$$

where $d_1(x)$ and $d_2(x)$ are two functions that calculate the distance to the border of nearest and second cells in their cell segmentation problem.

4.3 Focal loss

To reduce the contribution of easy examples so that the CNN focuses more on the difficult examples, Lin et al. (2017b) added the term $(1-\hat{p})^{\gamma}$ to the cross entropy loss as:

$$\mathrm{FL}(p,\hat{p}) = -(\alpha(1-\hat{p})^{\gamma} p \log(\hat{p}) + (1-\alpha)\hat{p}^{\gamma}(1-p)\log(1-\hat{p})). \tag{7}$$

Setting $\gamma = 0$ in this equation yields the BCE loss.

4.4 Overlap measure based loss functions

4.4.1 Dice loss/F1 score

Another popular loss function for image segmentation tasks is based on the Dice coefficient, which is essentially a measure of overlap between two samples and is equivalent to the F1 score. This measure ranges from 0 to 1, where a Dice coefficient of 1 denotes perfect and complete overlap. The Dice coefficient (DC) is calculated as:

$$\mathrm{DC} = \frac{2TP}{2TP+FP+FN} = \frac{2|X \cap Y|}{|X|+|Y|}. \tag{8}$$

Similarly, the Jaccard metric (intersection over union: IoU) is computed as:

$$\mathrm{IoU} = \frac{TP}{TP+FP+FN} = \frac{|X \cap Y|}{|X|+|Y|-|X \cap Y|} \tag{9}$$

where X and Y are the predicted and ground truth segmentation, respectively. TP is the true positives, FP false positives and FN false negatives. We can see that DC $\geq$ IoU.

To use this as a loss function the DC can be defined as a Dice loss (DL) function (Milletari et al. 2016):

$$\mathrm{DL}(p,\hat{p}) = \frac{2\langle p,\hat{p}\rangle}{\|p\|_1 + \|\hat{p}\|_1} \tag{10}$$

where $p \in \{0,1\}^n$ and $0 \leq \hat{p} \leq 1$. p and $\hat{p}$ are the ground truth and predicted segmentation and $\langle \cdot, \cdot \rangle$ denotes dot product.

4.4.2 Tversky loss

Tversky loss (TL) (Salehi et al. 2017) is a generalization of the DL. To control the level of FP and FN, TL weights them as the following:

$$\mathrm{TL}(p,\hat{p}) = \frac{\langle p,\hat{p}\rangle}{\langle p,\hat{p}\rangle + \beta\langle 1-p,\hat{p}\rangle + (1-\beta)\langle p,1-\hat{p}\rangle} \tag{11}$$

setting $\beta = 0.5$ simplifies the equation to Eq. 10.

4.4.3 Exponential logarithmic loss

Wong et al. (2018) proposed using a weighted sum of the exponential logarithmic Dice loss ($\mathcal{L}_{\text{eld}}$) and the weighted exponential cross-entropy loss ($\mathcal{L}_{\text{wece}}$) in order to improve the segmentation accuracy on small structures for tasks where there is a large variability among the sizes of the objects to be segmented.

$$\mathcal{L} = w_{\text{eld}}\mathcal{L}_{\text{eld}} + w_{\text{wece}}\mathcal{L}_{\text{wece}}, \tag{12}$$

where

$$\mathcal{L}_{\text{eld}} = \mathbf{E}\left[\left(-\ln\left(D_i\right)\right)^{\gamma_D}\right], \text{ and} \tag{13}$$

$$\mathcal{L}_{\text{wece}} = \mathbf{E}\left[\left(-\ln\left(p_l(\mathbf{x})\right)\right)^{\gamma_{CE}}\right]. \tag{14}$$

$\mathbf{x}$, i, and l denote the pixel position, the predicted label, and the ground truth label. D_i denotes the smoothed Dice loss (obtained by adding an $\epsilon = 1$ term to the numerator and denominator in Eq. 10 in order to handle missing labels while training, and γ_D and γ_{CE} are used to control the non-linearities of the respective loss functions.

4.4.4 Lovász-softmax loss

Since it has been shown that the Jaccard loss (IoU loss) is submodular (Berman et al. 2018a), Berman et al. (2018b) proposed using the Lovász hinge with the Jaccard loss for binary segmentation, and proposed a surrogate of the Jaccard loss, called the Lovász-Softmax loss, which can be applied for the multi-class segmentation task. The Lovász-Softmax loss is, therefore, a smooth extension of the discrete Jaccard loss, and is defined as

$$\mathcal{L}_{\text{LovaszSoftmax}} = \frac{1}{|\mathcal{C}|}\sum_{c\in\mathcal{C}} \overline{\Delta_{J_c}}(\boldsymbol{m}(c)), \tag{15}$$

where $\Delta_{J_c}(\cdot)$ denotes the convex closure of the submodular Jaccard loss, $\bar{}$ denotes that it is a tight convex closure and polynomial time computable, $\mathcal{C}$ denotes all the classes, and J_c and $\boldsymbol{m}(c)$ denote the Jaccard index and the vector of errors for class c respectively.

4.4.5 Boundary loss

Kervadec et al. (2019a) proposed to calculate boundary loss $\mathcal{L}_B$ along with the generalized Dice loss $\mathcal{L}_{GD}$ function as

$$\alpha\mathcal{L}_{GD}(\theta) + (1-\alpha)\mathcal{L}_B(\theta), \tag{16}$$

where the two terms in the loss function are defined as

$$\mathcal{L}_{GD}(\theta) = 1 \; - 2\frac{w_G \sum_{p\in\Omega} g(p)s_\theta(p) + w_B \sum_{p\in\Omega}(1-g(p))\left(1-s_\theta(p)\right)}{w_G \sum_{p\in\Omega}\left[s_\theta(p)+g(p)\right] + w_B \sum_{p\in\Omega}\left[2-s_\theta(p)-g(p)\right]}, \text{ and} \tag{17}$$

$$\mathcal{L}_B(\theta) = p \in \Omega\phi_G(p)s_\theta(p), \tag{18}$$

where $\phi_G(p) = -\|p - z_{\partial G}(p)\|$ if $p \in G$ and $\phi_G(p) = \|p - z_{\partial G}(p)\|$, otherwise. The general form integral $\sum_\Omega g(p)f\left(s_\theta(p)\right)$ is for foreground and $\sum_\Omega (1 - g(p))f\left(1 - s_\theta(p)\right)$ for background. $w_G = 1/\left(\sum_{p\in\Omega} g(p)\right)^2$ and $w_B = 1/\left(\sum_\Omega (1 - g(p))\right)^2$.Ω shows the spatial domain.

4.4.6 Conservative loss

Zhu et al. (2018) proposed the Conservative Loss for in order to achieve a good generalization ability in domain adaptation tasks by penalizing the extreme cases and encouraging the moderate cases. The Conservative Loss is defined as

$$CL(p_t) = \lambda(1 + \log_a(p_t))^2 * \log_a(-\log_a(p_t)), \tag{19}$$

where p_t is the probability of the prediction towards the ground truth and a is the base of the logarithm. a and λ are empirically chosen to be e (Euler's number) and 5 respectively.

Other works also include approaches to optimize the segmentation metrics (Nowozin 2014), weighting the loss function (Roy et al. 2017), and adding regularizers to loss functions to encode geometrical and topological shape priors (BenTaieb and Hamarneh 2016; Mirikharaji and Hamarneh 2018).

A significant problem in image segmentation (particularly in medical images) is to overcome class imbalance for which overlap measure based methods have shown reasonably good performance in overcoming the imbalance. In Sect. 5, we summarize the approaches which use new loss functions, particularly for medical image segmentation or use the (modified) loss functions mentioned above.

In Fig. 12, we visualize the behavior of different loss functions for segmenting large and small objects. For the parameters of the loss functions, we use the same parameters as reported by the authors in their respective papers. Therefore, we use $\beta = 0.3$ in Eq. 11, $\alpha = 0.25$ and $\gamma = 2$ in Eq. 7, and $\gamma_D = \gamma_{CE} = 1$, $w_{\text{eld}} = 0.8$, and $w_{\text{wece}} = 0.2$ in Eq. 12. Moving from the left to the right for each plot, the overlap of the predictions and ground truth mask becomes progressively smaller, i.e., producing more false positives and false negatives. Ideally, the loss value should monotonically increase as more false positives, and

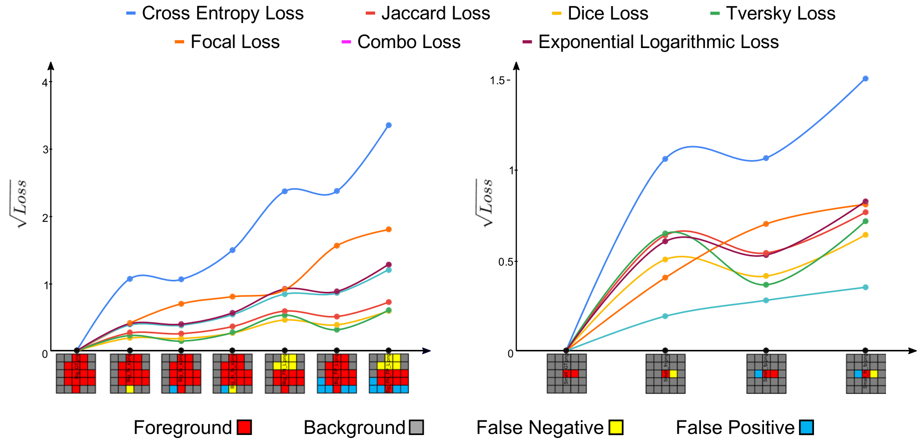


Fig. 12 A comparison of seven loss functions for different extends of overlaps for a large (left) and a small (right) object

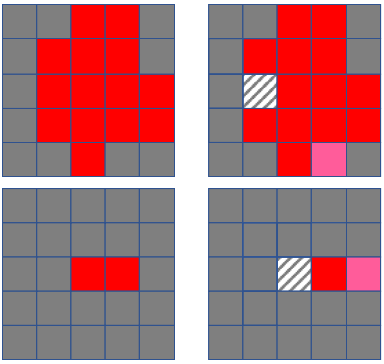


Fig. 13 Comparison of cross entropy and Dice losses for segmenting small and large objects. The red pixels show the ground truth and the predicted foregrounds in the left and right columns respectively. The striped and the pink pixels indicate false negative and false positive, respectively. For the top row (i.e., large foreground), the Dice loss returns 0.96 for one false negative and for the bottom row (i.e., small object) returns 0.66 for one false negative, whereas the cross entropy loss function outputs 0.83 for both the cases. By considering a false negative and false positive, the output value drops even more in case of using Dice but the cross entropy stays smooth (i.e., Dice value of 0.93 and 0.50 for large and small object versus cross entropy loss value of 1.66 for both.)

negatives are predicted. For large objects, almost all the functions follow this assumption; however, for the small objects (right plot), only combo loss and focal loss penalize *monotonically* more for larger errors. In other words, the overlap-based functions highly fluctuate while segmenting small and large objects (also see Fig. 13), which results in unstable optimization. The loss functions which use cross-entropy as the base and the overlap measure functions as a weighted regularizer show more *stability* during training.

5 Optimization function based improvements applied to medical images

The standard CE loss function and its weighted versions, as discussed in Sect. 4, have been applied to numerous medical image segmentation problems (Isensee et al. 2019; Li et al. 2019b; Lian et al. 2018; Ni et al. 2019; Nie et al. 2018; Oktay et al. 2018; Schlemper et al. 2019). However, Milletari et al. (2016) found that optimizing CNNs for DL (Eq. 10) in some cases, e.g., in the case of having very small foreground objects in a large background, works better than the original cross-entropy.

Li et al. (2019c) proposed adding the following regularization term to the cross entropy loss function to encourage smooth segmentation outputs.

$$R = \sum_{i=1}^{N} \mathbb{E}_{\xi',\xi} \left\| f\left(x_i;\theta,\xi'\right) - f\left(x_i;\theta,\xi\right) \right\|^2 \tag{20}$$

where ξ' and ξ are different perturbation (e.g., Gaussian noise, network dropout, and randomized data transformation) applied to the input image x_i.

Chen et al. (2019) proposed leveraging traditional active contour energy minimization into CNNs via the following loss function.

$$\text{Loss} = \text{Length} + \lambda \cdot \text{Region} \tag{21}$$

$$\text{Length} = \sum_{\Omega}^{i=1,j=1} \sqrt{\left| \left(\nabla u_{x_{i,j}}\right)^2 + \left(\nabla u_{y_{i,j}}\right)^2 \right| + \epsilon} \tag{22}$$

where x and y from $u_{x_{i,j}}$ and $u_{y_{i,j}}$ are horizontal and vertical directions, respectively.

$$\text{Region} = \left| \sum_{\Omega}^{i=1,j=1} u_{i,j}\left(c_1 - v_{i,j}\right)^2 \right| + \left| \sum_{\Omega}^{i=1,j=1} \left(1 - u_{i,j}\right)\left(c_2 - v_{i,j}\right)^2 \right| \tag{23}$$

where u and v are represented as prediction and a given image, respectively. c1 is set to 1 and c2 to 0. Similar to, Li et al. (2019c), Zhou et al. (2019a) proposed adding a contour regression term to the weighted cross entropy loss function.

Karimi and Salcudean (2019) optimized Hausdorff distance based function between a predicted and ground truth segmentation as follows.

$$f_{\text{HD}}(p, q) = \text{Loss}(p, q) + \lambda\left(1 - \frac{2\sum_{\Omega}(p \circ q)}{\sum_{\Omega}\left(p^2 + q^2\right)} \right) \tag{24}$$

where the second term is the Dice loss function and the first term can be replaced with three different versions of the Hausdorff distance for p and q i.e. ground truth and predicted segmentations respectively, as follows;

$$\text{Loss}(q, p) = \frac{1}{|\Omega|} \sum_{\Omega} \left((p - q)^2 \circ \left(d_p^{\alpha} + d_q^{\alpha} \right) \right) \tag{25}$$

The parameter α determines the level of penalty for larger errors. d_p is the distance map of the ground-truth segmentation as the unsigned distance to the boundary δp. Similarly, d_q is defined as the distance to δq. The $\circ$ is Hadamard operation.

$$\text{Loss}(q, p) = \frac{1}{|\Omega|} \sum_{k=1}^{K} \sum_{\Omega} \left((p - q)^2 \ominus_k B \right) k^{\alpha} \tag{26}$$

where $\ominus_k$ denotes k successive erosions. where

$$B = \begin{pmatrix} 0 & 1/5 & 0 \\ 1/5 & 1/5 & 1/5 \\ 0 & 1/5 & 0 \end{pmatrix} \tag{27}$$

$$\begin{aligned} \text{Loss}(q, p) = & \frac{1}{|\Omega|} \sum_{r \in R} r^{\alpha} \sum_{\Omega} \left[f_s\left(B_r * \overline{p}^C\right) \circ f_{\overline{q} \backslash \overline{p}} + f_s\left(B_r * \overline{p}\right) \circ f_{\overline{p} \backslash \overline{q}} \right. \\ & \left. + f_s\left(B_r * \overline{q}^C\right) \circ f_{\overline{p} \backslash \overline{q}} + f_s\left(B_r * \overline{q}\right) \circ f_{\overline{q} \backslash \overline{p}} \right] \end{aligned} \tag{28}$$

where $f_{\overline{q} \backslash \overline{p}} = (p - q)^2 q$. f_s indicates soft thresholding. B_r denotes a circular-shaped convolutional kernel of radius r. Elements of B_r are normalized such that they sum to one. $\overline{p}^C = 1 - \overline{p}$. Ground-truth and predicted segmentations, denoted with $\overline{p}$ and $\overline{q}$, Caliva et al. (2019) proposed to measure distance of each voxel to the boundaries of the objects and use the weight matrices to penalize a model for error on the boundaries. Kim and Ye (2019) proposed using level-set energy minimization as a regularizer summed with standard multi-class cross entropy loss function for semi-supervised brain MRI segmentation as:

$$\mathcal{L}_{\text{level}}(\Theta;x) = \sum_{n=1}^{N} \int_{\Omega} \left| x(r) - c_n^{\Theta} \right|^2 y_n^{\Theta}(r)dr + \lambda \sum_{n=1}^{N} \int_{\Omega} \left| \nabla y_n^{\Theta}(r) \right| dr \tag{29}$$

with

$$c_n^{\Theta} = \frac{\int_{\Omega} x(r) y_n^{\Theta}(r)dr}{\int_{\Omega} y_n^{\Theta}(r)dr} \tag{30}$$

where $x(r)$ is the input, $y_n^{\Theta}(r)$ is the output of softmax layer, Θ refers to learnable parameters.

Taghanaki et al. (2019e) discussed the risks of using solo overlap based loss functions and proposed to use them as regularizes along with a weighted cross entropy to explicitly handle input and output imbalance as follows;

$$\begin{aligned} \mathit{Combo\ Loss} =& \alpha\left(-\frac{1}{N}\sum_{i=1}^{N} \beta\left(t_i - \ln p_i\right) + (1-\beta)\left[\left(1-t_i\right)\ln\left(1-p_i\right)\right]\right) \\ &+ (1-\alpha)\sum_{i=1}^{K}\left(-\frac{2\sum_{i=1}^{N} p_i t_i + S}{\sum_{i=1}^{N} p_i + \sum_{i=1}^{N} t_i + S}\right) \end{aligned} \tag{31}$$

where α controls the amount of Dice term contribution in the loss function L, and $\beta \in [0, 1]$ controls the level of model penalization for false positives/negatives: when β is set to a value smaller than 0.5, FP are penalized more than FN as the term $(1 - t_i)\ln(1 - p_i)$ is weighted more heavily, and vice versa. In their implementation, to prevent division by zero, the authors perform add-one smoothing (a specific instance of the additive/Laplace/Lidstone smoothing; Russell and Norvig 2016), i.e., they add unity constant S to both the denominator and numerator of the Dice term.

The majority of the methods discussed in Sect. 5 have attempted to handle the class imbalance issue in the input images i.e., small foreground versus large background with providing weights/penalty terms in the loss function. Other approaches consist of first identifying the object of interest, cropping around this object, and then performing the task (e.g., segmentation) with better-balanced classes. This type of cascade approach has been applied for the segmentation of multiple sclerosis lesions in the spinal cord (Gros et al. 2019).

6 Image synthesis based methods

Deep CNNs are heavily reliant on big data to avoid overfitting and class imbalance issues, and therefore this section focuses on data augmentation, a data-space solution to the problem of limited data. Apart from standard online image augmentation methods such as geometric transformations (LeCun et al. 1998; Simard et al. 2003; Cireşan et al. 2011, 2012; Krizhevsky et al. 2012), color space augmentations (Galdran et al. 2017; Yuan 2017; Abhishek et al. 2020), etc., in this section, we discuss image synthesis methods, the output of which are novel images rather than modifications to existing images. GANs based augmentation techniques for segmentation tasks have been used for a wide variety of problems - from remote sensing imagery (Mohajerani et al. 2019) to filamentary anatomical structures (Zhao et al. 2017a). For a more detailed review of image augmentation strategies in deep learning, we direct the interested readers to Shorten and Khoshgoftaar (2019).

6.1 Image synthesis based methods applied to natural image segmentation

Neff et al. (2018) trained a Wasserstein GAN with gradient penalty (Gulrajani et al. 2017) to generate labeled image data in the form of image-segmenation mask pairs. They evaluated their approach on a dataset of chest X-ray images and the Cityscapes dataset, and found that the WGAN-GP was able to generate images with sufficient variety and that a segmentation model trained using GAN-based augmentation only was able to perform better than a model trained with geometric transformation based augmentation. Cherian and Sullivan (2019) proposed to incorporate semantic consistency in image-to-image translation task by introducing segmentation functions in the GAN architecture and showed that the semantic segmentation models trained with synthetic images led to considerable performance improvements. Other works include GAN-based data augmentation for domain adaptation (Huang et al. 2018; Choi et al. 2019) and panoptic data augmentation (Liu et al. 2019c). However, the majority of GAN based data augmentation has been applied to medical images (Shorten and Khoshgoftaar 2019). Next, we discuss the GAN based image synthesis for augmentation in the field of medical image analysis.

6.2 Image synthesis based methods applied to medical image segmentation

Chartsias et al. (2017) used a conditional GAN to generate cardiac MR images from CT images. They showed that utilizing the synthetic data increased the segmentation accuracy and that using only the synthetic data led to only a marginal decrease in the segmentation accuracy. Similarly, Zhang et al. (2018c) proposed a GAN based volume-to-volume translation for generating MR volumes from corresponding CT volumes and vice versa. They showed that synthetic data improve segmentation performance on cardiovascular MRI volumes. Huo et al. (2018) proposed an end-to-end synthesis and segmentation network called EssNet to simultaneously synthesize CT images from unpaired MR images and to segment CT splenomegaly on unlabeled CT images and showed that their approach yielded better segmentation performance than even segmentation obtained using models trained using the manual CT labels. Abhishek and Hamarneh (2019) trained a conditional GAN to generate skin lesion images from and confined to binary masks, and showed that using the synthesized images led to a higher skin lesion segmentation accuracy. Zhang et al. (2018b) trained a GAN for translating between digitally reconstructed radiographs and X-ray images and achieved similar accuracy as supervised training in multi-organ segmentation. Shin et al. (2018) proposed a method to generate synthetic abnormal MRI images with brain tumors by training a GAN using two publicly available data sets of brain MRI. Similarly, other works (Han et al. 2019; Yang et al. 2018; Yu et al. 2018a) have leveraged GANs to synthesize brain MR images.

7 Weakly supervised methods

Collecting large-scale accurate pixel-level annotation is time-consuming and financially expensive. However, unlabeled and weakly-labeled images can be collected in large amounts in a relatively fast and cheap manner. As shown in Fig. 2, varying levels of supervision are possible when training deep segmentation models, from pixel-wise annotations (supervised learning) and image-level and bounding box annotations (semi-supervised learning) to no annotations at all (unsupervised learning), the last two of which comprise weak supervision. Therefore, a promising direction for semantic image segmentation is to develop weakly supervised segmentation models.

7.1 Weakly supervised methods applied to natural images

Kim and Hwang (2016) proposed a weakly supervised semantic segmentation network using unpooling and deconvolution operations, and used feature maps from the deconvolutions layers to learn scale-invariant features, and evaluated their model on the PASCAL VOC and chest X-ray image datasets. Lee et al. (2019) used dropout (Srivastava et al. 2014) to choose features at random during training and inference and combine the many different localization maps to generate a single localization map, effectively discovering relationships between locations in an image, and evaluated their proposed approach on the PASCAL VOC dataset.

7.2 Weakly supervised methods applied to medical images

The scarcity of richly annotated medical images is limiting supervised deep learning-based solutions to medical image analysis tasks (Perone and Cohen-Adad 2019), such as localizing discriminatory radiomic disease signatures. Therefore, it is desirable to leverage unsupervised and weakly supervised models. Kervadec et al. (2019b) introduced a differentiable term in the loss function for datasets with weakly supervised labels, which reduced the computational demand for training while also achieving almost similar performance to full supervision for segmentation of cardiac images. Afshari et al. (2019) used a fully convolutional architecture along with a Mumford-Shah functional Mumford and Shah (1989) inspired loss function to segment lesions from PET scans using only bounding box annotations as supervision. Mirikharaji et al. (2019) proposed to learn spatially adaptive weight maps to account for spatial variations in pixel-level annotations and used noisy annotations to train a segmentation model for skin lesions. Taghanaki et al. (2019d) proposed to learn spatial masks using only image-level labels with minimizing mutual information between the input and masks, and at the same time maximizing the mutual information between the masks and image labels. Peng et al. (2019) proposed an approach to train a CNN with discrete constraints and regularization priors based on the alternating direction method of multipliers (ADMM). Perone and Cohen-Adad (2018) expanded the semi-supervised mean teacher (Tarvainen and Valpola 2017) approach to segmentation tasks on MRI data, and show that it can bring important improvements in a realistic small data regime. In another work, Perone et al. (2019) extended the method of unsupervised domain adaptation using self-ensembling for the semantic segmentation task. They showed how this approach could improve the generalization of the models even when using a small amount of unlabeled data.

8 Multi-task models

Multi-task learning (Caruana 1997) refers to a machine learning approach where multiple tasks are learned simultaneously, and the learning efficiency and the model performance on each of the tasks are improved because of the existing commonalities across the tasks. For visual recognition tasks, it has been shown that there exist relations between various tasks in the task space (Zamir et al. 2018), and multi-task models can help exploit these relationships to improve performance on the related tasks.

8.1 Multi-task models applied to natural images

Bischke et al. (2019) proposed a cascaded multi-task loss to preserve boundary information from segmentation masks for segmenting building footprints and achieved state-of-the-art performance on an aerial image labeling task. He et al. (2017) extended Faster R-CNN (Ren et al. 2015) by adding a new branch to predict the object mask along with a class label and a bounding box, and the proposed model was called Mask R-CNN. Mask R-CNN has been used extensively for multi-task segmentation models for a wide range of application areas (Abdulla et al. 2017), such as adding sports fields to OpenStreetMap (Remillard 2018), detection and segmentation for surgery robots (SUYEgit 2018), understanding climate change patterns from aerial imagery of the Arctic (Zhang et al. 2018a), converting satellite imagery to maps (Mohanty 2018), detecting image forgeries (Wang et al. 2019d), and segmenting tree canopy (Zhao et al. 2018).

8.2 Multi-task models applied to medical images

Chaichulee et al. (2017) extended the VGG16 architecture (Simonyan and Zisserman 2014) to include a global average pooling layer for patient detection and a fully convolutional network for skin segmentation. The proposed model was evaluated on images from a clinical study conducted at a neonatal intensive care unit, and was robust to changes in lighting, skin tone, and pose. He et al. (2019) trained a U-Net (Ronneberger et al. 2015)-like encoder-decoder architecture to simultaneously segment thoracic organs from CT scans and perform global slice classification. Ke et al. (2019) trained a multi-task U-Net architecture to solve three tasks - separating wrongly connected objects, detecting class instances, and pixelwise labeling for each object, and evaluated it on a food microscopy image dataset. Other multi-task models have also been proposed for segmentation and classification for detecting manipulated faces in images and video (Nguyen et al. 2019) and diagnosis of breast biopsy images (Mehta et al. 2018) and mammograms (Le et al. 2019).

Mask R-CNN has also been used for segmentation tasks in medical image analysis such as automatically segmenting and tracking cell migration in phase-contrast microscopy (Tsai et al. 2019), detecting and segmenting nuclei from histological and microscopic images (Johnson 2018; Vuola et al. 2019; Wang et al. 2019a, b), detecting and segmenting oral diseases (Anantharaman et al. 2018), segmenting neuropathic ulcers (Gamage et al. 2019), and labeling and segmenting ribs in chest X-rays (Wessel et al. 2019). Mask R-CNN has also been extended to work with 3D volumes and has been evaluated on lung nodule detection and segmentation from CT scans and breast lesion detection and categorization on diffusion MR images (Jaeger et al. 2018; Kopelowitz and Engelhard 2019).

9 Segmentation evaluation metrics and datasets

9.1 Evaluation metrics

The quantitative evaluation of segmentation models can be performed using pixel-wise and overlap based measures. For binary segmentation, pixel-wise measures involve the construction of a confusion matrix to calculate the number of true positive (TP), true negative (TN), false positive (FP), and false negative (FN) pixels, and then calculate various metrics such as precision, recall (also known as sensitivity), specificity, and overall pixel-wise accuracy. They are defined as follows:

$$\text{Precision} = \frac{TP}{TP + FP}, \tag{32}$$

$$\text{Recall or Sensitivity} = \frac{TP}{TP + FN}, \tag{33}$$

$$\text{Specificity} = \frac{TN}{TN + FP}, \quad \text{and,} \tag{34}$$

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}. \tag{35}$$

Two popular overlap-based measures used to evaluate segmentation performance are the Sørensen–Dice coefficient (also known as the Dice coefficient) and the Jaccard index (also known as the intersection over union or IoU). Given two sets $\mathcal{A}$ and $\mathcal{B}$, these metrics are defined as:

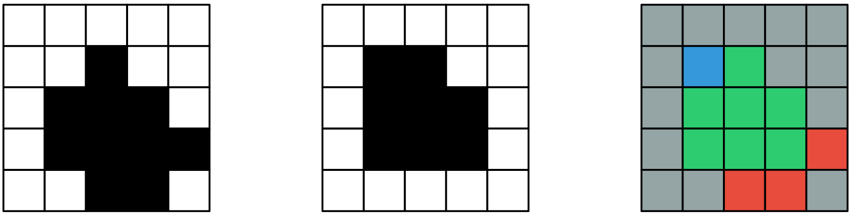
$$\text{Dice coefficient, Dice}(\mathcal{A}, \mathcal{B}) = 2\,\frac{|\mathcal{A} \cap \mathcal{B}|}{|\mathcal{A}| + |\mathcal{B}|}, \quad \text{and,} \tag{36}$$

$$\text{Jaccard index, Jaccard}(\mathcal{A}, \mathcal{B}) = \frac{|\mathcal{A} \cap \mathcal{B}|}{|\mathcal{A} \cup \mathcal{B}|}. \tag{37}$$

For binary segmentation masks, these overlap-based measures can also be calculated from the confusion matrix as shown in Eqs. 8 and 9 respectively. The two measures are related by:

$$\text{Jaccard} = \frac{\text{Dice}}{2 - \text{Dice}}. \tag{38}$$

Figure 14 contains a simple overlap scenario, with the ground truth and the predicted binary masks with a spatial resolution 5×5. Let black pixels denote the object to be segmented. The confusion matrix for this can be constructed as shown in Table 1. Using the expressions above, we can calculate the metrics as precision $= \frac{7}{8} = 0.875$,



(a) Ground truth binary mask **(b)** Predicted binary mask **(c)** Overlap between the masks.

Fig. 14 A 5 × 5 overlap scenario with **a** the ground truth, **b** the predicted binary masks, and **c** the overlap. In **a** and **b**, black and white pixels denote the foreground and the background respectively. In **c**, green, grey, blue, and red pixels denote TP, TN, FP, and FN pixels respectively

Table 1 Confusion matrix for the overlap scenario shown in Fig. 14

		Ground truth	
		Background	Object
Prediction	Background	14	3
	Object		7

recall $= \frac{7}{10} = 0.7$, specificity $= \frac{14}{15} = 0.9333$, pixel-wise accuracy $= \frac{21}{25} = 0.84$, Dice coefficient $= \frac{7}{9} = 0.7778$, and Jaccard index $= \frac{7}{11} = 0.6364$.

9.2 Semantic segmentation datasets for natural images

Next, we briefly discuss the most popular and widely used datasets for the semantic segmentation of natural images. These datasets cover various categories of scenes, such as indoor and outdoor environments, common objects, urban street view as well as generic scenes. For a comprehensive review of the natural image datasets that segmentation models are usually benchmarked upon, we direct the interested readers to Lateef and Ruichek (2019).

- *Pascal VOC datasets* The PASCAL Visual Object Classes (VOC) Challenge (Everingham et al. 2010) was an annual challenge that ran from 2005 through 2012 and had annotations for several tasks such as classification, detection, and segmentation. The segmentation task was first introduced in the 2007 challenge and featured objects belonging to 20 classes. The last offering of the challenge, the PASCAL VOC 2012 challenge, contained segmentation annotations for 2913 images across 20 object classes (Everingham et al. 2015).
- *PASCAL Context* The PASCAL Context dataset (Mottaghi et al. 2014) extended the PASCAL VOC 2010 Challenge dataset by providing pixel-wise annotations for the images, resulting in a much larger dataset with 19,740 annotated images and labels belonging to 540 categories.
- *Cityscapes* The Cityscapes dataset (Cordts et al. 2016) contains annotated images of urban street scenes. The data was collected during daytime from 50 cities and exhibits variance in the season of the year and traffic conditions. Semantic, instance wise, and dense pixel-

wise annotations are provided, with 'fine' annotations for 5,000 images and 'coarse' annotations for 20,000 images.
- *ADE20K* The ADE20K dataset (Zhou et al. 2017) contains 25,210 images from other existing datasets, e.g, the LabelMe (Russell et al. 2008), the SUN (Xiao et al. 2010), and the Places (Zhou et al. 2014) datasets. The images are annotated with labels belonging to 150 classes for "scenes, objects, parts of objects, and in some cases even parts of parts".
- *CamVid* The Cambridge-driving Labeled Video Database (CamVid) (Brostow et al. 2008, 2009) contains 10 min of video captured at 30 frames per second from a driving automobile's perspective, along with pixel-wise semantic segmentation annotations for 701 frames and 32 object classes.

Table 2 lists a summary of selected papers from this review, the nature of their proposed contributions, and the datasets that they were evaluated on. For the papers that evaluated their models on the PASCAL VOC 2012 dataset (Everingham et al. 2012), one of the most popular image semantic segmentation dataset for natural images, we also list their reported mean IoU scores. As can be seen in Table 2, the focus has been mostly on architectural improvements. Comparing the first deep learning-based model (i.e., FCN Long et al. 2015) to the state-of-the-art model (i.e., DeepLabV3+ Chen et al. 2018b) there is a large improvement (i.e. ∼ 27%, i.e., 62.2–89.0%) in terms of mean IoU. The latter model leverages a more sophisticated decoder, dilated convolutions, and feature pyramid pooling.

9.3 Semantic segmentation datasets for medical images

In contrast to natural images, it is difficult to tabulate and summarize the performance of medical image segmentation methods because of the vast number of (a) medical imaging modalities and (b) medical image segmentation datasets. Figure 15 presents a breakdown of the various attributes of the medical image segmentation papers surveyed in this review, color coded similar to Fig. 1. As shown in Fig. 15b, the papers covered in this review use 13 medical imaging modalities. Figure 15c shows the distribution of the number of samples across datasets from multiple modalities. We observe that modalities which are expensive to acquire and annotate (such as electron microscopy (EM), PET, and MRI) have smaller dataset sizes than relative cheaper to acquire modalities such as RGB images (e.g., skin lesion images), ultrasound (US) and X-ray images. We also present a summary of the popular medical image segmentation papers in Table 3 and include the entire table in the Supplementary Material.

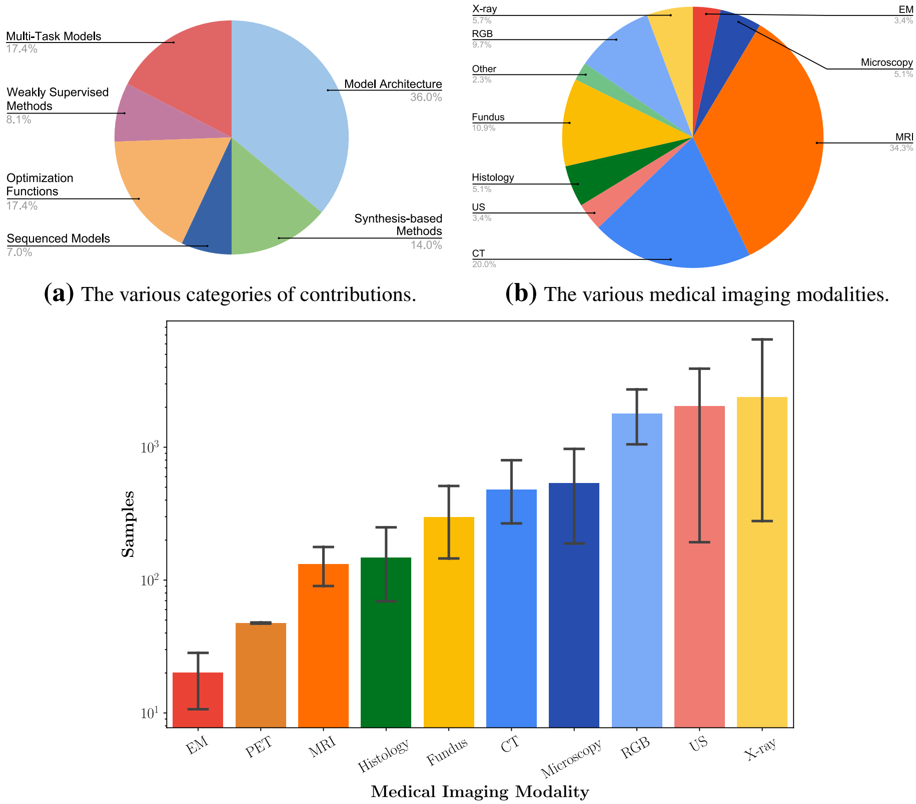
A similar observation can be made by looking at the medical image segmentation competitions. Grand Challenges in Biomedical Image Analysis (Challenge 2020) provides a comprehensive but not exhaustive list of publicly available medical image segmentation challenges, and since 2007, there have been 94 segmentation challenges for medical images and volumes from as many as 12 imaging modalities. Figure 16 shows the number of these challenges for every year since 2007, and it can be seen that this number has been on the rise in the past few years.

10 Discussion and future directions

In the following sections, we discuss in detail the potential future research directions for semantic segmentation of natural and medical images.

Table 2 A summary of papers for semantic segmentation of natural images applied to PASCAL VOC 2012 dataset

Paper	Type of improvement	Dataset(s) evaluated on	PASCAL VOC 2012 mean (%)
SegNet (Noh et al. 2015)	Architecture	PASCAL VOC, CamVid, SUN RGB-D	59.1
FCN (Long et al. 2015)	Architecture	PASCAL VOC, NYUDv2, SIFT Flow	62.2
Luc et al. (2016)	Adversarial segmentation	PASCAL VOC, stanford background	73.3
Lovász-Softmax loss Berman et al. (2018b)	Loss	PASCAL VOC, Cityscapes	76.44
Large kernel matters (Peng et al. 2017)	Architecture	PASCAL VOC, Cityscapes	82.2
Deep layer cascade (Li et al. 2017)	Architecture	PASCAL VOC, Cityscapes	82.7
TuSimple (Wang et al. 2018a)	Architecture	PASCAL VOC, KITTI Road Estimation	83.1
RefineNet (Lin et al. 2017a)	Architecture	PASCAL VOC, PASCAL Context, Person-Part, NYUDv2, SUN RGB-D, Cityscapes, ADE20K	84.2
ResNet-38 (Wu et al. 2019)	Architecture	PASCAL VOC, PASCALContext, Cityscapes	84.9
PSPNet (Zhao et al. 2017b)	Architecture	PASCAL VOC, Cityscapes	85.4
Auto-DeepLab (Liu et al. 2019a)	Architecture search	PASCAL VOC, ADE20K, Cityscapes	85.6
IDW-CNN (Wang et al. 2017b)	Architecture	PASCAL VOC	86.3
SDN+ (Fu et al. 2019b)	Architecture	PASCAL VOC, CamVid, Gatech	86.6
DIS (Luo et al. 2017)	Architecture	PASCAL VOC	86.8
DeepLabV3 (Chen et al. 2017b)	Architecture	PASCAL VOC	86.9
MSCI (Lin et al. 2018)	Architecture	PASCAL VOC, PASCAL Context, NYUDv2, SUN RGB-D	88.0
DeepLabV3+ (Chen et al. 2018b)	Architecture	PASCAL VOC, Cityscapes	89.0



(a) The various categories of contributions.

(b) The various medical imaging modalities.

(c) The distribution of dataset sizes across multiple imaging modalities.

Fig. 15 Analyzing the attributes of the medical image segmentation papers discussed in this review. The large number of medical imaging modalities (b) as well as the smaller average dataset sizes for medical image segmentation datasets (c) as compared to natural images (as discussed in Sect. 9.2) make it difficult to benchmark the performance of various approaches. In (b), PET (1.1%), OCT (0.6%), and topogram (0.6%) make up the 'Other' label

10.1 Architectures

Encoder-decoder networks with long and short skip connections are the winning architectures according to the state-of-the-art methods. Skip connections in deep networks have improved both segmentation and classification performance by facilitating the training of deeper network architectures and reducing the risks for vanishing gradients. They equip encoder-decoder-like networks with richer feature representations, but at the cost of higher memory usage, computation, and possibly resulting in transferring non-discriminative feature maps. Similar to Taghanaki et al. (2019c), one future work direction is to optimize the amount of data is being transferred through skip connections. As for the cell level architectural design, our study shows that atrous convolutions with feature pyramid pooling modules are highly being used in the recent models. These approaches are somehow modifications of the classical convolution blocks. Similar to the radial basis function layers in Meyer et al. (2018) and Taghanaki et al. (2019a), a future work focus can be designing new layers that capture a new aspect of data as opposed to convolutions or transform the

Table 3 A summary of medical image segmentation papers along with their type of proposed improvement

Paper	Type of improvement	Training data [Modality (Site, Count[a])]
Ronneberger et al. (2015)	Architecture optimization	EM (Drosophilia, 30) Microscopy (Cells, 35) Microscopy (HeLa cells, 20)
Milletari et al. (2016)	Architecture optimization	MRI (Prostate, 50)
BenTaieb and Hamarneh (2016)	Optimization	Histology (Colon, 70)
Yang et al. (2017b)	Sequenced models	US (Prostate, 17)
Salehi et al. (2017)	Optimization	MRI (Multiple sclerosis lesions. 15)
Chartsias et al. (2017)	Synthesis-based	CT (Heart, 20); MRI (Heart, 20)
Chaichulee et al. (2017)	Multi-task models	RGB (Skin, 4603)
Drozdzal et al. (2018)	Architecture	EM (Drosophilia, 30) CT (Liver, 77); MRI (Prostate, 50)
Zhou et al. (2018)	Architecture	Microscopy (Cell nuclei, 670) RGB (Colon polyp, 7379) CT (Liver, 331) CT (Lung nodule, 1012)
Oktay et al. (2018)	Architecture	CT (Abdominal, 150) CT (Pancreas, 82)
Xue et al. (2018)	Architecture	MRI (Brain, 246)
Zhang et al. (2018c)	Synthesis-based	CT (Heart, 4354); MRI (Heart, 142)
Shin et al. (2018)	Synthesis-based	MRI (Brain, 211)
Johnson (2018)	Multi-task models	Microscopy (Cell nuclei, 664)
Jaeger et al. (2018)	Multi-task models	CT (Lung nodule, 1035) MRI (Breast, 331)
Mehta et al. (2018)	Multi-task models	Histology (Breast, 428)
Huo et al. (2018)	Synthesis-based	MRI (Spleen, 60); CT (Spleen, 19)
Alom et al. (2019)	Sequenced models	fundus (Retinal vessel, 20) Fundus (Retinal vessel, 20) Fundus (Retinal vessel, 28) RGB (Skin, 1250); X-ray (Lung, 373)
Kervadec et al. (2019b)	Weakly Supervised optimization	MRI (Heart, 75) MRI (Vertebral body, 15) MRI (Prostate, 40)
Perone et al. (2019)	Weakly supervised	MRI (Spinal cord, 40)

[a] Indicates the count at the highest level. For example, if a paper reports counts of patients, volumes, slices, etc., we report the count of patients

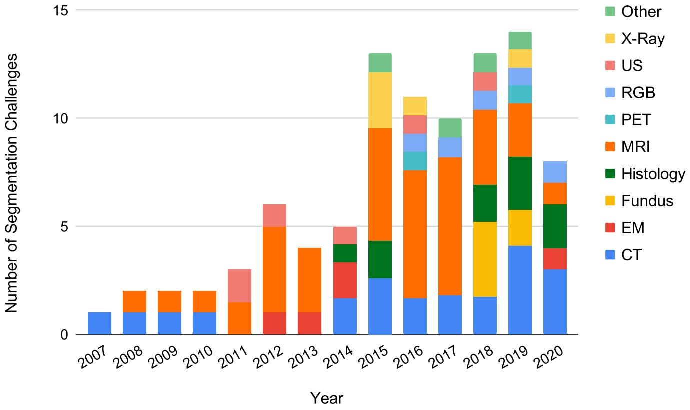


Fig. 16 The number of medical image segmentation challenges every year since 2007 listed on Grand Challenges (Challenge 2020), along with a imaging modality-wise breakdown. Note that for many challenges, the data is multi-modal, and therefore the breakdown takes that into account

convolution features into a new manifold. Another useful research direction is using neural architecture search (Zoph and Le 2016) to search for optimal deep neural network architectures for segmentation (Liu et al. 2019a; Zhu et al. 2019; Shaw et al. 2019; Weng et al. 2019b).

10.2 Sequenced models

For image segmentation, sequenced models can be used to segment temporal data such as videos. These models have also been applied to 3D medical datasets, however the advantage of processing volumetric data using 3D convolutions versus the processing the volume slice by slice using 2D sequenced models. Ideally, seeing the whole object of interest in a 3D volume might help to capture the geometrical information of the object, which might be missed in processing a 3D volume slice by slice. Therefore a future direction in this area can be through analysis of sequenced models versus volumetric convolution-based approaches.

10.3 Optimization functions

In medical image segmentation works, researchers have converged toward using classical cross-entropy loss functions along with a second distance or overlap based functions. Incorporating domain/prior knowledge (such as coding the location of different organs explicitly in a deep model) is more sensible in the medical datasets. As shown in Taghanaki et al. (2019e), when only a distance-based or overlap-based loss function is used in a network, and the final layer applies sigmoid function, the risk of gradient vanishing increases. Although overlap based loss function are used in case of a class imbalance (small foregrounds), in Fig. 13, we show how using (*only*) overlap based loss functions as the main term can be problematic for smooth optimization where they highly penalize a model under/over-segmenting a small foreground. However, the cross-entropy loss returns a reasonable score for the same cases. Besides using integrated cross-entropy based loss functions, future work can be exploring a single loss function that follows the behavior of the cross-entropy and at the same time, offers more features such capturing contour distance. This can be achieved by revisiting the current distance and overlap based loss functions. Another future path can be exploring auto loss function (or regularization term) search similar to the neural architecture search mentioned above. Similarly, gradient based optimizations based on Sobolev (Adams and Fournier 2003) gradients (Czarnecki et al. 2017), such as the works of Goceri (2019b, 2020) are an interesting research direction.

10.4 Other potential directions

- Going beyond pixel intensity-based scene understanding by incorporating prior knowledge, which have been an active area of research for the past several decades (Nosrati and Hamarneh 2016; Xie et al. 2020). Encoding prior knowledge in medical image analysis models is generally more possible as compared to natural images. Currently, deep models receive matrices of intensity values, and usually, they are not forced to learn prior information. Without explicit reinforcement, the models might still learn object relations to some extent. However, it is difficult to interpret a learned strategy.
- Because of the large number of imaging modalities, the significant signal noise present in imaging modalities such as PET and ultrasound, and the limited amount of medical imaging data mainly because of high acquisition cost compounded by legal, ethical, and privacy issues, it is difficult to develop universal solutions that yield acceptable performances across various imaging modalities. Therefore, a proper research direction would be along the work of Raghu et al. (2019) on image classification models, studying the risks of using non-medical pre-trained models for medical image segmentation.
- Creating large 2D and 3D publicly available medical benchmark datasets for semantic image segmentation such as the Medical Segmentation Decathlon (Simpson et al. 2019). Medical imaging datasets are typically much smaller in size than natural image datasets (Jin et al. 2020), and the curation of larger public datasets for medical imaging modalities will allow researchers to accurately compare proposed approaches and make incremental improvements for specific datasets and problems.
- A possible solution to address the paucity of sufficient annotated medical data is the development and use of physics based imaging simulators, the outputs of which can be used to train segmentation models and augment existing segmentation datasets. Several platforms (Marion et al. 2011; Glatard et al. 2013) as well as simulators already exist for various imaging modalities such as SIMRI (Benoit-Cattin et al. 2005) and POSSUM (Drobnjak et al. 2006, 2010) for magnetic resonance imaging (MRI), PET-SORTEO (Reilhac et al. 2005) and SimSET (Harrison and Lewellen 2012) for emission tomography, SINDBAD (Tabary et al. 2007) for computed tomography (CT), and FIELD-II (Jensen and Svendsen 1992; Jensen 1996) and SIMUS (Shahriari and Garcia 2018) for ultrasound imaging as well as anatomical regions of interest such as VascuSynth (Hamarneh and Jassi 2010) for vascular trees.
- Medical images, both 2D and volumetric, have in general, larger file sizes than natural images, which inhibits the ability to load them entirely onto the memory for processing. As such, they need to be processed either as patches or sub-volumes, making it difficult for the segmentation models to capture spatial relationships in order to perform accurate segmentation. Therefore, an interesting and potentially very useful research direction would be coming up with architectures and training methods that can incorporate spatial relationships from large medical images and volumes in the models.

- Exploring reinforcement learning approaches similar to Song et al. (2018) and Wang et al. (2018c) for semantic (medical) image segmentation to mimic the way humans delineate objects of interest. Deep CNNs are successful in extracting features of different classes of objects, but they lose the local spatial information of where the borders of an object should be. Some researchers resort to traditional computer vision methods such as conditional random fields (CRFs) to overcome this problem, which however, add more computation time to the models.
- Studying the causes for some models and datasets being prone to false positive and false negative predictions in the image segmentation context as found by Berman et al. (2018b) and Taghanaki et al. (2019e).
- Exploring segmentation-free approaches (Zhen and Li 2015; Hussain et al. 2017; Taghanaki et al. 2018; Mukherjee et al. 2019; Proenca and Neves 2019), i.e., bypassing the segmentation step according to the target problem.
- Weakly supervised segmentation using image-level labels versus a few images with segmentation annotations. Most new weakly supervised localization methods apply attention maps or region proposals in a multiple instance learning formulations. While attention maps can be noisy, leading to erroneously highlighted regions, it is not simple to decide on an optimal window or bag size for multiple instance learning approaches.
- While most deep segmentation models for medical image analysis rely on only clinical images for their predictions, there is often multi-modal patient data in the form of other imaging modalities as well as patient metadata that can provide valuable information, which most deep segmentation models do not use. Therefore, a valuable research direction for improving segmentation performance of medical images would be to develop models which are able to leverage multi-modal patient data.
- Modifying input instead of the model, loss function, and adding more train data. Drozdzal et al. (2018) showed that attaching a pre-processing module at the beginning of a segmentation network improves the network performance. Taghanaki et al. (2019b) leveraged the gradients of a trained segmentation network with respect to the input to transfer it to a new space where the segmentation accuracy improves.
- Deep neural networks are trained using error backpropagation (Rumelhart et al. 1986) and gradient descent for optimizing the network weights. However, there have been many neural network optimization techniques which do not rely on backpropagation, such as credit assignment (Bengio and Frasconi 1994), neuroevolution (Stanley and Miikkulainen 2002), difference target propagation (Lee et al. 2015), training with local error signals (Nøkland and Eidnes 2019) and several other techniques (Amit 2019; Bellec et al. 2019; Ma et al. 2019). Exploring these and similar other techniques to optimize deep neural networks for semantic segmentation would be another valuable research direction.

References

Abdulla W (2017) Mask R-CNN for object detection and instance segmentation on Keras and TensorFlow. https://github.com/matterport/Mask_RCNN

Abhishek K, Hamarneh G (2019) Mask2Lesion: mask-constrained adversarial skin lesion image synthesis. In: Medical image computing and computer-assisted intervention workshop on simulation and synthesis in medical imaging, pp 71–80

Abhishek K, Hamarneh G, Drew MS (2020) Illumination-based transformations improve skin lesion segmentation in dermoscopic images. arXiv:200310111

Adams RA, Fournier JJ (2003) Sobolev spaces. Elsevier, Amsterdam

Afshari S, BenTaieb A, Mirikharaji Z, Hamarneh G (2019) Weakly supervised fully convolutional network for PET lesion segmentation. In: Medical imaging 2019: image processing, international society for optics and photonics, vol 10949, p 109491K

Alom MZ, Yakopcic C, Hasan M, Taha TM, Asari VK (2019) Recurrent residual U-Net for medical image segmentation. J Med Imag 6(1):14006

Amirul Islam M, Rochan M, Bruce ND, Wang Y (2017) Gated feedback refinement network for dense image labeling. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 3751–3759

Amit Y (2019) Deep learning with asymmetric connections and hebbian updates. Front Comput Neurosci. https://doi.org/10.3389/fncom.2019.00018

Anantharaman R, Velazquez M, Lee Y (2018) Utilizing Mask R-CNN for detection and segmentation of oral diseases. In: 2018 IEEE international conference on bioinformatics and biomedicine, pp 2197–2204

Badrinarayanan V, Handa A, Cipolla R (2015) Segnet: a deep convolutional encoder-decoder architecture for image segmentation. arXiv:151100561

Bai W, Suzuki H, Qin C, Tarroni G, Oktay O, Matthews PM, Rueckert D (2018) Recurrent neural networks for aortic image sequence segmentation with sparse annotations. In: International conference on medical image computing and computer-assisted intervention. Springer, pp 586–594

Bellec G, Scherr F, Hajek E, Salaj D, Legenstein R, Maass W (2019) Biologically inspired alternatives to backpropagation through time for learning in recurrent neural nets. arXiv:190109049

Bengio Y, Frasconi P (1994) Credit assignment through time: alternatives to backpropagation. In: Advances in neural information processing systems, pp 75–82

Benoit-Cattin H, Collewet G, Belaroussi B, Saint-Jalmes H, Odet C (2005) The SIMRI project: a versatile and interactive MRI simulator. J Magn Reson 173(1):97–115. https://doi.org/10.1016/j.jmr.2004.09.027

BenTaieb A, Hamarneh G (2016) Topology aware fully convolutional networks for histology gland segmentation. In: International conference on medical image computing and computer assisted intervention. Springer, pp 460–468

Berman M, Blaschko MB, Triki AR, Yu J (2018a) Yes, IoU loss is submodular-as a function of the mispredictions. arXiv:180901845

Berman M, Rannen Triki A, Blaschko MB (2018b) The Lovász-Softmax loss: a tractable surrogate for the optimization of the intersection-over-union measure in neural networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 4413–4421

Bischke B, Helber P, Folz J, Borth D, Dengel A (2019) Multi-task learning for segmentation of building footprints with deep neural networks. In: 2019 IEEE international conference on image processing. IEEE, pp 1480–1484

Bonta LR, Kiran NU (2019) Efficient segmentation of medical images using dilated residual networks. In: Computer aided intervention and diagnostics in clinical and medical images. Springer, pp 39–47

Borji A, Cheng MM, Hou Q, Jiang H, Li J (2019) Salient object detection: a survey. Comput Vis Media 5(2):117–150. https://doi.org/10.1007/s41095-019-0149-9

Brostow GJ, Shotton J, Fauqueur J, Cipolla R (2008) Segmentation and recognition using structure from motion point clouds. In: Lecture notes in computer science. Springer, Berlin, pp 44–57. https://doi.org/10.1007/978-3-540-88682-2_5

Brostow GJ, Fauqueur J, Cipolla R (2009) Semantic object classes in video: a high-definition ground truth database. Pattern Recognit Lett 30(2):88–97. https://doi.org/10.1016/j.patrec.2008.04.005

Brügger R, Baumgartner CF, Konukoglu E (2019) A partially reversible U-Net for memory-efficient volumetric image segmentation. arXiv:190606148

Caliva F, Iriondo C, Martinez AM, Majumdar S, Pedoia V (2019) Distance map loss penalty term for semantic segmentation. In: International conference on medical imaging with deep learning

Caruana R (1997) Multitask learning. Mach Learn 28(1):41–75

Chaichulee S, Villarroel M, Jorge J, Arteta C, Green G, McCormick K, Zisserman A, Tarassenko L (2017) Multi-task convolutional neural network for patient detection and skin segmentation in continuous non-contact vital sign monitoring. In: 2017 12th IEEE international conference on automatic face & gesture recognition. IEEE, pp 266–272

Chakravarty A, Sivaswamy J (2018) RACE-Net: a recurrent neural network for biomedical image segmentation. IEEE J Biomed Health Inform 23(3):1151–1162

Challenge G (2020) Grand challenges in biomedical image analysis. https://grand-challenge.org/challenges/

Chartsias A, Joyce T, Dharmakumar R, Tsaftaris SA (2017) Adversarial image synthesis for unpaired multimodal cardiac data. In: International workshop on simulation and synthesis in medical imaging. Springer, pp 3–13

Chen LC, Yang Y, Wang J, Xu W, Yuille AL (2016) Attention to scale: scale-aware semantic image segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 3640–3649

Chen LC, Papandreou G, Kokkinos I, Murphy K, Yuille AL (2017a) Deeplab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE Trans Pattern Anal Mach Intell 40(4):834–848

Chen LC, Papandreou G, Schroff F, Adam H (2017b) Rethinking atrous convolution for semantic image segmentation. arXiv:170605587

Chen LC, Collins M, Zhu Y, Papandreou G, Zoph B, Schroff F, Adam H, Shlens J (2018a) Searching for efficient multi-scale architectures for dense image prediction. In: Advances in neural information processing systems, pp 8699–8710

Chen LC, Zhu Y, Papandreou G, Schroff F, Adam H (2018b) Encoder-decoder with atrous separable convolution for semantic image segmentation. In: Proceedings of the European conference on computer vision, pp 801–818

Chen X, Williams BM, Vallabhaneni SR, Czanner G, Williams R, Zheng Y (2019) Learning active contour models for medical image segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 11632–11640

Cherian A, Sullivan A (2019) Sem-GAN: semantically-consistent image-to-image translation. In: 2019 IEEE winter conference on applications of computer vision (WACV). IEEE. https://doi.org/10.1109/wacv.2019.00196

Choi J, Kim T, Kim C (2019) Self-ensembling with gan-based data augmentation for domain adaptation in semantic segmentation. In: Proceedings of the IEEE international conference on computer vision, pp 6830–6840

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

Cireşan D, Meier U, Schmidhuber J (2012) Multi-column deep neural networks for image classification. In: 2012 IEEE conference on computer vision and pattern recognition. IEEE, pp 3642–3649

Cireşan DC, Meier U, Masci J, Gambardella LM, Schmidhuber J (2011) High-performance neural networks for visual object classification. arXiv:11020183

Cohen JP, Luck M, Honari S (2018) Distribution matching losses can hallucinate features in medical image translation. In: Medical image computing and computer assisted intervention – MICCAI 2018. Springer, pp 529–536. https://doi.org/10.1007/978-3-030-00928-1_60

Cordts M, Omran M, Ramos S, Rehfeld T, Enzweiler M, Benenson R, Franke U, Roth S, Schiele B (2016) The cityscapes dataset for semantic urban scene understanding. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 3213–3223

Costa P, Galdran A, Meyer MI, Abràmoff MD, Niemeijer M, Mendonça AM, Campilho A (2017) Towards adversarial retinal image synthesis. arXiv:170108974

Couprie C, Farabet C, Najman L, LeCun Y (2013) Indoor semantic segmentation using depth information. arXiv:13013572

Czarnecki WM, Osindero S, Jaderberg M, Swirszcz G, Pascanu R (2017) Sobolev training for neural networks. In: Advances in neural information processing systems, pp 4278–4287

Dai W, Dong N, Wang Z, Liang X, Zhang H, Xing EP (2018) SCAN: structure correcting adversarial network for organ segmentation in chest X-rays. In: Deep learning in medical image analysis and multimodal learning for clinical decision support. Springer, pp 263–273. https://doi.org/10.1007/978-3-030-00889-5_30

Drobnjak I, Gavaghan D, Süli E, Pitt-Francis J, Jenkinson M (2006) Development of a functional magnetic resonance imaging simulator for modeling realistic rigid-body motion artifacts. Magn Reson Med 56(2):364–380. https://doi.org/10.1002/mrm.20939

Drobnjak I, Pell GS, Jenkinson M (2010) Simulating the effects of time-varying magnetic fields with a realistic simulated scanner. Magn Reson Imaging 28(7):1014–1021. https://doi.org/10.1016/j.mri.2010.03.029

Drozdzal M, Chartrand G, Vorontsov E, Shakeri M, Di Jorio L, Tang A, Romero A, Bengio Y, Pal C, Kadoury S (2018) Learning normalized inputs for iterative estimation in medical image segmentation. Med Image Anal 44:1–13

Everingham M, Gool LV, Williams CKI, Winn J, Zisserman A (2010) The pascal visual object classes (VOC) challenge. Int J Comput Vis 88(2):303–338. https://doi.org/10.1007/s11263-009-0275-4

Everingham M, Van Gool L, Williams CKI, Winn J, Zisserman A (2012) The PASCAL visual object classes challenge 2012 (VOC2012) results. http://www.pascal-network.org/challenges/VOC/voc2012/workshop/index.html

Everingham M, Eslami SA, Van Gool L, Williams CK, Winn J, Zisserman A (2015) The PASCAL visual object classes challenge: a retrospective. Int J Comput Vis 111(1):98–136

Fu J, Liu J, Tian H, Li Y, Bao Y, Fang Z, Lu H (2019a) Dual attention network for scene segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 3146–3154

Fu J, Liu J, Wang Y, Zhou J, Wang C, Lu H (2019b) Stacked deconvolutional network for semantic segmentation. IEEE Trans Image Process

Galdran A, Alvarez-Gila A, Meyer MI, Saratxaga CL, Araújo T, Garrote E, Aresta G, Costa P, Mendonça AM, Campilho A (2017) Data-driven color augmentation techniques for deep skin image analysis. arXiv:170303702

Gamage H, Wijesinghe W, Perera I (2019) Instance-based segmentation for boundary detection of neuropathic ulcers through Mask-RCNN. In: International conference on artificial neural networks. Springer, pp 511–522

Gao Y, Phillips JM, Zheng Y, Min R, Fletcher PT, Gerig G (2018) Fully convolutional structured LSTM networks for joint 4D medical image segmentation. In: 2018 IEEE 15th international symposium on biomedical imaging. IEEE, pp 1104–1108

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

Glatard T, Lartizien C, Gibaud B, da Silva RF, Forestier G, Cervenansky F, Alessandrini M, Benoit-Cattin H, Bernard O, Camarasu-Pop S, Cerezo N, Clarysse P, Gaignard A, Hugonnard P, Liebgott H, Marache S, Marion A, Montagnat J, Tabary J, Friboulet D (2013) A virtual imaging platform for multi-modality medical image simulation. IEEE Trans Med Imaging 32(1):110–118. https://doi.org/10.1109/tmi.2012.2220154

Goceri E (2019a) Challenges and recent solutions for image segmentation in the era of deep learning. In: 2019 ninth international conference on image processing theory, tools and applications (IPTA). IEEE. https://doi.org/10.1109/ipta.2019.8936087

Goceri E (2019b) Diagnosis of alzheimerś disease with sobolev gradient-based optimization and 3d convolutional neural network. Int J Numer Methods Biomed Eng. https://doi.org/10.1002/cnm.3225

Goceri E (2020) CapsNet topology to classify tumours from brain images and comparative evaluation. IET Image Process 14(5):882–889. https://doi.org/10.1049/iet-ipr.2019.0312

Goceri E, Goceri N (2017) Deep learning in medical image analysis: recent advances and future trends. In: Proceedings of the IADIS international conference computer graphics, visualization, computer vision and image processing (CGVCVIP) 2017, pp 305–310

Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y (2014) Generative adversarial nets. In: Advances in neural information processing systems, pp 2672–2680

Gros C, De Leener B, Badji A, Maranzano J, Eden D, Dupont SM, Talbott J, Zhuoquiong R, Liu Y, Granberg T et al (2019) Automatic segmentation of the spinal cord and intramedullary multiple sclerosis lesions with convolutional neural networks. Neuroimage 184:901–915

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 38:2281–2292

Gulrajani I, Ahmed F, Arjovsky M, Dumoulin V, Courville AC (2017) Improved training of wasserstein gans. In: Advances in neural information processing systems, pp 5767–5777

Guo Y, Liu Y, Georgiou T, Lew MS (2018) A review of semantic segmentation using deep neural networks. Int J Multimed Inf Retr 7(2):87–93

Hamarneh G, Jassi P (2010) VascuSynth: simulating vascular trees for generating volumetric image data with ground-truth segmentation and tree analysis. Comput Med Imaging Graphics 34(8):605–616. https://doi.org/10.1016/j.compmedimag.2010.06.002

Han C, Murao K, Satoh S, Nakayama H (2019) Learning more with less: GAN-based medical image augmentation. Med Imaging Technol 37(3):137–142

Han S, Liu X, Mao H, Pu J, Pedram A, Horowitz MA, Dally WJ (2016) EIE: efficient inference engine on compressed deep neural network. In: 2016 ACM/IEEE 43rd annual international symposium on computer architecture. IEEE, pp 243–254

Haralick RM, Shapiro LG (1992) Computer and robot vision. Addison-Wesley, Boston

Harrison R, Lewellen T (2012) The SimSET program. In: Monte Carlo calculations in nuclear medicine, Second Edition. Taylor & Francis, pp 87–110. https://doi.org/10.1201/b13073-7

He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 770–778

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

He T, Guo J, Wang J, Xu X, Yi Z (2019) Multi-task learning for the segmentation of thoracic organs at risk in CT images. In: SegTHOR@ISBI

Hesamian MH, Jia W, He X, Kennedy P (2019) Deep learning techniques for medical image segmentation: achievements and challenges. J Digit Imaging 32:582–596

Hochreiter S, Schmidhuber J (1997) Long short-term memory. Neural Comput 9(8):1735–1780

Honari S, Yosinski J, Vincent P, Pal C (2016) Recombinator networks: learning coarse-to-fine feature aggregation. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 5743–5752

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:170404861

Hu J, Shen L, Sun G (2018a) Squeeze-and-excitation networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 7132–7141

Hu Y, Chen Z, Lin W (2018b) RGB-D semantic segmentation: a review. In: 2018 IEEE international conference on multimedia & expo workshops. IEEE, pp 1–6

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

Huang SW, Lin CT, Chen SP, Wu YY, Hsu PH, Lai SH (2018) AugGAN: Cross domain adaptation with GAN-based data augmentation. In: Proceedings of the European conference on computer vision (ECCV). Springer, Berlin, pp 731–744. https://doi.org/10.1007/978-3-030-01240-3_44

Huo Y, Xu Z, Bao S, Assad A, Abramson RG, Landman BA (2018) Adversarial synthesis learning enables segmentation without target modality ground truth. In: 2018 IEEE 15th international symposium on biomedical imaging. IEEE, pp 1217–1220

Hussain MA, Amir-Khalili A, Hamarneh G, Abugharbieh R (2017) Segmentation-free kidney localization and volume estimation using aggregated orthogonal decision CNNs. In: International conference on medical image computing and computer-assisted intervention. Springer, pp 612–620

Isensee F, Petersen J, Klein A, Zimmerer D, Jaeger PF, Kohl S, Wasserthal J, Koehler G, Norajitra T, Wirkert S, et al. (2019) nnU-Net: self-adapting framework for U-Net-based medical image segmentation. In: Bildverarbeitung für die Medizin 2019. Springer, pp 22–22

Jaeger PF, Kohl SA, Bickelhaupt S, Isensee F, Kuder TA, Schlemmer HP, Maier-Hein KH (2018) Retina U-Net: embarrassingly simple exploitation of segmentation supervision for medical object detection. arXiv:181108661

Jégou S, Drozdzal M, Vazquez D, Romero A, Bengio Y (2017) The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops, pp 11–19

Jensen J, Svendsen N (1992) Calculation of pressure fields from arbitrarily shaped, apodized, and excited ultrasound transducers. IEEE Trans Ultrason Ferroelectr Freq Control 39(2):262–267. https://doi.org/10.1109/58.139123

Jensen JA (1996) Field: A program for simulating ultrasound systems. In: 10th Nordic-Baltic conference on biomedical imaging, Volume 34, Supplement 1, Part 1, pp 351–353

Jin D, Xu Z, Tang Y, Harrison AP, Mollura DJ (2018) CT-realistic lung nodule simulation from 3D conditional generative adversarial networks for robust lung segmentation. In: International conference on medical image computing and computer-assisted intervention. Springer, pp 732–740

Jin W, Fatehi M, Abhishek K, Mallya M, Toyota B, Hamarneh G (2020) Artificial intelligence in glioma imaging: challenges and advances. J Neural Eng 17(2):021002. https://doi.org/10.1088/1741-2552/ab8131

Johnson JW (2018) Adapting mask R-CNN for automatic nucleus segmentation. arXiv:180500500

Karimi D, Salcudean SE (2019) Reducing the Hausdorff distance in medical image segmentation with convolutional neural networks. arXiv:190410030

Karimi D, Dou H, Warfield SK, Gholipour A (2019) Deep learning with noisy labels: exploring techniques and remedies in medical image analysis. arXiv:191202911

Ke R, Bugeau A, Papadakis N, Schütz P, Schönlieb CB (2019) A multi-task U-Net for segmentation with lazy labels. arXiv:1906.12177

Kervadec H, Bouchtiba J, Desrosiers C, Granger E, Dolz J, Ben Ayed I (2019a) Boundary loss for highly unbalanced segmentation. In: Proceedings of the 2nd international conference on medical imaging with deep learning, PMLR, London, United Kingdom, proceedings of machine learning research, vol 102, pp 285–296. http://proceedings.mlr.press/v102/kervadec19a.html

Kervadec H, Dolz J, Tang M, Granger E, Boykov Y, Ayed IB (2019) Constrained-CNN losses for weakly supervised segmentation. Med Image Anal 54:88–99. https://doi.org/10.1016/j.media.2019.02.009

Khosravan N, Mortazi A, Wallace M, Bagci U (2019) PAN: projective adversarial network for medical image segmentation. arXiv:190604378

Kim B, Ye JC (2019) Multiphase level-set loss for semi-supervised and unsupervised segmentation with deep learning. arXiv:190402872

Kim HE, Hwang S (2016) Deconvolutional feature stacking for weakly-supervised semantic segmentation. arXiv:160204984

Kim YD, Park E, Yoo S, Choi T, Yang L, Shin D (2015) Compression of deep convolutional neural networks for fast and low power mobile applications. arXiv:151106530

Kopelowitz E, Engelhard G (2019) Lung nodules detection and segmentation using 3D Mask R-CNN. arXiv:1907.07676

Krizhevsky A, Sutskever I, Hinton GE (2012) ImageNet classification with deep convolutional neural networks. In: Advances in neural information processing systems, pp 1097–1105

Kuntimad G, Ranganath H (1999) Perfect image segmentation using pulse coupled neural networks. IEEE Trans Neural Netw 10(3):591–598. https://doi.org/10.1109/72.761716

Lateef F, Ruichek Y (2019) Survey on semantic segmentation using deep learning techniques. Neurocomputing 338:321–348

Le TLT, Thome N, Bernard S, Bismuth V, Patoureaux F (2019) Multitask classification and segmentation for cancer diagnosis in mammography. arXiv:190905397

LeCun Y, Bottou L, Bengio Y, Haffner P (1998) Gradient-based learning applied to document recognition. Proc IEEE 86(11):2278–2324. https://doi.org/10.1109/5.726791

Lee DH, Zhang S, Fischer A, Bengio Y (2015) Difference target propagation. In: Machine learning and knowledge discovery in databases. Springer, pp 498–515. https://doi.org/10.1007/978-3-319-23528-8_31

Lee J, Kim E, Lee S, Lee J, Yoon S (2019) Ficklenet: weakly and semi-supervised semantic image segmentation using stochastic inference. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 5267–5276

Leroux S, Molchanov P, Simoens P, Dhoedt B, Breuel T, Kautz J (2018) IamNN: iterative and adaptive mobile neural network for efficient image classification. arXiv:180410123

Li H, Xiong P, An J, Wang L (2018) Pyramid attention network for semantic segmentation. arXiv:180510180

Li H, Li J, Lin X, Qian X (2019a) Pancreas segmentation via spatial context based U-Net and bidirectional LSTM. arXiv:190300832

Li S, Dong M, Du G, Mu X (2019b) Attention dense-U-Net for automatic breast mass segmentation in digital mammogram. IEEE Access 7:59037–59047

Li X, Liu Z, Luo P, Change Loy C, Tang X (2017) Not all pixels are equal: difficulty-aware semantic segmentation via deep layer cascade. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 3193–3202

Li X, Yu L, Chen H, Fu CW, Heng PA (2019c) Transformation consistent self-ensembling model for semi-supervised medical image segmentation. arXiv:190300348

Lian S, Luo Z, Zhong Z, Lin X, Su S, Li S (2018) Attention guided U-Net for accurate iris segmentation. J Vis Commun Image Represent 56:296–304

Lin D, Ji Y, Lischinski D, Cohen-Or D, Huang H (2018) Multi-scale context intertwining for semantic segmentation. In: Proceedings of the European conference on computer vision, pp 603–619

Lin G, Milan A, Shen C, Reid I (2017a) RefineNet: Multi-path refinement networks for high-resolution semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 1925–1934

Lin TY, Goyal P, Girshick R, He K, Dollár P (2017b) Focal loss for dense object detection. arXiv:170802002

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 Anal 42:60–88

Liu C, Chen LC, Schroff F, Adam H, Hua W, Yuille AL, Fei-Fei L (2019a) Auto-deeplab: hierarchical neural architecture search for semantic image segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 82–92

Liu L, Ouyang W, Wang X, Fieguth P, Chen J, Liu X, Pietikäinen M (2019b) Deep learning for generic object detection: a survey. Int J Comput Vis 128(2):261–318. https://doi.org/10.1007/s11263-019-01247-4

Liu Y, Perona P, Meister M (2019c) Panda: panoptic data augmentation. arXiv:191112317

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

Luc P, Couprie C, Chintala S, Verbeek J (2016) Semantic segmentation using adversarial networks. arXiv:161108408

Luo P, Wang G, Lin L, Wang X (2017) Deep dual learning for semantic image segmentation. In: Proceedings of the IEEE international conference on computer vision, pp 2718–2726

Ma WDK, Lewis J, Kleijn WB (2019) The hsic bottleneck: Deep learning without back-propagation. arXiv:190801580

Marion A, Forestier G, Benoit-Cattin H, Camarasu-Pop S, Clarysse P, da SilvaRF, Gibaud B, Glatard T, Hugonnard P, Lartizien C, Liebgott H, Specovius S,Tabary J, Valette S, Friboulet D (2011) Multimodality medical image simulation of biological models with the virtual imaging platform (VIP). In: 2011 24th international symposium on computer-based medical systems(CBMS). IEEE. https://doi.org/10.1109/cbms.2011.5999141

Mehta S, Mercan E, Bartlett J, Weaver D, Elmore JG, Shapiro L (2018) Y-Net: Joint segmentation and classification for diagnosis of breast biopsy images. In: International conference on medical image computing and computer-assisted intervention. Springer, pp 893–901

Meyer BJ, Harwood B, Drummond T (2018) Deep metric learning and image classification with nearest neighbour Gaussian kernels. In: 2018 25th IEEE international conference on image processing. IEEE, pp 151–155

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. IEEE, pp 565–571

Mirikharaji Z, Hamarneh G (2018) Star shape prior in fully convolutional networks for skin lesion segmentation. In: International conference on medical image computing and computer assisted intervention. Springer, pp 737–745

Mirikharaji Z, Yan Y, Hamarneh G (2019) Learning to segment skin lesions from noisy annotations. In: International workshop on medical image learning with less labels and imperfect data

Moeskops P, Veta M, Lafarge MW, Eppenhof KA, Pluim JP (2017) Adversarial training and dilated convolutions for brain MRI segmentation. In: Deep learning in medical image analysis and multimodal learning for clinical decision support. Springer, pp 56–64

Mohajerani S, Asad R, Abhishek K, Sharma N, van Duynhoven A, Saeedi P (2019) Cloudmaskgan: a content-aware unpaired image-to-image translation algorithm for remote sensing imagery. In: 2019 IEEE international conference on image processing. IEEE, pp 1965–1969

Mohanty SP (2018) Crowdai mapping challenge 2018: baseline with mask RCNN. https://github.com/crowdai/crowdai-mapping-challenge-mask-rcnn

Mottaghi R, Chen X, Liu X, Cho NG, Lee SW, Fidler S, Urtasun R, Yuille A (2014) The role of context for object detection and semantic segmentation in the wild. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 891–898

Mukherjee S, Cheng I, Miller S, Guo T, Chau V, Basu A (2019) A fast segmentation-free fully automated approach to white matter injury detection in preterm infants. Med Biol Eng Comput 57(1):71–87

Mumford D, Shah J (1989) Optimal approximations by piecewise smooth functions and associated variational problems. Commun Pure Appl Math 42(5):577–685

Neff T, Payer C, Stern D, Urschler M (2017) Generative adversarial network based synthesis for supervised medical image segmentation. In: Proceedings of OAGM and ARW joint workshop

Neff T, Payer C, Štern D, Urschler M (2018) Generative adversarial networks to synthetically augment data for deep learning based image segmentation. In: Proceedings of the OAGM workshop 2018: medical image analysis. Verlag der Technischen Universität Graz, pp 22–29

Nguyen HH, Fang F, Yamagishi J, Echizen I (2019) Multi-task learning for detecting and segmenting manipulated facial images and videos. arXiv:1906.06876

Ni ZL, Bian GB, Xie XL, Hou ZG, Zhou XH, Zhou YJ (2019) RASNet: segmentation for tracking surgical instruments in surgical videos using refined attention segmentation network. arXiv:190508663

Nie D, Gao Y, Wang L, Shen D (2018) ASDNet: Attention based semi-supervised deep networks for medical image segmentation. In: International conference on medical image computing and computer-assisted intervention. Springer, pp 370–378

Noh H, Hong S, Han B (2015) Learning deconvolution network for semantic segmentation. In: Proceedings of the IEEE international conference on computer vision, pp 1520–1528

Nøkland A, Eidnes LH (2019) Training neural networks with local error signals. In: Chaudhuri K, Salakhutdinov R (eds) Proceedings of the 36th international conference on machine learning, PMLR, Long Beach, California, USA, proceedings of machine learning research, vol 97, pp 4839–4850. http://proceedings.mlr.press/v97/nokland19a.html

Nosrati MS, Hamarneh G (2016) Incorporating prior knowledge in medical image segmentation: a survey. arXiv:160701092

Nowozin S (2014) Optimal decisions from probabilistic models: the intersection-over-union case. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 548–555

Oktay O, Schlemper J, Folgoc LL, Lee M, Heinrich M, Misawa K, Mori K, McDonagh S, Hammerla NY, Kainz B, et al. (2018) Attention U-Net: learning where to look for the pancreas. arXiv:180403999

Paschali M, Gasperini S, Roy AG, Fang MYS, Navab N (2019) 3DQ: compact quantized neural networks for volumetric whole brain segmentation. arXiv:190403110

Peng C, Zhang X, Yu G, Luo G, Sun J (2017) Large kernel matters–improve semantic segmentation by global convolutional network. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 4353–4361

Peng J, Kervadec H, Dolz J, Ayed IB, Pedersoli M, Desrosiers C (2019) Discretely-constrained deep network for weakly supervised segmentation. arXiv:190805770

Perone CS, Cohen-Adad J (2018) Deep semi-supervised segmentation with weight-averaged consistency targets. In: Deep learning in medical image analysis and multimodal learning for clinical decision support, pp 12–19

Perone CS, Cohen-Adad J (2019) Promises and limitations of deep learning for medical image segmentation. J Med Artif Intell 2. http://jmai.amegroups.com/article/view/4659

Perone CS, Calabrese E, Cohen-Adad J (2018) Spinal cord gray matter segmentation using deep dilated convolutions. Sci Rep 8(1). https://doi.org/10.1038/s41598-018-24304-3

Perone CS, Ballester P, Barros RC, Cohen-Adad J (2019) Unsupervised domain adaptation for medical imaging segmentation with self-ensembling. Neuroimage 194:1–11

Pohlen T, Hermans A, Mathias M, Leibe B (2017) Full-resolution residual networks for semantic segmentation in street scenes. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 4151–4160

Proenca H, Neves JC (2019) Segmentation-less and non-holistic deep-learning frameworks for iris recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops

Qin Y, Kamnitsas K, Ancha S, Nanavati J, Cottrell G, Criminisi A, Nori A (2018) Autofocus layer for semantic segmentation. In: International conference on medical image computing and computer-assisted intervention. Springer, pp 603–611

Raghu M, Zhang C, Kleinberg J, Bengio S (2019) Transfusion: understanding transfer learning with applications to medical imaging. arXiv:190207208

Reddick W, Glass J, Cook E, Elkin T, Deaton R (1997) Automated segmentation and classification of multispectral magnetic resonance images of brain using artificial neural networks. IEEE Trans Med Imaging 16(6):911–918. https://doi.org/10.1109/42.650887

Reilhac A, Batan G, Michel C, Grova C, Tohka J, Collins D, Costes N, Evans A (2005) PET-SORTEO: validation and development of database of simulated PET volumes. IEEE Trans Nucl Sci 52(5):1321–1328. https://doi.org/10.1109/tns.2005.858242

Remillard J (2018) Images to OSM. https://github.com/jremillard/images-to-osm

Ren S, He K, Girshick R, Sun J (2015) Faster R-CNN: towards real-time object detection with region proposal networks. In: Advances in neural information processing systems, pp 91–99

Rezaei M, Harmuth K, Gierke W, Kellermeier T, Fischer M, Yang H, Meinel C (2017) A conditional adversarial network for semantic segmentation of brain tumor. In: International conference on medical image computing and computer assisted intervention, Brainlesion Workshop. Springer, pp 241–252

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

Roy AG, Conjeti S, Sheet D, Katouzian A, Navab N, Wachinger C (2017) Error corrective boosting for learning fully convolutional networks with limited data. In: International conference on medical image computing and computer-assisted intervention. Springer, pp 231–239

Rumelhart DE, Hinton GE, Williams RJ (1986) Learning representations by back-propagating errors. Nature 323(6088):533–536. https://doi.org/10.1038/323533a0

Russell BC, Torralba A, Murphy KP, Freeman WT (2008) LabelMe: a database and web-based tool for image annotation. Int J Comput Vis 77(1–3):157–173. https://doi.org/10.1007/s11263-007-0090-8

Russell SJ, Norvig P (2016) Artificial intelligence: a modern approach. Pearson Education Limited, Kuala Lumpur

Salehi SSM, Erdogmus D, Gholipour A (2017) Tversky loss function for image segmentation using 3D fully convolutional deep networks. In: International workshop on machine learning in medical imaging. Springer, pp 379–387

Sandler M, Howard A, Zhu M, Zhmoginov A, Chen LC (2018) MobileNetV2: inverted residuals and linear bottlenecks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 4510–4520

Saxena S, Verbeek J (2016) Convolutional neural fabrics. In: Advances in neural information processing systems, pp 4053–4061

Schlemper J, Oktay O, Schaap M, Heinrich M, Kainz B, Glocker B, Rueckert D (2019) Attention gated networks: learning to leverage salient regions in medical images. Med Image Anal 53:197–207

Shahriari S, Garcia D (2018) Meshfree simulations of ultrasound vector flow imaging using smoothed particle hydrodynamics. Phys Med Biol 63(20):205011. https://doi.org/10.1088/1361-6560/aae3c3

Shaw A, Hunter D, Landola F, Sidhu S (2019) SqueezeNAS: fast neural architecture search for faster semantic segmentation. In: Proceedings of the IEEE international conference on computer vision workshops

Shin HC, Tenenholtz NA, Rogers JK, Schwarz CG, Senjem ML, Gunter JL, Andriole KP, Michalski M (2018) Medical image synthesis for data augmentation and anonymization using generative adversarial networks. In: International workshop on simulation and synthesis in medical imaging. Springer, pp 1–11

Shorten C, Khoshgoftaar TM (2019) A survey on image data augmentation for deep learning. J Big Data. https://doi.org/10.1186/s40537-019-0197-0

Sifre L (2014) Rigid-motion scattering for image classification. PhD thesis, CMAP, Ecole Polytechnique

Simard PY, Steinkraus D, Platt JC (2003) Best practices for convolutional neural networks applied to visual document analysis. In: Proceedings of the seventh international conference on document analysis and recognition—Volume 2. IEEE Computer Society, USA, ICDAR '03, p 958

Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. arXiv:14091556

Simpson AL, Antonelli M, Bakas S, Bilello M, Farahani K, van Ginneken B, Kopp-Schneider A, Landman BA, Litjens G, Menze B, et al. (2019) A large annotated medical image dataset for the development and evaluation of segmentation algorithms. arXiv:190209063

Sinha A, Dolz J (2019) Multi-scale guided attention for medical image segmentation. arXiv:190602849

Son J, Park SJ, Jung KH (2017) Retinal vessel segmentation in fundoscopic images with generative adversarial networks. arXiv:170609318

Song G, Myeong H, Mu Lee K (2018) Seednet: automatic seed generation with deep reinforcement learning for robust interactive segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 1760–1768

Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R (2014) Dropout: a simple way to prevent neural networks from overfitting. J Mach Learn Res 15(1):1929–1958

Srivastava RK, Greff K, Schmidhuber J (2015) Highway networks. arXiv:150500387

Stanley KO, Miikkulainen R (2002) Evolving neural networks through augmenting topologies. Evol Comput 10(2):99–127

SUYEgit (2018) Mask R-CNN for surgery robot. https://github.com/SUYEgit/Surgery-Robot-Detection-Segmentation/

Tabary J, Hugonnard P, Mathy F (2007) SINDBAD: a realistic multi-purpose and scalable X-ray simulation tool for NDT applications. In: DIR 2007: international symposium on digital industrial radiology and computed tomography

Taghanaki SA, Duggan N, Ma H, Hou X, Celler A, Benard F, Hamarneh G (2018) Segmentation-free direct tumor volume and metabolic activity estimation from pet scans. Comput Med Imaging Graphics 63:52–66

Taghanaki SA, Abhishek K, Azizi S, Hamarneh G (2019a) A kernelized manifold mapping to diminish the effect of adversarial perturbations. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 11340–11349

Taghanaki SA, Abhishek K, Hamarneh G (2019b) Improved inference via deep input transfer. In: International conference on medical image computing and computer-assisted intervention. Springer, pp 819–827

Taghanaki SA, Bentaieb A, Sharma A, Zhou SK, Zheng Y, Georgescu B, Sharma P, Grbic S, Xu Z, Comaniciu D, et al. (2019c) Select, attend, and transfer: light, learnable skip connections. In: Medical image computing and computer-assisted intervention workshop on machine learning in medical imaging

Taghanaki SA, Havaei M, Berthier T, Dutil F, Di Jorio L, Hamarneh G, Bengio Y (2019d) InfoMask: masked variational latent representation to localize chest disease. In: International conference on medical image computing and computer assisted intervention

Taghanaki SA, Zheng Y, Zhou SK, Georgescu B, Sharma P, Xu D, Comaniciu D, Hamarneh G (2019e) Combo loss: handling input and output imbalance in multi-organ segmentation. Comput Med Imaging Graphics 75:24–33

Tajbakhsh N, Jeyaseelan L, Li Q, Chiang JN, Wu Z, Ding X (2020) Embracing imperfect datasets: a review of deep learning solutions for medical image segmentation. Med Image Anal 63:101693. https://doi.org/10.1016/j.media.2020.101693

Tarvainen A, Valpola H (2017) Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. In: Advances in neural information processing systems, pp 1195–1204

Tsai HF, Gajda J, Sloan TF, Rares A, Shen AQ (2019) Usiigaci: instance-aware cell tracking in stain-free phase contrast microscopy enabled by machine learning. SoftwareX 9:230–237

Vorontsov E, Molchanov P, Byeon W, De Mello S, Jampani V, Liu MY, Kadoury S, Kautz J (2019) Towards semi-supervised segmentation via image-to-image translation. arXiv:190401636

Vuola AO, Akram SU, Kannala J (2019) Mask R-CNN and U-net ensembled for nuclei segmentation. arXiv:190110170

Wang EK, Zhang X, Pan L, Cheng C, Dimitrakopoulou-Strauss A, Li Y, Zhe N (2019a) Multi-path dilated residual network for nuclei segmentation and detection. Cells 8(5):499

Wang F, Jiang M, Qian C, Yang S, Li C, Zhang H, Wang X, Tang X (2017a) Residual attention network for image classification. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 3156–3164

Wang G, Luo P, Lin L, Wang X (2017b) Learning object interactions and descriptions for semantic image segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 5859–5867

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. IEEE, pp 1451–1460

Wang S, Rong R, Yang DM, Cai L, Yang L, Luo D, Yao B, Xu L, Wang T, Zhan X, et al. (2019b) Computational staining of pathology images to study tumor microenvironment in lung cancer. Available at SSRN 3391381

Wang W, Lai Q, Fu H, Shen J, Ling H (2019c) Salient object detection in the deep learning era: an in-depth survey. arXiv:190409146

Wang X, Girshick R, Gupta A, He K (2018b) Non-local neural networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 7794–7803

Wang X, Wang H, Niu S, Zhang J (2019d) Detection and localization of image forgeries using improved mask regional convolutional neural network. Math Biosci Eng 16:4581–4593

Wang Z, Sarcar S, Liu J, Zheng Y, Ren X (2018c) Outline objects using deep reinforcement learning. arXiv:180404603

Wen W, Wu C, Wang Y, Chen Y, Li H (2016) Learning structured sparsity in deep neural networks. In: Advances in neural information processing systems, pp 2074–2082

Weng Y, Zhou T, Li Y, Qiu X (2019a) NAS-Unet: neural architecture search for medical image segmentation. IEEE Access 7:44247–44257

Weng Y, Zhou T, Li Y, Qiu X (2019b) NAS-unet: neural architecture search for medical image segmentation. IEEE Access 7:44247–44257. https://doi.org/10.1109/access.2019.2908991

Wessel J, Heinrich MP, von Berg J, Franz A, Saalbach A (2019) Sequential rib labeling and segmentation in chest X-ray using Mask R-CNN. In: International conference on medical imaging with deep learning—extended abstract track, London, United Kingdom. https://openreview.net/forum?id=SJxuHzLjFV

Wojna Z, Ferrari V, Guadarrama S, Silberman N, Chen LC, Fathi A, Uijlings J (2017) The devil is in the decoder. arXiv:170705847

Wong KC, Moradi M, Tang H, Syeda-Mahmood T (2018) 3D segmentation with exponential logarithmic loss for highly unbalanced object sizes. In: International conference on medical image computing and computer assisted intervention. Springer, pp 612–619

Wu Y, He K (2018) Group normalization. In: Proceedings of the European conference on computer vision, pp 3–19

Wu Z, Shen C, Van Den Hengel A (2019) Wider or deeper: revisiting the resnet model for visual recognition. Pattern Recognit 90:119–133

Xiao J, Hays J, Ehinger KA, Oliva A, Torralba A (2010) SUN database: Large-scale scene recognition from abbey to zoo. In: 2010 IEEE computer society conference on computer vision and pattern recognition. IEEE, pp 3485–3492

Xie S, Tu Z (2015) Holistically-nested edge detection. In: Proceedings of the IEEE international conference on computer vision, pp 1395–1403

Xie X, Niu J, Liu X, Chen Z, Tang S (2020) A survey on domain knowledge powered deep learning for medical image analysis. arXiv:200412150

Xu X, Lu Q, Yang L, Hu S, Chen D, Hu Y, Shi Y (2018) Quantization of fully convolutional networks for accurate biomedical image segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 8300–8308

Xue Y, Xu T, Zhang H, Long LR, Huang X (2018) SegAN: adversarial network with multi-scale L1 loss for medical image segmentation. Neuroinformatics 16(3–4):383–392

Yang D, Xu D, Zhou SK, Georgescu B, Chen M, Grbic S, Metaxas D, Comaniciu D (2017a) Automatic liver segmentation using an adversarial image-to-image network. In: International conference on medical image computing and computer-assisted intervention. Springer, pp 507–515

Yang Q, Li N, Zhao Z, Fan X, Chang EI, Xu Y, et al. (2018) MRI cross-modality neuroimage-to-neuroimage translation. arXiv:180106940

Yang X, Yu L, Wu L, Wang Y, Ni D, Qin J, Heng PA (2017b) Fine-grained recurrent neural networks for automatic prostate segmentation in ultrasound images. In: Thirty-first AAAI conference on artificial intelligence

Yu B, Zhou L, Wang L, Fripp J, Bourgeat P (2018a) 3D cGAN based cross-modality MR image synthesis for brain tumor segmentation. In: 2018 IEEE 15th international symposium on biomedical imaging. IEEE, pp 626–630

Yu C, Wang J, Peng C, Gao C, Yu G, Sang N (2018b) BiSeNet: Bilateral segmentation network for real-time semantic segmentation. In: Proceedings of the European conference on computer vision (ECCV). Springer, pp 334–349. https://doi.org/10.1007/978-3-030-01261-8_20

Yuan Y (2017) Automatic skin lesion segmentation with fully convolutional-deconvolutional networks. arXiv:170305165

Zamir AR, Sax A, Shen W, Guibas LJ, Malik J, Savarese S (2018) Taskonomy: disentangling task transfer learning. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 3712–3722

Zhang P, Zhong Y, Deng Y, Tang X, Li X (2019) A survey on deep learning of small sample in biomedical image analysis. arXiv:190800473

Zhang W, Witharana C, Liljedahl A, Kanevskiy M (2018a) Deep convolutional neural networks for automated characterization of arctic ice-wedge polygons in very high spatial resolution aerial imagery. Remote Sens 10(9):1487

Zhang Y, Yang L, Chen J, Fredericksen M, Hughes DP, Chen DZ (2017) Deep adversarial networks for biomedical image segmentation utilizing unannotated images. In: International conference on medical image computing and computer-assisted intervention. Springer, pp 408–416

Zhang Y, Miao S, Mansi T, Liao R (2018b) Task driven generative modeling for unsupervised domain adaptation: application to X-ray image segmentation. In: International conference on medical image computing and computer-assisted intervention. Springer, pp 599–607

Zhang Z, Yang L, Zheng Y (2018c) Translating and segmenting multimodal medical volumes with cycle-and shape-consistency generative adversarial network. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 9242–9251

Zhang Z, Zhang X, Peng C, Xue X, Sun J (2018d) Exfuse: enhancing feature fusion for semantic segmentation. In: Proceedings of the European conference on computer vision, pp 269–284

Zhao H, Li H, Cheng L (2017a) Synthesizing filamentary structured images with GANs. arXiv:170602185

Zhao H, Shi J, Qi X, Wang X, Jia J (2017b) Pyramid scene parsing network. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 2881–2890

Zhao M, Hamarneh G (2019a) Retinal image classification viavasculature-guided sequential attention. In: International conference on computer vision workshop on visual recognition for medical images

Zhao M, Hamarneh G (2019b) Tree-LSTM: using LSTM to encode memory in anatomical tree prediction from 3D images. In: Medical image computing and computer-assisted intervention workshop on machine learning in medical imaging

Zhao T, Yang Y, Niu H, Wang D, Chen Y (2018) Comparing U-Net convolutional network with Mask R-CNN in the performances of pomegranate tree canopy segmentation. In: Asia-pacific remote sensing

Zhao ZQ, Zheng P, Xu ST, Wu X (2019) Object detection with deep learning: a review. IEEE Trans Neural Netw Learn Syst 30(11):3212–3232. https://doi.org/10.1109/tnnls.2018.2876865

Zhen X, Li S (2015) Towards direct medical image analysis without segmentation. arXiv:151006375

Zhou B, Lapedriza A, Xiao J, Torralba A, Oliva A (2014) Learning deep features for scene recognition using places database. In: Advances in neural information processing systems, pp 487–495

Zhou B, Zhao H, Puig X, Fidler S, Barriuso A, Torralba A (2017) Scene parsing through ADE20K dataset. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 633–641

Zhou S, Nie D, Adeli E, Yin J, Lian J, Shen D (2019a) High-resolution encoder-decoder networks for low-contrast medical image segmentation. IEEE Trans Image Process 29:461–475

Zhou T, Ruan S, Canu S (2019b) A review: deep learning for medical image segmentation using multimodality fusion. Array 3–4:100004. https://doi.org/10.1016/j.array.2019.100004

Zhou Z, Siddiquee MMR, Tajbakhsh N, Liang J (2018) 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 X, Zhou H, Yang C, Shi J, Lin D (2018) Penalizing top performers: conservative loss for semantic segmentation adaptation. In: Proceedings of the European conference on computer vision, pp 568–583

Zhu Z, Liu C, Yang D, Yuille A, Xu D (2019) V-NAS: neural architecture search for volumetric medical image segmentation. In: 2019 international conference on 3D vision (3DV). IEEE. https://doi.org/10.1109/3dv.2019.00035

Zoph B, Le QV (2016) Neural architecture search with reinforcement learning. arXiv:161101578

Zou Z, Shi Z, Guo Y, Ye J (2019) Object detection in 20 years: a survey. arXiv:190505055