Recent Advances in the Applications of Convolutional Neural Networks to Medical Image Contour Detection

Article · August 2017

Zizhao Zhang
University of Florida

Hai Su
University of Florida

Xiaoshuang Shi
Tsinghua University

Recent Advances in the Applications of Convolutional Neural Networks to Medical Image Contour Detection

Zizhao Zhang[1a], Fuyong Xing[a], Hai Su[a], Xiaoshuang Shi[a], Lin Yang[a]

[a] *University of Florida*

Abstract

The fast growing deep learning technologies have become the main solution of many machine learning problems for medical image analysis. Deep convolution neural networks (CNNs), as one of the most important branch of the deep learning family, have been widely investigated for various computer-aided diagnosis tasks including long-term problems and continuously emerging new problems. Image contour detection is a fundamental but challenging task that has been studied for more than four decades. Recently, we have witnessed the significantly improved performance of contour detection thanks to the development of CNNs. Beyond purusing performance in existing natural image benchmarks, contour detection plays a particularly important role in medical image analysis. Segmenting various objects from radiology images or pathology images requires accurate detection of contours. However, some problems, such as discontinuity and shape constraints, are insufficiently studied in CNNs. It is necessary to clarify the challenges to encourage further exploration. The performance of CNN based contour detection relies on the state-of-the-art CNN architectures. Careful investigation of their design principles and motivations is critical and beneficial to contour detection. In this paper, we first review recent development of medical image contour detection and point out the current confronting challenges and problems. We discuss the development of general CNNs and their applications in image contours (or edges) detection. We compare those methods in detail, clarify their strengthens and weaknesses. Then we review their recent applications in medical image analysis and point out limitations, with the goal to light some potential directions in medical image analysis. We expect the paper to cover comprehensive technical ingredients of advanced CNNs to enrich the study in the medical image domain.

[1] E-mail: zizhaozhang@ufl.edu

Keywords:
Medical image, Contour detection, Deep learning, Convolution neural networks

1. Introduction

Medical image analysis is the foundation of a computer-aided diagnosis (CAD) system. The analysis of medical images with different modalities usually requires accurate segmentation to isolate abnormal objects (cells or organs) to support efficient quantization. Contour detection is a fundamental prerequisite for medical image segmentation, with the aim to detect the edges from images and further collect the knowledge of the contours of objects. Accurate and fast contour detection is a long-term study in this domain, which suffers from many difficulties. In recent years, we have witnessed inspirational renovation in medical image analysis, particularly image segmentation, due to the development of advanced machine learning technologies.

Edge is the basic components of images. Detecting object edges and contours is critical in many practical computer vision and medical image computing tasks. The research in contour detection is a huge family comprised by a large number of directions using various computer vision, image processing, and machine learning techniques [156]. Early contour detection methods are dominated by unsupervised approaches, with the aim to estimate the local gradient changes. In the past five years, supervised methods gradually dominate this area as a result of both accuracy and efficiency advantages. The main idea is to train a machine learning classifier to predict central pixel labels (edge or non-edge) of local patches. Both directions require heavy hand-crafted features to accurately represent the local gradient information. Contour detection includes the detection of edges but can simultaneously outline the continuous edges belong to object contours. Therefore, contour detection is substantially more challenging than edge detection since it models both low-level gradients and high-level object information. Under conventional directions, the integration of low-level and high-level cues is difficult, which often results in complex and computationally demanding frameworks, including pre-processing, feature engineering, classifier training, and post-processing. A fast and highly-integrated method that can accept raw images and output contour maps is quite hypothetical, and the advantages of deep learning give hopes to the demand.

There is world-wide recognition that deep learning has advanced the artificial intelligence (AI) to the next generation [103, 61, 179]. The family of deep learning is comprised of a number of unsupervised and supervised learning models

[75, 107, 108, 62], such as Restricted Boltzmann Machine (RBM) [177, 149, 176] and Recurrent Neural Networks (RNNs). CNN is a supervised model widely used for image understanding. Compared with conventional machine learning models, such as support vector machines (SVM) [37], random forests [113]. CNN has a deep layer-wise structure, making it proficient at learning hierarchical and nonlinear representations to represent and discover complex and intricate high-dimensional data structures to support discriminative classification. Therefore, CNN is also a representation learning method [9].

In edge detection, there are multiple remarkable studies using standard CNNs [186, 56, 12] have achieved marginal improvement over conventional methods. However, the standard CNN suffers from the computationally bottleneck of their patch-to-pixel dense prediction paradigm. The development of contour detection continuously takes benefits from the development of semantic segmentation [151, 31, 70, 40, 117, 217, 159, 225, 138]. One of the most critical techniques for dense prediction tasks is end-to-end CNN, proposed by [126, 185], which performs pixel-wise prediction in a single feedforward. At present, end-to-end training becomes the standard for most kinds of structured outputs, such as bounding boxes [165, 57], shape [87], orientations [138], which offers much flexibility to CNN beyond the form of image labels. The end-to-end training manner for dense pixel prediction [217, 225, 138] dramatically improves the performance of contour detection, surpassing a significant margin over preceding standard CNN based methods and even surpassing the empirical accuracy of human annotators. These advantages dramatically affect the medical image domain.

Different from natural images, medical images do not have rich semantic information, effective usage of low-level edge information is the key to support accurate segmentation. Moreover, the failure of detecting edges will cause huge issues in diagnostic precision, for example, the failure of segmenting touching cells will cause abnormal cell size statistics. Therefore, contour detection is usually treated as an intermediate step for image segmentation. Some segmentation studies have implicit contour detection because the main challenging of object segmentation is the accurate location of edges (this paper will take such kinds of segmentation methods into the consideration), such as segmenting neuronal membranes in electron microscopy images [34] and vessels in retinal images [56]. We have witnessed numerous state-of-the-art CNN based methods being successfully applied to medical image contour detection and segmentation [28, 139, 171]. However, direct technical transferring sometimes conceals several critical problems in the medical images but may not being concerned in the natural image domain, such as the detection of weak edges, the processing of discontinuity of

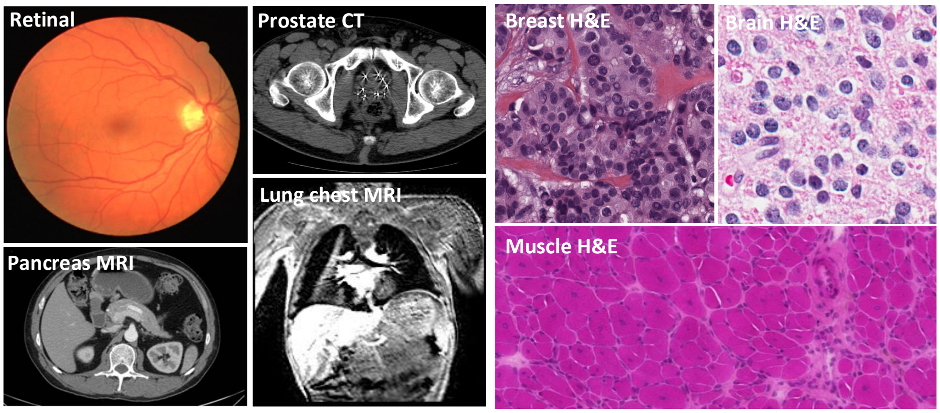

Figure 1: The left side shows four kinds of CT, MRI, or tomography organ images. The right side shows three kinds of microscopic images. Segmenting the objects (e.g. pancreas or nucleus) needs clear detection of object contours. In breast or muscle images, detecting contours is obviously very challenging due to the severe touching objects and artifacts.

contours, and the detection of edges from high signal-to-noise ratio (SNR) images. The solutions to these problems are significant in medical image analysis.

We start by discussing the difficulties of edge/contour detection in medical image computing and review some conventional approaches in Section 2 To better understand the development of CNNs in image contour detection, we briefly introduce the principle of CNNs in Section 3 and discuss most state-of-the-art CNN based methods for edge detection in Section 5, with the goal to clarity the key problems they are addressing and their advantages for medical image usage. After understanding the principle of CNNs for contour detection, in Section 6, we review recent methods for medical image contour detection and segmentation to help understand underlying technical basics of current methods in medical image analysis and build the connection to the state-of-the-art method in the computer vision community, and also show the limitations of current methods. Section 7 discusses some key problems and potential directions. Section 8 concludes the paper. We expect this paper can cover the necessary technical advances in CNNs that is useful for contour detection and, more importantly, can attract attentions of the underlying problems and lead to further exploration of the CNN technologies in medical image analysis.

2. Overview of Contour Detection for Medical Images

In this section, we provide an overview of the challenges and significance of image contour detection in medical images. Then we review several kinds of conventional directions for medical image contour detection, with the goal to encourage inspirations in CNN designing.

2.1. Challenges and significance

Compared with natural images containing all kinds of semantic objects, medical images are more modality-specific such that in one modality, there is less semantic and texture information inside or between objects. In radiological data like pancreas MRI or CT or ultrasound images, the targeting objects are organs or bones. In pathological images like hematoxylin and eosin (H&E) stained lung cell specimens [50, 218, 130], the targeting objects are cells and diseased regions. Organs and cells usually have consistent appearance. Therefore, object shapes and structures play a key role for medical image segmentation or detection. Figure 1 shows some types of medical images that need careful process of object contours to achieve accurate segmentation.

The image quality defection due to the acquisition and imaging processes is common and significant [202, 226]. Noises bring obstacles to edge detection because it reduces the contrast of real edges and also introduce spurious edges due to noisy contrast. Although applying denoising algorithms before some local edge detection can reduce the effects to some extent, this approach has been shown not very promising [152]. A global method with some prior knowledge of targeting objects is supposed to overcome the effects of local noises. The fine detection of edges and global object contours are equally important and require discrimination when prediction. For example, in retinal images, the detection and segmentation of blood vessels require very fine detection of subtle edges. While in pathological images, ignoring gradients caused by staining noises and boundaries of small cells is critical.

Detecting weak or broken edges due to occlusion or staining artifacts between touching or overlapping objects is a long-term studied problem in medical image analysis. Sometimes detecting these kinds of edges which are even visually in-discriminative seems impossible, but learning to link the broken edges is a remedial measure [69]. We believe this problem is an active research topic of the contour detection. In the earlier stage, deformable models [91, 23, 223, 22] are popular techniques to guarantee the continuity and smoothness of object contours, because active contour models use a parametric or non-parametric representation to initialize object boundaries and transform these closed contours to align with objects. However, currently this area is not active because active contour is built on particular assumptions. Recently, we have seen work [174] that train CNNs to predict the movement of active contours.

Besides direct contour detection, [214, 197, 109] have studied how to detect global, closed, or convex object contours from a set of broken contour pieces containing interesting edges and noisy edges. However, these methods are difficult to generalize to real datasets due to the relied assumptions, such as bilateral symmetry of objects. Additionally, [167] have studied contour completion using conditional random field (CRF) models, and [120] extends the technique to handle medical images. [227] and [198] use RNNs and autoencoders [78] to achieve contour completion.

High-level reasoning is an important factor for contour completion. Human can easily recognize the broken edges between objects and identity touching or overlapping objects, although the actual gradient in the broken edges is hardly seen. We believe there are two reasons at least [182, 178, 81, 101]:

1. Human vision has a strong reasoning ability by observing surrounding object contours connecting to the broken points, and thus it can easily recognize and predict the existence of broken edges.
2. Human vision has high-level prior knowledge about the observing objects' appearance [101], so it can estimate rough object appearance and reject unfamiliar appearance (touched and connected multiple objects).

Both need to take advantage of the context information. However, as we previously discussed, most of previous CNN based contour detection and segmentation methods do not focus on this kind of context information. In semantic segmentation tasks, we have noticed a lot of work to model the context information using methods like CRF and markov random field (MRF) [117, 119, 3, 110, 31, 235, 31, 125]. The inference of edges is supposed to be more difficult because it contains less semantic knowledge and require more complex understanding of shapes and structures of objects. For example, [55] have studied occlusion boundary detection by exploring deep context from CNNs, including local structural boundary patterns, observations from surrounding regions, and temporal context.

In addition, the effective learning from limited annotated data is an significant topic because of the difficulty in collecting large-scale medical image datasets. Usually CNNs require large-scale datasets to train. Semi-supervised and unsupervised learning methods [112] and transfer learning [189, 206] are recently been discussed in the study of CNNs. In addition, a better CNN architecture design can significantly increase the efficiency of parameter utilization. We will discuss the details in the following.

2.2. Previous medical image contour detection

Conventional contour detection and segmentation methods before the prevalence of deep learning have numerous directions in medical image computing, and a comprehensive review can be found in [222].

Intensity thresholding [60] is one of the early-stage approaches for medical image segmentation. For some specific image modalities, such as fluorescence microscopy images, where the target objects (e.g., nuclei or cells) are usually brightly stained in a dark background, it is effective to apply intensity thresholding to object segmentation [32]; however, it is difficult for thresholding to handle other image modalities (e.g., H&E stained images), especially for touching or partially overlapped objects.

Watershed transform [170] is a popular segmentation method in medical images [226], which pursues the 'catchment basins' (gradient discontinuity points). It can be used to find the continuous contours of objects, and therefore it is popular at segmenting multiple objects like cells in pathological images [143]. However, watershed usually suffers from over-segmentation, and thus marker-controlled watershed [60] has been proposed for effective contour detection and segmentation. An alternative method to handle over-segmentation is to merge falsely segmented regions with certain criteria [115].

One widely-studied direction for medical image contour detection and segmentation is deformable models [42]. Deformable model based methods focus on deforming an initial active contour to align the object boundary by solving an energy function. Representative deformable models include geodesic models or level-set models (such as Chan-Vese model [23]) [22], parametric models (Snake [91] and GVF [223]). Many related work has been proposed for object contour delineation [220, 233, 111, 39] and some of them are combined with shape prior modeling for touching object segmentation in medical images [2, 221]. One potential limitation of deformable models is the requirement of proper contour initialization, which might be achieved using effective object detection methods [222].

Graph-based methods are another popular category of methods for medical image contour detection and segmentation. In graph partition, the max-flow/min-cut algorithm [17, 16] is usually used to minimize an energy function, and it has been successfully applied to contour detection in medical images [1, 24]. Normalized cut [188] is proposed to avoid the bias of favoring small sets in the global

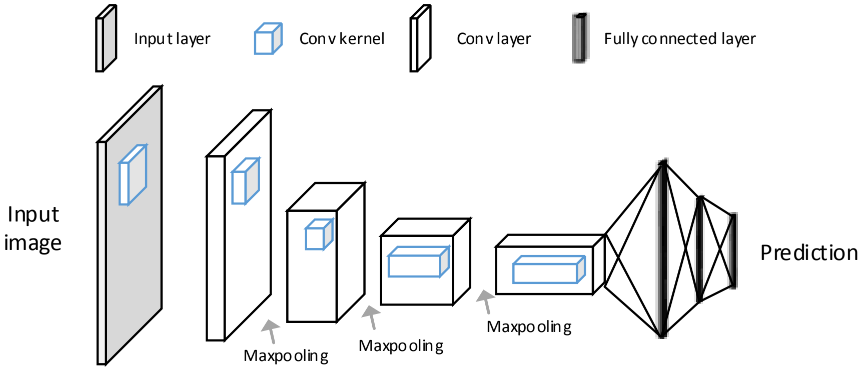

Figure 2: An illustration of the architecture of a CNN, including five convolutional layers, three max pooling layers, and two three fully connected layers. The first convolutional layer takes an input image and the last fully connected layer predicts the label.

minimum cut, and a generalized normalized cut [11] has been proposed for object segmentation in microscopy images. Some other graph partitioning methods such as random walk [64] and isoperimetric partitioning [65] have been also reported for object segmentation in medical image data.

Conventional machine learning methods have been applied to medical image contour detection and segmentation. For pixel-wise classification-base segmentation, it is usually necessary to conduct further processing to split touching objects [98]; for superpixel-wise classification, it would improve the computational efficiency, but the pre-generated superpixels need to well adhere real object boundaries [88]. In addition, the conventional machine learning approaches require manual feature representation design, which is not a trivial task in some medical applications.

3. Convolutional Neural Networks

In this section, we briefly introduce the basic concept of CNNs. As the predecessor of CNN, ANN is originally inspired by the goal of modeling biological neural systems. Its organization simulates the mechanism of information transmission in the brain neuron. The computation unit contains a linear transformation $\boldsymbol{z}_i = \sum_i w_i \boldsymbol{x}_i + b$ plus an activation function $\boldsymbol{y}_i = \frac{1}{1+e^{-\boldsymbol{z}_i}}$ (e.g. Sigmoid) on the input $\boldsymbol{x}$ and generate an output $\boldsymbol{y}$. $\boldsymbol{w} = [w_1, ..., w_n]^T$ is the weights function connecting previous neurons to next neutron. Computation units are connected one after another, which compose a layer-wise organized architecture. When connecting to the output layer, a Softmax or Sigmoid function is often used to map the output to $[0, 1]$, representing label probabilities. The above formulation is a very basic ANN design, which primarily models the biological neuron system. After that, the design of ANN towards to the machine learning and engineering guidance.

3.1. Architecture

CNN has a very similar architecture with ANN, which is composed by a series of computational layers [104]. Each layer contains a set of neurons and may have learnable parameters to decide how the neutrons between layers are connected. Therefore, the overall architecture is structured by cascaded layers. Figure 2 shows a CNN architecture with eight layers.

The main building brick is the convolutional (Conv) layer, whose Conv kernels (or filters) perform convolutional operation across the whole image spatial locations to generate output image representations (i.e. feature maps). The spatial extent of kernels refer to as the receptive field. This local connectivity through the receptive field is originally inspired by the brain neuron science [49]. Pooling layer is inserted between Conv layers for the purpose of downsampling the feature map dimension. The most common used pooling layer is max pooling, which keeps the highest response value in an image extent and discard the rest, and perform this operation crossing the whole image. Activation layers map data points non-linearly. The appropriate settings of activations is critical to the behaviors of CNN training. The most common used activation at present for CNN is ReLU [149], which simply performs $y = \max(0, x)$. It is applied after a Conv layer or a fully connected layer. ReLU is an important technique for modern CNNs. Previous activation functions such as Sigmoid and Tanh suffer from strong gradient vanishing (or gradient saturation) problems [59], which can be accumulated and getting severer as the layer increases.

Apart from the basic layers, there are a variety of layers proposed by modern CNNs. For example, Dropout [195] and batch normalization [85] are now standard configurations in the CNN design patterns. Modified/generalized convolutional layers [228, 126], ReLU layers [129], and pooling layers [231] are also specifically designed for various applications. We will discuss some of them in the following.

3.2. Network training

Recently, the improvement of the optimization algorithm, hardware capacity, and functional deep learning libraries, make the training more easier than before. A beginner with moderate experiences can deploy the training of very deep networks in a few lines of code on a GPU. The main component of CNN training is stochastic gradient descent (SGD) [15] and backpropagation [105].

SGD minimizes the empirical risk of the loss function $J(\theta)$ by updating the parameters θ:

$$\boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t - \gamma \sum_{1}^{m} \nabla_\theta E[J(\theta)] \tag{1}$$

where n is the number of observed training data and γ is the learning rate. SGD randomly sample a mini-batch of m samples from total training samples and update the parameters using the averaged gradients generated by mini-batch samples. Standard SGD could easily trap at local optima and lead to slow convergence [201]. Then momentum method is introduced to resolve this problem by controlling the gradient velocity during optimization [201]. The SGD with momentum is defined as

$$\boldsymbol{v}_{t+1} = \lambda \boldsymbol{v}_t - \gamma \sum_{1}^{m} \nabla_\theta E[J(\theta)] \tag{2}$$

$$\boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t - \boldsymbol{v}_{t+1} \tag{3}$$

where $\lambda \in [0, 1]$ is the momentum coefficient. Based on the basic SGD with momentum, there are also new algorithms to improve the training efficiency and lead to better convergence, such as Nesterov momentum [201], Adagrad [48], RMSprop [76], Adam [95]. Selecting a appropriate learning rate is tricky. The last three can adaptively adjust the learning rate per parameter at each update, which have been shown to lead to faster convergence. Current state-of-the-art methods still use different optimization algorithms based on specific applications. The optimal choice depends on specific problems.

Backpropagation propagates the errors computed in the loss layers back to all proceeding layers. Every computational layer will generate the gradient w.r.t. its own parameters accordingly. The overall procedure follows the basic chain rule. Let's define the object function J as

$$J(\boldsymbol{x}, \boldsymbol{g}; \theta) = \frac{1}{m} \sum_{i=1}^{m} loss(\boldsymbol{g}_i, f(\boldsymbol{x}_i; \theta)), \tag{4}$$

where $loss$ computes the difference between the prediction $f(x_i;\theta)$ and the groundtruth $\boldsymbol{g}_i$. There are various types of loss functions, such as Cross-entropy, Softmax, Euclidean distance, Max-margin, etc, for botch classification and regression purposes. f denotes overall function of CNN, so f takes an input image x and outputs y_L as its predicted label. Suppose y_L is computed by a fully connected layer (the L- layer of the network), defined as

$$y_L = \sigma(W_L y_{L-1} + b_L) \tag{5}$$

where y_{L-1} is the output of the $(L-1)$-th layer. σ is the activation function. The gradient of J w.r.t W_L and y_{L-1} is defined as:

$$\frac{\partial J}{\partial W_L} = \frac{\partial J}{\partial y_L} \cdot \frac{\partial y_L}{\partial \sigma} \cdot \frac{\partial \sigma}{\partial W_L}, \tag{6}$$

$$\frac{\partial J}{\partial y_{L-1}} = \frac{\partial J}{\partial y_L} \cdot \frac{\partial y_L}{\partial \sigma} \cdot \frac{\partial \sigma}{\partial y_{L-1}} \tag{7}$$

Eq. (6) computes the gradients respecting to the weights of layer L and Eq. (7) computes the gradients respecting to the input y_{L-1} of layer L.

Successfully training CNN networks is not as simple as its mathematical definition. The overall optimization is highly non-convex and the process is difficult to visualize. Overfitting is one of the long-term challenge actively studied in the community. There are a wide range of well-investigated approaches that substantially alleviate CNN training difficulties, for instance, data augmentation, weight initialization [77, 201, 58, 72], regularizations [195, 213], activations [149, 59, 63, 129, 35], normalization [85], and skip-connection [73]. We refer readers to [68] for more details.

4. State-of-the art CNN Architectures

In this section, we introduce several well-known CNN architectures, which are recognized as the milestone in the CNN development and the basement of various computer vision tasks, specially image edge detection. We also discuss the related variations to address specific problem in CNNs, such as ensembling, generalization, etc. The performance of these CNNs is publicly validated on the annual ImageNet Large Scale Image Recognition Challenge (ILSIRC). Figure 3 shows the winner networks of ImageNet challenges in the past five years. From 2010 to 2015, the classification error rate has decreased by more than nine times.

4.1. CNN architecture benchmarks

LeNet stands for the first successful application of CNN. It is proposed by [104] and used for hand-written digit recognition. ConvNet to allow the weight sharing between neurons, which dramatically decreases the heavy parameters needed in ANNs. This network is much shallower compared with recent architectures, including 2 Conv layers with an intermediate subsampling layer and 2 fully connected layers. The main contribution of this work is the usage of local connection to replace the fully connection between network neurons of conventional ANNs.

AlexNet is recognized as the first deep CNN which is successfully applied onto large-scale image recognition, which is proposed by [102]. It won the 2012 ILSIRC. AlexNet has a quite similar architecture with LeNet but has more Conv layers. AlexNet has some better solutions to prevent the overfitting and gradient vanishing problem. First, AlexNet uses the ReLU [149] activation to replace Sigmoid. Second, it applies local response normalization (LRN) scheme before each ReLU to further prevent gradient vanishing effect. LRN is an another way to normalize the data for model generalization. Basically, it normalizes the response value at each receptive fields (divided by the sum of the same spatial location), which force the response value to be in a relative small range. Third, it uses overlapped pooling kernels. General max pooling applies non-overlapping kernels subsample the feature map. Overlapped pooling is just changing the kernel size larger than stride. The paper argues that overlapping kernels is helpful to prevent overfitting. Fourth, AlexNet uses Dropout to prevent overfitting. Overall, AlexNet has 5 Conv layers and 3 fully connected layers with totally 60M parameters (dominated by the three fully connection layers). This number is very large compared with recent CNNs.

ZFNet is proposed by Zeiler and Fegus [231], which won the 2013 ILSIRC. This network has very similar architecture with AlexNet but more detailed hyperparameter tunning and smaller kernels of bottom Conv layers. ZFNet has 75M parameters. The main contribution of this paper is unpooling and deconvolution, which enable the visualization of the hidden layers. Unpooling and deconvolution are quite novel and 'uncommon' combination in CNNs. The usage of these two layers can project deep feature maps to the image space, so as to deeply visualize image features highlighted. Nowadays it has become a very popular research topic [183, 25, 154, 131].

VGGNet [191] is very popular network not only in image classification but also in many other applications [217, 225, 138]. It obtains the second best results in

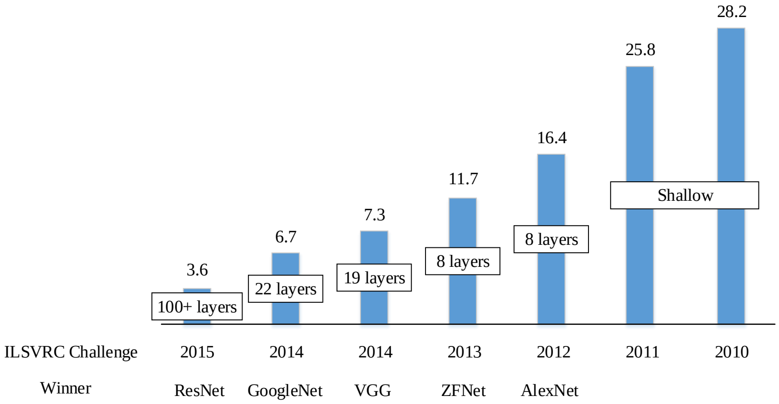

Figure 3: The ImageNet ILSVRC challenge results (top-5 error $(\%)$)) from 2010 to 2015. AlexNet achieves a very large margin improvement with deeper network than before. The number of layers is continuously increasing as the accuracy increases. ResNet conquers the barrier of training network over 100 layers, much deeper than previous winners.

ILSIRC 2014, right behind GoogleNet. Its architecture is quite neat and unique compared with GoogleNet. It has five sets of Conv units. All feature maps in each unit has the same dimension and number. Units are connected by max-pooling layers. VGGNet has a 16-layer and a 19-layer version. VGG has $140M$ parameters. Thanks to its clean and regular architecture and available pre-trained models, VGG gives researches flexibility to manipulate to internal layer representations. Thus it is the mostly widely-used architecture for high-level computer vision tasks, such as semantic segmentation and edge detection.

GoogleNet is proposed by [204]. It won the ILSIRC 2014 and largely outperforms AlexNet. Moreover, GoogleNet only has 12X fewer parameters than AlexNet yet much deeper (22 layers). The main contribution of this paper is the introduction of the *Inception module*, with the aim to better utilize the representations in network layers. The basic *Inception module* takes an input from an layer and pass to multiple different and independent layers (such as 3×3 Conv layer, 5×5 Conv layer, max pooling layers) in parallel, then the layers' output are merged together. Different kernel sizes capture multi-scale information. The inception module has been extended as four progressive versions.

Inception V2 [85] introduces batch normalization. Inception V3 [205] discusses and summarizes the design principles of the inception module in detail.

For instance, factoring 5×5 kernels to two small 3×3 kernels increases computational efficiency and training speed. Inception V4 [203] is the latest version (including multiple variants). This study braces the idea of residual networks (ResNet) [73] into their design. It also introduces the residual scaling to prevent the instabilities when number of filters exceed 1000.

GoogLeNet uses a global average pooling (averaging the value in the window) to transforms the feature maps of the last Conv layer to a vector and only one fully connected layer is used to perform prediction. Average pooling layer has been recognized as a good alternative to fully connected layer after the last Conv layer since it saves the majority of parameters coming from the fully connected layer and has intrinsic regularization effects for modal generalization. This configuration is first applied by Network in Network (NIN) [118], another interesting network architecture which builds multilayer perceptron between Conv layers to enhance the local region information abstraction. Most recent CNNs use this configuration as the classification module.

ResNet is the most successful CNN in recent two years, developed by [73, 74]. It is the first CNN that overcomes the barrier of training networks with more than 1,000 layers. ResNet won the 1st place in ImageNet classification, detection, localization, COCO detection and COCO segmentation. The idea of ResNet has been largely extended and widely used in image classification [212, 83, 82, 207, 164, 203, 181], contour detection [138], object detection [122].

We have a common understanding that deeper network can give rise to better abstraction, but when depth increases to some level, extra layers will hurt the performance. The initial motivation of ResNet is raised by a common question: why it is difficult to train very deep networks?

For examples, suppose we can train a 30 layer network, when the depth increases to $30 + 10$. The error rate will rise up [71, 196]. However, intuitively, if we setting the extra 10 layers as identity mapping, i.e., passing the same output of 30-th layer the next 10 layers. The error rate should be the same. However, this simple identity mapping operation seems difficult for CNN to learn directly. So the author argues that the difficult of training very deep network is due to optimization issue but not architecture itself. To overcome this difficult, ResNet suggests to instead allow CNNs to learn residual mapping, which is expected to be easier than learning identity mapping.

Let's define the ideal underline mapping as $\mathcal{H}(x)$. ResNet lets the computational units (convolutions) model the mapping of $\mathcal{F}(x) = \mathcal{H}(x) - x$. Therefore, the targeting mapping becomes $\mathcal{F}(x) + x$. This formulation is converted to a novel

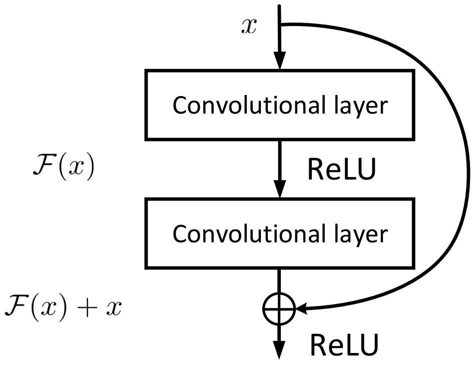

Figure 4: The illustrate of identity mapping in ResNet. The figure shows an residual unit including an identity mapping and two convolutional layers.

architecture computational unit with skip-connection, as illustrated in Figure 4. This unit is called "residual unit". The overall network is constructed by stacking such computational unit. The concept of skip-connection in neural networks stems from [163] and Highway Network [196], which acts like a gate function to selectively allow the information pass to the following layers.

The general form of the residual unit is defined as

$$\begin{aligned} \boldsymbol{y}_l &= \mathcal{F}(\boldsymbol{x}_l) + h(\boldsymbol{x}_l) \\ \boldsymbol{x}_{l+1} &= f(\boldsymbol{y}_l) \end{aligned} \tag{8}$$

where h is identity mapping and f is ReLU in the original ResNet architecture [73]. $\mathcal{F}_l$ is composed by a set of Conv units associated with batch normalization, activations and optionally Dropout. As can be observed, x_l is not completely identity mapping but projected by ReLU after addition. A follow-up paper [74] proposes 'pre-activation' to allow complete gradient backpropogation during training (explained as follows). 'pre-activation' makes the ReLU inside $\mathcal{F}$, which results in the new residual unit definition:

$$\boldsymbol{x}_{l+1} = \mathcal{F}(\boldsymbol{x}_l) + \boldsymbol{x}_l \tag{9}$$

This new modification actually is a simple solution to the long-term gradient vanishing issue in CNN training. Specifically, in l-th of L residual units of ResNet, the forward output y_l and the gradient of the loss $\mathcal{L}$ w.r.t y_l is defined as

$$\boldsymbol{y}_l = \mathcal{F}_l(\boldsymbol{y}_{l-1}) + \boldsymbol{y}_{l-1} \tag{10}$$

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{y}_l} = \frac{\partial \mathcal{L}}{\partial \boldsymbol{y}_L}(1 + \frac{\partial}{\partial \boldsymbol{y}_l}\sum_{i=l}^{L-1}\mathcal{F}(\boldsymbol{y}_i)) \tag{11}$$

Thanks to the addition scheme, the information from prior layers (i.e., $\boldsymbol{y}_{l-1}$ in forward and $\frac{\partial \mathcal{L}}{\partial \boldsymbol{y}_L}$ in backward) can flow directly to previous layers without passing to any weight layer. Since the weight layers can vanish the gradient, this property is able to deal with the gradient vanishing effects when training the depth of the network increases [157, 74].

This simple solution provides subsequent advantages. Several follow-up studies [74, 212, 83] gradually reveal them as discussed in the following. An remarkable paper worth to mention is stochastic depth network [83] built on ResNet. Since training deep network suffers from overfitting and very deep network is usually difficult and inefficient to train. Stochastic depth network trains network with random depth during training stage. Since in each residual network the data from bottom has two paths: $\mathcal{F}(\boldsymbol{x})$ and $\boldsymbol{x}$, if the data does not pass some $\mathcal{F}(\boldsymbol{x})$, the actual network depth decreases by some ratios. So during the training, the idea is to randomly block some $\mathcal{F}(\boldsymbol{x})$ in every mini-batch forward with probability p (i.e. output zeros). Each residual unit could have individual p (named survival rate). The method has a comprehensive discussion of the settings of p. While during testing, all $\mathcal{F}(\boldsymbol{x})$ functions are applied (scaled by $1-p$). Intuitively, this design braces the idea of Dropout and ensemble learning, which significantly improves the generalization of ResNet.

[212] argues that ResNet is actually the ensembles of exponential number of relative shallower networks (also mentioned by [83]). As mentioned in the last paragraph, each residual unit has two paths to allow data flow to next layers. Suppose we have n such residual units, totally there are 2^n number of paths, yielding 2^n plain networks with different depths. This paper has experimentally verified its argument. Moreover, it has shown the independence between residual units, in other word, removing some residual units does not influence the results substantially. However, removing a single layer from VGG will cause a huge issue. Swapout [192] pushes the ensemble into an extreme, by combining ResNet, the stochastic depth network, and Dropout.

DenseNet [82] replaces the addition of residual unit with concatenation to allow dense connection between layers, which results in better better feature usage. This strategy implicitly uses multi-layer representations to increase the performance. Wide ResNet [230] introduces a widen factor and shows that increasing width of layers rather than only depth gives rise to better performance, higher training efficiency, and less memory usage. ResNet of ResNet (RoR) adds an-

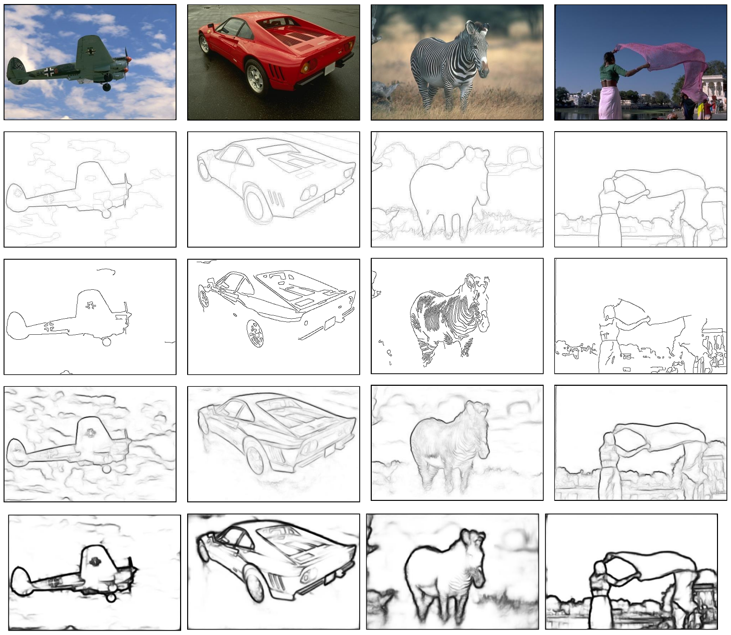

Figure 5: The visualization of contour detection results on the BSDS500 dataset. The first and second rows show the input image and groundtruth. The third, fourth, fifth rows show the results of basic Canny detector (1989) [21], SE (2013) [44], and HED (2015) [217] detectors, respectively. SE and HED are learning based methods. As can be observed, current methods can generate more clear edge maps and be aware of object contours and internal or background edges.

other level of identity mapping and shows better performance. ResNeXt [216] introduces a cardinality factor inside ResNet by repeating homogeneous residual transformation inside a residual unit.

Generalization The addition operation is effective in practice for general network training. One main reason is because the addition operation of skip-connection can intrinsically ensemble outputs of modules during forward and equally split the gradients to two paths and may merge them later on during backward. This 'averaging' behavior can stabilize the gradients. We have observed various kinds of applications that take benefits from this skip-connection mechanism [127, 158, 93]. Skip-connection encourages the multi-scale feature fusion to prevent small information loss and makes the network training for efficiency due to better gradient backpropagation. Both characteristics are favorable to medical images. [47] specifically discusses the importance of skip-connection in medical image analysis.

Table 1: Comparison of edge detection on the BSDS500 dataset with standard evaluation metrics[141], including F-measure, precision/recall (PR) curves, and average precision (AP). The F-measure score is reported at fixed optimal threshold (ODS) and per-image threshold (OIS). AP is the area under the PR curve. Human annotator accuracy is shown in the first block. The second block shows several early-state unsupervised methods. The third blocks shows conventional supervised methods. The last block shows recent CNN based methods. As can be observed, CNN based methods improve previous approaches by a large margin.

	ODS	OIS	AP
Human	.80	.80	-
Canny [21]	.60	.64	.58
Felz-Hutt [53]	.61	.64	.56
Normalized Cuts [38]	.64	.68	.48
Mean Shift [36]	.64	.68	.56
ISCRA [169]	.72	.75	.46
gPb-owt-ucm [5]	.73	.76	.70
Sketch Tokens [114]	.73	.75	.78
SE [44]	.74	.76	.78
SE-Var [45]	.75	.77	.80
MCG [6]	.75	.78	.76
SE-u [112]	.72	.75	.76
PMI [86]	.74	.77	.78
SemiContour [234]	.73	.75	.78
DeepNet [96]	.74	.76	.76
N^4-field [56]	.75	.77	.78
DeepEdge [12]	.75	.77	.81
DeepContour [186]	.76	.77	.80
CSCNN [84]	.76	.78	.80
HFL [14]	.77	.79	.80
HED [217]	.79	.81	.84
HED-u[112]	.73	.75	.76
CEDN [225]	.79	.80	.82
RDS [123]	.79	.81	.82
COB [138]	.79	.82	.85
PixelNet [8]	.80	.81	.83
RCF [124]	.81	.83	-

5. Image Contour Detection

In this section, we first briefly review the history literature of edge/contour detection. Then, we introduce the pioneer work of CNN based contour detection methods. Next, we introduce an important end-to-end CNN, supporting present state-of-the-art contour detection methods, and the details of some other breakthrough contour detection methods. Finally, we highlight the shortcomings of existing methods. Figure 5 qualitatively compares the contour detection results of several strongly foundational contour detection methods on the image contour detection and segmentation benchmark (Berkeley segmentation dataset), BSDS500 [5]. Better edge detection methods have stronger ability to detect object contours while ignoring internal edges inside objects or background. Some advanced methods are widely used in medical images but some are barely used. We discuss the strengthens and weakness and clarity their favors to medical images.

In Table 1, we compare the contour detection performance of several remarkable contour detection methods under standard evaluation metrics. We discuss the compared methods in the following.

5.1. Historical overview of contour detection

There is a very long and rich history of literature for edge detection [21, 99, 161, 140, 137, 156, 190, 236]. The early-stage Canny detector [21] computes the local brightness discontinuities and produces continuous edges. Estimating the local changes using global and local cues with multiscales is critical of many following edge detection methods. There are a variety of directions for contour detection using local oriented filters [147, 160, 54], spectral clustering [38, 208, 5, 136], sparse reconstruction [132, 215], supervised learning [43, 132], and so on [128, 36, 142]. We refer readers to [156, 136] for detailed categorization.

There are several remarkable edge detection methods still used by recent state of the arts. Global probability of boundary (gPb) detector [5, 136], which is developed by Arbelaez and Malik, computes local orientated gradient features (brightness, color, and texture) based on [141] and the multiscaling strategy [166], and then computes global spectral partitions with normalized cut to achieve the globally oriented contour information [133]. This globalization mechanism differs from the earlier Pb detector [141]. This method produces quite clear and effective edges than previous methods. Various methods based on gPb is developed [215, 135, 6, 94, 229] for contour detection and segmentation with the focus on both efficiency and accuracy. More importantly, [5] also presents an edge based segmentation method called oriented watershed and ultrametric contour

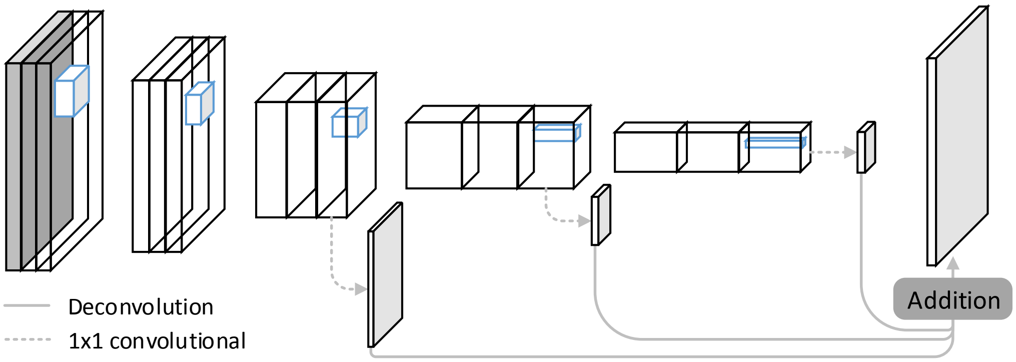

Figure 6: The architecture of FCN. The main part of the CNN architecture adopts the VGG architecture. FCN uses side outputs to extract intermediate layer feature maps. The 1×1 convolution is used to generate multi-class segmentation masks. All predictions are unsampled to the image space and added as the final result.

map (OWT-UCM) (first published in [4]). The overall method for both contour detection and segmentation is well recognized as gPb-owt-ucm in following literature. The OWT-UCM technique and its multi-scaling improved version, Multi-scale combinatorial grouping (MCG), are still used by latest CNN based contour detection method to generate object proposals or computing thin edges. However, gPb is very inefficient for practical usage. After that, the study of edges tends to learning based [43], which offers much efficiency and accuracy gains. A critical method is the Structured Edge (SE) detector developed by [44, 45], which is an excellent improvement of SketchToken [114]. SE can be recognized as the most successfully edge detector using random forest with structured outputs [100]. The main idea is to use structured learning (i.e.structured random forests [100]) to densely predict inherent structures of patches, such as straight lines, parallel lines, curves, T-junctions, Y-junctions, etc. The patch with inherent edge structures is called sketch token in the community [168, 114]. SE is very efficient (60 FPS) compared with gPb-owt-ucm 1/240 FPS and leads the performance of the contour detection field for quite a while, before it is largely surpassed by the introduction of CNN. Many variants of SRF are proposed after then for image edge detection and segmentation [148, 6, 211, 210, 186, 45, 234], and it is also popular in medical images [121].

5.2. Pioneer CNN-based contour detection

The performance of edge detection increases significantly in recent two years. We outline the discussed CNN based methods in this section as pioneer work because these methods mostly adopt conventional CNN with the inspirations from previous edge detection methods to build the conception of edge patterns and structures, multi-scaling, and so on. Following the convention neutral network, the CNN based edge/contour detection takes a local patch around a centering pixel as input and predicts label indicting whether the centering pixel is edge or non-edge.

[180] propose multiple networks' outputs to perform segmentation and edges. [84] (denoted as CSCNN) use a CNN as feature extractors and train a SVM to classify edges. [56] propose N^4 field, i.e., using neural networks and nearest neighbor search to retrieve the best matching edge pattern of local patches. [96], instead of using CNNs, extracts feature using unsupervised generative models (RBM and DBN) and train classifiers to predict edges. [13] propose DeepEdge, which use multi-scale features from multiple layers and separate two task branches with two losses, one is called classification branch which learns to predict the edge likelihood (with a classification objective) whereas the other regression branch is trained to learn the fraction of human labelers agreeing about the edge presence at a given pixel. The outputs are combined to predict the edges.

Later on, [186] propose DeepContour to extract visual feature of local patches and use the extracted features as additional features to the used features by the SE detector. The local patches can be categorized into a limited number edge patterns (tokens). Accurately recognizing these patterns is critical for better contour detection. DeepContour uses this property to supervise CNN training to generate rich and discriminative features, then it trains structured random forests to predict edges. This step can be viewed as an edge refinement process. DeepContour achieved leading performance than previous methods.

The efficiency of contour detection is priority. The above methods still suffer from the heavy dense prediction computations, making the edge prediction particularly slower than the SE detector.

5.3. Fully convolutional network (FCN)

The introduction of FCN [126, 185] changes the standard of using CNN for (pixel-wise) dense prediction, making the real-time prediction becomes possible and as well largely improved performance, hence it benefits later contour detection methods.

The technique of FCN is actually straightforward and simple. FCN allows the network directly outputs a segmentation mask having the same dimension of the input. Suppose the input RGB image has dimension $3 \times H \times W$. The output of image will have the size $C \times H \times W$ for C class semantic segmentation. In other word, each spatial location of the segmentation mask predicts the probability of its semantic label. Figure 6 illustrates the FCN architecture.

FCN uses VGG as the basic network architecture. We have discussed the details of VGG previously. It contains 5 convolutional sets, with 5 2×2 max pooling in between, which totally resizes the original input by 2^5. So the feature map dimension of the last convolutional set is $512 \times \lfloor \frac{H}{32} \rfloor \times \lfloor \frac{W}{32} \rfloor$. These feature maps keep coarse spatial location where each spatial location stands for 32×32 region in the original image.

Instead of using the fully connected layer after the last convolution layer as CNNs for image classification, FCN directly applies an 1×1 convolutional to transform the 512-dimensional vector of each spatial location to the label space, i.e., C-dimensional in this example, as the semantic label probability distribution. Next, FCN adds an deconvolution (used as an upsampling operation) layer to enlarge feature maps from $C \times \lfloor \frac{H}{32} \rfloor \times \lfloor \frac{W}{32} \rfloor$ to $C \times H \times W$, resulting a pixel-wise segmentation mask. It is worth to mention that the terminology of deconvolution used here is debatable, because it is not the conventional deconvolution operation [232]. The one used here is actually a 'backward convolution' operation, i.e., each spatial pixel performs element-wise product with all the weights of the kernel and expand the predictions (one for each weight value) as an image extent. Figure 7 explains this 'deconvolution' operations. The weights of the deconvolutional layer can be initialized as a bilinear interpolation kernel and allow this deconvolution to act upsampling behavior. We name it upsampling in this paper because we will introduce another strategy (i.e, unpooling plus deconvolution) next to achieve structured outputs (see bottom of Figure 7). Training the network is straightforward. In the loss layer, every pixel contributes to the loss, all spatial locations are summed. This variation does not break direct backpropagation. This training strategy, i.e., inputting an image and outputting a pixel-wise prediction map, is termed as end-to-end training.

This simple approach achieves surprising good results compared with previous methods and the prediction is very efficient (less 1 second for an 500×500 image) because the network does not need computational expensive fully connected layers and only once forward is needed to obtain the final results. To generate more precise and robust prediction, FCN also proposes to build side outputs to make use of the feature maps from multiple convolutional layers. Since convolutional layers have different feature map dimensions, this approach in nature utilizes multi-scale information. FCN-32s uses the one side-output to predict the segmentation. FCN-

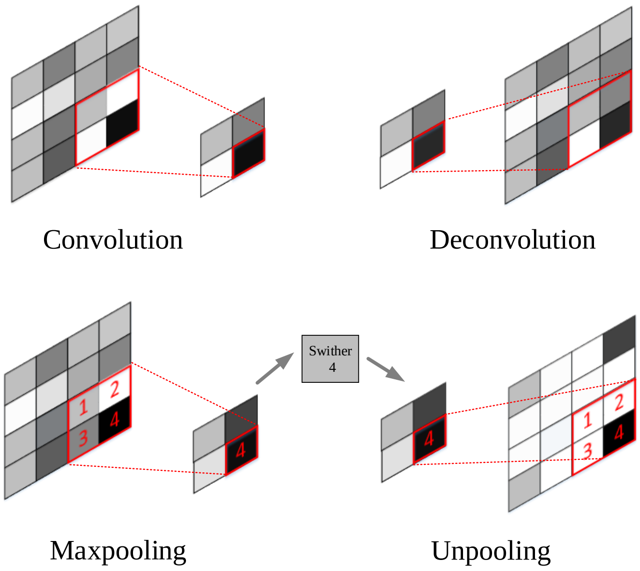

Figure 7: The illustrate of convolution and deconvolution operation and maxpooling and unpooling layer. The kernel size is 2 × 2. Convolution operation averages the values of a 2 × 2 region. Deconvolution computes the values of the region value given one value. Maxpooling selects the maximum response in each region. Unpooling uses the switcher information (saved the selected location of the region in the maxpooling layer) to fill a value in the region, yielding a sparse feature map.

8s uses three side-outputs and merge as the final results as shown in Figure 6.

This idea becomes the foundation of CNN based image segmentation and contour detection crossing various image domains. Wide improvements have been proposed in recent years [162, 155, 235, 117, 116, 70, 184, 31]

5.4. Holistically-nested edge (HED)

HED is the first method successfully applied FCN into contour detection. It significantly improved previous methods. The main contribution of this paper is the usage of FCN with side outputs and deep supervision [106]. Actually, the concept of side output uses extra branches from intermediate layers to encourage multi-scale feature reuse. Differently, the side outputs of HED are not directly combined together to output a single mask (edge map here) as FCN does or discrete label as DeepEdge does. Each side output is directly used to predict the edge map. Since side-outputs connecting intermediate layers have different feature map dimensions, deconvolution layer (upsampling layer) is used to resize the feature map to the original input image size. Each mapped output is passed to a Sigmoid cross-entropy loss to perform pixel-wise binary prediction. The loss of the network is define as

$$\mathcal{L}_{side}(\boldsymbol{W}, \boldsymbol{w}) = \sum_{i=1}^{S} \alpha_i \mathcal{L}_{side}^{i}(\boldsymbol{W}, \boldsymbol{w}^{(i)}), \tag{12}$$

where $\mathcal{L}_{side}^{i}$ is the loss from i-th side output. $\boldsymbol{W}$ is the core network parameters. α_i is the loss weight. S is the number of slide outputs. Sine HED uses VGG16 as the base network architecture. $\boldsymbol{W}$ refers to the parameters of its 5 convolutional units. $\boldsymbol{w}^{(i)}$ is the parameters of i-th side output. It is actually a 1×1 convolutional to map the feature map to the label space (edge or non-edge).

Since edge pixel has much small portion than non-edge pixel in one image, the loss is imbalanced per image. From machine learning perspective, unbalanced training data is not undesirable for model optimization. HED introduces weight-balanced loss, defined as follows:

$$\begin{aligned} \mathcal{L}_{side}^{(i)}(\boldsymbol{W}, \boldsymbol{w}) = -\beta \sum_{y \in Y_+} logPr(y_j = 1 | X; \boldsymbol{W}, \boldsymbol{w}^{(i)}) \\ -(1-\beta) \sum_{y \in Y_-} logPr(y_j = 0 | X; \boldsymbol{W}, \boldsymbol{w}^{(i)}), \end{aligned} \tag{13}$$

where $\beta = |Y_-|/|Y|$ and $1 - \beta = |Y_+|/|Y|$. $|Y_+|$ denotes the number of pixels belong to edges according to groundtruth edge map Y. β is the weight balancing coefficient. Besides individual loss, all slide outputs are then fused together and generate a fused output associated with a loss:

$$\mathcal{L}_{fuse} = Distance(Y, \sum_{i}^{S} \gamma_i \hat{Y}^{(i)}), \tag{14}$$

where $\hat{Y}^{(i)}$ is the predicted edge map by i-th side output. and γ_i is the learnable weights. $Distance$ is also a Sigmoid cross-entropy loss to measure the pixel-wise prediction error.

The training of this method is relatively light. It uses pre-trained VGG-16 as initialization and train with a few thousands iteration to obtain the results. Note that BSDS500 only has 200 training image. HED sample 100 test images and test on the rest 100 images. In addition, the paper also discuss the inconsistence of groundtruth. The outputs at deeper layers are coarse which will in nature ignore detailed edges (such as background and object internal edges). To prevent very fine annotations (containing many 'noisy' edges) of groundtruth from affecting the convergence of supervision of deeper layers, the paper treats a pixel as edges only it is labeled as edge in at least three annotations.

Side-output with deep supervision is fairly effective combination to boost the performance of dense prediction tasks, because it maximizes the reuse of rich hierarchical representations at different layers with different scale. The multi-scaling property is also a well-known approach to improve the contour detection accuracy. Most following work lies on the feature map re-usage to push the performance. For example, [97] also trains multi-scale HED based on multi-scale inputs to further boost BSDS500 accuracy. It has outperformed the empirical accuracy of human annotator (.80 ODS). RCF [124] generalizes HED by using richer features from all convolutional layers of a CNN, pushing the performance to .81ODS.

HFL [14] uses object-level features to accomplish low-level edge prediction, because the human vision system uses object-level reasoning to locate edge points. Specifically, it extracts object-level deep features from multiple layers of VGG-16 and use a MLP to classify edges. This method can be viewed as a special way to use rich features from a pre-trained CNN. More interestingly, HFL extends its network to the application of semantic boundary labeling and semantic segmentation to show that low-level boundaries have positive effects to high-level vision tasks.

Another analogous method is PixelNet [8], which highlights the usage of all feature maps and uses a specially designed predictor (a MLP) for coherent semantic segmentation and contour detection. PixelNet also discusses the sampling of predicted pixels and mini-batch to reduce the unbalance of edge and non-edge pixel ratio and the memory consumption.

In addition, HED has been applied and improved for different tasks. [187] uses HED to extract object skeleton. [112] even train HED in an unsupervised manner base don video optical flow by iteratively refining the model. Beyond the natural image domain, HED is widely welcomed in medical image domain because of its efficiency and multi-scaling scheme to handle resolution and scale problems ubiquitous in medical images. We will discuss them in the following section. There also several notable papers generalize HED [224, 20, 209, 224] for various computer vision tasks in both natural image and medical image domains.

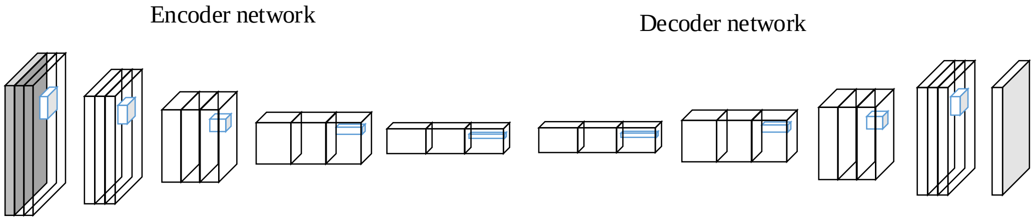

Figure 8: The architecture of DeconvNet. DeconvNet includes an encoder to encode an input image to a set of feature maps and the decoder to decode the encoding to a segmentation mask through a set of unpooling layers (illustrated in Figure 7) and convolutional layers.

5.5. *Encoder-Decoder network*

Another noticeable network for contour detection is called Encoder-Decoder network [225]. This network has very similar architecture with DeconvNet [151] for semantic segmentation. The difference is that the encoder and decoder architectures are asymmetric. The encoder part is identical to DeconvNet borrowed from VGG-16, but the decoder part has light computational units. Instead of using deconvolution layer, every unpooling layer is followed by a convolutional operation. Figure 8 illustrates the network architecture. This network is a successful application of encoder-decoder style CNN architectures on contour detection.

Different from methods using side outputs or deep supervision, this method directly uses the last convolutional layer of VGG-16 (this encoder part is fixed during training) and uses a set of unpooling and convolutional to generate unsampled prediction map. The feature map dimension of the last convolutional layer is 32 times smaller than the input image. Small scale coarse prediction ignores the detailed information such as short and weak edges, thus the generated contour map ignores the background and object internal edges but only retains occlusion boundaries [200]. So this method achieves significantly improved performance than HED (57.0 ODS vs. 44.0 ODS) on the PASCAL val2012 dataset. Note that this dataset contains only groundtruth of object contours. However, when applied onto BSDS500 dataset where groundtruth contains fine edge annotations, this method performs slightly worser than HED.

Compared with HED, this network is better capture overall object contours using content information rather than small edges due to gradient changes. This property is suitable to detect contours of medical image objects, such as organs.

5.6. Oriented contour detection

Using the global orientation to predict edges has been well studied previously [188, 5]. Affinity matrices well capture pairwise local intensity changes by using a global graph embedding and the generated graph partitions preserve strong edge information where each map highlights edges with one particular orientation. There are a few work considers the orientation information inside the CNN training.

Affinity CNN [134] directly trains a network to output the affinity matrix. To achieve that, this method trains 48 predictors (composed by convolutional layers) to predict the affinity of each pixel to its 8 neighbors at 3 different scales (distances of 1, 4, and 16 pixels). There are two losses functions supervised by pre-computed affinity matrix and image edge groundtruth respectively.

Another remarkable work is called Convolutional Oriented Boundaries (COB) [138]. Firstly, COB generalizes the side output and deep supervision scheme of HED to obtain fine and coarse contour maps. To estimate the contour orientation, COB connects multiple small sub-networks to predict oriented edge maps for each orientation bin. Each subnetwork is access to all side outputs with different scales. To decide the final orientation of each pixel, COB computes the max response of sub-network outputs respecting to different bins and may average the two orientations if both oriented sub-networks have high responses. COB uses a strong 50-layer ResNet as the basic network while most other methods we discussed use VGG-16. COB shows that using ResNet improves the performance by a quite large margin. It demonstrates the important role of the basic CNN for high-level applications.

Using orientation information is useful for medical images. For example, in Lung X-ray image diagnosis, healthy images contain clear rib cage contour [41]. In this way, we can predict contours with semantic information to help diagnosis.

5.7. Weakly-supervised, semi-supervised and unsupervised edge detection

Obtaining the annotated contour detection dataset is very labor expensive. At present, only the BSDS500 dataset has fine edge annotations. Dense prediction tasks can obtain many pixel-wise label from a single image to optimize the parameters of CNNs, however we can still observe that the size of training data is determinate [225, 138]. There are some yet limited work that studied how to introduce unlabeled data to train or improve an edge detector [112, 234, 123, 92]. However, we only witnessed growing related literatures for semantic segmentation [79, 80, 155, 40].

[112] propose an unsupervised learning method of edges, which utilizes the property of discontinuity around the edge pixels of motions to generate weak annotations from a large video dataset. This method uses a quite sophisticated process to generate clean annotations based on motion estimation techniques including motion estimation, motion contour detection, and edge alignment. Using motion cues for contour detection is proposed by [200]. Then, it treats the motion estimation and contour detection as an iterative optimization process. At each iteration, the estimated edges are used as supervision to train an edge detector. The trained edge detector is used to generate edge maps for better motion estimation for next iteration. The paper conducts experiments using both SE and HED as the base edge detector, showing than this kind of coarse-to-fine interaction training strategy does improve the performance of CNNs and structured random forest classifiers. However, as the author stated, the generated annotations have noises and have unable to generate fine edges, so it can hardly obtain the same or even outperform the edge detector trained with strong supervision with human annotations although more images are available. Besides unsupervised training, [234] propose a semi-supervised structured ensemble learning method, which is built on SE, to train an edge detector with only 3 labeled images and outperforms this unsupervised method (see Table 1). However, this semi-supervised learning method can not be used for CNNs. There is no work study about semi-supervised CNNs for contour detection. [92] generate object contour supervision under multi-level supervision and test the SE and HED detectors.

The coarse-to-fine CNN training paradigm for contour detection is embedded into intermediate layers of a single CNN by [123]. This supervision is denoted as relaxed deep supervision (RDS), which is used to improve a pre-trained HED with a large set of coarse edge annotations. The motivation behind is that RDS relaxes the human annotations (edge and non-edge points) to get more relaxed labels, which is used to adapt to the diversities of intermediate layers. Relaxing labels is produced by Canny, SE or HED detectors. The benefit of RDS is that it processes the false positives using a "delayed strategy" to allow more discriminative layers (deeper layers) handle difficult points and leave these difficult points ignored in early layers. This is a validate way to achieve better network convergence.

Using less annotated data for training CNN is obviously essential for medical image analysis. Unfortunately, this area is waiting to be explored.

Table 2: Overview of the literature using deep learning for various kinds of medical image segmentation that are contour aware. Most methods use end-to-end CNNs (FCN or HED) scheme to perform pixel-wise classification. See text for more detailed explanations.

Reference	Task	Method
[34]	Neuronal membrane segmentation	Standard CNNs with patch-wise pixel classification.
[56]	Retinal vessel segmentation	Using CNN features to model edge patterns and apply nearest search to detect edges.
[171]	Biomedical image segmentation	A novel end-to-end CNN (Unet) for cell segmentation with special designs for contour preserving.
[194]	Cervical cell segmentation	Extracting CNN and different kinds of features to learn to localize object contours.
[198]	Cell segmentation	An autoencoder based method to recover broken edges between touching or overlapping cells.
[28]	Gland segmentation	An end-to-end CNN uses multi-layer feature maps to perform contour detection and segmentation.
[29]	Neuronal membrane segmentation	An deeply supervised CNN to capture contextual information.
[139]	Retinal image segmentation	Using mixed multi-layer feature maps to segment vessels and optic disc.
[224]	Gland segmentation	Combining FCN and HED to conduct segmentation and edge detection together.
[172]	Pancreas segmentation	Spatial aggregation with random forest to combine edge and interior CNN cues of organs.
[19]	Pancreas segmentation	Aggregating organ contour detection and segmentation results of HED and FCN through conditional random fields.
[46]	Liver segmentation	A 3D deep supervised FCN with an extra contour refinement process.
[30]	Biomedical image segmentation	Using RNNs to model the intra-slice and inter-slice context represented by FCN to separate touching objects.
[29]		
[33]	Volumetric Segmentation	The 3D version of Unet.
[145]	Volumetric Segmentation	VNet with new dice score driven loss function to trade-off imbalance of labels.
[174]	MRI segmentation	Using CNNs to predict evolution of active contours for contour localization.
[144]	Ultrasound and MRI segmentation	A hough-voting CNN has robust contour extraction of anatomy.
[227]	Ultrasound segmentation	Using recurrent memory networks for contour completion.
[146]	Cardiac MRI	Combing a dynamical system and a CNN to model the contours of objects.
[153]	Multi-modal cardiac	Incorporating prior anatomical knowledge (e.g. boundaries and shape) into network training.

6. Medical Image Contour Detection and Segmentation with CNNs

Deep learning has became the mainstream of medical image contour detection and segmentation methods [89, 222, 67, 189, 173, 10, 153, 18]. In this section, we review recent CNN based methods for medical image contour detection and segmentation. Most reviewed paper aims at segmentation. We focus on the methods that use the abovementioned methods or are designed to be contour aware to achieve more accurate segmentation. Table 2 summarizes the reviewed methods. Detailed are discussed in the following.

[34] adopt the standard CNN as a patch-wise pixel classifier to segment the neuronal membranes (EM) of electron microscopy images. This study is a pioneer work of using CNN for medical image segmentation. It won for ISBI 2012 EM image segmentation challenge and significantly outperforms other competing methods. In [52], a CNN with an architecture specifically optimized for EM image segmentation is presented. Compared to the original CNN, smaller receptive field in the upper layers and deeper architecture are employed. Such optimized design enables the network to learn better features from the local context with increased non-linearity. Significant improvement in performance is validated on the ISBI 2012 challenge [7]. [194] propose a segmentation system for cervical cytoplasm and nuclei in which pixel-wise classification is obtained by multiple convolutional networks trained for images at different scales. Then they use various kinds of features to learn to localize object contours and split the touching objects.

The majority of contour detection and segmentation methods follows the structure of the FCN [126] and HED [217] networks. [171] propose U-net, an end-to-end CNN that can take advantage of information from different layers. To handle touching objects, a weighted loss is introduced to penalize the errors around the boundary margin between objects. The proposed U-net achieved the best performance on ISBI 2012 EM challenge dataset [7]. The state-of-the-art segmentation performance on the EM dataset is achieved by a new deep contextual network proposed in [29]. The deep contextual network adopts an architecture that is similar to HED. The difference is that the final segmentation result is a combined version of the segmentation results derived from different layers through an auxiliary classification layer. In the forward propagation, such design can more efficiently exploit the contextual information from different layers for edge detection. In return, the lower layers can be deeply supervised through these auxiliary classification layers. This is because the classification layers provide a short cut between the lower layers and final segmentation error. [28, 27] propose a deep contour-aware network for gland image segmentation. This method uses side outputs as multi-tasking deep supervision. The detected contour map is merged with the segmented binary mask to prevent touching of glands, which is a special treatment to cell contours. This method won the 2015 MICCAI Gland Segmentation Challenge [193]. In the following, [224] propose a multichannel side supervision CNN for gland segmentation. This network can be treated as a combination of HED and FCN for simultaneous segmentation and contour detection. Similarly, [150] propose a lymph node cluster segmentation algorithm based on HED, FCN and structured optimization to address the contour appearances. [19] propose a data fusion step using CRF to adaptively consider the segmentation mask generated by FCN and the contour map generated by HED for pancreas segmentation. [172] propose to use random forest based spatial aggregation to integrate semantic mid-level cues of deeply-learned organ interior and boundary maps to segment pancreas with HED.[139] explores the combination of multi-layers' feature maps to perform multi-task learning on vessel segmentation of retinal images. [153] proposes to incorporate anatomical priors on anatomy (e.g. shape and label) structure into CNNs, so as to make the predictions anatomically meaningful, especially for the case when input images have missing boundaries.

CNN based methods for 3D medical image segmentation have been attracting attentions in recent two years. Most existing methods are extensions of known 2D CNNs. [46] propose a 3D deeply supervised network for Liver segmentation. It can be viewed as a 3D extension of HED. Moreover, it uses a fully connected CRF to refine the object contours. 3D Unet [33] is proposed by the same group with U-net for 3D volumetric segmentation. [145] propose V-Net, which contains a new loss function based on Dice coefficient to resolve the strong imbalance between foreground and background. It uses the skip-connection strategy of Unet to prevent the detail information loss which will affect fine contour prediction.

Besides the direct application of end-to-end CNNs for pixel-wise classification, there are a number of interesting studies exploring the usage of CNNs or RNNs to achieve better context information modeling (e.g. contour completion). [198] propose to use stacked denoising autoencoder to restore the broken cell boundaries for cellular segmentation in brain tumor and lung cancer pathology images. Similar to [198], [90] propose a breast density segmentation method based on a multi-scale CNN that is trained in an unsupervised way in which an autoencoder is trained. During the unsupervised training, image patches and their corresponding segmentation masks are randomly cropped and the network is trained to reconstruct the segmentation masks. The unsupervisedly learned model is used to extract features for pixel level classification. Therefore, the different components are delineated. [227] use RNNs to achieve completion for ultrasound images where the contours are unclear and broken. The RNN it used is called bidirectional Long Short-Term Memory (BiLSTM) networks, which is able to leverage past and future information to make prediction. [30] use RNN model and propagate the contextual information of the third dimension of the 2D image planes. A 2D CNN (i.e. U-net) extracts the hierarchy of contexts from 2D images and pass the information to RNN to leverage the inter-slice correlation for 3D segmentation. Their good results on fungus images, which contain very weak boundaries between objects, demonstrate its ability of be aware of object contours. [174] trains class-specific convolutional neural network to predict the evolution of active contours as a vector field. This method is a new way to formulate the structured output of CNNs. [144] propose hough-CNN, a CNN with voting mechanism along the contours of objects to localize anatomy centroid. A recent work [146] combines CNNs with dynamic system theory for Cardiac organ contour detection. This method takes advantage of an important concept in dynamical system, i.e., limit cycle, to represent the contours of the target object. Instead of classifying pixels into label classes, they propose to predict a vector for each pixel and thus a vector field is formed for an entire image. Based on the vector field, the organ contour is detected through dynamic theory in which a limit cycle is detected as the finally detected contour. The method needs very limited training data to train the model.

7. Discussion

We have discussed the state-of-the-art image edge or contour detection methods in the computer vision community and we review their applications in the medical image domain. Based on the discussed methods above, it can be observed that the usage of CNNs in medical image contour detection and segmentation is relatively crowded into a narrow line. Most work leverages on end-to-tend CNNs for direct dense prediction. Extension towards to wide and new perspectives to solve specific problems would be necessary for CNN development in medical image analysis, but only a few literature exists. We discuss some interesting topics and outline potential directions.

1) Multi-scaling Medical images intrinsically contain rich multi-scale information such as the nucleoli and tumor regions in microscopic images. Using fine features without much spatial information loss is important for contour detection. From the popularity of HED in medical images, sufficient usage of the layers' feature maps is as always promising. Moreover, the aggregation of HED is a way of model resembling [97]. Ensembling offers a multi-scaling and averaging mechanism, which is important to generate smooth contours. Detailed explorations of state-of-the-art CNNs, for example, ResNet [73] and DenseNet [82], are necessary, which have skip-connection to strengthen feature map usage. Appropriate usage of skip-connection to build very deep FCN is studied by [47]. RCF [124] shows an extreme of using features from convolutional layers. Most studies enable multi-scaling with side outputs and optional deep supervision. From our experience, the supervision at shallow layers usually has large losses which is very difficult to overfit even on training data. Large losses will result in large gradients and thereby disturb the error backpropagation of deeper layers. This is also discussed by [123]. We think there should be more careful studies on the consideration of effective deep supervision mechanism.

2) Transfer learning Transfer learning is gaining popularity in the medical imaging domain. Several literature [206, 189, 67] have shown that fine-tuning CNNs trained on natural image datasets helps improve the performance. Designing highly effective network architectures needs rich experience, but borrowing or modifying existing architectures alleviates the pains. Transfer learning addresses the insufficient dataset problem in the medical domain. In fact, dense prediction tasks with end-to-end CNNs can implicitly gather many training data (one for each pixel). That is one of the reasons that some large architectures like U-net [171] can be trained from scratch using only 30 images. However, we believe transfer learning will give further improvement, with careful consideration of the usage of shallower layers and deeper layers. Shallower layers capture fine edge information which are shared between natural images and medical image, while the deeper layers capture content information which are completely difficult. Therefore, the appropriate usage of feature maps of earlier layers is important (which also recalls the problem of multi-scaling), for example, extra links to combine shallower layers and deeper layers [145].

3) Discontinuity of broken edges Although the CNN is powerful to detect edges. The severe and common touching and overlapping phenomenon between objects in medical images are still challenging. One common way is to deploy a remedy process on CNN outputs as discussed above, by integrating conventional methods [121] or extra deep learning models attempting to reconstruct a better contour map [198]. RNN has shown the potential to achieve contour completion [227] with its ability to capture long-term dependence of inputs. However, the semantic information of edges is very limited. To enrich the features of edge as the input of RNN, better usage of feature maps would be helpful. In addition, instead of conventional RNNs, advanced memory networks [199, 66] could be useful to handle contour reasoning. In addition, using shape priors [219] is popular to main the structure of objects in medical image segmentation because organs or cells usually have similar shapes. [26, 51] use deep Boltzmann machines to model hierarchical structures to constrain the evolution of the shape-driven variational models. [175] considers using CNNs to achieve a similar task. However, we haven't seen studied to incorporate shape priors into CNN training. We think there are several direction can be considered. The first is adding shape prior constraints in the final loss layer. The second is formulating other structured outputs to maintain the shape and continuity of predicted contours, such as [174, 146].

4) Miscellaneous Collecting large-scale medical image dataset is extremely difficult. Exploring using less-labeled training data is essential. [92, 112] have show ways of using CNN on natural images and show promising results. However, the study on medical images is not seen. Low-quality or high SNR images are also common in medical images. Specific methods to resist such situation is necessary [152].

8. Conclusion

This paper discusses the key components and technical ingredients of CNNs specific to medical image contour detection. Specifically, we review several mainstream CNN architectures and clarify how these approaches overcome the difficulties of CNN training and promote the CNN development. The advantages and disadvantages are analyzed in details. We believe those details are important for the research of CNN based image contour detection. Next, we discuss several state-of-the-art methods for image contour detection using CNNs with comprehensive analysis and discussions, with the goal to show the problems current state-of-the-art methods are trying to solve. We discuss the challenges and significance of contour detection in medical images and review the historical approaches to solve these problems. Then we review the CNN based medical image contour detection and segmentation methods that leverage on recent advances in CNNs for contour detection and segmentation. Finally, we discuss the problems of existing methods and point out potential directions.

This paper attempts to cover necessary technical ingredients of state-of-the-art CNNs and connect their applications in the medical image domain. Compared with the various methods in the computer vision community, we point out the current diversity deficiency in the medical image contour detection and segmentation and provide potential directions and technical suggestions.

References

References

[1] Al-Kofahi, Y., Lassoued, W., Lee, W., Roysam, B., April 2010. Improved automatic detection and segmentation of cell nuclei in histopathology images. IEEE Transactions on Biomedical Engineering 57 (4), 841–852.

[2] Ali, S., Madabhushi, A., 2012. An integrated region-, boundary-, shape-based active contour for multiple object overlap resolution in histological imagery. IEEE Transactions on Medical Imaging 31 (7), 1448–1460.

[3] Alvarez, J. M., LeCun, Y., Gevers, T., Lopez, A. M., 2012. Semantic road segmentation via multi-scale ensembles of learned features. In: European Conference on Computer Vision. pp. 586–595.

[4] Arbelaez, P., Maire, M., Fowlkes, C., Malik, J., 2009. From contours to regions: An empirical evaluation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 2294–2301.

[5] Arbelaez, P., Maire, M., Fowlkes, C., Malik, J., 2011. Contour detection and hierarchical image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence 33 (5), 898–916.

[6] Arbelaez, P., Pont-Tuset, J., Barron, J., Marques, F., Malik, J., 2014. Multiscale combinatorial grouping. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 328–335.

[7] Arganda-Carreras, I., Turaga, S. C., Berger, D. R., Cireşan, D., Giusti, A., Gambardella, L. M., Schmidhuber, J., Laptev, D., Dwivedi, S., Buhmann, J. M., et al., 2015. Crowdsourcing the creation of image segmentation algorithms for connectomics. Frontiers in neuroanatomy 9.

[8] Bansal, A., Chen, X., Russell, B., Gupta, A., Ramanan, D., 2016. Pixelnet: Towards a General Pixel-level Architecture. arXiv preprint arXiv:1609.06694.

[9] Bengio, Y., Courville, A., Vincent, P., 2013. Representation learning: A review and new perspectives. IEEE Transactions on Pattern Analysis and Machine Intelligence 35 (8), 1798–1828.

[10] 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.

[11] Bernardis, E., Yu, S. X., 2010. Finding dots: segmentation as popping out regions from boundaries. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 199–206.

[12] Bertasius, G., Shi, J., Torresani, L., 2015. Deepedge: A multi-scale bifurcated deep network for top-down contour detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 4380–4389.

[13] Bertasius, G., Shi, J., Torresani, L., 2015. Deepedge: A multi-scale bifurcated deep network for top-down contour detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 4380–4389.

[14] Bertasius, G., Shi, J., Torresani, L., 2015. High-for-low and low-for-high: Efficient boundary detection from deep object features and its applications to high-level vision. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 504–512.

[15] Bottou, L., 2010. Large-scale machine learning with stochastic gradient descent. In: Proceedings of COMPSTAT. pp. 177–186.

[16] Boykov, Y., Kolmogorov, V., 2004. An experimental comparison of min-cut/max- flow algorithms for energy minimization in vision. IEEE Transactions on Pattern Analysis and Machine Intelligence 26 (9), 1124–1137.

[17] Boykov, Y., Veksler, O., Zabih, R., 2001. Fast approximate energy minimization via graph cuts. IEEE Transactions on Pattern Analysis and Machine Intelligence 23 (11), 1222–1239.

[18] Cai, J., Lu, L., Xie, Y., Xing, F., Yang, L., 2017. Improving deep pancreas segmentation in ct and mri images via recurrent neural contextual learning and direct loss function. International Conference on Medical Image Computing and Computer-Assisted Intervention.

[19] Cai, J., Lu, L., Zhang, Z., Xing, F., Yang, L., Yin, Q., 2016. Pancreas segmentation in mri using graph-based decision fusion on convolutional neural networks. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 442–450.

[20] Cai, Z., Fan, Q., Feris, R. S., Vasconcelos, N., 2016. A unified multi-scale deep convolutional neural network for fast object detection. In: European Conference on Computer Vision. pp. 354–370.

[21] Canny, J., 1986. A computational approach to edge detection. IEEE Transactions on pattern Analysis and Machine Intelligence (6), 679–698.

[22] Caselles, V., Kimmel, R., Sapiro, G., 1997. Geodesic active contours. International Journal of Computer Vision 22 (1), 61–79.

[23] Chan, T. F., Vese, L. A., 2001. Active contours without edges. IEEE Transactions on Image Processing 10 (2), 266–277.

[24] Chang, H., Han, J., Borowsky, A., Loss, L., Gray, J., Spellman, P., Parvin, B., 2013. Invariant delineation of nuclear architecture in glioblastoma multiforme for clinical and molecular association. IEEE Transactions on Medical Imaging 32 (4), 670–682.

[25] Chatfield, K., Simonyan, K., Vedaldi, A., Zisserman, A., 2014. Return of the devil in the details: Delving deep into convolutional nets. arXiv preprint arXiv:1405.3531.

[26] Chen, F., Yu, H., Hu, R., Zeng, X., 2013. Deep learning shape priors for object segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1870–1877.

[27] Chen, H., Qi, X., Yu, L., Dou, Q., Qin, J., Heng, P.-A., 2017. Dcan: Deep contour-aware networks for object instance segmentation from histology images. Medical Image Analysis 36, 135–146.

[28] Chen, H., Qi, X., Yu, L., Heng, P.-A., 2016. Dcan: Deep contour-aware networks for accurate gland segmentation. In: Proceedings of the IEEE conference on Computer Vision and Pattern Recognition. pp. 2487–2496.

[29] Chen, H., Qi, X. J., Cheng, J. Z., Heng, P. A., 2016. Deep contextual networks for neuronal structure segmentation. In: Thirtieth AAAI Conference on Artificial Intelligence.

[30] Chen, J., Yang, L., Zhang, Y., Alber, M., Chen, D. Z., 2016. Combining fully convolutional and recurrent neural networks for 3d biomedical image segmentation. In: Advances in Neural Information Processing Systems. pp. 3036–3044.

[31] Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A. L., 2014. Semantic image segmentation with deep convolutional nets and fully connected crfs. arXiv preprint arXiv:1412.7062.

[32] Chen, X., Zhou, X., Wong, S. T. C., April 2006. Automated segmentation, classification, and tracking of cancer cell nuclei in time-lapse microscopy. IEEE Transaction Biomedical Engineering 53 (4), 762–766.

[33] Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T., Ronneberger, O., 2016. 3d u-net: learning dense volumetric segmentation from sparse annotation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 424–432.

[34] Ciresan, D., Giusti, A., Gambardella, L. M., Schmidhuber, J., 2012. Deep neural networks segment neuronal membranes in electron microscopy images. In: Advances in Neural Information Processing Systems. pp. 2843–2851.

[35] Clevert, D.-A., Unterthiner, T., Hochreiter, S., 2015. Fast and accurate deep network learning by exponential linear units (elus). arXiv preprint arXiv:1511.07289.

[36] Comaniciu, D., Meer, P., 2002. Mean shift: A robust approach toward feature space analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence 24 (5), 603–619.

[37] Cortes, C., Vapnik, V., 1995. Support-vector networks. Machine learning 20 (3), 273–297.

[38] Cour, T., Benezit, F., Shi, J., 2005. Spectral segmentation with multiscale graph decomposition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Vol. 2. pp. 1124–1131.

[39] Cremers, D., Rousson, M., Deriche, R., 2007. A review of statistical approaches to level set segmentation: integrating color, texture, motion and shape. International Journal of Computer Vision 72 (2), 195–215.

[40] Dai, J., He, K., Sun, J., 2015. Boxsup: Exploiting bounding boxes to supervise convolutional networks for semantic segmentation. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 1635–1643.

[41] Dai, W., Doyle, J., Liang, X., Zhang, H., Dong, N., Li, Y., Xing, E. P., 2017. Scan: Structure correcting adversarial network for chest x-rays organ segmentation. arXiv preprint arXiv:1703.08770.

[42] Delgado-Gonzalo, R., Uhlmann, V., Schmitter, D., Unser, M., 2015. Snakes on a plane: a perfect snap for bioimage analysis. IEEE Signal Processing Magazine 32 (1), 41–48.

[43] Dollar, P., Tu, Z., Belongie, S., 2006. Supervised learning of edges and object boundaries. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Vol. 2. pp. 1964–1971.

[44] Dollár, P., Zitnick, C. L., 2013. Structured forests for fast edge detection. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 1841–1848.

[45] Dollár, P., Zitnick, C. L., 2015. Fast edge detection using structured forests. IEEE Transactions on Pattern Analysis and Machine Intelligence 37 (8), 1558–1570.

[46] Dou, Q., Chen, H., Jin, Y., Yu, L., Qin, J., Heng, P.-A., 2016. 3d deeply supervised network for automatic liver segmentation from ct volumes. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 149–157.

[47] Drozdzal, M., Vorontsov, E., Chartrand, G., Kadoury, S., Pal, C., 2016. The importance of skip connections in biomedical image segmentation. In: International Workshop on Large-Scale Annotation of Biomedical Data and Expert Label Synthesis. pp. 179–187.

[48] Duchi, J., Hazan, E., Singer, Y., 2011. Adaptive subgradient methods for online learning and stochastic optimization. Journal of Machine Learning Research 12, 2121–2159.

[49] Duhamel, J.-R., Bremmer, F., BenHamed, S., Graf, W., 1997. Spatial invariance of visual receptive fields in parietal cortex neurons. Nature 389 (6653), 845–848.

[50] El-Baz, A., Beache, G. M., Gimel'farb, G., Suzuki, K., Okada, K., Elnakib, A., Soliman, A., Abdollahi, B., 2013. Computer-aided diagnosis systems for lung cancer: challenges and methodologies. International journal of biomedical imaging 2013.

[51] Eslami, S. A., Heess, N., Williams, C. K., Winn, J., 2014. The shape boltzmann machine: a strong model of object shape. International Journal of Computer Vision 107 (2), 155–176.

[52] Fakhry, A., Peng, H., Ji, S., 2016. Deep models for brain em image segmentation: novel insights and improved performance. Bioinformatics, 2352–2358.

[53] Felzenszwalb, P. F., Huttenlocher, D. P., 2004. Efficient graph-based image segmentation. IJCV 59 (2), 167–181.

[54] Freeman, W. T., Adelson, E. H., 1991. The design and use of steerable filters. IEEE Transactions on Pattern analysis and machine intelligence 13 (9), 891–906.

[55] Fu, H., Wang, C., Tao, D., Black, M., 2016. Occlusion boundary detection via deep exploration of context. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 234–241.

[56] Ganin, Y., Lempitsky, V., 2014. Nˆ 4-fields: Neural network nearest neighbor fields for image transforms. In: Asian Conference on Computer Vision. pp. 536–551.

[57] Girshick, R., Donahue, J., Darrell, T., Malik, J., 2014. Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 580–587.

[58] Glorot, X., Bengio, Y., 2010. Understanding the difficulty of training deep feedforward neural networks. In: Artificial Intelligence and Statistics Conference. Vol. 9. pp. 249–256.

[59] Glorot, X., Bordes, A., Bengio, Y., 2011. Deep sparse rectifier neural networks. In: Artificial Intelligence and Statistics Conference. Vol. 15. p. 275.

[60] Gonzalez, R. C., Woods, R. E., 2008. Digital image processing. Pearson Education, Inc., Upper Saddle River, NJ, USA.

[61] Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep Learning. MIT Press, http://www.deeplearningbook.org.

[62] 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.

[63] Goodfellow, I. J., Warde-Farley, D., Mirza, M., Courville, A. C., Bengio, Y., 2013. Maxout networks. International Conference on Machine Learning 28, 1319–1327.

[64] Grady, L., 2006. Random walks for image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence 28 (11), 1768–1783.

[65] Grady, L., Schwartz, E. L., 2006. Isoperimetric graph partitioning for image segmetentation. IEEE Transactions on Pattern Analysis and Machine Intelligence 28 (1), 469–475.

[66] Graves, A., Wayne, G., Reynolds, M., Harley, T., Danihelka, I., Grabska-Barwińska, A., Colmenarejo, S. G., Grefenstette, E., Ramalho, T., Agapiou, J., et al., 2016. Hybrid computing using a neural network with dynamic external memory. Nature 538 (7626), 471–476.

[67] Greenspan, H., van Ginneken, B., Summers, R. M., 2016. Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique. IEEE Transactions on Medical Imaging 35 (5), 1153–1159.

[68] Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., Liu, T., Wang, X., Wang, G., 2015. Recent advances in convolutional neural networks. arXiv preprint arXiv:1512.07108.

[69] Hajjar, A., Chen, T., 1997. A new real time edge linking algorithm and its vlsi implementation. In: Computer Architecture for Machine Perception, 1997. CAMP 97. Proceedings. 1997 Fourth IEEE International Workshop on. pp. 280–284.

[70] Hariharan, B., Arbeláez, P., Girshick, R., Malik, J., 2015. Hypercolumns for object segmentation and fine-grained localization. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 447–456.

[71] He, K., Sun, J., 2015. Convolutional neural networks at constrained time cost. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 5353–5360.

[72] He, K., Zhang, X., Ren, S., Sun, J., 2015. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 1026–1034.

[73] 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.

[74] He, K., Zhang, X., Ren, S., Sun, J., 2016. Identity mappings in deep residual networks. In: European Conference on Computer Vision. pp. 630–645.

[75] Hinton, G., 2011. Deep belief nets. In: Encyclopedia of Machine Learning. Springer, pp. 267–269.

[76] Hinton, G., Srivastava, N., Swersky, K., 2012. Neural networks for machine learning lecture 6a overview of mini–batch gradient descent.

[77] Hinton, G. E., Salakhutdinov, R. R., 2006. Reducing the dimensionality of data with neural networks. Science 313 (5786), 504–507.

[78] Hinton, G. E., Zemel, R. S., 1994. Autoencoders, minimum description length and helmholtz free energy. In: Advances in Neural Information Processing Systems. pp. 3–10.

[79] Hong, S., Noh, H., Han, B., 2015. Decoupled deep neural network for semi-supervised semantic segmentation. In: Advances in Neural Information Processing Systems. pp. 1495–1503.

[80] Hong, S., Oh, J., Lee, H., Han, B., 2016. Learning transferrable knowledge for semantic segmentation with deep convolutional neural network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 3204–3212.

[81] Hsieh, P.-J., Vul, E., Kanwisher, N., 2010. Recognition alters the spatial pattern of fmri activation in early retinotopic cortex. Journal of neurophysiology 103 (3), 1501–1507.

[82] Huang, G., Liu, Z., Weinberger, K. Q., 2017. Densely connected convolutional networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.

[83] Huang, G., Sun, Y., Liu, Z., Sedra, D., Weinberger, K. Q., 2016. Deep networks with stochastic depth. In: European Conference on Computer Vision. pp. 646–661.

[84] Hwang, J.-J., Liu, T.-L., 2015. Pixel-wise deep learning for contour detection. arXiv preprint arXiv:1504.01989.

[85] Ioffe, S., Szegedy, C., 2015. Batch normalization: Accelerating deep network training by reducing internal covariate shift. arXiv preprint arXiv:1502.03167.

[86] Isola, P., Zoran, D., Krishnan, D., Adelson, E. H., 2014. Crisp boundary detection using pointwise mutual information. In: European Conference on Computer Vision. pp. 799–814.

[87] Jaderberg, M., Simonyan, K., Zisserman, A., et al., 2015. Spatial transformer networks. In: Advances in Neural Information Processing Systems. pp. 2017–2025.

[88] Janssens, T., Antanas, L., Derde, S., Vanhorebeek, I., den Berghe, G. V., Grandas, F. G., December 2013. Charisma: an integrated approach to automatic h&e-stained skeletal muscle cell segmentation using supervised learning and novel robust clump splitting. Medical Image Analysis 17 (8), 1206–1219.

[89] Jiang, J., Trundle, P., Ren, J., 2010. Medical image analysis with artificial neural networks. Computerized Medical Imaging and Graphics 34 (8), 617–631.

[90] Kallenberg, M., Petersen, K., Nielsen, M., Ng, A. Y., Diao, P., Igel, C., Vachon, C. M., Holland, K., Winkel, R. R., Karssemeijer, N., et al., 2016. Unsupervised deep learning applied to breast density segmentation and mammographic risk scoring. IEEE Transactions on Medical Imaging 35 (5), 1322–1331.

[91] Kass, M., Witkin, A., Terzopoulos, D., 1988. Snakes: Active contour models. International Journal of Computer Vision 1 (4), 321–331.

[92] Khoreva, A., Benenson, R., Omran, M., Hein, M., Schiele, B., 2016. Weakly supervised object boundaries. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 183–192.

[93] Kim, J., Kwon Lee, J., Mu Lee, K., 2016. Accurate image super-resolution using very deep convolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1646–1654.

[94] Kim, T. H., Lee, K. M., Lee, S. U., 2013. Learning full pairwise affinities for spectral segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence 35 (7), 1690–1703.

[95] Kingma, D., Ba, J., 2014. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

[96] Kivinen, J. J., Williams, C. K., Heess, N., Technologies, D., 2014. Visual boundary prediction: A deep neural prediction network and quality dissection. In: Artificial Intelligence and Statistics Conference. Vol. 1. p. 9.

[97] Kokkinos, I., 2015. Pushing the boundaries of boundary detection using deep learning. arXiv preprint arXiv:1511.07386.

[98] Kong, H., Gurcan, M., Belkacem-Boussaid, K., 2011. Partitioning histopathological images: an integrated framework for supervised color-texture segmentation and cell splitting. IEEE Transactions on Medical Imaging 30 (9), 1661–1677.

[99] Konishi, S., Yuille, A. L., Coughlan, J. M., Zhu, S. C., 2003. Statistical edge detection: Learning and evaluating edge cues. IEEE Transactions on Pattern Analysis and Machine Intelligence 25 (1), 57–74.

[100] Kontschieder, P., Bulo, S. R., Bischof, H., Pelillo, M., 2011. Structured class-labels in random forests for semantic image labelling. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 2190–2197.

[101] Kourtzi, Z., Kanwisher, N., 2001. Representation of perceived object shape by the human lateral occipital complex. Science 293 (5534), 1506–1509.

[102] Krizhevsky, A., Sutskever, I., Hinton, G. E., 2012. Imagenet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems. pp. 1097–1105.

[103] LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521 (7553), 436–444.

[104] LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., Jackel, L. D., 1989. Backpropagation applied to handwritten zip code recognition. Neural computation 1 (4), 541–551.

[105] LeCun, Y. A., Bottou, L., Orr, G. B., Müller, K.-R., 2012. Efficient backprop. In: Neural networks: Tricks of the trade. pp. 9–48.

[106] Lee, C.-Y., Xie, S., Gallagher, P., Zhang, Z., Tu, Z., 2015. Deeply-supervised nets. In: Artificial Intelligence and Statistics Conference. pp. 562–570.

[107] Lee, H., Grosse, R., Ranganath, R., Ng, A. Y., 2009. Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations. In: Proceedings of the 26th annual international conference on machine learning. pp. 609–616.

[108] Lee, H., Pham, P., Largman, Y., Ng, A. Y., 2009. Unsupervised feature learning for audio classification using convolutional deep belief networks. In: Advances in Neural Information Processing Systems. pp. 1096–1104.

[109] Levinshtein, A., Sminchisescu, C., Dickinson, S., 2010. Optimal contour closure by superpixel grouping. In: European Conference on Computer Vision. pp. 480–493.

[110] Li, C., Wand, M., 2016. Combining markov random fields and convolutional neural networks for image synthesis. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 2479–2486.

[111] Li, K., Miller, E. D., Chen, M., Kanade, T., Weiss, L. E., Campbell, P. G., 2008. Cell population tracking and lineage construction with spatiotemporal context. Medical Image Analysis 12 (5), 546–566.

[112] Li, Y., Paluri, M., Rehg, J. M., Dollár, P., 2016. Unsupervised learning of edges. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1619–1627.

[113] Liaw, A., Wiener, M., 2002. Classification and regression by random forest. R news 2 (3), 18–22.

[114] Lim, J. J., Zitnick, C. L., Dollár, P., 2013. Sketch tokens: A learned mid-level representation for contour and object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 3158–3165.

[115] Lin, G., Chawla, M. K., Olson, K., Guzowski, J. F., Barnes, C. A., Roysam, B., 2005. Hierarchical, model-based merging of multiple fragments for improved three-dimensional segmentation of nuclei. Cytometry A 63 (1), 20–33.

[116] Lin, G., Shen, C., Reid, I., van den Hengel, A., 2015. Deeply learning the messages in message passing inference. In: Advances in Neural Information Processing Systems. pp. 361–369.

[117] Lin, G., Shen, C., van den Hengel, A., Reid, I., 2016. Efficient piecewise training of deep structured models for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 3194–3203.

[118] Lin, M., Chen, Q., Yan, S., 2014. Network in network. International Conference on Learning Representations.

[119] Liu, F., Lin, G., Shen, C., 2015. Crf learning with cnn features for image segmentation. Pattern Recognition 48 (10), 2983–2992.

[120] Liu, F., Xing, F., Su, H., Yang, L., 2014. Touching adipocyte cells decomposition using combinatorial optimization. In: IEEE 11th International Symposium on Biomedical Imaging (ISBI). pp. 1340–1347.

[121] Liu, F., Xing, F., Zhang, Z., Mcgough, M., Yang, L., 2015. Robust muscle cell quantification using structured edge detection and hierarchical segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 324–331.

[122] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., 2015. Ssd: Single shot multibox detector. pp. 21–37.

[123] Liu, Y., Cheng, M.-m., Hu, X., Wang, K., Bai, X., 2016. Learning relaxed deep supervision for better edge detection. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.

[124] Liu, Y., Cheng, M.-M., Hu, X., Wang, K., Bai, X., 2017. Richer convolutional features for edge detection. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.

[125] Liu, Z., Li, X., Luo, P., Loy, C.-C., Tang, X., 2015. Semantic image segmentation via deep parsing network. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 1377–1385.

[126] 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.

[127] Lu, J., Xiong, C., Parikh, D., Socher, R., 2016. Knowing when to look: Adaptive attention via a visual sentinel for image captioning. arXiv preprint arXiv:1612.01887.

[128] Ma, W.-Y., Manjunath, B. S., 2000. Edgeflow: a technique for boundary detection and image segmentation. IEEE Transactions on Image Processing 9 (8), 1375–1388.

[129] Maas, A. L., Hannun, A. Y., Ng, A. Y., 2013. Rectifier nonlinearities improve neural network acoustic models. In: International Conference on Machine Learning. Vol. 30.

[130] Madabhushi, A., Lee, G., 2016. Image analysis and machine learning in digital pathology: Challenges and opportunities.

[131] Mahendran, A., Vedaldi, A., 2015. Understanding deep image representations by inverting them. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 5188–5196.

[132] Mairal, J., Leordeanu, M., Bach, F., Hebert, M., Ponce, J., 2008. Discriminative sparse image models for class-specific edge detection and image interpretation. In: European Conference on Computer Vision. pp. 43–56.

[133] Maire, M., Arbeláez, P., Fowlkes, C., Malik, J., 2008. Using contours to detect and localize junctions in natural images. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1–8.

[134] Maire, M., Narihira, T., Yu, S. X., 2016. Affinity cnn: Learning pixel-centric pairwise relations for figure/ground embedding. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 174–182.

[135] Maire, M., Stella, X. Y., Perona, P., 2014. Reconstructive sparse code transfer for contour detection and semantic labeling. In: Asian Conference on Computer Vision. pp. 273–287.

[136] Maire, M. R., 2009. Contour detection and image segmentation. Ph.D. thesis, Citeseer.

[137] Malik, J., Belongie, S., Leung, T., Shi, J., 2001. Contour and texture analysis for image segmentation. International Journal of Computer Vision 43 (1), 7–27.

[138] Maninis, K.-K., Pont-Tuset, J., Arbeláez, P., Van Gool, L., 2016. Convolutional oriented boundaries. In: European Conference on Computer Vision. pp. 580–596.

[139] Maninis, K.-K., Pont-Tuset, J., Arbeláez, P., Van Gool, L., 2016. Deep retinal image understanding. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 140–148.

[140] Marr, D., Hildreth, E., 1980. Theory of edge detection. Proceedings of the Royal Society of London B: Biological Sciences 207 (1167), 187–217.

[141] Martin, D. R., Fowlkes, C. C., Malik, J., 2004. Learning to detect natural image boundaries using local brightness, color, and texture cues. IEEE Transactions on Pattern Analysis and Machine Intelligence 26 (5), 530–549.

[142] Meer, P., Georgescu, B., 2001. Edge detection with embedded confidence. IEEE Transactions on pattern Analysis and Machine Intelligence 23 (12), 1351–1365.

[143] Meijering, E., 2012. Cell segmentation: 50 years down the road. IEEE Signal Processing Magazine 29 (5), 140–145.

[144] Milletari, F., Ahmadi, S.-A., Kroll, C., Plate, A., Rozanski, V., Maiostre, J., Levin, J., Dietrich, O., Ertl-Wagner, B., Bötzel, K., et al., 2017. Hough-cnn:

Deep learning for segmentation of deep brain regions in mri and ultrasound. Computer Vision and Image Understanding.

[145] Milletari, F., Navab, N., Ahmadi, S.-A., 2016. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In: 3D Vision (3DV), 2016 Fourth International Conference on. pp. 565–571.

[146] Mo, Y., Liu, F., Zhang, J., Yang, G., He, T., Guo, Y., 2017. Deep poincaré map for robust medical image segmentation. In: https://arxiv.org/pdf/1703.09200.pdf.

[147] Morrone, M. C., Owens, R. A., 1987. Feature detection from local energy. Pattern recognition letters 6 (5), 303–313.

[148] Myers, A., Teo, C. L., Fermüller, C., Aloimonos, Y., 2015. Affordance detection of tool parts from geometric features. In: International Conference on Robotics and Automation.

[149] Nair, V., Hinton, G. E., 2010. Rectified linear units improve restricted boltzmann machines. In: Proceedings of the 27th International Conference on Machine Learning. pp. 807–814.

[150] Nogues, I., Lu, L., Wang, X., Roth, H., Bertasius, G., Lay, N., Shi, J., Tsehay, Y., Summers, R. M., 2016. Automatic lymph node cluster segmentation using holistically-nested neural networks and structured optimization in ct images. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 388–397.

[151] 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.

[152] Ofir, N., Galun, M., Nadler, B., Basri, R., 2016. Fast detection of curved edges at low snr. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.

[153] Oktay, O., Ferrante, E., Kamnitsas, K., Heinrich, M., Bai, W., Caballero, J., Guerrero, R., Cook, S., de Marvao, A., O'Regan, D., et al., 2017. Anatomically constrained neural networks (acnn): Application to cardiac image enhancement and segmentation. arXiv preprint arXiv:1705.08302.

[154] Oquab, M., Bottou, L., Laptev, I., Sivic, J., 2014. Learning and transferring mid-level image representations using convolutional neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1717–1724.

[155] Papandreou, G., Chen, L.-C., Murphy, K., Yuille, A. L., 2015. Weakly- and semi-supervised learning of a dcnn for semantic image segmentation. Proceedings of the IEEE International Conference on Computer Vision.

[156] Papari, G., Petkov, N., 2011. Edge and line oriented contour detection: State of the art. Image and Vision Computing 29 (2), 79–103.

[157] Pascanu, R., Mikolov, T., Bengio, Y., 2013. On the difficulty of training recurrent neural networks. ICML (3) 28, 1310–1318.

[158] Paszke, A., Chaurasia, A., Kim, S., Culurciello, E., 2016. Enet: A deep neural network architecture for real-time semantic segmentation. arXiv preprint arXiv:1606.02147.

[159] Pathak, D., Shelhamer, E., Long, J., Darrell, T., 2014. Fully convolutional multi-class multiple instance learning. International Conference on Learning Representations.

[160] Perona, P., Malik, J., 1990. Detecting and localizing edges composed of steps, peaks and roofs. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 52–57.

[161] Perona, P., Malik, J., 1990. Scale-space and edge detection using anisotropic diffusion. IEEE Transactions on Pattern Analysis and Machine Intelligence 12 (7), 629–639.

[162] Pinheiro, P. O., Collobert, R., Dollar, P., 2015. Learning to segment object candidates. In: Advances in Neural Information Processing Systems. pp. 1990–1998.

[163] Raiko, T., Valpola, H., LeCun, Y., 2012. Deep learning made easier by linear transformations in perceptrons. In: Artificial Intelligence and Statistics Conference. Vol. 22. pp. 924–932.

[164] Rastegari, M., Ordonez, V., Redmon, J., Farhadi, A., 2016. Xnor-net: Imagenet classification using binary convolutional neural networks. arXiv preprint arXiv:1603.05279.

[165] 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.

[166] Ren, X., 2008. Multi-scale improves boundary detection in natural images. In: European Conference on Computer Vision. pp. 533–545.

[167] Ren, X., Fowlkes, C. C., Malik, J., 2005. Scale-invariant contour completion using conditional random fields. In: Proceedings of the IEEE International Conference on Computer Vision. Vol. 2. pp. 1214–1221.

[168] Ren, X., Fowlkes, C. C., Malik, J., 2006. Figure/ground assignment in natural images. In: European Conference on Computer Vision. pp. 614–627.

[169] Ren, Z., Shakhnarovich, G., 2013. Image segmentation by cascaded region agglomeration. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 2011–2018.

[170] Roerdink, J. B., Meijster, A., 2000. The watershed transform: Definitions, algorithms and parallelization strategies. Fundamenta informaticae 41 (1, 2), 187–228.

[171] 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. pp. 234–241.

[172] Roth, H. R., Lu, L., Farag, A., Sohn, A., Summers, R. M., 2016. Spatial aggregation of holistically-nested networks for automated pancreas segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 451–459.

[173] Roth, H. R., Lu, L., Liu, J., Yao, J., Seff, A., Cherry, K., Kim, L., Summers, R. M., 2016. Improving computer-aided detection using convolutional neural networks and random view aggregation. IEEE Transactions on Medical Imaging 35 (5), 1170–1181.

[174] Rupprecht, C., Huaroc, E., Baust, M., Navab, N., 2016. Deep active contours. arXiv preprint arXiv:1607.05074.

[175] Safar, S., Yang, M. H., 2015. Learning shape priors for object segmentation via neural networks. In: IEEE International Conference on Image Processing. pp. 1835–1839.

[176] Salakhutdinov, R., Hinton, G. E., 2009. Deep boltzmann machines. In: Proceedings of the International Conference on Artificial Intelligence and Statistics. Vol. 5. pp. 448–455.

[177] Salakhutdinov, R., Mnih, A., Hinton, G., 2007. Restricted boltzmann machines for collaborative filtering. In: Proceedings of the 24th international conference on Machine learning. pp. 791–798.

[178] Sanguinetti, J. L., Allen, J. J., Peterson, M. A., 2013. The ground side of an object perceived as shapeless yet processed for semantics. Psychological Science.

[179] Schmidhuber, J., 2015. Deep learning in neural networks: An overview. Neural Networks 61, 85–117.

[180] Schulz, H., Behnke, S., 2012. Learning object-class segmentation with convolutional neural networks. In: European Symposium on Artificial Neural Networks.

[181] Shah, A., Kadam, E., Shah, H., Shinde, S., Shingade, S., 2016. Deep residual networks with exponential linear unit. In: Proceedings of the Third International Symposium on Computer Vision and the Internet. ACM, pp. 59–65.

[182] Shapley, R., Tolhurst, D., 1973. Edge detectors in human vision. The Journal of physiology 229 (1), 165.

[183] Sharif Razavian, A., Azizpour, H., Sullivan, J., Carlsson, S., 2014. Cnn features off-the-shelf: an astounding baseline for recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. pp. 806–813.

[184] Sharma, A., Tuzel, O., Jacobs, D. W., 2015. Deep hierarchical parsing for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 530–538.

[185] Shelhamer, E., Long, J., Darrell, T., 2016. Fully convolutional networks for semantic segmentation. In: IEEE Transactions on Pattern Analysis and Machine Intelligence. Vol. 39. pp. 640–651.

[186] Shen, W., Wang, X., Wang, Y., Bai, X., Zhang, Z., 2015. Deepcontour: A deep convolutional feature learned by positive-sharing loss for contour detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 3982–3991.

[187] Shen, W., Zhao, K., Jiang, Y., Wang, Y., Zhang, Z., Bai, X., 2016. Object skeleton extraction in natural images by fusing scale-associated deep side outputs. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 222–230.

[188] Shi, J., Malik, J., 2000. Normalized cuts and image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence 22 (8), 888–905.

[189] Shin, H.-C., Roth, H. R., Gao, M., Lu, L., Xu, Z., Nogues, I., Yao, J., Mollura, D., Summers, R. M., 2016. Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning. IEEE Transactions on Medical Imaging 35 (5), 1285–1298.

[190] Shrivakshan, G., Chandrasekar, C., et al., 2012. A comparison of various edge detection techniques used in image processing. IJCSI International Journal of Computer Science Issues 9 (5), 272–276.

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

[192] Singh, S., Hoiem, D., Forsyth, D., 2016. Swapout: Learning an ensemble of deep architectures. In: Advances in Neural Information Processing Systems. pp. 28–36.

[193] Sirinukunwattana, K., Pluim, J. P., Chen, H., Qi, X., Heng, P.-A., Guo, Y. B., Wang, L. Y., Matuszewski, B. J., Bruni, E., Sanchez, U., et al., 2017. Gland segmentation in colon histology images: The glas challenge contest. Medical Image Analysis 35, 489–502.

[194] Song, Y., Zhang, L., Chen, S., Ni, D., Lei, B., Wang, T., 2015. Accurate segmentation of cervical cytoplasm and nuclei based on multiscale convolutional network and graph partitioning. IEEE Transactions on Biomedical Engineering 62 (10), 2421–2433.

[195] Srivastava, N., Hinton, G. E., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., 2014. Dropout: a simple way to prevent neural networks from overfitting. Journal of Machine Learning Research 15 (1), 1929–1958.

[196] Srivastava, R. K., Greff, K., Schmidhuber, J., 2015. Highway networks. arXiv preprint arXiv:1505.00387.

[197] Stahl, J. S., Wang, S., 2008. Globally optimal grouping for symmetric closed boundaries by combining boundary and region information. IEEE Transactions on Pattern Analysis and Machine Intelligence 30 (3), 395–411.

[198] Su, H., Xing, F., Kong, X., Xie, Y., Zhang, S., Yang, L., 2015. Robust cell detection and segmentation in histopathological images using sparse reconstruction and stacked denoising autoencoders. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 383–390.

[199] Sukhbaatar, S., Weston, J., Fergus, R., et al., 2015. End-to-end memory networks. In: Advances in Neural Information Processing Systems. pp. 2440–2448.

[200] Sundberg, P., Brox, T., Maire, M., Arbeláez, P., Malik, J., 2011. Occlusion boundary detection and figure/ground assignment from optical flow. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 2233–2240.

[201] Sutskever, I., Martens, J., Dahl, G. E., Hinton, G. E., 2013. On the importance of initialization and momentum in deep learning. International Conference on Machine Learning 28, 1139–1147.

[202] Suzuki, K., Horiba, I., Sugie, N., 2003. Neural edge enhancer for supervised edge enhancement from noisy images. IEEE Transactions on Pattern Analysis and Machine Intelligence 25 (12), 1582–1596.

[203] Szegedy, C., Ioffe, S., Vanhoucke, V., 2016. Inception-v4, inception-resnet and the impact of residual connections on learning. arXiv preprint arXiv:1602.07261.

[204] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A., 2015. Going deeper with convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1–9.

[205] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z., 2016. Rethinking the inception architecture for computer vision. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 2818–2826.

[206] Tajbakhsh, N., Shin, J. Y., Gurudu, S. R., Hurst, R. T., Kendall, C. B., Gotway, M. B., Liang, J., 2016. Convolutional neural networks for medical image analysis: Full training or fine tuning? IEEE Transactions on Medical Imaging 35 (5), 1299–1312.

[207] Targ, S., Almeida, D., Lyman, K., 2016. Resnet in resnet: Generalizing residual architectures. arXiv preprint arXiv:1603.08029.

[208] Taylor, C. J., 2013. Towards fast and accurate segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1916–1922.

[209] Teikari, P., Santos, M., Poon, C., Hynynen, K., 2016. Deep learning convolutional networks for multiphoton microscopy vasculature segmentation. arXiv preprint arXiv:1606.02382.

[210] Teo, C. L., Fermüller, C., Aloimonos, Y., 2015. Fast 2d border ownership assignment. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 5117–5125.

[211] Uijlings, J. R., Ferrari, V., 2015. Situational object boundary detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 4712–4721.

[212] Veit, A., Wilber, M. J., Belongie, S., 2016. Residual networks behave like ensembles of relatively shallow networks. In: Advances in Neural Information Processing Systems. pp. 550–558.

[213] Wan, L., Zeiler, M., Zhang, S., Cun, Y. L., Fergus, R., 2013. Regularization of neural networks using dropconnect. In: International Conference on Machine Learning. pp. 1058–1066.

[214] Wang, S., Stahl, J. S., Bailey, A., Dropps, M., 2007. Global detection of salient convex boundaries. International Journal of Computer Vision 71 (3), 337–359.

[215] Xiaofeng, R., Bo, L., 2012. Discriminatively trained sparse code gradients for contour detection. In: Advances in Neural Information Processing Systems. pp. 584–592.

[216] Xie, S., Girshick, R., Dollár, P., Tu, Z., He, K., 2017. Aggregated residual transformations for deep neural networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.

[217] Xie, S., Tu, Z., 2015. Holistically-nested edge detection. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 1395–1403.

[218] Xie, Y., Xing, F., Kong, X., Su, H., Yang, L., 2015. Beyond classification: structured regression for robust cell detection using convolutional neural network. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 358–365.

[219] Xing, F., Shi, X., Zhang, Z., Cai, J., Xie, Y., Yang, L., 2016. Transfer shape modeling towards high-throughput microscopy image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 183–190.

[220] Xing, F., Su, H., Neltner, J., Yang, L., 2014. Automatic Ki-67 counting using robust cell detection and online dictionary learning. IEEE Transactions on Biomedical Engineering 61 (3), 859–870.

[221] Xing, F., Xie, Y., Yang, L., February 2016. An automatic learning-based framework for robust nucleus segmentation. IEEE Transactions on Medical Imaging 35 (2), 550–566.

[222] Xing, F., Yang, L., 2016. Robust nucleus/cell detection and segmentation in digital pathology and microscopy images: A comprehensive review. IEEE reviews in biomedical engineering.

[223] Xu, C., Prince, J. L., 1998. Snakes, shapes, and gradient vector flow. IEEE Transactions on Image Processing 7 (3), 359–369.

[224] Xu, Y., Li, Y., Liu, M., Wang, Y., Lai, M., Eric, I., Chang, C., 2016. Gland instance segmentation by deep multichannel side supervision. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 496–504.

[225] Yang, J., Price, B., Cohen, S., Lee, H., Yang, M.-H., 2016. Object contour detection with a fully convolutional encoder-decoder network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 193–202.

[226] Yang, L., Qiu, Z., Greenaway, A. H., Lu, W., July 2012. A new framework for particle detection in low-snr fluorescence live-cell images and its application for improved particle tracking. IEEE Transactions on Biomedical Engineering 59 (7), 2040–2050.

[227] Yang, X., Yu, L., Wu, L., Ni, D., Heng, P.-A., 2017. Shape completion with recurrent memory.

[228] Yu, F., Koltun, V., 2016. Multi-scale context aggregation by dilated convolutions. International Conference on Learning Representations.

[229] Yu, Y., Fang, C., Liao, Z., 2015. Piecewise flat embedding for image segmentation. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 1368–1376.

[230] Zagoruyko, S., Komodakis, N., 2016. Wide residual networks. British Machine Vision Conference.

[231] Zeiler, M. D., Fergus, R., 2014. Visualizing and understanding convolutional networks. In: European Conference on Computer Vision. pp. 818–833.

[232] Zeiler, M. D., Krishnan, D., Taylor, G. W., Fergus, R., 2010. Deconvolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 2528–2535.

[233] Zhang, B., Zimmer, C., Olivo-Marin, J.-C., 2004. Tracking fluorescent cells with coupled geometric active contours. In: Biomedical Imaging: Nano to Macro, 2004. IEEE International Symposium on. pp. 476–479.

[234] Zhang, Z., Xing, F., Shi, X., Yang, L., 2016. Semicontour: A semi-supervised learning approach for contour detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 251–259.

[235] Zheng, S., Jayasumana, S., Romera-Paredes, B., Vineet, V., Su, Z., Du, D., Huang, C., Torr, P. H., 2015. Conditional random fields as recurrent neural networks. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 1529–1537.

[236] Ziou, D., Tabbone, S., et al., 1998. Edge detection techniques-an overview. Pattern Recognition and Image Analysis C/C of Raspoznavaniye Obrazov I Analiz Izobrazhenii 8, 537–559.