Image Segmentation and Classification Using Deep Learning

Chapter · November 2022

DOI: 10.1002/9781119861850.ch2

Abhisek Ray
Indian Institute of Technology Patna

Maheshkumar Kolekar
Indian Institute of Technology Patna

2

Image Segmentation and Classification Using Deep Learning

Abhisek Ray and Maheshkumar H. Kolekar

Video Surveillance Lab, Department of Electrical Engineering, Indian Institute of Technology, Patna, Bihar 801103, India

2.1 Introduction

Among many, three major concerns surfaced which propel researchers to concentrate more on image processing [1] and video processing [2–4] and especially in the digital domain. These problems can be elaborated as transmission, storage, and printing of images, effective interpretation of an image, and machine vision of an image. Many image-processing techniques had been proposed toward these problems, for example, image coding, digitization, and coding for first; picture restoration, enhancement, and restoration for the second; image description and segmentation for the last one.

A panchromatic image can be interpreted as a two-dimensional light intensity function, $f(x, y)$, in which x and y are the spatial coordinates. The value at this coordinate (x, y) is proportional to the image brightness at that location. For a multispectral image, $f(x, y)$ is a vector with each component specifying the luminance of the scene at point (x, y) in the associated spectral band. A digital picture is a discretized image $f(x, y)$ in both spatial and brightness coordinates. An array of 2D integers, one for each color band, is used to describe it. Gray level refers to a digitized brightness value. Pixel or pel is a term derived from the phrase “picture element” for each element of the array. The dimension of such an array is usually a few hundred pixels by a few hundred pixels, and there are dozens of different gray levels to choose from. The notable contribution of our article can be interpreted as follows:

- We try to cover and discuss in detail all two important domains of image processing, i.e. image segmentation and image classification [5].
- This chapter focuses on various state-of-the-art deep-learning (DL) techniques that should be acknowledged by the research fraternity before diving into image processing.
- Comprised state-of-the-art models are described based on their application through domain-specific research articles. To the best of our knowledge, we only selected the introductory article for a specific application in a particular domain.
- Through this chapter, we try to convey the advantages and limitations of different popular models while applying them in these two domains of image processing.

The organization of the rest of the chapter is based on the above-mentioned four applications in image processing. While Section 2.1 have already discussed the introductory part of image processing, Section 2.2 introduces the domain of image segmentation along with its types, advantages, and applications. In a very similar fashion, Section 2.3 intuitively elaborates the image classification. In the very last part, this chapter is concluded with concluding remarks which are briefed in Section 2.4.

Machine Learning Algorithms for Signal and Image Processing, First Edition.
Edited by Deepika Ghai, Suman Lata Tripathi, Sobhit Saxena, Manash Chanda, and Mamoun Alazab.

2.2 Image Segmentation

Image segmentation, also known as pixel-level categorization, is the process of grouping those portions together that correspond to the same object class. Detection is the process of locating and recognizing objects. Image segmentation is a pixel-level predictive process, where each pixel is differentiated into its classification category. Literature can be consulted for more information. Semantic picture segmentation can be used for a variety of tasks, including recognizing road signs, segmenting colon crypts, and classifying land use and land cover. It's also employed in the medical industry for things like detecting brains and tumors, as well as detecting and monitoring medical tools during procedures. Several medical implications of segmentation are listed. Scene parsing is critical in self-driving cars and advanced driver assistance systems (ADASs) because it mainly relies on image segmentation. The accuracy of segmentation has improved dramatically since the re-advent of the deep neural network (DNN) [6]. The methods used before DNN are referred to as traditional approaches. In this sections (Section 2.2), we briefly discuss some popular and widely implemented DL-based segmentation algorithms.

2.2.1 Types of DL-Based Segmentation

2.2.1.1 Instance Segmentation Using Deep Learning

Instance segmentation has surfaced as one of the demanding and complex computer vision research topics. It locates various classes of object instances existing in different photos by anticipating the object class label prediction and the pixel-based object instance-mask anticipation. Instance segmentation application is intended to aid robotics, autonomous vehicles, and surveillance. Many instance segmentation approaches were suggested with the introduction of DL, notably convolutional neural networks (CNNs) [7], for example, in which the segmentation accuracy continued to grow. Mask regional convolutional neural network (R-CNN) is a simple and effective method for segmenting instances. A fully convolutional network (FCN) has been used to predict segmentation masks alongside box-regression and object classification, following the lead of the fast/faster R-CNN. To extract stage-wise network features with high efficiency, a feature pyramid network (FPN) was developed, in which a top-down network path with lateral connections was used to produce semantically strong features. Some relatively recent datasets provide ample opportunity for the proposed methodologies to be improved. The common objects in context (COCO) collection from Microsoft contains 200k pictures. In this dataset's photos, many examples with complicated spatial layouts have been documented. Additionally, the cityscapes dataset and the mapillary vistas dataset (MVD) both offer street scene photos with a huge number of traffic objects per image. These dataset's photos contain blurring, occlusion, and minute examples. Many principles for network architecture for image classification have been offered. Object recognition can benefit greatly from the same. Narrowing the information path, employing dense connections, and improving the versatility and diversity of the information path by creating parallel paths are other examples.

2.2.1.2 Semantic Segmentation Using Deep Learning

The objective of semantic-image segmentation, also known as pixel-level categorization, is to group those sections of a picture that correspond to the very same object class. Modern semantic segmentation algorithms are built on the foundation of fully convolutional networks (FCNs). They frequently forecast outcomes with lower resolution than the input grid and retrieve the additional 8–16 resolution using bilinear upsampling. Dilated/atrous convolutions, which substitute some subsampling layers at the cost of greater memory and processing, may improve outcomes. Encoder–decoder architectures that subsample the grid representation in the encoder and then upsampled it in the decoder, employing skip connections to retrieve filtered features, are an alternative technique. Before performing bilinear interpolation, current techniques combine dilated convolutions with an encoder–decoder architecture to perform operations on four finer grids than the input grids. We present a method for efficiently predicting fine-level details on a grid as dense as the input grid in our chapter.

2.2.2 Advantages and Applications of DL-Based Segmentation

Satellite pictures have been effectively segmented using DL-based segmentation methods in remote sensing applications and mostly in urban planning and precision agriculture. To solve major environmental issues, such as climate change or any meteorological issues, aerial images are taken by drones and airborne equipment and gone under DL-based segmentation algorithms. The larger image size and the scarcity of ground-truth data required to assess the accuracy of the segmentation algorithms are the primary stumbling block. Similarly, in the assessment of building materials, DL-based segmentation approaches confront issues due to a large number of relevant pictures and the lack of reference information. Last but not least, biological imaging is an essential application sector for DL-based segmentation. The potential here is to create standardized image databases that can be used to evaluate novel infectious diseases and track pandemics.

2.2.3 Types and Literature Survey Related to DL-Based Segmentation

DL-based image segmentation techniques can be grouped into nine categories according to their model architecture.

2.2.3.1 Fully Convolution Model

A milestone step toward convolutional segmentation is proposed in [8], as shown in Figure 2.1, comprising a fully connected network ~~(FC Network)~~, which includes only convolutional layers to output a segmentation map from the same sized input layer.

AQ1

This convention was modified and resulted in VGG16, ParseNet, and GoogleNet like frameworks where the final FC layer is cut off from the mainframe to give spatial segmentation map from an arbitrary sized input image instead of a classification score. The skip connection technique came into play where upsampled final layer feature maps got fused with earlier layer resulting in semantic information recombination to achieve a detailed and accurate segmentation technique. The state-of-the-art performance can be achieved by testing on PNYUDv2, PASCAL VOC, and SHIFT Flow in the domain of brain tumor segmentation, iris segmentation, skin lesion segmentation, and instance aware segmentation.

2.2.3.2 CNN with Graphical Model

Due to scene-level semantic context ignorance of FCN, many researchers combine FCN with a probabilistic graphical model like conditional random fields (CRFs) and Markov random fields (MRFs) that can localize image

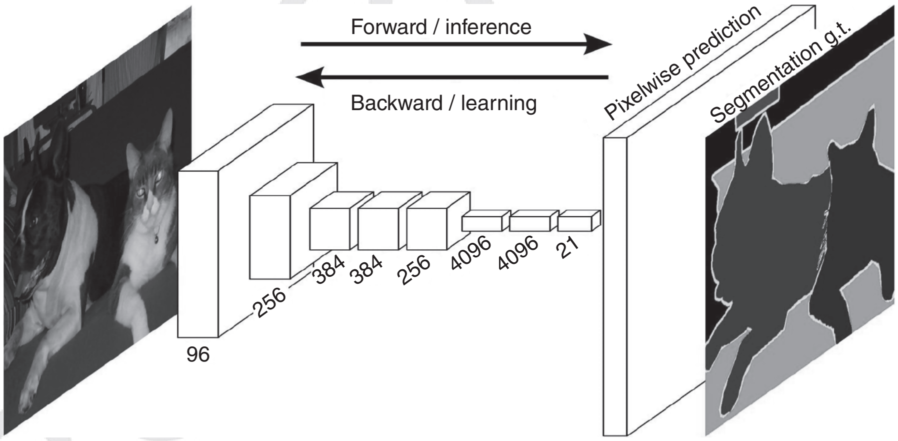

Figure 2.1 A pixel predictive model based on FCN. g.t, means the ground truth. Source: Long et al. [8], IEEE.

segment boundaries more accurately than previous methods. The invariance features of CNNs make deep CNN suitable for high-level tasks like classification. And due to this property, the responses from the deeper layers are not sufficiently localized enough for successful object segmentation task. Chen et al. [9] presented a segmentation approach that integrates both CNN and CRF techniques to address this flaw. Other related proposed architectures are

- Joint training method using CNNs and fully connected CRFs proposed by Schwing and Urtasun [10].
- Contextual deep CRFs model explores the patch-background context and patch–patch context in semantic segmentation proposed by Lin et al. [11].
- Parsing network, a CNN model, enables end-to-end analysis in a single go proposed by Liu et al. [12]. Latter this rich information incorporates optimized MRFs to yield satisfactory results.

2.2.3.3 Dilated Convolution Model

The idea behind dilated convolution is to boost the convolutional mask receptive field without an increase in the computational cost. Some real-time segmentation models use this dilated technique e.g. DeepLab family [13], efficient network (Enet) [14], dense atrous spatial pyramid pooling (DenseAPP) [15], dense up-sampling convolution and hybrid dilated convolution (DUC-HDC) [16], and multiscale context aggregation [17].

- Proposed by Chen et al., DeepLabV1 [9] and DeepLabV2 [13] are registered as an efficient and popular image-segmentation technique that has three special features as shown in Figures 2.2 and 2.3 respectively. Dilated convolution mitigates the decreasing resolution due to striding and max pooling, strous spatial pyramid pooling (ASPP) filters the convolution layer at a distinct sampling rate, and combining the structure of DCNN (ResNet or VGGNet) with probabilistic graphical model improves the segmenting mask localization.
- In a very similar fashion, Chen et al. again proposed DeepLabV3 [19] and DeepLabV3+ [18] by combining cascaded and parallel dilated units and by incorporating encoder–decoder architecture with dilated separable convolution unit, respectively.

2.2.3.4 Encoder–Decoder Model

This is the most popular approach in semantic segmentation encompassing DeConvNet, SegNet, HRNet, stacked deconvolutional networks (SDNs), LinkNet, W-Net, U-Net, and V-Net.

- Noh et al. [20] proposed the encoder–decoder DeConvNet model, as shown in Figure 2.4, which recognizes pixel-wise class labels and predicts segmentation layer. This model has multiple convolutions and deconvolution layers adopted from VGG16 along with the pooling and unspooling layer to find pixel-accurate class probability.

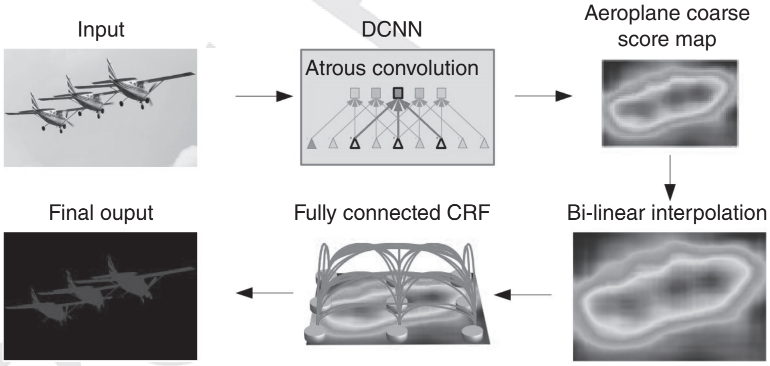

Figure 2.2 The DeepLab architecture. Source: Chen et al. [13], IEEE.

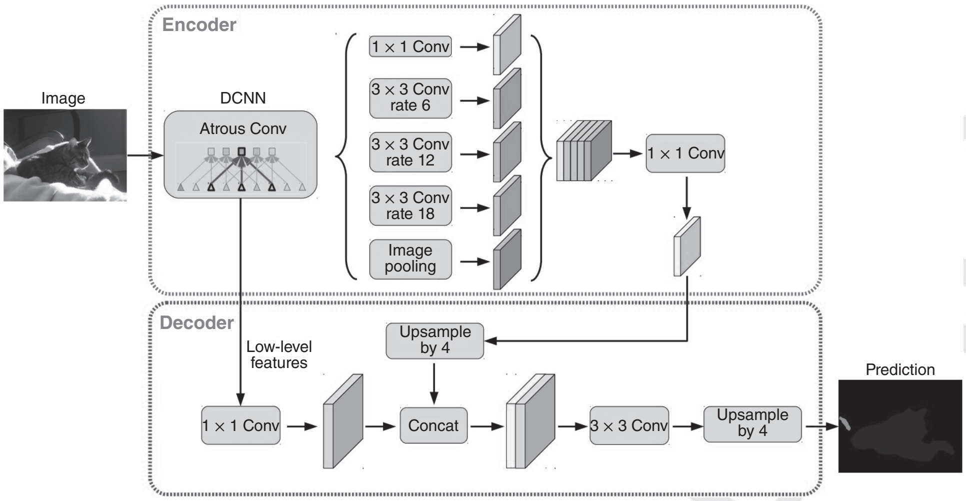

Figure 2.3 The DeepLab-V3+ model architecture. Source: Visin et al. [18], IEEE.

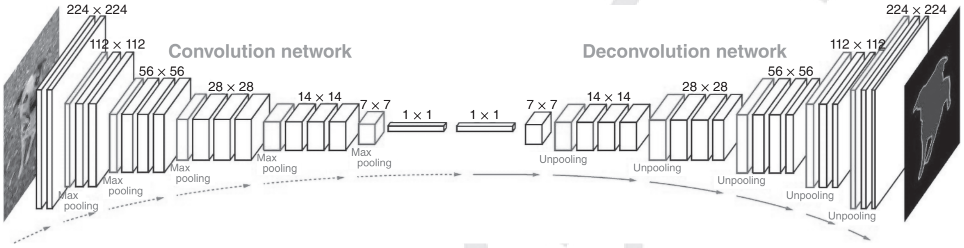

Figure 2.4 Deconvolutional segmentation (semantic). Source: Noh et al. [20], IEEE.

- Another encoder–decoder architecture derived from VGG16 has a novelty of low-resolution input feature upsampling by max-pooling layer. This model, SegNet as shown in Figure 2.5, is proposed in [21]. It has 13 Conv–Deconv layers and a pixel classification layer.
- Recovering from fine-grained information loss in image resolution during encoding as seen in DeConvNet and SegNet, HRNet not only parallelly connects high-to-low resolution Conv. Stream but also exchange information among themselves.
- For biological microscopy segmenting images, U-Net was proposed by Ronneberger et al. [22] which comprises two paths; one contracting path which is responsible for context capturing and one asymmetric expanding path responsible for localization.
- Milletari et al. [23] proposed V-Net to eradicate the problems that arise due to the voxel count difference between foreground and background by introducing a new loss function taking dice co-efficient.

2.2.3.5 R-CNN Based Model

Initially proposed for object detection, R-CNN uses a region proposal network that extracts region of interest (ROI) and ROI pool layer for coordinate class interference.

- Mask R-CNN, as shown in Figure 2.6, runs on three parallel lines of processing concept which incorporates bounding box computing line, class prediction line, and binary mask computing line. Initially proposed in [24], it shows its efficient capability in object detection and segmentation.

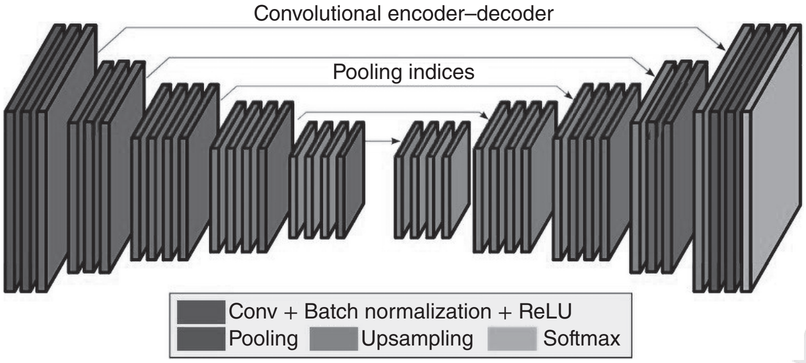
Convolutional encoder–decoder
Pooling indices
Conv + Batch normalization + ReLU
Pooling
Upsampling
Softmax

Figure 2.5 The SegNet framework. Source: Badrinarayanan et al. [21].

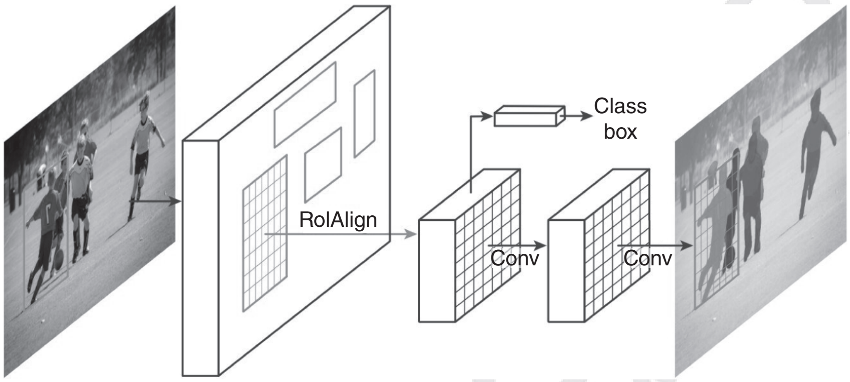

Figure 2.6 Mask R-CNN model architecture. Source: He et al. [24], IEEE.

- Referring to the concept of FPN and masked R-CNN, Liu et al. [25] proposed path aggregation network (PANet) which uses augmented bottom-up pathway with FPN backbone for lower layer feature improvement.
- Multitask network by Dai et al. [26], MaskLab by Chen et al. [27], DeepMask [28], TensorMask by Chen et al. [29], PolarMask [30], ~~RFCN~~ [31], and CenterMask [32] are among popular R-CNN techniques which do not only yield an efficient detection technique but also produce an accurate segmenting mask.

2.2.3.6 Multiscale Pyramid Based Model

Many neural network (NN) models apply multiscale and pyramid-like structures for segmenting images. Among, FPN [33], dynamic multiscale filter network (DM-Net), pyramid scene parsing network (PSPN) [34], Laplacian pyramid structure [35], adaptive pyramid context network (APC-Net) [36], context contracted network (CCN) [37], salient object segmentation [38], and multiscale context intertwining (MSCI) [39] are most popular.

- FPN was initially proposed for object detection by Lin et al. [33] and later applied for segmentation purposes by merging low and high resolution features through a top-down pathway, a bottom-up pathway, and lateral connections.
- Aiming for global context representation of an image residual network (ResNet) is used as feature extractor in PSPN, as shown in Figure 2.7, which was proposed by Zhao et al. [34]. Extracted feature map first passed through pooling module for distinct (four) pattern study, then a 1 × 1 Conv layer for dimensionality reduction, and lastly through an upsampling and concatenating unit to yield local as well as global context information. Above mention, three-layer is considered as pyramid pooling modules which later combine with a convolution layer to generate pixel-wise predicted feature maps.

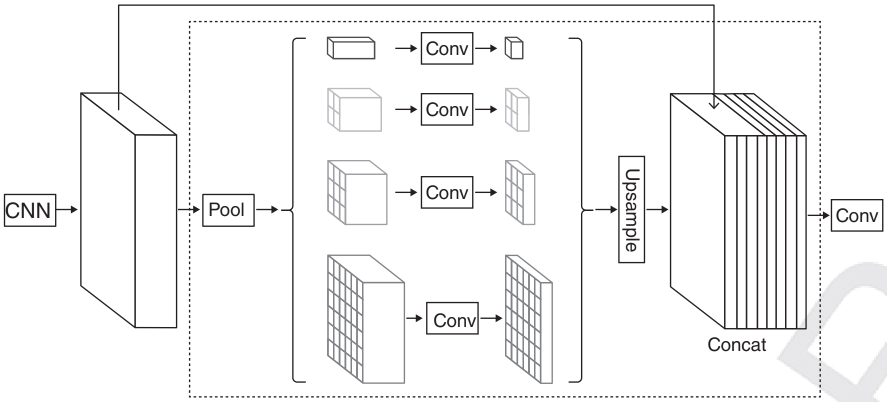

Figure 2.7 The PSPN model architecture. Source: Zhao et al. [34].

2.2.3.7 RNN Based Model

Like CNN, recurrent neural network (RNN) also can improve estimated segmenting maps by exploiting the pixel level short/long term dependencies.

- ReSeg, as shown in Figure 2.8, is a popular RNN based model proposed in [18] which is composed of an image classification model, ReNet [40]. The combination of four RNNs in ReNet is responsible for vertical and horizontal image sweeping, patch/activation encoding, and global information gathering. ReNet is stacked upon the VGG16 convolution layer to extract local descriptions. After that, the initial image resolution is recovered through up-sampling.
- 2D ~~LSTM~~ models, proposed in [41], have a greater role in per-pixel image segmentation by learning complex spatial dependencies and textures. This model successfully carries out context integration, segmentation, and classification as well.
- Another approach toward semantic segmentation framework is the graph-LSTM network, as shown in Figure 2.9, proposed in [42] that can yield more structural and global information by augmenting the convolutional layers through graph-LSTM layers. Array structured uniformly arranged data are generalized to graph-structured nonuniform data by taking arbitrary superpixels as consistent nodes and existing relation between superpixels as the edge in an undirected graph.

Apart from these models, data associated recurrent neural network (DA-RNN) proposed by Xiang and Fox [43] in semantic labeling and 3D joint mapping and semantic segmentation algorithm after combining spatial CNN image description and temporal LSTM linguistic description proposed by Hu et al. [44] are popular RNN based algorithm. The sequential nature of RNN does not permit for parallel execution and therefore RNN based models are slower than their CNN counterparts.

2.2.3.8 Generative Adversarial Network (GAN) Based Model

Nowadays generative adversarial networks (GANs) are widely accepted DL techniques in the computer vision domain, not excluding segmentation. Luc et al. [45] proposed a conjoint model by incorporating a combination of

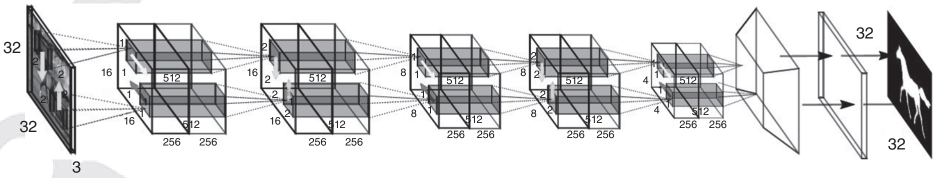

Figure 2.8 The ReSeg framework. Source: Visin et al. [18], IEEE.

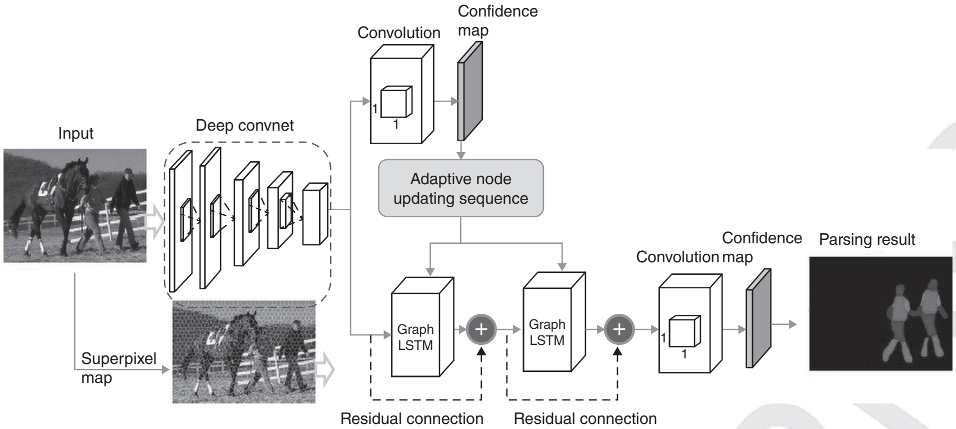

Figure 2.9 The graph-LSTM semantic segmentation model. Source: Liang et al. [42], Springer Nature.

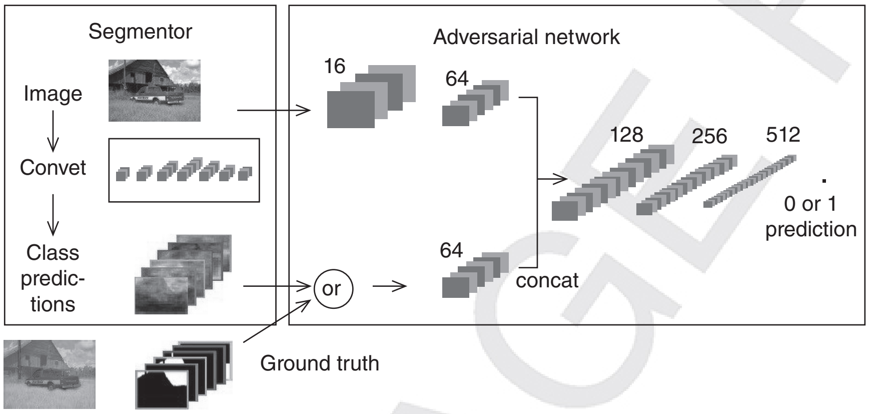

Figure 2.10 The GAN-based semantic segmentation model. Source: Luc et al. [45], Cornell University.

convolutional segmentation network and adversarial network which distinguishes between generated segmentation map and ground truth segmentation map as shown in Figure 2.10.

- A multiclass classifier or otherwise act as a discriminator using a semi-weekly supervised GAN network was proposed by Souly et al. [46] that either assigns a foreign sample a possible label from trained classes or strike it as an irregular class.
- Another semi-supervised semantic segmentation GAN framework is surfaced by Hung et al. [47] after designing an FCN discriminator that differentiates between predicted segmentation map and ground truth. This spatial resolution-based network has three-loss function terms e.g. adversarial loss, cross-entropy loss, and semi-supervised loss of the segmentation ground truth, discriminator network, and confidence map output respectively.
- Other GAN-based approaches include multiscale adversarial networks [48], cell image GAN segmentation [49], and invisible part segmenting and generating network [50] mostly in the medical image domain.

2.2.3.9 Segmentation Model Based on Attention Mechanism

- Chen et al. [51] weighted multiscale features at every pixel location and proposed a powerful attention-based semantic segmentation model, as shown in Figure 2.11, which assigns distinct weights based on their importance at different scales and positions.

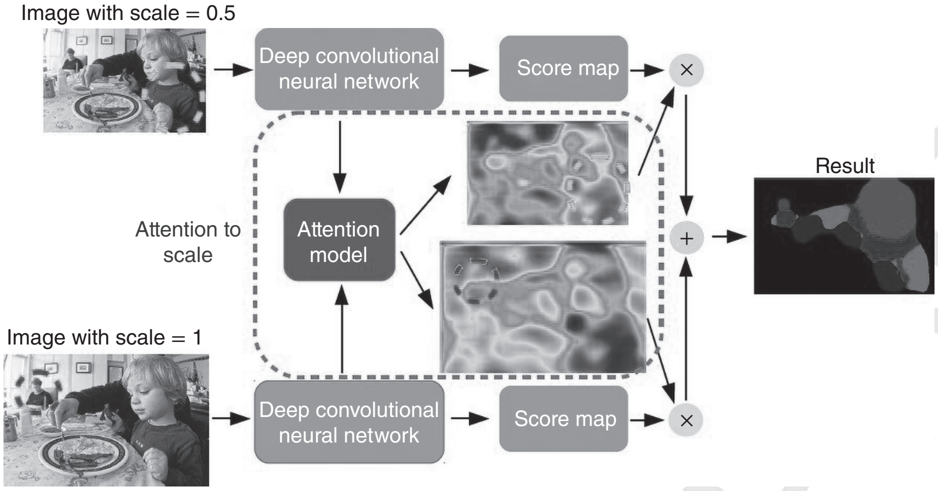

Figure 2.11 Attention-based model for semantic segmentation. Source: Chen et al. [51], IEEE.

- Li et al. [52] incorporated attention mechanism with spatial pyramid technique, which yields more precise dense features for image-pixel labeling. This model is otherwise known as pyramid attention network (PAN).
- Reverse attention network (RAN) studies the reverse concept features which are not linked with the target class by training the model through a reverse attention mechanism. Proposed by Huang et al. [52], RAN performs both direct and reverse attention to study semantic features.
- Due to persistent exploration, many models like dual attention network (DAN) [53], OCNet [54], ResNest: split attention network [55], criss cross attention network (CCNet) [56], discriminative feature network (DFN) [57], expectation–maximization attention network (EMANet) [58], and height driven attention network (HAN) shows prominent accuracy in attention-based semantic segmentation.

2.3 Image Classification

Image-level classification assigns a particular label to every input image rather than doing pixel-wise prediction. Unsupervised feature learning such as sparse auto-encoders (SAEs) and GANs has been utilized to solve various computer vision problems, and it can be used in medical imaging. The use of neural networks as classifiers, which directly output an individual prediction for one image, is a simple technique to classify images. Alternatively, a network can be employed as feature extractor to build data representations that are supplied to other target classifiers after they have been trained with large-scale data sets.

2.3.1 Types and Schemes in Image Classification

In DL, unsupervised learning and supervised learning have been two intertwined significant themes. Pretraining a DNN, which was subsequently fine-tuned with supervised tasks, was one of the important applications of deep unsupervised learning over the last decade. Stack (denoising) autoencoders, deep belief networks, sparse encoder–decoders, and deep Boltzmann machines are only a few of the deep unsupervised models presented. When the number of available labels was limited, these methods dramatically enhanced the performance of neural networks on supervised tasks.

However, in recent years, supervised learning without any unsupervised pretraining has outperformed unsupervised learning, and it has emerged as the most popular method for training DNNs for practical works like image categorization and object recognition. Purely supervised learning permitted network topologies, such as the inception unit and residual structure, to be more flexible, since they were not constrained by the modeling norms of unsupervised methods. Furthermore, the batch normalization approach has made neural-network learning much easier.

There have been several efforts to combine supervised and unsupervised learning in the same chapter, allowing unsupervised ideas to influence network training after supervised learning. Although these procedures have opened up new possibilities for unsupervised learning, they have yet to be demonstrated to escalate to large counts of labeled and unlabeled data. Many recent papers projected an architecture that is trouble-free to couple with a classification network by outspreading the stacked denoising autoencoder with lateral links, i.e. from the encoder to the same layers of the decoder, and their methods yield promising semi-supervised learning results.

2.3.2 Types and Literature Survey Related to DL-Based Image Classification

Various structures for completing the image-classification problem are enlisted as follows: CNN-based classifier, CNN–RNN-based classifier, auto-encoder-based classifier, and GAN-based classifier.

2.3.2.1 CNN Based Image Classification

- **LeNet:** The CNN was developed by LeCun et al. [59] in 1998 to categorize handwritten digits. LeNet-5 CNN model includes seven weighted (trainable) layers. Three layers of the convolutional block, two layers of average pooling block, an FC layer, and an output layer are among them. Non-linear features from an image are exploited first, followed by the down-sampling operation through the pooling layer. Several consecutive units of Euclidean radial basis function (RBF) are used to categorize 10 digits at the output end of the framework. LeCun et al. [59] used LeNet-5 to train and test the MNIST handwritten digits dataset. MNIST dataset comprises 60k and 10k records for training and testing, respectively. The authors used a stochastic gradient descent (SGD) technique to train various variants of LeNet-5 architecture was set with 20 iterations, a momentum of 0.02, and a lower global learning rate for training in each session.
- **AlexNet:** In 2012, Krizhevsky et al. [60] proposed a deep CNN network, AlexNet, for classifying ImageNet data. AlexNet has the same design as LeNet-5, but it is substantially larger. It consists of eight trainable layers. Five Conv layers and three FC layers are included among them. After convolutional and FC layers, they used rectified linear unit (ReLU) non-linearity to assist their model train quicker than equivalent networks with the units. After the first and second convolutional layers, they employed local response normalization (LRN), also known as "brightness normalization," to enhance generalization. The LRN layer and convolutional layer (fifth) are followed by a max-pooling layer that down-sample the dimension of each feature. For the ILSVRC – 2010 and ILSVRC – 2012, [60] created AlexNet to classify 1.2 million photos into 1000 classes. It used 256 × 256 pixels images after down-sampling and centering from ImageNet's variable resolution image. They employed runtime data augmentation and a regularization method termed dropout to minimize the overfitting issue. Both ~~PCA~~ and data augmentation techniques are employed for feature reduction and new sample generation purposes, respectively. These translated and horizontally sifted augmented images are retrieved in the form of 10 random patches of dimension 224 × 224. The authors used SGD to learn AlexNet, which had a batch size of 128, weight decay of 0.0005, and momentum of 0.9.
- **ZFNet:** Zeiler and Fergus [61] presented a CNN architecture called ZFNet in 2014. AlexNet and ZFNet are congruent architectures, except the filter size of the first layer reduce from 11 × 11 to 7 × 7. To keep more features, the convolution operation of stride 2 is employed in the first two convolutional layers. It explained the extraordinary performance of deep CNN in the image domain. Another visualization technique, called DeconvNet, is used to reconstruct the activations at higher layers back to the space of the input pixel where it originates. The ImageNet dataset comprising 1.3 million training, 50k validation, and 100k testing images, was used by ZDNet. The hierarchical nature of extracted features in the CNN network can be visualized by projecting each layer in increasing order.

- **VGGNet:** Simonyan and Zisserman proposed VGGNet based on a deeper configuration of AlexNet [62]. This deeper network has kept all the parameters constant except the filter size of the convolution layer. The size of the convolutional filter is fixed to 3 × 3 throughout the deep network. Initially, they employed four CNN configurations, such as A, A-LRN, B, and C, having four distinct weighted layers of count 11, 22, 13, and 16, respectively. These four CNN configurations are followed by two variants of VGG configuration named D or VGG16 and E or VGG19. To improve non-linearity in model C, three convolutional filters of size 1 × 1 each is replaced in the 6th, 9th, and 12th position. A pack of three 3 × 3 convolutional filters has a similar effective receptive field but equipped with more additive non-linearity than a single 7 × 7. Therefore single 7 × 7 filter is replaced by a group of three 3 × 3 filters. Unlike AlexNet, the input image is resized to a dimension of 224 × 224 in VGGNet.
- **GoogLeNet:** Szegedy et al. [63] developed a novel design for GoogLeNet, which differs from traditional CNN. They used parallel filters termed inception modules of sizes 1 × 1, 3 × 3, and 5 × 5 in each convolution layer to increase the number of units in each convolution layer. The total number of layers in GoogLeNet has become increased to 22 compared to 19 in VGGNet. They kept the computational budget in mind when building this model. To handle several scales, they implemented a sequence of weighted Gabor filters of varied sizes in the inception design. Instead of using the naïve form of the inception module, they employed the inception module with a dimension reduction technique to make the architecture computationally efficient. The disbelief distributed machine learning framework was used to train GoogLeNet. The training process has been accomplished with a reasonable amount of structure and data-processing parallelization. During optimization, the following network has used SGD (asynchronous) as an optimizer with a momentum of 0.9 at a constant learning rate.
- **ResNet:** The vanishing gradient has become a very prominent problem while increasing the number of layers in CNN model. Normalization and intermediate initialization techniques combinedly address these issues very effectively. But, the performance still degrades and this is not due to overfitting. To solve this issue, He et al. [64] introduced a pre-trained shallower model that coupled with an extra layer for identity mapping. The combination of both are sequentially executed to form ResNet architecture. The residual mapping, represented by $H(x) = F(x) + x$, replaces the previous underlying mapping $H(x)$. ResNet model optimization is carried out by SGD as an optimizer with a batch size of 128. The other training parameters, such as weight decay, learning rate, and momentum, are set as 0.0001, 0.1, and 0.9, respectively. At 32k and 48k iterations, the learning rate is manually updated by lowering the initial rate (0.1) and finally stopped at 64k iterations. After each Conv layer, they employed weight initialization and batch normalization. The dropout regularization approach was not used.
- **DenseNet:** Dense Convolutional Networks (DenseNets) were proposed by Huang et al. [65] and included dense blocks in standard CNN. In a dense block, the input of one layer is the concatenation of the outputs of all of the previous levels. The features of earlier layers are reused iteratively to improve the feature quality and lower the vanishing gradient problems. The number of parameters was also minimized by using a modest number of filters. The non-linear transformation functions in a dense block are a combination of batch normalization, ReLU, and the 3 × 3 convolution operation. To minimize dimensionality, they used the 1 × 1 bottleneck layer. The DenseNet framework is pre-trained on CIFAR and ImageNet datasets. This architecture uses SGD as an optimizer with batch sizes of 256 and 64 on the SVHN and CIFAR datasets, respectively. The initial learning rate was 0.1 and reduced by 1/10 twice. They employed 0.0001 weight decay, a 0.9 Nesterov momentum, and a 0.2 dropout.
- **CapsNet:** Conventional CNNs has two flaws, as detailed above. The sub-sampling, for starters, eliminates spatial information among higher-level features. Second, it has a hard time generalizing to new viewpoints. It can handle translation but not affine transformations of different dimensions. Geoffrey E. Hinton and coworkers proposed CapsNet [66] in 2017 as a solution to these issues. CapsNet includes capsule components. A capsule is made up of a collection of neurons. CapsNet layers are essentially made up of layered neurons. Sabour et al. introduced the CapsNet, which is made up of three layers: two Conv levels and one FC layer. The first convolutional layer employs ReLU as an activation function and consists of 256 convolutional units with 9 × 9 kernels

of stride 1. This layer identifies local features and delivers them as input to the second layer's major capsules. Each primary capsule holds eight CU, with a stride of 2 kernels. The primary capsule layer comprises a total of 32 6 × 6 8D capsules. Each digit class has one 16D capsule in the final layer. Between both the primary layer and the DigitCaps layer, the authors employed routing. The MNIST photos are used to train CapsNet. As a regularization strategy, they employed reconstruction loss.

2.3.2.2 CNN–RNN Based Image Classification

Two alternative structures for completing the image-classification problem are: CNN-based classifier and CNN–RNN classifier. Unlike CNN-based generators, CNN–RNN generators can efficiently manipulate the hierarchical labels' dependency, resulting in improved classification results for both fine and coarse classes. As can be observed, this generator outperforms the previous CNN-based generator in terms of fine and coarse grade predictions with and without the help of the data augmentation technique. In particular, this CNN–RNN generator generates out classes the previous CNN-based generator by at least 5.05% for coarse predictions, while by over 6.7% for fine predictions. Figure 2.12 shows the framework of the CNN–RNN pipeline.

- Guo et al. [67] used the dropout version, a bigger mini-batch size (200), and more iterations (7 × 104 in total, with the learning rate dropping at 2 × 104, 4 × 104, and 6 × 104 iterations) to train the network as shown in Figure 2.12. The following are some more experimental setups. This chapter uses the pre-activation residual block and trains the models for 7 × 104 iterations with a mini-batch size of 200, a momentum of 0.9, and a weight decay of 0.0005. The learning rate starts at 0.1 and decreases by 0.01 in every 4 × 104 and 6 × 104 iterations.
- The proposed CNN–RNN model is a CNN and RNN combined model for image categorization presented by [68]. The convolution computation is a filtering procedure that treats image data as two-dimensional wave data. It can remove non-critical band information from an image while leaving crucial aspects intact. The RNN module is used to calculate the dependency features of intermediate layer output and connect the characteristics of these intermediary tiers to the final full-connection network for classification prediction to improve classification accuracy. At the same time, to satisfy the RNN model's limitation on the size of the input sequence and avoid gradient explosion or disappearance in the network, this research filters the input data using the wavelet transform (WT) method in combination with the Fourier transform. This research will put the suggested CNN–RNN model to the test on the CIFAR-10 dataset, which is extensively utilized.

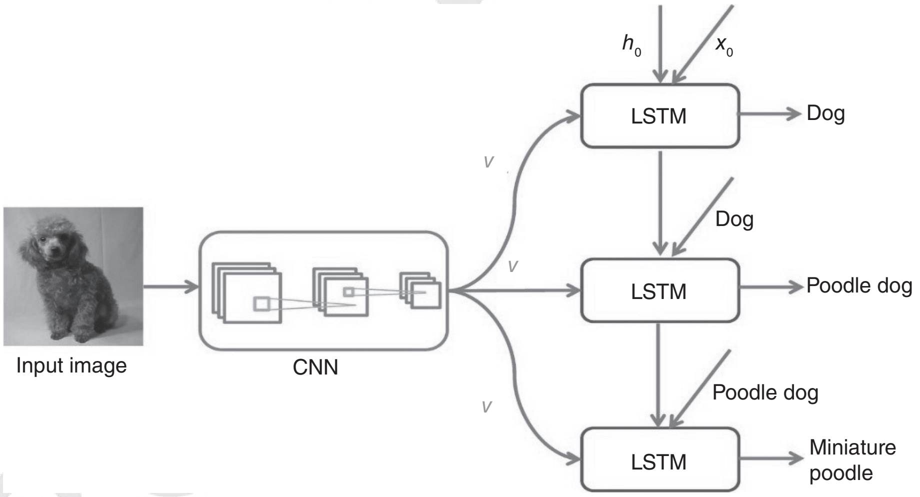

Figure 2.12 The CNN–RNN pipeline framework. Source: Guo et al. [67], CC BY, Springer Nature.

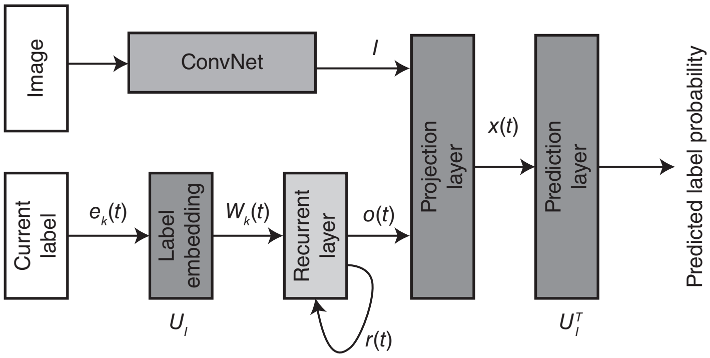

Figure 2.13 The RNN model for multi-label classification. Source: Wang et al. [69].

- Wang et al. [69] used an end-to-end approach to take advantage of semantic redundancy and co-occurrence dependency, which are both necessary for effective multi-label classifications (Figure 2.13). In this challenge, the recurrent neurons model of high-order label co-occurrence dependency is more concise and efficient than other label co-occurrence approaches. The recurrent neurons' implicit attention mechanism adapts picture characteristics to accurately predict small things that require more context.

2.3.2.3 Auto-encoder Based Image Classification

When compared to patch-based sparse coding, convolutional sparse coding (CSC) can simulate local links between visual content and minimize code redundancy. CSC, on the other hand, requires a time-consuming optimization technique to infer the codes (i.e. feature maps).

- Luo et al. [70] introduced a convolutional sparse auto-encoder (CSAE) in this paper, which takes advantage of the convolutional ~~AE's~~ structure and integrates max-pooling to empirically sparsify feature maps for feature learning. This basic sparsifying method, along with competition over feature channels, allows the SGD algorithm to perform quickly for CSAE training; consequently, no complicated optimization process is required. The features learned in the CSAE were used to initialize CNNs for classification, and the results were competitive on benchmark data sets. Lou et al. also offered an approach for constructing local descriptors from the CSAE for categorization by establishing linkages between the CSAE and the CSC. Experiments on Caltech-101 and Caltech-256 proved the efficacy of the suggested strategy and confirmed that the CSAE as a CSC model can investigate connections between nearby image information for classification tasks.
- Liang et al. [71] offered a new approach for image classification remotely built upon stacked denoising autoencoder due to the accuracy bottleneck nature of conventional remote image classification methods. The deep network model is first constructed using denoising autoencoder's stacked layers. The unsupervised greedy layer-wise training algorithm is then used with noised input to train every layer in turn for a more robust expression. Information is retrieved in supervised learning using a back propagation (BP) neural network, and the entire network is optimized using error backpropagation. The total and kappa accuracy are both greater than those of the ~~SVM~~ and back propagation NNs by a margin of 95.7% and 0.955 respectively processed on Gaofen-1 satellite data. The experimental outcomes recommend that this strategy may effectively increase remote sensing picture categorization accuracy.

2.3.2.4 GAN Based Image Classification

- Kong et al. [72] presented a unique technique for obtaining data labels cost-effectively without querying an oracle. To evaluate the uncertainty level, a new reward for each sample is developed in the model. It is generated through a trained classifier on older labeled data. A conditional GAN is guided by this reward to create useful samples with a greater probability for the specific label. The efficiency of the model has been proven through comprehensive assessments, demonstrating that the formed samples are competent enough to increase performance in common picture classification assignments. To evaluate the proposed model, it is employed on four datasets: CIFAR-10, Fashion-MNIST, MNIST, and a large size dataset Tiny-ImageNet. For each of the 10 broad classes, CIFAR-10 features colorful pictures.

- Zhu et al. [73], for the first time, investigate the utility and efficacy of GAN for hyperspectral image classification. A CNN is utilized to distinguish the inputs in the proposed GAN. With the help of another CNN, false inputs are also generated. The generating and discriminative CNNs are trained at the same time. The generative CNN seeks to create false inputs as realistic as possible, while the other one tries to differentiate true and false inputs. The adversarial training enhances the discriminative CNN's generalization capabilities, which is critical when training samples are restricted. This work proposes two strategies for spectral classifiers: (i) a well-designed 1D-GAN as a spectral classifier; and (ii) a robust 3D-GAN as a spectral–spatial classifier. In addition, the generated adversarial samples are combined with real training data to fine-tune the discriminative CNN, resulting in improved classification performance. Three commonly used hyperspectral data sets are used to test the suggested classifiers: Salinas, Indiana Pines, and Kennedy Space Center. In comparison to state-of-the-art approaches, the collected findings show that the proposed models deliver competitive results. Furthermore, the suggested GANs indicate the enormous potential of GAN-based approaches for the analysis of such complex and intrinsically nonlinear data in the remote sensing community, as well as the immense potential of GAN-based methods for the tough problem of ~~HSI~~ categorization.

2.4 Conclusion

This chapter went through the numerous DL approaches that are employed in various tasks along the image processing chain in detail. We discuss different models for image segmentation and image classification deeply procured from related articles. Researchers have created a large number of neural architectures in recent decades, and for each model, we explained the methodologies used for certain image-processing applications. In comparison to conventional models, hybrid models have been demonstrated to perform better in case studies. Each task of image segmentation and classification used in the model was detailed in detail, and the model that sets the highest standard was discussed briefly. Despite the excellent accuracy of traditional models, newly designed hybrid models are adapted to a specific use case. Integrating these traditional models according to their use case improves accuracy and boosts the models' performance. As a result, DL algorithms with a highly iterative and learning-based approach perform much better for image processing applications. Through this chapter, we try to convey the advantages and limitations of different popular models while applying them in these two domains of image processing. In the future, we can compare the performance of discussed architecture to conclude a more efficient model for a specific task.

References

1 Chen, Q., Xu, J., and Koltun, V. (2017). Fast image processing with fully-convolutional networks. In: *Proceedings of the IEEE International Conference on Computer Vision*, 2497–2506. IEEE. AQ2

2 Kolekar, M.H., Palaniappan, K., Sengupta, S., and Seetharaman, G. (2009). Semantic concept mining based on hierarchical event detection for soccer video indexing. *Journal of Multimedia* 4 (5).

3 Kolekar, M.H. and Sengupta, S. (2010). Semantic concept mining in cricket videos for automated highlight generation. *Multimedia Tools and Applications* 47 (3): 545–579.

4 Kolekar, M.H. and Sengupta, S. (2015). Bayesian network-based customized highlight generation for broadcast soccer videos. *IEEE Transactions on Broadcasting* 61 (2): 195–209.

5 Bhatnagar, S., Ghosal, D., and Kolekar, M.H. (2017). Classification of fashion article images using convolutional neural networks. In: *2017 Fourth International Conference on Image Information Processing (ICIIP)*, 1–6. IEEE.

6 Liu, W., Wang, Z., Liu, X. et al. (2017). A survey of deep neural network architectures and their applications. *Neurocomputing* 234: 11–26.

7 Albawi, S., Mohammed, T.A., and Al-Zawi, S. (2017). Understanding of a convolutional neural network. In: *2017 International Conference on Engineering and Technology (ICET)*, 1–6. Elsevier.

8 Long, J., Shelhamer, E., and Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In: *IEEE Conference on Computer Vision and Pattern Recognition*, 3431–3440. IEEE.

9 Chen, L.C., Papandreou, G., Kokkinos, I., et al. 2014. Semantic image segmentation with deep convolutional nets and fully connected CRFs. arXiv preprint arXiv:1412.7062.

10 Schwing, A.G. and Urtasun, R. 2015. Fully connected deep structured networks. arXiv preprint arXiv:1503.02351.

11 Lin, G., Shen, C., Van Den Hengel, A. et al. (2016). Efficient piecewise training of deep structured models for semantic segmentation. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3194–3203. IEEE.

12 Liu, Z., Li, X., Luo, P. et al. (2015). Semantic image segmentation via deep parsing network. In: *Proceedings of the IEEE International Conference on Computer Vision*, 1377–1385. IEEE.

13 Chen, L.C., Papandreou, G., Kokkinos, I. et al. (2017). Deeplab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 40 (4): 834–848.

14 Paszke, A., Chaurasia, A., Kim, S. et al. 2016. Enet: a deep neural network architecture for real-time semantic segmentation. arXiv preprint arXiv:1606.02147.

15 Yang, M., Yu, K., Zhang, C. et al. (2018). Denseaspp for semantic segmentation in street scenes. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3684–3692. IEEE.

16 Wang, P., Chen, P., Yuan, Y. et al. (2018). Understanding convolution for semantic segmentation. In: *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 1451–1460. IEEE.

17 Yu, F. and Koltun, V. 2015. Multi-scale context aggregation by dilated convolutions. arXiv preprint arXiv:1511.07122.

18 Visin, F., Ciccone, M., Romero, A. et al. (2016). ReSeg: a recurrent neural network-based model for semantic segmentation. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 41–48. IEEE.

19 Chen, L.C., Papandreou, G., Schroff, F. et al. 2017. Rethinking atrous convolution for semantic image segmentation. arXiv preprint arXiv:1706.05587.

20 Noh, H., Hong, S., and Han, B. (2015). Learning deconvolution network for semantic segmentation. In: *Proceedings of the IEEE International Conference on Computer Vision*, 1520–1528. IEEE.

21 Badrinarayanan, V., Kendall, A., and Cipolla, R. (2017). SegNet: a deep convolutional encoder–decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39 (12): 2481–2495.

22 Ronneberger, O., Fischer, P., and Brox, T. (2015). U-Net: convolutional networks for biomedical image segmentation. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 234–241. IEEE.

23 Milletari, F., Navab, N., and Ahmadi, S.A. (2016). V-Net: fully convolutional neural networks for volumetric medical image segmentation. In: *2016 Fourth International Conference on 3D Vision (3DV)*, 565–571. IEEE.

24 He, K., Gkioxari, G., Dollár, P., and Girshick, R. (2017). Mask R-CNN. In: *Proceedings of the IEEE International Conference on Computer Vision*, 2961–2969. IEEE.

25 Liu, S., Qi, L., Qin, H. et al. (2018). Path aggregation network for instance segmentation. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 8759–8768. IEEE.

26 Dai, J., He, K., and Sun, J. (2016). Instance-aware semantic segmentation via multi-task network cascades. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3150–3158. IEEE.

27 Chen, L.C., Hermans, A., Papandreou, G. et al. (2018). MaskLab: instance segmentation by refining object detection with semantic and direction features. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 4013–4022. IEEE.

28 Pinheiro, P.O., Collobert, R., and Dollár, P. 2015. Learning to segment object candidates. arXiv preprint arXiv:1506.06204.

29 Chen, X., Girshick, R., He, K., and Dollár, P. (2019). TensorMask: a foundation for dense object segmentation. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2061–2069. IEEE.

30 Xie, E., Sun, P., Song, X. et al. (2020). Polarmask: single shot instance segmentation with polar representation. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 12193–12202. IEEE.

31 Dai, J., Li, Y., He, K., and Sun, J. (2016). R-FCN: object detection via region-based fully convolutional networks. In: *Advances in Neural Information Processing Systems*, 79–387. IEEE.

32 Lee, Y. and Park, J. (2020). CenterMask: real-time anchor-free instance segmentation. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 13906–13915. IEEE.

33 Lin, T.Y., Dollár, P., Girshick, R. et al. (2017). Feature pyramid networks for object detection. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2117–2125. IEEE.

34 Zhao, H., Shi, J., Qi, X. et al. (2017). Pyramid scene parsing network. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2881–2890. IEEE.

35 Ghiasi, G. and Fowlkes, C.C. (2016). Laplacian pyramid reconstruction and refinement for semantic segmentation. In: *European Conference on Computer Vision*, 519–534. Springer.

36 He, J., Deng, Z., Zhou, L. et al. (2019). Adaptive pyramid context network for semantic segmentation. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 7519–7528. IEEE.

37 Ding, H., Jiang, X., Shuai, B. et al. (2018). Context contrasted feature and gated multi-scale aggregation for scene segmentation. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2393–2402. IEEE.

38 Li, G., Xie, Y., Lin, L., and Yu, Y. (2017). Instance-level salient object segmentation. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2386–2395. IEEE.

39 Lin, D., Ji, Y., Lischinski, D. et al. (2018). Multi-scale context intertwining for semantic segmentation. In: *Proceedings of the European Conference on Computer Vision (ECCV)*, 603–619. Springer.

40 Visin, F., Kastner, K., Cho, K. et al. 2015. ReNet: a recurrent neural network based alternative to convolutional networks. arXiv preprint arXiv:1505.00393.

41 Byeon, W., Breuel, T.M., Raue, F., and Liwicki, M. (2015). Scene labeling with LSTM recurrent neural networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3547–3555. IEEE.

42 Liang, X., Shen, X., Feng, J. et al. (2016). Semantic object parsing with graph LSTM. In: *European Conference on Computer Vision*, 125–143. Springer.

43 Xiang, Y. and Fox, D. 2017. DA-RNN: semantic mapping with data associated recurrent neural networks. arXiv preprint arXiv:1703.03098.

44 Hu, R., Rohrbach, M., and Darrell, T. (2016). Segmentation from natural language expressions. In: *European Conference on Computer Vision*, 108–124. Springer.

45 Luc, P., Couprie, C., Chintala, S. et al. 2016. Semantic segmentation using adversarial networks. arXiv preprint arXiv:1611.08408.

46 Souly, N., Spampinato, C., and Shah, M. (2017). Semi supervised semantic segmentation using generative adversarial network. In: *Proceedings of the IEEE International Conference on Computer Vision*, 5688–5696. IEEE.

47 Hung, W.C., Tsai, Y.H., Liou, Y.T. et al. 2018. Adversarial learning for semi-supervised semantic segmentation. arXiv preprint arXiv:1802.07934.

48 Xue, Y., Xu, T., Zhang, H. et al. (2018). SegAN: adversarial network with multi-scale L_1 loss for medical image segmentation. *Neuroinformatics* 16 (3): 383–392.

49 Majurski, M., Manescu, P., Padi, S. et al. 2019. Cell image segmentation using generative adversarial networks, transfer learning, and augmentations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops,* Long Beach, CA (16–20 June 2019).

50 Ehsani, K., Mottaghi, R., and Farhadi, A. (2018). SegAN: segmenting and generating the invisible. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 6144–6153. IEEE.

51 Chen, L.C., Yang, Y., Wang, J. et al. (2016). Attention to scale: scale-aware semantic image segmentation. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3640–3649. IEEE.

52 Huang, Q., Xia, C., Wu, C., Li, S., Wang, Y., Song, Y., & Kuo, C. C. J. (2017). Semantic segmentation with reverse attention. arXiv preprint arXiv:1707.06426.

53 Fu, J., Liu, J., Tian, H. et al. (2019). Dual attention network for scene segmentation. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 3146–3154. IEEE.

54 Yuan, Y., Huang, L., Guo, J. et al. 2018. OCNet: object context network for scene parsing. arXiv preprint arXiv:1809.00916.

55 Zhang, H., Wu, C., Zhang, Z. et al. 2020. ResNeSt: Split-attention networks. arXiv preprint arXiv:2004.08955.

56 Huang, Z., Wang, X., Huang, L. et al. (2019). CCNet: criss-cross attention for semantic segmentation. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 603–612. IEEE.

57 Yu, C., Wang, J., Peng, C. et al. (2018). Learning a discriminative feature network for semantic segmentation. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1857–1866. IEEE.

58 Li, X., Zhong, Z., Wu, J. et al. (2019). Expectation–maximization attention networks for semantic segmentation. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 9167–9176. IEEE.

59 LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE* 86 (11): 2278–2324.

60 Krizhevsky, A., Sutskever, I., and Hinton, G.E. (2012). ImageNet classification with deep convolutional neural networks. In: *Advances in Neural Information Processing Systems*, vol. 25, 1097–1105. IEEE.

61 Zeiler, M.D. and Fergus, R. (2014). Visualizing and understanding convolutional networks. In: *European Conference on Computer Vision*, 818–833. Springer.

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

63 Szegedy, C., Liu, W., Jia, Y. et al. (2015). Going deeper with convolutions. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1–9. IEEE.

64 He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 770–778. IEEE.

65 Huang, G., Liu, Z., Van Der Maaten, L., and Weinberger, K.Q. (2017). Densely connected convolutional networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 4700–4708. IEEE.

66 Sabour, S., Frosst, N., and Hinton, G.E. 2017. Dynamic routing between capsules. arXiv preprint arXiv:1710.09829.

67 Guo, Y., Liu, Y., Bakker, E.M. et al. (2018). CNN–RNN: a large-scale hierarchical image classification framework. *Multimedia Tools and Applications* 77 (8): 10251–10271.

68 Yin, Q., Zhang, R., and Shao, X. (2019). CNN and RNN mixed model for image classification. In: *MATEC Web of Conferences*, vol. 277, 02001. EDP Sciences.

69 Wang, J., Yang, Y., Mao, J. et al. (2016). CNN–RNN: a unified framework for multi-label image classification. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2285–2294. IEEE.

70 Luo, W., Li, J., Yang, J. et al. (2017). Convolutional sparse autoencoders for image classification. *IEEE Transactions on Neural Networks and Learning Systems* 29 (7): 3289–3294.

71 Liang, P., Shi, W., and Zhang, X. (2018). Remote sensing image classification based on stacked denoising autoencoder. *Remote Sensing* 10 (1): 16.

72 Kong, Q., Tong, B., Klinkigt, M. et al. (2019). Active generative adversarial network for image classification. In: *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. 01, 4090–4097. IEEE.

73 Zhu, L., Chen, Y., Ghamisi, P., and Benediktsson, J.A. (2018). Generative adversarial networks for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing* 56 (9): 5046–5063.

Author Queries

AQ1	Please check whether FC, D-CNN, RFCN, LSTM, MNIST, ILSVRC, PCA, CIFAR, SVHN, CU, AE, SVM, HSI need to be expanded. If yes please provide expansion for the same.	The missing acronyms are expanded in the body of this chapter.
AQ2	Please note that to maintain sequential order references have been renumbered. Kindly check and confirm.	Yes, the order of the references are correct.