International Journal of Engineering Sciences Paradigms and Researches (IJESPR)
(Vol. 32, Issue 01) and (Publishing Month: July 2016)
(An Indexed, Referred and Impact Factor Journal)
ISSN: 2319-6564
www.ijesonline.com

Seeing Deep Learning Techniques for Image Segmentation

Pravat Mallick[1]
Aryan Institute of Engineering & Technology, Bhubaneswar, Odisha
Bandana Behera[2]
Raajdhani Engineering College, Bhubaneswar, Odisha

Abstract: The AI people group has been overpowered by a plenty of profound learning–based methodologies. Many testing PC vision errands, like identification, confinement, acknowledgment, and division of items in an unconstrained climate, are by and large productively tended to by different sorts of profound neural net-works, for example, convolutional neural organizations, intermittent organizations, ill-disposed organizations, and autoencoders. Despite the fact that there have been a lot of insightful investigations in regards to the article identification or acknowledgment area, numerous new profound learning strategies have surfaced concerning picture division methods. This arti-cle approaches these different profound taking in strategies of picture division according to a logical point of view. The principle objective of this work is to give a natural comprehension of the significant methods that have made a huge commitment to the picture division area. Beginning from a portion of the conventional picture seg-mentation draws near, the article advances by depicting the impact that profound learning has had on the picture division area. From that point, a large portion of the significant division calculations have been legitimately ordered with sections devoted to their interesting commitment. With a sufficient measure of instinctive clarifications, the peruser is relied upon to have a further developed capacity to picture the inner elements of these cycles.CCS Concepts: • Computing methodologies → Image segmentation; Neural networks;

Keywords: Deep learning, semantic image segmentation, convolutional neural networks.

1 INTRODUCTION

Image segmentation can be defined as a specific image processing technique that is used to divide an image into two or more meaningful regions. Image segmentation can also be seen as a process of defining boundaries between separate semantic entities in an image. From a more technical perspective, image segmentation is a process of assigning a label to each pixel in the image such

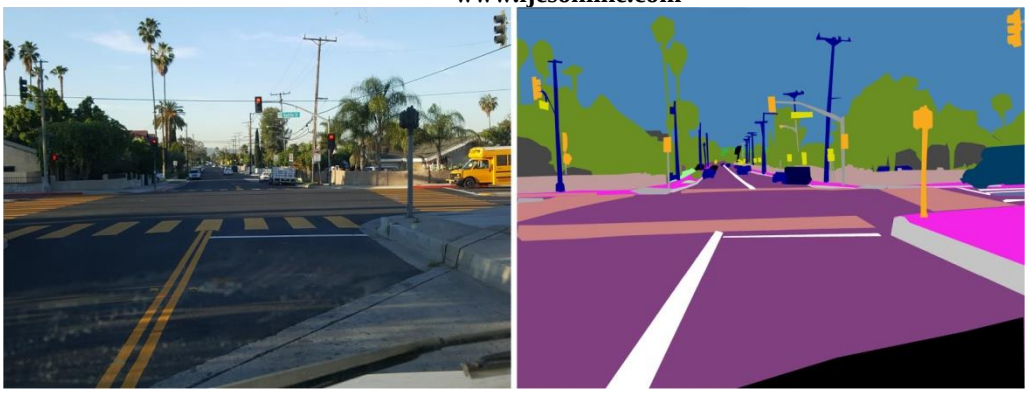

Fig. 1. Semantic image segmentation (samples from the Mapillary Vistas dataset [126]).

that pixels with the same label are connected with respect to some visual or semantic property (Figure 1).

Image segmentation subsumes a large class of finely related problems in computer vision. The most classic version is semantic segmentation [53]. In semantic segmentation, each pixel is classified into one of the predefined set of classes such that pixels belonging to the same class belong to a unique semantic entity in the image. It is also worthy to note that the semantics in question depend not only on the data but also the problem that needs to be addressed. For example, for a pedestrian detection system, the whole body of person should belong to the same segment; however, for an action recognition system, it might be necessary to segment different body parts into different classes. Other forms of image segmentation can focus on the most important object in a scene. A particular class of problem called *saliency detection* [14] is born from this. Other variantsof this domain can be foreground-background separation problems. In many systems, such as im- age retrieval or visual question answering, it is often necessary to count the number of objects. Instance-specific segmentation addresses that issue. Instance-specific segmentation is often cou- pled with object detection systems to detect and segment multiple instances of the same object[35] in a scene. Segmentation in the temporal space is also a challenging domain and has various applications. In object tracking scenarios, pixel-level classification is not only performed in the spatial domain but also across time. Other applications in traffic analysis or surveillance need to perform motion segmentation to analyze paths of moving objects. In the field of segmentation witha lower semantic level, over-segmentation is also a common approach where images are divided into extremely small regions to ensure boundary adherence, at the cost of creating a lot of spu- rious edges. Over-segmentation algorithms are often combined with region merging techniques to perform image segmentation. Even simple color or texture segmentation also finds its use in various scenarios. Another important distinction between segmentation algorithms is the needfor interactions from the user. Although it is desirable to have fully automated systems, a littlebit of interaction from the user can improve the quality of segmentation to a large extent. This is especially applicable when we are dealing with complex scenes or we do not possess an ample amount of data to train the system.

Segmentation algorithms have several applications in the real world. In medical image processing [99], we also need to localize various abnormalities such as aneurysms [39], tumors [118], cancerous elements like melanoma detection [157], or specific organs during surgeries [172]. Another domain where segmentation is important is surveillance. Many problems, such as pedestrian detection [89] and traffic surveillance [49], require the segmentation of specific objects (e.g.,

persons or cars). Other domains include satellite imagery [9, 12], guidance systems in defense [95], and forensics (e.g., face [5], iris [41], and fingerprint [117] recognition). Generally, traditional meth- ods, such as histogram thresholding [162], hybridization [69, 161] feature space clustering [32], region-based approaches [48], edge detection approaches [152], fuzzy approaches [31], entropy- based approaches [38], neural networks (Hopfield neural network [28], self-organizing maps [20]),and physics-based approaches [129], are used popularly in this purpose. However, such feature- based approaches have a common bottleneck that they are dependent on the quality of feature extracted by the domain experts. Generally, humans are bound to miss latent or abstract features for image segmentation. Yet deep learning in general addresses this issue of automated feature learning. In this regard, one of the most common techniques in computer vision was introducedby the name of convolutional neural networks (CNNs) [87] that learned a cascaded set of convo- lutional kernels through back-propagation [150]. Since then, it has been improved significantly, with features such as layerwise training [11], rectified linear activations [124], batch normaliza- tion [66], auxiliary classifiers [42], atrous convolutions [177], skip connections [63], and better optimization techniques [78]. In addition, there have been a large number of new types of image segmentation techniques. Various such techniques have drawn inspiration from popular networks such as AlexNet [82], convolutional autoencoders [115], recurrent neural networks (RNNs) [116],and residual networks [63].

2 MOTIVATION

There have been many reviews and surveys regarding the traditional technologies associated with image segmentation [50, 131]. Although some of them have specialized in application areas [84, 99, 153], others have focused on specific types of algorithms [14, 15, 48]. With the arrival of deeplearning techniques, many new classes of image segmentation algorithms have surfaced. Earlier studies [185] have shown the potential of deep learning–based approaches. There have been more recent studies [55] that cover several methods and compare them on the basis of their reportedperformance. The work of Garcia-Garcia et al. [53] lists a variety of deep learning–based seg- mentation techniques. They have tabulated the performance of various state-of-the-art networks on several modern challenges. The resources are incredibly useful for understanding the currentstate of the art in this domain. Knowing the available methods is quite useful to develop products; however, to contribute to this domain as a researcher, one needs to understand the underlying mechanics of the methods that make them confident. In the present work, our main motivation is to answer the question of why the methods are designed in the way they are. Understanding the mechanics of modern techniques would make it easier to tackle new challenges and develop better algorithms. Our approach carefully analyzes each method to understand why the methods succeed at what they do and also why they fail in certain situations. Being aware of the pros andcons of such methods, new designs can be initiated that reap the benefits of the pros and overcome the cons. We recommend the work of Garcia-Garcia [53] for an overview of some of the best imagesegmentation techniques using deep learning, whereas our focus is to understand why, when, and how these techniques perform in various challenges.

Contribution

The article has been designed in a way such that new researchers reap the most benefit. Initially,some of the traditional techniques have been discussed to uphold the frameworks before the deep learning era. Gradually, the various factors governing the onset of deep learning has been dis- cussed so that readers have a good idea of the current direction in which machine learning is pro- gressing. In subsequent sections, the major deep learning algorithms have briefly been described in a generic way to establish a clearer concept of the procedures in the mind of the readers. The

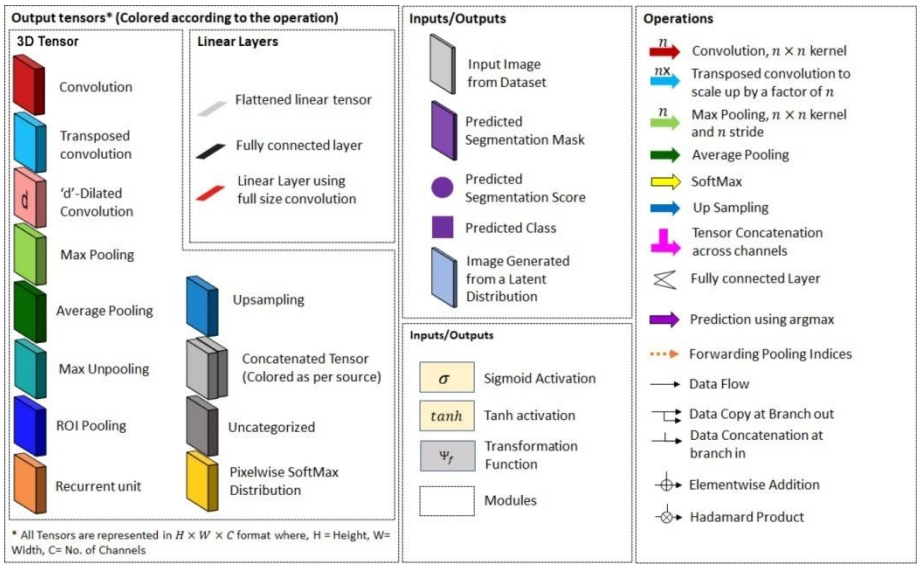

Fig. 2. Legends for subsequent diagrams of popular deep learning architectures.

image segmentation algorithms discussed thereafter have been categorized into the major families of algorithms that have governed the past few years in this domain. The concepts behind all of the major approaches have been explained through a very simple language with a minimum amount of complicated mathematics. Almost all of the diagrams corresponding to major networks have been drawn using a common representational format as shown in Figure 2. The various approaches that have been discussed come with different representations for architectures. The unified representation scheme allows the user to understand the fundamental similarities and differences between networks. Finally, the major application areas have been discussed to help new researchers pursue a field of their choice.

3 IMPACT OF DEEP LEARNING ON IMAGE SEGMENTATION

The development of deep learning algorithms like CNNs or deep autoencoders not only affect typical tasks like object classification but are also efficient in other related tasks like object detection, localization, and tracking, or, as in this case, image segmentation.

Effectiveness of Convolutions for Segmentation

As an operation, convolution can be simply defined as the function that performs a sum of product between kernel weights and input values while convoluting the smaller kernel over a larger image. For a typical image with k channels, we can convolute a smaller-size kernel with k channels along the x and y direction to obtain an output in the format of a 2D matrix. It has been observed that after

training a typical CNN, the convolutional kernels tend to generate activation maps with respect to certain features of the objects [180]. Given the nature of activations, it can be seen as segmentation masks of object-specific features. Hence, the key to generating requirement-specific segmentation is already embedded within these output activation matrices. Most of the image segmentation algorithms use this property of CNNs to somehow generate the segmentation masks as required to solve the problem. As shown in Figure 3, the earlier layers capture local features like the contour or a small part of an object. In the later layers, more global features are activated, such as field,

Fig. 3. Input image and sample activation maps from a typical CNN. (Top row) Input image and two acti-vation maps from earlier layers showing part objects like t-shirts and features like contours. (Bottom row)Activation maps from later layers with more meaningful activations like fields, people, and sky, respectively.

It can also be noted from the figure that the earlier layers show sharper activations compared to the later ones.

Impact of Larger and More Complex Datasets

The second impact that deep learning has brought to the world of image segmentation is the plethora of datasets, challenges, and competitions. These factors have encouraged researchers across the world to come up with various state-of-the-art technologies to implement segmentation across various domains. A list of many such datasets has been provided in Table 1.

5 APPLICATIONS

Image segmentation is one of the most commonly addressed problems in the domain of computer vision. It is often augmented with other related tasks like object detection, object recognition, scene parsing, image description generation. Hence this branch of study finds extensive use in various real-life scenarios.

Content-Based Image Retrieval

With the ever-increasing amount of structured and unstructured data on the Internet, development of efficient information retrieval systems is of utmost importance. Content-Based Image Retrieval (CBIR) systems have hence been a lucrative area of research. Similar interests also exist in many other related problems like visual question answering, interactive query–based image processing, description generation. Image segmentation is useful in many cases, as it is representative of spatial relations among various objects [10, 102]. Instance-level segmentation is essential for handling

numeric queries [183]. Unsupervised approaches [72] are particularly useful for handling a bulk amount of non-annotated data, which is very common in this field of work.

Medical Imaging

Another major application area for image segmentation is in the domain of health care. Many kinds of diagnostic procedures involve working with images corresponding to different types of imaging sources and various parts of the body. Some of the most common types of tasks are segmentation of organic elements such as vessels [47], tissues [73], and nerves [107]. Other kinds of problems include localization of abnormalities like tumors [118, 181] and aneurysms [39, 106]. Microscopic images [67] also need various kinds of segmentations, such as cell or nuclei detection, counting numbers of cells, and cell structure analysis for cancer detection. The primary challenges with this domain is the lack of a bulk amount of data for challenging diseases and variety in the quality of images due to the different types of imaging devices involved. Medical procedures not only involve human beings but also animals and plants.

Object Detection

With the success of deep learning algorithms, there has also been a surge in research areas related to automatic object detection, such as robotic maneuverability [90], autonomous driving [163], intelligent motion detection [160], and tracking systems [170]. Extremely remote regions, such as the deep sea [83, 158] or space [149], can efficiently be explored with the help of intelligent robots making autonomous decisions. In sectors like defense, unmanned aerial vehicles (UAVs) [125] are used to detect anomalies or threats in remote regions [95]. Segmentation algorithms have significant use in satellite images for various geo-statistical analysis [86]. In fields like image or video post-production, it is often essential to perform segmentation for various tasks, such as image matting [91], compositing [17], and rotoscoping [2].

Forensics

Biometric verification systems that use the iris [52, 100], fingerprints [76], finger veins [140], and dental records [75] involve segmentation of various informative regions for efficient analysis.

Surveillance

Surveillance systems [71, 77, 120] are associated with various issues, such as occlusion, lighting, or weather conditions. Moreover, surveillance systems can also involve analysis of images from hyper-spectral sources [4]. Surveillance systems can also be extended to applications such as object tracking [64], searching [3], anomaly detection [143], threat detection [111], and traffic control [174]. Image segmentation plays a vital role in segregating objects of interest from the clutter present in natural scenes.

6 DISCUSSION AND FUTURE SCOPE

Throughout this article, various methods have been discussed in an effort to highlight their key contributions, pros, and cons. With so many different options, it is still hard to choose the right approach to a problem. The most optimal way to choose a correct algorithm is to first analyze the variables that affect the choice.

One of the most important aspects that affect the performance of deep learning–based approaches is the availability of datasets and annotations. In that regard, a concise list of datasets belonging to various domains was provided in Table 1. When working on other small-scale datasets, it is a common practice to pre-train the network on a larger dataset of a similar domain. Sometimes an ample amount of samples are available, yet pixel-level segmentation labels may not be available considering that creating them is a taxing problem. Even in those cases, pre-training parts of networks on other related problems like classification or localization can also help in the process of learning a better set of weights.

A related decision that one must make in this regard is to choose among supervised, unsupervised, or weakly supervised algorithms. In the current scenario, a large number of supervised approaches exist; however, unsupervised and weakly supervised algorithms are still far from reaching a level of saturation. This is a legitimate concern in the field of image segmentation because data collection can be carried out through many automated processes, but annotating them perfectly requires manual labor. It is one of the most prominent areas where a researcher can contribute in terms of building end-to-end scalable systems that can model data distribution, decide on the optimal number of classes, and create accurate pixel-level segmentation maps in a completely unsupervised domain. The area of weakly supervised algorithms is also highly demanding. It is much easier to collect annotations corresponding to problems like classification or localization. Using those annotations to guide the image segmentation problem is also a promising domain.

The next important aspect of building deep learning models for image segmentation is the selection of the appropriate approaches. Pre-trained classifiers can be used for various fully convolutional approaches. Most of the time, some kind of multi-scale feature fusion can be carried out by combining information from different depths of the network. Pre-trained classifiers like VGGNet or ResNet or DenseNet are also often used for the encoder part of a encoder-decoder architecture. Here also, information can be passed from various layers of encoders to corresponding similar-size layers of the decoder to obtain multi-scale information. Another major benefit of encoder-decoder architectures is that if the down-sampling and up-sampling operations are de-

signed carefully, outputs can be generated that are of the same size as the input. It is a major benefit over simple convolutional approaches like FCN or DeepMask. This removes the dependency on the input size and hence makes the system more scalable. These two approaches are the most common in case of a semantic segmentation problem. However, if a finer level of instance-specific segments are required, it is often necessary to couple with other methods corresponding to object detection. Utilizing bounding box information is one way to address these problems, whereas

other approaches use attention-based models or recurrent models to provide output as sequence of segments for each instance of the object.

There can be two aspects to consider while measuring the performance of the system. One is speed, and the other is accuracy. CRF is one of the most commonly used post-processing modules for refining outputs from other networks. CRFs can be simulated as an RNN to create end-to-end trainable modules to provide very precise segmentation maps. Other refinement strategies include the use of over-segmentation algorithms like super-pixels, or using human interactions to guide segmentation algorithms. In terms of gain in speed, networks can be highly compressed using strategies like depthwise separable convolutions, kernel factorizations, and reducing the number of spatial convolutions. These tactics can reduce the number of parameters to a great extent without reducing the performance too much. Lately, generative adversarial networks have seen a tremendous rise in popularity. However, their use in the field of segmentation is still pretty thin, with only a handful of approaches addressing the avenue. Given the success that they have gained, they certainly have the potential to improve existing systems by a great margin.

The future of image segmentation largely depends on the quality and quantity of available data. Although there is an abundance of unstructured data on the Internet, the lack of accurate annotations is a legitimate concern. Pixel-level annotations especially can be incredibly difficult to obtain without manual intervention. The most ideal scenario would be to exploit the data distribution itself to analyze and extract meaningful segments that represent concepts rather than content. This is an incredibly challenging task especially if we are working with a huge amount of unstructured data. The key is to map a representation of the data distribution to the intent of the problem statement such that the derived segments are meaningful in some way and contribute to the overall purpose of the system.

7 CONCLUSION

Image segmentation has seen a new rush of deep learning–based algorithms. Starting with the evolution of deep learning–based algorithms, we have thoroughly explained the pros and cons of the various state-of-the-art algorithms associated with image segmentation based on deep learning. The simple explanations allow the reader to grasp the most basic concepts that contribute to the success of deep learning–based image segmentation algorithms. The unified representation scheme followed in the figures can highlight the similarities and differences of various algorithms. In the future, this theoretical survey work can be accompanied by empirical analysis of the discussed methods.

ACKNOWLEDGMENT

The authors would like to thank the reviewers for their valuable suggestions which helped improve the quality of the article.

REFERENCES

[1] Radhakrishna Achanta, Appu Shaji, Kevin Smith, Aurelien Lucchi, Pascal Fua, Sabine Süsstrunk, et al. 2012. SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 34, 11 (2012), 2274–2282.

[2] Aseem Agarwala, Aaron Hertzmann, David H. Salesin, and Steven M. Seitz. 2004. Keyframe-based tracking for rotoscoping and animation. 23, 584–591.

[3] Jamil Ahmad, Irfan Mehmood, and Sung Wook Baik. 2017. Efficient object-based surveillance image search using spatial pooling of convolutional features. *Journal of Visual Communication and Image Representation* 45 (2017), 62–76.

[4] Fahim Irfan Alam, Jun Zhou, Alan Wee-Chung Liew, and Xiuping Jia. 2016. CRF learning with CNN features for hyperspectral image segmentation. In *Proceedings of the 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS'16)*. IEEE, Los Alamitos, CA, 6890–6893.