IJCSE International Journal of Computer Sciences and Engineering Open Access
Survey Paper Vol.-6, Issue-12, Dec 2018 E-ISSN: 2347-2693

Survey On Semantic Segmentation

P.S.Gunde[1*], S.K.Shirgave[2]

[1]Dept of Computer Science and Engineering, D.K.T.E. Society's Textile & Engineering Institute, Ichalkaranji (An Autonomous Institute), India.

[2]Dept of Computer Science and Engineering, D.K.T.E. Society's Textile & Engineering Institute, Ichalkaranji (An Autonomous Institute), India.

[*]*Corresponding Author: priyankagunde0192@gmail.com*

Available online at: www.ijcseonline.org

Accepted: 20/Dec/2018, Published: 31/Dec/2018

Abstract— The Image segmentation is referred to as one of the most important processes of image processing. Image segmentation is the technique of dividing or partitioning an image into parts, called segments. It is mostly useful for applications like image compression or object recognition, because for these types of applications, it is inefficient to process the whole image. So, image segmentation is used to segment the parts from image for further processing. Semantic image segmentation is a vast area for computer vision and machine learning researchers. Many vision applications need accurate and efficient image segmentation and segment classification mechanisms for assessing the visual contents and perform the real-time decision making. There exist several image segmentation techniques, which partition the image into several parts based on certain image features like pixel intensity value, color, texture, etc. These all techniques are categorized based on the segmentation method used. The application area includes remote sensing, autonomous driving, indoor navigation, video surveillance and virtual or augmented reality systems etc. This survey paper provides a review of different traditional methods of image segmentation.

Keywords—Image segmentation, Conditional Random Field, Deep learning, semantic video segmentation

I.INTRODUCTION

Semantic segmentation is task of clustering parts of images together which belongs to the same category. Many vision applications need accurate and efficient image segmentation and segment classification mechanisms for assessing the visual contents and perform the real-time decision making. The application area includes detecting road signs [1], detecting tumors [2], detecting medical instruments in operations [3], colon crypts segmentation [4], land use and land cover classification [5].

At the same time, by using the general characteristics of a single object, non-semantic segmentation only clusters pixels together.

Object Detection is an important problem in computer vision and Robotics. It involves placing tight bounding boxes around areas of interest. Semantic Segmentation takes this one step further and deals with detecting the exact region that the object occupies. It involves classifying each segment in the image into one of the given classes.

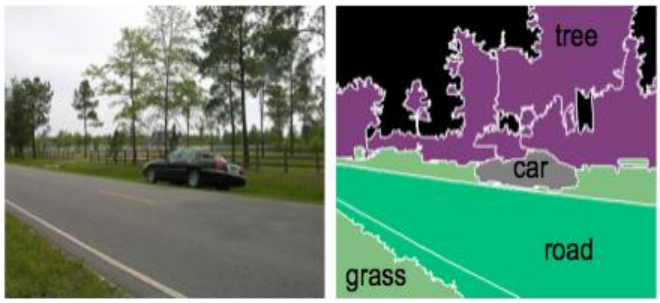

Figure 1: Example: left is input image and right is desired output image

Semantic segmentation is also known as image labeling or scene parsing, which relate to the problem of giving semantic labels to every pixel in the image. And it is very challenging task in computer vision and one of the most crucial steps towards scene understanding. This survey paper mainly focused on different techniques of segmentation. Figure 1 shows the example of segmentation of image in which left image shows the input to the system and right side gives the result of the semantic segmentation.

II. OUTLINE

Traditional methods for semantic segmentation used features like SIFT, HoG etc with classification algorithms like SVM, Random Decision Forest. Other class of algorithms [6] used Conditional Random Fields defined on individual image pixels or a group of pixels(superpixels) More recently, deep learning methods have become more prevalent and most of the state-of-the-art algorithms use these methods. The first significant attempt to solve this problem using deep learning methods was made by Long et.al proposing a Fully Convolutional Network [7]. Parallely, Hariharan et.al proposed hypercolumns [8], based on the idea of collecting information from all levels of a net-work to give equal importance to semantic and localization information. Chen et.al combined the traditional CRF approach with FCNs [9] obtaining further improvements. This idea was further refined by Zheng et.al by formulating CRFs as RNNs [10].

The CRF based approach proposed in [6], is extended to videos by incorporating temporal dependence and learning new spatial features, by Kundu et.al [11].

III. CONDITIONAL RANDOM FIELD(CRF)

The key idea of CRF inference for semantic labelling is to formulate the label assignment such as the label agreement between similar pixels or image regions based on position or color intensities.

The main contribution of the paper was the proposal of an efficient algorithm for approximate inference, in a CRF model defined on superpixels. Before this, CRFs were mainly defined on individual pixels instead of superpixels. Doing so made the task computationally more cheaper. But at the same time, the performance of the algorithm to a large extent, depended on the boundaries of the proposed image regions with the object boundaries.

IV. DEEP LEARNING METHODS

4.1 Fully Convolutional network (FCN)

The approach employ a fully convolutional neural network (FCN), which is trained end-to-end for pixel level annotation. A series of convolution, pooling and deconvolution layers are used.

Conversion to Fully Convolutional Network: Fully connected layers are discarded to maintain some amount of local information, which is otherwise completely lost in favor of learning global features. A deep network is modified to fully convolutional network as shown in figure 2. One advantage of using a fully convolutional network is that an image of any dimension can be fed into the network.

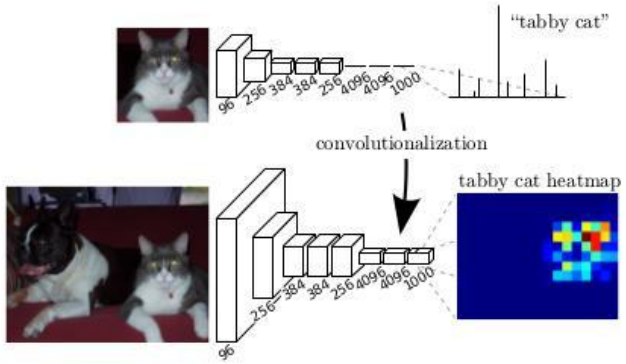

Figure 2: Transforming fully connected layers into convolution layers enables the network to take inputs of any size [7]

Skip Architecture: The convolutional network too suffers from the loss of locality, in part due to down sampling and pooling layers. This problem is overcome by the introduction of a skip architecture, where features from layers less deep down in the network are used. These features are used in addition to the features learnt deeper down, that are more semantically rich and hence help in classifying a region.

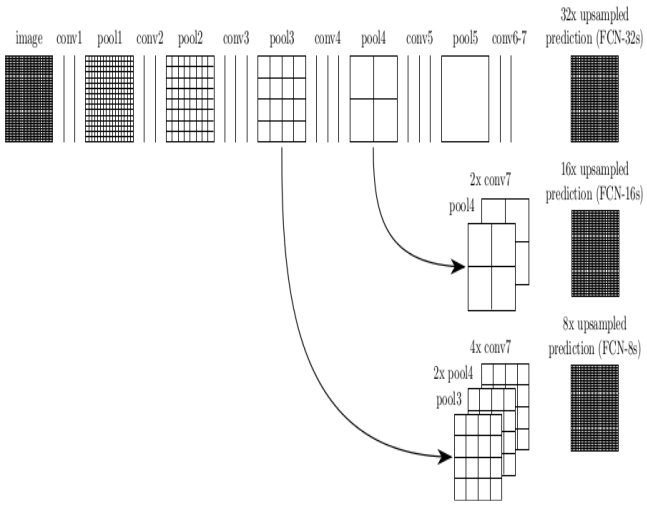

Figure 3: Using features from both shallower and deeper layers to aid in localization and classification [7]

4.2 Hypercolumns

Recognition Algorithms are based on features extracted from the higher layers of a Deep Convolutional Network. These features are very good for the semantic representation of a given image. But, in addition to semantic information, localization information also holds importance for tasks like semantic segmentation. The top level features, however are too course spatially to capture localization, which can be extracted more effectively from lower layers.

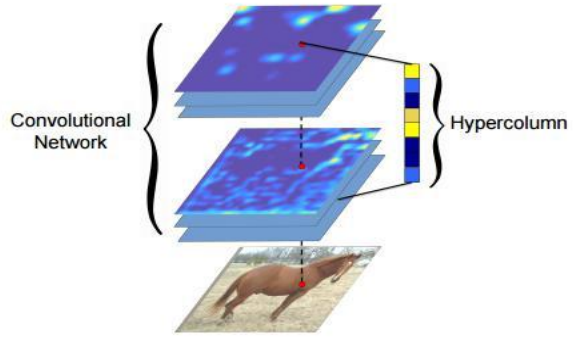

Figure 4: Hypercolumns [8]

Based on these observations, a novel way of feature representation for each pixel, called hypercolumn (Figure 4), is proposed for pixel level classification. The hypercolumn at a given location is defined as a vector containing the outputs of all units above that location at all layers of the CNN.

4.3 Fully Convolutional Networks with Fully Connected Conditional Random Fields (CRF)

The paper proposes some changes to the network in [7] and then builds on top of it, by introducing a CRF based segmentation algorithm.

There is a natural trade off between classification accuracy and localization accuracy with convolutional networks. Fully connected CRF is introduced to overcome this challenge of localization. At the same time, it also smoothes the region boundaries, without compromising on the use of the robust features computed by the DCNN. The algorithm is capable of segmenting 8 images per second on a modern GPU.

Changes to FCN: The last two max-pooling layers are skipped to maintain a stride of 8 at the end, as opposed to a stride of 32. This enables more dense prediction at the target. At the same time the filters in the last few layers are applied with an input stride of 2 or 4 (this is referred to as the 'hole' algorithm) (figure 5). This is done to retain as much local information as possible, at the same time allowing them to capture as much semantic information as possible. The loss function is the sum of the cross entropy loss for each pixel (after the original image has been appropriately subsample to the size of the FCN's output)

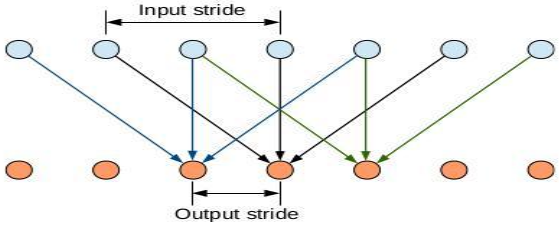

Figure 5: Illustration of the hole algorithm in 1-D, when kernel size = 3, input stride = 2, and output stride = 1[9]

Fully Connected CRF: The algorithm proposed in [6] is used as it is, using the output of the DCNN as the probabilities in the unary potentials. The DCNN is trained separately and the unary potentials provided by it are considered fixed at the time of training the CRF.

4.4 Conditional Random Fields as Recurrent Neural Networks (CRF-RNN)

The main contribution of the paper is the formulation of the CRF as a Recurrent Neural Network (CRF as RNN). This network is concatenated at the network designed in [7], thus obtaining a deep network that has the desirable properties of both CNNs and CRFs. The network, unlike [9] can be trained end-to-end using the usual back-propagation algorithm.

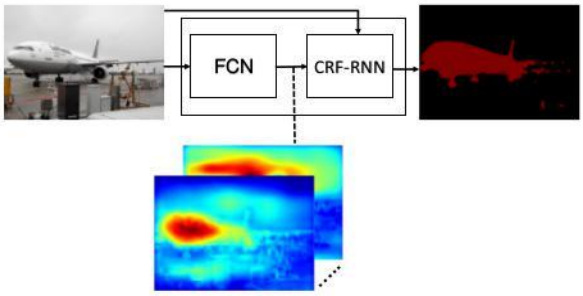

Figure 6: Schematic visualization of the full network which consists of a CNN and the CNN-CRF network [10]

V. SEMANTIC VIDEO SEGMENTATION [11]

CRF Model: A dense CRF based approach is adopted. A series of CRFs are defined over all pixels in segments of space-time volume called blocks (figure 7). As in image segmentation using CRFs, all pixels over multiple blocks are labelled jointly and labelling coherence is explicitly enforced. This alleviates the noise and inconsistency that can arise when pixels are classified independently. This is done by defining pairwise potentials, which coerce similar (in space, time and colour intensity) pixels to have similar labels. The CRF also requires unary potentials that represent the probability of individual pixels taking on particular labels. These unary potentials can be provided by any standard image semantic segmentation algorithm.

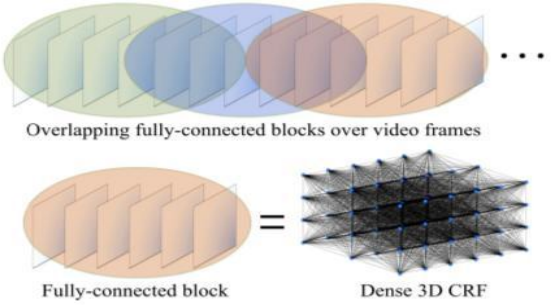

Figure 7: The temporal structure of the model. The video is covered by overlapping blocks. A dense CRF is defined over each block [11]

The significant attempt to solve this problem using deep learning method with a Fully Convolutional Network. Hypercolumn is based on the idea of collecting information from all levels of a net-work to give equal importance to semantic and localization information. Traditional CRF approach with FCNs obtaining further improvements and in further by formulating CRFs as RNNs.

VI. CONCLUSION

The popular CRF framework is used for segmentation in the CRF and deep learning methods. The CRF is based on both pixel level and image level. It is mainly used for labelling the superpixels with same features and regions.

REFERENCES

[1]S.Maldonado-Bascon, S.Lafuente-Arroyo, P. Gil-Jimenez, H. Gomez-Moreno, and F. Lopez-Ferreras, "Road-sign detection and recognition based on support vector machines," Intelligent Transportation Systems, IEEE Transactions on,vol. 8, no. 2, pp. 264–278, Jun. 2007.

[2]N.Moon, E. Bullitt, K. Van Leemput, and G. Gerig, "Automatic brain and tumor segmentation," in Medical Image Computing and Computer-Assisted Intervention MICCAI 2002. Springer, 2002, pp.372–379.

[3]G.-Q. Wei, K. Arbter, and G. Hirzinger, "Automatic tracking of laparoscopic instruments by color coding," in CVRMed-MRCAS'97, ser. 1997, vol. 1205, pp. 357–366.

[4]"Memory based active contour algorithm using pixel-level classified images for colon crypt segmentation," Computerized Medical Imaging and Graphics, Nov. 2014.

[5] C. Huang, L. Davis, and J. Townshend, "An assessment of support vector machines for land cover classification," International Journal of remote sensing, vol. 23, no. 4, pp. 725–749, 2002.

[6] Krahenbuhl, P. and Koltun, V. E_cient inference in fully connected crfs with gaussian edge potentials.In NIPS, 2011.

[7]Jonathan Long, Evan Shelhamer, Trevor Darrell Fully Convolutional Networks for Semantic Segmentation CVPR 2015 arXiv:1411.4038

[8]B. Hariharan, P. Arbelaez, R. Girshick, and J. Malik. Hypercolumns for object segmentation and _new grained localization. In Computer Vision and Pattern Recognition, 2015

[9]L.C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille. Semantic image segmentation with deep convolutional nets and fully connected crfs. In ICLR, 2015

[10] S. Zheng, S. Jayasumana, B. Romera-Paredes, V. Vineet, Z. Su, D. Du, C. Huang, and P. Torr. "Conditional random fields as recurrent neural networks". In ICCV, 2015

[11]Kundu, Abhijit, Vineet, Vibhav, and Koltun, Vladlen. "Feature space optimization for semantic video segmentation". In CVPR, 2016

[12]M. Everingham, L. Van Gool, C.K.I. Williams, J. Winn, A. Zisserman, "The PASCAL Visual Object Classes Challenge 2007 (VOC2007) Results", 2010

[13]G. J. Brostow, J. Fauqueur, and R. Cipolla. "Semantic object classes in video: A high-de nition ground truth database". Pattern Recognition Letters, 30(2), 2009

[14]M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele. "The Cityscapes dataset for semantic urban scene understanding". In CVPR, 2016

[15]A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision meets robotics The kitti dataset," 2013

[16]A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? the kitti vision benchmark suite," in CVPR, 2012

Authors Profile

Priyanka S. Gunde pursed Bachelor of Engineering from SIT College of Engineering, Yadrav, India University, in year 2013, She is currently pursuing Master of Technology from DKTE's TEI, (An Autonomous Institute), Ichalkaranji, India. Her research work focuses on recommendation systems.

Suresh Shirgave is currently Associate Professor of Computer Science and Engineering at DKTE Society's Textile and Engineering Institute, Ichalkaranji, Maharashatra, India. He has published many research papers in national andinternational conferences and Journals. His research interests include data mining, Web usage mining,recommendation systems, social networks and Internet security.

