A Survey on Semantic Image Segmentation

[1]Mohsina Ansar, [2]Shiny.B

Image Processing
College of Engineering Chengannur Kerala, India

***Abstract*: The field of digital image processing has experienced continuous and notable expansion in recent years. Nowadays, the advancement and wide availability of image processing hardwares has further enhanced the usefulness of image processing. Image Segmentation is an important technique in this area and is essential for image analysis tasks. Here, we are considering the concept of semantic image segmentation which describes the process of associating each pixel of an image with a class label. Also, the aim of semantic segmentation is that it clusters parts of images which belong to same object class. This paper deals with detailed survey on different semantic image segmentation techniques.**

***Index Terms*: Semantic Segmentation, Image Segmentation, Hierarchical Models, Deep Learning, Convolutional Neural Network.**

I. INTRODUCTION

Image segmentation is the method of dividing an image into different sets of pixels. The goal of segmentation is to simplify or change the representation of an image into a form that is more meaningful and easier to analyse. Semantic segmentation is the process of assigning each pixel of an image with a class label. It is a high-level task that covers the way towards complete scene understanding. The significance of scene understanding in computer vision problem is mentioned by the fact that an increasing number of applications nourish from inferring knowledge from imagery. Some of those application includes self-driving vehicles, human computer interaction, virtual reality etc. With the popularity of deep learning, many semantic segmentation problems are being solved using deep architectures, mostly Convolutional Neural Nets, which exceeds other approaches in terms of accuracy and efficiency.

One of the important challenges in the history of computer vision has been semantic segmentation which is the ability to segment an unknown image into different parts and objects. In semantic image segmentation the main goal is to label each pixel of an image in to corresponding class of what is being represented. This is also referred to as dense prediction since we predict every pixel in the image. Also, here exist a different class of models, known as instance segmentation model. That distinguish between separate objects of the same class. This paper provides different techniques used in semantic image segmentation

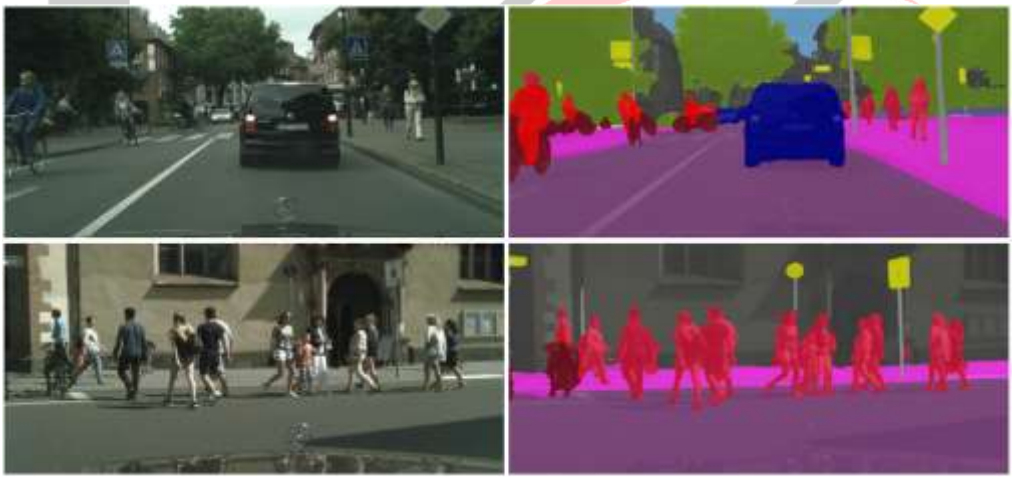

Fig. 1. Examples for Semantic Segmentation

II. LITERATURE SURVEY

This section includes various methods used for semantic image segmentation.

A. Color Classification

Color Classification is mainly achieved within the HSV colour space, and is commonly used for image classification. Unlike RGB, HSV is mainly designed to make human interpretation easier, separating colour data into channels reflecting how human vision functions[3]. Here, the image intensity or value is separated from the colour information. This classification mainly works in a

sequential process. A colour based Bayesian Network classifier is trained and used to semantically classify each segmented cluster. Here, the 'first' channel to process is Value, which represents distance from black within the cylindrical co-ordinates. Also, the Saturation colour channel denotes the intensity of colour and is applicable for the classification of every region which is not identified as black. Here a novel approach called NRL is used. A probabilistic BN was used to fuse both HSV and NRL to give an accurate estimate of each cluster class.

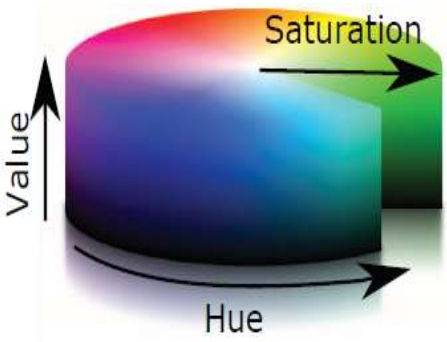

Fig. 2. HSV cylindrical colour space

B. FCN-CRF Model

CRF is a discriminative undirected probabilistic graph model. CRF models have strong reasoning ability, and thus able to train and inference with complex, overlapping and non-independent features, which make full use of the context information. The FCN is adopted to automatically learn effective features directly from original image data. It then combine a deep, coarse layer with shallow, fine layer to get more accurate and detailed information[4]. Thus it avoid the need of constructing handcrafted feature. Then we combine FCN with CRF to incorporate image feature learning and then dense predictions for per-pixel in a unified framework. Also, the training of FCN-CRF is in end-to-end fashion by computing transferring the sensitivity of neurons, which makes the errors transfer from top to bottom and makes the model data consistency

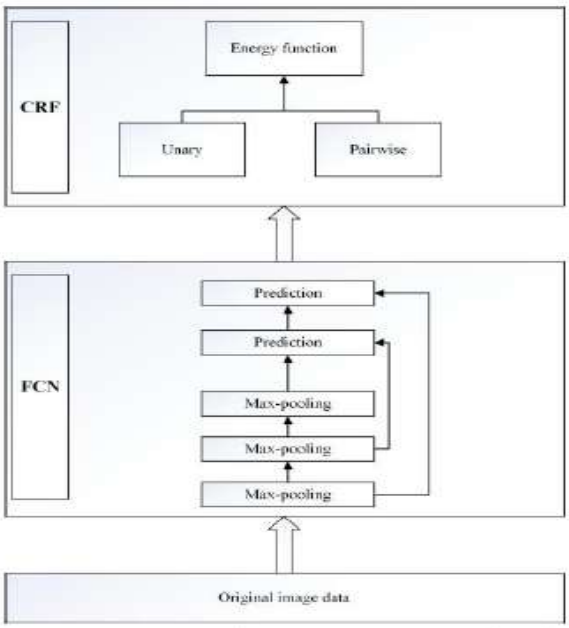

Fig. 3. The process of FCN-CRF model and main points

C. Ontology Based Semantic Image Segmentation Method

OBSIS is an important method which efficiently employs different types of information at the proper levels. It bridges low- and high-level features by incorporating semantic knowledge in a gradual process from the beginning. The low-level visual space is transformed into an intermediate semantic space of reduced dimensionality using the Dirichlet process mixture models and multiple CRFs. The Dirichlet process mixture model can be seen as a limit of a parametric mixture model. The hierarchical Gaussian mixture

model formulation is used by approaching the limit of the number of mixture components to infinity. Let the Dirichlet process model can hence be defined as:

$$p(\vec{Y}|\vec{\pi},\vec{\alpha},\vec{\beta}) = \sum_{j=1}^{\infty} \pi_j Dir(\vec{Y}|\vec{\alpha}_j,\vec{\beta}_j),$$

The Dirichlet mixture models are used to transform the low level visual space into a higher level space with the CRF. The visual features in this higher level space include intermediate labels which are used for the region labeling. From the, semantic concepts, higher level features in the intermediate space and their relationships a semantic ontology is constructed. And the final inference is performed by this model. After that cluster the visual space using the Dirichlet process, and learn the cluster representations using CRFs. Thus, the problem of image segmentation is thereby reduced to that of a classification task where the CRFs individually classify image regions into suitable labels for each visual feature. Also, the visual features are learned within the respective contexts by the CRFs where training and predicting are facilitated using the higher level information with less dimensionality.

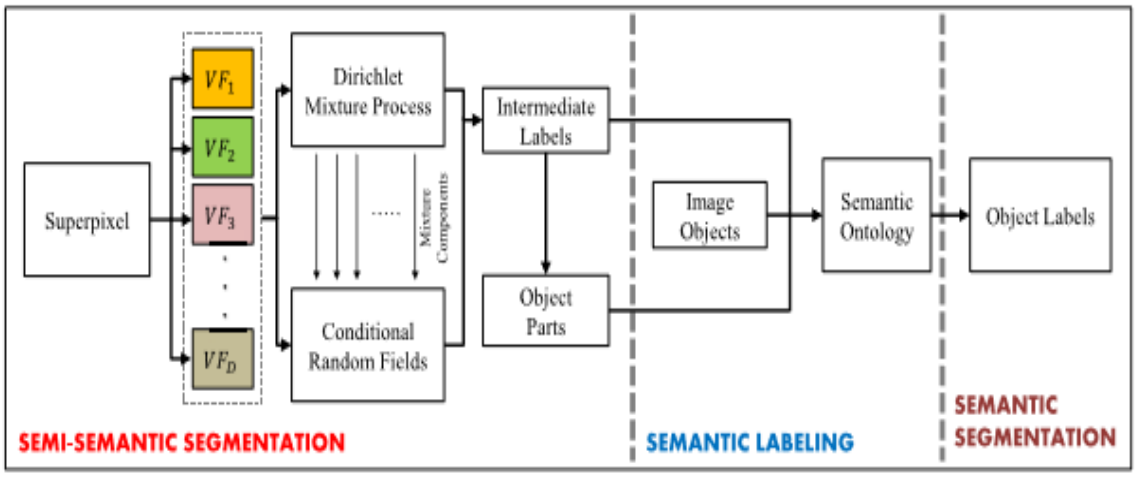

Fig. 4. OBSIS

Using the ontologies for the final inference enables reduction of the semantic labels over the inferred information and knowledge in the previous steps. This captures the interactions between semantic concepts from both the semantic level and visual level with an indispensable semantic context modelling[2]. Another important point is the fuzzy descriptions of the relationships and assignments. Appropriate weighing of different features by ontologies is another added advantage of this method, which causes differences in discriminant features for each region and also the human-like image labeling.

D. Contextual Hierarchical Model

CHM learns contextual information in a hierarchical framework for semantic segmentation. Here, the classifiers are trained at multiple resolutions in a hierarchy. These classiers are trained based on the downsampled input images and outputs of previous levels. CHM method mostly targets a posterior probability at multiple resolutions and maximizes it greedily through the hierarchy. Here at first, a multi-resolution representation of the input image is obtained by applying downsampling sequentially. Then, a series of classiers are trained at different resolutions starting from the nest resolution to the coarsest resolution. At each resolution, the classier is trained based on the outputs of the previous classifiers in the hierarchy and the input image at that resolution[1]. Finally, the outputs of the classifiers are used to train a new classifier at original resolution. This classifier utilize the rich contextual information from multiple resolutions. The entire training process targets a joint posterior probability at multiple resolutions. The whole process is illustrated in fig.5[1]

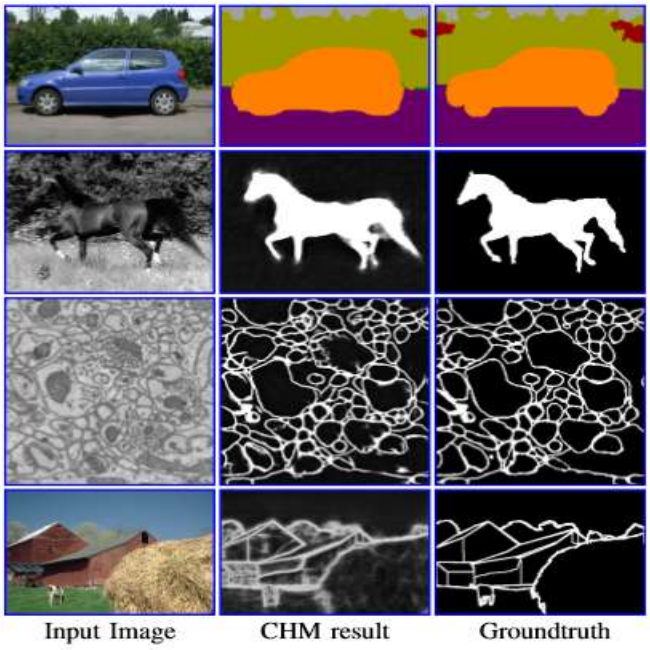

Fig. 5. Results of CHM on different tasks

E. Deep Convolutional Neural Network

Deep CNN method for semantic segmentation separates an input image into multiple regions corresponding to predetermined object classes[5]. This is an end-to-end structure that provides pixel-wise segmentation. Also, here a new class extension scheme is developed based on a pre-trained model and a partly labeled training data. Here, an encoder-decoder structure is used which aims to improve in convergence speed and segmentation accuracy by adding shortcuts between network layers. This architecture is illustrated in fig.6.

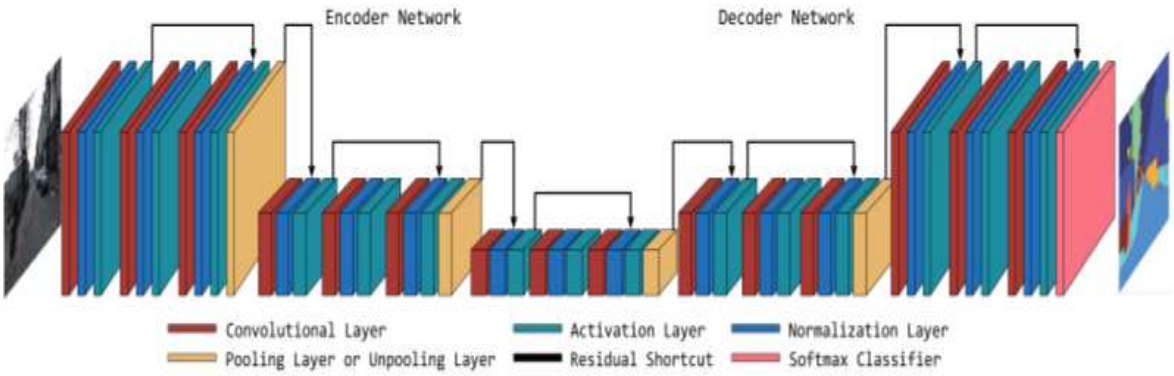

Fig. 6. Results of CHM on different tasks

F. DeepLab

Deep CNN method for semantic segmentation separates an input image into multiple regions corresponding to predened object classes. This is an end-to-end structure that provides pixel-wise segmentation. Also, here a new class extension scheme is developed based on a pre-trained model and a partly labeled training data. Here, an encoder-decoder structure is used which aims to improve in convergence speed and segmentation accuracy by adding shortcuts between network layers. By applying the 'atrous convolution' for dense feature extraction[6], DeepLab system re-purposes networks trained on image classification to the task of semantic segmentation with upsampled filters. Further extend it to atrous spatial pyramid pooling, that encodes objects as well as image context at multiple scales. To produce semantically detailed segmentation and accurate predictions maps along object boundaries, we combine ideas from deep convolutional neural networks and fully-connected conditional random fields. Compared to regular convolution with

larger filters, atrous convolution allows to effectively enlarge the field of view of filters without increasing the number of amount of computation or the parameters. The method is illustrated in fig.7.

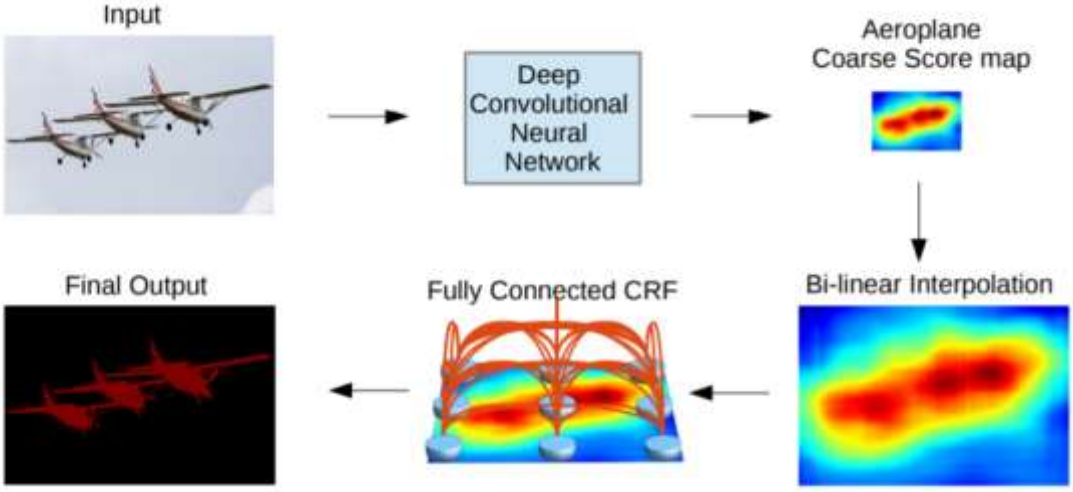

Fig. 7. Model Illustration

III. CONCLUSION

In computer vision, image segmentation plays a very important role and is the process of segregating a digital image into multiple segments. Mostly, in image segmentation we assign a label to each pixel of an image. Thus pixels with same labels share same characteristics. Actually, semantic segmentation is a classic computer vision problem that takes an input and convert them into a mask with different regions of interest. However, in this work we consider semantic image segmentation which paves a way towards complete scene understanding. So semantic image segmentation is more efficient. In this paper, we have briefed various methods for semantic image segmentation. Different methods discussed includes colour classication, ontology based method, FCN-CRF model, CHM model, DeepLab,Deep CNN. Also, semantic segmentation provides several applications in different fields which includes geosensing , autonomous driving, facial segmentation etc.

REFERENCES

[1] Tolga Tasdizen Mojtaba Seyedhosseini. ”semantic image segmentation with contextual hierarchicalmodels”. in IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE,26(9):4446 – 4456, June 2016.

[2] Mohsen Zand, Shyamala Doraisamy, Alan Abdul Halin,and Mas Rina Mustaffa,”Ontology-Based Semantic Image Segmentation Using Mixture Models and Multiple CRFs, ”, in IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 25, NO. 7, JULY 2016.

[3] Matthew Coombes, William Eaton, Wen-Hua Chen , ”Colour Based Semantic Image Segmentation and Classication for Unmanned Ground Operations”, in International Conference on Unmanned Aircraft Systems (ICUAS) June 7-10, 2016.

[4] Hao Zhou*, Jun Zhang, Jun Lei, Shuohao Li, Dan Tu, ”Image Semantic Segmentation Based on FCN-CRF Model”,in 2016 International Conference on Image, Vision and Computing.

[5] ChunlaiWang,LukasMauch,ZeGuoandBinYang,”OnSemanticImage Segmentation Using Deep Convolutional Neural Network with Shortcuts and Easy Class Extension ,”, in Institute of Signal Processing and System Theory, University of Stuttgart, Germany

[6] Liang-Chieh Chen , George Papandreou,Iasonas Kokkinos,Kevin Murphy, and Alan L. Yuille,”DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs”,in I IEEE transactions on pattern analysis and machine intelligence,2016.

[7] C. Farabet, C. Couprie, L. Najman, and Y. LeCun,”Learning hierarchical features for scene labeling”,TdPAMI,2013.

[8] Z. Ren and G. Shakhnarovich, “Image segmentation by cascaded region agglomeration,”, CVPR, 2013.

[9] Z. Tu and X. Bai, ”Auto-context and its application to high-level vision tasks and 3d brain image segmentation,”, TPAMI, 2010.

[10] G. Heitz, S. Gould, A. Saxena, and D. Koller,"Cascaded classication models: Combining models for holistic scene understanding,",NIPS, 2008.

[11] Z. Ren and G. Shakhnarovich, "Image segmentation by cascaded region agglomeration,", CVPR, 2013.

[12] M. Seyedhosseini and T. Tasdizen "Multi-class multi-scale series contextual model for image segmentation,",Image Processing, IEEE Transactions on 4486–4496, 2013.

[13] Z. D. Kuettel and V. Ferrari, "Figure-ground segmentation by transferring window masks,", in CVPR, 2012.

[14] E. Borenstein, E. Sharon, and S. Ullman "Combining top-down and bottom-up segmentation," Proc. of CVPRW, pp. 46 –46, 2004.

[15] S. Gould, R. Fulton, and D. Koller, "Decomposing a scene into geometric and semantically consistent regions," in ICCV, 2009.

[16] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in ICCV, 2001.

[17] N. Silberman and R. Fergus, " Indoor scene segmentation using a structured light sensor ," in ICCV Workshop, 2011.

[18] O. Sporns, G. Tononi, and R. Ktter, "The human connectome: a structural description of the human brain," in PLoS Computational Biology, 2005.

[19] M. Seyedhosseini, M. Sajjadi, and T. Tasdizen, "Image segmentation with cascaded hierarchical models and logistic disjunctive normal networks," in ICCV, 2013.

[20] D. Larlus and F. Jurie, "Combining appearance models and markov random elds for category level object segmentation," in CVPR, 2008.

[21] M. P. Kumar and D. Koller, "Efciently selecting regions for scene understanding," in CVPR, 2010.

[22] X. He, R. Zemel, and M. Carreira-Perpinan, ""Multiscale conditional random elds for image labeling," in CVPR, 2004.

[23] J. Long, E. Shelhamer, and T. Darrel, "Fully convolutional networks for semantic segmentation," in CVPR, 2015.

[24] N. Silberman and R. Fergus, "Indoor scene segmentation using a structured light sensor" in ICCV Workshop, 2011.