Exploiting Contextual Information of Color Images through Feature Extraction Techniques for Semantic Segmentation

Cheruku Sandesh Kumar, Vinod Kumar Sharma, Rekha Chaturvedi

Abstract*: In our findings we try to explore spatial context to obtained good results of semantic segmentation. Spatial context has patch-to-patch and patch-to-background. Patch-to-patch context has semantic relationships on visual patterns of two stuffs of a image. Patch-to-background context had semantic relations in image patch and whole background region. In our research we have explored contextual relations based on CRF. CNNs pair-wise potential captures semantic correlation on nearby patches. Researchers in the past used CNN-Sparse CRF.In our model we used CNN-Dense CRF technique to refine our samples to sharpen the object boundaries. CNN-Dense CRF use pair-wise potentials for local smoothness of images. Pair-Wise potentials are log-linear functions for semantic compatibility in image regions. CRF Pairwise is to develop coarse-level prediction. CRF and Potts-model-based pair-wise potential are jointed to obtained good results for semantic segmentation.

Index Terms*: Deep learning, FCNN, ANN, Adaboost, CRF, SS-Semantic Segmentation.

I. INTRODUCTION

Image SS has a category label for image pixel that plays important role in the complete scene understanding of an image. The related approaches like CNNs have pixel-level labeling [151 1] [2] [3]. There are many CNNs methods FCNNs [2] [3] is widely used. Context information or data has the main cues for scene understanding areas. Considering on a highway, a bottle on a stool, context encodes incompatibility relations would be the example a boat on the highway. Context information is mainly in finding sign for isolated object that has visual uncertainty. The spatial context is a broad area of research has given in [4].

Fig. 1 shows prediction method. The patch-background context is traversed in this regard. CNN based techniques of multi-scale image network gives good output when compared to recent semantic segmentation methods [1] [5]. In this model the use of multi-scale networks to encode background data and then slide pyramid pooling on feature maps is applied to encapsulate intelligence from background regions of various sizes. Generally Pairwise potentials have rich computational inference in CRF training. The piece-wise learning of CRF ignores continuous inference on back propagation learning of deep model [6].

The major points involved in the design of FCNN and CRF is shown in Fig. 1 and the inference and learning process of FCNN is shown the below Fig. 2. Semantic Segmentation is a major area in digital image processing for complete scene understanding. In deep learning pixel level labeling tasks are done. We combined FCNNs and CRF to achieve semantic segmentation. The FCNNs model reads image features from the original color image size of 960×720 of portable network graphics type and is resized to 224×224 to develop local prognostics and global structure consistency by combining good and coarse layers. The CRFs are probabilistic graphs for exploiting contextual information or data. This model does end to end training with back propagation algorithm and maximum likelihood estimation. The combining of FCNNs and CRF is related to the sensitivity of neurons.

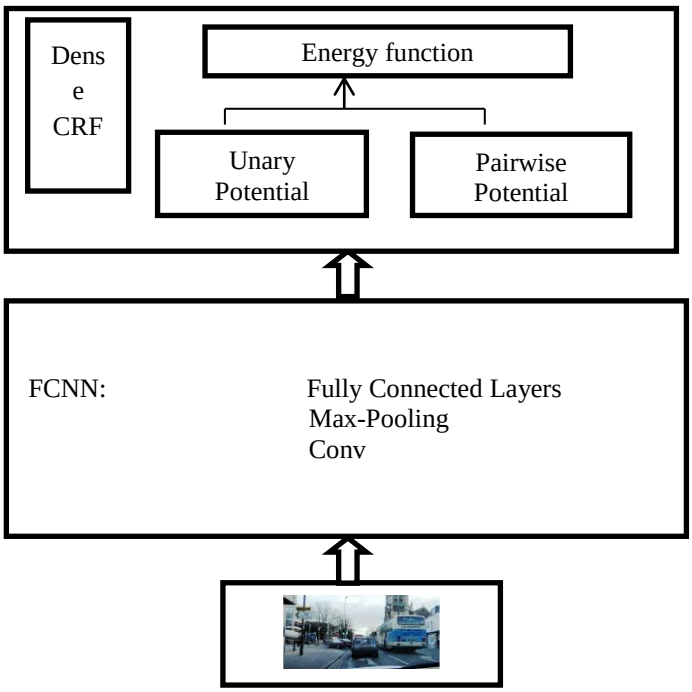


Fig. 1. Proposed deep model.

Revised Manuscript Received on March 10, 2019.

Dr. Cheruku Sandesh Kumar, ECE, Amity University Rajasthan, Jaipur.
Mr Vinod Kumar Sharma, ECE, Amity University Rajasthan, Jaipur.
Rekha Chaturvedi, CSE, Amity University Rajasthan, Jaipur.

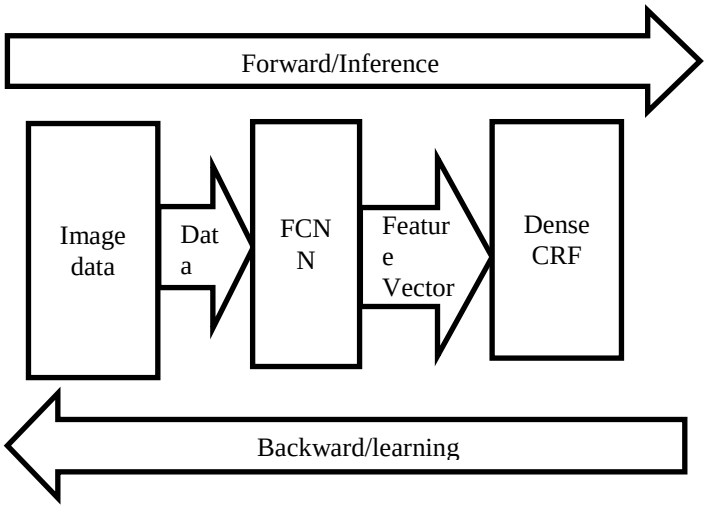


Fig. 2. Training and Inference.

II. PRE-PROCESSING

Pre-processing is done after the database is created. Pre-processing includes various sub processing units. As soon as the digital image is processed it looks as a raw material and to make use of this material pre-processing is needed. Feature extraction is done after pre-processing and without pre-processing features extracted is negligible. Pre-processing makes the digital images more adaptable, appropriate, intensified and significant for features extraction. In this research the pre-processing steps are namely, De-noising, Smoothening, Sharpening, Histogram Equalization etc. [7].

III. DEEP LEARNING

Deep learning is a process in which structured learning is used. This learning is an application of ANNs. In ANNs multiple hidden layers are used. Deep learning is considered as an important part of machine learning techniques, where the learning data representation is used. In these kind of learning techniques, learning may be supervised, unsupervised, reinforcement and semi-supervised. In our model we are involved with supervised learning.

Interpretation of information processing is based on representation, for example communication patterns depend on the representation of nervous system. Deep learning models are of various types such as, DNNs, deep belief and RNNs. They can be applied for various applications such as computer vision, augmented reality, speech recognition etc. The performances of these architectures are very good as compared to humans [8].

IV. CONTEXT DEEP CRF

The design of deep CRF model is given below. Let us consider $i \in I$ one original image and $j \in J$ labeling mask that labels of every node in CRF graph. As we know, energy function is given as $E(j,i;\theta)$ and it models the compatibility of the input-output pairs with a small output value denoting high confidence in prediction j.The networks are denoted by θ that the classifiers need to study.

As per probability theory, the conditional likelihood for the given image is

$$P(j/i) = \frac{1}{S(i)} \exp\left[-E(j,i\right] \quad (1)$$

where 'S' is partition function, and is given as

$$S(i) = \sum_j \exp\left[-E(j,i\right].$$

The energy function which is a set of unary and pairwise potentials is given by

$$E(j,i) = \sum_{K \in k} \sum_{c \in M_K} K(j_c, i_c) + \sum_{T \in t} \sum_{(c,d) \in F_T} T(j_c, j_d, i_{cd}) \quad . \quad (2)$$

Here K is unary potential function. Simplify expression, assume multiple type of unary potential k - the set of all unary potentials. M_K denotes the set of nodes of potential K. T is a pairwise potential function with 't' is the set of all types of pair-wise potentials. F_T is the set of edges for the potential T. i_c and i_{cd} indicates image regions associated with node and edge.

A. Unary Potentials

The unary potential functions for feature maps and FCN by stack of FeatMap-Net is known as unary network. To get the final output of Unary Potential Functions (UPF) it is written as

$$K(j_c, i_c; \theta_K) = -w_{c,j_c}(i; \theta_K) \quad (3)$$

Here w_{c,j_c} is the output value of unary net for c^{th} node and j_c^{th} class.

We assign feature vector of one node. Input-output of unary network is node feature vector from feature map. From feature map formulate one CRF node feature vector. Dimension of unary net output vector of one node is H classes [9].

B. Pairwise Potentials

Pairwise potential function in comparison with unary potentials by stack FeatMap-Net, develops feature maps and FCN, known as pairwise network to give last output of pairwise potential function. Pairwise potential function is given

$$T(j_c, j_d, i_{cd}; \theta_T) = -w_{c,d,j_c,j_d}(i; \theta_T) \quad (4)$$

Here w_{c,d,j_c,j_d} is the output efficacy of pairwise net.

Confidence efficacy of node pair (c,d) is labeled with class value (j_c, j_d) for the original image i. θ_T

correspondes to a set of CNN parameters for potential T. Feature vectors of two nodes are added to get CRF edge feature vector. The pairwise network has H_2 output classes to sink the number of label combinations of a pair of nodes [10]. The pairwise potential nodes constitutes semantic similarity relations between two nodes with output for each feasible efficacy of labeled pairs is (j_c, j_d) obtained by FCNNs [11] [12]. After coarse level prediction, there is still work to be done of refining end prognostic. In this thesis we applied dense CRF technique at the stage of prediction refinement.

V. FEATURE CLASSIFICATION USING DENSE CRF

The CRF approach is to either maximize the likelihood or minimize the negative log likelihood. This is represented for each image is as:

$$-\log P(j/i;\theta) = E(j,i;\theta) + \log S(i;\theta) \quad (5)$$

The CNN parameters θ optimized for CRF learning is given by

$$\min_{\theta} \frac{\lambda}{2}\|\theta\|_2^2 - \sum_{b=1}^{G} \log P(j^{(b)} | i^{(b)};\theta) \quad (6)$$

The $i^{(b)}, j^{(b)}$ represent the b -th training image and the segmentation mask, G is number of training images, λ is weight decay parameter.

Substituting equation (5) into equation (6) we get

$$\min_{\theta} \frac{\lambda}{2}\|\theta\|_2^2 + \sum_{b=1}^{G}\left[E(j^{(b)},i^{(b)};\theta) + \log S(i^{(b)};\theta\right] \quad (7)$$

The classifiers used Stochastic Gradient Decent (SGD) techniques to optimize the above problem for learning θ .

Energy function $\mathrm{E(j,i;}\theta)$ constructed from CNNs and its gradient $\nabla_\theta E(j,i;\theta)$ can be computed by applying the chain rule in conventional CNNs. Partition function S gives the degree of difficulty for optimization. Its corresponding gradient is given by

$$\nabla_\theta \log S(i;\theta) = \nabla_\theta \log \sum exp\left[-E(j,i)\right] \quad (8)$$

$$= \sum_j \frac{exp\left[-E(j,i;\theta\right]}{\sum_{j'} exp\left[-E(j',i;\theta\right]} \nabla_\theta\left[-E(j,i;\theta)\right] \quad (9)$$

$$= -E_j \sim P_{(\frac{j}{i};\theta)} \nabla_\theta E(j,i;\theta) \quad (10)$$

Similarly the range of output area J that has exponential amount of nodes, prevent straight computation of S and its gradient. CRF graph is loop graph has extensive quantity of nodes. In this case for 224×224 image, the number of nodes is around 600, and the number of connections is around 100 for each node. Therefore for this image, we need to process 600×100 pairwise relations for generating edge features. Loopy graph is accompanied by massive number of nodes and edges. A substantial unit of SGD iterations are required for learning CNNs. There are thousands of SGD iterations, perform inference at every SGD utterance is expensive computation.

VI. PROGNOSTIC CONSUMMATION PHASE/POST PROCESSING

Generating score map for coarse prediction is done by marginal distribution can be obtained by mean field inference. The size of two feature maps are upsampled to the size of third one by bilinear interpolation for coarse prediction. We use dense CRF technique for post processing [19] to scarp object boundary for generating last prediction. In last stage, phase boundary refinement is done which leverages low level pixel intensity data and low resolution prediction [20] [21] [22]. Example of Refinement methods are coarse training of deconvolution networks, multiple coarse for fine learning and exploring middle layer features for resolution prediction [23][24][25].

Refinement approach is used for performance improvement. Feature maps are explored from middle layers to refine the coarse prediction practically.

VII. PERFORMANCE EVALUATION

The semantic segmentation performance is calculated through IoU. Assumption of $P_{x,y}$ is the confusion matrix that is amount of pixel of x^{th} class a ground truth, y^{th} class a prognostic. v_x entire quantity pixels of x^{th} class of ground truth. H aggregate of classes. IoU score that is $\frac{1}{H}\sum_x \frac{P_{xx}}{v_x + \sum_y P_{xy} - P_{xx}}$ measures ground truth of prognostic.

VIII. RESULTS ON SEMANTIC SEGMENTATION THROUGH ANN

The data set consists of 1,464, 1,449, and 1,456 images - training, validation and testing, execution is done on MATLAB contextual modelling of visual object classes. ANN method with classifiers is trained using images that gave accuracy score of 64.2%. The accuracy scores are shown in Table I.

TABLE I. Individual category results of ANN on the MATLAB Contextual Modeling accuracy scores.

Classifier	Intersecrion Over Union Accuracy (%)	
ANN	**Road**	**64.2**
	Pedestrain	50.1
	Building	39.2
	Sofa	32.2
	Chair	24.7

IX. RESULTS ON SEMANTIC SEGMENTATION BY ADABOOST

The Adaboost method with classifiers is trained using images that gave accuracy score of 53.2%. The accuracy scores are shown in Table II.

TABLE II. Individual category results of Adaboost on the MATLAB Contextual Modeling accuracy scores.

Classifier	Intersecrion Over Union Accuracy (%)	
Adaboost	Road	55.7
	Pedestrain	58.7
	Building	44.4
	Sofa	**53.2**
	Chair	31.5

X. RESULTS ON MATLAB FOR SEMANTIC SEGMENTATION BY DEEP MODEL

The simulation of the introduced technique are verified on challenging SS test sets. They are MATLAB contextual modelling that unfolds different kind of scene images like counting indoor and outdoor scenes etc. The simulation of the proposed technique got outperforming performance on the above stated test sets.

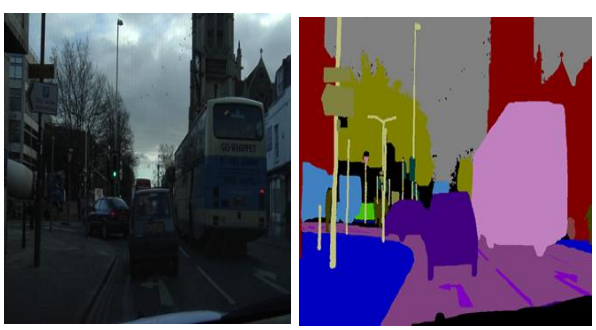
A. OUTCOME ON THE MATLAB CONTEXTUAL MODELLING TEST SET

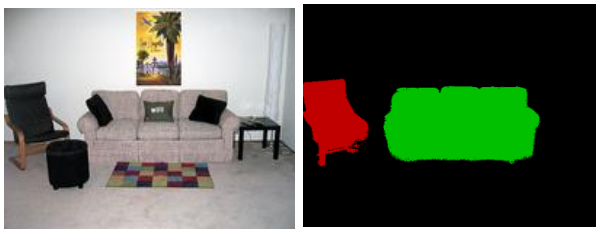
The comparison has been done for various techniques with outstanding performance. The proposed model is trained using images that gave accuracy score of 71.2%, outperforming other techniques. The accuracy scores are shown in Table III.

TABLE III. Individual category results of Deep model on the MATLAB Contextual Modeling accuracy scores.

Classifier	Intersecrion Over Union Accuracy (%)	
CRF	Road	66.3
	Pedestrain	34.6
	Building	53.2
	Sofa	**71.2**
	Chair	35.1

The prognostic outputs of method on Matlab contextual modeling given in Fig. 3 – original image and prognostic.





(a) Original image (b) Prognostic

Fig. 3. Some Prognostic examples of Deep model on MATLAB contextual modeling.

XI. CONCLUSION

Though there are abundant hardware designed for Deep model; it becomes a challenging task for a researcher to design a new structured model for real time applications. SS has many applications in real world. Researchers in this field always diagnose various possibilities for design of effective SS. This thesis attempts at improvising by combining various available techniques to increase the intersection over union accuracy. We achieve an accuracy of 71.2%. According to the literature review a single technique can satisfy all the needs required by a Deep model. The research was done to pick out the relevant techniques suited for elevating accuracy and SS performance.

REFERENCES

1. C. Farabet, C. Couprie, L. Najman, and Y. LeCun, Learning hierarchical features for scene labeling. IEEE T. Pattern Analysis & Machine Intelligence, 2013.
2. J. Long, E. Shelhamer, and T. Darrell. Fully convolutional networks for semantic segmentation. In Proc. IEEE Conf. Comp. Vis. Pattern Recogn., 2015.
3. L. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille. Semantic image segmentation with deep convolutional nets and fully connected CRFs. In Proc. Int. Conf. Learning Representations, 2015.
4. G. Heitz and D. Koller. Learning spatial context: Using stuff to find things. In Proc. European Conf. Computer Vision, 2008.
5. M. Mostajabi, P. Yadollahpour, and G. Shakhnarovich. Feedforward semantic segmentation with zoom-out features. In Proc. IEEE Conf. Comp. Vis. Pattern Recogn., 2015.
6. C. A. Sutton and A. McCallum. Piecewise training for undirected models. In Proc. Conf. Uncertainty Artificial Intelli, 2005.
7. Mark S. Nixon, Alberto S.Aguado. Feature Extraction and Image Processing for Computer Vision, 2012.
8. A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems, 1097–1105, 2012.
9. Guosheng Lin, Chunhua Shen, Anton van den Hengel and Ian Reid. Efficient Piecewise Training of Deep Structured Models for Semantic Segmentation. Computer Vision and Pattern Recognition, 3195-3203, 2016.
10. A. Kolesnikov, M. Guillaumin, V. Ferrari, and C. H. Lampert. Closed-form training of conditional random fields for large scale image segmentation. In Proc. Eur. Conf. Comp. Vis., 2014.

11. L. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille. Semantic image segmentation with deep convolutional nets and fully connected CRFs. In Proc. Int. Conf. Learning Representations, 2015.
12. S. Zheng, S. Jayasumana, B. Romera-Paredes, V. Vineet, Z. Su, D. Du, C. Huang, and P. Torr. Conditional random fields as nrecurrent neural networks. In Proc. Int. Conf. Comp. Vis., 2015.
13. K. Hornik, M. Stinchcombe and H. White. Multilayer feedforward networks are universal approximators. Neural Networks, 2: 359-366, 1989.
14. S. Sunny, D. Peter, K. Jacob. Performance of different classifiers in speech recognition. International Journal of Research and Engineering Technology, 2(4), 2013.
15. Rashidul Hasan, Mustafa Jamil, Golam Rabbani and Saifur Rahman. Speaker Identification Using Mel Frequency Cepstral Coefficients. 3rd International Conference on Electrical and Computer Engineering, Dhaka, Bangladesh, 2004.
16. J. L. McClelland and D. E. Rumelhart. Explorations in Parallel Distributed Processing. MIT Press, Cambridge, MA, 1988.
17. G. Gudj onsson. Theory and Applications of Neural Circuits. Independent Study, Electrical Engineering Department, University of North Dakota, 1989.
18. J. Park, F. Diehl, M.J.F. Gales, M. Tomalin and P.C. Woodland. Training and Adapting MLP features for Arabic Speech Recognition. In Proc. of IEEE ICASSP, 2009.
19. P. Krahenbuhl and V. Koltun. Efficient inference in fully connected CRFs with Gaussian edge potentials. In Proc. Adv. Neural Info. Process. Syst., 2012.
20. L. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille. Semantic image segmentation with deep convolutional nets and fully connected CRFs. In Proc. Int. Conf. Learning Representations, 2015.
21. S. Zheng, S. Jayasumana, B. Romera-Paredes, V. Vineet, Z. Su, D. Du, C. Huang, and P. Torr. Conditional random fields as nrecurrent neural networks. In Proc. Int. Conf. Comp. Vis., 2015.
22. J. Dai, K. He, and J. Sun. BoxSup: Exploiting bounding boxes to supervise convolutional networks for semantic segmentation. In Proc. Int. Conf. Comp. Vis., 2015.
23. D. Eigen and R. Fergus. Predicting depth, surface normals and semantic labels with a common multi-scale convolutional architecture. In Proceedings of the IEEE International Conference on Computer Vision, 2015.
24. H. Noh, S. Hong, and B. Han. Learning deconvolution network for semantic segmentation. In Proc. Int. Conf. Comp. Vis., 2015.
25. B. Hariharan, P. Arbel´aez, R. Girshick, and J. Malik. Hypercolumns for object segmentation and fine-grained localization. In Proc. IEEE Conf. Comp. Vis. Pattern Recogn., 2014.