Research Paper on Point Based Rendering System for Self Driving Cars

Anjani Pandey[1], Asawari Durape[2], Ananta Kumari[3], Dr Pankaj Chandre[4]
[1,2,3]*Students, Computer Science & Engineering, MIT School of Engineering, MIT ADT University, Pune*
[4]*Associate Professor, Computer Science & Engg, MIT School of Engineering, MIT ADT University, Pune*

Abstract: **This research paper encapsulates all the processes required to make a Point Based Rendering System for Self Driving Cars. Our project has two different scopes, first scope is concept based scope, while the second scope is application based.**
For the scope concerned with the concept, we have developed such a neural network module that can be applied to both instance as well as semantic segmentation. Using this module, we can successfully achieve more clear mask resolution.
Another scope that we have worked on is the application scope, here we have used the scope of self driving cars. More and more people are buying self driving cars everyday, and this makes the scope of self driving cars very intriguing. Since self driving cars' processing is very much dependent on processing of images and segmentation of the same. Much of the processing of the self driving cars is dependent of processing of images every second. We believe using this project will tremendously benefit the scope of self driving cars.

Index Term **- Iterative subdivision, Point based image segmentation, Point based image segmentation for self-driving cars, Point based rendering, algorithm**

I INTRODUCTION

Image segmentation is a process where segmented to different sets. It is often the case where these methods oversample unnecessary sections and under sample important sections. This is were we can use Subdivision Iterative Algorithm.
Map pixels on a regular grid to a label map, or a series of label maps, on the same grid in image segmentation tasks. The label map indicates the expected category for each pixel in semantic segmentation. For each recognised object, a binary foreground vs. background map is projected using instance segmentation. Convolutional neural networks are the modern tools of choice for these problems (CNNs). [1, 2]
In the processing of CNNs, regular grids are used. Basically the image that passes through the processing is of regular grid which consists of pixels. Furthermore, the elusive representations are also on a regular grid, and for the last point the outputs are based on label maps, and these maps are also regular grids.
Points stated so far are a not necessarily efficient for image segmentation. The reason why these networks are not ideal is because, the label maps that are predicted by these networks are quite even and straightforward. This happens because the neighbouring pixels often take up the same label. In overall scope of things, this results in oversampling of the unwanted or featureless areas, and undersampling of important areas.
Our neural network, Point Based Rendering System on the other hand use Subdivision Iterative Algorithm (SIA). By using the basis of this algorithm, we have introduced our neural network. Point Based Rendering System only choose the ambiguous point from the coarse grid, and adaptively make the prediction grid more and more refined with each step.

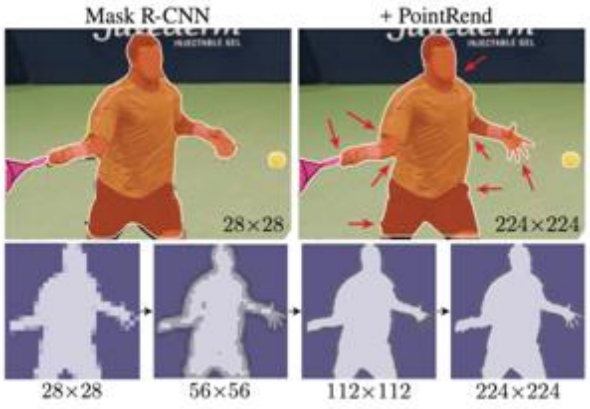

Figure 1: In this network, we start with a coarse grid of 7x7 and interpolated bi-linearly for 5-steps till the mask grid is of 224x224. In the above given figure, the mask head of Mask R-CNN is replaced [3] with Point Based Rendering System's mask.
Use of SIA assures that the ambiguous edges are interpolated instead of the entire grid going through the interpolation. This reduces the computational power required for the segmentation and along with that it also reduces the time required for the processing. It can also be seen in the above given images that instance segmentation with the our module yields better and refined masks.
For complete comparison and visualisation we are going to use Cityscapes [4] and COCO [5] datasets.

II RELATED WORK

We can look at the scope of rendering related with our problem statement. The algorithms that are used to solve rendering problems, usually output regular grid of pixels. The process that these algorithms follow process the prediction over random points.

Different useful processes like subdivision [6] and adaptive sampling [7, 8] refine coarse prediction with points that have different values, making it random.

One of the popular image segmentation techniques is instance segmentation. This technique is based on Mask R-CNN architecture [9]. Instance segmentation only predicts masks on 28x28 grid. This could be an issue for large and complex images. It also gives output that is not refined and the mask predict is not accurate upto the mark. This would raise and issue when dealing with feature-loaded images, and also the scope where the features should be mapped out clearly. There are different methods that do produce better outputs that are detailed. One such type of technique is called bottom-up approach, this technique processes group of pixels to form object masks [10, 11, 12]. There is another technique called TensorMask [13], this technique yields better mask outputs, but it still is not efficient when it comes to yielding high-resolution masks.

Another very important image segmentation technique is semantic segmentation. Semantic segmentation primarily uses FCNs [14]. This technique predicts masks that are of lower quality than the input image. The results later on are made better dilated convolution that replace some subsampling layers [15, 16]. There are techniques that can be used to improve the mask outputs, one such architecture is encoder-decoder architecture [17, 18, 19, 20]. This technique subsamples the grid in the endear and then upsample it in the decoder, using skip connections [19] to yield finer grid masks.

III METHOD

We have used Subdivision Iterative Algorithm, this algorithm helps with upsampling the relevant regions. Neural network built on this principle would retrieve and output better and clear and refined masks for both semantic as well as instance segmentation.

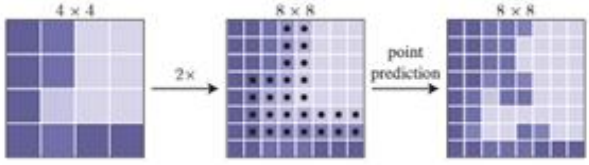

Figure 2: In the figure above, we can see the basic working of Subdivision Iterative Algorithm. A 4x4 grid predicts 8x8 grid. This is achieved by upsampling the 4x4 grid using bilinear interpolation. The SIA works on the most ambiguous points, in this case the dots seen of the grid 2. It can also be noticed the instead of interpolating the entire, interpolating the dots will yield better and fined edge.

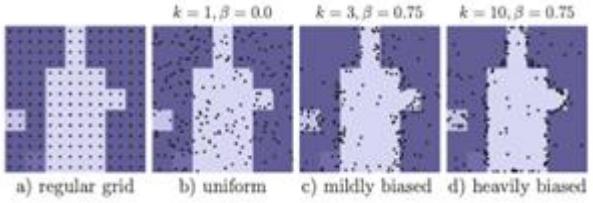

Figure 3: For training we use combination of k and beta to achieve different biasing and uniformity. The strategy is also used with the underlying coarse prediction. To achieve efficiency during the training, the best option to choose is madly biased along with the coarse underlying prediction.

For Point Based Rendering System we have used points as the basis of the traversal. This module accepts one or more CNN features maps that are used in the processing of C channels, where $f \in RC \times H \times W$. Here each channel is defined over a grid that is 4x to 16x coarser than the image grid [21].

Point Based Rendering System consists of three main processes: (i) Taking the basis of Subdivision Iterative Algorithm into consideration, this module chooses limited number of points for making predictions on. This would result in only necessary extraction and prediction. Otherwise it would have resulted in oversampling of unnecessary points and increase the computational power and time. (ii) Once the relevant predictions are made, the next step to further interpolate the predicted points. This is done bilinear interpolated to the predicted points, and using the 4 nearest neighbours of the each predicted points. Another important consideration is to predict such a grid that is more refined than the previous step. (iii) For each predicted points, a mask head that is made specially for Point Based Rendering System comes into play.

Unique feature about Point Based Rendering System is that, it can be applied to instance segmentation as well as semantic segmentation. Processing of instance segmentation is based on Mask R-CNN [3] and processing of semantic segmentation is based on FCNs [22].

Training: For training we have used different procedure than inference, since inference is based on training rather than it being the other way around. For training, we had to pay attention to uncertain points and also uniform coverage. Now both of these things can not be achieved

if we used SIA, for this we have used similar approach though. We have used non-repetitive strategy with basis on random sampling.
This strategy selects N points to start with, these points are based on feature map. This strategy is designed to select uncertain regions and at the same time gained uniform coverage. This can be see in the fig 3.

Inference: The first step for inference would be to select such points where the values are very different from other points. Basically choosing such points that have drastically different values from that of neighbouring points. This is how we select new points, since we have used Subdivision Iterative Algorithm, other points have be obtained from interpolating of already computed points. For every region, whatever the output mask is, it will be iteratively refined. The output mask will go from coarse to fine mask. The first mask output or rather the first level prediction is the coarsest. This is done by using standard coarse segmentation head. There are multiple iteration needed to compute the result of the relevant resolution. For each iteration, Point Based Rendering System upsamples the previous segmentation prediction. This is done by using bilinear interpolation, after the process of bilinear interpolation it chooses N points which are the most ambiguous, now these N points are available on much denser grid. Labels for these N points are predicted. This entire procedure is iterated over and over again till we have desired output resolution.

Point-head: Our model, Point Based Rendering System processes point-wise segmentation. The prediction of this point-wise segmentation is done using a very simple multi-layer perceptron. Using multi-layer perceptron helps in equally sharing the weights across all the predicted points.

Architecture:The processing uses Mask R-CNN [3] with ResNet-50 [23] + FPN [24] backbone. For comparison, the module first uses the default mask head of Mask R-CNN, this mask head is a region wise FCN. For the later part of comparison we use a different mask head that specially designed for Point Based Rendering System.

Datasets: For datasets, we have used two different datasets COCO [5] and CityScapes [4]. For our module we would us AP metric for instance segmentation and mIoU for semantic segmentation.
COCO dataset has many different categories while Cityscapes primarily has street view of different instances. Cityscapes has 8 categories to be specific. Cityscapes has more detailed images when compares to COCO, hence Cityscapes has finer resolution when compared to COCO.

IVIMPLEMENTATION

A coarse prediction to be predict first, this is done by replacing the default mash head from the architecture with a lighter box-head. This new mask head produces a 7x7 mask. In second level of FPN a 14x14 grid is predicted. This continues till stride 5, when the grid prediction is 224x224.
At each level, the prediction grid gets finer, that is in 5 strides it goes from 7x7 to 224x224 grid prediction. The prediction points are further analyse with their neighbouring points in the next stride. This is done by MLP, this MLP has 3 different hidden layers with 256 channels.
For training schedule we 1x training schedule from Detetctron2 [25]. For further processing we have used different combinations of k and β. While training, we have used Mask R-CNN, here the mask heads runs at the same time as the box. But it has been deduced that Point Based Rendering system would yield better results only if the mask head and the box are processed in order.
We have used Subdivision Iterative Algorithm for inference. Here the prediction starts with a coarse grid of 7x7 and moves up to 224x224 in five steps. Use of Subdivision Iterative Algorithm uses less FLOPs when compared to other methodologies. Point Based Rendering System is able to do so my ignoring such areas that does not need refining and can be deduced with coarse prediction. When it comes to complex edges and more boundaries it make sense to use more points. This would result better and more crisper images.
Point Based Rendering System is dependent on the process of creating point-wise features at few selected points. The features extractions is done by depending on two different features: fine-grained and coarse prediction features. Basically for features predictions, this model is dependent on two different features, fine-grained and coarse predictions.

Fine-grained features: Fine-grained features are used to help Point Based Rendering System to extract finer details. This is done by extracting a feature vector at every sampled point from the CNN feature maps. At every sampled point, we perform bilinear interpolation, since each point is a real-value 2D coordinated, we can use bilinear interpolation on the feature maps to produce the feature vector using the standard procedure [3, 26, 27]. These features can be simply extracted using one feature map or multiple feature map or feature pyramid can also be used to extract features.

Coarse prediction features: There are a few shortcomings when we consider fine-grained features. Fine-grained features are optimal for refined details but it would not contain region-specific information. Since it lacks in region-specific information, it results in same points being overlapped by two different bounding box of instances. But the point can be in boundary of only one instance, this would cause different regions predicting different labels for the same point. This can be resolved if we had additional information.

For resolving this issue, the second prediction feature could be of coarse type. This coarse prediction happens from the network that is K-dimensional vector prediction for each point in the region. This coarse prediction produces very globalised information and at same time the channels convey the semantic classes. These predictions which are of coarse kind are similar to the results given out by existing architectures.

IVMAIN RESULTS

Point Based Rendering System so far is processable with instance and semantic segmentation, but not limited to them in any ways. We have tried to demonstrated how our module would help instance and semantic segmentation. In future, this module can also be considered helpful in other fields of applications and also with other modules likewise.

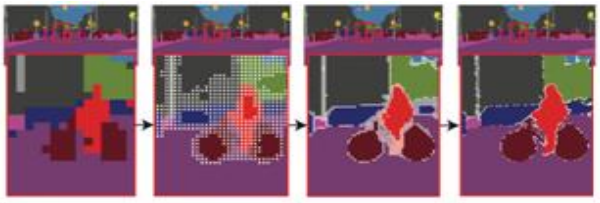

Figure 4: This is Point Based Rendering System used for semantic segmentation.

Figure 5-8: This is Point Based Rendering System used for instance segmentation.

REFERENCES

[1] Alex Krizhevsky, Ilya Sutskever, and Geoff Hinton. ImageNet classification with deep convolutional neural networks. In NIPS, 2012.

[2] Yann LeCun, Bernhard Boser, John S Denker, Donnie Henderson, Richard E Howard, Wayne Hubbard, and Lawrence D Jackel. Back propagation applied to handwritten zip code recognition. Neural computation, 1989.

[3] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask R-CNN. In ICCV, 2017.

[4] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. The Cityscapes dataset for semantic urban scene understanding. In CVPR, 2016.

[5] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft COCO: Common objects in context. In ECCV, 2014.

[6] Turner Whitted. An improved illumination model for shaded display. In ACM SIGGRAPH Computer Graphics, 1979.

[7] Don P Mitchell. Generating antialiased images at low sampling densities. ACM SIGGRAPH Computer Graphics, 1987.

[8] Matt Pharr, Wenzel Jakob, and Greg Humphreys. Physically based rendering: From theory to implementation, chapter 7. Morgan Kaufmann, 2016.

[9] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask R-CNN. In ICCV, 2017.

[10] Shu Liu, Jiaya Jia, Sanja Fidler, and Raquel Urtasun. SGN: Sequential grouping networks for instance segmentation. In CVPR, 2017.

[11] P. R. Chandre, P. N. Mahalle and G. R. Shinde, "Machine Learning Based Novel Approach for Intrusion Detection and Prevention System: A Tool Based Verification," 2018 IEEE Global Conference on Wireless Computing and Networking (GCWCN), 2018, pp. 135-140, doi: 10.1109/GCWCN.2018.8668618.

[12] Alexander Kirillov, Evgeny Levinkov, Bjoern Andres, Bog- dan Savchynskyy, and Carsten Rother. InstanceCut: from edges to instances with multicut. In CVPR, 2017.

[13] Chandre, P.R., Mahalle, P.N., Shinde, G.R. (2020). Deep Learning and Machine Learning Techniques for Intrusion Detection and Prevention in Wireless

Sensor Networks: Comparative Study and Performance Analysis. In: Das, S., Samanta, S., Dey, N., Kumar, R. (eds) Design Frameworks for Wireless Networks. Lecture Notes in Networks and Systems, vol 82. Springer, Singapore. https://doi.org/10.1007/978-981-13-9574-1_5

[14] Deshmukh, S. S., & Chandre, P. R. Survey on: Naive Bayesian and AOCR Based Image and Text Spam Mail Filtering System. International Journal of Emerging Technoogy and Advanced Engineering.

[15] 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. PAMI, 2018.

[16] Liang-Chieh Chen, George Papandreou, Florian Schroff, and Hartwig Adam. Rethinking atrous convolution for semantic image segmentation. arXiv:1706.05587, 2017.

[17] Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation. In ECCV, 2018.

[18] Alexander Kirillov, Ross Girshick, Kaiming He, and Piotr Dollár. Panoptic feature pyramid networks. In CVPR, 2019.

[19] Chandre, P.R., Mahalle, P.N., Shinde, G.R. (2021). Intrusion Detection and Prevention Using Artificial Neural Network in Wireless Sensor Networks. In: Patil, V.H., Dey, N., N. Mahalle, P., Shafi Pathan, M., Kimbahune, V.V. (eds) Proceeding of First Doctoral Symposium on Natural Computing Research. Lecture Notes in Networks and Systems, vol 169. Springer, Singapore. https://doi.org/ 10.1007/978-981-33-4073-2_12

[20] Ke Sun, Yang Zhao, Borui Jiang, Tianheng Cheng, Bin Xiao, Dong Liu, Yadong Mu, Xinggang Wang, Wenyu Liu, and Jingdong Wang. High-resolution representations for labeling pixels and regions. arXiv:1904.04514, 2019.

[21] Alexander Kirillov, Yuxin Wu, Kaiming He, Ross Girshick. PointRend: Image Segmentation as Rendering.

[22] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In CVPR, 2015.

[23] Chandre, P.R., Mahalle, P.N., Shinde, G.R. (2021). Intrusion Detection and Prevention Using Artificial Neural Network in Wireless Sensor Networks. In: Patil, V.H., Dey, N., N. Mahalle, P., Shafi Pathan, M., Kimbahune, V.V. (eds) Proceeding of First Doctoral Symposium on Natural Computing Research. Lecture Notes in Networks and Systems, vol 169. Springer, Singapore. https://doi.org/10.1007/978-981-33-4073-2_12

[24] Tsung-Yi Lin, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie. Feature pyramid networks for object detection. In CVPR, 2017.

[25] Yuxin Wu, Alexander Kirillov, Francisco Massa, Wan-Yen Lo, and Ross Girshick. Detectron2.

[26] Max Jaderberg, Karen Simonyan, Andrew Zisserman, and Koray Kavukcuoglu. Spatial transformer networks. In NIPS, 2015.

[27] Jifeng Dai, Haozhi Qi, Yuwen Xiong, Yi Li, Guodong Zhang, Han Hu, and Yichen Wei. Deformable convolutional networks. In ICCV, 2017.