3D POINT CLOUD CLASSIFICATION USING DEEP LEARNING

Sangeetha Kalyanaraman
Associate Prof. Computer Science
Panimalar Engineering College
(Affiliated to Anna University)
Chennai, India

Jennisha Christina Martin
B.E Computer Science
Panimalar Engineering College
(Affiliated to Anna University)
Chennai, India

Sivasankari R
B.E Computer Science
Panimalar Engineering College
(Affiliated to Anna University)
Chennai, India

Sakthi R V
B.E Computer Science
Panimalar Engineering College
(Affiliated to Anna University)
Chennai , India

***Abstract* — In the previous couple of years, the provision of 3D content continues to be less than 2D counterpart. Hence many 2D-to-3D image conversion methods are proposed. Methods involving human operators are most successful but also tedious and expensive. Automatic methods that make use of a deterministic 3D scene model, have not yet achieved the identical level of quality for they depend on assumptions that are often violated in practice. Here two types of methods are developed. The primary is predicated on learning some extent mapping from local image/ attributes, like color, spatial position. The second method relies on globally estimating the complete depth map of a question image directly from a repository of 3D images employing a Nearest-neighbour regression type idea. It demonstrates the flexibility and therefore the computational efficiency of the methods on numerous 2D images and discusses their pitfall and benefits**

I. INTRODUCTION

It demonstrates the flexibility and also the computational efficiency of the methods on numerous 2D images and discusses their drawbacks and benefits. Point cloud learning has lately attracted increasing attention thanks to its wide applications in many areas, like computer vision, autonomous driving, and robotics. As a dominating technique in AI, deep learning has been successfully wont to solve various 2D vision problems. However, deep learning on point clouds remains in its infancy thanks to the unique challenges faced by the processing of point clouds with deep neural networks. Recently, deep learning on point clouds has become even thriving, with numerous methods being proposed to deal with different problems during this area. Video segmentation is a vital building block for prime level applications, like scene understanding and interaction analysis. While outstanding results are achieved during this field by the state-of-the-art learning and model-based methods ,they are restricted to certain styles of scenes or require an outsized amount of annotated training data to realize object segmentation in generic scenes. Various deep learning algorithms are proposed for the segmentation of point clouds. The point transformation applied basically in real time because it's supported purely local image attributes, like color, spatial position, and motion at each pixel. A second method is developed that estimates the world depth map of a question image or video frame directly from a repository of 3D images (image + depth pairs or stereopair) employing a nearest-neighbor regression type

idea.

The key observation is that among innumerable 3D images available on-line, there likely exist many whose 3D content matches that of a 2Dinput . An assumption is created that two images that are photometrically similar even have similar 3D structure (depth).Given monocular query image Q, assumed to be the left image of a stereopair that's to be computed, relies on the above observation and assumption to “learn” the complete depth from a repository of 3D images. Search for representative depth fields: Find k 3D images within the repository I that have most similar depth to the query image, for instance by performing a k nearest-neighbor (kNN) search employing a metric supported photometric properties. Depth fusion: Combine the k representative depth fields, for instance, by means of median filtering across depth field. Depth smoothing: Process the fused depth field to get rid of spurious variations, while preserving depth discontinuities, for instance, by means of cross-bilateral filtering. Stereo rendering: Generate the correct image of a fictitious stereopair using the monocular query image and also the smoothed depth field followed by suitable processing of occlusions and newly-exposed areas. In the previous few years, the supply of 3D content continues to be but 2D counterpart. Hence many 2D-to-3D image conversion methods are proposed. Methods involving human operators are most successful but also time- consuming and costly. Automatic methods, that make use of a deterministic 3D scene model, haven't yet achieved the identical level of quality for they depend on assumptions that are often violated in practice. Here two varieties of methods are developed. The first is predicated on learning some extent mapping from local image/attributes, like color, spatial position. The second method is predicated on globally estimating the whole depth map of a question image directly from a repository of 3D images (image + depth pairs or stereo pairs) employing a nearest-neighbour regression type idea. Point cloud semantic segmentation plays a critical role in autonomous driving, robot navigation, augmented reality and 3D reconstruction. Currently, with the event of deep learning technology, semantic segmentation has made great progress, but it also faces many difficulties. The unordered and unstructured properties of 3D point clouds make it difficult to be presented as 2D images. the most contributions of our work are listed below: We design a brand new network MH Net, which may fully incorporate both local and global multiscale hierarchical features for highly accurate point cloud semantic segmentation.

II. RELATED WORKS

Some of the research works carried out by researches as related to 3D Point Cloud Classification are discussed in the succeeding Sub-sections.

2.1 Yuxing Xie; Jiaojiao Tian; Xiao Xiang Zhu

3D Point Cloud Semantic Segmentation(2019)

Here, the merits are :

.

1. In color (RGB, multi-spectral) information suitable for large area. High accuracy suitable for large area not affected by weather global data is available compared to ALS complete Building facade information is available 4D Information middle accuracy not affected by the weather

And , the Demerits are :

1. Affected by mirror reflection long scanning time Close-range limited accuracy.

2.2 Xiaoli Liang; Zhongliang Fu.
MH Net: Multiscale Hierarchical Network for 3D Point Cloud Semantic Segmentation (2019).

Here , the merits are :
To take fully advantage of the correlations of Propagated information between the different scale coarse layers and the original points, the Local features of each scale are characterized by feature propagation to obtain the features of the original point clouds at the corresponding scale.

And the Demerits are :
Although there is considerable noise points in the segmentation result of MH Net, the number of noise point’s only accounts for a small part of the total point clouds. Even If all the noise points are corrected, the overall accuracy of the point clouds will not be greatly improved

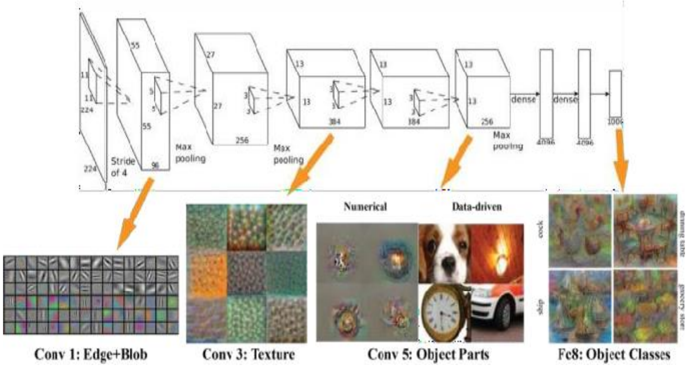
2.3 Yulan Guo;Hanyun Wang;Qingyong Hu; Hai Liu ;Li Liu ; Mohammed Bennamoun. Deep Learning for 3D Point Clouds (2020)

Here the merits are :
A promising solution is to address the raw point clouds with the ConvNets. Since ConvNets has the advantage of overlapping during Convolutional operation may benefit the future**.**

And the demerits are:
It is very expensive and it has included some challenges to over it the irregular network connection will reduce the performance level in the system.

III. METHODOLOGY

In recent years, there has been an increasing demand for applications to monitor the targets related to land-use, using remote sensing images. Proposed the automatic approach to localize and detect building footprints, road networks and vegetation areas. Automatic interpretation of visual data is a comprehensive task in computer vision field. The Deep learning approaches improve the capability of classification in an intelligent way. Deep Learning algorithms gives high accuracy compared to the semi supervised machine learning algorithms.



SYSTEM ARCHITECTURE

IV. FUNCTIONALITIES

MODULE 1 :

SEARCH FOR REPRESENTATIVE DEPTH FIELDS :

Find k 3D images in the repository I that have most similar depth to the query image, for example by performing a k nearest-neighbor(kNN) search using a metric based on photometric Properties.

MODULE 2: *DEPTH FUSION*

Combine the k representative depth fields, for example, by means of median filtering across depth field.

MODULE 3 : *DEPTH SMOOTHING*

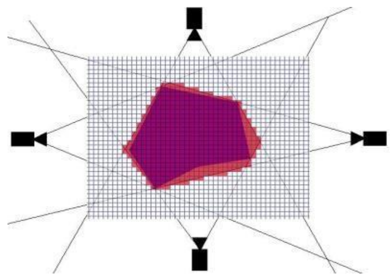
Process the fused depth field to remove spurious variations, while preserving depth discontinuities, for example, by means of cross- bilateral filtering.

MODULE 4 : *STEREO RENDERING*

Generate the right image of a fictitious stereopair using the monocular query image and the smoothed depth field followed by suitable processing of occlusions and newly-exposed areas.

TESTING

A. Idea: Find a shape consistent with images
   I. Shape: voxel grid
   II. For each voxel -> compute occupied or free
B. Shape from Silhouette
   I. Extract silhouette images from different views
   II. Project voxel on each image
      Lies within silhouette - occupied
      Otherwise - free
C. Other methods: voxel coloring, shape carving
   But requires images to segmented, cameras calibrated.



Performance:

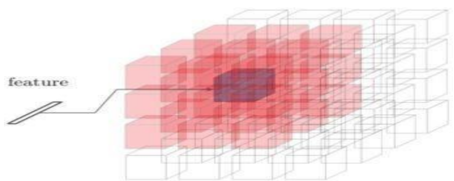
A. Enables estimating 3D shape with fewer viewpoints

I. Extreme case: a single view point
II. Earliest work: shape from shading
III. Assumes lambertian reflectance&constant Albedo.

B. Includes prior knowledge on

I . reflectance, shape, viewpoints,
II. learnt from data

C. Less reliant on finding accurate feature correspondences



3D Recurrent Unit:

A . LSTM units arranged in 3D grid structure
B . A unit receives
   I. A feature vector from encoder
   II. Hidden states of neighbour by convolution
C. Each unit reconstructs a part of final output.

V. IMPLEMENTATION

Training data:

I. 3D CAD for input images and ground truth occupancy grid

II. ShapeNets, PASCAL 3D, Online Products
III. images augmented with random crops from Pascal
IV. Viewpoints sampled randomly.

Training:

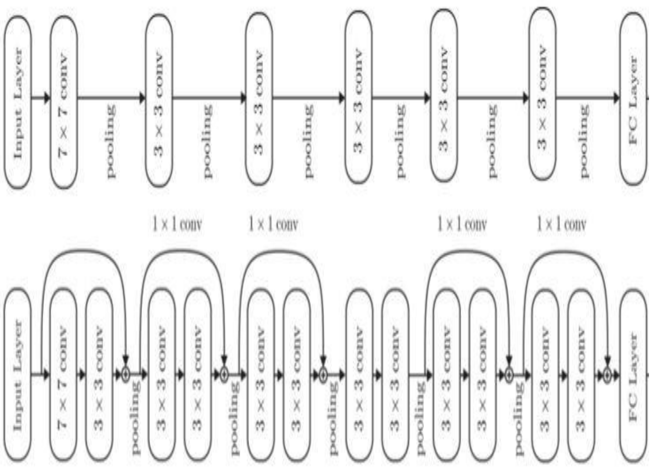
I. Variable length inputs across different mini Batches.
II. Combine single & multi view reconstruction.

Network:

I. Input 127x127, Output 32x32x32
II. 60k iterations, leaky relu (slope: 0.1)
III. Implemented in Theano, Adam for SGD update

Encoder:

A. Encode image into low dim features .
B. Involves standard convolution, pooling and fully connected layers.
C. Uses leaky Relu as activation function D. Also has a residual variant .



Decoder:

A.The hidden state is passed to the decoder.
B.Decoder up samples to target output resolution,
I. Applies nonlinearities, 3D deconvolution,
Unpooling.
C.Output of last activation is converted to occupancy probability,
I.Uses voxel-wise softmax.
D.Finally we minimize the following cross entropy loss and back propagate.

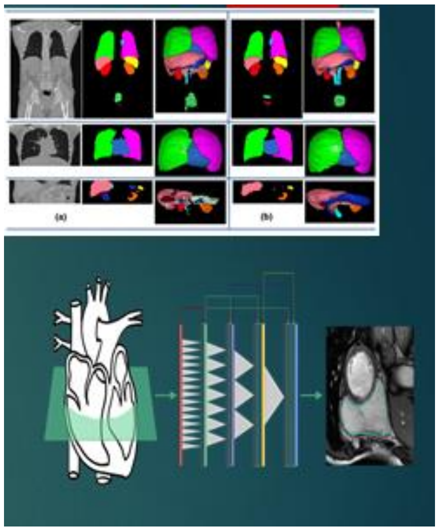
$$L(\mathcal{X}, y) = \sum_{i,j,k} y_{(i,j,k)} \log(p_{(i,j,k)}) + (1 - y_{(i,j,k)}) \log(1 - p_{(i,j,k)})$$

GRU Variant:

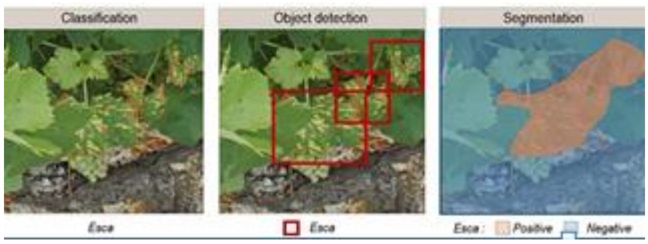
$$u_t = \sigma(W_{fx}\mathcal{T}(x_t) + U_f * h_{t-1} + b_f)$$
$$r_t = \sigma(W_{ix}\mathcal{T}(x_t) + U_i * h_{t-1} + b_i)$$
$$h_t = (1 - u_t) \odot h_{t-1} + u_t \odot \tanh(W_h \mathcal{T}(x_t) + U_h * (r_t \odot h_{t-1}) + b_h)$$

VI. FUTURE USE

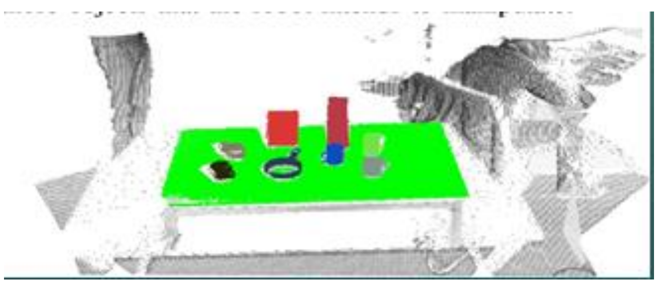
In medical science for identifying cancer, tumors ,by classifying the 3d points rendering high dimensional internal views. .



In Agriculture imaging, it might used for crop detection



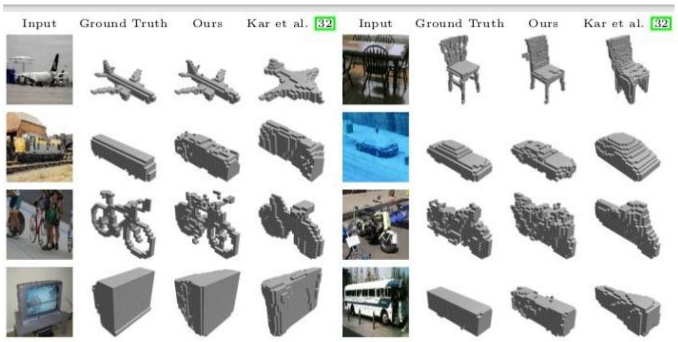
In Robotics , to enable high level vision for robots to understands and perceive things as humans do.



VII. RESULT ANALYSIS

A typical 2D-to-3D conversion process consists of two steps: depth estimation for a given 2D image and depth based rendering of a new image in order to form a stereo pair. While the rendering step is well understood, the challenge is in estimating depth from a single image. Therefore, throughout the focus is on depth recovery.

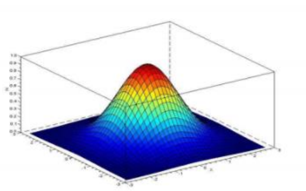
SINGLE VIEW CONSTRUCTION

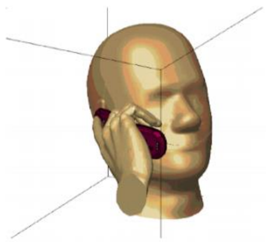


MULTI VIEW 3D RECONSTRUCTION

ESTIMATE A 3D SHAPE GIVEN A SET OF IMAGES

Stereo reconstruction obtain point correspondences and camera calibration with estimate internal and external params. Calculate projection rays with intersection gives 3D point.





VIII. CONCLUSION

A new class of methods is proposed to aim at 2D-to-3D image conversion that is based on the radically different approach of learning from examples. One method that is proposed is based on learning a point mapping from local image attributes to scene-depth. The other method is based on globally estimating the entire depth field of a query directly from a repository of image + depth pairs using nearest-neighbor-based regression. It objectively validates the algorithms' performance against state-of-the-art algorithms. While the local method was outperformed by other algorithms, it is extremely fast as it is, basically, based on table look-up. However, the global method performed better than the state-of-the-art algorithms in terms of cumulative performance across two datasets and two testing methods, and has done so at a fraction of CPU time.

IX. REFERENCES

[1] A. Boulch, B. Le Saux, and N. Audebert, ''Unstructured point cloud semantic labeling using deep segmentation networks,'' in Proc. Eurographics Workshop 3D Object Retr., 2017, pp. 17–24.

[2] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, ''DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs,'' IEEE Trans. Pattern Anal. Mach. Intell., vol. 40, no. 4, pp. 834–848, Apr. 2018.

[3] E. Shelhamer, J. Long, and T. Darrell, ''fully convolutional networks for semantic segmentation,'' IEEE Trans. Pattern Anal. Mach. Intell., vol. 39, no. 4, pp. 640–651, Apr. 2017.

[4] A. Garcia-Garcia, S. Orts-Escolano, S. Oprea, V. Villena-Martinez, and J. Garcia-Rodriguez, ''A review on deep learning techniques applied to semantic segmentation,'' Apr. 2017, arXiv:1704.06857. [Online]. Available: https://arxiv.org/abs/1704.06857.

[5] L. P. Tchapmi, C. B. Choy, I. Armeni, J. Gwak, and S. Savarese, ''SEGCloud: Semantic segmentation of 3D point clouds,'' Oct. 2017, arXiv: 1710.07563. [Online]. Available: https://arxiv.org/abs/1710.07563.

[6] D. Maturana and S. Scherer, ''VoxNet: A 3D convolutional neural network for real-time object recognition,'' in Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS), Sep./Oct. 2015, pp. 922–928.

[7] Z. Wu, S. Song, A. Khosla, F. Yu, L. Zhang, X. Tang, and J. Xiao, ''3D ShapeNets: A deep representation for volumetric shapes,'' in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2015, pp. 1912–1920. [8] C. R. Qi, H. Su, M. Kaichun, and L. J. Guibas, ''PointNet: Deep learning on point sets for 3D classification and segmentation,'' in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jul. 2017, pp. 77–85.

[9] R. Zhang, ''Research on polymorphic object semantic segmentation of complex 3D scenes based on laser point clouds,'' Ph.D. dissertation, Dept. Survey Map, PLA Strategic Support Force Inf. Eng. Univ., Zhengzhou, China, 2018.

[10] L. Wang, ''Intelligent segmentation of point cloud based on PointNet network,'' M.S. thesis, Dept. RS Inf. Eng., Wuhan Univ., Wuhan, China, 2018.