Taylor & Francis
Taylor & Francis Group

Journal of Interdisciplinary Mathematics

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tjim20

Mathematical model for 3D object reconstruction using OccNet (CNN)

A. Shruthiba & R. Deepu

To cite this article: A. Shruthiba & R. Deepu (2022) Mathematical model for 3D object reconstruction using OccNet (CNN), Journal of Interdisciplinary Mathematics, 25:7, 1961-1970, DOI: 10.1080/09720502.2022.2148360

To link to this article: https://doi.org/10.1080/09720502.2022.2148360

Published online: 08 Dec 2022.

Journal of Interdisciplinary Mathematics
ISSN: 0972-0502 (Print), ISSN: 2169-012X (Online)
Vol. 25 (2022), No. 7, pp. 1961–1970
DOI : 10.1080/09720502.2022.2148360

Mathematical model for 3D object reconstruction using OccNet (CNN)

A. Shruthiba *
Department of Artificial Intelligence and Machine Learning
Bangalore Institute of Technology
Bengaluru
Karnataka
India

R. Deepu †
Department of Computer Science and Engineering
ATME College of Engineering
Mysore
Karnataka
India

Abstract

The input 2D image is used by the encoder to first understand the geometrical restrictions in compressed representation. Second, in the straightforward AI method, the latent representation of the input image is acquired during encoding. On the other hand, the suggested OccNet (CNN) technique computes two encoded vectors of mean and standard deviation during the encoding stage from input. The acquired encoded representation is then transformed into a three-dimensional model via the decoding process. The same decoding process is used by both of the suggested solutions. The reconstruction of a complex 3D object with colourful effects from a single 2D shot may also be the subject of future research. Unlike other methods, our representation doesn't need a lot of memory to encode a description of the 3D output at infinite resolution. We show that our representation effectively encodes three-dimensional structure and can be deduced from a variety of inputs. Our experiments show competitive results for the difficult challenges of 3D reconstruction from single images, noisy point clouds and coarse discrete voxel grids, both qualitatively and numerically.

Subject Classification: *68T07.*

Keywords: *OccNet, 2D, 3D, CNN.*

* *E-mail:* `shruthibaa@bit-bangalore.edu.in (Corresponding Author)`
† *E-mail:* `rdeepusingh21@gmail.com`

1. Introduction

Learned models, unlike typical multi-view stereo algorithms, can encode rich previous information about the space of 3D shapes, which aids in the resolution of ambiguities in the input. The concept of 3D object reconstruction is an age-old problem that has been researched for ages in the field concerned with image enhancement, graphics design and artificial intelligence. The application of 3D object reconstruction began a decade ago using the method of convolution neural networks and has piqued popularity among researchers till now. In this era of speedy development, this field has also been approached in various ways by people working in multiple areas of technological applications. We attempt to present a theoretical graph technique of solving the problems dealing with 3D object reconstruction. Finally, for modeling the surface of this object, we take a weighted graph having two distinct parts that are nothing but bipartite over two contours or boundaries that lie next to each other. We reform the structure model by linking all the parts that we formed from the matching of the bigraphs to obtain a complete prototype.

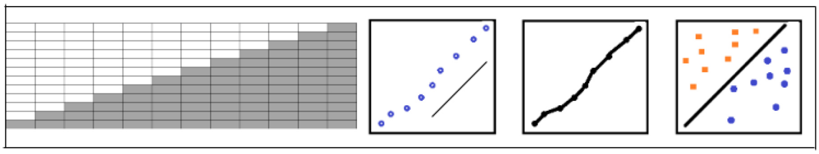

Figure 1

3D representations discretize the output space

2. Related Work with Point Representations and Mesh Representations

It has been investigated in the past to rebuild 3D geometry from a single image using 3D convolutional neural networks that function on voxel grids. Meshes were initially investigated for discriminative 3D segmentation or classification tasks by applying convolutions to the graph covered by the vertices and edges of the mesh. Due to the high memory needs of voxel representations, recent work has favoured reconstructing 3D objects using a multi-resolution technique. Our approach differs from the others in that it generates high-resolution closed surfaces devoid of self-intersections and does not demand the input of template meshes belonging to the same object class. Our approach makes use of deep learning to generate a more expressive representation that is simple to incorporate into a learning pipeline from beginning to end.

3. ML-based Algorithms

As an illustration of 3D geometry, we introduce OccNet. Then, we show how point clouds, single images and low-resolution voxel demos can be used to train a model that infers this demonstration from various inputs. We show how to extract high-quality 3D models from our model while testing as a conclusion. Our key assumption is that by using a neural network to assign each object an occupancy probability between 0 and 1, we can estimate this 3D function. This network is comparable to a neural network for binary classification, with the exception that we are interested in the decision border, which implicitly represents the object's surface. If the marching cubes method were used at the current declaration, these voxels would intersect the mesh. Each active voxel is divided into eight smaller ones and any extra grid points that emerge from this subdivision are added to the occupancy grid and evaluated. We keep repeating these procedures until we reach the desired result.

4. Formation of contours

The editing of each structure's outlines can be split into three parts. This initial step is the normalized digital surface modeling (nDSM), a pixelated portrayal of proportional height objective data above the ground. The distinction between the digital surface modeling (DSM) and the digital elevation model (DEM) is used to calculate it (DEM). The DSM is achieved using the linear triangulated irregular network (TIN) approximation approach from aerial high spatial resolution point clouds. A progressive morphological filter is then used to classify the ground points, which are further incorporated into the DEM. The choice of cell size is also important in the production of DSM and DEM. According to certain investigations, the following formula can be used to determine the size of the cell represented by 's' :

$$s = \sqrt{1} / \rho \tag{1}$$

where ρ is the pulse density of the high spatial resolution point clouds in the unit (returns/meter square).

Next we move towards synthesis of contours. To eliminate signal disturbance, a Gaussian filter for smoothing is used on nDSM prior producing tensor outlines:

Gaussian Smoothing:

$$\text{Gaussian Smoothing}: \quad GS(a,b) = \frac{1}{2\pi\delta^2} e^{\frac{-(a^2+b^2)}{2\delta^2}} \tag{2}$$

where, GS(a,b) is the point where the point of application of the Gaussian smoothing, δ represents the standard deviation of the Gaussian kernel. This facilitates the generation of boundaries with the help of nDSM that has been smoothed.While creating a curve/boundary, two factors—the base contour (c0) and the contour interval (xi)—should be taken into account. The base contour c0 is set to 0 m as the nDSM captures comparative elevation data of surface items. The contour interval di can be chosen to be a little bit greater than the vertical truthfulness of the nDSM produced by LiDAR. Constructing outlines is the last stage. Following the previous phase, the outline has a multitude of outlines, particularly trees and automobiles. We use the data sets of 2D object outlines, which were subjectively assessed from slightly elevated overhead pictures, to remove the quasi outlines. Particular object outlines can be derived and recognized under the direction of the structure blueprints. The succeeding studies then make use of the calculated specific object outlines.

5. Synthesis of Centralized Outline Trees Depending on Graphs

Following the extraction of object contours, objects are shown as contour clusters, where tight outlines at one height are surrounded by additional close outlines at a lower elevation. According to other research, such cluster types may be further described via graph-based local outline trees. Few earlier works also offer thorough information on the conceptual foundations of the limited outline tree theory. Finding and including the adjacent outline B as a child node of contour A is the final stage. This recurrent process is continued up until the highest contour F is used as the terminal node. The sole structure has a tree outline with a single branch at the end. Similar to how an inter-hybrid object's outline tree is created, Figure 2's inter-branched contour tree is produced. Figure 2 serves as an example. The inter outline tree is composed of the parent vertex (A1), six internal vertices (A2, B1, B2, B3, B4 and C1) and two end vertices (B5 and C2).

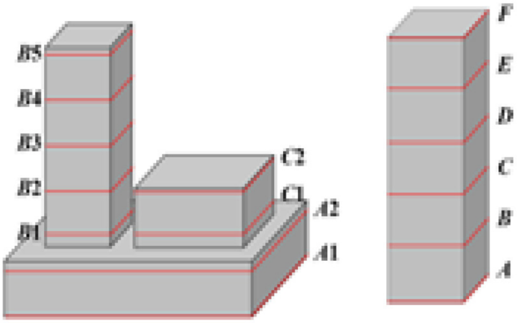

Figure 2

The local outline tree

It is possible to divide an object's structure into collections of smaller components. The dataset of the regional outline tree can be used to find and analyze structures and configurations. A single structural object's object outlines all belong to the same structure and have the same topological relationships. As a result, the single structure is represented by a single outline tree. The one vertex contour tree in Figure 1b shows no topological changes, which implies that the comparable objects in Figure 2 have a single, uncomplicated structure. The matching trees for the intricately structured structures are undergoing fundamental topological alterations. For the inter outline tree shown in Figure 1d, the vertex A2 contains two child vertices, B1 and C1, which together make up a detachment connection in the context of topological modeling. The tortuous inter tree, which has its base at B1, represents a piece of the object. In fact, the tree branch that is fixed at C1 is a structural element. The item also comprises of the inter-trees A1 and A2, which form the base of the building. As a result, A1–A2, B1–B5 and C1–C2 make up the three sections of the structure in Figure 1c. It is clear that the local contour tree accurately reproduces the topographical and geographic layout of objects.

The outline tree design allows for the division of the object into a number of discrete components, each with a standardized and repeating layout. More information is provided on how to create a geometric model for each component.

6. Graph Fitting using Bipartite Graphs

In this work, the outlines are used to recreate the 3D object models. The goal is to combine all of the independent features that are created from the outlines found in an inter-tree to build a comprehensive representation. We take into consideration two successive outlines, C_B and C_U, to make the issue simpler (see Figure 2). By resolving the feature congruence issue between these two outlines, the exterior model is created in this case. The feature relationship issue is then solved using a graph conceptual method called scaled bipartite graph fitting.

The preservation of the semantic relation is not guaranteed by the point similarity of 2-point sets because the number of nodes in outlines C_B and C_U will not be identical. Consequently, contours C_B and C_U are shared equally into n portions, giving rise to 2-point sets B and U that, when combined, form contours C_B and C_U, in both. Throughout further sections, we'll go over further specifics about the choice of an outline interval n. Eventually, we move towards the construction of a bipartite graph K represented by K={B,U,E} having E as the set of edges that are connecting the vertices of surface B and U. The cardinality (number of vertices) of this graph K can be shown as $|K|=|B|+|U|$ where $|.|$ represents cardinality of individual contours B and U. The number of vertices that is cardinality is equal for both the contours hence $|B|=|U|=|n|$. The range in the image pixels determines the weight of the links/edges connecting the vertices in B and U.

Let us consider two vertices in the B and U of form

$$v_i = [a_i, b_i, C_{B0}]^T \tag{3}$$

and

$$v_j = [a_j, b_j, C_{U0}]^T \tag{4}$$

with 1≤i,j≤n respectively, where C_{B0} and C_{U0} represents contour height value in the contours C_B and C_U respectively. The distance value is measure in terms of Euclidean norm $|.||.|_2$:

$$d_{i,j} = \| v_i - v_j \|_2 \tag{5}$$

Applying this value to find the weight $W_{i,j}$ from v_i to v_j as;

$$W_{i,j} = e^{\frac{-d^2}{\varsigma^2}} \tag{6}$$

Here, a shrinkage factor that regulates the spacing of the vertices which is obtained by the technique of trial and error method. The value considered to be 15. The application will result in an output with complete maximum matching value max for given bipartite graph K={B,U,E} and n n cost matrix for edges.The outcome max-weighted bipartite matching measures the groups of similar vertices as well as the matching of optimum cardinality (contours). Large K **values** indicate more similarity, demonstrating how close the contour forms are. We construct the correlation between locations in subsequent outlines using the Kuhn-Munkres method to produce proper modeling of the surface. The aforementioned procedure may produce the contour from any two neighboring outlines in a binary-contour tree. We combine all of the modeling into a hybrid geometric model in the end.

7. Renovation of an Object Model

One-layered models might not be adequate to depict structures since object models can be intricate and sophisticated. The objects are divided into multiple pieces in the method we provide and each piece is rebuilt separately. By following the steps outlined above, the entire object structure may be rebuilt. In order to create a full model, all of the component models must be composited. The connectors between the base model and higher model may not come into contact directly while joining the unique part models but instead have a slight space between them. This results from the contour's defining feature, the contour interval. The contours placed at the intersections may be missed because of the contour interval.

The node matching problem is being addressed by the construction of the bipartite graph with weights calculating its optimum matching. In our scenario the optimum matching return will offer this subjective (one to one) mappings from B through U if the is maximized. Let us consider the graph matching for the graph K= {B, U, E} be represented χ_K and χ_{asbe} all possible matching index for bipartite graphs. The maximum matching value χmax that represents the matching B to U in one to one manner that is one vertex of set B is linked to only vertex of set U. Next, we solve the maximum matching problem by a squared matrix minimization or maximization technique for generic allocation issues given Kuhn-Munkers **algorithm. The methodology for application of this algorithm is stated below:**

Create an n x n square matrix that is the cost of matrix for the edges with (i, j) = W_{ij}

Next generation of permutation (Qt; t=0, 1,…, n-1) of integers 0, 1, …, n-1 for the maximization of the $\sum_{t=0}^{n-1} M[t, Q_t]$

Representation of maximum target function of the weighted graph matching for Kis: $\chi_{\max} = \operatorname{argmax}_{\chi_K} \sum_{1\le i,j\le n} (i,j) M_{i,j}$

The application will result in an output with complete maximum matching value $\chi_{\max}$ **for given bipartite graph K= {B, U, E} and n** x n cost matrix for edges.

5. Experimental Result

In our first research, we test how well possession networks characterize 3D geometry, regardless of the faults in the input encoding. In this experiment, we check how well our network can display in visuals the IoU of a voxelization relative to the ground truth mesh and the number of parameters per model required for the two representations. Watch as our model outperforms a low-resolution voxel representation in terms of IoU compared to the ground truth mesh. While occupancy networks have a parameter count that is independent of resolution, voxel representations have a parameter count that grows cubically with resolution. Figure 3 shows Graphical representation of IoU vs. Resolution.

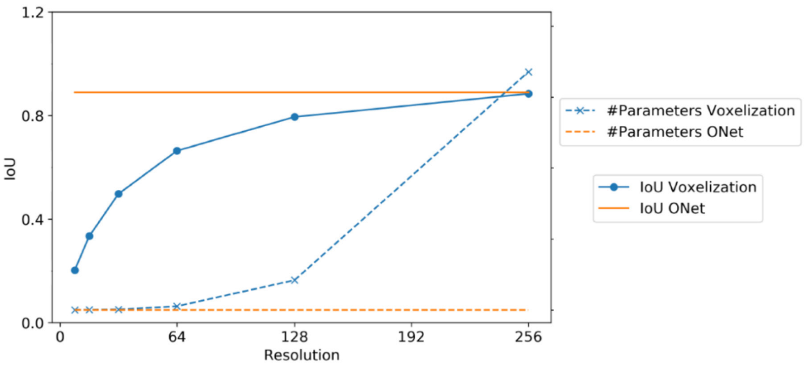

Figure 3

Graphical representation of IoU vs. Resolution

8. Conclusions

The input 2D image is used by the encoder to first understand the geometrical restrictions in compressed representation. Second, in the straightforward AI method, the latent representation of the input image is acquired during encoding. On the other hand, the suggested OccNet (CNN) technique computes two encoded vectors of mean and standard deviation during the encoding stage from input. The acquired encoded representation is then transformed into a three-dimensional model using the decoding technique. The same decoding process is used by both of the suggested solutions. The reconstruction of a complex 3D object with colourful effects from a single 2D shot may also be the subject of future research. Our representation, in contrast to earlier methods, does so without consuming a lot of memory and while encoding a explanation of the 3D productivity at endless pledge. We show that our representation effectively encodes three-dimensional construction and can be deduced from a variety of inputs.

References

[1] Z. Wu, T. Zhao and C. Nguyen, "3D Reconstruction and Object Detection for HoloLens," 2020 Digital Image Computing: Techniques and Applications (DICTA), 2020, pp. 1-2, doi: 10.1109/DICTA51227.2020.9363378.

[2] H. Chen, W. Chen and T. Gao, "Ground 3D Object Reconstruction Based on Multi-View 3D Occupancy Network using Satellite Remote Sensing Image," 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, 2021, pp. 4826-4829, doi: 10.1109/IGARSS47720.2021.9554612.

[3] B. Li, Y. Zhang, B. Zhao and H. Shao, "3D-ReConstnet: A Single-View 3D-Object Point Cloud Reconstruction Network," in IEEE Access, vol. 8, pp. 83782-83790, 2020, doi: 10.1109/ACCESS.2020.2992554.

[4] C. Sui, K. He, C. Lyu and Y. -H. Liu, "Accurate 3D Reconstruction of Dynamic Objects by Spatial-Temporal Multiplexing and Motion-Induced Error Elimination," in IEEE Transactions on Image Processing, vol. 31, pp. 2106-2121, 2022, doi: 10.1109/TIP.2022.3150297.

[5] X. Li, K. Ping, X. Gu and M. He, "3D Shape Reconstruction of Furniture Object from a Single Real Indoor Image," 2020 17th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP), 2020, pp. 101-104, doi: 10.1109/ICCWAMTIP51612.2020.9317479.

[6] K. Patoommakesorn, F. Vignat and F. Villeneuve, "The 3D Edge Reconstruction from 2D Image by Using Correlation Based Algorithm," 2019 IEEE 6th International Conference on Industrial Engineering and Applications (ICIEA), 2019, pp. 372-376, doi: 10.1109/IEA.2019.8714965.

[7] Z. Zhou, Q. Lai, S. Ding and S. Liu, “Joint 2D Object Detection and 3D Reconstruction via Adversarial Fusion Mesh R-CNN,” 2021 IEEE International Symposium on Circuits and Systems (ISCAS), 2021, pp. 1-5, doi: 10.1109/ISCAS51556.2021.9401668.

[8] E. Ouyang, J. Wu, B. Li, L. Zhao and W. Hu, “Band Regrouping and Response-Level Fusion for End-to-End Hyperspectral Object Tracking,” in IEEE Geoscience and Remote Sensing Letters, vol. 19, pp. 1-5, 2022, Art no. 6005805, doi: 10.1109/LGRS.2021.3137606.

[9] J. Xue, Y. Tu and D. Sun, “3D Reconstruction and Understanding of Indoor Scene Based on Single Image,” 2020 International Conference on Virtual Reality and Visualization (ICVRV), 2020, pp. 292-293, doi: 10.1109/ICVRV51359.2020.00072.

[10] E. Tunçer and Ş. Gümüştekin, “3D Object Reconstruction Using Sequentially Activated Multiple Depth Cameras,” 2020 28th Signal Processing and Communications Applications Conference (SIU), 2020, pp. 1-4, doi: 10.1109/SIU49456.2020.9302191.