ESANN 2019 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence
and Machine Learning. Bruges (Belgium), 24-26 April 2019, i6doc.com publ., ISBN 978-287-587-065-0.
Available from http://www.i6doc.com/en/.

Deep hybrid approach for 3D plane segmentation

Felipe Gomez Marulanda, Pieter Libin, Timothy Verstraeten and Ann Nowé

Artificial Intelligence Lab - Vrije Universiteit Brussel - Brussels, Belgium

Abstract. We address the limitations of Deep learning models for 3D geometry segmentation by using Conditional Random fields (CRF). We show that CRFs can take advantage of the neighbouring structure of point clouds to assist the learning of the Deep Learning models (DL). Our hybrid PN-CRF model is able to learn more optimal weights by taking advantage of equal-segmentation assignments to neighbouring points. As a result, it increases the robustness in the model specially for segmentation tasks where correctly detecting the boundaries between segmentations is very important.

1 Introduction

The ability to learn directly from unordered data such as 3D geometries remains a challenge specifically for classification and segmentation. Most techniques transform 3D geometries into ordered representations so machine learning algorithms such as Deep Learning (DL) can operate on it. This is typically achieved by summarising the 3D shapes into geometrical features (i.e., characteristics). However, performing these transformations may induce information loss which leads to a significant decrease with respect to the learning accuracy. Ideally, we would like to circumvent these transformations and directly learn on 3D geometrical spaces. In this paper we will describe how to accurately learn to segment 3D geometries without manually transforming the input space.

Our research improves upon three prior studies [1, 2, 3] that introduced the concept of symmetric function approximation to achieve order equivariance and invariance to geometrical transformations (i.e., rotation) in 3D point spaces. They designed a DL architecture that learns to approximate symmetric functions that allow to find optimal cuts in feature space. In this work, we show that the architecture of [2] called PointNet dismisses the spatial information that exist between points, leading it to make wrong predictions between the boundaries labels or outliers within a segmentation class. Having smooth and accurate predictions at the boundaries of label segments is very important for different fields of research. For example, segmentation on point clouds generated by sensors in self driving cars, requires an accurate boundary between the pavement and the road. To overcome the PointNet limitations, we introduce a technique based on Conditional Random Fields (CRF). CRFs have been proven to be very successful in the field of image segmentation and classification. We found this technique increases the robustness of the model and reduces noise at the boundaries between label segments. Furthermore, it speeds up the learning process on geometries that have complex boundary agreements.

2 Related Work

The majority of methods that deal with unordered 3D models such a point clouds accomplish automatic segmentation by transforming the original unordered 3D format onto an ordered feature space which is then used as input to segmentation methods or machine learning algorithms. For example, the concept of shape distributions is introduce in [4], where a shape signature of a geometry is generated by a probability distribution sampled from a shape function that measures the geometrical properties of the 3D model [4]. Similarly, local diameter functions introduced in [5] measures the diameter of a 3D shape in the neighbourhood of each vertex and then constructs a histogram from this function as a signature of the 3D shape. A different approach was shown in [6] which rendered a collection of images taken from different view-angles of the 3D space and used it in an ensemble of Convolutional Neural Networks (CNN) to extract higher order features from the 3D geometries. Equivalently, in [7] a voxelised representation was created from the 3D geometry to extract higher order features using a CNN. As these techniques transform the data, important information that is useful to perform high quality segmentation is lost in the process.

3 Methods

3.1 3D point cloud learning

Point clouds are the one of the most basic forms of representing a 3D object. Formally, a point cloud $\mathcal{P}$ is an unordered collection of points $\{\vec{p_i}\}_{i=1}^{N}$ in Euclidean space $\mathbb{R}^3$. This collection of points can be obtained from 3D scanners or by sampling continuous surfaces. In [2], direct learning from 3d point clouds was first introduced. They used a Deep learning algorithm called PointNet to first approximate symmetric functions to achieve order equivariance and to extract higher order features from the transformed point cloud. This results into a point cloud which is invariant to geometric transformations such as rotation, translation and other rigid transformations. The first part of PointNet approximates the symmetric function and transformation function in $\mathbb{R}^{3\times 3}$ which is then used to transform the input space. The second part repeats the same operations as in the first part of the architecture, however, the transformation matrix $R^{64\times 64}$ is performed on the feature space. The last part of the PointNet consists of extracting the global signature of the input space and aggregating it with local features to compute the likelihood that a point p_i belongs to segmentation label l_i. As the second transformations matrix have a higher number of dimensions $R^{64\times 64}$ compared to the input transformations $\mathbb{R}^{3\times 3}$, the feature transformation matrix is constrained to be close to an orthogonal matrix allowing the preservation of its symmetric inner product. To achieve this, the cost function is regularised by the following equation:

$$L_{reg} = \left\| I - AA^T \right\|_F^2 \tag{1}$$

where A is the approximated feature transformation matrix, I is the identity matrix, and $||\cdot||_F$ is the Frobenius norm. Adding the L_{reg} regularisation term to the cost function renders the optimisation more stable. PointNet accomplishes on average a accuracy of 86% per segmentation class on the ShapeNet benchmark dataset [8]. During our visual analysis of the PointNet predictions, we found that for most well segmented geometries several prediction inconsistencies can be observed at the edges of segment boundaries and at individual segments (i.e., outliers within the segment). We hypothesise that this is due to two reasons which we will confirm in section 4. Firstly, the architecture in [2] does not encode the label agreement between boundary segments and the discretisation density of point clouds may be insufficient to correctly extract the boundaries between segmentations.

To overcome PointNet limitations highlighted in this section, we propose to use a technique called Conditional Random Field (CRF) to improve the learning process by backpropagating the label agreement that exist between neighbouring points. We will show that introducing information about the inter-features that exist between points will help the Neural Network to come up with better kernels that are able to improve feature space partitioning.

3.2 Conditional Random Fields for 3D point learning

A Conditional Random Field (CRF) is a undirected probabilistic graphical (UGM) model that belongs to the group of generative models. UGMs define the joint distribution of a set of random variables over the structure of an undirected graph. Besides describing a probability distribution of segmentation labels for each point, it also models neighbouring information that favours equality of labels amongst spatial proximal points. Formally, let $\mathcal{Z}$ be defined over a set of random variables $\{z_i\}_{i=1}^N$ where each variable can take any value from a predefined set of labels $\mathcal{L} = \{l_i\}_{i=1}^K$, and a set of observations $\mathcal{P} = \{\vec{p}_i\}_{i=1}^N$ which represent the points coordinates in a point cloud. The CRF is structured as an undirected graph $G = (\mathcal{V}, \mathcal{E})$ where the vertices $\mathcal{V}$ index the pair of variables $(z_i, \vec{p}_i)$, and the edges $\mathcal{E}$ correspond to the dependencies between the neighbouring variables (z_i, z_j). According to the Hammersley-Clifford's theorem, [9] the joint distribution $\mathbb{P}(z,p)$ can be directly modelled by its conditional $\mathbb{P}(z|p)$ which can be factorised over a product neighbouring factors. Conditioning on the neighbours of a variable makes it independent from the rest of the other variables in the random field. Such a conditional probability distribution is often referred to as a Gibbs measure.

The simplest model of a CRF is the pairwise-CRF which express the total energy $E(\mathcal{Z}, \mathcal{P})$ as a sum of unary and pairwise factors.

$$E(\mathcal{Z}, \mathcal{P}) = \sum_{i \in \mathcal{V}} \psi(z_i, \vec{p}_i) + \sum_{(i,j) \in \mathcal{E}} \psi(z_i, z_j, \vec{p}_i, \vec{p}_j), \tag{2}$$

where the *unary* factor $\psi(z_i, \vec{p}_i)$ provides the energy cost of assigning a label to a point and the *pairwise* factor $\psi(z_i, z_j, \vec{p}_i, \vec{p}_j)$ provides the energy cost of

ESANN 2019 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2019, i6doc.com publ., ISBN 978-287-587-065-0. Available from http://www.i6doc.com/en/.

assigning a labels to a variable pair. In Neural Networks, we can use the cross-entropy cost along with a logistic regression (i.e., softmax or sigmoid functions) to compute the unary term of this equation. However, the pairwise factors are computed separately over the entrired point cloud and added to the CNN architecture in order to match the full energy cost of the CRF. The pairwise factor $\psi(z_i, z_j, \vec{p}_i, \vec{p}_j)$ we used in our experiments is as follows:

$$\psi(z_i, z_j, \vec{p}_i, \vec{p}_j) = \theta \; \delta(z_i, z_j) \exp(-||\vec{p}_i - \vec{p}_j||^2) \tag{3}$$

where $\delta(z_i, z_j)$ is a Kronecker delta function that provides the cost of assigning equal or different labels to a pair of random variables. This means that is only when z_i and z_j are different is when the model gets penalised. The second factor is a similarity cost that modulates the penalty by favouring the equal-label assignments to spatial proximal points. Incorporating this measure into the loss function influences the cost such that when the Euclidean distance between two points is small it becomes likely that their segmentation labels are the same. Note that the pairwise factor is bounded by θ which restricts it to take control over the full loss.

More complex global co-occurrence factors $\delta(z_i, z_j)$ can be derived from domain knowledge of the family of 3D geometries. For example, in some families of aircraft, there is a low probability that a point labelled as an engine is next to a point labelled as a fuselage. In this case, $\delta(z_i, z_j)$ will be close to zero, rendering the entire equation close to zero even though these two points may be spatially alongside each other. If such domain knowledge is unavailable, δ can be approximate by computing the co-occurrence frequencies of the labels found in the data and by adding a decay during training, such that it will not bias the learning process in the limit.

To introduce CRF into the Deep learning model, we assume that the distances between neighbouring points can be pre-computed before the training phase. However, it is not a requirement during the prediction phase, as this measures are only used to guide the learning of the network and CRF.

4 Experiments and Results

Our PointNet-CRF (PN-CRF) model was trained on the geometries of the annotated version of the *ShapeNet* dataset [10, 8] which contains $16,881$ shapes from 16 categories. The ground truth annotations are labelled on the point sampled from the original shapes. These datasets where partitioned into training, test and validation sets to evaluate the generalisation accuracy of PN-CRF. Both PointNet and PN-CRF were setup with the same hyperparameters as in [2].Furthermore, Table 1 illustrates the IoU (%) accuracy after running the models on the annotated-ShapeNet dataset. In this table we observe that on average the both models are very similar in performance. However, when we visually inspected the boundaries for most geometries in the ShapeNet, We saw that most of the boundary segmentation problems in PointNet disappeared. An example, of these validations are shown in Figure 1. These prediction differences are not emphasised in Table 1 as the corrections made by PN-CRF are to small relative to the overall point cloud.

Table 1: PN-CRF validation results for the ShapeNet dataset . The metric used is the IoU(%) on the points.

	Aircraft	**Bag**	**Cap**	**Car**	**Chair**	**Earphone**	**Guitar**	**Knife**	**Lamp**	**Laptop**	**Motor**	**Mug**	**Pistol**	**Rocket**	**Skateboard**	**Table**
# shapes	2690	76	55	898	3758	69	787	392	1547	451	202	184	283	66	152	5271
PointNet-[2]	72.0	75.5	**90.0**	75.3	71.8	81.7	**91.2**	**87.5**	-	82.8	**63.9**	88.5	**86.9**	48.1	**90.7**	84.7
Ours	**72.8**	**80.4**	83.5	**77.3**	**72.0**	**82.4**	91.0	83.5	-	**87.7**	56.9	**90.0**	82.3	**51.1**	89.7	**84.9**

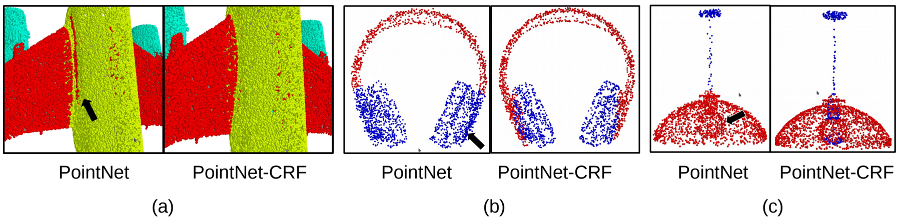

Fig. 1: Visual Comparison between vanilla PointNet and PN-CRF. These images shows how the CRF managed to correct the boundary issues between neighbouring surfaces.

4.1 Conclusion

We proposed a 3D Deep learning algorithm that uses Conditional Random Fields to influence the learning by favouring equal-segmentation assignments to neighbouring points. We show that PN-CRF adjusted the inconsistencies found at the boundaries and decreased noise on single segmentations. Furthermore, we analysed the reason of the accuracy loss in the Motor dataset after the CRF. We found that this was due to wheel fenders. As the CRF enforces equal label assignments at this location, it makes the points of the wheel to become a fender label. In contrast, PointNet labels all these points as a wheel which is also incorrect. The ground truth is a mixture of both where the wheel is more abundant. We hypothesise that this problem arises because we are computing the CRF over the whole point cloud and not over a embedding space as done in [11]. Consequently, this can be potentially corrected by learning the weights of the DL model and the CRF together. This will be analysed in the near future.

Acknowledgments

Felipe Gomez Marulanda's work was supported by Doctiris-innoviris brussels grant and Pieter Libin and Timothy Verstraeten were supported by a PhD grant of the FWO (Fonds Wetenschappelijk Onderzoek-Vlaanderen)

References

[1] Sander Dieleman, Jeffrey De Fauw, and Koray Kavukcuoglu. Exploiting cyclic symmetry in convolutional neural networks. ICML'16, pages 1889–1898. JMLR.org, 2016.

[2] Charles R Qi, Hao Su, Kaichun Mo, and Leonidas J Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. *Proc. Computer Vision and Pattern Recognition (CVPR), IEEE*, 1(2):4, 2017.

[3] Felipe Gomez Marulanda, Pieter Libin, Timothy Verstraeten, and Ann Nowé. Ipc-net: 3d point-cloud segmentation using deep inter-point convolutional layers. In *IEEE Tools with Artificial Intelligence (ICTAI)*, pages 293–301. IEEE, 2018.

[4] Robert Osada, Thomas Funkhouser, Bernard Chazelle, and David Dobkin. Shape distributions. *ACM Transactions on Graphics (TOG)*, 21(4):807–832, 2002.

[5] Ran Gal, Ariel Shamir, and Daniel Cohen-Or. Pose-oblivious shape signature. *IEEE computer graphics*, 13(2):261–271, 2007.

[6] Hang Su, Subhransu Maji, Evangelos Kalogerakis, and Erik Learned-Miller. Multi-view convolutional neural networks for 3d shape recognition. In *IEEE computer vision*, pages 945–953, 2015.

[7] André Brock, Theodore Lim, James M. Ritchie, and Nick Weston. Generative and discriminative voxel modeling with convolutional neural networks. *CoRR*, abs/1608.04236, 2016.

[8] Angel X Chang, Thomas Funkhouser, Leonidas Guibas, Pat Hanrahan, Qixing Huang, Zimo Li, Silvio Savarese, Manolis Savva, Shuran Song, Hao Su, et al. Shapenet: An information-rich 3d model repository. *arXiv preprint arXiv:1512.03012*, 2015.

[9] John M Hammersley and Peter Clifford. Markov fields on finite graphs and lattices. 1971.

[10] Li Yi, Vladimir G Kim, Duygu Ceylan, I Shen, Mengyan Yan, Hao Su, Cewu Lu, Qixing Huang, Alla Sheffer, Leonidas Guibas, et al. A scalable active framework for region annotation in 3d shape collections. *ACM Transactions on Graphics (TOG)*, 35(6):210, 2016.

[11] 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. *IEEE*, 40(4):834–848, 2018.