

**VG-VAE: A Venatus Geometry Point-Cloud Variational Auto-Encoder**

Tejas Anvekar, Ramesh Ashok Tabib, Dikshit Hegde, Uma Mudengudi  
Center of Excellence in Visual Intelligence (CEVI), KLE Technological University.  
Vidya Nagar, Hubballi, Karnakata, India.  
anvekartejas@gmail.com, ramesh_t@kletech.ac.in, dikshithegde@gmail.com, uma@kletech.ac.in

---

**Abstract**

In this paper, we propose VG-VAE: Venatus Geometric Variational Auto-Encoder for capturing unsupervised hierarchical local and global geometric signatures in pointcloud. Recent research emphasises the significance of the underlying intrinsic geometry for pointcloud processing. Our contribution is to extract and analyse the morphology of the pointcloud using the proposed Geometric Proximity Correlator (GPC) and variational sampling of the latent. The extraction of local geometric signatures is facilitated by the GPC, whereas the extraction of global geometry is facilitated by variational sampling. Furthermore, we apply a naive mix of vector algebra and 3D geometry to extract the basic per-point geometric signature, which assists the unsupervised hypothesis. We provide statistical analyses of local and global geometric signatures. The impacts of our geometric features are demonstrated on pointcloud classification as downstream task using the classic pointcloud feature extractor PointNet. We demonstrate our analysis on ModelNet40 a benchmark dataset, and compare with state-of-the-art techniques.

---

**1. Introduction**

In recent years, 3D pointcloud has grown in relevance in a variety of fields, including 3D printing, Metaverse, and self-driving automobiles as sensors capture depth in addition with other visual signals. Pointclouds are beginning to play a crucial role in many real-world applications, such as simultaneous localization and mapping (SLAM) [14] [5], 3D object detection, digitization of heritage sites towards presentation in AR/VR/XR/MR. Towards presentation of 3D data in digital space, there is a need for efficient methods to analyze, process and derive huge volume of 3D pointclouds [15] [20] [25] [22] [23]. In contrast to human vision, supervising a machine to infer geometric information
is challenging. Representing 3D object as a collection of intrinsic structure provides morphology as a geometric signature. This abbreviated representation of a 3D object is critical for improving shape understanding, processing, and analysis. To address the lack of geometric information in 3D pointclouds, various studies attempt to decompose a 3D pointcloud into meaningful structures, and infer a topological graph by modelling correlations between these parts \[16\] \[9\] \[10\] as geometric signatures. However, the decomposition is completely perceptual. While, some approaches attempt to recover a parametric form of the input pointclouds using defined geometric shapes \[12\] \[24\] \[9\] \[10\]. Authors in \[10\] \[9\] discuss about deriving fundamental geometric signatures and use them as a plugin towards downstream tasks. These techniques find challenges generalizing across datasets and hybrid geometries. Towards this, we propose self-supervised approach to eliminate the need for generalization across datasets, and generate redesigned hybrid geometric signatures for a chosen dataset unlike \[10\] \[9\].

The main contributions of the work are:

- We propose a self-supervised / unsupervised technique to derive geometric features of pointcloud and name it as VG-VAE: Venatus Geometry Point-Cloud Variational Auto-Encoder.
  - A novel Geometric Proximity Correlator (GPC) to co-relate intrinsic and empirical geometric signatures of a pointcloud.
  - An Intrinsic Geometric Interpreter (IGI) to derive the features of pointcloud in existing and higher dimensional space.
  - A KNN prior based geometric feature aware encoder.

- We propose to consider variational sampling to generalize over extrinsic geometric features across the samples.

- We model a geometry aware pointcloud classifier using the derived geometric features as a plugin towards classification of pointclouds.

- We demonstrate the effect of derived geometric signatures on benchmark dataset (ModelNet40) and compare the results with state-of-the-art techniques.

In Section 2 we discuss the proposed VG-VAE: Venatus Geometry Point-Cloud Variational Auto-Encoder. We discuss the ablation and results of proposed model and compare with state-of-the-art techniques in Section 3, and conclude in Section 4.
2. Venatus Geometry Point-Cloud Variational Auto-Encoder (VG-VAE)

We model VG-VAE, Venatus Geometry Point-Cloud Variational Auto-Encoder for learning unsupervised hierarchical geometric features in pointcloud data. Geometric aspects of a pointcloud characterise its morphology which facilitates abstraction, generation, and description of the pointcloud as shown in Figure 2. Towards this, we propose a Geometric Proximity Correlator (GPC) for extracting hierarchical local and global geometric signatures. Further we estimate the underlying per-point variations in higher dimensional space with nearest neighbour as a prior using Variational sampling.

### 2.1. Geometric Proximity Correlator (GPC)

Geometric Proximity Correlator facilitates to capture the geometric signatures. Geometric signatures are computed via a seven step process as shown in Algorithm 1. Initially, we search for the prior i.e., indices \(idx_{knn}\) using K-Nearest neighbours. Intrinsic Geometry Interpreter \(\psi\) is estimated on the given pointcloud \(pc\). Towards extracting the empirical (learnt) features \(PC_G\), we emphasise on weight shared mlp that projects \(\psi\) to a higher dimensional space. The geometric posterior \(X\) is captured by gathering the nearest neighbour \(gathered_{neighbors}\) in higher dimensional space aided by prior \(idx_{knn}\) as shown in Figure 3. To boost the learning, we compute the statistical features (Mean, Standard deviation, Min, and Max) of \(gathered_{neighbors}\) with respect to its K nearest neighbours and project them to higher dimensional space using a weight shared mlp. The proposed geometric signatures \(G\) are simply obtained by computing clusters \(G\). The idiocentric clusters are inspired from Attention Based Decomposition network [9] and PointDCCNet [10].

#### Algorithm 1: Geometric Proximity Correlator

<table>
<thead>
<tr>
<th>Input: Point Cloud (PC); // (B, N, C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Output: Geometric Proximity Correlator’s weights (f_\theta); Geometry aware pointcloud features (X); Geometric Signatures (G) and Geometric Cluster labels (G).</td>
</tr>
</tbody>
</table>

1. Initialize \(K\), \(k\), \(C_{in}\), \(C_{geo}\), and \(C_{out}\) according to Section 2.
2. \(idx_{knn} = \text{Knn index}(PC, K)\). /* \(B, N, C\) */
3. \(\psi = \text{intrinsic geometry interpreter}(PC)\). /* \(B, N, C_{in}\) */
4. \(PC_G = \text{shared mlp}(C_{in} + C_1, C_{geo})(\psi)\). /* \(B, N, C_{geo}\) */
5. \(gathered_{neighbors} = \text{gather operation}(idx_{knn}, PC_G)\). /* \(B, N, C_{geo}, K\) */
6. \(\text{min, max, std, and mean} \leftarrow gathered_{neighbors}\). /* \(B, N, C_{out}\) */
7. \(X \leftarrow \text{shared mlp}(C_{geo}, C_{out})(\text{Concat}(\text{min, max, std, and mean}))\). /* \(B, N, C_{out}\) */
8. \(G = \text{mean}(PC_G, \text{dim=1})\). /* \(B\) */
9. \(G = \text{KMeans}(k, G)\). /* \(B\) */

#### 2.1.1 Intrinsic Geometry Interpreter (IGI)

The unsupervised hypothesis of our network is notably aided by joint learning, variational sampling of latent and
a basic Intrinsic Geometry Interpreter $\psi_i$. The variational sampling of latent guarantees that the global signatures of the pointcloud is captured, whilst the geometric interpreter enables encapsulating the local signatures. The Intrinsic Geometry Interpreter $\psi_i$ is a naive combination of vector algebra and 3D geometric as shown in Figure 3. Pointclouds are widely used to produce meshes by triangulation. By forming a triangle with respect to two nearest neighbors $(pc_{j1}, pc_{j2})$ of pointcloud $pc_i$, we estimate Intrinsic Geometry Interpreter $\psi_i$ by:

$$\psi_i = \begin{cases} \frac{pc_i = x, y, z; \quad pc_i \in \mathbb{R}^n,}{|v_1| = \ell_2(v_1); \quad |v_1| \in \mathbb{R}^3,} \\
\frac{|v_2| = \ell_2(v_2); \quad |v_2| \in \mathbb{R}^3,}{\hat{n} = \frac{v_1 \times v_2}{|v_1 \times v_2|}; \quad \hat{n} \in \mathbb{R}^3.} \end{cases}$$ (1)

where $(\hat{v}_1, \hat{v}_2)$ represents edge $(1, 2)$ respectively of given point $pc_i$, $(|v_1|, |v_2|)$ represents edge lengths and $\hat{n}$ represent normals of pointcloud $pc_i$. Unlike Geometric Backpropagation Net [19], we propose utilizing $\psi_i$ beyond three dimensions, with the exception of excluding $\hat{n}$, as the cross-product has the orthogonality characteristic only in three and seven dimensional spaces. The set of per-point geometry aids in the robustness of our hypothesis, since normals and edge length would immediately reveal outliers in the supplied point set.

Towards interpretability and hierarchical learning of local to global geometric signatures, we propose to utilize Geometric Proximity Correlator in a dolly chain fashion multiple times. The intermediate $G_i$’s and $G_i$’s correspond to local to global geometric signatures.

### 2.2. Variational Sampling

The hierarchical global symmetry $\mathcal{S}$ is captured by pooling the global geometric features $X$ through a weight shared mlp inspired by PointNet [17]. The variational sample space $z$ captures the extrinsic geometry of the pointcloud space through reparametrization technique and ELBO loss [11].

Variational Autoencoders (VAE) are generative models capable of learning approximated data distributions via variational inference. We consider the stochastic latent space $z$ and optimize the upper-bound on the negative log-likelihood of $x$:

$$\mathbb{E}_{x \sim p_d(x)}[- \log p(x)] < \mathbb{E}_x[\mathbb{E}_{z \sim q(z|x)}] + \mathbb{E}_x[D]$$

where $D = KL(q(z|x)||p(z))$.

Towards tuning of geometric features by optimizing the weights of encoder $f_\theta$ and decoder $g_\theta$ as shown in Figure 2, we propose to utilize Chamfer Distance as reconstruction loss given by:

$$CD(S_1, S_2) = \sum_{x \in S_1} \min_{y \in S_2} ||x-y||^2 + \sum_{y \in S_2} \min_{x \in S_1} ||x-y||^2 (2)$$

and KL Divergence as a regularizer.

$$KL(p, q) = \log \frac{\sigma}{\sigma_c} + \frac{\sigma^2 + (\mu_c - \mu)^2}{2\sigma^2} - \frac{1}{2} (3)$$

where $(\mu, \sigma)$ are the expected mean and standard deviation as shown in Figure 2. $(\mu_c, \sigma_c)$ are expected mean and standard deviation for a given pointcloud.

The proposed decoder is feature extrapolation net comprised of mlp’s similar to an image based VAE [11].

### 2.3. Pointcloud Classification

We incorporate the produced unsupervised geometric signatures $\mathcal{G}$ to the input pointcloud and feed it through the PointNet architecture to demonstrate the impact of our proposed methodology. We demonstrate our analysis of the impact made by hierarchical local and global geometric signatures on ModelNet40 [30] a benchmark dataset and compare them with state-of-the-art techniques.

### 3. Results and Discussions

In this section, we discuss about the dataset used, evaluation metrics and comparison of our methodology with state-of-the-art methods.
3.1. Datasets

To evaluate the performance of the proposed methodology we use benchmark ModelNet40 dataset and compare the classification accuracy with state-of-the-art methods PointNet [17], PointNet++ [18], KCNet [6], MRT-Net [3], SpecGCN [27], DGCCN [28], PCNN [1], ABDNet [9], PointConv [29], PointDCCNet [10], PointTransformer [32], RSCNN [13], PCT [4]. We compare our geometric results with ABDNet [9].

- **ModelNet 40 [30]**: dataset consists of CAD models belonging to 40 categories. These CAD models are sampled to 2048 points to form a pointcloud.

- **ShapeNet [2]**: ShapeNet is a single clean 3D models and manually verified category and alignment annotations. It covers 55 common object categories with about 51,300 unique 3D models.

3.2. Experimentation Details

In this section, we discuss the experimental setup as two major parts.

- **Architectural Details**: The details for architecture of VG-VAE, GPC, and variational sampling decoder are shown in Table 1. To make it more understandable table is colour coded with respect Figure 2. The architecture of GPC is elaborated in Algorithm 1.

- **Training Setup VG-VAE**: Initially during training the weights of VG-VAE’s Encoder $f_θ$ and Decoder $g_θ$ set to uniform distribution and expected mean $μ_e$ and standard deviation $σ_e$ from Equation 3, is set to (0, 0.02) respectively. The weights $θ$ of VG-VAE are tuned using proposed variational loss given in Equation 3 and chamfer pseudo distance as reconstruction loss given in Equation 1. The variational sampling is explained in Section 2.2. We train VG-VAE for 2000 epochs until KL divergence has reached to minima. The intermediate geometric signatures $G_i$ are saved for computing geometric labels $G_i$ using KMeans clustering. We find number of cluster $K = 4$ yields better results visually and also facilitates pointcloud classification.

- **Training Setup VG-VAE + PointNet**: We train a geometry aware 3D classifier by merging geometric features of VG-VAE with pointcloud, we concatenate the pointcloud and geometric feature and train PointNet, the hyperparameters for classification are set as per PointNet [17] with an exception of replacing in channels of 1st layer to dimensional of geometric features and pointcloud together. During training VG-VAE’s weights are freezed so that the discriminative loss wont affect extracted geometric features.

<table>
<thead>
<tr>
<th>Methods</th>
<th>Input</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>PointNet <a href="2017">17</a></td>
<td>xyz</td>
<td>89.2</td>
</tr>
<tr>
<td>PointNet++ <a href="2017">18</a></td>
<td>xyz</td>
<td>90.7</td>
</tr>
<tr>
<td>MRTNet <a href="2018">3</a></td>
<td>xyz</td>
<td>91.2</td>
</tr>
<tr>
<td>Spec-GCN <a href="2018">27</a></td>
<td>xyz</td>
<td>91.5</td>
</tr>
<tr>
<td>Spec-GCN <a href="2018">27</a></td>
<td>xyz, nor</td>
<td>91.8</td>
</tr>
<tr>
<td>PCNN <a href="2018">1</a></td>
<td>xyz</td>
<td>92.3</td>
</tr>
<tr>
<td>PointConv <a href="2018">29</a></td>
<td>xyz, nor</td>
<td>92.5</td>
</tr>
<tr>
<td>DGCCNN <a href="2019">28</a></td>
<td>xyz</td>
<td>92.2</td>
</tr>
<tr>
<td>RSCNN <a href="2019">13</a></td>
<td>xyz</td>
<td><strong>92.9</strong></td>
</tr>
<tr>
<td>3D-GCNN <a href="2019">26</a></td>
<td>xyz</td>
<td>83.5</td>
</tr>
<tr>
<td>Point Transformer <a href="2020">32</a></td>
<td>xyz</td>
<td><strong>93.7</strong></td>
</tr>
<tr>
<td>Point Transformer <a href="2020">32</a></td>
<td>xyz, nor</td>
<td>92.8</td>
</tr>
<tr>
<td>KCNet <a href="2021">6</a></td>
<td>xyz</td>
<td>91.0</td>
</tr>
<tr>
<td>PointDCCNet <a href="2021">10</a></td>
<td>xyz, nor</td>
<td>92.5</td>
</tr>
<tr>
<td>ABD-Net <a href="2021">9</a></td>
<td>xyz</td>
<td>92.2</td>
</tr>
<tr>
<td>ABD-Net <a href="2021">9</a></td>
<td>xyz, nor</td>
<td>92.8</td>
</tr>
<tr>
<td>PCT <a href="2021">4</a></td>
<td>xyz</td>
<td>93.2</td>
</tr>
<tr>
<td><strong>Ours VG-VAE + PointNet</strong></td>
<td>xyz</td>
<td><strong>92.9</strong></td>
</tr>
<tr>
<td><strong>Ours VG-VAE + Linear SVM</strong></td>
<td>xyz</td>
<td>84.5</td>
</tr>
</tbody>
</table>

3.3. Results

In this section, we analyse the inference of our experiments. Initially we discuss the implications of unsupervised hierarchical geometrical signatures as shown in Figure 1. We can infer that the Geometrical Signatures $G$ of VG-VAE change due to unsupervised hypothesis. Towards this, we propose to the cluster the Geometrical signatures $G$ to attain inferable Geometrical Labels $G$ as shown in highlighted regions in Figure 1. The Geometrical labels are the obtained from KMeans [7], where cluster indexes are computed using distance rather than density. The Geometric
Figure 4. Visual supremacy of our proposed VG-VAE’s unsupervised geometric labels $G_d$ achieved over supervised Attention Based Decomposition Network (ABD-Net) labels. Left side shows the results of ABD-Net and VG-VAE towards right side. We highlighted area of the given pointclouds (“Bottle, Knife and Chair”) at the specific area with maximal change in geometric signatures. We compare ours against ABD-Net therefore green colour boxes represents better performance of VG-VAE over ABD-Net and red colour indicates vice-versa.

Features of VG-VAE utilise the property of KMeans, as $G$ is distance function in a higher dimensional space.

We compare our geometric features with state-of-the-art decomposition network ABD-Net [9]. We find that supervised methodology do not generalize to capture geometric signatures across various datasets. The issue lies with overfitting of supervised methods on a single dataset, and finds challenges with other data distribution. We show similar scenario where, ABD-Net achieve state-of-the-art 99% decomposition accuracy on ANSI [21] dataset, but fail to capture basic geometric signature, when tested on ShapeNet [2]. In Figure 4, we analyse the performance of both algorithms. We compare ours against ABD-Net therefore green colour highlights in Figure 4, represents better performance of VG-VAE over ABD-Net and red colour indicates vice-versa. By considering results of ABD-Net as reference, as it is supervised method, we highlight specific regions of 6 pointclouds ("Bunny", "Chair", "Bottle", "Table", "Knife", and "Airplane") where maximal geometric change could be observed. Comparing ABD-Net we observe:

- VG-VAE captures better Geometric Signatures ("Tail", "Wings", and "Engine") on Airplane as shown in Figure 4, compared to ABD-Net.
- On observing Bottle VG-VAE is able to capture the change in geometry at bottle-head and at base of bottle, unlike ABD-Net.
- Even comparing Knife ABD-Net fails to capture the required geometric features unlike VG-VAE.
- Comparing Table, Both the methods yield similar results.
- On Chair ABD-Net is more generalized compared VG-VAE in capturing flat surfaces.
- On Bunny its hard to infer, tho VG-VAE captures change in geometries at ears.

Comprehensively VG-VAE is more robust in capturing geometrical features. We do not perform evaluation on obtained features as they are unsupervised trained.
We evaluate the effect of geometric signature by merging these as a clue to PointNet classifier and a Linear SVM. We compare the classification accuracy of our method with state-of-the-art supervised pointcloud classification method as shown in Table 2. We can infer that our method compared to discriminative methods such as ABD-Net and PointDC-CNet, our method outperforms the both.

The experiments are conducted on NVIDIA RTX 3090 GPU with 24GB RAM and AMD RYZEN threadripper 3970x CPU using PyTorch framework.

4. Conclusions

In this work, we have proposed VG-VAE: Venatus Geometric Variational Auto-Encoder for capturing unsupervised hierarchical local and global geometric signatures in pointcloud. The extraction of local geometric signatures is facilitated by the Geometric Proximity Correlator, whereas the extraction of global geometry is facilitated by variational sampling. We have provided statistical analyses of local and global geometric signatures. The impacts of our geometric features are demonstrated on pointcloud classification as downstream task using the classic pointcloud feature extractor PointNet. We have demonstrated our analysis on ModelNet40 a benchmark dataset and compare them with state-of-the-art techniques.

5. Acknowledgement

This work is partly carried out under the AICTE-RPS project “Shape Representation, Reconstruction and Rendering of 3D Models” (8-247/ RIFD/ RPS (POLICY-r)/20LB-19) and Department of Science and Technology (DST) through the ICPS programme - Indian Heritage in Digital Space for the project “Crowd-Sourcing” (DST/ ICPS/ IHDS/ 2018 (General)).

References


[18] Xiaojuan Ning, Er Li, Xiaopeng Zhang, and Yinghui Wang. Shape decomposition and understanding of point cloud objects based on perceptual information. VRCAI ’10, page 199–206, New York, NY, USA, 2010. Association for Computing Machinery. 2


[21] TraceParts S.A.S. 6


