Hybrid Vector-Occupancy Field for Robust Implicit 3D Surface Reconstruction

Yue Wu[1,2,*], Zhigang Gao[1,2], Tengfei Xiao[1,2], Can Qin[3], Yongzhe Yuan[1,2], Hao Li[1,2], Kaiyuan Feng[1,2], Wenping Ma[1,2]

[1]Xidian University
[2]MoE Key Lab of Collaborative Intelligence Systems, Xidian University
[3]Northeastern University
ywu@xidian.edu.cn,{zhiganggao,tfxiao}@stu.xidian.edu.cn qin.ca@northeastern.edu
yyz@stu.xidian.edu.cn,haoli@xidian.edu.cn,fkylwl@gmail.com,wpma@mail.xidian.edu.cn

Abstract

We introduce the Hybrid Vector-Occupancy Field (HvoF), a new implicit 3D representation for reconstructing both open and closed surfaces from sparse point clouds. Existing approaches, such as occupancy field and signed distance field, face severe limitations. They struggle with open surfaces, while unsigned distance field and neural vector field exhibit directional instability in complex topologies and ridge regions. HvoF addresses these challenges by incorporating a smoothly decaying occupancy field around the surface, while capturing precise local geometry using truncated displacement vectors, naturally mitigating direction-field ambiguities near ridge regions. This unified design forms a robust hybrid representation that leverages both occupancy and vector field. To fulfill it, we design a Hybrid Field variational autoencoder including a hierarchical cross-attention encoder and dual-branch decoder that jointly learn occupancy and vector field through continuous weighting. Extensive experiments demonstrate that HvoF consistently outperforms state-of-the-art methods across ShapeNet, ABC, and MGN datasets, accurately reconstructing both open and closed surfaces while preserving fine geometric details in complex regions.

Introduction

Reconstructing continuous surfaces from sparse point clouds plays a critical role in computer graphics, 3D vision, and robotics (Peng et al. 2021; Hoppe 2008; Mittendorfer and Cheng 2012). These methods can be broadly categorized into explicit and implicit approaches based on their output representations. Explicit representations directly predict the precise surface locations, such as meshes (Chen et al. 2024) and voxels (Du et al. 2024), but they are often constrained by resolution and topology (Chen and Zhang 2019; Jiang et al. 2020; Peng et al. 2020). Neural implicit representations, such as the occupancy field (Peng et al. 2020) and the signed distance field (SDF) (Park et al. 2019), have established a new state-of-the-art in their ability to represent surface with complex topologies at arbitrary resolutions. However, the occupancy field and SDF divide space into interior and exterior regions, so they can only handle watertight objects and cannot be used for open surface reconstruction.

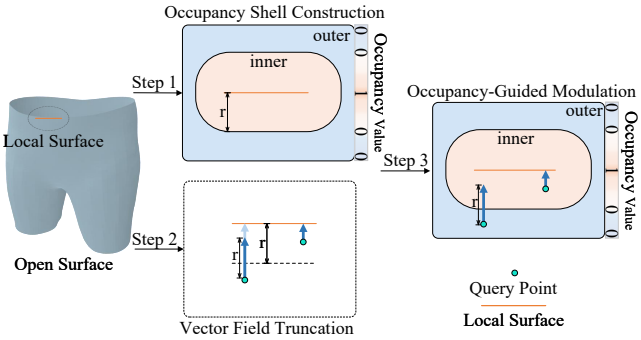

Figure 1: The HvoF pipeline begins with occupancy-shell construction, proceeds to vector-field truncation, and concludes with occupancy-guided modulation that produces the hybrid vector–occupancy field.

To address the need for representing open surfaces, the Unsigned Distance Field (UDF) was proposed. This approach reconstructs complex open surfaces by predicting continuous scalar distances from query points to the nearest surface points, along with calculating gradient directions. However, UDF suffers from gradient ambiguities at the "ridge" points, where gradients do not precisely direct toward target points. This limitation manifests as jagged edges in reconstructed surfaces, particularly at boundary regions. To address this issue, Neural Vector Field (NVF) (Yang et al. 2023) is proposed to represent shape as a vector field by directly predicting the displacement vector from a query point to the nearest surface point. However, NVF suffers from two significant limitations. First, it combines both distance and direction into a single three-dimensional vector, which increases optimization complexity during training. Second, NVF exhibits discontinuities when handling open surfaces or complex topologies, particularly near "ridge" points where the vector field undergoes abrupt changes.

In this work we introduce the Hybrid Vector–Occupancy Field (HvoF; Fig. 1) as a unified implicit representation for both closed and open surfaces. We first construct a continuous occupancy shell of thickness r, whose value decays linearly from 1 on the surface to 0 at distance r and stays 0 beyond. Conditioned on this shell, we predict a displacement vector field that preserves full vectors inside the shell and

*Corresponding author.

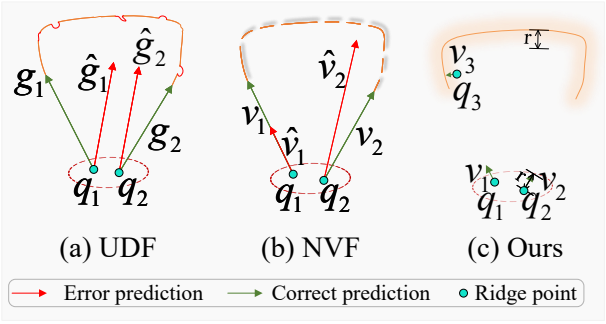

Figure 2: Ridge point challenges in open surface representation and our solution. (a) UDF suffers from gradient ambiguities (b) NVF exhibits discontinuities near boundarie (c) Our HvoF reduces ridge point effects by excluding them from the occupancy field.

clips their magnitude to r outside, keeping only directional information where large distances are unnecessary. The two components are then combined through occupancy-guided modulation, yielding an HvoF that retains the training stability of occupancy field while achieving the geometric accuracy of distance-based vectors.

Unlike previous methods that regress an exact distance or displacement for every query point, we first enclose the surface in a thin occupancy shell whose value decays linearly with distance. The network predicts full displacement vectors only for points inside this shell. Queries outside it receive direction-only vectors whose length is fixed to the shell thickness. Ambiguous ridge points therefore fall outside the shell and are supervised solely by this coarse, magnitude-clamped signal, so they cannot destabilise the fine-grained vector predictions learned near the surface. This separation markedly improves reconstruction fidelity, as illustrated in Fig. 2. The hybrid field is trained end-to-end with a variational auto-encoder in which a hierarchical encoder compresses the input point cloud into a compact latent features, and a dual-branch decoder jointly recovers the occupancy map and the occupancy-modulated vector field. Our key contributions are as follows:

- We propose a novel Hybrid Vector-Occupancy Field (HvoF) that effectively addresses the limitations of existing representations by integrating a continuous occupancy field with a truncated vector field in a region-adaptive approach.
- We develop a Hybrid Field variational autoencoder with hierarchical geometric encoder that transform point clouds into latent features, jointly decoding occupancy and vector field through an occupancy-guided modulation mechanism.
- We conduct extensive experiments demonstrating our method consistently outperforms state-of-the-art approaches on both closed and open surface across multiple datasets, with strong generalization capabilities to unseen shape categories and domains.

Related Works

Occupancy and SDF Representation. Implicit representations have emerged as a powerful paradigm for 3D shape modeling in recent years. They define shape by encoding the spatial relationship between points and surface through implicit functions, where surface boundaries emerge naturally as level sets or decision boundaries of these functions. For instance, occupancy field (Mescheder et al. 2019; Chibane, Alldieck, and Pons-Moll 2020; Peng et al. 2020; Boulch and Marlet 2022; Wang et al. 2023b; Zhang et al. 2023; Wu et al. 2024) and signed distance field (SDF) (Ben-Shabat, Koneputugodage, and Gould 2022; Park et al. 2019; Jiang et al. 2020; Chou, Bahat, and Heide 2023; Chou, Chugunov, and Heide 2022) represent shape by partitioning space into interior and exterior regions. The occupancy field is defined as a binary classification problem for points in space, determining whether each point is occupied or unoccupied (Mescheder et al. 2019). OccNet (Peng et al. 2020) integrates a convolutional encoder with an implicit occupancy decoder to enhance large-scale 3D scene reconstruction. ALTO (Wang et al. 2023b) improves reconstruction using an attention-based architecture that refines features through intelligent interpolation. Occupancy field also serve as fundamental shape representation for 3D generation (Zhang et al. 2023; Zhao et al. 2023; Petrov et al. 2024; Wu et al. 2024). 3DS2V (Zhang et al. 2023) encodes neural field into a vector set, designing a learnable occupancy representation for efficient spatial representation. To address gradient discontinuities near the surface, Direct3D (Wu et al. 2024) introduces continuous occupancy field, assigning smoothly varying values to nearby points. In contrast to occupancy field, SDF offers a more expressive representation by predicting the signed distance to the nearest surface. DeepSDF (Park et al. 2019) leverages this formulation by representing shape boundaries as the zero level set of a learned function, while DeepLS (Chabra et al. 2020) introduces grid-based latent codes for enhanced localized representation. GenSDF (Chou, Chugunov, and Heide 2022) employs a two-stage semi-supervised meta-learning framework, enabling shape priors to be transferred for reconstructing unseen categories.

UDF Representation for Open Surface. However, occupancy field and SDF can only characterize closed shape and are unable to address open surface. Recent research has explored unsigned distance field (UDF) (Chibane, Mir, and Pons-Moll 2020; Ye et al. 2022; Yang et al. 2023; Ren et al. 2023; Hu et al. 2024) to represent open surface. NDF (Chibane, Mir, and Pons-Moll 2020) designs a hierarchical neural network to learn UDF with real-valued distance supervision. GIFS (Ye et al. 2022) further generalizes shape representation by modeling multi-layer surface through spatial relationships between points. To improve UDF optimization and stability, CAP-UDF (Zhou et al. 2024) introduces consistency-aware constraints and level set projections to stabilize UDF optimization and generate more precise geometries. GeoUDF (Ren et al. 2023) takes a geometry-guided approach, representing unsigned distances as a learnable affine combination of tangent plane distances from neighboring points. In contrast to recent ap-

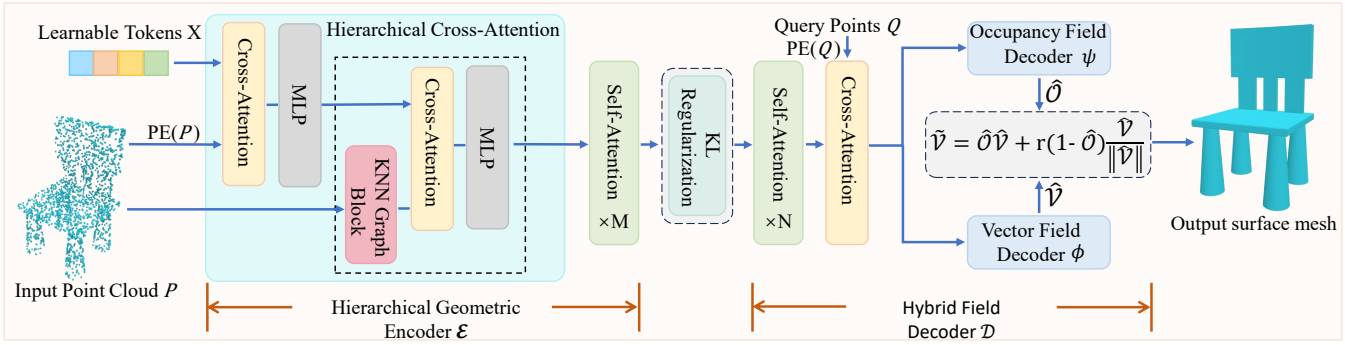

Figure 3: Architecture of our Hybrid Field VAE. Left: Hierarchical Geometric Encoder processes input point cloud P. Right: Hybrid Field Decoder predicts occupancy and vector field from query points Q, producing a hybrid representation for accurate 3D surface reconstruction.

proaches (De Luigi et al. 2023; Yu et al. 2025) that threshold UDF values within predefined ranges for learning supervision, our method preserves geometric fidelity by truncating displacement vector magnitudes while maintaining their directional information. Beyond distance-based methods, NVF (Yang et al. 2023) reformulated the distance field into a vector field, encoding both distance and directional information while resolving ambiguities at ridge points.

Approach

Given a point cloud $P \in \mathbb{R}^{N\times 3}$ and query points $Q \in \mathbb{R}^{M\times 3}$, we propose HvoF, a hybrid vector-occupancy field framework that predicts a continuous occupancy field $\mathcal{O}(q)$ and a truncated vector field $\mathcal{V}(q)$ for each query point $q \in Q$. To define our representation, we establish a continuous occupancy field within a thin shell of thickness r around the surface where full displacement vectors are maintained, while only directional information with truncated magnitude is preserved beyond this region. This hybrid representation is learned through a transformer-based Hybrid Field VAE architecture (Fig. 3) consisting of a hierarchical geometric encoder for point cloud processing and a dual-branch decoder with occupancy-guided field modulation.

Hybrid Vector-Occupancy Field

For robust representation of complex 3D shapes, we introduce the Hybrid Vector-Occupancy Field (HvoF), combining continuous occupancy and truncated vector field. To formalize HvoF, we define the following components: For any query point $q \in Q$, let $p \in P$ denote the nearest surface point, and define the displacement vector $\mathbf{\Delta q} = p - q$. The magnitude $\|\mathbf{\Delta q}\|$ equals the distance d between q and p, which defines the unsigned distance function value $udf(q)$. The unit vector $\mathbf{g} = \frac{\mathbf{\Delta q}}{\|\mathbf{\Delta q}\|}$ represents the direction from q to p.

Specifically, we design our approach around a thin shell of thickness r near the surface. We extend the semi-continuous occupancy concept introduced in Direct3D (Wu et al. 2024) with a more generalized formulation capable of handling both open and closed surfaces:

$$\mathcal{F}_o(q) = \begin{cases} 1, & \text{if } d = 0, \\ 1 - \frac{d}{r}, & \text{if } 0 < d \le r, \\ 0, & \text{if } d > r. \end{cases} \tag{1}$$

This piecewise function creates a smooth gradient from 1 at the surface ($d = 0$) to 0 beyond distance r. This continuous decay explicitly defines near-surface regions as occupied ($\mathcal{F}_o(q) > 0$) while marking distant regions as non-occupied ($\mathcal{F}_o(q)) = 0$).

Leveraging the stability benefits of constraining distance values within predefined ranges (De Luigi et al. 2023; Yu et al. 2025), we propose a new truncated vector field $\mathcal{F}_v(q)$ that adaptively integrates with our occupancy field:

$$\mathcal{F}_v(q) = \begin{cases} \mathbf{\Delta q}, & \text{if } \mathcal{F}_o(q) > 0, \\ r\mathbf{g}, & \text{if } \mathcal{F}_o(q) = 0. \end{cases} \tag{2}$$

Here, query points inside the shell ($d \le r$) retain the full displacement vector $\mathbf{\Delta q}$, capturing accurate geometry in the local region. Conversely, query points far from the surface ($d > r$) rely solely on directional information $\mathbf{g}$ with fixed magnitude r. This truncation serves two purposes: (i) it prevents large, unbounded displacement values from dominating the learning process, and (ii) it reduces unnecessary complexity in regions that are too distant to provide fine-grained geometric detail.

This hybrid representation naturally handles the challenges of complex geometries, particularly at ridge points and open boundaries, by maintaining accurate displacement vectors in critical near-surface regions while providing consistent directional guidance elsewhere.

Hybrid Field VAE

To effectively learn our hybrid vector-occupancy representation, we propose a hybrid field variational autoencoder (VAE) framework, as illustrated in Fig. 3. Our architecture adopts a transformer-based design that processes both input point cloud and query points, inspired by the perceiver architecture (Zhang et al. 2023; Li et al. 2025). As shown in our framework overview, the VAE consists of: (1) a Hierarchical Geometric Encoder $\mathcal{E}$ that maps the input point cloud

P to a latent space $\mathbf{z} \in \mathbb{R}^{L \times c}$, where L denotes the latent tokens length and c is the dimension of each token, and (2) a Hybrid Field Decoder $\mathcal{D}$ with dual branches that conditions on query points Q to jointly predict the occupancy field through an Occupancy Field Decoder ψ and the vector field through a Vector Field Decoder ϕ. This design enables end-to-end learning of detailed geometric features while maintaining global shape consistency.

Hierarchical Geometric Encoder. A key challenge in learning implicit representation lies in effectively encoding the inherent structural information of 3D shape. To address this, we propose a hierarchical attention-based encoder $\mathcal{E}$ that systematically aggregates geometric features through a progressive fusion strategy. We extract both local and global geometric features from the input surface point cloud P. We begin by mapping the point cloud into a continuous position-aware representation through Fourier feature embedding (Jaegle et al. 2021), yielding $\mathrm{PE}(P)$. To capture localized information, we introduce learnable tokens $\mathbf{X} \in \mathbb{R}^{L \times e}$ that interact with local point features through cross-attention. For capturing global context, we employ a KNN graph network (Wang et al. 2019) that dynamically connects each point with its neighbors, obtaining the global descriptor through element-wise max-pooling. This global representation is then integrated with the local features through a second cross-attention operation. The hierarchical attention design enables comprehensive feature fusion at different scales. The final latent representation $\mathbf{z}$ is obtained through iterative self-attention refinement that captures long-range dependencies and spatial correlations across the shape geometry.

Hybrid Field Decoder. The decoder $\mathcal{D}$ reconstructs the continuous occupancy field and vector field from the latent representation $\mathbf{z}$. It consists of a self-attention module followed by cross-attention and two independent decoding branches. First, self-attention layers operate on $\mathbf{z}$ to refine spatial correlations and capture long-range dependencies. The refined features are then processed through cross-attention, where query points Q interact with $\mathbf{z}$, enabling each query to extract relevant geometric information.

Following this interaction, the decoder $\mathcal{D}$ splits into two specialized branches: the Occupancy Field Decoder ψ and the Vector Field Decoder ϕ, both implemented using linear projection layers. The Occupancy Field Decoder predicts the continuous occupancy field $\hat{\mathcal{O}}$, while the Vector Field Decoder estimates the displacement vector field $\hat{\mathcal{V}}$:

$$\begin{aligned} \hat{\mathcal{O}} &= \psi\left(\mathrm{CrossAttn}\left(\mathrm{PE(Q)}, \mathrm{SelfAttn}^{(i\ldots n)}\left(\mathbf{z}\right)\right)\right), \\ \hat{\mathcal{V}} &= \phi\left(\mathrm{CrossAttn}\left(\mathrm{PE(Q)}, \mathrm{SelfAttn}^{(i\ldots n)}\left(\mathbf{z}\right)\right)\right), \end{aligned} \tag{3}$$

where n denotes the number of self-attention layers.

The dual decoders produce two independent field predictions: $\hat{\mathcal{O}}$ and $\hat{\mathcal{V}}$. Although both capture essential geometric information, nothing yet enforces our hybrid-field constraints. In particular, $\hat{\mathcal{V}}$ exhibits unconstrained vector magnitudes across all regions, and the occupancy and vector fields are not coordinated to realize the intended spatial behaviour. To bridge this gap, we introduce occupancy-guided vector modulation:

$$\tilde{\mathcal{V}} = \hat{\mathcal{O}}\hat{\mathcal{V}} + r\left(1 - \hat{\mathcal{O}}\right)\frac{\hat{\mathcal{V}}}{\parallel \hat{\mathcal{V}} \parallel}. \tag{4}$$

This formulation creates a continuous transition across our hybrid field: when $\hat{\mathcal{O}} = 1$ (near surfaces), $\tilde{\mathcal{V}} = \hat{\mathcal{V}}$, preserving full displacement vectors; when $\hat{\mathcal{O}} = 0$ (far from surfaces), $\tilde{\mathcal{V}} = r\frac{\hat{\mathcal{V}}}{\|\hat{\mathcal{V}}\|}$, retaining only directional information with fixed magnitude r. For intermediate occupancy values, the equation produces a weighted blend between these behaviors, ensuring smooth transitions while conforming to our hybrid field definition and effectively integrating information from both predicted field to achieve high-fidelity representation of complex 3D surfaces.

Optimization

To effectively train our hybrid field VAE framework, we design a comprehensive loss function that ensures accurate prediction of both the continuous occupancy field and the truncated vector field. For occupancy prediction, we employ Binary Cross-Entropy (BCE) loss:

$$\mathcal{L}_o = \mathrm{BCE}(\hat{\mathcal{O}}, \mathcal{O}). \tag{5}$$

Here, $\hat{\mathcal{O}} = \mathcal{D}_\psi(\mathcal{E}(P), Q)$ represents the predicted occupancy field generated from input point cloud P and query points Q via our hierarchical geometric encoder $\mathcal{E}$ and Occupancy Field Decoder ψ, while $\mathcal{O} = \mathcal{F}_o(Q)$ denotes the ground-truth occupancy values derived from our hybrid field formulation. For vector field supervision, we implement L2 loss between the modulated vector field and ground-truth vectors:

$$\mathcal{L}_v = \left\| \tilde{\mathcal{V}} - \mathcal{V} \right\|_2^2, \tag{6}$$

where $\tilde{\mathcal{V}}$ is the occupancy-guided modulated vector field as defined in Eq. 4, with $\hat{\mathcal{V}} = \mathcal{D}_\phi(\mathcal{E}(P), Q)$ being the raw output from the Vector Field Decoder ϕ. This ensures that our supervision aligns with the hybrid field properties defined in Sec. , where $\mathcal{V} = \mathcal{F}_v(Q)$ corresponds to the ground-truth vectors.

Our complete loss function integrates these supervision terms with VAE regularization:

$$\mathcal{L} = \mathcal{L}_o + \mathcal{L}_v + \lambda\mathcal{L}_{KL}, \tag{7}$$

where $\mathcal{L}_{KL}$ represents the KL divergence loss for latent space regularization, with $\lambda = 1 \times 10^{-3}$ controlling its strength. This joint optimization strategy preserves the correct relationship between occupancy and vector field while maintaining our desired hybrid field properties.

Experiments

Implementation Details. During preprocessing, we normalize shapes to [-0.8, 0.8]. For training, we sample 2,048 surface points as input and 4,096 query points within [-1, 1]. Our network architecture incorporates learnable tokens $\mathbf{X} \in \mathbb{R}^{512 \times 512}$ and latent representation $\mathbf{z} \in \mathbb{R}^{512 \times 64}$ with

Methods	Base Category				Novel Category			
	CD↓	EMD↓	$F1_{2.5\times10^{-5}}$↑	$F1_{1\times10^{-4}}$↑	CD↓	EMD↓	$F1_{2.5\times10^{-5}}$↑	$F1_{1\times10^{-4}}$↑
GridFormer(Li et al. 2024)	0.364	1.876	79.859	92.025	0.374	1.928	81.409	89.873
GEM3D(Petrov et al. 2024)	0.400	**1.332**	77.528	95.270	0.446	1.630	75.869	93.516
GeoUDF(Ren et al. 2023)	0.465	1.444	66.463	93.000	0.451	1.609	70.270	92.879
3DS2V(Zhang et al. 2023)	0.895	1.970	59.783	84.020	1.076	2.528	53.933	80.043
NVF(Yang et al. 2023)	0.452	1.406	69.653	93.012	0.514	**1.537**	73.042	92.602
Ours	**0.299**	1.671	**87.605**	**98.159**	**0.323**	1.744	**84.713**	**96.593**

Table 1: Quantitative comparison between Seen and Zero-shot reconstruction on ShapeNet. "Base" categories represent Seen reconstruction results, while "Novel" categories show Zero-shot reconstruction performance.

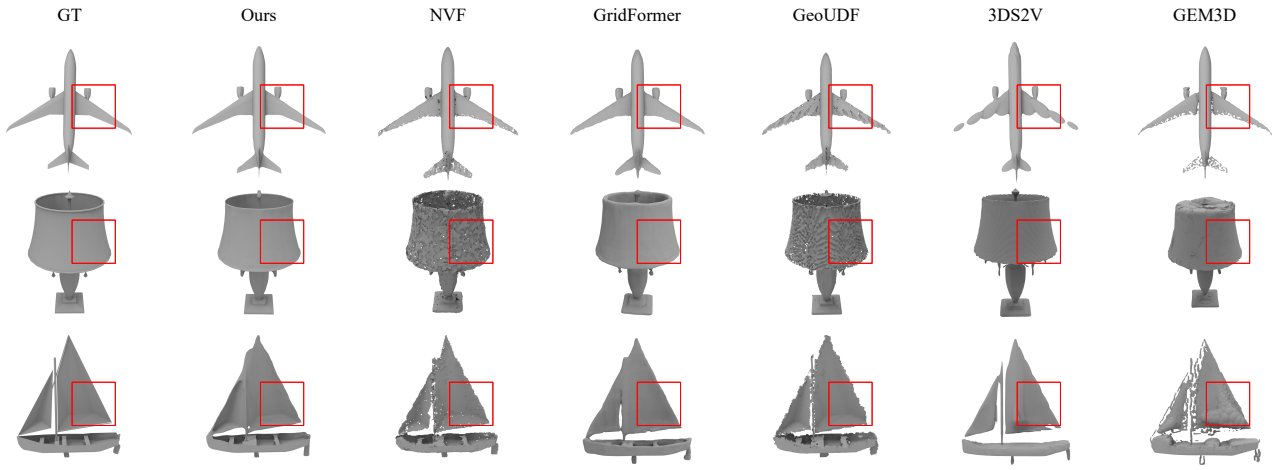

Figure 4: Qualitative comparison on ShapeNet dataset. The first row shows reconstruction results on "Base" categories, while the last two rows present "Novel" categories.

truncation threshold $r = 0.1$. The VAE structure comprises $m = 8$ self-attention layers in the encoder and $n = 16$ in the decoder. We optimize the model using the Adam optimizer with a learning rate of 1×10^{-4} on an NVIDIA RTX 4090D GPU. For surface reconstruction, we implement DoubleCoverUDF (Hou et al. 2023) to efficiently exploit our hybrid field representation.

Datasets. Following the experimental protocol in PEIF (Hu et al. 2024), we evaluate our method on three representative datasets: ShapeNet (Chang et al. 2015), ABC (Koch et al. 2019), and MGN (Bhatnagar et al. 2019). **(1) ShapeNet** is a large-scale repository of 3D CAD models spanning 55 object categories. For our experiments, we designate cars, chairs, planes, and tables as "base" categories, while speakers, benches, lamps, and watercraft serve as "novel" categories for generalization testing. **(2) ABC** dataset contains a diverse collection of one million CAD models, predominantly comprising mechanical parts. We follow the same data splits as (Wang et al. 2023a) for our experiments. **(3) MGN** dataset consists of five categories of open garment meshes, which we exclusively employ for testing our model's generalization capability to deformable surface.

Evaluation Metrics and Comparisons. We evaluate using Chamfer-L1 distance (CD, $\times 10^{-2}$), Earth Mover Distance (EMD, $\times 10^{-2}$), and F-score ($\times 10^{-2}$) with thresholds 2.5×10^{-4} and 1×10^{-4}. We compare our method against UDF-based approaches (GeoUDF (Ren et al. 2023), NVF (Yang et al. 2023)), and occupancy field-based methods (3DS2V (Zhang et al. 2023), GEM3D (Petrov et al. 2024), GridFormer (Li et al. 2024)). For fair comparison, all baselines were retrained on identical datasets under the same conditions.

Seen and Zero-shot Reconstruction

We evaluate our method on ShapeNet for both base (Seen) and novel (Zero-shot) categories, and on the ABC dataset for Seen reconstruction. As shown in Tab. 1 and 2, HvoF achieves superior performance in most metrics, particularly excelling in CD and F1-scores while maintaining competitive performance in EMD. Fig. 4 and 5 illustrate the advantages of our hybrid representation over baseline methods. In ShapeNet examples, NVF (Yang et al. 2023) suffers from vector prediction errors that create holes in reconstructed surfaces, as evidenced in the watercraft sails. GeoUDF (Ren et al. 2023) produces jagged edges due to gradient ambiguities at ridge points where UDF methods struggle with directional guidance. Our thin-shell design and occupancy-guided vector modulation effectively address these issues by preserving accurate displacement vectors near surfaces while providing stable directional infor-

mation elsewhere. This approach enables HvoF to reconstruct complex topological features with smooth transitions across boundaries, particularly evident in the mechanical structures of the ABC dataset. The combination of continuous occupancy field with truncated vector field allows our method to maintain consistent reconstruction quality in both Seen and Zero-shot scenarios, showing robust generalization capability with moderate performance differences between base and novel categories.

Method	CD↓	EMD↓	$F1_{1\times10^{-4}}$↑
3DS2V(Zhang et al. 2023)	0.465	1.263	95.435
GEM3D(Petrov et al. 2024)	0.356	**1.100**	96.595
GeoUDF(Ren et al. 2023)	<u>0.335</u>	<u>1.124</u>	<u>97.506</u>
NVF(Yang et al. 2023)	0.390	1.179	94.369
Ours	**0.266**	1.137	**98.753**

Table 2: Quantitative comparison of surface reconstruction performance on ABC dataset.

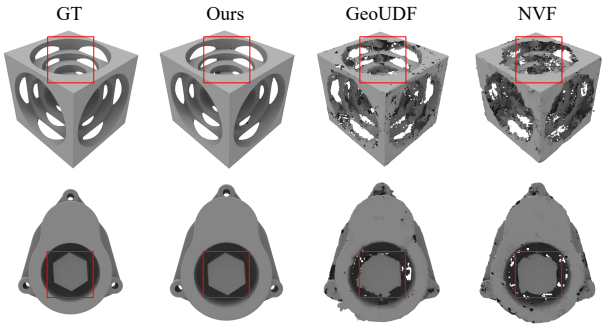

Figure 5: Comparison on ABC dataset shows our method produces smoother surface and clearer geometric details on topologically complex shapes.

Single-category Reconstruction

To systematically evaluate the performance of our method in single-category reconstruction, we conduct comprehensive experiments on the car category from ShapeNet. Cars represent an ideal testbed for single-category reconstruction due to their complex geometric structures, including intricate surface details (e.g., mirrors, wheels) and varying topological features. All methods were retrained on the car category under identical settings for fair comparison.

As shown in Tab. 3, our method achieves the best CD score and competitive results on other metrics. While NVF achieves the lowest EMD, and GEM3D (Petrov et al. 2024) shows the highest F1 score, our method offers the best overall balance across all evaluation metrics. Fig. 6 illustrates the qualitative advantages of our approach. Occupancy-based methods 3DS2V (Zhang et al. 2023), GEM3D (Petrov et al. 2024) struggle with structural completeness, particularly in thin regions such as side mirrors and wheel structures due to their inherent binary space partitioning limitation. NVF (Yang et al. 2023) exhibits significant detail loss in geometric features, while GeoUDF (Ren et al. 2023) produces boundary blurring at critical junctions where directional gradient ambiguities occur. In contrast, our hybrid representation effectively preserves intricate car details with precise geometry reconstruction. The occupancy-guided vector modulation maintains the full displacement vectors near surfaces while stabilizing predictions in other regions, enabling our method to capture sharp features and smooth transitions simultaneously in complex automotive structures.

Method	CD ↓	EMD ↓	$F1_{1\times10^{-4}}$↑
GEM3D(Petrov et al. 2024)	0.556	1.302	**96.234**
3DS2V(Zhang et al. 2023)	0.859	1.684	78.491
GeoUDF(Ren et al. 2023)	<u>0.494</u>	<u>1.275</u>	91.163
NVF(Yang et al. 2023)	0.511	**1.136**	90.005
Ours	**0.413**	1.384	<u>94.528</u>

Table 3: Quantitative comparison of different methods on ShapeNet cars reconstruction.

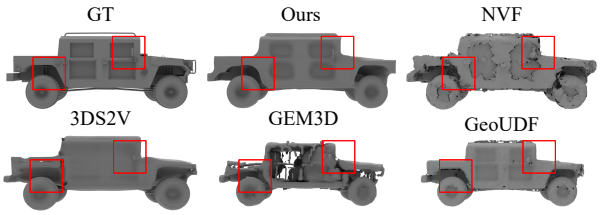

Figure 6: Qualitative comparison on ShapeNet cars. Our method preserves fine details (e.g., mirrors, wheels) with smoother surface than baselines, closely approximating ground-truth geometry.

Cross-domain Reconstruction

We evaluate cross-domain generalization capability by applying our model trained on ShapeNet-base to the MGN dataset, which consists of real-world scanned garments with challenging open surfaces. Tab. 4 shows our method achieves the best CD score while maintaining competitive performance in EMD and F1 metrics. As illustrated in Fig. 7, baseline methods struggle with complex regions like collar openings and pant hems. NVF (Yang et al. 2023) suffers from vector prediction errors that create holes in reconstructed surfaces, while GeoUDF (Ren et al. 2023) produces jagged edges due to gradient ambiguities at ridge points. In contrast, our hybrid representation effectively captures intricate clothing features with smoother transitions and more complete details in these challenging areas. This demonstrates robust generalization from synthetic to real-world data, particularly for complex open surfaces.

Ablation Studies

We perform ablation studies on the ABC dataset to evaluate our field representation design and shell thickness config-

Method	CD↓	EMD↓	$F1_{1\times10^{-4}}$↑
GeoUDF(Ren et al. 2023)	0.341	0.742	97.547
NVF(Yang et al. 2023)	0.311	**0.734**	**97.978**
Ours	**0.310**	0.740	97.388

Table 4: Quantitative comparison of cross-domain reconstruction on MGN dataset.

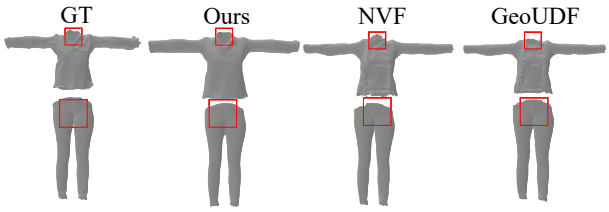

Figure 7: Open-surface garment reconstruction on MGN

uration, validating component effectiveness and framework robustness.

Field Representation Analysis. We compare three field representation strategies using the same model. As shown in Fig. 8, the occupancy field reconstruction exhibits clear structural outlines but lacks detailed features on the mechanical part, especially at complex junctions and edges. The untruncated vector field shows significant degradation near surface boundaries and ridge points, where abrupt changes in displacement vectors lead to blurred geometric structures. This is primarily due to the inherent difficulty in simultaneously predicting accurate distance and direction information in a coupled vector representation. Our hybrid design addresses these limitations through occupancy-guided truncation, as evidenced by the sharp feature preservation and clean surface reconstruction. The quantitative results in Tab. 5 further support these observations, with our method achieving superior EMD and F1 scores while maintaining competitive CD performance.

Method	CD ↓	EMD ↓	$F1_{1\times10^{-4}}$↑
Occupancy Field	**0.234**	2.182	98.547
Vector Field	0.349	2.391	93.916
Ours	0.266	**1.137**	**98.753**

Table 5: Ablation study on different field representations.

Shell Thickness Analysis. The shell thickness parameter r in our hybrid field representation directly influences surface detail capture. We conduct experiments with different r values ranging from 0.05 to 0.15 as shown in Tab. 6. Our method achieves optimal performance at $r = 0.10$, demonstrating the best balance across all metrics. A very small r creates too narrow a region for effective surface representation, while larger values cause over-smoothing. As shown in Fig. 9, the moderate value $r = 0.10$ achieves optimal reconstruction quality, particularly in preserving detailed geometric features at complex junctions.

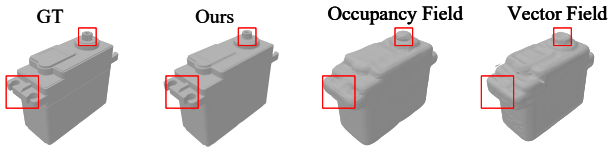

Figure 8: Comparison of different field representations on a mechanical part.

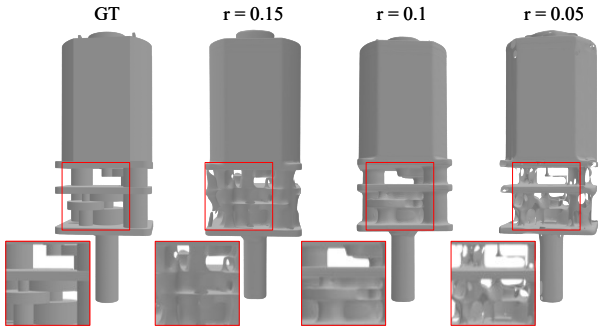

Figure 9: Visual comparison of reconstruction quality with different shell thickness values r

Shell Thickness (r)	CD ↓	EMD ↓	$F1_{1\times10^{-4}}$↑
0.05	0.279	1.316	98.684
0.10	**0.266**	**1.137**	**98.753**
0.15	0.339	1.489	96.009

Table 6: Impact of shell thickness on reconstruction quality.

Conclusion

In this paper, we introduced Hybrid Vector-Occupancy Field (HvoF), a novel 3D shape representation that combines a continuous occupancy field within a thin shell with a truncated vector field for unified modeling of both closed and open surface. Our approach effectively resolves limitations of existing implicit methods through a region-selective strategy that preserves complete displacement vectors near surfaces while maintaining only directional information elsewhere. The dual-branch VAE with hierarchical cross-attention encoder extracts multi-scale geometric features and implements occupancy-guided modulation for smooth field transitions. Extensive experiments on ShapeNet, ABC, and MGN datasets demonstrate HvoF consistently outperforms state-of-the-art methods in reconstruction accuracy and generalization capability across diverse object categories and domains.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (62036006, 62276200) and the Innovation Capability Support Plan of Shaanxi Province (2023KJXX-144).

References

Ben-Shabat, Y.; Koneputugodage, C. H.; and Gould, S. 2022. Digs: Divergence guided shape implicit neural representation for unoriented point clouds. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 19323–19332.

Bhatnagar, B. L.; Tiwari, G.; Theobalt, C.; and Pons-Moll, G. 2019. Multi-garment net: Learning to dress 3d people from images. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 5420–5430.

Boulch, A.; and Marlet, R. 2022. POCO: Point Convolution for Surface Reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 6302–6314.

Chabra, R.; Lenssen, J. E.; Ilg, E.; Schmidt, T.; Straub, J.; Lovegrove, S.; and Newcombe, R. 2020. Deep local shapes: Learning local sdf priors for detailed 3d reconstruction. In *Proceedings of the European conference on computer vision*, 608–625. Springer.

Chang, A. X.; Funkhouser, T.; Guibas, L.; Hanrahan, P.; Huang, Q.; Li, Z.; Savarese, S.; Savva, M.; Song, S.; Su, H.; et al. 2015. Shapenet: An information-rich 3d model repository. *arXiv:1512.03012*.

Chen, Y.; He, T.; Huang, D.; Ye, W.; Chen, S.; Tang, J.; Chen, X.; Cai, Z.; Yang, L.; Yu, G.; Lin, G.; and Zhang, C. 2024. MeshAnything: Artist-Created Mesh Generation with Autoregressive Transformers. arXiv:2406.10163.

Chen, Z.; and Zhang, H. 2019. Learning implicit fields for generative shape modeling. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 5939–5948.

Chibane, J.; Alldieck, T.; and Pons-Moll, G. 2020. Implicit functions in feature space for 3d shape reconstruction and completion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 6970–6981.

Chibane, J.; Mir, A.; and Pons-Moll, G. 2020. Neural Unsigned Distance Fields for Implicit Function Learning. In *Advances in Neural Information Processing Systems*.

Chou, G.; Bahat, Y.; and Heide, F. 2023. Diffusion-SDF: Conditional Generative Modeling of Signed Distance Functions. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2262–2272.

Chou, G.; Chugunov, I.; and Heide, F. 2022. Gensdf: Two-stage learning of generalizable signed distance functions. *Advances in Neural Information Processing Systems*, 35: 24905–24919.

De Luigi, L.; Li, R.; Guillard, B.; Salzmann, M.; and Fua, P. 2023. DrapeNet: Garment Generation and Self-Supervised Draping. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 1451–1460.

Du, H.; Yan, X.; Wang, J.; Xie, D.; and Pu, S. 2024. Arbitrary-Scale Point Cloud Upsampling by Voxel-Based Network with Latent Geometric-Consistent Learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, 1626–1634.

Hoppe, H. 2008. Poisson surface reconstruction and its applications. In *Proceedings of the 2008 ACM Symposium on Solid and Physical Modeling*, 10–10.

Hou, F.; Chen, X.; Wang, W.; Qin, H.; and He, Y. 2023. Robust Zero Level-Set Extraction from Unsigned Distance Fields Based on Double Covering. *ACM Trans. Graph.*, 42(6).

Hu, X.; Tang, X.; Yu, R.; and Sun, J. 2024. Learning 3D equivariant implicit function with patch-level pose-invariant representation. In *Advances in Neural Information Processing Systems*.

Jaegle, A.; Gimeno, F.; Brock, A.; Vinyals, O.; Zisserman, A.; and Carreira, J. 2021. Perceiver: General perception with iterative attention. In *International Conference on Machine Learning*, 4651–4664. PMLR.

Jiang, C.; Sud, A.; Makadia, A.; Huang, J.; Nießner, M.; Funkhouser, T.; et al. 2020. Local implicit grid representations for 3d scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 6001–6010.

Koch, S.; Matveev, A.; Jiang, Z.; Williams, F.; Artemov, A.; Burnaev, E.; Alexa, M.; Zorin, D.; and Panozzo, D. 2019. Abc: A big cad model dataset for geometric deep learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 9601–9611.

Li, S.; Gao, G.; Liu, Y.; Liu, Y.-S.; and Gu, M. 2024. GridFormer: Point-Grid Transformer for Surface Reconstruction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, 3163–3171.

Li, W.; Liu, J.; Yan, H.; Chen, R.; Liang, Y.; Chen, X.; Tan, P.; and Long, X. 2025. CraftsMan3D: High-fidelity Mesh Generation with 3D Native Diffusion and Interactive Geometry Refiner. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 5307–5317.

Mescheder, L.; Oechsle, M.; Niemeyer, M.; Nowozin, S.; and Geiger, A. 2019. Occupancy Networks: Learning 3D Reconstruction in Function Space. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 4455–4465.

Mittendorfer, P.; and Cheng, G. 2012. 3D surface reconstruction for robotic body parts with artificial skins. In *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 4505–4510. IEEE.

Park, J. J.; Florence, P.; Straub, J.; Newcombe, R.; and Lovegrove, S. 2019. Deepsdf: Learning continuous signed distance functions for shape representation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 165–174.

Peng, S.; Jiang, C.; Liao, Y.; Niemeyer, M.; Pollefeys, M.; and Geiger, A. 2021. Shape as points: A differentiable poisson solver. *Advances in Neural Information Processing Systems*, 34: 13032–13044.

Peng, S.; Niemeyer, M.; Mescheder, L.; Pollefeys, M.; and Geiger, A. 2020. Convolutional occupancy networks. In *Proceedings of the European Conference on Computer Vision*, 523–540. Springer.

Petrov, D.; Goyal, P.; Thamizharasan, V.; Kim, V.; Gadelha, M.; Averkiou, M.; Chaudhuri, S.; and Kalogerakis, E. 2024. GEM3D: GEnerative Medial Abstractions for 3D Shape Synthesis. In *ACM SIGGRAPH 2024 Conference*. Association for Computing Machinery.

Ren, S.; Hou, J.; Chen, X.; He, Y.; and Wang, W. 2023. Geoudf: Surface reconstruction from 3d point clouds via geometry-guided distance representation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 14214–14224.

Wang, Y.; Huang, Z.; Shamir, A.; Huang, H.; Zhang, H.; and Hu, R. 2023a. Aro-net: learning implicit fields from anchored radial observations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 3572–3581.

Wang, Y.; Sun, Y.; Liu, Z.; Sarma, S. E.; Bronstein, M. M.; and Solomon, J. M. 2019. Dynamic graph cnn for learning on point clouds. *ACM Transactions on Graphics*, 38(5): 1–12.

Wang, Z.; Zhou, S.; Park, J. J.; Paschalidou, D.; You, S.; Wetzstein, G.; Guibas, L.; and Kadambi, A. 2023b. Alto: Alternating latent topologies for implicit 3d reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 259–270.

Wu, S.; Lin, Y.; Zhang, F.; Zeng, Y.; Xu, J.; Torr, P.; Cao, X.; and Yao, Y. 2024. Direct3D: scalable image-to-3D generation via 3D latent diffusion transformer. In *Advances in Neural Information Processing Systems*. Curran Associates Inc.

Yang, X.; Lin, G.; Chen, Z.; and Zhou, L. 2023. Neural Vector Fields: Implicit Representation by Explicit Learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 16727–16738.

Ye, J.; Chen, Y.; Wang, N.; and Wang, X. 2022. Gifs: Neural implicit function for general shape representation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 12829–12839.

Yu, Z.; Dou, Z.; Long, X.; Lin, C.; Li, Z.; Liu, Y.; Müller, N.; Komura, T.; Habermann, M.; Theobalt, C.; et al. 2025. Surf-D: Generating High-Quality Surfaces of Arbitrary Topologies Using Diffusion Models. In *Proceedings of the European Conference on Computer Vision*, 419–438. Springer.

Zhang, B.; Tang, J.; Niessner, M.; and Wonka, P. 2023. 3dshape2vecset: A 3d shape representation for neural fields and generative diffusion models. *ACM Transactions on Graphics*, 42(4): 1–16.

Zhao, Z.; Liu, W.; Chen, X.; Zeng, X.; Wang, R.; Cheng, P.; FU, B.; Chen, T.; YU, G.; and Gao, S. 2023. Michelangelo: Conditional 3D Shape Generation based on Shape-Image-Text Aligned Latent Representation. In *Advances in Neural Information Processing Systems*.

Zhou, J.; Ma, B.; Li, S.; Liu, Y.-S.; Fang, Y.; and Han, Z. 2024. CAP-UDF: Learning Unsigned Distance Functions Progressively From Raw Point Clouds With Consistency-Aware Field Optimization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(12): 7475–7492.