NEF: Neural Edge Fields for 3D Parametric Curve Reconstruction from Multi-view Images

Yunfan Ye Renjiao Yi Zhirui Gao Chenyang Zhu Zhiping Cai Kai Xu*
National University of Defense Technology

Abstract

We study the problem of reconstructing 3D feature curves of an object from a set of calibrated multi-view images. To do so, we learn a neural implicit field representing the density distribution of 3D edges which we refer to as Neural Edge Field (NEF). Inspired by NeRF [20], NEF is optimized with a view-based rendering loss where a 2D edge map is rendered at a given view and is compared to the ground-truth edge map extracted from the image of that view. The rendering-based differentiable optimization of NEF fully exploits 2D edge detection, without needing a supervision of 3D edges, a 3D geometric operator or cross-view edge correspondence. Several technical designs are devised to ensure learning a range-limited and view-independent NEF for robust edge extraction. The final parametric 3D curves are extracted from NEF with an iterative optimization method. On our benchmark with synthetic data, we demonstrate that NEF outperforms existing state-of-the-art methods on all metrics. Project page: https://yunfan1202.github.io/NEF/.

1. Introduction

Feature curves “define” 3D shapes to an extent, not only geometrically (surface reconstruction from curve networks [15, 16]) but also perceptually (feature curve based shape perception [4, 35]). Therefore, feature curve extraction has been a long-standing problem in both graphics and vision. Traditional approaches to 3D curve extraction often work directly on 3D shapes represented by, e.g., polygonal meshes or point clouds. Such approaches come with a major difficulty: Sharp edges may be partly broken or completely missed due to imperfect 3D acquisition and/or reconstruction. Consequently, geometrically-based methods, even the state-of-the-art ones, are sensitive to parameter settings and error-prune near rounded edges, noise, and sparse data. Recently, learning-based methods are proposed to address these issues but with limited generality [18,19,33,39].

In many cases, edges are visually prominent and easy to detect in the 2D images of a 3D shape. To resolve occlusion, one may think of 3D curve reconstruction from multi-view edges. This solution, however, relies strongly on cross-view edge correspondence which itself is a highly difficult problem [28]. This explains why there is rarely a work on multi-view curve reconstruction even in the deep learning era. We ask this question: Can we learn 3D feature curve extraction directly from the input of multi-view images?

In this work, we try to answer the question through learning a neural implicit field representing the density distribution of 3D edges from a set of calibrated multi-view images, inspired by the recent success of neural radiance field (NeRF) [20]. We refer to this edge density field as Neural Edge Field (NEF). Similar to NeRF, NEF is optimized with a view-based rendering loss where a 2D edge map is rendered at a given view and is compared to the ground-truth edge map extracted from the image of that view. The
volumetric rendering is based on edge density and color (gray-scale) predicted by MLPs along viewing rays. Different from NeRF, however, our goal is merely to optimize the NEF which is later used for extracting parametric 3D curves; no novel view synthesis is involved. The rendering-based differentiable optimization of 3D edge density fully exploits 2D edge detection, without needing a 3D geometric operator or cross-view edge correspondence. The latter is implicitly learned with multi-view consistency.

Directly optimizing NEF as NeRF-like density is problematic since the range of density can be arbitrarily large and different from scene to scene, and it is hard to select a proper threshold to extract useful geometric shapes (e.g., 3D surfaces for NeRF and 3D edges for NEF). Moreover, NeRF density usually does not approximate the underlying 3D shape well due to noise. Therefore, we seek for confining the edge density in the range of [0, 1] through learning a mapping function with a learnable scaling factor to map the edge density to the actual NEF density. By doing so, we can easily choose a threshold to extract edges robustly from the optimized edge density.

Another issue with NEF optimization is the incompatible visibility of the edge density field and the edges detected in images. While the former is basically a wireframe representation of the underlying 3D shape and all edges are visible from any view (i.e., no self-occlusion), edges in 2D images can be occluded by the object itself. This leads to inconsistent supervisions of different views with different visibility and may cause false negative: An edge that should have been present in NEF according to one view visible to the edge may be suppressed by other views invisible. To address this issue, we opt to 1) impose consistency between density and color in NEF and 2) give less punishment on non-edge pixels in the rendering loss, to allow the NEF to keep all edges seen from all views. This essentially makes NEF view-independent which is reasonable.

Having obtained the edge density, we fit parametric curves by treating the 3D density volume as a point cloud of edges. We optimize the control points of curves in a coarse-to-fine manner. Since initialization is highly important to such a non-convex optimization, we first apply line fitting to cover most points. Based on the initialization, we then upgrade lines to cubic Bézier curves by adding extra control points and optimize all curves simultaneously with an extra endpoint regularization.

We build a benchmark with a synthetic dataset consisting of 115 CAD models with complicated shape structures from ABC dataset [14] and utilize BlenderProc [7] to render posed images. Extensive experiments on the proposed dataset show that NEF, which is self-trained with only 2D supervisions, outperforms existing state-of-the-art methods on all metrics. Our contributions include:

- A self-supervised 3D edge detection from multi-view 2D edges based neural implicit field optimization.
- Several technical designs to ensure learning a range-limited and view-independent NEF and an iterative optimization strategy to reconstruct parametric curves.
- A benchmark for evaluating and comparing various edge/curve extraction methods.

2. Related Work

Neural Radiance Fields. NeRF [20] have demonstrated the remarkable ability for novel view synthesis. The basic idea of NeRF to represent the geometry and appearance of a scene as a radiance field, allowing querying color and volume density in continuous spatial positions and viewing directions for rendering. Many extensions are designed on the NeRF backbone, such as speeding up the training [25, 30] and the inference [3, 13, 38], editing [17, 31, 40], generative models [21, 27] and model reconstruction [22, 32, 37], and more are discussed in [5, 9]. However, there are not many works that utilize NeRFs to extract 3D skeletons.curves. We propose Neural Edge Fields (NEF) to reconstruct 3D edges from 2D images. The closest NeRF-based works to ours are for model reconstruction [32, 37], and we all recover the precise shape geometry by defining the original density as a transformed new representation. The difference is, they represent the surface by zero-level sets of the signed distance function (SDF) and focus on surface reconstruction; while we introduce edge density to represent the edge probability at every spatial position by learning a NEF.

3D Parametric Curve Reconstruction. The basis of 3D parametric curve reconstruction is point cloud edge detection algorithms. Traditional (non-learning) methods focus on multi-view images [23], or local geometric properties of point clouds such as normals [6, 34], curvatures [36], and hierarchical clustering [8]. Recent data-driven methods often adopt edge detection as a binary classification for point clouds. For each possible edge point, its neighborhood attributes are taken as the learning features. With the progress of network architectures, the classifier for edge detection ranges from random forests [10, 11], Pointwise Multilayer Perceptron (MLP) [33, 39] based on PointNet++ [24], to capsule networks [2].

Representing point cloud edges as parametric curves is more challenging. PIE-NET [33] learns to detect edges and corners from point clouds, and generates parametric curve proposals using networks, suppressing the invalid ones at last. PC2WF [18] is composed of a sequence of feed-forward blocks to sample point cloud as patches, to classify if the patch contains a corner. They regress the location of corners, and connect all corners as parametric curves. DEF [19] calculates estimates of the truncated distance-to-feature field for each input point cloud by an extra set of depth images in a patch-based manner, and fitting curves...
after corner detection and clustering. Unlike those works requiring at least point clouds as input and training from labeled datasets, our method is self-supervised by 2D edges.

3. Method

To obtain 3D parametric curves from multi-view images, the method consists of two steps: building neural edge fields (NEF) and reconstructing parametric curves. As illustrated in Fig. 1, 2D edge maps are predicted by a state-of-the-art edge detection network PiDiNet [29], and NEF is built from these multi-view edge maps, as introduced in Sec. 3.1. Adopting NeRF directly on edge maps is problematic, and there are many differences between edge maps and original images. Edges are sparse, and inconsistent among views due to occlusions. To deal with them, we introduce several training losses specifically designed for NEF. To reconstruct the parametric curves from 3D edge points, we introduce two-stage coarse-to-fine optimization in Sec. 3.2. In the coarse-stage, we simplify curves to straight lines, and fit a group of lines to 3D edge points in a fit-and-delete strategy. In the fine-stage, we upgrade straight lines into cubic Bézier curves by adding extra control points.

3.1. Reconstructing 3D Edge Points

In this section, we learn a neural implicit filed representing the spatial distribution of 3D edges, named neural edge field (NEF). We first introduce preliminaries about NeRF in Sec. 3.1.1. The design of NEF is introduced in Sec. 3.1.2. Training NEF requires specific loss designs, as introduced in Sec. 3.1.3.

3.1.1 Preliminaries

NeRF [20] represents a continuous scene with an MLP network, which maps 5D coordinates (location \((x, y, z)\) and viewing direction \((\theta, \phi)\)) among camera rays, to color \((r, g, b)\) and volume density \(\sigma\). After training, novel views can be rendered from arbitrary camera poses, following volume rendering. The camera origin \(o\) and ray direction \(d\) with near and far bounds \(t_n\) and \(t_f\), the predicted pixel color \(\hat{C}\) of camera ray \(r(t) = o + td\) is defined as follows:

\[
\hat{C}(r) = \int_{t_n}^{t_f} T(t)\sigma(r(t))c(r(t), d)dt, \tag{1}
\]

where \(T(t) = \exp \left( -\int_{t_n}^{t} \sigma(r(s))ds \right)\), and densities \(\sigma\) and colors \(c\) are predictions of the MLP network. The loss function of NeRF is a re-rendering loss defined by the mean squared error (MSE) between the rendered color \(\hat{C}\) and the ground truth pixel color \(C\):

\[
\mathcal{L}_{\text{color}} = \sum_{r \in R} \| (C(r) - \hat{C}(r)) \|_2^2, \tag{2}
\]

where \(R\) is the set of rays in each training batch.

In our scenario, we adopt the structure of NeRF as backbone, but modify the color \(c = (r, g, b)\) to one-dimensional gray value as \(c = \text{(gray)}\) to represent edge intensities.

3.1.2 Neural Edge Fields

We introduce neural edge fields (NEF), training from 2D edge maps to represent the edge probability at every spatial positions. While NeRFs synthesize photorealistic novel views images based on the differentiable volume rendering, their volume density does not approximate the actual 3D shape very well. Similar cases also exist for NEF. The NEF density does not approximate the actual 3D edges. The range and scale of NEF density also varies from scene to scene, making the 3D edges difficult to extract from them. Recently, NeuS [32] and VolSDF [37] represent the object surface by the signed distance function (SDF) and mapping SDF to volume density of NeRFs by a distribution function. Similarly, we introduce an intermediate density field, called edge density, before NEF densities, as illustrated in Fig. 3. During training, with proper supervisions/constraints on edge densities, they are expected to approximate the 3D edges well. Edge density describes the edge probability at each position. It is in the range of \([0, 1]\), to be unified with 2D edges (1 represents edge and 0 is non-edge). After mapping functions, we can transform edge densities to NEF densities, which are used for volume rendering. Given \(x \in \mathbb{R}^3\) represent the space occupied by the object in \(R^3\), and \(E(x)\) represent the value of edge density in location \(x\), the NEF density \(\sigma\) is calculated by:

\[
\sigma(x) = \alpha(1 + e^{-g(E(x) - \beta)})^{-1}, \tag{3}
\]

where \(\alpha\) is a trainable parameter to control the density scale, \(\beta\) is the mean to control the function position, and \(g\) adjust the distribution around \([0, 1]\). The edge density is expected to adaptively match the distribution of NEF density, and should also be easily binarized by a unified threshold. Thus, to ensure a proper mapping from edge density to NEF density, we set \(\beta = 0.8\) and \(g = 10\) in all experiments, and \(\alpha\) is a trainable parameter. As illustrated in Fig 2, \(\alpha\) value varies from scene to scene. The edge density is obtained by adding an extra sigmoid layer after the original NeRF MLPs. We add another MLPs of 4 hidden layers with a size of 256 to predict the gray value \(c\) by edge densities and view directions. The network architecture is shown in Fig. 3.

3.1.3 Training NEFs

Training NEF as NeRF is problematic due to several problems. Firstly, 3D edges are similar to 3D skeletons of the object, and in volume rendering, samples are sparse along rays, making the network easily stuck in local optima. Secondly, 2D edge maps do not match the actual 3D wireframe, due to occlusions. Edges may not be visible in all
Figure 2. Examples of the transformation and variety trend of $\alpha$ in Eqn 3. As in (a), with edge density ranging from 0 to 1, we can adaptively match NEF density. Adapting to different scenes is controlled by the trainable parameter $\alpha$ as in (b).

views, leading to inconsistencies among views. We introduce the weighted mean squared error loss (W-MSE) and consistency loss to solve these two problems. Furthermore, to encourage the sparsity of points in NEF, we also introduce a sparsity loss.

With three loss designs, we are able to stably train NEF supervised by 2D images. The final loss function is represented as:

$$L = \lambda_1 L_{\text{color}} + \lambda_2 L_{\text{consistency}} + \lambda_3 L_{\text{sparsity}},$$

where the balancing parameters $\lambda_1, \lambda_2, \lambda_3$ are set to 1, 1 and 0.01 respectively in all experiments.

Once NEF is trained, we set a fixed threshold of 0.7 to binarize edge densities, to extract 3D edge points from NEF.

**W-MSE loss.** We obtain 2D edge maps from a lightweight edge detector PiDiNet [29]. On edge maps, edge pixels are in white color and often sparsely distributed. Therefore, when training NEF, edge and non-edge pixels are highly imbalanced, which leads to very sparse samples along rays. In this case, the network is easy to degenerate to local optima. A most common degeneration case is to predict all densities and colors are zeros, and the rendered images are all black. Therefore, we modify the original color loss by adding an adaptive weight $W(r)$ in each batch. The weighted mean squared error loss (W-MSE) is defined as:

$$L_{\text{color}} = \sum_{r \in R_i} W(r) \|(C(r) - \hat{C}(r))\|^2,$$

where $C^+$ and $C^-$ denote the number of edge and non-edge pixels in each batch decided by the threshold $\eta$. We set $\eta$ to 0.3 throughout the paper. The adaptive weight is simple yet effective by enforcing the network to focus more on edge pixels/rays, avoiding degenerating.

**Consistency loss.** The edge map of each view does not match the real 3D wireframe. On a 2D edge map, not all edges are visible due to occlusions. It means the “ground truths” are not exactly correct, missing those invisible edges in each view.

To successfully reconstruct 3D edge points from these 2D edge maps, we should recover the complete edge map of each view by integrating the information from other views where the occluded edges are visible, as shown in Fig 4. For each view, occluded edges are invisible in the image, as well as the edge map. NEF will be confused during training due to such inconsistencies among views. In each view, there are many false-negative pixels which on “ground truth” are non-edges. Such inconsistency brings noisy NEF around the object surface in spatial positions. For these occluded edges, the value of edge density is close to 1, but the color is close to 0 to fit those false-negative samples. Therefore, we enforce the value of edge density and color intensity (both are within $[0, 1]$) to be consistent for all samples along rays, to reduce the false-negative pixels. The consistency loss is also calculated by mean squared error, and defined as:

$$L_{\text{consistency}} = \sum_{r \in R_i} \|(E(r) - c(r))\|^2.$$
views. After training, the re-rendered edge maps by NEF successfully recover those invisible edges, as illustrated in Fig. 4. Thus, adopting different 2D edge detectors have limited influences on NEF reconstruction. With the consistency loss, NEF automatically corrects missing 2D edges, no matter they are occluded or missed by detectors (See Sec. 4.3).

**Sparsity loss.** As mentioned, edges are sparse in both 2D and 3D spaces. To encourage spatial sparsity, as well as accelerating the convergence, we add an extra regularizer, sparsity loss, to penalize unnecessary edge densities along rays of non-edge pixels during training. We adopt Cauchy loss [1] as the sparsity regularizer, which is highly robust to outliers. The sparsity loss is defined as:

$$L_{\text{sparsity}} = \sum_{i,j} \log \left( 1 + \frac{E(r_i(t_j))^2}{s} \right),$$

where $i$ indexes non-edge pixels of input edge maps, $j$ indexes samples along the corresponding rays, $s$ control the scale of the regularizer. We fix $s = 0.5$.

### 3.2. Extracting 3D Parametric Curves

With 3D edge point clouds from NEF, we further extract parametric curves. We extract Bézier curves from 3D edges in a coarse-to-fine manner. The objective of optimization is introduced in Sec. 3.2.1. The coarse-to-fine pipeline is introduced in Sec. 3.2.2.

#### 3.2.1 Bézier Curves Optimization

We adopt cubic Bézier curves to represent the geometrical shape of 3D edges. For each curve, we optimize the positions of four control points. The first and last control points define the beginning and ending positions, and the other two control points lead to different curvatures. Straight lines can be considered as linear Bézier curves with two control points. The goal is to optimize a set of parameters (positions of four control points) for all curves $\{\text{curve}_i\}_{i=1}^n = \{\{p^4_i\}_{j=1}^n\}_{i=1}^n$ to fit the 3D point cloud. The number of curves $n$ varies for different objects. To optimize the curve fitting, we sample 100 points on each curve, and dilate them up to 500 by adding Gaussian noise around them. We apply the widely-used Chamfer Distance (CD) to compute the distances between the curve points and 3D edge points. $P_c$ and $P_t$ represent the sampled points from curves and the target 3D edge points, respectively, the Chamfer loss is defined as:

$$L_{\text{CD}}(P_c, P_t) = \frac{1}{|P_c|} \sum_{x \in P_c} \min_{y \in P_t} \|x - y\|_2^2 + \frac{1}{|P_t|} \sum_{y \in P_t} \min_{x \in P_c} \|x - y\|_2^2,$$

where $\gamma$ is the parameter to control the tendency for each side ($\gamma = 1$ for original Chamfer loss). Each point in $P_c$ finds the closest point in $P_t$ (and vice-versa), and calculates the average pair-wise point-level distance. A bigger $\gamma$ means the optimization focus more on minimizing the distance from $P_t$ to $P_c$.

By minimizing CD, we fit Bézier curves to 3D edge points. However, the optimization of CD is insensitive to endpoint details, and we find that many curves are not connected. To encourage curve connections, we add a regularizer in the objective function, to encourage endpoints which are close in space to meet. Two endpoints of each Bézier curve are the first and the last control points, and the endpoints of all curves $\{\text{curve}_i\}_{i=1}^n$ are termed as $P_E = \{\{p^4_i\}_{j=1}^n\}_{i=1}^n$. The endpoint regularizer is defined as:

$$L_{\text{EP}} = \sum_{x,y \in P_E} M \|x - y\|_2^2,$$

in which

$$M = \begin{cases} 1, & \text{if } x - y \leq d, \\ 0, & \text{if } x - y > d, \end{cases}$$

$M$ is a mask to ensure the endpoints loss only regularize those endpoints that are close enough to each other (within the distance $d$). At last, the objective function to optimize all curves:

$$\arg\min_{\{p^4_i\}_{j=1}^n} (L_{\text{CD}} + \lambda L_{\text{EP}}),$$

where $\lambda$ is set to 0.01 in all experiments.

#### 3.2.2 Coarse-to-fine Scheme

The objective function of optimization is highly non-convex, making it easy for control points to converge to local minima. Therefore, the initialization of Bézier curves has a significant impact on the final result of the optimization. It is also difficult to select a proper number of curves that is suitable for all objects. Thus, we design a coarse-to-fine pipeline to extract curves. At coarse-level, we downgrade cubic Bézier to straight lines, and fit a set of lines to 3D edge points. At fine-level, we upgrade lines to cubic Bézier curves, and connect the endpoints of curves.
Coarse-level optimization. Instead of optimizing all lines simultaneously, we iteratively optimize the lines one by one with a fit-and-delete strategy. Specifically, during each iteration, we enlarge $\gamma$ to 5 in Chamfer loss (Eqn. 9), to encourage one line (linear curve) to fit partial 3D edge points to the utmost. After one line is decided, we delete the 3D edge points around the line and record its parameters. The process continues until there are not many 3D edge points left (i.e., <20). The group of lines fitted works as initialization to the fine level. We demonstrate the process of coarse-level optimization in Fig 5. Since the lines are fitted one by one, we do not consider endpoints regularization at this level.

Fine-level optimization. Coarse-level optimization initializes the number of lines, as well as the beginning and ending positions. In fine-level, we upgrade all straight lines back to cubic Bézier curves by interpolating another two control points between the endpoints pair, and solve the optimization in Eqn. 12. The resulting parametric curves precisely match the 3D edge points as demonstrated in Fig. 6.

4. Experiments

We compare with state-of-the-art methods and conduct ablations on the contributed ABC-NEF dataset. More experiments, discussions, training details and video demos are in the Supplementary.

4.1. ABC-NEF Dataset.

As previous works [18, 19, 33], we conduct experiments on the ABC dataset [14] which consists of more than one million CAD models with edge annotations. To evaluate our pipeline, we provide a dataset called ABC-NEF, consisting of 115 distinct and challenging CAD models. They include various types of surfaces and curves, from the first chunk of the ABC dataset. We adopt BlenderProc [7] to render posed images facing the center of the object. We sample 50 views of 800 × 800 image rays for each object. The 50 views are sampled by evenly placing cameras on a sphere by Fibonacci sampling [12]. Statistical analysis of ABC-NEF dataset and ablations about the number of views are included in the supplementary material.

4.2. Comparisons with state-of-the-arts

Comparison Settings. We compare the proposed method with three most state-of-the-art data-driven methods of parametric curve reconstruction, PIE-NET [33], PC2WF [18] and DEF [19]. All three methods require point clouds as inputs, while ours require only 2D images. Following the settings in their papers, for PIE-NET, we apply the farthest point sampling method to uniformly sample 8096 clean point clouds that represent the object shape as the inputs, the outputs of PIE-NET contains closed and open curves. For PC2WF, we sub-sample 200,000 points for each object from surface meshes as the inputs, it outputs pairs of endpoints representing straight lines. For DEF, the inputs contain 128 views of depth maps and point clouds. In DEF, depth maps are collected to build a distance-to-feature field, which is used to detect corners on point clouds and extract spline curves.

We adopt their pre-trained models to reconstruct parametric curves for evaluation. Since PC2WF is designed to detect straight lines, we also make comparisons on a subset of the proposed ABC-NEF which contains 26 CAD models of only lines, named ABC-NEF-Line.

Evaluation Metrics. We sample points on reconstructed parametric curves and evaluate distances between the sampled points and ground truth edge points. To ensure the
points are evenly distributed, we down-sample the points on voxel grids, so there is at most one point per voxel.

To measure the location of reconstructed 3D edges, we adopt the Intersection over Union (IoU), precision, recall and their F-score. However, a small shift between two point clouds may lead to large changes on above metrics. We also adopt the Chamfer Distance (CD) between point clouds, to measure the geometric accuracy of reconstructed parametric curves. A small shift between point clouds would not affect much for CD. Before comparison, we normalize and align all ground truth edges and curve predictions into the range of $[0, 1]$. After normalization, when evaluating IoU, precision, recall and F-score, points are considered matched if there exists at least one ground truth point with the L2 distance smaller than 0.02.

Comparison Results. As reported in Table 1, our self-supervised method with only 2D supervisions significantly outperforms other state-of-the-arts in all metrics and datasets. We observe that PIE-NET and PC2WF both achieve much higher precision than recall, which means they often miss curves, but the detected curves locate precisely. Although PC2WF is designed to detect straight lines, we notice on ABC-NEF-Line dataset, our method still achieves better performance.

We illustrate qualitative performance in Fig. 6. The results show that PIE-NET and DEF can detect and locate most curves well, and PC2WF is proficient in reconstructing line structures. However, limited by the design, PC2WF can only detect lines and struggles in capturing any other types of curves. Since PIE-NET is exactly trained on sharp features, leading to incompetent performance around ellipse edges and areas with relatively weak curvature. Meanwhile, DEF reconstructs curves mainly based on a continuous and smooth distance-to-feature field, thus has trouble in discriminating close curves and tends to incorrectly connect adjacent curves. We also notice these methods reconstruct curves heavily relying on corner detection, thus failing to cover all edges if some corners are missed. Essentially, these data-driven methods, may suffer from reconstructing curves for out-of-distribution shapes. On the contrary, our method benefits from the self-supervised pipeline, and can be trained on natural images. More comparisons and discussions are in the supplementary material.

4.3. Ablation Studies

We perform ablation studies to verify the inclusion of each loss and design. W-MSE loss in Eqn. 5 is essential to learn the NEF due to the imbalanced edge and non-edge pixels (rays). Without W-MSE loss, the training of NEF would suffer from degenerating of predicting all-zero fields. Therefore, we take NEF with W-MSE loss as the baseline version, and evaluate sparsity and consistency losses.

For better visualization, we compare the quality of edge densities by illustrating the rendered depth maps. Since a depth map is essentially rendered by accumulated NEF densities along rays, and exactly conveys the spatial distribution of edge density. As shown in Fig. 7, the network may generate random noisy densities in the scene without sparsity regularization. Without consistency loss, the network is trained to fit the incorrect “ground truth”, missing occluded edges, thus overfit 2D edges in each view and fails to reconstruct consistent 3D edges.

After getting 3D edge points, we reconstruct paramet-
ric curves in a coarse-to-fine manner. We demonstrate the necessity of our designs by removing each part individually. We show the optimization results with and without the coarse-level initialization, the line-to-curve strategy, and the endpoint loss in Fig. 8. Quantitative results of the selected samples are shown in Table 2. The curves are quite noisy without the coarse-level initialization. In the coarse-level, if we try to fit cubic Bézier curves without initializing from straight lines, one cubic Bézier curve may fit multiple connected straight lines. Therefore, without the line-to-curve strategy, the total number of curves may be insufficient for global optimization and further influence the endpoints loss. The endpoints loss works to refine all curves to be compactly connected. With all designed strategies, our full results are clean, compact and fit the geometrical shapes.

<table>
<thead>
<tr>
<th>Method</th>
<th>CD $\downarrow$</th>
<th>F-score $\uparrow$</th>
<th>IoU $\uparrow$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Without Initialization</td>
<td>0.0734</td>
<td>0.5016</td>
<td>0.3216</td>
</tr>
<tr>
<td>Without Line-to-curve</td>
<td>0.0202</td>
<td>0.9715</td>
<td>0.9387</td>
</tr>
<tr>
<td>Without Endpoints Loss</td>
<td>0.02</td>
<td>0.9805</td>
<td>0.959</td>
</tr>
<tr>
<td>Full result</td>
<td><strong>0.0189</strong></td>
<td><strong>0.9935</strong></td>
<td><strong>0.9851</strong></td>
</tr>
</tbody>
</table>

Table 2. Quantitative results of data in Fig. 8. Initialization improves performance on all metrics significantly. Although the line-to-curve strategy and the endpoints loss seem to bring little improvement, they help to refine the curves to match the real geometrical shape and to be visually plausible.

We also conduct ablations on other edge detector (i.e. Canny), noisy 2D edge maps detected on blurred images (Gaussian Blur with a $9 \times 9$ kernel size), and randomly dropout 30% and 50% image (edge) pixels for all views to test the robustness of our method. As in Fig. 9, all alternatives perform reasonably. Even if edge maps are badly broken, it still restores the rough 3D shape.

**4.4. Real-world Scene**

We also test the performance of NEF for several collected toys with sharp geometry in the real-world scene. We took a video surrounding and looking at the target toys, and cut about 60 frames as the input. We apply COLMAP [26], a well-known structure-from-motion (SFM) solver, to estimate the camera poses for the input images. We still apply the pre-trained PiDiNet [29] to extract 2D edge maps, train the NEF, and reconstruct curves from extracted edge points. The process is illustrated in Fig. 10. The reconstructed results show the potential of our method to extract 3D edge points and reconstruct parametric curves in real-world scenes, even with camera poses that are not completely correct.

**5. Conclusions**

We presented the first self-supervised pipeline for 3D parametric curve reconstruction by learning a neural edge field. Self-supervised by only 2D supervisions, our method achieves comparable and even better curve reconstruction than alternatives taking clean and complete point clouds as inputs. Our method shows the potential of generalization ability and leveraging advantages of multi-modal information. The method has limitations in dealing with textured objects, edges inside the objects, and the network architecture would be optimized to be simpler. More discussions are in the Supplementary.

**Acknowledgements:** We thank the anonymous reviewers for their valuable comments. This work is supported in part by the National Key Research and Development Program of China (2018AAA0102200), NSFC (62132021, 62002375, 62002376), Natural Science Foundation of Hunan Province of China (2021JJ40696, 2021RC3071, 2022RC1104, 2022RC3061) and NUDT Research Grants (ZK19-30, ZK22-52).
References


