NEURAL FACE IDENTIFICATION IN A 2D WIREFRAME PROJECTION OF A MANIFOLD OBJECT

Anonymous authors
Paper under double-blind review

ABSTRACT

In computer-aided design (CAD) systems, 2D line drawings are commonly used to illustrate 3D object designs. To reconstruct the 3D models depicted by a single 2D line drawing, an important key is finding the edge loops in the line drawing which correspond to the actual faces of the 3D object. In this paper, we approach the classical problem of face identification from a novel data-driven point of view. We cast it as a sequence generation problem: starting from an arbitrary edge, we adopt a variant of the popular Transformer model to predict the edges associated with the same face in a natural order. This allows us to avoid searching the space of all possible edges loops with various hand-crafted rules and heuristics as most existing methods do, deal with challenging cases such as curved surfaces and nested edge loops, and leverage additional cues such as face types. We further discuss how possibly imperfect predictions can be used for 3D object reconstruction.

1 INTRODUCTION

In this paper, we revisit a classical problem in computer-aided design (CAD), namely the conversion of 2D line drawings into 3D objects. In a traditional CAD pipeline, design engineers commonly draw a 2D wireframe[1] of the desired object when creating their ideas and when communicating their ideas to others. Therefore, there is a strong need to develop algorithms which can convert 2D line drawings into 3D solid models for analysis, simulation, and manufacturing.

An important step of 3D reconstruction from 2D line drawings is face identification, that is, finding loops of the edges which correspond to faces of the 3D object (Figure 1). If the correct face configuration of an object can be obtained, the number of degrees of freedom in its reconstruction will be greatly reduced. Many studies have been performed to achieve this goal, and in certain cases, a satisfactory solution exists. For example, it is well known that there is a unique planar embedding when the object has genus 0 and the drawing is 3-connected, from which the set of faces can be determined (Shpitalni & Lipson, 1996a). However, for a general manifold object with complex geometry (*e.g.*, curved surfaces) and topology (*e.g.*, high genus), the task remains difficult.

A close look at current methods reveals two primary sources of challenges. The first one is the algorithmic complexity. As no analytical solution is available for a general manifold object, these methods need to search through the space of all possible face loops, which grows exponentially with the number of vertices. In order to make the solution efficient, much effort has been made to design heuristic search algorithms (Liu et al., 2002; Liu & Tang, 2005; Varley & Company, 2010). The second one is the inherent ambiguity in reconstructing 3D objects from 2D projections. For example, it is impossible for a topological algorithm to tell if two nested cycles (Figure 1(a)) are coplanar, thus are two loops of the same face. Besides, a single 2D projection may yield multiple topologically correct solutions (Figure 1(b)). In such cases, additional heuristics (*e.g.*, number of cycles, geometric regularities) or human intervention is required to choose the final output.

Nevertheless, humans can effortlessly "see" the real faces in a typical line drawing, including those in Figure 1. Rather than performing a tedious search for the face loops, our ability to quickly identify the faces seems to be attributed to our past experience interacting with 3D objects. This makes us wonder: Is it possible for a computer to learn to recognize faces in a data-driven fashion? This work represents a first attempt to answer the question. We train a deep neural network to detect

[1]We use the terms "wireframe" and "line drawing" interchangeably in this paper.

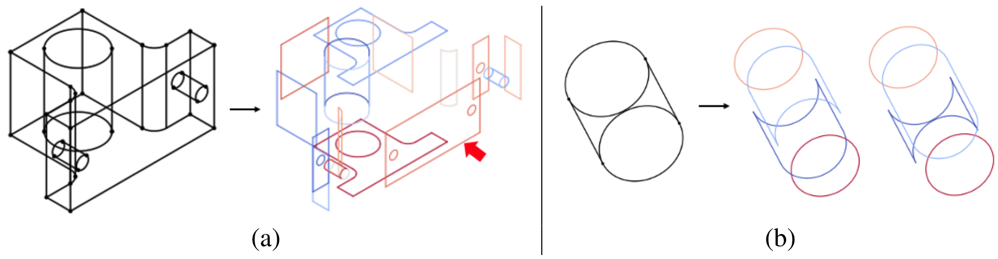


Figure 1: Given a 2D line drawing, face identification aims to find edge loops which correspond to actual faces of the 3D object. (a) A case where several faces (including the one indicated by a red arrow) are enclosed by two or more loops. (b) A case where multiple topologically correct interpretations exist.

faces using a large collection of 3D objects and their 2D projections. This way, we are able to avoid exploring a search space of exponential order. Moreover, by learning from ground truth 3D data, our method implicitly learns to generate the most plausible solutions, resolving the inherent ambiguity associated with the problem.

To this end, we cast face identification as a sequence generation problem, leveraging the natural order of co-edges in a face loop (see Section 3 for a detailed explanation). Given a set of co-edges in a 2D line drawing, we train a variant of the popular Transformer model (Vaswani et al., 2017) to predict one co-edge index that forms the face at one timestamp. For each detected face, we further classify it into different types such as planes and cylindrical surfaces. On the public ABC dataset (Koch et al., 2019), our model achieves 93.8% and 95.9% in precision and recall, respectively. Finally, with the detected faces, we develop a simple convex optimization scheme to reconstruct structured 3D models from a single 2D line drawing.

In summary, the **main contributions** of this work are two-fold: (i) We propose a first data-driven approach to face identification and study its advantages and disadvantages over existing geometry- and topology-based methods. (ii) We develop a simple scheme to reconstruct a 3D model from a single 2D line drawing using the face identification results. With this work, we discuss new opportunities for incorporating learning-based approaches into established CAD pipelines, such as identifying conflicting face loops in a geometric constraint system for 3D reconstruction.

2 Related Work

Face identification is a long-standing problem in the area of automatic interpretation of line drawings. Research in this field dates back to the '70s and '80s. Early work (Markowsky & Wesley, 1980) proposes a geometric approach to detecting planar faces in a 3D wireframe: it first generates possible planes at a vertex joined by two non-collinear edges, then searches for other vertices lying on each such plane and tries to use the vertices to form cycles. Subsequent work (Abbasinejad et al., 2011; Zhuang et al., 2013) can deal with more general curve networks with highly curved faces and complex topology (*i.e.*, manifolds with high genus). But these methods rely on 3D coordinates of the wireframe, therefore cannot be applied to 2D projections.

Another line of work addresses the problem from a graph-topological point of view. To find the faces of a vertex-edge graph, Hanrahan (1982); Dutton & Brigham (1983) propose to compute the planar embedding of the graph, where the resulting regions represent the faces. Ganter & Uicker (1983); Brewer & Courter (1986) generate an initial cycle basis from the spanning tree of the graph, then perform a cycle reduction procedure to find the faces. But these methods are only suitable for genus-0 manifolds. For manifold objects of a non-zero genus or non-manifold objects, Shpitalni & Lipson (1996a) present a method which searches through the set of all possible sets of face loops and use various geometric criteria to assess them. The method could be slow as the number of possible face loops is $O(e^v)$ and any topological algorithm for finding faces of objects of genus > 0 has exponential complexity (Bagali & Waggenspack, 1995). To reduce the search complexity, Liu et al. (2002) propose a depth-first search that is primarily guided by topological constraints; Liu & Tang (2005) develop a genetic algorithm which uses 2D geometric information in its assessment criteria; and Varley & Company (2010) use a heuristic search based on the shortest-path and Dijkstra's algorithm.

Given the face topology, several studies (Shpitalni & Lipson, 1996b; Liu et al., 2008; Wang et al., 2009) attempt to reconstruct the 3D object from a single 2D line drawing. These works solve for the 3D shape in an optimization framework, using various constraints based on structural regularities, such as minimum standard deviation of angles (MSDA), face planarity, line parallelism, and corner orthogonality. In this work, we introduce a different method that makes use of the predicted face types – information previous methods do not have access to. Further, we study the impact of imperfect face detection results on 3D reconstruction.

Finally, our technical approach is inspired by the recent advance in sequence-to-sequence modeling (Bahdanau et al., 2015; Vaswani et al., 2017), which has produced state-of-the-art results in a wide range of NLP and vision tasks lately. Specifically, our network design follows Pointer Net (Vinyals et al., 2015), which proposes an autoregressive model to generate a distribution on a given input data set. Recently, PolyGen (Nash et al., 2020) extends this idea to generate 3D polygon meshes by sequentially predicting the vertices and faces using a Transformer-based architecture.

3 Problem Formulation

In this paper, a 2D wireframe projection is assumed to be an orthographic projection where all the edges (including silhouettes) and vertices of the object are visible. We make a few assumptions about the input line drawing: *First*, the hidden lines and vertices are given. *Second*, the crossing point of two edges in a line drawing is not vertex and cannot be used to form faces. As such, the input line drawing can be represented as an edge-vertex graph $\mathcal{G} = (V, E)$, where each edge (or vertex) of the graph corresponds to exactly one edge (or vertex) of the object. Note that the graph may contain one or more connected components. In practice, these graphs may be a result of previous processing of a rough sketch or a scanned-in drawing (Shpitalni & Lipson, 1997).

Given the graph $\mathcal{G} = (V, E)$, the goal of face identification is to find all faces $F = \{f_1, \dots, f_M\}$ of the object, where each face can be written as the set of enclosing edges: $f_i = \{e_{i_1}, \dots, e_{i_n}\}$.

Same as prior work (Liu et al., 2002; Varley & Company, 2010), we focus on *manifold objects* only. A manifold object is defined as a solid where every point on its surface has a neighborhood topologically equivalent to an open disk in the 2D Euclidean space. A key property of manifold objects is that each edge of a manifold is shared exactly by two faces. This property is best expressed in terms of *co-edges*, $C = \{c_1, c_2, \dots\}$, an important type of topological entities in the B-rep. As illustrated in Figure 2, there is a co-edge pointing from vertex v_i to v_j if and only if there is an edge connecting v_i and v_j. For example, c_3 and c_{12} are two mutually mating co-edges associated with edge e_8.

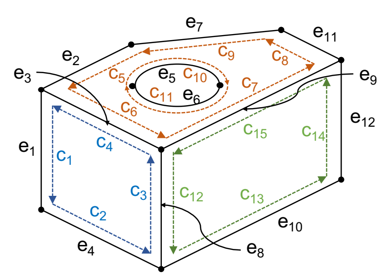


Figure 2: B-rep of an object. Hidden lines are omitted in this example.

A face can be conveniently represented as a loop (*i.e.*, closed path) of co-edges. For example, in Figure 2, the loops (c_1, c_2, c_3, c_4) and $(c_5, c_6, c_7, c_8, c_9, c_{10}, c_{11})$ represent two faces of the object. We follow the conventional definition of loop direction: if a loop is viewed along its direction, with the face normal pointing upwards, then the face that owns the loop is to the left.

4 Face Identification via Sequence Generation

The natural order of co-edges corresponding to a face motivates us to treat face identification as a sequence generation problem. Specifically, a face with n co-edges can be written as: $f_i = (c_{i_1}, \dots, c_{i_n})$, where each index i_t, $t = 1, 2, \dots, n$, is an integer between 1 and N, and N is the total number of co-edges. Thus, starting with an arbitrary co-edge c_{i_1}, our goal is to grow it into a sequence of co-edges $(c_{i_1}, \dots, c_{i_n})$ on which c_{i_1} lies, as shown in Figure 3. To detect all the faces

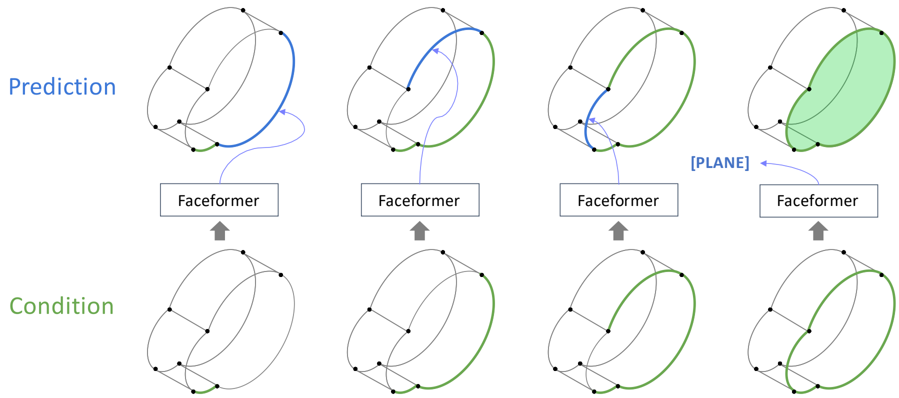


Figure 3: Our model, Faceformer, takes as input the set of all co-edges, and current sequence of co-edge indices, and outputs a distribution over the co-edge indices.

$F = \{f_1, \ldots, f_M\}$, we may use every co-edge in C as the starting co-edge and repeat the process for N times.

In the following, we describe how to generate a single face f_i given the starting co-edge c_{i_1}.

4.1 Face Identification Model

Since our goal is to select a subset of the input co-edges to form the output face, we first briefly review Pointer Net (Vinyals et al., 2015), a sequence-to-sequence (seq2seq) model which uses the attention mechanism to create pointers to input elements. Given the input P, the main idea is to learn the conditional probability $p(\mathcal{I} \mid P)$ using a parametric model, such as LSTM (Hochreiter & Schmidhuber, 1997) or Transformer (Vaswani et al., 2017), to estimate the terms of the probability chain rule: $p(\mathcal{I} \mid P) = \prod_{t=1}^{T} p(i_t \mid i_1, \ldots, i_{t-1}, P)$. Here, $P = \{\mathbf{p}_1, \mathbf{p}_2, \ldots\}$ is a sequence of vectors and $\mathcal{I} = (i_1, \ldots, i_T)$ is a sequence of T indices, each between 1 and $|P|$.

Similar to most seq2seq models, Pointer Net adopts an encoder-decoder architecture. First, it obtains a contextual embedding $\mathbf{w}_k$ for each input using an encoder. At each decoding time step t, the decoder outputs a pointer vector $\mathbf{u}_t$, which is then compared to the input embeddings via a dot-product. The resulting scores are normalized using a softmax to form a valid distribution over the input set:

$$\{\mathbf{w}_k\}_{k=1}^{N} = \textsc{Encoder}(P; \theta) \tag{1}$$

$$\mathbf{u}_t = \textsc{Decoder}(\mathcal{I}_{<t}, P; \theta) \tag{2}$$

$$p(i_t = k \mid \mathcal{I}_{<t}, P; \theta) = \mathrm{softmax}_k(\mathbf{u}_t^T \mathbf{w}_k) \tag{3}$$

Finally, the parameters of the model are learned by maximizing the conditional probabilities on a training set.

Input sequence and embeddings. In our problem, the input sequence consists of all co-edges C. To further leverage geometric cues, we propose to classify the faces into different types – a benefit of data-driven approaches. In this work we only consider two special face types, namely planar surface and cylinder surface, but the method can be easily extended to other types. To this end, we add three special tokens, `[PLANE]`, `[CYLINDER]` and `[OTHERS]`, to indicate different face types. We replace regular stop tokens with the face type tokens, expecting the network to predict one of the three tokens immediately after generating all co-edges of the face.

We jointly embed the special tokens with the co-edges, to obtain a total of $N + 3$ input embeddings. We use two embeddings for each input co-edge, including (i) value embedding, representing the coordinate value of the edge, and (ii) position embedding, indicating the token location in the sequence. Due to the varying edge length, we uniformly sample a fixed number of edge points to represent the co-edge as in Das et al. (2020). The points are ordered based on the co-edge direction. Then, we flatten the edge points and apply two linear layers to obtain a 512-dimensional embedding.

Output sequence and embeddings. As mentioned above, we represent each face as a sequence of co-edges $f_i = (c_{i_1}, \ldots, c_{i_n})$. A face type token is added (i) to predict the face type and (ii) to indicate the end of the sequence. When a face consists of multiple loops, special care needs to be taken to ensure a unique co-edge sequence. In such cases, except for the loop to which the starting co-edge belongs, we order the co-edges in each other loop from lowest to highest first by its x-coordinate, followed by y-coordinate. Take Figure 2 as an example, if we use c_{11} as the starting co-edge, the desired face sequence should be $(c_{11}, c_{10}, c_5, c_6, c_7, c_8, c_9)$.

Similar to input embeddings of the encoder, we use learned position and value embeddings for the inputs to the decoder. We use the co-edge's contextual embedding from the encoder output as its value embedding.

Network architecture. Inspired by recent success of Transformer-based architectures (Vaswani et al., 2017), we adopt it as the basic block of our face identification model. Given the input co-edge embeddings, the encoder encodes them into contextual embeddings, then the decoder outputs pointers based on the contextual embeddings and the decoder input. The model consists of 6 Transformer encoder and decoder layers, with a feed-forward dimension of 1024 and 8 attention heads.

4.2 Training and Inference

We implement our model with PyTorch (Paszke et al., 2017). We use Adam optimizer (Kingma & Ba, 2015) with a learning rate of 1×10^{-3}. The batch size is set to 4. The model is trained with $400\,000$ iterations, taking about 30 hours to converge on a single NVIDIA RTX 3090 GPU device.

At inference time, since our parallel model predicts faces from all co-edges independently, there are many duplicated faces in the result. We take three post-processing steps to remove invalid and duplicated predictions. First, because a face is defined as a closed path of co-edges, we filter out any unclosed face predictions. Second, we remove predictions that contain both co-edges associated with the same edge, as this never happens in real faces. Finally, we identify duplicated predictions by comparing the set of edges generated in each face prediction. Note that these duplicated predictions may have different face type classifications. We count the number of times each face type is predicted, and take the face type of the highest count as its predicted face type.

4.3 Experiments

4.3.1 Experimental Setup

Dataset. We build a benchmark for this novel task using a subset of CAD mechanical models from ABC dataset (Koch et al., 2019). We use pythonOCC[2], a Python 3D development framework built upon the Open CASCADE Technology, to project CAD models into 2D line drawings. We first normalize the shape such that the half diagonal length of the bounding box is equal to 1. Then, the camera is placed at a random distance between 1.25 to 1.5 away from the object center, pointing towards the object. The viewpoints are randomly sampled on a hemisphere. The dataset consists of 9370/202/504 samples for training/validation/testing.

To eliminate cases that are unnecessarily complicated, we filter out shapes with more than 42 faces or 37 edges in a face. Since the ABC dataset has many duplicate shapes, we also run additional filters based on the shape's topology and three orthogonal views to remove duplicates.

Evaluation metrics. We develop several metrics to evaluate the performance of our model.

First, we compute the *prediction accuracy* at the co-edge sequence level. For each starting co-edge, the sequence prediction is regarded as correct if it exactly matches a ground truth sequence. Then, for an object with N co-edges, the average accuracy can be computed as $\text{acc} = \frac{1}{N}\sum_k \mathbf{1}(\mathcal{I}_k = \mathcal{I}_k^*)$.

Second, we compute the *precision* and *recall* at the face level. We treat each face as an unordered set of edges. For our model, this is the result after the post-processing step as described in Section 4.2. A prediction f is said to match a ground truth face f^* if and only if the two sets are equal. Then, let F^* denote the set of ground truth faces, and F denote the set of faces detected by any method, the precision and recall are defined as: $\text{precision} = |F \bigcap F^*|/|F|$, $\text{recall} = |F \bigcap F^*|/|F^*|$.

[2] https://github.com/tpaviot/pythonocc

Table 1: Experiment results on network design.

models	precision	recall	runtime (s)
seq2seq	77.6	76.8	2.72
seq2seq + co-edge	**94.9**	90.5	2.84
ours (seq2seq + co-edge + parallel)	93.8	**95.9**	**0.44**

Table 2: Experiment results on input data.

input data	pred. accuracy	precision	recall	cls. rate
ours	92.1	93.8	95.9	97.5
ours - perspective	92.3	93.6	96.2	97.6
ours - fixed viewpoint	95.6	95.9	97.3	97.6

Finally, we report the *face classification rate*, which is the percentage of correctly classified faces among all faces correctly detected by our model.

4.3.2 Experimental Results

Experiment on network design. In this experiment, we compare our model with two variants of it to illustrate the benefits of (i) using co-edges and (ii) face-wise parallel prediction.

The first variant, seq2seq, directly operates on the edges (instead of co-edges) and generates all faces in a sequential fashion. In this variant, the input sequence would be the set of all edges, ordered from lowest to highest first by its x-coordinate, then by y-coordinate. For the output sequence, we introduce three additional special tokens, `[SOS]`, `[EOS]`, and `[SEP]`, to indicate the start and end of sequence, and the separation between faces, respectively. During training, we order faces according to the edge indices and then concatenate all faces into a single sequence. Take Figure 2 as an example, the desired output sequence would be $\{\mathtt{SOS}, \mathtt{e}_1, \mathtt{e}_3, \mathtt{e}_4, \mathtt{e}_8, \mathtt{SEP}, \mathtt{e}_2, \mathtt{e}_3, \mathtt{e}_5, \mathtt{e}_6, \mathtt{e}_7, \mathtt{e}_9, \mathtt{e}_{11}, \mathtt{SEP}, \mathtt{e}_8, \mathtt{e}_9, \mathtt{e}_{10}, \mathtt{e}_{12}, \mathtt{EOS}\}$.

The second variant, seq2seq + co-edge, also generates all faces sequentially but uses co-edges. In this variant, the input sequence would include all the co-edges and three special tokens `[SOS]`, `[EOS]`, and `[SEP]`. We order the faces in the same way as the first variant. Thus, the desired output sequence for the example in Figure 2 is $\{\mathtt{SOS}, \mathtt{c}_1, \mathtt{c}_2, \mathtt{c}_3, \mathtt{c}_4, \mathtt{SEP}, \mathtt{c}_5, \mathtt{c}_6, \mathtt{c}_7, \mathtt{c}_8, \mathtt{c}_9, \mathtt{c}_{10}, \mathtt{c}_{11}, \mathtt{SEP}, \mathtt{c}_{12}, \mathtt{c}_{13}, \mathtt{c}_{14}, \mathtt{c}_{15}, \mathtt{EOS}\}$.

Table 1 shows the quantitative results on ABC dataset. The seq2seq model using edges performs substantially worse because (i) unlike co-edges, edges do not encode directional information, so the model cannot take advantage of the natural order of face loops; and (ii) an edge is shared by two faces, so there is ambiguity about which face to predict given an edge. Comparing the two models using co-edges, the one with parallel prediction has slightly lower precision (-1.1%) but higher recall ($+5.4\%$). This is because it makes a prediction starting from each co-edge, thus would cover each face several times. In contrast, the model with sequential prediction only generates each face once. We also record the average runtime of the models for one shape in the test set. As one can see in Table 1, the model with parallel prediction is much faster.

Experiment on input data. In the next experiment, we study the performance of our model with different types of input data. First, we replace orthographic projection with perspective projection when generating the 2D line drawings. As shown in Table 2, this change has very small impact on all the metrics. Second, we fix the camera viewpoint to generate an isometric drawing for each shape. In CAD, an isometric view is commonly used to reveal as much information about the 3D shape as possible, and to avoid situations where the object's edges or vertices coincide (or appear as joined) accidentally. Therefore, such a viewpoint is considered easier than random viewpoints. By employing a fixed viewpoint, our model achieves even higher accuracies (Table 2).

Qualitative results. Figure 4 visualizes the incorrect predictions made by our model for various shapes. Some common problems are (i) incomplete prediction when a face consists of multiple

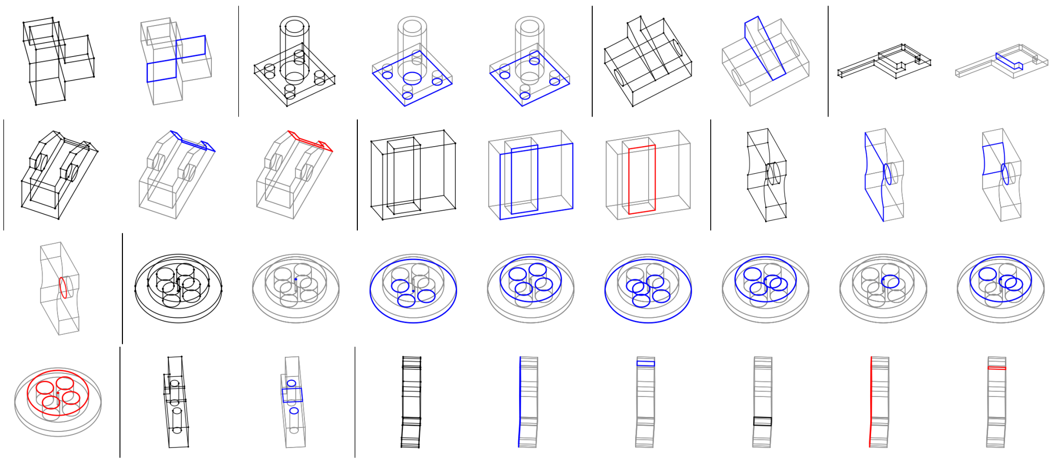


Figure 4: Incorrect predictions made by our face identification model. For each case, we show the input line drawing, followed by predictions with wrong face loops (blue), predictions with wrong face types (black), and missed faces (red).

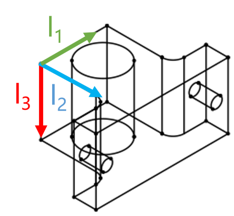
nested edge loops, and (ii) grouping edges or loops which belong to different faces. But as we will see in the next section, these errors have varying impact on the reconstruction of the 3D models.

5 3D Object Reconstruction

We now tackle the problem of 3D reconstruction with the predicted face loops and types. In the literature, this is often formulated as recovering the missing depths of the vertices of a line drawing. While knowledge about face topology significantly reduces the number of degrees of freedom, such information itself is not enough to uniquely determine the 3D geometry. Prior work (Shpitalni & Lipson, 1996b; Liu et al., 2008; Wang et al., 2009) resort to additional optimization criteria such as MSDA, face planarity, line parallelism, and corner orthogonality. We observe a few problems with these approaches: (i) the criteria are designed to emulate the human perception of a 2D line drawing as a 3D object, but it is uncommon for them to be violated in practice; (ii) the search for optimal solutions could get stuck at local minimum; (iii) they do not use information about face types.

Our primary goal is to build 3D models from the output of a deep face identification model – something never considered in prior work. Thus, we prefer a solution which decouples the impact of mistakes made by the network from that of other constraints or algorithms employed.

To this end, we develop a simple method that relies on only one type of constraints: line parallelism. As shown in the figure to the right, we assume that there are three mutually orthogonal directions $\{\mathbf{l}_1, \mathbf{l}_2, \mathbf{l}_3\}$ in the 3D scene, *i.e.*, $\mathbf{l}_j^T \mathbf{l}_k = 0, \forall j \neq k$. While these directions are provided *a priori* in this work, techniques for automatic estimation exist: for orthographic projections, these directions can be found by grouping parallel lines in a line drawing; for perspective projections, these directions correspond to the three dominant vanishing points. Afterward, the 3D directions $\{\mathbf{l}_1, \mathbf{l}_2, \mathbf{l}_3\}$ can be obtained via camera calibration (Hartley & Zisserman, 2000).



With this assumption, we are able to align the faces to the dominant directions according to their enclosing edges and solve for the 3D geometry via convex optimization. Below we first describe our method for objects with planar faces, then extend it to curved surfaces.

5.1 3D Reconstruction of Planar Objects

Suppose we are given an object with M faces $F = \{f_1, \ldots, f_M\}$ and L vertices $V = \{v_1, \ldots, v_L\}$. For each vertex, we write $v_l = [x_l, y_l, z_l]^T$, where z_l is the unknown depth of v_l in the camera coordinate system.

If the object is planar, then each face f_i can be represented by the plane equation $a_i x + b_i y + z + c_i = 0$. If a vertex v_l lies on two or more faces, for each pair of faces, say (f_1, f_2), we have $a_1 x_l + b_1 y_l + c_1 - a_2 x_l - b_2 y_l - c_2 = 0$. For all vertices in V, we can write similar constraints for pairs of faces. Rewriting all these linear equations in matrix form, we have: $\mathbf{P}_1 \mathbf{f} = 0$, where $\mathbf{f} = [a_1, b_1, c_1, \ldots, a_M, b_M, c_M]^T$ is a vector consisting of all the parameters of the faces.

For each dominant direction $\mathbf{l}_j, j = \{1, 2, 3\}$, we can identify all edges in the line drawing which are parallel to it, thus also find the faces which align with it. For such a face, say f_1, we have $a_1 l_x^j + b_1 l_y^j + l_z^j = 0$. Rewriting all these linear equations in matrix form, we have: $\mathbf{P}_2 \mathbf{f} = 0$.

Combining the above constraints, we can find $\mathbf{f}$ by solving a convex optimization problem:

$$\min_{\mathbf{f}} \quad \|\mathbf{P}\mathbf{f}\|_1, \quad \text{s.t.} \quad z_l = -(a_i x_l + b_i y_l + c_i) > 0, \forall \mathbf{v}_l \text{ on } f_i. \tag{4}$$

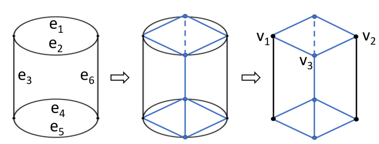
Here, the constraints enforce that the 3D vertices lie in front of the camera.

5.2 Handling Curved Surfaces

To deal with curved surfaces, we follow the approach proposed in (Wang et al., 2009). The main idea is to replace a curve with straight line segments so that the object becomes a polyhedron. Then, methods for planar objects, such as the one described in Section 5.1, can be applied. Finally, the curved surface is recovered by fitting general Bézier curves to the corresponding 3D vertices.

In our case, because the face types are known, the original approach can be substantially simplified, as we (i) do not need to employ an algorithm to distinguish between curved and planar faces; (ii) can develop face approximation and fitting methods for each specific face type.

The figure to the right shows an example of converting a curved surface on a cylinder into planar faces. For face (e_2, e_3, e_5, e_6), we find the singular points on curves e_2 and e_5 and then replace each curve with two straight lines. Once the 3D geometry of the polyhedron is reconstructed, the original curves can be recovered by fitting a 3D circle to the three vertices (*e.g.*, v_1, v_2, and v_3 in the figure).



5.3 Experiments

Before presenting the 3D reconstruction results based on the predicted faces, we point out that in rare cases ($< 5\%$) the reconstruction may not be perfect even with ground truth faces, due to the simple assumption (*i.e.*, line parallelism) we employ in the pipeline. The figure to the right shows two examples where major parts of the model do not align with any dominant direction, making the reconstruction under-constrained. As a result, the 3D model may appear to be distorted (first row) or partly incorrect (second row). In practice, these may be addressed by introducing additional constraints. But we choose to keep the reconstruction method simple so that the impact of errors in face identification can be better analyzed.

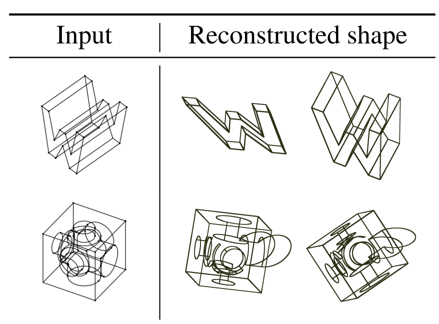


Figure 5 shows 3D reconstruction results for various shapes using the predicted faces. In the *first row*, we show two cases in which all faces are correctly identified and the 3D model is fully reconstructed. As a comparison, we also train AtlasNet (Groueix et al., 2018), a popular deep learning methods for single-view 3D reconstruction, on our dataset and include the test results in Figure 5. Unlike our method, most existing deep learning methods take an image as input and generate unstructured point cloud or meshes in an end-to-end fashion. As one can see, such a method struggles to learn with the wireframe inputs because the features are very sparse when treated as an image. In contrast, we propose to detect structures (*i.e.*, face topology) in the wireframe and use geometric reasoning for 3D inference. Therefore, our method is able to output clean, structured 3D models.

The *second row* and *third row* of Figure 5 show examples in which our face identification model makes some incorrect predictions (see Figure 4) but the 3D reconstruction results are unaffected.

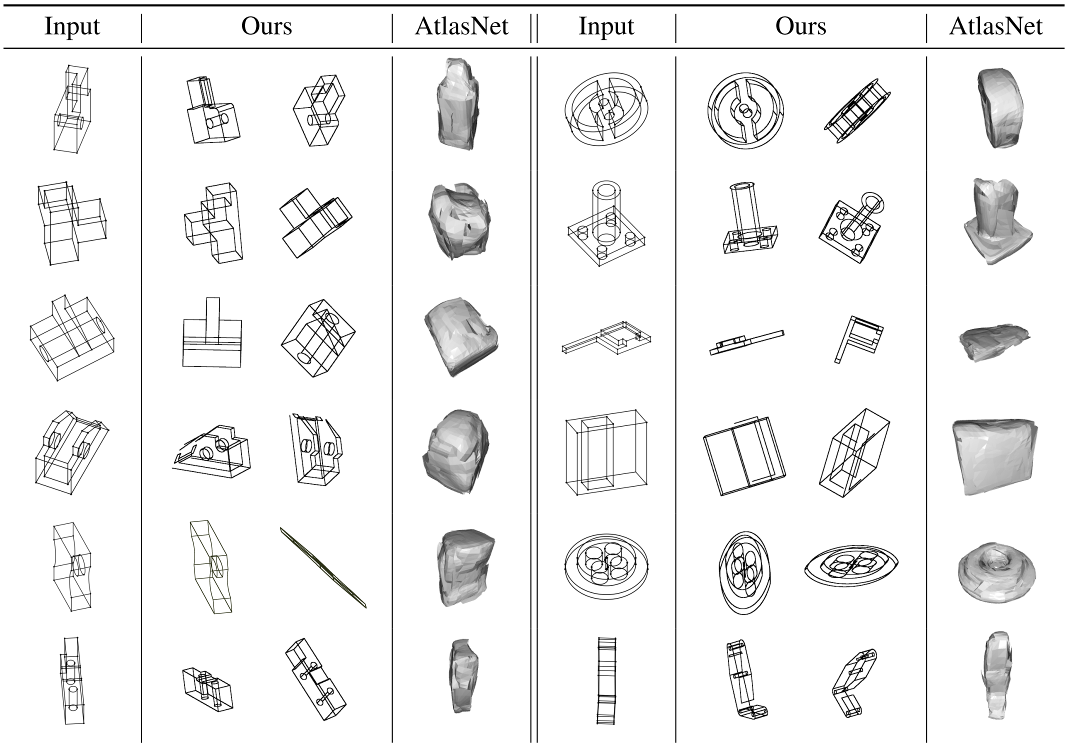


Figure 5: 3D reconstruction results.

For the two cases in the second row, our model generates edge loops which are not part of the true face topology, but all included edges fall onto the same surface (*e.g.*, a plane). For the two cases in the third row, the incorrect faces are filtered because they align to more than two dominant directions. Since the number of constraints created by the faces is typically larger than the number of unknowns in Equation 4, the 3D model can be recovered even if some faces are missed.

The *fourth row* shows two examples in which incorrectly predicted faces affect local part of the 3D model. And the *fifth row* shows two examples in which 3D reconstruction completely fails and the recovered shapes appear flat. The latter typically occurs when our model makes multiple mistakes (*e.g.*, grouping edges on the opposite sides of the 3D shape into one face). Again, we refer readers to Figure 4 for visualization of the mistakes made by our model.

In the *last row* of Figure 5, we show two interesting cases in which even humans have trouble inferring the 3D geometry due to the chosen viewpoints. In contrast, our method does a decent job using the extracted face loops. This suggests that our method is not sensitive to the viewpoints.

6 DISCUSSION

In this work, we present a data-driven approach to face identification. We conclude by pointing out that the proposed method should not be treated as a replacement or competitor to the geometry- and topology-based methods. Instead, we have found that our method complements existing techniques in several aspects, such as capturing the intent of designers, handling curved surfaces and disjoint components in the edge-vertex graph. Meanwhile, our model could make wrong predictions, while geometry- and topology-based methods are guaranteed to succeed when all conditions are met.

Our work opens up several directions for future work. One direction we are particularly excited about is identifying conflicting constraints (due to incorrect predictions made by a deep network) in a geometric constraint system for 3D reconstruction. In this work, we treat all predicted faces equally. A better solution will not only improve the 3D reconstruction results, but may also generalize to other types of structural constraints in CAD.

REFERENCES

Fatemeh Abbasinejad, Pushkar Joshi, and Nina Amenta. Surface patches from unorganized space curves. *Comput. Graph. Forum*, 30(5):1379–1387, 2011.

Siddarameshwar Bagali and Warren N. Waggenspack. A shortest path approach to wireframe to solid model conversion. In *ACM Symposium on Solid Modeling and Applications*, pp. 339–350. ACM, 1995.

Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. In *Int. Conf. Learn. Represent.*, 2015.

John A. Brewer and S. Mark Courter. Automated conversion of curvilinear wire-frame models to surface boundary models; a topological approach. *SIGGRAPH Comput. Graph.*, 20(4):171–178, 1986.

Ayan Das, Yongxin Yang, Timothy Hospedales, Tao Xiang, and Yi-Zhe Song. Béziersketch: A generative model for scalable vector sketches. In *Eur. Conf. Comput. Vis.*, pp. 632–647, 2020.

Ronald D. Dutton and Robert C. Brigham. Efficiently identifying the faces of a solid. *Comput. Graph.*, 7(2):143–147, 1983.

Mark A. Ganter and John J. Uicker. From wire-frame to solid geometric: automated conversion of data representations. *Computer in Mechanical Eng.*, 22(2):40–45, 1983.

Thibault Groueix, Matthew Fisher, Vladimir G Kim, Bryan C Russell, and Mathieu Aubry. A papier-mâché approach to learning 3d surface generation. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pp. 216–224, 2018.

Wenyu Han, Siyuan Xiang, Chenhui Liu, Ruoyu Wang, and Chen Feng. Spare3d: A dataset for spatial reasoning on three-view line drawings. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pp. 14690–14699, 2020.

Patrick M. Hanrahan. Creating volume models from edge-vertex graphs. *SIGGRAPH Comput. Graph.*, 16(3):77–84, 1982.

Richard Hartley and Andrew Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2000.

Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural Comput.*, 9(8):1735–1780, 1997.

Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *Int. Conf. Learn. Represent.*, 2015.

Sebastian Koch, Albert Matveev, Zhongshi Jiang, Francis Williams, Alexey Artemov, Evgeny Burnaev, Marc Alexa, Denis Zorin, and Daniele Panozzo. Abc: A big cad model dataset for geometric deep learning. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pp. 9601–9611, 2019.

Jianzhuang Liu and Xiaoou Tang. Evolutionary search for faces from line drawings. *IEEE Trans. Pattern Anal. Mach. Intell.*, 27(6):861–872, 2005.

Jianzhuang Liu, Yong Tsui Lee, and Wai-Kuen Cham. Identifying faces in a 2d line drawing representing a manifold object. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24(12):1579–1593, 2002.

Jianzhuang Liu, Liangliang Cao, Zhenguo Li, and Xiaoou Tang. Plane-based optimization for 3D object reconstruction from single line drawings. *IEEE Trans. Pattern Anal. Mach. Intell.*, 30(2): 315–327, 2008.

George Markowsky and Michael A. Wesley. Fleshing out wire frames. *IBM J. Res. Dev.*, 24(5): 582–597, 1980.

Charlie Nash, Yaroslav Ganin, S. M. Ali Eslami, and Peter W. Battaglia. Polygen: An autoregressive generative model of 3d meshes. In *Int. Conf. Mach. Learn.*, pp. 7220–7229, 2020.

Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. Automatic differentiation in pytorch. In *Adv. Neural Inform. Process. Worksh.*, 2017.

Moshe Shpitalni and Hod Lipson. Identification of faces in a 2d line drawing projection of a wireframe object. *IEEE Trans. Pattern Anal. Mach. Intell.*, 18(10):1000–1012, 1996a.

Moshe Shpitalni and Hod Lipson. Optimization-based reconstruction of a 3D object from a single freehand line drawing. *Comput. Aided Des.*, 28(8):651–663, 1996b.

Moshe Shpitalni and Hod Lipson. Classification of sketch strokes and corner detection using conic sections and adaptive clustering. *J. Mech. Des*, 119(1):131–135, 1997.

Peter A.C. Varley and Pedro P. Company. A new algorithm for finding faces in wireframes. *Comput. Aided Des.*, 42(4):279–309, 2010.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Adv. Neural Inform. Process. Syst.*, pp. 5998–6008, 2017.

Oriol Vinyals, Meire Fortunato, and Navdeep Jaitly. Pointer networks. In *Adv. Neural Inform. Process. Syst.*, pp. 2692–2700, 2015.

Yingze Wang, Yu Chen, Jianzhuang Liu, and Xiaoou Tang. 3D reconstruction of curved objects from single 2d line drawings. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pp. 1834–1841, 2009.

Yixin Zhuang, Ming Zou, Nathan Carr, and Tao Ju. A general and efficient method for finding cycles in 3d curve networks. *ACM Trans. Graph.*, 32(6):180:1–180:10, 2013.

A ABC DATASET

Since ABC dataset (Koch et al., 2019) is created from Onshape public repository, it contains many duplicate shapes, such as simple boxes and cylinders that were created when novice users tested the Onshape software. There also exist many over-complicated shapes that do not help our model generalize. To make our model learn more effectively, we use the following set of heuristics to filter the ABC dataset when constructing our own dataset.

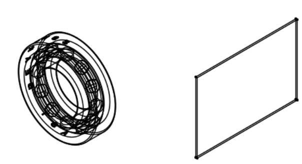


Figure 6: Examples of over-complicated shapes and thin shapes.

Topology. Same shapes must share the same topology. To find duplicate shapes, we first group shapes of matching topology features together, and then find duplicates within each group. When grouping, we consider the following topology features: the number of parts, surfaces, edges of a shape, and its face and edge type distribution.

Visual similarity. We then further cluster within each group based on the visual similarity of the shapes. We use three orthogonal views of a shape as its visual feature, and run agglomerative clustering[3] based on the Jaccard distance of three views between two shapes.

Thickness. Because the ABC dataset contains a lot of primitives which only consist of a plane, we filter out shapes that are too thin. The thickness of a shape is defined as the minimum distance of all pair-wise distances between two edges within a shape. We remove shapes that have thickness smaller than 0.05.

Complexity. We also filter out shapes with more than 42 faces or 37 edges in a face to eliminate over-complicated shapes that result in multiple small, overlapping faces in the same area of the line drawing.

[3]We use `cluster.AgglomerativeClustering` from SciPy.

B ATLASNET

In order to train AtlasNet, we convert the vectorized line drawings to bitmap images. Following SPARE3D (Han et al., 2020), the hidden lines are drawn in red. We adopt the original implementation[4] of AtlasNet, and use the same optimization settings.

C ADDITIONAL RESULTS

In the folder attached to the supplementary material, we provide visualization of (i) all incorrect predictions made by our face identification model and (ii) 3D reconstruction results on the entire test set.

[4]https://github.com/ThibaultGROUEIX/AtlasNet