

GFPACK++: Improving 2D Irregular Packing by Learning Gradient Field with Attention

TIANYANG XUE, Shandong University, China

LIN LU*, Shandong University, China

YANG LIU, Microsoft Research Asia, China

MINGDONG WU, Peking University, China

HAO DONG, Peking University, China

YANBIN ZHANG, Shandong University, China

RENMIN HAN, Shandong University, China

BAOQUAN CHEN, Peking University, China

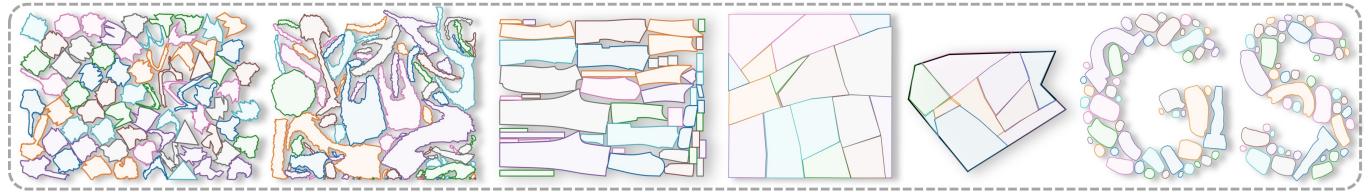


Fig. 1. The proposed attention-based diffusion model for 2D irregular packing, GFPACK++, supports continuous rotation and accommodating arbitrary boundaries.

2D irregular packing is a classic combinatorial optimization problem with various applications, such as material utilization and texture atlas generation. This NP-hard problem requires efficient algorithms to optimize space utilization. Conventional numerical methods suffer from slow convergence and high computational cost. Existing learning-based methods, such as the score-based diffusion model, also have limitations, such as no rotation support, frequent collisions, and poor adaptability to arbitrary boundaries, and slow inferring. The difficulty of learning from teacher packing is to capture the complex geometric relationships among packing examples, which include the spatial (position, orientation) relationships of objects, their geometric features, and container boundary conditions. Representing these relationships in latent space is challenging. We propose GFPACK++, an attention-based gradient field learning approach that addresses this challenge. It consists of two pivotal strategies: *attention-based geometry encoding* for effective feature encoding and *attention-based relation encoding* for learning complex relationships. We investigate the utilization distribution between the teacher and inference data and design a weighting function to prioritize tighter teacher data during training, enhancing learning effectiveness. Our diffusion model supports continuous rotation and outperforms existing methods on various datasets. We achieve higher space utilization over several widely used baselines, one-order faster than the previous diffusion-based method, and promising generalization for arbitrary boundaries. We plan to release our code and datasets to support further research in this direction.

CCS Concepts: • Computing methodologies → Shape modeling; Neural networks; Graphics systems and interfaces.

*Corresponding author.

Authors' addresses: Tianyang Xue, Shandong University, China, timhsue@gmail.com; Lin Lu, Shandong University, China, llu@sdu.edu.cn; Yang Liu, Microsoft Research Asia, China, yangliu@microsoft.com; Mingdong Wu, Peking University, China, wmingd@pku.edu.cn; Hao Dong, Peking University, China, hao.dong@pku.edu.cn; Yanbin Zhang, Shandong University, China, yanbinzhang@mail.sdu.edu.cn; Renmin Han, Shandong University, China, hanrenmin@sdu.edu.cn; Baoquan Chen, Peking University, China, baoquan@pku.edu.cn.

Additional Key Words and Phrases: irregular packing, gradient field, continuous rotation

1 INTRODUCTION

Irregular shape packing is a classic challenge in operation research, with significant implications for various applications such as texture atlas generation and material utilization in manufacturing [Wang et al. 2021]. Its NP-hard nature requires practical and effective packing algorithms to optimize space utilization.

Traditional numerical optimization methods are often slow and costly. Recently, learning-based methods have emerged as a promising alternative for irregular packing, mainly using reinforcement learning [Xu et al. 2023; Yang et al. 2023]. Another promising approach is the diffusion model, which learns geometric relationships from teacher packing examples and establishes a gradient field for all objects, guiding them to move simultaneously [Hosseini et al. 2023; Xue et al. 2023]. In particular, the work of [Xue et al. 2023], namely GFPACK, achieves a notable trade-off between computational cost and space utilization, demonstrating scalability in terms of polygon count and generalizability to rectangular container boundaries. However, it still faces unresolved issues, notably the lack of support for rotation, a relatively high occurrence of object collisions in results, and limited adaptability to arbitrary boundaries.

The main idea of the learning-based packing framework is to capture the geometric relationships between shapes, which are essential for the gradient computation. Such relationships in latent space encompass multiple facets: the *spatial (position, orientation) relationship* of the shapes, the *geometric features relationship* among them, and the *container boundary conditions* that constrain them. However, GFPACK falls short in fully representing these relationships. It solely focuses on the positional aspect, neglecting orientation, thus hindering rotational capabilities. Furthermore, it relies

on PointNeXt [Qian et al. 2022] for geometric feature extraction, ignoring critical details like polygon edges and resulting in a loss of topological and local features. Thus, we argue that the full potential of gradient learning has not been thoroughly exploited in GFPACK.

In this study, we introduce two primary strategies to tackle the challenge of learning packing policies. Firstly, we use a graph to represent each object and encode it with an *attention-based geometric encoder*, which can capture polygonal geometry information effectively. Secondly, we design an *attention-based relation encoding* mechanism to encode intricate relationships both between polygons and among polygons and boundaries, eliminating the need to explicitly represent the relationships with a graph. This attention-based diffusion model enhances the understanding of how objects interact with each other and the boundaries, resulting in superior performance in both training and inference. We also investigate the utilization distribution between the teacher and inference data and design a weighting function to prioritize tighter teacher data during training, confirmed to be effective.

This paper makes the following contributions:

- We propose GFPACK++, a novel packing framework that employs attention-based learning of gradient fields, supporting continuous rotation and enhancing generalization.
- Advancing beyond GFPACK, we encapsulate geometric relationships in an attention-based encoder, incorporating mutual spatial (position, orientation) relationships of objects, their geometric features, and container boundary conditions, and a weighting function to facilitate effective learning policies.
- GFPACK++ achieves superior irregular packing results, demonstrating higher space utilization over several widely used baselines, one order of magnitude faster than the previous diffusion-based method, coupled with promising generalization for arbitrary boundaries.

2 RELATED WORK

Traditional packing approaches. Traditional methods usually involve placement optimization and sequence planning (*c.f.* surveys [Bennell and Oliveira 2009; Leao et al. 2020]). Most research focuses on placement optimization with heuristic rules (e.g., BL, BLF), NFP-based search [Bennell and Song 2008], or mathematical models [Hopper and Turton 2001]. Novel approaches include SDF-based placement [Pan et al. 2023] and spectral domain correlation [Cui et al. 2023], but they only support limited rotations. Sequence planning combines heuristics [Lopez et al. 2013] with global search strategies like genetic algorithms [Junior et al. 2013], simulated annealing [Gomes and Oliveira 2006], beam search [Bennell and Song 2010], or particle swarm optimization [Shalaby and Kashkoush 2013]. However, these methods are computationally expensive and could improve space utilization by considering continuous rotation [Romanova et al. 2018].

Reinforcement-learning-based packing. Packing problems have attracted increasing interest in the field of machine learning, especially reinforcement learning (RL). RL methods can handle online 3D bin packing, where the object order is either fixed [Zhao et al. 2021, 2022] or optimized [Hu et al. 2020; Xu et al. 2023; Zhang et al.

2021], also known as the transport-and-packing (TAP) problem. RL methods can also deal with irregular packing, where the objects have complex shapes and orientations [Huang et al. 2023; Zhao et al. 2023]. A recent RL-based pipeline [Yang et al. 2023] integrates high-level object selection and low-level pose estimation and achieves a state-of-the-art packing ratio improvement of 5% to 10% over XAtlas [Young 2022] and NFP. However, RL-based packing methods share a commonality in considering object sequence and positioning separately, indicating room for improvement in achieving a better balance between space utilization and computational efficiency.

Diffusion-based packing and arrangement. As a powerful generative model, diffusion models can effectively learn spatial relationships. Recent studies [Liu et al. 2023; Wei et al. 2023; Wu et al. 2022] show that object rearrangement tasks can implicitly learn object arrangements from training set, enabling object movement without explicit goals. However, object rearrangement does not require complex shape comparisons. Hosseini et al. [2023] tackle this challenge by proposing a spatial puzzle solver using diffusion models. Nevertheless, puzzle solving focuses on pairwise object relationships as the ground truth solution exists, whereas packing problems lack such priors on the global optima, indicating a larger search space and necessitating more intricate consideration of geometric relationships. GFPACK [Xue et al. 2023] explores the use of diffusion models for 2D irregular packing problems. It learns geometric relationships from teacher examples and creates a gradient field for all objects, moving them simultaneously. While effective, GFPACK does not support rotation, exhibits frequent collisions, has limited adaptability to arbitrary boundaries, and is time-consuming. Our work addresses these issues by incorporating attention mechanisms into gradient field learning and designing an efficient learning framework.

3 OVERVIEW

3.1 Problem Statement

The 2D irregular packing problem aims to fit a set of 2D irregular polygons $P = \{P_i\}_{i=0}^n$ into a polygonal container B with minimal wasted space. This problem can be formulated as an optimization problem, where the variables are the rigid transformations of each polygon, denoted as $A = \{A_i\}_{i=0}^n$. The objective is to maximize the space utilization: $u = \bigcup_i \text{Area}(A_i(P_i)) / \text{Area}(B)$, subject to the constraints that each transformed polygon must be inside the container and not overlap with any other transformed polygon. In this paper, we mainly set the container as an infinite strip that extends horizontally with a given height. Our objective is to optimize the packing of P to achieve the minimal length of the strip [Baker et al. 1980], a practical consideration in most manufacturing scenarios.

3.2 Learning Gradient Field with Attention

Following GFPACK, we formulate the packing problem as a Markov Process. In this process, at each state, a velocity vector is generated for each polygon from the preceding state. This velocity includes both linear and angular components, as demonstrated in Fig. 2 (a). Our goal is to learn a neural network that predicts the velocity at each timestep, using a teacher (training) set.

To generate the velocity vector, a state is encoded into a latent vector, which is then ‘translated’ by decoders into a velocity, as

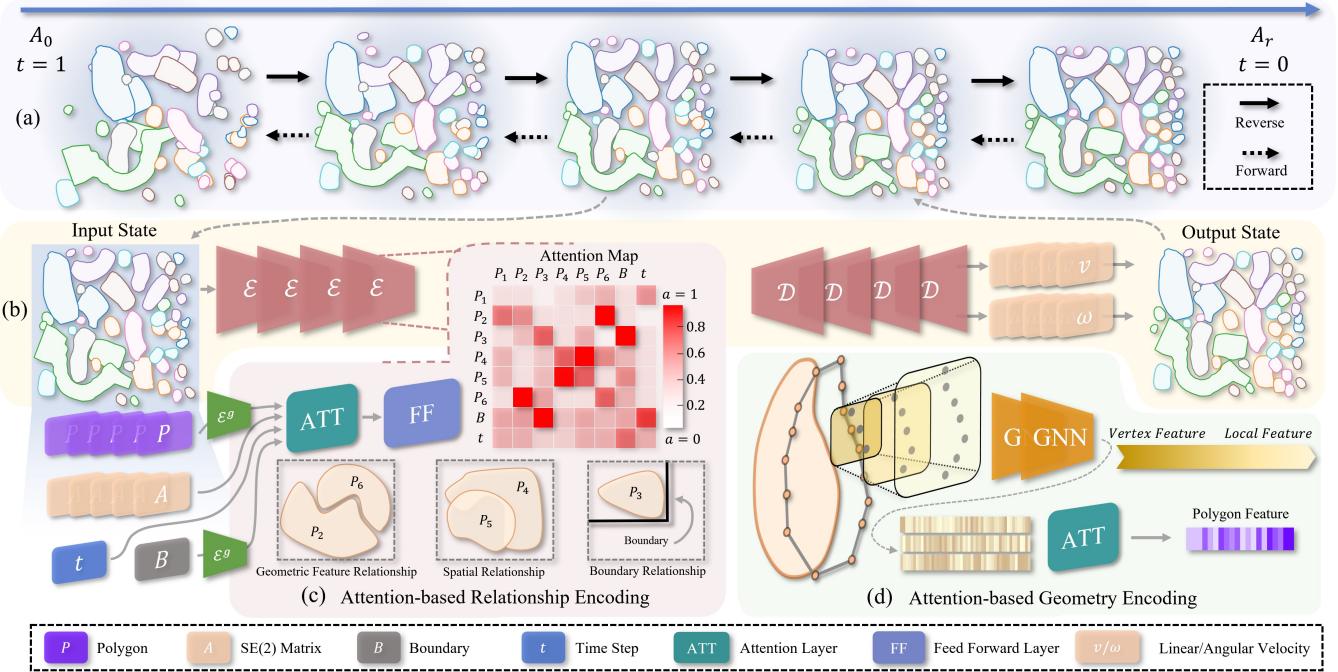


Fig. 2. (a) Our approach is a diffusion-based generation method. (b) We use a sequence-to-sequence model to encode and decode the state and velocities at each time step and generate the next state accordingly. Our model consists of two components: an attention-based relationship encoder (c) and an attention-based geometry encoder (d). The former encodes the geometric feature, spatial, and boundary relationships among the input variables using the attention mechanism. The latter computes the geometric features \mathcal{E}^g using a multi-level GCN feature aggregation network and Attention Pooling.

shown in Fig. 2 (b). The accuracy of encoding implicit relationships within the state significantly influences efficiency. This is because, although the teacher data may not always be optimal, it often contains optimal local layouts. Therefore, an effective learning policy should be capable of discerning and learning such latent relationships in the teacher data. In GFPACK [Xue et al. 2023], a graph network determined by the distances between polygons is employed to facilitate state encoding. However, this approach overlooks the geometric relationships among the polygons. Moreover, the graph has to be rebuilt at each sampling step, leading to low efficiency.

To address these issues, we introduce two key designs. We developed an attention-based relationship encoder that learns the encoding of multifaceted relationships between polygons and between polygons and boundaries, achieving high efficiency without explicitly updating the graph like GFPACK, as illustrated in Fig. 2 (c). Additionally, we created a feature representation module to accurately capture the local and global geometric features of shapes, depicted in Fig. 2 (d). This module ensures the comprehensive acquisition of geometric information, crucial for encoding the geometric relationships among shapes. Through the aforementioned designs, we have trained a score-based diffusion model to learn the gradient field for packing problems, accommodating continuous rotation and irregular boundaries. Accompanied by a weighting function that compels the model to focus on high-utilization data and utilizing an enhancement strategy for utilization, this results in the generation of high-utilization packing instances.

4 METHOD

We tackle the packing problem by encoding a state as polygon velocity, where the state $S = \{A, t, P, B\}$. Initially, polygons, composed of contour points present in both P and B , are encoded as geometric features (Sec. 4.1). Subsequently, these features, in conjunction with additional information A and t , are fed into a relation encoding module to generate velocity (Sec. 4.2). This entire process is trained through a score-based diffusion model (Sec. 4.3, 4.4) and undergoes the utilization enhancement (Sec. 4.5).

4.1 Attention-based Geometry Encoding

For a polygon contour point set $N = \{N_i\}_{i=0}^m$, where $N_i \in \mathbb{R}^2$ and all vertices are given in order, we can readily construct a graph $G = \{N, E = \{(i, i+1) | i \in [0, m]\}\}$.

Furthermore, for each vertex N_i , we compute its internal angle α and together with its coordinates, we form a set $f_i^{(0)} = (x_i, y_i, \alpha_i)$ as initial vertex feature. The graph G and initial feature $f^{(0)}$ is then input into a GCN [Kipf and Welling 2017], and undergoes the following aggregation process:

$$h_\Theta(H^{(l)}) = \sum_{(i,j) \in E} \text{ReLU}(\hat{D}^{-\frac{1}{2}} \hat{\mathcal{A}} \hat{D}^{-\frac{1}{2}} H^{(l)} W^{(l)}), \quad (1)$$

where $H^{(l)}$ is the features at layer l , $H^{(l)} = \{f^{(l)}\}$, $\hat{\mathcal{A}} = \mathcal{A} + I$ is the adjacency matrix with added self-connections, \hat{D} is degree matrix of $\hat{\mathcal{A}}$ and W is a learnable weight matrix. For each layer of GCN, we

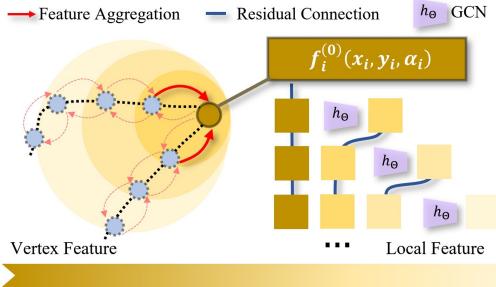


Fig. 3. Local Feature Representation. Our method leverages the information aggregation capability of GCNs combined with the multi-scale abilities of residual connections. It aggregates the geometric features of each point and its neighboring area with Eq. 1.

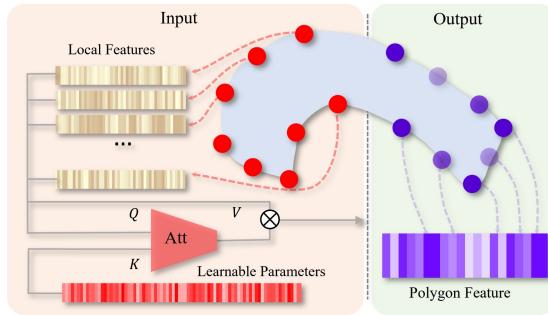


Fig. 4. Attention pooling for local features. For the local features of each point, we aggregate them into the shape features of the polygon using an attention mechanism.

connect them as depicted in Fig. 3, employing a residual connection [He et al. 2016]:

$$H^{(l+1)} = h_\phi(h_\theta(H^{(l)}), H^{(l)}) \quad (2)$$

We achieved the local geometry encoding $H^{(L)}$ through the aggregation of L layers.

Inspired by [Lee et al. 2019; Vaswani et al. 2017; Veličković et al. 2018], we introduce an *attention pooling* method to integrate local geometry features into a polygon geometry feature. The polygon geometry feature is computed as follows (Fig. 4):

$$\text{ATT}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_K}}\right)V \quad (3)$$

Here, Q (Query), K (Key), and V (Value) represent the key components of the attention mechanism, and d_K is the dimension of K . In our design, $K = W_K$ is a learnable matrix, and $Q = V = H^{(L)}$.

Polygons \mathbf{P} and boundary B are computed as geometric features \mathcal{F}_P and \mathcal{F}_B using the aforementioned method.

4.2 Attention-based Relation Encoding

Given the input state $S_i = \{\mathbf{A}_i, t_i; \mathbf{P}, \mathbf{B}\}$ the output velocities v is used to calculate the \mathbf{A}_{i+1} in the next state with $\mathbf{A}_{i+1} = \mathbf{A}_i + v(S_i)$. We trained a Transformer [Vaswani et al. 2017] model to encode the

state and generate the velocity, employing an attention mechanism to encode the geometric relationships mentioned above.

Initially, the state is encoded as a uniform-shaped latent feature vector with the following: Polygons and boundary represented in contour are encoded through Sec. 4.1 as $\mathcal{F}_P \in \mathbb{R}^{|\mathbf{P}| \times d_p}$ and $\mathcal{F}_B \in \mathbb{R}^{d_b}$. $SE(2)$ matrices are encoded through an MLP as $\mathcal{F}_A \in \mathbb{R}^{|\mathbf{P}| \times d_a}$. Time step t is processed using Gaussian Fourier Projection:

$$(t_x, t_y) = (\sin(2\pi w t), \cos(2\pi w t)), \quad (4)$$

where w is initialized as $w \sim \mathcal{N}(0, I)$, $I \in \mathbb{R}^{d_t \times d_t}$, t_x and t_y are feature vectors of shape d_t calculated by t . This projection transforms t into a $2d_t$ -dimensional feature vector.

We devise three attention-based relation encoding modules with Eq. 3. The first is *geometric feature relationships* module, derived from the geometric features among polygons, achieved through $Q = K = V = \mathcal{F}_P$. For encoding the *spatial relationships*, we design $Q = V = (\mathcal{F}_P, \mathcal{F}_A), K = \mathcal{F}_A$. Ultimately, we configure $Q = V = (\mathcal{F}_P, \mathcal{F}_A), K = \mathcal{F}_B$ to process the *boundary relationships*. All these encoded features are fed into the next layer with a Feed Forward layer.

4.3 Packing with Gradient Field

We leverage the gradient field method from GFPACK to generate velocity with score-based diffusion. Specifically, we learn the log-density $\psi \approx \nabla_{\mathbf{A}} \log p_{opt}(\mathbf{A})$ from sample distributions generated by a teacher packing algorithm.

We adopt gradient fields to global and local optimization phases through conditional time step $\{t_i \in [0, 1]\}_{i=0}^r$ as noise scale. We incorporate Variance-Exploding Stochastic Differential Equation (VE-SDE) proposed by [Song et al. 2021] for noising control.

$$\sigma(t) = \sigma_{\min} \left(\frac{\sigma_{\max}}{\sigma_{\min}} \right)^t \quad (5)$$

VE-SDE is capable of preserving the original data distribution, enhancing precision control for packing problems. This enables the calculation of the distribution at any time step t :

$$p(\mathbf{A}(t)|C) = \int \mathcal{N}(\mathbf{A}(t)|\mathbf{A}(0), \sigma^2(t)\mathbf{I}) \cdot p(\mathbf{A}(0)|C) d\mathbf{A}(0) \quad (6)$$

where $C = \{\mathbf{P}, \mathbf{B}\}$ and $p(\mathbf{A}(t)|C)$ represents the marginal distribution of $\mathbf{A}(t)$ conditional on C . The gradient field based on time step and boundary can be expressed as:

$$\psi(\mathbf{A}, t; C) \approx \nabla_{\mathbf{A}} \log p(\mathbf{A}(t)|C) \quad (7)$$

Conditioned on any $t \sim \mathcal{U}(0, 1)$, we can train ψ via Denoising Score Matching (DSM) [Vincent 2011]:

$$\mathcal{L}(\psi) = \mathbb{E}_{\mathbf{A}(t) \sim p(\mathbf{A}(t))} \left[\left\| \psi(\mathbf{A}, t; C) - \frac{\mathbf{A}(0) - \mathbf{A}(t)}{\sigma(t)^2} \right\|_2^2 \right] \quad (8)$$

We acquire the final solution by solving the following reverse-time Random Stochastic Differential Equation (RSDE):

$$d\mathbf{A} = -g^2(t) \nabla_{\mathbf{A}} \log p(\mathbf{A}(t)|C) dt + g(t) \sqrt{dt} \epsilon, \quad (9)$$

where $\sqrt{dt} \epsilon$ is the standard Wiener process with $\epsilon \sim \mathcal{N}(0, 1)$, $g(t) = \sigma(t) \sqrt{2 \log(\frac{\sigma_{\max}}{\sigma_{\min}})}$ and $\nabla_{\mathbf{A}} \log p(\mathbf{A}(t)|C)$ is approximated by the learned gradient field $\psi(\mathbf{A}, t; C)$. The solution path of the RSDE

mentioned above can be exactly conceptualized as the velocities of the generation process $\{v(\mathcal{S}_i)\}_{i=1}^r$.

4.4 Training Optimization

We observed that the utilization rates of the training data influence the generation results, with higher utilization rates leading to better outcomes. To prioritize high utilization data during learning, we apply a *sigmoid* weighting to the loss function in Eq. 8. $\lambda(\mathbf{A}, \mathbf{C})$ directs the network's focus towards optimal layouts:

$$\lambda(\mathbf{A}, \mathbf{C}) = \text{sigmoid}\left(\frac{u(\mathbf{A}, \mathbf{C}) - U_{avg}}{U_{max} - U_{min}} \times 10\right), \quad (10)$$

where $U_{min|avg|max}$ represents the Min|Avg|Max utilization ratio across the entire teacher set, respectively.

This weighting function effectively balances the quality and diversity of the training data.

4.5 Utilization Enhancement

Diffusion-based methods fall short of achieving strict constraints but offer approximate optimal results. To address this, we developed a local utilization enhancement algorithm for resolving minor collisions and further minimizing polygon spacing in strip packing. Initially, we calculate separation vectors $v_{i,j}$ between polygon pairs using the Separating Axis Theorem [Ericson 2004]. To prevent polygons from exceeding the boundary B , we introduce an offset vector o_i between the polygon and the boundary. The displacement v_i of polygon i is then computed as:

$$v_i = \sum_{j>i} \frac{S_j v_{i,j}}{S_i + S_j} - \sum_{j< i} \frac{S_j v_{i,j}}{S_i + S_j} + o_i, \quad (11)$$

where S_i is the area of P_i . We iterate to update v_i to A_i . Subsequently, we implement a rapid gap elimination method inspired by [Cui et al. 2023]. Following these iterations, binary search is employed to determine the minimum distance polygons can move in both $-x$ and $-y$ directions. This distance is then utilized to reduce gaps between polygons. Refer to Fig. 9 for an illustration.

5 IMPLEMENTATION

5.1 Datasets and Teacher Algorithm

We validated our algorithm using six datasets (Figs. 5, 6). The Garment and Dental datasets stem from real manufacturing scenarios [Xue et al. 2023]. We introduced a dataset, named Dental (alphabet), generated by packing dental data according to alphabetical boundaries. We also employed Atlas datasets from [Yang et al. 2023], which contain the UV patches of 3D models.

For these four datasets, we utilized SVGNest [Qiao 2017] as the teacher algorithm to generate exemplar packing results (training data distributions are provided in the appendix Table 8). SVGNest employs a packing position selection based on NFP and a genetic algorithm for sequence and discrete angle selection [Lopez et al. 2013; Valvo 2017]. In the rotational learning study, we set the range of discrete angle selection to 32 and applied random rotations to the polygons before inputting them into the teacher algorithm. This strategy helps the model learn the distribution of arbitrary rotation angles. The SVGNest's population size and number of iterations

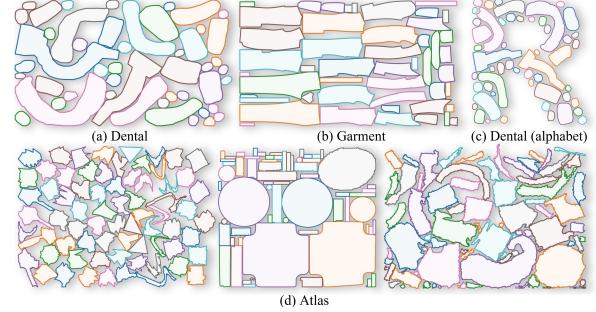


Fig. 5. Samples from the teacher datasets.

were both set to 8. We ran the algorithm 10 times on the same set of polygon combinations and selected the best result.

We introduced two new puzzle datasets, Puzzle (square) and Puzzle (arbitrary), to assess the method's generalizability across various polygon shapes and boundaries. Both datasets feature randomly generated polygon shapes and irregular boundaries, creating "ground-truth" packing solutions with 100% utilization. Puzzle (square) comprises 16 polygon fragments per item, while Puzzle (arbitrary) consists of 10 fragments. Additionally, Puzzle (arbitrary) includes 8192 boundaries randomly selected from pieces generated in Puzzle (square). The generation algorithm for each data item involves randomly dividing a polygon boundary into multiple parts, ensuring no intersection.

The training data for each dataset comprises approximately 100,000 teacher items of random, non-colliding polygon combinations. Detailed polygon information and utilization ratio distributions for each dataset are provided in the appendix Sec. A.

5.2 Experimental Setup

In GFPACK++, we employed a four-layer ($L = 4$) Graph Convolutional Network (GCN) in Sec. 4.1 for local feature extraction, with attention pooling, ultimately generating feature vectors F_p and F_b with $d_p = 64$, $d_b = 128$. The $SE(2)$ matrices \mathbf{A} and time step t are encoded to F_A with $d_a = 64$, $d_t = 64$.

The feature vectors F_p and F_A are concatenated together to form a 128-dimensional feature vector. Along with F_A and F_t , they are fed into the transformer structure with 8-layer encoders and 8-layer decoders (Sec. 4.2). Through the attention mechanism, relationship information is encoded and the velocity is outputted.

At each training step, a random t is selected from the range $[0.01, 1]$. The corresponding perturbed A' with a noise level at t is then input into the network, calculated based on the formulas in Sec. 4.3 with $\sigma_{min} = 0.1$ and $\sigma_{max} = 1000$, aligned with packing results. Additionally, for computational convenience, the polygon's translations and rotations are represented as $(x, y, \cos(\theta), \sin(\theta))$ before being input into the network. This ensures uniform data scales in both rotation and translation spaces.

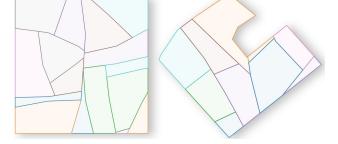


Fig. 6. Puzzle dataset.

Table 1. Statistics of packing ratios (Min|Avg|Max) and time consumption for the average packing results. All of these results were optimized with our utilization enhancement algorithm.

Dataset	XAtlas	NFP	SVGnest	GFPACK++ ($b = 128$)	GFPACK++ ($b = 512$)
Garment	62.26 69.53 74.70 1.01 s	61.49 65.49 67.94 6.13 s	69.38 72.75 75.13 80.21 s	69.37 74.22 77.83 3.25 s	74.15 77.04 80.62 10.5 s
Dental	63.18 70.86 73.59 1.25 s	60.42 66.30 68.13 8.22 s	67.22 73.64 76.49 96.73 s	70.57 75.21 78.67 4.28 s	74.06 77.53 79.79 14.5 s
Puzzle (square)	56.16 66.48 76.15 1.21 s	56.90 62.53 68.13 4.52 s	61.00 67.22 73.21 42.55 s	82.22 93.99 98.64 6.23 s	84.80 95.12 98.74 32.1 s
Atlas (building)	50.56 71.21 87.62 0.96 s	41.97 67.32 83.31 2.76 s	44.97 74.51 90.54 52.43 s	58.96 78.78 98.47 8.12 s	66.03 80.16 98.88 25.3 s
Atlas (object)	38.58 60.98 76.09 1.43 s	32.18 57.75 83.12 45.3 s	41.53 63.87 83.12 402.2 s	31.76 65.11 86.08 15.5 s	41.95 67.47 86.81 65.4 s

Finally, we compute the *Mean Squared Error* (MSE) between the network output and the perturbation, as described in Eq. 8, optimizing using the AdamW optimizer [Loshchilov and Hutter 2019] at a learning rate of 2×10^{-4} . During generation, for a given polygon set, we initialize $A(t_0) \sim N(0, \sigma(t_0))$ with $t_0 = 1$. We then sample $r = 128$ steps using the Euler-Maruyama method [Kloeden and Platen 2011] to generate A_r through RSDE in Eq. 9. The results are processed with a parallel-accelerated utilization enhancement algorithm described in Sec. 4.5, implemented in C++ with a 100-step iteration limit. GFPACK++ is capable of generating a batch of data in parallel under a specified batch size b . We select the collision-free result that exhibits the highest utilization ratio from one batch.

6 RESULTS

6.1 Experimental Analysis

Baselines. We applied strip packing to the Garment, Dental, and Puzzle datasets. The strip heights were randomly set for Garment, 1200 for Dental, and 2000 for Puzzle. GFPACK++ was trained and tested on datasets containing 48 polygons for Dental and Garment, and 16 polygons for Puzzle (square). The Atlas datasets were tested without fixed boundaries, with the number of polygons varying from 5 to 200. Following model training, we randomly generated new test sets, each comprising 128 instances, to evaluate the algorithms. The statistics are summarized in Table 1. Among these, XAtlas [Young 2022] and NFP are non-sequential optimization algorithms known for their efficiency in single-step position selection, making them relatively fast. XAtlas was configured in brute-force mode and adapted to support strip packing, with details in the appendix Sec. B. SVGnest [Qiao 2017], which also serves as our teacher algorithm, combines sequential optimization and achieves superior space utilization at the expense of increased time consumption. GFPACK++ demonstrates superior space utilization by 4%-32% compared to the three baseline algorithms, attributed in part to our learning of optimal layouts and the ability to explore a continuous optimal distribution of rotation. These results are further visualized in Figs. 11, 12, and 13.

Utilization distribution shift. We observed that the utilization ratio of teacher data is enhanced by GFPACK++, and is further improved by our weighting function introduced in Sec. 4.4. The experimental setup was as follows: we selected 1000 data randomly from the teacher dataset and conducted inference by GFPACK++ trained with and without the weighting function. The utilization distribution of the dental data, depicted in Fig. 7, shows a shift towards greater utilization when the weighting function is applied, indicating that

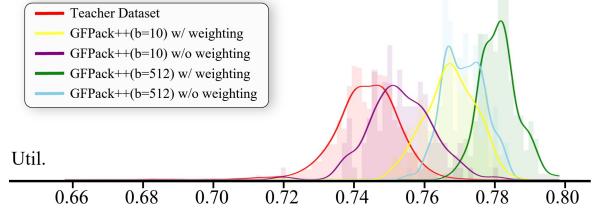


Fig. 7. Utilization distribution shift for the dental data. The figure plots the histogram of utilization ratios of 1000 random data.

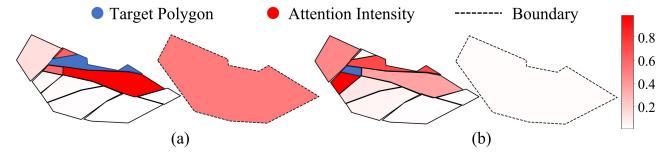


Fig. 8. Visualization of the attention intensity of the 7th-layer decoder at $t = 0.2$ using data from the Puzzle (arbitrary) dataset. Attention intensity of each polygon is normalized by $\frac{a_i - a_{\min}}{a_{\max} - a_{\min}}$.

training data with higher utilization are more effective to the training. Moreover, employing strategies that allow for the simultaneous generation of multiple solutions on the batch size b enhances utilization. We chose two b values, as the teacher data is chosen from 10 samples (Sec. 5.1) and GFPACK++ inference is performed with $b = 512$.

Attention-based relation encoding visualization. To better understand the relations represented by the attention mechanism, we visualize an attention weight from our model. Fig. 8 illustrates the attention intensity at $t = 0.2$, closing to the convergence. Polygons close to the boundary exhibit stronger attention towards the boundary, whereas polygons fully enveloped by surrounding polygons demonstrate lower boundary attention intensity. Additionally, polygons display near-zero attention intensity towards those polygons with which they are not adjacent.

Utilization enhancement. As illustrated in Fig. 9, enhancement algorithm is employed to eliminate minor overlaps and remove gaps. The impact of enhancement on other algorithms is not significant (about 1%, see Table 2), as they do not contain overlaps, have relatively small gaps, and offer little room for local optimization.

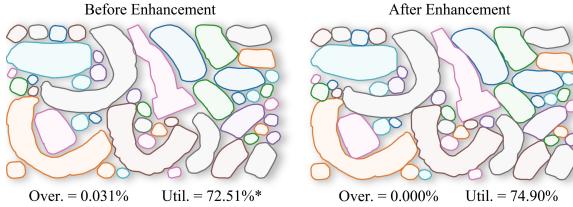


Fig. 9. A comparative result before and after enhancement on the Dental dataset. *Due to overlapping, this instance is infeasible. However, for reference, we calculated the ‘utilization’ by dividing the total area of all polygons by the area of the bounding box.

Table 2. Average improvements achieved through utilization enhancement.

Dataset	XAtlas		SVGnest	
	Before	After	Before	After
Garment	68.90%	69.53%	70.98%	72.75%
Dental	69.88%	70.86%	72.55%	73.64%

6.2 GFPACK++ v.s. GFPACK

In Table 3, GFPACK++ is compared with GFPACK, showcasing notable enhancements in utilization, overlap reduction, and computational efficiency. GFPACK++ exhibits a 5% to 7% increase in utilization when both methods employ utilization enhancement techniques, and it operates an order of magnitude faster than GFPACK.

Regarding the overlapping area ratio, GFPACK++ consistently reduces it to under 0.1% of the total area, effectively eliminated by the enhancement algorithm. In contrast, GFPACK not only generates larger overlaps but also requires more time for their elimination. The feasibility of solutions, measured as the average number of feasible solutions per batch—was also evaluated. Despite the use of enhancement algorithms, GFPACK yielded only 64.2% feasible solutions in the garment dataset and 34.2% in the dental dataset. Meanwhile, GFPACK++ achieved 100%, 98.4%, and 96.3% average rates of feasible

Table 3. Comparison of GFPACK and GFPACK++ on Garment and Dental. ‘E’ indicates enhancement of model outputs, ‘R’ denotes training with arbitrary rotations. ‘Util.’ refers to the average spatial utilization, ‘Over.’ is the average overlap of generated outcomes as a percentage of total polygon area, and ‘Time’ represents the average generation time.

Dataset	Algo.	Util.(%)	Over.(%)	Time(s)
Garment	GFPACK	69.82	0.85	81.2
	GFPACK (E)	72.17	0.23	124.0
	GFPACK++	74.25	0.08	7.9
	GFPACK++ (E)	77.04	0.00	10.5
Dental	GFPACK	66.59	1.43	80.4
	GFPACK++	69.31	0.06	8.0
	GFPACK (E)	70.11	0.78	132.1
	GFPACK++ (E)	73.93	0.00	12.1
	GFPACK++ (R)	74.25	0.08	7.9
	GFPACK++ (R+E)	77.53	0.00	14.5

Table 4. Performance comparison of GFPACK++ and GFPACK using different geometric encoders (first column) and relational encoders (second column) on the dental dataset. The results do not include enhancement.

Geometric Enc.	Relation Enc.	Over.	Util.
GFPACK	GFPACK	13.2%	-
GFPACK++	GFPACK	8.38%	-
GFPACK	GFPACK++	1.89%	72.84%
GFPACK++ (AvgPool)	GFPACK++	0.85%	73.11%
GFPACK++	GFPACK++	0.08%	74.25%

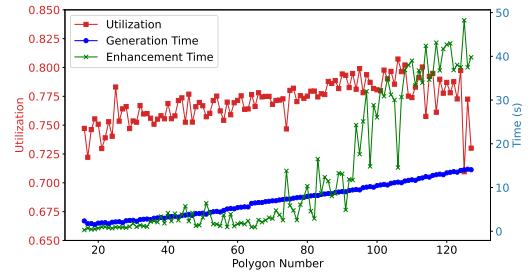


Fig. 10. Polygon number scalability of GFPACK++ (16 to 128).

Table 5. Generalization test on Dental (alphabet).

Batch Size	Valid Solutions	Time
GFPACK++ ($b = 128$)	38.53%	4.21 s
GFPACK++ ($b = 512$)	67.54%	7.95 s

solutions across the garment, dental, and rotation-inclusive dental datasets, respectively. Significantly, GFPACK was unable to produce feasible solutions for the rotation-inclusive dental dataset, with the learning policy struggling to determine reasonable velocities amidst the greatly increased complexity of shape relationships.

To further clarify the disparities between GFPACK and GFPACK++, we conducted ablation experiments on the proposed encoders. Table 4 shows that an accurate representation of polygonal geometric features helps to reduce overlap areas.

6.3 Generalization and Scalability

Number of polygons. We tested the scalability of GFPACK++ to the number of polygons in the Dental dataset, by training the model with a training set 24 to 64 polygons and validating it on a test set with 16 to 128 polygons. As shown in Fig. 10, our method maintains a low level of time consumption while avoiding significant decreases in space utilization when the number of polygons increases.

Boundary and polygon shape variation. The generalization capabilities of GFPACK++ were assessed for polygon shapes and arbitrary boundaries using the Dental and Puzzle datasets. For the Dental dataset, 128 groups of valid data were generated, and the model attained a packing success ratio of 67.54% with a batch size of 512, as

Table 6. Generalization test on Puzzle (arbitrary).

Batch Size	IoU (Min Avg Max)	Time
GFPACK++ ($b = 128$)	80.39 % 93.75 % 96.80 %	3.27 s
GFPACK++ ($b = 512$)	82.43 % 94.91 % 98.86 %	4.23 s

Table 7. Comparison of packing ratios (Min|Avg|Max) between GFPACK++ and [Yang et al. 2023]. GFPACK++* means our method trained on the general dataset.

Algo.	Building	Object
[Yang et al. 2023]	68.3 82.7 98.0	37.7 68.7 86.2
GFPACK++	66.0 80.2 98.9	42.0 67.5 86.8
GFPACK++*	61.5 78.4 97.4	36.2 66.2 83.9

detailed in Table 5. In the case of the Puzzle dataset, 128 new boundaries with corresponding test data were created, approaching packing as a puzzle-solving task without enhancements. Performance was gauged using the Intersection over Union (IoU), defined as $IoU = \frac{B \cap (UP)}{B \cup (UP)}$, which quantifies the proportion of non-overlapping area within the boundary. This metric, akin to the overlap score in [Hosseini et al. 2023], underscores the model’s proficiency. Table 6 demonstrates GFPACK++’s high IoU, underscoring its applicability in puzzle-solving contexts, while Fig. 11 provides some visual results.

Comparison with [Yang et al. 2023]. Table 7 shows that our method achieves results comparable to those in [Yang et al. 2023], which is designed for Atlas data. Our training data are generated by SVGnest, and we believe using the data from [Yang et al. 2023] (currently unavailable) could further improve our model. Training exclusively on the general dataset resulted in only a slight performance decrease. Refer to Fig. 13 for visualizations.

7 CONCLUSIONS AND PERSPECTIVES

We present GFPACK++, a novel packing framework utilizing attention-based learning of gradient fields. GFPACK++ supports continuous rotation, substantially improving generalization capabilities. It outperforms baseline algorithms in terms of space utilization and delivers efficiency that is an order of magnitude faster than prior diffusion-based methods. We will release the source code and datasets to facilitate further research.

Future work. Inspired by the observed utilization distribution shift, we speculate that the incorporation of GFPACK++’s generation results into the teacher data could further enhance the learning. We are also interested in developing a unified model for general 2D irregular packing, with challenges potentially arising in generating sufficiently diverse geometric features concisely and handling domain discrepancies. Furthermore, GFPACK++ demonstrates potential in puzzle-solving, prompting an exploration of its adaptation to fragment reassembling problems [Huang et al. 2006; Sellán et al. 2022]. Lastly, we plan to extend the proposed diffusion model to address 3D irregular packing problems.

REFERENCES

- Brenda S. Baker, E. G. Coffman, Jr., and Ronald L. Rivest. 1980. Orthogonal Packings in Two Dimensions. *SIAM J. Comput.* 9, 4 (1980), 846–855. <https://doi.org/10.1137/0209064> arXiv:<https://doi.org/10.1137/0209064>
- J A Bennell and J F Oliveira. 2009. A tutorial in irregular shape packing problems. *Journal of the Operational Research Society* 60, sup1 (may 2009), S93–S105. <https://doi.org/10.1057/jors.2008.169>
- Julia A Bennell and Xiang Song. 2008. A comprehensive and robust procedure for obtaining the nofit polygon using Minkowski sums. *Computers & Operations Research* 35, 1 (2008), 267–281. <https://doi.org/10.1016/j.cor.2006.02.026> Part Special Issue: Applications of OR in Finance.
- Julia A Bennell and Xiang Song. 2010. A beam search implementation for the irregular shape packing problem. *Journal of Heuristics* 16 (2010), 167–188.
- Qiaodong Cui, Victor Rong, Desai Chen, and Wojciech Matusik. 2023. Dense, Interlocking-Free and Scalable Spectral Packing of Generic 3D Objects. *ACM Trans. Graph.* 42, 4, Article 141 (jul 2023), 14 pages. <https://doi.org/10.1145/3592126>
- Christer Ericson. 2004. *Real-time collision detection*. Crc Press.
- A Miguel Gomes and José F Oliveira. 2006. Solving irregular strip packing problems by hybridising simulated annealing and linear programming. *European Journal of Operational Research* 171, 3 (2006), 811–829.
- Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep Residual Learning for Image Recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770–778. <https://doi.org/10.1109/CVPR.2016.90>
- E Hopper and B.C.H Turton. 2001. An empirical investigation of meta-heuristic and heuristic algorithms for a 2D packing problem. *European Journal of Operational Research* 128, 1 (jan 2001), 34–57. [https://doi.org/10.1016/s0377-2217\(99\)00357-4](https://doi.org/10.1016/s0377-2217(99)00357-4)
- Sepidehsadat Hosseini, Mohammad Amin Shabani, Saghaf Irandoust, and Yasutaka Furukawa. 2023. PuzzleFusion: Unleashing the Power of Diffusion Models for Spatial Puzzle Solving. In *Thirty-seventh Conference on Neural Information Processing Systems*.
- Ruizhen Hu, Juzhan Xu, Bin Chen, Minglun Gong, Hao Zhang, and Hui Huang. 2020. TAP-Net: Transport-and-Pack Using Reinforcement Learning. *ACM Trans. Graph.* 39, 6, Article 232 (nov 2020), 15 pages. <https://doi.org/10.1145/3414685.3417796>
- Qi-Xing Huang, Simon Flöry, Natasha Gelfand, Michael Hofer, and Helmut Pottmann. 2006. Reassembling fractured objects by geometric matching. *ACM Trans. Graph.* 25, 3 (jul 2006), 569–578. <https://doi.org/10.1145/1141911.1141925>
- Sichao Huang, Ziwei Wang, Jie Zhou, and Jiwen Lu. 2023. Planning Irregular Object Packing via Hierarchical Reinforcement Learning. *IEEE Robotics and Automation Letters* 8, 1 (2023), 81–88. <https://doi.org/10.1109/LRA.2022.3222996>
- Bonfim A. Junior, Plácido R. Pinheiro, and Rommel D. Saraiva. 2013. A Hybrid Methodology for Tackling the Irregular Strip Packing Problem. *IFAC Proceedings Volumes* 46, 7 (May 2013), 396–401. <https://doi.org/10.3182/20130522-3-br-4036.00041>
- Thomas N. Kipf and Max Welling. 2017. Semi-Supervised Classification with Graph Convolutional Networks. In *International Conference on Learning Representations*. <https://openreview.net/forum?id=SUJ4ayYgI>
- P.E. Kloeden and E. Platen. 2011. *Numerical Solution of Stochastic Differential Equations*. Springer Berlin Heidelberg. <https://books.google.com/books?id=BCvtssom1CMC>
- Aline A.S. Leao, Franklina M.B. Toledo, José Fernando Oliveira, Maria Antónia Carravilla, and Ramón Alvarez-Valdés. 2020. Irregular packing problems: A review of mathematical models. *European Journal of Operational Research* 282, 3 (may 2020), 803–822. <https://doi.org/10.1016/j.ejor.2019.04.045>
- Junhyun Lee, Inyeop Lee, and Jaewoo Kang. 2019. Self-Attention Graph Pooling. In *Proceedings of the 36th International Conference on Machine Learning (Proceedings of Machine Learning Research, Vol. 97)*, Kamalika Chaudhuri and Ruslan Salakhutdinov (Eds.). PMLR, 3734–3743. <https://proceedings.mlr.press/v97/lee19c.html>
- Weiyu Liu, Yilun Du, Tucker Hermans, Sonia Chernova, and Chris Paxton. 2023. Struct-Diffusion: Language-Guided Creation of Physically-Valid Structures using Unseen Objects. In *RSS 2023*.
- Eunice Lopez, Gabriela Ochoa, Hugo Terashima-Marín, and Edmund Burke. 2013. An effective heuristic for the two-dimensional irregular bin packing problem. *Annals of Operations Research* 206 (07 2013), 241–264. <https://doi.org/10.1007/s10479-013-1341-4>
- Ilya Loshchilov and Frank Hutter. 2019. Decoupled Weight Decay Regularization. In *International Conference on Learning Representations*. <https://openreview.net/forum?id=Bkg6RiCqY7>
- Jia-Hui Pan, Ka-Hei Hui, Xiaojie Gao, Shize Zhu, Yun Liu, Pheng-Ann Heng, and Chi-Wing Fu. 2023. SDF-Pack: Towards Compact Bin Packing with Signed-Distance-Field Minimization. In *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 10612–10619. <https://doi.org/10.1109/IROS55552.2023.10341940>
- Guocheng Qian, Yuchen Li, Houwen Peng, Jinjie Mai, Hasan Hammoud, Mohamed Elhoseiny, and Bernard Ghanem. 2022. PointNeXt: Revisiting PointNet++ with Improved Training and Scaling Strategies. In *Advances in Neural Information Processing Systems (NeurIPS)*.
- Jack Qiao. 2017. SVGnest. <https://github.com/Jack000/SVGnest>
- T. Romanova, J. Bennell, Y. Stoyan, and A. Pankratov. 2018. Packing of concave polyhedra with continuous rotations using nonlinear optimisation. *European Journal of*

- Operational Research* 268, 1 (2018), 37–53. <https://doi.org/10.1016/j.ejor.2018.01.025>
- Silvia Sellán, Yun-Chun Chen, Ziyi Wu, Animesh Garg, and Alec Jacobson. 2022. Breaking Bad: A Dataset for Geometric Fracture and Reassembly. In *Thirty-sixth Conference on Neural Information Processing Systems Datasets and Benchmarks Track*. <https://openreview.net/forum?id=mJWt6pOcHNy>
- Mohamed A. Shalaby and Mohamed Kashkoush. 2013. A Particle Swarm Optimization Algorithm for a 2-D Irregular Strip Packing Problem. *American Journal of Operations Research* 03, 02 (2013), 268–278. <https://doi.org/10.4236/ajor.2013.32024>
- Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. 2021. Score-Based Generative Modeling through Stochastic Differential Equations. In *International Conference on Learning Representations*. <https://openreview.net/forum?id=PxTIG12RRHS>
- Ernesto Lo Valvo. 2017. Meta-heuristic Algorithms for Nesting Problem of Rectangular Pieces. *Procedia Engineering* 183 (2017), 291–296. <https://doi.org/10.1016/j.proeng.2017.04.041> 17th International Conference on Sheet Metal, SHEMET17.
- Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is All you Need. In *Advances in Neural Information Processing Systems*, I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (Eds.), Vol. 30. Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ec243547dee91fb0d053c1c4a845aa-Paper.pdf
- Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. 2018. Graph Attention Networks. In *International Conference on Learning Representations*. <https://openreview.net/forum?id=rJXMpikCZ>
- Pascal Vincent. 2011. A connection between score matching and denoising autoencoders. *Neural computation* 23, 7 (2011), 1661–1674.
- Ziqi Wang, Peng Song, and Mark Pauly. 2021. State of the Art on Computational Design of Assemblies with Rigid Parts. *Computer Graphics Forum* 40, 2 (may 2021), 633–657. <https://doi.org/10.1111/cgf.142660>
- QiuHong Anna Wei, Sijie Ding, Jeong Joon Park, Rahul Sajnani, Adrien Poulenard, Srinath Sridhar, and Leonidas Guibas. 2023. LEGO-Net: Learning Regular Rearrangements of Objects in Rooms. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE. <https://doi.org/10.1109/cvpr52729.2023.01825>
- Mingdong Wu, Fangwei Zhong, Yulong Xia, and Hao Dong. 2022. TarGF: Learning Target Gradient Field to Rearrange Objects without Explicit Goal Specification. In *Advances in Neural Information Processing Systems*, S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh (Eds.), Vol. 35. Curran Associates, Inc., 31986–31999. https://proceedings.neurips.cc/paper_files/paper/2022/file/cf5a019ae9c11b4be88213ce3f85d85c-Paper-Conference.pdf
- Juzhan Xu, Minglun Gong, Hao Zhang, Hui Huang, and Ruizhen Hu. 2023. Neural Packing: from Visual Sensing to Reinforcement Learning. *ACM Trans. Graph.* 42, 6, Article 267 (dec 2023), 11 pages. <https://doi.org/10.1145/3618354>
- Tianyang Xue, Mingdong Wu, Lin Lu, Haoxuan Wang, Hao Dong, and Baoquan Chen. 2023. Learning Gradient Fields for Scalable and Generalizable Irregular Packing. In *SIGGRAPH Asia 2023 Conference Papers* (, Sydney, NSW, Australia.) (SA '23). Association for Computing Machinery, New York, NY, USA, Article 105, 11 pages. <https://doi.org/10.1145/3610548.3618235>
- Zeshi Yang, Zherong Pan, Manyi Li, Kui Wu, and Xifeng Gao. 2023. Learning Based 2D Irregular Shape Packing. *ACM Trans. Graph.* 42, 6, Article 266 (dec 2023), 16 pages. <https://doi.org/10.1145/3618348>
- Jonathan Young. 2022. XAtlas. <https://github.com/jpcy/xatlas>.
- Jingwei Zhang, Bin Zi, and Xiaoyu Ge. 2021. Attend2Pack: Bin Packing through Deep Reinforcement Learning with Attention. *arXiv e-prints*, Article arXiv:2107.04333 (July 2021), arXiv:2107.04333 pages. <https://doi.org/10.48550/arXiv.2107.04333> arXiv:2107.04333 [cs.LG]
- Hang Zhao, Zherong Pan, Yang Yu, and Kai Xu. 2023. Learning Physically Realizable Skills for Online Packing of General 3D Shapes. *ACM Trans. Graph.* 42, 5, Article 165 (jul 2023), 21 pages. <https://doi.org/10.1145/3603544>
- Hang Zhao, Qijin She, Chenyang Zhu, Yin Yang, and Kai Xu. 2021. Online 3D Bin Packing with Constrained Deep Reinforcement Learning. *Proceedings of the AAAI Conference on Artificial Intelligence* 35, 1 (May 2021), 741–749. <https://doi.org/10.1609/aaai.v35i1.16155>
- Hang Zhao, Yang Yu, and Kai Xu. 2022. Learning Efficient Online 3D Bin Packing on Packing Configuration Trees. In *International Conference on Learning Representations*. <https://openreview.net/forum?id=bfuGjlCwAq>

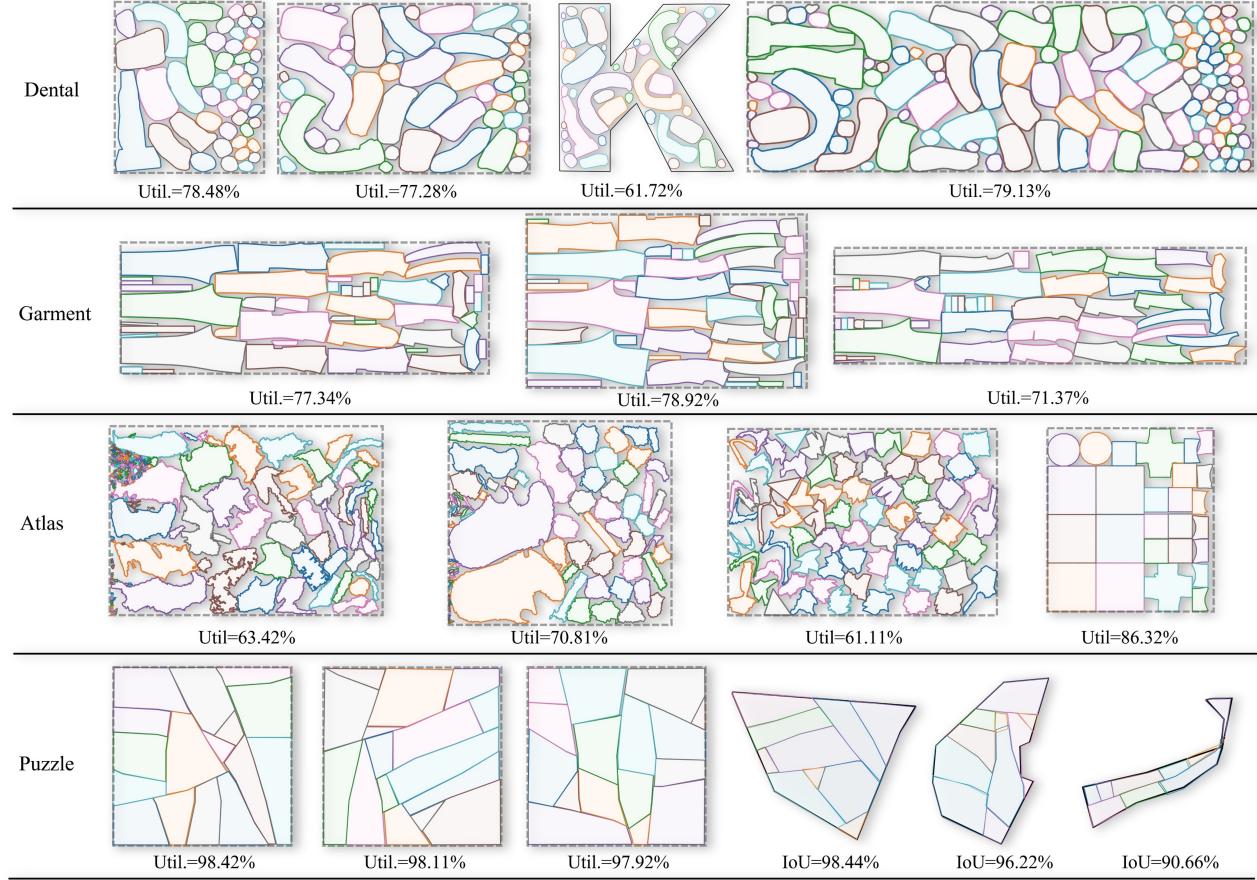


Fig. 11. Some results from GFPack++ across different datasets. Note that the utilization enhancement only applies to regular boundaries, so small overlaps may be observed in the puzzle (arbitrary) results.

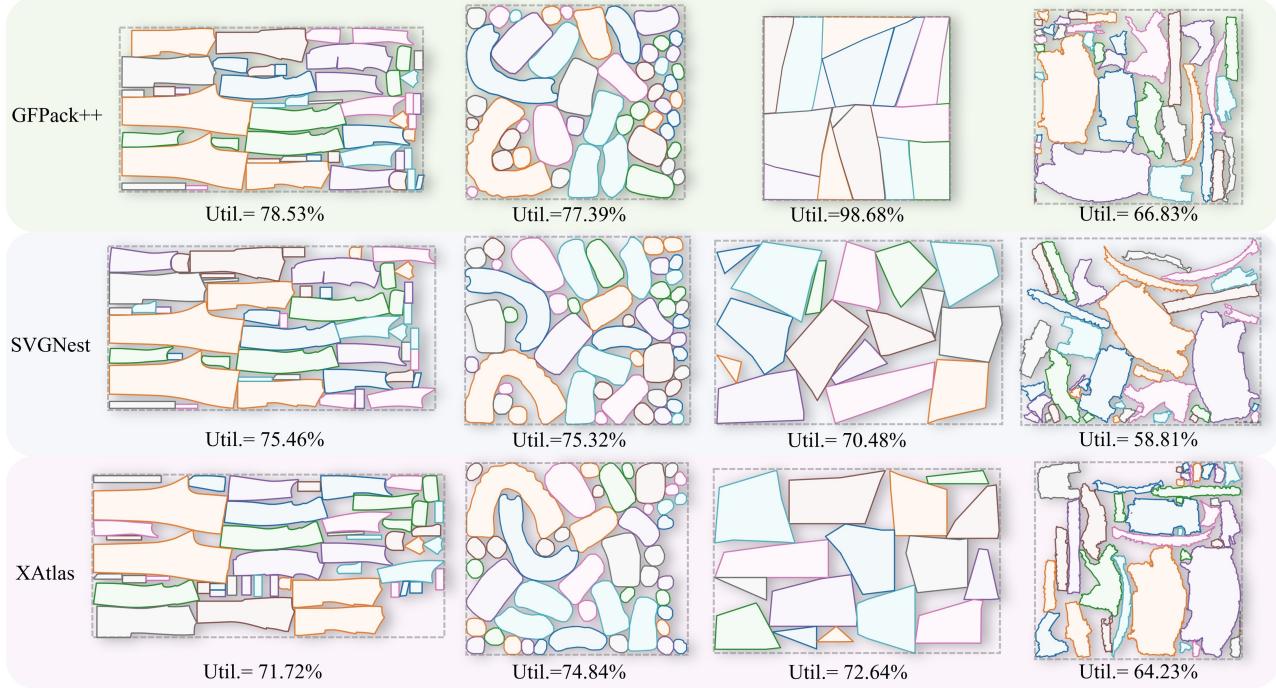


Fig. 12. Comparative results from GFPack++, SVGnest, and XAtlas applied to Garment, Dental, Puzzle, and Atlas datasets. In the lower left corner of the Garment dataset, a shared similar local layout pattern between GFPack++ and SVGnest can be observed. For the Puzzle dataset, GFPack++ outperforms the others, nearly achieving a global optimum due to its continuous rotation solving.

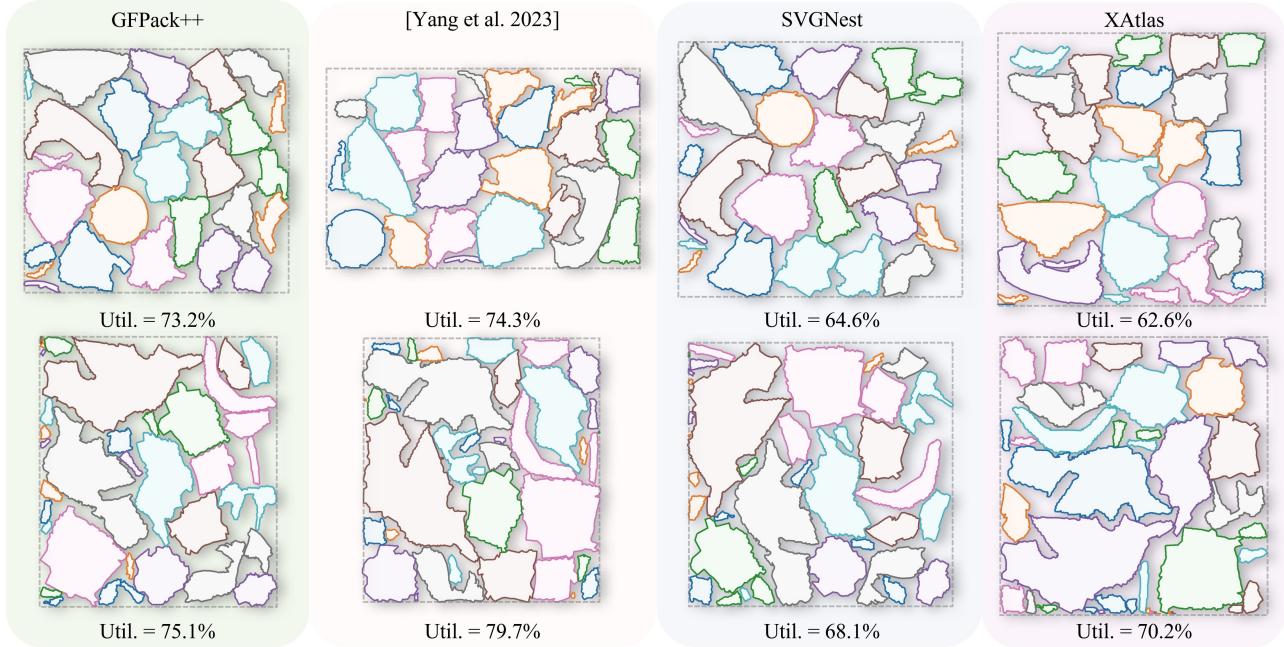


Fig. 13. Comparative results from GFPack++, [Yang et al. 2023], SVGnest, and XAtlas applied to Atlas datasets. The results of [Yang et al. 2023] are taken directly from the paper.

A DETAILS OF DATASETS

The distribution of our training data is described in Table 8. The distribution of vertex numbers in the polygons of each dataset is presented in Table 9 and Fig. 14. Every polygon in the Puzzle dataset is unique. Therefore, we randomly sampled 1000 polygons from the Puzzle dataset to calculate the vertex counts. The number of polygons in each dataset is displayed in Table 10. We also provide visualizations of the polygons in the ‘data/Datasets’ folder. Some generated reutls are presented in ‘data/Validations’ folder.

B TRAINING SETTINGS

We conducted our training and testing on a Linux server equipped with 4 NVIDIA RTX 4090 GPUs and an Intel(R) Xeon(R) Platinum 8163 CPU. GFPACK++ converged in 50 hours. GFPACK converged after 168 hours of training. We continued training for an additional 48 hours after convergence for both methods.

We quadrupled the number of parameters in GFPACK from 10(M) to match GFPACK++ at 40(M). This increase resulted in excessive VRAM usage and extended convergence times, with the model not converging within 240 hours. Despite this, GFPACK still failed to produce collision-free results on datasets with rotations. We did not further increase the number of training parameters since the training time was already too long.

C BASELINE ALGORITHMS

We modified the code of XAtlas to enable its application to non-atlas packing problems. Key modifications include: (a) the introduction of a packing height constraint to facilitate strip packing operations; (b) adjustments to padding, alignment, and scaling methods to prevent distortion. Detailed XAtlas settings are provided here:

```
s_atlas.options.pack.height = height;
s_atlas.options.pack.rotation = true;
s_atlas.options.pack.rotateChartsToAxis = align;
s_atlas.options.pack.bruteForce = true;
s_atlas.options.pack.padding = 0.5f;
s_atlas.options.pack.bilinear = true;
s_atlas.options.pack.blockAlign = false;
s_atlas.options.scale = 1.0f;
```

In this context, ‘height’ specifies the stripe height, with a setting of -1 indicating unfixed boundaries. XAtlas supports rotations of 0 and 90 degrees when both brute force and rotation are enabled. The align parameter determines whether polygons are aligned to their bounding boxes. We have disabled this setting for the dental dataset, while it is enabled for other datasets.

Additionally, for SVGNest, we have enhanced its parallel efficiency to improve its performance within the same time.

Table 8. Teacher dataset utilization distributions.

Teacher dataset	Min	Avg	Max
Garment	63.02%	72.39%	78.90%
Dental	67.33%	74.16%	77.36%
Atlas (building)	39.92%	74.55%	95.74%
Atlas (object)	24.69%	63.25%	85.18%
Atlas (general)	40.76%	67.54%	92.41%

Table 9. Polygon Vertex counts.

Dataset	Min	Avg	Max	Std
Garment	5	81.57	284	76.83
Dental	17	34.47	123	14.50
Puzzle	4	8.25	29	2.86
Atlas (building)	4	6.52	70	3.87
Atlas (object)	4	67.41	1120	102.38

Table 10. Polygon counts in each dataset.

Dataset	Polygon Count
Garment	313
Dental	440
Atlas (building)	3262
Atlas (object)	6764

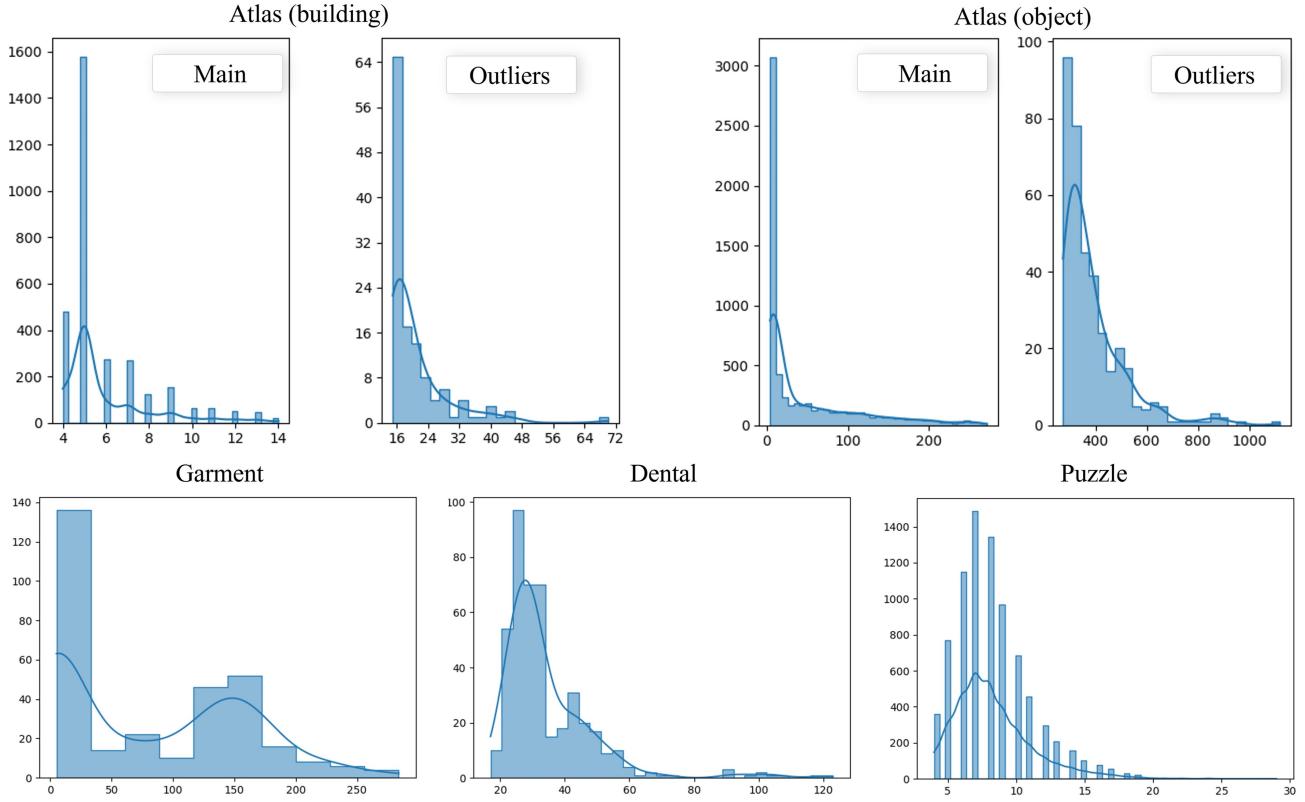


Fig. 14. The statistics of polygon vertex counts. The horizontal axis represents the number of vertices in a polygon, while the vertical axis indicates the frequency of these vertex counts. Due to the wide distribution of vertex counts in the Atlas dataset, we have divided it into two parts: the main distribution and the outliers.