

# Reinforcement learning for automatic quadrilateral mesh generation: a soft actor-critic approach

Jie Pan<sup>a</sup>, Jingwei Huang<sup>b</sup>, Gengdong Cheng<sup>c</sup>, Yong Zeng<sup>a,\*</sup>

<sup>a</sup>*Concordia Institute for Information Systems Engineering, Concordia University, Montreal, H3G 1M8, Quebec, Canada*

<sup>b</sup>*Department of Engineering Management & Systems Engineering, Old Dominion University, Norfolk, 23529, Virginia, United States*

<sup>c</sup>*Department of Engineering Mechanics, Dalian University of Technology, Dalian, 116023, Liaoning, China*

## Abstract

This paper proposes, implements, and evaluates a reinforcement learning (RL)-based computational framework for automatic mesh generation. Mesh generation plays a fundamental role in numerical simulations in the area of computer aided design and engineering (CAD/E). It is identified as one of the critical issues in the NASA CFD Vision 2030 Study. Existing mesh generation methods suffer from high computational complexity, low mesh quality in complex geometries, and speed limitations. These methods and tools, including commercial software packages, are typically semiautomatic and they need inputs or help from human experts. By formulating the mesh generation as a Markov decision process (MDP) problem, we are able to use a state-of-the-art reinforcement learning (RL) algorithm called “soft actor-critic” to automatically learn from trials the policy of actions for mesh generation. The implementation of this RL algorithm for mesh generation allows us to build a fully automatic mesh generation system without human intervention and any extra clean-up operations, which fills the gap in the existing mesh generation tools. In the experiments to compare with two representative commercial software packages, our system demonstrates promising performance with respect to scalability, generalizability, and effectiveness.

*Keywords:* Reinforcement learning, mesh generation, soft actor-critic,

\*Corresponding author

Email address: [yong.zeng@concordia.ca](mailto:yong.zeng@concordia.ca) (Yong Zeng )

## 1. Introduction

Reinforcement learning (RL) has been researched and applied in many fields, such as games (Silver et al., 2016), health care (Gottesman et al., 2019), natural language processing (Wang et al., 2018b), and combinatorial optimization, particularly for NP-hard problems (Mazyavkina et al., 2021). However, it has rarely been applied to the area of computational geometry, especially in the field of mesh generation (Pan et al., 2021). Mesh generation is a fundamental step in conducting numerical simulations in the area of Computer-Aided Design and Engineering (CAD/E) such as finite element analysis (FEA), computational fluid dynamics (CFD), and graphic model rendering (Gordon and Hall, 1973; Yao and Stillman, 2019). Mesh generation techniques have been identified as one of the six basic capacities to build in NASA’s Vision 2030 CFD Study (Slotnick et al., 2014).

Mesh generation discretizes a complex geometric domain (see Fig. 1 (a)) into a finite set of (geometrically simple and bounded) elements, such as 2D triangles or quadrilaterals (see Fig. 1 (b) in 2D geometries) and tetrahedra or hexahedra (in 3D geometries). Since the reliable automation and high quality of mesh representation matter significantly to the numerical simulation results, mesh generation has continued to be a significant bottleneck in those fields due to algorithm complexities, inadequate error estimation capabilities, and complex geometries (Slotnick et al., 2014).

Strongly favored by engineering analysis, quadrilateral mesh generation has been a critical research problem for decades. However, existing quadrilateral mesh generation methods require heuristic knowledge in algorithm development and severely depend on preprocessing or postprocessing operations to maintain high mesh quality. Preprocessing includes decomposing complex domains into multiple regular components (Liu et al., 2017) and generating favorable vertex locations (Remacle et al., 2013). Postprocessing (i.e., clean-up operations) is used to handle where the mesh consists of elements other than quadrilaterals (e.g., triangular elements), flat or inverted elements, and irregularly connected elements. The operations include reducing the singularity (Verma and Suresh, 2017), performing iterative topological changes (e.g., splitting, swapping, and collapsing elements) (Docampo-Sanchez and Haines, 2019), and mesh adaptation (Verma and Suresh, 2018).

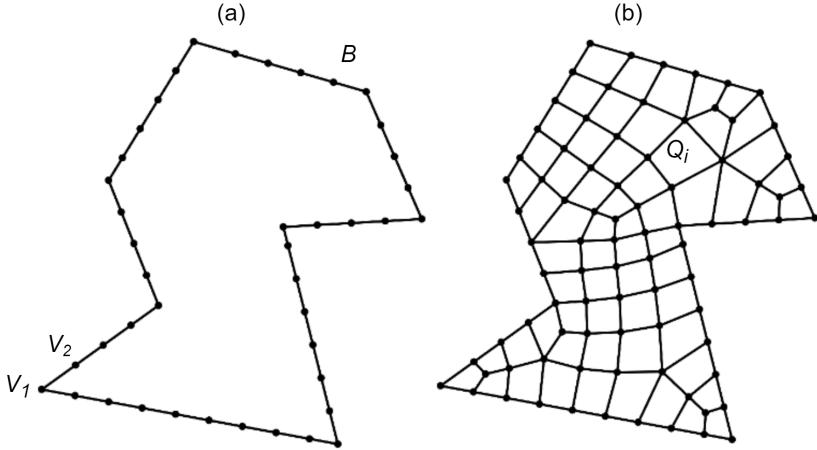


Figure 1: Meshing problem. (a) The initial geometric boundary,  $B$ , is composed of piecewise linear segments, represented as a sequence of vertices  $[V_1, V_2, \dots, V_N]$ ; (b) The final mesh,  $\Omega$ , is composed of a set of quadrilateral elements  $[Q_1, Q_2, \dots, Q_M]$ .

Improving mesh quality with extra operations increases computational complexity and slows the meshing speed. Many machine learning-based methods are proposed to support mesh optimization and reconstruction (Zhang et al., 2020; Yang et al., 2021; Gupta, 2020; Chen et al., 2018; Wang et al., 2018a; Defferrard et al., 2016). Nechaeva (2006) proposed an adaptive mesh generation algorithm based on self-organizing maps (SOMs), which adapts a given uniform mesh onto a target physical domain through mapping. Pointer networks can generate triangular meshes, but the result is not robust and contains intersecting edges (Vinyals et al., 2015). Papagiannopoulos et al. (2021) proposed a triangular mesh generation method with three neural networks. The training data came from the constrained Delaunay triangulation algorithm (Chew, 1989). The trained model could predict the number of candidate inner vertices (to form a triangular element), the coordinates of those vertices, and their connective relations with existing segments on the boundary. Adapting to arbitrary and complex geometries is a shortcoming because of the fixed input scale and constrained diversity of training samples.

To achieve generalizability for arbitrarily shaped geometries, Zeng and Cheng (1993) used a recursive algorithm to generate one element at a time by three primitive rules to avoid taking the whole contour as the input. Yao et al. (2005) improved the recursive method by using an artificial neural network (ANN) to automate the rule learning. However, the quality and diversity of

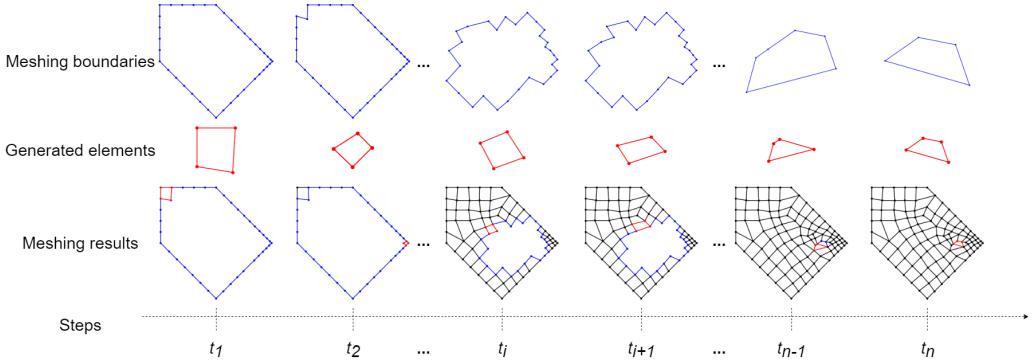


Figure 2: A sequence of actions taken by the mesh generator to complete a mesh. At each time step  $t_i$ , an element (in red) is extracted from the current boundary (in blue). The boundary is then updated by cutting off the element and serves as the meshing boundary in the next time step  $t_{i+1}$ . This process continues until the updated boundary becomes an element.

training data cannot be achieved and limit the model performance.

To resolve this issue, Pan et al. (2021) formulated the mesh generation algorithm by Zeng and Cheng (1993) into a Markov decision problem (MDP). The formalized problem framework was addressed by reinforcement learning (Sutton and Barto, 1998, 2018), as illustrated in Fig. 2. At each time step, a quadrilateral element is generated from the domain boundary. The boundary is inwardly updated by cutting off the generated element. In each iteration, the meshing boundary evolves into a new boundary to generate new elements, which is naturally a sequential decision-making process. In the previous work (Pan et al., 2021), the present authors only used the RL method to sample training data for a feedforward neural network model, which did not fully exploit the potential of the RL models and resulted in an extra sampling selection phase. This heuristic selection can easily cause an imbalanced dataset and compromise model performance.

The present article resolves the abovementioned problem and provides an RL-based computational framework, FreeMesh-RL, for quadrilateral mesh generation that automatically collects balanced samples through trial-and-error learning, as illustrated in Fig. 3. It aims to automatically provide high-quality meshes for various complex geometries without human intervention. A few challenges must be addressed: 1) the RL method needs to be robust and less hyperparameter-sensitive; 2) the trade-off between the current element quality and the remaining boundary quality shall be made to maintain overall

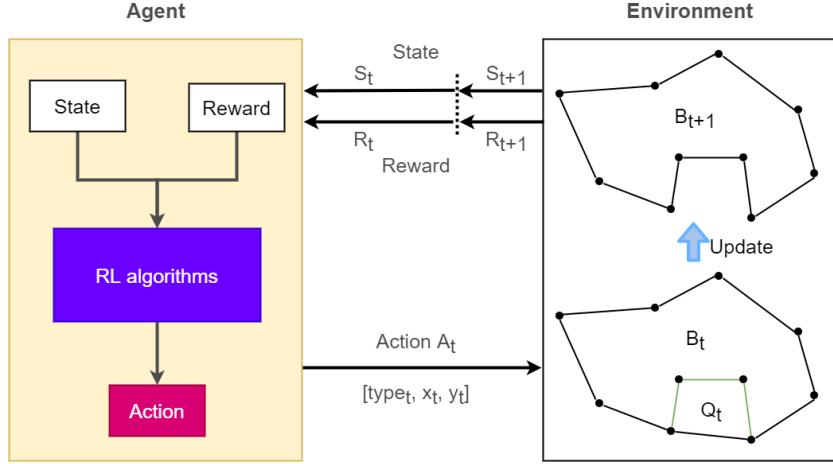


Figure 3: The RL-based computational framework for automatic mesh generation. The agent acts as the mesh generator by implementing various RL techniques. It generates an element after a state is perceived and improves the action by observing the rewards. The environment models the meshing boundary and updates the boundary by cutting off the element generated via the action.

high mesh quality; and 3) the meshing process should be finished in finite steps. Compared with other RL algorithms (Lillicrap et al., 2015; Mnih et al., 2016; Schulman et al., 2017; Fujimoto et al., 2018), soft actor-critic (SAC) (Haarnoja et al., 2018a,b) is selected to address the mesh generation problem, because of its capability allowing reuse of previous experience, good balance between exploration and exploitation, stable learning efficiency, and being less hyperparameter-sensitive.

The rest of the present paper is organized as follows. Section 2 introduces the detailed formulation of mesh generation as an RL problem, including action, state, and reward, and proposes an RL-based meshing architecture with SAC. Section 3 explores the implementation details of each key concepts in applying SAC to mesh generation, and evaluates the performance of the proposed method in comparison with two state-of-the-art meshing approaches that share the similar algorithmic strategy yet with the rules and knowledge developed based on expert experience. Section 4 discusses the main findings based on the experiment results and explains how the method can be applied to a broader area both in RL and mesh generation. Section 5 concludes this article.

## 2. RL based mesh generation

This section proposes an RL-based computational framework for automatic mesh generation. The action formulation and state representation are detailed first. A reward function is then designed to meet mesh quality requirements. Finally, the architecture of the proposed framework is explained.

### 2.1. Action formulation

In the mesh generation process (see Fig. 2), two important decisions should be made: 1) how to choose a vertex (called the reference vertex in this paper) from the boundary; and 2) how to decide on the other three vertices to form a quadrilateral element.

The first decision is to select a local reference region that has the least boundary angle from the boundary. It reduces the formation of narrow regions in the remaining boundary, and thus increases high-quality elements generated by subsequent actions. The reference vertex  $V_i^*$  is selected using the following equation:

$$V_i^* = \arg \min_{V_i} \frac{1}{n_{rv}} \sum_{j=1}^{n_{rv}} \angle V_{l,j} V_i V_{r,j}, i \in N_B, \quad (1)$$

where  $N_B$  is the number of vertices contained in the present boundary  $B$ ;  $V_{l,j}$  and  $V_{r,j}$  denote the  $j$ -th vertices at the left and right side (considering clockwise orientation) of the reference vertex  $V_i$  along the boundary, respectively;  $n_{rv}$  represents how many surrounding vertices should be included; the  $V_i$  is the  $i$ -th vertex on the boundary. An example of this selection is shown in Fig. 4. The detailed selection is discussed in the experiment section.

The second decision is to choose an correct action to form a quadrilateral element from three basic action types: adding zero, one, or two more vertices (Zeng and Cheng, 1993). As shown in Fig. 5, the action is represented by  $[type, V_1, V_2]$ , where  $type \in \{0, 1, 2\}$ , corresponding to the three possible actions;  $V_1$  and  $V_2$  are the coordinates of the newly added vertices. The coordinate space for the vertices is constrained to a fan-shaped area (in light blue) with radius  $r$ , which is calculated as follows:

$$\begin{aligned} r &= \alpha * L, \\ L &= \frac{1}{2n} \sum_{j=0}^n |V_{l,j} V_{l,j+1}| + |V_{r,j} V_{r,j+1}|, 0 < n < N/2 \end{aligned} \quad (2)$$

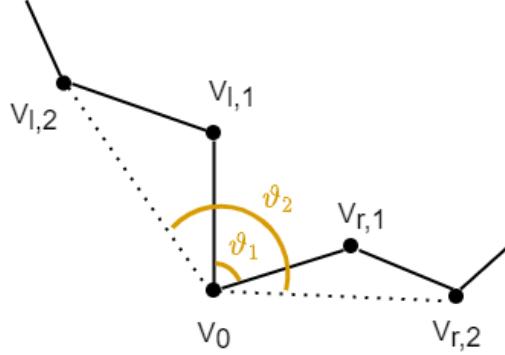


Figure 4: The angle calculation in reference vertex selection. For example, two surrounding vertices of  $V_0$ ,  $n_{rv} = 2$ , are used to calculate the angles  $\vartheta_1$  and  $\vartheta_2$ , where  $\vartheta_1 = \angle V_{l,1}V_0V_{r,1}$ ,  $\vartheta_2 = \angle V_{l,2}V_0V_{r,2}$ . This calculation iterates along the boundary. The vertex with the least averaged angle by  $\vartheta_1$  and  $\vartheta_2$  is selected as the reference vertex.

where  $\alpha$  is a factor to amplify the base length  $L$ ;  $V_{l,j}$  and  $V_{r,j}$  denote the  $j$ -th vertex at the left and right side of the reference vertex along the boundary;  $V_{l,0} = V_i = V_{r,0}$  is the reference vertex; and  $|V_aV_b|$  is the Euclidean distance between two vertices. Usually, the action  $type = 2$  is needed only on special occasions (e.g., circular domains) at limited times. It is implemented in the environment if these situations occur. The action  $[type, V_1]$  is eventually used in this article.

## 2.2. State representation

A state is an observation of the environment by the agent. The environment here is the meshing boundary (see Fig. 2). The full information for the state of the environment at time  $t$  consists of all the vertices along the boundary. However, not all the vertices are relevant and necessary to deciding what action to take. Therefore, a partial observation of the boundary environment is designed, which consists of a reference vertex determining where the agent should start generating an element, and its surrounding vertices providing a local environment.

A state at time  $t$  is denoted as  $s_t$ , as shown in Fig. 6, and is composed of five components: (1) a reference vertex,  $V_0$ , which is used as the relative origin to generate the new element with an action  $a_t$ , and is calculated by Equation 1; (2)  $n$  neighboring vertices on the right side of the reference vertex; (3)  $n$  neighboring vertices on the left side; (4)  $g$  neighboring points  $V_{\zeta_1}, \dots, V_{\zeta_g}$ , the closest vertices in the fan-shaped area  $\zeta_1, \dots, \zeta_g$  within a radius, which is

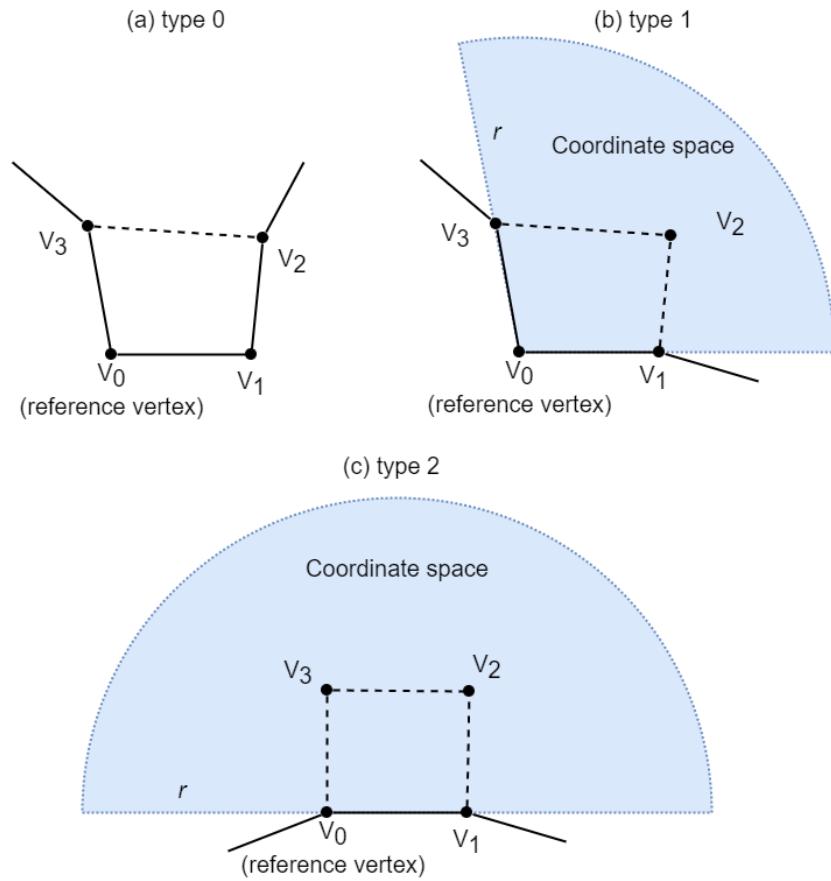


Figure 5: Action space for each rule type. Subfigures (a)-(c) correspond to three types of actions, respectively. The blue area is the space to select the candidate vertices with a radius  $r$  and the origin vertex  $V_0$  (reference vertex). The candidate vertices are  $V_2$  (in type 1), and  $V_2$  and  $V_3$  (in type 2), respectively.

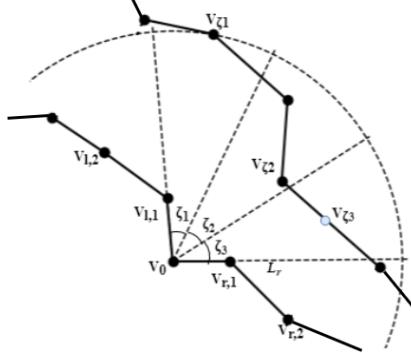


Figure 6: Partial observation of the meshing boundary. For example, the partial boundary, where  $\beta = 4, n = 2, g = 3$ , is represented as the state. First, two vertices on the left and right sides of the reference vertex  $V_0$  are selected. Second, the angle  $\angle V_{l,1}V_0V_{r,1}$  is evenly split into three angles  $\zeta_1, \zeta_2$ , and  $\zeta_3$ ; three fan-shaped areas are hence formed with these angles and a radius  $L_r = 4 * L$ . Then, the closest vertex in each area is selected.  $V_{\zeta_3}$  is an intersection vertex between the bisector and the boundary segment.

calculated as:

$$L_r = \beta * L, \quad (3)$$

where  $\beta$  is a factor to amplify the base length and  $L$  is calculated by Equation 2. When  $g = 3$ , the fan-shaped area is evenly divided into three parts,  $\zeta_1 = \zeta_2 = \zeta_3$ , as shown in Fig. 6. If there are no vertices in a sliced fan-shaped area (e.g.,  $\zeta_3$ ), the furthest bisector vertex in the slice or the intersection vertex between the bisector of the slice and the boundary edge is selected, such as  $V_{\zeta_3}$  in Fig. 6; and (5)  $\rho_t$ , the area ratio between the updated domain and the original domain, which is used to indicate the meshing progress.

The current state  $S_t$ , representing the partial boundary around the reference point and meshing progress, is denoted as follows:

$$S_t = \{V_{l,n}, \dots, V_{l,1}, V_0, V_{r,1}, \dots, V_{r,n}, V_{\zeta_1}, \dots, V_{\zeta_g}, \rho_t\}. \quad (4)$$

All the vertices are represented by a polar coordinate system with  $V_0$  as the origin and  $\overrightarrow{V_0V_{r,1}}$  as the reference direction. There are a few geometrical operations (i.e., rotation, scaling, and transit) to keep only the relative information of the contained vertices (Yao et al., 2005).

### 2.3. Reward function

The criteria for high-quality meshing are as follows: 1) each element is a quadrilateral; (2) each element should be as close to a square as possible, and

minimally, the inner corners of each element should be between 45° and 135°; (3) the aspect ratio (the ratio of opposite edges) and taper ratio (the ratio of neighboring edges) of each quadrilateral should be within a predefined range; and (4) the transition between a dense mesh and a coarse mesh should be smooth (Zeng and Cheng, 1993; Zeng and Yao, 2009).

The reward function shall guarantee the quality requirements and step-wisely measure the performance of each action. The action can cause three situations: 1) if it forms an invalid element or intersects with the boundary, the reward is set to -0.1; 2) if it generates the last element, the reward is set to 10; and 3) if it constructs a valid element, the reward is a joint measurement of element quality, the quality of the remaining boundary, and the density. Consequently, the reward function is represented as follows:

$$r_t(s_t, a_t) = \begin{cases} -0.1, & \text{invalid element;} \\ 10, & \text{the element is the last element;} \\ m_t, & \text{otherwise.} \end{cases} \quad (5)$$

The measurement  $m_t$  is calculated by the following equation:

$$m_t = \eta_t^e + \eta_t^b + \mu_t. \quad (6)$$

The element quality  $\eta_t^e$  is measured by its edges and internal angles, and is calculated as follows:

$$\begin{aligned} \eta_t^e &= \sqrt{q^{edge} q^{angle}}, \\ q^{edge} &= \frac{\sqrt{2} \min_{j \in \{0,1,2,3\}} \{l_j\}}{D_{max}}, \\ q^{angle} &= \frac{\min_{j \in \{0,1,2,3\}} \{angle_j\}}{\max_{j \in \{0,1,2,3\}} \{angle_j\}}, \end{aligned} \quad (7)$$

where  $q^{edge}$  refers to the quality of edges of this element;  $l_j$  is the length of the  $j$ th edge of the element;  $D_{max}$  is the length of the longest diagonal of the  $t$ th element;  $q^{angle}$  refers to the quality of the angles of the element; and  $angle_j$  is the degree of the  $j$ th inner angle of the element. The quality  $\eta_t^e$  will range from 0 to 1, which is the greater the better. Examples of various element qualities are shown in Fig. 7.

The quality of the remaining boundary  $\eta_t^b$  is measured by both the quality of the angles formed between the newly generated element and the boundary,

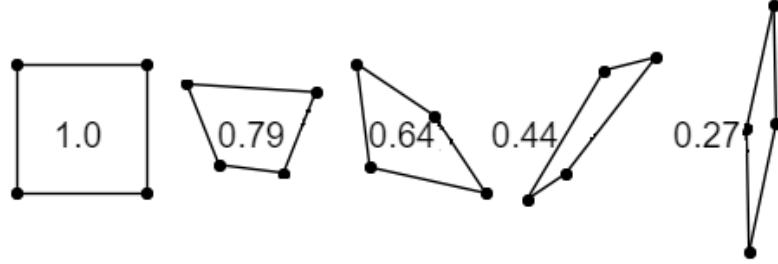


Figure 7: Element quality varies with different shapes. The quality value ranges from 0 to 1. The element with the best quality 1 is a square.

and the shortest distance of the generated vertex to its surrounding edges, and is denoted as follows:

$$\begin{aligned} \eta_t^b &= \sqrt{\frac{\min_{k \in \{1,2\}} \{\min(\varsigma_k, M_{angle})\}}{M_{angle}} q^{dist}} - 1, \\ q^{dist} &= \begin{cases} \frac{d_{min}}{(d_1 + d_2)/2}, & \text{if } d_{min} < (d_1 + d_2)/2; \\ 1, & \text{otherwise.} \end{cases} \end{aligned} \quad (8)$$

where  $\varsigma_k$  refers to the degrees of the  $k$ th generated angle and  $d_{min}$  is the distance of  $V_2$  to its closest edge. The details are shown in Fig. 8. The quality  $\eta_t^b$  ranges from -1 to 0, with a larger value representing better quality. It serves as a penalty term to decrease the reward if the quality of the remaining boundary worsens. We set  $M_{angle} = 60^\circ$ . When the formed new angles are less than  $M_{angle}$ , the quality will decrease. It penalizes the generation of sharp angles that are harmful to the overall mesh quality and may even fail the meshing process.

These two qualities together represent the trade-off between the generated element and the remaining boundary, ensuring the overall mesh quality. The last term, density  $\mu_t$ , secures the completion of the meshing process within finite steps by controlling mesh density, which is calculated as follows:

$$\mu_t = \begin{cases} -1, & \text{if } \mathcal{A}_t < \mathcal{A}_{min}; \\ \frac{\mathcal{A}_t - \mathcal{A}_{min}}{\mathcal{A}_{max} - \mathcal{A}_{min}}, & \text{if } \mathcal{A}_{min} \leq \mathcal{A}_t < \mathcal{A}_{max}; \\ 0, & \text{otherwise.} \end{cases} \quad (9)$$

where  $\mathcal{A}_t$  is the area of the element generated at time  $t$ ;  $\mathcal{A}_{min}$  is the estimated minimum area that an element should meet, and is calculated by

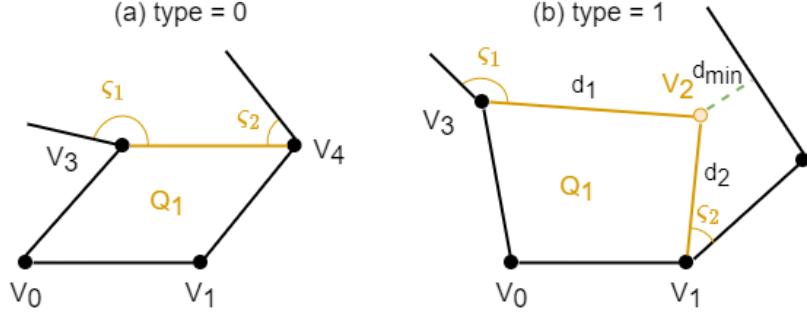


Figure 8: The quality of the remaining boundary for two action types.  $Q_1$  is the newly generated element. Once it is removed, it forms two angles  $s_1$  and  $s_2$  with existing boundary segments. (a) When action type = 0, the boundary quality only hinges on these two angles, assuming  $q^{dist} = 1$ . (b) When action type = 1, the boundary quality is jointly measured by the two angles, and the closest Euclidean distance  $d_{min}$  of the newly added vertex  $V_2$  to the existing segments.  $d_1$  and  $d_2$  are the Euclidean lengths of segments  $V_2V_3$  and  $V_1V_2$ .

$\mathcal{A}_{min} = v \cdot e_{min}^2$ ;  $\mathcal{A}_{max}$  is the estimated maximum area of the element, and is calculated by  $\mathcal{A}_{max} = v \left( \frac{e_{max}-e_{min}}{\kappa} + e_{min} \right)^2$ ;  $e_{max}$  and  $e_{min}$  are the lengths of the longest and shortest edges in the boundary, respectively;  $\kappa$  adjusts the estimated maximum area and is independent of the unit of edges ( $\kappa = 4$  in our experiments); and  $v$  is a weight and values in  $(0, 10]$ , for which a smaller value means a greater density, and vice versa. We set  $v = 1$  in our experiments for the medium density.

#### 2.4. Meshing scheme via SAC

The formulated RL-based meshing architecture, FreeMesh-RL, is shown in Fig. 3. We formulate meshing process as an MDP process, consisting of a set of boundary environment states  $\mathcal{S}$ , a set of possible actions  $\mathcal{A}(s)$ , a set of rewards  $\mathcal{R}$ , and a state transition probability  $P(S_{t+1}, R_{t+1}|S_t, A_t)$ . The agent, at each time step  $t$ , observes a state  $S_t$  from the environment, and conducts an action  $A_t$  applied to the environment. The environment responds to the action and transitions into a new state  $S_{t+1}$ . It then reveals the new state and provides a reward  $R_t$  to the agent. This process iterates until a given condition is satisfied (i.e., the RL problem is solved). The extraction process will produce a sequence  $[S_0, A_0, S_1, R_1, A_1, \dots]$ . The goal of the meshing agent is to complete the meshing process for any given geometric object while maintaining high mesh quality.

The RL algorithm deployed in this computational framework is the SAC method. SAC is one of the state-of-the-art RL algorithms for continuous action control problems (Haarnoja et al., 2018a,b). To overcome the sample complexity and hyperparameter-sensitivity, it adds an entropy term in addition to the reward in the objective function, and maximizes the reward return while maximizing the randomness of the policy.

Following (Haarnoja et al., 2018a), the objective function of the policy is correspondingly denoted as

$$J(\pi) = \sum_{t=0}^T \mathbb{E}_{(s_t, a_t) \sim \rho_{\pi_\theta}} [r(s_t, a_t) + \alpha \mathcal{H}(\pi_\theta(\cdot | s_t))], \quad (10)$$

where  $\mathcal{H}(\cdot)$  is the entropy measure (Ziebart, 2010);  $\rho_{\pi_\theta}$  is the state-action marginal distribution of policy  $\pi$  parameterized by  $\theta$ ; and  $\alpha$  indicates the significance of the entropy term, known as the temperature parameter. Entropy maximization allows the learned policy to act as randomly as possible while guaranteeing task completion, which gains a trade-off between exploration and exploitation and thus accelerates learning. This randomness is especially important for a partially observable environment.

SAC combines Q-learning with stochastic policy gradient learning. Q-learning is based on the Bellman equation

$$Q(s_t, a_t) = r(s_t, a_t) + \gamma \mathbb{E}_{s_{t+1} \sim P} (\max_{a'} Q(s_{t+1}, a')). \quad (11)$$

As we already know, in SAC, policy entropy maximization is introduced in the RL objective of maximizing the expected return. Correspondingly, we have the following form of the soft Bellman equation.

$$Q(s_t, a_t) = r(s_t, a_t) + \gamma \mathbb{E}_{s_{t+1} \sim \rho_\pi(s)} [V(s_{t+1})], \quad (12)$$

and

$$V(s_t) = \mathbb{E}_{a_t \sim \pi} [Q(s_t, a_t) - \alpha \log \pi(a_t | s_t)], \quad (13)$$

where  $\rho_\pi(s)$  is the state marginals of the trajectory distribution induced by a policy  $\pi(a_t | s_t)$ .

SAC uses this soft Bellman equation to estimate the target soft Q-values, and soft Q-learning in SAC aims to minimize the difference between the Q-value estimated by a Q-function approximator with parameters  $\theta$  and the target soft Q-value as follows.

$$J_Q(\theta) = \mathbb{E}_{s_t, a_t \sim \mathcal{D}} \left[ \frac{1}{2} (Q_\theta(s_t, a_t) - \hat{Q}(s_t, a_t))^2 \right]. \quad (14)$$

In training, all samples are taken from the replay buffer  $\mathcal{D}$ , which is the collection of previous experience, i.e.  $(s, a, r, s')$  tuples. SAC uses neural networks to approximate the soft Q-function and the policy.

To train Q-function neural network with parameters  $\theta$  by using gradient descent, we can estimate the gradient as follows,

$$\hat{\nabla}_\theta J_Q(\theta) = \nabla_\theta Q_\theta(s_t, a_t)(Q_\theta(s_t, a_t) - (r(s_t, a_t) + \gamma(Q_{\bar{\theta}}(s_{t+1}, a_{t+1}) - \alpha \log(\pi_\phi(a_{t+1} | s_{t+1})))), \quad (15)$$

where  $\pi_\phi$  is the current policy parameterized by a neural network with parameters  $\phi$ . To simplify gradient descent, the target Q-value is calculated with a different neural network with parameters  $\bar{\theta}$ , which is simply the corresponding soft Q-function network with delayed parameter updates. Practically,  $\bar{\theta}_i$  is obtained as an exponential moving average of  $\theta_i$ .

As in the TD3 (Twin Delayed DDPG) model (Fujimoto et al., 2018), SAC has a neural network with parameters  $\phi$  to approximate policy, two neural networks, each with parameters  $\theta_i$  ( $i = 1, 2$ ), working together to approximate the soft Q-function, and two neural networks, each with parameters  $\bar{\theta}_i$  ( $i = 1, 2$ ), to estimate target Q-values. The latter two neural networks are simply the former two networks with the delayed updates of parameters  $\theta_i$ . In order to overcome the overestimation bias (Fujimoto et al., 2018), the minimum of the two values estimated by the twin networks is used as the estimated Q-value, i.e.,  $Q_\theta(s_t, a_t) = \min\{Q_{\theta_1}(s_t, a_t), Q_{\theta_2}(s_t, a_t)\}$ .

In the soft policy improvement stage, following (Haarnoja et al., 2018a), the policy parameter can be updated by minimizing the expected Kullback-Leibler (KL)-divergence. The learning objective can be expressed as follows,

$$J(\phi) = \mathbb{E}_{s_t \sim \mathcal{D}} [\mathbb{E}_{a_t \sim \pi_\phi} [\alpha \log(\pi_\phi(a_t | s_t)) - Q_\theta(s_t, a_t)]]. \quad (16)$$

The policy is reparameterized with a neural network transformation  $a_t = f_\phi(\epsilon_t; s_t)$ , where  $\epsilon_t$  is an input noise vector and can be sampled from a fixed distribution.

Finally, the temperature  $\alpha$  is updated by minimizing the objective

$$J(\alpha) = \mathbb{E}_{a_t \sim \pi} [-\alpha \log \pi_\phi(a_t | s_t) - \alpha \bar{\mathcal{H}}]. \quad (17)$$

The details of the algorithm are shown in Algorithm 1. The parameters used are detailed in the experiment section.

---

**Algorithm 1** SAC for mesh generation.

---

```
1: Initialize parameters for Q networks and policy network,  $\theta_1, \theta_2, \phi$ ; initialize replay buffer  $\mathcal{D}$ ; initialize environment
2: Set target Q networks  $\bar{\theta}_1 \leftarrow \theta_1, \bar{\theta}_2 \leftarrow \theta_2$ 
3: for time step  $t$  in range  $(1, N_T)$ , do
4:   Select action  $a_t \sim \pi_\phi(\cdot | s_t)$ 
5:   Observe reward  $r_{t+1}$  and new state  $s_{t+1}$ 
6:   Store  $(s_t, a_t, r_{t+1}, s_{t+1})$  in replay buffer  $\mathcal{D}$ 
7:   if  $s_{t+1}$  is terminal, reset environment state
8:   if the size of replay buffer  $\mathcal{D} > m$  then
9:     for gradient step  $j$  in range  $(1, N_G)$ , do
10:      Sample a batch from replay buffer  $\mathcal{D}$  with size  $m$  and calculate
11:         $\hat{\nabla}_{\theta_i} J_Q(\theta_i), \hat{\nabla}_\phi J_\pi(\phi)$ 
12:      Update soft Q-function  $\theta_i \leftarrow \theta_i - \lambda_Q \hat{\nabla}_{\theta_i} J_Q(\theta_i)$  for  $i \in \{1, 2\}$ 
13:      Update policy network  $\phi \leftarrow \phi - \lambda_\pi \hat{\nabla}_\phi J_\pi(\phi)$ 
14:      Update target network  $\bar{\theta}_i \leftarrow \tau \theta_i + (1 - \tau) \bar{\theta}_i$  for  $i \in \{1, 2\}$ 
15:      Adjust temperature  $\alpha \leftarrow \alpha - \lambda \nabla_\alpha J(\alpha)$ 
16:    end for
17:  end if
18: end for
```

---

### 3. Experimental results

In this section, we conducted two categories of experiments: 1) to identify the optimal parameter settings for the proposed method, FreeMesh-RL and 2) to demonstrate the framework performance in scalability, generalizability, and mesh quality.

#### 3.1. Implementation settings

This section examines the optimal settings of RL algorithm selection, training domain selection, SAC implementation, state representation, action space, and reward function design. All the experiments are conducted on a computer with an i7-8700 CPU and an Nvidia GTX 1080 Ti GPU with 32 GB of RAM. To evaluate the learning performance, we calculate the average return for ten evaluation episodes at every 10k time steps.

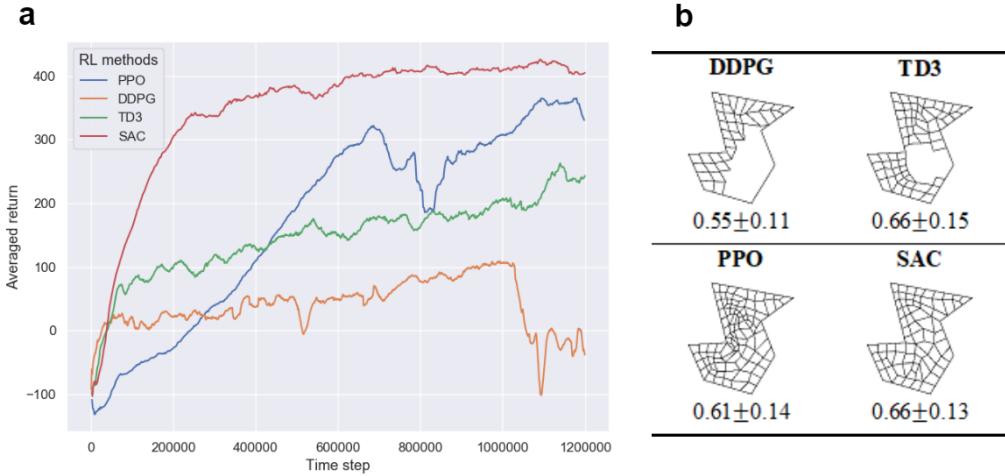


Figure 9: Learning efficiency and performance comparison of various RL algorithms. (a) Four types of RL algorithms, PPO, DDPG, TD3, and SAC, are chosen to compare the learning efficiency over the same training domain. (b) The performance of the policies learned by the four methods is examined on a domain. Their element quality in each mesh is computed. DDPG and TD3 cannot complete the mesh because of the early convergence of their meshing policy.

### 3.1.1. RL method selection

To be applied to real engineering problems, mesh generation requires a robust and stable policy to be learned without complicated hyperparameter tuning. We compared the learning efficiency of four types of RL algorithms (i.e., PPO, DDPG, TD3, and SAC) over the same training domain. With same hyperparameter settings in their original literature (Raffin et al., 2021), the results are compared in Fig. 9 (a). In the early stage of training ( $< 5e4$ ), DDPG, TD3, and SAC achieve faster learning performance than the PPO method. Then the learning speeds of DDPG and TD3 decrease, while SAC maintains high speed and converges to a stable meshing policy. Although PPO gradually achieves suboptimal performance, its policy severely oscillates.

We also tested all the learned policies on a domain, as illustrated in Fig. 9 (b). The policies learned by DDPG and TD3 do not converge and cannot successfully mesh the domain. PPO can easily generate irregular elements, causing poor mesh quality, whereas SAC has successfully meshed the domain. Without further hyperparameter adjustment, the SAC method

achieves the fastest learning performance and the most stable meshing policy. SAC appears to be an optimal method for mesh generation.

### *3.1.2. Training domain selection*

In a real engineering environment, the meshing domains have diverse shapes and topological structures. The basic domain features include sharp angles, bottleneck regions, and unevenly distributed boundary segments for a single connected 2D geometry. Therefore, the training domain should match those criteria to ensure the richness of the samples. The meshing boundary changes constantly during element generation, as shown in Fig. 2. The total number of intermediate boundaries is equivalent to the number of elements generated. This process will also increase the sample diversity.

To find the appropriate training domain, we designed three candidate domains, T1, T2, and T3, based on the above criteria, as shown in Fig. 10 (a-c). The training results using SAC are shown in Fig. 10(d). The agents in both domains T1 and T2 achieve fast learning speed and converge to a stable meshing policy, but T1 has less learning fluctuation. The agent spends more learning time in domain T3 because of the difficulties of having more sharp angles and bottlenecks. We choose domain T1 as the training domain because of its optimal steadiness and learning efficiency. As shown in the rest of experiments, the model trained with T1 has strong generalizability.

### *3.1.3. SAC implementation*

To identify the optimal NN hidden layer setting (i.e., the number of hidden layers and neurons in each layer) in SAC, we compared four different configurations, S1-S4, from the perspective of learning efficiency. The results are compared in Fig. 11 (a). Configuration S1 has the poorest performance in learning the policy. The network structure S2 with three hidden layers, [128, 128, 128], achieves the best learning performance while having fewer parameters than S3 and S4. It is used in the article to approximate the soft Q-function, policy, and target networks.

A random seed is a common parameter in RL, which often affects learning performance. We selected three random seeds and tested their impact on policy learning, as shown in Fig. 11 (b). It turns out that the SAC learning process is not sensitive to those seeds, which is one of its advantages compared with other RL methods. The other hyperparameters used for SAC are listed in Table 1.

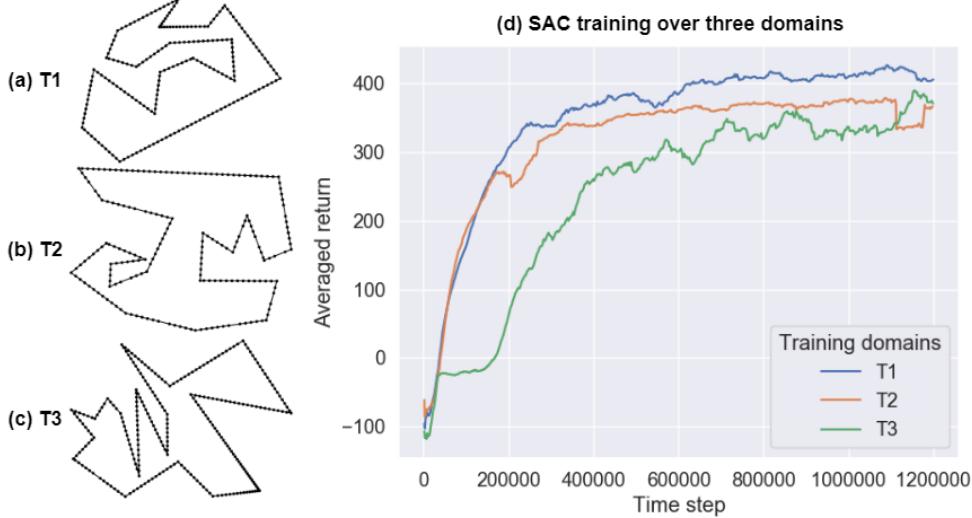


Figure 10: Learning difficulty comparison over different training domains. (a)-(c) Three types of training domains, T1, T2, and T3, are designed based on the identified criteria. (d) The training results over three domains using SAC. Domain T1 achieves a fast and stable learning efficiency which has an easier meshing policy to learn than the other two domains, whereas domain T3 is the hardest one to converge to a stable policy because multiple sharp angles exist.

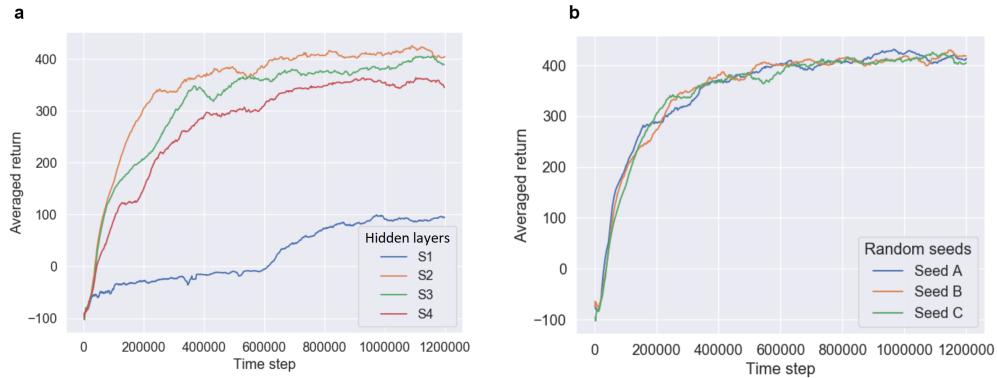


Figure 11: Comparisons of hidden layer configurations and random seeds for the SAC algorithm. (a) Comparison of four types of hidden layer configurations: S1-S4 represent four different neural network structures in the hidden layers, including [32, 32], [128, 128, 128], [64, 64, 64, 64], and [32, 128, 128, 128, 64, 32], respectively. Configuration S2 achieves a fast and stable learning result. (b) Random seed comparison: three different random seeds, A (356), B (567), and C (999), are chosen to examine the learning difference. Certainly, SAC is not sensitive to those random seeds and could achieve stable learning results.

Table 1: Training hyperparameters for the SAC algorithm.

Parameter	Description	Value
$N_D$	Experience pool size	1e6
$m$	Minibatch size	256
$\gamma$	Discount factor	0.99
$\lambda_Q, \lambda_\pi$	Learning rate	3e-4
$N_T$	Total time steps	1.2e6
$N_G$	Gradient steps	1
$\tau$	Soft update factor	5e-3

### 3.1.4. Agent’s view of environmental state

The state is the agent’s observation of the environment and provides a decision basis for the agent. As partial observation is adopted in this method, it is necessary to decide the observation range for the agent to learn the meshing policy effectively. The range of the observation is controlled by three parameters,  $\beta$  (in Equation 3),  $n$  and  $g$  (in Equation 4). The parameter  $\beta$  controls how far in the fan-shaped area the agent can observe from the selected reference vertex while the other two parameters determine how many vertices the agent perceives around the reference vertex. The learning performance is compared with three types of settings,  $O1$  ( $\beta = 4, n = 2, g = 3$ ),  $O2$  ( $\beta = 6, n = 2, g = 3$ ), and  $O3$  ( $\beta = 6, n = 3, g = 4$ ). The results are compared in Fig. 12. By comparing  $O1$  and  $O2$ , it can be found that further observation contributes to more return. This is because the agent could adjust the position of the candidate vertex in advance to avoid a conflict with the remaining boundary. On the other hand, the more vertices in the fan-shaped areas are considered, the more information will be embedded in the state. Therefore, increasing  $g$  could slow the learning speed and delay policy convergence. This phenomenon can be observed when comparing  $O2$  and  $O3$ . The more information the agent observes, the more time is needed to build the correlation.

We also evaluated the influence of the observation ranges on the meshing performance from the aspects of element quality ( $\eta^e$  in Equation 7), the number of elements generated, and time cost. Table 2 presents those performance comparisons. The policies learned by the observations of  $O1$  and  $O2$  have a small difference in element quality, whereas the one learned by  $O3$  is

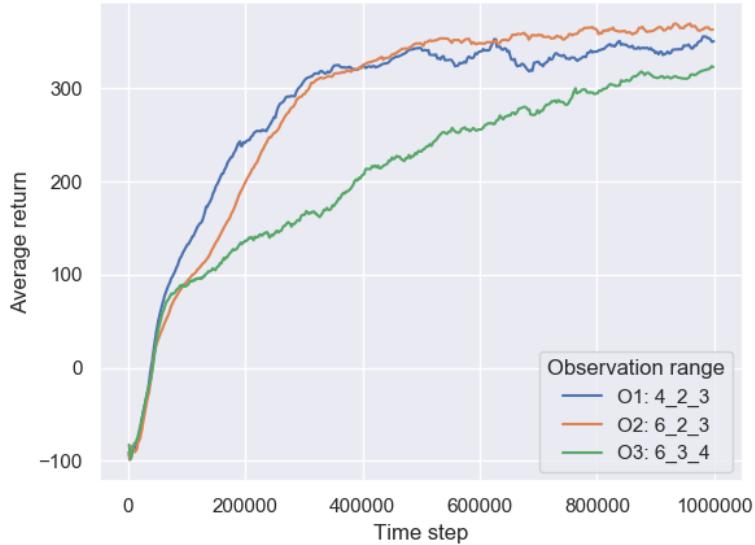
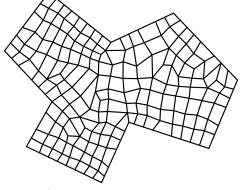
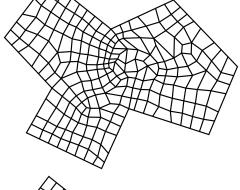
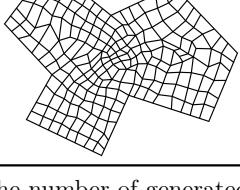


Figure 12: Learning efficiency comparison of different agent's observation ranges. The observation range is formed by  $\beta_n g$ , which represents the radius of the fan shape in the state, the number of neighboring vertices on the left and right sides of the reference vertex, and the number of vertices in the fan-shaped area. This range determines how far and how much information the agent will observe in the meshing environment.

relatively poorer because of the slow learning speed. With the increase in observed information, the number of generated elements also increases. This is reasonable because the agent could make prudent decisions to avoid collision with the sensed boundary in a fan shape when it observes farther and broader. As illustrated by the sample meshes, the interior area has smaller and more regular elements by  $O2$  and  $O3$  as the boundary is updated inwardly. The meshing speed hinges on the number of generated elements, which is indirectly influenced by the observation ranges because the agent tends to make smaller elements if it observes a potential collision with the sensed boundary. Ideally, the more information the agent observes, the more accurate the decision it can make and the more learning time it requires to converge. To achieve a trade-off between the computational cost and mesh quality, the observation range  $O2$  is used to represent the state in this article.

Table 2: Meshing results comparison of different agents' observation ranges. The observation ranges are  $O1$ ,  $O2$ , and  $O3$ . The performance is evaluated from three indicators, including element quality, the number of generated elements, and time cost. A sample mesh (without smoothing) for each range is also represented. All the experiments are repeated ten times over the same domain for each observation.

Observation	Sample mesh	Element quality	#elements	Time cost (s)
$O1: 4\_2\_3$		$0.722 \pm 0.14$	$198.8 \pm 22.11$	$0.6 \pm 0.09$
$O2: 6\_2\_3$		$0.712 \pm 0.13$	$212.8 \pm 24$	$0.64 \pm 0.08$
$O3: 6\_3\_4$		$0.689 \pm 0.137$	$217.1 \pm 11.16$	$0.7 \pm 0.13$

#elements - the number of generated elements.

### 3.1.5. Action space

The agent's action space defines the shape and size scale of each element to be generated, which influences the mesh quality and policy's learning efficiency. This section will examine the appropriate size of the search space, i.e., the coordinate space to select the candidate vertex in forming a quadrilateral element. The coordinate space, specified by the radius in Equation 2, is defined by weight  $\alpha$  and a base length  $L$ . The number of vertices in calculating the base length is equivalent to the number of neighboring vertices of the reference vertex in the state (see the parameter  $n$  in Equation 4). The ideal observation range (i.e., state) is determined as  $O2$  in the previous section. The parameter  $n$  is hence set to 2.

The size of the search space is defined by the radius of the viewing fan shape, determined by the weight  $\alpha$ . To find the appropriate radius, three types of settings,  $R1$  ( $\alpha = 1$ ),  $R2$  ( $\alpha = 2$ ), and  $R3$  ( $\alpha = 3$ ), are examined during meshing policy learning over the same domain (see Fig. 10 (a)). Their learning results are shown in Fig. 13. The radius  $R1$  achieves the fastest learning speed and converges earlier than the other two settings. Because its search space is the smallest, action exploration is less needed. Although the radius  $R2$  is slightly slower in convergence than  $R1$ , it facilitates the highest reward return indicating optimal mesh quality, due to the larger search space. Unfortunately, the agent fails to learn an efficient policy by radius  $R3$ . The large search space contains sparse efficient actions, and makes agent's exploration harder. Therefore, the radius  $R2$  is used in the present article.

### 3.1.6. Reward function

There are three terms in the reward function: element quality  $\eta_t^e$ , the quality of the remaining boundary  $\eta_t^b$ , and density  $\mu_t$ . The first two terms guarantee the element quality and the ease of continuous meshing. The last term controls the meshing density by the parameter  $v$ , which adjusts the minimum element size tolerated, and ensures the mesh termination within finite steps. The first two terms are necessary and not changeable once the quality requirement is determined, whereas the third term can be adjusted according to different requirements for mesh density.

To achieve the optimal mesh density, three different parameters are compared, including sparse ( $v = 1.5$ ), medium ( $v = 1$ ), and dense ( $v = 0.5$ ) settings. The results are compared in Fig. 14 (a)-(c). We also examine the number of elements generated by each density, as shown in Fig. 14 (d). The

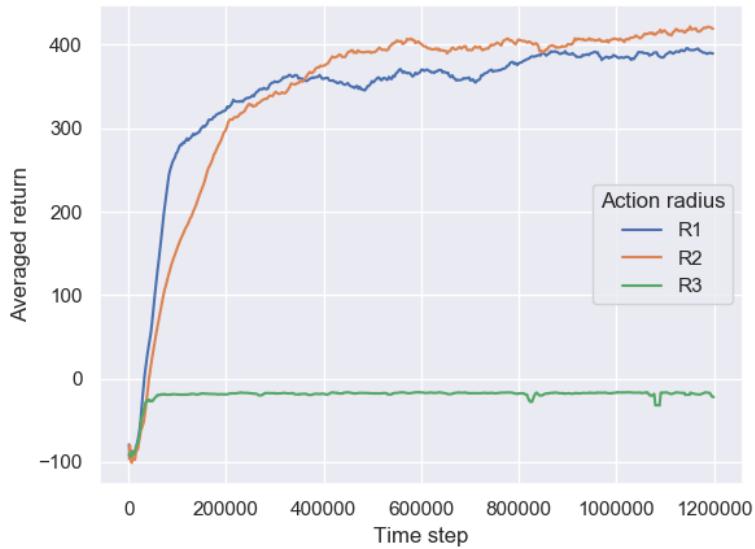


Figure 13: Learning efficiency comparisons of three types of radius settings in the action. The appropriate radius is necessary for optimal mesh quality and learning efficiency. These settings are  $R1$  ( $\alpha = 1$ ),  $R2$  ( $\alpha = 2$ ), and  $R3$  ( $\alpha = 3$ ). A larger  $\alpha$  indicates a larger coordinate space for searching a candidate vertex used to form a quadrilateral element.

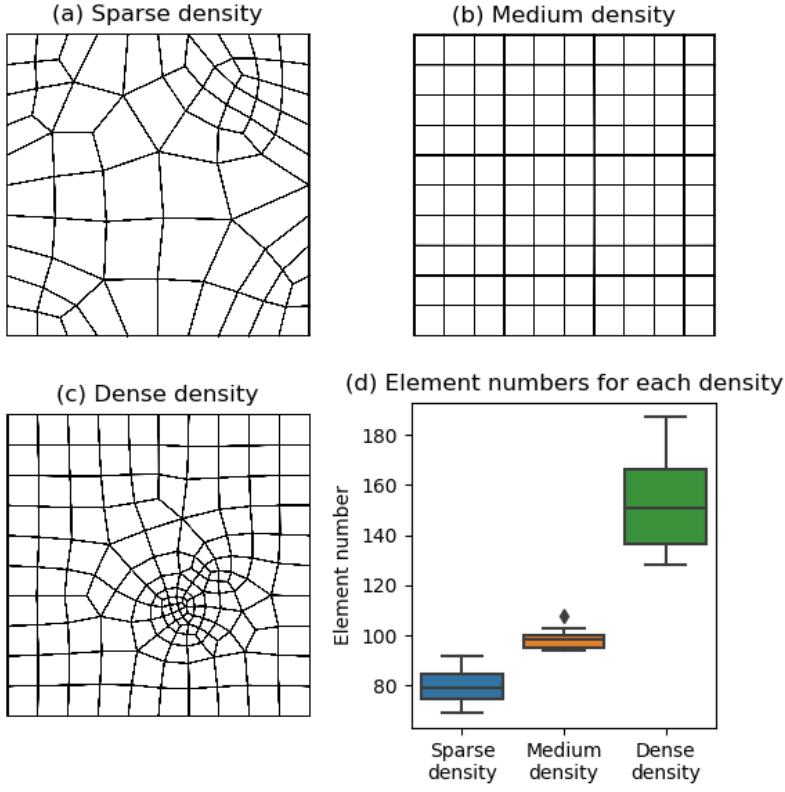


Figure 14: Different mesh densities controlled by the reward function. Three different densities from sparse to dense, (a)-(c), are controlled by the parameters  $v = 1.5$ ,  $v = 1$ , and  $v = 0.5$ , respectively. Subfigure (d) shows the results of the number of generated elements by three types of densities averaged over ten episodes.

results are averaged over 10 episodes. The difference in the number of elements between sparse and medium is approximately 20, while the difference between medium and dense density is approximately 50 elements in the testing domain. The medium density is ideal and used across all the remaining experiments.

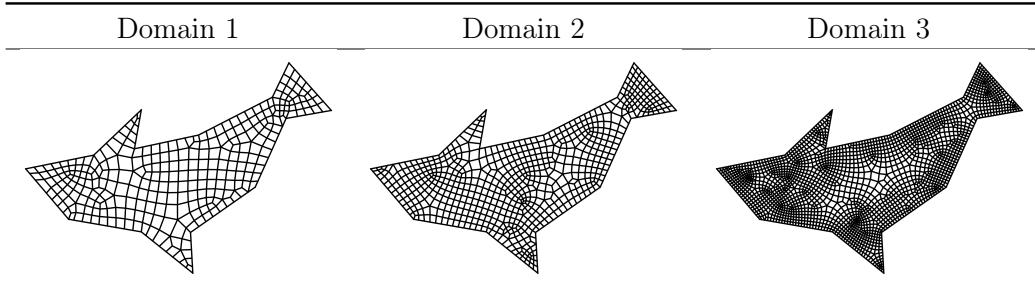
### 3.2. Effectiveness evaluation

The effectiveness of FreeMesh-RL is evaluated from the perspectives of scalability, generalizability, and mesh quality. The model used in the following experiments is trained on the same domain (i.e., domain T1 in Fig. 10 (a)) without any adjustment.

### 3.2.1. Scalability verification

To validate the scalability of the learned meshing policy, we constructed three geometry domains with the same shape but different vertex densities (i.e., 6.8: 9.9: 20.1, as shown in Table 4) on the boundaries. Three domains are meshed by the same RL model, and the results are shown in Table 3. It can be seen that all the domains have successfully meshed; the elements on the boundaries are denser than in the interior area, which is beneficial for reducing computational burden; and the transitions between dense and coarse meshes are smooth. The meshing speed over three domains is approximately 237 elements per second on average. Consequently, the results show that the learned mesh policy achieves good scalability to different scales of the meshing problem and is not constrained to the density requirements of the training domain. This scalability is a special manifestation of the broader generalizability to be discussed next.

Table 3: Meshing the same domain with different boundary segment densities by FreeMesh-RL. The boundary shapes of domains 1-3 are the same, but the number of vertices on the boundary is from low to high. The scalability is examined by meshing all the domains with the same trained model.



### 3.2.2. Generalizability verification

To verify the generalizability of the obtained meshing model, we meshed four different domains, domains 4-7, using the same model, trained with domain T1 as shown in Fig. 10. The shapes of the domains are from simple to complex. The meshing process and results are shown in Fig. 15. The general meshing process starts from the boundary and advances inwardly to the central area of the domain. Regardless of the complexity of the initial domains, the boundary will evolve into various intermediate shapes with the

Table 4: Geometrical details of three domain boundaries and their generated meshes. The perimeter, the number of vertices and vertices per unit length on the initial boundaries of three domains are compared as well as their number of generated elements and meshing time.

	Domain 1	Domain 2	Domain 3
#vertices	102	150	304
Perimeter	15.1	15.1	15.1
#vertices per unit length	6.8	9.9	20.1
#elements	289	665	2157
Execution time (s)	0.9	2.6	15.3

#vertices - the number of vertices on the boundary.

#elements - the number of generated elements.

meshing process guided by the learned meshing policy. The dynamic changes of the domain boundary also reflects the generalizability, i.e., how good the policy to handle different shapes.

The number of actions executed in meshing each of the four domains is shown in Fig. 16. The meshing policy mainly consists of two types of actions (see type 0 and type 1 in Fig. 5), and the execution number of type 1 is approximately 80% of the total number. Each action, a single execution of the policy, will produce an element. The number of intermediate boundaries equals the number of elements generated. For example, in domain 7, there are 1489 (the sum of execution times of type 0 and type 1) intermediate boundaries in total. The meshing policy succeeds in meshing 1489 different domain boundaries. It can be noted that all the intermediate boundaries of the four domains gradually become an oval shape in Fig. 15. This demonstrates that the meshing policy can solve complex boundary situations with sharp angles and create smooth and easy-to-handle situations for future element generation.

Although the geometry shape can be diverse and infinite in real engineering applications, our state representation with a partial boundary and the use of a reference vertex as the origin make our model general and able to handle various boundary shapes. As demonstrated above, the obtained policy is able to mesh hard situations while maintaining the high quality of the remaining boundaries. Therefore, the most challenging part of the whole meshing process is to mesh the initial boundary situations of a given domain. However, it can be easily solved by providing training domains with

sufficient difficulties (in Fig. 10 (a-c)). To conclude, the proposed method can be applied to arbitrary domains and is general to various geometries.

### 3.2.3. Comparison to conventional methods

To evaluate the meshing performance, we compare the quality of meshes by FreeMesh-RL with two other representative meshing approaches over three predefined 2D domains (i.e., domains D7-9). These domains possess different features, including sharp angles, bottleneck regions, unevenly distributed edges, and holes, to increase the testing diversity. The two conventional meshing approaches are Blossom-Quad (Remacle et al., 2012) and Pave (Blacker and Stephenson, 1991; White and Kinney, 1997). Blossom-Quad is an indirect method that generates quadrilateral elements by finding the perfect matching of a pair of triangles generated in advance. The method is implemented by an open source generator Gmsh (Geuzaine and Remacle, 2009). Pave is another state-of-the-art meshing method for directly generating quadrilateral elements, implemented by the CUBIT software (Blacker et al., 2016).

The meshing results are shown in Table 5. Although all the methods can complete the meshes for the three domains, there are some subtle differences. Only FreeMesh-RL generates fully quadrilateral meshes. The other two methods have difficulties in discretizing the domains into full quadrilaterals, containing triangles in domains (marked in yellow color in each domain). Specifically, Blossom-Quad has a problem in handling domains with sharp angles along the boundary, while Pave has issues in the interior of the domain. Extra operations (e.g., clean-ups) are thus usually required to eliminate those triangles or bad elements. Another advantage of FreeMesh-RL is that the generated mesh can smoothly transition from very small to large elements over a short distance, as shown in Table 5 (domain 9), which is beneficial in reducing the computational burden during simulations (Shewchuk, 2012).

To quantitatively analyze the meshing results of the three methods, we selected eight common quality metrics, including 1) singularity, the number of irregular nodes whose number of incident edges in the interior of a mesh is not equal to four; 2) element quality,  $\eta^e$  in Equation 7; 3)  $|MinAngle - 90^\circ|$ ; 4)  $|MaxAngle - 90^\circ|$ ; 5) scaled Jacobian, the minimum Jacobian (Knupp, 2000) at each corner of an element divided by the lengths of the two edge vectors; 6) stretch, the degree of deformation; 7) taper, the maximum absolute difference between the value one and the ratio of two triangles' areas are separated by a diagonal within a quadrilateral element; and 8) the num-

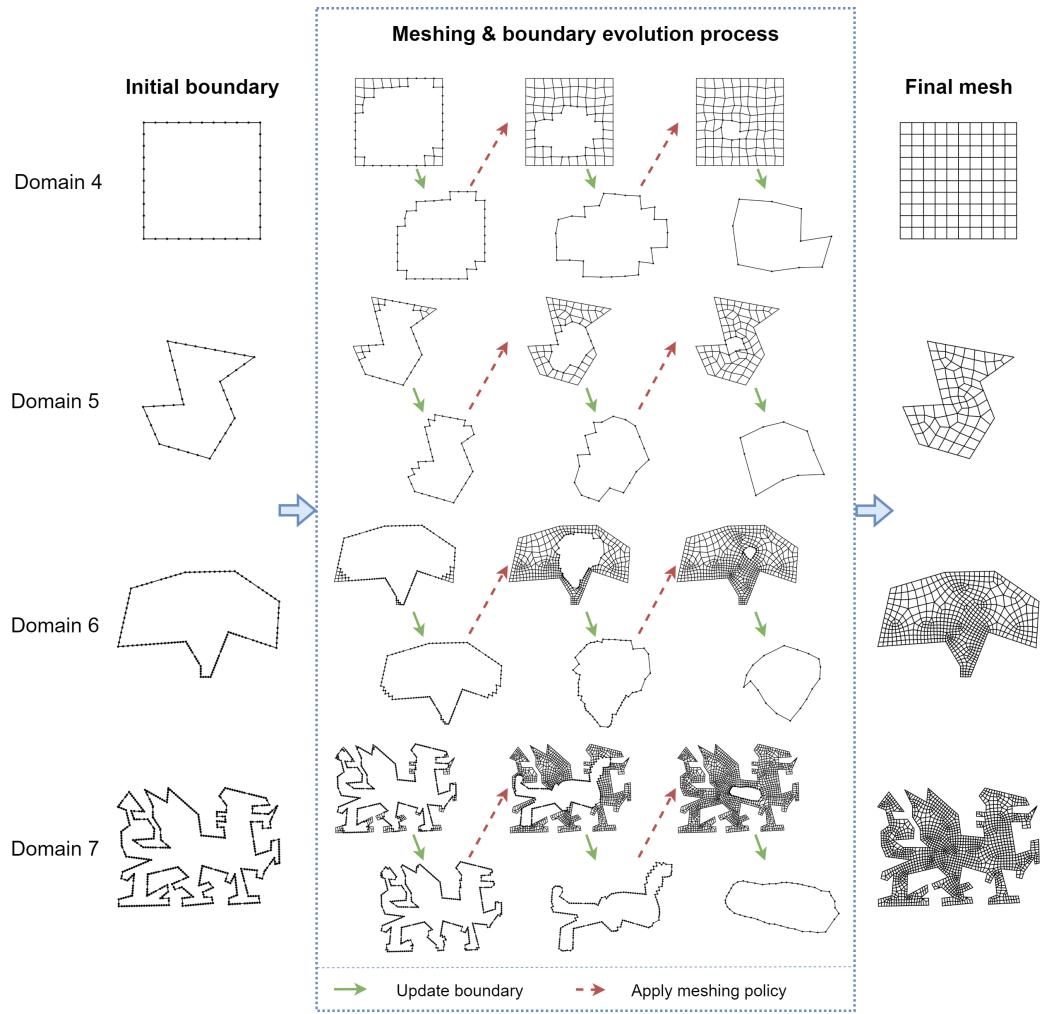


Figure 15: Model generalizability verification. Four different domains with shapes from simple to complex are selected; their meshing procedures are presented to exhibit the boundary evolution process and general meshing pattern by the learned policy. No matter how simple an initial shape was, the intermediate shapes can be complex; no matter how complex the shape of the initial boundary was, it gradually evolves into a smooth oval shape in the center of the domain.

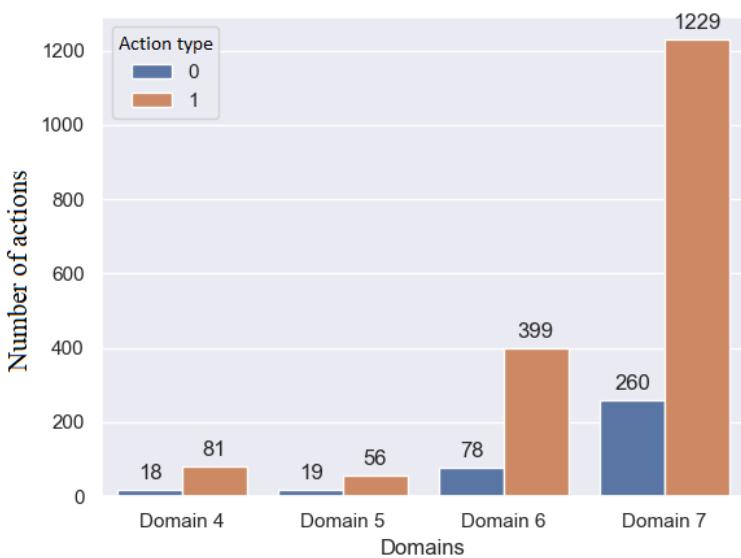
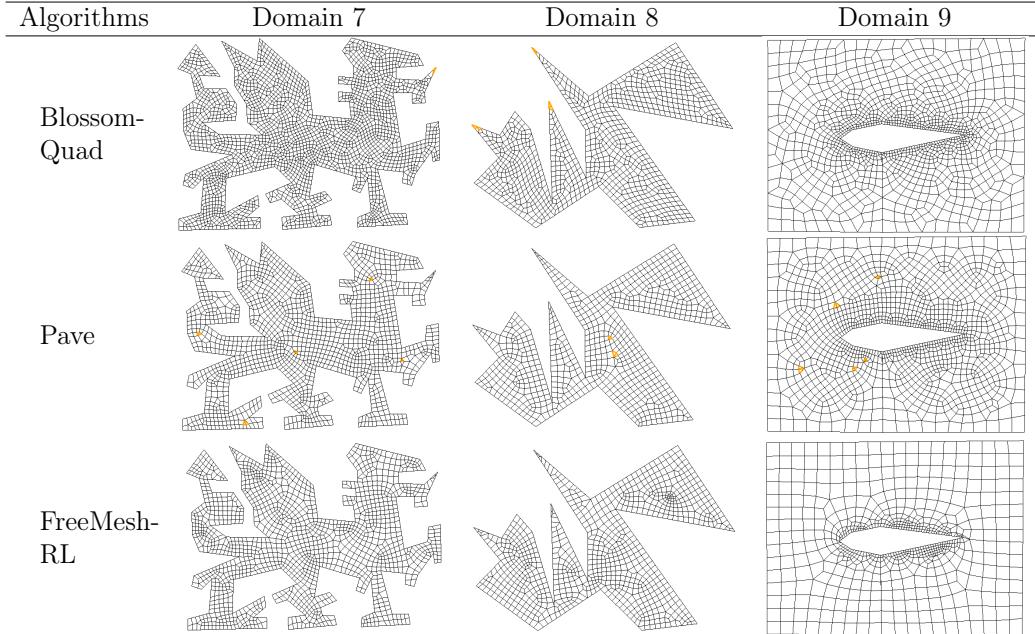


Figure 16: The execution times of the meshing policy over the four domains. The meshing policy mainly consists of two types of actions: types 0 and 1. The number over each bar indicates the execution number of each action type.

Table 5: Meshing results comparison. Three domains are designed to examine the meshing performance based on the four identified criteria: a. sharp angle, b. narrow region, c. unevenly distributed boundary segments, and d. having a hole inside. Both domains 7 and 8 possess features a and b, whereas domain 9 has features c and d. Two representative methods, Blossom-Quad and Pave, are chosen to compare the meshing performance with the proposed method, FreeMesh-RL.



The elements in yellow represent existing triangles in the meshes.

ber of triangles (#triangles) (Pan et al., 2021; Knupp et al., 2006). Smaller singularity and taper mean better regularity, which can provide more accurate results for numerical simulations. A larger scaled Jacobian indicates higher convexity of the mesh, containing less inverted and flat quadrilateral elements. The existence of triangles indicates that the domain is not fully meshed by quadrilaterals and usually require extra treatments to eliminate.

All the measurement results are averaged over three domains and are shown in Table 6. The proposed method, FreeMesh-RL, outperforms other methods in the indices of singularity, taper, scaled Jacobian, and number of triangles, while being comparable with other best performing indices by Pave. Pave achieves the best performance in the remaining indices (i.e., element quality, min and max angles, and stretch), whereas Blossom-Quad has the lowest quality and is only slightly better than Pave at having fewer triangles that are not expected. It is common for indirect methods (i.e., Blossom-Quad) to have suboptimal performance because of the dependence on prior

triangulation. The computational complexities for the three methods are all  $O(n^2)$  (Pan et al., 2021).

Table 6: Quantitative measurement of the meshing performance of the three methods. All the quality metrics are averaged over the three domains (i.e., domains 7-9).

Metrics	Blossom-Quad	Pave	FreeMesh-RL
Singularity (L)	$388 \pm 209.50$	$146.70 \pm 51.50$	<b><math>132 \pm 50</math></b>
Element quality (H)	$0.72 \pm 0.12$	<b><math>0.79 \pm 0.12</math></b>	$0.79 \pm 0.13$
$ MinAngle - 90^\circ $ (L)	$6.55 \pm 6.91$	<b><math>3.69 \pm 4.60</math></b>	$4.02 \pm 5.10$
$ MaxAngle - 90^\circ $ (L)	$22.16 \pm 11.14$	<b><math>15.69 \pm 14.71</math></b>	$15.73 \pm 12.48$
Scaled jacobian (H)	$0.91 \pm 0.08$	$0.94 \pm 0.13$	<b><math>0.94 \pm 0.10</math></b>
Stretch (H)	$0.79 \pm 0.08$	<b><math>0.84 \pm 0.10</math></b>	$0.83 \pm 0.11$
Taper (L)	$0.15 \pm 0.11$	$0.12 \pm 0.14$	<b><math>0.11 \pm 0.11</math></b>
#Triangle (L)	$2.70 \pm 2.50$	$8 \pm 2.80$	<b><math>0 \pm 0</math></b>

L and H indicate whether the lower value or higher value is preferred, respectively. The value in bold means the best among other approaches in that specific metric.

The boxplots of the quality metrics are compared in Fig. 17, which shows the steadiness and extremely poor situations in each quality measure. Pave has the most outliers in all the quality metrics except singularity and number of triangles, indicating that more elements exist with extremely low quality than others, although it maintains the optimal averaged performance in some metrics. Although Blossom-Quad has the lowest average performance, its generated meshes have the steadiest quality. FreeMesh-RL is in the middle regarding quality steadiness, generating less extreme low quality elements than Pave and maintaining high averaged performance. The more unstable and poor situations require more postprocessing operations.

We also analyzed the knowledge dependence during the algorithm development, as shown in Table 7. The development of the meshing method can be briefly divided into three stages: preprocessing, element generation, and postprocessing. Since Blossom-Quad is an indirect quadrilateral mesh generation method, it requires the preprocessing of triangulation, which is unnecessary for the other two methods. To generate elements, the development of Blossom-Quad requires more complex geometric knowledge than Pave and FreeMesh-RL (see details in (Pan et al., 2021)). Pave, however, relies on considerable heuristic knowledge, which is time-consuming and inefficient for designers. FreeMesh-RL is the simplest method and only demands

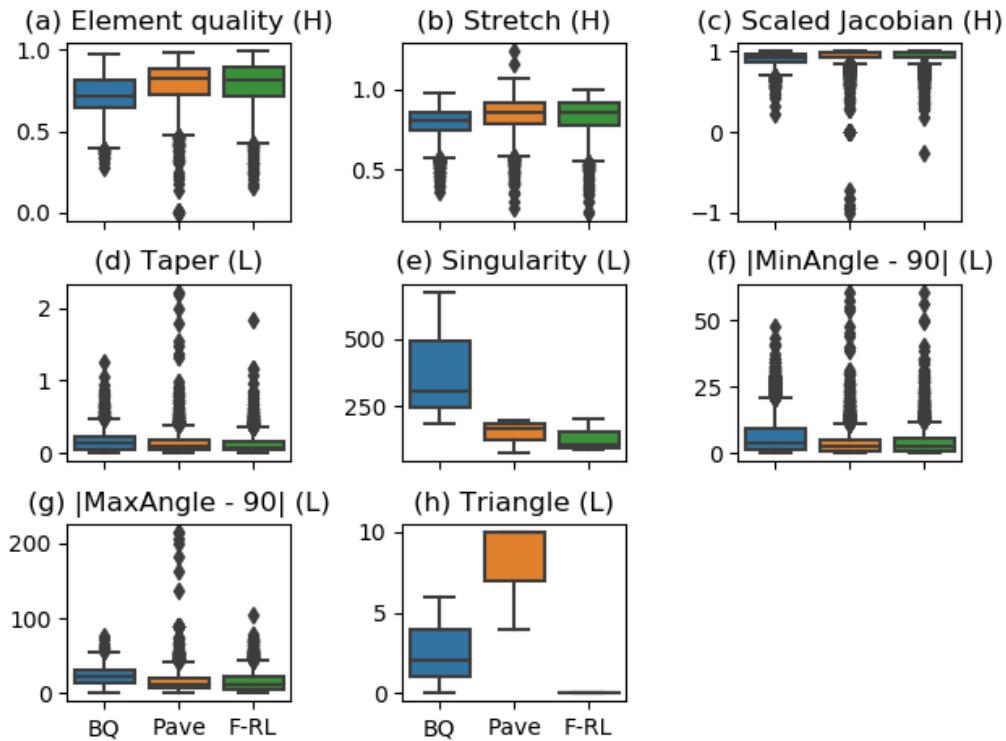


Figure 17: Meshing performance comparison results over eight types of quality indices. BQ represents the Blossom-Quad method; F-RL represents the FreeMesh-RL method. L, H indicate if the lower value or higher value is preferred, respectively.

basic geometry concepts. Extra treatments such as clean-ups are necessary for Blossom-Quad and Pave, whereas FreeMesh-RL is a clean-up free method.

Table 7: Comparison of knowledge dependence of the three methods during algorithm development. The algorithm development procedure is divided into three stages: preprocessing, element generation, and postprocessing.

Steps	Blossom-Quad	Pave	FreeMesh-RL
Preprocessing	Delaunay triangulation	—	—
Element generation	Complex geometry knowledge	Heuristic knowledge	Basic geometry concepts
Postprocessing	Clean-ups	Clean-ups	—

## 4. Discussion

Herein, we present a fully automatic quadrilateral mesh generation algorithm, FreeMesh-RL, by using SAC reinforcement learning. With minimal knowledge inputs, this algorithm automatically learns mesh generation policy and performs with comparable quality to commercial systems that conduct quadrilateral mesh generation. This section will discuss the factors behind the performance of the proposed algorithm.

*What are the big challenges in mesh generation?*

The big challenges in mesh generation come from two aspects. (1) In real engineering problems, the geometric domains are very complex, having various shapes, diverse scales, and different mesh requirements. (2) In mesh generation tools, the rules and knowledge manually acquired are insufficient to tackle all of the situations in a mesh generation problem, especially for large and complex geometries. For example, postprocessing is mandatory to guarantee good quality mesh in Pave and Blossom-Quad, the two representative state-of-the-art systems that we discussed in the last section. The fundamental problem is how to make a mesh generation algorithm general enough to handle various geometric domains with different engineering requirements.

*How does FreeMesh-RL tackle the challenges?*

To tackle the challenges in mesh generation, particularly to achieve generalizability for arbitrarily shaped geometries, Zeng and Cheng (1993) first proposed an element-wise approach by using a recursive algorithm to generate one element at a time with three primitive rules shown in Figure 5. This element-wise approach exhibits a similar problem structure in reinforcement

learning. As such, we formulated the mesh generation problem as an RL problem (Pan et al., 2021), as illustrated in Figure 3. Generally, RL is a machine learning paradigm whose success essentially depends on two aspects: (1) the domain-dependent problem formulation, that is, how to represent the environment, states, actions, and rewards; (2) a suitable RL method. This article presents our work on (1) how to optimally design the representation of states, actions, and rewards; (2) how to apply SAC RL to solve the mesh generation problem effectively.

The problem formulation needs insightful abstraction. The three primitive element extraction rules proposed in Zeng and Cheng (1993) shown in Figure 5 lays a foundation for the problem formulation. The use of reference vertex makes it possible to have a general representation of the problem, avoiding to deal different shapes and dynamically changing boundaries. We use a partial observation of the environment as the state, thus focusing on the most relevant information of the environment for actions. The reward function design plays a critical role in mesh quality and optimal trade-off of the current reward (the quality of the element to generate) and long-term return (the overall mesh quality). A general observation is that it is critical to make the information represented by states well cover the factors of rewards.

With respect to RL methods, we select SAC mainly for two reasons. First, its off-policy mechanism allows us to reuse previous experience, thus reducing sampling complexity and making learning more efficient. Secondly and importantly, its distinguishing feature of maximizing policy randomness while maintaining policy performance offers a natural mechanism for an effective trade-off between exploration and exploitation in policy search. This feature of the stochastic policy is particularly beneficial for learning in partially observable environments.

Our problem formulation, together with the SAC RL algorithm, has demonstrated strong generalizability. As presented in last section, the model trained with a single domain (see Fig. 10 (a)) can mesh various other unseen domains, as shown in Table 3, Table 5, and Fig. 15.

The major advantages of the proposed FreeMesh-RL algorithm lie in two aspects: (1) it can automatically acquire and refine knowledge for mesh generation, which is generally labor-intensive and time-consuming, and (2) the automatic knowledge acquisition process does not need predefined samples and labeled data. These advantages effectively tackle some practical challenges in commercial mesh generation software systems.

*What is the distinguishing feature of this work?*

The presented research demonstrates how we achieve the integrated intelligence exhibited by the strong generalizability in mesh generation by combining SAC reinforcement learning with the rule-based knowledge representation for states, actions, and rewards.

In Zeng and Cheng (1993), the authors developed a rule-based knowledge system for mesh generation. They created three primitive element extraction rules, which provide a framework for mesh generation problem representation. However, to deal with various boundaries, many rules are needed, but knowledge acquisition is a bottleneck. To address this issue, Yao et al. (2005) attempted to acquire knowledge for element extraction with MLP neural networks. They introduced the use of reference vertex, which makes the mesh generation problem able to be represented in the same relative space, thus being a MIMO mapping and suitable for using MLP neural networks. Since MLP is a supervised machine learning model, the availability of a large number of samples became a new bottleneck. To solve this problem, we formulated mesh generation as an RL problem in (Pan et al., 2021), took RL as a mechanism to generate meshing samples, selected good quality samples with rules, then fed the selected samples to MLP for training.

In the present article, we further integrate the rule-based knowledge for representing states, actions, and rewards into the framework of SAC reinforcement learning and realize a fully automatic mesh generation system. In this system, rules govern the location of reference vertex, the construction of states, and the quality evaluation of the generated element and the rest of the shape. SAC does the job of exploring possible actions to generate a new element in the current shape, learning to evaluate the quality of the actions, and learning the optimal policy to generate a new element based on the current shape of the geometric domain to mesh.

*What's next?*

As observed in the last section, the performance of the SAC RL algorithm varies with not only its hyperparameters but also the problem formulation and representation and the associated parameter settings. On the other hand, 2D mesh generation is a visible and easily understood problem. We believe this study of automatic 2D mesh generation with RL can be developed into an RL testbed for RL research.

FreeMesh-RL is currently limited to 2D domains. We will apply the proposed framework to 3D mesh generation by reformulating the state rep-

resentation and a few primitive actions for hexahedron elements as well as the reward function.

In the fourth industrial revolution, engineering is transforming into digital engineering (Huang et al., 2021; Huang, 2022), where digital data and models will be shared across the engineering lifecycle. This will make it possible to collect a large volume of meshing examples. Currently, our model is trained with just a single simple geometric domain and shows the capability to mesh in comparable quality to representative commercial software. In real-world engineering for 3D meshing, there will be various challenging geometric domains and meshing requirements for different purposes. We intend to leverage transfer learning (Bozinovski and Fulgosi, 1976; Zhuang et al., 2021) and the available big data of real-world mesh samples to develop deep neural networks, which will be pretrained with excellent samples and can be repurposed by further training with a few samples for a specific task with specific meshing requirements.

## 5. Conclusion

This article presented our research on applying soft actor-critic reinforcement learning for automatic mesh generation. We designed and implemented FreeMesh-RL, a fully automatic quadrilateral mesh generation algorithm with SAC reinforcement learning. With minimal knowledge inputs, this algorithm automatically learns mesh generation policy and performs with comparable quality to commercial systems that conduct quadrilateral mesh generation. Further research can go in several directions. We will extend our current 2D automatic mesh generation algorithm with RL for 3D mesh generation. Also, based on this study of automatic 2D mesh generation with SAC RL, we will develop it into an RL testbed for RL research.

## Acknowledgment

The support of the NSERC Discovery Grant (RGPIN-2019-07048) is gratefully acknowledged.

## References

- Blacker, T.D., Owen, S.J., Staten, M.L., Quadros, W.R., Hanks, B., Clark, B.W., Meyers, R.J., Ernst, C., Merkley, K., Morris, R., et al., 2016. CUBIT

- geometry and mesh generation toolkit 15.2 user documentation. Technical Report SAND2016-1649 R. Sandia National Laboratory. Albuquerque, New Mexico.
- Blacker, T.D., Stephenson, M.B., 1991. Paving: a new approach to automated quadrilateral mesh generation. International Journal for Numerical Methods in Engineering 32, 811–847.
- Bozinovski, S., Fulgosi, A., 1976. The influence of pattern similarity and transfer learning upon training of a base perceptron b2, in: Proceedings of Symposium Informatica, pp. 121–126.
- Chen, R.T., Rubanova, Y., Bettencourt, J., Duvenaud, D., 2018. Neural ordinary differential equations. arXiv:1806.07366 .
- Chew, L.P., 1989. Constrained delaunay triangulations. Algorithmica 4, 97–108.
- Defferrard, M., Bresson, X., Vanderghenst, P., 2016. Convolutional neural networks on graphs with fast localized spectral filtering. Advances in Neural Information Processing Systems 29, 3844–3852.
- Docampo-Sanchez, J., Haimes, R., 2019. Towards fully regular quad mesh generation, in: AIAA Scitech 2019 Forum, p. 1988.
- Fujimoto, S., Hoof, H., Meger, D., 2018. Addressing function approximation error in actor-critic methods, in: International Conference on Machine Learning, PMLR. pp. 1587–1596.
- Geuzaine, C., Remacle, J.F., 2009. Gmsh: a 3-D finite element mesh generator with built-in pre-and post-processing facilities. International Journal for Numerical Methods in Engineering 79, 1309–1331.
- Gordon, W.J., Hall, C.A., 1973. Construction of curvilinear co-ordinate systems and applications to mesh generation. International Journal for Numerical Methods in Engineering 7, 461–477.
- Gottesman, O., Johansson, F., Komorowski, M., Faisal, A., Sontag, D., Doshi-Velez, F., Celi, L.A., 2019. Guidelines for reinforcement learning in healthcare. Nature Medicine 25, 16–18.

- Gupta, K., 2020. Neural mesh flow: 3D manifold mesh generation via diffeomorphic flows. Master's thesis. University of California, San Diego.
- Haarnoja, T., Zhou, A., Abbeel, P., Levine, S., 2018a. Soft actor-critic: off-policy maximum entropy deep reinforcement learning with a stochastic actor, in: International Conference on Machine Learning, PMLR. pp. 1861–1870.
- Haarnoja, T., Zhou, A., Hartikainen, K., Tucker, G., Ha, S., Tan, J., Kumar, V., Zhu, H., Gupta, A., Abbeel, P., et al., 2018b. Soft actor-critic algorithms and applications. arXiv:1812.05905 .
- Huang, J., 2022. Digital engineering transformation with trustworthy AI for industry 4.0. *Journal of Integrated Design and Process Science* 26, 1–20, (in print).
- Huang, J., Beling, P., Freeman, L., Zeng, Y., 2021. Trustworthy AI for digital engineering transformation. *Journal of Integrated Design and Process Science* 25, 1–7.
- Knupp, P.M., 2000. Achieving finite element mesh quality via optimization of the jacobian matrix norm and associated quantities. Part II—a framework for volume mesh optimization and the condition number of the jacobian matrix. *International Journal for Numerical Methods in Engineering* 48, 1165–1185.
- Knupp, P.M., Ernst, C., Thompson, D.C., Stimpson, C., Pebay, P.P., 2006. The verdict geometric quality library. Technical Report. Sandia National Laboratories.
- Lillicrap, T.P., Hunt, J.J., Pritzel, A., Heess, N., Erez, T., Tassa, Y., Silver, D., Wierstra, D., 2015. Continuous control with deep reinforcement learning. arXiv:1509.02971 .
- Liu, C., Yu, W., Chen, Z., Li, X., 2017. Distributed poly-square mapping for large-scale semi-structured quad mesh generation. *Computer-Aided Design* 90, 5–17.
- Mazyavkina, N., Sviridov, S., Ivanov, S., Burnaev, E., 2021. Reinforcement learning for combinatorial optimization: a survey. *Computers & Operations Research* 134, 105400.

- Mnih, V., Badia, A.P., Mirza, M., Graves, A., Lillicrap, T., Harley, T., Silver, D., Kavukcuoglu, K., 2016. Asynchronous methods for deep reinforcement learning, in: International Conference on Machine Learning, PMLR. pp. 1928–1937.
- Nechaeva, O., 2006. Composite algorithm for adaptive mesh construction based on self-organizing maps, in: International Conference on Artificial Neural Networks, Springer. pp. 445–454.
- Pan, J., Huang, J., Wang, Y., Cheng, G., Zeng, Y., 2021. A self-learning finite element extraction system based on reinforcement learning. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing* 35, 180–208.
- Papagiannopoulos, A., Clausen, P., Avellan, F., 2021. How to teach neural networks to mesh: application on 2-D simplicial contours. *Neural Networks* 136, 152–179.
- Raffin, A., Hill, A., Gleave, A., Kanervisto, A., Ernestus, M., Dormann, N., 2021. Stable-baselines3: reliable reinforcement learning implementations. *Journal of Machine Learning Research* 22, 1–8.
- Remacle, J.F., Henrotte, F., Carrier-Baudouin, T., Béchet, E., Marchandise, E., Geuzaine, C., Mouton, T., 2013. A frontal delaunay quad mesh generator using the  $\ell_\infty$  norm. *International Journal for Numerical Methods in Engineering* 94, 494–512.
- Remacle, J.F., Lambrechts, J., Seny, B., Marchandise, E., Johnen, A., Geuzainet, C., 2012. Blossom-quad: a non-uniform quadrilateral mesh generator using a minimum-cost perfect-matching algorithm. *International Journal for Numerical Methods in Engineering* 89, 1102–1119.
- Schulman, J., Wolski, F., Dhariwal, P., Radford, A., Klimov, O., 2017. Proximal policy optimization algorithms. arXiv:1707.06347 .
- Shewchuk, J.R., 2012. Unstructured mesh generation. *Combinatorial Scientific Computing* 12, 2.
- Silver, D., Huang, A., Maddison, C.J., Guez, A., Sifre, L., Van Den Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M.,

- et al., 2016. Mastering the game of go with deep neural networks and tree search. *Nature* 529, 484–489.
- Slotnick, J., Khodadoust, A., Alonso, J., Darmofal, D., Gropp, W., Lurie, E., Mavriplis, D., 2014. CFD vision 2030 study: a path to revolutionary computational aerosciences. Technical Report CR-2014-218178. National Aeronautics and Space Administration, Langley Research Center.
- Sutton, R.S., Barto, A.G., 1998. Reinforcement learning: an introduction. MIT press.
- Sutton, R.S., Barto, A.G., 2018. Reinforcement learning: an introduction, second edition. MIT press.
- Verma, C.S., Suresh, K., 2017. A robust combinatorial approach to reduce singularities in quadrilateral meshes. *Computer-Aided Design* 85, 99–110.
- Verma, C.S., Suresh, K., 2018.  $\alpha$ mst: a robust unified algorithm for quadrilateral mesh adaptation in 2D and 3D. *Computer-Aided Design* 103, 47–60.
- Vinyals, O., Fortunato, M., Jaitly, N., 2015. Pointer networks, in: Advances in Neural Information Processing Systems, pp. 2692–2700.
- Wang, N., Zhang, Y., Li, Z., Fu, Y., Liu, W., Jiang, Y.G., 2018a. Pixel2mesh: generating 3D mesh models from single rgb images, in: Proceedings of the European Conference on Computer Vision (ECCV), pp. 52–67.
- Wang, W.Y., Li, J., He, X., 2018b. Deep reinforcement learning for nlp, in: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics: tutorial abstracts, pp. 19–21.
- White, D.R., Kinney, P., 1997. Redesign of the paving algorithm: robustness enhancements through element by element meshing, in: 6th International Meshing Roundtable, Citeseer. p. 830.
- Yang, J., Dzanic, T., Petersen, B., Kudo, J., Mittal, K., Tomov, V., Camier, J.S., Zhao, T., Zha, H., Kolev, T., et al., 2021. Reinforcement learning for adaptive mesh refinement. arXiv:2103.01342 .
- Yao, J., Stillman, D., 2019. An angular method with position control for block mesh squareness improvement. Springer International Publishing. pp. 129–147.

- Yao, S., Yan, B., Chen, B., Zeng, Y., 2005. An ann-based element extraction method for automatic mesh generation. *Expert Systems with Applications* 29, 193–206.
- Zeng, Y., Cheng, G., 1993. Knowledge-based free mesh generation of quadrilateral elements in two-dimensional domains. *Computer-Aided Civil and Infrastructure Engineering* 8, 259–270.
- Zeng, Y., Yao, S., 2009. Understanding design activities through computer simulation. *Advanced Engineering Informatics* 23, 294–308.
- Zhang, Z., Wang, Y., Jimack, P.K., Wang, H., 2020. Meshingnet: a new mesh generation method based on deep learning, in: *International Conference on Computational Science*, Springer. pp. 186–198.
- Zhuang, F., Qi, Z., Duan, K., Xi, D., Zhu, Y., Zhu, H., Xiong, H., He, Q., 2021. A comprehensive survey on transfer learning. *Proceedings of the IEEE* 109, 43–76. doi:10.1109/JPROC.2020.3004555.
- Ziebart, B.D., 2010. Modeling purposeful adaptive behavior with the principle of maximum causal entropy. Ph.D. thesis. Carnegie Mellon University.