Generalization in Deep RL for TSP Problems via Equivariance and Local Search

A Preprint

Wenbin Ouyang∗, Yisen Wang∗
University of Michigan
Shanghai Jiao Tong University
{oywenbin, yiwenw}@umich.edu

Paul Weng†, Shaochen Han
Shanghai Jiao Tong University
{paul.weng, sc.han}@sjtu.edu.cn

October 8, 2021

Abstract

Deep reinforcement learning (RL) has proved to be a competitive heuristic for solving small-sized instances of traveling salesman problems (TSP), but its performance on larger-sized instances is insufficient. Since training on large instances is impractical, we design a novel deep RL approach with a focus on generalizability. Our proposition consisting of a simple deep learning architecture that learns with novel RL training techniques, exploits two main ideas. First, we exploit equivariance to facilitate training. Second, we interleave efficient local search heuristics with the usual RL training to smooth the value landscape. In order to validate the whole approach, we empirically evaluate our proposition on random and realistic TSP problems against relevant state-of-the-art deep RL methods. Moreover, we present an ablation study to understand the contribution of each of its components.

1 Introduction

The traveling salesman problem (TSP) has numerous applications from supply chain management [Snyder and Shen, 2019] to electronic design automation [Gerez, 1999] or bioinformatics [Jones and Pevzner, 2004]. As an NP-hard problem, solving large-sized problem instances is generally intractable. Deep reinforcement learning (RL)-based heuristic solvers have been demonstrated [Bello et al., 2016, Dai et al., 2017, Kool et al., 2019, Ma et al., 2019] as being able to provide competitive solutions while being fast, which is crucial in many domains such as logistics for real-time operations. However, this ability has only been demonstrated on small-sized problems where RL-based solvers are usually trained and tested on small instances. When evaluated instead on larger instances, such solvers perform poorly. Since the computational cost of training on large instances is prohibitive, increasing the size of the training instances is not a practical option for obtaining efficient solvers for larger instances. To overcome this limitation, this paper investigates techniques to increase the generalization capability of deep RL solvers, which would allow to train on small instances and then solve larger ones.

Generalization in deep RL [François-Lavet et al., 2018] can be improved by adjusting the following three components: input/representation space, objective function, and learning algorithm. We propose to achieve this by exploiting equivariance, local search, and curriculum learning, which we explain next. Although our approach is demonstrated on TSP problems, we believe it can be naturally extended to other combinatorial optimization problems, but also to more classic RL domains. We leave this for future work.

Problems with spatial information, such as TSP problems, enjoy many spatial symmetries which can be exploited for RL training and generalization. An RL solver is invariant with respect to some symmetry, if its outputs remain unchanged for symmetric inputs. For instance, translating all the city positions leave optimal solutions intact. More generally, an RL solver is equivariant, if its outputs are also transformed with a corresponding symmetry for symmetric inputs. For instance, if cities are permuted, a corresponding permutation is required on the outputs of the RL solver.

∗Equal contribution
†Corresponding author
Invariance and equivariance of the solver with respect to some symmetries allows a smaller input space to be considered during training and more abstract representation insensitive to those symmetries to be learned while the trained solver still covers the whole input space.

Previous work has considered using local search as a simple additional step to improve the feasible solution output by a RL-based solver. In contrast to previous work, we further employ local search as a tool to smooth the objective function optimized during RL training. To that aim, we interleave RL training with local search using the improved solution provided by the latter to train the RL solver via a modified policy gradient. Moreover, we propose a novel simple baseline called the policy rollout baseline to reduce the variance of its estimation.

Curriculum learning [Soviany et al., 2021] can accelerate training, but also improve generalization [Weinshall et al., 2018]. By modulating the difficulty of the training instances, as described by instance sizes, the RL-based solver can learn more easily and faster. Since the solver sees various sizes of instances, this may prevent it to overfit to one particular instance size.

The architecture of our model includes a graph neural network (GNN), a multi-layer perceptron (MLP), and an attention mechanism. Due to all the used techniques, we name our model as eMAGIC (equivariance for e, MLP for M, Attention for A, Graph neural network for G, Interleaved local search for I, and Curriculum learning for C). We demonstrate that it can be trained on small random instances (up to 50 cities) and perform competitively on larger random or realistic instances (up to 1000 cities). The only learning methods that can perform better than ours require learning on the instance to be solved (see section 2 for a discussion).

2 Related Work

Research on exploiting deep learning [Vinyals et al., 2015, Li et al., 2018, Joshi et al., 2019a, Prates et al., 2019, Xing and Tu, 2020] or RL [Bello et al., 2016, Dai et al., 2017, Kool et al., 2019, Deudon et al., 2018, Ma et al., 2019, Da Costa et al., 2020, Zheng et al., 2021, Wu et al., 2021a] to design heuristics for solving TSP problems has become very active. The current achievements demonstrate the potential of machine-learning-based approaches, but also reveal their usual limitation on solving larger-size instances, which has prompted much research on generalization [Joshi et al., 2019b, Fu et al., 2021, Wu et al., 2021a]. Some experimental work [Joshi et al., 2019b] suggests that deep RL may provide more generalizable solvers than supervised learning, which is a further motivation for our own work. Another advantage of deep RL is that it does not require optimal solvers to train. We discuss next the related work that exploits equivariance or local search like our method. To the best of our knowledge, no other work uses curriculum learning for designing an RL-based solver for TSP. For space reasons, our discussion below emphasizes the approaches using deep RL as a constructive heuristic (which builds a solution iteratively).

Invariance and equivariance have been important properties to exploit for designing powerful deep learning architectures [Bengio and Lecun, 1997, Gens and Domingos, 2014, Cohen and Welling, 2016]. They have also inspired some of the previous work on solving TSP problems, although they were often not explicitly discussed like in our work. For instance, Deudon et al., 2018 use principal component analysis to exploit rotation invariance as a single preprocessing step. In contrast to their work, we compose various preprocessing transformations to be applied at every solving step. To the best of our knowledge, we are the first to propose such technique. In addition, permutation invariance is the motivation for using attention models [Kool et al., 2019, Deudon et al., 2018] or GNNs [Ma et al., 2019] in RL solvers. Contrary to Deudon et al., 2018, Kool et al., 2019 select the next city to visit based on the first and last visited cities while Ma et al., 2019 propose to use relative positions for translation invariance. Our proposition combines those two ideas, but compared to Kool et al., 2019, our model is based on GNN, while compared to Ma et al., 2019, it has a simpler architecture that does not require an LSTM. To summarize, compared to all previous work, our paper investigates in a more systematic fashion the exploitation of equivariance to help training and improve generalization.

Several previous studies combine deep learning or RL with various local search techniques to find better solutions: 2-opt [Deudon et al., 2018, Ma et al., 2019, Da Costa et al., 2020, Wu et al., 2021a], k-opt [Zheng et al., 2021], or Monte-Carlo tree search [Fu et al., 2021]. In contrast to these methods, we use a combined local search technique, which applies several heuristics in an efficient way. Moreover, to the best of our knowledge, no previous work considers interleaving local search with policy gradient updates to obtain a more synergetic final method.

All these machine-learning-based methods can be categorized as learning from a batch of instances or as directly learning on the instance to be solved. Like our work, most previously-discussed methods are part of the first category. However, some recent work belongs to the second category [Zheng et al., 2021] or are hybrid [Fu et al., 2021]. For instance, Zheng et al., 2021 apply RL to make the LKH algorithm [Helsgaun, 2017], a classic efficient heuristic for TSP, adaptive to the instance to be solved. In Fu et al., 2021’s method, evaluations learned in a supervised way from batch of instances are used in Monte-Carlo tree search so that solutions can be found in an adaptive for the current
instance. Being adaptive to the current instance can undoubtedly boost the performance of the solver. We leave for future work the improvement of our solver to become more adaptive to the instance to be solved.

3 Preliminaries

Following previous work, we focus on the symmetric 2D Euclidean TSP, which we recall below. We then explain how a TSP instance can be tackled with RL. We first provide some notations:

For any positive integer \( N \in \mathbb{N} \), \([N]\) denotes the set \( \{1, 2, \ldots, N\} \). Vectors and matrices are denoted in bold (e.g., \( x \) or \( X \)). For a set of subscripts \( I \), \( X_I \) denotes the matrix formed by the rows of \( X \) whose indices are in \( I \).

**Traveling Salesman Problem** A symmetric 2D Euclidean TSP instance is described by a set of \( N \) cities (identified to the set \([N]\) and their coordinates in \( \mathbb{R}^2 \). The goal in this problem is to find the shortest tour that visits each city exactly once. The coordinates of city \( i \) are denoted \( x_i \in \mathbb{R}^2 \). The matrix whose rows correspond to city coordinates is denoted \( X \in \mathbb{R}^{N \times 2} \). A feasible solution (i.e., tour) of a TSP instance is a permutation \( \sigma \) over \([N]\) with length equal to:

\[
L_\sigma(X) = \sum_{t=1}^{N} \|x_{\sigma(t)} - x_{\sigma(t+1)}\|_2
\]

where \( \| \cdot \|_2 \) denotes the \( L_2 \)-norm, \( \sigma(t) \in [N] \) is the \( t \)-th visited city in tour \( \sigma \), and by abuse of notation, \( \sigma(N+1) \) denotes \( \sigma(1) \). Therefore, solving a TSP instance consists in finding the permutation \( \sigma \) that minimizes the tour length \( L_\sigma(X) \) defined in Equation 1. Since TSP solutions are invariant to scaling, we assume that the city coordinates are in the square \([0,1]^2\). In Appendix A, we recall some heuristics (e.g., insertion, \( k \)-opt) that have been proposed for solving TSP.

**RL as a Constructive Heuristic** RL can be used to construct a TSP tour \( \sigma \) sequentially. Intuitively, at iteration \( t \in [N] \), an RL solver (i.e., policy) selects the next unvisited city \( \sigma(t) \) to visit based on the current partial tour and the description of the TSP instance (i.e., coordinates of cities). Therefore, this RL problem corresponds to a repeated \( N \)-horizon sequential decision-making problem. As noticed by [Kool et al., 2019], the decision for the next city to visit only depends on the description of the TSP instance and the first and last visited cities. In addition, we update the description of the TSP instance by removing the coordinates of visited cities. Surprisingly, to the best of our knowledge, no previous work exploits this simplified RL model.

Formally, in the RL language, at time step \( t \in [N] \), an action \( a_t \) represents the next city to visit, i.e., \( a_t = \sigma(t) \). Let \( I_1 = [N] \) and \( I_{t+1} = [N] \setminus \{\sigma(1), \ldots, \sigma(t)\} \) be the set of remaining cities after \( t \) cities have been already visited. Moreover, let \( J_1 = [N] \) and \( J_t = I_t \cup \{\sigma(1), \sigma(t)\} \) be the set of unvisited cities in addition of the first and last visited cities (\( \sigma(1) \) and \( \sigma(t) \)). Therefore, \( a_t \in I_t \) for all \( t \in [N] \). A state \( s_t \) can be represented by a matrix \( X_{I_t} \) with flags indicating the first and last visited cities. This matrix includes the coordinates of unvisited cities \((X_{I_1})\) in addition to the coordinates of the first and last visited cities \((x_{\sigma(1)})\) and \((x_{\sigma(t)})\). Note that at \( t = 1 \), no city has been chosen yet, so the initial state \( s_1 \) only contains the list of city coordinates. A state \( s_N \) contains the coordinates of the last unvisited city and those of the first and last visited cities. After choosing the last city to visit in state \( s_N \), the tour \( \sigma \) is completely generated. The immediate reward \( r(s_t, a_t) \) for an action \( a_t \) in a state \( s_t \) can be defined as the negative length between the last visited city and the next chosen city, since we want the tour length to be small:

\[
r(s_t, a_t) = \begin{cases} 0 & \text{for } t = 1 \\ -\|x_{\sigma(t)} - x_{\sigma(t-1)}\|_2 & \text{for } t = 2, \ldots, N - 1 \\ -\|x_{\sigma(N)} - x_{\sigma(N-1)}\|_2 - \|x_{\sigma(1)} - x_{\sigma(N)}\|_2 & \text{for } t = N \end{cases}
\]

After choosing the first city \( a_1 = \sigma(1) \), the reward is zero since no length can be computed. After the final action \( a_N \), an additional reward \(-\|x_{\sigma(1)} - x_{\sigma(N)}\|_2\) is added to complete the tour length.

In deep RL, a policy \( \pi_\theta \) is represented as a neural network parametrized by \( \theta \). The goal is then to find \( \theta^* \) that maximizes the objective function \( J(\theta) \):

\[
\theta^* = \arg\max_\theta J(\theta) = \arg\max_\theta \mathbb{E}_{\tau \sim \rho_\theta(r)} \left[ \sum_{t=1}^{N} r_t \right],
\]

where for all \( t \in [N] \), \( r_t = r(s_t, a_t) \), \( \tau = (s_1, a_1, s_2, a_2, \ldots, s_N, a_N) \) is a complete trajectory, and \( \rho_\theta \) is the probability over trajectories induced by policy \( \pi_\theta \). This objective function \( J(\theta) \) can be optimized by a policy gradient [Williams, 1992] or actor-critic method [Sutton and Barto, 1998].
4 Equivariant Model

In this section, we present several techniques to exploit invariances and equivariances in the design of an RL solver. Formally, a mapping \( f : A \rightarrow B \) from a set \( A \) to a set \( B \) is invariant with respect to a symmetry \( \rho^A : A \rightarrow A \) iff \( f(x) = f(\rho^A(x)) \) for any \( x \in A \). More generally, given a symmetry that acts on \( A \) with \( \rho^A : A \rightarrow A \) and on \( B \) with \( \rho^B : B \rightarrow B \), a mapping \( f : A \rightarrow B \) is equivariant with respect to this symmetry iff \( \rho^B(f(x)) = f(\rho^A(x)) \) for any \( x \in A \). This definition shows that invariance is a special case of equivariance when \( \rho^B \) is the identity function. Intuitively, equivariance for an RL solver (resp. its value function) means that if a transformation is applied to its input, its output (resp. its value) can be recovered by a corresponding transformation.

In the remaining of the paper, for simplicity, we often use equivariance to refer to both equivariance and invariance, since the latter is a special case of the former. Equivariance can be exploited in RL in various ways. We consider some equivariant preprocessing methods on the description of TSP instances to standardize the kinds of instances the solver is trained on and evaluated on. In addition, we propose a simple deep learning model for which we apply some other equivariant preprocessing methods on its inputs to further reduce the input space.

4.1 Equivariant Preprocessing of TSP Instance

An RL solver should be invariant with respect to any Euclidean symmetry (rotation, reflection, translations and their compositions) and to any positive scaling transformation applied on city positions. To enforce these invariances, we can apply these transformations to preprocess the inputs of the solver such that the transformed inputs are always in a standard form. Doing so allows the solver to be trained on more similar inputs.

Concretely, for rotation invariance, we rotate and scale the city positions such that they are mostly distributed along the first diagonal of the square \([0, 1]^2\). This can be achieved by performing a principal component analysis, rotating the first found axis by \(45^\circ\) anti-clockwise and scaling to fit the cities in the \([0, 1]^2\) square. This transformation, which is slightly different from [Deudon et al., 2018], allows the cities to be as spread as possible in \([0, 1]^2\). For scaling and translation invariance, we apply a scaling and translation transformation to the city positions such that there are a maximum number of cities (i.e., 2 or 3 depending on configuration) on the border of the square \([0, 1]^2\). For reflections, we only consider horizontal, vertical, and diagonal flips (with respect to the coordinate axes) for simplicity. Cities are reflected such that a majority of them are in a fixed chosen region.

Since the symmetries can be composed, these preprocessing methods can be applied sequentially. Theoretically, it would be beneficial to apply all of them in combination, however this has a computational cost. It is therefore more effective to only use a selection of the most effective ones. Moreover, these preprocessing methods can be applied either once on the initial TSP instance (\(X\)), or at each solving step \(t\) on the remaining cities (\(X_{I_t}\)). Theoretically, performing these preprocessings iteratively should help most, since the RL solver is only trained on standardized inputs. This is confirmed by our experimental analysis. We find out that the combination that provides the best results are rotation followed by scaling and translation, which will be used in the experimental evaluation of our method. However, our overall approach is generic and could include any other symmetries for which equivariance would hold for the RL solver. In Appendix E we present the details of the evaluation of the different preprocessing methods.

In the remaining, we assume that the set of equivariant transformations is fixed. For any set of indices \(I \subseteq [N]\), we denote the matrix of positions \(X_I\) after preprocessing by \(\widetilde{X}_I\) and any coordinates \(\bar{x}\) after preprocessing by \(\bar{x}\). Note that the preprocessing step (e.g., if it includes the rotation transformation described above) may depend on the initial matrix \(X_I\), but we prefer not to reflect it in the notations to keep them simple.

4.2 Equivariant Model

Our proposed model, which represents the RL solver, has an encoder-decoder architecture (see Figure 1). In our model, the first city to visit is fixed arbitrarily, since the construction of the tour should be invariant to the starting city. Therefore, the decisions of the solver starts at time step \(t \geq 2\). Our model is designed to take into account other equivariances that are known to hold in the problem. Invariance with respect to translation can be further exploited by considering relative positions with respect to the last visited city instead of the original absolute positions, as suggested by [Ma et al., 2019]. Therefore, we define the input of our RL solver to only include two pieces of information. First, the information about the current partial tour now only needs the relative preprocessed position of the first visited city \((\widetilde{x}_{\sigma(1)} - \widetilde{x}_{\sigma(t)})\). Since the last visited city is always represented by the origin, it can be dropped. Second, the information about the remaining TSP instance corresponds to the relative preprocessed positions of the unvisited cities \((\widetilde{x}_{\sigma(n)} - \widetilde{x}_{\sigma(t-1)}\) for \(n \in I_t\) and of the first and last visited cities \((\widetilde{x}_{\sigma(1)}\) and the origin \(\widetilde{x}_{\sigma(t)} = 0\). We denote the matrix of those relative preprocessed positions by \(\widetilde{X}_{I_t}\), i.e., the matrix whose rows are the rows of \(\widetilde{X}_{I_t}\).
(obtained from $X_{J_t}$) minus $\bar{x}_{\sigma(t)}$. The last visited city needs to be kept in this matrix, since it represents the labels of the nodes of the remaining graph on which the tour should be completed.

**Encoder** We denote the dimension of the embeddings by $H$. The encoder of our model is composed of a graph neural network (GNN) \cite{battaglia2018relational}, which computes an embedding $\hat{X}_{J_t} \in \mathbb{R}^{|J_t| \times H}$ of the relative city positions $\bar{X}_{J_t} \in \mathbb{R}^{|J_t| \times 2}$, and a multilayer perceptron (MLP), which computes an embedding $\hat{x}_{\sigma(1)} \in \mathbb{R}^H$ of the relative position of the first visited city $\bar{x}_{\sigma(1)} \in \mathbb{R}^2$. The GNN encodes the information about the graph describing the remaining TSP problem. As a GNN, it is invariant with respect to the order of its inputs and its outputs depend on the graph structure and information (i.e., city positions). The MLP encodes the information about visited cities. Since we exploit the independence with respect to the visited cities between the first and last cities (not included), our model does not require any complex architecture, like LSTM \cite{hochreiter1997long} or GNN, for encoding the positions of the visited cities, in contrast to most previous work. A simple MLP suffices since only the relative position of the first city is required.

Formally, the GNN computes its output $\hat{X}_{J_t} \in \mathbb{R}^{|J_t| \times H}$ from inputs $\bar{X}_{J_t} \in \mathbb{R}^{|J_t| \times 2}$ through $n_{\text{GNN}}$ layers, which allows each city to get information from its neighbors up to $n_{\text{GNN}}$ edges away:

\[
X^{(0)} = \bar{X} \Theta^{(0)}
\]

\[
X^{(\ell)} = \lambda \cdot X^{(\ell-1)} \Theta^{(\ell)} + (1 - \lambda) \cdot F^{(\ell)} \left( \frac{X^{(\ell-1)}}{|J_t| - 1} \right)
\] (5)

where $\Theta^{(0)} \in \mathbb{R}^{2 \times H}$ and $\Theta^{(\ell)} \in \mathbb{R}^{H \times H}$ are learnable weights, $X^{(\ell-1)} \in \mathbb{R}^{|J_t| \times H}$ is the output of the $\ell$th layer of the GNN for $\ell \in \{1, \ldots, n_{\text{GNN}}\}$, $X^{(\text{gnn})} = \hat{X}_{J_t}$, $F^{(\ell)} : \mathbb{R}^{|J_t| \times H} \rightarrow \mathbb{R}^{|J_t| \times H}$ is the aggregation function, which is implemented as a neural network, and $\lambda \in [0, 1]$ is another trainable parameter. Visual illustration of the GNN is deferred to Appendix $\ref{app:gnn}$ for space reasons.

**Decoder** Once the embeddings $\hat{X}_{J_t}$ and $\hat{x}_{\sigma(0)}$ are computed, the probability of selecting a next city to visit is obtained via a standard attention mechanism \cite{bello2016neural} using $\hat{X}_{J_t}$ and $\hat{x}_{\sigma(0)}$ as the keys and query respectively. Formally, the decoder outputs a vector $u \in \mathbb{R}^N$ expressed as:

\[
u_j = \begin{cases} -\infty & \forall j \in I_t \\
w \cdot \tanh \left( \hat{X}_{J_t,j} \Theta_g + \Theta_m \hat{x} \right) & \text{otherwise} \end{cases}
\] (6)

where $\nu_j$ is the $j$th entry of the vector $u$, $\hat{X}_{J_t,j}$ is the $j$th row of the matrix $\hat{X}_{J_t}$, $\Theta_g$ and $\Theta_m$ are trainable matrices with shape $H \times H$, $w \in \mathbb{R}^N$ is another trainable weight vector. Then, a softmax transformation turns $\nu$ into a probability distribution $p = (p_j)_{j \in [N]}$ over unvisited cities:

\[
p = \text{softmax}(\nu) = \left( \frac{e^{\nu_j}}{\sum_{j=1}^N e^{\nu_j}} \right)_{j \in [N]}
\] (7)

where $p_j$ is the $j$th entry of the probability distribution $p$ and $\nu_j$ is the $j$th entry of the vector $\nu$. Note that the probability of any visited city $j \in I_t$ is zero since $\nu_j = -\infty$ in Equation $\ref{eq:softmax}$.
5 Algorithm & Training

We introduce three innovative training techniques that we apply during our training process: combined local search, smoothed policy gradient, and stochastic curriculum learning. Overall, these techniques can help train a fast and generalizable policy, which is able to generate tours that can easily be improved by local search. The overall algorithm can be found in Appendix C.

Combined Local Search In contrast to previous work, which generally only considers 2-opt, we propose to use a combination of several (possibly random) local search methods (i.e., local insertion heuristic, random 2-opt, search 2-opt and search random 3-opt) to improve a tour generated by the RL solver. For space reasons, we defer their presentations in Appendix A. A local search method can heuristically improve a tour, but may get stuck in a local optimum. Using a combination of them alleviates this issue, since different heuristics usually have different local optima. Another important distinction with previous work is that we also use local search during training, not only testing.

Smoothed Policy Gradient For simplicity, we train our model with the REINFORCE algorithm [Williams, 1992], which leverages the policy gradient for optimization. However, instead of using the standard policy gradient, which is based on the value of the tour generated by the policy, we compute the policy gradient with the value of the tour improved by our combined local search. While standard RL training may yield policies whose outputs may not easily be improved by local search, our new definition directly trains the RL solver to find solutions that can be improved by local search, which allows RL and local search to have synergetic effects.

Intuitively, this new policy gradient amounts to smoothing the objective function $J(\theta)$ that is optimized. Recall $J(\theta) = -\mathbb{E}[L_\sigma(X)]$, where the expectation is taken with respect to $\sigma$, which is a random variable corresponding to the tour generated by policy $\pi_\theta$. In our approach, this usual objective function is replaced by $J^+(\theta) = -\mathbb{E}[L_{\sigma'}(X)]$, where $\sigma'$ is a random variable corresponding to the improved tour obtained by our combined local search from $\sigma$. Therefore, this last expectation is taken with respect to the probability distributions generated by policy $\pi_\theta$ and our combined local search. This objective function can be understood as $J^+(\theta) = -\mathbb{E}[\min_{\sigma' \in N(\sigma)} L_{\sigma'}(X)]$ where $N(\sigma)$ is a neighborhood of $\sigma$ defined by local search. This stochastic min operation has a smoothing effect on $J(\theta)$. That is why, we call the gradient of $J^+(\theta)$ a smoothed policy gradient.

Formally, this novel policy gradient can be estimated on a batch of TSP instances $X^{(1)}, \ldots, X^{(B)}$:

$$\nabla_\theta J^+(\theta) \approx -\frac{1}{B} \sum_{b=1}^B \left( \sum_{t=2}^N \nabla_\theta \log \pi_\theta(a_t^{(b)}|s_t^{(b)}) \right) \left( L_{\sigma'^{(b)}}(X^{(b)}) - l^{(b)} \right),$$

(8)

where $\sigma'^{(b)}$ is the tour generated by the current policy $\pi_\theta$ on instance $X^{(b)}$, $\sigma'^{(b)}$ is the improved tour obtained by our combined local search starting from $\sigma'^{(b)}$, and $l^{(b)} = -L_{\sigma'^{(b)}}(X^{(b)})$ is a novel baseline, which we call policy rollout baseline, used to reduce the variance of the policy gradient estimation. It enjoys the nice property that it does not require additional calculations, since $L_{\sigma'^{(b)}}(X^{(b)})$ is computed when $\sigma'^{(b)}$ is generated. In our experiments, the policy rollout baseline easily outperforms the previous greedy baselines [Ma et al., 2019; Kool et al., 2019].

Note that by construction, $L_{\sigma'^{(b)}}(X^{(b)}) \leq L_{\sigma'^{(b)}}(X^{(b)}) = -\sum_{t=1}^N s_t^{(b)}$. Therefore, the smoothed policy gradient updates more if our combined local search can make more improvements upon this certain policy. For completeness, we provide more details in Appendix B.

Stochastic Curriculum Learning Curriculum Learning (CL) is widely used in machine learning [Soviany et al., 2021]. Its basic principle is to control the increase of the difficulty of training instances. CL can speed up learning and improve generalization [Weinshall et al., 2018]. We apply stochastic CL to train our model. Instead of a deterministic process, stochastic CL increases the difficulty according to a probability distribution. We choose the TSP size (i.e., number of cities) as a measure of difficulty for a TSP instance. We assume it to be in the range of $R = \{10, 11, \ldots, 50\}$. For each epoch $e$, we define the vector $g^{(e)} \in \mathbb{R}^{41}$ (since $|R| = 41$) to be:

$$g^{(e)}_k = \frac{1}{\sqrt{2\pi}\sigma_N} \exp\left(-\frac{1}{2} \frac{(k+10) - e}{\sigma_N}^2\right),$$

(9)

where $g^{(e)}_k$ is the $k$-th entry of $g^{(e)}$ and hyperparameter $\sigma_N$ is the standard deviation of this Gaussian density function. Then, $g^{(e)}$ is turned into a categorical distribution $p^{(e)} \in [0, 1]^{41}$ via a softmax:

$$p^{(e)} = \text{softmax}(g^{(e)})$$

(10)

Our implementation takes advantage of GPU acceleration when possible. The source code will be shared after publication.
The $k$-th entry of $p^{(e)}$ gives the probability of choosing a TSP instance of size $(k + 10)$ at epoch $e$.

6 Experimental Results

We present three sets of experiments. First, to validate the effectiveness of eMAGIC, we train our model on randomly-generated TSP instances (using CL with sizes up to 50), and test the model on other randomly-generated TSP instances (TSP$_n$ where size $n = 20$ up to 1000). Second, to further prove of its generalization capability, we directly evaluate models trained on random instances on realistic symmetric 2D Euclidean TSP instances with sizes range from 51 to 1002 in TSPLIB [Reinelt, 1991]. Third, we conduct an ablation study to show the significance of every component of eMAGIC (i.e., equivariance, stochastic CL, policy rollout baseline, combined local search, and RL). We evaluate three versions of our model: eMAGIC, eMAGIC(s), and eMAGIC(S). In the first version, only one improved solution is sampled (i.e., generated from the RL policy and improved by our combined local search). In the second and third versions, we sample respectively 10 and 100 improved solutions and keep the best tours. The details about the experimental settings and the used hyperparameters, which are the same for all experiments, are provided in Appendix D.

Performance on Randomly Generated TSP  We compare the performance of our model with 12 other methods in Table 1 and 2, which covers various types of TSP solvers including exact solvers, heuristics, and learning-based approaches. Column 1 and 2 of both Table 1 and 2 corresponds to the method name and the method type, respectively. Columns 3, 4, and 5 provide the average tour length, gap to the optimal (provided by Concorde [Applegate et al., 2004]), and computational time, respectively.

Table 1 and 2 show that the computational times of exact solvers become prohibitive as the TSP size increases, with Gurobi not being to solve TSP instances larger than 200 under a reasonable time budget. The classic heuristic methods are relatively fast, but their performances are not satisfactory. Among all learning methods, Fu et al. [2021] provide excellent results, but it is based on Monte-Carlo tree search, which adapts to the instance to be solved. Without this adaptivity, our model with or without sampling provides competitive results. It can be better than Fu et al. [2021] up to TSP200. This somewhat suggests the limit of our approach, which trains on small instances and directly generalizes to large ones, without learning on the instance to be solved. We expect that our approach could be improved by training on slightly bigger instances (e.g., up to 100) or adding an adaptive component like in Fu et al. [2021].

Table 1: Results of eMAGIC vs baselines, tested on 10,000 instances for TSP 20, 50 and 100.

| Method   | Type | TSP20 | TSP50 | TSP100 |
|----------|------|-------|-------|--------|
|          |      | Len. | Gap  | Time   | Len. | Gap  | Time   | Len. | Gap  | Time   |
| Concorde*| ES   | 3.830| 0.00%| 2.3m   | 5.691| 0.00%| 13m   | 7.761| 0.00%| 1.0h   |
| Gurobi*  | ES   | 3.830| 0.00%| 2.3m   | 5.691| 0.00%| 26m   | 7.761| 0.00%| 3.6h   |
| LKH3*    | H    | 3.830| 0.00%| 21m   | 5.691| 0.00%| 27m   | 7.761| 0.00%| 50m    |
| 2-opt    | H    | 4.082| 6.56%| 0.3s  | 6.444| 13.24%| 2.3s  | 9.100| 17.26%| 9.3s   |
| Random†  | H    | 4.005| 4.57%| 3.3m  | 6.128| 7.69%| 12m   | 8.511| 9.66%| 21m    |
| Nearest† | H    | 4.332| 13.10%|3.8m  | 6.780| 19.14%| 10m  | 9.462| 21.92%| 21m    |
| Farthest†| H    | 3.932| 2.64%| 4.0m  | 6.010| 5.62%| 10m   | 8.360| 7.71%| 21m    |
| GCN†     | SL(G)| 3.855| 0.65%| 19s  | 5.893| 3.56%| 2.0m  | 8.413| 8.40%| 11m    |
| AGCRN+M‡ | SL+M | 3.830| 0.00%| 1.6m  | 5.691| 0.01%| 7.9m  | 7.764| 0.04%| 15m    |
| GAT‡     | RL(S)| 3.874| 1.14%| 10m  | 6.109| 7.34%| 20m   | 8.837| 13.87%| 48m    |
| GAT‡     | RL(G)| 3.841| 0.29%| 6.0s  | 5.785| 1.66%| 34s   | 8.101| 4.38%| 1.8m   |
| GAT‡     | RL(S)| 3.832| 0.05%| 17m  | 5.719| 0.49%| 23m   | 7.974| 2.74%| 1.2m   |
| GPN      | RL   | 4.074| 6.35%| 0.8s  | 6.059| 6.47%| 2.5s  | 8.885| 14.49%| 6.2s   |
| eMAGIC   | RL(LS)| 3.844| 0.37%| 3.0s  | 5.763| 1.27%| 17s  | 7.964| 2.61%| 1.3m   |
| eMAGIC(s)| RL(LS)| 3.831| 0.03%| 9.0s  | 5.728| 0.67%| 49s  | 7.852| 1.17%| 2.9m   |
| eMAGIC(S)| RL(LS)| 3.830| 0.00%| 38s  | 5.691| 0.01%| 3.5m  | 7.762| 0.02%| 14.6m  |

† Joshi et al. [2019a], ‡ Fu et al. [2021], § Deudon et al. [2018], ¶ Kool et al. [2019], § Ma et al. [2019]

ES - Exact Solver; H - Heuristic; SL - Supervised Learning; RL - Reinforcement Learning; G - Greedy; S - Sampling; M - Monte-Carlo Tree Search; LS - Combined Local Search.

Random - Random Insertion; Nearest - Nearest Insertion; Farthest - Farthest Insertion.
Table 2: Results of eMAGIC vs baselines, tested on 10,000 instances for TSP 200, 500, and 1000.

| Method       | Type†     | TSP200 | TSP500 | TSP1000 |
|--------------|-----------|--------|--------|---------|
|              | Len.      | Gap    | Time   | Len.    | Gap    | Time   | Len.     | Gap    | Time   |
| Concorde     | ES        | 10.72  | 0.00%  | 3.4m    | 16.55  | 0.00%  | 38m     | 23.12  | 0.00%  | 7h     |
| Gurobi       | ES        |        |        |         |        |        |         |        |        |        |
| LKH3         | H         | 10.72  | 0.00%  | 2.0m    | 16.55  | 0.00%  | 11m     | 23.12  | 0.00%  | 38m    |
| 2-opt        | H         | 12.84  | 19.80% | 34s     | 20.44  | 23.51% | 3.3m    | 28.95  | 25.23% | 14m    |
| Random       | H         | 11.84  | 10.47% | 27s     | 18.59  | 12.34% | 1.1m    | 26.12  | 12.98% | 38m    |
| Nearest      | H         | 13.19  | 23.03% | 29s     | 20.61  | 24.59% | 1.3m    | 28.97  | 25.32% | 2.9m   |
| Farthest     | H         | 11.64  | 8.63%  | 33s     | 18.31  | 10.64% | 1.4m    | 25.74  | 11.35% | 2.9m   |
| GCN          | 1SL(G)    | 17.01  | 58.73% | 59s     | 29.72  | 79.61% | 7m      | 48.62  | 110.3% | 29m    |
| AGCRN+M      | 2SL+M     | 10.81  | 0.88%  | 2.5m    | 16.97  | 2.54%  | 5.9m    | 23.86  | 3.22%  | 12m    |
| GAT          | 3RL(S)    | 13.18  | 22.91% | 4.8m    | 28.63  | 73.03% | 20m     | 50.30  | 117.6% | 38m    |
| GAT          | 4RL(G)    | 11.61  | 8.23%  | 4.5m    | 22.64  | 36.84% | 16m     | 42.80  | 85.15% | 11h    |
| GAT          | 4RL(S)    | 11.45  | 6.83%  | 5.0s    | 20.02  | 20.99% | 1.5m    | 31.15  | 34.75% | 3.2m   |
| GPN          | 5RL      | -      | -      | -       | 19.61  | 18.49% | -       | 28.47  | 23.15% | -      |
| GPN+2opt     | 5RL+2opt  | -      | -      | -       | 18.36  | 10.95% | -       | 26.13  | 13.02% | -      |
| eMAGIC       | RL(LS)    | 11.14  | 3.89%  | 34s     | 17.54  | 6.07%  | 1.8m    | 24.74  | 7.05%  | 4.9m   |
| eMAGIC(s)    | RL(LS)    | 10.95  | 2.12%  | 1.1m    | 17.29  | 4.50%  | 4.1m    | 24.50  | 5.99%  | 11m    |
| eMAGIC(S)    | RL(LS)    | 10.77  | 0.50%  | 2.4m    | 17.03  | 2.92%  | 9.7m    | 24.12  | 4.36%  | 27m    |

See footnotes of Table 1.

Performance on Realistic TSP Table 3 compares the performances of a variety of learning-based TSP solvers on instances from TSPLIB. Each column of Table 3 represents the average gap to the optimal solution over the instances indicated by the corresponding rows. We treat the tested instances as small cases if their sizes are under 200 and big cases otherwise. And we extend the testing to larger instances with size up to 1002 and leave the performances of other models as empty for size ranging from 400 to 1002 since in the their papers, the testings stop at instances with sizes around 400. We can observe from Table 3 that our model can perform much better than the other learning-based solvers not only for small problems but also large ones, which demonstrates the strong ability and the practical significance of eMAGIC to tackle realistic TSP problems. When the testing size increases, most models suffer from a relatively big increasing of the average gap while ours only increases less than 1% and remains in a good absolute value (3.40%) for even larger instances, which indicates a strong generalization ability of our model. Further details and experimental results are provided in Appendix H.

Table 3: Gap to optimal for different ranges of instances in TSPLIB.

| Size Range | eMAGIC(S) | Wu et al. | S2VDQN | OR | AM | L2OPT | Furthest |
|------------|-----------|-----------|--------|----|----|-------|----------|
| 50 – 199   | 0.46%     | 15.00%    | 3.77%  | 3.49% | 78.73% | 6.54% | 7.60%    |
| 200 – 399  | 1.37%     | 23.49%    | 6.87%  | 3.61% | 293.76% | 12.17% | 9.54%    |
| 400 – 1002 | 3.40%     | -         | -      | -    | -    | -     | -        |

1 [Wu et al., 2021b], 2 [Khalil et al., 2017], 3 [Perron and Furnon], 4 [Kool et al., 2019], 5 [de O. da Costa et al., 2020], 6 [Furthest Insertion].

Ablation Study We demonstrate the strength of all the techniques we applied (including equivariance, stochastic CL, policy rollout baseline, combined local search, and RL) using an ablation study. We turn off each feature one at a time to see if the performance drops compared to the full version of eMAGIC. For equivariance, we remove all the equivariant procedures during training/testing (e.g., deleting visited cities, preprocessing, and using relative positions). For the policy rollout baseline, we replace it with the self-critic baseline, which is a greedy baseline implemented in [Ma et al., 2019]. For combined local search, we only apply it during testing to check if it can help improve the training process. For RL, we directly apply combined local search without performing any learning. Table 4 demonstrates that each technique plays a role in our model. For space reasons, the ablation study for stochastic CL and individual equivariant components is deferred to Appendix F.
Table 4: Ablation study on equivariance, policy rollout baseline, combined local search, and RL, tested on 10,000 instances for TSP 20, 50 and 100, and 128 instances for TSP 200, 500 and 1000.

| TSP Size | Full | w/o Equiv. | w/o BL | w/o LS | w/o RL |
|----------|------|------------|--------|--------|--------|
|          | Len. | Gap        | Len.   | Gap    | Len.   | Gap    | Len.   | Gap    | Len.   | Gap    |
| TSP20    | 3.844| 0.37%      | 3.857  | 0.69%  | 3.875  | 1.17%  | 3.874  | 1.15%  | 3.879  | 1.27%  |
| TSP50    | 5.763| 1.27%      | 5.808  | 2.07%  | 5.859  | 2.96%  | 5.837  | 2.58%  | 5.901  | 3.70%  |
| TSP100   | 7.964| 2.61%      | 8.086  | 4.19%  | 8.124  | 4.68%  | 8.100  | 4.37%  | 8.178  | 5.38%  |
| TSP200   | 11.14| 3.89%      | 11.27  | 5.16%  | 11.33  | 5.71%  | 11.32  | 5.64%  | 11.43  | 6.67%  |
| TSP500   | 17.54| 6.01%      | 17.72  | 7.10%  | 17.74  | 7.22%  | 17.82  | 7.71%  | 17.86  | 7.96%  |
| TSP1000  | 24.75| 7.05%      | 24.94  | 7.86%  | 24.99  | 8.10%  | 25.12  | 8.67%  | 25.15  | 8.80%  |

Equiv. - equivariance; BL - Baseline; LS - Combined Local Search.

7 Conclusion

We presented a combination of novel techniques (notably, equivariance, combined local search, smoothed policy gradient) for designing an RL solver for TSP, which shows a good generalization capability. We demonstrated its effectiveness both on random and realistic instances, which shows that our model can reach state-of-the-art performance. For future work, the approach can be further improved in various ways, e.g., extending it to the actor-critic scheme and exploiting invariances with the critic; or making it adaptive and learn on the instance to be solved. The approach could also be applied on other combinatorial optimization problems and other RL problems.
References

L.V. Snyder and Z.-J.M. Shen. *Fundamentals of Supply Chain Theory*, chapter The Traveling Salesman Problem, pages 403–461. Wiley, 2019.

Sabih H. Gerez. *Algorithms for VLSI Design Automation*, chapter Routing. Wiley, 1999.

Neil C. Jones and Pavel A. Pevzner. *An Introduction to Bioinformatics Algorithms*. MIT Press, 2004.

Irwan Bello, Hieu Pham, Quoc V. Le, Mohammad Norouzi, and Samy Bengio. Neural combinatorial optimization with reinforcement learning. *CoRR*, 2016.

Hanjun Dai, Elias B. Khalil, Yuuy Zhang, Bistra Dilkina, and Le Song. Learning Combinatorial Optimization Algorithms over Graphs. *Advances in Neural Information Processing Systems*, DECEM2017:6349–6359, April 2017.

Wouter Kool, Herke van Hoof, and Max Welling. Attention, learn to solve routing problems!, 2019.

Qiang Ma, Suwen Ge, Danyang He, Darshan Thaker, and Iddo Drori. Combinatorial optimization by graph pointer networks and hierarchical reinforcement learning. *CoRR*, 2019.

Vincent François-Lavet, Peter Henderson, Riashat Islam, Marc G. Bellemare, and Joelle Pineau. An introduction to deep reinforcement learning. *Foundations and Trends in Machine Learning*, 11(3-4), 2018.

Petru Soviany, Radu Tudor Ionescu, Paolo Rota, and Nicu Sebe. Curriculum learning: A survey, 2021.

Daphna Weinshall, Gad Cohen, and Dan Amir. Curriculum learning by transfer learning: Theory and experiments with deep networks. In *ICML*, 2018.

Oriol Vinyals, Meire Fortunato, and Navdeep Jaitly. Pointer networks. In *NIPS*, pages 2692–2700, 2015.

Zhuwen Li, Qifeng Chen, and Vladlen Koltun. Combinatorial Optimization with Graph Convolutional Networks and Guided Tree Search. In *Advances in Neural Information Processing Systems*, volume 2018-Decem, pages 539–548, 2018. URL [http://arxiv.org/abs/1810.10659](http://arxiv.org/abs/1810.10659).

Chaitanya K. Joshi, Thomas Laurent, and Xavier Bresson. An efficient graph convolutional network technique for the travelling salesman problem. *CoRR*, 2019a.

Marcelo O. R. Prates, Pedro H. C. Avelar, Henrique Lemos, Luis Lamb, and Moshe Vardi. Learning to solve np-complete problems - a graph neural network for decision tsp. In *AAAI*, 2019. URL [http://arxiv.org/abs/1809.02721](http://arxiv.org/abs/1809.02721).

Zhihao Xing and Shikui Tu. A graph neural network assisted monte carlo tree search approach to traveling salesman problem. *IEEE Access*, 8:108418–108428, 2020. doi [10.1109/ACCESS.2020.3000236](https://doi.org/10.1109/ACCESS.2020.3000236).

Michel Deudon, Pierre Cournut, Alexandre Lacoste, Yossiri Adulyasak, and Louis Martin Rousseau. Learning heuristics for the tsp by policy gradient. In *CPAIOR*, volume 10848 LNCS, pages 170–181. Springer Verlag, 2018.

Paulo R de O Da Costa, Jason Rhuggenaath, Yingqian Zhang, and Alp Akcay. Learning 2-opt heuristics for the traveling salesman problem via deep reinforcement learning. In *ACML*, 2020.

Jiongzhai Zheng, Kun He, Jianrong Zhou, Yan Jin, and Chu-Min Li. Combining reinforcement learning with Lin-Kernighan-Helsgaun algorithm for the traveling salesman problem. In *AAAI*, 2021.

Yaoxin Wu, Wen Song, Zhiguang Cao, Jie Zhang, and Andrew Lim. Learning improvement heuristics for solving the travelling salesman problem. *IEEE Transactions on Neural Networks and Learning Systems*, 2021a.

Zhang-Hua Fu, Kai-Bin Qiu, and Hongyuan Zha. Generalize a small pre-trained model to arbitrarily large TSP instances. In *AAAI*, 2021.

Ronald J. Williams. Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Machine Learning*, pages 229–256, 1992.

R.S. Sutton and A.G. Barto. *Reinforcement learning: An introduction*. MIT Press, 1998.
Generalization in Deep RL for TSP Problems via Equivariance and Local Search

Peter W. Battaglia, Jessica B. Hamrick, Victor Bapst, Alvaro Sanchez-Gonzalez, Vinicius Zambaldi, Mateusz Malinowski, Andrea Tacchetti, David Raposo, Adam Santoro, Ryan Faulkner, Caglar Gulcehre, Francis Song, Andrew Ballard, Justin Gilmer, George Dahl, Ashish Vaswani, Kelsey Allen, Charles Nash, Victoria Langston, Chris Dyer, Nicolas Heess, Daan Wierstra, Pushmeet Kohli, Matt Botvinick, Oriol Vinyals, Yujia Li, and Razvan Pascanu. Relational inductive biases, deep learning, and graph networks, 2018.

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

Gerhard Reinelt. TSPLIB—a traveling salesman problem library. *ORSA Journal on Computing*, 3(4):376–384, 1991.

David Applegate, Bixby Ribert, Chvatal Vasek, and Cook William. Concorde tsp solver. 2004. URL http://www.math.uwaterloo.ca/tsp/concorde

Yaoxin Wu, Wen Song, Zhiguang Cao, Jie Zhang, and Andrew Lim. Learning improvement heuristics for solving routing problems. *IEEE Transactions on Neural Networks and Learning Systems*, pages 1–13, 2021b. doi:10.1109/TNNLS.2021.3068828

Elias Khalil, Hanjun Dai, Yuyu Zhang, Bistra Dilkina, and Le Song. Learning combinatorial optimization algorithms over graphs. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017. URL https://proceedings.neurips.cc/paper/2017/file/d9896106ca98d3d05b8cbdf4fd8b13a1-Paper.pdf

Laurent Perron and Vincent Furnon. Or-tools. URL https://developers.google.com/optimization/

Paulo R. de O. da Costa, Jason Rhuggenaath, Yingqian Zhang, and Alp Akcay. Learning 2-opt heuristics for the traveling salesman problem via deep reinforcement learning, 2020.

Christos H. Papadimitriou. The euclidean travelling salesman problem is np-complete. *Theoretical Computer Science*, 4(3):237–244, 1977.
A Combined Local Search

We first provide a brief introduction to TSP heuristics and then give the detailed presentation of each local search method we used, together with our whole combined local search algorithm.

Since TSP is an NP-hard problem [1977], solving exactly large TSP instances is generally impractical. Hence, many heuristics have been proposed to solve TSP. Two important categories of heuristics are constructive heuristics and local search heuristics. On the one hand, constructive heuristics iteratively build a tour from scratch and therefore, do not rely on existing tours. As an example, insertion heuristics are constructive: they first choose a starting city randomly, then repeatedly insert an unvisited city into the partial tour that minimizes the increase of tour length until all the cities are included in the tour. The unvisited city can be selected in different ways leading to various versions of insertion heuristics: random insertion, nearest insertion, or farthest insertion (see Kool et al. [2019] for implementation details). Among them, farthest insertion usually yields the best results. On the other hand, local search heuristics try to improve a given complete tour by perturbing it. They are widely used as post optimization for TSP solvers, such as Deudon et al. [2018], Ma et al. [2019]. One important class of local search heuristic is k-opt, which improves an existing complete tour \( \sigma \) by repeatedly performing the following operation: remove \( k \) edges of the current tour and reconnect the obtained subtours in order to decrease the tour length. For instance, one 2-opt operation would replace:

\[
\sigma = (\sigma(1), \sigma(2), ..., \sigma(i), ..., \sigma(j), ..., \sigma(N))
\]

by

\[
\sigma' = (\sigma(1), ..., \sigma(i), \sigma(j), \sigma(j-1), ..., \sigma(i+1), \sigma(j+1), ..., \sigma(N))
\]

where \( i < j < N \) if \( L_{\sigma_t}(X) < L_{\sigma}(X) \). Based on k-opt, the LKH algorithm [2017] can often achieve nearly optimal solutions, but requires a long runtime.

Based on how many edges are removed and how the subtours are reconnected during one step, every local search heuristic can be seen as a special case of k-opt. Since the number of possible k-opt operations is \( O(N^k) \), 2-opt and 3-opt are usually preferred to efficiently search for fast improvements of existing tours, although they can get stuck in local optima. Hence, in our work, we use a combination of several local search methods, including local insertion heuristics, random 2-opt, search 2-opt, and search random 3-opt (see below for their definitions), to efficiently find potential 2-opt and 3-opt operations to improve the solution proposed by the RL solver. Instead of running each algorithm once for long enough, we run every local search shortly one by one and repeat this process for multiple times so that the combined local search are able to avoid more local optima. Intuitively, the rationale is that once a certain heuristic gets stuck in a local optimum, another can help get it out by trying a different operation. Next, we explain the above four heuristics, and then illustrate how we leverage them to form our combined local search.

A.1 Random 2-opt

For random 2-opt, we randomly try two edges for performing a 2-opt operation and repeat this process for \( \alpha \times N^\beta \) times, where \( \alpha, \beta > 0 \) are two hyperparameters controlling the strength of this heuristic as well as its runtime. Theoretically, random 2-opt can potentially cover all 2-opt operations and such a procedure makes random 2-opt much more flexible than trying all \( N(N-1)/2 \) possible pairs of edges. In this paper, we set \( \alpha = 0.5 \) and \( \beta = 1.5 \) for all experiments.

A.2 Local Insertion Heuristic

For local insertion heuristic, inspired by the insertion heuristic, we iterate through all cities and find the best positions to insert them from their original locations. Let \( \sigma \) be the current tour and \( \sigma_{t,t'} \) be the tour where we exchange the positions of \( \sigma(t) \) with \( \sigma(t') \). Namely,

\[
\sigma_{t,t'} = (\sigma(1), ..., \sigma(t'), \sigma(t), \sigma(t' + 1), ..., \sigma(t-1), \sigma(t+1), ..., \sigma(N)). \tag{11}
\]

for \( t' \neq t-1 \) and \( \sigma_{t,t-1} = \sigma \). The local insertion heuristic (see Algorithm 1) iterates over every \( t \in [N] \). For each \( t \), we need to find \( t^* \) such that:

\[
t^* = \arg \min_{t'} L_{\sigma_{t,t'}}(X), \tag{12}
\]

where the definition of \( L \) is given in Equation (11). Then we replace \( \sigma \) by \( \sigma_{t,t^*} \). Theoretically, the local insertion heuristic is a special case of 3-opt where two of the removed edges cover a same node.
We repeat this process for Algorithm 3 where 

Then we replace \( \sigma \) to \( \sigma^* \). Search random potentially covers all possible 2-opt operations.

### A.3 Search 2-opt

For search 2-opt, we first iterate through all edges and for each first edge, we search for the best second edge to do the 2-opt. Let \( \sigma \) be the current tour and \( \sigma(t, t') \) be the tour where we reverse the cities between \( \sigma(t) \) and \( \sigma(t') \). Namely,

\[
\sigma(t, t') = (\sigma(1), ..., \sigma(t - 1), \sigma(t'), \sigma(t' - 1), ..., \sigma(t + 1), \sigma(t), \sigma(t' + 1), ..., \sigma(N)).
\] (13)

where \( t < t' \), and \( \sigma(t, t) = \sigma \). Search 2-opt (see Algorithm 2) iterates over every \( t \in [N] \). For each \( t \), we find \( t^* \) such that:

\[
t^* = \arg \min_{t' \geq t} L_{\sigma(t, t')}(X)
\] (14)

Then we replace \( \sigma \) by \( \sigma(t, t^*) \). Search 2-opt potentially covers all possible 2-opt operations.

### A.4 Search Random 3-opt

For search random 3-opt, the algorithm first randomly picks two edges. Then for these two randomly-picked edges, similar to search 2-opt, we search to find the best third edge to apply 3-opt. Let \( \sigma \) be the current tour, edges from \( \sigma(t_1) \) to \( \sigma(t_1 + 1) \) and \( \sigma(t_2) \) to \( \sigma(t_2 + 1) \) be the two edges picked, and \( \sigma^*_{(t_1, t_2, t_3)} \) be the optimal tour that differs from \( \sigma \) only in edges from \( \sigma(t_1) \) to \( \sigma(t_1 + 1) \), \( \sigma(t_2) \) to \( \sigma(t_2 + 1) \) and \( \sigma(t_3) \) to \( \sigma(t_3 + 1) \). For randomly picked \( t_1 \) and \( t_2 \), search random 3-opt (see Algorithm 3) iterates over every \( t_3 \in [N] \) and find \( t^*_3 \) such that:

\[
t^*_3 = \arg \min_{t_3} L_{\sigma^*_{(t_1, t_2, t_3)}}(X)
\] (15)

We repeat this process for \( \alpha \times N^3 \) times, where \( \alpha, \beta > 0 \) are the same two hyperparameters as in random 2-opt. We set \( \alpha = 0.5, \beta = 1.5 \) for all experiments. Search random 3-opt potentially covers all possible 3-opt operation.

### Algorithm 3 Search random 3-opt

1. **Input**: A matrix of city coordinates \( X = (x_i)_{i \in [N]} \), current tour \( \sigma \), hyperparameters \( \alpha \) and \( \beta \)
2. **Output**: An improved tour \( \sigma \)
3. **for** iter = 1 to \( \alpha N^3 \) **do**
4. Randomly pick \( t_1, t_2 \in [N] \) such that \( t_1 \neq t_2 \)
5. \( t^*_3 = \arg \min_{t_3} L_{\sigma^*_{(t_1, t_2, t_3)}}(X) \)
6. \( \sigma \leftarrow \sigma^*_{(t_1, t_2, t_3)} \)
7. **end for**
8. **return** \( \sigma \)
A.5 Combined Local Search

With these 4 different local search heuristics, our combined local search applies them one by one sequentially and repeat this for $I$ times (see Algorithm 4), where $I$ is a hyperparameter, which we set to $I = 10$ for all experiments in this paper.

Algorithm 4 Combined Local Search Algorithm

1: **Input**: A matrix of city coordinates $X = (x_i)_{i \in [N]}$, current tour $\sigma$, hyperparameters $\alpha$, $\beta$ and $I$ for local search.
2: **Output**: An improved tour $\sigma$
3: for $t = 1$ to $I$ do
4: $\sigma \leftarrow$ apply local insertion heuristic($X$, $\sigma$)
5: $\sigma \leftarrow$ apply random 2-opt on $\sigma$ for $\alpha N^\beta$ times
6: $\sigma \leftarrow$ apply search 2-opt on $\sigma$
7: $\sigma \leftarrow$ apply search random 3-opt on $\sigma$ for $\alpha N^\beta$ times
8: end for
9: return $\sigma$

B Policy Gradient of eMAGIC

As promised in Section 5 - Smoothed Policy Gradient, we elaborate on the detailed mathematical derivation for Equation 8.

The objective function $J(\theta)$ in Equation 3 can be approximated with the empirical mean of the total rewards using $B$ trajectories sampled with policy $\pi_{\theta}$:

$$J(\theta) \approx \hat{E}_B \left[ \sum_{t=1}^{N} r^{(b)}_t \right] = \frac{1}{|B|} \sum_{b=1}^{B} \sum_{t=1}^{N} r^{(b)}_t,$$

where $\hat{E}_B$ represents the empirical mean operator and $r^{(b)}_t$ is the $t$-th reward of the $b$-th trajectory. The policy gradient to optimize $J(\theta)$ can be estimated by:

$$\nabla_\theta J(\theta) = E_{\tau} \left[ \left( \sum_{t=1}^{N} \nabla_\theta \log \pi_{\theta}(a_t|s_t) \right) \left( \sum_{t=1}^{N} r_t \right) \right]$$

$$\approx \hat{E}_B \left[ \sum_{t=1}^{N} \nabla_\theta \log \pi_{\theta}(a^{(b)}_t|s^{(b)}_t) \left( \sum_{t=1}^{N} r^{(b)}_t \right) \right].$$

However, instead of directly using Equation 17 to update our policy, we interleave the policy gradient with our combined local search, as well as the policy rollout baseline $l^{(b)}$ to update our policy:

$$\nabla_\theta J(\theta) \approx \hat{E}_B \left[ \sum_{t=1}^{N} \nabla_\theta \log \pi_{\theta}(a^{(b)}_t|s^{(b)}_t) \left( L_{\sigma^{(b)}}(X^{(b)} - l^{(b)}) \right) \right],$$

which is the same as Equation 8. Please refer to Section 5 and Appendix A for details regarding our combined local search and the policy rollout baseline $l^{(b)}$.

C Overall Algorithm

We provide the pseudo-code of our overall training algorithm in Algorithm 5.

D Settings and Hyperparameters

All our experiments are run on a computer with an Intel(R) Xeon(R) E5-2678 v3 CPU and a NVIDIA 1080Ti GPU. In consistency with previous work, all our randomly generated TSP instances are sampled in $[0, 1]^2$ uniformly. During training, stochastic CL chooses a TSP size in $\mathcal{R} = \{10, 11, \ldots, 50\}$ for each epoch $e$ according to Equation 9 and 10.
Algorithm 5 REINFORCE exploiting stochastic CL, equivariance, and smoothed policy gradient

1: **Input**: Total number of epochs \( E \), training steps per epoch \( T \), batch size \( B \), hyperparameters \( \alpha, \beta, \gamma \) and \( I \) for local search
2: Initialize \( \theta \)
3: for \( e = 1 \) to \( E \) do
4: \( N \leftarrow \) Sample from \( p(e) \) according to Stochastic CL
5: for \( t = 1 \) to \( T \) do
6: \( \forall b \in \{1, ..., B\} \; X(b) \leftarrow \) Random TSP instance with \( N \) cities
7: \( \forall b \in \{1, ..., B\} \; \sigma(b) \leftarrow \) Apply \( \pi(\theta) \) on \( X(b) \) after all the equivariant preprocessing steps
8: \( \forall b \in \{1, ..., B\} \; \sigma_+^{(b)} \leftarrow \) Apply the combined local search on \( \sigma(b) \)
9: Use \( \sigma(b) \) and \( \sigma_+^{(b)} \) to calculate \( \nabla_\theta J_+^{(\theta)} \) according to Equation 8
10: \( \theta \leftarrow \) Update in the direction of \( \nabla_\theta J_+^{(\theta)} \)
11: end for
12: end for

with \( \sigma_N \) set to be 3. We trained for 200 epochs in each experiment, with 1000 batches of 128 random TSP instances in each epoch. We set the learning rate to be \( 10^{-3} \) and the learning rate decay to be 0.96 in each experiment. In each experiment, we set \( \alpha = 0.5 \), \( \beta = 1.5 \), \( \gamma = 0.25 \) and \( I = 10 \) for the parameters of our combined local search. As for random TSP testing and the ablation study, we test on TSP instances with size 20, 50, 100, 200, 500 and 1000 to evaluate the generalization capability of our model; as for realistic TSP, we test on TSP instances with sizes up to 1002 from the TSPLIB library. With respect to our model architecture, our MLP encoder has an input layer with dimension 2, two hidden layers with dimension 128 and 256, respectively, and an output layers with dimension 128; for the GNN encoder, we set \( H = 128 \) and \( n_{GNN} = 3 \).

E Evaluation of Equivariant Preprocessing Methods

E.1 Comparisons between different symmetries used during preprocessing

We present the comparison results of applying rotation, translation and reflection. The comparisons are done by testing on random TSP cases followed the same setting in the experiment section. As in table E.1, the rotation has the best performance on small and large TSP instances. By this, we choose rotation for our final algorithm.

| TSP Size | Rotation Len. | Rotation Gap | Reflection Len. | Reflection Gap | Translation Len. | Translation Gap |
|----------|---------------|--------------|-----------------|--------------|-----------------|----------------|
| 20       | 3.844         | 0.37%        | 3.850           | 0.51%        | 3.848           | 0.47%          |
| 50       | 5.763         | 1.27%        | 5.786           | 1.67%        | 5.807           | 2.05%          |
| 100      | 7.964         | 2.61%        | 8.050           | 3.72%        | 8.020           | 3.33%          |
| 200      | 11.14         | 3.89%        | 11.22           | 4.71%        | 11.19           | 4.39%          |
| 500      | 17.54         | 6.01%        | 17.65           | 6.69%        | 17.68           | 6.85%          |
| 1000     | 24.75         | 7.05%        | 24.81           | 7.33%        | 24.86           | 7.55%          |

E.2 Comparisons between one preprocessing application and iteration preprocessing applications

We present the comparison results of one preprocessing application and iteration preprocessing applications. The comparisons are done by testing on random TSP cases followed the same setting in the experiment section. As in table E.2, the iteration preprocessing applications has better performance on small and large TSP instances. By this, we choose iteration preprocessing applications for our final algorithm.

F More Ablation Studies

This section is the continuation of Section 6- Ablation Study. In this section, we first perform an ablation study of stochastic CL, meaning we fixed our TSP size to be 50 during training. Moreover, we perform an ablation study on each
Table 6: Comparisons between iteration preprocessing and one preprocessing

| TSP Size | Iteration† | One‡ |
|----------|------------|------|
|          | Len.       | Gap  | Len.       | Gap  |
| 20       | 3.844      | 0.37%| 3.873      | 1.12%|
| 50       | 5.763      | 1.27%| 5.871      | 3.17%|
| 100      | 7.964      | 2.61%| 8.102      | 4.40%|
| 200      | 11.136     | 3.89%| 11.347     | 5.85%|
| 500      | 17.541     | 6.01%| 17.742     | 7.23%|
| 1000     | 24.749     | 7.05%| 24.978     | 8.05%|

† Iteration preprocessing application.
‡ Once preprocessing application.

component of our equivalent model, which includes deleting the visited cities during training, equivariant preprocessing operation, and using relative positions during training, as mentioned in Section 6 - Ablation Study. For the ablation study of these three components, we remove them during training and testing. Table 7 illustrates the results of our ablation study (including the full version for comparison), which demonstrate that each component is effective in our model.

Table 7: Ablation study on CL, Deleting, Preprocessing and Relative Position, tested on 10,000 instances for TSP 20, 50 and 100, and 128 instances for TSP 200, 500 and 1000.

| TSP Size | Full       | w/o CL      | w/o Delete | w/o Pre† | w/o RP‡ |
|----------|------------|-------------|------------|----------|---------|
|          | Len.       | Gap         | Len.       | Gap      | Len.    |
| TSP20    | 3.844      | 0.37%       | 3.861      | 0.79%    | 3.855   |
| TSP50    | 5.763      | 1.27%       | 5.821      | 2.30%    | 5.824   |
| TSP100   | 7.964      | 2.61%       | 8.062      | 3.88%    | 8.064   |
| TSP200   | 11.145     | 3.89%       | 11.295     | 5.29%    | 11.28   |
| TSP500   | 17.54      | 6.01%       | 17.669     | 6.93%    | 17.67   |
| TSP1000  | 24.75      | 7.05%       | 24.889     | 7.68%    | 24.888  |

† Pre - Preprocessing; RP - Relative Position.

G  GNN Encoder Details

Figure 2 is a detailed illustration of the GNN architecture introduced in Section 4.2. Please notice that Figure 2 could be regarded as a zoom-in version of the GNN part in Figure 1. The aggregation function used in GNN is represented by a neutral network followed by a ReLU function on each entry of the output.

![Figure 2: Detailed Architecture of GNN](image)

H  Realist TSP Instances in TSPLIB

As promised in Section 6, Table 9 and 10 list the performances of various learning-based TSP solvers and heuristic approaches upon realistic TSP instances in TSPLIB. The bold numbers show the best performance among all the approaches. We can observe that most bold numbers are provided by eMAGIC, meaning our approach provides...
excellent results for TSPLIB. In addition, we also provide our results for the larger instances (with size up to 1002) from TSPLIB in Table 8. Table 8 only contains the experiments of our model since no other model can generalize to large size TSP instances with adequate performance.

Table 8: Performance of eMAGIC on large size TSP instances in TSPLIB

| Problems | OPT  | Length | Gap   |
|----------|------|--------|-------|
| rd400    | 15.281 | 15.707 | 2.79% |
| f1417    | 11.861 | 11.961 | 0.84% |
| pr439    | 107.217 | 109.610 | 2.23% |
| pbe442   | 50.778 | 51.868 | 2.15% |
| d493     | 35.002 | 36.184 | 3.38% |
| u574     | 36.905 | 38.486 | 4.28% |
| rat575   | 6.773  | 7.105  | 4.90% |
| p654     | 34.643 | 35.001 | 1.03% |
| d657     | 48.912 | 50.826 | 3.91% |
| u724     | 41.910 | 43.853 | 4.64% |
| rat783   | 8.806  | 9.324  | 5.88% |
| pr1002   | 259.045 | 271.370 | 4.76% |
Table 9: Comparison of Performances on TSPLIB - Part I

| Problems   | OPT | eMAGIC(S) | Wu et al. | S2V-DQN | OR-Tools |
|------------|-----|-----------|-----------|---------|----------|
|            | Len. | Gap | Len. | Gap | Len. | Gap | Len. | Gap | Len. | Gap |
| ei151      | 426  | 429    | 0.70% | 438  | 2.82% | 439  | 3.05% | 436  | 2.35% |
| berlin52   | 7,542| 7,544   | 0.03% | 8,020| 6.34% | 7,542| 0.00% | 7,945| 5.34% |
| st70       | 675  | 677     | 0.30% | 706  | 4.50% | 696  | 3.11% | 683  | 1.19% |
| ei176      | 538  | 546     | 1.49% | 575  | 6.88% | 564  | 4.83% | 561  | 4.28% |
| pr76       | 108,159| 108,159| 0.00% | 109,668| 1.40% | 108,446| 0.27% | 111,104| 2.72% |
| rat99      | 1.211| 1,223   | 0.99% | 1,419| 17.18%| 1,280| 5.70% | 1,232| 1.73% |
| kroA100    | 21,282 | 21,285  | 0.01% | 25,196| 18.39%| 21,897| 2.89% | 21,448| 0.78% |
| kroB100    | 22,141 | 22,141  | 0.00% | 26,563| 19.97%| 22,692| 2.49% | 23,006| 3.91% |
| kroC100    | 20,749 | 20,749  | 0.00% | 25,343| 22.14%| 21,074| 1.57% | 21,583| 4.02% |
| kroD100    | 21,294 | 21,361  | 0.31% | 24,771| 16.33%| 22,102| 3.79% | 21,636| 1.61% |
| kroE100    | 22,068 | 22,068  | 0.00% | 26,903| 21.91%| 22,913| 3.83% | 22,598| 2.40% |
| rd100      | 7,910 | 7,916   | 0.08% | 7,915| 0.06% | 8,159| 3.15% | 8,189| 3.53% |
| el1101     | 629  | 636     | 1.11% | 658  | 4.61% | 659  | 4.77% | 664  | 5.56% |
| lin105     | 14,379| 14,379  | 0.00% | 18,194| 26.53%| 15,023| 4.48% | 14,824| 3.09% |
| pr107      | 44,303| 44,346  | 0.10% | 53,056| 19.76%| 45,113| 1.83% | 46,072| 3.99% |
| pr124      | 59,030| 59,075  | 0.08% | 66,010| 11.82%| 61,623| 4.39% | 62,519| 5.91% |
| bier127    | 118,282| 119,151 | 0.73% | 142,707| 20.65%| 121,576| 2.78% | 122,733| 3.76% |
| ch130      | 6,110 | 6,162   | 0.85% | 7,120| 16.53%| 6,270| 2.62% | 6,284| 2.85% |
| pr136      | 96,772| 97,264  | 0.51% | 105,618| 9.14% | 99,474| 2.79% | 102,213| 5.62% |
| pr144      | 58,537| 58,537  | 0.00% | 71,006| 21.30%| 59,436| 1.54% | 59,286| 1.28% |
| ch150      | 6,528 | 6,592   | 0.98% | 7,916| 21.26%| 6,985| 7.00% | 6,729| 3.08% |
| kroA150    | 26,524| 26,727  | 0.77% | 31,244| 17.80%| 27,888| 5.14% | 27,592| 4.03% |
| kroB150    | 26,130| 26,282  | 0.58% | 31,407| 20.20%| 27,209| 4.13% | 27,572| 5.52% |
| pr152      | 73,682| 73,682  | 0.00% | 85,616| 16.20%| 75,283| 2.17% | 75,834| 2.92% |
| ui159      | 42,080| 42,080  | 0.00% | 51,327| 21.97%| 45,433| 7.97% | 45,778| 8.79% |
| rat195     | 2,323 | 2,377   | 2.32% | 2,913| 25.40%| 2,581| 11.11%| 2,389| 2.84% |
| d198       | 15,780| 15,874  | 0.60% | 17,962| 13.83%| 16,453| 4.26% | 15,963| 1.16% |
| kroA200    | 29,368| 29,840  | 1.61% | 35,958| 22.44%| 30,965| 5.44% | 29,741| 1.27% |
| kroB200    | 29,437| 29,743  | 1.04% | 36,412| 23.69%| 31,692| 7.66% | 30,516| 3.67% |
| ts225      | 126,643| 126,939 | 0.23% | 158,748| 25.35%| 136,302| 7.63% | 128,564| 1.52% |
| tsp225     | 3,916 | 3,981   | 1.66% | 4,701| 20.05%| 4,154| 6.08% | 4,046| 3.92% |
| pr226      | 80,369| 80,436  | 0.08% | 97,348| 21.13%| 81,873| 1.87% | 82,968| 3.23% |
| gi1262     | 2,378 | 2,417   | 1.64% | 2,963| 24.60%| 2,537| 6.69% | 2,519| 5.93% |
| pr264      | 49,135| 49,908  | 1.57% | 65,946| 34.21%| 52,364| 6.57% | 51,954| 5.74% |
| as280      | 2,579 | 2,635   | 2.17% | 2,989| 15.90%| 2,867| 11.17%| 2,713| 5.20% |
| pr299      | 48,191| 48,905  | 1.48% | 59,786| 24.06%| 51,895| 7.69% | 49,447| 2.61% |
| lin318     | 42,029| 42,948  | 2.19% | –    | –    | 45,375| 7.96% | –    | –    |

1 Wu et al. [2021b], 2 Khalil et al. [2017], 3 Perron and Furnon
| Problems | OPT | L2OPT<sup>1</sup> | AM<sup>2</sup> | Furthest<sup>3</sup> | 2-opt |
|----------|-----|-----------------|-------------|----------------|-------|
|          | Len. | Gap | Len. | Gap | Len. | Gap | Len. | Gap | Len. | Gap |
| ei151    | 426  | 427  | 0.23% | 435 | 2.11% | 467 | 9.62% | 446  | 4.69% |
| berlin52 | 7,542 | 7,974 | 5.73% | 7,668 | 1.67% | 8,307 | 10.14% | 7,788 | 3.26% |
| st70     | 675  | 680  | 0.74% | 690 | 2.22% | 712 | 5.48% | 753  | 11.56% |
| ei176    | 538  | 552  | 2.60% | 563 | 4.64% | 583 | 8.36% | 591  | 9.85% |
| pr76     | 108,159 | 111,085 | 2.71% | 111,250 | 2.85% | 119,692 | 10.66% | 115,460 | 6.75% |
| rat99    | 1,211 | 1,388 | 14.62% | 1,319 | 8.91% | 1,314 | 8.51% | 1,390 | 14.78% |
| kroA100  | 21,282 | 23,751 | 11.60% | 38,200 | 79.49% | 23,356 | 9.75% | 22,876 | 7.49% |
| kroB100  | 22,141 | 23,790 | 7.45% | 35,511 | 60.38% | 23,222 | 4.88% | 23,496 | 6.12% |
| kroC100  | 20,749 | 22,672 | 9.27% | 30,642 | 47.67% | 21,699 | 4.58% | 23,445 | 12.99% |
| kroD100  | 21,294 | 23,334 | 9.58% | 32,211 | 51.60% | 22,034 | 3.48% | 23,967 | 12.55% |
| kroE100  | 22,068 | 23,253 | 5.37% | 27,164 | 23.09% | 23,516 | 6.56% | 22,800 | 3.32% |
| rd100    | 7,910 | 7,944 | 0.43% | 8,152 | 3.05% | 8,944 | 13.07% | 8,757 | 10.71% |
| ei1101   | 629  | 635  | 0.95% | 667 | 6.04% | 673 | 7.00% | 702  | 11.61% |
| l1in105  | 14,379 | 16,156 | 12.36% | 51,325 | 256.94% | 15,193 | 5.66% | 15,536 | 8.05% |
| pr107    | 44,303 | 54,378 | 22.74% | 205,519 | 363.89% | 45,905 | 3.62% | 47,058 | 6.22% |
| pr124    | 59,030 | 59,516 | 0.82% | 167,494 | 183.74% | 65,945 | 11.71% | 64,765 | 9.72% |
| bier127  | 118,282 | 121,122 | 2.40% | 207,600 | 75.51% | 129,495 | 9.48% | 128,103 | 8.30% |
| ch130    | 6,110 | 6,175 | 1.06% | 6,316 | 3.37% | 6,498 | 6.35% | 6,470 | 5.89% |
| pr136    | 96,772 | 98,453 | 1.74% | 102,877 | 6.36% | 105,361 | 8.88% | 110,531 | 14.22% |
| pr144    | 58,537 | 61,207 | 4.56% | 183,583 | 213.61% | 61,974 | 5.87% | 60,321 | 3.05% |
| ch150    | 6,528 | 6,597 | 1.06% | 6,877 | 5.34% | 7,210 | 10.45% | 7,232 | 10.78% |
| kroA150  | 26,524 | 30,078 | 13.40% | 42,335 | 59.61% | 28,658 | 8.05% | 29,666 | 11.85% |
| kroB150  | 26,130 | 28,169 | 7.80% | 35,511 | 60.38% | 27,404 | 4.88% | 29,517 | 12.96% |
| pr152    | 73,682 | 75,301 | 2.20% | 103,110 | 39.93% | 75,396 | 2.33% | 77,206 | 4.78% |
| u159    | 42,080 | 42,716 | 1.51% | 115,372 | 174.17% | 46,789 | 11.19% | 47,664 | 13.27% |
| rat195   | 2,323 | 2,955 | 27.21% | 3,661 | 57.59% | 2,609 | 12.31% | 2,605 | 12.14% |
| d198    | 15,780 | - | - | 68,104 | 331.57% | 16,138 | 2.27% | 16,596 | 5.17% |
| kroA200  | 29,368 | 32,522 | 10.74% | 58,643 | 99.68% | 31,949 | 8.79% | 32,760 | 11.55% |
| kroB200  | 29,437 | - | - | 50,867 | 72.79% | 31,522 | 7.08% | 35,107 | 12.47% |
| ts225    | 126,643 | 127,731 | 0.86% | 141,628 | 11.83% | 140,626 | 11.04% | 138,101 | 9.05% |
| tsp225   | 3,916 | 4,354 | 11.18% | 24,816 | 533.70% | 4,280 | 9.30% | 4,278 | 9.24% |
| pr226    | 80,369 | 91,560 | 13.92% | 101,992 | 26.90% | 84,130 | 4.68% | 89,262 | 11.07% |
| gi1262   | 2,378 | 2,490 | 4.71% | 2,683 | 13.24% | 2,623 | 10.30% | 2,597 | 9.21% |
| pr264    | 49,135 | 59,109 | 20.30% | 338,506 | 588.93% | 54,462 | 10.84% | 54,547 | 11.01% |
| a280     | 2,579 | 2,898 | 12.37% | 11,810 | 357.92% | 3,001 | 16.36% | 2,914 | 12.99% |
| pr299    | 48,191 | 59,422 | 23.31% | 513,673 | 938.83% | 51,903 | 7.70% | 54,914 | 13.95% |
| lin318   | 42,029 | - | - | - | - | 45,918 | 9.25% | 45,263 | 7.69% |

<sup>1</sup> [de O. da Costa et al., 2020], <sup>2</sup> Kool et al., 2019. <sup>3</sup> [Furthest Insertion Heuristic]