

# Robot Motion Planning in Learned Latent Spaces

Brian Ichter, Marco Pavone

Stanford University, Stanford CA 94305, USA  
`{ichter,pavone}@stanford.edu`

**Abstract.** This paper presents Latent Sampling-based Motion Planning (L-SBMP), a methodology towards computing motion plans for complex robotic systems by learning a plannable latent representation. Recent works in control of robotic systems have effectively leveraged local, low-dimensional embeddings of high-dimensional dynamics. In this paper we combine these recent advances with techniques from sampling-based motion planning (SBMP) in order to design a motion planning framework for high-dimensional robotic systems (e.g., humanoids, or even systems where planning occurs in the visual space). Specifically, the learned latent space is constructed through an autoencoding network, a dynamics network, and a collision checking network, which mirror the three main algorithmic primitives of SBMP, namely state sampling, local steering, and collision checking. Notably, these networks can be trained through only raw data of the system’s states and actions along with a supervising collision checker. Building upon these networks, an RRT-based algorithm is used to plan motions *directly in the latent space* – we refer to this exploration algorithm as Learned Latent RRT (L2RRT). This algorithm globally explores the latent space and is capable of generalizing to new environments. The overall methodology is demonstrated on two planning problems that are well beyond the reach of standard SBMP, namely a visual planning problem, whereby planning happens in the visual (pixel) space, and a humanoid robot planning problem.

**Keywords:** motion and path planning, machine learning, learned latent space, plannable spaces, sampling-based motion planning

## 1 Introduction

Motion planning is a fundamental problem in robotics whereby one seeks to compute a dynamically-feasible trajectory connecting an initial state to a goal region while avoiding obstacles. This problem quickly increases in difficulty with the dimensionality of the system and the complexity of the dynamics [1]. For some high-dimensional systems, and even differentially-constrained systems, sampling-based motion planning (SBMP) techniques have emerged as a particularly successful approach. SBMP avoids the explicit construction of the state space and instead maintains an implicit representation of it through three algorithmic primitives: sampling the state space, steering (or connecting) to nearby sampled states, and collision checking such connections [1, 2]. This implicit representation is global and increasingly accurate as samples are added [1]. However, applying

SBMP to robotic systems with, say, more than ten dimensions, quickly becomes intractable. This precludes its application (at least directly) to cases where, for example, plans are computed from visual inputs alone [3, 4] or the dynamics are very high-dimensional and complex, as for humanoids [5].

A promising solution to address this challenge is represented by recent works on learning-based control, which leverage learned, low-dimensional manifolds embedded within the state space to which the dynamics of a robotic system are mainly restricted to [3, 6–8]. For example, a bipedal humanoid walking robot is generally in gait and upright, and a visual representation of a robot has a “true” underlying lower-dimensional dynamical representation. These works, however, primarily focus only in local regions, and can not be directly applied to compute motion plans whereby one needs to *globally* explore the state space. Our key insight to scale these techniques to robot motion planning is to leverage the local effectiveness of low-dimensional latent representations to learn a latent space in which one can directly use SBMP. In other words, the idea is to build a *global* implicit representation of the latent space through only local connections. The latent space is constructed through an autoencoding network, a dynamics network, and a collision checking network, which mirror the three aforementioned algorithmic primitives of SBMP, namely state sampling, local steering, and collision checking. Specifically, the space is learned through an autoencoding network with learned latent dynamics, which provides the ability to sample the latent space and connect local states through dynamically-feasible trajectories. Notably, this network only requires raw data of the system’s states and control inputs to be trained. Separately, a collision checking network is trained to predict whether a trajectory between two latent states is collision free given a specific problem environment. Collectively, these networks provide a methodology to globally search the latent space (in our case, through an RRT-based algorithm), in a way that generalizes well to new environments and is general enough to apply to a broad class of problems, including visual planning and complex dynamics problems. The final trajectory can then be projected to the full state space through the autoencoding network and used in conjunction with the aforementioned local control methods to effectively control the robotic system along the planned trajectory. We refer to this methodology as Latent Sampling-based Motion Planning (L-SBMP).

*Related Work:* The field of representation learning seeks to learn representations of data to extract useful information for classifiers or other predictors [9–11]. This often takes the form of manifold learning, i.e., learning lower-dimensional manifolds embedded within high-dimensional data. In recent years representation learning has been applied along the entire robotic stack – from perception [12], to decision making and planning [13, 14], to control and dynamics [3, 6, 7, 15]. Recently, these methods have seen particular success in the field of control. Embed to Control [3] and Robust Locally-Linear Controllable Embeddings [6] present approaches to learning to control robotic systems directly from raw images. Both methods learn a locally-linear embedding via a variational autoencoder and directly perform stochastic optimal control on the latent space. [4] learns to predict the results of actions in visual spaces and uses model-predictive control to plan local manipulator actions. In contrast, our work learns

a *non-linear* latent space (allowing generalization to more complex systems) and a collision checking network (allowing generalization to new environments).

In the context of robot motion planning, Universal Planning Networks (UPN) [8] learn a plannable latent representation in which gradient descent can be used to compute a plan that minimizes an imitation loss (learned from an expert planner). Our methodology, L-SBMP, learns a similar latent space to UPN (by enforcing dynamics in the latent space), but performs a *global* search of the latent space to find motion plans. In addition, L-SBMP does not rely on an expert planner, which for many problems may not be available. [16] uses a Gaussian process latent variable model to learn a low-dimensional representation and then uses inference to solve the planning problem via a particle-based approach. Compared to L-SBMP, the latent variable model must be well initialized and using inference does not allow a global search of the latent space. Motion Planning Networks (MPNet) [17] learns a low-dimensional representation of the obstacle space and uses a feedforward neural network to predict the next step an optimal planner would take given an initial state, goal state, and the obstacle representation (trained via an asymptotically optimal planner). MPNet approaches the problem of accelerating solutions to planning problems for which solutions to similar problems can be provided by demonstration, whereas our work attempts to solve previously intractable planning problems. Lastly, within sampling-based planning, learning has been used to replace parts of the algorithmic stack, such as the collision checker [18] or the sample distribution [19]. L-SBMP too learns a collision checking network, though directly in the latent space, and, in a manner, learns where to sample by choosing samples for the exploration algorithm directly from encoded training states.

It is important to note that, in practice, the L-SBMP methodology can be, and should be, combined with many of the aforementioned approaches. In many cases, L-SBMP may be an initialization for the latent spaces in [3, 6, 8]. Furthermore, these approaches can be used to improve the trajectory generated by L-SBMP and to control the motion along it.

*Statement of Contributions:* The contribution of this paper is threefold. First, we present the Latent Sampling-based Motion Planning (L-SBMP) methodology, which comprises an autoencoding network, a dynamics network, and a collision checking network to enable planning through the latent space. These networks can be trained through only raw data of the system’s states and control inputs along with a supervising collision checker. Second, building upon L-SBMP, we design an RRT-based algorithm, termed Learned Latent Rapidly-exploring Random Trees (L2RRT), to plan motions *directly in the latent space*. L2RRT efficiently and globally explores the learned manifold by maintaining an implicit representation of the space formed through sampling encoded states and dynamically propagating nearby trajectories in the tree (under the supervision of the collision checking network). The final latent trajectory is then projected back to the full state space through the autoencoding network. Third, we demonstrate our approach on two planning problems that are well beyond the reach of standard SBMP, namely a visual planning problem, whereby planning happens in the visual (pixel) space, and a humanoid robot planning problem.

*Organization:* The remainder of the paper is organized as follows. Section 2 introduces the problem we are approaching and provides necessary background.

Section 3 outlines the L-SBMP methodology for learning the plannable latent space. Section 4 presents L2RRT, the sampling-based exploration strategy used to globally search the latent space. Section 5 demonstrates the performance of L-SBMP on a visual planning problem and a humanoid planning problem. Section 6 summarizes our findings and suggests several avenues for future work.

## 2 Problem Statement

In this section we formulate the motion planning problem we wish to solve. We then briefly overview sampling-based motion planning algorithms in order to give necessary background and to motivate our methodology. Lastly, we present the problem of learning plannable latent spaces.

### 2.1 Robot Motion Planning Problem

Let  $\mathcal{X} \subseteq \mathbb{R}^{d_x}$  and  $\mathcal{U} \subseteq \mathbb{R}^{d_u}$  denote the state space and control input space, respectively, of a robotic system. Let the dynamics of the robot be defined by

$$\dot{\mathbf{x}}(\tau) = f_{\mathcal{X}}^{(c)}(\mathbf{x}(\tau), \mathbf{u}(\tau)), \quad (1)$$

where  $\mathbf{x}(\tau) \in \mathcal{X}$  and  $\mathbf{u}(\tau) \in \mathcal{U}$  denote the state and control input of the system at time  $\tau$  and  $f_{\mathcal{X}}^{(c)}$  denotes the continuous-time system dynamics. To make the dynamics learning problem tractable we consider a discrete-time formulation. Note however that we still consider the continuous-time underlying dynamics to determine collisions. Let the discrete-time dynamics of the robot be defined by

$$\mathbf{x}_{t+1} = f_{\mathcal{X}}(\mathbf{x}_t, \mathbf{u}_t), \quad (2)$$

where  $\mathbf{x}_t \in \mathcal{X}$  and  $\mathbf{u}_t \in \mathcal{U}$  denote the state and control input of the system at time step  $t$  and  $f_{\mathcal{X}}$  denotes the discrete-time system dynamics.

Let  $\mathcal{X}_{\text{free}} \subseteq \mathcal{X}$  define the free state space of the robot and let  $\mathcal{X}_{\text{obs}} \subset \mathcal{X}$  define the obstacle space, such that  $\mathcal{X}_{\text{free}} = \mathcal{X} \setminus \mathcal{X}_{\text{obs}}$ . Let us be given an initial state  $\mathbf{x}_{\text{init}} \in \mathcal{X}_{\text{free}}$  and a goal region  $\mathcal{X}_{\text{goal}} \subset \mathcal{X}_{\text{free}}$ . A trajectory  $\pi$  for the robot is defined as a series of states and control inputs  $(\mathbf{x}_0, \mathbf{u}_0, \dots, \mathbf{x}_T, \mathbf{u}_T)$ , where  $T$  is the number of time steps in the trajectory. Further, let  $\overline{\mathbf{x}_0, \dots, \mathbf{x}_T}$  denote the continuous curve traced by the robot's trajectory (connecting  $\mathbf{x}_0, \dots, \mathbf{x}_T$ ). We make this distinction here to clarify that collision checking should be done along the continuous trajectory connecting discrete states. A trajectory is said to be *collision-free* if  $\overline{\mathbf{x}_0, \dots, \mathbf{x}_T} \cap \mathcal{X}_{\text{obs}} = \emptyset$ . A trajectory is said to be *feasible* if it connects the initial state and goal region, i.e.,  $\mathbf{x}_0 = \mathbf{x}_{\text{init}}$  and  $\mathbf{x}_T \in \mathcal{X}_{\text{goal}}$ , is collision-free, and is dynamically feasible, i.e.,  $\mathbf{x}_t = f_{\mathcal{X}}(\mathbf{x}_{t-1}, \mathbf{u}_{t-1})$  for  $t = 1, \dots, T$ . In this work, we are interested in the following problem,

*Problem 1 (Motion Planning Problem).* Given a motion planning problem  $(\mathbf{x}_{\text{init}}, \mathcal{X}_{\text{goal}}, \mathcal{X}_{\text{free}})$ , find a feasible trajectory  $\pi$ . If no such trajectory exists, report failure.

## 2.2 Sampling-based Motion Planning

Sampling-based motion planning has emerged as a highly effective algorithmic paradigm for solving complex motion planning problems. SBMP avoids the explicit construction of the state space and instead maintain an implicit, searchable representation. This representation is formed by a set of probing samples which are locally connected and verified via a “black box” collision checking module [1]. This methodology enables a global search and guarantees on probabilistic completeness and often asymptotic optimality [20, 21].

One particularly successful family of sampling-based motion planning algorithms is the rapidly-exploring random tree (RRT) [2]. As our exploration algorithm is built from this, we now overview RRT to clarify each algorithmic primitive. The algorithm begins by initializing a tree  $\mathcal{T}$  with the initial state  $\mathbf{x}_{\text{init}}$ , and subsequently iteratively builds a tree through the state space. At each iteration, a state in the free state space is sampled,  $\mathbf{x}_{\text{sample}} \in \mathcal{X}_{\text{free}}$ . The nearest node within the tree to  $\mathbf{x}_{\text{sample}}$  is selected as the expansion node,  $\mathbf{x}_{\text{expand}}$ . The node is then expanded towards  $\mathbf{x}_{\text{sample}}$ , forming a new sample  $\mathbf{x}_{\text{new}}$ . If the trajectory connecting  $\mathbf{x}_{\text{expand}}$  and  $\mathbf{x}_{\text{new}}$  does not intersect  $\mathcal{X}_{\text{obs}}$ , i.e., is collision-free, then  $\mathbf{x}_{\text{new}}$  is added as a node to the tree. This continues until either time elapses or a pre-determined number of samples is reached. The approach is capable of globally searching the state space only through three algorithmic primitives: (1) sampling, (2) locally connecting, and (3) collision checking. The key idea in this work is that, by ensuring each of these algorithmic primitives can be applied in a learned, low-dimensional latent space representing the robotic system, we can directly use SBMP to compute plans for the full state space.

## 2.3 Plannable Latent Spaces

The robotic systems we consider in this work may have complex-dynamics and high-dimensional state spaces. For example, these may be pixel representations of the robotic system and environment or they may be highly-dynamic, high-dimensional robots. In these cases standard approaches to solving motion planning problems are computationally intractable. Even approaches like SBMPs, which were initially created for high-dimensional systems, have difficulty scaling well beyond ten dimensions even in the absence of dynamic constraints.

Our goal is to find the underlying manifold of the problem by learning a low-dimensional latent space model in which SBMP can be performed. Due to the complexity of the system, this space should be learned only through sequences of states and control inputs of the robot in operation and a separate local collision checker. Let  $\mathcal{Z} \subseteq \mathbb{R}^{d_z}$  denote the learned latent space and  $\mathbf{z}_t \in \mathcal{Z}$  a latent state at time step  $t$ . We seek to learn a mapping  $m$  from  $\mathcal{X}$  to  $\mathcal{Z}$  and a mapping  $n$  from  $\mathcal{Z}$  to  $\mathcal{X}$  for which latent dynamics  $f_{\mathcal{Z}}$  are enforced over  $\mathcal{Z}$ . We further learn a “black-box” collision checking function  $g$  in the latent space. We thus wish to learn the following:

$$\begin{aligned} m : \mathcal{X} &\rightarrow \mathcal{Z}, & n : \mathcal{Z} &\rightarrow \mathcal{X}, & \mathbf{z}_{t+1} &= f_{\mathcal{Z}}(\mathbf{z}_t, \mathbf{u}_t), \\ g(\mathbf{z}_t, \mathbf{z}_{t+1}, \mathcal{X}_{\text{obs}}) &= \{1 \text{ if } \overline{\mathbf{x}_t, \mathbf{x}_{t+1}} \cap \mathcal{X}_{\text{obs}} = \emptyset, 0 \text{ otherwise}\}. \end{aligned} \tag{3}$$

### 3 Learning a Latent Space for Robotic Motion Planning

In order to plan on a low-dimensional latent space model, one needs: (1) that the latent space can be sampled, (2) that the dynamics are known to connect nearby latent states, and (3) that a local, latent trajectory can be classified as collision-free. It is also desirable that the latent samples can be mapped to the full space once a solution is found, so that the robotic system can execute the motion plan. Finally, we do not assume we have access to the true system dynamics  $f_{\mathcal{X}}$ , i.e., only observations of states and actions over time and a collision check.

The Latent Sampling-based Motion Planning (L-SBMP) architecture consists of three networks: (1) an autoencoder, (2) a latent local dynamics model, and (3) a collision classifier. The autoencoder projects the high-dimensional full state ( $\mathbf{x}_t$ ) into a lower-dimensional latent state ( $\mathbf{z}_t$ ), and also allows decoding to form inputs for a subsequent controller. The dynamics model enforces local dynamics on the encoded latent space and allows the dynamics to be propagated. Finally, the collision classifier identifies if the connection between two nearby latent states will be possible given the problem environment. The latent space is trained from a series of trajectories from the robotic system in operation; i.e., a sequence of states  $\mathbf{x}_t$  and actions  $\mathbf{u}_t$ . The collision checking network is trained from a data set of nearby states and a label denoting if the connecting trajectory is in collision.

#### 3.1 Learned Latent Space and Dynamics

The first network in our architecture follows a standard treatment for autoencoders. Given a robotic system with state  $\mathbf{x}_t$  and latent state  $\mathbf{z}_t$ , we define an encoding network  $h_{\phi}^{\text{enc}}(\mathbf{x}_t)$  and a decoding network  $h_{\theta}^{\text{dec}}(\mathbf{z}_t)$ , with parameters  $\phi$  and  $\theta$ , respectively. These networks project the full state into the lower-dimensional latent space and then reconstruct the full state. Depending on the robotic system, such networks may be convolutional or fully connected networks. For this work we use an  $\ell_2$  reconstruction loss.

The second network in our architecture learns, and subsequently enforces, local dynamics in the latent space. Let  $h_{\psi}^{\text{dyn}}(\mathbf{z}_t, \mathbf{u}_t)$  denote the dynamics network and  $\psi$  its parameters. Such network predicts the next state in the latent space,  $\hat{\mathbf{z}}_{t+1}$ . The accuracy of this prediction is enforced through two losses. First, the prediction is decoded  $h_{\theta}^{\text{dec}}(\hat{\mathbf{z}}_{t+1})$ , and compared to the full state  $\mathbf{x}_{t+1}$  with an  $\ell_2$  reconstruction loss. Second, the prediction is compared in the latent space to the next encoded state  $\mathbf{z}_{t+1} = h_{\phi}^{\text{enc}}(\mathbf{x}_{t+1})$ . In order to have a dynamically-consistent loss, we linearize the dynamics around  $\mathbf{z}_t$  and  $\mathbf{u}_t$  and compute the weighted controllability Gramian  $G_t$  (also known as the reachability Gramian) as

$$A_t = \frac{\partial h_{\psi}^{\text{dyn}}}{\partial \mathbf{x}}(\mathbf{z}_t, \mathbf{u}_t), \quad B_t = \frac{\partial h_{\psi}^{\text{dyn}}}{\partial \mathbf{u}}(\mathbf{z}_t, \mathbf{u}_t), \quad G_t = A_t B_t B_t^\top A_t^\top. \quad (4)$$

The loss is then computed as the minimum energy between the prediction and the true next latent state, that is  $\mathcal{L}_t^{\text{dyn}} = (\mathbf{z}_{t+1} - \hat{\mathbf{z}}_{t+1})^\top G_t^{-1} (\mathbf{z}_{t+1} - \hat{\mathbf{z}}_{t+1})$ . Essentially this enforces that the loss in prediction is representative of the dynamics of the system, e.g., for a car, missing the latent state prediction forward or backward is less important than missing left and right. Note however, that initially, before

the dynamics network has been trained, the weighted controllability Gramian from Eq. 4 may be ill-conditioned. We thus begin with an  $\ell_2$  norm and gradually shift to the  $G_t^{-1}$  norm. We also add a small positive term to the diagonal of  $G_t$  to ensure invertibility.

### 3.2 Learning a Collision Checker

The third network in our architecture learns to identify if the continuous trajectory between two latent states will be in collision; we denote such network as  $h_{\zeta}^{cc}(\mathbf{z}_t, \mathbf{z}_{t+1}, \mathcal{X}_{obs})$ , parameterized by  $\zeta$ . This is a supervised learning process entailing binary classification; meaning, we assume we have a collision checker in the full state space capable of reporting if the trajectory is in collision – arguably, a very mild assumption. We note here our use of  $\mathcal{X}_{obs}$  as an input to the network. In SBMP  $\mathcal{X}_{obs}$  is generally only implicitly defined through the collision checker. We use it here to denote any environmental information available to the state space collision checker. This representation can be quite general as the neural network architecture has the capacity to arbitrarily project this information as necessary. For example, this may be workspace coordinates of the obstacles, an occupancy grid, or an image of the problem. This network is learned separately, after the encoder and dynamics have been learned. It is possible to learn this network through unsupervised learning by identifying regions of the state space that are rarely or never connected dependent on the environment. We leave this for future work.

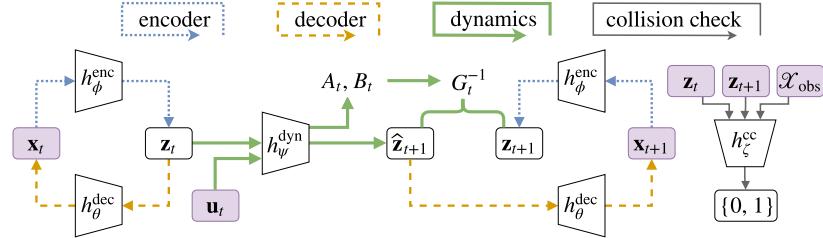


Fig. 1: The L-SBMP network architecture. The encoder network is shown in blue, the decoder network in orange, the dynamics network in green, and the collision checking network in gray. The inputs are shown in purple.

## 4 Planning in the Latent Space

We now present the Learned Latent Rapidly-exploring Random Trees (L2RRT) algorithm used to globally search the latent space learned as described in Section 3. L2RRT is a SBMP algorithm within the rapidly exploring random trees (RRT) family of algorithms [2]. Briefly, the algorithm builds a tree,  $\mathcal{T}$ , through the latent space by sampling latent samples from the learned manifold and propagating dynamics from nearby tree nodes (under the supervision of the collision checking network). As the learned dynamics may be quite complex, the algorithm must rely on a dynamics propagation strategy, rather than a steering function, to connect local latent states.

The L2RRT algorithm is similar to the RRT-BestNear algorithm presented in [22], with samples restricted to the latent space and dynamic propagation through latent dynamics. Specifically, the algorithm is outlined in Alg. 1 and visualized in Fig. 2. The algorithm takes as input the planning problem in the full state space: the obstacle representation,  $\mathcal{X}_{\text{obs}}$  initial state,  $\mathbf{x}_{\text{init}}$ , and goal region,  $\mathcal{X}_{\text{goal}}$ . Note, we abuse notation here with  $\mathcal{X}_{\text{obs}}$ . Traditionally in SBMP,  $\mathcal{X}_{\text{obs}}$  is only implicitly defined through the collision-checker. Here we use it to refer to any environmental information used for the state space collision checker. This may be, for example, workspace coordinates of the obstacles, an occupancy grid, or an image of the problem.

To begin, the initial state and goal region are encoded into the latent space as  $h_{\phi}^{\text{enc}}(\mathbf{x}_{\text{init}}) = \mathbf{z}_{\text{init}}$  and  $h_{\phi}^{\text{enc}}(\mathcal{X}_{\text{goal}}) = \mathcal{Z}_{\text{goal}}$ , respectively (line 1). The tree is initialized with the initial latent state (line 2). The algorithm then begins expanding the tree outwards into the latent space. At each iteration a sample,  $\mathbf{z}_{\text{sample}}$ , is selected from a sample set  $\mathcal{Z}_{\text{samples}}$  (line 4). This set is comprised of latent samples generated by encoding a randomly drawn set of the robot in operation. That is, given a sequence of states  $(\mathbf{x}_0, \dots, \mathbf{x}_N)$  (where  $N$  is large to well cover the operational state space of the robot), we randomly select a sample set  $\mathcal{X}_{\text{samples}}$  and encode it as  $\mathcal{Z}_{\text{samples}} = h_{\phi}^{\text{enc}}(\mathcal{X}_{\text{samples}})$ . By selecting samples only from this set, L2RRT ensures that the exploration remains nearby the learned manifold in the latent space. L2RRT then selects a node in the tree for propagation by using the Best First Selection strategy outlined in [22] (line 5). This strategy selects the tree node  $\mathbf{z}_{\text{prop}}$  within a ball radius,  $(\mathbf{z}_{\text{prop}} - \mathbf{z}_{\text{sample}})^T G_t^{-1} (\mathbf{z}_{\text{prop}} - \mathbf{z}_{\text{sample}}) < \delta_{G_t^{-1}}$ , with the lowest-cost trajectory<sup>1</sup>. This parameter ball radius  $\delta_{G_t^{-1}}$  represents a tradeoff between exploration and exploitation of good trajectories; its value is discussed in the subsequent paragraph and at length in [22]. With  $\mathbf{z}_{\text{prop}}$  in hand, L2RRT then forward propagates dynamics through  $h_{\psi}^{\text{dyn}}$  with a randomly chosen control action  $\mathbf{u}_{0, \dots, T_{\text{prop}}-1} \in \mathcal{U}$  for time  $T_{\text{prop}}$ , where  $T_{\text{prop}}$  is randomly chosen between 1 and  $T_{\text{max}}$ , following the Monte Carlo propagation procedure outlined in [22] (line 6). This gives a trajectory in latent space  $(\mathbf{z}_0, \mathbf{u}_0, \dots, \mathbf{z}_{T_{\text{prop}}-1}, \mathbf{u}_{T_{\text{prop}}-1}, \mathbf{z}_{T_{\text{prop}}}, 0)$ , where  $\mathbf{z}_0 = \mathbf{z}_{\text{prop}}$  (line 7). Each pair of waypoints is then collision checked by  $h_{\zeta}^{\text{cc}}(\mathbf{z}_t, \mathbf{z}_{t+1}, \mathcal{X}_{\text{obs}})$  for  $t = 0, \dots, T_{\text{prop}} - 1$ . The probability that a local connection is collision-free can be computed as sigmoid( $h_{\zeta}^{\text{cc}}$ ) (line 8). Thus, acceptable probability threshold  $\alpha$  can be tuned to the required level of safety. Finally, if the propagated trajectory is collision free,  $\mathbf{z}_{T_{\text{prop}}}$  is added to the tree (line 9). The exploration ends after a set amount of time or number of samples. If no trajectories end in the goal region  $\mathcal{Z}_{\text{goal}}$ , the algorithm reports failure (line 13). If any trajectories reach the goal region, the best trajectory<sup>1</sup> is then selected and projected back to the full state space through  $h_{\theta}^{\text{dec}}$  (line 15-16). This gives both the trajectory in the full state space, which may be used as input to a controller or a trajectory optimizer, as well as nominal control inputs.

Briefly we provide discussion on how to choose values for the algorithm's parameters. The  $\delta_{G_t^{-1}}$  ball radius represents a tradeoff between exploiting good trajectories and exploring with new ones. This value should be tuned according to the system and from discussion in [22]. The maximum dynamic propagation

---

<sup>1</sup>This cost can be fairly general, e.g., control cost or the number of time steps.

**1** Learned Latent RRT Outline

---

**Parameters:**  $\delta_{G_t^{-1}}$ ,  $T_{\max}$ ,  $\mathcal{Z}_{\text{samples}}$ ,  $\alpha$

**Inputs:**  $(\mathbf{x}_{\text{init}}, \mathcal{X}_{\text{goal}}, \mathcal{X}_{\text{free}})$

- 1 Encode problem input into latent space:  $\mathbf{z}_{\text{init}} = h_{\phi}^{\text{enc}}(\mathbf{x}_{\text{init}})$ ,  $\mathcal{Z}_{\text{goal}} = h_{\phi}^{\text{enc}}(\mathcal{X}_{\text{goal}})$
- 2 Add  $\mathbf{z}_{\text{init}}$  to the tree  $\mathcal{T}$
- 3 **while** time remains **do**
- 4   Sample  $\mathbf{z}_{\text{sample}} \in \mathcal{Z}_{\text{samples}}$
- 5   Select  $\mathbf{z}_{\text{prop}} \in \mathcal{T}$  as best tree node within a  $\delta_{G_t^{-1}}$  ball radius of  $\mathbf{z}_{\text{sample}}$   
(if the  $\delta_{G_t^{-1}}$  ball is empty, select nearest tree node to  $\mathbf{z}_{\text{sample}}$ , weighted by  $G_t^{-1}$ )
- 6   Sample  $T_{\text{prop}} \in \{1, \dots, T_{\max}\}$  and  $\mathbf{u}_{0, \dots, T_{\text{prop}}-1} \in \mathcal{U}$
- 7   Propagate dynamics  $\mathbf{z}_{t+1} = h_{\psi}^{\text{dyn}}(\mathbf{z}_t, \mathbf{u}_t)$ , where  $\mathbf{z}_0 = \mathbf{z}_{\text{prop}}$  and  $t = 0, \dots, T_{\text{prop}} - 1$ , to generate latent trajectory  $(\mathbf{z}_0, \mathbf{u}_0, \dots, \mathbf{z}_{T_{\text{prop}}-1}, \mathbf{u}_{T_{\text{prop}}-1}, \mathbf{z}_{T_{\text{prop}}}, 0)$
- 8   **if**  $\text{sigmoid}(h_{\zeta}^{\text{cc}}(\mathbf{z}_t, \mathbf{z}_{t+1}, \mathcal{X}_{\text{obs}})) > \alpha \forall t = 0, \dots, T_{\text{prop}} - 1$  **then**
- 9     Add  $\mathbf{z}_T$  to the tree  $\mathcal{T}$
- 10   **end if**
- 11 **end while**
- 12 **if**  $\mathcal{T} \cap \mathcal{Z}_{\text{goal}} = \emptyset$  **then**
- 13   Report failure
- 14 **else**
- 15   Select lowest-cost trajectory with its end node in  $\mathcal{T} \cap \mathcal{Z}_{\text{goal}}$
- 16   Project back to full state space through  $h_{\theta}^{\text{dec}}$  and return  $(\mathbf{x}_{\text{init}}, \mathbf{u}_0, \dots, \mathbf{x}_T, \mathbf{u}_T)$ , where  $T$  is the number of time steps to  $\mathbf{z}_{\text{goal}}$
- 17 **end if**

---

time  $T_{\max}$  controls, roughly, the length of edge. For subsequent results this was approximately set so the robot with optimal control can traverse the full state space in  $4T_{\max}$ . The value of  $\alpha$  reflects the required level of safety in the collision checking network. We found most collision classifications yielded probabilities less than 1% or greater than 99%, however, for subsequent results this threshold was set as  $\alpha = 0.9$ . Finally, we also use a technique called goal-biasing, in which we sample more heavily from the goal region. For subsequent results we sample a state within the goal region 10% of the time.

Due to the approximate nature of the learned latent space, L2RRT does not enjoy formal completeness or optimality guarantees. An interesting avenue for

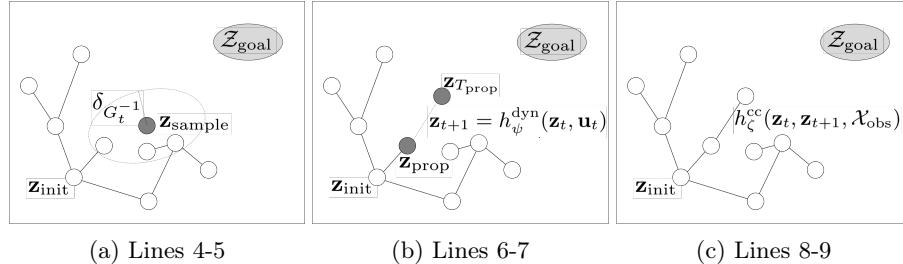


Fig. 2: Learned Latent RRT exploring the latent space, as described in Alg. 1. (2a) Sample  $\mathbf{z}_{\text{sample}}$  and select lowest-cost tree node within a  $\delta_{G_t^{-1}}$  ball. (2b) Sample time  $T_{\text{prop}}$  and controls  $\mathbf{u}_{0, \dots, T_{\text{prop}}-1}$ . Propagate dynamics with  $h_{\psi}^{\text{dyn}}$ . (2c) Verify the new edge is collision free with  $h_{\zeta}^{\text{cc}}$ . Add the node to the tree.

future research is to investigate conditions under which theoretical guarantees for L2RRT can be derived.

## 5 Experimental Results

In this section we demonstrate the effectiveness of our methodology on two planning problems. The first is a visual planning problem entailing a point robot with single integrator dynamics navigating a cluttered environment based on visual inputs. The second is a humanoid planning problem where a humanoid robot has to maneuver amidst obstacles. The methodology was implemented in Python and Tensorflow [23] with dynamics from MuJoCo [24]. All data was collected on a Unix system with a 3.4 GHz CPU and an NVIDIA GeForce GTX 1080 Ti GPU. Example code can be found at <https://github.com/StanfordASL/LSBMP>.

### 5.1 Visual Planning Problem

We consider a visual space planar problem similar to the one outlined in [3], though with a more complex obstacle space. This setup allows us to demonstrate the performance of the L-SBMP methodology to generalize to new environments by leveraging the algorithmic prior of separating collision checking from the dynamic propagation process. As we have access to the true underlying state, and the underlying planning problem is relatively simple, this problem allows us to compare our methodology against sampling-based motion planning algorithms with access to the true state of the system.

The full state of the visual problem is a  $32 \times 32$  pixel image rendered in MuJoCo showing both the robot and the obstacles in the scene. The robot is a small point robot with single integrator dynamics. The obstacles are a random number of circles and squares with randomly generated locations and sizes. An example problem's initial and goal states are shown in Figs. 3a-3b. The training data for the autoencoder and dynamics network consists of 10,000 environments, each with a trajectory of 10 successive states and control inputs. The training data for the collision checking network consists of 25,000 environments, again with 10 pairs of states and a collision label.

The full state is projected down to a two dimensional latent space. The methodology applied to an image-based system requires some architectural changes to make the learning problem more tractable. The encoder network,  $h_\phi^{\text{enc}}$ , is a deep-spatial autoencoder [7], which encourages learning important visual features by using a convolutional neural network followed by a spatial soft arg-max. To create accurate reconstructions of the full state the decoder network is given as input an image of the environment in addition to the latent state. This additional input image may be the environment without the robot present or it could be an image with the robot from a different time step (e.g., the initial state for a planning problem).

Fig. 3c shows the learned latent space for the problem in Fig. 3a-3b. The methodology learns a latent space that captures the position of the robot and effectively removes latent regions within obstacles, though this obstacle set has never been seen before. Fig. 3d shows the inverse weighted controllability Gramian

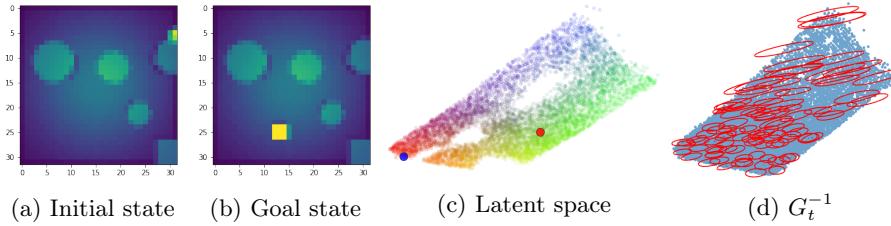


Fig. 3: (3a-3b) The full state of the visual robot with obstacles (robot is shown in yellow) and initial and goal states for input to our methodology. (3c) The learned latent space colored by the position of the robot in the true underlying state space. States from the lower-left corner are green, lower-right are yellow, upper-right are red, and upper-left are blue. The encoded initial state is shown in blue and the goal state in red. States in collision are not visualized (empty white regions). (3d) The inverse of the weighted controllability Gramian in the state space (an ellipse representing equal energy costs). The learned latent space in the upper right is less dense so the ellipses are larger in the direction of the spread, thus penalizing errors less.

in the latent space. As we know the true dynamics of are system are the same everywhere in the state space, this gives an opportunity to evaluate this weighted loss function. As desired, in regions of the state space that are more spread (i.e., the same control input allows more movement), errors are penalized less.

Fig. 4 shows two planning problems for this system, the planned trajectories in the full state space, and the L2RRT in latent space. L-SBMP is able to successfully generalize to these problems and find trajectories even through narrow passages. These global plans can then be fed into a local controller, such as [6], or a trajectory optimizer [8], to be executed.

Fig. 5 shows the success rate and cost convergence curve (in terms of control input, i.e., distance) versus the number of samples. The success rate increases quickly at first before leveling below 100% (though all problems are solvable). The cost convergence curve follows a standard curve for SBMP, converging quickly at first before leveling out. The confusion matrix shows nearly 90% of collisions were correctly classified, with only 4% of true collisions classified as free. This false positive ratio could likely be reduced by including conservatism in the full state space collision checker. With 1000 samples, the algorithm took on average 18 seconds.

We additionally compare these plans to plans computed in the true state space (given a perfect representation of the robot and obstacles in two dimensions) by the Fast Marching Tree (FMT\*) algorithm [21], an asymptotically optimal sampling-based motion planning algorithm, and by L2RRT (again, run in the true state space). Each algorithm is run with 2000 samples and over 100 different planning problems (varying the initial states, goal regions, and obstacle sets). We note that FMT\* relies on a two point boundary value problem solver to make local connections. This allows FMT\* to explore the state space more efficiently than L2RRT (which only uses dynamics propagation). In terms of success rate of finding a solution, normalized to FMT\*, L2RRT run in the latent space was able to solve 92% of the problems while L2RRT in the true state space solved 96%. In terms of solution trajectory cost, L2RRT solved these problems with a 13% higher cost than FMT\* and L2RRT in the true state space solved

with a 5% higher cost than FMT\*. From these results we make two observations. First, planning in the latent space incurs some loss (in terms of success rate and cost), however, this loss is relatively small. Second, if the latent space were forced into a more restrictive class of dynamical systems (e.g., locally-linear) with a two point boundary value problem solver available, the L-SBMP methodology could leverage algorithms like FMT\* to more efficiently explore the state space. We leave this to future work.

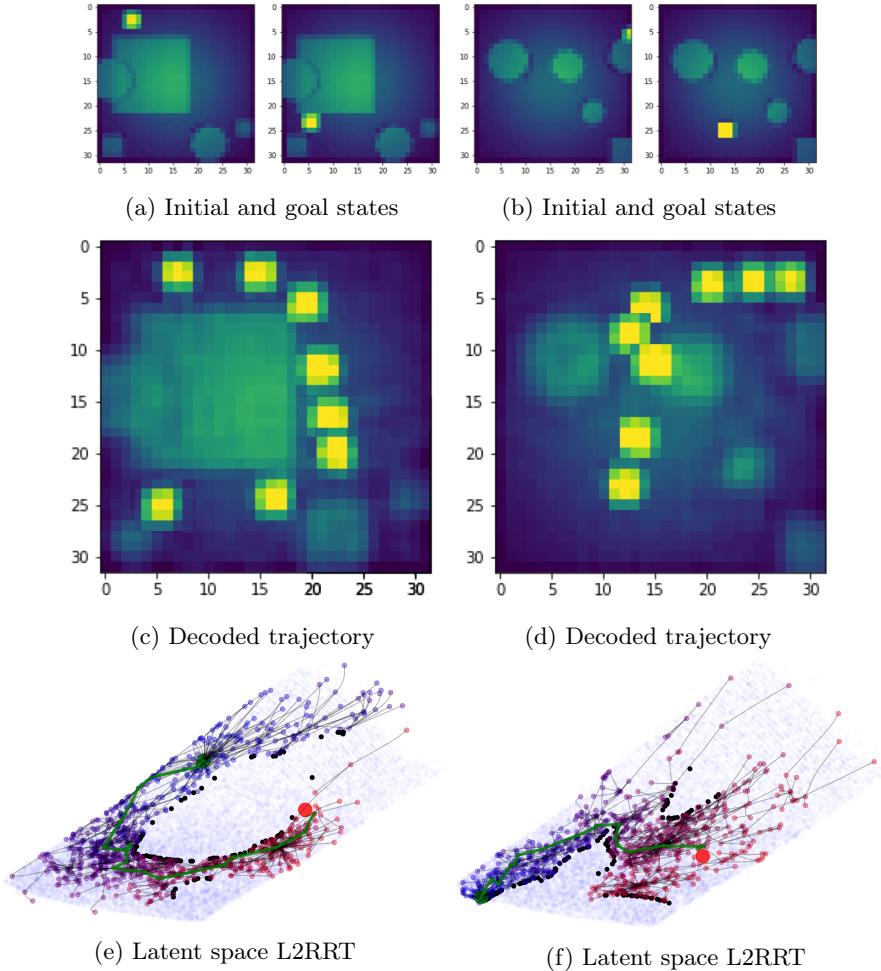


Fig. 4: (4a-4b) Initial and goal states for the visual planning problem. (4c-4d) Planned trajectories decoded into the full state space. (4e-4f) L2RRT in the latent space, where the initial state and final trajectory are shown in green, the goal state in red, and collisions between states in black. States within the tree are color-coded by their cost (blue denotes low cost, red denotes high cost). Lastly, samples from the training data are used as samples in L2RRT and are shown in blue.

Our second numerical experiment shows the methodology with a high-dimensional dynamical system, a MuJoCo humanoid in OpenAI Roboschool [5]. The system

consists of 50 states with dynamics simulated with MuJoCo [24] and a learned controller from OpenAI Roboschool which takes as input a target position. The obstacles are randomly generated spheres or cubes of random sizes and positions. The initial and goal states too are randomly generated through the state space.

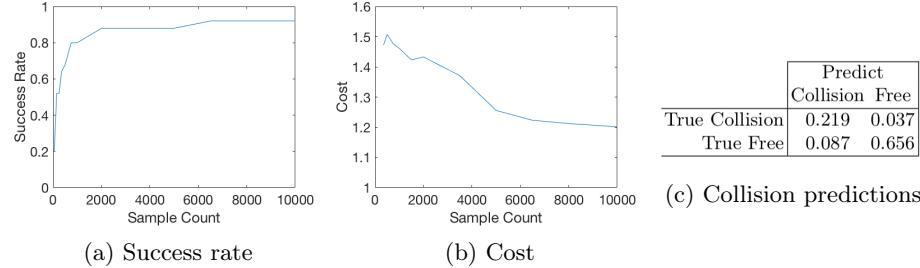


Fig. 5: Convergence curves and collision confusion table for visual planning problem.

## 5.2 Humanoid Robot

The state is encoded into a four dimensional latent space. The dimensionality was chosen inspired by [16]. The network architecture is fully connected. The collision checking network takes as input two latent states and a vector of the centers of each obstacle, the type of obstacle (sphere or cube), and the sphere radius or cube width of the obstacle. The autoencoder and dynamics network was trained on 1,000,000 data points (states and control input), sampled at 10 Hz. The collision checking network was trained on 1,000,000 data points (pairs of states and the location, type, and size of each obstacle in the workspace).

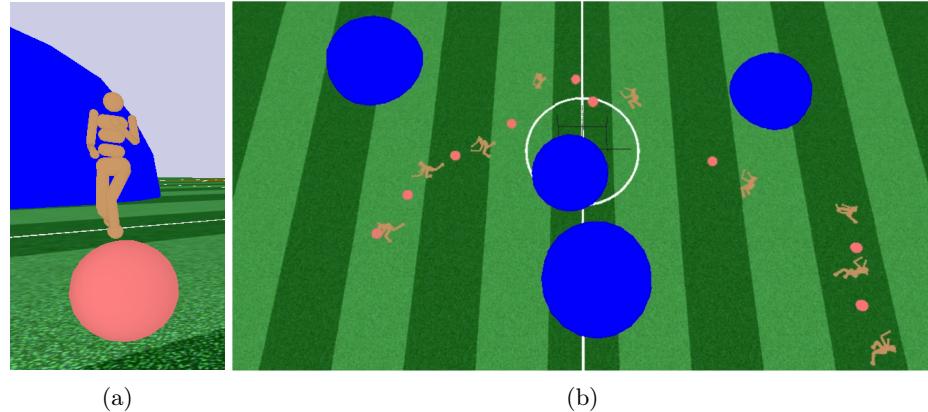


Fig. 6: (6a) The humanoid robot with an obstacle in blue and the control target in red. (6b) The humanoid robot executing the planned trajectory in the full state space.

Fig. 7 shows the structure learned within the latent space. The latent space captures the intuitively important dynamics of the problem: the robot's position and yaw. These states most significantly govern the humanoid's dynamics, i.e., it

is often traveling in a straight line towards the control target. The other dimensions of the state space, such as joint angles and velocities, are fairly structured within the gait of the humanoid and only vary small amounts.

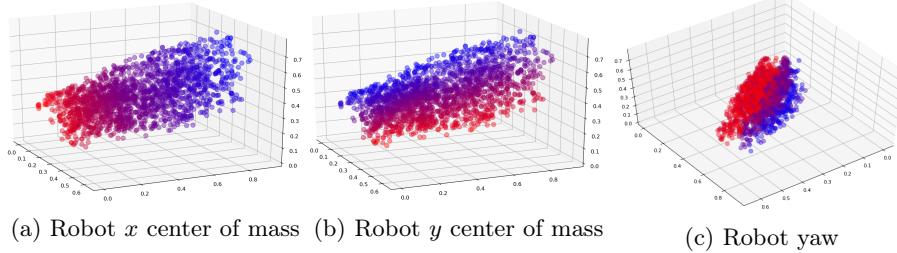


Fig. 7: A 3D representation of the structure learned within the humanoid latent space, with each plot visualized by the robot’s  $x$ ,  $y$ , yaw respectively (where red denotes a low value and blue a high value).

For a single problem, L2RRT for the humanoid (projected into two and three dimensions) is shown in Fig. 8. The latent search is able to remain near the learned manifold, and the collision checker has effectively encapsulated the obstacles. The final trajectory is shown in Fig. 6b with the control input (the target location) set as a half second ahead on the planned trajectory. A video is available at <https://goo.gl/wrm9JR>. Fig. 9 details L2RRT’s performance in terms of success rate, convergence, and collision network confusion. After only a few hundreds of samples, the algorithm was able to solve 100% of the planning problems. The trajectory cost again showed convergence as the number of samples increased. The collision checking network performed remarkably well, classifying nearly 95% correct, with only a 0.7% false positive rate. With 1000 samples, the algorithm took on average 30 seconds.

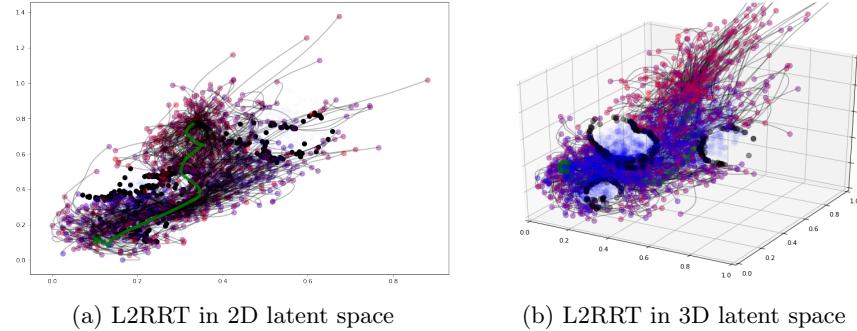


Fig. 8: L2RRT in latent space with nodes colored by cost (blue is low cost, red is high cost), collisions are shown in black, and the final trajectory is shown in green.

## 6 Conclusions

We have presented the Latent Sampling-based Motion Planning (L-SBMP) methodology, which leverages the effectiveness of local, latent representations for robotic

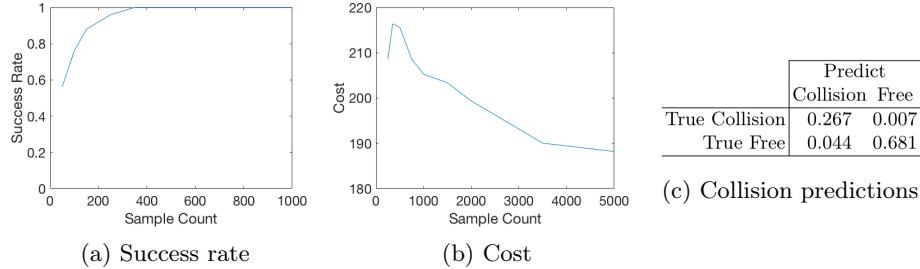


Fig. 9: Convergence curves and collision confusion table for humanoid problem.

systems with techniques from sampling-based motion planning (SBMP) to compute motion plans for high-dimensional, complex systems. In particular, this latent space is learned through an autoencoder, a dynamics network, and a separate collision checking network, each of which enforces the main algorithmic primitives of SBMP on the latent space (sampling, local connections, and collision checking). We further have presented the Learned Latent RRT (L2RRT) algorithm, for globally exploring this latent space and computing motion plans directly in it. Through two experiments, one planning within visual space and one planning with a humanoid robot, we demonstrate the methodology’s overall generality and ability to generalize to new environments.

This work leaves several avenues for future work. Specific to L-SBMP, we first plan to investigate how much data is required to learn the necessary representation. Second, we plan to investigate learning more restricted classes of dynamics for which a steering function is available, e.g., locally-linear dynamics. This would allow more optimal latent space exploration techniques. Third, we plan to investigate using unsupervised learning to learn the collision checking network. Fourth, we plan to investigate planning problems in which the topology of the latent space is dependent on the environment and obstacle set specific to a planning problem, e.g., if stairs are involved for a humanoid. Fifth, we plan to use a similar methodology within a task and motion planning framework by state augmentation in the full state space. Finally, we plan to investigate conditions under which theoretical guarantees for L2RRT can be derived. Beyond L-SBMP, the approach of learning plannable latent spaces through enforced algorithmic primitives can be quite powerful. We believe a similar methodology could be used to learn a space directly for trajectory optimization, or instead for lower-dimensional problems, combinatorial motion planning.

## Acknowledgments

This work was supported by NASA under the Space Technology Research Grants Program, Grant NNX12AQ43G, and by the King Abdulaziz City for Science and Technology (KACST). The GPUs used for this research were donated by the NVIDIA Corporation. The authors wish to thank Edward Schmerling and Ani Majumdar for their helpful discussions and feedback on this work.

## References

1. S. M. LaValle, *Planning Algorithms*. Cambridge Univ. Press, 2006.
2. S. M. LaValle and J. J. Kuffner, “Randomized kinodynamic planning,” *Int. Journal of Robotics Research*, 2001.
3. M. Watter, J. Springenberg, J. Boedecker, and M. Riedmiller, “Embed to control: A locally linear latent dynamics model for control from raw images,” in *NIPS*, 2015.
4. C. Finn and S. Levine, “Deep visual foresight for planning robot motion,” in *ICRA*, 2017.
5. G. Brockman, V. Cheung, L. Pettersson, J. Schneider, J. Schulman, J. Tang, and W. Zaremba, “OpenAI Gym,” 2016.
6. E. Banijamali, R. Shu, M. Ghavamzadeh, H. Bui, and A. Ghodsi, “Robust locally-linear controllable embedding,” *arXiv preprint arXiv:1710.05373*, 2017.
7. C. Finn, X. Y. Tan, Y. Duan, T. Darrell, S. Levine, and P. Abbeel, “Deep spatial autoencoders for visuomotor learning,” in *ICRA*, 2016.
8. A. Srivivas, A. Jabri, P. Abbeel, S. Levine, and C. Finn, “Universal planning networks,” *arXiv preprint arXiv:1804.00645*, 2018.
9. Y. Bengio, A. Courville, and P. Vincent, “Representation learning: A review and new perspectives,” *TPAMI*, 2013.
10. S. T. Roweis and L. K. Saul, “Nonlinear dimensionality reduction by locally linear embedding,” *Science*, 2000.
11. G. E. Hinton and R. R. Salakhutdinov, “Reducing the dimensionality of data with neural networks,” *Science*, 2006.
12. J. Piater, S. Jodogne, R. Detry, D. Kraft, N. Krüger, O. Kroemer, and J. Peters, “Learning visual representations for perception-action systems,” *IJRR*, 2011.
13. J. D. Co-Reyes, Y. Liu, A. Gupta, B. Eysenbach, P. Abbeel, and S. Levine, “Self-consistent trajectory autoencoder: Hierarchical reinforcement learning with trajectory embeddings,” *arXiv preprint arXiv:1806.02813*, 2018.
14. E. Schmerling, K. Leung, W. Vollprecht, and M. Pavone, “Multimodal probabilistic model-based planning for human-robot interaction,” in *Proc. IEEE Conf. on Robotics and Automation*, 2018.
15. N. Chen, M. Karl, and P. van der Smagt, “Dynamic movement primitives in latent space of time-dependent variational autoencoders,” in *IEEE-RAS Humanoids*, 2016.
16. J.-S. Ha, H.-J. Chae, and H.-L. Choi, “High-dimensional motion planning using latent variable models via approximate inference,” *arXiv preprint arXiv:1711.08275*, 2017.
17. A. H. Qureshi, M. J. Bency, and M. C. Yip, “Motion planning networks,” *arXiv preprint arXiv:1806.05767*, 2018.
18. N. Das, N. Gupta, and M. Yip, “Fastron: A learning-based configuration space model for rapid collision detection for gross motion planning in changing environments,” in *RSS Workshop Data-Driven Manipulation*, 2017.
19. B. Ichter, J. Harrison, and M. Pavone, “Learning sampling distributions for robot motion planning,” in *Proc. IEEE Conf. on Robotics and Automation*, 2018.
20. S. Karaman and E. Frazzoli, “Sampling-based algorithms for optimal motion planning,” *Int. Journal of Robotics Research*, 2011.
21. L. Janson, E. Schmerling, A. Clark, and M. Pavone, “Fast Marching Tree: a fast marching sampling-based method for optimal motion planning in many dimensions,” *Int. Journal of Robotics Research*, 2015.
22. Y. Li, Z. Littlefield, and K. E. Bekris, “Asymptotically optimal sampling-based kinodynamic planning,” *Int. Journal of Robotics Research*, 2016.
23. M. Abadi *et al.* (2015) TensorFlow: Large-scale machine learning on heterogeneous systems. Software available from tensorflow.org. [Online]. Available: <https://www.tensorflow.org/>
24. E. Todorov, T. Erez, and Y. Tassa, “Mujoco: A physics engine for model-based control,” in *IROS*, 2012.