nature COMMUNICATIONS

ARTICLE

https://doi.org/10.1038/s41467-022-31795-2 OPEN

Check for updates

Mechanical intelligence for learning embodied sensor-object relationships

Ahalya Prabhakar[1]✉ & Todd Murphey[2]✉

Intelligence involves processing sensory experiences into representations useful for prediction. Understanding sensory experiences and building these contextual representations without prior knowledge of sensor models and environment is a challenging unsupervised learning problem. Current machine learning methods process new sensory data using prior knowledge defined by either domain knowledge or datasets. When datasets are not available, data acquisition is needed, though automating exploration in support of learning is still an unsolved problem. Here we develop a method that enables agents to efficiently collect data for learning a predictive sensor model—without requiring domain knowledge, human input, or previously existing data—using ergodicity to specify the data acquisition process. This approach is based entirely on data-driven sensor characteristics rather than predefined knowledge of the sensor model and its physical characteristics. We learn higher quality models with lower energy expenditure during exploration for data acquisition compared to competing approaches, including both random sampling and information maximization. In addition to applications in autonomy, our approach provides a potential model of how animals use their motor control to develop high quality models of their sensors (sight, sound, touch) before having knowledge of their sensor capabilities or their surrounding environment.

---

[1] École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland. [2] Northwestern University, Chicago, IL, USA. ✉email: ahalya.prabhakar@epfl.ch; t-murphey@northwestern.edu

A defining characteristic of intelligence and learning is the ability to process sensory experiences into concise representations from which one can predict and anticipate sensory experience at new states[1–3]. Humans recognize and encode complex scenes and objects in memory with remarkable accuracy, able to imagine them from different viewpoints or places[1–3]. Recent efforts have focused on developing neural networks and models for learning concise representations from sensory data[4,5], but these approaches require that datasets for learning already exist. The challenge of data acquisition in support of learning, when these datasets are not available, remains unsolved. This lack of a theory explaining where data should come from affects all data-driven science[6] since the quality of machine intelligence and resulting human behavior depends on quality of data. Information-based search strategies have been used for both modeling animal behavior[7,8] and synthesizing behavior[9] for artificial systems. However, there are no principles explaining how mechanical motion supports general models of learning. This paper creates a general model of how agents use motor capability to support learning when the data for learning needs to be physically acquired—a typically necessary step that is often not considered in the machine learning literature that uses canned datasets acquired elsewhere. When training data acquired for learning is dependent on exploratory strategies throughout the environment, principles for motion control become a cornerstone of principles for learning.

In this paper, we use the challenge of learning predictive sensor models without predefined knowledge of the sensor or environment as a setting for deriving physics-constrained motion strategies for learning. Our method works well without modification for both near-field and far-field sensor examples. We accomplish this using a motor principle based on ergodicity—a physically important property from statistical mechanics[10] that relates temporal behavior of trajectories to spatial distributions—with respect to information-rich regions of a physical domain. This approach results in dramatically higher quality learned models with lower energy expenditure than approaches that rely on randomized sampling of the environment, the dominant data acquisition strategy in nearly all machine learning. Moreover, the approach presented here outperforms entropy-based information maximization in terms of learning performance (see Supplementary Information).

Our learning framework is based on autoencoders; these learn an encoder-decoder framework that encodes input data into a lower-dimensional latent representation from which signals are decoded for prediction[11]. Conditional autoencoders learn a predictive model by learning a latent space conditioned on external factors (e.g., agent pose). Specifically, we learn a sensory model of the form ($y' = f(\theta_e, \theta_d, x')$)—where θ_e and θ_d parameterize an encoder and decoder in an autoencoder that has a latent space z determined from a seed observation at robot state x_0—in order to predict what the agent would experience at a new state x'. Active data acquisition is specified by minimizing the distance from ergodicity[12] between the time-averaged statistics of an agent along a trajectory and a target coverage spatial distribution (that represents how information about the autoencoder is distributed spatially). Ergodicity with respect to the entropy of variables in a learning representation is a principle of motion, much like energy minimization[13,14] and entropy[7].

Metrics on ergodicity for synthesizing motion for active learning. Unsupervised learning for predictive sensory models (e.g., scene prediction using data from sensors in self-driving cars) relies on processing and learning from batches of data received continuously and analyzed intermittently. This episodic framework benefits from a data acquisition approach that is simultaneously responsive to changes in the learning representation (e.g., changes to a deep neural network) and persistent in its data collection while waiting for those changes. Unlike other prediction-based reward functions that focus on acquisition of unusual measurements[15], metrics on ergodicity—measures that indicate how well a particular trajectory $x(t)$ covers a specified distribution $\phi(x)$—enable the formulation of a controlled response to the need for data that seeks out informative regions based on the entropy of the learned model over the exploration space, generating mechanical motion that actively aids intelligent learning. Ergodic control—based on a metric on ergodicity—enables synthesis of exploratory motions that adapt in response to data collected and evolving information distributions.

Autoencoders provide an estimate of a latent space variable z, explicitly dependent on a conditional variable x, based on a sensor observation y. As learning progresses, the parameters of the encoder and decoder networks—θ_e and θ_d, respectively—are optimized to interpret input data to best predict sensory data, such that it outputs both a prediction of the sensor observation y' and its entropy $\mathbb{H}_y$. As illustrated in Fig. 1, the state of a sensor is the conditional variable in this framework; the learned representation changes as the sensor state changes (e.g., pose of a camera or pressure and location of a tactile sensor). Conditional autoencoders are the representation of learning we focus on here, but their key characteristic is that the learning representation is explicitly conditioned on something over which an agent has control; the arguments used here in the context of conditional autoencoders would apply to other learning methodologies with this same property.

Since the entropy estimate approximates how information in the autoencoder depends on the state of the sensor, we maximize the ergodicity of the sensor trajectory with respect to entropy of the autoencoder model to dictate how the sensor moves in order

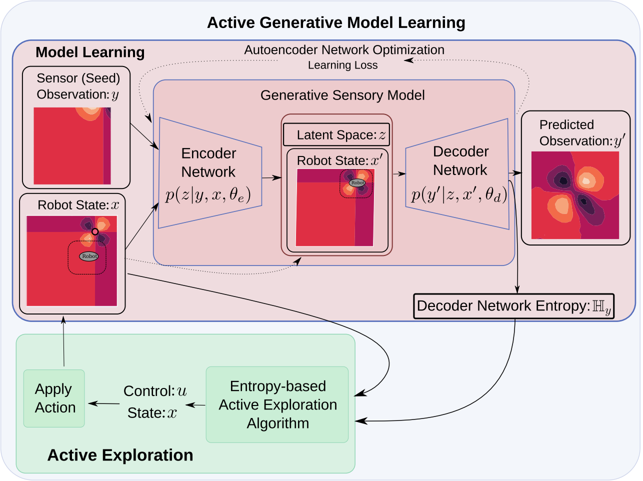

Fig. 1 Model learning approach. The learning approach consists of a learning algorithm for creating generative sensory models from data and an active exploration framework that collects data based on the conditional entropy of the learned model. Conditional autoencoders are used to learn a low-dimensional latent encoding of the high-dimensional sensor input. Sensor output at novel states—the generative aspect of the model—is then predicted using this low-dimensional encoding. We develop an active exploration approach for data acquisition based on ergodic coverage of the entropy of the latent space to improve model uncertainty. This framework demonstrates a feasible approach for real-time, generalizable sensor model learning without the need for predefined domain knowledge of the environment or sensor structure.

to acquire more information about the unknown or partially known autoencoder. We assume that the conditional variable x is a state of a controllable system dependent on an input u, and that the dynamics $\dot{x} = f(x, u)$ of the sensor are known (see Supplementary Information). We numerically optimize u to be maximally ergodic with respect to $\mathbb{H}_y$, the entropy of the decoder network by minimizing a metric on ergodicity, where in this case $\phi(x)$ is $\mathbb{H}_y$ conditioned on x and the latent space representation z derived from the current x, y (also discussed in the Supplementary Information). We refer to this control as *ergodic sampling*.

Results

Near-field sensors (e.g., touch, electrosense) and far-field sensors (e.g., vision, hearing) depend differently on actions taken by an agent. In the case of vision, actions that support learning involve getting a target within the field of view at a sufficient distance to identify visual features, avoiding motion to avoid blur. In contrast, touch requires proximity to a target and often requires motion relative to the target; information content drops dramatically as soon as a sensor is out of contact. The ergodic sampling method applies to both near-field and far-field sensory learning without modification. We demonstrate this using electrosense physics as a near-field sensory model and vision as a far-field model.

First, we explore the benefits of ergodic sampling for learning an electrosensory model. The electrosensory modality—commonly referred to as electrolocation—is used by a group of South American weakly electric freshwater nocturnal fish that hunts in darkly tinted waters where visibility is low. Understanding their unique sensing modality and decision-making has been of interest both for understanding biological behavior and for using these sensor modalities in robotics[16–18]. Electrosense is a near-field sensory modality that only produces electrical perturbations in voltage observations when near the object[16], highlighting the importance of active data acquisition—like touch, model generation is impossible without movement. In the case of weakly electric fish, electrosense measurements are distributed across the surface of the fish body, yielding a sensor where each voltage measurement can be thought of as a pixel in an image. However, these sensory activity readings are proportional to the voltage difference between inside and outside the skin at the sensor, so the geometry of the pixels does not directly correspond to the geometry of the environment, the way a camera's pixels do. Similar to a camera image, the signal space of the electrosensory modality is high dimensional. We construct an electrosense measurement model (described in the Supplementary Information) of an underwater conductor in an environment with a spherical object at an unknown location. In Fig. 2 we compare the results of electrosensory model learning using ergodic sampling and compare to random sampling as a benchmark—reflecting the ideal sampling around the workspace without having to move according to the laws of physics (see Supplementary Information for details of numerical experiments).

Trajectories spend time close to the object, as seen in Fig. 2 (bottom). Without any prior knowledge of the sensor measurement model or the object, this method automatically detects key aspects of near-field sensing without those features being prespecified. The sensor model has a 441-dimensional sensor observation y—which consist of voltage differentials evenly distributed in a 21×21 grid (covering the sensor coverage region indicated in Fig. 2) over the center of the robot—and the robot state x and encodes y into a 2-dimensional latent space representation z conditioned on x. The latent space representation was determined from a single seed observation (top right) with the agent sufficiently near the object to register an electrical perturbation—an observation sufficient for determining features of perturbations but insufficient to otherwise describe the location of the conductor (other details are provided in the Supplementary Information).

The sensor model learned using the ergodicity-based dataset results in a better reconstruction of the analysis-based measurement model (right, second from top) than the sensor model learned with randomly sampled data (right, second from bottom). Moreover, the data acquisition technique spends time in locations that cover regions of high information about voltage perturbations (bottom), so that the resulting predicted sensor output (bottom right) accurately captures the object's location and electrical signals experienced at a novel agent state.

We next consider the learning problem for high-dimensional vision sensors in environments with complex, nonparameterizable objects (e.g., objects with arbitrary shapes and visual textures) and multiple regions of interest (e.g., when there are multiple objects present). We construct a vision-based model of a three dimensional object for a 3-channel 75×75 pixel RGB camera attached to a robot arm in a simulated Pybullet[19] environment as shown in Fig. 3. The $75 \times 75 \times 3 = 16{,}875$-dimensional sensor output uses a CNN-CVAE (from ref. [20] and described in the Supplementary Information) for learning and encoding a 16-dimensional latent space representation. The simulated environment consists of a tabletop with a rubber duck on it—an object with complicated shapes and textures.

The random sampling generates samples over the entire workspace, while the ergodic sampling controller spends more time acquiring data near—but not on top of—the object in the environment, where the highest-value information in the environment is distributed (Fig. 3 and Supplementary Movie 1). Unlike the near-field electrosense example, where the robot needed to be very close to the object to acquire informative data, the ergodic sampling algorithm samples further away from the object for vision modeling. Exploration is not guided by any domain knowledge—the exploration behavior arises entirely based on the entropy of the unknown sensory model during the learning process and the maximally ergodic response with respect to that entropy.

The learned vision model of the object predicts the resulting camera image at every agent location within the domain, many of which were not in the training dataset generated during the data acquisition process. With ergodic sampling the learned model generates a higher quality reconstruction of the object in the camera image than the one generated from random sampling, reflecting the object's location accurately and capturing the complex shape and lighting of the object in the RGB camera model. Furthermore, the energy used to explore the environment and acquire data is lower over time for ergodic sampling than uniform random sampling, indicating that one can both improve energetic efficiency while also improving learning by collecting high-quality data.

Lastly, an example of modeling multiple objects using a simulated black and white camera is treated in detail in the Supplementary Information. There we see two important features of ergodic sampling with respect to the entropy of the learned model —that the exploratory trajectories inspect each object individually, without requiring a representation of the objects or their number, and that the resulting predictive model is able to discriminate between the objects based on the learning updates when random sampling-based learning cannot. Moreover, we use this example to illustrate the advantages of ergodic sampling over entropy-based information maximization. In the case of information maximization, the active learning strategy fixates on just one object, whereas the ergodic sampling strategy collects information about all the features about both objects, leading to substantially lower testing loss.

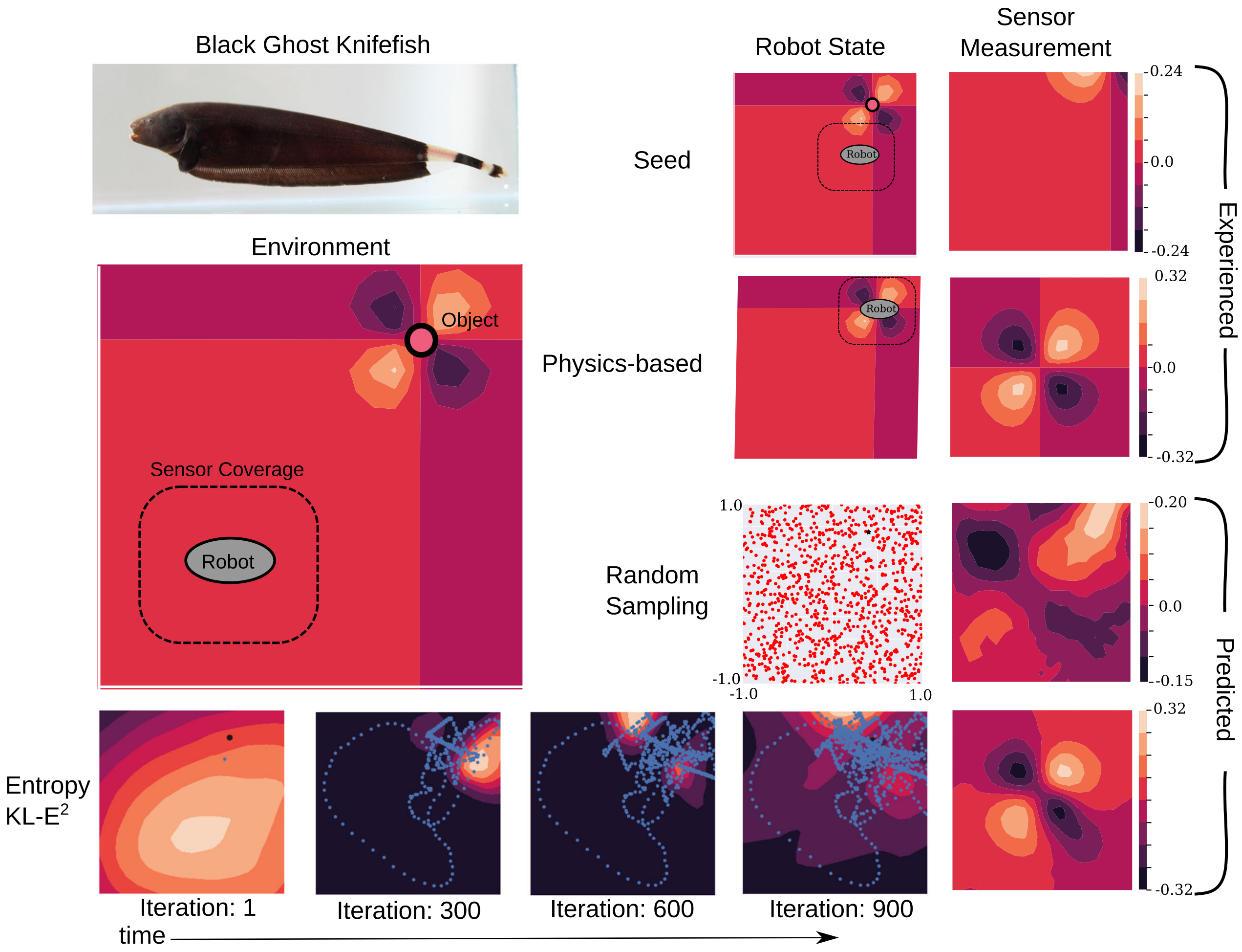

Fig. 2 Electrosense model learning. We compare the predicted measurement of the learned electrosensory model to the physics-based electrosensory model. The model is an approximation of the sensory system of a weakly electric fish (the black ghost knifefish[16]). Using a seed image, we show the reconstruction of the physics-based measurement field estimate near the object interest, using a conditional autoencoder with active learning using random sampling, and the proposed ergodic sampling strategy. The proposed approach explores the environment based on the evolution of the entropy of the latent space over time (shown evolving over time along with the resulting trajectory shown in blue). Despite no a priori information about the object's location (indicated with black markers) or measurement model, as time evolves the entropy of the latent space—representing where the information affecting the model estimate is highest—converges near the object location and the ergodic sampling spends significantly more time exploring that region without fixating on a particular state. As a result, the ergodic sampling approach results in a significantly better reconstruction of the predicted sensor observation with much lower energy expenditure needed during exploration.

Discussion

The use of learning-for-control dominates much of modern robotics and has its roots in reinforcement learning[21]. Here, we develop a theory of control-for-learning, based on a principle of maximal ergodicity with respect to entropy of a learning algorithm.

Sensor-environment model learning often relies on domain knowledge[22–26] or is hand tailored to specific domains[27–29]. Other unsupervised model learning approaches assume pre-defined knowledge of sensor structure[30] or assume large pre-existing datasets[4,5]. Oftentimes these environments are simulation-based—permitting sampling the state space without adhering to the laws of physics or even continuity of the state in time—so they do not examine the data acquisition process or optimization of the learning representation. Other work examines the quality of sensor data, but does so independently from the learning representation. For instance, Bayesian surprise in sensory data is used to identify data needed to guide human behavior[31–33] and to generate classifiers and event detection in sensors and robots[34–36], but these techniques focused on analyzing the value of existing data rather than guiding exploration and data acquisition. Such methods do not aim to provide approaches to actively acquire informative data in the physical world subject to physical constraints. Even reinforcement learning techniques treat data acquisition as an afterthought (discussed in the Supplementary Information).

This work proposes a domain-independent framework that integrates data acquisition and learning processes using ergodicity—familiar to most as a measure from statistical mechanics. We demonstrate that an ergodicity-based approach enables active acquisition of informative sensor data to efficiently learn a sensor model. The approach is amenable to high-dimensional sensor outputs and complex learning models in a variety of scenes and environments; here we demonstrate with both near-field and far-field sensor types with different physical properties. The ergodic sampling algorithm capable of exploring high-dimensional search spaces in real-time was developed in ref. [37]. Here, we further developed it for exploring with respect to entropy of learning models. Critically, we show how ergodicity with respect to the state of the neural network can be used to actively acquire

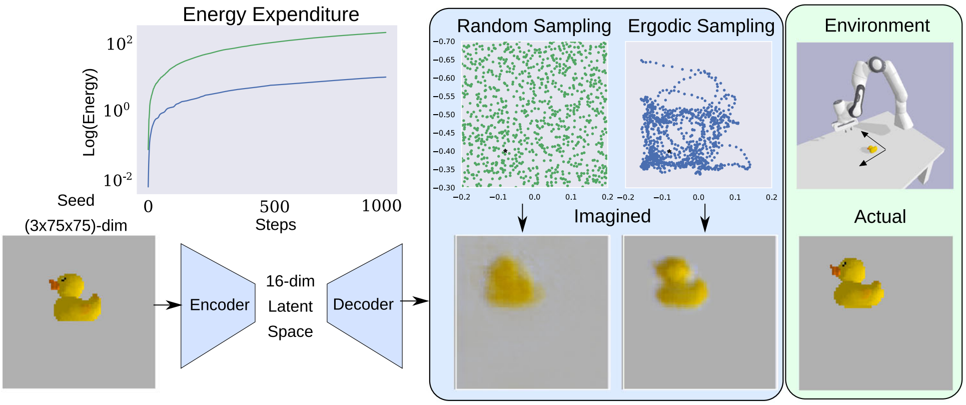

Fig. 3 RGB camera model learning. We compare the predicted measurement generated from the learned 3-channel RGB camera model. Using a seed image, we show the reconstruction of the visual camera field estimate near the object interest, using a CVAE with active learning using random sampling, and the proposed ergodic sampling strategy. Despite no a priori information about the object's location or measurement model, as time evolves, ergodic sampling spends significantly more time exploring the region around the object in the environment. However, unlike the near-field electrosensory model, the camera characteristics generates exploration further away from the object, where the object appears in the periphery of the camera view, instead of remaining directly over the object. The ergodic sampling approach (shown in blue) requires much lower energy expenditure compared to random sampling (shown in green). Furthermore, the ergodic sampling approach results in a significantly better reconstruction of the predicted sensor observation, better capturing the object's characteristics (i.e, shape, color) and the characteristics of the sensor-environment interactions (i.e., camera lighting). More detail can be found in the Supplementary Information.

informative data within an unsupervised learning framework, enabling active data acquisition that evolves along with the unsupervised model learning. The ergodic framework enables exploration without requiring physics-based knowledge of a sensor or the environment (but does assume ability to move the sensor). This work provides a framework for all sensor-environment learning that does not rely on physics-based models to learn and predict experiences, but rather uses experienced observations along with the ability to move in an environment to guide and inform model learning and future behavior.

Although this paper has focused on synthetic forms of mechanical motion that approximate ergodicity as a principle for learning, mechanical systems that are automatically ergodic as a result of their statistical mechanical properties would learn without needing this synthesis. This indicates how learning may be a consequence of mechanical coverage combined with computing updates, in some cases making learning as much a mechanical property[38] as an algorithmic property of a system.

Methods

In the following section, we describe the methods for learning a generative sensory model with no a priori information, human input, or sensor domain knowledge. We describe a Kullback–Leibler (KL)-ergodic measure, a sample-based approximation of the ergodic metric, and an optimal controller to enable ergodic exploration that extends to high-dimensional exploration spaces. Using this, we present an exploration strategy for acquiring data from regions of high information for efficient learning.

Assumptions. Assume we have an agent with system dynamics that can be modeled as

$$\dot{x}(t) = f(x(t), u(t)) \tag{1}$$

$$= g(x(t)) + h(x(t))u(t) \tag{2}$$

where $x \in X \subset \mathbb{R}^n$ is the state of the agent, $u \in \mathcal{U} \subset \mathbb{R}^m$ is a control vector, $f(x, u)$ is the (possibly nonlinear) transition model, which we split into $g(x) \subset \mathbb{R}^n \rightarrow \mathbb{R}^n$, the free (unactuated) dynamics, and $h(x) : \mathbb{R}^n \rightarrow \mathbb{R}^{n \times m}$, the input vector fields. It has a sensor that outputs data $y \in Y \subset \mathbb{R}^o$, which here comes from simulated sensors.

The agent explores a workspace $\mathcal{X}$ with no a priori knowledge of environmental physical characteristics (e.g. objects in the environment or textures).

Model learning. We use a conditional variational autoencoder (CVAE) to construct a generative sensor model of the sensing modality. We define the inputs to the CVAE encoder as the sensor data $y(t)$ and agent state $x(t)$. We model both the encoder and decoder as neural networks. The encoder network is a probability distribution $p_{\theta_e}(z|y, x)$ parameterized by θ_e that compresses the high-dimensional sensor measurement into a low-dimensional compressed latent space z. The decoder network is a probability distribution $p_{\theta_d}(y|z, x)$ parameterized by θ_d that has as its input z_x—generated by sampling the normally-distributed latent space z such that $z_x = z_\mu + z_\sigma \epsilon$, where $\epsilon \sim \text{Normal}(0, 1)$ and the agent state $x(t)$. The outputs from the decoder are the expected sensor readings $\tilde{y}(t)$ conditioned on the agent state $x(t)$.

Electrosense. The electrosense network's input is the 441-dimensional sensor output $y(t)$ and the 2-dimensional agent state $x(t) = [p_x, p_y]$, where p_x, p_y are the positions in space at time t. The encoder network compresses the input with a series of linear transformation and ReLU layers into a 2-dimensional latent space z and its log-variance $\log[\mathbb{V}(z|y, x)]$. The decoder network takes in samples of the reparametrized z latent space and the agent state x and generates a prediction of the 441-dimensional measurement $\tilde{y_n}$ and a scalar value representing the log-variance $\log[\mathbb{V}(y|z, x)]$ representing the uncertainty with a series of linear transformation and ReLU layers.

Vision. Intensity camera model: The camera model network's input is a 1-channel 38x38 intensity image (1444-dimensional) sensor output $y(t)$ and the 2-dimensional agent state $x(t) = [p_x, p_y]$. The encoder network compresses the input with a series of linear transformation and ReLU layers into a 16-dimensional latent space z and its log-variance $\log[\mathbb{V}(z|y, x)]$. The decoder network takes in samples of the reparametrized z latent space and the agent state x and generates a prediction of the 1444-dimensional image measurement $\tilde{y}(t)$ and its scalar log-variance $\log[\mathbb{V}(y|z, x)]$ with a series of linear transformation and ReLU layers.

RGB camera model: The CVAE for RGB camera model learning consists of a combination of convolutional neural networks and a conditional autoencoder networks (CNN-CVAE). For the encoder network, the 3-channel 75×75 image is first an input into a convolutional neural network (CNN) consisting of 2d convolutional layers and max pooling layers to condense the image into a lower-dimensional feature space. The flattened output of the CNN and the 2-dimensional agent state $x = [p_x, p_y]$ is then an input into a second neural network that compresses the input with a series of linear transformation and ReLU layers into a 16-dimensional latent space z and its log-variance $\log[\mathbb{V}(z|y, x)]$. The decoder network consists of two different decoder networks, an image-decoder network and variance-decoder network, that both take as input samples of the reparametrized z latent space and the agent state x. The image-decoder network generates a prediction of the image by first using a neural network with a series of linear transformation and ReLU layers and then a deconvolutional neural network to generate the 3-channel image prediction. The variance-decoder network takes in the samples of the reparametrized z latent space and the agent state x with a series of linear transformation and ReLU layers generates a scalar value representing the log-variance, or the uncertainty, of the predicted image.

Our objective is to develop a method that is general to any form of sensor model learning that requires an agent to dynamically move to seek out measurements through actions. In the next section, we derive a KL-ergodic measure and the derived optimal controller (KL-E^2) that specifies the active exploration task as an area-coverage objective, developed in[37,39].

Kullback–Leibler (KL)-ergodic measure. We define active exploration for informative data acquisition by generating actions that actively seek out informative data. The expected informativeness of a measurement at a particular state is defined to be the entropy of the learning model at that state. This is accomplished by specifying the active data acquisition task using an area coverage objective where we minimize the KL-divergence between the time average statistics of the agent along a trajectory and a spatial distribution defining the current coverage requirement. As the exact time-averaged statistics, as described in ref. [9], are a collection of delta functions parameterized by time, we define an approximation of the spatial statistics of the agent. To do so, we approximate the delta function as a Gaussian distribution with covariance Σ, converging to the delta function as $\|\Sigma\| \to 0$. We define the approximation to the spatial statistics of the agent as follows:

Definition 1 Given a search domain $\mathcal{X}^v \subset \mathbb{R}^{n+m}$ where $v \le n+m$, the Σ-approximated time-averaged statistics of the agent, i.e., the time the agent spends in regions of the search domain $\mathcal{X}^v$, is defined by

$$q(s|x(t)) = \frac{1}{T_r}\int_{t_i-t_r}^{t_i+T} \mu(s \parallel \bar{x}(t))dt \tag{3}$$

$$= \frac{1}{T_r}\int_{t_i-t_r}^{t_i+T} \frac{1}{\eta}\exp\left[-\frac{1}{2}\parallel (s-\bar{x}(t)\|^2_{\Sigma^{-1}}\right]dt \tag{4}$$

where $s \in \mathcal{X}^v \subset \mathbb{R}^{n+m}$ is a point in the search domain $\mathcal{X}^v$, $\bar{x}(t)$ is the component of the agent's state $x(t)$ that intersects the search domain $\mathcal{X}^v$, $\Sigma \in \mathbb{R}^{v\times v}$ is a positive definite matrix that specifies the width of the Gaussian, η is a normalization factor, t_i is the i^{th} sampling time, and $T_r = T + t_r$ is sum of the time horizon T and amount of time t_r the agent remembers $x_v(t)$ into the past.

Using this approximation, we are able to approximate the ergodic area-coverage objective in[9] using the following KL-divergence objective[40]:

$$D_{\text{KL}}(p \parallel q) = \int_{\mathcal{X}^v} p(s)\ln\frac{p(s)}{q(s)}ds \tag{5}$$

$$= \underbrace{\int_{\mathcal{X}^v} p(s)\ln p(s)ds}_{\text{does not depend on } x(t)} - \underbrace{\int_{\mathcal{X}^v} p(s)\ln q(s)ds}_{\text{depends on } x(t)} \tag{6}$$

$$= -\int_{\mathcal{X}^v} p(s)\ln q(s)ds \tag{7}$$

$$= -\int_{\mathcal{X}^v} p(s)\ln q(s)ds \tag{8}$$

$$= -\mathbb{E}_{p(s)}\left[\ln q(s)\right], \tag{9}$$

where $\mathbb{E}$ is the expectation operator, $q(s) = q(s|x(t))$, and $p(s)$, $p(s) > 0, \int_{\mathcal{X}^v} p(s)ds = 1$, is a distribution that describes where in the search domain an informative measurement is likely to be acquired. Note that we drop the first term in the expanded KL-divergence because it does not depend on the trajectory of the agent $x(t)$.

We approximate the expectation operator using

$$D_{\text{KL}}(p \parallel q) = -\mathbb{E}_{p(s)}\ln q(s) \tag{10}$$

$$\approx -\sum_{i=1}^{N} p(s_i)\ln q(s_i) \tag{11}$$

$$\approx -\sum_{i=1}^{N} p(s_i)\ln\int_{t_0}^{t_f} \exp\left[-\frac{1}{2}\parallel (s-\bar{x}(t)\|^2_{\Sigma^{-1}}\right]dt \tag{12}$$

where N is the number of samples in the search domain drawn from a uniform distribution. With this formulation, we are able to obtain the benefits of indirectly sampling from the spatial distribution $p(s)$ without having to directly compute derivatives to generate an optimal control signal. Furthermore, this sample-based measure prevents the computation from scaling exponentially with the number of exploration states.

KL-ergodic exploration (KL-E^2). We now synthesize optimal control using the derivative of the measure in Eq. (10) with respect to the duration time λ of control $\mu_*(t)$, or the mode insertion gradient, which calculates the sensitivity of the measure with respect to duration.

The first order sensitivity of (10) with respect to the control duration λ of the applied control $\mu_\star(\tau)$ is

$$\frac{\partial J}{\partial \lambda}|_{t=\tau} = \rho(\tau)^\top(f_2 - f_1) \tag{13}$$

where $f_2 = f(x(t), \mu_\star(t))$ and $f_1 = g(x(t))$ (or the unactuated dynamics), and $\rho(t) \in \mathbb{R}^n$ is the adjoint, or co-state variable is the solution to the differential equation:

$$\dot{\rho} = -\left(-\frac{\eta}{T_r}\sum_i\frac{p(s_i)}{q(s_i)}\left(\frac{\partial\mu}{\partial x}\right)\right) - \left(\frac{\partial f}{\partial x}\right)^\top\rho. \tag{14}$$

Given the mode insertion gradient, we need to find the control $\mu_\star(t)$ that most significantly decreases (13) by writing it as an unconstrained optimization problem:

$$J_2 = \int_{t_i}^{t_i+T}\frac{\partial}{\partial\lambda}J|_{t=\tau} + \frac{1}{2}\parallel \mu_\star(t))\|^2_R \tag{15}$$

where $R \in \mathbb{R}^{m\times m}$ is a positive definite matrix that penalizes the deviation from the policy $\mu(x)$. The control vector that minimizes J_2 is given by

$$\mu_\star(t) = -R^{-1}h(x(t))^\top\rho(t). \tag{16}$$

Supplementary Fig. 4 shows the results of the KL-E^2 controller exploring a target distribution. The controller generates a trajectory that ergodically explores with respect to the target distribution (i.e., it explores proportionally with respect to the target distribution). However, the Fourier-based ergodic metric, which scales exponentially $\mathbb{O}(n * |k|^n)$, where $|k|$ is the maximum-integer Fourier term and n is the dimensions of the exploration space. On the other hand, our proposed approach scales linearly $\mathbb{O}(|k| * n)$, where $|k|$ is the number of samples, making it more computationally efficient for real-time, high-dimensional exploration. Given this control formulation, all that is needed is to define a target spatial distribution that specifies which measurements are more informative for estimating the sensor model.

Importance measure for sensor model estimation. To generate a target coverage distribution for the agent, we construct a measure that allows the agent to quantify where in the search domain will be useful data to best improve its learning of the sensor model. Here, we use the CVAE network to generate an estimate of data importance over the search space.

Definition 2 The importance measure $\nu : \mathcal{X} \times \mathcal{X}^v \to \mathbb{R}^+$ for a new state s in the search domain when the agent is at its current state x is given by

$$\nu(s,x) = \exp\left(\mathbb{H}(p_{\theta_d(y|z_x,s)})\right)^k \tag{17}$$

where $\mathbb{H}$ is the entropy of the decoder network p_{θ_d} conditioned on a given robot state s and the latent space representation z_x generated from the robot's current state x and k is an effective temperature parameter for the distribution that exponentially weighs regions of high importance.

As we assume a normal distribution, the entropy does not depend on the mean of $p_\phi(y|z,x)$[11]. Since the CVAE converts the inputs into a statistical latent space, we use the estimate of the decoder network entropy as a function of state to represent model uncertainty over the workspace—generating an expected information density. We use this expected information density distribution to specify ergodic coverage, such that the robot prioritizes regions with the most informative data to best improve the latent space representation.

Definition 3 The importance distribution $\nu \in \mathbb{R}^+$ is defined as the entropy of the normal distribution,

$$\nu(s,x) \propto \exp\left(\frac{1}{2}\ln[\sigma_y^2(s,x)]\right)^k \tag{18}$$

where points $s, x \in \mathcal{X}^v$ and the robot's current state is x.

Algorithm overview. We provide an outline of our KL-E^2 method in Algorithm 1 for real-time active learning for sensor model estimation.

Algorithm 1. KL-Ergodic exploration (KL-E^2) for variational autoencoders

```
initialize: local dynamics model $f(x,u)$, initial condition $x(t_0)$, CVAE
network $C$, Learned parameters $\theta, \phi$ of network $C$, importance distribution
$p(s,x)$, number of CVAE optimizations $j_{opt}$, batch size $b$.
for $i = 0, \ldots, \infty$ do
    Calculate optimal control $\mu_\star(\tau)$ for $t \in [t_i, t_i + T]$
    Calculate application time and duration $\tau, \lambda$ that minimizes $\frac{\partial}{\partial\lambda}J$
    Apply $\mu_\star(\tau)$ if $t \in [\tau, \tau+\lambda]$ else apply $\mu(x(t))$ to robot
    Sample state $x(t_i)$ and measurement $y(t_i)$
    for $j = 0, \ldots, j_{\text{opt}}$ do
        Generate minibatch $M$ by randomly sampling $b$ data points from full set
        Generate random noise $\epsilon \sim p(\epsilon)$ for each data point in $M$
        Compute loss $L_{\theta,\phi}$ and its gradients $\nabla_{\theta,\phi}L_{\theta,\phi}$
        Update $\theta$ and $\phi$ using SGD optimizer
    end for
    Generate latent space representation $z_x$ at current robot state $x_i$
    Update importance distribution $p(s,x)$
end for
```

Data availability

The data for all experiments and figures in the paper are publicly available. A downloadable version of the datasets used for the different sensory simulations in this paper is archived and links for them can be found on the github reposity https://github.com/apr600/mechanical-intelligence.

Code availability

Implementations of the active sensory learning algorithm with the different environment examples are available at https://github.com/apr600/mechanical-intelligence. This package also contains code reproducing the main results from the paper.

Received: 25 June 2021; Accepted: 4 July 2022;
Published online: 15 July 2022

References

1. Quiroga, R. Q., Reddy, L., Kreiman, G., Koch, C. & Fried, I. Invariant visual representation by single neurons in the human brain. *Nature* **435**, 1102–1107 (2005).
2. Park, S., Intraub, H., Yi, D.-J., Widders, D. & Chun, M. M. Beyond the edges of a view: boundary extension in human scene-selective visual cortex. *Neuron* **54**, 335–342 (2007).
3. Wolbers, T., Klatzky, R., Loomis, J., Wutte, M. & Giudice, N. Modality-independent coding of spatial layout in the human brain. *Curr. Biol.* **21**, 984 – 989 (2011).
4. Ha, D. & Schmidhuber, J. Recurrent world models facilitate policy evolution. In *Advances in Neural Information Processing Systems*, **31** 2451–2463 (NeurIPS, 2018).
5. Eslami, S. A. et al. Neural scene representation and rendering. *Science* **360**, 1204–1210 (2018).
6. Rahwan, I. et al. Machine behaviour. *Nature* **568**, 477–486 (2019).
7. Vergassola, M., Villermaux, E. & Shraiman, B. I. "Infotaxis" as a strategy for searching without gradients. *Nature* **445**, 406–409 (2007).
8. Chen, C., Murphey, T. D. & MacIver, M. A. Tuning movement for sensing in an uncertain world. *eLife* **9**, e52371 (2020).
9. Miller, L. M., Silverman, Y., MacIver, M. A. & Murphey, T. D. Ergodic exploration of distributed information. *IEEE Trans. Robot.* **32**, 36–52 (2016).
10. Sethna, J. *Statistical Mechanics: Entropy, Order Parameters, and Complexity* (Oxford University Press, 2021).
11. Bishop, C. M. *Pattern Recognition and Machine Learning* (Springer, 2006).
12. Mathew, G. & Mezić, I. Metrics for ergodicity and design of ergodic dynamics for multi-agent systems. *Phys. D: Nonlinear Phenom.* **240**, 432–442 (2011).
13. Srinivasan, M. & Ruina, A. Computer optimization of a minimal biped model discovers walking and running. *Nature* **439**, 72–75 (2006).
14. Collins, S., Ruina, A., Tedrake, R. & Wisse, M. Efficient bipedal robots based on passive-dynamic walkers. *Science* **307**, 1082–1085 (2005).
15. Pathak, D., Agrawal, P., Efros, A. A. & Darrell, T. Curiosity-driven exploration by self-supervised prediction. In *International Conference on Machine Learning*. 2778–2787 (ACM, 2017).
16. Nelson, M. E. & Maciver, M. A. Prey capture in the weakly electric fish apteronotus albifrons: sensory acquisition strategies and electrosensory consequences. *J. Exp. Biol.* **202**, 1195–1203 (1999).
17. Nelson, M. E., MacIver, M. A. & Coombs, S. Modeling electrosensory and mechanosensory images during the predatory behavior of weakly electric fish. *Brain Behav. Evol.* **59**, 199–210 (2002).
18. Bai, Y., Snyder, J. B., Peshkin, M. & MacIver, M. A. Finding and identifying simple objects underwater with active electrosense. *Int. J. Robot. Res.* **34**, 1255–1277 (2015).
19. Coumans, E. & Bai, Y. *Pybullet, A Python Module for Physics Simulation for Games, Robotics and Machine Learning.* http://pybullet.org (2016–2021).
20. Sohn, K., Lee, H. & Yan, X. Learning structured output representation using deep conditional generative models. In *Advances in Neural Information Processing Systems*, **28** 3483–3491 (NIPS, 2015).
21. Sutton, R. S. & Barto, A. G. *Reinforcement learning: An introduction* (MIT press, 2018).
22. Mottaghi, R., Rastegari, M., Gupta, A. & Farhadi, A. "What happens if..." learning to predict the effect of forces in images. In *European Conference on Computer Vision*, 269–285 (Springer, 2016).
23. Mottaghi, R., Bagherinezhad, H., Rastegari, M. & Farhadi, A. Newtonian scene understanding: Unfolding the dynamics of objects in static images. In *IEEE Conference on Computer Vision and Pattern Recognition*, 3521–3529 (IEEE, 2016).
24. Bhat, K. S., Seitz, S. M., Popović, J. & Khosla, P. K. Computing the physical parameters of rigid-body motion from video. In *European Conference on Computer Vision*, 551–565 (Springer, 2002).
25. Wu, J., Lim, J. J., Zhang, H., Tenenbaum, J. B. & Freeman, W. T. Physics 101: Learning physical object properties from unlabeled videos. In *Proceedings of the British Machine Vision Conference (BMVC)*, vol. 2, 7 (BMVA, 2016).
26. Brubaker, M. A., Sigal, L. & Fleet, D. J. Estimating contact dynamics. In *IEEE International Conference on Computer Vision*, 2389–2396 (IEEE, 2009).
27. Fragkiadaki, K., Agrawal, P., Levine, S. & Malik, J. Learning visual predictive models of physics for playing billiards. *International Conference on Learning Representations (ICLR), poster* (2015).
28. Stewart, R. & Ermon, S. Label-free supervision of neural networks with physics and domain knowledge. In *Thirty-First AAAI Conference on Artificial Intelligence* (AAAI, 2017).
29. Watters, N. et al. Visual interaction networks: Learning a physics simulator from video. In *Advances in Neural Information Processing Systems 30*, 4539–4547 (MLR, 2017).
30. Trinh, H. & McAllester, D. Unsupervised learning of stereo vision with monocular cues. In *Proc. of the British Machine Vision Conference (BMVC)*, vol. 4 (BMVA, 2009).
31. Itti, L. & Baldi, P. Bayesian surprise attracts human attention. *Vis. Res.* **49**, 1295–1306 (2009).
32. Itti, L. & Baldi, P. F. Bayesian surprise attracts human attention. In *Advances in Neural Information Processing Systems*, 547–554 (MLR, 2006).
33. Ostwald, D. et al. Evidence for neural encoding of Bayesian surprise in human somatosensation. *NeuroImage* **62**, 177–188 (2012).
34. Itti, L. & Baldi, P. A principled approach to detecting surprising events in video. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 631–637 (IEEE, 2005).
35. Schauerte, B. & Stiefelhagen, R. "Wow!" Bayesian surprise for salient acoustic event detection. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, 6402–6406 (IEEE, 2013).
36. Ranganathan, A. & Dellaert, F. Bayesian surprise and landmark detection. In *IEEE International Conference on Robotics and Automation*, 2017–2023 (IEEE, 2009).
37. Abraham, I., Prabhakar, A. & Murphey, T. D. An ergodic measure for active learning from equilibrium. *IEEE Transactions on Automation Science and Engineering* 1–15 (IEEE, 2021).
38. Yasuda, H. et al. Mechanical computing. *Nature* **598**, 39–48 (2021).
39. Abraham, I., Prabhakar, A. & Murphey, T. D. Active area coverage from equilibrium. *Workshop on the Algorithmic Foundations (WAFR)* (Mérida, Mexico, 2018).
40. Ayvali, E., Salman, H. & Choset, H. Ergodic coverage in constrained environments using stochastic trajectory optimization. In *International Conference on Intelligent Robots and Systems*, 5204–5210 (IEEE, 2017).

Acknowledgements

We thank Malcolm Maciver for valuable discussions on this work. We thank Allison Pinosky, Jake Ketchum, and Ian Abraham for comments on the manuscript and accompanying code. This material is based on work supported by the United States National Science Foundation grant CNS 1837515 (A.P., T.M.) and the United States Office of Naval Research grant N00014-21-1-2706 (T.M.). The views, opinions, and/or findings expressed are those of the authors, and should not be interpreted as representing the views or policies of the agency.

Author contributions

A.P. and T.M. conceived and designed research; A.P. developed simulations and drafted manuscript; A.P. and T.M. edited and revised manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-022-31795-2.

Correspondence and requests for materials should be addressed to Ahalya Prabhakar or Todd Murphey.

Peer review information *Nature Communications* thanks Dalin Zhang and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2022