2023 IEEE International Conference on Systems, Man, and Cybernetics (SMC)
October 1-4, 2023, Oahu, Hawaii, USA

Robotic Information Gathering via Deep Generative Inpainting

Tamim Khatib, O. Patrick Kreidl, Ayan Dutta, Ladislau Bölöni, Swapnoneel Roy

***Abstract*— In today's era of automation, mobile robots are being used for collecting meaningful information about an ambient phenomenon such as temperature or moisture distribution in an agricultural field. Most of the studies in the literature assume that the underlying information field is Gaussian, and therefore, Gaussian Process (GP)-based models are extremely popular. Furthermore, we have found that due to the inherent computational complexity of such naive GP-based techniques, most studies in the literature do not scale well beyond small-size environments, i.e., where the number of informative points $n < 1000$. These render such a predictive model more or less useless in many practical applications. In this paper, we posit that a different technique, Generative Adversarial Network-based inpainting, for robotic information gathering can be useful. The state-of-art inpainting techniques 1) do not assume that the underlying data is Gaussian, and 2) easily scale to $n \gg 1000$. Thus, they eliminate the two bottlenecks posed by the GP-based solutions. We have tested our hypothesis on a synthetic and a real-world crop dataset. Results show that while the inpainting technique easily scales to 1024×1024, GP-based predictions cannot. On the other hand, their solution qualities are shown to be comparable.**

I. Introduction

The reduced cost and increasing sensing capabilities of drones brought robotic information gathering, once the domain of military and law enforcement applications, to a much wider audience. Farmers can use them to survey fields for plant diseases or invasive species, contractors can survey roofs to generate construction or repair cost estimates and homeowners on vacation can use them to monitor their premises. All these applications can be seen as instances of the information gathering problem, where a mobile robot takes observations from different spatial locations to be used to create an information model of the phenomena of interest. The quality of the final model depends on two factors: the *path* of the robot, which determines what observations are made, and the *inference process* that determines how the information model is built from the observations.

Traditionally, the focus of research in this field was on path planning, which trades cost factors (e.g. time, fuel, number of robots) against the quantity and quality of observations [1], [2], [3], [4]. If the user can afford enough resources to cover the whole area of interest with observations, the inference step becomes a straightforward step of stitching observations together. However, in many applications, aiming for complete coverage is impractical. This is especially the case for civilian or personal applications, where resources are at a premium. In these cases, the inference step needs to estimate the missing information based on the available observations. Missing information may also result from unexpected sensor malfunctions or environmental disturbances that result in outliers during post-processing. Fig. 1 illustrates the impact of incomplete coverage or identified outliers in the context of grayscale imagery over a farm field[1].

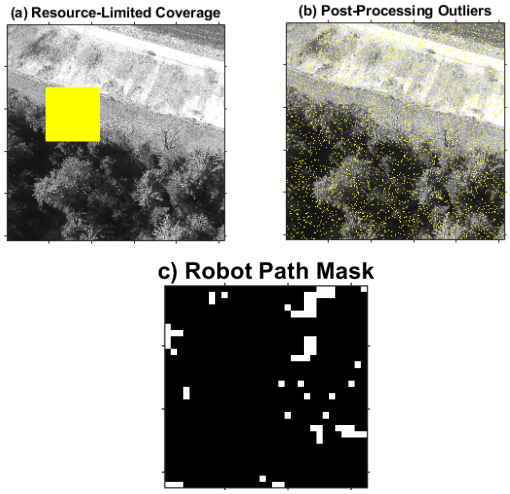

Fig. 1. The impact of missing information (rendered in yellow over an otherwise grayscale image) within an agricultural imagery collection scenario: (a) a block set of missing values due to limited collection resource relative to the area-of-interest or (b) a speckled set of missing values due to spurious environmental disturbances or sensor malfunctions. (c) An example of a robot *path mask* where the robot has visited and sensed information from the black locations only whereas the sensor measurements in the white cells need to be inferred.

One of the most popular models for the inference step in this problem is based on the idea of Gaussian inference. The information to be acquired is represented as a scalar field covering the geographical area, the observations are distinct measurements made at particular locations. An assumption of a normal distribution of the uncertainty at unobserved locations allows the inference to be performed through a Gaussian Process (GP) – a standard approach used in most recent research projects [1], [3], [5], [6], [7], [8], due to several desirable attributes. Under certain assumptions, a rigorous theoretical foundation supports that the model is an optimal representation [9]. Furthermore, as part of the calculations already done, the model returns an estimate of the uncertainty at each point [10]. Finally, while the algorithm is complex, high-quality publicly available implementations

T. Khatib, O.P. Kreidl, A. Dutta, and S. Roy are with the University of North Florida, USA. Emails: {n01379084, patrick.kreidl, a.dutta, s.roy}@unf.edu
L. Bölöni is with the University of Central Florida, USA. Email: {ladislau.boloni}@ucf.edu
This work is supported in part by NSF CPS Grants #1932300 and #1931767.

[1]Original image taken from https://ageagle.com/resources/?filter_by=data-set

2023 IEEE International Conference on Systems, Man, and Cybernetics (SMC) | 979-8-3503-3702-0/23/$31.00 ©2023 IEEE | DOI: 10.1109/SMC53992.2023.10394444

exist that can be readily deployed.

As the deployment of robotic information gathering broadens its users, GP-based approaches are confronted with scalability problems. The $O(n^3)$ computational complexity of the canonical GP problem makes it very difficult to scale beyond several hundred observations and query points on commonly accessible hardware. For instance, in a precision agriculture application, the area of interest can be millions of square feet for a typical farm. This motivates a challenging research direction of high practical importance to find inference methods that approximate the performance of Gaussian inference but have much better scaling properties.

The approach we take in this paper is to reverse the way in which we are usually thinking about the inference problem. Instead of inferring a map out of discrete observations, we consider that the observations represent the final map, which, however, has some missing parts. The goal of the inference is thus to infer the missing parts of the map. This reformulation allows us to take advantage of the significant recent progress made in the computer vision community. The general concept of *image inpainting* refers to a technique to fill in missing information from an image which allows the removal of scratches or editing of distracting objects by replacing them with the inferred background [11]. The objectives of image inpainting are to ensure that the filled-in area is semantically correct and visually realistic. Both of these desiderata can be formalized as a probability distribution of the output. In contrast to Gaussian inference, the probability models are significantly more complex and are not suitable for closed-form expressions. In recent years, however, the use of deep neural networks as representations of the distribution and the use of training techniques such as generative adversarial networks and variational autoencoders provided efficient solutions to the image inpainting problem [12], [13].

In this paper, we adapt DeepFill (v2) [14], a recent and well-established image inpainting model, to the task of data model inference in robotic information gathering. The application domain we are considering involves information collection in precision agriculture where the size of the areas involved makes the use of external models difficult to use. We test our models both on real-world as well as synthetic data and compare them with Gaussian Process regression in terms of both scalability and model quality.

The primary contributions of this paper are as follows:

1) We propose a novel technique of inference in the robotic information gathering problem based on techniques inspired by image inpainting literature.
2) We show that the proposed inference model provides comparative solution quality but scales significantly better than a GP regression technique, or the standard approach to inference in robotic information gathering.

II. Related Work

The task of information gathering asks a robot to go to k locations in an unknown environment and measure information using onboard sensing capabilities such as a camera or a humidity sensor. The sensor measurements in the unvisited locations are predicted based on the measurements in the visited locations [5]. In the offline version of the problem, the hyper-parameters of the underlying ambient phenomena are assumed to be known [2], [10]. Machine learning models can be deployed to learn the parameters in such offline setups and the robot can be deployed to follow the best path found by the offline planner [8], [15]. Such offline assumptions do not usually hold in a real-world information gathering application such as precision agriculture [1] and ocean monitoring [3]. On the other hand, the online variant does not involve such a strong assumption, and the parameters of the process are learned and tuned with the sampling of new unseen information [16], [17]. Gaussian process (GP) [18] is the most commonly used technique to model the information field [10], [3], [16], [5], [17]. Primarily, two main performance metrics are used to evaluate the quality of the paths: Entropy [16] and Mutual Information (MI) [10]. Singh et al. have shown that the problem of optimal information gathering is NP-hard [5], and therefore, greedy (or, myopic) strategies have been popularly deployed sometimes with provable worst-case performance guarantees [19], [10]. Non-myopic solutions plan the paths for longer horizons using techniques such as dynamic programming [20], [21], [22]. Due to the use of compute-intensive calculations including matrix inversions in GPs, the researchers in the literature have mostly used small-scale environments for testing their proposed algorithms. Table I lists some of the recent papers in this domain and the corresponding number of locations (nodes) their test environments have. However, we would like to point out that there is a sub-field in machine learning that is dedicated to finding efficient solutions using sparse GPs unlike the classic GP models employed in robotics [23], [24], [25]. It is worth mentioning that the most noted work in the field of information gathering that scales beyond 1000 locations is due to Solin et al. [7], who have used a reduced-rank GP model to achieve this.

Reference	Environment Size (# of nodes)
Solin et al. [7]	9261
Viseras, Shutin, and Merino [17]	400
Viseras, Shutin, and Merino [4]	400
Ma, Liu, and Sukhatme [3]	144
Ma, Liu, and Sukhatme [26]	100
Wei and Zheng [8]	60
Our paper (Synthetic data)	1024
Our paper (Real-world data)	1,048,576

TABLE I

RECENT STUDIES ON GP-BASED INFORMATION SAMPLING AND THE ENVIRONMENT SIZES USED IN THEM.

We argue that a different approach can be taken to solve the information gathering without relying on a careful design of sophisticated sparse-GP techniques. Specifically, we investigate a technique called inpainting [11], which is most prominent in the field of image processing. The goal here is to reconstruct missing parts of an image [27]. Recently, due to the advancements in deep learning, image inpainting

has received significant attention, and a plethora of new techniques are now available in the literature [12], [28], [29], [13], [14] that are based on Generative Adversarial Networks [30]. Recently, Variational Auto-Encoder (VAE) networks [31] has been used by Shrestha et al. [32] for map configuration prediction. Li et al. [33] have used a deep learning schema using CNN and RNN for active information sensing and inference. The closest study to our presented work is due to Caley and Hollinger [34]. The main difference with their work is that they need careful network design unlike ours. In this paper, we use a state-of-art inpainting technique [14] for *predicting* the missing sensor measurements in the unvisited locations and compare the performance against a traditional GP-based solution from the perspectives of solution quality and prediction time.

III. Approach

This section discusses the two primary approaches to robotic information gathering that will be compared experimentally in Section IV. The first approach described, referred to as *DeepFill*, leverages an inpainting technique from Google based on Generative Adversarial Networks (GANs) for image processing. To our knowledge, this technique has not before been applied in the robotic information gathering context. The second approach described, referred to as *GaussFill*, leverages canonical inference/learning tools within MATLAB based on Gaussian Processes (GPs), which have become a standard for the robotic information gathering problem. DeepFill promises advantages in two main ways: scalability and applicability.[2] DeepFill's scalability advantage comes by virtue of inpainting being fundamentally an image processing problem, where typical image sizes have millions of pixels, which correspond to collection locations in the navigation environment. DeepFill's applicability advantage arises from the extent to which uncertainties in real-world robotic environments only questionably befit Gaussian Process assumptions e.g., in high-resolution agricultural imagery as shown in Fig.1 via low-altitude drones.

A. The DeepFill Image Inpainting Technique

Early approaches to image inpainting aimed to expand the current texture of the image in a local neighborhood. While this technique can be used successfully to fill in small areas such as scratches on an old photograph or noisy pixels, it fails when larger areas in the image need to be filled in, where the technique cannot rely on neighboring or nearby pixels [11], [40], [41].

More recent approaches consider global information in the image and are suitable for filling in larger areas. These techniques usually rely on generative models which, given a condition expressed in the form of an incomplete image, generate a complete image drawn from a plausible distribution. Many of these techniques are based on the general concept of a generative adversarial network (GAN). GANs train simultaneously two components, the generator that creates an image starting from a random seed while taking into account the specific condition and a discriminator that is trained to distinguish between the "fake", that is, generated images, and the "real" images taken from the training set. In this application, only the generator component is retained after the training. The competition between the increasingly better discriminator and the generator forces the generator to increasingly better approximate the probability distribution of the training data [30].

The application of image inpainting to the robot information gathering problem is not immediate. While we are interested in a semantically correct in-fill, visual realism is usually not a requirement, thus global reasoning is more important than retaining local texture. Given the complex trajectory of robotic exploration, the missing areas might be relatively large patches with unpredictable shapes. For instance, in the case of generating missing information from a map, the system needs to handle both map cells for which nearby cells have valid observations, cells where the neighbors have no observations, and cells at the corners of the map where the neighboring cells are not defined.

One of the recent in-fill algorithms which aims to solve in the image processing domain problems similar to the ones faced in the robotic exploration is the DeepFill v2 [14] model. The technique uses a model called gated convolutions, which learns the contributions of the various types of nearby pixels alongside the different features for each spatial location. To handle the missing areas of arbitrary shape, the system uses a variant of GANs called Spectral-Normalized Markovian Discriminator (SN-PatchGAN), which generalizes the concept of local and global GANs. For more details, the readers are referred to [14].

B. The GaussFill Information Modeling Technique

Gaussian Process (GP) models of environmental uncertainties [42], [10], [9] assume that all the collection locations generate information according to Gaussian random vector $\mathbf{X}$ with known (prior) mean vector μ and covariance matrix Σ. Typical navigation planning constraints do not permit full coverage of the environment, decomposing the set of all collection locations into two disjoint subsets, U and V, corresponding to the unvisited and visited locations, respectively. Under negligible sensor noise assumptions, the Gaussian random vector $\mathbf{X}_U$ characterizing the uncollected information has (posterior) mean vector and covariance matrix given by equations

$$\begin{aligned} \mu_{U|\mathbf{X}_V} &= \mu_U + \Sigma_{UV}\Sigma_{VV}^{-1}(\mathbf{x}_V - \mu_V) \\ \Sigma_{UU|\mathbf{X}_V} &= \Sigma_{UU} - \Sigma_{UV}\Sigma_{VV}^{-1}\Sigma_{VU} \end{aligned} \tag{1}$$

with $\mathbf{X}_V = \mathbf{x}_V$ denoting the values measured in the visited locations and the prior statistics organized into the corresponding block forms with respect to U and V i.e.,

$$\mu = \begin{bmatrix} \mu_U \\ \mu_V \end{bmatrix} \quad \text{and} \quad \Sigma = \begin{bmatrix} \Sigma_{UU} & \Sigma_{UV} \\ \Sigma_{VU} & \Sigma_{VV} \end{bmatrix}.$$

[2]It should be noted that there is, albeit beyond this paper's scope, a rich literature on large-scale approximate inference/learning (e.g., [35], [36], [37], [38], [39]) that may challenge the stated advantages of DeepFill.

For brevity, we omit the analogous equations in the case of non-negligible sensor noise, which given any set of measurements $\mathbf{Y} = \mathbf{y}$ renders posterior statistics $\mu_{\mathbf{X}|\mathbf{Y}}$ and $\Sigma_{\mathbf{X}|\mathbf{Y}}$ still involving all collection locations. In any case, the resulting posterior mean is known to be the minimum-mean-square-error predictor of the process given the measurements i.e., in the context of (1), it is known that

$$\mu_{U|\mathbf{X}_V} = \arg \min_{\hat{x} \in \mathbb{R}^{|U|}} \mathbf{E}\left[||\mathbf{X}_U - \hat{x}||^2 \mid \mathbf{X}_V = \mathbf{x}_V \right]$$

with $|\cdot|$ denoting set cardinality, $\mathbf{E}[\cdot]$ denoting the expectation operator, and $||\cdot||$ denoting the vector 2-norm and, thus, $||\cdot||^2$ is the sum over the squared elements of its vector argument.

In practice, the principal challenge of Gaussian prediction is to obtain accurate prior statistics for the environmental information (and accurate noise statistics when sensing errors are non-negligible). Such priors are typically derived from training data via statistical learning methods (e.g., maximum-likelihood [9]) and, for spatially-distributed Gaussian processes, usually also leverage domain-specific environmental considerations. A length-p Gaussian process has $d = 2p + p(p-1)/2$ degrees-of-freedom, in general, where requirements that the number of training samples $n \gg d$ are often formidable in robotics applications. In such situations, it is common to assume a reduced-order structure for the Gaussian process. For example, the so-called "homogeneous and isotropic Gaussian Markov random field using an exponential kernel" defines the covariance matrix using just two hyper-parameters: given any pair of locations i and j at spatial positions $\mathbf{q}_i$ and $\mathbf{q}_j$, respectively, the kernel function is given by

$$\sigma_{ij}^2 = \beta^2 \exp\left(-||\mathbf{q}_i - \mathbf{q}_j||/\ell\right)$$

where $\beta > 0$ is the local standard deviation and $\ell > 0$ is the exponential rate of diminishing covariance between increasingly-distant locations. Observe that such a process, when $\beta = 0$, will deterministically render only the mean field, whose p degrees-of-freedom can analogously exhibit a reduced-order structure. Fig 2 illustrates a realization from such a structured length-1024 Gaussian process, rendering the information field over spatial positions on a 32x32 uniform grid within the two-dimensional, unit-square region. Such Gaussian realizations become compatible with (8-bit) grayscale image processing conventions by linearly scaling the surface heights to span the interval [0, 255] and then quantize. Indeed, Fig. 2 is an example of the synthetic data considered in our experiments to be described next.

IV. Experiments

This section describes the experiments through which we explore various information inference techniques described in Section III. These experiments involve two different sets of imagery in PNG format, which we label "Synthetic" and "Cropland," as well as three different sets of randomly selected missing data values, or *masks*, which we label "Block", "Speckled", and "Path". Fig. 1 has already motivated the consideration of these masks, illustrating their

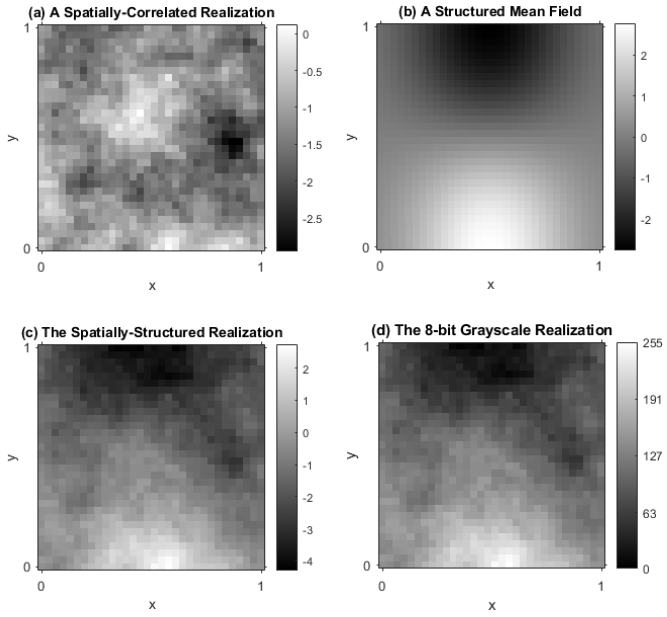

Fig. 2. (a) A length-1024 realization from a zero-mean Gaussian process (via MATLAB's `sqrtm` and `randn` functions), mapped onto locations over a 32x32 uniform grid in the unit square, with a spatially-correlated covariance structure induced by exponential kernel parameters $\beta = \ell = 1$. (b) A mean-field that within the spatial region is proportional to the sum of an upper "valley" (with a peak depth of -5) and a lower "hill" (with a peak height of +5), each a Gaussian-shaped surface centered by its length-2 mean vector (taking values $[0.50; 0.75]$ and $[0.50; 0.25]$, respectively) with elliptical contours defined by its 2-by-2 diagonal covariance matrix (taking values 0.0625 and 0.2500 in x and y, respectively). (c) The addition of the spatially-correlated realization in (a) and the structured mean field in (b). (d) The identical spatial pattern in (c) after linearly scaling the surface heights according to grayscale image processing conventions.

distinct characters, while Fig. 2 displays the formation of an actual image from the Synthetic set.

A. Setup and Configuration

1) Image Sets: Our Synthetic image set consists of 3000 different 32x32 8-bit grayscale images in PNG format, each formed via first sampling from the same length-1024 Gaussian process and then proceeding exactly as outlined in Fig. 2. Note that, by construction, our experiments thus have perfect knowledge of the prior statistics governing this Synthetic image set—in turn, the GaussFill technique employing (1) using these prior statistics produces the optimal prediction in the minimum-mean-square-error sense. Our Cropland image sets are twofold, one set of 32x32 images and the other of 1024x1024 images, each consisting of 1011 different 8-bit grayscale images in PNG format. Both sets of Cropland images derive from the same original set of 24-bit color images in JPEG format, which exhibited varying sizes (in pixels) and varying spatial resolutions (in feet per pixel).[3] Across all of these original images, the coarsest spatial resolution was 15 ft, and thus, as Fig. 3's step from (a) to (b) illustrates, after all the higher-resolution images were appropriately downsized the smallest pixel dimensions across all images became 1382x1843; the step from (b) to (c)

[3]The URL of the image repository was noted when first discussing Fig.1.

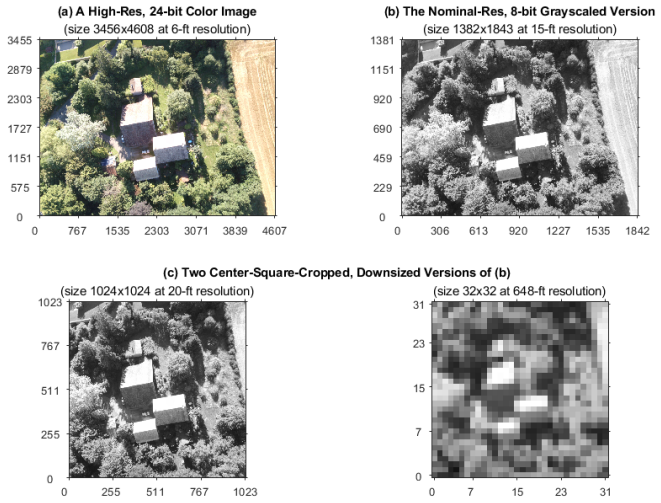

Fig. 3. The pre-processing steps of downloaded cropland imagery in preparation for the experimental comparison of DeepFill versus GaussFill techniques: (a) a 24-bit color image as rendered from its original JPEG format, which has rectangular pixel dimensions and a higher spatial resolution than others in the set; (b) the same image rendered after converting to 8-bit grayscale and downsizing its pixel dimensions (respectively via MATLAB's `rgb2gray` and `imresize` functions using default settings) to match the lowest spatial resolution in the set; and (c) a pair of square images, each further downsizing the center-square portion of the image in (b) to have pixel dimensions 1024x1024 and 32x32. The reduced spatial resolution of the 1024x1024 version (bottom-left) is hardly noticeable to the human eye, while that of the 32x32 version (bottom-right) is not only noticeable but its textural pattern is arguably more similar to the synthetic imagery in Fig. 2.

illustrates the cropping and further downsizing to yield image sets with the desired pair of square dimensions. It is worth noting that our direct implementation of GaussFill does not scale to Cropland's 1024x1024 images; even for Cropland's 32x32 images, GaussFill is no longer equipped with perfect knowledge of the prior statistics nor is there even a guarantee that Gaussian assumptions are truly applicable.

2) Mask Types: As was motivated in Fig.1, our experiments will compare fill techniques assuming three distinct types of patterns for the missing data values. The generated patterns, or masks, are represented as binary images where white (black) indicates a location with missing (non-missing) data. In the GaussFill notation of (1), for example, the black (white) pixels define the set V of visited locations (the set U of unvisited locations). All masks designate (by white pixels) that data is missing from $\frac{1}{16}$th grid cells in the environment. Block masks are constrained to themselves be square with a randomly-selected center such that it remains entirely within the area of interest. Speckled masks select their locations from the area of interest at random and without replacement. Finally, the robot path masks represent a battery-constrained robotic information gathering scenario, where the robot could not visit all the locations in the environment, and therefore, inference is needed to "know" the sensor measurements at the unvisited locations. The robot randomly explores and collects sensor measurements until $\frac{1}{16}$th grid cells are left to be visited whose measurements we infer. This creates irregular mask shapes.

3) Fill Techniques: Section III discussed the two primary fill techniques of interest, DeepFill and GaussFill. Our implementation of DeepFill uses the Pytorch v2 version[4], provided as trained by its developers. Quoting the GitHub site:[5] "The training dataset is a collection of images from Places365-Standard in which spatial sizes are larger than 512x512. (It will be freer to crop images with a larger resolution during training)." While the exact number of images used in the training dataset is unknown, the Places365-Standard dataset[6] has approximately 1.8 million images. We all also evaluate a third technique called *AverageFill*, but its implementation is straightforward: replace all missing data values with the empirical average of the non-missing values. Its score serves only as a lowest-fidelity method against which our DeepFill versus GaussFill comparisons can be calibrated. Recall that these comparisons also have a highest-fidelity calibration by the mean-square-error optimality of GaussFill when equipped with perfect knowledge of the prior statistics, which by construction is available for the Synthetic image set. On the Cropland image set, however, our implementation of GaussFill must first determine suitable prior statistics. We accomplish this through a maximum-likelihood fit (via MATLAB's `fitrgp` function using default settings) given the unmasked data values. It assumes a reduced-order Gaussian process relying upon identical structural assumptions that the model parameters assigned in Fig. 4 are an instance of. That is, when presented with a masked image from either (32x32) set, the first step of GaussFill is to use the unmasked data with maximum-likelihood to estimate the structured mean field $\hat{\mu}$, the exponential kernel parameters $(\hat{\beta}, \hat{\ell})$ and then construct the associated homogeneous and isotropic covariance matrix $\hat{\Sigma}$. The GaussFill technique for predicting the missing data values then proceeds in the same manner as when true prior statistics are known, only using the fitted statistics. On the Synthetic data, we thus can evaluate two variations of the GaussFill technique, *Gaussfill (optimal)* using the true prior statistics and *GaussFill (fitted)* using the estimated prior statistics.

B. Results

Fig. 4 considers the 32x32 Cropland image of Fig. 3 against a particular pair of masks, one block and one speckled, and visualizes the fill results of our three techniques: DeepFill, GaussFill (fitted) and AverageFill. On the other hand, Fig. 5 shows an illustrative realization of the filled results on a 32x32 synthetic dataset using the above-mentioned techniques. The indicated mean-square error (MSE) between an original and a filled information field represents how close-to-reality the modeling was, i.e., lower is better.

Table II presents the MSE values (in terms of averages and standard deviations) for robot path masks and Synthetic data samples. Our proposed DeepFill approach performs

[4]See `https://github.com/csqiangwen/DeepFillv2_Pytorch/tree/04d44efce277f6d67001e6ae0a187bcaf0f24860`

[5]See `http://places2.csail.mit.edu/download.html`

[6]See `https://github.com/CSAILVision/places365`

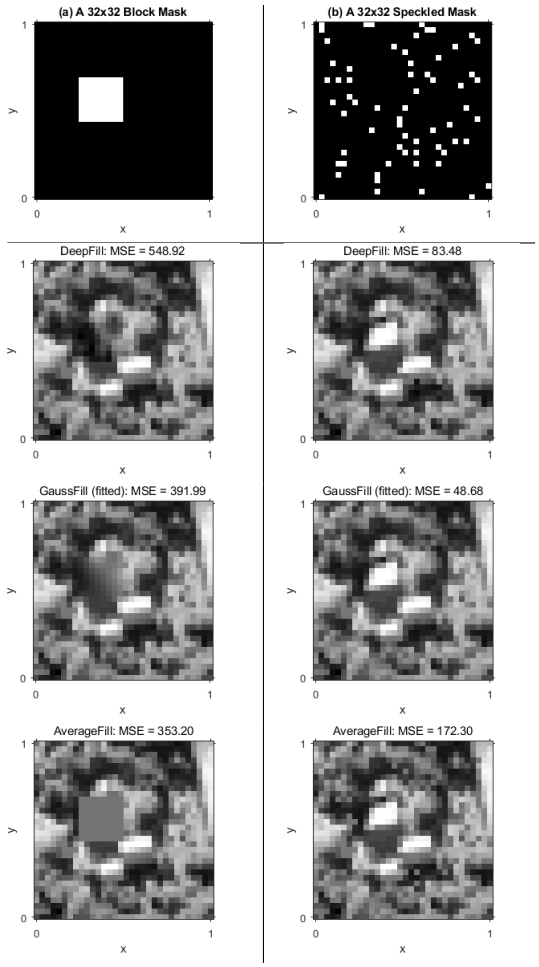

Fig. 4. A realization for two types of masks, namely (a) Block and (b) Speckled, compatible with our 32x32 image sets. Both masks indicate 64 locations of missing data values, but the Block type always renders them as a square region whereas the Speckled type renders them without adjacency constraints. (The analogous two types of masks compatible with Cropland's 1024x1024 image set, which each indicate 65,536 locations of missing data values, are visualized in Fig. 1.) Also shown under each mask, given the 32x32 data of Fig. 3, are the fills using the three different techniques and their corresponding MSE values.

almost 12 times better than the AverageFill approach while performing only about twice as worse as the GaussFill technique. Table III summarizes the associated MSE values over the Synthetic image set while using block and speckled masks. As expected for the Synthetic data, GaussFill (optimal) performed the best because of its knowledge of the true prior statistics. Similarly expected is that Gaussfill (fitted) performed close to the DeepFill technique. Note that the relative differences between the MSE scores of DeepFill and Gaussfill (fitted) is nearly negligible–the maximum being 1.72 times worse than GaussFill (fitted) with speckled masks on the Synthetic dataset. On the other hand, the AverageFill technique performs the worst—up to 33.19 times worse than DeepFill on the synthetic dataset. Table IV presents the same summary but over both Cropland image sets. On the 32x32 versions, DeepFill and GaussFill are again close in performance relative to where AverageFill sits. For example, with speckled masks, DeepFill performs only 1.73 times

Fig. 5. An example of a robot path mask and the inferred information fields using three tested techniques are shown. a) Original spatial information field, b) path mask (white cells indicate the unvisited locations – $1/16$-th of the environment), and the inferred data using c) DeepFill (MSE = 17.50), d) GaussFill (Fitted) (MSE = 3.68), and e) AverageFill (MSE = 329.57) techniques.

TABLE II

MEAN SQUARE ERRORS BETWEEN GROUND TRUTH AND FILLED DATA OVER THE 8-BIT (32X32) SYNTHETIC IMAGE SET

Technique	Robot Path Masks
DeepFill	16.54 ± 12.15
GaussFill (Fitted)	7.50 ± 9.87
AverageFill	197.18 ± 121.28

TABLE III

MEAN SQUARE ERRORS BETWEEN GROUND TRUTH AND FILLED DATA OVER THE 8-BIT (32X32) SYNTHETIC IMAGE SET

Technique	Block Masks	Speckled Masks
DeepFill	15.78 ± 10.87	5.89 ± 2.04
GaussFill (Fitted)	9.33 ± 16.69	3.42 ± 4.67
AverageFill	200.21 ± 174.05	195.48 ± 40.87
GaussFill (Optimal)	6.85 ± 4.48	2.66 ± 1.05

worse than the GaussFill (Fitted). In case of block masks, DeepFill performs even better—1.13 times worse than the GaussFill (Fitted). Note that when an original 1024x1024 image is resized to 32x32 for the sake of testing against the GaussFill technique, the (trilinear) interpolation essentially averages the fine-grained neighborhood pixel values to populate each coarse-grained pixel. While the resulting 32x32 image is more likely to obey Gaussian assumptions, prompting the GaussFill technique to perform better than expected, the greatly reduced spatial resolution is practically useless for any real-world farming application (see Fig. 3 for an example). With 1024x1024 Cropland data, GaussFill does not scale while DeepFill continues to perform notably better than AverageFill, especially on speckled masks (up to 9.17 times better). We notice high standard deviations in some cases due to a few outliers. For example, on the Synthetic dataset with block masks, the GaussFill (Fitted) technique has a mean MSE score of 9.33, but a standard deviation of 16.69. The highest MSE value in this case was 269.37 whereas the median was 5.84.

On all data sets, MSE scores are better with speckled masks than with block and path masks. This is because filling is essentially a neighborhood-extrapolation problem. In block regions, there are increased chances of compounding errors from border inwards while filling speckles has less of that risk [11], [13]. This is also true for robot path masks as one region (e.g., a corner) in the environment was completely unexplored by the robot, which created a block-like mask in essence. One should note that the relatively poorer perfor-

TABLE IV

MEAN SQUARE ERRORS BETWEEN GROUND TRUTH AND FILLED DATA OVER THE 8-BIT CROP IMAGE SETS (BOTH 32X32 AND 1024X1024)

Technique	Block Masks	Speckled Masks
The 32x32 Versions (spatial resolution of meters/pixel)		
DeepFill	32.99 ± 59.74	9.41 ± 12.13
GaussFill (Fitted)	29.16 ± 60.29	5.43 ± 8.77
AverageFill	63.12 ± 82.28	65.45 ± 62.55
The 1024x1024 Versions (spatial resolution of meters/pixel)		
DeepFill	65.04 ± 56.86	10.72 ± 6.90
AverageFill	96.82 ± 90.42	98.33 ± 71.90

mance of DeepFill on the Cropland (1024x1024) dataset with block masks is because filling in a missing 256x256 region of a high-resolution image is hard not only for any technique but for human brains as well. In terms of run time, prediction of the sensor measurements in the missing locations using DeepFill was extremely fast, taking on average only 0.02 sec. and 0.62 sec. on, respectively, the 32x32 and 1024x1024 information fields.

V. CONCLUSION

Gaussian Process modeling is the most popular approach in the robotic information gathering literature. However, such modeling techniques usually suffer from scalability and applicability issues. In this paper, we argue that we can leverage the recent advances in deep neural networks and image inpainting techniques for such information modeling tasks. Unlike Gaussian-based models, the inpainting techniques do not require the underlying information field of the ambient phenomena to be Gaussian. Furthermore, they easily scale beyond millions of pixels (informative locations in our setting), which is infeasible for standard Gaussian Process-based inference methods. To this end, we have employed a state-of-the-art inpainting technique for such an information modeling task on both synthetic and real-world crop datasets. Experimental results show that the inpainting-based inference technique is extremely fast and it performed comparably against the popular Gaussian regression technique while outscaling the regression method for large (1024x1024) crop field modeling. In the future, we plan to adopt an active sensing model, where the robot's actively sensed information will determine the inferred field, which in turn, will drive the robot to the next location, where the robot will sense information again. The process will continue until the robot's battery lasts.

REFERENCES

[1] A. Dutta, S. Roy, O. P. Kreidl, and L. Bölöni, "Multi-robot information gathering for precision agriculture: Current state, scope, and challenges," *IEEE Access*, vol. 9, pp. 161 416–161 430, 2021.

[2] G. A. Hollinger and G. S. Sukhatme, "Sampling-based robotic information gathering algorithms," *The International Journal of Robotics Research*, vol. 33, no. 9, pp. 1271–1287, 2014.

[3] K.-C. Ma, L. Liu, and G. S. Sukhatme, "Informative planning and online learning with sparse gaussian processes," in *2017 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2017, pp. 4292–4298.

[4] A. Viseras, D. Shutin, and L. Merino, "Online information gathering using sampling-based planners and gps: An information theoretic approach," in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2017, pp. 123–130.

[5] A. Singh, A. Krause, C. Guestrin, and W. J. Kaiser, "Efficient informative sensing using multiple robots," *Journal of Artificial Intelligence Research*, vol. 34, pp. 707–755, 2009.

[6] A. Singh, A. Krause, and W. J. Kaiser, "Nonmyopic adaptive informative path planning for multiple robots," in *IJCAI 2009, Proceedings of the 21st International Joint Conference on Artificial Intelligence, Pasadena, California, USA, July 11-17, 2009*, C. Boutilier, Ed., 2009, pp. 1843–1850. [Online]. Available: http://ijcai.org/Proceedings/09/Papers/306.pdf

[7] A. Solin, M. Kok, N. Wahlström, T. B. Schön, and S. Särkkä, "Modeling and interpolation of the ambient magnetic field by gaussian processes," *IEEE Transactions on robotics*, vol. 34, no. 4, pp. 1112–1127, 2018.

[8] Y. Wei and R. Zheng, "Informative path planning for mobile sensing with reinforcement learning," in *IEEE INFOCOM 2020-IEEE Conference on Computer Communications*. IEEE, 2020, pp. 864–873.

[9] C. E. Rasmussen and C. K. Williams, *Gaussian processes for machine learning*. MIT press Cambridge, 2006, vol. 1.

[10] A. Krause, A. Singh, and C. Guestrin, "Near-optimal sensor placements in gaussian processes: Theory, efficient algorithms and empirical studies," *Journal of Machine Learning Research*, vol. 9, no. Feb, pp. 235–284, 2008.

[11] M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester, "Image inpainting," in *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, 2000, pp. 417–424.

[12] W. Cai and Z. Wei, "Piigan: generative adversarial networks for pluralistic image inpainting," *IEEE Access*, vol. 8, pp. 48 451–48 463, 2020.

[13] J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, and T. S. Huang, "Generative image inpainting with contextual attention," in *Proc. of the IEEE Conference on Computer vision and Pattern Recognition (CVPR-2018)*, 2018, pp. 5505–5514.

[14] ——, "Free-form image inpainting with gated convolution," in *Proc. of the IEEE/CVF Int. Conference on Computer Vision (ICCV-2019)*, 2019, pp. 4471–4480.

[15] T. Said, J. Wolbert, S. Khodadadeh, A. Dutta, O. P. Kreidl, L. Bölöni, and S. Roy, "Multi-robot information sampling using deep mean field reinforcement learning," in *2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, 2021, pp. 1215–1220.

[16] T. Samman, J. Spearman, A. Dutta, O. P. Kreidl, S. Roy, and L. Bölöni, "Secure multi-robot adaptive information sampling," in *2021 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR)*, 2021, pp. 125–131.

[17] A. Viseras, D. Shutin, and L. Merino, "Robotic active information gathering for spatial field reconstruction with rapidly-exploring random trees and online learning of gaussian processes," *Sensors*, vol. 19, no. 5, p. 1016, 2019.

[18] C. E. Rasmussen, "Gaussian processes in machine learning," in *Summer School on Machine Learning*. Springer, 2003, pp. 63–71.

[19] N. Cao, K. H. Low, and J. M. Dolan, "Multi-robot informative path planning for active sensing of environmental phenomena: a tale of two algorithms," in *International Conference on Autonomous Agents and Multi-Agent Systems, AAMAS '13, Saint Paul, MN, USA, May 6-10, 2013*, M. L. Gini, O. Shehory, T. Ito, and C. M. Jonker, Eds. IFAAMAS, 2013, pp. 7–14. [Online]. Available: http://dl.acm.org/citation.cfm?id=2484926

[20] Y. Kantaros, B. Schlotfeldt, N. Atanasov, and G. J. Pappas, "Sampling-based planning for non-myopic multi-robot information gathering," *Autonomous Robots*, vol. 45, no. 7, pp. 1029–1046, 2021.

[21] A. Krause and C. Guestrin, "Nonmyopic active learning of gaussian processes: an exploration-exploitation approach," in *Proceedings of the 24th international conference on Machine learning*, 2007, pp. 449–456.

[22] J. L. Nguyen, N. R. Lawrance, and S. Sukkarieh, "Nonmyopic planning for long-term information gathering with an aerial glider," in *2014 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2014, pp. 6573–6578.

[23] J. Quinonero-Candela and C. E. Rasmussen, "A unifying view of sparse approximate gaussian process regression," *The Journal of Machine Learning Research*, vol. 6, pp. 1939–1959, 2005.

[24] A. Ranganathan, M.-H. Yang, and J. Ho, "Online sparse gaussian process regression and its applications," *IEEE Transactions on Image Processing*, vol. 20, no. 2, pp. 391–404, 2010.

[25] M. W. Seeger, C. K. Williams, and N. D. Lawrence, "Fast forward selection to speed up sparse gaussian process regression," in *International Workshop on Artificial Intelligence and Statistics*. PMLR, 2003, pp. 254–261.

[26] K.-C. Ma, L. Liu, and G. S. Sukhatme, "An information-driven and disturbance-aware planning method for long-term ocean monitoring," in *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2016, pp. 2102–2108.

[27] O. Elharrouss, N. Almaadeed, S. Al-Maadeed, and Y. Akbari, "Image inpainting: A review," *Neural Processing Letters*, vol. 51, no. 2, pp. 2007–2028, 2020.

[28] S. Iizuka, E. Simo-Serra, and H. Ishikawa, "Globally and locally consistent image completion," *ACM Transactions on Graphics (ToG)*, vol. 36, no. 4, pp. 1–14, 2017.

[29] D. Pathak, P. Krahenbuhl, J. Donahue, T. Darrell, and A. A. Efros, "Context encoders: Feature learning by inpainting," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2536–2544.

[30] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," *Communications of the ACM*, vol. 63, no. 11, pp. 139–144, 2020.

[31] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," *arXiv preprint arXiv:1312.6114*, 2013.

[32] R. Shrestha, F.-P. Tian, W. Feng, P. Tan, and R. Vaughan, "Learned map prediction for enhanced mobile robot exploration," in *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 2019, pp. 1197–1204.

[33] T. Li, C. Wang, M. Q.-H. Meng, and C. W. de Silva, "Attention-driven active sensing with hybrid neural network for environmental field mapping," *IEEE Transactions on Automation Science and Engineering*, vol. 19, no. 3, pp. 2135–2152, 2021.

[34] J. A. Caley and G. A. Hollinger, "Environment prediction from sparse samples for robotic information gathering," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 10 577–10 583.

[35] A. S. Willsky, "Multiresolution markov models for signal and image processing," *Proceedings of the IEEE*, vol. 90, no. 8, pp. 1396–1458, 2002.

[36] J. K. Johnson and A. S. Willsky, "A recursive model-reduction method for approximate inference in gaussian markov random fields," *IEEE transactions on image processing: a publication of the IEEE Signal Processing Society*, vol. 17, no. 1, pp. 70–83, 2008.

[37] M. J. Wainwright, M. I. Jordan *et al.*, "Graphical models, exponential families, and variational inference," *Foundations and Trends® in Machine Learning*, vol. 1, no. 1–2, pp. 1–305, 2008.

[38] T. Hastie, R. Tibshirani, and M. Wainwright, "Statistical learning with sparsity," *Monographs on statistics and applied probability*, vol. 143, p. 143, 2015.

[39] A. Spantini, D. Bigoni, and Y. Marzouk, "Inference via low-dimensional couplings," *The Journal of Machine Learning Research*, vol. 19, no. 1, pp. 2639–2709, 2018.

[40] C. Guillemot and O. Le Meur, "Image inpainting: Overview and recent advances," *IEEE signal processing magazine*, vol. 31, no. 1, pp. 127–144, 2013.

[41] M. Richard and M. Y.-S. Chang, "Fast digital image inpainting," in *Appeared in the Proceedings of the International Conference on Visualization, Imaging and Image Processing (VIIP 2001), Marbella, Spain*, 2001, pp. 106–107.

[42] C. Guestrin, A. Krause, and A. P. Singh, "Near-optimal sensor placements in gaussian processes," in *Proceedings of the 22nd international conference on Machine learning*. ACM, 2005, pp. 265–272.