Disentangled Multi-Fidelity Deep Bayesian Active Learning

Dongxia Wu [1 2] **Ruijia Niu** [1] **Matteo Chinazzi** [3 4] **Yi-An Ma** [2 1] **Rose Yu** [1 2]

Abstract

To balance quality and cost, various domain areas of science and engineering run simulations at multiple levels of sophistication. Multi-fidelity active learning aims to learn a direct mapping from input parameters to simulation outputs at the highest fidelity by actively acquiring data from multiple fidelity levels. However, existing approaches based on Gaussian processes are hardly scalable to high-dimensional data. Deep learning-based methods often impose a hierarchical structure in hidden representations, which only supports passing information from low-fidelity to high-fidelity. These approaches can lead to the undesirable propagation of errors from low-fidelity representations to high-fidelity ones. We propose a novel framework called Disentangled Multi-fidelity Deep Bayesian Active Learning (`D-MFDAL`), which learns the surrogate models conditioned on the distribution of functions at multiple fidelities. On benchmark tasks of learning deep surrogates of partial differential equations including heat equation, Poisson's equation and fluid simulations, our approach significantly outperforms state-of-the-art in prediction accuracy and sample efficiency.

1. Introduction

Mathematical modeling and simulations play a crucial role in various scientific and engineering fields, ranging from diffusion modeling to epidemic simulation. These models can often be simulated at different levels of sophistication. High-fidelity models provide highly accurate results but require more computational resources, while low-fidelity models offer less accuracy but are less computationally expensive. Multi-fidelity modeling, as outlined in (Peherstorfer et al., 2018), aims to strike a balance between computation cost and prediction accuracy by using data from multiple levels of fidelity to learn an accurate high-fidelity surrogate. The learned surrogate can replicate the behavior of the original model to eliminate the complex numerical integration.

While Gaussian processes (GPs) remain to be predominant tools in multi-fidelity modeling (Perdikaris et al., 2016; Wang et al., 2021), deep learning arises as a more scalable alternative for high-dimensional data (Cutajar et al., 2019; Wang & Lin, 2020; Hebbal et al., 2021; Wu et al., 2022). These methods use a deep neural network to learn a direct mapping from input parameters to simulation outputs using multi-fidelity data. However, they also require simulating massive training data beforehand, which is expensive to obtain, especially for high-fidelity simulation.

Multi-fidelity deep active learning (MFDAL) (Li et al., 2022b;a) proposes a framework to acquire data at different fidelity levels with deep learning and to reduce the cost of data simulation. Such models pass information from low-fidelity to high-fidelity hidden representations through a neural network (NN). This design requires accurate hidden representations at each fidelity to propagate useful information from low-fidelity to high-fidelity levels. However, in multi-fidelity active learning, these hidden representations can be easily erroneous when the number of training data is highly unbalanced at each fidelity and the data distribution is dramatically shifted during the beginning stage of active learning. Moreover, the trained surrogate model will also have the overfitting issue at the beginning stage with limited training data at each fidelity level. These overfitted hidden representations are less accurate and their error will propagate from low-fidelity to high-fidelity.

To alleviate the overfitting problem, (Wu et al., 2022) propose a unified neural latent variable model for multi-fidelity surrogate modeling called Multi-fidelity Hierarchical Neural Processes (MFHNP). They introduce latent variables to learn the distributions over functions at each fidelity level. However, this model still requires a hierarchical structure to pass information from low-fidelity to high-fidelity levels via hidden representations of a NN. Therefore, the error propagation issue remains.

[1]Department of Computer Science and Engineering, University of California San Diego, La Jolla, USA [2]Halıcıoğlu Data Science Institute, University of California San Diego, La Jolla, USA [3]The Roux Institute, Northeastern University, Portland, USA [4]Network Science Institute, Northeastern University, Boston, USA. Correspondence to: Rose Yu <roseyu@ucsd.edu>.

Proceedings of the 40^{th} International Conference on Machine Learning, Honolulu, Hawaii, USA. PMLR 202, 2023.

In this work, we design a novel framework called Disentangled Multi-fidelity Deep Bayesian Active Learning (D-MFDAL) to learn the multi-fidelity representations in the functional space. D-MFDAL is able to solve both error propagation and overfitting issues mentioned above. Specifically, D-MFDAL belongs to the Neural Process (NP) family (Garnelo et al., 2018b;a) to learn the latent variables from the individual latent representations of the input-output pairs in the context set. The latent variables are used to represent the distributions over functions at each fidelity level. D-MFDAL disentangles these individual latent representations into two parts for global-local separation. The global representations are treated as the samples generated from latent representations among all fidelity levels, while the local ones are samples generated from latent representations at individual fidelity level. In this way, D-MFDAL avoids the hierarchical model architecture.

We design a unified evidence lower bound (ELBO) for the joint distribution among all fidelity levels as the training loss and introduce the multi-fidelity regularization term to enforce similar global representations across the fidelity levels for the same sample. Furthermore, we extend the acquisition function, latent information gain (Wu et al., 2023), designed for Bayesian active learning on NP-based models to multi-fidelity setting and design an efficient algorithm for budget-constrained batch active learning.

In summary, our contributions include:

- A scalable Disentangled Multi-fidelity Deep Bayesian Active Learning framework (D-MFDAL). The disentangled representation makes it flexible and efficient to share global information across all fidelity levels.
- A novel acquisition function called Multi-fidelity Latent Information Gain (MF-LIG) and an efficient algorithm for budget-constrained greedy-based batch active learning implementation.
- Superior performance in multiple benchmark studies of learning deep surrogates of partial differential equations and complex fluid prediction task in both passive learning and active learning settings.

2. Background

Muti-Fidelity Modeling. Formally, given input domain $\mathcal{X} \subseteq \mathbb{R}^{d_x}$ and output domain $\mathcal{Y} \subseteq \mathbb{R}^{d_y}$, a model is a (stochastic) function $f : \mathcal{X} \to \mathcal{Y}$. The evaluations of f incur computational costs $c > 0$. The computational costs c are higher at higher fidelity level ($c_1 < ... < c_K$). In multi-fidelity modeling, we have a set of functions $\{f_1, \cdots, f_K\}$ that approximate f with increasing accuracy and computational cost. Our target is to learn a deep surrogate model $\hat{f}_K$ based on data from K fidelity levels and N different parameter settings (scenarios) $\{x_{k,n}, y_{k,n}\}_{k=1,n=1}^{K,N}$.

Neural Processes. Neural processes (NPs) (Garnelo et al., 2018b) are a family of conditional latent variable models for implicit stochastic processes ($\mathcal{SP}s$) (Wang & Van Hoof, 2020). NPs combine GPs and neural networks (NNs). Like GPs, NPs can represent distributions over functions and can estimate the uncertainty of the predictions. But they are more scalable in high dimensions and allow continual and active learning out-of-the-box (Jha et al., 2022).

According to Kolmogorov Extension Theorem (Øksendal, 2003), NPs meet exchangeability and consistency conditions to define $\mathcal{SP}s$. Formally, NP includes latent variables $z \in \mathbb{R}^{d_z}$ and model parameters θ and is trained by the context set $\mathcal{D}^c \equiv \{x_n^c, y_n^c\}_{n=1}^N$ and target sets $\mathcal{D}^t \equiv \{x_m^t, y_m^t\}_{m=1}^M$. Here $\mathcal{D}^c$ and $\mathcal{D}^t$ are randomly split from the training set $\mathcal{D}$. Learning the posterior of z and θ is equivalent to maximizing the following posterior likelihood:

$$p(y_{1:M}^t | x_{1:M}^t, \mathcal{D}^c, \theta) = \int p(z|\mathcal{D}^c, \theta) \prod_{m=1}^{M} p(y_m^t | z, x_m^t, \theta) dz \tag{1}$$

Since marginalizing over the latent variables z is intractable, the NP family (Garnelo et al., 2018b; Kim et al., 2019) uses approximate inference and derives the corresponding evidence lower bound (ELBO):

$$\log p(y_{1:M}^t | x_{1:M}^t, \mathcal{D}^c, \theta) \geq$$
$$\mathbb{E}_{q_\phi(z|\mathcal{D}^c \cup \mathcal{D}^t)} \big[\sum_{m=1}^{M} \log p(y_m^t | z, x_m^t, \theta) + \log \frac{q_\phi(z|\mathcal{D}^c)}{q_\phi(z|\mathcal{D}^c \cup \mathcal{D}^t)} \big] \tag{2}$$

Note that this variational approach approximates the intractable true posterior $p(z|\mathcal{D}^c, \theta)$ with the approximate posterior $q_\phi(z|\mathcal{D}^c)$. This approach is also an amortized inference method as the global parameters ϕ are shared by all context data points. It is efficient during the test time (no per-data-point optimization) (Volpp et al., 2020).

3. Methodology

Our proposed D-MFDAL is presented in two sections. First, we describe the disentangled neural processes architecture, specifically designed for multi-fidelity surrogate modeling and the associated training procedure. Secondly, we introduce a new acquisition function (MF-LIG) for multi-fidelity active learning, which extends Latent Information Gain (Wu et al., 2023). Additionally, we present a greedy-based algorithm for batch active learning under budget constraints.

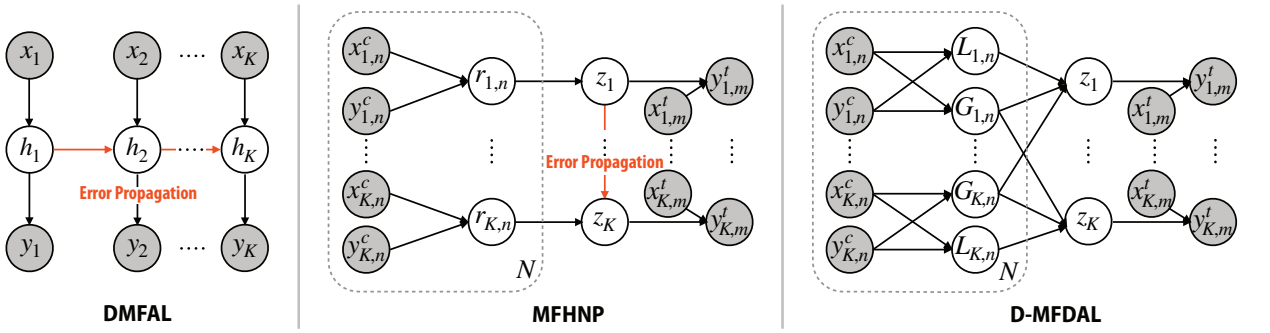

Figure 1. Graphical model: Left and Middle: two multi-fidelity surrogate modeling baselines. Both have hierarchical structures. They use the hidden variable h_k or the latent variable z_k to pass information from low-fidelity to high-fidelity levels and therefore suffer from the error propagation issue. Right: `D-MFDAL` disentangles the latent representations $r_{k,n}$ shown in MFHNP into local representations $L_{k,n}$ and global representations $G_{k,n}$, and directly uses them to infer the latent variable z_k. z_k are conditionally independent of each other given the local and global representations. Shaded circles denote observed variables and hollow circles represent latent variables. The directed edges represent conditional dependence.

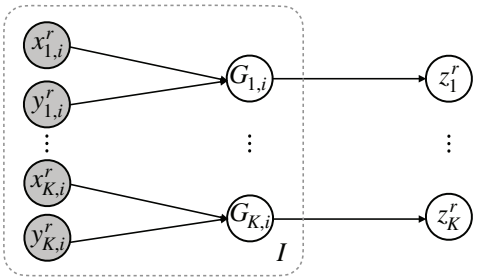

Figure 2. Graphical model: Inference graph for the reference context pairs $\{x^r_{k,i}, y^r_{k,i}\}$. Shaded circles denote observed variables and hollow circles represent latent variables. The directed edges represent conditional dependence.

3.1. Disentangled Multi-fidelity Neural Processes

We design a NP based model, Disentangled Multi-fidelity Neural Processes (DMFNP), to efficiently integrate information from multiple fidelity levels without the hierarchical structure.

Local and Global Latent Representations. The key idea of the `D-MFDAL` model is to disentangle latent representations $r_{k,n}$ into local representations $L_{k,n}$ and global representations $G_{k,n}$, see Figure 1 right. Intuitively, $G_{k,n}$ embeds the information from the context pair $\{x^c_{k,n}, y^c_{k,n}\}$ that can be shared to all fidelity levels, where k is the fidelity level of the context pair and n is the scenario index. On the other hand, $L_{k,n}$ embeds the information from the context pair $\{x^c_{k,n}, y^c_{k,n}\}$ that is only for the fidelity level k.

Multi-fidelity Bayesian Context Aggregation. We extend Bayesian aggregation (BA) (Volpp et al., 2020) to infer latent variables z_k. We learn the local and global representation $L_{k,n}, G_{k,n}$ together with the corresponding variance $\sigma^2_{L_{k,n}}, \sigma^2_{G_{k,n}}$. The local representation $L_{k,n}$ can be considered as a sample of $p(z_k)$. On the other hand, we treat the global representation $G_{k,n}$ as K copies of samples of $p(z_k)$ across all fidelity levels. Then we aggregate local and global representations of context data pairs to infer z following the graph in Figure 1. We implement it using the factorized Gaussian observation model with the following form:

$$\begin{aligned} p(L_{k,n}|z_k) &= \mathcal{N}(L_{k,n}|z_k, \text{diag}(\sigma^2_{L_{k,n}})), \\ L_{k,n} &= \text{enc}_\phi(x^C_{k,n}, y^C_{k,n}). \\ p(G_{k,n}|z_k) &= \mathcal{N}(G_{k,n}|z_k, \text{diag}(\sigma^2_{G_{k,n}})), \\ p(G_{k,n}|z_m) &= \mathcal{N}(G_{k,n}|z_m, \text{diag}(\sigma^2_{G_{k,n}})), \text{for all } m \neq k \\ G_{k,n} &= \text{enc}_\phi(x^C_{k,n}, y^C_{k,n}). \end{aligned} \tag{3}$$

We use factorized Gaussian priors

$$p_0(z_k) := \mathcal{N}(z_k|\mu_{z_{k,0}}, \text{diag}(\sigma^2_{z_{k,0}}))$$

to derive a multi-fidelity Gaussian aggregation model and update the parameters of the posterior distribution

$q_\phi(z_k|\mathcal{D}^c)$ in a closed form:

$$\sigma^2_{z_k} = \big[(\sigma^2_{z_{k,0}})^\ominus + \sum_{n=1}^{N}(\sigma^2_{L_{k,n}})^\ominus) + \sum_{j=1}^{K}[\sum_{n=1}^{N}(\sigma^2_{G_{j,n}})^\ominus)]\big]^\ominus,$$

$$\mu_{z_k} = \mu_{z_{k,0}} + \sigma^2_{z_k} \odot \big[\sum_{n=1}^{N}(L_{k,n} - \mu_{z_{k,0}}) \oslash (\sigma^2_{L_{k,n}}) + \sum_{j=1}^{K}[\sum_{n=1}^{N}(G_{j,n} - \mu_{z_{k,0}}) \oslash (\sigma^2_{G_{j,n}})]\big]. \quad (4)$$

where $\ominus$, $\odot$ and $\oslash$ denote element-wise inversion, product, and division, respectively.

Unified ELBO. We design a unified ELBO based on the D-MFDAL model. For multi-fidelity surrogate modeling, we infer the latent variables z_k at each fidelity level. Therefore, we use K encoders $q_{\phi_k}(z_k|\mathcal{D}^c)$ and K decoders $p_{\theta_k}(y^t_k|z_k, x^t_k)$ for $k \in \{1, ..., K\}$. When $K = 2$, we can derive the corresponding ELBO containing 4 terms as:

$$\begin{aligned}
&\log p(y^t_1, y^t_2|x^t_1, x^t_2, \mathcal{D}^c, \theta)\\
\geq&\mathbb{E}_{q_\phi(z_1,z_2|\mathcal{D}^c\cup\mathcal{D}^t)}\big[\log p(y^t_1, y^t_2|z_1, z_2, x^t_1, x^t_2, \theta) + \log\frac{q_\phi(z_1, z_2|\mathcal{D}^c)}{q_\phi(z_1, z_2|\mathcal{D}^c \cup \mathcal{D}^t)}\big]\\
=&\mathbb{E}_{q_{\phi_2}(z_2|\mathcal{D}^c\cup\mathcal{D}^t)q_{\phi_1}(z_1|\mathcal{D}^c\cup\mathcal{D}^t)}\big[\log p(y^t_2|z_2, x^t_2, \theta_2) + \log p(y^t_1|z_1, x^t_1, \theta_1) + \log\frac{q_{\phi_2}(z_2|\mathcal{D}^c)}{q_{\phi_2}(z_2|\mathcal{D}^c \cup \mathcal{D}^t)} + \frac{q_{\phi_1}(z_1|\mathcal{D}^c)}{q_{\phi_1}(z_1|\mathcal{D}^c \cup \mathcal{D}^t)}\big]
\end{aligned} \quad (5)$$

Such a unified ELBO objective can be generalized to accommodate any desired number of fidelity levels.

Multi-Fidelity Regularization. Since $G_{k,n}$ is the global representation, any $(G_{k_1,i}, G_{k_2,i})$ pair should be similar across fidelity levels for the same scenario i. However, since the output dimensions are different at each fidelity level, D-MFDAL cannot share the encoder at different fidelity levels. Therefore, we introduce reference context data $\mathcal{D}^r_k = \{x^r_{k,i}, y^r_{k,i}\}^I_{i=1}$, which is shared across all fidelity levels (see Figure 2 for the inference graph). I is the total number of reference scenarios. We design the multi-fidelity regularization term to minimize the Jensen–Shannon divergence between the inferred posterior z^r_k distribution from $(x^r_{k,i}, y^r_{k,i})$ pairs (where $k < K$) and the posterior z^r_K distribution from $(x^r_{K,i}, y^r_{K,i})$ pairs. Note that D-MFDAL does not require additional data as we use the initial training data as reference data for fair comparison.

We use factorized Gaussian priors for reference latent representations z^r_k:

$$p_0(z^r_k) := \mathcal{N}(z^r_k|\mu_{z^r_{k,0}}, \text{diag}(\sigma^2_{z^r_{k,0}}))$$

The posterior distribution $q_\phi(z^r_k|\mathcal{D}^r_k)$ can be written as:

$$\begin{aligned}
\sigma^2_{z^r_k} &= \big[(\sigma^2_{z^r_{k,0}})^\ominus + \sum_{n=1}^{N}(\sigma^2_{G_{k,n}})^\ominus)\big]^\ominus,\\
\mu_{z^r_k} &= \mu_{z^r_{k,0}} + \sigma^2_{z^r_k} \odot \big[\sum_{n=1}^{N}(G_{k,n} - \mu_{z^r_{k,0}}) \oslash (\sigma^2_{G_{k,n}})\big].
\end{aligned} \quad (6)$$

We further derive the multi-fidelity regularization using the sum of Jensen-Shannon divergence between the highest fidelity level K and all other lower fidelity levels k as:

$$\begin{aligned}
&\sum_{k=1}^{K}\text{JSD}(q_\phi(z^r_k|\mathcal{D}^r_k), q_\phi(z^r_K|\mathcal{D}^r_K))\\
=&\frac{1}{2}\sum_{k=1}^{K}\mathbb{E}_{q_\phi(z^r_k|\mathcal{D}^r_k)}\big[\log\frac{q_\phi(z^r_K|\mathcal{D}^r_K)}{q_\phi(z^r_k|\mathcal{D}^r_k)}\big]\\
+&\frac{1}{2}\sum_{k=1}^{K}\mathbb{E}_{q_\phi(z^r_K|\mathcal{D}^r_K)}\big[\log\frac{q_\phi(z^r_k|\mathcal{D}^r_k)}{q_\phi(z^r_K|\mathcal{D}^r_K)}\big]
\end{aligned} \quad (7)$$

Training Procedure. D-MFDAL is designed for scalable training, which means the model inference time should scale at most linearly with respect to the number of fidelity levels. It can be realized by using the disentangled latent representations to share the information across the fidelity levels. In this way, the latent variables z_k are conditionally independent to each other given the global representations G and the local representations L. Therefore, we no longer require nested Monte Carlo (MC) sampling of z_k from low-fidelity to high-fidelity levels as in previous models with hierarchical structures.

For the training loss including ELBO in Equation 5 and multi-fidelity regularization in Equation 7, we use MC sampling to optimize the following objective function:

$$\begin{aligned}
\mathcal{L}_{MC} = \sum_{k=1}^{K}\Big[&\tfrac{1}{S}\sum_{s=1}^{S}\log p(y^t_k|x^t_k, z^{(s)}_k)\\
&-\text{KL}[q(z_k|, \mathcal{D}^c, \mathcal{D}^t))\|p(z_k|\mathcal{D}^c)]\\
&+\text{JSD}(q(z^r_k|\mathcal{D}^r_k), q(z^r_K|\mathcal{D}^r_K))\Big]
\end{aligned} \quad (8)$$

where the latent variables $z^{(s)}_k$ is sampled by $q_{\phi_1}(z_k|\mathcal{D}^c)$. The sampling time scales linearly w.r.t. the number of fidelity levels.

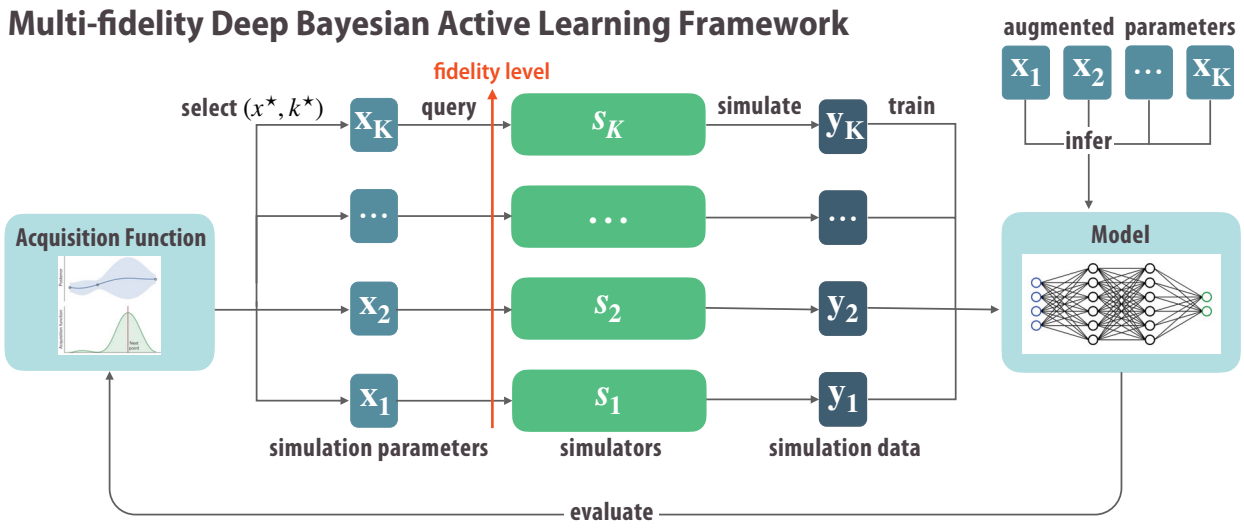

Figure 3. Illustration of the multi-fidelity deep Bayesian active learning framework (D-MFDAL). Given simulation parameters and data, D-MFDAL trains a deep surrogate model to infer the latent variables at each fidelity level. The inferred latent variables allow prediction and uncertainty quantification. The uncertainty is used to calculate the acquisition function (e.g. MF-LIG) to select the next set of parameters to query and simulate more data to add to the training set.

3.2. Multi-Fidelity Active Learning

In this section, we propose the novel acquisition function MF-LIG based on the model architecture of D-MFDAL for multi-fidelity active learning. Furthermore, we design a greedy batch multi-fidelity active learning algorithm with budget constraints for data efficiency.

Weighted Information Gain (IG). Define the search space as $S = \{(x_{k,n}, y_{k,n})\}_{k=1,n=1}^{K,N}$ with K fidelity levels and N input parameters for each fidelity. We flatten the search space and define the acquisition function as:

$$\text{IG}(x_{k,n}, y_{k,n}) = \frac{1}{c_k}[H(w) - H(w|x_{k,n}, y_{k,n})] \quad (9)$$

where c_k is the computational cost for level k. This is a naive implementation of IG for Bayesian active learning. In this paper, we study the continuous input parameter and discrete fidelity level setting:

$$\text{IG}(y_k(x_k)) = \frac{1}{c_k}[H(w) - H(w|y_k(x_k))]. \quad (10)$$

In practice, we do not know $y_k(x_k)$ before querying the simulator. The best we can do is to use the weighted information gain (EIG) to replace the weighted IG:

$$\text{EIG}(x_k) = \frac{1}{c_k}\mathbb{E}_{p(y_k(x_k))}[H(w) - H(w|y_k(x_k)]. \quad (11)$$

Latent Information Gain for Multi-Fidelity Active Learning. For multi-fidelity active learning, our goal is to improve the model performance at the highest fidelity level. Therefore, weighted IG/EIG is suboptimal as it treats all the model parameters w at each fidelity level equally important. To find the optimal solution, we design a new acquisition function called Multi-Fidelity Latent Information Gain (MF-LIG).

We start by searching for an x_k to optimize the EIG with respect to the model parameters used at the highest fidelity level. We can write the corresponding acquisition function:

$$\text{MF-EIG}(x_k) = \frac{1}{c_k}\mathbb{E}_{p(y_k(x_k))}[H(w_K) - H(w_K|y_k(x_k))]. \quad (12)$$

where w_K are the model parameters at the fidelity level K.

The next step is to use the inferred latent variable z_k of D-MFDAL to replace w_k as they are learned from the context set $\{x^c_{k,n}, y^c_{k,n}\}_{n=1}^N$ to represent $f_k(.)$ of the ground truth simulators and are capable of performing conditional modeling $p(y^t_{k,m}(x^t_{k,m})|z_k)$ at each fidelity level k. We then propose a new acquisition function MF-LIG measuring the weighted expected information gain between the prediction and the latent variables at the highest fidelity level:

$$\begin{aligned} a_s(x_k) &= \text{MF-LIG}(x_k) \\ &= \frac{1}{c_k}\mathbb{E}_{p(y_k(x_k))}\text{KL}[p(z_K|y_k(x_k))\|p(z_K))]. \end{aligned} \quad (13)$$

Algorithm 1 Batch MF-LIG

Input: costs $\{c_1, ..., c_K\}$, budget B, training set $\mathcal{D}$.
Initialize the current selected data index $j \leftarrow 0$, selected data set $\mathcal{D}_j^q \leftarrow \emptyset$, current cost $C_j \leftarrow 0$.
while $C_j \leq B$ **do**
  $(x^*, k^*) = \text{argmax}_{(x,k)}\text{MF-LIG}(x_k)$
  $j \leftarrow j + 1$
  $\mathcal{D}_j^q \leftarrow \mathcal{D}_{j-1}^q \cup \{((x^*, k^*, \hat{y}(x^*, k^*))\}$
  $\mathcal{D} \leftarrow \mathcal{D} \cup \mathcal{D}_j^q$
  $C_j \leftarrow C_{j-1} + c_{k^*}$
end while
Return $\mathcal{D}_j^q$

Batch Multi-Fidelity Active Learning Algorithm. We follow the greedy active learning algorithm by (Li et al., 2022a) using our proposed MF-LIG for budget-constrained batch active learning. Since MF-LIG is also a mutual information based acquisition function, the guaranteed near $(1 - 1/e)$ approximation for the greedy algorithm also applies in our case. Our approach is summarized in Algorithm 1 and the overall framework is visualized in Figure 3.

4. Related Work

Multi-fidelity Modeling. Multi-fidelity surrogate modeling is widely used in science and engineering fields, from aerospace systems (Brevault et al., 2020) to climate science (Hosking, 2020; Valero et al., 2021) (Valero et al., 2021). The pioneering work of (Kennedy & O'Hagan, 2000) uses GPs to relate models at multiple fidelity with an autoregressive model. (Le Gratiet & Garnier, 2014) proposed recursive GP with a nested structure in the input domain for fast inference. (Perdikaris et al., 2015; 2016) deals with high-dimensional GP settings by taking the Fourier transformation of the kernel function. (Perdikaris et al., 2017) proposed multi-fidelity Gaussian processes (NARGP) but assumes a nested structure in the input domain to enable a sequential training process at each fidelity level. Wang et al. (2021) proposed a Multi-Fidelity High-Order GP model to speed up the physical simulation. They extended the classical Linear Model of Coregionalization (LMC) to nonlinear case and placed a matrix GP prior on the weight functions. Deep Gaussian processes (DGPs) (Cutajar et al., 2019) design a single objective to optimize kernel parameters at each fidelity level jointly. However, DGPs are not scalable for applications with high-dimensional data.

Deep learning has been applied to multi-fidelity modeling. For example, (Guo et al., 2022) uses deep neural networks to combine parameter-dependent output quantities. (Meng & Karniadakis, 2020) propose a composite neural network for multi-fidelity data from inverse PDE problems. (Meng et al., 2021) propose Bayesian neural nets for multi-fidelity modeling. (De et al., 2020) use transfer learning to fine-tune the high-fidelity surrogate model with the deep neural network trained with low-fidelity data. (Cutajar et al., 2019; Hebbal et al., 2021) propose deep GPs to capture nonlinear correlations between fidelities, but their method cannot handle the case where different fidelities have data with different dimensions. Tangentially, multi-fidelity methods have also recently been investigated in Bayesian optimization, active learning and bandit problems (Li et al., 2020b; 2022a; Perry et al., 2019; Kandasamy et al., 2017).

Neural Processes (NPs) (Garnelo et al., 2018a; Kim et al., 2018; Louizos et al., 2019; Singh et al., 2019) provide scalable and expressive alternatives than GPs for modeling stochastic processes. It lies between GPs and NN. However, none of the existing NP models can efficiently incorporate multi-fidelity data. Previous work by (Raissi & Karniadakis, 2016) combines multi-fidelity GP with deep learning by placing a GP prior on the features learned by deep neural networks. Their model, however, remains closer to GPs. Quite recently, (Wang & Lin, 2020) proposed multi-fidelity neural process with physics constraints (MFPC-Net). They use NP to learn the correlation between multi-fidelity data by mapping both the input and output of the low-fidelity model to high-fidelity model output. But their model requires paired data and cannot utilize the remaining unpaired data at the low-fidelity level.

Bayesian Active Learning. Bayesian active learning is well studied in statistics and machine learning (Chaloner & Verdinelli, 1995; Cohn et al., 1996). GPs are popular for posterior estimation, e.g. (Houlsby et al., 2011; Zimmer et al., 2018), but often struggle in high dimension. Deep neural networks provide scalable solutions for active learning. Deep active learning has been applied to discrete problems such as image classification (Gal et al., 2017) and sequence labeling (Siddhant & Lipton, 2018). The data are queried based on different types of acquisition functions, such as predictive entropy and Bayesian Active Learning by Disagreement (BALD) (Houlsby et al., 2011). Kirsch et al. (2019) further developed BatchBALD, a greedy approach that incrementally selects a set of unlabeled images based on BALD score to issue batch queries for active learning. This batch acquisition function based on BALD is submodular, and therefore its corresponding greedy approach achieves a $1-\frac{1}{e}$ approximation. Similarly, (Li et al., 2020a) propose the optimization-based method DMFAL which is optmization-based and supports mutli-fidelity surrogate modeling, and BMFAL (Li et al., 2022a) uses greedy approach to further extend DMFAL to support batch active-learning.

Task	Setting	DMFAL	NARGP	MFHNP	D-MFDAL
Heat 2	Nested	0.177±2.94e-6	0.313 ±3.47e-6	0.115±8.34e-5	**0.1 ±4.92e-5**
	Non-nested	0.170±1.21e-6	0.311±1.71e-7	0.078±1.02e-4	**0.04 ±6.4e-9**
	Full	0.138 ± 4.0e-8	0.31±2.12e-6	0.026±4.01e-5	**0.015±1.42e-5**
Heat 3	Nested	0.173 ± 1.6e-7	0.311±2.56e-6	0.145±5.11e-5	**0.13 ±2.32e-5**
	Non-nested	0.162±2.35e-6	0.31 ±1.05e-6	0.152±8.86e-5	**0.112±2.06e-5**
	Full	0.137±1.23e-7	0.309±3.46e-6	0.111±4.82e-6	**0.108±4.85e-8**
Poisson 2	Nested	0.179 ± 3.9e-7	0.595±8.71e-8	0.107±7.07e-5	**0.097±5.63e-5**
	Non-nested	0.157±4.56e-5	0.596±1.74e-5	0.102±4.25e-4	**0.084±5.74e-4**
	Full	0.107 ± 6.58e-5	0.585±9.84e-5	0.093±2.55e-4	**0.07 ±2.99e-4**
Possion 3	Nested	0.177±3.99e-5	0.594±6.3e-6	0.281±2.85e-5	**0.126±1.03e-5**
	Non-nested	**0.129±6.51e-5**	0.592±3.77e-5	0.317±8.67e-5	0.131±3.22e-5
	Full	0.121±1.47e-5	0.58 ±1.02e-4	0.335±2.37e-5	**0.101±1.81e-4**
Fluid	Nested	0.294±8.02e-8	0.358±1.26e-3	0.26 ±1.11e-6	**0.21 ±5.13e-6**
	Non-nested	0.331±6.86e-7	0.371±2.41e-3	0.263±1.67e-5	**0.237±3.14e-6**
	Full	0.275±4.59e-7	0.353±9.28e-4	0.234±4.82e-6	**0.207±1.31e-5**

Table 1. Passive learning performance (nRMSE) comparison of 4 different methods applied to the Heat and Poisson simulators with two and three fidelities and fluid simulation with Navier-Stokes equation. Each set of data is restructured into three settings to mimic different stages during active learning.

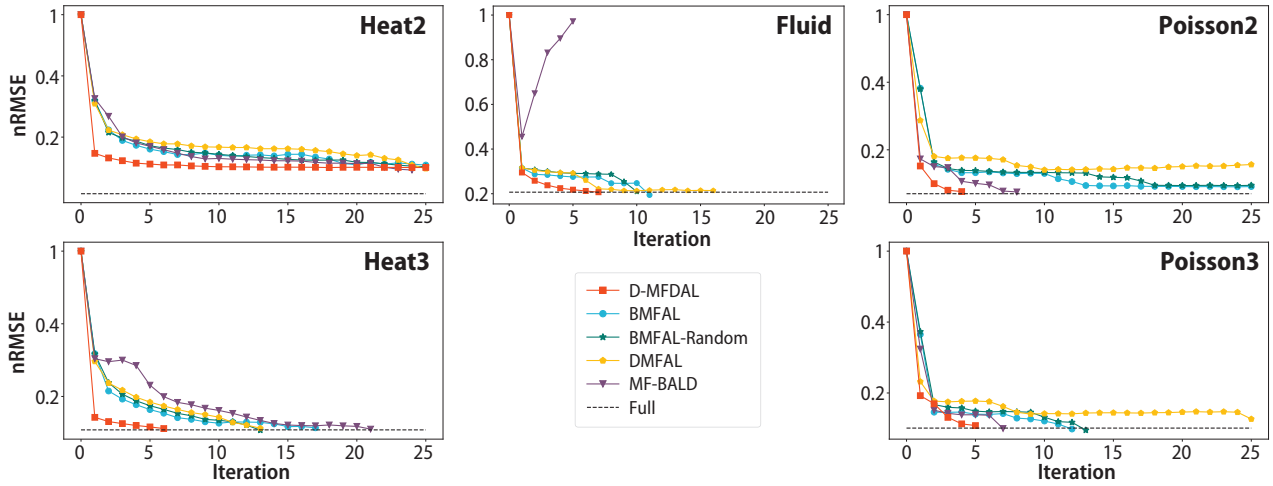

Figure 4. Active learning performance comparison for Heat and Poisson simulation with two and three fidelity levels, fluid simulation with two fidelity levels using Navier-Stokes equation. Performance is measured at the highest fidelity level.

5. Experiments

5.1. Datasets

We evaluate our methods on learning surrogate models of partial differential equations (PDE) benchmark, and a more complex fluid dynamics prediction task.

Partial Differential Equations. We include 4 benchmark tasks in computational physics. The goal is to predict the spatial solution fields of 2 PDEs, including Heat and Poisson's equations (Olsen-Kettle, 2011). The ground-truth data is generated from the numerical solver. High-fidelity and low-fidelity examples are generated by solvers running with dense and coarse meshes, respectively. The output dimension is the same as the flattened mesh points. For both Heat and Poisson's equation with two-fidelity setting, they have 16×16 meshes at low fidelity level and 32×32 meshes at high fidelity level. For three-fidelity setting, they both have additional 64×64 meshes at the highest fidelity level. We calculate the relative cost of querying at each fidelity level c_k based on the averaged computation time for data generation. We always set $c_1 = 1$ as a reference.

Fluid Simulation. We also test D−MFDAL on a more challenging fluid dynamics simulation task. This computationally challenging simulation is based on the Navier-Stokes equation and the Boussinesq approximation (Holl et al., 2020). We obtain the ground truth data by simulating the velocity field of smoke dynamics in a 50×50 grid. Initially, a static incompressible smoke cloud of radius 5 is placed at the lower center of the domain together with a consistent inflow force is applied to the center at the initial position of the smoke. The inflow force varies in magnitude and direction for different scenarios. The two-dimensional input controls the magnitude of the inflow force at x and y directions. The output is the first component of the velocity field by applying the inflow for 30 time stamps. We simulated the low fidelity ground truth with a 32×32 mesh and high fidelity with a 64×64 mesh.

5.2. Experiment Setup

We consider two groups of experiments:

- **Passive Learning**: model accuracy and robustness test by comparing the performance between D−MFDAL versus other baseline models using the entire training dataset.
- **Active Learning**: budget-constrained batch multi-fidelity active learning comparison between D−MFDAL with the MF-LIG acquisition function versus other multi-fidelity active learning frameworks.

For passive learning, we evaluate the performance of our model under three settings: nested, non-nested, and full. Let $\mathcal{X}_1$ and $\mathcal{X}_2$ to be two training input sets at 2 fidelity levels. The "full" setting means that $\mathcal{X}_1 = \mathcal{X}_2$ and both sets have a large number of scenarios uniformly distributed in the input space, mimicking the final and convergent stage of active learning. The "nested" setting means that $\mathcal{X}_2 \subset \mathcal{X}_1$ and the "nonnested" settings means that $\mathcal{X}_1 \wedge \mathcal{X}_q = \mathcal{X}^r$, where $\mathcal{X}^r$ includes the inputs for the reference set. These two settings are used to mimic the early stage of active learning where the number of low-fidelity data points is much larger than the high-fidelity data points. We use these three settings to test the robustness of D−MFDAL and other baselines. For comparison, we consider state-of-the-art baselines for multi-fidelity surrogate modeling, including DMFAL (Li et al., 2020a), NARGP (Perdikaris et al., 2017), and MFHNP (Wu et al., 2022).

For active learning, we use the same 8 uniformly sampled data points across all fidelity levels as the reference data for initial training. We run 25 iterations and at each iteration, the active learning framework queries the simulator for the input with the highest acquisition function score until it reaches the budget limit of 20 per iteration. We compare our method against DMFAL (Li et al., 2020a), BMFAL-Random (Li et al., 2022a), BMFAL (Li et al., 2022a) and MF-BALD (Gal et al., 2017) as baselines, using the same hyperparameter settings as in the literature.

For both passive and active learning, we randomly generate 512 data points as the test set for 4 benchmark tasks and 256 data points as the test set for fluid simulation. We use the normalized Root Mean Squared Error (nRMSE) to measure prediction performance at the highest fidelity level, as our goal is to mimic the dynamics at the highest fidelity level. All experiment results are averaged over 3 random runs.

Our code is available at https://github.com/Rose-STL-Lab/Multi-Fidelity-Deep-Active-Learning.

5.3. Experimental Results

Passive Learning Performance. We test the passive learning performance of D−MFDAL and baselines across 5 tasks and 3 settings. The results are shown in Table 1. It can be seen that our model consistently outperforms all baselines across all settings and tasks. Furthermore, D−MFDAL performs particularly well under challenging nested and disjoint settings where the number of training data available at the highest fidelity level is limited. For example, in the complex fluid simulation, we find D−MFDAL with only 8 data points at the high fidelity level under the nested setting outperforming all other baselines in the full setting.

The results show that D−MFDAL is capable of utilizing the information from the low fidelity levels to make good predictions at the highest fidelity level. D−MFDAL is also quite robust as it almost has the best model performance under all three representative active learning settings. These advantages show that D−MFDAL is suitable for Bayesian active learning throughout the training process.

Active Learning Performance. Figure 4 shows the nRMSE versus the number of iterations in active training. Our proposed D−MFDAL with MF-LIG always has the best nRMSE performance throughout the active learning process. Furthermore, D−MFDAL converges to offline performance iterations faster than all other baselines for the Poisson2, Poisson3, Heat3 and Fluid experiments. Figure 5 is the visualization of prediction residuals for D−MFDAL, as well as 4 other baselines. We visualize the residual between the predictions and the truth to highlight the performance difference across 5 datasets. A higher residual value indicates lower accuracy. We randomly select 3 samples from the test set for each task. It can also be found that D−MFDAL with MF-LIG outperforms other baselines as it successfully predicts the true patterns among all 15 samples.

Ablation Study. In Figure 6, we compare active learning performance at 3 fidelity levels on the Heat3 dataset. We

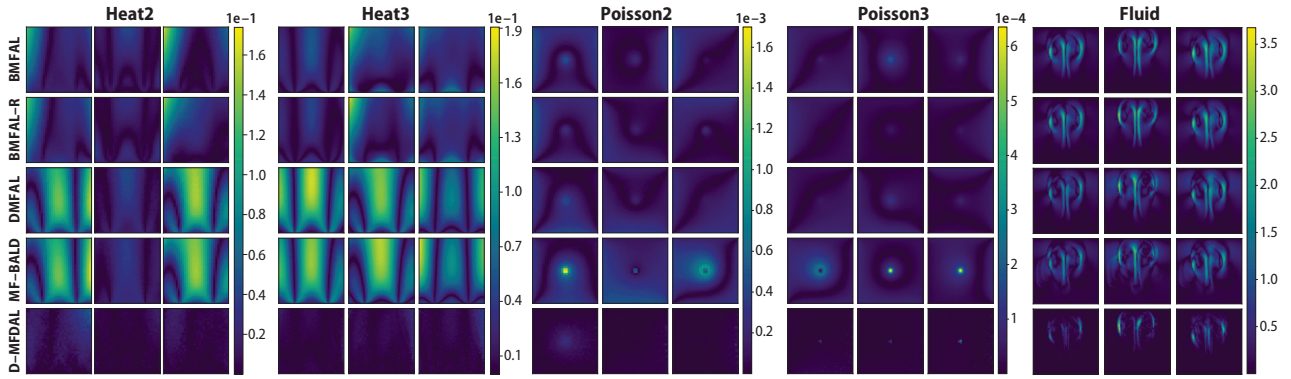

Figure 5. Prediction residual visualizations at the highest fidelity level for D−MFDAL and 4 baselines for Heat equation and Poisson's equation simulation with two and three fidelity levels, fluid simulation with two fidelity levels. For each simulation scheme presented, we randomly select three samples to visualize. Better performance is indicated by a darker color.

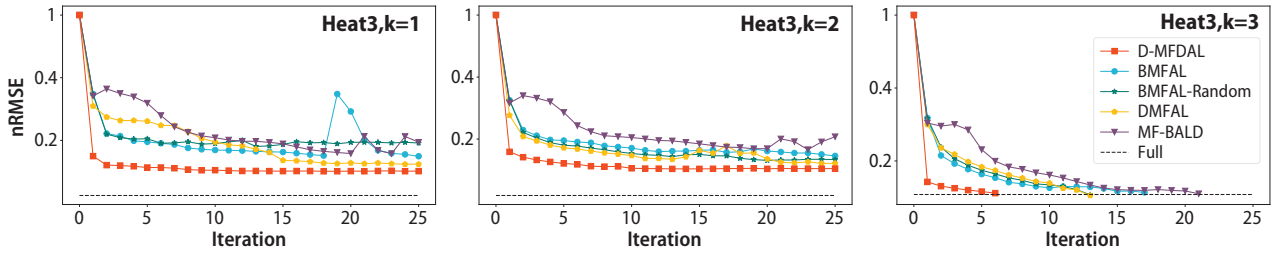

Figure 6. Active learning performance comparison for Heat3 simulation at three fidelity levels. Performance is measured at each fidelity level. k represents the fidelity level. D−MFDAL outperforms the baselines across all fidelity levels.

find that the performance of D−MFDAL is always the best at each fidelity level, although the MF-LIG is designed to optimize the surrogate modeling performance at the highest fidelity level. Specifically, we find that the performance gap between D−MFDAL and the other baselines is consistently evident across all active learning iterations and fidelity levels. It shows one of the other advantages of our proposed D−MFDAL. That is, we can utilize the data at the high fidelity level to reversely improve the model performance at the low fidelity level. Although it is not the goal to improve surrogate modeling performance at lower fidelity levels in our tasks, it makes D−MFDAL flexible to be applied to general setups such as multi-task surrogate modeling where multiple tasks are considered.

6. Conclusion

To conclude, we design a multi-fidelity deep active learning framework, D−MFDAL, to learn functional relationships across multiple fidelity levels. D−MFDAL disentangles the individual latent representations, separating them into global and local terms to tackle issues of error propagation and overfitting. We design a unified ELBO over the joint distribution across all fidelity levels to serve as the training loss and include a multi-fidelity regularization term to infer the global representations across different levels of fidelity. Additionally, we generalize the acquisition function, latent information gain, used in Bayesian active learning for NP-based models to multi-fidelity settings and design an efficient algorithm for budget-constrained batch active learning. We conduct extensive empirical evaluations on several benchmark studies and complex spatiotemporal simulations to demonstrate the superior performance of our proposed D−MFDAL for both passive learning and active learning. For future work, we plan to extend this method for multi-task active learning.

7. Acknowledgments

This work was supported in part by U.S. Department Of Energy, Office of Science, Facebook Data Science Research Awards, U. S. Army Research Office under Grant W911NF-20-1-0334, and NSF Grants #2134274 and #2146343, as well as NSF-SCALE MoDL (2134209) and NSF-CCF-2112665 (TILOS). M.C. acknowledges support from grant HHS/CDC 5U01IP0001137.

References

Brevault, L., Balesdent, M., and Hebbal, A. Overview of gaussian process based multi-fidelity techniques with variable relationship between fidelities, application to aerospace systems. *Aerospace Science and Technology*, 107:106339, 2020.

Chaloner, K. and Verdinelli, I. Bayesian experimental design: A review. *Statistical Science*, pp. 273–304, 1995.

Cohn, D. A., Ghahramani, Z., and Jordan, M. I. Active learning with statistical models. *Journal of artificial intelligence research*, 4:129–145, 1996.

Cutajar, K., Pullin, M., Damianou, A., Lawrence, N., and González, J. Deep gaussian processes for multi-fidelity modeling. *arXiv preprint arXiv:1903.07320*, 2019.

De, S., Britton, J., Reynolds, M., Skinner, R., Jansen, K., and Doostan, A. On transfer learning of neural networks using bi-fidelity data for uncertainty propagation. *International Journal for Uncertainty Quantification*, 10(6), 2020.

Gal, Y., Islam, R., and Ghahramani, Z. Deep bayesian active learning with image data. In *International Conference on Machine Learning*, pp. 1183–1192. PMLR, 2017.

Garnelo, M., Rosenbaum, D., Maddison, C., Ramalho, T., Saxton, D., Shanahan, M., Teh, Y. W., Rezende, D., and Eslami, S. A. Conditional neural processes. In *International Conference on Machine Learning*, pp. 1704–1713. PMLR, 2018a.

Garnelo, M., Schwarz, J., Rosenbaum, D., Viola, F., Rezende, D. J., Eslami, S., and Teh, Y. W. Neural processes. *arXiv preprint arXiv:1807.01622*, 2018b.

Guo, M., Manzoni, A., Amendt, M., Conti, P., and Hesthaven, J. S. Multi-fidelity regression using artificial neural networks: efficient approximation of parameter-dependent output quantities. *Computer methods in applied mechanics and engineering*, 389:114378, 2022.

Hebbal, A., Brevault, L., Balesdent, M., Talbi, E.-G., and Melab, N. Multi-fidelity modeling with different input domain definitions using deep gaussian processes. *Structural and Multidisciplinary Optimization*, 63(5):2267–2288, 2021.

Holl, P., Koltun, V., and Thuerey, N. Learning to control pdes with differentiable physics. *arXiv preprint arXiv:2001.07457*, 2020.

Hosking, S. Multifidelity climate modelling, github. `https://github.com/scotthosking/mf_modelling`, 2020.

Houlsby, N., Huszár, F., Ghahramani, Z., and Lengyel, M. Bayesian active learning for classification and preference learning. *arXiv preprint arXiv:1112.5745*, 2011.

Jha, S., Gong, D., Wang, X., Turner, R. E., and Yao, L. The neural process family: Survey, applications and perspectives. *arXiv preprint arXiv:2209.00517*, 2022.

Kandasamy, K., Dasarathy, G., Schneider, J., and Póczos, B. Multi-fidelity bayesian optimisation with continuous approximations. In *International Conference on Machine Learning*, pp. 1799–1808. PMLR, 2017.

Kennedy, M. C. and O'Hagan, A. Predicting the output from a complex computer code when fast approximations are available. *Biometrika*, 87(1):1–13, 2000.

Kim, H., Mnih, A., Schwarz, J., Garnelo, M., Eslami, A., Rosenbaum, D., Vinyals, O., and Teh, Y. W. Attentive neural processes. In *International Conference on Learning Representations*, 2018.

Kim, H., Mnih, A., Schwarz, J., Garnelo, M., Eslami, A., Rosenbaum, D., Vinyals, O., and Teh, Y. W. Attentive neural processes. *arXiv preprint arXiv:1901.05761*, 2019.

Kirsch, A., Van Amersfoort, J., and Gal, Y. Batchbald: Efficient and diverse batch acquisition for deep bayesian active learning. *Advances in neural information processing systems*, 32, 2019.

Le Gratiet, L. and Garnier, J. Recursive co-kriging model for design of computer experiments with multiple levels of fidelity. *International Journal for Uncertainty Quantification*, 4(5), 2014.

Li, S., Kirby, R. M., and Zhe, S. Deep multi-fidelity active learning of high-dimensional outputs. *arXiv preprint arXiv:2012.00901*, 2020a.

Li, S., Xing, W., Kirby, R., and Zhe, S. Multi-fidelity bayesian optimization via deep neural networks. *Advances in Neural Information Processing Systems*, 33: 8521–8531, 2020b.

Li, S., Phillips, J., Yu, X., Kirby, R., and Zhe, S. Batch multi-fidelity active learning with budget constraints. In *Advances in Neural Information Processing Systems*, 2022a.

Li, S., Wang, Z., Kirby, R., and Zhe, S. Deep multi-fidelity active learning of high-dimensional outputs. In *International Conference on Artificial Intelligence and Statistics*, pp. 1694–1711. PMLR, 2022b.

Louizos, C., Shi, X., Schutte, K., and Welling, M. The functional neural process. *Advances in Neural Information Processing Systems*, 2019.

Meng, X. and Karniadakis, G. E. A composite neural network that learns from multi-fidelity data: Application to function approximation and inverse pde problems. *Journal of Computational Physics*, 401:109020, 2020.

Meng, X., Babaee, H., and Karniadakis, G. E. Multi-fidelity bayesian neural networks: Algorithms and applications. *Journal of Computational Physics*, 438:110361, 2021.

Øksendal, B. Stochastic differential equations. In *Stochastic differential equations*, pp. 65–84. Springer, 2003.

Olsen-Kettle, L. Numerical solution of partial differential equations. *Lecture notes at University of Queensland, Australia*, 2011.

Peherstorfer, B., Willcox, K., and Gunzburger, M. Survey of multifidelity methods in uncertainty propagation, inference, and optimization. *Siam Review*, 60(3):550–591, 2018.

Perdikaris, P., Venturi, D., Royset, J. O., and Karniadakis, G. E. Multi-fidelity modelling via recursive co-kriging and gaussian–markov random fields. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 471(2179):20150018, 2015.

Perdikaris, P., Venturi, D., and Karniadakis, G. E. Multifidelity information fusion algorithms for high-dimensional systems and massive data sets. *SIAM Journal on Scientific Computing*, 38(4):B521–B538, 2016.

Perdikaris, P., Raissi, M., Damianou, A., Lawrence, N. D., and Karniadakis, G. E. Nonlinear information fusion algorithms for data-efficient multi-fidelity modelling. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 473(2198):20160751, 2017.

Perry, D. J., Kirby, R. M., Narayan, A., and Whitaker, R. T. Allocation strategies for high fidelity models in the multifidelity regime. *SIAM/ASA Journal on Uncertainty Quantification*, 7(1):203–231, 2019.

Raissi, M. and Karniadakis, G. Deep multi-fidelity gaussian processes. *arXiv preprint arXiv:1604.07484*, 2016.

Siddhant, A. and Lipton, Z. C. Deep bayesian active learning for natural language processing: Results of a large-scale empirical study. *arXiv preprint arXiv:1808.05697*, 2018.

Singh, G., Yoon, J., Son, Y., and Ahn, S. Sequential neural processes. *Advances in Neural Information Processing Systems*, 32:10254–10264, 2019.

Valero, M. M., Jofre, L., and Torres, R. Multifidelity prediction in wildfire spread simulation: Modeling, uncertainty quantification and sensitivity analysis. *Environmental Modelling & Software*, 141:105050, 2021.

Volpp, M., Flürenbrock, F., Grossberger, L., Daniel, C., and Neumann, G. Bayesian context aggregation for neural processes. In *International Conference on Learning Representations*, 2020.

Wang, Q. and Van Hoof, H. Doubly stochastic variational inference for neural processes with hierarchical latent variables. In *International Conference on Machine Learning*, pp. 10018–10028. PMLR, 2020.

Wang, Y. and Lin, G. Mfpc-net: Multi-fidelity physics-constrained neural process. *arXiv preprint arXiv:2010.01378*, 2020.

Wang, Z., Xing, W., Kirby, R., and Zhe, S. Multi-fidelity high-order gaussian processes for physical simulation. In *International Conference on Artificial Intelligence and Statistics*, pp. 847–855. PMLR, 2021.

Wu, D., Chinazzi, M., Vespignani, A., Ma, Y.-A., and Yu, R. Multi-fidelity hierarchical neural processes. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 2029–2038, 2022.

Wu, D., Niu, R., Chinazzi, M., Vespignani, A., Ma, Y.-A., and Yu, R. Deep bayesian active learning for accelerating stochastic simulation. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2023.

Zimmer, C., Meister, M., and Nguyen-Tuong, D. Safe active learning for time-series modeling with gaussian processes. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, pp. 2735–2744, 2018.