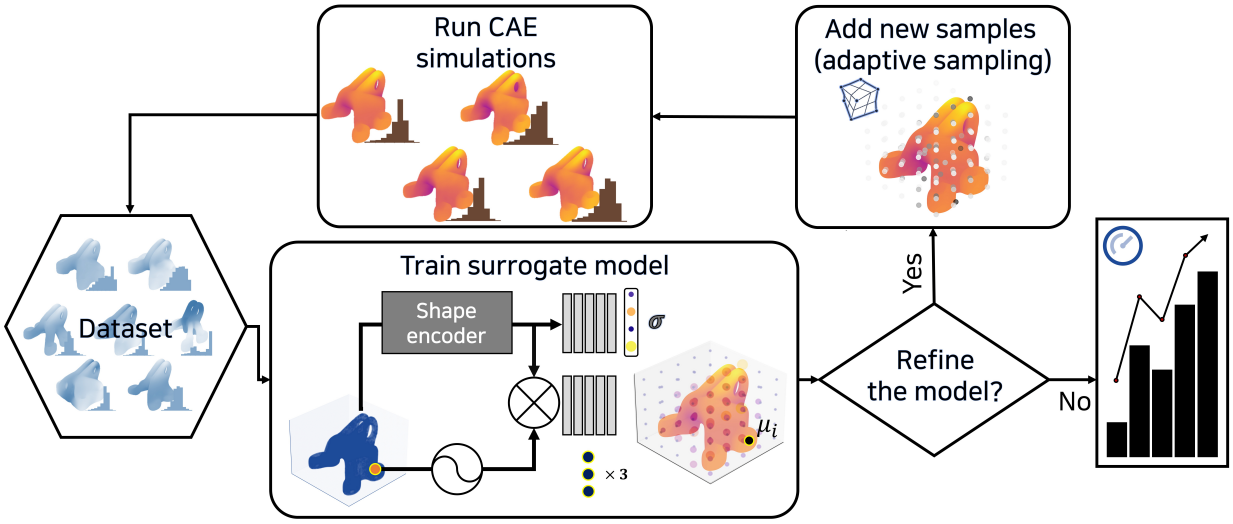
Graphical Abstract

Three-dimensional Geometric Augmentation Guided by Engineering Uncertainty

Yongmin Kwon, Namwoo Kang



Highlights

Three-dimensional Geometric Augmentation Guided by Engineering Uncertainty

Yongmin Kwon, Namwoo Kang

- Proposes a framework integrating uncertainty quantification and free-form deformation.
- Features an explicit lattice-based deep ensemble uncertainty quantification architecture.
- Offers performance-aware three-dimensional shape augmentation.
- Leverages epistemic uncertainty to guide active exploration.

Three-dimensional Geometric Augmentation Guided by Engineering Uncertainty

Yongmin Kwon[a,b], Namwoo Kang[a,b,*]

[a] *Cho Chun Shik Graduate School of Mobility, Korea Advanced Institute of Science and Technology, 193, Munji-ro, Yuseong-gu, 34051, Daejeon, Republic of Korea*

[b] *Narnia Labs, 193, Munji-ro, Yuseong-gu, 34051, Daejeon, Republic of Korea*

Abstract

In the conceptual design stage, fast and accurate shape evaluation is crucial for efficient product development. While simulation-based performance evaluation is commonly adopted, its high computational cost has motivated the development of more efficient meta-models that leverage deep learning for rapid inference and effective handling of nonlinear, high-dimensional data. However, acquiring large-scale three-dimensional (3D) datasets in industry and academia is challenging. Augmentation-based active learning emerges as a viable alternative. However, mere geometric diversity does not guarantee corresponding performance diversity, underscoring the need for data augmentation that considers both geometry and performance. This study addresses this challenge by proposing an integrated framework that combines geometry-based uncertainty quantification with free-form deformation augmentation. Uncertainty quantification (UQ) is achieved using a deep ensemble model that incorporates a novel architecture comprising two regressors: an explicit lattice-based uncertainty regressor to quantify aleatoric uncertainty and a performance regressor to predict performance. The epistemic uncertainty extracted from this model is then leveraged to actively explore the regions effective for the practical training of a meta-model. To simulate a common industrial data-scarce scenario, meta-models were initially built with small subsets of the public three-dimensional bracket dataset and an automotive aerodynamic dataset, and the datasets were subsequently augmented with samples generated by the framework. This approach employs an active learning framework that accounts for both geometric and performance diversity. Experimental results confirmed that the augmented datasets exhibit superior performance, demonstrating the framework's robustness across diverse physical domains (structural mechanics and fluid dynamics) without requiring domain-specific modifications. Performance-aware 3D shape augmentation has strong potential for industrial applications.

Keywords:
data augmentation, artificial intelligence, active learning, uncertainty quantification

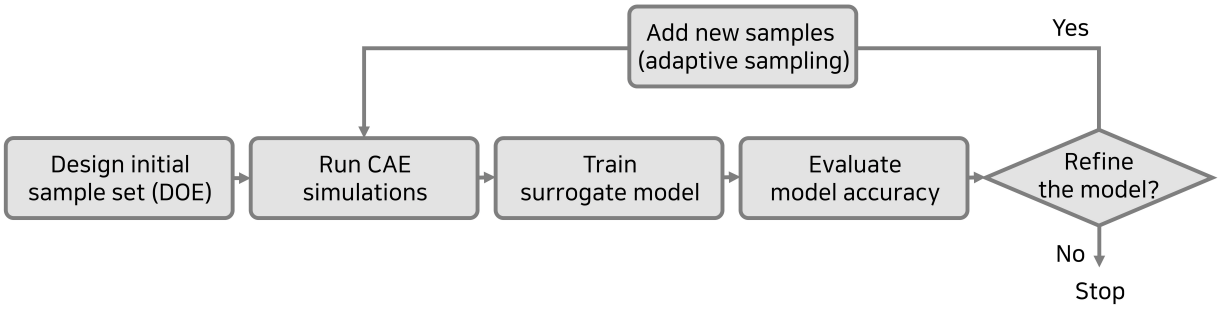


Figure 1: Meta-modeling workflow

1. Introduction

Performance evaluation for shapes is the most crucial area in the conceptual design stage of industrial product development [1, 2]. Numerical analysis methods such as Finite Element Method (FEM) and Computational Fluid Dynamics (CFD) have emerged to address this, significantly reducing time and cost compared to traditional physical experiments. However, performance evaluation through numerical analysis remains the most time-intensive and costly step in the overall design process. This high computational expense creates a significant bottleneck, as illustrated by the standard meta-modeling workflow in Fig. 1.

This workflow typically begins with designing an initial sample set using the design of experiments. Each sample in this set must then be evaluated using high-fidelity numerical analyses (e.g., FEM/CFD) to obtain performance labels. A surrogate model is then trained on this initial data, and its accuracy is assessed. However, this is rarely a single-step process. To achieve sufficient accuracy, the model must be refined, triggering a costly iterative loop. This refinement loop requires adding new samples (adaptive sampling), and crucially, each new sample requires running another full, high-cost Computer-Aided Engineering (CAE) simulation. This iterative dependency on the 'Run CAE simulations' step is precisely what constitutes the significant bottleneck in design optimization.

Research into data-driven meta-models has advanced to address the time and cost limitations of numerical analysis [3, 4, 5]. Among various data-driven meta-models, deep learning is the field where research is progressing most actively. Deep learning offers several advantages over traditional meta-models, particularly its suitability for complex engineering problems.

First, deep learning eliminates the need to explicitly parameterize input shapes. Conventionally, design optimization using a meta-model begins by defining the shape through a set of parameters. Defining data with a parameter set that encompasses all possible shapes requires substantial effort and can introduce subjective design elements into the parameter definition process. In contrast, deep learning automatically extracts features from input data and optimizes internal model parameters, thereby eliminating the need to separately parameterize a meta-model.

*Corresponding author

Second, deep learning models can predict high-dimensional outputs, such as full field values composed of nodal values, and can also predict field values across the time domain. For instance, a deep learning-based meta-model can predict the complete stress distribution across every node of a complex 3D structure or precise pressure and velocity fields at every point within a fluid domain [6, 7]. This capability extends to demonstrating excellent predictive performance in spatio-temporal problems. The ability to handle high-dimensional data isn't limited to prediction; it also excels in readily accommodating UQ for predicted values at each node [8, 9].

Finally, deep learning provides superior capabilities in processing complex data and modeling nonlinear relationships. It effectively learns complex nonlinear mappings, making it well-suited for simulations involving 3D shapes or unstructured grids. Unlike traditional reduced-order methods that require predefined bases, deep learning's universal approximation ability enables the end-to-end learning of key features from data, resulting in high generalization and flexibility for new designs or extrapolation [10, 11].

Despite these advantages, most research on data-driven meta-models, particularly deep learning studies, has predominantly focused on two-dimensional (2D) data, such as metamaterials, rather than the 3D Computer-Aided Design (CAD) data commonly used in industrial settings. There is an absolute lack of high-quality 3D engineering datasets for research purposes. Conversely, 2D images or structural data, which are relatively easy to create and measure, are readily available, leading to their greater use in research.

Moreover, 3D data involves diverse representation methods (e.g., meshes, point clouds, or voxels), and measuring or generating physical shapes is complex and costly, making it very difficult to acquire large, diverse datasets. Existing research leveraging 3D engineering data is limited to a few cases that utilize benchmark datasets [7, 12, 13, 14, 15]. Existing research leveraging 3D engineering data is limited to a few cases that utilize benchmark datasets. However, these research attempts are confined mainly to limited datasets, which makes it challenging to generate a wide variety of shapes. Recently, studies utilizing foundation models are also underway; however, generating feasible data from an engineering perspective remains a challenge. Even these benchmark datasets are mostly acquired not by designing and simulating a wide variety of new shapes from scratch, but primarily by augmenting a small number of representative seed data through specific deformations or parameter variations [14, 16, 17, 18]. Similarly, in industrial practice, the standard method for complex shapes, such as ship hulls or aircraft wings, is to generate and augment data by manipulating predefined basic shape parameters [19, 20, 21]. Methods such as Space Filling Sampling are predominantly used to expand the data within the given parameterized space to find optimal combinations [20].

However, for 3D data explicitly defined by meshes, unlike simple parameterized shapes, their high dimensionality and interdependent nodes make them prone to generating infeasible shapes when sampled randomly or with standard augmentation techniques. To address this challenge and effectively generate valid data candidates for meta-modeling, parameterization using Free Form Deformation (FFD) is widely employed [7, 13, 22, 23]. FFD maps complex 3D shapes to a lower-dimensional space via a control lattice, allowing intuitive manipulation using control points. This process generates feasible, smoothly varying shape candidates

while preserving structural validity, thereby mitigating the issue of unrealistic shapes that can arise from arbitrary sampling. FFD effectively maps complex 3D shape data to a relatively low-dimensional parameter space defined by the control lattice of a predefined (N, N, N) size and the coordinate shifts of the corresponding control points. It has the advantage of efficiently controlling the entire shape by manipulating a small number of control points and intuitively representing nonlinear shape deformations. Moreover, understanding and applying shape deformation via control-point movement helps engineers easily interpret optimization or model-prediction results and apply them to actual designs. Furthermore, recent research is actively exploring hybrid approaches that combine the structural and controllable deformation capabilities provided by FFD with the complex-data-distribution learning and novel data-generation capabilities of deep learning-based generative models, enabling more automated and precise shape generation and optimization [21].

In summary, while data augmentation methods like FFD have been widely utilized, a critical flaw identified in current research lies in their conventional application. Most existing studies focus primarily on shape diversity while neglecting the performance information of existing seed data [24, 7, 13, 22, 23]. This represents a significant oversight, as effective engineering design data generation requires consideration of the diversity of both design and performance spaces [25, 26]. Consequently, to construct a high-performance, generalized meta-model, there is a distinct need for data augmentation methodologies that actively incorporate performance information.

This creates a critical need for an intelligent augmentation strategy that actively seeks out data that will most improve the meta-model. This strategy is known as active learning. However, a fundamental challenge remains: how to identify which new shapes will be the most informative?

A model might be highly confident about some new shapes (which are redundant or easy to predict) but highly uncertain about others. Wasting expensive CAE simulations on these easy shapes is the very inefficiency this study aims to solve. This is the core motivation for adopting UQ in this framework.

UQ provides a formal methodology for quantifying the model's ignorance—specifically, epistemic uncertainty, which represents a lack of knowledge due to insufficient data. By quantifying this uncertainty, a metric is gained to guide the FFD augmentation process. UQ is not merely an add-on; it is the essential component that enables a truly performance-aware active learning loop, guiding the framework on where to explore the design space to find the most valuable data points for model training.

This study proposes a new data augmentation and model training framework that combines UQ-guided active learning using FFD to efficiently and accurately construct meta-models for complex engineering shapes. The proposed framework operates as follows: First, using the currently trained meta-model, the prediction results and prediction uncertainty for each generated shape candidate are quantified. Next, based on the quantified uncertainty information at the FFD control points, exploration is performed in areas expected to significantly improve performance from the meta-model's perspective, thereby augmenting the data. Finally, numerical analysis simulations (labeling) are performed on the shapes

augmented through FFD. This new data is added to the dataset, and the meta-model is retrained. This active-learning-based sampling goes beyond the simple FFD augmentation, which considers only the existing shape space. It acquires the most efficient data for model learning by considering the engineering performance space through the UQ. As a result, this framework enables the efficient construction of a highly accurate meta-model with minimal execution of high-cost simulations. To rigorously validate the effectiveness of the framework under these exact conditions, the experiments were designed to simulate an initialization from a data-sparse regime. This approach is demonstrated by constructing an initial meta-model from a very small subset of data (50 samples), mimicking a typical starting point in the conceptual design stage where high-fidelity simulations are strictly budget-constrained.

Key contributions can be summarized as follows:

1) Seamless UQ Architecture Tailored for Data Augmentation:

This study proposes an engineering performance-guided data augmentation technique. To facilitate this, an explicit lattice-based UQ methodology is introduced, providing the necessary information concerning engineering performance. This UQ methodology seamlessly integrates with the FFD approach employed in this research, demonstrating strong predictive performance.

2) Uncertainty-driven Data Augmentation:

In this study, uncertainty signifies the reliability of the predicted performance for a given shape. Specifically, regions with high epistemic uncertainty indicate insufficient data in the training dataset, suggesting opportunities to improve model performance through data augmentation. By aligning the control points of the FFD with the lattice-based UQ regressor's coordinates, the proposed method leverages meta-model uncertainty to augment existing seed data, thereby achieving superior performance.

This work offers a comprehensive, data-centric solution that integrates uncertainty-driven active learning with FFD-based augmentation to overcome data scarcity challenges in engineering applications.

This paper is organized as follows: Section 2 provides background explanations on uncertainty quantification, shape morphing, and active learning techniques used in this study. Section 3 describes the comprehensive methodology of the framework proposed in this study. Section 4 explains the experimental setup and presents the results and their analysis to verify the effectiveness of the proposed framework. Section 5 summarizes this study and discusses future work.

2. Related Works

2.1. Shape Morphing

Shape morphing refers to the continuous transformation from one shape to another and is utilized as a data augmentation technique in various domains. Unlike simple geometric transformations (e.g., rotation, scaling), shape morphing can generate more diverse data samples by including structural changes such as shape distortion and deformation of detailed

structures. In data augmentation using shape morphing, FFD and Radial Basis Function (RBF) interpolation are frequently employed as deformation methods.

FFD is a widely used technique for manipulating the shape of an object by deforming the space in which the object is embedded [27, 28]. The object itself is conceptually enclosed within a deformable control lattice, typically a cuboid defined by a grid of control points. Deforming this control lattice then smoothly and implicitly deforms the object within it.

Mathematically, the deformation maps an original point $\mathbf{x}$ within the control lattice, represented by its local coordinates (u, v, w) within the $[0,\ 1]^3$ normalized space of the lattice, to its new position $\mathbf{x}'$ in space. For a control lattice with $(L+1) \times (M+1) \times (N+1)$ control points, this mapping is expressed as a trivariate tensor product of Bernstein polynomials, weighted by the displaced positions of the control points. The deformed position $\mathbf{x}'(u, v, w)$ is given by:

$$\mathbf{x}'(u, v, w) = \sum_{l=0}^{L} \sum_{m=0}^{M} \sum_{n=0}^{N} B_l(u)\, B_m(v)\, B_n(w)\, \mathbf{P}_{lmn} \tag{1}$$

$\mathbf{x}'(u, v, w)$ represents the deformed position of a point originally at local coordinates (u, v, w) within the lattice space. The terms $B_l(u)$, $B_m(v)$, and $B_n(w)$ represent the Bernstein basis functions defined over the local parameters u, v, and w, respectively, which determine the weight of each control point's influence. $\mathbf{P}_{lmn}$ denotes the coordinates of the control points of the grid after they have been adjusted to define the deformation. The formula computes the deformed position as a weighted sum, with weights determined by the point's local position and the control points' positions. This formulation enables versatile shape transformations, ranging from simple geometric changes to complex structural modifications, by adjusting the control points $\mathbf{P}_{lmn}$.

RBF is a powerful technique for performing continuous deformations via spatial interpolation based on specific points [29, 30, 31, 32]. RBF defines deformations at designated control points and then propagates this transformation across the entire set of data points within the domain. This approach offers smooth, nonlinear deformations, making it especially suitable for complex data transformations. In particular, the transformation $\mathbf{T}(\mathbf{x})$ for an original point $\mathbf{x}$ can be written as the original position plus a weighted sum of radial basis functions centered at the control points:

$$\mathbf{T}(\mathbf{x}) \;=\; \mathbf{x} \;+\; \sum_{i=1}^{N} w_i\, \phi\big(|\mathbf{x} - \mathbf{c}_i|\big) \tag{2}$$

where $\mathbf{x}$ denotes the original coordinates, $\mathbf{c_i}$ are the center points (typically the control points where deformation is specified), w_i are the associated weights, and $\phi(r)$ is the chosen radial basis function, with $r = |\mathbf{x} - \mathbf{c_i}|$ representing the Euclidean distance between the point $\mathbf{x}$ and the center $\mathbf{c_i}$. The distance computation within ϕ can be adjusted depending on the problem requirements. Common choices for the radial basis function $\phi(r)$ include thin-plate spline ($r^2 \log(r)$), Gaussian (e^{-cr^2}), polyharmonic splines (r^k), and multiquadrics ($\sqrt{r^2 + c^2}$).

Moreover, RBF is well-suited for local deformations, as the influence of each center point $\mathbf{c_i}$ decays with distance, inducing smooth transitions throughout the domain. This technique allows for precise adjustments by directly manipulating the control points $\mathbf{c_i}$, where the corresponding weights w_i are typically determined by solving a linear system of equations that ensures the transformation maps the control points to their desired deformed locations.

Such deformation methods are highly versatile because they can easily modify explicit representations, such as meshes. Many studies have leveraged FFD [33, 34, 35, 36, 37] or RBF [29, 30, 31, 32, 38] to transform seed geometries into parameterized shapes, enabling morphing-based augmentation through design-of-experiments approaches such as Latin Hypercube Sampling (LHS). In particular, several works combine dimension-reduction models with optimization algorithms to achieve high-performance shape transformations. Morphing techniques are especially effective because they can generate feasible data even from a limited number of seed data points. With the rapid advancement of data-driven methodologies, recent research has increasingly focused on using morphing methods for data augmentation [7, 13]. Integrating more sophisticated sampling strategies is becoming crucial.

However, these morphing techniques, particularly FFD, are not without limitations. Their primary drawback is that they are generally restricted to local deformations and design refinement, rather than generating entirely new global shapes or topologies.

More importantly, a critical flaw identified in current research lies not just in the methods' limitations, but in their conventional application. Most existing morphing studies focus primarily on shape diversity, neglecting the crucial consideration of performance information. This is a significant oversight because, for effective engineering design data generation, it is necessary to consider the diversity of both design and performance spaces [25, 26].

Since performance information serves as the meta-model's output, its distribution and characteristics influence the model's overall accuracy and robustness just as much as the distribution of the input shape data. Therefore, a performance-aware approach based on model uncertainty is essential for efficiently building an accurate meta-model that is sensitive to critical performance regions. Without incorporating such considerations into the sampling process, the resulting dataset may lead to issues like overfitting or inefficient learning in these key areas.

Simply focusing on sampling diverse shapes in the design space—often done blindly using standard sampling like LHS on the FFD control points—makes it challenging to identify where the meta-model struggles with predictions (i.e., exhibits high uncertainty) or which areas of the performance space require further learning to improve model fidelity. This analysis reveals a clear research gap: while FFD provides an excellent mechanism for deforming shapes, it lacks an intelligent guide. There is a clear need for a framework that directs the FFD process, moving it beyond random geometric exploration to one that is actively guided by the meta-model's performance and uncertainty.

2.2. Uncertainty Quantification

Engineers across various fields are increasingly leveraging data-driven methods to address complex engineering challenges. Neural networks, possessing a universal approximation ca-

pability [10, 11], offer significant flexibility in handling intricate nonlinear relationships and processing high-dimensional data. These challenges have long served as bottlenecks in traditional approaches. Such data-driven techniques are applicable to the exploratory phases of design optimization and extend to tasks such as dimensionality reduction and model order reduction for high-dimensional predictive modeling.

Although regression models offer substantial potential as surrogates for costly simulations, their data-driven nature inherently introduces uncertainty into their predictions. In real engineering applications, ensuring the interpretability of model predictions and quantifying the associated uncertainty is indispensable. Failure to adequately consider the uncertainties arising from limited training data or model imperfections can lead to unforeseen challenges in risk assessment and management. Beyond simple risk assessment, UQ's primary role in data-driven engineering is to enable efficient data acquisition. In the context of active learning (as detailed in Sec. 2.3), UQ provides the query strategy. It is the tool that answers the critical question: "Given a limited budget, which data point should be labeled next?". Recent advances have led to the development of scalable yet straightforward approaches to approximate Bayesian inference, which have garnered significant attention among engineers for quantifying uncertainty in neural network predictions, including those used in meta-models. In particular, methods such as deep ensemble and Monte Carlo dropout (MC-dropout) have been widely adopted in engineering applications because they require only minimal modifications to existing network architectures.

MC-dropout estimates predictive uncertainty by keeping dropout activated during inference, while the deep ensemble approach utilizes an ensemble of networks trained in parallel [8]. These methods are valued for delivering robust uncertainty estimates, even with out-of-distribution data. Its effectiveness has been demonstrated across various architectures and datasets in studies such as those by [39, 40, 41]. Deep ensembles capture uncertainty information by training models with the Negative Log-Likelihood (NLL) loss rather than the traditional mean squared error. The NLL loss quantifies the negative log probability of the observed data given the predicted probability distribution, providing a robust metric for evaluating probabilistic models. For example, in a regression task that assumes a Gaussian likelihood, where the model predicts a mean $\mu(\mathbf{x})$ and variance $\sigma^2(\mathbf{x})$ for an input $\mathbf{x}$, the NLL loss for an observation y is expressed as:

$$\mathrm{NLL}(y, x) = \frac{1}{2}\log(2\pi\sigma^2(x)) + \frac{(y-\mu(x))^2}{2\sigma^2(x)} \tag{3}$$

This formulation penalizes errors between predicted and actual values, as well as misestimation of uncertainty, thereby enabling a quantitative assessment of the model's probabilistic outputs.

However, a single model is typically dominated by the inherent aleatoric uncertainty in the data, making it challenging to accurately quantify its epistemic uncertainty. Deep ensemble methods train multiple neural networks in parallel using the same dataset to address this issue. An ensemble composed of M models is utilized, each producing a predicted mean μ_i and variance σ_i^2. The ensemble's overall predictive mean $\hat{\mu}$ is given by:

$$\hat{\mu} = \frac{1}{M}\sum_{i=1}^{M}\mu_i \tag{4}$$

while the total predictive variance $\hat{\sigma}^2$ decomposes into aleatoric and epistemic components:

$$\hat{\sigma}^2 = \underbrace{\frac{1}{M}\sum_{i=1}^{M}\sigma_i^2}_{\text{aleatoric}} + \underbrace{\left(\frac{1}{M}\sum_{i=1}^{M}\mu_i^2 - \hat{\mu}^2\right)}_{\text{epistemic}}. \tag{5}$$

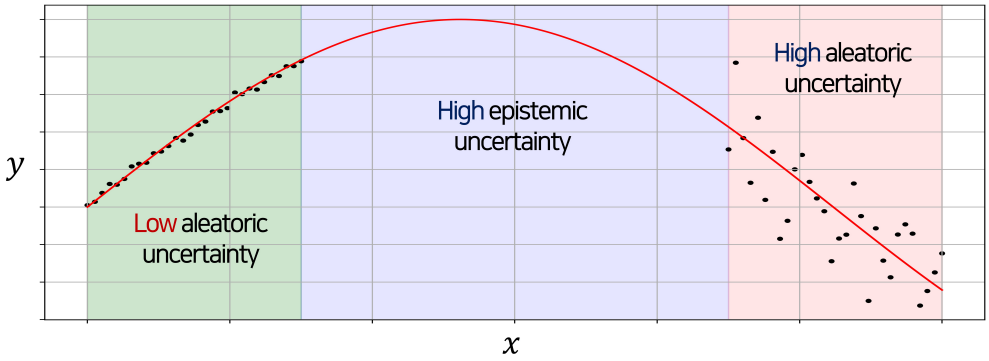


Figure 2: Illustration of aleatoric and epistemic uncertainty in regression

In other words, the averaged variance term ($\frac{1}{M}\sum_{i=1}^{M}\sigma_i^2$) captures the noise inherent in the data (aleatoric uncertainty), while the spread of the individual model means ($\frac{1}{M}\sum_{i=1}^{M}\mu_i^2 - \hat{\mu}^2$) reflects the uncertainty attributable to limited knowledge of the model parameters (epistemic uncertainty). The two uncertainties are schematically represented in Fig. 2. By combining the predictions from multiple networks in this manner, deep ensembles not only produce stable overall predictions but also provide a reliable measure of both data-driven and model-driven uncertainties. The effectiveness of the deep ensemble technique has been demonstrated through various studies [42, 43, 44, 45].

2.3. Active Learning for Engineering

Active learning is a prominent learning strategy that enables machine learning models to achieve high performance with the minimum possible amount of labeled data [46]. While conventional supervised learning typically requires large amounts of labeled data, real-world applications often face significant constraints on the cost or time required for data labeling. Active learning addresses this by maximizing data efficiency; it intelligently selects the most informative data points (i.e., those about which the model is uncertain or those that offer high information gain) and requests an oracle, such as a human expert or a high-fidelity simulation, to provide labels for these selected instances.

A core aspect of active learning is the Query Strategy, which determines how the model selects the next data point to be labeled from a pool of unlabeled data. The fundamental

idea is to query the instance that is most confusing or beneficial to the model's learning. Various strategies have been developed to quantify this notion of informativeness. Furthermore, reliable UQ is indispensable for guiding advanced data-acquisition strategies such as active learning, which enables the selection of informative samples that effectively reduce model uncertainty. However, applying general UQ methods within an active learning loop to complex engineering problems involving high-dimensional, sparse data poses challenges. These include accurately estimating uncertainty in under-sampled regions and managing the computational cost of UQ within sequential-sampling frameworks, highlighting the need for tailored UQ approaches.

Prominent uncertainty sampling strategies select instances where the model exhibits the lowest confidence in its predictions. One common measure is Shannon Entropy, calculated as:

$$H(p) = -\sum_{i=1}^{K} p_i \log p_i \tag{6}$$

where p_1, p_2, ..., p_K are the model's predicted probabilities for each class. A higher entropy value indicates a more uniform predicted probability distribution, and thus higher model uncertainty. Another straightforward strategy is Least Confidence, which uses $1-p_{(1)}$ as the uncertainty measure, where $p_{(1)}$ is the highest predicted probability for any class. A lower $p_{(1)}$ directly correlates with lower model confidence and higher perceived uncertainty. Beyond simple confidence measures, other query strategies include Query-by-Committee, which selects instances exhibiting the most significant disagreement among an ensemble of models, and Information Gain-based methods, which quantify the expected reduction in model uncertainty upon labeling a particular data point.

Active learning is particularly valuable in engineering fields where acquiring labeled data through physical experiments or high-fidelity numerical analysis (e.g., CFD, FEM) incurs very high computational costs [47, 48, 49, 50, 51]. Active learning enables the development of high-performance models under strict budget constraints by minimizing the number of necessary, expensive simulations. Optimization frameworks that utilize meta-models based on numerical analysis results actively leverage active learning techniques [50]. Furthermore, given its flexibility, the core principle can be applied to any algorithm, provided that "what data is necessary for learning" can be defined; active learning is widely applicable beyond optimization. It is frequently used when acquiring high-reliability data is resource-intensive, such as multi-fidelity modeling [50, 51, 52, 53] and uncertainty quantification tasks [54, 55, 56]. As demonstrated across various studies, active learning is a practical approach for training meta-models by efficiently sampling data.

2.4. Taxonomy and Research Gap

To clearly position this work within the current research landscape and identify the specific research gap, Table 1 provides a taxonomy of relevant 3D data augmentation approaches. This comparative analysis highlights the evolution from geometric-centric to performance-aware methodologies.

Table 1: Comparison of 3D augmentation methods, highlighting the research gap

Study / Method	Augmentation method	Focus on geometric diversity	Performance-aware
Conventional Augmentation [7, 13, 22, 23, 33, 34, 35, 36, 37]	FFD / Parametric	Yes	No
Generative Models [12, 18, 57, 58, 26, 59]	DGM	Yes	Partial [26]
This Work	FFD	Yes	Yes

As summarized in Table 1, existing methods show a clear limitation in how they guide data acquisition. Conventional FFD methods focus solely on geometric diversity, blindly exploring the design space without feedback. While recent generative frameworks like t-METASET aim to address performance awareness, they rely on regressor-predicted mean values to estimate property diversity, explicitly disregarding the model's sampling uncertainty [26]. This reliance on predictions can be risky, as the model may confidently predict incorrect values in unexplored regions.

Consequently, the critical research gap is the lack of a framework that uses epistemic uncertainty as the direct driver of spatial augmentation. Unlike previous works that rely on random sampling or deterministic predictions, this work proposes a novel framework where the FFD control points are actively guided by the meta-model's uncertainty. This ensures that the augmentation process explicitly targets the unknown regions of the performance space, leading to more efficient and robust model training.

3. Methodology

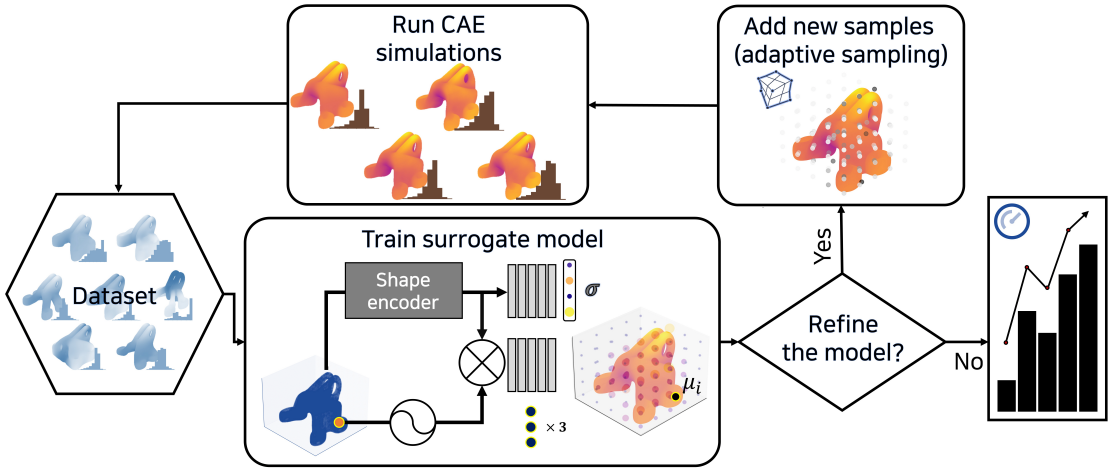


Figure 3: Overall framework integrating FFD, UQ, and Active Learning

The overall framework of this study is shown in Fig. 3 , which strategically links the three key technologies of FFD, UQ, and Active Learning. In this framework, FFD is utilized as the core generative mechanism for localized deformation. This mechanism is guided by an Active Learning strategy. The query for this strategy is provided by UQ: the N^3 FFD control

points are explored based on uncertainty information quantified by a deep-ensemble-based meta-model. Here, the focus is on leveraging epistemic uncertainty, which can be reduced by augmenting training data.

Within a single shape instance, regions of high and low uncertainty regarding the engineering performance are spatially distributed. For example, if displacement is defined as the prediction performance, points near a fixed boundary condition will likely have similar values across different shapes, resulting in a less diverse data distribution. Conversely, points near a load condition are significantly influenced by shape variations, leading to a more diverse data distribution. Thus, uncertainty can vary spatially even for a single instance, indicating regions that are valid and informative for exploration from a data perspective.

This spatial variation in uncertainty was confirmed through uncertainty quantification, and the exploration space for the FFD control points was defined based on epistemic uncertainty. The strategy involves sampling less around regions of low uncertainty and exploring a broader range in regions of high uncertainty. This approach aims to secure data diversity from a performance perspective, ensuring that the augmented data effectively covers regions where the model is less certain. Among several uncertainty-based sampling criteria, the Least Confidence is adopted in this study.

3.1. Field Prediction with UQ

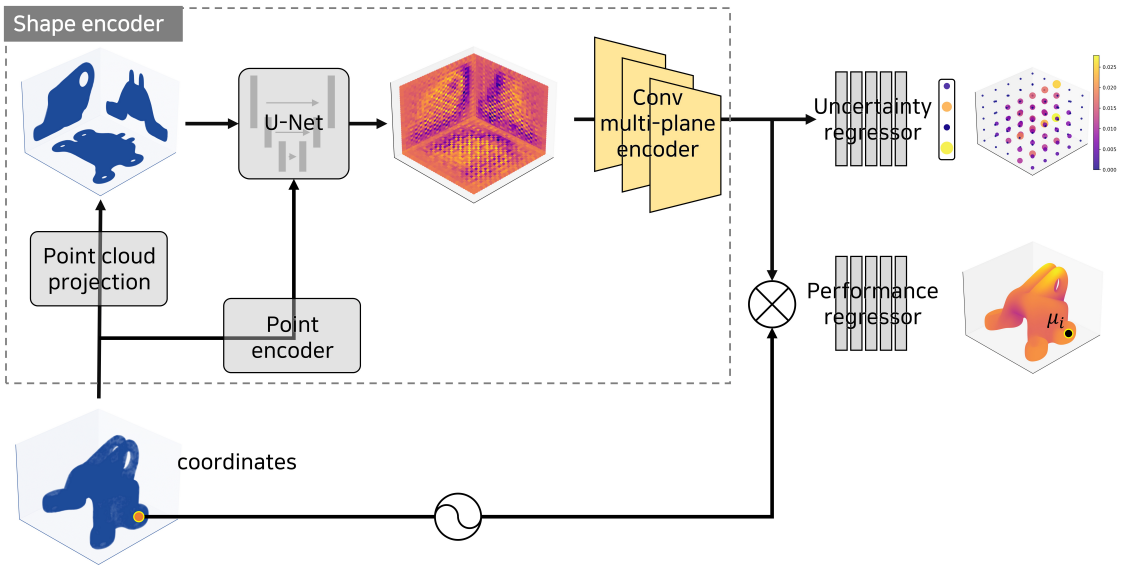


Figure 4: Architecture of the proposed explicit lattice-based UQ meta-model

In this study, Implicit Neural Representation (INR), a widely used meta-model in scientific machine learning, is adopted. INR is a coordinate-based neural network that parameterizes physical properties within space and time. Unlike general explicit representation methods, INR is continuous and adaptive, offering memory efficiency and enabling the representation of high-resolution data. Due to these advantages, INR is effectively utilized for processing information such as signed distance functions, which represent the distance of a point from the surface of a shape, occupancy fields that define whether a specific coordinate

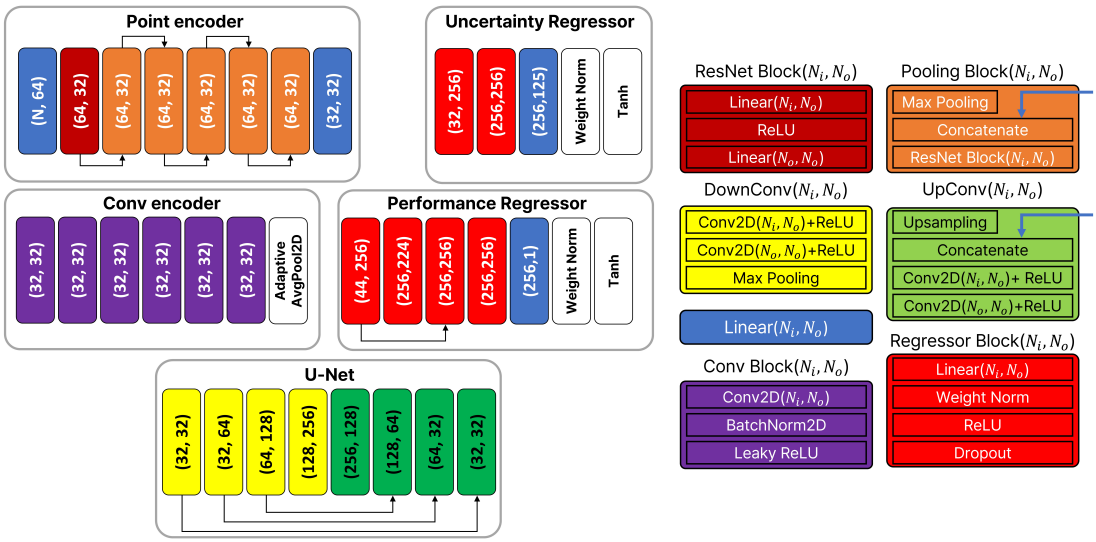


Figure 5: Detailed parametric configuration of the explicit lattice-based UQ meta-model

is within an occupied region, and physical properties [4, 60, 61, 62, 63, 64]. A particular strength is its robust performance even with less data [18]. Data from methods such as FEM/CFD often contain a large number of volume-mesh nodes to achieve high accuracy. The DeepJEB data utilized in this study, for instance, has approximately 200,000 nodes per data point, making learning explicitly from meshes considerably time-consuming [12].

The architecture is shown in Fig. 4. It comprises a shape encoder that processes explicit point cloud data, along with two regressors for uncertainty quantification and performance prediction. The detailed parameters of this architecture are presented in Fig. 5, and the *Adam* optimizer was configured with a learning rate of 0.005 [65]. To implement the deep ensemble approach, an ensemble size of $M = 3$ was used.

Shape Encoder:

This study utilizes the multi-plane point cloud encoder from ConvONet as the shape encoder [66]. This method is known for effectively capturing local information using Convolutional Neural Networks and possesses translational equivariance. This characteristic is well-suited to feature extraction in mechanical design components, where similar shapes recur rather than arbitrary geometries.

To process 3D point clouds, the encoder comprises two branches designed to capture global and local information. A point encoder is designed to extract the global features of the point cloud, and it is configured to maintain permutation invariance, similar to PointNet [67, 68]. The 3D point cloud is projected onto 2D planes (xy, yz, and zx) to capture local information more precisely, as shown in Fig. 6. The projected tri-plane features are fed into a U-Net, and the global features are encoded by the point encoder to obtain features across the 2D multi-planes [69]. Since these features are represented as point locations, convolutional layers can effectively extract location-based local features. Finally, a single global feature can be obtained by applying convolutional layers to the individual planes.

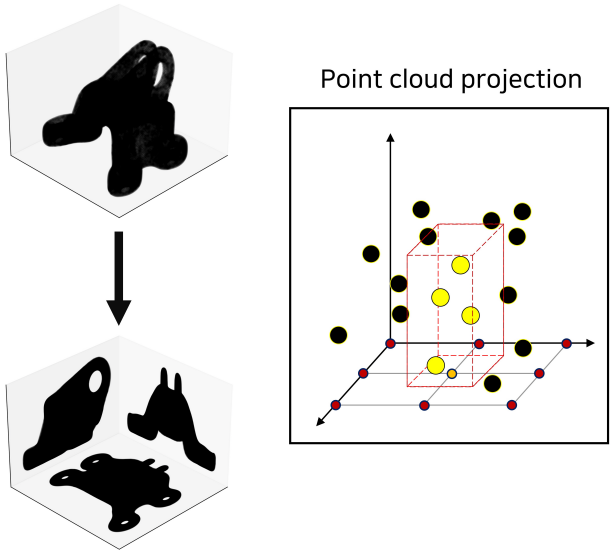


Figure 6: Process of point cloud projection within the encoder

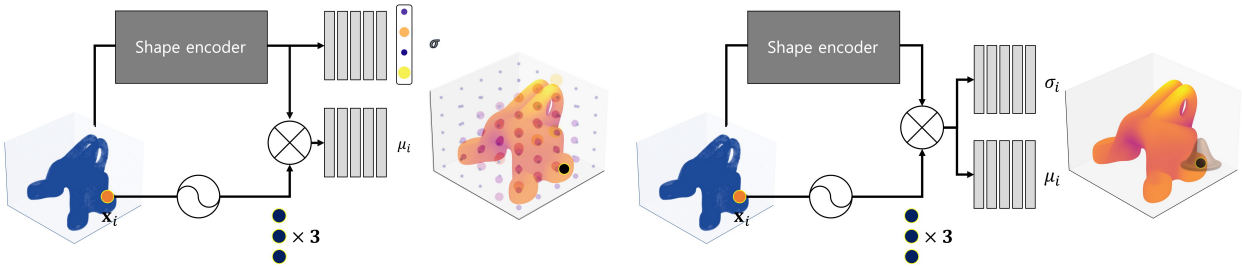
Performance Regressor:

The performance regressor receives feature information from the point cloud (the shape encoder's output) and the locations of individual point coordinates. Here, these coordinates undergo positional encoding. The positional encoding can be expressed as follows:

$$\gamma(p) = (sin(2^0\pi p), cos(2^0\pi p), ..., sin(2^{L-1}\pi p), cos(2^{L-1}\pi p)) \tag{7}$$

Positional encoding projects input coordinates into high-dimensional Fourier space using sinusoidal transformations [61, 62]. This enables models to capture both high- and low-frequency spatial features effectively. It facilitates the learning of fine details in complex spatial data, thereby improving expressiveness and precision. The performance regressor is constructed by concatenating the positional encoded coordinates and the global features from the shape encoder. In this study, the parameter L was set to 2. This combined representation is then processed by the regressor to predict the performance value at the given coordinates.

Uncertainty Regressor:



(a) Explicit lattice-based UQ meta-model

(b) Point-wise UQ meta-model

Figure 7: Architectures of the compared meta-models

This study proposes a framework that seamlessly integrates FFD and the meta-model for uncertainty quantification. The most critical component here is the uncertainty regressor. To quantify the performance uncertainty of the shape, a structure was designed to measure uncertainty in the space around the shape, using feature information from the point cloud (the output of the shape encoder) [8, 45]. This can be briefly represented as shown in Fig. 7a. This architecture demonstrates improved predictive performance compared to the point-wise uncertainty regressor (Fig. 7b) and can estimate overall uncertainty in the locations of FFD control points. However, unlike existing work that can process only a single instance, the model used here adds a component that takes point cloud features as input and outputs N^3 vectors, enabling the processing of multiple instances [45].

Uncertainty regression is performed based on the global features obtained from the shape encoder. The uncertainty regressor receives only the point cloud's feature information as input, enabling uncertainty to be inferred across the entire point cloud rather than at individual points.

$$\sigma_p = \sum_{a=0}^{1}\sum_{b=0}^{1}\sum_{c=0}^{1}[w_x^{(a)} w_y^{(b)} w_z^{(c)}]\sigma(i+a, j+b, k+c) \tag{8}$$

$$w_x^{(0)} = 1 - dx, \quad w_x^{(1)} = dx, w_y^{(0)} = 1 - dy, \quad w_y^{(1)} = dy, w_z^{(0)} = 1 - dz, \quad w_z^{(1)} = dz \tag{9}$$

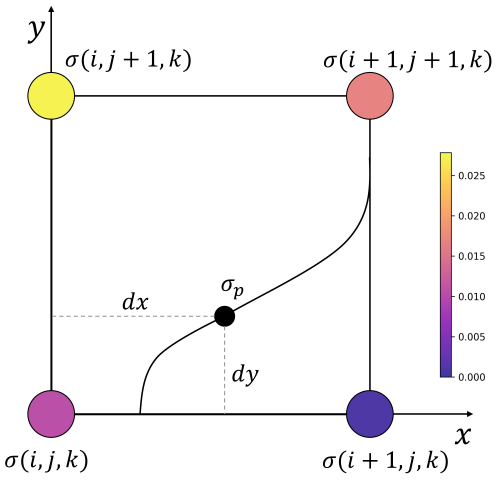


Figure 8: Trilinear interpolation of explicit uncertainty

Points for uncertainty estimation are defined by dividing the pre-defined domain, where the point cloud is defined, $x \in [-1, 1], y \in [-1, 1], z \in [-1, 1]$, into N equal divisions along each axis. The uncertainty of an individual point is then calculated as the trilinear interpolation of the uncertainties assigned to the eight points that define the unit cube containing it. This can be mathematically expressed as Eqs. 8 and 9, and a schematic representation of this process is shown in Fig. 8. Through this architecture, uncertainty regarding performance can be estimated in the space surrounding the 3D shape.

Algorithm 1: Grid Coordinate Optimization via Trilinear Interpolation

Input: 3D Field Tensor $\mathbf{F} \in \mathbb{R}^{1\times1\times D\times H\times W}$, Target Value $t \in \mathbb{R}$

Output: Optimized Grid Coordinate $\mathbf{g} \in [-1, 1]^3$

Initialize grid coordinate $\mathbf{g} \sim \mathcal{U}(-1, 1)^3$ as trainable parameter;

for $i \leftarrow 1$ **to** 3000 **do**

  Perform trilinear interpolation: $s \leftarrow \text{TrilinearInterp}(\mathbf{F}, \mathbf{g})$;

  Compute loss: $\mathcal{L} \leftarrow \|s - t\|^2$;

  Update $\mathbf{g}$ via gradient descent using *Adam* optimizer;

return optimized coordinate $\mathbf{g}$

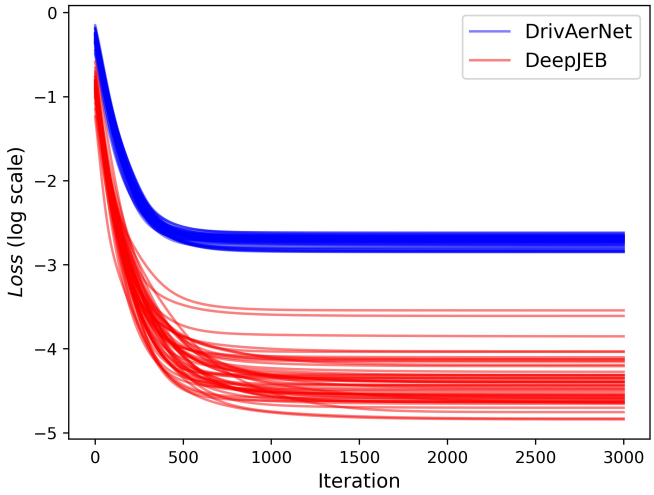
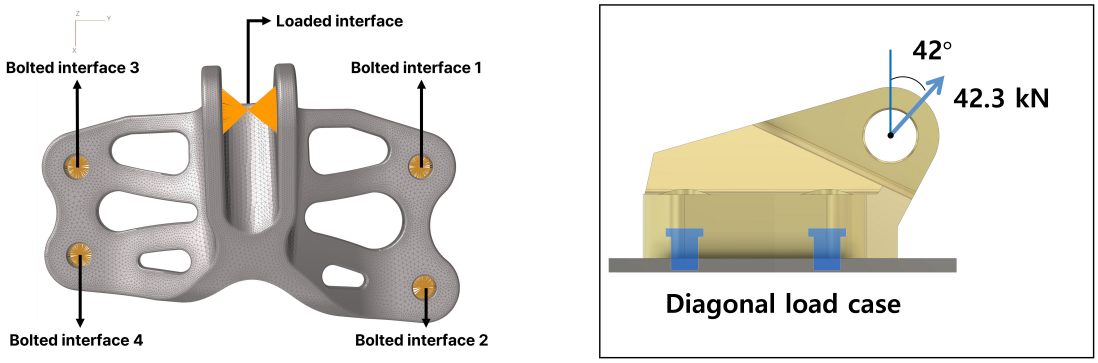


Figure 9: Optimization loss history during the inverse estimation (Algorithm 1)

This process extracts N^3 values that best represent the entire field for each model. Based on these values, aleatoric uncertainty is calculated. However, the estimated aleatoric uncertainty derived from the ensemble in this manner is not the primary focus of this study. To effectively estimate the epistemic uncertainty, which arises from a lack of data and is the focus here, the variance of the data points obtained from the ensemble members must be quantified. The objective is to determine the epistemic uncertainty not point-wise but at the discrete control points of the FFD lattice. Quantifying this precisely by directly inverting the trilinear interpolation, which was used to quantify aleatoric uncertainty, is impossible. This is because the inverse problem is non-injective, making it impossible to uniquely define the grid values from a given set of N^3 uncertainty values. Therefore, to acquire the grid values that fit the trilinear interpolation used for aleatoric uncertainty, an optimization-based inverse estimation is performed. The objective function in Algorithm 1 is formally grounded in the theory of inverse problem regularization. Given that the mapping from lattice control points to spatial field values is non-injective, a direct analytical inversion is unattainable.

Therefore, the estimation is formulated as a least-squares optimization problem to minimize the residual between the interpolated uncertainty and the target field. This formulation leverages the Bernstein basis functions of the FFD lattice as a spatial filter. Effectively, the lattice structure acts as a low-pass regularizer, suppressing high-frequency predictive noise and ensuring that the generated uncertainty map maintains C^2 continuity with the underlying physical domain. The *Adam* optimizer is used as the optimization algorithm for this process [65]. This procedure is further detailed in Algorithm 1 below. Figure 9 visualizes the optimization loss history. The number of iterations was set to 3,000, as the loss trajectories for both the DeepJEB and DrivAerNet datasets (on a $\log_{10}$ scale) demonstrate that the optimization consistently stabilizes and reaches a plateau well before this limit. Specifically, the loss values for DeepJEB converge to approximately 10^{-5}, while DrivAerNet stabilizes around 10^{-3}, confirming that 3,000 iterations are sufficient to ensure robust inverse estimation for both physical domains. Based on these values, epistemic uncertainty is calculated.

4. Experiments



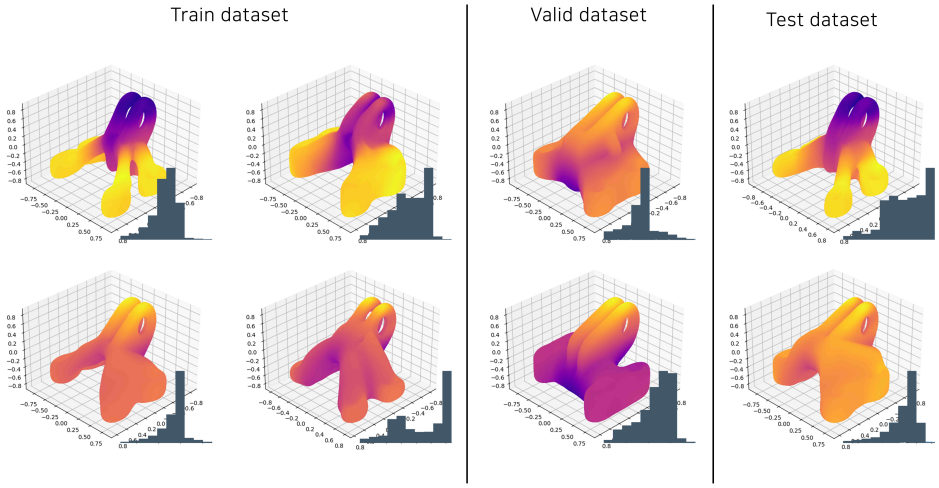
(a) Location of load and fixed conditions

(b) Description of the diagonal load case

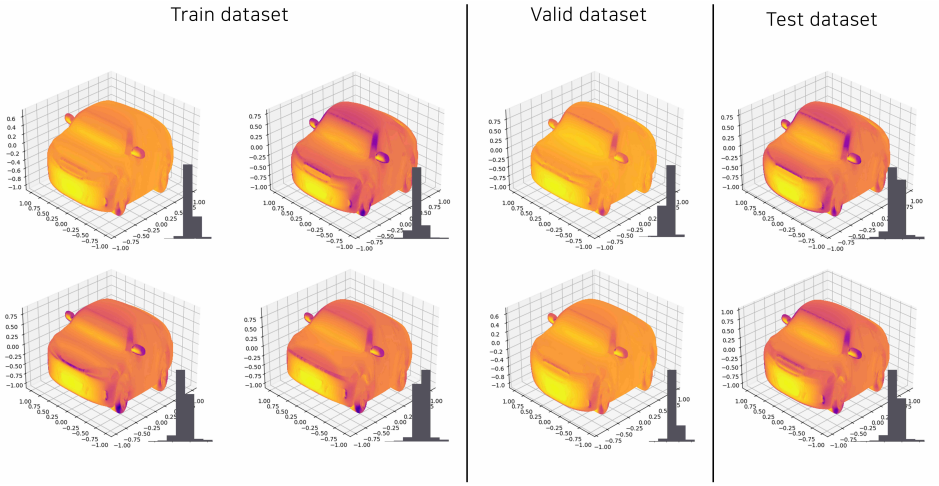
Figure 10: Experimental setup for the DeepJEB dataset

To confirm the usefulness and generalizability of the framework proposed in this study, two distinct 3D engineering datasets were utilized: the DeepJEB bracket dataset [12, 17] and the DrivAerNet vehicle dataset [7]. DeepJEB is a synthetic dataset generated by learning from SimJEB data, containing 2,138 data points for a single bracket class [12, 17]. Properties of the Ti-6Al-4V bracket material include an elastic modulus (E) of 113.8 GPa, a Poisson's ratio (ν) of 0.342, and a density (ρ) of 4.47×10^{-3} g/mm^3. The dataset includes high-fidelity FEM analysis results obtained using 2nd-order tetrahedral elements. Notably, it is a dataset with consistent boundary conditions, featuring bolted interfaces at each corner and a loaded interface, as shown in Fig. 10a. As shown in Fig. 10b, the x-direction displacement resulting from applying a 42.3 kN force diagonally was selected as the label. The DrivAerNet dataset comprises 4,000 parametric car shapes with diverse aerodynamic properties evaluated via high-fidelity CFD simulations using OpenFOAM with the $k-\omega$ SST turbulence model at $Re \approx 9.39 \times 10^6$ [70]. In this study, the surface pressure distribution was selected as the primary label to capture detailed aerodynamic characteristics. Consequently, the entire vehicle geometry was defined as the design domain.

The experimental setup was specifically designed to simulate the common industrial challenge of data scarcity, which serves as the core theoretical background and motivation for this work. Instead of leveraging the entire dataset, which would be unrealistic during the conceptual design stage, a small subset of only 50 data points was selected from the DeepJEB and DrivAerNet datasets to serve as the initial $Data_{Original}$. This initialization from a minimal sample set provides the critical basis for this study, as the primary objective is to evaluate the framework's ability to efficiently construct a robust meta-model from a realistic, budget-constrained starting point rather than from an already data-rich environment. To ensure statistical reliability and mitigate the influence of random initialization, all experiments were conducted over five independent runs ($N = 5$) with distinct random seeds. Accordingly, all performance metrics reported in Tables 3, 4 and 7 are presented as the mean value followed by the standard deviation (mean $\pm$ std).



(a) DeepJEB bracket samples



(b) DrivAerNet vehicle samples

Figure 11: Visualization of representative samples and performance distributions

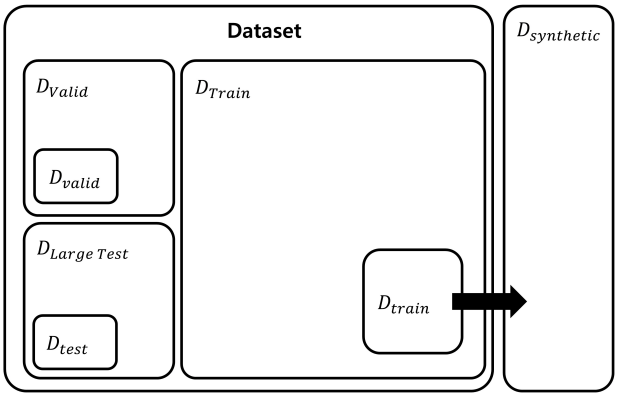


Figure 12: Venn diagram of the datasets used in the experiment

Dataset		R^2 score	MAE	RMSE
DeepJEB	D_{Train}	0.9821	1.4275 ($\times 10^{-3}$)	0.2036 ($\times 10^{-3}$)
	D_{Valid}	0.9373	2.2834 ($\times 10^{-3}$)	3.7481 ($\times 10^{-3}$)
	$D_{Large\ Test}$	0.9265	2.2924 ($\times 10^{-3}$)	4.1055 ($\times 10^{-3}$)
DrivAerNet	D_{Train}	0.9474	2.6028 ($\times 10^{+1}$)	1.4689 ($\times 10^{+1}$)
	D_{Valid}	0.9481	2.5808 ($\times 10^{+1}$)	1.4826 ($\times 10^{+1}$)
	$D_{Large\ Test}$	0.9484	2.5789 ($\times 10^{+1}$)	1.4712 ($\times 10^{+1}$)

Table 2: Predictive performance of the pre-trained meta-models for the DeepJEB and DrivAerNet datasets

4.1. Pre-trained Meta-model

Typically, in an active learning setting, the augmented data would require new labels, often generated by re-performing high-fidelity analysis (FEM or CFD) for each augmented instance. However, in this study, pre-trained meta-models were developed for both the DeepJEB and DrivAerNet datasets to enable rapid auto-labeling. The data augmentation method proposed in this study focused on exploring the performance space using these surrogates.

The DeepJEB meta-model was trained on 2,060 data points, while the DrivAerNet model used 4,000. Both datasets were divided in an 8:1:1 ratio for D_{Train}, D_{Valid}, and D_{Test}, respectively. Representative samples included in each of these sets are visualized in Fig. 11. The histograms at the bottom right of each sample show the node value distributions (displacement for DeepJEB, pressure for DrivAerNet), confirming that the data exhibit diverse distributions and shapes.

As shown in Fig. 12, D_{Train} comprises D_{train} (used for augmentation) and is distinct from D_{test} and $D_{Large\ Test}$ (used for model evaluation). The models were evaluated using standard regression metrics, including Root Mean Squared Error (RMSE), Mean Absolute

Error (MAE), and the R^2 score. The results, summarized in Table 2, demonstrate high accuracy and generalization across both physics domains.

In particular, the DeepJEB meta-model achieved an R^2 score of 0.9265 on the test data, and the DrivAerNet model achieved an even higher R^2 score of 0.9484. These high R^2 scores demonstrate that the meta-models effectively capture the underlying relationships in the data and generalize well to unseen samples, confirming their robustness and reliability in predicting the target responses.

4.2. Experimental Results

4.2.1. Hyperparameter Sensitivity Analysis

To ensure optimal performance across diverse physical domains, hyperparameter tuning was performed independently for each dataset using the Tree-Structured Parzen Estimator sampler within the Optuna framework [71, 72]. To provide a rigorous understanding of the model's training dynamics beyond this initial optimization, a quantitative sensitivity analysis was conducted using the Functional ANOVA (F-ANOVA) method [73]. This approach decomposes the objective function's variance into contributions from individual hyperparameters, offering insights into which factors are critical for convergence. The relative importance of each hyperparameter for both the proposed explicit lattice-based model and the comparative point-wise model across both datasets is summarized in Fig. 13.

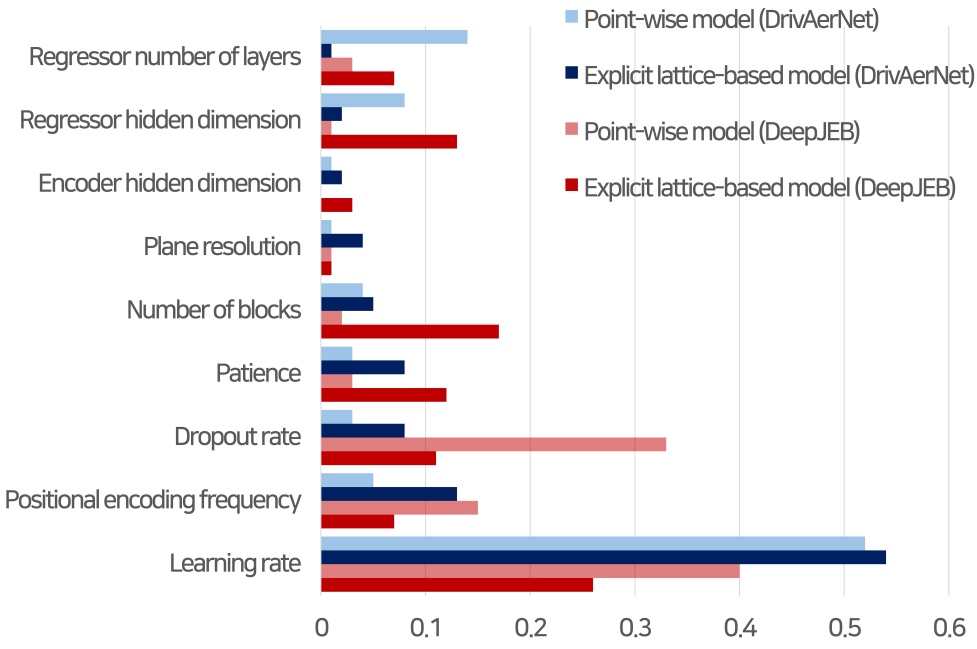


Figure 13: Hyperparameter importance analysis using F-ANOVA

The comparative analysis reveals distinct behaviors for each architecture depending on the task:

- **Dominance of Learning Rate:** The **learning rate** consistently emerges as a critical factor.In the DrivAerNet case, it is the most dominant factor for both architectures

(importance > 0.5). In the DeepJEB case, while its relative importance is lower than that of DrivAerNet, it remains a primary factor (0.40 for Point-wise, 0.26 for Explicit), indicating that the optimization step size is universally fundamental to convergence.

- **Inconsistency of the Point-wise Model:** The point-wise model exhibits an inconsistent sensitivity profile across domains. In the DrivAerNet case (aerodynamics), structural parameters, such as the **regressor number of layers** (0.14), are influential. However, in the DeepJEB case (structural mechanics), the model becomes extremely sensitive to regularization, with **dropout rate** (0.33) being a major factor. This suggests that the point-wise model requires fundamentally different tuning strategies—focusing on capacity for complex flows rather than on regularization for structural data.
- **Structural Utilization of the Explicit Lattice-based Model:** The explicit model demonstrates a more stable utilization of its architectural features. In the DeepJEB case, the **number of blocks** (0.17) and **regressor hidden dimension** (0.13) appear as important secondary factors. This indicates that the model effectively leverages its explicit lattice structure (represented by the multi-plane encoder blocks) to capture geometric features. In the DrivAerNet case, the model shows robustness to these structural parameters, with secondary importance shifting to training stability factors like **positional encoding frequency** (0.13).

In summary, the point-wise model shows varying sensitivity to network depth and regularization depending on the physics, whereas the explicit lattice-based model tends to rely on its lattice structure and training stability parameters.

4.2.2. Case Study I: Structural Analysis (DeepJEB)

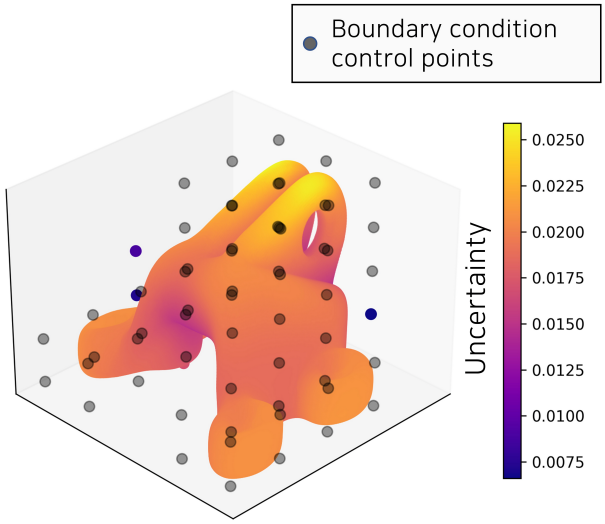


Figure 14: Uncertainty at FFD control points

A small dataset was separately constructed from D_{Train}, D_{Valid}, and $D_{Large\ Test}$. This dataset was used to verify the effectiveness of the augmentation framework, which utilizes

a pre-trained meta-model. Specifically, a small dataset was configured for framework validation, utilizing data points from D_{train}, D_{valid}, and D_{test}, as shown in Fig. 12, with 50, 10, and 10 data points, respectively. To assess generalization, $D_{Large\,Test}$ was utilized in the evaluation. Furthermore, since the data points in the dataset share identical boundary conditions, post-processing was performed, as shown in Fig. 14, to ensure that control points near the boundary conditions were not moved during augmentation.

As shown in Table 3, experiments were conducted using a total of four datasets: Original ($Data_{Original}$), Random Augmentation ($Data_{Random}$), UQ-considered Augmentation ($Data_{UQ}$), and Add Data ($Data_{Add}$).

$Data_{Original}$ is the simplest, consisting of 50 training and 10 validation data points, and it represents the lower bound of the meta-model's performance achievable with a minimal dataset size.

$Data_{Add}$ comprises 50 original data points and 500 additional data points from the DeepJEB dataset. Since 500 data points were added directly as arbitrary shape data rather than through augmentation, this dataset spans the most expansive design space among comparable datasets of similar size. This dataset represents the upper bound of the meta-model's performance achievable by simply adding more data points of the same type without intelligent augmentation.

$Data_{Random}$ was augmented in a manner consistent with studies that use the LHS of control points for data augmentation. This dataset was constructed by augmenting each of the 50 original data points 10 times, yielding 500 augmented data points for a total of 550.

$Data_{UQ}$ was augmented using the methodology proposed in this study. Similar to $Data_{Random}$, it was constructed by augmenting each of the 50 original data points 10 times.

The experimental results are presented in Table 3. Overall, the methodology proposed in this study demonstrated superior performance, with an R^2 improvement of 16.5% compared to $Data_{Original}$. As shown in Table 3, the proposed $Data_{UQ}$ strategy achieved an R^2 score of 0.813 ± 0.006 on the $D_{Large\ Test}$ set. This improvement remained consistent across all trials, and the narrow confidence interval demonstrates a robust and reliable performance boost that significantly exceeds the random baseline (0.698 ± 0.007). The performance difference compared to the $Data_{Add}$ was 10.6 %, indicating that the proposed augmentation approach, with significantly fewer original data points, can achieve performance comparable to simply adding a large number of arbitrary data points.

In particular, one of the novelties presented in this study, the explicit lattice-based uncertainty regressor, showed overall high performance improvement compared to a general point-wise uncertainty regressor. The point-wise regressor, as shown in Fig. 7b, was evaluated under the same experimental setup, and the results are presented in Table 3. Overall, it can be confirmed that the performance is superior to that of experiments in the same category (i.e., using a point-wise regressor). Furthermore, a performance improvement of 21.0 % was observed compared to the experiment using the point-wise regressor on $Data_{Original}$.

Model	Data type	# of train data		R^2 score	MAE ($\times 10^{-3}$)	RMSE ($\times 10^{-3}$)
Explicit (Ours)	$Data_{Original}$	50	D_{train}	**0.964 ± 0.000**	**1.89 ± 0.02**	**2.57 ± 0.02**
			D_{valid}	0.819 ± 0.006	4.12 ± 0.07	5.49 ± 0.10
			D_{test}	0.624 ± 0.011	5.14 ± 0.05	7.44 ± 0.07
			$D_{Large\ Test}$	0.679 ± 0.008	5.26 ± 0.09	7.55 ± 0.11
	$Data_{Random}$	50 (+500)	D_{train}	0.906 ± 0.000	2.98 ± 0.01	4.51 ± 0.01
			D_{valid}	0.830 ± 0.007	4.00 ± 0.07	5.78 ± 0.12
			D_{test}	0.669 ± 0.005	5.21 ± 0.05	7.42 ± 0.06
			$D_{Large\ Test}$	0.698 ± 0.007	5.27 ± 0.07	7.64 ± 0.09
	$Data_{UQ}$	50 (+500)	D_{train}	0.934 ± 0.001	2.62 ± 0.02	3.39 ± 0.03
			D_{valid}	0.839 ± 0.004	4.11 ± 0.05	5.41 ± 0.06
			D_{test}	0.718 ± 0.010	4.56 ± 0.07	6.12 ± 0.10
			$D_{Large\ Test}$	0.813 ± 0.006	4.54 ± 0.05	5.93 ± 0.06
	$Data_{Add}$	50 (+500)	D_{train}	0.939 ± 0.001	2.44 ± 0.02	3.44 ± 0.03
			D_{valid}	**0.878 ± 0.003**	**3.36 ± 0.07**	**4.83 ± 0.07**
			D_{test}	**0.899 ± 0.004**	**2.78 ± 0.04**	**4.06 ± 0.06**
			$D_{Large\ Test}$	**0.899 ± 0.004**	**2.96 ± 0.04**	**4.30 ± 0.06**
Point-wise	$Data_{Original}$	50	D_{train}	0.936 ± 0.001	2.52 ± 0.01	3.42 ± 0.01
			D_{valid}	0.794 ± 0.006	4.53 ± 0.05	6.07 ± 0.06
			D_{test}	0.571 ± 0.007	5.37 ± 0.03	7.65 ± 0.04
			$D_{Large\ Test}$	0.642 ± 0.005	5.53 ± 0.04	7.83 ± 0.07
	$Data_{Random}$	50 (+500)	D_{train}	0.832 ± 0.001	3.95 ± 0.03	5.21 ± 0.04
			D_{valid}	0.816 ± 0.003	4.17 ± 0.03	5.54 ± 0.04
			D_{test}	0.658 ± 0.006	5.60 ± 0.07	7.54 ± 0.08
			$D_{Large\ Test}$	0.686 ± 0.004	5.72 ± 0.05	7.69 ± 0.07
	$Data_{UQ}$	50 (+500)	D_{train}	0.850 ± 0.001	4.04 ± 0.01	5.69 ± 0.02
			D_{valid}	0.750 ± 0.002	5.15 ± 0.03	6.98 ± 0.03
			D_{test}	0.530 ± 0.010	5.92 ± 0.20	8.37 ± 0.37
			$D_{Large\ Test}$	0.676 ± 0.006	5.59 ± 0.05	7.90 ± 0.08
	$Data_{Add}$	50 (+500)	D_{train}	0.861 ± 0.002	3.77 ± 0.02	5.11 ± 0.03
			D_{valid}	0.777 ± 0.009	4.56 ± 0.09	6.43 ± 0.15
			D_{test}	0.839 ± 0.006	3.69 ± 0.07	5.15 ± 0.10
			$D_{Large\ Test}$	0.826 ± 0.005	4.00 ± 0.05	5.61 ± 0.08

Table 3: Experimental results for the DeepJEB dataset

4.2.3. Case Study II: Aerodynamic Analysis (DrivAerNet)

To further validate the generalizability of the proposed framework, additional experiments were conducted using the DrivAerNet dataset [7]. Consistent with the DeepJEB experiment, a small data-scarce scenario was simulated. Subsets were extracted from the original splits

D_{Train}, D_{Valid}, and $D_{Large\ Test}$ to create D_{train}, D_{valid}, and D_{test}, containing 50, 10, and 10 data points, respectively.

A key difference in this experiment lies in the boundary conditions. Unlike the bracket dataset, which features fixed bolted interfaces, the DrivAerNet vehicle data does not possess specific constrained boundary conditions. Consequently, the entire vehicle geometry was defined as the design domain. However, considering the inherent left-right symmetry of automotive shapes, the performance-aware augmentation process was explicitly configured to maintain symmetry across the xz-plane.

The quantitative results for this aerodynamic analysis are summarized in Table 4. Overall, the methodology proposed in this study demonstrated superior performance, achieving the highest accuracy among all strategies. $Data_{UQ}$ recorded an R^2 score of 0.927 ± 0.003 on the $D_{Large\ Test}$, representing a distinct improvement over $Data_{Random}$, which achieved an R^2 of 0.916 ± 0.003. In this complex aerodynamic domain, $Data_{UQ}$ recorded an R^2 score of 0.927 ± 0.003 on the $D_{Large\ Test}$ (Table 4). The minimal variance observed across $N = 5$ runs confirms that the lattice-based UQ architecture effectively and stably captures spatial uncertainty, even in high-dimensional fluid-dynamics problems. Notably, in this domain, the performance-aware approach also exceeded the performance of the brute-force approach ($Data_{Add}$, $R^2 = 0.906 \pm 0.002$), which utilized 500 entirely new shapes.

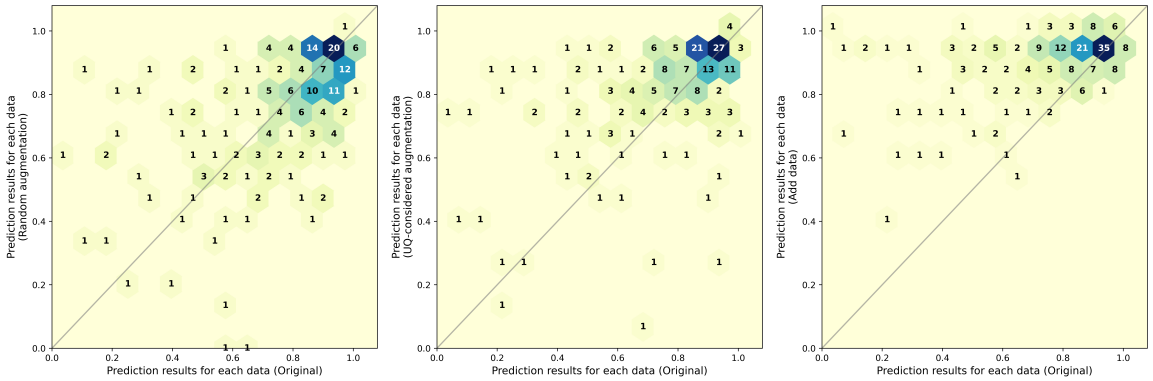
Consistent with the structural analysis results, the proposed explicit lattice-based uncertainty regressor demonstrated substantial performance gains compared to the general point-wise uncertainty regressor in this aerodynamic domain. As presented in Table 4, the point-wise model trained with the same $Data_{UQ}$ strategy achieved an R^2 of only 0.779 ± 0.005. The explicit lattice-based model outperformed the point-wise counterpart by a significant margin (approximately 16.0% improvement). This confirms that the proposed architecture is crucial for effectively capturing spatial uncertainty and guiding meaningful deformations in complex fluid dynamics problems.

This distinct trend regarding $Data_{Add}$ is closely related to the inherent characteristics of the DrivAerNet dataset. Unlike DeepJEB, which was constructed via sampling from a generative model resulting in distinct distributions, DrivAerNet was created by augmenting specific seed geometries. Consequently, rather than clear boundaries between the training, validation, and test sets, there is significant geometric similarity across splits. This explains the consistent error margins observed across datasets, where the test set performance occasionally surpasses the training set results, a contrast to the apparent generalization gaps observed in DeepJEB. Nevertheless, the fact that $Data_{UQ}$ outperforms $Data_{Add}$ despite these characteristics is a significant achievement.

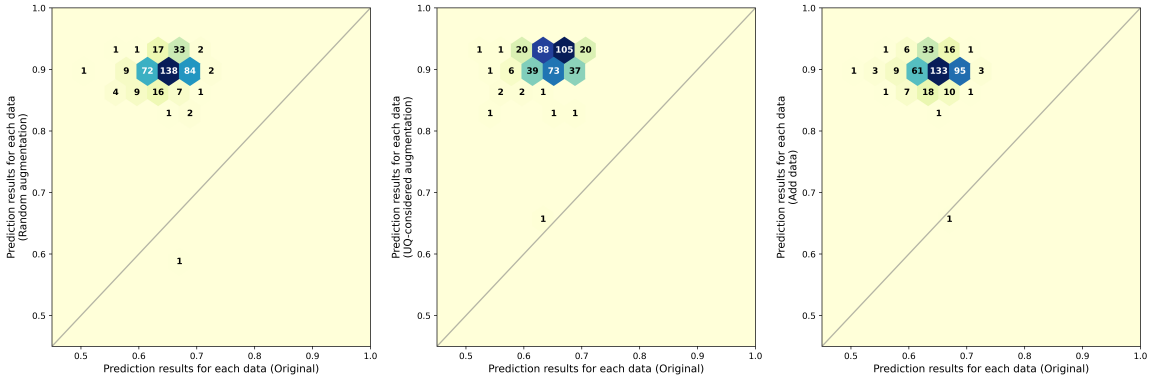
This result suggests that in high-dimensional and complex design spaces like full-vehicle aerodynamics, targeted exploration guided by epistemic uncertainty can be significantly more data-efficient and effective than simply adding random global variations.

4.3. Discussion of Results

This section synthesizes the experimental findings from both the DeepJEB and DrivAerNet datasets to address the core research gaps identified in this study: the necessity of



(a) DeepJEB ($Data_{Random}$) (b) DeepJEB ($Data_{UQ}$) (c) DeepJEB ($Data_{Add}$)



(d) DrivAerNet ($Data_{Random}$) (e) DrivAerNet ($Data_{UQ}$) (f) DrivAerNet ($Data_{Add}$)

Figure 15: R^2 comparison (Hexbin plots) on the large test set ($D_{Large\ Test}$)

Model	Data type	# of train data		R^2 score	MAE ($\times 10^{+1}$)	RMSE ($\times 10^{+1}$)
Explicit (Ours)	$Data_{Original}$	50	D_{train}	0.649 ± 0.002	3.55 ± 0.01	6.81 ± 0.03
			D_{valid}	0.660 ± 0.002	3.52 ± 0.02	6.65 ± 0.03
			D_{test}	0.651 ± 0.002	3.57 ± 0.04	6.72 ± 0.03
			$D_{Large\ Test}$	0.648 ± 0.005	3.48 ± 0.03	6.76 ± 0.04
	$Data_{Random}$	50 (+500)	D_{train}	0.909 ± 0.004	1.84 ± 0.00	3.46 ± 0.07
			D_{valid}	0.907 ± 0.001	1.90 ± 0.00	3.49 ± 0.02
			D_{test}	0.909 ± 0.001	1.82 ± 0.01	3.42 ± 0.04
			$D_{Large\ Test}$	0.916 ± 0.003	1.80 ± 0.01	3.32 ± 0.04
	$Data_{UQ}$	50 (+500)	D_{train}	**0.921 ± 0.003**	**1.73 ± 0.00**	**3.20 ± 0.05**
			D_{valid}	**0.913 ± 0.001**	**1.85 ± 0.00**	**3.37 ± 0.03**
			D_{test}	**0.916 ± 0.002**	**1.75 ± 0.01**	**3.30 ± 0.04**
			$D_{Large\ Test}$	**0.927 ± 0.003**	**1.68 ± 0.02**	**3.06 ± 0.06**
	$Data_{Add}$	50 (+500)	D_{train}	0.900 ± 0.000	1.88 ± 0.00	3.61 ± 0.01
			D_{valid}	0.901 ± 0.001	1.89 ± 0.00	3.58 ± 0.02
			D_{test}	0.904 ± 0.002	1.84 ± 0.01	3.54 ± 0.05
			$D_{Large\ Test}$	0.906 ± 0.002	1.81 ± 0.01	3.52 ± 0.02
Point-wise	$Data_{Original}$	50	D_{train}	0.516 ± 0.006	4.74 ± 0.07	8.01 ± 0.07
			D_{valid}	0.501 ± 0.019	4.89 ± 0.19	8.07 ± 0.14
			D_{test}	0.482 ± 0.007	5.13 ± 0.07	8.14 ± 0.06
			$D_{Large\ Test}$	0.535 ± 0.009	4.41 ± 0.16	7.80 ± 0.10
	$Data_{Random}$	50 (+500)	D_{train}	0.790 ± 0.004	2.82 ± 0.01	5.26 ± 0.08
			D_{valid}	0.771 ± 0.003	2.92 ± 0.01	5.46 ± 0.05
			D_{test}	0.788 ± 0.002	2.85 ± 0.02	5.25 ± 0.05
			$D_{Large\ Test}$	0.771 ± 0.005	3.06 ± 0.06	5.48 ± 0.03
	$Data_{UQ}$	50 (+500)	D_{train}	0.802 ± 0.005	2.70 ± 0.01	5.13 ± 0.09
			D_{valid}	0.782 ± 0.004	2.86 ± 0.02	5.33 ± 0.06
			D_{test}	0.801 ± 0.002	2.77 ± 0.02	5.07 ± 0.04
			$D_{Large\ Test}$	0.779 ± 0.005	2.91 ± 0.05	5.39 ± 0.04
	$Data_{Add}$	50 (+500)	D_{train}	0.714 ± 0.001	3.20 ± 0.00	6.12 ± 0.02
			D_{valid}	0.712 ± 0.001	3.22 ± 0.01	6.13 ± 0.01
			D_{test}	0.719 ± 0.003	3.15 ± 0.01	6.03 ± 0.03
			$D_{Large\ Test}$	0.712 ± 0.006	3.26 ± 0.01	6.15 ± 0.07

Table 4: Experimental results for the DrivAerNet dataset

performance-aware augmentation over mere geometric diversity, and the challenge of integrating uncertainty information with spatial deformation techniques like FFD.

	Data type	Case Study I: DeepJEB		Case Study II: DrivAerNet	
		Mean	Median	Mean	Median
D_{train}	$Data_{Original}$	-1.666	-1.595	-6.561	-7.180
	$Data_{Random}$	-1.814	-1.627	-6.906	-7.497
	$Data_{UQ}$	**-2.084**	**-1.802**	**-7.421**	**-7.863**
	$Data_{Add}$	-1.934	-1.742	-6.868	-7.366
D_{valid}	$Data_{Original}$	-1.634	-1.509	-7.794	-8.260
	$Data_{Random}$	-1.882	-1.686	-8.112	-8.424
	$Data_{UQ}$	**-1.934**	**-1.754**	**-8.502**	**-8.450**
	$Data_{Add}$	-1.900	-1.571	-8.304	-8.226
D_{test}	$Data_{Original}$	-1.531	-1.490	-6.520	-7.082
	$Data_{Random}$	-2.050	-1.779	-7.087	-7.447
	$Data_{UQ}$	**-2.240**	**-1.884**	**-7.282**	-7.343
	$Data_{Add}$	-2.096	-1.672	-7.129	**-7.624**
$D_{Large\ Test}$	$Data_{Original}$	-1.611	-1.544	-6.795	-7.497
	$Data_{Random}$	-1.979	-1.749	-6.903	-7.425
	$Data_{UQ}$	**-2.084**	**-1.646**	**-8.550**	**-9.739**
	$Data_{Add}$	-2.025	-1.823	-7.419	-8.207

Table 5: Comparison of mean and median epistemic uncertainty (log scale) across augmentation strategies

4.3.1. Performance-aware Augmentation vs. Geometric Diversity

Model	Data type	Case Study I: DeepJEB			Case Study II: DrivAerNet		
		R^2 score	MAE	RMSE	R^2 score	MAE	RMSE
Explicit (Ours)	$Data_{Original}$	26.2	14.0	25.7	200	83.8	141
	$Data_{Random}$	30.7	26.6	36.3	5.92	11.2	4.58
	$Data_{Add}$	25.0	47.0	38.5	8.25	15.9	7.92
Point-wise	$Data_{Original}$	43.3	29.2	43.7	113	90.4	138
	$Data_{Random}$	35.6	32.9	39.2	101	85.5	63.2
	$Data_{UQ}$	32.6	27.3	40.9	82.3	108	67.3
	$Data_{Add}$	3.35	14.2	6.3	111	224	113

Table 6: Statistical significance testing of performance on DeepJEB and DrivAerNet datasets ($N = 5$)

The experimental results from both case studies (Tables 3 & 4) provide strong validation for the central hypothesis. In the structural analysis (DeepJEB), the proposed $Data_{UQ}$ method achieved an R^2 score of 0.813 ± 0.006, representing a significant 14.1 % improvement over the random baseline ($Data_{Random}$). This trend was further amplified in the aerodynamic analysis (DrivAerNet), where $Data_{UQ}$ achieved the highest rank with an R^2 of 0.927 ± 0.003. To verify the reliability of these findings, statistical significance testing was performed across both datasets as summarized in Table 6 to further validate these improvements. Notably, all results confirmed in Table 6 yielded a p-value of less than 0.0001, securing strong statistical significance. This confirms that the Explicit lattice-based $Data_{UQ}$ approach provides statistically significant gains in R^2, MAE, and RMSE over both Point-wise and Random baselines, proving that the observed performance enhancements are due to the integrated framework rather than stochastic chance.

These consistent improvements are further corroborated by the sample-wise performance analysis in Fig. 15. This figure presents hexbin plots showing the changes in R^2 performance for individual samples in the $D_{Large\ Test}$ set. Data points above the $y = x$ line indicate that the specific augmentation strategy (y-axis) improved performance compared to the baseline.

In the DeepJEB dataset (Figs. 15a–15c), $Data_{Add}$ clearly exhibits an overall performance boost, with the majority of samples positioned above the $y = x$ line, implying consistent improvement without significant performance loss. For $Data_{UQ}$, a high density of data points is clustered in the top-right region compared to $Data_{Random}$. This indicates that the performance-aware strategy further refines predictions for samples where the model already performs reasonably well. While $Data_{UQ}$ certainly has fewer samples below the $y = x$ line compared to $Data_{Random}$, it still falls short of the extensive coverage provided by $Data_{Add}$ in this structural domain.

Conversely, in the DrivAerNet dataset (Figs. 15d–15f), which involved augmentation from a small amount of seed data, a distinct trend emerges. The overall center of mass for the $Data_{UQ}$ distribution is located notably higher than that of $Data_{Add}$. This aligns with the quantitative performance shown in Table 4, confirming that the strategy effectively configures the data to further improve high-performing predictions. On the other hand, $Data_{Random}$

and $Data_{Add}$ display strikingly similar distributions. This similarity is likely attributable to the nature of the DrivAerNet dataset itself, which was originally constructed using a random augmentation methodology similar to $Data_{Random}$. These observations confirm that $Data_{UQ}$ achieves effective performance enhancement, particularly in complex domains where random expansion yields diminishing returns.

Collectively, these results demonstrate that strategically selecting informative samples is a method to integrate uncertainty with spatial augmentation effectively. The $Data_{Random}$ approach, which relies solely on geometric diversity, fails to capture critical performance boundaries, generating shapes that are geometrically distinct but performantly redundant. The superior performance of $Data_{UQ}$ demonstrates that a performance-aware strategy guided by epistemic uncertainty is universally more efficient for training robust meta-models, thereby addressing the critical research gap in existing augmentation methodologies.

4.3.2. Mechanism: Uncertainty Reduction as the Driver for Performance

This confirms that the active learning loop successfully identified regions within the design space characterized by low prediction confidence and actively acquired data to resolve these knowledge deficits. This mechanism of uncertainty reduction is observable not only in the generalization capability on the large test set but also throughout the training and validation phases across diverse physical domains.

In the case of the DeepJEB dataset, the $Data_{UQ}$ strategy achieved the lowest mean epistemic uncertainty in both the training set (D_{train}: -2.084) and the test set ($D_{Large\ Test}$: -2.084), surpassing the reduction achieved by random augmentation (D_{train}: -1.814, $D_{Large\ Test}$: -1.979). This indicates that the framework effectively calibrates the model's confidence during learning, rather than merely improving final predictions by chance.

This trend is similarly observed in the DrivAerNet experiment. As detailed in Table 5, $Data_{UQ}$ demonstrated the most significant reduction in uncertainty across all splits—training, validation, and testing. Specifically, for the training set, the mean uncertainty was reduced to -6.306, and for the validation set, it reached -6.973, consistently outperforming the brute-force $Data_{Add}$ approach. This consistent reduction across all data partitions demonstrates that the proposed method actively constructs a more reliable and confident meta-model by systematically illuminating epistemic voids in the high-dimensional design space, regardless of the underlying physics. The framework successfully identified the regions where the model was unaware and actively acquired data to fill those gaps, enhancing the model's overall reliability, which in turn led to higher predictive accuracy on unseen data (as shown in Table 3 & 4).

4.3.3. Justification for the Novel UQ Architecture

A second critical research gap was how to effectively integrate uncertainty with spatial augmentation. The comparison between the proposed explicit lattice-based model and the standard point-wise model provides a definitive answer. Under the same $Data_{UQ}$ strategy, the explicit model consistently outperformed the point-wise model on the large test set ($D_{Large\ Test}$).

- **DeepJEB:** Explicit ($R^2 = 0.813 \pm 0.006$) vs. Point-wise ($R^2 = 0.676 \pm 0.006$)

- **DrivAerNet:** Explicit ($R^2 = 0.927 \pm 0.003$) vs. Point-wise ($R^2 = 0.779 \pm 0.005$)

The performance gap is particularly pronounced in the DrivAerNet case (approx.30% difference). This suggests that for complex physics such as fluid dynamics, where flow features (e.g., pressure gradients, wakes) are inherently spatial and regional, a point-wise estimate fails to capture the necessary correlations. This finding is further corroborated by the hyperparameter sensitivity analysis in Sec. 4.2.1. The point-wise model's high sensitivity to architectural depth suggests it struggles to find a stable representation of this complex field, whereas the explicit model's robustness indicates its structural suitability.

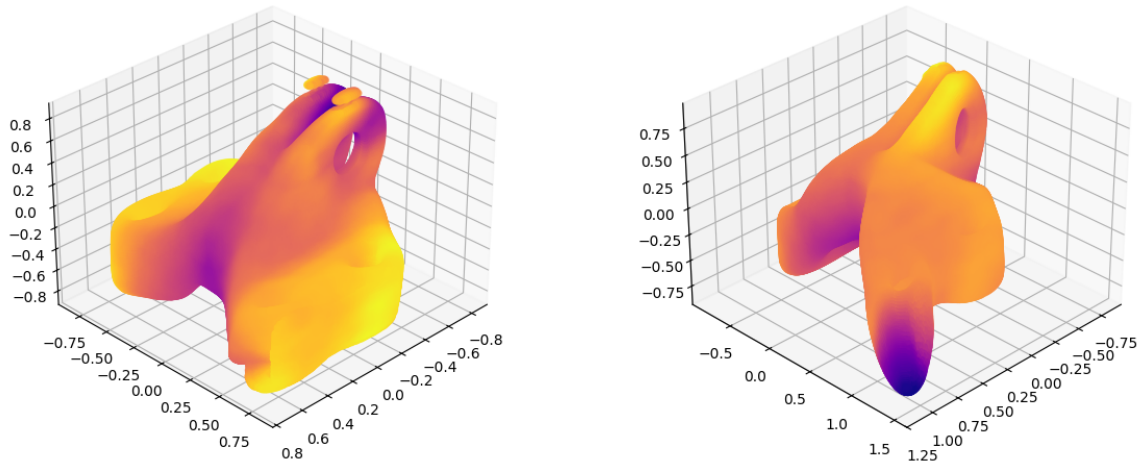
The proposed lattice-based architecture inherently aggregates spatial information, aligning perfectly with the FFD control structure. Mathematically, the lattice acts as a spatial regularizer. In data-scarce scenarios, point-wise ensembles often produce noisy uncertainty maps with high-frequency fluctuations due to lack of constraints. The explicit lattice filters this noise by projecting the uncertainty onto a lower-dimensional continuous grid. This validates the proposed architecture as a necessary innovation for enabling effective spatial augmentation, ensuring that deformations are guided by meaningful trends rather than local noise.

4.3.4. Ablation Study

Data type	# of train data		R^2 score	MAE ($\times 10^{-3}$)	RMSE ($\times 10^{-3}$)
$Data_{RBF}$	50 (+500)	D_{train}	0.867 ± 0.001	3.50 ± 0.01	4.91 ± 0.02
		D_{valid}	0.772 ± 0.006	4.46 ± 0.06	6.43 ± 0.09
		D_{test}	0.509 ± 0.011	5.52 ± 0.03	7.96 ± 0.05
		$D_{Large\ Test}$	0.646 ± 0.005	5.25 ± 0.03	7.54 ± 0.05
$Data_{MSE}$	50 (+500)	D_{train}	0.896 ± 0.001	3.21 ± 0.02	4.05 ± 0.02
		D_{valid}	0.793 ± 0.002	4.53 ± 0.04	6.19 ± 0.05
		D_{test}	0.430 ± 0.029	6.22 ± 0.08	8.32 ± 0.09
		$D_{Large\ Test}$	0.713 ± 0.005	5.54 ± 0.03	7.14 ± 0.06
$Data_{UQ100}$	100 (+500)	D_{train}	**0.943 ± 0.001**	**2.14 ± 0.01**	**2.96 ± 0.02**
		D_{valid}	**0.882 ± 0.003**	**3.27 ± 0.06**	**4.60 ± 0.08**
		D_{test}	**0.796 ± 0.005**	**3.86 ± 0.05**	**5.39 ± 0.07**
		$D_{Large\ Test}$	**0.855 ± 0.004**	**3.80 ± 0.06**	**5.23 ± 0.08**
$Data_{Ctrl10}$	50 (+500)	D_{train}	0.937 ± 0.001	2.45 ± 0.01	3.70 ± 0.02
		D_{valid}	0.874 ± 0.002	3.41 ± 0.03	4.98 ± 0.03
		D_{test}	0.738 ± 0.003	4.29 ± 0.02	6.32 ± 0.02
		$D_{Large\ Test}$	0.815 ± 0.002	4.08 ± 0.03	5.96 ± 0.03

Table 7: Ablation study of the proposed framework on the DeepJEB dataset

To investigate the individual contribution of the technical components within the proposed framework, an ablation study was conducted on the DeepJEB dataset, comparing



(a) Local topological discontinuity

(b) Abnormal geometric stretching

Figure 16: Representative failure modes in 3D shape augmentation

the integrated framework against several variants: $Data_{RBF}$, $Data_{MSE}$, $Data_{UQ100}$, and $Data_{Ctrl10}$. The quantitative results of these experiments are summarized in Table 7.

Effect of Morphing Mechanism: FFD vs. RBF The contribution of the spatial morphing mechanism was evaluated by substituting the proposed FFD approach with RBF interpolation ($Data_{RBF}$). This variant was implemented using the same lattice-based architecture and pre-trained models to isolate the effect of the morphing technique. Unlike FFD, which deforms the entire embedded space via a structured lattice to ensure smooth geometric transitions, RBF induces deformations where control points act independently. This lack of spatial coupling often leads to geometric distortions, such as localized stretching or mesh self-intersections, which inherently limit the effective range of permissible shape variations. Consequently, $Data_{RBF}$ achieved an R^2 value of 0.646 ± 0.005 on the $D_{Large\ Test}$. This performance is not only inferior to FFD-based augmentation but also indicates that the limitations of RBF morphing can effectively negate the potential benefits of data augmentation in complex 3D engineering domains.

The failure mode associated with RBF is illustrated in Fig. 16a. Because RBF control points exert independent influence without the global spatial constraints of a lattice, the resulting deformations often induce local topological discontinuities. This fragmentation of the mesh surface prevents the generation of smooth, continuous geometries, thereby rendering the augmented data ineffective for model training.

Effect of Query Strategy: Epistemic Uncertainty vs. MSE To verify the advantage of uncertainty-driven sampling, an alternative active learning strategy guided by point-wise MSE ($Data_{MSE}$) was examined. Because this strategy relies on direct prediction residuals from the training data, a simplified architecture that outputs only distance values

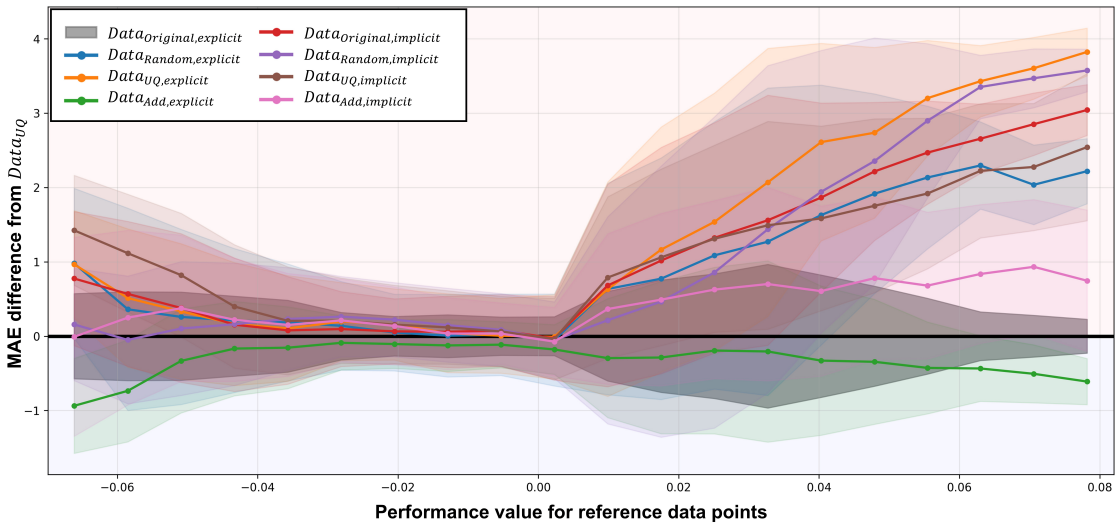
was utilized instead of the explicit lattice-based UQ architecture. While point-wise MSE provides an intuitive approach to exploring regions with known high errors, it does not capture epistemic uncertainty, the model's lack of knowledge in unexplored regions. The $Data_{MSE}$ strategy yielded an R^2 value of 0.713 ± 0.005 on the $D_{Large\ Test}$. Although this outperformed the $Data_{Random}$ baseline, the result remained significantly lower than the R^2 score achieved by the proposed $Data_{UQ}$ configuration (0.813 ± 0.006, detailed in Table 3). This confirms that epistemic uncertainty is a more reliable and efficient indicator of informative regions in data-scarce design spaces than deterministic error metrics.

Influence of Seed Dataset Diversity While the primary focus of this research is intentionally directed toward highly data-scarce regimes (50 samples) to reflect realistic industrial scenarios where computational budgets for high-fidelity simulations are strictly limited, the impact of initial data diversity on model scalability was further examined. The performance of the framework was evaluated by increasing the number of seed geometries from 50 to 100 ($Data_{UQ100}$).

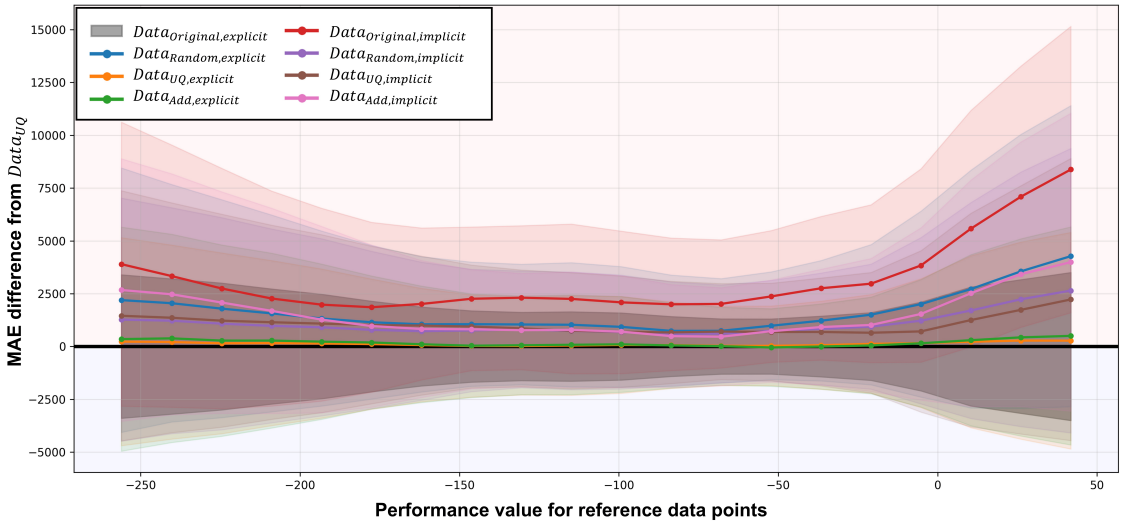
While previous experiments utilized a configuration of 50 samples augmented 10 times each, $Data_{UQ100}$ utilized 100 seed samples with only 5 augmentations per sample to maintain a similar total training dataset size. This configuration achieved an R^2 score of 0.855 ± 0.004 on the $D_{Large\ Test}$, representing the highest performance among all tested variants. This finding demonstrates that while the proposed augmentation method effectively enhances model accuracy in minimal-data environments, the inherent diversity of the initial seed data remains a critical factor for achieving superior model generalization. The results underscore that as the variety of the seed dataset expands, the information gain from targeted augmentation is further amplified, leading to more robust predictive performance across complex design spaces.

Impact of Lattice Resolution Finally, the influence of FFD lattice resolution was investigated by increasing the grid size to $10 \times 10 \times 10$ ($Data_{Ctrl10}$). Although this variant yielded an R^2 value of 0.815 ± 0.002 on the $D_{Large\ Test}$, indicating performance comparable to the default $Data_{UQ}$ configuration, it highlighted potential risks to geometric consistency. As illustrated in Fig. 16b, increasing the number of control points without a corresponding adjustment to the displacement variance can precipitate abnormal geometric stretching. At higher resolution, while offering finer control, the likelihood of cumulative localized shifts producing exaggerated features or self-intersections increases, potentially diverging from the valid design space. To prioritize structural validity and ensure that generated shapes remain physically plausible while maintaining optimal accuracy, a more conservative lattice resolution was adopted for the final framework.

To proactively mitigate these potential geometric inconsistencies, the framework was refined to incorporate an adaptive range adjustment mechanism. This procedure utilizes the quantified epistemic uncertainty to dynamically scale the augmentation range at each control point. By ensuring that displacement magnitudes are appropriately bounded according to local uncertainty levels, the framework maintains a critical balance between exploration diversity and the geometric validity of the augmented samples.



(a) DeepJEB bracket dataset



(b) DrivAerNet vehicle dataset

Figure 17: Performance-dependent point-wise MAE errors on $D_{Large\ Test}$

4.3.5. Point-wise Error Distribution Analysis

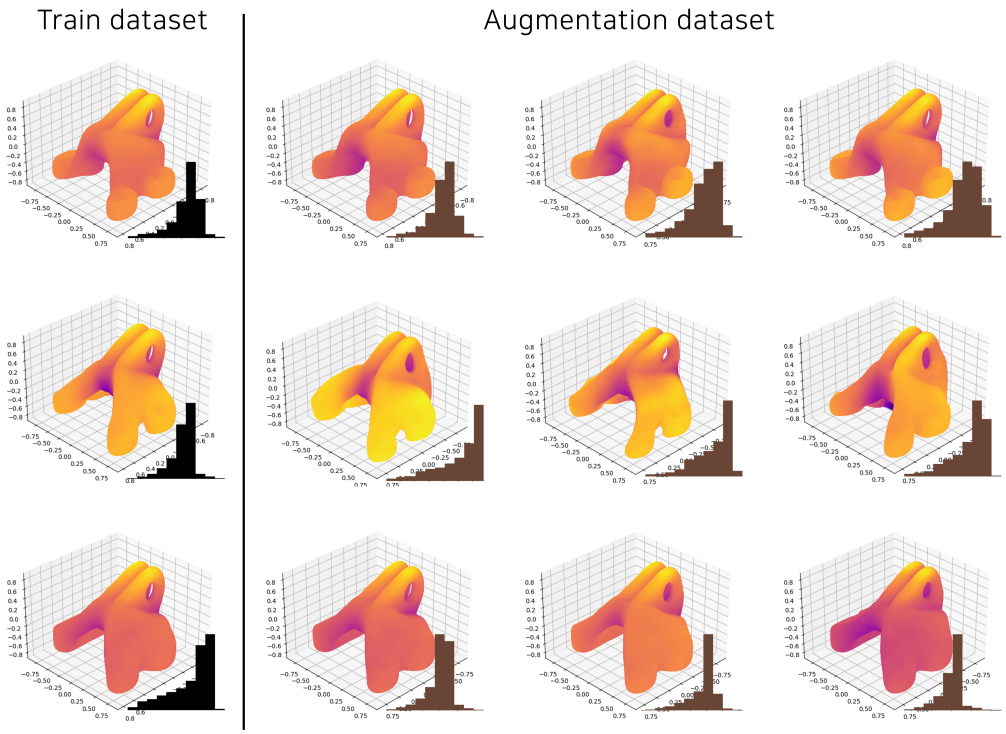
To address the limitations of aggregate metrics like R^2 and further scrutinize the reliability of the Baseline ($Data_{UQ}$) strategy, a point-wise error analysis was conducted on the $D_{Large\ Test}$ set (Fig. 17). This analysis evaluates the difference in MAE relative to the $Data_{UQ}$ model across the entire spectrum of performance values. Critically, data points below the horizontal baseline ($y = 0$) indicate a lower MAE than the model trained from $Data_{UQ}$, indicating an improvement in predictive accuracy. For the DeepJEB dataset (Fig. 17a), the $Data_{Add}$ approach continues to demonstrate superior overall suppression of prediction errors, maintaining its lead even in regions with high performance values where sample data is comparatively scarce. This confirms that simple data addition provides a robust performance upper bound in structural domains. Conversely, in the DrivAerNet dataset (Fig. 17b), the Baseline ($Data_{UQ}$) strategy yields the most consistent and significant error reduction across all performance regimes. Notably, the fact that these point-wise error distributions closely mirror the global performance trends observed in Tables 3 and 4 validates the framework's effectiveness. The results demonstrate that the proposed uncertainty-guided augmentation not only improves average metrics but also ensures stable and desired accuracy refinements at a granular level across the design space. Furthermore, as evidenced by Fig. 17, the framework achieves outstanding performance improvements across the entire performance spectrum, including regions with extreme data scarcity. Consequently, these granular-level refinements naturally enhance the reliability of domain-specific metrics derived from these field predictions, such as peak stress accuracy in structural mechanics and drag coefficient precision in aerodynamics.

5. Conclusion

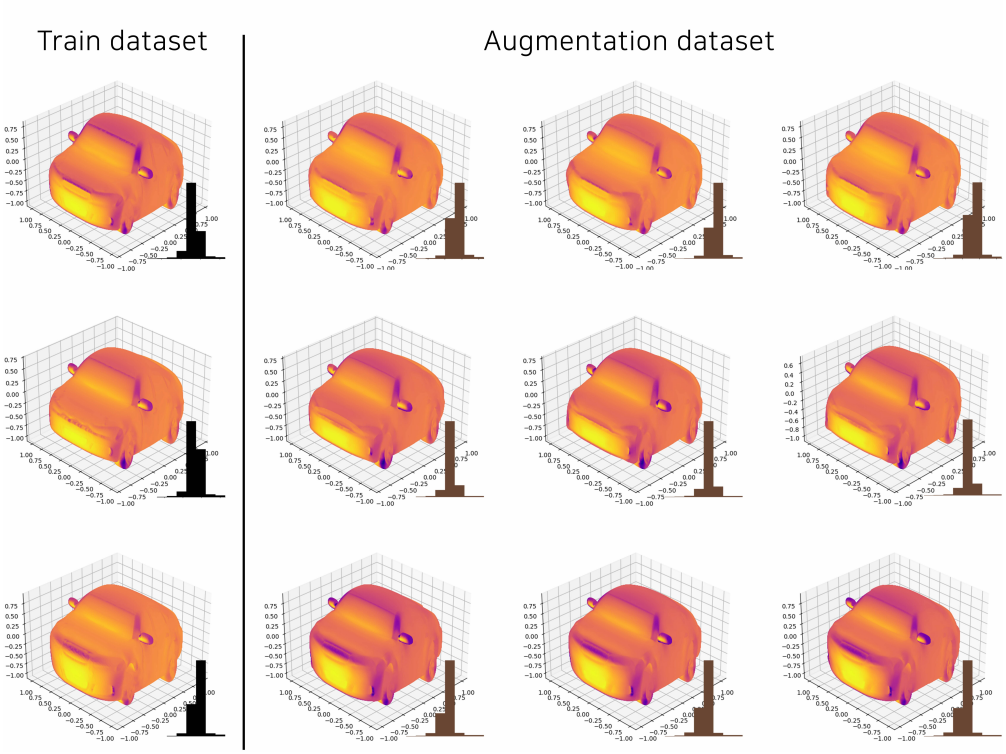
5.1. Summary of Contributions

This study aimed to redefine data augmentation methodologies to address the practical data scarcity phenomenon encountered in the industrial field, thereby enabling the efficient construction of meta-models. A new data sampling and model training framework was proposed, coupling performance and shape through the combination of geometry-based engineering performance uncertainty quantification and FFD-based data augmentation. By simultaneously considering the meta-model's input (shape) and output (performance), this framework offers greater freedom from overfitting than existing methods that only consider shape diversity, and it can achieve better performance and shape coverage. The effectiveness of this approach was confirmed through experiments that augmented data from a set of 50 original data points in DeepJEB. The results, compared to training with 500 arbitrary data points, quantitatively demonstrated that the augmented data is more competitive than randomly added data.

In particular, a novel UQ architecture was proposed to seamlessly connect the UQ of the INR-based meta-model and FFD augmentation. Compared to a point-wise regressor on the same data, this architecture showed overall performance improvements. Specifically, for the DeepJEB bracket dataset, compared to the performance of the point-wise regressor trained on the $Data_{Original}$, the proposed UQ-considered meta-model showed a significant performance improvement of 21.0 % and exhibited excellent performance, differing by only 10.6



(a) DeepJEB brackets



(b) DrivAerNet vehicles

Figure 18: Visualized examples of augmented shape data

% from the $Data_{Add}$. Furthermore, in the aerodynamic analysis (DrivAerNet), the proposed method demonstrated even superior capabilities, effectively outperforming the brute-force $Data_{Add}$ approach. This suggests that effective data augmentation based on performance distributions is feasible and well-integrated with the FFD lattice grid. It was confirmed that the data obtained through this augmentation methodology significantly improves the performance of the meta-model. Furthermore, statistical significance tests confirmed that all performance improvements achieved by the proposed framework are highly significant, with p-values consistently below 0.0001. This underscores the reliability of the proposed approach in efficiently constructing meta-models even under extreme data scarcity.

5.2. Engineering Impact and Practical Guidance

These technical contributions translate directly into significant practical engineering impact. The proposed framework provides a tangible pathway for engineers in the conceptual design stage to build highly accurate meta-models while minimizing the number of high-fidelity simulations, which are often the most expensive and time-consuming bottleneck in the entire design process. By achieving competitive performance with a strategically augmented dataset rather than a much larger, randomly collected one, this method directly enables mitigating this bottleneck. Ultimately, this allows for more rapid and reliable design optimization cycles, accelerates the overall product development timeline, and reduces R&D costs associated with expensive physical prototyping and computational analysis.

For practitioners seeking to implement this framework, several considerations must be addressed. First, the FFD lattice resolution selection involves a trade-off between fidelity and cost. A low resolution (e.g., $3 \times 3 \times 3$) is computationally cheap but may fail to capture local uncertainty, while a high resolution (e.g., $10 \times 10 \times 10$) offers greater precision in capturing local uncertainty but introduces two significant challenges. First, it increases the cost of uncertainty quantification and optimization exponentially (N^3). Second, the dense control grid allows sharp, localized deformations, increasing the risk of generating infeasible or unrealistic shapes (e.g., self-intersections) that deviate significantly from the original seed data's valid geometry.

Second, computational resource requirements are a key factor.The deep ensemble (M=3) is inherently more costly than a single model. Experiments were conducted on an NVIDIA RTX 4090 ($24GB$ VRAM), with the ensemble training taking approximately 30-50 minutes. The process of augmenting 10 samples for each of the 50 seed shapes (totaling 500 augmented data points) and applying Algorithm 1 optimization required approximately 80 minutes. It is important to note that this processing time is directly proportional to the number of points in each shape's point cloud.

Third, this approach is recommended for data-scarce scenarios where simulations are prohibitively expensive. As shown in Table 3 & 4, while the proposed method ($Data_{UQ}$) is significantly more data-efficient than $Data_{Random}$, the generalized approach of simply adding more data ($Data_{Add}$) can still yield superior performance in certain contexts. This confirms that the proposed framework is best suited for "efficient refinement" when starting with minimal data, rather than for "global exploration" when data is abundant.

5.3. Limitations and Future Work

Although promising results were obtained, several limitations of this framework warrant further investigation and define clear directions for future research.

A key limitation lies in how uncertainty is handled. The uncertainty addressed in this framework is epistemic, broadly divided into spatial uncertainty (where to deform within a single shape) and shape uncertainty (which seed shape to select for augmentation). In this study, data were augmented by the same number for each training data point, making it unnecessary to separate these uncertainties. However, the potential contribution of individual data points to improving meta-model performance is not uniform. For instance, a single data point may effectively represent a large portion of the shape and performance space. At the same time, other regions may be challenging to cover, even with dozens of data augmentations. This highlights the necessity for methodologies capable of decomposing spatial and shape uncertainty.

Building upon this research, future work could propose several advanced methodologies. One direction is to develop a hybrid DGM-FFD framework. This would leverage DGMs for global, topological exploration to create novel design concepts, while our method would then act as an efficient refiner, using its performance-aware FFD to rapidly and locally optimize these new promising shapes.

A related application is to use this framework as a repair mechanism. Instead of just generating new shapes, this method could take the geometrically inconsistent synthetic data from 2D-to-3D foundation models and apply performance- and geometry-aware deformations to transform these "unusable" shapes into "usable", physically plausible data for engineering applications [74, 75].

Finally, the framework could be expanded by integrating modern foundation models, such as Vision Language Models. These models could be used to objectively extract high-level shape features and characteristics (e.g., "aggressive", "streamlined"). This information could then be used to build a more sophisticated multi-agent system capable of quantifying not only performance-based uncertainty and shape diversity, but also subjective aesthetic values. This would expand the framework from purely engineering-driven optimization to a more holistic design process that balances performance with human-centric metrics.

CRediT authorship contribution statement

Yongmin Kwon: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Resources; Software; Validation; Visualization; and Roles/Writing - original draft. **Namwoo Kang**: Project administration, Supervision, Investigation, Funding acquisition, Writing – Review & Editing.

Acknowledgements

This work was supported by the Ministry of Science and ICT of Korea grant (No.2022-0-00969, No.2022-0-00986, and No.GTL24031-000) and the Ministry of Trade, Industry & Energy grant (RS-2024-00410810, and RS-2025-02317327).

References

[1] C. Yuan, T. Marion, M. Moghaddam, Leveraging end-user data for enhanced design concept evaluation: A multimodal deep regression model, Journal of Mechanical Design 144 (2) (2022) 021403.

[2] X. Li, C. Xie, Z. Sha, Design representation for performance evaluation of 3d shapes in structure-aware generative design, Design Science 9 (2023) e27.

[3] N. Kang, Generative ai-driven design optimization: eight key application scenarios, JMST Advances (2025) 1–7.

[4] S. Kim, M. Seo, N. Kang, Decoupled dynamics framework with neural fields for 3d spatio-temporal prediction of vehicle collisions, arXiv preprint arXiv:2503.19712 (2025).

[5] J. Kim, J. Park, N. Kim, Y. Yu, K. Chang, C.-S. Woo, S. Yang, N. Kang, Physics-constrained graph neural networks for spatio-temporal prediction of drop impact on oled display panels, Expert Systems with Applications 274 (2025) 126907.

[6] J. Park, N. Kang, Point-deeponet: A deep operator network integrating pointnet for nonlinear analysis of non-parametric 3d geometries and load conditions, arXiv preprint arXiv:2412.18362 (2024).

[7] M. Elrefaie, A. Dai, F. Ahmed, Drivaernet: A parametric car dataset for data-driven aerodynamic design and graph-based drag prediction, in: International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, Vol. 88360, American Society of Mechanical Engineers, 2024, p. V03AT03A019.

[8] B. Lakshminarayanan, A. Pritzel, C. Blundell, Simple and scalable predictive uncertainty estimation using deep ensembles, Advances in neural information processing systems 30 (2017).

[9] S. Yang, K. Yee, Towards reliable uncertainty quantification via deep ensemble in multi-output regression task, Engineering Applications of Artificial Intelligence 132 (2024) 107871.

[10] K. Hornik, M. Stinchcombe, H. White, Multilayer feedforward networks are universal approximators, Neural networks 2 (5) (1989) 359–366.

[11] A. R. Barron, Universal approximation bounds for superpositions of a sigmoidal function, IEEE Transactions on Information theory 39 (3) (1993) 930–945.

[12] S. Hong, Y. Kwon, D. Shin, J. Park, N. Kang, Deepjeb: 3d deep learning-based synthetic jet engine bracket dataset, Journal of Mechanical Design 147 (4) (2025).

[13] M. Elrefaie, F. Morar, A. Dai, F. Ahmed, Drivaernet++: A large-scale multimodal car dataset with computational fluid dynamics simulations and deep learning benchmarks, Advances in Neural Information Processing Systems 37 (2025) 499–536.

[14] J. Brüning, F. Hellmeier, P. Yevtushenko, T. Kühne, L. Goubergrits, Uncertainty quantification for non-invasive assessment of pressure drop across a coarctation of the aorta using cfd, Cardiovascular engineering and technology 9 (2018) 582–596.

[15] S. Yoo, N. Kang, Deepwheel: Generating a 3d synthetic wheel dataset for design and performance evaluation, arXiv preprint arXiv:2504.11347 (2025).

[16] A. I. Heft, T. Indinger, N. A. Adams, Experimental and numerical investigation of the drivaer model, in: Fluids Engineering Division Summer Meeting, Vol. 44755, American Society of Mechanical Engineers, 2012, pp. 41–51.

[17] E. Whalen, A. Beyene, C. Mueller, Simjeb: simulated jet engine bracket dataset, in: Computer Graphics Forum, Vol. 40, Wiley Online Library, 2021, pp. 9–17.

[18] Y. Kwon, N. Kang, Three-dimensional deep shape optimization with a limited dataset (2025). arXiv:2506.12326.
URL https://arxiv.org/abs/2506.12326

[19] N. J. Bagazinski, F. Ahmed, Ship-d: Ship hull dataset for design optimization using machine learning, in: International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, Vol. 87301, American Society of Mechanical Engineers, 2023, p. V03AT03A028.

[20] S. Khan, K. Goucher-Lambert, K. Kostas, P. Kaklis, Shiphullgan: A generic parametric modeller for ship hull design using deep convolutional generative model, Computer Methods in Applied Mechanics and Engineering 411 (2023) 116051.

[21] W. Chen, A. Ramamurthy, Deep generative model for efficient 3d airfoil parameterization and generation, in: AIAA Scitech 2021 Forum, 2021, p. 1690.

[22] D. Sieger, S. Menzel, M. Botsch, On shape deformation techniques for simulation-based design optimization, New Challenges in Grid Generation and Adaptivity for Scientific Computing (2015) 281–303.

[23] F. Salmoiraghi, A. Scardigli, H. Telib, G. Rozza, Free-form deformation, mesh morphing and reduced-order methods: enablers for efficient aerodynamic shape optimisation, International Journal of Computational Fluid Dynamics 32 (4-5) (2018) 233–247.

[24] J. Li, M. Zhang, J. R. Martins, C. Shu, Efficient aerodynamic shape optimization with deep-learning-based geometric filtering, AIAA journal 58 (10) (2020) 4243–4259.

[25] C. Picard, J. Schiffmann, F. Ahmed, Dated: Guidelines for creating synthetic datasets for engineering design applications, in: International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, Vol. 87301, American Society of Mechanical Engineers, 2023, p. V03AT03A015.

[26] D. Lee, Y.-C. Chan, W. W. Chen, L. Wang, A. van Beek, W. Chen, t-metaset: Tailoring property bias of large-scale metamaterial datasets through active learning, arXiv preprint arXiv:2202.10565 (2022).

[27] T. W. Sederberg, S. R. Parry, Free-form deformation of solid geometric models, in: Proceedings of the 13th annual conference on Computer graphics and interactive techniques, 1986, pp. 151–160.

[28] R. MacCracken, K. I. Joy, Free-form deformations with lattices of arbitrary topology, in: Proceedings of the 23rd annual conference on Computer graphics and interactive techniques, 1996, pp. 181–188.

[29] M. Botsch, L. Kobbelt, Real-time shape editing using radial basis functions, in: Computer graphics forum, Vol. 24, Blackwell Publishing, Inc Oxford, UK and Boston, USA, 2005, pp. 611–621.

[30] B. S. Morse, T. S. Yoo, P. Rheingans, D. T. Chen, K. R. Subramanian, Interpolating implicit surfaces from scattered surface data using compactly supported radial basis functions, in: ACM SIGGRAPH 2005 Courses, 2005, pp. 78–es.

[31] N. Kojekine, V. V. Savchenko, M. Senin, I. Hagiwara, Real-time 3d deformations by means of compactly supported radial basis functions., in: Eurographics (Short Presentations), 2002.

[32] V. V. Savchenko, A. A. Pasko, O. G. Okunev, T. L. Kunii, Function representation of solids reconstructed from scattered surface points and contours, in: Computer graphics forum, Vol. 14, Wiley Online Library, 1995, pp. 181–188.

[33] H. Zhao, D. Kamensky, J. T. Hwang, J.-S. Chen, Automated shape and thickness optimization for non-matching isogeometric shells using free-form deformation, Engineering with Computers (2024) 1–24.

[34] M. Gomez, M. C. Galbraith, Analysis driven shape design using free-form deformation of parametric cad geometry, in: AIAA SCITECH 2024 Forum, 2024, p. 1317.

[35] G. Padula, F. Romor, G. Stabile, G. Rozza, Generative models for the deformation of industrial shapes with linear geometric constraints: Model order and parameter space reductions, Computer Methods in Applied Mechanics and Engineering 423 (2024) 116823.

[36] M. Tezzele, F. Salmoiraghi, A. Mola, G. Rozza, Dimension reduction in heterogeneous parametric spaces with application to naval engineering shape design problems, Advanced Modeling and Simulation in Engineering Sciences 5 (2018) 1–19.

[37] H. Gagnon, D. Zingg, Two-level free-form deformation for high-fidelity aerodynamic shape optimization, in: 12th AIAA Aviation Technology, Integration, and Operations (ATIO) Conference and 14th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference, 2012, p. 5447.

[38] F. Gagliardi, K. C. Giannakoglou, Rbf-based morphing of b-rep models for use in aerodynamic shape optimization, Advances in Engineering Software 138 (2019) 102724.

[39] A. Ashukha, A. Lyzhov, D. Molchanov, D. Vetrov, Pitfalls of in-domain uncertainty estimation and ensembling in deep learning, arXiv preprint arXiv:2002.06470 (2020).

[40] F. K. Gustafsson, M. Danelljan, T. B. Schon, Evaluating scalable bayesian deep learning methods for robust computer vision, in: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops, 2020, pp. 318–319.

[41] S. Fort, H. Hu, B. Lakshminarayanan, Deep ensembles: A loss landscape perspective, arXiv preprint arXiv:1912.02757 (2019).

[42] L. Cao, Y. Zhao, Uncertainty quantification for structural response field with ultra-high dimensions, International Journal of Mechanical Sciences 271 (2024) 109110.

[43] D. Q. Nguyen, K. Q. Tran, T. D. Le, M. A. Wahab, H. Nguyen-Xuan, A data-driven uncertainty quantification framework in probabilistic bio-inspired porous materials (material-uq): An investigation for rottmps plates, Computer Methods in Applied Mechanics and Engineering 435 (2025) 117603.

[44] J. Shen, R. Ren, A. Ruiz, F. Moreno-Noguer, Estimating 3d uncertainty field: Quantifying uncertainty for neural radiance fields, in: 2024 IEEE International Conference on Robotics and Automation (ICRA), IEEE, 2024, pp. 2375–2381.

[45] Z. Feng, H. Zhan, Z. Chen, Q. Yan, X. Xu, C. Cai, B. Li, Q. Zhu, Y. Xu, Naruto: Neural active reconstruction from uncertain target observations, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024, pp. 21572–21583.

[46] B. Settles, Active learning literature survey (2009).

[47] T. Qu, S. Guan, Y. Feng, G. Ma, W. Zhou, J. Zhao, Deep active learning for constitutive modelling of granular materials: From representative volume elements to implicit finite element modelling, International Journal of Plasticity 164 (2023) 103576.

[48] Z. Meng, Z. Zhang, H. Zhou, H. Chen, B. Yu, Robust design optimization of imperfect stiffened shells using an active learning method and a hybrid surrogate model, Engineering Optimization 52 (12) (2020) 2044–2061.

[49] C. Zheng, T. Ji, F. Xie, X. Zhang, H. Zheng, Y. Zheng, From active learning to deep reinforcement learning: Intelligent active flow control in suppressing vortex-induced vibration, Physics of Fluids 33 (6) (2021).

[50] M. Tripathi, S. Kumar, Y. B. Desale, R. S. Pant, Active learning-cfd integrated surrogate-based framework for shape optimization of lta systems, in: AIAA AVIATION FORUM AND ASCEND 2024, 2024, p. 4120.

[51] A. Hernandez-Garcia, N. Saxena, M. Jain, C.-H. Liu, Y. Bengio, Multi-fidelity active learning with gflownets, arXiv preprint arXiv:2306.11715 (2023).

[52] R. Pellegrini, J. Wackers, R. Broglia, A. Serani, M. Visonneau, M. Diez, A multi-fidelity active learning method for global design optimization problems with noisy evaluations, Engineering with Computers 39 (5) (2023) 3183–3206.

[53] C. Hai, W. Qian, W. Wang, L. Mei, Active learning-assisted multi-fidelity surrogate modeling based on geometric transformation, Computer Methods in Applied Mechanics and Engineering 426 (2024) 116990.

[54] T.-T. Nguyen, V.-H. Dang, D.-M. Hoang, X.-D. Pham, T.-H. Nguyen, V.-T. Dinh, Robust active learning framework for structural reliability analysis using uncertainty quantification and flexible meta-model, in: Structures, Vol. 63, Elsevier, 2024, p. 106465.

[55] H. Mertens, F. Zhu, Comparative analysis of uncertainty quantification models in active learning for efficient system identification of dynamical systems, in: 2024 IEEE 20th International Conference on Automation Science and Engineering (CASE), IEEE, 2024, pp. 1869–1876.

[56] Y. Shi, P. Wei, K. Feng, D.-C. Feng, M. Beer, A survey on machine learning approaches for uncertainty quantification of engineering systems, Machine Learning for Computational Science and Engineering 1 (1) (2025) 1–39.

[57] S. Yoo, S. Lee, S. Kim, K. H. Hwang, J. H. Park, N. Kang, Integrating deep learning into cad/cae system: generative design and evaluation of 3d conceptual wheel, Structural and multidisciplinary optimization 64 (4) (2021) 2725–2747.

[58] W. Chen, F. Ahmed, Padgan: A generative adversarial network for performance augmented diverse designs, in: International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, Vol. 84003, American Society of Mechanical Engineers, 2020, p. V11AT11A010.

[59] Y.-C. Chan, F. Ahmed, L. Wang, W. Chen, Metaset: Exploring shape and property spaces for data-driven metamaterials design, Journal of Mechanical Design 143 (3) (2021) 031707.

[60] J. J. Park, P. Florence, J. Straub, R. Newcombe, S. Lovegrove, Deepsdf: Learning continuous signed distance functions for shape representation, in: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2019, pp. 165–174.

[61] V. Sitzmann, J. Martel, A. Bergman, D. Lindell, G. Wetzstein, Implicit neural representations with periodic activation functions, Advances in neural information processing systems 33 (2020) 7462–7473.

[62] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, R. Ng, Nerf: Representing scenes as neural radiance fields for view synthesis, Communications of the ACM 65 (1) (2021) 99–106.

[63] A. Kashefi, D. Rempe, L. J. Guibas, A point-cloud deep learning framework for prediction of fluid flow fields on irregular geometries, Physics of Fluids 33 (2) (2021).

[64] J. He, S. Koric, D. Abueidda, A. Najafi, I. Jasiuk, Geom-deeponet: A point-cloud-based deep operator network for field predictions on 3d parameterized geometries, Computer Methods in Applied Mechanics and Engineering 429 (2024) 117130.

[65] D. P. Kingma, J. Ba, Adam: A method for stochastic optimization, arXiv preprint arXiv:1412.6980 (2014).

[66] S. Peng, M. Niemeyer, L. Mescheder, M. Pollefeys, A. Geiger, Convolutional occupancy networks, in: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part III 16, Springer, 2020, pp. 523–540.

[67] C. R. Qi, H. Su, K. Mo, L. J. Guibas, Pointnet: Deep learning on point sets for 3d classification and segmentation, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 652–660.

[68] C. R. Qi, L. Yi, H. Su, L. J. Guibas, Pointnet++: Deep hierarchical feature learning on point sets in a metric space, Advances in neural information processing systems 30 (2017).

[69] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18, Springer, 2015, pp. 234–241.

[70] H. Jasak, A. Jemcov, Z. Tukovic, et al., Openfoam: A c++ library for complex physics simulations, in: International workshop on coupled methods in numerical dynamics, Vol. 1000, Dubrovnik, Croatia), 2007, pp. 1–20.

[71] T. Akiba, S. Sano, T. Yanase, T. Ohta, M. Koyama, Optuna: A next-generation hyperparameter optimization framework, in: Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining, 2019, pp. 2623–2631.

[72] J. Bergstra, R. Bardenet, Y. Bengio, B. Kégl, Algorithms for hyper-parameter optimization, Advances in neural information processing systems 24 (2011).

[73] F. Hutter, H. Hoos, K. Leyton-Brown, An efficient approach for assessing hyperparameter importance, in: International conference on machine learning, PMLR, 2014, pp. 754–762.

[74] Z. Lai, Y. Zhao, H. Liu, Z. Zhao, Q. Lin, H. Shi, X. Yang, M. Yang, S. Yang, Y. Feng, et al., Hunyuan3d 2.5: Towards high-fidelity 3d assets generation with ultimate details, arXiv preprint arXiv:2506.16504 (2025).

[75] Y. Li, Z.-X. Zou, Z. Liu, D. Wang, Y. Liang, Z. Yu, X. Liu, Y.-C. Guo, D. Liang, W. Ouyang, et al., Triposg: High-fidelity 3d shape synthesis using large-scale rectified flow models, arXiv preprint arXiv:2502.06608 (2025).