Solar-GECO: Perovskite Solar Cell Property Prediction with Geometric-Aware Co-Attention

Lucas Li[1], **Jean-Baptiste Puel**[2,3], **Florence Carton**[1], **Dounya Barrit**[1,3], **Jhony H. Giraldo**[4]

[1]TotalEnergies OneTech, 91120 Palaiseau, France
[2]EDF, Electricité de France, 91120 Palaiseau, France
[3]IPVF, Institut Photovoltaïque d'Ile-de-France, 91120 Palaiseau, France
[4]LTCI, Télécom Paris, Institut Polytechnique de Paris, 91120 Palaiseau, France

Abstract

Perovskite solar cells are promising candidates for next-generation photovoltaics. However, their performance as multi-scale devices is determined by complex interactions between their constituent layers. This creates a vast combinatorial space of possible materials and device architectures, making the conventional experimental-based screening process slow and expensive. Machine learning models try to address this problem, but they only focus on individual material properties or neglect the important geometric information of the perovskite crystal. To address this problem, we propose to predict perovskite **solar** cell power conversion efficiency with a **ge**ometric-aware **co**-attention (Solar-GECO) model. Solar-GECO combines a geometric graph neural network (GNN)—that directly encodes the atomic structure of the perovskite absorber—with language model embeddings that process the textual strings representing the chemical compounds of the transport layers and other device components. Solar-GECO also integrates a co-attention module to capture intra-layer dependencies and inter-layer interactions, while a probabilistic regression head predicts both power conversion efficiency (PCE) and its associated uncertainty. Solar-GECO achieves state-of-the-art performance, significantly outperforming several baselines, reducing the mean absolute error (MAE) for PCE prediction from 3.066 to 2.936 compared to semantic GNN (the previous state-of-the-art model). Solar-GECO demonstrates that integrating geometric and textual information provides a more powerful and accurate framework for PCE prediction.

1 Introduction

Machine learning has transformed materials science by enabling fast prediction of isolated material properties, such as bandgap, formation energy, or carrier mobility [1, 2, 3]. While these efforts have accelerated the discovery of promising compounds, they often focus on single-scale problems where the target property is intrinsic to the material itself. However, in real-world applications, optimal device performance arises from the coupled behavior of multiple components across diverse scales [4]. For complex optoelectronic devices such as perovskite solar cells, efficiency depends not only on the properties of the perovskite absorber, but also on the underlying interactions between transport layers, electrodes, and their interfaces [5]. This multiscale interdependence poses challenges that go beyond conventional single-material property prediction [6].

Perovskite solar cells have achieved rapid progress in laboratory efficiency [7], but their commercialization encounters a fundamental bottleneck: the extensive combinatorial space of potential materials, device architectures, and processing conditions. Each device layer—such as the hole transport layer (HTL), electron transport layer (ETL), and encapsulation—can be produced using several prospective materials, each with its own variants in stoichiometry, morphology, and processing [8, 9]. Therefore,

the total number of possible configurations grows exponentially. Physics-guided and intuition-driven design cannot explore this design space at the pace required by current innovation cycles [10, 11]. The result is a gap between the growing diversity of prospect materials and the rate at which optimal full-device architectures can be identified.

Conventional development pipelines often adopt a sequential strategy: first identifying high-performing materials in isolation, then attempting to integrate them into full devices [12]. However, this approach can be misleading because many promising combinations only exhibit their full potential when considered holistically, as interactions between layers can enhance—or severely degrade—performance [13]. This phenomenon is not unique to perovskite solar cells; similar issues arise in other multicomponent systems such as batteries, catalysts, and thermoelectrics [14]. Bridging the gap between layer-level and full-device performance prediction requires models that can represent both intra-layer properties and inter-layer relationships within the same device.

In this work, we propose to predict perovskite **solar** cell power conversion efficiency with a **ge**ometric-aware **co**-attention (Solar-GECO) model. Solar-GECO is a hybrid algorithm where we explicitly integrate crystal-level information from the perovskite layer with device-level architectural context. Unlike prior work on semantic device graphs [6], which rely solely on text embeddings from a large language model (LLM), our method processes the crystal structure of the perovskite absorber directly with a geometric graph neural network (GNN). This allows the extraction of physically grounded features from the crystal structure as shown in Figure 1. Other layers in the device are encoded using LLM-derived molecular embeddings.

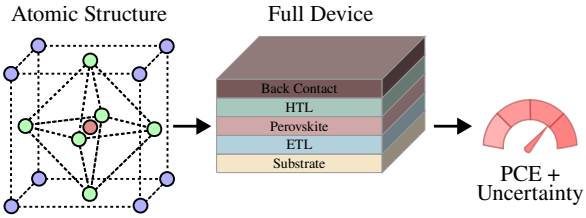

Figure 1: Our model integrates the atomic structure of the perovskite absorber (left) with the device context of stacked layers (center) through a co-attention module. The fused representation is used to predict the power conversion efficiency (PCE) of the device and its associated uncertainty (right).

In Solar-GECO, we also introduce self-attention and cross-attention mechanisms to jointly model intra-layer dependencies and inter-layer interactions, capturing how the interaction between atoms in the perovskite layer and the context layer of the device propagates to the device performance. Significant variability in the PCE may derive from fabrication process sensitivities, including parameters like humidity and annealing temperature, and unmodeled factors such as human error, material insolubility, and poor wettability. To account for some of this inherent uncertainty, our model predicts the PCE using a Gaussian negative log-likelihood (NLL) loss function. We evaluate Solar-GECO on a curated subset of the Perovskite Database [15] and the Materials Project [16], achieving state-of-the-art performance. Our main contributions are as follows:

- We propose a novel model, Solar-GECO, that combines crystal-level geometric GNN encoding of the perovskite absorber with LLM-based molecular embeddings for the context device layers.
- We introduce a co-attention module that combines self-attention within layers and cross-attention across layers, enabling mutual refinement of graph and text representations.
- Solar-GECO models uncertainty in PCE predictions by training with a Gaussian NLL loss.
- Our model achieves state-of-the-art for PCE prediction in perovskite solar devices.

2 Related Work

Property prediction in materials. Several machine learning approaches have been applied to predict material properties. Early efforts focused on composition-to-property regression using algorithms such as gradient boosting regression tree, kernel ridge regression, and support vector machines to estimate bandgaps and stability metrics [17, 18, 19]. More recent methodologies include structure-based models that only rely on chemical composition. For example, CrabNet [20] uses a Transformer-based architecture to learn the relative importance of elements for target property prediction, without requiring any explicit structure information. With the rise of geometric deep learning, GNNs have become the dominant paradigm for predicting material properties from structure,

as they can easily represent atomic interactions and capture complex composition–structure–property relationships [21, 22]. In particular, geometric GNNs, which incorporate 3D atomic coordinates and respect physical symmetries such as $\mathrm{E}(3)$-invariance/equivariance, have shown superior data efficiency and generalization across materials domains [23]. For crystalline materials, specialized architectures such as Matformer [24] explicitly encode periodic boundary conditions and lattice geometry, enabling improved prediction of electronic and structural properties.

Multimodal and attention based fusion. Traditional early- and late-fusion methods rely on explicit feature concatenation or decision-level averaging, which limits their ability to capture rich cross-modal interactions [25, 26]. More recently, attention-based multimodal architectures have been proposed to model interactions directly across modalities[27].

Device-level perovskite property prediction. While most prior work targets isolated material properties, the recent *semantic device graphs* method [6] models the entire device stack, representing each layer and its interfaces as a heterogeneous graph. LLMs are used to embed the textual strings representing the chemical compounds in the device, and a GNN is applied to capture inter-layer relationships. This holistic view enables the identification of high-performing layer combinations that might be overlooked by sequential screening of individual materials. However, this method represents the perovskite absorber only through text embeddings, ignoring its crystallographic structure.

Contrary to previous works, we explicitly include the atomic structure information in the prediction of PCE in perovskite devices. To this end, we directly process the perovskite's atomic structure with a geometric GNN, incorporating physically grounded features that improve PCE prediction accuracy.

3 Methodology: Solar-GECO

3.1 Preliminaries

Notations. In this work, we denote sets using calligraphic letters (*e.g.*, $\mathcal{V}$), with their cardinality written as $|\mathcal{V}|$. Bold uppercase letters (*e.g.*, $\mathbf{H}$) represent matrices, and bold lowercase letters (*e.g.*, $\mathbf{x}$) represent vectors. The transpose operator is denoted by $(\cdot)^\top$ and concatenation by $[\,\cdot\,;\,\cdot\,]$.

Perovskite solar device. A perovskite solar cell is a multi-layered photovoltaic device. A typical architecture consists of five different layers as shown in the center of Figure 1. The substrate is the foundational base, typically glass with an FTO (fluorine-doped tin oxide) or ITO (indium tin oxide) layer that acts as the front electrode. We also have two charge transport layers: *i*) the ETL which extracts electrons, and *ii*) the HTL that extracts the holes. In the core of the device, we have the perovskite absorber, which absorbs sunlight to generate electron-hole pairs. Finally, the back contact is an electrode that completes the electrical circuit.

Graph representation of crystals. In materials science, a crystal structure can be naturally represented as a graph $G = (\mathcal{V}, \mathcal{E})$, where the set of vertices $\mathcal{V}$ corresponds to the atoms in the crystal unit cell, and the set of edges $\mathcal{E}$ represents the connections or bonds between them. Typically, an edge exists between two atoms if their interatomic distance is within a specified cutoff radius. The features of each node can include atomic properties (*e.g.*, electronegativity, atomic mass), and edge features can represent the distances between connected atoms.

Geometric graph neural networks. GNNs are deep learning models designed to operate on graph-structured data by learning node representations through iterative message passing [28]. For 3D atomic systems—such as molecules, proteins, and crystalline materials—a specialized class of models known as geometric GNNs has become the state-of-the-art architecture [23].

These models represent atomic systems as geometric graphs, where atoms are nodes endowed with coordinates in 3D Euclidean space ($\mathbb{R}^3$). The core principle of a geometric GNN is the incorporation of fundamental physical symmetries as a powerful inductive bias. Specifically, the learned representations are constrained to respect the symmetries of the Euclidean group $\mathrm{E}(3)$, which encompasses all rotations, reflections, and translations, as well as permutation invariance with respect to node indexing. This property is known as $\mathrm{E}(3)$-equivariance. Formally, if the input atomic coordinates are transformed by an operation $g(\cdot) \in \mathrm{E}(3)$, the model's feature vectors at each layer must also transform predictably according to a corresponding group representation of $g(\cdot)$. A special case of equivariance is invariance, where the output remains unchanged under these transformations , such as the total potential energy of a system, which is invariant to rotations and translations.

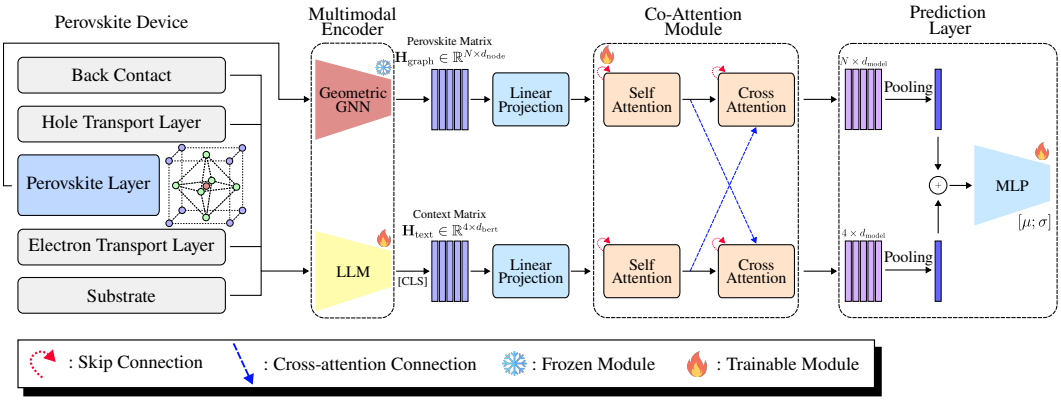

Figure 2: Overview of Solar-GECO. Our model combines a geometric GNN encoder for the perovskite absorber's 3D atomic structure with a device text encoder (LLM) for chemical descriptions of the substrate, ETL, HTL, and back contact. Their outputs are fused in a co-attention module with self- and cross-attention layers to model intra-layer dependencies and inter-layer interactions. Pooled features are concatenated and passed to a probabilistic regression head that predicts the mean and variance of PCE, trained with a Gaussian negative log-likelihood loss.

Attention mechanism. Solar-GECO uses the multi-head scaled dot-product attention mechanism [29] to process intra- and inter-layer information within the perovskite device. In general terms, this mechanism weighs the importance of different elements in a sequence or set of tokens. An attention function is a mapping of a query and a set of key-value pairs to an output. The output is computed as a weighted sum of the values, where the weight assigned to each value is computed by a compatibility function of the query with the corresponding key.

Formally, given a query matrix $\mathbf{Q} \in \mathbb{R}^{N \times d_k}$, a key matrix $\mathbf{K} \in \mathbb{R}^{N \times d_k}$, and a value matrix $\mathbf{V} \in \mathbb{R}^{N \times d_v}$, the scaled dot-product attention is calculated as:

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^\top}{\sqrt{d_k}}\right)\mathbf{V}, \tag{1}$$

where d_k is the dimension of the queries and keys, d_v is the dimension of the values, and N is the number of tokens. The matrix product $\mathbf{Q}\mathbf{K}^\top$ computes the dot product between every possible pair of rows in $\mathbf{Q}$ and $\mathbf{K}$. In other words, this inner product computes the similarity between each query and all keys. The result is scaled by $1/\sqrt{d_k}$, and a softmax function is applied to obtain the so-called attention weights. Finally, we weigh the values matrix with the attention weights, which assign different levels of importance (attention) to the set of values. Our model uses multi-head attention, which applies this mechanism multiple times in parallel with different, learned linear projections of $\mathbf{Q}$, $\mathbf{K}$, and $\mathbf{V}$, defined as follows:

$$\begin{aligned} \text{MultiHead}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) &= [\mathbf{M}_1; \ldots; \mathbf{M}_h]\,\mathbf{W}^O, \\ \text{where } \mathbf{M}_i &= \text{Attention}(\mathbf{Q}\mathbf{W}_i^Q, \mathbf{K}\mathbf{W}_i^K, \mathbf{V}\mathbf{W}_i^V). \end{aligned} \tag{2}$$

Here, h is the number of heads, and $\mathbf{W}_i^Q \in \mathbb{R}^{d_{\text{model}} \times d_k}$, $\mathbf{W}_i^K \in \mathbb{R}^{d_{\text{model}} \times d_k}$, $\mathbf{W}_i^V \in \mathbb{R}^{d_{\text{model}} \times d_v}$, $\mathbf{W}^O \in \mathbb{R}^{hd_v \times d_{\text{model}}}$ are learnable projection matrices for the head i, where d_{model} is the dimensionality of the input embeddings. The multi-head attention allows the model to jointly attend to information from different representation subspaces at different positions.

Problem definition. Given a dataset of device-PCE pairs $\mathcal{D} = \{\mathbf{x}_i, \mathbf{y}_i\}_{i=1}^M$, where $\mathbf{x}_i$ is the ith perovskite solar device and $\mathbf{y}_i$ is its PCE, our objective is to train a model $F_{\boldsymbol{\theta}}$ with parameters $\boldsymbol{\theta}$ to predict the PCE of each device. In practice, since our datasets come from experimental work, one single $\mathbf{x}_i$ might have different values of PCE. In this work, we use neural networks to parametrize $F_{\boldsymbol{\theta}}$.

3.2 Overview

Figure 2 shows the pipeline of our method. Solar-GECO is designed to predict the PCE of perovskite cells by integrating information from the perovskite's crystal structure and the device's text representations. The pipeline consists of three main stages. First, we employ two specialized, pre-trained

encoders to independently extract features: a crystal graph convolutional neural network (CGCNN) [30] processes the atomic structure of the perovskite, while a MaterialsBERT model [31] processes the text string representing the chemical compounds of the device's functional layers. Second, these distinct feature sets are fed into a multi-layer co-attention module, which processes the interdependent relationships between the material's structural properties and the device's architectural context. Finally, the fused, context-aware representations are passed to a probabilistic regression head that predicts the PCE and the inherent uncertainty of its prediction, optimized using a Gaussian NLL loss.

3.3 Multi-Modal Feature Extraction

The initial step of our model involves extracting high-quality representations from the two distinct input modalities: the perovskite's crystal structure and the device's text representations.

Crystal graph encoder. We use a pre-trained CGCNN [30] to encode the perovskite's atomic structure. Given a crystal graph G, the input representation for each node is obtained based on its atomic number. CGCNN applies a series of graph convolution operations where each atom's feature vector is updated based on the features of its neighboring atoms and the distances between them. This process yields a feature vector for each atom (node) in the graph, resulting in a sequence of node features $\mathbf{H}_{\text{graph}} \in \mathbb{R}^{N \times d_{\text{node}}}$, where N is the number of atoms in the perovskite layer.

Device text encoder. For the device architecture, we use a pre-trained MaterialsBERT model [31] to encode the textual strings representing the chemical compounds of the four context layers: the substrate, ETL, HTL, and back contact. The input string of each layer is individually tokenized and processed by the model. Following standard practice [32], we extract the hidden state of the special [CLS] token from the encoder as the representative embedding for each component. These four embedding vectors are then stacked to form a sequence of text-based features $\mathbf{H}_{\text{text}} \in \mathbb{R}^{4 \times d_{\text{bert}}}$.

3.4 Co-Attention Fusion Module

To effectively model the relationships between the perovskite and the context layers of the device, we introduce a co-attention fusion module. To this end, we define a co-attention fusion layer by stacking a self-attention and a cross-attention operation for each branch (graph and text) as shown in Figure 2. This module enables the graph and text representations to be mutually refined.

Intra-modal self-attention. The first step within each fusion layer is to refine the representation of each modality independently using self-attention. In this case, the queries, keys, and values are all derived from the same input sequence, allowing each element (atom or text token) to attend to all other elements within its modality. For the graph representation, let $\mathbf{H}_{\text{graph}}^{(l-1)}$ be the output of the previous $l-1$ co-attention fusion layer of the graph branch, where $1 \leq l \leq L$, and $\mathbf{H}_{\text{graph}}^{(0)} = \mathbf{H}_{\text{graph}}\mathbf{W}_{\text{graph}}$, *i.e.*, the output of the CGCNN after linear projection with $\mathbf{W}_{\text{graph}} \in \mathbb{R}^{d_{\text{node}} \times d_{\text{model}}}$. The self-attention block is then defined as:

$$\mathbf{H'}_{\text{graph}}^{(l)} = \gamma\left(\mathbf{H}_{\text{graph}}^{(l-1)} + \text{MultiHead}\left(\mathbf{H}_{\text{graph}}^{(l-1)}, \mathbf{H}_{\text{graph}}^{(l-1)}, \mathbf{H}_{\text{graph}}^{(l-1)}\right)\right), \tag{3}$$

where $\gamma(\cdot)$ is a normalization function. An analogous operation is performed for the text branch, where $\mathbf{H}_{\text{text}}^{(0)} = \mathbf{H}_{\text{text}}\mathbf{W}_{\text{text}}$, *i.e.*, the output of the BERT model after linear projection. This self-attention step updates the embedding for each element by aggregating information from all other elements within the same modality before the cross-attention module.

Inter-modal cross-attention. The second step is the core of the fusion process: inter-modal bidirectional cross-attention. Here, representations from the two modalities query each other. To update the graph representation, its self-attended features $\mathbf{H}'_{\text{graph}}$ act as the queries, while the self-attended text features $\mathbf{H}'_{\text{text}}$ provide the keys and values. This allows each atom's representation to be updated with information from the entire device stack. The operation is thus defined as:

$$\mathbf{H}_{\text{graph}}^{(l)} = \gamma\left(\mathbf{H'}_{\text{graph}}^{(l)} + \text{MultiHead}\left(\mathbf{H'}_{\text{graph}}^{(l)}, \mathbf{H'}_{\text{text}}^{(l)}, \mathbf{H'}_{\text{text}}^{(l)}\right)\right). \tag{4}$$

Similarly, for the text branch, we have:

$$\mathbf{H}_{\text{text}}^{(l)} = \gamma\left(\mathbf{H'}_{\text{text}}^{(l)} + \text{MultiHead}\left(\mathbf{H'}_{\text{text}}^{(l)}, \mathbf{H'}_{\text{graph}}^{(l)}, \mathbf{H'}_{\text{graph}}^{(l)}\right)\right). \tag{5}$$

The entire process of self-attention followed by cross-attention is repeated for L layers, allowing for progressively deeper integration of the multi-modal information. This architecture allows the model to iteratively learn which atoms in the crystal are most relevant to specific device layers, and conversely, which device layers are most influenced by specific crystal structure.

3.5 Prediction Layer

Layer-wise fusion and pooling. After the final cross-attention layer, we apply masked average pooling on the graph features $\mathbf{H}_{\text{graph}}^{(L)} \in \mathbb{R}^{N \times d_{\text{model}}}$ and text features $\mathbf{H}_{\text{text}}^{(L)} \in \mathbb{R}^{4 \times d_{\text{model}}}$. Therefore, we obtain single fixed-size vectors $\mathbf{v}_{\text{graph}} \in \mathbb{R}^{d_{\text{model}}}$ and $\mathbf{v}_{\text{text}} \in \mathbb{R}^{d_{\text{model}}}$ for the perovskite and context layers, respectively. These features are concatenated to form the final feature vector, *i.e.*, $\mathbf{v}_{\text{final}} = [\mathbf{v}_{\text{graph}}; \mathbf{v}_{\text{text}}] \in \mathbb{R}^{2d_{\text{model}}}$, which is then passed to the regression head.

Final prediction head. To account for the inherent uncertainty in PCE prediction, Solar-GECO is designed to predict a probability distribution that represents the possible range of PCEs. In this work, we assume this distribution to be Gaussian. Therefore, we add a head to Solar-GECO for mapping the fused representation to the parameters of the target probability distribution. More precisely, the vector $\mathbf{v}_{\text{final}}$ serves as the input to the final prediction head, which is implemented as a Multi-Layer Perceptron (MLP). This output layer produces two scalar values, which are interpreted as the predicted mean $\mu(x)$ and the standard deviation $\sigma(x)$ of the Gaussian distribution $y \sim \mathcal{N}(\mu(x), \sigma(x)^2)$:

$$[\mu(x); \sigma(x)] = \text{MLP}(\mathbf{v}_{\text{final}}). \tag{6}$$

3.6 Loss Function

We frame the prediction task as a probabilistic regression to enable the model to predict the target PCE value and quantify the prediction uncertainty. To this end, the model is trained to predict the parameters of a Gaussian distribution, $\mathcal{N}(\mu, \sigma^2)$, for each sample [33].

Formally, we train the model to minimize the Gaussian NLL of the true target values given the predicted distributions. The probability density function for a single true target value y given a predicted mean μ and variance σ^2 is given by:

$$P(y|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right). \tag{7}$$

The training objective is to maximize the log-likelihood of the data, which is equivalent to minimizing its negative. The NLL loss is therefore:

$$\mathcal{L}_{\text{NLL}} = \frac{1}{B}\sum_{i=1}^{B} -\log P(y_i|\mu_i, \sigma_i^2) = \frac{1}{2B}\sum_{i=1}^{B}\left(\log(2\pi\sigma_i^2) + \frac{(y_i-\mu_i)^2}{\sigma_i^2}\right), \tag{8}$$

where B is the batch size. During optimization, the constant term $\frac{1}{2}\log(2\pi)$ can be disregarded as it does not affect the location of the minimum. This yields the final loss function used in our work, averaged over a batch of samples:

$$\mathcal{L} = \frac{1}{2B}\sum_{i=1}^{B}\left(\log(\sigma_i^2) + \frac{(y_i-\mu_i)^2}{\sigma_i^2}\right). \tag{9}$$

This loss function consists of two terms. The term $\frac{(y-\mu)^2}{\sigma^2}$ is the variance-scaled squared error, which encourages the predicted mean μ to be close to the true value y. The term $\log(\sigma^2)$ acts as a regularization term that penalizes the model for predicting excessively high variance. Building on the NLL loss formulation, we could assess the calibration of predicted uncertainties by grouping samples into quantile bins and analyzing the relationship between predicted σ and observed errors [34]. The details are provided in Appendix A.

4 Experiments and Results

In this section, we detail our experimental setup, covering datasets, implementation details, and evaluation metrics. Then, we compare Solar-GECO against four baseline architectures designed for

Table 1: Statistical summary of the dataset curation process, showing the reduction in total records and material diversity after filtering.

Metric	Original	Final	Reduction (%)
Total records	43,398	29,344	32.4
Unique perovskite formulas	465	34	92.7
Unique ETL materials	1,468	1,159	21.0
Unique HTL materials	1,978	1,416	28.4
Unique back contacts	290	244	15.9
Unique substrates	194	159	18.0

materials and device property prediction. First, we consider the prior state of the art in perovskite device prediction, semantic GNN [6]. This model represents the entire solar cell as a heterogeneous graph and uses a GNN to learn inter-layer relationships from text-based material representations. Second, we adapt CrabNet [20] for our device-level task, where individual layer representations are aggregated via mean pooling to form a single feature vector for the entire device, which is then used for the final PCE prediction. We also implement a BERT+MLP baseline, where the textual descriptions of all device layers are encoded using a pre-trained MaterialsBERT model, and then these features are mean-pooled and fed to an MLP head for the regression task. We also implement a CGCNN+BERT+MLP baseline. This model first generates a global feature vector for the perovskite crystal using the CGCNN encoder and a separate feature vector for the device context by mean-pooling the BERT embeddings of the text layers. These two vectors are then concatenated and passed to a final MLP head for prediction. In addition, we introduce a LLM+Co-Attention baseline, where chemical formulas describing each layer are processed by a large language model and fused with crystal features via a co-attention mechanism before prediction. We also conduct a series of ablation studies on *i*) the LLM, *ii*) the geometric GNN, *iii*) the type of attention mechanism, and *iv*) the type of loss function. Finally, to further examine the impact of different data splitting strategies, we compare Solar-GECO against LLM+Co-Attention and semantic GNN under a group split based on device material configurations.

4.1 Experimental Setup

Dataset. We construct our dataset by combining device-level specifications from the Perovskite Database [15] with the Materials Project [16], extracting functional layer descriptions (HTL, ETL, absorber, back contact, substrate) and the corresponding PCE values.

To incorporate atomic-level structural information, the Materials Project was cross-referenced to retrieve the corresponding crystallographic information. This hybrid data collection strategy results in a multi-modal dataset where each sample is represented by two distinct data types: *i*) a crystallographic information for the geometric construction of the crystal graph, and *ii*) a set of textual strings describing the device context. This filtering process, while essential for structural encoding, significantly refines the dataset as summarized in Table 1. The total number of device records was reduced by 32.4%, from an initial 43,398 to a final 29,344 data points after matching. More critically, this process substantially narrows the chemical space of the perovskite absorbers. The requirement crystal structure in the Materials Project results in a 92.7% reduction in the diversity of unique perovskite formulas, from 465 to 34.

Table 2: Training hyperparameters for the Solar-GECO model.

Hyperparameter	Value
Hidden dimension (co-attention)	64
MLP layer dimensions	$[128, 64, 2]$
Co-attention layers	3
Normalization on co-attention $\gamma(\cdot)$	LayerNorm
Attention heads	4
Dropout rate	0.2
Batch size	16
Learning rate (main modules)	1×10^{-4}
Learning rate (MaterialsBERT)	1×10^{-6}
Learning rate schedule	Cosine
Warm-up epochs	10
Weight decay	1×10^{-5}
Total epochs	200
Early stopping patience	30

Table 3: Performance comparison of Solar-GECO with baseline models. Results are reported as mean $\pm$ standard deviation over three runs with different random seeds. The best results are highlighted in **bold**. Stars indicate that a baseline's performance is statistically significantly different from Solar-GECO: $^{*}p < 0.05$ (significant difference from Solar-GECO), $^{***}p < 0.001$ (highly significant difference from Solar-GECO).

Model	R^2 Score ↑	MAE ↓	Spearman's ρ ↑
BERT+MLP	$0.3863 \pm 0.0043^{***}$	$3.0436 \pm 0.0164^{***}$	$0.5944 \pm 0.0136^{*}$
CGCNN+BERT+MLP	$0.4009 \pm 0.0067^{*}$	$3.0111 \pm 0.0324^{*}$	$0.6109 \pm 0.0047^{*}$
CrabNet	$0.2090 \pm 0.0058^{***}$	$3.3655 \pm 0.0072^{***}$	$0.3807 \pm 0.0024^{***}$
LLM Co-attention	$0.4048 \pm 0.0049^{*}$	$2.9812 \pm 0.0104^{*}$	$0.6120 \pm 0.0003^{*}$
Semantic GNN	$0.3907 \pm 0.0105^{*}$	$3.0668 \pm 0.0471^{*}$	$0.5943 \pm 0.0071^{*}$
Solar-GECO	$\mathbf{0.4179 \pm 0.0042}$	$\mathbf{2.9361 \pm 0.0179}$	$\mathbf{0.6192 \pm 0.0034}$

Metrics. We evaluate Solar-GECO using standard regression metrics (R^2 and MAE), the Spearman's rank correlation coefficient (Spearman's ρ) to assess ranking capability, and the prediction interval coverage probability (PICP) [35] to asses uncertainty calibration.

Implementation details. We train Solar-GECO with the AdamW [36] optimizer and a cosine learning rate schedule (10-epoch warm-up), freezing the pre-trained CGCNN encoder. The dataset is split into 80%/10%/10% for training, validation, and testing, and experiments are repeated with three random seeds. Full hyperparameters are listed in Table 2.

4.2 Comparison with the State of the Art

Table 3 compares Solar-GECO with the baseline models. Across both evaluation metrics, Solar-GECO consistently achieves the best performance, demonstrating significant improvements over existing approaches. We verify the statistical significance of these results using t-tests against all baselines. Stars in Table 3 mark cases where differences are significant ($p < 0.05$ or $p < 0.001$), confirming that the observed gains are robust rather than random fluctuations.

The largest performance gap in Table 3 is relative to CrabNet, showing the limitations of composition-only methods for modeling complex, multi-scale device properties. The BERT+MLP baseline benefits from contextual embeddings provided by the MaterialsBERT model, yet Solar-GECO's statistically significant gains show that explicitly modeling the interactions between atomic structures and device layers is more effective. The performance metrics reveal a statistically significant advantage for the Solar-GECO over the CGCNN+BERT+MLP baseline, achieving a higher R^2 score (0.4179 vs. 0.4009) and a significantly lower MAE of 2.9361 compared to the baseline's 3.0111.

Furthermore, the improved performance of Solar-GECO over the BERT+MLP baseline highlights the significant benefit of incorporating the CGCNN to model the core perovskite layer. This performance improvement can be also attributed to the integration of the co-attention module, since it helps to align the two modalities and capture interactions between device layers. Solar-GECO also outperforms the semantic GNN, it reduces the MAE from 3.0668 to 2.9361 and improves R^2 from 0.3907 to 0.4179, highlighting the importance of incorporating geometric information from the crystal lattice, which text-based embeddings alone cannot capture. In terms of ranking ability, our model

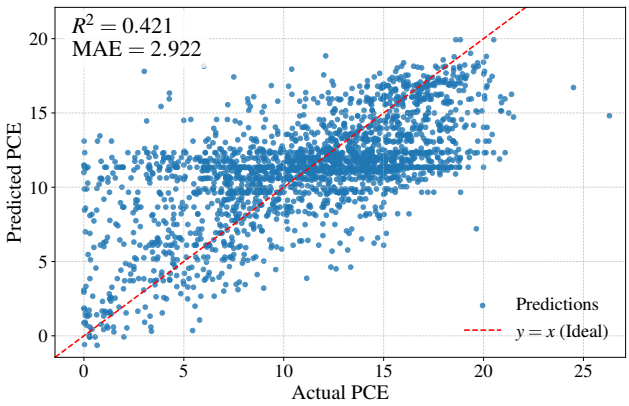

Figure 3: Plot of predicted PCE versus true PCE for the original model on the test set. The diagonal line represents perfect prediction.

Solar-GECO also demonstrates superior performance, achieving a Spearman's ρ of 0.6192 $\pm$ 0.0034. We also evaluate the LLM+Co-Attention baseline, which achieves an MAE of 2.9812 $\pm$ 0.0104, R^2 of 0.4048 $\pm$ 0.0049, and Spearman's ρ of 0.6120 $\pm$ 0.0003. While competitive, its performance is slightly below Solar-GECO, showing that co-attention improves modality alignment but lacks geometric crystal information for optimal accuracy.

Figure 3 further illustrates our model performance (shown for one seed only) through a parity plot of predicted versus actual PCE values. Most predictions cluster tightly around the ideal diagonal, reflecting the model's overall accuracy and stability across the test set. However, deviations are more pronounced for low-PCE devices, where larger relative errors suggest that accurately predicting the PCE of underperforming architectures is inherently more difficult. Nevertheless, Solar-GECO maintains strong global consistency, reinforcing the robustness of our approach.

To evaluate the quality of the model's predicted uncertainty, we construct prediction intervals with a nominal confidence level of 95% and calculate the PICP on the test set. Our model achieved a PICP of 0.9593. This empirical coverage is in agreement with the 95% nominal level (an absolute difference of 0.93%), which indicates that the predicted uncertainty is well-calibrated. In Figure 4, we test the hypothesis that the expected absolute error $\mathbb{E}[|y - \mu(x)|]$ scales linearly with predicted uncertainty $\sigma(x)$ via $c = \sqrt{2/\pi}$. This is supported by the 95% confidence interval (blue band), which reflects sampling variability. The theoretical line (red dashed) lying almost entirely within this band confirms that deviations from the empirical mean (blue line) are not statistically significant, indicating well-calibrated σ values across the output range.

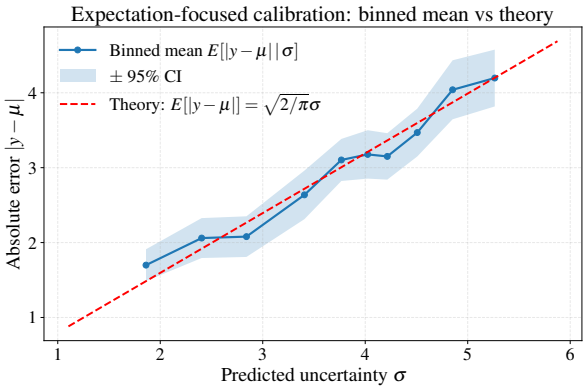

Figure 4: Calibration plot: quantile-binned mean $\overline{|e|}$ vs. $\overline{\sigma}$ with 95% confidence intervals (CIs) for the *mean*, overlaid with the theoretical line $c\,\sigma$, $c = \sqrt{2/\pi}$.

4.3 Ablation Studies

To assess the contribution of each component in Solar-GECO, we perform several ablation experiments by systematically replacing key modules of our model. The results of the ablation studies are shown in Tables 4, 5, 6, and 7.

First, we compare MaterialsBERT [31] with MatSciBERT [37] as the text encoder in Table 4. MaterialsBERT yields a slightly lower MAE (2.922 vs. 2.924), which might be attributed to its broader and more diverse pre-training corpus. Second, we test the crystal graph encoder by replacing the pre-trained CGCNN [30] with CHGNet [38] in Table 5. The performance drop suggests that representations optimized for bulk property prediction are less transferable to the downstream task of device-level PCE estimation, whereas CGCNN features seem to generalize more effectively. Also, we replace the standard cross-attention with a gated contextual cross-attention mechanism [39] in Table 6. This leads to a marked degradation in performance, indicating that the added parametric complexity was not beneficial for our dataset size and may have introduced overfitting. Finally, we also perform an ablation on the loss function, comparing the standard mean squared error (MSE) with the probabilistic Gaussian NLL loss. The results in Table 7 show that both loss functions achieve an identical MAE of 2.922, while the Gaussian NLL loss yields a higher R^2 score (0.421 vs. 0.415). Overall, the ablation results indicate that the combination of MaterialsBERT, CGCNN, and the vanilla attention mechanism provides the best trade-off between model complexity and predictive accuracy, supporting the overall design choices of Solar-GECO.

4.4 Sensitivity Studies

To analyze sensitivity, we adopt an alternative data splitting strategy based on device-layer material groups, in addition to the standard random split. Using this group split, we con-

Table 4: Ablation on the LLM encoder.

Text Encoder	R^2 Score ↑	MAE ↓
MatSciBERT	**0.421**	2.924
MaterialsBERT	**0.421**	**2.922**

Table 5: Ablation on the geometric GNN encoder.

Geometric GNN	R^2 Score ↑	MAE ↓
CHGNet	0.394	3.032
CGCNN	**0.421**	**2.922**

Table 6: Ablation on the attention mechanism.

Attention Function	R^2 Score ↑	MAE ↓
Gated	0.372	3.108
Vanilla	**0.421**	**2.922**

Table 7: Ablation on the loss function.

Loss Function	R^2 Score ↑	MAE ↔
MSE	0.415	2.922
Gaussian NLL	**0.421**	2.922

duct experiments on Solar-GECO, along with the Semantic GNN and LLM+Co-Attention baselines for comparison. The results of this experiment are summarized in Table 8. This analysis reveals that the group-based split, which is designed to test for the generalization to unseen material combinations, presents a significantly more challenging benchmark for all models. Under this split, Solar-GECO demonstrates the highest robustness, achieving the best performance with the lowest MAE (3.1271) and the highest R^2 score (0.3724), which shows its current architecture also effective at generalizing to novel device-layer material groups compared to the baseline approaches.

Table 8: Sensitivity analysis comparing Solar-GECO with Semantic GNN and LLM+Co-Attention under group split.

Model	R^2 Score	MAE
Semantic GNN	0.3700	3.2740
LLM+Co-Attention	0.3374	3.2820
Solar-GECO	**0.3724**	**3.1271**

4.5 Limitations

Despite its strong performance, Solar-GECO has some limitations that suggest directions for future work. First, the geometric data for the perovskite absorber is primarily obtained from the Materials Project [16], which constrains the diversity of available crystal structures and may limit exposure to less-represented materials. Moreover, the current framework focuses on compositional and structural features but does not explicitly account for fabrication-related parameters, such as annealing temperature or deposition methods. These parameters are known to strongly influence device performance, so considering them in the prediction of PCE is an important step towards having a reliable machine learning methodology that could be later used for exploration purposes. Another challenge we observed is the low PCE bias , balancing strategies such as adaptive sampling, importance weighting, or contrastive loss could alleviate this issue.

5 Conclusions

In this work, we introduced Solar-GECO, a multi-modal geometric deep learning framework for predicting the PCE of perovskite solar cells. By combining a geometric GNN encoding of the perovskite crystal structure with LLM-based embeddings of device layers, and fusing them through a co-attention module, our model captures both intra- material properties and inter-device interactions. This representation leads to state-of-the-art accuracy, outperforming established baselines such as CrabNet and semantic GNN. Our results demonstrate that explicitly incorporating crystal geometry alongside device context is essential for modeling complex, multi-scale perovskite devices. Beyond accuracy, Solar-GECO provides a practical tool for screening candidate device architectures, accelerating the discovery of efficient perovskite solar cells.

Future work will proceed on two fronts. First, we will integrate fabrication parameters, such as deposition procedures available in the Perovskite Database, as explicit input features. Second, we will seek to expand the structural data diversity. These combined efforts hold promise for further improving the model's generalization and applicability in the discovery of new perovskite devices.

References

[1] A. Dunn, Q. Wang, A. Ganose, D. Dopp, and A. Jain, "Benchmarking materials property prediction methods: the Matbench test set and Automatminer reference algorithm," *npj Computational Materials*, vol. 6, no. 1, p. 138, 2020.

[2] K. Choudhary, D. Wines, K. Li, K. F. Garrity, V. Gupta, A. H. Romero, J. T. Krogel, K. Saritas, A. Fuhr, P. Ganesh, *et al.*, "JARVIS-leaderboard: a large scale benchmark of materials design methods," *npj Computational Materials*, vol. 10, no. 1, p. 93, 2024.

[3] K. L. K. Lee, C. Gonzales, M. Nassar, M. Spellings, M. Galkin, and S. Miret, "MatSciML: A broad, multi-task benchmark for solid-state materials modeling," in *NeurIPS 2023 Workshop on Machine Learning for Materials Modeling*, 2023.

[4] S. M. Moosavi, K. M. Jablonka, and B. Smit, "The role of machine learning in the understanding and design of materials," *Journal of the American Chemical Society*, vol. 142, no. 48, pp. 20273–20287, 2020.

[5] M. Stolterfoht, P. Caprioglio, C. M. Wolff, J. A. Marquez, J. Nordmann, S. Zhang, D. Rothhardt, U. H"ormann, Y. Amir, A. Redinger, *et al.*, "The impact of energy alignment and interfacial recombination on the internal and external open-circuit voltage of perovskite solar cells," *Energy & Environmental Science*, vol. 12, no. 9, pp. 2778–2788, 2019.

[6] A. Aneesh, N. Alampara, J. A. Marquez, and K. M. Jablonka, "Semantic device graphs for perovskite solar cell design," in *AI for Accelerated Materials Design - ICLR 2025 workshop*, 2025.

[7] J. Guo, B. Wang, D. Lu, T. Wang, T. Liu, R. Wang, X. Dong, T. Zhou, N. Zheng, Q. Fu, *et al.*, "Ultralong carrier lifetime exceeding 20 μs in lead halide perovskite film enable efficient solar cells," *Advanced Materials*, vol. 35, no. 28, p. 2212126, 2023.

[8] F. Liu, Q. Li, and Z. Li, "Hole-transporting materials for perovskite solar cells," *Asian Journal of Organic Chemistry*, vol. 7, no. 11, pp. 2182–2200, 2018.

[9] Z. Liu, N. Rolston, A. C. Flick, T. W. Colburn, Z. Ren, R. H. Dauskardt, and T. Buonassisi, "Machine learning with knowledge constraints for process optimization of open-air perovskite solar cell manufacturing," *Joule*, vol. 6, no. 4, pp. 834–849, 2022.

[10] T. Unold, "Accelerating research on novel photovoltaic materials," *Faraday Discussions*, vol. 239, pp. 235–249, 2022.

[11] P. J. Dale and M. A. Scarpulla, "Efficiency versus effort: A better way to compare best photovoltaic research cell efficiencies?," *Solar Energy Materials and Solar Cells*, vol. 251, p. 112097, 2023.

[12] L. Mao and C. Xiang, "A comprehensive review of machine learning applications in perovskite solar cells: Materials discovery, device performance, process optimization and systems integration," *Materials Today Energy*, vol. 47, p. 101742, 2025.

[13] J. Liang, T. Wu, Z. Wang, Y. Yu, L. Hu, H. Li, X. Zhang, X. Zhu, and Y. Zhao, "Accelerating perovskite materials discovery and correlated energy applications through artificial intelligence," *Energy Mater*, vol. 2, no. 3, p. 200016, 2022.

[14] C. Charalambous, E. Moubarak, J. Schilling, E. Sanchez Fernandez, J.-Y. Wang, L. Herraiz, F. Mcilwaine, S. B. Peh, M. Garvin, K. M. Jablonka, *et al.*, "A holistic platform for accelerating sorbent-based carbon capture," *Nature*, vol. 632, no. 8023, pp. 89–94, 2024.

[15] T. J. Jacobsson, A. Hultqvist, A. Garcia-Fernandez, *et al.*, "An open-access database and analysis tool for perovskite solar cells based on the fair data principles," *Nature Energy*, vol. 7, no. 1, pp. 107–115, 2022.

[16] A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, and K. A. Persson, "The Materials Project: A materials genome approach to accelerating materials innovation," *APL Materials*, vol. 1, no. 1, p. 011002, 2013.

[17] N. Parikh, M. Karamta, N. Yadav, M. M. Tavakoli, D. Prochowicz, S. Akin, A. Kalam, S. Satapathi, and P. Yadav, "Is machine learning redefining the perovskite solar cells?," *Journal of Energy Chemistry*, vol. 66, pp. 74–90, 2022.

[18] J. Im, S. Lee, T.-W. Ko, H. W. Kim, Y. Hyon, and H. Chang, "Identifying Pb-free perovskites for solar cells by machine learning," *npj Computational Materials*, vol. 5, no. 1, p. 37, 2019.

[19] J. Li, B. Pradhan, S. Gaur, and J. Thomas, "Predictions and strategies learned from machine learning to develop high-performing perovskite solar cells," *Advanced Energy Materials*, vol. 9, no. 46, p. 1901891, 2019.

[20] A. Y.-T. Wang, S. K. Kauwe, R. J. Murdock, and T. D. Sparks, "Compositionally restricted attention-based network for materials property predictions," *npj Computational Materials*, vol. 7, no. 1, p. 77, 2021.

[21] J. Gilmer, S. S. Schoenholz, P. F. Riley, O. Vinyals, and G. E. Dahl, "Neural message passing for quantum chemistry," in *International Conference on Machine Learning*, pp. 1263–1272, PMLR, 2017.

[22] K. T. Schütt, H. E. Sauceda, P.-J. Kindermans, A. Tkatchenko, and K.-R. Müller, "SchNet - a deep learning architecture for molecules and materials," *The Journal of Chemical Physics*, vol. 148, no. 24, 2018.

[23] A. Duval, J. Gasteiger, J.-N. Joswig, P. Harris, O.-E. Ganea, T., L. Pattanaik, *et al.*, "A hitchhiker's guide to geometric gnns for 3D atomic systems," *arXiv preprint arXiv:2312.07511*, 2023.

[24] K. Yan, Y. Liu, Y. Lin, and S. Ji, "Periodic graph transformers for crystal material property prediction," *arXiv preprint arXiv:2209.11807*, 2022.

[25] J. Ngiam, A. Khosla, M. Kim, J. Nam, H. Lee, and A. Y. Ng, "Multimodal deep learning," in *Proceedings of the 28th International Conference on Machine Learning (ICML-11)*, pp. 689–696, 2011.

[26] S. Poria, E. Cambria, D. Hazarika, N. Majumder, A. Zadeh, and L.-P. Morency, "Context-dependent sentiment analysis in user-generated videos," in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 873–883, 2017.

[27] Y.-H. H. Tsai, S. Bai, P. P. Liang, J. Z. Kolter, L.-P. Morency, and R. Salakhutdinov, "Multimodal transformer for unaligned multimodal language sequences," in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (ACL)*, pp. 6558–6569, 2019.

[28] F. Scarselli, M. Gori, A. C. Tsoi, M. Hagenbuchner, and G. Monfardini, "The graph neural network model," *IEEE Transactions on Neural Networks*, vol. 20, no. 1, pp. 61–80, 2008.

[29] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in Neural Information Processing Systems*, vol. 30, 2017.

[30] T. Xie and J. C. Grossman, "Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties," *Physical Review Letters*, vol. 120, no. 14, p. 145301, 2018.

[31] P. Shetty, A. C. Rajan, C. Kuenneth, S. Gupta, L. P. Panchumarti, L. Holm, C. Zhang, and R. Ramprasad, "A general-purpose material property data extraction pipeline from large polymer corpora using natural language processing," *npj Computational Materials*, vol. 9, no. 1, p. 52, 2023.

[32] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Conference of the North American Chapter of the Association for Computational Linguistics*, pp. 4171–4186, 2019.

[33] C. M. Bishop, "Mixture density networks," 1994.

[34] D. Levi, L. Gispan, N. Giladi, and E. Fetaya, "Evaluating and calibrating uncertainty prediction in regression tasks," *Sensors*, vol. 22, no. 15, p. 5540, 2022.

[35] N. Dewolf, B. D. Baets, and W. Waegeman, "Valid prediction intervals for regression problems," *Artificial Intelligence Review*, vol. 56, no. 1, pp. 577–613, 2023.

[36] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," in *International Conference on Learning Representations*, 2017.

[37] T. Gupta, M. Zaki, N. A. Krishnan, and Mausam, "MatSciBERT: A materials domain language model for text mining and information extraction," *npj Computational Materials*, vol. 8, no. 1, p. 102, 2022.

[38] B. Deng, P. Zhong, K. Jun, J. Riebesell, K. Han, C. J. Bartel, and G. Ceder, "CHGNet as a pretrained universal neural network potential for charge-informed atomistic modelling," *Nature Machine Intelligence*, vol. 5, no. 9, pp. 1031–1041, 2023.

[39] D. Zhang, R. Nayak, and M. A. Bashar, "Pre-gating and contextual attention gate - a new fusion method for multi-modal data tasks," *Neural Networks*, vol. 179, p. 106553, 2024.

A Uncertainty Calibration Analysis Details

To assess the calibration of predicted uncertainties, we group samples into quantile bins and analyze the relationship between predicted standard deviation $\sigma(x)$ and observed errors, which follows for each input x, a predictive Gaussian distribution $Y \mid X = \mathbf{x} \sim \mathcal{N}\big(\mu(\mathbf{x}),\, \sigma^2(\mathbf{x})\big)$,. Given a dataset $\mathcal{D} = \{(\mathbf{x_i}, y_i)\}_{i=1}^{M}$,we construct N quantile bins on the scalar $\{\sigma_i\}_{i=1}^{M}$,let $\{\mathcal{I}_b\}_{b=1}^{N}$ be a partition of $\{1, \ldots, M\}$ such that if $i \in \mathcal{I}_b$ and $j \in \mathcal{I}_{b'}$ with $b < b'$, then $\sigma_i \leq \sigma_j$, denote $n_b = |\mathcal{I}_b|$. For each bin b, we compute the average predicted uncertainty and the average absolute error:

$$\overline{\sigma}_b := \frac{1}{n_b} \sum_{i \in \mathcal{I}_b} \sigma_i, \qquad \overline{|e|}_b := \frac{1}{n_b} \sum_{i \in \mathcal{I}_b} |e_i|, \tag{10}$$

Define the residual $e := y - \mu(\mathbf{x})$ and the conditional residual $\varepsilon := y - \mu(\mathbf{x}) \mid \mathbf{x}$. Assume the conditional residual is Gaussian:

$$\varepsilon \,\big|\, \mathbf{x} \;\sim\; \mathcal{N}\big(0,\, \sigma(\mathbf{x})^2\big). \tag{11}$$

Equivalently, write $\varepsilon = \sigma(\mathbf{x}) Z$ with $Z \sim \mathcal{N}(0, 1)$. Then:

$$\mathbb{E}[\,|y - \mu(\mathbf{x})| \,|\, \mathbf{x}] = \mathbb{E}[\,|\varepsilon| \,|\, \mathbf{x}] = \sigma(\mathbf{x})\, \mathbb{E}|Z| = c\, \sigma(\mathbf{x}), \tag{12}$$

where the half-normal mean constant is:

$$\mathbb{E}|Z| = \int_{-\infty}^{+\infty} |z| \, \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \, dz = 2 \int_{0}^{+\infty} z \, \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \, dz = \sqrt{\frac{2}{\pi}}. \tag{13}$$

The expected absolute error conditioned on $\mathbf{x}$ scales linearly with the predicted uncertainty $\sigma(\mathbf{x})$ with slope $c = \sqrt{2/\pi}$. We plot points $\big(\overline{\sigma}_b,\, \overline{|e|}_b\big)$ and overlay the theoretical line $c\, \sigma(\mathbf{x})$ from (12).

To express the uncertainty of the bin mean, we draw a 95% confidence interval for the mean using the standard error:

$$\mathrm{SE}_b \;=\; \frac{s_b}{\sqrt{n_b}}, \quad \text{where} \quad s_b^2 = \frac{1}{n_b - 1} \sum_{i \in \mathcal{I}_b} \Big(|e_i| - \overline{|e|}_b \Big)^2, \tag{14}$$

and the band $\overline{|e|}_b \pm 1.96\, \mathrm{SE}_b$.