MATFUSION: A MULTI-MODAL FRAMEWORK BRIDGING LLMS AND STRUCTURAL EMBEDDINGS FOR EXPERIMENTAL MATERIALS PROPERTY PREDICTION

Yuwei Wan[*], **Yuqi An**[*] **& Tong Xie**[†]
GreenDynamics
Sydney, NSW, Australia
tong@greedynamics.com.au

Dongzhan Zhou
Shanghai AI Lab
Shanghai, China

Jiahao Dong
Distyl AI
San Francisco, CA, USA

Chunyu Kit
City University of Hong Kong
Hong Kong, China

Wenjie Zhang & Bram Hoex
University of New South Wales
Kensington, NSW, Australia

Yingheng Wang[†]
Cornell University
Ithaca, NY, USA
yingheng@cs.cornell.edu

ABSTRACT

The scarcity of experimental data in materials science often necessitates property predictions based on large-scale simulations, which may suffer from accuracy and reliability limitations. Uni-modal representations derived from simulated structures inherently incorporate approximations—such as the choice of exchange-correlation functional in Density Functional Theory (DFT)—which constrain machine learning models in capturing complex experimental characterizations. In this work, we propose a novel multi-modal framework, **MatFusion**, that integrates embeddings from domain-specific large language models (LLMs) and structural models to enhance the prediction of experimental material properties. Our approach combines LLM-derived embeddings of material compositions with graph-based structural representations, achieving a 9.15% reduction in mean absolute error (MAE) for experimental bandgap prediction. By leveraging both experiential knowledge from materials science literature and first-principles structural information, our framework transcends traditional representation constraints, offering a powerful paradigm for improving experimental materials property predictions.

1 INTRODUCTION

Materials discovery has traditionally been slow and costly, relying on experimental trial and error, which inherently limits the pace of innovation. Efficient and accurate property prediction is crucial for accelerating advancements in fields such as transistors (Radisavljevic et al., 2011), photovoltaics (Polman et al., 2016), and light-emitting diodes (LEDs)(Schubert & Kim, 2005). Recent breakthroughs in deep learning have shown great promise in expediting this process by providing rapid, scalable predictions of material properties, often leveraging large-scale simulation datasets such as the Materials Project (MP)(Jain et al., 2013) and JARVIS (Choudhary et al., 2020). While these datasets offer valuable insights, they may not fully capture real-world material performance due to inherent approximations, system errors, and microstructural variations (Zhuo et al., 2018; Billah et al., 2025). Beyond simulation-driven approaches, an alternative yet underutilized source of information is materials science literature, which contains a wealth of experiential knowledge. Experimental results recorded in text data can be effectively encoded by large language models (LLMs), generating semantically meaningful embeddings that capture domain-specific insights.

[*]Equal contribution
[†]Corresponding author

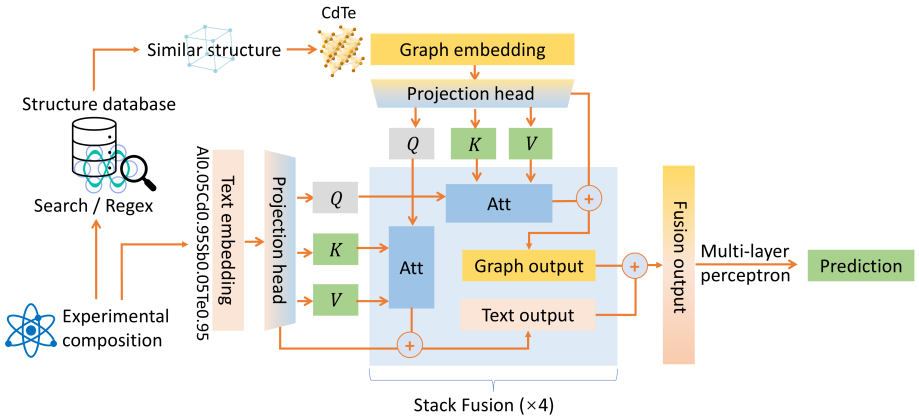

Figure 1: Overview of MatFusion. Materials composition are used to retrieve structure (either exact match or fuzzy match through Regex) in the database. Text and graph embeddings are obtained from LLMs and GNNs. The embeddings are fused through cross-attention and learnable concatenation.

In this work, we introduce MatFusion, a novel multi-modal learning framework that bridges large language model (LLM) embeddings with structural representations to advance experimental materials property prediction (see Figure 1). MatFusion leverages text embeddings from DARWIN (Xie et al., 2024), an LLM designed specifically for materials science, and structural embeddings from the powerful CrystalFormer (Wang et al., 2024) pre-trained on 55,722 computational structures to encode crystal graph representations. We retrieve the structure data based on materials composition using either exact match or fuzzy match through Regex. To effectively integrate these modalities, we develop an enhanced cross-attention module that fuses text and structural embeddings, providing unified representations for downstream predictions.

We evaluate MatFusion on the bandgap prediction task, as bandgap is a fundamental material property. To this end, we curate two datasets: `small` and `large`, where small includes compositional context information, while large consists of compositions alone. Both datasets incorporate experimental and computational data. Through extensive experiments, we find that (1) MatFusion performs better on downstream tasks when using composition embeddings alone rather than incorporating additional context information; (2) MatFusion significantly enhances experimental bandgap prediction but offers little improvement for computational bandgap prediction, where structural embeddings alone are sufficient; (3) domain-specific text embeddings from DARWIN outperform general LLMs (BERT, T5, LLaMA2) in experimental property prediction. Overall, MatFusion improves experimental bandgap prediction accuracy by 23.72% compared to DARWIN's standalone text embeddings and 9.15% compared to CrystalFormer's structural embeddings.

2 Related Work

Language models for materials science. Natural language processing (NLP) has been widely used in materials science. Mat2Vec (Tshitoyan et al., 2019) and MatBERT (Walker et al., 2021) are pre-trained on materials science literature, which can be further fine-tuned to tackle specific tasks, such as crystal property prediction (Korolev & Protsenko, 2023; Rubungo et al., 2023), crystal recommendation and ranking (Qu et al., 2024), and synthesis action retrieval (Song et al., 2023). These approaches offer an alternative to crystal graph representations by modeling structures based on text descriptions (Ganose & Jain, 2019).

GNNs for materials. Crystal structure can be constructed as graphs (Xie & Grossman, 2018; Huang et al., 2022), which crystal GNNs (Chen et al., 2019; Schütt et al., 2021; Batzner et al., 2022; Chen & Ong, 2022; Liao & Smidt, 2022; Yan et al., 2022; Du et al., 2023) are tailored to learn their representations. Recently, CrystalFormer (Wang et al., 2024) proposes a graph construction method that preserves periodic invariance and a crystal graph transformer that can encode angular information, achieving state-of-the-art performance in various property prediction tasks.

Multi-modal models for molecule. MoleculeBind Mirza et al. (2024) aligned SMILES, molecular graphs, and 3D structures. Chemeleon (Park et al., 2024) integrates text descriptions in a diffusion model trained on crystal structures. Ock et al. (2024) integrates transformer-based language models with GNNs in latent space to improve adsorption energy predictions. Both studies utilize

contrastive learning (Chen et al., 2020; You et al., 2020; Wang et al., 2021a;b; Luo et al., 2023) for self-supervised training.

3 MATFUSION

As illustrated in Figure 1, materials composition is used to match and retrieve the structure in the database. If no exact match can be found, we use regular expressions (Regex) to extract the core composition and perform a next round of search. If multiple structures exist in the database for the same composition, the one with the lowest formation energy will be selected. Graph and text embeddings are obtained from a GNN model and an LLM (in our case, CrystalFormer and DARWIN), respectively. A cross-attention module will be then employed to fuse the text and graph embeddings.

Specifically, we first project the embedding from each modality to a shared dimension d. Let $\mathbf{H}_{\text{m1}} \in \mathbb{R}^{B\times d}$ and $\mathbf{H}_{\text{m2}} \in \mathbb{R}^{B\times d}$ denote the projected embeddings of two modalities. In each MatFusion layer, the scaled dot-product m1-to-m2 cross-attention will be computed as follows:

$$\mathbf{Q}^{(\text{m1})} = \mathbf{W}_q^{(\text{m1})}\,\mathbf{H}_{\text{m1}}, \quad \mathbf{K}^{(\text{m2})} = \mathbf{W}_k^{(\text{m2})}\,\mathbf{H}_{\text{m2}}, \quad \mathbf{V}^{(\text{m2})} = \mathbf{W}_v^{(\text{m2})}\,\mathbf{H}_{\text{m2}}.$$

$$\text{CrossAtt}\big(\mathbf{Q}^{(\text{m1})}, \mathbf{K}^{(\text{m2})}, \mathbf{V}^{(\text{m2})}\big) = \text{softmax}\Big(\frac{\mathbf{Q}^{(\text{m1})}\,\mathbf{K}^{(\text{m2})\top}}{\sqrt{d}}\Big)\,\mathbf{V}^{(\text{m2})}.$$

The other branch of the MatFusion layer computes the m2-to-m1 attention from another direction, allowing both representations to attend to each other.

These cross-attended embeddings are then passed into subsequent layers, including a residual connection layer and a feed-forward layer. We conduct preliminary experiments to explore the impact of the number of stacked MatFusion layers on model performance. The results show that stack 4 MatFusion layers and a following linear achieve the best performance in our task. Based on this observation, we adapt this configuration to obtain the cross-attended outputs and then concatenate them as the input to a linear layer to get the final fused embeddings. The fused embeddings can be simply fed into a MLP for downstream tasks.

4 EXPERIMENTAL SETUP

Dataset Curation We constructed two datasets: `large` (24k samples) and `small` (4.4k samples, subset of `large`). Both datasets consist of experimental data from `matbench_expt_gap` and computational data from `matbench_mp_gap` (Jain et al., 2013) (see Table 4 for details). Each experimental data sample includes a materials composition and its experimental bandgap. Each computational data sample contains a materials composition, its structure (CIF) and bandgap calculated using Perdew–Burke–Ernzerhof (PBE) (Perdew et al., 1996) DFT. When curating dataset, we retain experimental data and remove duplicate entries from the computational data if necessary. The key distinction between these two datasets is that each experimental data in the `small` dataset is associated with a set of context sentences extracted from materials science literature containing the corresponding composition.

Other Setups. The baselines are the prediction performance with uni-modal embeddings. Structural embeddings are derived from CrystalFormer, while text embeddings are from DARWIN. We also test text embeddings from various LLMs including BERT (Devlin, 2018), T5 (Ni et al., 2021), LLaMA2 (Touvron et al., 2023), and MatBERT (Wan et al., 2024) (see details in Figure 2). The MLP used for bandgap prediction has some hyper-parameters: learning rate (η), batch size (B), dropout rate (δ), and layer configurations (L). We use AdamW optimizer with decoupled weight decay regularization (λ). Early stopping is applied. We use **mean absolute error (MAE)** as the metric and also report **standard deviation (STD)** from 5 runs (see subsection A.2).

5 RESULTS

Baseline and Fusion methods. Table 1 shows the uni-modal (fine-tuned MatBERT and CrystalFormer) and MatFusion performance. On `small`, the formula-based text embeddings achieve an MAE of 0.8250 for experimental data, outperforming the context-based approach, which reaches 0.8733. A simpler fusion baseline using direct concatenation of uni-modal embeddings yields suboptimal results, achieving an MAE of 0.5799 (see Appendix Table 5), suggesting the need for a

more expressive fusion mechanism. MatFusion further improves prediction accuracy, consistently surpassing the baseline models, with the best performance at 0.5558. For computational data, the formula-based approach performs slightly worse than the context-based method, but the overall trends align with those observed for experimental materials. As the training set scales up, the formula-based fusion approach achieves the highest accuracy for experimental data, with an MAE of 0.3601. Similarly, MatFusion significantly outperforms baseline text embeddings for computational data, achieving the best MAE of 0.2780, though a slight gap remains compared to using graph embeddings. These results highlight the critical role of structural information in predicting material properties.

Table 1: Uni-modal and MatFusion comparison on experimental and computational data.

Dataset	Embedding Type	MAE_{exp}	STD_{exp}	MAE_{comp}	STD_{comp}
Small	Text@Context	0.8733	0.0869	0.7657	0.0071
	Text@Formula	0.8250	0.0626	0.7943	0.0185
	Fusion@Context	0.5745	0.0405	0.3273	0.0065
	Fusion@Formula	**0.5558**	0.0542	0.3297	0.0214
	Graph@Structure	0.6348	0.0689	**0.3155**	0.0105
Large	Text@Formula	0.5658	0.0536	0.6742	0.0193
	Fusion@Formula	**0.3601**	0.0312	0.2780	0.0055
	Graph@Structure	0.3958	0.0279	**0.2776**	0.0031

Model Comparison. As illustrated in Table 2, we also test text embeddings derived from other LLMs. The results show that the embeddings generated by autoregressive models, such as the LLaMA series, consistently outperform non-autoregressive models like BERT. Furthermore, models fine-tuned on materials science literature exhibit enhanced predictive power compared to their general counterparts. Using text embeddings extracted by DARWIN, MatFusion achieves the highest performance (0.3596) in predicting experimental materials.

Table 2: Text-modal and fusion performance across different text models.

Model	Embedding Type	MAE_{exp}	STD_{exp}	MAE_{comp}	STD_{comp}
BERT	Text@Formula	0.6201	0.0266	0.6978	0.0288
	Fusion@Formula	0.3720	0.0251	0.2751	0.0037
T5	Text@Formula	0.5585	0.0663	0.6353	0.0276
	Fusion@Formula	0.3803	0.0278	**0.2698**	0.0018
LLaMA2	Text@Formula	0.4844	0.0360	0.5514	0.0062
	Fusion@Formula	0.3782	0.0195	0.2781	0.0036
DARWIN	Text@Formula	0.4714	0.0298	0.5409	0.0131
	Fusion@Formula	**0.3596**	0.0170	0.2857	0.0062

Dataset Scaling. We incrementally scale up the training set of one data modality from `small` to `large` while fixing the other modality. Table 3 shows that such scaling up enhances the prediction performance on the scaled modality but deteriorates the performance on the other one.

Table 3: Scaling up the training set of one data modality but fixing the other.

Embedding Type	Scaled Modality	MAE_{exp} Trend	MAE_{comp} Trend
Text@Formula	Exp.	$0.8341 \rightarrow 0.6687$	$0.7846 \rightarrow 0.8058$
	Comp.	$0.8112 \rightarrow 0.7864$	$0.7538 \rightarrow 0.6813$
Graph@Structure	Exp.	$0.5957 \rightarrow 0.4085$	$0.3118 \rightarrow 0.3265$
	Comp.	$0.6064 \rightarrow 0.6267$	$0.2989 \rightarrow 0.2727$
Fusion@Formula	Exp.	$0.5396 \rightarrow 0.3850$	$0.3418 \rightarrow 0.3525$
	Comp.	$0.5916 \rightarrow 0.6043$	$0.3102 \rightarrow 0.2771$

6 Conclusion

We propose **MatFusion**, a multi-modal framework that bridges LLMs and structural embeddings for experimental materials property prediction. Through extensive evaluations on bandgap prediction tasks, we demonstrate that our framework significantly improves prediction accuracy by 23.72% compared to standalone text embeddings and 9.15% compared to structural embeddings. The improvement is more pronounced for experimental bandgap prediction, likely because text-based representations encapsulate empirical knowledge and experimental intuition, which may aid

experimental property prediction but be less relevant—or even disruptive—for computational ones governed by physical relationships. Additionally, context information does not enhance performance as much as expected compared to using only material composition, possibly due to noise, irrelevant details, or data imbalance from limited context sentences for some materials. For future work, we plan to explore standardized GPT-generated descriptions for more uniform and informative contextual representation, use more advanced structure matching algorithms, and extend to predict other material properties or support different downstream tasks.

REFERENCES

Simon Batzner, Albert Musaelian, Lixin Sun, Mario Geiger, Jonathan P Mailoa, Mordechai Kornbluth, Nicola Molinari, Tess E Smidt, and Boris Kozinsky. E (3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nature communications*, 13(1): 2453, 2022.

Parishad BehnamGhader, Vaibhav Adlakha, Marius Mosbach, Dzmitry Bahdanau, Nicolas Chapados, and Siva Reddy. Llm2vec: Large language models are secretly powerful text encoders. *arXiv preprint arXiv:2404.05961*, 2024.

Md Maruf Billah, Mohamed Elleithy, Waris Khan, Saltuk Yıldız, Zekeriya Ender Eğer, Sheng Liu, Matthew Long, and Pınar Acar. Uncertainty quantification of microstructures: a perspective on forward and inverse problems for mechanical properties of aerospace materials. *Advanced Engineering Materials*, 27(2):2401299, 2025.

Chi Chen and Shyue Ping Ong. A universal graph deep learning interatomic potential for the periodic table. *Nature Computational Science*, 2(11):718–728, 2022.

Chi Chen, Weike Ye, Yunxing Zuo, Chen Zheng, and Shyue Ping Ong. Graph networks as a universal machine learning framework for molecules and crystals. *Chemistry of Materials*, 31(9):3564–3572, 2019.

Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *International conference on machine learning*, pp. 1597–1607. PMLR, 2020.

Kamal Choudhary, Kevin F Garrity, Andrew CE Reid, Brian DeCost, Adam J Biacchi, Angela R Hight Walker, Zachary Trautt, Jason Hattrick-Simpers, A Gilad Kusne, Andrea Centrone, et al. The joint automated repository for various integrated simulations (jarvis) for data-driven materials design. *npj computational materials*, 6(1):173, 2020.

Jacob Devlin. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.

Yuanqi Du, Yingheng Wang, Yining Huang, Jianan Canal Li, Yanqiao Zhu, Tian Xie, Chenru Duan, John Gregoire, and Carla P Gomes. m^2 hub: Unlocking the potential of machine learning for materials discovery. *Advances in Neural Information Processing Systems*, 36:77359–77378, 2023.

Alex M Ganose and Anubhav Jain. Robocrystallographer: automated crystal structure text descriptions and analysis. *MRS Communications*, 9(3):874–881, 2019.

Zhongyu Huang, Yingheng Wang, Chaozhuo Li, and Huiguang He. Going deeper into permutation-sensitive graph neural networks. In *International Conference on Machine Learning*, pp. 9377–9409. PMLR, 2022.

Anubhav Jain, Shyue Ping Ong, Geoffroy Hautier, Wei Chen, William Davidson Richards, Stephen Dacek, Shreyas Cholia, Dan Gunter, David Skinner, Gerbrand Ceder, et al. Commentary: The materials project: A materials genome approach to accelerating materials innovation. *APL materials*, 1(1), 2013.

Vadim Korolev and Pavel Protsenko. Accurate, interpretable predictions of materials properties within transformer language models. *Patterns*, 4(10), 2023.

Yi-Lun Liao and Tess Smidt. Equiformer: Equivariant graph attention transformer for 3d atomistic graphs. *arXiv preprint arXiv:2206.11990*, 2022.

Dongsheng Luo, Wei Cheng, Yingheng Wang, Dongkuan Xu, Jingchao Ni, Wenchao Yu, Xuchao Zhang, Yanchi Liu, Yuncong Chen, Haifeng Chen, et al. Time series contrastive learning with information-aware augmentations. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pp. 4534–4542, 2023.

Adrian Mirza, Sebastian Starke, Erinc Merdivan, and Kevin Maik Jablonka. Bridging chemical modalities by aligning embeddings. In *AI for Accelerated Materials Design-Vienna 2024*, 2024.

Jianmo Ni, Gustavo Hernandez Abrego, Noah Constant, Ji Ma, Keith B Hall, Daniel Cer, and Yinfei Yang. Sentence-t5: Scalable sentence encoders from pre-trained text-to-text models. *arXiv preprint arXiv:2108.08877*, 2021.

Janghoon Ock, Srivathsan Badrinarayanan, Rishikesh Magar, Akshay Antony, and Amir Barati Farimani. Multimodal language and graph learning of adsorption configuration in catalysis. *Nature Machine Intelligence*, pp. 1–11, 2024.

Hyunsoo Park, Anthony Onwuli, and Aron Walsh. Exploration of crystal chemical space using text-guided generative artificial intelligence. 2024.

John P Perdew, Kieron Burke, and Matthias Ernzerhof. Generalized gradient approximation made simple. *Physical review letters*, 77(18):3865, 1996.

Albert Polman, Mark Knight, Erik C Garnett, Bruno Ehrler, and Wim C Sinke. Photovoltaic materials: Present efficiencies and future challenges. *Science*, 352(6283):aad4424, 2016.

Jiaxing Qu, Yuxuan Richard Xie, Kamil M Ciesielski, Claire E Porter, Eric S Toberer, and Elif Ertekin. Leveraging language representation for materials exploration and discovery. *npj Computational Materials*, 10(1):58, 2024.

Branimir Radisavljevic, Aleksandra Radenovic, Jacopo Brivio, Valentina Giacometti, and Andras Kis. Single-layer mos2 transistors. *Nature nanotechnology*, 6(3):147–150, 2011.

Nils Reimers and Iryna Gurevych. Sentence-bert: Sentence embeddings using siamese bert-networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, 11 2019. URL https://arxiv.org/abs/1908.10084.

Andre Niyongabo Rubungo, Craig Arnold, Barry P Rand, and Adji Bousso Dieng. Llm-prop: Predicting physical and electronic properties of crystalline solids from their text descriptions. *arXiv preprint arXiv:2310.14029*, 2023.

E Fred Schubert and Jong Kyu Kim. Solid-state light sources getting smart. *Science*, 308(5726): 1274–1278, 2005.

Kristof Schütt, Oliver Unke, and Michael Gastegger. Equivariant message passing for the prediction of tensorial properties and molecular spectra. In *International Conference on Machine Learning*, pp. 9377–9388. PMLR, 2021.

Yu Song, Santiago Miret, and Bang Liu. Matsci-nlp: Evaluating scientific language models on materials science language tasks using text-to-schema modeling. *arXiv preprint arXiv:2305.08264*, 2023.

Hugo Touvron, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov, Soumya Batra, Prajjwal Bhargava, Shruti Bhosale, et al. Llama 2: Open foundation and fine-tuned chat models. *arXiv preprint arXiv:2307.09288*, 2023.

Vahe Tshitoyan, John Dagdelen, Leigh Weston, Alexander Dunn, Ziqin Rong, Olga Kononova, Kristin A Persson, Gerbrand Ceder, and Anubhav Jain. Unsupervised word embeddings capture latent knowledge from materials science literature. *Nature*, 571(7763):95–98, 2019.

Nicholas Walker, Amalie Trewartha, Haoyan Huo, Sanghoon Lee, Kevin Cruse, John Dagdelen, Alexander Dunn, Kristin Persson, Gerbrand Ceder, and Anubhav Jain. The impact of domain-specific pre-training on named entity recognition tasks in materials science. *Available at SSRN 3950755*, 2021.

Yuwei Wan, Tong Xie, Nan Wu, Wenjie Zhang, Chunyu Kit, and Bram Hoex. From tokens to materials: Leveraging language models for scientific discovery. *arXiv preprint arXiv:2410.16165*, 2024.

Yingheng Wang, Yaosen Min, Xin Chen, and Ji Wu. Multi-view graph contrastive representation learning for drug-drug interaction prediction. In *Proceedings of the web conference 2021*, pp. 2921–2933, 2021a.

Yingheng Wang, Yaosen Min, Erzhuo Shao, and Ji Wu. Molecular graph contrastive learning with parameterized explainable augmentations. In *2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 1558–1563. IEEE, 2021b.

Yingheng Wang, Shufeng Kong, John M Gregoire, and Carla P Gomes. Conformal crystal graph transformer with robust encoding of periodic invariance. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pp. 283–291, 2024.

Tian Xie and Jeffrey C Grossman. Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties. *Physical review letters*, 120(14):145301, 2018.

Tong Xie, Yuwei Wan, Yixuan Liu, Yuchen Zeng, Wenjie Zhang, Chunyu Kit, Dongzhan Zhou, and Bram Hoex. Darwin 1.5: Large language models as materials science adapted learners. *arXiv preprint arXiv:2412.11970*, 2024.

Keqiang Yan, Yi Liu, Yuchao Lin, and Shuiwang Ji. Periodic graph transformers for crystal material property prediction. *Advances in Neural Information Processing Systems*, 35:15066–15080, 2022.

Yuning You, Tianlong Chen, Yongduo Sui, Ting Chen, Zhangyang Wang, and Yang Shen. Graph contrastive learning with augmentations. *Advances in neural information processing systems*, 33: 5812–5823, 2020.

Ya Zhuo, Aria Mansouri Tehrani, and Jakoah Brgoch. Predicting the band gaps of inorganic solids by machine learning. *The journal of physical chemistry letters*, 9(7):1668–1673, 2018.