Transfer Learning in Materials Informatics: structure-property relationships through minimal but highly informative multimodal input

Dario Massa
IDEAS-NCBR Chmielna 69 00-801 Warsaw, Poland
NOMATEN Center of Excellence in Multifunctional Materials,
ul. Andrzeja Sołtana, 05-400 Otwock, Świerk, Poland
dario.massa@ncbj.gov.pl

Grzegorz Kaszuba
IDEAS-NCBR Chmielna 69 00-801 Warsaw, Poland
Poznań University of Technology, Poland

Stefanos Papanikolaou
NOMATEN Center of Excellence in Multifunctional Materials,
ul. Andrzeja Sołtana, 05-400 Otwock, Świerk, Poland

Piotr Sankowski
IDEAS-NCBR Chmielna 69 00-801 Warsaw, Poland
MIM Solutions, Poland
piotr.sankowski@mim.ai

Abstract

In this work we propose simple, effective and computationally efficient transfer learning approaches for structure-property relation predictions in the context of materials, with highly informative input from different modalities. As materials properties stand from their electronic structure, representations are extracted directly from datasets of electronic charge density profile images using Neural Networks. We demonstrate transferability of the existing pre-trained Convolutional Neural Networks and Large Language Models knowledge to physics domain data, exploring a wide set of compositions for the regression of energetics- or structure-related properties, and the role of semantic crystallographic information in the context of multimodal approaches. We test the applicability of the CLIP multimodal model, and employ as well a training protocol for building a more interpretable and versatile stacked custom solution from different pre-trained modalities. The study offers a promising avenue for enhancing the effectiveness of descriptor identification in physical systems, shedding light on the power of multimodal transfer learning for materials property prediction. Instead of using the well-established GNN-based approaches, we explore the transfer learning of image- and text-based architectures, which can impact decision making for new low-cost AI methods in the field of Materials and Chemoinformatics.

1 Introduction

Machine Learning (ML) [1, 2] has initiated a remarkable transformation in Materials Science, with the emergence of Materials Informatics (MI) playing a pivotal role in pushing the boundaries of innovation in the context of multiple technological research challenges [3, 4, 5, 6, 7, 8, 9, 10, 11]. A key aspect is the identification of efficient descriptors, capable of capturing the essential features of complex physical systems. In the realm of materials simulations, a wide variety of descriptors have been proposed, both based on translationally and rotationally invariant functions of atomic coordinates [12, 13, 14, 15, 16], or involving fixed-length feature vectors of atomic or electronic properties [17, 18, 19], as well as more advanced feature extraction approaches through Graph Convolutional Neural Networks (GCNNs) [20, 21, 22, 23, 24, 25, 26]. Wide open-source databases [27, 28, 29, 30], originating from high-throughput [31, 32] ab-initio Density-Functional-Theory (DFT) [33, 34] calculations, represent the data source for the development of such models and MI in general.

In real-life research settings, it is very common to come across low-quantity but highly informative data about the systems of interest: in the field of Molecular Science, an example can be found in molecular orbitals, encoding reactivity and optical properties; in the field of Materials Science, for example, phonon dispersion relations encode information on thermal conductivity. Recent works have demonstrated the power of deploying Electron Charge Density (ECD) profiles in ML for unsupervised studies of defects-related effects in metals [35], or prediction of density fields in unseen samples via graph-based models [36]. The interest in such quantity lies in its encoding of the necessary information regarding the properties of the system – in terms of constitutive atoms and their interactions - regardless of dimensionality, stability or geometry. In this view, the ECD profiles represent an example of highly informative data, which could be efficiently exploited for structure-property relations also with the simplest available solutions.

In this work, we aim to propose to utilize the computational efficiency of multimodal transfer learning solutions to predict materials properties from ECD profiles [37]. We build an image and text dataset based on the Materials Project (MP) [28], which could stand as a representative example of a relatively small-sized but very diverse problem to be learned. We successfully test the transferability of the known multimodal Contrastive Language-Image Pre-training (CLIP) model from OpenAI [38]. Furthemore, we employ a protocol for building a custom, more interpretable and versatile stacked multimodal solution starting from single pre-trained modalities. The proposed setting stands as an impactful solution involving minimal information and computational effort which could lay the basis for novel low-cost AI models for inverse materials design and databases in the field of Materials and Cheminformatics.

2 Methods

2.1 Multimodal dataset

We build our multimodal dataset as a subset of the Materials Project [28], restricting it to stable face-centered cubic (FCC) crystals. Their MP-IDs, the associated target properties and ECD fields are collected, resulting in 781 samples. Data-cleaning is performed to filter out outliers in the distribution of maxima and minima of the collected ECD fields, through definition of acceptance boundaries with the interquartile range (IQR), resulting in 592 samples. As shown in Fig.(3) of the Appendix, the majority of the collected samples is represented by ternary crystals, and their quantum-level characterization is very demanding from a computational point of view. We use the dedicated Py-Rho [39] to unify the grid dimensions to $60 \times 60 \times 60$. To further simplify the problem without losing the descriptive power of the data, we slice the three-dimensional data along specific bi-dimensional plane, characterized by the highest atomic density. Having restricted ourselves to FCC crystals, such plane corresponds to the $\langle 111 \rangle$ Miller indices. As shown in Fig.(1), we perform the slicing in a supercell made by an $8 \times 8 \times 8$ repetition of the unit-cell field data, and focus the extraction of a squared image from such slice. Even though both total and differential ECDs are available, we restrict the dataset on the latter type for the enhanced variability within and among samples originating from the highlighting of interatomic interactions, rather than local atomic electron densities. A comparison between the slices along the z-axis of the total and differential ECD fields is presented in Appendix in Fig.(4). For regression experiments, we target energetics, with the formation (E_F) and Fermi energies (E_{fermi}), as well as a structural property, the bulk modulus (K_{vrh}). Before application of the models,

mp-865869, Be2PtCu is Heusler structured and crystallizes in the cubic Fm-3m space group.
Be(1) is bonded in a body-centered cubic geometry to four equivalent Pt(1) and four equivalent Cu(1) atoms.
All Be(1)-Pt(1) bond lengths are 2.39 Å. All Be(1)-Cu(1) bond lengths are 2.39 Å.
Pt(1) is bonded in a body-centered cubic geometry to eight equivalent Be(1) atoms.
Cu(1) is bonded in a distorted body-centered cubic geometry to eight equivalent Be(1) atoms.
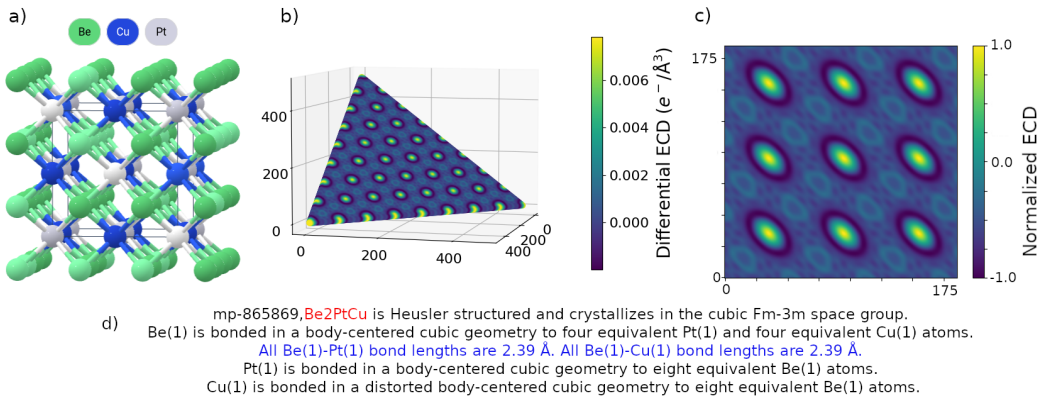
Figure 1: Visualization of the creation steps of the multimodal dataset. a) A bulk crystal is considered together with its electron charge density field, its target properties and its structure file (.cif). b) The differential electron charge density field is sliced along the maximum atomic density plane from an 8x8x8 repetition of the sample. c) A squared focus from the slice is extracted, and properly normalized. d) The crystallographic textual information about the sample is produced. During experiments, we also consider fractions of text as: 'Formula' (only the composition name), 'Keywords' (the composition name and the bond lengths information), and 'All' (the entire text).

the target properties undergo standardization, and the ECD isoline values a z-score normalization and rescaling in the $[-1; 1]$ range: this allows to preserve peculiar patterns of zero-level oscillations across electron accumulation and depletion isolines. Finally, with the help of the Robocrystallographer [40], a tool capable of generating concise text descriptions of crystalline structures by analyzing their symmetry, local environment, and extended connectivity from their Pymatgen structures, we build a set of text attributes associated to each MP-ID, an example of which is reported in Fig. (1).

2.2 Multimodal models

We use two foundational models to incorporate the textual and image information about the crystal structures, resulting in a stacked custom multimodal model. The schematic representation of the architecture is reported in Appendix in Fig.(5) To process the images, we use InceptionV3, a model pre-trained on ImageNet. For text, we use Roberta, trained on language modeling task. We use the same procedure to adapt either of the models. First, we remove model's classification head and replace it with a readout network composed of 4 dense layers, whose width gradually decreases from pre-trained model's latent dimension to 1 output channel. We first fit the new dense readout, to then unfreeze the rest of the model. After training the image- and text-based property regressors, we load them both to create a multimodal model. We remove the final layer of both readout dense networks. We concatenate the outputs of penultimate layers of our models, each having 256 channels, and create two new readout layers. Once again, we first fit the newly created layers to then unfreeze the remainder of the model. All dense readout layers, apart from the ultimate single-channel layer, are followed by ReLU activation, BatchNorm and dropout. To contrast this approach, we also conduct experiments with CLIP, a transformer-based model pre-trained on a text-image multimodal task. We add a single linear layer on top of the pre-trained model to get the regression output.

3 Experiments

First, we verify the transferability of the CLIP multimodal model to our physics domain problem. To do so, we use 85% of the dataset for cross-validation experiments with five folds, and keep the remaining 15% for further testing. Preliminary experiments have been conducted to optimize the hyperparameters of the model, and are reported in the Appendix Table(2). In Table(1), we report the performances of the CLIP model averaged over various trials. The higher and more stable performances are found for the regression of the bulk modulus, but overall we observe promising generalization capabilities in terms of R^2, on average around 0.8. We report examples of learning curves over different folds in the Appendix Fig.(6). Additional experiments have been conducted by

Table 1: Statistical results for regression of the target properties with the CLIP multimodal model showing the mean R^2 across validation folds, as well as the test set R^2 for each run, along with the average values. CLIP used the full textual description available ('All').

		E_F	E_{fermi}	K_{vrh}
		R^2/RMSE[eV/atom]	R^2/RMSE[eV]	R^2/RMSE[GPa]
#1	Val. Folds Average	0.91/0.21	0.90/0.89	0.91/21.9
	Test	0.72/0.36	0.87/0.90	0.88/24.4
#2	Val. Folds Average	0.88/0.23	0.90/0.88	0.88/24.1
	Test	0.89/0.24	0.82/1.25	0.80/35.3
#3	Val. Folds Average	0.77/0.35	0.90/0.86	0.92/21.1
	Test	0.67/0.35	0.51/2.07	0.80/36.5
$\langle\#\rangle$	**Val. Folds Average**	0.85/0.26	0.9/0.86	0.90/22.4
	Test set	0.76/0.31	0.73/1.41	0.83/32.1

taking the Fourier Transform of ECD images to test the importance of periodicity. The Appendix Fig(9) and Table(3) respectively report examples of such data and the related performances of the model, which however did not underline specific improvements.

Fig.(2) reports the regression performance of the stacked custom multimodal model during the different stages of its implementation. Already at the stage of InceptionV3 (image-only) finetuning, the bulk modulus prediction reaches high R^2 values in generalization, comparable to the CLIP model, and the subsequent finetuning of the RoBERTa (text-only) model or the composition of the semantic and visual embeddings from the finetuned models do not seem to provide visibile improvements. The prediction of the Fermi energy based on ECD images only, instead, does not generalize well, and remarkably benefits from the use of textual crystallographic information at all its levels (Formula, Keywords and All). The formation energy is a challenging prediction target, though it shows slight improvement as text input is added.

4   Discussion, Limitations and Future Work

In this work we propose computationally efficient approaches to property prediction in materials through the use of multiple modalities in transfer learning, with ECDs as highly informative input data. In particular, we show the dual approach of a ready-to-use pre-trained multimodal model like CLIP and the creation of a custom stack of single-modality pre-trained models. Even though the performances of the two approaches are similar, the latter stands as i) a more interpretable solution, because the single-modality stages convey the importance of each model and data type in the final combined performance; ii) more versatile solution, in it being composed of arbitrary single-modality models, which might be chosen appropriately for the planned application. We underline the highly informative nature of the data: the image of a single slice out of a three-dimensional field could be enough to build expressive features for the prediction of computationally expensive material properties properties like the bulk modulus, which generally requires calculating the total energy of the system under different volumes and then fitting to an equation of state. Furthermore, we observe remarkable regression improvements through inclusion of semantic embeddings, as opposed to only using visual information. It is an encouraging sign of the possibilities that multimodal approaches offer in such context.

The work presents some limitations and potential for future work. As we run independent trials for different variants of textual descriptions, we notice that multiple instances of image-only finetuning show palpable variance in performance, even though they are not influenced by varying text input (Fig. (2, Image-FT). This points at the need for multi-run experiments to quantify the variance in performance. An exciting avenue already under consideration could be the implementation of multimodal feature estimation in an inverse design procedure, which will be able to employ ECD fields and basic crystallographic information in a search for desired set of properties. Such approach would represent a remarkable leap in the field, connecting low-cost approaches, generative models and ECD prediction.

a) Formation energy: R^2/RMSE[eV/atom]

	Train	Val	Test
Formula Image	0.71/0.42	0.62/0.36	0.72/0.17
Formula Text	0.88/0.27	0.65/0.35	0.64/0.29
Formula Image+Text	0.95/0.17	0.59/0.38	0.68/0.27
Keywords Image	0.61/0.49	0.55/0.40	0.49/0.35
Keywords Text	0.94/0.19	0.75/0.30	0.84/0.19
Keywords Image+Text	0.94/0.19	0.68/0.33	0.78/0.23
All Image	0.61/0.49	0.53/0.40	0.65/0.29
All Text	0.92/0.22	0.71/0.32	0.60/0.31
All Image+Text	0.95/0.17	0.70/0.32	0.73/0.25

b) Fermi energy: R^2/RMSE[eV]

	Train	Val	Test
Formula Image	0.77/1.47	0.63/1.76	0.60/1.56
Formula Text	0.91/0.92	0.71/1.55	0.86/0.92
Formula Image+Text	0.99/0.31	0.78/1.35	0.88/0.86
Keywords Image	0.82/1.30	0.72/1.53	0.63/1.50
Keywords Text	0.92/0.87	0.84/1.16	0.86/0.92
Keywords Image+Text	0.97/0.53	0.84/1.16	0.87/0.89
All Image	0.80/1.37	0.64/1.73	0.56/1.64
All Text	0.94/0.75	0.77/1.38	0.87/0.89
All Image+Text	0.97/0.53	0.80/1.29	0.84/0.99

c) KVRH: R^2/RMSE[GPa]

	Train	Val	Test
Formula Image	0.77/36.61	0.84/31.04	0.84/28.09
Formula Text	0.90/24.14	0.81/33.82	0.82/29.79
Formula Image+Text	0.97/13.22	0.87/27.98	0.87/25.32
Keywords Image	0.80/34.14	0.85/30.05	0.84/28.09
Keywords Text	0.94/18.70	0.76/38.01	0.83/28.95
Keywords Image+Text	0.95/17.07	0.81/33.82	0.81/30.61
All Image	0.79/34.98	0.83/31.99	0.79/32.18
All Text	0.80/34.14	0.75/38.80	0.75/35.11
All Image+Text	0.92/21.59	0.75/38.80	0.71/37.81

d) Average across labels: R^2

	Train	Val	Test
Formula Image	0.75	0.70	0.72
Formula Text	0.90	0.72	0.77
Formula Image+Text	0.97	0.75	0.81
Keywords Image	0.74	0.71	0.65
Keywords Text	0.93	0.78	0.84
Keywords Image+Text	0.95	0.78	0.82
All Image	0.73	0.67	0.67
All Text	0.89	0.74	0.74
All Image+Text	0.95	0.75	0.76

Figure 2: Regression performance of the different stages of a custom stacked multimodal model towards the three target properties: the formation energy, E_F, the Fermi energy, E_{fermi}, and the bulk modulus, K_{vrh}. The 'Image' only stage represents the fine-tuning (FT) of the InceptionV3 model. The 'Text' only stage represents the fine-tuning of the RoBERTa-Base model. The 'Image+Text' stage represents the fine-tuning of the embedding concatenation from the two modalities. Each of the stages involving text, also report the extent of textual information used, being it only the 'Formula', the 'Keywords' (composition name and bond lenghts information) or 'All' available text.

Acknowledgments and Disclosure of Funding

References

[1] Ryszard Stanislaw Michalski, Jaime Guillermo Carbonell, and Tom M Mitchell. *Machine learning: An artificial intelligence approach*. Springer Science & Business Media, 2013.

[2] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553):436–444, 2015.

[3] Rampi Ramprasad, Rohit Batra, Ghanshyam Pilania, Arun Mannodi-Kanakkithodi, and Chiho Kim. Machine learning in materials informatics: recent applications and prospects. *npj Computational Materials*, 3(1):54, 2017.

[4] Logan Ward and Chris Wolverton. Atomistic calculations and materials informatics: A review. *Current Opinion in Solid State and Materials Science*, 21(3):167–176, 2017.

[5] Krishna Rajan. Materials informatics. *Materials Today*, 8(10):38–45, 2005.

[6] Li-Chiang Lin, Adam H Berger, Richard L Martin, Jihan Kim, Joseph A Swisher, Kuldeep Jariwala, Chris H Rycroft, Abhoyjit S Bhown, Michael W Deem, Maciej Haranczyk, et al. In silico screening of carbon-capture materials. *Nature materials*, 11(7):633–641, 2012.

[7] Christopher E Wilmer, Michael Leaf, Chang Yeon Lee, Omar K Farha, Brad G Hauser, Joseph T Hupp, and Randall Q Snurr. Large-scale screening of hypothetical metal–organic frameworks. *Nature chemistry*, 4(2):83–89, 2012.

[8] Grégoire Montavon, Matthias Rupp, Vivekanand Gobre, Alvaro Vazquez-Mayagoitia, Katja Hansen, Alexandre Tkatchenko, Klaus-Robert Müller, and O Anatole Von Lilienfeld. Machine learning of molecular electronic properties in chemical compound space. *New Journal of Physics*, 15(9):095003, 2013.

[9] Matthias Rupp, Alexandre Tkatchenko, Klaus-Robert Müller, and O Anatole Von Lilienfeld. Fast and accurate modeling of molecular atomization energies with machine learning. *Physical review letters*, 108(5):058301, 2012.

[10] Wesley T Hong, Roy E Welsch, and Yang Shao-Horn. Descriptors of oxygen-evolution activity for oxides: a statistical evaluation. *The Journal of Physical Chemistry C*, 120(1):78–86, 2016.

[11] Jeff Greeley, Thomas F Jaramillo, Jacob Bonde, IB Chorkendorff, and Jens K Nørskov. Computational high-throughput screening of electrocatalytic materials for hydrogen evolution. *Nature materials*, 5(11):909–913, 2006.

[12] Matthias Rupp, Alexandre Tkatchenko, Klaus-Robert Müller, and O Anatole Von Lilienfeld. Fast and accurate modeling of molecular atomization energies with machine learning. *Physical review letters*, 108(5):058301, 2012.

[13] Jörg Behler. Atom-centered symmetry functions for constructing high-dimensional neural network potentials. *The Journal of chemical physics*, 134(7), 2011.

[14] Fabio Pietrucci and Wanda Andreoni. Graph theory meets ab initio molecular dynamics: atomic structures and transformations at the nanoscale. *Physical review letters*, 107(8):085504, 2011.

[15] Sandip De, Albert P Bartók, Gábor Csányi, and Michele Ceriotti. Comparing molecules and solids across structural and alchemical space. *Physical Chemistry Chemical Physics*, 18(20):13754–13769, 2016.

[16] Ali Sadeghi, S Alireza Ghasemi, Bastian Schaefer, Stephan Mohr, Markus A Lill, and Stefan Goedecker. Metrics for measuring distances in configuration spaces. *The Journal of chemical physics*, 139(18), 2013.

[17] Atsuto Seko, Atsushi Togo, Hiroyuki Hayashi, Koji Tsuda, Laurent Chaput, and Isao Tanaka. Prediction of low-thermal-conductivity compounds with first-principles anharmonic lattice-dynamics calculations and bayesian optimization. *Physical review letters*, 115(20):205901, 2015.

[18] D Xue, PV Balachandran, J Hogden, J Theiler, D Xue, and T Lookman. Accelerated search for materials with targeted properties by adaptive design. *Nat Commun*, 7:11241, 2016.

[19] Olexandr Isayev, Corey Oses, Cormac Toher, Eric Gossett, Stefano Curtarolo, and Alexander Tropsha. Universal fragment descriptors for predicting properties of inorganic crystals. *Nature communications*, 8(1):15679, 2017.

[20] T. Xie and J. C. Grossman. Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties. *Phys. Rev. Lett.*, 120:145301, 2018.

[21] C. Chen et al. Graph net- works as a universal machine learning framework for molecules and crystals. *Chem. Mater.*, 31(9):3564, 2019.

[22] S. Back et al. Convolutional neural network of atomic surface structures to predict binding energies for high-throughput screening of catalysts. *J. Phys. Chem. Lett.*, 10:4401–4408, 2019.

[23] J. Jorgensen et al. Neural message passing with edge updates for predicting properties of molecules and materials. *arXiv.org e-Print archive*, page arXiv:1806.01261, 2018.

[24] K. Schütt et al. Schnet: a continuous-filter convolutional neural network for modeling quantum interactions. *NeurIPS*, 30:991–1001, 2017.

[25] D. K. Duvenaud et al. Convolutional networks on graphs for learning molecular fingerprints. *NeurIPS*, page 2224–2232, 2015.

[26] Z. Wu et al. Moleculenet: a benchmark for molecular machine learning. *Chem. Sci.*, 9:513–530, 2018.

[27] E. Saal et al. Materials design and discovery with high-throughput density functional theory: The open quantum materials database (oqmd). *JOM*, 65(1501), 2013.

[28] A. Jain et al. The materials project: A materials genome approach to accelerating materials innovation. *APL Mater*, 1(011002), 2013.

[29] S. Curtarolo. A distributed materials properties repository from high-throughput ab initio calculations. *Comput. Mater. Sci.*, 58(227), 2012.

[30] J. Hachmann et al. The harvard clean energy project: Large-scale computational screening and design of organic photovoltaics on the world community grid. *J. Phys. Chem. Lett.*, 2(2241), 2011.

[31] Stefano Curtarolo, Gus LW Hart, Marco Buongiorno Nardelli, Natalio Mingo, Stefano Sanvito, and Ohad Levy. The high-throughput highway to computational materials design. *Nature materials*, 12(3):191–201, 2013.

[32] James E Saal, Scott Kirklin, Muratahan Aykol, Bryce Meredig, and Christopher Wolverton. Materials design and discovery with high-throughput density functional theory: the open quantum materials database (oqmd). *Jom*, 65:1501–1509, 2013.

[33] Walter Kohn, Axel D Becke, and Robert G Parr. Density functional theory of electronic structure. *The journal of physical chemistry*, 100(31):12974–12980, 1996.

[34] Aron J Cohen, Paula Mori-Sánchez, and Weitao Yang. Challenges for density functional theory. *Chemical reviews*, 112(1):289–320, 2012.

[35] Dario Massa, Efthimios Kaxiras, and Stefanos Papanikolaou. Alloy informatics through ab initio charge density profiles: Case study of hydrogen effects in face-centred cubic crystals. *Acta Materialia*, 268:119773, 2024.

[36] Teddy Koker, Keegan Quigley, Eric Taw, Kevin Tibbetts, and Lin Li. Higher-order equivariant neural networks for charge density prediction in materials. *npj Computational Materials*, 10(1):161, 2024.

[37] Lisa Torrey and Jude Shavlik. Transfer learning. In *Handbook of research on machine learning applications and trends: algorithms, methods, and techniques*, pages 242–264. IGI global, 2010.

[38] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. Learning transferable visual models from natural language supervision. *CoRR*, abs/2103.00020, 2021.

[39] Jimmy-Xuan Shen, Jason M. Munro, Matthew K. Horton, Patrick Huck, Shyam Dwaraknath, and Kristin A. Persson. A representation-independent electronic charge density database for crystalline materials, 2021.

[40] A. M. Ganose and A. Jain. Robocrystallographer: automated crystal structure text descriptions and analysis. *MRS Communications*, 9, 2019.