Residual Neural Network Architectures to Improve Prediction Accuracy of Properties of Materials

David Tatis
Laboratory for Applied Remote-Sensing, Imaging and Photonics.
University of Puerto Rico Mayaguez

Heidy Sierra
Laboratory for Applied Remote-Sensing, Imaging and Photonics.
University of Puerto Rico Mayaguez

Emmanuel Arzuaga
Laboratory for Applied Remote-Sensing, Imaging and Photonics.
University of Puerto Rico Mayaguez

***Abstract*—Properties in material composition and crystal structures have been explored by density functional theory (DFT) calculations, using databases such as the Open Quantum Materials Database (OQMD). Databases like these have been used currently for the training of advanced machine learning and deep neural network models, the latter providing higher performance when predicting properties of materials. However, current alternatives have shown a deterioration in accuracy when increasing the number of layers in their architecture (over-fitting problem). As an alternative method to address this problem, we have implemented residual neural network architectures based on Merge and Run Networks, IRNet and UNet to improve performance while relaxing the observed network depth limitation. The evaluation of the proposed architectures include a 9:1 ratio to train and test as well as 10 fold cross validation. In the experiments we found that our proposed architectures based on IRNet and UNet are able to obtain a lower Mean Absolute Error (MAE) than current strategies. The full implementation (Python, Tensorflow and Keras) and the trained networks will be available online for community validation and advancing the state of the art from our findings.**

***Index Terms*—material properties, elemental compositions, residual neural networks.**

I. Introduction

The discovery of new materials originally consisted of trial and error strategies, since there was no precise methodology of knowing the physical or chemical characteristics that these would have when they were created [1]. Computational advances during the past decades have allowed researchers to systematically predict properties of chemical compounds by the development of mathematical methods that use collected data from prior experimentation in materials science or molecular chemistry. Among the methods used for the prediction of material properties, Density Functional Theory (DFT) [2], which is based mainly on The Hohenberg-Kohn theorem [3], establishes that the ground state of an electronic system is a functional of the electronic density. This means that it is only necessary to know the density to calculate all the other properties of the system [4]. DFT allows to predict properties of crystalline solids such as formation enthalpy which can then be used to determine if a compound is stable, and therefore to improve the ability to discover new materials.

For many years the DFT method could not be practically applied, due to its computational complexity and the available computing hardware that limited its effective implementations. However, with the increase of computational power, it is now possible to use the DFT method at a large scale in the order of hundreds of thousands of structures in a reasonable time frame. As a result, databases such as the Open Quantum Materials Database (OQMD) [5], contains collections of consistently calculated DFT total energies and relaxed crystal structures are available. The OQMD database includes DFT calculated thermodynamic and structural properties of 637,644 materials, and those numbers are increasing with time. Furthermore, the OQMD has open access, allowing researchers from different fields to use it and develop new algorithms and computational tools that advance new materials development [6]. For example, research work conducted on the development of AI methods based on Machine Learning (ML) and Artificial Neural Networks (ANN) to accelerate the discovery of new materials using the OQMD is emerging and positively impacting this area [7]–[9].

ML algorithms such as Linear Regression, SGD Regression, Decision Trees, Extra Trees, Random Forest, among others have been used for property prediction of materials properties by using the OQMD database where the output of the model depends on the property to be predicted [7] [8]. An example, is the prediction of the formation of enthalpy, which is an important property to determine if a material is stable and thus can be used for different applications [10]. Moreover, ML methods have been trained using not only elemental compositions but also to the domain knowledge, which results in more precise results, but with the cost of increasing the computational complexity [5] [7].

ANN and Deep Neural Networks (DNN) have been used to predict the properties of different compounds showing an improvement in accuracy. In addition, with the use of graphics processing units (GPUs), such predictions have seen speedups of up to 35x when compared to prediction of ML methods [7]. OQMD data begin with the molecular formulas of a compound, and a decomposition of individual elements is carried out to create 2D arrays, where the rows represent the compounds and each column represents the portion of atoms found in the compound.

ElemNet is one of the DNN architectures used for the prediction of material properties, which consists of 17 fully connected layers with ReLU activation functions, and 4 dropout layers that help reduce the problem of over-fitting and provides higher accuracy than traditional ML methods such as Random

978-1-7281-6251-5/20/$31.00 ©2020 IEEE

Forest [7]. However, the efforts to improve the accuracy by increasing the number of layers and training with larger data sets (256,622 compounds) resulted in network saturation and over-fitting problems [7].

In this work, we propose the use of Residual Neural Networks (RNN) based on ResNet [11] that integrate shortcuts between layers and allow mapping different characteristics of the data within the network avoiding the problem of over-fitting encountered by the ElemNet architecture. Particularly, we construct DNN architectures based on Merge and Run (MnR) Networks [12], IRNet [9] and UNet [13], and use them to evaluate their effectiveness in predicting properties of materials. These architectures can be trained to predict (with different accuracy) any of the properties of the compounds found in OQMD such as: energy pa, volume pa, magmmon pa, bandgap, delta e and stability. The rest of the paper is organized as follows: in section 2 provides a brief description of the OQMD and presents a description of the proposed DNN architectures. Section 3 presents the experimental setup while section 4 presents the results and provides a discussion. Finally, section 5 presents the conclusions of this work and proposes future directions.

II. Methods

A. Dataset

We use the OQMD database using 342k compounds, which is the same organization presented in ElemNet [8] to allow for comparisons between the architectures and limit the impact of the training or test data. This selection allows to provide more precise comparisons of the different proposed architectures.

Each of our experiments used 90% of the data for training ($\sim$307k compounds) and 10% for testing ($\sim$34k compounds). The input data of the networks are vectors containing raw elemental compositions in the compound -86 elements of the periodic table - and formation enthalpy in eV/atom as output labels.

The output of the models would be a numerical value in units of eV/atom. The accuracy of the proposed model is measured by using the Mean Absolute Error (MAE) and the results are compared to some of the existing methods. The MAE is defined as (1):

$$MAE = \frac{1}{N}\sum |y - \widehat{y}| \tag{1}$$

where y is the formation enthalpy, $\widehat{y}$ is the formation enthalpy predicted by the model and N is the total number of predicted compounds.

We performed the training with 10 fold cross validation. Each fold is used once as a validation while the 9 remaining folds form the training set. At the end of the training, we obtain 10 different weights for the model. To measure the accuracy of each of the 10 weights, we load the weights and then predict the test data. Finally the weights of the fold that obtained the lowest MAE in the test data are taken.

Based on how residual networks allow us to solve gradient and over-fitting problems, we decided to carry out experiments with architectures that implement skip connections and that have had good results in other prediction problems [14]. After doing experiments with residual networks, we noticed that there were still over-fitting problems - We obtained very low MAE with the training data but high with the test data. Among the methods to reduce this problem we tried with different dropout values in the network, however it was not enough to improve the accuracy and we resorted to including batch normalization (BN) between the different layers. By implementing BN in the network, the dropout can be removed without over-fitting problems [15] [16], which ultimately improves the accuracy of some models.

B. FCMnR

This architecture is mainly inspired by MNRCSNet [17], a network based on Merge and run which has verified previous uses in the recovery of images from their compressive measurements [18] [19]. Its main characteristic is the small additive residual contributions within the network. For our experiments we obtained the best results using a version with five FCMnR blocks. Additionally, we made changes to the original architecture replacing convolutional layers to fully connected layers, allowing us to train it with preprocessed OQMD data. Among the network configurations, we experimented with a variation in the number of neurons in the FCMnR block, the location and number of layers that include batch normalization. The architecture and parameters presented in 3 are the most optimal when predicting the test data.

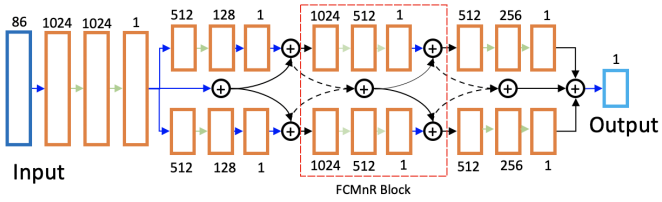

Fig. 1. Network diagram of the FCMnR ANN based architecture. The modular MnR block is highlighted in red. First blue arrow represents ReLU connection, and green arrows represent Batch Normalization + ReLU connections. Orange rectangles represent fully connected layers with the number of units above/below them. The solid circle denotes the sum operation and dot lines are averaged.

C. IRNet with cross validation

Among the DNN methods for the prediction of properties in materials, IRNet consists of an individual residual network: a succession of fully connected layers with intermediate jumps between the layers, as described in figure (2) [9].

The main feature of IRNet is the shortcut connections between layers which is done individually - the output of one layer is concatenated to the output of the following making the regression learning task easy - [9]. Using batch normalization before applying the activation function substantially reduces over-fitting and gradient fading problems. Our implementation of IRNet includes cross validation in its training phase which is not performed in the original IRNet implementation.

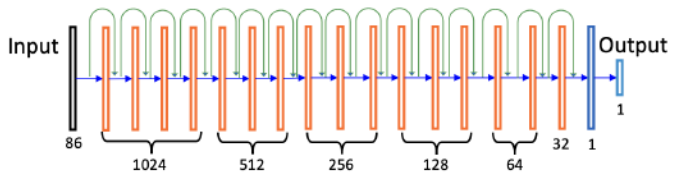

Fig. 2. Network diagram of IRNet, a Residual network architecture. Orange rectangles with the number of units inside represent fully connected layers + Batch Normalization + ReLU. Blue rectangle represents a fully connected layer + Linear activation. Skips connections are represented with green lines and are concatenated with the output of the layer, represented with blue lines.

D. FCUNet

UNet is a network architecture created for the segmentation of biomedical images. Its main characteristic is that it is formed by a sequence of layers with decreasing number of neurons to capture context in the data. Subsequently, a symmetric sequence of layers that allows location of characteristics given the previous context. This architecture was originally designed for image segmentation, receiving two-dimensional data as input and outputting the segmented image, which is also organized in two dimensions. ' For the purposes of predicting material properties and following the same approach of our modified FCMnR network, we substitute convolutional layers for fully connected layers in the original UNet to create FCUNet. In addition, after observing gradient and over-fitting problems, we also included batch normalization before the activation functions. Finally, we connect a last layer with a linear activation function and a single unit, which allows us to obtain the formation enthalpy predicted by the network.

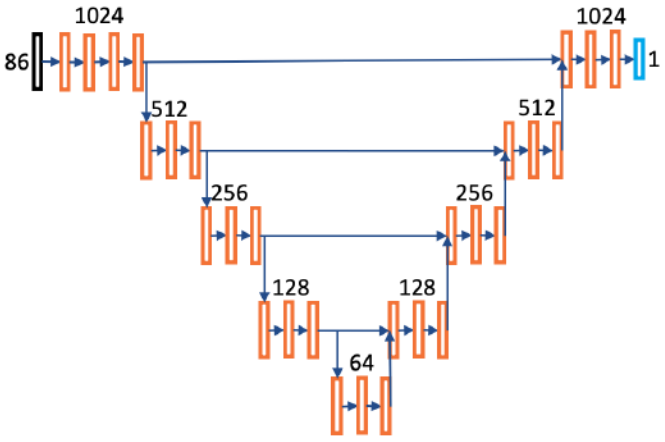

Fig. 3. Network diagram of FCU-Net architecture. Orange rectangles represent fully connected layers + Batch Normalization + ReLU and the number of units of each layer above. Blue rectangle represents a fully connected layer + Linear activation.

III. Experimental Setup

All models were trained and tested using the Tesla V100 32GB. The implementation was done using python 3.7.4 and Tensorflow 2.0.0 [20] with Keras API. We used scripts shared by ElemNet for OQMD preprocessing which are implemented using python 3.6 and sklearn [21].

TABLE I
MAE OF EACH ARCHITECTURE EVALUATING THE TEST DATA

Network architecture	MAE (ev/atom)
FCMnR	**0.0511**
IRNet with cross-validation	**0.0338**
FCU-net	**0.0397**

TABLE II
TRAINING AND PREDICTION TIME

Network architecture	Training time(sec)	Prediction time(sec)
FCMnR	∼87	∼9.4
IRNet with cross-validation	∼57	∼5.4
FCUnet	∼79	∼7.7

A. Networks training

For training we used a batch size of 64, an Adam optimizer with a learning rate of 10^{-4} and Mean Absolute Error loss function and metric. The number of epochs used was varied to a patience. In the case of networks based on FCMnR, a patience of 300 epochs has been found to be optimal. In the case of IRNet with cross validation and FCUNet, the patience was 200 epochs.

IV. Results and Discussion

Table (I) shows the best accuracy that we obtained when predicting the test data in each of the 10 fold cross validation. These can be compared with the lowest MAE reported by the authors of ElemNet which uses the same dataset as us and is 0.0437 ev/atom. Similarly, the original IRNet presents results of 0.0411 ev/atom. Based on these results we can observe that IRNet with cross validation and FCUNet provide a higher accuracy than the state of the art methods. In the case of FCMnR, we obtained less accurate results, but given the complexity in parameter configuration of this type of architectures, more work is needed in order to identify a proper configuration capable of reducing MAE.

An important factor in predicting material properties is the required time to make a prediction since one of the applications in this area is the discovery of new materials, requiring the prediction of thousands of compounds [7]. Table (II) shows the training time per epoch in seconds and the prediction time of the total test data (∼34k compounds). These times are obtained with the settings mentioned below. Notice that FCMnR requires the less amount of training time while FCUNet require the most. In terms of prediction time, IRNet with cross validation and FCMnR required similar time while again FCUNet requires a higher amount of time.

V. Conclusions

After applying some variations of residual networks, better results were obtained than those reported by state of the art architectures such as ElemNet and IRNet. The FCUNet architecture stands out because it was originally raised for an

image segmentation problem, however with the changes in the layers we obtained a very good result for this prediction problem. This work shows how architectures that reach the state of the art can be adapted in problems of different areas of research such as prediction and classification of images, in areas of Materials Science and Chemical Engineering.

There are many variations to be made in the parameters of FCMnR and FCUNet based architectures, such as increasing the depth of the fully connected blocks and vary the batch sizes. Due to the complexity of the architectures that are being experimented as well as the cross-validation method to obtain the results; the search for the optimal parameters in the models requires hours of computation. Future work in this area includes the identification of such parameters that further improve the prediction accuracy of FCMnR and FCUNet.

VI. Acknowledgements

This material is based upon work supported in part by the National Science Foundation under Grant No. OIA-1849243.

References

[1] A. R. Oganov, C. J. Pickard, Q. Zhu, and R. J. Needs, "Structure prediction drives materials discovery," 2019.

[2] G. Vignale and M. Rasolt, "Density-functional theory in strong magnetic fields," *Phys. Rev. Lett.*, vol. 59, pp. 2360–2363, Nov 1987. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.59.2360

[3] P. Hohenberg and W. Kohn, "Inhomogeneous electron gas," *Phys. Rev.*, vol. 136, pp. B864–B871, Nov 1964. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRev.136.B864

[4] I. B. Obot, D. D. Macdonald, and Z. M. Gasem, "Density functional theory (DFT) as a powerful tool for designing new organic corrosion inhibitors: Part 1: An overview," 2015.

[5] J. E. Saal, S. Kirklin, M. Aykol, B. Meredig, and C. Wolverton, "Materials design and discovery with high-throughput density functional theory: The open quantum materials database (OQMD)," *JOM*, 2013.

[6] S. Kirklin, J. E. Saal, B. Meredig, A. Thompson, J. W. Doak, M. Aykol, S. Rühl, and C. Wolverton, "The Open Quantum Materials Database (OQMD): Assessing the accuracy of DFT formation energies," *npj Computational Materials*, 2015.

[7] D. Jha, L. Ward, A. Paul, W. keng Liao, A. Choudhary, C. Wolverton, and A. Agrawal, "ElemNet: Deep Learning the Chemistry of Materials From Only Elemental Composition," *Scientific Reports*, 2018.

[8] D. Jha, K. Choudhary, F. Tavazza, W. keng Liao, A. Choudhary, C. Campbell, and A. Agrawal, "Enhancing materials property prediction by leveraging computational and experimental data using deep transfer learning," *Nature Communications*, 2019.

[9] D. Jha, C. Wolverton, L. Ward, I. Foster, Z. Yang, L. Wei-Keng, A. Choudhary, and A. Agrawal, "IRNet: A general purpose deep residual regression framework for materials discovery," in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2019.

[10] B. Meredig, A. Agrawal, S. Kirklin, J. E. Saal, J. W. Doak, A. Thompson, K. Zhang, A. Choudhary, and C. Wolverton, "Combinatorial screening for new materials in unconstrained composition space with machine learning," *Physical Review B - Condensed Matter and Materials Physics*, 2014.

[11] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *CoRR*, vol. abs/1512.03385, 2015. [Online]. Available: http://arxiv.org/abs/1512.03385

[12] L. Zhao, J. Wang, X. Li, Z. Tu, and W. Zeng, "On the connection of deep fusion to ensembling," *CoRR*, vol. abs/1611.07718, 2016. [Online]. Available: http://arxiv.org/abs/1611.07718

[13] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2015.

[14] M. S. Ebrahimi and H. K. Abadi, "Study of residual networks for image recognition," *arXiv preprint arXiv:1805.00325*, 2018.

[15] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *32nd International Conference on Machine Learning, ICML 2015*, 2015.

[16] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *Journal of Machine Learning Research*, 2014.

[17] F. Arias, H. Sierra, and E. Arzuaga, "A Framework for An Artificial Neural Network Enabled Single Pixel Hyperspectral Imager," in *Workshop on Hyperspectral Image and Signal Processing, Evolution in Remote Sensing*, 2019.

[18] L. Gondara, "Medical Image Denoising Using Convolutional Denoising Autoencoders," in *IEEE International Conference on Data Mining Workshops, ICDMW*, 2016.

[19] J. Masci, U. Meier, D. Cireşan, and J. Schmidhuber, "Stacked convolutional auto-encoders for hierarchical feature extraction," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2011.

[20] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, "TensorFlow: Large-scale machine learning on heterogeneous systems," 2015, software available from tensorflow.org. [Online]. Available: https://www.tensorflow.org/

[21] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.