PIXEL CHEM: A REPRESENTATION FOR PREDICTING MATERIAL PROPERTIES WITH NEURAL NETWORK

Anonymous authors
Paper under double-blind review

ABSTRACT

In this work we developed a new representation of the chemical information for the machine learning models, with benefits from both the real space (R-space) and energy space (K-space). Different from the previous symmetric matrix presentations, the charge transfer channel based on Paulings electronegativity is derived from the dependence on real space distance and orbitals for the hetero atomic structures. This representation can work for the bulk materials as well as the low dimensional nano materials, and can map the R-space and K-space into the pixel space (P-space) by training and testing 130k structures. P-space can well reproduce the R-space quantities within error 0.53. This new asymmetric matrix representation double the information storage than the previous symmetric representations.This work provides a new dimension for the computational chemistry towards the machine learning architecture.

1 INTRODUCTION

The most common way of representing material structures is using the coordinates of atoms together with the periodic unit cell vectors and categorized these information as unstructured data (i.e. `cif` files)(Lam Pham et al., 2017). However, those unstructured data cannot put into neural network directly and need to be reformatted as matrices for the advanced GPU based neural network computational architecture, which are extremely robust in pixel processing.(Hadji & Wildes, 2018) As Scholkopf & Smola (2002) indicated, when using Kernel-based machine learning models, in order to reach a fast and accurate prediction of first principle properties, a single Hilbert space of atomistic system is required, in which regression is carried out. In all above situation, an ideal representation needs to possess several essential characters like invariant, unique, continuous, general and efficient etc.(Rupp, 2015; Bartok et al., 2013; von Lilienfeld et al., 2015): In addition, apart from the R-space information, other K-space information might be more important for the overall property of material, especially when the research target relates with the band gap or excitation.

Various kinds of representation have been developed and reported. The Coulomb matrix representation (CM, Rupp (2015); Faber et al. (2015); Rupp et al. (2012)) coded nuclear charge and atom displacement inside to predict atomization energies and formation energies. Based on it, bag of bond (BOB)(Hansen et al., 2015) representation constructed an effective pairwise interatomic potential to strip down pair-wise variant of CM, which improved the predictive model remarkably. Many-body tensor representation (MBTR, Huo & Rupp (2017)) is derived from 2 above and reached a significant drop in error rate. Symmetry functions(Behler, 2011; Eshet et al., 2010; 2012) is another method which focuses on the representation of local chemical environment of atoms and maps this representation to atomic energy. Recently, based on Simplified molecular-input line-entry system (SMILES, Weininger (1988)) representation, a new 3N-MCTS module based on Mote-Carlo tree search is proposed by Segler et al. (2018). This module keeps only the structure-connection information and trained large dataset to provide supplementary instructions for retrosynthetic analysis. In addition, representations like bonding-angular machine learning(Hansen et al., 2015) and others(Huang & von Lilienfeld, 2016; Schutt et al., 2014) have also been reported and tested. However, most of these representations only satisfy those requirements partially, in which unique and continuous principles are often violated.

The most severe shortcoming for both CM and BoB is that they focus on predicting the extentive quantaty,in R space like structural energy but left other intensive quantities in K space like band gap.

Though some reported researches also included other properties(Huang & von Lilienfeld, 2016), their error rates are way larger comparing with energy. Overall, the R-space and K-space are still separated for this kind of represenation. The possible improvements could be made by endorsing more R/K space information to the targeted representations. Recall that CM only coded atomic number together with distance, BoB added roughly designed bonds energy in compliment, both failed to represent most of the properties in K-space. Therefore, defining a new representation with enough physical identity from R/K space is a vital task for material informatics to ensure the uniqueness and continuity.

2 DEFINITION

2.1 PIXEL CHEM REPRESENTATION FOR FINITE STRUCTURES

Here in this work,we proposed a representations with projected both R and K space chemical information into the pixel space (Pixel Chem). We start from the property of different atoms, mainly binding strength and charge transfer ability. To describe these, we first introduce two new matrices, $\boldsymbol{C}$ and $\boldsymbol{B}$. $\boldsymbol{C}$, which mainly reflects the charge transfer ability, is based on electron affinity energy and ionization energy of atoms. Inspired by Mullikan Electronegativity(Mulliken, 1934), it is defined as $\boldsymbol{C}_{i,j} = (\boldsymbol{I}\boldsymbol{E}_{i,:} + E\boldsymbol{A}_{:,j})/2$. As for binding strength, $\boldsymbol{B}_{i,j}(U_{\text{Bond}})$ is imported, which represents the typical bond energy formed by two atoms. Note that though $\boldsymbol{B}_{i,j} = \boldsymbol{B}_{j,i}$, $\boldsymbol{C}_{i,j} \neq \boldsymbol{C}_{j,i}$, which means this matrix is not symmetric between diagonal. Finally, we proposed a displacement matrix $\boldsymbol{D}_{i,j} = \boldsymbol{D}_{j,i} = \exp(-|\boldsymbol{r}_i - \boldsymbol{r}_j|/a)$ to represent the relative position of different atoms, where $|\boldsymbol{r}_i - \boldsymbol{r}_j|$ represent the Euclidean distance between two atoms and a is the interaction strength coefficient trained by machine learning methods. To mix all the things up, we proposed a charge & energy matrix $\boldsymbol{E}_{i,j}$ which can be produced by multiplying $\boldsymbol{D}_{i,j}$ with $\boldsymbol{C}_{i,j}^T$ and $\boldsymbol{B}_{i,j}^T$ as follows:

$$\boldsymbol{E} = \boldsymbol{D} \times \boldsymbol{C}^T \times \boldsymbol{B}^T \tag{1}$$

This operation ensures that for every element $\boldsymbol{E}'$ in $\boldsymbol{E}_{i,j}$, there will be a product of the $\boldsymbol{D}'$ in $\boldsymbol{D}_{i,j}$ and $\boldsymbol{B}'$ in $\boldsymbol{B}_{i,j}^T$ corresponding to the same atom. Therefore, for all the atoms, we have:

$$\boldsymbol{E}'_{i,j} = \boldsymbol{D}'_{i,j} \cdot \boldsymbol{C}'_{i,j} \cdot \boldsymbol{B}'_{i,j} \tag{2}$$

Atomic orbits information, including arrangement, position and number, is another important factor that highly influence the property of a structure. In order to represent those information, we first construct a 1×10 sized orbital charge vector $\boldsymbol{o}_i = [N_{\sum core}, N_s, N_{px}, N_{py} \dots N_{dyz}, N_{dxz}]$ (core stands for all the core electrons) is created to represent (mainly) the electron configuration of every atom in valance electron orbitals. The binding energy B_E of each orbital also needs to be considered to distinguish those orbitals in different chemical environments. We therefore product the $\exp(-B_E^{\text{orbit}})$ of every orbital in $\boldsymbol{o}_i$ to get an energy encoded orbital matrix $\boldsymbol{o}_{\text{BE}i}$, where:

$$\boldsymbol{o}_{\text{BE}i} = [N_{\sum \text{core}} \cdot \exp(-B_E^{\sum \text{core}}), N_s \cdot \exp(-B_E^s), \cdots N_{dyz} \cdot \exp(-B_E^{d_{x^2-y^2}})] \tag{3}$$

To better represent the chemical environment of each atom, especially the different condition of each orbital, we hereby introduce two closely related methods, tight-binding method and linear combination of atomic orbitals (LCAO, Slater & Koster (1954)) method, to our representation model. With the existence of atomic Hamiltonian, this matrix can precisely describe the interaction between two atoms orbital by orbital. In detail, a 10×10 sized angular interaction matrix $\boldsymbol{A}_{i,j}$, representing from all core electrons to dz orbital electrons, is implemented for all atoms in one specific structure, which formats as:

$$\begin{bmatrix} \mathit{Int}(i[\sum \text{core}], j[\sum \text{core}]) & \mathit{Int}(i[\sum \text{core}], b[s]) & \cdots & \mathit{Int}(i[\sum \text{core}], j[d_{x^2-y^2}]) \\ \mathit{Int}(i[s], j[\sum \text{core}]) & \mathit{Int}(i[s], j[s]) & \cdots & \mathit{Int}(i[s], j[d_{x^2-y^2}]) \\ \vdots & \vdots & \ddots & \vdots \\ \mathit{Int}(i[d_{x^2-y^2}], j[\sum \text{core}]) & \mathit{Int}(i[d_{x^2-y^2}], j[s]) & \cdots & \mathit{Int}(i[d_{x^2-y^2}], j[d_{x^2-y^2}]) \end{bmatrix}_{10\times 10} \tag{4}$$

which $i[m]$ represent the m orbital of i atom.

Recalled 3, every element inside the matrix of atoms, shown as $\textit{Int}(A[\text{orbit}a], B[\text{orbit}b])$, represent the interaction strength between two orbitals which belong to two atoms respectively, it is yielded out as follows:

$$\textit{Int}(i[m], j[n]) = N_{i[m]} \cdot \exp(-B_E^{i[m]}) \cdot N_{j[n]} \cdot \exp(-B_E^{j[n]}) \cdot \textit{Coeff} \tag{5}$$

Which can be summarized as:

$$\textit{Int}(i[m], j[n]) = \boldsymbol{o}_{\text{BE}i}[m] \cdot \boldsymbol{o}_{\text{BE}j}[n] \cdot \textit{Coeff} \tag{6}$$

The coefficient inside this formula is defined by the position of two atoms, the direction of two target orbits and the energy of them. To be specific, the calculation was primarily based on the Slater-Koster Approximation(Hehre et al., 1972), which was designed to solve the overlapping area between two selected orbits. Some orbits, such as p_x, p_y and p_z orbits for the same atom, are perpendicular to each other, which will result in 0 at the corresponding element. The calculation detail could be found in supplementary notes. In all, this angular interaction matrix expanded the previous matrix for 10 times, which encodes more physical meanings, especially the Hamiltonian of each atom, which could benefit the learning efficiency and accuracy of our training model.

To combine all the angular matrices $\boldsymbol{A}_{i,j}$ with the previous $\boldsymbol{E}$. We product every element $\boldsymbol{E}_{i,j}$ with its corresponding $\boldsymbol{A}_{i,j}$ to get new element $\boldsymbol{M}_{i,j}$. Note that new element $\boldsymbol{M}_{i,j}$ is also a 10×10 matrix. Therefore, $\boldsymbol{M}$ will be the final Pixel Chem representation, denoted as:

$$\boldsymbol{M}_{i,j} = \boldsymbol{E}_{i,j} \cdot \boldsymbol{A}_{i,j} \tag{7}$$

Which can also be written as:

$$\boldsymbol{M}_{i,j} = \boldsymbol{D}_{i,j} \cdot \boldsymbol{C}_{i,j} \cdot \boldsymbol{B}_{i,j} \cdot \boldsymbol{A}_{i,j} \tag{8}$$

which included electron position configuration $\boldsymbol{A}_{i,j}$, charge transfer ability $\boldsymbol{C}_{i,j}$ and position information $\boldsymbol{D}_{i,j}$. $\boldsymbol{M}_{i,j}$ has two meanings, depending on whether $i = j$. When $i = j$, it corresponds to the diagonal parts, where $\boldsymbol{D}_{i,j} = 1$. Without the influence of $\boldsymbol{D}_{i,j}$, the diagonal parts are served as an element list which contains orbital and energy information. When $i \neq j$, it corresponds to the off-diagonal parts, where $\boldsymbol{D}_{i,j}$ keeps decreasing when displacement is increasing. This part in the matrix represents the interaction strength of two selected atoms. Moreover, though the different input order of atoms will yield out a different $\boldsymbol{M}$, they can be easily transferred into the same arrangement by switching the column and row of two target atoms correspondingly. This will not break the order because one atom will only influence its corresponding column and row.

Moreover, when implementing Pixel Chem into organic structures, this representation could be simplified because most of the elements in organic structures are from the first 3 periods which means they do not have d orbitals. This simplification results in a smaller matrix size and could handle the representation of organic molecules in a faster way. When considering organic structures, hybridization, especially for carbon and nitrogen, the type of chemical bonds need to be considered. We hereby implement an adjoint classifier to estimate the chemical state for each carbon, which will return the bond type and hybridization state of carbon atoms. Take sp3 as an example, instead of which proposed above, the $\boldsymbol{A}_{i,j}$ for carbon atom will be transferred to:

$$\begin{bmatrix} \textit{Int}(a[\sum \text{core}], b[\sum \text{core}]) & \textit{Int}(i[\sum \text{core}], j[s]) & \cdots & \textit{Int}(i[\sum \text{core}], j[pz]) \\ \textit{Int}(i[s], j[\sum \text{core}]) & \textit{Int}(i[s], j[s]) & \cdots & \textit{Int}(i[s], j[pz]) \\ \vdots & \vdots & \ddots & \vdots \\ \textit{Int}(i[pz], j[\sum \text{core}]) & \textit{Int}(i[pz], j[s]) & \cdots & \textit{Int}(i[pz], j[pz]) \end{bmatrix}_{5\times 5} \tag{9}$$

Hereby the $\boldsymbol{o}_{\text{BE}i} = [N_{\sum \text{core}} \cdot \exp(-B_E^{\sum \text{core}}), N_s \cdot \exp(-B_E^s), \cdots, N_{pz} \cdot \exp(-B_E^{pz})]$.

as 4 sp3 orbitals present the same property. B_E^{sp3} will also take the place of their initial values. Figure 1 shows different orbital matrix for different hybridization state.

To illustrate this representation in a straight forward way, we visualized the pixel chem using `matplotlib` in Figure 2. Each colored square represents a numerical value. Black color stands for absolute 0, which also means there is no electron in the corresponding orbital. Some of the off-diagonal and diagonal parts are carried out with explanation, which are used to represent atoms properties (for diagonal parts) or interaction strength between atoms (for off-diagonal parts).

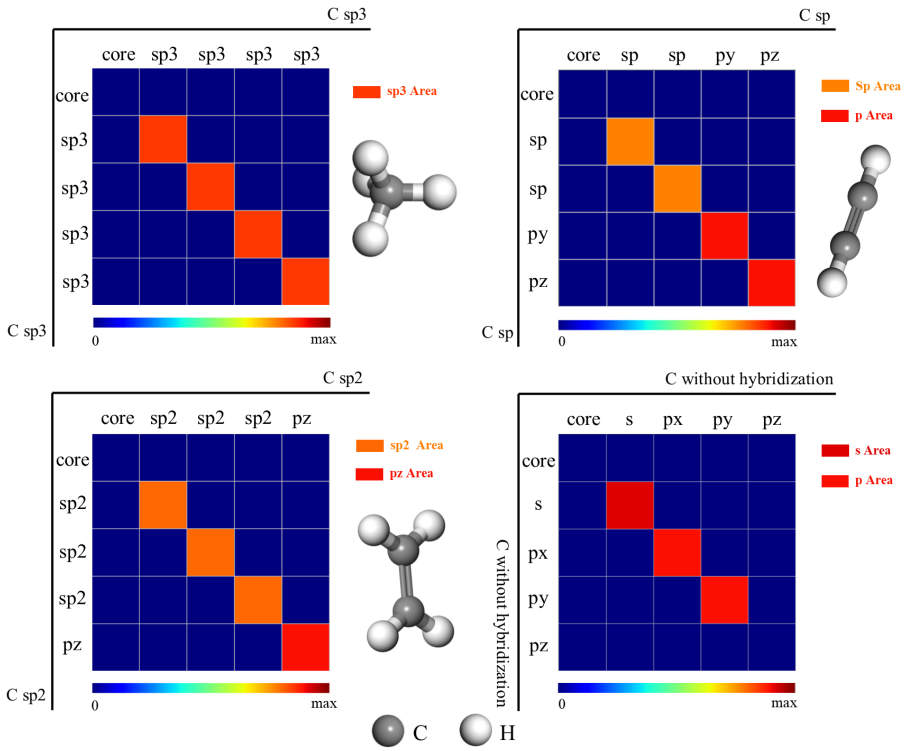

Figure 1: visualized orbital matrix representation for different hybridization states of carbon. We use a color spectrum from blue to red to represent the value of different element. These kinds of matrices lie on the diagonal parts of Pixel Chem. On off-diagonal parts, elements distance matrix is smaller, so the color will be bluer in most of the cases. All the s and p orbits, including sp/sp2/sp3 hybridization orbits, have no overlap area between each other, which yields out all 0 values in off diagonal parts of four matrices.

2.2 FROM FINITE TO PERIODIC STRUCTURES

Apart from finite structures like fullerene, the bulk periodic structures like diamond and quasi 2D layered periodic structures such as graphite or multi-layer graphene could also be represented by Pixel Chem representation,as shown in Figure 3 (a) to (c). In this situation, we calculate the interaction including atoms belonging to itself and two adjacent primitive cells, $-z$ and $+z$ respectively. For those adjacent cells, $\boldsymbol{D}_{i,j}$ is the only difference and it was altered as $\boldsymbol{D}_{i,j} = \boldsymbol{D}_{j,i} = \sum_{k=-1}^{1} \exp(-|\boldsymbol{r}_i - \boldsymbol{r}_j| + kc)$ (c is the lattice parameter, $k = \{-1, 0, 1\}$ for adjacent cells) to include all interaction parts. Note that the $\boldsymbol{D}_{i,j}$ in diagonal part will also be larger, which simply indicates whether this matrix stands for a periodic structure or a finite structure. By measuring this part, this structures lattice constant, C, could be determined. Moreover, this representation comes together with an adjacent cell interaction strength that could be measured directly from the matrix. Generally, this methodology also works for periodic structures which have quansi 0 to 2-dimensional periodicity. In those cases, the adjacent cells in corresponding dimension will be calculated in the same way. Figure 3 lists 4 other different structures for a better view of the unity of Pixel Chem. Among those 4 structures, all except C_{60} are periodic and diamond has periodicity for all three dimensions. This can be directly seen from the diagonal parts where the corresponding orbital interaction value of periodic dimensions will increase. Note that for the bottom two isoelectronic structures, apart from the orbital difference, graphite is strictly symmetric across diagonal while BN is asymmetric because of the different charge transfer ability.

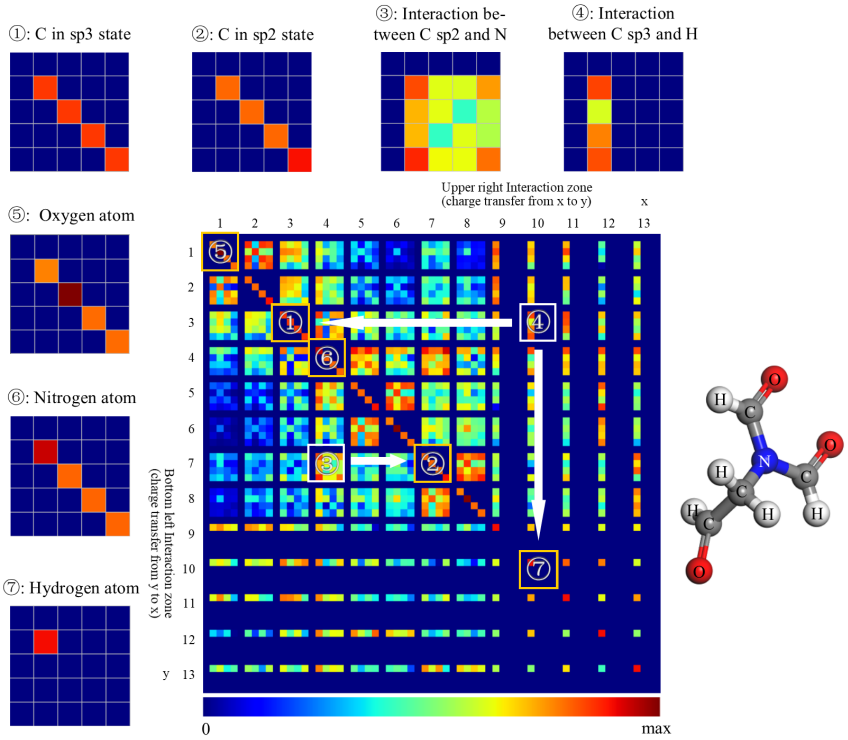

Figure 2: Visualized Pixel Chem for a typical molecule in QM9 dataset, $C_4H_5O_3N$. Note that upper right and lower left parts are not perfectly symmetric because the charge transfer ability $\boldsymbol{C}_{Ei,j}$ differs. The color of different pixels is together influenced by orbitals, bonds, angles and charges.

2.3 Benefits

Comparing with other representations in previous works, Pixel Chem have enormous benefits. Firstly, it satisfied all the requirements indicated by Rupp (2015); Bartok et al. (2013); von Lilienfeld et al. (2015), which are (1) invariant, (2) unique, (3) continuous, (4) general, (5) fast and efficient for calculation. For the invariant property, the neural network we designed is order uncorrelated. Meanwhile, the various of position and atom property of different structures could secure the uniqueness even for isomers. For (3), though the representation is discrete, the interaction between each two atoms and their major orbitals are considered with no information gap. For (4), this representation could handle both finite and periodic structures with only a little bit difference in definition. On the contrary, the CM and most of previous modules could only handle finite structures. For (5), comparing with conventional methods, which typically needs more than 10 minutes for a structure with 20 atoms, our neural network could output the predict value in less than 1 second if the model has been well trained. Morover,the matrix representation in the Pixel Chem is asymmetrical mirrored by the diagonals due to the charge transfer directions in hetero atomic structures, doubled of the information storage size compared with the previous matrix representation.

3 Network

What we designed is a Pixel Chemistry network (PCnet) which can learn a representation for the prediction of molecular properties such as energy and band gap. The network reflect basic physical laws in molecule including invariance to atom indexing. All the prediction for different properties are rotationally invariant.

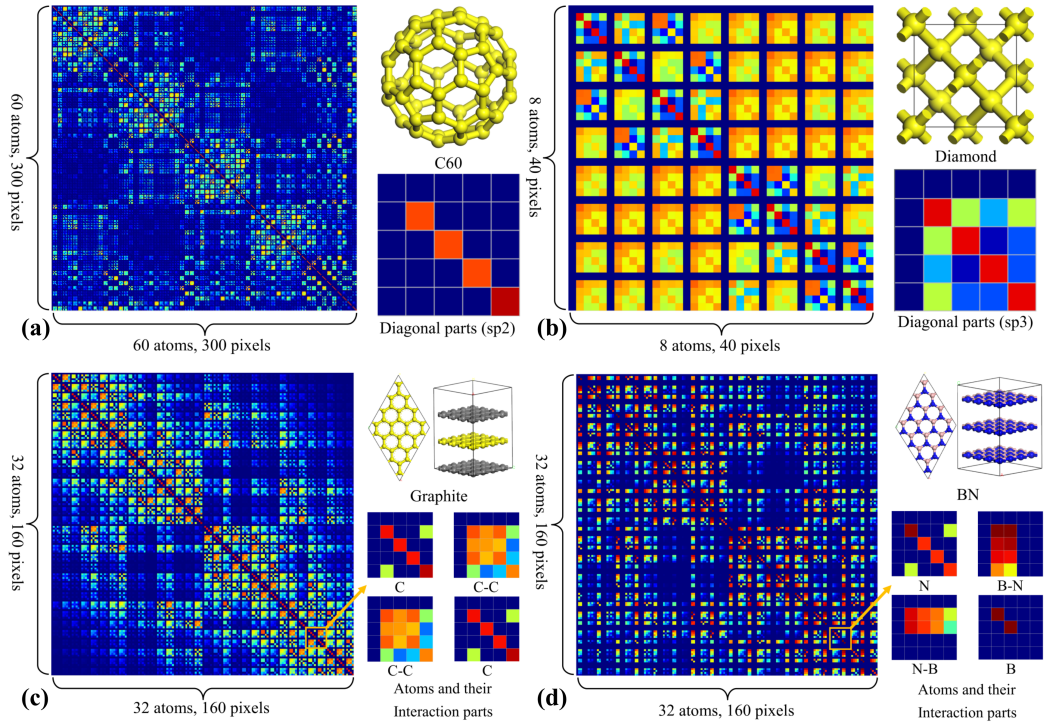

Figure 3: (a) (b) (c) (d): 3 carbon allotropes and a boron nitride (BN) structure generated by Pixel Chem. (a) C_{60}, 60 Carbon atoms without periodicity, symmetric, (b) Diamond, 8 Carbon atoms with periodicity on 3 dimensions, symmetric, (c) Graphite, 32 Carbon atoms with z direction periodicity, symmetric, (d) BN, 16 Boron atoms and 16 Nitrogen atoms with z direction periodicity, a graphite isoelectronic, asymmetric.

3.1 ARCHITECTURE

The figure 4 shows an overview of the architecture of PCnet, which is inspired by Deep Tensor Neural Network (DTNN, Schütt et al. (2017)). Interactions between atoms are modeled by three interaction blocks. The final prediction of the properties is obtained after atom-wise updates of the feature representation and pooling of the resulting atom-wise energy. We will discuss the different components of the network in the following.

Molecular representation A molecule in a certain conformation can be described uniquely by a set of n atoms with data of atoms $\boldsymbol{A} = (\boldsymbol{A}_1, \ldots, \boldsymbol{A}_n)$ and interaction between each two of all atoms. We divided the interaction into upper part $\boldsymbol{U} = (\boldsymbol{U}_1, \boldsymbol{U}_2, \ldots, \boldsymbol{U}_n)$ and lower part $\boldsymbol{L} = (\boldsymbol{L}_1, \boldsymbol{L}_2, \ldots, \boldsymbol{L}_n)$ by the position in the matrix. $\boldsymbol{U}_i = (\boldsymbol{U}_{i,1}, \cdots, \boldsymbol{U}_{i,(i-1)}, \boldsymbol{U}_{i,(i+1)}, \cdots, \boldsymbol{U}_{i,n})$ represent all the upper part interaction between atom i and other atoms, $\boldsymbol{L}_i = (\boldsymbol{L}_{i,1}, \cdots, \boldsymbol{L}_{i,(i-1)}, \boldsymbol{L}_{i,(i+1)}, \cdots, \boldsymbol{L}_{i,n})$ represent all the lower part interaction between atom i and other atoms.

Through the layers of PCnet, we use a tuple of features $\boldsymbol{X}^l = (\boldsymbol{x}_1^l, \ldots, \boldsymbol{x}_n^l)$ to represent the atoms or interaction data, where $\boldsymbol{x}_i^l \in \mathbb{R}^F$ with the number of atoms n, the current layer l, the number of features map F.

Atom-wise layer A dense layer that are applied to the feature $\boldsymbol{x}_i^l$ of atom i separately with learnable weight $\boldsymbol{W}^l$ and learnable bias $\boldsymbol{b}^l$ in layer l:

$$\boldsymbol{x}_i^{l+1} = \boldsymbol{W}^l \boldsymbol{x}_i^l + \boldsymbol{b}^l$$

The atom-wise layer are designed to represent the combination of the origin feature map. PCnet can work for molecules with any size since weights of atom-wise layers are shared across atoms.

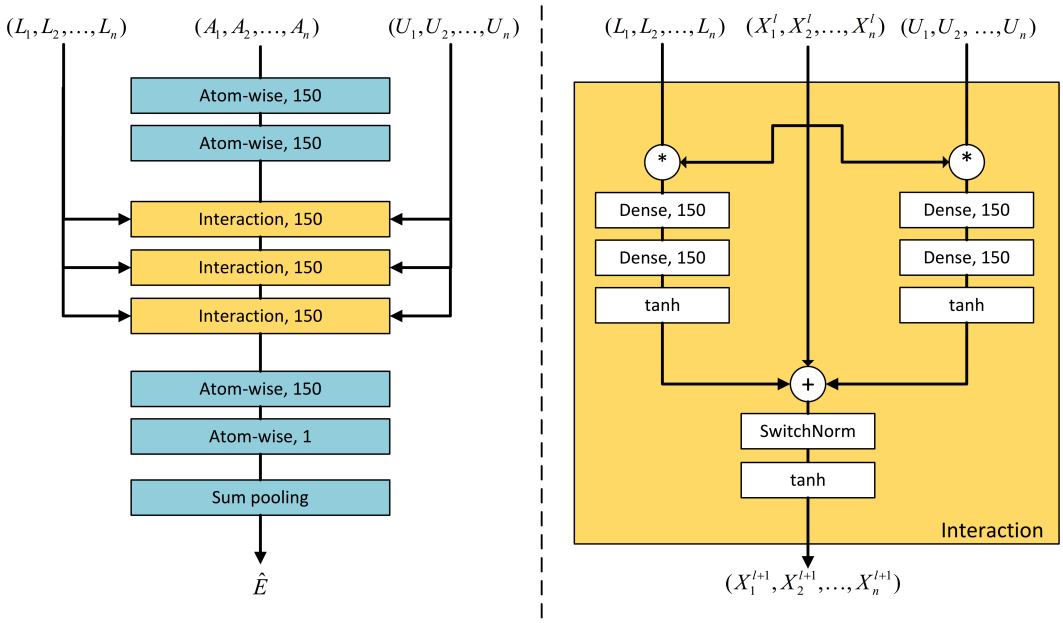

Figure 4: Illustration of PCnet with an architectural overview(left) and the interaction block(right).

Interaction The interaction block are designed to update the representations of atoms based on the interaction with other atoms. The interaction block use residual connection inspired by ResNet(He et al., 2016):

$$\boldsymbol{x}_i^{l+1} = \boldsymbol{x}_i^l + \boldsymbol{r}_{i1}^l + \boldsymbol{r}_{i2}^l$$

As shown in the interaction block in figure, the residual $\boldsymbol{r}_{i1}^l$ and $\boldsymbol{r}_{i2}^l$ are computed through two dense layers and a $\tanh$ non-linearity with lower and upper part interaction as input.

After the residual, we use switchable normalization layer (SwitchNorm, Luo et al. (2018)) to normalize the feature $\boldsymbol{x}_i^l$:

$$\boldsymbol{x}_i^{l+1} = \gamma \frac{\boldsymbol{x}_i^l - \mu}{\sqrt{\sigma^2 + \epsilon}} + \beta$$

where γ and β are a learnable scale and a learnable shift parameter respectively. ϵ is a small constant to prevent numerical instability. The SwitchNorm is continuing with a $\tanh$ non-linearity.

We keep the number of feature map to be constant $F = 150$ throughtout the whole network except the last atom-wise layer. All atom-wise layers and dense layers are followed by a learnable PReLU non-linearity(Xu et al., 2015). The last atom-wise layer is followed by a hyperbolic tangent non-linearity. The feature output from the last atom-wise layer represents the differences between the ground state energy and the energy contributed in the molecule, followed by a sum pooling layer which sum all atoms' contributed energy to get the predicted energy.

3.2 TRAINING

We apply PCnet to the quantum chemistry data set QM9(Ramakrishnan et al., 2014). QM9 has 133885 molecules. We remove 3423 unreasonable molecules with largest absolute cohesive energy, which may have N-N-N bonds in a single molecule. The cohesive energy is defined as the energy of molecule subtracts to the sum of atoms' energies in the molecule. Cohesive energy is useful in filtrating molecules by their stability. This energy is usually negative, and because of the subtraction operation, it can show the part of energy used to combine atoms together. A molecule is considered as more stable if less energy is used to combine atoms together.

We randomly chose 10000 molecules for validation, 10000 molecules for testing, and used remains for training. For energy prediction, the model was trained using SGD with the ADAM optimizer(Kingma & Ba, 2014) with 32 molecules per mini-batch for 1 million steps. We used 10^{-3}

Table 1: Comparison of mean absolute error of previous approaches (left) and our network (right) in $\mathrm{kCal/mol}$.

Target	BAML	BOB	CM	ECFP4	HDAD	GC	GG-NN	DTNN	PCnet
U0	1.21	1.43	2.98	85.01	0.58	3.02	0.83	0.84[1]	0.53
gap	3.28	3.41	5.32	3.86	2.49	1.78	1.70	N/A	2.13
HOMO	2.20	2.20	3.09	2.89	1.54	1.18	1.17	N/A	1.44
LUMO	2.76	2.74	4.26	3.10	1.96	1.10	1.08	N/A	1.26

as the initial learning rate and an exponential learning rate decay with rate 0.96 every 2 epoches (220924 steps). All targets were normalized to mean 0 and variance 1. We minimized the mean absolute error between predicted energy $\hat{E}$ and true energy E.

3.3 RESULT

In Table 1 we compare the performance of PCnet and the previous approaches. These baselines include 5 different hand engineered molecular representations, which then get fed through a standard, off-the-shelf classifier. These input representations include the Coulomb Matrix (CM, Neese (2003)), BoB, Bonds Angles, Machine Learning (BAML, Huang & von Lilienfeld (2016)), Extended Connectivity Fingerprints (ECPF4, Rogers & Hahn (2010)), and Projected Histograms (HDAD, Faber et al. (2017a)) representations. In addition to these hand engineered features we include the Molecular Graph Convolutions model (GC, Kearnes et al. (2016)), the original GG-NN model(Li et al., 2015) trained with distance bins and DTNN.

We can see from Table 1 that PCnet can achieve the lowest error bars (0.53) in the R-space U0 energy quantity due to the directional coulomb interaction in the representation without any additional load for imposing more geometric information than HDAD. The K-space accuracy is already below the DFT(B3LYP, Faber et al. (2017b)) level errors.

4 CONCLUSION

We proposed a new representation of material structures, named as Pixel Chem and designed a PCnet neural network to process this module. We also introduced a new pixel space apart from R-space and K-space in conventional chemistry representations. Distinct from those representations which focus mainly on numerical approach, our module covers both finite and periodic systems. In the future, this representation has large prospect. First, the combination of parameters can still be optimized to carry more information with less space. Moreover, more physical and chemical properties, such as Young's modulus and melting point, can be predicted using the same way. Finally, it can be wildly implemented in other field. One example is computational biology, where it can be used to construct larger material structures.

REFERENCES

Albert P Bartok, Risi Kondor, and Gabor Csanyi. On representing chemical environments. *Physical Review B*, 87(18):184115, 2013.

Jorg Behler. Atom-centered symmetry functions for constructing high-dimensional neural network potentials. *The Journal of chemical physics*, 134(7):074106, 2011. ISSN 0021-9606.

Hagai Eshet, Rustam Z Khaliullin, Thomas D Kuhne, Jorg Behler, and Michele Parrinello. Ab initio quality neural-network potential for sodium. *Physical Review B*, 81(18):184107, 2010.

Hagai Eshet, Rustam Z Khaliullin, Thomas D Kuhne, Jorg Behler, and Michele Parrinello. Microscopic origins of the anomalous melting behavior of sodium under high pressure. *Physical review letters*, 108(11):115701, 2012.

[1] As reported in DTNN. The model was trained on a different train/test split with 100k training samples vs about 110k used in our experiments.

FA Faber, L Hutchison, B Huang, Ju Gilmer, SS Schoenholz, GE Dahl, O Vinyals, S Kearnes, PF Riley, and OA von Lilienfeld. Fast machine learning models of electronic and energetic properties consistently reach approximation errors better than dft accuracy. *arXiv preprint arXiv:1702.05532*, 2017a.

Felix Faber, Alexander Lindmaa, O Anatole von Lilienfeld, and Rickard Armiento. Crystal structure representations for machine learning models of formation energies. *International Journal of Quantum Chemistry*, 115(16):1094–1101, 2015. ISSN 1097-461X.

Felix A Faber, Luke Hutchison, Bing Huang, Justin Gilmer, Samuel S Schoenholz, George E Dahl, Oriol Vinyals, Steven Kearnes, Patrick F Riley, and O Anatole von Lilienfeld. Prediction errors of molecular machine learning models lower than hybrid dft error. *Journal of chemical theory and computation*, 13(11):5255–5264, 2017b.

Isma Hadji and Richard P Wildes. What do we understand about convolutional networks? *arXiv preprint arXiv:1803.08834*, 2018.

Katja Hansen, Franziska Biegler, Raghunathan Ramakrishnan, Wiktor Pronobis, O Anatole Von Lilienfeld, Klaus-Robert Müller, and Alexandre Tkatchenko. Machine learning predictions of molecular properties: Accurate many-body potentials and nonlocality in chemical space. *The journal of physical chemistry letters*, 6(12):2326–2331, 2015. ISSN 1948-7185.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

Warren J Hehre, Robert Ditchfield, and John A Pople. Selfconsistent molecular orbital methods. xii. further extensions of gaussiantype basis sets for use in molecular orbital studies of organic molecules. *The Journal of Chemical Physics*, 56(5):2257–2261, 1972.

Bing Huang and O Anatole von Lilienfeld. Communication: Understanding molecular representations in machine learning: The role of uniqueness and target similarity, 2016.

Haoyan Huo and Matthias Rupp. Unified representation for machine learning of molecules and crystals. *arXiv preprint arXiv:1704.06439*, 2017.

Steven Kearnes, Kevin McCloskey, Marc Berndl, Vijay Pande, and Patrick Riley. Molecular graph convolutions: moving beyond fingerprints. *Journal of computer-aided molecular design*, 30(8): 595–608, 2016.

Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

Tien Lam Pham, Hiori Kino, Kiyoyuki Terakura, Takashi Miyake, Koji Tsuda, Ichigaku Takigawa, and Hieu Chi Dam. Machine learning reveals orbital interaction in materials. *Science and technology of advanced materials*, 18(1):756–765, 2017. ISSN 1468-6996.

Yujia Li, Daniel Tarlow, Marc Brockschmidt, and Richard Zemel. Gated graph sequence neural networks. *arXiv preprint arXiv:1511.05493*, 2015.

Ping Luo, Jiamin Ren, and Zhanglin Peng. Differentiable learning-to-normalize via switchable normalization. *arXiv preprint arXiv:1806.10779*, 2018.

Robert S Mulliken. A new electroaffinity scale; together with data on valence states and on valence ionization potentials and electron affinities. *The Journal of Chemical Physics*, 2(11):782–793, 1934. ISSN 0021-9606.

Frank Neese. An improvement of the resolution of the identity approximation for the formation of the coulomb matrix. *Journal of computational chemistry*, 24(14):1740–1747, 2003.

Raghunathan Ramakrishnan, Pavlo O Dral, Matthias Rupp, and O Anatole Von Lilienfeld. Quantum chemistry structures and properties of 134 kilo molecules. *Scientific data*, 1:140022, 2014.

David Rogers and Mathew Hahn. Extended-connectivity fingerprints. *Journal of chemical information and modeling*, 50(5):742–754, 2010.

Matthias Rupp. Machine learning for quantum mechanics in a nutshell. *International Journal of Quantum Chemistry*, 115(16):1058–1073, 2015. ISSN 1097-461X.

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

Bernhard Scholkopf and Alexander J Smola. *Learning with kernels: support vector machines, regularization, optimization, and beyond*. MIT press, 2002. ISBN 0262194759.

Kristof T. Schütt, Farhad Arbabzadah, Stefan Chmiela, Klaus R. Müller, and Alexandre Tkatchenko. Quantum-chemical insights from deep tensor neural networks. *Nature Communications*, 8:13890, 2017. doi: 10.1038/ncomms13890https://www.nature.com/articles/ncomms13890#supplementary-information. URL `http://dx.doi.org/10.1038/ncomms13890`.

KT Schutt, H Glawe, F Brockherde, A Sanna, KR Muller, and EKU Gross. How to represent crystal structures for machine learning: Towards fast prediction of electronic properties. *Physical Review B*, 89(20):205118, 2014.

Marwin H. S. Segler, Mike Preuss, and Mark P. Waller. Planning chemical syntheses with deep neural networks and symbolic ai. *Nature*, 555:604, 2018. doi: 10.1038/nature25978https://www.nature.com/articles/nature25978#supplementary-information. URL `http://dx.doi.org/10.1038/nature25978`.

John C Slater and George F Koster. Simplified lcao method for the periodic potential problem. *Physical Review*, 94(6):1498, 1954.

O Anatole von Lilienfeld, Raghunathan Ramakrishnan, Matthias Rupp, and Aaron Knoll. Fourier series of atomic radial distribution functions: A molecular fingerprint for machine learning models of quantum chemical properties. *International Journal of Quantum Chemistry*, 115(16):1084–1093, 2015. ISSN 1097-461X.

David Weininger. Smiles, a chemical language and information system. 1. introduction to methodology and encoding rules. *Journal of chemical information and computer sciences*, 28(1):31–36, 1988. ISSN 0095-2338.

Bing Xu, Naiyan Wang, Tianqi Chen, and Mu Li. Empirical evaluation of rectified activations in convolutional network. *arXiv preprint arXiv:1505.00853*, 2015.