

A deep learning model for predicting selected organic molecular spectra

Received: 27 May 2023

Accepted: 6 October 2023

Published online: 13 November 2023

 Check for updates

Zihan Zou¹, Yujin Zhang¹✉, Lijun Liang², Mingzhi Wei¹, Jiancai Leng¹, Jun Jiang¹✉, Yi Luo¹✉ & Wei Hu¹✉

Accurate and efficient molecular spectra simulations are crucial for substance discovery and structure identification. However, the conventional approach of relying on the quantum chemistry is cost intensive, which hampers efficiency. Here we develop DetaNet, a deep-learning model combining E(3)-equivariance group and self-attention mechanism to predict molecular spectra with improved efficiency and accuracy. By passing high-order geometric tensorial messages, DetaNet is able to generate a wide variety of molecular properties, including scalars, vectors, and second- and third-order tensors—all at the accuracy of quantum chemistry calculations. Based on this we developed generalized modules to predict four important types of molecular spectra, namely infrared, Raman, ultraviolet-visible, and ¹H and ¹³C nuclear magnetic resonance, taking the QM9S dataset containing 130,000 molecular species as an example. By speeding up the prediction of molecular spectra at quantum chemical accuracy, DetaNet could help progress toward real-time structural identification using spectroscopic measurements.

Molecular spectroscopy, a collection of techniques for capturing the electronic or vibrational ‘fingerprint’ of molecules, is widely used in many scientific fields including physics, chemistry, biology, materials, medicine, food, environmental studies, astronomy, meteorology, chemical engineering and substance discovery. Usually, rapid and accurate theoretical prediction plays an important role in identifying molecular structure from such experimentally measured spectroscopic fingerprints. Until now, quantum chemical simulations have given the most accurate spectral predictions. However, this method requires extensive electronic structure calculations on hundreds of thousands of ‘trial’ molecules, thus limiting the usefulness of quantum chemistry in applications requiring real-time spectral identification. A cost-effective theoretical approach that can precisely predict molecular spectra is highly desirable.

Deep-learning methods have gained increasing popularity for predicting molecular properties^{1,2}, by learning and later predicting

structure–property and structure–spectroscopic relationships from reference data without invoking computationally demanding electronic structure calculations. For instance, a multilayer perceptron and convolutional neural network model has been developed³ for predicting molecular electronic-excitation spectra. A machine learning protocol that uses structural descriptors to rapidly predict amide infrared (IR), ultraviolet-visible (UV-Vis) and electronic circular dichroism spectra of various proteins has also been presented⁴. A machine learning framework has been developed⁵ to probe the electronic state properties based on molecular orbital image. A deep learning approach has been used⁶ to study the complex spectroscopic changes that occur in cross-linked polyethylene pipe. An evolutionary deep-learning algorithm (HDNNP)⁷, whereby the local (within a cutoff radius, R_c) chemical environment has been used in a neural network to predict molecular properties from atomic contributions. Recently, message-passing neural networks^{8–12} (MPNNs) have been developed that have the advantage

¹School of Chemistry and Chemical Engineering, Qilu University of Technology (Shandong Academy of Science), Jinan, China. ²College of Automation, Hangzhou Dianzi University, Hangzhou, China. ³Key Laboratory of Precision and Intelligent Chemistry, School of Chemistry and Materials Science, University of Science and Technology of China, Hefei, China. ⁴Hefei National Laboratory, University of Science and Technology of China, Hefei, China.

⁵Hefei National Research Center for Physical Sciences at the Microscale, University of Science and Technology of China, Hefei, China.

✉e-mail: zhangyujin@qlu.edu.cn; jiangj1@ustc.edu.cn; yiluo@ustc.edu.cn; weihu@qlu.edu.cn

of permutation invariance. They treat atoms and bonds (or interatomic distances) as graphs and propagate messages of atoms and chemical environments in combination. The tensor field neural network model¹³ uses inter-atomic position vectors as features to predict vector properties (for instance, atomic forces) at low computational cost while maintaining rotational and translational equivariance.

Here, we developed a deep equivariant tensor attention network (DataNet, Extended Data Figs. 1 and 2) based on the E(3)-equivariant MPNN framework^{14,15}, to describe molecular properties that are invariant and equivariant to rotation, translation and reflection. By passing higher order tensorial messages between atoms, DataNet uses structural features to rapidly predict scalars (for example, energy and partial charges), vectors (for example, electric dipole and atomic forces), second-order tensors (for example, Hessian matrix, quadrupole moment and polarizability) and third-order tensors (for example, octupole moment and first hyperpolarizability) with high accuracy. Derived from these, DataNet includes modules to predict IR absorption, Raman scattering, UV-Vis absorption, and ¹H and ¹³C nuclear magnetic resonance (NMR) spectra. Based on the widely used QM9 (ref. 16) dataset consisting of 130,000 organic molecules, DataNet predicted UV-Vis spectra with high (92%) accuracy, and IR, Raman, and NMR spectra with over 99% accuracy. Importantly, this method improves the prediction efficiency of spectra by three to five orders of magnitude compared with traditional quantum chemical methods employing density functional theory (DFT). Moreover, the results demonstrate transferability both to molecules with different size and when considering varying environments such as applied electric fields, solvents and surface adsorption, which will further expand the real-time application of DataNet in various spectra.

Results and discussion

Property prediction

Based on 130,000 organic molecules extracted from the QM9 dataset, we performed geometrical re-optimization and spectral simulations to construct the QM9S dataset (see ‘Reference data’ section in Methods). Combining QM9S with the partial QM7-X¹⁷ dataset (20 randomly selected configurations for each molecule from the dynamic sampling), we tested the accuracy with which DataNet predicts molecular equivariant and invariant properties. Note that an ablation experiment should be performed first to test the impact of each module and obtain the optimal parameters for DataNet (Extended Data Fig. 3). We then calculated mean absolute error (MAE) and the root mean square error (RMSE) to assess the error between DataNet and the reference data; coefficient of determination (R^2) was used to indicate prediction accuracy. As shown in Extended Data Fig. 4, DataNet achieves extremely high accuracy for scalars (energy and natural population charge), vectorial properties (atomic forces and electric dipole moment) and second-order tensorial properties (electric quadrupole and polarizability), with R^2 greater than 99.9%. The energy and atomic forces were learned from the QM7-X dataset because these properties play important roles in geometrical optimization and molecular dynamics. Furthermore, we predicted the third-order tensorial properties electric octupole and first hyperpolarizability with R^2 accuracies of 97.03% and 99.40%, respectively. Many other typical quantum chemical properties were also predicted precisely by DataNet using QM9 and QM7-X data (Supplementary Table 1).

Vibrational spectroscopy

Combining the precise equivariant property-prediction capabilities of DataNet with the harmonic oscillator approximation, we developed standard approaches to predict molecular vibrational spectra. The training and prediction of a Hessian matrix in deep learning is difficult because calculating second-order cross-partial derivatives in existing frameworks is computationally expensive. Alternatively, a few previous studies^{18,19} predicted the Hessian matrix by training atomic

forces, but with poor accuracy and at high cost. Here we propose an algorithm for calculating Hessian matrices. As shown in Fig. 1a, the Hessian matrix is divided into an atomic tensor H_i and an interatomic tensor H_{ij} , for which training and prediction are conducted separately, and are calculated as:

$$\mathbf{H}_i = \sum_j \frac{\partial^2 \mathbf{s}_i}{\partial \vec{\mathbf{r}}_j \partial \vec{\mathbf{r}}_i^T} \quad \mathbf{H}_{ij} = \frac{\partial^2 \mathbf{s}_i}{\partial \vec{\mathbf{r}}_j \partial \vec{\mathbf{r}}_i^T} \quad (1)$$

where $\vec{\mathbf{r}}_i$ and $\vec{\mathbf{r}}_j$ represent the coordinates of central atoms i and neighboring atoms j involved in the interaction layers, \mathbf{s}_i is the output scalar properties and T represents the transpose operation.

Figure 1b compares DataNet predictions with DFT calculations for the Hessian matrices. The absolute error of Hessian matrix values is only 0.073 eV Å⁻², achieving an accuracy of 99.94%. The vibrational frequencies ω and normal coordinates P are calculated by diagonalizing the Hessian matrix. Further, the IR and Raman intensities can be calculated based on the partial derivatives of the dipole moment μ and polarizability tensor α with respect to the normal coordinates:

$$I_{\text{IR}} \propto \left(\frac{\partial \mu}{\partial P} \right)^2 I_{\text{Raman}} \propto \frac{(\omega - \omega_{\text{in}})^4}{\omega} \left(45 \left(\frac{\partial \alpha_{\text{iso}}}{\partial P} \right)^2 + 7 \left(\frac{\partial \alpha_{\text{aniso}}}{\partial P} \right)^2 \right) \frac{1}{1 - \exp\left(\frac{-hc\omega}{KT}\right)} \quad (2)$$

where ω_{in} and ω represents the frequency of the incident and scattering light in cm⁻¹, h and K are the Planck and Boltzmann constants, respectively, T is the temperature and c is the speed of light; α_{iso} and α_{aniso} are the isotropic and anisotropic parts of the polarizability. The whole process of calculating the vibrational spectra is illustrated in Extended Data Fig. 5. As shown in Fig. 1c,d, the derivatives of the dipole moment and polarizability predicted by DataNet agree very well with the DFT reference data.

Taking cyclohexanone and 2-methylpyrazine from the evaluation set and caffeine from the QM9S dataset as examples (Fig. 1e), we tested the performance of DataNet’s calculations of their IR and Raman spectra. As shown in Fig. 1f,g, the IR and Raman spectra predicted by DataNet are almost identical to the results of the DFT calculations and fit the experimental spectra well. Spearman coefficients were calculated to assess the accuracy of DataNet’s reproduction of the DFT-simulated vibration spectra. For the 6,500 molecules in the evaluation sets, average Spearman coefficients for the IR and Raman spectra were 99.42% and 99.14%, respectively, demonstrating the high accuracy of DataNet.

UV-Vis and NMR spectroscopy

The standard algorithm for simulating UV-Vis spectra requires estimation of absorption energies and transition dipole moments. We found DataNet performs well in predicting absorption energies, but poorly in oscillator strength as it is based only on ground state structures (Supplementary Section 4). As a result, we used an alternative method to directly predict the broad UV-Vis spectra over the range 1.5–13.5 eV (92–827 nm). As shown in Fig. 2e, the DataNet-predicted UV-Vis spectra of five representative molecules (cyclohexanone, 2-methylpyrazine, hepta-3,5-diyne-2-one, aniline and 5-methoxy-1,3-oxazole-2-carbaldehyde; structures are shown in Fig. 2d) are almost identical to the DFT results. Moreover, we found the MAE for the DataNet-predicted absorption strengths compared to DFT results was less than 0.012 a.u., with an R^2 of approximately 92.04% (Fig. 2a), demonstrating DataNet’s accuracy for simulating UV-Vis spectra.

For NMR spectra, we predicted the isotropic magnetic shielding tensors for hydrogen, carbon, nitrogen and oxygen atoms and the corresponding chemical shift values. As shown in Fig. 2f,g, the DataNet-predicted ¹³C and ¹H NMR spectra of the five representative molecules were extremely close to the DFT reference data. Notably, the MAE of

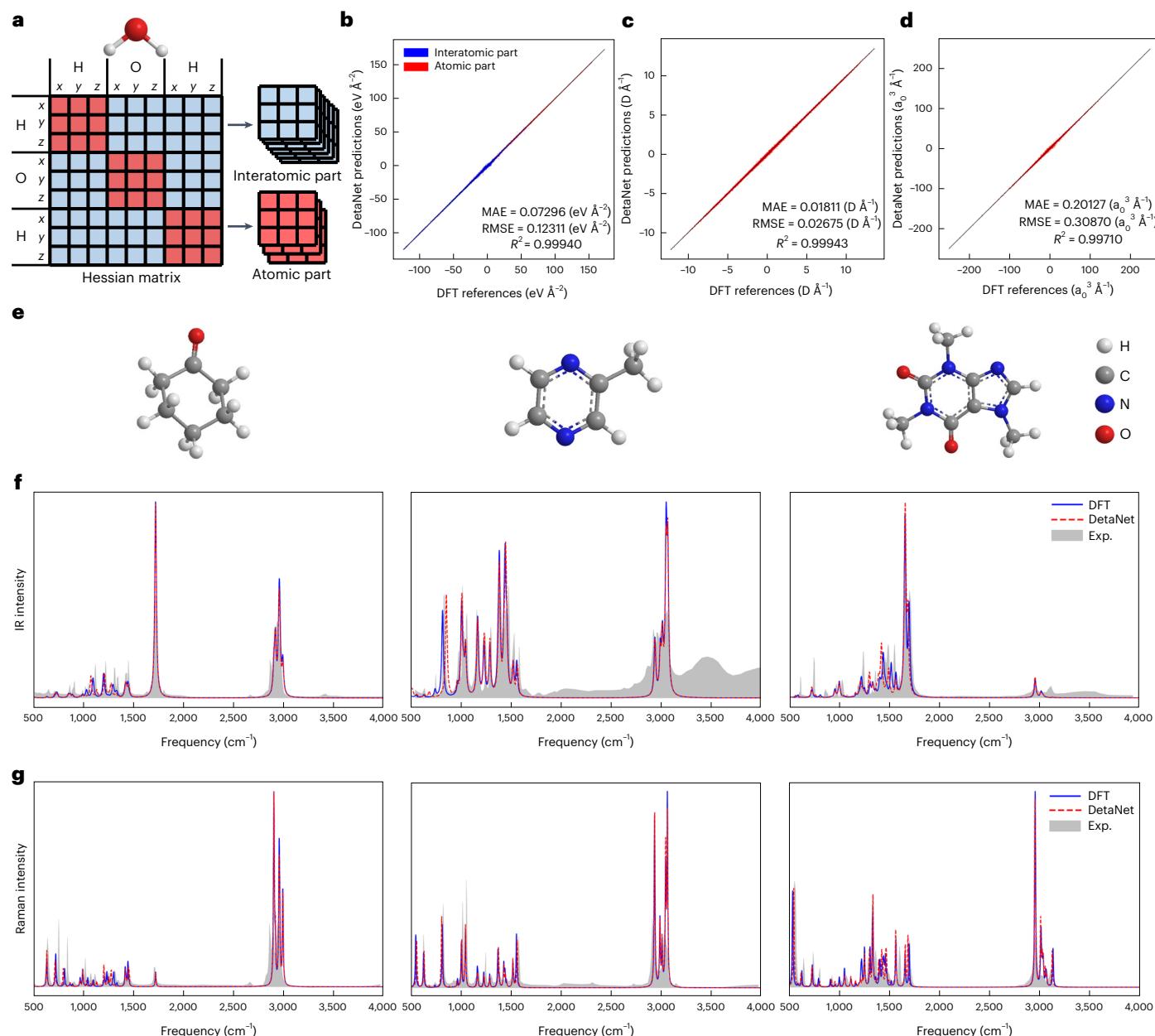


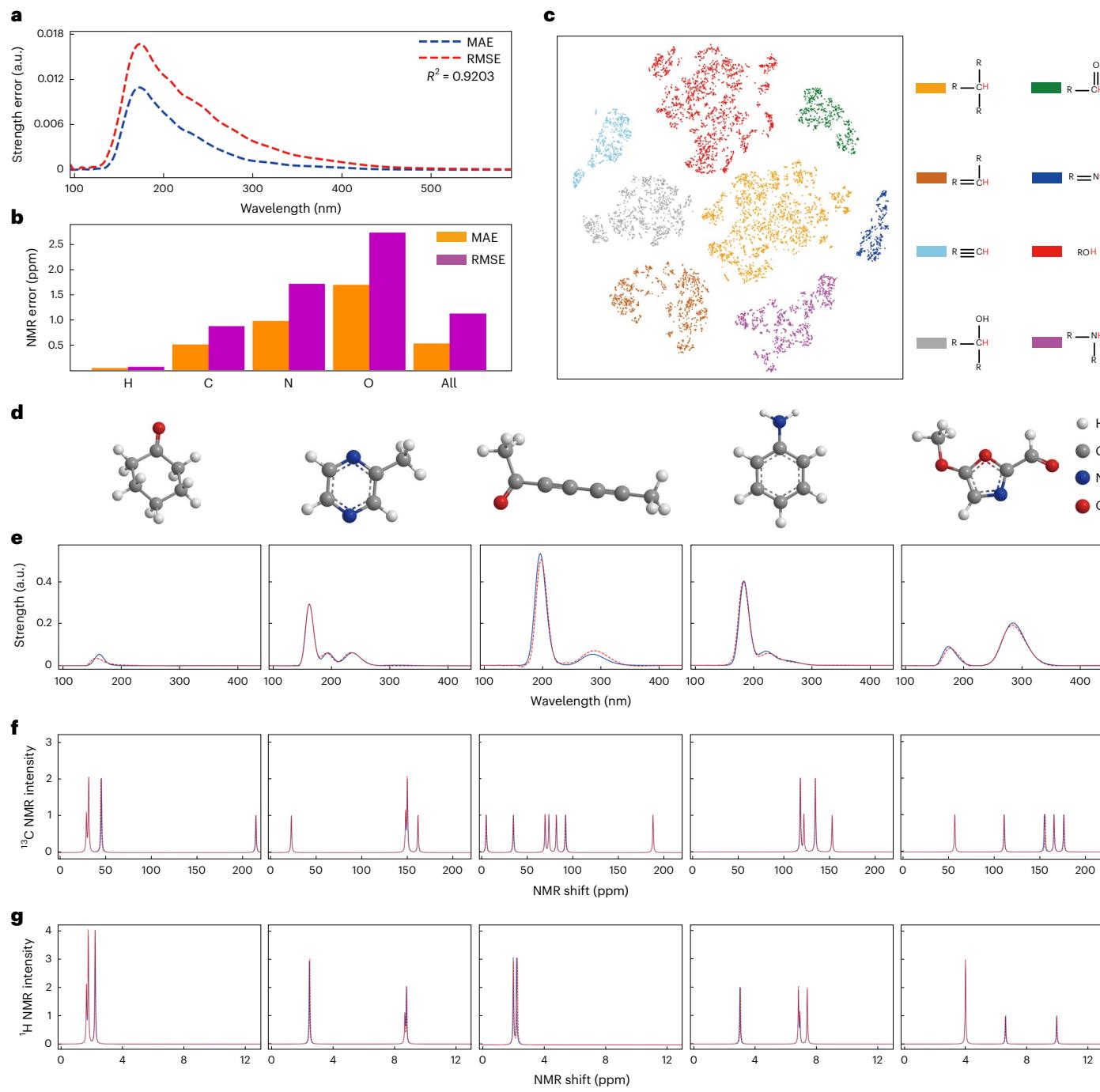
Fig. 1 | DetaNet-predicted IR and Raman spectra and their comparison with DFT reference and experimental data. **a**, Schematic diagram for calculating the Hessian matrix taking water as an example, where x, y and z represent the Cartersian coordinates of each atom. **b**, Regression plots of the DetaNet-predicted atomic and interatomic parts of the Hessian matrix. **c**, Regression plots of the DetaNet-predicted derivatives of dipole moment. **d**, Regression plots of the DetaNet-predicted derivatives of polarizability. **e**, Schematic

structures of cyclohexanone (left), 2-methylpyrazine (middle) and caffeine (right). **f**, Comparison of the DetaNet-predicted IR spectra to DFT simulations and experimental observations for the molecules shown in **e**. **g**, Comparison of the DetaNet-predicted Raman spectra to DFT simulations and experimental observations at a temperature of 298 K and Raman excitation at 500 nm for cyclohexanone (left), 2-methylpyrazine (middle) and Raman excitation at 532 nm for caffeine (right).

the DetaNet-predicted ^{13}C NMR shift was 0.5203 ppm ($R^2 = 0.9997$), much better than the MAE of 1.92 ppm produced previously by FCHL²⁰. Moreover, DetaNet also precisely simulated ^1H NMR spectra, with an MAE of 0.054 ppm ($R^2 = 0.9985$; Fig. 2b). Calculation of t-distributed stochastic neighbor embedding (t-SNE)²¹ of the 128-dimensional latent space into a 2D space for hydrogen atoms in different chemical environments provides a visual assessment of DetaNet's ability to learn NMR spectra. The well-separated and tight clusters shown in Fig. 2c for the 100,000 hydrogen atoms with the eight different hydrogen environments demonstrate that DetaNet can accurately and faithfully simulate the chemical environment and thus precisely predict NMR spectra.

Transferability

We tested DetaNet's transferability by varying molecular types, sizes and geometries. We randomly selected 5,500 organic molecules from the PubChem database²² and found DetaNet and DFT obtain almost identical IR and Raman spectra even for large molecules not included in QM9S. Moreover, we used both implicit (such as electric field or polarizable continuum) and explicit models (such as explicit water environment or adsorbed on metal surface) to introduce molecular deformation and found DetaNet could also reproduce the IR/Raman spectra of distorted molecules (Supplementary Section 5).

**Fig. 2 | Performance of DataNet in predicting UV-Vis and NMR spectra.**

a, Absorption strength prediction error as a function of wavelength. **b**, Prediction error of NMR shielding for different elements. **c**, Visualization of feature vectors of hydrogen atoms in different chemical environments created with t-SNE. **d**, Left to right, schematic structures of cyclohexanone, 2-methylpyrazine, hepta-

3,5-diyne-2-one, aniline and 5-methoxy-1,3-oxazole-2-carbaldehyde. **e**, Comparison of the DataNet-predicted (red) UV-Vis spectra with reference data (blue) for the molecules shown in **d**. **f**, Comparison of the DataNet-predicted ^{13}C NMR spectra with reference data for the molecules shown in **d**. **g**, Comparison of the DataNet-predicted ^1H NMR spectra with reference data for the molecules shown in **d**.

Computational efficiency

We tested the computational efficiency of DataNet on both a CPU (Intel-i7 8700K) and a GPU (NVIDIA RTX 3080Ti). As shown in Extended Data Fig. 6a, the average prediction times for vibrational, UV-Vis and NMR spectra using DataNet on a CPU device were 0.98 s, 0.020 s and 0.029 s, respectively; these are about 10^3 , 10^5 and 10^4 times faster than the traditional DFT (Gaussian 16 B.01; ref. 23) calculations. DataNet efficiency is improved further when the GPU is used, as shown in Extended Data Fig. 6a. Furthermore, the batch prediction strategy

used in DataNet greatly accelerates simulation efficiency. For example, the average times for simulating one vibrational spectrum in a batch of 16 are 0.23 s and 0.059 s on the CPU and GPU devices, respectively; these times are fourfold and eightfold faster than the simulation time for a single molecule. Moreover, the time consumed by DataNet is merely linearly proportional to the number of atoms contained in the molecule (shown as the blue line in Extended Data Fig. 6b). By contrast, the computational cost of a conventional DFT calculation scales cubically with the number of electrons. Thus, DataNet has a

great advantage when simulating spectra of large molecules compared with DFT calculations.

Overall, DetaNet accelerates the property calculation and spectra simulations by three to five orders compared with traditional quantum chemistry, indicating its potential application in high-throughput screening processes such as catalyst design and drug discovery. However, the usage of DFT results as the benchmark also limits the accuracy of DetaNet. Moreover, DetaNet predicts molecular properties and spectra based on the static geometries and harmonic approximation, resulting the anharmonic effects are ignored. Even though, the vector and high-order tensor prediction capability of DetaNet allow to develop the deep-learning models to simulate some other spectra such as x-ray spectra, electronic circular dichroism, phosphorescence and fluorescence spectra. Furthermore, we could also develop deep-learning methods for the molecular geometrical optimization, transition state search and molecular dynamic simulations based on the DetaNet-predicted energy, atomic force and Hessian matrix.

Methods

Reference data

We constructed the QM9S dataset using 130,000 organic molecules based on the QM9 dataset. First, we re-optimized molecular geometries using the Gaussian16 package (B.01 version)²³ at the B3LYP/def-TZVP level of theory^{24,25}. Then the molecular properties including scalars (for example, energy, partial charges), vectors (for example, electric dipole), second-order tensors (for example, Hessian matrix, quadrupole moment, polarizability) and third-order tensors (for example, octupole moment, first hyperpolarizability) were calculated at the same level. The frequency analysis and time-dependent DFT were carried out at the same level to obtain the IR, Raman and UV-Vis spectra. All these calculated molecular properties and spectra are collected in the QM9S dataset.

Apart from the QM9S, we used other reference data from the literature and databases. For example, we the QM9-NMR dataset²⁶ developed to train the ¹H and ¹³C NMR prediction module. The QM7-X¹⁷ dataset including molecular dynamics sampling structures for 7,211 molecules was used to enable DetaNet to accurately learn the geometries that deviate from global or local minima. Here only 20 configurations randomly selected from the molecular dynamics sampling are used for each molecule. For every dataset, we selected 90%, 5% and 5% of molecules as the training, test and evaluation sets, respectively.

E(3) equivariance

Given a set of transformations that act on a vector space X as $T_g : X \rightarrow X$ to which we associate an abstract group G , a function $\phi_e : X \rightarrow Y$ is said to be equivariant with respect to G if

$$\phi_e(T_g(\mathbf{x})) = S_g(\phi_e(\mathbf{x})) \quad g \in G, \mathbf{x} \in X \quad (4)$$

where $S_g : Y \rightarrow Y$ is an equivalent transformation on the output space. By contrast, if the function has non-transformation on its output space as follows:

$$\phi_e(T_g(\mathbf{x})) = \phi_e(\mathbf{x}) \quad g \in G, \mathbf{x} \in X \quad (5)$$

we say that the function is invariant to G . In this work, we focus on the equivariance and invariance in 3D Euclidean space of operations, namely rotations, translations and reflections, that form the E(3) group.

Equivariant linear layer

All linear layers used in DetaNet are equivariant and can be expressed as:

$$L(\mathbf{x}_{l,p}^m) = \begin{cases} \varphi(W\mathbf{x} + b) & l = 0, p = 1 (\text{scalar}) \\ W_{l,p}\mathbf{x}_{l,p}^m & \text{else} \end{cases} \quad (6)$$

where $\mathbf{x}_{l,p}^m \in \mathbb{R}^{(2l+1) \times F_{in}}$ is any input feature with degree l and parity p ; \mathbb{R} is the set of real numbers; $W, W_{l,p} \in \mathbb{R}^{F_{in} \times F_{out}}$ and $b \in \mathbb{R}^{F_{out}}$ are the learnable weight matrices and biases; φ is a non-linear activation function; and F_{in} and F_{out} are the dimensions of the input and output features. $W_{l,p}$ varies depending on the degree l and parity p , while sharing the weight matrices of different order m of the same degree l and parity p . The bias b and activation function φ are non-equivariant and act only on scalar but not on irreps features. Further, the first layer of the embedding and radial embedding modules does not use bias and activation functions and is only used to map feature dimensions.

A multilayer perceptron (MLP) was used at the end of the model to obtain the molecular property outputs:

$$\text{MLP}(\mathbf{x}) = L_{NI}(\dots L_2(L_1(\mathbf{x}))) \quad (7)$$

where NI is the number of linear layers in the MLP. The last layer of the MLP is used for output properties; thus no bias or activations are included.

DetaNet architecture

DetaNet takes as inputs atomic numbers $\{Z_1, \dots, Z_N | Z_i \in \mathbb{N}\}$ and Cartesian coordinates $\{\mathbf{r}_1, \dots, \mathbf{r}_N | \mathbf{r}_i \in \mathbb{R}^3\}$ specifying the element identity and position of each atom in a molecule. Similar to previous MPNNs, DetaNet builds structure–property relationships using interaction layers that are defined in terms of ResNet-style message functions and update functions (Extended Data Fig. 1a). Unlike most previous MPNNs, DetaNet's atomic feature for node i at hidden state n is expressed as an irreducible representation (irreps) tensor $\mathbf{T}_{i,l,p}^{n,m}$, where the integers $l \in 0, 1, 2, 3 \dots$ and $m \in [-l, l]$ specify the rotation degree and order of the irreps tensor, and $p \in [-1, 1]$ indicates even or odd parity. The special case when $l = 0$ and $p = 1$ represents the scalar features \mathbf{S}_i^0 that are treated separately in DetaNet. Initially the atomic irreps features $\mathbf{T}_{i,l,p}^{0,m}$ are set to zero and the initial scalar features \mathbf{S}_i^0 are obtained from the embedding layer (Extended Data Fig. 1b):

$$\mathbf{S}_i^0 = L_{emb}(L_z(O(Z_i)) + L_Q(Q(Z_i))) \quad (8)$$

where $O(Z_i)$ is the commonly used one-hot feature, representing nuclear types. $Q(Z_i)$ represents the inherent electronic structure of an isolated atom (Supplementary Table 3). The learnable linear layers L_z and L_Q map the nuclear and electronic features to the F -dimensional atomic features, while L_{emb} integrates the two parts.

As shown in Extended Data Fig. 1c, the message module generates the atomic message for atom i using the \mathbf{S}_i^n from the upper layer and the position vector $\vec{\mathbf{r}}_{ij} = \mathbf{r}_j - \mathbf{r}_i$ of all neighboring atoms j within an R_c . To be specific, the absolute distance $r = \|\vec{\mathbf{r}}_{ij}\|$ is embedded as two different radial weights w_k and w_v (Extended Data Fig. 1d) for the next self-attention module. Then the edge features e_{ij} are constructed through a self-attention mechanism²⁷ (Extended Data Fig. 1e):

$$\begin{aligned} \mathbf{Mq} &= L_{Mq}(\mathbf{S}_i^n) \quad \mathbf{Mk} = L_{Mk}(\mathbf{S}_j^n) \cdot w_k \quad \mathbf{Mv} = L_{Mv}(\mathbf{S}_j^n) \cdot w_v \\ \mathbf{e}_{ij} &= L_{out}\left(\text{Softmax}\left(\frac{\mathbf{Mq} \times \mathbf{Mk}^T}{\sqrt{F^{Mq}}}\right) \times \mathbf{Mv}\right) \end{aligned} \quad (9)$$

where Mq , Mk and Mv represent the query, key and value features of the self-attention mechanism; L_{Mq} , L_{Mk} , L_{Mv} and L_{out} are learnable linear layers; and F^{Mq} is the dimension of Mq . The self-attention mechanism module reinforces the feature of the central atom \mathbf{S}_i^n and the radial interaction from neighboring atoms in the form of a continuous filter¹⁰ $L(\mathbf{S}_j^n) \cdot w(\|\vec{\mathbf{r}}_{ij}\|)$. Subsequently, the output of the message layer is obtained by invariant and equivariant convolution:

$$\Delta_M \mathbf{S}_i^n = L_S\left(\sum_j \mathbf{e}_{ij}\right) \quad \Delta_M \mathbf{T}_{i,l,p}^{n,m} = L_T\left(\sum_j \mathbf{e}_{ij} \otimes Y_{l,p}^m\left(\frac{\vec{\mathbf{r}}_{ij}}{\|\vec{\mathbf{r}}_{ij}\|}\right)\right) \quad (10)$$

where $\Delta_M \mathbf{S}_i^n$ and $\Delta_M \mathbf{T}_{i,l,p}^{n,m}$ represent the residuals for the \mathbf{S}_i^n and $\mathbf{T}_{i,l,p}^{n,m}$ features of the upper layer, and L_s and L_T indicate linear layers. $Y_{l,p}^m$ is the spherical harmonic function that maps vectors into primitives of the irreps tensor and performs a tensor product with edge feature e_{ij} to produce the irreps features. After passing the interatomic message, an equivariant update module is used to describe the interaction between the features with different degrees and parity. Similarly, we use the attention mechanism to obtain the residuals of equivariant and invariant features (Extended Data Fig. 1f):

$$\begin{aligned} A &= L_a (\sum_{l=0,p=1} Uq \otimes Uk) \\ \Delta_U \mathbf{S}_i^n &= \text{Softmax}(A) \cdot UV_{(S)} \quad {}_U \mathbf{T}_{i,l,p}^{n,m} = \text{Softmax}(A) \otimes UV_{(T)} \end{aligned} \quad (11)$$

where Uq , Uk and $UV_{(T)}$ are generated by adding $\mathbf{T}_{i,l,p}^{n,m}$ and $\Delta_M \mathbf{T}_{i,l,p}^{n,m}$ from the message layer via linear layers. $UV_{(S)}$ is generated by adding \mathbf{S}_i^n and $\Delta_M \mathbf{S}_i^n$ from the message layer via a linear layer. A is an attention feature produced as a scalar obtained by the tensor product of two linear combinations of irreps features through the softmax function. The residual result of the update process is obtained by multiplying the attention features with the linear combination of the irreps and scalar feature.

After reaching the maximum interaction layer N , an MLP is used to output the properties:

$$\begin{aligned} \mathbf{s}_i &= \text{MLP}(\mathbf{S}_i^N) \quad \mathbf{t}_{i,l,p}^m = \text{MLP}(\mathbf{T}_{i,l,p}^{N,m}) \\ y &= \sum_i \phi(\mathbf{s}_i, \mathbf{t}_{i,l,p}^m) \end{aligned} \quad (12)$$

where ϕ is an output function that produces a particular molecular property.

Translation equivariant radial embedding module

The radial embedding module (Extended Data Fig. 1d) takes the absolute distances $r = \|\mathbf{r}_{ij}\|$ between each atom i and its neighbouring atoms j as inputs. Then radial basis functions map distance r to high-dimensional sparse features $B(r)$ scaled between [0,1]. Finally, the radial weights w_k and w_v used in the self-attention mechanism are generated from the learnable linear layers. All interatomic distances share the same linear layer; consequently the module is 2D translation equivariant. To express the $B(r)$ features, we modified the learnable Bessel radial function¹² as:

$$B(r) = \sqrt{\frac{2}{R_c}} \times \frac{\sin\left(\frac{b \cdot \pi \cdot r}{R_c}\right)}{a \cdot r} \quad (13)$$

where r is the distance between two atoms, R_c is the cut-off radius to ensure the radial basis decays smoothly to 0 as the distance increases, a and b are learnable parameters initially set to 2.0, and $[1, 2, \dots, N_b]$ (N_b is the number of radial basis functions). Since the Bessel function contains R_c itself, we do not use any additional cut-off functions.

Property prediction functions

DataNet applies various functions to predict molecular properties. We have demonstrated through ablation experiments that these functions are precise for the prediction of molecular properties (Extended Data Figs. 3 and 4). The strategies for predicting these properties can be summarized as:

- (1) Scalar properties. Each single-atom scalar property is calculated directly from the appropriate scalar feature, while whole-molecule scalar properties are predicted by summing up single-atom contributions. For example, atomic charges are predicted from their scalar outputs and

molecular energy is calculated from single-atom contributions as: $E = \sum_i E_i$.

- (2) Vectorial properties. Vector properties that relate to the whole molecule are predicted by considering both local and non-local contributions. For example, the dipole moment is calculated as: $\mu = \sum_i \mathbf{t}_{i,1,-1}^m + \mathbf{s}_i \times \vec{\mathbf{r}}_{ia}$, where $\vec{\mathbf{r}}_{ia} = \mathbf{r}_i - \mathbf{r}_a$ represents the vector from the center-of-mass to atom i and the $\mathbf{s}_i \times \vec{\mathbf{r}}_{ia}$ represent non-local contributions; each $\mathbf{t}_{i,1,-1}^m$ can be regarded as the local dipole moment contributed by the redistribution of electrons near that atom²⁸. The atomic force vector is obtained by taking the first-order derivative of the predicted energy with respect to the atomic coordinates¹⁸: $\mathbf{f}_i = -\frac{\partial E}{\partial \mathbf{r}_i}$.
- (3) Second-order tensorial properties. According to the group of theory, a second-order Cartesian tensor in 3D space can be generated from the 0th-order subspace (the tensor trace) and the second-degree subspace (termed the traceless tensor). For example, we calculated polarizability as: $\alpha = \sum_i \mathbf{s}_i^{(1)} \oplus (\mathbf{t}_{i,2,1}^m + \mathbf{s}_i^{(2)} \times Y_{2,1}^m(\vec{\mathbf{r}}_{ia}))$. Here \oplus denotes direct sum. The learnable $\mathbf{s}_i^{(1)}$ represents the trace of the tensor (the isotropic component). The second-degree irreps tensor $\mathbf{t}_{i,2,1}^m$ can be regarded as consisting of local contributions to the traceless tensor from every single atom. The other learnable $\mathbf{s}_i^{(2)}$ together with the spherical harmonics $Y_{2,1}^m(\vec{\mathbf{r}}_{ia})$ form the non-local traceless tensor. Expressions for calculating the quadrupole moment and inertia tensor moment were generated similarly.
- (4) Third-order tensorial properties. According to the group of theory, third-order tensors in 3D space can be generated from the first-degree and third-degree subspaces. For example, the first-order hyperpolarizability is predicted as: $\beta = \sum_i (\mathbf{t}_{i,1,-1}^m + \mathbf{s}_i^{(1)} \times \vec{\mathbf{r}}_{ia}) \oplus (\mathbf{t}_{i,3,-1}^m + \mathbf{s}_i^{(2)} \times Y_{3,-1}^m(\vec{\mathbf{r}}_{ia}))$. Here the irreps tensors $\mathbf{t}_{i,1,-1}^m$ and $\mathbf{t}_{i,3,-1}^m$ represent local contributions. Scalars together with spherical harmonics of center-of-mass vector generate the non-local part. A similar strategy is performed to create an expression for the octupole moment.
- (5) NMR. The shielding tensor σ_i is an asymmetric tensor that represents the chemical environment around each atom and is calculated as: $\sigma_i = \mathbf{s}_i \oplus \mathbf{t}_{i,1,1}^m \oplus \mathbf{t}_{i,2,1}^m$. Here the trace \mathbf{s}_i and traceless tensor $\mathbf{t}_{i,2,1}^m$ form the symmetric part of the tensor and the pseudo-vector $\mathbf{t}_{i,1,1}^m$ forms the asymmetric part. Since the NMR shifts are determined only by the isotropic values of the shielding tensor, it is also possible to obtain the NMR shifts by directly predicting the traces of the tensor $\sigma_i^{\text{iso}} = \mathbf{s}_i$.
- (6) Dipole and polarizability derivatives. DataNet provides a direct calculation module for position-related derivative properties. For example, derivatives of the dipole moment $\frac{\partial \mu}{\partial \mathbf{r}_i}$ and polarizabilities $\frac{\partial \alpha}{\partial \mathbf{r}_i}$ are derived from the output dipole moments $\vec{\mu}$ and polarizabilities α with respect to the input coordinates $\vec{\mathbf{r}}_i$.

Activations

The learnable sigmoid linear unit²⁹ (also known as Swish) was selected as the non-linear activation function to learn the non-linear mapping between feature spaces, defined as:

$$\text{Swish}(\mathbf{x}) = \frac{\gamma_1 \mathbf{x}}{1 + e^{-\gamma_2 \mathbf{x}}} \quad (14)$$

where $\gamma_1, \gamma_2 \in \mathbb{R}^{F_{\text{in}}}$ are learnable parameters and were initially set to 1.0 and 1.702 according to SpookyNet¹¹ and other previous works.

The softmax activation function in the attention mechanism was defined as:

$$\text{Softmax}(\mathbf{x}) = \frac{\exp(\mathbf{x})}{\sum \exp(\mathbf{x})} \quad (15)$$

Training hyperparameters

All DetaNet models in this work use $N=3$ interaction layers, $F=128$ features of each order m (total 2,048 features) and $R_c=5.0\text{ \AA}$, unless otherwise specified. The degree of the irreps features is set to $l=1, 2, 3$. A Bessel function with $N_b=32$ is used as the radial basis function (Extended Data Fig. 2).

We optimized DetaNet with Adam³⁰ and the AMSGrad³¹ variant implemented in PyTorch³². The dataset was divided into batches (64 molecules for normal properties and 16 molecules for the derivative properties) with the initial learning rate set to 10^{-3} . Tests were performed every 50 epochs of training and the learning rate was reduced by 50% if the loss did not decrease. Training ended when the loss dropped to 10^{-5} or upon reaching 1,000,000 epochs. All training was performed using the mean squared error (also known as l2loss) as the loss function. For training of the NMR shift, we first pre-trained a model including all chemical elements at once and then subsequently trained by including only one of the elements H, C, N or O at a time.

Data availability

All datasets used in this document are publicly available. The QM9S dataset including optimized structures, various properties, and IR, Raman, and UV-Vis spectra of the 133,885 molecules is available at figshare³³: <https://doi.org/10.6084/m9.figshare.24235333> or code ocean for reference³⁴: <https://doi.org/10.24433/CO.5808137.v3>. The original QM9 dataset¹⁶ is available from <http://quantum-machine.org/datasets/>. We used both the gas and solvent phase NMR dataset²⁶ obtained at mPW1PW91/6-311+G(2d,p) level, which is available at <https://moldis.tiffr.res.in/data/QM9NMR>. The QM7-X dataset¹⁷ is available from <https://zenodo.org/record/3905361>. Infrared and Raman experimental spectra for comparison are from Spectral Database for Organic Compounds³⁵ (SDBS) at <https://sdbs.db.aist.go.jp>. The experimental Raman spectrum of caffeine comes from rruff database³⁶ at <https://rruff.info/Ca/D120006>. Source data are provided with this paper.

Code availability

The DetaNet model and trained parameters are available from Code Ocean³⁴. The program used for spectrum prediction based on DetaNet is also included in the code. All code is written using the PYG³⁷ and e3nn³⁸ libraries on Python 3.9.

References

- Chen, D., Wang, Z., Guo, D., Orekhov, V. & Qu, X. Review and prospect: deep learning in nuclear magnetic resonance spectroscopy. *Chem. Eur. J.* **26**, 10391–10401 (2020).
- Qu, X. et al. Accelerated nuclear magnetic resonance spectroscopy with deep learning. *Angew. Chem.* **132**, 10383–10386 (2020).
- Ghosh, K. et al. Deep learning spectroscopy: neural networks for molecular excitation spectra. *Adv. Sci.* **6**, 1801367 (2019).
- Ye, S. et al. A machine learning protocol for predicting protein infrared spectra. *J. Am. Chem. Soc.* **142**, 19071–19077 (2020).
- Chen, Z. & Yam, V. W.-W. Machine-learned electronically excited states with the MolOrbImage generated from the molecular ground state. *J. Phys. Chem. Lett.* **14**, 1955–1961 (2023).
- Grossutti, M. et al. Deep learning and infrared spectroscopy: representation learning with a β -variational autoencoder. *J. Phys. Chem. Lett.* **13**, 5787–5793 (2022).
- Behler, J. Atom-centered symmetry functions for constructing high-dimensional neural network potentials. *J. Chem. Phys.* **134**, 074106 (2011).
- Schütt, K. T., Arbabzadah, F., Chmiela, S., Müller, K. R. & Tkatchenko, A. Quantum-chemical insights from deep tensor neural networks. *Nat. Commun.* **8**, 13890 (2017).
- Schütt, K., Unke, O. & Gastegger, M. Equivariant message passing for the prediction of tensorial properties and molecular spectra. In *Int. Conf. Machine Learning* 9377–9388 (PMLR, 2021).
- Schütt, K. T., Sauceda, H. E., Kindermans, P.-J., Tkatchenko, A. & Müller, K.-R. SchNet—a deep learning architecture for molecules and materials. *J. Chem. Phys.* **148**, 241722 (2018).
- Unke, O. T. et al. SpookyNet: learning force fields with electronic degrees of freedom and nonlocal effects. *Nat. Commun.* **12**, 7273 (2021).
- Gasteiger, J., Groß, J. & Günnemann, S. Directional message passing for molecular graphs. Preprint at <https://doi.org/10.48550/arXiv.2003.03123> (2020).
- Thomas, N. et al. Tensor field networks: rotation-and translation-equivariant neural networks for 3D point clouds. Preprint at <https://doi.org/10.48550/arXiv.1802.08219> (2018).
- Batzner, S. et al. E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nat. Commun.* **13**, 2453 (2022).
- Atz, K., Grisoni, F. & Schneider, G. Geometric deep learning on molecular representations. *Nat. Mach. Intell.* **3**, 1023–1032 (2021).
- Ramakrishnan, R., Dral, P. O., Rupp, M. & Von Lilienfeld, O. A. Quantum chemistry structures and properties of 134 kilo molecules. *Sci. Data* **1**, 140022 (2014).
- Hoja, J. et al. QM7-X, a comprehensive dataset of quantum-mechanical properties spanning the chemical space of small organic molecules. *Sci. Data* **8**, 43 (2021).
- Zubatyuk, R., Smith, J. S., Leszczynski, J. & Isayev, O. Accurate and transferable multitask prediction of chemical properties with an atoms-in-molecules neural network. *Sci. Adv.* **5**, eaav6490 (2019).
- Chen, W.-K., Zhang, Y., Jiang, B., Fang, W.-H. & Cui, G. Efficient construction of excited-state Hessian matrices with machine learning accelerated multilayer energy-based fragment method. *J. Phys. Chem. A* **124**, 5684–5695 (2020).
- Christensen, A. S., Bratholm, L. A., Faber, F. A. & Anatole von Lilienfeld, O. FCHL revisited: faster and more accurate quantum machine learning. *J. Chem. Phys.* **152**, 044107 (2020).
- Van der Maaten, L. & Hinton, G. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **9**, 106005 (2008).
- Kim, S. et al. PubChem 2023 update. *Nucleic Acids Res.* **51**, D1373–D1380 (2023).
- Frisch, M. et al. Gaussian 16, Revision B. 01 (Gaussian, 2016).
- Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77**, 3865 (1996).
- Schäfer, A., Huber, C. & Ahlrichs, R. Fully optimized contracted Gaussian basis sets of triple zeta valence quality for atoms Li to Kr. *J. Chem. Phys.* **100**, 5829–5835 (1994).
- Gupta, A., Chakraborty, S. & Ramakrishnan, R. Revving up ^{13}C NMR shielding predictions across chemical space: benchmarks for atoms-in-molecules kernel machine learning with new data for 134 kilo molecules. *Mach. Learn. Sci. Technol.* **2**, 035010 (2021).
- Vaswani, A. et al. Attention is all you need. In *Advances in Neural Information Processing Systems* Vol. 30 (2017).
- Veit, M., Wilkins, D. M., Yang, Y., DiStasio, R. A. Jr. & Ceriotti, M. Predicting molecular dipole moments by combining atomic partial charges and atomic dipoles. *J. Chem. Phys.* **153**, 024113 (2020).
- Elfwing, S., Uchibe, E. & Doya, K. Sigmoid-weighted linear units for neural network function approximation in reinforcement learning. *Neural Networks* **107**, 3–11 (2018).
- Kingma, D. P. & Ba, J. Adam: a method for stochastic optimization. Preprint at <https://doi.org/10.48550/arXiv.1412.6980> (2014).
- Reddi, S. J., Kale, S. & Kumar, S. On the convergence of adam and beyond. Preprint at <https://doi.org/10.48550/arXiv.1904.09237> (2019).
- Paszke, A. et al. PyTorch: an imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems* Vol. 32 (2019).

33. Zihan Zou, W. H. QM9S dataset. Figshare <https://doi.org/10.6084/m9.figshare.24235333> (2023).
34. Deep equivariant tensor attention Network(DataNet). Code Ocean <https://doi.org/10.24433/CO.5808137.v3> (2023).
35. Saito, T. et al. *Spectral Database for Organic Compounds (SDBS)* (National Institute of Advanced Industrial Science and Technology, 2006).
36. Lafuente, B., Downs, R. T., Yang, H. & Stone, N. in *Highlights in Mineralogical Crystallography* (eds. Armbruster, T. & Danisi, R. M.) Ch. 1, 1–30 (De Gruyter, 2015).
37. Fey, M. & Lenssen, J. E. Fast graph representation learning with PyTorch Geometric. Preprint at <https://doi.org/10.48550/arXiv.1903.02428> (2019).
38. Geiger, M. & Smidt, T. e3nn: Euclidean neural networks. Preprint at <https://doi.org/10.48550/arXiv.2207.09453> (2022).

Acknowledgements

We acknowledge grants from the National Natural Science Foundation of China (22073053 (W.H.), 22025304 (J.J.), 22033007 (J.J.)), the Young Taishan Scholar Program of Shandong Province (tsqn201909139 (W.H.)), the Natural Science Foundation of Shandong Province (ZR2023MA089 (Y.Z.)), the Program of New Collegiate 20 Items in Jinan (2021GXRC042 (W.H.), 202228031 (Y.Z.)) and the Qilu University of Technology (Shandong Academy of Sciences) Basic Research Project of Science, Education and Industry Integration Pilot (2023PY046 (Y.Z.)). The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Author contributions

Z.Z. and W.H. conceived the research, designed the DataNet model and performed all data analysis. W.H., Y.L., J.J. and Y.Z. jointly supervised the work from the model design to data analysis. Y.Z., L.L., M.W. and J.L. interpreted the data. All authors contributed to the writing of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s43588-023-00550-y>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s43588-023-00550-y>.

Correspondence and requests for materials should be addressed to Yujin Zhang, Jun Jiang, Yi Luo or Wei Hu.

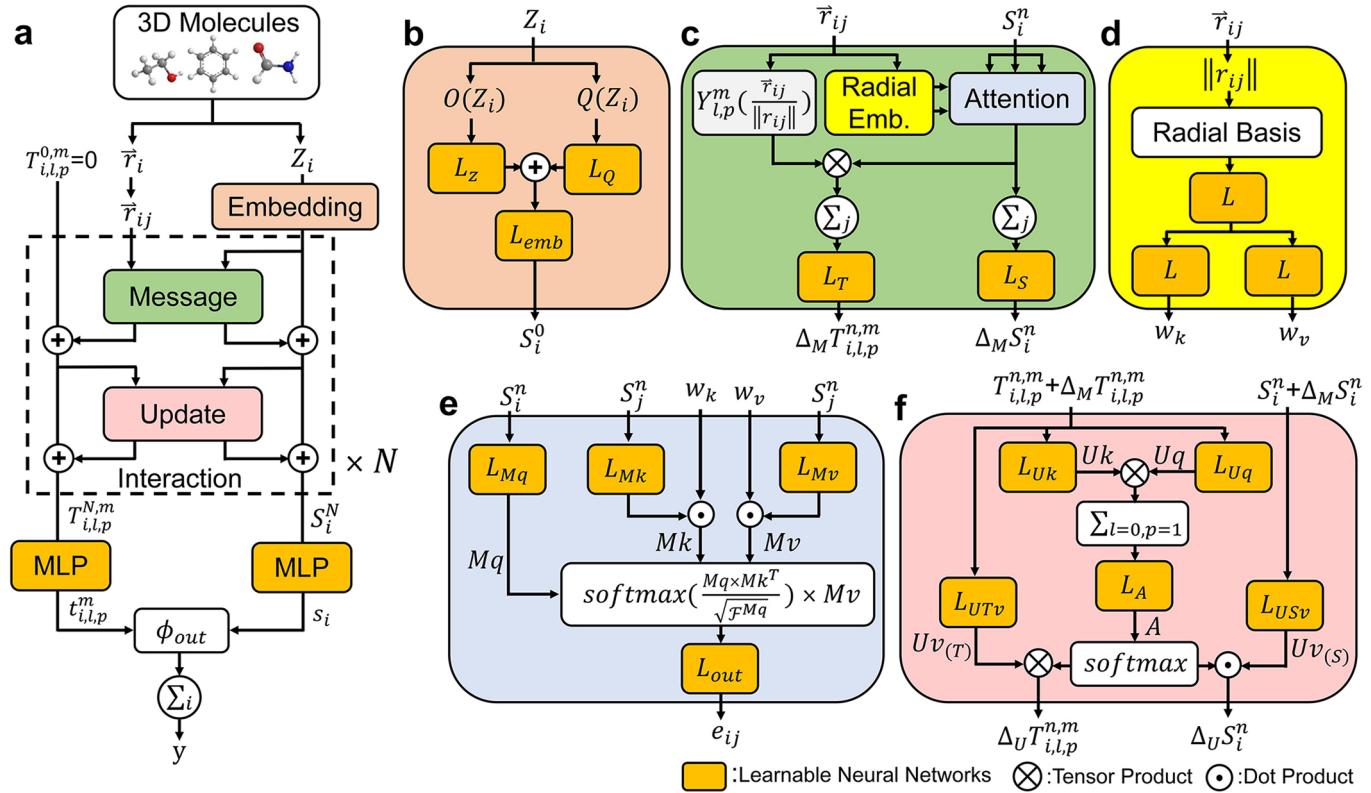
Peer review information *Nature Computational Science* thanks Conrard Tetsassi Feugmo, Feng Wang and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editor: Kaitlin McCurdle, in collaboration with the *Nature Computational Science* team.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2023

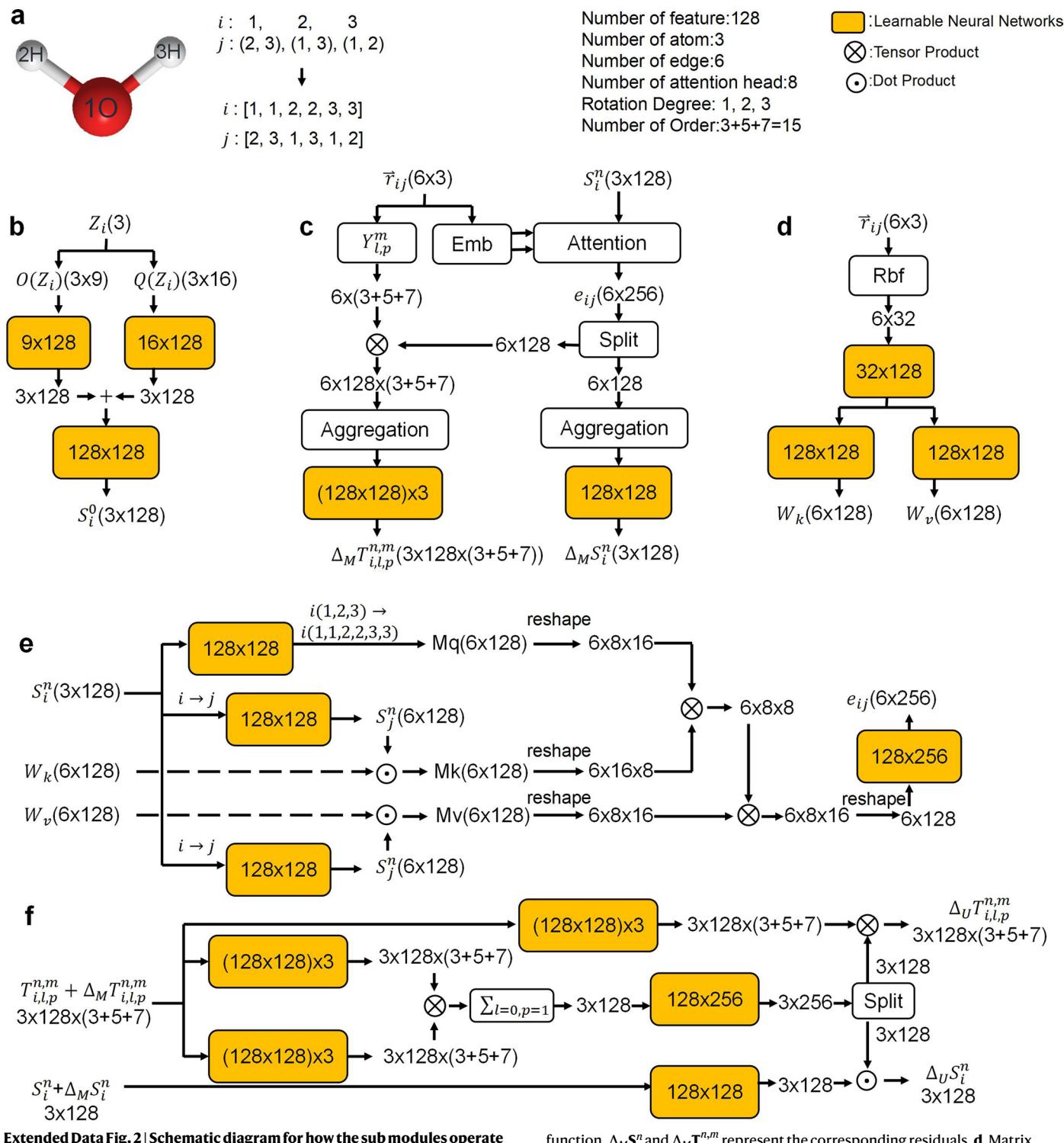


Extended Data Fig. 1 | Schematic diagram of DetaNet’s architecture with a color-coded view of individual components. a. Architecture overview.

Irreducible representation features ($T_{i,l,p}^{n,m}$) and scalar features (S_i^n) are used as messages in the interaction layer, where n, p, i, l and m represent the order of the passing interaction layer, the even/odd parity, the atomic number, the rotation degree and order. N and \vec{r}_{ij} is the maximum interaction layer and the position vector from atom j to i . $t_{i,l,p}^m$ and s_i is output irreps tensor and scalar. **b.** Atomic embedding module based on nuclear and electronic features, where $O(Z_i)$ and $Q(Z_i)$ represent the nuclear types and the inherent atomic electronic structure.

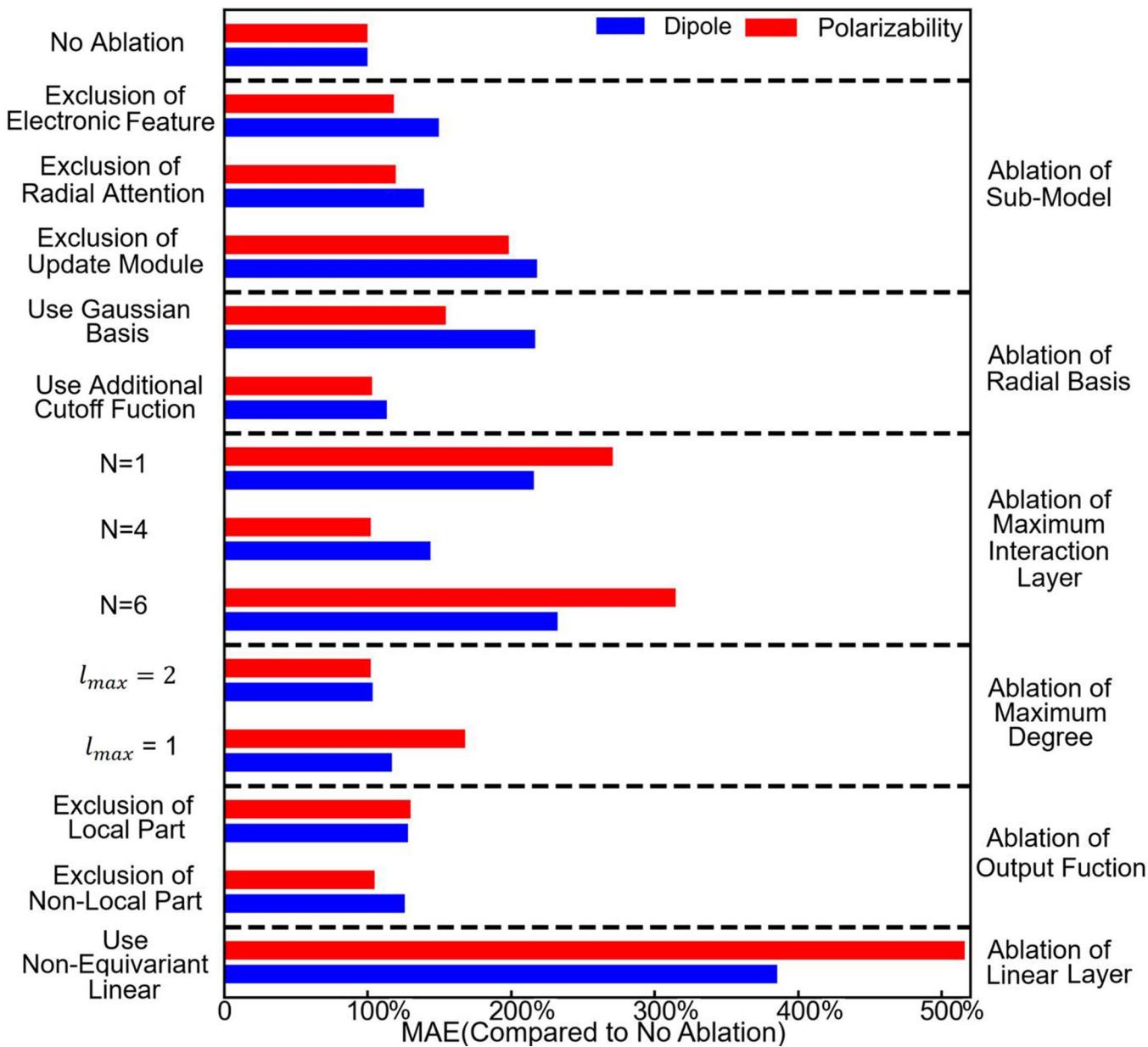
L_z, L_Q and L_{emb} are linear layers to integrate a F -dimensional atomic features.

c. Message module architecture. $Y_{l,p}^m$ is the spherical harmonic function. $\Delta_M S_i^n$ and $\Delta_M T_{i,l,p}^{n,m}$ represent the corresponding residuals. **d.** A radial embedding module to generate the key (w_k) and the value (w_v) weights for the next self-attention module. **e.** A radial self-attention module. Mq , Mk and Mv represents the query, key and value features. F^{Mq} and e_{ij} represent the dimension of Mq and the output edge features. **f.** Atomwise self-attention update module. Uq , Uk , and $Uv_{(T)}$ are query, key and value features of the update module. All symbol of L with subscript indicate the learnable linear layers.



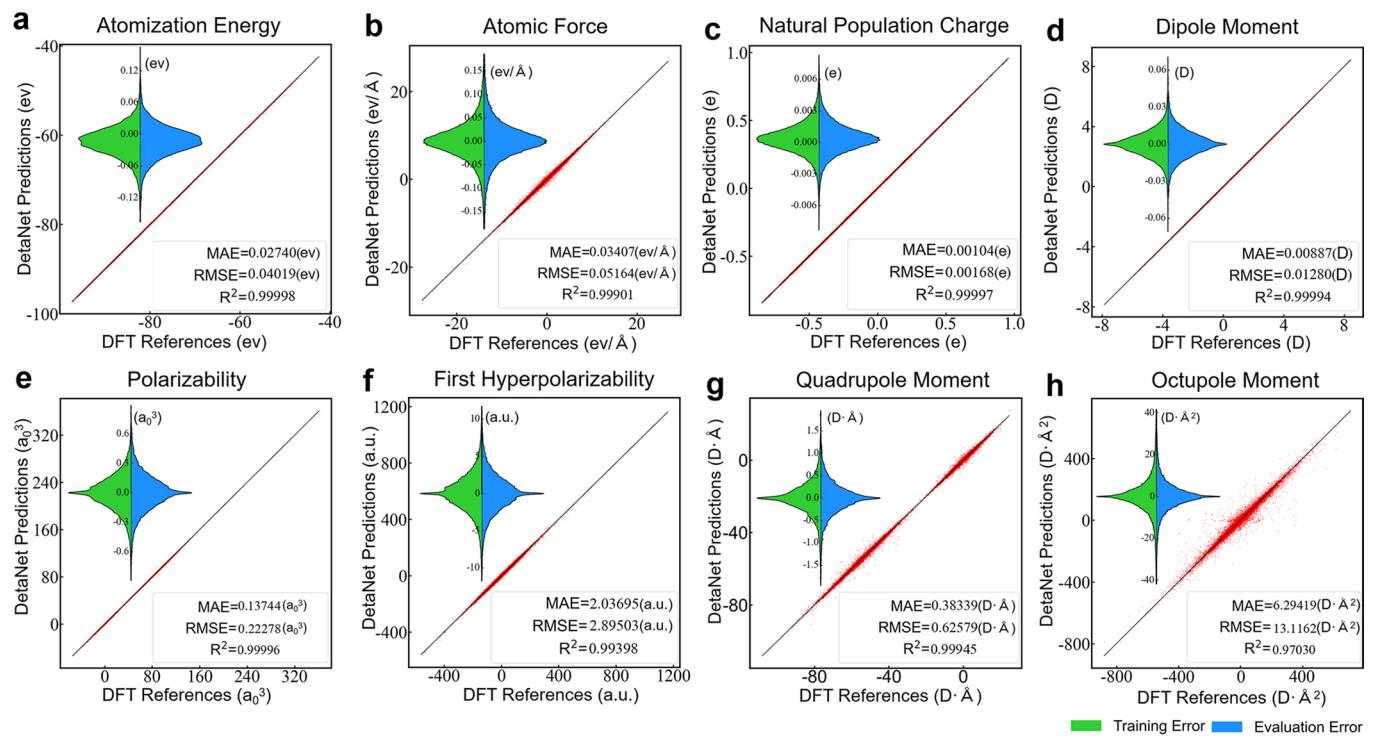
Extended Data Fig. 2 | Schematic diagram for how the sub modules operate using the matrix representations, taking water as a example. **a.** Illustration of H2O molecule and the definition of central (i) and neighboring (j) atom. **b.** Matrix representation for the sub module of the atomic embedding. $O(Z_i)$ and $Q(Z_i)$ represent the nuclear types and the inherent atomic electronic structure. (S_i^0) is the generated scalar properties. **c.** Matrix representation for the message module. \vec{r}_{ij} is position vector from atom j to i . $Y_{l,p}^m$ is the spherical harmonic

function. $\Delta_M S_i^n$ and $\Delta_M T_{i,l,p}^{n,m}$ represent the corresponding residuals. **d.** Matrix representation for the radial embedding module, where w_k and the w_v are the weight corresponding key and value. **e.** Matrix representation for the radial attention module, where Mq , Mk and Mv are the corresponding query, key and value features. **f.** Matrix representation for the atomwise attention update module. $T_{i,l,p}^{n,m}$ and S_i^n are the irreps and scalar features.



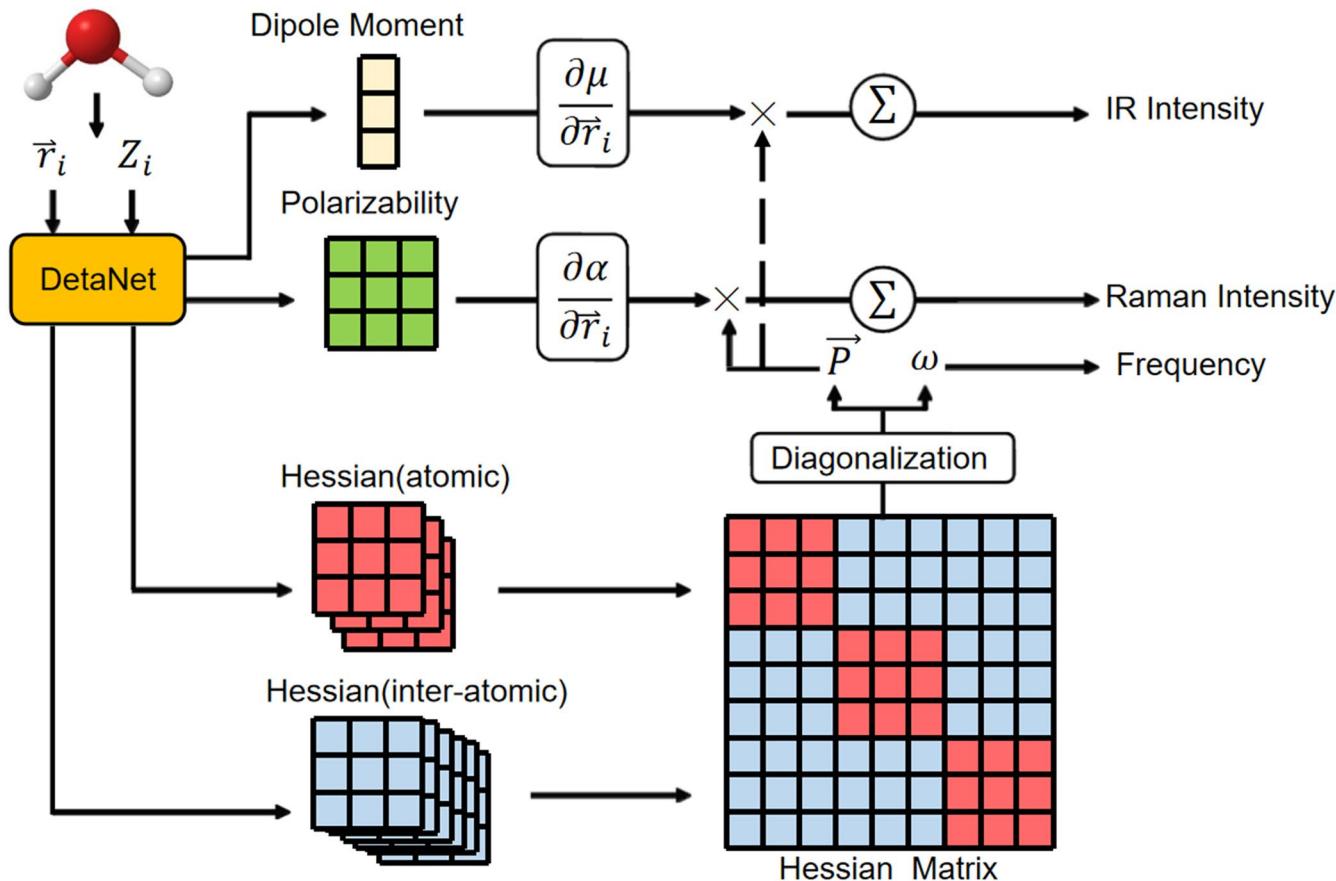
Extended Data Fig. 3 | Ablation studies of DataNet. Ablation experiments were performed to test the impact of each module on the model. We list the MAE of the dipole moment vectors and polarizability tensors if we exclude any given module. The model has the best performance with parameter values of $N=3$,

$l_{max}=3$, when excluding the cutoff function while keeping the electronic features, the radial self-attention module, the update module and the local part in the output function. N and l_{max} are the maximum interaction layers and the maximum degrees, respectively.



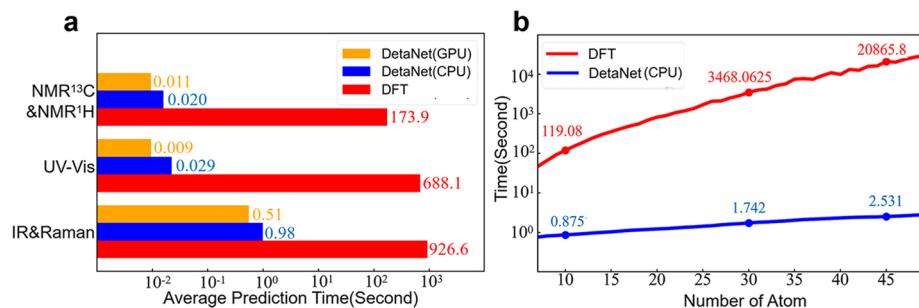
Extended Data Fig. 4 | Error distributions and regression plots of DeltaNet's predictions for eight properties. **a.** Energy learned from partial QM7-X datasets. **b.** Atomic forces learned from partial QM7-X datasets. **c.** Natural Population Charge learned from QM9S datasets. **d.** Electric dipole Moment learned from QM9S datasets. **e.** Polarizability learned from QM9S datasets.

f. First hyperpolarizability learned from QM9S datasets. **g.** Electric quadrupole moment learned from QM9S datasets. **h.** Electric octupole moment learned from QM9S datasets. The MAE, RMSE and R² represent the mean absolute errors, the root mean square errors and the coefficients of determination.



Extended Data Fig. 5 | Complete program for predicting infrared and Raman spectra using DetaNet. We firstly performed the frequency analysis by diagonalizing the DetaNet-predicted Hessian matrix to obtain the vibrational frequencies and the corresponding normal coordinates. Then the infrared adsorption intensities and Raman scattering activities were calculated as the first

derivatives of the polarizability and dipole moment with respect to the normal coordinates using the chain rule. Here \vec{r}_i is the atomic position and Z_i is the atomic number. ω represents the frequency of the adsorption/scattering light and \vec{P} is the normal coordinates. μ and α are the dipole moment and polarizability tensor, respectively.



Extended Data Fig. 6 | Computational efficiency of DetaNet. **a.** Comparison of average times in seconds for the prediction of the vibrational, UV-Vis and NMR spectra. **b.** Average computational times for prediction of vibrational spectra

using DFT and DetaNet (CPU) with increasing molecular size. Here DFT and DetaNet (CPU) indicates the time consumed on an Intel-i7 8700 K device, while DetaNet (GPU) on NVIDIA RTX 3080Ti.