

基于 SE(3)-Transformer 的晶体电子态密度预测

刘家轩 岳佳仪

2024 年 5 月 28 日

1 引言

1.1 晶体的电子能带与电子态密度

晶体是一类具有周期性结构的固体，由于晶体平移一个晶格矢后的结构完全相同，晶体内原子实与电子的相互作用势也完全相同，即：

$$\hat{T}(R_l)\hat{V}(x) = \hat{V}(x - R_l) = \hat{V}(x) \quad (1)$$

因此，我们可以知道在晶体中，整晶格矢量的平移算符与电子的哈密顿量（能量算符）是对易的：

$$[\hat{H}, \hat{T}(R_l)] = \hat{V}\hat{T}(R_l) - \hat{T}(R_l)\hat{V} = 0 \quad (2)$$

于是晶体中电子的能量和平移算符的本征值 e^{ikR_l} 可以同时确定，我们可以将电子能量 E 写成波矢 k 的函数 $E_n(k)$ ，这就是晶体的电子能带。波矢 k 是电子波包的准动量，所以电子能带 $E_n(k)$ 也就对应于经典物理中的能量-动量色散关系 $E(p)$ 。如果能够知道电子能带结构，我们就可以知道晶体电子的大部分信息。态密度则是在能带上根据 k 均匀撒点，并统计每个能量所占有的点的数目，反映的是每个能量上电子可能的状态的数目（图1）。从电子态密度中，我们可以获知能带的总体信息，例如，我们可以判断晶体是绝缘体还是金属。另外，在进行材料宏观性质的研究时，（如研究电子电导、电子热容等），往往是使用求和后的态密度进行研究。

利用计算机进行密度泛函（Density Functional Theory, DFT）计算，已经可以较好地从第一性原理计算出晶体材料的能带和态密度。然而，由于 DFT 计算过程中需要自洽求解电荷密度；而每一步自洽都需要数值求解薛定谔方程，因此 DFT 的计算相对较慢，代价较高。

1.2 神经网络的不变性 (invariant) 与等变性 (equivariant/covariant)

对于拟合一些实际物理问题而言，简单的图神经网络则会遇到一些问题。例如给定一些分子结构及其总能的数据，我们使用图网络来拟合一个从分子结构到总能量的映射。当这个拟合好的图网络遇到了一个与数据集中一模一样的分子，但是输入的数据旋转了一个角度，即使实质上是一样的结构，神经网络预测出来的总能量依旧会非常差。这是因为图网络并不保证拟合函数和 $SO(3)$ 群操作是对易的；而实际分子结构到能量的映射应当保证对 $SO(3)$ 群操作对易。因此，在实际使用神经网络解决物理问题时，我们往往要让神经网络具有一定的对称性；更具体地说，我们希望神经网络的每一层都和 $SO(3)$ 的群操作对易：

$$F^i(gX) = D(g)F^i(X), g \in SO(3) \quad (3)$$

$$(4)$$

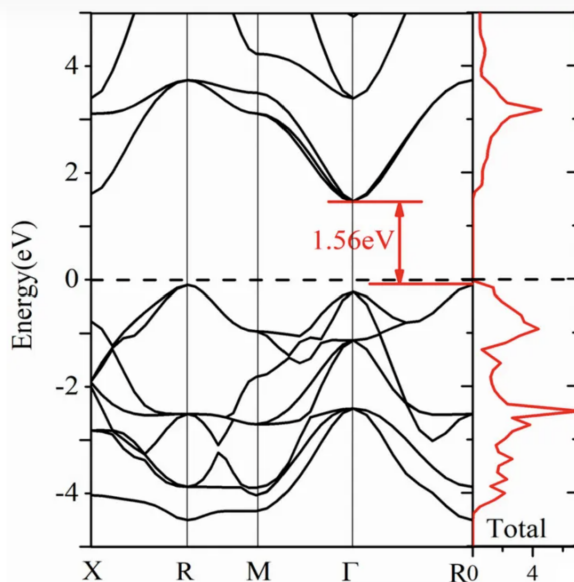


图 1: 能带-态密度示意图, 左黑线为能带示意图, 右红线为对应的态密度。截自知乎<https://zhuanlan.zhihu.com/p/654133143>

满足上式的神经网络则称为等变的 (equivariant) 神经网络。

对于拟合能量的任务而言, 由于分子整体旋转一个角度, 其能量与原来相等, 因此 $D(g)$ 可以是恒等表示:

$$F^i(gX) = F^i(X), g \in SO(3) \quad (5)$$

$$(6)$$

这种特殊的等变网络, 则称为不变 (invariant) 网络。即使我们想要力、位力这类等变的信息, 我们可以对能量做原子位置的偏导。如此看来, 似乎我们只要构建好一个不变的神经网络, 当我需要等变的特性时, 我只需要对输出做相应的偏导, 然后 (如果需要的话) 再通过分导表示的方法将偏导化到各个不可约表示对应的空间即可, 并不需要建立等变神经网络。但是, 在 [1] 中, Batzner 发现, 对于拟合能量和力的任务来说, 等变神经网络性能远好于不变神经网络的性能。对于拟合分子能量的任务来说, 即使最后我们想要的物理量是不变的, 但是影响分子能量的许多变量 (例如原子轨道波函数) 却是等变的。因此, 在中间层启用等变神经网络是具有一定物理意义的, 我们需要一个对于任意阶的不可约表示空间都是等变的神经网络。

2 方法

2.1 特征设计

在一般的分子特征学习任务中, 输入输出符合 3D 点云图结构。3D 点云图以分子中原子为结点, 点云的位置为原子的坐标 r_α , $\alpha = 1, \dots, N$ 。原子的化学环境信息 F_α 为结点特征, 并通过坐标和结点信息的得到的边信息进行更新。以单个神经网络层的映射为例, 输入和输出具有以下形式

$$(r, F_i) \xrightarrow{\text{layer}} (r, F_o)$$

其中 $r \in R^{3N}$, $F_i \in R^{c_i d_i N}$, $F_o \in R^{c_o d_o N}$ 。N 为原子数目, d 为特征维数, c 为通道数, i, o 表示输入和输出。

该图中的结点特征 f_α 一般是局域的信息, 例如元素种类、受力等。除了局部特征, 全局特征能够反映分子性质, 如形成能等。

在物理学中, 局域态密度为

$$dos_\alpha = \sum_{n=1}^N |\phi_n(r_\alpha)|^2 \delta(\epsilon - \epsilon_n)$$

即原子处波函数处于在相应能量本征态下的概率。将总态密度看成局域态密度的贡献之和

$$dos = \sum_{\alpha} dos_\alpha$$

而 dos_α 为图神经网络输出层的结点特征。因此我们采用 MPNN[2] 结构, 先由若干层 SE(3)transformer 得到 dos_α , 再由 meanpooling 消息汇集的方式得到总 dos。网络中特征提取的方式如下

$$f_0 = \text{embedding}(A) \in R^{\text{types} \times N}$$

$$f_{k+1} = \phi(r, f_k, f_0)$$

$$f_{out} = \text{silu}(\text{Linear}(f_K))$$

$$dos = \text{meanpooling}(f_{out})$$

其中 $f^l \in R^{d_l c_l N}$, d_l, c_l 分别为特征维度和通道数。初始原子编码采用简单的 onehot 编码经过可学习的 MLP 映射得到, 其中 types 为元素种类个数, 在后续被认为是输入的通道数, 对应特征维度为 1。K 为总 layer 数目, 即网络的最后一层输出特征为 $ldos$, $f^K \in R^{c_{dos} N}$, 即通道数为 dos 维度, 特征维度为 1。需要说明的是, 在 SE(3)transformer 层 ϕ 中每层信息更新都参考了初始原子编码 f^0 , 这样可以防止信息的过度平滑。

此处 dos 特征被认为窗口区间内 dos 采样点对应的值向量, 即 $dos_c = \text{density}(\epsilon_c)$, $c = 1, \dots, c_{dos}$ 。其具体维度数与数据处理方法相关。实验中选择 (-10, 10)eV 费米面附近, 400 个均匀采点的到的 400 维向量。目前有文章对 dos 特征的表示方法做了比较 [3], 证明了现有的几种基于 dos 曲线的特征在训练任务中的表现没有显著的区别, 因此选择简单的 dos 值作为特征是合理的。

2.2 等变网络 backbone

结合输入输出特征, 以及网络等变的要求, 我们选择 multi-edge-SE(3)transformer 作为网络更新的层, 并基于 resnet 选择了 self-interaction 的更新方式。

现已有许多关于 SE(3) 或 E(3) 等变的神经网络, 主要可以将它们分成三种, 不变的 (包含 $l = 0$ 特征), 等变的 (包含 $l = 1$ 特征) 和 SE(3) 等变的 (包含 $l > 1$ 特征)。不变网络, 如 SchNet[4], Dimenet[5]。 $l = 1$ 等变, 如 EGNN[6], $l > 1$ 等变, 如 Tensor Field Network[7] 和 SE(3)Transformer[8]。此处 l 为不可约张量维度 (以转动群元作用下的变换方式为基准)。为了标记方便, 在以下的更新中, 统一采用 $f^{l_i}, f^{l_o}, f^{l_f}$ 表示层的输入、输出、滤波器特征的某子空间特征。总特征 $F = \bigoplus_{l \leq l_{max}} f^l$, $f^l \in R^{2^{l+1}}$ 。

对于晶体结构, 除了一般的 SE(3) 对称性, 还有周期对称性。针对此, CGCNN 方法 [9] 提出采用 multi-edgegraph 构造晶体图, 该方法结合了晶体的周期性信息, 通过在一定截断半径内捕捉邻居原子 (包括跨胞作用) 来实现周期对称性。

我们的网络结合了以上两种网络的特点。首先，我们考虑晶体周期结构在一定截断半径内的全部邻居构造 multi-edge 晶体图，然后对晶体图的每个结点赋予高阶张量特征并整体上采用 TFN 的架构更新特征。针对选择 TFN 的原因，主要考虑到 dos 预测问题本身的复杂性。由于态密度反应电子结构性质，该性质一定受到原子轨道相互作用和元素种类的共同影响，因此考虑高阶张量是有必要的。并且由于在一般的物理模型中，对原子轨道的讨论也不会涉及 l 过高的轨道，因此选择低于一定 l_{max} 的等变特征也能够符合物理模型。

TFN 中由邻居的特征卷积更新结点特征方法如下

$$f_{\alpha}^{l_o} = \sum_{\beta, l_i} W^{l_o l_i}(\mathbf{r}_{\alpha\beta}) f_{\beta}^{l_i} \quad (7)$$

$$W^{l_o l_i}(\mathbf{r}_{\alpha\beta}) = \sum_{\beta, n, l_f} CG^{l_o l_i l_f} \hat{w}^{l_o l_i l_f n} R^n(|r_{\alpha\beta}|) Y^{l_f}(\hat{\mathbf{r}}_{\alpha\beta}) \quad (8)$$

对于固定的输入输出和中间层滤波器维度, $W^{l_o l_i} \in R^{(2l_o+1) \times (2l_i+1)}$, $CG^{l_o l_i l_f} \in R^{(2l_o+1) \times (2l_i+1) \times (2l_f+1)}$, $\hat{w}^{l_o l_i l_f n} \in R^b$, $R^{n l_f} \in R^{b \times (2l_f+1)}$, $Y^{l_f} \in R^{2l_f+1}$, 其中 b 为径向基数目。注意不同于 CG , F , Y 的在不可约指标下还有内部自由度, \hat{w} 在 l_i, l_o, l_f 下只是一个常数, 例如对于 l_f , \hat{w} 与 Y 之间为普通数值乘积而非矩阵乘积。这说明卷积操作的滤波器径向为 $\sum_n \hat{w}^{l_o l_i l_f n} R^n(|r_{\alpha\beta}|)$ 函数, 该函数在不同特征耦合路径中对 R^n 基的展开系数 \hat{w} 为 TFN 网络全部自由度。

将矩阵内部自由度完整写出, 特征更新如下。

$$f_{\alpha m_o}^{l_o} = \sum_{\beta, l_i, m_o, m_i} W_{m_o m_i}^{l_o l_i}(\mathbf{r}_{\alpha\beta}) f_{\beta m_i}^{l_i} \quad (9)$$

$$W_{m_o m_i}^{l_o l_i}(\mathbf{r}_{\alpha\beta}) = \sum_{\beta, n, l_f, m_o, m_f, m_i} CG_{m_o m_f m_i}^{l_o l_i l_f} \hat{w}_n^{l_o l_i l_f} R^n(|r_{\alpha\beta}|) Y_{m_f}^{l_f}(\hat{\mathbf{r}}_{\alpha\beta}) \quad (10)$$

以上的讨论中尚未明确通道数的概念, 此处说明应将多通道看作有多个同维度的不可约表示 l_i 。根据以上变换, 多通道情况下一个特征可以和每份相同 1 进行以上变换并最终进行由 w 控制的加权和。因此以上更新方式中也包含了多通道下的变换。

除了卷积变换外, TFN 网络能实现的最一般操作是按照不可约特征的耦合分解, 其定义为

$$f^{l_o} = \sum_{l_1, l_2} W^{l_o l_1 l_2} f^{l_1} f^{l_2} \quad (11)$$

$$W^{l_o l_1 l_2} = \hat{w}^{l_o l_1 l_2} CG_{l_1 l_2}^{l_o} \quad (12)$$

即两个等变特征依赖 CG 系数耦合得到新的等变特征, 其中路径权重是可训练参数。卷积变换可以看作将径向维度考虑后, 以球谐函数和结点特征作为输入的一般张量耦合变换。

此外, 网络中还使用了 TFN 的线性变换 Linear, 线性变换只在相同维度的不可约特征间进行。

$$f^l = W^l f^l \quad (13)$$

$$W^l = \hat{w}^l \quad (14)$$

因为 TFN 只给出了一般的 3D 点云等变特征的更新方案。为了增强网络的表达能力，我们在 TFN 基础上结合了深度学习中的注意力机制和残差链接思想。注意力机制的设计参考了 SE(3)Transformer[8]，将 f_{out} 更新分为自注意力和邻居卷积两部分。（由于此时输入输出为完整的不可约表示的直和，因此此处特征维度标记不再使用 i, o 。）

$$f_{out,\alpha}^l = W_V^{ll} f_{in,\alpha}^l + \sum_{k \leq 0} \sum_{j \in N_\alpha/\alpha} a_{\alpha\beta} W_V^{lk} (x_\beta - x_\alpha) f_{in,\beta}^k$$

$$a_{\alpha\beta} = \frac{q_\alpha k_{\alpha\beta}^T}{\sum_{\beta' \in N_\alpha/\alpha} \exp\{q_\alpha k_{\alpha\beta'}^T\}}$$

其中 k, q, v 的概念与一般的注意力一致，但采用了保持等变性的更新方式

$$q_\alpha = \bigoplus_{l \geq 0} \sum_{l_1 \geq 0} W_Q^{ll_1} f_{in,\alpha}^{l_1}$$

$$k_{\alpha\beta} = \bigoplus_{l \geq 0} \sum_{l_1 \geq 0} W_K^{ll_1} (x_\beta - x_\alpha) f_{in,\beta}^{l_1}$$

$$v_{\alpha\beta}^l = \sum_{l_1 \leq 0} W_V^{ll_1} (x_\beta - x_\alpha) f_{in,\alpha}^{l_1}$$

可以看出

$$Atten_\alpha = \sum_{k \leq 0} \sum_{\beta \in N_\alpha/\alpha} a_{\alpha\beta} W_V^{lk} (x_\beta - x_\alpha) f_{in,\beta}^k$$

相比原先的信息更新只是完全取决于训练好的径向函数，SE(3)Transformer 使得更新考虑到结点之间本身相似性的关联。由于 $a_{\alpha\beta}$ 是转动不变的，因此网络的协变性的得到保持。

在 SE(3)Transformer 的基础上，我们借鉴了 ResNet，使得自注意力能够保留原子的初始嵌入信息，而不只是上一个输入层的信息。最终的网络层更新函数如下：

$$f_{k+1} = SelfInteraction(f_k, f_0) + Interaction(f_k, r)$$

其中

$$SelfInteraction(f_k, f_0) = \bigoplus_{l \geq 0} \sum_{l_1 l_2} \hat{w}^{l, l_1, l_2} CG_{l_1 l_2}^l f_k^{l_1} f_0^{l_2}$$

为结点特征与初始嵌入的 TFN 耦合。

$$Interaction(f_{in}, r) = Linear(Atten(f_{in}, r))$$

即首先计算 $atten$ 输出，然后进行 TFN 线性变换。由于网络中所有变换都是 TFN 形式，因此网络协变性被保证。

3 实验

3.1 数据集准备

material project 是由多所大学和研究机构维护的材料数据库，提供了超过 15w 个材料的结构和物理化学信息。其中拥有态密度的材料结构超过 9w 个，并且在不断更新中。我们从 material project 上下载了 48330 个符合实验要求的数据，并且在预处理后将其中 70% 作为训练集，30%

部分作为测试集。我们选择的这些结构满足形成能在 $(-500, 100)\text{eV}$ 之间，并且 dos 数据的 ϵ 取点数目为 2000。

数据经过预处理后被用于训练。由于我们更关心 fermi 面附近的 dos，我们对数据在费米面窗口附近进行了 Gaussian 差值，最终得到 $(-10, 10)\text{eV}$ 之间 400 个数据点的 dos。进一步，考虑到 dos 受到原子种类和化学环境的影响，并且受到 DFT 计算的网格密度影响，导致不同结构 dos 的量级不同，我们对不同材料的原子局域 dos 进行了归一化，保证每个结构的总 dos 在区间内的最大值为 1。这些归一化的 dos_{norm} 被作为训练的目标，训练的 loss 设置为 MPNN 预测的全局 \hat{dos} 和归一化后 dos 的 mse 损失函数。

$$loss = mse(\hat{dos}, dos_{norm})$$

3.2 网络参数和结构

我们使用了 pytorch, pytorch geometric 和 e3nn[10] 实现网络中的 SelfInteraction 和 Interaction。pytorch geometric 是几何深度学习普遍使用的 python 包。e3nn 是针对不可约张量网络设计的 python 包，它包括不可约张量类 (Irreps)，不可约张量耦合类 (FullTensorProduct)，球谐函数 (SphericalHarmonics) 等。可以采用 e3nn 设定输入输出张量的维度初始化对应类，然后通过传入数据张量，得到经过 CG 系数变换后的输出不可约张量。可训练参数在类初始化的时候可以被定义。

网络的 ϕ 变换采用的 e3nn 模块如下 (输入输出层都是简单的不变张量，因此对应的输入输出层网络和一般网络一致，在此不做展开)：

Self Interaction：采用 FullTensorProduct 结构。

Interaction：Attention 部分，q 采用对应单结点的线性变换 Linear，k 和 v 采用 FullyConnectedTensorProduct。其中 Filter 中球谐函数采用 SphericalHarmonics。外层的 Linear 采用了 e3nn 中 Linear 函数。

以上网络结构中的参数均可在训练时通过 config 设置。在我们的实验中，考虑到数据包含的元素种类为 118，输出特征维度为 400，因此选择中间层的表示维数为 $l_{max} = 3$ ，中间层通道数为 259。

4 结论

我们使用了来自 Materials Project 的 48330 个晶体材料及其态密度数据进行训练，使用 4 张 h800 训练了约 6 个小时，损失函数如图2所示。最后，我们得到最小的验证集损失为 0.0192。

我们统计了测试集中 9666 个数据损失函数的分布，如图3.(a) 所示。我们从测试集中不同位置随机选取了一个材料，并对比其计算数据和神经网络的预测，见图3.(b)-(f)。可以看到，我们的网络可以预测测试集中绝大部分材料的态密度。即使是对于误差较大的材料，我们的网络仍然能预测得到态密度的大致形状。

4.1 相似的工作

在课程作业进行到一半时，我们注意到一篇相似的工作 [11] 使用 Tensor Field Network 的结构预测声子态密度。声子态密度是晶体原子振动的色散关系 $E(k)$ 的态密度；尽管声子与电子在理论上使用了相似的方法进行讨论 (因而看上去拥有相似的形式)，它们拥有截然不同的物理来

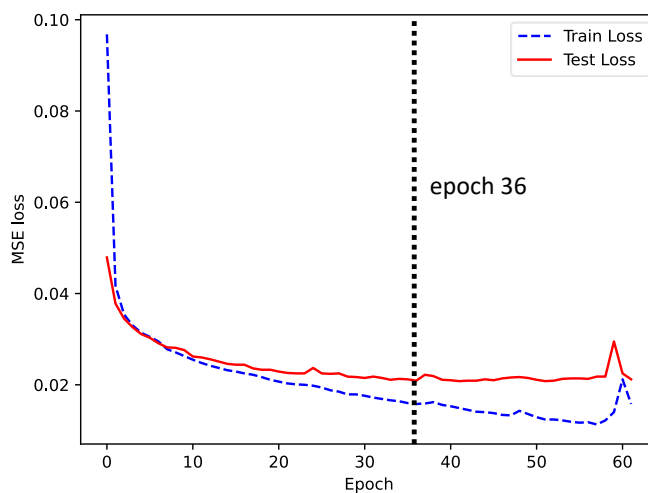


图 2: 训练过程中的损失函数, 蓝色虚线与红色实线分别为训练集、测试集的损失函数。第 36 个 epoch 的测试集损失函数最小。

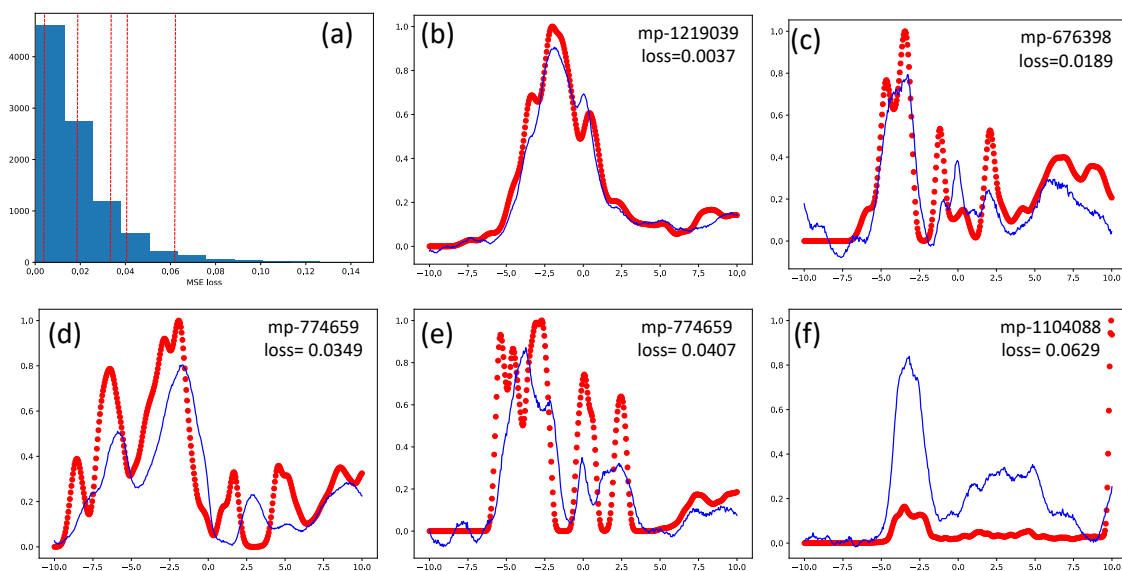


图 3: 测试集损失的分布及测试集中的几个例子。(a), 测试集的方均损失分布, 图中红色虚线从左至右分别表示 (b)-(f) 图在测试集的位置。(b)-(f) 测试集中不同材料的计算数据与神经网络的预测数据。红点表示从 Materials Project 上下载得到的数据, 蓝线表示神经网络的预测。

源。声子是晶体内原子振动的量子，因此声子态密度很大程度上与晶体结构直接相关；从晶体结构出发预测声子态密度也相对容易。电子态密度则涉及到不同原子轨道间的杂化耦合，同时受到晶体结构和原子元素种类的影响，因此预测电子态密度是更为困难的任务。

然而，对比 [11] 的结果，我们在使用了相同标度的情况下，得到了更低的损失函数。我们认为这有两方面原因：首先是我们使用的 SE(3)-Transformer 结构更好，能够更高效地学习结构与态密度之间的映射。其次，[11] 中的训练数据较少，仅有约一千五百个材料及其声子态密度。而我们则使用了来自 Materials Project 的约五万个材料的结构及其电子态密度。

5 分工及声明

本次小组作业由刘家轩和岳佳怡共同完成。刘家轩负责模型代码的编写和调试；模型的训练以及网络结构的调整，完成了约 50% 的工作。岳佳怡同学负责训练数据的收集和预处理；模型代码的编写、调试以及说明文档的编写，完成了约 50% 的工作。

模型主题部分的代码参考了 NequIP[1] 以及 e3nn[10] 给出的代码；而数据预处理部分则参考了 MatDeepLearn_DOS[12] 的代码。

本次课程作业的所有代码放在了 github.com/JiaxuanLiu-Arsko/DosNet 上；训练好的模型文件和配置文件则在该代码仓库的 [Release:v0.0.1-full](#) 中。

参考文献

- [1] Simon Batzner, Albert Musaelian, Lixin Sun, Mario Geiger, Jonathan P. Mailoa, Mordechai Kornbluth, Nicola Molinari, Tess E. Smidt, and Boris Kozinsky. E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nature Communications*, 13(1):2453, May 2022.
- [2] Justin Gilmer, Samuel S. Schoenholz, Patrick F. Riley, Oriol Vinyals, and George E. Dahl. Neural message passing for quantum chemistry, 2017.
- [3] Chiheb Ben Mahmoud, Andrea Anelli, Gábor Csányi, and Michele Ceriotti. Learning the electronic density of states in condensed matter. *Physical Review B*, 102(23), December 2020.
- [4] Kristof T. Schütt, Pieter-Jan Kindermans, Huziel E. Sauceda, Stefan Chmiela, Alexandre Tkatchenko, and Klaus-Robert Müller. Schnet: A continuous-filter convolutional neural network for modeling quantum interactions, 2017.
- [5] Johannes Gasteiger, Janek Groß, and Stephan Günnemann. Directional message passing for molecular graphs, 2022.
- [6] Victor Garcia Satorras, Emiel Hoogeboom, and Max Welling. E(n) Equivariant Graph Neural Networks, February 2022. arXiv:2102.09844 [cs, stat].
- [7] Nathaniel Thomas, Tess Smidt, Steven Kearnes, Lusann Yang, Li Li, Kai Kohlhoff, and Patrick Riley. Tensor field networks: Rotation- and translation-equivariant neural networks for 3D point clouds, May 2018. arXiv:1802.08219 [cs].

- [8] Fabian B. Fuchs, Daniel E. Worrall, Volker Fischer, and Max Welling. SE(3)-Transformers: 3D Roto-Translation Equivariant Attention Networks, November 2020. arXiv:2006.10503 [cs, stat].
- [9] Tian Xie and Jeffrey C. Grossman. Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties. *Physical Review Letters*, 120(14), April 2018.
- [10] Mario Geiger and Tess Smidt. e3nn: Euclidean Neural Networks, July 2022. arXiv:2207.09453 [cs].
- [11] Zhantao Chen, Nina Andrejevic, Tess Smidt, Zhiwei Ding, Qian Xu, Yen-Ting Chi, Quynh T. Nguyen, Ahmet Alatas, Jing Kong, and Mingda Li. Direct Prediction of Phonon Density of States With Euclidean Neural Networks. *Advanced Science*, 8(12):2004214, June 2021.
- [12] Victor Fung, P. Ganesh, and Bobby G. Sumpter. Physically Informed Machine Learning Prediction of Electronic Density of States. *Chemistry of Materials*, 34(11):4848–4855, June 2022.