

Published as a conference paper at ICLR 2024

EQUIFORMER V2: IMPROVED EQUIVARIANT TRANSFORMER FOR SCALING TO HIGHER-DEGREE REPRESENTATIONS

Yi-Lun Liao¹ **Brandon Wood²** **Abhishek Das^{2*}** **Tess Smidt^{1*}**

¹Massachusetts Institute of Technology ²FAIR, Meta *Equal contribution

AI赋能的生物大分子模拟和计算

EQUIFORMER V2

汇报人：王柯人/周子航/吴典蔚/袁雅萱 Group 10

日期：2026.05.27

目录

01- 生物背景 (benchmark)

王柯人

02- 模型架构

周子航

03- 训练参数和过程

吴典蔚

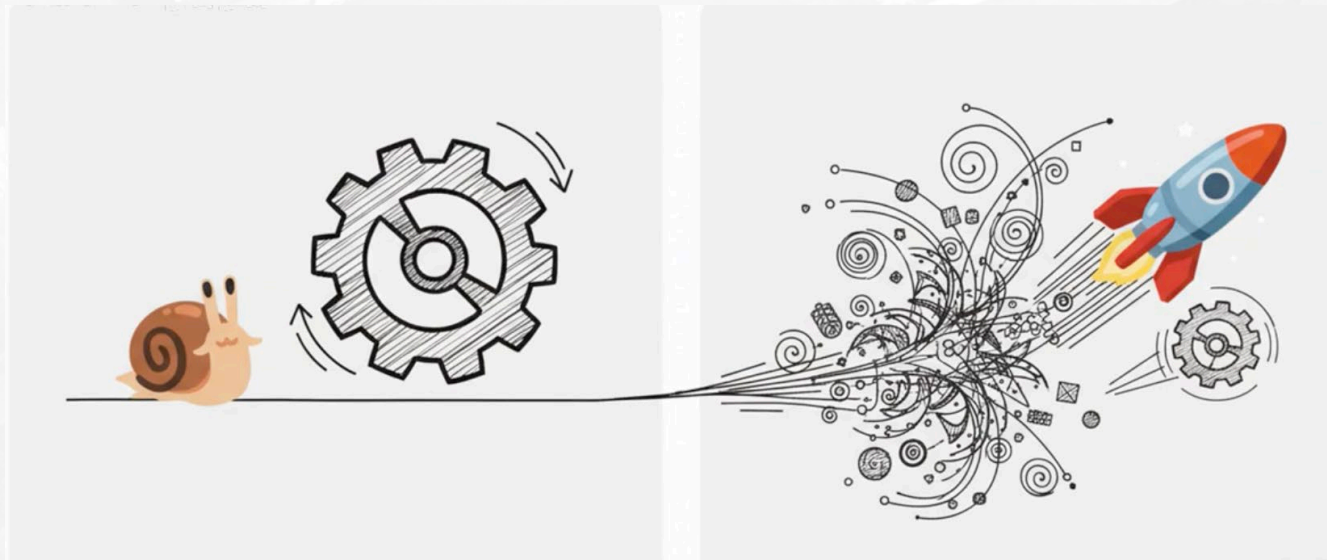
04- 训练结果可视化分析

袁雅萱

01

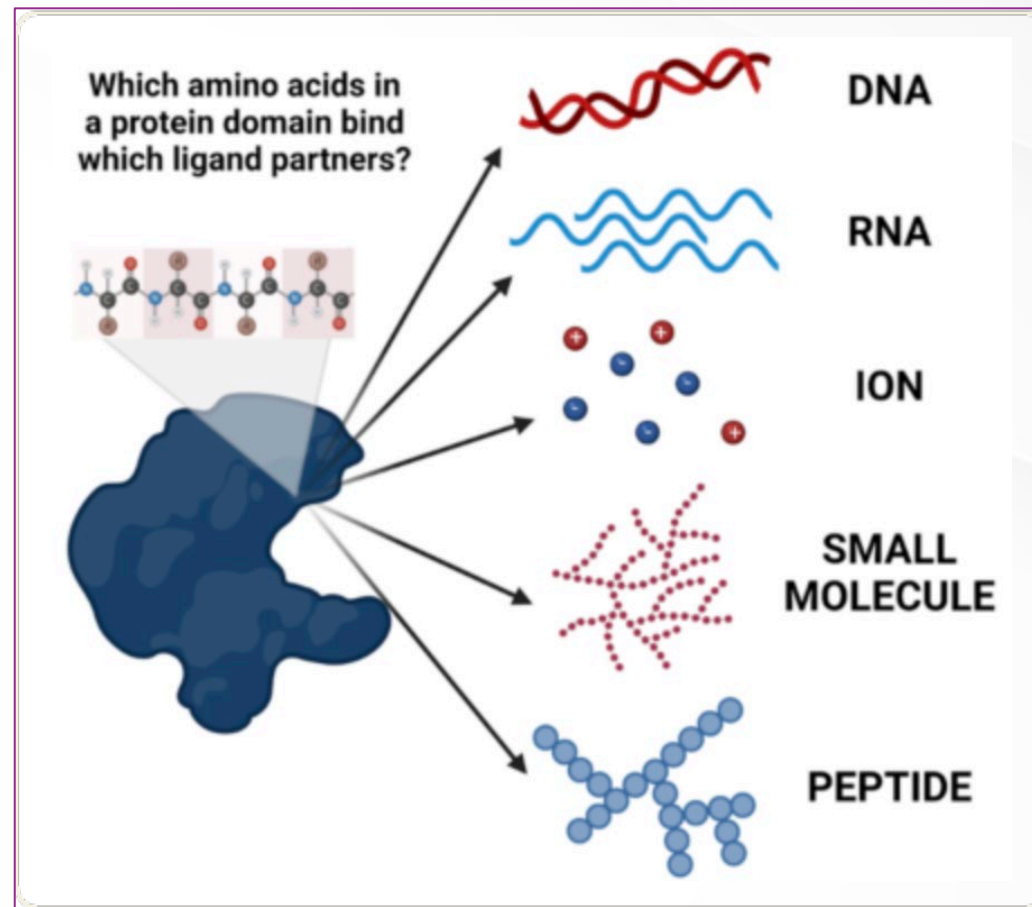
生物背景 (benchmark)

| 王柯人



从催化到蛋白质构象

- **精确度瓶颈：**传统的经验力场（如 AMBER /CHARMM）无法描述复杂的化学键断裂与生成。
- **计算成本：**第一性原理（DFT）精度极高，但对生物大分子而言计算成本呈爆炸式增长。
- **几何等变性：**原子系统具有旋转和平移不变性，模型必须满足等变性（Equivariance）才能实现数据高效。

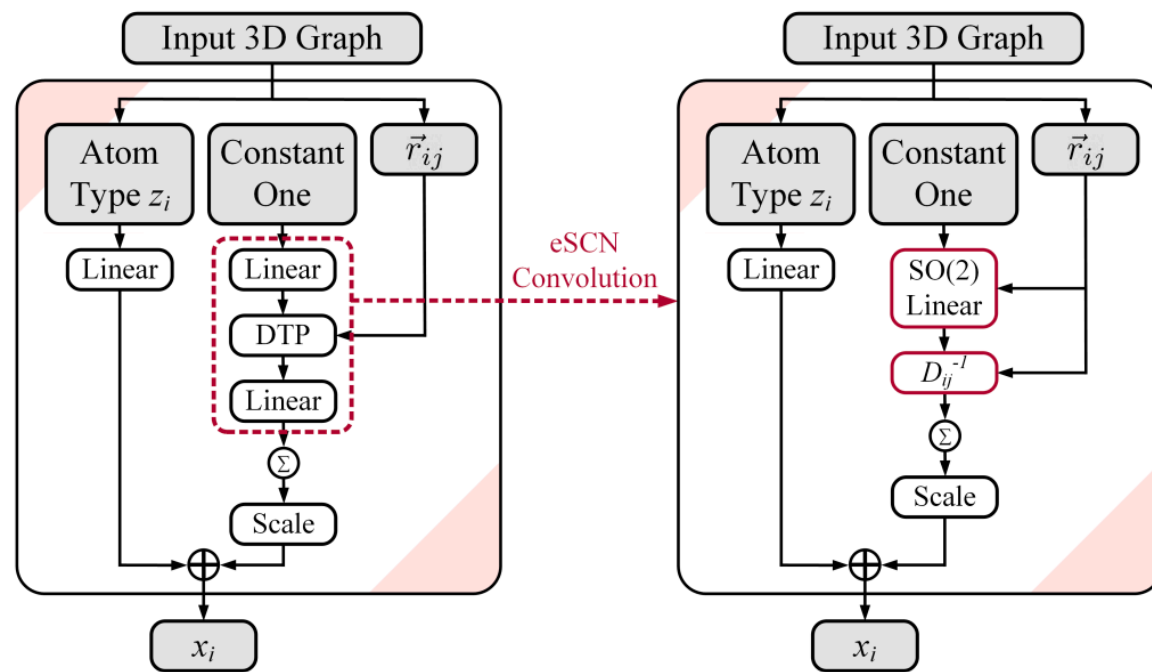


核心挑战：高阶表示的扩展性

为了捕捉复杂的角信息，我们需要高阶的不可约表示 (irreps)，但面临严重的计算复杂度问题：

$$\text{复杂度: } O(L_{\max}^6) \rightarrow O(L_{\max}^3)$$

- **EquiformerV1 限制：** 只能使用低阶 ($L \leq 2$)，限制了力的预测精度。
- **EquiformerV2 突破：** 引入 eSCN 卷积，支持 $L=6$ 甚至 $L=8$ 的高阶张量。



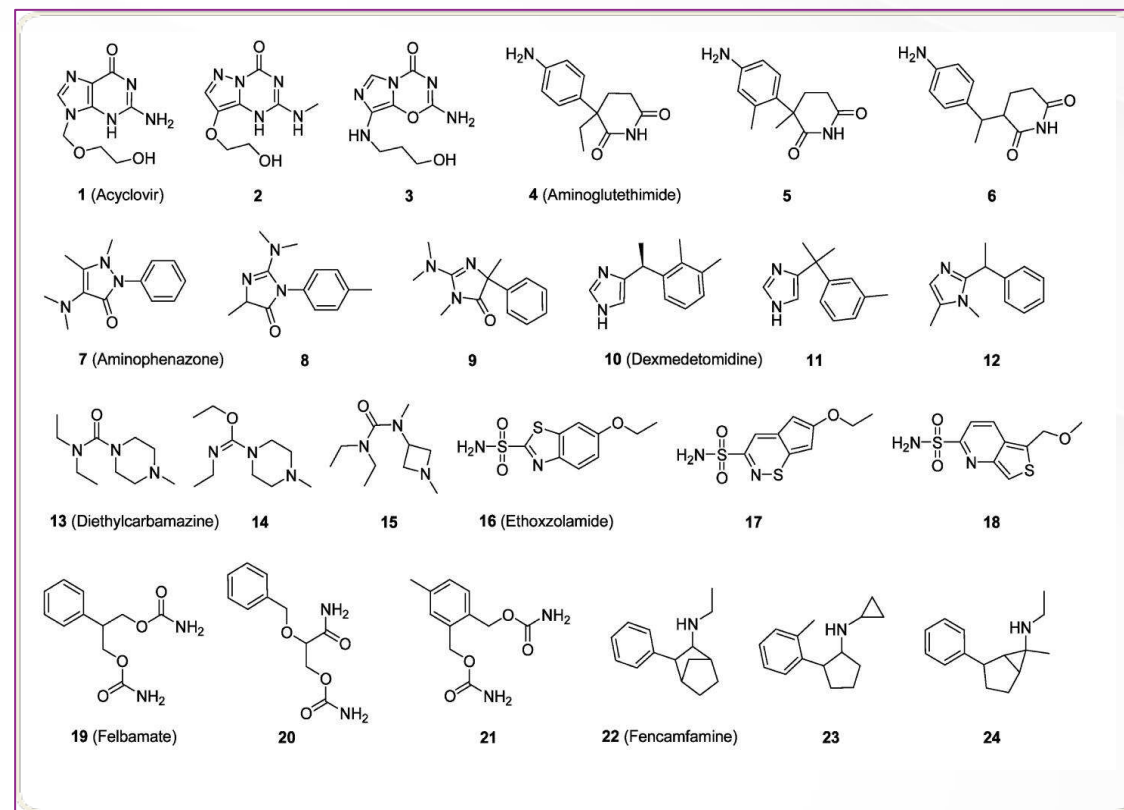
(c) Embedding (Atom + Edge-Degree)

Benchmark 1: QM9 (小分子属性)

入门级测试集

包含约 13.4 万个平衡态的小分子，涵盖 12 种物理属性：

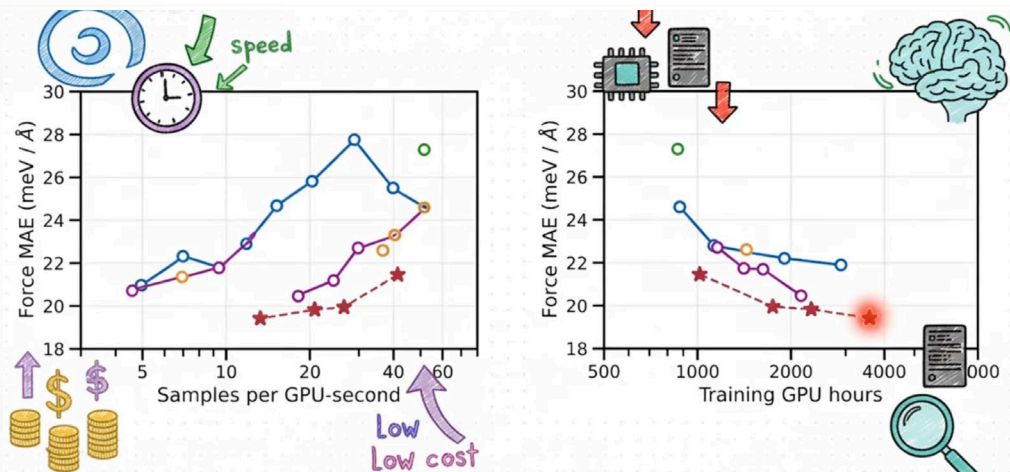
- **目标：** 验证模型捕捉分子基本性质（如偶极矩、焓等）的能力。
- **表现：** 在 12 项任务中，EquiformerV2 在 9 项上达到了最优表现。
- **结论：** 即使在原子数较少、角度变化不大的系统中，高阶表示依然能提供额外的增益。



Benchmark 2: OC20 (大型多样化系统)

目前最大的原子模拟数据集

- 1.3 亿个原子结构，模拟吸附质与催化剂表面相互作用。
- 不仅有平衡态，还有非平衡态的构象，对 MD 模拟至关重要。



OC20 的三项全能测试

1.S2EF

从结构预测能量和力，MD 模拟最直接的任务。

2.IS2RS

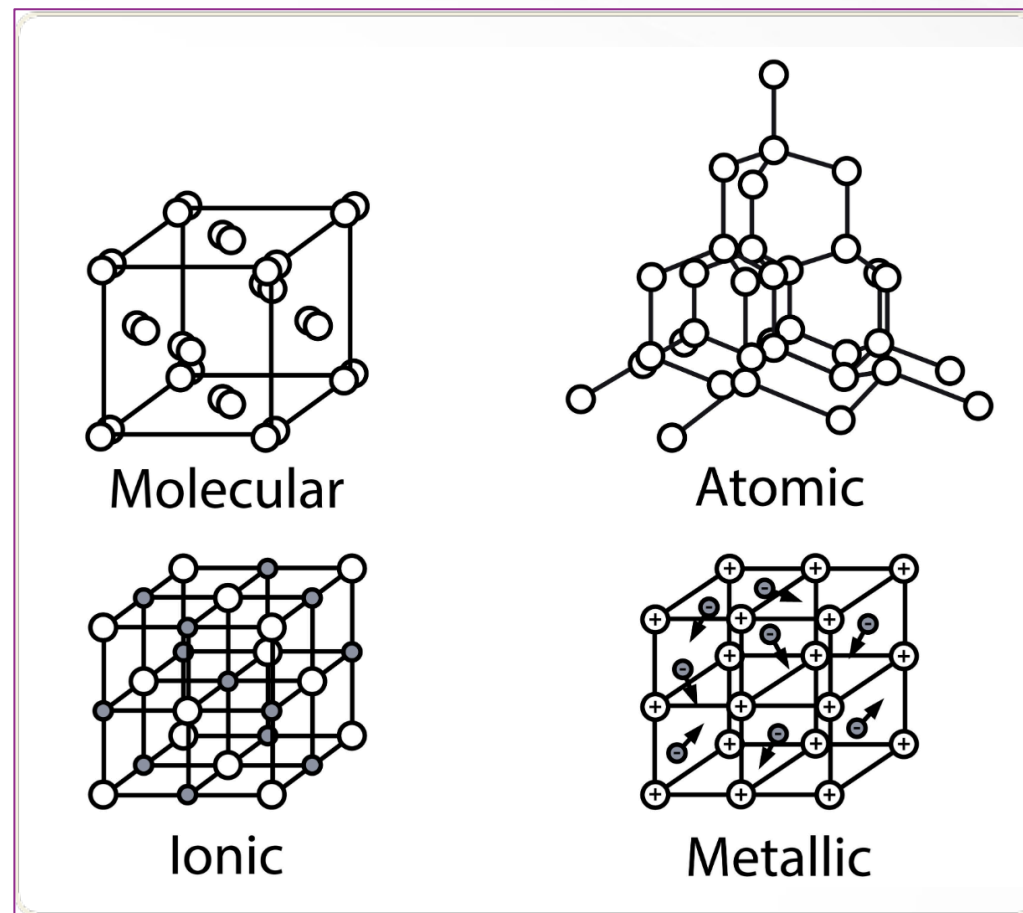
从初始结构预测弛豫后的结构。考验模型的长程优化能力。

3.IS2RE

直接预测弛豫后的能量，类似于预测结合亲和力

超越数据量的胜利

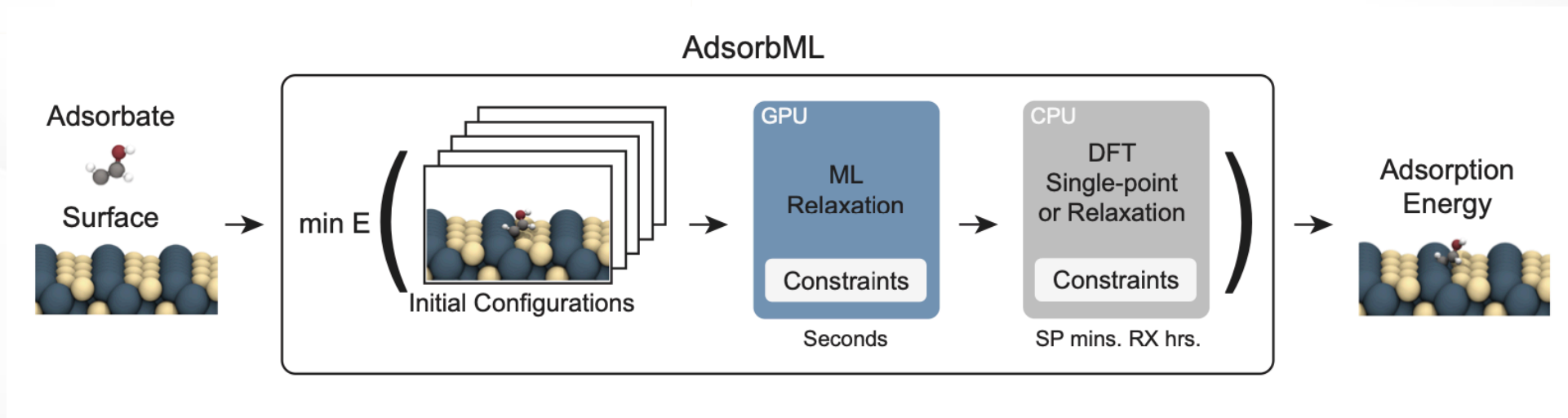
- **任务：** 预测 DFT 全能量（Total Energy），比吸附能更难。
- **提升：** EquiformerV2 仅在 OC22 上训练，就打败了在 OC20+OC22 混合训练的其他模型。
- **启示：** 更好的数学架构（高阶表示）可以补偿训练数据的不足。



2 倍加速与精度平衡

在寻找最低能量构象的过程中：

- EquiformerV2 将所需的 DFT 单点计算次数减少了 50%。
- 成功率大幅提升。
- **结构生物学类比：** 类似于用更少的计算步数找到蛋白与配体最稳定的结合模式。



OC20 All+MD

指标	GemNet-OC	SCN	EquiformerV2 (153M)
能量 MAE (meV)	252	243	230 (↓9%)
力 MAE (meV/Å)	20.0	17.1	14.6 (↓17%)
弛豫成功率 (AdsorbML)	84.4%	87.8%	91.6% (↑)

1. 阶数 (Degree) 决定精度

对于预测“力”这种矢量信息，高阶张量展开 ($L=6^+$) 是不可或缺的。

2. 效率与精度的权衡

eSCN 卷积解决了高阶扩展的计算代价问题，使得模型在 1 亿级数据上可训练。

3. 领域迁移潜力

在催化剂和氧化物上的成功，预示着该架构在蛋白质构象空间建模中的巨大潜力。

4. 归一化与稳定性

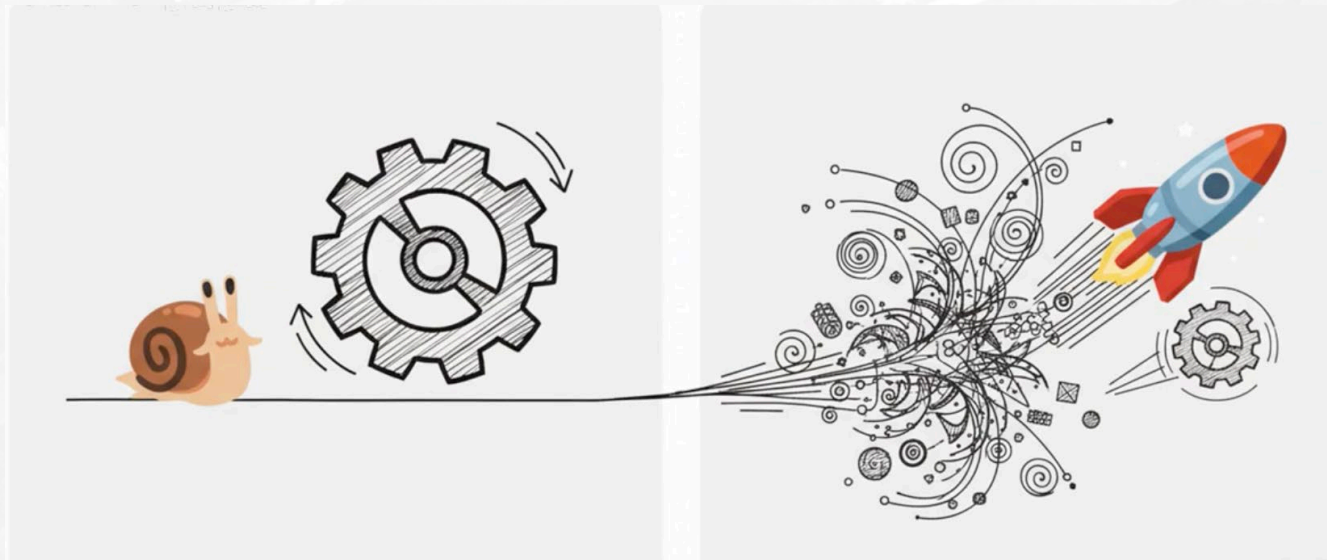
新的归一化技术使得在大规模数据集上训练时能保持收敛

小结

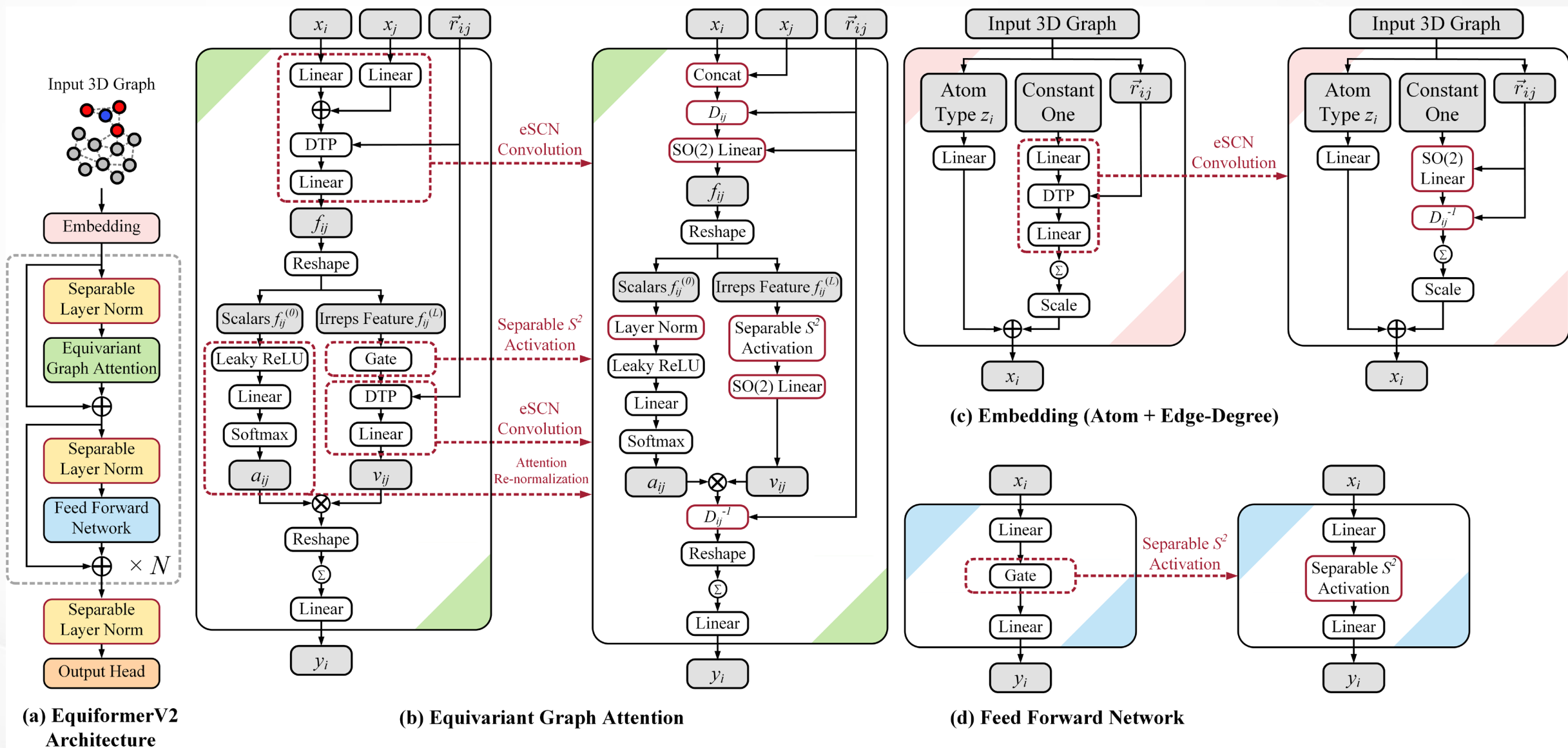
02

模型架构

| 周子航



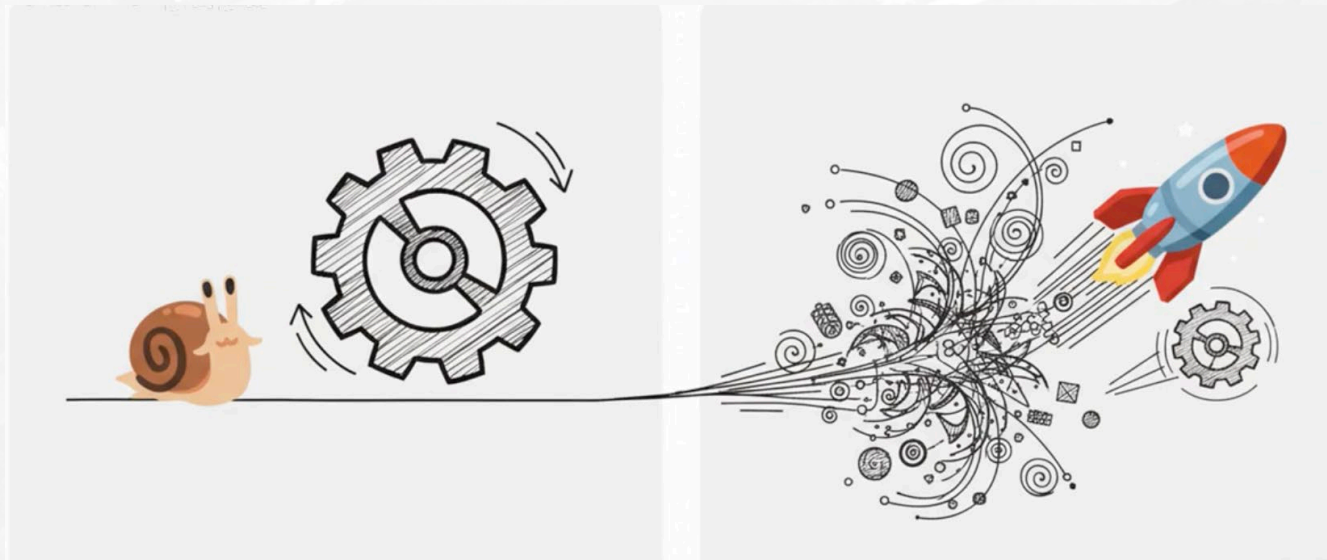
模型架构 (equiformer v2)



03

训练参数和 过程

| 吴典蔚



数据集参数

trainer: forces_v2

训练器类型

dataset:

```
- src: /home/wudianwei/Equiformer_V2/equiformer_v2-main/datasets/train_lmdb
  normalize_labels: True
  target_mean: -0.7554450631141663
  target_std: 2.887317180633545
  grad_target_mean: 0.0
  grad_target_std: 2.887317180633545
- src: /home/wudianwei/Equiformer_V2/equiformer_v2-main/datasets/val_lmdb
```

开启标签标准化

能量

力

logger: wandb

task:

```
dataset: trajectory_lmdb_v2
description: "Regressing to energies and forces for DFT trajectories from OCP"
type: regression
metric: force_mae
labels:
  - potential energy
grad_input: atomic forces
train_on_free_atoms: True
eval_on_free_atoms: True
```

hide_eval_progressbar: False

模型参数

```
model:
  name: equiformer_v2

  use_pbc: False      用孤立分子训练，关闭周期性边界条件
  regress_forces: True
  otf_graph: False

  max_neighbors: 50    每个原子最多连接的邻居原子数量
  max_radius: 6.0
  max_num_elements: 90

  num_layers: 4        Equiformer 块（网络层数）的堆叠深度
  sphere_channels: 128
  attn_hidden_channels: 64      # [64, 96] This determines the hidden size of message passing. Do not necessarily use 96.
  num_heads: 8
  attn_alpha_channels: 64      # Not used when `use_s2_act_attn` is True.
  attn_value_channels: 16
  ffn_hidden_channels: 128
  norm_type: 'layer_norm_sh'   # ['rms_norm_sh', 'layer_norm', 'layer_norm_sh']

  lmax_list: [1]          球谐函数的最大角动量和磁量子数
  mmax_list: [1]
  grid_resolution: 18     # [18, 16, 14, None] For `None`, simply comment this line.

  num_sphere_samples: 16  球面采样点的数量

  edge_channels: 128
  use_atom_edge_embedding: True
  share_atom_edge_embedding: False      # If `True`, `use_atom_edge_embedding` must be `True` and the atom edge embedding will be shared across all blocks.
  distance_function: 'gaussian'
  num_distance_basis: 512      # not used

  attn_activation: 'silu'
  use_s2_act_attn: False      # [False, True] Switch between attention after S2 activation or the original EquiformerV1 attention.
  use_attn_renorm: True      # Attention re-normalization. Used for ablation study.
  ffn_activation: 'silu'     # ['silu', 'swiglu']
  use_gate_act: False      # [True, False] Switch between gate activation and S2 activation
  use_grid_mlp: True      # [False, True] If `True`, use projecting to grids and performing MLPs for FFNs.
  use_sep_s2_act: True      # Separable S2 activation. Used for ablation study.

  alpha_drop: 0.1          # [0.0, 0.1]
  drop_path_rate: 0.05     # [0.0, 0.05]
  proj_drop: 0.0

  weight_init: 'uniform'   # ['uniform', 'normal']
```

Loaded EquiformerV2_OC20 with 4418242 parameters.

The original setting

```
num_layers: 12
sphere_channels: 128
attn_hidden_channels: 64
num_heads: 8
attn_alpha_channels: 64
attn_value_channels: 16
ffn_hidden_channels: 128
norm_type: 'layer_norm_sh'

lmax_list: [6]
mmax_list: [2]
grid_resolution: 18

num_sphere_samples: 128
```

4h/epoch -> 10min/epoch

优化器参数

```
optim:
  batch_size:          64      # 6
  eval_batch_size:     64      # 6
  grad_accumulation_steps: 1      # gradient accumulation: effective batch size = `grad_accumulation_steps` * `batch_size` * (num of GPUs)
  load_balancing: atoms
  num_workers: 0
  lr_initial:          0.0004   # [0.0002, 0.0004], eSCN uses 0.0008 for batch size 96

  optimizer: AdamW
  optimizer_params:
    weight_decay: 0.001
  scheduler: LambdaLR
  scheduler_params:
    lambda_type: cosine
    warmup_factor: 0.2
    warmup_epochs: 0.1
    lr_min_factor: 0.01

  max_epochs: 30
  force_coefficient: 100
  energy_coefficient: 2
  clip_grad_norm: 100
  ema_decay: 0.999
  loss_energy: mae
  loss_force: l2mae

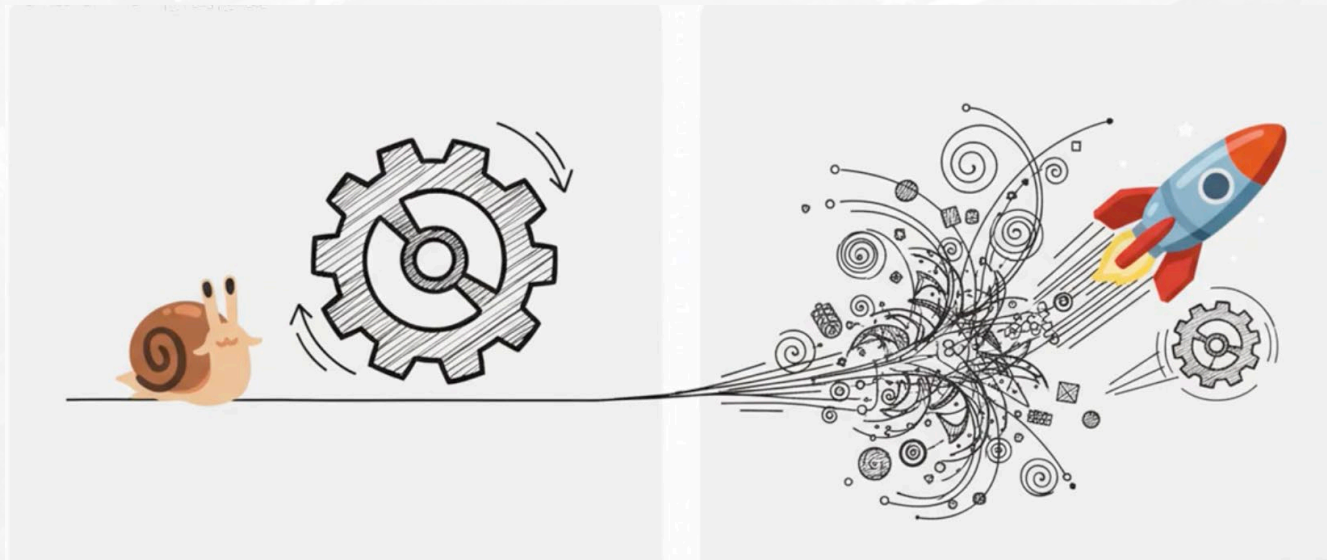
  eval_every: 157518
```

梯度裁剪阈值，防止梯度爆炸

04

训练结果 可视化分析

| 袁雅萱



200k 二肽数据集 · 能量与原子力回归

20 万

二肽结构

16 / 30

已完成 epoch

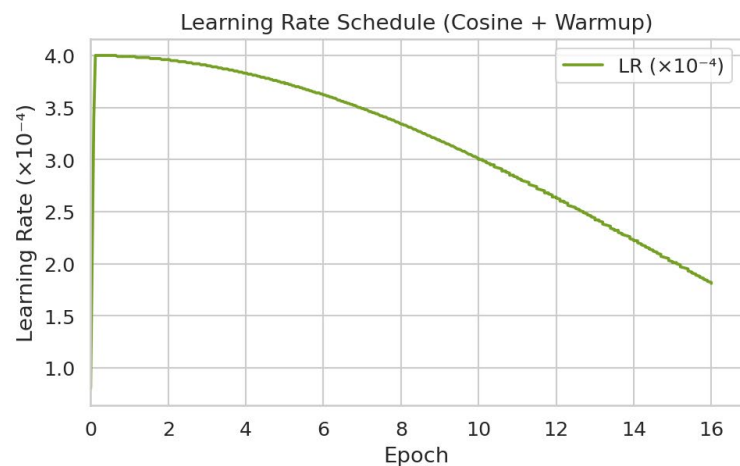
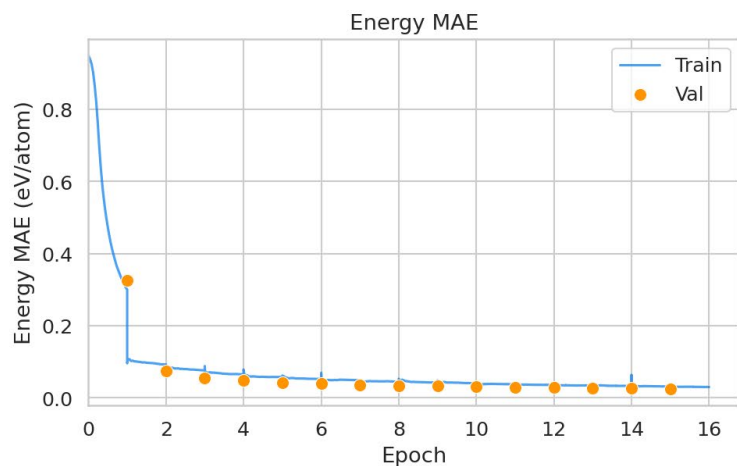
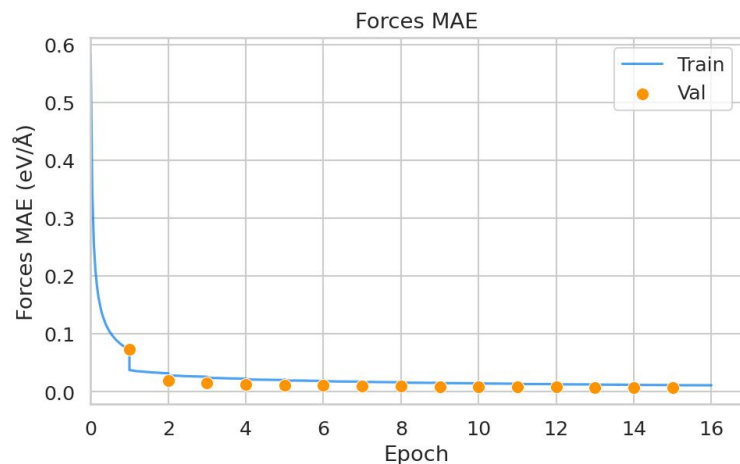
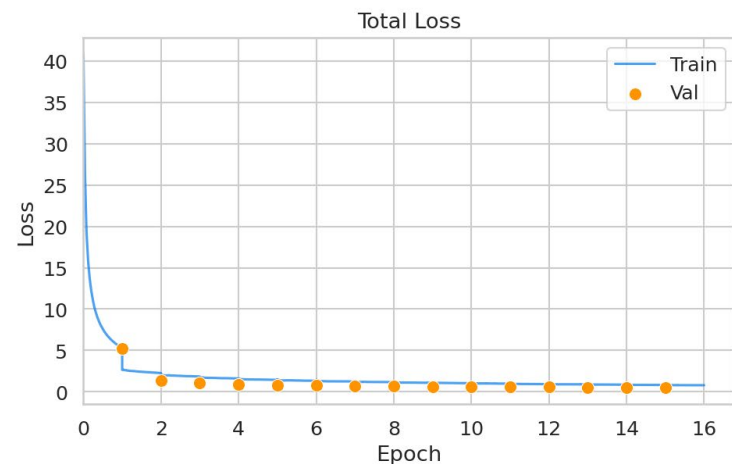
3.9 万

评测结构

0.999

力方向 cos

训练收敛



Loss 下降

40 $\rightarrow$ 0.8

最终 val 力 MAE

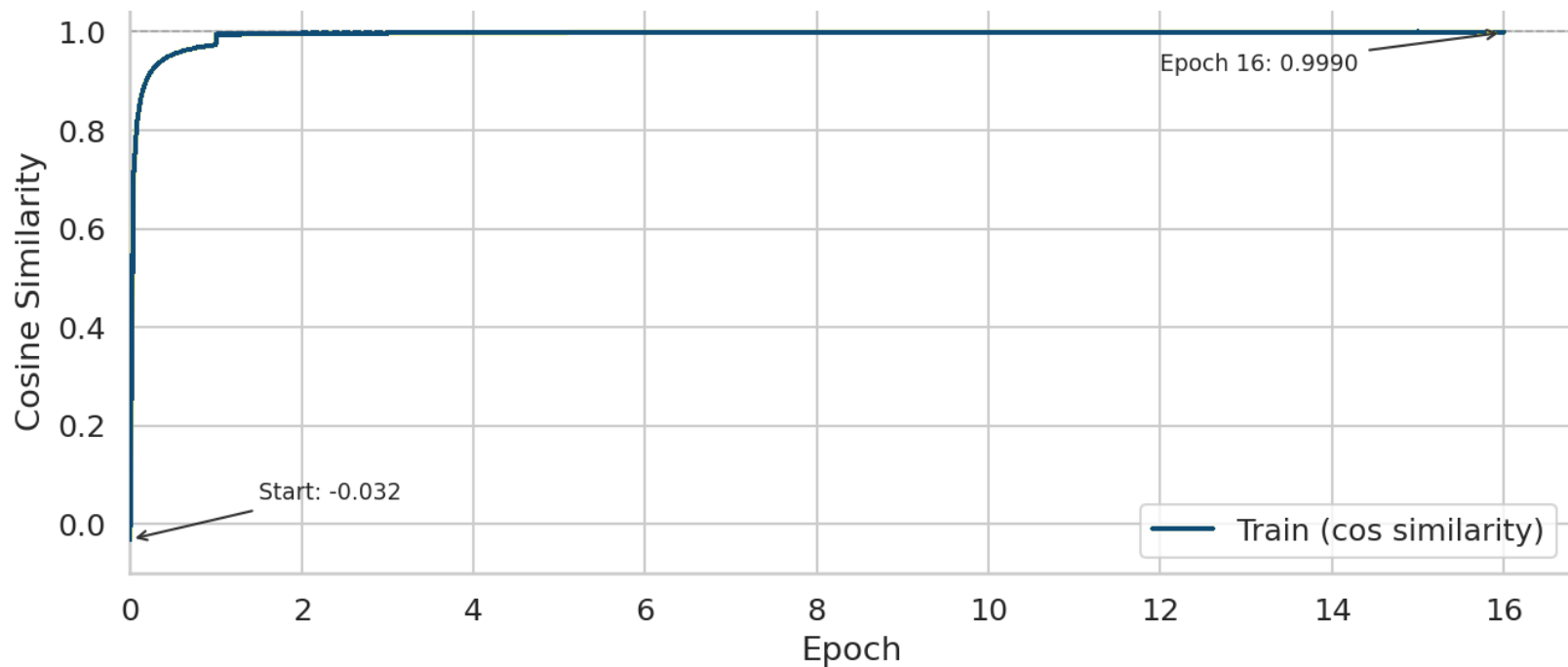
0.013 eV/Å

最终 val 能量 MAE

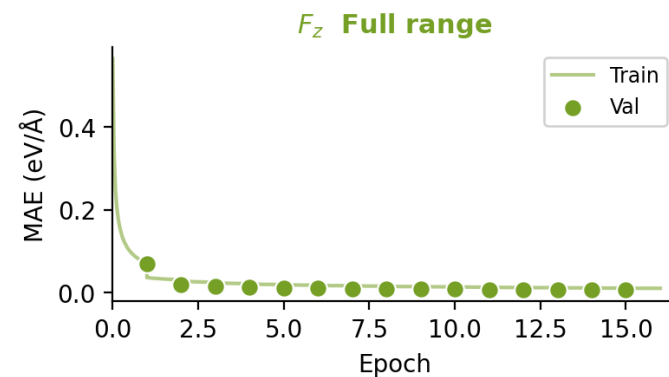
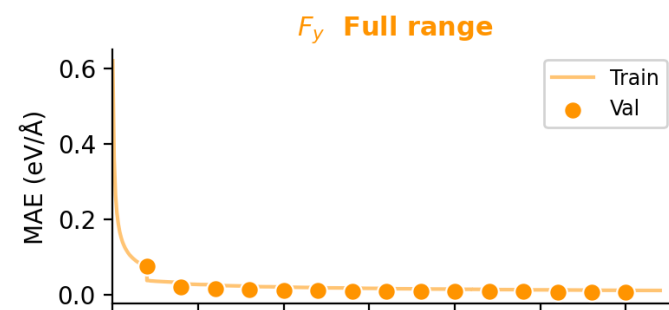
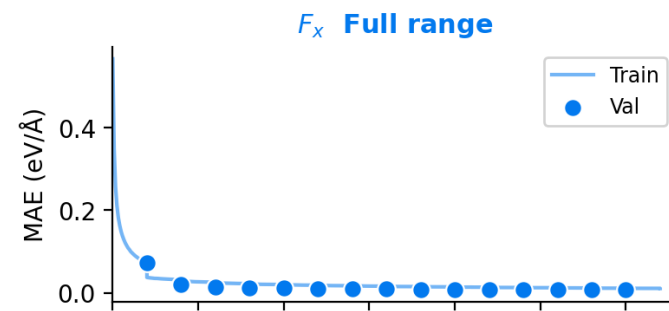
0.036 eV/atom

Val 紧跟 train, 无过拟合前 2 个 epoch 完成大部分学习 LR cosine 衰减稳定后期

力方向预测质量



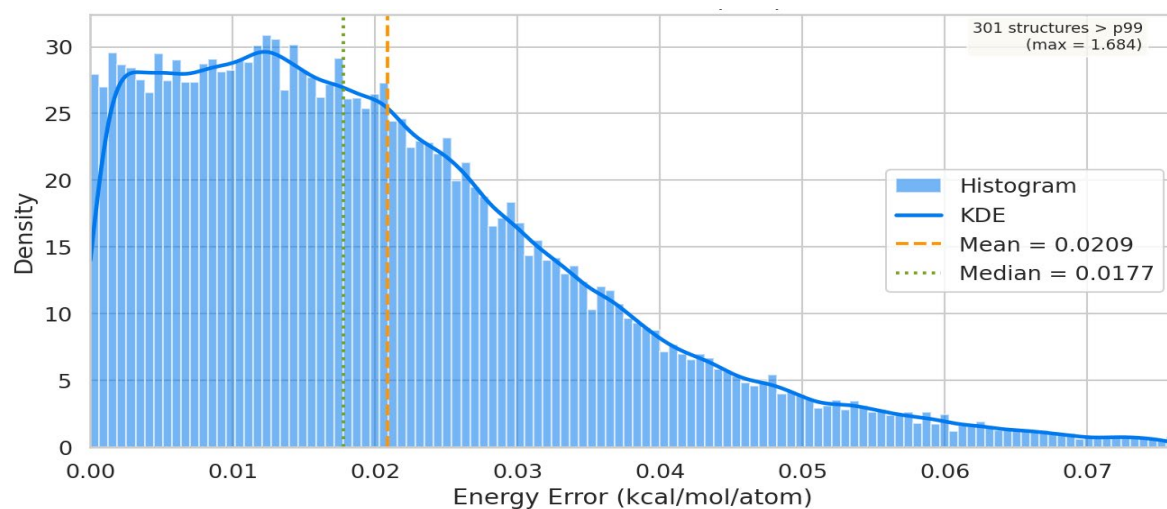
Cosine similarity: 只看预测力与真实力的方向是否一致 (无关大小)



单结构误差分布

Energy Error

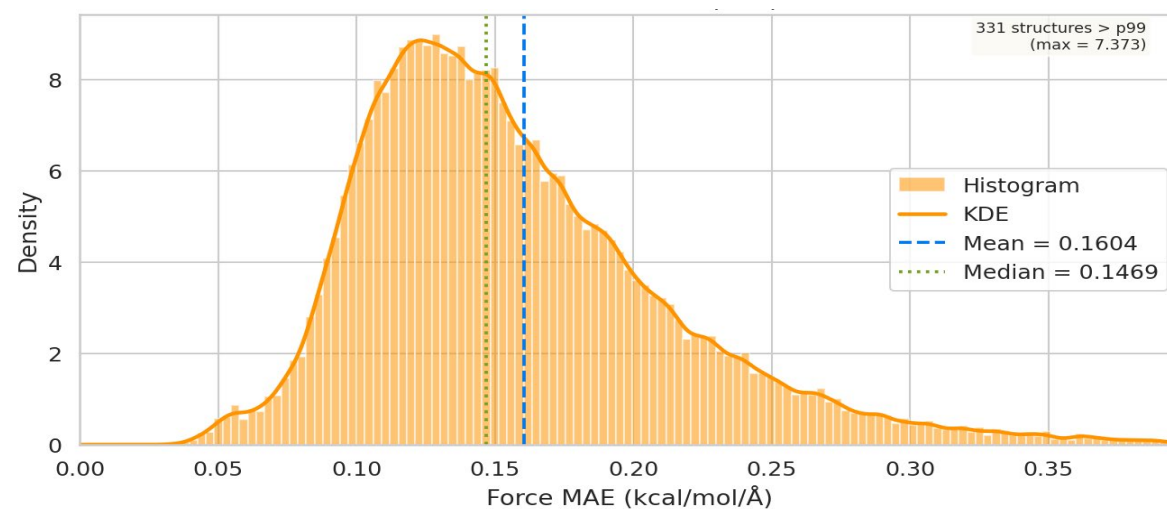
(kcal / mol / atom)



MEAN **0.0209** MEDIAN **0.0177** P99 **0.0725**

Force MAE

(kcal / mol / Å)



MEAN **0.1604** MEDIAN **0.1469** P99 **0.3782**

两个分布都是**右偏**：均值 > 中位数，少数大误差样本拉高了平均。



谢谢， 敬请批评指正！



清华大学
Tsinghua University