# A IMPLEMENTATION OF HAMNET

## A.1 FEATURIZATION

We use the identical featurization as Attentive FP (Xiong et al., 2020). In total, 39-dimensional atom features (including atom types, atom degree, indicators of aromaticity and chirality *et al*) and 10-dimensional bond features (including bond types, indicator of conjugation *et al*) are derived from the molecules. Specific details are shown in Table 1

Table 1: Featurization details (Table from (Xiong et al., 2020).)

| Atom Features | Size | Description |
|---|---|---|
| atom symbol | 16 | [B, C, N, O, F, Si, P, S, Cl, As, Se, Br, Te, I, At, metal] (one-hot) |
| degree | 6 | number of covalent bonds [0,1,2,3,4,5] (one-hot) |
| formal charge | 1 | electrical charge (integer) |
| radical electrons | 1 | number of radical electrons (integer) |
| hybridization | 6 | [sp, sp2, sp3, sp3d, sp3d2, other] (one-hot) |
| aromaticity | 1 | whether the atom is part of an aromatic system [0/1] (one-hot) |
| hydrogens | 5 | number of connected hydrogens [0,1,2,3,4] (one-hot) |
| chirality | 1 | whether the atom is chiral center [0/1] (one-hot) |
| chirality type | 2 | [R, S] (one-hot) |
| **Bond Features** | **Size** | **Description** |
| bond type | 4 | [single, double, triple, aromatic] (one-hot) |
| conjugation | 1 | whether the bond is conjugated [0/1] (one-hot) |
| ring | 1 | whether the bond is in ring [0/1] (one-hot) |
| stereo | 4 | [StereoNone, StereoAny, StereoZ, StereoE] (one-hot) |

## A.2 HAMILTONIAN ENGINE

Table 2 concludes all details of the implementation of the Hamiltonian Engine.

Table 2: Hyperparameters in the Hamiltonian Engine.

| Module | Hyperparameter | Value |
|---|---|---|
| Pos. & Mom. Init. | GCN layers | 2 |
| | GCN hidden dim. | 128 |
| | LSTM hidden dim. | 128 |
| | LSTM output dim. | 32 |
| Ham. Dyn. | $d_f$ | 32 |
| | depth ($T$) | 20 |
| | step size ($\eta$) | 0.025 |
| | $\dim(\boldsymbol{W}_U)$ | $32 \times d_f$ |
| | $\dim(\boldsymbol{W}_T)$ | $32 \times d_f$ |
| | $\dim(\boldsymbol{W}_\phi)$ | $32 \times d_f$ |
| | loss factor ($\lambda$) | 5 |
| Training | learning rate | 1e-4 |
| | exponential decay | 4e-5 |
| | batch size | 32 |
| | dropout rate | 0.2 |

**Positions & Momentum Initialization.** For the initialization of positions and momentums $\boldsymbol{q}^{(0)}, \boldsymbol{p}^{(0)}$, we use 2 layers of GCNs of $\dim = 128$ and connect $\boldsymbol{v}$, $\boldsymbol{h}^1$ and $\boldsymbol{h}^2$ as the input of the LSTM. In the LSTM, the hidden dimensionality is set as 128, and the output dimensionality is 32, which equals to $d_f$.

**Hamiltonian Dynamics.** Except for the experiments of testing the effects of corresponding hyperparameters, we set $T = 20$, $d_f = 32$, and $\eta = 0.025$. The rank of the quadratic forms are all set as 32, *i.e.*

$$\boldsymbol{W}_U, \boldsymbol{W}_T, \boldsymbol{W}_\phi \in \mathbb{R}^{32\times 32}. \tag{1}$$

In the loss $L_{HE}$, we use $\lambda = 5$.

**Training details.** We train the Hamiltonian Engine with learning rate $1 \times 10^{-4}$ with exponential decay $\delta = 4 \times 10^{-5}$ per step, *i.e.* $\eta_{t+1} = (1 - \delta)\eta_t$. The batch size is set as 32 and a dropout rate 0.2 [1] is used.

### A.3 Fingerprint Generator

Table 3 summarizes the hyperparameters we used in the Fingerprint Generator, including model depths, hidden dimensionalities and training configurations. Note that the choices of hyperparameters are roughly the same than those in (Xiong et al., 2020).

Table 3: Hyperparameters in the Fingerprint Generator.

| Hyperparameter | Tox21 | QM9 | Lipop | FreeSolv | ESOL |
|---|---|---|---|---|---|
| # atom-level attentive layers ($L$) | 3 | 2 | 2 | 2 | 2 |
| # global attentive layers ($M$) | 3 | 2 | 4 | 2 | 2 |
| atom hidden dim. ($\dim(\boldsymbol{h})$)<br>atom message dim. ($\dim(\boldsymbol{m})$)<br>bond information dim. ($\dim(\boldsymbol{f})$)<br>global hidden dim. ($\dim(\boldsymbol{h}_g)$)<br>global message dim. ($\dim(\boldsymbol{s}_g)$) | 200 | 280 | 200 | 120 | 200 |
| learning rate | 1e-4 | 1e-5 | 1e-4 | 1e-4 | 1e-4 |
| exponential decay | 1e-4 | 1e-5 | 1e-4 | 1e-3 | 1e-3 |
| batch size | 20 | 32 | 20 | 16 | 16 |
| dropout rate | 0.0 | 0.5 | 0.5 | 0.5 | 0.5 |

## B Implementation of HamNet Variants

### B.1 Conformation Reconstruction

The ablation analysis in the conformation reconstruction task is implemented as follow:

- Ham. Eng. *w/o LSTM* takes the output of GCNs ($\tilde{\boldsymbol{q}}, \tilde{\boldsymbol{p}}$) and transform them to corresponding dimensions with a tanh-activated single layer perceptron to derive initial positions and momentums of atoms ($\boldsymbol{q}^{(0)}, \boldsymbol{p}^{(0)}$), i.e.

$$\boldsymbol{q}^{(0)} \oplus \boldsymbol{p}^{(0)} = \tanh(\boldsymbol{W}^T(\tilde{\boldsymbol{q}} \oplus \tilde{\boldsymbol{p}})).$$

- Ham. Eng. *w/o dyn.* removes the Hamiltonian dynamics and directly treats the initial positions and momentums $\boldsymbol{q}^{(0)}, \boldsymbol{p}^{(0)}$ as the final quantities $\boldsymbol{q}^{(T)}, \boldsymbol{p}^{(T)}$, i.e.

$$\boldsymbol{q}^{(T)} = \boldsymbol{q}^{(0)}, \boldsymbol{p}^{(T)} = \boldsymbol{p}^{(0)}.$$

- Ham. Eng. *w/o* $\Phi$ removes the dissipation when iterating in the Hamiltonian Engine, i.e. $\Phi \equiv 0$.
- Ham. Eng. *w/o ADJ-3* removes ADJ-3 loss in the optimization, *i.e.* $\lambda = 0$.

[1] NOT the keep probability.

### B.2 MOLECULAR PROPERTY PREDICTION

- HamNet *(w/o conf.)* takes no account of conformation when generating the fingerprints ($d_f = 0$). That is, the Hamiltonian Engine is removed as well as all $\boldsymbol{q}, \boldsymbol{p}$-dependent components in the Fingerprint Generator. Note that the architecture of HamNet w/o conf. is similar to, but not the same as Attentive FP, as the *Bond information* $\boldsymbol{f}$ still functions.
- HamNet *(real conf.)* also removes the Hamiltonian Engine, and takes real 3D conformations as inputs and projects them linearly to $2 \times d_f$-dimensional vectors, which are then regarded as $\boldsymbol{q} \oplus \boldsymbol{p}$. On datasets without available conformations, we use conformations generated by the RDKit package.

## REFERENCES

Z. Xiong, D. Wang, X. Liu, F. Zhong, X. Wan, X. Li, Z. Li, X. Luo, K. Chen, H. Jiang, and M. Zheng. Pushing the boundaries of molecular representation for drug discovery with the graph attention mechanism. *Journal of medicinal chemistry*, 63(16):8749–8760, 2020.