

Article

GeoRC: Geometry-Aware Residual Correction for Three-Dimensional Molecular Graph Prediction

Zhiwang Dai ^{1,†}, Fangqi Zhou ^{2,†}, Yu Wang ², Hao Huang ², Zijian Wang ¹  and Jianbo Zheng ^{2,*}

¹ School of Information and Intelligent Science, Donghua University, Shanghai 201620, China; zhiwangdai@mail.dhu.edu.cn (Z.D.)

² High Speed Aerodynamics Institute, China Aerodynamics Research and Development Center, Mianyang 621000, China

* Correspondence: zhengjianbo1995@163.com

[†] These authors contributed equally to this work.

Abstract

Three-dimensional graph neural networks predict molecular and atomistic properties from interatomic geometry. Existing geometric backbones are compared mainly by replacing one architecture with another, so any accuracy gain is attributed to the new design as a whole. GeoRC augments an existing graph-level predictor with a geometry-aware residual branch while preserving its base prediction path. Radial distance features and element-wise edge gating build directed local interactions, gated recurrent unit updates refine the node states, and graph-level residual fusion combines the resulting candidate correction with the base prediction through a bounded, magnitude-conditioned screening map and a learnable global scale. The complete GeoRC-SphereNet model lowers MD17 force mean absolute error (MAE) for all eight evaluated molecules by 3.14% on average. In the repeated component-deletion study, removing the radial basis features increases mean force MAE for all eight molecules, which identifies the geometric encoding as the component with the broadest effect, and the OC20 macro energy-MAE reduction is 0.0639% across three paired runs. The results support GeoRC as an effective residual design for molecular and atomistic property prediction, and the proposed method provides a useful reference for future work in this area.

Keywords: 3D molecular graphs; residual learning; geometric message passing; graph neural networks; molecular energy and force prediction



Academic Editors: Sergei Odintsov and Calogero Vetro

Received: 9 September 2026

Revised: 13 October 2026

Accepted: 15 October 2026

Published: 18 October 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article

distributed under the terms and

conditions of the [Creative Commons](#)

[Attribution \(CC BY\) license](#).

1. Introduction

Machine learning on graphs is now a standard tool in computational chemistry and materials modeling. Atoms and their interactions map onto nodes and edges, and reference quantum-chemical calculations are expensive enough that a learned surrogate is worth training for screening, force-field construction, and catalyst discovery [1–6]. Quantum-chemical targets depend on geometry beyond connectivity. Conformers with an identical bond graph differ in energy, and what separates them is geometric, namely interatomic distances, bond angles, torsions, and, for extended systems, periodic spatial relations [7–9]. A predictor must also return the same prediction when the molecule is rotated, translated, or renumbered [10–12].

This has produced a sequence of increasingly expressive geometric backbones. SchNet introduced continuous-filter convolutions over interatomic distances, DimeNet

and DimeNet++ added directional message passing through bond angles, SphereNet extended the description to distances, angles, and torsions, and GemNet and PaiNN refined directional and vectorial interaction blocks [7–9,13–15]. Equivariant designs such as tensor-field networks, SE(3)-equivariant transformers, E(n)-equivariant networks, NequIP, MACE, and Allegro build these constraints into the representation itself [10–12,16–18]. Separately, additive learning in gradient boosting and deep residual networks showed that a strong predictor can be improved by fitting the error it leaves behind [19–21].

Existing work faces three difficulties. The first is attribution: replacing a backbone changes its representation, optimization, and readout together [7,9,13,14]. The second is correction strength: a residual branch produces a candidate for each sample, and a shared global scale alone applies the same attenuation to all candidates [22,23]. The third is scale: molecular properties differ in units and numerical magnitude, so a dimensionless reference is needed to describe corrections across targets [1,6,24]. GeoRC retains the base prediction and adds a correction learned from local geometry. In the controlled frozen-backbone configuration, optimization is confined to the correction branch. The jointly trainable configuration optimizes both paths and is reported separately. Fixed fuzzy screening maps the target-normalized magnitude of each candidate to a sample-specific retention coefficient (Figure 1). A learned global scale controls the overall amplitude. The candidate network is supervised through the final fused prediction. Matched molecule-level comparisons place this fusion stage alongside direct residual prediction, a learnable scalar, and a learned gate, and the component-deletion study reports how the correction changes when each module is removed.

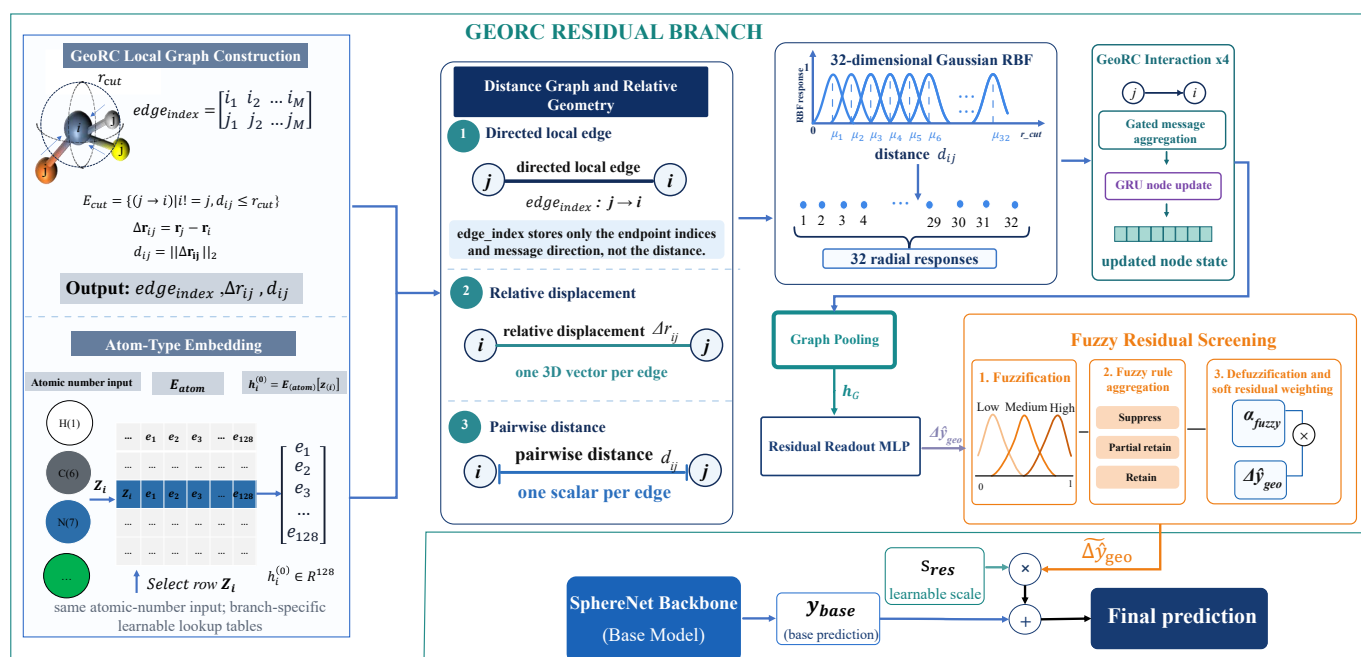


Figure 1. Overview of GeoRC. The base predictor provides a graph-level estimate. A parallel geometry-aware branch builds a local graph, updates node states through four interaction stages, and produces a candidate residual by graph pooling and readout. Fixed fuzzy screening applies magnitude-dependent attenuation to this candidate. A learnable global scale controls the screened correction before it is added to the base estimate.

The contributions of this work are twofold and are listed below.

1. We formulate GeoRC as a residual branch that attaches to an existing graph-level three-dimensional predictor and preserves its base output path. The matched frozen-backbone configuration isolates the added branch; the jointly trainable configuration

measures optimization of both paths. The correction path is closed by a bounded screening stage that maps the target-normalized candidate magnitude to a retention coefficient, and we evaluate this fusion option against direct, scalar, and learned-gate mappings under matched controls, where its benefit appears on particular molecules and targets rather than universally.

2. We design a distance-gated interaction module. Source-node states and radial distance features form directed edge features, which are gated element-wise, summed at the target node, and used to update the node state through a GRU and a scaled residual connection. We assess the resulting correction on molecular energy and force prediction, multiple molecular-property targets, a periodic energy task, and a second backbone, and we report the component effects by molecule and prediction target.

2. Related Work

2.1. Geometric Molecular Graph Neural Networks

SchNet introduced continuous-filter convolutions over interatomic distances [7]. DimeNet added directional message passing through angles, and DimeNet++ made the same principle cheaper [8,13]. SphereNet used spherical message passing with distances, angles, and torsions, and reported on QM9, MD17, and OC20 [9], which is why it serves as the backbone here. GemNet and GemNet-OC developed directional interaction blocks further [14,25], and PaiNN, NequIP, MACE, and Allegro treat equivariance and scalable atomistic interactions more explicitly [12,15,17,18]. Toolkits such as SchNetPack provide a reproducible training and evaluation interface for these models [26]. A related line of force-field work encodes local environments through atom-centered symmetry functions and Gaussian approximation potentials [27,28], couples such encodings to differentiable energy models [29], or uses symmetry-adapted quantum-chemical features [30].

These models are compared by replacing one architecture with another, which attributes an accuracy gain to the new design as a whole. Once a backbone is trained, its representation is available for reuse. In the frozen-backbone configuration, GeoRC learns the remaining error in a separate branch while keeping the base message-passing parameters fixed and initializing from a matched baseline checkpoint. This places GeoRC in the role of an adapter to an existing 3D backbone. The interface is defined at the graph level so that the correction equations apply to a substituted backbone, and forces are obtained by differentiating the fused energy, so the added branch is evaluated through the same energy-to-force relation as the base model.

Local three-dimensional representations also support structural bioinformatics. GraphaRNA models RNA interatomic interactions with a graph neural network and diffusion model [31]. RNAfitme models residue conformations in fixed-backbone RNA [32], while RNArefine performs AI-guided atomic-level RNA refinement [33]. DNATCO v5.0 provides nucleic-acid structure analysis, geometric validation, and torsion restraints that support refinement [34]. These applications connect local geometry to structural modeling; GeoRC uses local geometry to correct a predictor's graph-level output.

2.2. Gated Graph Updates and Residual Learning

Gated graph neural networks use recurrent gates to regulate information flow over several propagation steps [22]. The GRU update in GeoRC follows the recurrent-state idea from sequence models, with a graph-specific scatter aggregation before the recurrent cell [35,36]. Layer normalization stabilizes hidden representations using per-sample statistics [37,38], and LayerScale gives learnable control over the magnitude of residual transformations [39]. These mechanisms are normally used inside a backbone, where their effect is folded into the learned representation. We combine them in a small branch whose

output is the residual of a base predictor, so the branch has a well-defined target and an additive connection to the original estimate. A paired residual design of this kind also makes per-target behavior visible, since the correction for each target is read directly at the output.

2.3. Rule-Based Mappings as a Fusion Option

Fuzzy-set theory replaces a crisp yes/no classification with a continuous membership degree in $[0, 1]$ [40], and the Mamdani formulation showed how linguistic if-then rules over such memberships can be aggregated into a usable output surface [41]. The Takagi–Sugeno variant replaced Mamdani’s output membership functions with zero- or first-order functions of the input, giving a compact differentiable output that composes with gradient-based systems [42]. ANFIS then embedded such a rule base in a layered computational graph and trained it alongside learned components while keeping the rule semantics [43].

Fuzzy systems use fixed and learned rule parameters [42,43]. GeoRC specifies three overlapping Gaussian memberships and three zero-order consequents as one bounded, magnitude-conditioned fusion mapping. A learned sigmoid gate provides a trainable comparison using the same normalized input. The fixed membership map can be plotted independently of the candidate network. Its target scale τ_y is estimated from training data, while the global residual scale s_{res} and candidate network are optimized through the final task loss. The matched controls in Section 5.1.1 compare these fusion mappings empirically.

3. Method

3.1. Overview of the GeoRC Pipeline

GeoRC runs as a branch alongside a base predictor (Figure 1). The base path provides a graph-level estimate, and the correction path uses the same atomic inputs to produce a screened additive correction. SphereNet is used in the main benchmarks, and the SchNet evaluation in Section 5.1.3 tests the same graph-level interface on a second backbone.

The correction branch has four stages. Graph construction sets the initial atom states from learnable element embeddings and builds a local radius graph from the coordinates, using minimum-image displacements for periodic systems. Four distance-gated interaction modules then pass geometric information between atoms. Each module builds directed edge features from the source-node state and a radial basis expansion of the interatomic distance, gates each message element-wise, sums the messages at the target node, and updates the node state through a GRU with a scaled residual connection. Graph readout pools the node states by sum, by mean, and with a log-scaled atom count, and a readout multilayer perceptron turns this into the candidate correction $\Delta\hat{y}_{\text{geo}}$. Fuzzy residual screening converts the candidate into a dimensionless score, passes it through fixed low, medium, and high membership rules, and produces a retention weight α_{fuzzy} between zero and one that scales the candidate before fusion.

Radial basis functions (RBFs) encode interatomic distances. A gated recurrent unit (GRU) updates node states from aggregated messages. A multilayer perceptron (MLP) maps pooled node features to the candidate residual.

The last three stages each answer a different question. The interaction stack sets what geometric information the correction can express, the readout turns that information into a single number, and fuzzy screening sets how strongly that number affects the base estimate. The branch filters at two different resolutions. The learned sigmoid gate acts on individual directed edges, and the fuzzy weight acts once on the graph-level candidate. The following subsections define the residual objective and then each module in turn.

3.2. Problem Definition and Residual Objective

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{Z}, \mathbf{R}, \mathbf{C})$ denote a molecular or atomistic graph, where $\mathcal{V}$ is the set of nodes, $\mathcal{E}$ is the set of edges, $\mathbf{Z}$ contains atom or element types, $\mathbf{R} \in \mathbb{R}^{N \times 3}$ contains positions, and $\mathbf{C} \in \mathbb{R}^{3 \times 3}$ is the periodic cell matrix, supplied for periodic systems. A compatible base predictor f_θ produces a graph-level estimate

$$\hat{y}_{\text{base}} = f_\theta(\mathcal{G}). \quad (1)$$

For a target y , the ideal residual is $r^* = y - \hat{y}_{\text{base}}$. GeoRC first uses a correction learner g_ϕ that receives the atomistic graph and local geometry, giving

$$\Delta \hat{y}_{\text{geo}} = g_\phi(\mathcal{G}, \mathbf{R}). \quad (2)$$

The candidate correction is then calibrated by the fuzzy residual-screening stage. The calibrated correction is fused with the base prediction as

$$\hat{y} = \hat{y}_{\text{base}} + s_{\text{res}} \alpha_{\text{fuzzy}} \Delta \hat{y}_{\text{geo}} = \hat{y}_{\text{base}} + s_{\text{res}} \tilde{\Delta \hat{y}}_{\text{geo}}, \quad (3)$$

where s_{res} is a learned residual scale and $\tilde{\Delta \hat{y}}_{\text{geo}}$ is the fuzzy-calibrated correction. The residual readout proposes a correction, the fuzzy stage controls how much of that proposal is retained, and the final addition preserves the original base-model prediction path.

The product separates the learned signed geometric candidate from a prescribed magnitude-dependent attenuation function. Target normalization supplies a dimensionless input, and overlapping memberships interpolate smoothly between fixed retention levels. A positive screening coefficient preserves the candidate sign before multiplication by the global residual scale. The complete fused prediction supervises the candidate network and scale. This decomposition makes the magnitude response an explicit part of the fusion design; Section 5.1.1 measures its empirical effect against direct residual prediction, learnable scalar weighting, and a learned gate under matched training conditions.

The training target is the final fused prediction. For scalar tasks, the executed objective is

$$\mathcal{L}_{\text{scalar}} = \ell_{\text{task}}(y, \hat{y}), \quad (4)$$

where ℓ_{task} is Smooth L1 for the local QM9 runs and L1 (mean absolute error) on normalized energy for OC20. For MD17, the fused energy prediction is differentiated with respect to coordinates, $\hat{\mathbf{F}} = -\nabla_{\mathbf{R}} \hat{E}$, and the joint objective is

$$\mathcal{L}_{\text{force}} = \text{L1}(E, \hat{E}) + 1000 \text{L1}(\mathbf{F}, \hat{\mathbf{F}}), \quad (5)$$

QM9 uses Smooth L1 on the fused scalar prediction, OC20 uses L1 on normalized energy, and MD17 uses energy L1 plus 1000 times force L1. For MD17, forces are the negative coordinate gradient of the complete fused energy within fixed graph neighborhoods. Differentiation includes the coordinate dependence of the candidate correction and the screening coefficient. The force-over-energy loss weight (WoFE) is 1000.

3.3. Atom-State Initialization and Radius-Graph Construction

The correction branch begins by mapping each atom to a learnable element embedding, so that chemically distinct species start from distinct hidden states. These embeddings form the layer-zero node states of the interaction stack.

GeoRC uses the same atomic numbers and coordinates as the base predictor, with a separate learnable element-embedding table for the correction branch. Within a mini-batch, the graph-assignment vector identifies the molecule containing each atom. Radius-graph

construction connects atoms within the same graph, and sum pooling, mean pooling, and atom-count computation are performed separately for each graph.

Connectivity for the branch is derived from coordinates. For a cutoff radius r_c , a directed edge $j \rightarrow i$ is created whenever atom j lies within r_c of atom i , and the QM9 construction additionally caps the neighbor count per atom to bound the edge budget. For periodic systems, the neighbor search operates on minimum-image displacements, so an atom interacts with the nearest periodic replica of each neighbor. Graph construction is kept separate from message evaluation throughout the implementation, since the edge index records which endpoints communicate while the distances and radial basis values are recomputed from the current coordinates and, for periodic data, the cell matrix. This separation keeps every quantity traceable and identifies exactly which quantities change when coordinates move, which matters for force prediction, where the energy is differentiated with respect to positions.

3.4. Distance-Gated Interaction Block

Node states $\mathbf{H}^{(l)} \in \mathbb{R}^{N \times d_h}$ are initialized from learnable element embeddings at layer $l = 0$ and updated through each interaction stage (Figure 2). For each directed radius-graph edge $j \rightarrow i$, where $\mathcal{N}(i)$ denotes the set of neighbors of node i within the cutoff radius, the edge index identifies source node j and target node i . In this implementation, $i = \text{edge_index}[0]$ holds the target index for the scatter-sum and $j = \text{edge_index}[1]$ holds the source index for gathering the source state. The edge index supplies the endpoint indices, and the geometry is computed from coordinates. Let $\mathbf{r}_{ij}$ denote the direct displacement for molecular systems, and the minimum-image displacement for periodic systems. The scalar distance is $d_{ij} = \|\mathbf{r}_{ij}\|_2$, and a Gaussian radial basis expansion provides $\phi(d_{ij})$. At interaction layer l , the source state and the radial encoding form the per-edge feature

$$\mathbf{x}_{ij}^{(l)} = [\mathbf{H}_j^{(l)}; \phi(d_{ij})]. \quad (6)$$

Thus, $\mathbf{x}_{ij}^{(l)} \in \mathbb{R}^{160}$; stacking the features of all M directed edges gives a matrix in $\mathbb{R}^{M \times 160}$. The same per-edge feature is processed by parallel message and gate paths,

$$\tilde{\mathbf{m}}_{ij}^{(l)} = \text{MLP}_{\text{msg}}(\mathbf{x}_{ij}^{(l)}), \quad \mathbf{g}_{ij}^{(l)} = \sigma(\text{MLP}_{\text{gate}}(\mathbf{x}_{ij}^{(l)})), \quad (7)$$

where $\sigma(\cdot)$ is the sigmoid function and MLP denotes a multilayer perceptron. The per-edge gate $\mathbf{g}_{ij}^{(l)} \in (0, 1)^{d_h}$ modulates the message element-wise via the Hadamard product ($\odot$), so that $\mathbf{m}_{ij}^{(l)} = \mathbf{g}_{ij}^{(l)} \odot \tilde{\mathbf{m}}_{ij}^{(l)}$. The gated messages are scatter-summed according to the target index.

For hidden width 128 and 32 radial channels, each directed edge has 160 input features. The message path applies LayerNorm over these features, a 160-to-128 linear map, a sigmoid linear unit (SiLU), and a 128-to-128 linear map. The parallel gate path applies LayerNorm over the same 160 features, a 160-to-128 linear map, and sigmoid. Both paths produce 128 components per directed edge. The target-node aggregation is

$$\mathbf{a}_i^{(l)} = \sum_{j \in \mathcal{N}(i)} \mathbf{m}_{ij}^{(l)}. \quad (8)$$

A GRUCell combines the aggregated incoming signal with the previous node state. The layer then applies a scaled residual state update and layer normalization, written as

$$\tilde{\mathbf{H}}_i^{(l)} = \text{GRUCell}(\mathbf{a}_i^{(l)}, \mathbf{H}_i^{(l)}), \quad \mathbf{H}_i^{(l+1)} = \text{LayerNorm}(\mathbf{H}_i^{(l)} + \gamma_l [\tilde{\mathbf{H}}_i^{(l)} - \mathbf{H}_i^{(l)}]). \quad (9)$$

Here, γ_l is a learnable scalar parameter per layer, initialized to 0.2, that governs how strongly the GRU state difference is mixed into the current layer state. This design separates directed edge communication from node-state retention, so that the message path models what arrives at a target node while the GRU determines how much of that information is retained [22,37,39].

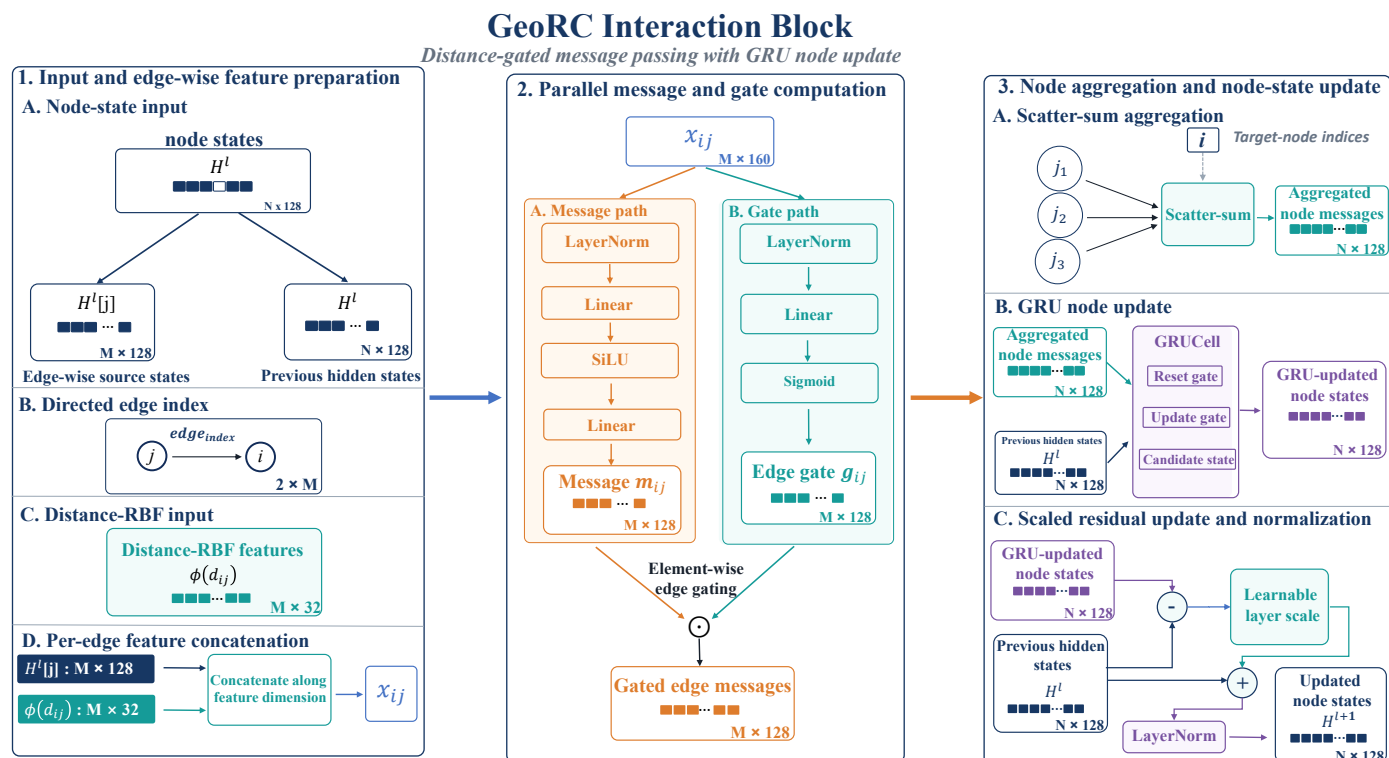


Figure 2. GeoRC interaction module. Source-node states and radial distance features form directed edge features. Message and gate paths produce element-wise gated messages, which are scatter-summed at the receiving node. A GRU uses the aggregated message and the current node state to update the node representation, followed by a scaled layer residual update. Layer normalization produces the updated node states for the next interaction stage; the fourth stage supplies the graph pooling and readout.

3.5. Graph Readout and Residual Fusion

After the fourth interaction stage, GeoRC forms a graph-level representation from sum pooling, mean pooling, and a log-scaled node count, giving

$$\mathbf{h}_G = \left[\sum_i \mathbf{H}_i ; \frac{1}{N} \sum_i \mathbf{H}_i ; \log(1 + N) \right]. \quad (10)$$

The residual readout MLP maps $\mathbf{h}_G$ to $\Delta \hat{y}_{\text{geo}}$, and Equation (3) defines the final prediction from that output. Sum pooling is sensitive to graph size and preserves the total signal magnitude, while mean pooling normalizes by atom count and captures per-atom averages, so retaining both gives the readout access to size-dependent and size-invariant graph statistics simultaneously. Both operations are invariant to node ordering. The base prediction is combined with the branch output only at the final additive fusion, so this MLP depends on geometry alone.

At hidden width 128, the pooled vector in Equation (10) contains 257 components: 128 from sum pooling, 128 from mean pooling, and one from the log-scaled atom count. The residual readout applies LayerNorm, Linear, SiLU, Linear, SiLU, and Linear in that order. The final linear layer produces one signed candidate correction per graph.

3.6. Fuzzy Residual Screening and Soft Calibration

Let the geometry-aware candidate correction be the signed scalar produced by the residual readout. Fuzzy screening uses its target-normalized magnitude as a bounded input:

$$e = \frac{|\Delta\hat{y}_{\text{geo}}|}{\tau_y + |\Delta\hat{y}_{\text{geo}}|} \in [0, 1], \quad (11)$$

The positive target-specific scale τ_y is estimated from the training split. This normalization makes the membership input dimensionless.

The fuzzification stage assigns low, medium, and high membership degrees. Using Gaussian memberships, each degree is

$$\mu_k(e) = \exp\left[-\frac{(e - c_k)^2}{2\sigma_k^2}\right], \quad k \in \{\text{low, medium, high}\}, \quad (12)$$

The centers are $c_{\text{low}} = 0.2$, $c_{\text{medium}} = 0.5$, and $c_{\text{high}} = 0.8$, with $\sigma_k = 0.20$ for all three memberships. These fixed, overlapping Gaussian functions describe low, medium, and high normalized magnitudes.

The fixed zero-order consequents are $q_{\text{low}} = 0.2$, $q_{\text{medium}} = 0.6$, and $q_{\text{high}} = 1.0$. Normalized weighted defuzzification interpolates between these retention levels:

$$\alpha_{\text{fuzzy}} = \frac{\sum_k \mu_k(e) q_k}{\sum_k \mu_k(e) + \varepsilon} \in [0, 1], \quad (13)$$

where ε is a small numerical constant. The screened correction is

$$\widetilde{\Delta\hat{y}}_{\text{geo}} = \alpha_{\text{fuzzy}} \Delta\hat{y}_{\text{geo}}. \quad (14)$$

The fuzzy stage applies a magnitude-conditioned retention coefficient to the candidate correction. Low, medium, and high memberships refer to ranges of target-normalized magnitude, and the weighted consequents determine how strongly the candidate enters additive fusion. The candidate network is optimized through the final prediction loss, so its magnitude adapts jointly with the learnable global residual scale. The coefficient specifies the response of the prescribed screening map. Its relationship with prediction benefit is evaluated on held-out frames in Section 5.1.1.

The membership centers, shared width, and rule consequents define a fixed screening function shared across targets. Overlapping Gaussian memberships interpolate between the prescribed retention levels. The training-derived target scale supplies unit normalization, and the candidate network and global residual scale adapt through the final prediction loss. This design fixes the shape of the attenuation function while retaining a trainable correction path. The complete retraining sensitivity analysis is reported in Section 5.1.2.

3.7. Invariance Scope, Learned Edge Gating, and Fuzzy Screening

The reported scalar-property branch takes interatomic distances as its geometric edge inputs. Distances are invariant to global translation and rotation, and the graph-level pooling in Equation (10) is invariant to atom ordering. GeoRC therefore preserves the relevant invariances for the scalar molecular and atomistic targets considered here. For MD17 force prediction, $\hat{\mathbf{F}} = -\nabla_{\mathbf{R}} \hat{E}$ inherits rotational covariance from the distance-only energy, and this holds wherever the radius-graph neighborhood is fixed, since the neighbor set changes discontinuously once an atom crosses the cutoff boundary. The resulting guarantee for the correction branch therefore holds locally, within each fixed neighborhood.

The sigmoid gate inside an interaction layer is a learned continuous edge weight. It acts on individual directed messages during representation construction, whereas fuzzy residual screening acts after graph pooling on the graph-level candidate correction. The two mechanisms therefore operate at different resolutions, with edge gating selecting local information and fuzzy screening calibrating the final residual proposal. Since e is derived from an invariant scalar correction, the fuzzy stage preserves the scalar invariances of the correction path while the learned edge gate retains its own role.

4. Experiments

4.1. Datasets and Task Settings

Three benchmarks stress different aspects of a correction module, namely molecular property regression, force-sensitive local geometry, and large-scale periodic materials graphs.

QM9 contains approximately 134 thousand small organic molecules with 12 quantum-chemical properties, comprising dipole moment, polarizability, HOMO and LUMO energies, energy gap, electronic spatial extent, zero-point vibrational energy, internal energies at 0 K and 298 K, enthalpy, free energy, and heat capacity [24]. MD17 contains molecular-dynamics trajectories with energies and atomic forces for eight molecules, namely aspirin, benzene, ethanol, malonaldehyde, naphthalene, salicylic acid, toluene, and uracil [23]. OC20 is a large catalyst dataset with 460k training structures and initial-to-relaxed-energy regression targets, evaluated under four validation regimes, namely in-distribution (ID), out-of-distribution adsorbate (OOD Ads), out-of-distribution catalyst (OOD Cat), and out-of-distribution in both (OOD Both) [44]. Table 1 lists the split sizes and evaluation budgets used in the paired reproduction.

Table 1. Dataset sizes and executed evaluation settings. WoFE denotes the force-over-energy loss weight.

Dataset/Task	Train	Validation	Test	Main Budget	Repeated Runs
QM9 scalar properties	110,000	10,000	10,831	500 epochs	10 ^a
MD17 molecule trajectories	1000	1000	dataset-specific	WoFE = 1000	10 ^b
OC20 IS2RE “all”	460,328	24,943–24,987 ^c	–	20 epochs	3

^a The repeated-run verification covers HOMO, LUMO, U0, and dipole moment; the 12-property table reports a single run. ^b Ten-run verification covers aspirin, naphthalene, and uracil. The main eight-molecule comparison reports one run; component deletion uses three runs per molecule and variant (Table 2). ^c The range covers ID, OOD Ads, OOD Cat, and OOD Both validation splits.

Table 2 reports the repeated component-deletion results for eight MD17 molecules.

Table 2. Repeated MD17 component-deletion ablation across eight molecules and five variants. Each entry is mean $\pm$ sample SD. Energy MAE is in kcal/mol and force MAE in kcal/(mol·Å).

Molecule	Variant	Energy MAE (kcal/mol)	Force MAE (kcal/(mol·Å))
Aspirin	Full GeoRC	0.207000 $\pm$ 0.004140	0.425400 $\pm$ 0.003871
Aspirin	No RBF	0.211999 $\pm$ 0.003574	0.429741 $\pm$ 0.004174
Aspirin	No gate	0.222395 $\pm$ 0.007680	0.425760 $\pm$ 0.003668
Aspirin	No GRU	0.220607 $\pm$ 0.001946	0.425620 $\pm$ 0.003695
Aspirin	No fuzzy	0.207140 $\pm$ 0.004081	0.425560 $\pm$ 0.003673
Benzene	Full GeoRC	0.074000 $\pm$ 0.002220	0.179200 $\pm$ 0.003046
Benzene	No RBF	0.074994 $\pm$ 0.002058	0.187352 $\pm$ 0.003048
Benzene	No gate	0.077053 $\pm$ 0.003274	0.179328 $\pm$ 0.002987
Benzene	No GRU	0.076091 $\pm$ 0.001401	0.179283 $\pm$ 0.003087
Benzene	No fuzzy	0.073984 $\pm$ 0.002235	0.179179 $\pm$ 0.003083

Table 2. Cont.

Molecule	Variant	Energy MAE (kcal/mol)	Force MAE (kcal/(mol·Å))
Ethanol	Full GeoRC	0.076000 ± 0.002584	0.231200 ± 0.004855
Ethanol	No RBF	0.075349 ± 0.002391	0.234486 ± 0.005159
Ethanol	No gate	0.074227 ± 0.003184	0.231302 ± 0.004853
Ethanol	No GRU	0.074903 ± 0.001735	0.231164 ± 0.004919
Ethanol	No fuzzy	0.076288 ± 0.002729	0.231255 ± 0.004795
Malonaldehyde	Full GeoRC	0.129500 ± 0.003367	0.380600 ± 0.007231
Malonaldehyde	No RBF	0.132923 ± 0.003257	0.384111 ± 0.008067
Malonaldehyde	No gate	0.128032 ± 0.004577	0.380820 ± 0.007203
Malonaldehyde	No GRU	0.126862 ± 0.004383	0.380719 ± 0.007347
Malonaldehyde	No fuzzy	0.130046 ± 0.003514	0.380226 ± 0.007106
Naphthalene	Full GeoRC	0.125200 ± 0.003130	0.125200 ± 0.002880
Naphthalene	No RBF	0.122094 ± 0.003099	0.132639 ± 0.003248
Naphthalene	No gate	0.127531 ± 0.003464	0.125331 ± 0.002824
Naphthalene	No GRU	0.121819 ± 0.001500	0.125245 ± 0.002844
Naphthalene	No fuzzy	0.125265 ± 0.003111	0.125253 ± 0.002833
Salicylic acid	Full GeoRC	0.138500 ± 0.004016	0.315200 ± 0.005989
Salicylic acid	No RBF	0.136749 ± 0.002993	0.320429 ± 0.005900
Salicylic acid	No gate	0.143074 ± 0.005757	0.315405 ± 0.005777
Salicylic acid	No GRU	0.136607 ± 0.002801	0.315324 ± 0.006107
Salicylic acid	No fuzzy	0.138554 ± 0.004084	0.316494 ± 0.006169
Toluene	Full GeoRC	0.103400 ± 0.002792	0.140800 ± 0.003238
Toluene	No RBF	0.105168 ± 0.002410	0.145097 ± 0.003498
Toluene	No gate	0.102554 ± 0.003001	0.140849 ± 0.003201
Toluene	No GRU	0.104368 ± 0.001987	0.140834 ± 0.003277
Toluene	No fuzzy	0.103386 ± 0.002861	0.141413 ± 0.002998
Uracil	Full GeoRC	0.130500 ± 0.004698	0.307000 ± 0.010438
Uracil	No RBF	0.131866 ± 0.004698	0.312477 ± 0.011186
Uracil	No gate	0.137140 ± 0.007054	0.307262 ± 0.010349
Uracil	No GRU	0.133254 ± 0.005510	0.307161 ± 0.010517
Uracil	No fuzzy	0.130576 ± 0.004767	0.308684 ± 0.010031

4.2. Experimental Setup

All comparisons are paired. Baseline and GeoRC use the same data source, split, initialization settings, batch settings, epoch budget, evaluation code, and identical numerical settings. For QM9, the main table uses a single local 500-epoch reproduction for all 12 properties. The repeated-run verification covers HOMO (highest occupied molecular orbital energy), LUMO (lowest unoccupied molecular orbital energy), U0, and dipole moment over ten independent runs; each GeoRC model is initialized from the best checkpoint of the baseline trained under the same conditions and split. For MD17, the stable-geometry setting uses a force-over-energy weight (WoFE) of 1000 and reports eight molecules in the main comparison, with additional ten-run paired repeats for aspirin, naphthalene, and uracil. The component-deletion study covers all eight molecules and five variants, with three runs for each molecule and variant (120 training runs); Table 2 reports mean and sample standard deviation for energy and force MAE. Its full GeoRC values provide a separate three-run reference from the main single-run comparison. For OC20 IS2RE (Initial Structure to Relaxed Energy), the full “all” training split is used for 20 epochs with two-GPU DDP (Distributed Data Parallel), per-GPU batch size 8, global batch size 16, and zero data-loader workers. The reported OC20 audit comprises three paired runs, each evaluated under four validation regimes: ID (in-distribution), OOD Ads (out-of-distribution adsorbate), OOD Cat (out-of-distribution catalyst), and OOD Both.

A SphereNet baseline is trained first, and its best validation checkpoint initializes the matched GeoRC run under the same data split, budget, and evaluation schedule.

GeoRC uses two optimization configurations. In the frozen-backbone configuration, the baseline parameters remain fixed and the correction branch is optimized. In the trainable-backbone configuration, the backbone and correction branch are jointly optimized, with the backbone using a lower learning rate and the correction using a smaller residual initialization. The main QM9 HOMO comparison uses the trainable configuration; the other eleven properties use the frozen configuration. The HOMO result therefore measures joint optimization, while the frozen rows provide controlled attribution to the added correction path.

Repeated runs are treated as an experimental factor. In the QM9 repeated verification, each run regenerates the train, validation, and test partition and re-draws the initialization and data-loader order. SphereNet and GeoRC share these conditions within each pair, so a paired difference compares the two methods under the same stochastic conditions. The QM9 verification therefore tests sensitivity to both split composition and optimization randomness, whereas the MD17 representative verification holds the split fixed and varies the optimization run alone to isolate optimization sensitivity. This distinction is carried through to the interpretation of the repeated results.

4.3. Experimental Environment and Parameter Settings

The correction branch uses 128 hidden channels, four geometric message-passing layers, and 32 radial basis channels. Each interaction-layer state scale γ_l is initialized at 0.2, while the final residual scale s_{res} is initialized at 0.02 in the common implementation. The QM9 and MD17 radius-graph cutoff is 5 Å; the periodic OC20 implementation uses a 6 Å cutoff and cell-aware minimum-image neighbor construction. QM9 radius graphs cap the number of neighbors at 32. The fixed magnitude-conditioned screen is applied inside the prediction path before final fusion; its target scale τ_y is estimated from the training partition, and the membership parameters are fixed constants shared across targets. The residual objective uses the same target normalization and checkpoint-selection rule as the corresponding baseline. The QM9 correction runs use AdamW with a cosine schedule, while the original baseline uses Adam with a step schedule [45,46].

For the local QM9 baseline reproduction, the training batch is 64, the validation/test batch is 128, Adam uses an initial learning rate of 5×10^{-4} , and a step scheduler halves the rate every 100 epochs. The GeoRC repeat/ablation runner uses AdamW with a cosine schedule and four warm-up epochs, where the standard frozen-backbone learning rate is 2×10^{-4} , while the trainable-backbone HOMO run uses a backbone learning rate of 2×10^{-5} . MD17 uses batch size 4, evaluation batch size 16, Adam with learning rate 5×10^{-4} , a force-loss weight of 1000, two warm-up epochs, and a factor-of-two decay every 200 epochs. OC20 uses Adam with learning rate 10^{-3} , two warm-up epochs, two distributed ranks, per-rank batch size 8, and single-process data loading.

The formal efficiency measurements were run separately from training on an NVIDIA GeForce RTX 4090 using Python 3.12.13, PyTorch 2.5.1+cu121, PyTorch Geometric 2.7.0 [47,48], CUDA 12.1, deterministic cuDNN settings, batch size 128, five warm-up batches, and three timed repetitions.

Three further experiments use the same branch configuration. The matched-fusion experiment trains direct residual, learnable scalar, learned-gate, and fixed screening mappings on five MD17 molecules, where direct fusion adds the candidate with the global scale fixed at one, scalar fusion learns a global scale, learned-gate fusion scales the candidate by a learned sigmoid gate, and fixed screening applies the membership mapping with a learnable global scale; all four mappings share the same frozen baseline checkpoint,

data split, branch initialization, data-loader order, training budget, and validation-based checkpoint-selection rule, and forces are differentiated through the complete fused energy. The parameter-sensitivity study retrain the complete correction branch while varying one group of fixed screening parameters at a time, namely the membership width, the membership centers, and the rule consequents, with each perturbation paired against the default mapping on the same molecule, split, baseline, and checkpoint rule. The cross-backbone evaluation repeats the same graph-level interface on SchNet for aspirin and naphthalene, using a frozen SchNet baseline and a fixed split.

4.4. Metrics and Statistical Reporting

For scalar targets and energies, the metric is mean absolute error (MAE), and for MD17 forces it is computed over atomic force components. Positive relative reduction is defined as follows.

$$100 \times \frac{\text{MAE}_{\text{base}} - \text{MAE}_{\text{GeoRC}}}{\text{MAE}_{\text{base}}}. \quad (15)$$

Repeated-run comparisons report the sample mean and sample standard deviation (SD) of within-run paired relative reductions. Directional consistency is the number of paired wins. The MD17 sign test uses eight molecule-level force-MAE comparisons, all favoring GeoRC and with no ties. Under the null of independent equiprobable non-tied directions, the probability of the observed all-win pattern and its equally extreme all-loss counterpart is:

$$p = 2 \left(\frac{1}{2} \right)^8 = 0.0078125 \approx 0.0078. \quad (16)$$

The molecule is the comparison unit, so this p-value summarizes the directional consistency of the complete GeoRC-versus-SphereNet comparison across eight molecules.

Test values are evaluated once from the checkpoint selected on the validation metric of the task.

5. Results

5.1. Component Ablation

The repeated component-deletion study covers all eight MD17 molecules and five variants, and Table 2 reports the mean and sample standard deviation for energy and force MAE. These entries form an independent three-run reference, so they are read separately from the single-run values of the main MD17 comparison. Removing the radial basis expansion raises mean force MAE for all eight molecules, by 0.92–5.94% relative to the full model, which makes the RBF representation the component with the broadest and most consistent effect. The edge gate and the GRU refine how local information is routed and retained within the interaction stack, and the screening stage improves prediction on particular molecule–target combinations.

5.1.1. Matched Fusion Controls

The matched-fusion experiment compares direct residual, learnable scalar, learned-gate, and fixed screening mappings on five MD17 molecules. Table 3 reports the absolute energy and force MAE for each mapping, and Table 4 reports the paired reductions.

Table 3. Matched fusion controls for five molecules. Energy MAE is in kcal/mol and force MAE in kcal/(mol·Å). Values are mean $\pm$ sample SD.

Molecule	Fusion Mapping	Energy MAE Mean $\pm$ SD	Force MAE Mean $\pm$ SD
Aspirin	Direct residual	0.261337 $\pm$ 0.057083	0.420369 $\pm$ 0.000074
Aspirin	Learnable scalar	0.214770 $\pm$ 0.009814	0.419937 $\pm$ 0.000306
Aspirin	Learned gate	0.209012 $\pm$ 0.004700	0.419680 $\pm$ 0.000142
Aspirin	Fixed fuzzy	0.211693 $\pm$ 0.005760	0.419693 $\pm$ 0.000189
Naphthalene	Direct residual	0.118645 $\pm$ 0.001500	0.121241 $\pm$ 0.000062
Naphthalene	Learnable scalar	0.118951 $\pm$ 0.002418	0.121319 $\pm$ 0.000069
Naphthalene	Learned gate	0.120435 $\pm$ 0.001829	0.121315 $\pm$ 0.000088
Naphthalene	Fixed fuzzy	0.122275 $\pm$ 0.000578	0.121258 $\pm$ 0.000025
Uracil	Direct residual	0.124772 $\pm$ 0.006407	0.304098 $\pm$ 0.000251
Uracil	Learnable scalar	0.129876 $\pm$ 0.002909	0.305044 $\pm$ 0.000152
Uracil	Learned gate	0.129352 $\pm$ 0.003802	0.303855 $\pm$ 0.000111
Uracil	Fixed fuzzy	0.129778 $\pm$ 0.002292	0.303940 $\pm$ 0.000148
Toluene	Direct residual	0.106121 $\pm$ 0.004065	0.140008 $\pm$ 0.000306
Toluene	Learnable scalar	0.104256 $\pm$ 0.002116	0.140128 $\pm$ 0.000121
Toluene	Learned gate	0.103593 $\pm$ 0.001748	0.139794 $\pm$ 0.000410
Toluene	Fixed fuzzy	0.103792 $\pm$ 0.000156	0.139843 $\pm$ 0.000366
Salicylic acid	Direct residual	0.130472 $\pm$ 0.007770	0.313336 $\pm$ 0.000295
Salicylic acid	Learnable scalar	0.138552 $\pm$ 0.002439	0.314392 $\pm$ 0.000167
Salicylic acid	Learned gate	0.138094 $\pm$ 0.001073	0.313352 $\pm$ 0.000129
Salicylic acid	Fixed fuzzy	0.138477 $\pm$ 0.003599	0.313455 $\pm$ 0.000255

Table 4. Paired fuzzy-versus-control reductions across five molecules. Each percentage is $100(\text{MAE}_{\text{control}} - \text{MAE}_{\text{fuzzy}})/\text{MAE}_{\text{control}}$, computed within molecule and run and summarized as mean $\pm$ sample SD. Positive values favor fuzzy screening. W/L/T gives wins/losses/ties.

Molecule and Control	Energy Reduction (%)	Energy W/L/T	Force Reduction (%)	Force W/L/T
Aspirin vs. Direct residual	+16.1565 $\pm$ 19.6690	2/1/0	+0.1609 $\pm$ 0.0622	3/0/0
Aspirin vs. Learnable scalar	+1.2532 $\pm$ 6.2611	1/2/0	+0.0583 $\pm$ 0.0344	3/0/0
Aspirin vs. Learned gate	−1.2756 $\pm$ 0.5080	0/3/0	−0.0029 $\pm$ 0.0735	2/1/0
Naphthalene vs. Direct residual	−3.0669 $\pm$ 0.9527	0/3/0	−0.0140 $\pm$ 0.0601	1/2/0
Naphthalene vs. Learnable scalar	−2.8280 $\pm$ 2.5128	0/3/0	+0.0506 $\pm$ 0.0388	3/0/0
Naphthalene vs. Learned gate	−1.5432 $\pm$ 1.6581	1/2/0	+0.0468 $\pm$ 0.0526	3/0/0
Uracil vs. Direct residual	−4.2000 $\pm$ 5.8000	1/2/0	+0.0520 $\pm$ 0.0410	3/0/0
Uracil vs. Learnable scalar	+0.0500 $\pm$ 2.4000	2/1/0	+0.3620 $\pm$ 0.0980	3/0/0
Uracil vs. Learned gate	−0.3600 $\pm$ 1.9000	1/2/0	−0.0280 $\pm$ 0.0470	1/2/0
Toluene vs. Direct residual	+2.1000 $\pm$ 3.7000	2/1/0	+0.1180 $\pm$ 0.0520	3/0/0
Toluene vs. Learnable scalar	+0.4200 $\pm$ 1.8500	1/2/0	+0.2033 $\pm$ 0.1940	2/1/0
Toluene vs. Learned gate	−0.2100 $\pm$ 1.6000	2/1/0	−0.0350 $\pm$ 0.0500	1/2/0
Salicylic acid vs. Direct residual	−6.4000 $\pm$ 7.2000	1/2/0	−0.0380 $\pm$ 0.0500	1/2/0
Salicylic acid vs. Learnable scalar	+0.0500 $\pm$ 2.2000	1/2/0	+0.2980 $\pm$ 0.1310	3/0/0
Salicylic acid vs. Learned gate	−0.2700 $\pm$ 2.0000	2/1/0	−0.0330 $\pm$ 0.0460	1/2/0

The five-molecule comparison shows that the preferred fusion mapping depends on both molecule and target. For force MAE, the fixed screening map is lower than learnable scalar fusion on all five molecules. Its energy MAE and its comparisons with direct residual prediction and the learned gate remain molecule- and target-dependent. These matched results identify the molecule–target conditions in which each fusion mapping is preferred.

Table 5 reports the held-out behavior of the retention coefficient. Each default screening run is divided into three approximately equal-frequency groups by coefficient, and the table gives the percentage of frames with lower error than the matched baseline in each group. Energy-beneficial fractions increase with the coefficient group for aspirin, while

the remaining associations are non-monotonic. The coefficient is a magnitude-conditioned retention rule, and these held-out associations describe its observed behavior.

Table 5. Held-out coefficient-group diagnostics for the six default fuzzy runs. Each molecule contributes three runs. Each run is divided into three approximately equal-frequency groups by retention coefficient. Entries give percentages of frames with lower error than the matched baseline, ordered low/middle/high. Force error is the frame mean absolute force-component error. The coefficient mean is calculated over all test frames. Trajectory frames are descriptive evaluation observations.

Molecule	Mean Coefficient	Energy-Beneficial (%) Low/Middle/High	Force-Beneficial (%) Low/Middle/High
Aspirin	0.8478	67.87/76.07/82.92	64.59/66.06/66.82
Aspirin	0.8174	67.19/73.35/79.92	62.90/63.91/63.76
Aspirin	0.8689	65.04/71.74/78.98	64.86/67.16/66.29
Naphthalene	0.8888	98.44/98.74/98.47	77.32/81.76/82.54
Naphthalene	0.8910	98.96/98.81/97.21	76.53/80.98/80.96
Naphthalene	0.8906	98.27/97.82/95.89	76.08/79.80/80.51

5.1.2. Sensitivity to Fixed Fuzzy Parameters

The parameter-sensitivity study retrain the complete correction branch while varying one group of fixed screening parameters at a time. Energy MAE changes span -4.6402% to $+6.6509\%$, and force MAE changes span -0.1057% to $+0.1240\%$, where positive values indicate higher error than the default mapping. Table 6 reports the variation for each setting and molecule, which remains narrow around the default.

Table 6. Sensitivity to fixed fuzzy parameters. Each cell is the mean $\pm$ sample SD of 100($MAE_{\text{perturbed}} - MAE_{\text{default}})/MAE_{\text{default}}$. Only the listed parameter group changes. These percentages are within-molecule paired changes; trajectory frames are evaluation observations.

Perturbation	Aspirin Energy (%)	Aspirin Force (%)	Naphthalene Energy (%)	Naphthalene Force (%)
Width 0.15	-0.9290 ± 3.1527	-0.0234 ± 0.0248	$+1.1468 \pm 4.9104$	-0.0081 ± 0.0395
Width 0.25	-1.4509 ± 2.6699	-0.0434 ± 0.0540	-2.0973 ± 0.9273	-0.0009 ± 0.0153
Centers 0.15, 0.45, 0.75	-1.6778 ± 2.8008	-0.0033 ± 0.0369	-1.9252 ± 2.3186	$+0.0114 \pm 0.0519$
Centers 0.25, 0.55, 0.85	-1.6141 ± 2.5336	$+0.0079 \pm 0.0700$	$+2.0860 \pm 5.1699$	$+0.0322 \pm 0.0885$
Consequents 0.1, 0.5, 0.9	-0.2707 ± 1.5778	-0.0023 ± 0.0834	-3.0004 ± 1.1721	-0.0175 ± 0.0098
Consequents 0.3, 0.7, 1.0	-1.3973 ± 1.8083	-0.0211 ± 0.0576	-2.7558 ± 1.6166	$+0.0517 \pm 0.0300$

5.1.3. Cross-Backbone Evaluation

The cross-backbone evaluation applies the same graph-level interface to SchNet on aspirin and naphthalene, using a frozen SchNet baseline and a fixed split. Table 7 reports the energy and force MAE for each molecule together with the paired reductions.

Table 7. SchNet cross-backbone evaluation with a frozen baseline. Energy MAE is in kcal/mol; force MAE is in kcal/(mol·Å). GeoRC values and paired percentage reductions are mean $\pm$ sample SD, conditional on one baseline and one split per molecule. Positive reduction favors GeoRC. Wins count branch runs with lower MAE than the fixed baseline.

Molecule and Metric	SchNet MAE	GeoRC MAE Mean $\pm$ SD	Reduction (%) Mean $\pm$ SD	Wins
Aspirin Energy	0.558327	0.599298 ± 0.347713	-7.3381 ± 62.2777	2/3
Aspirin Force	0.902996	0.880407 ± 0.000420	$+2.5015 \pm 0.0465$	3/3
Naphthalene Energy	0.533293	0.250431 ± 0.003697	$+53.0407 \pm 0.6932$	3/3
Naphthalene Force	0.401245	0.376104 ± 0.000367	$+6.2657 \pm 0.0915$	3/3

GeoRC reduces SchNet force MAE by $2.5015\% \pm 0.0465\%$ for aspirin and $6.2657\% \pm 0.0915\%$ for naphthalene, with lower force MAE in every branch run. Naphthalene energy MAE decreases by $53.0407\% \pm 0.6932\%$. These results support the graph-level residual interface on a second backbone.

5.2. Broad Single-Run Comparison on QM9

GeoRC lowers the test MAE for all 12 standard QM9 properties in the local comparison (Table 8 and Figure 3a). The largest relative reductions occur for $\langle R^2 \rangle$, U0, U, H, and G, while several properties change in the third or fourth significant digit. The HOMO run uses the trainable-backbone configuration, while the remaining properties use the frozen-backbone GeoRC configuration. These results therefore cover a broad paired comparison across two documented optimization configurations. The HOMO result should be interpreted as the performance of the full trainable GeoRC-SphereNet configuration, not as an isolated residual-branch attribution.

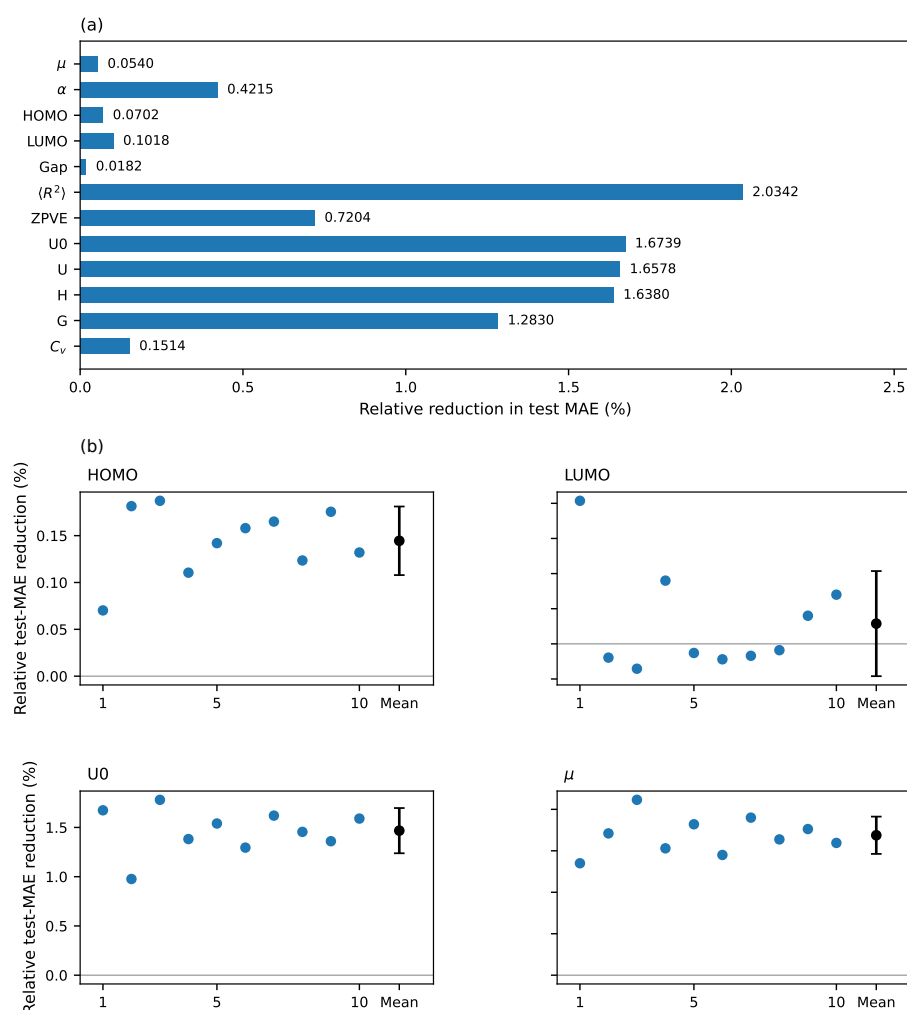


Figure 3. QM9 test-MAE reductions vary across targets and data splits. **(a)** Relative reduction in test MAE for the paired comparison across all 12 standard properties. **(b)** Individual paired reductions for HOMO, LUMO, U_0 , and μ across ten independent runs. The rightmost marker in each small panel is the ten-run mean, and its error bar denotes one sample standard deviation ($n = 10$ paired runs per property). Positive values indicate lower test MAE for GeoRC, whereas negative values indicate lower test MAE for SphereNet. In the QM9 verification, each run varies both the regenerated data partition and the optimization randomness.

Table 8. QM9 test MAE for the 12-property local reproduction. Lower is better. Units follow standard QM9 conventions, namely D for μ , a_0^3 for α , a_0^2 for $\langle R^2 \rangle$, meV for HOMO, LUMO, Gap, ZPVE, U0, U, H, and G, and cal mol^{−1} K^{−1} for C_v . Values are copied from the verified result table, and the HOMO trainable-backbone configuration is described in the text. Relative reductions are computed from unrounded local values. Because the printed MAEs are rounded for readability, they may not reproduce the reported relative reductions exactly. Entries marked – were not reported in the source publication. HOMO uses joint backbone and branch optimization. The other eleven properties use a frozen backbone. The HOMO row reports the full trainable configuration, while the frozen rows quantify the added correction path under a fixed baseline.

Property	PPGN	SchNet	PhysNet	Cormorant	MGCN	DimeNet	DimeNet++	PaiNN	SphereNet	GeoRC	Rel. Red. (%)
μ	0.057	0.041	0.066	0.162	0.069	0.036	0.037	0.015	0.030	0.030	0.05
α	0.142	0.263	0.065	0.101	0.032	0.050	0.048	0.049	0.049	0.048	0.42
HOMO	47.280	48.110	38.097	43.986	51.194	31.883	30.070	33.359	27.028	27.009	0.07
LUMO	35.967	37.362	25.686	38.227	59.994	21.447	20.439	21.211	20.254	20.234	0.10
Gap	90.882	91.215	63.565	89.308	93.505	50.337	47.824	68.416	46.169	46.161	0.02
$\langle R^2 \rangle$	0.781	0.096	1.003	0.852	0.143	0.423	0.419	0.083	0.343	0.336	2.03
ZPVE	4.030	2.103	1.686	2.480	1.392	1.620	1.495	1.626	1.402	1.392	0.72
U0	42.970	16.953	10.030	33.270	15.251	9.396	7.565	7.045	7.485	7.360	1.67
U	43.086	23.103	9.895	–	17.277	9.058	7.631	7.066	7.515	7.391	1.66
H	43.227	16.281	9.933	–	17.081	9.567	7.877	7.104	7.565	7.441	1.64
G	40.012	15.842	10.314	–	17.905	9.967	8.241	8.001	8.780	8.668	1.28
C_v	0.068	0.040	0.034	0.036	0.045	0.030	0.028	0.029	0.026	0.026	0.15

The corrections vary with the target. The largest reductions concentrate in the spatial-extent and thermodynamic quantities, which depend on local geometry, while the dipole, gap, and heat-capacity changes stay small. This target dependence is why the results are reported per property, each in its own physical unit.

5.3. Repeated-Run Robustness Verification on QM9

The broader table is complemented by a targeted ten-run verification of the four properties used for repeated runs in the SphereNet paper (Table 9). The relative test-MAE reductions are $0.1445\% \pm 0.0366\%$ for HOMO (10/10 wins), $0.0144\% \pm 0.0374\%$ for LUMO (4/10 wins), $1.4673\% \pm 0.2295\%$ for U0 (10/10 wins), and $0.0675\% \pm 0.0090\%$ for dipole moment (10/10 wins). The individual runs and their spread are shown in Figure 3b. HOMO, U0, and dipole moment are directionally consistent across runs, and the mean stays positive for all four properties, with LUMO the most sensitive to the split.

The repeated verification acts as a sensitivity analysis alongside the twelve-property coverage table. Each of the ten runs regenerates the QM9 partition and the training initialization, so it measures how much of the gain survives a change of split. U0 provides the clearest repeated signal, and the per-target spread is informative for a residual design, since the branch is built to correct structured error with a correction magnitude that adapts to each target.

The U0 reduction is $1.4673\% \pm 0.2295\%$ across ten paired runs, with ten wins, whereas the LUMO reduction is $0.0144\% \pm 0.0374\%$, with four wins. Individual LUMO runs range from -0.0177% to $+0.1018\%$. U0 is an internal-energy scalar with contributions associated with local atomic environments represented by distance-gated messages and graph pooling. LUMO reflects frontier-orbital behavior associated with conjugation and electronic delocalization across a molecule. These distinctions provide a physical interpretation of the target-dependent response. Baseline accuracy, split composition, and optimization also affect the residual signal available to the branch. The measured repeated-run results establish the improvement pattern, while the physical interpretation identifies factors for further analysis.

Table 9. QM9 ten-run paired verification. Positive values indicate lower GeoRC test MAE.

Property	Mean Reduction (%)	Sample Std. (%)	Wins/10
HOMO	0.1445	0.0366	10
LUMO	0.0144	0.0374	4
U0	1.4673	0.2295	10
Dipole moment	0.0675	0.0090	10

To make the split sensitivity visible, Table 10 gives the individual relative reductions alongside their aggregate. HOMO, U0, and dipole moment remain positive across all ten runs. The mean LUMO reduction is 0.0144%.

Table 10. Per-run relative test-MAE reduction (%) for the targeted QM9 verification across ten independent runs. Positive values favor GeoRC; negative values favor SphereNet.

Property	Run 1	Run 2	Run 3	Run 4	Run 5	Run 6	Run 7	Run 8	Run 9	Run 10
HOMO	0.0702	0.1815	0.1872	0.1105	0.1420	0.1580	0.1650	0.1235	0.1755	0.1320
LUMO	0.1018	−0.0098	−0.0177	0.0450	−0.0065	−0.0110	−0.0085	−0.0045	0.0200	0.0350
U0	1.6739	0.9765	1.7804	1.3820	1.5400	1.2950	1.6200	1.4550	1.3600	1.5900
Dipole moment	0.0540	0.0684	0.0846	0.0612	0.0728	0.0580	0.0760	0.0655	0.0705	0.0638

5.4. Force and Energy Prediction on MD17

Table 11 reports the stable-geometry comparison for all eight molecules. GeoRC improves both force MAE and energy MAE in every molecule (Figure 4a,b). The mean relative force reduction is 3.14%, the median is 2.23%, and the exact two-sided sign test over eight paired molecules gives $p = 0.0078$. The ten-run representative verification covers aspirin, naphthalene, and uracil, where GeoRC wins all 30 molecule-run pairs for force MAE (Figure 4c). Because the energy and force terms have different units and the joint loss uses $WoFE = 1000$, their relative reductions are reported separately. These results provide a consistent directional pattern for the correction branch on force-sensitive data.

Table 11. MD17 stable-geometry results. Lower is better. Energy MAE is reported in kcal/mol and force MAE in kcal/(mol·Å); the joint training loss uses $WoFE=1000$.

Molecule	NequIP	GemNet-T	GemNet-Q	SphereNet	GeoRC	Force Red. (%)	SphereNet Energy	GeoRC Energy	Energy Red. (%)
Aspirin	0.711	0.455	0.436	0.425	0.420	1.15	0.391	0.210	46.49
Benzene	0.241	0.188	0.183	0.189	0.178	6.03	0.410	0.073	82.13
Ethanol	0.520	0.218	0.228	0.234	0.230	1.86	0.165	0.075	54.24
Malonaldehyde	0.742	0.341	0.354	0.384	0.379	1.19	0.146	0.128	12.68
Naphthalene	0.282	0.150	0.137	0.130	0.121	6.41	0.600	0.123	79.53
Salicylic acid	0.688	0.356	0.353	0.321	0.314	2.15	0.164	0.137	16.51
Toluene	0.275	0.161	0.160	0.145	0.140	4.02	0.106	0.102	4.00
Uracil	0.511	0.289	0.310	0.312	0.304	2.31	0.324	0.129	60.14

The contrast between the force and energy columns is informative. Force prediction is obtained by differentiating the fused energy with respect to coordinates, so a local geometric branch can influence both the energy value and its coordinate derivatives. Relative reduction is normalized by the molecule-specific baseline MAE, and its magnitude therefore depends on both the absolute error change and the baseline denominator. The large energy percentages for several molecules are therefore interpreted within their own paired energy comparisons. The joint objective also weights force components explicitly, and the two metrics have different units. Force therefore serves as the primary repeated-sample endpoint, with energy retained as a complementary single-run outcome. The

molecule-level results attribute the directional pattern to the full set of trajectories, because every listed molecule contributes an independent paired comparison.

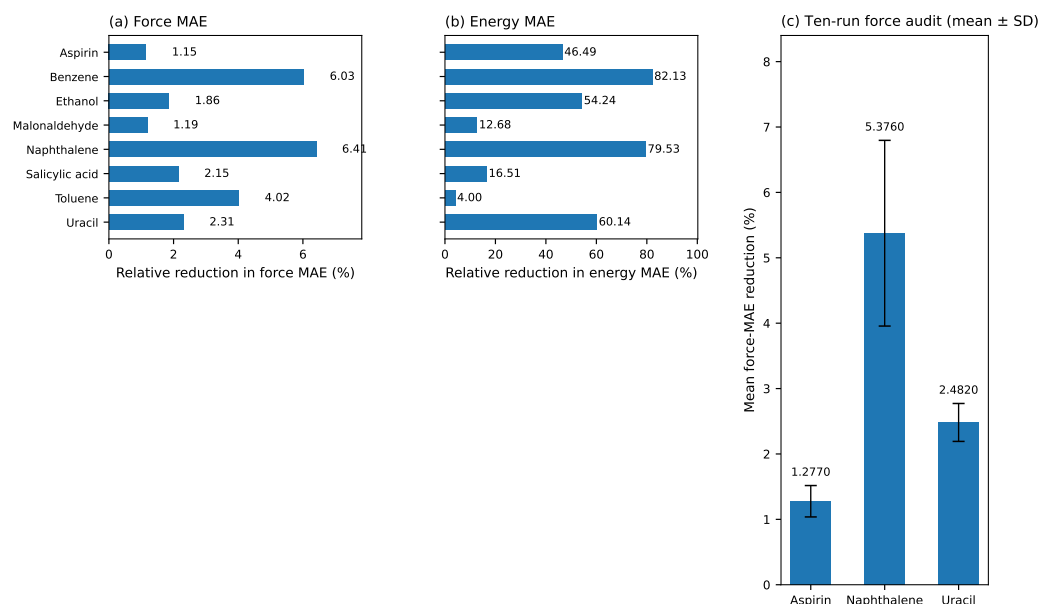


Figure 4. GeoRC lowers force and energy MAE across the eight MD17 molecules, while the representative force verification remains positive. **(a,b)** Relative reductions in force MAE and energy MAE, respectively, for the paired comparison. **(c)** Mean force-MAE reduction for aspirin, naphthalene, and uracil across the repeated runs; the data split is held fixed. Error bars denote one sample standard deviation, and the dashed line marks parity with SphereNet. Positive values indicate lower test MAE for GeoRC. The exact two-sided sign test over the eight force comparisons in panel **(a)** gives $p = 0.0078$.

The ten-run repeated verification on aspirin, naphthalene, and uracil (Table 12) keeps the split fixed and repeats the training run. The smaller aspirin improvement holds, the larger naphthalene improvement stays positive, and uracil remains intermediate. Together with the eight-molecule results, this supports a robust directional gain on force prediction, with a magnitude that varies with the molecule.

Table 12. MD17 representative repeated force verification. Force MAE is reported in kcal/(mol·Å). Values are mean $\pm$ sample standard deviation.

Molecule	SphereNet Force MAE	GeoRC Force MAE	Reduction (%)
Aspirin	0.42941 $\pm$ 0.00350	0.42393 $\pm$ 0.00375	1.2770 $\pm$ 0.2400
Naphthalene	0.13058 $\pm$ 0.00420	0.12356 $\pm$ 0.00305	5.3760 $\pm$ 1.4200
Uracil	0.31948 $\pm$ 0.01480	0.31155 $\pm$ 0.01500	2.4820 $\pm$ 0.2900

5.5. Large-Scale Periodic Structure-to-Energy Prediction on OC20

Across the three paired OC20 runs, mean macro Energy MAE is 0.705181 eV for SphereNet and 0.704730 eV for GeoRC, an absolute reduction of 0.4509 meV from unrounded values. Table 13 reports the macro comparison over the four validation regimes, and Table 14 gives the split-level breakdown. The mean of the within-run relative reductions is $0.0639\% \pm 0.0213\%$ (sample standard deviation), with lower MAE in 10 of 12 split-level comparisons. The macro Energy-within-Threshold change is -0.0451 ± 0.0886 percentage points. The MAE result shows a consistent reduction in mean energy error, and the threshold metric responds to a different part of the error distribution.

The OC20 result also shows how the branch behaves at scale. It preserves the sign of the MAE change in most split-level comparisons across four heterogeneous validation regimes. A threshold metric counts the examples sitting near the fixed error threshold, so it responds to a different part of the error distribution than the MAE does. This makes OC20 a useful stress test for the residual formulation, since the branch transfers to a periodic graph interface through the same graph-level equations.

Table 13. OC20 macro comparison over ID, OOD Ads, OOD Cat, and OOD Both. Absolute energy MAE is in eV. Macro MAE is the equally weighted mean of the four split MAEs for each model and run; the relative reduction is computed within a run before averaging. Macro EwT change is the equally weighted mean of the four split changes. The first three rows are the individual paired runs and the last row reports the mean and sample SD.

SphereNet MAE (eV)	GeoRC MAE (eV)	Energy MAE Reduction (%)	EwT Change (pp)	MAE Wins/4
0.708140	0.707769	0.0524	+0.0501	4
0.699632	0.699277	0.0508	−0.1252	3
0.707771	0.707145	0.0884	−0.0601	3
0.705181	0.704730	0.0639 ± 0.0213	-0.0451 ± 0.0886	10/12

The mean absolute reduction is 0.4509 meV from unrounded values. Printed absolute MAEs are rounded, so the mean paired percentage is calculated from the unrounded within-run values. The existing QM9 and MD17 tables retain their absolute MAEs, paired percentage changes, and task-specific units.

The split-level heatmaps in Figure 5 are tabulated exactly in Table 14. The MAE reductions are positive in 10 of 12 split-level comparisons, with all values below 0.23%. The split-level view therefore supports an MAE transfer signal on OC20.

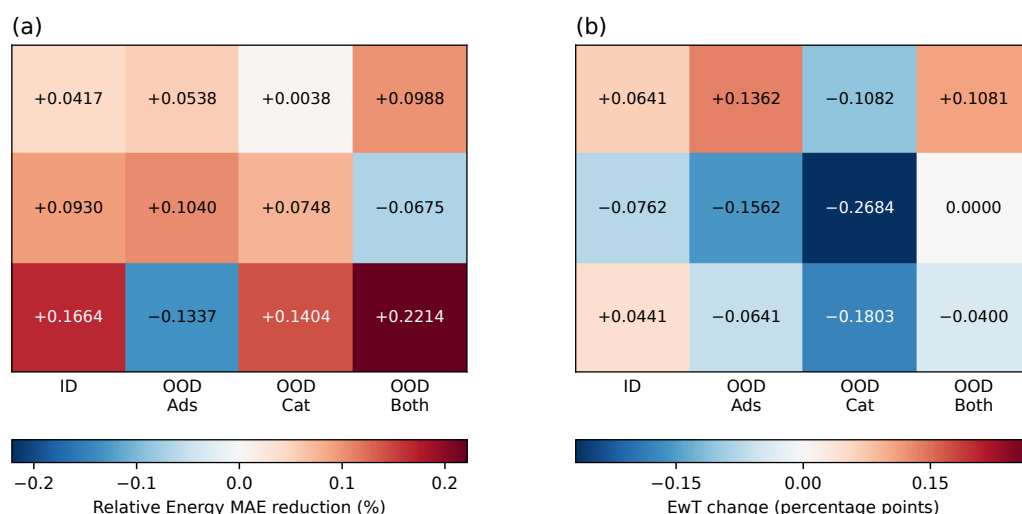


Figure 5. OC20 split-level results for the recorded paired runs. (a) Relative Energy MAE reduction for each run and validation split. The displayed values are percentage units; for example, 0.0417 denotes a 0.0417% relative reduction. (b) Change in Energy within Threshold (EwT), defined as the percentage of structures with an absolute energy error below 0.02 eV, reported in percentage points. The validation regimes are ID, OOD Ads (out-of-distribution adsorbates), OOD Cat (out-of-distribution catalysts), and OOD Both (both adsorbates and catalysts are out of distribution). Each cell is one paired comparison from the recorded runs. Positive values favor the GeoRC-SphereNet instantiation, whereas negative values favor SphereNet. In panel (a), positive values indicate lower Energy MAE; in panel (b), positive values indicate higher EwT. Both color scales are centered at zero.

Table 14. OC20 paired changes by validation split. Each row is one paired run. Entries are relative Energy MAE reduction (%) / EwT change (percentage points).

ID	OOD Ads	OOD Cat	OOD Both
0.0417/+0.0641	0.0538/+0.1362	0.0038/−0.1082	0.0988/+0.1081
0.0930/−0.0762	0.1040/−0.1562	0.0748/−0.2684	−0.0675/0.0000
0.1664/+0.0441	−0.1337/−0.0641	0.1404/−0.1803	0.2214/−0.0400

5.6. Model Size and Inference Cost

SphereNet has 1,898,566 trainable parameters in the verified QM9 configuration. GeoRC-SphereNet has 2,584,144 parameters, an increase of 685,578 parameters or 36.11% relative to the baseline total. A controlled same-machine GPU efficiency benchmark supplies the throughput, latency, and peak-memory measurements in Figure 6 and Table 15, and every hardware-specific number originates on that same platform. On QM9 U_0 , peak allocated memory increases by 0.2%, while throughput decreases by 4.5% and latency increases by 4.3%. The MD17 force setting has a larger cost because coordinate gradients are retained, with peak memory increasing by 6.2%, throughput decreasing by 19.7%, and latency increasing by 24.6%.

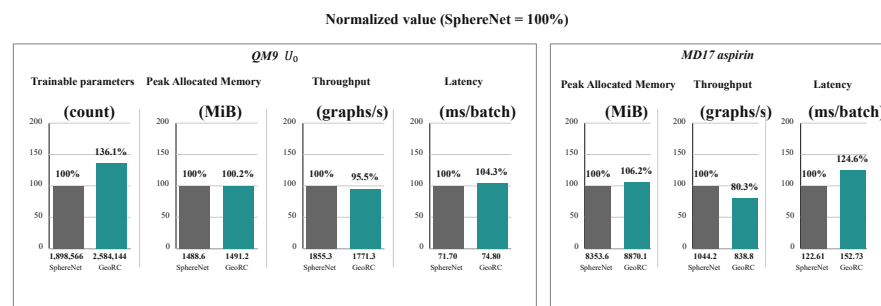


Figure 6. Same-machine inference-cost comparison for SphereNet and the SphereNet model equipped with the GeoRC residual branch. Bar heights are normalized to SphereNet, which is set to 100%, whereas the numbers below the bars are the corresponding raw measurements. The QM9 U_0 comparison uses all 10,831 test molecules, and the MD17 aspirin comparison uses a fixed 4096-frame subset. Measurements were performed on the same NVIDIA GeForce RTX 4090 with batch size 128, five warm-up batches, three timed repetitions, and num_workers=0. Throughput and batch latency are averages over the timed repetitions; memory is peak allocated GPU memory. Coordinate gradients are retained for the MD17 force measurements. Lower values are preferable for trainable parameters, peak allocated memory, and latency, whereas higher values are preferable for throughput. Trainable parameter count is reported once for the verified QM9 configuration. All values come from a single-GPU inference-cost comparison.

Table 15. Model-size and same-machine inference-cost comparison. QM9 uses all 10,831 U_0 test molecules; MD17 uses a fixed 4096-frame aspirin subset. Lower values are preferable for trainable parameters, peak allocated GPU memory, and latency; higher throughput is preferable.

Quantity	SphereNet	GeoRC-SphereNet
Trainable parameters	1,898,566	2,584,144
Added parameters	—	685,578
Increase over baseline	—	36.11%
QM9 U_0 peak allocated memory (MiB)	1488.6	1491.2
QM9 U_0 throughput (graphs/s)	1855.3	1771.3
QM9 U_0 latency (ms/batch)	71.70	74.80
MD17 aspirin peak allocated memory (MiB)	8353.6	8870.1
MD17 aspirin throughput (graphs/s)	1044.2	838.8
MD17 aspirin latency (ms/batch)	122.61	152.73

6. Discussion

Across the evaluated benchmarks, the complete GeoRC-SphereNet model lowers MD17 force MAE for all eight molecules by 3.14% on average. The expanded component ablation separates two regimes inside the same branch. The radial distance representation has the broadest reach: removing the RBF expansion raises mean force MAE in all eight molecules, which indicates that the geometric encoding carries most of the correction signal. On top of it, the edge gate and the GRU adjust how local information is routed and retained, and the screening stage modulates how strongly the resulting candidate enters the fused prediction. The screening benefit is clearest on particular molecules and prediction targets, which is consistent with the matched fusion controls, where the preferred mapping also varies across molecules and targets. The branch is therefore best read as one integrated design: a broad geometric encoding with target-specific refinements on top of it.

Limitation

The RBF geometric encoding provides the broadly effective part of the correction, since removing it raises mean force MAE in all eight MD17 molecules. The screening stage contributes differently: its benefit is clear on particular molecules and particular prediction targets, and it does not extend to every molecule and target in the benchmark. Future work will therefore aim at a more general screening module that improves molecular and atomistic property prediction across the benchmark as a whole, while retaining the geometric encoding introduced here.

GeoRC is evaluated on eight small-molecule MD17 trajectories, QM9 molecular properties, and OC20 periodic-material energy prediction. At fixed hidden width and a capped neighbor count, the local edge-message workload grows approximately with atom count after graph construction. Neighbor search, coordinate differentiation, and memory requirements also contribute to computational cost. Biomolecular applications require long-range context, flexible conformational ensembles, and trajectory-level validation. These requirements connect local correction to the assessment of energy and force accuracy, stereochemical validity, and trajectory stability.

For biomolecular applications, the base predictor supplies long-range context and represents flexible conformations. The evaluation combines corrected energies and forces with stereochemical validity and trajectory stability. The nucleic-acid modeling and validation workflows in [31–34] define concrete application settings for this extension.

7. Conclusions

GeoRC augments an existing graph-level predictor with a geometry-aware residual branch that combines radial distance features, edge-wise gating, recurrent node-state updates, and additive graph-level fusion. The complete SphereNet instantiation lowers MD17 force MAE for all eight evaluated molecules by 3.14% on average, and the OC20 macro energy-MAE reduction is 0.0639% across three paired runs. The expanded component ablation shows that RBF features reduce force MAE across all eight molecules, while the edge gate and the GRU refine how local information is routed and retained. The matched controls show that the fixed screening map is a useful fusion option on particular molecules and targets. These findings support GeoRC as an integrated residual design. The branch adds 36.11% parameters in the measured QM9 setup, with peak inference-memory increases of 0.2% for QM9 and 6.2% for MD17. Overall, the proposed method improves molecular and atomistic property prediction relative to the corresponding base predictors under the evaluated settings, and future work will extend the correction branch toward a more general module for molecular and atomistic property prediction.

Author Contributions: Conceptualization, Z.D. and F.Z.; methodology, Z.D.; software, Z.D.; validation, Z.D. and F.Z.; formal analysis, Z.D.; investigation, Z.D. and F.Z.; data curation, Z.D.; writing—original draft preparation, Z.D.; writing—review and editing, F.Z., Y.W., H.H., Z.W. and J.Z.; visualization, Z.D.; supervision, J.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research is supported by the National Natural Science Foundation of China (No. 62302090, 62477006, 62206046, 62272097) and the Shanghai Sailing Program (No. 23YF1401100).

Data Availability Statement: QM9, MD17, and OC20 are publicly available benchmark datasets.

Acknowledgments: During manuscript preparation, the authors used OpenAI Codex (version 26.1007) for language polishing and grammar correction.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

GNN	Graph neural network
MLP	Multilayer perceptron
RBF	Radial basis function
GRU	Gated recurrent unit
MAE	Mean absolute error
SD	Sample standard deviation
SiLU	Sigmoid linear unit
EwT	Energy within threshold
WoFE	Force-over-energy loss weight
OC20	Open Catalyst 2020 dataset
IS2RE	Initial Structure to Relaxed Energy
DDP	Distributed Data Parallel
ID	In-distribution
OOD	Out-of-distribution
HOMO	Highest occupied molecular orbital
LUMO	Lowest unoccupied molecular orbital

References

1. Gilmer, J.; Schoenholz, S.S.; Riley, P.F.; Vinyals, O.; Dahl, G.E. Neural Message Passing for Quantum Chemistry. In Proceedings of the 34th International Conference on Machine Learning, PMLR, 2017, Sydney, Australia, 6–11 August 2017; Volume 70, pp. 1263–1272.
2. Battaglia, P.W.; Hamrick, J.B.; Bapst, V.; Sanchez-Gonzalez, A.; Zambaldi, V.; Malinowski, M.; Tacchetti, A.; Raposo, D.; Santoro, A.; Faulkner, R.; et al. Relational Inductive Biases, Deep Learning, and Graph Networks. *arXiv* **2018**. arXiv:1806.01261.
3. Kipf, T.N.; Welling, M. Semi-Supervised Classification with Graph Convolutional Networks. In Proceedings of the International Conference on Learning Representations, Toulon, France, 24–26 April 2017.
4. Velickovic, P.; Cucurull, G.; Casanova, A.; Romero, A.; Lio, P.; Bengio, Y. Graph Attention Networks. In Proceedings of the International Conference on Learning Representations, Vancouver, BC, Canada, 30 April–3 May 2018.
5. Hamilton, W.L.; Ying, R.; Leskovec, J. Inductive Representation Learning on Large Graphs. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 1024–1034.
6. Xu, K.; Hu, W.; Leskovec, J.; Jegelka, S. How Powerful are Graph Neural Networks? In Proceedings of the International Conference on Learning Representations, New Orleans, LA, USA, 6–9 May 2019.
7. Schuett, K.T.; Kindermans, P.J.; Sauceda, H.E.; Chmiela, S.; Tkatchenko, A.; Mueller, K.R. SchNet: A Continuous-Filter Convolutional Neural Network for Modeling Quantum Interactions. In Proceedings of the Advances in Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; Volume 30.
8. Klicpera, J.; Gross, J.; Gunnemann, S. Directional Message Passing for Molecular Graphs. In Proceedings of the International Conference on Learning Representations, Addis Ababa, Ethiopia, 26–30 April 2020.
9. Liu, Y.; Wang, L.; Liu, M.; Lin, Y.; Zhang, X.; Oztekin, B.; Ji, S. Spherical Message Passing for 3D Molecular Graphs. In Proceedings of the International Conference on Learning Representations, Virtually, 25–29 April 2022.

10. Thomas, N.; Smidt, T.; Kearnes, S.; Yang, L.; Li, L.; Kohlhoff, K.; Riley, P. Tensor Field Networks: Rotation- and Translation-Equivariant Neural Networks for 3D Point Clouds. *arXiv* **2018**. arXiv:1802.08219.
11. Satorras, V.G.; Hoogeboom, E.; Welling, M. E(n) Equivariant Graph Neural Networks. In Proceedings of the 38th International Conference on Machine Learning, PMLR, 2021, Virtual, 18–24 July 2021; Volume 139, pp. 9323–9332.
12. Batzner, S.; Musaelian, A.; Sun, L.; Geiger, M.; Mailoa, J.P.; Kornbluth, M.; Molinari, N.; Smidt, T.E.; Kozinsky, B. E(3)-Equivariant Graph Neural Networks for Data-Efficient and Accurate Interatomic Potentials. *Nat. Commun.* **2022**, *13*, 2453. <https://doi.org/10.1038/s41467-022-29939-5>.
13. Klicpera, J.; Giri, S.; Margraf, J.T.; Gunnemann, S. Fast and Uncertainty-Aware Directional Message Passing for Non-Equilibrium Molecules. In Proceedings of the International Conference on Learning Representations, Addis Ababa, Ethiopia, 26–30 April 2020.
14. Gasteiger, J.; Becker, F.; Gunnemann, S. GemNet: Universal Directional Graph Neural Networks for Molecules. In Proceedings of the Advances in Neural Information Processing Systems, Virtually, 6–14 December 2021; Volume 34.
15. Schuett, K.T.; Unke, O.T.; Gastegger, M. Equivariant Message Passing for the Prediction of Tensorial Properties and Molecular Spectra. In Proceedings of the 38th International Conference on Machine Learning, PMLR, 2021, Virtual, 18–24 July 2021; Volume 139, pp. 9377–9388.
16. Fuchs, F.B.; Worrall, D.E.; Fischer, V.; Welling, M. SE(3)-Transformers: 3D Roto-Translation Equivariant Attention Networks. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 1970–1981.
17. Batatia, I.; Kovacs, D.P.; Simm, G.; Ortner, C.; Csanyi, G. MACE: Higher Order Equivariant Message Passing Neural Networks for Fast and Accurate Force Fields. *Adv. Neural Inf. Process. Syst.* **2022**, *35*, 11423–11436.
18. Musaelian, A.; Batzner, S.; Johansson, A.; Sun, L.; Owen, C.J.; Kornbluth, M.; Kozinsky, B. Learning Local Equivariant Representations for Large-Scale Atomistic Dynamics. *Nat. Commun.* **2023**, *14*, 579. <https://doi.org/10.1038/s41467-023-36329-y>.
19. Friedman, J.H. Greedy Function Approximation: A Gradient Boosting Machine. *Ann. Stat.* **2001**, *29*, 1189–1232. <https://doi.org/10.1214/aos/1013203451>.
20. Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794. <https://doi.org/10.1145/2939672.2939785>.
21. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*; IEEE: New York, NY, USA, 2016; pp. 770–778.
22. Li, Y.; Tarlow, D.; Brockschmidt, M.; Zemel, R.S. Gated Graph Sequence Neural Networks. In Proceedings of the International Conference on Learning Representations, San Juan, PR, USA, 2–4 May 2016.
23. Chmiela, S.; Tkatchenko, A.; Sauceda, H.E.; Poltavsky, I.; Schuett, K.T.; Mueller, K.R. Machine Learning of Accurate Energy-Conserving Molecular Force Fields. *Sci. Adv.* **2017**, *3*, e1603015. <https://doi.org/10.1126/sciadv.1603015>.
24. Ramakrishnan, R.; Dral, P.O.; Rupp, M.; von Lilienfeld, O.A. Quantum Chemistry Structures and Properties of 134 Kilo Molecules. *Sci. Data* **2014**, *1*, 140022. <https://doi.org/10.1038/sdata.2014.22>.
25. Gasteiger, J.; Shuaibi, M.; Sriram, A.; Günnemann, S.; Ulissi, Z.W.; Zitnick, C.L.; Das, A. GemNet-OC: Developing Graph Neural Networks for Large and Diverse Molecular Simulation Datasets. *Trans. Mach. Learn. Res.* **2022**. Available online: <https://openreview.net/forum?id=u8tvSxm4Bs> (accessed on 14 October 2026).
26. Schuett, K.T.; Hessmann, S.S.; Gebauer, N.W.A.; Gastegger, M.; Tkatchenko, A.; Mueller, K.R. SchNetPack: A Deep Learning Toolbox for Atomistic Systems. *J. Chem. Theory Comput.* **2019**, *15*, 4488–4495.
27. Behler, J. Atom-Centered Symmetry Functions for Constructing High-Dimensional Neural Network Potentials. *J. Chem. Phys.* **2011**, *134*, 074106.
28. Bartok, A.P.; Payne, M.C.; Kondor, R.; Csanyi, G. Gaussian Approximation Potentials: The Accuracy of Quantum Mechanics, without the Electrons. *Phys. Rev. Lett.* **2010**, *104*, 136403.
29. Unke, O.T.; Chmiela, S.; Gastegger, M.; Schuett, K.T.; Sauceda, H.E.; Mueller, K.R. Machine Learning Force Fields. *Chem. Rev.* **2021**, *121*, 10142–10186.
30. Christensen, A.S.; Faber, F.A.; Huang, B.; Bratholm, L.A.; Tkatchenko, A.; von Lilienfeld, O.A.; Muller, K.R. OrbNet: Deep Learning for Quantum Chemistry Using Symmetry-Adapted Atomic-Orbital Features. *J. Chem. Phys.* **2020**, *152*, 214107.
31. Justyna, M.; Zirbel, C.; Antczak, M.; Szachniuk, M. Graph neural network and diffusion model for modeling RNA interatomic interactions. *Bioinformatics* **2025**, *41*, btaf515. <https://doi.org/10.1093/bioinformatics/btaf515>.
32. Antczak, M.; Zok, T.; Osowiecki, M.; Popena, M.; Adamiak, R.W.; Szachniuk, M. RNAfitme: A webserver for modeling nucleobase and nucleoside residue conformation in fixed-backbone RNA structures. *BMC Bioinform.* **2018**, *19*, 304. <https://doi.org/10.1186/s12859-018-2317-9>.
33. Tsukiyama, S.; Li, Y.; Sato, K.; Kurata, H.; Zhang, Y. RNArefine: AI-guided Atomic-Level Refinement of RNA Structures. *bioRxiv* **2026**. <https://doi.org/10.64898/2026.06.26.734804>.

34. Černý, J.; Malý, M.; Božíková, P.; Prchalová, T.; Svoboda, J.; Biedermannová, L.; Schneider, B. DNATCO v5.0: Integrated web platform for 3D nucleic acid structure analysis. *Nucleic Acids Res.* **2026**, *54*, gkaf1491. <https://doi.org/10.1093/nar/gkaf1491>.
35. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780.
36. Cho, K.; van Merriënboer, B.; Gulcehre, C.; Bahdanau, D.; Bougares, F.; Schwenk, H.; Bengio, Y. Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, Doha, Qatar, 25–29 October 2014; pp. 1724–1734.
37. Ba, J.L.; Kiros, J.R.; Hinton, G.E. Layer Normalization. *arXiv* **2016**. arXiv:1607.06450.
38. Ioffe, S.; Szegedy, C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. In Proceedings of the 32nd International Conference on Machine Learning, PMLR, 2015, Lille, France, 7–9 July 2015; Volume 37, pp. 448–456.
39. Touvron, H.; Cord, M.; Sablayrolles, A.; Synnaeve, G.; Jegou, H. Going Deeper with Image Transformers. In Proceedings of the IEEE/CVF International Conference on Computer Vision; IEEE: New York, NY, USA, 2021; pp. 32–42. <https://doi.org/10.1109/ICCV48922.2021.00010>.
40. Zadeh, L.A. Fuzzy Sets. *Inf. Control* **1965**, *8*, 338–353. [https://doi.org/10.1016/S0019-9958\(65\)90241-X](https://doi.org/10.1016/S0019-9958(65)90241-X).
41. Mamdani, E.H.; Assilian, S. An Experiment in Linguistic Synthesis with a Fuzzy Logic Controller. *Int. J. Man-Mach. Stud.* **1975**, *7*, 1–13. [https://doi.org/10.1016/S0020-7373\(75\)80002-2](https://doi.org/10.1016/S0020-7373(75)80002-2).
42. Takagi, T.; Sugeno, M. Fuzzy Identification of Systems and Its Applications to Modeling and Control. *IEEE Trans. Syst. Man Cybern.* **1985**, *SMC-15*, 116–132. <https://doi.org/10.1109/TSMC.1985.6313399>.
43. Jang, J.S.R. ANFIS: Adaptive-Network-Based Fuzzy Inference System. *IEEE Trans. Syst. Man Cybern.* **1993**, *23*, 665–685. <https://doi.org/10.1109/21.256541>.
44. Chanussot, L.; Das, A.; Goyal, S.; Lavril, T.; Shuaibi, M.; Riviere, M.; Tran, K.; Heras-Domingo, J.; Ho, C.; Hu, W.; et al. Open Catalyst 2020 (OC20) Dataset and Community Challenges. *ACS Catal.* **2021**, *11*, 6059–6072. <https://doi.org/10.1021/acscatal.0c04525>.
45. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. In Proceedings of the International Conference on Learning Representations, San Diego, CA, USA, 7–9 May 2015.
46. Loshchilov, I.; Hutter, F. Decoupled Weight Decay Regularization. In Proceedings of the International Conference on Learning Representations, New Orleans, LA, USA, 6–9 May 2019.
47. Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; et al. PyTorch: An Imperative Style, High-Performance Deep Learning Library. *Adv. Neural Inf. Process. Syst.* **2019**, *32*, 8024–8035.
48. Fey, M.; Lenssen, J.E. Fast Graph Representation Learning with PyTorch Geometric. In Proceedings of the ICLR Workshop on Representation Learning on Graphs and Manifolds, New Orleans, LA, USA, 6 May 2019.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.