Extended Abstract Track

Symmetry-Aware Generative Modeling through Learned Canonicalization

Editors: List of editors' names

Abstract

Generative modeling of symmetric densities has a range of applications in AI for science, from drug discovery to physics simulations. The existing generative modeling paradigm for invariant densities combines an invariant prior with an equivariant generative process. However, we observe that this technique is not necessary and has several drawbacks resulting from the limitations of equivariant networks. Instead, we propose to model a learned slice of the density so that only one representative element per orbit is learned. To accomplish this, we learn a group-equivariant canonicalization network that maps training samples to a canonical pose and train a non-equivariant generative model over these canonicalized samples. We implement this idea in the context of diffusion models. Our preliminary experimental results on molecular modeling are promising; demonstrating improved sample quality and faster inference time.

Keywords: Generative modeling, equivariance, symmetry, deep learning.

1. Introduction

Equivariant models have emerged as a popular family of models for generative modeling tasks with symmetry, see e.g. (Köhler et al., 2020; Shi et al., 2021; Hoogeboom et al., 2022a; Vignac et al., 2023; Bose et al., 2024). These models are especially prevalent in physical domains, which are subject to Euclidean symmetries (rotations and translations) and permutation symmetries.

However, equivariant models have some limitations. One issue is computational complexity, as specialized operations like group convolutions or spherical harmonics are expensive (Kondor et al., 2018; Cohen et al., 2018; Weiler and Cesa, 2019b). Additionally, there are expressivity constraints, since limiting models to equivariant functions may restrict their ability to capture complex patterns not perfectly aligned with the symmetry group (Maron et al., 2019a; Ravanbakhsh et al., 2017; Zhou and Feng, 2020). There are generally trade-offs between expressivity and computational cost with equivariant models (Maron et al., 2019b; Ravanbakhsh, 2020; Joshi et al., 2023; Xie et al., 2024). Furthermore, they can be cumbersome to design and utilize. More flexible models with fewer rigid assumptions are expected to scale better as available compute grows.

Despite these shortcomings, we see two common motivations for imposing equivariance in generative models. First, using an equivariant generator guarantees that we model an invariant distribution if we start from an invariant prior distribution (Köhler et al., 2020; Xu et al., 2022b). Second, equivariance can serve as a useful inductive bias for the generator: if a data distribution over X is invariant to a group G, then an equivariant model can generalize what it learns for sample $x \in X$ to $g \cdot x, \forall g \in G$ making it more sample efficient.

For many tasks, having an invariant output distribution isn't necessary. Consider molecules: generating novel and stable molecules is more important than ensuring molecules

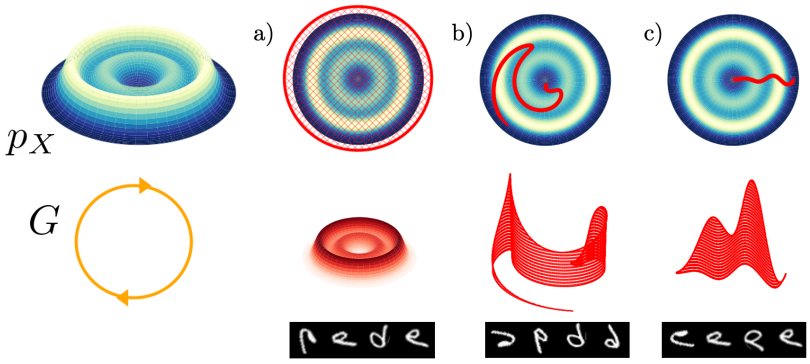

Figure 1: Left: A distribution p_X that's invariant to the group of rotations G. Right: a) A non-equivariant generative model must learn the whole distribution p_X. b) A fixed canonicalization method outputs group elements that map different samples of p_X onto a single slice (red), which may be learned by a generative model. c) A learned canonicalizer can map samples onto a slice that yields a simpler distribution for the generative model to learn.

can be generated in all orientations. Instead of modelling the density on X, it suffices to model over X/G, the set of orbits of X with respect to a group G. To do this we need a map $c : X \to X/G$. Then, we can model a distribution on X/G using any non-equivariant denoiser, giving the added benefit of increased flexibility and speed. Following recent works (Kaba et al., 2023; Kim et al., 2023; Mondal et al., 2024; Allingham et al., 2024), we accomplish this by using a canonicalization function, which transforms data into a standardized or "canonical" form, which is a representative element in its orbit under the group. By parameterizing c as a neural network, the model can discover an optimal way to standardize data, leading to improved representations and performance. This maintains the inductive bias benefit of equivariance by ensuring similarly structured data is oriented similarly for the generator.

We demonstrate this approach in our experiments using denoising diffusion models (Ho et al., 2020). We show on a molecular dataset that using a simple non-equivariant denoising network along with a learned canonicalizer results in higher-quality samples than using an equivariant network, while halving inference time.

2. Methods

Our goal is to learn a distribution over the orbits X/G instead of directly learning a distribution over X. For this, we define the *quotient map* $c : X \to X/G$ as $c(x) = h(x)^{-1} x$, following Bloem-Reddy and Teh (2020). The function $h : X \to G$ is a relaxed equivariant canonicalization function Kaba and Ravanbakhsh (2023). The invariant quotient map chooses a representative from the orbit $G \cdot x$ of any sample x.

Given the quotient map c and a data distribution $p_X(x)$, we then consider modelling the distribution over the orbit representatives as $p_{X/G}(c(x))$.

This essentially corresponds to projecting the distributions over slices defined by orbit-representatives, as shown in Figure 1. The distribution $p_{X/G}$ is then modelled using a non-equivariant generator network ϕ. The resulting training procedure for a denoising diffusion model is described in Algorithm 1.

Algorithm 1: Training a denoiser to produce canonical samples.

Input: Data point x, denoising network ϕ, equivariant network h ;
Define canonicalizer $c : X \to X/G$ by $c(x) = h(x)^{-1}x$;
While not converged:
  Sample $t \sim U(0, ..., T), \epsilon \sim \mathcal{N}(0, I)$;
  Compute $z_t = \alpha_t c(x) + \sigma_t \epsilon$;
  Update h, ϕ to minimize $\|\epsilon - \phi(z_t, t)\|^2$;

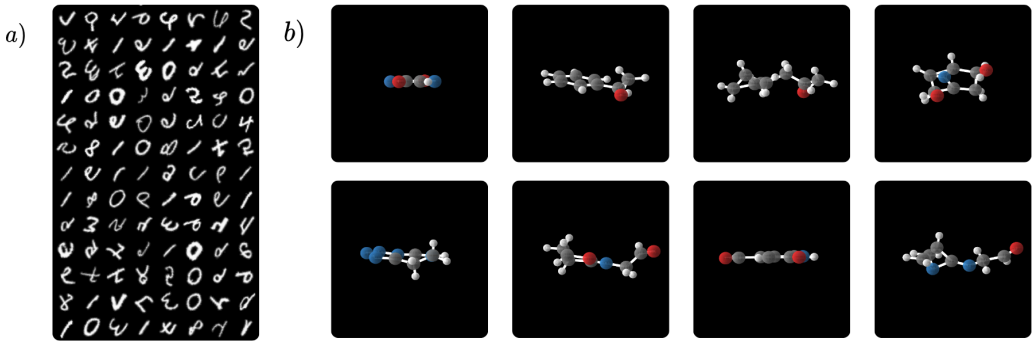

Figure 2: a) Output of the learned canonicalizer on RotMNIST after training showing aligned digits. b) Select molecules generated from canon+GDM in a near-canonical pose.

The central hypothesis of our framework is that projecting the distribution over orbits should result in a considerably simpler learning problem compared to modelling the full distribution. Furthermore, learning the quotient map via the canonicalization function h, should slice the distribution in a way that is easier to model for the generator ϕ.

It is common practice to set the center of mass to be zero when modelling 3D molecules in order to be invariant to translations, as one cannot define a translationally-invariant prior distribution in Euclidean space (Köhler et al., 2020). Our method can be seen as a generalization of the center of this mass removal procedure to any group, and we therefore similarly circumvent the issue of needing an invariant prior.

3. Experiments

3.1. Rotated MNIST Generation

We start with a simple task on images to demonstrate our framework. A UNet (Ronneberger et al., 2015) is trained to denoise rotated MNIST digits (Larochelle et al., 2007), a common classification benchmark for equivariant architechtures. We then train the same network with discrete image canonicalization using an ESCNN (Weiler and Cesa, 2019a) with 8, 16 and 32 discretizations. Results in Table 2 (appendix B) show an improved loss on the test set after canonicalization is applied. This method gives some visual intuition for what canonicalization is doing. Figure 2 shows the output of a trained canonicalization network. Digits across all classes are aligned, resulting in an easier task for the denoiser.

Table 1: QM9 Results.

Model	NLL	Mol stable	At stable	Valid	Unique	s/sample
GDM	-105.8	70.6	97.5	88.6	99.8	**0.55**
EDM	-110.7 ± 1.7	82.0 ± 0.4	**98.7** ± 0.1	91.9 ± 0.5	90.7 ± 0.6	1.16
canon(fr)+GDM	-104.1	82.1	97.8	92.9	99.5	**0.55**
canon+GDM	**-117.4** ± 1.2	**84.6** ± 0.4	98.4 ± 0.2	**94.4** ± 0.3	99.7 ± 0.2	**0.55**

3.2. QM9 Molecule Generation

Experiment Setup. Models are trained to generate molecules in the QM9 dataset (Ramakrishnan et al., 2014). As baselines, an Equivariant Diffusion Model (EDM) (Hoogeboom et al., 2022a) and non-equivariant GNN Diffusion Model (GDM) with identical hyperparameters are trained on the task. Then, a multi-channel EGNN (Levy et al., 2023; Satorras et al., 2021) is introduced as the canonicalization network, outputting a rotation matrix that is applied to the molecule before denoising with the GDM backbone (canon+GDM). We show that even an unlearned canonicalizer with frozen weights (canon(fr)+GDM) aids the GDM in producing stable molecules.

Evaluation Setup. A similar evaluation setup to Satorras et al. (2022) and Hoogeboom et al. (2022a) is used. We report the test NLL in addition to a number of desirable features of generated molecules relevant to the drug discovery pipeline: atom stability, molecule stability, validity, uniqueness and inference time.

Results. Our findings (Table 1) show that canon+GDM outperforms EDM across several metrics while halving inference time. Additionally, the unlearned canonicalizer, canon(fr)+GDM, is still able to generate stable molecules with results similar to EDM. Figure 2 showsf samples from the canon+GDM model. These molecules are generated in a near-canonical orientation: planar molecules are entirely confined to the xy plane while other oblong molecules are primarily horizontal rather than vertical. This supports the notion that the canonicalizer is able to learn to transform to poses that make denoising easier and that the denoiser is able to generate molecules in such poses.

Although we used a simple non-equivariant counterpart of the equivariant generative model from Hoogeboom et al. (2022b) as a denoising network in this experiment, our flexible method can accommodate any arbitrarily expressive non-equivariant model.

4. Conclusion

We show that contrary to previous work, modelling the entire invariant distribution p_X is unnecessary and simply modeling $p_{X/G}$ via canonicalization is a more tractable task. Using a learned canonicalizer allows us to maintain the inductive bias benefit of equivariant neural networks while gaining flexibility and efficiency by the use of a fast expressive non-equivariant generator. In future work, we expect the flexibility of this method will allow for the use of even more powerful generators for generative modelling tasks across domains with symmetry.

References

James Urquhart Allingham, Bruno Kacper Mlodozeniec, Shreyas Padhy, Javier Antorán, David Krueger, Richard E Turner, Eric Nalisnick, and José Miguel Hernández-Lobato. A generative model of symmetry transformations. *arXiv preprint arXiv:2403.01946*, 2024.

Brandon Anderson, Truong Son Hy, and Risi Kondor. Cormorant: Covariant molecular neural networks. In *Advances in Neural Information Processing Systems*, volume 32, pages 14510–14519, 2019.

Benjamin Bloem-Reddy and Yee Whye Teh. Probabilistic symmetries and invariant neural networks. *The Journal of Machine Learning Research*, 21(1):3535–3595, 2020.

Joey Bose, Tara Akhound-Sadegh, Guillaume Huguet, Kilian Fatras, Jarrid Rector-Brooks, Cheng-Hao Liu, Andrei Cristian Nica, Maksym Korablyov, Michael M. Bronstein, and Alexander Tong. SE(3)-stochastic flow matching for protein backbone generation. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=kJFIH23hXb.

Johann Brandstetter, Andrea Grisafi, Michele Ceriotti, Clémence Corminboeuf, and Matthias Rupp. Geometric deep learning on molecular representations. *arXiv preprint arXiv:2206.03083*, 2022.

Chi Chen, Weike Ye, Yunxing Zuo, Chen Zheng, and Shyue Ping Ong. Equivariant message passing for the prediction of tensorial properties and molecular spectra. *arXiv preprint arXiv:2003.03123*, 2020.

Taco Cohen and Max Welling. Group equivariant convolutional networks. In *International Conference on Machine Learning*, pages 2990–2999, 2016.

Taco S. Cohen, Mario Geiger, Jonas Köhler, and Max Welling. Spherical cnns. In *International Conference on Learning Representations (ICLR)*, 2018.

Congyue Deng, Or Litany, Yueqi Duan, Adrien Poulenard, Andrea Tagliasacchi, and Leonidas J Guibas. Vector neurons: A general framework for so (3)-equivariant networks. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 12200–12209, 2021.

Carlos Esteves, Christine Allen-Blanchette, Ameesh Makadia, and Kostas Daniilidis. Learning so(3) equivariant representations with spherical cnns. In *Advances in Neural Information Processing Systems (NeurIPS)*, pages 6541–6550, 2018.

Luca Falorsi, Taco S. Cohen, and Max Welling. Explorations in homeomorphism learning for equivariant diffusion models. *arXiv preprint arXiv:1807.10306*, 2018.

Marc Finzi, Samuel Stanton, Pavel Izmailov, and Andrew Gordon Wilson. Generalizing convolutional neural networks for equivariance to lie groups on arbitrary continuous data. In *Proceedings of the 37th International Conference on Machine Learning (ICML)*, pages 3165–3176, 2020.

Fabian B Fuchs, Daniel E Worrall, Volker Fischer, and Max Welling. SE(3)-transformers: 3d roto-translation equivariant attention networks. In *Advances in Neural Information Processing Systems*, volume 33, pages 1970–1981, 2020.

Justin Gilmer, Samuel S. Schoenholz, Patrick F. Riley, Oriol Vinyals, and George E. Dahl. Neural message passing for quantum chemistry. *arXiv preprint arXiv:1704.01212*, 2017.

Jiaxu Han, Haizhao Yang, and Jianfeng Lu. Uniformly stable deep neural networks for dynamical systems. *arXiv preprint arXiv:1912.01328*, 2019.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems*, pages 6840–6851, 2020.

Emiel Hoogeboom, Victor Garcia Satorras, Clément Vignac, and Max Welling. Equivariant diffusion for molecule generation in 3d, 2022a. URL https://arxiv.org/abs/2203.17003.

Emiel Hoogeboom, Vıctor Garcia Satorras, Clément Vignac, and Max Welling. Equivariant diffusion for molecule generation in 3d. In *International Conference on Machine Learning*, pages 8867–8887. PMLR, 2022b.

Chaitanya K. Joshi, Cristian Bodnar, Simon V Mathis, Taco Cohen, and Pietro Lio. On the expressive power of geometric graph neural networks. In Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt, Sivan Sabato, and Jonathan Scarlett, editors, *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pages 15330–15355. PMLR, 23–29 Jul 2023. URL https://proceedings.mlr.press/v202/joshi23a.html.

Sékou-Oumar Kaba and Siamak Ravanbakhsh. Symmetry breaking and equivariant neural networks. *arXiv preprint arXiv:2312.09016*, 2023.

Sékou-Oumar Kaba, Arnab Kumar Mondal, Yan Zhang, Yoshua Bengio, and Siamak Ravanbakhsh. Equivariance with learned canonicalization functions. In *International Conference on Machine Learning*, pages 15546–15566. PMLR, 2023.

Nicolas Keriven and Gabriel Peyré. Universal invariant and equivariant graph neural networks. *Advances in Neural Information Processing Systems*, 32, 2019.

Jinwoo Kim, Dat Nguyen, Ayhan Suleymanzade, Hyeokjun An, and Seunghoon Hong. Learning probabilistic symmetrization for architecture agnostic equivariance. In A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine, editors, *Advances in Neural Information Processing Systems*, volume 36, pages 18582–18612. Curran Associates, Inc., 2023. URL https://proceedings.neurips.cc/paper_files/paper/2023/file/3b5c7c9c5c7bd77eb73d0baec7a07165-Paper-Conference.pdf.

Jonas Köhler, Leon Klein, and Frank Noé. Equivariant flows: exact likelihood generative learning for symmetric densities. In *International conference on machine learning*, pages 5361–5370. PMLR, 2020.

Risi Kondor, Zhen Lin, and Shubhendu Trivedi. Clebsch–gordan nets: a fully fourier space spherical convolutional neural network. In *Advances in Neural Information Processing Systems*, volume 31, pages 10117–10126, 2018.

Hugo Larochelle, Dumitru Erhan, Aaron Courville, James Bergstra, and Yoshua Bengio. An empirical evaluation of deep architectures on problems with many factors of variation. In *Proceedings of the 24th International Conference on Machine Learning*, ICML '07, page 473–480, New York, NY, USA, 2007. Association for Computing Machinery. ISBN 9781595937933. doi: 10.1145/1273496.1273556. URL https://doi.org/10.1145/1273496.1273556.

Daniel Levy, Sékou-Oumar Kaba, Carmelo Gonzales, Santiago Miret, and Siamak Ravanbakhsh. Using multiple vector channels improves e(n)-equivariant graph neural networks, 2023. URL https://arxiv.org/abs/2309.03139.

Haggai Maron, Heli Ben-Hamu, Nadav Shamir, and Yaron Lipman. Universality of invariant networks. In *International Conference on Machine Learning*, pages 4363–4371, 2019a.

Haggai Maron, Ethan Fetaya, Nimrod Segol, and Yaron Lipman. On the universality of invariant networks. In Kamalika Chaudhuri and Ruslan Salakhutdinov, editors, *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 4363–4371. PMLR, 09–15 Jun 2019b. URL https://proceedings.mlr.press/v97/maron19a.html.

Arnab Kumar Mondal, Siba Smarak Panigrahi, Sékou-Oumar Kaba, Sai Rajeswar, and Siamak Ravanbakhsh. Equivariant adaptation of large pretrained models, 2023. URL https://arxiv.org/abs/2310.01647.

Arnab Kumar Mondal, Siba Smarak Panigrahi, Oumar Kaba, Sai Rajeswar Mudumba, and Siamak Ravanbakhsh. Equivariant adaptation of large pretrained models. *Advances in Neural Information Processing Systems*, 36, 2024.

Raghunathan Ramakrishnan, Pavlo O. Dral, Matthias Rupp, and O. Anatole von Lilienfeld. Quantum chemistry structures and properties of 134 kilo molecules. *Scientific Data*, 1 (1), aug 2014. doi: 10.1038/sdata.2014.22. URL https://doi.org/10.1038%2Fsdata.2014.22.

Siamak Ravanbakhsh. Universal equivariant multilayer perceptrons. In Hal Daumé III and Aarti Singh, editors, *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pages 7996–8006. PMLR, 13–18 Jul 2020. URL https://proceedings.mlr.press/v119/ravanbakhsh20a.html.

Siamak Ravanbakhsh, Jeff Schneider, and Barnabas Poczos. Equivariance through parameter-sharing. In *International Conference on Machine Learning*, pages 2892–2901. PMLR, 2017.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation, 2015. URL https://arxiv.org/abs/1505.04597.

Victor Garcia Satorras, Emiel Hoogeboom, and Max Welling. E(n) equivariant graph neural networks. In *International Conference on Machine Learning*, pages 9323–9332, 2021.

Victor Garcia Satorras, Emiel Hoogeboom, Fabian B. Fuchs, Ingmar Posner, and Max Welling. E(n) equivariant normalizing flows, 2022. URL https://arxiv.org/abs/2105.09016.

Kristof Schütt, Pieter-Jan Kindermans, Huziel Enoc Sauceda Felix, Stefan Chmiela, Alexandre Tkatchenko, and Klaus-Robert Müller. Schnet: A continuous-filter convolutional neural network for modeling quantum interactions. In *Advances in Neural Information Processing Systems*, pages 991–1001, 2017.

Kristof T. Schütt, Michael Gastegger, Alexandre Tkatchenko, Klaus-Robert Müller, and Reinhard J. Maurer. Equivariant message passing for the prediction of tensorial properties and molecular spectra. *arXiv preprint arXiv:2102.03150*, 2021.

Chence Shi, Shitong Luo, Minkai Xu, and Jian Tang. Learning gradient fields for molecular conformation generation. In *International conference on machine learning*, pages 9558–9568. PMLR, 2021.

Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In *International Conference on Machine Learning*, pages 2256–2265, 2015.

Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=PxTIG12RRHS.

Nathaniel Thomas, Tess Smidt, Steven Kearnes, Lusann Yang, Li Li, Karthik Kohlhoff, and Patrick Riley. Tensor field networks: Rotation-and translation-equivariant neural networks for 3d point clouds. *arXiv preprint arXiv:1802.08219*, 2018.

Oliver Thorsten Unke and Markus Meuwly. Physnet: A neural network for predicting energies, forces, dipole moments, and partial charges. *Journal of Chemical Theory and Computation*, 15(6):3678–3693, 2019.

Clement Vignac, Nagham Osman, Laura Toni, and Pascal Frossard. Midi: Mixed graph and 3d denoising diffusion for molecule generation, 2023. URL https://arxiv.org/abs/2302.09048.

Yatao Wang, Taco Cohen, and Max Welling. Equivariant diffusion for molecule generation in 3d. *arXiv preprint arXiv:2202.05116*, 2022.

Maurice Weiler and Gabriele Cesa. 3d steerable cnns: Learning rotationally equivariant features in volumetric data. In *Advances in Neural Information Processing Systems (NeurIPS)*, pages 10402–10413, 2018.

Maurice Weiler and Gabriele Cesa. General E(2)-Equivariant Steerable CNNs. In *Conference on Neural Information Processing Systems (NeurIPS)*, 2019a.

Maurice Weiler and Gabriele Cesa. General E(2)-equivariant steerable CNNs. In *Advances in Neural Information Processing Systems (NeurIPS)*, pages 14334–14345, 2019b.

Daniel E. Worrall, Stephan J. Garbin, Daniyar Turmukhambetov, and Gabriel J. Brostow. Harmonic networks: Deep translation and rotation equivariance. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5028–5037, 2017.

YuQing Xie, Ameya Daigavane, Mit Kotak, and Tess Smidt. The price of freedom: Exploring tradeoffs between expressivity and computational efficiency in equivariant tensor products. In *ICML 2024 Workshop on Geometry-grounded Representation Learning and Generative Modeling*, 2024. URL https://openreview.net/forum?id=0HHidbjwcf.

Minkai Xu, Lantao Yu, Yang Song, Chence Shi, Stefano Ermon, and Jian Tang. Geodiff: A geometric diffusion model for molecular conformation generation. In *International Conference on Learning Representations*, 2022a. URL https://openreview.net/forum?id=2J3PcoZnqh6.

Minkai Xu, Lantao Yu, Yang Song, Chence Shi, Stefano Ermon, and Jian Tang. Geodiff: a geometric diffusion model for molecular conformation generation, 2022b. URL https://arxiv.org/abs/2203.02923.

Yanzhi Zhou and Jiashi Feng. Group-theory based decomposition of convolutional networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6530–6539, 2020.

Appendix A. Related Work

Equivariant neural networks Equivariant neural networks have been successfully applied in fields such as image processing Cohen and Welling (2016); Worrall et al. (2017); Weiler and Cesa (2019b), 3D data processing Thomas et al. (2018); Fuchs et al. (2020); Weiler and Cesa (2018); Deng et al. (2021); Esteves et al. (2018), graph analysis Maron et al. (2019a); Keriven and Peyré (2019); Chen et al. (2020), physical simulations Finzi et al. (2020); Han et al. (2019); Brandstetter et al. (2022), and AI for science applications such as molecular modeling and computational chemistry Schütt et al. (2017); Anderson et al. (2019); Schütt et al. (2021); Satorras et al. (2021); Gilmer et al. (2017); Unke and Meuwly (2019), leading to improved generalization and sample efficiency. Despite their benefits, existing techniques to build equivariant neural networks face several challenges. One significant issue is ***computational complexity***, as specialized operations like group convolutions or spherical harmonics are computationally expensive Fuchs et al. (2020); Kondor et al. (2018); Weiler and Cesa (2019b); Cohen et al. (2018). Additionally, there are ***expressivity constraints***, since limiting models to equivariant functions may restrict their ability to capture complex patterns not perfectly aligned with the symmetry group Maron et al. (2019a); Ravanbakhsh et al. (2017); Zhou and Feng (2020). Moreover, the design of these networks inherently presents ***significant complexities and challenges***, necessitating tailored architectural solutions that add layers of difficulty to both development and practical implementation.

Canonicalization for architecture-agnostic equivariance Canonicalization offers an alternative approach to designing specialized neural networks by transforming data into a standardized or "canonical" form, effectively removing symmetry-related variability. The canonicalization function maps each data point to a representative element in its orbit under the group. Recent works propose learning the canonicalization function jointly with the main task-specific network Kaba et al. (2023); Kim et al. (2023); Mondal et al. (2024); Allingham et al. (2024). By parameterizing c as a neural network, the model can discover an optimal way to standardize data, leading to improved representations and performance.

Equivariance and Diffusion Models Diffusion models have emerged as a powerful class of generative models, achieving remarkable success in generating high-fidelity samples across various domains Sohl-Dickstein et al. (2015); Ho et al. (2020); Song et al. (2021). These models learn data distributions by reversing a diffusion process that incrementally adds Gaussian noise to the data. Incorporating equivariance into diffusion models can enhance their performance on data with inherent symmetries Hoogeboom et al. (2022b); Xu et al. (2022a); Falorsi et al. (2018); Wang et al. (2022). Equivariant diffusion models ensure that the generative process respects the symmetries of the data, leading to more efficient learning and better generalization. However, integrating equivariance into diffusion models also inherits the challenges associated with equivariant neural networks.

Appendix B. Experimental Details

B.1. RotMNIST Dataset

A denoising U-Net is trained with a base dimension of 64. The discrete image canonicalizer is implemented using EquiAdapt (Mondal et al., 2023). An Equivariant Steerable Neural Network (ESCNN) (Weiler and Cesa, 2019a) with 16 channels, 5 layers and kernel size 5 is used for canonicalization. Both networks are trained for 1000 epochs on an NVIDIA A100-SXM4-80GB over roughly 9 hours.

Table 2: RotMNIST Results: reconstruction L2 distance of denoising model over the test set with randomly sampled noising timesteps.

Model	Test Reconstruction Loss
UNet	0.0312
canon(p32)+UNet	0.0264
canon(p16)+UNet	**0.0255**
canon(p8)+UNet	0.0264
canon(p16, fr)+UNet	0.0264

B.2. QM9 Dataset

The Equivaraint Diffusion Model (EDM) and Graph Diffusion Model (GDM) are trained with hyperparameters equivalent to Hoogeboom et al. (2022b), namely 1000 diffusion steps with polynomial noise schedule and precision 1e-5. An L2 denoising loss is used with batch

size 64. The denoiser has 256 node features and 9 layers. An EMA decay of 0.9999 is used. For the canonicalization network, a multi-channel EGNN (Levy et al., 2023; Satorras et al., 2021) is used to output two rotation vectors which are then orthonomalized into a rotation matrix via modified Gram-Schmidt. The network has 9 layers and 64 node features. A single run on a Quadro RTX 8000 takes roughly 4 days.