CaloRex: From Self-Supervised Representation Learning to Fast Calorimeter Simulation

Cheng Jiang
School of Physics and Astronomy
University of Edinburgh
Edinburgh EH9 3FD, United Kingdom
chjiang@cern.ch

Sitian Qian
Department of Physics
University of Wisconsin-Madison
Madison, WI 53706
sitian.qian@cern.ch

Huilin Qu
State Key Laboratory of Dark Matter Physics
Tsung-Dao Lee Institute & School of Physics and Astronomy
Shanghai Jiao Tong University
Shanghai 200240, China
hqu@cern.ch

Abstract

Self-supervised learning provides a promising approach for extracting generalizable features from unlabeled calorimeter data, yet its benefits beyond standard classification tasks remain largely unexplored. We introduce CaloRex, a DINO-inspired framework that learns spatially coherent, dense representations from high granularity calorimeter showers by enforcing agreement between local and global views. We first validate these representations through linear probing and robustness tests under input perturbations, demonstrating significant improvements over models trained from scratch. We then show that the learned features provide a strong inductive bias for conditional generative modeling: when used as conditioning variables for a flow matching model, they yield lower training losses and more accurate physics distributions compared to baseline generative models, with representation alignment further guiding the learning process. These results establish self-supervised pretraining as a practical bridge between representation learning and fast calorimeter simulation.

1 Introduction

Modern machine learning in high energy physics (HEP) increasingly relies on high dimensional calorimeter data for both reconstruction and fast simulation tasks. The standard workflow depends on Monte Carlo (MC) [1–3] simulations for labeled training data, but imperfect modeling of detector responses can limit model performance. Meanwhile, generative approaches to fast calorimeter simulation [4–8] have advanced rapidly, yet they typically learn entirely from scratch without leveraging the rich structure already present in shower data.

Self-supervised learning (SSL) offers a way to extract reusable features directly from unlabeled showers. In computer vision, methods such as DINO [9–11] learn representations that cluster by semantic content without explicit supervision, producing features that transfer effectively to diverse downstream tasks. The multi-scale, spatially structured nature of calorimeter data makes it a natural fit for such approaches.

We introduce CaloRex, a DINO-inspired SSL framework applied to high granularity calorimeter showers from the International Large Detector (ILD) [12, 13]. We first show that the learned embed-

40th Conference on Neural Information Processing Systems (NeurIPS 2026). Workshop: PAI26.

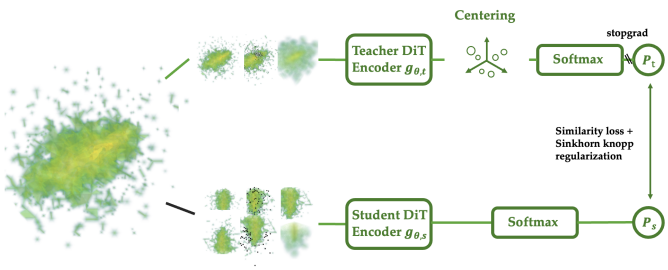


Figure 1: Schematic architecture of the CaloRex.

dings cluster by physical factors and exhibit strong invariance under a range of input transformations. An additional improvement is that these dense features, when used as conditioning for generative models, yield faster convergence and more accurate physics distributions than training from scratch. These findings show that self-supervised pretraining successfully connects learned representations with efficient calorimeter simulation.

2 Related Work

SSL in high energy physics and computer vision SSL methods like JetCLR, MPM, and J-JEPA [14–19] extract structure from particle data without labels. In computer vision, contrastive and self-distillation methods such as SimCLR, MoCo, SwAV, and DINO [20–24, 9, 10] learn transferable representations by matching views of the same input. However, standard image augmentations do not transfer directly to calorimeter showers, where global spatial correlations and structured energy deposition are crucial.

Generative models for fast simulation Extensive prior work uses generative models [25–45] to produce fast, high fidelity simulated showers. These approaches learn entirely from scratch and focus on data synthesis. An open question is whether pretrained representations can serve as a more reliable basis for generative models. In this work, we bridges this gap by leveraging learned dense shower features, effectively streamlining convergence and improving the fidelity of the generated output.

3 Method

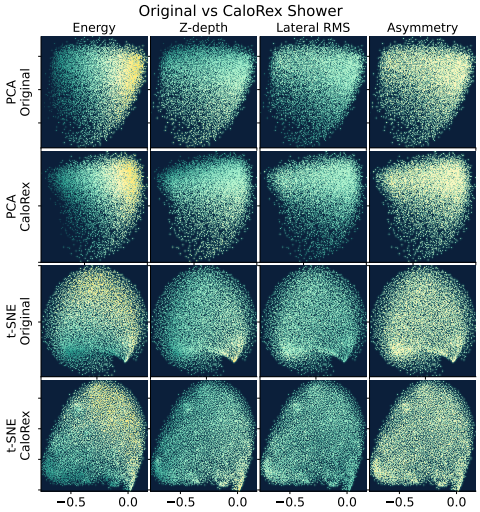
Inputs Each shower is a high granularity 3D voxel grid of energy deposits in the ILD calorimeter. We normalize each voxel x_i by the incident energy ($\tilde{x}_i = x_i/E_{\text{inc}}$), apply a logit transform to reduce skewness, and standardize using training set statistics.

DINO objective We use a Diffusion Transformer (DiT) [46] backbone in a dual student–teacher network. A student encoder g_θ and an EMA teacher encoder g_ξ process multiple stochastic views of each shower: global views V_g and local views V_l. Training minimizes the cross entropy between student predictions and temperature scaled teacher targets:

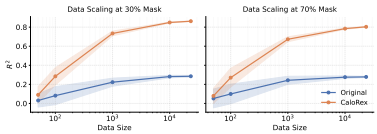
$$\mathcal{L}_{\text{CE}} = -\sum_{v \in V_g} \sum_{\substack{v' \in V \\ v' \neq v}} P_t(v) \log P_s(v'),$$

where P_s uses a softmax with temperature τ_s and P_t is obtained via Sinkhorn-Knopp [47, 48] normalization at temperature τ_t. We add a KoLeo entropic regularizer [49, 50] to encourage uniform feature density: $\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{CE}} + \lambda \mathcal{L}_{\text{KoLeo}}$.

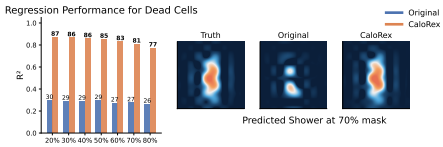
Augmentations Global and local views are constructed using physically motivated augmentations: discrete spatial symmetries (90° rotations, x-y flips), energy smearing via Gaussian noise, low energy cell masking, 3D cutouts, and a multi-crop strategy in which local views are smaller contiguous fixed-size spatial windows extracted from different spatial origins within each calorimeter sub-layer (e.g. core and periphery), while global views the span full shower.



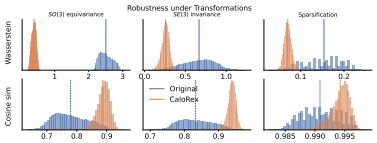
(a) PCA and t-SNE projections colored by physical quantity (columns: energy, depth, lateral spread, asymmetry) using a continuous colormap (low = dark, high = bright). CaloRex rows show smoother color gradients, indicating that the embedding space organizes by these quantities.



(b) Linear probing R^2 vs. number of labeled samples used for regression, at 30% and 70% cell masking.



(c) R^2 across masked cell percentages, with example shower reconstructions at 70% masking.



(d) Wasserstein distance and cosine similarity under $SO(3)$, $SE(3)$, and sparsification.

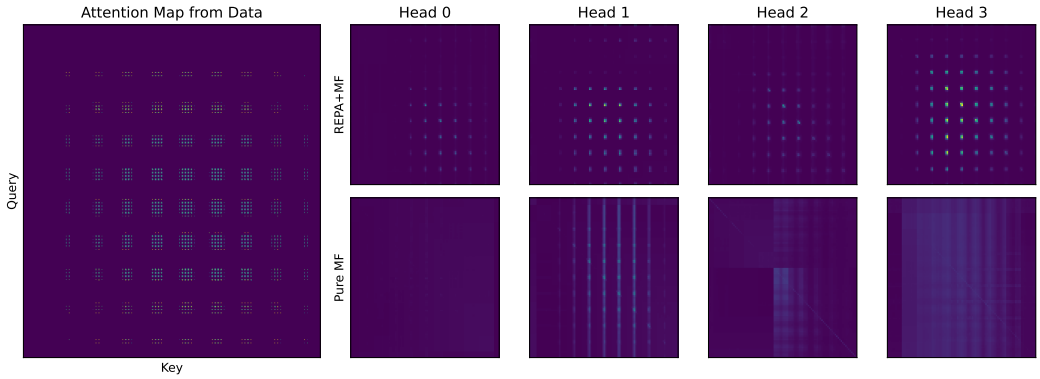
Figure 2: **Analysis of learned representations.** (a) PCA/t-SNE embeddings; a continuous colormap encodes each physical quantity (see subcaption). (b–d) Quantitative comparisons where blue = original showers and orange = CaloRex.

4 Experiments

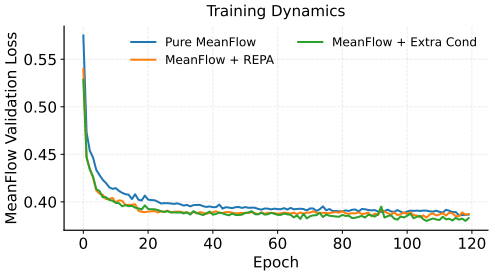
Training. We use approximately 50,000 simulated photon showers from the ILD electromagnetic calorimeter, covering an energy range from 1 GeV to 1 TeV. The dataset is split into 80% for training, 10% for validation, and 10% for testing. The model is trained for 200 epochs using the AdamW optimizer with a learning rate of 1×10^{-4}. For each shower, we extract 3 global views and 12 local views. The student temperature is set to $\tau_s = 0.1$, while the teacher temperature is linearly increased from 0.04 to 0.07 over the first 40 epochs. The KoLeo regularization weight is set to $\lambda = 0.1$. All experiments are conducted on a single NVIDIA A100 GPU.

Representation quality (Figure 2). PCA and t-SNE projections (Figure 2(a)) show that CaloRex embeddings evolve smoothly across physical parameters such as energy, depth, lateral spread, and asymmetry. Each quantity forms a coherent gradient within the embedding space, whereas these patterns are less apparent in the baseline representations. Quantitatively, linear probing via Ridge regression (Figure 2b,c), trained to reconstruct the masked-region energy from the visible patches, reveals a significant disparity in how each representation scales with available statistics. CaloRex exhibits strong data scaling and robustness, maintaining high predictive performance with R^2 values of 0.87 at 20% masking and 0.77 even under extreme 80% masking. In contrast, the raw voxel baseline demonstrates poor scaling, with performance staying around $R^2 \approx 0.30$ to 0.26 under the same conditions. The geometric robustness of the representation is further evaluated by comparing Wasserstein distances and cosine similarities under various transformations to mimic detector misalignment and geometric mismodelling. When subjected to SO(3) rotations and SE(3) transformation, CaloRex learned embeddings demonstrate significantly lower Wasserstein distances compared to the original shower representations, indicating a more stable and invariant latent manifold.

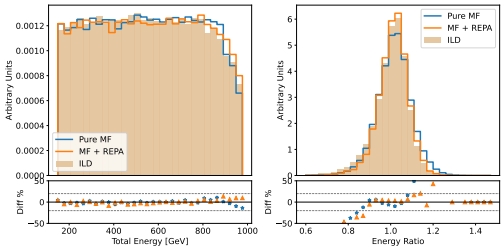
Generative performance (Figure 3). We test whether CaloRex features improve downstream generative modeling via two strategies with a MeanFlow [8] approach: (1) representation alignment (REPA) [51], which aligns intermediate features with frozen CaloRex embeddings using a cosine



(a) Left: Reference attention map from data showing block diagonal structure reflecting spatial locality in the shower. Right: REPA+MF heads recover sparse, localized patterns; Pure MF heads are diffuse.



(b) Validation loss curves. MF+REPA (orange) converges faster and reaches a lower loss than Pure MF (blue).



(c) Total energy and energy ratio distributions. MF+REPA (orange) tracks the ILD reference (gold) more closely than Pure MF (blue), with smaller residuals in the ratio panels.

Figure 3: **Generative modeling with CaloRex features.** Representation alignment (REPA) with MeanFlow consistently outperforms the pure MeanFlow baseline in attention structure, convergence speed, and physics fidelity on the ILD dataset.

similarity loss,

$$\mathcal{L}_{\cos} = -\frac{1}{N}\sum_{i=1}^{N} \frac{z_i \cdot \tilde{z}_i}{\|z_i\|_2 \, \|\tilde{z}_i\|_2},$$

where z_i and $\tilde{z}_i$ denote the normalized student and target feature vectors, respectively, and (2) extra conditioning, which concatenates CaloRex features as additional input to the flow model.

The attention maps (Figure 3(a)) provide insight into why the model should outperform the baseline: reference attention maps from data exhibits a block diagonal structure reflecting spatial locality where nearby voxels attend strongly to each other. REPA+MeanFlow heads recover sparse, localized patterns consistent with this structure, while baseline MeanFlow model heads produce diffuse, unstructured attention. Both REPA and extra-conditioning variants converge faster and reach lower validation loss (Figure 3(b)), with REPA showing the fastest early convergence. Most importantly, MF+REPA yields more accurate physics: the ILD total energy and energy ratio distributions (Figure 3(c)) are more closely aligned with the reference, with consistently improved agreement than baseline model.

5 Conclusion

We presented CaloRex, a self supervised framework that extracts robust and transferable representations from calorimeter showers, with the aim of reducing reliance on extensively labeled simulations in high energy physics. Our evaluations show that CaloRex provides a strong inductive bias that scales well and remains robust to dataset variations and domain shifts, addressing a key limitation

of supervised models. Under severe input masking, linear probing accuracy improves from 30% to 87%. Furthermore, leveraging the learned dense features leads to faster convergence and more accurate modeling of physics distributions, improving both representation learning and conditional generative modeling. Open directions include domain adaptation for real detector data, augmentation ablations across detector geometries, benchmarking MeanFlow against other flow-based backbones, and extensions to point cloud or graph representations for sparse collision environments.

Acknowledgments and Disclosure of Funding

References

[1] S. Agostinelli et al. Geant4 – a simulation toolkit. *Nucl. Instrum. Meth. A*, 506:250–303, 2003.

[2] J. Allison et al. Geant4 developments and applications. *IEEE Trans. Nucl. Sci.*, 53(1):270–278, 2006.

[3] J. Allison et al. Recent developments in geant4. *Nucl. Instrum. Meth. A*, 835:186–225, 2016.

[4] Michela Paganini, Luke de Oliveira, and Benjamin Nachman. Accelerating science with generative adversarial networks: An application to 3d particle showers in multilayer calorimeters. *Phys. Rev. Lett.*, 120(4):042003, 2018.

[5] Claudius Krause and David Shih. Fast and accurate simulations of calorimeter showers with normalizing flows. *Physical Review D*, 107(11), June 2023.

[6] Oz Amram and Kevin Pedro. Denoising diffusion models with geometry adaptation for high fidelity calorimeter simulation. *Physical Review D*, 108(7):072014, 2023.

[7] Luigi Favaro, Ayodele Ore, Simon P. Schweitzer, and Tilman Plehn. Calodream – detector response emulation via attentive flow matching. 2024.

[8] Zhengyang Geng, Mingyang Deng, Xingjian Bai, J. Zico Kolter, and Kaiming He. Mean flows for one-step generative modeling. 2025.

[9] Mathilde Caron, Hugo Touvron, Ishan Misra, Hervé Jégou, et al. Emerging properties in self-supervised vision transformers. In *International Conference on Computer Vision (ICCV)*, 2021.

[10] Maxime Oquab, Timothée Darcet, et al. Dinov2: Learning robust visual features without supervision. *arXiv preprint arXiv:2304.07193*, 2023.

[11] Oriane Siméoni, Huy V. Vo, et al. Dinov3, 2025.

[12] Halina Abramowicz et al. International large detector: Interim design report. 2020.

[13] José Repond et al. Design and electronics commissioning of the physics prototype of a si-w electromagnetic calorimeter for the international linear collider. *JINST*, 3:P08001, 2008.

[14] Barry Dillon, Gregor Kasieczka, Hans Olischlager, Tilman Plehn, Peter Sorrenson, and Lorenz Vogel. Symmetries, safety, and self-supervision. *SciPost Physics*, 12(6), June 2022.

[15] Tobias Golling, Lukas Heinrich, Michael Kagan, Samuel Klein, Matthew Leigh, Margarita Osadchy, and John Andrew Raine. Masked particle modeling on sets: Towards self-supervised high energy physics foundation models, 2024.

[16] Subash Katel, Haoyang Li, Zihan Zhao, Raghav Kansal, Farouk Mokhtar, and Javier Duarte. Learning symmetry-independent jet representations via jet-based joint embedding predictive architecture, 2024.

[17] Joschka Birk, Anna Hallin, and Gregor Kasieczka. Omnijet-α: the first cross-task foundation model for particle physics. *Machine Learning: Science and Technology*, 5(3):035031, August 2024.

[18] Jonas Birk, Frank Gaede, Anna Hallin, Gregor Kasieczka, Martina Mozzanica, and Hjalte Rose. Omnijet-α_c: Learning point cloud calorimeter simulations using generative transformers. *JINST*, 20(07):P07007, 2025.

[19] Zichun Hao, Raghav Kansal, Abhijith Gandrakota, Chang Sun, Jennifer Ngadiuba, Javier Duarte, and Maria Spiropulu. RINO: Renormalization Group Invariance with No Labels. 9 2025.

[20] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *International Conference on Machine Learning (ICML)*, 2020.

[21] Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. Momentum contrast for unsupervised visual representation learning. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.

[22] Mathilde Caron, Ishan Misra, Julien Mairal, Priya Goyal, Piotr Bojanowski, and Armand Joulin. Unsupervised learning of visual features by contrasting cluster assignments. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2020.

[23] Mahmoud Assran, Quentin Duval, Ishan Misra, Piotr Bojanowski, Pascal Vincent, Michael Rabbat, and Yann LeCun. Self-supervised learning from images with a joint-embedding predictive architecture. *arXiv preprint arXiv:2301.08243*, 2023.

[24] Randall Balestriero and Yann LeCun. Lejepa: Provable and scalable self-supervised learning without the heuristics, 2025.

[25] Michela Paganini, Luke de Oliveira, and Benjamin Nachman. Calogan: Simulating 3d high energy particle showers in multilayer electromagnetic calorimeters with generative adversarial networks. *Phys. Rev. D*, 97(1):014021, 2018.

[26] Luke de Oliveira, Michela Paganini, and Benjamin Nachman. Controlling physical attributes in gan-accelerated simulation of electromagnetic calorimeters. *J. Phys. Conf. Ser.*, 1085(4):042017, 2018.

[27] Martin Erdmann, Lukas Geiger, Jonas Glombitza, and David Schmidt. Generating and refining particle detector simulations using the wasserstein distance in adversarial networks. *Comput. Softw. Big Sci.*, 2(1):4, 2018.

[28] Martin Erdmann, Jonas Glombitza, and Thorsten Quast. Precise simulation of electromagnetic calorimeter showers using a wasserstein generative adversarial network. *Comput. Softw. Big Sci.*, 3(1):4, 2019.

[29] Jesse C. Cresswell, Brendan L. Ross, Gabriel Loaiza-Ganem, Humberto Reyes-Gonzalez, Marco Letizia, and Anthony L. Caterini. Caloman: Fast generation of calorimeter showers with density estimation on learned manifolds. In *36th Conference on Neural Information Processing Systems (NeurIPS 2022), Workshop on Machine Learning and the Physical Sciences*, 2022.

[30] Sharif Hoque, Honglu Jia, Akash Abhishek, Mohammad Fadaie, José Q. Toledo-Marín, Tiago Vale, Roger G. Melko, Mikolaj Swiatlowski, and Wojciech T. Fedorko. Caloqvae: Simulating high-energy particle-calorimeter interactions using hybrid quantum-classical generative models. *Eur. Phys. J. C*, 84(12):1244, 2024.

[31] Qi Liu, Chase Shimmin, Xinyu Liu, Eli Shlizerman, Shih-Chieh Li, and Shih-Chieh Hsu. Calo-vq: Vector-quantized two-stage generative model in calorimeter simulation. 2024.

[32] Claudius Krause and David Shih. Caloflow ii: Even faster and still accurate generation of calorimeter showers with normalizing flows. 2023.

[33] Jesse C. Cresswell, Brendan Leigh Ross, Gabriel Loaiza-Ganem, Humberto Reyes-González, Marco Letizia, and Anthony L. Caterini. Caloman: Fast generation of calorimeter showers with density estimation on learned manifolds. In *NeurIPS 2022 Workshop on Machine Learning and the Physical Sciences*, 2022.

[34] Vinicius Mikuni and Benjamin Nachman. Score-based generative models for calorimeter shower simulation. *Phys. Rev. D*, 106(9):092009, 2022.

[35] Fernando Torales Acosta, Vinicius Mikuni, Benjamin Nachman, Miguel Arratia, Bishnu Karki, Ryan Milton, Piyush Karande, and Aaron Angerami. Comparison of point cloud and image-based models for calorimeter fast simulation. 2023.

[36] Vinicius Mikuni and Benjamin Nachman. Caloscore v2: Single-shot calorimeter shower simulation with diffusion models. *JINST*, 19(02):P02001, 2024.

[37] Cheng Jiang, Sitian Qian, and Huilin Qu. Choose your diffusion: Efficient and flexible ways to accelerate the diffusion model in fast high energy physics simulation. *SciPost Phys.*, 18(6):195, 2025.

[38] Dmytro Kobylianskii, Natan Soybelman, Eric Dreyer, and Eilam Gross. Graph-based diffusion model for fast shower generation in calorimeters with irregular geometry. *Phys. Rev. D*, 110(7):072003, 2024.

[39] Cheng Jiang, Sitian Qian, and Huilin Qu. Buff: Boosted decision tree based ultra-fast flow matching. 2024.

[40] Johann Brehmer, Victor Bresó, Pim de Haan, Tilman Plehn, Huilin Qu, Jonas Spinner, and Jesse Thaler. A lorentz-equivariant transformer for all of the lhc. *SciPost Phys.*, 19(4):108, 2025.

[41] Matthew Leigh, Debajyoti Sengupta, Guillaume Quétant, John Andrew Raine, Knut Zoch, and Tobias Golling. Pc-jedi: Diffusion for particle cloud generation in high energy physics. *SciPost Physics*, 16(1), January 2024.

[42] Matthew Leigh, Debajyoti Sengupta, John Andrew Raine, Guillaume Quétant, and Tobias Golling. Faster diffusion model with improved quality for particle cloud generation. *Physical Review D*, 109(1), January 2024.

[43] Piyush Raikwar, Anna Zaborowska, Peter McKeown, Rodrigo Cardoso, Mateusz Piorczynski, and Kian Yeo. A generalisable generative model for multi-detector calorimeter simulation. 2025.

[44] Frank Gaede, Gregor Kasieczka, and Lorenzo Valente. Cross-geometry transfer learning in fast electromagnetic shower simulation. 2025.

[45] Thorsten Buss, Henry Day-Hall, Frank Gaede, Gregor Kasieczka, and Katja Krüger. Allshowers: One model for all calorimeter showers. 2026.

[46] William Peebles and Saining Xie. Scalable diffusion models with transformers. *ICCV*, 2023.

[47] Mathilde Caron, Ishan Misra, Julien Mairal, Priya Goyal, Piotr Bojanowski, and Armand Joulin. Unsupervised learning of visual features by contrasting cluster assignments. In *Advances in Neural Information Processing Systems*, 2020.

[48] Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. In *Advances in Neural Information Processing Systems*, 2013.

[49] Leonenko Kozachenko. Sample estimate of the entropy of a random vector. *Problems of Information Transmission*, 1987.

[50] Alexandre Sablayrolles, Matthijs Douze, Cordelia Schmid, and Hervé Jégou. Spreading vectors for similarity search. In *International Conference on Learning Representations*, 2019.

[51] Sihyun Yu, Sangkyung Kwak, Huiwon Jang, Jongheon Jeong, Jonathan Huang, Jinwoo Shin, and Saining Xie. Representation alignment for generation: Training diffusion transformers is easier than you think, 2025.