1. Supplementary Material

1.1. Empirical Analysis on the impact of our loss on bounding the target domain error

To the best of our knowledge, there are no works that consider domain distributions as interpolations between distributions of domains in *domain generalization* (DG). However, there are few works (Cui et al., 2014; Gopalan et al., 2011) in *domain adaptation* (unlabelled target domain is available) which assume that source and target domains lie on a manifold and create *intermediate domains* through interpolations between them. The intermediate domains aid in bridging the distribution mismatch between the source and target domains. In DG, the target domain is not available. Prior works (Albuquerque et al.; Pham et al., 2023) have upper bounded the error on the target domain as a combination of training error on the source domains, pairwise divergences among the source domains, and divergence between the source and the target domain. A more formal theory for our method is an avenue for future work; however, we do show empirically that our interpolated-based representation learning can be used to bridge the gap between domains. Our model (DNT) has lower Wasserstein divergence among the source domains compared to the baseline DeepAll in Fig. 1. Due to the creation of more interpolated domains, the divergence between a target domain and its closest source domain is reduced too. Due to the reduced source-source and source-target divergences, our models can lower the error on the target domain.

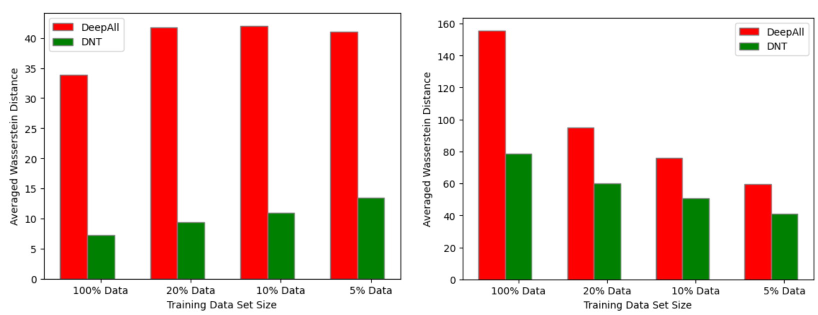

Figure 1: Wasserstein divergence among the source domain representations (left). Wasserstein divergence between the source and target domain representations (right). Models trained on PACS.

1.2. Experimental Settings.

1.2.1. Implementation Details.

We adopted the following experimental setup to match the baselines: learning algorithm: SGD, learning rate (η): 0.001, momentum: 0.9, minibatch size: 64, weight decay: 0.001 for all the experiments. We display the dimension of the latent space $\mathcal{Z}$ in Table 1. We tuned the regularization weight parameter λ over various values ranging from 1 to 10^{-4} at intervals of 10^{-1}. The detailed set up of λ values is provided in Table 2. *It can be noted from Table 2 that as the size of the training data is reduced, our loss is given a higher weight-age for better performance indicating its importance in a limited data setting.* We trained every method for 100 epochs and report the average and standard errors over 5 different seeds. Along with training our models, we also re-trained all baselines for a fair comparison.

Table 1: Dimension of the latent space $\mathcal{Z}$

Dataset	Encoder	Dimension of $\mathcal{Z}$
RotatedMNIST	MNIST CNN	64
PACS	Resnet-18	256
OfficeHome	Resnet-18	512
VLCS	Resnet-18	512
VLCS	Alexnet	512
PACS	Resnet-50	2048
VLCS	Resnet-50	2048

Table 2: Best λ across Datasets and Models

	PACS				VLCS				RotatedMNIST				OfficeHome		
	100%	20%	10%	5%	100%	20%	10%	5%	100%	20%	10%	5%	100%	20%	10%
DNT	1	1	1	1	0.0001	1	1	1	1	1	1	1	1	1	1
DRINT	1	1	1	1	0.1	1	1	1	1	1	1	1	1	1	1
DGNT	0.1	1	1	1	0.0001	0.01	0.1	1	1	1	1	1	0.001	0.1	1

1.3. Experimental Results

1.3.1. Test Accuracy on OfficeHome Dataset

- **OfficeHome** (Venkateswara et al., 2017) dataset consists of $15,500$ images from 65 classes. The four domains are: Art – artistic images in the form of sketches, paintings, ornamentation, etc.; Clipart – collection of clipart images; Product – images of objects without a background and Real-World – images of objects captured with a regular camera.

We provide the mean accuracy and standard deviation the for OfficeHome dataset in Table 3. For the limited data setting, we reduce the training data into 20% and 10%. We did not reduce it to 5% as OfficeHome has 65 classes and there were not enough samples to get at least 2 representative samples from each class through proportional sampling. Results on OfficeHome were not reported by Nguyen et al. (2021) in the DIRT paper and by Zhao et al. (2020) in DGER paper. We have trained the models for our analysis. It can be seen from Table 3 that our methods DNT, DIRT, and DRINT enhance the performance of all the baselines as the previous datasets.

We only compare Mixup with DNT for the full data setting as we were unable to reproduce the results of the original paper Gulrajani and Lopez-Paz (2021) with the code provided for the OfficeHome dataset. We fell short by 6% accuracy in the reproduced version. Hence, we did not train it for a limited data setting and did not adapt it for Manifold Mixup too. Instead, we directly report the results of Mixup on the OfficeHome dataset from the paper.

Table 3: Prediction accuracy % on OfficeHome. Our methods: DNT, DRINT, and DGNT outperform DeepAll, DIRT, and DGER, respectively.

OfficeHome					
Model	E_ϕ	**100%** Acc ± Std Err	**20%** Acc ± Std Err	**10%** Acc ± Std Err	**Average** Acc
DeepAll	Resnet 18	62.48 ± 0.2	53.49 ± 0.3	48.23 ± 0.5	54.73
DNT		**63.25** ± 0.4	**53.72** ± 0.4	**49.00** ± 0.6	**55.32**
DIRT	Resnet 18	62.60 ± 0.2	53.15 ± 0.3	47.78 ± 0.4	54.51
DRINT		**63.81** ± 0.2	**54.49** ± 0.4	**49.82** ± 0.4	**56.04**
DGER	Resnet 18	**64.17** ± 0.1	55.18 ± 0.4	48.38 ± 0.5	55.91
DGNT		64.08 ± 0.1	**55.55** ± 0.4	**50.91** ± 0.5	**56.85**
Mixup	Resnet 50	67.00 ± 0.2			
DNT		**69.68** ± 0.3			

1.3.2. Sensitivity analysis on the hyper-parameter λ

Fig. 2 shows the sensitivity analysis on λ. This analysis shows that maintaining a non-zero lambda is important especially when the dataset size is reduced.

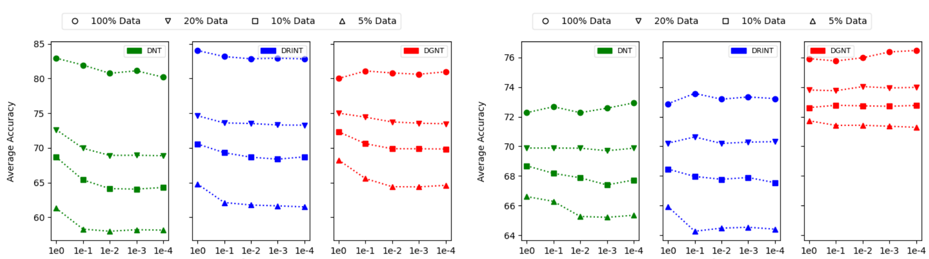

Figure 2: Sensitivity analysis of the hyper-parameter λ on PACS (left) and VLCS (right) datasets. On average, across both datasets, as the training data size was reduced, higher values of λ resulted in better accuracy, emphasizing the importance of robust interpolation in the limited data setting.

1.3.3. Domain wise results

As discussed in the main paper, we present the leave-one-domain-out results of each domain for PACS, VLCS and RotatedMNIST datasets across different sizes in Tables 4, 5 and 6 respectively. One interesting aspect of domain wise results can be observed between DIRT, DNT and DRINT. For example, consider Table 4, DIRT (79.45) performs better than DNT (75.17) in the *sketch* domain, but not as well in the *art painting* domain. DNT (83.18) performs better than DIRT (80.67) in *art painting* , domain, but does not do as well in *sketch* domain. However, DRINT, the combination of both, DNT and DIRT and performs well in both *sketch* (79.85) and *art painting* (82.82) domains.

Table 4: PACS - Domain wise results

Domain	A				C			
Data size:	**100%**	**20%**	**10%**	**5%**	**100%**	**20%**	**10%**	**5%**
DeepAll	78.82 ± 0.76	68.95 ± 0.56	61.68 ± 1.62	54.52 ± 2.26	76.33 ± 0.45	62.98 ± 1.39	59.19 ± 0.98	51.60 ± 1.74
DNT	83.18 ± 0.41	73.92 ± 0.63	68.30 ± 1.38	59.23 ± 2.11	77.63 ± 0.52	67.04 ± 1.16	63.86 ± 1.41	56.90 ± 1.42
DIRT	80.67 ± 0.47	68.69 ± 1.16	64.49 ± 0.80	56.44 ± 2.05	76.26 ± 0.30	67.29 ± 0.45	59.87 ± 0.59	50.69 ± 2.26
DRINT	82.82 ± 0.66	73.64 ± 1.57	68.59 ± 1.11	60.66 ± 0.78	78.13 ± 0.41	69.40 ± 0.85	64.89 ± 1.87	60.39 ± 3.23
DGER	78.83 ± 0.69	71.71 ± 0.75	68.86 ± 0.97	60.31 ± 1.57	71.98 ± 1.75	68.83 ± 1.52	62.86 ± 1.61	46.02 ± 3.45
DGNT	79.34 ± 0.40	73.85 ± 1.24	72.77 ± 0.66	64.67 ± 1.59	75.53 ± 0.47	70.28 ± 1.19	68.04 ± 1.26	64.66 ± 1.17
Domain	P				S			
Data size:	**100%**	**20%**	**10%**	**5%**	**100%**	**20%**	**10%**	**5%**
DeepAll	95.92 ± 0.22	90.79 ± 0.35	87.61 ± 0.63	80.15 ± 1.92	69.93 ± 0.65	53.94 ± 0.93	50.59 ± 2.18	47.86 ± 3.13
DNT	95.73 ± 0.15	92.19 ± 0.28	89.66 ± 0.07	84.75 ± 0.64	75.17 ± 0.70	57.31 ± 2.77	52.82 ± 2.46	44.29 ± 2.94
DIRT	94.85 ± 0.12	90.83 ± 0.34	87.40 ± 1.27	81.06 ± 1.56	79.45 ± 0.49	65.26 ± 0.88	54.57 ± 3.70	51.73 ± 4.49
DRINT	95.31 ± 0.04	91.85 ± 0.22	90.20 ± 0.37	86.05 ± 1.28	79.85 ± 0.46	63.63 ± 1.27	58.56 ± 1.79	52.10 ± 2.68
DGER	93.74 ± 0.63	93.16 ± 0.26	91.71 ± 0.31	89.16 ± 0.95	70.59 ± 1.76	61.49 ± 1.72	55.76 ± 2.02	48.60 ± 2.05
DGNT	96.50 ± 0.12	92.56 ± 0.24	91.48 ± 0.20	89.97 ± 0.68	72.94 ± 0.75	63.33 ± 1.37	56.94 ± 1.84	53.67 ± 2.56

Table 5: VLCS - Domain wise results

Domain	C				L			
Data size:	**100%**	**20%**	**10%**	**5%**	**100%**	**20%**	**10%**	**5%**
DeepAll	95.12 ± 0.48	92.48 ± 0.21	87.25 ± 1.30	80.42 ± 2.13	56.86 ± 0.75	53.67 ± 0.54	56.16 ± 1.84	52.75 ± 0.46
DNT	95.58 ± 0.28	91.63 ± 1.18	88.67 ± 0.94	79.88 ± 2.38	59.40 ± 0.50	57.39 ± 2.67	57.84 ± 1.60	54.89 ± 1.99
DIRT	94.96 ± 0.43	92.15 ± 0.57	87.32 ± 2.58	77.66 ± 1.81	60.09 ± 0.66	55.96 ± 1.33	57.45 ± 1.93	55.10 ± 1.13
DRINT	95.69 ± 0.24	92.84 ± 0.44	90.46 ± 2.10	81.56 ± 2.24	61.30 ± 0.72	56.14 ± 0.48	54.27 ± 1.07	57.64 ± 1.05
DGER	96.89 ± 0.40	95.74 ± 0.31	93.10 ± 0.75	92.84 ± 0.99	62.68 ± 0.08	60.43 ± 0.44	61.04 ± 0.62	59.29 ± 1.41
DGNT	96.93 ± 0.45	95.93 ± 0.51	93.89 ± 0.52	92.87 ± 1.53	63.30 ± 0.36	60.48 ± 0.63	60.82 ± 0.90	60.46 ± 0.98
Domain	S				V			
Data size:	**100%**	**20%**	**10%**	**5%**	**100%**	**20%**	**10%**	**5%**
DeepAll	66.05 ± 0.96	63.07 ± 2.74	64.06 ± 1.21	61.03 ± 1.38	68.23 ± 0.99	65.62 ± 0.18	63.77 ± 0.92	58.32 ± 0.62
DNT	67.39 ± 0.41	64.39 ± 0.21	64.15 ± 1.15	59.69 ± 0.70	69.36 ± 0.36	66.11 ± 0.87	64.07 ± 0.93	59.93 ± 1.08
DIRT	67.79 ± 0.59	65.13 ± 1.87	64.83 ± 1.57	61.32 ± 1.79	69.62 ± 0.23	66.39 ± 0.29	63.14 ± 0.56	60.08 ± 0.61
DRINT	68.05 ± 0.41	66.33 ± 1.60	64.85 ± 1.85	62.36 ± 0.93	69.26 ± 0.55	65.52 ± 0.52	64.27 ± 0.62	62.17 ± 0.81
DGER	69.80 ± 0.65	68.16 ± 0.51	67.35 ± 0.65	65.42 ± 0.77	74.14 ± 0.48	71.01 ± 0.63	69.40 ± 0.76	67.39 ± 0.68
DGNT	70.53 ± 0.44	68.60 ± 0.43	67.18 ± 0.54	65.34 ± 0.75	75.09 ± 0.43	71.12 ± 0.42	69.17 ± 0.81	68.21 ± 0.94

Table 6: RotatedMNIST - Domain wise results

Domain		0°			15°	
Data size:	**100%**	**20%**	**10%**	**100%**	**20%**	**10%**
DeepAll	82.96 ± 0.99	65.10 ± 1.64	57.80 ± 0.72	98.40 ± 0.19	86.54 ± 0.37	79.46 ± 0.84
DNT	93.60 ± 0.25	71.32 ± 0.57	64.82 ± 1.41	99.46 ± 0.14	87.00 ± 1.68	83.02 ± 0.41
DIRT	97.34 ± 0.12	69.32 ± 0.92	60.56 ± 0.64	99.82 ± 0.06	90.56 ± 0.38	82.84 ± 0.33
DRINT	97.46 ± 0.13	72.54 ± 0.51	63.94 ± 1.25	99.84 ± 0.05	91.08 ± 0.33	84.10 ± 0.52
DGER	88.78 ± 0.25	63.42 ± 0.67	56.56 ± 0.69	98.48 ± 0.09	85.50 ± 0.51	79.80 ± 0.64
DGNT	90.88 ± 0.35	71.14 ± 0.41	61.68 ± 0.89	98.86 ± 0.06	88.62 ± 0.49	82.66 ± 0.35
Domain		30°			45°	
Data size:	**100%**	**20%**	**10%**	**100%**	**20%**	**10%**
DeepAll	98.32 ± 0.15	87.64 ± 0.31	83.00 ± 0.87	98.12 ± 0.11	87.32 ± 0.18	82.38 ± 0.81
DNT	99.30 ± 0.08	90.94 ± 0.38	86.84 ± 0.41	99.20 ± 0.03	91.06 ± 0.22	86.74 ± 0.54
DIRT	99.54 ± 0.07	91.10 ± 0.42	85.18 ± 0.99	99.48 ± 0.07	90.86 ± 0.09	86.80 ± 0.72
DRINT	99.62 ± 0.05	91.96 ± 0.31	86.94 ± 0.36	99.66 ± 0.07	91.70 ± 0.18	87.72 ± 0.31
DGER	98.10 ± 0.12	87.40 ± 0.24	82.54 ± 0.18	98.12 ± 0.12	87.62 ± 0.77	83.12 ± 0.37
DGNT	98.84 ± 0.06	90.50 ± 0.45	85.78 ± 0.18	98.66 ± 0.14	90.00 ± 0.59	86.06 ± 0.31
Domain		60°			75°	
Data size:	**100%**	**20%**	**10%**	**100%**	**20%**	**10%**
DeepAll	98.42 ± 0.11	86.10 ± 0.39	80.02 ± 0.85	85.94 ± 0.24	72.12 ± 0.73	61.04 ± 0.56
DNT	99.64 ± 0.08	90.56 ± 0.47	84.40 ± 0.24	92.96 ± 0.39	75.98 ± 0.46	67.52 ± 0.45
DIRT	99.70 ± 0.08	90.60 ± 0.32	83.78 ± 0.79	96.64 ± 0.45	77.28 ± 0.32	64.56 ± 1.17
DRINT	99.78 ± 0.09	91.16 ± 0.38	84.72 ± 0.20	96.64 ± 0.32	78.44 ± 0.41	68.74 ± 0.60
DGER	98.08 ± 0.07	85.86 ± 0.29	80.78 ± 0.42	86.08 ± 0.28	69.54 ± 0.45	59.34 ± 0.39
DGNT	98.66 ± 0.10	88.74 ± 0.39	83.20 ± 0.29	89.60 ± 0.23	74.10 ± 0.47	64.24 ± 0.37

1.3.4. Accuracy on test domains with linear interpolation.

We perform linear interpolation for PACS, VLCS and RotatedMNIST in Tables 7, 8 and 9. The results are better than baselines, but slightly lower than the nonlinear version. In the case of VLCS, linear interpolation performs slightly better as the dataset is more scene centric. Hence, we assume that the domains are the linear interpolations of each other more than that of the object centric datasets.

Table 7: Prediction accuracy % on PACS.

PACS									
	100%		**20%**		**10%**		**5%**		**Average**
Model	Acc ± Std Err	Gain %	Acc ± Std Err	Gain %	Acc ± Std Err	Gain %	Acc ± Std Err	Gain %	
DeepAll	80.25 ± 0.52	0.00	69.17 ± 0.81	0.00	64.77 ± 1.35	0.00	58.53 ± 2.26	0.00	68.18
DNT Linear	81.99 ± 0.49	1.74	70.57 ± 0.82	1.41	66.71 ± 1.29	1.94	61.12 ± 2.25	2.59	70.10
DNT	**82.92** ± 0.44	2.68	**72.62** ± 1.21	3.45	**68.66** ± 1.33	3.89	**61.29** ± 1.78	2.76	**71.37**
DIRT	82.81 ± 0.34	0.00	73.02 ± 0.71	0.00	66.58 ± 1.59	0.00	59.98 ± 2.59	0.00	70.60
DRINT Linear	83.52 ± 0.45	0.71	74.75 ± 0.79	1.73	69.54 ± 1.47	2.96	63.96 ± 2.28	3.98	72.94
DRINT	**84.03** ± 0.39	1.22	**74.63** ± 0.98	1.61	**70.56** ± 1.28	3.98	**64.80** ± 1.99	4.82	**73.50**
DGER	80.85 ± 0.43	0.00	73.80 ± 1.06	0.00	69.79 ± 1.23	0.00	65.00 ± 1.54	0.00	72.36
DGNT Linear	80.91 ± 0.44	0.06	74.42 ± 0.97	0.62	70.62 ± 1.36	0.82	66.01 ± 1.41	1.01	72.99
DGNT	**81.08** ± 0.39	0.23	**75.00** ± 1.01	1.21	**72.31** ± 0.99	2.51	**68.24** ± 1.50	3.24	**74.16**

Table 8: Prediction accuracy % on VLCS.

VLCS									
	100%		**20%**		**10%**		**5%**		**Average**
Model	Acc ± Std Err	Gain %	Acc ± Std Err	Gain %	Acc ± Std Err	Gain %	Acc ± Std Err	Gain %	
DeepAll	71.56 ± 0.80	0.00	68.71 ± 0.92	0.00	67.81 ± 1.32	0.00	63.13 ± 1.15	0.00	67.80
DNT Linear	72.60 ± 0.52	1.04	**70.10** ± 0.69	1.38	67.98 ± 0.76	0.17	**65.66** ± 1.39	2.53	**69.09**
DNT	**72.93** ± 0.39	1.37	69.88 ± 1.23	1.17	68.68 ± 1.15	0.87	63.60 ± 1.54	0.47	68.77
DIRT	73.11 ± 0.48	0.00	69.91 ± 1.01	0.00	68.19 ± 1.66	0.00	63.54 ± 1.33	0.00	68.69
DRINT Linear	73.53 ± 0.38	0.41	**70.49** ± 0.49	0.58	67.54 ± 0.52	-0.65	**66.66** ± 1.20	3.12	**69.55**
DRINT	**73.57** ± 0.48	0.46	70.21 ± 0.76	0.30	**68.46** ± 1.41	0.28	65.93 ± 1.26	2.39	69.54
DGER	75.87 ± 0.40	0.00	73.84 ± 0.47	0.00	72.72 ± 0.70	0.00	71.24 ± 0.96	0.00	73.42
DGNT Linear	**76.59** ± 0.49	0.72	**74.20** ± 0.65	0.36	72.77 ± 0.74	0.05	71.67 ± 0.96	0.43	**73.81**
DGNT	76.47 ± 0.42	0.60	74.03 ± 0.50	0.19	**72.77** ± 0.69	0.05	**71.72** ± 1.05	0.48	73.74

Table 9: Prediction accuracy % on RotatedMNIST.

RotatedMNIST							
	100%		**20%**		**10%**		**Average**
Model	Acc ± Std Err	Gain %	Acc ± Std Err	Gain %	Acc ± Std Err	Gain %	
DeepAll	92.69 ± 0.30	0.00	80.80 ± 0.60	0.00	73.95 ± 0.78	0.00	82.48
DNT Linear	94.52 ± 0.22	1.83	81.40 ± 0.70	0.60	75.69 ± 0.66	1.74	83.87
DNT	**97.36** ± 0.16	4.67	**84.48** ± 0.63	3.67	**78.89** ± 0.58	4.94	**86.91**
DIRT	98.75 ± 0.14	0.00	84.95 ± 0.41	0.00	77.29 ± 0.77	0.00	87.00
DRINT Linear	98.88 ± 0.09	0.12	84.69 ± 0.57	-0.26	77.42 ± 0.70	0.13	86.99
DRINT	**98.83** ± 0.12	0.08	**86.15** ± 0.35	1.19	**79.36** ± 0.54	2.07	**88.11**
DGER	95.61 ± 0.15	0.00	79.89 ± 0.49	0.00	73.69 ± 0.45	0.00	83.06
DGNT Linear	95.52 ± 0.16	-0.09	80.38 ± 0.62	0.49	74.77 ± 0.57	1.08	83.55
DGNT	**95.92** ± 0.16	1.31	**83.85** ± 0.47	3.96	**77.27** ± 0.40	3.58	**85.68**

References

Isabela Albuquerque, João Monteiro, Mohammad Darvishi, Tiago H Falk, and Ioannis Mitliagkas. Generalizing to unseen domains via distribution matching. *arXiv preprint arXiv:1911.00804*.

Zhen Cui, Wen Li, Dong Xu, Shiguang Shan, Xilin Chen, and Xuelong Li. Flowing on riemannian manifold: Domain adaptation by shifting covariance. *IEEE Transactions on Cybernetics*, 44(12):2264–2273, 2014. doi: 10.1109/TCYB.2014.2305701.

Raghuraman Gopalan, Ruonan Li, and Rama Chellappa. Domain adaptation for object recognition: An unsupervised approach. In *2011 International Conference on Computer Vision*, pages 999–1006, 2011. doi: 10.1109/ICCV.2011.6126344.

Ishaan Gulrajani and David Lopez-Paz. In search of lost domain generalization. In *International Conference on Learning Representations*, 2021.

A. Tuan Nguyen, Toan Tran, Yarin Gal, and Atilim Gunes Baydin. Domain invariant representation learning with domain density transformations. In *Advances in Neural Information Processing Systems*, volume 34, pages 5264–5275, 2021.

Thai-Hoang Pham, Xueru Zhang, and Ping Zhang. Fairness and accuracy under domain generalization. In *The Eleventh International Conference on Learning Representations*, 2023.

Hemanth Venkateswara, Jose Eusebio, Shayok Chakraborty, and Sethuraman Panchanathan. Deep hashing network for unsupervised domain adaptation, 2017.

Shanshan Zhao, Mingming Gong, Tongliang Liu, Huan Fu, and Dacheng Tao. Domain generalization via entropy regularization. *Advances in Neural Information Processing Systems*, 33:16096–16107, 2020.