Domain Generalizing DINO for Visual Regression via Latent Distractor Subspace Consistency

Nikhil Reddy[1]   Chetan Arora[2]   Mahsa Baktashmotlagh[3]

[1]UQ-IIT Delhi Research Academy (UQIDRA)   [2]IIT Delhi, India   [3]UQ, Australia

Hyper Parameters. We implement our framework in PyTorch [3] and select hyperparameters through cross-validation. We leverage SisPCA [4] to extract the orthogonal subspaces in our method, using a kernel bandwidth of $\sigma = 80$. Table 1 reports the hyperparameters for the `PovertyMap`, `RCF-MNIST`, and `dSprites` datasets.

Dataset	`PovertyMap`	`RCF-MNIST`	`dSprites`
Learning rate	1×10^{-4}	5×10^{-4}	1×10^{-3}
Weight decay	1×10^{-5}	5×10^{-5}	5×10^{-5}
Scheduler	Cosine decay	StepLR	Cosine decay
Batch size	64	128	256

Table 1. Hyperparameters used for different datasets in the `LDSC` framework.

Ablation on Lambda. To demonstrate the impact of the learning rate and λ_{LDSC} on Average RMSE performance, we consider the `dSprites` dataset [2] and vary λ_{LDSC} values while training the `LDSC` framework. Results are reported below.

Method	λ_{ldsc}	Average RMSE
• **LDSC (ours)**	1×10^{-3}	1.239
• **LDSC (ours)**	1×10^{-1}	**0.263**
• **LDSC (ours)**	1	1.122

Table 2. Ablation on λ_{ldsc} values for the `dSprites` dataset. A moderate weighting ($\lambda_{\text{ldsc}} = 10^{-1}$) achieves the best performance.

Ablation study on μ. controlling the HSIC regularization strength in the `LDSC` framework. We consider the `dSprites` dataset[2] to report its results. Results are reported in Tab. 3 **Ablation study on μ.** controlling the HSIC regularization strength in the `LDSC` framework. We consider the `PovertyMap` dataset [5] to report its results. Results are reported in Tab. 3

Ablation Study on d_1, d_2. We conduct an ablation study to analyze the impact of different choices of d_1 and d_2 on the performance of the `LDSC` framework. Using the `SkillCraft` dataset [1], we evaluate the framework with

Method	μ	Average RMSE
• **LDSC (ours)**	0.1	1.301
• **LDSC (ours)**	0.4	0.774
• **LDSC (ours)**	0.5	1.847
• **LDSC (ours)**	1	1.772

Table 3. Ablation study on μ, controlling the HSIC regularization strength in the `LDSC` framework. We use the `PovertyMap` dataset.

various values of d_1 and d_2, measuring the average RMSE for each setting. The results, presented in Tab. 4, highlight the sensitivity of the framework to these dimensions.

Ablation Study on d_1, d_2. We conduct an ablation study to analyze the impact of different choices of d_1 and d_2 on the performance of the `LDSC` framework. Using the `PovertyMap` dataset [5], we evaluate the framework with various values of d_1 and d_2, measuring the average RMSE for each setting. The results, presented in Tab. 4, highlight the sensitivity of the framework to these dimensions.

Method	d_1	d_2	Average RMSE
• **LDSC (ours)**	5	10	1.445
• **LDSC (ours)**	2	5	1.479
• **LDSC (ours)**	5	5	**0.774**
• **LDSC (ours)**	10	5	1.645

Table 4. Ablation study on d_1 and d_2 in the `LDSC` framework using the `PovertyMap` dataset.

Ablation study on learning rate. We perform an ablation study on the impact of the learning rate on the `LDSC` framework. We consider the `PovertyMap` dataset [5] and train the `LDSC` framework with different learning rates. Results are reported in Tab. 5.

Ablation Study - Learning rate . We analyze the impact of the learning rate on the loss function formulation. Results are reported in Fig. 1. For this experiment, we analyze the impact of the learning rate using the `RCF-MNIST` dataset and its effect on the test mean square error. We observe that

Method	Learning rate	Average RMSE
• **LDSC (ours)**	$1e-5$	1.312
• **LDSC (ours)**	$1e-4$	1.986
• **LDSC (ours)**	$1e-3$	**0.774**

rule[1.5pt]

Table 5. Ablation study on learning rate used in the loss formulation of the `LDSC` framework with the `PovertyMap` dataset.
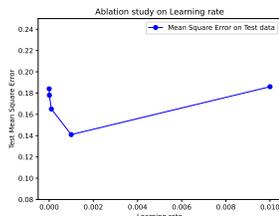

Figure 1. Impact of the learning rate on the test mean square error. We consider the `RCF-MNIST` dataset and experiment settinng.

$1 \times e^{-3}$ is the optimal learning rate for our experiments.

Ablation Study on LoRA Rank. We further conduct an ablation study to analyze the impact of the LoRA rank on the performance of the `LDSC` framework. We vary the rank parameter r while keeping all other hyperparameters fixed, and evaluate on the `PovertyMap` [5] dataset. As shown in Tab. 6, a moderate rank ($r = 8$) achieves the best trade-off between model capacity and generalization, while both very small and very large ranks lead to degraded performance due to under- and over-parameterization, respectively.

Method	LoRA Rank r	Average RMSE
• **LDSC (ours)**	2	1.428
• **LDSC (ours)**	4	0.921
• **LDSC (ours)**	8	**0.774**
• **LDSC (ours)**	16	1.107
• **LDSC (ours)**	32	1.589

Table 6. Ablation study on the LoRA rank in the `LDSC` framework using the `PovertyMap` dataset. A moderate rank ($r = 8$) achieves the best performance.

References

[1] Blair, Thompson Mark, Henrey Joe, Chen Andrew, and Bill. SkillCraft1 Master Table Dataset. UCI Machine Learning Repository, 2013. DOI: https://doi.org/10.24432/C5161N.

[2] Loic Matthey, Irina Higgins, Demis Hassabis, and Alexander Lerchner. dsprites: Disentanglement testing sprites dataset. `https://github.com/google-deepmind/dsprites-dataset`, 2017. DeepMind, accessed 2025-09-10.

[3] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32:8026–8037, 2019.

[4] Jiayu Su, David A Knowles, and Raul Rabadan. Disentangling interpretable factors with supervised independent subspace principal component analysis. *Advances in Neural Information Processing Systems*, 37:37408–37438, 2024.

[5] Christopher Yeh, Anthony Perez, Anne Driscoll, George Azzari, Zhongyi Tang, David Lobell, Stefano Ermon, and Marshall Burke. Using publicly available satellite imagery and deep learning to understand economic well-being in africa. *Nature communications*, 11(1):2583, 2020.