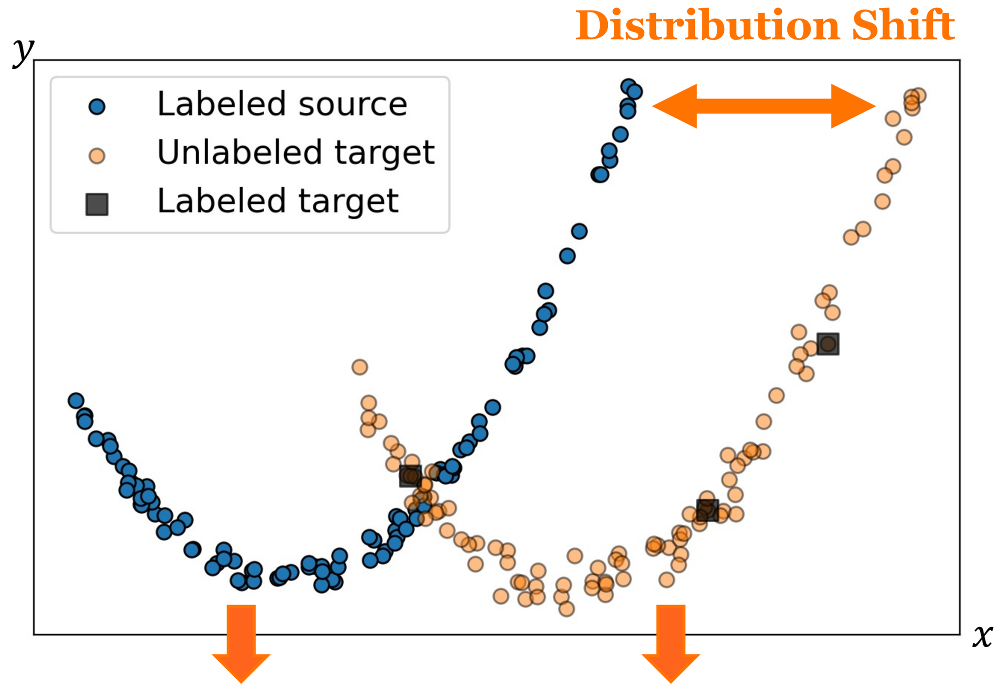
Distribution-Informed Neural Networks for Domain Adaptation Regression

Jun Wu, Jingrui He, Sheng Wang, Kaiyu Guan, Elizabeth Ainsworth
University of Illinois at Urbana-Champaign
junwu3@illinois.edu, jingrui@illinois.edu, sheng12@illinois.edu, kaiyug@illinois.edu, ainswort@illinois.edu

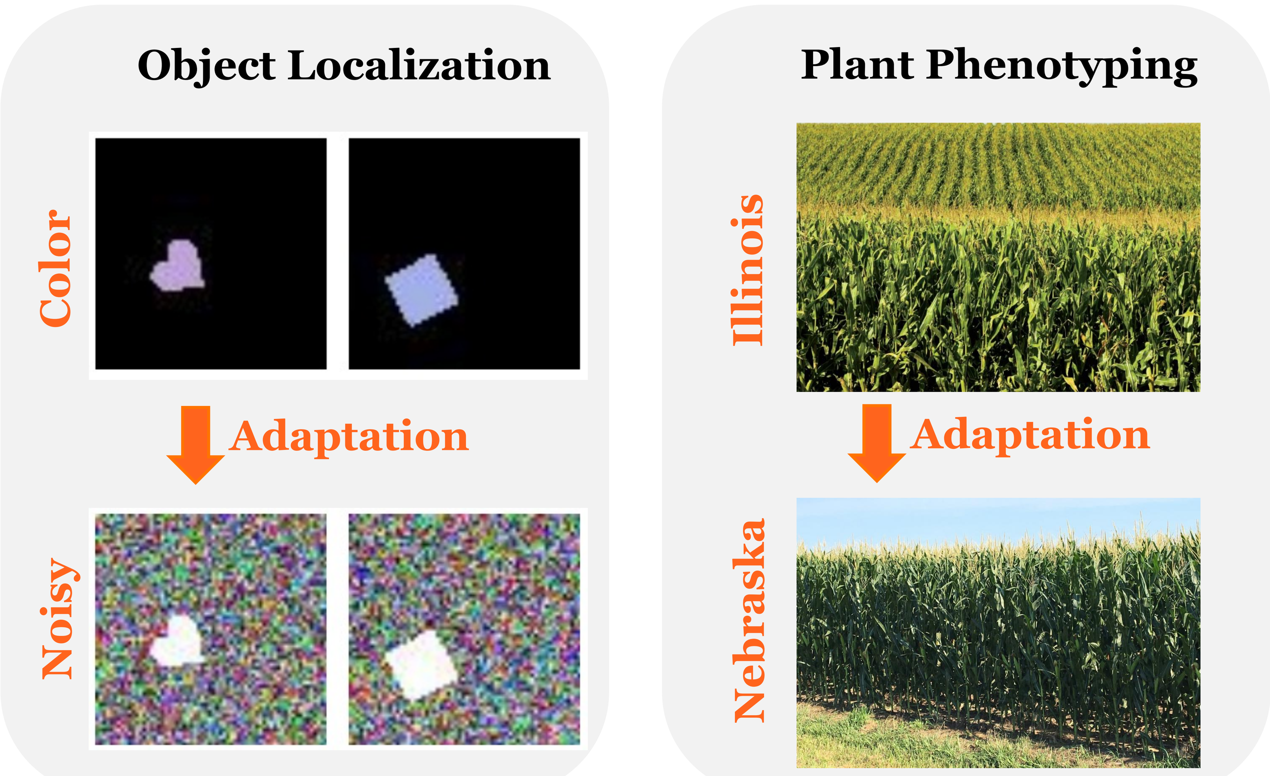
AIFARMS Artificial Intelligence for Future Agricultural Resilience, Management, and Sustainability

Background

❑ Domain adaptation regression



❑ Applications

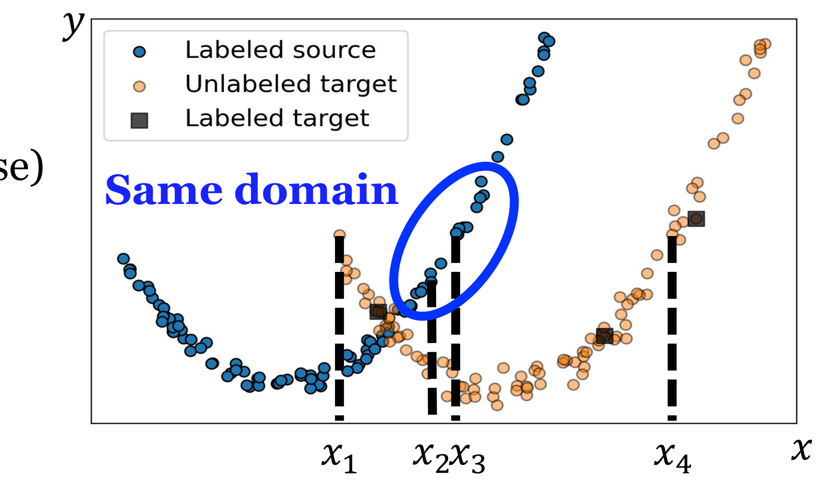


Distribution-Informed Neural Network (DINO)

❑ Motivation

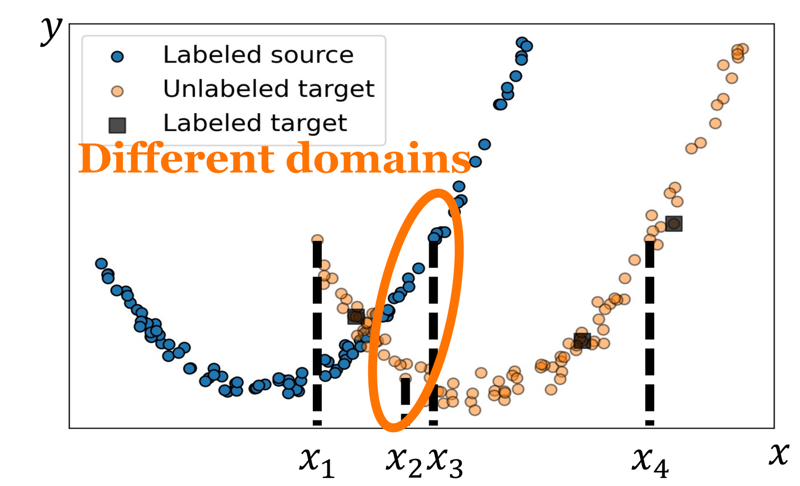
$f(x_2) \approx f(x_3)$ if $x_2 \approx x_3$ (Homogeneous Case)

➢ x_2 and x_3 from source domain



$y_2 \neq y_3$ if $x_2 \approx x_3$ (Heterogeneous Case)

➢ x_2 from target domain
➢ x_3 from source domain



❑ Definition

DINO: $$\tilde{f}(x, \mathbb{P}) \coloneqq f_\theta(x) \cdot g_{w_g}(\mathbb{P}|x) = \left(\phi_{\theta^{<L}}(x)^T w\right) \cdot \left(\Phi_x(\mathbb{P})^T w_g\right) = w^T\left(\phi_{\theta^{<L}}(x)\Phi_x(\mathbb{P})^T\right) w_g$$

Input representation learning

A fully-connected NN: $f_\theta(x) = \phi_{\theta^{<L}}(x)^T w$

($\theta^{<L}$: Parameters of the first $L-1$ layers; w: Parameters of the output layer)

Input-oriented distribution representation learning

Infinitely-wide $f_\theta(\cdot)$ ➡ NNGP kernel space K_X

➡ $\Phi_x(\mathbb{P}) = \sum_{i=1}^{n} \beta_{x,\tilde{x}_i} \langle \cdot, \tilde{x}_i \rangle_{K_X}$ ➡ $g_{w_g}(\mathbb{P}|x) = \Phi_x(\mathbb{P})^T w_g$

Algorithms

❑ DINO-INIT

➢ At initialization, DINO is a Gaussian process with adaptive NNGP kernel

Under random initialization, we have $\tilde{f}(\cdot) \sim \mathcal{N}(0, K^{DA})$ with

$$K^{DA}\left((x,\mathbb{P}),(x',\mathbb{P}')\right) = K_X(x,x') \cdot K_{\mathcal{P}|X}(\mathbb{P},\mathbb{P}'|x,x')$$

where $K_X(\cdot,\cdot)$ is the NNGP kernel, and $K_{\mathcal{P}|X}(\cdot,\cdot)$ is a distribution kernel, i.e.,

$$K_{\mathcal{P}|X}(\mathbb{P},\mathbb{P}'|x,x') = \sum_{i=1}^{n}\sum_{j=1}^{n'} \beta_{x,\tilde{x}_i}\beta_{x',\tilde{x}_j} K_X(\tilde{x}_i,\tilde{x}_j)$$

➢ Adaptive Gaussian process
- Prior GP $\tilde{f}(\cdot) \sim \mathcal{N}(0, K^{DA})$
- Prediction function $p(Y|X_*^{tgt}) = \mathcal{N}(\bar{\mu}, \bar{\Sigma})$

$$\bar{\mu} = K^{DA}(X_*^{tgt}, X)C^{-1}Y \qquad \bar{\Sigma} = K^{DA}(X_*^{tgt}, X_*^{tgt}) - K^{DA}(X_*^{tgt}, X)C^{-1}K^{DA}(X_*^{tgt}, X)^T$$

❑ DINO-TRAIN

➢ Gradient descent training with the following objective function

$$\mathcal{L}(\theta) = \underbrace{\frac{\alpha}{2n_{src}}\sum_{i=1}^{n_{src}}\left(\tilde{f}(x_i^{src},\mathbb{P}^{src}) - y_i^{src}\right)^2 + \frac{1-\alpha}{2n_{tgt}^l}\sum_{j=1}^{n_{tgt}^l}\left(\tilde{f}(x_j^{tgt},\mathbb{P}^{tgt}) - y_j^{tgt}\right)^2}_{\text{Supervised loss over labeled examples}} + \underbrace{\frac{\mu}{2}\mathrm{MMD}^2_{\Theta_{DA}}(\mathbb{P}^{src},\mathbb{P}^{tgt})}_{\text{Empirical MMD-NTK}}$$

➢ Empirical Maximum Mean Discrepancy (MMD) over training dynamics

$$\mathrm{MMD}^2_{\Theta_{DA}}(\mathbb{P}^{src},\mathbb{P}^{tgt}) = \left\| \frac{1}{n_{src}}\sum_{i=1}^{n_{src}} \nabla_\theta \tilde{f}(x_i^{src},\mathbb{P}^{src}) - \frac{1}{n_{tgt}}\sum_{j=1}^{n_{tgt}} \nabla_\theta \tilde{f}(x_j^{tgt},\mathbb{P}^{tgt}) \right\|^2_{\mathcal{H}_{DA}}$$

Experiments

Methods	RL → RC	RL → T	RC → RL	RC → T	T → RL	T → RC	Avg.
NNGP [34]	$0.313_{\pm 0.001}$	$0.438_{\pm 0.004}$	$0.356_{\pm 0.005}$	$0.515_{\pm 0.008}$	$0.367_{\pm 0.001}$	$0.324_{\pm 0.004}$	0.386
NTKGP [25]	$0.396_{\pm 0.001}$	$0.365_{\pm 0.001}$	$0.200_{\pm 0.007}$	$0.390_{\pm 0.003}$	$0.390_{\pm 0.000}$	$0.354_{\pm 0.003}$	0.349
AT-GP [7]	$0.214_{\pm 0.011}$	$0.209_{\pm 0.002}$	$0.227_{\pm 0.010}$	$0.198_{\pm 0.002}$	$0.236_{\pm 0.000}$	$0.249_{\pm 0.000}$	0.222
TL-NTK [38]	$0.206_{\pm 0.004}$	$0.200_{\pm 0.002}$	$0.213_{\pm 0.000}$	$0.197_{\pm 0.000}$	$0.226_{\pm 0.001}$	$0.218_{\pm 0.000}$	0.210
DINO-INIT (ours)	$0.204_{\pm 0.001}$	**$0.185_{\pm 0.006}$**	**$0.207_{\pm 0.003}$**	**$0.182_{\pm 0.004}$**	**$0.218_{\pm 0.001}$**	**$0.212_{\pm 0.001}$**	**0.201**
DINO-TRAIN (ours)	**$0.193_{\pm 0.001}$**	$0.194_{\pm 0.003}$	**$0.207_{\pm 0.003}$**	$0.188_{\pm 0.002}$	$0.226_{\pm 0.001}$	$0.218_{\pm 0.001}$	0.204

Results on dSprites

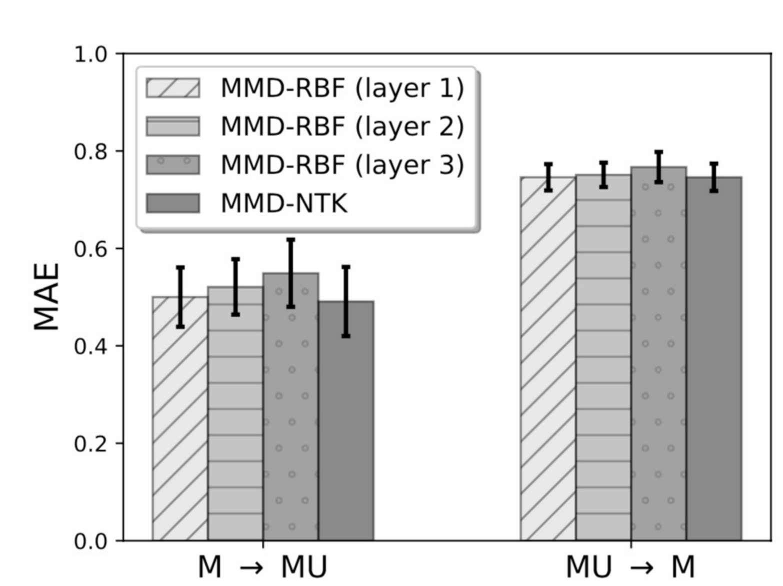
Methods	C → N	C → S	N → C	N → S	S → C	S → N	Avg.
NNGP [34]	$2.041_{\pm 0.001}$	$1.823_{\pm 0.001}$	$0.445_{\pm 0.002}$	$0.624_{\pm 0.001}$	$0.197_{\pm 0.002}$	$0.459_{\pm 0.002}$	0.932
NTKGP [25]	$1.345_{\pm 0.002}$	$1.227_{\pm 0.000}$	$0.323_{\pm 0.002}$	$0.529_{\pm 0.004}$	$0.248_{\pm 0.001}$	$0.425_{\pm 0.002}$	0.683
AT-GP [7]	$0.194_{\pm 0.005}$	$0.259_{\pm 0.002}$	**$0.104_{\pm 0.001}$**	$0.252_{\pm 0.005}$	$0.118_{\pm 0.003}$	$0.189_{\pm 0.006}$	0.186
TL-NTK [38]	$0.164_{\pm 0.001}$	**$0.231_{\pm 0.000}$**	$0.124_{\pm 0.005}$	$0.242_{\pm 0.002}$	$0.125_{\pm 0.001}$	$0.197_{\pm 0.004}$	0.181
DINO-INIT (ours)	$0.128_{\pm 0.001}$	$0.233_{\pm 0.003}$	$0.114_{\pm 0.002}$	**$0.227_{\pm 0.002}$**	**$0.112_{\pm 0.001}$**	**$0.181_{\pm 0.005}$**	**0.166**
DINO-TRAIN (ours)	**$0.127_{\pm 0.002}$**	$0.240_{\pm 0.003}$	$0.127_{\pm 0.000}$	$0.243_{\pm 0.000}$	$0.128_{\pm 0.001}$	$0.194_{\pm 0.001}$	0.177

Results on MPI3D

Lower is better

Methods	M → MU	MU → M
NNGP [34]	$0.562_{\pm 0.001}$	$0.672_{\pm 0.010}$
NTKGP [25]	$0.562_{\pm 0.004}$	$0.702_{\pm 0.010}$
AT-GP [7]	**$0.308_{\pm 0.006}$**	$0.593_{\pm 0.025}$
TL-NTK [38]	$0.316_{\pm 0.008}$	$0.488_{\pm 0.027}$
DINO-INIT (ours)	$0.316_{\pm 0.007}$	$0.645_{\pm 0.017}$
DINO-TRAIN (ours)	$0.314_{\pm 0.009}$	**$0.443_{\pm 0.030}$**

Results on Plant Phenotyping



Conclusion

❑ **Problem**: We study the domain adaptation regression problem in term of convergence and generalization when using deep neural networks.
❑ **Algorithm**: Distribution-informed neural network (DINO) is proposed for learning domain heterogeneity, followed by two instantiated algorithms based on random initialization and gradient descent training.
❑ **Evaluation**: The efficacy of the proposed algorithms is verified on several domain adaptation regression tasks.

Acknowledgments

This work is supported by National Science Foundation under Award No. IIS-1947203, IIS-2117902, IIS-2137468, and Agriculture and Food Research Initiative (AFRI) grant no. 2020-67021-32799/ project accession no.1024178 from the USDA National Institute of Food and Agriculture. The views and conclusions are those of the authors and should not be interpreted as representing the official policies of the funding agencies or the government.