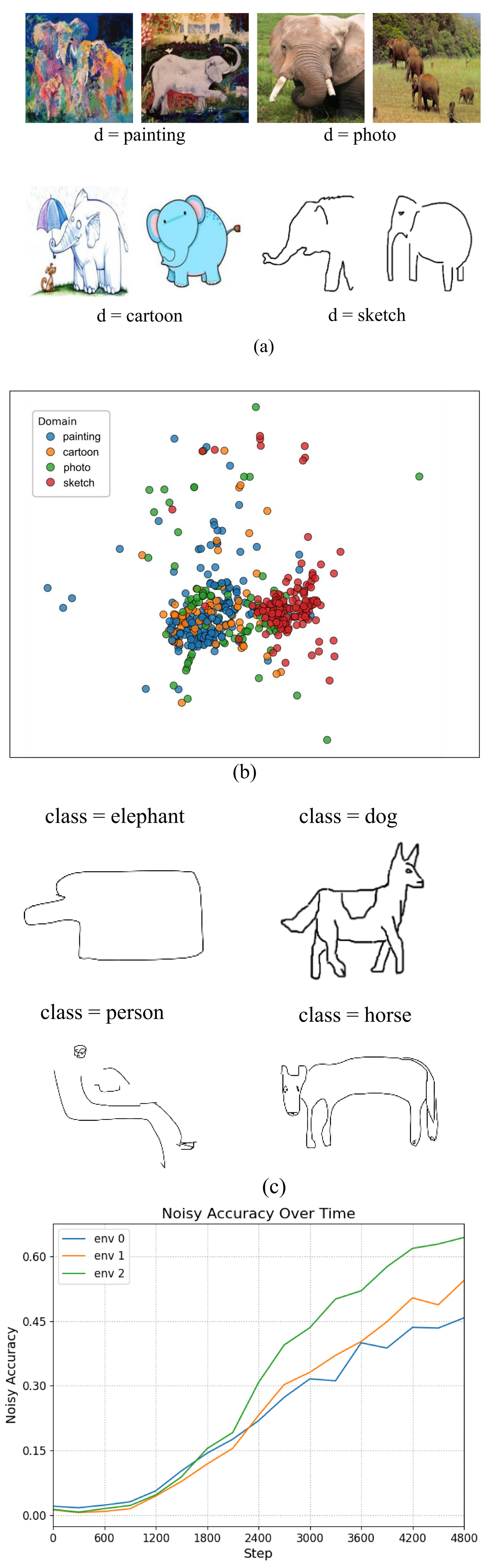
A^{3}W: Language Anchor-Guided Method for Robust Noisy Domain Generalization

Authors: Zilin Dai, Lehong Wang, Fangzhou Lin, Yidong Wang, Zhigang Li, Kazunori D Yamada, Ziming Zhang, Wang Lu*

Advisor: Ziming Zhang

Introduction



Problem & Challenge

- ***Domain Shift***: Training and test distributions differ, leading to performance drops of ML models.
- ***Intra-Class Variability***: Changes in lighting and backgrounds challenge invariant feature extraction.
- ***Spurious Correlations***: Models can latch onto domain-specific cues.
- ***Label Noise***: High mislabeling (8–38.5%) in real-world datasets further degrades performance.
- ***Current DG Limitations***: Struggle with noisy data and weak domain-invariant signals.

Key Contributions

- Developed an **iterative update algorithm** that ***aligns semantic and image features*** for robust generalization under noise.
- Introduced **NLP anchors** from large-scale language model for ***domain-invariant constraints***.
- Proposed a **weighted loss** that adjusts sample's contribution by its NLP ***anchor similarity***.
- Demonstrated **superior performance** on multiple domain generalization benchmarks with ***injected noise***.

Methodology

NLP Anchor Setting

- Computing semantic anchors using ***CLIP*** as text encoder. The anchors are then stacked into an anchor matrix and are used to initialize a set of ***linear projectors*** that map image features into the same semantic space.

Algorithm 1 Training Outline for A^3W

Require: Dataset $\mathcal{D}$ with classes $\{1,\ldots,C\}$, hyperparameters $\lambda,\tau,\ldots$

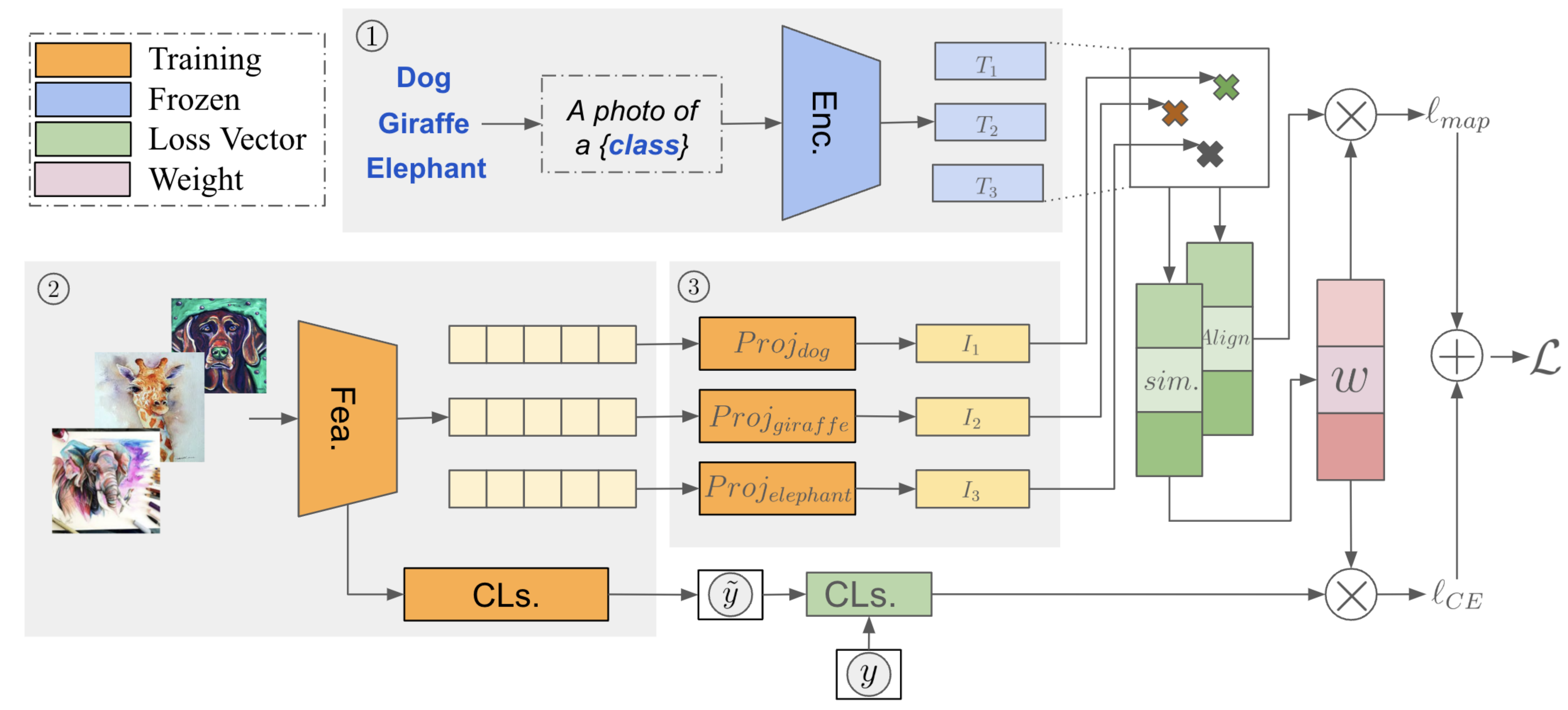
Initialize: featurizer $f(\cdot)$, classifier $g(\cdot)$, empty mapping layers $\{Proj_1,\ldots,Proj_c\}$
Set NLP anchors: $\{\mathbf{a_1},\ldots,\mathbf{a_c}\}$ via CLIP)
Warm-up Training:
for 10% of steps **do**
  Sample mini-batch $\{(\mathbf{x_i},y_i)\}$ from $\mathcal{D}$
  Update parameters with $\mathcal{L}_{\text{warm-up}}$
end for
Main Training:
for step = 1 to n_steps **do**
  Sample mini-batch $\{(\mathbf{x_i},y_i)\}$ from $\mathcal{D}$
  if *condition for maplayer update is met* **then**
    Update mapping layers with $\mathcal{L}$
  else
    Update featurizer and classifier and layers with $\mathcal{L}$
  end if
end for

Main Model Update

- Update the featurizer and classifier to improve classification performance using the ***combined loss*** of the weighted alignment and cross-entropy. During inference, the ***exponential moving average*** (EMA) network, initialized as a deep copy of the primary network, is employed to produce predictions.

Mapping Layer Optimization

- Each class is assigned a trainable mapping layer, which projects feature embeddings into the NLP anchor space. The mapping layers are ***updated iteratively*** using our weighted loss function.



$$\mathcal{L} = \sum_{i=1}^{N} w_i \left[\lambda\left(-\cos\left(Proj_{y_i}(f(\mathbf{x}_i)),\, \mathbf{a}_{y_i}\right)\right) + \ell_{CE}\left(g(f(\mathbf{x}_i)),\, y_i\right)\right] \qquad w_i = \frac{\exp\left(\tau\, \cos\left(Proj_{y_i}(f(\mathbf{x}_i)),\, \mathbf{a}_{y_i}\right)\right)}{\sum_{j=1}^{N} \exp\left(\tau\, \cos\left(Proj_{y_j}(f(\mathbf{x}_j)),\, \mathbf{a}_{y_j}\right)\right)}.$$

Result

Table 2: Cross-test accuracy (%) for domain shifts under a noise level of $\eta = 0.25$. Best results in **bold**.

Method	PACS	VLCS	Office-Home	SVIRO	DomainNet
ERM	74.5 ± 0.6	71.9 ± 0.6	54.9 ± 0.3	77.9 ± 0.9	65.0 ± 0.4
GroupDRO	74.7 ± 0.4	71.2 ± 0.2	54.1 ± 0.3	75.4 ± 0.5	68.4 ± 0.2
IRM	71.0 ± 1.9	70.3 ± 0.3	53.9 ± 1.8	62.7 ± 14.7	37.6 ± 15.5
VREx	73.5 ± 0.7	71.8 ± 0.6	53.0 ± 0.9	73.6 ± 2.6	64.9 ± 0.8
Mixup	75.2 ± 0.9	71.9 ± 0.5	57.5 ± 0.3	83.3 ± 1.5	67.9 ± 0.2
A^3W **(ours)**	$\mathbf{82.1 \pm 0.5}$	$\mathbf{76.1 \pm 0.3}$	$\mathbf{65.2 \pm 0.2}$	$\mathbf{93.9 \pm 0.1}$	$\mathbf{73.9 \pm 0.2}$

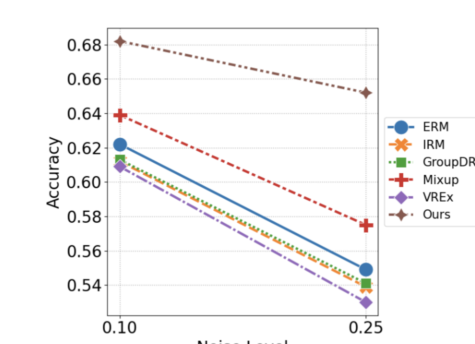
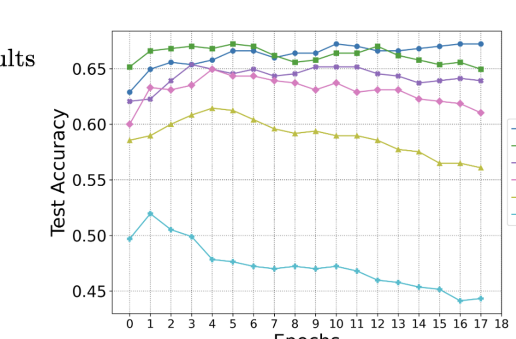
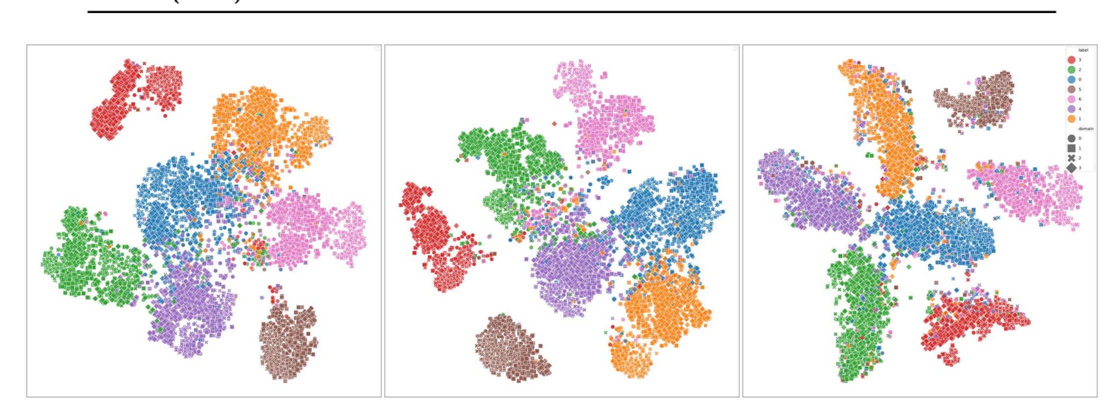
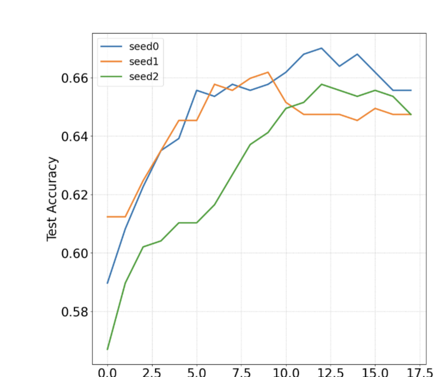


Table 3: Cross-test accuracy (%) for domain shifts under a noise level of $\eta = 0.1$. Best results in **bold**.

Method	PACS	VLCS	Office-Home	SVIRO	DomainNet
ERM	82.0 ± 0.5	75.0 ± 0.3	62.2 ± 0.1	80.8 ± 1.9	70.8 ± 0.5
GroupDRO	82.4 ± 0.3	75.1 ± 0.1	61.3 ± 0.4	78.6 ± 2.1	71.9 ± 0.1
IRM	80.4 ± 1.2	74.6 ± 0.3	61.2 ± 1.2	72.5 ± 8.1	38.9 ± 19.8
VREx	81.4 ± 0.2	75.0 ± 0.1	60.6 ± 0.5	81.6 ± 3.5	66.9 ± 0.4
Mixup	83.6 ± 0.1	75.5 ± 0.2	63.9 ± 0.1	84.1 ± 0.7	69.2 ± 0.6
A^3W **(ours)**	$\mathbf{85.2 \pm 0.2}$	$\mathbf{78.5 \pm 0.2}$	$\mathbf{68.2 \pm 0.3}$	$\mathbf{94.8 \pm 0.3}$	$\mathbf{76.1 \pm 0.1}$

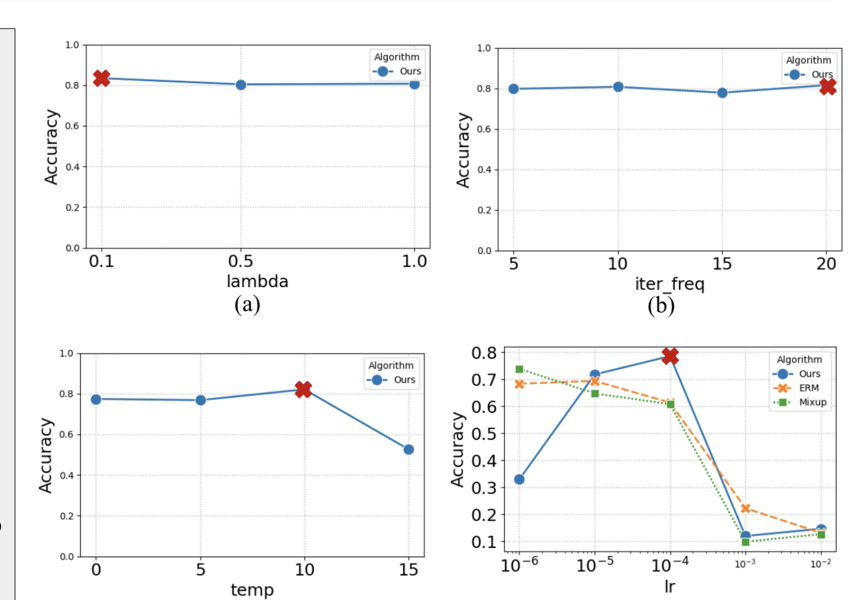






Ablation

- ***NLP Anchor Alignment***: Removing semantic anchors deprives the model of critical external guidance → forces the feature extractor to rely solely on internal cues.
- ***Adaptive Weighting***: Switching from dynamic, softmax-based weights to uniform weighting → disrupts the model's ability to prioritize cleaner samples.



Method	PACS	VLCS	Office-Home	SVIRO	DomainNet
w/out NLP anchor	80.7	75.5	65.1	91.7	85.5
w/out weighted loss	79.7	74.2	64.7	93.6	85.8
A^3W **(baseline)**	**82.1**	**76.1**	**65.2**	**93.9**	**86.5**

Discussion and Futrue Work

- ***Computation Cost***: Extra mapping layers and cosine loss add ~10% runtime → scaling to many classes is challenging.
- ***Dynamic Anchors***: Replace fixed text embeddings with *adaptive, evolving* anchors to enhance visual-text alignment.
- ***Domain-Aware Prompts***: *Tailor text descriptions* to specific domain characteristics for improved adaptation.

Reference: Dai, Zilin, et al. "A Language Anchor-Guided Method for Robust Noisy Domain Generalization." arXiv preprint arXiv:2503.17211 (2025). https://arxiv.org/abs/2503.17211.