MVDG: Multi-View Framework for Domain Generalization

Sujeet Mehta (21D070076)
Course Code: EE451
Guide: Prof. Amit Sethi
Department of Electrical Engineering
Indian Institute of Technology, Bombay

***Abstract*—Domain generalization (DG) focuses on training models that generalize well to unseen domains, addressing the challenges posed by domain shifts. Overfitting in both training and test stages has been a major challenge in DG. This paper proposes a novel multi-view framework, MVDG, which integrates multi-view regularized meta-learning (MVRML) to resist overfitting during training and multi-view prediction (MVP) to stabilize predictions during testing. Experimental results on VLCS datasets demonstrate state-of-the-art (SOTA) performance and validate the effectiveness of our approach. Our framework enhances generalization by leveraging task-based augmentation during training and sample-based augmentation during testing.**

I. Introduction

Traditional supervised learning assumes training and test data share the same distribution, an assumption often violated in real-world applications. Domain generalization (DG) tackles this issue by training models to generalize directly to unseen domains without retraining. Existing DG methods often suffer from overfitting during training and instability in predictions during testing. Meta-learning approaches, while popular, are limited by their reliance on single-task optimization trajectories, resulting in biased optimization directions.

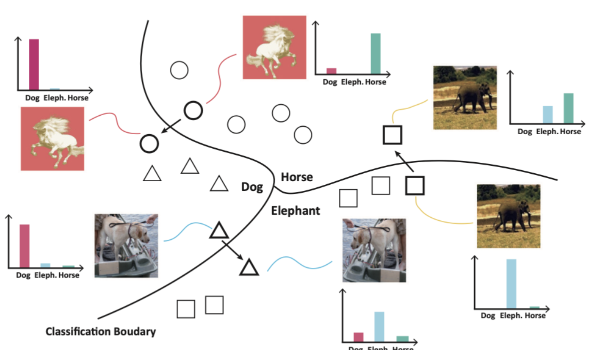

Fig. 1. Discrepancies in Domains[1]

To address these limitations, we propose MVDG, a unified multi-view framework that leverages:

- **Multi-View Regularized Meta-Learning (MVRML):** Employs multiple optimization trajectories to find a robust weight space and enhance generalization.
- **Multi-View Prediction (MVP):** Utilizes weak augmentations to stabilize test predictions by aggregating predictions from multiple views.

This paper presents the design, implementation, and evaluation of MVDG, demonstrating its superiority over SOTA methods.

II. Related Work

A. Domain Generalization

DG methods are categorized into:

- **Domain-Invariant Feature Learning:** Extracts features invariant across domains using techniques like adversarial training.
- **Data Augmentation:** Reduces domain discrepancies by augmenting unseen data using generative models or feature manipulation.
- **Regularization-Based Methods:** Uses auxiliary losses or ensemble learning to mitigate overfitting during training.

B. Meta-Learning

Meta-learning optimizes models for rapid adaptation to new tasks. In DG, episodic training simulates domain shifts for regularization. However, current methods are limited by single-task trajectories, which can lead to biased optimization directions.

III. Proposed Framework

A. Multi-View Regularized Meta-Learning

MVRML enhances generalization by:

- Sampling multiple tasks per iteration to create diverse optimization trajectories.
- Averaging the weights of temporary models to achieve a flat minimum in the loss surface, promoting stability and robustness.

The training process updates model parameters as:

$$\theta_{j+1} = \theta_j + \beta \left(\frac{1}{T} \sum_{t=1}^{T} \theta_{t,\text{tmp}} - \theta_j \right)$$

Algorithm 1 Multi-view Regularized Meta-Learning.

Input: source data $\mathcal{D}_{src}$, network parametrized by θ, hyperparameters: inner loop learning rate α, outer loop learning rate β, the number of optimization trajectories T and the number of sampled tasks s.
Output: the trained parameter

$\theta_0 \leftarrow \theta; j \leftarrow 0$ `// initialize parameter`
while not converged **do**
  $\theta_j^0 \leftarrow$ Initialized by θ_j
  for $t \in \{1, ..., T\}$ **do** `// train temporary models`
    for $i \in \{1, ..., s\}$ **do**
      $\mathcal{D}_{tr}, \mathcal{D}_{te} \leftarrow$ Random split $\mathcal{D}_{src}$ // $\mathcal{D}_{tr} \cup \mathcal{D}_{te} = \mathcal{D}_{src}$, $\mathcal{D}_{tr} \cap \mathcal{D}_{te} = \emptyset$
      Sample mini-batch $\mathcal{B}_{tr}, \mathcal{B}_{te}$ from $\mathcal{D}_{tr}, \mathcal{D}_{te}$
      $\theta_j^{i-1} \leftarrow \theta_j^{i-1} - \alpha \nabla_\theta (\mathcal{L}(\mathcal{B}_{tr} | \theta_j^{i-1}))$ `// train with meta-train data`
      $\theta_j^{i} \leftarrow \theta_j^{i-1} - \alpha \nabla_\theta (\mathcal{L}(\mathcal{B}_{te} | \theta_j^{i-1}))$ `// train with meta-test data`
    end for
    $\theta_{tmp}^t \leftarrow \theta_j^i$ `// assign current temporary parameter`
  end for
  $\theta_{j+1} \leftarrow \theta_j + \beta(\frac{1}{T}\sum_{t=1}^{T}\theta_{tmp}^t - \theta_j)$ `// update the original parameter`
  $j \leftarrow j + 1$
end while

Fig. 2. MVRML Algorithm[1]

B. Multi-View Prediction

MVP stabilizes predictions by:

- Generating weakly augmented views of test images.
- Aggregating predictions across multiple views to produce a robust final output:

$$p = \text{softmax}\left(\frac{1}{m}\sum_{i=1}^{m} f(T(x)|\theta)\right)$$

IV. Experimental Setup

A. Datasets

I evaluated MVDG on VLCS dataset. These datasets represent varying domain gaps and complexities, with VLCS having the smallest gap.
The VLCS dataset consists of 10,729 images across 5 classes: bird, car, chair, dog, and person. It spans 4 domains: VOC2007, LabelMe, Caltech101, and SUN09, making it ideal for domain adaptation and generalization tasks.

Sample images from each domain of the VLCS dataset: Caltech, LabelMe, Sun09, and Pascal 2007.

Fig. 3. VLCS dataset, https://shorturl.at/NIPmB

B. Implementation Details

- Backbone: ResNet-50 pretrained on ImageNet.
- The initial learning rates are set to 0.05 (outer loop) and 0.001 (inner loop) for the first 24 epochs.
- The learning rates are then reduced to 5×10^{-3} (outer loop) and 1×10^{-4} (inner loop) for the final 6 epochs.
- Testing: MVP applied with weak augmentations (random crop and flip).

VLCS Dataset	Caltech(%)	LabelMe(%)	VOC2007(%)	SUN09(%)
Bird	99	85	96.36	80.01
Car	99	98.92	95.99	97.42
Chair	99	42.05	90.89	92.76
Dog	99	47.62	97.14	50.01
Person	99	62.91	98.2	93.67
Domain-Wise Avg	99	79.03	96.05	93.97
Total Avg = 92.01				

Fig. 4. Evaluation on trained dataset[Self]

V. Results and Insights

A. Comparison with SOTA Methods

MVDG outperforms SOTA methods across all datasets:

- On VLCS, MVDG delivers top performance on CALTECH, LABELME, and PASCAL domains.

Method	C	L	P	S	Avg.
DeepAll	96.98	62.00	73.83	68.66	75.37
JiGen [7]	96.17	62.06	70.93	71.40	75.14
MMLD [34]	97.01	62.20	73.01	**72.49**	76.18
RSC [19]	96.21	62.51	73.81	72.10	76.16
MVDG (Ours)	**98.40**	**63.79**	**75.26**	71.05	**77.13**

Fig. 5. Results of experiment with SOTA[1]

B. Insights

- Increasing the number of tasks and trajectories improves generalization but plateaus after a threshold.
- Weak augmentations in MVP yield better results compared to strong augmentations, which may drift images off the original manifold.

VI. Conclusion

MVDG effectively addresses the overfitting and prediction instability challenges in DG by combining task-based augmentation during training and sample-based augmentation during testing. Extensive experiments validate its effectiveness, setting a new benchmark for robustness and generalization in DG tasks.

References

[1] J. Zhang, L. Qi, Y. Shi, and Y. Gao, ”MVDG: A Unified Multi-View Framework for Domain Generalization,” 2022.
[2] J. S. Yoon, K. Oh, Y. Shin, M. A. Mazurowski, and H.-I. Suk, “Domain Generalization for Medical Image Analysis: A Survey ’2024

Code and Resources

The code and implementation details for this work can be accessed at the following link:

- **Repository:** https://www.kaggle.com/code/codingbysujeet/vlcs-sre