Proceedings of the Annual General Donald R. Keith Memorial Conference
West Point, New York, USA
April 24, 2025
A Regional Conference of the Society for Industrial and Systems Engineering

Unsupervised Domain Adaptation with Neuro-Symbolic Artificial Intelligence

John Beggs[1], Sean Coffey[2], and Nathaniel Bastian[1, 2]

[1]Department of Mathematical Sciences, United States Military Academy, West Point, New York 10996

[2]Department of Electrical Engineering & Computer Science, United States Military Academy, West Point, New York 10996

Corresponding author's Email: nathaniel.bastian@westpoint.edu

Author Note: This work was supported by the U.S. Army Combat Capabilities Development Command (DEVCOM) Army Research Laboratory under Support Agreement No. USMA 21050 and the Defense Advanced Research Projects Agency under Support Agreement No. USMA 23004. The views expressed in this paper are those of the authors and do not reflect the official policy or position of the U.S. Military Academy, the U.S. Army, the U.S. Department of Defense, or the U.S. Government.

Abstract: Artificial intelligence (AI) enabled systems for military use on the battlefield must be able to self-adapt to a variety of domains without requiring extensive re-training, especially in resource-constrained and communication-limited environments. This paper proffers a neuro-symbolic AI-enabled system capable of self-adapting to domain shifts at inference time. We conduct unsupervised object detection on simulated overhead drone imagery to eliminate reliance on any ground-truth labels, as well as incorporate multi-modal language models to proportionally merge multiple domain-specific models when inferencing. In conducting multi-domain testing, our system's proportionally merged model outperforms single-domain models. While further work in the space is necessary, we contribute a feasible means of maintaining system performance in multi-domain scenarios.

Keywords: Unsupervised Domain Adaptation, Neuro-Symbolic Artificial Intelligence, Model Merging, Object Detection

1. Introduction

The United States military must develop a decisive advantage over its adversaries in artificial intelligence (AI) technological advancement and operational integration. Modern AI-enabled systems fail to maintain performance across multiple domains, struggling to adapt when environmental conditions (e.g., weather) change, for example, due to their reliance on large, labeled datasets and frequent re-training requirements (Rawat, 2022). This brittleness is particularly problematic in battlefield environments, where operational effectiveness depends on AI-enabled systems that can dynamically adjust to new domains.

To mitigate this challenge, Unsupervised Domain Adaptation (UDA) aims to improve AI model generalization across diverse operating environments without requiring labeled target domain data. A domain shift is considered as when environmental conditions change. In theory, UDA should enable machine learning based AI models for objective detection to function reliably even when the domain shifts. In practice, however, existing UDA approaches are inadequate, as they struggle to self-adapt and fail when assumptions about domain similarity break down (Shamitha & Ilango, 2024). Without mechanisms to actively "reason" about shifting environments, these AI models suffer near immediate performance degradation, limiting their applicability to real-world military aerial reconnaissance tasks, for example.

To overcome these limitations, we proffer a neuro-symbolic AI (NSAI) enabled system capable of self-adapting to domain shifts at inference time, which enhances UDA for object detection by combining symbolic reasoning with model merging. Specifically, our self-adaptive NSAI-enabled system leverages:

1. A novel methodology that integrates a deep neural network (DNN) with multi-modal language models, a vision language model (VLM) and large language model (LLM), to infer symbolic representations within an image, and
2. A weighted magnitude selection algorithm for model merging that fuses learned information about separate domains.

Our novel UDA for object detection approach enables continuous self-adaptation to new environmental conditions without re-training or human supervision. We demonstrate the NSAI-enabled system's effectiveness on simulated overhead drone imagery, showing its potential as an application for military aerial reconnaissance in multi-domain battlefield settings. This gives the warfighter access to an AI system that remains reliable as weather and terrain rapidly shift in combat operations.

ISBN: 97819384969-9-8

2. Literature Review

UDA attempts to help AI models adapt to new domains without labeled target domain data, but current approaches fail to deliver reliable performance in real-world settings. Self-adaptation is a critical component of any model deployed on the future battlefield as frequent re-training and labeled data are unrealistic to have access to. Adversarial learning methods, such as Joint Adversarial Domain Adaptation and Adversarial Discriminative Domain Adaptation, work by aligning distributions between domains with adversarial loss functions. However, these methods often break down due to unstable training and sensitivity to hyper-parameters, ultimately failing when domain shifts are too severe (Li et al., 2019; Tzeng, Hoffman, Saenko, & Darrell, 2017). Hybrid approaches, such as two-stage adaptation frameworks, combine different techniques to improve self-adaptation but are often so complex resulting in inconsistent performance (Yu, Zhai, & Zhang, 2022). Optimal transport-based approaches, like Robust Deep Adaptation via Optimal Transport, aim to bridge domain gaps by capturing local feature structures but are limited by high computational costs and poor scalability (Gilo, Mathew, Mondal, & Sanodiya, 2023).

Many UDA methods for object detection are built on the assumption that domain shifts will be gradual and predictable. In practice, however, environments do not reflect this assumption, particularly in military battlefield settings, leading to system feature alignments failing to maintain across domains (Zhang & Zhang, 2022). Additionally, current UDA methods lack the ability to self-adapt at the point of model inference. As such, there is need for a fundamentally different approach. NSAI presents the opportunity to combine the strengths of traditional, non-symbolic DNNs with AI models for symbolic reasoning (Jalaian & Bastian, 2023). A system with symbolic reasoning incorporated allows for auto-adjustment of its predictions in real-time. Still, many current NSAI-enabled systems are computationally inefficient, thus limiting their operability in real-world environments.

In the context of UDA for object detection, only using pure NSAI for self-adaptation may still experience computational inefficiency. Thus, model merging allows for the continual learning necessary in an NSAI-enabled system but eliminates the requirement of any retraining. Some techniques, such as task arithmetic and simple weight averaging, resolve conflicts between conflicting model parameters but require a more proportional way to merge. Alternative merging approaches, such as MagMax, enable continual learning, making them well-suited for environments where self-adaptation is critical (Marczak, Twardowski, Trzciński, & Cygert, 2024). Our approach builds on this effort to prevent AI models forgetting previously learned patterns.

3. Methodology

Our solution approach starts by estimating the number of objects in each image, generating rough outlines, training models for each environment, and finally merging them using insight from language models to improve multi-domain performance. This novel methodology integrates symbolic reasoning (using an integrated VLM-LLM approach) with DNN-based models for object detection to form a self-adaptive NSAI-enabled system. Figure 1 depicts our overall solution architecture, which consists of three stages. Each stage contributes to the self-adaptation process where domain-specific representations are extracted, mapped to symbolic embeddings, and proportionally merged at the point of inference. This NSAI-enabled system ensures UDA through dynamic re-weighting of the final merged model, while preserving the essential feature representations.

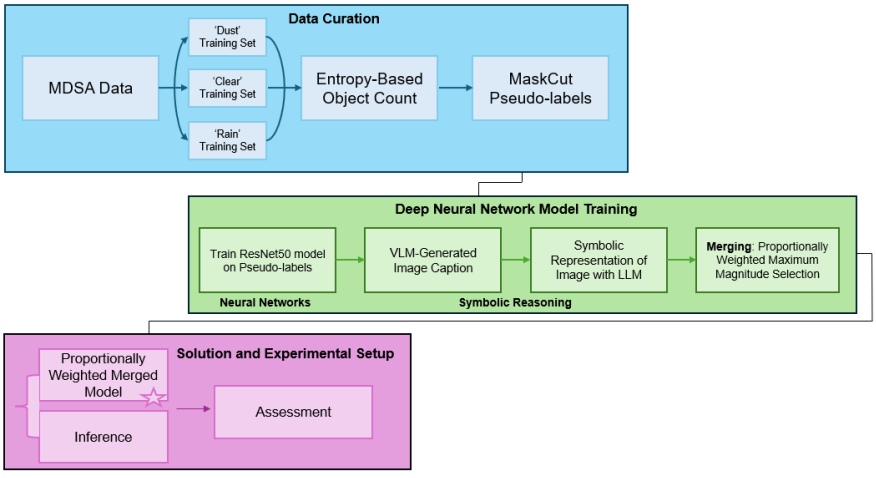

Figure 1: Solution Architecture - UDA for Objective Detection using a Self-Adaptive NSAI-enabled System

ISBN: 97819384969-9-8

3.1. Data Curation

Our NSAI-enabled system is trained and tested with simulated overhead drone images from the Multiple Distribution Shift - Aerial (MDSA) dataset, which provides imagery useful for overhead object detection tasks (Ngu et al., 2025). Further, use of this dataset is useful for simulating real-world variability in weather and visibility. Although all subsets in MDSA contain urban landscapes, they are differentiated into five distinct domains: Clear, Dust, Fog, Rain, and Snow. We selected three - Clear, Dust, and Rain - for training. Each domain contains 100 images, approximately half of which include vehicles (Figure 2).

To prepare the dataset for object detection, we first estimate the number of objects in each image using entropy-based object counting. Estimating object count enables pseudo-labeling without human input, which is critical to adapting to new, unlabeled domains. This method identifies regions in an image with high pixel variability, which correspond to areas most likely to contain objects (Figure 3). With this estimate, we better gain a quantitative understanding of the image's complexity without needing the label. Additionally, we can better inform the appropriate masking strategy for later stages of data curation.

With the object number estimates, we use MaskCut (Wang, Girdhar, Yu, & Misra, 2023) to generate masks to isolate the distinct objects in the unlabeled images. MaskCut provides pseudo-labels to allow for training models without needing manual annotation. It works by constructing an affinity matrix, measuring the cosine similarity between image patch feature vectors extracted with a vision transformer backbone. With the estimated count, we iteratively apply the process of matrix partitioning to signal background and foreground regions to ensure the masks are more granular. Initial masks may notice the outline of a vehicle, but after this refinement process, finer details like side mirrors are distinguishable (Figure 4). Creating these pseudo-labels for training models in any new environment helps to support adaptation when ground truth is not available.

Figure 2: MDSA Example Image

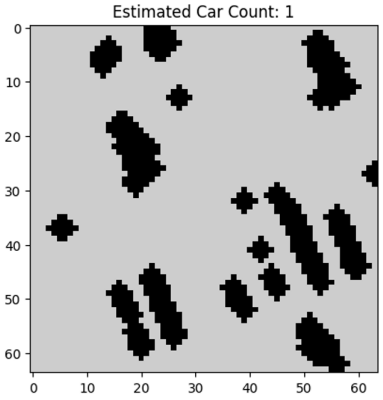

Figure 3: Fig. 2's Object Count

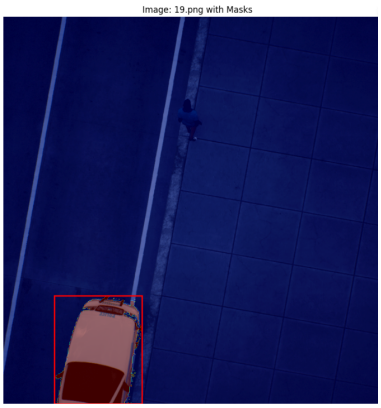

Figure 4: MaskCut Pseudo-Label

3.2. Deep Neural Network Model Training

For each domain, we train a ResNet50-based model using the MaskCut-generated pseudo-labels. Training separate models for each environment is essential to capturing the unique features in each domain, thus improving model performance in varied conditions. The ResNet50 backbone is initialized with pre-trained weights from a Distillation with No Labels model to leverage its feature extraction ability. Each model is trained with a single object class, aligning with the pseudo-label representation. For simplicity, the single class is cars. Training parameters are a base learning rate of 0.001, a batch size of two images per iteration, and a maximum of 1,000 iterations. The models are trained independently but have the same architecture.

We next use a VLM to incorporate symbolic reasoning into the solution architecture. The use of a VLM enables the system to 'describe' images in words at inference-time, helping it recognize environmental characteristics beyond the pixels. This capability is a core element of our NSAI-enabled system, enabling the model to adapt to severe domain shifts. The Bootstrapped Language-Image Pre-training (BLIP) model (Jian, Gao, & Vosoughi, 2024) is used due to its detailed and descriptive captions generated zero-shot when given a visual input. This process aims to provide a textual description that captures the domain-specific essence of images that may be considered as belonging to more than one domain. BLIP's encoder-decoder architecture extracts visual features and translates them into natural language descriptions. By integrating symbolic representations into the NSAI pipeline, the system's ability to generalize across domains is improved without solely relying on traditional feature alignment. The produced captions are designed to extend beyond generic descriptions, thus forming the foundation for symbolic reasoning. In turn, this enables the final merged model to make logical inferences about the domain of each image.

ISBN: 97819384969-9-8

To quantify a domain's proportion of the merged model, we use LLM-based textual embeddings generated from the Bidirectional Encoder Representation from Transformers (BERT) model (Kenton & Toutanova, 2019). This helps the system to understand how closely a new image matches each known environment, which guides the weighting of model contributions. Domain-specific captions are tokenized into embeddings using a pre-trained BERT model. Domain embeddings are then formed by averaging the encoded words and phrases, allowing direct comparison with new image captions. Its embedding is compared with the domain vectors using cosine similarity to calculate the relevance of the domain. These similarity scores generate the self-adaptive domain weights, which are used to dynamically adjust the contribution of each domain in model merging.

With trained models and domain weights, we merge models utilizing a novel approach inspired by MagMax (Marczak et al., 2024) to combine domain-specific models into a unified architecture. Merging ensures we combine the strengths from each domain-specific model, allowing generalization to unseen environments. The merging strategy balances proportional constraints from the textual embeddings with magnitude-based adjustments to preserve the models' critical features. For each shared model parameter, we compute a proportional merge as a weighted sum of the parameter values. To further refine the process, we conduct a magnitude-based adjustment of the merged model's final parameters by comparing the weighted magnitude of each parameter. The highest-magnitude parameter is selected as the dominant contributor, ensuring the model retains domain-specific features. This adjustment is the critical component to ensuring domain-specific features are not diluted. The final merged parameter is computed as a blend of the proportional merge and the magnitude-adjusted parameter. This weighted merging approach (Algorithm 1) retains each domain's strength while preventing significant overfitting to any one domain.

Algorithm 1: Weighted Magnitude Selection for Model Merging

Input: Domain models $\theta_d, \theta_c, \theta_r$, Image caption
Output: Merged model θ_M
Compute Domain Weights:
forall $D \in \{dust, clear, rain\}$ **do**
  Compute embedding v_D from domain phrases
end
Compute caption embedding v_C and similarity scores $s_D \leftarrow \cos(v_C, v_D)$
Normalize weights: $w_D \leftarrow s_D / \sum s_D$
Merge Model Parameters:
forall *parameters* k *in* θ_M **do**
  Compute proportional merge: $\theta_P^k \leftarrow \sum_D w_D \theta_D^k$
  Compute magnitude weights: $M_D^k \leftarrow w_D |\theta_D^k|$
  Select dominant parameter: $\theta_A^k \leftarrow \theta_{\arg\max(M_d^k, M_c^k, M_r^k)}$
  Final merge: $\theta_M^k \leftarrow 0.5\theta_P^k + 0.5\theta_A^k$
end
Load θ_M into model
return θ_M

3.3. Solution and Experimental Setup

To evaluate the effectiveness of the trained models, we measure performance using Precision and F1 Score. Performance is evaluated at the pixel level by comparing predicted object masks with ground-truth annotations. These metrics aim to provide a comprehensive understanding of the proposed approach's ability to detect objects in the unlabeled, cross-domain test set images. For evaluation, we randomly select 20 images and their corresponding annotations for processing. Each image's pixel values are normalized and the assessment is on just one object class, cars.

We compare the model's predicted binary masks against ground-truth masks derived from polygonal annotations. This allows for a direct assessment of the models. To ensure consistency across each variation, we manually label images for each evaluated model's test set. While ground-truth annotations may be applicable to our system for validation, they lack feasibility in the more practical setting the system aims to emulate. We apply a confidence threshold of 0.1 to refine the predicted masks which ensures that the system accounted for high-probability predictions while suppressing noise in uncertain regions. However, given the use of binary masks, we aggregate the results for each image then average each image's result to assess final performance.

4. Results and Discussion

First, we assessed the single-domain and merged models on the multi-domain test set (Figure 5). The individual single-domain models performed comparably to one another, but each suffered slight performance degradation when tested outside its training domain. Precision and F1 scores in cross-domain settings frequently dropped below 0.3 and 0.2, showing their inability to generalize beyond a single domain condition. By contrast, the merged model maintained precision and F1 scores exceeding 0.4, demonstrating improved stability across domains (Table 1). This confirms that weighted magnitude selection preserves critical domain-specific features while allowing the model to generalize effectively.

Next, we evaluated the models on their respective single-domain training sets. Unexpectedly, the single-domain models did not achieve the highest performance within their own domains (Table 2). Instead, the merged model outperformed almost all single-domain models, even in their original training domains (Figure 6). This suggests that model merging is not simply a self-adaptation mechanism for domain shift but also enhances performance within individual domains. The observed underfitting aligns with expectations for any unsupervised setting but remains rationale for the confirmation of our original hypothesis.

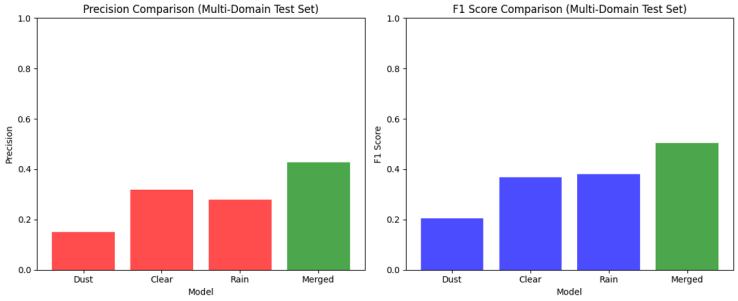

Figure 5: Performance on Multi-Domain Test Set

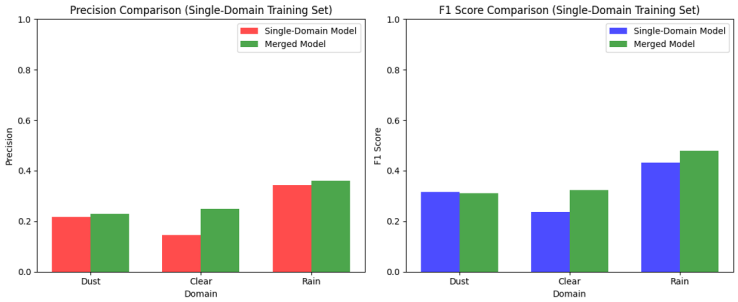

Figure 6: Performance on Single-Domain Training Sets

Table 1: Multi-Domain Test Set Results

Model	Precision	F1 Score
Dust	0.1496	0.2054
Clear	0.3182	0.3687
Rain	0.2790	0.3798
Merged	0.4266	0.5050

Table 2: Single-Domain Training Set Results

Domain	Precision		F1 Score	
	SDM	MM	SDM	MM
Dust	0.2162	0.2282	0.3166	0.3120
Clear	0.1442	0.2484	0.2355	0.3244
Rain	0.3430	0.3593	0.4320	0.4783

These results reveal the viability of our proposed NSAI approach for UDA, as existing UDA methods struggle when domain shifts are severe, often requiring explicit domain alignment or labels. Our approach achieves cross-domain generalization without retraining, meaning that models can dynamically self-adapt at inference time without significant prior exposure to unseen environments. For real-world military applications, such as aerial intelligence, surveillance, or reconnaissance, this

ability is critical to success due to rapidly changing environmental conditions, even in a singular location. The ability to merge models at inference time without any human intervention ensures that the system's ability to identify these domain shifts is strong. Our self-adaptive NSAI-enabled system shows that in unsupervised tasks, performance can increase across domains.

5. Conclusion

Our results indicate that weighted magnitude selection enables model merging without retraining while maintaining performance across domains. Even without ground-truth annotations for training, the merged model outperformed the single-domain models in both cross-domain and same-domain settings. These results highlight that potential of our NSAI framework to address a major limitation of traditional UDA methodologies: the inability to self-adapt at inference without domain supervision. With further development, this means AI systems can continue functioning in dynamic environments without needing human intervention or manual re-training. Several challenges still remain. The computational efficiency of our approach is a key concern, as the reliance on BERT for the self-adaptive component is bulky and the ResNet50 models are not as lightweight as will be feasible on some low size, weight and power platforms. While the merging strategy takes a positive step towards preventing forgetting, further refinement is necessary to ensure stability when incorporated into a wider range of domains with more single-domain models. Future will work focus on three areas: 1) we will reduce computational overhead by exploring lighter-weight object detection and multi-modal language models; 2) we will extend the approach to handle real-world aerial imagery with lower-resolution data, and 3) we will integrate our approach into a hierarchical federated learning paradigm for decentralized self-adaptation. These improvements will ensure that the system more closely aligns with expectations of future battlefield settings. Additional effort will incorporate military equipment imagery, varied terrain, and diverse domains to better simulate real-world battlefield conditions. Real-world AI-enabled systems must function in complex, dynamic operating environments where domain shifts are constant and unpredictable. By eliminating reliance on labels and re-training, our NSAI-based approach takes a step forward to bridge the gap in UDA. Continued refinement will be necessary to move this methodology from proof-of-concept to operational readiness, ensuring that the military's AI-enabled systems hold an advantage on the battlefield.

References

Gilo, O., Mathew, J., Mondal, S., & Sanodiya, R. (2023). Rdaot: Robust unsupervised deep sub-domain adaptation through optimal transport for image classification. *IEEE Access*.

Jalaian, B., & Bastian, N. D. (2023). Neurosymbolic ai in cybersecurity: Bridging pattern recognition and symbolic reasoning. In *Milcom 2023-2023 ieee military communications conference (milcom)* (pp. 268–273).

Jian, Y., Gao, C., & Vosoughi, S. (2024). Bootstrapping vision-language learning with decoupled language pre-training. *Advances in Neural Information Processing Systems*, *36*.

Kenton, J. D. M.-W. C., & Toutanova, L. K. (2019). Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of naacl-hlt* (Vol. 1, p. 2).

Li, S., Liu, C. H., Xie, B., Su, L., Ding, Z., & Huang, G. (2019). Joint adversarial domain adaptation. In *Acm multimedia.*

Marczak, D., Twardowski, B., Trzciński, T., & Cygert, S. (2024). Magmax: Leveraging model merging for seamless continual learning. *arXiv preprint arXiv:2407.06322*.

Ngu, N., Taparia, A., Simari, G. I., Leiva, M., Corcoran, J., Senanayake, R., … Bastian, N. D. (2025). *Multiple distribution shift – aerial (mds-a): A dataset for test-time error detection and model adaptation.* Retrieved from `https://arxiv.org/abs/2502.13289`

Rawat, D. B. (2022). Artificial intelligence meets tactical autonomy: Challenges and perspectives. In *2022 ieee 4th international conference on cognitive machine intelligence (cogmi)* (pp. 49–51).

Shamitha, S., & Ilango, V. (2024). Applications of deep learning algorithms for object detection in real-time drone surveillance: Systematic review, recent developments and open issues. In *2024 fourth international conference on advances in electrical, computing, communication and sustainable technologies (icaect)* (pp. 1–6).

Tzeng, E., Hoffman, J., Saenko, K., & Darrell, T. (2017). Adversarial discriminative domain adaptation. In *Computer vision and pattern recognition.*

Wang, X., Girdhar, R., Yu, S. X., & Misra, I. (2023). Cut and learn for unsupervised object detection and instance segmentation. In *Proceedings of the ieee/cvf conference on computer vision and pattern recognition* (pp. 3124–3134).

Yu, Y., Zhai, Y., & Zhang, Y. (2022). Align and adapt: A two-stage adaptation framework for unsupervised domain adaptation. In *Acm multimedia.*

Zhang, C., & Zhang, J. (2022). Transferable regularization and normalization: Towards transferable feature learning for unsupervised domain adaptation. *Information Sciences*.