Empirical Evaluation and Analysis of State-of-the-Art Methods in Unsupervised Domain Adaptation

Grace Vincent[1], Purushothaman Saravanan[2], Iype Eldho[3]
gmvincen@ncsu.edu[1], psarava3@ncsu.edu[2], ieldho@ncsu.edu[3]
ECE 792-061: Advanced Topics in Machine Learning

Background

In real-world applications, data collected from one source or environment may have different characteristics or distributions from the data collected from another source or environment. This can lead to poor generalization performance of machine learning models that are trained on one domain and tested on another. Unsupervised domain adaptation (UDA) is a technique that addresses this problem by allowing models to adapt to new domains using only unlabeled data. The Office-31[1] dataset provides a testbed for evaluating the performance of UDA methods, as it contains over 4,600 images from three different domains with different resolutions, noise levels, and lighting conditions, for a total of 31 different office-based objects with distributions depicted in Fig. 1 and sample images presented in Fig. 4. This dataset allows us to examine the adaptation from web-based images to realistic office and home environments and evaluate six domain shift configurations for each implemented technique.

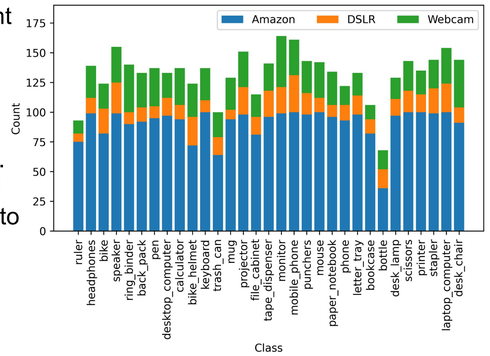

Figure 1: The distribution of class labels for the OFFICE-31[1] dataset. Each bar represents a different class and the different colors represent the different domains within the dataset.

Model Selection

- These domain adaptation techniques fall into the categories of adversarial training, feature alignment, and self-training.
- We have implemented four domain adaptation models: Domain Adversarial Neural Network (DANN) [2], Multi-Adversarial Domain Adaptation (MADA) [3], Cross-domain transformer (CDTrans) [4], and Contrastive Adaptation Network (CAN) [5]. DANN and MADA belong to the adversarial training category, while CDTrans and CAN fall under the category of feature alignment.

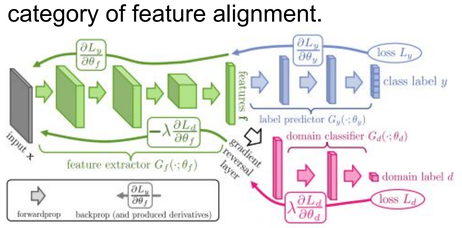

Figure 2: DANN Architecture

- The architecture of DANN with the gradient reversal layer is shown in Fig. 2. Gradient reversal ensures that the feature distributions over the two domains are made similar.
- The source and target images are sent to source branch and target branch as shown in Fig. 3. Self-attention module is involved to learn the domain-specific representations.

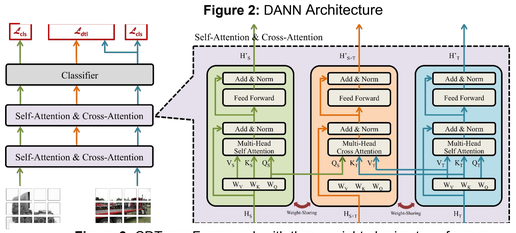

Figure 3: CDTrans Framework with three weight-sharing transformers

Methods

- Experiments showed that using a ResNet-50 encoder for feature extraction significantly improved the results of DANN, MADA, and CAN. A DeiT is used as the backbone for CDTrans.
- Hyperparameter tuning was conducted to optimize the performance of the models. We tuned parameters such as optimizer, learning rates, and regularization strength to achieve the best results. A learning rate scheduler was used to improve convergence and accuracy.
- Table 1 notes the model sizes and their average inforencing time for a single image.

Model	Number of Parameters	Avg Inference Time
DANN	27.2M	0.9 ms
MADA	26.2M	2.8 ms
CDTrans	86.6M	0.2ms
CAN	23.6M	0.3 ms

Table 1: Statistics for model size based on number parameters. We also highlight the average inference time of a single image for the four models.

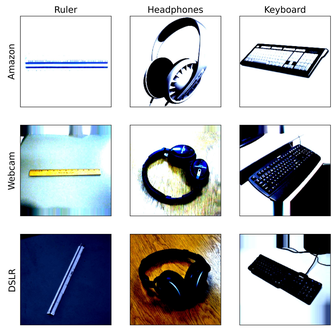

Figure 4: Sample images of three different classes from the three different domains.

- Robustness with respect to noise is important since real-world data often contains various types of noise, which can impact the performance and generalization of machine learning models.
- To assess the robustness of our models, we conducted a thorough robustness survey, where we added random Gaussian noise to 75% of the images in the target domain with varying standard deviations from 0.1 to 5.0 (Fig. 5), and evaluated the models' generalization performance.

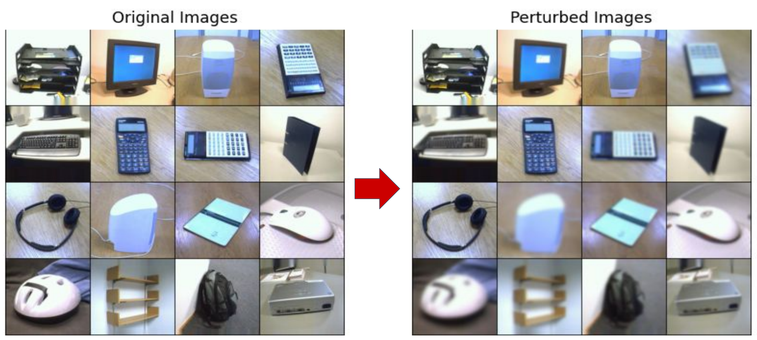

Figure 5: Example images from the Webcam domain of the effects of random Gaussian Blur applied to the images. The zero-mean Gaussian noise was added randomly to the 75% of the images and had a random standard deviation ranging from 0.1 to 5.0.

References
[1] K. Saenko, B. Kulis, M. Fritz, and T. Darrell, “Adapting visual category models to new domains,” in Computer Vision – ECCV 2010, K. Daniilidis, P. Maragos, and N. Paragios, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 213–226
[2] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. Marchand, and V. Lempitsky, “Domain-adversarial training of neural networks,” J. Mach. Learn. Res., vol. 17, no. 1, p. 2096–2030, jan 2016
[3] Z. Pei, Z. Cao, M. Long, and J. Wang, “Multi-adversarial domain adaptation,” 2018.
[Online]. Available: https://arxiv.org/abs/1809.02176
[4] T. Xu, W. Chen, P. Wang, F. Wang, H. Li, and R. Jin, “Cdtrans: Cross-domain transformer for unsupervised domain adaptation,” arXiv preprint arXiv:2109.06165, 2021.
[5] G. Kang, L. Jiang, Y. Yang and A. Hauptmann, "Contrastive Adaptation Network for Unsupervised Domain Adaptation," in 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 2019 pp. 4888-4897.

Results

- The Accuracy and F1 score results for UDA for the Amazon to Webcam, DSLR to Webcam and Webcam to Amazon are displayed in Table 2.
- Webcam domain proves to be difficult to classify given its a low image resolution which is highlighted in Fig. 4.
- The models perform in the following order: DANN < MADA < CAN < CDTrans.

Model	Amazon to Webcam		DSLR to Webcam		Webcam to Amazon	
	Acc	F1	Acc	F1	Acc	F1
DANN	0.786	0.780	0.969	0.969	0.606	0.596
MADA	0.764	0.762	0.957	0.957	0.656	0.635
CAN	0.949	**0.968**	0.964	**0.992**	0.773	**0.846**
CDTrans	**0.967**	0.965	**0.991**	0.989	**0.819**	0.820

Table 2: Accuracy and F1-Score as a result of Domain Adaptation for our four models, DANN, MADA, CDTrans, and CAN. We present three domain adaptations.

Model	Amazon to Webcam	DSLR to Webcam	Webcam to Amazon
	Acc	Acc	Acc
DANN	0.567	0.728	0.463
MADA	0.556	0.703	0.477
CDTrans	0.504	0.570	0.494
CAN	0.124	0.167	0.113

Table 3: Results of the robustness survey. The accuracy for the four models is presented for 3 different domain adaptations.

Conclusions

- The DANN model is more robust compared to the others, primarily due to its comparatively simpler architecture.
- CDTrans outperforms all other models in terms of Accuracy. It demonstrates that Transformers have better generalization ability over Convolutional Networks
- The results show that certain domains like Amazon and DSLR, are better for training with before transferring for UDA, due to the better image quality.
- We note that CDTrans produces higher accuracy but CAN returns a higher F1-Score. Depending on the vision task and domains, the F1-Score would be more indicative of a high performing model.