AFRL-RI-RS-TR-2023-192

LEARNING FROM SMALL LABELED SETS BY USING TASK AND DOMAIN STRUCTURE

CORNELL UNIVERSITY

NOVEMBER 2023

FINAL TECHNICAL REPORT

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED

STINFO COPY

AIR FORCE RESEARCH LABORATORY
INFORMATION DIRECTORATE

■ **AIR FORCE MATERIEL COMMAND** ■ **UNITED STATES AIR FORCE** ■ **ROME, NY 13441**

NOTICE AND SIGNATURE PAGE

AFRL-RI-RS-TR-2023-192 HAS BEEN REVIEWED AND IS APPROVED FOR PUBLICATION IN ACCORDANCE WITH ASSIGNED DISTRIBUTION STATEMENT.

FOR THE CHIEF ENGINEER:

/ S /
NANCY A. SEPULVEDA
Work Unit Manager

/ S /
SCOTT D. PATRICK
Deputy Chief
Intelligence Systems Division
Information Directorate

REPORT DOCUMENTATION PAGE

1. REPORT DATE NOVEMBER 2023	**2. REPORT TYPE** FINAL TECHNICAL REPORT	**3. DATES COVERED** **START DATE** JUNE 2019	**END DATE** MAY 2023

4. TITLE AND SUBTITLE
LEARNING FROM SMALL LABELED SETS BY USING TASK AND DOMAIN STRUCTURE

5a. CONTRACT NUMBER FA8750-19-2-0039	**5b. GRANT NUMBER** N/A	**5c. PROGRAM ELEMENT NUMBER** 61101E
5d. PROJECT NUMBER	**5e. TASK NUMBER**	**5f. WORK UNIT NUMBER** R2T2

6. AUTHOR(S)
Bharath Hariharan

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Cornell University
107 Hoy Road
Ithaca NY 14850

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Air Force Research Laboratory/RIEA
525 Brooks Road
Rome NY 13441-4505

10. SPONSOR/MONITOR'S ACRONYM(S)
AFRL/RI

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
AFRL-RI-RS-TR-2023-192

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited. This report is the result of contracted fundamental research deemed exempt from public affairs security and policy review in accordance with SAF/AQR memorandum dated 10 Dec 08 and AFRL/CA policy clarification memorandum dated 16 Jan 09.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

This report details their approach and results for the DARPA LwLL or Learning with Less Labels program. The key problem tackled in this report is the problem of learning from small labeled datasets. This is a problem that is frequent in many application areas. It is especially an issue for defense applications where labeled data might be classified and therefore limited. Unfortunately, modern deep learning systems need large numbers of labeled training examples, making them ineffective in the few label regime. In this report, they describe their efforts to produce new kinds of learning machinery that can learn effectively from few labels.

Their key insight for this problem is to leverage structure in the task and in the domain. This structure may take the form of domain knowledge about needed invariances. It might also be more amorphous and difficult to specify, but nevertheless it might dictate which classifiers work well in this domain.

15. SUBJECT TERMS
Deep learning, self-supervised learning, few-shot learning, metal-learning, image classification, object detection

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**
a. REPORT U	**b. ABSTRACT** U	**C. THIS PAGE** U	SAR	22

19a. NAME OF RESPONSIBLE PERSON
NANCY SEPULVEDA

19b. PHONE NUMBER *(Include area code)*
N/A

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1.0 SUMMARY

This report details our approach and results for the DARPA LwLL or Learning with Less Labels program.

The key problem tackled in this report is the problem of learning from small labeled datasets. This is a problem that is frequent in many application areas. It is especially an issue for defense applications where labeled data might be classified and therefore limited. Unfortunately, modern deep learning systems need large numbers of labeled training examples, making them ineffective in the few label regime. In this report, we describe our efforts to produce new kinds of learning machinery that can learn effectively from few labels.

Our key insight for this problem is to leverage structure in the task and in the domain. This structure may take the form of domain knowledge about needed invariances. It might also be more amorphous and difficult to specify, but nevertheless it might dictate which classifiers work well in this domain. We explored three ways of incorporating this idea of domain structure:

1. **Incorporating known invariances:** We explored new architectures for machine learning that explicitly incorporate known invariances. In particular, we designed recognition architectures that allowed for features to move around spatially, to incorporate pose invariance.
2. **Learning across domains:** We explored new techniques for transferring learned models across domains, taking into account the similarity between domains.
3. **Domain-specific learning algorithms:** We also explored specialized learning techniques for two domains: satellite images and self-driving cars. For these domains, we used the special structure of the data to learn models without labels.

We experimented with both standard benchmarks as well as DARPA's evaluation. We found generally that:

1. When nothing is known about the domain, the best strategy is one of **self-training**, where we use the network's own outputs, appropriately filtered, as "ground truth" labels. Here we identified new ways of performing self-training even with networks trained on a completely different problem domain.
2. If we know the domain we are working with, we can even build completely unsupervised recognition models that outperform supervised ones. The key is to use knowledge of the underlying data generation processes. We demonstrated results on both satellite images and with self-driving cars.

2.0 INTRODUCTION

Modern recognition systems based on deep learning are extremely data hungry. On the one hand, this means that as the amount of labeled data increases, they get more and more accurate. On the other hand, when labeled data is limited, their accuracy drops significantly. This is a problem for many application domains where labeling requires expertise and is therefore difficult to acquire. It is also a problem for applications of national security, where data might be classified and therefore cannot be farmed out to external labelers for annotation. This is the motivation behind the Learning with Less Labels program, which aimed at maintaining the accuracy of contemporary systems while substantially reducing the number of labels required.

From the outset, we realized that when labeled data is very limited in a particular application, there simply isn't enough signal in the labeled data to train a good model. We therefore need to bring in information that is **external** to the available labeled data. We hypothesize that there are at least two such sources of external information:

1. Information about the target application that a domain expert might have, but that is difficult to specify as labels, and
2. Knowledge/understanding that the system might have learnt in other domains which are related to the problem at hand, but where labeled data is plentiful.

To concretize this insight, let us consider a particular example. Suppose we want to build a recognition model to recognize different kinds of military aircrafts, but we have very limited training data. By contrast, aviation enthusiasts have collected large labeled datasets of civilian, or old military aircrafts; a model trained on this public data may have learnt just the right kind of feature representation to perform our task of interest. This corresponds to the second source of information above. In addition, we know that aircrafts may be seen from many different viewpoints, but the viewpoint should be irrelevant to the classification. Thus, a small number of views of a new kind of aircraft should be enough to tell us what the aircraft looks like from any other view, and thus should suffice to build a recognition model. This corresponds to the first insight. In our approach, we looked for ways to incorporate exactly these insights.

Past work: There has been a lot of work on reducing the number of labels required for recognition. This past work primarily falls into the following three buckets:

Few-shot learning techniques aim to build recognition systems that can learn novel concepts from very few examples and can be found in [1-5]. However, these typically assume that a large labeled dataset for a very related problem is available. Many few-shot learning techniques use the idea of *meta-learning,* where the recognition system uses the large labeled dataset to learn how to learn: concretely, it estimates parameters of a learning algorithm so that the learning algorithm can learn future concepts from very few labeled examples. Some of the related work in this direction includes [1-19].

Instead of using labeled data from a related problem, *self-supervised learning* techniques try to use unlabeled data. The dominant approach here is to use the unlabeled data to learn a feature representation that can then be used in conjunction with a linear classifier for the task of interest. While many self-supervised learning techniques have been proposed, the current state-of-the-art is contrastive learning [20-27]. Contrastive learning techniques generally aim to spread data points out in feature space while making sure that small variations of the same data point are embedded close to each other.

Another way of using unlabeled data is to combine it with labeled data and perform *semi-supervised learning* [28]. Techniques here use the unlabeled data to enforce some regularization over the learnt model. A straightforward but effective approach is to use the available labeled data to train a classifier, then use this classifier to *pseudo-label* the unlabeled data, and then promote the most confident predictions to ground truth. State-of-the-art expands upon this by enforcing that predicted labels on unlabeled data points are robust to minor variations (*augmentations*) of the data point.

Crucially, all previous work focuses primarily on the learning techniques themselves, and fails to use relevant knowledge about the problem domain.

3.0 METHODS, ASSUMPTIONS, AND PROCEDURES

We will first look at the problem setup that was exemplified by the evaluation in the DARPA LwLL program, and discuss our contributions in this setup. We will then discuss a broader problem setup that is nevertheless relevant to the goals of LwLL.

3.1 Learning recognition systems for arbitrary new domains with few labels

3.1.1 Problem setup

We are provided a small labeled dataset for a new task, along with a larger unlabeled dataset from the same domain. In addition to this data, we have access to a set of pre-trained models that have not be trained in this data. One of the key challenges therefore is to best use these pre-trained models to learn a recognition system for this new task.

3.1.2 Approach

Our approach has two steps: (1) Finding the right pre-trained model to use, and (2) adapting it to the domain of interest. We address the second question first, since that will yield us valuable insight.

Adapting pre-trained models to a new domain:

Suppose that we have a pre-trained model M_s that we want to now adapt to a new target task. This pre-trained model may be trained on a very different domain. For example, the pre-trained model may be trained on internet images, but the target domain might involve satellite imagery. The key question is how best to adapt the source model to the target domain.

If one had labels in the target domain, this adaptation is easy: we simply use the pre-trained source model M_s as an initialization for the target domain model M and optimize the standard "cross entropy" loss function:

$$L^{labelled} = \frac{1}{n}\sum_{i=1}^{N} L_{ce}\left(M(x_i), y_i\right) \tag{1}$$

Here, L_{ce} denotes the cross entropy loss (or the KL divergence) between the predicted class probabilities from the model M and the true label. But we have very few labeled examples. When training a model with very few examples, even when starting from a good initialization, one is bound to overfit and produce a model that works well for the small training set, but fails otherwise.

To regularize the problem better, therefore, we used the unlabeled data. To do so, we looked at the predictions of the source domain model M_s on images from the target domain. A priori, these labels should be meaningless, since they come from a very different domain. For example, an internet image classifier may have a label "banana" that makes no sense for a satellite image. However, we found generally that the *grouping* of data points induced by the source classifier is in fact meaningful. That is, target domain data points from the same target class tend to receive similar predictions from the source domain classifier (even though the prediction itself may be meaningless).

This in turn suggests that we can use the predictions of the source classifier on the target domain unlabeled data as "pseudo-labels" that are informative of the unknown true labels.

We use this insight and add an additional "head" (i.e., another layer using the same backbone) on the target domain model that tries to predict these pseudo-labels produced by the source domain classifier. In other words, we construe our model M( ) as h(f( )), where f is a feature extractor and h is a linear classifier. We use a separate linear classifier h' operating on the same feature space to reproduce the pseudo-labels from the source classifier. We then add the following training objective:

$$L_S^{distill} = \frac{1}{m}\sum_{i=1}^{m} L_{kl}\left(h'\left(f(x_i)\right), M_s(x_i)\right) \tag{2}$$

Where L_{kl} is the KL divergence. We call it "distill" because of its close relationship to knowledge distillation. We can then optimize a combined loss:

$$L = \lambda L_{labeled} + (1-\lambda) L_{unlabeled} \tag{3}$$

Extensions and variants:

1. When multiple source models are available, this objective is easy to extend by simply adding a pseudo-label based objective for each source model. This objective can be weighted by how similar a source model is to the target domain (this similarity is described in more detail in the next section). This variant is shown in the figure below.
2. Observe that when no labeled examples are available, we can discard the first term. In this case after training we will obtain a feature representation f that can then be used to learn a classifier for the new domain when labels are available. This is in fact the paradigm we

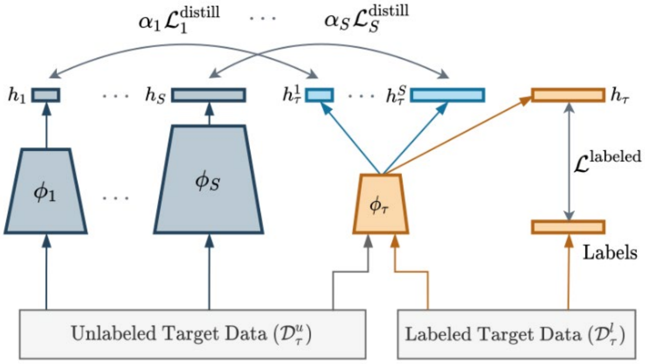

Figure 1: Variant

used for the LwLL evaluation (although in practice the combined objective will likely work better).

3. We also note that our training objective can easily be augmented with additional objectives, such as those from semi- or self-supervised learning. Our system for the evaluation, for example, used an additional SimCLR[21] objective.

Further details about this approach are available in two papers [11,30].

Finding the right pre-trained model:

For the approach described above to work, we need a source model that groups target domain data points similar to the true underlying class labels. How can we identify such a pre-trained model?

If some labels are available, many approaches are available to do this. The simplest class of approaches is as follows. We can use the source model to extract predictions or feature vectors fi for each target data point xi, and compute pairwise correlations between these. We can then compare the resulting matrix of correlations to the correlations between the labels themselves. Depending on the details, this yields many metrics. We found PARC [31] to be most useful. The figure below shows the correlation between this metric and the test accuracy after using the corresponding source model to finetune.

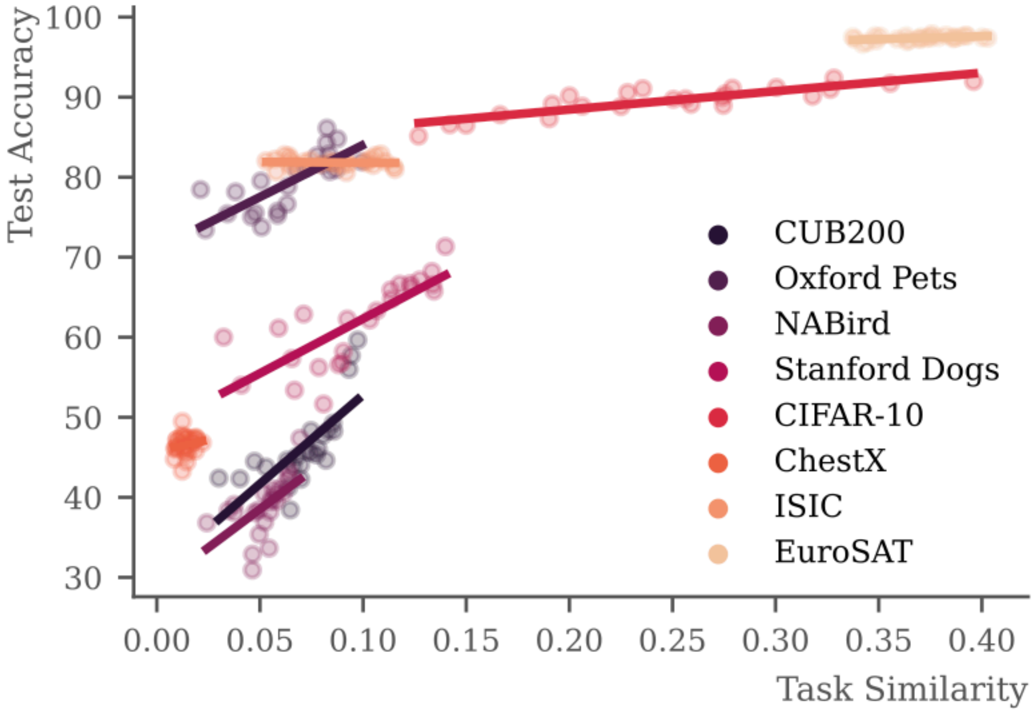

Figure 2: Correlation between this metric and the test accuracy after using the corresponding source model to finetune

For the evaluation, we found that models trained on ImageNet were generally the best for a variety of downstream tasks. This is because of the size and diversity of the data. We therefore skipped this and used ImageNet-trained models as the source. However, we anticipate that as more and more large-scale models become available, this source selection becomes important.

3.2 Incorporating domain knowledge when learning from a few labels

The approaches described above yielded significant improvements in accuracy in the regime of limited labeled data. However, we still couldn't achieve the LwLL goal of reducing the number of labeled examples required by 3 orders of magnitude. The challenge is that our previous approach does not incorporate any information about the domain.

In this section, we look at the broader problem of learning from a few labels when we do have access to information or knowledge about the domain. Here we present three case studies where we used such knowledge to learn recognition systems from very few labels.

Pose invariant recognition

Many objects are deformable and can appear in very different poses. We want a recognition system that is invariant to pose changes. Pose changes typically lead to features move spatially around the image. To incorporate this invariance, we reformulate few-shot classification as a reconstruction problem in latent space. The ability of the network to reconstruct a query feature map from support features of a given class predicts membership of the query in that class. We introduced a novel mechanism for few-shot classification by regressing directly from support features to query features in closed form, without introducing any new modules or large-scale learnable parameters. We found the resulting Feature Map Reconstruction Networks to be both more performant and computationally efficient than previous approaches. We demonstrate consistent and substantial accuracy gains on four fine-grained benchmarks with varying neural architectures. We found this model to also be competitive on the non-fine-grained mini-ImageNet and tiered-ImageNet benchmarks with minimal bells and whistles.

This work is described in [18].

Satellite image analysis

Automatic remote sensing tools can help inform many large-scale challenges such as disaster management, climate change, etc. While a vast amount of spatio-temporal satellite image data is readily available, most of it remains unlabelled. Without labels, this data is not very useful for supervised learning algorithms. Self-supervised learning instead provides a way to learn effective representations for various downstream tasks without labels.

We developed new techniques that leverage characteristics unique to satellite images to learn better self-supervised features. Specifically, we use the temporal signal to contrast images with long-term and short-term differences, and we leverage the fact that satellite images do not change frequently. Using these characteristics, we formulated a new loss contrastive loss called Change-Aware Contrastive (CACo) Loss. Further, we also developed a novel method of sampling different geographical regions. We showed that leveraging these properties leads to better performance on diverse downstream tasks. For example, we saw a 6.5% relative improvement for semantic segmentation and an 8.5% relative improvement for change detection over the best-performing baseline with our method.

This work is described in [32]

Object discovery for self-driving cars

Current 3D object detectors for autonomous driving are almost entirely trained on human-annotated data. Although of high quality, the generation of such data is laborious and costly, restricting them to a few specific locations and object types. We developed an alternative approach entirely based on unlabeled data, which can be collected cheaply and in abundance almost everywhere on earth. Our approach leverages several simple common-sense heuristics to create an initial set of approximate seed labels. For example, relevant traffic participants are generally not persistent across multiple traversals of the same route, do not fly, and are never underground. We found that these seed labels are highly effective to bootstrap a surprisingly accurate detector through repeated self-training without a single human annotated label.

This work is described in [33]

4.0 RESULTS AND DISCUSSION

In this section, we focus on results from our approach for the LwLL evaluation, which is described in Section 3.1.

4.1 LwLL results

Below we present results from our approach on the LwLL evaluation performed by JPL. These are results from all 4 classification tasks we participated in as part of the third phase of evaluation. We focus on the first 4 checkpoints, which represent training with 1, 2, 4 and 8 training examples per class.

For this evaluation, we did not perform intelligent pre-training task selection because of the paucity of diverse pre-training datasets. We consistently used ImageNet for images and Kinetics for video. Each row represents a task on two related datasets. In each row, for the first dataset, we pretrain on ImageNet (or Kinetics for video), while for the second dataset, we use the model trained on the first dataset as initialization. Thus, the pretraining datasets are the same for all plots on the left, but the pretraining dataset for plots on the right change with each task. This is true for both our approach (lavender solid line) and the baseline (black solid line).

Generally, we observe gains on 5/8 datasets, especially in the few shot regime (first four checkpoints; note that for computational reasons, we did not do any training for checkpoints 5 to 7, and as such displayed results for those checkpoints are not an accurate representation of our approach). The gains are especially large on the adaptation datasets (the second dataset in each task) indicating that our approach is particularly good when we pretrain on a related task.

The plots also shows results from all other performers. Note that unlike many other performers, we used the same network architecture as the baseline, which means that all gains come from our novel training strategy; these gains are likely to be higher with more modern network architectures.

These results are representative of what we observed in general in our own experiments, namely that our work yields higher accuracies than baselines in the extreme low-shot setting, as long as the task of interest is close enough to ImageNet, which is our pre-training task.

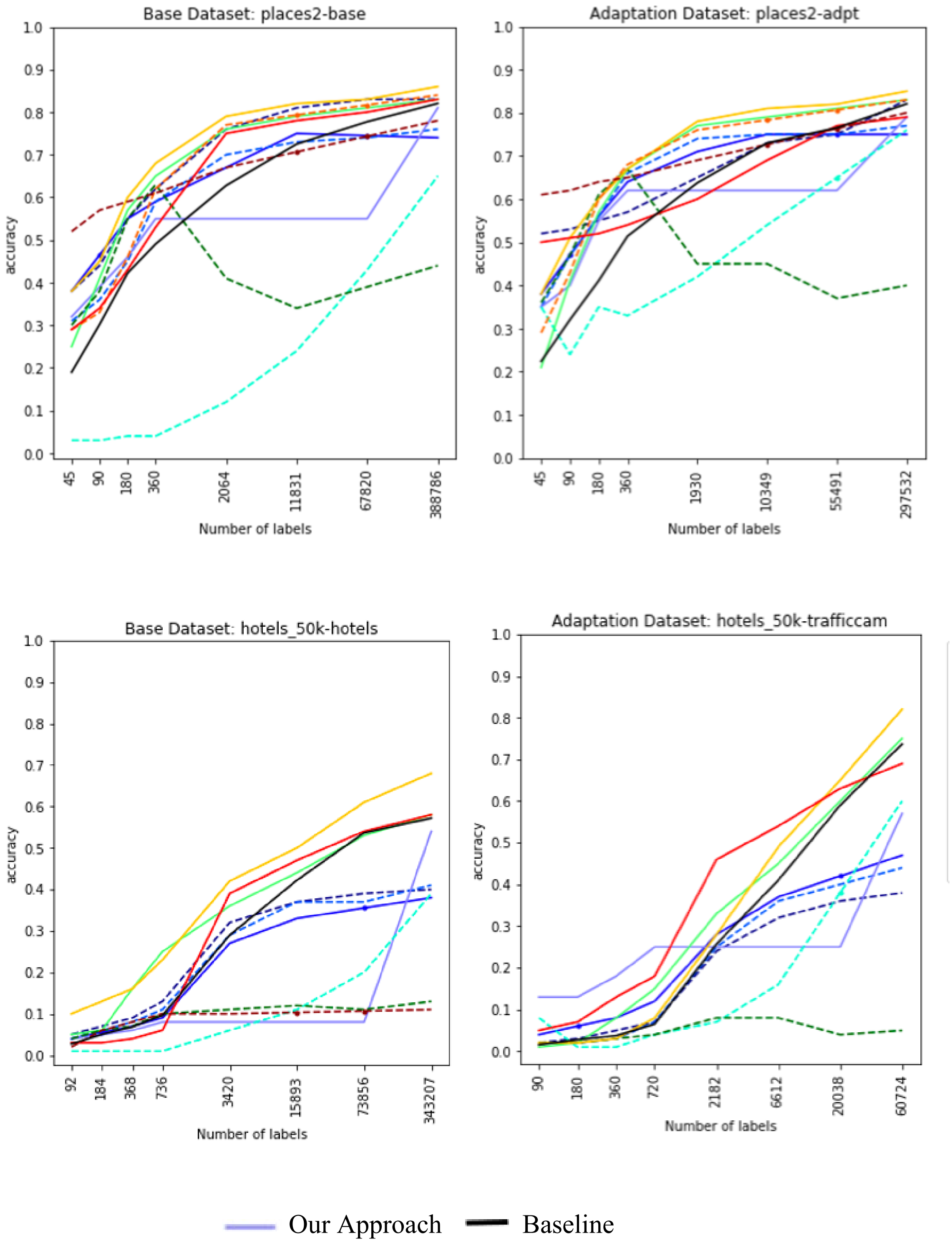

Figure 3: Results from our evaluation for Places, and Hotels

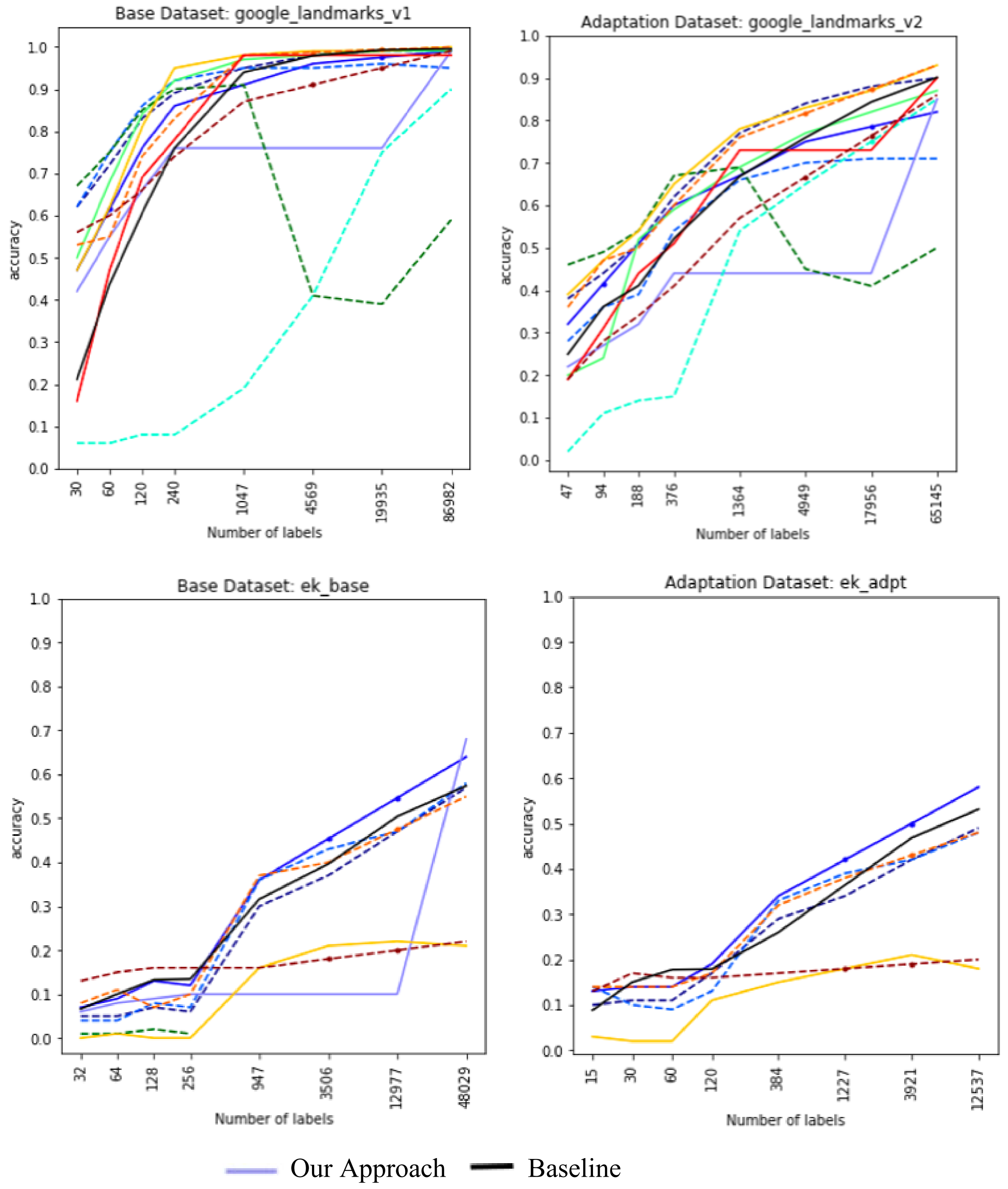

Figure 4: Results from our evaluation for Google Landmarks, and Epic Kitchens

The table below shows our results from our paper with five training examples per class [11]:

Table 1: Results for five training examples per class

	Chest X Rays	*ISIC*	*EuroSat*	*CropDisease*
Naïve transfer	26.7	43.1	80.3	90.1
Self-supervised learning	25.0	36.1	59.0	92.6
STARTUP	26.9	47.2	82.3	93.0

Our approach yields robust and significant improvements across a wide range of domains. Since our approach primarily also allows us to use models that are much smaller and yet yield noticeable improvements as shown in the plot below (here we used source models that are much larger than the target domain model. The figure shows average accuracy across 8 tasks).

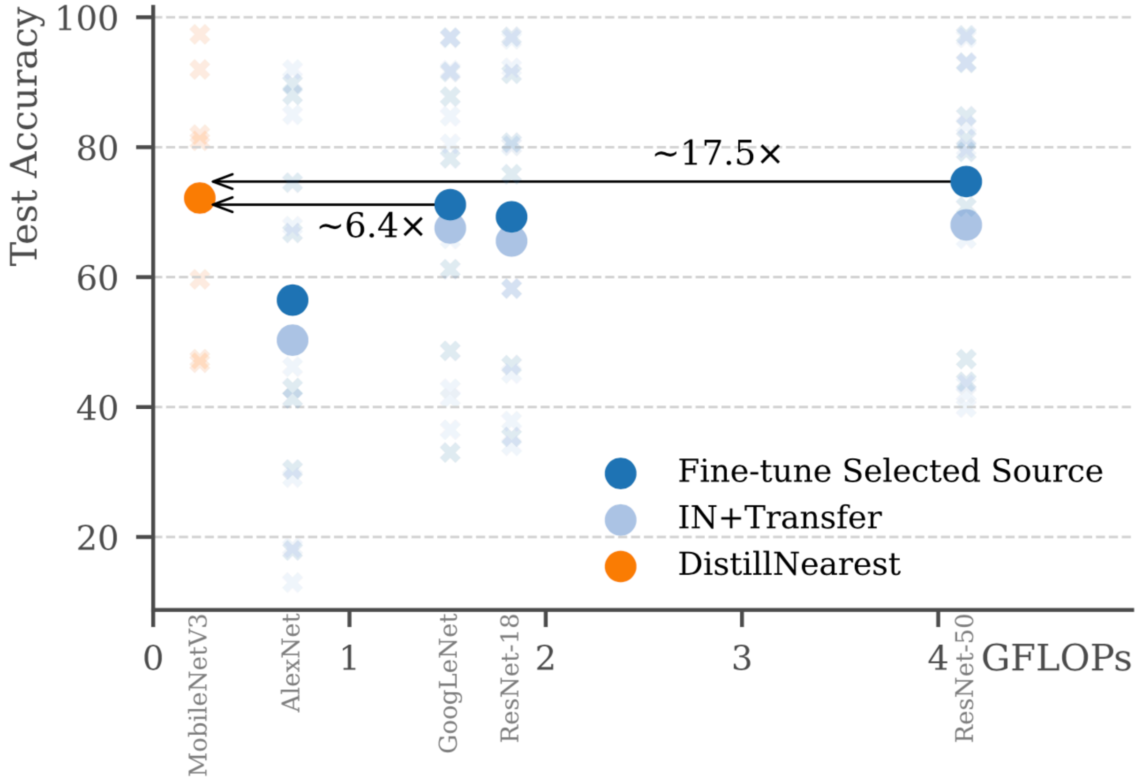

Figure 5: Average Accuracy across 8 tasks

5.0 CONCLUSIONS

We have developed techniques to improve accuracy in the limited data regime that rely on automatically selecting source models that are similar to the target task, and using their predictions as supervisory labels for training. Our approach yields improvements of upto 10 points in the regime of very few training examples, and reduces the number of training examples required for a given accuracy by a factor of 4. In addition, we explored other methods for using domain knowledge to further improve accuracy significantly, even beating supervised technique without a single label.

6.0 REFERENCES

References

1. Wei-Yu Chen, Yen-Cheng Liu, Zsolt Kira, Yu-Chiang Frank Wang, and Jia-Bin Huang. A closer look at few-shot classification. In International Conference on Learning Representations, 2018.
2. Zitian Chen, Yanwei Fu, Yu-Xiong Wang, Lin Ma, Wei Liu, and Martial Hebert. Image deformation meta-networks for one-shot learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 8680-8689, 2019.
3. Mandar Dixit, Roland Kwitt, Marc Niethammer, and Nuno Vasconcelos. Aga: Attribute-guided augmentation. In CVPR, 2017.
4. Li Fei-Fei, Rob Fergus, and Pietro Perona. One-shot learning of object categories. TPAMI, 2006.
5. Chelsea Finn, Pieter Abbeel, and Sergey Levine. Model-agnostic meta-learning for fast adaptation of deep networks. In ICML, 2017.
6. Spyros Gidaris and Nikos Komodakis. Dynamic few-shot visual learning without forgetting. In CVPR, 2018.
7. Bharath Hariharan and Ross Girshick. Low-shot learning by shrinking and hallucinating features. In ICCV, 2017.
8. Brenden M Lake, Ruslan Salakhutdinov, and Joshua B Tenenbaum. Human-level concept learning through probabilistic program induction. Science, 2015.
9. Bo Liu, Mandar Dixit, Roland Kwitt, and Nuno Vasconcelos. Feature space transfer for data augmentation. In CVPR, 2018.
10. Erik G Miller, Nicholas E Matsakis, and Paul A Viola. Learning from one example through shared densities on transforms. In CVPR, 2000.
11. Cheng Perng Phoo and Bharath Hariharan. Self-training for few-shot transfer across extreme task differences. In ICLR, 2021.

12. Jake Snell, Kevin Swersky, and Richard S Zemel. Prototypical networks for few-shot learning. In *NIPS*, 2017.
13. Jong-Chyi Su, Subhransu Maji, and Bharath Hariharan. When does self-supervision improve few-shot learning? In ECCV, 2020.
14. Luming Tang, Davis Wertheimer, and Bharath Hariharan. Revisiting pose-normalization for fine- grained few-shot recognition. In CVPR, 2020.
15. Yonglong Tian, Yue Wang, Dilip Krishnan, Joshua B Tenenbaum, and Phillip Isola. Rethinking few-shot image classification: a good embedding is all you need? In Computer Vision-ECCV 2020: 16th European Conference, Glasgow, UK, August 23-28, 2020, Proceedings, Part XIV 16, pages 266-282. Springer, 2020.
16. Hung-Yu Tseng, Hsin-Ying Lee, Jia-Bin Huang, and Ming-Hsuan Yang. Cross-domain few-shot classification via learned feature-wise transformation. In International Conference on Learning Representations, 2020.
17. Davis Wertheimer and Bharath Hariharan. Few-shot learning with localization in realistic settings. In CVPR, 2019.
18. Davis Wertheimer, Luming Tang, and Bharath Hariharan. Fine-grained few-shot classification with feature map reconstruction networks. In CVPR, 2021.
19. Yu xiong Wang, Ross Girshick, Martial Herbert, and Bharath Hariharan. Low-shot learning from imaginary data. In CVPR, 2018.
20. Mathilde Caron, Ishan Misra, Julien Mairal, Priya Goyal, Piotr Bojanowski, and Armand Joulin. Unsupervised learning of visual features by contrasting cluster assignments. 2020.
21. Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In International conference on machine learning, pages 1597-1607. PMLR, 2020.
22. Hirokatsu Kataoka, Kazushige Okayasu, Asato Matsumoto, Eisuke Yamagata, Ryosuke Yamada, Nakamasa Inoue, Akio Nakamura, and Yutaka Satoh. Pre-training without natural images. In Proceedings of the Asian Conference on Computer Vision, 2020.
23. Alexander Kolesnikov, Xiaohua Zhai, and Lucas Beyer. Revisiting self-supervised visual repre- sentation learning. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pages 1920-1929, 2019.
24. Mehdi Noroozi and Paolo Favaro. Unsupervised learning of visual representations by solving jigsaw puzzles. In European conference on computer vision, pages 69-84. Springer, 2016.
25. Deepak Pathak, Ross Girshick, Piotr Dolla ÃÅr, Trevor Darrell, and Bharath Hariharan. Learning features by watching objects move. In CVPR, 2017.
26. Bram Wallace and Bharath Hariharan. Extending and analyzing self-supervised learning across domains. In ECCV, 2020.
27. Tongzhou Wang and Phillip Isola. Understanding contrastive representation learning through alignment and uniformity on the hypersphere. 119:9929-9939, 13-18 Jul 2020.

28. Kihyuk Sohn, David Berthelot, Nicholas Carlini, Zizhao Zhang, Han Zhang, Colin A Raffel, Ekin Dogus Cubuk, Alexey Kurakin, and Chun-Liang Li. Fixmatch: Simplifying semi-supervised learning with consistency and confidence. Advances in Neural Information Processing Systems, 33, 2020.
29. Bram Wallace, Ziyang Wu, and Bharath Hariharan. Can we characterize tasks without labels or features? In CVPR, 2021.
30. Borup, K., Phoo, C. P., & Hariharan, B. (2023). Distilling from Similar Tasks for Transfer Learning on a Budget. arXiv preprint arXiv:2304.12314.
31. Bolya, D., Mittapalli, R., & Hoffman, J. (2021). Scalable diverse model selection for accessible transfer learning. Advances in Neural Information Processing Systems, 34, 19301-19312.
32. Mall, U., Hariharan, B., & Bala, K. (2023). Change-Aware Sampling and Contrastive Learning for Satellite Images. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 5261-5270).
33. You, Y., Luo, K., Phoo, C. P., Chao, W. L., Sun, W., Hariharan, B., ... & Weinberger, K. Q. (2022). Learning to detect mobile objects from lidar scans without labels. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 1130-1140).

APPENDIX A – PUBLICATIONS AND PRESENTATIONS

List the dates, times, title, event and speakers of any presentations made under this effort and the title author and publication information for any publication made under this effort.

1. **Unsupervised Adaptation from Repeated Traversals for Autonomous Driving**
   Yurong You*, Cheng Perng Phoo*, Katie Z Luo*, Travis Zhang, Wei-Lun Chao, Bharath Hariharan, Mark Campbell, Kilian Q. Weinberger
   In *NeurIPS*, 2022
2. **Learning to Detect Mobile Objects from LiDAR Scans Without Labels**
   Yurong You*, Katie Luo*, Cheng Perng Phoo, Wei-Lun Chao, Wen Sun, Bharath Hariharan, Mark Campbell, Kilian Weinberger
   In *CVPR*, 2022

3. **Coarsely-labeled Data for Better Few-shot Transfer**
   Cheng Perng Phoo, Bharath Hariharan
   In *ICCV*, 2021

4. **Field Guide-inspired Zero-Shot Learning**
   Utkarsh Mall, Bharath Hariharan, Kavita Bala
   In *ICCV*, 2021

5. **Few-Shot Classification with Feature Map Reconstruction Networks**
   Davis Wertheimer, Luming Tang, Bharath Hariharan
   In *CVPR*, 2021

6. **PiCIE: Unsupervised Semantic Segmentation using Invariance and Equivariance in Clustering**
   Jang Hyun Cho, Utkarsh Mall, Kavita Bala, Bharath Hariharan
   In *CVPR*, 2021

7. **Self-training For Few-shot Transfer Across Extreme Task Differences**
   Cheng Perng Phoo, Bharath Hariharan
   In *ICLR*, 2021 (**Oral**)

8. **When Does Self-supervision Improve Few-shot Learning?**
   Jong-Chyi Su, Subhransu Maji, Bharath Hariharan
   In *ECCV*, 2020

9. **Extending and Analyzing Self-Supervised Learning Across Domains**
   Bram Wallace, Bharath Hariharan
   In *ECCV*, 2020

10. **Learning Feature Descriptors using Camera Pose Supervision**
    Qianqian Wang, Xiaowei Zhou, Bharath Hariharan, Noah Snavely
    In *ECCV*, 2020 (**Oral**)

11. **Revisiting Pose-Normalization for Fine-Grained Few-Shot Recognition**
    Luming Tang, Davis Wertheimer, Bharath Hariharan
    In *CVPR*, 2020
12. **Distilling from Similar Tasks for Transfer Learning on a Budget**
    Kenneth Borup, Cheng Perng Phoo, Bharath Hariharan
    In *ICCV*, 2023

LIST OF SYMBOLS, ABBREVIATIONS, AND ACRONYMS

CACo	Change-Aware Contrastive
DARPA	Defense Advanced Research Projects Agency
JPL	Jet Propulsion Laboratory
KL	Kullback Leibler
LwLL	Learning with Less Labels