ProtoKD: Learning from Extremely Scarce Data for Parasite Ova Recognition

Shubham Trehan[1], Udhav Ramachandran[2], Ruth Scimeca[2], and Sathyanarayanan N. Aakur[1]
[1] Department of Computer Science and Software Engineering, Auburn University, Auburn, AL, 36849
[2] Department of Veterinary Pathobiology, Oklahoma State University, Stillwater, OK, 74078

***Abstract*—Developing reliable computational frameworks for early parasite detection, particularly at the ova (or egg) stage, is crucial for advancing healthcare and effectively managing potential public health crises. While deep learning has significantly assisted human workers in various tasks, its application in diagnostics has been constrained by the need for extensive datasets. The ability to learn from an extremely scarce training dataset, i.e., when fewer than 5 examples per class are present, is essential for scaling deep learning models in biomedical applications where large-scale data collection and annotation can be expensive or not possible (in case of novel or unknown infectious agents). In this study, we introduce ProtoKD, one of the first approaches to tackle the problem of multi-class parasitic ova recognition using extremely scarce data. Combining the principles of prototypical networks and self-distillation, we can learn robust representations from only one sample per class. Furthermore, we establish a new benchmark to drive research in this critical direction and validate that the proposed ProtoKD framework achieves state-of-the-art performance. Additionally, we evaluate the framework's generalizability to other downstream tasks by assessing its performance on a large-scale taxonomic profiling task based on metagenomes sequenced from real-world clinical data.**

***Index Terms*—Learning from Extremely Scarce Data, Ova Detection, Microscope Image Analysis**

I. Introduction

Parasitic infections pose a significant threat to human and animal health, often leading to severe illness and even fatalities. These infections can be transmitted through various means, including contaminated food and water sources and common disease vectors such as mosquitoes. For instance, certain zoonotic diseases can be transmitted to humans through the consumption of infected livestock, such as cows and pigs. Many infections, particularly gastrointestinal, can cross the species barrier, and can further amplify the public health risks associated with them. Detecting these parasites early, especially at the ova stage, is crucial for preventing outbreaks of parasitic diseases. Parasitic ova (eggs or cysts) have unique characteristics that allow them to be a distinguishing factor between different kinds of parasitic infections. The typical identification process requires the isolation of the egg from fecal samples collected from the infected host which are subsequently analyzed by a parasitologist under a microscope. Genus-level identification can enable the development of treatments to prevent severe infections and large-scale outbreaks.

Deep learning has enabled the development of "smart" systems that have made immense progress in many fields. However, enormous amounts of historical data (hundreds if not thousands of samples per class) have been necessary to derive insights for downstream applications. Advances in disease diagnostics have been sparse due to the need to learn associations from highly limited, constrained data. Acquiring clinical data to train such deep learning models can be expensive, both in terms of the data acquisition cost and the time and effort of highly skilled human users to provide high-quality, large-scale annotations, which curbs the applications of deep learning models trained under a traditional supervised setting. Transfer learning [1] and few-shot learning [2], [3], [4] have somewhat alleviated this dependency but still require significant amounts of quality data. Generative models such as diffusion models [5] and generative adversarial networks (GANs) [6] show potential in generating synthetic samples for training data augmentation but have been prone to memorizing and replicating training data [7], and prone to hallucinating artifacts [8]. Hence, their application in biomedical settings is inhibited due to privacy concerns [9], where patient identity and data integrity could be compromised.

In this work, we present *ProtoKD*, a framework designed to work with extremely scarce data, e.g., where less than 5 training samples are present per class. Such settings are common in biomedical and diagnostic applications where labeled training data can be expensive. There is a need for rapid learning from a few samples for novel or unseen classes of interest. An illustration of the approach is shown in Figure 1. Our approach was based on the idea that learning robust representations of limited training data and using *domain-specific* augmentations to create an auxiliary dataset that can, together, capture intra-class variation. Through a cyclical, two-phase process, we aim to align representations from original and augmented images to capture variations in the decision boundary across closely related classes. First, a matching loss is introduced to learn robust representations to distinguish *between* classes using a prototypical network. Second, a self-distillation loss is introduced to help capture the intra-class variations in the data by presenting the network with heavily augmented data and training with pseudo-labels generated by the prototypical network. This step has a two-fold effect: (i) it adds a level of regularization that prevents the networks from overfitting, and (ii) it allows us to introduce other learning losses that help discriminate between fine-grained representations. By extending the idea of prototypical networks and self-distillation, we learned robust representations from extremely sparse data that was capable of capturing the intra-class variations through

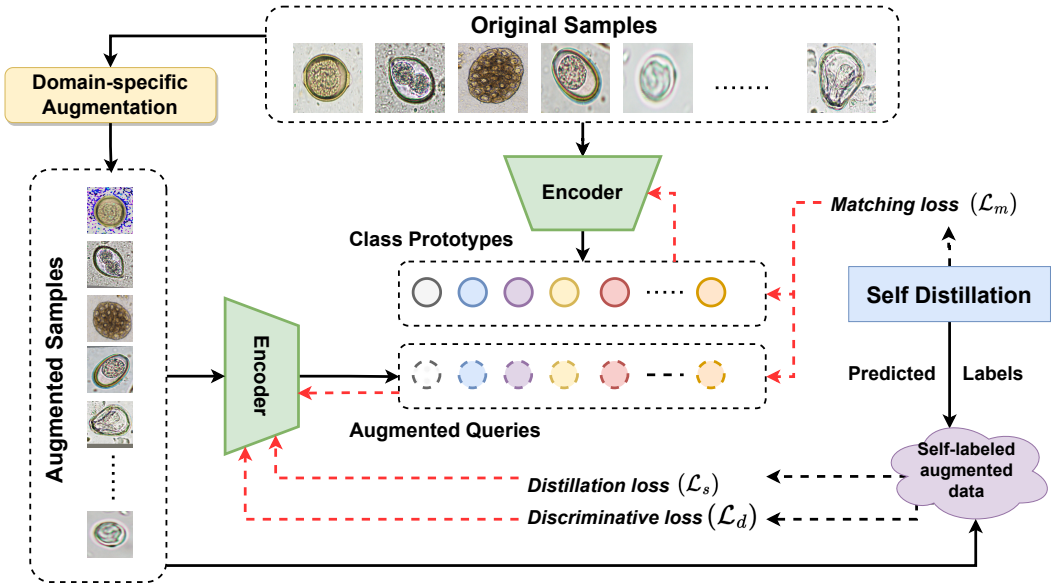

Fig. 1. The **overall architecture** of the proposed ProtoKD is illustrated. Based on prototypical networks and self-distillation, a two-stage process captures the intra-class variations that can occur in biological data in a robust representation.

domain-specific data augmentation.

The **contributions** of our work are three-fold: (i) we present one of the first works to address the problem of multi-class parasitic ova recognition from microscopic images, (ii) we develop a framework to learn from extremely scarce data (from a single example per class) in a multi-class classification setting, and finally (iii) we demonstrate its generalization to other biomedical tasks by evaluating on metagenome profiling.

A. Related Work

There have been very few **automatic parasitic ova detection** frameworks explored in literature. Supervised transfer learning has been explored in detecting and classifying seven species of *Eimeria* spp. in chickens [10] through the analysis of curated large-scale microscopic imagery [11], [12]. Data augmentation techniques, such as image flipping, adding Gaussian noise and histogram normalization, and transfer learning from ImageNet [13] pre-trained models have enabled the training of large deep learning models for ova detection. However, the dependency on large-scale training data was not alleviated, which limits their generalization to other biomedical applications. Weakly supervised approaches such as those based on Multiple Objects Feature Fusion (MOFF) [14] and traditional image processing techniques [15] have been used to reduce the dependency on densely annotated data, yet still assume access to large-scale datasets for learning associations between the input and target classifications. Advances in generative models such as GANs provided a viable mechanism to generate additional training data for **learning from limited data.** DADA [16] explores training with small samples from Cifar-10 and SVNH. They use GANs to augment the original dataset for more diversity with the same labels. Barz *et al.* [17] showed that the Cosine Loss is better than categorical cross-entropy whenever only a few samples are present per class. Ishikawa *et al.* [18] use conditional GAN for augmentation to improve efficiency in generating training samples but increase the computational cost. Brigato *et al.* [19] use Auxiliary-Classifier GANs for image synthesis and classification in low data settings. Meta-learning approaches such as MAML [2] and Prototypical Networks [4], [20] have enabled *few-shot* learning where only a few samples per class are required for making inferences. However, such or similar approaches [20], [21], [22] assume a reasonably large training corpus exists to create "meta-tasks" for learning robust representations for downstream classification.

II. ProtoKD: Learning from Scarce Data

Problem Statement. In this work, we consider the task of learning from extremely scarce samples $n \leq 5$ per class. During training, the model can access a set of N training examples $X=\{x_1, x_2, x_3, \ldots x_N\}$ drawn from C classes with n samples each. The model is presented with samples from any of the C classes at test time. In contrast, meta-learning approaches [4], [3], [2] consider the training and evaluation phases to consist of c-way classification tasks, where $c \ll C$. Our setup is more challenging since we have a C-way classification task and have access to extremely scarce data.

1) Prototypical Networks for Scarce Data: Prototypical networks aim to construct a D-dimensional vector representation of each class, called the *prototype*, that captures its underlying characteristics. Ideally, each prototype p_j represents the typical example from that class and captures the intra-class variations through an embedding function $f_\theta : x_i \longrightarrow F_i$ that projects a sample x_i to a D-dimensional vector F_i. The prototype is computed as the mean representation of all n training samples in the class $c_j \in C$. In a typical meta-

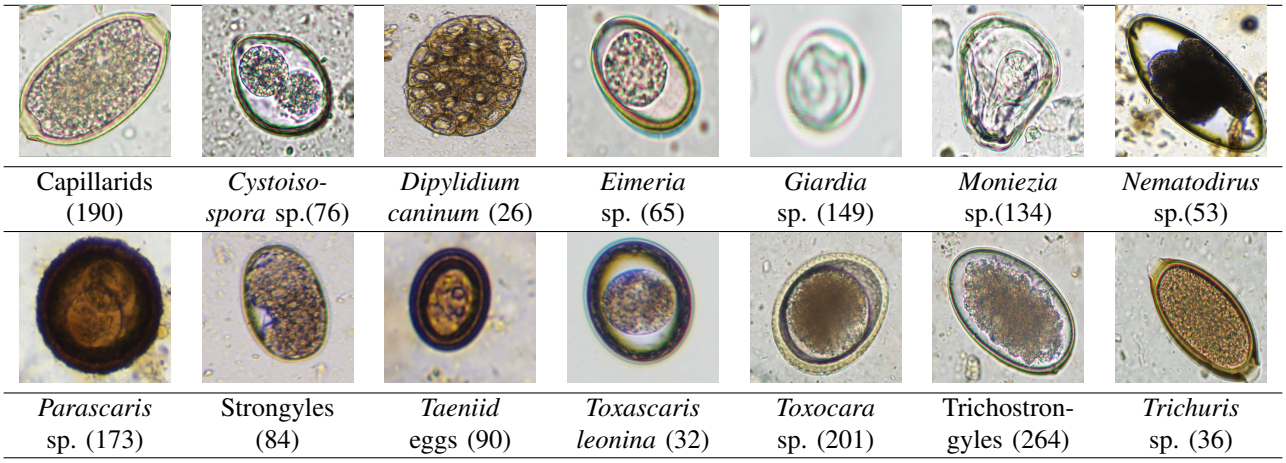

Fig. 2. **Dataset Characteristics.** Our parasite ova dataset has 1573 instances across 14 ova classes, imaged from 594 clinical samples. An exemplar image from each class and the number of samples in each class are presented. It is highly imbalanced, with high inter- and intra-class variation.

learning setup, a *support* set S consisting of k examples from c classes is sampled to construct these prototypes. A function ϕ provides the distance between each example in the *query* set Q and each prototype. A softmax function over these distances provides a probability distribution to label each example $x_i \in Q$. This setup assumes that (i) each sample in the query comes from the c classes sampled in the episode and (ii) there exists a reasonably large dataset to sample the support and query sets during training to learn good representations for c-way classification. However, in biomedical applications, the number of training samples can be sparse due to the high data acquisition cost or the need to adapt rapidly to a changing scenario. Similarly, assuming a smaller subset of classes at inference is unrealistic since this requires prior knowledge about each test sample.

We extend the prototypical network formulation to the extremely scarce data regime to overcome these limitations by proposing to learn the embedding function f_θ through domain-specific data augmentation. Given a support set S_i, each example x_i is augmented through a controlled, predetermined scheme to create k samples $\{x'_1, x'_2, x'_3, \dots x'_k\}$ that become part of the query set. Hence, each minibatch consists of examples from *all* C classes, with the prototypes constructed from original, uncorrupted samples and the query samples providing samples that cover the plausible intra-class variations for each class. The training objective is to minimize the negative log probability of each query sample x'_i belonging to its true class c_j represented by its prototype p_j. Hence, we minimize the matching loss ($\mathcal{L}_m$) given by

$$-logp(y=j|x'_i,\theta) = -log\left(\frac{-exp(\phi(f_\theta(x'_i,p_j)))}{\sum_k^C exp(\phi(f_\theta(x'_i,p_k)))}\right) \quad (1)$$

where x'_i is an example from the query set Q; $\phi(\cdot)$ is a distance function that provides the distance (in the range $[0,+\infty)$) between each sample x'_i and a prototype p_j; and f_θ is an embedding function, defined as a Wide ResNet [23], and $\phi(\cdot)$ is the Euclidean distance. The data augmentation scheme is specific to each use case and must be designed based on prior, domain-specific knowledge. In our experiments with the parasite ova data, we apply augmentation mechanisms such as random zoom, rotation, contrast change, flip, shear, and solarization. However, these are not plausible augmentation schemes for the genome data to capture the intra-class variations. Hence, we design genome-specific augmentation schemes based on base flipping to simulate observation error [24]. We add noise drawn from a normal distribution (with 0 mean and variance of 1) to 5% of the pixels randomly selected symmetrically along the diagonal. This augmentation mechanism mimics the observation errors commonly found in genome sequencing [24], [25] and provides a natural augmentation scheme to capture the intra-class variation and additionally helps preserve the symmetry of the pseudo-image-based k-mer representations. We refer the reader to [26] for more details on the pseudo-imaging for genomics.

2) Learning Intra-class Variations with Self-Distillation: The second step is to enhance the prototypes $P=\{p_1, p_2, p_3, \dots p_C\}$ to capture the intra-class variations. First, we generate *pseudo-labels* for each query sample $x'_i \in Q$ using the probability distribution defined in Equation 1 to create labels for the augmented samples. This *self-labeled* data is then used to fine-tune the encoder network using a distillation loss given by

$$\mathcal{L}_s = D_{KL}(p_t(y=j|x_i,\theta_k)/\tau, p_s(y=j|x_i,\theta_s)) \quad (2)$$

where $D_{KL}(\cdot)$ is the Kullback–Leibler divergence [27] between the probability distributions of the prototypical network defined in Section II-1 and a linear layer trained on top of representations F_i from f_θ; τ is a temperature parameter that controls how closely the linear layer's output should match that of the prototypical network; and θ_t and θ_s are the trainable parameters of the prototypical network (i.e., the embedding function f_θ) and the linear layer, respectively. We set τ to 5, chosen based on a grid search between 0 and 10. This formulation could be extended to leverage unlabeled data into a semi-supervised learning setting in future works. We leave

TABLE I

QUANTITATIVE RESULTS ON THE PARASITE OVA DATASET. ONE EXAMPLE FROM EACH CLASS WAS RANDOMLY SAMPLED FOR TRAINING. AVERAGE PERFORMANCE FROM 10 RUNS IS REPORTED.

	Supervised		ProtoNet		ProtoKD	
	Precision	**Recall**	**Precision**	**Recall**	**Precision**	**Recall**
Average	0.436	0.448	0.472	0.494	**0.523**	**0.544**
Capillarids	0.558	0.468	0.554	**0.680**	**0.730**	0.460
***Cystoisospora* spp.**	0.166	0.140	0.174	0.188	**0.360**	**0.260**
Dipylidium caninum	0.116	0.424	**0.236**	**0.448**	0.080	0.180
***Eimeria* spp.**	0.292	0.152	0.120	0.222	**0.480**	**0.290**
***Giardia* spp.**	0.908	0.978	0.934	0.918	**0.945**	**1.000**
***Moniezia* spp.**	**0.528**	**0.580**	0.492	0.528	0.515	0.575
***Nematodirus* spp.**	0.272	0.414	**0.352**	0.464	0.250	**0.790**
***Parascaris* spp.**	0.768	0.746	**0.858**	0.716	0.755	**0.785**
Strongyles	0.268	0.464	0.362	0.484	**0.565**	**0.650**
***Taeniid* spp.**	0.574	0.334	0.650	0.320	**0.785**	**0.715**
Toxascaris leonina	0.226	0.670	**0.342**	**0.834**	0.255	0.630
***Toxocara* spp.**	0.738	0.438	0.790	0.592	**0.795**	**0.610**
Trichostrongyles	0.562	0.174	0.560	0.180	**0.665**	**0.265**
***Trichuris* spp.**	0.150	0.300	**0.184**	0.322	0.140	**0.400**

that exploration to future work and focus only on the extremely scarce data setting.

In addition to the distillation loss described above, we introduce a simple discriminative loss to help increase the separability of the decision boundary between the classes. To this end, we employ a similarity-based contrastive loss that reduces the difference between the features of each query sample x'_i and its corresponding prototype p_j while increasing the distance to other prototypes. The intuition behind this loss is that increasing the distance between the query sample and other prototypes can capture the variability within each class since we widen its decision boundary based on its prototype p_j. We define this to be a discriminative loss given by

$$\mathcal{L}_d = -log\left(\frac{exp(p_j^T \cdot F_i)}{\sum_{k=1}^{C} \mathbb{1}_{j\neq k} exp(p_k^T \cdot F_i)}\right) \tag{3}$$

where F_i is the feature representation of a query sample x'_i; $\mathbb{1} \in \{0, 1\}$ indicates whether the sample is from its true class c_j; and both p_j and F_i are ℓ_2 normalized vectors. This formulation allows us to leverage our proposed domain-specific augmentation setup for a contrastive learning mechanism without having to sample new positive and anchor examples or triplet mining, as with other contrastive learning mechanisms such as SimCLR [28] or triplet losses [29].

Implementation Details. The framework is trained end-to-end in two alternating phases. Every epoch consists of 10 iterations of training only with the matching loss, followed by 10 iterations of training using both distillation and discriminative losses. The matching loss cycle allows us to build prototypes of each class, while the self-distillation phase allows us to refine the prototypes by capturing the intra-class variation explicitly. We use a Wide-ResNet [23], trained from scratch, as the backbone for our mapping function, with a depth of 28 layers, a width of 2, and a dropout rate of 0.3. The features are projected to a 128-dimension vector using a linear layer. The images are resized to 128×128 and pre-processed as done in ResNet [30]. All networks are trained for 100 epochs with a support set size of 1 and a query set of 5 and converge in 90 minutes. All experiments were conducted on a workstation

TABLE II

ABLATION STUDIES TO DETERMINE THE IMPACT OF EACH COMPONENT ON THE PERFORMANCE OF PROTOKD ON THE PARASITE OVA RECOGNITION TASK.

$\mathcal{L}_m$	$\mathcal{L}_s$	$\mathcal{L}_d$	**Precision**	**Recall**	**F1-Score**
✓	✗	✗	0.472	0.494	0.483
✗	✓	✗	0.436	0.448	0.442
✓	✓	✗	0.497	0.501	0.499
✓	✗	✓	0.508	0.516	0.512
✓	✓	✓	**0.523**	**0.544**	**0.533**

server with an AMD ThreadRipper CPU with 64 cores, 128 GB RAM, and an NVIDIA Titan RTX (24GB).

III. EXPERIMENTAL EVALUATION

In this section, we present the experimental setup and evaluation results for the proposed ProtoKD approach. We begin by describing the data collection process to curate the parasitic ova recognition dataset, one of the first attempts at a comprehensive benchmark for the task. We discuss the metrics and baselines used for evaluation and present the quantitative results. We conclude by demonstrating the generalization capabilities of the proposed ProtoKD framework to other biomedical applications, such as genome classification.

Data Collection. We collected 1573 ova examples, curated from 594 clinical samples at Oklahoma State University's Veterinary Diagnostics laboratory. Each type of egg was collected by performing a centrifugal fecal flotation on samples from various hosts infected with different parasites and were subsequently observed under a microscope by a certified parasitologist. Following identification, images were captured using the Olympus BX43 microscope (Tokyo, Japan) and the Olympus Cell Sens Entry software v1.18. Based on their frequency of occurrence in samples received at both local and national diagnostics labs, the following 14 parasitic ova were considered: Capillarids, *Cystoisospora* spp., *Dipylidium caninum*, *Eimeria* spp., *Giardia* spp., *Moniezia* spp., *Nematodirus* spp., *Parascaris* spp., Strongyles, *Taeniid eggs*, *Toxascaris leonina*, *Toxocara* spp., Trichostrongyles, and *Trichuris* spp. Figure 2 provides examples and statistics. Each parasite and its relative size determined where the magnification would be 40x, 20x, or 10x objective. Most of the images captured are at 10x objective magnification which is available on a large majority of microscopes. Clinical samples were collected until each class had at least 25 examples to provide a comprehensive benchmark for evaluating machine learning frameworks for ova recognition with extremely scarce data.

Metrics and Baselines. Due to the highly imbalanced nature of the data, we choose precision and recall per class as our evaluation metric since accuracy can be highly skewed towards classes with more examples. We evaluate the performance of each algorithm with a training set with 1 example per class and report the mean results from 10 random trials to avoid conflating the results due to the choice of the example from each class. We choose a fully supervised Wide-ResNet as our backbone network for all baselines, including a fully

TABLE III
EVALUATION ON METAGENOME CLASSIFICATION. HOST AND AVERAGE PATHOGEN F1-SCORES ARE USED FOLLOWING PRIOR WORK [26].

# Samples	MG-NET		ProtoNet		ProtoKD	
	Host F1	Pathogen F1	Host F1	Pathogen F1	Host F1	Pathogen F1
1	0.230	0.026	0.596	0.108	**0.699**	**0.121**
5	0.330	0.029	0.690	0.127	**0.804**	**0.139**
10	0.320	0.035	**0.840**	0.081	0.780	**0.129**
15	0.360	0.047	**0.850**	0.109	0.824	**0.127**
20	0.370	0.051	0.780	0.147	**0.799**	**0.151**
25	0.371	0.056	0.720	**0.200**	**0.775**	0.193

supervised one, the modified ProtoNet from Section II-1, and ProtoKD. All hyperparameters for each baseline are kept constant for each trial and trained for 100 epochs with early stopping based on a validation set of 5 samples per class.

Performance on Parasite Ova Data. We first evaluate and compare our approach against baselines on the parasite ova dataset. Table I summarizes the results when training with only one example per class. The proposed ProtoKD approach performs well across classes, with an average precision of 0.523 and recall of 0.544, outperforming the supervised and ProtoNet baselines. Of particular interest is the performance of the baselines on the two classes with the least intra-class variation (*Giardia* spp.) and the highest intra-class variation (*Nematodirus* spp.). With *Nematodirus* spp., ProtoKD's recall (0.790) was almost 1.5 times that of ProtoNet (0.464) and the supervised (0.414) baselines, which indicates that the self-distillation and discriminative losses played their role in learning robust prototypes. All three baselines performed well on Giardia, with ProtoKD achieving 100% recall with a high precision of 0.945. On average, we find that the ProtoKD achieves high recall at the cost of precision, particularly in the case of classes with a large number of samples, such as *Parascaris* spp. and *Trichuris* spp. We hypothesize that this is an effect of the contrastive, discriminative loss intended to expand each class's decision boundary. Interestingly, the ProtoKD severely fails on *Dipylidium caninum*. Upon close inspection, *Dipylidium caninum* was highly confused with Tirchostrongyle eggs, which, though visually similar (see Figure 2), are functionally different. The strong augmentation scheme resulted in overlapping decision boundaries due to the extreme variation in Trichostrongyle examples. We anticipate using super-resolution mechanisms [31], [32] to enhance the images will help find better representations.

Ablation Studies. We perform ablation studies to systematically evaluate the impact of the different components of the approach on its performance. The three loss functions, defined in Equations 1, 2, and 3, are the major components of the approach. Hence, we ablate over the impact of using $\mathcal{L}_s$ and $\mathcal{L}_d$ in combination with $\mathcal{L}_m$ and report results in Table II. Using only the matching loss (Equation 1), the approach degenerates to a standard ProtoNet, one of our baselines (Row 1). When τ in $\mathcal{L}_s$ (defined in Equation 2) is set to 1, it becomes the standard cross-entropy loss and hence is equivalent to our fully supervised baseline (Row 2). It can be seen that using knowledge distillation loss ($\mathcal{L}_c$) alone or matching loss ($\mathcal{L}_m$) performs reasonably well, although not as much as the proposed ProtoKD. Adding the discriminative loss ($\mathcal{L}_d$) with the matching loss ($\mathcal{L}_m$) provides a higher increase in performance than combining the matching loss ($\mathcal{L}_m$) with self-distillation ($\mathcal{L}_s$). Combining all three provides a higher increase overall, indicating the subtle balance between learning inter-class and intra-class variations provided by the alternating training methodology proposed in ProtoKD. Note that this formulation can naturally be extended to semi-supervised learning where the self-distillation loss ($\mathcal{L}_m$) can be used to train on unlabeled data. We leave that to future work since our focus is on tackling the problem of learning from scarcely available (< 5 samples per class) training data.

Extension to Other Biomedical Applications. In addition to our experiments on parasite ova recognition, we evaluate the generalizability of the proposed ProtoKD formalism to other biomedical applications by evaluating its ability to learn representations from an entirely separate application: metagenome sequences. We evaluate the ProtoKD framework on the data provided by MG-NET, which has $31,580$ sequence reads across seven classes - Bovine (host), *B. trehalosi*, *H. somni*, *M. bovis*, *M. haemolytica*, *P. multocida* and *T. pyogenes*. Specifically, we use the pseudo-images generated by the MG-NET framework as input and evaluate it by training with varying samples per class. The test set was fixed with 8192 samples for a fair comparison with MG-NET in all the scenarios. Average performance from 10 trials is reported. Table III summarizes the result. We can see that the proposed ProtoKD framework and ProtoNet outperform the supervised MG-NET at very low samples, i.e., less than 25 samples per class. It takes MG-NET at least 500 samples per class to outperform ProtoKD with 25 samples, achieving an overall host F1-score of 0.906 and an average pathogen F1-score of 0.319. Interestingly, the performance initially reduces as the number of samples per class is increased. We attribute this phenomenon to the fact that fine-grained recognition requires *highly distinct* samples for learning robust features. Genomes between closely related species have shown to have similar genome sequences [33]; hence, larger amounts of data do not necessarily translate into better performance. For example, ProtoNet achieves a higher pathogen F1-score (0.200) than ProtoKD (0.193) at 25 samples. It has 100% precision and 0% recall on two pathogen classes, indicating that the model does not make balanced predictions and fails on edge cases. We anticipate that including structural information [26], [33] and other metadata will improve the performance.

IV. CONCLUSION AND FUTURE WORK

In this work, we presented *ProtoKD*, one of the first works to tackle the problem of learning from extremely scarce training samples. Using a benchmark dataset of parasitic ova, we demonstrate its strong ability to learn robust representations from just one example per class. Experiments on large-scale metagenome-based taxonomic profiling data demonstrated its generalizability to other downstream applications. We anticipate using super-resolution to enhance the images will help find better representations. We aim to extend this framework

for scaling deep learning frameworks to work with highly constrained data typical in biomedical applications such as disease diagnostics and the Internet of Medical Things.

Acknowledgement. This work was partially supported by the US National Science Foundation (NSF) grant IIS 1955230.

REFERENCES

[1] M. Ghafoorian, A. Mehrtash, T. Kapur, N. Karssemeijer, E. Marchiori, M. Pesteie, C. R. Guttmann, F.-E. de Leeuw, C. M. Tempany, B. Van Ginneken *et al.*, "Transfer learning for domain adaptation in mri: Application in brain lesion segmentation," in *Medical Image Computing and Computer Assisted Intervention- MICCAI 2017: 20th International Conference, Quebec City, QC, Canada, September 11-13, 2017, Proceedings, Part III 20*. Springer, 2017, pp. 516–524.

[2] C. Finn, P. Abbeel, and S. Levine, "Model-agnostic meta-learning for fast adaptation of deep networks," in *International conference on machine learning*. PMLR, 2017, pp. 1126–1135.

[3] O. Vinyals, C. Blundell, T. Lillicrap, k. kavukcuoglu, and D. Wierstra, "Matching networks for one shot learning," in *Advances in Neural Information Processing Systems*, D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett, Eds., vol. 29. Curran Associates, Inc., 2016. [Online]. Available: https://proceedings.neurips.cc/paper/2016/file/90e1357833654983612fb05e3ec9148c-Paper.pdf

[4] J. Snell, K. Swersky, and R. S. Zemel, "Prototypical networks for few-shot learning," 2017. [Online]. Available: https://arxiv.org/abs/1703.05175

[5] J. Ho, C. Saharia, W. Chan, D. J. Fleet, M. Norouzi, and T. Salimans, "Cascaded diffusion models for high fidelity image generation." *J. Mach. Learn. Res.*, vol. 23, no. 47, pp. 1–33, 2022.

[6] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," *Communications of the ACM*, vol. 63, no. 11, pp. 139–144, 2020.

[7] C. A. Corneanu, S. Escalera, and A. M. Martinez, "Computing the testing error without a testing set," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 2677–2685.

[8] J. P. Cohen, M. Luck, and S. Honari, "Distribution matching losses can hallucinate features in medical image translation," in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part I*. Springer, 2018, pp. 529–536.

[9] M. Fredrikson, S. Jha, and T. Ristenpart, "Model inversion attacks that exploit confidence information and basic countermeasures," in *Proceedings of the 22nd ACM SIGSAC conference on computer and communications security*, 2015, pp. 1322–1333.

[10] P. He, Z. Chen, Y. He, J. Chen, K. Hayat, J. Pan, and H. Lin, "A reliable and low-cost deep learning model integrating convolutional neural network and transformer structure for fine-grained classification of chicken eimeria species," *Poultry Science*, vol. 102, no. 3, p. 102459, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0032579122007532

[11] K. Ray, S. Saharia, and N. Sarma, "Detection and identification of ascaris lumbricoides and necator americanus eggs in microscopic images of faecal samples of pigs," *International Journal of Automation and Control*, vol. 15, p. 378, 01 2021.

[12] Y. Wang, Z. He, S. Huang, and H. Du, "A robust ensemble model for patasitic egg detection and classification," 2022. [Online]. Available: https://arxiv.org/abs/2207.01419

[13] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein *et al.*, "Imagenet large scale visual recognition challenge," *International journal of computer vision*, vol. 115, pp. 211–252, 2015.

[14] P. Manescu, C. Bendkowski, R. Claveau, M. Elmi, B. J. Brown, V. Pawar, M. J. Shaw, and D. Fernandez-Reyes, "A weakly supervised deep learning approach for detecting malaria and sickle cells in blood films," in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part V 23*. Springer, 2020, pp. 226–235.

[15] J. E. Arco, J. M. Górriz, J. Ramírez, I. Álvarez, and C. G. Puntonet, "Digital image analysis for automatic enumeration of malaria parasites using morphological operations," *Expert Systems with Applications*, vol. 42, no. 6, pp. 3041–3047, 2015.

[16] X. Zhang, Z. Wang, D. Liu, and Q. Ling, "Dada: Deep adversarial data augmentation for extremely low data regime classification," in *IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2019, pp. 2807–2811.

[17] B. Barz and J. Denzler, "Deep learning on small datasets without pre-training using cosine loss," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2020, pp. 1371–1380.

[18] T. Ishikawa and S. Stent, "Boosting supervised learning in small data regimes with conditional gan augmentation," in *2022 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2022, pp. 1351–1355.

[19] A. Dravid, F. Schiffers, Y. Wu, O. Cossairt, and A. K. Katsaggelos, "Investigating the potential of auxiliary-classifier gans for image classification in low data regimes," in *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2022, pp. 3318–3322.

[20] J. Li, P. Zhou, C. Xiong, and S. C. Hoi, "Prototypical contrastive learning of unsupervised representations," *arXiv preprint arXiv:2005.04966*, 2020.

[21] F. Sung, Y. Yang, L. Zhang, T. Xiang, P. H. Torr, and T. M. Hospedales, "Learning to compare: Relation network for few-shot learning," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 1199–1208.

[22] C.-Y. Chuang, J. Robinson, Y.-C. Lin, A. Torralba, and S. Jegelka, "Debiased contrastive learning," *Advances in Neural Information Processing Systems*, vol. 33, pp. 8765–8775, 2020.

[23] S. Zagoruyko and N. Komodakis, "Wide residual networks," in *British Machine Vision Conference 2016*. British Machine Vision Association, 2016.

[24] T. Laver, J. Harrison, P. O'neill, K. Moore, A. Farbos, K. Paszkiewicz, and D. J. Studholme, "Assessing the performance of the oxford nanopore technologies minion," *Biomolecular Detection and Quantification*, vol. 3, pp. 1–8, 2015.

[25] V. Indla, V. Indla, S. Narayanan, A. Ramachandran, A. Bagavathi, V. L. Ramnath, and S. N. Aakur, "Sim2real for metagenomes: Accelerating animal diagnostics with adversarial co-training," in *Pacific-Asia Conference on Knowledge Discovery and Data Mining*. Springer, 2021, pp. 164–175.

[26] S. N. Aakur, S. Narayanan, V. Indla, A. Bagavathi, V. Laguduva Ramnath, and A. Ramachandran, "Mg-net: leveraging pseudo-imaging for multi-modal metagenome analysis," in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part V 24*. Springer, 2021, pp. 592–602.

[27] J. M. Joyce, "Kullback-leibler divergence," in *International encyclopedia of statistical science*. Springer, 2011, pp. 720–722.

[28] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A simple framework for contrastive learning of visual representations," in *International conference on machine learning*. PMLR, 2020, pp. 1597–1607.

[29] W. Ge, "Deep metric learning with hierarchical triplet loss," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 269–285.

[30] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.

[31] B. Liao, Y. Chen, Z. Wang, C. D. Smith, and J. Liu, "Comparative study on 1.5 t-3t mri conversion through deep neural network models," in *2022 21st IEEE International Conference on Machine Learning and Applications (ICMLA)*. IEEE, 2022, pp. 963–968.

[32] Q. Zhu, P. Li, and Q. Li, "Attention retractable frequency fusion transformer for image super resolution," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 1756–1763.

[33] S. N. Aakur, V. Indla, V. Indla, S. Narayanan, A. Bagavathi, V. L. Ramnath, and A. Ramachandran, "Metagenome2vec: Building contextualized representations for scalable metagenome analysis," in *2021 International Conference on Data Mining Workshops (ICDMW)*. IEEE, 2021, pp. 500–507.