Prototypical Networks for Few shot Learning

Aratrika Das (Student Number: 5298121)[1], Kenzo Boudier (Student Number: 5333024)[1], Swetha Balaram (Student Number: 6019552)[1], and Yash Mundhra (Student Number: 5017874)[1]

[1]Faculty of Electrical Engineering, Mathematics and Computer Science, TU Delft

April 2024 - CS4240 Deep Learning

1 Introduction

1.1 Background

In the realm of machine learning, the ability of models to generalize from limited data is a fundamental challenge. Traditional supervised learning paradigms typically rely on vast amounts of labelled data to train models effectively. However, in many real-world scenarios, obtaining such extensive datasets may be infeasible due to factors like data scarcity, high annotation costs, or the need for domain expertise.

Few-shot learning represents a paradigm shift in addressing this challenge by focusing on training models with only a small number of labelled examples per class. This approach mirrors the human capacity to learn new concepts with minimal exposure, making it particularly valuable in domains where labelled data is scarce or costly to obtain.

In few-shot learning, the goal is to equip models with the ability to generalize from a few examples of each class to accurately classify previously unseen instances. This requires models to learn robust representations of classes and effectively leverage the available information to make accurate predictions.

1.2 Introduction to the Paper

The paper "Prototypical Networks for Few-shot Learning" by Jake Snell, Kevin Swersky, and Richard S. Zemel presents a novel approach to address the few-shot classification problem [4]. At its core, the paper introduces the concept of prototypical networks, which leverage the idea of computing distances in a metric space to prototype representations of each class.

Prototypical networks offer a principled framework for few-shot learning by learning a metric space in which classification can be performed effectively. By computing distances to prototype representations of each class, the model can generalize from a small number of examples per class to accurately classify new instances.

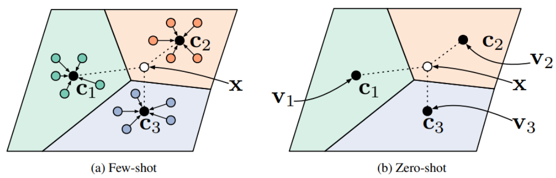

Figure 1: Prototypical networks using few-shot and zero-shot learning respectively [4].

1.3 Summary of Key Concepts and Methodology

The paper proposes prototypical networks as a solution for few-shot learning tasks, where a classifier must adapt to accommodate new classes with limited data. Key concepts and methodologies presented in the paper include:

- **Prototype Representation:** Prototypical networks compute prototype representations for each class by averaging embedded support examples.
- **Distance Metric:** Classification is performed by computing Euclidean distances in a metric space to prototype representations of each class.
- **Training Procedure:** The paper outlines a training procedure for prototypical networks, including the computation of prototype representations and the classification process.
- **Evaluation:** The efficacy of prototypical networks is evaluated on benchmark datasets, demonstrating state-of-the-art performance in few-shot classification tasks.

In the subsequent sections of this blog post, we will delve into the reproduction of the paper's results, conduct additional experiments and assess their effects on few-shot learning.

2 Experiments With Omniglot Dataset

The first dataset that is used in the paper to experiment with is the Omniglot dataset [2]. The Omniglot dataset contains 1623 different handwritten characters from 50 different alphabets. The Omniglot dataset used in the paper was pre-processed using the procedure of Vinyals et al. [5] "resizing the grayscale images to 28×28 and augmenting the character classes with rotations in multiples of 90 degrees".

To evaluate the performance of the prototypical networks approach on the Omniglot dataset the authors used "Euclidean distance in the 1-shot and 5-shot scenarios with training episodes containing 60 classes and 5 query points per class" [4]. In our attempt at verifying the results from the paper we decided to utilize an existing implementations which we found on a public GitHub repositories [1].

By using the implementation from this GitHub repository and adhering to the training parameters specified in the original paper, such as the number of epochs and the size of each epoch we managed to reproduce the results of the paper quite closely. These outcomes not only validate the reproducibility of the prototypical networks but also reinforce the robustness of this approach in the field of few-shot learning.

Table 1 compares the differences between the original results from the paper and the reproduced results for the Omniglot dataset.

	5-way accuracy		20-way accuracy	
	1-shot	5-shot	1-shot	5-shot
Paper Results	98.8%	99.7%	96%	98.9%
Reproduced Results	98.5%	99.54%	94.77%	98.53%

Table 1: Comparison of reproduced result and paper results for the **Omniglot** dataset

Figure 2 presents the comparison of the accuracy in a bar graph.

3 Experiments With Mini_ImageNet Dataset

The experiments conducted with the prototypical networks on the Mini_ImageNet dataset [3], which is derived from the larger ILSVRC-12 dataset, presented more challenges than those conducted on the Omniglot dataset. In the experiments the authors use the following split: "different set of 100 classes, divided into 64 training, 16 validation, and 20 test classes". The primary obstacle we faced was that the majority of available implementations did not support this specific dataset. To address this issue, we adapted an existing implementation from a GitHub repository [2]. By making slight modifications to the implementation and updating the deprecated libraries, we were able to successfully run the prototypical network on the Mini_ImageNet dataset.

For our experiments, we employed 30-way episodes for 1-shot classification and 20-way episodes for 5-shot classification, similar to the experimental setup described in the original paper. This approach allowed us to directly compare our results with the benchmarks set in the existing literature.

[1] https://github.com/swetha2410/Prototypical-Networks-for-Few-Shot-Learning
[2] https://github.com/yashmundhra/prototypical-networks-ymundhra

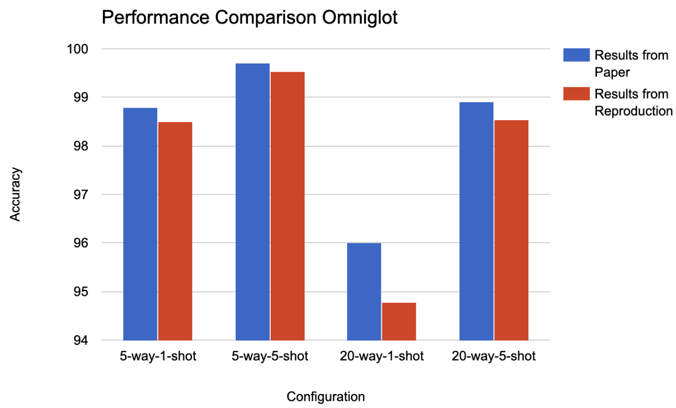

Figure 2: Comparison of reproduced accuracy and paper accuracy on Omniglot dataset. The x-axis indicates configuration of the training episodes (way, distance, and shot), and the y-axis indicates test accuracy for the corresponding shot.

Upon comparing our experimental results with those documented in the original research, as shown in Table 2, we observed that the discrepancies between our reproduced results and the original paper's results were slightly more significant than those we noted with the Omniglot dataset. This variation is likely attributable to the fewer number of training epochs employed in our replication, which possibly hindered the network's ability to train as comprehensively as in the original study. Despite this, the results we achieved were still quite close to those presented in the original paper, demonstrating the robustness of the prototypical network approach.

	20-way accuracy	
	1-shot	5-shot
Paper Results	$49.42 \pm 0.78\%$	$68.20 \pm 0.66\%$
Reproduced Results	$47.07 \pm 0.19\%$	$63.05 \pm 0.23\%$

Table 2: Comparison of reproduced result and paper results for the **Mini_ImageNet** dataset

Figure 3 presents the comparison of the accuracy in a bar graph.

4 Changing the Distance Metric - Kenzo

In the initial paper [4], the authors use L2 norm also called Euclidean distance to compute their distance metric. The distance metric in this case is used to compute the proximity from each samples' features to each one of the barycen-

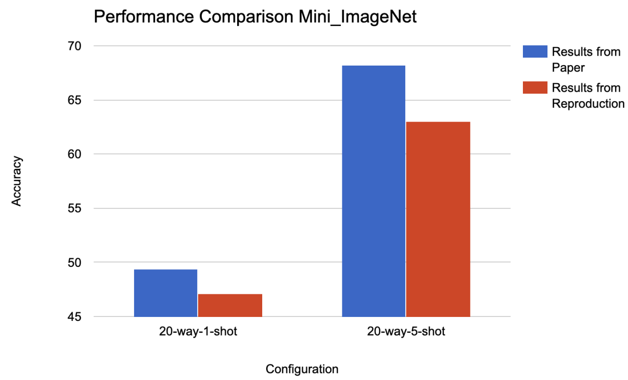

Figure 3: Comparison of reproduced accuracy and paper accuracy on Mini_ImageNet dataset. The x-axis indicates configuration of the training episodes (way, distance, and shot), and the y-axis indicates test accuracy for the corresponding shot.

ters. Given the high dimensionality of features that each sample contains, we thought that a lower distance metric could lead to better results. According to [1] Charu C Aggarwal, Alexander Hinneburg and Daniel A Keim, a L1 norm is preferable to an L2 norm for high dimensional data. Therefore, one of the aims in the reimplementation of this paper is to see the effects on few shot learning using a smaller norm in this case L1 norm. It was decided to use the Omniglot dataset as it obtained more results overall and therefore would make it easier to compare results.Additionally we are going to compare the results of this change in distance metric with our own results as to prevent bias in the equipment. From both implementations we got the following table 3.

Model	5-Way Accuracy				10-Way Accuracy			
	1 shot		5 shot		1 shot		5 shot	
Distance Metric	L2	L1	L2	L1	L2	L1	L2	L1
Prototypical Network Mode	98.5%	98.4%	99.54%	99.5%	94.77%	94.64%	98.53%	98.37%

Table 3: Table comparing different models and distance metric for prototypical network mode.

From the given table, it can be observed that the L1 norm (Manhattan distance) performs constantly worse than the L2 norm (Euclidean distance). Therefore it can be concluded that in this scenario, L2 norm allows for a slightly better performance than L1 norm. However it should be pointed out that the performance is almost not affected my the distance metric. This means if one wants to further optimise the model they should not focus on the distance metric as it does not have a consequent effect on the performance of the model.

5 Ablation Study: Impact of Embedding Size

In this section, we investigate the impact of varying the embedding size on the performance of our few-shot classification model. The paper proposes a method based on prototypical networks, where each class is represented by a prototype computed as the mean of the embedded support examples. The embedding size refers to the dimensionality of the feature space in which examples are represented. We conduct experiments with different embedding sizes and evaluate their effect on both 5-way and 20-way classification tasks, considering both 1-shot and 5-shot scenarios.

Table 4 presents the results of our ablation study, examining different embedding sizes and their effect on both 5-way and 20-way classification tasks, considering both 1-shot and 5-shot scenarios.

Embedding Size	5-way Accuracy		20-way Accuracy	
	1-shot	5-shot	1-shot	5-shot
0	20%	20%	5%	5%
8	96.1%	98.5%	88.2%	95.2%
16	98.1%	99.4%	93.6%	98.03%
32	98.2%	99.5%	94.7%	98.44%
64	98.2%	99.6%	94.77%	98.53%
128	98.3%	99.6%	94.53%	98.53%
256	98.3%	99.6%	94.38%	98.5%
512	97.9%	99.6%	94.56%	98.54%

Table 4: Change in the accuracy of the best model with variation in the size of the embeddings for the Omniglot Dataset)

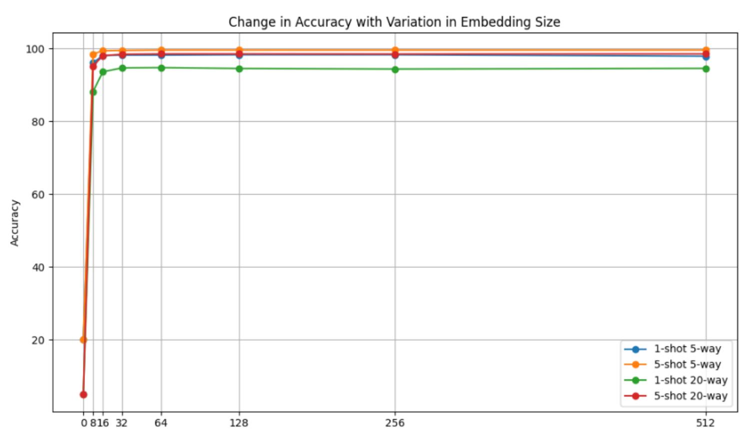

Figure 4: Variation of the accuracy of the model with respect to increase in embedding size

From the results, we observe that embedding sizes of 16 and more tend to yield similar high accuracy on 5-way 5-shot classification. For the 20-way 5-shot task, a similar observation is made for embedding sizes of 32 and more. Hence, a larger representation may not have more advantages over a more compact representation. For the 1-shot methods, we see a similar result with the variation being slightly more pronounced. This is consistent with the idea of maintaining a compact and informative feature representation space. However, further statistical testing would be required to draw more detailed conclusions. All in all, this experiment aligns with the paper's emphasis on the importance of a suitable embedding space for effective few-shot learning.

6 Contributions

Overall the work load was divided evenly amongst the team. Initially, Swetha set up the project on Kaggle. All the members reproduced the results. Yash worked on replicating the results for MiniImageNet and Kenzo experimented with changing the loss function. Both Aratrika and Swetha worked on the ablation study to analyse the impact of the embedding size. For the report, Yash worked on the sections about the experiments with the Omniglot and MiniImageNet Datasets. Kenzo wrote the section about changing the distance metric and Swetha worked on the ablation study. Aratrika wrote the introduction and conclusion and worked on the ablation study as well.

7 Conclusion

In this blog post, we reproduced and expanded upon the findings presented in the paper "Prototypical Networks for Few-shot Learning". By first replicating the experiments conducted on the Omniglot dataset, we verified the reproducibility of the prototypical network approach and confirmed its efficacy in the few-shot learning domain.

Building upon this foundation, we extended our investigation to the MiniImageNet dataset, which presented new challenges and opportunities for experimentation. Despite the differences in dataset characteristics, our adapted implementation of the prototypical network yielded promising results, showcasing the robustness of the approach across varied data domains.

Furthermore, we explored variations of the prototypical network method, such as changing the distance metric and evaluating the impact of embedding size through an ablation study. These additional experiments provided valuable insights into the factors influencing the performance of few-shot learning models, aligning with the goals of the original paper.

Through our experiments, we not only validated the reproducibility of the prototypical networks but also reinforced the robustness of this approach in the field of few-shot learning. Our findings contribute to the growing body of research in this area and highlight the potential of prototypical networks as a promising solution for addressing the challenges of few-shot learning tasks.

As we conclude our exploration, we recognize the importance of continued experimentation and refinement to further advance the state-of-the-art in few-shot learning. By building upon the foundations laid by seminal works like "Prototypical Networks for Few-shot Learning", we can continue to push the boundaries of what is possible in this exciting field of machine learning.

References

[1] Charu C Aggarwal, Alexander Hinneburg, and Daniel A Keim. On the surprising behavior of distance metrics in high dimensional space. In *Database Theory—ICDT 2001: 8th International Conference London, UK, January 4–6, 2001 Proceedings 8*, pages 420–434. Springer, 2001.

[2] Brenden M. Lake, Ruslan Salakhutdinov, and Joshua B. Tenenbaum. Human-level concept learning through probabilistic program induction. *Science*, 350(6266):1332–1338, 2015.

[3] Sachin Ravi and H. Larochelle. Optimization as a model for few-shot learning. In *International Conference on Learning Representations*, 2016.

[4] Jake Snell, Kevin Swersky, and Richard S. Zemel. Prototypical networks for few-shot learning. *CoRR*, abs/1703.05175, 2017.

[5] Oriol Vinyals, Charles Blundell, Timothy P. Lillicrap, Koray Kavukcuoglu, and Daan Wierstra. Matching networks for one shot learning. *CoRR*, abs/1606.04080, 2016.