Reimplementation of Simple Shot (Nearest-Neighbor Classification for Few-Shot Learning)

Kevin Chen, Sophia Wang, Katherine Wei, Anya Yerramilli

Introduction, Background, & Motivation

Introduction

Few-shot learning requires models to recognize new object categories using only 1–5 labeled examples per class, leading to data scarcity. While meta-learning approaches have been proposed to address this, they often involve complex training regimes. *SimpleShot* (Wang et al.) explores if a simpler solution can perform comparably with less computational overhead.

Background

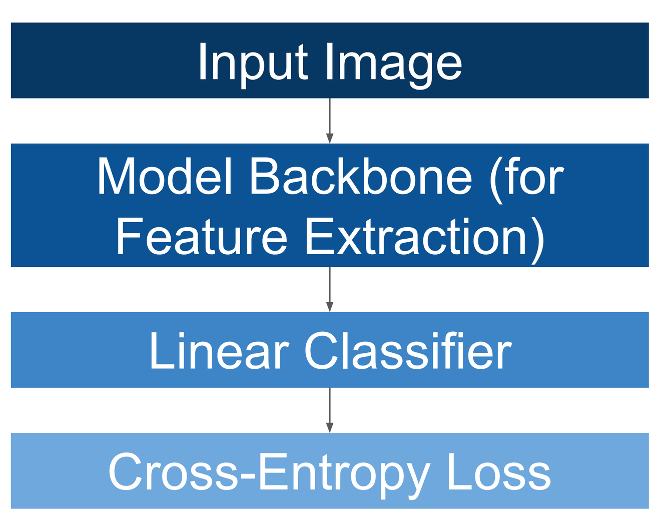
From *SimpleShot* (Wang et al.), we focus on:

- One-shot accuracy
- Five-shot accuracy

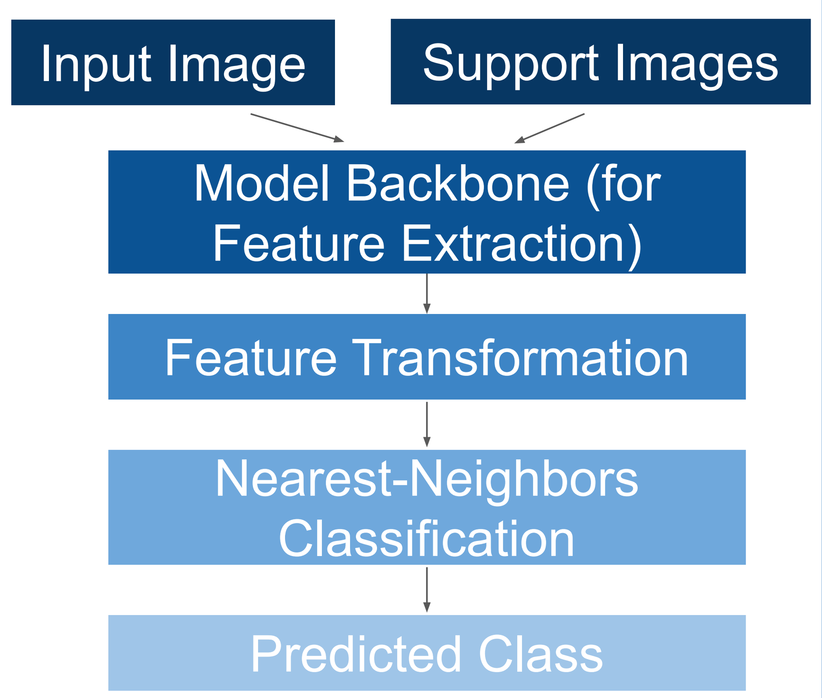
We aimed to reproduce the following results:

Approach	Network	One shot	Five shots
SimpleShot (CL2N)	Conv-4	49.69 ± 0.19	66.92 ± 0.17
SimpleShot (CL2N)	ResNet-10	60.85 ± 0.20	78.40 ± 0.15
SimpleShot (CL2N)	ResNet-18	**62.85 ± 0.20**	**80.02 ± 0.14**

Training Architecture



Testing Architecture



Motivation

Goal: We wanted to validate whether simple feature transformations (centering + L2-normalization) could make nearest-neighbor classifiers competitive with meta-learning.

Hypothesis: “Vanilla” nearest-neighbor classifiers, with proper feature normalization, match or exceed the performance of meta-learning techniques.

References

- Wang, Y., Chao, W.-L., Weinberger, K. Q., & van der Maaten, L. (2019). SimpleShot: Revisiting nearest-neighbor classification for few-shot learning. arXiv. https://arxiv.org/abs/1911.04623
- Zagoruyko, S., & Komodakis, N. (2016). Wide Residual Networks. arXiv. https://arxiv.org/abs/1605.07146

Methodology

Dataset Preparation

- *miniImageNet:* 100 classes, 600 examples/class
- 64 base classes, 16 val classes, 20 test classes
- Resized images to be 84 x 84 pixels (rescaling + center cropping)

Feature Extraction and Transformation

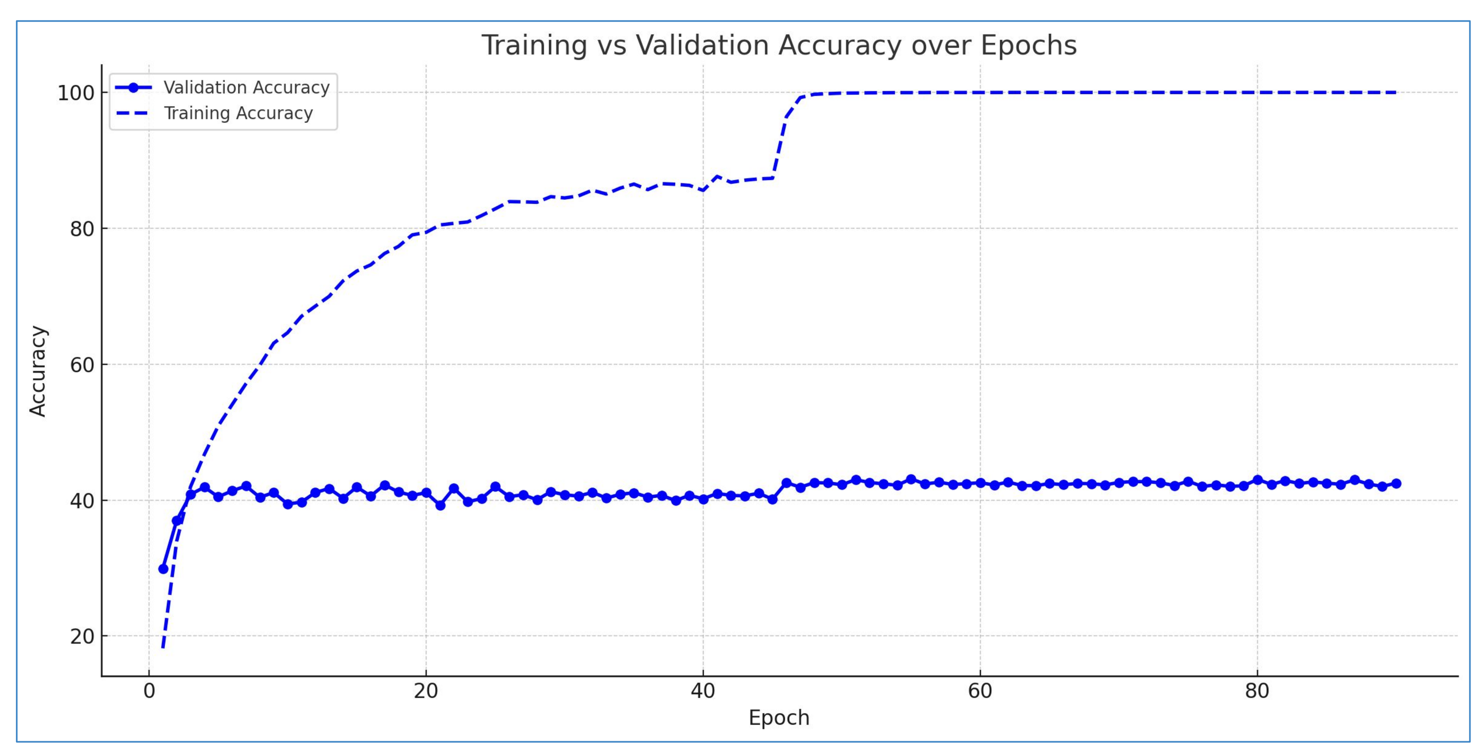
- Trained CNN to extract features from images
  - Conv-4, ResNet-10/18, WRN-28-10, MobileNet, DenseNet-121
- Applied the following feature transformations
  - *L2-normalization **(L2N)**:* $x \leftarrow x / ||x||_2$ for feature x
  - *Centering + L2-normalization **(CL2N)**:* Center the data via $x \leftarrow x - \bar{x}$ and apply L2-normalization.

Nearest-Neighbor Classification

- Classification Rule: for a test image feature vector $\hat{x}$
  - One-shot: assigns the class of the nearest support image
    - $y(\hat{x}) = \arg\min_{c \in \{1, \dots, C\}} ||\hat{x} - \hat{x}_c||_2$
  - Multi-shot: Applies the one-shot rule to the class centroids
- Distance Metric:
  - Euclidean distance with transformed features
  - CL2N: $d(\hat{x}, \hat{x}') = || ((\hat{x}-\bar{x}) / ||\hat{x}-\bar{x}||_2) - ((\hat{x}'-\bar{x}) / ||\hat{x}'-\bar{x}||_2) ||_2$

Training and Evaluation

- 90 epochs, batch size 256, with SGD + Cross Entropy loss
- LR: 0.1, then / 10 at epoch 45 and 66
- 1-shot and 5-shot 5-way tasks (5 novel classes)
- Tried: UN, L2N, CL2N (centered then L2N), validated with CL2N



Training accuracy steadily rose to ~80% by epoch 45, then jumped to >95% after the LR drop. Validation accuracy hovered around 40% over all epochs, peaking at 43.17% after the LR drop.

Modifications & Design Choices

- Only used *miniImageNet* dataset.
- Some CNN descriptions from the paper were unclear and referenced other papers, so we tried our best to replicate the exact CNN designs.

Results & Conclusions

Testing

We evaluated our models on 5-way 1-shot and 5-way 5-shot classification tasks on using three different feature transformations: UN (Unnormalized), L2N (L2 Normalization), and CL2N (Centered L2 Normalization). The resulting accuracies are as follows,

Conv-4:

	5-way 1-shot	5-way 5-shot
UN	30.76% ± 0.15%	58.41 ± 0.17%
L2N	43.53% ± 0.18%	61.09% ± 0.17%
CL2N	44.65% ± 0.18%	61.19% ± 0.17%

ResNet-10:

	5-way 1-shot	5-way 5-shot
UN	50.70% ± 0.20%	70.09% ± 0.16%
L2N	52.41% ± 0.20%	70.96% ± 0.16%
CL2N	54.84% ± 0.20%	70.92% ± 0.16%

ResNet-18:

	5-way 1-shot	5-way 5-shot
UN	48.58% ± 0.20%	69.91% ± 0.16%
L2N	53.08% ± 0.20%	71.10% ± 0.16%
CL2N	55.71% ± 0.20%	71.46% ± 0.16%

Our results reinforce the findings of the original *SimpleShot* (Wang et al.) paper: feature normalization, particularly CL2N, significantly improves few-shot classification performance. Despite slightly lower testing accuracies for all settings compared to the original paper, our results still confirm the impact of feature transformation techniques and number of shots on model performance. With a Conv-4 backbone, accuracy improved from 31.51% (UN) to 46.15% (CL2N) on 5-way 1-shot tasks. While ResNet-10 and ResNet-18 achieve higher accuracy overall between 69.91% and 71.46%), Conv-4 remains competitive given its simplicity. The impact of normalization is most pronounced in 1-shot settings, where structured embeddings are critical.

Conclusion

Our results support the hypothesis that simple nearest-neighbor classifiers, when paired with proper feature normalization (such as Centering + L2N or CL2N), can achieve comparable performance to meta-learning methods in few-shot learning tasks. The SimpleShot methods provide significant advantages in training efficiency and simplicity compared to meta-learning.