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

Kevin Chen, Sophia Wang, Katherine Wei, Anya Yerramilli

Introduction, Background, & Motivation

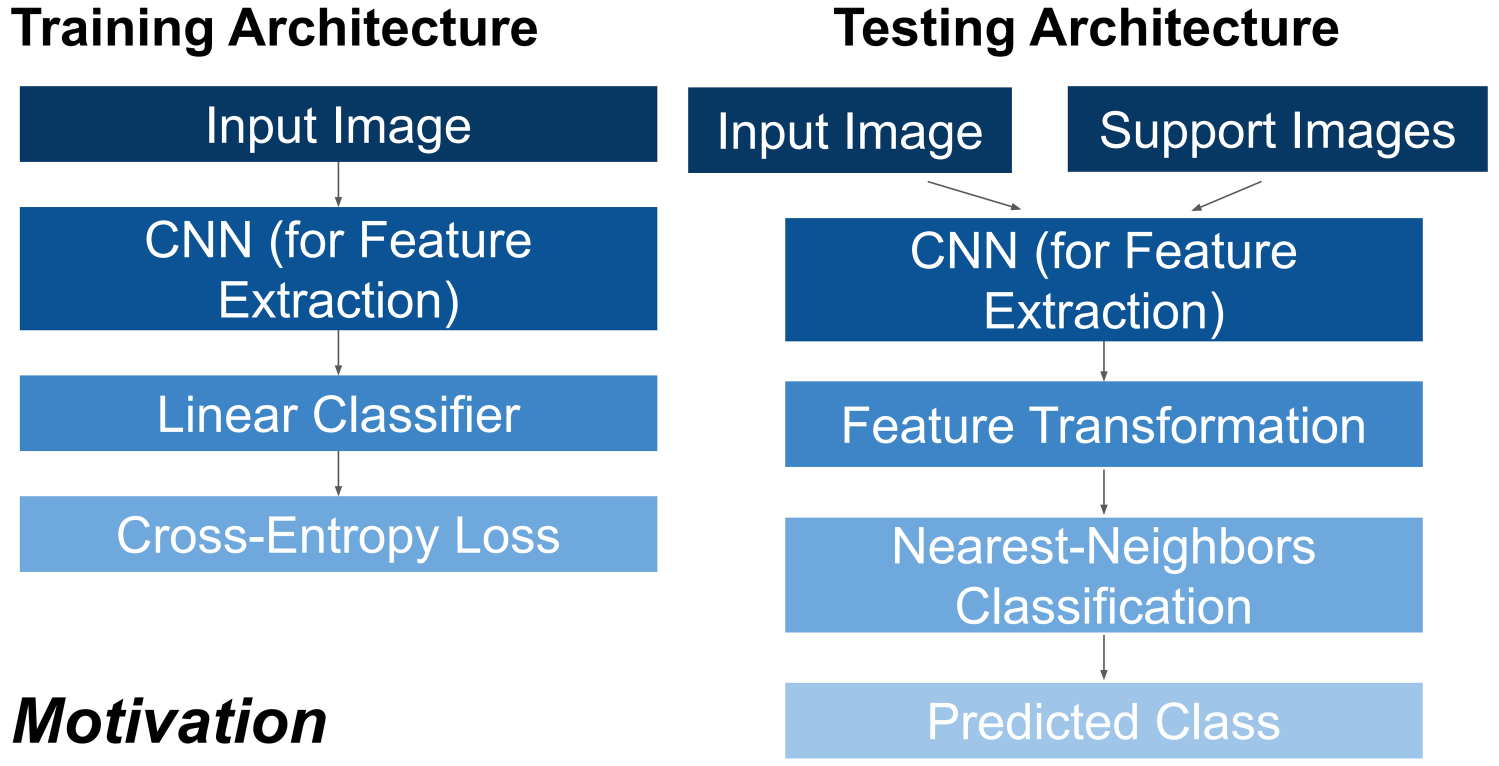
Introduction & Background

- Task: Apply few-shot learning to classify images.
- Traditionally, meta-learning has been used for these tasks.
- However, our paper [1] proposes a **new** approach called SimpleShot that is more simple yet yields similar performance.

The results in blue use SimpleShot, and everything else uses meta-learning. They show the accuracy of 1-shot and 5-shot for 5-way classification. Our **goal** is to reproduce the blue 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**
SimpleShot (CL2N)	WRN	63.50 ± 0.20	80.33 ± 0.14
MatchingNet [34]	Conv-4	43.56 ± 0.84	55.31 ± 0.73
MatchingNet [34]†	ResNet-18	52.91 ± 0.88	68.88 ± 0.69
MatchingNet [34]♯	WRN	64.03 ± 0.20	76.32 ± 0.16

SimpleShot Design:



Motivation

We want to validate whether simple feature transformations (centering + L2-normalization) can make SimpleShot competitive with meta-learning for few-shot image classification.

References

[1] 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

[2] O. Vinyals, C. Blundell, T. Lillicrap, D. Wierstra, et al. Matching networks for one shot learning. arXiv. https://arxiv.org/abs/1606.04080

[3] J. Snell, K. Swersky, and R. Zemel. Prototypical networks for few-shot learning. arXiv. https://arxiv.org/abs/1703.05175

[4] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. arXiv. https://arxiv.org/abs/1512.03385

[5] Zagoruyko, S., & Komodakis, N. (2016). Wide Residual Networks. arXiv. https://arxiv.org/abs/1605.07146

Methodology

Dataset Preparation

- *mini*ImageNet [2]: 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

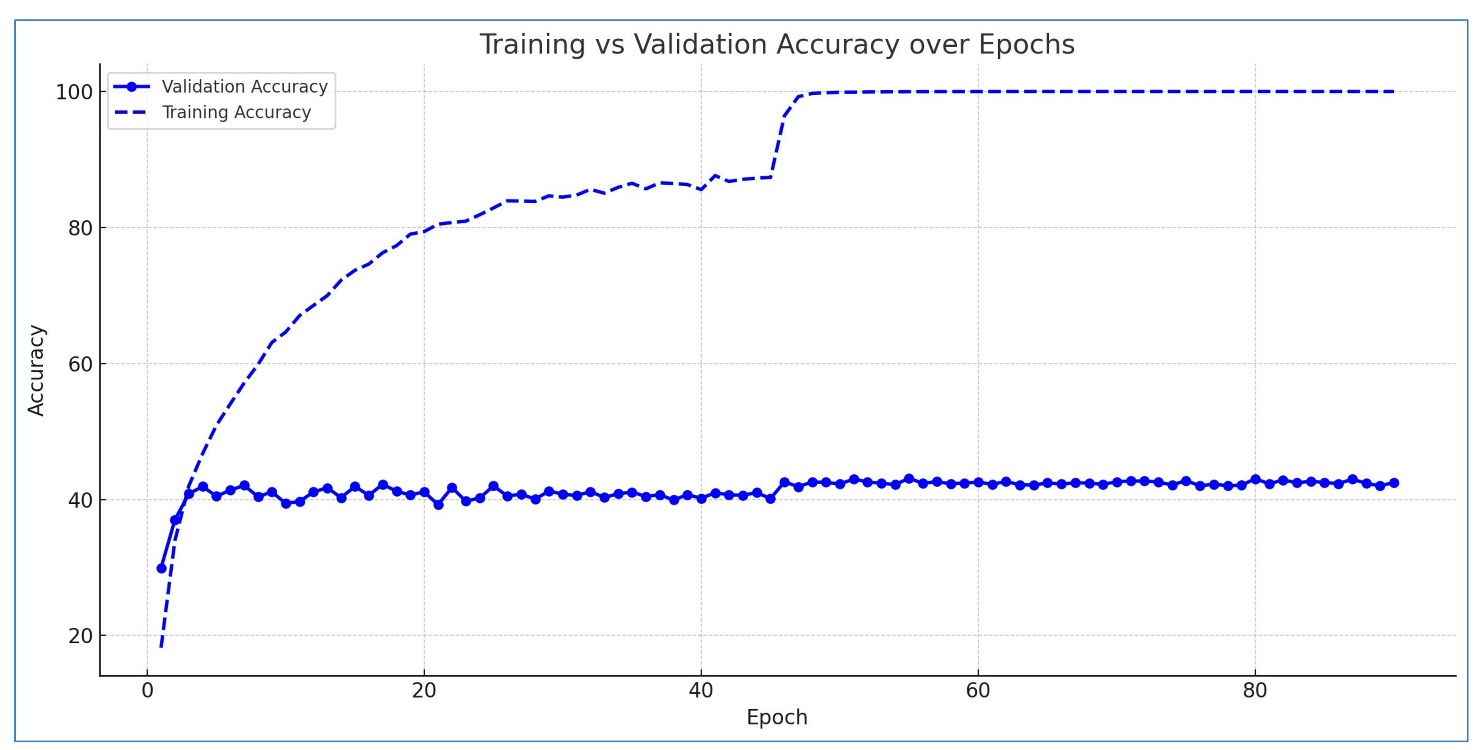
- Trained CNN to extract features from images
  - Conv-4 [3], ResNet-10/18 [4], WRN-28-10 [5]
- Applied the following feature transformations
  - *L2-normalization* ***(L2N)***: $x \leftarrow x / ||x||_2$ for feature x
  - *Centering + L2-normalization* ***(CL2N)***: Center the data via $x \leftarrow x - \bar{x}$ and apply L2-normalization.

Nearest-Neighbor Classification

- Classification Rule for a test image with feature vector $\hat{x}$:
  - One-shot: assigns the class of the nearest support image
    - $y(\hat{x}) = \arg\min_{c \in \{1, \ldots C\}} ||\hat{x} - \hat{x}_c||_2$
  - 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

- Applied linear classifier on features with cross entropy loss to train the CNN
- 90 epochs, batch size 256, used SGD on cross entropy loss
- Learning rate: starts at 0.1, then divide by 10 at epochs 45 and 66
- Used early stopping with CL2N applied to validation dataset



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

Results

We recorded 5-way 1-shot and 5-way 5-shot accuracies using UN (Unnormalized), L2N, and CL2N.

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%

WRN-28-10:

	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 are slightly **lower** than those from the original paper.
- But our reimplementation is still comparable with other state-of-the-art few-shot learners, such as MatchingNet shown by the table in the leftmost panel..
- Our results show how performance increases when going from UN → L2N → CL2N.

Conclusion

- Our results support that SimpleShot can achieve comparable performance to meta-learning methods in few-shot learning tasks.
- SimpleShot provides significant advantages in training efficiency and offering a more simple architecture compared to meta-learning.