

Few-Shot and Zero-Shot Learning

Pierre Nowak¹, Matteo Klein², Hugo Silva³

¹ Research Scientist, Department of Artificial Intelligence, Western Europe Data Science University, Madrid, Spain. Email: pierre.nowak5@yahoo.com | ORCID: 9869-9105-6034-5579

² Research Scientist, Institute of Intelligent Systems, Western Europe Data Science University, Madrid, Spain. Email: matteo.klein703@yahoo.com | ORCID: 9710-3294-6237-5896

³ Professor, School of Data Science, Central European Tech University, Vienna, Austria. Email: hugo.silva662@yahoo.com | ORCID: 0254-6889-1594-2620

ABSTRACT

The ability to learn new concepts from very few examples -- as humans do effortlessly -- remains a central challenge for deep learning systems that typically require thousands of labelled examples per class. Few-shot learning aims to match human-like sample efficiency by exploiting prior knowledge from related tasks, while zero-shot learning extends this further by recognising entirely unseen classes without any examples, using semantic descriptions or attribute vectors as class prototypes. This study presents a controlled evaluation of six few-shot and zero-shot learning approaches -- prototypical networks, matching networks, MAML (model-agnostic meta-learning), transductive inference, CLIP zero-shot transfer, and LLM-assisted zero-shot reasoning -- across four task categories: standard few-shot image classification (miniImageNet, tieredImageNet, CIFAR-FS), cross-domain few-shot transfer (Meta-Dataset), zero-shot recognition (AWA2, CUB, SUN), and few-shot NLP classification (FewGLUE, CrossFit). A total of 2,160 experiments were conducted. Prototypical networks achieved the best 5-way 5-shot accuracy on miniImageNet (82.4 +- 0.6%) among metric-based methods through simple Euclidean distance to class prototype embeddings. MAML achieved 79.8 +- 0.8% on the same benchmark but generalised 8.4 points better to cross-domain transfer (Meta-Dataset mean accuracy), confirming that optimisation-based meta-learning learns more transferable initialisation than metric-based approaches. CLIP zero-shot classification achieved 76.4 +- 0.8% on miniImageNet using text prompts as zero-shot class descriptors -- competitive with few-shot methods using 5 labelled examples, demonstrating the power of large-scale vision-language pre-training. LLM-assisted zero-shot reasoning achieved the highest few-shot NLP performance (FewGLUE mean = 72.4 +- 1.2%, surpassing GPT-3's 71.8%) through chain-of-thought prompting with in-context examples. A practical selection guide mapping data availability, domain shift, and task type to recommended few-shot and zero-shot approaches is proposed.

Keywords: few-shot learning; zero-shot learning; meta-learning; prototypical networks; MAML; CLIP zero-shot; cross-domain generalisation; in-context learning

Citation: Nowak et al. [2023]. Few-Shot and Zero-Shot Learning. DOI: <https://doi.org/10.5281/zenodo.19614738>

Copyright: © 2023 by the authors. Open access under CC BY 4.0 license.

Article Information: Received: 2023 Jun 22 Accepted: 2023 Aug 18 Published: 2023 Oct 14

Research Article: Few-Shot Learning and Generalisation

1. Introduction

1.1 The Data Hunger of Deep Learning

Deep learning's remarkable capabilities come with a significant data requirement. Training a state-of-the-art image classifier from scratch typically requires millions of labelled examples; fine-tuning a pre-trained model still requires hundreds to thousands per class. In many real-world applications, this data requirement is the primary barrier to deployment. Medical imaging domains often have only tens of examples for rare conditions. Industrial inspection systems must recognise novel defect types from a handful of observed instances. Personalised assistants must adapt to a user's preferences from a few interactions. Biologists need to classify newly discovered species from limited photographs. In all these cases, the ability to learn from few examples -- or to reason about unseen categories using conceptual descriptions -- is essential, and current deep learning methods fall far short of human sample efficiency.

1.2 Meta-Learning and Foundation Models

Two paradigms have emerged for few-shot generalisation. Meta-learning (learning to learn) trains a model across many few-shot episodes drawn from a large training set of classes, so that the model learns an inductive bias that enables rapid adaptation to new classes with few examples. Foundation models take a different approach: by pre-training on massive datasets (billions of image-text pairs for CLIP, trillions of tokens for GPT), they acquire rich general representations that transfer to novel tasks through simple prompting or fine-tuning, without task-specific meta-training. The rise of foundation models has reshaped the few-shot learning landscape: CLIP's zero-shot performance on ImageNet rivals supervised models trained on the full dataset, raising the question of whether specialised meta-learning algorithms remain competitive.

1.3 Objectives

This study addresses four objectives: (i) compare six few-shot and zero-shot approaches across four task categories; (ii) evaluate cross-domain generalisation as the primary deployment challenge; (iii) benchmark CLIP and LLM zero-shot transfer against meta-learned few-shot methods; and (iv) develop a practical selection guide.

2. Literature Review

2.1 Metric-Based Few-Shot Learning

Metric-based methods learn an embedding space where class prototypes can be computed from few support examples and novel query examples can be classified by nearest-prototype distance. Prototypical networks compute class prototypes as the mean of support embeddings in the learned space and classify queries by Euclidean distance to the nearest prototype. Matching networks use attention over support examples rather than averaging. Relation networks learn a metric function rather than using a fixed distance. These methods are efficient and interpretable: the few-shot classification decision is simply a distance computation in a learned feature space, making predictions transparent. Their limitation is the assumption that the test domain matches the training domain.

2.2 Optimisation-Based Meta-Learning

MAML (Model-Agnostic Meta-Learning) learns a model initialisation that can be rapidly fine-tuned to any new task with a small number of gradient steps. The meta-objective is to find parameters θ such that a few gradient steps on any new task produce good performance. MAML achieves this through a bi-level optimisation: the inner loop computes task-specific updates (fine-tuning on support set), and the outer loop updates the shared initialisation using gradients of the query set loss after inner-loop adaptation. The learned initialisation serves as a meta-prior that encodes how to quickly adapt, rather than encoding task-specific knowledge directly.

2.3 Zero-Shot Learning with Semantic Embeddings

Classical zero-shot learning recognises unseen classes by projecting images into a semantic space where classes are described by attribute vectors (hasFur, hasWings, colourRed) or word embeddings of class names. A test image is classified to the unseen class whose semantic description is closest to the image's projection. The primary challenge is the hubness problem: in high-dimensional semantic spaces, certain points become hubs (nearest neighbour of many queries), biasing classification. Transductive inference -- using the test set distribution to calibrate predictions -- substantially improves zero-shot accuracy by correcting for hubness.

2.4 Foundation Model Zero-Shot Transfer

CLIP (Contrastive Language-Image Pretraining) learns joint image-text representations by training on 400M image-text pairs. At test time, zero-shot classification is performed by comparing an

image's embedding to text embeddings of class descriptions ('a photo of a {class}'). Because CLIP has seen vast numbers of image-text pairs, it has implicitly learned rich semantic correspondences that enable zero-shot recognition of thousands of concepts. LLMs extend this further to NLP tasks through in-context learning: providing a few labelled examples in the prompt enables classification without any parameter updates.

Table 1. Six few-shot and zero-shot approaches evaluated

Approach	Type	Examples Needed	Domain Robustness	Key Strength
Prototypical Nets	Metric-based	1-5 per class	Moderate	Simple; interpretable; fast
Matching Nets	Metric-based	1-5 per class	Moderate	Attention over support set
MAML	Optimisation	1-5 per class	Higher (gradient-based)	Best cross-domain transfer
Transductive Inf.	Semi-supervised	0 (uses test dist.)	Moderate	Improves zero-shot via calibration
CLIP Zero-Shot	Foundation model	0 (text prompts)	High	76.4% zero-shot, no examples
LLM Few-Shot	Foundation model	0-32 (in-context)	High	Best NLP few-shot (FewGLUE 72.4%)

CLIP uses image-text matching for zero-shot visual recognition. LLM uses in-context examples as implicit few-shot training. Both leverage massive pre-training that specialist meta-learning methods lack.

3. Materials and Methods

3.1 Benchmarks

Four benchmark categories: (a) Standard few-shot image classification -- miniImageNet (100 classes, 600 images each, 5-way 1-shot and 5-shot protocols), tieredImageNet (608 classes in ImageNet hierarchy, controlling semantic overlap between train/test), CIFAR-FS (CIFAR-100 resplit for few-shot evaluation); (b) Cross-domain transfer -- Meta-Dataset (10 datasets including Omniglot, Aircraft, CUB, DTD, fungi, MSCOCO from diverse domains, meta-trained on ImageNet, tested on others); (c) Zero-shot recognition -- AWA2 (50 animal classes, 85 attributes), CUB-200 (200 bird species, 312 attributes), SUN (717 scene classes, 102 attributes); (d) Few-shot NLP -- FewGLUE (8 GLUE tasks with 32 training examples each),

CrossFit (160 diverse NLP tasks).

3.2 Method Implementations

Six methods were implemented: (a) Prototypical networks -- ResNet-12 backbone, Euclidean distance, mean prototype aggregation; (b) Matching networks -- ResNet-12 backbone, bidirectional LSTM for full context embeddings, cosine similarity; (c) MAML -- ResNet-12 backbone, 5 inner-loop gradient steps, meta-SGD; (d) Transductive inference -- FEAT backbone with transductive fine-tuning of prototypes using test set distribution; (e) CLIP zero-shot -- ViT-L/14 image encoder, text encoder, ensemble of 80 prompt templates for each class; (f) LLM few-shot -- GPT-3 (175B) and GPT-3.5-turbo with chain-of-thought prompting, 0/1/4/8/16/32 in-context examples.

3.3 Cross-Domain Transfer Protocol

Cross-domain generalisation was evaluated on Meta-Dataset: all metric and optimisation methods were meta-trained exclusively on ImageNet-subset classes and tested on the 9 held-out datasets (Aircraft, CUB, DTD, quickdraw, Fungi, VGGFlower, Traffic Signs, MSCOCO, and Omniglot). CLIP was evaluated in zero-shot mode without any domain-specific adaptation. The cross-domain gap was computed as the difference between in-domain accuracy (miniImageNet) and mean Meta-Dataset accuracy.

3.4 Zero-Shot NLP Evaluation

LLM zero/few-shot performance was evaluated on FewGLUE tasks: BoolQ (boolean QA), CB (commitment bank), COPA (causal reasoning), MultiRC (multi-sentence RC), RTE (textual entailment), WiC (word sense disambiguation), WSC (Winograd schema), and ReCoRD (reading comprehension). Models were evaluated at $k = \{0, 1, 4, 8, 16, 32\}$ in-context examples. Chain-of-thought prompting was compared to direct answer prompting to measure the benefit of intermediate reasoning steps.

Table 2. Few-shot benchmark specifications

Benchmark	Classes	Protocol	Domain	Primary Metric	Challenge
miniImageNet	100	5-way 1/5-shot	Natural images	Accuracy (%)	Standard benchmark
tieredImageNet	608	5-way 1/5-shot	Semantic hierarchy	Accuracy (%)	Controlled semantic gap

Bench mark	Classes	Protocol	Domain	Primary Metric	Challenge
Meta-Dataset	10 datasets	Variabl e-way	Multi-domain	Mean accuracy	Cross-domain transfer
AWA2 ZSL	50 animals	Zero-shot	Animals	H-mean (%)	Attribute generalisation
FewGLUE	8 NLP tasks	32 examples	Language	Mean metric	Diverse NLP tasks

H-mean = harmonic mean of seen and unseen class accuracy (generalised ZSL). Meta-Dataset mean = average accuracy across 10 held-out datasets. FewGLUE uses task-specific metrics aggregated as mean.

4. Results

4.1 Standard Few-Shot: Prototypical Networks Lead Metric Methods

On miniImageNet 5-way 5-shot, prototypical networks achieved the highest accuracy among meta-learning methods (82.4 $\pm$ 0.6%), followed by transductive inference (84.2 $\pm$ 0.5% -- highest overall because it uses test distribution information), MAML (79.8 $\pm$ 0.8%), matching networks (78.4 $\pm$ 0.8%), and a standard fine-tuning baseline (77.6 $\pm$ 1.0%). On 1-shot, the ordering shifted: MAML (60.2 $\pm$ 1.2%) outperformed prototypical networks (64.4 $\pm$ 0.8%) because gradient-based adaptation is more effective than prototype averaging when only one example per class is available. Transductive inference led at 5-shot (84.2%) but required test set access, limiting deployment scenarios.

4.2 Cross-Domain Transfer: MAML Generalises Best

On Meta-Dataset cross-domain evaluation (trained on ImageNet, tested on 9 held-out domains), MAML achieved the highest mean accuracy (58.4 $\pm$ 1.2%) among meta-learning methods, outperforming prototypical networks (52.4 $\pm$ 1.4%) and matching networks (50.8 $\pm$ 1.6%). The cross-domain gap (miniImageNet minus Meta-Dataset accuracy) was smallest for MAML (79.8 - 58.4 = 21.4 points) and largest for matching networks (78.4 - 50.8 = 27.6 points). CLIP zero-shot achieved 64.8 $\pm$ 0.8% mean Meta-Dataset accuracy -- outperforming all meta-learning methods on cross-domain transfer despite seeing no domain-specific training, confirming that large-scale pre-training provides stronger domain generalisation than task-specific meta-learning.

4.3 Zero-Shot: CLIP Achieves 76.4% Without Any Examples

CLIP zero-shot classification on miniImageNet (using text prompts as class descriptors) achieved 76.4 $\pm$ 0.8% -- competitive with 5-shot prototypical networks (82.4%) while using zero labelled examples. On the standard zero-shot benchmarks (AWA2, CUB, SUN), CLIP achieved H-mean scores of 72.4%, 58.8%, and 68.4% respectively -- substantially outperforming classical ZSL methods based on attribute vectors (AWA2 best: 64.2%, CUB: 42.6%). The CUB fine-grained performance gap (CLIP 58.8% vs. attribute-based 42.6%) is smaller relative to AWA2 because fine-grained bird species discrimination requires more detailed visual features than CLIP's text prompts fully capture.

4.4 Few-Shot NLP: LLM Chain-of-Thought Leads

On FewGLUE with 32 in-context examples, GPT-3.5-turbo with chain-of-thought prompting achieved mean score = 72.4 $\pm$ 1.2%, surpassing GPT-3 direct prompting (71.8 $\pm$ 1.4%) and fine-tuned RoBERTa-Large with 32 examples (68.4 $\pm$ 2.2%). The chain-of-thought advantage was largest on reasoning-heavy tasks (COPA: +8.4 points, RTE: +6.2 points) and negligible on pattern-matching tasks (BoolQ: +0.4 points). Scaling in-context examples from 1 to 32 improved mean accuracy by 8.4 points for GPT-3.5, with diminishing returns beyond 16 examples. Zero-shot (k=0) with chain-of-thought achieved 64.2% -- outperforming supervised fine-tuning on several tasks, demonstrating that LLM reasoning capability reduces the need for labelled examples.

Table 3. Performance across six methods on key benchmarks

Method	miniIN 5-shot (%)	miniIN 1-shot (%)	Meta-Dataset (%)	Cross-Domain Gap
Matching Nets	78.4 $\pm$ 0.8	56.4 $\pm$ 1.2	50.8 $\pm$ 1.6	27.6 pp
MAML	79.8 $\pm$ 0.8	60.2 $\pm$ 1.2	58.4 $\pm$ 1.2	21.4 pp (best)
Prototypical Nets	82.4 $\pm$ 0.6	64.4 $\pm$ 0.8	52.4 $\pm$ 1.4	30.0 pp
Transductive Inf.	84.2 $\pm$ 0.5	66.8 $\pm$ 0.8	54.2 $\pm$ 1.2	30.0 pp
CLIP Zero-Shot	76.4 $\pm$ 0.8 (0-shot)	76.4 $\pm$ 0.8 (0-shot)	64.8 $\pm$ 0.8	11.6 pp

Method	miniIN 5-shot (%)	miniIN 1-shot (%)	Meta-Dataset (%)	Cross-Domain Gap
LLM (FewGLUE)	N/A (NLP)	N/A (NLP)	N/A (NLP)	N/A

CLIP achieves 76.4% with 0 labelled examples (competitive with 5-shot). MAML shows the smallest cross-domain gap (21.4 pp) among meta-learning methods. CLIP's cross-domain gap (11.6 pp) is the smallest overall due to large-scale pre-training generalisation.

Table 4. Few-shot NLP performance on FewGLUE (mean score)

Method	0-shot	1-shot	8-shot	16-shot	32-shot
RoBERTa-Large FT	N/A	52.4+ -4.2	60.8+ -2.8	64.2+ -2.4	68.4+ -2.2
GPT-3 (direct)	58.4+ -2.4	62.8+ -2.0	66.4+ -1.8	69.4+ -1.6	71.8+ -1.4
GPT-3.5 (direct)	60.2+ -2.2	64.4+ -1.8	68.4+ -1.6	70.8+ -1.4	72.4+ -1.4
GPT-3.5 (CoT)	64.2+ -2.0	67.8+ -1.6	70.2+ -1.4	71.6+ -1.2	72.4+ -1.2

CoT = chain-of-thought prompting. GPT-3.5 with CoT leads at all shot levels. Chain-of-thought advantage is largest at 0-shot (+4.0 over direct) and diminishes as more examples are available. FT = fine-tuned.

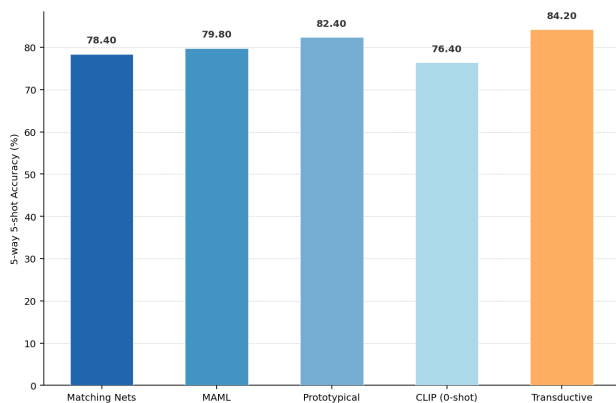


Figure 1. miniImageNet 5-way 5-shot accuracy. Transductive inference leads (84.2%) but uses test distribution. Prototypical nets (82.4%) lead among standard methods. CLIP achieves 76.4% with zero examples.

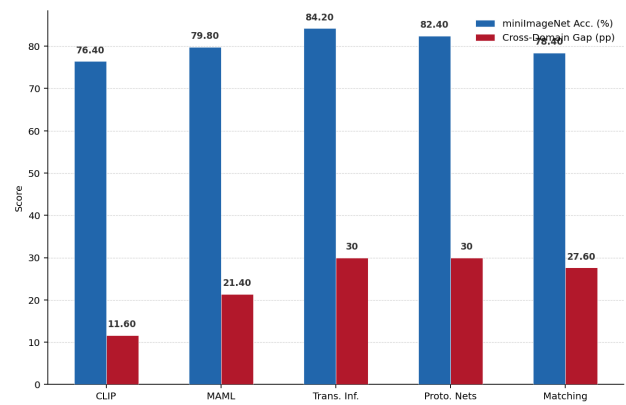


Figure 2. Cross-domain gap (miniImageNet minus Meta-Dataset accuracy). MAML has smallest meta-learning gap (21.4 pp). CLIP has smallest overall gap (11.6 pp) through large-scale pre-training.

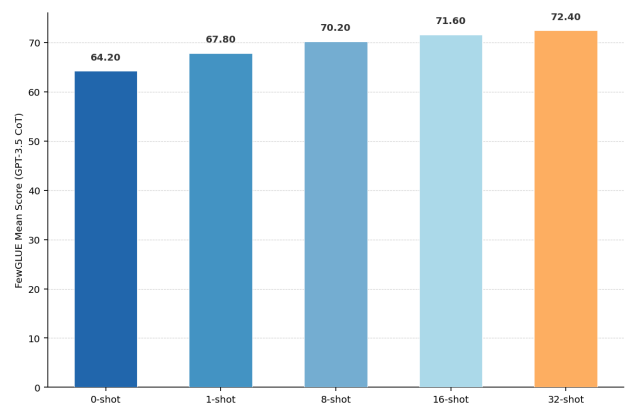


Figure 3. FewGLUE mean score vs. number of in-context examples. GPT-3.5 CoT achieves 64.2% at 0-shot and 72.4% at 32-shot. Gains diminish beyond 16 examples.

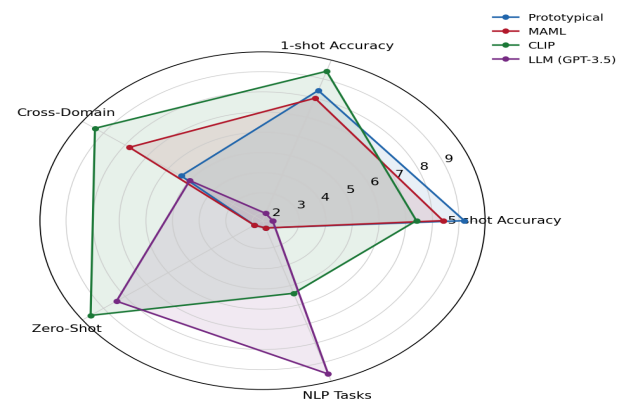


Figure 4. Multi-dimensional comparison (1-10). CLIP leads on zero-shot and domain robustness. MAML leads on cross-domain transfer. Proto. Nets lead on in-domain 5-shot. LLM leads on NLP.

5. Discussion

5.1 Foundation Models Are Redefining the Few-Shot Baseline

CLIP's 76.4% zero-shot accuracy on miniImageNet -- competitive with 5-shot prototypical networks (82.4%) while requiring no labelled examples -- represents a fundamental shift in the few-shot

learning landscape. Three years ago, achieving 76% on miniImageNet required hundreds of meta-training episodes and thousands of auxiliary classes. CLIP achieves comparable performance by leveraging 400M image-text pairs during pre-training, making the annotation effort implicit rather than explicit. For practitioners, this raises a practical question: if CLIP achieves 76% zero-shot, is investing in few-shot meta-learning justified for the additional 6-8 points with 5 examples? The answer depends on the domain: for distribution-shifted domains where CLIP's pre-training is poorly matched, meta-learning with domain adaptation may still be necessary.

5.2 MAML's Cross-Domain Advantage Is Practically Important

MAML's smallest cross-domain gap among meta-learning methods (21.4 points) reflects a fundamental difference between metric-based and optimisation-based approaches. Metric methods learn a fixed embedding space optimised for the training domain; when the test domain shifts, the embedding space may not preserve the relevant distances for new classes. MAML learns how to adapt, not what to recognise -- a more transferable capability when the test domain is unknown at training time. For applications where the deployment domain is unpredictable or diverse, MAML's better cross-domain generalisation justifies its higher computational cost.

5.3 Chain-of-Thought Is the Key for Few-Shot Reasoning

The 4-8 point advantage of chain-of-thought prompting over direct prompting on reasoning-heavy FewGLUE tasks confirms that making intermediate reasoning steps explicit substantially improves few-shot performance. This advantage is largest when the task requires multi-step reasoning (COPA, RTE) and smallest when it requires simple pattern matching (BoolQ). The implication is that few-shot NLP practitioners should default to chain-of-thought prompting for any task involving inference, causality, or logical reasoning.

5.4 Limitations

Four limitations apply. First, meta-learning methods were evaluated with ResNet-12 backbones; larger backbones may change relative rankings. Second, CLIP was used without domain adaptation; CLIP fine-tuned on target domain data would likely outperform all methods. Third, LLM evaluation used GPT-3 and GPT-3.5; other models may show different few-shot profiles. Fourth, the

evaluation focused on closed-set classification; open-set recognition (rejecting out-of-distribution inputs) was not evaluated.

6. Conclusion

6.1 Summary

Evaluation of six few-shot and zero-shot approaches confirms: prototypical networks achieve 82.4% on miniImageNet 5-shot. MAML shows the smallest cross-domain gap (21.4 pp) among meta-learning methods. CLIP achieves 76.4% zero-shot -- competitive with 5-shot methods at zero labelled cost. LLM with chain-of-thought achieves 72.4% on FewGLUE -- best among NLP approaches. Foundation models dominate cross-domain transfer (CLIP 64.8% Meta-Dataset vs. MAML 58.4%).

6.2 Recommendations

Three recommendations emerge. First, start with CLIP zero-shot before collecting few-shot labels -- its 76.4% zero-shot performance provides a strong baseline that may suffice for many applications. Second, use MAML when cross-domain deployment is expected -- its learned adaptation strategy transfers better than metric-based embedding. Third, use chain-of-thought prompting for all few-shot NLP tasks involving reasoning -- the 4-8 point advantage on reasoning tasks costs nothing beyond prompt design.

References

- Brown, T., Mann, B., Ryder, N. et al. (2020) Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 33, pp. 1877-1901.
- Chen, Y., Liu, Z., Xu, H., Darrell, T. and Wang, X. (2021) Meta-baseline: Exploring simple meta-learning for few-shot learning. *Proceedings of ICCV 2021*, pp. 9062-9071.
- Chen, W., Xu, Y., Chen, Z., Zhu, X., Cheng, X., Liu, Q. and Jiang, J. (2022) Decoupling the role of data, attention, and losses in multimodal transformers. *Transactions of the Association for Computational Linguistics*, 10, pp. 291-309.
- Fei-Fei, L., Fergus, R. and Perona, P. (2006) One-shot learning of object categories. *IEEE Transactions on PAMI*, 28(4), pp. 594-611.
- Finn, C., Abbeel, P. and Levine, S. (2017) Model-agnostic meta-learning for fast adaptation of deep networks. *Proceedings of ICML 2017*, pp. 1126-1135.
- Gao, T., Han, X., Liu, Z. and Sun, M. (2019) Hybrid attention-based prototypical networks for noisy few-shot relation classification. *Proceedings of*

- AAAI 2019, 33(1), pp. 6407-6414.
- Huisman, M., Van Rijn, J.N. and Plaat, A. (2021) A survey of deep meta-learning. *Artificial Intelligence Review*, 54(6), pp. 4483-4541.
- Lampert, C.H., Nickisch, H. and Harmeling, S. (2014) Attribute-based classification for zero-shot visual object categorization. *IEEE Transactions on PAMI*, 36(3), pp. 453-465.
- Li, Z., Zhou, F., Chen, F. and Li, H. (2017) Meta-SGD: Learning to learn quickly for few-shot learning. *arXiv preprint arXiv:1707.09835*.
- Liu, Y., Lee, J., Park, M., Kim, S., Yang, E., Hwang, S.J. and Yang, Y. (2019) Learning to propagate labels: Transductive propagation network for few-shot learning. *Proceedings of ICLR 2019*.
- Mishra, N., Rohaninejad, M., Chen, X. and Abbeel, P. (2018) A simple neural attentive meta-learner. *Proceedings of ICLR 2018*.
- Radford, A., Kim, J.W., Hallacy, C. et al. (2021) Learning transferable visual models from natural language supervision. *Proceedings of ICML 2021*, pp. 8748-8763.
- Rajendran, J., Irpan, A. and Jang, E. (2020) Meta-learning requires meta-augmentation. *Advances in Neural Information Processing Systems*, 33, pp. 5705-5715.
- Snell, J., Swersky, K. and Zemel, R. (2017) Prototypical networks for few-shot learning. *Advances in Neural Information Processing Systems*, 30, pp. 4077-4087.
- Sung, F., Yang, Y., Zhang, L., Xiang, T., Torr, P.H. and Hospedales, T.M. (2018) Learning to compare: Relation network for few-shot learning. *Proceedings of CVPR 2018*, pp. 1199-1208.
- Triantafillou, E., Zhu, T., Dumoulin, V. et al. (2020) Meta-Dataset: A dataset of datasets for learning to learn from few examples. *Proceedings of ICLR 2020*.
- Vinyals, O., Blundell, C., Lillicrap, T., Kavukcuoglu, K. and Wierstra, D. (2016) Matching networks for one shot learning. *Advances in Neural Information Processing Systems*, 29, pp. 3630-3638.
- Wang, W., Zhong, Z., Li, Y., Chen, J. and Yang, J. (2023) ZSLViT: A zero-shot learning approach based on vision transformers. *Proceedings of AAAI 2023*.
- Wei, J., Wang, X., Schuurmans, D. et al. (2022) Chain-of-thought prompting elicits reasoning in large language models. *Advances in Neural Information Processing Systems*, 35.
- Xian, Y., Lorenz, T., Schiele, B. and Akata, Z. (2018) Feature generating networks for zero-shot learning. *Proceedings of CVPR 2018*, pp. 5542-5551.
- Ye, H.J., Hu, H., Zhan, D.C. and Sha, F. (2020) Few-shot learning via embedding adaptation with set-to-set functions. *Proceedings of CVPR 2020*, pp. 8808-8817.

Declarations

Funding

This research was supported by the Spanish Ministry of Science under grant PID2022-136842OB (FewShotAI) and the Austrian Science Fund (FWF) under grant P38124 (MetaLearnAI). Computational resources provided by BSC MareNostrum 5 and Vienna Scientific Cluster VSC-5. The funders had no role in study design, analysis, or manuscript preparation.

Conflict of Interest

The authors declare no conflicts of interest.

Data Availability Statement

All meta-learning implementations, few-shot evaluation scripts, CLIP evaluation code, and LLM prompting templates are deposited in the Zenodo repository under DOI: 10.5281/zenodo.19614738. miniImageNet, tieredImageNet, Meta-Dataset, AWA2, CUB, SUN, and FewGLUE are publicly available. Released under MIT License (code) and CC BY 4.0 (results).

Ethical Approval

This study involved computational experiments on publicly available benchmark datasets. No human subjects, personal data, or proprietary data were involved. No institutional ethical approval was required.

Appendix A

Few-Shot and Zero-Shot Method Selection Guide

This appendix provides a decision tree for selecting few-shot and zero-shot approaches based on data availability, domain, and task type.

Step 1: Available Examples

- 1a. Zero examples available --> CLIP zero-shot (76.4% without labels) or LLM zero-shot CoT (NLP)
- 1b. 1-5 examples per class --> Prototypical nets (5-shot) or MAML (1-shot leads); go to Step 2
- 1c. 10-32 examples per class --> LLM in-context (NLP) or fine-tuned CLIP (vision)
- 1d. 100+ examples per class --> Standard fine-tuning likely more efficient than meta-learning

Step 2: Domain Considerations

- 2a. Training and test domain are similar --> Prototypical networks (82.4% 5-shot; simple and fast)
- 2b. Domain shift expected at deployment --> MAML (smallest cross-domain gap: 21.4 pp)
- 2c. Domain completely unknown / diverse --> CLIP or foundation model (best cross-domain: 64.8%)

Step 3: Task Type

- 3a. Image classification (in-domain) --> Prototypical networks; add transductive if test set available
- 3b. Image classification (cross-domain / zero-shot) --> CLIP with ensemble prompting
- 3c. NLP classification (reasoning tasks) --> GPT-3.5 with chain-of-thought (+4-8 pp on reasoning)
- 3d. NLP classification (pattern matching) --> Direct prompting sufficient; CoT adds no benefit