Pushing the Limits of Simple Pipelines for Few-Shot Learning: External Data and Fine-Tuning Make a Difference

Shell Xu Hu, Da Li, Jan Stühmer, Minyoung Kim and Timothy M. Hospedales
Does few-shot classification need fancy meta-learning?

Best meta-learning results in the past 5 years:
- DINO pre-training + ProtoNet (transfer learning)
- ProtoNet with no pre-training
What is our recipe for a transfer learning pipeline?

1. **External data**
2. **Domain A**
   - Class 1
   - Class 2
   - Class 3
3. **Domain B**
   - Class 4
   - Class 5
   - Class 6

- **Pre-trained ViT**
- **Meta-training** (e.g., ProtoNet)
- **Fine-tuning** with validated lr

**Support set**
**Augmented support set**
Questions behind the recipe

• Q1: How does pre-training regime affect few-shot learning?
• Q2: Is ViT better suited for few-shot learning?
• Q3: How to best exploit fine-tuning for meta-testing?
Q1: How does pre-training regime affect FSL?

| Training Configuration | Benchmark Results |
|------------------------|-------------------|
|                        | MD     | miniIN | CIFAR |
| ID    Arch  Pre Train  MetaTr |       |       |       |
| 0     ViT-small DINO (IN1K) - | 67.4   | 97.0   | 79.8  |
| 1     ViT-small DeIT (IN1K) - | 67.5   | 98.8   | 84.6  |
| 2     ResNet50 DINO (IN1K) - | 63.8   | 91.5   | 76.1  |
| 3     ResNet50 Sup. (IN1K) - | 62.4   | 96.4   | 82.3  |
| 4     ViT-small DINO (IN1K) PN | 78.4   | 98.0   | 92.5  |
| 5     ViT-small DEIT (IN1K) PN | 79.3   | 99.4   | 93.6  |
| 6     ViT-small - PN  | 52.8   | 49.1   | 59.8  |
| 7     ResNet50 DINO (IN1K) PN | 72.4   | 92.0   | 84.0  |
| 8     ResNet50 Sup. (IN1K) PN | 70.2   | 97.4   | 87.6  |
| 9     ResNet50 - PN  | 62.9   | 72.2   | 68.4  |
| 10    ResNet18 - PN   | 63.3   | 73.7   | 70.2  |
| 11    ViT-base DINO (IN1k) PN | 79.2   | 98.4   | 92.2  |
| 12    ViT-base CLIP (YFCC) PN | 80.0   | 98.1   | 93.2  |
| 13    ViT-base Sup (IN21K) PN | 81.4   | 99.2   | 96.7  |
| 14    ViT-base BEIT (IN21K) PN | 82.8   | 99.0   | 97.5  |
| 15    ResNet50 CLIP (YFCC) PN | 75.0   | 92.2   | 82.6  |

Pre-training alone may be > ProtoNet (PN) baseline

Without pre-training larger networks can be worse: e.g., ResNet50 < ResNet18

Pre-training offers a strong feature to boost PN baseline
Q2: Is ViT better suited for FSL?

| ID | Arch     | Pre Train | MetaTr | Benchmark Results |
|----|----------|-----------|--------|------------------|
| 0  | ViT-small| DINO (IN1K) | -      | 67.4 97.0 79.8    |
| 1  | ViT-small| DeiT (IN1K) | -      | 67.5 98.8 84.6    |
| 2  | ResNet50 | DINO (IN1K) | -      | 63.8 91.5 76.1    |
| 3  | ResNet50 | Sup. (IN1K) | -      | 62.4 96.4 82.3    |
| 4  | ViT-small| DINO (IN1K) | PN     | 78.4 98.0 92.5    |
| 5  | ViT-small| DEIT (IN1K) | PN     | 79.3 99.4 93.6    |
| 6  | ViT-small| -         | PN     | 52.8 49.1 59.8    |
| 7  | ResNet50 | DINO (IN1K) | PN     | 72.4 92.0 84.0    |
| 8  | ResNet50 | Sup. (IN1K) | PN     | 70.2 97.4 87.6    |
| 9  | ResNet50 | -         | PN     | 62.9 72.2 68.4    |
| 10 | ResNet18 | -         | PN     | 63.3 73.7 70.2    |
| 11 | ViT-base | DINO (IN1k) | PN     | 79.2 98.4 92.2    |
| 12 | ViT-base | CLIP (YFCC) | PN     | 80.0 98.1 93.2    |
| 13 | ViT-base | Sup (IN21K) | PN     | 81.4 99.2 96.7    |
| 14 | ViT-base | BEIT (IN21K) | PN     | 82.8 99.0 97.5    |
| 15 | ResNet50 | CLIP (YFCC) | PN     | 75.0 92.2 82.6    |

ViT-small > ResNet50

Yes, DINO ViT yields a stronger FSL baseline

Better foundation models make the baseline stronger
Q3: How to best exploit fine-tuning for meta-testing?

| M | Arch     | PreTr | MetaTr | MetaTe   | Avg   | Out-D  |
|---|----------|-------|--------|----------|-------|--------|
| 1 | ViT-small| DINO  | PN (IN)| PN       | 68.380| 67.679 |
| 2 | ViT-small| DINO  | PN (IN)| PN+FT(lr=0.01) | 76.051| 76.536 |
| 3 | ViT-small| DINO  | PN (IN)| PN+FT(lr=0.001) | 74.469| 74.509 |
| 4 | ViT-small| DINO  | PN (IN)| PN+FT(Tuned)  | 77.532| 77.848 |
| 5 | ViT-small| DINO  | PN (MD)| PN       | 78.428| 55.705 |
| 6 | ViT-small| DINO  | PN (MD)| PN+FT(lr=0.01) | 76.094| 73.26  |
| 7 | ViT-small| DINO  | PN (MD)| PN+FT(lr=0.001) | 74.642| 69.965 |
| 8 | ViT-small| DINO  | PN (MD)| PN+FT(Tuned)  | 83.133| 75.72  |

Fine-tuning during meta-testing improves substantially for Out-D.

Validating the best learning rate for each domain is important.

Fine-tuning of Out-D ≈ meta-training of In-D.
## Comparison with SOTA: Meta-Dataset

| 8 in-domain datasets | In-domain | Out-of-domain |
|----------------------|-----------|---------------|
|                      | INet | Omglot | AcrAft | CUB | DTD | QDraw | Fungi | Flower | Sign | COCO | Avg |
| **ProtoNet [60]**    | 67.01 | 44.5   | 79.56  | 71.14 | 67.01 | 65.18 | 64.88 | 40.26  | 86.85 | 46.48 | 63.287 |
| **CNAAPs [52]**      | 50.8  | 91.7   | 83.7   | 73.6  | 59.5  | 74.7  | 50.2  | 88.9   | 56.5  | 39.4  | 66.9 |
| **SUR [54]**         | 56.1  | 93.1   | 84.6   | 70.6  | 71    | 81.3  | 64.2  | 82.8   | 53.4  | 50.1  | 70.72 |
| **Trans. CNAPS [7]** | 57.9  | 94.3   | 84.7   | 78.8  | 66.2  | 77.9  | 48.9  | 92.3   | 59.7  | 42.5  | 70.32 |
| **URT [42]**         | 55.7  | 94.4   | 85.8   | 76.3  | 71.8  | 82.5  | 63.5  | 88.2   | 51.1  | 52.2  | 72.15 |
| **FLUTE [59]**       | 51.8  | 93.2   | 87.2   | 79.2  | 68.8  | 79.5  | 58.1  | 91.6   | 58.4  | 50    | 71.78 |
| **URL [40]**         | 57.51 | 94.51  | 88.59  | 80.54 | 76.17 | 81.94 | 68.75 | 92.11  | 63.34 | 54.03 | 75.749 |
| **ITA [39]**         | 57.35 | **94.96** | 89.33  | 81.42 | 76.74 | **82.01** | 67.4  | 92.18  | 83.55 | 55.75 | 78.069 |

### In-domain = ImageNet

|                      | In-domain | Out-of-domain |
|----------------------|-----------|---------------|
|                      | INet | Omglot | AcrAft | CUB | DTD | QDraw | Fungi | Flower | Sign | COCO | Avg |
| **ProtoNet [60]**    | 50.5  | 59.98  | 53.1   | 68.79 | 65.66 | 48.96 | 39.71 | 85.27  | 47.12 | 41    | 56.099 |
| **ALFA+fo-Proto-MAML [5]** | 52.8 | 61.87  | 63.43  | 69.75 | 70.78 | 59.17 | 41.49 | 85.96  | 60.78 | 48.11 | 61.14 |
| **BOHB [54]**        | 51.92 | 67.57  | 54.12  | 70.69 | 68.34 | 50.33 | 41.38 | 87.34  | 51.8  | 48.03 | 59.152 |
| **CTX [23]**         | 62.76 | 82.21  | 79.49  | 80.63 | 75.57 | 72.68 | 51.58 | **95.34** | 82.65 | 59.9  | 74.281 |
| **DINO > PN > FT (RN50)** | 67.08 | 75.33  | 75.39  | 72.08 | 86.42 | 66.79 | 50.53 | 94.14  | 86.54 | 58.2  | 73.25 |
| **DINO > PN > FT (ViT-small)** | 74.69 | 80.68  | 76.78  | 85.04 | 86.63 | 71.25 | 54.78 | 94.57  | 88.33 | 62.57 | 77.532 |
| **DINO > PN > FT (ViT-base)** | 76.69 | **81.42** | **80.33** | **84.38** | **86.87** | **75.43** | **55.93** | **95.14** | **89.68** | **65.01** | **79.088** |
## Comparison with SOTA: Cross-domain FSL

|                | ChestX | ISIC  | EuroSAT | CropDisease |
|----------------|--------|-------|---------|-------------|
|                | 5w5s   | 5w20s | 5w50s  | 5w5s  | 5w20s | 5w50s | 5w5s  | 5w20s | 5w50s | 5w5s | 5w20s | 5w50s |
| ProtoNet [55]  | 24.05  | 28.21 | 29.32  | 39.57 | 49.5  | 51.99 | 73.29 | 82.27 | 80.48 | 79.72 | 88.15 | 90.81 |
| RelationNet [57] | 22.96  | 26.63 | 28.45  | 39.41 | 41.77 | 49.32 | 61.31 | 74.43 | 74.91 | 68.99 | 80.45 | 85.08 |
| MetaOptNet [38] | 22.53  | 25.53 | 29.35  | 36.28 | 49.42 | 54.8  | 64.44 | 79.19 | 83.62 | 68.41 | 82.89 | 91.76 |
| Finetune [29]  | 25.97  | 31.32 | 35.49  | 48.11 | 59.31 | 66.48 | 79.08 | 87.64 | 90.89 | 89.25 | 95.51 | 97.68 |
| CHEF [1]       | 24.72  | 29.71 | 31.25  | 41.26 | 54.3  | 60.86 | 74.15 | 83.31 | 86.55 | 86.87 | 94.78 | 96.77 |
| STARTUP [47]   | 26.94  | 33.19 | 36.91  | 47.22 | 58.63 | 64.16 | 82.29 | 89.26 | 91.99 | 93.02 | 97.51 | 98.45 |
| DINO > PN > FT (RN50) | 27.13  | 31.57 | 34.17  | 43.78 | 54.06 | 57.86 | **89.18** | **93.08** | **96.06** | **95.06** | **97.25** | **97.77** |
| DINO > PN > FT (ViT-small) | **27.27** | **35.33** | **41.39** | **50.12** | **65.78** | **73.5** | 85.98 | 91.32 | 95.4  | 92.96 | **98.12** | **99.24** |

+ 4.5%  + 7.0%  + 4.0%  + 0.8%
Thank you for your attention!

Please come to visit our poster on **June 22nd at 2:30 PM**
**Session 2.2: transfer / low-shot / long-tail learning**
**ID 110b**