# Adaptive manifold for imbalanced transductive few-shot learning


Michalis Lazarou[1], Yannnis Avrithis[2], Tania Stathaki[1]

[1]Imperial College London [2]Institute of Advanced Research on Artificial Intelligence


## Background

- **Transductive few-shot learning:** Labeled support examples and unlabeled queries are all available at test time.
- Main lines of research include class centroid approaches and data manifold approaches.
- Transductive few-shot learning benchmarks use perfectly class-balanced tasks.
- **Problem:** Several methods exploit this bias by encouraging class-balanced predictions.
- We investigate the more realistic imbalanced transductive few-shot learning setting where the number of queries per class is different.
- **Contributions:** Propose a novel algorithm that combines the merits of both class centroid and data manifold approaches named $(AM)$.
- New state of the art performance on the imbalanced transductive few-shot setting.
- On par or even outperform many state of the art methods in the standard balanced transductive few-shot setting.

## Problem definition

### Pre-training:

- We use publicly available pre-trained networks from published works.
- Base class dataset: $D_{\text{base}} := \{(x_i, y_i)\}_{i=1}^{I}$ where $y_i \in C_{\text{base}}$.
- Network $f_\theta : \mathcal{X} \to \mathbb{R}^d$ is trained on $D_{\text{base}}$.

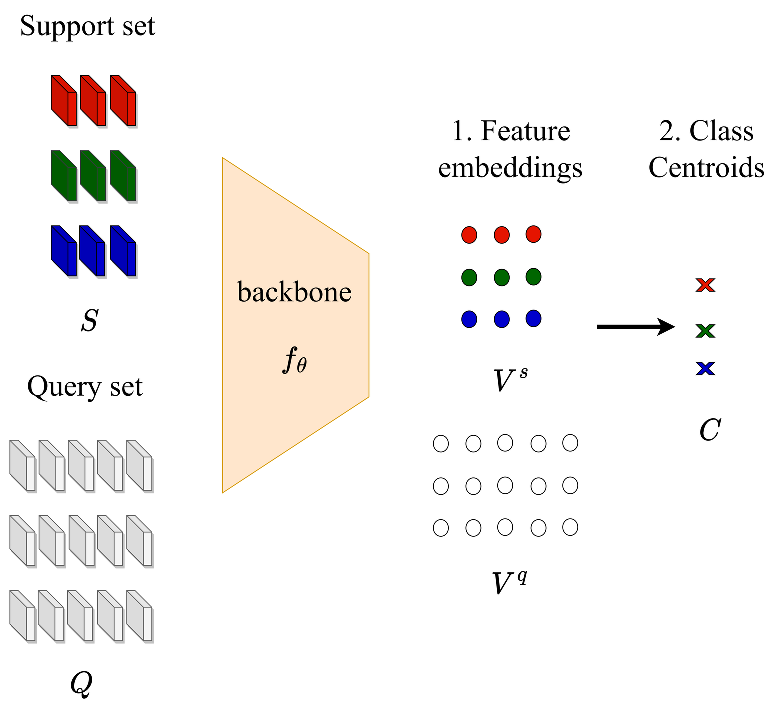
### Inference stage

- Novel class dataset $D_{\text{novel}}$ with $C_{\text{novel}}$ disjoint from $C_{\text{base}}$.
- Assume access to $f_\theta$, a support set, $S$, a query set, $Q$.
- We focus on imbalanced transductive few-shot learning.

## Manifold Centroids

- Embed all examples from $S$ and $Q$ into feature vectors and $\ell_1$-normalize them.
- Calculate the manifold class centroids using the labeled support vectors of every class. For class $j$ the manifold centroid is:

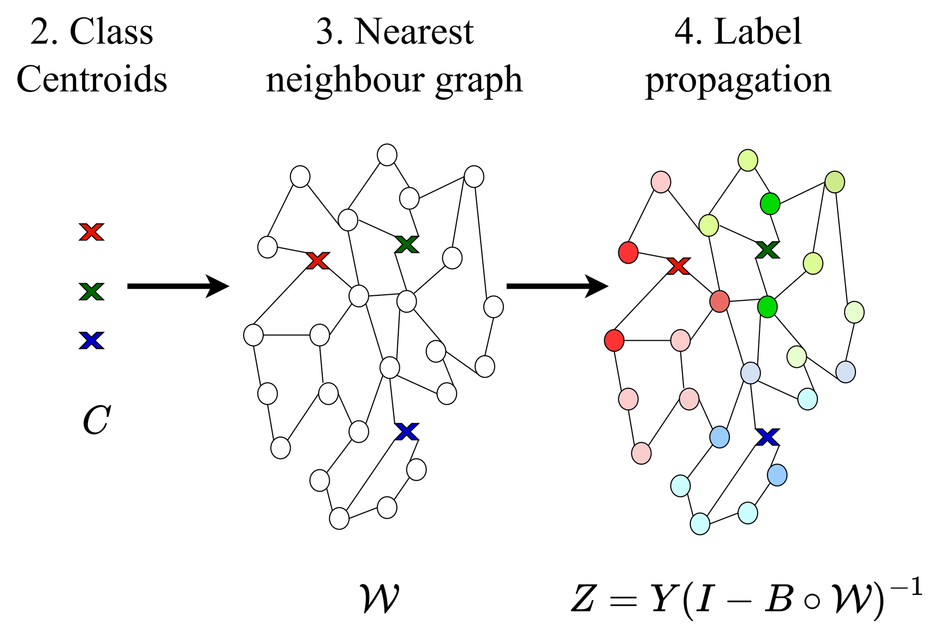
$$\mathbf{c}_j = \frac{1}{K} \sum_{\mathbf{v}_i^s \in \mathcal{V}^s} y_{ji}^s \mathbf{v}_i^s \tag{1}$$




## Label propagation

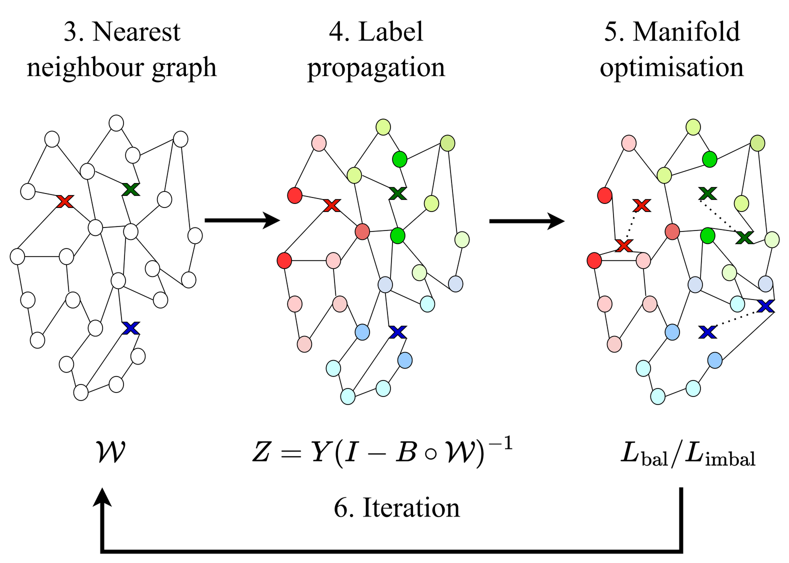
- Construct $k$-nearest neighbour graph, and obtain its adjacency matrix $\mathcal{W}$.
- Define the *label matrix* $Y$.
- Label propagation to obtain a class probability distribution for every query

$$Z := (I - \alpha\mathcal{W})^{-1} Y. \tag{2}$$




## Manifold Adaptation

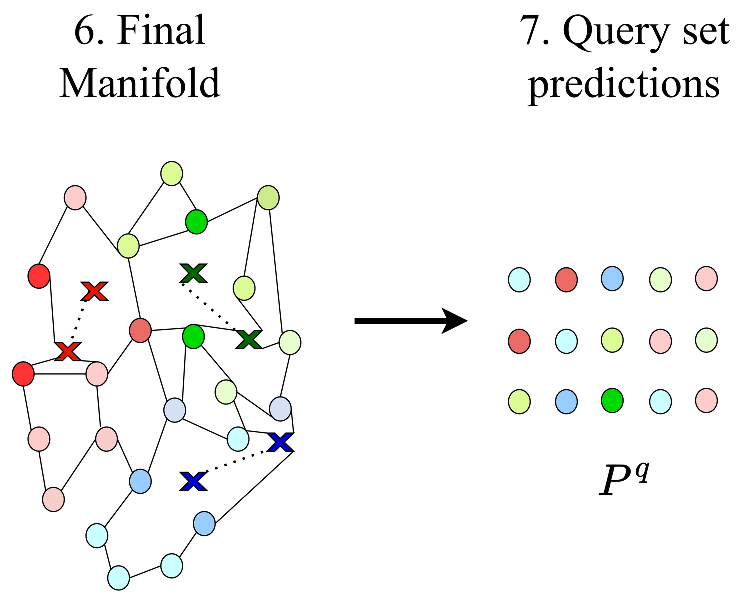
- Iteratively adapt the manifold centroids along with the manifold parameters using either the balanced or imbalanced loss function proposed by [1].




## Query set predictions

- Exploit the final Manifold to make predictions about the queries in the query set $Q$.
- Every query is classified to the class with the highest manifold similarity.

$$\hat{y}_i^q = \arg\max_j p_{ji}^q \tag{3}$$




## Implementation details

- Our Implementation is based on PyTorch.

## Ablation study

Table: *Ablation study of algorithmic components* of both balanced and imbalanced versions of our method AM on *mini*ImageNet. $\text{NN}_k$: $k$-nearest neighbour graph; otherwise, complete graph. $C$: learnable class centroids. $G$: learnable pairwise scaling factors $G$. $B$: learnable adjacency matrix $B$. PLC: feature pre-processing.

| Components | | | | | Imbalanced | | | | Balanced | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | ResNet-18 | | WRN-28-10 | | ResNet-18 | | WRN-28-10 | |
| $\text{NN}_k$ | $C$ | $G$ | $B$ | PLC | 1-shot | 5-shot | 1-shot | 5-shot | 1-shot | 5-shot | 1-shot | 5-shot |
| | | | | | 60.21±0.27 | 74.24±0.21 | 63.34±0.27 | 76.19±0.21 | 59.09±0.21 | 71.54±0.19 | 62.38±0.21 | 73.46±0.19 |
| ✓ | | | | | 63.95±0.27 | 81.15±0.17 | 67.14±0.27 | 83.40±0.16 | 63.82±0.22 | 80.47±0.15 | 67.22±0.21 | 82.58±0.16 |
| ✓ | ✓ | | | | 68.57±0.28 | 82.69±0.16 | 71.22±0.26 | 84.74±0.16 | 73.43±0.23 | 84.37±0.14 | 75.94±0.22 | 86.55±0.13 |
| ✓ | ✓ | ✓ | | | 70.16±0.29 | 82.62±0.17 | 72.89±0.28 | 84.89±0.16 | 75.59±0.27 | 84.80±0.15 | 78.72±0.25 | 87.11±0.13 |
| ✓ | ✓ | | ✓ | | 69.11±0.29 | 82.97±0.16 | 71.64±0.28 | 85.16±0.15 | 74.85±0.25 | 84.66±0.14 | 77.70±0.23 | 86.91±0.13 |
| ✓ | ✓ | ✓ | ✓ | | **70.24**±0.29 | 82.71±0.17 | **73.22**±0.29 | 85.00±0.16 | 76.06±0.28 | 84.82±0.15 | 79.37±0.26 | 87.12±0.13 |
| ✓ | ✓ | ✓ | ✓ | ✓ | 69.97±0.29 | **83.31**±0.17 | 71.98±0.29 | **85.66**±0.15 | **77.35**±0.27 | **85.47**±0.14 | **80.99**±0.26 | **87.86**±0.13 |

## State of the art comparisons

### Imbalanced transductive:

Table: The results are reported from $\alpha$-TIM. Our reproduction of the imbalanced ProtoLP used the official code.

| Method | *mini*ImageNet 1-shot | *mini*ImageNet 5-shot | *tiered*ImageNet 1-shot | *tiered*ImageNet 5-shot |
|---|---|---|---|---|
| ResNet-18 | | | | |
| PT-MAP | 60.10 | 67.10 | 64.10 | 70.00 |
| LaplacianShot | 65.40 | 81.60 | 72.30 | 85.70 |
| BD-CSPN | 67.00 | 80.20 | 74.10 | 84.80 |
| ProtoLP | 65.42 | 78.48 | 71.12 | 82.51 |
| TIM | 67.30 | 79.80 | 74.10 | 84.10 |
| $\alpha$-TIM | 67.40 | 82.50 | 74.40 | 86.60 |
| $\alpha$-TIM$_{\text{PLC}}$ | 63.38 | 82.80 | 70.17 | 86.82 |
| $\alpha$-AM | **70.24** | 82.71 | **77.28** | 86.97 |
| $\alpha$-AM$_{\text{PLC}}$ | 69.97 | **83.31** | 76.44 | **87.19** |

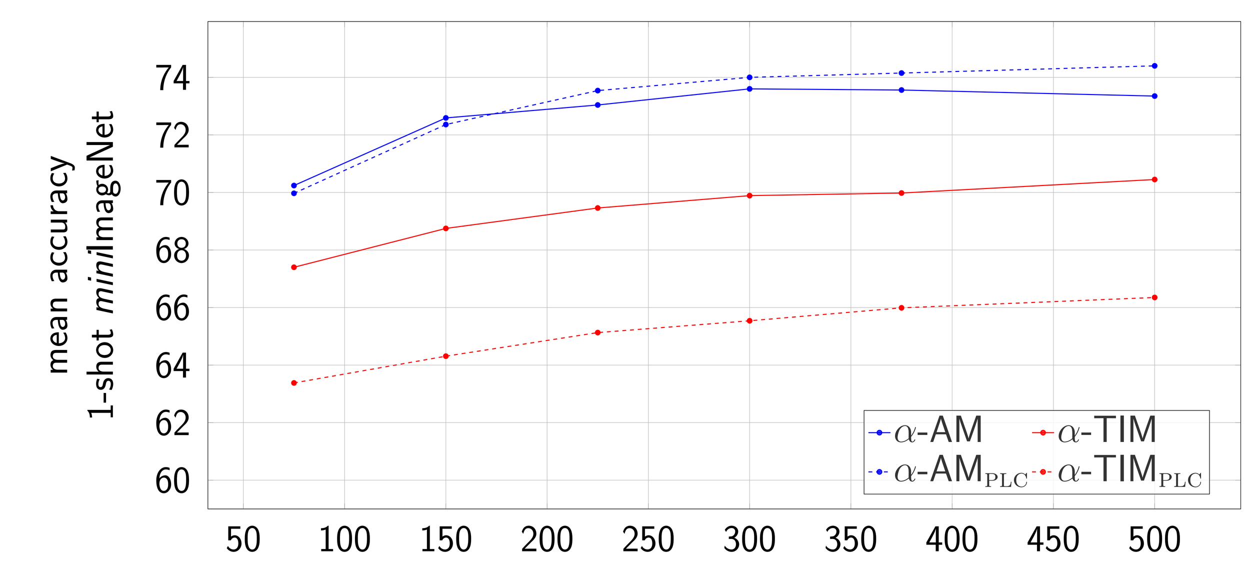
### Balanced transductive:

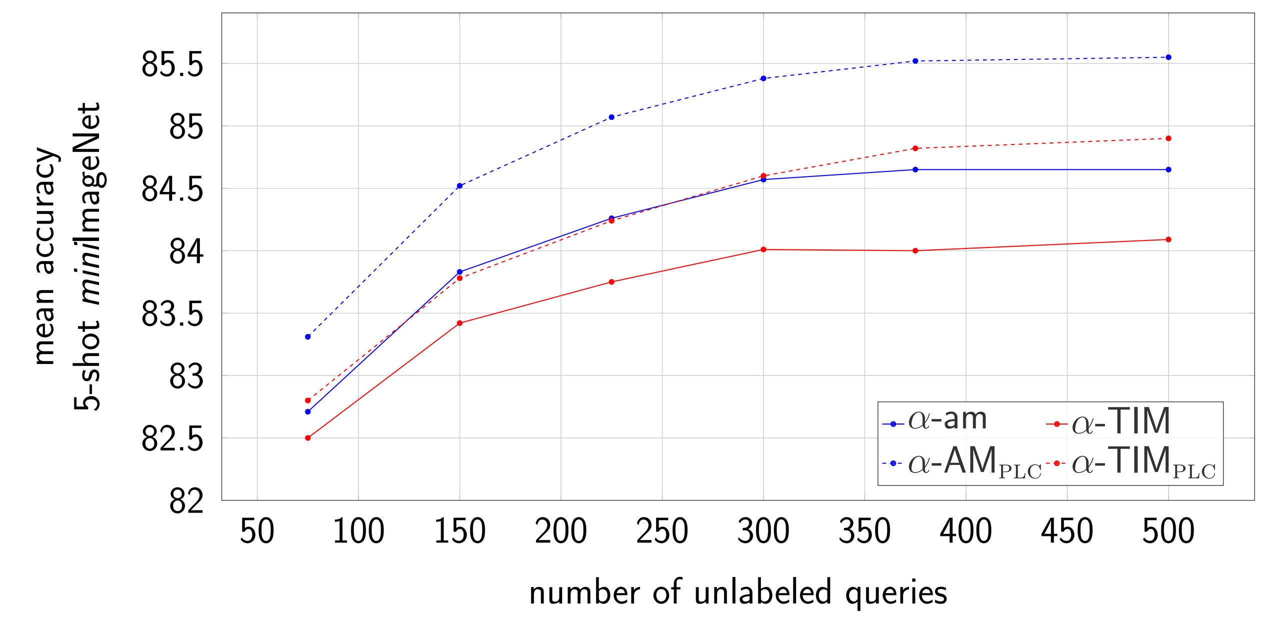
Table: All results were reproduced using the official code provided by $\alpha$-TIM. †: Our reproduction using the official code of protoLP.

| Method | *mini*ImageNet 1-shot | *mini*ImageNet 5-shot | *tiered*ImageNet 1-shot | *tiered*ImageNet 5-shot |
|---|---|---|---|---|
| ResNet-18 | | | | |
| LaplacianShot | 70.24 | 82.10 | 77.28 | 86.22 |
| BD-CSPN | 69.36 | 82.06 | 76.36 | 86.18 |
| PT-MAP | 76.88 | 85.18 | 82.89 | 88.64 |
| protoLP† | 76.96 | 84.90 | 83.06 | 88.55 |
| TIM | 73.81 | 84.91 | 80.13 | 88.61 |
| TIM$_{\text{PLC}}$ | 69.33 | 84.53 | 76.36 | 88.33 |
| AM | 76.06 | 84.82 | 82.42 | 88.61 |
| AM$_{\text{PLC}}$ | **77.35** | **85.47** | **83.40** | **89.07** |

## Using more unlabeled examples

*Effect of number of unlabeled queries $M$ on $\alpha$-iLPC and $\alpha$-TIM using ResNet-18.*




- The performance gap tends to increase as the number of unlabeled queries increase.
- $\alpha$-AM Exloits the data manifold through the k-nearest neighbour graph while $\alpha$-TIM works in Euclidean space.


[1] O. Veilleux, M. Boudiaf, P. Piantanida, and I. Ben Ayed. Realistic evaluation of transductive few-shot learning. *Advances in Neural Information Processing Systems*, 34:9290–9302, 2021.





Contact: michalis.lazarou14@imperial.ac.uk