Self-Attention Message Passing for Contrastive Few-Shot Learning

Prototypical Contrastive Learning with Graphs

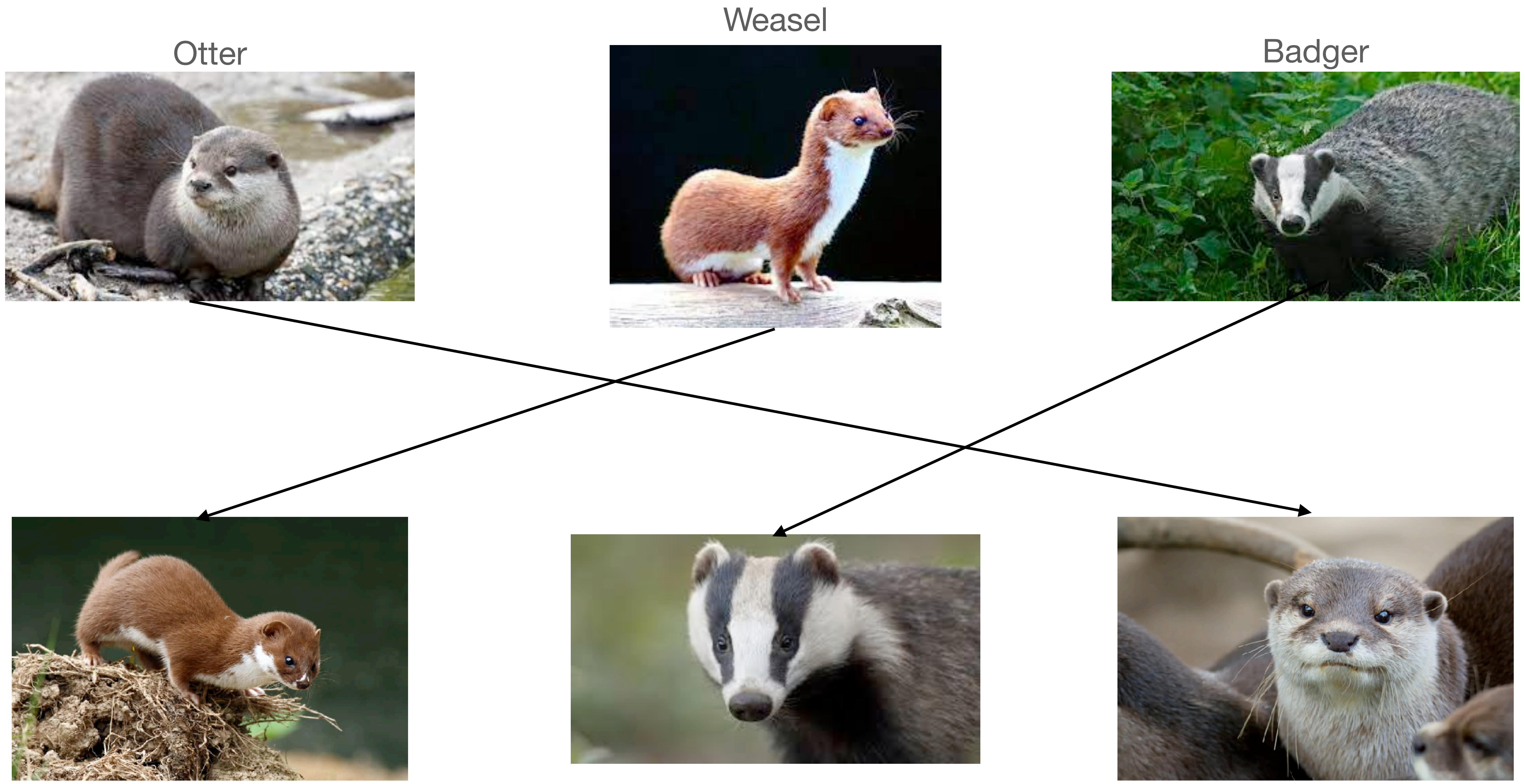
Ojas Shirekar, Anuj Singh, Dr. Hadi Jamali-Rad
TU Delft

Delft University of Technology

$ whoami

- First year PhD Student
  - Started a month ago
  - Supervised by Dr. Chirag Raman
- Previously worked at Shell
  - Graphs for wind power forecasting
- Masters from Delft, 2022

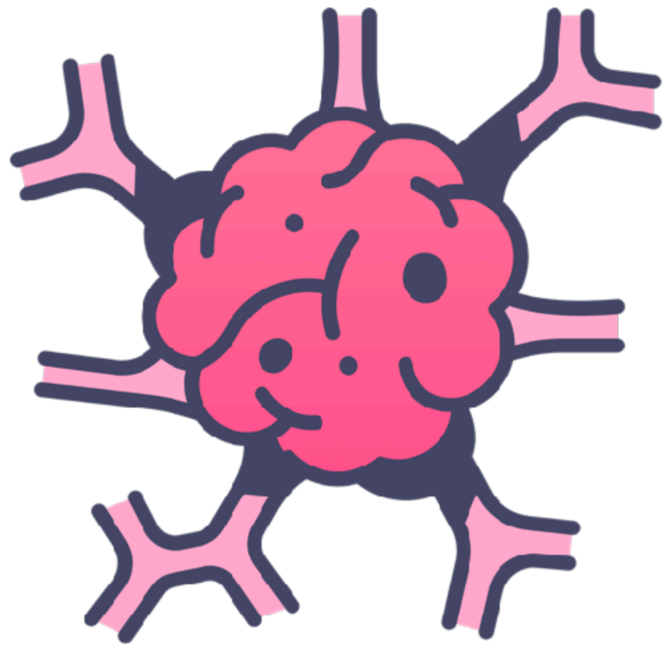
Let’s talk about human understanding



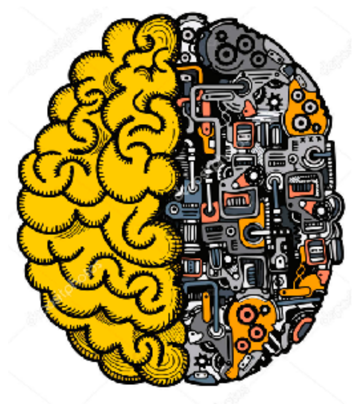
Humans are good at generalising from few examples*.

*Lake, Brenden, et al. "One shot learning of simple visual concepts." *Proceedings of the annual meeting of the cognitive science society*. Vol. 33. No. 33. 2011.

How can machines generalise from few samples?



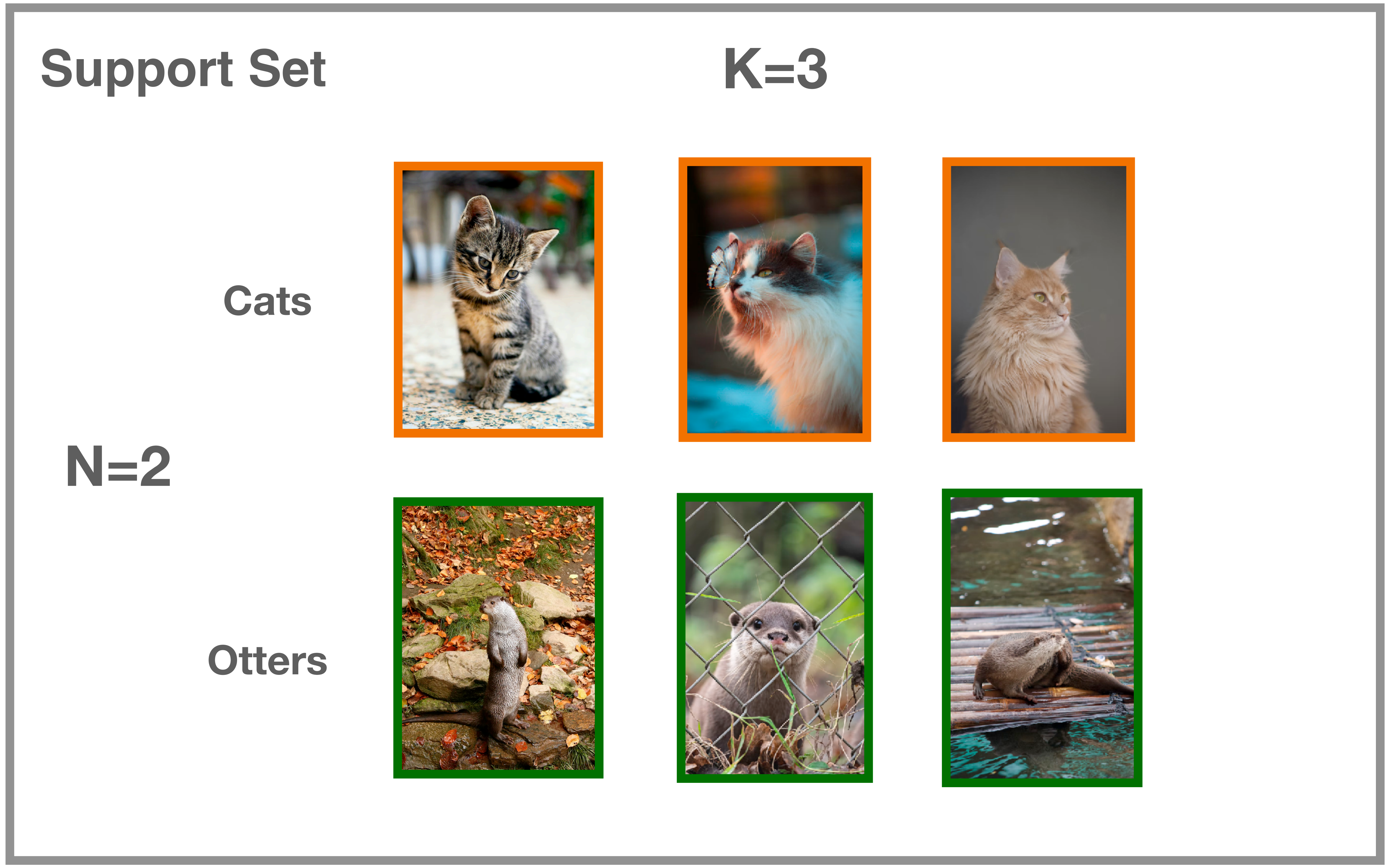
- Sometimes there isn't enough data available
    - Eg: Cancer classification
- Cost of training is high! (Think millions)
- Close the gap between machines and humans
    - **Few Shot Learning!**
- Specifically, generalisation in few-shot **classification** (not AGI)
- In an unsupervised way
    - Closer to human ability (at least that's the assumption)



Few-Shot Classification

In a standard setting

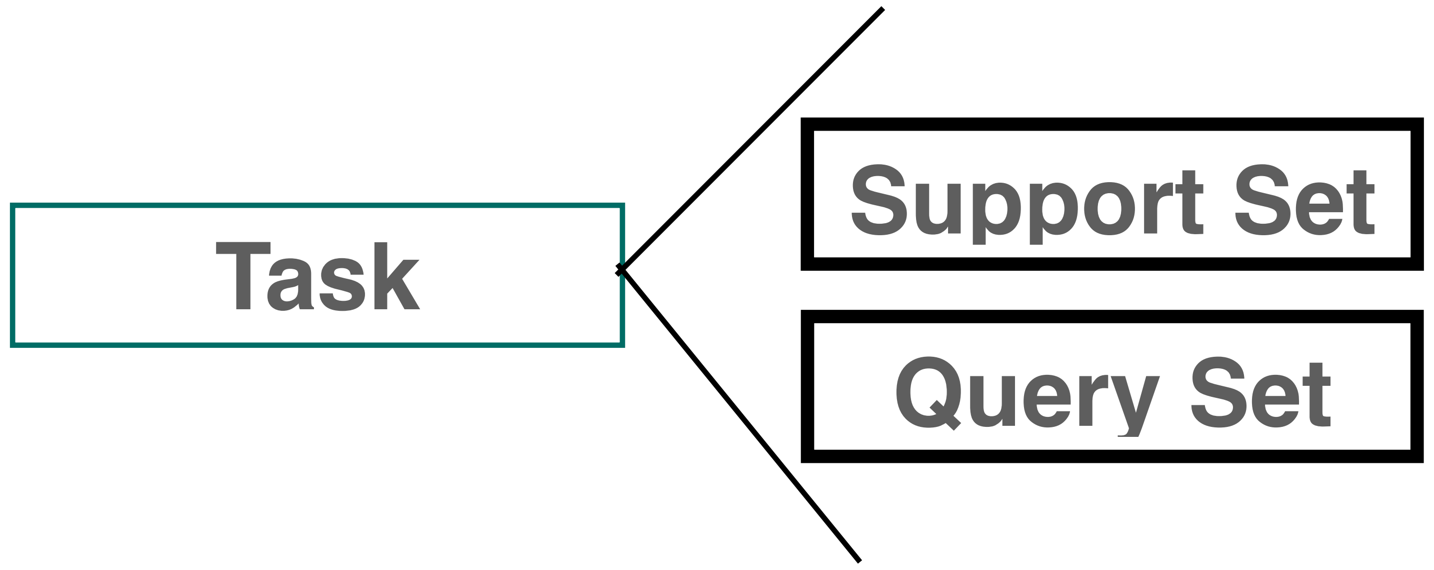
- N-classes
- K-examples per class
- Total images = N * K
- **Support set** = labelled samples
- No labels in unsupervised training



Few-Shot Classification

In a standard Setting

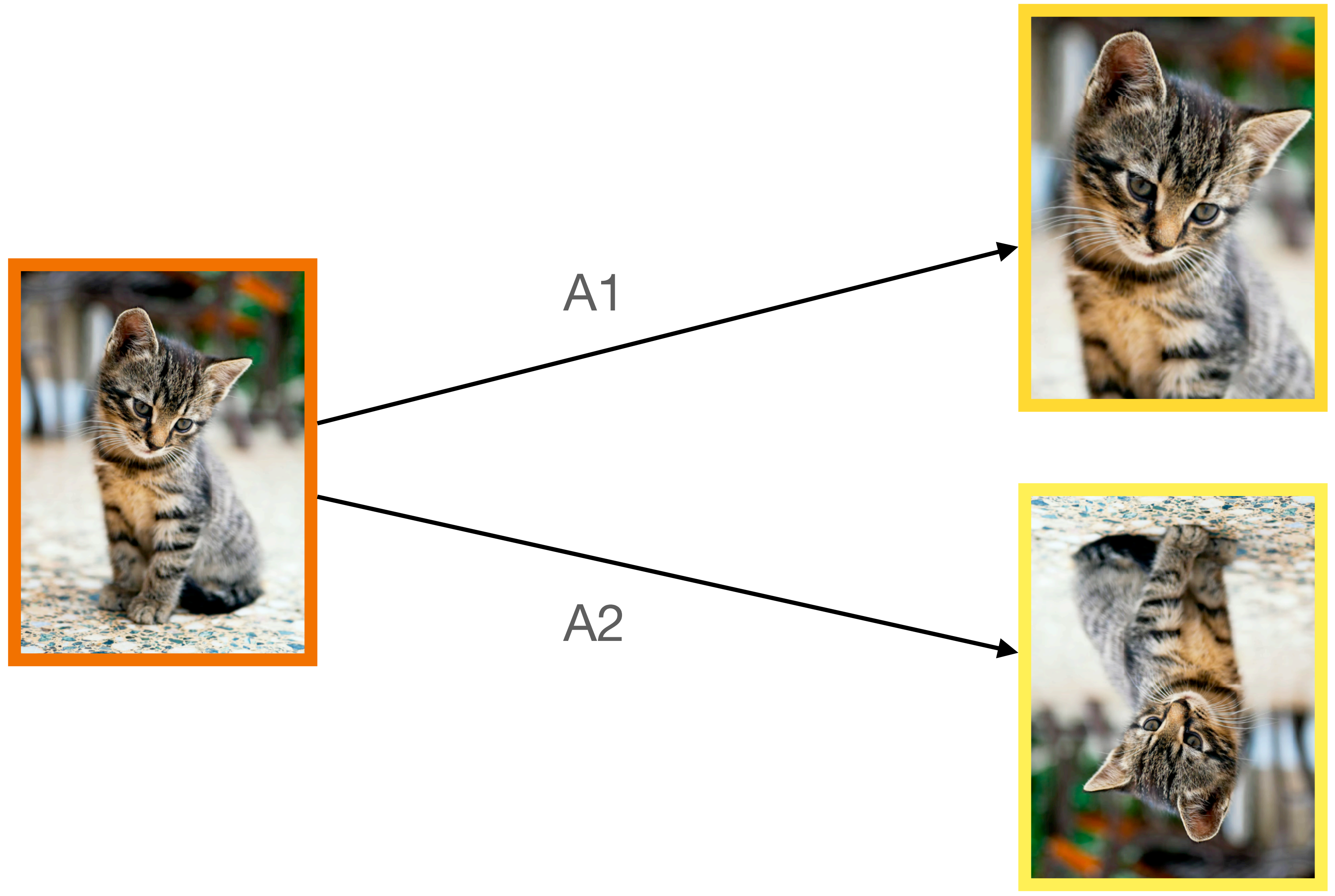
- Model learns from a support set
- We need **query** images that are unlabelled
  - For evaluation after learning from support set
  - This is not a test set!
- Classes in query set ≡ classes in support set
- **Test time classes are completely different from training time**



How do we teach “similarity”?

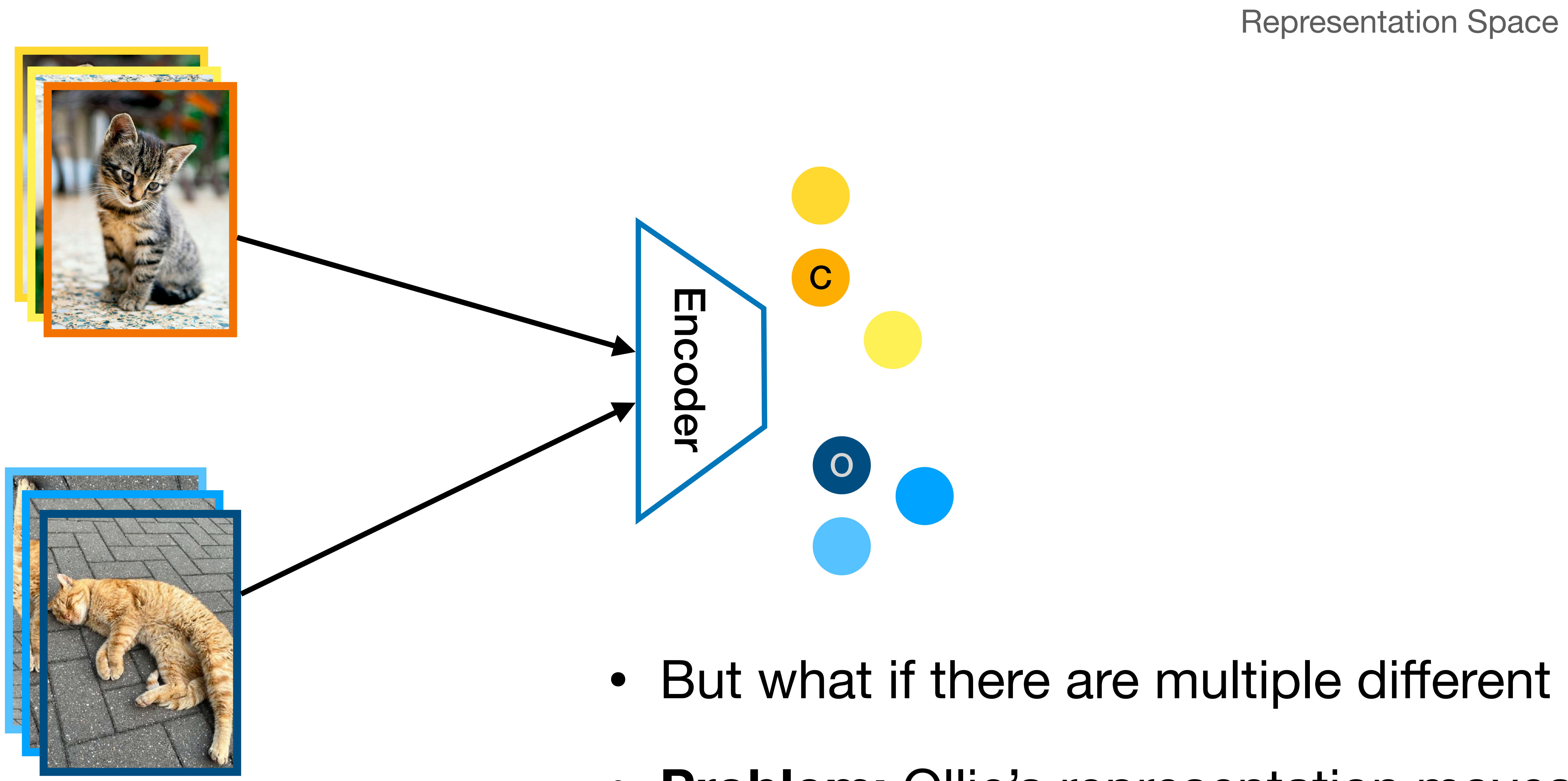
In an unsupervised manner

- We transform an image twice
  - Make network predict similar representation for the two
- **Augmentations** form the core of contrastive learning
- We augment an image twice (or as many as we need)



How do we teach “similarity”?

Contrastive Learning



- But what if there are multiple different cats in a batch?
- **Problem:** Ollie’s representation moves away from the other cat!
- Ideally, we want all cat representations to be in the same vicinity

Solution: SAMPTransfer

What is SAMP Transfer?

- It is a combination of a pre-training and fine-tuning method
  - Tailored for the **few-shot classification** problem
- SAMP = Self Attention based Message Passing
- Uses **graph** refined representations
- Idea: help the network learn relationships in representations of data

Graphs can approximate how humans model the world

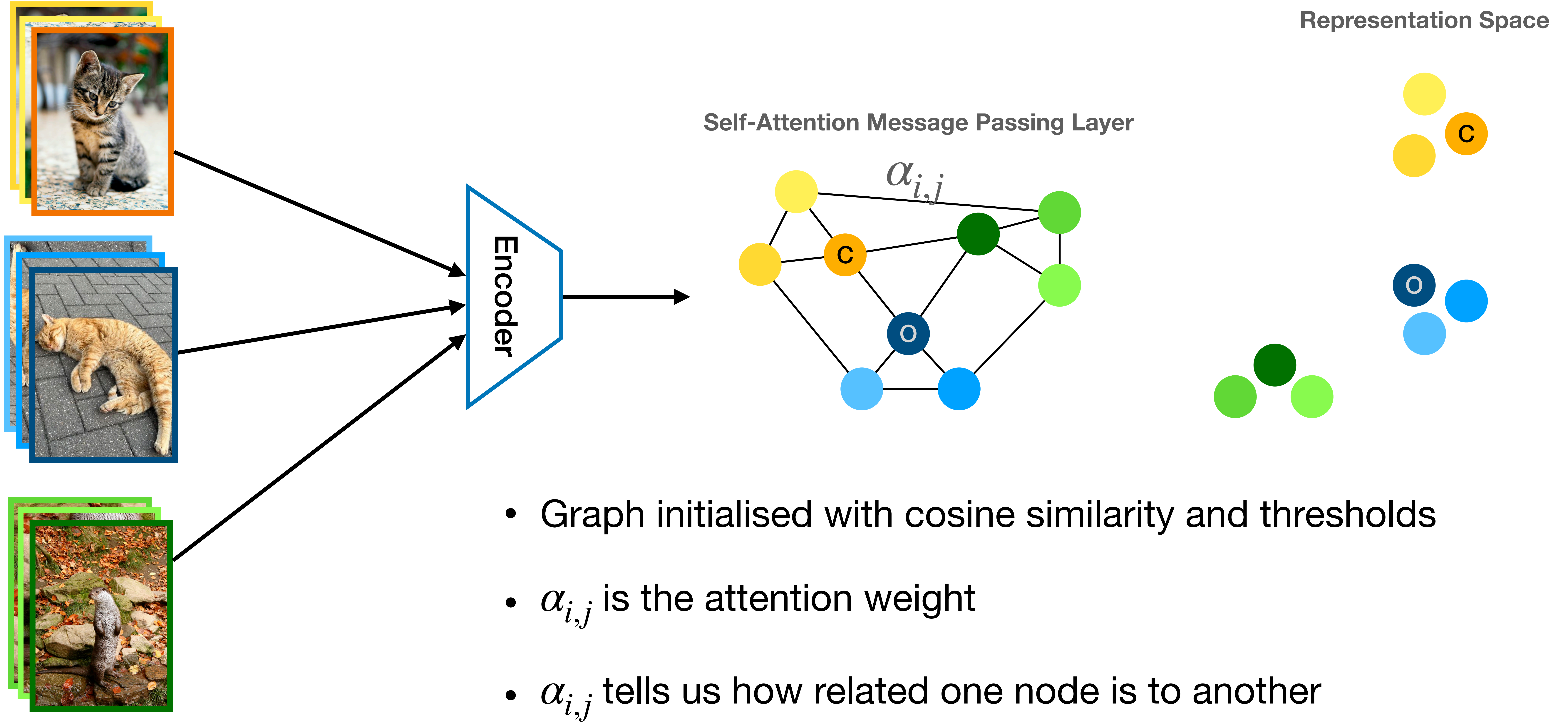
"The image of the world around us, which we carry in our head, is just a *model*. Nobody in his head imagines all the world, government or country. He has only selected *concepts*, and *relationships* between them, and uses those to represent the real system."

Jay Wright Forrester

Inventor of RAM

Network Architecture

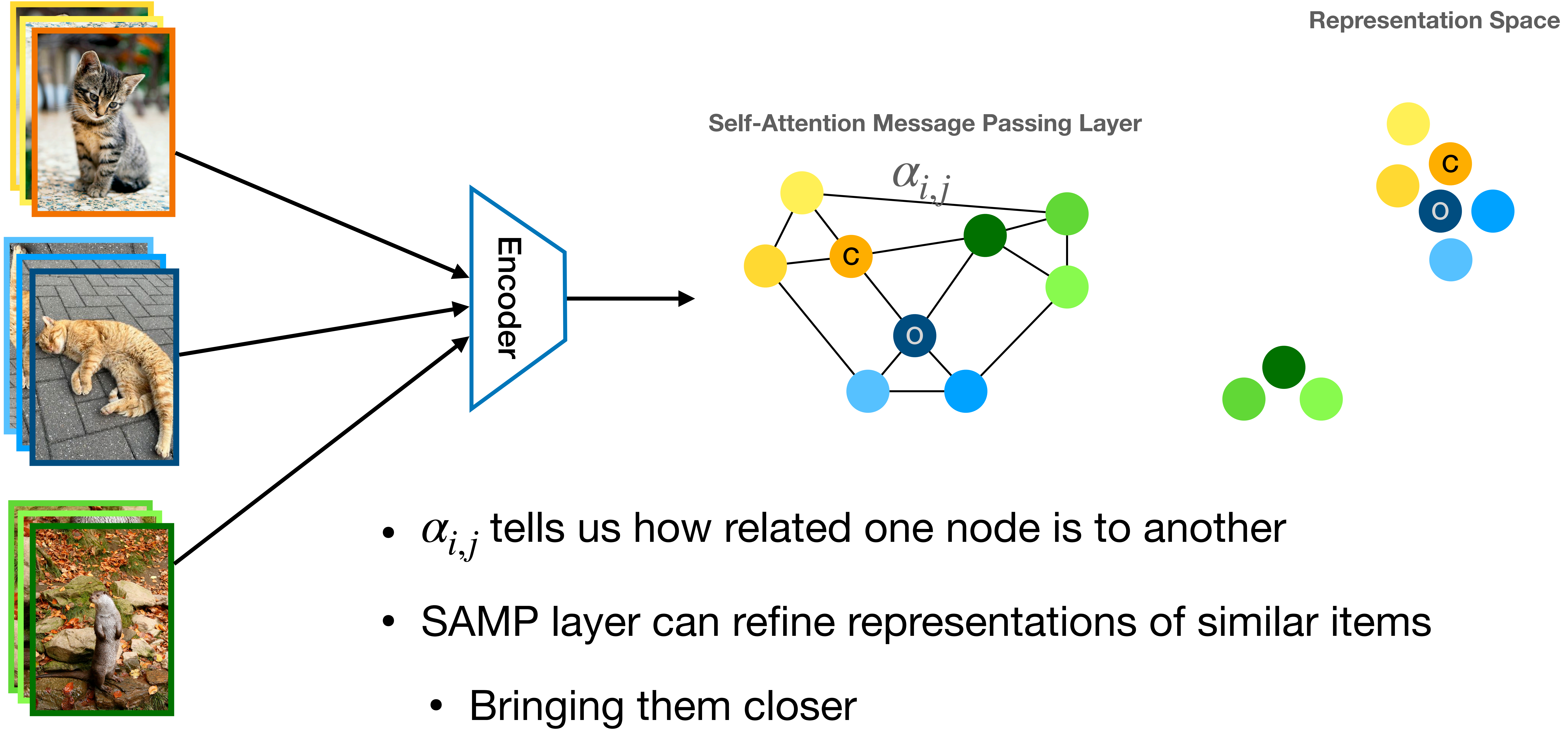
Now in SAMP Transfer



- Graph initialised with cosine similarity and thresholds
- $\alpha_{i,j}$ is the attention weight
- $\alpha_{i,j}$ tells us how related one node is to another

Pre-Training Scheme

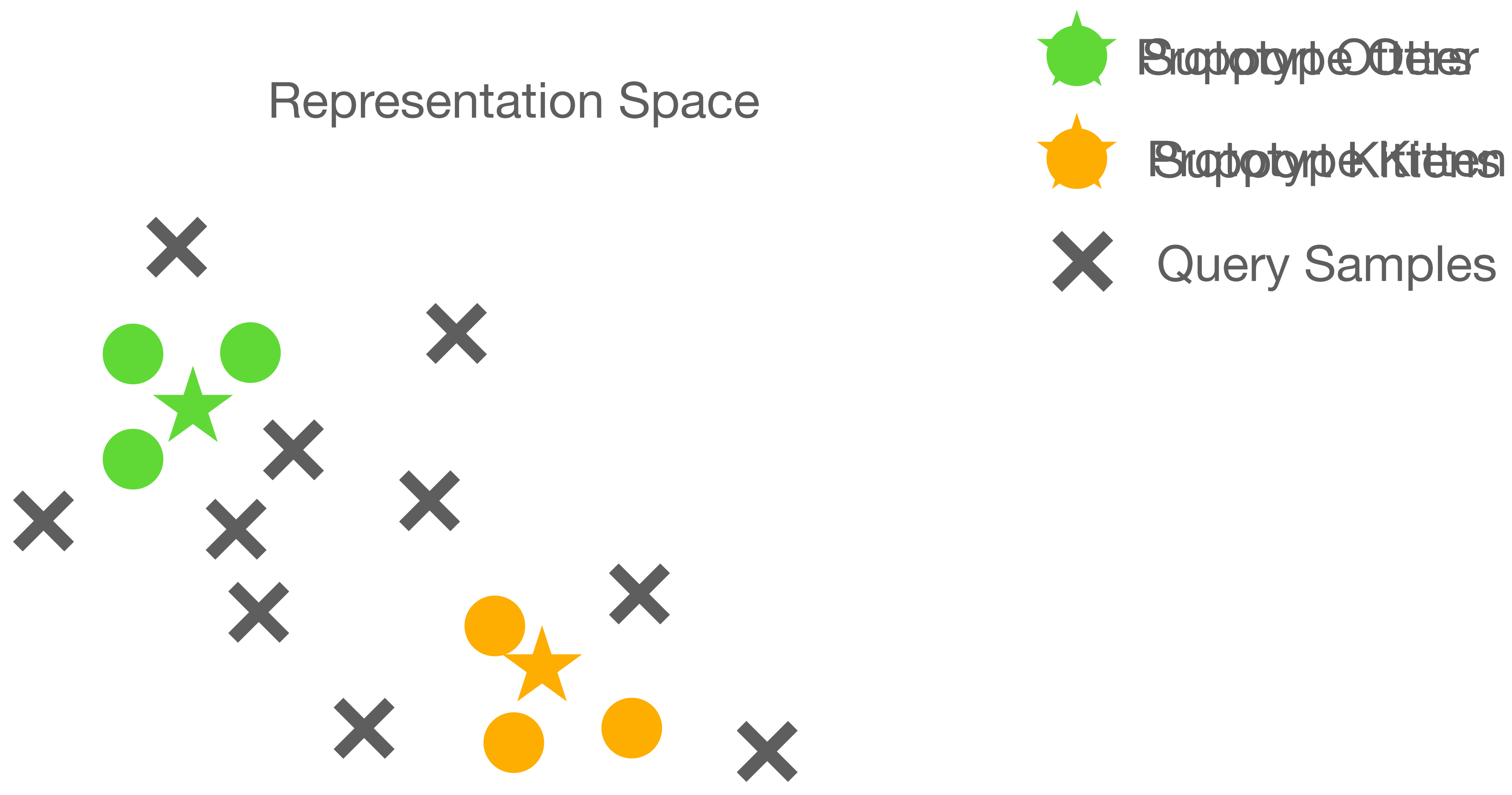
SAMP-CLR



- $\alpha_{i,j}$ tells us how related one node is to another
- SAMP layer can refine representations of similar items
  - Bringing them closer

Classification at test time

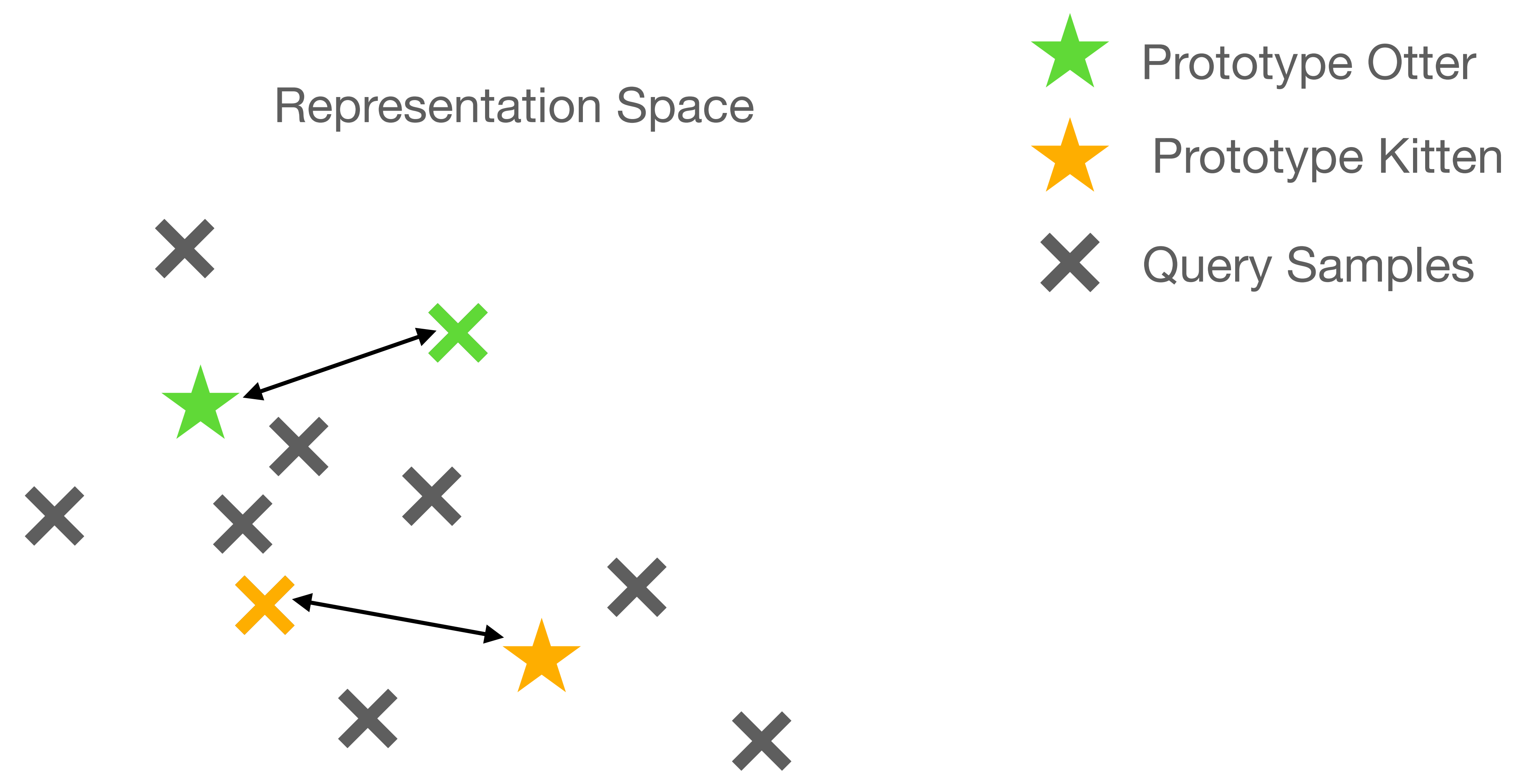
- Support representations are averaged
- Average of the supports is called a **prototype**
- It is a generic representative of a otter or kitten



Classification at test time

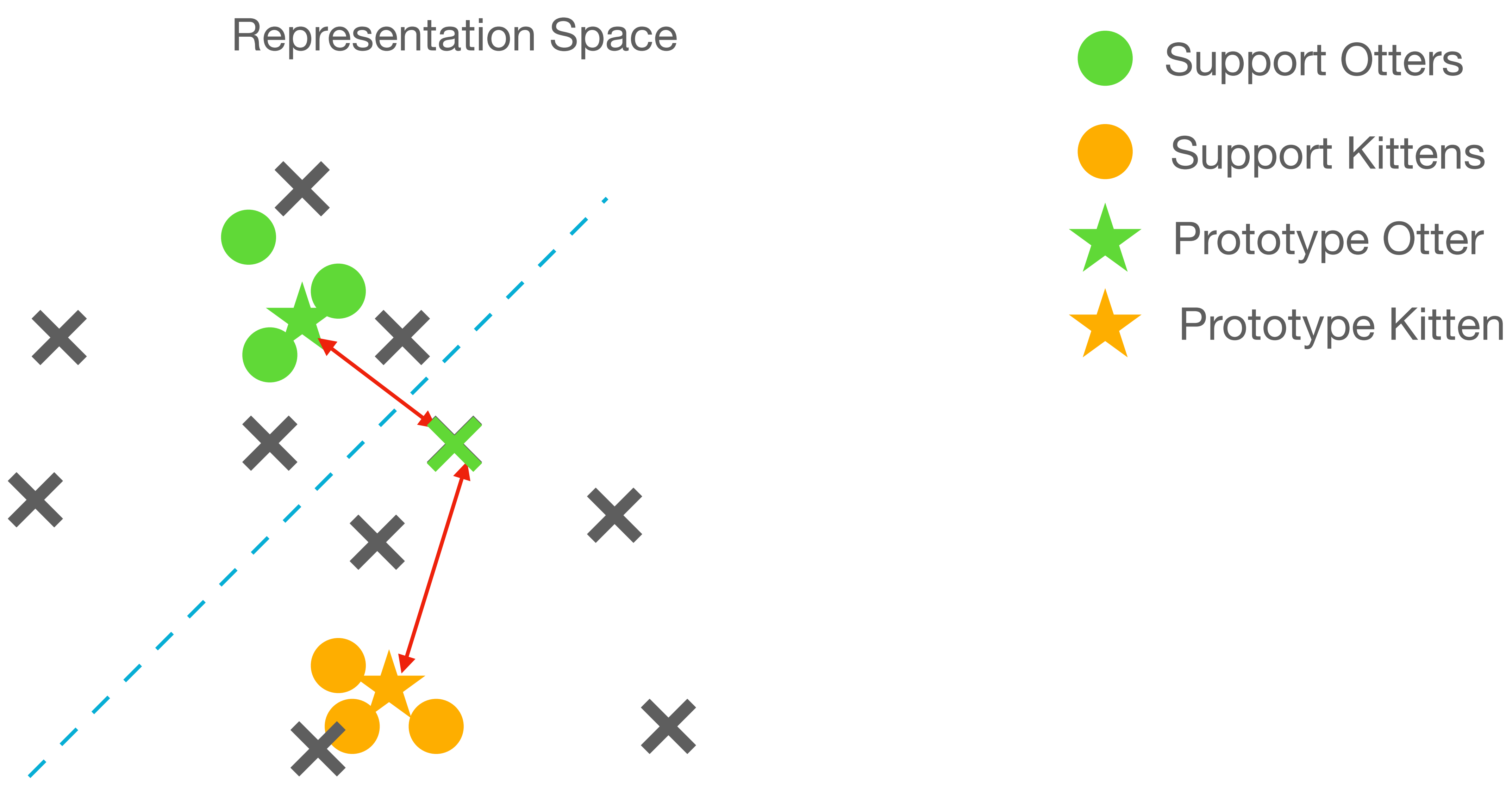
Prototypical Classification

- Distance between prototype and query measured
- Euclidean distance
- If query is closer to the otter prototype it will be tagged as “otter”
- If the query is closer to the kitten prototype it will be tagged as “kitten”



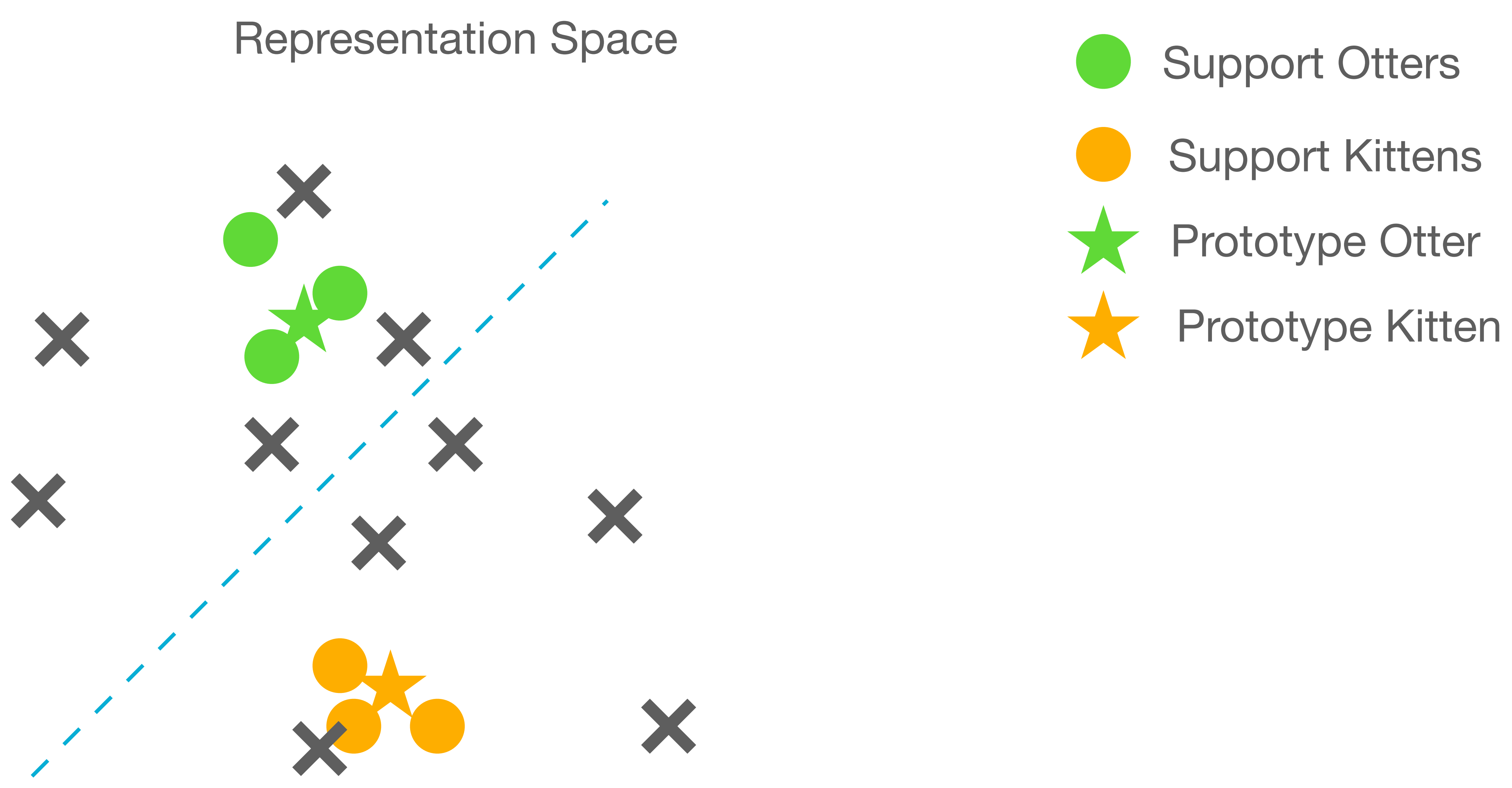
Another problem here

- Supports are **grouped** up in one side
- Prototype is no longer representative
- Leads to misclassification
  - Green cross is actually orange



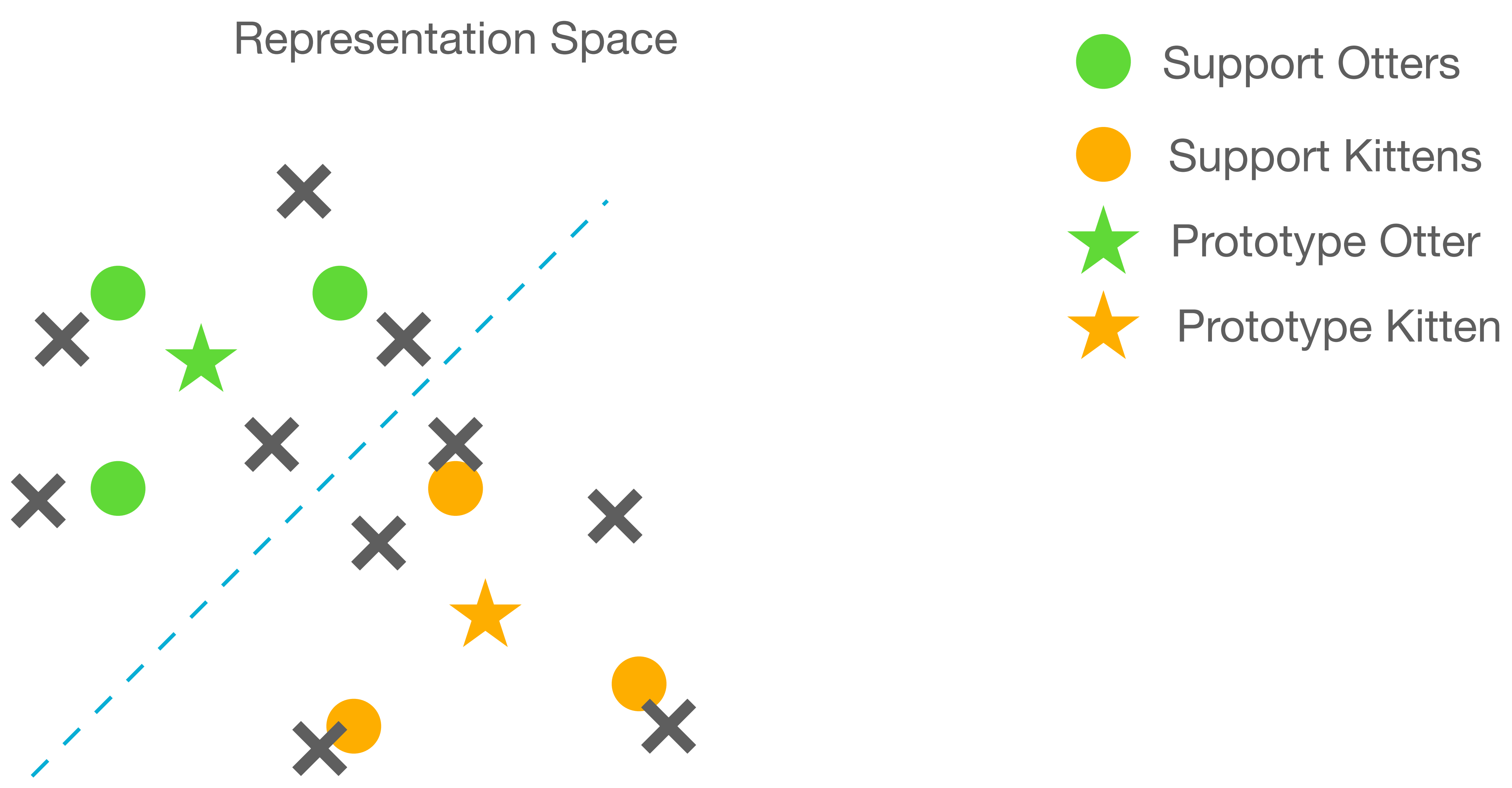
How do we fix it?

OpT-Tune



- We move the supports!
- Using **Optimal Transport**, we call it <u>OpT-Tune</u>
- *Works if features learnt by encoder are good

How do we fix it?



- Increases the spread across the queries
- Gives a better prototype
- **Works only if features learnt by encoder are good

SAMP Transfer Performance

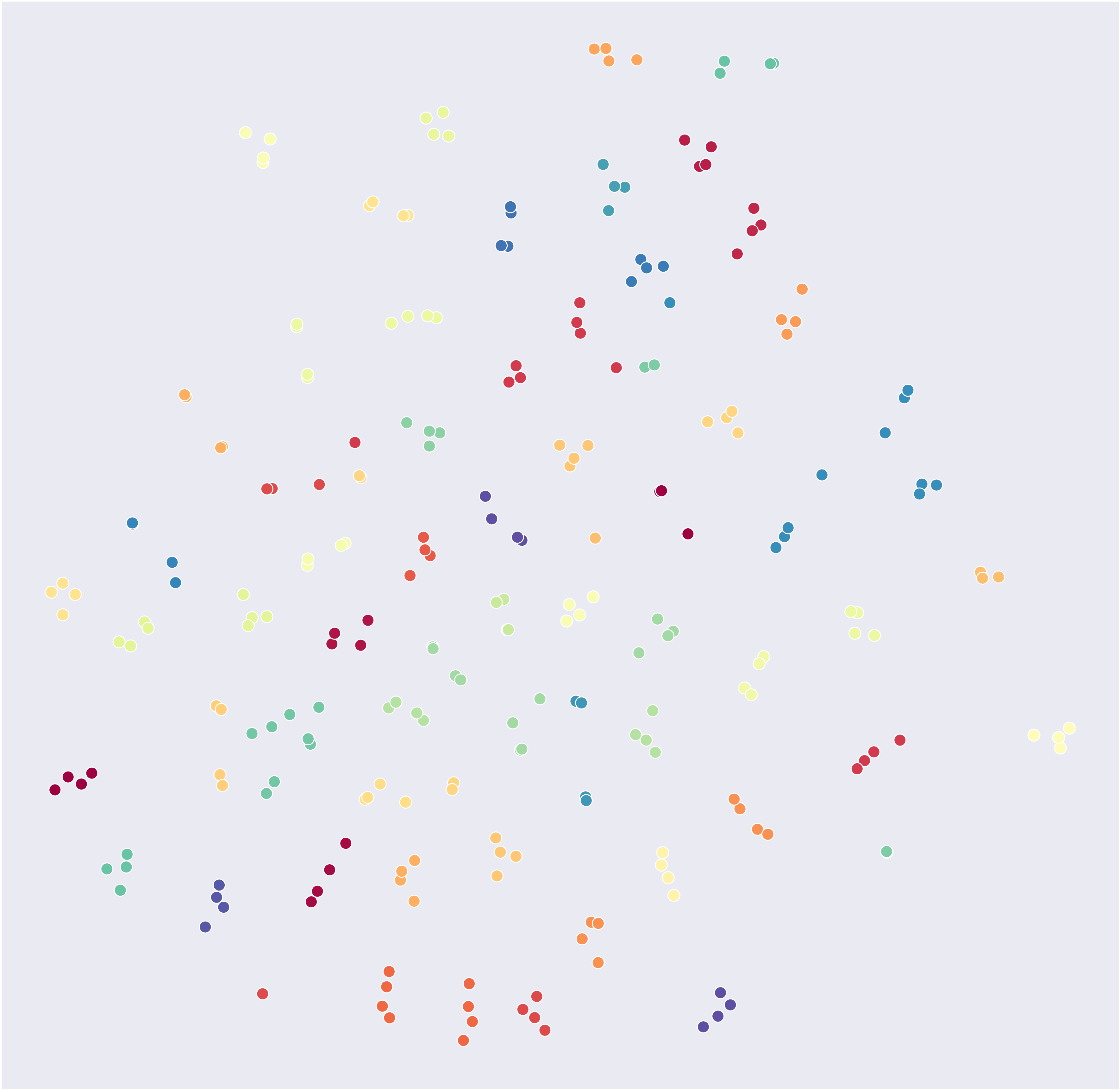
Dataset	Mini-Imagenet	
Method (N,K)	(5, 1)	(5, 5)
Algorithm		
CACTUs-MAML	39.9 +- 0.74	53.97 +- 0.70
CACTUs-ProtoNet	39.18 +- 0.71	53.36 +- 0.70
UMTRA	39.93	50.73
AAL-ProtoNet	37.67 +- 0.39	40.29 +- 0.68
AAL-MAML++	34.57 +- 0.74	49.18 +- 0.47
UFLST	33.77 +- 0.70	45.03 +- 0.73
ULDA-ProtoNet	40.63 +- 0.61	55.41 +- 0.57
ULDA-MetaOptNet	40.71 +- 0.62	54.49 +- 0.58
U-SoSN+ ArL	41.08 +- 0.84	57.01 +- 0.79
U-MISo	41.09	55.38
Meta-GMVAE	42.82	55.73
Prototransfer	45.67 +- 0.79	62.99 +- 0.75
Revisiting UML	48.12 +- 0.19	65.33 +- 0.17
C^3LR	47.92 +- 1.2	64.81 +- 1.15
SAMPTransfer (ours)	**61.02 +- 1.0**	**72.52 +- 0.68**
MetaQDA (supervised)	56.41	72.64
Transductive CNAPS (sup.)	55.60	73.10

SAMPTransfer is on-par with some newer supervised methods

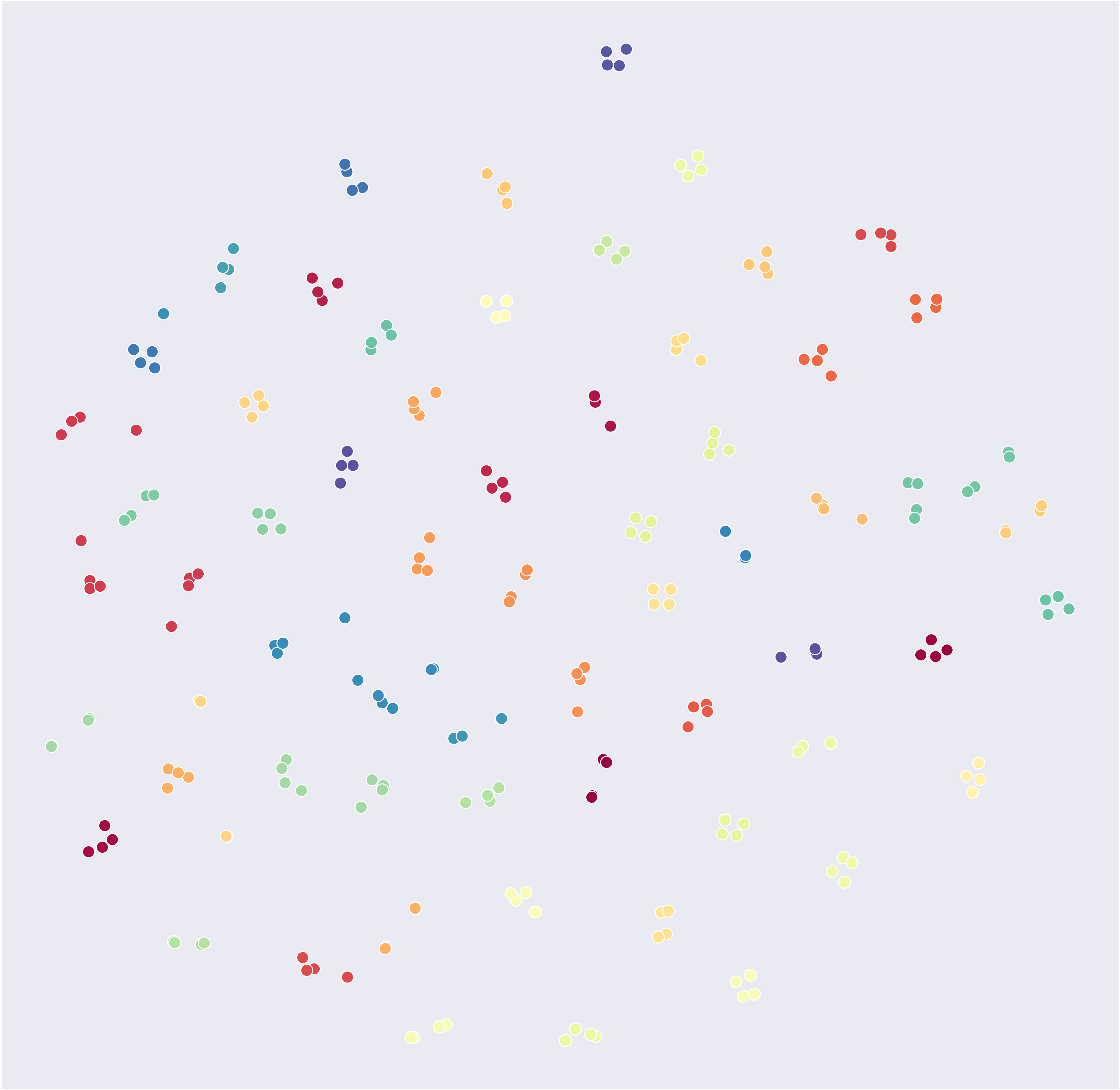
Dataset	Tiered-Imagenet	
Method (N, K)	(5,1)	(5, 5)
Algorithm		
ULDA-ProtoNet	41.60+-0.64	56.28+-0.62
ULDA-MetaOptNet	41.77+-0.65	56.78+-0.63
U-SoSN+ArL	43.68 +-0.91	58.56+-0.74
U-MISo	43.01+-0.91	57.53+-0.74
C^3LR	42.37+-0.77	61.77 +-0.25
SAMPTransfer (ours)	**49.10+-0.94**	**65.19+-0.82**

Visuals

SAMP refinement in action (pre-training)



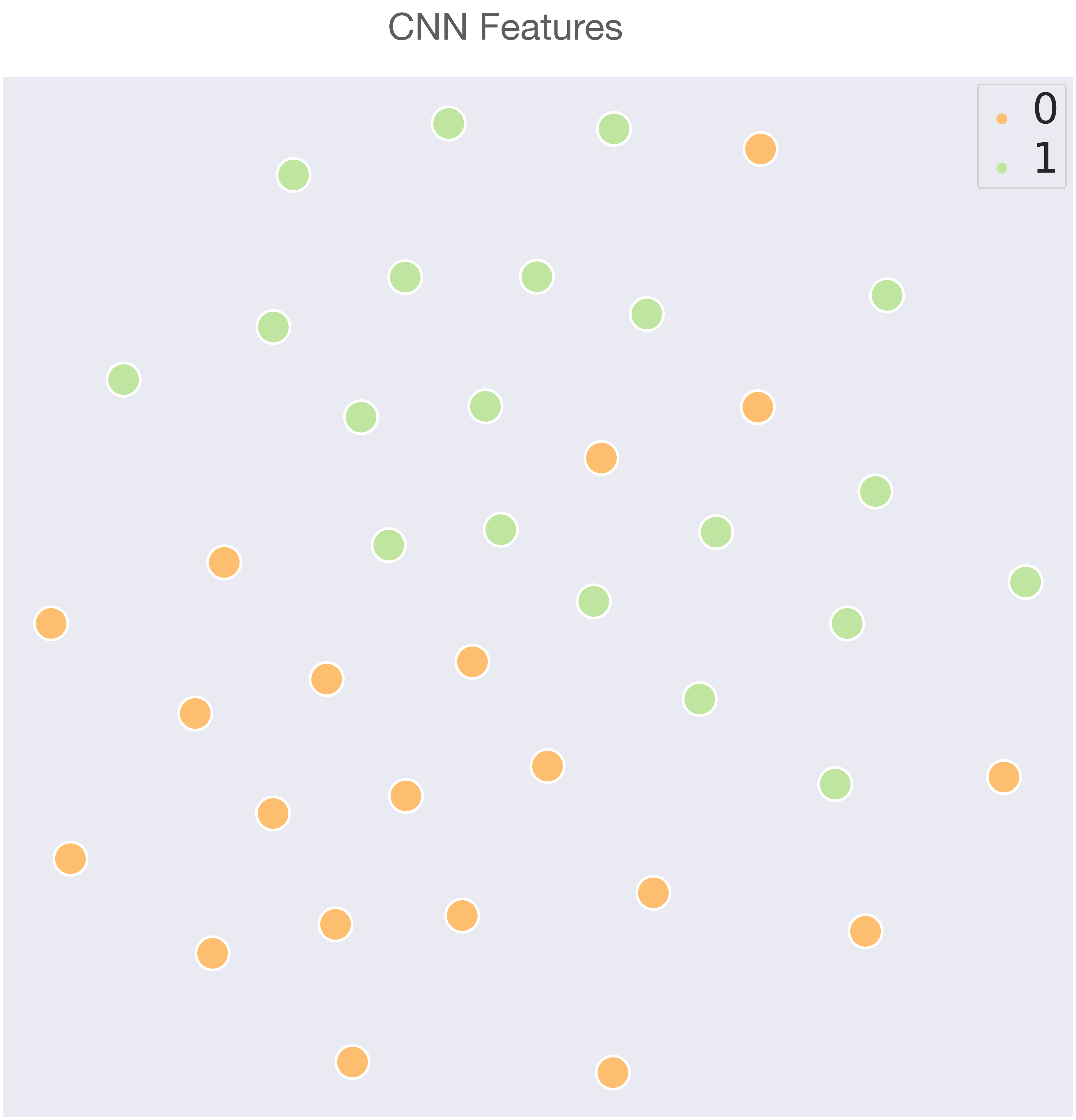
CNN features plotted with UMAP

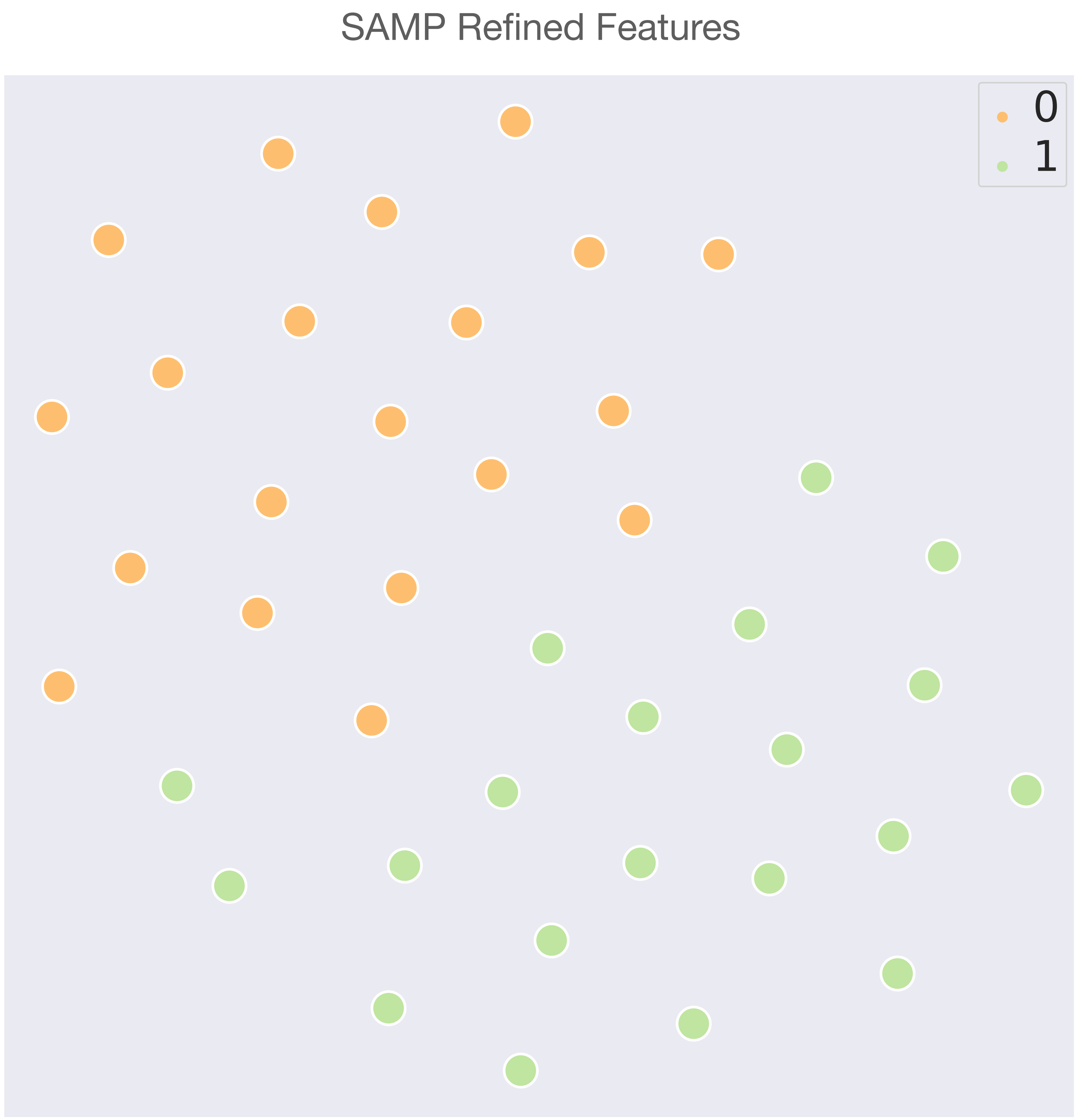


SAMP refined features plotted with UMAP

Visuals

SAMP refinement in action (evaluation 2w-5s)





Recap

- SAMPTransfer
  - Self-attention based message passing
  - Prototypical Classification
  - Optimal Transport to move the supports around
- Analysed performance on common benchmarks and show SOTA performance
- Code is publicly available: https://github.com/ojss/SAMPTransfer/
- Check out our paper: https://arxiv.org/abs/2210.06339

What’s next?

Where are we planning to apply GNNs?

- Chirag and I are at the intersection of **cognitive psychology, sociology and AI**
  - Graphs for modelling agent interactions in social settings
  - Generation of interactive social scenarios
- Any ideas from you?

Extras

OpT-Tune in action