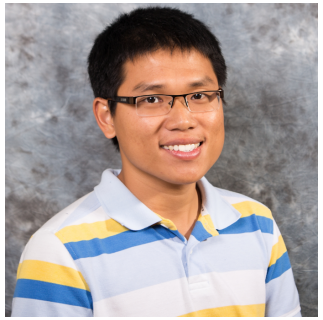
A Self-supervised GAN for Unsupervised Few-shot Object Recognition



Khoi Nguyen Sinisa Todorovic

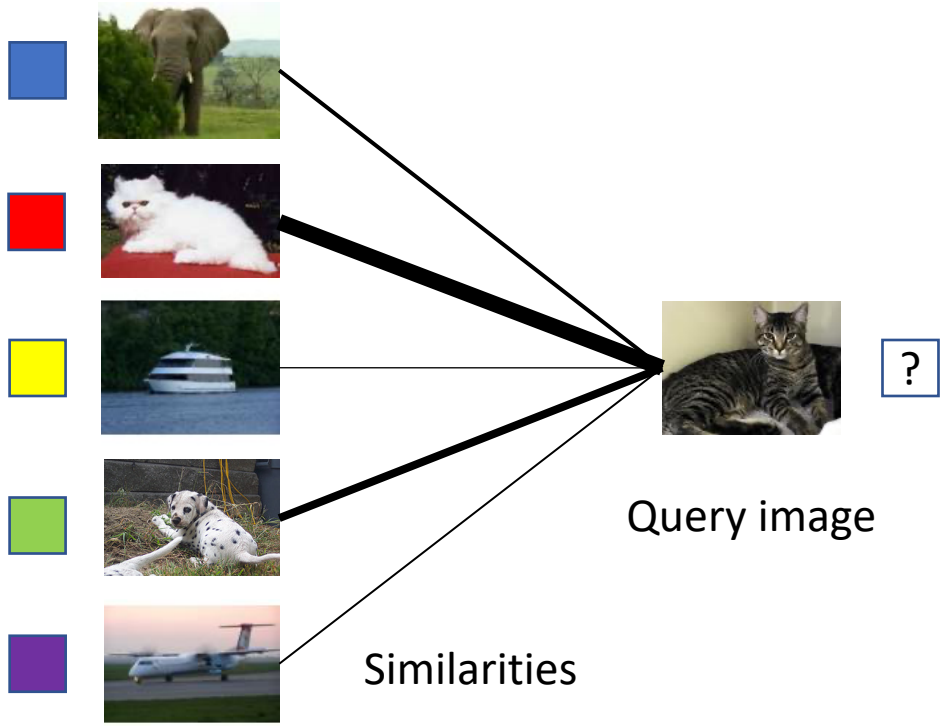
Paper ID: 970

Problem Statement

Training

Unlabeled images

Testing



A few labeled support images

Images from Imagenet dataset: http://image-net.org

Challenges

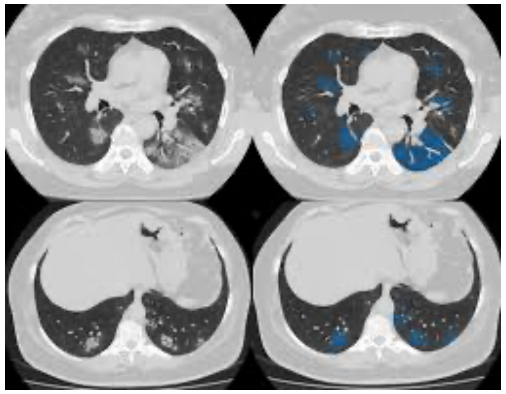
Our setting differs from standard few-shot classification which requires a large set of labeled examples for the episodic training.

Key challenges:

- Learn a deep image representation on unlabeled data such that it would generalize well to unseen classes in testing
- Estimate an accurate similarity between query and support images

Applications

- In domains with very few labeled images per class:
  - Medical images
  - Unsupervised video object segmentation
  - Online tracking



https://medicalxpress.com

https://liuziwei7.github.io/projects/VSReID.html

https://github.com/foolwood/SiamMask

Prior Work

- Generate pseudo labels for training set and apply episodic training as in standard few-shot classification:
  - Image clustering (Hsu et al., ICLR 2019)
  - Data augmentation where each training example is a unique class (Khodadadeh et al., NeurIPS 2019)

➔ Limitations of pseudo labels:

  - Are noisy and not reliable for learning a distance based representation
  - Do not respect the semantic similarity relationship between images

Proposed Unsupervised Training

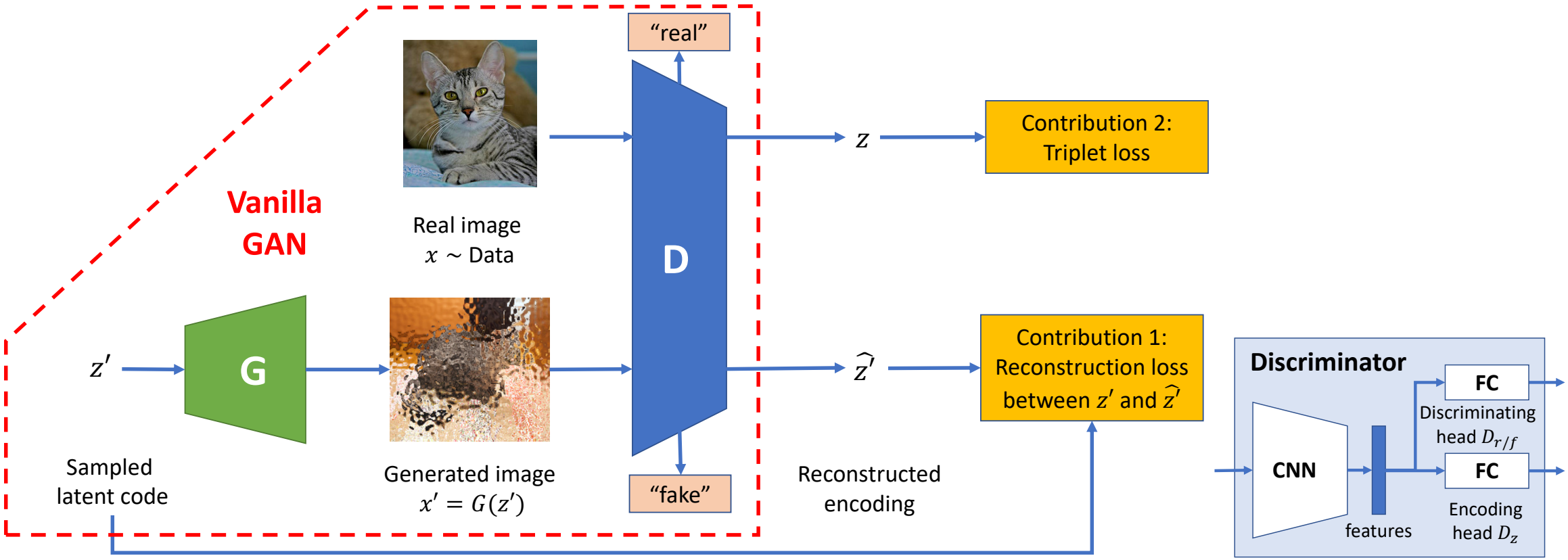
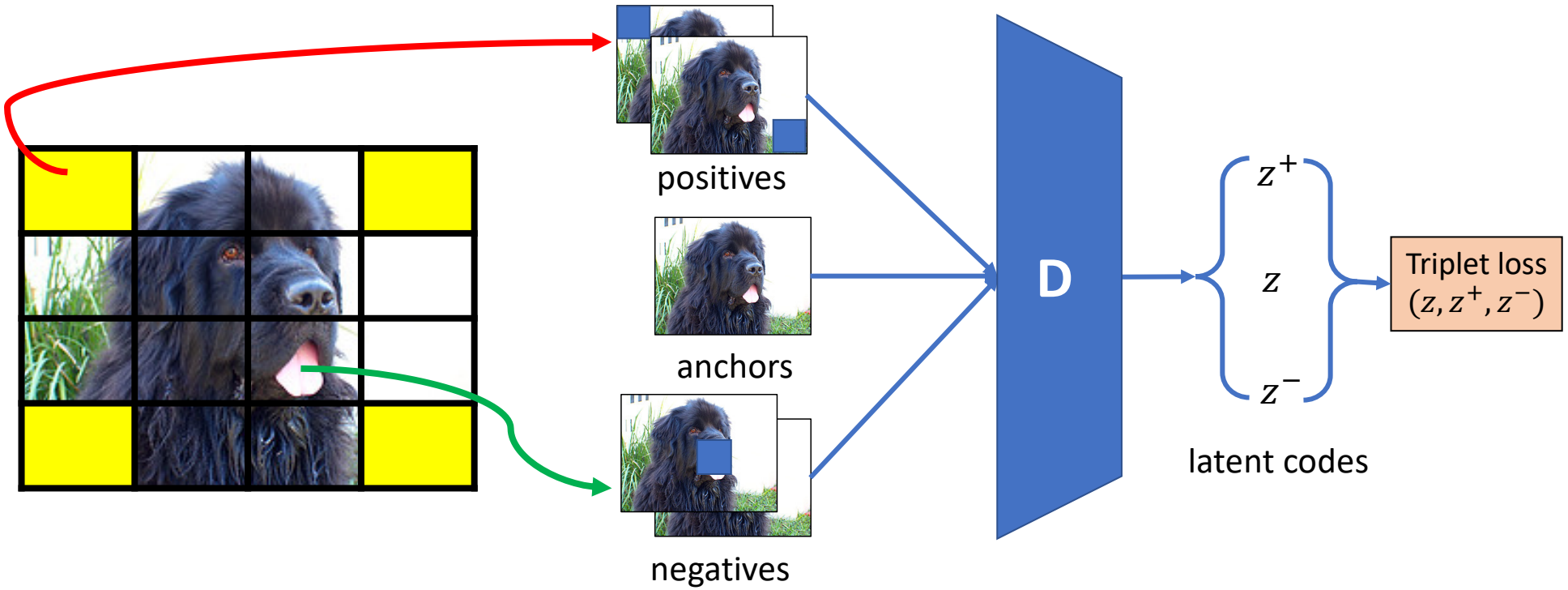
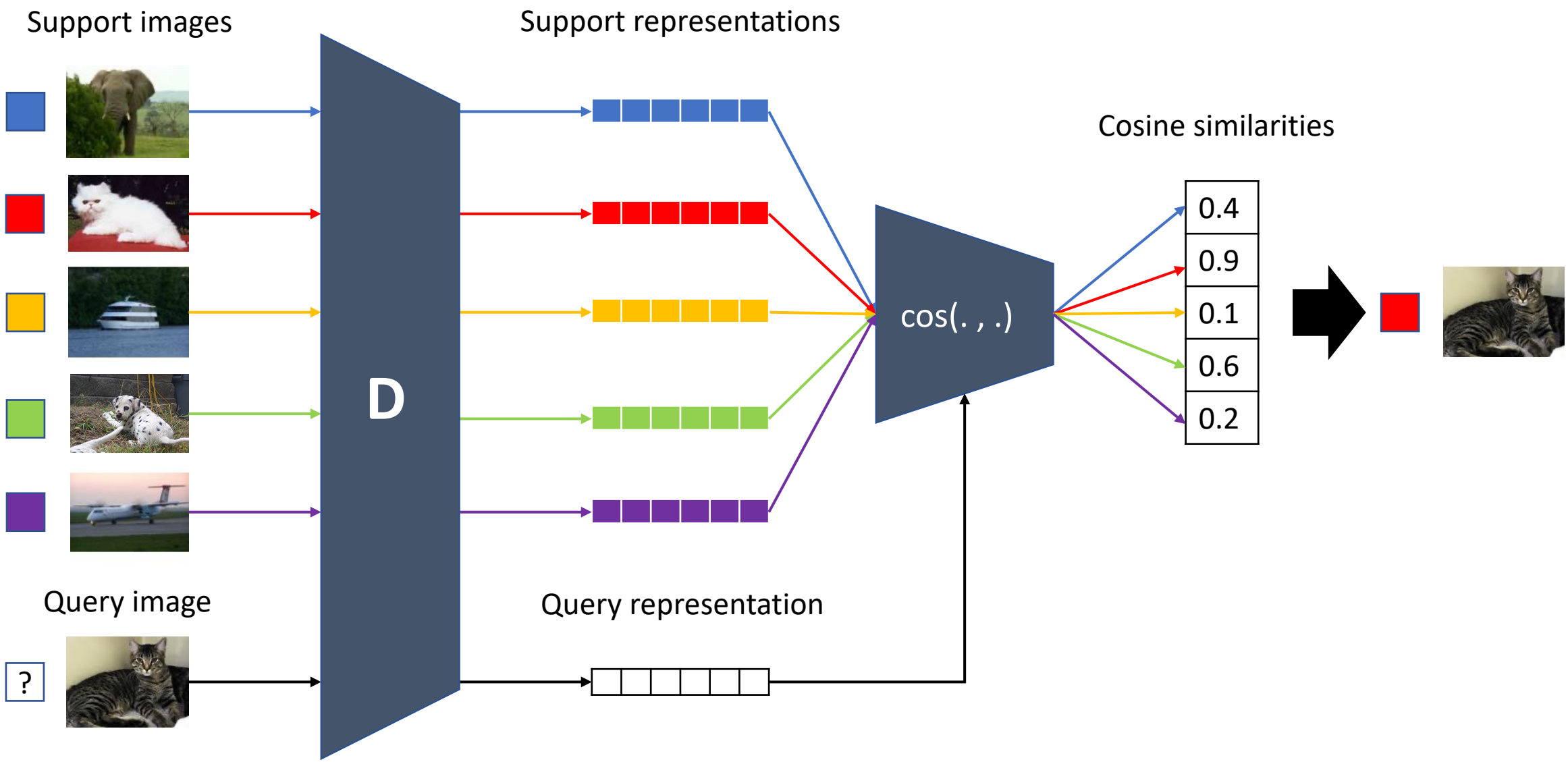


Image Masking and Triplet Loss



Few-shot Classification Testing



Experiments

- Datasets:
  - *Mini-Imagenet*: 100 classes: train: 64, validation: 16, and test: 20 classes
  - *Tiered-Imagenet*: 608 classes grouped into 34 high-level categories: train: 20, validation: 6, and test: 8 categories (minimize semantic similarity between the splits)
- Evaluation metrics:
  - Average N-way K-shot classification accuracy with 95% confidence interval over 1000 episodes where N is the number of classes and K is the number of labeled examples for each class

Few-shot Classification Results

	Mini-Imagenet, 5-way		**Tiered-Imagenet, 5-way**	
Unsupervised Methods	1-shot	5-shot	1-shot	5-shot
SN-GAN (Miyato et al., ICLR 2018)	34.84 $\pm$ 0.68	44.73 $\pm$ 0.67	35.57 $\pm$ 0.69	49.16 $\pm$ 0.70
AutoEncoder (Vincent et al., JMLR 2010)	28.69 $\pm$ 0.38	34.73 $\pm$ 0.63	29.57 $\pm$ 0.52	38.23 $\pm$ 0.72
Rotation (Gidaris et al., ICLR 2018)	35.54 $\pm$ 0.47	45.93 $\pm$ 0.62	36.90 $\pm$ 0.54	51.23 $\pm$ 0.72
BiGAN kNN (Donahue et al,. ICLR 2017)	25.56 $\pm$ 1.08	31.10 $\pm$ 0.63	-	-
AAL-ProtoNets (Antonious et al., Arxiv 2019)	37.67 $\pm$ 0.39	40.29 $\pm$ 0.68	-	-
CACTUs-ProtoNets (Hsu et al., ICLR 2019)	39.18 $\pm$ 0.71	53.36 $\pm$ 0.70	-	-
Our GdBT2	**48.28 $\pm$ 0.77**	**66.06 $\pm$ 0.70**	**47.86 $\pm$ 0.79**	**67.70 $\pm$ 0.75**
Fully-supervised Methods				
ProtoNets (Snell et al., NeurIPS 2017)	46.56 $\pm$ 0.76	62.29 $\pm$ 0.71	46.52 $\pm$ 0.72	66.15 $\pm$ 0.74

Similarity Ranking Relative to the Unmasked Image

Increasing similarity

Conclusions

- We have addressed unsupervised few-shot classification
- Our approach extends the vanilla GAN to integrate adversarial learning with two new strategies for self-supervised learning:
  - Latent code reconstruction loss
  - Image masking with triplet loss
- Outperform SOTAs with significant margin (9% on Mini-Imagenet)
- Beat a fully-supervised few-shot classification approach (ProtoNet)!