IBM Research AI

TEL AVIV UNIVERSITY

Δ-encoder: An effective sample synthesis method for few-shot object recognition

Eli Schwartz*[12], Leonid Karlinsky*[1], Joseph Shtok[1], Sivan Harary[1], Mattias Marder[1], Abhishek Kumar[1], Rogerio Feris[1], Raja Giryes[2], Alex M. Bronstein[3]

1 -IBM Research AI; 2 - School of Electrical Engineering, Tel-Aviv University, Tel-Aviv, Israel; 3 - Department of Computer Science, Technion, Haifa, Israel

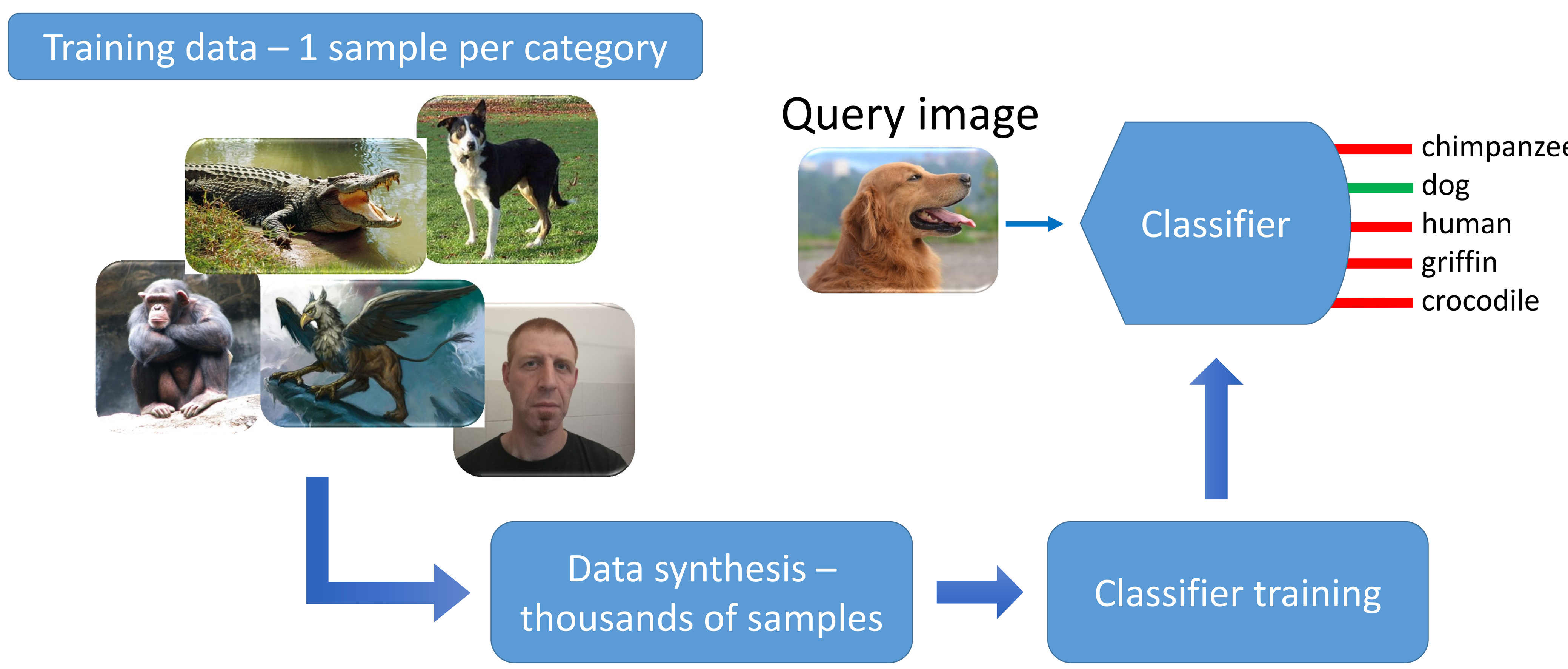
see our 3-minutes pitch

NIPS

Neural Information Processing Systems

Few-shot classification problem: presented with a few samples of K novel categories, produce a classifier for the K categories

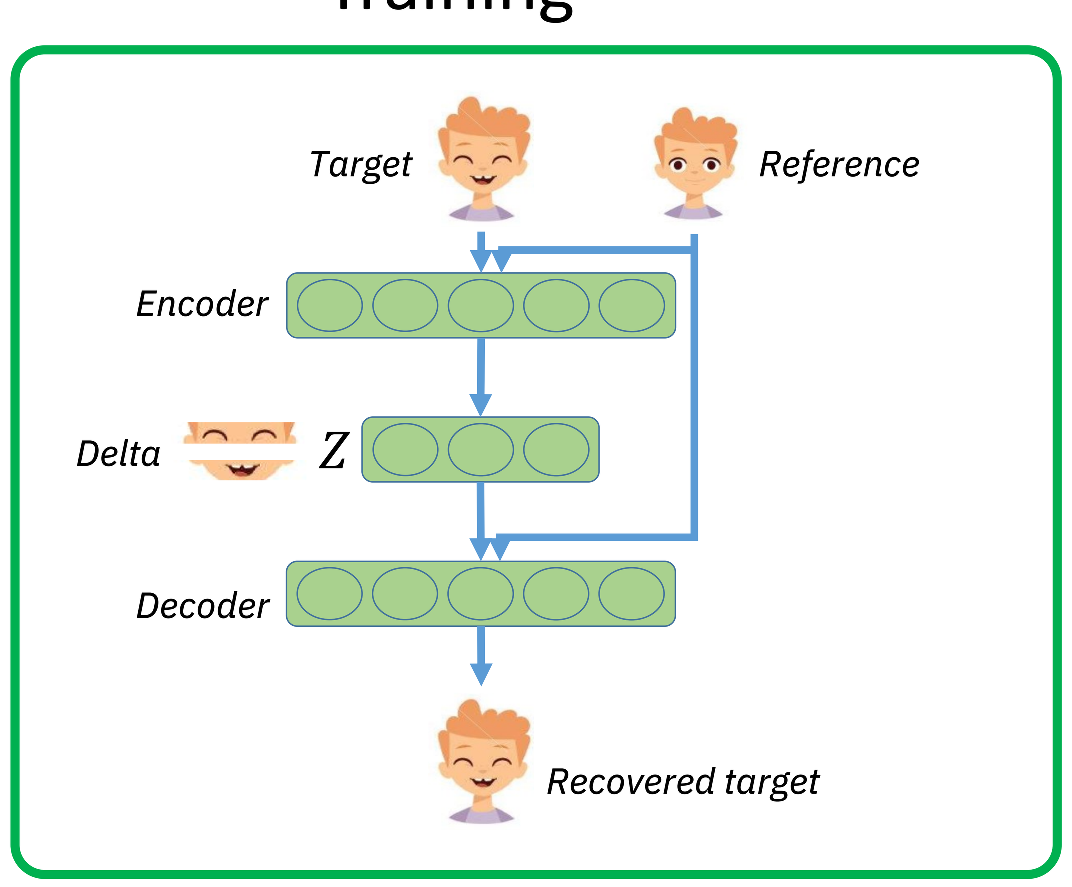
Generative approach: synthesize new samples from the few available examples to improve the classifier training



Our main idea

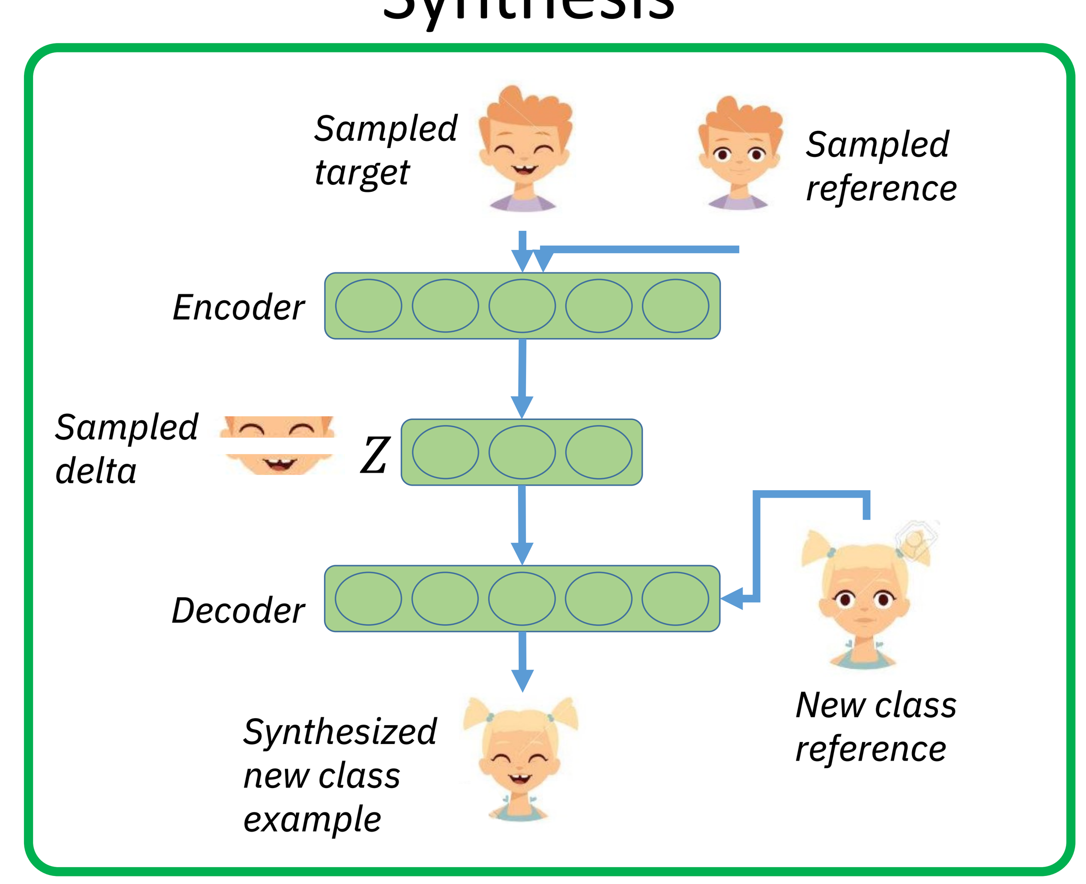
1. Learn to recover the relative transformation (the delta) between pairs of samples from the same category
2. Sample from a distribution of such transformations within the set of known categories
3. Learn to re-apply the transformations to examples of novel categories to produce additional samples thereof

Training



- A variant of an auto-encoder operating in feature space
- Network learns to encode the Δ between reference and target images
- This Δ is used to recover the target image as a (non-linear) combination of the reference and Δ

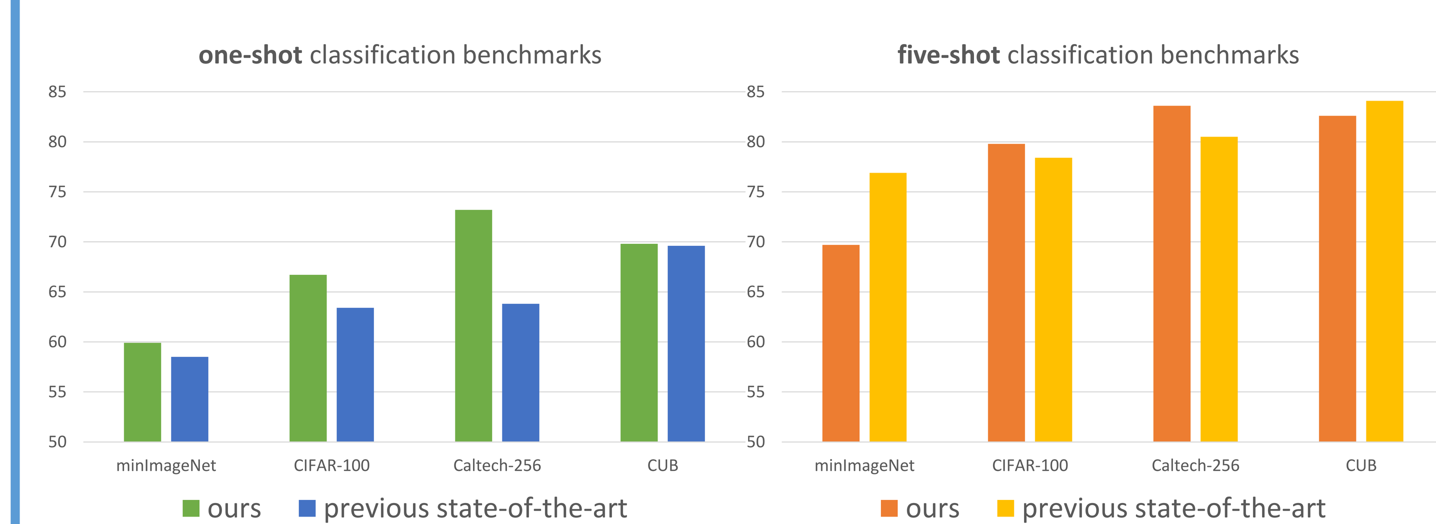
Synthesis



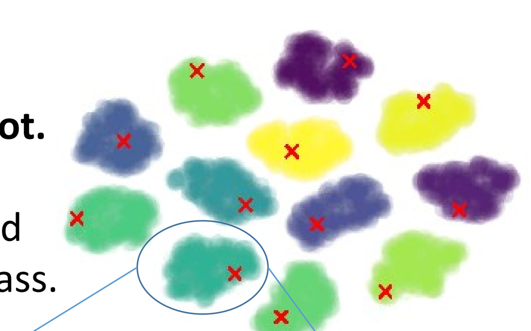
- At test time encoded Δs are sampled from random training image pairs belonging to the same category
- Sampled Δs are used to create samples for the new categories by combining them with the new category reference examples

Visualization of two-way one-shot classification trained on synthesized examples. *Correctly* classified images are framed in magenta (Golden retriever) and yellow (African wild dog). The only two images seen at training time and used for sample synthesis are framed in blue. Note the non-trivial relative arrangement of examples belonging to different classes handled successfully by our approach. Plotted using t-SNE applied to image features.

Dataset	Previous SOA 1-shot/5-shot	Ours 1-shot/5-shot
miniImageNet	58.5 / **76.9**	**59.9** / 69.7
CIFAR-100	63.4 / 78.4	**66.7 / 79.8**
Caltech-256	63.8 / 80.5	**73.2 / 83.6**
CUB	69.6 / **84.1**	**69.8** / 82.6



Generated samples for 12-way one-shot. The red crosses mark the original 12 examples (one per class). The generated points are colored according to their class.



Synthesized samples visualization. The single example image is framed in blue. All other images represent the synthesized samples visualized using their nearest "real image" neighbors in the feature space. The two-dimensional embedding was produced by t-SNE.

1-shot/5-shot 5-way accuracy using features externally pre-trained on a non-overlapping set of ImageNet categories

Method	AWA2	APY	SUN	CUB
Nearest neighbor (baseline)	65.9 / 84.2	57.9 / 76.4	72.7 / 86.7	58.7 / 80.2
Prototypical Networks	80.8 / 95.3	69.8 / 90.1	74.7 / **94.8**	71.9 / 92.4
Δ-encoder	**90.5 / 96.4**	**82.5 / 93.4**	**82.0** / 93.0	**82.2 / 92.6**