



Improving few shot object classification using contrastive learning

Abhishek Rajendra Prasad
Jishnu Jaykumar Padalunkal

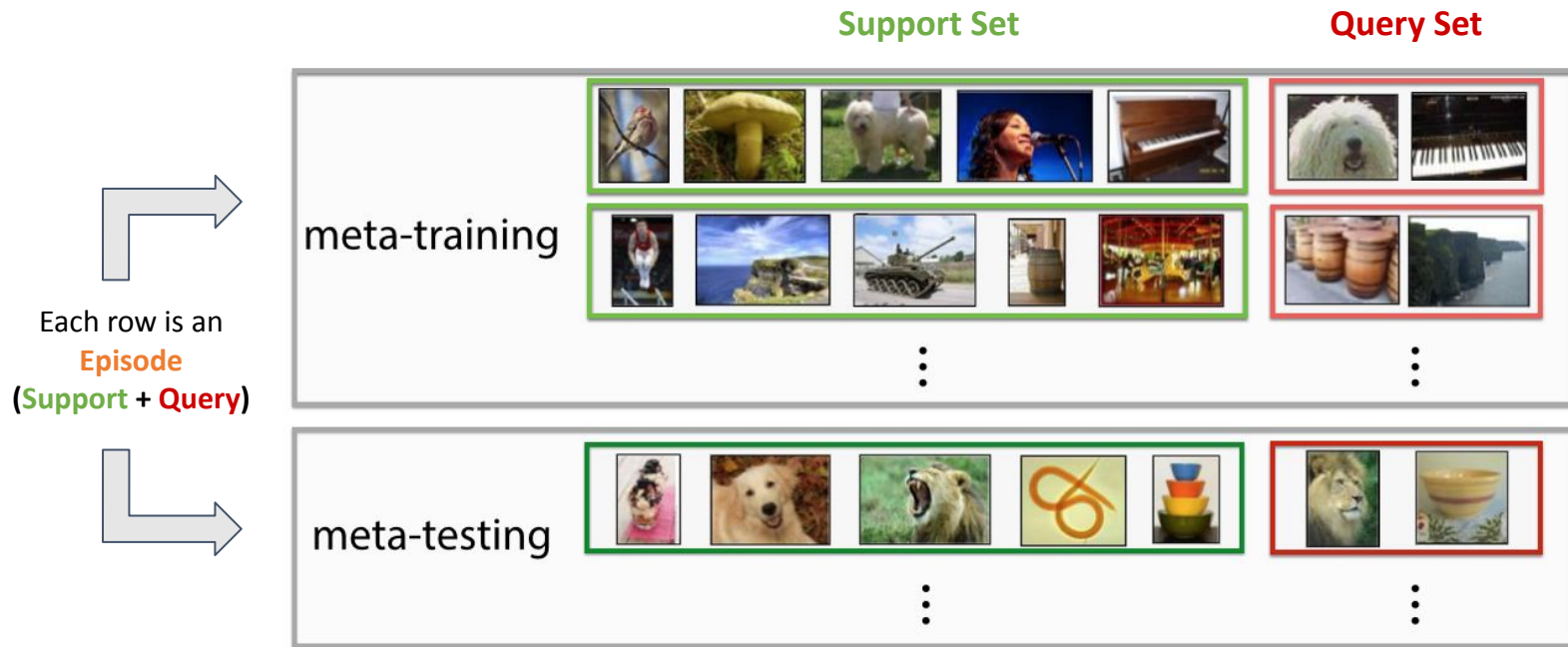
CS 6301.004 - Deep Learning For NLP

Group-11 | Spring 2023

Improving **few shot** object classification using contrastive learning

- Few-Shot Learning is a sub-area of machine learning. It's about classifying **new data** when you have only a **few training samples** with **supervised information** (neptune.ai).
- Formulated as an N-way-K-shot problem (**Episodes**)
 - N := number of classes
 - K := number of samples per class
 - In a fixed setup, this remains same for all classes
 - In a variable setup, this varies across classes

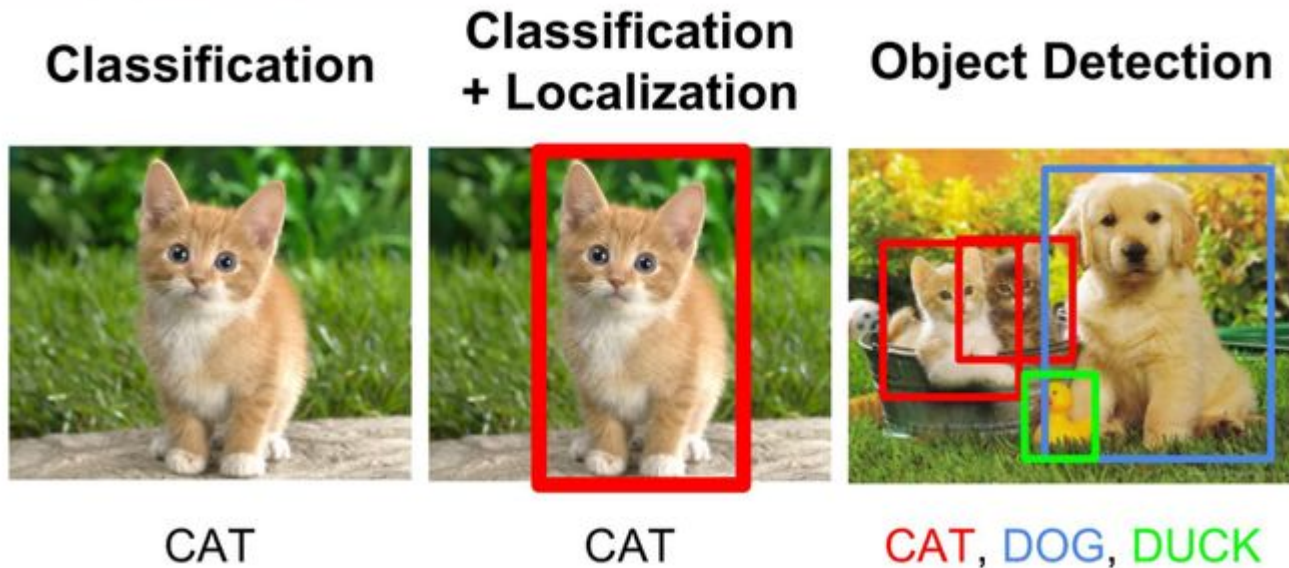
Improving **few shot** object classification using contrastive learning



Here, it's a **5-way-1-shot** setup (**fixed** episode variant)

Image: <https://bair.berkeley.edu/blog/2017/07/18/learning-to-learn>

Improving few shot **object classification** using contrastive learning

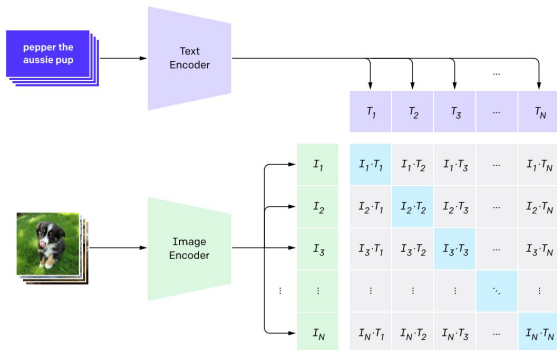


We will be dealing with **classification** only. i.e. Given an image containing a single object, classify it.

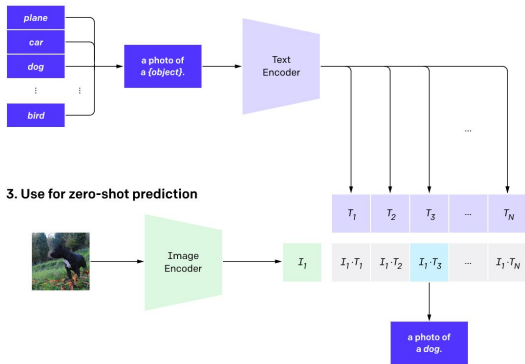
Image: <https://www.kaggle.com/getting-started/169984>

Related Works

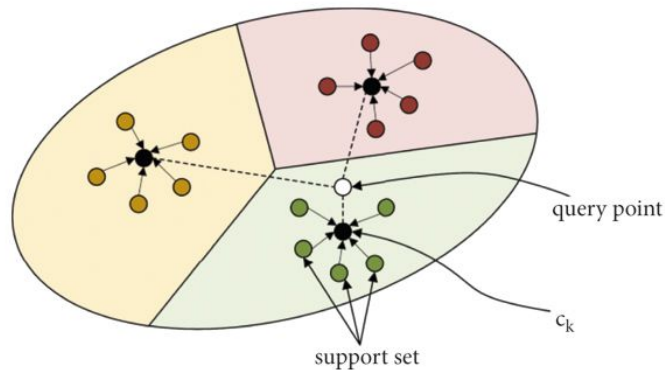
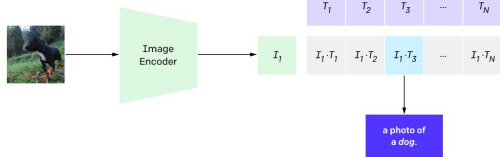
1. Contrastive pre-training



2. Create dataset classifier from label text



3. Use for zero-shot prediction



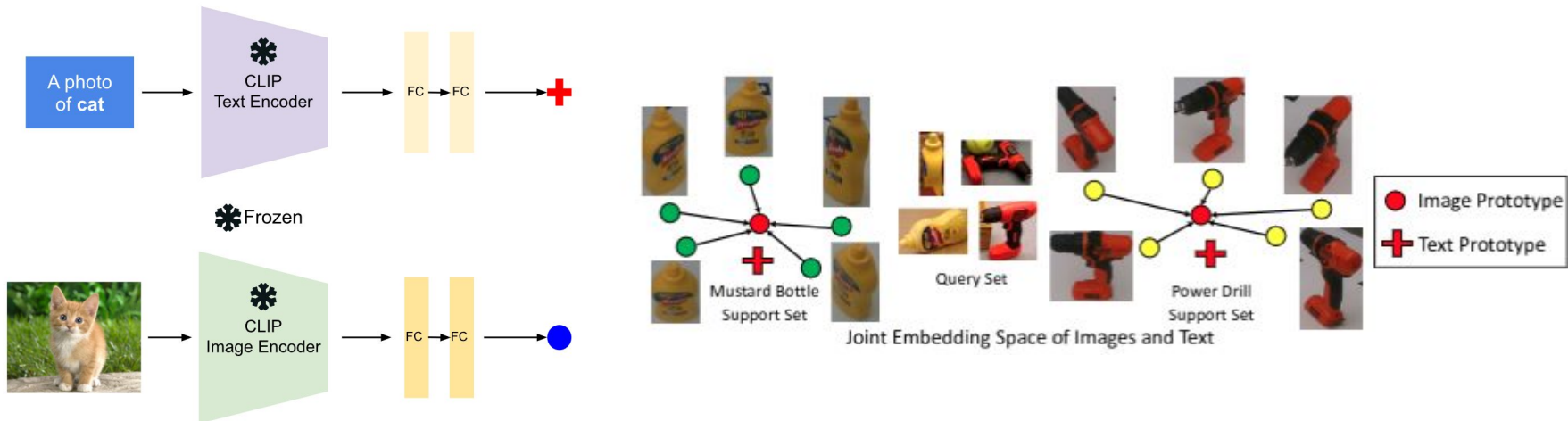
$$c_k = \frac{1}{|S_k|} \sum_{(x_i, y_i) \in S_k} f_\phi(x)$$

$$p_\phi(y = k | x) = \frac{\exp(-d(f_\phi(x), c_k))}{\sum_{k'} \exp(-d(f_\phi(x), c_{k'}))}$$

$$\min J(\phi) = -\log p_\phi(y = k | x)$$

Snell, Jake, Kevin Swersky, and Richard Zemel. "Prototypical networks for few-shot learning." NeurIPS 2017.

Proposed Method Proto-CLIP



$$P(y = k | \mathbf{x}^q, \mathcal{S}) = \alpha P(y = k | \mathbf{x}^q, \mathcal{S}_x) + (1 - \alpha) P(y = k | \mathbf{x}^q, \mathcal{S}_y)$$

Loss: Negative Log Likelihood

Our proposed **Proto-CLIP** model learns a *joint embedding space of images and text*, where *image prototypes* and *text prototypes* are learned using *support sets* for few-shot classification.

Dataset: ImageNet | **Metric:** Accuracy

Questions?