Contents

[Chapter 1 Contrastive Proposal Encoding ii](#_Toc161316488)

[1.1 Empirical Risk Minimization. ii](#_Toc161316489)

**C H A P T E R 1**

**Contrastive Proposal Encoding**

**CONTRASTIVE PROPOSAL ENCODING**

(CPE) Loss Inspired by supervised contrastive objectives in classification and identification, our CP E loss is defined as follows with considerations tailored for detection. Concretely, for a mini-batch of N RoI box features where is contrastive head encoded RoI feature for -th region proposal, denotes its Intersectionover-Union (IOU) score with matched ground truth bounding box, and denotes the label of the ground truth,

|  |  |  |  |
| --- | --- | --- | --- |
|  | | (1.1) | |
|  | | (1.2) |

is the number of proposals with the same label as , and is the hyper-parameter temperature as in In­foNCE.

We use unfrozen RPN and ROI with two modifications, (1.1) double the maximum number of proposals kept after NMS, this brings more foreground proposals for novel instances, and (1.2) halving the number of sam­pled proposals in RoI head used for loss computation

## 1.1 EMPIRICAL RISK MINIMIZATION

Given a hypothesis h, we want to minimize its expected risk R, which is the loss measured with respect to. Specifically,

|  |  |
| --- | --- |
|  | (1.3) |

As is unknown, the empirical risk (which is the average of sample losses over the training set of samples)

|  |  |
| --- | --- |
|  | (1.4) |

is usually used as a proxy for , leading to empirical risk minimization (with possibly some regularizers). For illustration, let:

* = be the function that minimizes the expected risk;
* = be the function in H that minimizes the expected risk;
* = be the function in H that minimizes the empirical risk.

As is unknown, one has to approximate it by some . is the best approximation for in , while is the best hypothesis in obtained by empirical risk minimization. For simplicity, we assume that, and are unique. The total error can be decomposed as:

|  |  |
| --- | --- |
|  | (1.5) |

where the expectation is with respect to the random choice of . The approximation error measures how close the functions in can approximate the optimal hypothesis, and the estimation error measures the effect of minimizing the empirical risk instead of the expected risk within .

Recently, more complicated task-invariant embedding models are learned via meta-learning 2 methods:

1. Matching Nets and its variants;
2. Prototypical Networks (ProtoNet) and its variants;
3. Other methods.

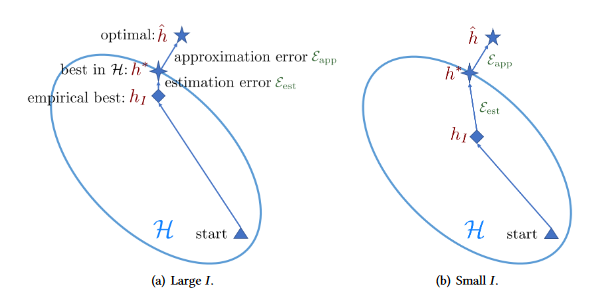


FIGURE 1.1 Comparison of learning with sufficient and few training samples.

However, in FSL, the number of available examples is small. The empirical risk may then be far from being a good approximation of the expected risk , and the resultant empirical risk minimizer over­fits. Indeed, this is the core issue of FSL supervised learning, i.e., the empirical risk minimizer is no longer reliable. Therefore, FSL is much harder. A comparison of learning with sufficient and few training sam­ples is shown in Figure 1.1.

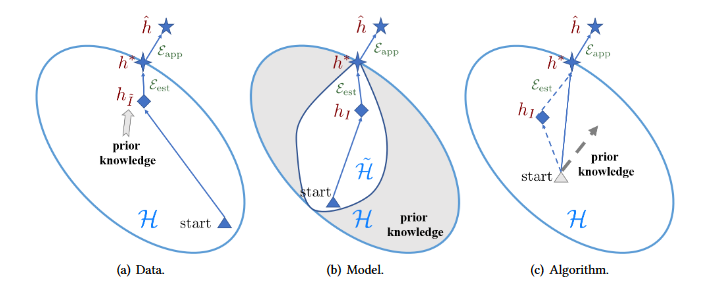


FIGURE 1.2 Different perspectives on how FSL methods solve the few-shot problem.

To alleviate the problem of having an unreliable empirical risk minimizer in FSL supervised learning, prior knowledge must be used. Based on which aspect is enhanced using prior knowledge, existing FSL works can be categorized into the following perspectives (Figure 1.2).

Depending on what samples are transformed and added to , we categorize these methods as shown in Table 1.1.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| TABLE 1.1 | Characteristics for FSL methods focusing on the data perspective. The transformer takes input and returns synthesized sample to augment the few-shot . | | | |
| category | | input | transformer | output |
| transforming samples from | | original | learned transformation function on |  |
| transforming samples from a weakly labeled or unla­beled data set | | weakly labeled or unla­beled | a predictor trained from |  |
| transforming samples from similar data sets | | samples from similar data sets | an aggregator to combine |  |

In terms of what prior knowledge is used, methods belonging to this category can be further classified into four types (Table 1.2).

|  |  |  |
| --- | --- | --- |
| TABLE 1.2 | Characteristics for FSL methods focusing on the model perspective. | |
| strategy | prior knowledge | how to constrain |
| multitask learning | other with their data sets | share/tie parameter |
| embedding learning | embedding learned from/together with other | project samples to a smaller embed­ding space in which similar and dissimilar samples can be easily discriminated |
| learning with external memory | embedding learned from other to interact with memory | refine samples using key-value pairs stored in memory |
| generative modeling | prior model learned from other | restrict the form of distribution |