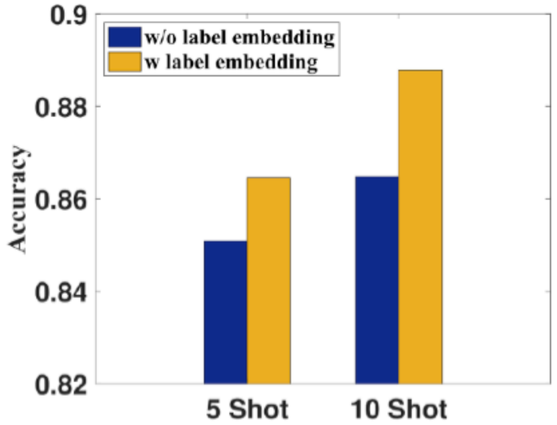
Experiments

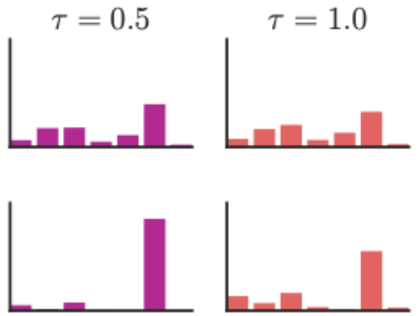
Ablation Study: w/o v.s. w/ Label Embedding



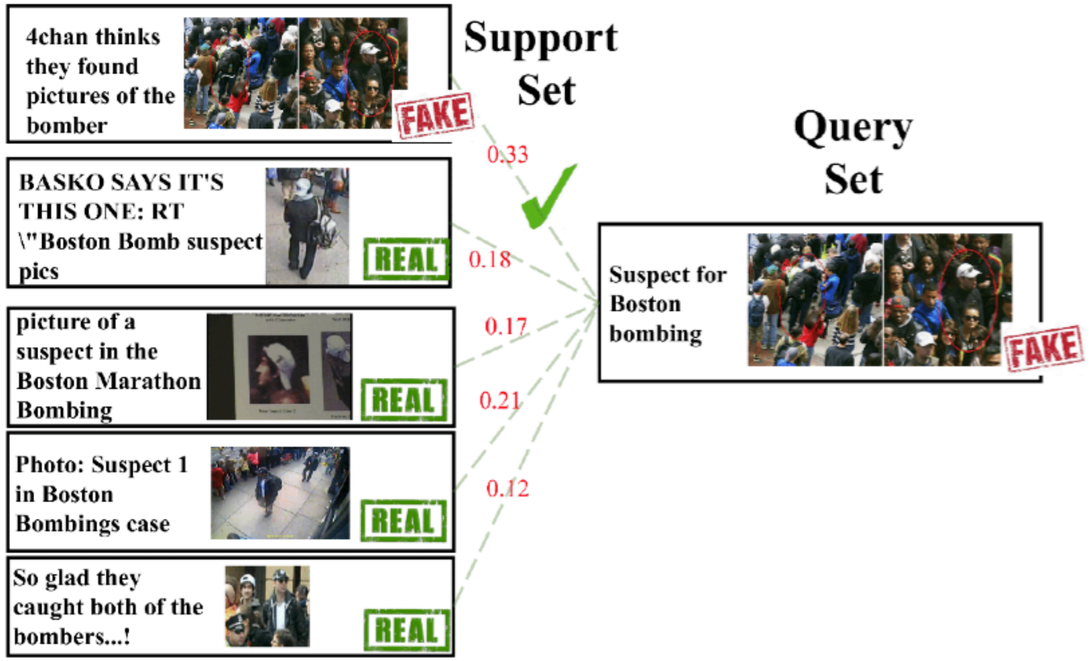
- Reduced model by replacing label embedding with label value 0 or 1.
- Change the multiplication between output with label embedding to a binary-class fully connected layer to directly output the probabilities.
- Observe that the accuracy score of w/ label embedding is greater than w/o label embedding in two shot settings.
  - Demonstrating the effectiveness of label embedding.

Case Study

Effective of Hard-Attention



https://arxiv.org/abs/1611.01144



- Although the first example with largest attention score value is most similar to news example in the query set, the majority of context information is from the other four examples due to imbalanced class distribution.
- Due to imbalanced class condition in the support set, it's difficult for Soft-Attention to provide correct prediction for news of interest in the query set.
- Hard-Attention can achieve correct result by focusing on the most similar sample in the support set.