Semi-Supervised Learning....

Ankan Bansal Joaquin Zepeda

1 Problem

Getting annotations is difficult. We want to leverage the large amount of unlabeled data in addition to the existing labeled data for training training deep networks.

Closest work?

2 Idea

$\mathbf{q_t}$ is the output probability distribution over the classes. T is the batch size. $\mathcal{L}_{cls}$ is the cross-entropy loss.

2.1 Mean Entropy Penalty (MEL)

$$\mathcal{L}_{MEL} = \frac{1}{T}\sum_{t=1}^{T} H(\mathbf{q_t}) \tag{1}$$

where $H(.)$ represents the entropy.

2.2 Negative Batch Entropy Penalty (NBEL)

$$\mathcal{L}_{NBEL} = -H(\frac{1}{T}\sum_{t=1}^{T} \mathbf{q_t}) \tag{2}$$

2.3 Locality Penalty

Class Activation Maps, C. Feature maps, F. Weights of the final layer, W. $C = A \times W$. The class with maximum probability is i. Calculate locality loss on C_i.

$$\mathcal{L}_{Loc} = \sum_{j=1}^{4} l_j(C_i) \tag{3}$$

where the sum is over the four groups (left-right, right-left, top-bottom, and bottom-top).

Consider one of these cases. Let, $C_i \in \mathbb{R}^{N\times N}$. We define the left-right group norms as $G_{lr} = [\left\|C_i^k\right\|_2 \mid C_i^k = C_i(1:k, 1:N), k = 1 \dots N]$. Then, $l_1(C_i) = \|G_{lr}\|_1$. Other group losses can be defined similarly.

Then the total loss is given as:

$$\mathcal{L} = \mathcal{L}_{cls} + \alpha\mathcal{L}_{MEL} + \beta\mathcal{L}_{NBEL} + \gamma\mathcal{L}_{Loc} \tag{4}$$

3 Evaluation

We use ImageNet as the training and testing dataset. We fix a set of about 64,000 images for which we have labels available (supervised set). We use an additional 200,000 images as unsupervised images, i.e. images for which labels are not available. We use the ImageNet validation set as the test set. We use the state-of-the-art DenseNet model.

3.1 Baseline

We train a network with only the supervised images and only supervised images.

3.2 Cross-entropy + MEL

We noticed that the performance of the model was close to the baseline but not quite there. When we had tried this case (cross-entropy + MEL) with only supervised data, we had achieved small improvements over baseline. This led us to give a higher weight to supervised data while sampling a batch.

We varied the ratio of supervised images in a batch from the default (0.25) to 1.0. The performance keeps improving till 0.8. But from 0.9, over-fitting becomes a problem. A concern could be that the gain in performance might be because of passing the supervised data so many times. However, we eliminated this.

3.3 Cross-entropy + MEL + NBEL

Now we add the negative batch-entropy loss to cross-entropy loss and MEL. This loss is based on the idea that uniform sampling from the dataset should lead to uniform sampling over the output classes. We varied β while keeping α and r fixed from the previous case.

3.4 Cross-entropy + Loc

We vary γ.

We are also currently trying to weigh the class activation maps by the corresponding probabilities instead of using only the CAM with the maximum probability.

3.5 Cross-entropy + MEL + BEL + Loc

TODO

References

1. Author, F.: Article title. Journal **2**(5), 99–110 (2016)