Semi-supervised learning using convolutional neural networks

Bhargavi Neti
Arizona State University
bneti@asu.edu

Kasturi Adep
Arizona State University
kadep@asu.edu

Abstract

It is an established fact that convolutional neural networks have revolutionised image classification. But most of these algorithms require large amounts of labelled data for learning. When we look at it, most data in the real world is not labelled. Also, training of models using labelled data alone creates situations of overfitting. These challenges call for advanced methods that help the model prepare better for unseen data. Lot of work is being done around semi-supervised learning methods, a technique that brings together the best of both supervised and unsupervised learning, in addition to regularization techniques to address the problems of overfitting and limited labelled data availability.

With this motivation, we experiment with semi-supervised learning using Entropy Minimization and data augmentation techniques such as CutMix and MixUp using the CIFAR-10 dataset. We conclude our study with a thorough analysis of each of these techniques and present a comparison of their performances.

1. Introduction

The scientific community has proposed several ideas on how best to make use of labelled and unlabelled data in training a model. Popularly listed among these are EM with generative mixture models, transductive SVM, co-training, self-training, and graph-based methods[4]. The method one chooses to employ on their dataset largely depends on the assumptions that fit their problem structure.

Generative models have major drawbacks as they use joint probability distribution of patterns and classes in the dataset. Computation of P(X, Y) is expensive and makes it more complicated than simply estimating the conditionals. As parameters need to be estimated this causes further uncertainty in the estimation process. To address these shortcomings of existing techniques, semi-supervised algorithms were enhanced.

Particularly for image classification, use of deep convolutional neural networks(CNN) with semi-supervised learning techniques is being widely explored. By training the model to focus on less discriminative parts of objects, regional dropout strategies enhanced performance of CNN classifiers, allowing the network to generalize better. One major drawback of these strategies is the removal of pixels containing valuable information resulting in inefficiency. CutMix and MixUp are techniques that try addressing this issue while producing models robust to corrupted input data.

1.1. Related Work

There have been many proposals to address the problems faced in processing unlabeled data in a supervised classification setting. Entropy minimization was one of the earliest methods used to model semi-supervised learning in a probabilistic framework[1]. This series of experiments suggest that with increase in class overlap, the amount of information content in unlabelled data decreases.

More recent techniques have chosen to work with neural networks in order to extract more relevant data from existing records and help the model generalize better. Any effective regularization technique aims to reduce the variance without trigering increase in bias. One among this is CutMix. As opposed to randomly dropping parts of image data, CutMix replaces removed portions with a patch from another image. This results in more efficient training of the neural network as there is no uninformative pixel left. Its performance has been evaluated on different set of data, CNN models and tasks[3].

MixUp is yet another data augmentation technique similar to CutMix that interpolates the image and labels. A series of experiments show the effects of this simple principle in alleviating undesirable behaviours of neural networks such as memorization and sensitivity to adversarial examples[2].

2. Method

In this set of experiments we consider the CIFAR-10 dataset. It is a labelled set of images containing 60000 images belonging to 10 different classes. A labelled dataset allows us to compare different techniques and evaluate their performance to the closest true value.

We start by implementing a deep convolutional neural network using only the labeled training data. Then we evaluate its performance on the test data. Considering this as the **Baseline** model we compare its performance with other models.

Following the baseline model, we build three additional models that use semi-supervised learning techniques.

- Entropy Minimization - Conditional entropy $H(Y|X)$, on the assumption that classes are well separated, measures the class overlap[1]. It helps identify the usefulness of data given the uncertainty associated with the label and trains the model to be confident regarding its predictions. The entropy loss calculations for unlabelled data are implemented using the following equation.

$$H(Y|X,Z) = -\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{K}P(\omega_k|x_i,z_i)logP(\omega_k|x_i,z_i) \quad (1)$$

- MixUp - In this method, two randomly selected samples (x_i, y_i) and (x_j, y_j) are combined to form a resultant virtual training examples such that[2]

$$\bar{x} = \lambda x_i + (1-\lambda)x_j \quad (2)$$
$$\bar{y} = \lambda y_i + (1-\lambda)y_j \quad (3)$$

where $\lambda \in (0, 1)$.

- CutMix - In this method we combine two samples (x_i, y_i) and (x_j, y_j) to generate a new image $(\bar{x}, \bar{y})$. The resultant image is used for training the model for the given loss function.

  We can define the combining operation as[3],

$$\bar{x} = M \cdot x_i + (1-M)\cdot x_j \quad (4)$$
$$\bar{y} = \lambda y_i + (1-\lambda)y_j \quad (5)$$

where $M \in (0,1)^{WxH}$ is a binary mask operation, $\lambda \in (0, 1)$.

3. Experiments

Though current research shows that Entropy Minimization, CutMix and MixUp algorithms improve image classification using semi-supervised techniques, we conducted a set of experiments that create models using these three techniques.

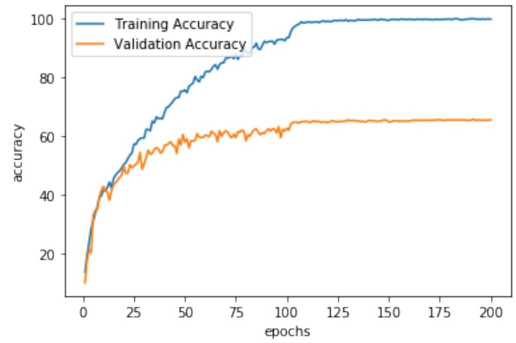

Figure 1. Training and validation accuracy trends for Baseline model

The CIFAR10 dataset was divided into four parts, labelled, unlabelled, validation and test, while ensuring equal representation for all classes.

We have used a ResNet18 Model architecture for 200 epochs with learning rate 0.1, decay rate 0.0001 and batch size as 256. The model parameters were modulated in order to produce maximum accuracy. We conducted separate observations and conclusions for Entropy Minimization and the data augmentation techniques.

1. Baseline:
   This model does not employ any special techniques while training. It takes an image as input and provides an output. This output is then compared with the target value. Finally the backward pass is initiated using cross entropy loss as a function for calculating net loss.

2. Entropy minimization :
   We varied the percentage of labelled and unlabelled samples in the training data. The accuracy stays constant at around 115 epochs. It was observed that, the increase of even a few unlabelled samples impacts the accuracy of the model a lot. The testing and training results have been summarized in table 1, figures 2 and 3

% of labelled data	Training accuracy	Testing accuracy
100%	99.5 %	84.6%
99.9%	98.5 %	85.4%
99.5%	29.51 %	28.5%

Table 1. Accuracy trends for Entropy minimization on CIFAR-10

3. MixUp:
   For the same set of data used in experiment 2, train a different model using MixUp strategy. The idea is to overlap two images with different opacity. For that

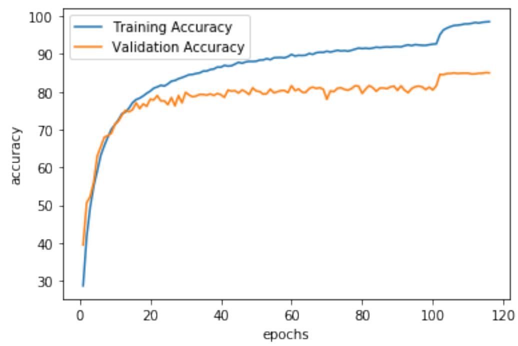

Figure 2. Variation of training vs validation accuracy for 99.9 percent labelled data

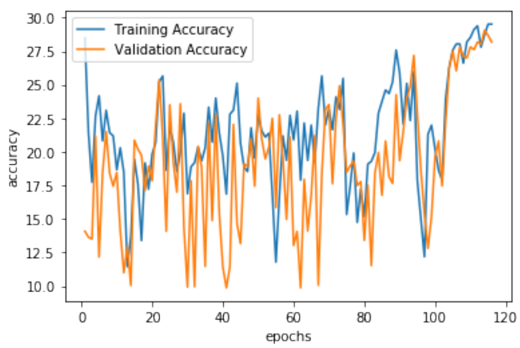

Figure 3. Variation of training vs validation accuracy for 99.5 percent labelled data

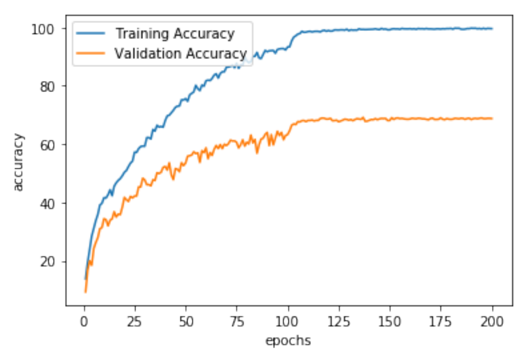

Figure 4. Training and validation accuracy trends for MixUp

we choose a factor λ and create a new image using the equations described in Section 2. On providing this image to the model we calculate the output to determine the loss. Estimated loss value for each label is weighted with respect to the contribution of each constituting image.

Fig 4 showcases the increase in accuracy of the model after every epoch.

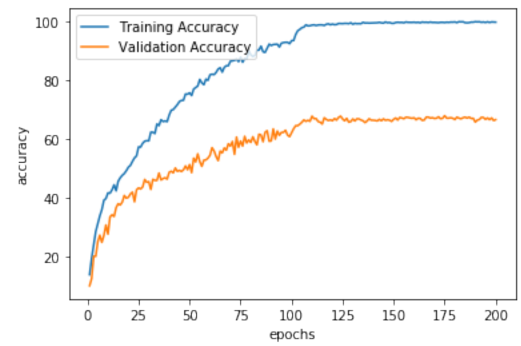

Figure 5. Training and validation accuracy trends for CutMix

4. CutMix:
   CutMix aims to maximally used deleted regions of an image. A particular section of an image is replaced with a part from another image. The ground truth labels are mixed in proportion of number of pixels of the combined two images[3]. This way the technique ensures that there is no uninformative pixel in the resulting training data. The network parameters are updated using cross entropy loss function. Observe fig 5 for trends in accuracy.

ResNet18	Training accuracy	Test accuracy
Baseline	99.58%	65.48%
+ Mixup	99.58%	67.06%
+ CutMix	99.56%	69.76%

Table 2. Accuracy trends on CIFAR-10[3]

4. Conclusions

As can be inferred from fig 1, 100% training accuracy does not guarantee an equivalent performance on unseen data. This validates the need for regularization techniques such as CutMix and MixUp.

Improvement in test accuracy has been the general observation throughout the course of these experiments. The training accuracy does not show any significant change but the increase in test accuracy is definitely noteworthy. This confirms that CutMix and MixUp work towards reducing the variance without impacting the bias. Another factor that largely affected the accuracy was the percentage of labelled data.

- It has been observed that Entropy Minimization brings about a 1% improvement in the accuracy when the percentage of unlabelled data is minimal. With an increase in the amount of unlabelled data, there is a drop in accuracy.

- MixUp samples can be perceived as unnatural and ambiguous as they are an overlap of two images. This is believed to confuse the model. But the technique shows 3-5% improvement on performance nonetheless.

- Shortcomings of MixUp lead researches to propose CutMix, where an image is divided into two sections. Each individual section still provides complete information as opposed to MixUp where the internal representations are entirely changed. CutMix works on a data level and shows better performance than MixUp. A 5-7% increase in test accuracy over the baseline model was observed for this method.

5. Division of Work

1. Kasturi Adep:
   Implementation and integration of MixUp and CutMix techniques.

2. Bhargavi Neti:
   Creation of baseline model and implementation of Entropy Minimization.

6. Self-Peer Evaluation

This section tabulates the self-evaluation score for self and peers based on their contributions towards this project.

Myself - 20	Bhargavi Neti - 20

Table 3. Self-Peer Evaluation Table

References

[1] Y. Grandvalet and Y. Bengio. Semi-supervised learning by entropy minimization. *Advances in neural information processing systems*, page 529–536, 2005.

[2] Y. N. Dauphin H. Zhang, M. Cisse and D. Lopez-Paz. mixup: Beyond empirical risk minimization. *arXivpreprint arXiv:1710.09412*, 2017.

[3] S. J. Oh S. Chun J. Choe S. Yun, D. Han and Y. Yoo. Cutmix: Regularization strategy to train strong classifiers with localizable features. *arXiv preprint arXiv:1905.04899*, 2019.

[4] Xiaojin Zhu and Andrew B. Goldberg. Introduction to semi-supervised learning. 2009.