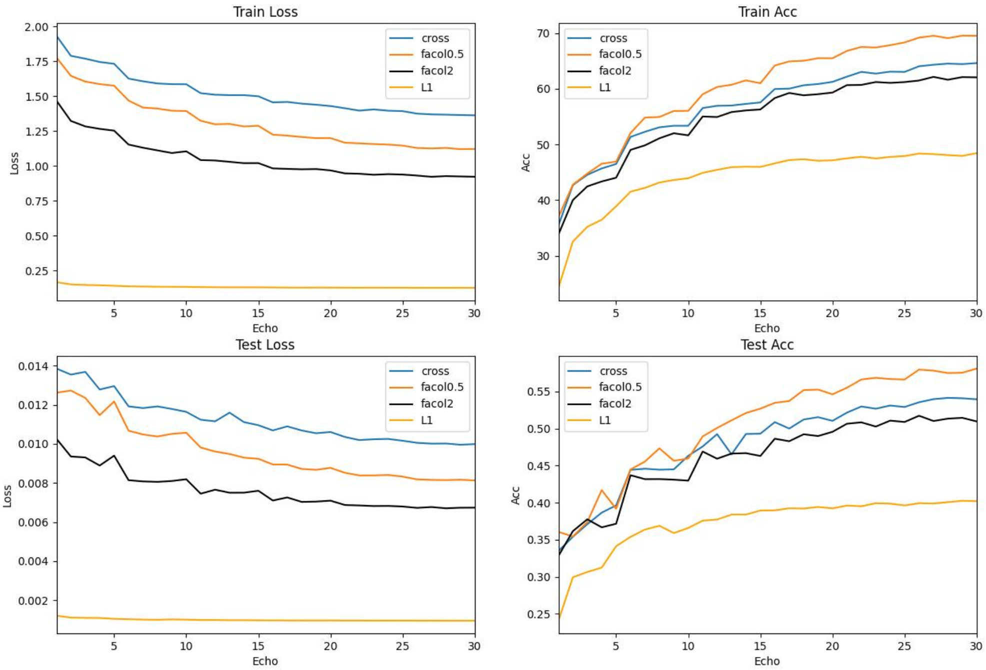
For four loss functions, MAE(L1) loss, CE loss, Focal loss (gamma=0.5) and Focal loss (gamma=2), I conducted training and testing respectively, and saved the loss and accuracy (Acc) of the training set and test set.



1. First of all, we conducted an overall image analysis and found that Focal(gamma=0.5) achieved the best effect. From the perspective of training set and test set, its final accuracy was higher than that of the other three models, and its convergence speed was almost the same as CE and Focal(gamma=2). They all converge at about 25 epochs.

2. L1 Loss:

Advantages: L1loss is more suitable for multi-variable linear regression problem, because it can eliminate some variables with little influence as much as possible.

Disadvantages: L1loss is not suitable for classification problems. In this multi-classification problem, it is found that the loss of L1loss is almost stable and does not change, because the setting of L1loss is used for linear regression problems, similar to lasso. Since we set our target as a unique thermal vector, it will make itself output as many zeros as possible, which leads to the loss calculation every time. The output has only two or three non-zero values, so its loss is relatively stable. This loss function thus dampens the network's enthusiasm for scoring different categories, which can improve its accuracy. The resulting accuracy, as shown in the figure above, does not even exceed 0.5, maintaining around 0.4.

3. CE Loss

CE Loss can also be viewed as Focal Loss(gamma=0), and the magnitude of the retrogression gradient for each class of CE Loss is proportional to the deviation from its predicted value. It encourages the model to score different categories and is more suitable for classification problems than L1 Loss.

Disadvantage:

• Sensitive to outliers: The CE loss is very sensitive to outliers, which means that when there is noise or labeling errors in the data, the CE loss may cause the performance of the model to degrade.

• Class imbalance problem: In the case of class imbalance, CE loss may cause the model to be biased in favor of predicting the majority class while ignoring the minority class. These problems are solved in Focal Loss.

• Difficulty dealing with missing data: CE losses usually require a valid category label for each sample. If there are missing labels or uncertainties in the data, CE loss may not be the best option.

4. Focal Loss

Advantages:

• It has the same advantage as CE loss, and the size of the back-propagation gradient of each type is proportional to the deviation of its predicted value. The model is also encouraged to give scores to different categories.

• Processing category imbalance, Focal Loss = has one more item than CE Loss =, which is also called modulation factor, which is used to reduce the loss contribution of easily divided samples. Whether it is foreground class or background class, the larger the item is, the easier the sample is to be distinguished, and the smaller the modulation factor is. The model can be concentrated to divide difficult samples.

Cons:

・ Parameter sensitivity: Focal Loss has hyperparameter beta that can be adjusted They need to be carefully adjusted to suit specific problems, the wrong parameter selection can cause performance degradation. You can refer to 2 and 0.5 in this experiment, and it is not that the larger the gamma, the better. In this experiment, 0.5 > 0 > 2, can be obtained from the accuracy curve of the test. My analysis reasons are as follows:

① Because the data set is more efficient for relatively simple tasks and the categories are more balanced, smaller values help the model converge more smoothly. High gamma results in too much focus on hard to classify samples, resulting in incorrect classification of simple samples.

② And because there is still some class imbalance, gamma=0.5 will make the classification to deal with such problems, so it is more reasonable and effective than CE Loss.