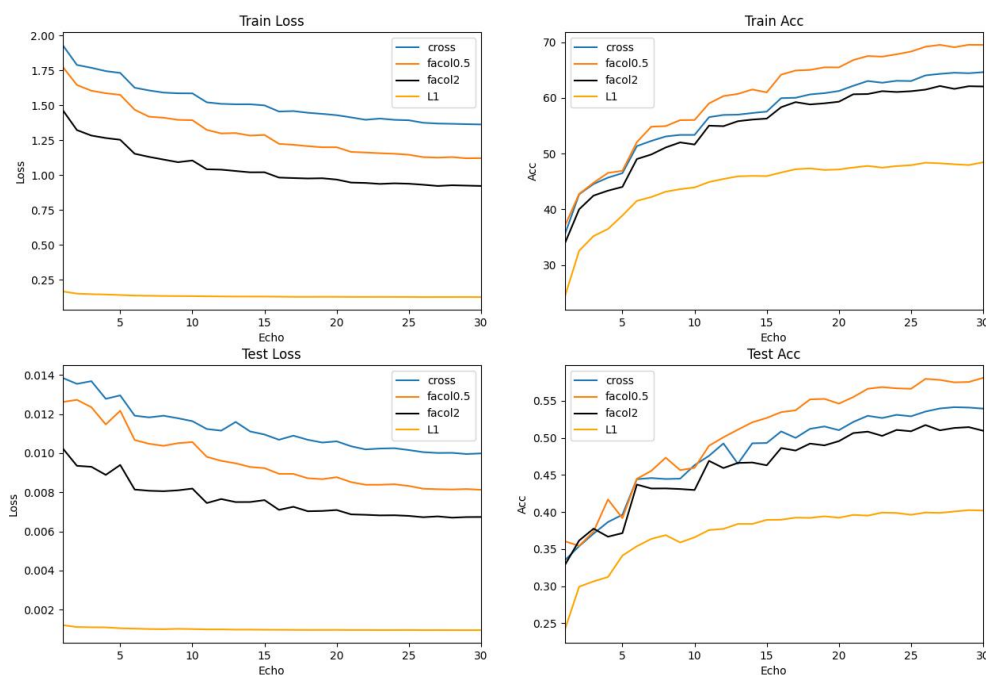


对于 MAE(L1) loss, CE loss, Focal loss (gamma=0.5), Focal loss (gamma=2)四种损失函数, 我分别进行训练和测试, 并且将训练集和测试集的损失(Loss)和准确率(Acc)保存下来。



1. 首先我们对于图像进行整体的分析, 发现 **Focal(gamma=0.5)**, 取得的效果是最好的, 从训练集和测试集来看, 它的最后准确率是高于其他三种模型的, 并且收敛速度于 CE 和 Focal(gamma=2)几乎一样, 都是在 25 个 epoch 左右几乎收敛。

## 2. L1 Loss:

**优点:** L1loss 比较适合进行多变量的线性回归问题, 因为可以尽可能的剔除掉一些影响不大的变量。

**缺点:** L1loss 并不适合分类问题。在此多分类问题当中, 发现 L1loss 的损失几乎稳定不发生变化, 是因为 L1loss 的设置是用于线性回归问题, 类似 lasso, 由于我们将我们的 target 设定成为独热向量, 所以它会使得自己尽可能输出多的 0, 这也就导致每次计算损失时, 输出只有二到三项非 0 值, 所以它的损失就会较为稳定。因此这个损失函数抑制了网络对于不同类别打分的热情, 而对不同类别打分可以提高自己的准确率。导致准确率如上图, 甚至没有超过 0.5, 维持在 0.4 左右。

## 3. CE Loss

**优点:** CE Loss 也可以看作为 Focal Loss(gamma=0), CE Loss 的每一类的回传梯度的大小都是与其预测值的偏差成正比。它鼓励模型对于不同类别进行打分, 相比于 L1 Loss 更加适合于分类问题。

**缺点:**

- 对异常值敏感: CE 损失对异常值 (outliers) 非常敏感, 这意味着当数据中存在噪声或标签错误时, CE 损失可能会导致模型的性能下降。

- 类不平衡问题：在类别不平衡的情况下，CE 损失可能会导致模型偏向于预测占多数的类别，而忽略了少数类别。此类问题在 Focal Loss 当中解决。

- 难以处理缺失数据：CE 损失通常要求每个样本都有一个有效的类别标签。如果数据中存在缺失标签或不确定性，CE 损失可能不是最佳选择。

#### 4. Focal Loss

##### 优点：

- 与 CE loss 拥有共同的优点，每一类的回传梯度的大小都是与其预测值的偏差成正比。并且鼓励模型给不同类别打分。

- 处理类别不平衡， $\text{Focal Loss} = -(1 - p_t)^\gamma \log(p_t)$  相比于  $\text{CE Loss} = -\log(p_t)$  多了一项  $(1 - p_t)^\gamma$ ，这一项我们又称为调制因子，它是用来减低易分样本的损失贡献，无论是前景类还是背景类， $p_t$  越大，就说明该样本越容易被区分，调制因子也就越小，可以让模型集中去分难分的样本。

##### 缺点：

参数敏感性：Focal Loss 有超参数  $\beta$  可以调节它们需要谨慎调整以适应特定问题，错误的参数选择可能会导致性能下降。可以参考本实验当中的 2 和 0.5，并不是 gamma 越大越好，此实验当中  $0.5 > 0 > 2$ ，由测试的准确率曲线可以得到。我分析理由有：

① 因为这个数据集是一些相对简单的任务以及类别是较为平衡的情况可能更有效，因此较小的值有助于模型更平稳地收敛。gamma 较大会导致过于关注难分类样本，而导致简单样本分类不正确。

② 再因为类别还是存在一些类别不平衡，所以 gamma=0.5, 会使得分类时处理此类问题，因此比 CE Loss 更为合理以及效果好。