

第三次训练结果

2022年8月15日 17:12

一、数据集

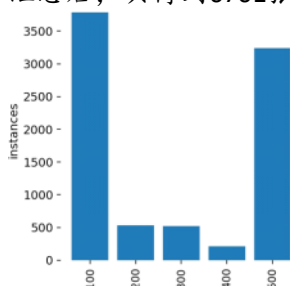
目前使用了四个数据集进行训练和测试：

1. 原始数据集（提供的已标注数据集）

数据集来源于'病害标注工作-公共数据集-日本部分'：

s (F:) > 路面检测 > 0730 > 病害标注工作-公共数据集-日本部分			
名称	修改日期	类型	大小
JOB-20220721-01	2022/07/29 17:33	文件夹	
JOB-20220721-02	2022/07/29 17:32	文件夹	
JOB-20220721-03	2022/07/21 11:44	文件夹	
JOB-20220721-04	2022/07/29 17:32	文件夹	
JOB-20220721-05	2022/07/21 14:14	文件夹	

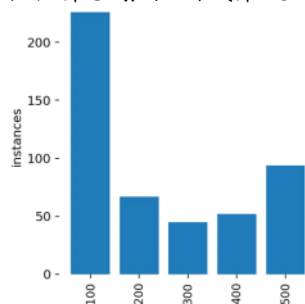
汇总后，共得到6701张已标注图片，训练集实例分布如下：



将数据集采用随机划分的方式，分成训练集5744张，验证集1149张，测试集957张。

2. 筛选的数据集1

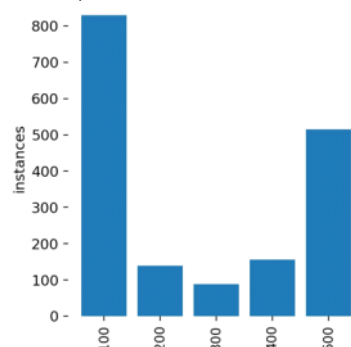
本数据集来源于已标注数据集中的512张图片，经过部分修改。划分为训练集共326张、验证集84张、测试集102张。训练集实例分布如下：



注：本数据集挑选了目标边界清晰、背景较为简单、图中实例个数较少的图片，图片质量较高，同时场景相对简单一些。

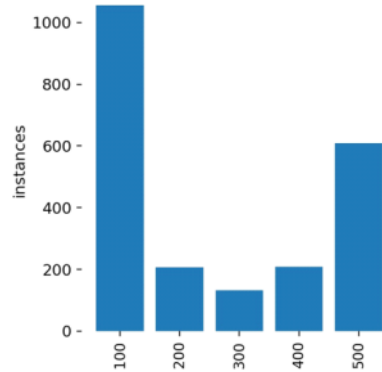
3. 筛选的数据集2

本数据集来源于已标注数据集中的1332张图片，存在修改。划分为训练集904张，验证集228张，测试集200张。实例分布如下：



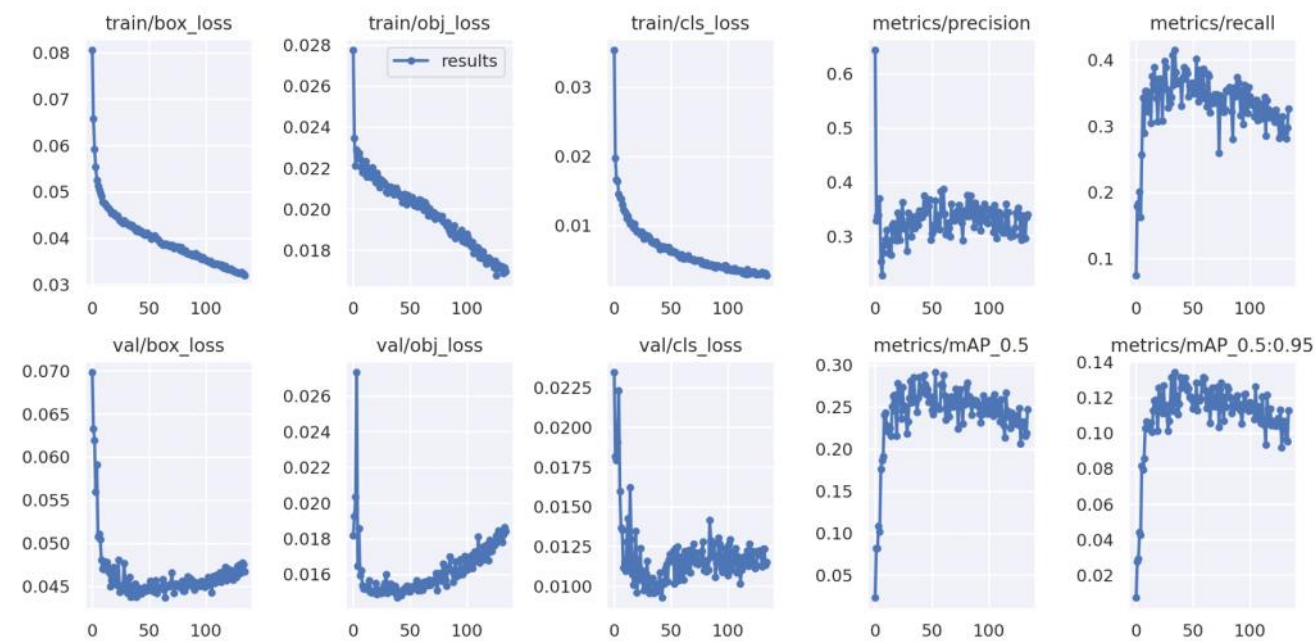
4. 筛选数据集1+筛选数据集2的合并数据集3

将筛选数据集1和筛选数据集2合并后得到数据集3，共1844张图片，训练集1230张，验证集312张，测试集302张。训练集实例分布如下：



二、训练结果

1、原始数据集（全部已标注数据集）训练结果

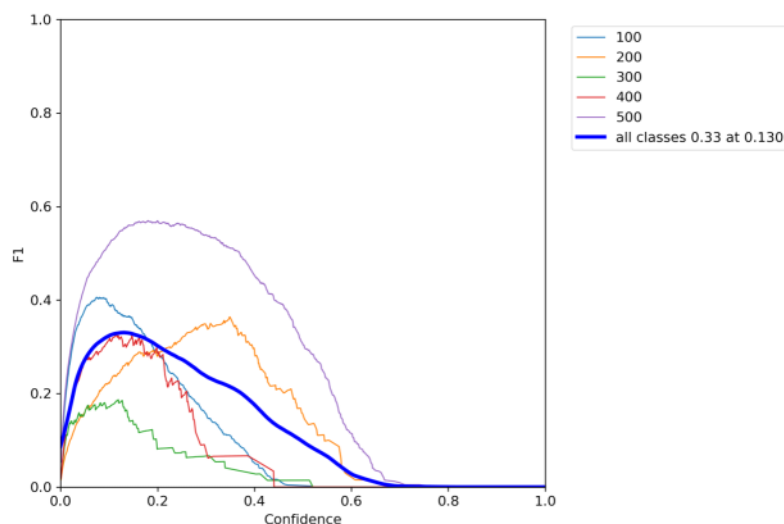


模型在训练第34轮获得最佳结果，各类别平均精度为0.32，平均召回率为0.414，map0.5为0.284。在不同类别中，类别‘100’的精度达到0.501，但召回率仅0.288。如下表所示：

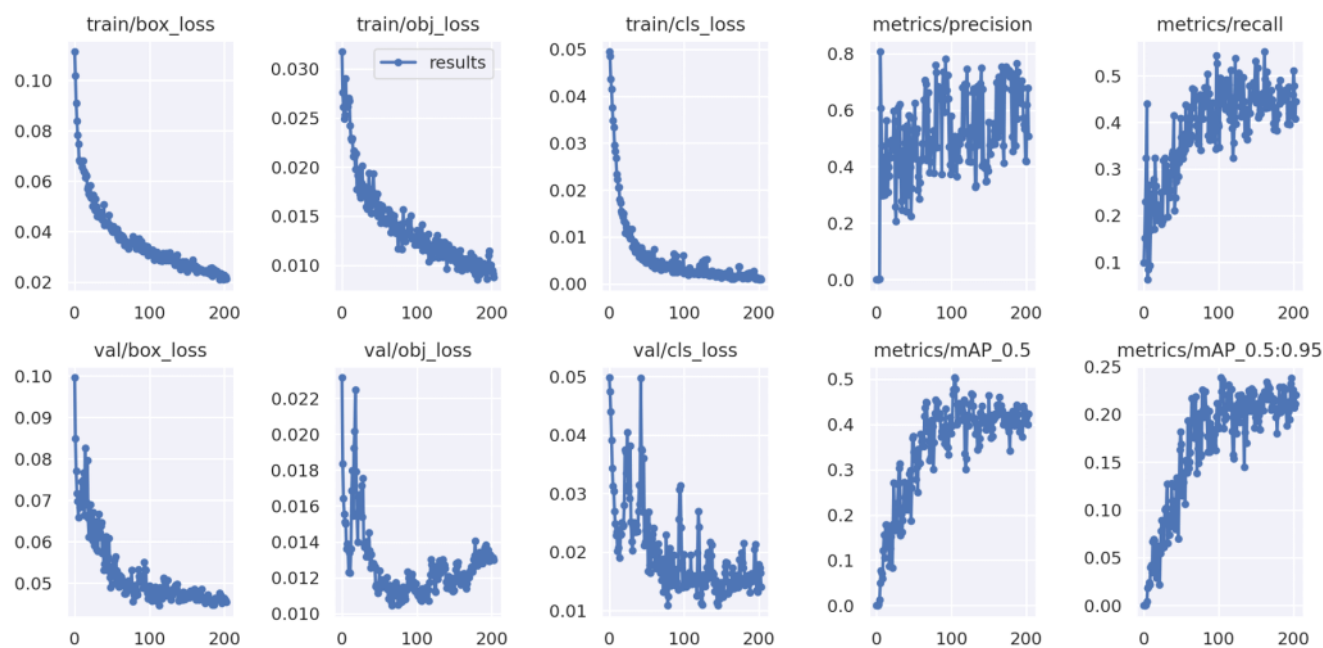
Class	Images	Labels	P	R	mAP@.5	mAP@.5:.95
all	1149	2110	0.32	0.414	0.284	0.134
100	1149	968	0.501	0.288	0.335	0.119
200	1149	129	0.181	0.465	0.222	0.0829
300	1149	137	0.233	0.124	0.0848	0.0368
400	1149	58	0.216	0.517	0.235	0.098
500	1149	818	0.469	0.675	0.54	0.333

模型存在难以拟合的现象，因为训练仅34轮就获得最佳结果，继续训练难以提高模型效果。

F1综合评价指标曲线如下：



2、筛选数据集1训练结果



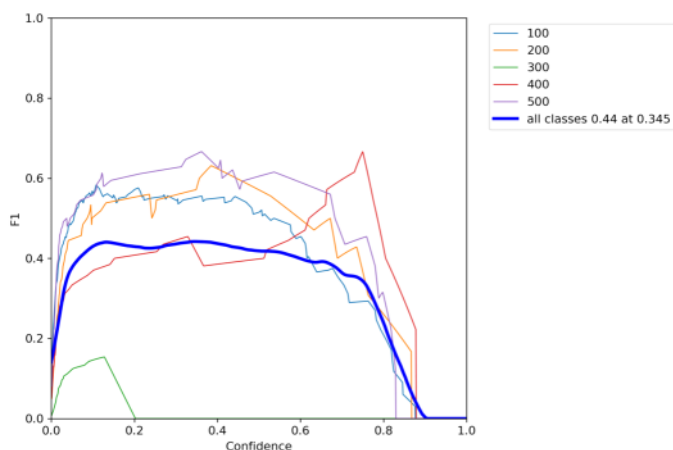
本数据集的在训练第103轮得到最佳结果。从上方的图中可以看到，准确率（precision）和召回率（recall）的结果不是很稳定，曲线颠簸较大，存在训练不稳定的问题，这是因为训练集和测试集的图片较少，增加数据集图片能够提高训练的稳定性。

各类别及平均表现如下表：

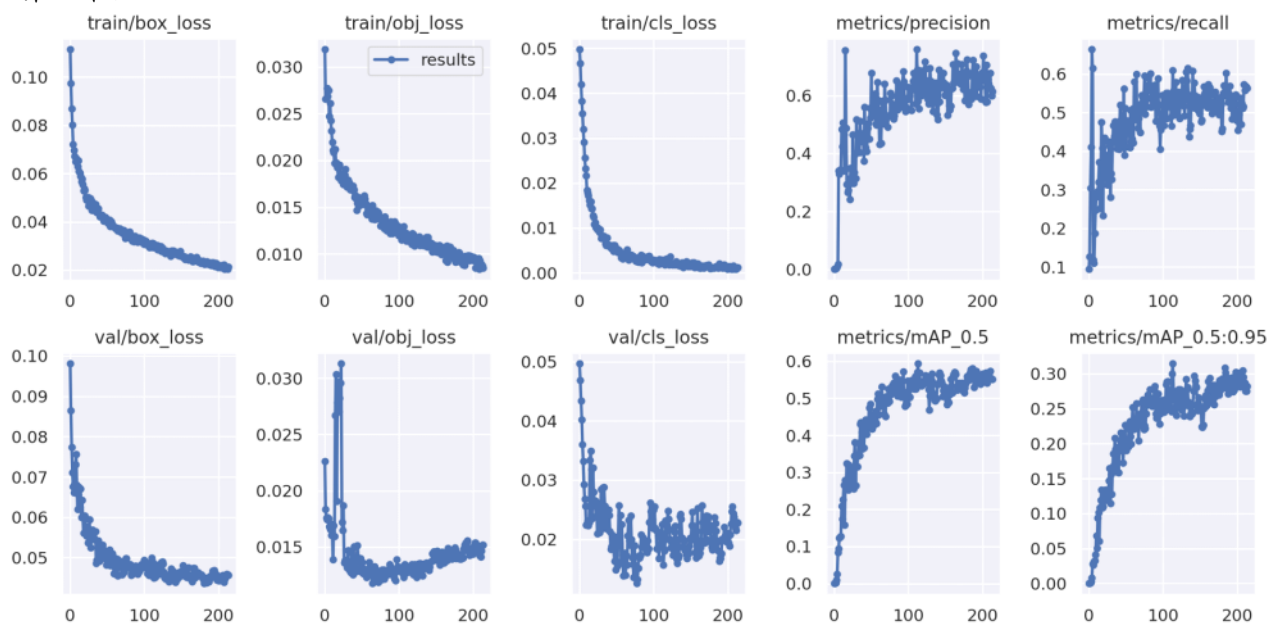
Class	Images	Labels	P	R	mAP@.5	mAP@.5:.95
all	84	109	0.463	0.445	0.477	0.238
100	84	64	0.68	0.469	0.58	0.248
200	84	11	0.675	0.545	0.568	0.278
300	84	10	0	0	0.0336	0.0134
400	84	8	0.317	0.524	0.565	0.329
500	84	16	0.645	0.688	0.636	0.322

虽然平均精度只有0.463，但在‘100’、‘200’、‘500’的类别中，精度均超过了0.6。但由于验证集数量不多，所以上述结果仅作参考，不应当作为模型的真实效果。需要继续增加数据来获得更准确的结果。

F1综合评价指标曲线如下：



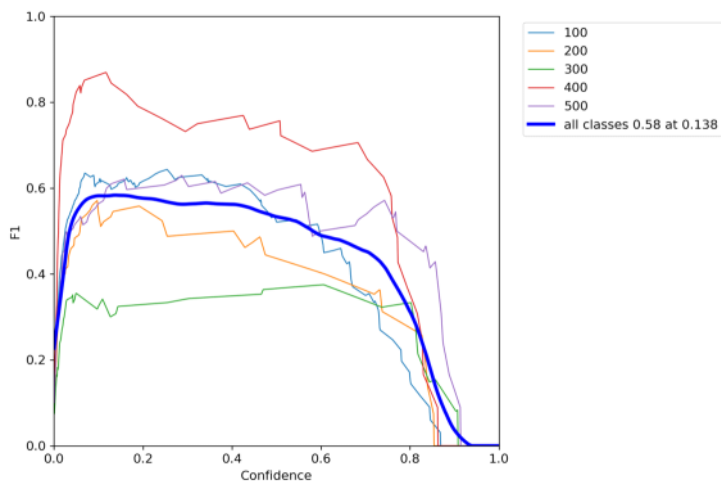
根据筛选数据集1第一次训练的上述结果，将原验证集和原训练集合并成新的训练集，原测试集作为验证集（训练集和验证集仍不存在交集），得到在此数据集上的第二次训练结果如下：



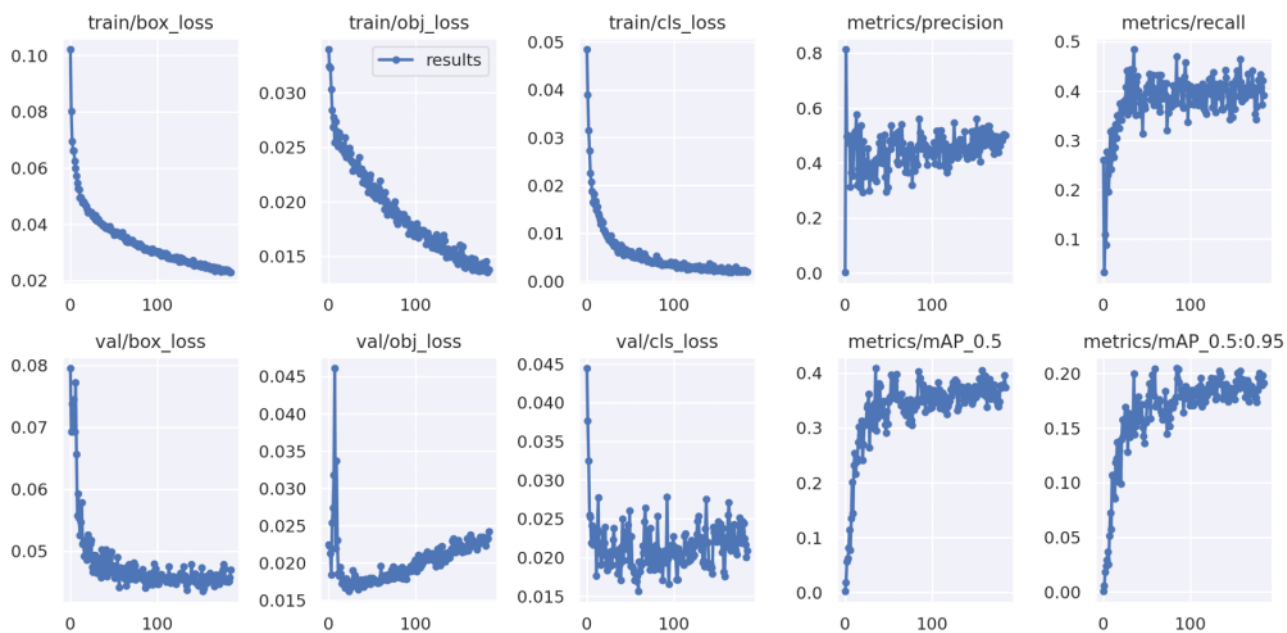
对比第一次在本数据集上训练，第二次结果更加稳定，可以得出筛选数据集1质量较高的结论。最佳模型是第113轮训练的结果。各类别及平均精确率、召回率的结果如下图。平均准确率达到了0.581，平均召回率达到了0.605。

Class	Images	Labels	P	R	mAP@.5	mAP@.5:.95
all	102	158	0.581	0.605	0.593	0.315
100	102	64	0.574	0.632	0.639	0.283
200	102	27	0.666	0.444	0.501	0.192
300	102	23	0.354	0.261	0.27	0.0965
400	102	22	0.827	0.869	0.895	0.542
500	102	22	0.484	0.818	0.66	0.46

F1综合评价指标曲线如下：



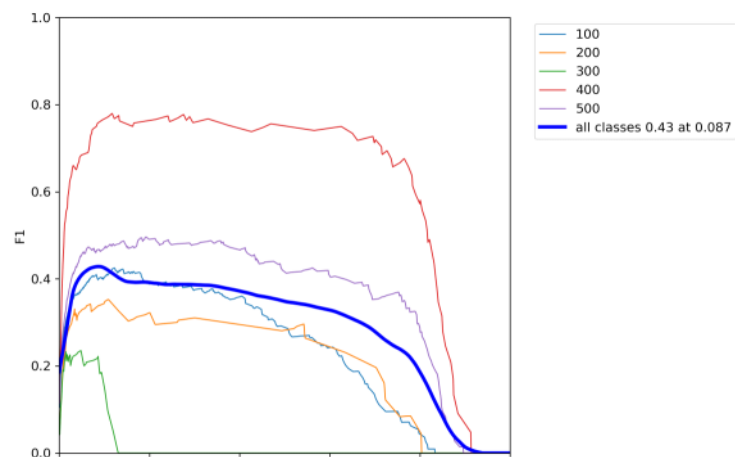
3、筛选数据集2的训练结果

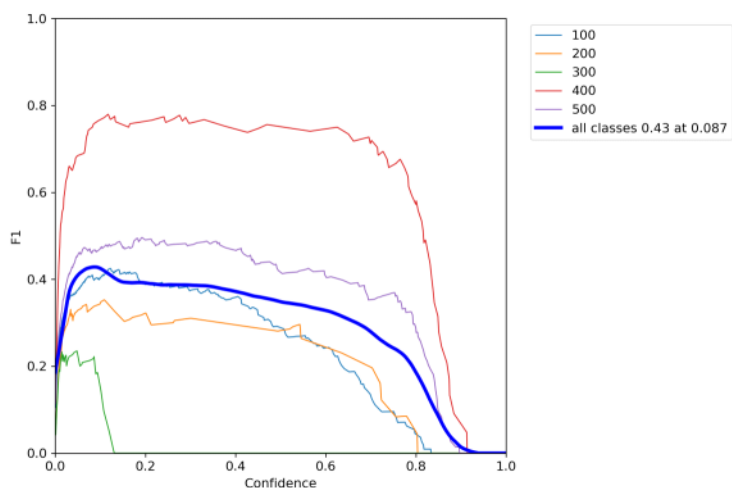


本数据集在训练第84轮得到最佳结果。在最佳结果中，平均精度为0.467，平均召回率为0.461，详情如下。各类别精度较为均匀，但召回率存在极端情况，如‘400’的召回率为0.929，很有可能是验证集中有部分没有标注，所以召回率高。

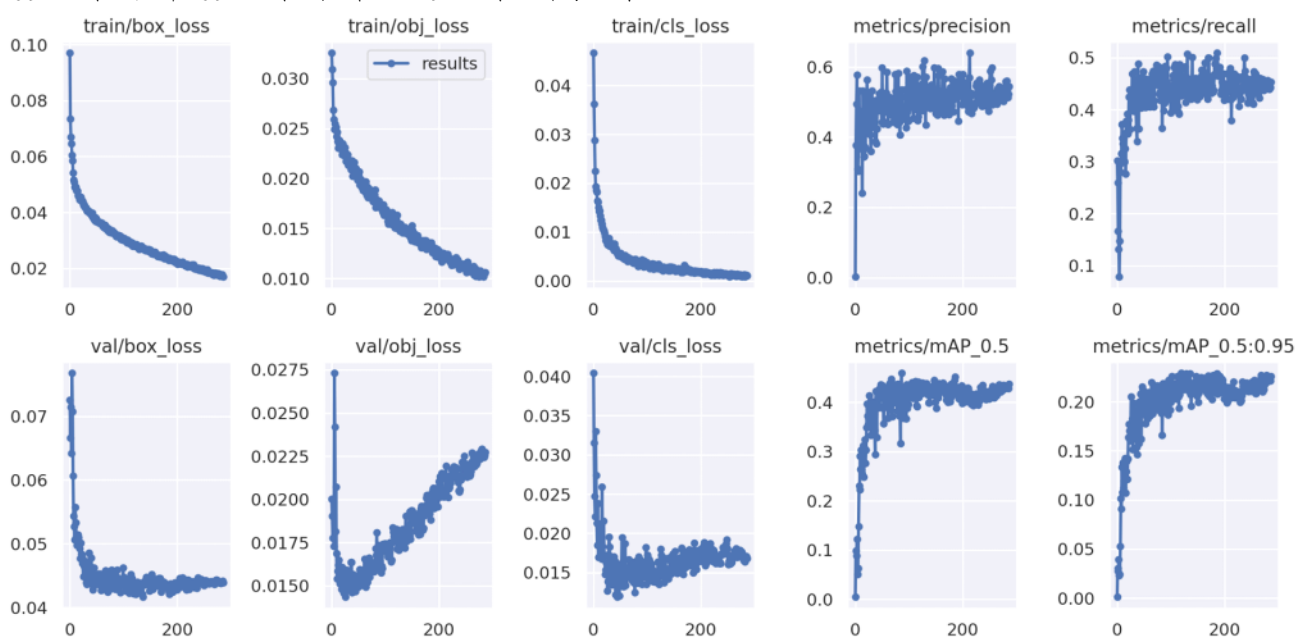
Class	Images	Labels	P	R	mAP@.5	mAP@.5:.95
all	228	477	0.467	0.461	0.402	0.205
100	228	216	0.382	0.426	0.356	0.153
200	228	45	0.467	0.267	0.257	0.123
300	228	35	0.448	0.114	0.119	0.061
400	228	42	0.645	0.929	0.811	0.441
500	228	139	0.395	0.568	0.465	0.246

F1综合评价指标如下：





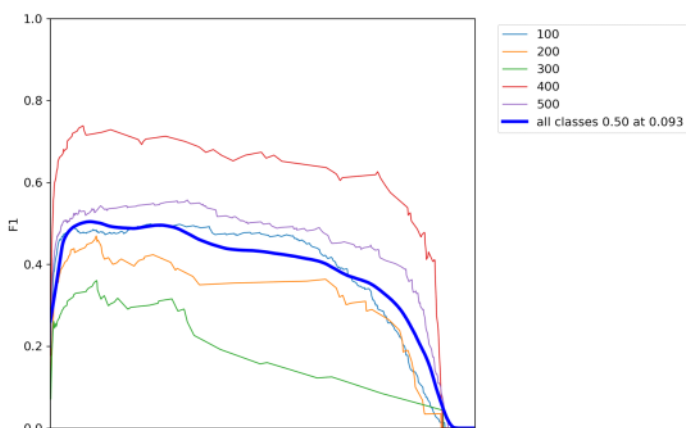
4、筛选数据集1和筛选数据集2合并后的数据集3训练结果：

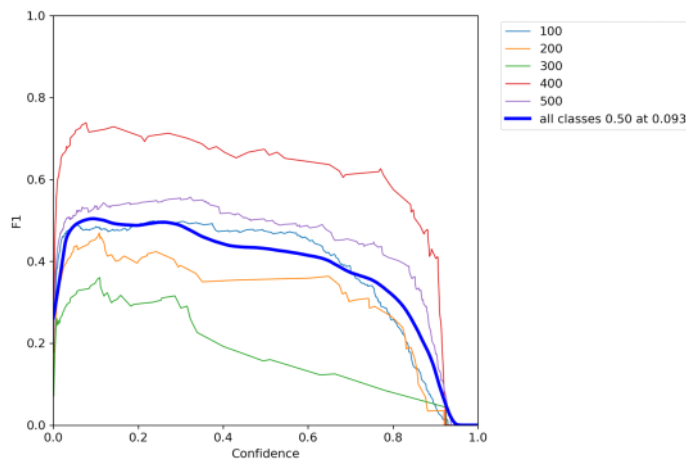


筛选数据集1和筛选数据集2经过合并后，对模型训练的稳定性会有所提高。最佳模型是第186轮的训练结果。对比筛选数据集2，总体表现均有所上涨。

Class	Images	Labels	P	R	mAP@.5	mAP@.5:.95
all	312	586	0.513	0.51	0.452	0.229
100	312	280	0.479	0.486	0.451	0.186
200	312	56	0.481	0.429	0.332	0.135
300	312	45	0.432	0.289	0.25	0.124
400	312	50	0.665	0.78	0.7	0.398
500	312	155	0.508	0.568	0.525	0.3

F1综合评价指标如下：





三、总结

筛选数据集1和筛选数据集2的整体质量会优于原始数据集。表现最佳的是在筛选数据集1上第二次训练的结果，但是筛选数据集1的图片场景复杂程度较低，有可能不符合真实使用场景。筛选数据集2的复杂程度较高，更接近原始数据集的场景。建议根据筛选数据集1和筛选数据集2的场景和各项指标综合改善数据集。