表XI
不同系统的结果

建筑	mAP	AUC	d-prime
CNN6	0.343	0.965	2.568
CNN10	0.380	0.971	2.678
CNN14	0.431	0.973	2.732
ResNet22	0.430	0.973	0.270
ResNet38	0.434	0.974	2.737
ResNet54	0.429	0.971	2.675
MobileNetV1	0.389	0.970	2.653
MobileNetV2	0.383	0.968	2.624
DaiNet[31]	0.295	0.958	2.437
LeeNet11[42]	0.266	0.953	2.371
LeeNet24	0.336	0.963	2.525
Res1dNet31	0.365	0.958	2.444
Res1dNet51	0.355	0.948	2.295
美国有线电视新闻网Wavegram	0.389	0.968	2.612
韦格拉姆·洛格尔CNN	0.439	0.973	2.720

表十二
不同系统的多重加法次数和参数

建筑	多添加	参数
CNN6	二十一点九八六一零	4,837,455
CNN10	二十八点一六六一零九	5,219,279
CNN14	四十二点二二零一零	80,753,615
ResNet22	三十点零八一一零九	63,675,087
ResNet38	四十八点九六二一零九	73,783,247
ResNet54	五十四点五六三一零九	104,318,159
MobileNetV1	三点六一四一零九	4,796,303
MobileNetV2	二点八一零一零	4,075,343
DaiNet	三十点三九五一零九	4,385,807
LeeNet11	4.741×10^9	748,367
LeeNet24	二十六点三六九一零九	10,003,791
Res1dNet31	三十二点六八八一零九	80,464,463
Res1dNet51	六十一点八三三一零九	106,538,063
美国有线电视新闻网Wavegram	四十四点二三四一零九	80,991,759
韦格拉姆·洛格尔CNN	五十三点五一零一零九	81,065,487

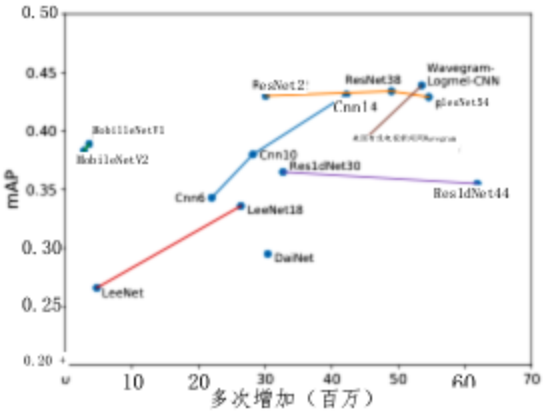

图5。AudioSet标签系统的多重添加与mAP。相同类型的架构以相同的颜色分组

9层CNN等CNN的表现优于深层CNN[33]。一种可能的解释是，较小的数据集可能会受到过拟合的影响，而AudioSet足够大，可以训练更深层次的CNN，至少可以训练到我们研究的14层CNN。

11） ResNets：我们应用ResNets来研究更深层次PANN的性能。表XI显示，ResNet22系统实现了与CNN14系统类似的0.430的mAP。ResNet38的mAP为0.434，略高于其他系统。ResNet54实现了0.429的mAP，这并没有进一步提高性能。。

12） MobileNets：上述系统表明，PANNs在AudioSet标记方面取得了良好的性能。然而，当在便携式设备上实现时，这些系统没有考虑计算效率。我们研究了第II-C节中描述的使用轻量级MobileNets构建PANN。表XI显示MobileNetV1实现了0.389的mAP，比CNN14系统0.431低9.7%。MobileNetV1系统的乘法和加法（多重加法）数量和参数分别仅为CNN14系统的8.6%和5.9%。MobileNetV2系统实现了0.383的mAP，比CNN14低11.1%，并且比MobileNetV1的计算效率更高，其中多重添加和参数的数量仅为CNN14系统的6.7%和5.0%。

13） 一维CNNs：表XI显示了一维CNNs的性能。具有18层的DaiNet[31]实现了0.295的mAP。具有11层的LeeNet11[42]实现了0.266的mAP。我们改进的24层LeeNet将LeeNet11的mAP提高到0.336。我们在第II-D3节中提出的Res1dNet31和Res1dNet 51分别实现了0.365和0.355的mAP，并在一维CNN系统中实现了最先进的性能。

14） Wavegram-Logmel-CNN：表XI的底行显示了我们提出的Wavegram-CNN和Wavegral-Logmel-CNN系统的结果。Wavegram CNN系统实现了0.389的mAP，优于之前最好的一维CNN系统（Res1dNet31）。这个表明波形图是一种有效的学习特征。此外，我们提出的Wavegram-Logmel CNN系统在所有PANN中实现了0.439的最先进mAP。

15） 复杂性分析：我们分析了用于推理的PANN的计算复杂性。多重加法的数量和参数是复杂性分析的两个重要因素。表XII的中间列显示了推断10秒音频片段的多重相加次数。表XII的右栏显示了不同系统的参数数量。CNN14系统的多加数和参数分别为42210万和8080万，大于CNN6和CNN10系统。ResNets22和ResNet38系统的多重添加次数略少于CNN14系统。ResNet54系统包含的多重加法最多，为54.6 10。一维CNN的计算成本与二维CNN相似。性能最好的一维系统Res1dNet31包含32.7 10个多重加法和8050万个参数。我们提出的Wavegram CNN系统包含44.2 10个多加法和8100万个参数，与CNN14相似。Wavegram-Logmel CNN系统将倍数略微增加到53.5 10，参数数量为8110万，与CNN14相似。为了减少多重加法和参数的数量，