|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| External resources | Datasets | Augmentation | Pretrained models | Size | Performance |
| PANN  [Paper](https://arxiv.org/pdf/1912.10211)  [Github](https://github.com/qiuqiangkong/audioset_tagging_cnn/tree/master)  [Download](https://zenodo.org/records/3576403) | Name: Youtube audio clips  Train: 1,934,187  Instances  Test: 18,887 instances  Classes: 527  Type: Timeseries  Fs: 32kHz | Uniform sample each mini batch (balancing)  Mixup  Time-, Frequency masking | Cnn10  Cnn14\_DecisionLevelAtt  Cnn14\_DecisionLevelMax  Cnn4\_emd128  Cnn14\_emd32  Cnn14\_emd512  Cnn14  Cnn6  DaiNet19  LeeNet11  LeeNet24  MobileNetV1  MobileNetV2  Res1dNet31  Res1dNet51  ResNet22  ResNet38  ResNet54  Wavegram\_Cnn14  Wavegram\_Logmel\_Cnn14 | 6,302,735  82,917,948  81,837,071  76,891,151  76,643,855  77,880,335  81,837,071  5,920,911  4,385,807  748,367  10,003,791  5,879,759  5,158,799  80,464,463  106,538,063  64,758,543  74,866,703  105,401,615  80,991,759  82,148,943 | mAP = 0.380  mAP = 0.425  mAP = 0.385  mAP = 0.412  mAP = 0.364  mAP = 0.420  mAP = 0.431  mAP = 0.343  mAP = 0.295  mAP = 0.266  mAP = 0.336  mAP = 0.389  mAP = 0.383  mAP = 0.365  mAP = 0.355  mAP = 0.430  mAP = 0.434  mAP = 0.429  mAP = 0.389  mAP = 0.439 |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |