|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| đcmExternal resources | Datasets | Augmentation | Pretrained models | Size | Performance  (mAP) |
| PANN  [Paper](https://arxiv.org/pdf/1912.10211)  [Github](https://github.com/qiuqiangkong/audioset_tagging_cnn/tree/master)  [Download](https://zenodo.org/records/3576403) | Source: [AudioSet](https://research.google.com/audioset/dataset/index.html)  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  Input:  Log-Melspectrogram  Time series | 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 | 0.380  0.425  0.385  0.412  0.364  0.420  0.431  0.343  0.295  0.266  0.336  0.389  0.383  0.365  0.355  0.430  0.434  0.429  0.389  0.439 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| External resources | Datasets | Augmentation | Pretrained models | Size | Performance  (mAP) | |
| AS-20K | AS-2M |
| CED  [Paper](https://arxiv.org/pdf/2308.11957)  [Github](https://github.com/RicherMans/CED)  [Download](https://github.com/RicherMans/CED) | Source: [AudioSet](https://research.google.com/audioset/dataset/index.html)  Train: 1,904,746  Instances  Test: 18,299 instances  Classes: 527  Type: Timeseries  Fs: 16kHz | Split train data into 2 sets:  Balance: AS-20K.  21,155 instances  Unbalance:  AS-2M.  1.9M instances  Mixup  Input:  Log-Melspectrogram  Time series | [CED-Tiny](https://zenodo.org/record/8275347/files/audiotransformer_tiny_mAP_4814.pt?download=1)  [CED-Mini](https://zenodo.org/record/8275347/files/audiotransformer_mini_mAP_4896.pt?download=1)  [CED-Small](https://zenodo.org/record/8275319/files/audiotransformer_small_mAP_4958.pt?downloa)  [CED-Base](https://zenodo.org/record/8275347/files/audiotransformer_base_mAP_4999.pt?download=1) | 5,5 M  10 M  22 M  86 M | 36.5  38.5  41.6  44.0 | 48.1  49.0  49.6  50.0 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| External resources | Datasets | Augmentation | Pretrained models | Size | Performance  (Accuracy) |
| VGGish  [Paper](https://arxiv.org/pdf/2104.06517)  [Github](https://github.com/tensorflow/models/tree/master/research/audioset/vggish)  (implement on Tensorflow) | [Youtube-8M](https://research.google.com/youtube8m/) | Input:  Log-Melspectrogram | VGGish |  | See table in Paper, only show in the target datasets |

Paper [DCASE-2023](https://dcase.community/documents/challenge2023/technical_reports/DCASE2023_Lv_119_t2.pdf)

Paper [icassp2024](https://ieeexplore.ieee.org/abstract/document/10447183)

Paper [KNN-AnomalyDetection](https://dcase.community/documents/challenge2023/technical_reports/DCASE2023_Jiang_124_t2.pdf)

Pretrained models:

[Wav2vec-XLS-R-300m](https://huggingface.co/facebook/wav2vec2-xls-r-300m)

[UniSpeech](https://huggingface.co/models?other=unispeech-sat)-[Github](https://github.com/microsoft/UniSpeech/tree/main)

[HuBERT](https://huggingface.co/collections/facebook/hubert-651fca95d57549832161e6b6)

[WavLM](https://huggingface.co/models?search=WavLM)

Preprocessing:

Sample random 2 seconds in training [1]

Outlier detector:

KNN: k = 2, [cosine distance](https://medium.com/@milana.shxanukova15/cosine-distance-and-cosine-similarity-a5da0e4d9ded) [2]

2 KNN models for source and target. [2]

Threshold as min distance of source and target [2]

Train configuration:

Input as segment 2s time series

Output as attribute ID

Loss L\_AAM

Fine-Tunning as continue from default checkpoint [email]

Classifier as LinearLayer [email]

Optional:

[Speed Perturbation](https://pytorch.org/audio/stable/generated/torchaudio.transforms.SpeedPerturbation.html)

Weighted sum [2]

Transformer pooling

Submission

Transformer Pooling (between embedding and classifier [1])trained after pre-trained model [email] to ensure same Embedding of the segments in a same time series [1][2]

Calculate w\_i\*S\_i [1]

With w\_i from transformer pooling in ensemble [1,2] ????

Result is mean(Hmean(machinen))(models)

|  |  |
| --- | --- |
| External resources | Download |
| [AudioSet](https://research.google.com/audioset/dataset/index.html) | Kaggle  [Train](https://www.kaggle.com/datasets/zfturbo/audioset)  [Test](https://www.kaggle.com/datasets/zfturbo/audioset-valid) |
| [IDMT-ISA-ELECTRIC-ENGINE](https://www.idmt.fraunhofer.de/en/publications/datasets/isa-electric-engine.html) | [Data](https://zenodo.org/records/7551261) |