List of Pretrained Models on Spectrogram

This document summarizes various pretrained models on spectrograms. These models are useful for tasks such as audio pattern recognition, classification, and feature extraction. Below is a summary of the models with details on input size, file size, pretrained weights availability, and whether they are available in PyTorch.

# 1. CNN6

Input Size: 1000×64 (frames × mel bins)

File Size: 23.7 MB

Description: A CNN model with 4 convolutional layers, average pooling, and a global pooling layer combining avg and max.

Pretrained Weights: Yes

Available in PyTorch: True

Pretrained Weights Filename: Cnn6\_mAP=0.343.pth

Pretrained Weights Link: [PANNs: Large-Scale Pretrained Audio Neural Networks for Audio Pattern Recognition (Pretrained Models) (zenodo.org)](https://zenodo.org/records/3987831)

# 2. CNN10

Input Size: 1000×64 (frames × mel bins)

File Size: 25.2 MB

Description: A deeper CNN with convolutional blocks and a fully connected layer of 2048 units.

Pretrained Weights: Yes

Available in PyTorch: True

Pretrained Weights Filename: Cnn10\_mAP=0.380.pth

Pretrained Weights Link: [PANNs: Large-Scale Pretrained Audio Neural Networks for Audio Pattern Recognition (Pretrained Models) (zenodo.org)](https://zenodo.org/records/3987831)

# 3. CNN14

Input Size: 1000×64 (frames × mel bins)

File Size: 327.4 MB

Description: An even deeper CNN with more convolutional blocks, embedding layers, and dropout for regularization.

Pretrained Weights: Yes

Available in PyTorch: True

Pretrained Weights Filename: Cnn14\_mAP=0.431.pth

Pretrained Weights Link: [PANNs: Large-Scale Pretrained Audio Neural Networks for Audio Pattern Recognition (Pretrained Models) (zenodo.org)](https://zenodo.org/records/3987831)

# 4. CNN14 (Embedding 128)

Input Size: 1000×64 (frames × mel bins)

File Size: 307.6 MB

Description: CNN14 variant with reduced embedding size to 128 units.

Pretrained Weights: Yes

Available in PyTorch: True

Pretrained Weights Filename: Cnn14\_emb128\_mAP=0.412.pth

Pretrained Weights Link: [PANNs: Large-Scale Pretrained Audio Neural Networks for Audio Pattern Recognition (Pretrained Models) (zenodo.org)](https://zenodo.org/records/3987831)

# 5. CNN14 (Embedding 512)

Input Size: 1000×64 (frames × mel bins)

File Size: 311.6 MB

Description: CNN14 variant with embedding size of 512 units.

Pretrained Weights: Yes

Available in PyTorch: True

Pretrained Weights Filename: Cnn14\_emb512\_mAP=0.420.pth

Pretrained Weights Link: [PANNs: Large-Scale Pretrained Audio Neural Networks for Audio Pattern Recognition (Pretrained Models) (zenodo.org)](https://zenodo.org/records/3987831)

# 6. ResNet22

Input Size: Log mel spectrograms

File Size: 259.1 MB

Description: ResNet architecture with 22 layers suitable for capturing detailed patterns in audio spectrograms.

Pretrained Weights: Yes

Available in PyTorch: True

Pretrained Weights Filename: ResNet22\_mAP=0.430.pth

Pretrained Weights Link: [PANNs: Large-Scale Pretrained Audio Neural Networks for Audio Pattern Recognition (Pretrained Models) (zenodo.org)](https://zenodo.org/records/3987831)

# 7. ResNet38

Input Size: Log mel spectrograms

File Size: 299.6 MB

Description: Deeper ResNet with additional residual blocks to enhance representation learning.

Pretrained Weights: Yes

Available in PyTorch: True

Pretrained Weights Filename: ResNet38\_mAP=0.434.pth

Pretrained Weights Link: [PANNs: Large-Scale Pretrained Audio Neural Networks for Audio Pattern Recognition (Pretrained Models) (zenodo.org)](https://zenodo.org/records/3987831)

# 8. ResNet54

Input Size: Log mel spectrograms

File Size: 421.9 MB

Description: A deeper ResNet architecture with more residual blocks for complex pattern recognition.

Pretrained Weights: Yes

Available in PyTorch: True

Pretrained Weights Filename: ResNet54\_mAP=0.429.pth

Pretrained Weights Link: [PANNs: Large-Scale Pretrained Audio Neural Networks for Audio Pattern Recognition (Pretrained Models) (zenodo.org)](https://zenodo.org/records/3987831)

# 9. MobileNetV1

Input Size: Log mel spectrograms

File Size: 23.6 MB

Description: MobileNet architecture with depthwise separable convolutions designed for efficiency on mobile devices.

Pretrained Weights: Yes

Available in PyTorch: True

Pretrained Weights Filename: MobileNetV1\_mAP=0.389.pth

Pretrained Weights Link: [PANNs: Large-Scale Pretrained Audio Neural Networks for Audio Pattern Recognition (Pretrained Models) (zenodo.org)](https://zenodo.org/records/3987831)

# 10. MobileNetV2

Input Size: Log mel spectrograms

File Size: 20.8 MB

Description: Improved MobileNet architecture with inverted residuals and linear bottlenecks for efficiency.

Pretrained Weights: Yes

Available in PyTorch: True

Pretrained Weights Filename: MobileNetV2\_mAP=0.383.pth

Pretrained Weights Link: [PANNs: Large-Scale Pretrained Audio Neural Networks for Audio Pattern Recognition (Pretrained Models) (zenodo.org)](https://zenodo.org/records/3987831)

# 11. Wavegram-Logmel-CNN14

Input Size: Log mel spectrograms + waveform

File Size: 328.7 MB

Description: Combines raw waveform and log mel spectrogram inputs using CNN14 architecture for improved performance.

Pretrained Weights: Yes

Available in PyTorch: True

Pretrained Weights Filename: Wavegram\_Logmel\_Cnn14\_mAP=0.439.pth

Pretrained Weights Link: [PANNs: Large-Scale Pretrained Audio Neural Networks for Audio Pattern Recognition (Pretrained Models) (zenodo.org)](https://zenodo.org/records/3987831)

# 12. VGGish

Input Size: 96×64 (frames × mel bins)

File Size: ~280 MB

Description: VGG-like CNN with multiple convolutional and max-pooling layers often used for feature extraction.

Pretrained Weights: Yes

Available in PyTorch: No

Pretrained Weights Filename: vggish\_model.ckpt

Pretrained Weights Link: https://storage.googleapis.com/audioset/vggish\_model.ckpt

# 13. YAMNet

Input Size: 64 mel bins

File Size: ~15 MB

Description: Based on MobileNetV1 architecture uses depthwise separable convolutions for audio tagging and classification.

Pretrained Weights: Yes

Available in PyTorch: No

Pretrained Weights Filename: yamnet.h5

Pretrained Weights Link: https://storage.googleapis.com/audioset/yamnet.h5

# 14. OpenL3

Input Size: Log mel spectrograms or linear spectrograms

File Size: ~30-100 MB

Description: Custom CNN model for extracting embeddings for various audio tasks available in music and environmental sound configurations.

Pretrained Weights: Yes

Available in PyTorch: No

Pretrained Weights Filename: openl3\_audio\_mel128\_music\_embedding512.h5

Pretrained Weights Link: Couldn’t find direct link, but official github repo is available