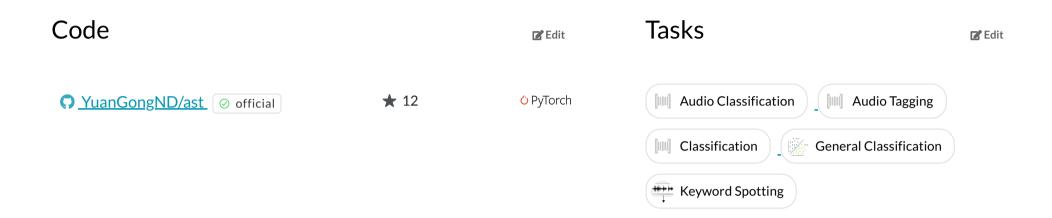
## **AST: Audio Spectrogram Transformer**

5 Apr 2021 · Yuan Gong, Yu-An Chung, James Glass · 🗷 Edit social preview

In the past decade, convolutional neural networks (CNNs) have been widely adopted as the main building block for end-to-end audio classification models, which aim to learn a direct mapping from audio spectrograms to corresponding labels. To better capture long-range global context, a recent trend is to add a self-attention mechanism on top of the CNN, forming a CNN-attention hybrid model... read more







## Results from the Paper

Ranked #1 on Keyword Spotting on Speech Commands (using extra training data)

Z Edit

→ Get a GitHub badge

Task	Dataset	Model	Metric Name	Metric Value	Global Rank	Uses Extra Training Data	Result	Benchmark
Audio Tagging	AudioSet	Audio Spectrogram Transformer	mean average precision	0.485	#1	~	Ð	Compare
Audio Classification	AudioSet	Audio Spectrogram Transformer	Test mAP	0.485	# 1	~	Ð	Compare
Audio Classification	ESC-50	Audio Spectrogram Transformer	Top-1 Accuracy	95.7	# 1	✓	Ð	Compare
			PRE-TRAINING DATASET	AudioSet, ImageNet	# 1	✓	Ð	Compare
			Accuracy (5-fold)	95.7	# 1	✓	Ð	Compare
Keyword Spotting	Google Speech Commands	Audio Spectrogram Transformer	Google Speech Commands V2 35	98.11	# 1	✓	Ð	Compare

Keyword Speech Audio Spectrogram Google Speech 98.11 #1 ✓ → Compare Spotting Commands Transformer Commands V2 35

Methods

<u>Adam • BPE • Dense Connections • Dropout • Label Smoothing • Layer Normalization • Multi-Head Attention • Residual Connection • Scaled Dot-Product Attention • Softmax • Transformer</u>

Contact us on: <u>■ hello@paperswithcode.com</u>. Papers With Code is a free resource with all data licensed under CC-BY-SA.

Terms Privacy Cookies policy