

Model	MTAT	MSD
Bag of multi-scaled features	0.8980	-
End-to-end	0.8815	-
Transfer learning	0.8800	-
Persistent CNN	0.9013	-
Time-Frequency CNN	0.9007	-
Timbre CNN	0.8930	-
2D CNN	0.8940	0.8510
CRNN	-	0.8620
Multi-level & multi-scale	0.9017 [†]	0.8878[†]
SampleCNN multi-features	0.9064 [†]	0.8842
SampleCNN	0.9055	0.8812
SE [This work]	0.9111	0.8840
ReSE [This work]	0.9113	0.8847