Continuous Training with Transfer Learning

Training Pipeline: Need Lots of Data



1000 Classes

Training Pipeline: Need Compute Resources







Compute





Training Pipeline: Need Compute Resources

Computationally Intensive Repeated Many Times (Epochs)

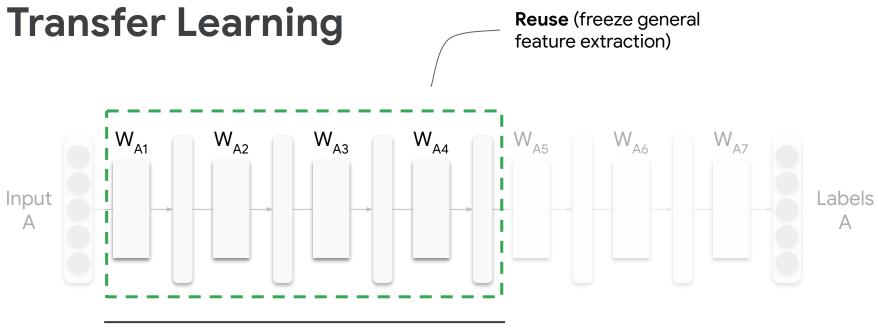


Memory

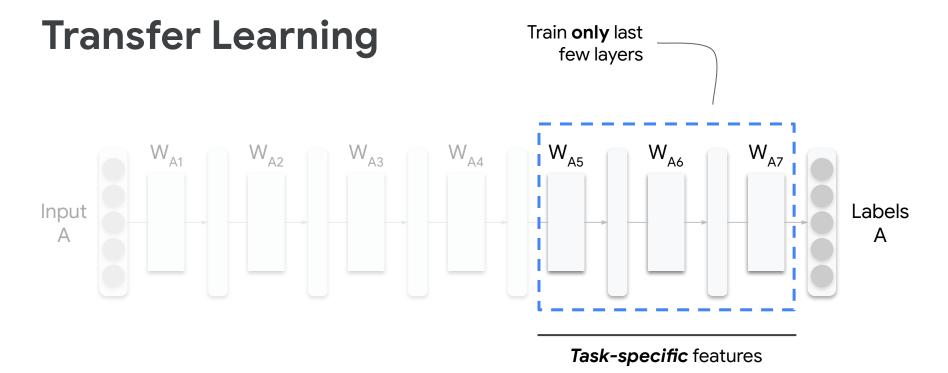


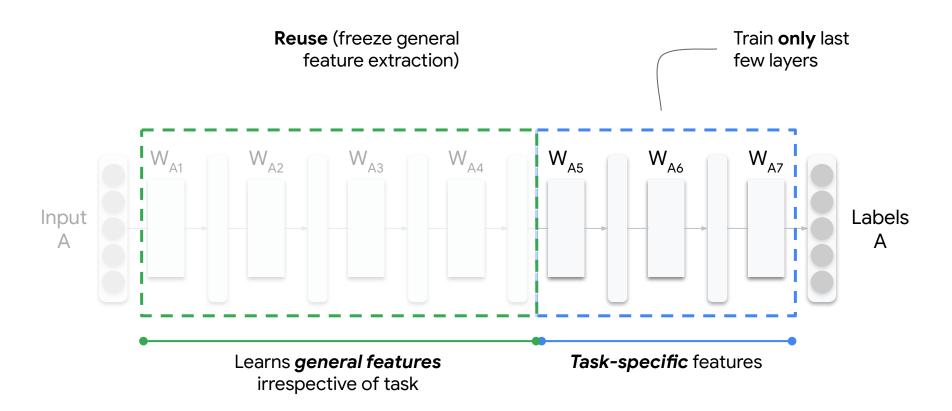


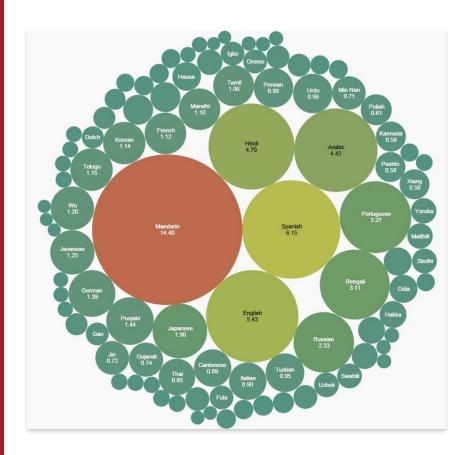




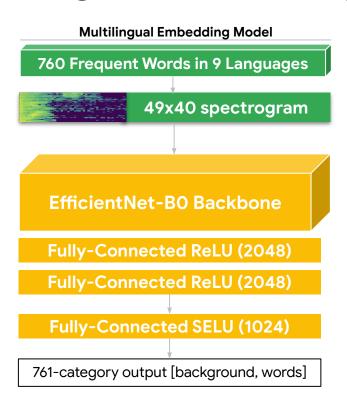
Learns *general features* irrespective of task







- Speech commands for the whole planet?
- For more than just voice assistants



Simple classifier for 760 frequent words in 9 languages

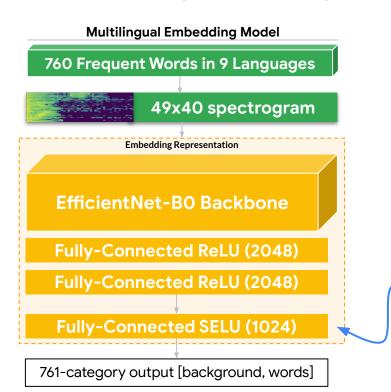
Multilingual Embedding Model

760 Frequent Words in 9 Languages 49x40 spectrogram **EfficientNet-BO Backbone Fully-Connected ReLU (2048) Fully-Connected ReLU (2048) Fully-Connected SELU (1024)**

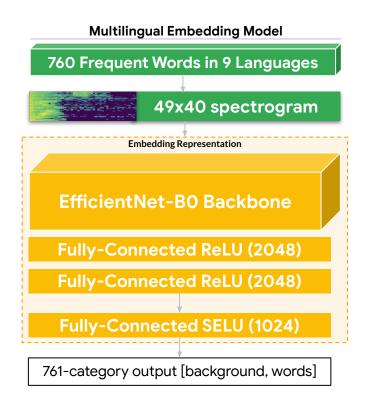
761-category output [background, words]

Language	# words	# train
English	265	518760
German	152	287100
French	105	205920
Kinyarwanda	68	134640
Catalan	80	132660
Persian	35	69300
Spanish	31	61380
Italian	17	31680
Dutch	7	13860
Model	760	1455300

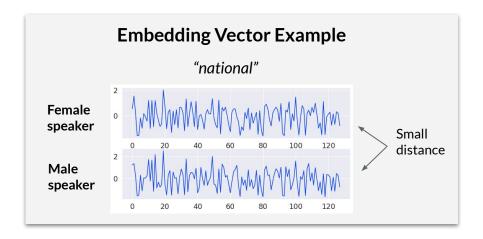
Few-Shot Keyword Spotting



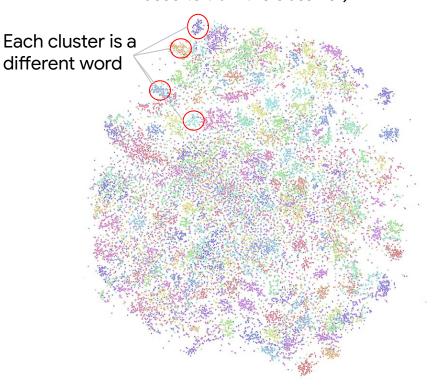
Use penultimate layer for embedding representation in keyword-spotting model



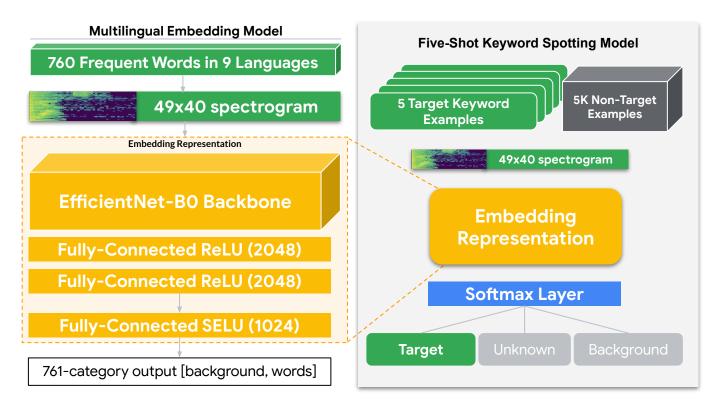
Penultimate layer **output** as embedding vector



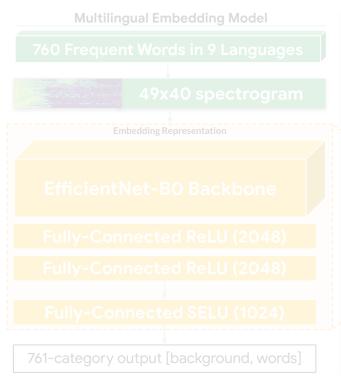
Multilingual Embedding Model 760 Frequent Words in 9 Languages 49x40 spectrogram **Embedding Representation EfficientNet-BO Backbone Fully-Connected ReLU (2048) Fully-Connected ReLU (2048) Fully-Connected SELU (1024)** 761-category output [background, words] t-SNE view of 165 novel words (not used to train the classifier)

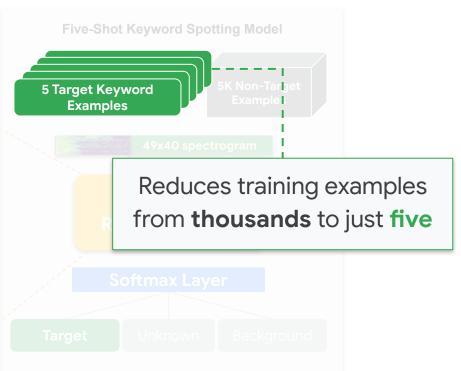


Few-Shot Keyword Spotting



Few-Shot Keyword Spotting





Transfer Learning

