Bespoke Neural Networks for Score-Informed Source Separation

Ethan Manilow and Bryan Pardo
Interactive Audio Lab
Northwestern University

Click here to Listen!

Motivation

Deep Learning Source Separation Systems work **really, really well**!!!

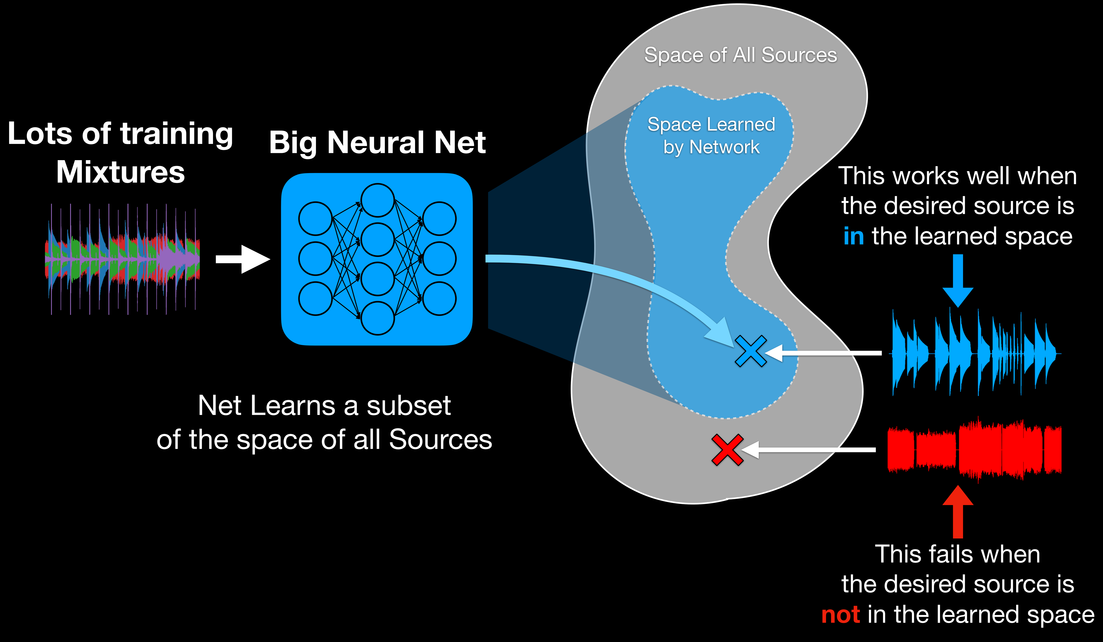
But…

Top performing source separation systems require a ***lot of data***!!!

So…

1. What if there's ***no*** training data for the source we want?
2. What if we don't have ***enough*** data?
3. What if we have the data, but the net ***still*** doesn't work?

Typical Training for Source Separation

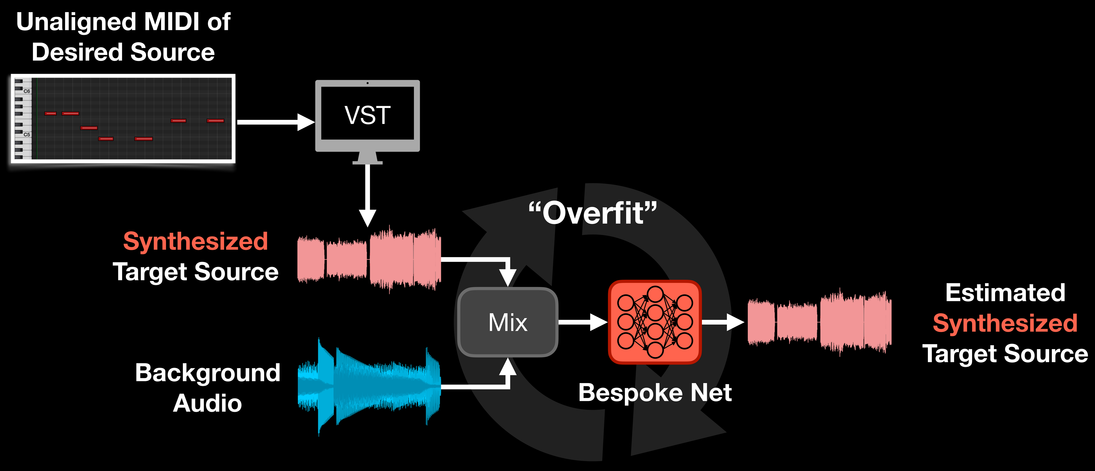


Why learn all that space?

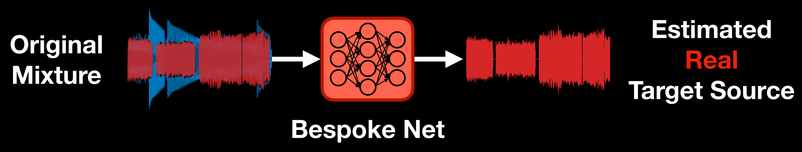
Bespoke Nets

- Given Unaligned MIDI for target source
- Synthesize MIDI & use as source labels
- Train a small network to “overfit”
- Run network on Original Mix

Training



Inference



Advantages

- Only need unaligned MIDI
- Separate any source type that you can synthesize
- Net can be small, few iterations ⟶ Much faster than training traditional net

Demos

Network

- 2 layer BLSTM Mask Inference net w/ 300 nodes
- 2000 iterations using tPSA objective

Mixing

- Target & Background gain chosen randomly & Mixed
- Dynamic Range Compression applied randomly to Mix

[Click here to Listen!]