

PAT 498/598 (Winter 2025)

Music & AI

Lecture 12: Source Separation

Instructor: Hao-Wen Dong



SCHOOL OF MUSIC, THEATRE & DANCE
PERFORMING ARTS TECHNOLOGY
UNIVERSITY OF MICHIGAN

(Recap) Music Classification Tasks

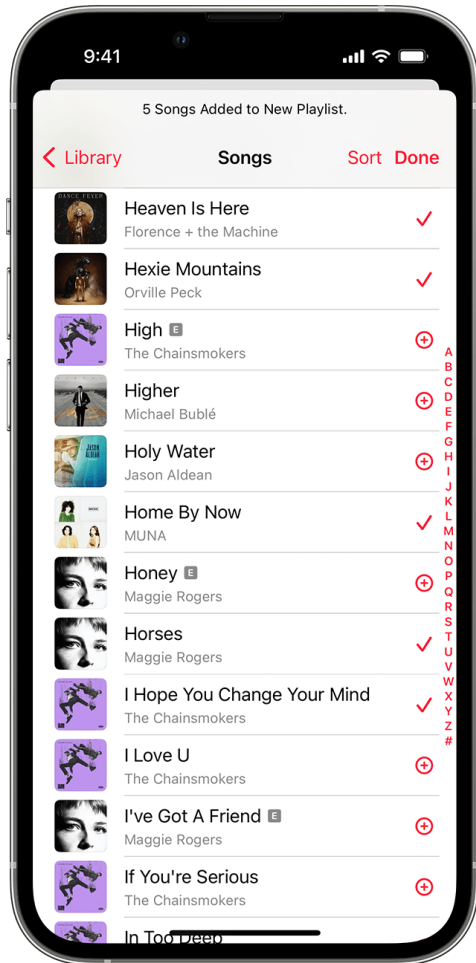
- **Genre** classification (pop, rock, r&b, jazz, hip-hop, classical, etc.)
- **Mood** classification (happy, sad, calm, aggressive, cheerful, etc.)
- **Instrument** recognition
- **Composer** identification
- **Key** detection
- **Chord** estimation
- **Music tagging** → **Can cover everything above!**

(Recap) Music Classification for Recommendation

What to play next?



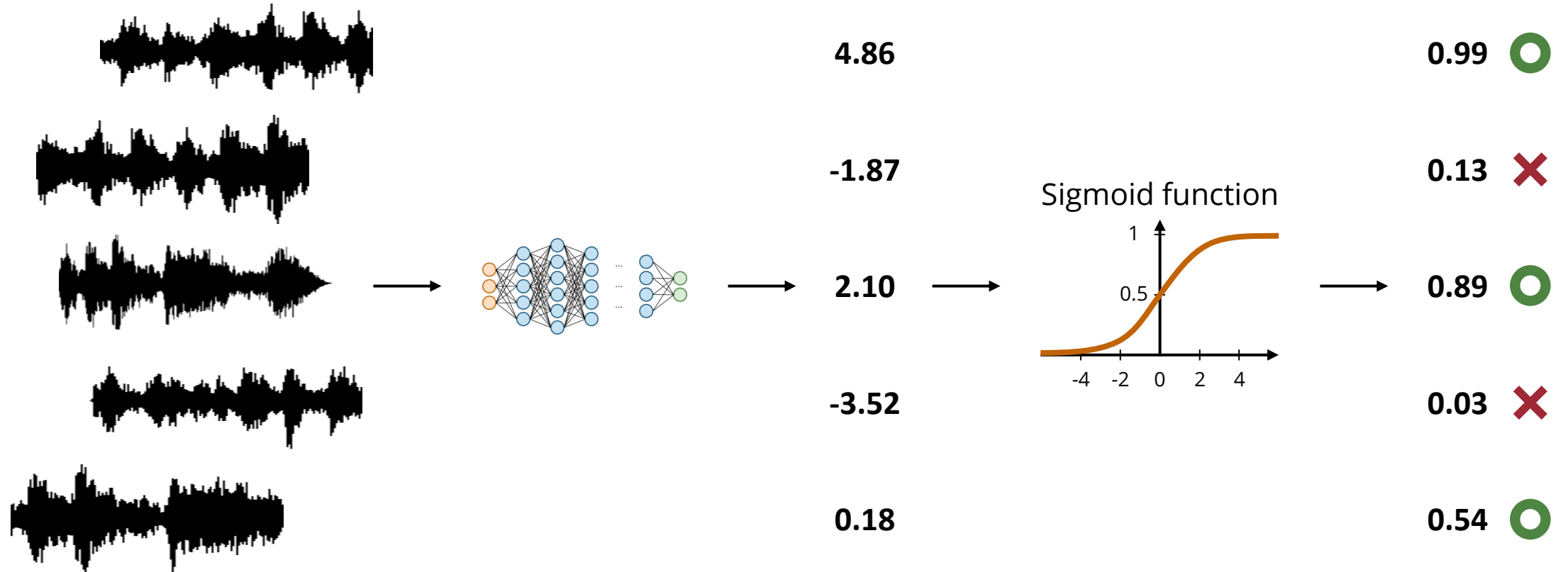
(Recap) Music Classification for Playlist Generation



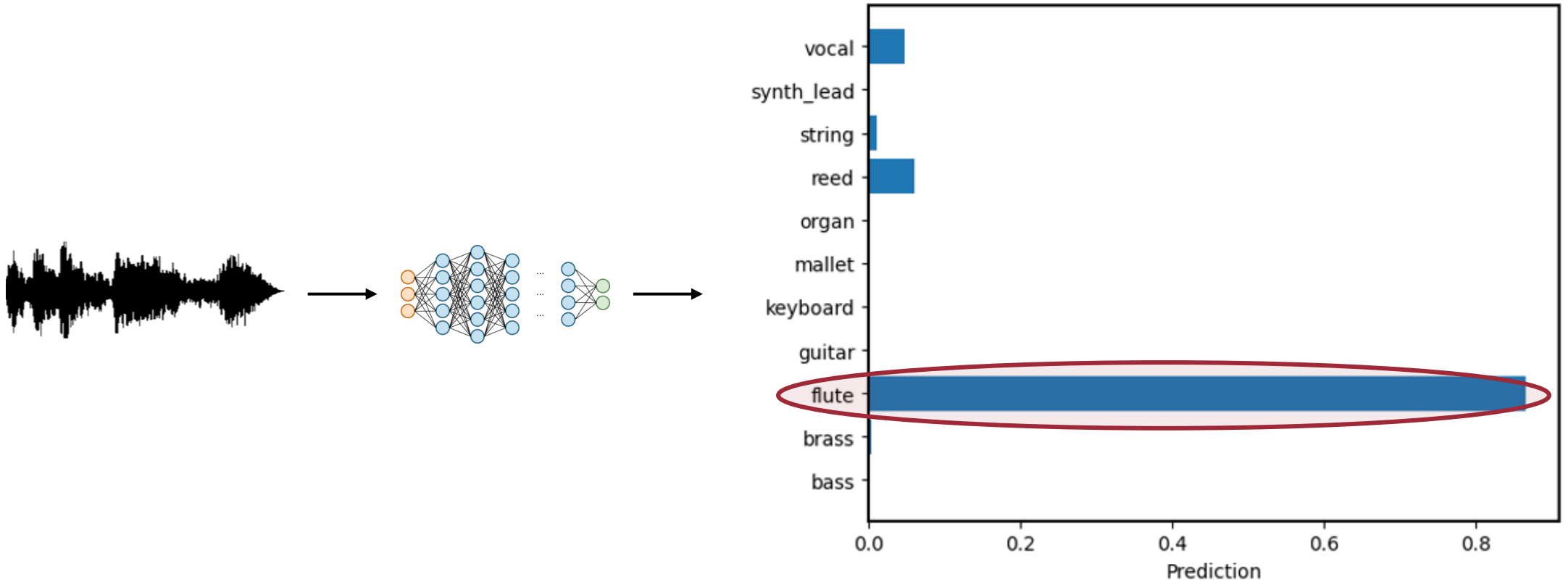
(Recap) Types of Classification Tasks

- **Binary** classification
- **Multiclass** classification
- **Multi-label** classification

(Recap) Binary Classification



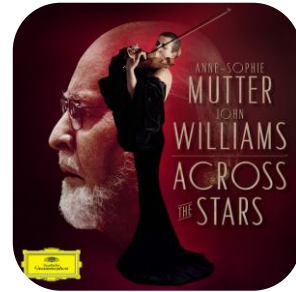
(Recap) Multiclass Classification



(Recap) Multi-label Classification



Soul, R&B, pop



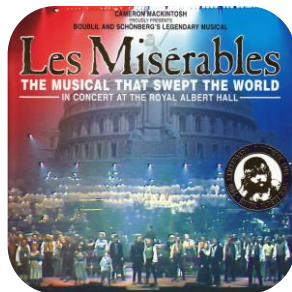
Soundtrack, classical



Rock



Classical



Musical



Alternative, rock

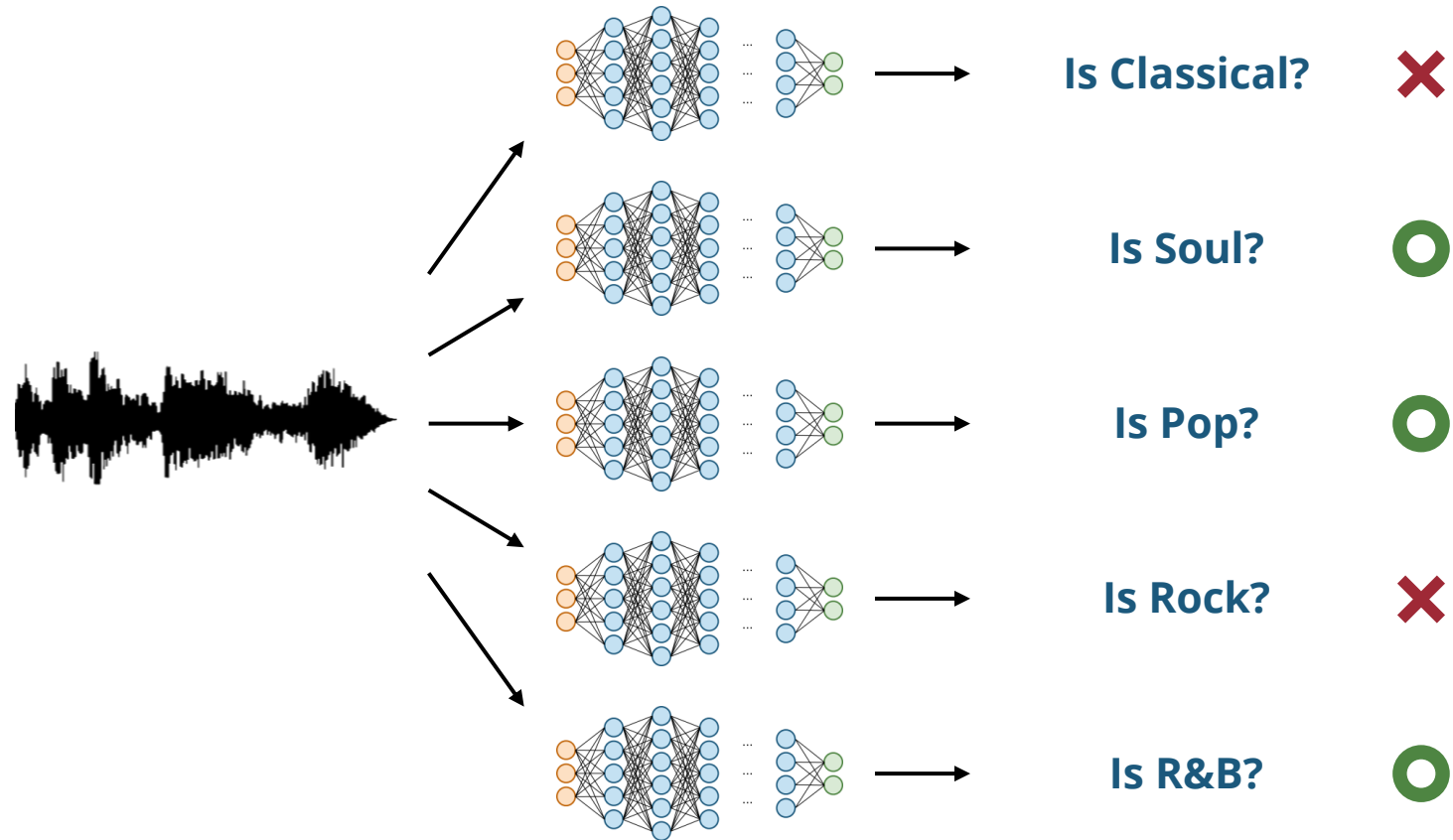


Pop



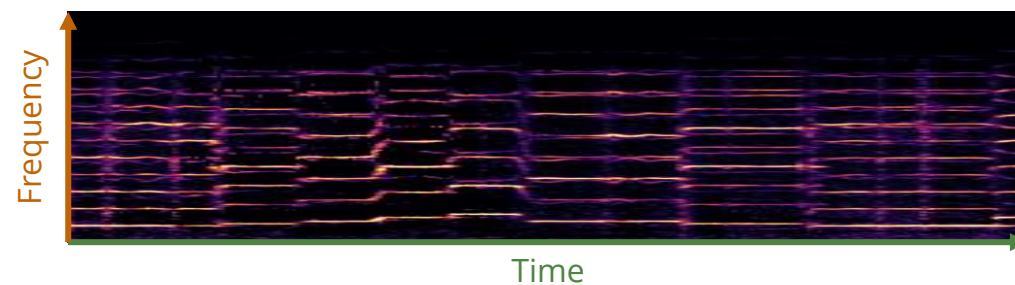
Pop, soul, R&B

(Recap) Multi-label Classification as Binary Classification



(Recap) Input Features

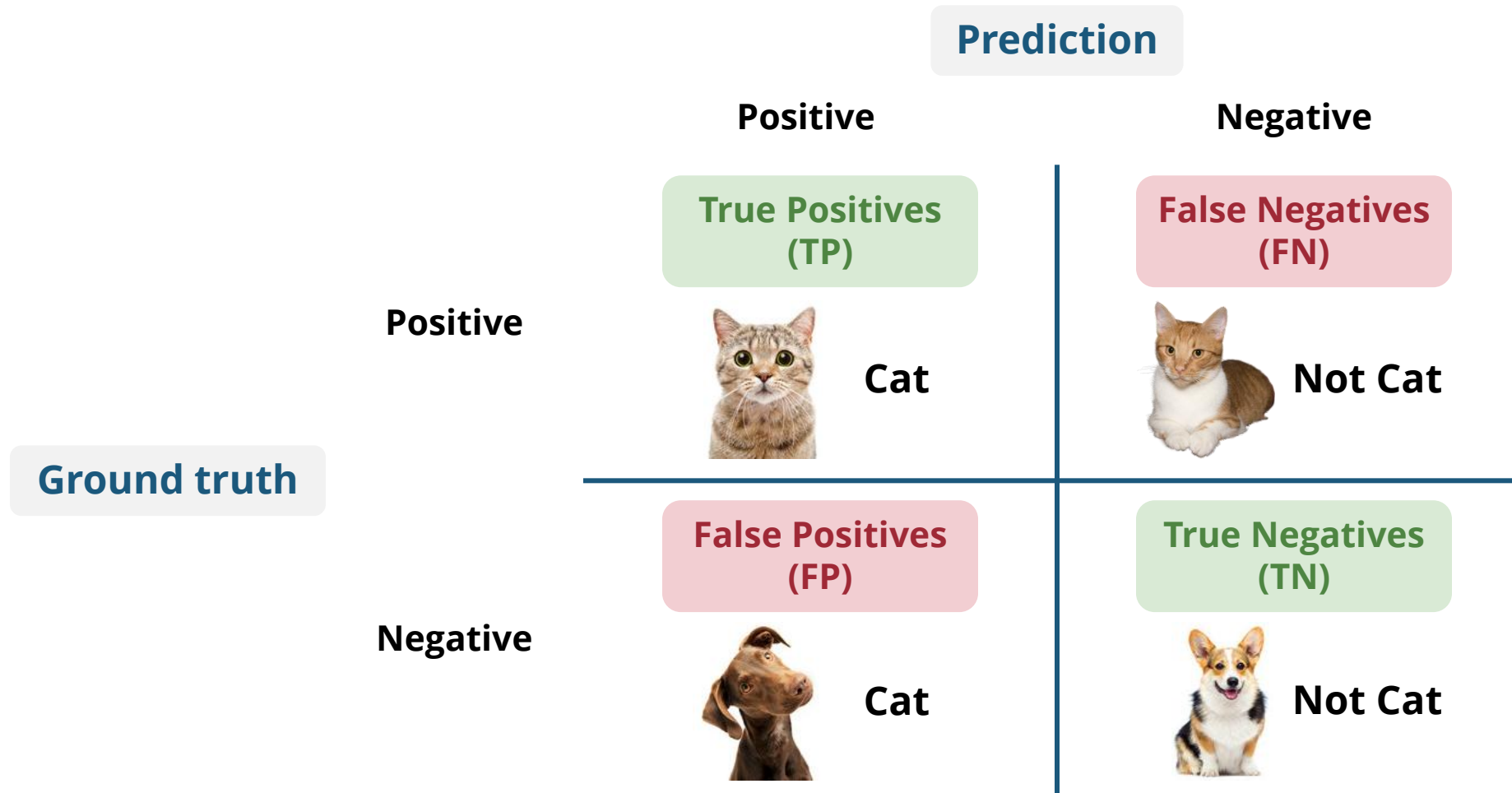
- **Waveform**
- Time-frequency representation (**spectrograms**)
- **Hand-crafted features** or **features provided in metadata**
 - Acoustic: loudness, pitch, timbre
 - Rhythmic: beat, tempo, time signature
 - Tonal: key, scale, chords
 - Instrumentation, expressions, structures, etc.



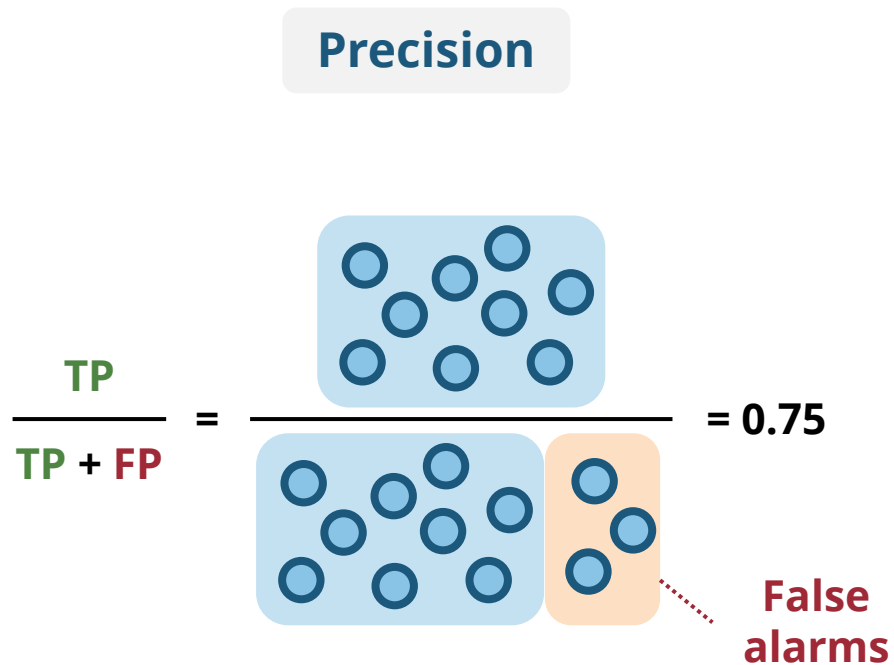
(Recap) Common Datasets

- **GTZAN**: 1,000 30-sec songs, 10 genres
- **MagnaTagATune**: 5,405 29-sec songs, 188 tags, 230 artists
- **Million Song Dataset (MSD)**: **1M** 30-sec songs, >500K tags, **tricky to access**
- **Free Music Archive (FMA)**: >10K **full** songs, 163 genres
- **MTG-Jamendo**: 55K **full** songs, 195 tags
- **AudioSet**: **1M** songs, YouTube URLs, **low-quality audio**
- **NSynth**: ~306K 4-sec instrument sounds

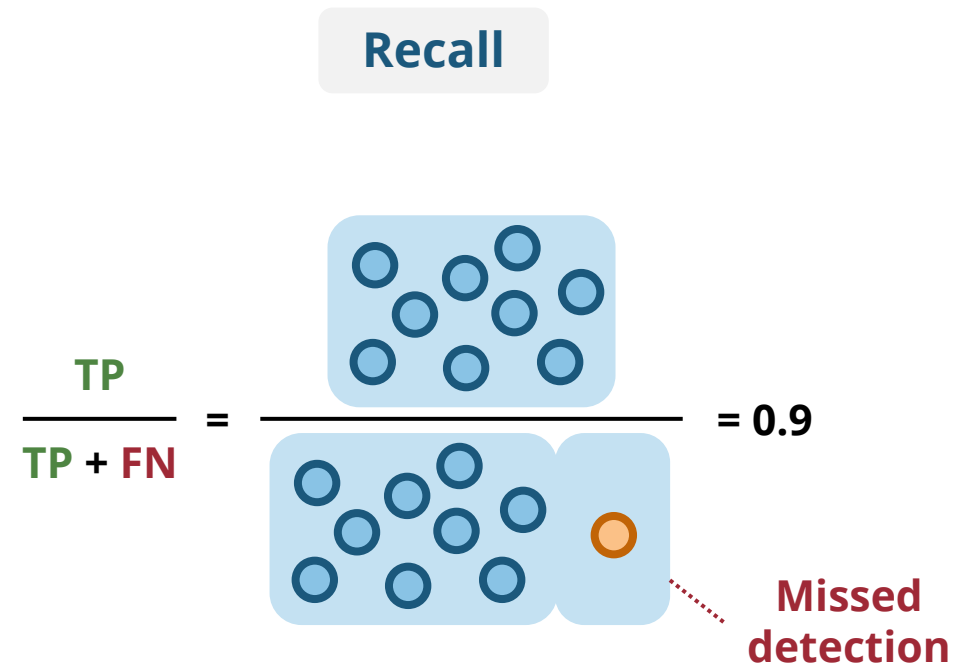
(Recap) Confusion Matrix for Binary Classification



(Recap) Precision vs Recall



How often predictions for the positive are correct



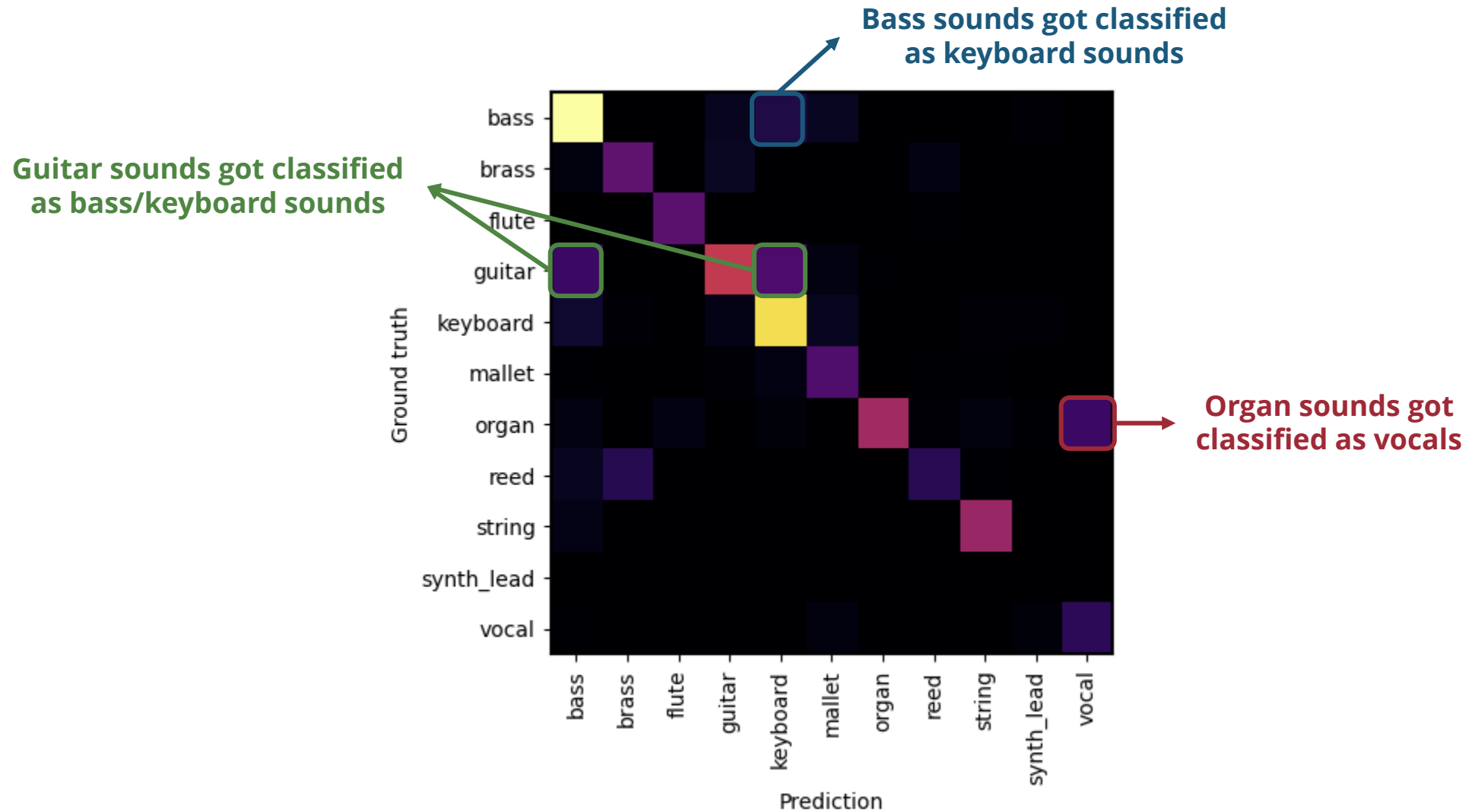
How well the model finds all positive instances in the dataset

(Recap) F1 Score: Considering both Precision & Recall

- Particularly useful for **imbalanced datasets**
 - Work better than accuracy when the dataset is imbalanced
 - For example, music search, retrieval, and recommendation

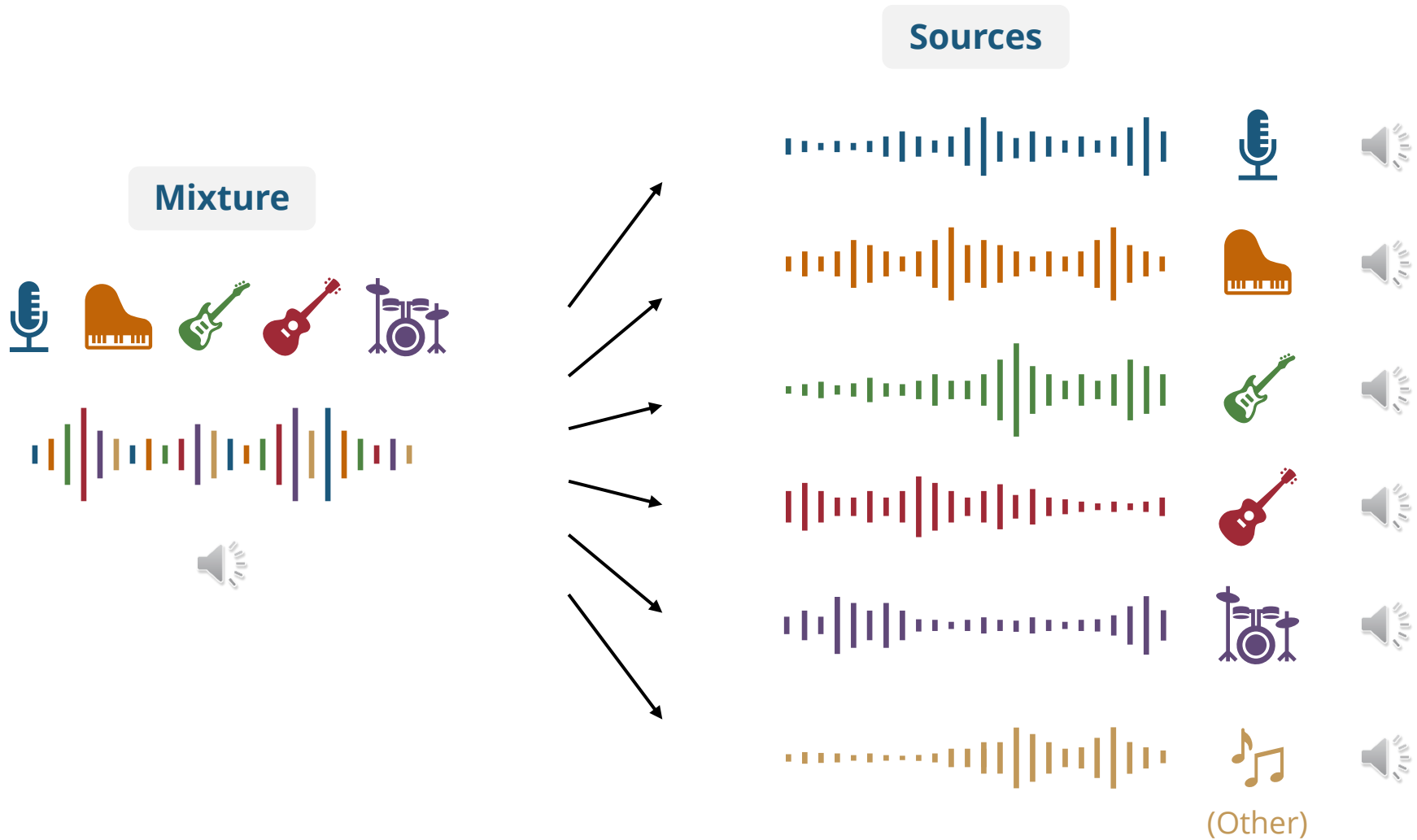
$$\begin{aligned} F_1 &= \frac{2}{\frac{1}{Precision} + \frac{1}{Recall}} \\ &= \frac{2 \cdot Precision \cdot Recall}{Precision + Recall} \end{aligned}$$

(Recap) Confusion Matrix for Multi-label Classification

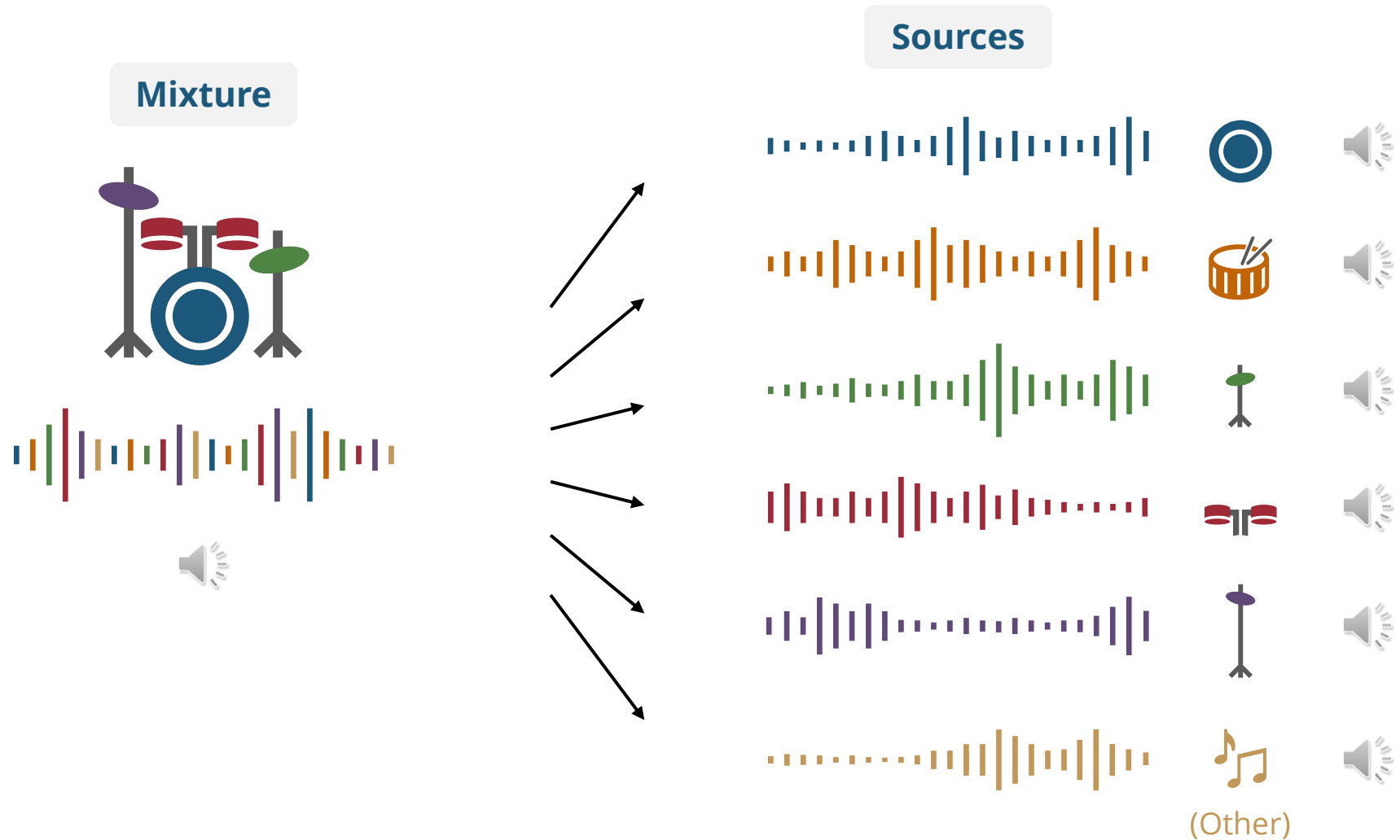


Source Separation

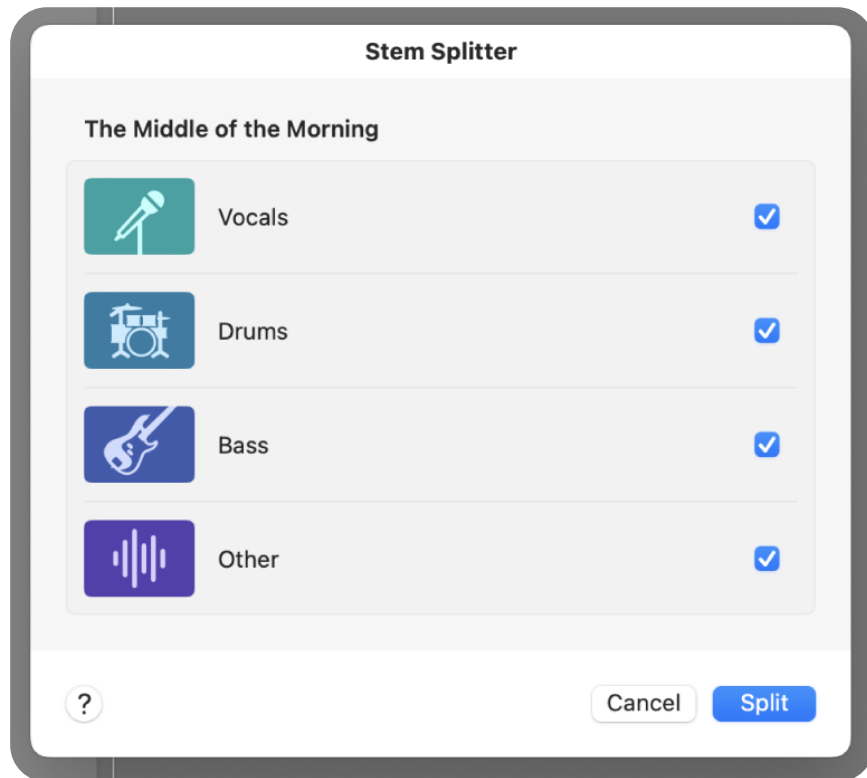
Source Separation



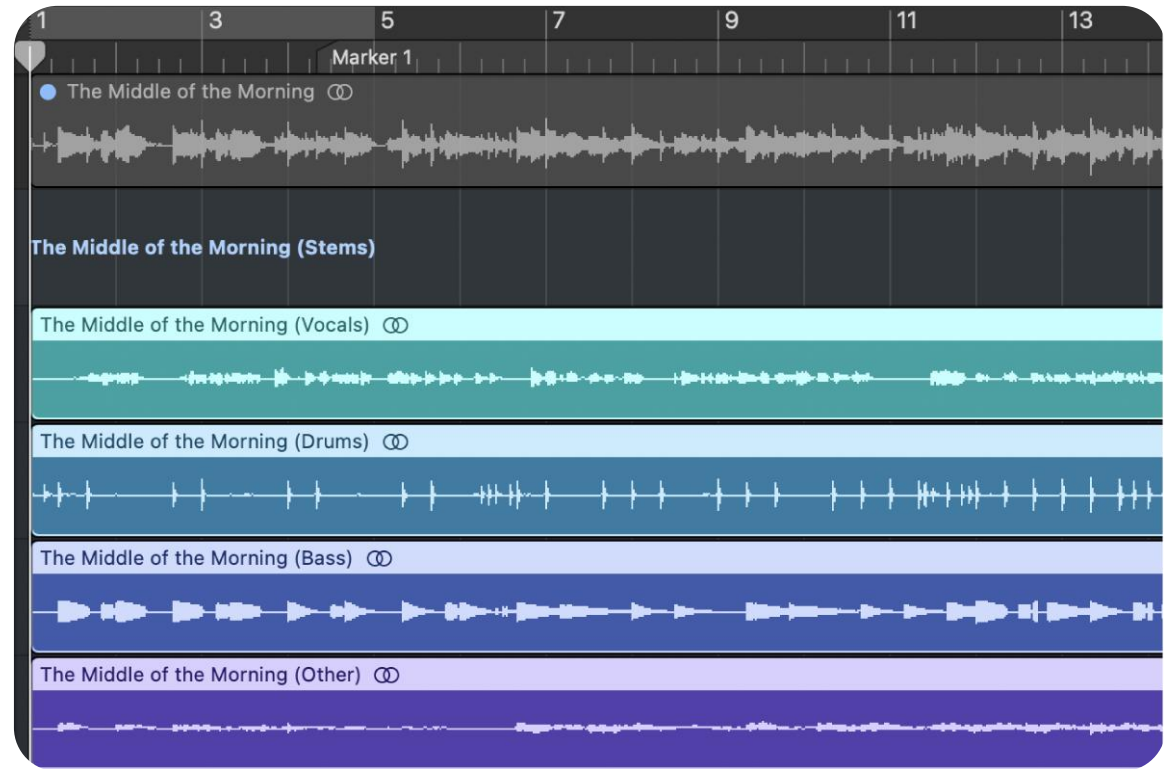
Source Separation



Stem Splitter in Logic Pro

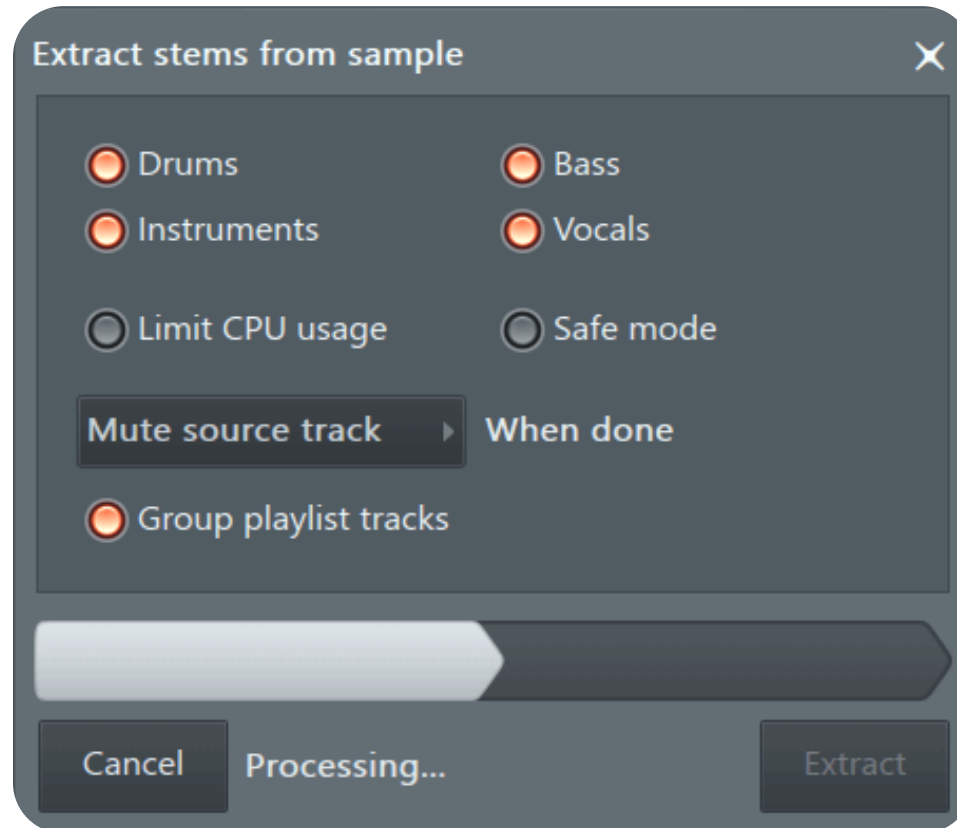


(Source: Logic Pro User Guide)



(Source: Logic Pro User Guide)

Extracting Stems from Sample in FL Studio

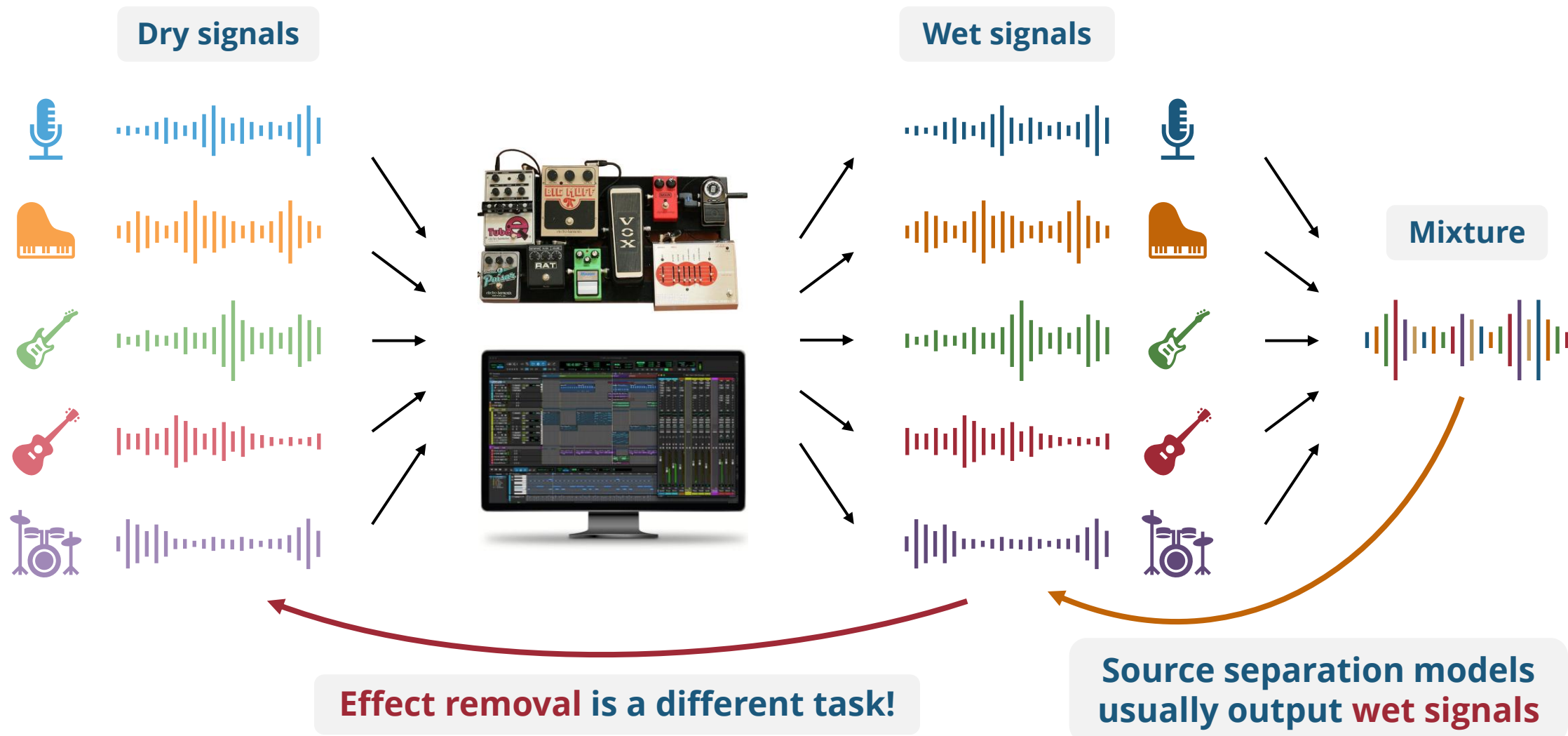


(Source: FL Studio Reference Manual)

Source Separation

- Also known as
 - **Stem separation**
 - **Stem splitter**
 - **Music demixing** → slightly different meaning
 - **Stem extraction** → slightly different meaning

Source Separation does NOT Remove Effects



Applications of Source Separation

- **Musical applications**
 - Remixing & sampling
 - Music practicing & education
 - Karaoke accompaniment generation
- **MIR tasks** (Oftentimes source separation is the first step)
 - Music transcription
 - Musical instrument & vocal detection
 - Singer identification
 - Lyric recognition
 - Lyric-to-music alignment

Moises

Moises

- 🎉 **Free Moises Pro license** until Summer 2025
- Register at studio.moises.ai/claim-trial/UMichFree/monthly/
 - Use your **U-M email** (@umich.edu)
 - Sign up in your **desktop browser**
 - Ignore the prompt to upgrade your account
 - **Deadline to sign up: March 14**

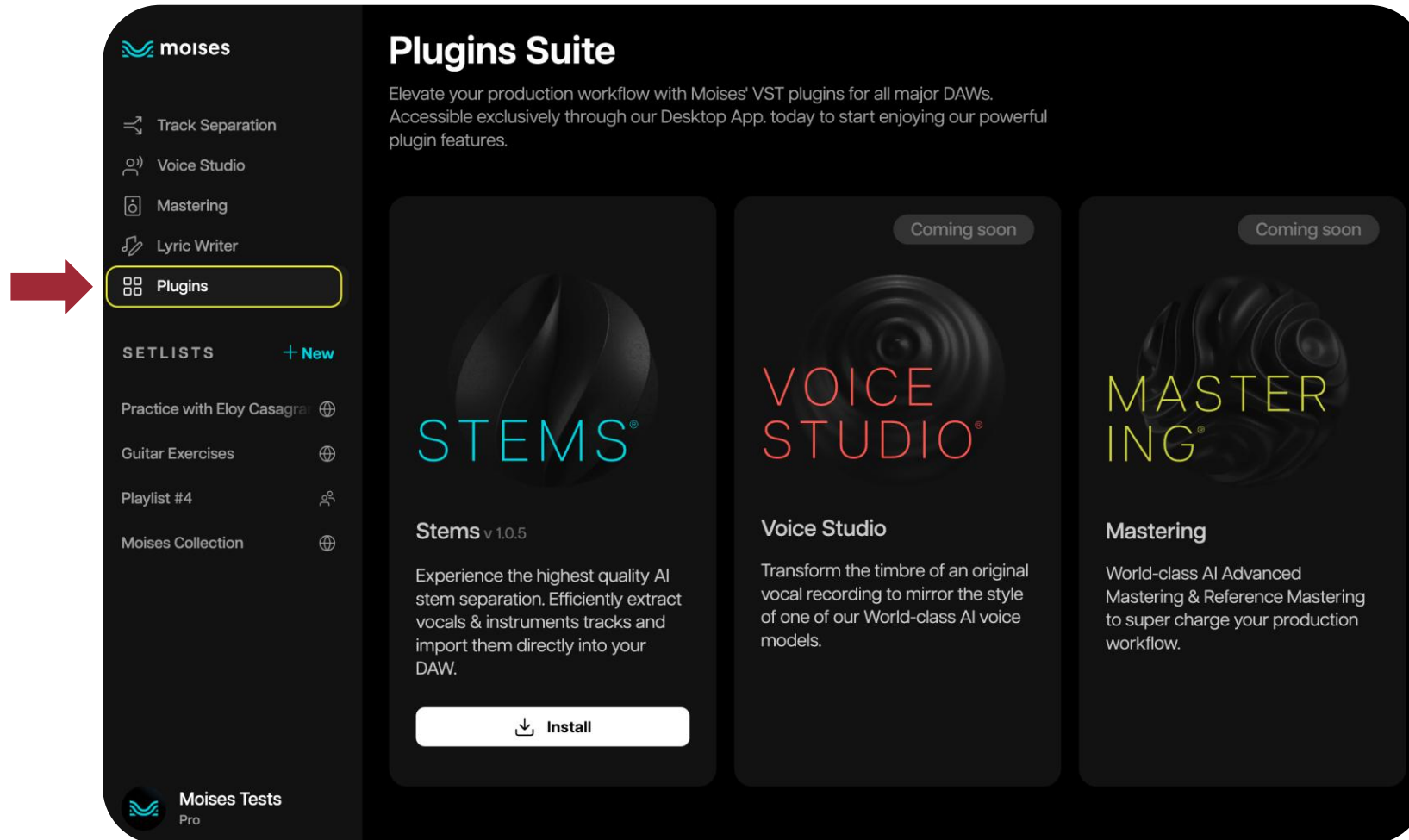


Moises Demo



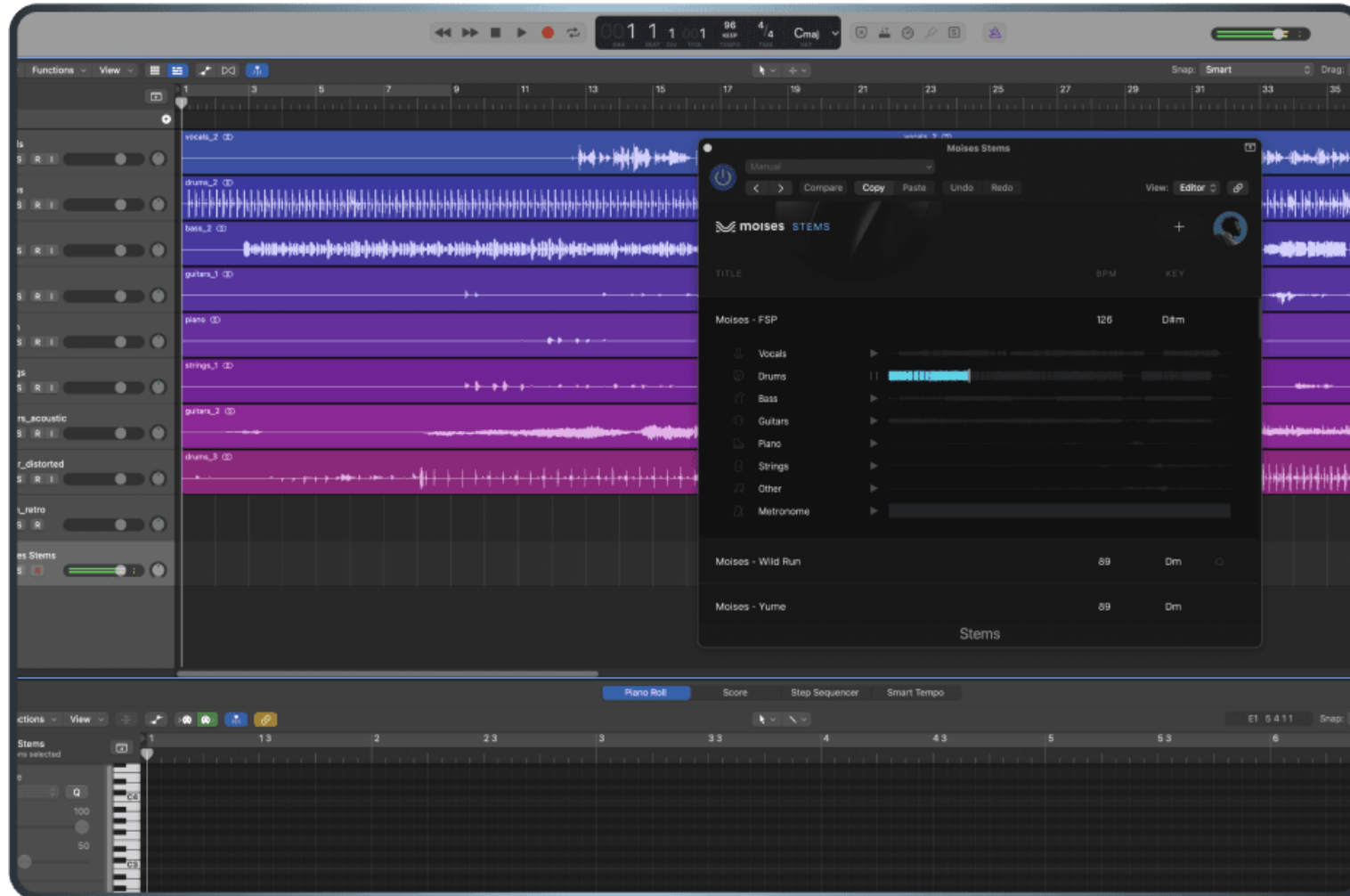
youtu.be/cyXPgU5UiB8

Moises VST Plugin



(Source: Moises)

Moises VST Plugin



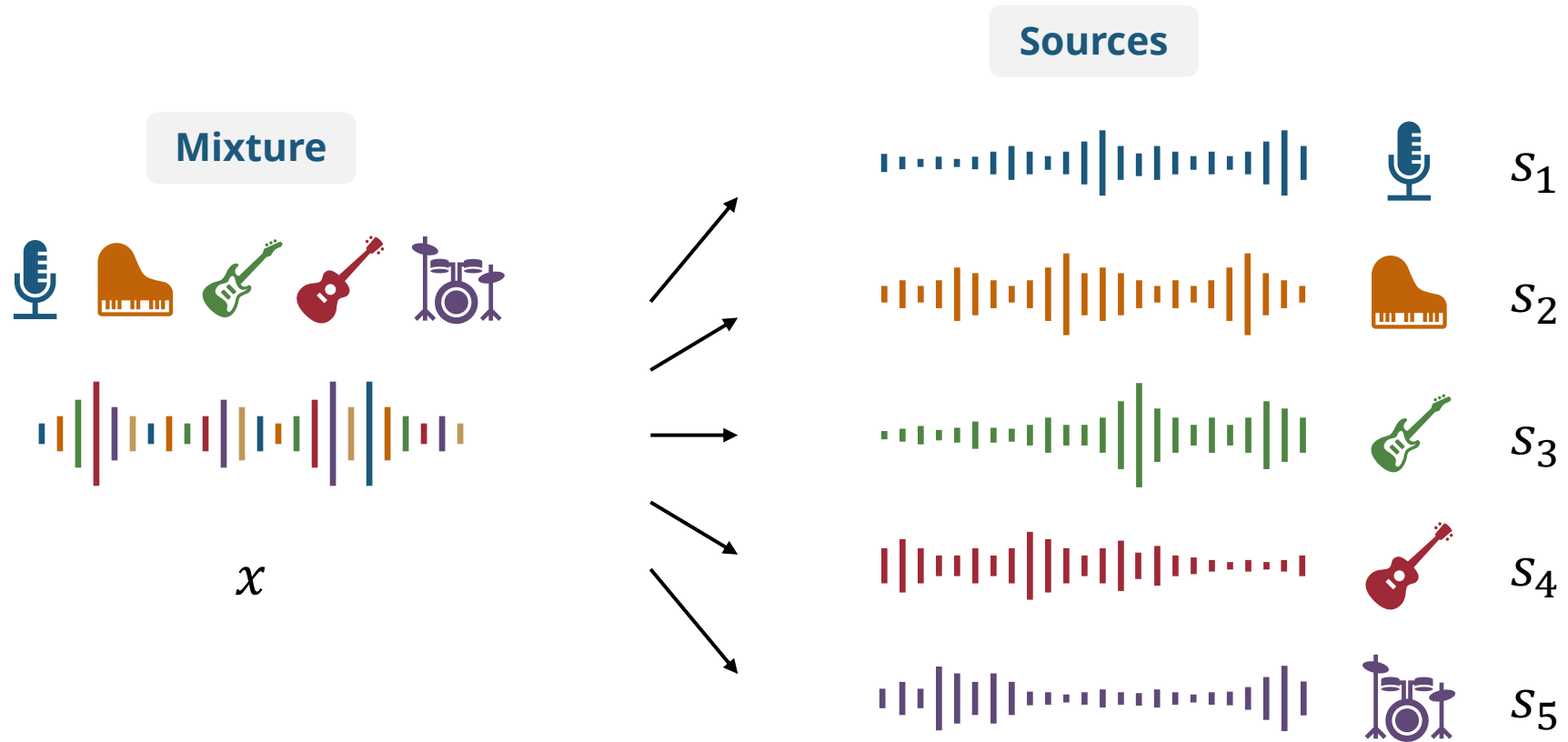
(Source: Moises)

How to Produce Music with Moises

- Part 1: “This is the way that a lot of tracks start”
- Part 2: Creating backing vocals from scratch
- Part 3: A new way to mix drums

How does it work?

Mathematical Formulation

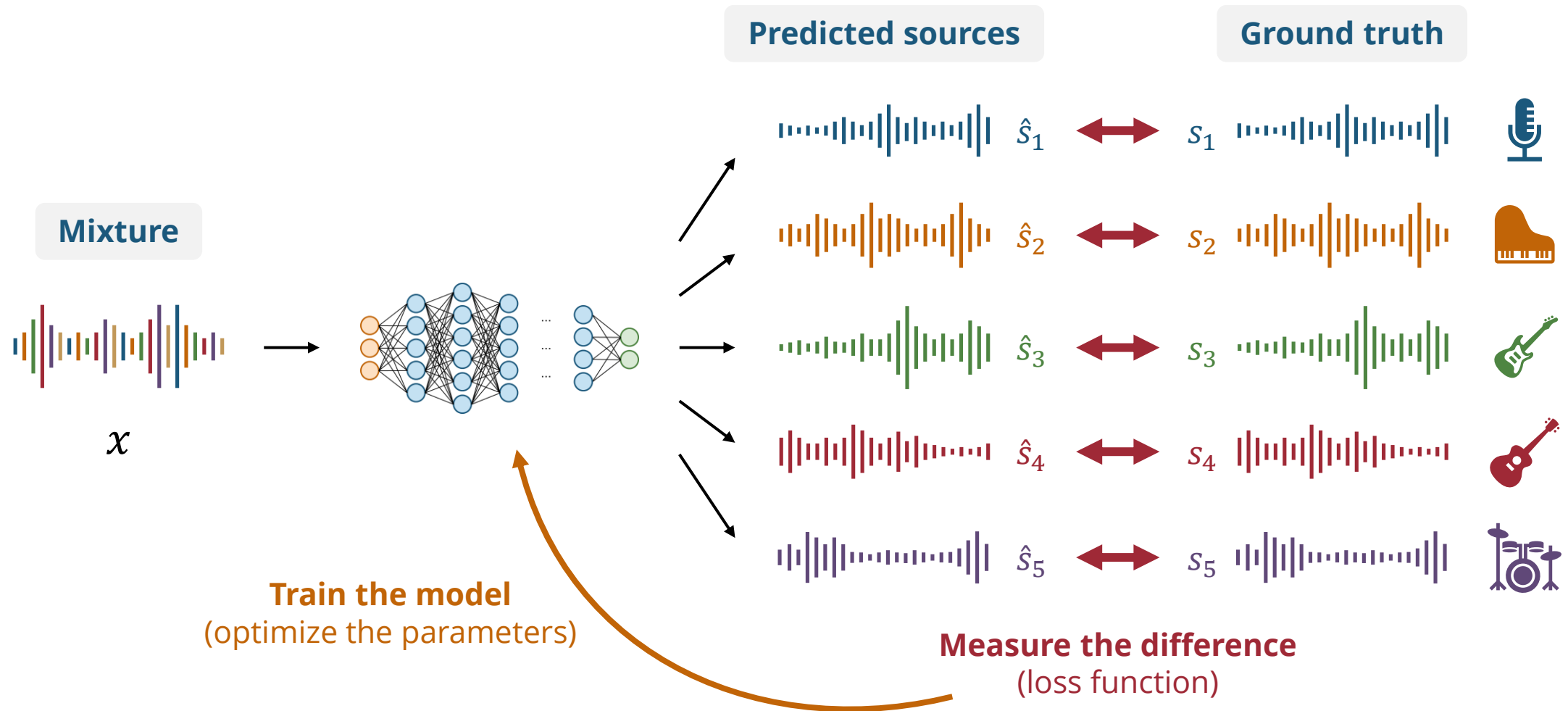


$$x = s_1 + s_2 + s_3 + s_4 + s_5 = \sum_i s_i$$

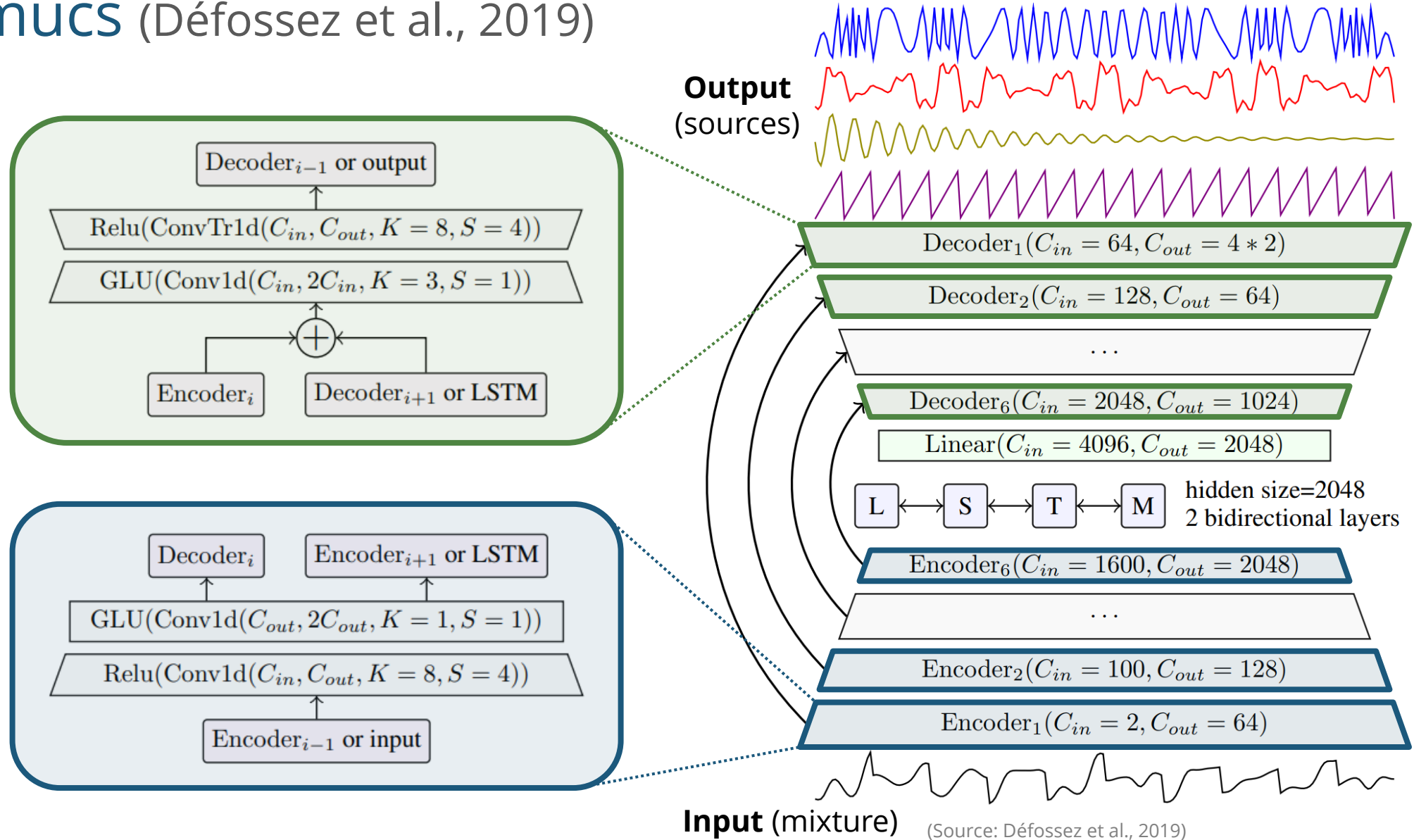
Source Separation is an Ill-posed Problem

- There are **more than one solution** to $x = s_1 + s_2 + \dots + s_N$
 - In fact, there are infinite possibilities
- However, we do know **what's more likely than another!**

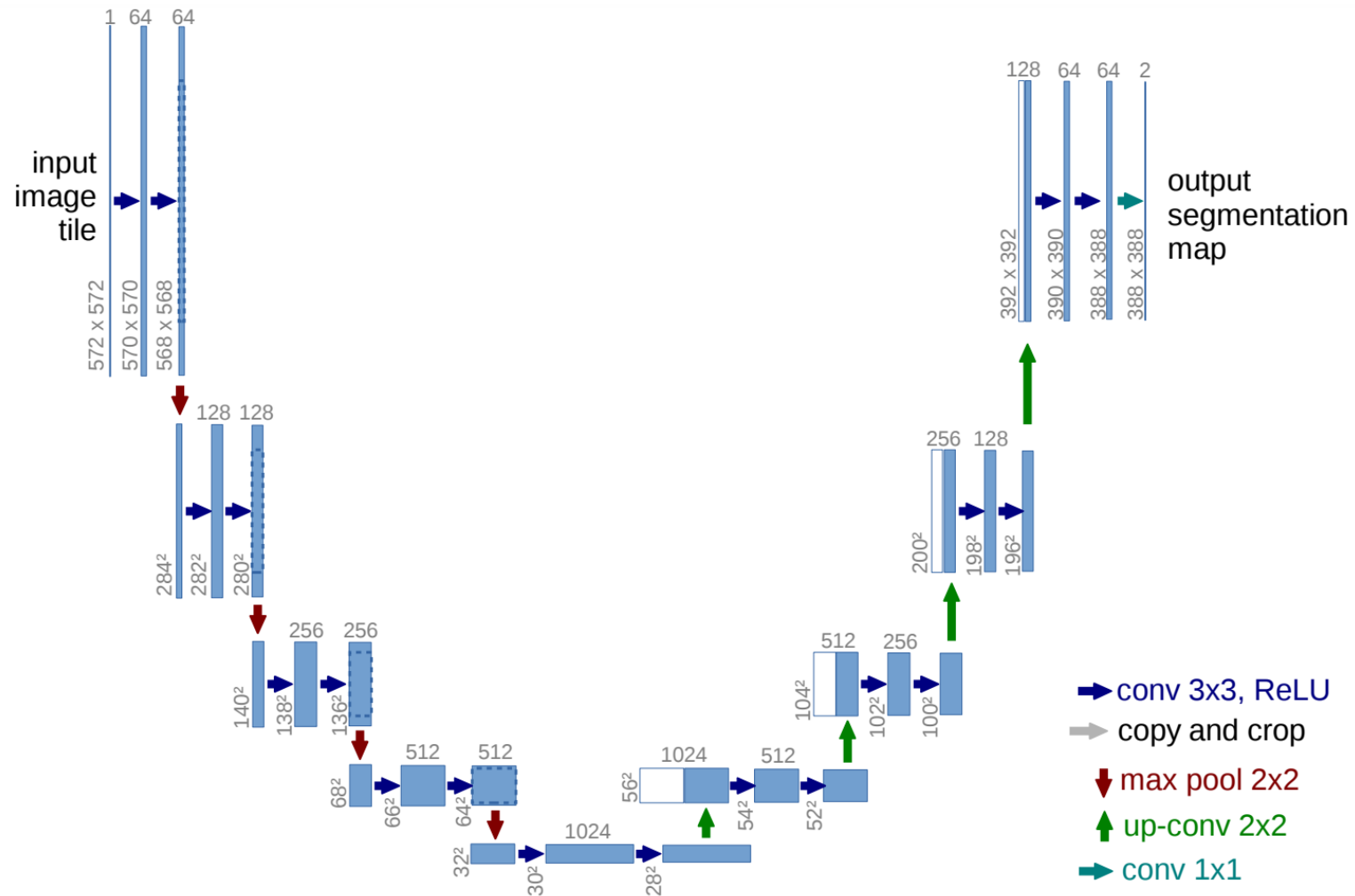
Deep Learning Based Source Separation



Demucs (Défossez et al., 2019)

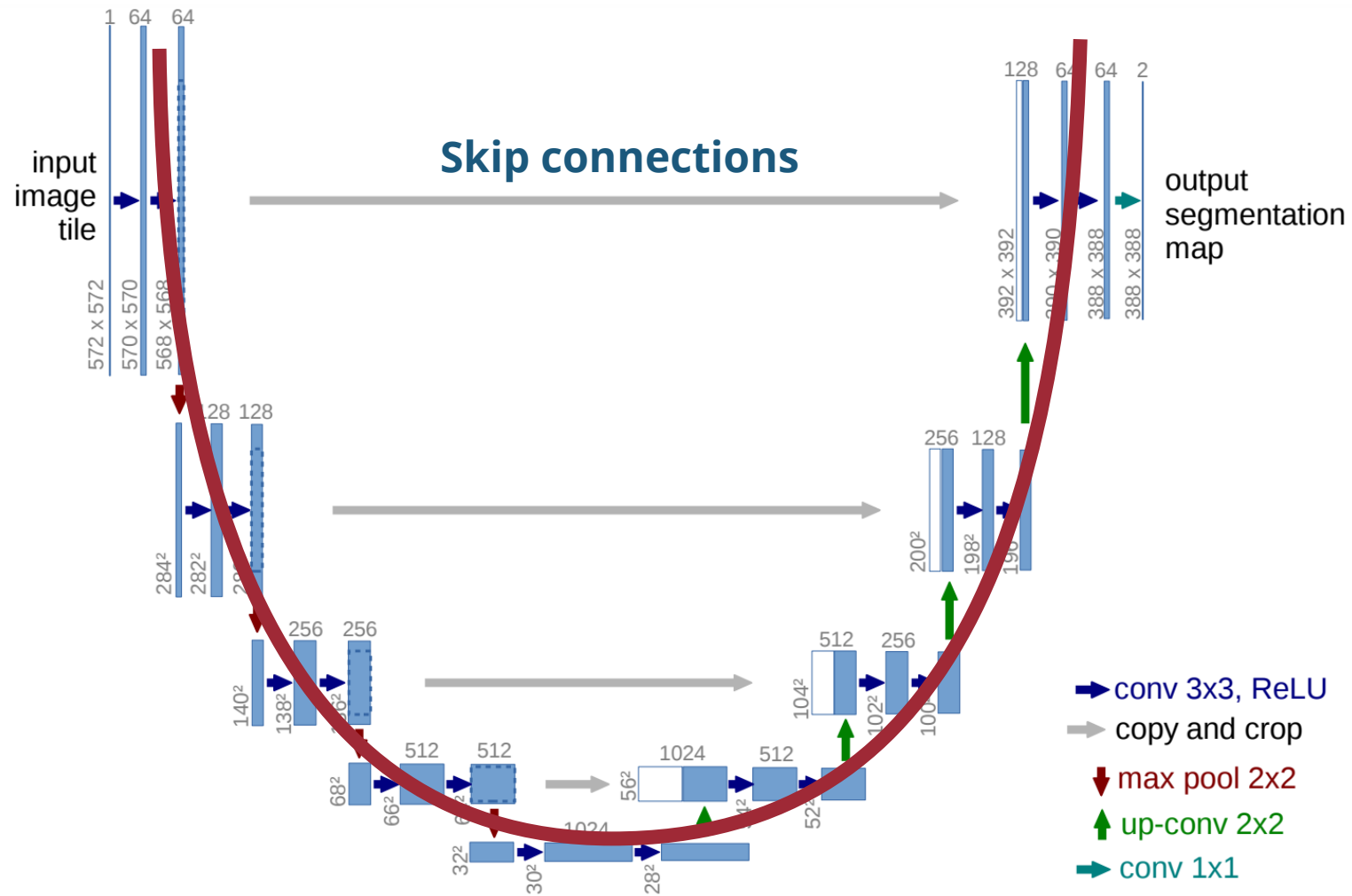


U-Net (Ronneberger et al., 2015)



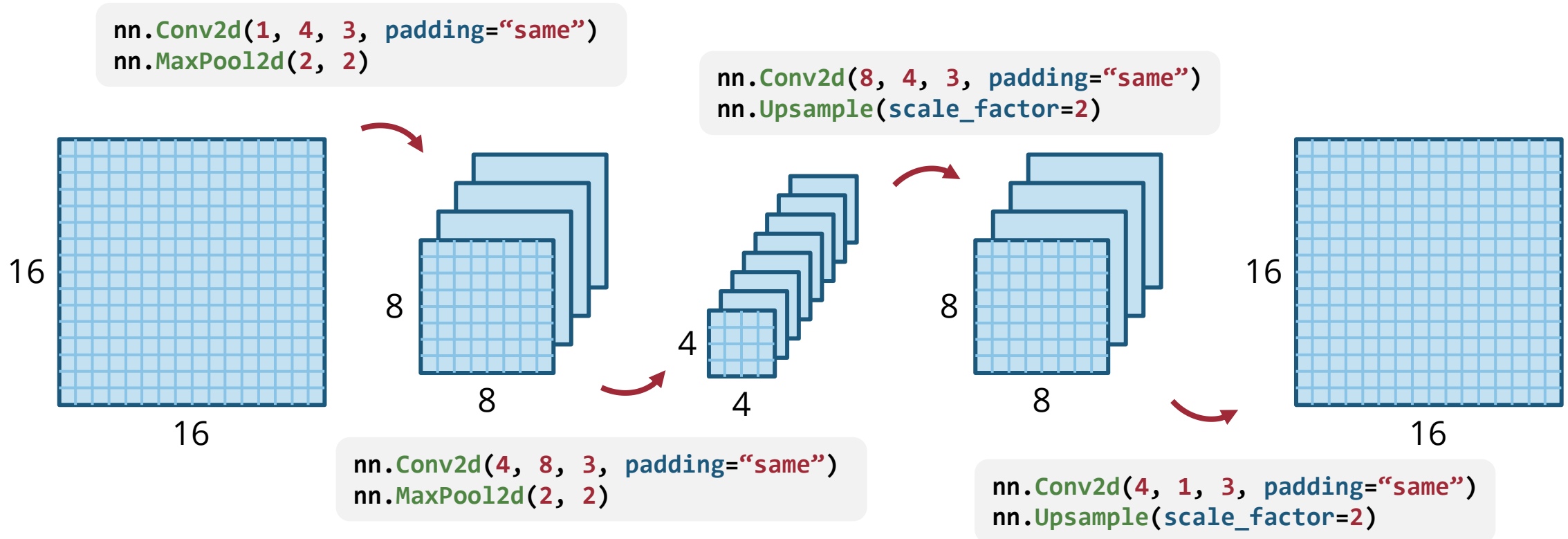
(Source: Ronneberger et al., 2015)

U-Net (Ronneberger et al., 2015)

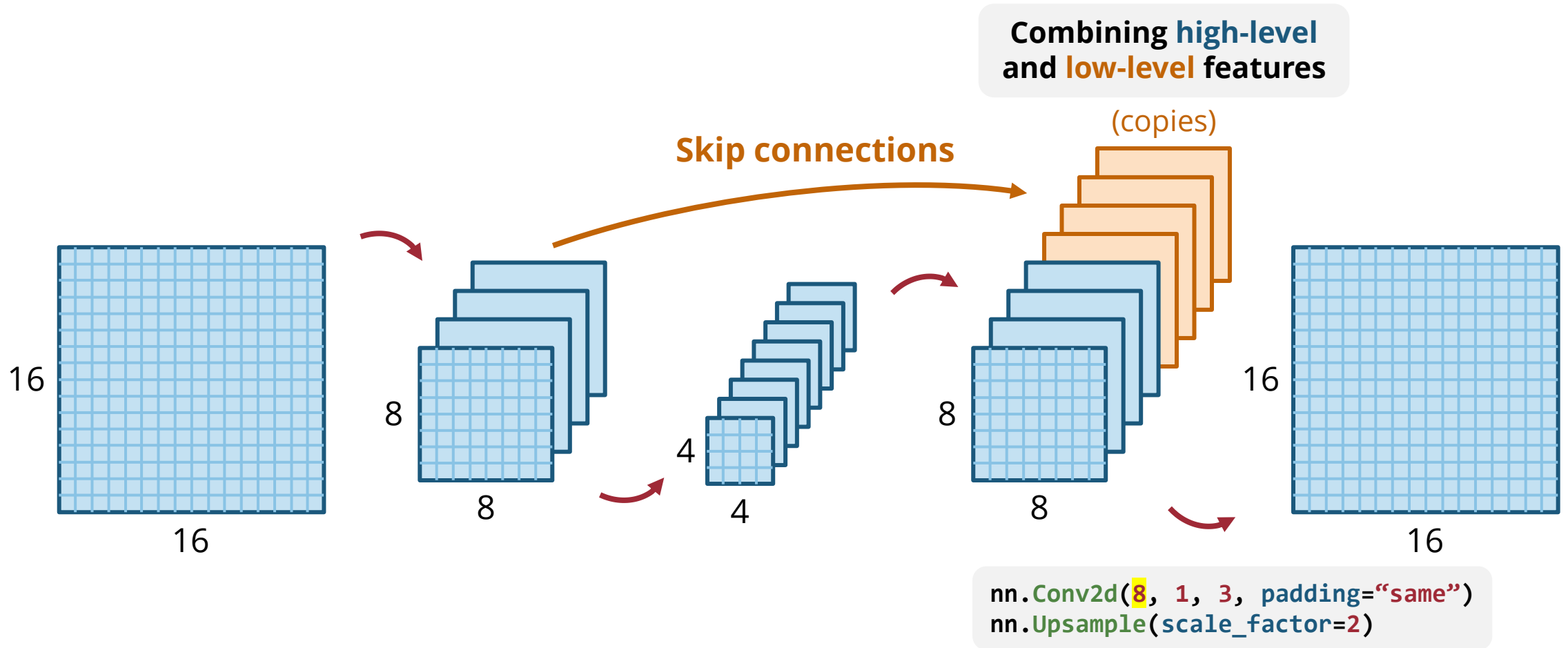


(Source: Ronneberger et al., 2015)

A Toy Example of U-Net

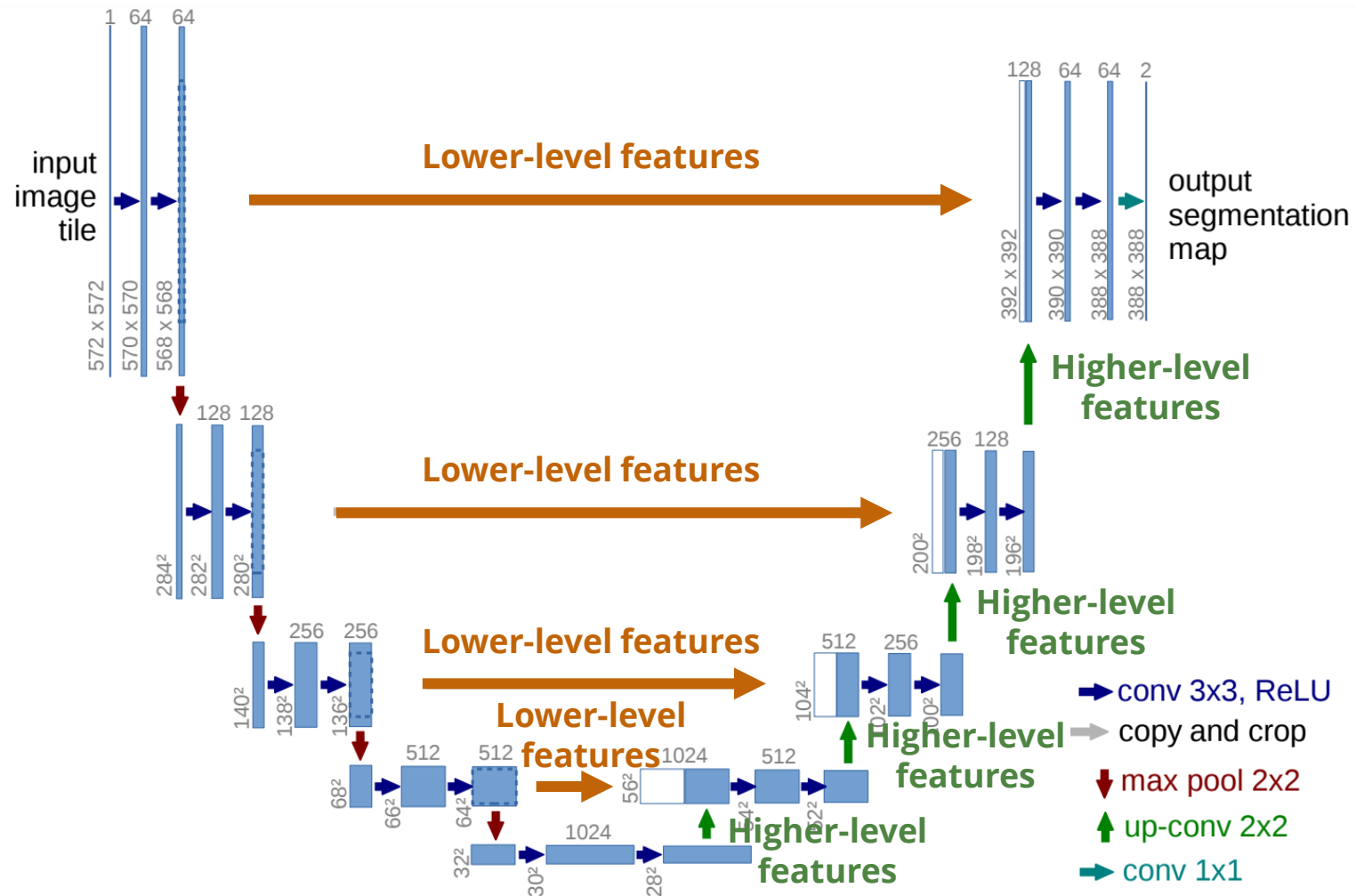


A Toy Example of U-Net



U-Nets are useful when the inputs and outputs have the same shape!

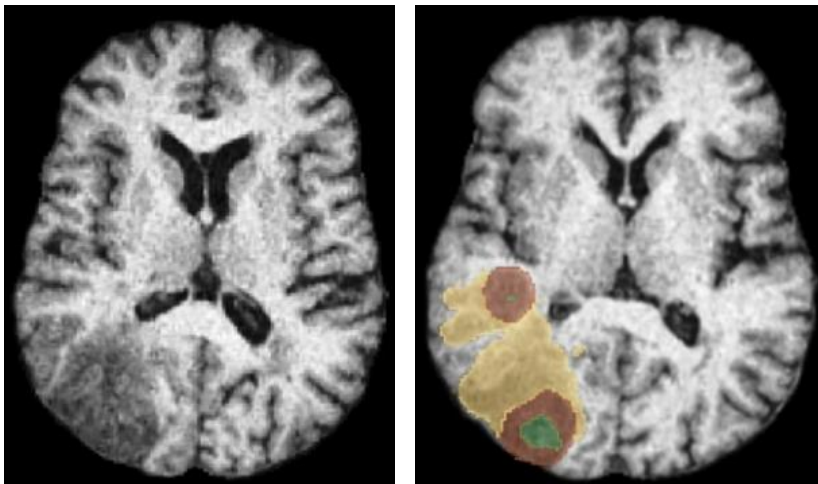
U-Net (Ronneberger et al., 2015)



(Source: Ronneberger et al., 2015)

Applications of U-Nets

Tumor Segmentation



(Source: Kharaji et al., 2024)

Depth Estimation



(Source: Barakat, 2018)

Image Segmentation



(Source: Kirillov et al., 2023)

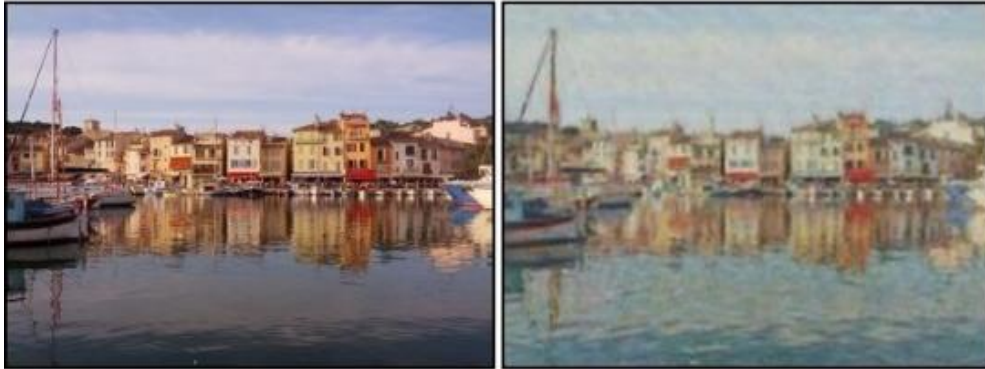
Omar Barakat, "[Depth estimation with deep Neural networks part 1](#)," *Medium*, January 11, 2018

Mona Kharaji, Hossein Abbasi, Yasin Orouskhani, Mostafa Shomalzadeh, Foad Kazemi, and Maysam Orouskhani, "[nnU-Net for Brain Tumor Segmentation](#)," *Neuroscience Informatics*, 2024.

Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C. Berg, Wan-Yen Lo, Piotr Dollár, and Ross Girshick, "[Segment Anything](#)," *ICCV*, 2023.

Applications of U-Nets

Style Transfer



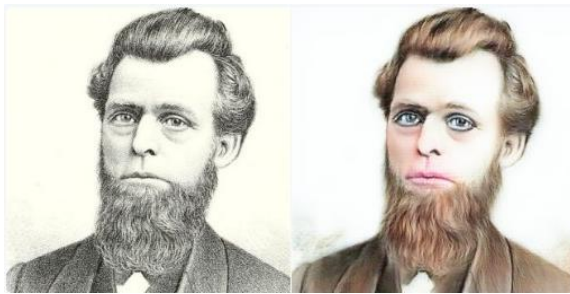
(Source: Zhu et al., 2018)

Sim2Real



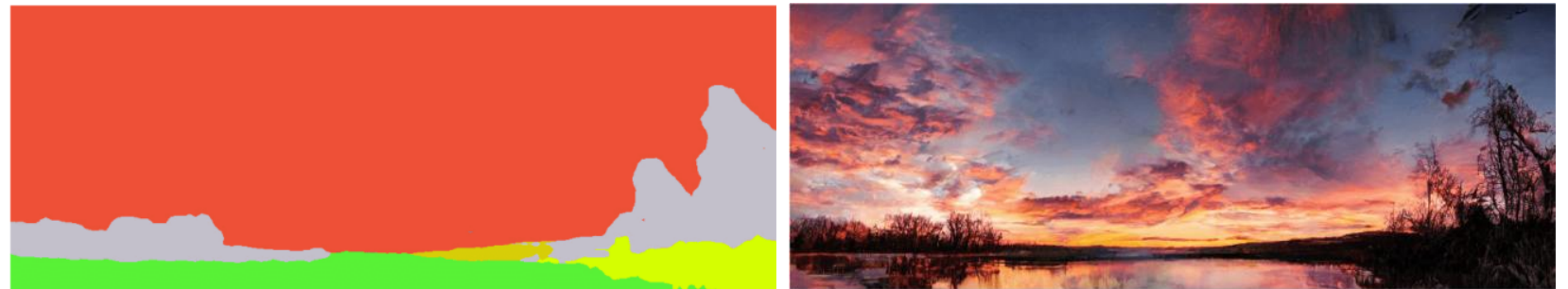
(Source: Zhu et al., 2018)

Colorization



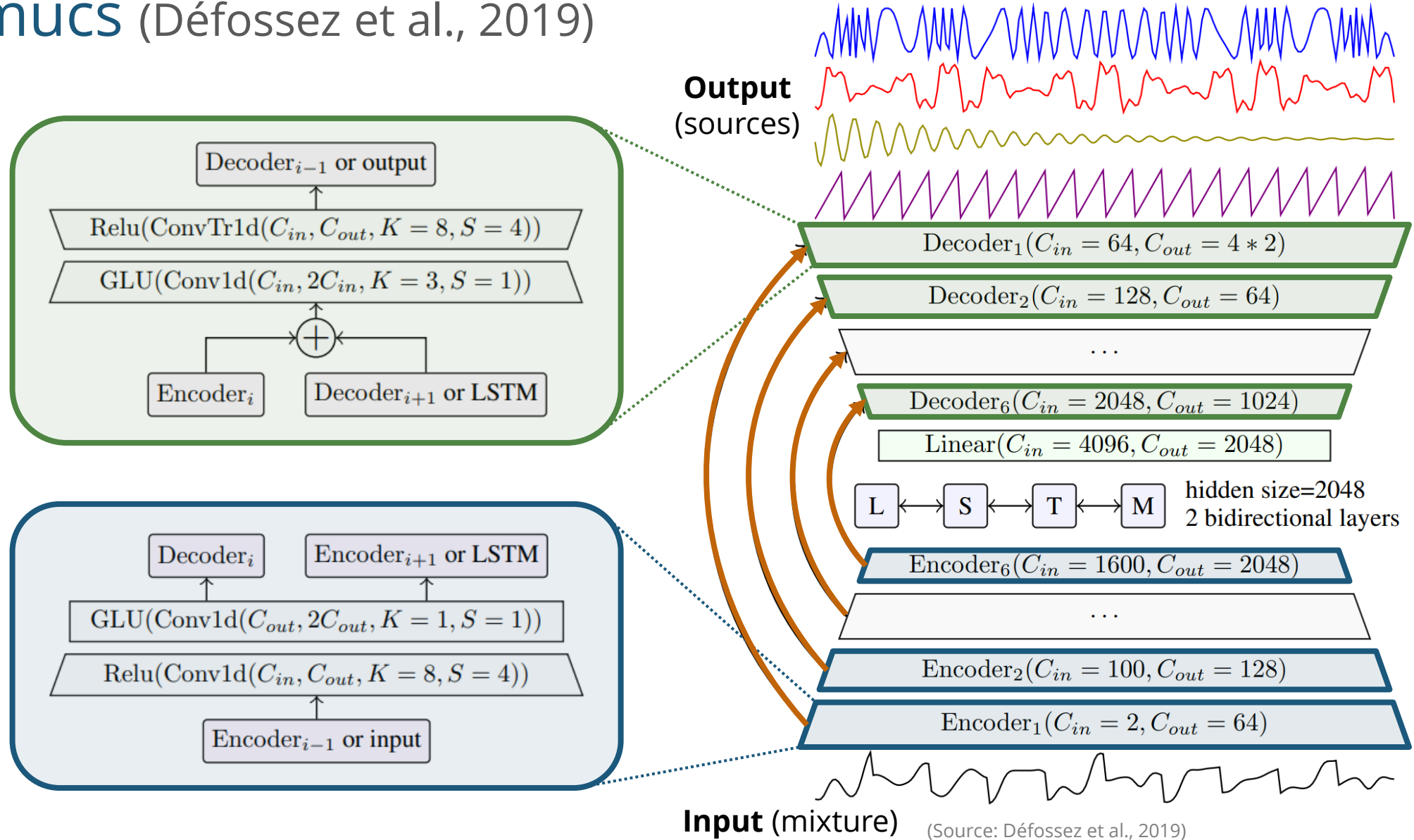
(Source: Zhu et al., 2018)

Semantic Synthesis

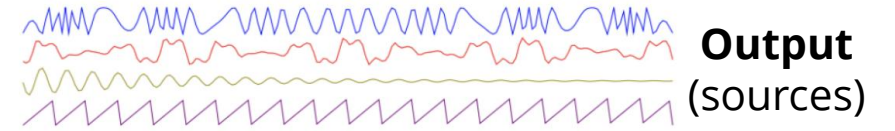


(Source: Rombach et al., 2022)

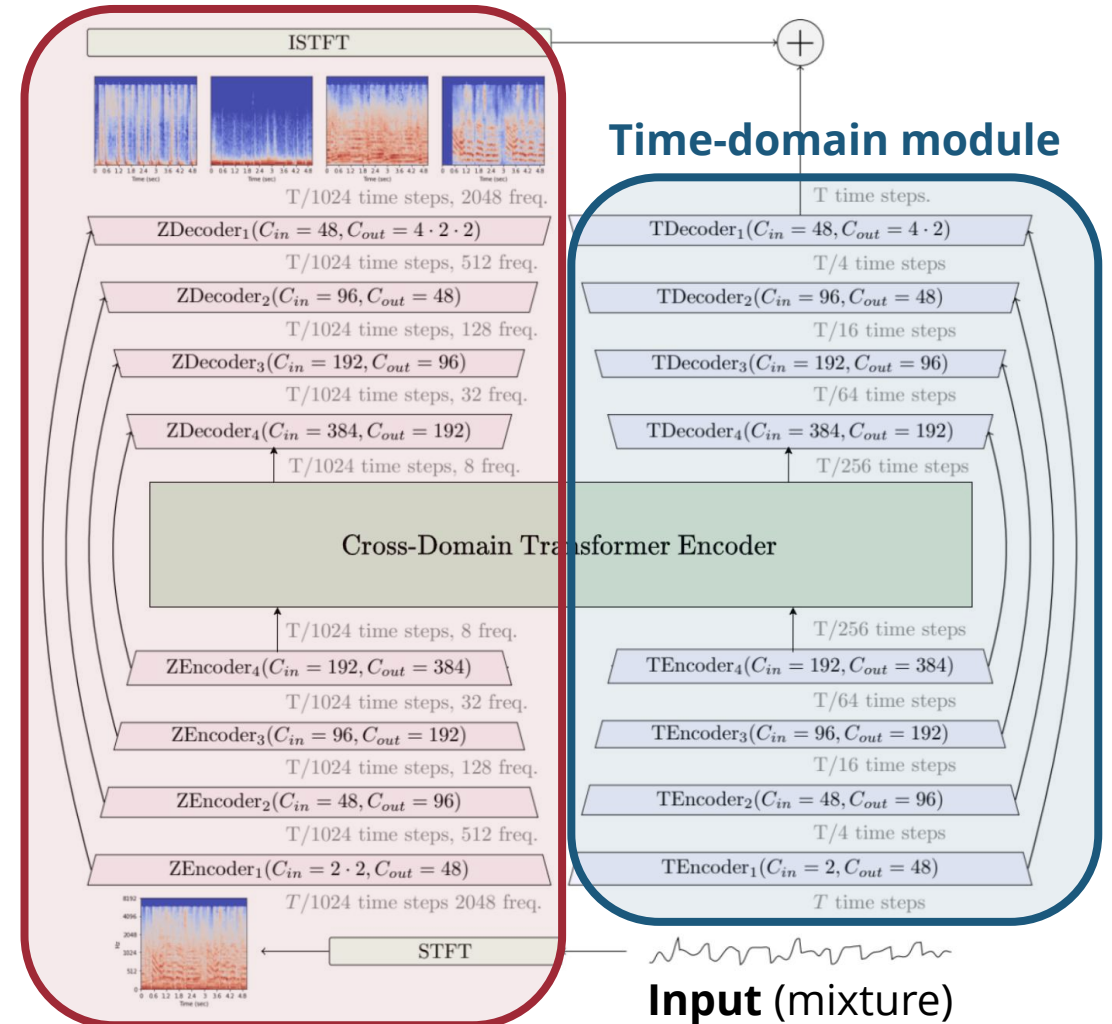
Demucs (Défossez et al., 2019)



Demucs-Hybrid (Rouard et al., 2023)



Frequency-domain module



(Source: Rouard et al., 2023)

Datasets

- [MIR-1K](#)
- [MedleyDB](#)
- [iKala](#)
- [DSD100](#)
- [MUSDB18](#) & [MUSDB18-HQ](#)
- [MoisesDB](#)
- Synthetic: [Slakh2100](#), [SynthSOD](#)

Datasets

| Dataset | Year | Tracks | Track duration (s) | Full/stereo? |
|--|------|--------|--------------------|--------------|
| MASS | 2008 | 9 | 16 ± 7 | ✗ / ✓ |
| MIR-1K | 2010 | 1,000 | (8 ± 8) | ✗ / ✗ |
| QUASI | 2011 | 5 | (206 ± 21) | ✓ / ✓ |
| ccMixer | 2014 | 50 | (231 ± 77) | ✓ / ✓ |
| MedleyDB | 2014 | 63 | (206 ± 121) | ✓ / ✓ |
| iKala | 2015 | 206 | 30 | ✗ / ✗ |
|  DSD100 | 2015 | 100 | (251 ± 60) | ✓ / ✓ |
|  MUSDB18 | 2017 | 150 | (236 ± 95) | ✓ / ✓ |
|  MUSDB18-HQ | 2019 | 150 | (236 ± 95) | ✓ / ✓ |

(Source: SigSep)

Choral Separation (Chen et al., 2022)

Demo

Mixture



Soprano



Alto



Tenor



Bass



Data Augmentation

SoundFont



Standard



Expressive
(vowels only)



Expressive
(words)



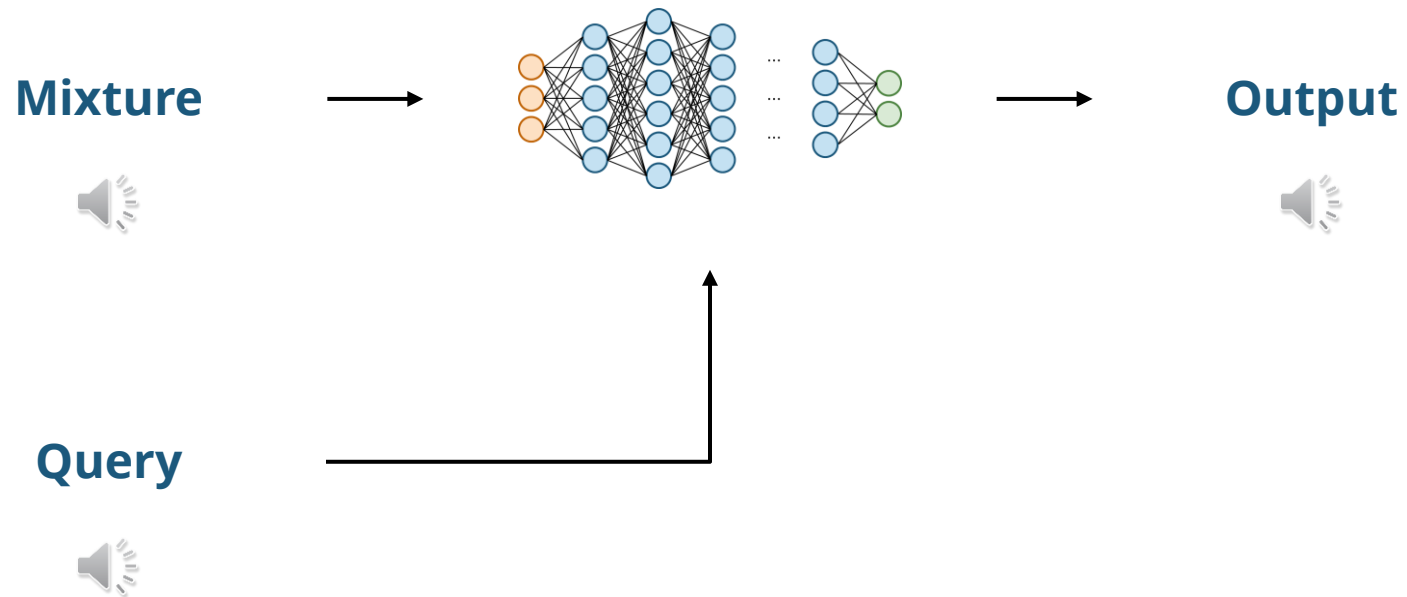
Blob Opera

- This is **NOT** based on source separation
- Sharing this simply **because it's cool!** 😎
- It's based on a **ML-based music harmonization model!**



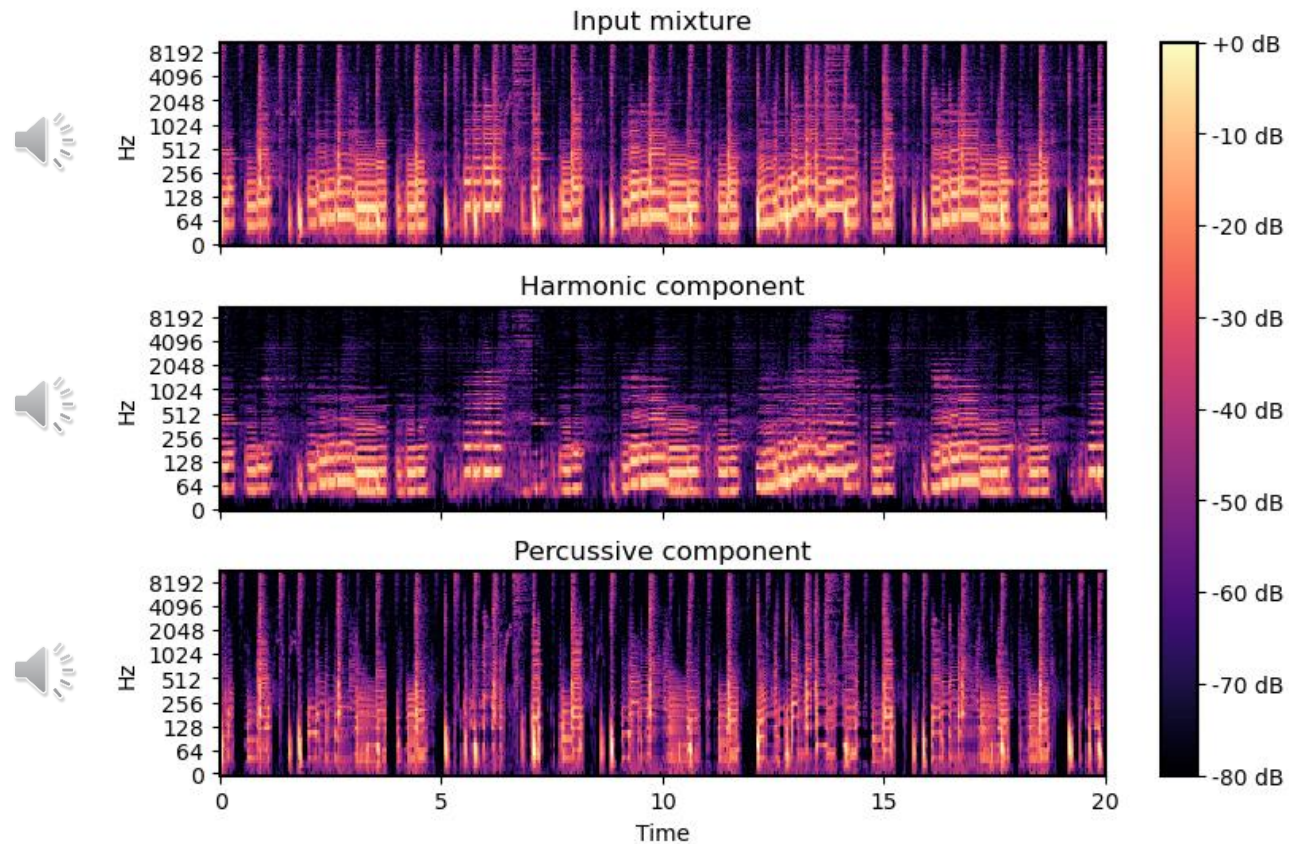
artsandculture.google.com/experiment/blob-opera/AAHWrq360NcGbw

Beyond Known Sources: Query by Audio



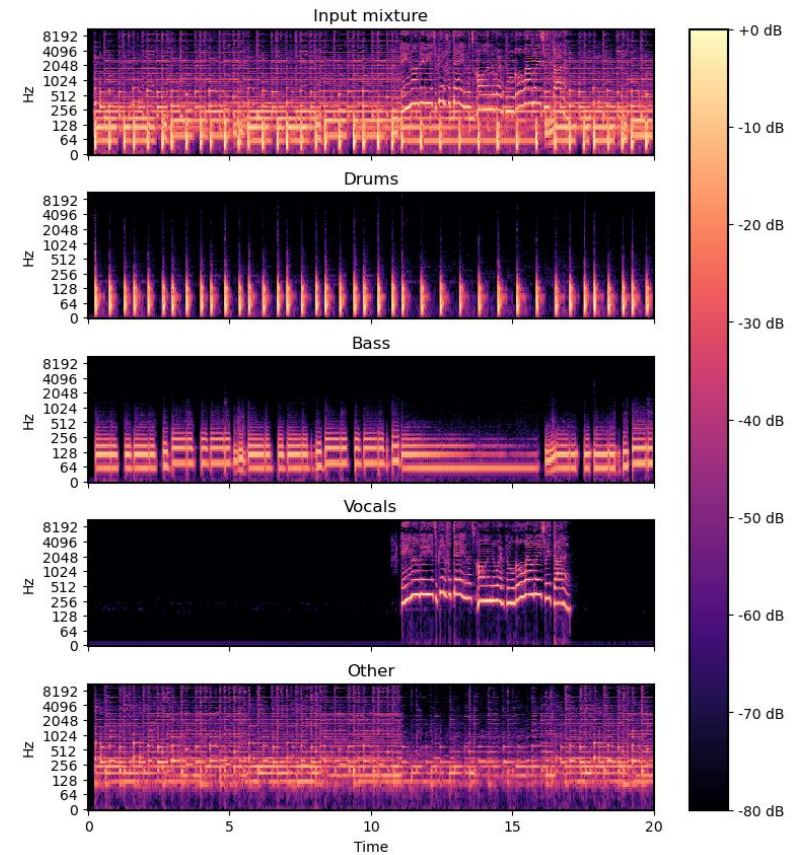
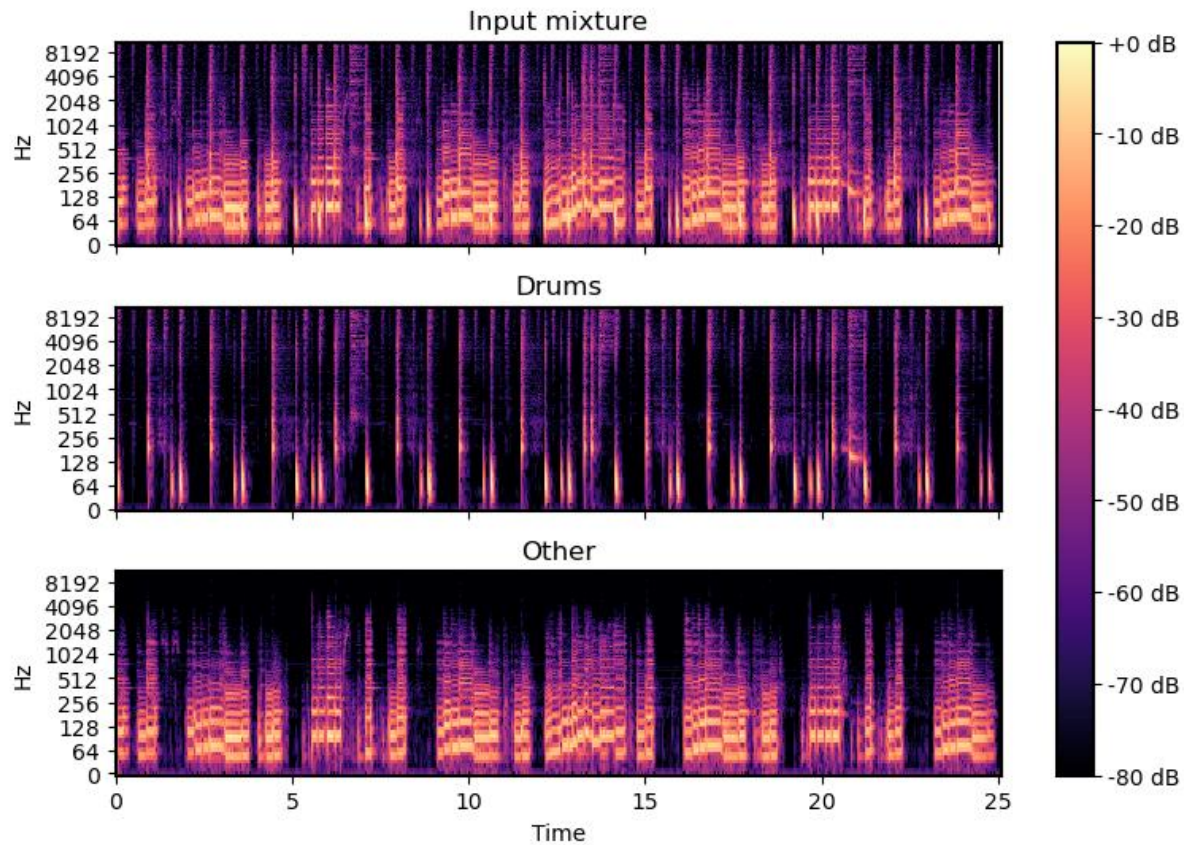
Homework 4: Source Separation

- **Part 1:** Harmonic-Percussive Source Separation (HPSS) using **librosa**



Homework 4: Source Separation

- **Part 2:** Music Source Separation using **Demucs**



Homework 4: Source Separation

- Instructions will be released on the [course website](#)
- Please submit your work to [Gradescope](#)
- Due at **11:59pm ET** on **February 26**
- Late submissions: **1 point deducted per day**
- No late submission is allowed a week after the due date

Optional Reading

- Ethan Manilow, Prem Seetharman, and Justin Salamon, "[Open Source Tools & Data for Music Source Separation](#)," *Tutorials of ISMIR*, 2020.