

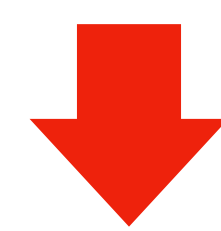
INTRODUCTION

Why music source separation

- Separated tracks are easy for transcription
- Can be used for audio remixing
- Instrument-wise equalization

Drawbacks of existing systems

1. Exclusively focus on source separation task itself
2. Limited types of separable instruments and limited training data

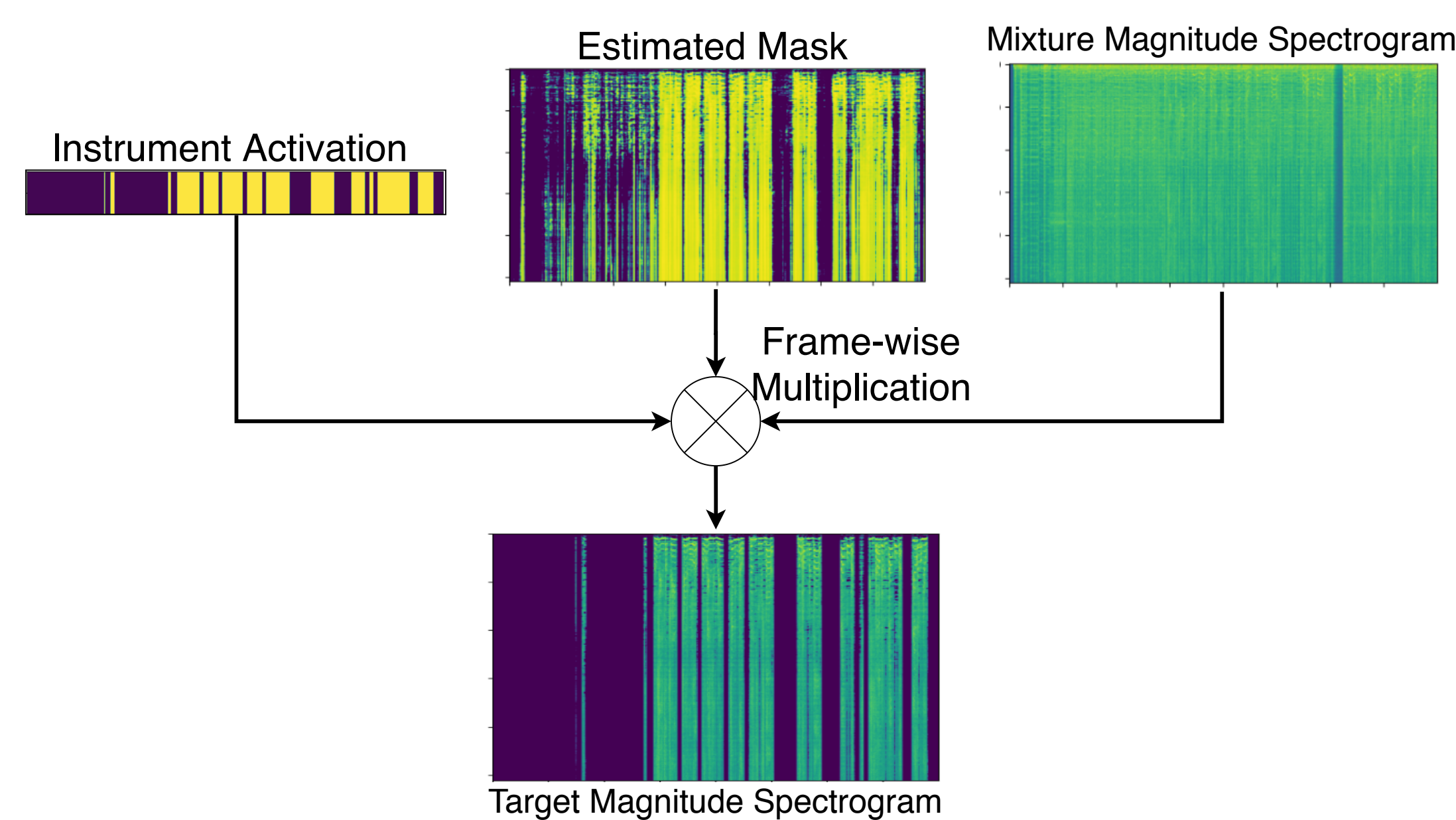


How we solve the first point

SYSTEM

We propose a multitask learning model to jointly estimate separated spectrogram and instrument activation

- We use U-net structure since it has been found useful in image segmentation and singing voice separation
- We use residual blocks to deepen the structure



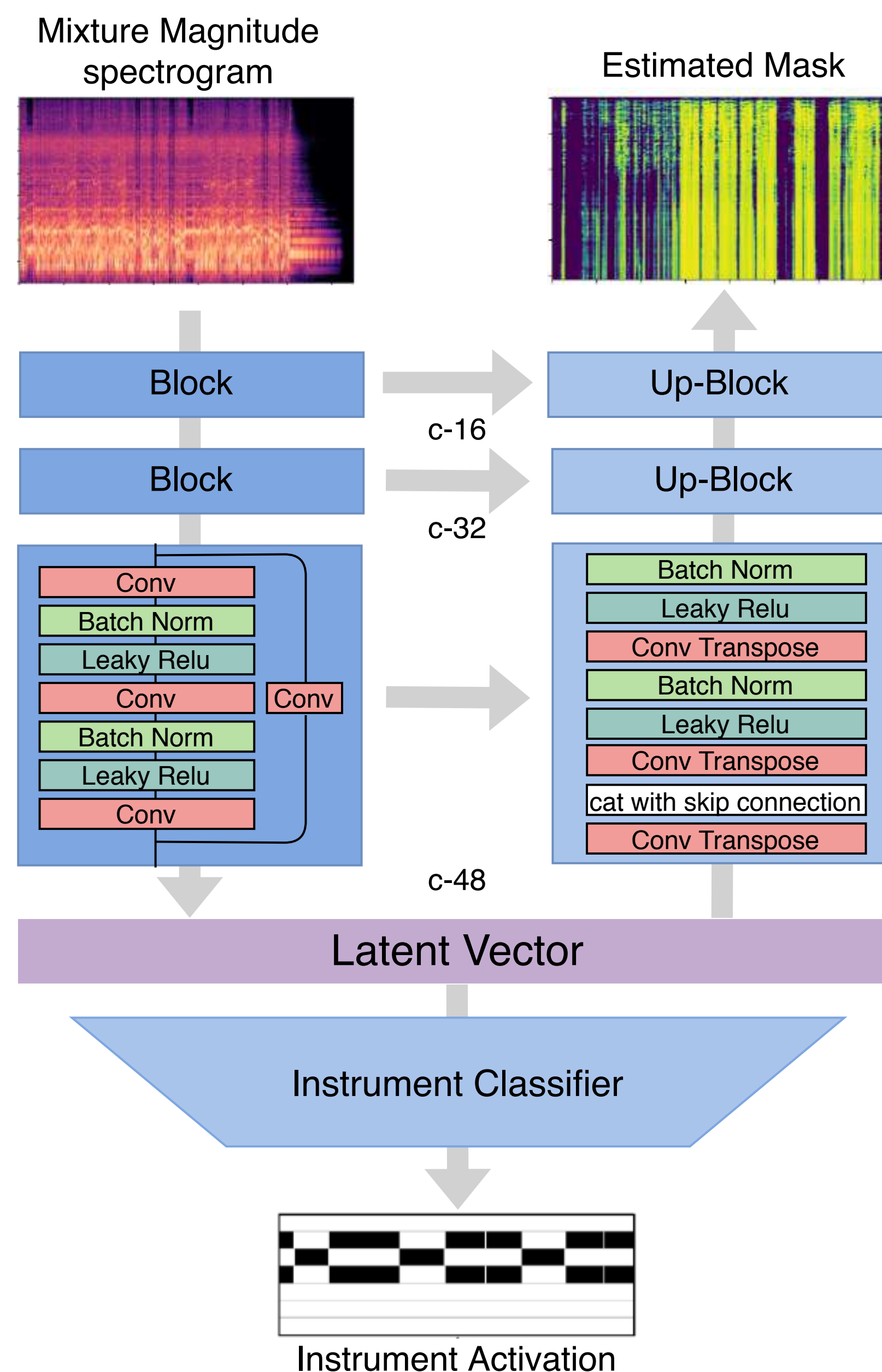
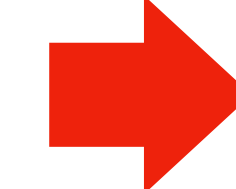
- During inference, instrument activation acts as a weight to filter estimated mask
- Median filter is used to smooth the estimated activation

DATASET

We use MedleyDB [1] + Mixing Secret [2] (MM dataset) to train and evaluate on more instrument categories

- 100+ instruments
- 400+ songs
- Provide baseline result on more instrument categories
- Select 6 most popular instruments for experiment

How we solve the second point



[1] Bittner, Rachel M., et al. "Medleydb: A multitrack dataset for annotation-intensive mir research." *ISMIR*. Vol. 14. 2014.
 [2] Gururani, Siddharth, and Alexander Lerch. "Mixing secrets: a multi-track dataset for instrument recognition in polyphonic music." *Proc. ISMIR-LBD*. 2017.
 [3] Stöter, Fabian-Robert, et al. "Open-unmix-a reference implementation for music source separation." (2019).

RESULT

We compared our model with Open-Unmix [3]

- Our approach performs equally on MUSDB-HQ dataset
- Our approach out-performs Open-Unmix on MM dataset

Training and testing on MUSDB-HQ (SDR)

	Method	SDR	SIR	SAR	ISR
Vocals	Open-Unmix	6.11	13.21	6.75	12.43
	IASS	6.46	14.70	6.98	14.30
Bass	Open-Unmix	4.48	8.23	5.40	10.29
	IASS	4.18	7.30	4.52	6.85
Drums	Open-Unmix	5.02	10.17	6.05	10.55
	IASS	5.56	10.74	6.86	10.92
Other	Open-Unmix	4.23	9.90	3.88	7.34
	IASS	4.19	8.78	4.70	9.32

Training and testing on MM dataset (SDR)

	Open-Unmix	IASS	IBM	input-SDR
Vocals	3.68	4.78	6.49	-6.24
Elecgtr	1.55	1.77	4.56	-5.90
Acgtr	0.95	1.29	3.38	-6.65
Piano	1.08	1.91	3.63	-6.31
Bass	4.04	5.26	5.34	-5.77
Drums	4.45	4.89	6.23	-6.05

*IBM: ideal binary mask

FUTURE WORK

- Combine synthesized datasets to increase training data and instrument categories
- Incorporate other tasks such as multi-pitch estimation
- Explore post-processing method for both instrument detection and source separation

Listening samples!

