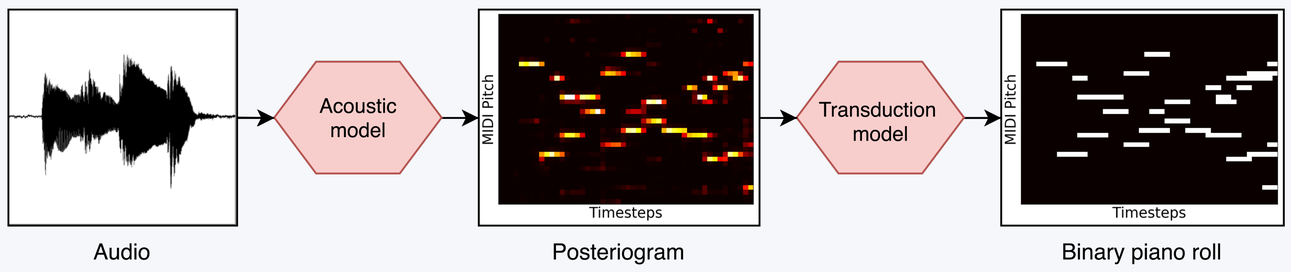
A Comparative Study of Neural Models for Polyphonic Music Sequence Transduction

Adrien Ycart, **Daniel Stoller**, Emmanouil Benetos

Centre for Digital Music, Queen Mary University of London

{a.ycart,d.stoller,emmanouil.benetos}@qmul.ac.uk

Problem statement



Audio — Posteriogram — Binary piano roll

We **compare** various **neural-network** approaches to learn a mapping from **posteriogram to piano-roll**

Dataset

- MAPS dataset: classical piano music
- Split: trained on synthetic pianos, tested on real pianos
- Inputs downsampled to 16th note timesteps using A-MAPS annotations

Comparison: 16 configurations

Acoustic Model

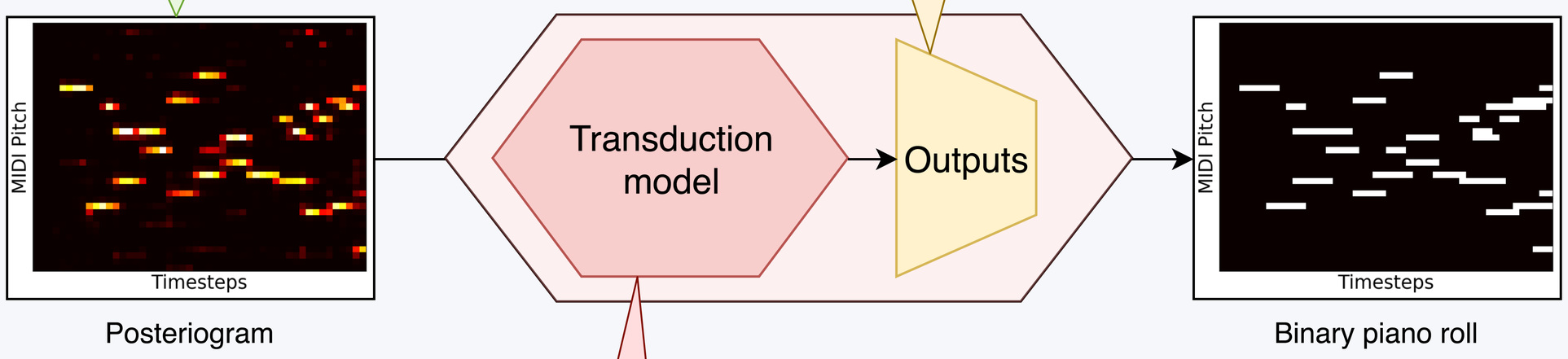
- *Kelz et al. (2016):*
  - Piano-specific CNN
- *Bittner et al. (2017):*
  - General-purpose multi-pitch detection system

Outputs

- Sigmoid outputs + threshold post-processing
  - Threshold tuned on validation dataset
- *Dong et al. (2018)*: Binary neurons
  - Forward: step function, backward: sigmoid
  - Good results for music generation with GAN

Training loss

- Frame-based:
  - Sigmoid outputs: Cross-entropy
  - Binary outputs: F-measure
- Adversarial: Wasserstein GAN
  - Conditional discriminator
  - Architecture inspired by DCGAN: (3x3 convolutions, stride 2x2)*4 + 2 dense layers

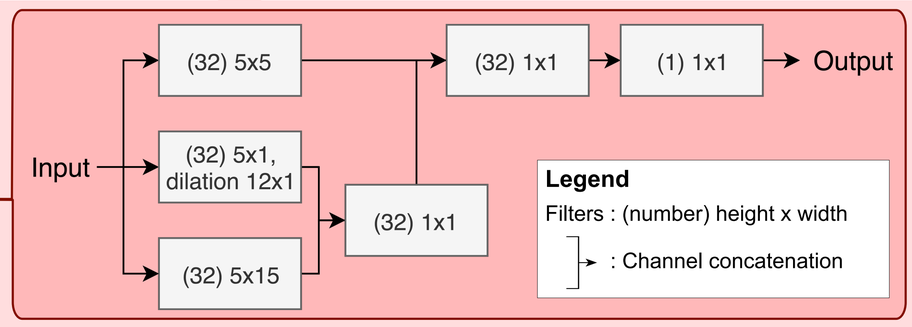


Posteriogram — Binary piano roll

Transduction Model

- *Ycart et al. (2018):*
  - Single-layer LSTM, 100 hidden nodes
- Newly-proposed CNN

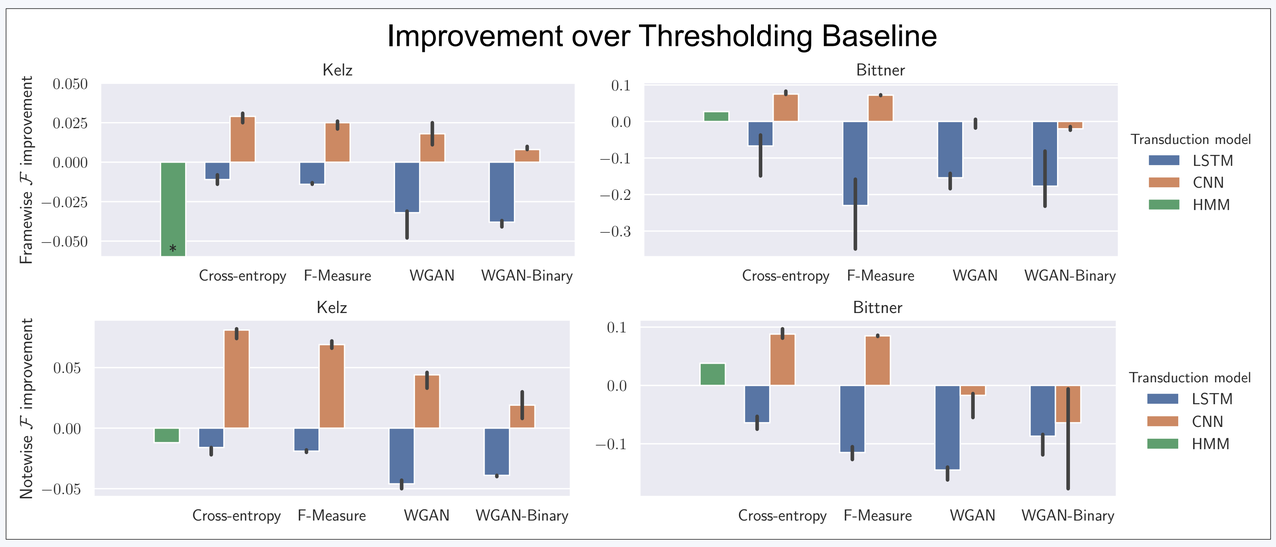
→ **Same number of parameters: ~80 000**



Baselines

- Thresholding posteriogram (threshold tuned on validation set)
- *Poliner et al. (2016) :* Pitchwise on/off HMM decoding

Results



Kelz

Metric		Thresh	HMM	Cross-entropy		F-measure		WGAN		WGAN-Binary	
				LSTM	CNN	LSTM	CNN	LSTM	CNN	LSTM	CNN
Frame	$\mathcal{F}$	67.9	49.6	66.8	**70.8**	66.5	70.4	64.7	69.7	64.1	68.7
	$\mathcal{P}$	70.9	**74.1**	72.6	73.4	70.2	72.2	72.5	**74.1**	74.4	73.9
	$\mathcal{R}$	66.7	40.1	63.2	69.6	64.6	**70.1**	59.8	67.2	57.8	65.5
Note	$\mathcal{F}$	45.0	43.8	43.4	**53.2**	43.1	51.6	40.4	49.4	41.1	46.9
	$\mathcal{P}$	44.0	**82.4**	42.8	50.9	39.3	50.7	39.5	50.1	40.5	44.6
	$\mathcal{R}$	47.5	31.3	45.6	**57.1**	49.4	53.9	43.0	50.0	43.6	51.0

Bittner

Metric		Thresh	HMM	Cross-entropy		F-measure		WGAN		WGAN-Binary	
				LSTM	CNN	LSTM	CNN	LSTM	CNN	LSTM	CNN
Frame	$\mathcal{F}$	58.8	61.5	52.1	**66.3**	35.8	66.0	43.4	58.8	41.1	56.8
	$\mathcal{P}$	59.6	52.6	48.0	68.5	26.9	**69.3**	43.9	58.7	35.9	61.9
	$\mathcal{R}$	61.5	**79.6**	60.5	66.1	65.4	65.3	47.7	60.9	50.8	56.4
Note	$\mathcal{F}$	44.6	48.4	39.3	**53.4**	31.9	53.1	30.1	43.2	36.2	44.0
	$\mathcal{P}$	42.2	**62.5**	35.7	49.3	25.2	50.7	27.6	40.8	34.4	44.8
	$\mathcal{R}$	48.6	40.4	45.1	**59.9**	45.6	57.3	34.5	47.2	39.2	44.5

Main conclusions

- **Overall best**: CNN, Sigmoid outputs, Cross-entropy loss
- **LSTM strongly overfits** on specific pianos in training set
- **Cross-entropy is better** than GAN and F-measure as loss
- **Binary neurons do not help** (neither with GAN nor F-measure loss)

R. Kelz, M. Dorfer, F. Korzeniowski, S. Bock, A. Arzt, and G. Widmer, "On the Potential of Simple Framewise Approaches to Piano Transcription," *17th International Conference on Music Information Retrieval (ISMIR)*, 2016.

R. M. Bittner, B. McFee, J. Salamon, P. Li, and J. P. Bello, "Deep Salience Representations for F0 Estimation in Polyphonic Music," *18th International Conference on Music Information Retrieval (ISMIR)*, 2017.

H.-W. Dong, Y.-H. Yang, "Convolutional Generative Adversarial Networks with Binary Neurons for Polyphonic Music Generation," *19th International Conference on Music Information Retrieval (ISMIR)*, 2018.

A. Ycart, E. Benetos, "Polyphonic Music Sequence Transduction with Meter-Constrained LSTM Networks," *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2018

G.E. Poliner, D.P.W Ellis, "A Discriminative Model for Polyphonic Piano Transcription," *EURASIP Journal on Advances in Signal Processing*, 2006

A. Ycart, E. Benetos, "A-MAPS: Augmented MAPS Dataset with Rhythm and Key Annotations," " *19th International Conference on Music Information Retrieval Late Breaking and Demos Papers*, 2018.

Queen Mary University of London

centre for digital music