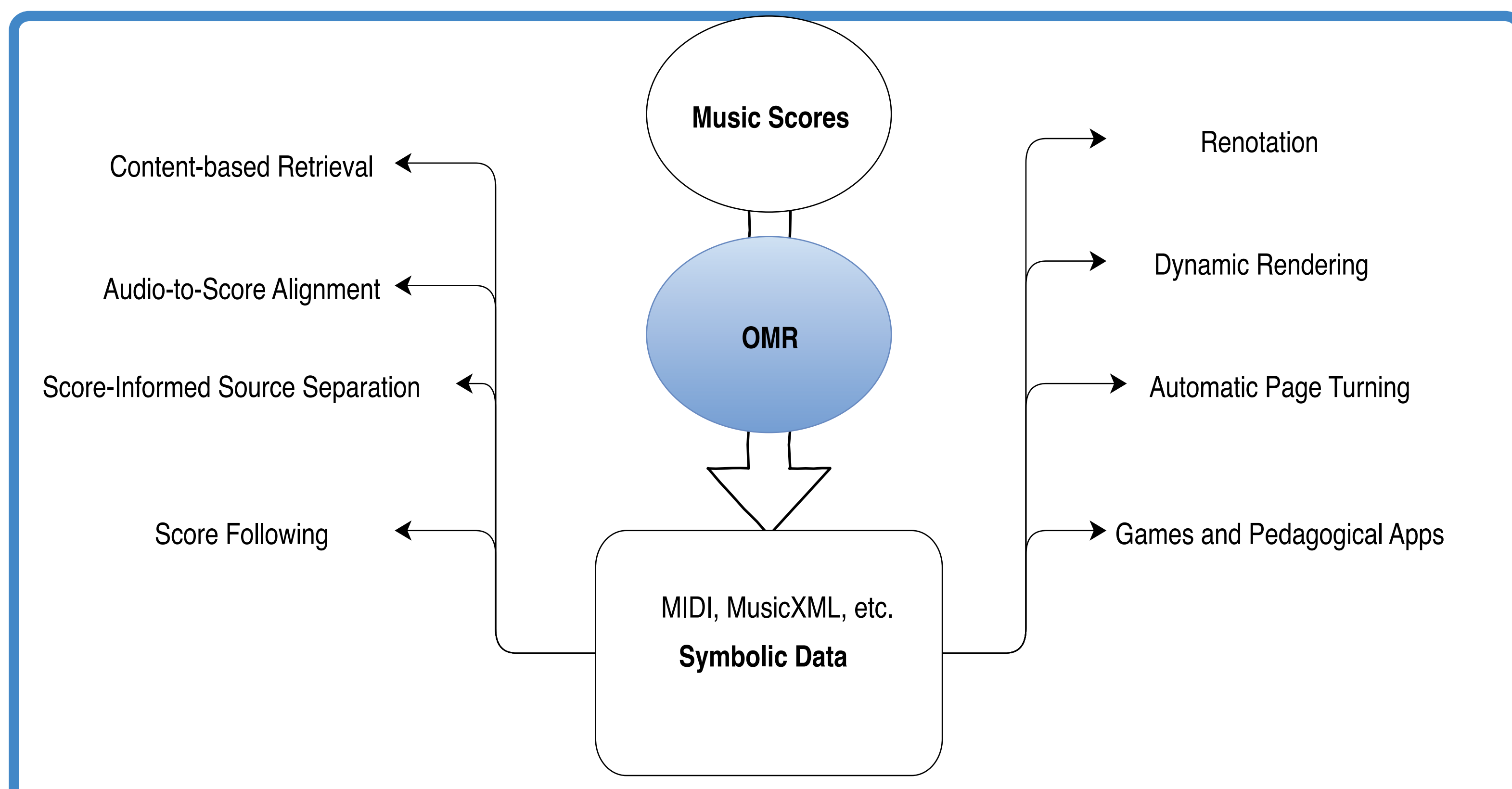
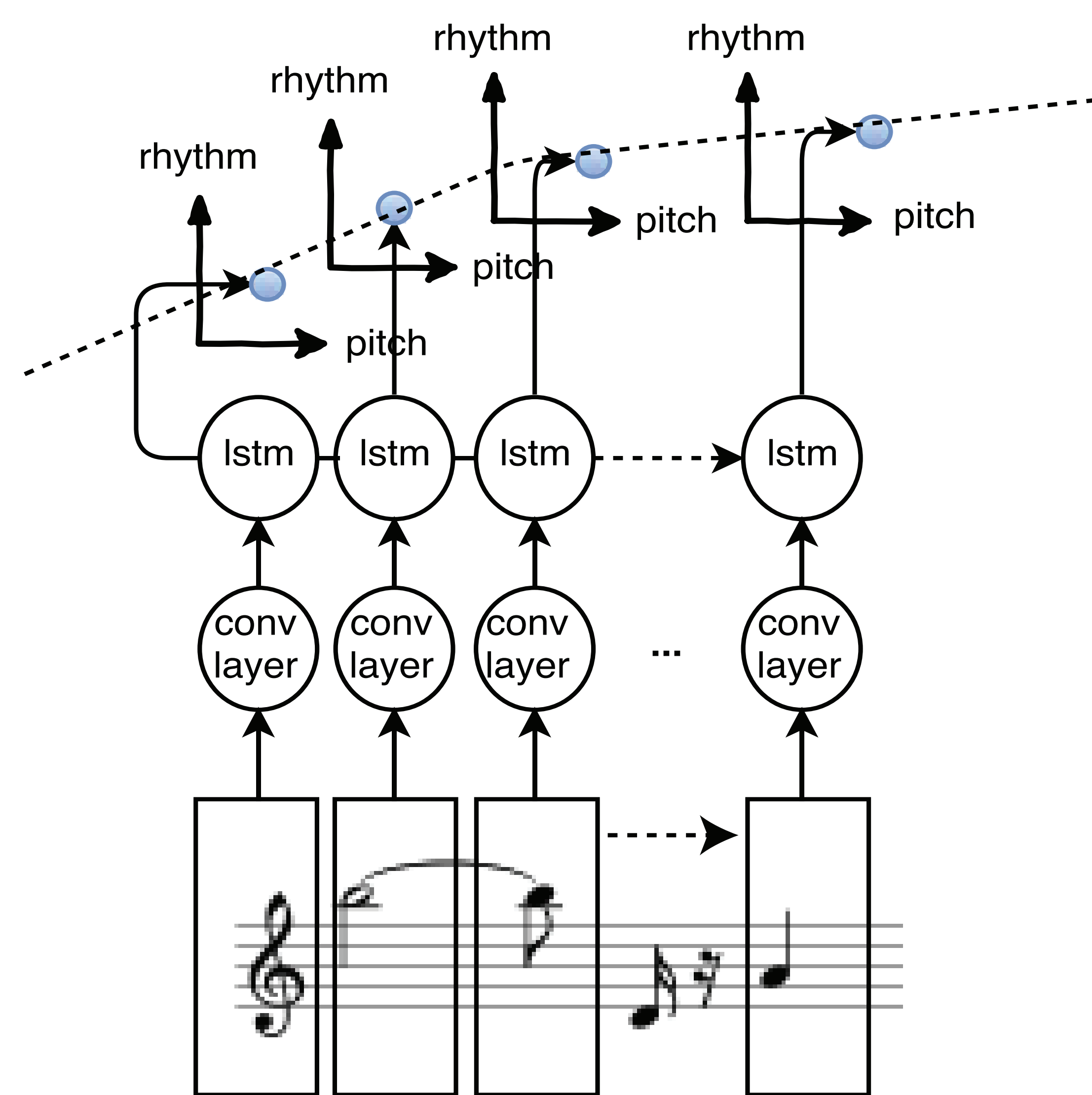


A Hybrid HMM-RNN for Optical Music Recognition

Liang Chen, Rong Jin, Simo Zhang, Stefan Lee, Zhenhua Chen, David Crandall
Indiana University Bloomington



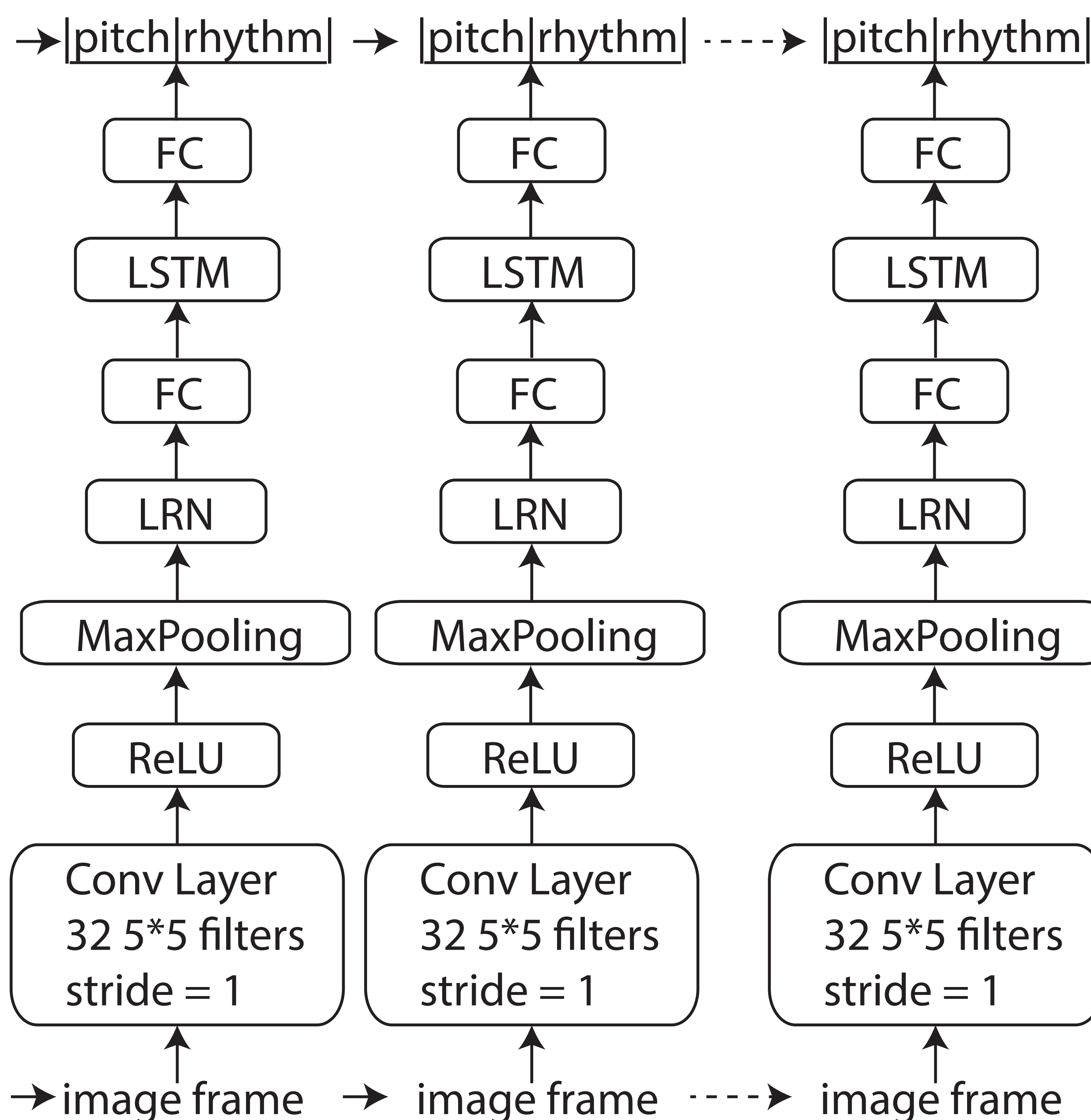
Our Model (Architecture)



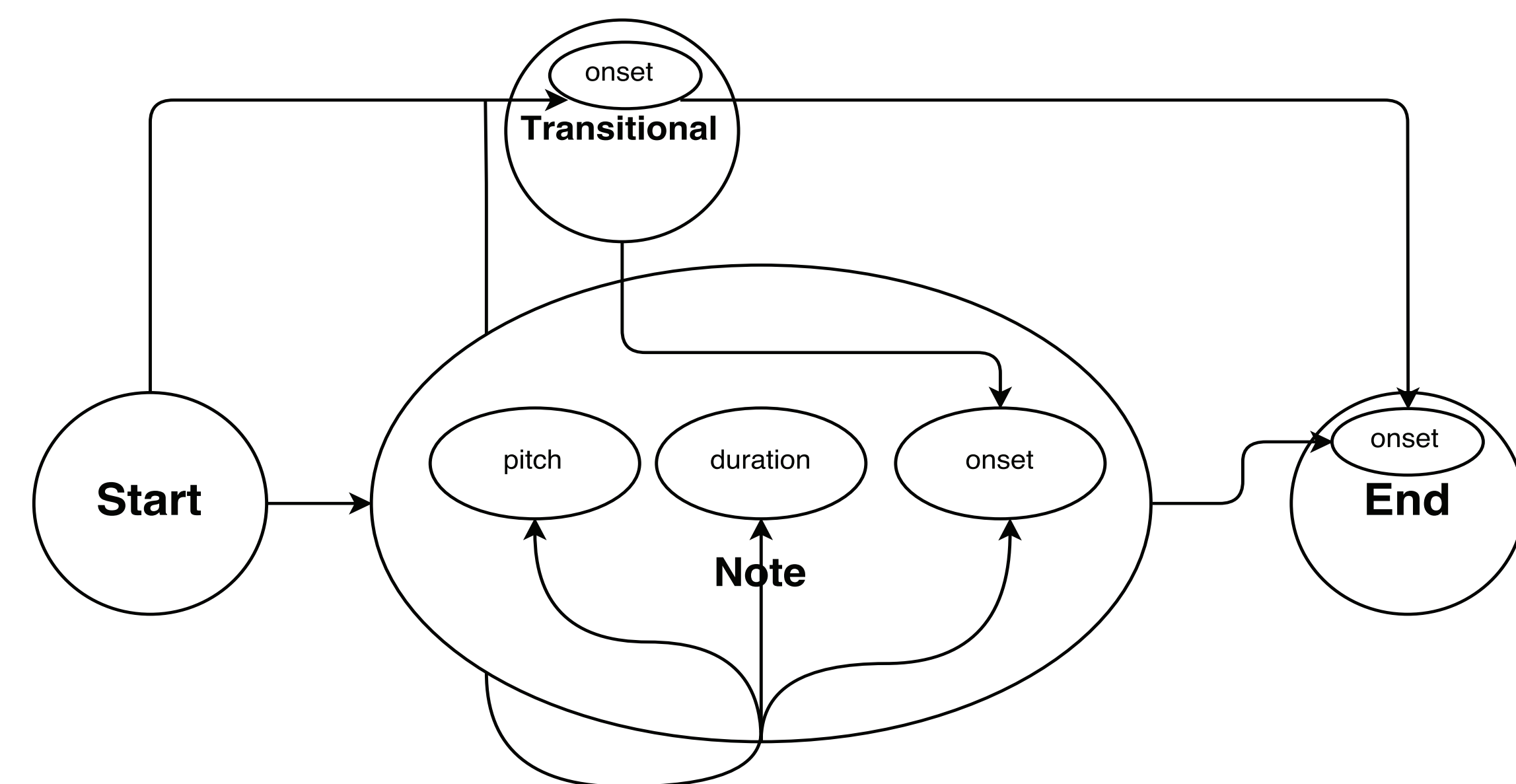
The Motivation:

- 1) Using deep neural networks to project score image to musical semantic space.
- 2) Using HMM to enforce rhythmic constraints and interpret music from semantic output.
- 3) Multi-label classification with CNN-RNN to jointly estimate pitch and rhythm.
- 4) Using CNN to extract discriminative features that account for different symbols.
- 5) Using RNN to learn long-term dependencies for semantic label prediction.

Net Configurations



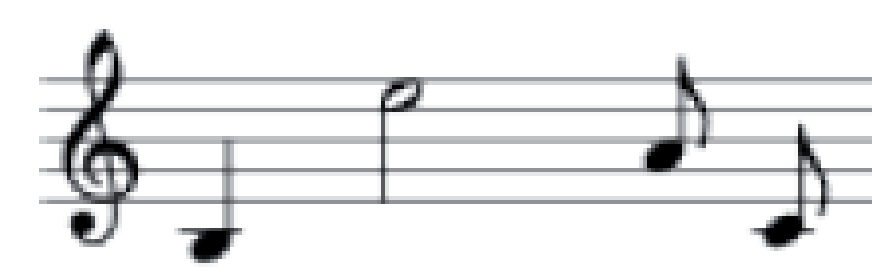
Our Model (Music Interpretation)



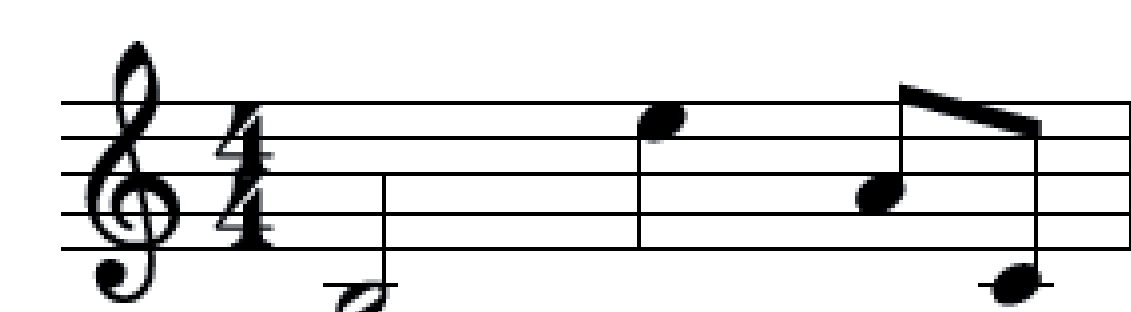
Preliminary Results

Pitch	Rhythm	Pitch (w/o bkg)	Rhythm (w/o bkg)
94.82%	88.29%	75.47%	43.17%

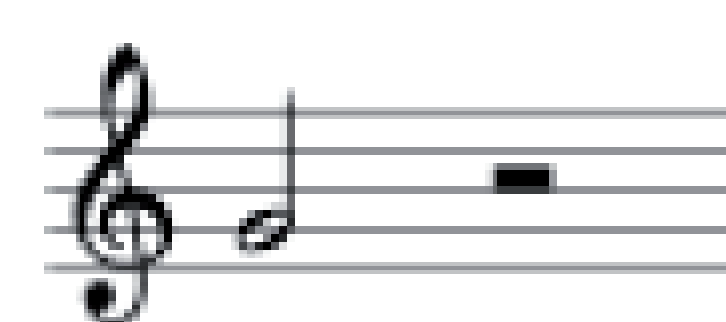
Per-frame accuracy for pitch and rhythm predictions. bkg represents background frames.



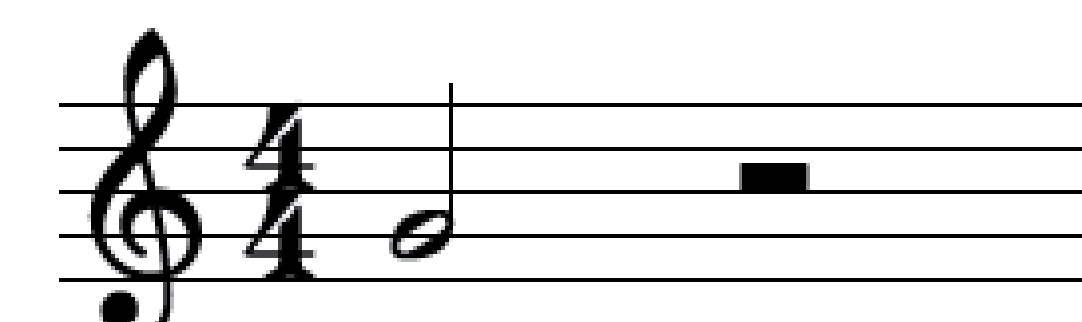
(a) Test image



(b) HMM reconstruction



(c) Test image



(d) HMM reconstruction