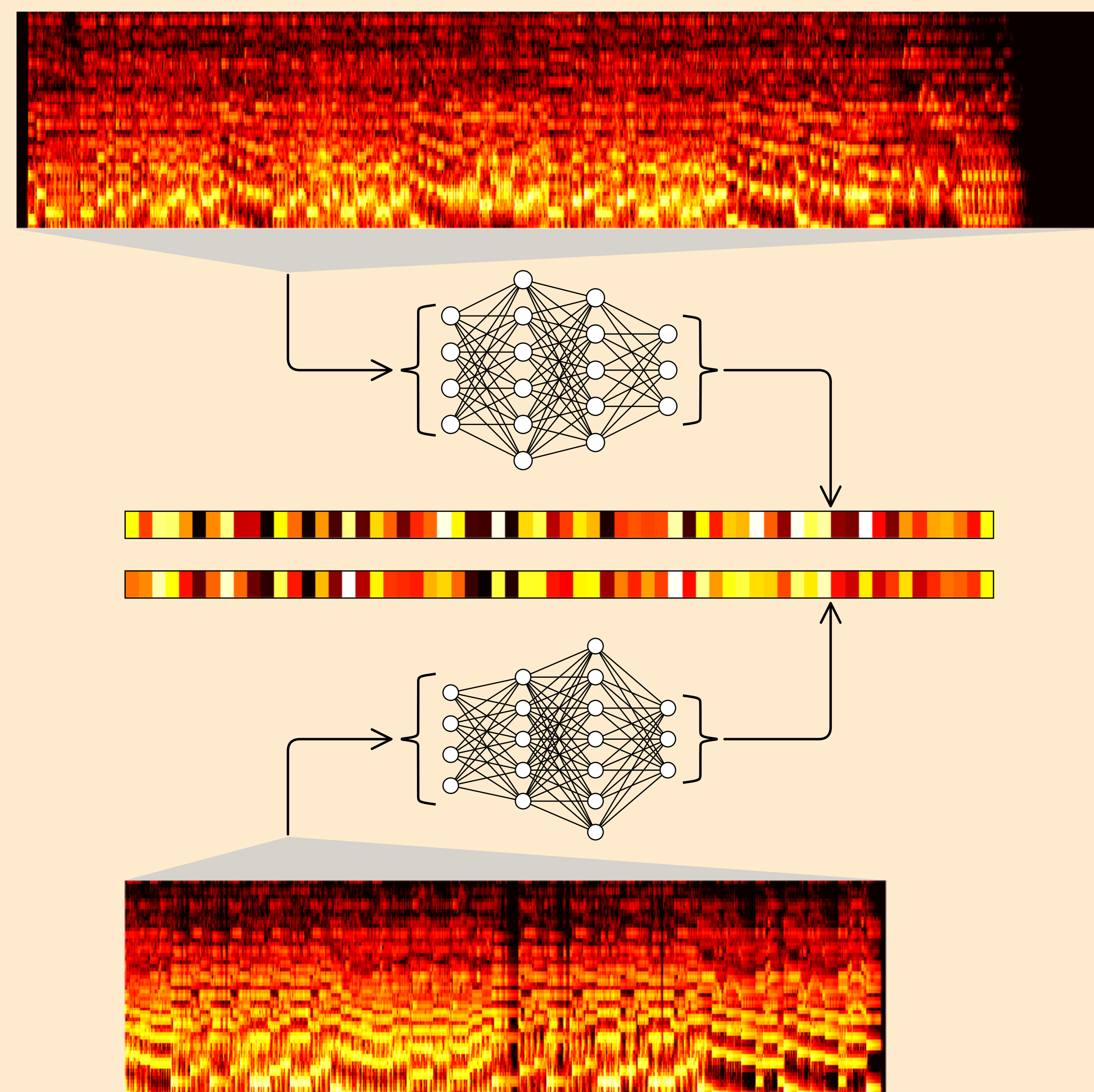
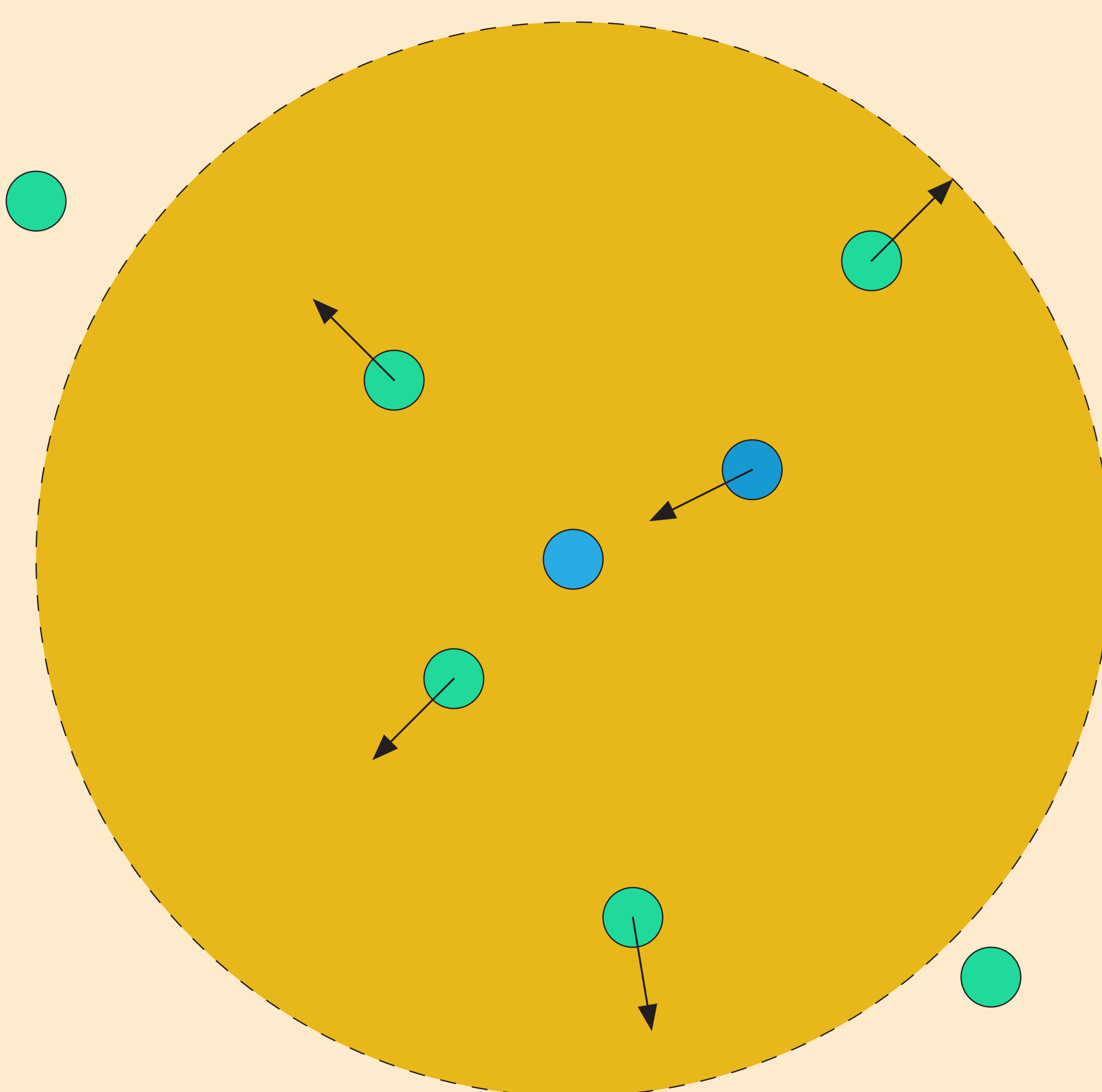


Searching a large database to find a sequence that is most similar to a query can be prohibitively expensive, particularly if individual sequence comparisons involve complex operations such as warping. To achieve scalability, “pruning” heuristics are typically employed to minimize the portion of the database that must be searched. We present an approximate pruning technique which embeds sequences in a Euclidean space where similar sequences have a small distance and dissimilar sequences are far apart. We propose the use of a convolutional network with a feedforward attention mechanism which integrates over time. By using fixed-length embeddings, our pruning method effectively runs in constant time, making it many orders of magnitude faster than full dynamic time warping-based matching. We demonstrate our approach on the task of matching synthesized MIDI transcriptions to a database of one million audio recordings.

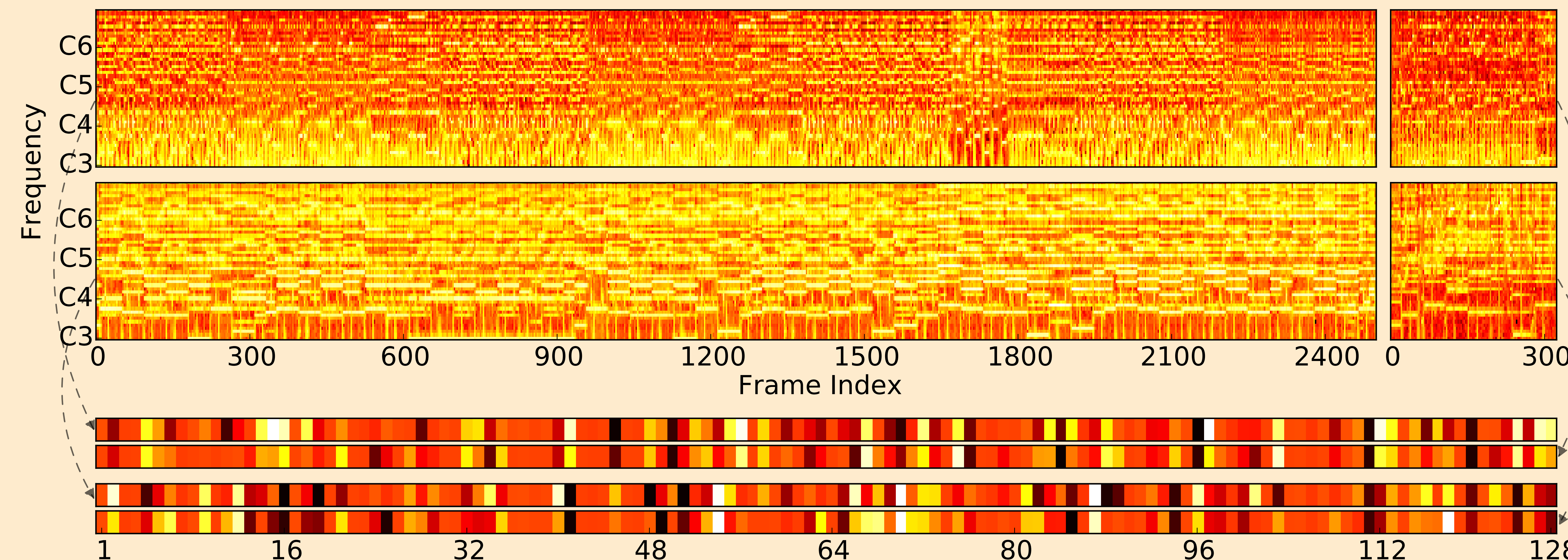
Pairwise sequence embedding



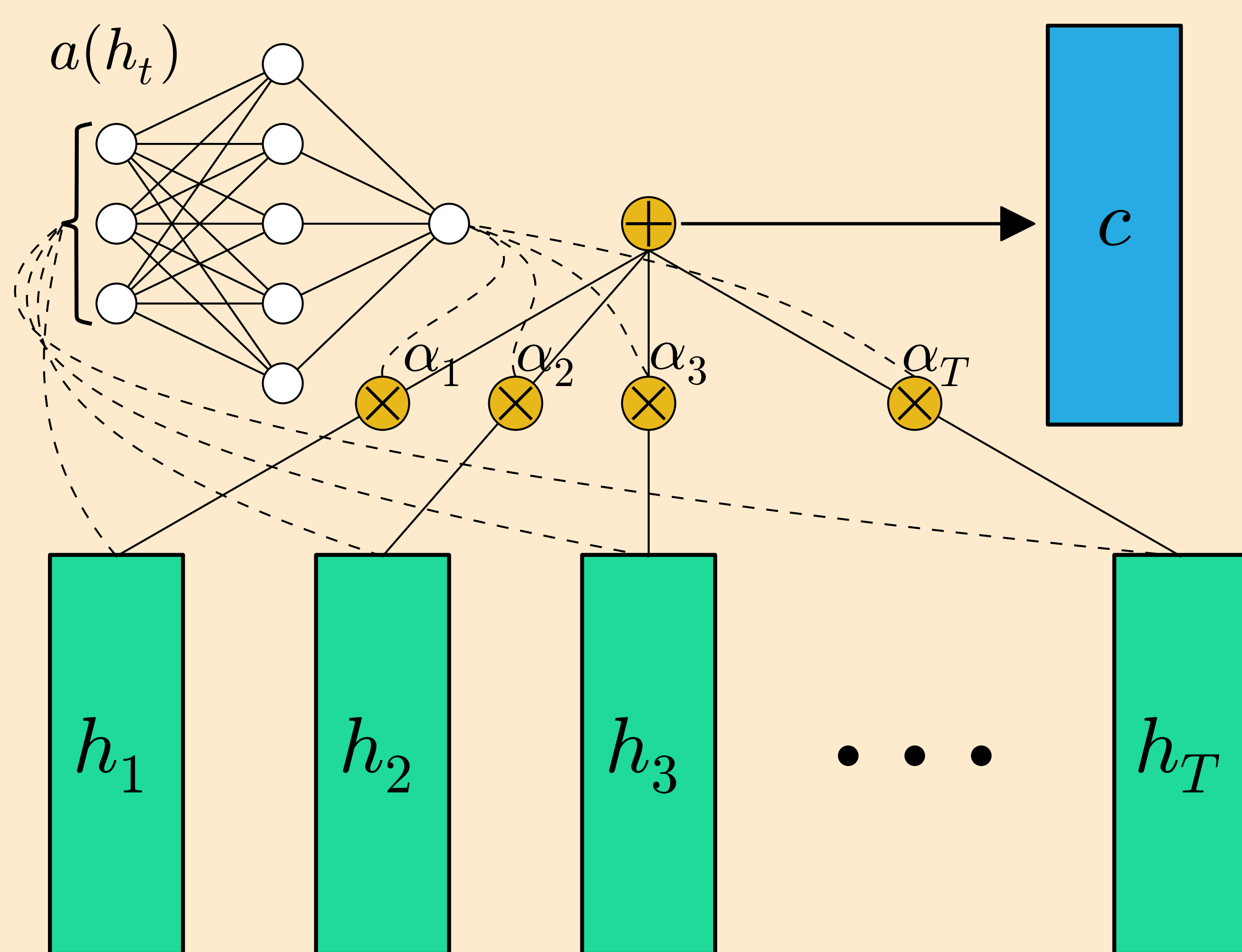
Objective function



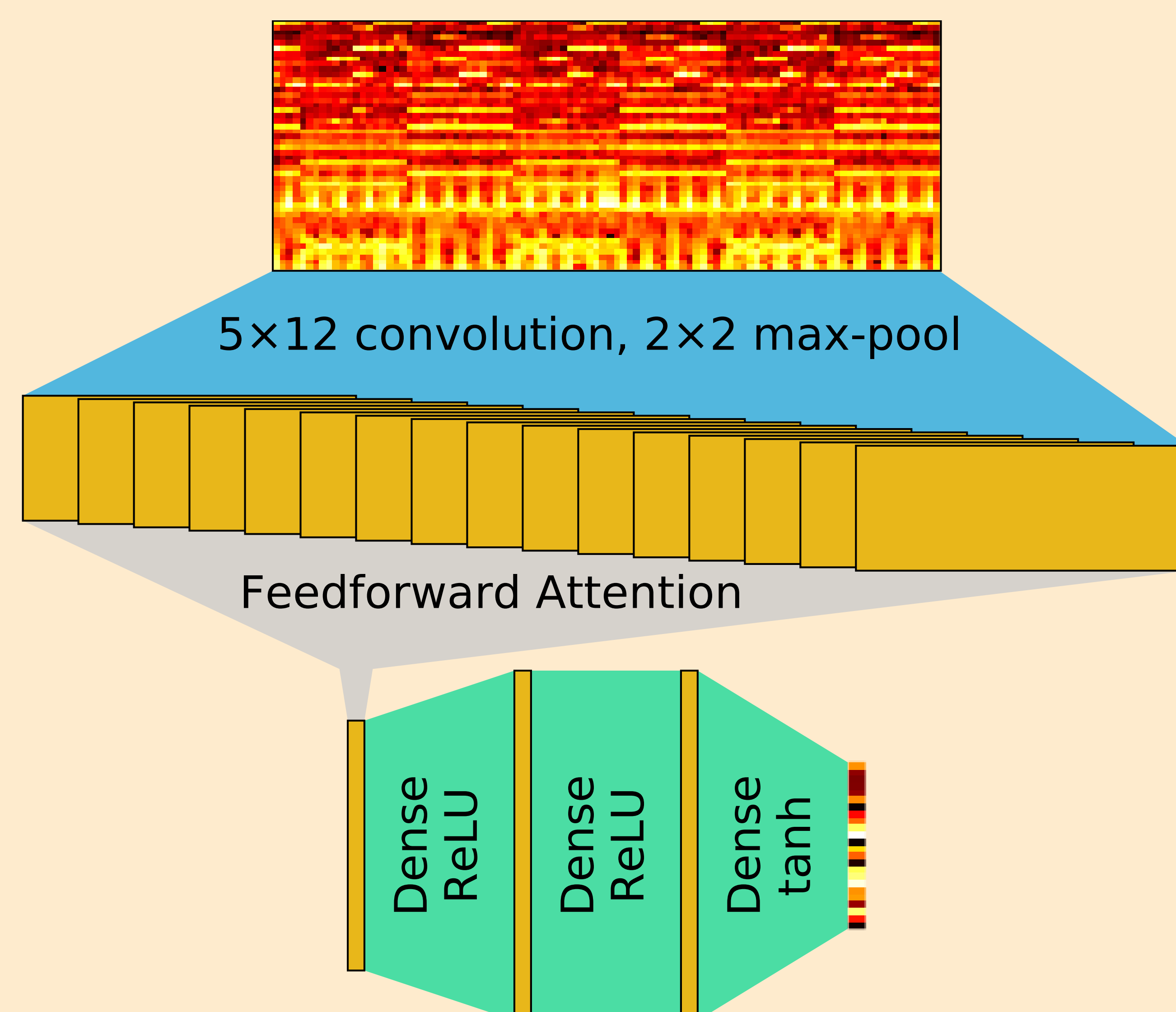
Example embeddings of matching spectrograms of MIDI transcriptions and audio recordings



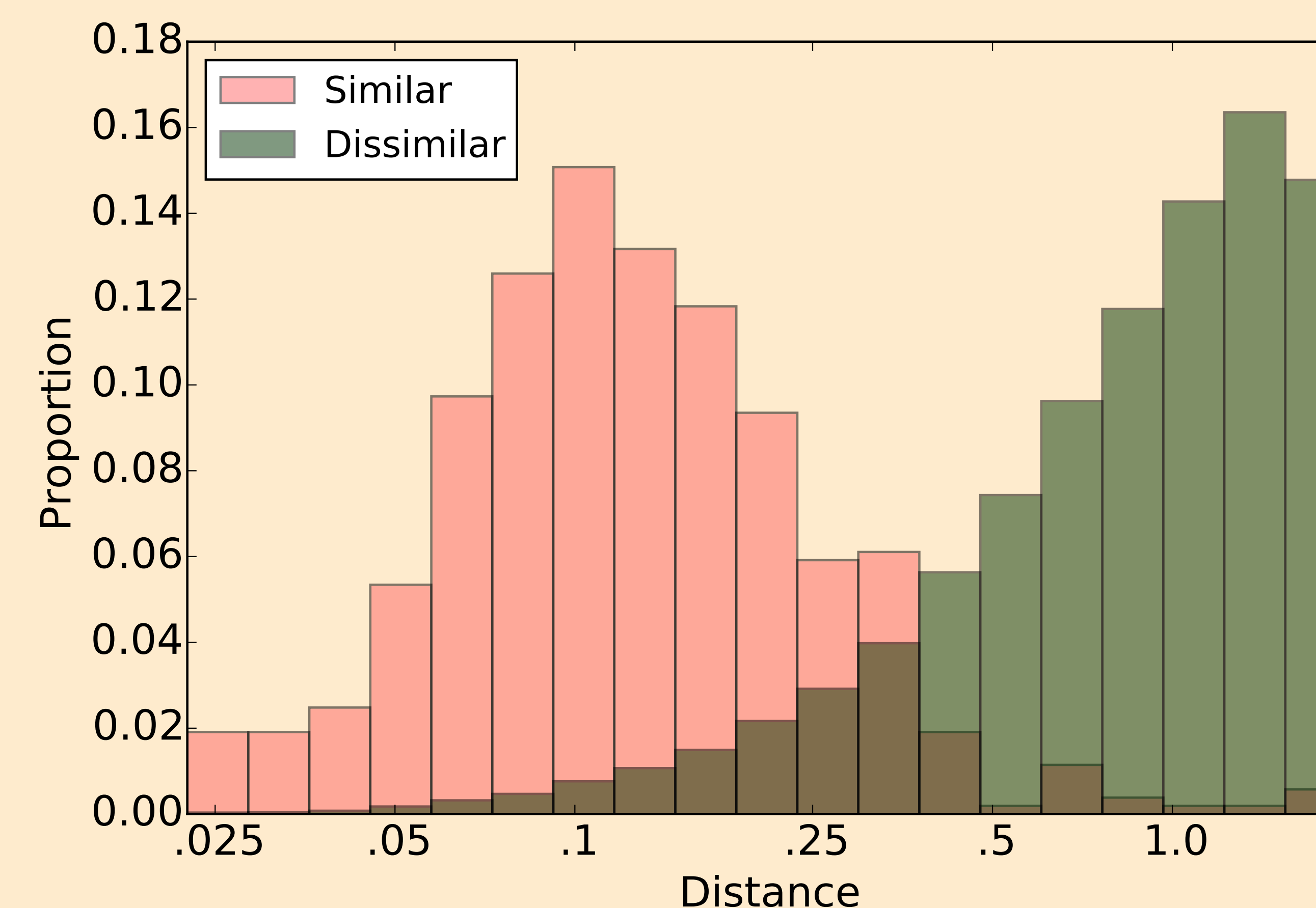
Feedforward attention



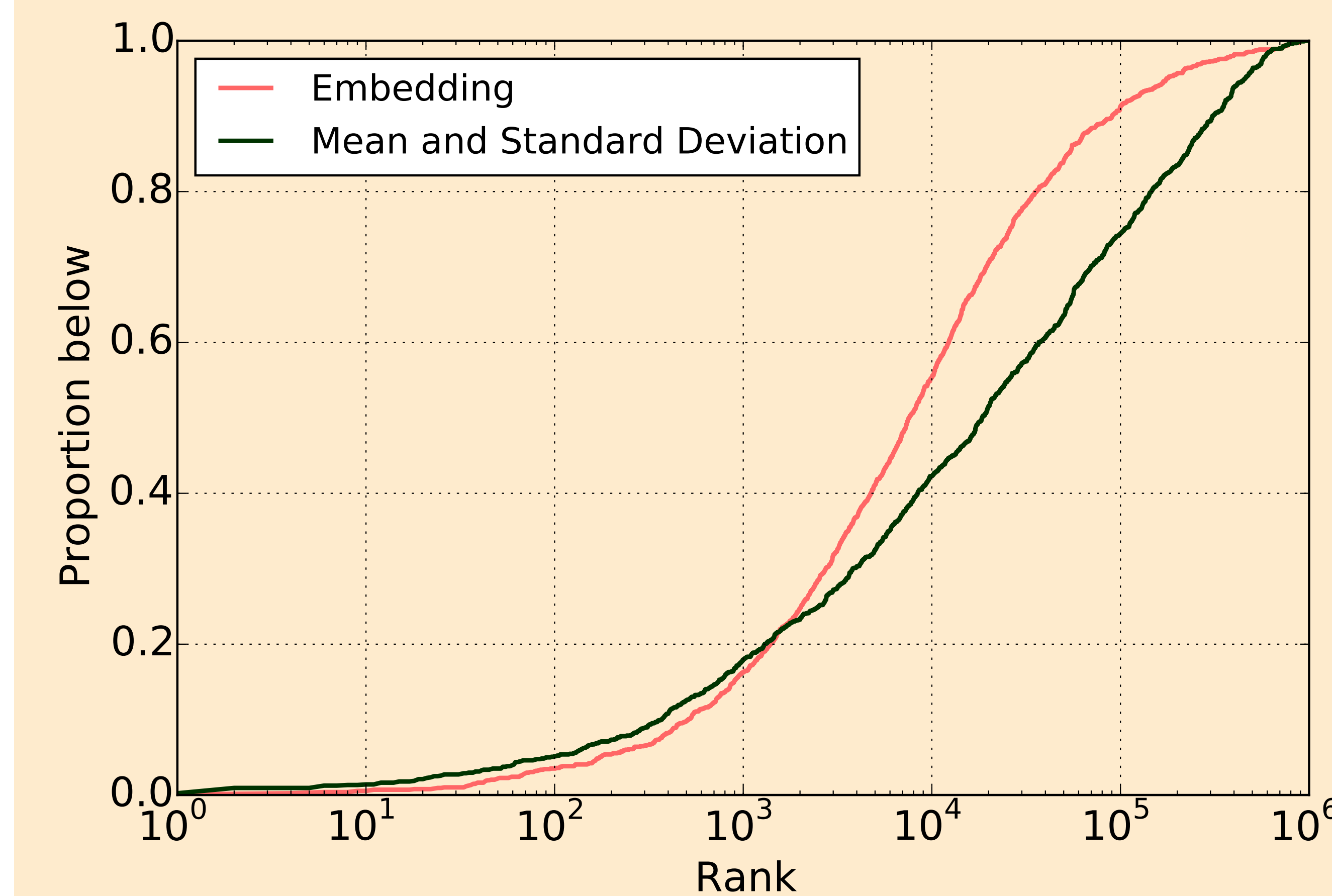
Network structure



Distance distributions



Ranks



See also

“Large-Scale Content-Based Matching of MIDI and Audio Files”, in *Proceedings of the 16th International Society for Information Retrieval Conference*, 2015.

“Feed-Forward Networks with Attention Can Solve Some Long-Term Memory Problems”, in *arxiv:1512.08756*, 2015.