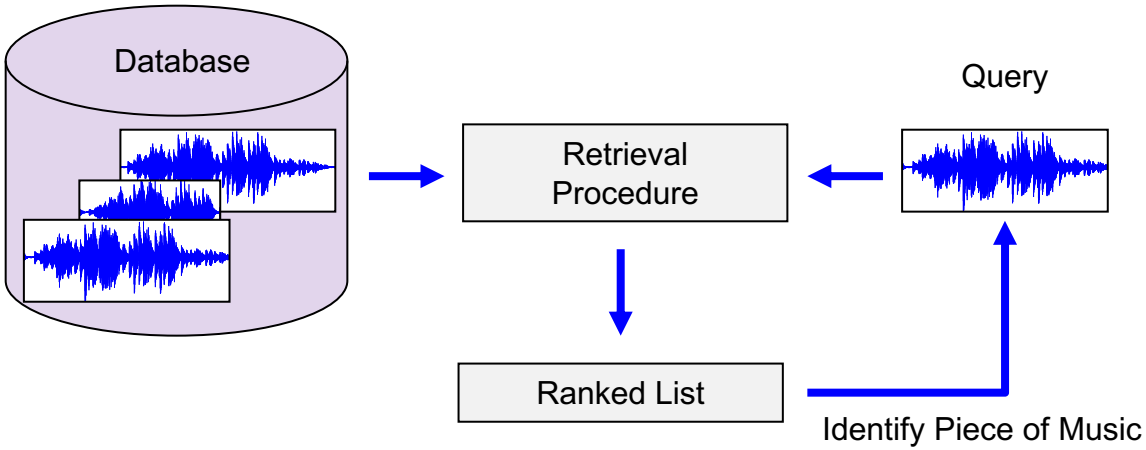
Comparison of PCA- and Autoencoder-Based Dimensionality Reduction of Feature Sequences for Efficient Music Retrieval

Frank Zalkow, Meinard Müller

International Audio Laboratories Erlangen

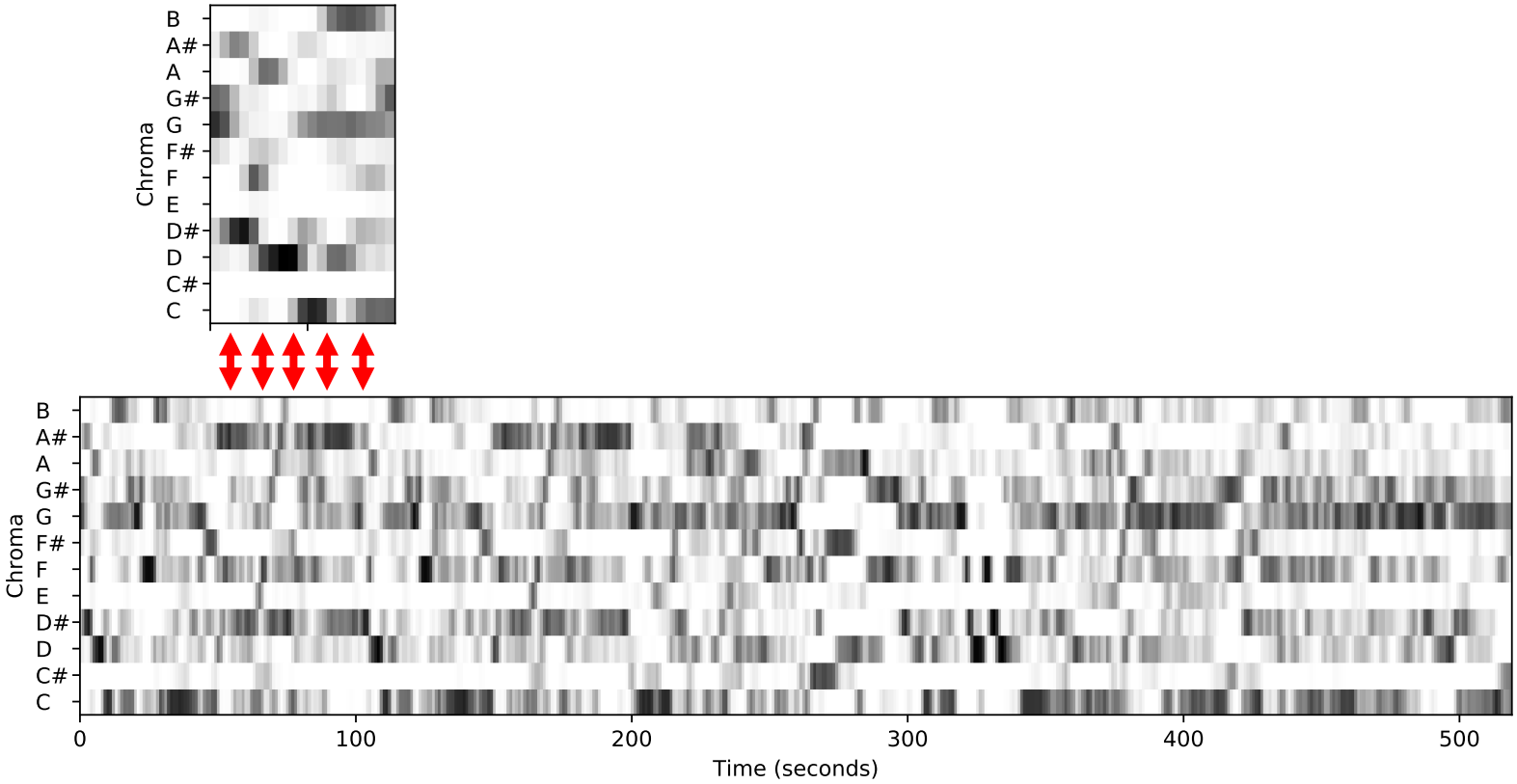
Application: Audio Matching for Classical Music

- Application: Identifying a classical piece of music by an audio recording, possibly incomplete
- Given: A database that contains this piece of music, but possibly in a different performance
- Challenges: Differences in tempo, instrumentation, articulation, etc.



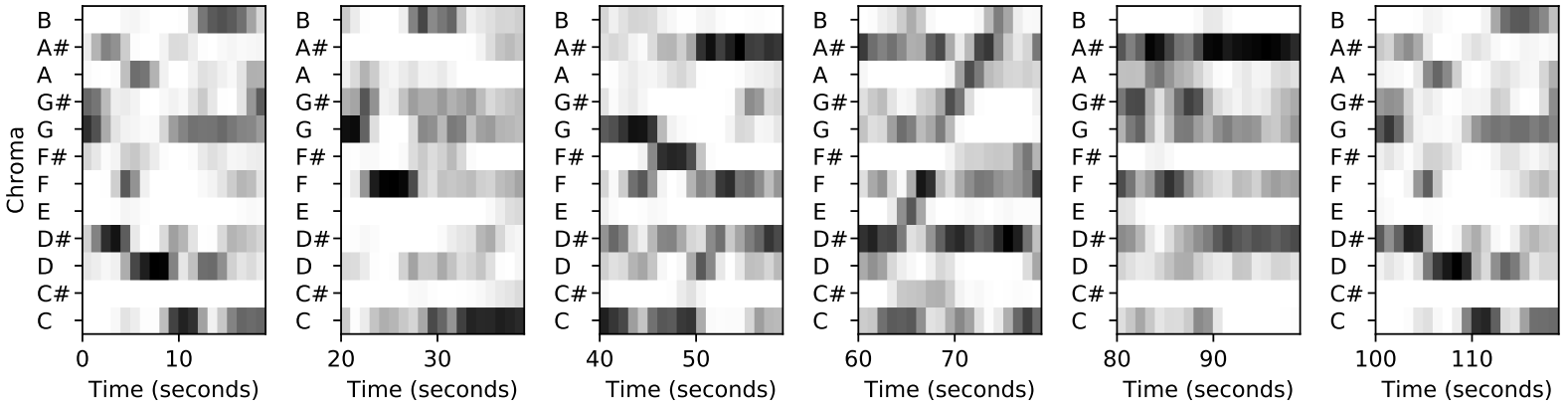
Previous Work

- Comparison of query and database on the basis of chroma features



Previous Work

- Ideal: Subsequence DTW, but prohibitive runtime for large databases
- Previous Work [1]: Shingle approach
- Turned out to be suited: 240 dimensional CENS shingles → the basis for our work



[1] Peter Grosche and Meinard Müller: *Toward Characteristic Audio Shingles for Efficient Cross-Version Music Retrieval*, Proc. of ICASSP, Kyoto, Japan, 2012, pp. 473–476.

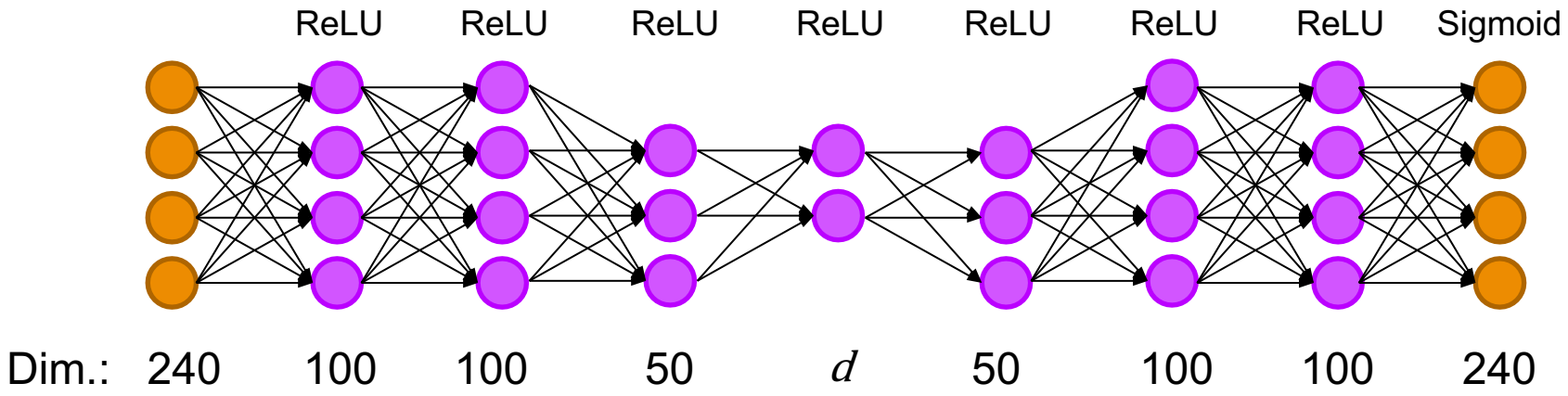
Data Set

Composer	Work	Movement	Recording	hh:mm:ss
Beethoven	Op. 67	1	10	01:12:07
		2	10	01:44:53
		3	10	01:02:53
		4	10	01:48:00
Chopin	Op. 17 No. 4		64	04:36:58
	Op. 24 No. 2		64	02:26:38
	Op. 30 No. 2		34	00:48:11
	Op. 63 No. 3		88	03:09:08
	Op. 68 No. 3		51	01:25:58
Vivaldi	RV 315	1	7	00:37:40
		2	7	00:17:23
		3	7	00:20:40
			362	19:30:28

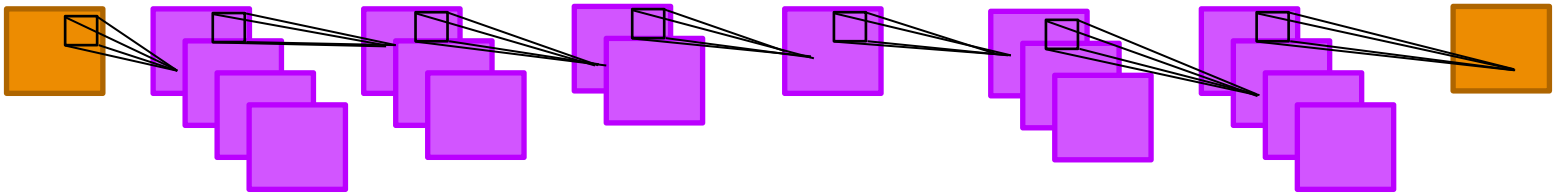
Our Approach

- Our goal: Reducing the dimensionality of the shingles while keeping good matching results
- PCA
- Fully-connected autoencoder
- Fully-convolutional autoencoder

Fully-Connected Autoencoder

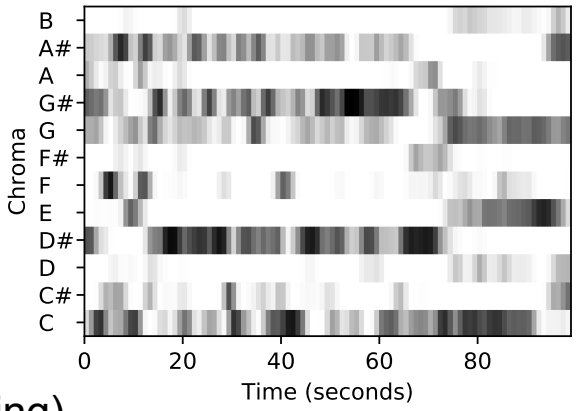


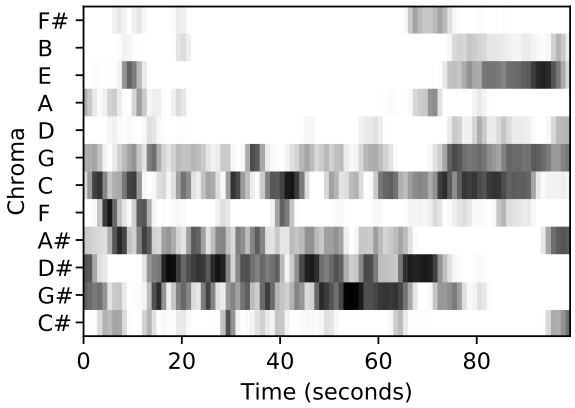
Fully-Convolutional Autoencoder



32 3x3 Conv Relu Maxpool	8 3x3 Conv Relu Maxpool	2 3x3 Conv Relu Maxpool	1 3x3 Conv **Relu**	UpSample 8 3x3 Conv Relu	UpSample 32 3x3 Conv Relu	UpSample 1 3x3 Conv Relu

Experiments

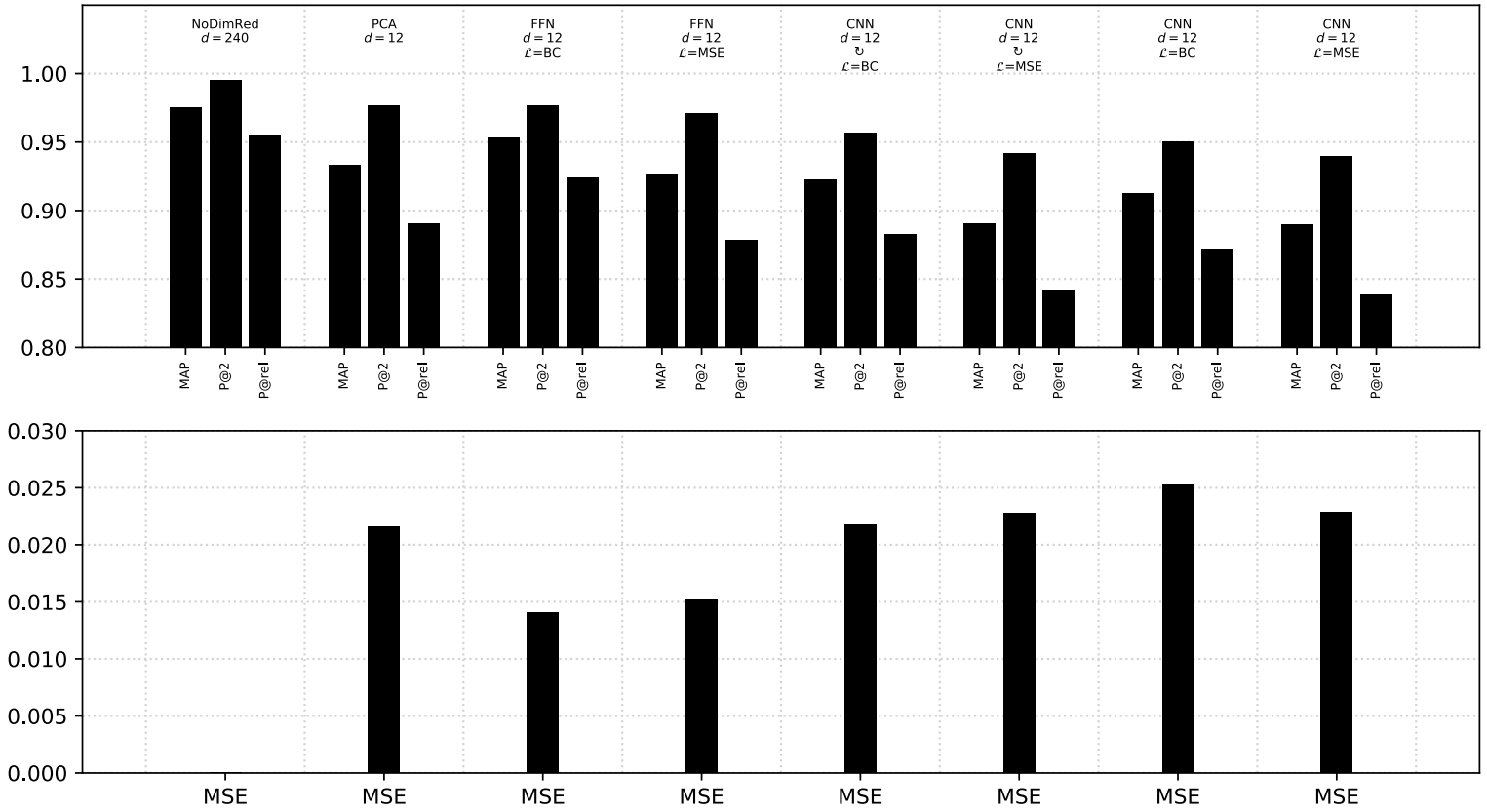




- PCA
  - Number of components
  - Preprocessing (standardizing)
- Deep autoencoder
  - Size of encoding
  - Preprocessing (standardizing)
  - Fully-connected and fully-convolutional
  - Loss function
  - For convolutional: Ordering of chroma bins in cycle of fifth
- Evaluation
  - MAP
  - P@2
  - P@rel

Evaluation

- Comparison of 12-dimensional versions, no standardizing



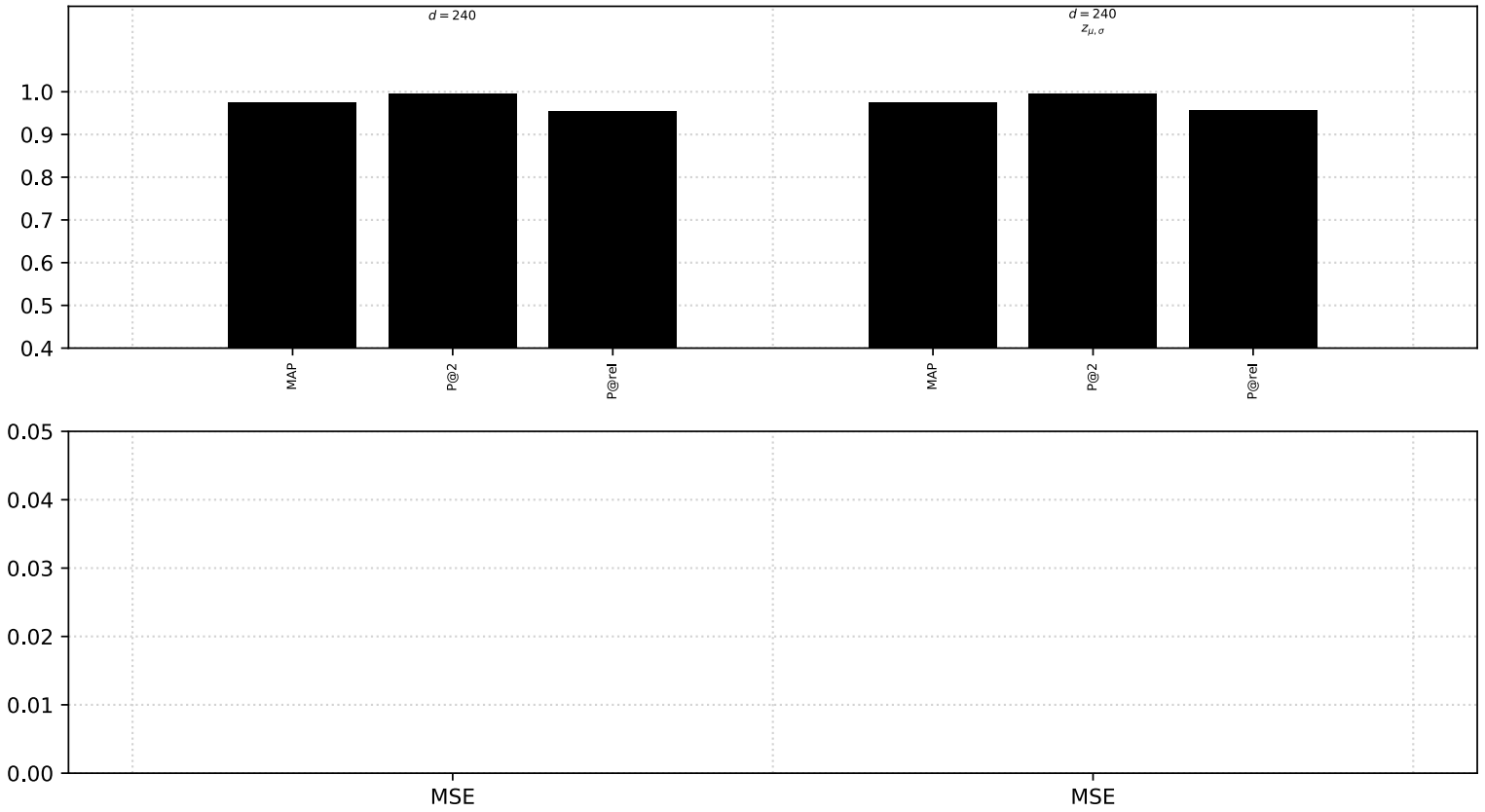
Future Work

- Different types of regularization
- Variational autoencoder
- Siamese networks
- Using a bigger data set

Backup Slides

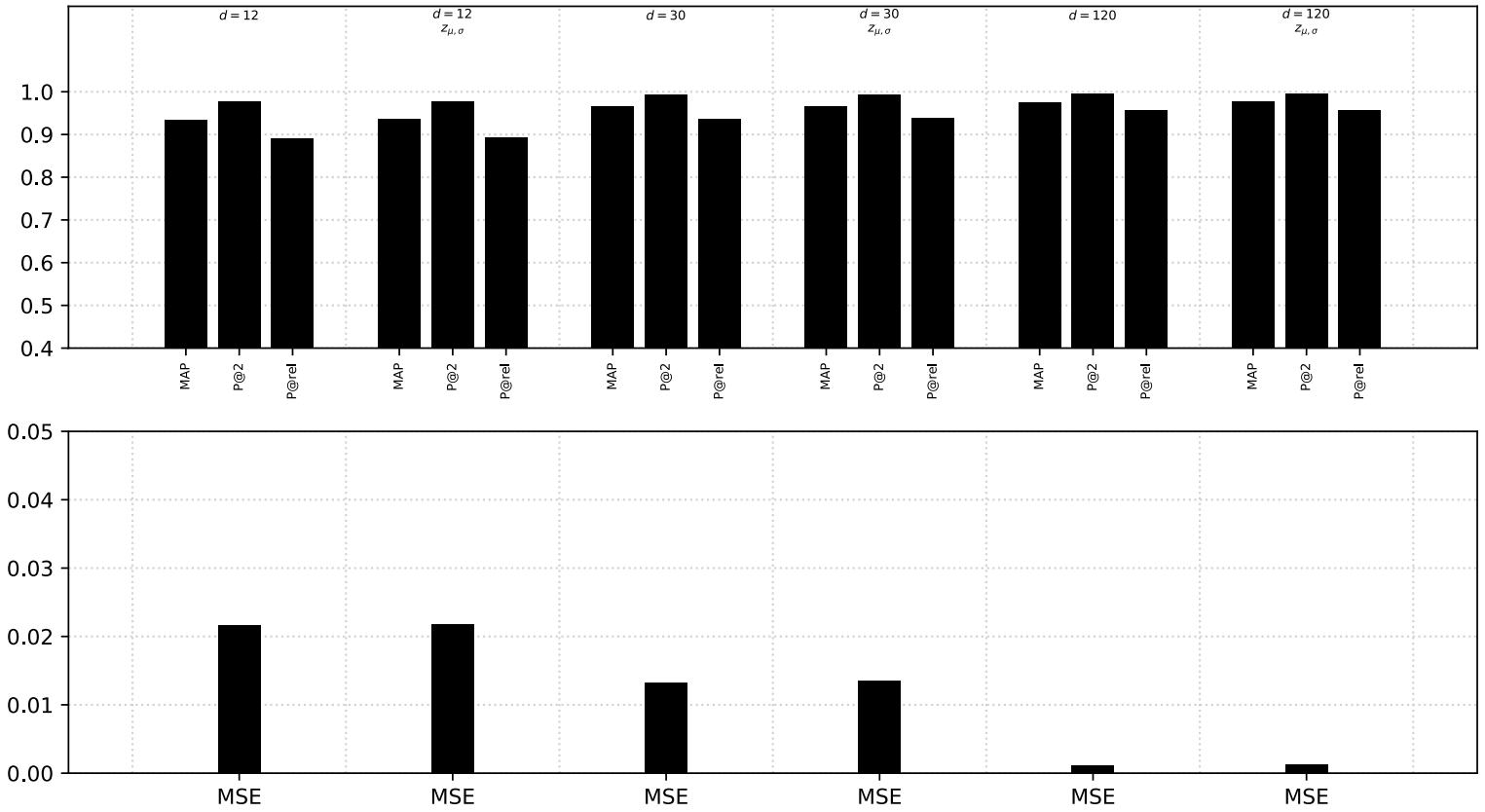
Evaluation

- No dimensionality reduction



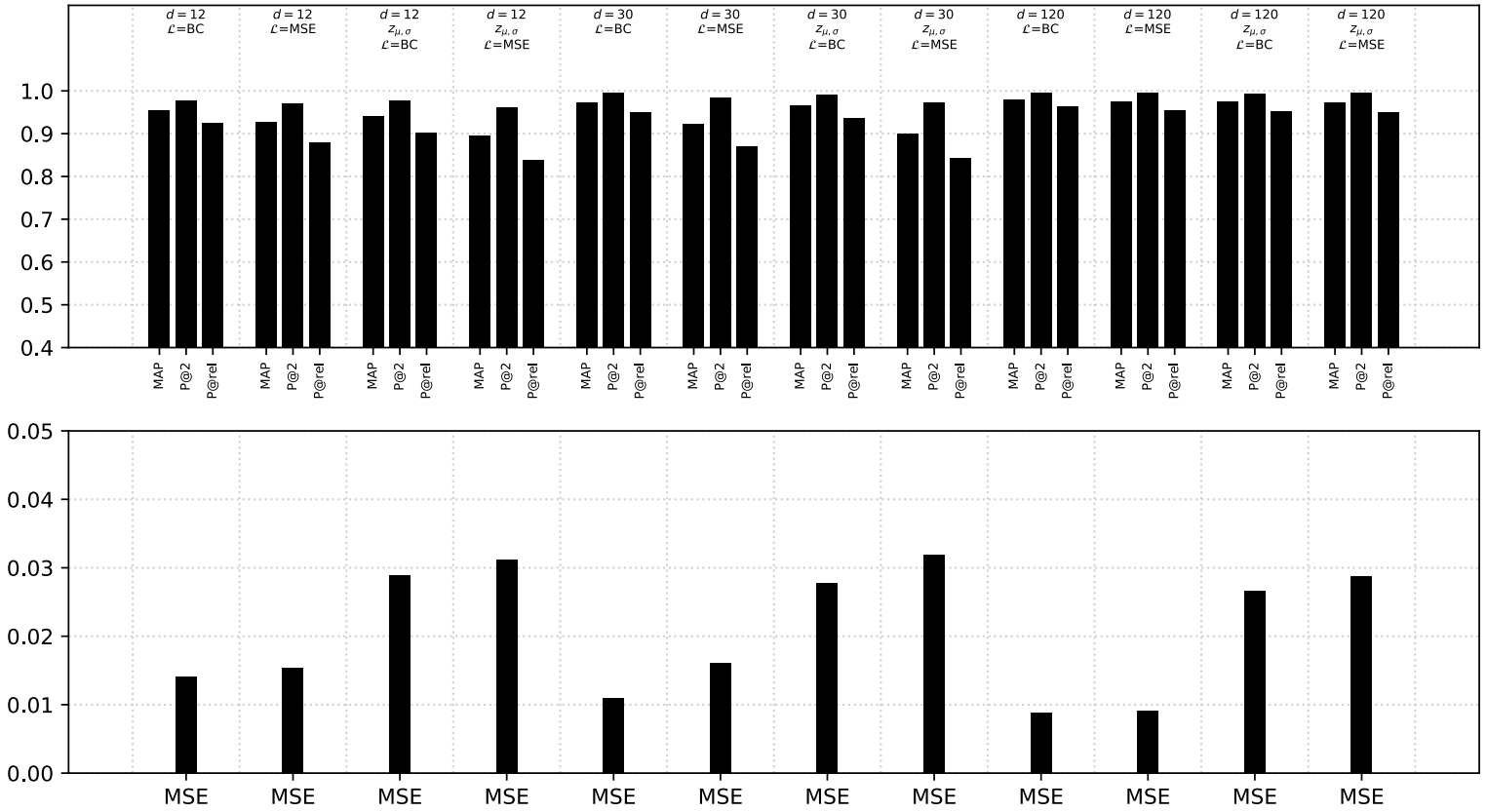
Evaluation

- PCA



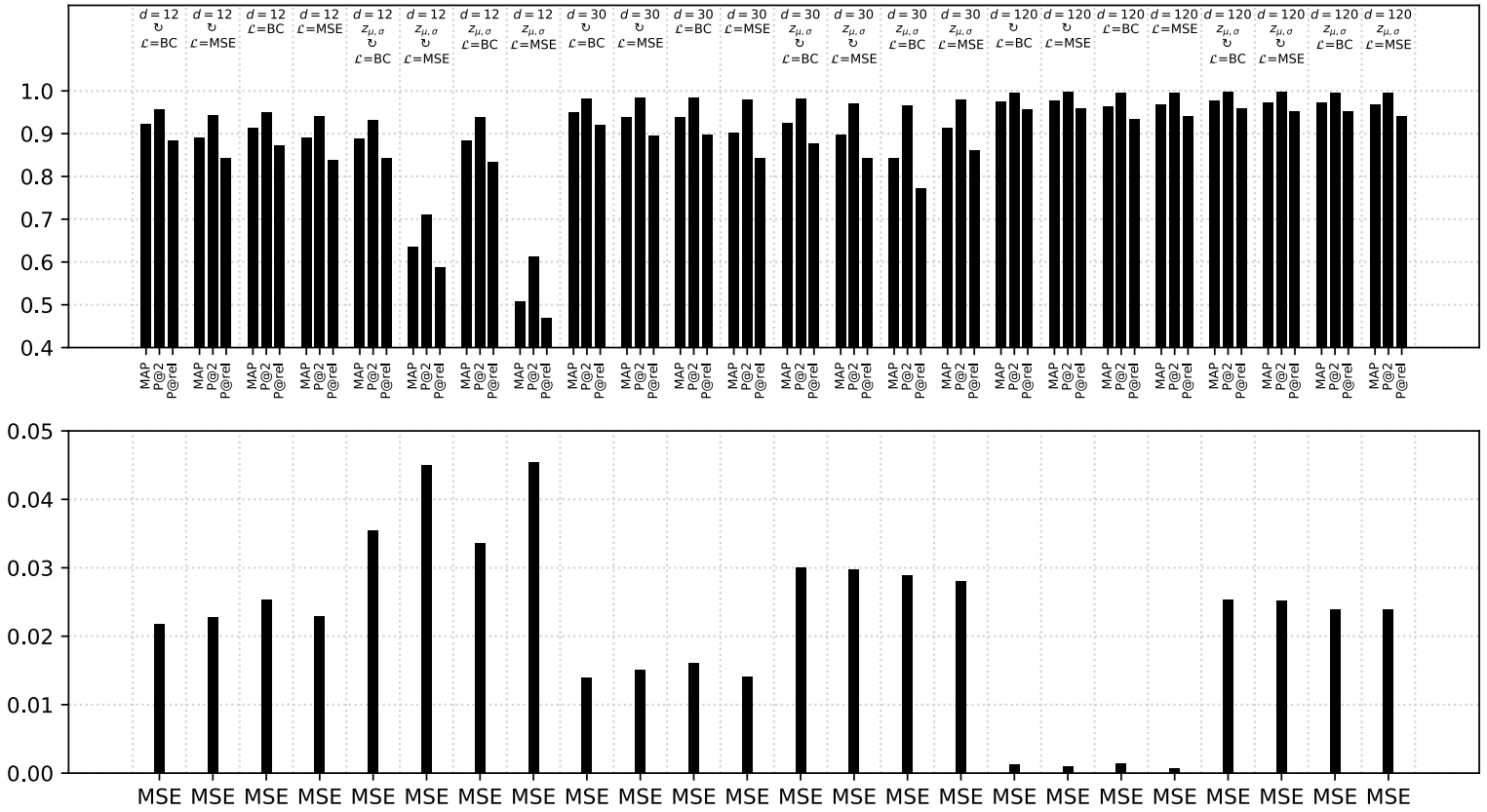
Evaluation

- Fully-connected network



Evaluation

- Fully-convolutional network



Evaluation

- Shallow autoencoder with PCA initialization