Deep Learning and Its application in Multimedia Retrieval

ZHIGUANG WANG

Roadmap

- Introduction
- Background of Stacked Auto-encoder
- Background of Convolutional Neural Network
- Application in Image Retrieval
- Application in Audio Retrieval
- Conclusion

Intro

- Modern IR laid more emphasis, as well as effort on multimedia retrieval
- IR tasks boost W.R.T image and audio content
- Image Retrieval
- Audio/Music Retreval
- Music Recommedation
- …

Intro

- Content based image retrieval

Intro

- Content based Music retrieval

	Artists with positive values	Artists with negative values
1	Justin Bieber, Alicia Keys, Maroon 5, John Mayer, Michael Bublé	The Kills, Interpol, Man Man, Beirut, the bird and the bee
2	Bonobo, Flying Lotus, Cut Copy, Chromeo, Boys Noize	Shinedown, Rise Against, Avenged Sevenfold, Nickelback, Flyleaf
3	Phoenix, Crystal Castles, Muse, Röyksopp, Paramore	Traveling Wilburys, Cat Stevens, Creedence Clearwater Revival, Van Halen, The Police

Intro

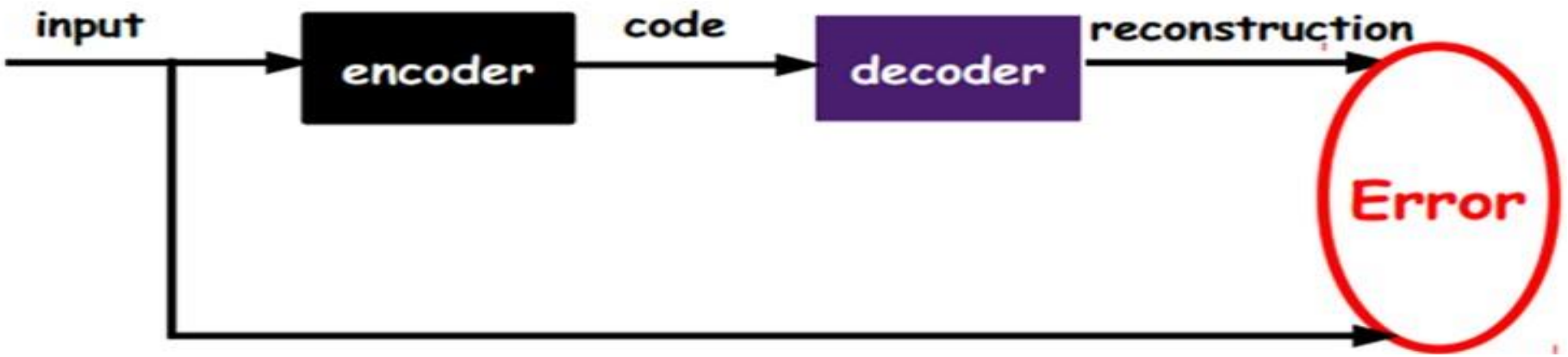
- Modern IR intertwines with ML and statistics.
- Inspires various methods for feature representation, language modeling, etc.
- Deep learning, new architecture
- Learn abstract/extensible features for IR tasks

Background - SAE

- Is it possible to learn features without label?
- SURE!
- Decompose and reconstruct itself!

Background - SAE

- Encode and decode
- code -> feature

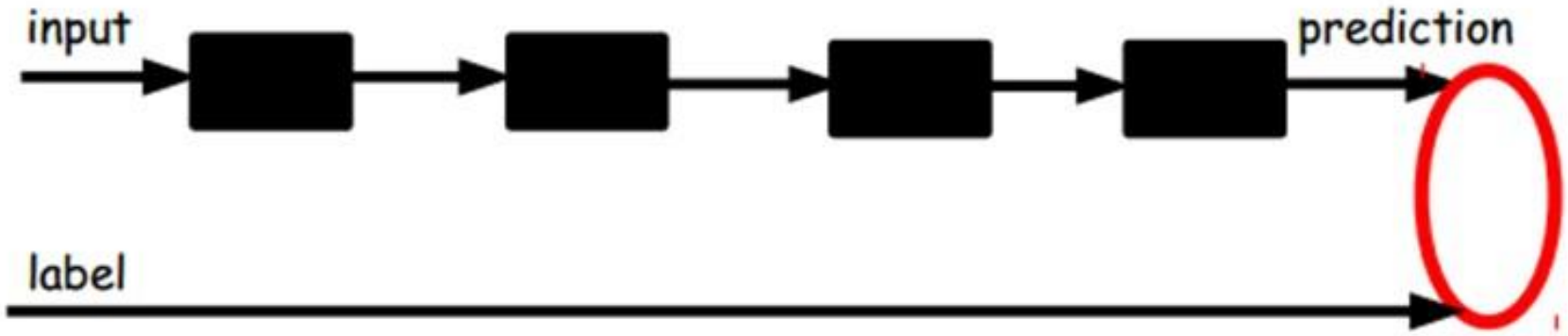


Background - SAE

- Fine tune the final search results/classifier/predictor

Background - SAE

- Fine tune the entire search model

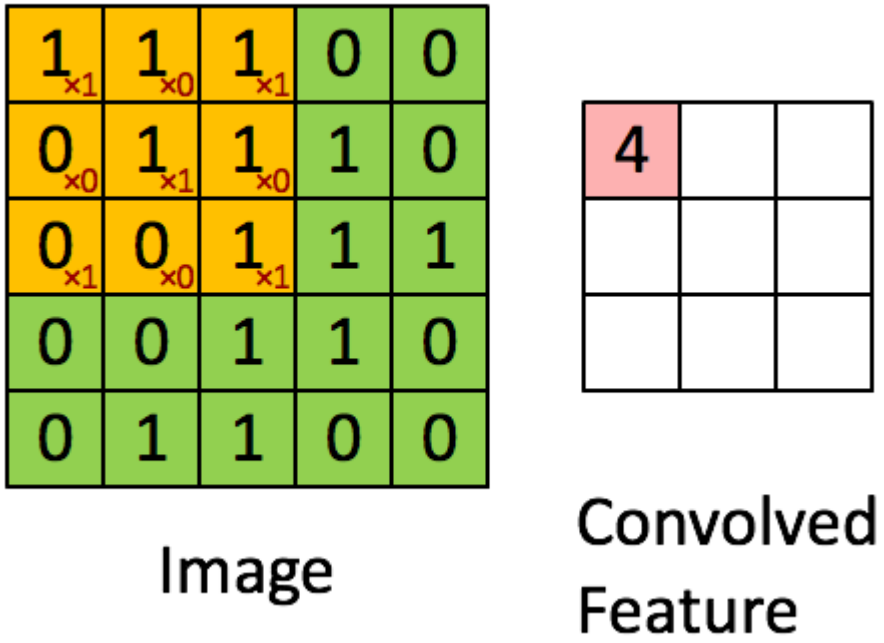


Background - SAE

- Feature could be:
- bag-of-words
- image pixels
- meta data vector
- …

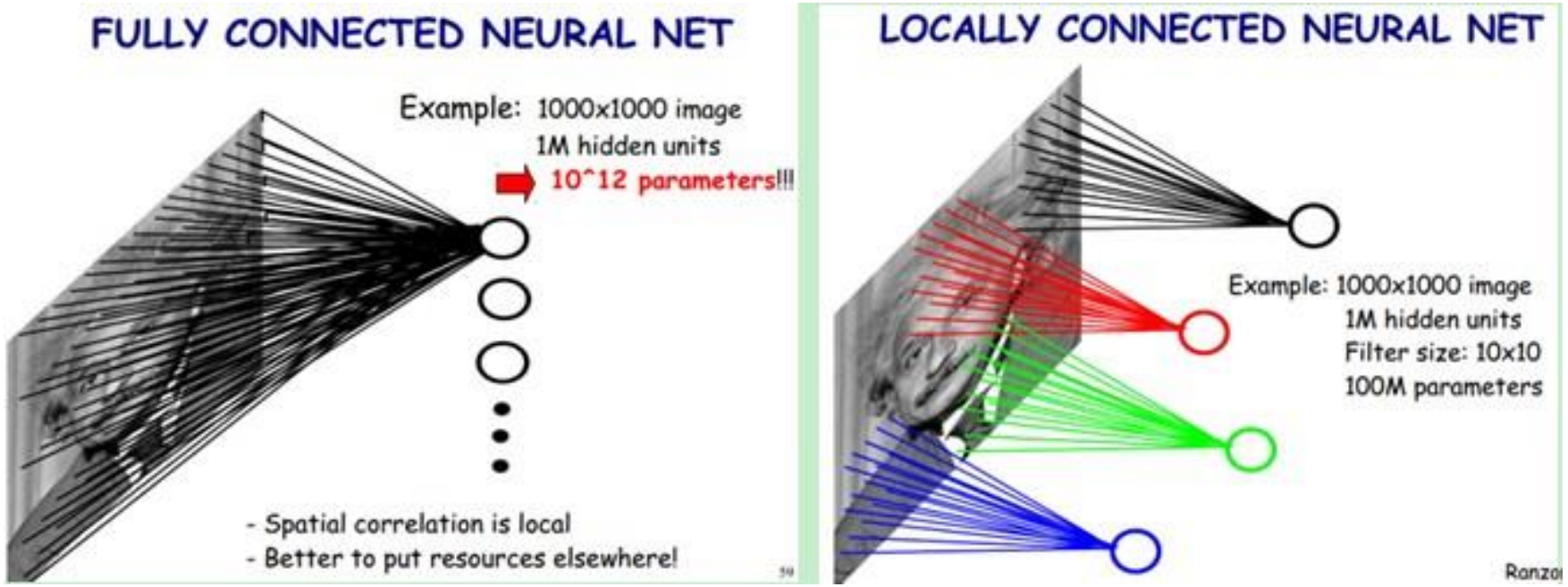
Background - CNN

- Convolution – extract features from multi-dimension



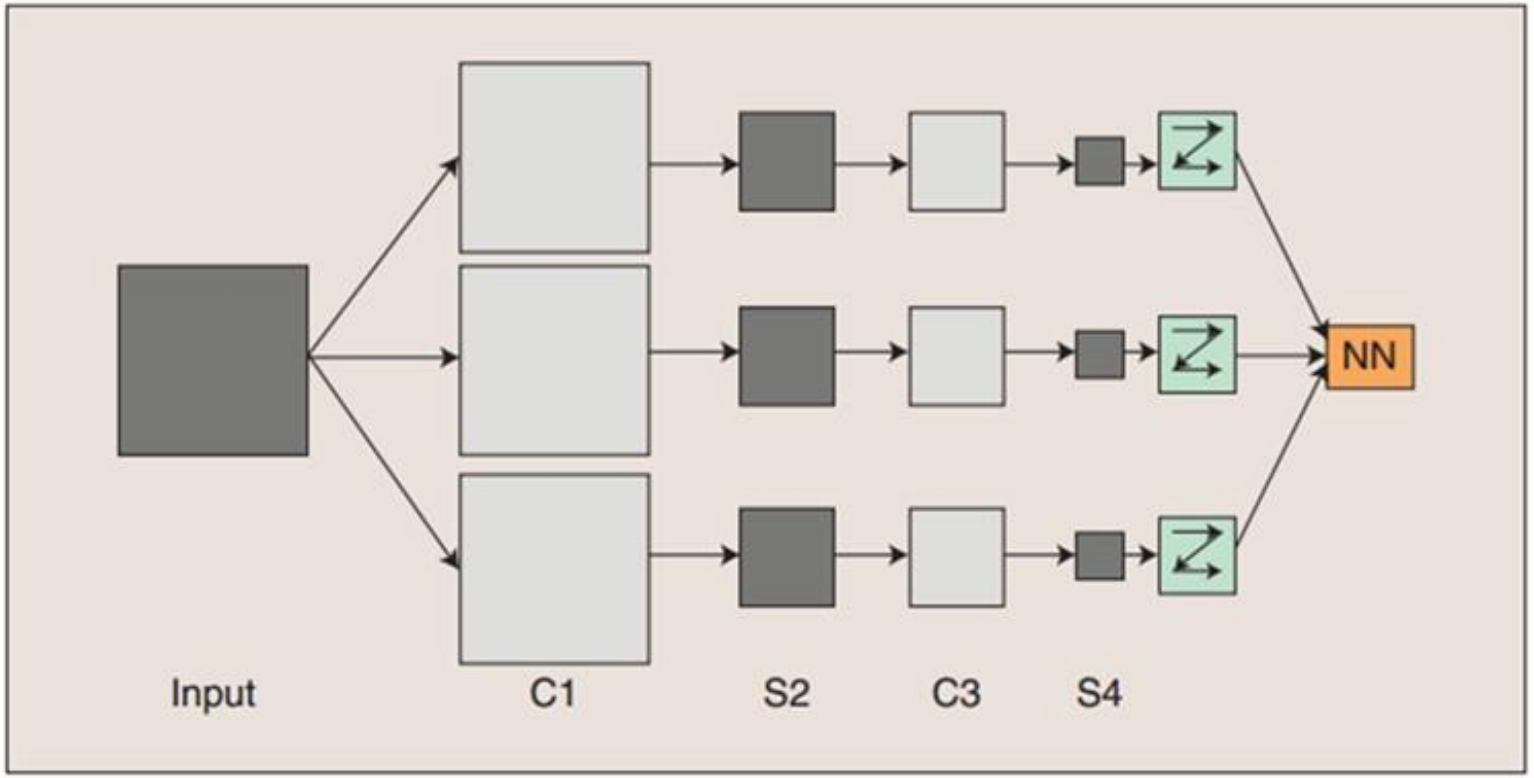
Background - CNN

- CNN – Weights sharing



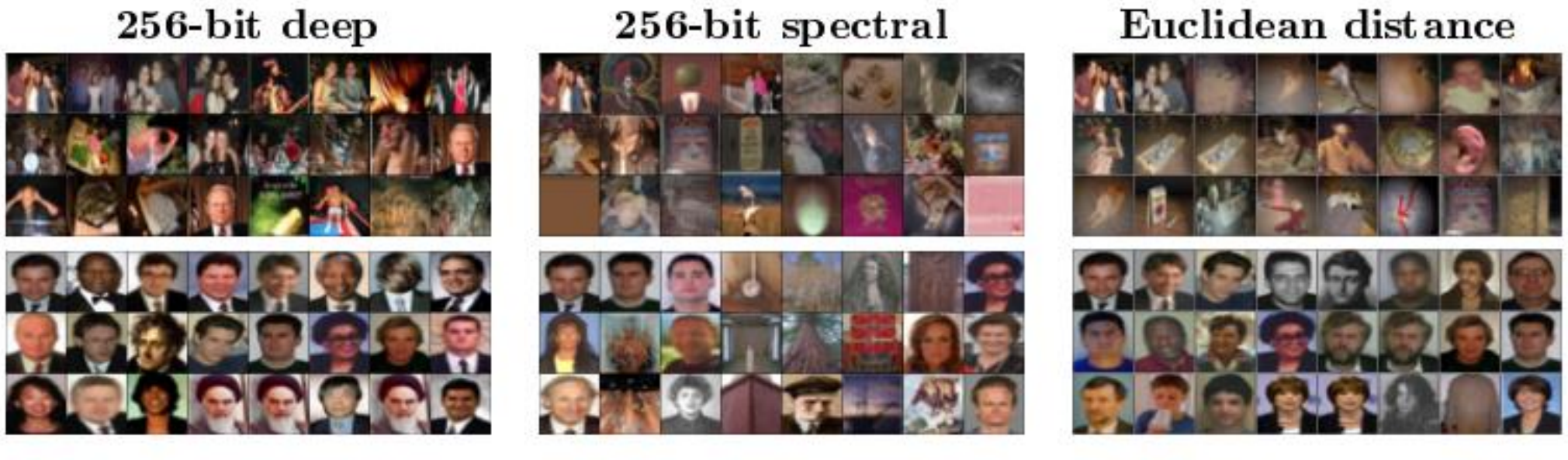
Background - CNN

- CNN – layer structure



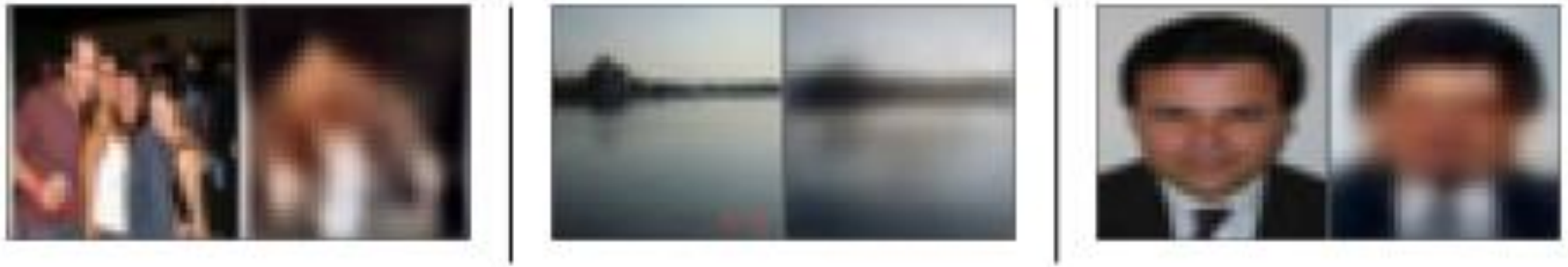
Application – SAE in IR

- Hinton, et al. 2011



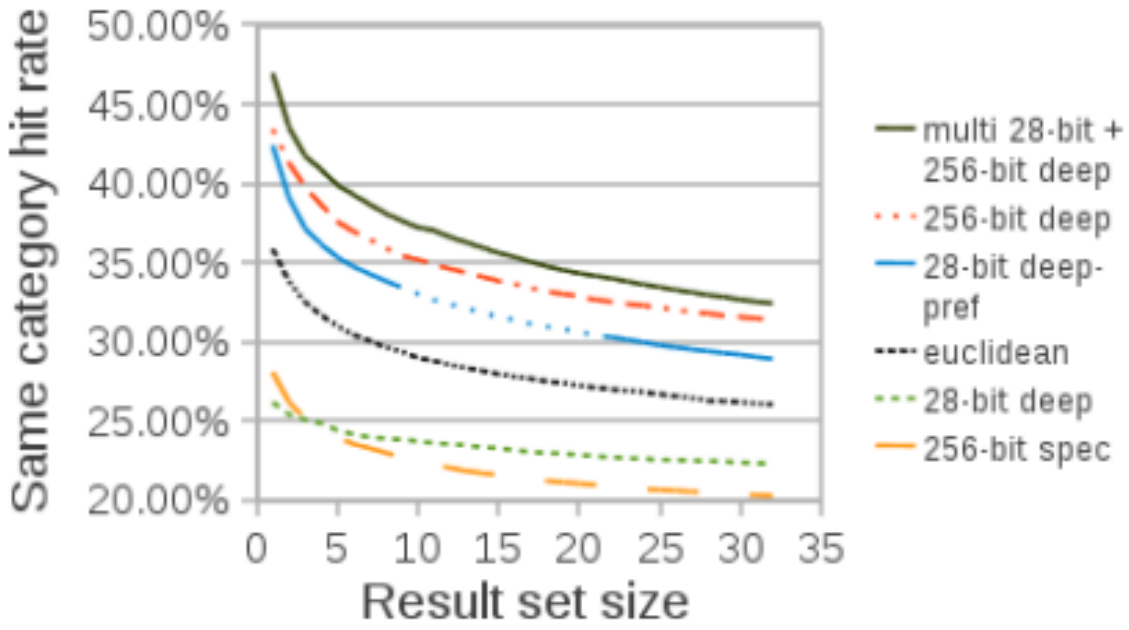
Application – SAE in IR

- Reconstruction of test images



Application – SAE in IR

- Result
- Deep codes >> Euclidean distance >> spectral codes



Application – CNN in Music Retrieval

- Content based music Retrieval/Recommedation
- Extract MFCCs from the audio signals
- Vector quantize the MFCCs
- Aggregate them into a bag-of-words representation
- Reduced the size of this representation using PCA (we kept enough components to retain 95% of the variance)
- Bag-of -words

Application – CNN in Music Retrieval

- aforementioned time-frequency representation as input
- Rectifier linear units

$$f(x) = \max(0, x)$$

Application – CNN in Music Retrieval

- CNN- based methods – ROC curve reach to 0.7+

Model	mAP	AUC
MLR	0.01801	0.60608
linear regression	0.02389	0.63518
MLP	0.02536	0.64611
CNN with MSE	0.05016	0.70987
CNN with WPE	0.04323	0.70101

Conclusion

- Deep Learning learn abstract features from unlabeled data
- Integrate with IR techniques as input space
- Work well with state-of-the-art performance for multimedia retrieval

Thanks!