

ONLINE MASTERS IN **DATA SCIENCE**

DSC 257R - UNSUPERVISED LEARNING

PICKING A GOOD REPRESENTATION

SANJOY DASGUPTA, PROFESSOR

UC San Diego

COMPUTER SCIENCE & ENGINEERING
HALICIOĞLU DATA SCIENCE INSTITUTE






"Solving a problem simply means representing it so as to make the solution transparent."

~Herbert Simon, Sciences of the Artificial


Representations for Text

- Bag of words
 - Latent semantic indexing
 - Brown clustering
 - Topic models
 - Word2Vec and Glove
 - BERT and beyond
- 
- A decorative teal triangle is located in the bottom right corner of the slide.

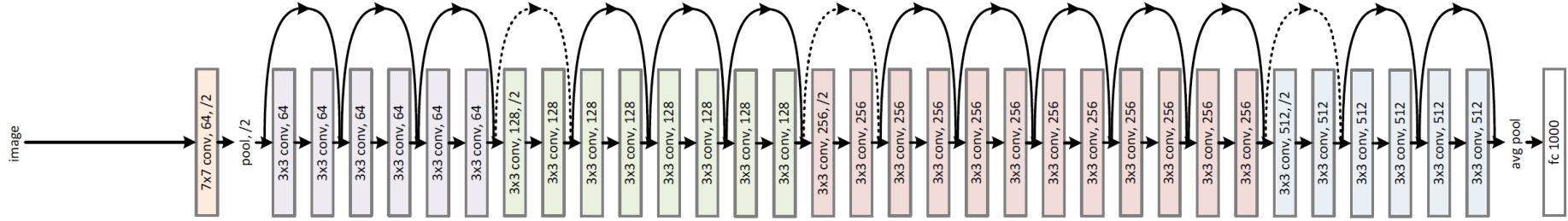
Example: Word2Vec




Representations for Images

- Principal component analysis, for images or image-patches
 - Wavelets
 - Sparse coding
 - SIFT
 - HOG
 - Deep belief nets
 - Self- or fully-supervised deep representations
- 
- A decorative teal triangle is located in the bottom right corner of the slide.

Example: Residual Network (ResNet)



Representations for Other Domains

- Audio: speech, music, animal sounds, etc.
 - Biological sequences: DNA, proteins, etc.
 - And many others.
- 
- A decorative teal triangle is located in the bottom right corner of the slide.