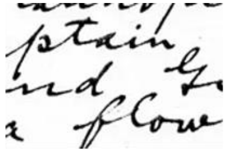
Intro to representational learning

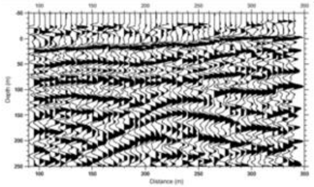
MLRG

February 2019

Data → vector

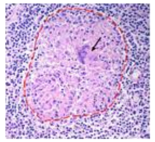


Handwritten scans



Seismic scans

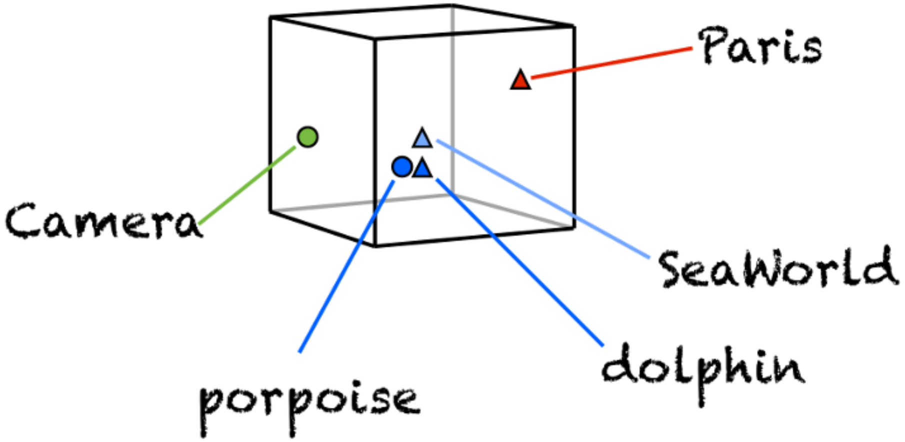
Raw images



Cell imaging

Astronomy

Stock market

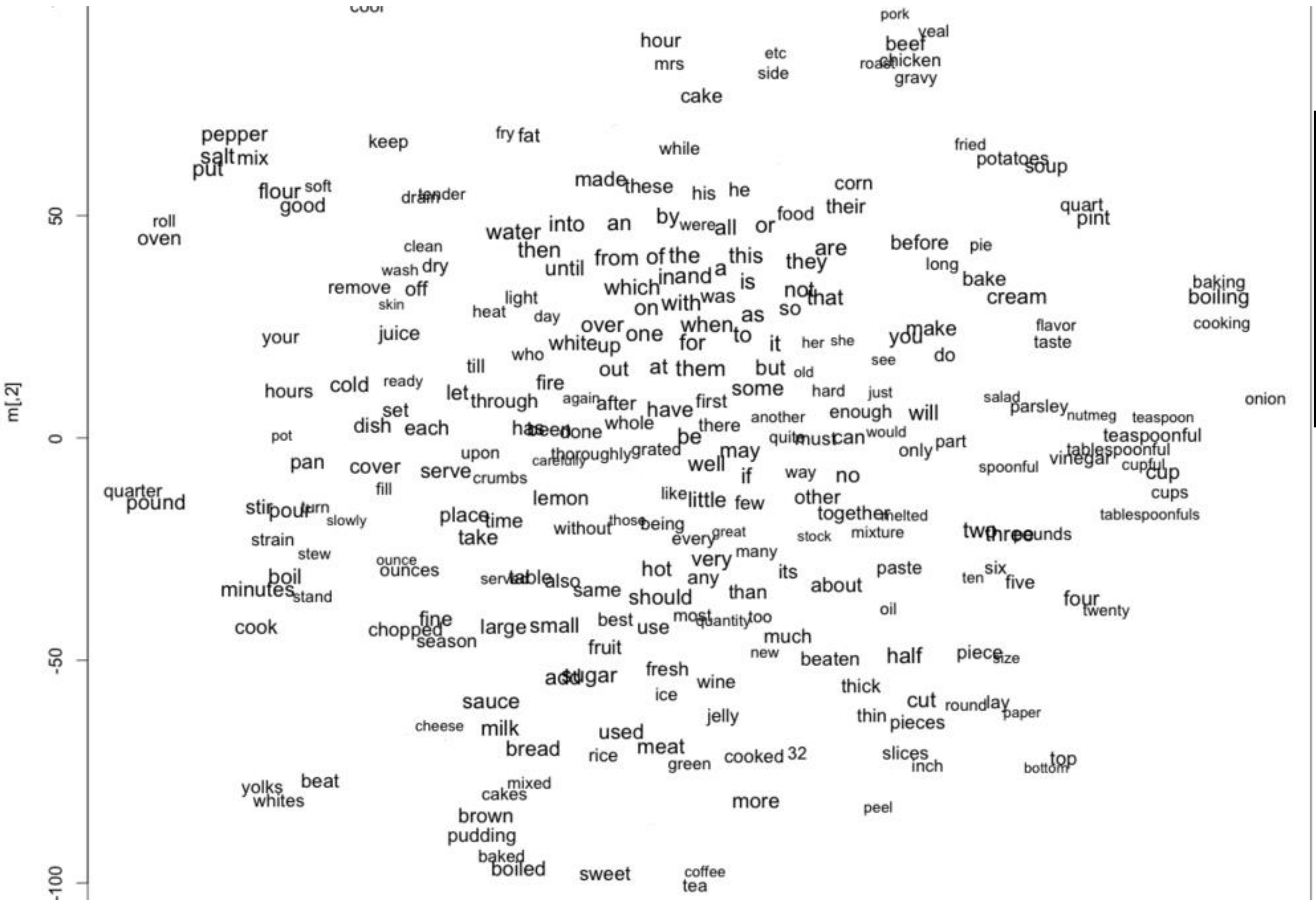


(Quora)

Why representational learning

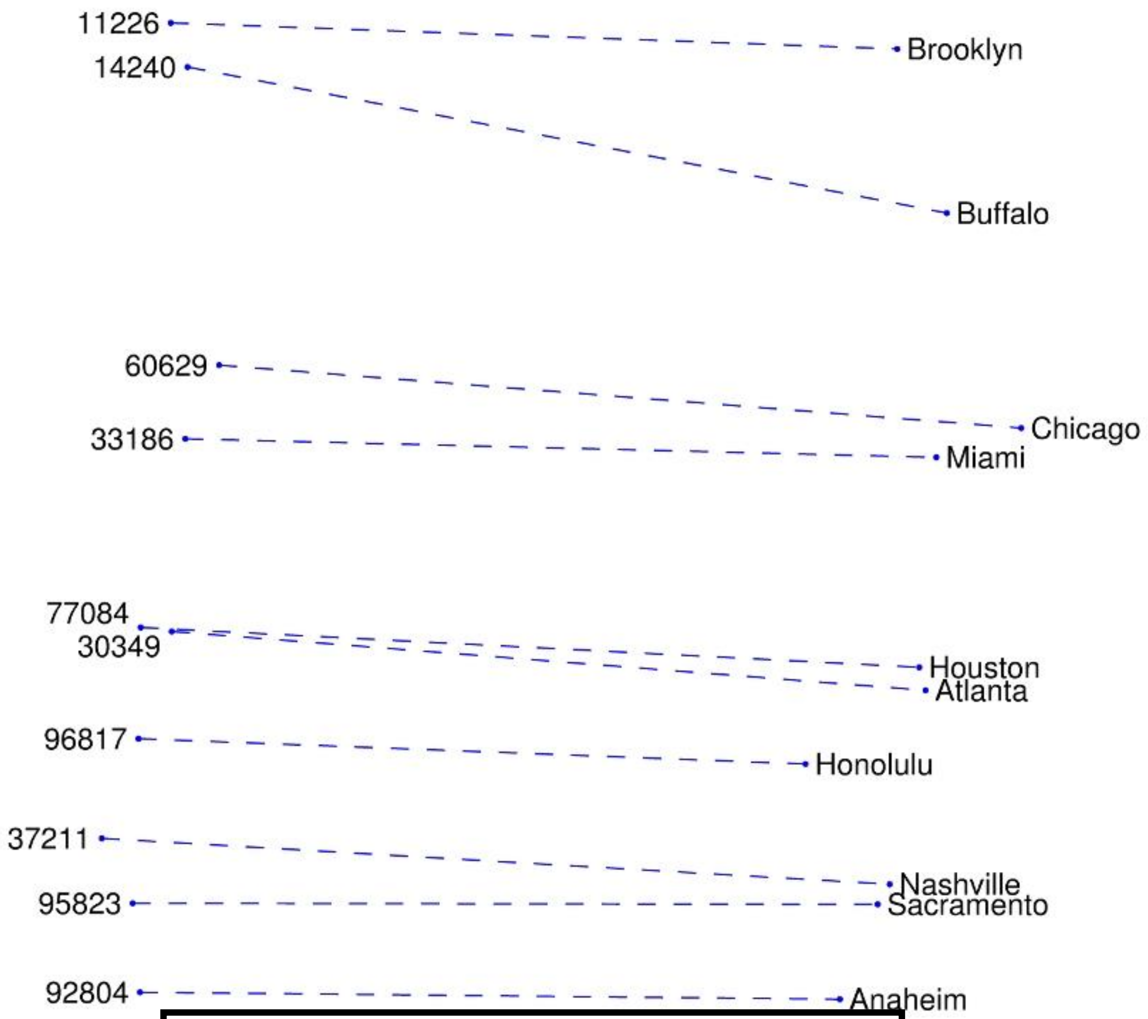
It looks pretty
Supervised learning
Unsupervised learning
Data mining

It looks pretty

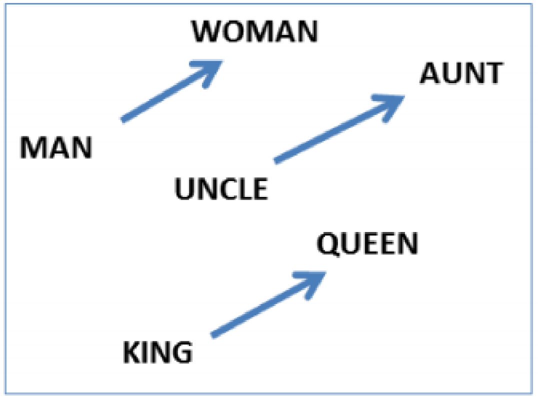


Mapping Word Embeddings with Word2vec:
Enhancing Natural Language Processing with Semantic and Syntactic Relationships between Word Vectors
Sam Liebman (2018)

It looks pretty

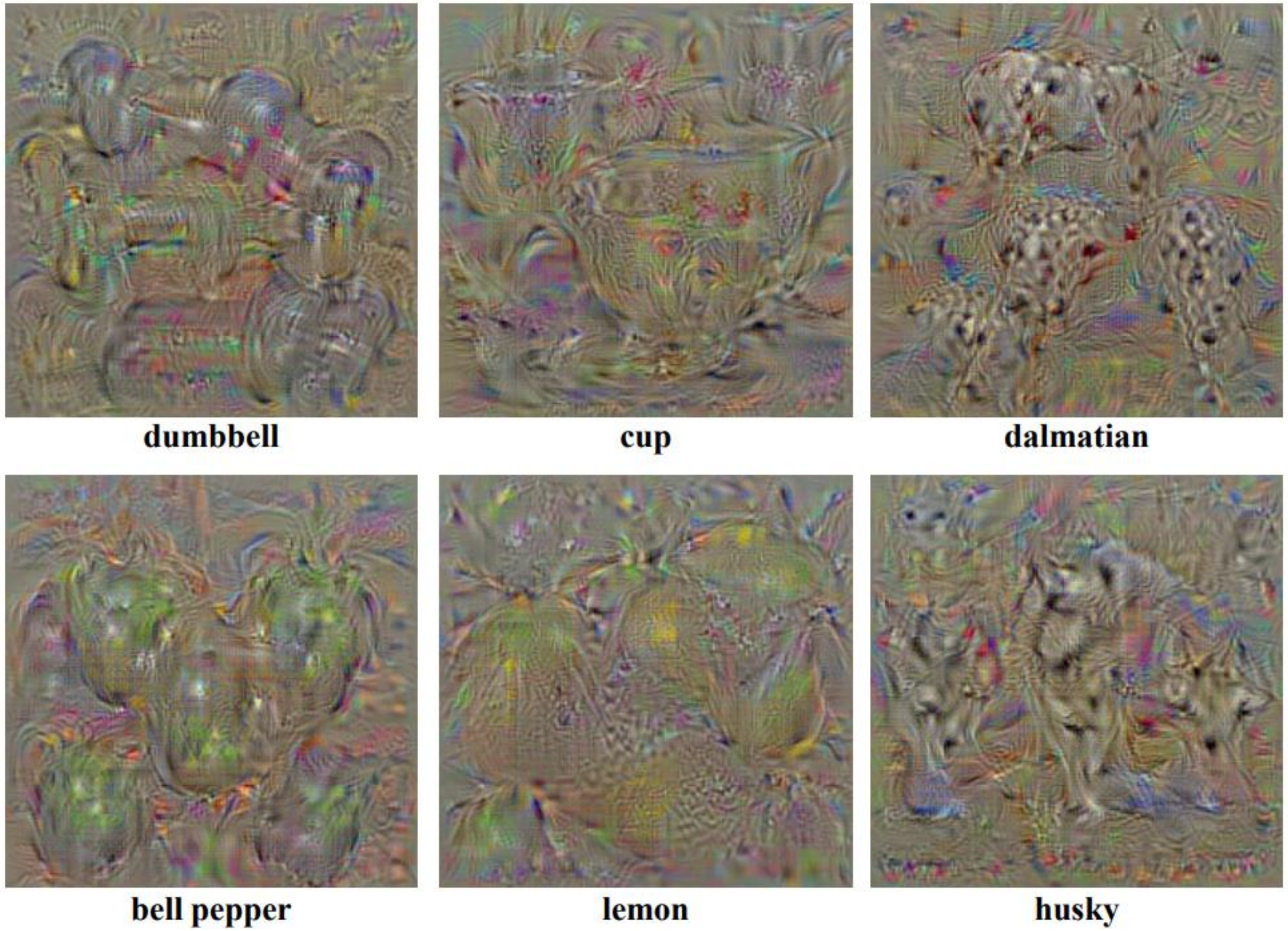


Pennington, Manning, Socher, 2014



(Mikolov et al., NAACL HLT, 2013)

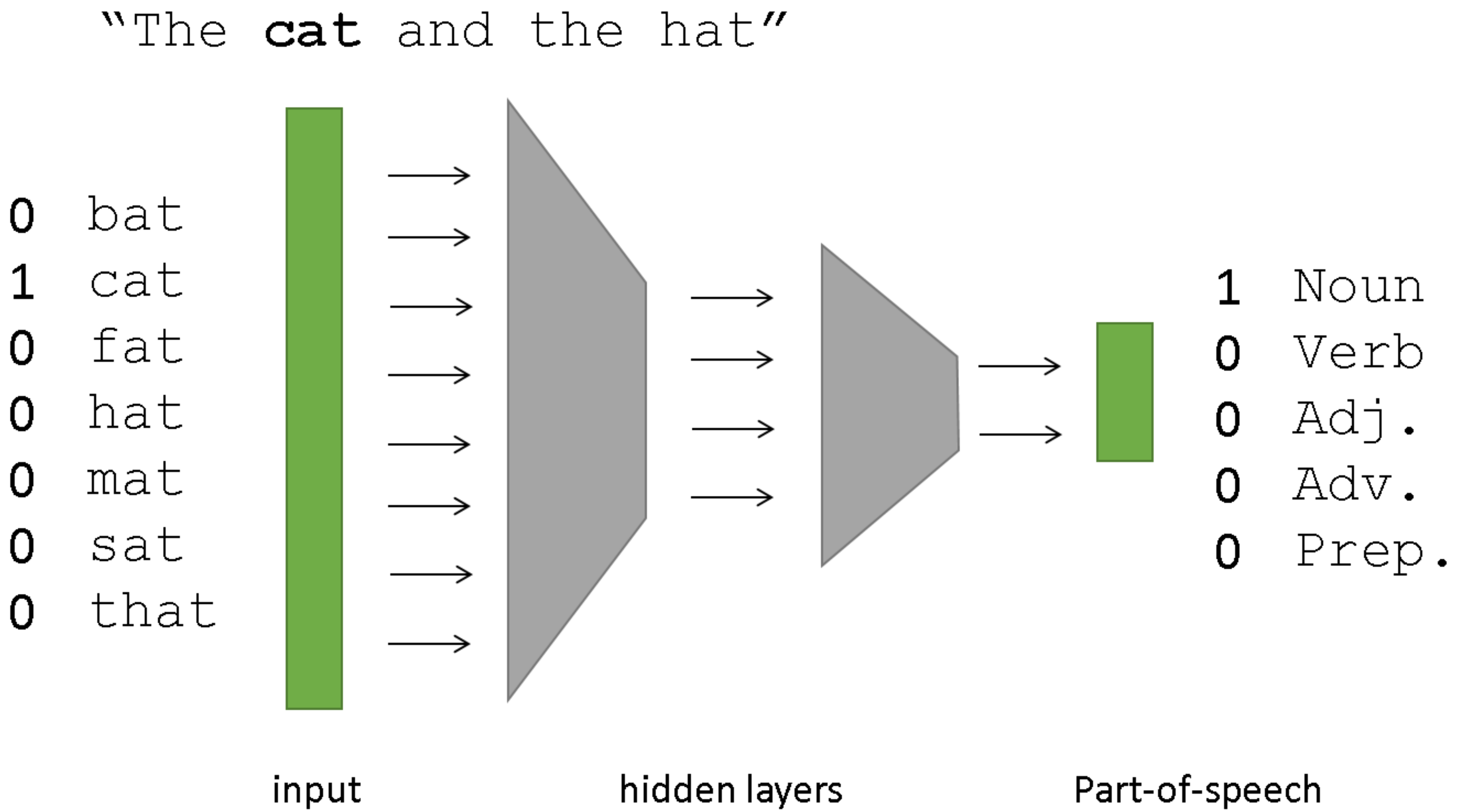
CNN layers



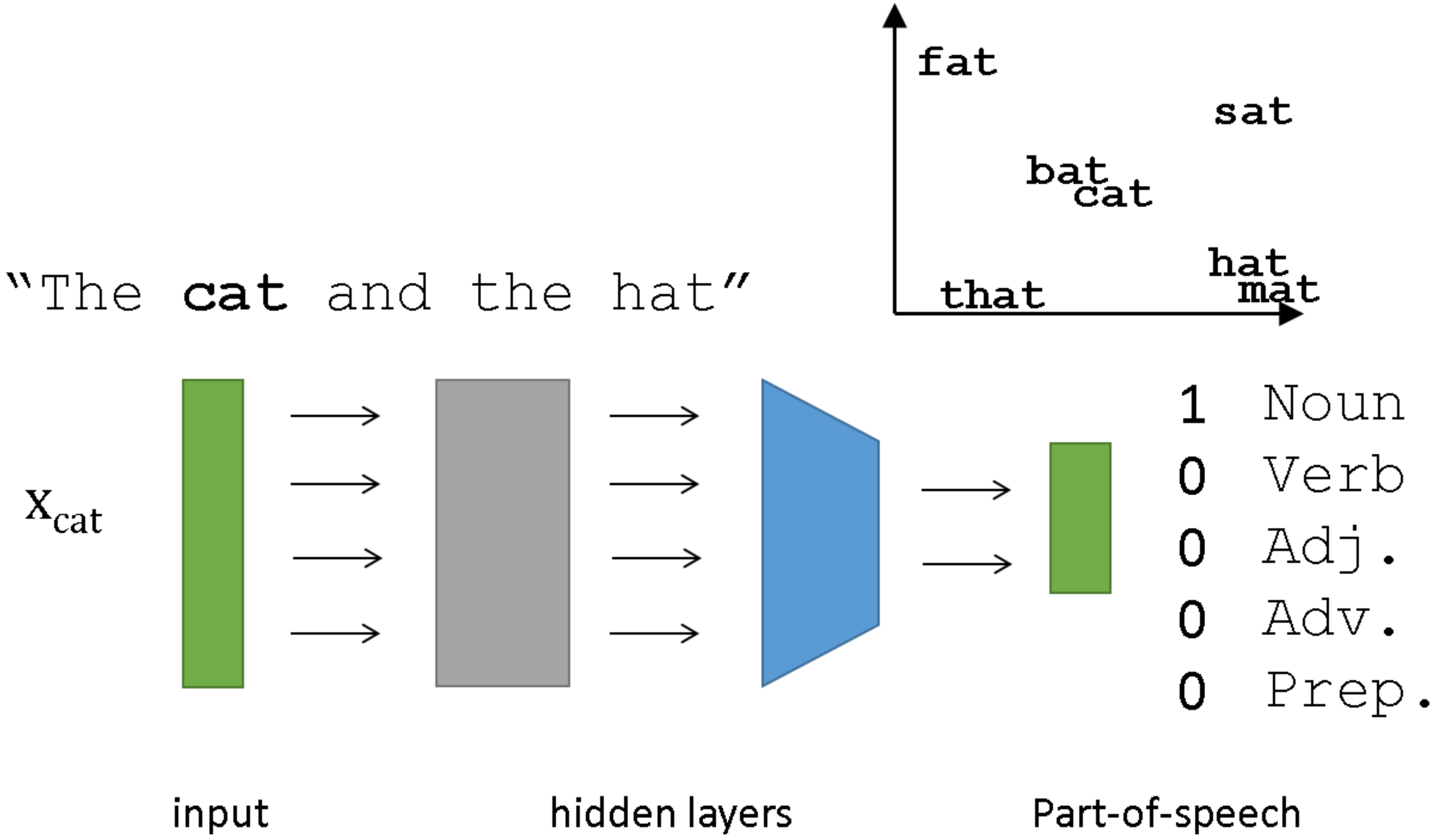
Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps

Simonyan, Vedaldi, Zisserman 2014

Supervised learning



Supervised learning



NLP named entity recognition

System	Dev	Test	MUC7
Baseline	90.03	84.39	67.48
Baseline+Nonlocal	91.91	86.52	71.80
HLBL 100-dim	92.00	88.13	75.25
Gazetteers	92.09	87.36	77.76
C&W 50-dim	92.27	87.93	75.74
Brown, 1000 clusters	92.32	**88.52**	**78.84**
C&W 200-dim	**92.46**	87.96	75.51
C&W+HLBL	92.52	88.56	78.64
Brown+HLBL	92.56	88.93	77.85
Brown+C&W	92.79	89.31	80.13
HLBL+Gaz	92.91	89.35	79.29
C&W+Gaz	92.98	88.88	81.44
Brown+Gaz	**93.25**	**89.41**	**82.71**
Lin and Wu (2009), 3.4B	-	88.44	-
Ando and Zhang (2005), 27M	93.15	89.31	-
Suzuki and Isozaki (2008), 37M	93.66	89.36	-
Suzuki and Isozaki (2008), 1B	**94.48**	89.92	-
All (Brown+C&W+HLBL+Gaz), 37M	93.17	90.04	82.50
All+Nonlocal, 37M	93.95	90.36	84.15
Lin and Wu (2009), 700B	-	**90.90**	-

Word representations: A simple and general method for semi-supervised learning

Turian, Ratinov, Bengio, 2010

Table 3: Final NER F1 results, showing the cumulative effect of adding word representations, non-local features, and gazetteers to the baseline. To speed up training, in combined experiments (C&W plus another word representation), we used the 50-dimensional C&W embeddings, not the 200-dimensional ones. In the last section, we show how many unlabeled words were used.

Clustering

Importance Weighted and
Adversarial Autoencoders
Kinsella (2017)

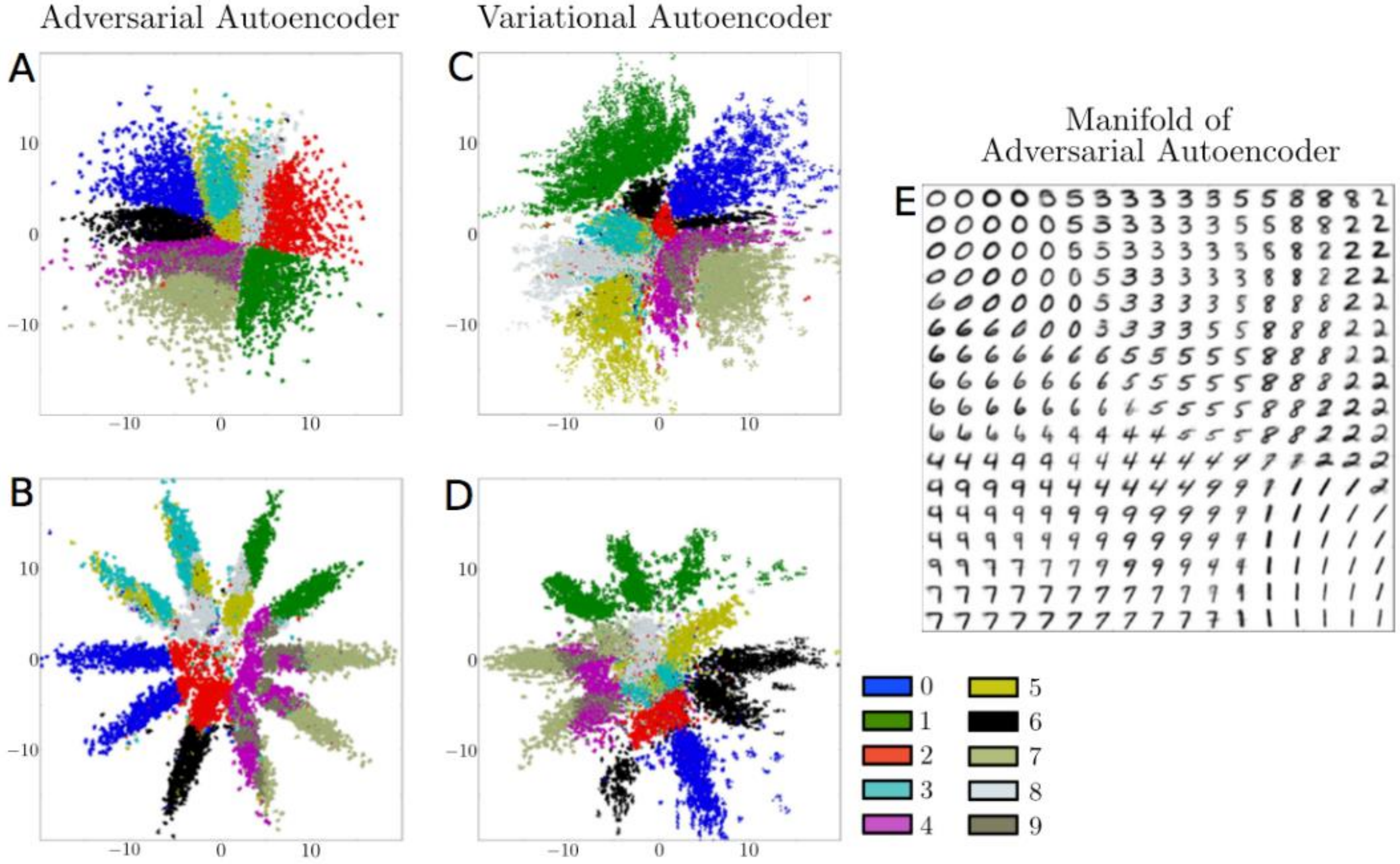


Figure 2: Comparison of adversarial and variational autoencoder on MNIST. The hidden code **z** of the *hold-out* images for an adversarial autoencoder fit to (a) a 2-D Gaussian and (b) a mixture of 10 2-D Gaussians. Each color represents the associated label. Same for variational autoencoder with (c) a 2-D gaussian and (d) a mixture of 10 2-D Gaussians. (e) Images generated by uniformly sampling the Gaussian percentiles along each hidden code dimension **z** in the 2-D Gaussian adversarial autoencoder.

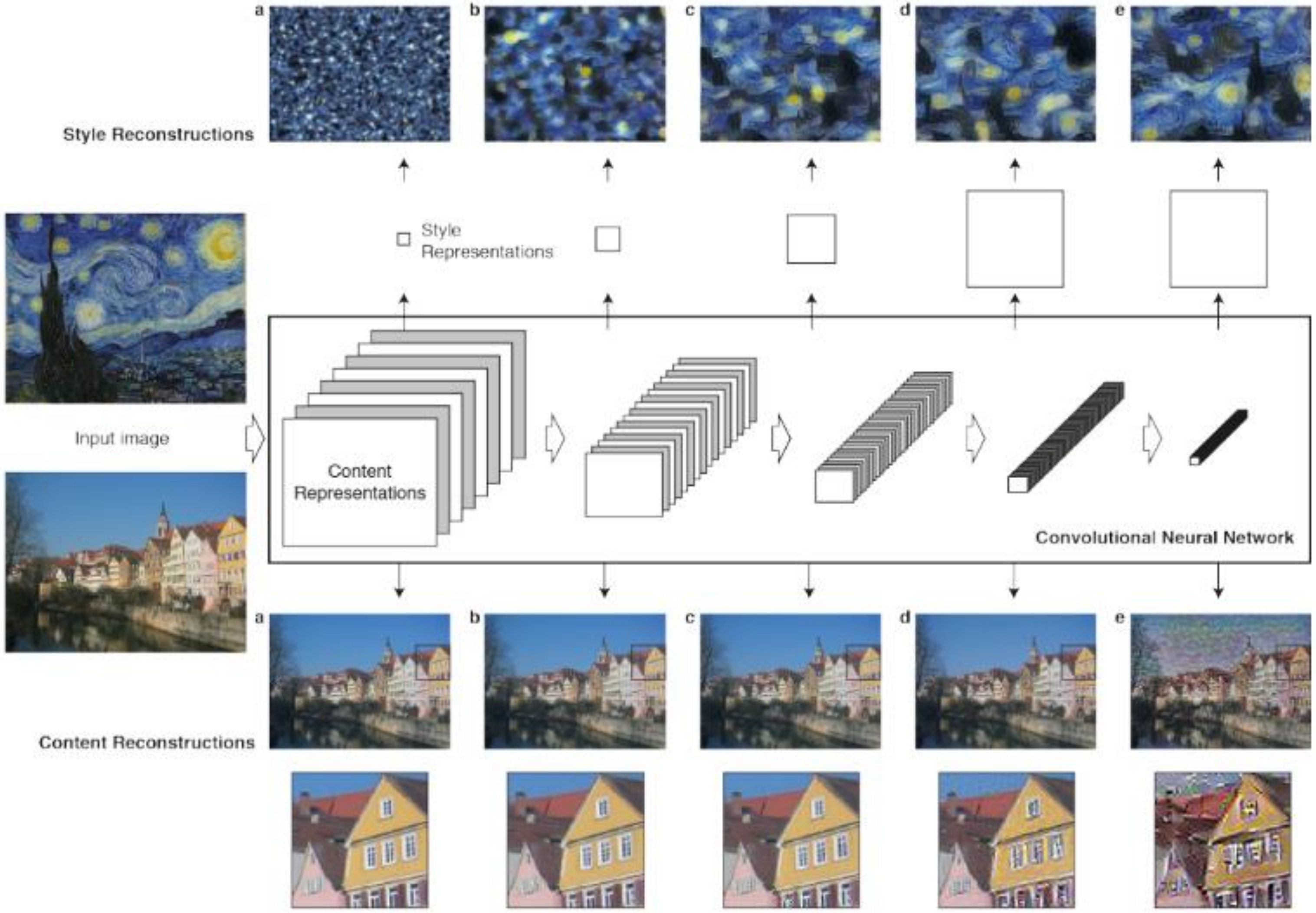
Disambiguation and style transfer

A Neural Algorithm of Artistic Style

Gatys, Ecker, and Bethge, 2015

Artistic Style Transfer

Doukkali, 2018



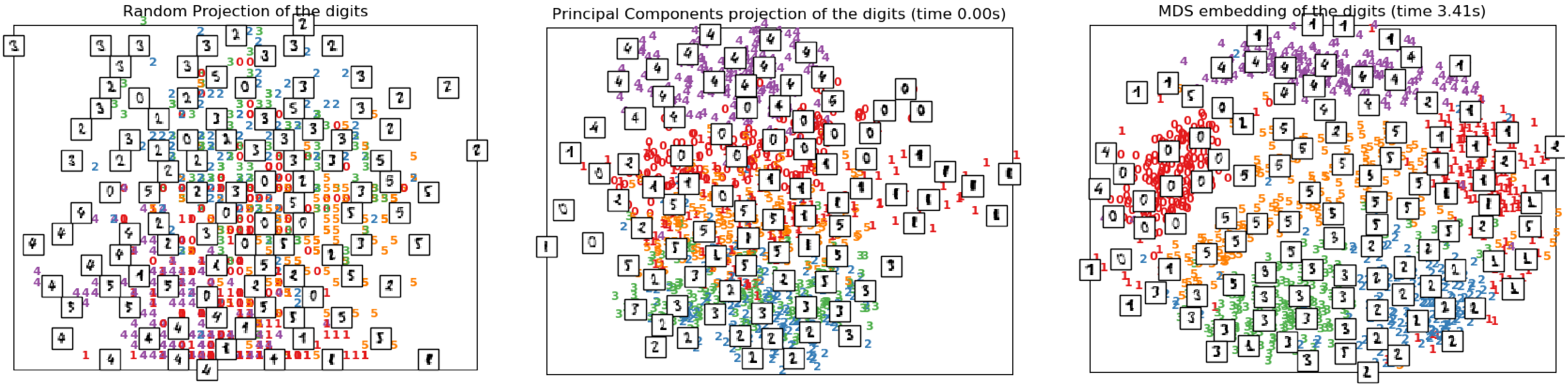
Past methods

Spectral methods

SIFT / HOG

An SDP approach

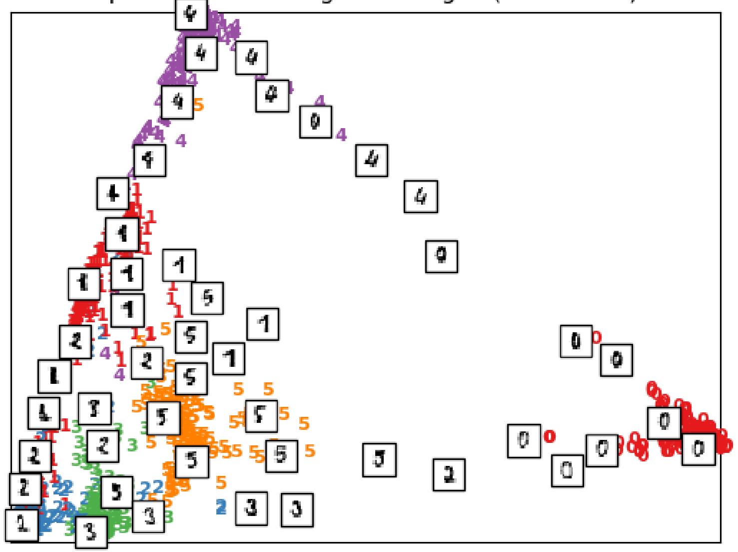
Also called manifold learning



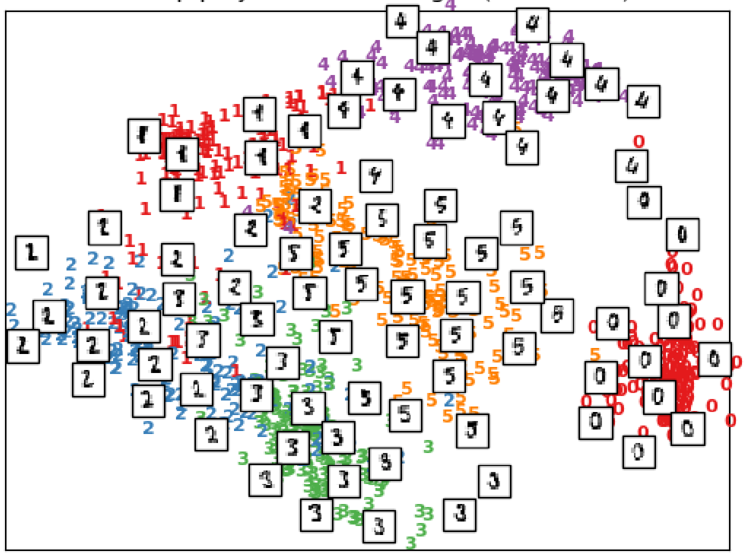
https://scikit-learn.org/stable/auto_examples/manifold/plot_lle_digits.html

Also called manifold learning

Spectral embedding of the digits (time 0.33s)



Isomap projection of the digits (time 0.90s)

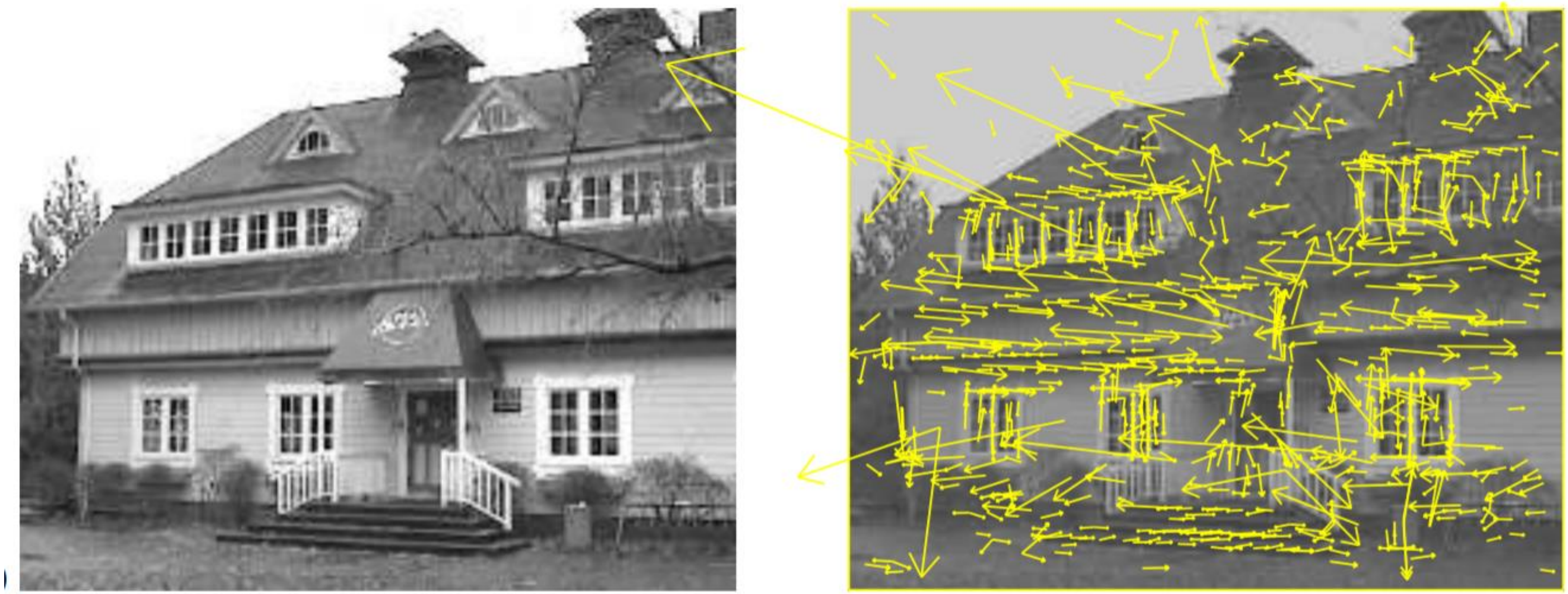


https://scikit-learn.org/stable/auto_examples/manifold/plot_lle_digits.html

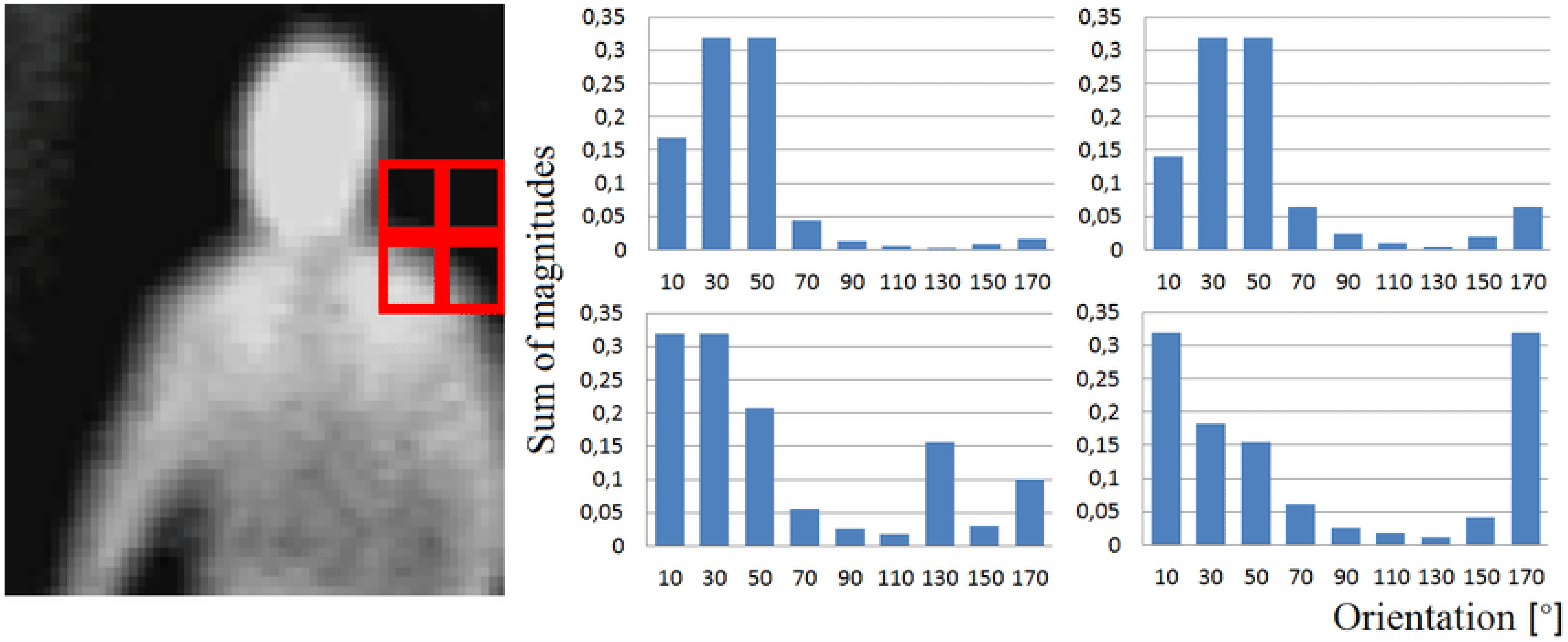
Scale Invariant Feature Transform (SIFT)

Distinctive Image Features from
Scale-Invariant Keypoints

Lowe, 2004



Histogram of oriented gradients



Method of and apparatus for pattern recognition

McConnell 1986

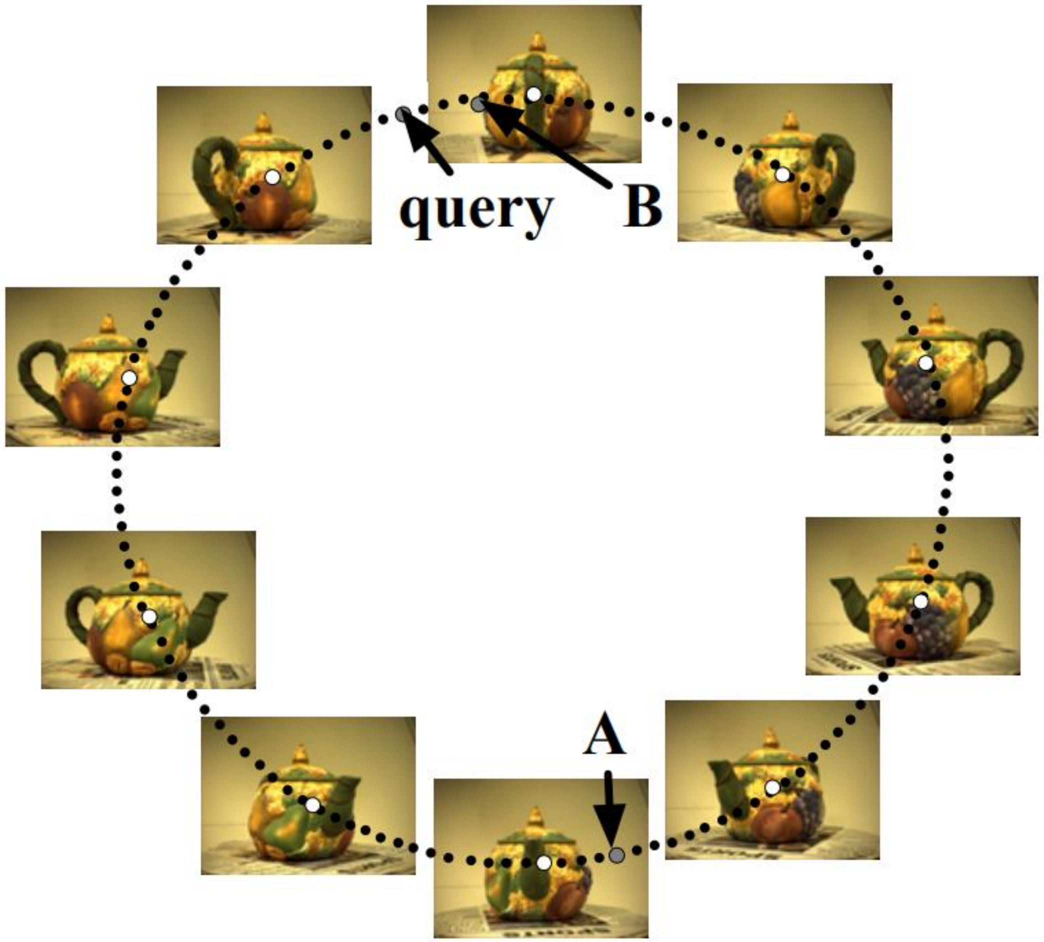
Histograms of Oriented Gradients for Human Detection

Dalal and Triggs 2005

Video Processing Algorithms for Detection of Pedestrians

Piniarski, Pawlowski, Dabrowski, 2015

Maximum variance unfolding (SDP approach)



An Introduction to Nonlinear Dimensionality Reduction by Maximum Variance Unfolding

Weinberger and Saul 2006

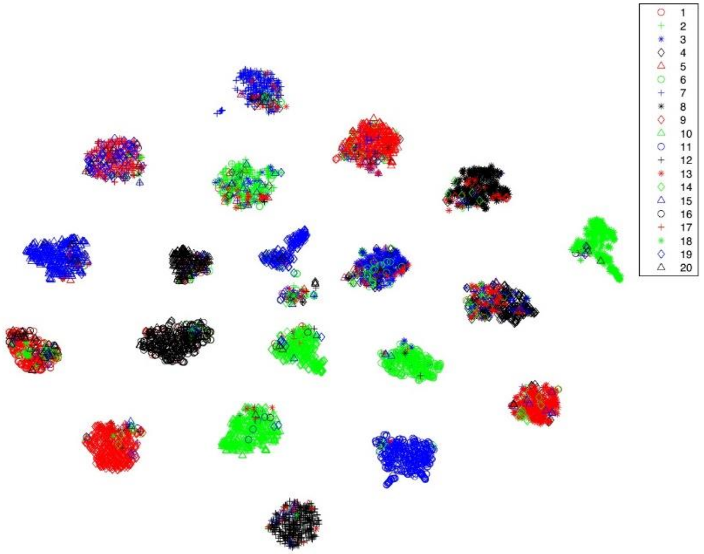
Neural network methods

t-SNE
Word2vec
Transfer learning
Autoencoders

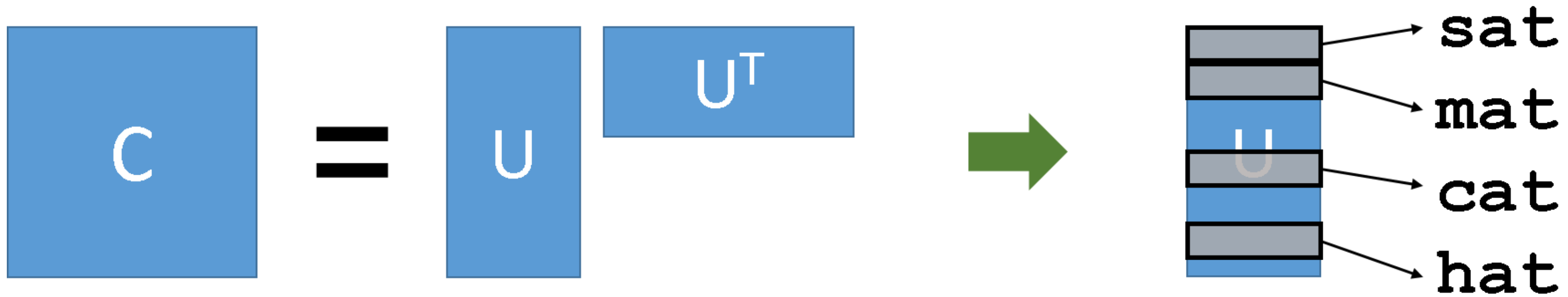
t-Distributed Stochastic Neighbor Embedding (t-SNE)

Visualizing High-Dimensional Data Using t-SNE

van der Maaten and Hinton 2008

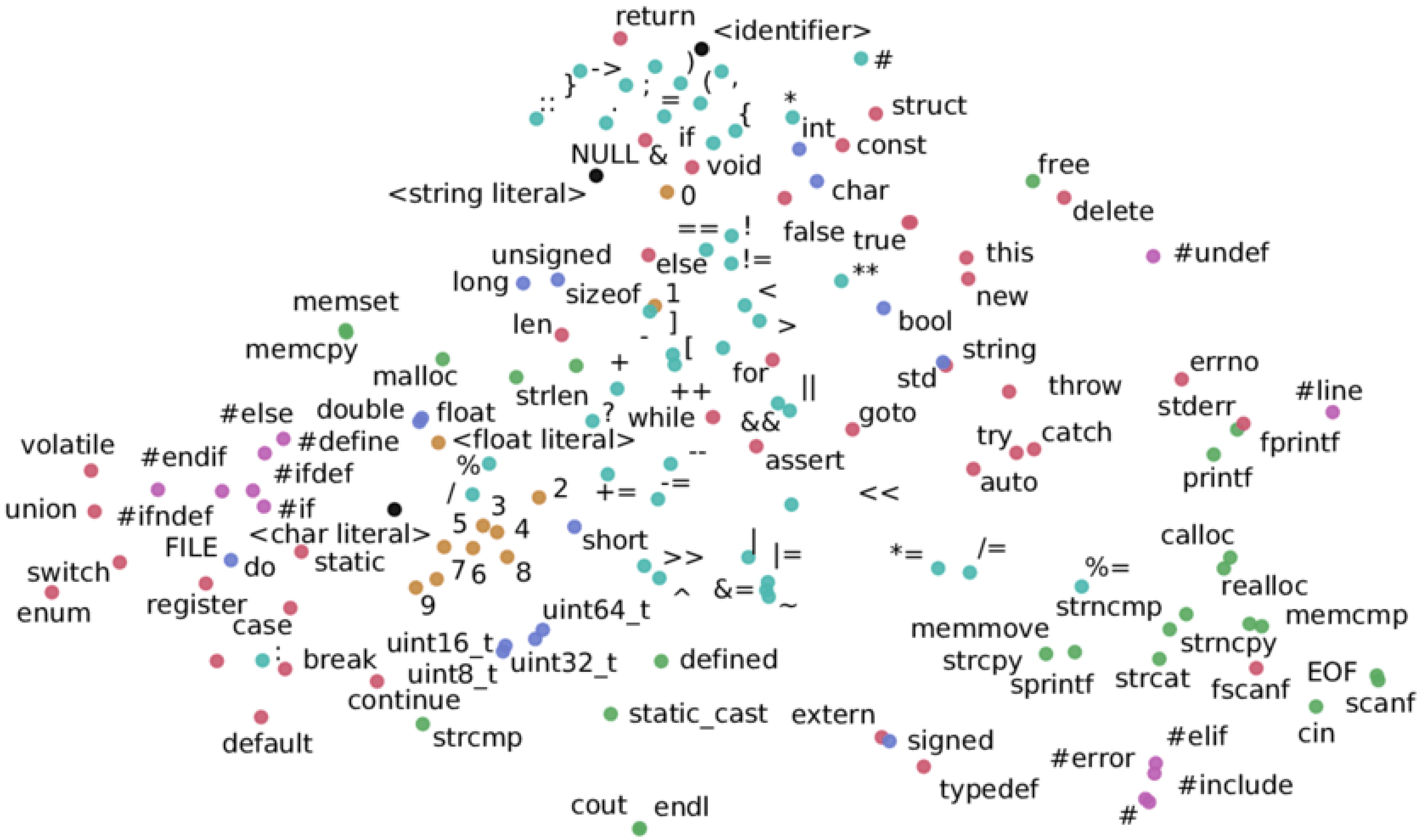


Word embeddings



- Low-rank SVD ...
  - ... on document-word counts (LSI, Deerwester et al 1990)
  - ... on cooccurance matrix (HAL, Lund / Burgess 1996)
  - ... on PPMI matrix (Bullinaria / Levy 2007)
- Shallow neural network
  - Bengio 2003, Collobert/Weston 2008, Mikolov et al 2013 (w2v)
- Least squares
  - GloVe (Pennington et al 2014)
- ... many unlisted ...

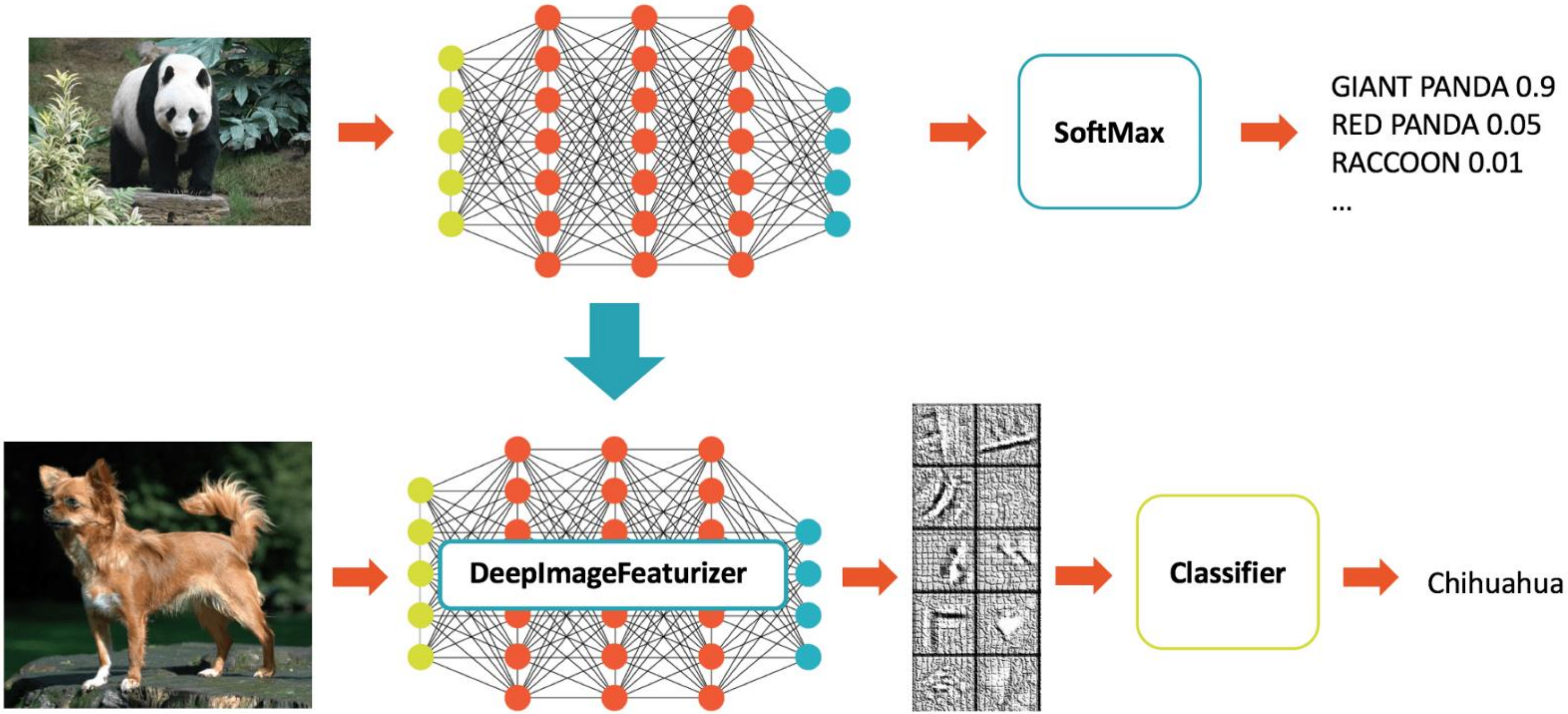
Similar technique on other data



Harer et al 2018

Transfer learning

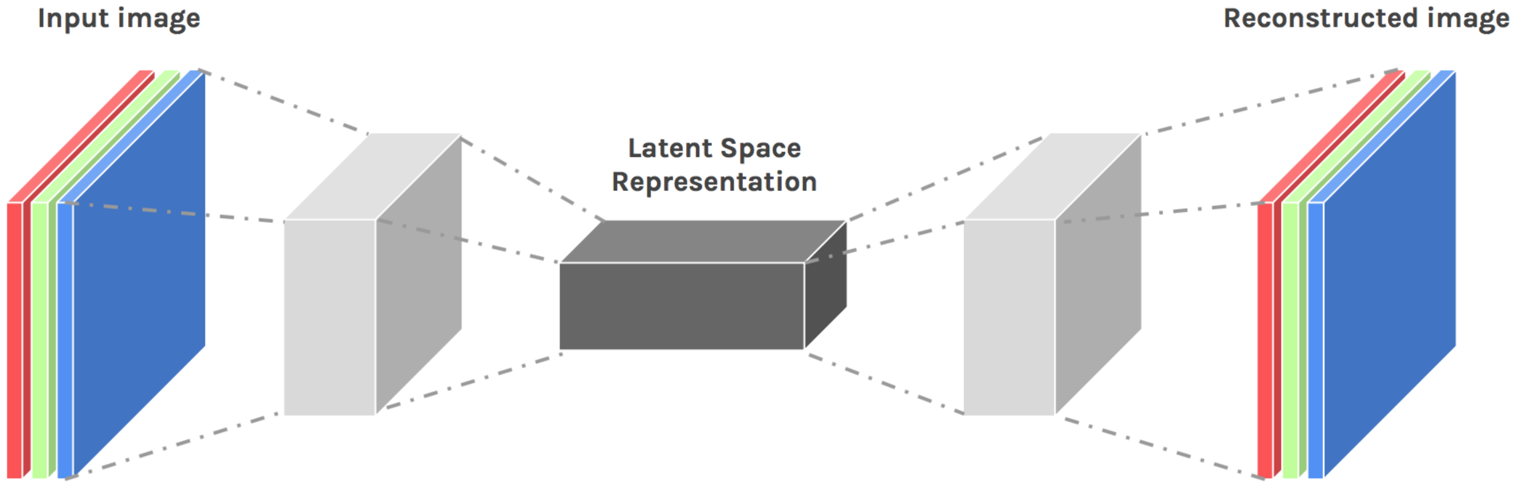
Transfer Learning for Natural Language Processing
Radhakrishnan, 2018

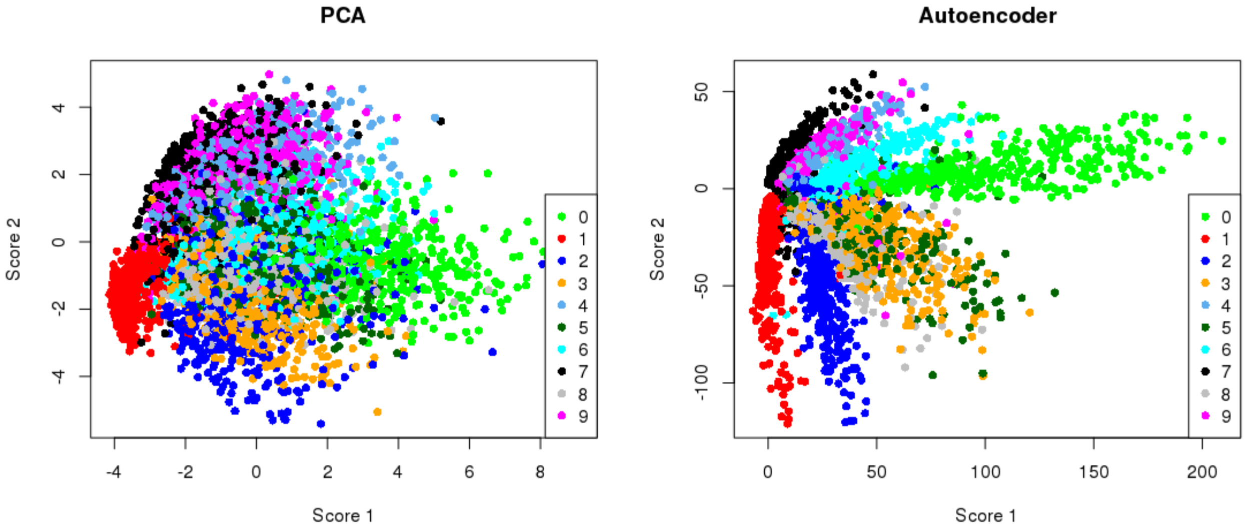


Autoencoders

Understanding Autoencoders -- Unsupervised Learning Technique

Mishra, 2018





Recommendation systems with deep autoencoders

Kashikar

Speakers

Week	Topic	Speaker
Feb 4	Intro	Yifan
Feb 11	Artistic style transfer	amir
Feb 18 (Holiday)		
Feb 25	GANS	aaron
Mar 4	autoencoders	vaden
Mar 11	Convolutional graph embeddings	cathy
Mar 18	Manifold learning	wilder
Mar 25	Graph and point cloud embeddings	marjan
Apr 1	Metric learning	mehrdad
Apr 8	disentanglement	Michael p
Apr 15 (AISTATS)	Dictionary learning	yihan

Topic Options:

Sparse coding / dictionary learning
Disentanglement
Convolutional graph embeddings
Style transfer
T-SNE
Variational autoencoders
GANs
Other types of embeddings (graphs, point clouds, video…)
Metric learning (learning similarities)
<other>