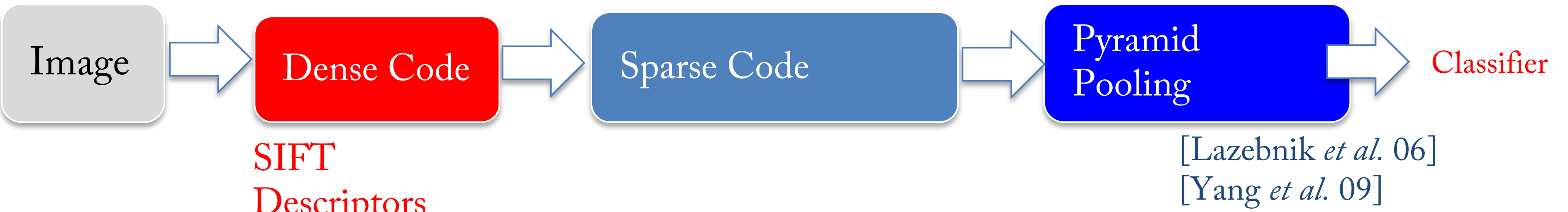
Unsupervised Feature Learning by Deep Sparse Coding

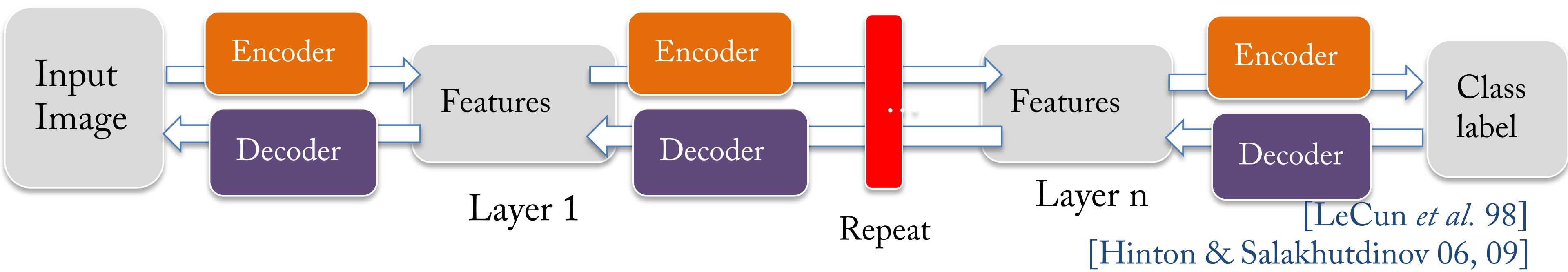
Yunlong He, Georgia Tech
Koray Kavukcuoglu, DeepMind Technologies
Yun Wang, Princeton University
Arthur Szlam, The City College of New York
Yanjun Qi, University of Virginia

Two lines of Previous Research

- Bag of Visual words

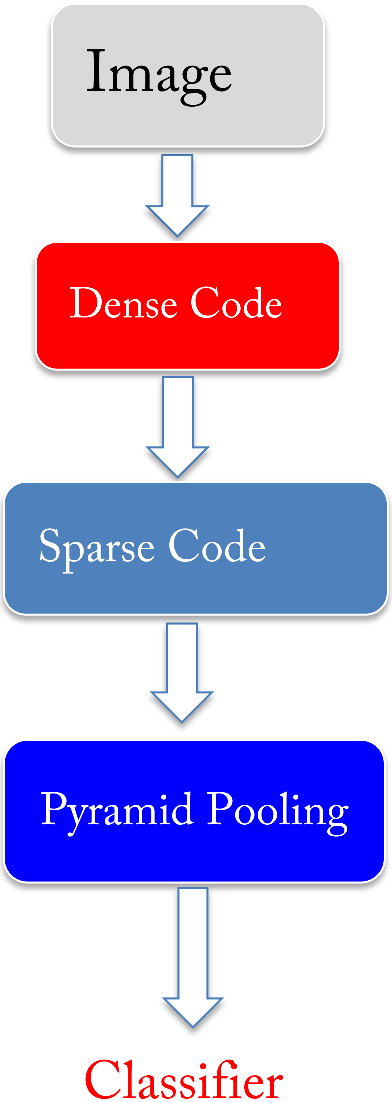


- Deep learning



- Motivation: Can we combine the power of two methods?

Background: Bag of Visual words

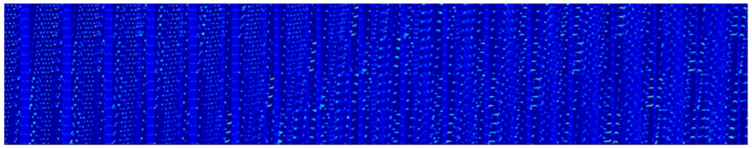


Background: Bag of Visual words

Image

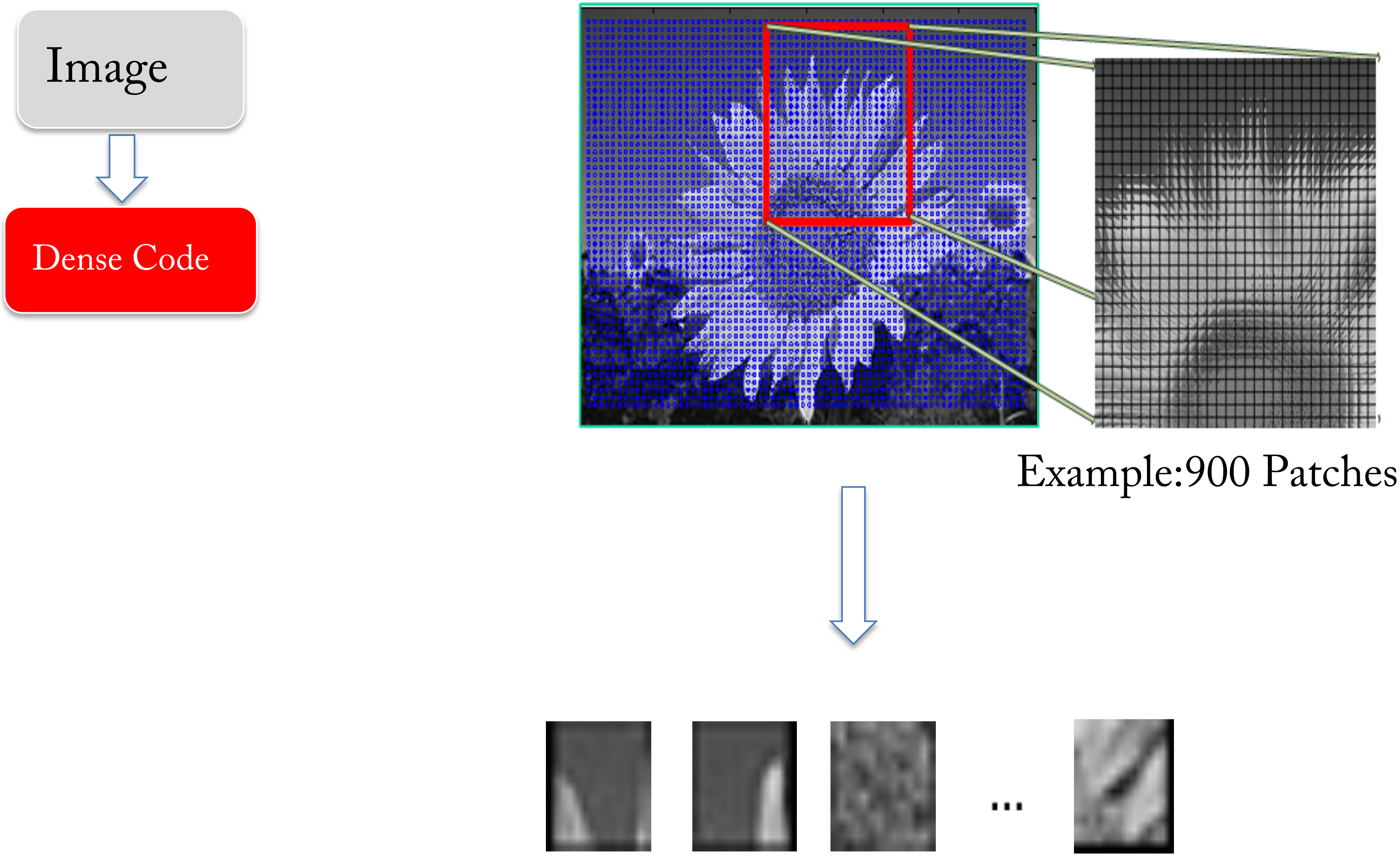
Dense Code

Image



SIFT Descriptors

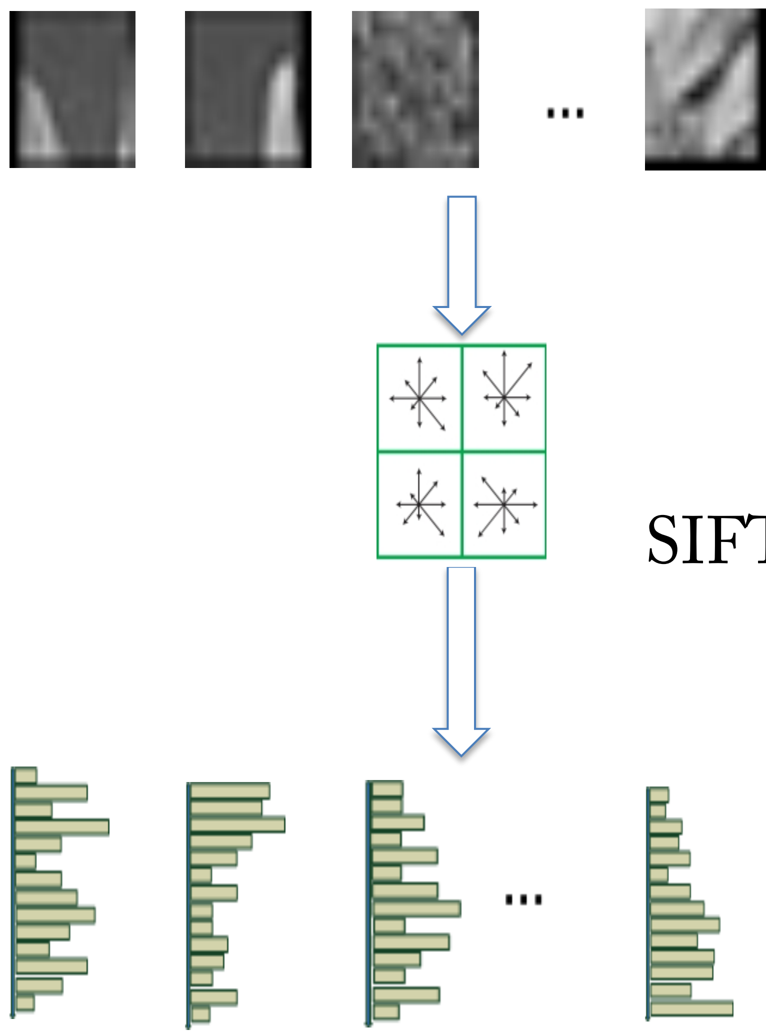
Background: Bag of Visual words



Background: Bag of Visual words

Image

Dense Code

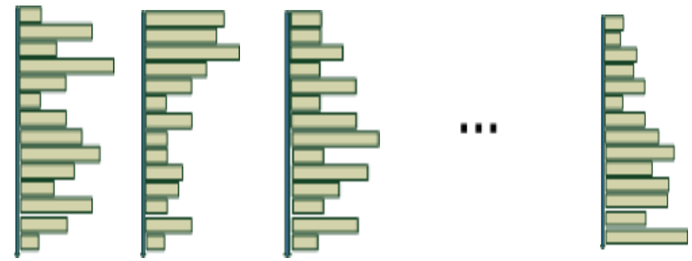


SIFT filter

SIFT Descriptors for every patch

Background: Bag of Visual words

SIFT Descriptors
$x \in R^{128}$, dense



Dense Code

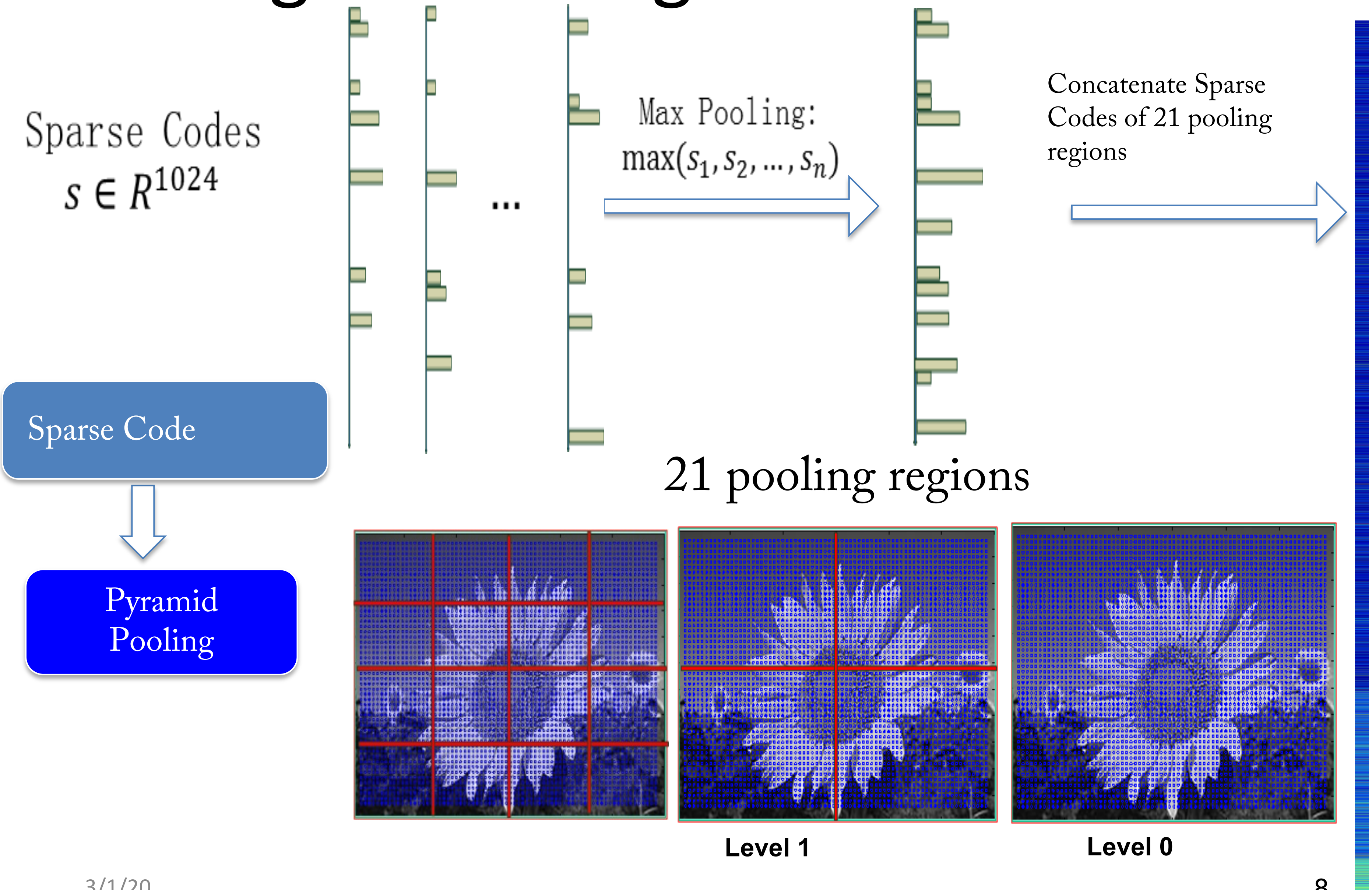
Sparse Code

$$s = argmin_s ||x - Bs||^2 + \lambda ||s||_1$$

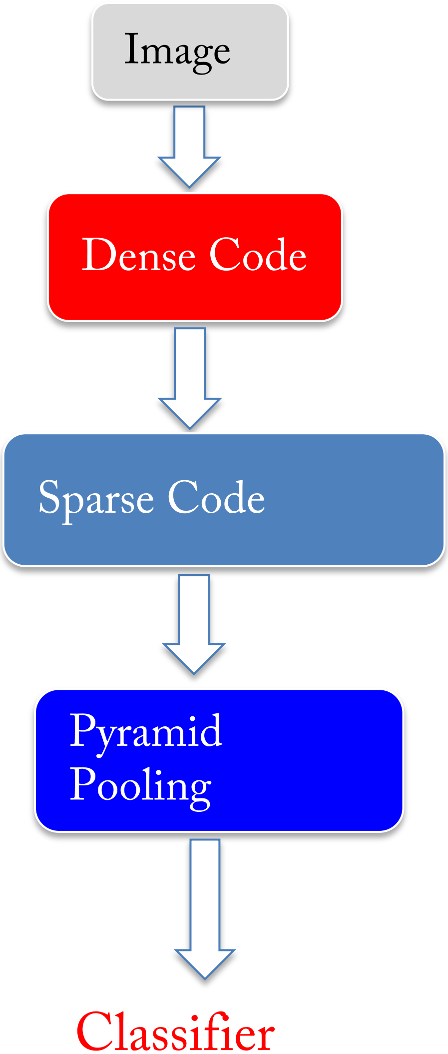
$B \in R^{128\times1024}$

Sparse Codes
$s \in R^{1024}$

Background: Bag of Visual words



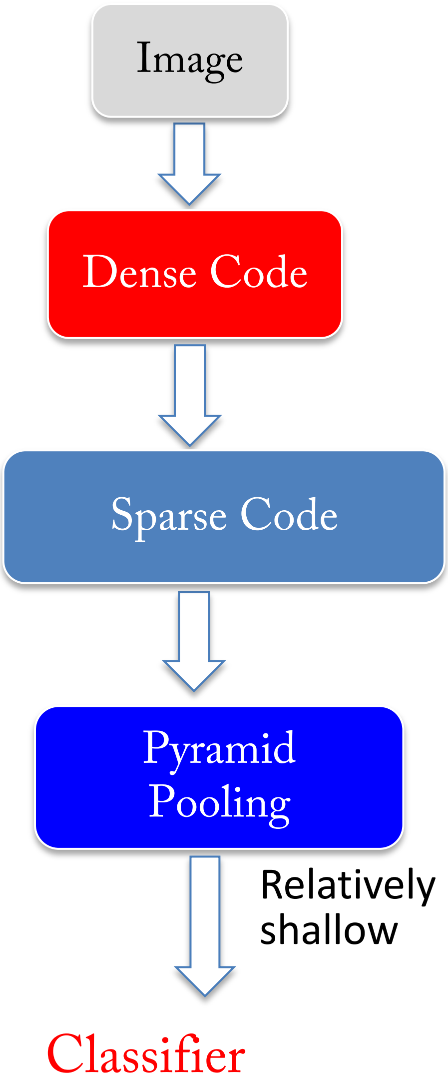
Background: Bag of Visual Words

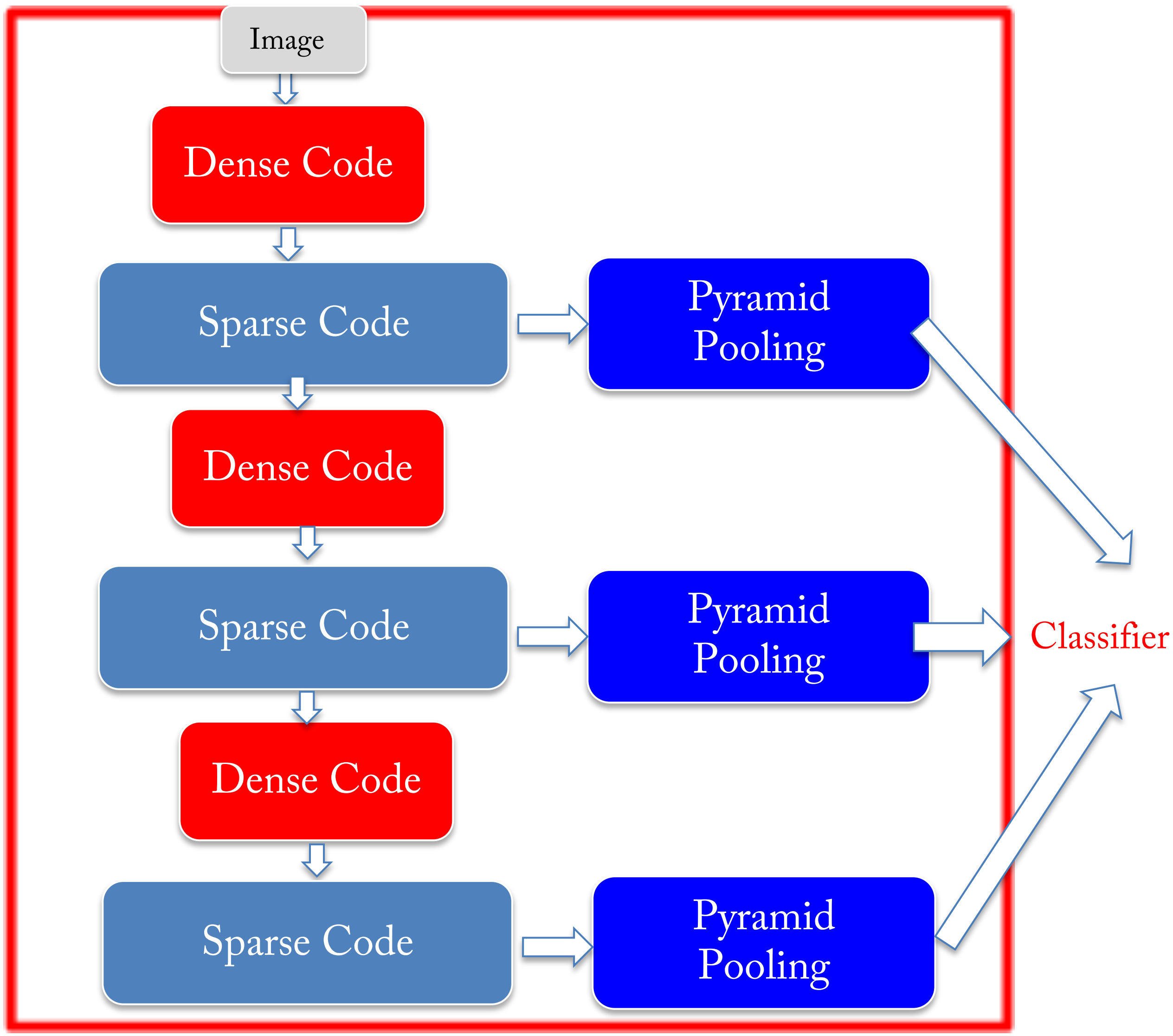


Relatively shallow

Can we learn multiple layers of sparse representations?

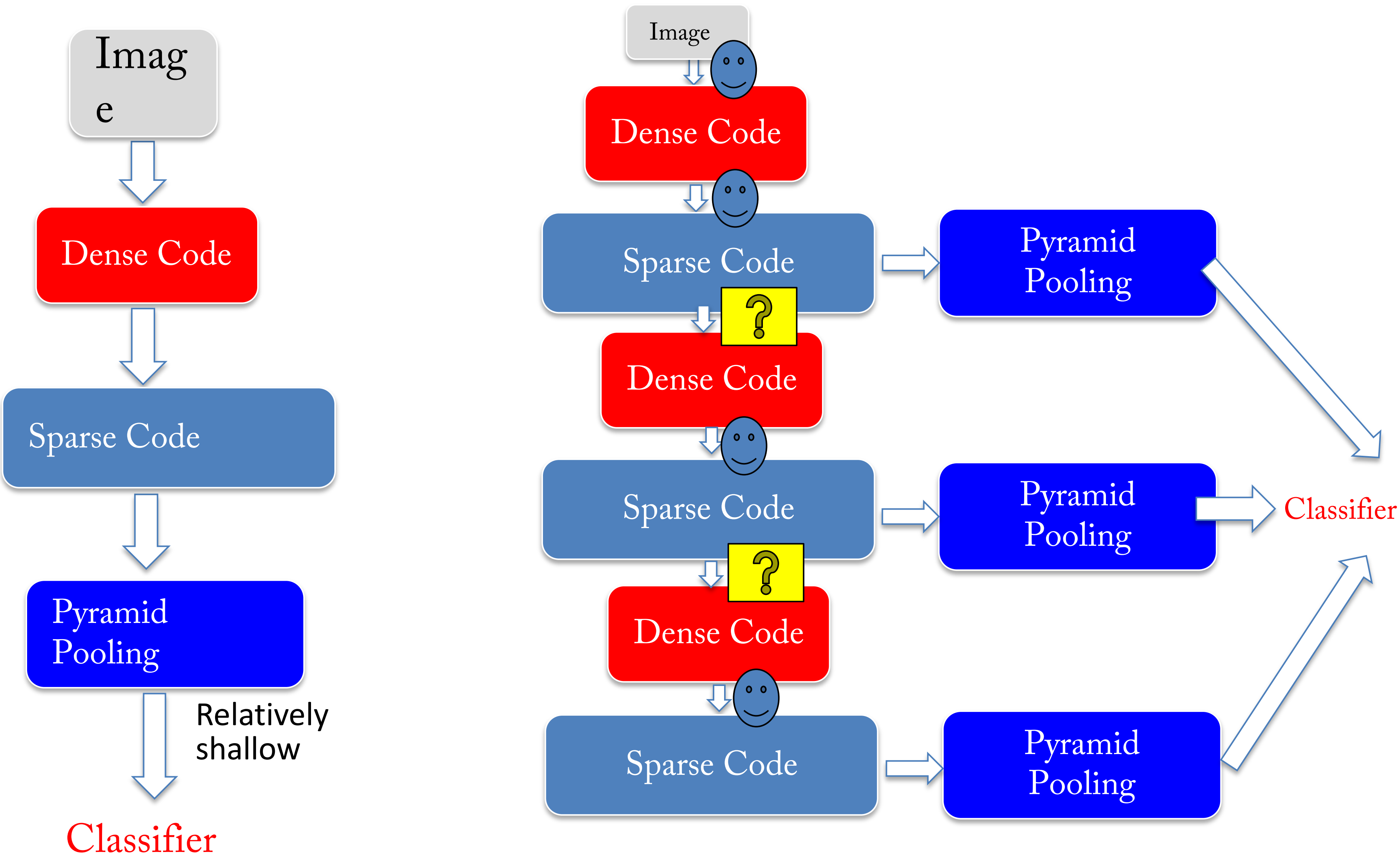
Deep Sparse Coding





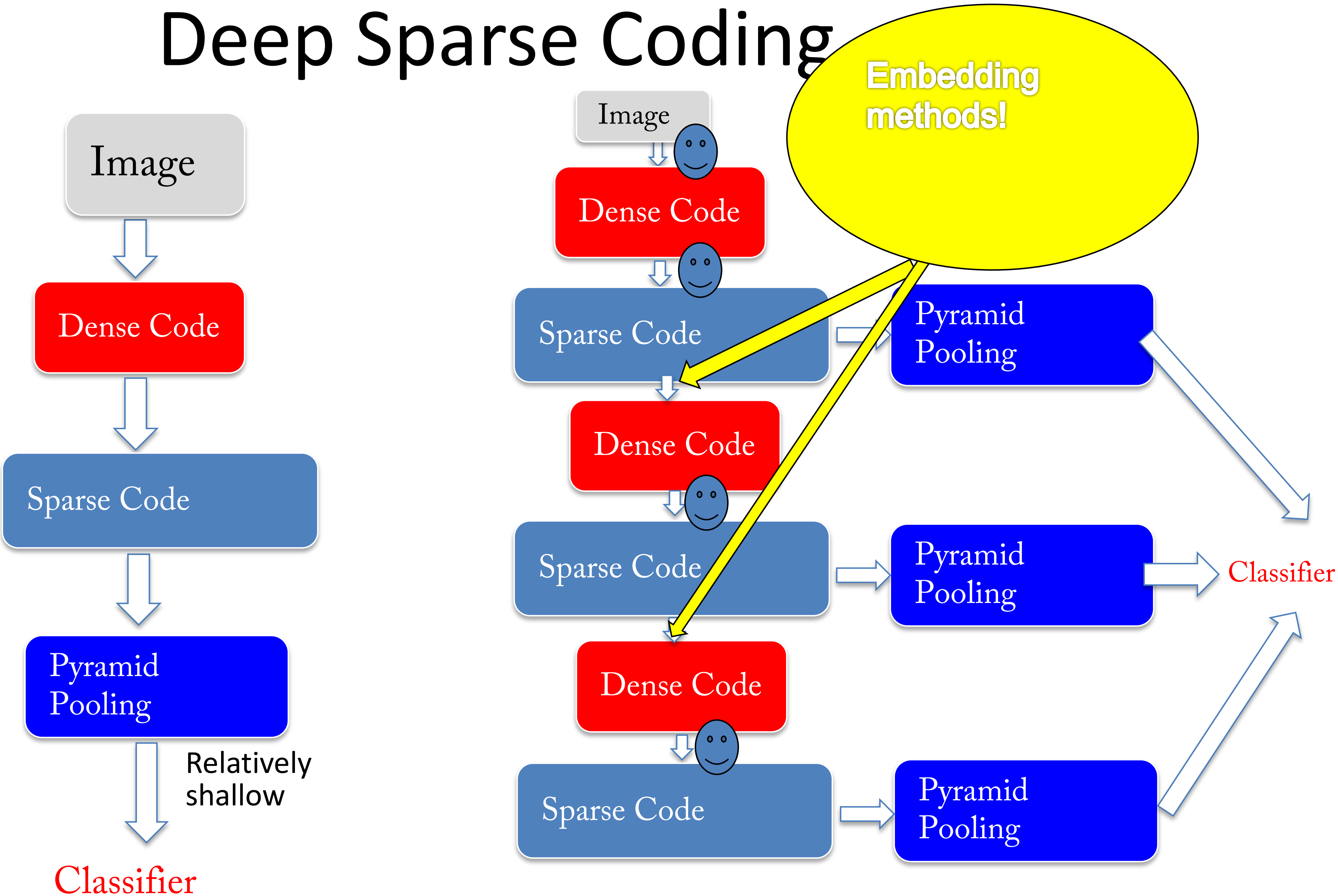
Unsupervised Deep Architecture!

Deep Sparse Coding



Unsupervised Deep Architecture!

Deep Sparse Coding

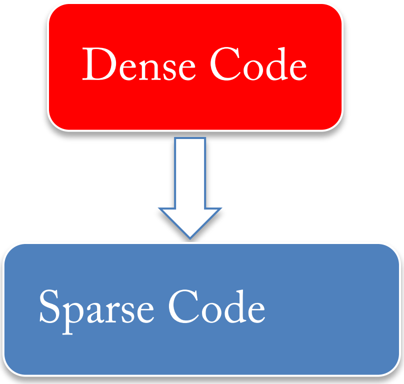


Unsupervised Deep Architecture!

Dimension expansion

Idea:

- Separation (for classification)
- Representation (bag of visual words)



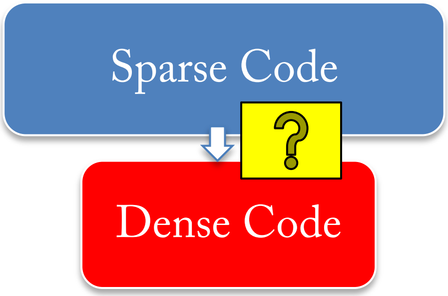
Method:

- Sparse coding

Dimension reduction

Idea:

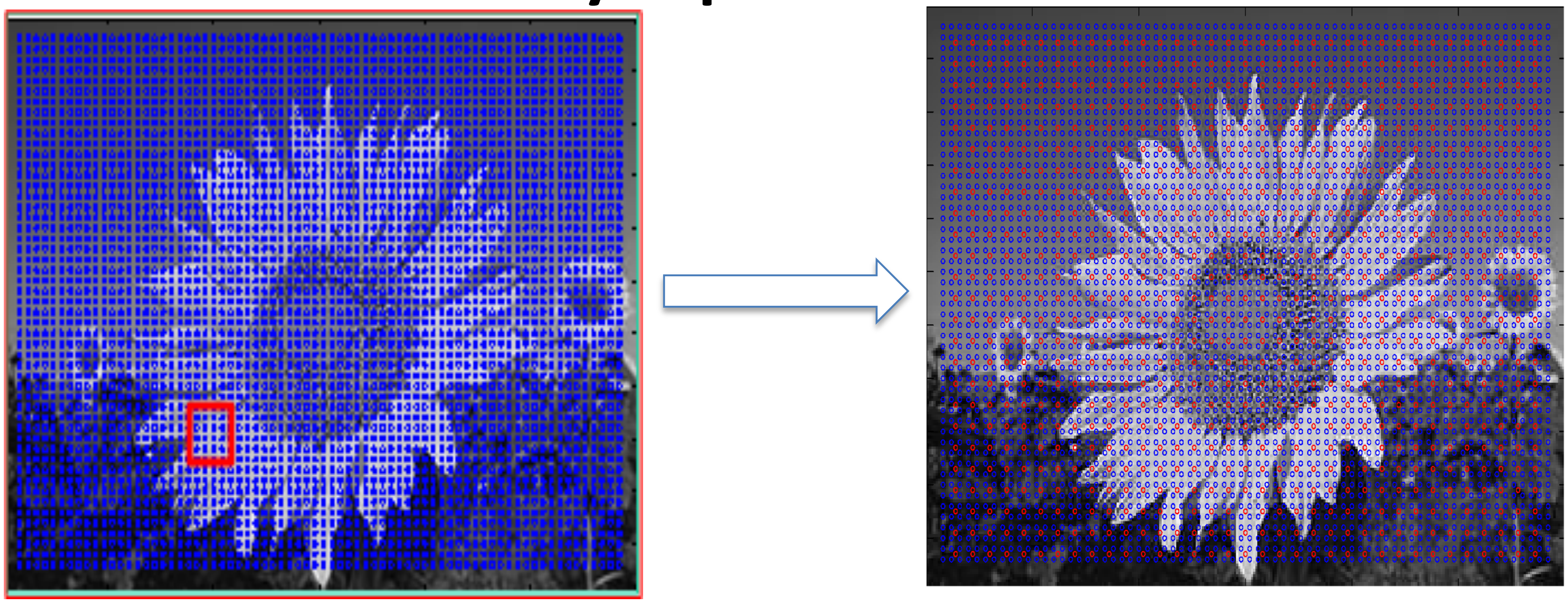
- Invariance
- Compositionality

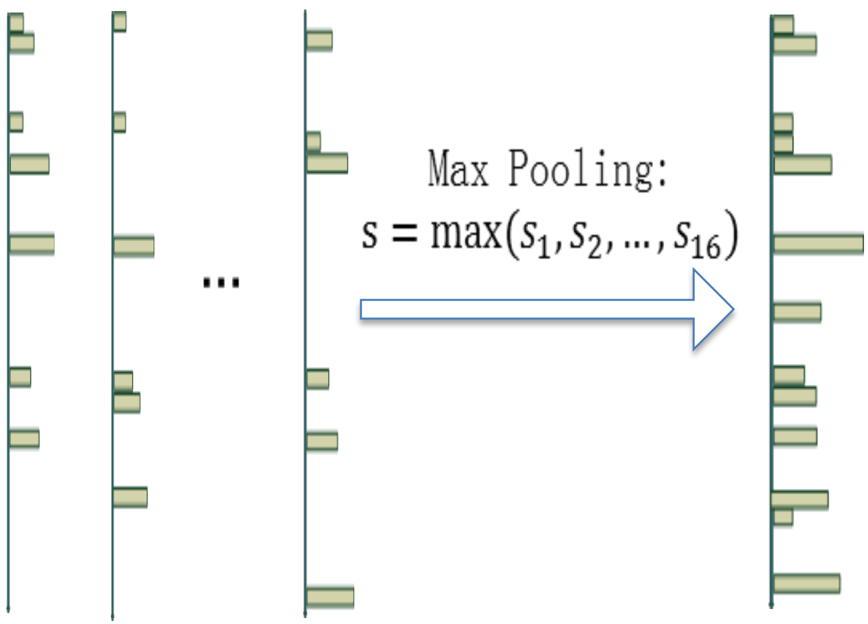


Method:

- Locally spatial pooling + Low dimensional embedding
- Both unsupervised

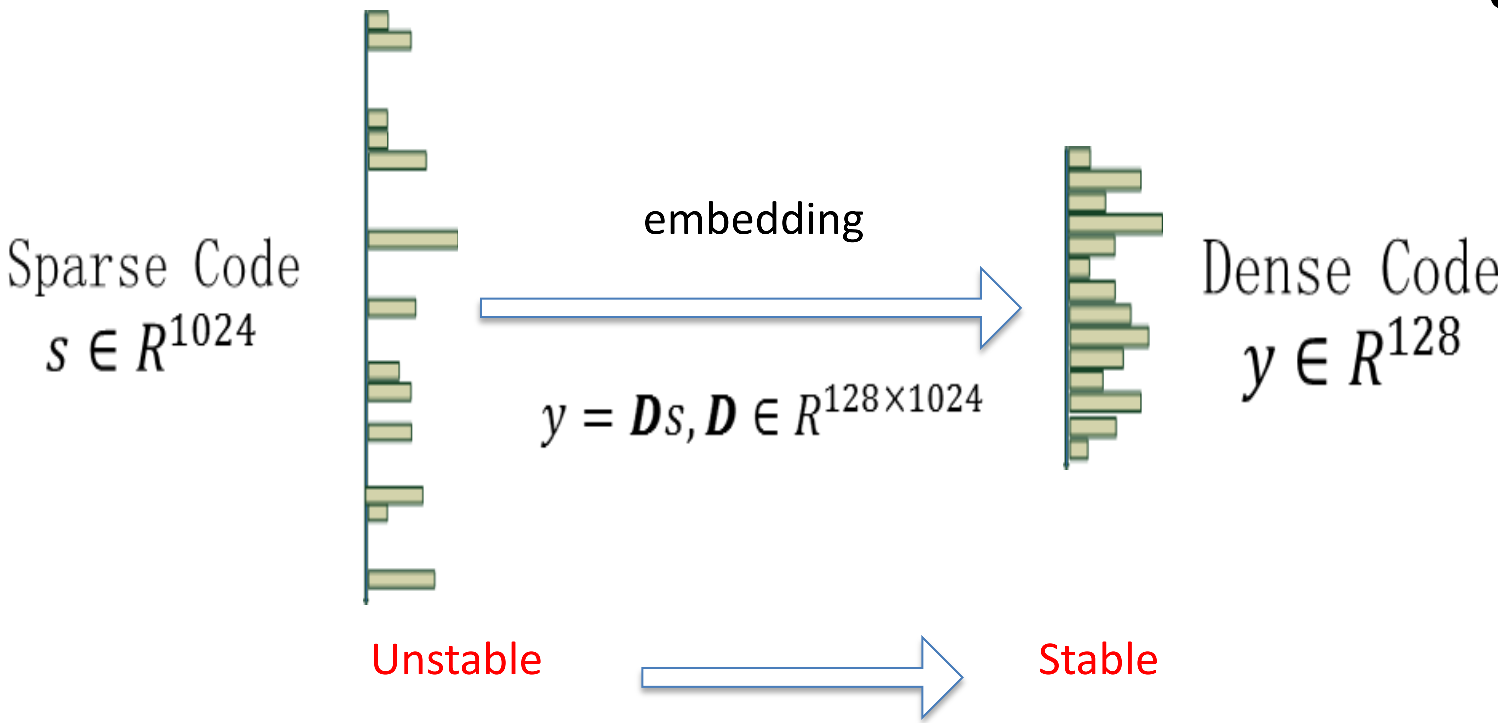
Locally spatial pooling



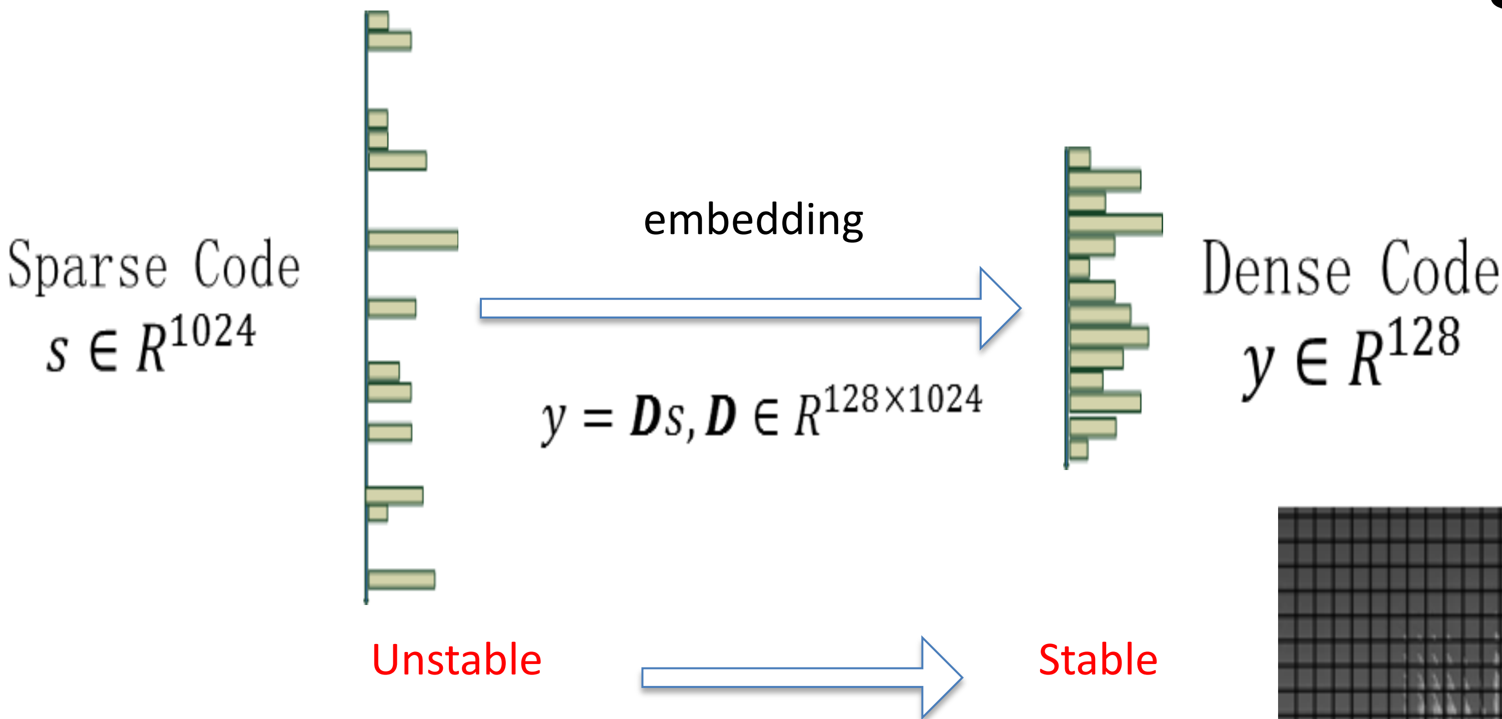


Covers larger area

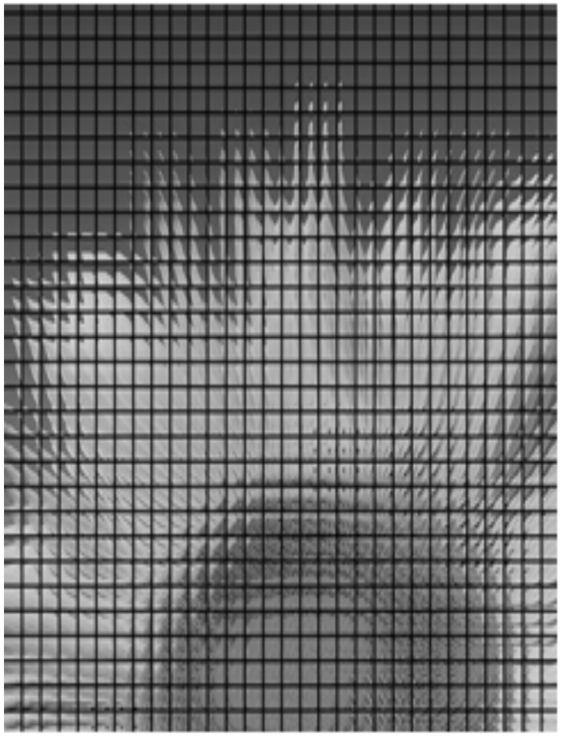
Low dimensional embedding



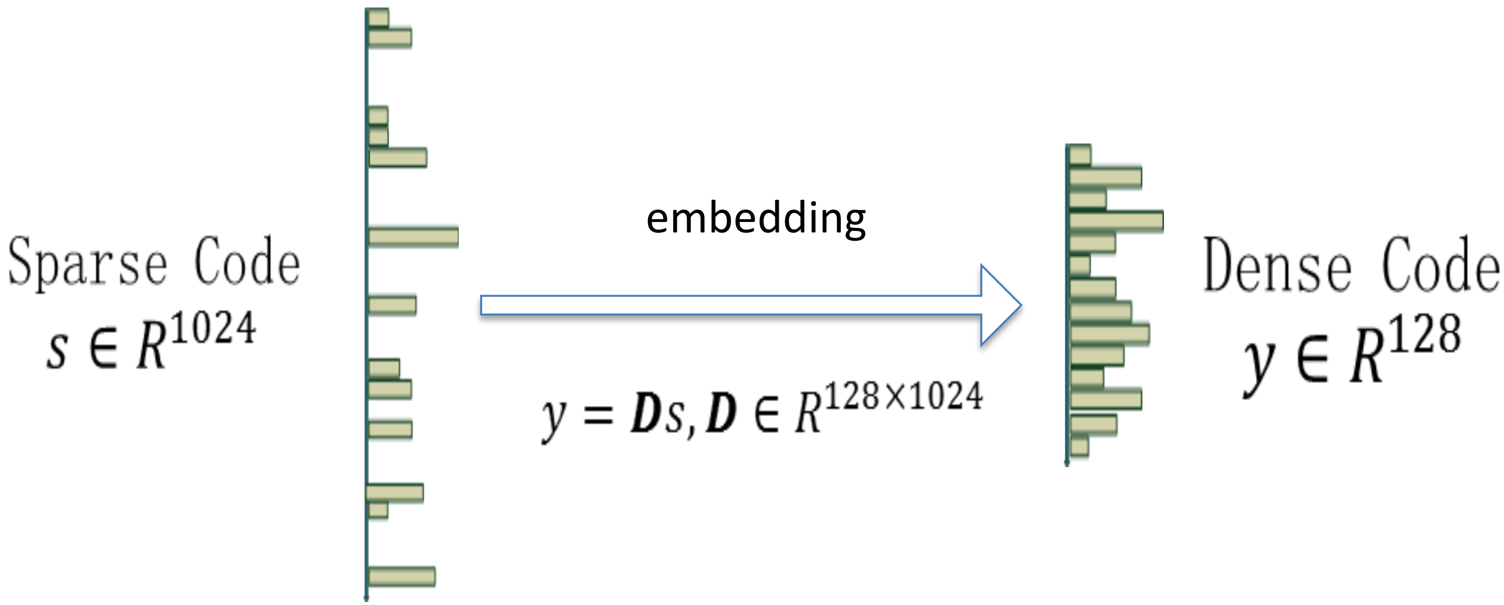
Low dimensional embedding



Idea: embedding with the help of spatial information



Low dimensional embedding



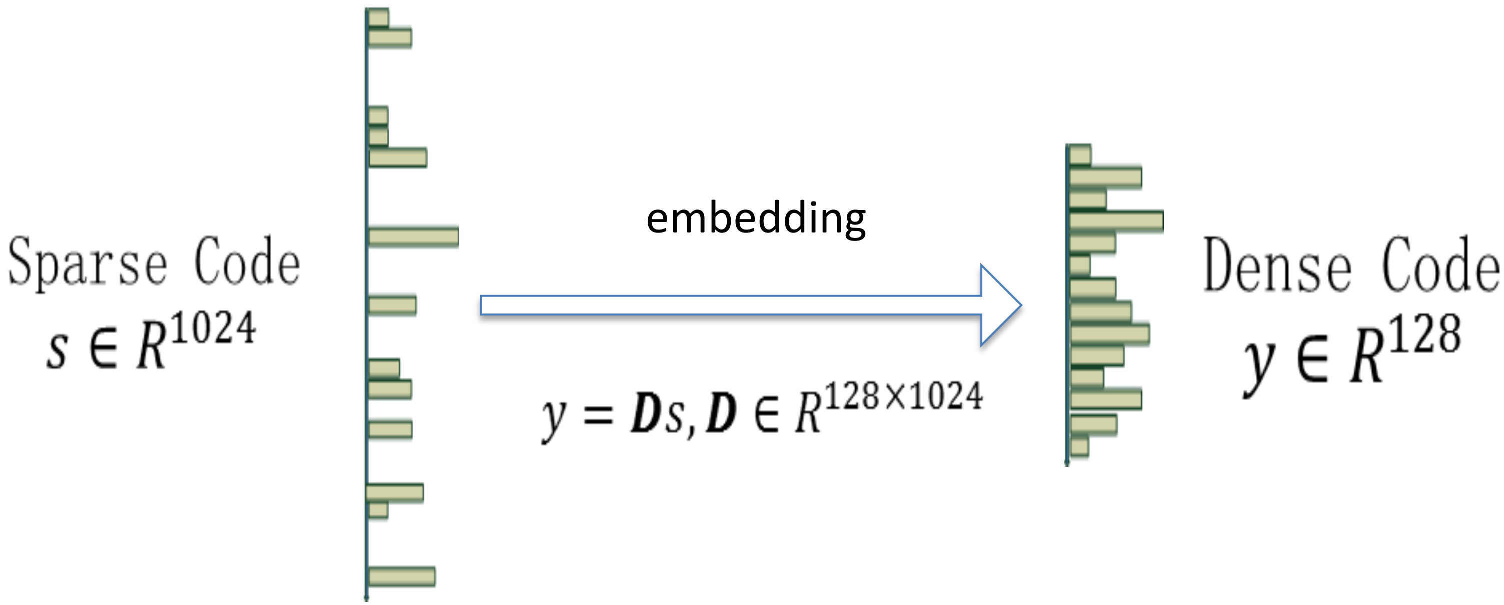
Dimensionality Reduction by Learning an Invariant Mapping (DR. LIM)

$$D = argmin_D \, \Sigma_{i,j} w_{ij} ||Ds_i - Ds_j||^2 + \Sigma_{i,j} (1 - w_{ij}) \max(0, \alpha - ||Ds_i - Ds_j||)^2$$

$w_{ij} = 1$, if s_i and s_j are overlapping neighbors

$w_{ij} = 0$, if s_i and s_j are non-overlapping neighbors

Low dimensional embedding



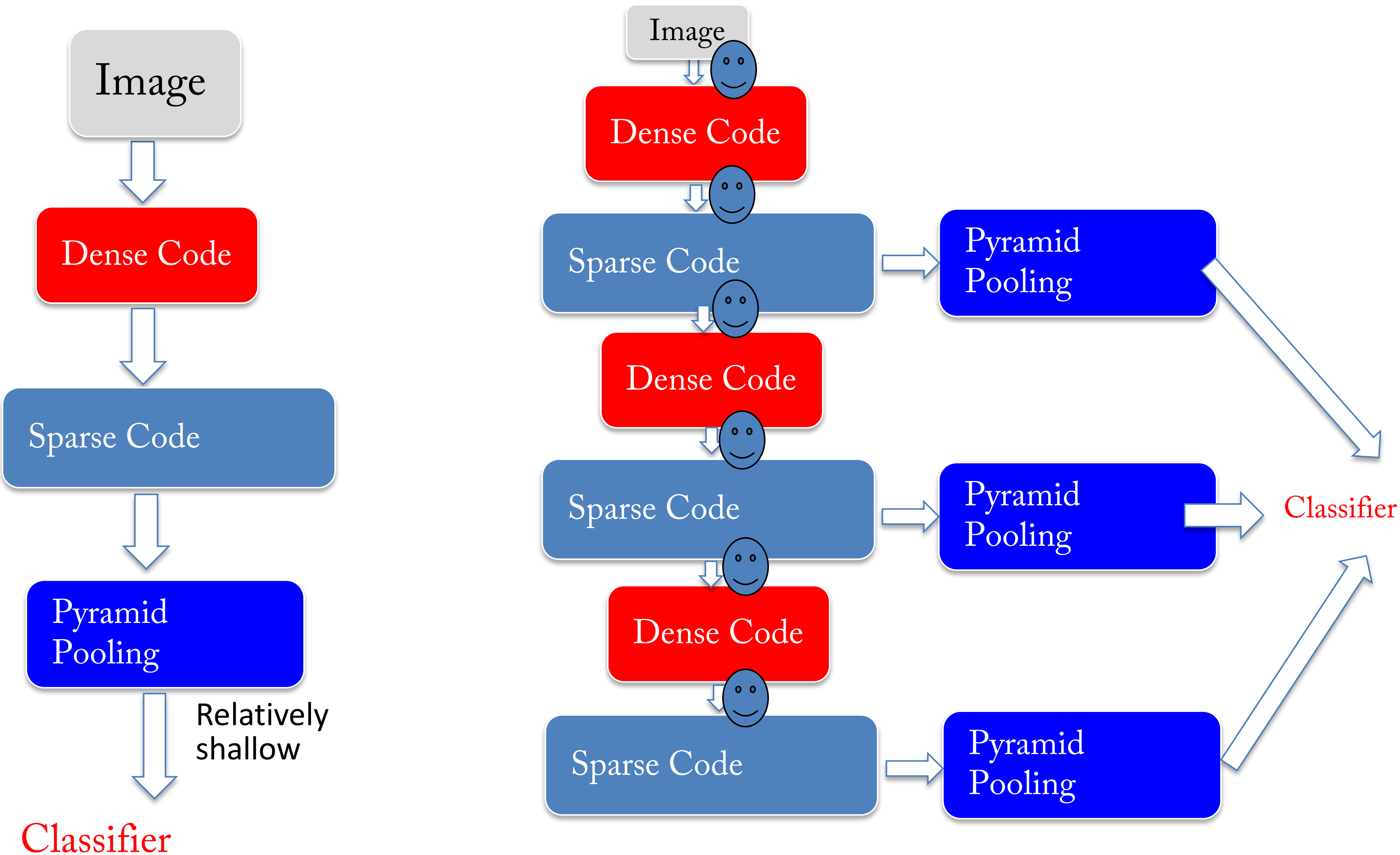
Dimensionality Reduction by Learning an Invariant Mapping (DR. LIM)

$$D = argmin_D \, \Sigma_{i,j} w_{ij} ||Ds_i - Ds_j||^2 + \Sigma_{i,j}\left(1 - w_{ij}\right) \max(0, \alpha - ||Ds_i - Ds_j||)^2$$

$w_{ij} = 1$, if s_i and s_j are overlapping neighbors

$w_{ij} = 0$, if s_i and s_j are non-overlapping neighbors

Deep Sparse Coding



Unsupervised Deep Architecture!

Conclusion

- Combine sparse coding with deep learning
- Combine dimension expansion and dimension reduction
- A connecting function is learned by embedding method
- Combining multiple layers of sparse code achieves state-of-the-art performance on image classification tasks