PlanetS

UNIVERSITÄT
BERN

Convolutional Neural Networks

Paolo Favaro

Workshop on Machine Learning - Observatoire de Geneve

Contents

- Convolutional Neural Networks
    - Convolutions (standard, unshared, tiled)
- Based on **Chapter 9** of Deep Learning by Goodfellow, Bengio, Courville

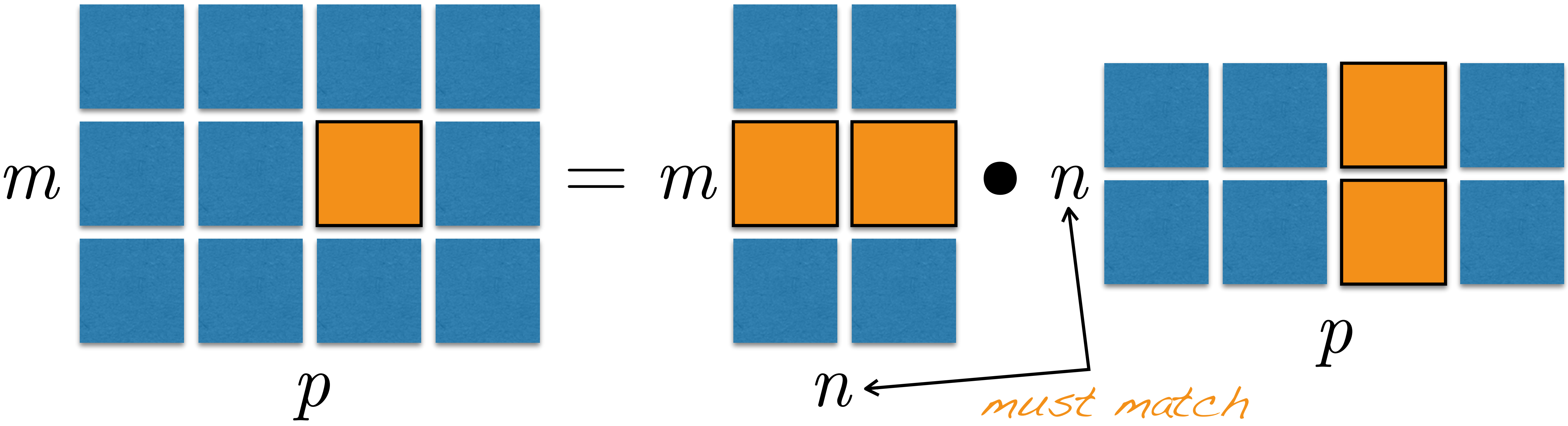
Convolutional Networks

- A specialized neural network for data arranged on a grid (e.g., audio signals, images)

- Allow neural networks to deal with high-dimensional data

- Key idea is to substitute fully connected layers with a convolution

Fully Connected Layers



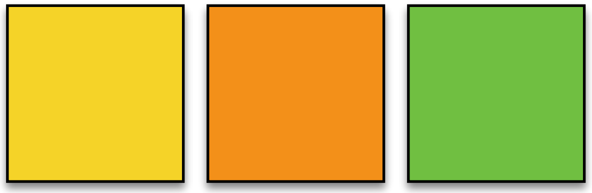
matrix product

The Convolution Operation

$$\overset{\substack{\text{feature map}\\ \downarrow}}{s}[m,n] = (x * w)[m,n] = \sum_{i,j} \overset{\substack{\text{input}\\ \downarrow}}{x}[m-i,n-j]\overset{\substack{\text{kernel}\\ \downarrow}}{w}[i,j]$$

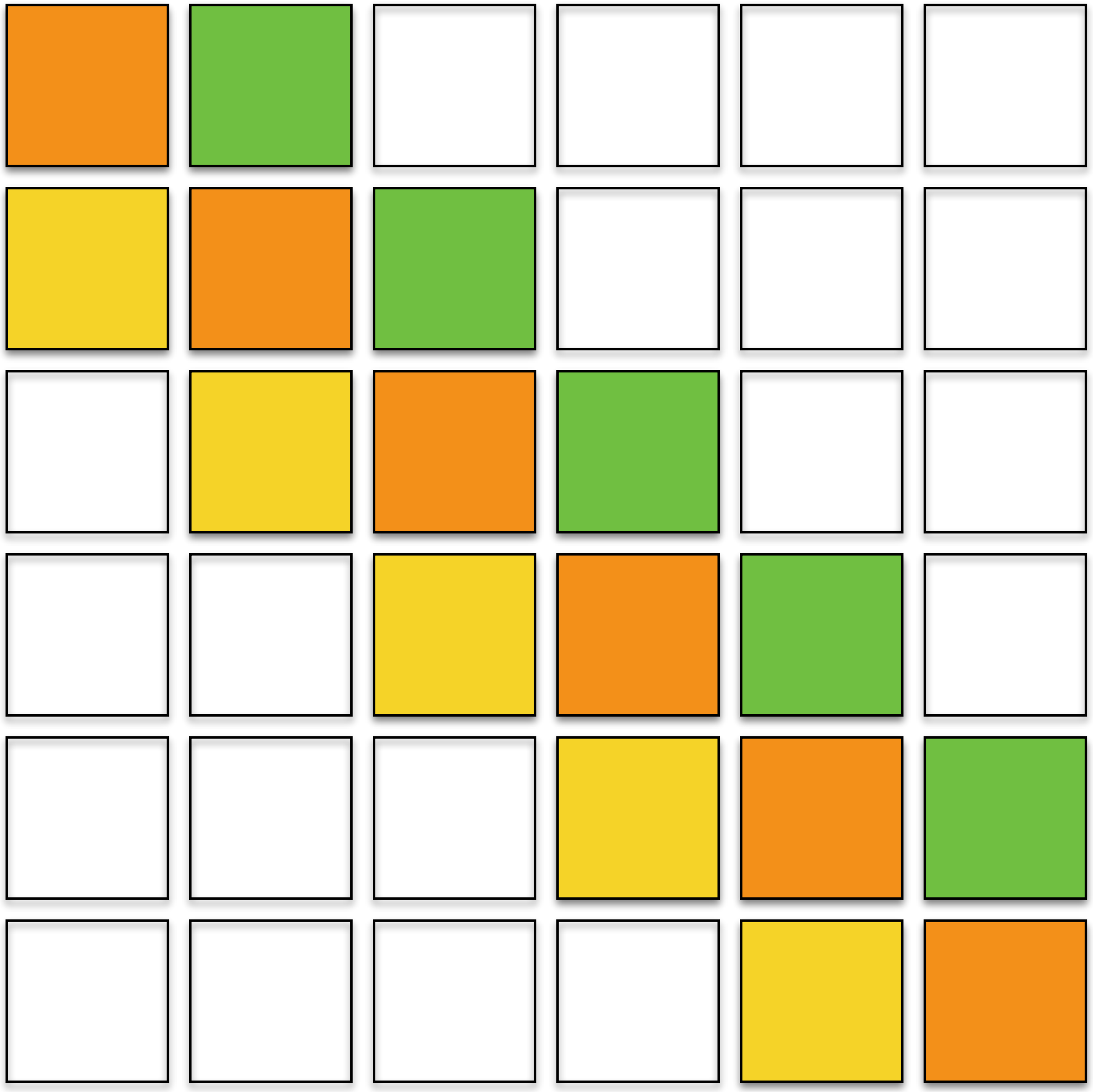
$$\text{symmetric} \longrightarrow = \sum_{i,j} w[m-i,n-j]\underset{\substack{\uparrow\\ \text{linear in } x\\ \text{with fixed } w}}{x}[i,j]$$

Toeplitz Matrix



kernel

$$s[n] = (x * w)[n]$$
$$= \sum_i A[n, i]x[i]$$



Toeplitz matrix

Variants

- Input data is typically a 4D tensor: 2 dimensions for the spatial domain, 1 dimension for the channels (e.g., colors), and 1 dimension for the batch

- The convolution (correlation) applies to the spatial domain only

$$\underset{\substack{\uparrow\\ \text{output}}}{Z_{i,j,k}} = \sum_{l,m,n} \underset{\substack{\uparrow\\ \text{input}}}{V_{l,j+m,k+n}} \underset{\substack{\uparrow\\ \text{kernel}}}{K_{i,l,m,n}}$$

Stride

- We can also skip outputs by defining a **stride** s larger than 1

$$Z_{i,j,k} = \sum_{l,m,n} V_{l,j\times s+m,k\times s+n} K_{i,l,m,n}$$

Padding

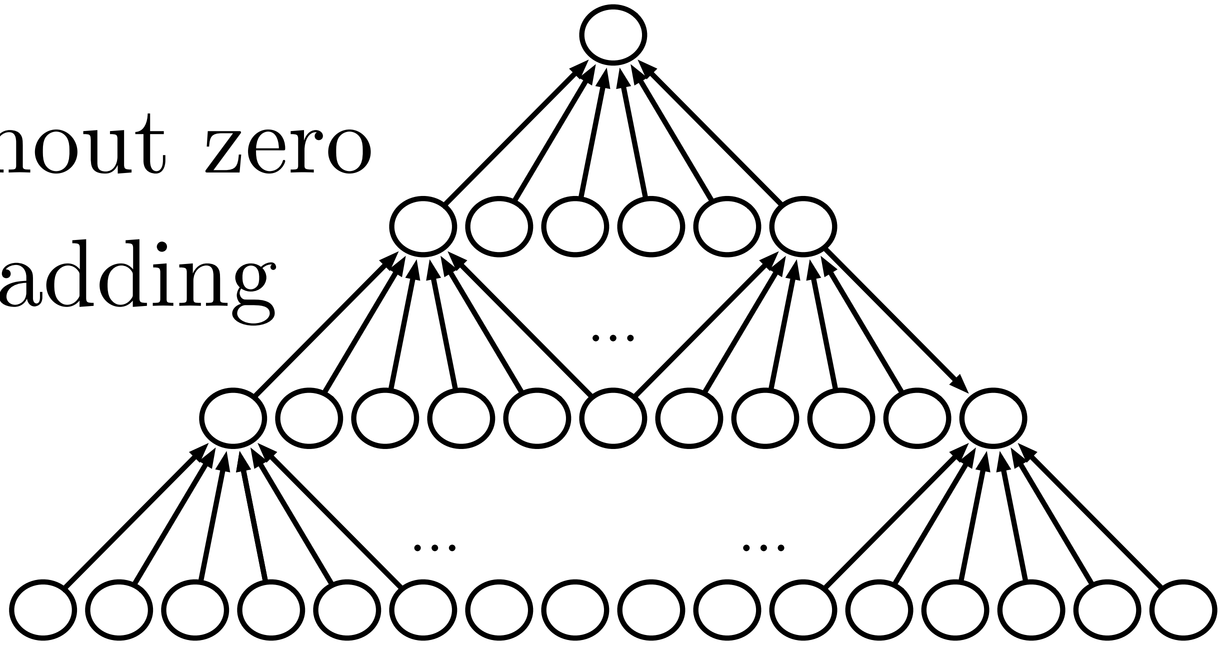
- The output of a convolution is valid as long as the summation uses available values

- In a convolution the valid output size is equal to:
  the input size - the size of the kernel + 1

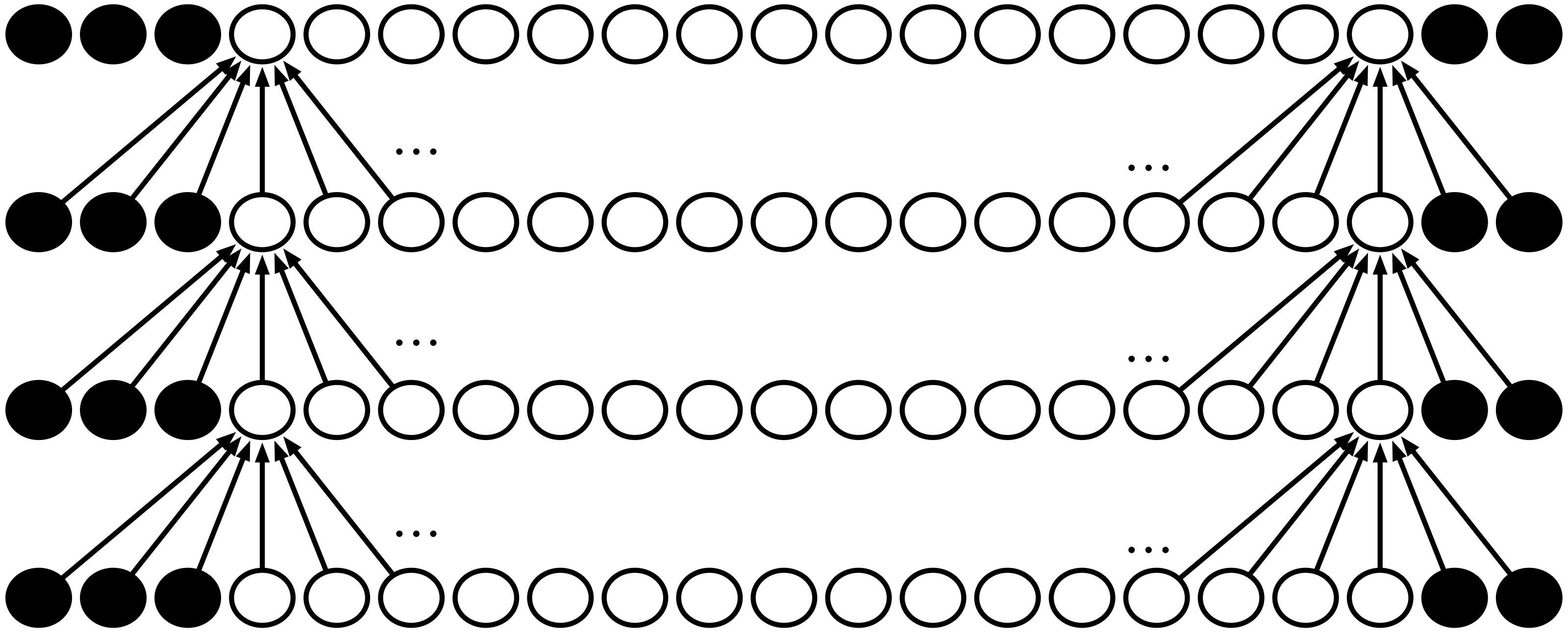
- Unless we make boundary assumptions, a convolution will lead to a progressive shrinking of the input

- **Padding** is the assumption that outside the given domain the input takes some fixed values (e.g., zero)

Padding



without zero padding



with zero padding

Data Types

- Input data can be in different formats

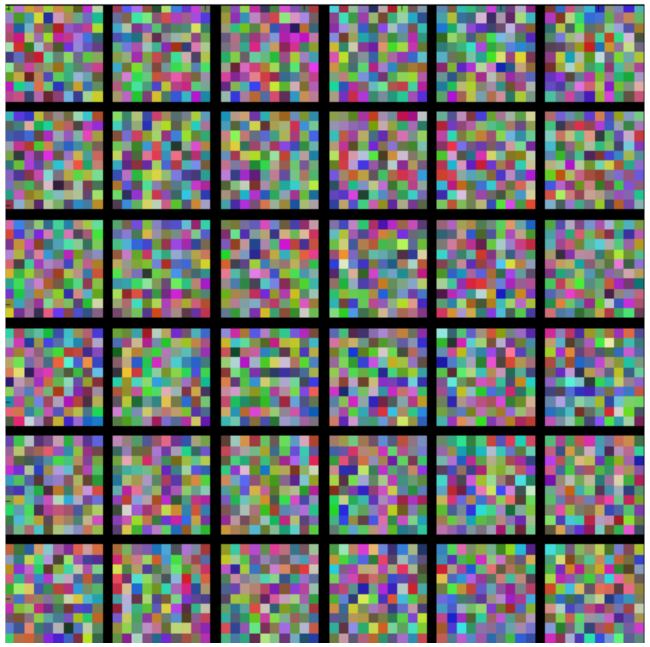
- 1D: Audio waveforms (single channel) and skeleton animation data/motion (multi-channel)

- 2D: Audio data preprocessed via Fourier (single channel), color image data (multi-channel)

- 3D: Volumetric data such as CT scans (single channel), color video data (multi-channel)

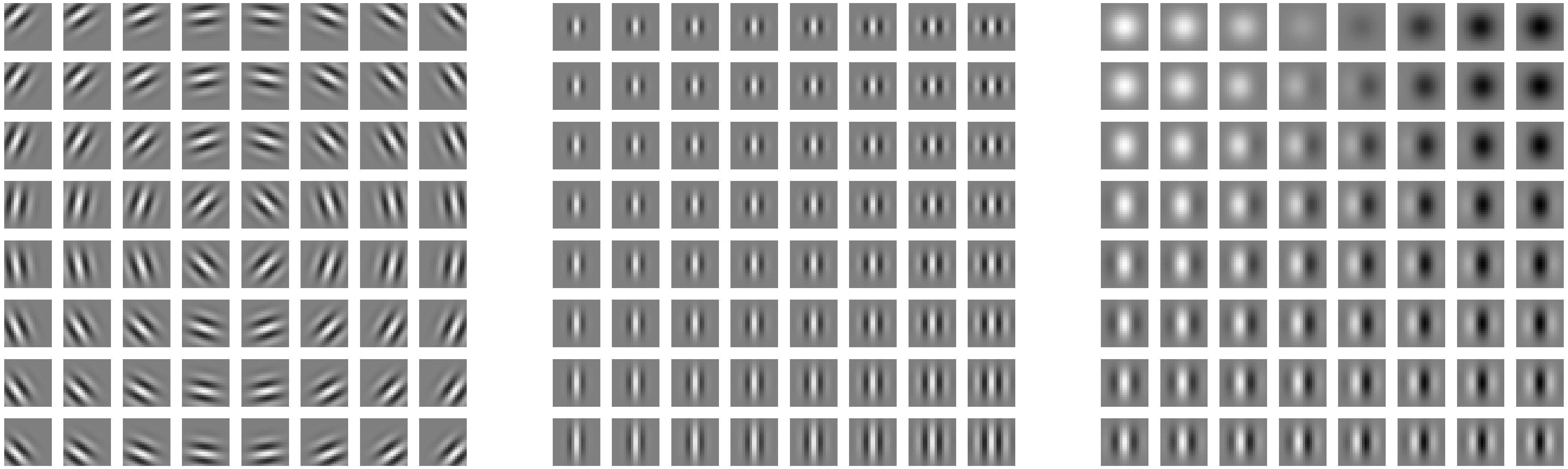
Random or Unsupervised Features

- Kernels can be initialized
  - with random weights



Random or Unsupervised Features

- Kernels can be initialized
  - with hand-designed features



Random or Unsupervised Features

- Kernels can be initialized
  - with unsupervised learning algorithms (e.g., apply k-means clustering to patches, then use centroids as kernels)