

- Spectral Networks and deep locally connected networks on graphs

CNN can exploit the following structure:

- 1) translation structure, allowing the use of filters instead of generic linear maps and hence weight sharing
- 2) metric on the grid, allowing compactly supported filters, whose support is typically much smaller than the size of the input signals
- 3) multiscale dyadic clustering of grid, allowing subsampling, implemented through stride convolutions and pooling

computational complexity of CNN: n input coordinates(?)

on a d -dim grid, a fully connected layer with m outputs requires $n*m$ parameters.
(independent of dimension?)

using small filter reduces complexity to $O(n)$ per feature map(?)

using multiscale dyadic clustering allows each successive layer to use a factor of 2^d less coordinates per filter(?)

Motivation of studying cnn on graphs

data defined on 3-D meshes, such as surface tension or temperature, measurements from a network of meteorological station, or data coming from social networks or collaborative filtering

another example: intermediate representation arising from deep neural networks(?)

spatial convolutional structure can be exploited at several layers, typically CNN don't assume any geometry in the feature dimension, resulting in 4-D tensor which are only convolutional along their spatial coordinates

two different constructions: first one (spatial construction) extend properties (2) and (3) to general graphs (so no weight sharing? Yes. No easy way to induce weight sharing across different locations of the graph), $O(n)$ parameters instead of $O(n^2)$

second one (spectral constructions) draw on properties on convolutions in Fourier domain. In \mathbb{R}^d , convolutions are linear operators diagonalised by the Fourier basis $\exp(i \cdot w \cdot t)$. we extend convolutions to general graphs by finding corresponding Fourier basis (Graph Laplacian arises). $O(n)$ parameters per feature map.

Smoothness and Sparsity

filters are multipliers on the eigenvalues of the laplacian Δ

functions that are smooth relative to the grid metric have coefficient with quick decay in the basis of eigenvectors of laplacian

The eigenvectors of the subsampled Laplacian are the low frequency eigenvectors of laplacian (why?)

$$x_{k+1,j} = h(V \Sigma_{i=1}^{f_{k-1}} F_{k,i,j} V^T x_{k,j})$$

Cons: most graphs have meaningful eigenvectors only for the very top of the spectrum; not obvious how to do either the forwardprop or backprop efficiently while applying the nonlinearity on space side.