CNN has been very successful
CNN has been very successful.
CNN nicely exploits the grid structure

grid metric  →  locally supported filters
CNN nicely exploits the grid structure

$G(i - i_0)$

$G(i)$

translation structure

allow the use of filters and weight sharing
CNN nicely exploits the grid structure

natural way to downsample multi-scale analysis
In many cases, data lies on less regular structures (generic graphs)

3D shape graph  social network  molecules
Moreover, conventional CNN doesn’t not assume any geometry in feature dimensions.
Geometry aware convolution can be important

convolutional along spatial coordinates

convolutional considering underlying geometry
Today’s topic

\[ \Omega: \text{set of input coordinates} \]
\[ W_{i,j}: \text{similarity between coordinates } i \text{ and } j \]
Agenda

• Challenges
• Background knowledge
• Spatial construction
  • Geodesic CNN
• Spectral construction
  • Spectral CNN
  • Anisotropic CNN
  • SyncSpecCNN
Agenda

- **Challenges**
- Background knowledge
- Spatial construction
  - Geodesic CNN
- Spectral construction
  - Spectral CNN
  - Anisotropic CNN
  - SyncSpecCNN
How to define convolution kernel on graphs?

- Desired properties:
  - locally supported (w.r.t graph metric)
  - allowing weight sharing across different coordinates

from Shuman et al. 2013
How to allow multi-scale analysis?

grid structure

graph structure

from Michaël Defferrard et al. 2016
How to allow multi-scale analysis?

grid structure

graph structure

hierarchical graph coarsening
structure aware? efficiency? can we do more?

from Michaël Defferrard et al. 2016
How to ensure generalizability across graphs?

grid structure has a natural alignment
How to ensure generalizability across graphs?

Graph structure does not have a natural alignment.
Agenda

- Challenges
- **Background knowledge**
- Spatial construction
  - Geodesic CNN
- Spectral construction
  - Spectral CNN
  - Anisotropic CNN
  - SyncSpecCNN
Laplacian

Smooth scalar field $f$

from Jonathan Masci et al
Laplacian

- **Gradient** $\nabla f(x) = \text{direction of the steepest increase of } f \text{ at } x$
Laplacian

- **Gradient** $\nabla f(x) = \text{‘direction of the steepest increase of } f \text{ at } x$'

- **Divergence** $\text{div}(F(x)) = \text{‘density of an outward flux of } F \text{ from an infinitesimal volume around } x$’

Smooth vector field $F$
Laplacian

- **Gradient** $\nabla f(x) = \text{‘direction of the steepest increase of } f \text{ at } x$’

- **Divergence** $\text{div}(F(x)) = \text{‘density of an outward flux of } F \text{ from an infinitesimal volume around } x$’

**Divergence theorem:**

$$\int_V \text{div}(F) dV = \int_{\partial V} \langle F, \hat{n} \rangle dS$$

Smooth vector field $F$

from Jonathan Masci et al
Laplacian

- **Gradient** $\nabla f(x) = \text{‘direction of the steepest increase of } f \text{ at } x$'

- **Divergence** $\text{div}(F(x)) = \text{‘density of an outward flux of } F \text{ from an infinitesimal volume around } x$'

**Divergence theorem:**

$$\int_V \text{div}(F)\,dV = \int_{\partial V} \langle F, \hat{n} \rangle\,dS$$

- **Laplacian** $\Delta f(x) = -\text{div}(\nabla f(x))$

  ‘difference between $f(x)$ and the average of $f$ on an infinitesimal sphere around $x$’ (consequence of the Divergence theorem)

from Jonathan Masci et al
Discrete Laplacian

One-dimensional

\[(\Delta f)_i \approx 2f_i - f_{i-1} - f_{i+1}\]

Two-dimensional

\[(\Delta f)_{ij} \approx 4f_{ij} - f_{i-1,j} - f_{i+1,j} - f_{i,j-1} - f_{i,j+1}\]
Physical application: heat equation

\[ f_t = -c \Delta f \]

*Newton’s law of cooling:* rate of change of the temperature of an object is proportional to the difference between its own temperature and the temperature of the surrounding.

\[ c \ [m^2/sec] = \text{thermal diffusivity constant} \ (\text{assumed} = 1) \]
Riemannian manifold

- Manifold $\mathcal{X} = \text{topological space}$
- No global Euclidean structure
- **Tangent plane** $T_x\mathcal{X} = \text{local Euclidean representation of manifold } \mathcal{X} \text{ around } x$
Riemannian manifold

- Manifold $\mathcal{X} = $ topological space
- No global Euclidean structure
- Tangent plane $T_x \mathcal{X} =$ local Euclidean representation of manifold $\mathcal{X}$ around $x$
- Riemannian metric

$$\langle \cdot, \cdot \rangle_{T_x \mathcal{X}} : T_x \mathcal{X} \times T_x \mathcal{X} \to \mathbb{R}$$
depending smoothly on $x$
Riemannian manifold

- Manifold $\mathcal{X} = \text{topological space}$
- No global Euclidean structure
- **Tangent plane** $T_x \mathcal{X} = \text{local Euclidean representation of manifold } \mathcal{X} \text{ around } x$
- **Riemannian metric**
  \[
  \langle \cdot, \cdot \rangle_{T_x \mathcal{X}} : T_x \mathcal{X} \times T_x \mathcal{X} \rightarrow \mathbb{R}
  \]
  depending smoothly on $x$
- **Isometry** = metric-preserving shape deformation

from Jonathan Masci et al
Riemannian manifold

- Manifold $\mathcal{X} = \text{topological space}$
- No global Euclidean structure
- **Tangent plane** $T_x \mathcal{X} = \text{local Euclidean representation of manifold } \mathcal{X} \text{ around } x$
- **Riemannian metric**
  \[
  \langle \cdot, \cdot \rangle_{T_x \mathcal{X}} : T_x \mathcal{X} \times T_x \mathcal{X} \to \mathbb{R}
  \]
  depending smoothly on $x$
- **Isometry** = metric-preserving shape deformation

- **Geodesic** = shortest path on $\mathcal{X}$ between $x$ and $x'$

from Jonathan Masci et al
Calculus on manifold

- Scalar field $f : \mathcal{X} \rightarrow \mathbb{R}$
- Vector field $F : \mathcal{X} \rightarrow T\mathcal{X}$
Calculus on manifold

- **Intrinsic gradient** operator
  \[ \nabla f : L^2(\mathcal{X}) \to L^2(T\mathcal{X}) \]
  "direction of steepest change of \( f \)"

- **Intrinsic divergence** operator
  \[ \text{div} : L^2(T\mathcal{X}) \to L^2(\mathcal{X}) \]
  "net flow of field \( F \) at \( x \)"

(from Jonathan Masci et al)
Calculus on manifold

- **Laplacian** $\Delta : L^2(\mathcal{X}) \rightarrow L^2(\mathcal{X})$

  $\Delta f = -\text{div}(\nabla f)$

  “difference between $f(x)$ and average value of $f$ around $x$”

- **Intrinsic** (expressed solely in terms of the Riemannian metric)
- **Isometry-invariant**
- **Positive semidefinite**

from Jonathan Masci et al
Discrete Laplacian

Undirected graph \((V, E)\)

\[(\Delta f)_i \approx \sum_{(i,j) \in E} w_{ij} (f_i - f_j)\]

Triangular mesh \((V, E, F)\)

\[(\Delta f)_i \approx \frac{1}{a_i} \sum_{(i,j) \in E} \frac{\cot \alpha_{ij} + \cot \beta_{ij}}{2} (f_i - f_j)\]

\(a_i = \text{local area element}\)

from Jonathan Masci et al
Fourier analysis - Euclidean space

A function \( f : [-\pi, \pi] \rightarrow \mathbb{R} \) can be written as **Fourier series**

\[
f(x) = \sum_{\omega} \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x')e^{i\omega x'} dx' e^{-i\omega x}
\]

= \alpha_1 + \alpha_2 + \alpha_3 + \ldots

from Jonathan Masci et al
Fourier analysis - Euclidean space

A function $f : [-\pi, \pi] \rightarrow \mathbb{R}$ can be written as Fourier series

$$f(x) = \sum_{\omega} \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x') e^{i\omega x'} dx' e^{-i\omega x}$$

$$\hat{f}(\omega) = \langle f, e^{-i\omega x} \rangle_{L^2([-\pi, \pi])}$$

$$= \alpha_1 + \alpha_2 + \alpha_3 + \ldots$$

from Jonathan Masci et al
Fourier analysis - Euclidean space

A function $f : [-\pi, \pi] \rightarrow \mathbb{R}$ can be written as Fourier series

$$f(x) = \sum_{\omega} \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x') e^{i\omega x'} dx' e^{-i\omega x}$$

$$\hat{f}(\omega) = (f, e^{-i\omega x})_{L^2([-\pi, \pi])}$$

$$= \alpha_1 + \alpha_2 + \alpha_3 + \ldots$$

Fourier basis = Laplacian eigenfunctions: $\Delta e^{-i\omega x} = \omega^2 e^{-i\omega x}$

from Jonathan Masci et al
Fourier analysis - non Euclidean space

A function $f : \mathcal{X} \to \mathbb{R}$ can be written as Fourier series

$$f(x) = \sum_{k \geq 0} \int_{\mathcal{X}} f(x') \phi_k(x') dx' \phi_k(x)$$

$$\hat{f}_k = \langle f, \phi_k \rangle_{L^2(\mathcal{X})}$$

Fourier basis = Laplacian eigenfunctions: $\Delta \phi_k(x) = \lambda_k \phi_k(x)$

from Jonathan Masci et al
Agenda

• Challenges
• Background knowledge
• Spatial construction
  • Geodesic CNN
• Spectral construction
  • Spectral CNN
  • Anisotropic CNN
  • SyncSpecCNN
How to define convolution kernel on graphs?

How to allow multi-scale analysis?

How to ensure generalizability across graphs?
Geodesic CNN

- Constructing convolution kernels:
  - Local system of geodesic polar coordinate
  - Extract a small patch at each point \( x \)
  - Radial coordinate \( \rho \) - geodesic distance (truncated)
  - Angular coordinate \( \theta \) - direction of geodesics (origin choice)
Geodesic CNN

- Local chart: bijective map

\[ \Omega(x) : B_{\rho_0}(x) \rightarrow [0, \rho_0] \times [0, 2\pi) \]

from manifold to local coordinates

\((\rho, \theta)\) around \(x\)
Geodesic CNN

- **Local chart**: bijective map
  \[ \Omega(x) : B_{\rho_0}(x) \rightarrow [0, \rho_0] \times [0, 2\pi) \]
  from manifold to local coordinates
  \((\rho, \theta)\) around \(x\)

- **Patch operator** applied to \(f \in L^2(X)\)
  interpolate \(f\) in the local coordinate
Geodesic CNN

\[
(D(x)f)(\rho, \theta) = \frac{\int_X v_\rho(x, \xi) v_\theta(x, \xi) f(\xi) d\xi}{\int_X v_\rho(x, \xi) v_\theta(x, \xi) d\xi}
\]

Radial weight

\[
v_\rho(x, \xi) \propto e^{-\frac{(d_X(x, \xi) - \rho)^2}{\sigma_\rho^2}}
\]

Angular weight

\[
v_\theta(x, \xi) \propto e^{-\frac{d_X^2(\Gamma(x, \theta), \xi)}{\sigma_\theta^2}}
\]
Geodesic CNN

- **Geodesic convolution** = apply filter $a$ to patches extracted from $f \in L^2(X)$ in local geodesic polar coordinates

\[
(f * a)(x) = \sum_{\theta, r} (D(x)f)(r, \theta) \ a(\theta, r)
\]
Geodesic CNN

- **Geodesic convolution** = apply filter $a$ to patches extracted from $f \in L^2(X)$ in local geodesic polar coordinates

$$
(f * a)(x) = \sum_{\theta, r} (D(x)f)(r, \theta) a(\theta, r)
$$

Jonathan Masci et al 2015
Geodesic CNN

- Geodesic convolution = apply filter $a$ to patches extracted from $f \in L^2(X)$ in local geodesic polar coordinates

\[
(f \ast a)(x) = \sum_{\theta, r} (D(x)f)(r, \theta) \cdot a(\theta, r)
\]

rotation ambiguity
Geodesic CNN

- Geodesic convolution = apply filter $a$ to patches extracted from $f \in L^2(X)$ in local geodesic polar coordinates

$$ (f \ast a)(x) = \sum_{\theta, r} (D(x)f)(r, \theta) a(\theta, r) $$

rotation ambiguity
Geodesic CNN
Geodesic CNN

• Issues:
  
  • The local charting method relies on a fast marching-like procedure requiring a triangular mesh.
  
  • The radius of the geodesic patches must be sufficiently small to acquire a topological disk.
  
  • No effective pooling, purely relying on convolutions to increase receptive field.
Agenda

• Challenges
• Background knowledge
• Spatial construction
  • Geodesic CNN
• Spectral construction
  • Spectral CNN
  • Anisotropic CNN
  • SyncSpecCNN
Fourier analysis

A function $f : [-\pi, \pi] \to \mathbb{R}$ can be written as Fourier series

$$f(x) = \sum_{\omega} \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x') e^{i\omega x'} dx' e^{-i\omega x}$$

$$f(\omega) = \langle f, e^{-i\omega x} \rangle_{L^2([-\pi, \pi])}$$

Fourier basis = Laplacian eigenfunctions: $\Delta e^{-i\omega x} = \omega^2 e^{-i\omega x}$

A function $f : \mathcal{X} \to \mathbb{R}$ can be written as Fourier series

$$f(x) = \sum_{k \geq 0} \int_{\mathcal{X}} f(x') \phi_k(x') dx' \phi_k(x)$$

$$f_k = \langle f, \phi_k \rangle_{L^2(\mathcal{X})}$$

Fourier basis = Laplacian eigenfunctions: $\Delta \phi_k(x) = \lambda_k \phi_k(x)$

Euclidean domain

non Euclidean domain

from Jonathan Masci et al
Convolution Theorem in Euclidean domain

Given two functions $f, g : [-\pi, \pi] \rightarrow \mathbb{R}$ their convolution is a function

$$(f \ast g)(x) = \int_{-\pi}^{\pi} f(\xi)g(x - \xi)d\xi$$

**Convolution Theorem:** Fourier transform diagonalizes the convolution operator $\Rightarrow$ convolution can be computed in the Fourier domain as

$$f \ast g = \mathcal{F}^{-1}(\mathcal{F}f \cdot \mathcal{F}g)$$

from Jonathan Masci et al
Convolution Theorem in Euclidean domain

Time Domain

\[ x[n] \ast h[k] = y[n] \]

Frequency Domain

\[ X[n] \cdot H[k] = Y[n] \]
Convolution Theorem in non Euclidean domain

Generalized convolution of $f, g \in L^2(X)$ can be defined by analogy:

$$(f \star g)(x) = \sum_{k \geq 1} \langle f, \phi_k \rangle_{L^2(X)} \langle g, \phi_k \rangle_{L^2(X)} \phi_k(x)$$

from Jonathan Masci et al
Convolution Theorem in non Euclidean domain

**Generalized convolution** of \( f, g \in L^2(X) \) can be defined by analogy

\[
(f \ast g)(x) = \sum_{k \geq 1} \langle f, \phi_k \rangle_{L^2(X)} \langle g, \phi_k \rangle_{L^2(X)} \phi_k(x)
\]

- product in the Fourier domain
- inverse Fourier transform
Convolution Theorem in non-Euclidean domain

Generalized convolution of $f, g \in L^2(X)$ can be defined by analogy

$$(f \ast g)(x) = \sum_{k \geq 1} \langle f, \phi_k \rangle_{L^2(X)} \langle g, \phi_k \rangle_{L^2(X)} \phi_k(x)$$

product in the Fourier domain

inverse Fourier transform

modified from Jonathan Masci et al
Convolution Theorem in non Euclidean domain

Generalized convolution of \( f, g \in L^2(X) \) can be defined by analogy

\[
(f \ast g)(x) = \sum_{k \geq 1} \langle f, \phi_k \rangle_{L^2(X)} \langle g, \phi_k \rangle_{L^2(X)} \phi_k(x)
\]

directly design convolution kernel in the spectral domain

modified from Jonathan Masci et al
Spectral CNN

- We can define the Laplacian on an undirected graph:
  \[ \Delta = (I - \tilde{W}) \text{, } \tilde{W} = D^{-1/2}WD^{-1/2} \text{, } D = \text{diag}(W1) \]
  \[ (\Delta x)_k = x_k - \sum_j \tilde{w}_{kj}x_j \]
  measures smoothness in the graph

- \( \Delta \) is positive definite and symmetric. \( \Delta = V \text{diag}(\lambda)V^T \)

- “Fourier basis” of the graph: \( V \): Eigenvectors of \( \Delta \)

Joan Bruna et al. 2013
Spectral CNN

• “Convolution” on a graph: Linear Operator commuting with $\Delta$:

$$x *_G h := V \text{diag}(h) V^T x$$

– Filter coefficients $h$ are specified in the spectral domain.

• Spectral Network: filter bank $(x *_G h_k)_{k \leq K}$
Spectral CNN

- "Convolution" on a graph: Linear Operator commuting with $\Delta$:
  \[ x \ast_G h := V \text{diag}(h) V^T x \]
  - Filter coefficients $h$ are specified in the spectral domain.

- Spectral Network: filter bank $(x \ast_G h_k)_{k \leq K}$

- Needs $O(n)$ parameters per filter
- There's no guarantee the filter will have local support on the graph
Spectral CNN

• Observation:

  In Fourier analysis, smoothness and sparsity are dual notions
Spectral CNN

• Use smooth interpolation kernels (spline, polynomial, heat kernel, etc.) to parameterize the filters
Spectral CNN

• Use smooth interpolation kernels (spline, polynomial, heat kernel, etc.) to parameterize the filters spatially locally concentrated
Spectral CNN

- Use smooth interpolation kernels (spline, polynomial, heat kernel, etc.) to parameterize the filters

spatially locally concentrated

control #parameter
Spectral CNN

- Issues:
  - Convolution kernels are not shift-invariant.

A heat kernel translated to different vertices
Spectral CNN

• Issues:

• Convolution kernels are not shift-invariant.

• No effective pooling

• Filter weights depend on Fourier basis, does not generalize well to new domains
Spectral CNN

image from Jonathan Masci et al
Agenda

• Challenges
• Background knowledge
• Spatial construction
  • Geodesic CNN
• Spectral construction
  • Spectral CNN
  • Anisotropic CNN
  • SyncSpecCNN
By far, we are using isotropic filters

- Less descriptive, in analogy to circular filters in image CNN

```
circular filters
```

```
edge filters
```
Consider a specific type of interpolation kernels

- Heat kernel

\[ h_t(x, \xi) = \sum_{k \geq 0} e^{-t \lambda_k} \phi_k(x) \phi_k(\xi). \]
Consider a specific type of interpolation kernels

• Heat kernel - isotropic diffusion

\[ f_t(x) = -\text{div}_X(c \nabla_X f(x)) \]

\( c = \text{thermal diffusivity constant} \) describing heat conduction properties of the material (diffusion speed is equal everywhere)
Consider a specific type of interpolation kernels

- Heat kernel - anisotropic diffusion

\[ f_t(x) = -\text{div}_x(A(x)\nabla_x f(x)) \]

\( A(x) = \text{heat conductivity tensor} \) describing heat conduction properties of the material (diffusion speed is position + direction dependent)
Anisotropic diffusion on manifold

\[ f_t(x) = -\text{div}_X \left( \begin{bmatrix} \alpha \\ 1 \end{bmatrix} \nabla_X f(x) \right) \]
Anisotropic diffusion on manifold

\[ f_t(x) = -\text{div}_X \left( \begin{bmatrix} \alpha & 1 \end{bmatrix} \nabla_X f(x) \right) \]

\[ f_t(x) = -\text{div}_X \left( \begin{bmatrix} R_\theta & 1 \\ A_{\alpha\theta}(x) \end{bmatrix} \nabla_X f(x) \right) \]
Anisotropic diffusion on manifold

\[ f_t(x) = -\text{div}_X \left( \begin{bmatrix} \alpha \\ 1 \end{bmatrix} \nabla_X f(x) \right) \]

\[ f_t(x) = -\text{div}_X \left( R_{\theta} \left[ \begin{bmatrix} \alpha \\ 1 \end{bmatrix} \right] R_{\theta}^T \nabla_X f(x) \right) \]

- **Anisotropic Laplacian** \( \Delta_{\alpha \theta} f(x) = \text{div}_X (A_{\alpha \theta}(x) \nabla_X f(x)) \)
- \( \theta \) = orientation w.r.t. max curvature direction
- \( \alpha \) = ‘elongation’

Davide Boscaini et al. 2016
Anisotropic heat kernels
Anisotropic heat kernels

- Using anisotropic heat kernels to parameterize spectral filters is more descriptive
Anisotropic heat kernels

• Sensitive to noise (computing the directions of principle curvatures)
• Does not tackle the generalization issue
• No pooling structure
Agenda

• Challenges
• Background knowledge
• Spatial construction
  • Geodesic CNN
• Spectral construction
  • Spectral CNN
  • Anisotropic CNN
  • SyncSpecCNN
Spectral CNN

• Issues:

  • Convolution kernels are not shift-invariant.
  • No effective pooling
  • Filter weights depend on Fourier basis, does not generalize well to new domains

SyncSpecCNN

• Introduce spectral counterpart for spatial pooling
  • Synchronize Fourier basis for better generalizability
Spectral CNN

• Issues:
  • Convolution kernels are not shift-invariant.
  • No effective pooling
  • Filter weights depend on Fourier basis, does not generalize well to new domains

SyncSpecCNN

• Introduce spectral counterpart for spatial pooling
• Synchronize Fourier basis for better generalizability
Dilated convolution

- Achieving large receptive field quickly without pooling

Yu et al. 2016
SyncSpecCNN: spectral dilated convolution

- Parameterize filters with interpolation kernels.
- Shrink kernel bandwidth to increase spatial support of filters
SyncSpecCNN: spectral dilated convolution

- Parameterize filters with interpolation kernels.
- Shrink kernel bandwidth to increase spatial support of filters.
Spectral CNN

- Issues:
  - Convolution kernels are not shift-invariant.
  - No effective pooling.
  - Filter weights depend on Fourier basis, does not generalize well to new domains.

SyncSpecCNN

- Introduce spectral counterpart for spatial pooling.
- Synchronize Fourier basis for better generalizability.
Cross domain discrepancy

Spectral Domain 1

Spectral domain is independently defined for each shape graph

The same spectral function would induce very different spatial functions on different graphs

Cross domain parameter sharing is not valid
Cross domain discrepancy

Li Yi et al. 2017
Different domain needs to be synchronized

Spectral Domain 1  Canonical Domain  Spectral Domain 2

Li Yi et al. 2017
Functional map for domain synchronization

Li Yi et al. 2017
Functional map for domain synchronization

Li Yi et al. 2017
Spectral transformer network

Li Yi et al. 2017
Spectral transformer network

• Generates high dimensional transformation, sensitive to initialization (15x45 matrix)

• Pre-trained to get a good starting point

• Fine tuned with the end task learning
Synchronization visualization

before synchronization

after synchronization

Li Yi et al. 2017
SyncSpecCNN

part segmentation

key point prediction
Discussion

• Spatial construction is usually more efficient but less principled

• Spectral construction is more principled but usually slow (computing Laplacian eigenvectors for large scale data could be painful)

• On going research tries to bridge the gap

Convolutional Neural Networks on Graphs with Fast Localized Spectral Filtering, Defferrard et al. 2016

no need to compute eigen decomposition; reduce filtering complexity from $O(|\mathcal{V}| \cdot |\mathcal{V}_{\text{trunc}}|)$ to $O(|\mathcal{E}| \cdot K)$
Discussion

• Spatial construction is usually more efficient but less principled

• Spectral construction is more principled but usually slow (computing Laplacian eigenvectors for large scale data could be painful)

• On going research tries to bridge the gap

• Generalization issue on generic graphs is still a challenge