Geometric deep learning: going beyond Euclidean data

Michael M. Bronstein, Universita della Svizzera Italiana (Facebook Research)
Joan Bruna, Univ. Politec. de Catalunya (Facebook Research)
Yann LeCun, Facebook Research
Arthur Szlam, Facebook Research
Pierre Vandergheynst, Swiss Federal Institute of Technology

IEEE Signal Processing Magazine, 2017, 34(4):18-42.

SPECTRAL METHODS

- **Spectral CNN (SCNN)**

特征向量 describing the smooth structure of the graph

signal/function

Mapping

$$\mathbf{g}_l = \xi\left(\sum_{l'=1}^{q} \mathbf{\Phi}_k \mathbf{\Gamma}_{l,l'} \boxed{\mathbf{\Phi}_k^\top \mathbf{f}_{l'}}\right)$$

non-linearity

Filter

$\mathbf{\Gamma}_{l,l'}$ is a $k \times k$ diagonal matrix of spectral multipliers

- Using only the first *k* eigenvectors in sets a cutoff frequency which depends on the intrinsic regularity of the graph and also the sample size.

SPECTRAL METHODS

- **Spectral CNN (SCNN)**

特征向量 describing the smooth structure of the graph

signal/function

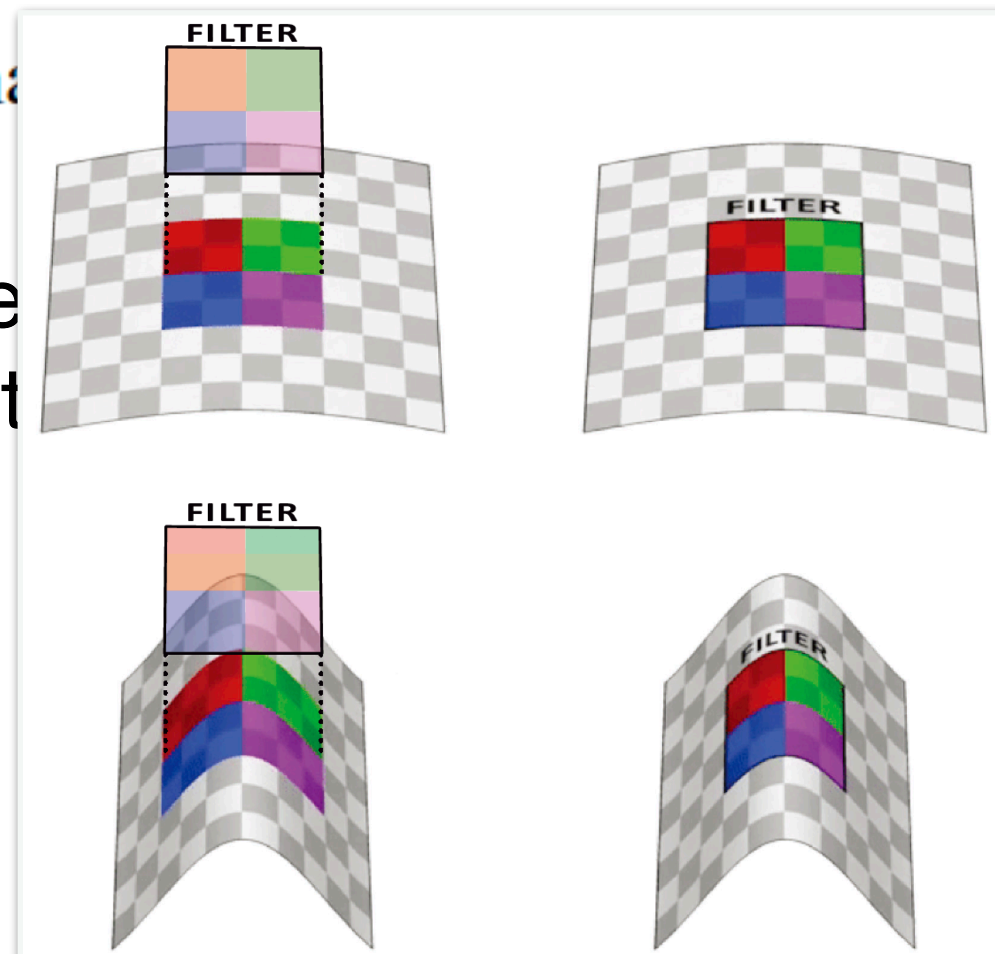
$$\mathbf{g}_l = \xi\left(\sum_{l'=1}^{q} \Phi_k \Gamma_{l,l'} \boxed{\Phi_k^\top \mathbf{f}_{l'}}\right)$$

Mapping

non-linearity

Filter

$\Gamma_{l,l'}$ is a $k \times k$ diagonal ma

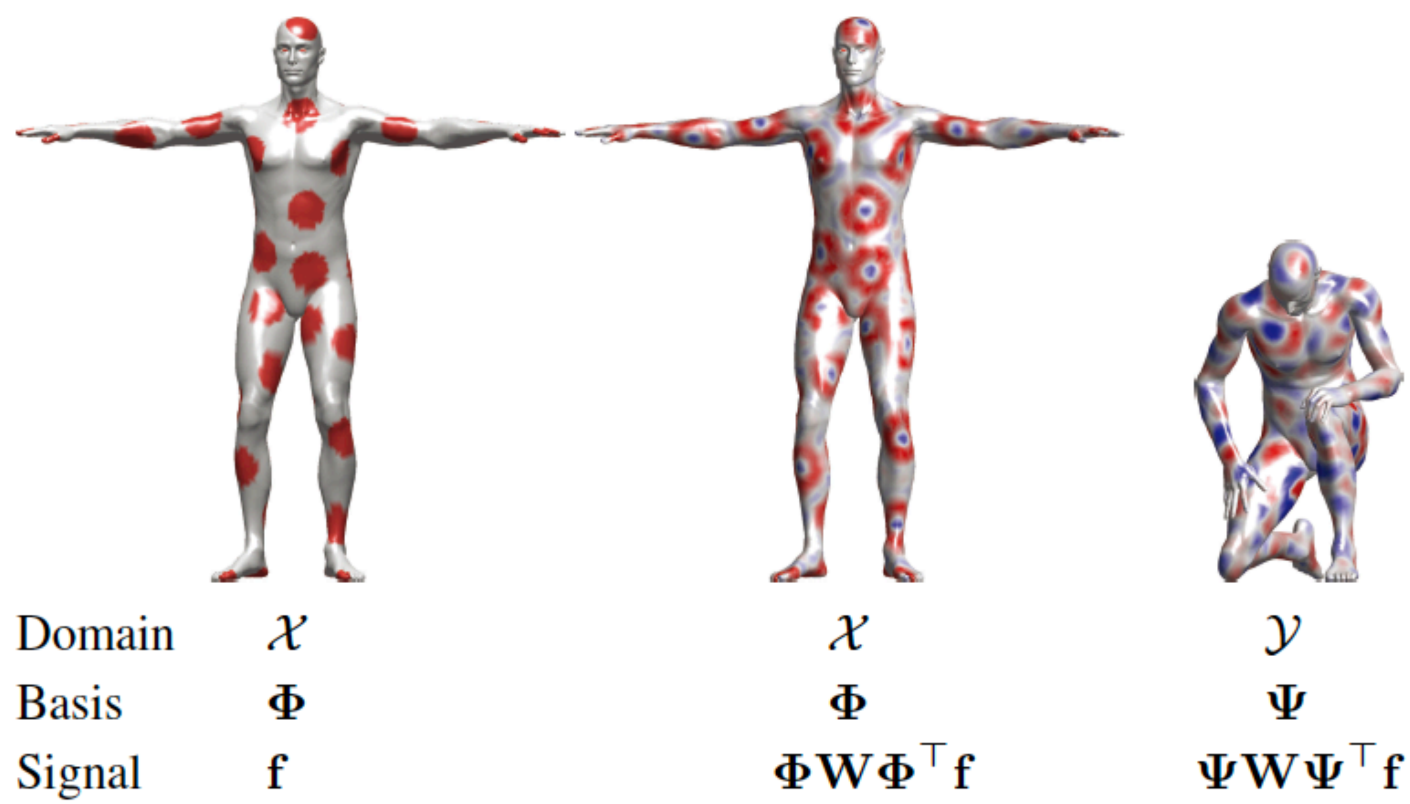


Euclidean CNN Geometric CNN

- Using only the first *k* eige equency which depends on the int ph and also the sample size.

SPECTRAL METHODS

- **Limitations of spectral CNN (SCNN)**
  - Basis dependent



  - Too much parameters

$$pqk = O(n)$$

SPECTRAL METHODS

- **Spectral CNN with smooth spectral multipliers**
  - To reduce the number of free parameters
  - Convolutional kernels with small spatial support (Euclidean domains)
    - The number of parameters independent of the input size
  - How to restrict the class of spectral multipliers?
- **Virtue of the Parseval Identity**

$$\int_{-\infty}^{+\infty} |x|^{2k} |f(x)|^2 dx = \int_{-\infty}^{+\infty} \left| \frac{\partial^k \hat{f}(\omega)}{\partial \omega^k} \right|^2 d\omega,$$

  - Smooth spectral multipliers = features (shared across locations and well localized)
  - Smoothness can be prescribed by learning only a subsampled set of frequency multipliers and using an interpolation kernel to obtain the rest, such as cubic splines.

SPECTRAL METHODS

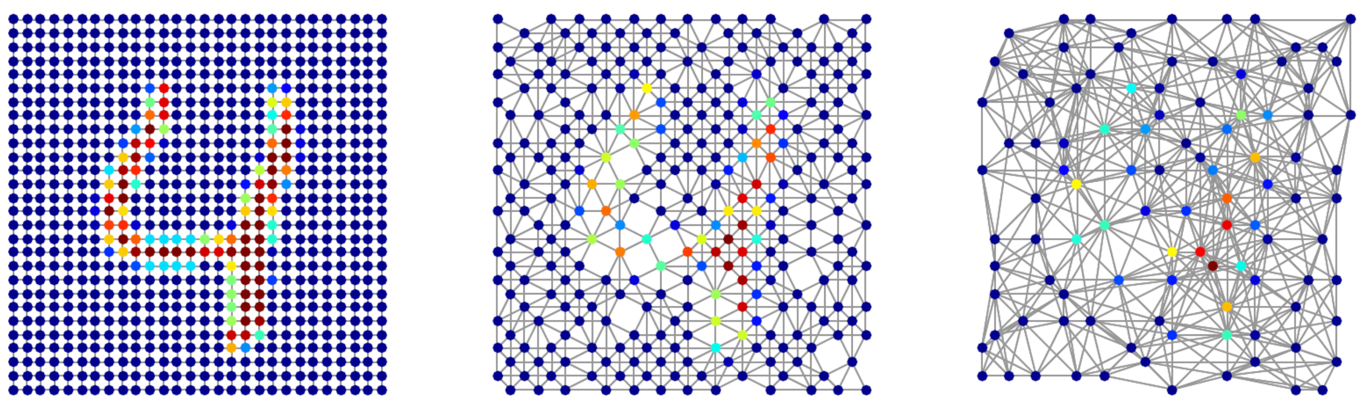
- **Spectral CNN with smooth spectral multipliers**
  - Interpolation kernel for the smoothness
    - to learn only a subsampled set of frequency multipliers and using an interpolation kernel to obtain the rest.

$$\mathrm{diag}(\underbrace{\boldsymbol{\Gamma}_{l,l'}}_{\text{Filter:frequency multipliers}}) = \underbrace{\mathbf{B}}_{\text{interpolation kernel}}\underbrace{\boldsymbol{\alpha}_{l,l'}}_{\text{interpolation coefficients}},$$

- Trainable parameters
  - O(log n), the same as CNNs on Euclidean grids

SPECTRAL METHODS

- **Graph coarsening**
  - The non-Euclidean analogy of pooling



$$\tilde{\mathbf{\Phi}} \approx \mathbf{P}\mathbf{\Phi} \begin{pmatrix} \mathbf{I}_{\alpha n} \\ \mathbf{0} \end{pmatrix}$$

Laplacian eigenvectors (Coarsening)

Laplacian eigenvectors (Original)

```
[1 1 0]
[1 1 0]
[0 0 4]
```

```
[4 0 0]
[0 1 1]
[0 1 1]
```

```
[ 0.70710678 -0.70710678  0.        ]
[ 0.70710678  0.70710678  0.        ]
[ 0.          0.          1.        ]
```

```
[ 0.          0.          1.        ]
[ 0.70710678  0.70710678  0.        ]
[ 0.70710678 -0.70710678  0.        ]
```

PECTRUM-FREE METHODS

- **Represent the filters via a polynomial expansion**

$$\boldsymbol{\Gamma}_{l,l'} = g_{\boldsymbol{\alpha}_{l,l'}}(\boldsymbol{\Lambda})$$

$$\mathbf{g}_l = \xi\left(\sum_{l'=1}^{q} \boldsymbol{\Phi}_k \boldsymbol{\Gamma}_{l,l'} \boldsymbol{\Phi}_k^\top \mathbf{f}_{l'}\right)$$

$$g_{\boldsymbol{\alpha}}(\boldsymbol{\Lambda}) \;=\; \mathrm{diag}(g_{\boldsymbol{\alpha}}(\lambda_1), \ldots, g_{\boldsymbol{\alpha}}(\lambda_n))$$

$$g_{\boldsymbol{\alpha}}(\lambda) = \sum_{j=0}^{r-1} \alpha_j \lambda^j$$

- It automatically yields localized filters
  - The Laplacian is a local operator, the action of its j-th power is constrained to j-hops.
  - The filter is a linear combination of powers of the Laplacian, overall behaves like a diffusion operator limited to r-hops around each vertex.

PECTRUM FREE METHODS

- **Graph CNN (GCNN) a.k.a. ChebNet**
  - Chebyshev polynomial

$$
\begin{aligned}
T_j(\lambda) &= 2\lambda T_{j-1}(\lambda) - T_{j-2}(\lambda); \\
T_0(\lambda) &= 1; \\
T_1(\lambda) &= \lambda.
\end{aligned}
$$

Recurrence relation

  - A filter can be parameterized uniquely via an expansion

$$
\begin{aligned}
g_{\boldsymbol{\alpha}}(\tilde{\boldsymbol{\Delta}}) &= \sum_{j=0}^{r-1} \alpha_j \boldsymbol{\Phi} T_j(\tilde{\boldsymbol{\Lambda}}) \boldsymbol{\Phi}^\top \\
&= \sum_{j=0}^{r-1} \alpha_j T_j(\tilde{\boldsymbol{\Delta}}),
\end{aligned}
$$

$$
\bar{\mathbf{f}}^{(j)} = T_j(\tilde{\boldsymbol{\Delta}})\mathbf{f}
$$

$$
\bar{\mathbf{f}}^{(j)} = 2\tilde{\boldsymbol{\Delta}}\bar{\mathbf{f}}^{(j-1)} - \bar{\mathbf{f}}^{(j-2)}
$$

Computational complexity is O(rn)
Does not require an explicit computation of the Laplacian eigenvectors

PECTRUM-FREE METHODS

- **Graph Convolutional Network**
  - Assuming r = 2 and $\lambda_n \approx 2$
  - One obtains filters represented by a single parameter

$$g_\alpha(\mathbf{f}) \quad = \quad \alpha(\mathbf{I} + \mathbf{D}^{-1/2}\mathbf{W}\mathbf{D}^{-1/2})\mathbf{f}.$$

  - ChebNet and GCN (spectral domain) boil down to applying simple filters acting on the r-or 1-hop neighborhood of the graph in the spatial domain.

PECTRUM-FREE METHODS

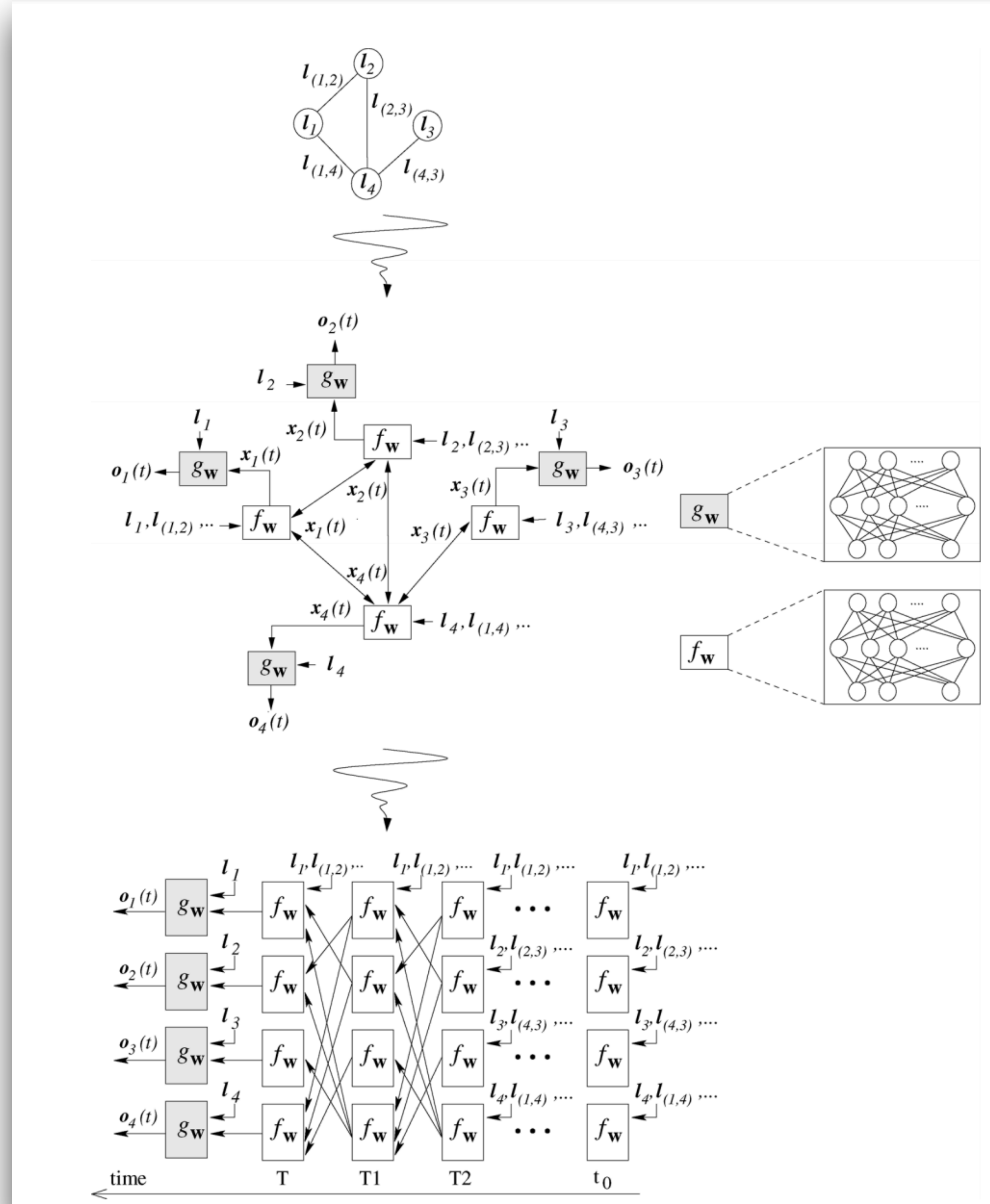
- **Graph Neural Network (GNN)**
  - Graph Neural Networks generalize the notion of applying the filtering operations directly on the graph via the graph weights.

$$\mathbf{g}_i = \eta_{\boldsymbol{\theta}}\left((\mathbf{W}\mathbf{f})_i, (\mathbf{D}\mathbf{f})_i\right)$$

  - Laplacian operator $\eta(\mathbf{a}, \mathbf{b}) = \mathbf{b} - \mathbf{a}$
  - Nonlinear choices for η yield trainable, task-specific diffusion operators

PECTRUM-FREE METHODS

- **Graph Neural Network**
  -



CHARTING-BASED METHODS

- Inherent drawback of inability to adapt the model across different domains
  - Resort to the convolution in the spatial domain
  - One of the major problems in applying the same paradigm to non-Euclidean domains is the lack of shift-invariance, implying that the '**patch operator**' extracting a local 'patch' would be position-dependent.

$$D_j(x)f = \int_{\mathcal{X}} f(x')v_j(x,x')dx', \quad j = 1, \ldots, J,$$

$$(f \star g)(x) = \sum_j g_j D_j(x) f,$$

CHARTING-BASED METHODS

- Geodesic CNN

$$v_{ij}(x, x') = e^{-(\rho(x') - \rho_i)^2 / 2\sigma_\rho^2} \, e^{-(\theta(x') - \theta_j)^2 / 2\sigma_\theta^2}$$

- Anisotropic CNN

$$\Delta_{\alpha\theta} f(x) = -\mathrm{div}(\mathbf{A}_{\alpha\theta}(x) \nabla f(x))$$

- Mixture model network (MoNet)

$$v_j(\mathbf{u}) = \exp\left(-\tfrac{1}{2}(\mathbf{u} - \boldsymbol{\mu}_j)^\top \boldsymbol{\Sigma}_j^{-1} (\mathbf{u} - \boldsymbol{\mu}_j)\right)$$

COMBINED SPATIAL /SPECTRAL METHODS

- Windowed Fourier transform

$$
\begin{aligned}
(Sf)(x,\omega) &= \int_{-\infty}^{\infty} f(x')\, \underbrace{g(x'-x)e^{-i\omega x'}}_{\overline{g_{x,\omega}}(x')}\, dx' \\
&= \langle f, g_{x,\omega} \rangle_{L^2(\mathbb{R})}.
\end{aligned}
$$

spatial location of the window x

modulation frequency w

window function

- Wavelets (Haar wavelets)
- Localized Spectral CNN (LSCNN)

CHARTING-BASED METHODS

Dichotomy of Geometric deep learning methods

Method	Domain	Data
Spectral CNN [52]	spectral	graph
GCNN/ChebNet [45]	spec. free	graph
GCN [77]	spec. free	graph
GNN [78]	spec. free	graph
Geodesic CNN [47]	charting	mesh
Anisotropic CNN [48]	charting	mesh/point cloud
MoNet [54]	charting	graph/mesh/point cloud
LSCNN [89]	combined	mesh/point cloud

APPLICATIONS

- Network analysis
  - Classification application (citation network)
  - Ranking (PageRank algorithm) and community detection.
  - Recommender systems (matrix completion)
    - Geometric matrix completion
    - Multi-Graph CNN
  - Computer vision and graphics
  - Particle physics and Chemistry (Classification)
    - Chemical properties of a molecule are determined by the relative positions of its atoms
  - Molecule design
    - A key problem in material- and drug design is predicting the physical, chemical, or biological properties of a novel molecule
  - Medical imaging (non-Euclidean domains)

OPEN PROBLEMS AND FUTURE DIRECTIONS

- **Generalization**
  - Spectral formulation of convolution allows designing CNNs on a graph, but the model learned this way on one graph cannot be straightforwardly applied to another one, since the spectral representation of convolution is domain-dependent.
    - Spatial transformer networks
  - The spatial methods, on the other hand, allow generalization across different domains
- **Time-varying domains**
- **Directed graphs**
- **Synthesis problems**
- **Computation**