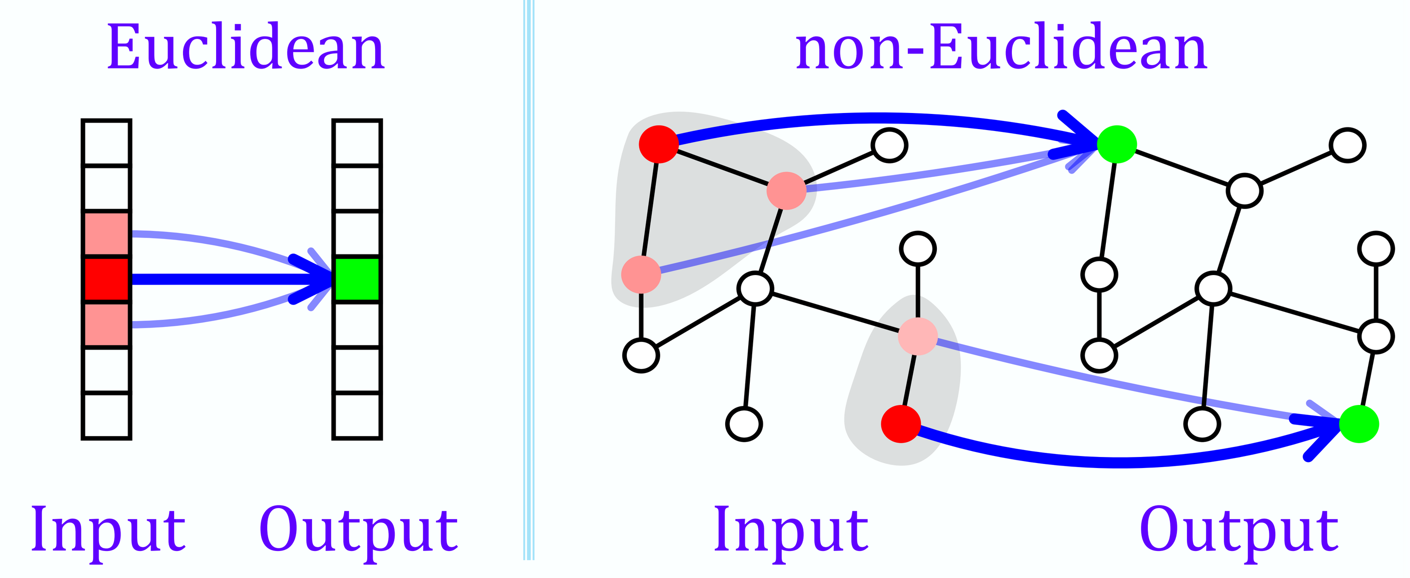
Structure-Aware Convolutional Neural Networks

32nd Conference on Neural Information Processing Systems (NIPS 2018), Montréal, Canada

Jianlong Chang, Jie Gu, Lingfeng Wang, Gaofeng Meng, Shiming Xiang, Chunhong Pan
{jianlong.chang, jie.gu, lfwang, gfmeng, smxiang, chpan}@nlpr.ia.ac.cn

Convolution

Intrinsically, convolution can be considered as an aggregation operation between local inputs and filters. By modeling the local structures into convolution, convolution is generalized on both Euclidean and non-Euclidean spaces.

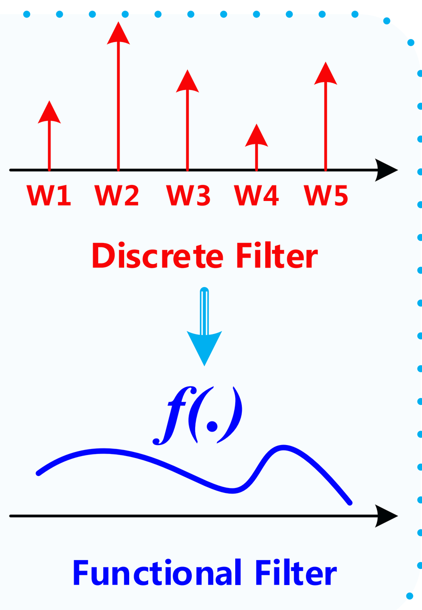


input $\mathbf{x} \in \mathbb{R}^n$ filter $\mathbf{w} \in \mathbb{R}^{2m-1}$

$$y_i = \mathbf{w}^{\mathrm{T}}\mathbf{x}_i = \sum_{i-m<j<i+m} w_{j-i+m} \cdot x_j, \quad i \in \{1, 2, \cdots, n\}$$

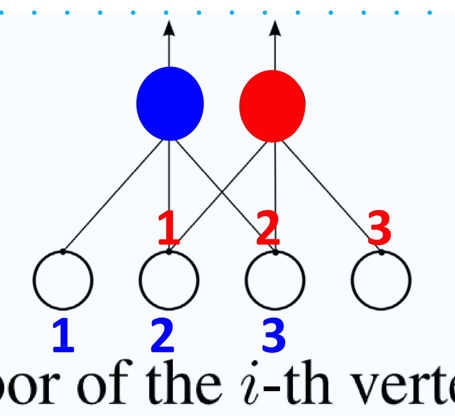
For any univariate function $f(\cdot)$ $\quad f(j-i+m) = w_{j-i+m}$

$$y_i = \mathbf{f}_{\mathcal{R}}^{\mathrm{T}}\mathbf{x}_i = \sum_{i-m<j<i+m} f(j-i+m) \cdot x_j, \quad i \in \{1, 2, \cdots, n\}$$



$$y_i = \mathbf{f}_{\mathcal{R}}^{\mathrm{T}}\mathbf{x}_i = \sum_{i-m<j<i+m} f(j-i+m) \cdot x_j, \quad i \in \{1, 2, \cdots, n\}$$

$$\mathcal{R} = \{j-i+m \mid i-m<j<i+m\} = \{1, 2, \cdots, 2m-1\}$$



$r \in \mathcal{R}$ means that the $(i-m+r)$-th vertex is the r-th neighbor of the i-th vertex

$\mathcal{R}$ encodes relationships between a vertex and its neighbors

$$y_i = \mathbf{f}_{\mathcal{R}_i}^{\mathrm{T}}\mathbf{x}_i = \sum_{e_{ji}\in\mathcal{E}} f(r_{ji}) \cdot x_j, \quad i \in \{1, 2, \cdots, n\}$$

Parametrization

The functional filter is learned with the Chebyshev polynomials basis functions, *i.e.*,

$$f(r_{ji}) = \sum_{k=1}^{t} v_k \cdot h_k(r_{ji})$$

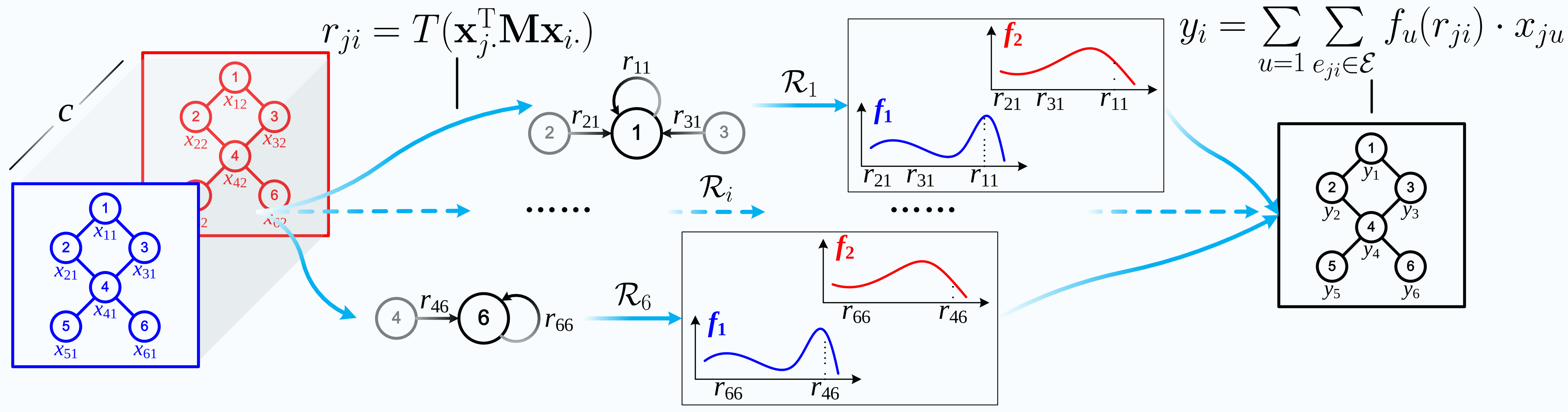
Theorem 1. *Under the Chebyshev polynomial basis, the structure-aware convolution is equivalent to*

$$y_i = \mathbf{v}^{\mathrm{T}}\mathbf{P}_i\mathbf{x}_i, \quad i \in \{1, 2, \cdots, n\},$$

where $\mathbf{v} \in \mathbb{R}^t$ *is the coefficients of the polynomials,* $\mathbf{P}_i \in \mathbb{R}^{t\times m}$ *is a matrix determined by the local structure representation* $\mathcal{R}_i$ *and the polynomials, and* $\mathbf{x}_i \in \mathbb{R}^m$ *is the local input at the* i*-th vertex.*

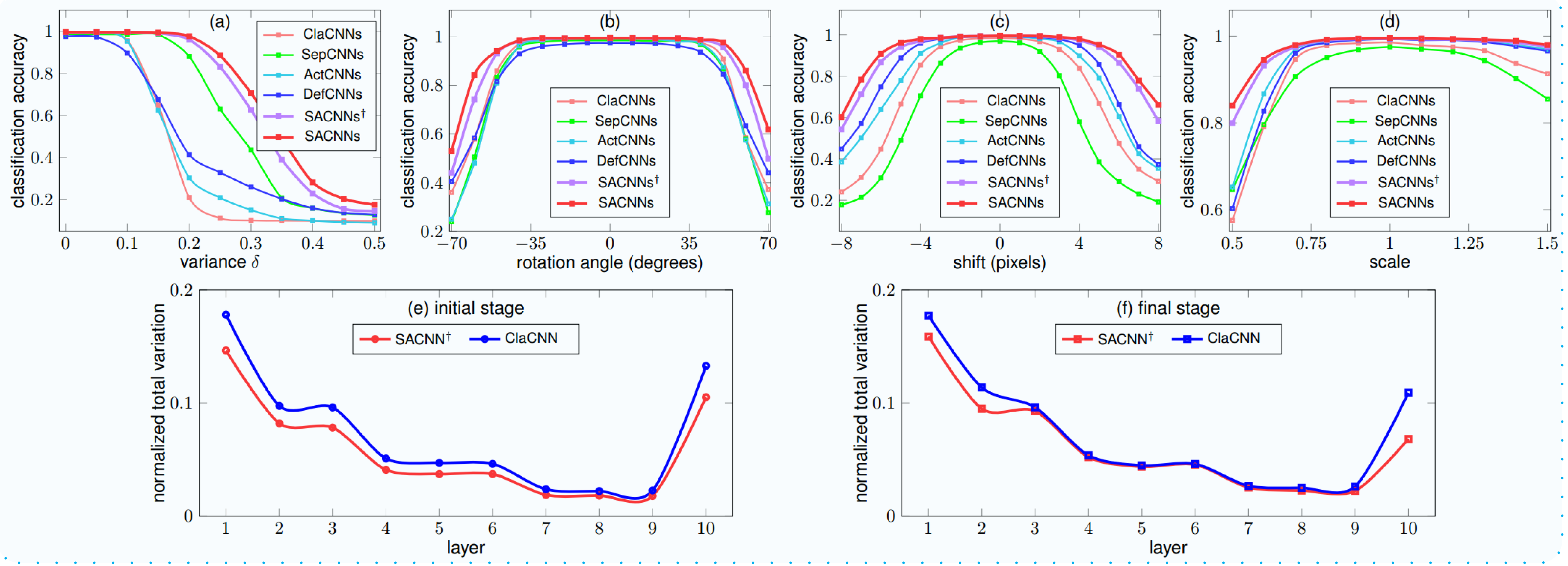
SACNNs

A structure-aware convolutional layer with two input channels and one output channel. First, the local structure is estimated by considering the relationships between vertices. Then, every local inputs can be aggregated based on local structures.



Robustness

Fig (a-d) prove that our SACNNs are in possession of excellent robustness to Gaussion noise, rotation, shift, and scale. Fig (e-f) show that the filters learned in our SACNNs are more smoother than in traditional CNNs, which implies that higher deformation stability will be achieved when smoother filters are learned.



Ablation study

(a) Impact of polynomial order. (b) Influence of channels. (c) Transfer learning from Reuters to 20News. (d) Impact of training samples. (e) Influence of basis functions. (f) Integration with recent networks. (g) Sensitivity to initialization. (h) Parameters distribution.