Few-shot learning with graph neural networks

At ICLR 2018

25 May 2018

Meta-learning Introduction

Meta-learning Introduction

- deep neural networks will severely overfit on one-shot
- humans spend a lifetime to classify things
- knowledge transfer from other tasks

Meta-learning Introduction

- The model is trained to learn tasks in the meta-training set.

Meta-learning Introduction

- two optimizations
  - the learner, which learns new tasks
  - the meta-learner, which trains the learner.

- three categories:
  - recurrent models
  - metric learning
  - learning optimizers

Problem Set-up

$$\mathcal{T} = \{\{(x_1, l_1), \ldots, (x_s, l_s)\}, \{\tilde{x}_1, \ldots, \tilde{x}_r\}, \{(\tilde{x}_1, \ldots, \tilde{x}_t)\}\}$$

$$Y = (y_1, \ldots, y_t) \in \{1, K\}^t$$

Labeled: $\{(x_1, l_1), \ldots, (x_s, l_s)\}$ for supervised learning

Unlabeled: $\{\tilde{x}_1, \ldots, \tilde{x}_r\}$ for semi-supervised and active learning

To classify: $\{(\tilde{x}_1, \ldots, \tilde{x}_t)\}$

Learning objective

$$\min_\Theta \frac{1}{L} \sum_{i \leq L} \mathcal{L}(\Phi(\mathcal{T}_i; \Theta), Y_i) + \mathcal{R}(\Theta)$$

Few-shot GNN

- Goal: propagate label information from labeled samples towards the unlabeled query image

- Idea: a posterior inference over a graphical model determined by the input images and labels

Few-shot GNN

- Contributions

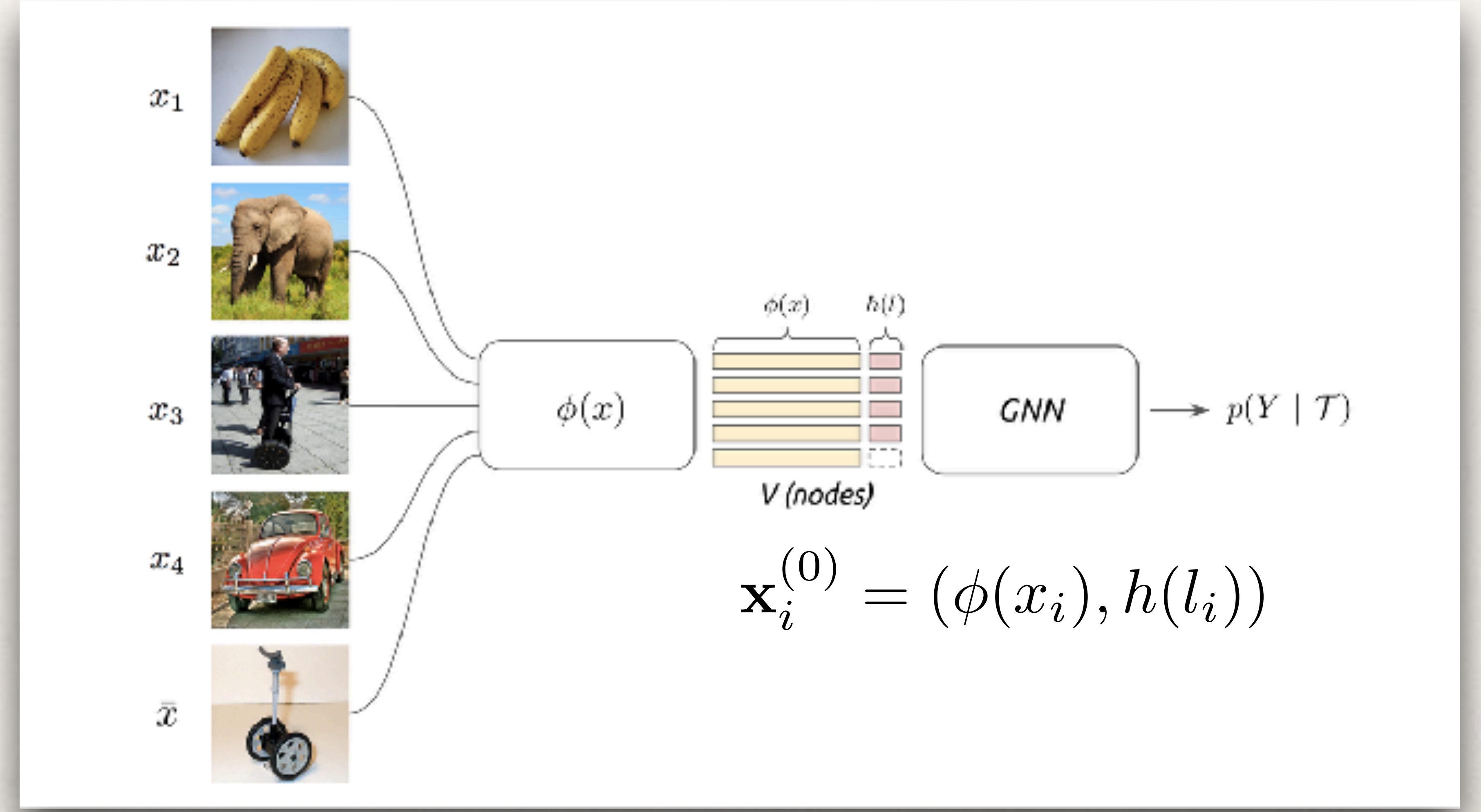
1. supervised message passing through end-to-end graph neural networks

2. state-of-the-art performance with fewer parameters

3. extend to semi-supervised and active learning

Model: node features

- Initial Node Features



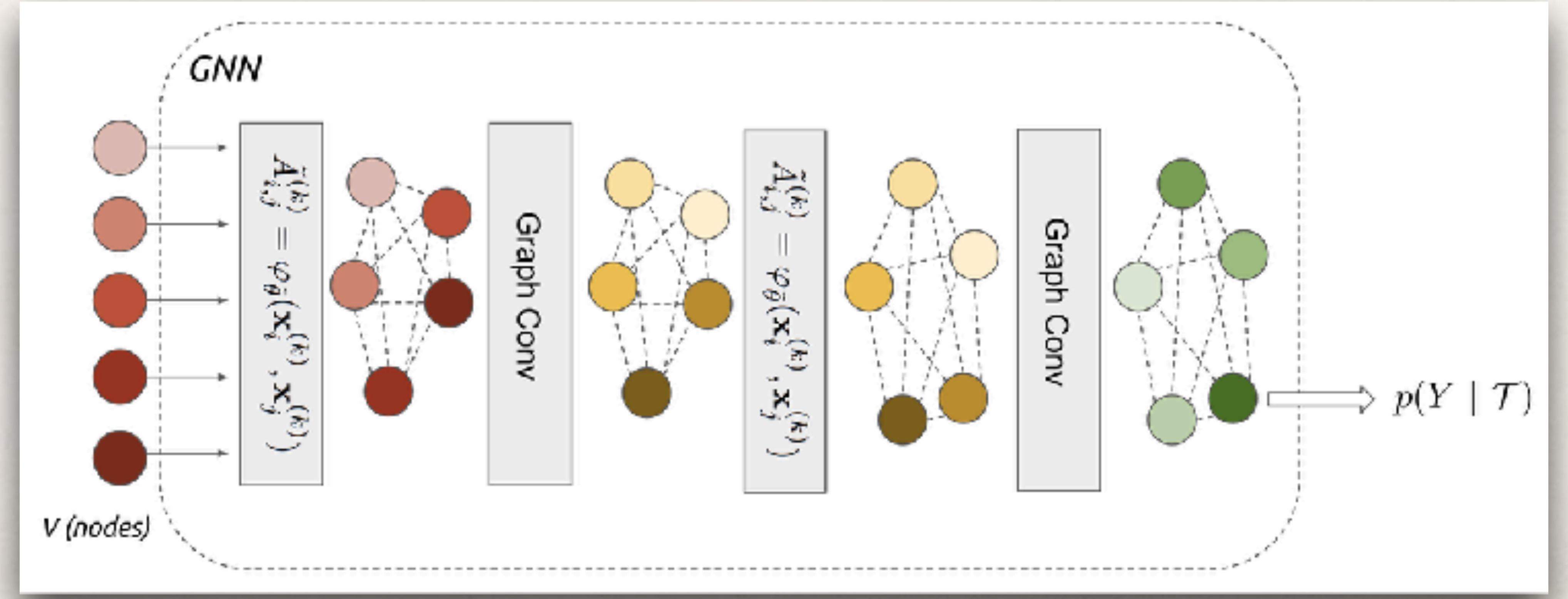
ϕ is CNN; $h(l) \in \mathbb{R}_+^K$ is one-hot encoding

Model: metric

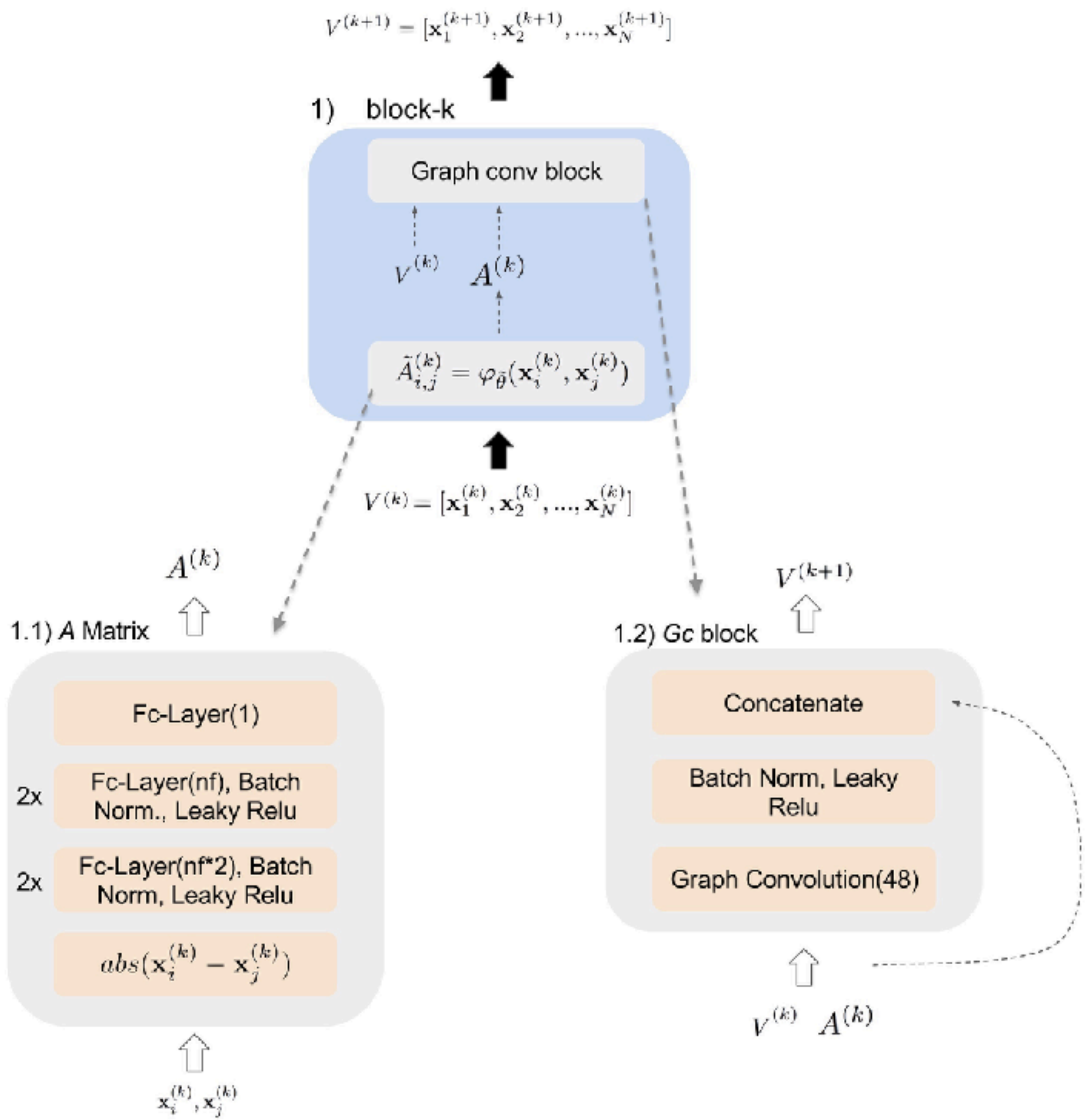
$$V = \{X_1, X_2, \ldots, X_N\}$$

$$\tilde{A}_{i,j}^{(k)} = \varphi_{\tilde{\theta}}(\mathbf{x}_i^{(k)}, \mathbf{x}_j^{(k)})$$

$$\varphi_{\tilde{\theta}}(\mathbf{x}_i^{(k)}, \mathbf{x}_j^{(k)}) = \mathrm{MLP}(abs(\mathbf{x}_i^{(k)} - \mathbf{x}_j^{(k)}))$$



Model: GNN



$V^{(k+1)} = [\mathbf{x}_1^{(k+1)}, \mathbf{x}_2^{(k+1)}, ..., \mathbf{x}_N^{(k+1)}]$
1) block-k
Graph conv block
$V^{(k)}$
$A^{(k)}$
$\tilde{A}_{i,j}^{(k)} = \varphi_{\tilde{\theta}}(\mathbf{x}_i^{(k)}, \mathbf{x}_j^{(k)})$
$V^{(k)} = [\mathbf{x}_1^{(k)}, \mathbf{x}_2^{(k)}, ..., \mathbf{x}_N^{(k)}]$
$A^{(k)}$
1.1) *A* Matrix
Fc-Layer(1)
2x
Fc-Layer(nf), Batch Norm., Leaky Relu
2x
Fc-Layer(nf*2), Batch Norm, Leaky Relu
$abs(\mathbf{x}_i^{(k)} - \mathbf{x}_j^{(k)})$
$\mathbf{x}_i^{(k)}, \mathbf{x}_j^{(k)}$
$V^{(k+1)}$
1.2) *Gc* block
Concatenate
Batch Norm, Leaky Relu
Graph Convolution(48)
$V^{(k)}$
$A^{(k)}$

Experiments

- Omniglot

  1623 hand drawn characters from 50 alphabets,
  every character has 20 examples, at resolution 105x105

	5-Way		20-Way	
Model	**1-shot**	**5-shot**	**1-shot**	**5-shot**
Pixels Vinyals et al. (2016)	41.7%	63.2%	26.7%	42.6%
Siamese Net Koch et al. (2015)	97.3%	98.4%	88.2%	97.0%
Matching Networks Vinyals et al. (2016)	98.1%	98.9%	93.8%	98.5%
N. Statistician Edwards & Storkey (2016)	98.1%	99.5%	93.2%	98.1%
Res. Pair-Wise Mehrotra & Dukkipati (2017)	-	-	94.8%	-
Prototypical Networks Snell et al. (2017)	97.4%	99.3%	95.4%	98.8%
ConvNet with Memory Kaiser et al. (2017)	98.4%	99.6%	95.0%	98.6%
Agnostic Meta-learner Finn et al. (2017)	98.7 ±0.4%	99.9 ±0.3%	95.8 ±0.3%	98.9 ±0.2%
Meta Networks Munkhdalai & Yu (2017)	98.9%	-	97.0%	-
TCML Mishra et al. (2017)	98.96% ±0.20%	99.75% ±0.11%	97.64% ±0.30%	99.36% ±0.18%
Our GNN	99.2%	99.7%	97.4%	99.0%

Experiments

- Mini-ImageNet

  60,000 colorful images of size 84 × 84 with 100 classes, each having 600 examples

Model	5-Way 1-shot	5-Way 5-shot
Matching Networks Vinyals et al. (2016)	43.6%	55.3%
Prototypical Networks Snell et al. (2017)	46.61% ±0.78%	65.77% ±0.70%
Model Agnostic Meta-learner Finn et al. (2017)	48.70% ±1.84%	63.1% ±0.92%
Meta Networks Munkhdalai & Yu (2017)	49.21% ±0.96	-
Ravi & Larochelle Ravi & Larochelle (2016)	43.4% ±0.77%	60.2% ±0.71%
TCML Mishra et al. (2017)	55.71% ±0.99%	68.88% ±0.92%
Our metric learning + KNN	49.44% ±0.28%	64.02% ±0.51%
Our GNN	50.33% ±0.36%	66.41% ±0.63%

Summary

- graph neural representation for few-shot, semi-supervised, and active learning
- two optimization: node metric + GNN learner
- comparable results