Edge-Labeling Graph Neural Network for Few-shot Learning

Jongmin Kim*[1,3], Taesup Kim[2,3], Sungwoong Kim[3], and Chang D.Yoo[1]

[1]Korea Advanced Institute of Science and Technology
[2]MILA, Université de Montréal
[3]Kakao Brain

QikuiZhu 09/25/2019

Kim, Jongmin, et al. "Edge-Labeling Graph Neural Network for Few-shot Learning." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019.

Contribution

1. The edge-labeling GNN (EGNN) is first proposed for few-shot learning with iteratively updating edge-labels with exploitation of both intra-cluster similarity and inter-cluster dissimilarity.

2. It consists of a number of layers in which each layer is composed of a node-update block and an edge-update block where the corresponding parameters are estimated under the episodic training framework.

3. On both of the supervised and semi-supervised few-shot image classification tasks with two benchmark datasets, the proposed EGNN significantly improves the performances over the existing GNNs. Additionally, several ablation experiments show the benefits from the explicit clustering as well as the separate utilization of intra-cluster similarity and inter-cluster dissimilarity.

Few-shot Classification

The few-shot classification aims to learn a classifier when only a few training samples per each class are given. Therefore, each few-shot classification task T contains a **support set S**, a labeled set of input-label pairs, and a **query set Q**, an unlabeled set on which the learned classifier is evaluated.

If the **support set S** contains K labeled samples for each of N unique classes, the problem is called **N-way K-shot** classification problem.

Meta-learning has become a standard methodology to tackle few-shot classification.

Episodic training is an efficient way of meta-learning, which is commonly employed in few-shot classification.

Episodic Training

Both training and test tasks are the **N-way K-shot problem.**

$$\mathcal{T} = \mathcal{S} \bigcup \mathcal{Q}$$

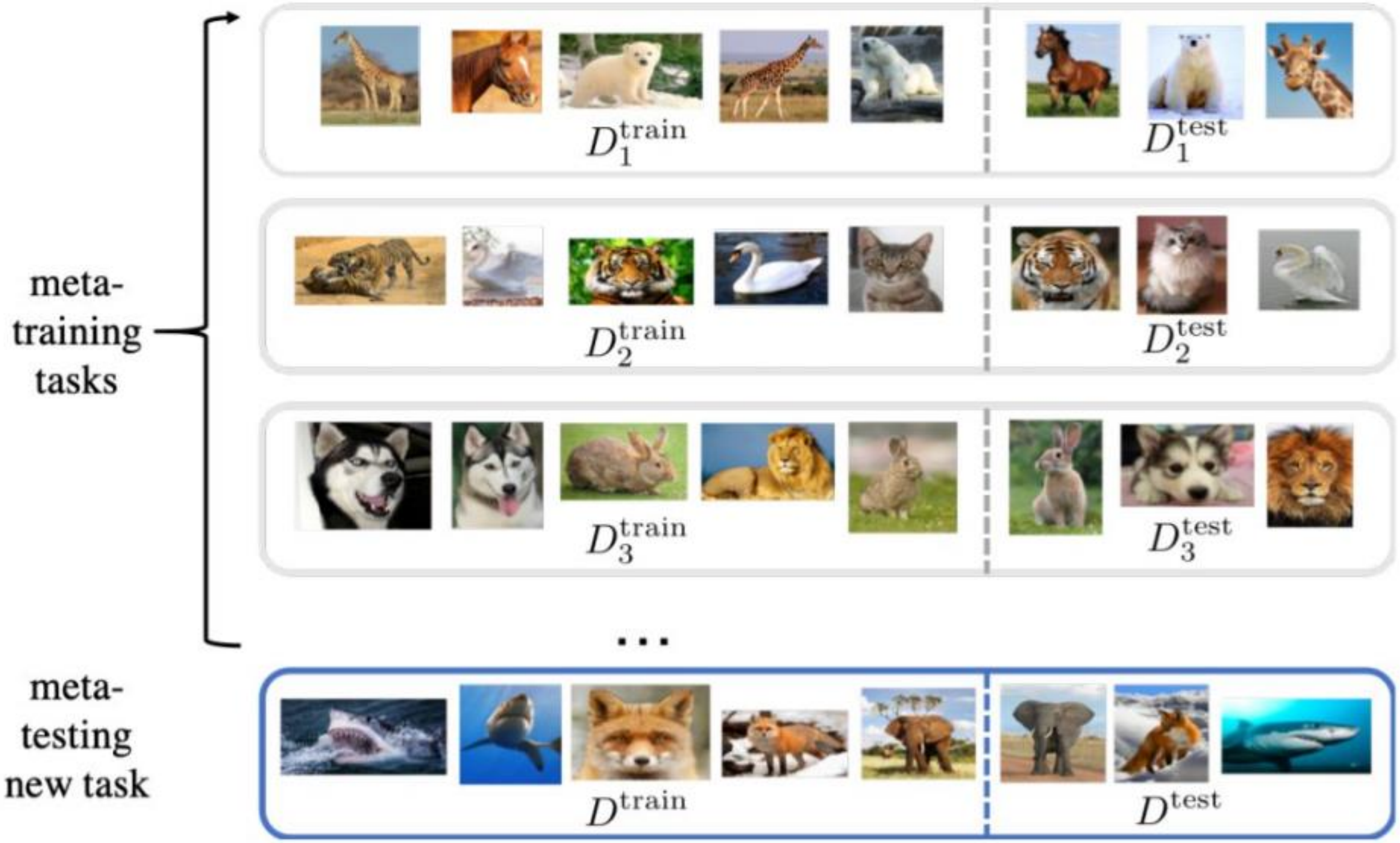
support set $\mathcal{S} = \{(\mathbf{x}_i, y_i)\}_{i=1}^{N \times K}$

query set $\mathcal{Q} = \{(\mathbf{x}_i, y_i)\}_{i=N \times K+1}^{N \times K+T}$ T is the number of query samples.

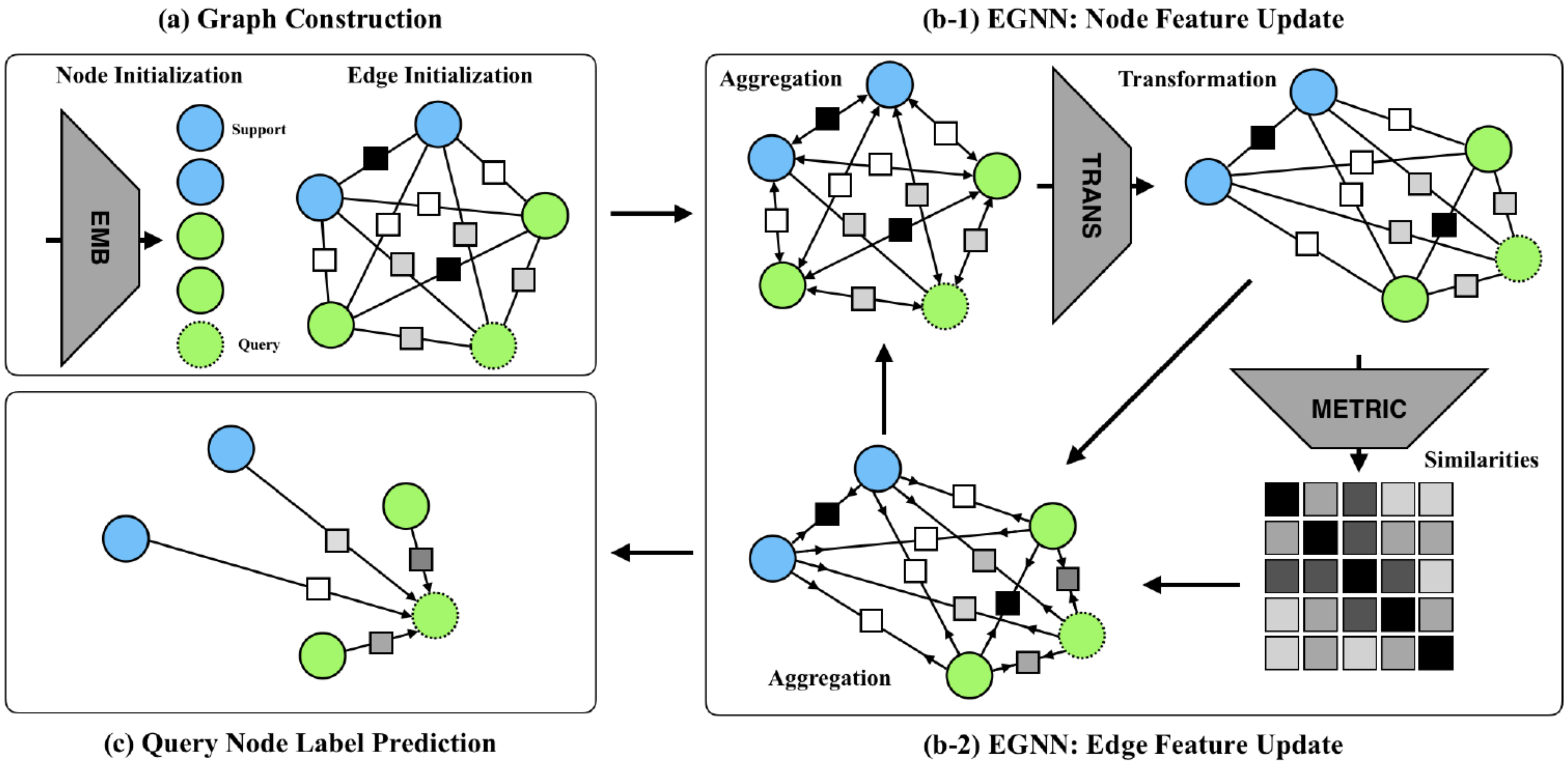
$y_i \in \{C_1, \cdots C_N\} = \mathcal{C}_\mathcal{T} \subset \mathcal{C}$ C is the set of all classes of either training or test dataset.

$\mathcal{C}_{train} \bigcap \mathcal{C}_{test} = \emptyset.$ Although both the training and test tasks are sampled from the common task distribution, the label spaces are mutually exclusive.

Episodic Training



EGNN



The overall framework of the proposed EGNN model. In this illustration, a 2-way 2-shot problem is presented as an example. Blue and green circles represent two different classes. Nodes with solid line represent labeled support samples, while a node with dashed line represents the unlabeled query sample. The strength of edge feature is represented by the color in the square. Note that although each edge has a 2-dimensional feature, only the first dimension is depicted for simplicity.

Methods

$$\mathcal{G} = (\mathcal{V}, \mathcal{E}; \mathcal{T})$$

$$\mathcal{V} := \{V_i\}_{i=1,\ldots,|\mathcal{T}|}$$

$$\mathcal{E} := \{E_{ij}\}_{i,j=1,\ldots,|\mathcal{T}|}$$

denote the set of nodes and edges of the graph

$\mathbf{v}_i$ $\mathbf{e}_{ij}$ are the node feature and the edge feature

$|\mathcal{T}| = N \times K + T$ is the total number of samples in the task

$$y_{ij} = \begin{cases} 1, & \text{if } y_i = y_j, \\ 0, & \text{otherwise.} \end{cases}$$

is the ground truth edge-label defined by the ground-truth node labels

Methods

Node features are initialized by the output of the pre-trained convolutional embedding network.

Edge feature

$$\mathbf{e}_{ij} = \{e_{ijd}\}_{d=1}^{2} \in [0,1]^2$$

is a 2-dimensional vector representing the (normalized) strengths of the intra- and inter-class relations of the two connected nodes. Edge features are initialized by edge labels.

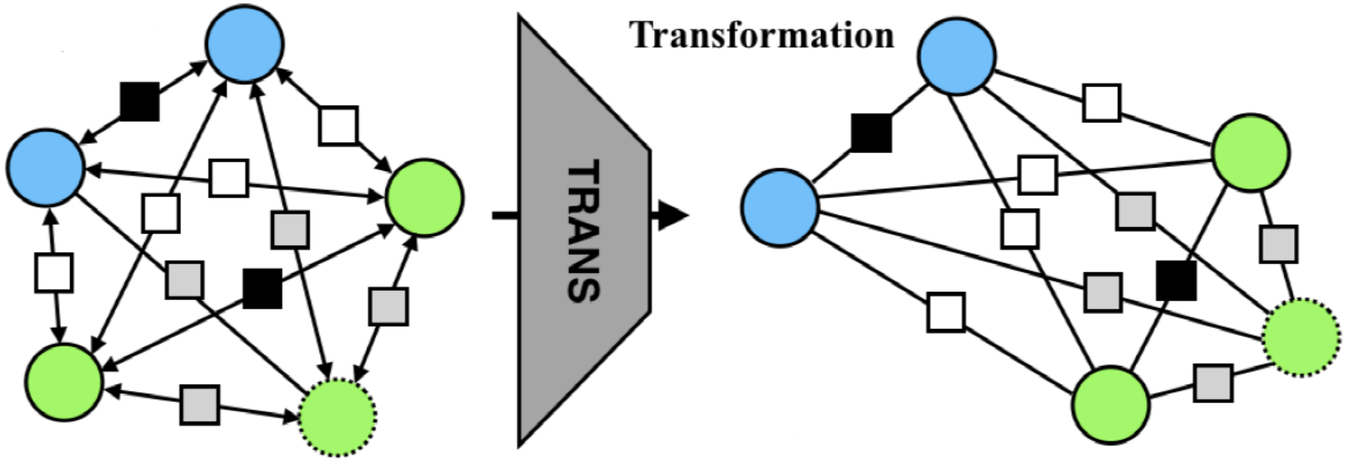
$$\mathbf{e}_{ij}^{0} = \begin{cases} [1||0], & \text{if } y_{ij} = 1 \text{ and } i,j \leq N \times K, \\ [0||1], & \text{if } y_{ij} = 0 \text{ and } i,j \leq N \times K, \\ [0.5||0.5], & \text{otherwise}, \end{cases}$$

$$y_{ij} = \begin{cases} 1, & \text{if } y_i = y_j, \\ 0, & \text{otherwise}. \end{cases}$$

is the ground truth edge-label defined by the ground-truth node labels

Methods

Node Features Update



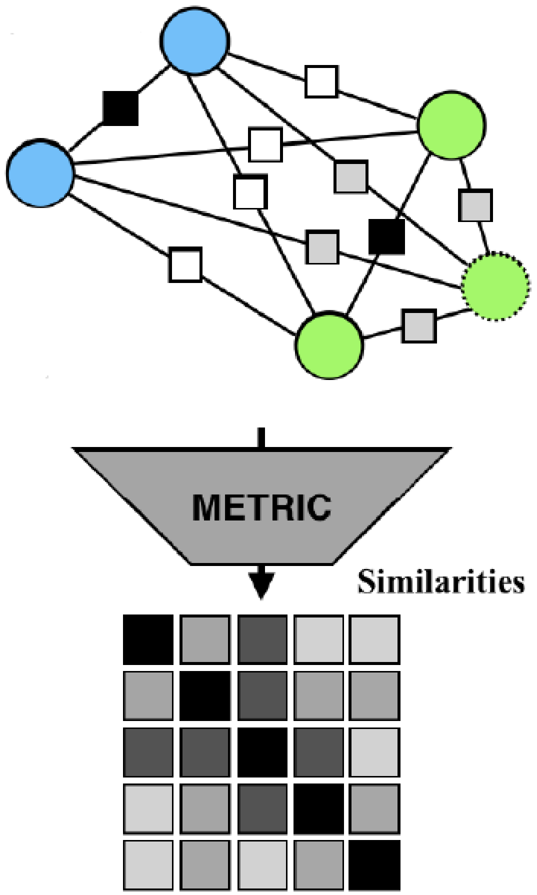
$$\mathbf{v}_i^\ell = f_v^\ell([\sum_j \tilde{e}_{ij1}^{\ell-1}\mathbf{v}_j^{\ell-1} || \sum_j \tilde{e}_{ij2}^{\ell-1}\mathbf{v}_j^{\ell-1}]; \theta_v^\ell)$$

$\tilde{e}_{ijd} = \frac{e_{ijd}}{\sum_k e_{ikd}}$ normalized edge feature

f_v^ℓ is the feature (node) transformation network

Methods

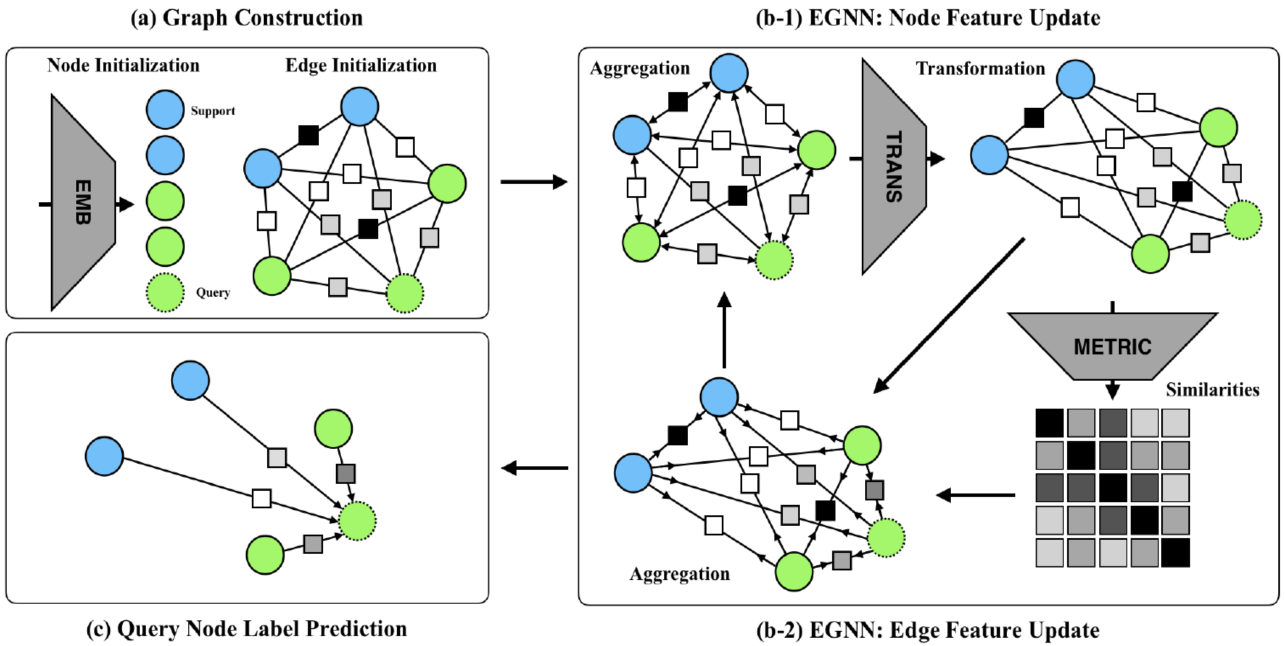
Edge Feature Update



$$
\begin{aligned}
\bar{e}^{\ell}_{ij1} &= \frac{f^{\ell}_e(\mathbf{v}^{\ell}_i, \mathbf{v}^{\ell}_j; \theta^{\ell}_e) e^{\ell-1}_{ij1}}{\sum_k f^{\ell}_e(\mathbf{v}^{\ell}_i, \mathbf{v}^{\ell}_k; \theta^{\ell}_e) e^{\ell-1}_{ik1} / (\sum_k e^{\ell-1}_{ik1})}, \\
\bar{e}^{\ell}_{ij2} &= \frac{(1 - f^{\ell}_e(\mathbf{v}^{\ell}_i, \mathbf{v}^{\ell}_j; \theta^{\ell}_e)) e^{\ell-1}_{ij2}}{\sum_k (1 - f^{\ell}_e(\mathbf{v}^{\ell}_i, \mathbf{v}^{\ell}_k; \theta^{\ell}_e)) e^{\ell-1}_{ik2} / (\sum_k e^{\ell-1}_{ik2})} \\
\mathbf{e}^{\ell}_{ij} &= \bar{\mathbf{e}}^{\ell}_{ij} / \|\bar{\mathbf{e}}^{\ell}_{ij}\|_1,
\end{aligned}
$$

f^{ℓ}_e is the metric network that computes similarity scores

EGNN



Algorithm 1: The process of EGNN for inference

Input: $\mathcal{G} = (\mathcal{V}, \mathcal{E}; \mathcal{T})$, where $\mathcal{T} = \mathcal{S} \bigcup \mathcal{Q}$,
$\mathcal{S} = \{(\mathbf{x}_i, y_i)\}_{i=1}^{N \times K}$, $\mathcal{Q} = \{\mathbf{x}_i\}_{i=N \times K+1}^{N \times K+T}$

Parameters: $\theta_{emb} \cup \{\theta_v^\ell, \theta_e^\ell\}_{\ell=1}^{L}$

Output: $\{\hat{y}_i\}_{i=N \times K+1}^{N \times K+T}$

Initialize: $\mathbf{v}_i^0 = f_{emb}(\mathbf{x}_i; \theta_{emb}), \mathbf{e}_{ij}^0, \ \forall i, j$

for $\ell = 1, \cdots, L$ **do**

  /* Node feature update */

  for $i = 1, \cdots, |V|$ **do**

    $\mathbf{v}_i^\ell \leftarrow \mathtt{NodeUpdate}(\{\mathbf{v}_i^{\ell-1}\}, \{\mathbf{e}_{ij}^{\ell-1}\}; \theta_v^\ell)$

  end

  /* Edge feature update */

  for $(i, j) = 1, \cdots, |E|$ **do**

    $\mathbf{e}_{ij}^\ell \leftarrow \mathtt{EdgeUpdate}(\{\mathbf{v}_i^\ell\}, \{\mathbf{e}_{ij}^{\ell-1}\}; \theta_e^\ell)$

  end

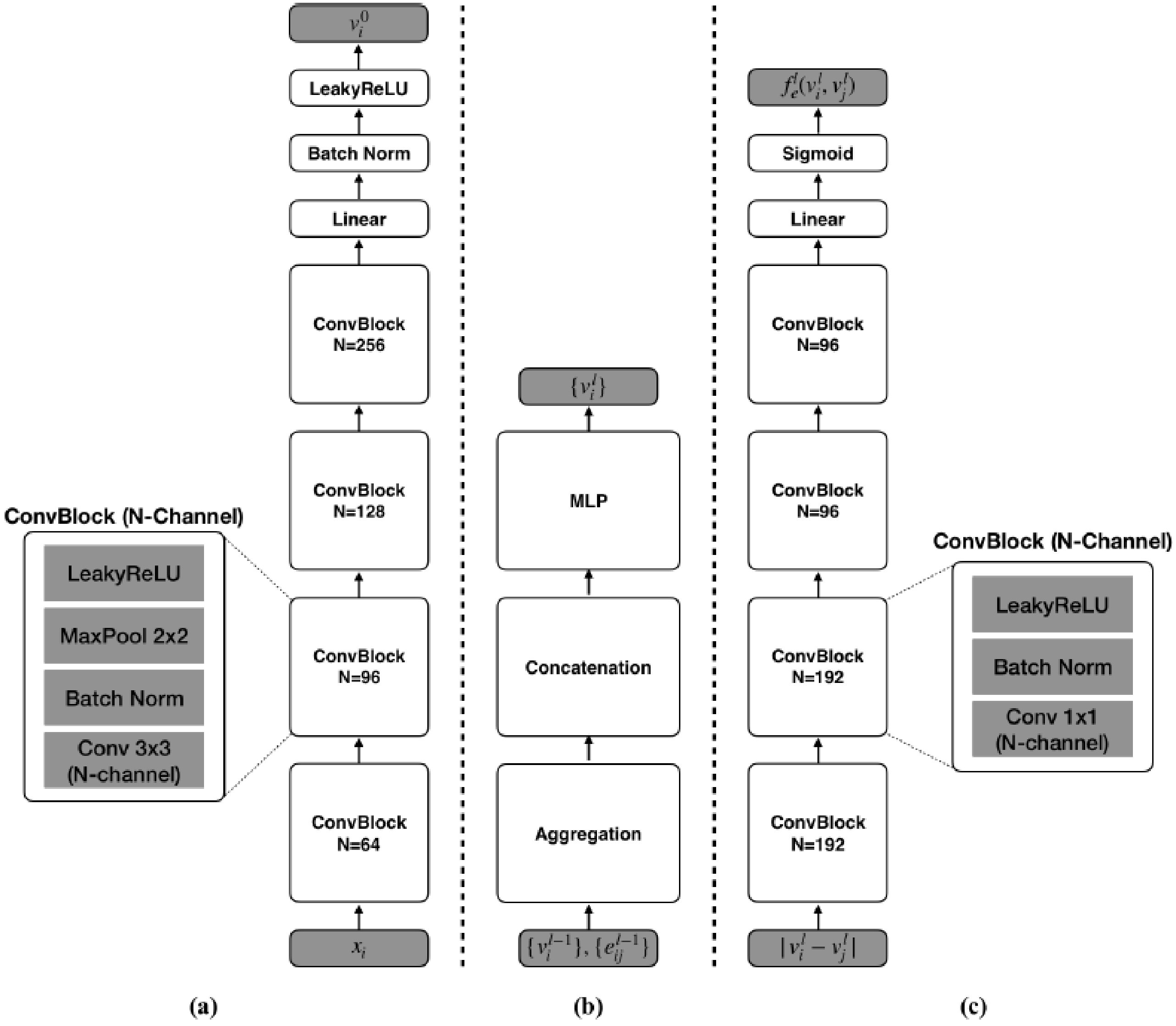
end

/* Query node label prediction */

$\{\hat{y}_i\}_{i=N \times K+1}^{N \times K+T} \leftarrow \mathtt{Edge2NodePred}(\{y_i\}_{i=1}^{N \times K}, \{\mathbf{e}_{ij}^L\})$

Methods

Network Architectures



Detailed network architectures used in EGNN. (a) Embedding network. (b) Feature (node) transformation Network. (c) Metric network

Methods

Training

$$\mathcal{L} = \sum_{\ell=1}^{L} \sum_{m=1}^{M} \lambda_\ell \mathcal{L}_e(Y_{m,e}, \hat{Y}_{m,e}^{\ell}),$$

$Y_{m,e}$ $\hat{Y}_{m,e}^{\ell}$ are the set of all ground-truth query edge-labels and the set of all (real-valued) query-edge predictions

Prediction

each node can be classified by simple weighted voting with support set labels and edge-label prediction results. The prediction probability of node can be formulated as

$$p_i^{(k)} = \text{softmax}\Big(\sum_{\{j : j \neq i \wedge (\mathbf{x}_j, y_j) \in \mathcal{S}\}} \hat{y}_{ij} \delta(y_j = \mathcal{C}_k) \Big)$$

Experiments

miniImageNet

It is the most popular few-shot learning benchmark proposed by derived from the original ILSVRC-12 dataset. All images are RGB colored, and of size 84 × 84 pixels, sampled from 100 different classes with 600 samples per class. We followed the splits 64, 16, and 20 classes for training, validation and testing, respectively.

tieredImageNet

It is also a subset of ILSVRC-12. Different from miniImageNet, tieredImageNet adopts hierarchical category structure where each of 608 classes belongs to one of 34 higher-level categories sampled from the high-level nodes in the ImageNet. Each higher-level category contains 10 to 20 classes, and divided into 20 training (351 classes), 6 validation (97 classes) and 8 test (160 classes) categories. The average number of images in each class is 1281.

Experiments

Table 1: Few-shot classification accuracies on miniImageNet and tieredImageNet. All results are averaged over 600 test episodes. Top results are highlighted.

(a) *mini*ImageNet

Model	Trans.	5-Way 5-Shot
Matching Networks [2]	No	55.30
Reptile [46]	No	62.74
Prototypical Net [3]	No	65.77
GNN [6]	No	66.41
EGNN	No	**66.85**
MAML [4]	BN	63.11
Reptile + BN [46]	BN	65.99
Relation Net [5]	BN	67.07
MAML+Transduction [4]	Yes	66.19
TPN [12]	Yes	69.43
TPN (Higher K) [12]	Yes	69.86
EGNN+Transduction	Yes	**76.37**

(b) *tiered*ImageNet

Model	Trans.	5-Way 5-Shot
Reptile [46]	No	66.47
Prototypical Net [3]	No	69.57
EGNN	No	**70.98**
MAML [4]	BN	70.30
Reptile + BN [46]	BN	71.03
Relation Net [5]	BN	71.31
MAML+Transduction [4]	Yes	70.83
TPN [12]	Yes	72.58
EGNN+Transduction	Yes	**80.15**

“No” means non-transductive method, where each query sample is predicted independently from other queries, “Yes” means transductive method where all queries are simultaneously processed and predicted together, and “BN” means that query batch statistics are used instead of global batch normalization parameters.

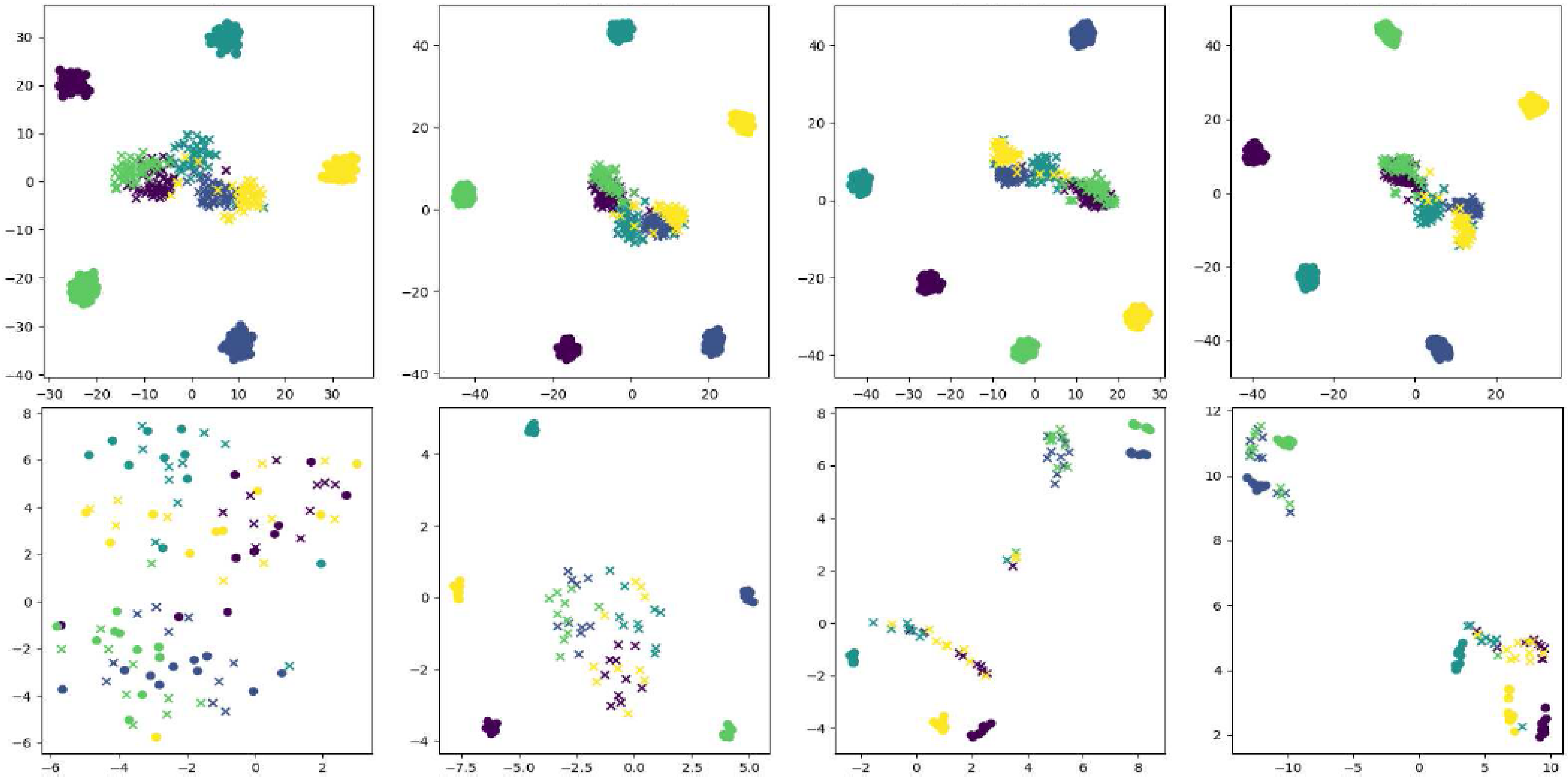
Experiments

Table 1: Semi-supervised few-shot classification accuracies on miniImageNet.

Training method	Labeled Ratio (5-way 5-shot) 20%	40%	60%	100%
GNN-LabeledOnly [6]	50.33	56.91	-	66.41
GNN-Semi [6]	52.45	58.76	-	66.41
EGNN-LabeledOnly	52.86	-	-	66.85
EGNN-Semi	**61.88**	**62.52**	**63.53**	66.85
EGNN-LabeledOnly(T)	59.18	-	-	76.37
EGNN-Semi(T)	**63.62**	**64.32**	**66.37**	76.37

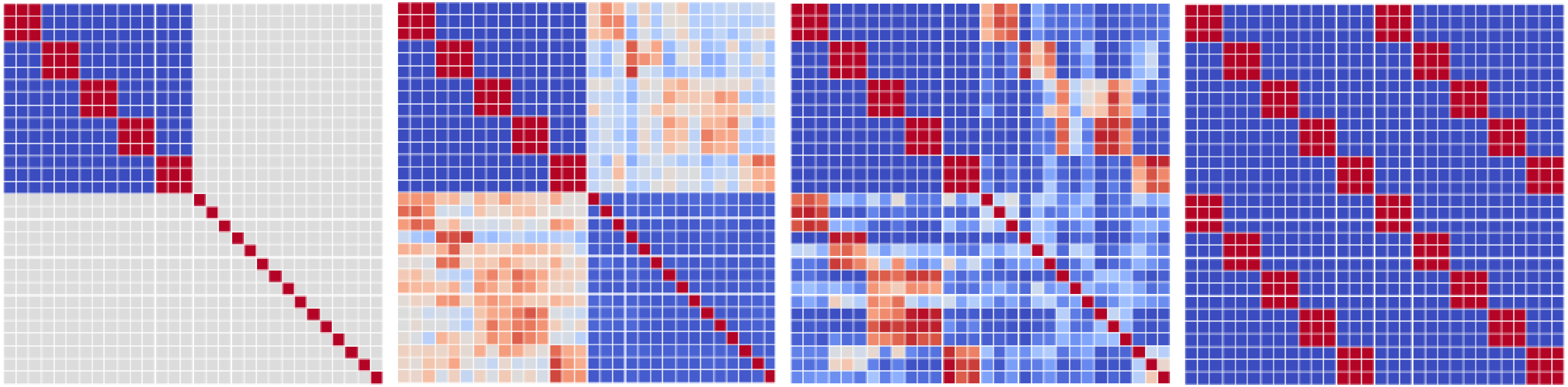
"LabeledOnly" denotes learning with only labeled support samples, and "Semi" means the semi-supervised. 'T' denotes transductive setting.

Ablation Studies



t-SNE visualization of node features. From top to bottom: GNN, EGNN. From left to right: initial embedding, 1st layer, 2nd layer, 3rd layer. ’x’ represents query, ’o’ represents support. Different colors mean different labels.

Ablation Studies



Visualization of edge feature propagation. From left to right: initial edge feature, 1st layer, 2nd layer, ground-truth edge labels. Red color denotes higher value (1), while blue color denotes lower value (e= 0). This illustration shows 5-way 3-shot setting, and 3 queries for each class, total 30 task-samples. The first 15 samples are support set, and latter 15 are query set