Exploiting Knowledge Embedded Soft Labels for Image Recognition

Lixian Yuan[1], Riquan Chen[1], Hefeng Wu[1*], Tianshui Chen[2], Wentao Wan[1], and Pei Chen[1]

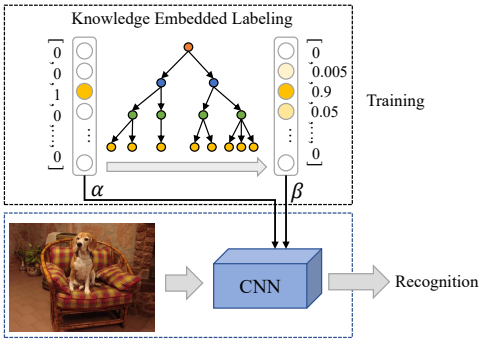
[1]Sun Yat-sen University, Guangzhou, China [2]DarkMatter AI Research, Guangzhou, China

Motivation

- Typical Image Recognition treats each class independently and the training labels are set as one-hot vectors.
- Ignores the correlations between different classes.

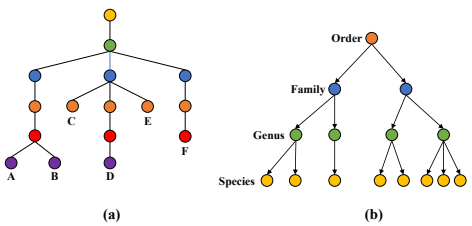
➢ Contributions

- We construct hierarchical knowledge graphs to reserve and explore the semantic relations across different classes.
- The knowledge embedded soft labels is generated based on the hierarchical knowledge graphs.
- The knowledge embedded soft labels serve as extra guidance during the training process.
- The experiment results show the effectiveness of our proposed method without changing the structure of the backbone network.



Proposed Method

➢ Hierarchical Knowledge Graphs



- We construct two possible hierarchical knowledge graphs by, (a) extracting a subnet from WordNet [23] for general image recognition or by, (b) manual categorization [9] based on biology taxonomy for fine-grained image recognition.

➢ Knowledge Embedded Soft Labels

- *The distance between class i and j in a graph*

$$d_{ij} = e_{ij} - 1 \ (\ e_{ij} \geq 2,\ i \neq j)$$

- *The similarity coefficients between class i and j in a graph*

$$c_{ij} = \lambda^{d_{ij}} \ (\ 0 < \lambda < 1,\ c_{ii} = 1\ )$$

- *The proposed knowledge embedded soft labels between class i and j in a graph*

$$g_{ij} = \frac{exp\ (\ T \cdot c_{ij}\ )}{\sum_{k=1}^{n} exp\ (\ T \cdot c_{ik}\ )}$$

➢ Training Loss

$$y'_{ij} = \frac{exp(\ s_{ij}\ )}{\sum_{k=1}^{n} exp(\ s_{ik}\ )}$$

$$\mathcal{L}_{CE} = -\sum_{k} y_{ik} \cdot log\ (\ y'_{ik}\ ) \qquad \mathcal{L}_{KL} = -\sum_{k} g'_{ik} \cdot log \frac{g_{ik}}{g'_{ik}}$$

$$\mathcal{L} = \alpha \cdot \mathcal{L}_{CE} + \beta \cdot \mathcal{L}_{KL}$$

Experiments

➢ Comparison with the state-of-the-art

Method	Accuracy	Network Backbone
Baseline	85.8	ResNet-50
DT-RAM [26]	86.0	ResNet-50
KERL [27]	86.3	VGG-16
MA-CNN [15]	86.5	VGG-19
KERL w/ bbox [27]	86.6	VGG-16
KERL w/ HR [27]	87.0	VGG-16
DPL-CNN [28]	87.1	ResNet-50
Ours	**87.1**	ResNet-50

- Comparison on the Caltech-UCSD Birds (CUB) [8] Dataset for fine-grained image recognition

Dataset	Baseline	Ours
Mini-ImageNet	78.4	**81.0**
Small-ImageNet	80.1	**80.6**

- Comparison on the Mini-ImageNet Dataset [25] and Small-ImageNet Dataset for general image recognition

Methods	Learning rate	Results
Label Smoothing [22]	0.01	85.8
	0.001	86.5
Ours	0.1	**87.1**
	0.01	86.7

- Comparison between our proposed method and Label Smoothing [22]

➢ Ablation Study

λ	T	α	β	Accuracy
-	-	1	0	85.8 (only $\mathcal{L}_{CE}$)
0.5	16	0	1	85.4 (only $\mathcal{L}_{KL}$)
0.5	2	1	1	86.7
	5	1	1	86.2
	16	1	1	85.8
0.5	16	1	1	85.8
	16	0.1	1	86.5
0.5	2	1	1	86.7
	2	1	0.1	**87.1**
	2	1	0.01	85.9

- Ablation study of hyper-parameters

Network Backbones	Baselines (w/o KESL)	Ours (w/ KESL)
MobileNetV2	81.7	82.1
ResNet-18	82.6	83.3
ResNet-50	**85.8**	**87.1**

- Ablation study of different network backbones