Exploiting Knowledge Embedded Soft Labels for Image Recognition

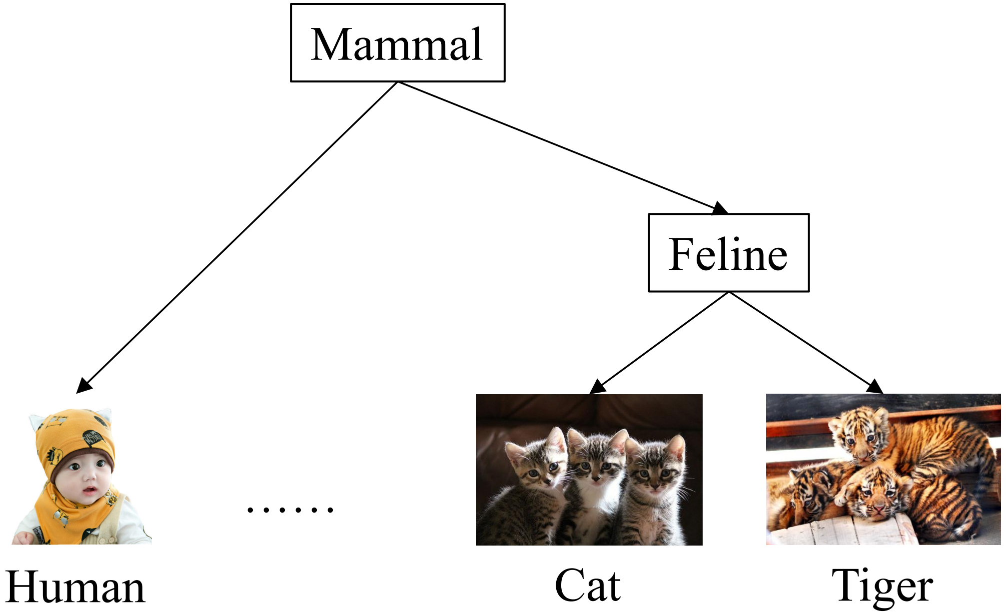
Lixian Yuan[1], Riquan Chen[1], Hefeng Wu[1*], Tianshui Chen[2], Wentao Wang[1] and Pei Chen[1]

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

ICPR 2020
25th INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION
Milan, Italy 10 | 15 January 2021

➢Background

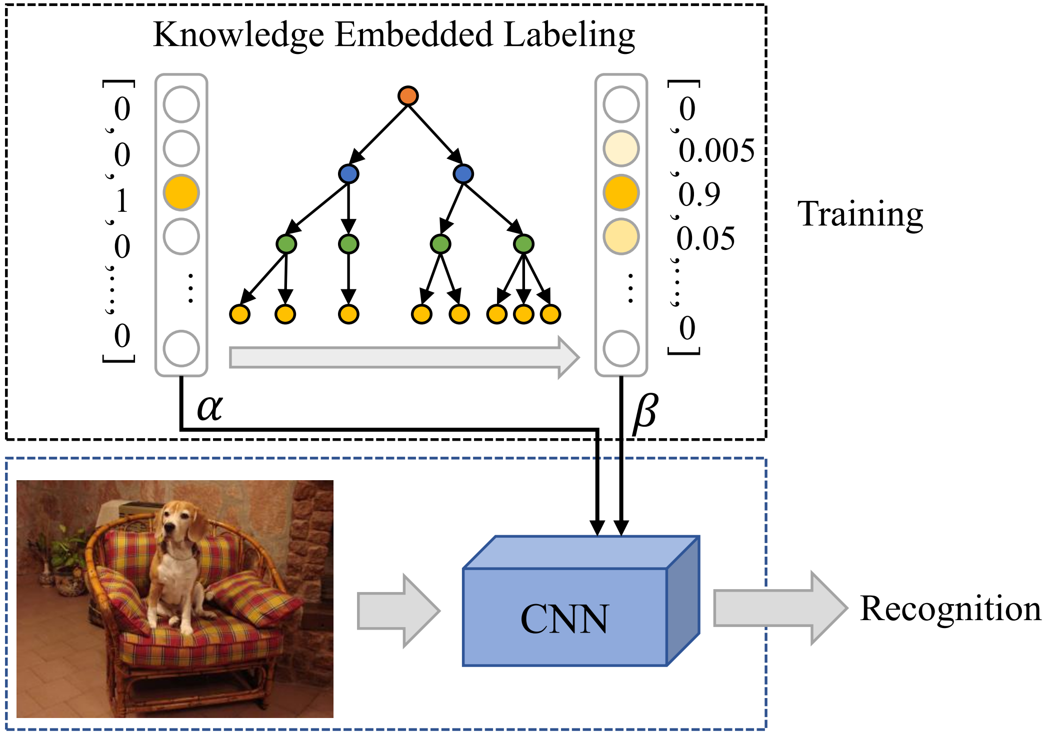
- Objects from correlated classes usually share highly similar appearance while objects from uncorrelated classes are very different.



➢Motivation

- Most existing approaches for image recognition treat each class independently.
- The training labels on the datasets are usually set as one-hot vectors.

  [1,0,0,…,0], [0,1,0,…,0], [0,0,1,…,0], …, [0,0,0,…,1]
- Ignore the correlations between different classes during training.



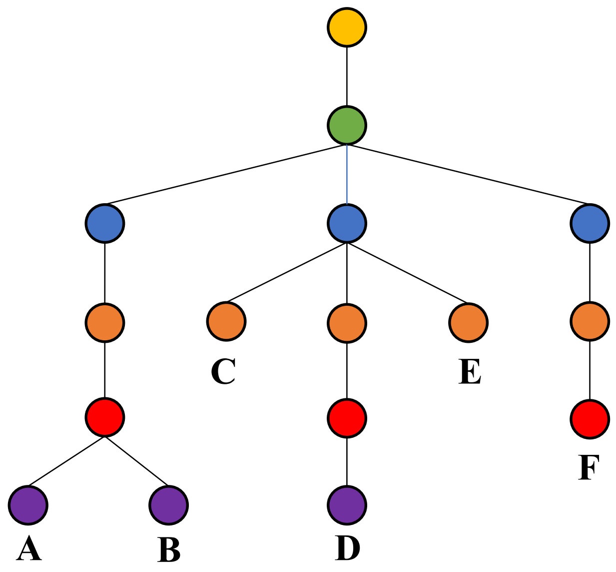
➢Contributions

- We construct hierarchical knowledge graphs to reserve and explore the semantic relations across different classes. With the constructed hierarchical knowledge graphs, we are able to calculate the distance between two classes thus regularize the semantic space across different classes.
- We generate the knowledge embedded soft labels based on the distance across different classes. The knowledge embedded soft labels serve as an extra guidance in training the classifier to better understand the semantic relations across different classes.
- We also devise a Small-ImageNet dataset, which consists of 400 classes objects from the original 1,000 classes in the full ImageNet dataset as an additional evaluation of our proposed method. The experiment results prove the effectiveness of our proposed method.

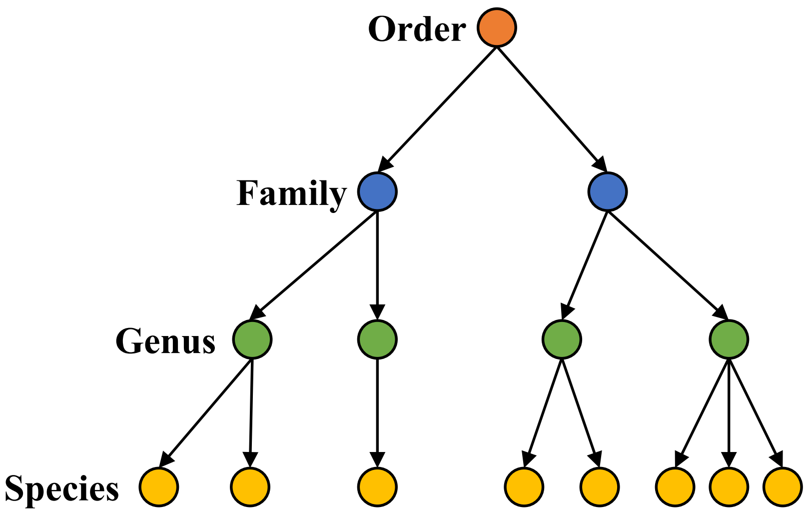
ICPR 2020
25th INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION
Milan, Italy 10 | 15 January 2021

➢Proposed Method

➢Hierarchical Knowledge Graphs



(a)



(b)

(a)Extract a subnet from WordNet (b)Manual Categorization

➢Proposed Method

➢Knowledge Embedded Soft Labels

-The distance between class (node) i and j:

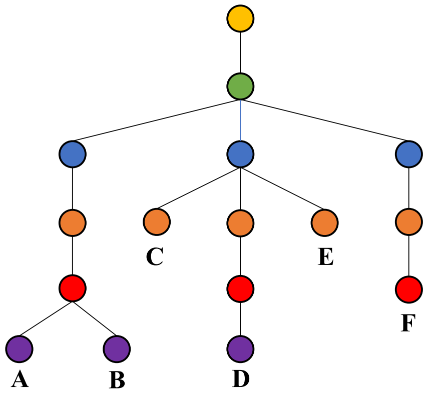
$$d_{ij} = e_{ij} - 1 \; (e_{ij} \geq 2, \; i \neq j)$$

-The similarity coefficients between class i and j:

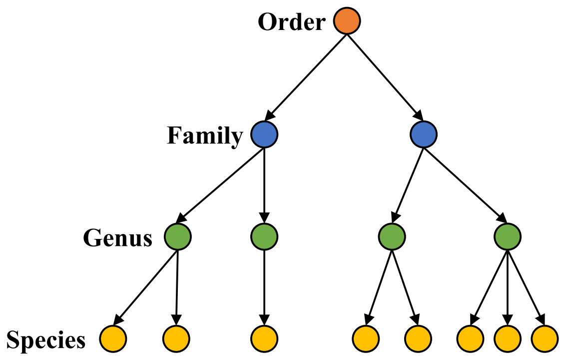
$$c_{ij} = \lambda^{d_{ij}} \; (0 < \lambda < 1, \; c_{ii} = 1)$$

-The proposed knowledge embedded soft labels between class i and j:

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



(a)



(b)

➢Proposed Method

➢Training Loss

-For the original hard one-hot labels, we employ cross-entropy loss as followings,

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

thus we can obtain the L_{CE} :

$$\mathcal{L}_{CE} = -\sum_{k} y_{ik} \cdot log\,(\, y'_{ik}\,)$$

-For the knowledge embedded soft labels, we calculate the Kullback-Leibler loss as:

$$\mathcal{L}_{KL} = -\sum_{k} g'_{ik} \cdot log \frac{g_{ik}}{g'_{ik}}$$

➢Finally, we have $\mathcal{L} = \alpha \cdot \mathcal{L}_{CE} + \beta \cdot \mathcal{L}_{KL}$

ICPR 2020
25th INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION
Milan, Italy 10 | 15 January 2021

Dataset:

Caltech-UCSD Birds (CUB) Dataset for fine-grained image recognition

Mini-ImageNet and **Small-ImageNet** Datasets for general image recognition

➢Experiments

➢Comparison with state-of-the-art in CUB dataset for fine-grained image recognition

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 with baseline (w/o knowledge embedded soft labels) on Mini-ImageNet and Small-ImageNet for general image recognition

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

➢Comparison with Label Smoothing

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

25th INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION

Milan, Italy 10 | 15 January 2021

➢Experiments

➢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

➢ The influences 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**

➢ The influences of different backbone networks

Thank you!

中山大學
SUN YAT-SEN UNIVERSITY

DMAI

ICPR 2020
25th INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION
Milan, Italy 10 | 15 January 2021