Received March 22, 2019, accepted April 12, 2019, date of publication April 24, 2019, date of current version May 14, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2912627

Knowledge Graph-Based Image Classification Refinement

DEHAI ZHANG[1], MENGLONG CUI[1], YUN YANG[1], PO YANG[1,2], (Member, IEEE), CHENG XIE[1], (Member, IEEE), DI LIU[1], BEIBEI YU[1], AND ZHIBO CHEN[1]

[1]School of Software, Yunnan University, Kunming, China
[2]Department of Computer Science, Liverpool John Moores University, Liverpool, U.K.

Corresponding author: Cheng Xie (xiecheng@ynu.edu.cn)

This work was supported in part by the Science Foundation of Yunnan University under Grant 2017YDQN11, in part by the Yunnan Provincial Science Research Project of the Department of Education under Grant 2018JS008, and in part by the National Natural Science Foundation of China under Grant 61876166 and Grant 61663046.

ABSTRACT Biologically inspired ideas are important in image processing. Not only does more than 80% of the information received by humans comes from the visual system, but the human visual system also gives its fast, accurate, and efficient image processing capability. In the current image classification task, convolutional neural networks (CNNs) focus on processing pixels and often ignore the semantic relationships and human brain mechanisms. With the development of image analysis and processing techniques, the information in the image is becoming increasingly complicated. Humans can learn about the characteristics of objects and the relationships that occur between them to classify the images. It is a significant characteristic that sets humans apart from the modern learning-based computer vision algorithms. How to make full use of the semantic relationships in categories and how to apply the knowledge of biological vision to image classification are our main concerns. In this view, we propose the concept of the image knowledge graph (IKG) to incorporate the semantic association and the scene association to fully consider the relations between objects (external and internal). We take full advantage of the reasoning model of the knowledge graph that is closer to the biological visual information-processing model. We conduct extensive experiments on large-scale image datasets (ImageNet), demonstrating the effectiveness of our approach. Furthermore, our method participates in ILSVRC 2017 challenges and obtains the new state-of-the-art results on the ImageNet (82.43%).

INDEX TERMS Biological vision, image classification, knowledge graph, convolutional neural network, semantic relationships.

I. INTRODUCTION

Biological vision mechanisms are important in visual recognition and have received increasing attention in the past few years. Recently, biologically inspired image analysis and processing systems have been built on the basis of further consideration of the learning mechanism of the human brain. In computer vision, image classification is a complex process that may be affected by many factors, e.g., whether the image is sufficiently clear, the complex background and a potential semantic relationship of the category. In human vision, the above problems rarely occur. Instead, humans can take advantage of potential associations to aid in image recognition. It is an advantage of biological vision compared to machine vision. Meanwhile, through the integration of neural biology, the biological perception mechanism, computer science and mathematical science, the related research can bridge biological vision and computer vision. In addition, the correct biological vision result delivers meaningful information that facilitates other related vision tasks, such as scene recognition, object detection [1]–[5], and action recognition.

With the development of deep learning, the correct rate of image recognition has increased yearly. In particular, the success of convolutional neural networks marks the beginning of a new era in image classification. Since 2012, a large deep convolutional neural network created by Alex Krizhevsky *et al.*(2012) on ILSVRC (ImageNet Large Scale Visual Recognition Challenge), AlexNet, won the championship, and by 2017, the top-5 (predict the correct

The associate editor coordinating the review of this manuscript and approving it for publication was Jiachen Yang.

2169-3536 © 2019 IEEE. Translations and content mining are permitted for academic research only.
Personal use is also permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

labels within 5 guesses) error rate was only 2.25% which exceeds human beings.

However, the accuracy of top-1 (predict the correct labels within 1 guess) is still under 80% on ImageNet [6] datasets. As shown in Figure 1, several categories with a low top-1 correct rate are listed. It can be seen that these categories have two main characteristics: the intraclass is diverse, and there is confusion between the categories. This is mainly because most classification approaches are based on per-pixel information without considering the potential knowledge of biological vision and semantic (and logical) associations between objects.

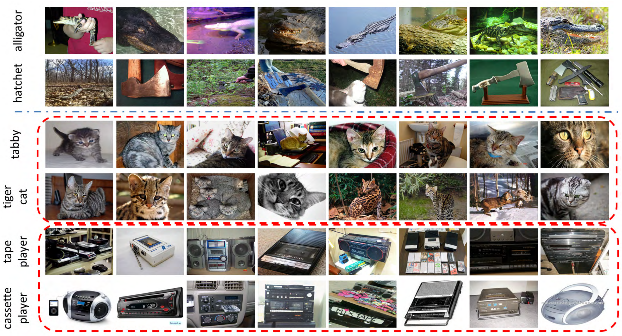


FIGURE 1. Image examples from the ImageNet dataset. Top Row: we show images from two separate object classes (i.e., *Hatchet* and *Alligator*). We notice that large intraclass variations are contained in these images. Bottom Row: there are two pairs of object categories (i.e., *Tabby* and *Tiger cat*, *Tape player* and *Cassette player*). As can be found, images from these pairs of ambiguous categories are highly confused.

Example 1: Alligator and *Hatchet*, the background objects and related objects (*Alligator*: *wood*, *wreck*, *sandbar*; *Hatchet*: *wood*, *bag*, *toolbox*) are contained in these prediction results and are often ignored.

Example 2: Tiger_cat and *Tabby*, images from these pairs of categories are confused. However, we can make a correct judgment based on the results of prediction (*Tiger_cat*: *tiger*, *lion*, *grass*, *wood*; *Tabby*: *keyboard*, *computer*, *toilet*, *rug*). From the examples, it can be found that there is a potential logical relationship between these predicted results (e.g., these objects, wood, wreck, sandbar, often occur around *Alligator*. In addition, *Hatchet* is related to wood and toolbox. Similarly, *Tiger_cat* is mostly in the wild, while *Tabby* is an indoor cat).

Further, with the development of AI and big data (BD) technologies, rich semantic and knowledge data are more useful for image big data analysis. We propose an image knowledge graph to establish the potential association between objects. Therefore, we proceed from the following four steps: (1) image categories extraction; (2) scene (logical) association and semantic association establishment; (3) image knowledge graph construction; (4) image classification refinement. Several experiments show that our method can improve the accuracy of image classification.

There are three main contributions to our work in this paper: (1) Entity extraction; (2) Image knowledge graph construction; (3) Image classification task refinement. We will introduce our work from the following sections. In Section II, some related works are briefly reviewed. Then, Section III presents the main implementation method of our approach. Then, the experimental results are reported, and different aspects of our method are analyzed in Section IV. Finally, concluding remarks are presented in Section V.

II. RELATED WORKS

This part briefly reviews previous works related to our method and clarifies the differences in them. Specifically, related studies are presented from three aspects: (1) image recognition; (2) semantics relationship; (3) knowledge graph.

A. IMAGE RECOGNITION

Since deep networks, such as AlexNet, performed well in ILSVRC 2012,many different deep learning techniques have been used to solve image classification problems. Several outstanding CNN models have been constructed, such as InceptionResNetV2 [7], Xception [8], InceptionV3 [9], ResNet50 [1], VGG19 [10], and VGG16 [10]. These models focus on designing deeper network architectures and contain hundreds of layers. At the same time, several techniques have been designed to reduce the effects of overfitting and data expansion, such as a smaller convolutional kernel size [11] and dropout [12]. For classification problems, boosting and bagging are the most popular sampling-based ensemble approaches. A novel hybrid sampling-based clustering ensemble by combining the strengths fo boosting and bagging is proposed by [13]. A fast image retrieval method which designed for big data has been proposed by [14]. In addition, some optimization techniques have also been proposed to improve image classification performance, such as boosting fuzzy classifiers [15] and relay backpropagation Shen2015Relay [16].

These works focus on the optimization of deep networks in the task of image classification without considering the specifics of knowledge association between different items in the picture. As discussed, complex scenes in the picture and object label ambiguity will increase the difficulty of the classification task. As a complement to previous works on object classification, we conduct a dedicated study on the difficulties of image classification and try to remedy the visual inconsistency problem. Table 1 shows that a comparison of the pros and cons of the most popular frameworks in recent years. In order to improve the training speed, this paper is based on the Tensorflow[2] and Keras[3].

B. SEMANTICS RELATIONSHIP

Semantic relationships are also known as context features. Context plays an important role in human scene understanding. There is a certain correlation between multiple categories in the same image,(e.g., a table and

[2] *https* : //*tensorflow.google.cn*
[3] *https* : //*keras.io*

TABLE 1. The advantages and disadvantages of the existing deep learning classification framework.

Framework	Year	Advantage	Disadvantages
TensorFlow	2015	Multi-GPU parallel computing Automatic derivation Flexible portability Parallel design	Code complexity Frequent interface changes Bad input channels
Caffe	2014	Robust and fast Component modularization	Not flexible enough Insufficient scalability
Keras	2017	Rich variety of models	Slow training speed
Theano	2008	Tight integration with Numpy Transparent use of a GPU Speed and stability optimizations	Low level framework Patchy support for pretrained models
Torch	2002	Support convolutional network	Not flexible enough
PyTorch	2017	Dynamic neural network Flexible expansion	Server architecture lacks flexibility

chair are probably present in the same images, whereas an elephant and a bed are not). It is similar to the *Probability* which is one of five different classes of relations between an object and its surroundings [17]. Early studies in psychology and cognition show that semantic context aids in visual recognition of human perception. Palmer [18] examined the influence of prior presentation of visual scenes on the identification of briefly presented drawings of real-world objects. For weak semantics of data itself, a spatiotemporal semantic scenario meta-model-based configurable platform is proposed by [19].

Recently, statistical methods have been used in object categorization. The study by Wolf and Bileschi [20] used semantic context obtained from a "semantic layer" available in training images. CoLA (for cooccurrence, location and appearance) [21] has been proposed. The model uses a conditional random field(CRF) to maximize object label agreement according to both semantic and spatial relevance and relative location between objects using simple pairwise features. Recently, contextual interaction has been proposed as a deep evaluator for image retargeting quality by [22].

In this paper, a wider range of areas was used, not confined to the field of expert knowledge. For example, the training corpus of works such as res [23], wup [24], lin [25] comes from different fields, so that the similarity between categories can be calculated from different semantics. This makes the image classification application more general.

C. KNOWLEDGE GRAPH

Recent years have witnessed rapid growth in knowledge graph (KG) construction and application. Numerous KGs, have been created and successfully applied to many real-world applications, from semantic parsing [26], [27] to information extraction [28], [29], such as Freebase [30], DBpedia [31], YAGO [32], and NELL [33]. There are some specific applications of knowledge graphs, such as KGs in hospital data [34], industrial manufacturing forms [35], social network data [36], and user behavior data [37]. There are also music cognition [38]. They designed a music cognition system based on knowledge graph. A KG is a multirelational graph composed of entities (nodes) and relations (different types of edges).

Using the idea of the knowledge graph in the image classification task, the correlation between categories can be regarded as a relationship in the knowledge graph. Wang *et al.* [39] proposed a method for using the knowledge correlation between categories to merge similar classes. A data clustering method have been proposed to deal with similar category fusion problems by [40]. In the semantics context, a knowledge graph can suggest pairwise spatial relations between objects that typically cooccur in the same scene. Additionally, knowledge graphs are defined on undirected graphs, and they are global probability distributions. KGs are better suited to handle interactions over image partitions.

In this paper, in combination with the knowledge graph method, we deeply studied the relationship between categories in the image, and combined the semantic calculation method (external knowledge) to guide the image classification task.

III. MOTIVATION

After the above discussion, it can be seen that the potential semantic relationship between objects has a considerable impact on the classification accuracy. Next, we will detail our method in the following four steps:

- **a.** The first step is association category extraction. We will extract association categories from the predictions of CNNs. Each category is a node in an image knowledge graph(IKG).
- **b.** After that, the relationships among the categories are divided into two kinds of associations according to different situations. One is semantics association, and another is scene association. The two relationships are merged into the IKG.
- **c.** Then, we use the confusion matrix from the CNN prediction results, combined with the information of the nodes and edges to construct an image knowledge graph(IKG).
- **d.** Finally, to verify the effect of the method, we use the image knowledge graph(IKG) as extra information to improve image classification.

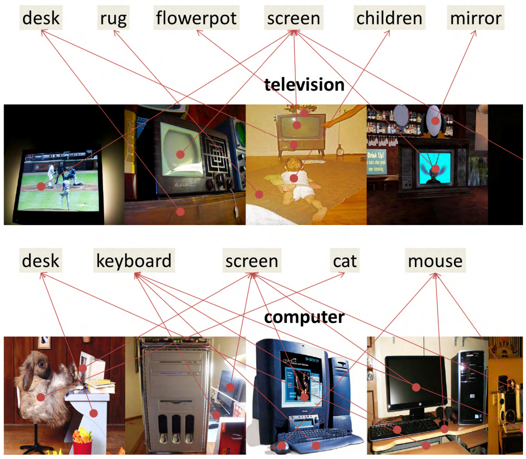


FIGURE 2. The surrounding objects in images.

A. IMAGE CATEGORIES EXTRACTION

In the image classification task, there is one subject category for each image, but its surrounding objects are ignored. It is worth noting that in the consciousness of the human brain, we can often obtain rich background information from surrounding objects. We use background information to identify the categories in the image more accurately.

- As shown in Figure2 showing computers and television, the background information of the computer includes keyboard, mouse, computer desk, and web-page, and the TVs include remote control, sofa, living room, and children. By comparing such background information, we can easily distinguish between images of computers and televisions.
- The images of the tiger and cat are also easily distinguished according to their background information. The former mostly includes a wild environment (e.g., tree, river, weeds, snake), while the latter mostly includes an indoor environment (e.g., table, computer, pillow, yarn ball).

It can be seen that the background information in the image classification task can provide deeper knowledge and should not be ignored. To take into account the background categories in image recognition, we associate the objects in the image, and the next item shows how to associate objects.

B. ESTABLISHING SCENE ASSOCIATION AND SEMANTICS ASSOCIATION

When constructing the relevance of objects, the associations are divided into two kinds according to different situations: one is semantics association, and the other is scene association. The former is also called semantics similarity. In the WordNet semantic web, the similarity between the tiger and the cat is high because both belong to the cat family in the WordNet hierarchy, but the cat has little relevance to the table. In the scene association, as shown in Figure 3 the similarity between categories is closely related to the environment in

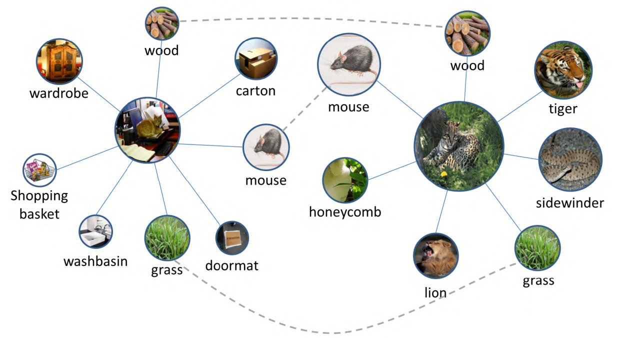


FIGURE 3. The scene association of objects in images.

which they are located, such as cat and table, dog and frisbee, and birds and wires.

As shown in Figure 4, using the association between objects, we build the image knowledge graph (IKG). The graph contains semantic associations and scene associations. Under the guidance of the knowledge graph, the recognition model uses background information as a factor when identifying a picture. This has a strong auxiliary correction effect on the image recognition task, which makes the machine recognition more accurate.

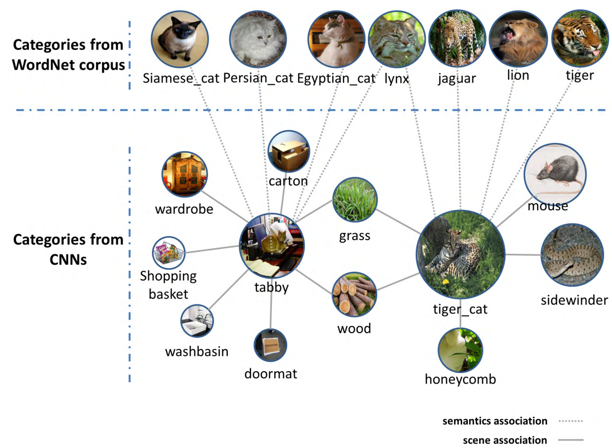


FIGURE 4. The combination of two types of associations.

C. IMAGE KNOWLEDGE GRAPH CONSTRUCTION

We constructed the image knowledge graph(IKG) using the category information(semantics association and scene association) in the image. In this graph, each category is related to the category of its own scene and is also related to its own similar category. There are four steps to building the graph. **First**, we obtain the prediction labels from the image and divide them into truth labels and surrounding labels. **Second**, we construct the "cell graph" with the truth label as the center node and the surrounding label as the side node. The association value between the "cell graph" is obtained from the adjacency matrix. **Third**, every related "cell graph" is combined into the subgraph. Each subgraph is a main category. **Fourth**, we combine all the subgraphs to construct a global graph. The graph contains knowledge information about the object-object association. The construction process is shown in Figure 5.

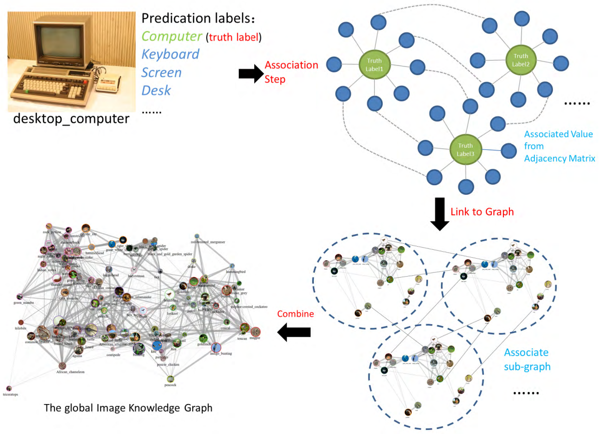


FIGURE 5. The process of constructing Image Knowledge Graph(IKG).

D. IMAGE CLASSIFICATION REFINEMENT

To increase the persuasiveness of the experiment, we use two methods to refine the image classification. One method uses external knowledge(WordNet semantic web) directly, and the other method uses IKG. In the refinement process, each prediction label has a weight, which is taken from IKG and word corpus. Then we make a full connection calculation for all labels of top-10. Finally, the label with the highest average correlation with each prediction label is the top-1. As shown in Figure 6, the prediction labels of a single image are obtained from CNN models. The confidence value of each label is the weight of itself and then is combined with the association weights obtained from the IKG or word corpus. The calculation details are described in Section III.

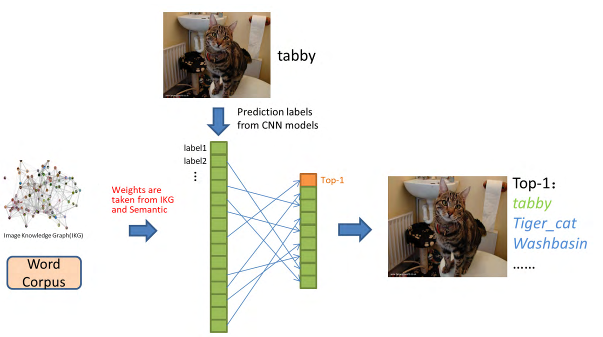


FIGURE 6. The process of refinement. After the weights are calculated for multiple labels of each image, the label with the highest weight will be top-1.

IV. THE METHOD

In this section, the processing of our method is discussed in detail. As shown in Figure 7, it is the framework of our model. Subsection A compares the results of several popular CNN image classifications and extracts the prediction labels in the image. After conducting the study, it shows that these prediction labels have a certain similarity. They are very similar to the ground truth, even some of which appear in the background. According to this idea, the training set is used to construct the IKG in subsection B. In the IKG, prediction labels are nodes, and similarities are edges. The similar categories are close to each other in the graph. The image knowledge graph(IKG) records similar labels for each image. In subsection C, our method aims to refine the top-1 results and reintegrate the confidence of labels with a word corpus.

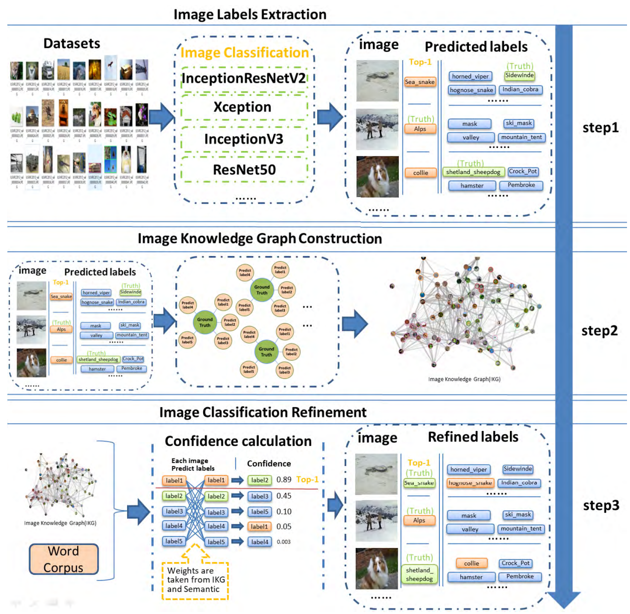


FIGURE 7. The framework of our model. In the model, we obtain the prediction labels from each image, which are built by CNN networks. Then, we construct the image knowledge graph. Finally, we improve the accuracy of the classification according to the knowledge of the IKG and the word corpus.

A. IMAGE CATEGORIES EXTRACTION

Before building an image knowledge graph(IKG), we first need to obtain the predicted categories in the image. As shown in Figure 8, for each image, the CNN models provide multiple predictions and sort them according to the confidence level. Therefore, the quality of CNN models has a great impact on graph construction, so first, we analyze several CNN models that are currently performing best.

The current popular CNNs are: Xception [8], InceptionV3 [9], InceptionResNetV2 [7], ResNet50 [1], VGG [10]. These models perform well in the image classification task with weights pretrained on ImageNet.

1) CNN MODELS

InceptionV3 [9] is trained with stochastic gradient utilizing the TensorFlow distributed machine learning system using

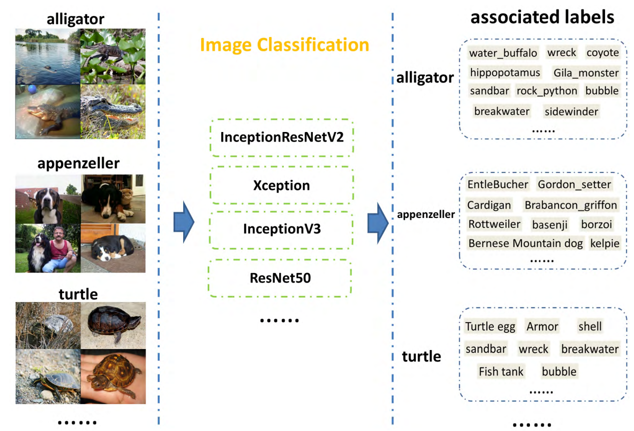


FIGURE 8. **Relevant categories extraction. The CNN models predict multiple images in the main category.**

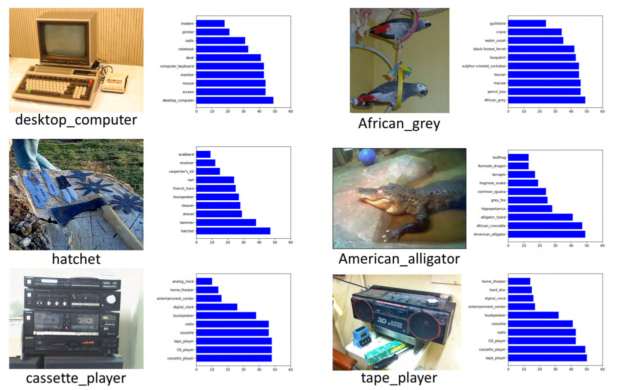


FIGURE 9. **The relevant labels. Several image examples with ground truth from the ImageNet dataset.**

50 replicas running each on a NVidia Kepler GPU with batch size 32 for 100 epochs. Furthermore, gradient clipping with threshold 2.0 was found useful for stabilizing the training. Compared to GoogLeNet and InceptionV2, InceptionV3 shows small gains in classification performance on the ImageNet dataset.

The Xception [8] architecture has 36 convolutional layers forming the feature extraction base of the network. In our experimental evaluation, because of the image classification task, the convolutional base is followed by a logistic regression layer. Compared to Inception V3, Xception shows small gains in classification performance on the ImageNet dataset.

The preprocessing of images in VGG [10] subtracts the mean RGB value, computed on the training set, from each pixel. In addition, it contains 1000 channels(one for each class) in the third fully connected layer which performs 1000 way object classification. Compared to GoogLeNet, VGG shows small gains in classification performance on the ImageNet dataset.

A *deep residual learning* framework is introduced to address the degradation problem in ResNet50 [1]. It is worth noticing that the ResNet50 has fewer filters and lower complexity than the VGG nets [10].

InceptionResNetV2 [7] is trained with a stochastic gradient utilizing the TensorFlow [41] distributed machine learning system using 20 replicas running each on a NVidia Kepler GPU. Model evaluations are performed using a running average of the parameters computed over time. Compared to the above models, InceptionResNetV2 [7] shows large gains in classification performance on the ImageNet dataset.

To ensure data consistency, we use the same training set to train on different models. the results of each model are consistently integrated into the main class set.

2) EXTRACTION RESULT

After processing by CNN models, each image obtains multiple predictions. To make the experiment representative, we combine the top-10 results of multiple models of a single image as the relevant categories.

As shown in Figure 9, each category of images is processed by multiple CNN models, and the results of top-10 are all integrated into the collection of the main category.

B. ESTABLISH SCENE ASSOCIATION AND SEMANTIC ASSOCIATION

1) SEMANTIC ASSOCIATION

In the WordNet semantic web, each category label is a corresponding word tree. The semantic trees can be used to establish semantic associations for categories in an image. Formally, we define the similarity between two words s_1 and s_2 as follows:

$$sim(s_1, s_2) = \frac{2 * \log P(s))}{\log P(s_1) + \log P(s_2)} \tag{1}$$

where s is the most specific synset in WordNet that subsumes s_1 and s_2. Where $P(x)$ is the probability of each synset in WordNet.

We define the semantic association between two categories c_1 and c_2 as follows:

$$associationA(c_1, c_2) = k \cdot sim(c_1, c_2) \tag{2}$$

where k is the threshold for the experiment adjustment. The value of association can be used as the weight in the IKG.

2) SCENE ASSOCIATION

Semantic association is external knowledge, and the scene association is the intraimage knowledge information. As shown in Figure 10, similar to the confusion matrix, the association matrix is computed based on the top-10 predicted categories.

This association matrix displays crossing errors between pairs of categories, which implicitly indicates the degree of similarity between them. Formally, the scene association is defined as follows:

$$associationB(c_i, c_j) = \frac{1}{2}(C_{ij} + C_{ij}^{\top}) \tag{3}$$

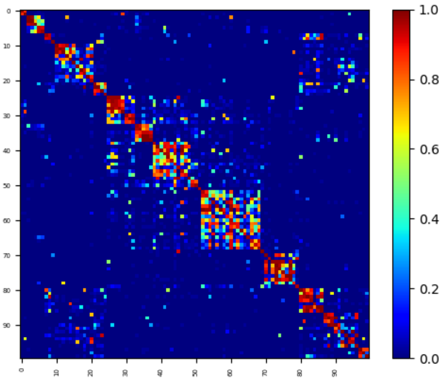


FIGURE 10. **The association matrix of predicted categories. The knowledge of association matrix displays crossing similarity between pairs of categories. Red represents high value of association.**

where $C \in \mathbb{R}^{N*N}$ is the association matrix, C_{ij} represents the probability of classifying the i^{th} class as the j^{th}, the larger this value indicates a higher association degree between two categories. N is the number of object categories. The equation ensures the association measure is a symmetric metric.

C. IMAGE KNOWLEDGE GRAPH CONSTRUCTION

In this section, the image knowledge graph is constructed by using the two types of associations. We build the IKG to associate as many objects as possible, and more object associations indicate the objectivity of the existence of objects in the images (reality). For this purpose, we consider both semantic and scene associations to establish object-object associations. The equation of object-object association is defined as follows:

$$\begin{aligned} &associationC(c_i, c_j) \\ &\quad = \alpha associationA(c_i, c_j) \\ &\qquad + \beta associationB(c_i, c_j) \end{aligned} \tag{4}$$

where α and β are the parameters balancing these two terms. c_i and c_j are the i^{th} and j^{th} category. A clustering algorithm that progressively merges relevant categories is shown in Algorithm 1.

In Algorithm 1, this iteration repeats until all relevant categories are associated with the graph. At the end, all predicted classes are merged into the image knowledge graph.

Algorithm 1 Construct Image Knowledge Graph

Data: Ground truth T, prediction labels P, association matrix M.
Result: Graph: G. -initialization: $G = \emptyset$.
while $t \in T$ **do**
  while $p \in P$ **do**
    Calculate $M(t, p)$
    if $M(t, p)>0$ **then**
      Add relevant t, p nodes in Graph G.
      Add line between t, p in Graph G.
    end
  end
end
return G.

D. IMAGE CLASSIFICATION REFINEMENT

1) PROCESS OF REFINEMENT

Two refinement experiments are designed to verify the validity of the method. First, refine the classification with the word corpus for making the comparison. At the same time, its similarity value between categories can provide semantic weights for the image knowledge graph(IKG) method. In the IKG method, the similarity in the graph and the semantic weights are used to refine the image classification.

2) REFINEMENT WITH WORD CORPUS

The methods to calculate the semantic similarity of the corpus include res [23], wup [24], lin [25], which can be exploited to calculate the similarity of the prediction labels. Then, each prediction label will determine a soft confidence(compared with the original confidence) as the result. The formula for calculating the confidence is as follows:

$$C = \frac{1}{N} \cdot \sum_{i=1}^{N} \sum_{j=i+1}^{N} sim(n_i, n_j) \tag{5}$$

In Equation 5, InceptionResNetV2 [7] produces 5 predicted class labels $n_i, i = 1, \cdots 5$, the semantic similarity between two words is S in the word corpus(WordNet [42]), and the average semantic similarity of each category is regarded as the soft confidence which is C.

The original confidence is used as the evaluation criterion. However, when the top two or three confidence levels are equal or lower, the classification accuracy declines rapidly. It can be seen that the confidence has a great influence on the classification.

In the top-5 results of the current picture, the smaller confidence of the first predicted category and the smaller difference between the confidence of the first two predicted categories, the greater the confusion value of the classification on this image. Then, the value is defined as follows:

$$v = \frac{1}{d_i(d_i - d_j) \cdot \alpha} \tag{6}$$

where d_i is the i^{th} result of the top-5. α is a parameter that is used to tune the result. In our experiment, α is assigned 100.

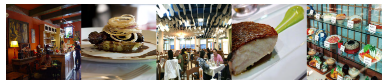
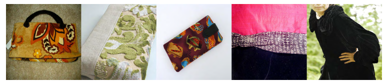
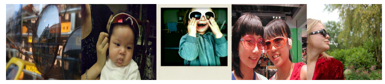
Next, taking into account the impact of the top-5 original category score, the v is used as the weights. Then, the soft confidence is C, and the expression is defined as follows:

$$C = \frac{1}{N} \cdot \sum_{i=1}^{N} \sum_{j=i+1}^{N} sim(n_i, n_j) * v \tag{7}$$

Algorithm 2 Calculate Category Similarity

Data: Graph G,prediction label P.
Result: Similarity S.
-initialization:
$N = G.node$,$L = node.link$,$Sum = 0$,$Count = 0$.
while $p \in P$ **do**
  while $n \in N$ **do**
    if $n{=}{=}p$ **then**
      while $l_i \in L$ **do**
        Calculate $T = G(n, l_i)$
        $Sum \leftarrow Sum + T$
        $Count \leftarrow Count + 1$
      end
    end
  end
end
$S \leftarrow Sum/Count$
return S.

TABLE 2. The diversity of categories in the data. We show example object categories along the range of that property.

categories	images	number
restaurant		50
velvet		50
sunglasses		50
hair slide		50
hook		50

3) REFINEMENT WITH IMAGE KNOWLEDGE GRAPH

The adjacency matrix is used as a context graph for the categories in the image. By calculating the relevance between the nodes in this graph, the confidence of the prediction categories in the image is determined, as shown in Algorithm 2. Specifically, the method is similar to the computational semantic similarity and employs this adjacency matrix for calculating the pairwise similarities of the object classes. Formally, the similarity is defined as follows: $S = \frac{1}{N} \cdot \sum_{i=1}^{N} \sum_{j=i+1}^{N} G(c_i, c_j)$. $G(c_i, c_j)$ is the value of relevance, where c_i and c_j are two different classes, the i^{th} and j^{th} class. N is the number of object classes. The impact of weights v is still taken into account. The formula is shown as follows:

$$C = \frac{1}{N} \cdot \sum_{i=1}^{N} \sum_{j=i+1}^{N} G(c_i, c_j) * v \tag{8}$$

It is worth noting that, different from other approaches, in our final expression, there is no "experience value" or "confidence value". Similarities among individuals are all calculated from the individuals' linkages. As we see in Section V, this framework further improves the performance of the CNN(baseline).

V. EXPERIMENT

In this section, the experimental setting is described, and the performance of our proposed method on the ImageNet [6] scene benchmark is reported. First, some detail of the datasets and evaluation are described in the data setup and evaluation metrics subsection. Second, the image label extraction experiment results are exhibited in experiment 1. Then, experiment 2 explores the effect of semantic information(WordNet [42]) by performing extensive experiments. After that, experiment 3 is conducted to explore the effect of the image knowledge graph on the ImageNet datasets. Finally, we analyze the experimental results and present several failure examples by our methods, and discuss possible reasons.

A. DATA SETUP

We first evaluate our method on ImageNet [6] classification datasets. The results are reported on their validation sets since the ground-truth labels of their test sets are not available. The ILSVRC 2012 dataset for the image classification task consists of photographs collected from Flickr and other search engines, manually labeled with the presence of one of 1000 object categories. Each image contains one ground truth label. The training data contains approximately $1,300,000$ images from these object categories. There are 50,000 images for the validation dataset and 100,000 images for testing. Table 2 shows the category distribution and averages one category corresponding to 50 images.

The dataset contains many fine-grained classes. For example, instead of the "dog"İ category, there are 120 different breeds of dogs. A total of 80 synsets are randomly sampled at every tree depth of the mammal and vehicle subtrees.

B. TRAINING DETAILS

All models were trained with stochastic gradient descent. We used a learning rate of 0.045, decayed every two epochs using an exponential rate of 0.94. Our models and output networks were single layer networks with sigmoid activations. All networks were trained for 20 epochs with a batch size of 16. To speed the training process, we use the multi-GPU extension of Caffe toolbox for our CNN training. For testing our models, we use the common 6 crops (5 corners and 1 center) and their horizontal flipping for each image at a single scale, thus having 10 crops in total for each image. The final score is obtained by taking average over the predictions of 10 crops.

TABLE 3. The first column shows the ground truth labeling on an example image, and the next three show three sample outputs with the corresponding evaluation scores.

Steel drum 	steel drum drum drumstick	Scale T-shirt steel drum drum drumstick	Scale drum drumstick T-shirt Folding chair
Ground truth	top-1 Acc:1 top-5 Acc:1	top-1 Acc:0 top-5 Acc:1	top-1 Acc:0 top-5 Acc:0

C. EVALUATION METRICS

1) TOP-1 EVALUATION

Top-1 accuracy can be evaluated by the algorithms. In this case, algorithms are penalized if their highest-confidence output label c_{i1} did not match ground truth class Ci.

Specifically, each image i has a single class label Ci. An algorithm is allowed to return j labels $c_{i1}, \cdots c_{ij}$ and is considered correct if $c_{i1} = C_i$ for some j. The accuracy $Top1_{acc}$ of a prediction $d_{ij} = d(c_{ij}, C_i)$ is 1 if $c_{i1} = C_i$ and 0 otherwise. In other words, we evaluate the method if the correct label is predicted within 1 guess.

$$Top1_{acc} = \frac{1}{N}\sum_{i=1}^{N}\min_{j} d_{ij} \tag{9}$$

Table 3 shows an example of an evaluation on an image. When the true output score is highest, the top-1 and top-5 accuracy is 1. When the true outputs score is in the top five and not in the first place, top-1 is 0, top-5 is 1. In addition, the true output scores are not in the top five, and both top-1 and top-5 are 0.

2) EXPERIMENTS OVERVIEW

Based on the above analysis, the next three experiments were conducted. First, image classification models were the baseline to explore the feature of images. After that, the second experiment used the word corpus as the semantic information between categories, and the highest confidence was the result of refinement. In the third experiment, the image knowledge graph was constructed by an adjacency matrix, and the relevance of the labels in the graph were used as the knowledge information between categories. As shown in Table 4, the method was employed in each experiment.

D. EXPERIMENT 1: IMAGE LABELS EXTRACTION

The category labels were obtained from image classification models, namely, InceptionResNetV2 [7], Xception [8], InceptionV3 [9], ResNet50 [1], VGG19 [10], VGG16 [10]. Table 5 shows the comparison of normal CNNs. It can be seen that the image classification models performed well. The top-1 accuracy was 70% or more. InceptionResNetV2 [7] preformed best and its top-1 accuracy was 80.4%.

TABLE 4. The name and abbreviation of each experiment. Three experiments were conducted to improve top-1 accuracy with CNN as the baseline.

Experiment	name	method
Experiment 1	Image Labels Extraction	CNNs
Experiment 2	Refine by Word Corpus	CNNs & WordNet
Experiment 3	Refine by Image Knowledge Graph(IKG)	CNNs & IKG

TABLE 5. The top-1 accuracy refers to the model's performance on the ImageNet validation dataset.

Model	Top-1 Accuracy
Xception	79.0%
InceptionV3	78.8%
ResNet50	75.9%
VGG19	72.7%
VGG16	71.5%
InceptionResNetV2	**80.4%**

TABLE 6. The results of refinement by the word corpus.

Method	Top-1 Accuracy	Compared
InceptionResNetV2	80.40%	–
InceptionResNetV2+res	75.98%	-4.42%
InceptionResNetV2+lin	76.97%	-3.43%
InceptionResNetV2+wup	79.44%	-0.96%
InceptionResNetV2+path	**80.71%**	**+0.31%**

The InceptionResNetV2 [7] method was chosen as the base structure for the next experiment. In addition, the prediction results of InceptionResNetV2 [7] were the baseline to construct the image's label.

E. EXPERIMENT 2: REFINE BY THE WORD CORPUS

In this subsection, the semantics of the categories are extracted from the word corpus of WordNet [42] according to the characteristics of the ImageNet dataset storage structure. The results of the top-1 in Table 6 show that the res [23], wup [24], lin [25]

and path[1] methods have considerable effects, but the effect is not optimistic. This happens because the word corpus is based on IC(information content), the degree of association with the category word of this dataset is small and resulting in the final result is not optimistic. In the next experiment, the relevant categories in the image knowledge graph are considered.

F. EXPERIMENT 3: REFINE BY THE IMAGE KNOWLEDGE GRAPH

As the results show in Table 7, the IKG refines the accuracy of different CNN models. Among them, IKG combined with InceptionResNetV2 has the highest result(82.43). The top-1 accuracy can be increased by approximately 2%, compared to the original 80.4%. The IKG correction algorithm

[1] *http://www.nltk.org*

TABLE 7. The results of refinement by the image knowledge graph.

Method	Top-1 Accuracy	Compared
Xception+IKG	80.43%	+1.43%
InceptionV3+IKG	80.31%	+1.51%
ResNet50+IKG	77.5%	+1.60%
VGG19+IKG	74.57%	+1.87%
VGG16+IKG	73.27%	+1.77%
InceptionResNetV2+IKG	**82.43%**	**+2.03%**

TABLE 8. The Score refers to the score of the first class description in the top-5.The Truth&Fault means truth or fault in Top-1.

θ	Number of Truth	Number of Fault	Difference
0.9	3205	1356	1849
0.8	1477	1113	364
0.7	985	930	55
0.6	655	740	-85
0.5	379	507	-128
0.4	162	269	-107

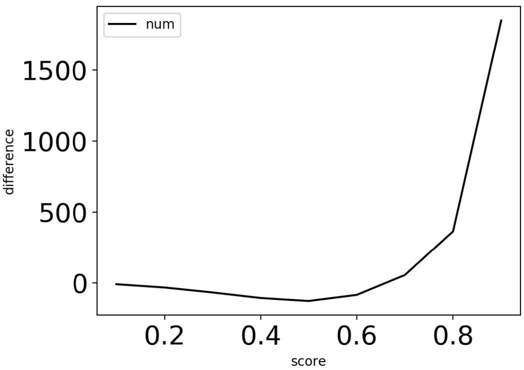


FIGURE 11. The linear relationship between the Score and the number of the Difference.

can improve the classification model accuracy. This happens because the IKG correction algorithm does not depend on the framework of the CNN models but uses predicate labels to correct the top-1 results. It also shows that the IKG algorithm has better generalization ability.

The model method we mentioned in the last few chapters uses the adjacency matrix of its associated graph to obtain the degree of association between the five categories, and we use this as a standard to correct the correction results.

G. THRESHOLD SETTING

The original confidence impact of the class label causes the correct class label to be fixed incorrectly. To eliminate this factor, we remove those categories with high confidence before obtaining the rank of refinement top-1 results. This setting is called "filt".

Our experiment obtains a picture, in which its confidence is as far as possible in the top-5, but not in the top-1 so that we can refine according to the semantic relationship of the 5 class labels description. We set θ as the parameter for selecting the fitting image. Only the image exceeding the θ parameter was selected.

From this result shown in Table 8, we see that when the scores of the five class labels are between 0.5 and 0.4, we can directly extract more data. In other words, the larger the negative value, the better the selection. To find a proper value for θ, the range should be set from 0.3 to 0.9.

Figure 11 shows that the number of the Difference changes with the change in the threshold(θ), and the lowest point is approximately 0.5. Therefore, in the remaining experiment, we fix the parameter θ as 0.5.

H. DISCUSSION

1) GENERALIZATION ANALYSIS

Extensive experimental results have demonstrated the effect of our proposed method on ImageNet single object classification datasets. The ImageNet datasets contain 1000 single categories and have a total of 50,000 images in the validation datasets, with at least 50 images per category. Following the original evaluation protocol, we use 40 images from each category for training, and another 10 images for testing. Figure 12 shows the difference in the average precision for each category in ImageNet between our IKG model and the detection baseline for the refinement experiment.

The semantic method and image knowledge graph are used to improve the top-1 recognition rate and achieve better results. Using our method, the top-1 accuracy can increase by approximately 2%, compared to the original 80.4%.

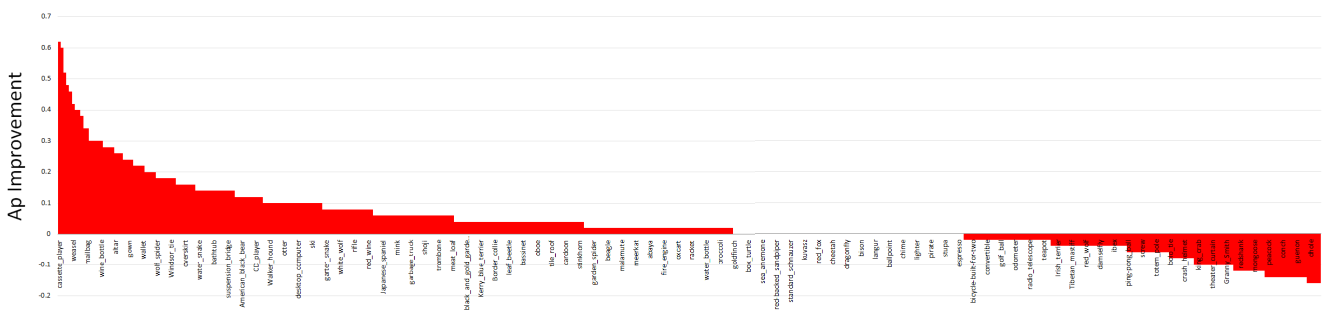


FIGURE 12. Difference in the average precision for each of the 1000 labels in ImageNet between our IKG model and the detection baseline for the refinement experiment. Top categories: Cassette player, tiger cat, weasel, mailbag, wallet. Bottom categories: Conch, peacock, flagpole, hook, mitten.

TABLE 9. Examples of images that our method successfully predicts the correct labels within 1 guess. We show 5 successive cases (under top-1 evaluation) on the validation set of the ImageNet dataset. The ground truth is in the top row. Bottom rows: We give the accuracy of CNN and our method. It can be seen that in some categories, our methods perform well.

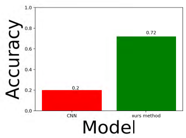
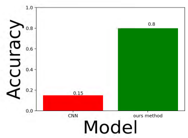
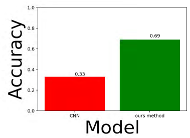
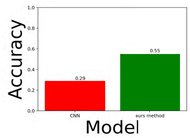
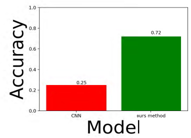
tiger cat	cassette player	maillot	sidewinder	Eskimo dog
				
				

FIGURE 13. Examples of images in which our method fail to predict the correct labels within 1 guess. The ground truth is marked with green, while the predicted label is marked with red. From these failed examples, we can explain that when there are fewer background objects in the picture and the necessary associations are missing, the refinement failure rate is high.

Specifically, in the recognition rate of some specific categories, the accuracy is increased by approximately 10%, as shown in Table 9.

2) FAILURE CASE ANALYSIS

Finally, a number of failure examples by our method from the datasets of ImageNet are presented. These examples are illustrated in Figure 13. From these examples, it is worth noticing that some object classes were easily confused with others. The categories of *wool* and *mitten*(made of wool), and the classes of *bridge* and *tower* were highly ambiguous in some cases. Some categories also had inclusion relationships; for instance, the category of *pole* was sometimes confused with the *flagpole* category, due to their semantic inclusion relationship. Overall, from these failure cases, it can be seen that object classification is still a challenging problem. Semantic variation is essentially a kind of problem in object classification, and in the future, we will consider how to handle the semantic variation problem.

VI. CONCLUSION

In this paper, we used the human brain reasoning mechanism to present the image knowledge graph(IKG) as a biological vision mechanisms to improve the performance of the image classification task. Two major problems in image classification were solved: background complexity(e.g., the surrounding of hatchet or alligator) and visual inconsistency(e.g., cat and tiger cat), and achieve good results on ImageNet datasets. We hope that this work can provide a step towards bringing the knowledge of biological vision into traditional computer vision frameworks. Our next steps will be to apply the IKG to other vision tasks, such as detection, scene recognition, and image captioning.

REFERENCES

[1] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[2] A. Torralba, "Contextual priming for object detection," *Int. J. Comput. Vis.*, vol. 53, no. 2, pp. 169–191, Jul. 2003. doi: 10.1023/A:1023052124951.

[3] L. Liu *et al.* (2018). "Deep learning for generic object detection: A survey," [Online]. Available: http://arxiv.org/abs/1809.02165

[4] A. Torralba, R. Fergus, and W. T. Freeman, "80 million tiny images: A large data set for nonparametric object and scene recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 11, pp. 1958–1970, Nov. 2008.

[5] S. K. Divvala, D. Hoiem, J. H. Hays, A. A. Efros, and M. Hebert, "An empirical study of context in object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2009, pp. 1271–1278.

[6] J. Deng, W. Dong, R. Socher, L. J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2009, pp. 248–255.

[7] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi. (Feb. 2016). "Inception-v4, inception-ResNet and the impact of residual connections on learning." [Online]. Available: https://arxiv.org/abs/1602.07261?context=cs

[8] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 1800–1807.

[9] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2818–2826. doi: 10.1109/CVPR.2016.308.

[10] K. Simonyan and A. Zisserman. (Sep. 2014). "Very deep convolutional networks for large-scale image recognition." [Online]. Available: https://arxiv.org/abs/1409.1556

[11] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, vol. 8689, 2014, pp. 818–833.

[12] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

[13] Y. Yang and J. Jiang, "Hybrid sampling-based clustering ensemble with global and local constitutions," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 5, pp. 952–965, May 2016.

[14] J. Yang, B. Jiang, B. Li, K. Tian, and Z. Lv, "A fast image retrieval method designed for network big data," *IEEE Trans Ind. Informat.*, vol. 13, no. 5, pp. 2350–2359, Oct. 2017.

[15] M. Korytkowski, L. Rutkowski, and R. Scherer, "Fast image classification by boosting fuzzy classifiers," *Inf. Sci.*, vol. 327, pp. 175–182, Jan. 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0020025515006180

[16] L. Shen, Z. Lin, and Q. Huang, "Relay backpropagation for effective learning of deep convolutional neural networks," in *Proc. Eur. Conf. Comput. Vis.*, vol. 7214, no. 1, 2015, pp. 467–482.

[17] I. Biederman, R. J. Mezzanotte, and J. C. Rabinowitz, "Scene perception: Detecting and judging objects undergoing relational violations," *Cogn. Psychol.*, vol. 14, no. 2, pp. 143–177, 1982. [Online]. Available: http://www.sciencedirect.com/science/article/pii/001002858290007X

[18] T. E. Palmer, "The effects of contextual scenes on the identification of objects," *Memory Cogn.*, vol. 3, no. 5, pp. 519–526, 1975.

[19] S. Huang, L. Li, H. Cai, B. Xu, G. Li, and L. Jiang, "A configurable WoT application platform based on spatiotemporal semantic scenarios," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 49, no. 1, pp. 123–135, Jan. 2019.

[20] L. Wolf and S. Bileschi, "A critical view of context," *Int. J. Comput. Vis.*, vol. 69, no. 2, pp. 251–261, 2006.

[21] C. Galleguillos, A. Rabinovich, and S. Belongie, "Object categorization using co-occurrence, location and appearance," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2008, pp. 1–8.

[22] J. Bin, Y. Jiachen, M. Qinggang, L. Baihua, and L. Wen, "A deep evaluator for image retargeting quality by geometrical and contextual interaction," *IEEE Trans. Cybern.*, to be published.

[23] P. Resnik. (2011). "Semantic similarity in a taxonomy: An information-based measure and its application to problems of ambiguity in natural language." [Online]. Available: http://arxiv.org/abs/1105.5444

[24] Z. Wu and M. Palmer, "Verbs semantics and lexical selection," in *Proc. Annu. Meeting Assoc. Comput. Linguistics (ACL)*, 1994, pp. 133–138.

[25] D. Lin, "An information-theoretic definition of similarity," in *Proc. Int. Conf. Mach. Learn.*, 1998, pp. 296–304.

[26] J. Berant, A. Chou, R. Frostig, and P. Liang, "Semantic parsing on freebase from question-answer pairs," in *Proc. Emnlp*, 2013, pp. 1533–1544.

[27] L. Heck, D. Hakkani-Tür, and G. Tur, "Leveraging knowledge graphs for Web-scale unsupervised semantic parsing," in *Proc. Int. Speech Commun. Assoc.*, 2013, pp. 1–5.

[28] R. Hoffmann, C. Zhang, X. Ling, L. Zettlemoyer, and D. S. Weld, "Knowledge-based weak supervision for information extraction of overlapping relations," in *Proc. Meeting Assoc. Comput. Linguistics, Hum. Lang. Technol.*, 2011, pp. 541–550.

[29] J. Daiber, M. Jakob, C. Hokamp, and P. N. Mendes, "Improving efficiency and accuracy in multilingual entity extraction," in *Proc. Int. Conf. Semantic Syst.*, 2013, pp. 121–124.

[30] K. Bollacker, C. Evans, P. Paritosh, T. Sturge, and J. Taylor, "Freebase: A collaboratively created graph database for structuring human knowledge," in *Proc. SIGMOD Conf.*, 2008, pp. 1247–1250.

[31] J. Lehmann *et al.*, "DBpedia—A large-scale, multilingual knowledge base extracted from Wikipedia," *Semantic Web*, vol. 6, no. 2, pp. 167–195, Jan. 2015. doi: 10.3233/SW-140134.

[32] F. Suchanek, G. Kasneci, and G. Weikum, "YAGO: A core of semantic knowledge," Max Planck Inst. Inform., Saarbrücken, Germany, Res. Rep. MPI-I-2006-5-006, Nov. 2006.

[33] A. Carlson, J. Betteridge, B. Kisiel, B. Settles, and T. M. Mitchell, "Toward an architecture for never-ending language learning," in *Proc. 24th AAAI Conf. Artif. Intell.*, 2010, pp. 1306–1313.

[34] C. Xie, P. Yang, and Y. Yang, "Open knowledge accessing method in IoT-based hospital information system for medical record enrichment," *IEEE Access*, vol. 6, pp. 15202–15211, 2018.

[35] C. Xie, H. Cai, L. Xu, L. Jiang, and F. Bu, "Linked semantic model for information resource service toward cloud manufacturing," *IEEE Trans. Ind. Informat.*, vol. 13, no. 6, pp. 3338–3349, Dec. 2017.

[36] C. Xie, G. Li, H. Cai, L. Jiang, and N. N. Xiong, "Dynamic weight-based individual similarity calculation for information searching in social computing," *IEEE Syst. J.*, vol. 11, no. 1, pp. 333–344, Mar. 2017.

[37] C. Xie, H. Cai, Y. Yang, L. Jiang, and P. Yang, "User profiling in elderly healthcare services in China: Scalper detection," *IEEE J. Biomed. Health Inform.*, vol. 22, no. 6, pp. 1796–1806, Nov. 2018.

[38] L. Lu, L. Xu, B. Xu, G. Li, and H. Cai, "Fog computing approach for music cognition system based on machine learning algorithm," *IEEE Trans. Comput. Social Syst.*, vol. 5, no. 4, pp. 1142–1151, Dec. 2018.

[39] L. Wang, S. Guo, W. Huang, Y. Xiong, and Y. Qiao, "Knowledge guided disambiguation for large-scale scene classification with multi-resolution CNNs," *IEEE Trans. Image Process.*, vol. 26, no. 4, pp. 2055–2068, Apr. 2017. doi: 10.1109/TIP.2017.2675339.

[40] Y. Yun and J. Jiang, "Adaptive Bi-weighting toward automatic initialization and model selection for hmm-based hybrid meta-clustering ensembles," *IEEE Trans. Cybern.*, vol. 49, no. 5, pp. 1657–1668, May 2019.

[41] M. Abadi *et al.* (2016). "TensorFlow: Large-scale machine learning on heterogeneous distributed systems." [Online]. Available: https://arxiv.org/abs/1603.04467

[42] C. Fellbaum and G. Miller, "WordNet: An electronic lexical database," *Library Quart. Inf. Community Policy*, vol. 25, no. 2, pp. 292–296, 1998.

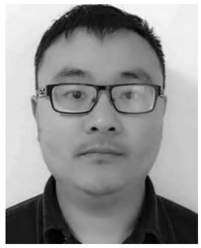
DEHAI ZHANG received the B.Sc. degree (Hons.) in computer science and the M.Sc. degree in basic mathematics from Yunnan Normal University, Kunming, China, in 1999 and 2002, respectively, and the Ph.D. degree in system analysis and integration from Yunnan University, Kunming, in 2009, where he is currently an Associate Professor of artificial intelligence with the National Pilot School of Software. His current research interests include NLP, machine learning, and data analysis.

MENGLONG CUI received the B.S. degree in software engineering from Qujing Normal University, Qujing, China, in 2017. He is currently pursuing the master's degree with Yunnan University, Kunming. His current research interests include image classification, knowledge graph, semantic web, and ontology.

YUN YANG received the B.Sc. degree (Hons.) in information technology and telecommunication from Lancaster University, Lancaster, U.K., in 2004, the M.Sc. degree in advanced computing from Bristol University, Bristol, U.K., in 2005, and the M.Phil. degree in informatics and the Ph.D. degree in computer science from The University of Manchester, Manchester, U.K., in 2006 and 2011, respectively. He was a Research Fellow with the University of Surrey, Surrey, U.K., from 2012 to 2013. He is currently a Full Professor of machine learning with the National Pilot School of Software, Yunnan University, Kunming, China. His current research interests include machine learning, data mining, pattern recognition, and temporal data process and analysis.

PO YANG (M'13) received the B.Sc. degree in computer science from Wuhan University, Wuhan, China, in 2004, the M.Sc. degree in computer science from the University of Bristol, Bristol, U.K., in 2006, and the Ph.D. degree in electronic engineering from the University of Staffordshire, Stoke-on-Trent, U.K., in 2010. He is currently a Senior Lecturer with the Department of Computing Science, Liverpool John Moores University, Liverpool, U.K. He holds a strong tracking of high-quality publications and research experiences. He has published over 60 papers. His current research interests include the Internet of Things, RFID and indoor localization, pervasive health, image processing, GPU, and parallel computing.

CHENG XIE (S'15–M'17) received the B.S. degree in software engineering from the Minzu University of China, Beijing, China, in 2009, and the M.S. and Ph.D. degrees in software engineering from Shanghai Jiao Tong University, Shanghai, China, in 2012 and 2017, respectively. From 2015 to 2016, he was a Visiting Scholar with the Data and Web Science Group, University of Mannheim, Germany. He is currently an Assistant Professor with the School of Software, Yunnan University, Kunming, China. He has published over 20 papers. His research interests include semantic web, linked open data, knowledge graph, and ontology. He received the Shanghai Science and Technology Progress Award (Second Prize), in 2014, the Shanghai Outstanding Graduates, in 2017, the Youth Support Project of China Association for Science and Technology, in 2018, and the Thousand Talents Program of Yunnan, in 2018.



DI LIU received the B.S. and M.S. degrees from Northwestern Polytechnical University, China, in 2007 and 2011, respectively, and the Ph.D. degree from Leiden University, The Netherlands, in 2017. He is currently an Assistant Professor with Yunnan University, China. His research interests include the fields of real-time systems, energy-efficient multicore/many core systems, and cyber-physical systems.

BEIBEI YU is currently pursuing the B.S. degree in information security from Yunnan University, Kunming, China. Her current research interests include knowledge graph and web security.

ZHIBO CHEN received the B.S. and M.S. degrees from Wuhan University, Wuhan, China, in 1984 and 1989, respectively. He is currently an Associate Professor with Yunnan University, China.