IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 7, July 2017

Part Based Image Representation for Find-Grained Category Detection

V. Saranya[1], Prof. M. Jothi[2]

Research Scholar, Dept of Computer Science, JJ College of Arts and Science (Autonomous), Pudukkottai, Tamil Nadu[1]

Asst Professor, Dept of Computer Science, JJ College of Arts and Science (Autonomous), Pudukkottai, Tamil Nadu[2]

Abstract: In this work, propose a system that can be sorted easily with fine-grained images. Do not use any comment objects / parts (weak supervisors) during training or testing, but only the training images of class labels. Sorting fine-grained target images for classification, only subtle differences between objects (e.g. dog's two breeds that look). Most of the existing works rely on the detector items / components to construct the correspondence between the parts of the object, At the very least, you need to train the exact annotation object or part of the object in the image. It points out the need for expensive items to prevent the widespread use of these methods. Instead, we propose to generate useful parts for the proposed size of the proposed component from the object, choose the recommended useful parts and the overall image representation for the calculation of the classification. This is specifically designed for classification fine-grained supervision because it has proven useful to play a key role in dependent works, but the exact detector is difficult to obtain part of the task. With Delegate suggestions can detect and visualize key parts (more discernible) Different types of object images. In the experiment, the proposed method of weak supervision to achieve a considerable or relatively weak advanced control method is more accurate, most of the existing methods Comments on the dependency of the three groups of challenging data. Its success shows that it is not always necessary to learn the expensive detector items / parts fine-grained image classification.

Keywords: classification, fine-grained, image, widespread, visualize key.

I. INTRODUCTION

Image processing is processing of images using mathematical operations by using any form of signal processing for which the input is an image, a series of images, or a video, such as a photograph or video frame; the output of image processing may be either an image or a set of characteristics or parameters related to the image. Most image-processing techniques involve isolating the individual color planes of an image and treating them as two-dimensional signal and applying standard signal-processing techniques to them. Images are also processed as three-dimensional signals with the third-dimension being time or the z-axis. Image processing usually refers to digital image processing, but optical and analog image processing also are possible. This article is about general techniques that apply to all of them. The acquisition of images (producing the input image in the first place) is referred to as imaging. easily recognize that the red box in Fig. 1 contains dogs while the blue box contains a kangaroo. Image representations that used to be useful for general image categorization may fail in fine-grained image categorization, especially when the objects are not well aligned, e.g., the two dogs are in different pose and the backgrounds are cluttered. Therefore, fine-grained categorization requires methods that are more discriminative than those for general image classification. Fine-grained categorization has wide applications in both industry and research societies. Different datasets have been constructed in different domains, butterflies cars etc. These datasets can have significant social impacts, butterflies are used to evaluate the forest ecosystem and climate change One important common feature of many existing fine-grained methods is that they explicitly use annotations of an object or even object parts to depict the object as precisely as possible. Bounding boxes of objects and / or object parts are the most commonly used annotations. Most of them heavily rely on object / part detectors to find the part correspondence.

II. LITERATURE SURVEY

Ms. Dipti S. Borade et al [1]. They described the intent of the image categorization process is to classify the digital image into one of the classes. General image categorization is comparatively easier than fine grained image categorization but it may fail to discriminate objects belonging to same class like birds, cars, plants etc. Fine grained image categorization needs to emphasize on the tiny details that helps to discriminate between similar objects. In the new system, object proposed to extracted from input image. From each object proposal, multi-scale part proposals are generated, from which many useful part proposals are selected. A global image representation is generated using selected useful part proposals. The global image representation is then used to train the classifier for image

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 7, July 2017

categorization. Application areas are forestry, agriculture, industry and research societies. Through survey, we found that both supervised and semi supervised approaches are used for categorization. Some of the methods reviewed in this paper use annotations to perform classification task but in many cases, these are not easily available. Part based methods can be used to improve the accuracy of image classification.

Jia-Lin Chen et al [2]. The discriminating features are obtained using a two-layer Linear Discriminate Analysis (LDA) classification to promise maximal reparability for parts and interactions respectively. Experimental results demonstrate that the proposed system is effective in learning part-based models in less annotated information and achieves comparable performance to state-of-the-art fully supervised approaches. They proposed an interaction prediction system to recognize interaction before it is fully executed. The flexible part-based model of interactions gives tolerance for the distance between each part. The two-layer LDA classification proposed promises maximal separability for parts and interactions respectively. Our experiments demonstrate that the proposed weakly supervised learning scheme is effective in learning the discriminative parts and achieves comparable performance to state-of-the-art fully supervised approaches.

Yu Zhang et al [3]. They proposed a fine-grained image categorization system with easy deployment. Fine-grained image categorization aims to classify objects with only subtle distinctions (e.g., two breeds of dogs that look alike). This is specially designed for the weakly supervised fine-grained categorization task, because useful parts have been shown to play a critical role in existing annotation dependent works but accurate part detectors are hard to acquire. They proposed weakly supervised method achieves comparable or better accuracy than state-of-the-art weakly-supervised methods and most existing annotation-dependent methods on three challenging datasets. Its success suggests that it is not always necessary to learn expensive object/part detectors in fine-grained image categorization. They proposed an efficient multimax pooling strategy to generate multi-scale part proposals by using the internal outputs of CNN on object proposals in each image. Then, we select useful parts from those part clusters which are important for categorization. Finally, they encode the selected parts at different scales separately in a global image representation. With the proposed image / part representation technique, we use it to detect the key parts of objects in different classes, whose visualization results are intuitive and coincide well with rules used by human experts.

Weixia Zhang et al [4]. They proposed a fine-grained image recognition framework by exploiting CNN as the raw feature extractor along with several effective methods including a feature encoding method, a feature weighting method, and a strategy to better incorporate information from multiscale images to further improve recognition ability. Besides, we investigate two dimension reduction methods and successfully merge them to our framework to compact the final image representation. Based on the discriminative and compact framework, we achieved the state-of-the-art performance in terms of classification accuracy on several fine-grained image recognition benchmarks based on weekly supervision. They proposed a novel multi-scale strategy which can be utilized efficiently alone or together with classical image pyramids strategy which has better performance in terms of classification accuracy; secondly, they proposed VLADFV to encode deep convolutional descriptors by concatenating Fisher Vectors and VLAD, resulting in better performance than only using either of them; thirdly, VLAD-FV is rather high-dimensional.

Luming Zhang et al [5]. They described a new fine-grained image categorization system that improves spatial pyramid matching is developed. In this method, multiple cells are combined into cellets in the proposed categorization model, which are highly responsive to an object's fine categories. image categorization can be formulated as the matching between the cellets of corresponding images. Toward an effective matching process, an active learning algorithm that can effectively select a few representative cells for constructing the cellets is adopted. A hierarchical sparse coding algorithm was used to represent each cellet by a linear combination of the basis cellets. The discrimination of a cellet can be selected for fine-grained categorization. Experimental results on three real-world data sets demonstrate that our proposed system outperforms the state of the art.

III. PROBLEM DESCRIPTION

In discriminative parts/modes are selected through the mean shift method on local patches in images for each class. In a set of representative parts are learned using an SVM (support vector machine) classifier with the group sparse constraint for each class in image recognition and segmentation. All these methods tried to evaluate each part, which may be very computationally expensive when the part number is very large. Part based methods have also been used in fine-grained image categorization for a long time. Detailed part annotations are provided with some datasets like CUB 200-2011 where each bird in the image has 15 part annotations. Some methods, for instance directly extract feature vectors from these annotated parts for recognition.also considers generating parts from aligned objects by dividing each object into several segments and assuming that each segment is a part in the object. Some works consider a more practical setup when part annotations are missing in the testing phase. They learn part detectors from annotated parts in the training images and apply them on testing images to detect parts. These part detectors include DPM or object

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 7, July 2017

classifiers learned for each object class used selective search to generate object/part proposals from each image, and applied the learned part detectors on them to detect the head and body in the bird.
The proposal which yields the highest response to a certain part detector is used as the detected part in the object. Convolutional neural networks (CNN) have been widely used in image recognition. The outputs from the inner convolutional (CONV) layers can be seen as the feature representations of sub-regions in the image. When CNN is used on an object proposal, the outputs from the inner convolutional layers can be seen as the part representations, e.g., used CNN on detected objects, and used the outputs from CONV4 (in Alex net) as the parts used the outputs from all layers in CNN and selected some important ones as parts. Recently, CNN aided by region proposal methods, has become popular in object recognition/detection, e.g., RCNN, fast-RCNN, faster-RCNN, and RCNN-minus-R. All these four methods focus on the supervised object detection, where object bounding boxes in training images are necessary to learn the object detectors. They cannot be directly used in our weakly-supervised fine-grained image categorization. These methods generate object level representations, while ours used fine-grained part level representations. In RCNN, CNN is applied on each object proposal (bounding box acquired by selective search on the input image) and the output from the fully connected layer is used as the feature vector, where CNN is applied multiple times on an image. In FastRCNN, CNN is only applied once on the whole image. The bounding boxes of object proposals are mapped to the final convolutional (CONV) layer to get the object feature. Similarly, RCNN-minus-R used sliding windows to map to the last CONV layer in CNN in order to get the object representation. In Faster-RCNN, instead of mapping object proposal from input images, sliding windows are directly used on the last CONV layer to get the object feature.

IV. METHODOLOGY

A. Part Based Methods:

Part representation has been investigated in general image recognition. In, over-segmented regions in images are used as parts and LDA (linear discriminant analysis) is used to learn the most discriminative ones for scene recognition. In discriminative parts/modes are selected through the mean shift method on local patches in images for each class. A set of representative parts are learned using an SVM (support vector machine) classifier with the group sparse constraint for each class in image recognition and segmentation. All these methods tried to evaluate each part, which may be very computationally expensive when the part number is very large. Part based methods have also been used in fine-grained image categorization for a long time. Detailed part annotations are provided with some datasets where each bird in the image has 15 part annotations. Some methods, for instance, directly extract feature vectors from these annotated parts for recognition. Also considers generating parts from aligned objects by dividing each object into several segments and assuming that each segment is a part in the object. When a CNN model is applied on an object bounding box in an image, the acquired receptive fields from MMP can be seen as the part candidates for the object. Thus, we can acquire a multi-scale representation of parts in objects with MMP. To compute the part proposals, we first generate object proposals from each image. Object proposals are those regions inside an image that have high objectness, i.e., having a higher chance to contain an object. Since no object/part annotations are utilized, we could only use unsupervised object detection methods. Selective search is used in our framework given its high computation efficiency, which has also been used to generate initial object/part candidates for object detectors.

a. Multi-scale Image Representation:

Considering our part proposals are generated at different scales aggregating all of them into a single image representation cannot highlight the subtle distinction in fine-grained images. Thus, we propose to encode part proposals in an image on different scales separately and we name it Scale Pyramid Matching (ScPM).

B .Weakly Supervised Fine-grained Categorization

Most existing fine-grained works heavily rely on the object / part annotations in categorization when the objects are in complex backgrounds is the first work which categorizes fine-grained images without using human annotations in any image (both training and testing), but with only image labels. In a CNN that is pre-trained from Image Net is first used as an object detector to detect the object from each image. Then, part features (outputs from CONV4 in CNN) are extracted from objects and clustered into several important ones by spectral clustering. For each part cluster, a part detector is learned to differentiate it from other clusters. Finally, these part detectors are used to detect useful parts in testing images. In each part is evaluated extensively by the learned part detectors and the detected ones are concatenated into the final image representation. In contrast, our method first encodes the large number of parts into a global image representation and then performs part selection on it, which can save much more computational effort than also categorized fine-grained images in the same setup.

C. Multi-scale Image Representation

Considering our part proposals are generated at different scales (with different M in Eq. 1), aggregating all of them into a single image representation cannot highlight the subtle distinction in fine-grained images. Thus, we propose to encode

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 7, July 2017

part proposals in an image on different scales separately and we name it Scale Pyramid Matching (ScPM). The steps are as follows:
Generate parts on different scales. Given an image I, which contains a set of object proposals I = {o1, ..., o|I|}, each object proposal oi contains a set of multi-scale part proposals oi = {z1, ..., z|oi|}. For part proposals in I on different scales M ∈ {1, ..., N}, we compute separate FVs. In practice, the scale number can be very large (N = 13 in the CNN setting), which may lead to a severe memory problem. Since the part proposals on neighboring scales are similar in size, we can divide all the scales into m (m ≤ N) non-overlapping groups {g(j), j = 1, . . . , m, g(j) ⊆ {1, . . . , N}}.

a. Part-based R-CNN classifiers

T he next step is to integrate the learned R-CNN detector results and use them to train fine-grained classifiers. In part-based R-CNN, proposed three types of geometric constraint to ensure that the relative location of detected objects and their semantic parts follow a geometric prior. Here, however, the strength and robustness of R-CNN part detectors result in geometric constraints that only play a minor role in detection, especially considering that fine-grained datasets usually contain only a relatively limited number of training images. Therefore, in our implementation, we only conduct a simple box constraint to ensure object parts do not fall outside the root bounding box. For an image I, let X = {x0, x1, ..., xp} be the predicted locations (bounding boxes) of an object and its parts, which are given during training, but unknown for both weakly supervised images and testing images. The final feature representation is then denoted as Φ(x) = [φ (0)(x0), ..., φ(n) (xn)], where φ (i) (xi) is the feature representation for part pi as the output of the fc7 layer of the i-th part-CNN. Beyond them, a one-versus-all linear SVM is trained for each fine-grained category.

V. EXPERIMENTAL RESULTS

A.SYSTEM ARCHITECTURE

The system's working is elaborated in the following steps:
1) Query image is given as an input to the system.
2) Image is segmented using graph based approach.
3) Features are extracted using the Selective Search method. It results in extracting object proposals from segmented image.
4) Filtering is performed on object proposals.
5) Principal Component Analysis is used for dimensionality reduction.
6) Clustering by using k-means clustering algorithm.
7) Fisher Vector encoding step.
8) Classify the input query image using SVM classifier.

a. Benchmark Datasets

1) CUB 200-2011 Dataset: The Caltech-UCSD Birds 200-2011 is an image dataset with the bird species (mostly North American). This dataset contains 200 bird categories. This dataset is challenging since it is difficult to classify the birds to one of the categories. It contains total 11788 images out of which 5994 images are used for training and 5794 are used for testing purpose.

Fig. Sample Images from Caltech USSD Birds 200-2011 dataset

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 7, July 2017

2) StanfordDogs Dataset: The Stanford Dogs dataset consist of 120 classes of dogs. There are 120 images per class. Total number of images are 20580. From which, 12000 images

b. Graph Based Image Segmentation

Input image is segmented by using graph based approach. Let G = (V, E) be an undirected graph with vertices vi 2 V , the set of elements to be segmented, and edges (vi, vj) 2 E corresponding to pairs of neighbouring vertices. Each edge (vi, vj) 2 E has a corresponding weight w((vi, vj)), which is a nonnegative measure of the dissimilarity between neighbouring elements vi and vj . In the case of image segmentation, the elements in V are pixels and the weight of an edge is some measure of the dissimilarity between the two pixels connected by that edge. Weight of an edge is the difference between pixel intensities. The region pairing is performed on the graph and result is the image with the regions formed. Each region refers to the set of pixels.

c. Feature Extraction

Features are extracted at the level of object proposals. An object proposal is the patch from an image which is most likely to contain an object. The concept of objectness is used here to define the term 'object proposal'. Selective Search method is used to extract object proposals from an input image. This method captures the object proposals effectively as well as it is fast to compute. The selective search use Hierarchical Grouping Method to group the regions. The set of regions generated in previous step is the input for Hierarchical Grouping Method. Using these regions as an input, Hierarchical Grouping algorithm is used to obtain object proposals. Initially, the set of similarities are calculated for the neighbouring regions.

d. Filtering

Part proposals are the sub regions within an object proposals. Part proposals are generated by using the Convolution and pooling layers. The object proposal generated by Selective Search[11] is the input for filtering process. The filtering process uses 3 filters 7_7, 5_5 and 3_3. The filters are convolved with the spatial regions. Pooling is performed after convolution operation. The MultiMax Pooling (MMP) technique is used to collaborate the information from all the spatial locations.

E. Optimization

A fixed length feature vector is generated as the output of filtering process. This feature vector is then reduced by using Principal Component Analysis. So that the memory required to store the features will be less. Classification is performed to cluster the object proposals. The mean and standard deviation for each cluster are used to compute one fisher vector for respective cluster.

B. EXPERIMENTAL SETUP

In the training phase, the system is trained using training image database and fisher vectors are generated as the output and are stored in a feature database. In testing phase, a query image will be taken as an input and the system generates image label for the query image.

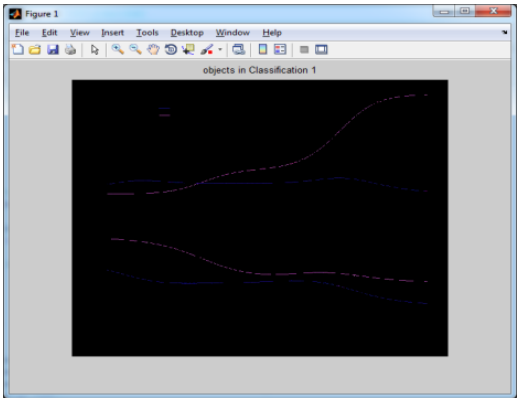
IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 7, July 2017

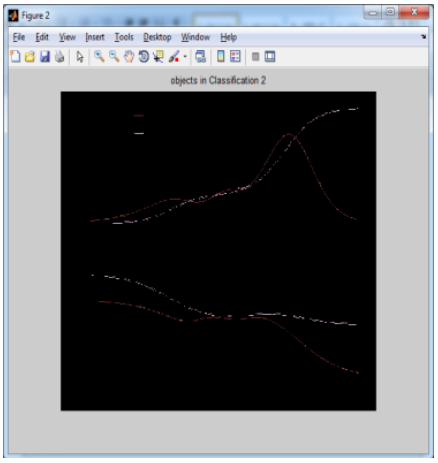
C. SYSTEM ANALYSIS

A. Performance Metrics Performance of the image classification system can be measured by using the parameters classification accuracy and time required for computation.



a. Classification Accuracy:

Classification accuracy for the image classification system can be defined as; the total number of images correctly classified divided by the total number of images classified. Total no. of images correctly classified is calculated based on ground truth image labels.

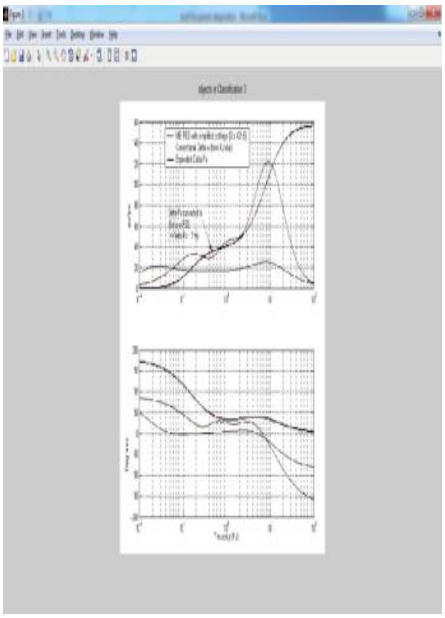


IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 7, July 2017

b. Computation Time: It can be defined as total time required classifying the query image.



VI. CONCLUSION

In this paper, we have proposed to categorize fine-grained images without using any object/part annotation either in the training or in the testing stage. Our basic idea is to select multiple useful parts from multi-scale part proposals and use them to compute a global image representation for categorization. This is specially designed for fine-grained categorization in the weakly-supervised scenario, because parts have been shown to play an important role in the existing annotation dependent works. Also, accurate part detectors are usually hard to acquire. Particularly, we propose an efficient multimax pooling strategy to generate multi-scale part proposals by using the internal outputs of CNN on object proposals in each image. Then, we select useful parts from those part clusters which are important for categorization. Finally, we encode the selected parts at different scales separately in a global image representation. With the proposed image / part representation technique, we use it to detect the key parts of objects in different classes, whose visualization results are intuitive and coincide well with rules used by human experts.

REFERENCES

[1] Ms. Dipti S. Borade and Prof. Nitin M. Shahane, "A Review on Fine Grained Categorization of an Image using Part Proposals", International Academy of Engineering and Medical Research, 2017.
[2] Jia-Lin Chen , "Weakly Supervised Learning of Part-based Models for Interaction Prediction via LDA", ACM, 2015.
[3] Yu Zhang, Xiu-Shen Wei, Jianxin Wu, Jianfei Cai, Jiangbo Lu, Viet-Anh Nguyen, and Minh N. Do, "Weakly Supervised Fine-Grained Categorization with Part-Based Image Representation", IEEE, 2015.
[4] Weixia Zhang , Jia Yan , Wenxuan Shi , Tianpeng Feng and Dexiang Deng, "Refining deep convolutional features for improving fine-grained image recognition", EURASIP Journal on Image and Video Processing, 2017.
[5] Luming Zhang, Yue Gao, Yingjie Xia, Qionghai Dai, and Xuelong Li, "A Fine-Grained Image Categorization System by Cellet-Encoded Spatial Pyramid Modeling", IEEE TRANSACTIONS ON INDUSTRIAL ELECTRONICS, 2015.
[6] A. Vedaldi, S. Mahindra, S. Tsogkas, S. Maji, B. Girshick, J. Kannala, E. Rahtu, I. Kokkinos, M. B. Blaschko, D. Weiss, B. Taskar, K. Simonyan, N. Saphra, and S. Mohamed, "Understanding objects in detail with fine-grained attributes," in Proc. IEEE Int'l Conf. on Computer Vision and Pattern Recognition, 2014, pp. 3622–3629.
[7] M.-E. Nilsback and A. Zisserman, "Automated flower classification over a large number of classes," in Indian Conf. on Computer Vision, Graphics and Image Processing, 2008, pp. 722–729.
[8] L. Bourdev, S. Maji, T. Brox, and J. Malik, "Detecting people using mutually consistent poselet activations," in Proc. European Conf. Computer Vision, vol. LNCS 6316, 2010, pp. 168–181.
[9] N. Zhang, R. Farrell, F. Iandola, and T. Darrell, "Deformable part descriptors for fine-grained recognition and attribute prediction," in Proc. IEEE Int'l Conf. on Computer Vision and Pattern Recognition, 2013, pp. 729–736.
[10] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan, "Object detection with discriminatively trained part based models," IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 32, pp. 1627– 1645, 2010.