Flower Detection using Weakly supervised learning

kartik vishwakarma
2017csm1001@iitrpr.ac.in

Department of Computer Science and Engineering,
Indian Institute of Technology Ropar, punjab

1 Abstract

In this project I implemented weakly supervised object detection, to detection flower (intra-class detection). 17-Flower data-set is used in this project. Unlike previous implementation, where others try to transform a classifier into detection using domain knowledge transfer from classifier to detection using available fully label categories. My approach detects object without any domain knowledge. In this approach it first finds most semantic similar class with respect to available fully label dataset and use their boundary box as ground truth for training network as detector. k-NN classifier used to find most semantic similar class further in class to get further more finer boundary Image Retrieval with Bag of Visual Words method used. Now based on threshold value for each class, weakly label data are mapped into strongly label dataset.

2 INTRODUCTION

Object classification, refer to differentiating two objects in an image saying what object they are. while Object Detection, is process to categorize and locate all known content in image. Object detection is difficult task than Object classification. In supervised Object detection involved image-level annotation and object-level annotation, while in Object Classification only image-level annotation is sufficient to classify. Recent CNN based object detection, such as DetectorNet [1], Faster R-CNN [8], YOLO [5] is dependent on large amount of training data manually labeled with object localizations(e.g PASCAL VOC [6] datasets). Previously many method proposed in weakly supervised object detection like large scale domain adaptation(LSDA) [4] [12], learning on multiple instance learning(MIL) [9], using object detection with posterior regularization [3]. All these method are use to detect inter-class object while my approach uses intra-class object detection. Generally detecting intra-class object is difficult that inter class detection since semantic feature in intra-class are similar in compare to inter-class.

3 RELATED WORK

In people so far worked on inter class weakly supervised learning and achieve performance (mAP %) to 51.8 very close to fully supervised object detection with mAP 61.8 % on PASCAL VOC 2007 [6] test set using mixed supervised. In this method it irst learn domain-invariant objectness knowledge from the existing fully labeled categories and utilize multiple instance learning (MIL) to model the concepts of both objects and distractors and to further improve the ability of rejecting distractors in weakly labeled images. Other appoaches [4] [11] [13] [2] are based on domain transfer.

4 Task definition

In my method task are pipe lined into 3-stages. Initially, 1) finding class similar to class of fully label class. 2) after class getting class it required further finding image with similar feature such a way that boundary box may best for that image. 3) after finding one-one matching train model for detection.

4.1 finding similar class

In 17-Flower dataset (17 class), Dataset consist of only 2-class with fully labeled and rest 15-class as weakly labeled, for better performance, i map one labeled class to only one unlabeled class, so i have 2 fully labeled class and with help of these data further expand 2 more class as follow:
from rest 15 weakly labeled dataset i apply KNN classifier trained on 2 fully label dataset using K=3, as found class-5 and class-15 are most closed to available fully labeled class-1 and class-2.

KNN-Classifier (K=3)		
Class(unlabeled)	class-7	class-16
0	11	69
1	27	53
2	33	47
3	51	29
4	36	44
5	56	24
6	22	58
8	28	52
9	9	71
10	21	59
11	9	71
12	13	67
13	17	63
14	16	64
15	25	55
16	57	23

KNN-classifier uses two feature to classify images, 1) using feature vector, 2) color histogram
I used class-5 and class-15 in my experiment and map class-5 to class-1 and class-15 to class-2.

KNN-Classification https://www.pyimagesearch.com/2016/08/08/k-nn-classifier-for-image-classification/

4.2 image to image mapping

To achieve better boundary boxes, i further find most similar image of unlabel class correspond to label class. I used SURF same feature extractor (using matlab) and for list of return value from feature extractor ("retrieveImages ()") choose top one with score information, Now to generate annotate data for unlabel images from label images, based on threshold value images are selected, threshold value changes per class since feature vector varies per class.

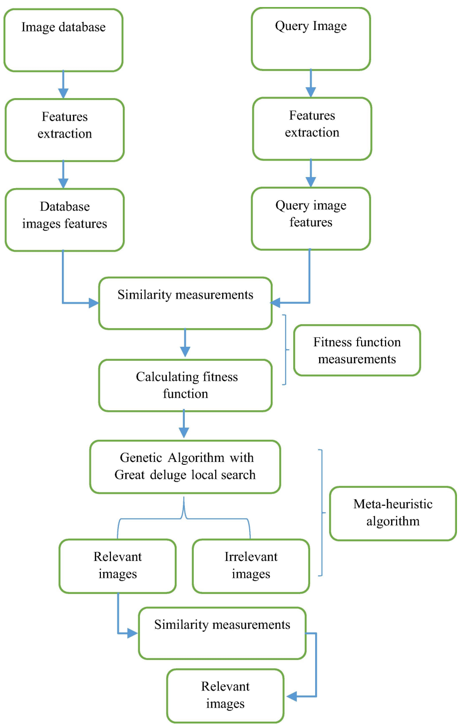


Figure 1: Flow diagram of content based image retrival

5 Training

Now from previous step i have transformed, weakly label dataset into strongly label dataset, now i can fed/train into any fully supervised object detector [7] fastecnn [5], [10]. For ease of understanding and observation, i train model on R-CNN model as follow:
1)R-CNN model trained as classifier on 17-class 80 images per class of image size 224x224x3 with batch size 64, learning rate 0.0001, decay at rate 0.99 after each 100 iterations for 20000 iteration total about 998 epochs.
R-CNN Architecture

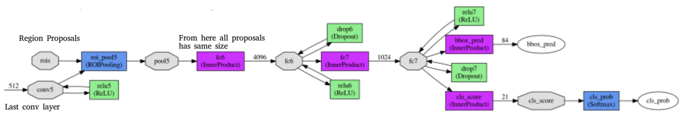


Figure 2: R-CNN network architecture for classification and detection.

After model trained , it finetune using 5 classes, 30 image per class (labeled) 1-class for background, 2-class from fully annotated data and 2-class generated using above methods using batch size 256, learning rate 0.001 decay at rate 0.99 after each 1000 iteration.
After fine tune SVM model trained with 0.3 and Regression trained threshold 0.6 to generate boundary box.

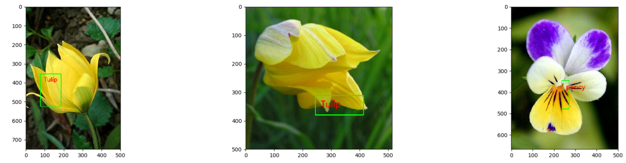
6 Experiment

For ease of observation and training and testing it experimented on very small dataset. Dataset contains only 17 classes, from 17 classes only 2 classes are fully label (tulip pancy), in training data 80 images per class and for fine tune 30 image per class where used.
Model experimented for various iteration values i.e. 10,000 iterations, 35,000 iterations, and 1,00,000 iteration and it found that model performing well only for 10,000 iteration value, for higher value model get biased along a class, reason may be very few images per class. Further, i train model using VGG-16 architecture, but while model didn't converges by tuning learning rate too.

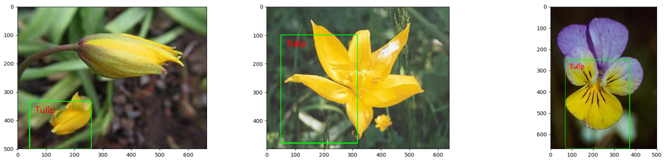
7 Result

In below figure shows correctly classified, localized and detected examples. Over all performance is 29% low, since i used R-CNN model as object detection and for fully supervised learning its performace it only 58% (mAP). Further Since images within same class have huge variation it is very hard to clearly classify on shallower network like Alexnet used in my experiment.
Because of very few training data and lots of variation of background model easily get distracted and wrongly classified images, training model with large data VOC PASCAL will greatly improve performance.

From experiment, i found that increase in training step does not help to improve accuracy and network detects particular sub-region rather than cover whole object region. Also model get biased along particular class and most of time it wrongly classify images. Below show image it clear that even for different class it find features of class along which it get biased.



Some of my failure cases where object either misclassified or poor localized



8 Further work

I have work on very few ans noisy dataset, working with huge dataset like PASCAL VOC 2012 or Imagenet2013 may increase performance. Also using object localization information will help to find better map of unlabel image into label images. Performance may further increase by rejecting background class and learn model using forward images, since background act as distractor for model while training and result in poor classification accuracy.
One way of rejecting background it, to use multiple instance learning (MIL) to select interested region and modeled as binary classifier 0 as background class and 1 as forward class.
Further feature of label data may be learn by unlabel data using reverse (-1 x g) gradient propagation to fill gap between these classes.

Replace R-CNN with more accurate object detection model like Faster-RCNN will improve performance significantly.

9 Conclusion

Aim of my experiment to try to detection object with help of boundary box information of other classes, in real world many object are very similar in shape e.g. dogs similar to dogs, bicycle similar to motorcycle etc. This method will failed with object with unique shape like human, chair etc.

10 References

[1] A. Toshev C. Szegedy and D. Erhan. Deep neural networks for object detection,âĂİ in neural information processing systems (nips), 2013.

[2] Tegan Maharaj Samira Ebrahimi Kahou4 Prabhat2 Christopher Pal Evan Racah, Christopher Beckham. âĂIJextremeweather: A large-scale climate dataset for semi-supervised detection, localization, and understanding of extreme weather events âĂİ, 2017.

[3] Tinne Tuytelaars Hakan Bilen, Macro Pedersoli. "weakly supervised object detection with posterior regularization".

[4] E. Tzeng R. Hu J. Donahue R. Girshick T. Darrell J. Hoffman, S. Guadarrama and K. Saenko. âĂIJlsda: Large scale detection through adaptation,âĂİ in neural information processing systems (nips), 2014.

[5] R. Girshick J. Redmon, S. Divvala and A. Farhadi. You only look once: Unified, real-time object detection,âĂİ in ieee conference on computer vision and pattern recognition (cvpr), 2016.

[6] C. Williams J. Winn M. Everingham, L. Van Gool and âĂIJThe Pascal visual object classes (VOC) challenge A. Zisserman. International journal of computer vision (ijcv), 2010.

[7] T. Darrell R. Girshick, J. Donahue and J. Malik. "rich feature hierarchies for accurate object detection and semantic segmentation".

[8] R. Girshick S. Ren, K. He and J. Sun. Faster r-cnn,âĂİ in neural information processing systems (nips), 2015.

[9] R. H. Lathrop T. G. Dietterich and T. Lozano-Perez. "solving the multiple instance problem with axis-parallel rectangles", 1997.

[10] Dumitru Erhan Christian Szegedy Scott Reed Cheng-Yang Fu Alexander C. Berg Wei Liu, Dragomir Anguelov. âĂIJssd: Single shot multibox detector âĂİ, 2015.

[11] Kaiqi Huang Jianguo Zhang Yan LI, Junge Zhang. "mixed supervised object detection with robust objectness transfer", 2018.

[12] Xiaofang Wang Boyang Gao Emmanuel DellandreÌĄa Robert Gaizauskas Yuxing Tang, Josiah Wang and Liming Chen Senior Member. âĂIJlarge scale semi-supervised object detection using visual and semantic knowledge transferâĂİ, 2018.

[13] Weishen Pan Jin Li Changshui Zhang Ziang Yan, Jian Liang. âĂIJweakly- and semi-supervised object detection with expectation-maximization algorithmâĂİ, 2017.