Object Proposals using Nonparametric Bounding Box Transfer

Neelima Chavali
Virginia Tech
Blacksburg VA, 24060
gneelima@vt.edu

Dhruv Batra
Virginia Tech
Blacksburg, VA, 24060
dbatra@vt.edu

Abstract

We present a novel nonparametric, category-independent approach for generating bounding box proposals which are likely to contain objects in an image. Given a query image our system finds its nearest neighbours from a large database containing images annotated with bounding boxes. We then establishe a dense correspondence between the query image and each of the nearest neighbours using the SIFT flow algorithm [9]. Based on the correspondences, bounding boxes in the neighbours are warped to the query image to produce candidate bounding boxes. When compared at the same number of proposals, our approach outperforms the state-of-the-art technique of Selective Search [10] when using DeCAF feature as well as traditional features like GIST and HOG.

1. Introduction

In recent years we have seen a marked shift towards algorithms which generate candidate regions or bounding boxes in object detection pipelines. [10, 1, 4, 2, 6]. However, most of these algorithms are learning based parametric algorithms. On the other hand, with emergence of large databases such as PASCAL [5] and ImageNet [3], a new family of non parametric data-driven methods have demonstrated impressive performance in several areas of computer vision [8, 7].

In this paper, we propose a novel, nonparametric bounding box transfer (NPBT) system that produces category-independent bounding box proposals for a test image by transferring over the annotated bounding boxes from similar images in a large dataset as illustrated in Figure 1. Our approach is similar in spirit to that of [8]; the key difference being that they attempt to copy over the entire image labelling, which is unlikely to be very accurate even with large datasets, while we copy over bounding box proposals which is more likely to succeed especially since these proposals simply need to be handed to a categorization algorithm downstream in the pipeline.

2. Approach

Figure 1 shows the pipeline of our approach, which involves the following steps:

- **Nearest neighbour retrieval**: Given a query image, retrieve a set of *k* nearest neighbors from a dataset in a particular feature space. We experiment with the DeCAF feature and other traditional features like GIST and HOG.
- **Dense image alignment**: Establish *dense pixelwise correspondence* between the query image and each of the retrieved nearest neighbors. Use the SIFT flow algorithm [9] for calculating correspondences.
- **Bounding box transfer**: Warp the bounding box annotations from the nearest neighbors to the query image according to the estimated dense correspondence. Perform non-maximal suppression on the transferred bounding boxes to get the bounding boxes of the query image.

We use Average Best Overlap (ABO) defined in [10] as a metric to evaluate our performance. To measure ABO, we calculate the best overlap between each ground truth annotation $g_i \in G$ and the object hypotheses L generated for the corresponding image and averaged over all ground-truth annotations:

$$ABO = \frac{1}{|G|} \sum_{g_i \in G} \max_{l_j \in L} \text{Overlap}(g_i, l_j) \tag{1}$$

where $Overlap$ is defined as

$$Overlap(g_i, l_j) = \frac{area(g_i) \cap area(l_j)}{area(g_i) \cup area(l_j)} \tag{2}$$

3. Results

We evaluate the performance of our method as follows. For a given query image from PASCAL VOC 2007 test set, we calculate $k = 5, 10, 15, 20$ nearest neighbors from PASCAL VOC 2007 train val set using DeCAF, GIST, HOG.

Figure 1. For a query image (a), our system finds the top matches (nine are shown here) (b) Using SIFT flow matching algorithm [9], the bounding boxes of the top matches are transferred to the input image, as shown in (c). For comparison, the ground-truth user annotation of (a) is shown in (d).

We compare the ABO values in each case with ABO of the top $n = 25, 50, 75, 100$ proposals of [10]. Figure 2 shows the ABO achieved by NPBT on PASCAL VOC test set as a function of the number of bounding boxes predicted. We compare NPBT with Selective Search(SS), which is widely considered the state-of-the-art technique for generating bounding box proposals. We use the publicly available implementation of SS in "Quality mode". We sorted the selective search bounding boxes with respect to the score reported by SS.

We observe that NPBT-DeCAF @ 67 bounding box proposals outperforms SS @ 100 proposals by 16%.

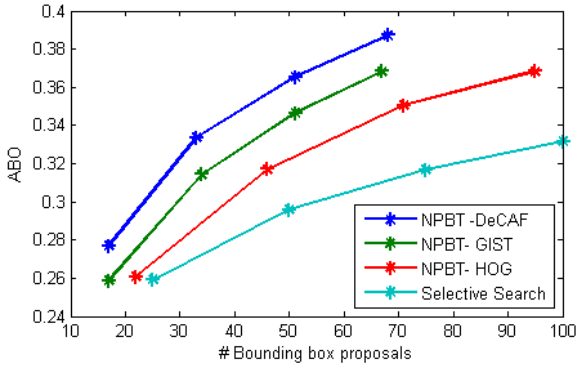

Figure 2. Comparison of NPBT against the Selective Search

4. Conclusion

We present a novel approach for generating object proposals by transfering bounding boxes using SIFT flow. Our approach outperforms the current state of art. We note that SS and other techniqeus typically report results at ˜2k bounding boxes. However, we argue that in order for recognition to scale, we must produce high ABO at small number of proposals. There are several avenues for future work, for instance how our performance varies as we increase the size of the dataset or with SIFT flow alignment;

Acknowledgements. This work was partially supported by the National Science Foundation under Grant No. IIS-1353694.

References

[1] B. Alexe, T. Deselaers, and V. Ferrari. What is an object? In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 73–80. IEEE, 2010. 1

[2] J. Carreira and C. Sminchisescu. Constrained parametric min-cuts for automatic object segmentation. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 3241–3248. IEEE, 2010. 1

[3] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. ImageNet: A Large-Scale Hierarchical Image Database. In *CVPR*, 2009. 1

[4] I. Endres and D. Hoiem. Category independent object proposals. In *Computer Vision–ECCV 2010*, pages 575–588. Springer, 2010. 1

[5] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The pascal visual object classes (voc) challenge. *IJCV*, 88(2):303–338, June 2010. 1

[6] R. Girshick, J. Donahue, T. Darrell, and J. Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In *Computer Vision and Pattern Recognition*, 2014. 1

[7] D. Kuettel and V. Ferrari. Figure-ground segmentation by transferring window masks. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 558–565. IEEE, 2012. 1

[8] C. Liu, J. Yuen, and A. Torralba. Nonparametric scene parsing via label transfer. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 33(12):2368–2382, 2011. 1

[9] C. Liu, J. Yuen, A. Torralba, J. Sivic, and W. T. Freeman. Sift flow: Dense correspondence across different scenes. In *Computer Vision–ECCV 2008*, pages 28–42. Springer, 2008. 1, 2

[10] J. Uijlings, K. van de Sande, T. Gevers, and A. Smeulders. Selective search for object recognition. *International journal of computer vision*, 104(2):154–171, 2013. 1, 2