Spatial Pooling of Heterogeneous Features for Image Applications

Lingxi Xie[1], Qi Tian[2], and Bo Zhang[1]
[1]Department of Computer Science and Technology, Tsinghua University, Beijing, China
[2]Department of Computer Science, University of Texas at San Antonio, Texas, USA

ABSTRACT

The Bag-of-Features (BoF) model has played an important role for image representation in many multimedia applications. Despite the advantages of this model, there are also notable drawbacks, including poor power of semantic expression of local descriptors, and lack of robust structures upon single visual words. To overcome these problems, various techniques have been proposed, such as multiple descriptors, spatial context modeling and interest region detection. Though they have been proven to improve the BoF model to some extent, there still lacks a coherent scheme to integrate each individual module.
To address the problems above, we propose a novel framework. Our model differs from the traditional ones on three aspects. First, we propose a new scheme for combining texture and edge based local features together at an early stage. Next, we build geometric visual phrases for mid-level representation of images. Finally, we perform a simple and effective spatial weighting scheme. We test our integrated framework on several benchmark datasets for image classification and retrieval applications. The extensive results show the superior performance of our algorithm over state-of-the-art methods.

NOVELTY

Three key observations for improving the BoF framework:

1. Enhancing descriptions of local patches.
2. Mid-level representation connecting low-level and high-level concepts.
3. Spatial weighting of images.

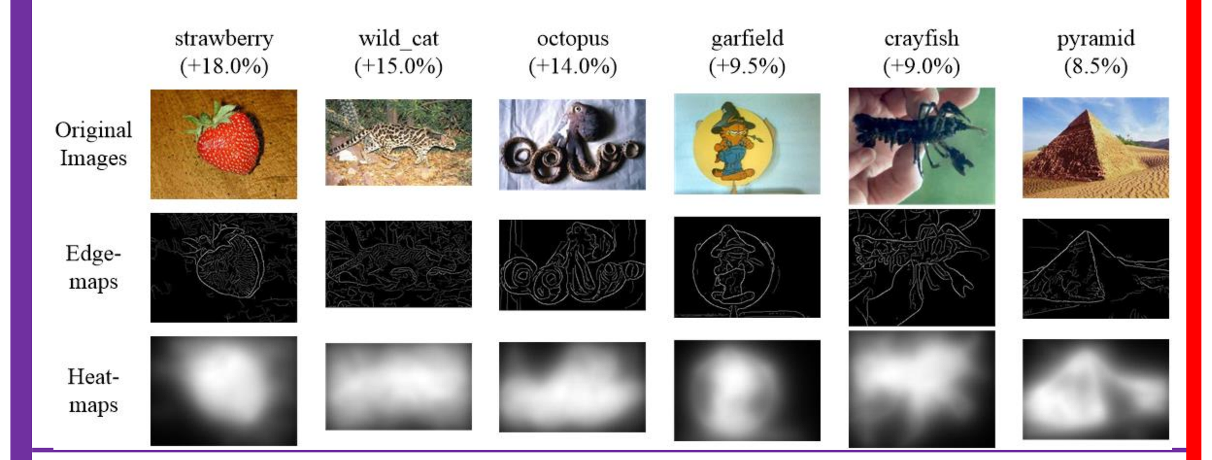
Based on the observations, we propose several novel algorithms from new aspects. We claim a THREEFOLD contribution:

1. We simultaneously extract SIFT and Edge-SIFT descriptors and combine them in the generation of the BoF model. Earlier fusion of descriptors makes it easier to mine complementary information from both descriptors.
2. We propose Geometric Visual Phrases (GVP) upon traditional visual words, and take them as mid-level image representation as well as apply a novel pooling algorithm, Geometric Phrase Pooling (GPP), to them.
3. We use naive Gaussian blur to obtain a weighting heatmap for spatial weighting on the image plane.

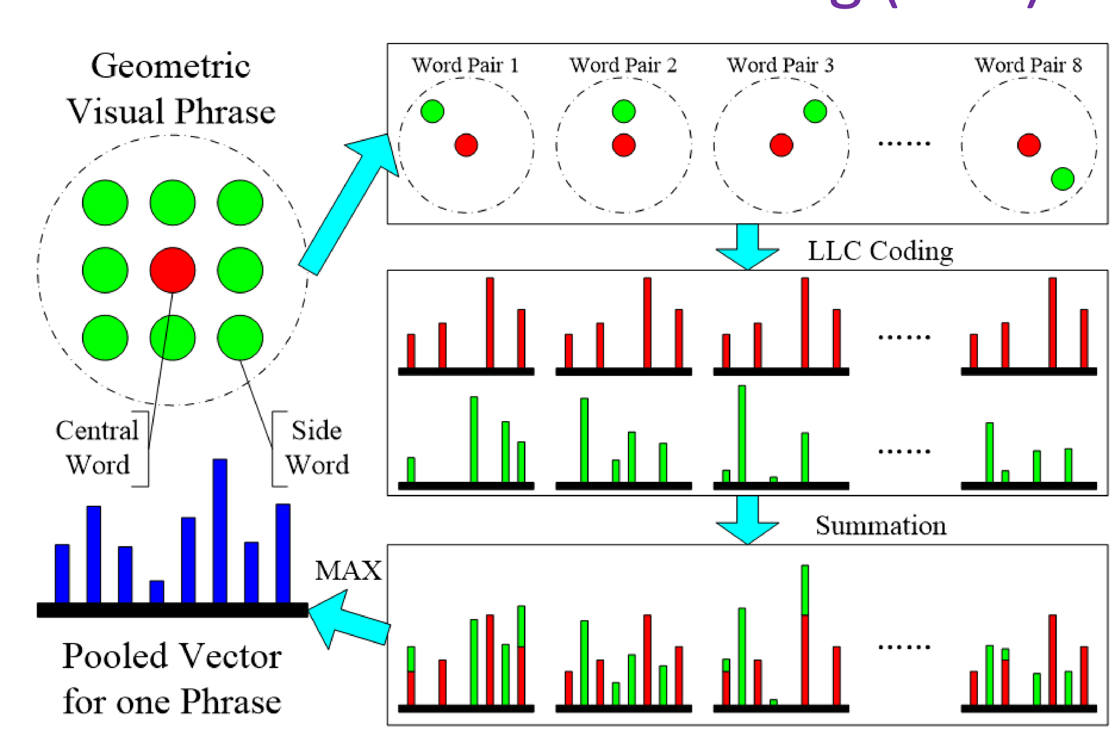
Integrating all the techniques produces a much more powerful framework, which outperforms state-of-the-art systems by a margin on various applications.

THE PROPOSED FRAMEWORK

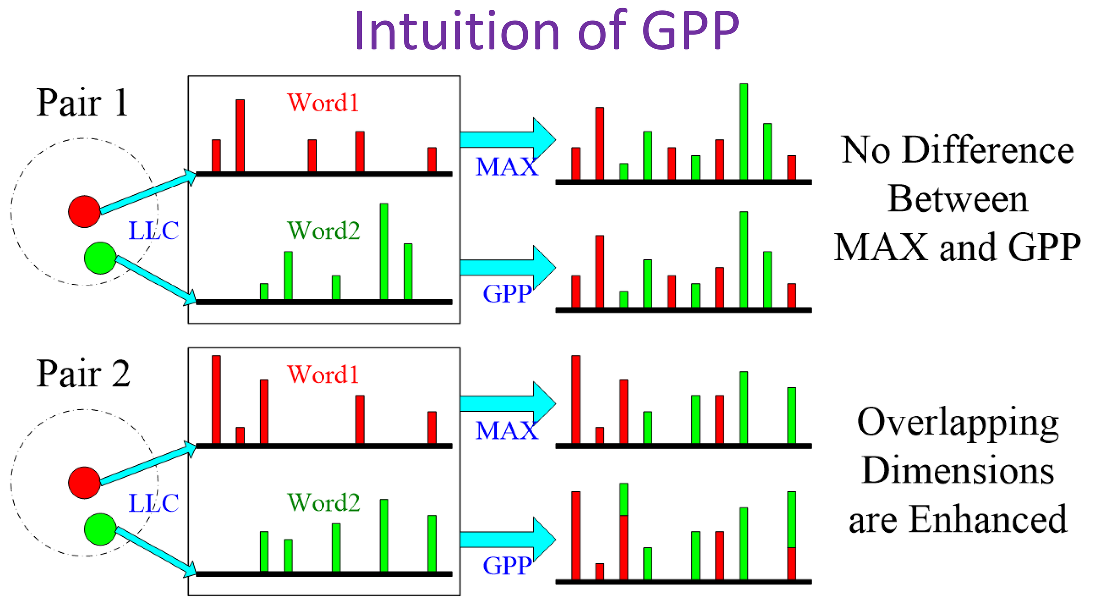
Edge-based Spatial Weighting



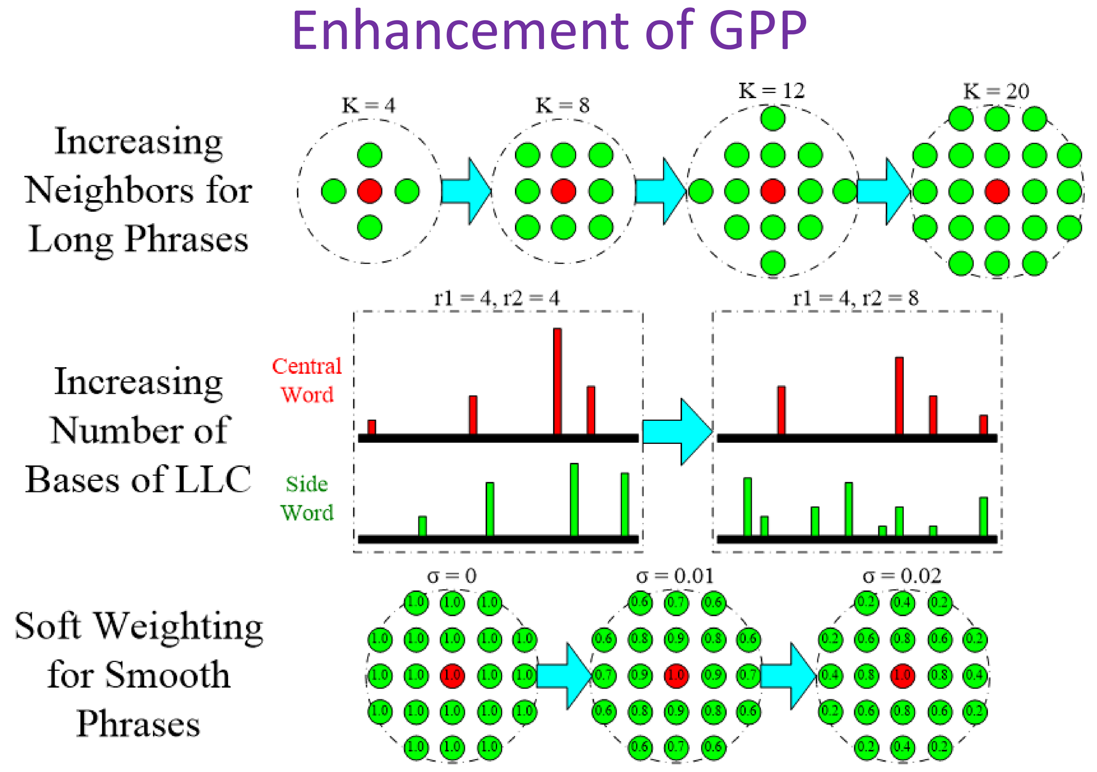
Geometric Phrase Pooling (GPP)

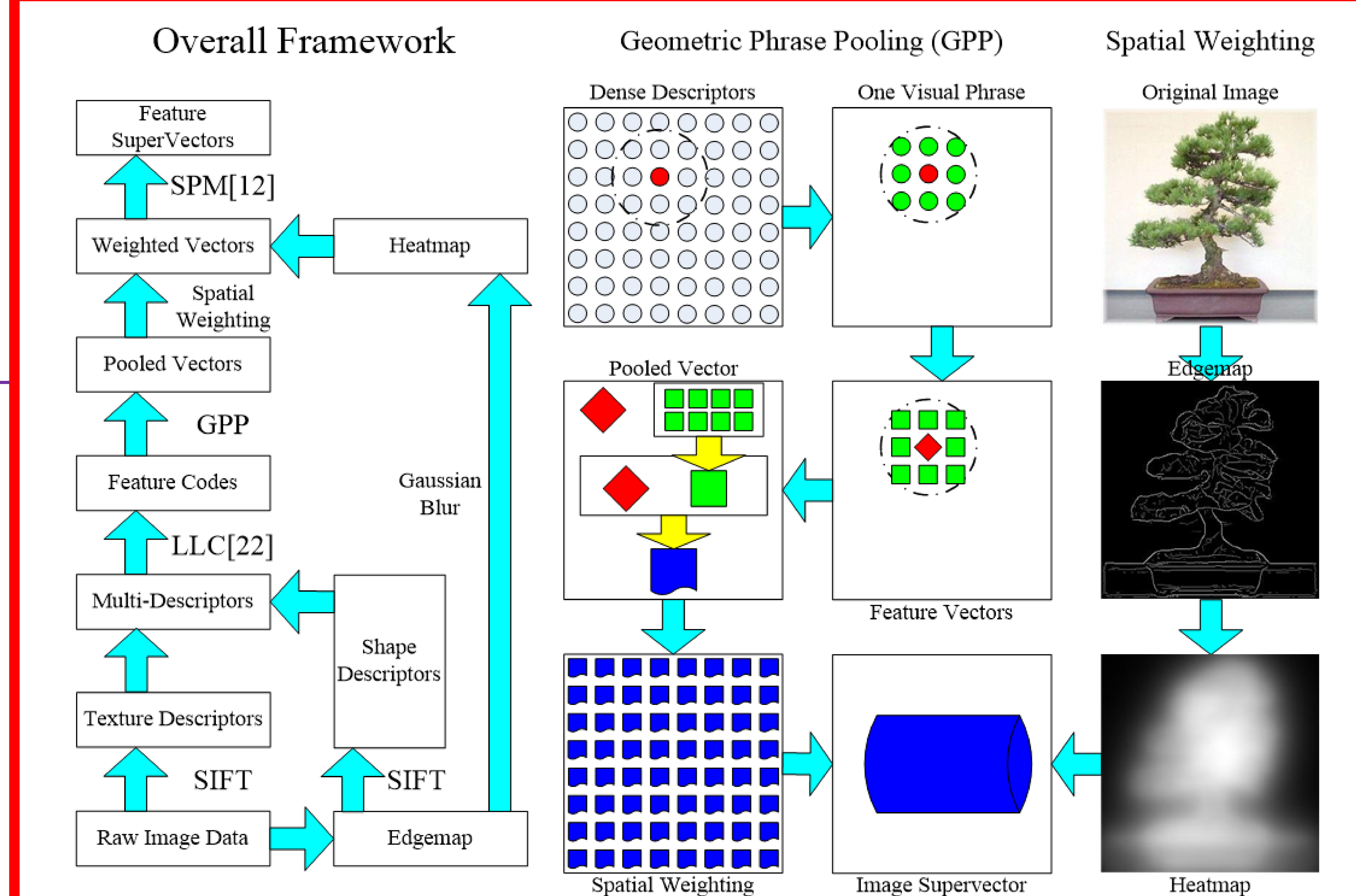


Intuition of GPP

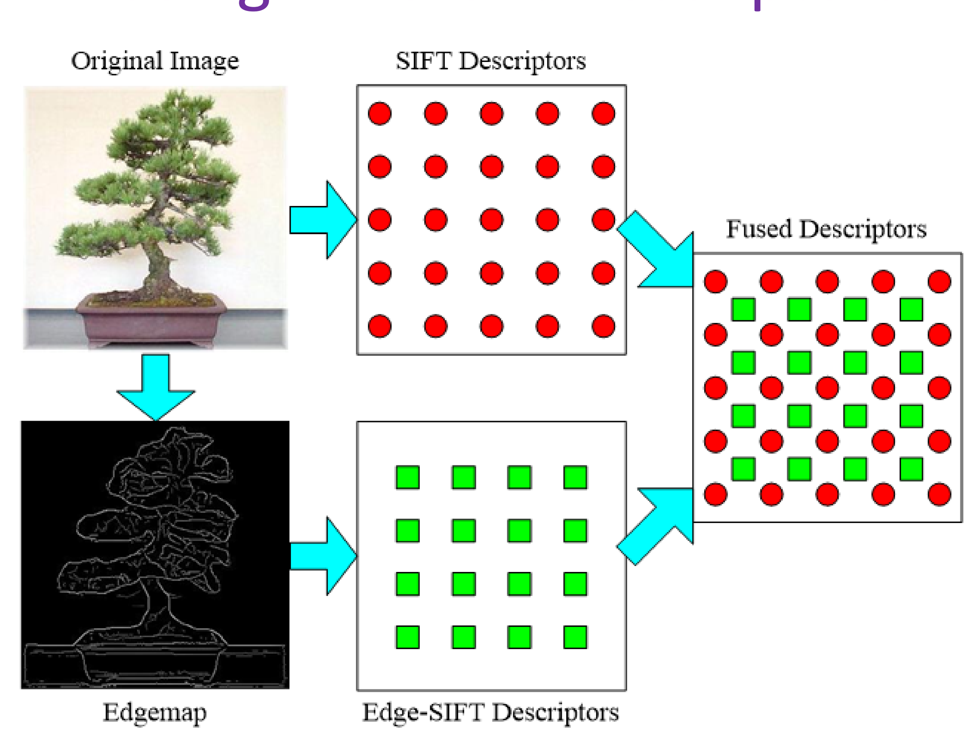


Enhancement of GPP

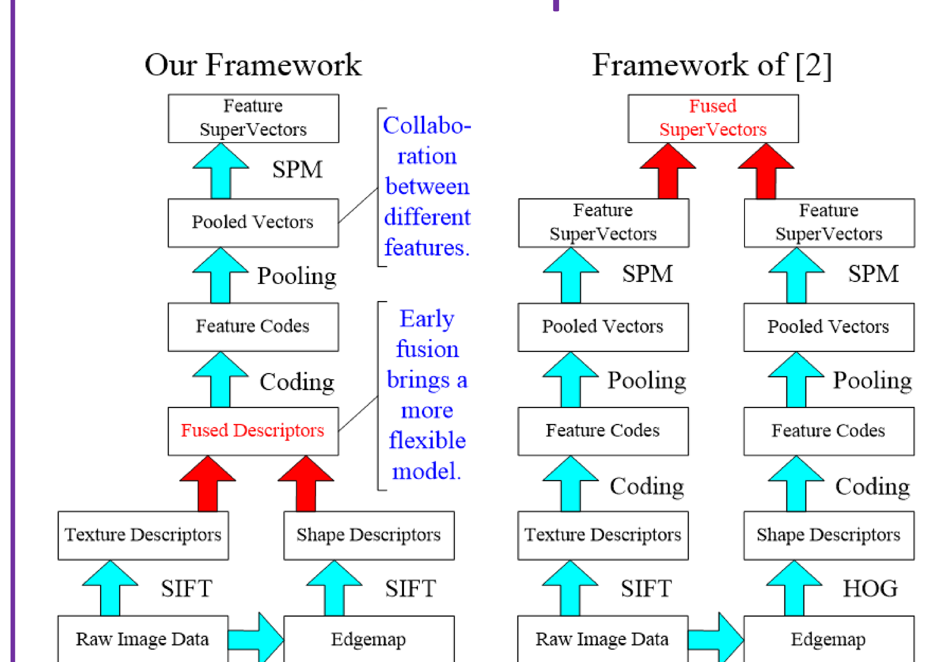




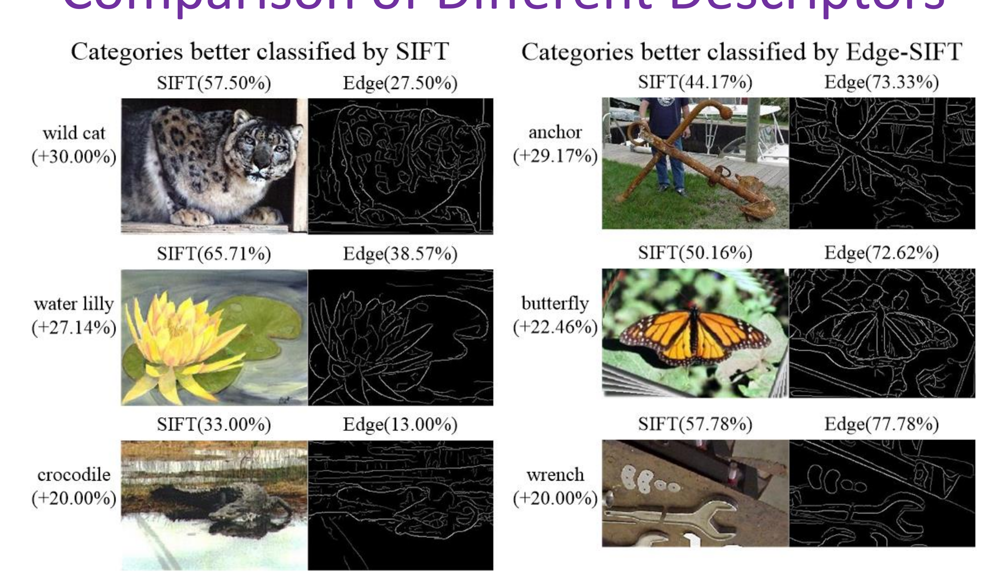
Fusing Different Descriptors



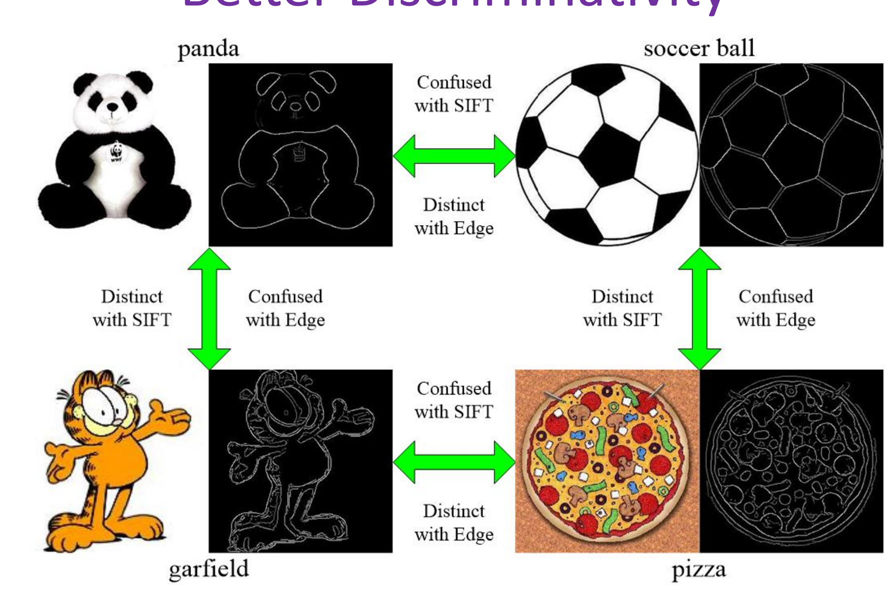
Model Comparison



Comparison of Different Descriptors



Better Discriminativity



RESULTS

Performance on Caltech101

#training	5	10	15	20	30
Lazebnik[12]			56.4		64.6
Wang[22]	51.15	59.77	65.43	67.74	73.44
Bosch[2]					81.3
Ours	61.95	71.75	76.03	78.53	82.45

Performance on Caltech256

#training	5	15	30	45	60
Wang[22]		34.36	41.19	45.31	47.68
Bosch[2]			44.0		
Ours	26.12	36.35	45.07	48.02	50.33

Conclusions

We propose a novel framework for image representation, and apply it for various image applications. By considering the key shortcomings of the BoF framework, we develop three novel modules for improvement. Integrating the new modules coherently, we obtain a very powerful model that outperforms state-of-the-art algorithms on image classifications, retrieval and understanding applications.

REFERENCES

References are numbered as they appear in the paper.
[2] A. Bosch, A. Zisserman, and X. Muoz. Image Classification using Random Forests and Ferns. International Conference on Computer Vision, pages 1 - 8, 2007.
[4] J. Canny. A Computational Approach to Edge Detection. Pattern Analysis and Machine Intelligence, (6):679 - 698, 1986.
[12] S. Lazebnik, C. Schmid, and J. Ponce. Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories. Computer Vision and Pattern Recognition, 2:2169 - 2178, 2006.
[16] D. G. Lowe. Distinctive Image Features from Scale-Invariant Keypoints. International Journal of Computer Vision, 60(2):91 - 110, 2004.
[22] J. Wang, J. Yang, K. Yu, F. Lv, T. Huang, and Y. Gong. Locality-Constrained Linear Coding for Image Classification. Computer Vision and Pattern Recognition, 3360 - 3367, 2010.

ACKNOWLEDGE.

This work was supported by the National Basic Research Program (973 Program) of China under Grant 2012CB316301, and Basic Research Foundation of Tsinghua National Laboratory for Information Science and Technology (TNList).

This work was also supported in part to Dr. Qi Tian by ARO grant W911NF-12-1-0057, NSF IIS 1052851, Faculty Research Awards by Google, NEC Laboratories of America and FXPAL, UTSA START-R award, respectively.