Semantic Context for Nonparametric Scene Parsing and Scene Classification

Gautam Singh Jana Košecká
George Mason University
Fairfax, VA
{gsinghc,kosecka}@cs.gmu.edu

Abstract

Our work focuses on different aspects of image representations as related to a variety of scene understanding tasks. We are interested in simple patch based representations as basic primitives and the role of semantic context as provided by different datasets. In our work, we have pursued a nonparametric approach for semantic parsing [5] which uses small patches and simple gradient, color and location features. We demonstrate the value of relevance of different features channels by learning a locally adaptive distance metric and the effect of feedback *in terms of semantic context, which greatly improves the performance, achieving state of the art results on different semantic parsing datasets. Here we report on an additional utility of the proposed representation for scene categorization on a subset of the scene attributes dataset introduced in [4].*

1. Introduction

With the increasing sizes of datasets and an increasing number of labels, the use of nonparametric approaches have shown notable success in semantic segmentation [7, 2]. They are appealing as they can utilize efficient approximate nearest neighbour search techniques e.g. k-d trees in nonparametric approaches like the k-nearest neighbour (k-NN) method. The issues of basic representations for semantic parsing still remain open. The state of the art approaches typically aggregate often complex image statistics over large regions yielding a rich set of features [7]. On the other hand, the recent insurgence of deep learning approaches shifts the focus from feature engineering towards unsupervised learning of basic image representations [3].

The work of [5] demonstrates a nonparametric approach for semantic parsing using small patches with simple features (gradient orientation histograms, color and location). While locally the aforementioned patches are highly ambiguous, it showed that learning the relevance of individual feature channels and the use of semantic context for nearest neighbor retrieval can achieve state of the art results on the task of semantic segmentation. In this paper, we further illustrate the utility of semantic context captured by a proposed semantic label descriptor for scene classification on a subset of the SUN-attribute dataset [4].

2. Semantic Segmentation

We formulate the semantic labelling of an image segmented into small watershed superpixels and compute gradient distribution (SIFT descriptor at superpixel centroid), color mean over pixels of the superpixel in Lab space and the location of the superpixel centroid. Initially when computing the superpixel likelihoods using k-NN method, we utilize a subset of images which are similar to the query image. We use three global image features for the dataset: (i) GIST, (ii) spatial pyramid of quantized SIFT and (iii) rgb-color histograms. Images in the training set are ranked in ascending order of individual global feature distances and the aggregate over individual feature ranks is used to select a subset of images. This subset of images serves as the source of image annotations to label the query image. It helps discard images which are dissimilar to the query image and provides a scene-level context. To compute the label likelihoods, we use a weighted k-NN method by adopting the locally adaptive metric approach of [1] for the weight computation. The initial labelling for the superpixels is obtained by inference in a Markov Random Field (MRF) where the data term is label likelihoods based on k-NN and the smoothness term combines Potts model (constant penalty) with a color difference based term. This is denoted WKNN-MRF in the experiments section.

Refined Retrieval Set The semantic labelling of an image provides a cue about the presence and absence of categories of different categories in the image. To summarize the initial semantic labeling, we aggregate the semantic information into a *semantic label descriptor* which captures semantic similarity of different images. The descriptor aggregates the evidence about presence of semantic labels in a manner similar to spatial pyramid and is described in detail in [5].

Each image of the training set is labelled by leave-one-

out-classification using the WKNN-MRF method. The resultant semantic image labelling is used to generate its corresponding semantic label descriptor. Similarily, for the query view, we generate its labelling and corresponding semantic label descriptor. We generate a ranking of the training set images based on the semantic label descriptor distance between them and the query. This ranking is combined with previously computed global image feature rankings (which used GIST, spatial pyramid over quantized SIFT and color histograms) to generate a refined retrieval set. This refined retrieval set is now used to label the query image. This method is denoted WAKNN-MRF in the experiments section. An overview of the system is provided in Figure 1.

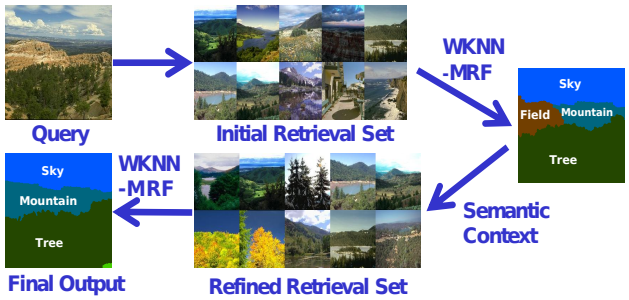

Figure 1. An overview of our semantic parsing approach.

3. Experiments

Semantic Segmentation For evaluating the performance of our semantic labelling method, experiments are performed on the SiftFlow and SUN09 datasets. The criterion for evaluation is the per pixel accuracy (percentage of pixels correctly labelled) and per class accuracy (the average of semantic category accuracies).

Dataset and System	Per-Pixel	Per-Class
SiftFlow		
Tighe et al. [7]	77.0	30.1
Eigen et al. [2]	77.1	32.5
WKNN-MRF	77.2	29.3
WAKNN-MRF	79.2	33.8
SUN09		
[6] CascALE Expert	49.3	16.7
[6] CascALE Sharing	52.8	15.2
WKNN-MRF	49.5	8.7
WAKNN-MRF	53.1	12.1

Table 1. Semantic labelling performance on SiftFlow and SUN09

Table 1 reports our performance on these datasets. On the SiftFlow dataset, our WKNN-MRF method performs on a comparable level with the other methods. After the use of semantic context to obtain a refined retrieval set, our system achieves the best performance. On the SUN09 dataset, the use of semantic context helped obtain an improvement of 3.6%. In comparison to [6], we perform better on per-pixel accuracy but trail on per-class accuracy.

Scene Classification Our experiments for using the semantic label descriptor for scene classification are performed on a subset of the SUN-Attribute dataset [4]. The attribute dataset has 717 scene categories with 20 images for each of them and we select a subset of categories from this. We choose all the scene categories whose occurrences in the SUN09 dataset exceeded 20. Using this criteria, we obtained a set of 40 scene categories. We use the WKNN-MRF method to label the images of these categories in the attribute dataset where the the training set is composed of the images corresponding to these scene categories in SUN09 dataset. We compute semantic label descriptors from the resultant labelling and use them to train scene classification SVMs. The SVMs are trained by selecting half of the 20 images for a scene category with the rest used for testing. The results are displayed in Table 2.

Descriptor	Classification Accuracy
Semantic label descriptor	30.25
Ground truth attributes	50.75
Combination	62.5

Table 2. Scene classification on subset of SUN-Attribute. The set of ground truth attributes was modified to remove all attributes which are in common with the semantic categories of SUN09.

Acknowledgements Supported by the Intelligence Advanced Research Projects Activity (IARPA) via Air Force Research Laboratory. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon. Disclaimer: The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of IARPA, AFRL, or the U.S. Government.

References

[1] C. Domeniconi, J. Peng, and D. Gunopulos. Locally adaptive metric nearest-neighbor classification. *PAMI*, 24(9):1281–1285, 2002.

[2] D. Eigen and R. Fergus. Nonparametric image parsing using adaptive neighbor sets. In *CVPR*, pages 2799–2806, 2012.

[3] C. Farabet, C. Couprie, L. Najman, and Y. LeCun. Scene parsing with multiscale feature learning, purity trees, and optimal covers. In *ICML*, pages 575–582, 2012.

[4] G. Patterson and J. Hays. SUN attribute database: Discovering, annotating, and recognizing scene attributes. In *CVPR*, pages 2751–2758, 2012.

[5] G. Singh and J. Košecká. Nonparametric scene parsing with adaptive feature relevance and semantic context. In *CVPR*, 2013.

[6] P. Sturgess, L. Ladicky, N. Crook, and P. Torr. Scalable cascade inference for semantic image segmentation. In *BMVC*, pages 62.1–62.10, 2012.

[7] J. Tighe and S. Lazebnik. Superparsing - scalable nonparametric image parsing with superpixels. *IJCV*, pages 1–21, 2012.