Dense Semantic Image Segmentation with Objects and Attributes

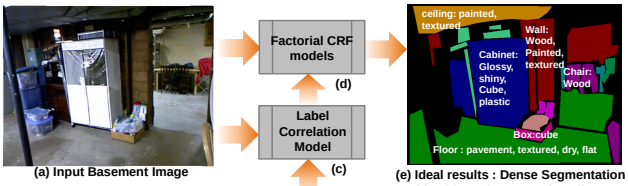
Shuai Zheng[1] Ming-Ming Cheng[1] Jonathan Warrell[2] Paul Sturgess[2]
Vibhav Vineet[2] Carsten Rother[3] Philip H. S. Torr[1]
[1]University of Oxford [2]Oxford Brookes University [3]TU Dresden
http://www.robots.ox.ac.uk/~tvg/ http://tu-dresden.de/inf/cvld

Abstract

*The concepts of objects and attributes are both important for describing images precisely, since verbal descriptions often contain both adjectives and nouns (e.g. 'I see a shiny red chair'). In this paper, we formulate the problem of joint visual attribute and object class image segmentation as a dense multi-labelling problem, where each pixel in an image can be associated with both an object-class and a set of visual attributes labels. In order to learn the label correlations, we adopt a boosting-based piecewise training approach with respect to the visual appearance and co-occurrence cues. We use a filtering-based mean-field approximation approach for efficient joint inference. Further, we develop a hierarchical model to incorporate region-level object and attribute information. Experiments on the a*PASCAL, CORE *and attribute augmented* NYU *indoor scenes datasets show that the proposed approach is able to achieve state-of-the-art results.*

1. Introduction

Using objects and attributes jointly provides a much more precise way to describe the content of a scene than using only one alone. *e.g.*, the image description *a shiny red chair* is more precise than the description *chair* on its own. Motivated by this fact, we introduce the problem of joint attribute-object image segmentation, where each image pixel is labelled with (i) an object label, such as car or road, (ii) visual attribute labels such as materials (wood, glass), and (iii) surface properties (shiny, glossy). We also make the distinction between things and stuff; where objects with a well defined shape and centroid are called things, and amorphous objects are refered to as stuff. This problem is well suited for being solved in a joint hierarchical model, as the attributes can help with the object predictions and vice versa in both region and pixel levels. In semantic image segmentation for object classes, existing approaches, *e.g.* [4, 7], treat the problem as a multi-class classification problem, where the goal is to associate each pixel with one of the object class labels. In this paper [1], we formulate the problem of joint visual attribute and object class image segmentation as a dense multi-labelling problem, where each pixel in an image can be associated with both an object-class and a set of visual attributes labels.



(a) Input Basement Image

(e) Ideal results : Dense Segmentation with Object Classes and Attributes

	Objects	Attributes
Region-Level	Things(e.g. Chair, TV monitor, Box, etc.)	Shape (e.g. Cube) , Material, Surface ...
Pixel-Level	Things and stuff (e.g. Chair, TV monitor, Box, Wall, ceiling, floor, etc.)	Material (e.g. plastic, wood), Surface (e.g. glossy), ...

(b) Semantic Space

Figure 1. **Illustration of the proposed approach.** (a) shows the input image, a scene image from NYU dataset. (b) represents the semantic label space including pixel-level objects and attributes, region-level objects and region attributes. (e) shows conceptual ideal results for dense semantic segmentation with objects and attributes. Best viewed in color.

2. Hierarchical Factorial CRF model for objects and attributes

We model scene images using a fully-connected multi-label conditional random field (CRF) with joint learning and inference. In our framework each image pixel is associated with both a set of attributes and a single object-class label, as illustrated in Fig. 1.

In order to efficiently tackle the multi-labelling problem, as shown in Fig 2, we break it down into manageable multi-class and binary subproblems using a factorial CRF framework. The structure of the factorial CRF we propose includes links between object and attribute factors that explicitly allow us to model correlations between these output variables.

[1]This is a published work [9].

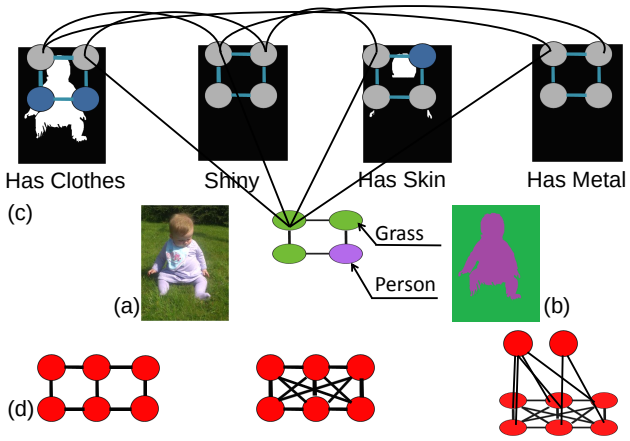


Figure 2. **Illustration of Factorial-CRF-based Semantic Segmentation for object classes and Attributes**. (a) shows the input image. (b) shows the ground truth mask image for object classes. (c) shows the attributes masks. (d) compares various CRF topologies including a grid CRF, a fully-connected CRF, and a hierarchial fully connected CRF. Best view in color.

In order to handle the use of attributes at different levels, we also propose a hierarchical model in which both objects and attributes are labelled at two levels, pixels and regions. Using the regions provided by the efficient object detector and the segmentation methods, we can predict attributes such as shape, which apply to object instances as a whole. This allows us to deal with attributes both for objects of fixed spatial extent, *i.e*. things that can be detected with deformable part based detector (*e.g*. chair, etc) as well as amorphous objects (stuff), *i.e*. ones that are more ambiguous (*e.g*. floor, etc).

To learn the correlations between factors we employ a boosting framework [6] that exploits both the visual similarity and co-occurrence relations between object and attributes labels. This provides an effective piecewise learning strategy to train the model. To perform joint inference we use a mean field based algorithm [3]. This allows us to use a fully-connected graph topology for both object and attribute factor CRFs, whilst maintaining efficiency through filtering.

3. aNYU Dataset

In this paper, we augment the annotations in NYU dataset [8]. Following the attribute annotation for scenes image [5], we added 8 additional attribute labels, *i.e.Wood, Painted, Cotton, Glass, Glossy, Plastic, Shiny, and Textured*. We asked 3 annotators to assign material, surface property attributes on each segmentation ground truth region. We then adopted the majority votes from 3 workers as our 8 additional attribute labels. We call this extended dataset the attribute NYU (aNYU) dataset. We will continue expanding the annotations and the data in the a NYU dataset.

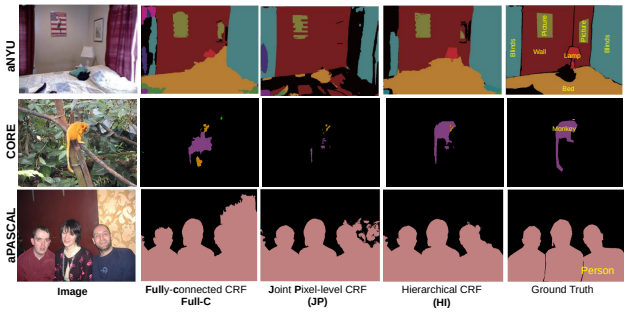


Figure 3. **Qualitative results.** Results on the aNYU, CORE [1] and aPASCAL [2] datasets. Best view in color.

4. Experiments

We evaluate our approach using three datasets: the Attribute Pascal (aPASCAL) dataset [2], the Cross-category Object REcognition (CORE) dataset [1], and the NYU indoor V2 dataset [8]. In this paper we only use the RGB images from the NYU dataset. We observe that the proposed hierarchical CRF approach achieve state-of-the-art performance in these datasets. Compared to the methods that only assign pixel with object labels, we observe significant improvements in object-based image segmentation when we use models that jointly assign attribute and object labels to pixels.

References

[1] A. Farhadi, I. Endres, and D. Hoiem. Attribute-centric recognition for cross-category generalization. In *CVPR*, 2010.

[2] A. Farhadi, I. Endres, D. Hoiem, and D. Forsyth. Describing objects by their attributes. In *CVPR*, 2009.

[3] P. Krähenbühl and V. Koltun. Efficient inference in fully connected crfs with gaussian edge potentials. In *NIPS*, 2011.

[4] L. Ladicky, C. Russell, P. Kohli, and P. H. S. Torr. Associative hierarchical crfs for object class image segmentation. In *ICCV*, 2009.

[5] G. Patterson and J. Hays. Sun attribute database: Discovering, annotating, and recognizing scene attributes. In *CVPR*, 2012.

[6] Y. Y. Sheng-Jun Huang and Z.-H. Zhou. Multi-label hypothesis reuse. In *KDD*, 2012.

[7] J. Shotton, J. Winn, C. Rother, and A. Criminisi. Textonboost for image understanding: Multiclass object recognition and segmentation by jointly modeling texture, layout, and context. *IJCV*, 81:2–23, 2009.

[8] N. Silberman, D. Hoiem, P. Kohli, and R. Fergus. Indoor segmentation and support inference from rgbd images. In *ECCV*, 2012.

[9] S. Zheng, M.-M. Cheng, J. Warrell, P. Sturgess, V. Vineet, C. Rother, and P. H. S. Torr. Dense semantic image segmentation with objects and attributes. In *CVPR*, 2014.