Constrained Parametric Min-Cuts for Automatic Object Segmentation

Joao Carreira and **Cristian Sminchisescu**
Computer Vision and Machine Learning Group, Institute for Numerical Simulation,
Faculty of Mathematics and Natural Sciences, University of Bonn
{carreira, cristian.sminchisescu}@ins.uni-bonn.de

Abstract

We present a novel framework for generating and ranking plausible objects hypotheses in an image using bottom-up processes and mid-level cues. The object hypotheses are represented as figure-ground segmentations, and are extracted automatically, without prior knowledge about properties of individual object classes, by solving a sequence of constrained parametric min-cut problems (CPMC) on a regular image grid. We then learn to rank the object hypotheses by training a continuous model to predict how plausible the segments are, given their mid-level region properties. We show that this algorithm significantly outperforms the state of the art for low-level segmentation in the VOC09 segmentation dataset. It achieves the same average best segmentation covering as the best performing technique to date [2], 0.61 when using just the top 7 ranked segments, instead of the full hierarchy in [2]. Our method achieves 0.78 average best covering using 154 segments. In a companion paper [18], we also show that the algorithm achieves state-of-the art results when used in a segmentation-based recognition pipeline.

1. Introduction

The challenge of organizing the elements of an image into plausible object regions, or segments, without knowing a-priori which objects are present in that image is one of the remarkable abilities of the human visual system, which we often take for granted. A more vivid conscious experience arises, perhaps, when observing abstract paintings. Clearly, in our perceived visual world not every hypothesis is equally likely, for example objects are usually compact, resulting in their projection in the image being connected; it is also common for strong contrast edges to mark objects boundaries.

The statistics of real-world object regions are not easy to incorporate into segmentation algorithms, making their performance sometimes brittle. One possibility would be to learn the parameters of the segmentation algorithm directly,

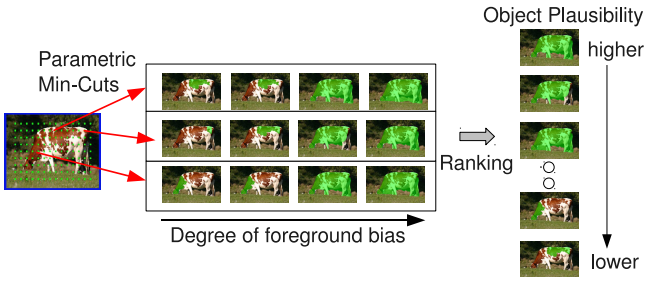

Figure 1. Our object segmentation framework. Segments are extracted around regularly placed foreground seeds, with various background seeds corresponding to image boundary edges, for all levels of foreground bias, which has the effect of producing segments with different scales. The resulting set of segments is filtered and ranked according to their plausibility of being good object hypotheses, based on mid-level properties.

by training a machine learning model using large amounts human annotated data. However, the local scope of dependencies and the intrinsically combinatorial nature of image segmentation diminishes the effectiveness of learning in such 'pixel spaces', at least with any moderately interesting model. On the other hand, once sufficient image support is available, learning to distinguish 'good' segments that represent plausible projections of real-world surfaces, from accidental image partitionings, is feasible.

A second issue is related to what can be expected from a bottom-up object hypothesis segmenter. Are segment hypotheses allowed to overlap with each other? Should one aim at multi-part image segmentations early? We argue that segmentation is already a sufficiently challenging problem without such constraints, and that global spatial consistency should be, perhaps, enforced at a later stage of processing, by higher-level routines that have better scope for this calculation. We argue that complex early multi-part consistency constraints disallow the speculative behavior necessary for segmentation algorithms, given the inherently ambiguous nature of the low-level cues they typically operate on. Hence, differently from most of the existing approaches to multiple segmentations, we derive methods to generate

several independent figure-ground partitionings, rather than a battery of segmentations of each image into multiple, non-overlapping regions.

The overall framework we pursue based on these two principles is depicted in fig. 1. We first solve a large number of independent binary problems on an image grid, at multiple scales. These are designed as energy functions efficiently solvable with parametric min-cut/max-flow techniques. The resulting pool of segments is minimally filtered to remove trivial solutions and ranked with a regressor trained to predict how object-like a segment is, based on its low and mid-level region properties.

We evaluate, objectively and empirically the quality of the list of object hypotheses returned by our algorithm, by measuring how accurate they are with respect to ground truth pixel-level annotations in object recognition datasets. We also record performance as a function of the number of segments. Results are reported on several publicly available benchmarks: MSRC [32], the Weizmann Segmentation Database [28] and VOC2009 [11] where the proposed method is shown to significantly outperform the state of the art using significantly fewer segments.

In a companion paper [18], we present state-of-the art results in object-class segmentation, detection and image categorization on several datasets. The recognition models in [18] use, as a front end, the ranked lists of object segment hypotheses produced by the algorithms proposed in this paper. Hence, the methodology is proven to be effective for both segmentation and recognition.

2. Related Work

Approaches that compute multiple segmentations can be roughly classified into hierarchical and independent *k*-partitionings. A *k*-partitioning is a division of the image into k segments, with every pixel constrained to belong to a single one. The idea of using multiple, fully independent *k*-partitionings was introduced by Hoiem *et al*. [15]. Russel *et al*. computed normalized cuts with different number of segments and image sizes [27]. By generating tens to hundreds of thousand of segments per image, Malisiewicz and Efros [21] produce very good quality segments on the MSRC dataset by merging pairs and triplets of segments obtained using the Mean Shift [9], Normalized Cuts [31] and Felzenszwalb-Huttenlocher's (FH) [12] algorithms. Stein *et al*. [1] solved Normalized Cuts with different number of segments, on a special affinity matrix derived from soft binary mattes, whereas Rabinovich *et al*. [24] select segmentations that tend to reoccur, hence are more stable.

Sharon *et al*. [30] proposed algebraic multigrid techniques for hierarchical segmentation, while Arbeláez *et al*. [2] derive an hierarchy using ultrametric contour maps. The quality of their algorithm is linked with the excellent globalPb [20] contour detector.

In interactive segmentation applications, single figure-ground segmentations are most common. They are usually computed using max-flow algorithms that find exact solutions to certain energy minimization problems [6]. In GrabCut [26], a seed is manually initialized and an observation model is iteratively fitted through expectation maximization (EM). Bagon *et al*. [3] also use EM but to estimate a sophisticated self-similarity energy. Interesting relaxation approaches also exist for energies where minimization is NP-hard [29].

Learning methods have been applied for segmentation, to distinguish real from apparent contours [13, 10, 16], similar from dissimilar superpixels [15], and good from bad regions [25]. Pen and Veksler [23] proposed, in the context of interactive segmentation, to learn to select the best segment among a small set generated for different values of one parameter.

3. Generating Multiple Figure-Ground Hypotheses

In order to generate a high-quality pool of segments, we solve multiple constrained parametric min-cuts (CPMC) with different seeds and unary terms. This gives us a large and diverse pool of segments of multiple scales. We then filter to remove high energy segments and discard all but the one with least energy, among the ones that are extremely similar. Our final, working set of segments is significantly reduced with the most accurate segments preserved.

3.1. Constrained Parametric Min-Cuts (CPMC)

For each image, alternative sets of pixels are hypothesized to belong to the foreground — the foreground seeds. Then, for each set multiple levels of foreground bias are exerted, by assigning different costs on all remaining pixels but the ones lying in negative seeds. The negative seeds are set along the image border. We compute the figure-ground segmentations resulting from minimum cuts respecting a seed hypothesis for several values of the foreground bias, including negative ones. We now formalize the optimization problem.

Let $I(\mathcal{V}) \rightarrow R^3$ be an image defined on a set of pixels $\mathcal{V}$. As commonly done in graph-based segmentation algorithms, the similarity between neighboring pixels is measured and encoded as edges of a weighted graph $G = (\mathcal{V}, \mathcal{E})$. Here, we use each pixel as a node and augment the node set $\mathcal{V}$ with two special nodes s and t that are required to be in separate partitions for any binary cut. They identify the foreground and the background, respectively. Given foreground and background seed pixels $\mathcal{V}_f$ and $\mathcal{V}_b$, our overall objective is to minimize over pixel labels $\{x_1, ..., x_k\}, x_i \in \{0, 1\}$, with k being the total number of pixels, the following energy function:

$$E^{\lambda}(X) = \sum_{u \in \mathcal{V}} D_{\lambda}(x_u) + \sum_{(u,v) \in \mathcal{E}} V_{uv}(x_u, x_v) \quad (1)$$

with $\lambda \in \mathbb{R}$, and with the unary potential function being:

$$D_{\lambda}(x_u) = \begin{cases} 0 & \text{if } x_u = 1, u \notin \mathcal{V}_b \\ \infty & \text{if } x_u = 1, u \in \mathcal{V}_b \\ \infty & \text{if } x_u = 0, u \in \mathcal{V}_f \\ f(x_u) + \lambda & \text{if } x_u = 0, u \notin \mathcal{V}_f \end{cases} \quad (2)$$

Two different functions $f(x_u)$ are used. The first one is constant and equal to zero, resulting in a uniform (variable) foreground bias. The second supplements this bias with a color term. It estimates RGB color distributions $p_f(x_u)$ on seed $\mathcal{V}_f$ and $p_b(x_u)$ on seed $\mathcal{V}_b$ and derives $f(x_u) = \ln p_f(x_u) - \ln p_b(x_u)$.

The pairwise term V_{uv} penalizes assigning different labels to similar neighboring pixels:

$$V_{uv}(x_u, x_v) = \begin{cases} 0 & \text{if } x_u = x_v \\ g(u,v) & \text{if } x_u \neq x_v \end{cases} \quad (3)$$

with the similarity between adjacent pixels given by $g(u,v) = \exp\left[-\frac{\max(gPb(u), gPb(v))}{\sigma^2}\right]$. gPb returns the output of the multi-cue contour detector globalPb [20] at a pixel. The square distance is also an option (that we experimented, with similar results), in place of the $\max$ operation.

The problem defined by eq. 1 is submodular. It can be solved exactly for each seed in the same complexity as a single max-flow problem, with a parametric solver [17]. In this work we used the pseudoflow solver [14]. The complexity of the CPMC procedure is $O(kN^2 \log(N))$ for k seeds and image graphs with N nodes.

3.2. The Grid Geometry

The foreground seeds consist of small squares and we experimented with three different strategies to place them automatically: a regular grid geometry, the centroids of large regions obtained with normalized cuts and the centroids of the superpixels, obtained using the algorithm in [12], closest to each grid position. We found that the differences were not significant, see table 1. Hence for all our experiments, we report results obtained with foreground seeds placed on a regular 5×5 grid. The background seeds are required in order to prevent trivial cuts that leave the background set empty. We used four different types: covering the full image boundary, just the vertical edges, just the horizontal edges and all but the bottom image edge. This is to allow for objects that are only partially inside the image. An alternative would be to have a single soft background seed, which we pursue as future work. 180 instances of problem (1) are solved for each image, for 30 λ values each, defined a priori on a logarithmic scale. The set of figure-ground segmentations is further enlarged by splitting the ones where the foreground consists of multiple connected components. The final pool has up to ten thousand segments.

Seed placement	MSRC score	Weizmann score
Grid	0.85 ± 0.1	0.93 ± 0.06
NCuts	0.86 ± 0.09	0.93 ± 0.07
FH	0.87 ± 0.08	0.93 ± 0.07

Table 1. Using superpixel segmentation algorithms (e.g. Normalized Cuts or FH [12]) to place the foreground seeds results in only minor improvements in the average best segmentation covering score on the MSRC dataset, when compared to placing the seeds on a regular grid. On Weizmann the average best F-measure is the same for all, probably because the objects are large and any seed placement will have some seeds inside the objects.

3.3. Fast Rejection

Generating a large segment set makes the algorithm very reliable, but many segments are redundant or not good, reflecting accidental image groupings. For images with large homogeneous regions, the method can produce many copies of the same segment instance because of the seeding strategy — every seed placed inside the region would tend to generate the same segment for the same λ. Moreover, sometimes visually arbitrary segments are created, that are artifacts of the foreground bias strength and the seed constraints employed.

We deal with these problems using a fast rejection step. We first filter out very small segments (up to 150 pixels in our implementation), then sort the segments using a simple energy (we use the ratio cut [33] as the energy because it is scale invariant and very effective) and keep the top up to 2000 segments. Then we hierarchically cluster the segments using their overlap as similarity, in order to form groups where all segments have at least 0.95 overlap. For each cluster, we keep the segment having the lowest energy.

The number of segments that pass the fast rejection step is rather small. It is indicative of how rich or cluttered an image is. In general, simple datasets have lower average number of segments. But even in the difficult PASCAL VOC dataset, the average number was 154.

4. Ranking Segments

Gestalt theorists [34, 22] argued that properties such as proximity, similarity and good continuation are key to visual grouping. With a segment set, it is interesting and useful to explore how the qualitative Gestalt theories can be implemented and how well they hold in practice. An important question is whether based on Gestalt properties, we can predict which segment is more likely to cover the object accurately, in a manner that does not require prior knowledge about the class of the object in the image. This is a challenging problem, since the visual aspects of objects

are extremely diverse. However, if this is solved, images could be represented by a handful of object-like segments, much easier to interpret and process by higher-level visual routines.

In this work, we take an empirical approach: we compile a large set of features and many annotated examples, and use machine learning tools to uncover their significance.

Three sets of features, 34 in total, are considered, related to graph, region and Gestalt properties. Graph properties, in particular variations of cut values, have long been used as cost functions in optimization formulations for segmentation. Region properties encode mainly the statistics of where and at what scale objects tend to appear in images. Finally Gestalt properties are mid-level cues related to convexity and non-accidental structure (e.g. continuity).

Graph partition properties (8 features) include the cut (sum of affinity along the segment boundary) [35], the ratio cut (sum of affinity along the boundary divided by their number) [33], the normalized cut (ratio of cut and affinity inside foreground, plus ratio of cut and affinity in background) [31], the unbalanced normalized cut (cut divided by affinity inside foreground) [30], and if the fraction of the cut larger than a threshold, normalized by segment perimeter, for 4 different thresholds.

Region properties (18 features) include area, perimeter, relative coordinates of the region centroid in the image, bounding box location and dimensions, major and minor axis lengths of the ellipse having the same normalized second central moments as the region, eccentricity, orientation, convex area, Euler number, diameter of a circle with the same area as the region, ratio of pixels in the region to pixels in the total bounding box, perimeter and absolute distance to the center of the image.

Gestalt properties (8 features) include inter-region texton similarity (chi-square distance between bag of textons computed over foreground and background), intra-region texton similarity (number of different textons present in foreground in quantity larger than a particular fraction of the area of the segment), inter-region brightness similarity (chi-square distance between intensity histograms of foreground and background regions), intra-region brightness similarity (measured in a similar way to texton homogeneity), inter-region contour energy (sum of edge energy inside foreground region, computed using globalPb, normalized by perimeter), intra-region contour energy (sum of edge energy along boundary, normalized by perimeter), curvilinear continuity (sum of consecutive angle differences of the line segment approximating the contour), convexity (ratio of area of foreground region and its convex hull).

The features were normalized by subtracting their mean and dividing by their standard deviation. Their importance, as learned by the random forests regressor, is shown in fig. 2.

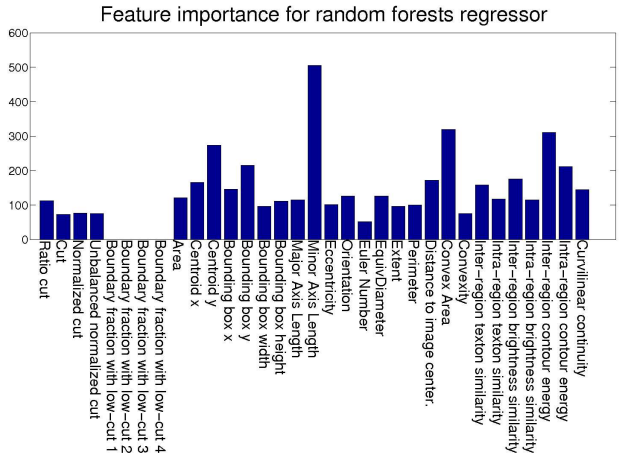

Figure 2. Feature importance for the random forests regressor learned on the VOC2009 segmentation training set. The minor axis of the ellipse having the same normalized second central moments as the segment (here "Minor Axis Length") is surprisingly the most important. This feature used in isolation results in poor rankings however (see fig. 4). The graph properties have small importance, and the "Boundary fraction of small cut" features, being binary, do not contribute at all. Gestalt features have above average importance, particularly the contour energies.

We regress on the largest **overlap** a segment has with a ground truth object from its features. definition of overlap is $O(S,G) = \frac{|S \cap G|}{|S \cup G|}$ [11]. This similarity function penalizes both under-segmentations and over-segmentations and has the advantage of being scale invariant. We used Random Forests [7], a popular non-linear model that predicts by averaging multiple regression trees.

The resulting ranking tends to put very similar segments in adjacent positions. An effective way to increase the quality of the first N segments is to **diversify** the ranking, by using Maximal Marginal Relevance (MMR) measures [8]. Starting with the originally top-scored segment, the next segment with MMR is the one maximizing the original score minus a redundancy measure. This procedure is iterated until all segments have been re-ranked. The redundancy measure employed is the overlap with the set of previous MMR segments.

Our procedure bears some similarity to the one of Ren and Malik [25], who use a random search algorithm to iteratively hypothesize segmentations by combining different superpixels, and train a classifier to distinguish good from bad segmentations. For each segment, a feature vector is extracted, a classification score is computed, and the segmentation having the highest average score is selected. Images from the Berkeley Segmentation Dataset were used, with positive examples being matched with the corresponding human segmentation, and the negative examples with a random human segmentation from a different image.

We differ from [25] in several essential aspects: we use a superset of previously proposed features, including graph and region properties or convexity, we aim at obtaining independent object-level segments, and learn directly from object class recognition datasets. We also train a regression model, not a classifier, hence we do not need to synthesize negative examples. We learn how object-like each segment is. Finally, we consider a precomputed set of figure-ground segmentations, making the process more efficient.

5. Experiments

We study both the quality of the pool of object hypotheses we generate, and the loss in quality incurred by working with the best N ranked object hypotheses (as opposed to a much larger pool).

We use three publicly available datasets: Weizmann's Segmentation Evaluation Database [28], MSRC [32] and the VOC2009 train and validation sets for the object-class segmentation problem [11].

Weizmann consists of 100 gray-valued images, with a single prominent foreground object in each. The goal is to generate a coverage of the whole object with a single segment as accurately as possible. We compare the performance of CPMC with published results for two state of the art segmentation algorithms. The results were reported using the best **F-measure** criterion, $F = \frac{2RP}{P+R}$, where P and R are the precision and recall of pixels in a segment relative to the ground truth [28]. Only the best F-measure in each image is relevant for the final score, because there is only one object in each image.

The MSRC dataset is quite different, featuring 23 different classes, including some *things*, such as water and grass. It has up to 11 objects present in each of its nearly 600 images. We use this one just to evaluate the quality of the pool of segments generated, not the individual rankings.

The VOC 2009 dataset is one of the most challenging for segmentation, featuring real-world images from Flickr, with 20 different classes of objects. In these two datasets, that have multiple ground-truth objects per image, we use the **segmentation covering** [2] as an accuracy measure. The covering of a ground truth segmentation by a machine segmentation is defined as:

$$C(S', S) = \frac{1}{N} \sum_{R \in S} |R| * \max_{R' \in S'} O(R, R') \qquad (4)$$

where N is the number of pixels in the image, $|R|$ is the number of pixels in the ground truth segment, and O is the overlap.

5.1. Segment Pool Quality

The automatic results obtained using CPMC on the Weizmann dataset are displayed in table 2 together with the previous best result, by Bagon et al [3], which requires the user to click a point inside the object. We also compare to Alpert *et al*. [28], which is also an automatic algorithm. Results for CMPC were obtained using an average of 53 segments per image. Visibly, it generates a very accurate pool of segments.

Results on MSRC and VOC2009 are compared in table 3 to Arbeláez *et al*. [2], which is arguably the state of the art method for low-level segmentation. Their methodology was followed, with the average of best coverings reported. We considered all of the unique segments in the hierarchy returned by the algorithm [2] to compute the score.

The pool of segments produced by CPMC is significantly more accurate and has an order of magnitude fewer hypotheses. A filtering procedure could be used for gPb-owt-ucm to reduce the number segments, but at the cost of quality. The dependency between the quality of segments with the size of the ground truth objects is shown in fig. 3.

Weizmann	F-measure
CPMC	0.93 ± 0.009
Bagon *et al*.	0.87 ± 0.010
Alpert *et al*.	0.86 ± 0.012

Table 2. Average of best segment F-measure scores over the entire dataset. Bagon's algorithm is interactive. Alpert's results were obtained automatically. The table shows that for each image among CPMC's pool of object hypotheses there is usually one which is extremely accurate. The average number of segments that passed the fast rejection step was 53 on this dataset.

MSRC	Covering	N Segments
CPMC	0.85 ± 0.1	57
gPb-owt-ucm	0.78 ± 0.15	670
VOC2009	**Covering**	**N Segments**
CPMC	0.78 ± 0.18	154
gPb-owt-ucm	0.61 ± 0.20	1286

Table 3. Average best image covering scores on MSRC and VOC2009 train+validation datasets, compared to Arbeláez *et al*. [2], here gPb-owt-ucm. Scores give the best covering of ground truth segments by the segments produced by the algorithm. CPMC results are before ranking.

5.2. Ranking Object Hypotheses

We evaluate the quality of our ranking method on both the validation set of the VOC2009 segmentation challenge, and on hold-out partitions of the Weizmann Segmentation Database. The training set of VOC2009 is composed of 750 images, resulting in 114000 training examples, one for each segment passing the fast rejection step. On the Weizmann Segmentation Database we select randomly 50 images, re-

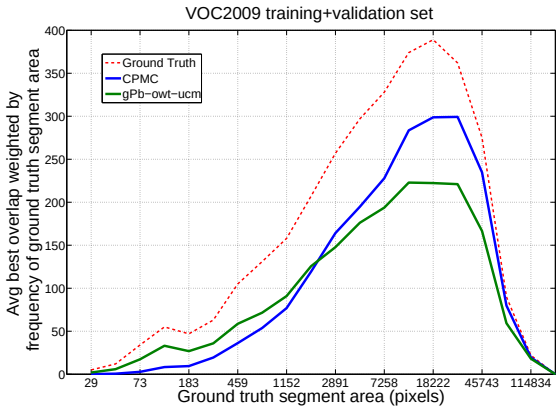

Figure 3. Quality of the segments in VOC2009 joint train and validation sets for the segmentation problem, as a function of the area of the ground truth segments. Medium and large size objects, that are more frequent, are segmented significantly more accurately by CPMC than by gPb-owt-ucm [2].

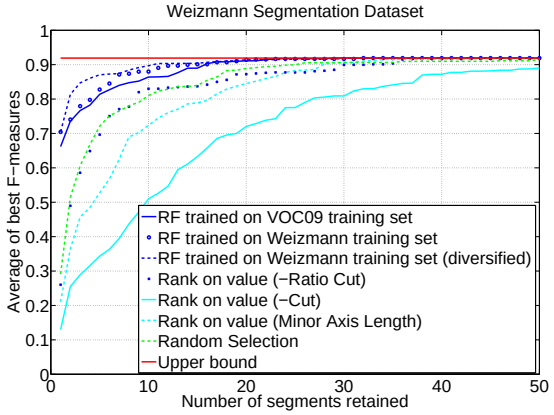

Figure 4. Average best segment F-measure as the number of retained segments from the ranking is varied. Results were averaged over three different splits of 50 training and 50 testing images. Note that when keeping 5 segments per image the results already equal the ones using the interactive method of Bagon *et al*. [3]. Note also that using this learned ranking procedure, it is possible to compress the original pool of segments to a fifth (10 segments), at negligible loss of quality.

sulting in 2500 training examples, and test on the remaining 50.

On Weizmann we compared a random forests regressor trained on Weizmann images with one trained on VOC2009. The results are similar, in fig. 4 showing that we are not overfitting to the statistics of the individual datasets. This also shows that it is possible to learn to rank the plausibility of segments containing arbitrary objects, by using object segments from only 20 classes. The learned models are significantly better than ranking using the value of any single feature such as the cut or the ratio cut.

On VOC2009 we have also run experiments where we have complemented the feature set with additional appearance and shape features — a bag of dense SIFT [19] features on the foreground mask, a bag of Local Shape Contexts [4] on its boundary, and a Pyramid HOG [5] with 3 levels — for a total of 1054 features. In this case we trained a linear regressor for ranking (this is significantly faster than random forests, that take about 8 hours to train using the basic 34 features). The results are shown in fig. 5. It is clear that the new features help somewhat, and results are slightly better than for linear regression that uses the basic feature set. However these are not better than the random forests model trained on the basic feature set. This shows that in conjunction with nonlinear models the basic features are already quite expressive. An interesting research question, here left unanswered, is whether the improvement using appearance and shape features would not be diminished as more classes are added to the 20 VOC classes. Notice that by using this ranking procedure, followed by diversification, we can obtain object hypotheses of equivalent quality to using the segmentation algorithm of [2]. In fact, by using the topmost ranked 7 segments, we obtain the same accuracy, 0.61, as

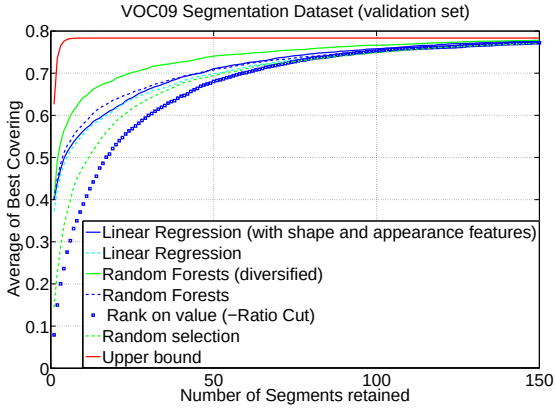

Figure 5. Complementing the basic descriptor set with appearance and shape features improves the ranking slightly, but with a more expressive regressor, random forests, the basic set is still superior. Diversifying the ranking improves the average best covering of the first top N segments significantly.

obtained using the full hierarchy of 1286 distinct segments in [2].

6. Conclusion

We have presented an algorithm that casts the automatic image segmentation problem as one of finding a set of plausible figure-ground object hypotheses. It does so by learning to rank figure-ground segmentations, using ground truth annotations available in object class recognition datasets and based on a set of low and mid-level properties. The

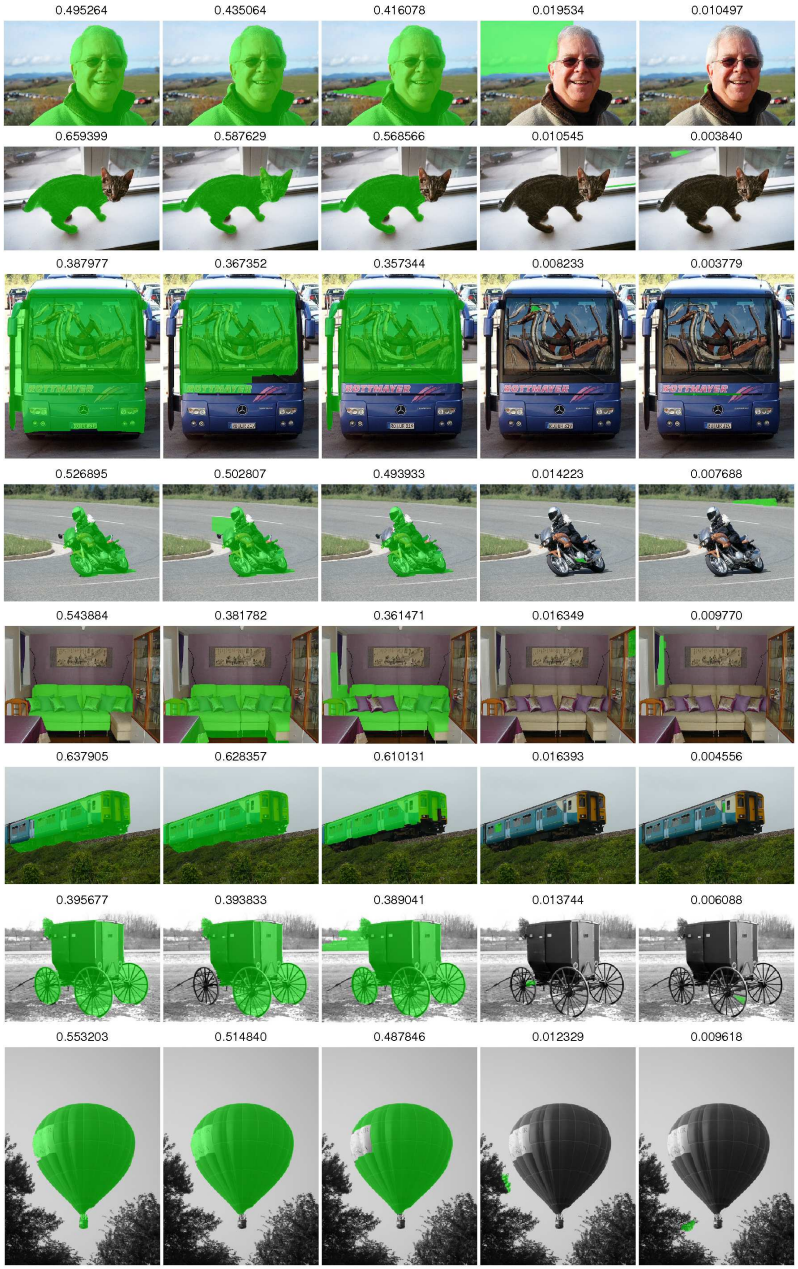

Figure 6. Ranking results obtained using the random forests model learned on the VOC2009 training set. The first three images on each row show the most plausible object hypotheses, the last two images show the least plausible segments. The segment scores are shown above each image. The green regions correspond to the foreground areas of the object hypotheses. The first six rows of images are from the VOC2009 segmentation validation set, the last two are from the Weizmann Segmentation Database. The lowest ranked object hypotheses are usually very small reflecting perhaps the statistics of the images in the VOC2009 training set. The algorithm shows a remarkable preference to segments with large overlap with the objects in the image, although it has no prior knowledge about their classes. There are neither chariots nor balloons in the training set, for example.

algorithm uses a very powerful new procedure to generate a pool of figure-ground segmentations, the Constrained Parametric Min-Cuts. This uses parametric max-flow to efficiently compute figure-ground hypotheses at multiple scales on an image grid. We have shown that the proposed framework is able to generate small sets of segments that represent the objects in an image more accurately than existing state of the art segmentation methods. These sets of segments have been used successfully in a segmentation-based recognition framework [18].

Acknowledgements

The authors want to thank David Fleet for valuable comments. This work was supported in part by the European Commission, under a Marie Curie Excellence Grant MCEXT-025481.

References

[1] T. S. A. Stein and M. Hebert. Towards unsupervised whole-object segmentation: Combining automated matting with boundary detection. In *CVPR*, June 2008. 2

[2] P. Arbelaez, M. Maire, C. Fowlkes, and J. Malik. From contours to regions: An empirical evaluation. *In CVPR*, 2009. 1, 2, 5, 6

[3] S. Bagon, O. Boiman, and M. Irani. What is a good image segment? a unified approach to segment extraction. *In ECCV*, pages 30–44, 2008. 2, 5, 6

[4] S. Belongie, J. Malik, and J. Puzicha. Shape context: A new descriptor for shape matching and object recognition. In *NIPS*, pages 831–837, 2000. 6

[5] A. Bosch, A. Zisserman, and X. Munoz. Representing shape with a spatial pyramid kernel. *CIVR*, pages 401–408, 2007. 6

[6] Y. Boykov and G. Funka-Lea. Graph cuts and efficient n-d image segmentation. *IJCV*, 70(2):109–131, 2006. 2

[7] L. Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001. 4

[8] J. Carbonell and J. Goldstein. The use of mmr, diversity-based reranking for reordering documents and producing summaries. *In RDIR*, pages 335–336, 1998. 4

[9] D. Comaniciu and P. Meer. Mean shift: A robust approach toward feature space analysis. *PAMI*, 24(5):603–619, 2002. 2

[10] P. Dollár, Z. Tu, and S. Belongie. Supervised learning of edges and object boundaries. In *CVPR*, June 2006. 2

[11] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The PASCAL Visual Object Classes Challenge 2009 (VOC2009) Results. http://www.pascal-network.org/challenges/VOC/voc2009/workshop/index.html. 2, 4, 5

[12] P. F. Felzenszwalb and D. P. Huttenlocher. Efficient graph-based image segmentation. *IJCV*, 59(2):167–181, September 2004. 2, 3

[13] C. Fowlkes, D. Martin, and J. Malik. Learning affinity functions for image segmentation: combining patch-based and gradient-based approaches. In *In CVPR*, volume 2, pages II–54–61 vol.2, June 2003. 2

[14] D. Hochbaum. The pseudoflow algorithm and the pseudoflow-based simplex for the maximum flow problem. In *IPCO*, pages 325–337, 1998. 3

[15] D. Hoiem, A. A. Efros, and M. Hebert. Geometric context from a single image. *In ICCV*, 1:654 – 661, October 2005. 2

[16] J. Kaufhold and A. Hoogs. Learning to segment images using region-based perceptual features. *In CVPR*, 2:954–961, 2004. 2

[17] V. Kolmogorov, Y. Boykov, and C. Rother. Applications of parametric maxflow in computer vision. *In ICCV*, pages 1–8, Oct. 2007. 3

[18] F. Li, J. Carreira, and C. Sminchisescu. Object Recognition as Ranking Holistic Figure-Ground Hypotheses. In *CVPR*, June 2010. 1, 2, 7

[19] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 60(2):91–110, 2004. 6

[20] M. Maire, P. Arbelaez, C. Fowlkes, and J. Malik. Using contours to detect and localize junctions in natural images. *In CVPR*, 0:1–8, 2008. 2, 3

[21] T. Malisiewicz and A. Efros. Improving spatial support for objects via multiple segmentations. *In BMVC*, September 2007. 2

[22] S. E. Palmer. *Vision Science: Photons to Phenomenology*. The MIT Press, 1999. 3

[23] B. Peng and O. Veksler. Parameter Selection for Graph Cut Based Image Segmentation. *BMVC*, 2008. 2

[24] A. Rabinovich, S. Belongie, T. Lange, and J. M. Buhmann. Model order selection and cue combination for image segmentation. *In CVPR*, 1:1130–1137, 2006. 2

[25] X. Ren and J. Malik. Learning a classification model for segmentation. *ICCV*, 1:10, 2003. 2, 4, 5

[26] C. Rother, V. Kolmogorov, and A. Blake. "grabcut": interactive foreground extraction using iterated graph cuts. *ACM Trans. Graph.*, 23(3):309–314, 2004. 2

[27] B. Russell, W. Freeman, A. Efros, J. Sivic, and A. Zisserman. Using multiple segmentations to discover objects and their extent in image collections. *In CVPR*, 2:1605–1614, 2006. 2

[28] R. B. S. Alpert, M. Galun and A. Brandt. Image segmentation by probabilistic bottom-up aggregation and cue integration. *In CVPR*, 0:1–8, 2007. 2, 5

[29] T. Schoenemann, F. Kahl, and D. Cremers. Curvature regularity for region-based image segmentation and inpainting: A linear programming relaxation. *ICCV*, 2009. 2

[30] E. Sharon, M. Galun, D. Sharon, R. Basri, and A. Brandt. Hierarchy and adaptivity in segmenting visual scenes. *Nature*, 442(7104):719–846, June 2006. 2, 4

[31] J. Shi and J. Malik. Normalized cuts and image segmentation. *PAMI*, 2000. 2, 4

[32] J. Shotton, J. Winn, C. Rother, and A. Criminisi. Textonboost: Joint appearance, shape and context modeling for multi-class object recognition and segmentation. In *ECCV*, pages 1–15, 2006. 2, 5

[33] S. Wang and J. M. Siskind. Image segmentation with ratio cut. *PAMI*, 25:675–690, 2003. 3, 4

[34] M. Wertheimer. Laws of organization in perceptual forms (partial translation). In *A sourcebook of Gestalt Psychology*, pages 71–88. 3

[35] Z. Wu and R. Leahy. An optimal graph theoretic approach to data clustering: Theory and its application to image segmentation. *PAMI*, 15(11):1101–1113, 1993. 4