PatchCut: Data-Driven Object Segmentation via Local Shape Transfer

Jimei Yang, Brian Price, Scott Cohen, Zhe Lin, and Ming-Hsuan Yang

Tayfun Ateş

Burak Ercan

Contents

Problem Statement

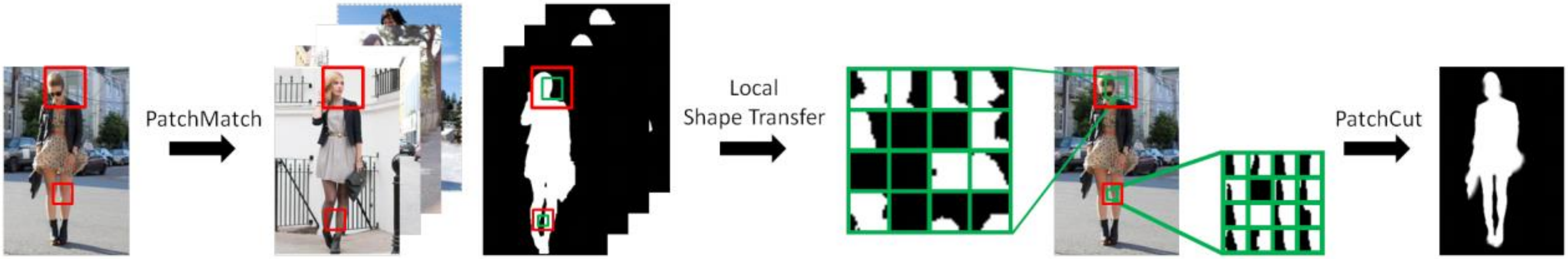
- Object segmentation is the task of separating a foreground object from its background

Image Object

Motivation

- Provides mid-level representations for high-level recognition tasks
  - Object recognition
  - Image classification
  - Semantic segmentation
  - Image captioning
- Has immediate applications to image and video editing
  - Adobe Photoshop and After Effects

Method Overview



- Object segmentation using examples
- Multiscale image matching in patches by PatchMatch
- Patch-wise segmentation candidates
- An algorithm based on higher order MRF energy function to produce the segmentation
- Coarse-to-fine approach

Main Contributions (1/2)

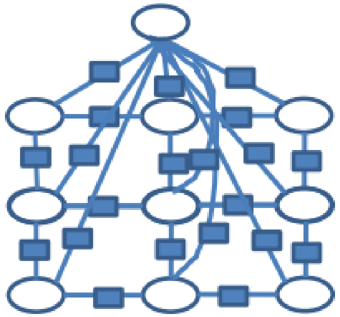
- A novel nonparametric high-order MRF model via patch-level label transfer for object segmentation
- An efficient iterative algorithm (PatchCut) that solves the proposed MRF energy function in patch-level without using graph cuts
- State-of-the-art performance on various object segmentation benchmark datasets

Main Contributions (2/2)

- Incorporating object shape information for segmentation
- No offline training
- No user interaction
- No prior knowledge on category specific object models
- Patch level local shape transfer scheme

Related Work (MRF)

- Binary labeling on Markov Random Fields (MRFs) with foreground/background appearance models:
  - **Y. Y. Boykov and M.-P. Jolly. Interactive graph cuts for optimal boundary & region segmentation of objects in n-d images. In *ICCV*, 2001.**



MRF with
global variables

$$E(x) = \sum_{i,j \in N_8} \theta_{ij}(x_i, x_j)$$

Related Work (Interactive Methods)

- Requires user input
- Color or texture cues to improve segmentation performance
  - Y. Y. Boykov and M.-P. Jolly. Interactive graph cuts for optimal boundary & region segmentation of objects in n-d images. In *ICCV*, 2001.
  - V. Lempitsky, P. Kohli, C. Rother, and T. Sharp. Image segmentation with a bounding box prior. In *ICCV*, 2009.
  - C. Rother, V. Kolmogorov, and A. Blake. Grabcut - interactive foreground extraction using iterated graph cuts. *ACM Transactions on Graphics (SIGGRAPH)*, 2004.
  - J. Wu, Y. Zhao, J.-Y. Zhu, S. Luo, and Z. Tu. Milcut: A sweeping line multiple instance learning paradigm for interactive image segmentation. In *CVPR*, 2014.

- **Incorporating object shape information for segmentation**
- **No offline training**
- **No user interaction**
- **No prior knowledge on category specific object models**
- **Patch level local shape transfer scheme**

Related Work (Salient Object Segmentation)

- Segmenting object(s) that grab(s) our attention most
- Requires high contrast
  - **F. Perazzi, P. Krahenb ¨ uhl, Y. Pritch, and A. Hornung. ¨ Saliency filters: Contrast based filtering for salient region detection. In *CVPR*, 2012.**
  - **R. Margolin, A. Tal, and L. Zelnik-Manor. What makes a patch distinct? In *CVPR*, 2013.**
  - **M.-M. Cheng, N. J. Mitra, X. Huang, P. H. S. Torr, and S.- M. Hu. Global contrast based salient region detection. *PAMI*, 2014.**

- **Incorporating object shape information for segmentation**
- **No offline training**
- **No user interaction**
- **No prior knowledge on category specific object models**
- **Patch level local shape transfer scheme**

Related Work (Model Based Algorithms)

- Offline learning based methods
  - **E. Borenstein and S. Ullman. Class-specific, top-down segmentation. In *ECCV*, 2002**
  - **D. Larlus and F. Jurie. Combining appearance models and markov random fields for category level object segmentation.
In *CVPR*, 2008.**
  - **M. P. Kumar, P. Torr, and A. Zisserman. Obj cut. In *CVPR*, 2005**
  - **L. Bertelli, T. Yu, D. Vu, and B. Gokturk. Kernelized structural svm learning for supervised object segmentation. In *CVPR*, 2011.**
  - **J. Yang, S. Safar, and M.-H. Yang. Max-margin Boltzmann machines for object segmentation. In *CVPR*, 2014.**

- **Incorporating object shape information for segmentation**
- **No offline training**
- **No user interaction**
- **No prior knowledge on category specific object models**
- **Patch level local shape transfer scheme**

Related Work (Data Driven Methods)

- <u>Global</u> shape transfer without online learning
- Image match by either window based or local feature based
- Less time efficient
  - **D. Kuettel and V. Ferrari. Figure-ground segmentation by transferring window masks. In *CVPR*, 2012.**
  - **E. Ahmed, S. Cohen, and B. Price. Semantic object selection. In *CVPR*, 2014.**
  - **J. Kim and K. Grauman. Shape sharing for object segmentation. In *ECCV*, 2012.**
  - **J. Tighe and S. Lazebnik. Finding things: Image parsing with regions and per-exemplar detectors. In *CVPR*, 2013.**

- **Incorporating object shape information for segmentation**
- **No offline training**
- **No user interaction**
- **No prior knowledge on category specific object models**
- **Patch level local shape transfer scheme**

Related Work (Structured Label Space)

- Forest based image labeling algorithms
- Each leaf node stores one example label patch
- These trained forests are used for
  - Edge Detection
  - Semantic Labeling
- **P. Kontschieder, S. R. Bulo, H. Bischof, and M. Pelillo. Structured class-labels in random forests for semantic image labelling. In *ICCV*, 2011.**
- **P. Dollar and C. Zitnick. Structured forests for fast edge detection. In *ICCV*, 2013.**

Revisiting Main Contributions

- **Incorporating object shape information for segmentation**
- **No offline training**
- **No user interaction**
- **No prior knowledge on category specific object models**
- **Patch level local shape transfer scheme**

Proposed Method

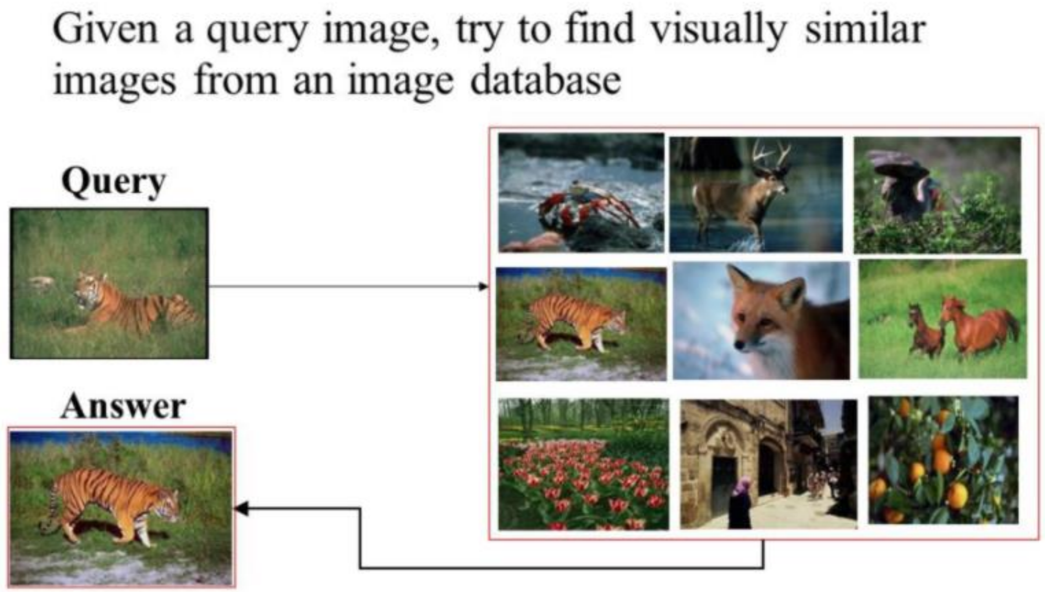
- A data driven approach

Proposed Method

- A data driven approach
- What is meant by being data driven?
  How the proposed method uses data?

Proposed Method

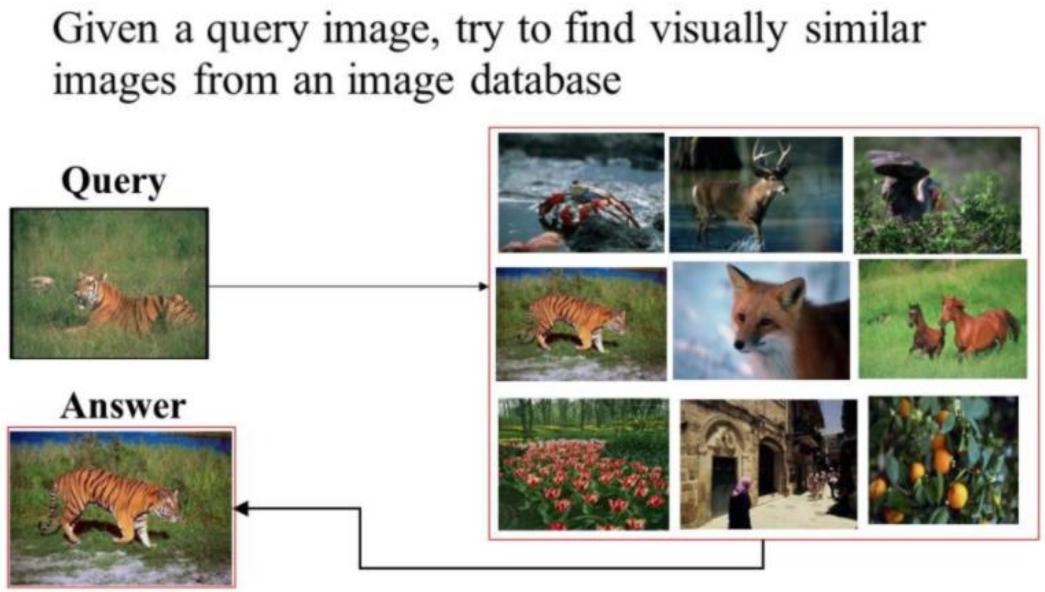
- A data driven approach
- What is meant by being data driven? How the proposed method uses data?
- For a single query image, it finds most similar M images (M is fixed as 16) with their segmentation masks and uses this information to create better segmentation results by proposing a multiscale patch based method.



From: svcl.ucsd.edu

Proposed Method

- A data driven approach
- What is meant by being data driven? How the proposed method uses data?
- For a single query image, it finds most similar M images (M is fixed as 16) with their segmentation masks and uses this information to create better segmentation results by proposing a multiscale patch based method.
- Image retrieval is done representing the query and dataset images either by using features from Bag-Of-Words, or 7th layer of convolutional networks (ConvNet)* trained with ImageNet.



From: svcl.ucsd.edu

*Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell. Caffe: Convolutional architecture for fast feature embedding. arXiv preprint arXiv:1408.5093, 2014.

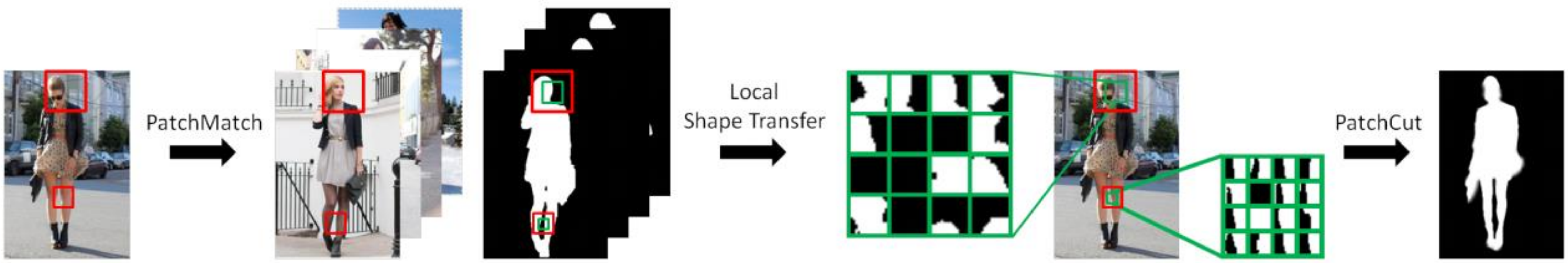
Proposed Method

$\mathbf{I}$ Test image

$\hat{\mathbf{Y}}$ Segmentation of the test image (we want to estimate this)

$\{\mathbf{I}_m, m = 1, 2, ..., M\}$ Example images (retrieved from the database)

$\{\mathbf{Y}_m, m = 1, 2, ..., M\}$ Segmentation ground truths of example images



Local Shape Transfer

$\{\mathbf{I}^s, s = 1, 2, 3\}$ Downsampled versions of the test image, with scale s

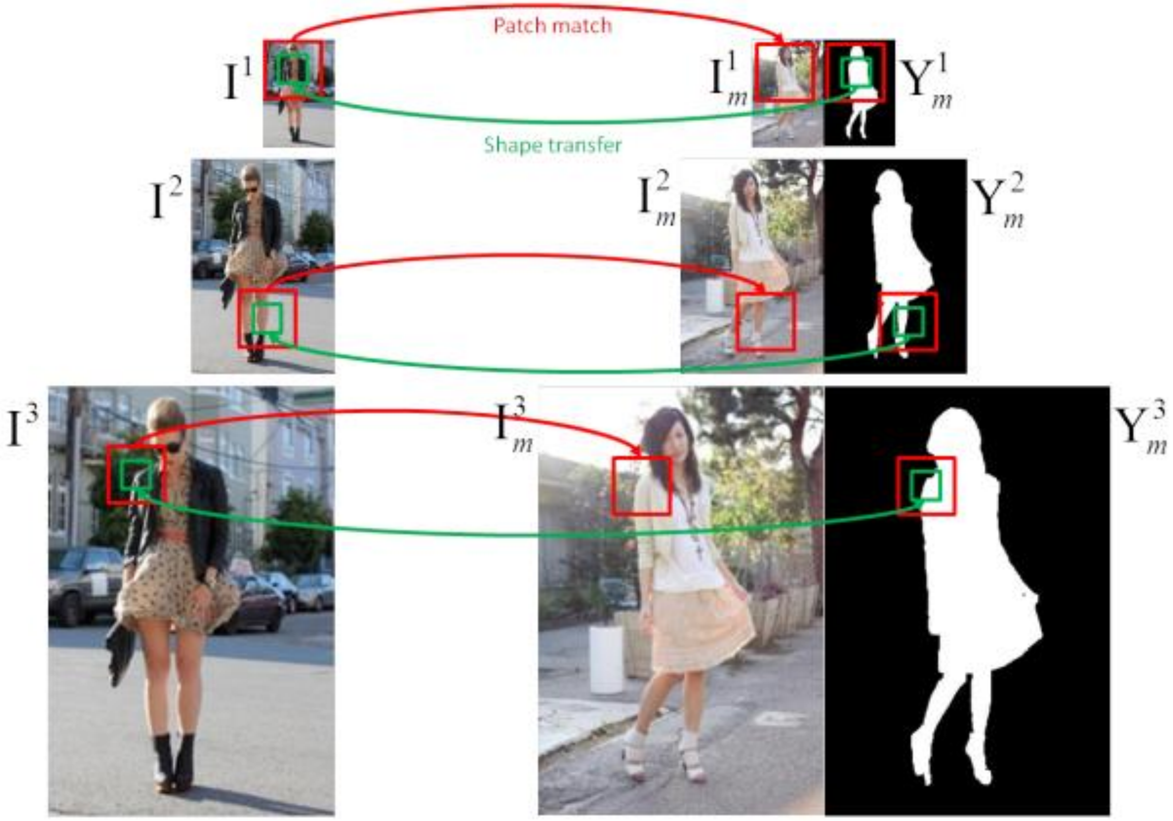
$\{\mathbf{I}^s_m, \mathbf{Y}^s_m, s = 1, 2, 3\}$ Downsampled versions of examples and their segmentations

$[h, w]$ Size of the original image

$[\frac{h}{2^{3-s}}, \frac{w}{2^{3-s}}]$ Sizes of the downsampled images

$\{\Delta^s_k, k = 1, 2, ..., K\}$ K number of 16x16 patches for scale s

How to Find Matches for a Patch?



$\mathbf{x}_k^s$ SIFT descriptor of 32x32 patches

Solve the matching problem:

$$\arg\min_{k'} \|\mathbf{x}_k^s - \mathbf{x}_{k'm}^s\|_1, \forall k = 1, 2, ..., K$$

PatchMatch efficiently solves this!

$\Delta_{k^*}^s$ Match of kth patch patch in mth example

$d_{km}^s = \|\mathbf{x}_k^s - \mathbf{x}_{k^*m}^s\|_1$ Cost of this match

Patch Match

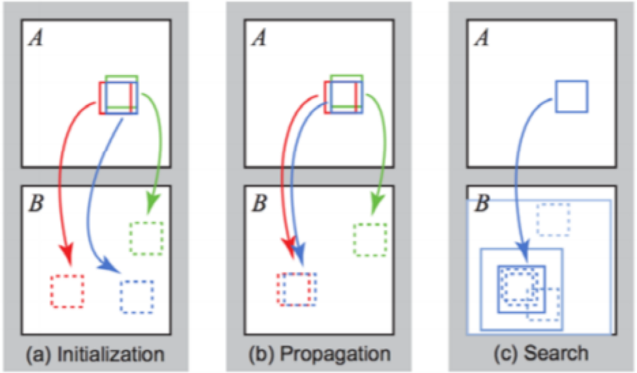
Algorithm

1. Initialize pixels with random patch offsets

2. Check if neighbors have better patch offsets

3. Search in concentric radius around the current offset for better better patch offsets

4. Go to Step 2 until converge.



O(mMlogM)

From: vis.berkeley.edu/courses/cs294-69-fa11

Solution Space for the Test Image

$\mathbf{z}^s_{km} = \mathbf{Y}^s_m(\Delta^s_{k^*})$ Local segmentation masks from the matched patches in *m*th example

Authors assume that:

- These masks constitute a patch-wise segmentation solution space for the test image
- The segmentation mask of test image can be well approximated by these masks

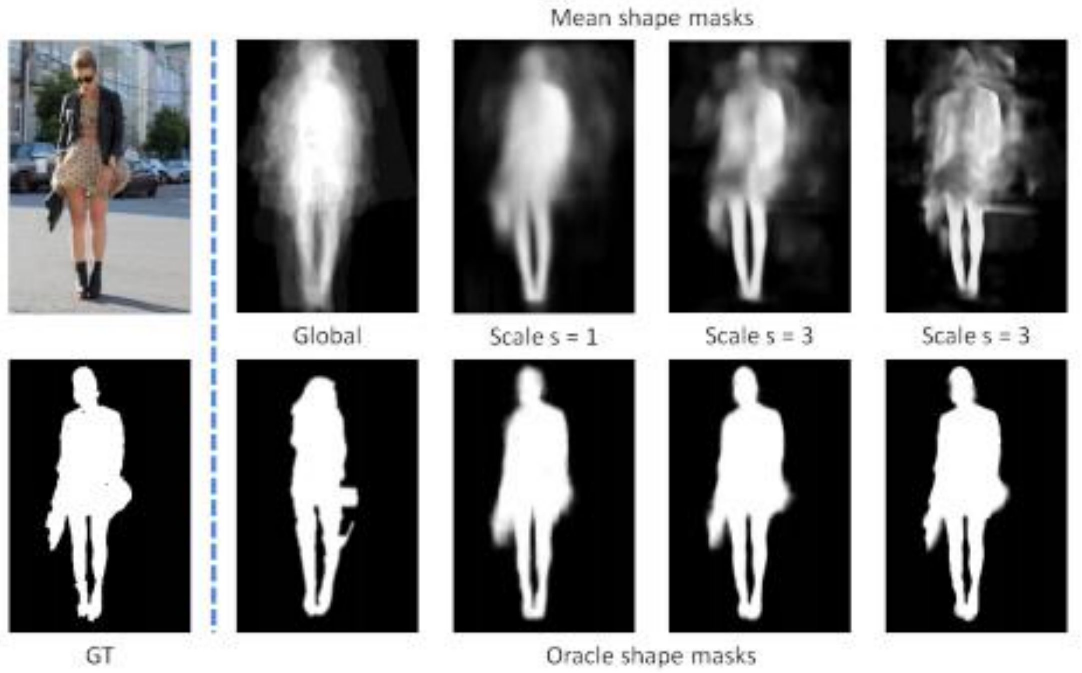
How can we validate this assumption?

Validation of the Assumption

$\bar{\mathbf{z}}_k^s = \frac{1}{M}\sum_m \mathbf{z}_{km}^s$ Let's calculate the mean of local masks over M example images

$\bar{\mathbf{Q}}^s$ Mean shape prior mask can then be calculated by adding up $\bar{\mathbf{z}}_k^s$

Also find the oracle shape prior mask $\tilde{\mathbf{Q}}^s$ from the best possible $\tilde{\mathbf{z}}_k^s$ (by using the ground truth as reference)



- Object is well located in the coarsest scale, but blurry
- In the finest scale, masks can become noisy
    - Background near the legs is mostly uniform
    - Background near the upper body is cluttered

- A coarse-to-fine strategy can be employed

PatchCut (Some Preliminaries)

$$E(\mathbf{Y}) = \sum_{i\in\mathcal{V}} U(y_i) + \gamma \sum_{i,j\in\mathcal{E}} V(y_i, y_j) + \lambda \sum_{i\in\mathcal{V}} S(y_i, q_i)$$

The energy function: Segmentation problem is solved by minimizing this function

$$U(y_i) = -\log P(y_i|\mathbf{c}_i, \mathbf{A}_1, \mathbf{A}_0)$$

The unary term: Negative log probability of the label y_i given the pixel color $\mathbf{c}_i$ and Gaussian Mixture Models (GMMs) $\mathbf{A}_1$ and $\mathbf{A}_0$ for foreground and background color

$$V(y_i, y_j) = \exp(-\alpha\|\mathbf{c}_i - \mathbf{c}_j\|^2)\mathbb{I}(y_i \neq y_j)$$

The pairwise term: Measures the cost of assigning different labels to two adjacent pixels (based on their color difference)

$$S(y_i, q_i) = -\log q_i^{y_i}(1-q_i)^{1-y_i}$$

The shape term: Measures the inconsistency with shape prior **Q**

This energy function can be minimized with alternating two steps similar to GrabCut:

1) $\{\mathbf{A}_1, \mathbf{A}_0\} \leftarrow \mathbf{Y}$
2) $\mathbf{Y} \leftarrow \{\mathbf{A}_1, \mathbf{A}_0\}$

High order MRF with Local Shape Transfer (1/2)

$$E'(\mathbf{Y}) = E(\mathbf{Y}) - \sum_k \log(P_{\text{cand}}(\mathbf{Y}(\Delta_k))) \longrightarrow$$

The modified energy function. The last term is the negative Expected Patch Log Likelihood (EPLL).

$$P_{\text{cand}}(\mathbf{Y}(\Delta_k)) \longrightarrow$$

Patch likelihood (this encourages the label patch for a patch in our test image to be similar to some candidate local shape mask)

$$\begin{aligned}
P_{\text{cand}}(\mathbf{Y}(\Delta_k)) &= \sum_{m=1}^{M} P(\mathbf{Y}(\Delta_k), m_k^* = m) \\
&= \sum_{m=1}^{M} P(\mathbf{Y}(\Delta_k)|m_k^* = m)P(m_k^* = m) \\
&= \sum_{m=1}^{M} \frac{\exp(-\eta||\mathbf{Y}(\Delta_k) - \mathbf{z}_{km}||_2^2)}{Z_1} \frac{\exp(-\tau d_{km})}{Z_2}
\end{aligned}$$

Assume η is large to encourage the output label patches to be as similar to the selected candidate patches as possible.

High order MRF with Local Shape Transfer (2/2)

For large η :

$$P_{\text{cand}}(\mathbf{Y}(\Delta_k)) \approx \begin{cases} \exp(-\tau d_{km})/Z_2 & \text{if } \mathbf{Y}(\Delta_k) = \mathbf{z}_{km} \\ 0 & \text{otherwise} \end{cases}$$

$$H(\mathbf{Y}(\Delta_k)) = \begin{cases} d_{km} & \text{if } \mathbf{Y}(\Delta_k) = \mathbf{z}_{km} \\ \infty & \text{otherwise} \end{cases}$$

$$E'(\mathbf{Y}) \approx E(\mathbf{Y}) + \tau \sum_k H(\mathbf{Y}(\Delta_k))$$

Is there a solution for this problem?

Approximate Optimization on Patches

$$E'(\mathbf{Y}) \approx E(\mathbf{Y}) + \tau \sum_k H(\mathbf{Y}(\Delta_k))$$

$$H(\mathbf{Y}(\Delta_k)) = \begin{cases} d_{km} & \text{if } \mathbf{Y}(\Delta_k) = \mathbf{z}_{km} \\ \infty & \text{otherwise} \end{cases}$$

The solution to this energy function do not exist when selected label patches disagree in any overlapping areas !

$$E'(\mathbf{Y}, \{\mathbf{z}_k\}) \approx E(\mathbf{Y}) + \tau \sum_k H(\mathbf{z}_k), \text{s.t. } \mathbf{Y}(\Delta_k) = \mathbf{z}_k$$

$\mathbf{z}_k$ denotes the selected label patch on kth patch

$$\mathbf{z}_k \in \{\mathbf{z}_{k1}, \mathbf{z}_{k2}, \ldots, \mathbf{z}_{kM}\}$$

$$E'(\mathbf{Y}, \{\mathbf{z}_k\}) \approx \kappa \sum_k E(\mathbf{z}_k) + \tau \sum_k H(\mathbf{z}_k), \text{s.t. } \mathbf{Y}(\Delta_k) = \mathbf{z}_k$$

Convert the constrained optimization problem to an unconstrained one by introducing a quadratic penalty on each patch.

$$E'(\mathbf{Y}, \{\mathbf{z}_k\}) \approx \sum_k (\kappa E(\mathbf{z}_k) + \tau H(\mathbf{z}_k) + \frac{\beta}{2} \|\mathbf{Y}(\Delta_k) - \mathbf{z}_k\|^2)$$

Choose β sufficiently large !

The Single Scale PatchCut Algorithm

$$\hat{\mathbf{z}}_k = \arg\min_{\mathbf{z}_k} \kappa E(\mathbf{z}_k) + \tau H(\mathbf{z}_k), \forall k$$

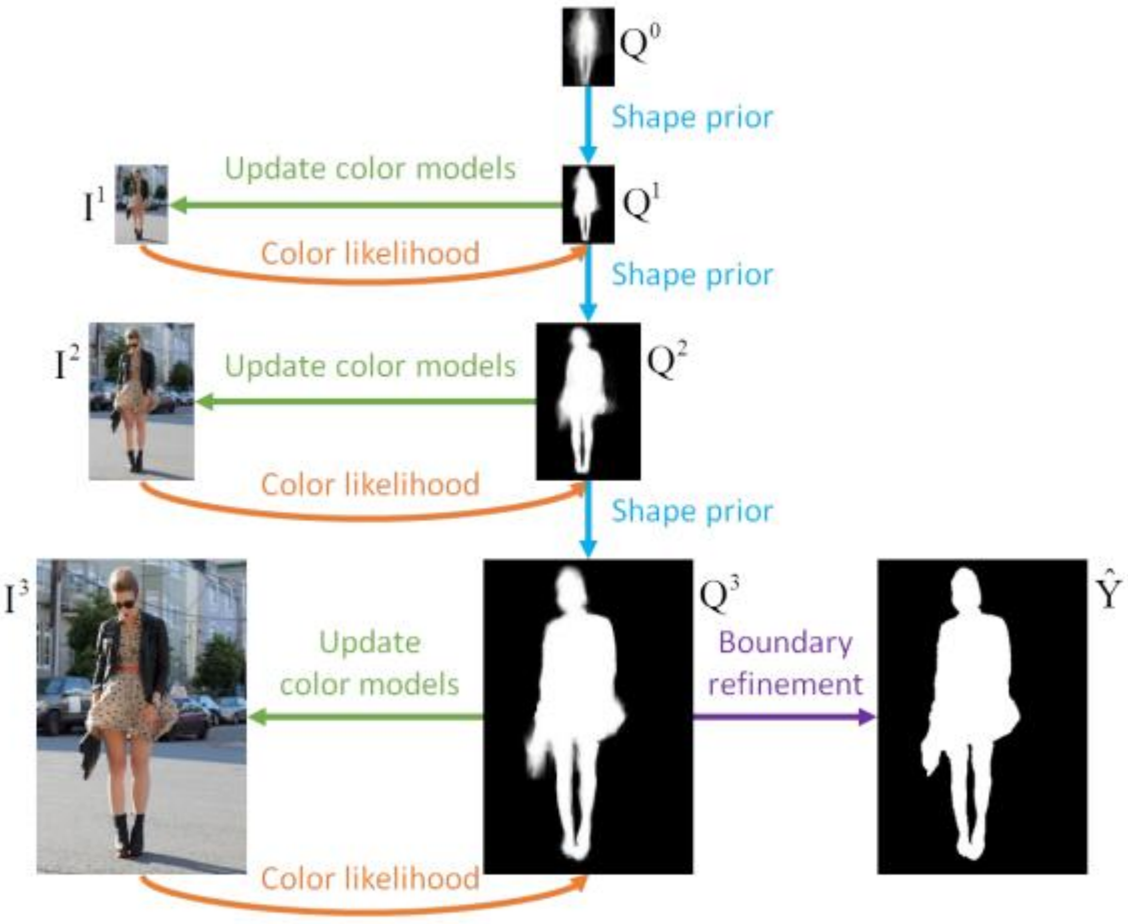
$$\hat{\mathbf{Y}} = \arg\min_{\mathbf{Y}} \sum_k \frac{1}{2} \|\mathbf{Y}(\Delta_k) - \hat{\mathbf{z}}_k\|^2$$

Algorithm 1 The single scale PatchCut algorithm.

1: **while** not converged **do**
2: for each patch Δ_k, select the candidate local shape mask $\hat{\mathbf{z}}_k$ by (10)
3: estimate the shape prior $\hat{\mathbf{Q}}$ by averaging $\hat{\mathbf{z}}_k$ and the segmentation $\hat{\mathbf{Y}}$ by (11)
4: update the foreground and background GMM color models $\{\mathbf{A}_1, \mathbf{A}_0\}$ by (2).
5: **end while**

This two step optimization states $\hat{\mathbf{Y}}$ as a binary function labeling a pixel as foreground or background. However, optimization is solved by finding a soft segmentation mask $\hat{\mathbf{Q}}$ having values between 0 and 1. This function can then be thresholded to find binary labeling function.

Multiscale Cascade Algorithm



Initialize shape prior from the segmentation maps of the examples

$$\hat{\mathbf{Q}}^0 = \frac{1}{M}\sum_m \mathbf{Y}^1_m$$

At each scale s=1, 2, 3 run the algorithm in the previous slide.

After calculating the last soft shape mask define:

$\hat{\mathbf{Y}}_t$	Thresholded version of soft shape mask
$\hat{\mathbf{Y}}_r$	Further refined version of shape mask with iterative graph cuts

Experiments (Fashionista*)

Fashionista Dataset:

- Consists of 700 street shots of fashion models
- Various poses, cluttered background and complex appearance
- Images are 600x400 pixels
- *Leave-one-out* tests are run: for each test image , remaining 699 images are used as database

* K. Yamaguchi, M. H. Kiapour, L. E. Ortiz, and T. L. Berg. Parsing clothing in fashion photographs. In *CVPR*, 2012.

Experiments (Fashionista)

Here are some qualitative results:

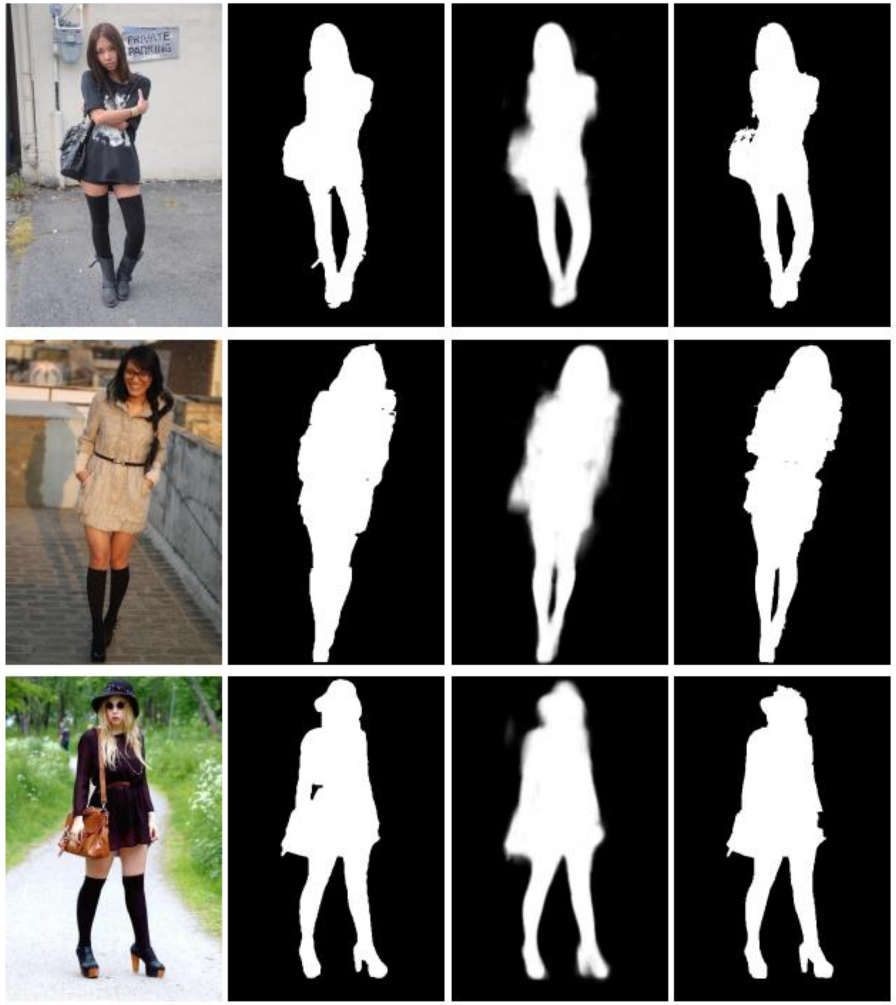


Image GT PatchCut_soft PatchCut

Experiments (Fashionista)

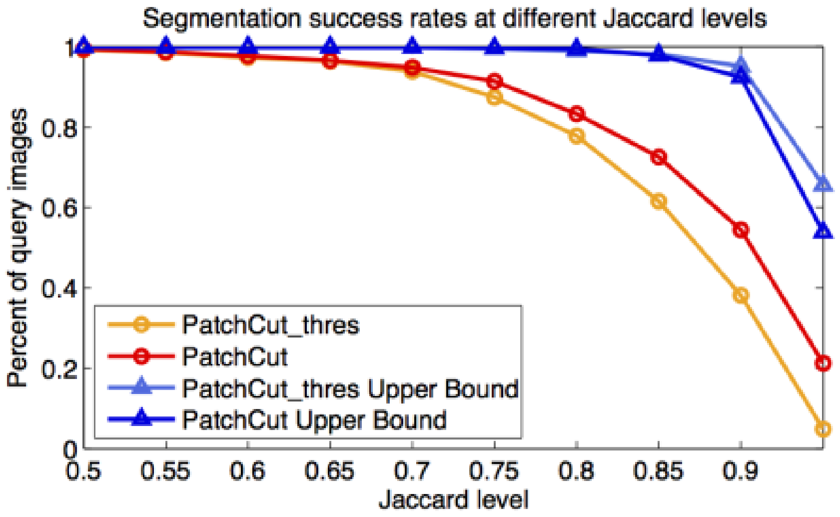
Here are the quantitative results:

Table 1: Segmentation performance on Fashionista.

	Jaccard (%)
GrabCut	64.23
PatchCut_thres	**86.25**
PatchCut	**88.33**
PatchCut_thres upper bound	95.72
PatchCut upper bound	95.20

Jackard (Intersection-over-Union) Score:

$$(|\hat{\mathbf{Y}} \cap \mathbf{Y}|/|\hat{\mathbf{Y}} \cup \mathbf{Y}|)$$



Estimating upper bound performance using ground truth segmentation by investigating different Jaccard levels

Experiments (Weizmann Horse*)

Weizmann Horse Dataset:

- 328 horse images with side views
- Widely used for benchmarking object segmentation algorithms
- 200 images are used for the database
- Remaining 128 images are used for the test set

* E. Borenstein and S. Ullman. Class-specific, top-down segmentation. In ECCV, 2002.

Experiments (Weizmann Horse)

Here are some qualitative results:

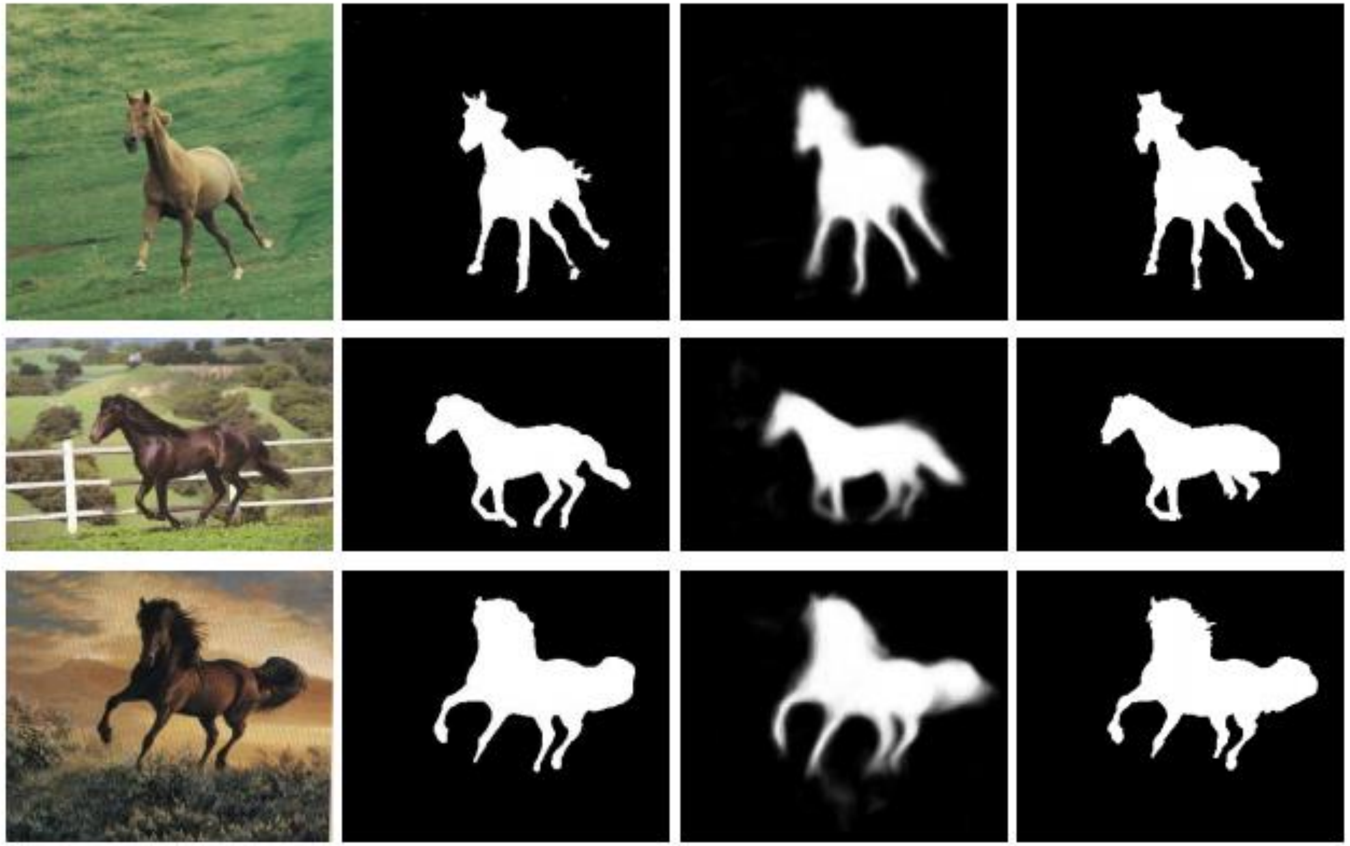


Image GT PatchCut_soft PatchCut

Experiments (Weizmann Horse)

Here are the quantitative results:

Table 2: Performance evaluation on Weizmann Horse.

	Jaccard (%)	Acc (%)
PatchCut_thres	**80.33**	**94.78**
PatchCut	**84.03**	**95.81**
Kernelized Structured SVM [4]	80.10	94.60
Fragment-based CRFs [21]	N/A	95.0
High-Order CRFs [23]	69.90	N/A
Max-Margin BMs [36]	75.78	90.71
Window Mask Transfer [17]	N/A	94.70

Comparison of the algorithms with Jaccard score and pixel-wise classification accuracy

Experiments (Object Discovery*)

Object Discovery Dataset:

- Consists of three object categories: airplane, car and horse
- Around 100 images in each category
- Images are collected from Internet
- Originally designed for evaluating object co-segmentation

* M. Rubinstein, A. Joulin, J. Kopf, and C. Liu. Unsupervise joint object discovery and segmentation in internet images. In CVPR, 2013.

Experiments (Object Discovery)

Here are some qualitative results:

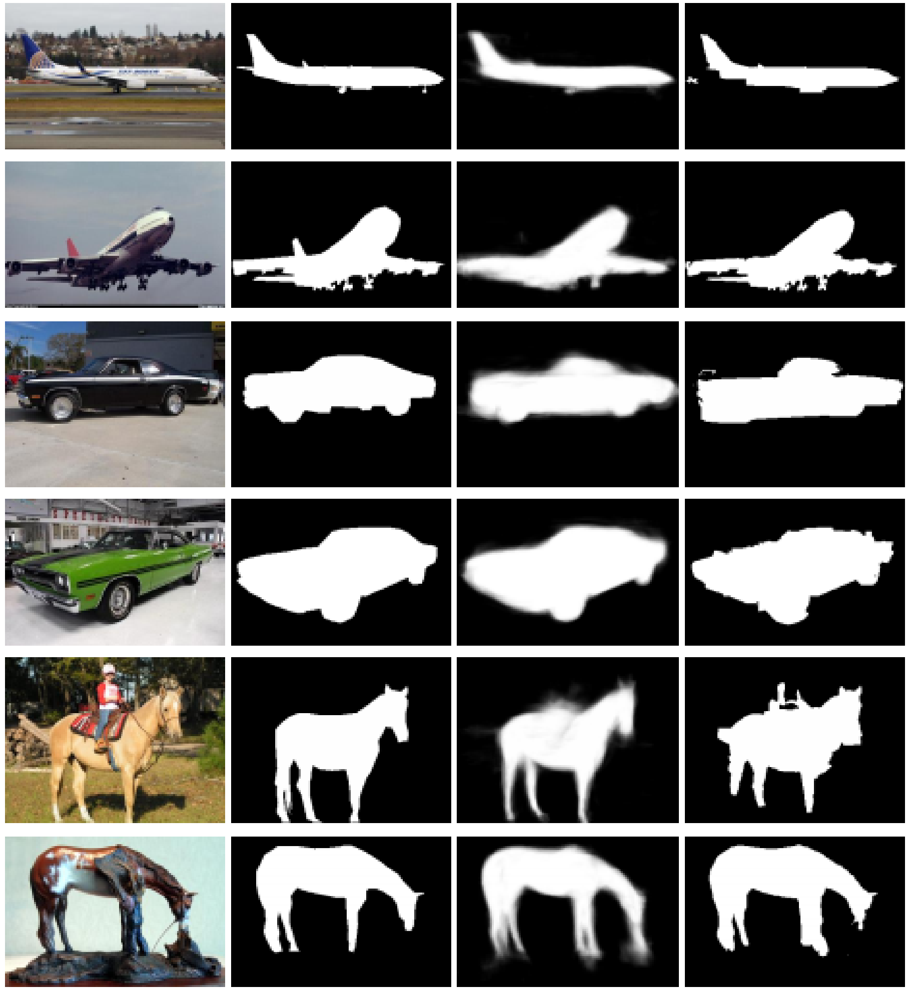


Image GT PatchCut_soft PatchCut

Experiments (Object Discovery)

Here are the quantitative results for different object categories:

Table 3: Jaccard scores on Object Discovery.

Jaccard (%)	Airplane	Car	Horse
GrabCut	63.29	67.63	50.32
Co-segmentation [30]	55.81	64.42	51.65
Ahmed et al. [2]	64.27	71.84	55.08
PatchCut_thres	**70.44**	**86.40**	**63.19**
PatchCut	**70.49**	**84.52**	**64.80**

Experiments (PASCAL*)

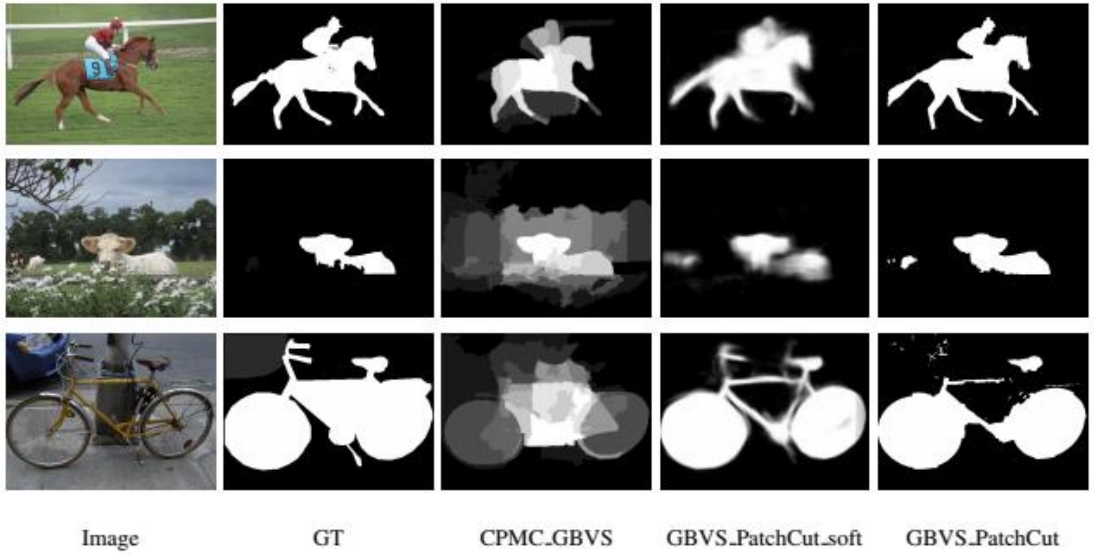
PASCAL VOC 2010 Dataset:

- Consists of 20 object classes
- Pose, shape and appearance variations and occlusions
- Training set images are used as database
- 850 images in the validation set are used as test set
- Salient object segmentation masks are collected for these sets

* M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The PASCAL Visual Object Classes Challenge 2010 (VOC2010) Results.

Experiments (PASCAL)

Here are some qualitative results:



This time, PatchCut is initialized with the saliency maps generated by the GBVS* and CPMC** algorithms

* J. Harel, C. Koch, and P. Perona. Graph-based visual saliency. In *NIPS*, 2006.

** Y. Li, X. Hou, C. Koch, J. M. Rehg, and A. L. Yuille. The secrets of salient object segmentation. In *CVPR*, 2014.

Experiments (PASCAL)

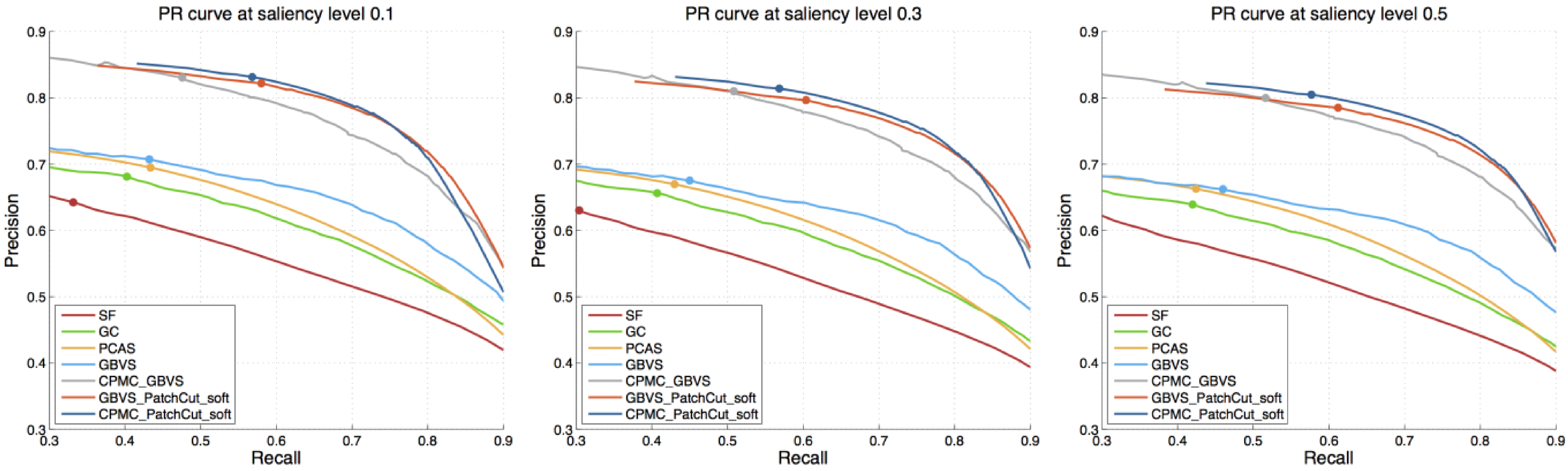
Here are the quantitative results, for different saliency levels:

Table 4: Jaccard scores on PASCAL.

Saliency level	0.1	0.3	0.5
GBVS_GrabCut	45.84	45.25	44.90
CPMC_GBVS [22]	59.43	60.58	60.75
GBVS_PatchCut_thres	60.08	60.22	59.27
GBVS_PatchCut	**62.02**	**62.15**	**61.14**
CPMC_PatchCut_thres	61.37	62.64	62.76
CPMC_PatchCut	**63.74**	**64.92**	**64.97**

Experiments (PASCAL)

Here are the quantitative results, as precision recall curves:



Conclusions

- A data driven object segmentation algorithm is presented
- MRF problem is decomposed into a set of independent label patch selection sub-problems, that are easier to solve in parallel
- A multiscale cascade algorithm in a coarse-to-fine manner
- Qualitative and quantitative evaluation on different datasets

Advantages

- No offline training
- Sub-problems can be solved in parallel
- No user interaction
- No prior knowledge on category specific object models

Disadvantages

- The effect of image retrieval on overall method performance is not evaluated
- Selection of some parameters such as number of scales (3) and size of patches (16x16) is not clarified well
- It is not clear when to refine the final mask using iterative graph cuts
- While claiming to be a category independent method, evaluations done on category specific datasets, such as Fashionista and Weizmann Horse

Disadvantages

- For multi-category datasets such as Object Discovery and PASCAL, comparisons done with methods suggested for different problems
- No qualitative results provided for other methods which are used for comparison
- While making quantitative comparisons with GrabCut, which is an interactive algorithm, a bad prior is provided to GrabCut

Future Work

- Generalized PatchMatch* can be used to increase the number of candidate patches from a single example image. This may improve the performance by eliminating noisy label patches.

*C. Barnes, E. Shechtman, D. Goldman, and A. Finkelstein. The generalized patchmatch correspondence algorithm. In ECCV, 2010.

Questions?

Questions?

Thank You…