Thesis: Interactive Object Segmentation using Binary Inputs*

Sivabalan Manivasagam[1], Gregory Canal[2], Shaoheng Liang[3], Dr. Christopher Rozell[4]

Spring 2018

Abstract

Every day, humans use their vision to process millions of pixels and select regions of interest. This task of highlighting and grouping pixels of interest in a scene is called image segmentation, and it is a fundamental method that humans use to communicate with each other ideas, concepts, and emotions. We develop an interactive image segmentation method that uses only binary inputs that allows users to adaptively select a region from the image. We propose two versions of our algorithm, and evaluate their ability to specify desired regions for the user with restricted inputs and noise on large, publicly available image data sets. We also compare our method to the previous best algorithm, developed by Rupprecht et al.

Introduction

One primary way humans interact with the world is through vision. Every day, humans must sense and process millions of pixels. In order to make sense of such a large amount of data, humans perceive and group pixels together into semantic, perceivable regions [1]. With the advent of technology such as SnapChat and Instagram, photo sharing and communication via image segmentation continues to increase and be the dominant form of online communication [2]. However, individuals dealing with [illegible] disabilities [illegible] such as amyotrophic lateral sclerosis, stroke, or spinal code injury are currently unable to fully participate online due to the lack of low-input, high-performing, and user-friendly hands-free methods for communicating and performing interactive (user-led) image segmentations. Recent hands-free technology such as brain computer interfaces and tongue drives have now opened the possibility of allowing these

* This work is based on: G. Canal, S. Manivasagam, S. Liang and C.J. Rozell. In Proceedings of the International Conference on Image Processing (ICIP), October 2018. Submitted.

individuals with muscle impairment to be able to start participating in these tasks [3-5]. This thesis proposal discusses previous work on interactive image segmentation, recent work on interactive image segmentation with low-dimensional binary input, and the space the proposed project hopes to fill.

Interactive image segmentation has been of significant interest in the computer vision community for professional image processing tasks [6, 7]. Most of these approaches involved detailed highlighting or tracing pixels on interest in the image (foreground) using a mouse or other free-form input mechanism. In addition, most approaches post-processed the input provided using image content dependent priors and heuristics. To reduce amount of human input required, Boykov and Jolly developed a graph cut based algorithm that requires coarser human strokes, which the algorithm then uses to infer foreground and background segmentation based on pixel color and gradient in the image [8, 9]. Additional work continued to automate the process by adding additional heuristics and priors on image content [10]. Others looked into using seed input (human selected pixels that are specified to be in the image segmentation) to create segmentations [11]. Rother and coworkers simplified image segmentation even further with an algorithm called GrabCut that only requires as input a bounding box [12-14], and then further post-processing [15, 16]. While these approaches are effective for image segmentation and processing, they still require fine-tuned user input, and our goal is to perform the task in an interactive hands-free manner.

An ideal hands-free image segmentation algorithm would have five main desirable properties:

1) Binary Input: We would like an image segmentation approach that only requires binary inputs, which can be communicated with brain-computer interfaces or other assisted communication devices.
2) Efficient: Fast communication rate of the image segmentation from user to computer
3) Simple: The approach should be intuitive and easy to use, and the user should be able to learn it quickly.
4) Robust: If a mistake is made in communicating the image segmentation, the approach should handle and correct for it
5) Feedback: The user should be able to understand and see the progression of the image segmentation

The only attempt at interactive binary image segmentation is a "N-Questions" based algorithm by Rupprecht and co-workers [17, 18]. The approach queries the user a simple question of whether a particular pixel is in the image segmentation or not, which the user answers with "yes" or "no." Each query pixel seed is chosen optimally to bisect the possible number of image segmentations ($2^{\#\text{ of pixels}}$), using tools such as Markov Chain Monte Carlo (MCMC) sampling and geodesic distance transform segmentation [19].

While the N-Questions algorithm performs well over several publicly available datasets such as the Berkeley 300 and Stanford Background data set, we have identified several areas for improvement based on the stated desirables above. First, while the N-Questions method may be resilient to input noise, such noise is not directly modeled in their algorithm. Secondly, the datasets used to test N-Questions emphasize arbitrary region selection, rather than object specification as is required in many real-world segmentation tasks. For instance, in the Stanford Background Dataset, ground truth segmentations (the segmentation of an image chosen by the actual user in the study) include entire foreground regions that merge multiple objects and may not be useful for communication of image segments of objects. Next, the use of the geodesic distance transform for determining optimal queries for segmentation creates a dependency between specified regions and inherent image topography, which may make it difficult to perform image segmentation of objects that have no relation to image content or color homogeneity. Finally, image segmentation from seed specification does not provide active feedback and may not keep the user engaged, since it is difficult to determine how specific answers will affect the resulting segmentation.

The system proposed in our lab uses an information theoretic approach with feedback previously applied in brain computer interface applications [20]. Rather than asking the user to specify the image segmentation, the algorithm proposes an image segmentation from a smaller ordered set of possible image segmentations (dictionary) to the user, and asks the user how "close" the algorithm. We believe that this new image segmentation system achieves the desirables of binary input, efficiency, simplicity, robustness, and feedback.

Author Attribution of Joint Project Work:

[1]Performing the image segmentation calculations

[2]Planned and designed experiment

[3]Implementing the posterior matching algorithm

[4]Principal Investigator

Literature Review

This literature review is broken into two main parts that are important to this work. We first discuss existing algorithms for object segmentation in an image, including those that use binary inputs. We then discuss studies that provided the motivation and application of posterior-matching information theoretic approaches used in this study's algorithm. In addition, we discuss existing image segmentation datasets that can be used to evaluate the performance of an image segmentation algorithm, and our image segmentation dataset choice designed for object selection.

Existing Algorithms for Object Segmentation

There are multiple studies that focus on the task of object segmentation, but most are in the computer vision community in the context of high performance image editing and computer graphics. A common approach is identifying foreground and background pixels via graph cuts [8-10, 14]. One of the first is Boykov and Jolly, which views the image in the context of a probabilistic framework by creating a graph using image pixels based on color and shading. The input provided by the user is demarcation of the image into foreground and background classes using "marker strokes" [8, 9]. This user-inputted demarcation of pixels is the "hard-constraint" on the generated image segmentation. They then devise an energy function ("soft constraints") like that in a Markov random field, where they penalize border pixels that are similar (borders should have sharp contrasts in neighboring pixel values), while encouraging pixels that are similar to have the same class. They then use an efficient min-cut/maxflow algorithm to derive pixel classifications of foreground and background regions. Inspired by their work, Rother and coworkers implemented an algorithm known as GrabCut, which combines the use of graph cuts with iterative optimization to get more accurate image segmentations and enhanced coverage of segments, such as those involving alpha matting. In addition, they improve usability of the tool by requiring only a bounding box for object identification, compared to the Bokyov and Jolly model which requires more detailed input. However, the model still requires specification of a region using a mouse cursor, and further segmentation when the algorithm is incorrect requires more input from the user [14]. Blake and coworkers build on and enhance the iterated GrabCut by including a Gaussian mixture model for more enhanced image segmentations and allowance of foreground regions outside the specified bounding box, improving usability and robustness for error [12]. Another method developed for object segmentation, developed by Ning and coworkers, merges small, user-provided input foreground and background regions, that are merged together to form the image segment. This method also require "marker-strokes" as user input, often difficult to communicate in a hands-free setting [11].

Some methods for object segmentation combine bounding boxes with saliency (the foreground in an image that the eye most likely saccades to first) metrics to identify the object of interest in an image using pixel colors and similar graph based approaches as discussed above [16, 21, 22]. These methods, while relying on simple user inputs and able to get a refined segmentation, require use of a mouse and are not hands free.

To our knowledge, the only work that has addressed interactive image segmentation with binary inputs is an algorithm developed by Rupprecht et al. [17]. This method, which we call "N-Questions," considers the set of all possible segmentations and probabilistically bisects the set based on user input of whether a pixel is foreground or not ("yes" or "no" response to the question). Proposed segmentations are generated using the geodesic distance transformation using the following set of parameters: color channel, smoothing, foreground and background seeds [19]. The optimal query the algorithm selects is with the Metropolis-Hashtings algorithm using Markov Chain Monte Carlo (MCMC) [23]. By appropriately querying the user the most optimal question, the space of possible segments is bisected. The authors extend their work to region selection in a 3-dimensional voxel space, which also uses binary responses [18]. While this algorithm addresses image segmentation in a binary input setting, there are several drawbacks that motivated this study. First, the N-Questions method demonstrates some resilience to input noise, such noise is not directly modeled in their algorithm. Secondly, the datasets used to test N-Questions emphasize arbitrary region selection, rather than object specification. Next, the use of the geodesic distance transform for determining optimal queries for segmentation creates a dependency between specified regions and inherent image topography; using N-Questions to specify segments whose geometry is independent with image content is difficult, since proposed segments naturally align with features in each image. Finally, image segmentation from seed specification does not provide active feedback and may not keep the user engaged, since it is difficult to determine how specific answers will affect the resulting segmentation. Therefore, a new approach that addresses these concerns is needed in interactive binary input segmentation.

Posterior-Matching and Information Theory Background

To develop a tool for object segmentation using solely binary inputs, we look to an information theoretic approach. By modelling the interactive object segmentation task as a user and computer communicating over a binary symmetric channel, we can utilize past work that allows for sending messages optimally, with robustness to noise [24]. Use of posterior-matching has been developed as a framework for designing algorithms in the brain-computer interface community [20, 25]. Omar and coworkers used posterior matching with feedback in brain computer interfaces for the use of making words and drawing out wheel chair paths [20]. The communication approach used in this study for posterior matching using noisy binary inputs is the one proposed for Shayevitz and Feder [26]. The main premise of this approach is that the error-convergence guarantees of the message being delivered correctly from encoder to decoder is simplified through the use of noiseless feedback [26]. In particular interest for this object segmentation task is the Burnashev-Zigangirov algorithm, which is adapted for a finite set of possible signals (like image segmentations in an image), and maintains the rate capacity and error-convergence properties discussed in the Shayevitz study [27].

Overview of Existing Image Segmentation Datasets and Selected Dataset

Two image segmentation datasets used by the N-Questions study are the Stanford Background Dataset [28] and the Berkeley Segmentation Dataset 300 [15], which were not chosen in this study for analysis.

The Stanford Background Dataset is composed of 725 images of outdoor scenes that have a horizon and a foreground object. The 725 images have a total of 3200 segmentations. The Stanford Background Dataset is an under-segmentation of images – it is intended for the purpose of scene understanding for 3D reconstruction. Natural outdoor scenes are decomposed into eight segments: sky, tree, road, grass, water, building, mountain, and foreground objects. The image below (taken from the Stanford Background Dataset) categorizes foreground objects into a single segmentation, even though the segment is composed of multiple cars, signs, and people. This is not an appropriate dataset to apply posterior matching since a subset of the 715 images used to construct the Stanford Background Dataset come from PASCAL-VOC, an object segmentation dataset, but segments are merged into a single region. The size of the dataset is inadequate for a rigorous testing of object segmentation. In addition, our method for object segmentation implies that a single segment should be one entity (one building) rather than several (multiple buildings).

Figure 2. Sample image from Stanford Background Dataset, with merged foreground region

The Berkeley Segmentation Dataset 300 is an over-segmentation of the images. The Berkeley segmentation dataset was created by taking 300 color images from the Correl dataset, and asking thirty students to decompose each image into semantic segments. The 300 color images produce 13,500 segmentations. Depending on the user, the resulting segmentations may look more like super pixels of the same color gradient than actual semantic objects (see below images).

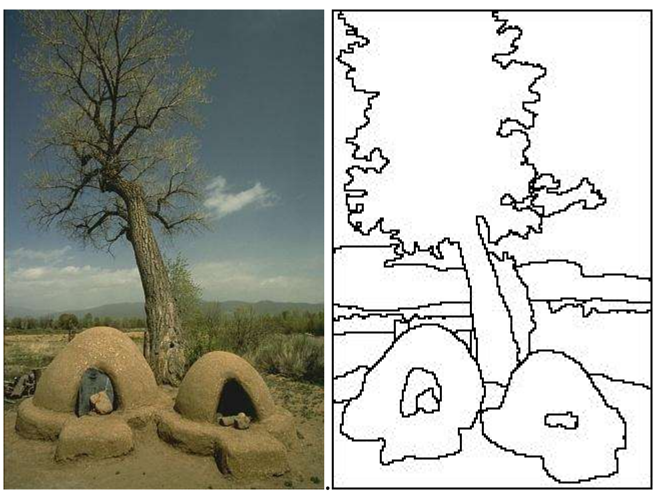


Figure 3. Sample BSDS image with 23 segments

We decided to apply our posterior matching algorithm in the context of object segmentation. Therefore, we evaluated the posterior matching algorithm's ability to identify and segment objects using the Microsoft Common Objects in Context (MS-COCO) validation dataset [29]. This dataset is suitable for this application as it contains over 280000 segmentations for 91 object classes (e.g. TV, dog, chair, racquet, cup) in both indoor and outdoor settings.

We chose MS-COCO over PASCAL-VOC [30], another object segmentation dataset to demonstrate the robustness of the method. The MS-COCO (~ 288000) validation dataset size is two orders of magnitude larger than PASCAL-VOC (~ 20000). The segmentation quality, higher number of instances per category and large number of categories are additional reasons we chose MS-COCO.

In addition to using the MS-COCO dataset, we do not do any pre-processing and filtering of viable segments. While in N-Questions segments that are less than 1% of the total image size are excluded, these segments less than 1% our important objects to extract from the image. For example, the following object segment is 0.46% of the total image.

Figure 4. Example of Image Segment < 1%

By incorporating aspects developed by the computer vision community, along with posterior matching theory, and an appropriate dataset for evaluation, we can contribute to the task of interactive object segmentation.

Methods and Materials

Proposed Algorithm

We now discussed the proposed method that attempts to do noisy binary image segmentation using posterior matching method. Figure 5 shows the model used to convert the problem of image segmentation into an information theoretic problem.

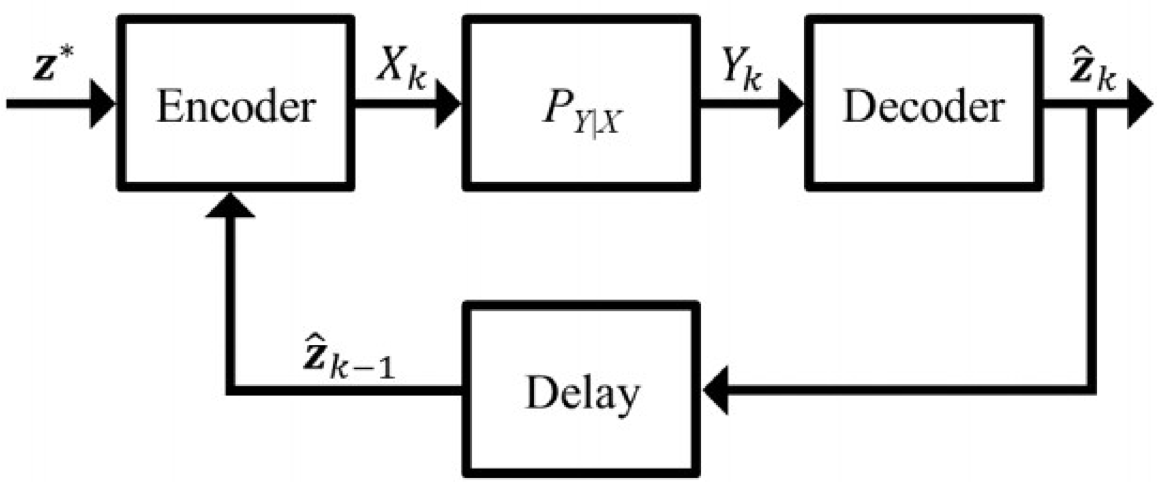


Figure 5. Model interactive image segmentation method as a communication channel

The user (the encoder), wants to select an object in an image, which can be simplified as a binary string; 1 denotes a pixel that should be included in the segmentation, while 0 excludes the pixel. Since the range of possible image segmentations is 2^n, this is an impractical search space. Furthermore, the set of reasonable image segmentations is sparse – as shown in the literature review, there may only be at most 100 segmentations of interest. If we constrain the set of image segmentations in the image and establish a lexicon or ordering of image segments, the user can then have a goal string in the lexicon that most closely matches the goal image segmentation. This goal string is $\boldsymbol{z}^*$. We then establish a communication protocol (an encoding scheme) with the user and the algorithm over a noisy binary symmetric channel: the decoder proposes an image segmentation from the lexicon $\hat{z}_{k-1}$ which is provided as feedback to the encoder. The encoder compares the feedback received with the goal segmentation and specifies a new binary input X_k (which can be noisy) depending on whether the proposed segmentation is higher or lower in the lexicon, which updates a posterior probability distribution over the set of strings in the lexicon $P_{Y|X}$. The algorithm then samples the median of this distribution, decodes the string into the image segmentation $\hat{z}_k$. Figure 6 shows a sample iteration of the image segmentation algorithm, which is the Burnashev-Zigangirov algorithm, applied to image segmentation.

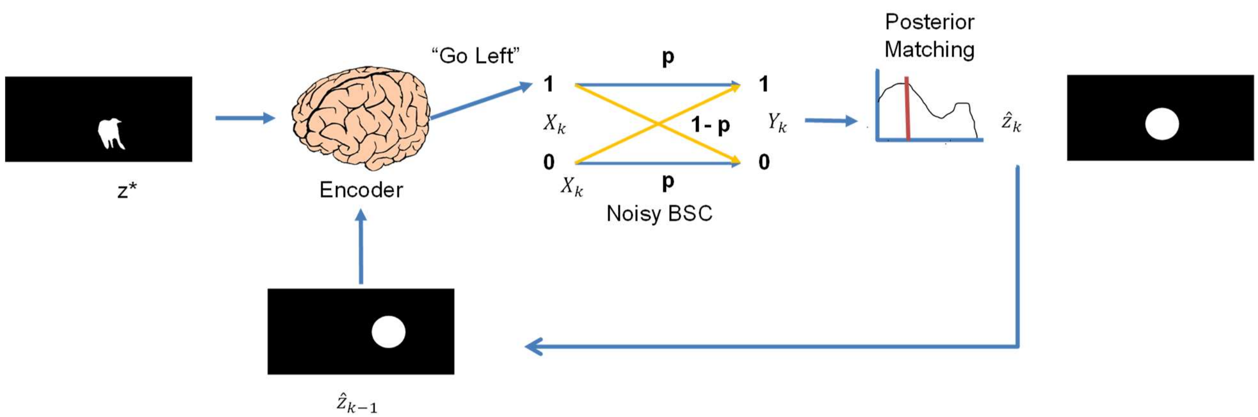
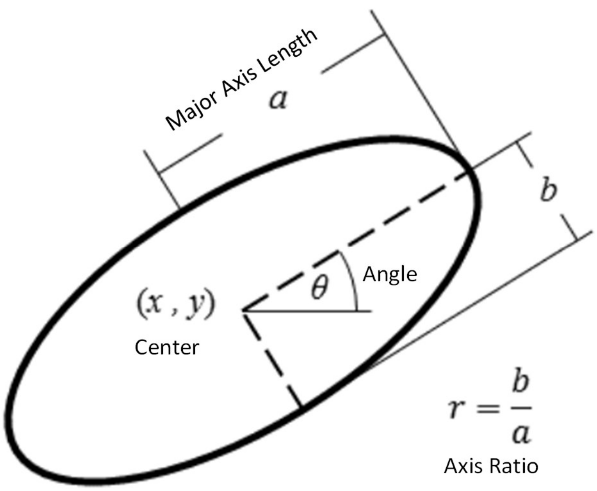


Figure 6. Example iteration of segmentation of sheep using Ellipse lexicon

Our innovation upon the B-Z algorithm and critical to the success of the image segmentation algorithm is choice of lexicon. We propose a finite lexicon of image segments, represented by tuple z = $\{z_1, z_2, \ldots, z_l\}$ instead of infinite length sequences. Our lexicon is a set of ellipse masks which span a given image with various locations, orientations, sizes, and aspect ratios. Ellipses were chosen since they specify a coarse approximation of a given region, but can also be tuned parametrically. Specifically, an ellipse is described by the tuple {**y**, **x**, **θ**, **a**, **r**} where **y** is the vertical position of ellipse center, **x** is the horizontal position of ellipse center, **θ** is the angle of the major axis from the horizon, **a** is the half-length of the major axis, and **r** is the aspect ratio of the major axis to minor axis (see example ellipse mask in Figure 7). Therefore, the possible image segments is discretized into a dictionary of possible ellipses. Each element in the ellipse tuple can take on a range of discrete values that are bounded. For example, **y** ranges from 0 to height of the image, **x** ranges from 0 to length of image, **θ** consists of m_θ values equally spaced from 0 to π radians (π is used instead of 2π due to the symmetry of ellipses). For our experiment the list of possible values is chosen uniformly, but expert domain knowledge can provide a more methodical choice in dictionary elements.



Letter	0	1
y	Higher < Lower	
x	Left < Right	
θ	Smaller < Larger	
a	Smaller < Larger	
r	Smaller < Larger	

Figure 7. Ellipse Lexicon Design

Experiment

While this algorithm requires a user to interactively choose a goal image segmentation, we wanted to evaluate the method in a controlled setting an apply it to hundreds of thousands of image segmentations. Therefore, we tested our algorithm on ground truth image segmentation regions from the MS-COCO dataset. These ground truth segments mimic a human's intention of objects she would want to select in an image. We applied the Ellipse-Lex algorithm to this dataset, testing various noise levels (0, 5, and 10%). We also compared it to N-Questions, the most recent algorithm in interactive binary image segmentation mentioned in the literature review. In order to do proper simulation of an interactive algorithm on a image segmentation dataset, we need to mimic the actions a human in order to reach the target segmentation. For N-questions, the ground truth can be queried for whether a pixel is in the image segmentation. For Ellipse-Lex, we generate the goal ellipse tuple that has the highest F1 score. Based on this “ground truth” ellipse we give the algorithm feedback. The Ellipse-Lex algorithm was run for 30 inputs, producing a segment mask after each input. For each of these produced segments, the F1 score (the harmonic mean of precision (number of pixels that match ground truth / number of pixels selected by algorithm) and recall (number of pixels that match ground truth / number of ground truth pixels) was calculated with respect to the original ground truth region. We evaluated quality of image segmentation with F1 score, which is the harmonic mean of precision and recall. Precision is fraction of pixels correctly predicted in the image segmentation out of the proposed set, while recall is the fraction of pixels correctly predicted from the ground truth. We tested two ellipse lexicons with low and high resolutions, referred to as “Ellipse-Lex-Low” and “Ellipse-Lex-High” respectively, with alphabet size parameterizations (m_y, m_x, m_θ, m_a, m_r) equal to (15, 20, 10, 20, 10) and (100, 100, 20, 20, 10). Choice of ellipse lexicon resolution occurred via trial and error. Implementation was done in MATLAB on the Georgia Tech’s PACE cluster.

Results

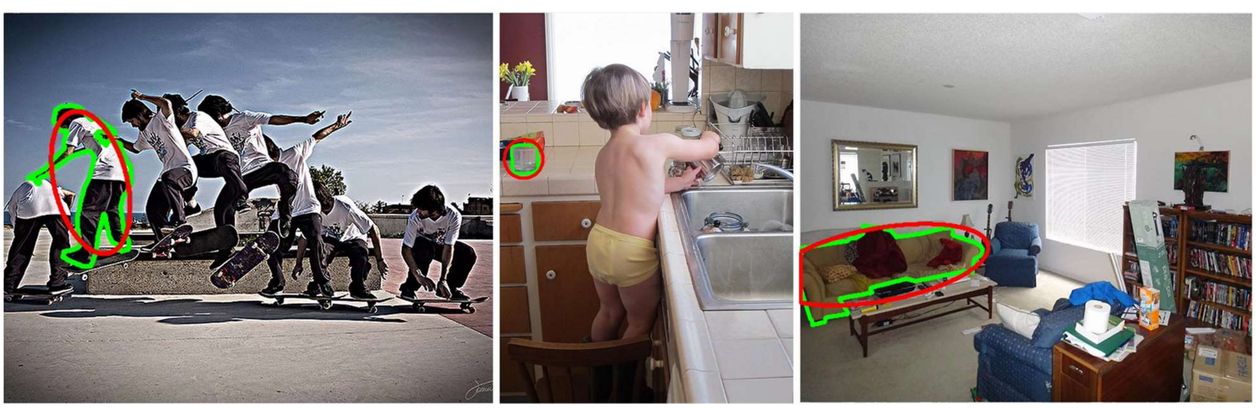


Figure 8. Qualitative Appearance of Image Segmentation is three different example images. Green is ground truth, red is achieved segmentation using EllipseBZ-High.

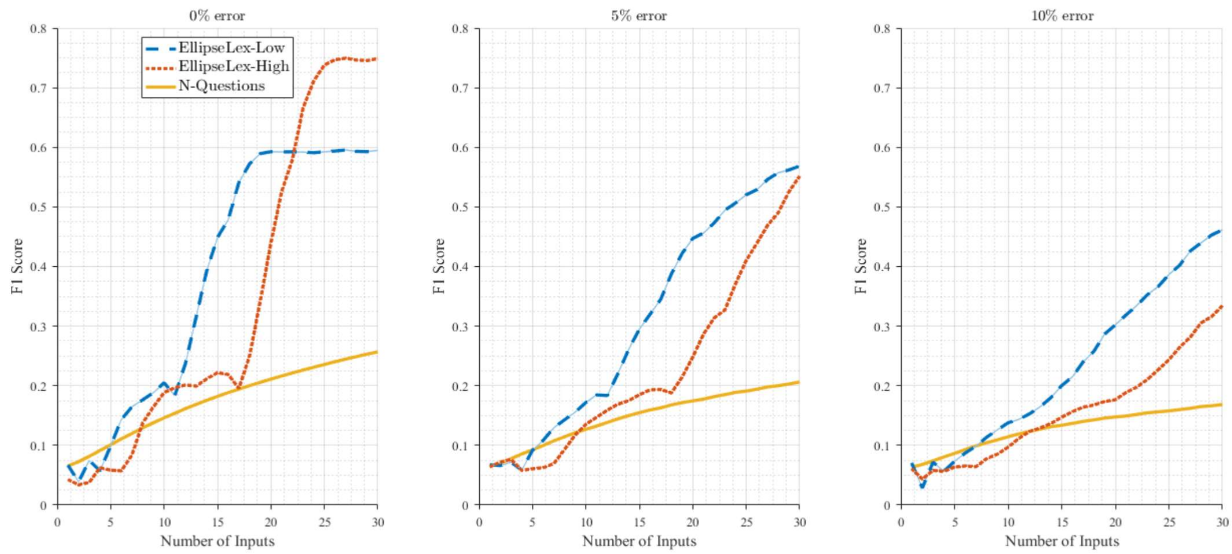


Figure 9. Performance vs. Number of Inputs

In Figure 9, mean F1 score from all ground truth regions in MS-COCO are plotted against number of binary inputs for methods across all noise levels. There is a spike in performance for the Ellipse-Lex-Low/High methods, after 12 and 17 inputs respectively.

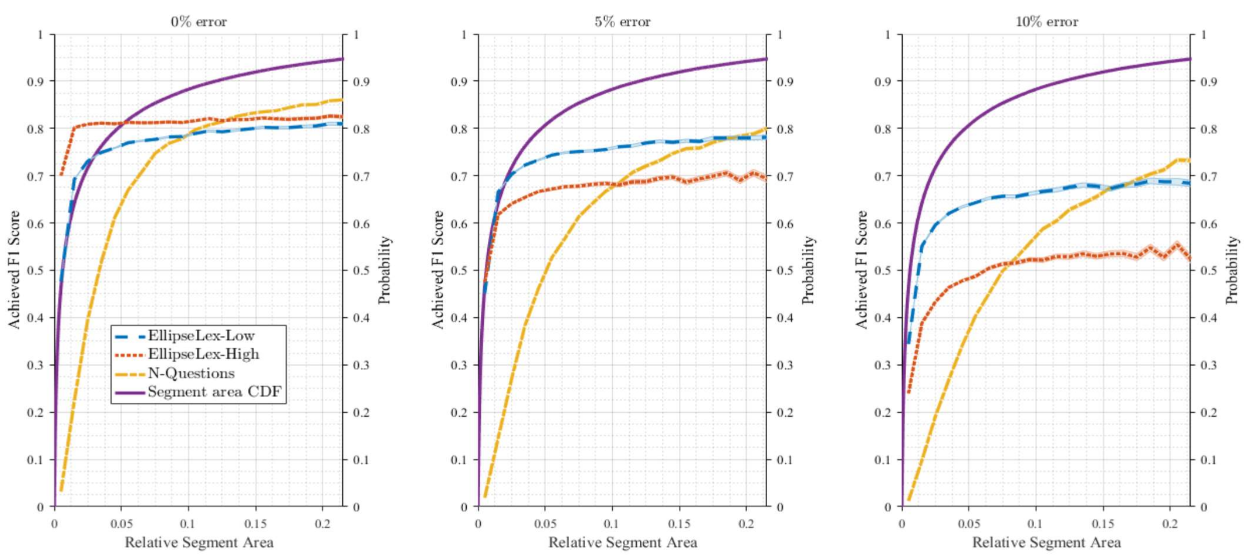


Figure 10. Performance vs. % of Relative Area in Image Segmentation

In Figure 10, mean F1 score is plotted against the relative segment area (the number of pixels in the segment / the total number of pixels in the image) for methods across all noise levels.

Discussion

The results demonstrate that, compared to N-Questions, Ellipse-Lex provides several quantitative advantages. First, Ellipse-Lex has higher F1 scores with ground truth segments compared to N-Questions after about 10 inputs, and reaches significantly higher F1 scores after about 25 inputs. Ellipse-Lex performs better than N-Questions across the three different noise levels evaluated, demonstrating its potential to be robust in a real-world setting. In addition, Figure 10 shows that Ellipse-Lex performs better than N-Questions on image segments that are small relative to the total image area. While N-Questions may be a strong algorithm for background segmentation (it achieves higher performance for segment areas larger than 0.15 of total image), these results suggest that Ellipse-Lex may be a better choice for object segmentation and highlighting small regions of interest, especially if the region does not have homogenous pixel properties. Ellipse-Lex also provides tunable parameters that allows for the user to balance communication speed with segment accuracy: Ellipse-Lex-High can capture up to 0.75 of the F1 score compared to the target segment, demonstrating that the ellipse is an ideal candidate for representing image segmentations. Ellipse-Lex-Low can reach the target segment in fewer than 20 inputs, demonstrating communication speed. Analyzing Ellipse-Lex and N-Questions by the 80 different object classes in MS-COCO, Ellipse-Lex performs better across all image classes compared to N-Questions.

Conclusion

The results in this work demonstrate the potential of the posterior matching method with an ellipse-based lexicon for specifying object segments using only noisy binary inputs. Our method performs well in noise levels that fall in the typical range of crossover probabilities in binary input brain-computer interface systems, indicating an ability to rapidly and precisely specify object segments in a manner resilient to input errors. Our method is robust because it explicitly models noise in posterior-matching as well as the convergence guarantees the algorithm provides.

Future Work

The use of an ellipse lexicon with posterior-matching presents several promising opportunities for future work and improvements. After an ellipse mask is produced, it can be input as a bounding box into an additional postprocessing stage that utilizes inherent structure in natural images to produce segmentations, such as the popular GrabCut segmentation algorithm. Furthermore, the posterior-matching framework can naturally incorporate a prior probability distribution over the lexicon into the segmentation algorithm. One approach might be to generate such priors using a saliency metric. Finally, the use of posterior-matching simply requires an ordered segment lexicon, allowing for the substitution of other segment classes besides ellipses. One approach might be to create a dynamic lexicon, which could be constructed from objects in the image itself by creating a set of bounding box proposals. Overall, posterior-matching in combination with a well-designed, ordered lexicon presents an algorithm of theoretical and practical interest for interactive object segmentation with noisy binary inputs.

References

1 Koffka, K.: ‘Principles of Gestalt psychology’ (Routledge, 2013. 2013)

2 Duggan, M.: ‘Photo and video sharing grow online’, Pew Research Internet Project, 2013

3 Park, H., Kiani, M., Lee, H.-M., Kim, J., Block, J., Gosselin, B., and Ghovanloo, M.: ‘A wireless magnetoresistive sensing system for an intraoral tongue-computer interface’, IEEE transactions on biomedical circuits and systems, 2012, 6, (6), pp. 571-585

4 Wolpaw, J.R., Birbaumer, N., McFarland, D.J., Pfurtscheller, G., and Vaughan, T.M.: ‘Brain–computer interfaces for communication and control’, Clinical neurophysiology, 2002, 113, (6), pp. 767-791

5 Pfurtscheller, G., Neuper, C., Guger, C., Harkam, W., Ramoser, H., Schlogl, A., Obermaier, B., and Pregenzer, M.: ‘Current trends in Graz brain-computer interface (BCI) research’, IEEE Transactions on Rehabilitation Engineering, 2000, 8, (2), pp. 216-219

6 Chuang, Y.-Y., Curless, B., Salesin, D.H., and Szeliski, R.: ‘A bayesian approach to digital matting’, in Editor (Ed.)^(Eds.): ‘Book A bayesian approach to digital matting’ (IEEE, 2001, edn.), pp. II-II

7 Mortensen, E.N., and Barrett, W.A.: ‘Intelligent scissors for image composition’, in Editor (Ed.)^(Eds.): ‘Book Intelligent scissors for image composition’ (ACM, 1995, edn.), pp. 191-198

8 Boykov, Y., and Kolmogorov, V.: ‘An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision’, IEEE transactions on pattern analysis and machine intelligence, 2004, 26, (9), pp. 1124-1137

9 Boykov, Y.Y., and Jolly, M.-P.: ‘Interactive graph cuts for optimal boundary & region segmentation of objects in ND images’, in Editor (Ed.)^(Eds.): ‘Book Interactive graph cuts for optimal boundary & region segmentation of objects in ND images’ (IEEE, 2001, edn.), pp. 105-112

10 Freedman, D., and Zhang, T.: ‘Interactive graph cut based segmentation with shape priors’, in Editor (Ed.)^(Eds.): ‘Book Interactive graph cut based segmentation with shape priors’ (IEEE, 2005, edn.), pp. 755-762

11 Ning, J., Zhang, L., Zhang, D., and Wu, C.: ‘Interactive image segmentation by maximal similarity based region merging’, Pattern Recognition, 2010, 43, (2), pp. 445-456

12 Blake, A., Rother, C., Brown, M., Perez, P., and Torr, P.: ‘Interactive image segmentation using an adaptive GMMRF model’, in Editor (Ed.)^(Eds.): ‘Book Interactive image segmentation using an adaptive GMMRF model’ (Springer, 2004, edn.), pp. 428-441

13 Lempitsky, V., Kohli, P., Rother, C., and Sharp, T.: ‘Image segmentation with a bounding box prior’, in Editor (Ed.)^(Eds.): ‘Book Image segmentation with a bounding box prior’ (IEEE, 2009, edn.), pp. 277-284

14 Rother, C., Kolmogorov, V., and Blake, A.: ‘Interactive foreground extraction using iterated graph cuts, 2004’, in Editor (Ed.)^(Eds.): ‘Book Interactive foreground extraction using iterated graph cuts, 2004’ (edn.), pp.

15 Marchesotti, L., Cifarelli, C., and Csurka, G.: ‘A framework for visual saliency detection with applications to image thumbnailing’, in Editor (Ed.)^(Eds.): ‘Book A framework for visual saliency detection with applications to image thumbnailing’ (IEEE, 2009, edn.), pp. 2232-2239

16 Jiang, H., Wang, J., Yuan, Z., Liu, T., Zheng, N., and Li, S.: ‘Automatic salient object segmentation based on context and shape prior’, in Editor (Ed.)^(Eds.): ‘Book Automatic salient object segmentation based on context and shape prior’ (2011, edn.), pp. 9

17 Rupprecht, C., Peter, L., and Navab, N.: ‘Image segmentation in twenty questions’, in Editor (Ed.)^(Eds.): ‘Book Image segmentation in twenty questions’ (2015, edn.), pp. 3314-3322
18 Dubost, F., Peter, L., Rupprecht, C., Becker, B.G., and Navab, N.: ‘Hands-Free Segmentation of Medical Volumes via Binary Inputs’, in Editor (Ed.)^(Eds.): ‘Book Hands-Free Segmentation of Medical Volumes via Binary Inputs’ (Springer, 2016, edn.), pp. 259-268
19 Toivanen, P.J.: ‘New geodosic distance transforms for gray-scale images’, Pattern Recognition Letters, 1996, 17, (5), pp. 437-450
20 Omar, C., Akce, A., Johnson, M., Bretl, T., Ma, R., Maclin, E., McCormick, M., and Coleman, T.P.: ‘A feedback information-theoretic approach to the design of brain–computer interfaces’, Intl. Journal of Human–Computer Interaction, 2010, 27, (1), pp. 5-23
21 Cheng, M.-M., Mitra, N.J., Huang, X., Torr, P.H., and Hu, S.-M.: ‘Global contrast based salient region detection’, IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37, (3), pp. 569-582
22 Srivatsa, R.S., and Babu, R.V.: ‘Salient object detection via objectness measure’, in Editor (Ed.)^(Eds.): ‘Book Salient object detection via objectness measure’ (IEEE, 2015, edn.), pp. 4481-4485
23 Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., and Teller, E.: ‘Equation of state calculations by fast computing machines’, The journal of chemical physics, 1953, 21, (6), pp. 1087-1092
24 Cover, T.M., and Thomas, J.A.: ‘Elements of information theory’ (John Wiley & Sons, 2012. 2012)
25 Akce, A., and Bretl, T.: ‘A probabilistic language model for hand drawings’, in Editor (Ed.)^(Eds.): ‘Book A probabilistic language model for hand drawings’ (IEEE, 2010, edn.), pp. 109-112
26 Shayevitz, O., and Feder, M.: ‘Optimal feedback communication via posterior matching’, IEEE Transactions on Information Theory, 2011, 57, (3), pp. 1186-1222
27 Castro, R., and Nowak, R.: ‘Active sensing and learning’, Foundations and Applications of Sensor Management, 2009, pp. 177-200
28 Gould, S., Fulton, R., and Koller, D.: ‘Decomposing a scene into geometric and semantically consistent regions’, in Editor (Ed.)^(Eds.): ‘Book Decomposing a scene into geometric and semantically consistent regions’ (IEEE, 2009, edn.), pp. 1-8
29 Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., and Zitnick, C.L.: ‘Microsoft coco: Common objects in context’, in Editor (Ed.)^(Eds.): ‘Book Microsoft coco: Common objects in context’ (Springer, 2014, edn.), pp. 740-755
30 Everingham, M., Van Gool, L., Williams, C.K., Winn, J., and Zisserman, A.: ‘The pascal visual object classes (voc) challenge’, International journal of computer vision, 2010, 88, (2), pp. 303-338