Çukurova Üniversitesi Mühendislik Fakültesi Dergisi, 37(2), ss. 569-576, Haziran 2022
*Cukurova University Journal of the Faculty of Engineering, 37(2), pp. 569-576, June 2022*

# Spatially Adaptive DGL Test for Robust User-Assisted Multilabel Segmentation

**Hüseyin AFŞER**[*1] **ORCID** *0000-0002-6302-4558*

[1]*Adana Alparslan Türkeş Bilim ve Teknoloji Üniversitesi Mühendislik Fakültesi, Elektrik-Elektronik Mühendisliği Bölümü, Adana*

*Geliş tarihi: 25.03.2022* *Kabul tarihi: 30.06.2022*

*Atıf şekli/ How to cite: AFŞER, H., (2022). Spatially Adaptive DGL Test for Robust User-Assisted Multilabel Segmentation. Çukurova Üniversitesi, Mühendislik Fakültesi Dergisi, 37(2), 569-576.*

## Abstract

Recently, the DGL test has been successfully applied to the user-assisted image segmentation problem where different types of user inputs, e.g. labeled pixels from ground truth masks, bounding boxes and pixel seeds, can be robustly leveraged to assist the segmentation process in a simple and effective way. However, in the baseline method the spatial information of the user inputs is not utilized and the test is implemented in the color domain. In this work, we propose a spatially adaptive version of the DGL test where the spatial information of the user-input regions is incorporated into the decision making process of the original test for an improved segmentation performance. We show that the proposed approach can be simply and seamlessly integrated into the baseline method without increasing its computational and algorithmic complexity. We demonstrate simulations on the Berkeley's BSDS500 image database that validate the effectiveness of the proposed method. We also present benchmarking results which indicate that the accuracy can be improved by about 3% compared to the baseline method.

**Keywords:** User-assisted segmentation, Multiple instance segmentation, Robust hypothesis testing, DGL test

## Kullanıcı Yardımına Kararlı, Uzlamsal Adaptif DGL Test Tabanlı Çoklu Görüntü Kesitleme

## Öz

Son dönemde DGL testi, kullanıcı yardımlı görüntü kesitleme problemine başarıyla uygulanmış ve etiketlenmiş pikseller, kesit çerçeveleri ve piksel tohumları gibi farklı kullanıcı girdileriyle kararlı bir şekilde çalışarak çoklu görüntü kesitleme problemine basit ve etkili bir çözüm olarak sunulmuştur. Fakat,

[*]Corresponding author (Sorumlu yazar): Hüseyin AFŞER, *afser@atu.edu.tr*

sunulan temel yöntemde kullanıcı girdilerinin görüntü kesitleri hakkında sağladığı uzlamsal bilgiden faydalanılmamış ve test sadece renk uzayında uygulanmıştır. Bu çalışmada, kullanıcı girdilerinin uzlamsal bilgilerinin daha iyi bir kesitleme performansı için temel karar verme mekanizmasına dahil edildiği, uzlamsal olarak duyarlı bir DGL testi sunulmuştur. Önerdiğimiz yöntemin, algoritmik ya da hesaplama karmaşıklığını arttırmadan, basit ve muntazam bir şekilde temel yönteme dahil edilebildiği gösterilmiştir. Berkeley BSDS500 görüntü veri tabanında yaptığımız betimlemeler önerilen yöntemin faydalarını göstermekte olup; performans betimlemeleri, temel yönteme göre %3 oranında kesitlemede iyileştirme elde edilebileceğini göstermektedir.

**Anahtar Kelimeler:** Kullanıcı yardımlı görüntü kesitleme, Çoklu görüntü kesitleme, Karalı hipotez testi, DGL testi

## 1. INTRODUCTION

Image segmentation, in general, is an NP-hard problem. Recently, several user-assisted segmentation methods have been developed in which different types of user inputs can be leveraged to assist the segmentation process. These inputs can be labeled pixels from the regions of interests, their bounding boxes, seeds points or scribbles. The mostly used methods include the Graph Cuts [1], the Random Walk [2] and the GrabCut [3] algorithms. For a detailed overview of these methods and their many variants we refer the reader to [4].

The methods cited above are often designed to work effectively under a particular user input type. For example, in Graph Cuts the user inputs are labeled seed points whereas in Random Walk the user assistance is provided via image scribbles. However, these algorithms can be sensitive to the amount or the precision of the user input and their performances may deteriorate if the user inputs are not delivered properly. Therefore, many variant algorithms (see [4]) are suggested in the literature to improve the robustness of these methods to user inputs.

The DGL test [5-7] is a robust multiple hypothesis testing procedure. Recently, this test has been successfully applied to the user-assisted multilabel image segmentation problem. In [8] it is shown that the inherent robustness of this test can be leveraged with a simple and effective segmentation method for robust operation under different types of user inputs. The proposed baseline method can be implemented with linear complexity in the number of pixels and quadratic in the number of image regions. Moreover, the method is algorithmically minimal in the sense that it can be implemented with around 30 lines of Matlab code. However, this baseline method only utilizes the empirical distributions, i.e. the histograms of the user input regions, and does not consider their spatial information which may be crucial to improve the segmentation process. For example, when the user inputs are bounding boxes of the image regions, then these boxes also provide a spatial information about the whereabouts of the regions of interest. In this paper we propose an adaptation of the baseline DGL test that can take advantage of this spatial information and provide performance improvements.

The main contributions of our work are as follows: we propose a spatially adaptive DGL test based, user-assisted segmentation method where the spatial information of the user inputs is incorporated in the DGL test in a seamless manner. This is accomplished by including the spatial information of the user inputs in the decision making process of the DGL test by combining the color and spatial domain disparity metrics. We show that such an approach does not increase the computational and algorithmic complexity of the original method while providing performance improvements. We validate the performance improvements compared to the baseline method on the Berkeley BSDS500 database [9]. Our benchmarking results indicate that the proposed method improves the segmentation accuracy around %3 for different types of user inputs such as fraction of labeled pixels form ground truth mask, bounding boxes, random seed points and perturbed bounding boxes.

## 2. PRELIMINARIES

### 2.1. Notation

We use capital letters X, Y for random variables and lowercase letters $x, y$ for their realizations. We let $\mathcal{X}, \mathcal{Y}$ to denote the alphabets such that $x \in \mathcal{X}$, and $y \in \mathcal{Y}$ where $|\mathcal{X}|$ and $|\mathcal{Y}|$ are used to denote the sizes of these alphabets. The cross product of the alphabets is denoted as $\mathcal{X} \times \mathcal{Y}$. The sequence of variables is denoted as $X_1$, $X_2$, ...., $X_N$ and we use the standard Landau notation o(N) and O(N) to denote the limiting values of functions.

### 2.2. Problem Statement

We assume the probabilistic formulation in [8], [10] and consider the segmentation process $I : \Omega \rightarrow \mathbb{R}^d$ where $\Omega$ is the pixel grid with size $|\Omega| = N$. We assume that the image consists of M disjoint segments $\Omega_1$, $\Omega_2$, ..., $\Omega_M$ where a segment may consist of different separated regions. Let X be an arbitrary pixel in $\Omega$ and I(X) denote its intensity. We assume that the pixel intensities are independent and identically distributed (i.i.d.) in the regions of interest, $\Omega_i$ , $i = 1, 2, \dots, M$ as (Equation 1):

$$\{I(X)|X \in \Omega_i\} \sim P_i, \qquad i = 1, 2, \dots, M, \tag{1}$$

where $P_i$ are the intensity distributions. The assumed image model is demonstrated in Figure 1. In this paper, we consider digitized images where the pixel intensities come from a discrete alphabet, $\mathcal{X}$, with dimension, $d$, as $\mathcal{X} = \mathcal{X}_1 \times \mathcal{X}_2 \times \dots \times \mathcal{X}_d$ so that $I : \Omega \rightarrow \mathcal{X}$ and $P_i : \mathcal{X} \rightarrow [0,1]$.

In multilabel segmentation one seeks a decision (labeling) rule D, that is of the form D: $\Omega \rightarrow \{1, 2, ..., M\}$ so that D(X)= $i$ is chosen provided that $X \in \Omega_i$.

### 2.3. DGL Test-Based Segmentation

DGL test by Devroye et al. [5] is a non-linear majority voting test that is suitable for robust hypothesis testing applications where the true distributions of the hypothesis are not readily available but one has access to a set of nominal distributions that are known to be close to true distributions in total variation distance. Recently, this test has been successfully adapted to user-assisted image segmentation in [8] where different types of user inputs such as bounding boxes and pixel seeds can be robustly utilized to aid the segmentation task. This is accomplished by using the empirical intensity distributions of the user inputs regions as nominal distributions in the DGL test.

Let $L_1, L_2, \dots, L_M$ denote M sets of image pixels that are labeled in accordance with the user inputs. This set of pixels may be gathered randomly from the bounding boxes, pixels seeds or ground truth mask. Let $Q_i : \mathcal{X} \rightarrow [0,1]$, $\mathcal{X} = \mathcal{X}_1 \times \mathcal{X}_2 \times \dots \times \mathcal{X}_d$, $i = 1, 2, \dots, M$, be the empirical distributions, i.e., histograms of $L_i$ as (Equation 2):

$$Q_i(q)_{q \in \mathrm{X}} = \frac{1}{|L_i|} \sum_{x \in L_i} \mathbb{I}_{\bar{I}(X)=q} \tag{2}$$

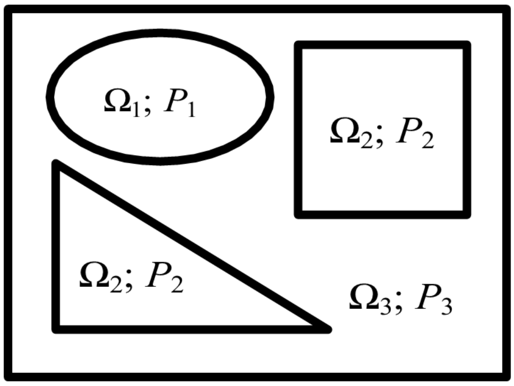


**Figure 1.** The depiction of the considered image model for M=3 region image segmentation problem. The pixel intensities $P_1, P_2, P_3$ are assumed to be different in the regions of interest $\Omega_1, \Omega_2, \Omega_3$

where $\mathbb{I}$ is an indicator function that takes value 1 when its argument is true, and $\bar{I}(X)$ denotes the rounded intensity of the pixel where rounding is performed in accordance with the edges of the histogram bin descriptor.

In the baseline method the segmentation process is implemented at the superpixel level by assuming the pixels in each superpixel are i.i.d and belong to some $\Omega_i$. This approach is illustrated in Figure 2. Notice that adopting superpixel level segmentation with the assumed image model effectively transforms the image segmentation problem into a hypothesis testing problem where robust hypothesis testing framework and the DGL is well suited. As the superpixels gather adjacent pixels into visually distinct pixel groups that adhere well to image boundaries, their differentiation with the DGL test provides a simple and effective solution to the segmentation problem.

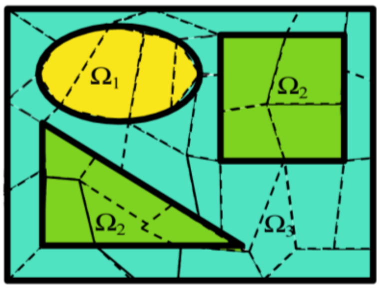


**Figure 2.** Perfoming the segmentation task at the superpixel level where dashed lines are the boundaries of the superpixels and the distinct image segments are denoted with different colors

Let $\Omega \rightarrow \{S_1, S_2, ...., S_K\}$, $K \leq N$, denote the partition of the pixel grid into superpixels. Before the application of the DGL test, M(M – 1)/2 Borel sets $A_{i,j}, A_{i,j} \in \mathcal{A}$ that have the following form (Equation 3):

$$A_{i,j} = \left\{ q : \hat{Q}_i(q) \geq \hat{Q}_j(q) \right\},\ 1 \leq i < j \leq M, \tag{3}$$

need to be calculated for all $q \in \mathcal{X}$. Then, if $\{X_1, X_2, ..., X_n\}$ are the set of pixels in $S_k$, the test decides on $D(S_k) = i$ provided that (Equation 4):

$$\max_{A \in \mathcal{A}} \left| \int_A \hat{Q}_i(A) - \mu_n(A) \right| = \min_{j=1,\ldots,M} \max_{A \in \mathcal{A}} \left| \int_A \hat{Q}_j(A) - \mu_n(A) \right| \tag{4}$$

Where (Equation 5):

$$\mu_n(A) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}_{X_i \in A} \tag{5}$$

The integrals in (4) can be calculated numerically for example using the trapz function in Matlab. This algorithm can be implemented with time complexity $O(M^2 \max\{|\mathcal{X}|, N\})$ and space complexity $O(M^3 \max\{|\mathcal{X}|, N\})$ [8]. When $|\mathcal{X}|$ is larger than $N$, the complexities are not linear in the number of pixels which is undesired for a practical segmentation algorithm. However, one can reduce the dimension of $\mathcal{X}$ and implement the algorithm on $\mathcal{X}'$ via dimensionality reduction techniques to have $|\mathcal{X}'| \leq N$. This ensures an algorithm with time complexity $O(M^2N)$ and space complexity $O(M^3N)$. In [8] the algorithm is implemented in the HSV color domain by considering only the H and S components where it is shown that the reduction in the performance might be negligible.

## 3. SPATIALLY ADAPTIVE DGL TEST-BASED IMAGE SEGMENTATION

### 3.1. Motivation

Notice that the decision rule in (4) in the baseline method aims to choose the hypothesis such that the mismatch between the nominal and the empirical histogram is minimum. Therefore, this method only takes advantage of the relative distances between the intensity histograms of the user input regions and does not consider their spatial properties. However, the spatial information of these regions can also be crucial for segmentation process. For example, if the bounding box of a region is provided by the user than the region of interest resides inside the bounding box and one should not search for an instance of that region in any other part of the image.

### 3.2. Implementation

The spatial information of the user input regions can be incorporated into the DGL by introducing spatially varying intensity distributions as in [11]. In this approach one augments the spatial dimension, i.e. the pixel grid Ω, with the alphabet of the intensity distributions, $\mathcal{X}$ , and considers a compound alphabet C = $\mathcal{X} \times \Omega$. With this alphabet the probability for the intensity of a pixel, I(X), is considered jointly with its spatial position, X, as Pr(I(X), X). Therefore, the resultant spatially varying distributions allow for discrimination both in the histogram and spatial domain. In order to adapt this method to the DGL the calculations of the

Borel sets A ∈ in (3) and implementation of the test in (4) must be performed over C instead of X. However, since the complexity of the DGL test depends on $|\mathcal{X}|$ such an approach would possibly increase the computational complexity of the baseline test. In this paper, we provide a simple alternative method based on the inherent mechanism of the DGL test that allows one to adapt it to handle spatial information as well.

The proposed approach is implemented as follows. We use the set of the labeled pixels $L_1, L_2, \ldots, L_M$ that are gathered from the vicinity of user input regions. We let (Equation 6):

$$C_j = \frac{1}{|L_i|}\sum_{x\in L_i} X \tag{6}$$

be the centroid of the pixels in $L_i$ and similarly define (Equation 7):

$$\mu(S_k) = \frac{1}{|S_k|}\sum_{x\in S_k} X \tag{7}$$

to be the centroid of the superpixel. When deciding a label for $S_k$ the baseline method tries to minimize the metric $\max_{A\epsilon \mathrm{A}} \left| \int_A \hat{Q}_j(A) - \mu_n(A) \right|$ in the color domain. This metric is the disparity between the (worst case) nominal and the empirical distribution on the support $A \in \mathcal{A}$. In an analogy, the term $|C_j - \mu(S_k)|$ is a spatial metric than can be regarded as the mismatch between the nominal location of the region Ωj and the centroid of the superpixel. In the proposed spatially adaptive DGL test we soften the decision rule based on the color metric by combining it with the proposed spatial metric and decide on D(Sk) = $i$ if (Equation 8):

$$i = \underset{j=1,2,\ldots,M}{\arg\min} \; \max_{A\epsilon A} \left| \int_A \hat{Q}_j(A) - \mu_n(A) \right| + w_j \left| C_j - \mu(S_k) \right| \tag{8}$$

where $w_j \geq 0$ are weight terms that can be adjusted to modify the relative dominance of the mismatch in the spatial or in the color domain. As $w_j \to 0$ the effect of the proposed spatial mismatch term vanishes and the proposed test becomes identical to the baseline test. Whereas, as $w_j \to \infty$ the effect of the mismatch in the color domain vanishes and the test becomes purely spatial. Therefore, we balance the two mismatches with a proper selection of the $w_j$ terms.

The improvements of the proposed method is depicted in Figure 3 for some images form the BSDS500 database where we have used the SLIC superpixels [12] with K = 500. Here, we have used the bounding boxes of image segments as user inputs and we have used the proposed algorithm by setting $w_j = 1$. Notice that in the first image the segmentation errors between the foreground and background grassy field, in the second image the errors in the background forest to the left and right of the tree, and in the last image the errors between the faces of the mother and the baby are almost eliminated with the proposed method. In the next section, we present our benchmarking results and show that similar improvements can be obtained with different types of user inputs as well.

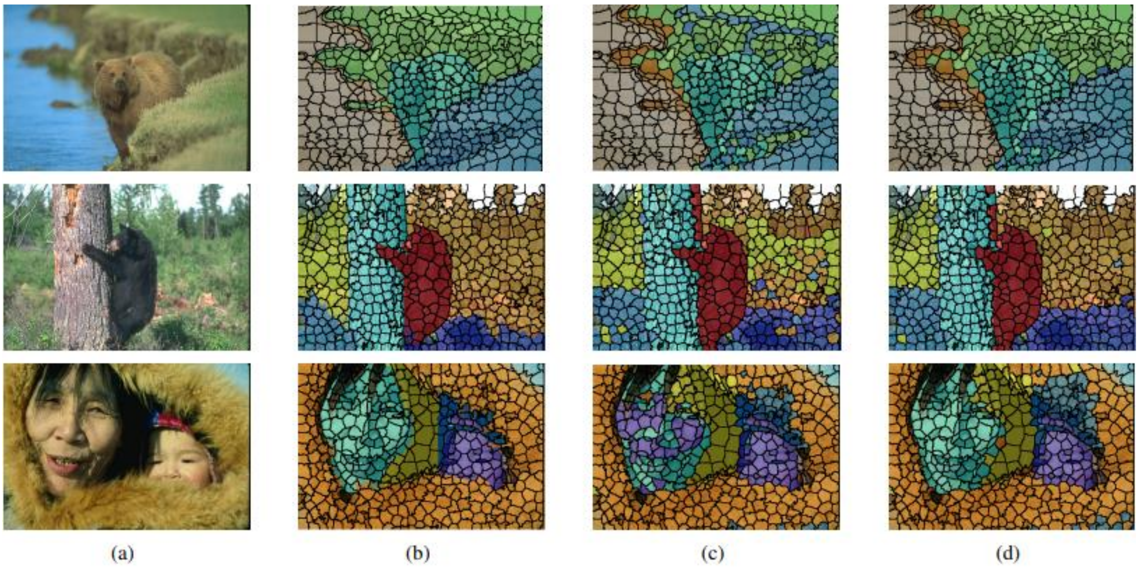

**Figure 3.** Performance comparison of the proposed method and the baseline method on some sample images from Berkeley's BSDS500 database where a) the original image, b) the ground truth, c) baseline DGL test d) spatially adaptive DGL test

### 3.3. Complexity

Notice that the proposed method only requires the calculations of equations in (6) and (7) which can be performed with complexity O(N). Therefore, the proposed method does not increase the computational complexity of the baseline test and can be implemented with time complexity $O(M^2N)$ and space complexity $O(M^3N)$.

## 4. SIMULATIONS

We have compared the performances of the proposed method and the baseline DGL test in [8]. The simulations are performed on Berkeley's BSDS500 database [9] test images. This dataset includes 200 natural RGB images where multiple ground truth annotations are provided for each image. The pixel grid has a size of 321×481 and each color intensity is encoded with 8 bits precision i.e. $|\mathcal{X}| = 256^3$. As a performance measure, we have used intersection over union (IoU) metric defined as (Equation 9):

$$\text{IoU}=\frac{\text{Area of Overlap}}{\text{Area of Union}} \tag{9}$$

where overlap and union are calculated for the considered superpixels and the ground truths. IoU is calculated by averaging it over multiple ground truths and over multiple images. Before calculating IoU, we have decreased the number of regions in each image and considered 90% of the labeled pixels, as in [8], since a majority of the annotations are Over-segmented.

We have compared the performances of both methods for all the considered user types in the baseline method in [8].

Following the same notation we let, $\text{GT}_i$ and $\text{BB}_i$ , i=1, 2,…, M, denote the ground truth masks and bounding boxes of the image segments. First, the set of labeled pixels, $\mathcal{T}_i$ are obtained from a percentage f%, f∈{25, 50, 75, 100}, of randomly selected labeled pixels from $\text{GT}_i$ and $\text{BB}_i$. These methods are respectively denoted by

$\mathrm{DGL}_{GT}^{f\%}$ and $\mathrm{DGL}_{\mathrm{BB}}^{\mathrm{f\%}}$ for the baseline test and by $\mathrm{DGLSPAT}_{GT}^{f\%}$ and $\mathrm{DGLSPAT}_{BB}^{f\%}$ in the proposed spatially adaptive DGL test. We have also considered t%, t∈{5, 10, 15}, random pixels seeds from each ground truth mask and investigated the case of p%, p∈{5, 10, 15}, randomly perturbed bounding boxes. These methods are respectively denoted by $\mathrm{DGL}_{GT}^{t\,pts}$, $\mathrm{DGL}_{GT}^{p\%pt}$ in the baseline method and by $\mathrm{DGLSPAT}_{GT}^{t\,pts}$, $\mathrm{DGLSPAT}_{GT}^{p\%pt}$ in the proposed method. While using random pixel seeds, the pixels in $L_i$ are gathered from square boxes with centers being the seed points in $\Omega_i$. The side-length of these squares are chosen to be 50 pixels as in [8]. For the case of perturbed bounding boxes, the two corner points, $(r_1, c_1)$, $(r_2, c_2)$, of the bounding boxes are uniformly translated $p\%$ to obtain $(\hat{r}_1, \hat{c}_1)$, $(\hat{r}_2, \hat{c}_2)$, respectively. Here, $\hat{r}_1$ is selected uniformly over $\mathrm{r1}-\frac{(\mathrm{r2}-\mathrm{r1})\mathrm{p}}{200}$, $\mathrm{r1}+\frac{(\mathrm{r2}-\mathrm{r1})\mathrm{p}}{200}$ the same perturbation method is applied to obtain $\hat{c}_1$, $\hat{r}_2$ and $\hat{c}_2$

Both algorithms are implemented in the HSV color space by considering only the H (hue) and S (saturation) components of the image. As in [8] we have applied dimensionality reduction by choosing $|\mathcal{X}_1|=|\mathcal{X}_2|=\sqrt{321\times 481}\approx 392$ so that the complexity of the both tests is linear complexity in the number of pixels. We have used the proposed algorithm by setting $w_j = 1$ in Eq. (8). This choice provided an accuracy improvement for the majority of the images in the BSDS500 database. We have also considered the injection of additional user inputs via a genie-aided user that relabels the mislabeled superpixels and observed the increase in the accuracy of segmentation versus the number of relabeled superpixels.

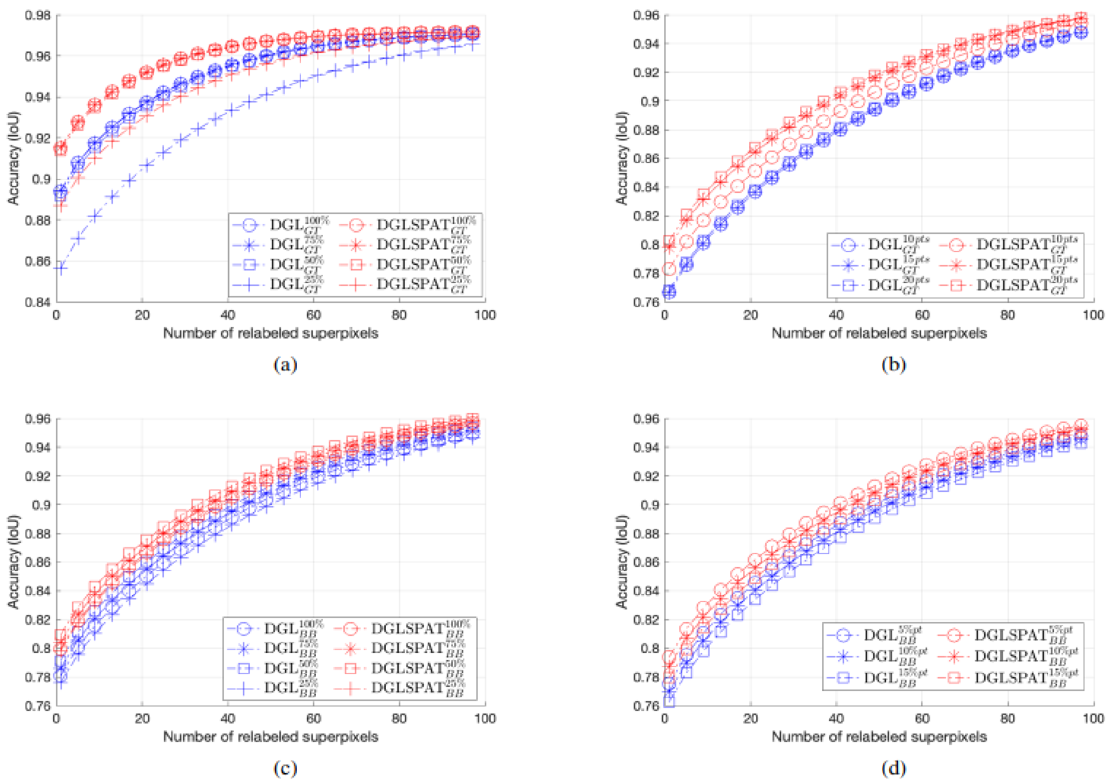


**Figure 4.** Benchmarking results of the proposed spatially adaptive DGL test and the baseline DGL test on Berkeley's BSDS500 database for different types of user inputs. In all the figures the blue curves represent the baseline method and the red curves represent the proposed method

The benchmarking results are presented in Figure 4. Here we compare $\mathrm{DGL}_{GT}^{f\%}$ and $\mathrm{DGLSPAT}_{GT}^{f\%}$ in Figure 4.a, $\mathrm{DGL}_{GT}^{t\,pts}$ and $\mathrm{DGLSPAT}_{GT}^{t\,pts}$ in Figure 4.b, $\mathrm{DGL}_{BB}^{f\%}$ and $\mathrm{DGLSPAT}_{BB}^{f\%}$ in Figure 4.c and $\mathrm{DGL}_{GT}^{p\%pt}$ and $\mathrm{DGLSPAT}_{GT}^{p\%pt}$ in Figure 4.d, respectively. From these figures we observe that the proposed spatially adaptive DGL test provides an accuracy improvement around 3% for all the considered user input types.

## 5. CONCLUSION

We have presented a spatially adaptive version of the DGL test to be used in user-assisted multilabel image segmentation problem. The proposed method offers a simple way to include the spatial information of the user inputs to the baseline, color domain based DGL test, for an improved segmentation performance. We have shown that the proposed method can be seamlessly integrated into the baseline method without increasing its complexity and it can provide performance improvements. We have also provided benchmarking results on the Berkeley's BSDS500 database and showed that an accuracy improvement of around 3% can be obtained compared the baseline method. Investigating the segmentation performance of the DGL test via spatially varying color distributions as in [11] is the topic of our upcoming work.

## 6. REFERENCES


**1.** Boykov, Y.Y., Jolly, M.P., 2001. Interactive Graph Cuts for Optimal Boundary & Region Segmentation of Objects in N-D Images, Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001, 1, 105-112.

**2.** Rother, C., Kolmogorov, V., Blake, A., 2004. Grabcut: Interactive Foreground Extraction Using Iterated Graph Cuts, ACM Transactions on Graphics, 23(3), 309–314.

**3.** Grady, L., 2006. Random Walks for Image Segmentation, IEEE Transactions on Pattern Analysis and Machine Intelligence, 28(11), 1768-1783.

**4.** Ramadan, H., Lachqar, C., Tairi, H., 2020. A Survey of Recent Interactive Image Segmentation Methods. Comp. Visual Media 6, 355–384.

**5.** Devroye, L., Gyorfi, L., Lugosi, G. A., 2002. A Note on Robust Hypothesis Testing. IEEE Trans. Inform. Theory, 48(7), 2111-2014.

**6.** Biglieri, E., Gyorfi, L., 2014. Some Remarks on Robust Binary Hypothesis Testing. IEEE Inter. Symp. on Inform. Theory, 566-570.

**7.** Afşer, H., 2021. Statistical Classification via Robust Hypothesis Testing: Non-Asymptotic and Simple Bounds. IEEE Signal Processing Letters, 28, 2112-2116.

**8.** Afşer, H., 2022. A Baseline Statistical Method for Robust User-assisted Multiple Segmentation. IEEE Signal Processing Letters, 29, 737-741.

**9.** Arbelaez, P., Maire, M., Fowlkes, C., Malik, J., 2011. Contour Detection and Hierarchical Image Segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 33(5), 898-916.

**10.** Junmo, Kim, Fisher, J.W., Yezzi, A., Cetin, M., Willsky, A.S., 2005. A Nonparametric Statistical Method for Image Segmentation Using Information Theory and Curve Evolution. IEEE Transactions on Image Processing, 14(10), 1486-1502.

**11.** Nieuwenhuis, C., Cremers, D., 2013. Spatially Varying Color Distributions for Interactive Multilabel Segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 35(5), 1234-1247.

**12.** Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Susstrunk, S., 2012. SLIC Superpixels Compared to State-of-the-Art Superpixel Methods. IEEE Transactions on Pattern Analysis and Machine Intelligence, 34(11), 2274-2282.