第 47 卷 第 9 期
2020 年 9 月

湖南大学学报（自然科学版）

Journal of Hunan University (Natural Sciences)

Vol. 47. No. 9.
September 2020

Novel Content Aware Pixel Abstraction for Image Semantic Segmentation

Mehak Maqbool Memon[1]*, Manzoor Ahmed Hashmani[2], Syed Sajjad Hussain Rizvi[3]

[1]Department of Computer and Information Sciences, Universiti Teknologi Petronas, Seri Iskandar, Malaysia,
[2]High-Performance Cloud Computing Center, Center for Research in Data Science, Department of Computer and Information Sciences, Universiti Teknologi Petronas, Seri Iskandar, Malaysia,
[3]Shaheed Zulfiqar Ali Bhutto Institute of Science & Technology, Karachi, Pakistan

Abstract: Image semantic segmentation is one of the recently researched topics due to the rise in visual deep learning-based applications. These applications work on the meaningful segments of the visual scene created by the application base. The literature identifies the current research status and highlights the existing problems of image semantic segmentation algorithms. These problems include the handling of complex images. Complex images can be of form high/low pixel intensities or dense structure regions of the image. Existing state-of-the-art deep learning algorithms fail to segment complex images semantically. For semantic segmentation of complex images, deep learning algorithms are proposed to be accompanied by pixel abstraction algorithm. The pixel abstraction algorithm creates atomic segments of the visual scene called super-pixels. Super-pixels generate feature vectors supporting the same regions. These feature vectors reduce the computational complexity to create semantic segments of the visual scene. The pixel abstraction algorithms lack functionality due to different aspects, one of which is the initial hand-crafted seed from the user to create super-pixels that do not work for all types of visual scenarios to create accurate semantic segments. The second aspect that limits pixel abstraction algorithms' functionality is the distance measure used for super-pixel (cluster) creation. The distance measures employed in existing algorithms do not capture content-aware information of visual scene; instead, end-up creating super-pixels based on Euclidean distance, which is based on straight line distance. Hence, the created pixels are distorted and irregular. For proving the flawed functionality of the existing super-pixel creation algorithm, detailed visual analysis is presented, uncovering the indicators for future research towards the development of a novel algorithm creating continuous and regular super-pixels. For creating content-aware super-pixels, the article describes an automatic super-pixel creation algorithm based on the idea of capturing image information in relevance to the content present in it. For example, we illustrate the proposed framework in detail as two modular approaches to improve the resulting super-pixels' quality. Firstly, to automate the entire process, the probability density function is proposed to initialize the cluster centers such that hand-crafted seed is not required from the user. Secondly, to retrieve fine-grained object boundaries, a novel distance measure with induced content-aware nature and complex image handling is proposed. The novel algorithm has the potential to tackle the problem of discontinuity and irregularity of retrieved segment boundaries.

Keywords: Super-Pixels, Simple Linear Iterative Clustering, Distance Measures.

用于图像语义分割的新型内容感知像素抽象

摘要：由於基於視覺深度學習的應用程序的興起，圖像語義分割是最近研究的主題之一。這些應用程序在應用程序庫創建的視覺場景的有意義的部分上工作。文獻確定了當前的研究現狀，並突出了圖像語義分割算法存在的問題。這些問題包括複雜圖像的處理。複雜圖像可以具有圖像的高/低像素強度或密集結構區域的形式。現有的最新深度學習算法無法在語義上分割複雜的圖像。對於復雜圖像的語義分割，提出了深度學習算法和像素抽象算法。像素抽象算法會創建視覺場景的原子片段，稱為超級像素。超像素生成支持相同區域的特徵向量。這些特徵向量降低了創建視覺場景語義段的計算複雜度。像素抽象算法由於不同方面而缺乏功能，其中之一是用戶最初手工製作的種子，這些種子是用戶創建的超級像素，不適用於所有類型的視覺場景以創建準確的語義段。限制像素抽象算法功能的第二個方面是用於創建超像素（群集）的距離度量。現有算法中採用的距離度量不能捕獲視覺場景的內容感知信息；相反，最終

Received (date): August 20, 2020
Corresponding Author’s name, E-mail: mehak_19001057@utp.edu.my

會根據基於直線距離的歐幾里得距離創建超像素。因此，創建的像素失真且不規則。為了證明現有超像素創建算法的功能存在缺陷，本文進行了詳細的視覺分析，揭示了開發新型連續和規則超像素算法的研究指標。為了創建可識別內容的超像素，本文介紹了一種基於捕獲與其中存在的內容相關的圖像信息的想法的自動超像素創建算法。例如，我們將提出的框架詳細說明為兩種模塊化方法，以提高所得超像素的質量。首先，為了使整個過程自動化，提出了概率密度函數來初始化聚類中心，這樣就不需要用戶手工製作的種子了。其次，為了獲取細粒度的對象邊界，提出了一種具有感應內容感知特性和復雜圖像處理的新型距離度量。該新算法具有解決檢索段邊界不連續性和不規則性的潛力。

关键词：超像素，簡單線性迭代聚類，距離測度。

Introduction

Image semantic segmentation is the process of recognizing and understanding the image data at a pixel level. Pixel-level understanding creates a raster mask of the visual scene. This raster mask represents different scene segments with the predefined class labels for each pixel [1]. The entire process of semantic image segmentation is automated due to advancements and the merger of Computer Vision (CV) and Deep Learning (DL) [2]. The recent advancements of the Deep Convolutional Neural Networks (DCNNs) have steered the application of image semantic segmentation towards more effective and dense predictive labels of the visual scene [2], [3], [4]. However, the existing techniques still suffer to create correct raster masks with object class labels for complex scenarios. The complex scenarios result in complex images having high/low pixel intensities (semi-dark images) and dense structure regions (multi-object images). If accompanied by pixel abstraction algorithms, the existing DCNNs used for semantic segmentation can result in accurate semantic segments [5], [6], [7]. Pixel abstraction algorithms reside in the early image segmentation techniques where the input is an image, and the output is regions or structures. The output regions created by pixel abstraction algorithms are called Super-pixels [5]. The accurate super-pixels generated by pixel-abstraction algorithms fed to DCCNs for further processing can generate accurate semantic segments in complex images, i.e., classify and recognize a specific object even if provided with complex images. Pixel abstraction algorithms are of two types: clustering-based and graph-based. These algorithms process the image and generate regions based on gradient information or color information [5, 8]. Clustering-based algorithms are recommended in the literature for their performance accuracy in terms of regularity, efficiency, boundary recall, and compactness [8], [9]. However, clustering-based algorithms are also limited in functionality to segment complex images. More efficient and content-aware pixel abstraction algorithms that capture fine-grained details from the complex scenes are required because inaccurate super-pixels can severely affect the final segmented results of DCNNs [5]. The desired properties of super-pixel algorithms include adherence of super-pixels for object boundaries, regularity, i.e., how close the created super-pixel is to the actual image content and efficiency in terms of reduced computational complexity as these algorithms are used as preprocessing step [8, 9]. Similarly, these properties apply to complex images. None of the super-pixel algorithms have been explicitly analyzed for complex images to the best of our knowledge. In this work, we have mainly focused on the problem of segmentation for complex images. We have proposed a novel pixel abstraction algorithm with integral content-aware nature for segmenting complex images. The content-aware nature is integrated in terms of using the relevant distance measures for the creation of super-pixels. The existing algorithms result in distortions in object boundaries and irrelevance to the image content. These problems lead to the formulation of research objectives to improve the results of segmented regions.

1. To investigate the factors that dominate the performance of pixel abstraction algorithms for segmentation of images

2. To identify mechanisms to automate the initialization process of super-pixels creation and label creation for semi-dark and multi-object images.

3. To propose an effective content-aware strategy for the creation of super-pixel in complex images.

The paper is organized as follows: Section 2 describes related work of the state-of-the-art pixel abstraction algorithms and analyzes their limitation, section 3 presents a proposal of a novel pixel abstraction algorithm, section 4 presents the tested results of the most recent super-pixel algorithm to prove the functionality flaw in existing approach, and finally, section 5 concludes by discussing contribution and future directions.

1 Related Work

With the continuous advancements in Machine Learning (ML) and Deep Learning (DL) technology, the entire process of segmenting and recognizing objects is automated. However, this automation does not support the segmentation schemes for complex images. Researchers are now considering the merger of traditional image semantic segmentation techniques with innovative deep learning techniques for accurate semantic segmentation. The traditional segmentation/pixel abstraction techniques are a crude method for creating atomic regions based on features like appearance and homogeneity because of human intervention in the process [3]. The atomic regions support the feature-based segmentation in relevance to the region-based features than the entire local window. The challenge with pixel abstraction algorithms, i.e., super-pixel algorithms, is to reduce the complexity and fetch the concrete homogenous information recognizing the intensity boundaries of the objects present in the images encompassing all types of visual scenarios. Super-pixel algorithms are divided into different categories based on different criteria. According to [8], [10], super-pixel algorithms are categorized into two categories, namely constrained and unconstrained algorithms based on the consideration of the function of an object concerning the compactness or not. Another study [11] categorizes super-pixel algorithms based on the super-pixel generation and names the categories as bottom-up or top-down approaches. The study [9] divides these algorithms as clustering-based and gradient-ascent based algorithms. The power of super-pixel lies in its representation of the scene's accurate information lies in its stability. The existing algorithms are designed to tackle different purposes, so they lack in one or other aspects, including control over super-pixel size, number, and compactness. In the presented study, we have categorized super-pixel algorithms in two types, namely: Graph-based and clustering-based algorithms. Graph-based algorithms represent image considering nodes as pixels and minimize the cost function defined on the graph. Clustering-based algorithms groups image regions to create homogenous clusters and iteratively refines them till convergence criteria are met [12].

1.1. Graph-based Algorithms

These algorithms represent the image as graph $G=\{V,E\}$, V being set of pixels/regions, and E representing edges connecting pixels/ regions to reflect the similarity. After representation, some optimization function is employed to create segments representing the grouped local features [13]. Normalized cuts [14] segment image regions by splitting the affinity graph. The algorithm tries to minimize global image information by recursively partitioning a given graph using contour and texture cues. The global cost function defined for edges is minimized in every iteration. The process is computationally expensive, and the performance for complex images is not analyzed. Felzenszwalb and Huttenlocher (GS04) [15] perform agglomerative clustering of pixel nodes, using the shortest spanning tree of pixels. The algorithm is faster than normalized cuts but provides no control over the number of super-pixels. Super-pixel lattice [16] generates optimal vertical/horizontal paths that cut the image. The horizontal or vertical paths are used as strips of pixels over the image, which provides control over the size, number, and super-pixel compactness. However, still dependent on the precomputed boundary maps. Pseudo-Boolean super-pixels [17] generate segmentation results as a multi-label problem. Here half overlapping horizontal strips are used. Each pixel might get assigned one of the latent strips. The strip decides the class or label. The super-pixel creation speed is independent of the number of pixels being generated, which is one of the problems present in other algorithms. The entropy rate super-pixel [18] handles super-pixel creation as a maximization problem on the graph. The image graph denotes pairwise similarities. The objective is selecting subset edges from the images graph resulting in K-connected graphs (K is the number of super-pixels). The working mechanics makes it highly dependent on the number of super-pixels to be generated, supplied in hand-crafted seed from the user. All these graph-based algorithms have their pros and cons, but one aspect is common: they have not been analyzed for the segmentation of complex images.

1.2. Clustering-based Algorithms

Clustering-based algorithms represent an image as a feature vector and apply feature space analysis methods. These methods include parametric and non-parametric approaches to create clusters in the feature space. The feature space is created on the local features, which drives these methods to have complex dependency over local statistics, which serve as the basis to segment image into a large number of small regions [13]. Watershed [19] performs gradient ascent from local minima in the image feature vector space to obtain watersheds, i.e., lines that separate catchment basins. This algorithm's short version applies graph-cuts to the graph build based on priorly created super-pixels by watershed. This is also named as pixel queuing. Turbopixels [20] dilates the number of seeds iteratively in image feature space using level-set based geometric information. The method relies on the local image gradient and evenly distributes super-pixels over the image plane. The method is constrained in its functionality to result in uniform size compactness and adherence to object boundaries, and it has a slow running time [9]. Meanshift [21] is one of the modes seeking algorithm. The algorithm recursively moves the kernel to the smoothed centroid for a data point in feature space. The input kernel can vary the size of the created super-pixels. The algorithm is limited in functionality by providing no control over the number, size, or super-pixel compactness. Quickshift [22] is another mode-seeking algorithm that tries to move each point in feature space to the nearest neighbor, increasing the parzen's density estimates. The algorithm results in numerous errors [5]. SLIC [5] identifies the maximum possible distance between two colors of feature space. The spatial distance in the xy plane depends on the size of the image. The algorithm uses a normalized form of Euclidean distance. The algorithm has unsatisfactory performance for complex color images [5],[23], along with constrained boundary recall [24] and poor performance for noisy images [25]. Structured Sensitive Super-pixels [26] employ geometric flow to compute the distance between pixels. Over-segmentation is adjusted using energy function, which inherently integrates color homogeneity, structure density, and compactness. The algorithm uses density function for analysis of similarity among pixels. Manifold SLIC [12] employs simplicity of SLIC and the content-aware nature of structure sensitive super-pixel. The proposed algorithm focuses on capturing boundaries in relevance to the object boundaries present in the scene. All the discussed clustering-based methods have different features to offer; however, none of the mentioned algorithms are analyzed for complex images. These super-pixel algorithms are probable to speed up the entire process of semantic segmentation via DL techniques.

1.3. Issue of inadequate Pixel Abstraction causing inefficient Image Semantic Segmentation

According to the conducted study, clustering-based algorithms have been the most promising option for creating super-pixels. The super-pixels result in the enhanced boundaries via clustering principle, i.e., regions or segments based on the color proximity and the distance. The created super-pixels result in pixel abstraction, which is then used by deep neural networks to create pixel-level masks of the image. The concerning issue is improper creation of the super-pixels, which generates improper pixel-level masks of the segmented image. The reason for improper results is a significant functionality flaw of super-pixel creation algorithms. For the conducted research, clustering-based algorithms are focused mainly on Simple Linear Iterative Clustering (SLIC). The SLIC algorithm used for super-pixel creation is fast and straightforward in execution. However, the distance measure used for the pixel cluster is inadequate in terms of no relevance to the actual content present in the image. Instead, the algorithm just blindly finds Euclidean distance, which is a straight line distance. At the same time, it is impossible to have uniform super-pixels for all the image regions. This functionality flaw results in the distorted super-pixels, and finally, the masks created by deep neural networks are also erroneous.

2 Methodology

Accurate segmentation as the additional information for further processing a complex scene by a Deep Convolutional Neural Network (DCNN) is crucial. The current state-of-the-art DCNNs work on per-pixel primitives to create semantic segments of images. The presented study proposed locally grouped primitives of the scene for further processing by the DCNNs. The proposed framework divides the entire digital enhancement preprocessing stage into two different modules. The first module handles the Image labeling process so that normal images are dropped out of the process immediately and passed to DCNN for further processing. This module makes sure that the images get the proper label of a semi-dark or multi-object image by using the thresholding process on the feature space. The thresholding process includes comparing two different threshold values with their relevant measure, i.e., pixel values and density estimate of the given input image. As the image consists of two-dimensional integer arrays representing individual components of the image, the pixel intensities are used to identify the darker or brighter image based on the threshold value for identifying semi-dark images. Simultaneously, for identifying multi-object images, the image's density estimates are employed because the density of image contents differs in different parts of the image, inferencing multi-object images by analyzing similarity measure among pixel values. After that, image processing techniques are applied to process the complex input images. Finally, the enhanced and labeled image is passed to the subsequent module. The second module maps this enhanced input image on a low dimensional manifold to reduce computational power consumption.

Manifolds create empirical mappings to retain topological properties of the image, making the entire process of learning accessible to handle in capacity limited Neural Networks. Further creates Centroidal Voronoi Tessellation (CVT) over the manifold, now having mass centroids. A Voronoi diagram partition the manifold with 'n' points into a convex polygon having only one generating point. Every point of the polygon is supposed to be at the least distance from its generating point. After that, initial cluster points are identified based on Parzen's Density Estimates of the CVT this way. The hand-crafted seed is not required from the user. Parzen's Density function is one of the non-parametric approaches of estimating probability density function, depicting class-conditional densities without any prerequisite knowledge of primary distribution. Finally, the iterative clustering process is initialized based on the image label with relevant distance measure for each pixel's cluster assignment. The iterative process is continued until the clusters become stable satisfy the condition for residual error.

2.1 Novel Pixel Abstraction Algorithm

The pixel abstraction takes M*N enhanced pixel grid and maps it on a manifold, which is a topological space resembling Euclidean space near each point. The 2D manifold is created in such a way that the self-intersection of objects does not take place. This is a process of hypothesizing the real scenarios' high dimensional data, over manifolds of lower dimension embedded in high-dimensional space to reduce computational resource usage. After that, Voronoi regions are created for the manifolds such that every point in high-dimensional space of manifolds is closest to generator w.r.t the generator. All the generators are supposed to be connected through Voronoi edges (line/half-line segment). The proposed abstraction module chooses to create Centroidal Voronoi Tessellations (CVTs), which has an additional mass centroid constraint as a Voronoi generator for the corresponding Voronoi region. For a reason, CVTs are more organized than the normal Voronoi tessellations. With its novel architecture, the proposed system removes the need for seed requirement from the user for cluster initialization and automates the entire process. The automation is made possible by applying Parzen's Density Estimates over feature space to find initial points for cluster creation. After cluster initialization, the subsequent pixel assignment process to a cluster is based on the relevant distance measure identified by the label created by the previously used image labeling module. The distance measure for cluster assignment is supposed to be different for multi-object and semi-dark images, discussed in detail in the upcoming section. Finally, clusters' stability is iteratively checked until the created clusters become stable, and the residual error condition is satisfied.

2.1.1 Distance Measure

The Euclidean distance used in the SLIC algorithm for cluster pixel assignment has some significant functionality flaws, resulting in irrelevance to the image contents, leading to discontinuous and irregular segments [26]. The witnessed functionality flaw is the result of constant distance measure, which remains the same regardless of whether there is a path along which the appearance transits smoothly. For avoiding this, Geodesic distance is proposed in which distance increases for the image point if the local density increases and vice versa. Usually, Geodesic distance has the shortest path; however, perturbation of the geodesic curve increases its length. Geodesic distance results in minimum distortions by calculating the shortest path between two points of a mesh graph where the length of an edge is associated with weights. The calculation of geodesic distance is based on the solution of the eikonal equation [27]. This way, content sensitivity is integrated into the novel pixel abstraction module. The geodesic distance for a color image is defined as

$$I(x\colon \Psi \to R^d)$$

where d=3 for color images

Considering binary region for simplicity,

$$\Psi \to R^2$$

is supposed to be continuous. Given a binary mask 'M'

$$M(x)\epsilon\{0,1\}, \forall\ x\epsilon\Psi$$

is associated with a seed (object) region.

$$\Omega = x\epsilon\Psi\colon \mathrm{M(x)} = 0$$

and the unsigned geodesic distance transform is defined as

$$D_0(\bullet; M\ \nabla I)$$

assigns each pixel 'x' its geodesic distance from Ω, and geodesic distance is defined as

$$D_0(x; M, \nabla I\ ) = \min_{\{x'\,|M(x')=0\}} d(x, x')$$

with

$$d(a,b) = \inf_{\Gamma\epsilon \mathrm{P}_{a,b}} \int_0^{\Gamma} \sqrt{1 + \gamma^2(\nabla I(s) \bullet \Gamma'(s))^2}\, ds,$$

where P(a,b) is set of all possible differentiable paths in *Ψ* between the points *a* and *b*,
Γ(s): R→R^2 indicates a path parameterized by arc length. The spatial derivative

$$\Gamma'(s) = d\Gamma(s)/ds$$

is the unit vector tangent to the direction of the path, dot product ensures maximum influence of gradient ∇I when it is parallel to path Γ, and γ is the geodesic factor weighing the contribution of image gradient versus spatial distance. Fig 1. shows a more intuitive elaboration of the entire concept of how Euclidean distance result in distortions and geodesic distance capture content-aware information of the image.

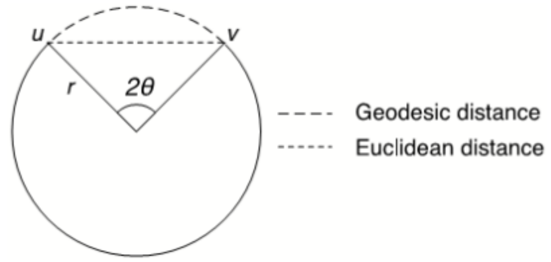


Fig.1 Difference between geodesic distance and Euclidean distance [27]

Moreover, for accurate segmentation of semi-dark images, geodesic distance is proposed to accompany the city block distance. It finds distance from a set of all black pixels targeting the dark pixel intensities in the image, employing four connected neighbors. The value of the distance always results in values zero and onwards. For similar points, the resultant is zero and higher values for a point which have less similarity. Again, for simplicity, consider a binary image. City blocks distance each point i,j of an image from a set of all black pixels and is defined as

$$B = \{(i,j): a_{ij} = 1\}$$

and

$$d_{i,j} = \min_{(x,y)\epsilon B} \{|i-x| + |j-y|\} \ 0 \le i,j \le n-1$$

Finally, for multi-object images, geodesic distance is proposed to accompany chessboard distance. It finds maximum differences between pixels targeting dark pixel intensities, employing eight connected neighbors targeting small object boundaries. For an image having arbitrary points 'i' and 'j' with coordinates (x1, y1) and (x2, y2), chessboard distance is defined as:

$$d_{i,j} = \max(|x_2 - x_1|, |y_2 - y_1|)$$

The novel solution is likely to overcome segment discontinuity and irrelevance for the present image content. Fig.2. shows the proposed framework's details for accurate semantic segmentation of complex images followed by the algorithm steps, which gives detailed insights into the proposed framework. The green color in the Fig.2. shows the significant contribution of the proposed framework.

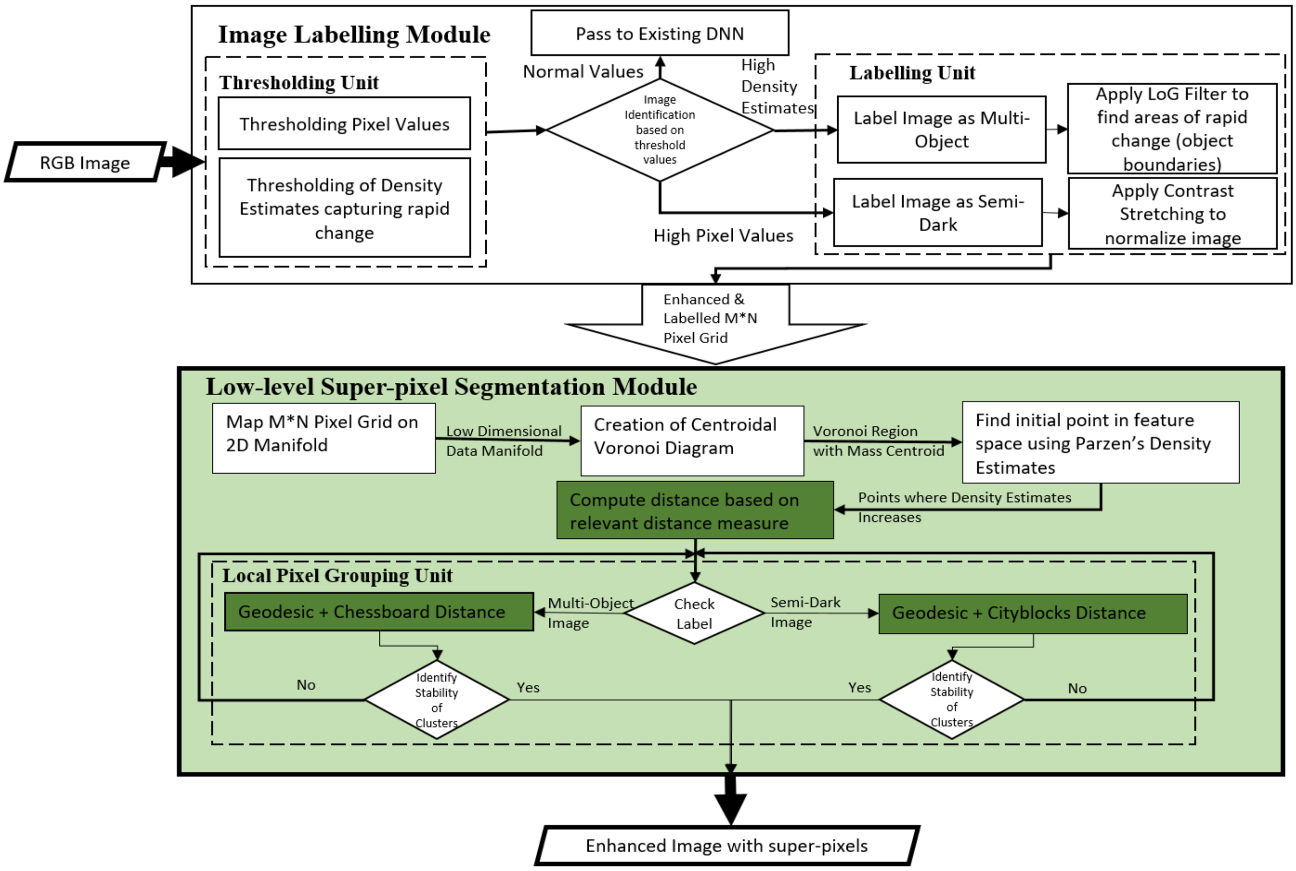


Fig.2 Detailed Pixel Abstraction Framework

The framework initializes with the RGB image having M*N Pixels. The input image is passed to the thresholding unit in the Image labeling module, which is the first module, where two different features of an image are analyzed for further processing. Suppose pixel intensities are identified to be greater than the threshold. In that case, the labeling unit labels the image as a semi-dark image and applies contrast stretching to normalize the image contents. Suppose the density estimates surpass the threshold fixed for intensities' density values. In that case, the labeling unit labels the image as a multi-object image and applies the LOG filter to identify immediate areas of change relevant to the object boundaries. Whereas, if the image is identified as a normal image based on the thresholding values, it is passed to the existing DNNs. The enhanced and labeled pixel grid is passed to the low-level super-pixel segmentation module. The pixel grid overlays on a 2D manifold, and the Centroidal Voronoi diagram having mass centroids is created for the pixel grid retaining its topological contents. The initial points for the iterative clustering process are identified based on Parzen's Density Estimates. The iterative process uses relevant distance measures based on the previously created label, i.e., semi-dark or multi-object. The stability of the relevant clusters is checked iteratively until stable clusters in the form super-pixels are created. Finally, an enhanced image with stable super-pixels is ready to be passed to the DNNs for further Semantic Segmentation.

Algorithm

Input: An image I of M*N pixels, the threshold for pixel values, the threshold for density estimates, and the convergence threshold ε.
Output: K Super-pixels of similar sizes.
STEP 1: Define RGB image as CIELAB color map
STEP 2: Identify image labels based on pixel values and density estimates
If Pixel value > Threshold
    Set label as Semi-dark
    Apply Contrast Stretching

$$I_{out} = (I_{in} - C)\left(\frac{b-a}{d-c}\right) + a$$

Where *(a,b)* are lower and upper limit of image type and its relevant data value (such as for RGB 8-bit, the limits are 0-255)
(c,d) are the lowest and highest pixel values in the provided image.
Elseif Density Estimates >Threshold
    Set label as Multi-object
    Apply LoG filter
    LoG Function centered on zero & Gaussian

Standard deviation δ has the form

$$LoG(x,y) = -\frac{1}{\pi\delta^4}[1 - \frac{x^2+y^2}{2\delta^2}]e^{\frac{x^2+y^2}{2\delta^2}}$$

Where

$$\delta = 1.4$$

Else Pixel value & Density Estimates ≈ Threshold
Set label as Normal
Pass to existing DCNN
End if
STEP 3: Map enhanced image (M*N pixel grid) on the 2D manifold
STEP 4: Create CVT
Given a set of Voronoi regions,

$$\{V_i\}_{i=1}^{k}$$

the mass centroid 'Ci' over a region with probability density p(y) and y is the vector in feature space defined as

$$C_i = \frac{\int V_i\, y\, p(y) dy}{\int V_i\, p(y)\, dy}$$

STEP 5: Initialize Cluster centers in feature space using Parzen's Density Estimates
STEP 6: Check the label
If label = Semi-dark
For each cluster center, do
Assign the best matching pixels around the center based on distance measure i.e.
Geodesic + City Blocks Distance
End For
Elseif label = Multi-object
For each cluster center, do
Assign the best matching pixels around the center based on distance measure i.e.
Geodesic + Chessboard Distance
End For
End if
STEP 7: Compute new clusters until stable and residual error ε<Threshold
STEP 8: Enforce Connectivity

3 Evaluation Criteria

The segmentation task is complex; it requires an adequately followed protocol for obtaining images and their relevant ground truth segmentations. Most of the time, only one ground truth segmentation, and researchers tend to work with multiple ground truth segmentations. In the study [8], the researchers attempted to work with five different ground truth segmentations per image.

Similarly, the evaluation criteria have also been evolving with the addition of new imagery manipulation methods. The evaluation criteria for super-pixels include coherence, compactness, and efficiency. In the literature, there have been many performance metrics used for the segmentation analysis. According to the study [8], these performance metrics are divided into three categories: segmentation quality evaluation, super-pixel quality, and computing efficiency. Segmentation quality measures access segmentation results over the adherence property for the boundaries and the pixel variations to generate the final segmentation. In this category, four different quality metrics can be used. *Precision recall* [28] identifies the boundary detection and segmentation as evaluative criteria by measuring the precision-recall curve. *Variation of information* [29] calculates the segmentations' averaged distance as the average conditional entropy based on the mutual segmentation information. *Probabilistic Rand Index* [28] checks the compatibility of cluster assignments between a pair of pixels assigned the same clusters. *Segmentation Covering* [8] identifies the intersection of the two segmentations. The second category is super-pixel quality performance, metrics residing in this category access super-pixels to retain image information as possible through coherence, compactness, and regularity. In this category, six different metrics can be used. *Under-segmentation error* [30] checks leakage of a super-pixel with the ground-truth segment. *Sum-of-squared error* [31] sums up the squared differences between each cluster's pixels and the cluster mean identifying variation in a cluster. *Achievable segmentation accuracy* assigns labels in relevance with the ground truth label, computed by the maximum fraction of correctly labeled pixels in the ground-truth. *Compactness* [32] identifies how well regular boundaries can be presented. *Explained Variation* identifies how well the color variation of the image is captured by the super-pixel, i.e., the difference between original pixels and super-pixels. *The regularity index* identifies the ratio of the area of the super-pixel and expected super-pixel area. The proposed super-pixel algorithm is expected to be evaluated using the mentioned performance metrics to identify further performance improvement to create super-pixels in complex images.

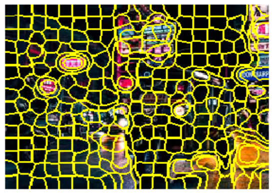
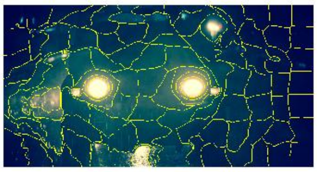
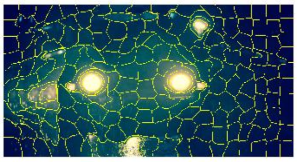
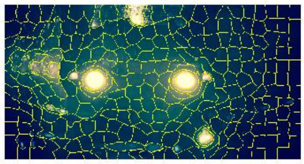
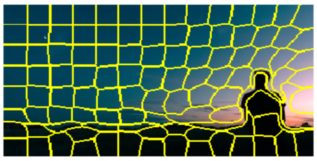
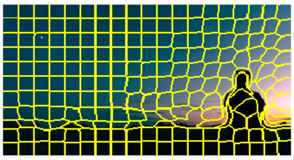
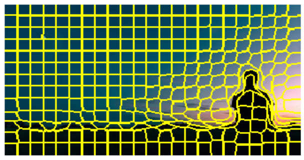
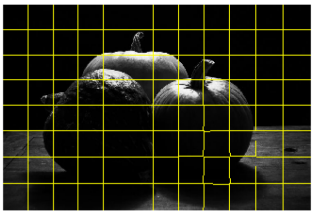
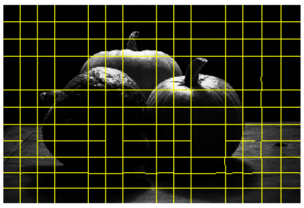
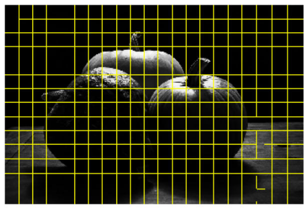
4 Experimental Deductions

The proposed novel algorithm is based on the SLIC [5] super-pixel algorithm. There are some major functionality flaws with the algorithm for the segmentation of semi-dark and multi-object images. For proving the flaws in the existing process flow, experiments are conducted implementing the SLIC algorithm for complex images. SLIC requires an initial seed to get the algorithm started, and the segment size is to be mentioned by the user. Segment size identification is a very crucial step for the later phases of the entire super-pixel creation scenario. Such size should be chosen so that the further process does not consume much of the computational resources while creating redundant segments. It should fetch at all the required object boundaries creating accurate super-pixels. The experimental study uncovering SLIC operating features' functionality flaws is conducted by using python programming and the skimage package's implementation of the SLIC algorithm.

The experimental protocol followed includes the variant size of input images having different complex scenarios, including multi-object and semi-dark images. The analysis's scope is focused on identifying the functional constraints of operating, for handling RGB images of a different resolution targeting the semi-dark and multi-object images. The purpose of using variant resolution images is to determine the effect of image resolution on the resulting super-pixels in terms of distortions.

Table 1. shows the experimental analysis of SLIC super-pixels containing the original image with the processed segments segment size 100, 200, 300.

Table 1 Experimental analysis depicting complex algorithm initialization and content irrelevant super-pixels

Case	*Original Images of variant sizes*	*Processed Image with segment size 100*	*Processed Image with segment size 200*	*Processed Image with segment size 300*
1				
2				
3				
4				
5				
6				

7				
8				

All the cases presented in Table 1 show that segment sizes significantly impact creating final super-pixel segmentation results. Cases 1-4 present semi-dark images and their relevant results. For case 1, the generated super-pixels are somewhat accurate in terms of boundary recall. However, with the increased segment size, a lot of redundant super-pixels are witnessed. Moreover, for some of the image portions, the super-pixels generated are irrelevant to the actual content. For case 2, semi-dark image poor boundary recall is witnessed, and this poor boundary recall does not change with increasing the segment sizes. For case 3, relatively better performance is witnessed in boundary recall, compactness, and other performance measures. However, segment size remains a crucial feature to select for proper operational functionality. In contrast, case 4 presents the worst-case scenario where the algorithm abruptly creates square super-pixels even in circular object boundaries depicting the restricted functionality while dealing with semi-dark images. Cases 5-8 present multi-object images. For segment size, 300 redundant super-pixels are generated, whereas for segment size 100, most of the significant object boundaries are skipped. For cases 5 and 8, a public place's crowded image is taken for super-pixel creation, and for different segment sizes, different performance is witnessed. However, one of the common patterns witnessed is poor boundary recall for all the segment sizes and somewhat irrelevant and distorted super-pixels. For case 6, the countryside image of crowded nature in many boats present in the scene is taken. Moreover, the image has less resolution as compared to other images used for the tests. The performance does not seem to change or get better or worse results as the computational power required to process such images is less, but no evident effect is seen. For case 7, an image crowded with bicycles is taken, and there are many circular objects in the scene. The performance is unsatisfactory in terms of capturing the content relevant information from the scene. These results are directed towards the need for a novel pixel abstraction algorithm for the segmentation of complex images.

5 Conclusion and Future Work

The presented research formulates an automated pixel abstraction framework that fetches fine-grained details from the complex input images. The proposed framework makes sure none of the additional computational resources are consumed to process normal images with a novel framework as such images can be accurately processed and segmented by the existing solutions. The proposed algorithm extends Manifold SLIC [12] by adding content-aware nature to the segmentation of complex (semi-dark, multi-object) images. The enhancement is in the form of identification of normal or complex images based on pixel intensities and density estimates of pixels. Moreover, two different distance measures for cluster assignments of pixel values of complex images are presented.

The future work includes the identification of threshold values for label creation of semi-dark and multi-object images. After that, the proposed framework would be implemented and analyzed for performance comparison with the state-of-the-art super-pixel algorithms.

References

[1] MARMANIS D., SCHINDLER K., WEGNER, J.D., and GALLIANI, S. Classification with an edge: Improving semantic image segmentation with boundary detection. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2016, 135: 158-172. doi: 10.1016/j.isprsjprs.2017.11.009

[2] ZHAO Hengshuang., QI Xiaojuan, SHEN Xiaoyong, SHI Jianping and JIA Jiaya. ICNet for real-time semantic segmentation on high-resolution images // *Proceedings of the 15th European Conference on Computer Vision, Munich, Germany, 8-14 Sept 2018*, pp. 405-420.

[3] SONG W., ZHENG N., ZHENG R., ZHAO X.-B. and WANG, A. Digital image semantic segmentation algorithms: A survey. *Journal of Information Hiding*

and Multimedia Signal Processing, 2019, 10(1):196-211.

[4] JAMEEL S. M., HASHMANI M., ALHUSSAIN, A. H. and BUDIMAN A. A fully adaptive image classification approach for industrial revolution 4.0. In: SAEED F., GAZEM N., MOHAMMED F., BUSALIM A. (eds) *Recent Trends in Data Science and Soft Computing. IRICT 2018. Advances in Intelligent Systems and Computing*, vol 843. Springer, Cham. https://doi.org/10.1007/978-3-319-99007-1_30

[5] ACHANTA, R., SHAJI A., SMITH K., LUCCHI A., FUA P. and SÜSSTRUNK S. SLIC superpixels, EPFL Technical Report 149300, June 2010.

[6] LI Yin, SUN Jian, CHI-KEUNG Tang and HEUNG-YEUNG Shum. Lazy snapping. *ACM Transactions on Graphics*, 2004, 23(3): 303-308. doi: 10.1145/1015706.1015719

[7] FULKERSON B., VEDALDI A. and SOATTO, S. Class segmentation and object localization with superpixel neighborhoods // *Proceedings of 2009 IEEE 12th international Conference on computer Vision, Kyoto, Japan, 29 Sept.-2 Oct. 2009*, pp. 670-677: doi:10.1109/ICCV.2009.5459175

[8] WANG M., LIU X., GAO Y., MA X. and SOOMRO I. C. Superpixel segmentation: A benchmark, *Signal Processing Image Communication*, 2017, 56:28-39

[9] ACHANTA R., SHAJI A., SMITH K., LUCCI A., FUA P. and SUSSTRUNK S. SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE Transactions on Software Engineering*, 2012, 34(11): 2274-2282, doi: 0.1109/TPAMI.2012.120.

[10] SHEN J., DU Y., WANG W., and LI X. Lazy random walks for superpixel segmentation. *IEEE Transactions on Software Engineering*, 2014, 23(4): 1451-1462. doi: 10.1109/TIP.2014.2302892

[11] XU L., WANG Z., and ZENG L. Saliency-based superpixels. *Signal Image and Video Processing*, 2014, 8(1): 181-190.

[12] LIU Yong-Jin, YU Cheng-Chi, YU Min-Jing and HE Ying. Manifold SLIC: A fast method to compute content-sensitive superpixels // *Proceedings of the I 2016 IEEE Conference on Computer Vision and Pattern Recognition*, *Las Vegas, Nevada, USA, 27-30 June, 2016*.pp.651-659. doi: 10.1109/CVPR.2016.77

[13] ZHU Hongyuan, MENG Fanman, CAI Jianfei and LU Shijian. Beyond pixels: A comprehensive survey from bottom-up to semantic image segmentation and cosegmentation. *Journal of Visual Communication and Image Representation*, 2015, 34:12-27 doi: 10.1016/j.jvcir.2015.10.012.

[14] SHI J., and MALIK J. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2000, 22(8): 888-905. doi: 10.1109/34.868688

[15] FELZENSZWALB P. F. and HUTTENLOCHER D.P. Efficient graph-based image segmentation, *International Journal of Computer Vision*, 2004, 59(2):167-181.

[16] MOORE A. P., PRINCE S. J. D., WARRELL J., MOHAMMED U. and JONES G. Superpixel lattices// *Proceedings of 2008 IEEE* conference *on computer vision and pattern recognition, Anchorage, Alaska, USA, 23-28 June 2008*, pp. 1-8. doi: 10.1109/CVPR.2008.4587471.

[17] ZHANG Y., HARTLEY R., MASHFORD J., and BURN S. Superpixels via pseudo-boolean optimization. // *Proceedings of 2011 IEEE International Conference on Computer Vision, Barcelona, Spain, 6-13 November, 2011*, pp. 1387-1394. doi:10.1109/ICCV.2011.6126393

[18] LIU M.-Y., TUZEL O., RAMALINGAM S., and CHELLAPPA R. Entropy rate superpixel segmentation// *Proceedings of 2011 IEEE Conference on Computer Vision and Pattern Recognition*, *Colorado Springs, USA, 21-23 June, 2011*, pp. 2097-2104.

[19] VINCENT L. and SOILLE, P. Watersheds in digital spaces: an efficient algorithm based on immersion simulations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1991, 13(6):583-598.

[20] LEVINSHTEIN A., STERE A., KUTULAKOS K., FLEET D., DICKINSON S. and SIDDIQI, K. TurboPixels: Fast superpixels using geometric flows. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2009, 31(12): 2290-2297. doi: 10.1109/TPAMI.2009.96

[21] COMANICIU D. and PETER M. Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2002, 25(5):603-619.

[22] VEDALDI A. and SOATTO S. Quick shift and kernel methods for mode seeking // *Proceedings of 10th European Conference on Computer Vision, Marseille, France, 12-18 October 2008*, pp. 705-718.

[23] FANG Zhuoqun, YU, Xiaosheng WU, Chengdong , CHEN Dongyue and JIA Tong. Superpixel Segmentation Using Weighted Coplanar Feature Clustering on RGBD Images. *Applied Sciences*, 2018, 8(6): 902. doi: 10.3390/app8060902.

[24] CONG Jinyu, WEI Benzheng, YIN Yilong, XI Xiaoming, and ZHENG Yuanjie. Performance evaluation of simple linear iterative clustering algorithm on medical image processing. *Bio-medical materials and engineering*. 2014, 24(6): 3231-3238. 10.3233/BME-141145.

[25] QIN Fachao, GUO Jiming and LANG Fengkai. Superpixel segmentation for polarimetric SAR imagery using local iterative clustering. *Geoscience and Remote Sensing Letters, IEEE*, 2014, 12(1): 13-17. doi: 10.1109/LGRS.2014.2322960.

[26] WANG Peng, ZENG Gang, GAN Rui, WANG Jingdong and ZHA Hongbin. Structure-Sensitive Superpixels via Geodesic Distance. *International Journal of Computer Vision*, 2013, 103(10: 1-21. doi: 10.1007/s11263-012-0588-6

[27] RAMU B. and THIRUPATHI D. A new methodology for diagnosis of appendicitis using sonographic images in image mining //*Proceedings of the 2nd International Conference on Computational Science, Engineering and Information Technology, Coimbatore, India, 26- 28 October,2012*, pp. 717-721.

[28] REN X. and MALIK J. Learning a classification model for segmentation //*Proceedings of the 9th IEEE International Conference on Computer Vision, St.*

Petersburg, Russia, 16-18 July, 2013, 1:10-17 doi:10.1109/ICCV.2003.1238308.

[29] ARBELAEZ P., MAIRE M., FOWLKES C., and MALIK J. Contour detection and hierarchical image segmentation *IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011*, 33(5):898-916. doi: 10.1109/TPAMI.2010.161.

[30] XU C. and CORSO J. J. Evaluation of super-voxel methods for early video processing //*Proceedings of 2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, Rhode Island, USA, 16-21 June, 2012* pp. 1202-1209, doi: 10.1109/CVPR.2012.6247802.

[31] STUTZ D. Superpixel segmentation: An evaluation. In: GALL P. GEHLER B. and LEIBE B. (Eds.) *Lecture Notes in Computer Science*, 2015, 9358:555-562. doi:10.1007/978-3-319-24947-6_46

[32] RAND W. Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical Association*, 1971, 66(336): 846-850. doi: 10.2307/2284239. 66

参考文:

[1] MARMANIS D.，SCHINDLER K.，WEGNER，J. D. 和 GALLIANI，S.帶有邊緣的分類：通過邊界檢測改進語義圖像分割。 ISPRS 攝影測量與遙感學報，2016，135：158-172。 doi：10.1016 / j.isprsjprs.2017.11.009

[2] Zhao H 榦霜., Q Ixia O 卷, she NX IA O 詠, shi J Ian 平安 DJ i AJ IA 呀. ICNet //第 15 屆歐洲計算機視覺會議論文集，德國慕尼黑，2018 年 9 月 8 日至 14 日，第 405-420 頁。

[3] SONG W., ZHENG N., ZHENG R., ZHAO X.-B. 和 WANG, A.數字圖像語義分割算法：調查。信息隱藏與多媒體信號處理學報，2019，10（1）：196-211。

[4] JAMEEL S.M.，HASHMANI M.，ALHUSSAIN，A.H. 和 BUDIMANA。工業革命的完全自適應圖像分類方法 4.0 在：SAEED F., GAZEM N., MOHAMMED F., BUSALIM A.（eds）最新趨勢數據科學和軟計算。 IRICT2018。《智能係統和計算的進展》，第 843 卷。https://doi.org/10.1007/978-3-319-99007-1_30

[5] ACHANTA R.，SHAJI A.，SMITH K.，LUCCHI A.，FUA P.和 SÜSSTRUNKS. SLIC 超像素，EPFL 技術報告 149300，2010 年 6 月。

[6]李茵，孫健，唐致強和沈香陽。偷懶。圖形上的 ACM 事務，2004，23（3）：303-308。 doi：10.1145 / 1015706.1015719

[7] FULKERSON B.，VEDALDI A. 和 SOATTO，S.具有超像素鄰域的類分割和對象定位// 2009 年 IEEE 第 12 屆計算機視覺國際會議論文集，日本京都，2009 年 9 月 29 日至 10 月 2 日，頁數 670-677：doi：10.1109 / ICCV.2009.5459175

[8]王敏。劉新。高野。馬新.SOOMRO I.C.超像素分割：基準測試，信號處理圖像通信，2017,56：28-39

[9]與最先進的超像素方法相比，ACHANTA R.，SHAJI A.，SMITH K.，LUCCI A.，FUA P. 和 SUSSTRUNK S. SLIC 超像素。 IEEE 軟件工程學報，2012，34（11）：2274-2282，doi：0.1109 / TPAMI.2012.120。

[10] SHEN J.，DU Y.，WANG W.，和 LI X.懶散隨機游動用於超像素分割。 IEEE 軟件工程學報，2014，23（4）：1451-1462。 doi：10.1109 / TIP.2014.2302892

[11] 徐 L，王 Z，曾 L 基於顯著性的超像素。信號圖像與視頻處理，2014，8（1）：181-190。

[12] l IU Yong-jin, Y u Cheng-χ, yum in-jin Gan 的 he ying。流形 SLIC：一種計算內容敏感超像素的快速方法// 2016 年 6 月 27 日至 30 日，美國內華達州拉斯維加斯，2016 年 I 屆 IEEE 計算機視覺和模式識別會議論文集，第 651-659 頁。 doi：10.1109 / CVPR.2016.77

[13] ZHU Hongyuan, MENG Fanman, CAI Jianfei 和 LU Shijian.。超越像素：從下至上到語義圖像分割和細分的全面調查。視覺傳達與圖像表示雜誌，2015，34：12-27 doi：10.1016 / j.jvcir.2015.10.012。

[14] SHI J. 和 MALIK J. 歸一化剪切和圖像分割。IEEE 模式分析與機器智能交易，2000，22（8）：888-905。 doi：10.1109 / 34.868688

[15] FELZENSZWALB P. F.和 HUTTENLOCHER D.P.基於圖的有效圖像分割，國際計算機視覺雜誌，2004，59（2）：167-181。

[16] MOORE A.P.，PRINCE S.J..D，WARRELL J.，MOHAMMED U.和 JONES G. 超像素 點陣// 2008 年 IEEE 計算機視覺和模式識別會議論文集，安克拉治，阿拉斯加，美國，2008 年 6 月 23 日至 28 日，頁數。 1-8。 doi：10.1109 / CVPR.2008.4587471。

[17] ZHANG Y.，HARTLEY R.，MASHFORD J.和 BURN S.通過偽布爾優化實現的超像素。 // 2011 年 IEEE 計算機視覺國際會議論文集，西班牙巴塞羅那，2011 年 11 月 6 日至 13 日，第 1387-1394 頁。 doi：10.1109 / ICCV.2011.6126393

[18] LIU M.-Y.，TUZEL O.，RAMALINGAM S.和 CHELLAPPA R.熵率超像素分割// 6 月 21 日至 23 日在美國科羅拉多斯普林斯舉行的 2011 IEEE 計算機視覺和模式識別會議論文集， 2011，第 2097-2104 頁。

19] VINCENT L. 和 SOILLE，P. 數字空間中的分水嶺：一種基於沉浸模擬的有效算法。IEEE 模式分析與機器智能交易，1991，13（6）：583-598。

[20] LEVINSHTEIN A.，STERE A.，KUTULAKOS K.，FLEET D.，DICKINSON S.和 SIDDIQI，K. TurboPixels：使用幾何流的快速超像素。 IEEE 模式分析和機器智能交易。 2009，31（12）：2290-2297。 doi：10.1109 / TPAMI.2009.96

[21] COMANICIU D. 和 PETER M. 均值平移：進行特徵空間分析的可靠方法。 IEEE 模式分析和機器智能交易。 2002，25（5）：603-619。

[22] VEDALDI A. 和 SOATTO S. 模式轉換的快速轉換和內核方法//第十屆歐洲計算機視覺會議論文集，

法國馬賽，2008 年 10 月 12 日至 18 日，第 705-718 頁。

[23] FANG Zhuoqun, YU, Xiaosheng WU, Chengdong , CHEN Dongyue 和 JIA Tong.。在 RGBD 圖像上使用加權共面特徵聚類進行超像素分割。應用科學，2018，8（6）：902. doi：10.3390 / app8060902。

[24] CONG Jinyu, WEI Benzheng, YIN Yilong, XI Xiaoming, 和 ZHENG Yuanjie。簡單線性迭代聚類算法在醫學圖像處理中的性能評估。生物醫學材料與工程。 2014，24（6）：3231-3238。 10.3233 / BME-141145。

[25] Q inf A 超, GU OJ i 名 and Lang F 翰開.。使用局部迭代聚類的極化 SAR 圖像的超像素分割。地學與遙感快報，IEEE，2014，12（1）：13-17。 doi：10.1109 / LGRS.2014.2322960。

[26] WANG Peng, ZENG Gang, GAN Rui, WANG Jingdong 和 ZHA Hongbin.。通過測地距離對結構敏感的超像素。國際計算機視覺雜誌，2013，103（10：1-21。doi：10.1007 / s11263-012-0588-6

[27] RAMU B. 和 THIRUPATHI D.在圖像挖掘中使用超聲圖像診斷闌尾炎的新方法//第二屆國際計算科學，工程和信息技術國際會議論文集，印度哥印拜陀，2012 年 10 月 26-28 日，第 717-721 頁。

[28] REN X.和 MALIK J.學習分割的分類模型//第 9 屆 IEEE 國際計算機視覺國際會議論文集，聖彼得堡，俄羅斯，2013 年 7 月 16-18 日，1：10-17 doi：10.1109 /ICCV.2003.1238308。

[29] ARBELAEZ P.，MAIRE M.，FOWLKES C.和 MALIK J.等高線檢測和分層圖像分割 IEEE 模式分析與機器智能交易，2011，33（5）：898-916。 doi：10.1109 / TPAMI.2010.161。

[30] XU C.和 CORSO JJ 對用於早期視頻處理的超級體素方法的評估// 2012 年 IEEE 計算機視覺和模式識別會議論文集，普羅維登斯，美國羅德島，2012 年 6 月 16 日至 21 日，第 1202-120 頁 1209，doi：10.1109 / CVPR.2012.6247802。

[31] STUTZ D.超像素分割：一種評估。在：GALL P.GEHLER B.和 LEIBE B.（編輯）計算機科學講座筆記，2015，9358：555-562 doi：10.1007 / 978-3-319-24947-6_46

[32] RAND W。評估聚類方法的客觀標準。美國統計協會雜誌，1971，66（336）：846-850。 doi：10.2307 / 2284239。 66