# A Novel Unsupervised Evaluation Metric Based on Heterogeneity Features for SAR Image Segmentation

Hang Yu, *Member, IEEE*, Xiangjie Yin [ID], Zhiheng Liu [ID], Suiping Zhou, Chenyang Li [ID], and Haoran Jiang

*Abstract*—The segmentation of synthetic aperture radar (SAR) images is vital and fundamental in SAR image processing, so evaluating segmentation results without ground truth (GT) is an essential part in segmentation algorithms comparison, parameters selection, and optimization. In this study, we first extracted the heterogeneous features (HF) of SAR images to adequately describe the SAR image targets, which were extracted by the proposed intensity feature extractor (IFEE) based on edge-hold and two fruitful methods. Then we proposed a novel and effective unsupervised evaluation (UE) metric $G$ to evaluate the SAR image segmentation results, which is based on HF and uses the global intrasegment homogeneity ($GHO$), global intersegment heterogeneity ($GHE$), and edge validity index ($EVI$) as local segmentation measures. The effectiveness of $GHO$, $GHE$, $EVI$, and $G$ was revealed by visual interpretation as qualitative analysis and supervised evaluation (SE) as quantitative analysis. In experiments, four segmentation algorithms are used to segment plenty of synthetic and real SAR images as the evaluation objects, and four widely used metrics are utilized for comparison. The results show the effectiveness and superiority of the proposed metric. Moreover, the mean correlation between the proposed UE metric and the SE metric is more than 0.67 and 0.99, which indicates that the proposed metric helps in choosing parameters of segmentation algorithms without GT.

*Index Terms*—Intersegment heterogeneity metric, intrasegment homogeneity metric, synthetic aperture radar (SAR) image segmentation, unsupervised evaluation (UE) metric.

## I. INTRODUCTION

SYNTHETIC aperture radar (SAR) is an active microwave imaging system that can provide high-resolution images day and night under all weather conditions [1], [2], [3]. SAR is used widely in many applications, such as environmental observations, crop monitoring, and military reconnaissance [4], [5], [6]. SAR image segmentation (SIS) methods divide the image into regions of different features without intersections, where every pixel in images gets the corresponding label [7]. Segmentation is vital for understanding and interpreting SAR

images [8], [9], [10]. After a SAR image is segmented, it is important to evaluate the quality of the segmentation results (seg-results) so that the image can further be processed, an optimal algorithm can be chosen, and related parameters can be properly adjusted [11], [12], [13], [14]. Thus, the research on SAR image segmentation evaluation metrics (SISEM) is significant to promote the development of SIS.

Plenty of scholars have studied SIS issues and achieved fruitful achievements. For example, Shang et al. [15], [16] conducted SIS research from the perspectives of modeling and optimization. Yu et al. [8], [17] launched SIS from the multifeature fusion of SAR images. Akbarizadeh et al. [18], [19], [20], [21] solved SIS problems from kurtosis, skewness wavelet energy, and the feature learning. In addition, Aghaei et al. [22], [23], [24], [25] also made achievements in object classification and detection of SAR images based on deep learning. This research can also provide new ideas for the research of SIS. However, compared to a large number of studies on SIS, there are few studies on SISEM. The existing SISEM can be divided into five categories [14]: subjective evaluation methods, system-level methods, analytical methods, supervised evaluation (SE) methods, and unsupervised evaluation (UE) methods [26], [27], [28], [29], [30], [31]. The subjective, system-level and analytical methods are more subjective and experience-dependent. Thus, the SISEM mainly relies on more objective SE methods [15], [16], [17].

SE methods [32] are designed to quantitatively measure the dissimilarity between the SIS results and the ground truth (GT) images to assess the quality of seg-results. However, the SE methods require the GT dataset to be composed manually [33]. Building the whole GT images for massive SAR images is tedious, time-consuming, involves subjectivity, and is hardly obtained in many cases. Overcoming the subjectivism in the GT building would save both time and effort [34].

UE methods [27] do not require GT images, and SIS results are evaluated by calculating human-recognized criteria representing good seg-results. UE is quantitative and objective and has apparent advantages. The most important benefit is that UE methods can evaluate different types of SIS without GT images [35], [36]. Therefore, it is of great significance to study UE methods that can replace SE methods. The UE become an inevitable trend in the study of SIS evaluation [37]. However, there are few studies dedicated to the UE of SAR image seg-results.

UE methods involve scoring and ranking multiple image segmentation using quality criteria, which are typically established

consonant human perceptions of what makes a good segmentation [38], [39]. The widely recognized definition of an ideal seg-result, given by Haralick and Shapiro [40], is as follows.

1) Regions should be uniform and homogeneous concerning some characteristic(s).
2) Adjacent regions should have some significant differences for the characteristic on which they are uniform.
3) Region interiors should be simple and without holes.
4) Boundaries should be simple, not ragged, and be spatially accurate.

Therefore, for SAR images, a good segmentation should maximize intrasegment homogeneity, intersegment heterogeneity, and edge validity. These elements are combined to assign an overall "goodness" score to the segmentation [11].

In this article, we aim to design a UE metric based on the heterogeneous features of SAR images for evaluating SIS results without GT. Establishing whole GT images for plenty of SAR images is tedious, time-consuming, and involves subjectivity. In SAR image processing research, the situation that SAR images without GT is widespread. And the current UE metrics only depend on a single feature to evaluate seg-results, especially. The metrics based on the single-feature design are valid for the seg-results of single-feature images but fail for the results of SAR images containing multiple features. Therefore, the key issue of designing for UE metric is to find features that can completely describe targets in SAR images to escape from the dependence on GT. Its difficulty is first to examine the "goodness" of intrasegments, intersegments, and segmented edges according to the features, and second, to design reasonable global metrics to evaluate the quality of seg-results reasonably. The similarity measures are used to measure the similarity between different objects and are mostly used in various evaluation and determination situations. Therefore, we consider examining homogeneity, heterogeneity, and edge validity according to characteristics of SAR image different feature data and reasonable metrics.

More importantly, it can hardly find a special evaluation metric designed for SIS results in existing UE methods. Therefore, to solve the existing intractable problems in designing of UE metric of SIS results, we proposed a novel global UE metric $G$ based on the heterogeneous features for SIS results in this article. We designed an effective feature extractor and extracted effective heterogeneous features of SAR images, proposed three local evaluation metrics, and designed a strategy to combine local metrics into a global evaluation metric, specifically. The main contributions of this article are as follows.

1) We conducted the SISEM design using a multifeature-based approach and extracted multiple features of SAR images, which provide effective information for designing $G$. In detail, we proposed an intensity feature extractor (IFEE) based on edge-hold to extract intensity features and used two fruitful methods to extract texture and edge features. IFEE has the properties of keeping edges while suppressing speckle noise. These heterogeneity features can describe the targets of SAR images completely.
2) Based on the heterogeneous features, we evaluate the SIS results in terms of homogeneity, heterogeneity, and segmentation edge to design local metrics for segmentation
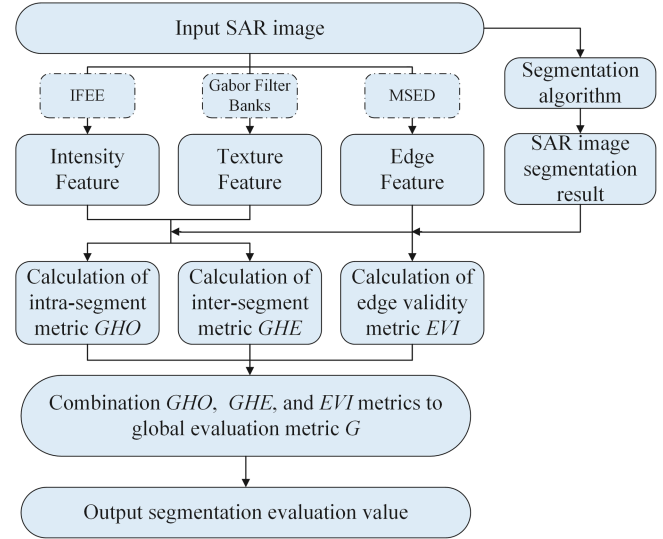


Fig. 1. Calculation flowchart of the proposed UE metric.

evaluation. We designed three local metrics to indicate the quality of SIS results in three local aspects: global intrasegment homogeneity ($GHO$), global intersegment heterogeneity ($GHE$), and edge validity index ($EVI$).

3) We designed a combination strategy that maximizes the intrasegment homogeneity, intersegment heterogeneity, and edge validity to fuse $GHE$, $GHO$, and $EVI$ into a global UE metric $G$. The UE metric $G$ with fused local metrics can accurately and quantitatively evaluate the seg-results of SAR images without GT.

The rest of this article is organized as follows. Section II discusses the proposed methods. The experiments and results are presented in Section III. Further analysis is discussed in Section IV. Finally, the main findings are concluded in Section V.

## II. METHODS

In this study, a global UE metric $G$ was designed for evaluating SAR image seg-results accurately and objectively. $G$ metric was developed based on heterogeneous features of SAR images and satisfied the principle of maximizing intrasegment homogeneity, intersegment heterogeneity, and edge validity. A calculation schematic of the metric $G$ is shown in Fig. 1 according to the feature criterion of UE methods discussed in Section I. First, input the SAR image $I$ and the corresponding seg-result $I_k$. The input image $I = \{x_1, x_2, x_3, \ldots, x_i, \ldots, x_N\}$ is of size $N_r \times N_c = N$. The input SAR image is divided into $M$ ($M > 1$) different segments ($A_1, \ldots, A_k, \ldots, A_M, 1 < k \leq M$) by a segmentation algorithm, and the seg-result is $J = \{y_1, y_2, y_3, \ldots, y_i, \ldots, y_N\}$, $y_i \in \{1, 2, 3, \ldots, M\}$. Then, the intensity, texture, and edge features of the SAR image were extracted by three effective feature extractors, which involved IFEE, Gabor filter banks, and multiscale edge detector (MSED). These features are heterogeneous and can adequately describe ground targets. Next, the intrasegment homogeneity metric $GHO$, the intersegment heterogeneity metric $GHE$, and the edge validity index $EVI$ were calculated according to heterogeneous
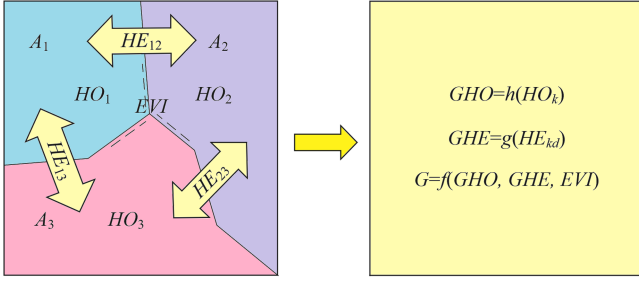
Fig. 2. Evaluation of the proposed metric on a 3-class seg-result.

features. Finally, the metric $G$ was calculated according to *GHO*, *GHE*, and *EVI* to evaluate SIS results. Fig. 2 illustrates a schematic of the evaluation process of the metric $G$ on a 3-Class seg-result.

### A. Extraction of Heterogeneous Features

Most existing UE metrics such as $F$ [33] are generally computed to evaluate the quality of seg-results based on a single image feature. These metrics are effective if each ground object has a uniform feature. Not only do SAR images show high spatial complexity, but also the ground objects are typically characterized by intensity, texture, and edge features [34]. Therefore, it is difficult for a uniform feature to describe all targets entirely and accurately. The heterogeneous features of SAR images were extracted, including intensity, texture, and edge features, in this study. Our previous study [8] showed that these features complement each other and can accurately and completely describe the various targets in SAR images. Thus, the UE metric $G$ based on the heterogeneous features can get an objective and accurate evaluation of SIS results.

*1) Extraction of Intensity Feature:* The magnitude of each pixel in SAR images represents the intensity of the reflected echo from the ground targets [41]. Therefore, the value of pixels in SAR images has an essential difference from general optical images, which fully reflect the radar echoes of ground objects [42]. The intensity is a very significant feature of SAR images. Speckle noise causes the intensity value to vary randomly in SAR images, which seriously affects the accurate extraction of intensity features. Therefore, we proposed IFEE to extract the intensity feature of SAR images.

IFEE assumes that the natural scene radar reflection cross section in SAR images obeys the Gamma distribution. IFEE first applies the maximum a posteriori (MAP) probability criterion [37] to suppress the speckle noise, and then combines the image spatial proximity and intensity similarity to perform region smoothing (RS) and edge preservation [34] to extract intensity features. The input SAR images are first normalized before applying IFEE.

Let $\mathbf{X}^B = \{x_1^B, x_2^B, \ldots, x_N^B\}$ represents normalized intensity values of an input SAR image $I$. Applying MAP to $\mathbf{X}^B$ can obtain $\mathbf{X}^M = \{x_1^M, x_2^M, \ldots, x_N^M\}$, which operates $x_i^B$ through a $3 \times 3$ window as follows:

$$x_i^M = \frac{(\alpha - L - 1)\mu + \sqrt{\mu^2(\alpha - L - 1)^2 + 4\alpha L \mu x_i^B}}{2\alpha} \quad (1)$$
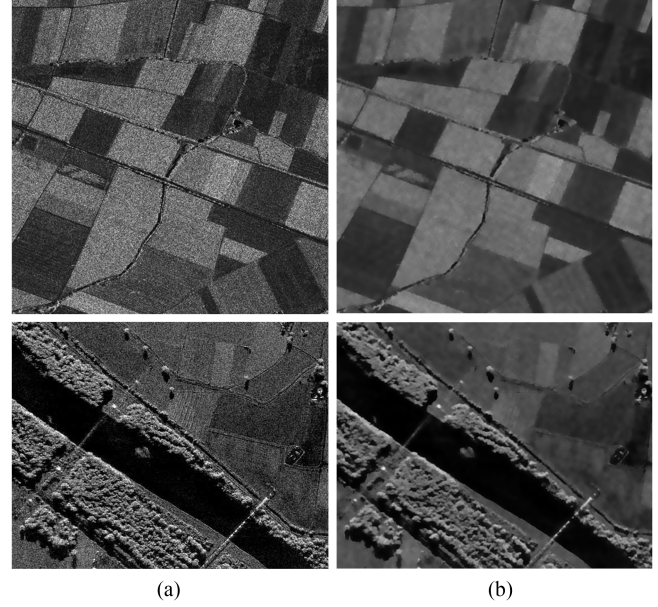


Fig. 3. Two real SAR images and the corresponding IFEE feature images.

where $\alpha = (1+1/L)/(\delta^2/\mu^2 - 1/L)$, $\delta^2$ is the variance of intensity values in the $3 \times 3$ window, $\mu$ is the mean value, and $L$ is the image Look. Then, using a convolution kernel $W_{ij}$ with spatial proximity, $\sigma_s$ and intensity similarity $\sigma_r$ to $\mathbf{X}^M$ can obtain the final intensity feature $\mathbf{X}^I = \{x_1^I, x_2^I, \ldots, x_N^I\}$. The $W_{ij}$ is represented as follows:

$$W_{ij}\left(\mathbf{X}^M\right) = \frac{1}{K_i} \exp\left(-\frac{|c_i - c_j|^2}{\sigma_s^2}\right) \cdot \exp\left(-\frac{|x_i^M - x_j^M|^2}{\sigma_r^2}\right) \quad (2)$$

where $c_i$ and $c_j$ are the coordinates of pixels $x_i$ and $x_j$, and $K_i$ is a normalizing parameter to ensure that $\Sigma_j W_{ij} = 1$. We set $\sigma_s = 5$ and $\sigma_r = 0.1$ in the actual application after plenty of verification.

Two real SAR images are shown in Fig. 3(a), and the corresponding intensity features extracted by IFEE are Fig. 3(b). The homogeneity within the target regions of the intensity features extracted by IFEE is obviously enhanced. Meanwhile, the excellent edge retention proves that instead of fusing the information of different target regions, the IFEE operation enhances the heterogeneity of different regions.

It can be seen that different surface objects are different in intensity, such as the farmland in Fig. 3(b). However, it is hard to distinguish these surface objects only according to the intensity feature. Therefore, we will introduce the extraction of other heterogeneous features in the subsequent sections.

*2) Extraction of Texture Feature:* The texture feature provides an accurate localization of boundaries and spatial information that can delineate the actual forms of ground objects adequately. This complementary information can discriminate various objects which are similar to intensity. The feature representation should be an excellent fit for the objects of human vision [43].
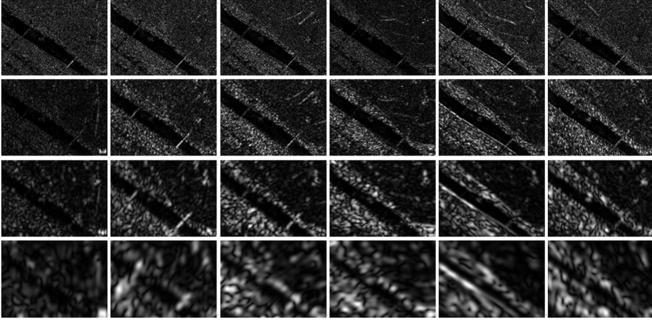
Fig. 4. Multiscale and multidirectional texture features of Fig. 3(a) second row.

The Gabor filter is widely used in SAR image texture expression due to the perceptual structure of the filter being similar to the human visual system [44], [45]. This study implements a two-dimensional Gabor wavelet to characterize the texture information, which is effective in our previous study [8]. The Gabor wavelet is expressed as follows [46]:

$$G(x, y) = \frac{\|k\|^2}{\sigma^2} \exp\left(-\frac{\|k\|^2 (x^2 + y^2)}{2\sigma^2}\right)$$

$$\cdot \exp\left(i k \begin{bmatrix} x \\ y \end{bmatrix} - \exp\left(-\frac{\sigma^2}{2}\right)\right) \quad (3)$$

$$k = \begin{bmatrix} k_x \\ k_y \end{bmatrix} = \begin{bmatrix} k_v \cos \varphi_u \\ k_v \sin \varphi_u \end{bmatrix} \quad (4)$$

where $\| \bullet \|$ donates the norm operator, $1 \leq x \leq N_r$ and $1 \leq y \leq N_c$ denote the image space horizontal and vertical coordinates, respectively, $k_v = 2^{-(v+2)/2}\pi$, $\varphi_u = u \cdot \pi/U$, $v$ and $u$ represent central frequency and orientation, $U$ is the number of the orientation parameters $u$, and $\sigma$ is the ratio of the Gaussian window width to the wavelength.

The study [43] demonstrated an efficient method of Gabor filter bank setting parameters and can adaptively extract the multiscale and multiorientational texture features of an input image. Applying the bank of Gabor filters with $V$ frequencies and $U$ orientations to a raw image $\mathbf{X}^B$ can generate $V \times U$ response images. Then, the pixel $x_i$ gets a Gabor texture feature of $V \times U$ dimension, which is defined as follows:

$$\mathbf{x}_i^T = \left[x_{i1}^T, x_{i2}^T, \ldots, x_{i(V \times U)}^T\right]. \quad (5)$$

The texture feature visualization results of Fig. 3(a) second row were shown in Fig. 4, which the Gabor filter banks with four scale parameters and six orientation parameters.

*3) Extraction of Edge Feature:* SAR images have abundant edge information. Edge features in SAR images not only help locate target edges but also assist in distinguishing quickly between different targets. These advantages are not afforded by intensity and texture features. Since SAR images are affected by strong speckle noise, the signal-to-noise ratio is low, which causes challenges for edge detection of SAR images. Traditional edge detection operators applied to SAR images usually produce more false alarm edges, which are poorly handled, and difficult
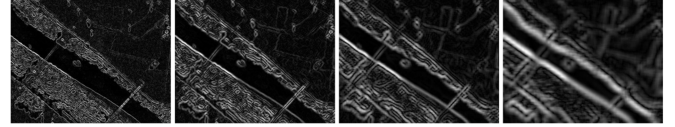


Fig. 5. Multiscale edge features of Fig. 3(a) second row.

to characterize the edge information at various scales in SAR images.

An MSED in our previous study [8] proves to be extremely effective in the SAR image edge detection process, which extracts multiscale SAR image edge features while dramatically reducing the effect of speckle noise. MSED is based on the Prewitt template, and the standard Prewitt template is expanded to contain $V$ scales by expanding the window size represented as follows:

$$g_m^r = \begin{bmatrix} -1 & \cdots & -1 & 0 & 1 & \cdots & 1 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ -1 & \cdots & -1 & 0 & 1 & \cdots & 1 \end{bmatrix}_V,$$

$$g_m^c = (g_m^r)^T, \ i = 1, \ldots, V \quad (6)$$

where $g_m^r$ and $g_m^c$ represent the vertical and horizontal edge detection templates in scale $m$, and $V$ is the number of scales. The number of scale parameters $V$ is the same as the number of scale parameters of the Gabor filter, which ensures that the edges have the same scale as the texture. Next, the normalized SAR image $\mathbf{X}^B$ is convolved with $g_m^r$ and $g_m^c$, respectively, to obtain vertical edge features $\mathbf{G}_m^r$ and horizontal edge features $\mathbf{G}_m^c$, which are performed as follows:

$$\begin{cases} \mathbf{G}_m^r = \mathbf{X}^B \otimes g_m^r \\ \mathbf{G}_m^c = \mathbf{X}^B \otimes g_m^c \end{cases}. \quad (7)$$

The Prewitt template increases as $m$ increases, allowing for the detection of larger scale edges and better removal of speckle noise. Finally, the vertical and horizontal edge intensities are used as the final edge features operating as

$$x_{im}^E = \sqrt{(G_{im}^r)^2 + (G_{im}^c)^2} \quad (8)$$

where $x_{im}^E$ is the $m$th dimension edge feature of the pixel $x_i$, $G_{im}^r \in \mathbf{G}_m^r$ and $G_{im}^c \in \mathbf{G}_m^c$ are the components of the vertical and horizontal edge features in the pixel $x_i$. Then, a set of multiscale edge features is obtained for each pixel $x_i$

$$\mathbf{x}_i^E = \left[x_{i1}^E, x_{i2}^E, \ldots, x_{iV}^E\right]. \quad (9)$$

Thus, the extraction of multiscale edge features by applying MSED is completed. Fig. 5 shows the edge feature images of Fig. 3(a) second row at different scales. It can be seen that the multiscale edge images can effectively represent the edge information at various scales in SAR images.

*B. Intrasegment Homogeneity, Intersegment Heterogeneity, and Edge Validity Metrics*

The widely recognized definition of ideal seg-results requires rules of high intrasegment homogeneity, intersegment heterogeneity, and edge validity. Considering the purpose of perfect

segmentation and heterogeneous features of SAR images, we respectively design the intrasegment homogeneity metric ($HO_k$) for each segment $A_k$ and the intersegment heterogeneity metric ($HE_{kd}$) between two segments $A_k$ and $A_d$. All $HO_k$ were combined into a global homogeneity metric ($GHO$) and all $HE_{kd}$ into a global heterogeneity metric ($GHO$). And we designed an $EVI$ to reveal the edge quality of seg-results. $GHO$, $GHE$, and $EVI$ metrics are designed based on intensity, texture, and edge features to evaluate the seg-results in different local areas.

*1) Global Intrasegment Homogeneity Metric:* The global intrasegment homogeneity metric $GHO$ is based on intrasegment homogeneity $HO_k$ to design. All $HO_k$ were summed up by area-weighted for $GHO$ due to segment with larger areas having a more significant impact on the global evaluation than the smaller ones. The $GHO$ is calculated as follows:

$$GHO = \frac{\sum_{k=1}^{M} N_k \cdot HO_k}{\sum_{k=1}^{M} N_k}. \tag{10}$$

The intrasegment homogeneity $HO_k$ of a segment $A_k$ is described by both intensity homogeneity $v_k^g$ and texture homogeneity $v_k^t$, which is calculated as follows:

$$HO_k = v_k^g \cdot v_k^t. \tag{11}$$

The $v_k^g$ calculates the variance of the intensity features in the segment $A_k$. The texture homogeneity $v_k^t$ is based on chi-squared test. The low variance of the intensity in the segment indicates high homogeneity for the intensity feature. Therefore, the smaller $v_k^g$ indicates the higher intensity homogeneity. The $v_k^g$ and $v_k^t$ are calculated as follows:

$$\begin{cases} v_k^g = \frac{1}{N_k-1}\sum_{x_i \in A_k}\left(x_i^I - \mu_k\right)^2 \\ v_k^t = \frac{\sum_{x_i \in A_k}\sum_{n=1}^{V \times U}\left(\frac{x_{in}^T - x_{n\,\max}^T}{x_{n\,\max}^T}\right)^2}{N_k} \end{cases} \tag{12}$$

where $N_k$ is the number of pixels in segment $A_k$, $x_i^I (i = 1, 2, 3, \ldots, N_k)$ is the intensity of the pixel $x_i$ in segment $A_k$, and $\mu_k$ is the intensity values mean of $A_k$. The $x_{in}^T$ is the $n$th dimension texture feature of the pixel $x_i$, $x_{n\,\max}^T$ is the maximum of all $x_{in}^T$ in $A_k$. We calculate the $x_{n\,\max}^T$ in each dimension of texture features and get a texture maximum vector $\mathbf{x}_{\max}^T$ with $V \times U$ dimension in $A_k$

$$\mathbf{x}_{\max}^T = [x_{1\,\max}^T, x_{2\,\max}^T, x_{3\,\max}^T, \ldots, x_{n\,\max}^T, \ldots, x_{V \times U\,\max}^T]. \tag{13}$$

Three types of ground targets are noted as A, B, and C in Fig. 6. The histograms of $\mathbf{x}_{\max}^T$ in A, B, and C have distinct differences, as shown in Fig. 7. Therefore, $\mathbf{x}_{\max}^T$ can represent the texture features of a single region for texture homogeneity calculation. The essence of texture homogeneity $v_k^t$ is the mean of the similarity measures, which are calculated from the chi-squared test measures of all $\mathbf{x}_i^T$ with the $\mathbf{x}_{\max}^T$ of $A_k$. It can be noted that the texture homogeneity $v_k^t$ computed by $v_k^t$ can compute not only the difference between the two sets of texture features in the range of values but also their similarity in the distribution. The smaller $v_k^t$ indicates the higher texture homogeneity for $A_k$.

The $HO_k$ was obtained by multiplying $v_k^g$ and $v_k^t$ shown in (11). The smaller $HO_k$ indicates the higher intrahomogeneity
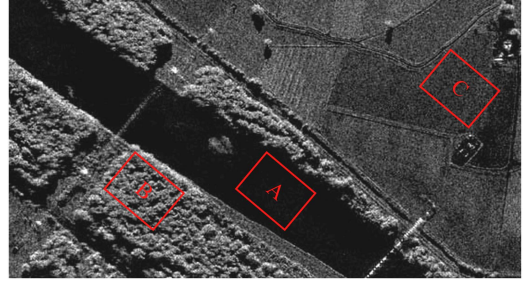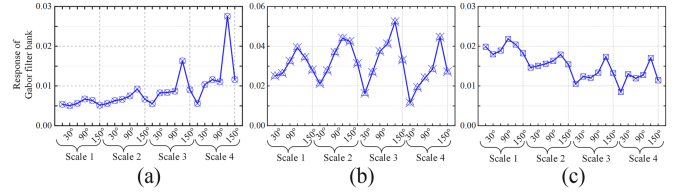


Fig. 6. Three types of ground targets in a SAR image.



Fig. 7. $\mathbf{x}_{\max}^T$ of three types ground targets are in Fig. 6.

of $A_k$ because the smaller $v_k^g$ and $v_k^t$ indicate the higher in both intensity and texture homogeneity for a segment $A_k$. Thus, a smaller $GHO$ represents a higher global homogeneity for seg-results.

*2) Global Intersegment Heterogeneity Metric:* The intersegment heterogeneity metric $GHE$ is based on intersegment heterogeneity $HE_{kd}$ to design. All $HE_{kd}$ were cumulated for global heterogeneity calculation

$$GHE = \frac{1}{2}\sum_{k \neq d} HE_{kd}. \tag{14}$$

The intersegment heterogeneity $HE_{kd}$ of any two segments $A_k$ and $A_d$ is described by both intensity heterogeneity $S_{kd}^g$ and texture heterogeneity $S_{kd}^t$

$$HE_{kd} = \frac{S_{kd}^g}{S_{kd}^t}. \tag{15}$$

The $S_{kd}^g$ calculates the similarity of the normalized intensity histograms of the segments $A_k$ and $A_d$, which is based on the Bhattacharyya coefficient (BhC) [47]. Therefore, the texture heterogeneity $S_{kd}^t$ is designed depending on the Canberra distance. The $S_{kd}^g$ and $S_{kd}^t$ are calculated as follows:

$$\begin{cases} S_{kd}^g = \sum_{q=1}^{256} \sqrt{\text{Hist}_{gk}^q \cdot \text{Hist}_{gd}^q} \\ S_{kd}^t = \sum_{q=1}^{V \times U} \frac{|f_k - f_d|}{|f_k| + |f_d|} \end{cases} \tag{16}$$

where $\text{Hist}_{gk}^q$ and $\text{Hist}_{gd}^q$ denote the normalized intensity histograms of $A_k$ and $A_d$, and $q$ represents the $q$th element of the histogram. The normalized histogram eliminates the pixel number difference between segments $A_k$ and $A_d$. Then, the BhC measures the similarity of the vectors in the range [0, 1], where 1 means perfectly similar and 0 means not similar at all. Thus, the smaller $S_{kd}^g$ indicates high heterogeneity of intensity features in $A_k$ and $A_d$. The $f_k$ is the texture feature descriptors of segments

$A_k$ and $f_d$ belongs to $A_d$

$$f_k = \{E_{k1}, \sigma_{k1}, E_{k2}, \sigma_{k2}, \ldots, E_{kn}, \sigma_{kn}, \ldots, E_{kV \times U}, \sigma_{kV \times U}\} \tag{17}$$

where $E_{kn}$ and $\sigma_{kn}$ are the mean and standard deviations of the $n$th dimension texture feature in $A_k$. The research [48] shows that the descriptor $f_k$ can effectively describe the global texture feature of the region and can be used to measure the texture similarity of two regions.

The large $S_{kd}^t$ indicates high texture heterogeneity between $A_k$ and $A_d$. And a small $S_{kd}^g$ indicates high heterogeneity of intensity features between $A_k$ and $A_d$. Therefore, the smaller $\text{HE}_{kd}$ represents higher heterogeneity between segments $A_k$ and $A_d$. A smaller $GHE$ value represents a higher global heterogeneity for seg-results.

*3) Edge Effective Index:* Edges with good quality seg-results should contain more edges of real targets and fewer false edges. Therefore, we propose the *EVI*. *EVI* aims to characterize the validity of the segmented edges between the segments in seg-results. According to the seg-result expression, $J = \{y_1, y_2, y_3, \ldots, y_i, \ldots, y_N\}$, $y_i \in \{1, 2, 3, \ldots, M\}$, then the edge of the seg-result is expressed as $I_E = \{z_1, z_2, z_3, \ldots, z_i, \ldots, z_N\}$, $z_i \in \{1, 0\}$. $z_i$ equal to 1 represents the edge pixel of the seg-result. Based on the multiscale edge features and $I_E$, the edge effective index is expressed as

$$\text{EVI} = \frac{\sum_{k=1}^{V} \sum_{i=1}^{N} z_i \cdot x_{ik}^E}{V \cdot \sum_{i=1}^{N} z_i} \tag{18}$$

where $V$ is the number of scales of edge features and $N$ is the number of SAR image pixels. The essence of *EVI* is the average of the edge features $x_{ik}^E$ of different scales corresponding to the seg-result edges. Valid edges correspond to larger feature values $x_{ik}^E$, and invalid edges correspond to smaller feature values $x_{ik}^E$. Poor quality edge seg-results usually contain false edges, which correspond to low $x_{ik}^E$ values and different values at each scale. Therefore, these false edges will balance out the real edge feature values when calculating the average value resulting in lower *EVI*.

In addition, good quality edge seg-results have higher mean values and thus higher *EVI* values since they are not affected by false edges. Therefore, higher *EVI* indicates higher edge validity of seg-results, and lower values indicate lower edge validity of seg-results. Fig. 8 shows the general procedure of *EVI* calculation. The color from blue to red in Fig. 8 indicates the edge feature values from small to large. Fig. 8(a) represents the segmented result and its segmented edges, (b) represents multiscale edge features, and (c) represents multiscale edge features under segmented edges.

### C. Combination of GHO GHE and EVI Metrics

The *GHO*, *GHE,* and *EVI* quantitatively evaluate seg-results from different aspects, which include global intrasegment homogeneity, global intersegment heterogeneity, and edge validity. However, it is meaningless to rely only on a single homogeneity, heterogeneity, or edge metric to measure the segmentation quality. Therefore, *GHO*, *GHE,* and *EVI* must be sufficiently considered to achieve an objective, accurate, and comprehensive global evaluation of seg-results. Combining the
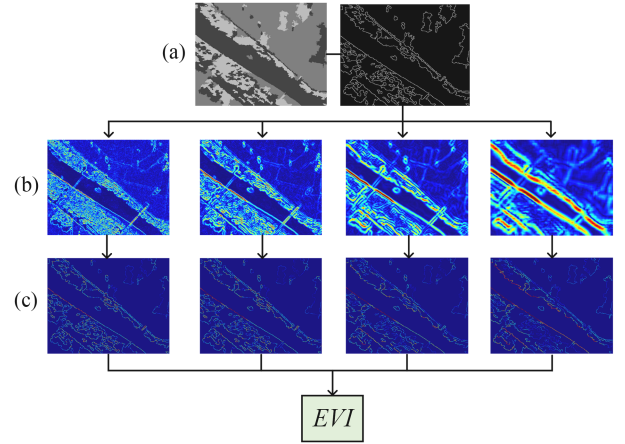


Fig. 8. General procedure of *EVI* calculation. (a) The SAR image segmented result and its segmented edges. (b) The multi-scale edge features of segmented SAR images. (c) The multi-scale edge features under segmented edges.

above design and analysis of homogeneity, heterogeneity, and edge metrics, the global metric $G$ was designed in the following manner:

$$G = \frac{GHO \cdot GHE}{EVI}. \tag{19}$$

As discussed in Section II-B, a smaller *GHO* and *GHE* indicate higher intrasegment homogeneity and intersegment heterogeneity, and a higher *EVI* indicates better edge validity. The principle of great seg-results requires high GHO, GHE, and effective target edges. Therefore, we multiply *GHO* by *GHE* using a strategy of minimizing intrahomogeneity and interheterogeneity, then an edge validity is added to this measure as the denominator to constraint. According to the combination strategy in (19), a smaller $G$ value means better segmentation quality.

## III. EXPERIMENTS

### A. Experimental Data and Setup

In this section, we utilize a wide range of real and synthetic SAR images [38], [39] to test the proposed method. Real SAR images are processed products of SAR data, and SAR images are generally composed of linear magnitude in terms of data. The synthetic SAR images are mainly single-look-complex (SLC) and multilook images. Therefore, the proposed method applies to diverse images used for SIS studies, such as SLC, multilook, and linear magnitude images. In addition, considering SAR images contain rich intensity, texture, and spatial information, the experiments require the synthetic objects to involve different intensities and multiple texture regions. The images used in the experiments and the corresponding experiment groups are shown in Fig. 9.

The real SAR images A1–A4 are shown in Fig. 10, which belong to magnitude images. Nördlinger Ries in Fig. 10 were taken in the middle of the Swabian Jura mountains in southwest Germany. The image size of A1–A3 is $512 \times 512$ and the A1 can be divided into three farmland areas and the A2 and A3 be four areas. The A4 is a Ku band image of the Rio Grande
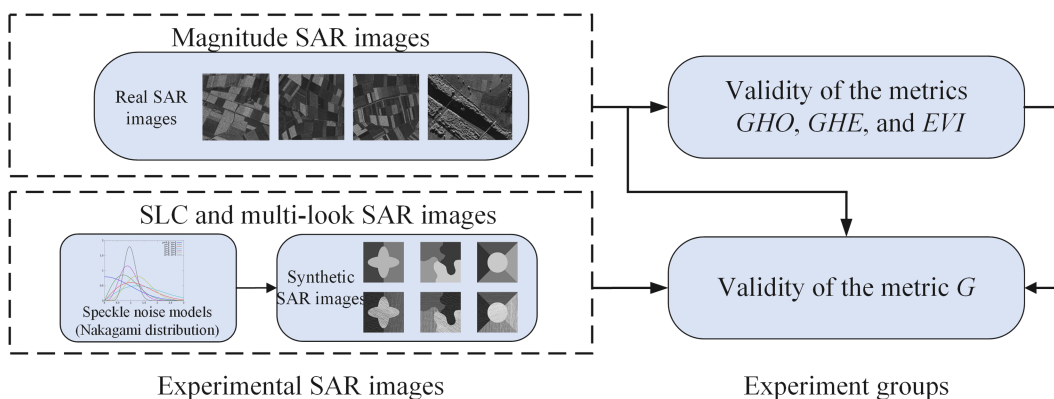
Fig. 9.   Images used in experiments and the corresponding experiment groups.
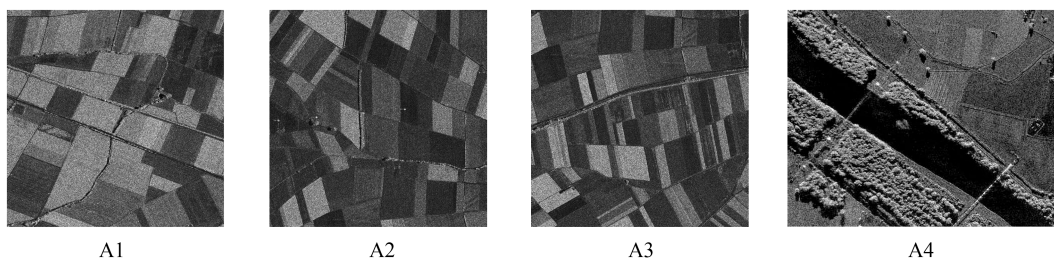


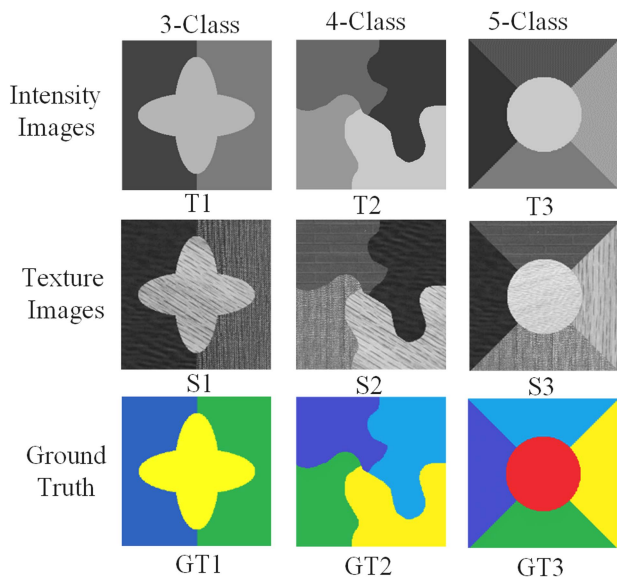Fig. 10.   Real SAR images used in experiments.



Fig. 11.   Synthetic SAR images and GT images with different classes.

basin near Albuquerque, New Mexico, USA, and can be divided into three classes of size 553 × 432. It is extremely difficult to build a rational GT for real SAR images. We did not provide GTs for A1–A4 due to the time-consuming building process and the differences in GTs marked by different people. Our UE metric G for SAR image seg-results is proposed to address this situation where GTs for a huge number of SAR images are difficult to obtain. The synthetic images T1–T3, S1–S3, and the GT are shown in Fig. 11, which includes 3–5 class images.

Intensity areas in T1–T3 are made up of different intensity values. And the texture areas used in S1–S3 were extracted from the USC texture database (https://sipi.usc.edu/database). The speckle noise widely exists in SAR images, and the noise is modeled by the multiplicative Nakagami distribution. Adding various speckle noises to synthetic SAR images can obtain SLC and multilook images for experiments, which have been proven effective in current studies [15], [49], [50]. The speckle noise was synthesized to T1–T3 and S1–S3, and then SLC (1Look) and multilook (2-10Look) images were obtained of size 512 × 512.

We designed two groups of experiments shown in Fig. 9 to fully validate the effectiveness of the proposed method, including GHO, GHE, EVI, and global metric G.

In the first experiments, metrics GHO, GHE, and EVI are tested by real SAR images. We base the RSLC segmentation algorithm with different classes to segment A1–A4 to obtain seg-results. Then, we calculate GHO, GHE, and EVI for each seg-result and combine these metrics into the G metric. These experiments can prove the local validity of the proposed method and the validity of the combined strategy.

In the second group of experiments, synthetic images T1–T3 and S1–S3 were segmented by three segmentation algorithms, and 180 seg-results were obtained. The real SAR images A1–A4 are segmented by the different algorithms with different parameters, and 20 seg-results of each image are obtained for the experiments. For each seg-result, different evaluation metrics are calculated using the SE method's segmentation accuracy (SA) if it has a GT image and multiple UE methods including the proposed method G. The experimental setup for seg-results evaluation is shown in Fig. 12.
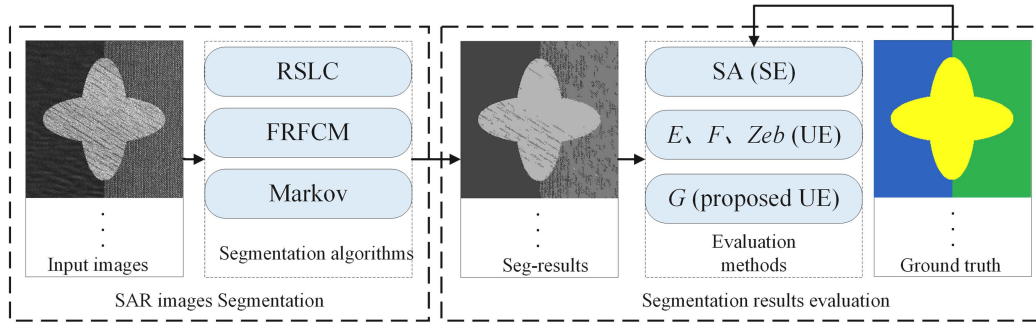
Fig. 12. Evaluation experiment setting of all seg-results.

TABLE I
DETAILS OF CURRENT UE METHODS AND OURS

| Metric | Intra-region | | Inter-region | | Intra- and Inter-region Combination |
| --- | --- | --- | --- | --- | --- |
| | Features | Combination | Features | Combination | |
| $E$ [52] | Entropy | $Sum_w$(size) | Entropy | — | Sum |
| $F$ [53] | Squared color error | Penal(sum) | — | — | — |
| $Zeb$ [27] | Color error | $Sum_w$(size) | Local color difference | $Sum_w$(length) | (Intra÷inter)\|inter |
| $G$ (ours) | Intensity variance and texture histogram | $GHO=g(HO_k)$ | Gray histogram and texture vector | $GHE=h(HE_{kd})$ | Penal(Intra×Inter) by $EVI$ |

"|" is the operation "or", and sum represents unweighted sum; sum$w(x)$ denotes summing per-region measures weighted by $x$; and penal(sum) denotes summary with some additional function applied to improve performance.

We conduct two application experiments in Sections IV-E and F after the above two experiments prove the effectiveness of the proposed method. One group is the application of the $G$ metric in the equivalent application of parameter selection, and another one is the evaluation of the $G$ metric in seg-results of deep neural networks (DNNs). The specific experimental setting is presented in the corresponding section.

## B. Existing Evaluation Metrics

There has been some research on segmentation evaluation metrics, including SE and UE metrics. This section will introduce the existing evaluation metrics, which will be used as comparison metrics with our method in next experiments.

A performance metric, $SA$, [51] is an SE metric, which is defined as the sum of the correctly classified pixels divided by the sum of the total number of pixels. The $SA$ is an extremely widely used SE metric in evaluating SIS results. In the studies of SIS such as [15], [16], and [51], the metric $SA$ is used to evaluate the quality of the SIS results. Although there are no UE metrics specifically for SIS evaluation, some general UE methods have been proposed. $E$ [52] is a UE method based on information theory and the minimum description length principle. $E$ uses the region entropy to measure intrasegment homogeneity, which measures the entropy of pixel intensities within each segment. $F$ [53] measures the average squared color error of the segments, punishing oversegmentation by weighting proportional to the square root of the number of segments. Zeboudj's contrast ($Zeb$) [27] is a UE criterion based on the internal and external contrasts

of the segments measured in the neighborhood of each pixel. The details of current UE methods and ours are given in Table I.

## C. Segmentation Algorithms Used in Experiments

In this study, four image segmentation algorithms are used for image segmentation to obtain seg-results, and the performance of the $G$ metric is verified based on the seg-results. The SIS method (RSLC) [15] using RS and label correction (LC) is specifically designed for SAR images. An improved FCM algorithm (FRFCM) [54] based on morphological reconstruction (MR) and membership filtering (MF) is a fast and robust image segmentation algorithm. Markov random field-based image segmentation (Markov) [55] is a statistical-based image segmentation algorithm with few parameters and solid spatial constraints, which is more widely used in image segmentation. Unsupervised image segmentation by backpropagation (UISB) [56] is an unsupervised segmentation algorithm based on a DNN.

The four algorithms will segment SAR images in the experiments to obtain the seg-results. Then the results are evaluated quantitatively using the proposed metrics and the existing metrics to evaluate the proposed method. The details of the four segmentation algorithms are shown in Table II.

## D. Results and Analysis

1) Effectiveness of the Global Intrahomogeneity, Global Interheterogeneity, Edge Metrics, and Combination Strategy: The RSLC algorithm was performed at A1–A4 on class parameters, ranging from 2 classes (under-segmentation) to 9 classes

TABLE II
MAIN PRINCIPLE AND APPLICATION FIELD OF SEGMENTATION ALGORITHMS

| Algorithm | Main principle | Application field |
|---|---|---|
| RSLC [15] | RS and LC | SAR image segmentation |
| FRFCM [54] | MR and MF | Color and gray image segmentation |
| Markov [55] | Markov Random Field Theory | Gray image segmentation |
| UISB [56] | DNN and Clustering | Color and gray image segmentation |



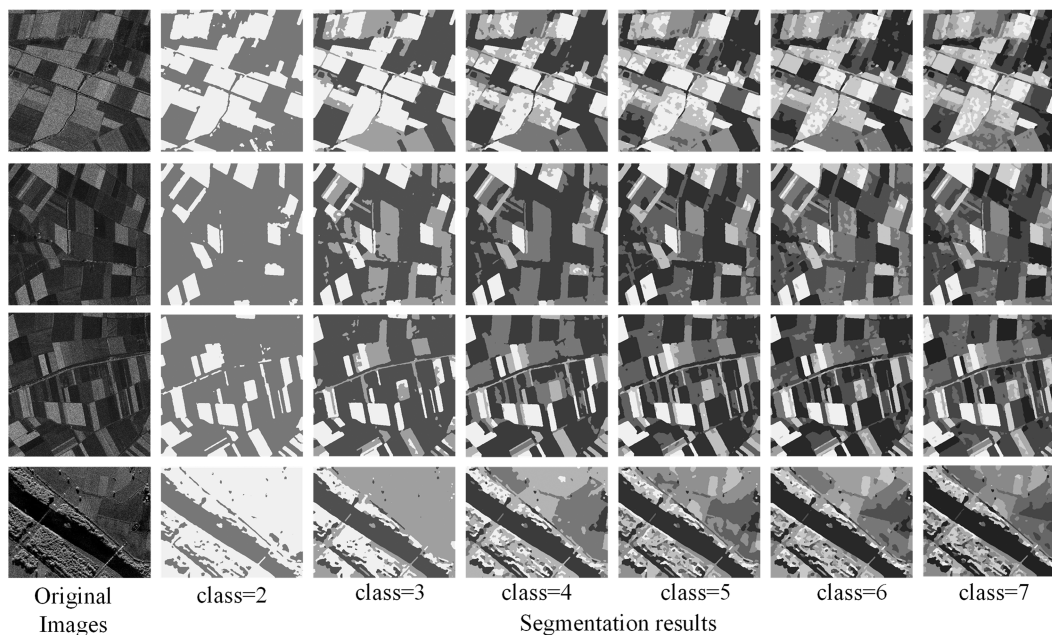| Original Images | class=2 | class=3 | class=4 | class=5 | class=6 | class=7 |
|---|---|---|---|---|---|---|

Segmentation results

Fig. 13.    Seg-results of A1–A4 with different classes by the RSLC.

(oversegmentation) to access the effectiveness of the proposed measure across SAR images. The part of seg-results of each testing image, corresponding to classes 2 and 7, is displayed in Fig. 13. Different gray levels in the seg-results of Fig. 13 indicate different segmentation classes.

We calculate the *GHO*, *GHE*, and *EVI* for all seg-results of A1–A4 with classes 2–10, according to (10), (14), and (18). Then, the *GHO*, *GHE*, and *EVI* of each seg-result are combined into *G* according to (19). Fig. 14 shows that these three indicators appropriately reflect the variation at different segmentation classes for seg-results. And *G* values in Fig. 14 also reflect the quality of seg-results. As the segmentation class increases, *GHO* and *EVI* keep decreasing and the value of *GHE* keeps increasing. The reason for generating the results in Fig. 14 is the seg-results of A1-A4 experienced a change from under-segmentation to over-segmentation as the class increased from 2 to 10. In Fig. 13, the number of segments increases in the process of tending to oversegmentation. The homogeneity within segments is increasing, the heterogeneity between segments is decreasing, and segmentation edges tend to occur with more false edges. The curves of each metric in Fig. 14 correctly and clearly reflect the variation in homogeneity and heterogeneity of the different seg-results as well as the segmentation edges.

Meanwhile, A1 and A4 have three class targets, and A2 and A3 have four in actuality. The *G* values in Fig. 14 also reflect the seg-results of the optimal class corresponding to each set of seg-results, which are consistent with human subjective conclusions. Thus, the validity of each component of the proposed metrics and the combination strategy are verified.

*2) Effectiveness of the G Metric:* The 1-10Look images of T1–T3 and S1–S3 are segmented using three segmentation algorithms to obtain 180 seg-results. The qualities of all seg-results were evaluated by UE metrics *E*, *F*, *Zeb*, and *G*, respectively. The SE metric *SA* was calculated for each SIS result as a precise reference to test the performance of the UE methods. The SE method is calculated based on GT1–GT3. Higher values of *SA* show better segmentation quality for the same image. However, the smaller the value of UE metric *G*, the better the segmentation quality. Therefore, the reciprocal of UE metric values is calculated to make the higher UE metric values correspond to better segmentation. Thus, from this part until Section IV, the values of *E*, *F*, *Zeb*, and *G* are represented as the reciprocal of their original values, respectively.

First, the experiments are performed on the seg-results of synthetic SAR images T1–T3. The part 1-10Look images of T1–T3 and the three seg-results of each image are shown in
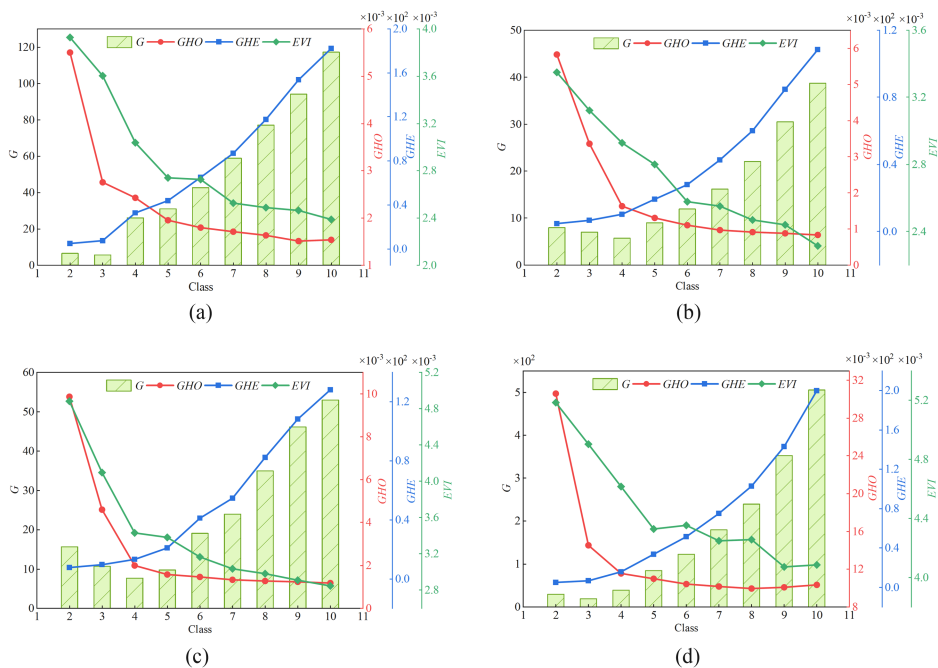
Fig. 14.    Results of *GHO*, *GHE*, *EVI*, and *G* for seg-results of Fig. 13. (a)–(d) Results of A1–A4.
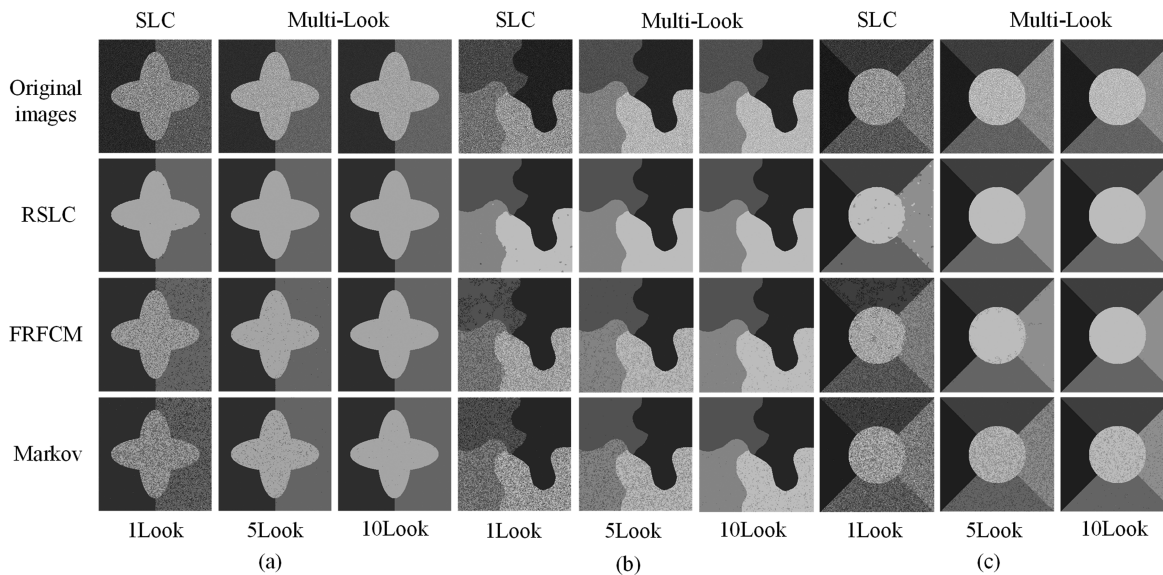


Fig. 15.    Seg-results of T1–T3 with different algorithms.

Fig. 15. The *SA*, *E*, *F*, *Zeb*, and *G* curves are plotted in Fig. 16 for all seg-results. As shown in Fig. 15, the seg-results of RSLC are the best when Look $< 5$, whereas the seg-results of RSLC are better than FRFCM, and the Markov seg-results are the worst. The seg-results are all acceptable and near to perfect segmentation when Look $\geq 5$.

The *SA* curves of the synthetic image seg-results are shown in Fig. 16 (*SA*). The *SA* curves of the seg-results accurately reflect the human observation segmentation quality. Therefore, the SE metric curves are a high reference to verify the UE methods. The *SA* values of RSLC results are the biggest in the same image.

For the multiple seg-results evaluations' of the same image, a well-performing UE metric deserves the same conclusion as the *SA* metric. The *F* and *G* values of RSLC results are higher than FRFCM, and the Markov results are the lowest when Look $< 6$. The *F* and *G* values of all results are almost equal when Look $\geq 6$. The evaluation results of *F* and *G* are similar to the *SA*. However, the relative magnitudes of the *E* and *Zeb* curves do not consist of the *SA* curve. Therefore, the *E* and *Zeb* metrics do not correctly evaluate the quality of the same image seg-results. The experiments are next performed on the seg-results of synthetic texture SAR images S1–S3. The part 1-10Look images of S1–S3
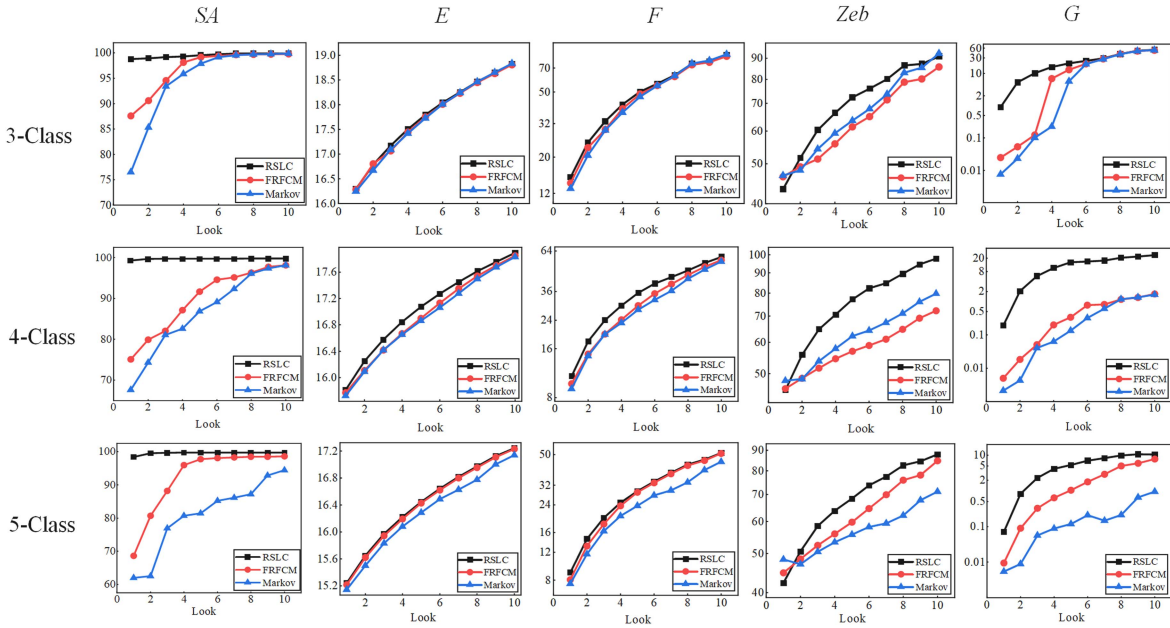
Fig. 16.    Segmentation evaluation results of T1–T3 with different algorithms.
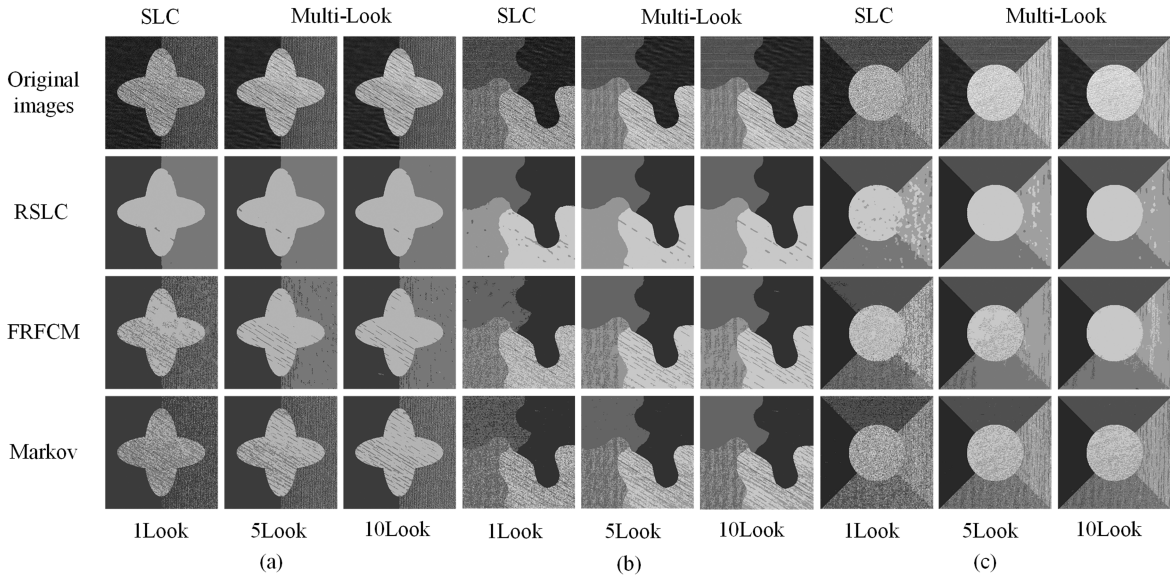


Fig. 17.    Seg-results of S1–S3 with different algorithms.

and the three seg-results of each image are shown in Fig. 17. The *SA*, *E*, *F*, *Zeb*, and *G* values of each seg-result are calculated and the curves are plotted in Fig. 18.

As shown in Fig. 17, the seg-results of RSLC are better than FRFCM, and the Markov seg-results are not superior. The *SA* curves of the images (3-Class) seg-results are shown in Fig. 18 (*SA*). The *SA* values of RSLC results are higher than FRFCM while the Markov is the lowest. The *G* values of RSLC results are higher than FRFCM, while the Markov is the lowest. Therefore, the relative magnitudes between the curves of *G* are similar to the *SA* curves. However, among the other three UE metrics *E*, *F*, and *Zeb,* none of them correctly evaluate the strengths and weaknesses of the same image seg-results.

The proposed UE metric demonstrates the superiority in evaluating seg-results by analyzing the evaluation result curves above. The *G* can obtain conclusions consistent with the SE method in qualitative analysis for each class of image segmentation. It is not difficult to find that the proposed metric *G* is independent of the number of segmentation classes. This further demonstrates the better applicability of our method. The Pearson correlation coefficient is implemented to further verify the accuracy of the proposed method quantitatively in evaluating seg-results. The coefficients calculate the correlation between the curves of UE metrics and SE metrics in Figs. 16 and 18. The correlation value statistics of each group is given in Table III. The correlation value statistics of each group is given in Table
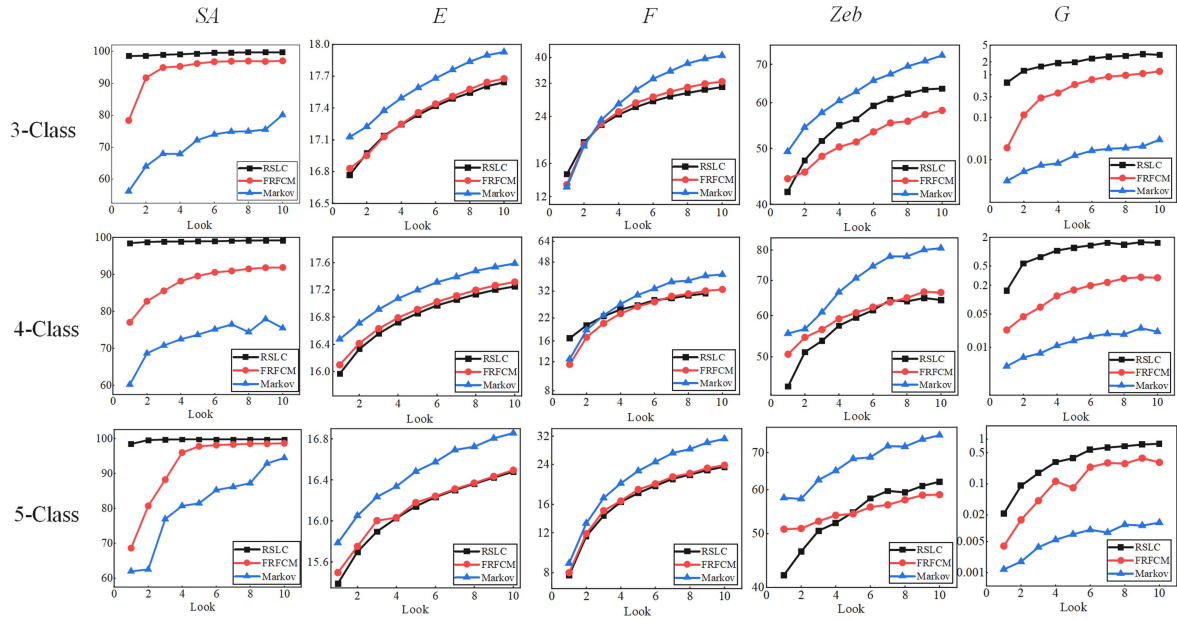
Fig. 18.    Segmentation evaluation results of S1–S3 with different algorithms.

TABLE III
PEARSON CORRELATION COEFFICIENT BETWEEN EACH UE METHOD AND SE METHOD *SA*

| Seg-results of 3 algorithms | | $E$ | $F$ | $Zeb$ | $G$ |
|---|---|---|---|---|---|
| Synthetic intensity SAR images (1-10Look) | T1 (3-Class) | 0.6005 | **0.6693** | 0.6301 | 0.6527 |
| | T2 (4-Class) | 0.6959 | 0.6969 | 0.6981 | **0.7011** |
| | T3 (5-Class) | 0.6343 | 0.6383 | 0.6502 | **0.6842** |
| Synthetic texture SAR images (1-10Look) | S1 (3-Class) | -0.0138 | 0.0416 | -0.2462 | **0.7214** |
| | S2 (4-Class) | -0.1137 | 0.0676 | -0.3204 | **0.7549** |
| | S3 (5-Class) | 0.0534 | 0.1737 | -0.3426 | **0.6899** |

III. The bold values in Table III indicate that the corresponding UE metric has the highest correlation with the SE metric for experimental results.

The proposed methods rank second in correlation with SE methods for the evaluation of gray images [T1(1-10Look)] seg-results while having the strongest correlation for all other categories (T2 and T3) of seg-results. Our method has the strongest correlation with the SE method in all segmentation categories for texture images, and the correlation of other UE methods with the SE method is a weakness.

Finally, experiments are conducted on real images to verify the validity of the proposed metric $G$. We apply the algorithm FRFCM to A1 and A3 to obtain the seg-results on real SAR images. The w is a significant parameter for FRFCM, which represents the size of the filtering window. The various w can affect the quality of seg-results.

We set the w to take values in the range [1], [19] with a step size of 2 in the process of obtaining the seg-results. FRFCM segmented A1 and A3 to get ten seg-results according to the images shown in Fig. 19. Obtaining the GT of a real SAR image is a certain degree subjective and time-consuming. We cannot evaluate the quality of seg-results with the SE metric

SA. Therefore, four UE metrics including the proposed method were used to evaluate the seg-results in experiments, and the results were shown in Fig. 20.

The curve of $G$ increases first and then decreases with increasing w in Fig. 20. It is obvious that the seg-results of A1 and A3 gradually become better from undersegmentation with w increasing from 1 to 5 and become oversegmentation after W > 7 in Fig. 23. The curve of $G$ is highly consistent with the results of the manual evaluation mentioned above in Fig. 20, which can evaluate the quality of segmentation accurately.

### E. Equivalent Application Experiments of the Proposed Method and SE Method in Parameter Selection

An important application of segmentation evaluation metrics is to help segmentation algorithms choose appropriate parameters. We hope the proposed UE metric is equivalent to the SE metric in helping parameter selection applications. UE is important for the automatic selection of optimal parameters in SIS [11]. RSLC is a novel segmentation algorithm explicitly designed for SAR images. There is a parameter of RSLC that needs to be selected: majority voting sliding window W. The accuracy of
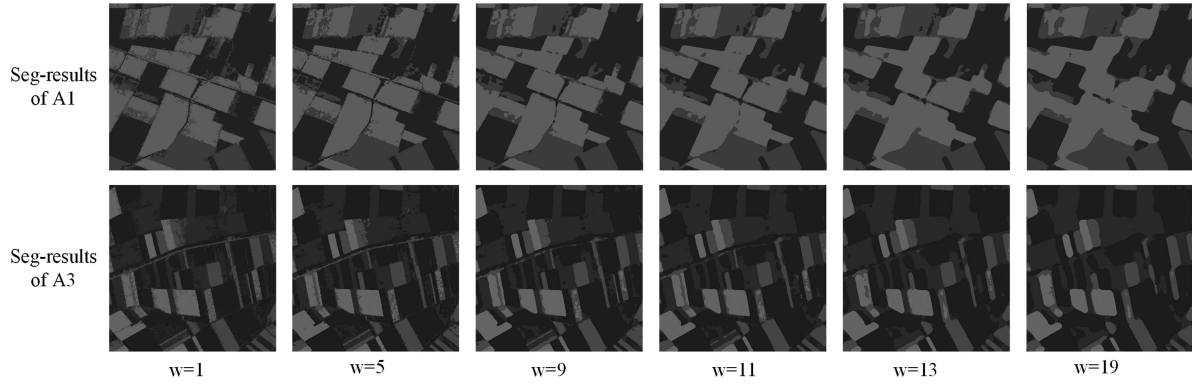
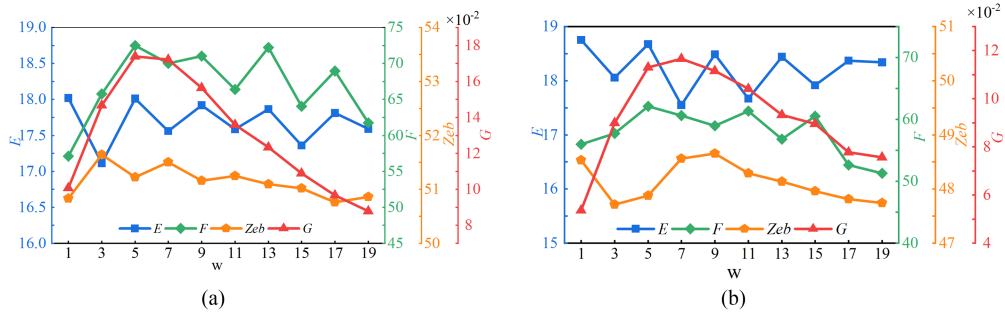Fig. 19. Seg-results obtained by FRFCM segmenting images A1 and A3 with different w.



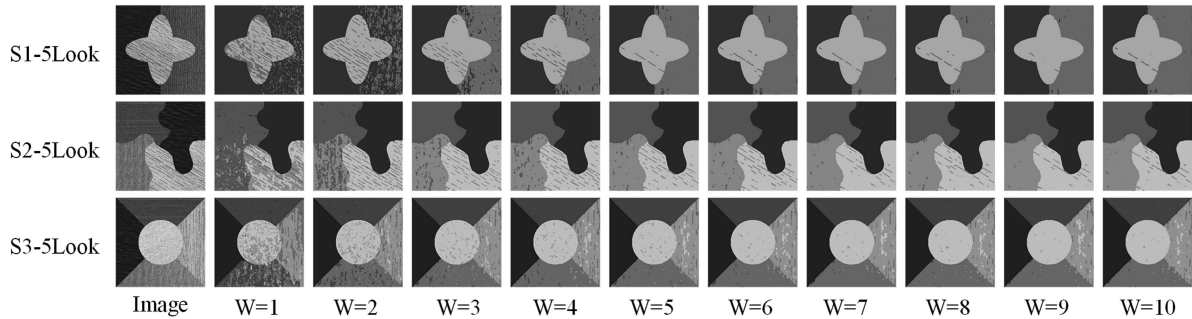Fig. 20. Comparative results of real SAR image A1 and A3 seg-results evaluation.



Fig. 21. Seg-results of RSLC with different W.

the seg-results is affected by W. Therefore, experiments were designed to verify the equivalence of using the $G$ metric to the SE metric in selecting the parameters of the segmentation algorithm.

In experiments, the RSLC is used to segment the image by varying the parameter W from 1 to 10 and keeping the other parameters constant. Then, ten seg-results with different W can be obtained for the same image. The above segmentation was performed on S1-5Look, S2-5Look, and S3-5Look. Then, 30 seg-results were obtained. The original images and seg-results with different W are shown in Fig. 21.

The $SA, E, F, Zeb$, and $G$ values of all seg-results are calculated and plotted as the curves in Fig. 22 to determine the setting of parameter W for the RSLC algorithm. As shown in Fig. 21, the quality of seg-results changes from poor to excellent and remains constant with the increase of W. The value of $SA$ increases when W < 5, and the value of $SA$ is almost constant when W $\geq$ 5. The

ideal seg-results are obtained at the same time. The value change in UE metric $G$ also shows that the value of $G$ hardly changes when W $\geq$ 5, and the ideal seg-results are obtained. Therefore, the parameter selection conclusions of the proposed metric are consistent with the conclusions obtained by the SE metric. The other UE metrics cannot correctly evaluate the pros and cons of the seg-results, as well as have different degrees of change with the increase of the W. The appropriate parameters cannot be obtained. To quantitatively compare the performance of $G$, $E, F$, and $Zeb$, the Pearson correlation coefficient was calculated between each UE method and SE method $SA$. The correlation experimental results are given in Table IV.

As shown in Fig. 22, methods $E$ and $Zeb$ perform poorly on the results of RSLC with different W for all testing images.

The values of $E$ and $Zeb$ are unsteady with the increase of W. Therefore, the correlation is weak and unsteady in Table IV.
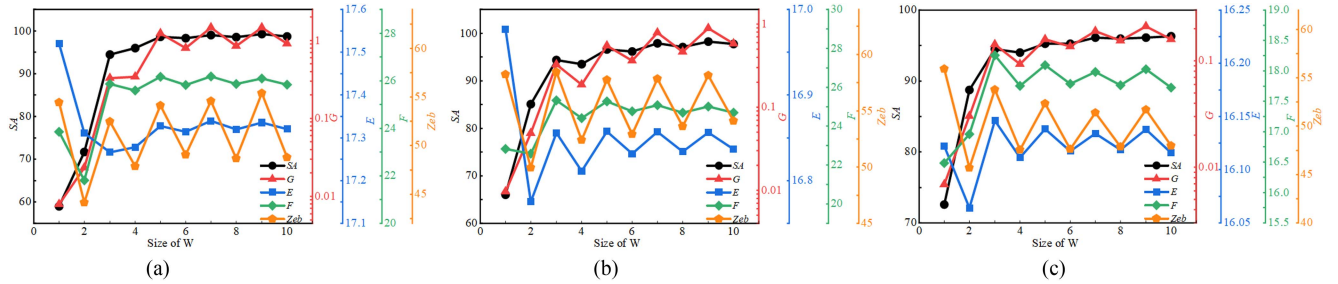
Fig. 22. Comparing $G$ with current evaluation methods for testing images. (a) Evaluation for S1-5Look seg-results. (b) Evaluation for S2-5Look seg-results. (c) Evaluation for S3-5Look seg-results.

TABLE IV
PEARSON CORRELATION COEFFICIENT BETWEEN EACH UE METHOD AND SE METHOD $SA$

| Testing Image | $E$ | $F$ | $Zeb$ | $G$ |
|---|---|---|---|---|
| S1-5Look | -0.72 | 0.84 | 0.11 | **0.99** |
| S2-5Look | -0.63 | 0.82 | -0.06 | **0.99** |
| S3-5Look | 0.22 | 0.89 | -0.47 | **0.99** |

The $F$ performs well on the results of RSLC with different W for S3-5Look [see Fig. 22(c)] but performs not enough on the results for S1-5Look and S2-5Look [see Fig. 22(a) and (b)]. The $F$ has the second-rank correlation with the SA metric. The curves of $G$ in Fig. 22 are similar to SA for all testing images. The $G$ always generates higher correlations with the SE method SA than the others for RSLC in experiments. Therefore, studies show that the proposed UE metric achieves the effects of the SE metric in the parameters set.

The "saw" shape of $G$ is not fine in Fig. 22 for the case of selecting the "W" parameter for RSLC when it is compared with the metric SA. The reason is that the UE metric does not provide better evaluation accuracy compared to SE methods due to the lack of comparison of GT images. The SIS result changes less with the increase in W, the SIS results quality looks similar (W $\geq$ 5) to Fig. 21, and in theory, the evaluation metrics should produce similar values. The SE metric SA can obtain high accuracy results with the GT, and its "saw" shape looks fine (W $\geq$ 5) in Fig. 22. However, the UE metric $G$ does not rely on GT, and the evaluation accuracy of the seg-results is almost lacking; its "saw" shape looks not fine (W $\geq$ 5) in Fig. 22. We wish all SAR images to have a suitable GT so that we can easily choose parameters with the help of SE metrics, but most of the SAR images without a suitable GT and, in this case, the SE metrics fail. The experimental results shown in Figs. 19 and 20 can show that the proposed UE metric can help the segmentation algorithm to select the appropriate parameters in the absence of GT.

### F. Experiments for Evaluating the Seg-Results of Deep Neural Networks Based Algorithm

The research on image segmentation based on DNN is a hot topic. Thus, we conduct experiments on the SAR image seg-results of the segmentation algorithm based on DNN. The UISB is an unsupervised segmentation algorithm based on DNN [56]. UISB first performs superpixel segmentation on the input image, then extracts features using a neural network, next performs superpixel merging by clustering and calculates the loss, and finally updates the parameters using stochastic gradient descent.

The authors of UISB use SILC [57] to generate superpixels, and a scholar has changed the original SILC to Felzenszwalb [58] algorithm to obtain fel-UISB. After experiments, fel-UISB has better segmentation quality on SAR images. Therefore, we use the proposed metric to evaluate the seg-results produced by fel-UISB for experiments. The scale parameter of Felzenszwalb affects the generated superpixels, which in turn affects the quality of the seg-results. We segmented the A3 by setting the scale to take a range of 32–76, step = 4, and obtained 12 seg-results, as shown in Fig. 23. The seg-results are evaluated using $G$ and other UE metrics, and the results are shown in Fig. 24.

Fig. 24 shows that the quality of the seg-results of A3 varies with the scale by different UE metrics. We can see the forest and farmland areas that are better segmented from the red box in Fig. 23. The seg-results of Fig. 23(b) and (k) are the best in all seg-results. And the $G$ values in Fig. 24 for scale = 72 and 36 are distinctly higher than the values of the other seg-results. The $G$ is low for scale = 68, indicating its poor segmentation quality. Since the forest region within the blue box in Fig. 23(j) is poorly segmented, splitting it incorrectly into three different classes results in less interheterogeneity, thus strongly affecting the $G$ value. The experimental results again show that $G$ can obtain conclusions that are consistent with human vision. Our method is more focused on the evaluation of seg-results and has no evident relationship with the segmentation algorithms.

### IV. DISCUSSION

It is worth noting that for different metrics we choose different distance measures, such as $v_k^g$, $v_k^t$, $S_{kd}^g$, and $S_{kd}^t$, which are
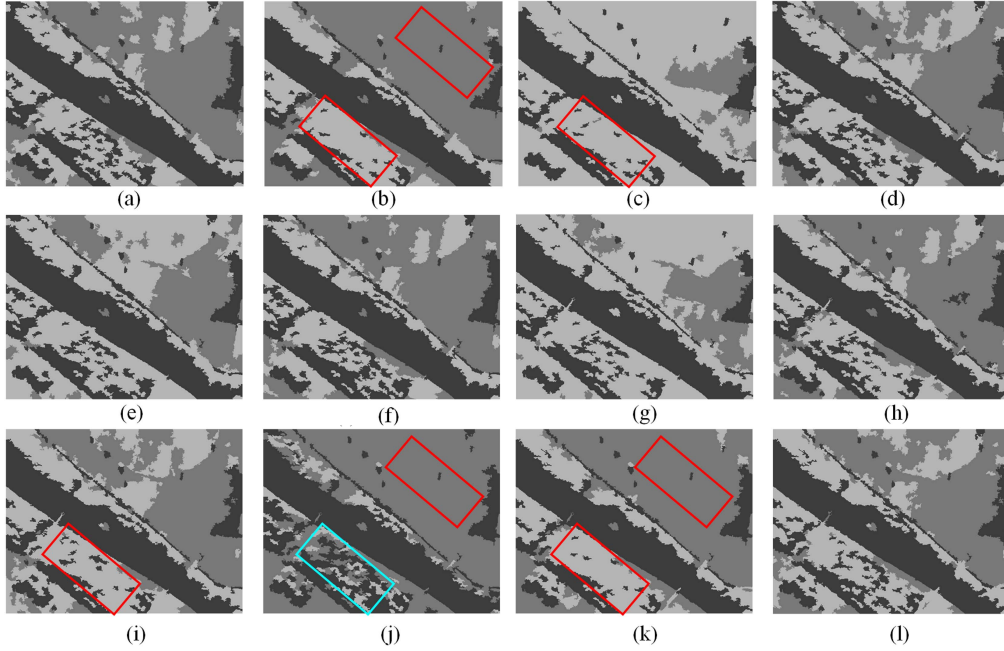
Fig. 23. Seg-results of A3 were obtained by fel-UISB with different parameters of Felzenszwalb scale. (a) Scale = 32. (b) Scale = 36. (c) Scale = 40. (d) Scale = 44. (e) Scale = 48. (f) Scale = 52. (g) Scale = 56. (h) Scale = 60. (i) Scale = 64. (j) Scale = 68. (k) Scale = 72. (l) Scale = 76.
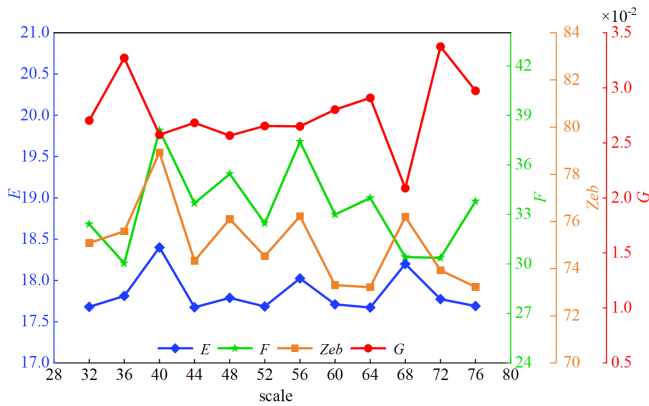


Fig. 24. Different UE metrics of seg-results are in Fig. 23.

chosen according to the data characteristics. The range of values and the distribution are different in texture features $\mathbf{x}_i^T$, and the texture homogeneity within the region can be described by chi-squared test measures after $\mathbf{x}_{\max}^T$ is obtained. The $f_k$ contains data with different scales, $E_{kn}$, and $\sigma_{kn}$, and the Canberra distance is insensitive to the data scale.

One of the issues with the proposed method is that it is sensitive to small segmentation regions due to the proposed metrics being represented based on global homogeneity and heterogeneity. This leads to the metric not paying enough attention to the local fine segmentation regions, especially the wrong segmentation regions. The issue needs continued focus and resolution in the next studies.

The proposed method is mainly based on SAR images that are widely used in SIS as stated in Section III-A. We establish a general framework for evaluating SIS results from the characteristics of SAR images. Therefore, for different types of SAR

data, such as multiband and multipolarized, the application of this general framework can extend the study of this article to other data types.

We proposed a UE metric that does not rely on GT to evaluate the quality of SIS results and helps in algorithm optimization and parameter selection. However, since it does not rely on GT images, the accuracy of our method is not as good as the SE method for similar seg-results. Therefore, developing the study of UE metric accuracy is important in the next work.

In further studies, more UE metrics and methods for SIS results should be paid more attention to as follows.

1) We can explore more ways to effectively characterize the heterogeneous features of SAR images and the methods of features fusion due to SAR images containing rich heterogeneous features.
2) We can design intrahomogeneity metrics and interheterogeneity metrics in different perspectives and more effective fusion strategies.
3) We can refine the evaluation contents, such as over- and undersegmentation evaluation, to focus on the local segmentation.
4) We can improve the accuracy of UE metrics in evaluating similar seg-results, which may rely on more refined features and combination strategies.

## V. CONCLUSION

This study aims to promote the development of SIS and segmentation evaluation. A heterogeneity-features-based novel UE metric was proposed for evaluating the SAR image seg-results and helping in algorithm optimization and parameter selection. In this method, we proposed an IFEE based on edge-hold and

used other two fruitful methods to extract heterogeneous features of SAR images, which include intensity, texture, and edge features. Then, we designed *GHO*, *GHE*, and *EVI* metrics based on heterogeneity features to reveal local segmentation quality. Finally, we use a strategy to fuse these metrics into a novel global evaluation metric ($G$) to indicate the quality of SAR image seg-results.

The RSLC, FRFCM, Markov, and UISB algorithms were applied to 60 synthesized SAR images and 4 real SAR images to produce over 200 seg-results for experiments. The effectiveness of the metric $G$ is further demonstrated by comparing three existing UE metrics and one SE metric. The results show the effectiveness and superiority of the proposed metric. Moreover, the mean correlation between the proposed UE metric and the SE metric is more than 0.67 and 0.98, which indicates that the proposed metric helps in choosing parameters of different algorithms without GT.

## REFERENCES

[1] A. Moreira, P. Prats-Iraola, M. Younis, G. Krieger, I. Hajnsek, and K. P. Papathanassiou, "A tutorial on synthetic aperture radar," *IEEE Geosci. Remote Sens. Mag.*, vol. 1, no. 1, pp. 6–43, Mar. 2013.

[2] D. Xiang, T. Tang, C. Hu, Y. Li, and Y. Su, "A kernel clustering algorithm with fuzzy factor: Application to SAR image segmentation," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 7, pp. 1290–1294, Jul. 2014.

[3] C. Oliver and S. Quegan, *Understanding Synthetic Aperture Radar Images*. Rijeka, Croatia: SciTech, 2004.

[4] F. Liu et al., "SAR image segmentation based on hierarchical visual semantic and adaptive neighborhood multinomial latent model," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 7, pp. 4287–4301, Jul. 2016.

[5] Y. Duan, F. Liu, L. Jiao, P. Zhao, and L. Zhang, "SAR image segmentation based on convolutional-wavelet neural network and Markov random field," *Pattern Recognit.*, vol. 64, pp. 255–267, 2017.

[6] R. H. Nobre, F. A. Rodrigues, R. C. P. Marques, J. S. Nobre, J. F. S. R. Neto, and F. N. Medeiros, "SAR image segmentation with Renyi's entropy," *IEEE Signal Process. Lett.*, vol. 23, no. 11, pp. 1551–1555, Nov. 2016.

[7] L. Jiao, M. Gong, S. Wang, B. Hou, Z. Zheng, and Q. Wu, "Natural and remote sensing image segmentation using memetic computing," *IEEE Comput. Intell. Mag.*, vol. 5, no. 2, pp. 78–91, May 2010.

[8] H. Yu, X. Zhang, S. Wang, and B. Hou, "Context-based hierarchical unequal merging for SAR image segmentation," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 2, pp. 995–1009, Feb. 2013.

[9] L. Jiao, X. Tang, B. Hou, and S. Wang, "SAR images retrieval based on semantic classification and region-based similarity measure for earth observation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 8, pp. 3876–3891, Aug. 2015.

[10] U. Javed, M. M. Riaz, A. Ghafoor, and T. A. Cheema, "SAR image segmentation based on active contours with fuzzy logic," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 52, no. 1, pp. 181–188, Feb. 2016.

[11] B. Johnson and Z. Xie, "Unsupervised image segmentation evaluation and refinement using a multi-scale approach," *ISPRS J. Photogramm. Remote Sens.*, vol. 66, no. 4, pp. 473–483, 2011.

[12] Y. Gao, J. F. Mas, N. Kerle, and J. A. Navarrete Pacheco, "Optimal region growing segmentation and its effect on classification accuracy," *Int. J. Remote Sens.*, vol. 32, no. 13, pp. 3747–3763, 2011.

[13] L. Drăguţ, O. Csillik, C. Eisank, and D. Tiede, "Automated parameterisation for multi-scale image segmentation on multiple layers," *ISPRS J. Photogramm. Remote Sens.*, vol. 88, pp. 119–127, 2014.

[14] H. Zhang, J. E. Fritts, and S. A. Goldman, "Image segmentation evaluation: A survey of unsupervised methods," *Comput. Vis. Image Understanding*, vol. 110, no. 2, pp. 260–280, 2008.

[15] R. Shang, J. Lin, L. Jiao, and Y. Li, "SAR image segmentation using region smoothing and label correction," *Remote Sens.*, vol. 12, no. 5, 2020, Art. no. 803.

[16] R. Shang, Y. Yuan, L. Jiao, B. Hou, A. M. G. Esfahani, and R. Stolkin, "A fast algorithm for SAR image segmentation based on key pixels," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 12, pp. 5657–5673, Dec. 2017.

[17] H. Yu, H. Jiang, Z. Liu, Y. Sun, S. Zhou, and Q. Gou, "SAR image segmentation by merging multiple feature regions," in *Proc. 3rd Int. Conf. Geol., Mapping Remote Sens.*, 2022, pp. 183–186.

[18] G. Akbarizadeh, G. A. Rezai-Rad, and S. B. Shokouhi, "A new region-based active contour model with skewness wavelet energy for segmentation of SAR images," *IEICE Trans. Inf. Syst.*, vol. 93, no. 7, pp. 1690–1699, 2010.

[19] Z. Tirandaz and G. Akbarizadeh, "A two-phase algorithm based on kurtosis curvelet energy and unsupervised spectral regression for segmentation of SAR images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 3, pp. 1244–1264, Mar. 2016.

[20] M. Rahmani and G. Akbarizadeh, "Unsupervised feature learning based on sparse coding and spectral clustering for segmentation of synthetic aperture radar images," *IET Comput. Vis.*, vol. 9, no. 5, pp. 629–638, 2015.

[21] Z. Tirandaz, G. Akbarizadeh, and H. Kaabi, "PolSAR image segmentation based on feature extraction and data compression using weighted neighborhood filter bank and hidden Markov random field-expectation maximization," *Measurement*, vol. 153, 2020, Art. no. 107432.

[22] N. Aghaei, G. Akbarizadeh, and A. Kosarian, "GreyWolfLSM: An accurate oil spill detection method based on level set method from synthetic aperture radar imagery," *Eur. J. Remote Sens.*, vol. 55, no. 1, pp. 181–198, 2022.

[23] F. Sharifzadeh, G. Akbarizadeh, and Y. Seifi Kavian, "Ship classification in SAR images using a new hybrid CNN–MLP classifier," *J. Ind. Soc. Remote Sens.*, vol. 47, pp. 551–562, 2019.

[24] N. Davari, G. Akbarizadeh, and E. Mashhour, "Corona detection and power equipment classification based on GoogleNet-AlexNet: An accurate and intelligent defect detection model based on deep learning for power distribution lines," *IEEE Trans. Power Del.*, vol. 37, no. 4, pp. 2766–2774, Aug. 2022.

[25] N. Aghaei, G. Akbarizadeh, and A. Kosarian, "Osdes_net: Oil spill detection based on efficient_shuffle network using synthetic aperture radar imagery," *Geocarto Int.*, vol. 37, pp. 1–22, 2022.

[26] P. L. Correia and F. Pereira, "Stand-alone objective segmentation quality evaluation," *EURASIP J. Adv. Signal Process.*, vol. 2002, 2002, Art. no. 431748.

[27] S. Chabrier, B. Emile, H. Laurent, C. Rosenberger, and P. Marche, "Unsupervised evaluation of image segmentation application to multi-spectral images," in *Proc. 17th Int. Conf. Pattern Recognit.*, 2004, vol. 1, pp. 576–579.

[28] M. C. Shin, D. B. Goldgof, K. W. Bowyer, and S. Nikiforou, "Comparison of edge detection algorithms using a structure from motion task," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 31, no. 4, pp. 589–601, Aug. 2001.

[29] J. S. Cardoso and L. Corte-Real, "Toward a generic evaluation of image segmentation," *IEEE Trans. Image Process.*, vol. 14, no. 11, pp. 1773–1782, Nov. 2005.

[30] S. Chabrier, H. Laurent, B. Emile, C. Rosenberger, and P. Marche, "A comparative study of supervised evaluation criteria for image segmentation," in *Proc. 12th Eur. Signal Process. Conf.*, 2004, pp. 1143–1146.

[31] P. Correia and F. Pereira, "Objective evaluation of relative segmentation quality," in *Proc. Int. Conf. Image Process.*, 2000, vol. 1, pp. 308–311.

[32] M. Neubert, H. Herold, and G. Meinel, "Assessing image segmentation quality—Concepts, methods and application," in *Object-Based Image Analysis*. Berlin, Germany: Springer, 2008, pp. 769–784.

[33] X. Zhang, X. Feng, P. Xiao, G. He, and L. Zhu, "Segmentation quality evaluation using region-based precision and recall measures for remote sensing images," *ISPRS J. Photogramm. Remote Sens.*, vol. 102, pp. 73–84, 2015.

[34] Y. Chen, D. Ming, L. Zhao, B. Lv, K. Zhou, and Y. Qing, "Review on high spatial resolution remote sensing image segmentation evaluation," *Photogramm. Eng. Remote Sens.*, vol. 84, no. 10, pp. 629–646, 2018.

[35] S. S. Chouhan, A. Kaul, and U. P. Singh, "Soft computing approaches for image segmentation: A survey," *Multimedia Tools Appl.*, vol. 77, no. 21, pp. 28483–28537, 2018.

[36] S. Böck, M. Immitzer, and C. Atzberger, "On the objectivity of the objective function—Problems with unsupervised segmentation evaluation based on global score and a possible remedy," *Remote Sens.*, vol. 9, no. 8, 2017, Art. no. 769.

[37] Z. Wang, E. Wang, and Y. Zhu, "Image segmentation evaluation: A survey of methods," *Artif. Intell. Rev.*, vol. 53, no. 8, pp. 5637–5674, 2020.

[38] S. Chabrier, B. Emile, C. Rosenberger, and H. Laurent, "Unsupervised performance evaluation of image segmentation," *EURASIP J. Adv. Signal Process.*, vol. 2006, 2006, Art. no. 096306.

[39] O. Kubassova, M. Boesen, and H. Bliddal, "General framework for unsupervised evaluation of quality of segmentation results," in *Proc. IEEE 15th Int. Conf. Image Process.*, 2008, pp. 3036–3039.

[40] R. M. Haralick and L. G. Shapiro, "Image segmentation techniques," *Comput. Vis., Graph., Image Process.*, vol. 29, no. 1, pp. 100–132, 1985.

[41] A. W. Doerry, "SAR image scaling dynamic range radiometric calibration and display," SAND-2019-2371, Sandia National Lab., Albuquerque, NM, USA, 2019.

[42] B. Zhengyao, L. Zhoufeng, and H. Peikun, "SAR image target segmentation based on entropy maximization and morphology," *J. Syst. Eng. Electron.*, vol. 15, no. 4, pp. 484–487, Dec. 2004.

[43] D. A. Clausi and M. E. Jernigan, "Designing Gabor filters for optimal texture separability," *Pattern Recognit.*, vol. 33, no. 11, pp. 1835–1849, 2000.

[44] D. A. Clausi and H. Deng, "Design-based texture feature fusion using Gabor filters and co-occurrence probabilities," *IEEE Trans. Image Process.*, vol. 14, no. 7, pp. 925–936, Jul. 2005.

[45] D. A. Clausi, "Comparison and fusion of co-occurrence, Gabor and MRF texture features for classification of SAR sea-ice imagery," *Atmos. Ocean*, vol. 39, no. 3, pp. 183–194, 2001.

[46] T. S. Lee, "Image representation using 2D Gabor wavelets," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 10, pp. 959–971, Oct. 1996.

[47] B. K. Patra, R. Launonen, V. Ollikainen, and S. Nandi, "A new similarity measure using Bhattacharyya coefficient for collaborative filtering in sparse data," *Knowl.-Based Syst.*, vol. 82, pp. 163–177, 2015.

[48] S. M. Singh and K. Hemachandran, "Content-based image retrieval using color moment and Gabor texture feature," *Int. J. Comput. Sci. Issues*, vol. 9, no. 5, 2012, Art. no. 299.

[49] J. Zhu, F. Wang, and H. You, "SAR image segmentation by efficient fuzzy C-means framework with adaptive generalized likelihood ratio nonlocal spatial information embedded," *Remote Sens.*, vol. 14, no. 7, 2022, Art. no. 1621.

[50] R. Shang et al., "SAR image segmentation based on constrained smoothing and hierarchical label correction," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5102216.

[51] R. Shang et al., "A spatial fuzzy clustering algorithm with kernel metric based on immune clone for SAR image segmentation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 4, pp. 1640–1652, Apr. 2016.

[52] H. Gao, Y. Tang, L. Jing, H. Li, and H. Ding, "A novel unsupervised segmentation quality evaluation method for remote sensing images," *Sensors*, vol. 17, no. 10, 2017, Art. no. 2427.

[53] M. Hussain, D. Chen, A. Cheng, H. Wei, and D. Stanley, "Change detection from remotely sensed images: From pixel-based to object-based approaches," *ISPRS J. Photogramm. Remote Sens.*, vol. 80, pp. 91–106, 2013.

[54] T. Lei, X. Jia, Y. Zhang, L. He, H. Meng, and A. K. Nandi, "Significantly fast and robust fuzzy c-means clustering algorithm based on morphological reconstruction and membership filtering," *IEEE Trans. Fuzzy Syst.*, vol. 26, no. 5, pp. 3027–3041, Oct. 2018.

[55] Z. Tu and S.-C. Zhu, "Image segmentation by data-driven Markov chain Monte Carlo," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 5, pp. 657–673, May 2002.

[56] A. Kanezaki, "Unsupervised image segmentation by backpropagation," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2018, pp. 1543–1547.

[57] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2274–2282, Nov. 2012.

[58] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient graph-based image segmentation," *Int. J. Comput. Vis.*, vol. 59, no. 2, pp. 167–181, 2004.

**Xiangjie Yin** received the B.E. degree in electronic information engineering from Linyi University, Linyi, China, in 2020. He is currently working toward the M.S. degree with the School of Aerospace Science and Technology, Xidian University, Xi'an, China.

His research interests include SAR image segmentation and evaluation metric designing, image processing, artificial intelligence, and remote sensing.



**Zhiheng Liu** received the Ph.D. degree in photogrammetry and remote sensing from the School of Geological Engineering and Geomatics, Chang'an University, Xi'an, China, in 2021.

He was a joint student with Satellite Surveying and Mapping Application Center in China from 2015 to 2016 and a visiting Ph.D. student with the University of Plymouth, United Kingdom, from 2019. He has been a Postdoc with the School of Aerospace Science and Technology, Xidian University, China, since 2021. His current research interests include remote sensing, computer vision, and deep learning.



**Suiping Zhou** received the B.E., M.A., and Ph.D. degrees in flight vehicle control, guidance, and simulation from Beihang University, Beijing, China, in 1989, 1992, and 1995, respectively.

He is currently a full Professor with the School of Aerospace Science and Technology, Xidian University, Xi'an, China. Previously, he was a full Professor with the Department of Computer Science, Middlesex University, United Kindom, and an Assistant Professor with the School of Computer Engineering, Nanyang Technological University, Singapore. His current research interests include large-scale distributed interactive applications, parallel/distributed systems, and big data in space science and technology.



**Chenyang Li** received the B.Eng. degree in detection guidance and control techniques from the North University of China, Taiyuan, China, in 2019. He is currently working toward the Ph.D. degree in instruments science and technology with the School of Aerospace Science and Technology, Xidian University, Xi'an, China.

His research interests include computer vision, image processing, and dehazing algorithm.



**Hang Yu** (Member, IEEE) received the B.E. degree in electronic and information engineering and the Ph.D. degree in circuits and systems from Xidian University, Xi'an, China, in 2005 and 2014, respectively.

Since 2017, he has been an Associate Professor with the School of Aerospace Science and Technology, Xidian University. He was a visiting Assistant Professor in Radiation Oncology, UT Southwestern Medical Center, from 2019 to 2020. His current research interests include image understanding and interpretation, computer vision, machine learning, information fusion, evolutionary computation and optimization, visual navigation, and SAR image processing.



**Haoran Jiang** received the B.E. degree in detection guidance and control techniques in 2018 from the School of Aerospace Science and Technology, Xidian University, Xi'an, China, where he is currently working toward the M.S. degree in mechanical engineering with the School of Aerospace Science and Technology.

His research interests include SAR image segmentation and evaluation metric designing, image processing, artificial intelligence, and remote sensing.