SAR image classification using deep features and neighborhood heterogeneity confidence refined spatial constraints

Anjun Zhang, Linsheng Huang, Lu Jia, Rui Ruan & Jun Wang

Published online: 14 Feb 2024.

Submit your article to this journal ↗

View related articles ↗

View Crossmark data ↗

Taylor & Francis
Taylor & Francis Group

Check for updates

SAR image classification using deep features and neighborhood heterogeneity confidence refined spatial constraints

Anjun Zhang[a], Linsheng Huang[a], Lu Jia [b], Rui Ruan[a] and Jun Wang[c]

[a]School of Internet, National Engineering Research Center for Agro-Ecological Big Data Analysis and Application, Anhui University, Hefei, Anhui, China; [b]School of Computer and Information, Anhui Province Key Laboratory of Industry Safety and Emergency and Technology, Hefei University of Technology, Hefei, Anhui, China; [c]School of Mechanical Engineering, Quzhou University, Quzhou, China

ABSTRACT

Recently, frameworks fusing the convolutional neural network (CNN) and Markov random field (MRF) models have been successfully used in synthetic aperture radar (SAR) image classification. Over-smoothing of the details and the incapability to eliminate regional-level misclassification are two common drawbacks of these frameworks, which are caused by the use of conventional neighbourhood consistency based spatial constraints. To address these problems, a neighbourhood heterogeneity confidence – Markov random field (NHC-MRF) and CNN-based framework is proposed for SAR image classification. In this framework, an NHC index is constructed in the binary energy function of the NHC-MRF to refine the conventional spatial constraints by adaptively allowing the existence of heterogeneity in the neighbourhood. The NHC index consists of a label coexisting probability term and a top-2 label term, which are constructed based on the outputs of the CNN. By using the label coexisting probability term, the details can be protected by allowing heterogeneity, and the regional-level misclassification can be eliminated by adjusting the impact of the neighbourhoods by using label coexisting probability based weighting. The top-2 label term is used as a correction of the coexisting probability term considering the disturbance of speckle noise. The NHC-MRF is further fused with a CNN by constructing the unary energy term and initial labels based on the outputs of the CNN. The effectiveness and superiority of the proposed framework are experimentally demonstrated using three SAR datasets. The experimental results demonstrate that the superiority of the NHC-MRF is derived from the simultaneous realization of retaining details and eliminating regional-level misclassification.

1. Introduction

Synthetic aperture radar (SAR) images can provide detailed information about land cover types under all-time and all-weather conditions. Due to this advantage, SAR images have been widely used in many fields. Consequently, it is important to effectively and accurately

interpret SAR images. Pixel-level SAR image classification can provide category labels for the pixels in a SAR image, which is the most fundamental step in SAR image classification. As a result, SAR image classification is a hot topic in the field of remote sensing (Dekker 2003; Ferro-Famil, Pottier, and Lee 2000; Fukuda and Hirosawa 1999; Maryam 2022; Modava, Akbarizadeh, and Soroosh 2019; Tison et al. 2004; Tzeng and Chen 1998).

SAR image classification algorithms can be broadly divided into two categories: supervised algorithms and unsupervised algorithms. Unsupervised algorithms generally explore clustering methods or the statistical properties of SAR images for classification (Xiang et al. 2014). Supervised algorithms are more frequently investigated because of their higher classification accuracies. Conventionally, supervised algorithms are performed by combining carefully designed or selected features and classifiers (Chapelle, Haffner, and Vapnik 1999; McNairn et al. 2014; Ressel, Frost, and Lehner 2015). The classification accuracy of supervised algorithms relies highly on manual feature designing. However, a common problem is that the carefully designed features are not discriminative and robust for SAR images with a high level of speckle noise (Lv et al. 2014). Aiming to resolve this problem, the strategy of automatic feature learning instead of feature designing should be applied. This deep learning technique has an outstanding feature learning ability, which has led to its extensive application in the field of SAR image classification. In recent years, various deep learning based SAR image classification algorithms have been proposed.

In the field of SAR image classification, convolutional neural networks (CNNs) (Krizhevsky, Sutskever, and Hinton 2012; Lécun et al. 1998; Szegedy et al. 2015) can automatically extract multiple local image features, which are more discriminative than manually designed features. Over the past few years, CNNs have been successfully applied in the field of SAR image classification (Duan et al. 2017; Jiang et al. 2021; Z. Zhang et al. 2017; Zhou et al. 2016). For example, Zhou et al. (2016) applied a CNN to SAR image classification for the first time and achieved a superior classification accuracy to that achieved using conventional algorithms. To improve the accuracy of locating the boundary, Duan et al. (2017) proposed the wavelet-pooling layer to reduce the destruction of structures. To accurately locate the boundaries, Zhang et al. (2019) proposed a speckle reducing anisotropic diffusion (SRAD) adaptive filtering layer to enhance the features of the boundaries. To utilize spatial information more abundantly, Jiang et al. (2021) proposed a hierarchical fusion CNN for SAR image classification. To explore both the spatial information and the polarimetric information in SAR images, Gao et al. (2017) proposed the dual-branch CNN to extract and fuse the deep spatial features and polarimetric features of SAR images. To fully use the polarimetric information, Zhang et al. (2017) and Cao et al. (2019) proposed the complex valued convolutional neural network (CV-CNN) and complex valued fully convolutional network (CV-FCN) to utilize the complex-value phase of the polarimetric components. The combination of the statistical information in SAR images with the deep spatial features of the CNN was also proposed. Liang et al. (2021) proposed the use of a covariance pooling layer to extract the second order global statistical features of SAR images. The global statistical features were then fused with the local spatial features for classification.

Although CNNs have been successfully used in the field of SAR image classification, there is a common drawback, that is, the current CNN framework is unable to explore the spatial neighbourhood constraints. To address this problem, several frameworks that fuse the CNN and the Markov random field (MRF) have been proposed (Duan et al. 2017; A. Zhang et al. 2020). These algorithms implement the complementation of the CNN and the MRF models. The CNN can provide a more accurate unary energy term for the MRF, and the MRF can provide the description of the spatial neighbourhood correlation for the CNN. Fusion with the MRF can clearly improve the classification accuracy via smoothing of the misclassification pixels using the spatial neighbourhood constraints. For example, Duan et al. (2017) proposed a SAR image classification algorithm that fuses the contourlet-CNN with the MRF for SAR image classification. We also proposed a super-pixel oriented SAR image classification algorithm that fuses the CNN and MRF. In this algorithm, a new MRF module, named the region category confidence-degree based MRF (RCC-MRF) (A. Zhang et al. 2020), is constructed by utilizing the probability distributions over all of the categories of pixels. The experimental results revealed that the exploration conducted by the RCC-MRF clearly improved the classification accuracy. These frameworks commonly utilize conventional neighbourhood consistency spatial constraints based smoothing, which assumes that the label of a pixel should be the same as the major label of its neighbours. However, this type of spatial constraint has two drawbacks: 1) some of the details and boundaries may be destroyed by the neighbourhood consistency – based smoothing, especially for regional-level MRF models; and 2) some of the regional-level misclassification cannot be eliminated, especially for the pixel-level MRF. These two common drawbacks limit the classification accuracy of CNN-MRF frameworks.

To solve these problems, a new SAR image classification framework that can simultaneously retain the boundaries and eliminate regional-level misclassifications is proposed in this paper. This framework fuses the CNN model with a newly proposed neighbourhood heterogeneity confidence (NHC)-MRF model. In the NHC-MRF model, a NHC index is proposed to build a new binary energy function, which describes an improved spatial constraint by adaptively allowing the existence of heterogeneous labels in the neighbourhood. The NHC term contains a label coexisting probability term and a top-2 label term, which are constructed based on the outputs of the CNN. By utilizing the NHC index, some details can be protected by allowing heterogeneity in the neighbourhood and the regional-level misclassification can be eliminated by adjusting the impact of the neighbourhoods in the smoothing using label coexisting probability based weighting. The NHC-MRF is further fused with the CNN by utilizing the outputs of the CNN to construct the model. The labels and the probability distributions over all of the categories of pixels output by the CNN are utilized to construct the initial labels and the unary energy function of the NHC-MRF. Experiments were conducted on two real SAR images with a 7 m resolution and a synthetic SAR image with exactly known boundaries. The experimental results demonstrate the superiority of the proposed algorithm over existing algorithms in terms of retaining details, eliminating regional-level misclassification, and improving of the classification accuracy. The highlights of this manuscript are as follows.

(1) A novel NHC-MRF model is proposed to describe improved spatial constraints, which adaptively allows the existence of heterogeneity of the category labels in the neighbourhood based on the outputs of the CNN.
(2) A new classification framework fusing the CNN and NHC-MRF is proposed for SAR image classification.
(3) The experimental results demonstrate that the proposed framework can simultaneously retain the details and eliminate regional-level misclassification, which is its superiority over conventional CNN-MRF frameworks.
(4) Background and related work

1.1. Convolutional neural network

The CNN is a type of deep neural network and has been widely used in the field of image processing. The first framework of the CNN, which is called LeNet-5, was proposed by Lécun et al. (1998) for the recognition of handwriting. In 2012, a new framework with a larger depth, called AlexNet, was proposed by Krizhevsky et al. (2012). In AlexNet, a new nonlinear active function, called the rectified linear unit (ReLU) activation function, was developed to increase the sparsity of the parameters. Subsequently, a series of more advanced CNN frameworks were proposed. To process the multi-scale features, the GoogleNet with inception modules was proposed by Szegedy et al. (2015). The inception modules consist of branches with convolutional kernels of different scales. Due to the elaborately designed structure, GoogleNet can effectively capture and process the multi-scale features. To solve the problem of gradient vanishing, ResNet was developed (K. He et al. 2016). It propagates the error of the higher layer directly to the lower layer, and the intermediate layers are skipped in this way. To emphasize the significance of each channel, SENet, which uses the global pooling results of each channel to generate the corresponding weights in an adaptive way, was proposed. Although a series of advanced CNN frameworks have been proposed, the basic components of the CNN are still the same. The structure of the standard CNN with the LeNet-5 framework is illustrated in Figure 1 as an example. The CNN shown in Figure 1 consists of two convolution layers, two pooling layers, one fully connected layer, and one classification layer with a softmax classifier.
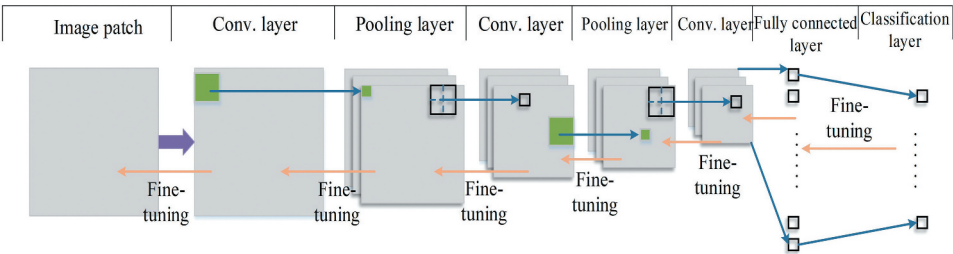

Figure 1. Structure of LeNet-5.

The convolution layer is a crucial component of the CNN and is used to extract the features from all of the input feature maps. It can be expressed as follows:

$$\mathbf{y}_i^l = ReLU(\sum_{i=1}^{m} \mathbf{x}_i^{l-1} \times \mathbf{w}_{ji}^l + b_j^l),$$ (1)

where $\mathbf{y}_j^l$ is the jth feature map in the lth layer, and $\mathbf{x}_i^{l-1}$ is the ith feature map in the $l\text{-}1^{th}$ layer. $\mathbf{w}_{ji}^l$ is the weight in the lth layer that connects the ith feature map in the $l\text{-}1^{th}$ layer and the jth feature-map in the lth layer. b_j^l is the bias in the lth layer. $ReLU()$ is the ReLU active function in the convolution layer.

The feature maps extracted by the convolution layers are then input into the pooling layers to perform sub-sampling. Usually, in the pooling layers, a $k\times k$ window slides within the feature map and a specific value is calculated to represent the image window. There are two commonly used methods of pooling: max-pooling and average-pooling. The pooling layer is used to reduce the data volume, enlarge the receptive field, and smooth the noise to some degree in the SAR image classification.

In the fully connected layer, the local features in the feature maps are integrated into a global feature vector. The feature vector may be processed by the following fully connected layer or directly input into the classification layer.

1.2. Image segmentation using Markov random field

Image segmentation can be regarded as a maximum a posterior (MAP) problem, in which the posterior $P(x|y)$ is maximized. Maximizing the posterior can be converted into maximizing $p(y|x)P(x)$ based on Bayes' theory. To analytically describe the terms $p(y|x)$ and $P(x)$, two models, i.e. the feature model (corresponding to $p(y|x)$) and the spatial context (corresponding to $P(x)$) model, are co2019nstructed. For the spatial context model, the joint probability distribution of the adjacent image sites, which considers the local interaction, can be described. The spatial context model is expressed using the energy function:

$$V_2(x_s, x_t) = \begin{cases} 0, x_s = x_t \\ \beta, x_s \neq x_t \end{cases}.$$ (2)

In Equation (2), s and t refer to two adjacent image sites that form the pair-site clique. The parameter β is a positive number called the balance coefficient. This model makes the Prior $P(x)$ small if the two neighbouring sites are dominated by a single class; otherwise, it is large.

In the feature model, y_s often refers to a constant grey level of site s and the corresponding label of s is denoted as x_s. Commonly, $p(y|x)$ is often assumed to have a Gaussian distribution for simplification. Under this condition, the energy function of the feature model can be written as follows:

$$V_1(x_s) = \frac{1}{2}[\ln(2\pi\sigma_{x_s}^2) + \frac{(y_s - \mu_s)^2}{\sigma_{x_s}^2}],$$ (3)

where μ and σ are the mean and variance of the intensity of class s. Based on the energy functions described in Equation (2) and Equation (3), the overall energy function of the MRF model is calculated as follow:

$$E_f = \sum_{s \in S} V_1(x_s), \tag{4}$$

$$E_s = \sum_{<s,t> \in C} V_2(x_s, x_t), \tag{5}$$

$$E = E_f + E_s, \tag{6}$$

where E_f is the energy function that corresponds to the feature model, and E_s is the energy function that represents the spatial context model. In Equation (5), C refers to the set of image sites. By substituting Equation (4) and Equation (5) into Equation (6), the total energy of the MRF can be described in detail as follows:

$$\arg\min\left\{\sum_{s \in S}\left\{\frac{1}{2}\ln(2\pi\sigma_{x_s}^2) + \frac{(y_s - \mu_{x_s})^2}{2\sigma_{x_s}^2}\right\} + \beta \sum_{<s,t> \in C}\left\{1 - \delta(x_s, x_t)\right\}\right\}. \tag{7}$$

To minimize the energy function, the simulated annealing (SA) algorithm is applied. After the optimization step, the optimized labels of the sites are output.

1.3. SAR image classification algorithms fusing CNN and MRF models

In this section, the algorithms that fuse the CNN and the MRF models are introduced. The CNN is fused with the MRF because that these two types of models can complement with each other. The CNN model can provide a more accurate unary energy function for the MRF model, and the MRF can provide an effective description of the spatial neighbourhood constraints for the CNN. Consequently, it is reasonable to combine the CNN and the MRF to improve the classification accuracy.

Based on the above discussion, a variety of algorithms that combine the CNN and MRF models have been proposed in recent years. Some of these algorithms combine the MRF with the CNN by utilizing the MRF for post-processing of the CNN. For example, Duan et al. (2017) explored the use of the MRF model to post-process the classification results produced by the wavelet-CNN. Geng et al. (2017) utilized the MRF to optimize the classification results produced by the dependency sensitive CNN (DSCNN). To improve the complementation between the CNN and the MRF, some algorithms fuse these two models into one framework. We propose a super-pixel region oriented SAR image classification framework that fuss the CNN and the RCC-MRF. In our proposed framework, the unary energy function of the RCC-MRF is designed to find the category that a super-pixel region most probably belongs to using the RCC term constructed based on the probability distributions over all of the categories predicted by the CNN. The binary energy function of the RCC-MRF explores the spatial constraints between the adjacent super-pixel regions. Compared with the conventional CNN, this framework achieves 3–9% improvement of the overall accuracy. Zhang et al. (2021) proposed a SAR image classification

framework that fuses the complex valued 3-D CNN (CV-3D-CNN) and a modified MRF designed according to the statistical distribution of the SAR image. Compared with CNN model – based algorithms, this framework can be used to improve the classification accuracy noticeably. Although the fused frameworks have superior performances over conventional CNN models, They have the common drawback that the heterogeneity of the labels in the neighbourhood are ignored, which may lead to over-smoothing of the boundaries and the incapability to eliminate regional-level misclassification. To address this problem, a novel framework that fuses the CNN and the NHC-MRF for SAR image classification is proposed in this paper. The details of this framework are described below.

2. Proposed method

2.1. Workflow of the proposed algorithm

In this paper, a new SAR image classification framework that fuses the CNN and the newly proposed NHC-MRF is introduced. This framework is proposed to address the problems of over-smoothing of the details and the incapability of eliminating regional-level misclassification of the conventional CNN-MRF frameworks. In the proposed framework, a new strategy that adaptively allows the existence of heterogeneity in the neighbourhood is proposed to refine the conventional neighbourhood consistency constraint. Accordingly, a new model, called the NHC-MRF, with a NHC index based binary energy term is developed. The NHC index consists of a coexisting probability term of the labels in the neighbourhood and a top-2 label correcting term, which are calculated based on the outputs of the CNN. By exploring the NHC term, the details can be protected by allowing the existence of heterogeneity in the neighbourhood and regional-level misclassification can be eliminated by adjusting the impact of the neighbourhoods using the coexisting probability based weighting. The NHC-MRF is further fused with a CNN by constructing the initial labels and unary energy term of this model based on the outputs of the CNN. The final classification results are obtained by minimizing the energy function of the NHC-MRF. A flowchart of the proposed framework is presented in Figure 2. The workflow of the proposed framework can be divided into three steps: initialization of the labels,
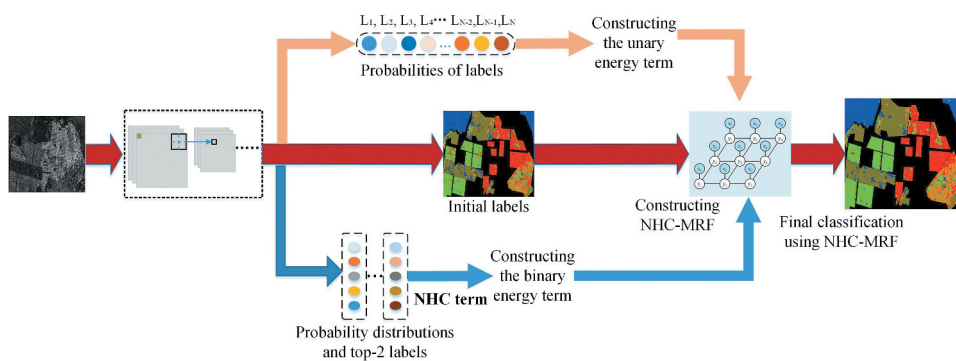

Figure 2. Workflow of the proposed NHC-MRF and the CNN based framework.

construction of the NHC-MRF model and, classification using the NHC-MRF model. These three steps will be introduced in detail in the following sections.

2.2. Initialization of labels

The CNN is used to output the initial labels of the pixels. The CNN model explored in this paper contains convolution layers, pooling layers, fully connected layers, and a softmax classification layer. The convolution layers are used to extract the deep features of the input images, and the pooling layers can enhance the receptive field by downsampling the feature maps. The fully connected layers capture the global features by stretching and concatenating the local features in the feature maps. After the fully connected layers are processed, a softmax classifier is utilized as the classification layer.

The mini-batch stochastic gradient descending (SGD) algorithm is utilized to optimize the parameters of the CNN. As the softmax classifier is used as the classification layer, the cost function of the CNN is as follows:

$$L = -\frac{1}{N}[\sum_{n=1}^{N}\sum_{j=1}^{K} 1\{y_n = j\} \frac{e^{\theta_j x_n}}{\sum_{j=1}^{K} e^{\theta_j x_n}}]. \tag{8}$$

In Equation (8), N is the number of samples in a mini-batch, and K is the total number of categories. The function $1\{y_s=j\}$ is a logic judgement function, which means that $1\{y_s=j\}$ is equal 1 when $y_s=j$ and $1\{y_s=j\}$ equal 0 and $y_s \neq j$. The vector θ denotes the parameters of the softmax classifier, and the vector x is the feature vector input into the softmax classifier. Based on the cost function, the parameters of the CNN can be optimized using a backpropagation (BP) algorithm. The weights and the biases in the CNN can be adjusted according to Equations (9)–(12). Equation (9) and Equation (11) are the analytical forms of the updating of the weights and the biases. According to the BP algorithm the numerical forms of Equation (9) and Equation (11) are illustrated by Equation (10) and Equation (12).

$$\mathbf{W}_{i+1}^{l} = \mathbf{W}_i^{l} - \alpha \frac{\partial L}{\partial \mathbf{W}_i^{l}}. \tag{9}$$

$$\Delta w_{mn}^{l} = rot180(conv(\mathbf{h}_m^{(l-1)}, rot180(\sigma_m^{l})))/N_{batchsize}. \tag{10}$$

$$b_{i+1}^{l} = b_i^{l} - \alpha \frac{\partial L}{\partial b_i^{l}}. \tag{11}$$

$$\Delta b = \sum_p \sum_q \delta_{j(p,q)}^{l}/N_{batchsize}. \tag{12}$$

In Equation (9), $\mathbf{W}_i^{l+1}$ refers to the weight matrices in the l+1 layer at the ith iteration. Similarly, in Equation (11), b_{i+1}^{l} represents the biases in the l+1 layer at the ith iteration. In Equation (10), the numbers m and n represent the mth and n^{th} feature maps in the l-1th and l^{th} layers. The function $rot180()$ denotes rotation of a matrix by 180°. The matrix

h represents the hidden layer, and the matrix σ represents the matrix of the error term. In Equation (12), (p, q) denotes the pixel at position (p, q).

After the training, the CNN with optimized parameters is obtained, and the initial labels of the input images can be obtained using the trained CNN. If the feeding forward function of the optimized CNN is denoted as $F()$, the label of an input image patch can be expressed as follows:

$$L_i = \max(F(\mathbf{X}_i)). \tag{13}$$

Using Equation (13), the initial labels of the pixels in the SAR image can be obtained. In Equation (13), the feeding forward function outputs the probability distributions over all of the categories corresponding to the input image. This can be expressed as follows:

$$[p_1, p_2, \ldots p_j, \ldots, p_k] = F(\mathbf{X}_i), \tag{14}$$

where p_j denotes the probability that the label of image patch $\mathbf{X}_i$ is assigned as j, and k is the total number of the categories.

2.3. Classification using the NHC-MRF

After the initialization of the labels, the NHC-MRF should then be constructed and utilized to produce the final labels of the pixels. In this way, the classification can be performed. In our proposed framework, the NHC-MRF is fused with a CNN, which means that the unary and binary energy terms of the NHC-MRF are constructed based on the output of the CNN. By minimizing the total energy function of the NHC-MRF model using the simulated annealing algorithm, the final labels of the pixels can be determined. The total energy function of the NHC-MRF is constructed by adding the unary and binary energy terms:

$$E_{total} = E_{unary} + E_{binary}. \tag{15}$$

The mechanisms and the details of the unary and the binary energy terms are presented in the following two sections.

2.3.1. Unary energy function of the NHC-MRF

The unary energy function is used to determine the category that a pixel most probably belongs to based on its own features. Conventionally, the unary energy function is constructed based on the mixed Gaussian distribution, which assumes that the label and intensity of a pixel have a mixed Gaussian distribution. However, the correlation between the label and the intensity of a pixel cannot be simply described by a mixed Gaussian distribution for the SAR image. Consequently, a new energy function is constructed based on the probability distributions of the pixels produced by the CNN, which describe the probabilities that the pixels belong to each category. Based on the above analysis, the unary energy function can be written as follows:

$$V_{unary} = -\ln(p(l_{x_i} = l|\mathbf{fv}_{CNN})). \tag{16}$$

$$E_{unary} = \sum_{l \in L} V_{unary}. \tag{17}$$

In Equation (16), the potential function of the energy function is given. $\mathbf{fv}_{CNN}$ is to the feature vector corresponding to pixel i, which is generated by CNN. The probability $p(l_{x_i} = l|\mathbf{fv}_{CNN})$ denotes the probability that pixel x_i belongs to category l under the condition that the corresponding input feature vector is $\mathbf{fv}_{CNN}$. By maximizing the probabilities in Equation (16), the potential function will be minimized and the optimized label of the category will be determined. In Equation (17), L is the set of category labels.

2.3.2. Binary energy function of the NHC-MRF

The binary energy term of the NHC-MRF is constructed to optimize the initial labels of the pixels based on the NHC index refined spatial constraint. As shown in Equation (2), the conventional binary energy term assumes that the pixels in a neighbourhood should have the same category and the dominant label of an $n \times n$ neighbourhood should be determined as the label of this neighbourhood. In fact, there may be some heterogeneous pixels in some neighbourhoods, such as at the boundaries, especially for a high-resolution SAR image. Taking Figure 3(a) as an example, there is a boundary in Figure 3(a). According to the conventionally explored spatial constraints, the labels of some of the pixels in the boundary region (highlighted by circles) will be improperly changed. The conventionally explored spatial constraints can also lead to the inability to eliminate the regional-level misclassification, which is demonstrated in Figure 3(b). In Figure 3(b), there is regional-level misclassification. In Figure 3(b), for the highlighted misclassified pixel, the majority of its neighbouring pixels are also misclassified pixels, which leads to the inability to correct the misclassification.

However, these two problems can be solved by allowing heterogeneity in the neighbourhood, which means that the label of a pixel can be different from the majority of its neighbours according to a certain degree of confidence. According to this idea, the NHC
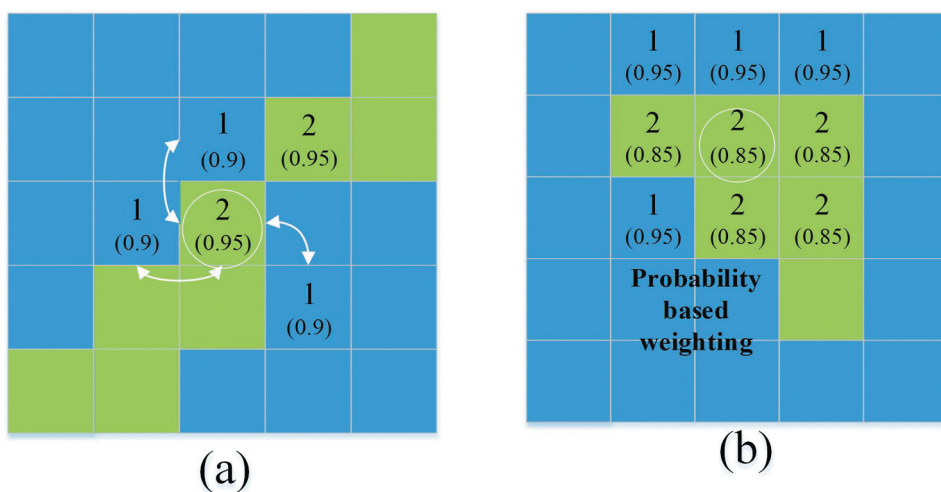

Figure 3. Illustration of the mechanism of the NHC term. (a) Retention of details (boundaries); and (b) elimination of regional-level misclassification.

index denotes the confidence of the heterogeneous labels in a neighbourhood in this paper. The NHC index consists of two parts: the neighbourhood label coexisting probability term and the top-2 label correcting term. The label coexisting probability term can preserve the details by giving the probability that the heterogeneity should be preserved and eliminating the regional-level misclassification by adaptively adjusting the impact of the neighbourhood pixels. The NHC based binary energy term, the corresponding potential function, and the details of the NHC index are quantitatively expressed in Equation (18), Equation (19), and Equation (20):

$$E_{binary} = \sum_{l \in L} V_{binary},$$

(18)

$$V_2(x_s, x_t) = \begin{cases} 0, l_s = l_t \\ \beta(1 - NHC_{CNN}^{(s,t)}), l_s \neq l_t \end{cases},$$

(19)

$$NHC_{CNN}^{(s,t)} = \alpha_{coe} P_{l_s} P_{l_t} . \alpha_{top2}^{\delta(l_{he}^{top2} = l_{NB})},$$

(20)

In Equation (18), *s* and *t are* two adjacent pixels in a neighbourhood. The corresponding labels of *s* and *t* are l_s and l_t. The NHC index range is (0,1). When the NHC index becomes larger, the binary potential function will become smaller and the heterogeneous labels are more likely to be maintained. In this way, some of the details and boundaries can be protected. The subscript *CNN* indicates that the NHC index is constructed based on the outputs of the CNN. The quantitative description of the NHC is illustrated in Equation (20). The NHC index is constructed by multiplying two terms: a coexisting probability term and a top-2 label term, which will be introduced below.

For two adjacent pixels *s* and *t*, as the corresponding labels (l_s and l_t) are separately predicted by the CNN, the assignment of labels to these two pixels can be regarded as being statistically independent. Consequently, the coexisting probability can be expressed as $P_{l_s} \times P_{l_t}$, where P_{l_s} and P_{l_t} denote the probability that the corresponding labels of pixels *s* and *t* are l_s and l_t. The coexisting probability term reflects the confidence of the existence of heterogeneous labels within a neighbourhood. It is more reasonable to regard the heterogeneous pixel as the boundary when the coexisting probability becomes larger (Figure 3(a)). By subtracting the NHC index term, the binary energy term will become smaller and the unary term will have more impact. In this way, smoothing of the heterogeneity will be hindered and the boundaries can be retained. In addition, regarding regional-level misclassification, the coexisting probability term can perform adaptive weighting of the neighbourhood pixels. In this adaptive weighting, the pixel that has a higher degree of confidence of its assigned label has a greater impact in determining the label of the central pixel. In this way, some of the regional-level misclassification can be corrected. As exemplified by the circled pixel in Figure 3(b), for the neighbouring pixels of the highlighted misclassified pixel (labelled as *2*), the correctly classified pixels (labelled as *1*) have higher degrees of confidence and the misclassification of this pixel is lower (because the misclassifications are generally caused by the disturbance of the speckle noise or some other factors). According to the denoted weights of the neighbouring pixels, the correctly classified pixels have a larger impact (described using the weighted sum) on the category label of the highlighted pixel. In this way, the

binary energy term will become smaller if the central pixel is labelled as *1*. Consequently, by using the coexisting probability term, the regional-level misclassification can be reduced. However, the label of the highlighted pixel will be kept unchanged by using the conventional spatial constraints because the highlighted pixel has the same label as the majority of its neighbours (labelled as *2*). The label coexisting probability term is weighted with a coefficient denoted as a_{coe}.

The top-2 label term, which serves as the correction of the label coexisting probability term, is illustrated in the form of logical judgement in Equation (20) using the function $\delta()$. As demonstrated by Equation (20), for a neighbourhood, $l_{he\text{-}top2}$ denotes the top-2 labels of the heterogeneous pixel, and l_{NB} is the label of the neighbourhood. The top-2 label term will be equal to *1* when $l_{he\text{-}top2} \neq l_{NB}$, while the top-2 label term will be equal to a_{top2} (a_{top2} is a positive constant within $(0,1)$) when $l_{he\text{-}top2} = l_{NB}$. This indicates that the heterogeneity in the neighbourhood may be caused by factors such as the speckle noise when $l_{he\text{-}top2} = l_{NB}$. In this way, the confidence of the heterogeneity in the neighbourhood should be reduced. As introduced above, the NHC index considers the confidence of the heterogeneous in the neighbourhood from two aspect: the exact existence of heterogeneity in the neighbourhood and the disturbance of the speckle noise. As discussed above, by exploring the NHC index, retention of the details and smoothing of the misclassification can be simultaneously realized.

3. Experiments and discussion

3.1. Experimental settings and datasets

To evaluate the performance of the proposed algorithm, three datasets are explored in experiments. These three datasets include two widely used real-world images and one synthetic image used in our former study (A. Zhang et al. 2020). The synthetic image is generated by adding the single-look speckle noise to the greyscale image. The synthetic image contains two types of areas: homogeneous areas and textured areas. In addition, the boundaries are exactly known in this image and regional-level misclassification frequently occurs in the textured areas. Consequently, this synthetic SAR image is suitable for evaluating the capability of the proposed framework to retain details and eliminate regional-level misclassification. In addition to the synthetic SAR image, two real-world SAR images are explored in the experiments. Both of these images were obtained by the Radarsat-2 system in the San Francisco Bay (N37°47′, W122°28′) and Flevoland (N52°22′, E5°27′) sites. Detailed descriptions of the explored images are presented in Table 1.

As reported in Table 1, these two images contain single-look speckle noise, which can lead to enhancement of the misclassification. In addition, the inter-class similarities

Table 1. Details of the explored real SAR images.

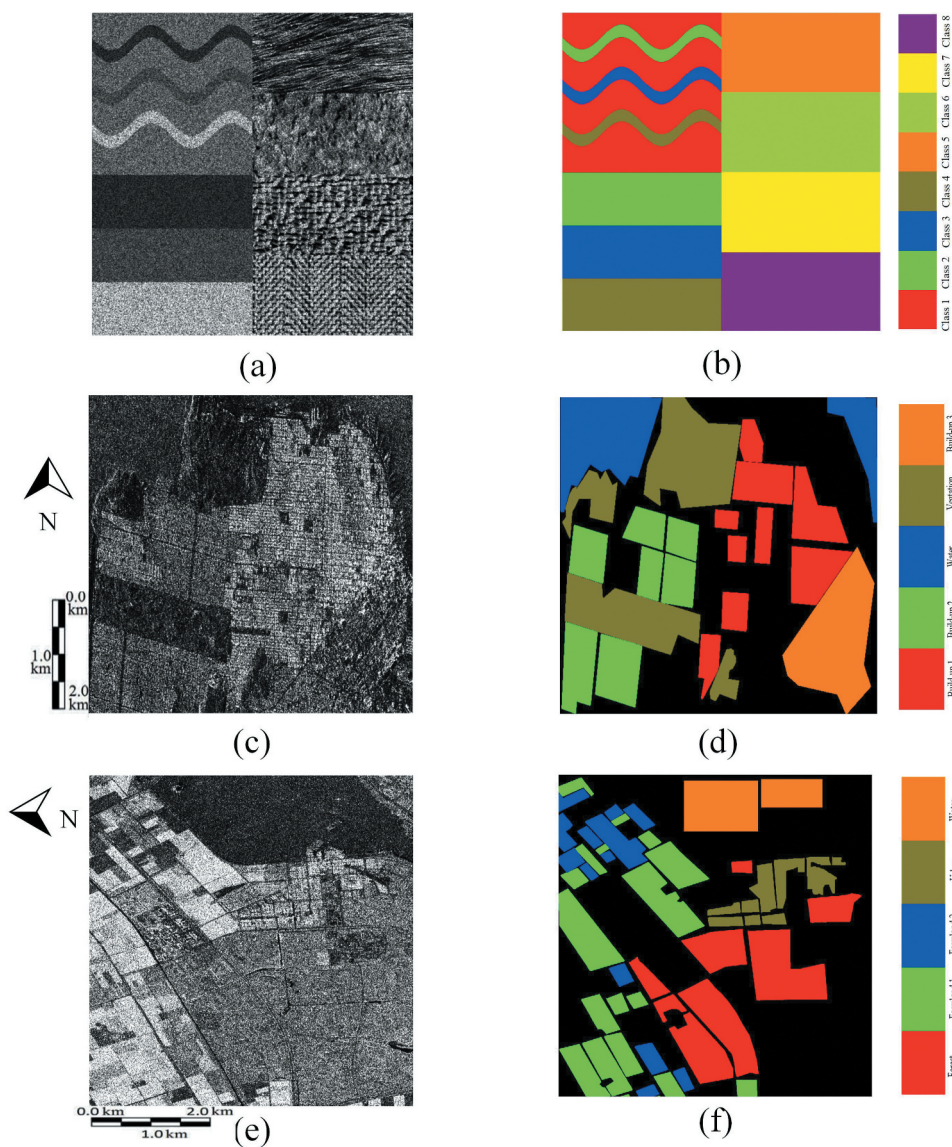| ID	Site	Latitude and longitude	Resolution	Format	Number of terrains
No.1	SanFransisco-Bay(USA)	N37°47′, W122°28′	7 m	SLC	5
No.2	Flevoland (Netherlands)	N52°22′ E5°27′	7 m	SLC	5

Figure 4. Images utilized and the corresponding ground truths.

between some of the terrains in these images can lead to regional-level misclassification. Consequently, these images are suitable for evaluating the ability of the NHC-MRF module to eliminate regional-level misclassification and retain details in real scenes. The explored SAR images and the corresponding ground truths are presented in Figure 4 (Zhao and Du 2016). For each image, the labelled pixels and their labels form the sample set. In each image, 1000 samples are randomly selected as the training sample to evaluate the performance of the proposed algorithm using limited training samples.

Before the comparison experiments, the parameters of the CNN should be determined. The size of the input image patch should be determined. According to our former study
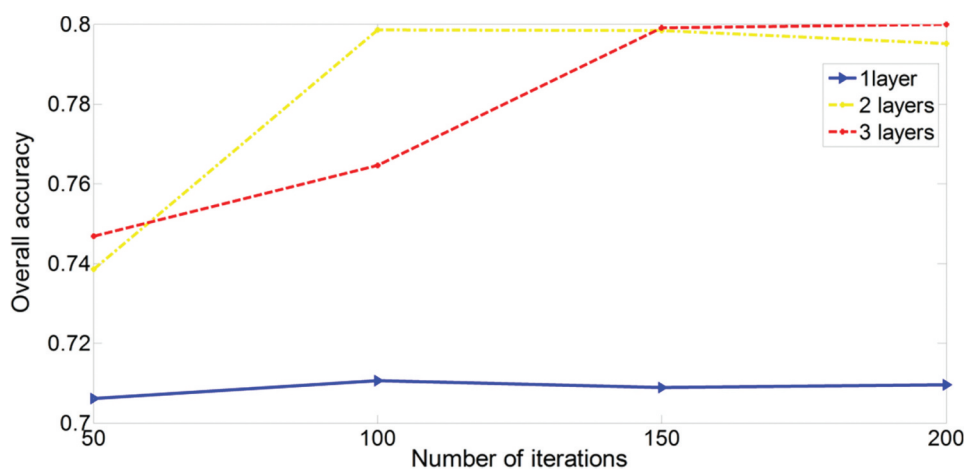
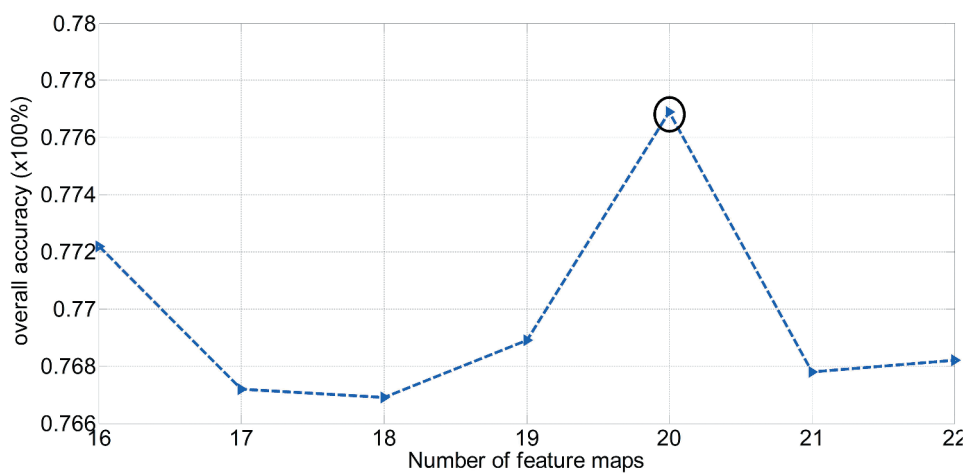**Figure 5.** Optimization of the number of layers and the number of iterations.

Figure 6. Optimization of the number of feature maps.

(A. Zhang et al. 2019), the size of the input image patch is set to 27×27, which shows a counter-balance between exploring the spatial information and smoothing the details of the features. The number of layers, the number of feature maps, and the number of iterations are then optimized using the grid research method. In this experiment, a Radarsat-2 San Francisco Bay real SAR image is explored for optimization.

The number of layers and the number of iterations are jointly optimized in the first step. In this step, the number of feature maps is fixed as a certain value. The results of this optimization are illustrated in Figure 5, and the optimized value of the number of layers is set to two (a layer refers to a couple of convolution-pooling layers). As indicated by Figure 6, the CNN obtains a much lower classification accuracy when only one layer is used because of the weak representation ability of the extracted features. The CNN with three layers obtains a slightly higher overall accuracy; however, the time cost is much higher. Consequently, the optimized value of the number of layers is set to two for the

Table 2. The optimized hyperparameters of compared algorithms.

CNN	Parameters	Size of input patch	Number of layers	Number of feature-maps	Number of iterations	Learning rate	Size of mini-batch
	Values	27 × 27	2	20	100	0.05	10
DBN	Parameters	Size of input patch	Number of hidden layers	Number of feature-maps	Number of iterations	Learning rate	Size of mini-batch
	Values	27 × 27	2	100/100	150	0.5	10
SAE	Parameters	Size of input patch	Number of hidden layers	Number of feature-maps	Number of iterations	Learning rate	Size of mini-batch
	Values	27 × 27	2	100/200	150	0.3	10
SVM	Parameters	Kernel function	Gamma parameter of kernel	Size of input feature vector			
	Values	Radial basis function	0.02	729			
RF	Parameters	Number of trees	Size of input feature vector	Depths of tress			
	Values	800	729	adaptive			
NHC-MRF /MRF	Parameters	Balance coefficients	Number of iterations				
	Values	15/10	20				

counter-balance between the classification accuracy and the time-cost. For the CNN with two layers, the optimized number of iterations should be 100. The CNN cannot be well trained using an insufficient number of iteration. However, over-fitting will occur and will lead to a lower classification accuracy when the number of iterations is too larger. Furthermore, the learning rate and the size of the mini-batch are also experimentally optimized using grid research.

The number of feature maps is then optimized. To reduce the amount of computation, the number of feature maps in each layer is set to the same value. The results of the optimization are presented in Figure 6. According to Figure 6, the optimized number of feature maps is 20, which is a tradeoff between the representative ability and the redundancy of the features. After this experiment, the optimized structural parameters of the CNN are determined. The CNN with optimized structural parameters is used in the evaluations in the following sections.

To quantitatively describe and compare the performance of the proposed algorithm and those of existing algorithms, several generally used metrics are explored. To measure the classification accuracy, the overall accuracy is determined first. The overall accuracy (OA) can be expressed as follow:

$$OA = \frac{N_{correct}}{N_{total}} \times 100\%, \tag{21}$$

where $N_{correct}$ is the number of correctly classified samples, and N_{total} is the total number of samples. In addition to the OA, the accuracy of each category is also given as a metric. To measure the ability to retain details, the figure of merit (FOM) (A. Zhang et al. 2019) is also used to describe the boundary preserving abilities of all of the evaluated algorithms.

The proposed algorithm is compared with the conventional CNN-MRF, CNN, stacked autoencoder (SAE), and support vector machine (SVM) algorithms. As limited

training samples are explored for experiments, the ensemble learning should be utilized for comparison to demonstrate the superiority of the proposed algorithm. As a result, the Random Forest(RF) which is a representative algorithm of ensemble learning is also used as a comparison.The RCC-MRF based algorithm, which achieves 3–9% improvement of the OA, is also used as a comparison algorithm to evaluate the superiority and limitations of the proposed algorithm. To further explain the mechanism of the proposed NHC-MRF module, an ablation experiment is also conducted as a part of the discussion to demonstrate the effect of the coexisting probability term and the top-2 label term. The optimized hyperparameters of the CNN and the other explored algorithms are summarized and listed in Table 2.

3.2. Results and discussion

3.2.1. Results on synthetic SAR image

The synthetic SAR image is used first to evaluate the performance of the proposed algorithm and the conventional algorithms. As the synthetic image contains exactly
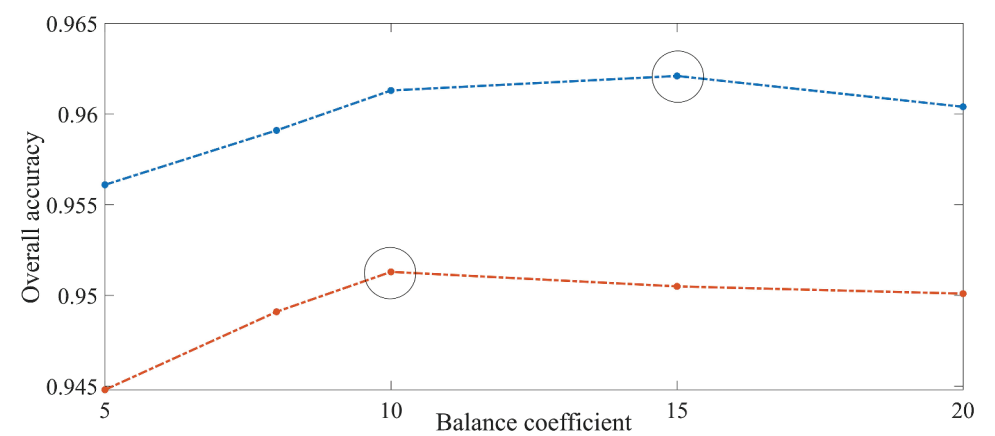

Figure 7. Optimization of the balance coefficients of the NHC-MRF and the conventional MRF for the synthetic SAR image.

Table 3. Classification accuracies for synthetic SAR image.

categories	Proposed algorithm(%)	CNN-MRF(%)	CNN(%)	DBN(%)	SAE (%)	SVM (%)	RF(%)
class 1	**96.76**	94.86	94.02	92.98	94.43	86.07	94.51
class 2	**97.30**	97.25	97.13	97.43	96.26	91.39	96.97
class 3	**98.40**	97.64	97.97	96.17	96.64	84.14	98.03
class 4	**98.71**	98.57	97.92	96.43	96.13	96.54	97.83
class 5	**94.47**	94.42	93.62	93.82	93.51	79.57	80.52
class 6	**90.71**	88.26	85.63	70.86	73.30	65.44	
class 7	**96.23**	95.96	92.33	84.21	77.78	82.08	
class 8	**96.87**	96.73	90.11	94.91	93.60	89.07	
OA	**96.21**	95.13	92.87	90.74	90.41	85.43	

known boundaries and narrow curving regions, it is suitable to utilize this image to describe the abilities of the proposed NHC-MRF and the conventional algorithms to retain details. The synthetic image also contains four textured regions, in which regional-level misclassifications frequently occurs because these four textured regions are similar to each other. Thus, this image is also suitable for evaluating the ability of the proposed algorithm to eliminate regional-level misclassification.

The optimization of the parameters of the NHC-MRF is performed before the comparison. The results of the optimization are illustrated in Figure 7, which shows that the optimal balance coefficient of the NHC-MRF is 15 and that of the conventional MRF is 10. The optimal coefficients of the label coexisting probability term and the top2 label term in the NHC index are 0.5 and 0.9.

The classification accuracies are listed in Table 3, and the classification results are shown in Figure 8. According to Table 3, the proposed algorithm achieves the highest overall accuracy. The CNN-MRF has the second highest overall accuracy (95.13%). As indicated by the result maps illustrated in Figures 18(a,b) both the MRF and the NHC-MRF can smooth the speckle noise, while the NHC-MRF can more effectively retain the details, which are highlighted by the white rectangles. In addition, the proposed algorithm can eliminate the regional-level misclassifications in the textured parts, which are highlighted by the white circles. In the textured parts, there are a series of misclassified regions, which may be caused by the speckle noise, inter-class similarity, and the other factors. By using the proposed algorithm, the regional-level misclassification can be effectively eliminated, which demonstrates the mechanism of the NHC-MRF. To further demonstrate the ability of the proposed NHC-MRF to eliminate regional-level misclassification, the confusion matrices of the proposed algorithm and the CNN-MRF are presented in Figure 9. In the confusion matrices of the explored algorithms, class 6 and class 7 are easily misclassified because of the regional-level misclassification. The accuracies of these two categories noticeably improve with the use of the proposed NHC-MRF, but not the CNN-MRF, especially class 6, for which the NHC-MRF results in a 2% improvement of the classification accuracy compared with the conventional MRF. Some boundaries are also retained better. The FOM of the proposed algorithm is 0.6640 and that of the CNN-MRF is 0.6140. This result demonstrates the ability of the proposed NHC-MRF to retain details. The superiority of the ability of the NHC-MRF to retain details is also demonstrated by the classification result maps. As highlighted by the white rectangle, the boundaries of the curving regions (classes 2–4) on the result map obtained using the proposed algorithm are clearer than those obtained using the CNN-MRF, which shows that the NHC-MRF can retain the boundaries better.

The OA of the CNN is 92.87, which is less than that of the CNN-MRF and the proposed algorithm but is much higher than those of the deep belief network (DBN), SAE, and the SVM. This can be explained by the fact that the CNN can smooth the misclassification produced by the speckle noise. As shown by the result maps of the CNN, DBN, SAE, and the SVM, the classification results of the CNN are much smoother due to the immunity of the CNN to speckle noise. The CNN also shows better performance than RF algorithm, which demonstrates the superiority of the deep feature learning strategy. As shown by Figure 8(g), there are a much larger portion of misclassification in the texture areas comparing with the results of CNN. This is because that the CNN can extract deep features with semantic information, which can reduce the misclassification caused by the inter-
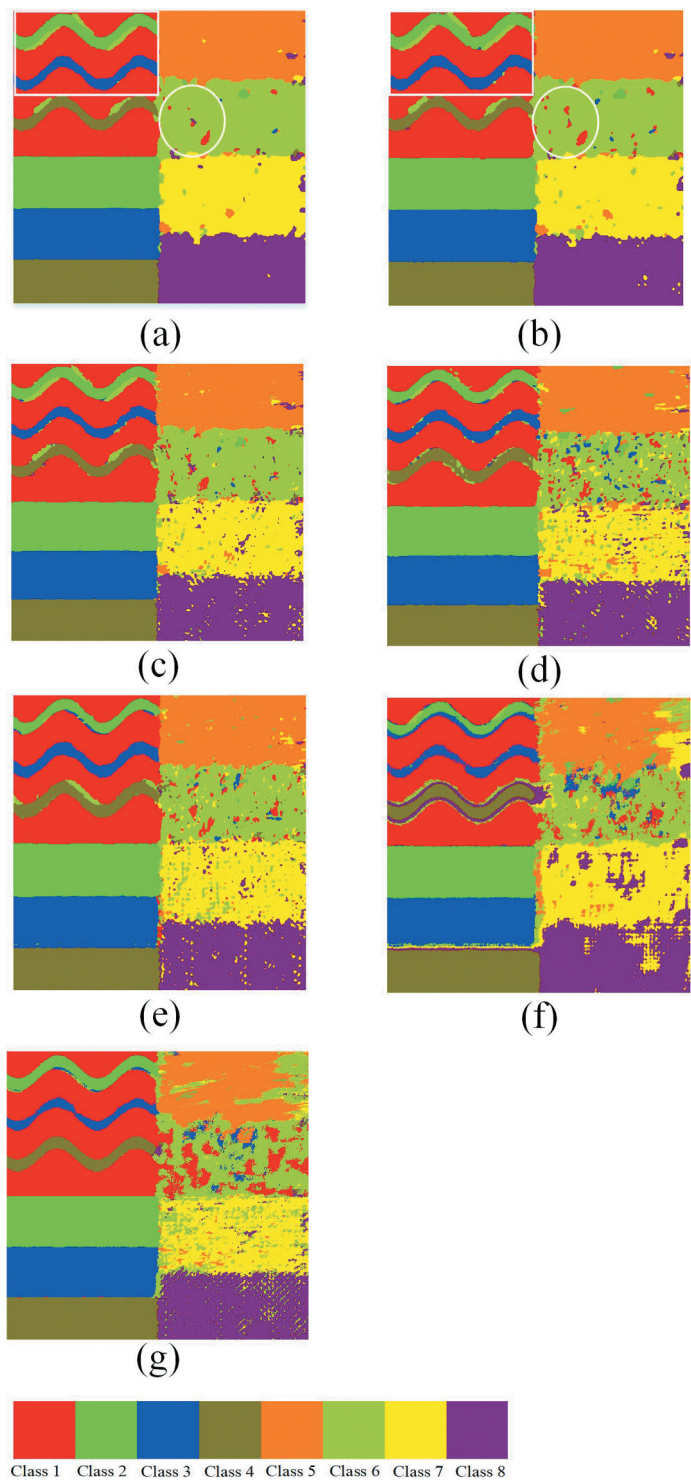
Figure 8. Classification result maps for the synthetic SAR image: (a) results of the proposed algorithm; (b) results of the CNN-MRF; (c) results of the CNN; (d) results of DBN; (e) results of the DBN; (f) results of the SAE; and (g) results of the SVM.
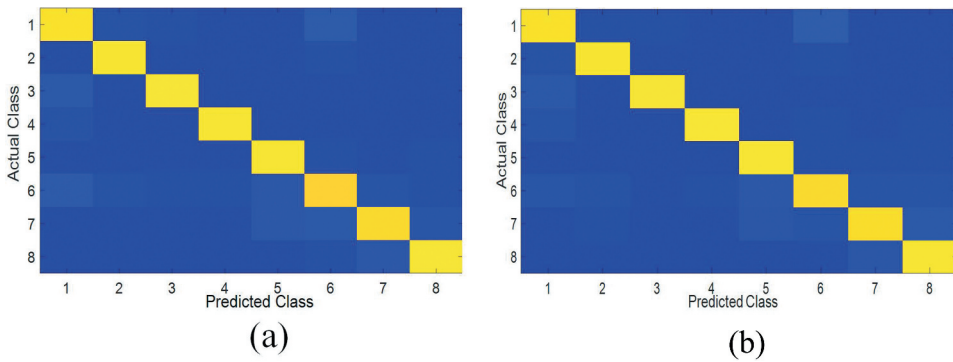
Figure 9. Confusion matrices of the proposed algorithm and the CNN-MRF for the synthetic SAR image.
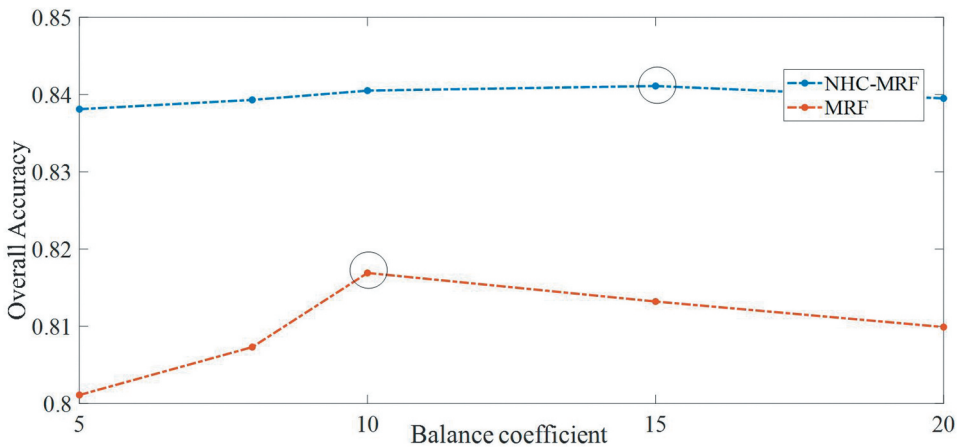

Figure 10. Optimization of the balance coefficients of the NHC-MRF and the conventional MRF for the radarsat-2 San Francisco Bay real SAR image.

similarity in the texture regions. However, the RF shows less misclassification than SAE, DBN and SVM. This is because that the RF can resist the speckle noise at some degree due to the voting mechanism when outputting the results.

3.2.2. Results on radarsat-2 San Francisco Bay

The Radarsat-2 San Francisco Bay image is then utilized to evaluate the performance of the proposed algorithm on real SAR images. This image contains single-look speckle noise, high inter-class similarities between the terrains of built-up land, and some boundaries. Consequently, the Radarsat-2 San Francisco Bay real SAR image is chosen to evaluate the performance of the proposed algorithm on a real SAR image.

This image contains five categories of terrains: water, forest, built-up 1, built-up 2, and built-up 3. Among these terrains, the three types of built-up land are similar to each other, and this inter-class similarity may lead to the misclassification of these terrains, especially in built-up3 where regional-level misclassification frequently occurs. In this SAR image, there
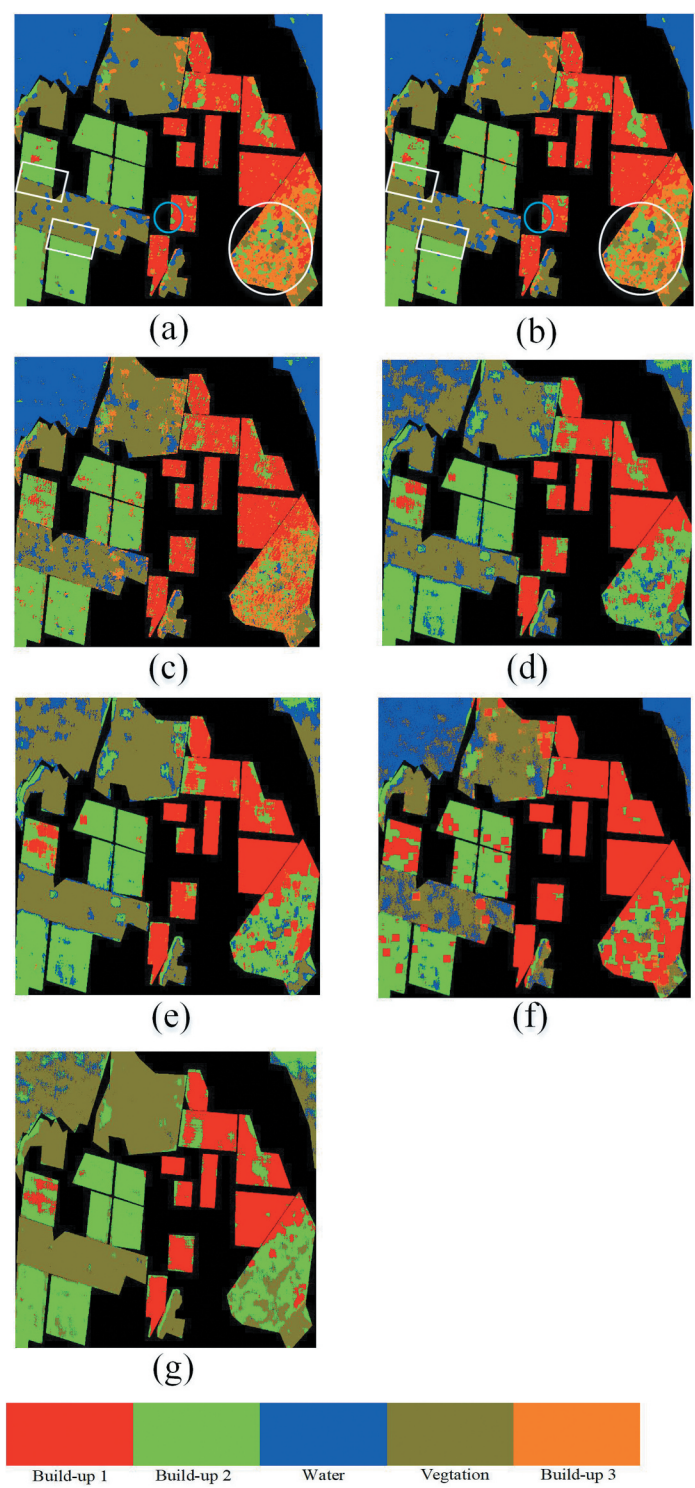
Figure 11. Classification result maps for the San Francisco Bay SAR image: (a) classification results of the proposed algorithm; (b) classification results of the CNN-MRF; (c) classification results of the CNN; (d) classification results of the DBN; (e) classification results of the SAE; and (f) the classification results of the SVM.

**Table 4.** Classification accuracies of all of the compared algorithms for the radarsat-2 San Francisco Bay real SAR image.

categories	Proposed algorithm(%)	CNN-MRF(%)	CNN(%)	DBN(%)	SAE (%)	SVM (%)	RF(%)
built-up 1	82.73	81.16	80.61	89.63	90.65	97.38	86.86
built-up 2	93.91	91.22	85.94	71.36	71.26	75.15	87.48
water	98.06	98.01	95.44	75.88	76.33	77.64	11.44
vegetation	85.84	83.19	77.73	70.84	71.35	70.43	94.65
built-up 3	**54.25**	**47.18**	**45.27**	5.12	4.97	5.78	0.8
OA	84.11	81.69	79.33	64.61	64.97	73.1	64.85

are also some boundary regions where some details exist. Because of the characteristics described above, the Radarsat-2 San Francisco Bay image is utilized to evaluate the ability of the proposed framework to eliminate regional-level misclassification and retain details.

The parameters of the NHC-MRF are optimized before the comparison experiments. The balance coefficient between the unary and binary terms of the NHC-MRF are optimized, and the results of the optimization are presented in Figure 10. According to the results of the optimization, the optimized balance coefficient of the NHC-MRF in this experiment is 15, and that of the MRF is also 10. The optimized coefficients of the label coexisting probability term and the top2 label term in the NHC index are 0.5 and 0.9. The parameters of the CNN-MRF, CNN, and other algorithms are also optimized to reach the highest classification accuracy for this image.

The comparison experiment is conducted after the optimization of the parameters. The classification accuracies of all of the compared algorithms are listed in Table 4, and the result maps are presented in Figure 11. According to Table 4, the proposed algorithm still achieves the highest accuracy, which is about 3% higher than that of the CNN-MRF. This result demonstrates the superiority of the proposed algorithm for a real SAR image with a high-level of noise and inter-class similarity. The confusion matrices of the proposed algorithm and the CNN-MRF are presented in Figure 12.

According to the confusion matrices, the proposed algorithm can significantly improve the classification accuracy on the built-up 3 which is easily confused with the other types of terrain. By comparing Figures 12(a,b) it can be seen that a series of regional-level misclassifications are eliminated by the proposed NHC-MRF, which leads to more than 7% enhancement of the classification accuracy (Table 3) for the
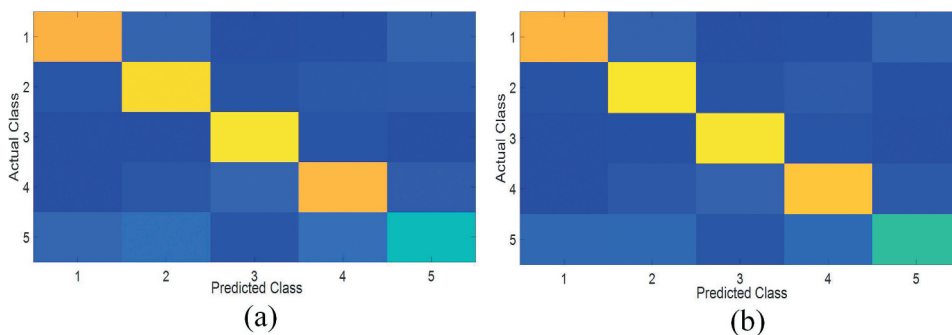
Figure 12. Confusion matrices of the proposed algorithm and the CNN-MRF for the radarsat-2 San Francisco Bay SAR image.

built-up 3. This conclusion is supported by the results highlighted by the white circles in Figure 11(a,b). The superiority of the NHC-MRF over the CNN-MRF in terms of retaining details is illustrated by the classification results highlighted by the white rectangles. It can be seen that the highlighted boundaries of the results of the CNN-MRF are vaguer than those of the proposed algorithm, which demonstrates that the proposed NHC-MRF model can retain the boundaries better due to the effect of the NHC term. The corresponding classification results of the proposed and compared algorithms demonstrate the ability of the proposed NHC-MRF model to eliminate regional-level misclassification and retain boundaries for a real SAR image. The proposed algorithm also has some disadvantages. As highlighted by the blue circles, the proposed NHC-MRF yields several regional-level misclassification results. This is because the NHC term is constructed based on the probability distribution output by the CNN. For some regions with severe interference, the output of the CNN is not reliable, which leads to misclassification by the NHC-MRF. However, as the CNN can resist the speckle noise to some degree by using local spatial information, the results of the CNN are reliable for most regions.The RF achieves 64.90% OA with the number of decision trees optimized as 800. In this image, the classification accuracy of RF is still much lower than that of CNN because of the high-level inter-class similarity. It can be seen in Figure 12 that the RF cannot distinguish water from vegetation, which also occurs in the results of SAE. And also, for the built-up3 which is mostly confused with the other terrains, the RF shows very poor classification performance. This results indicates that drawback of the RF is that it cannot distinguish terrains with high inter class similarities.

3.2.3. Results for the radarsat-2 Flevoland real SAR image

The Radarsat-2 Flevoland real SAR image with a 7-m resolution is also used to evaluate the performances of the proposed algorithm and the compared algorithms. This image contains five types of terrains: forest, water, farmland1, farmland2, and urban land. The image contains single-look speckle noise and the farmland2 is similar to the forest, which
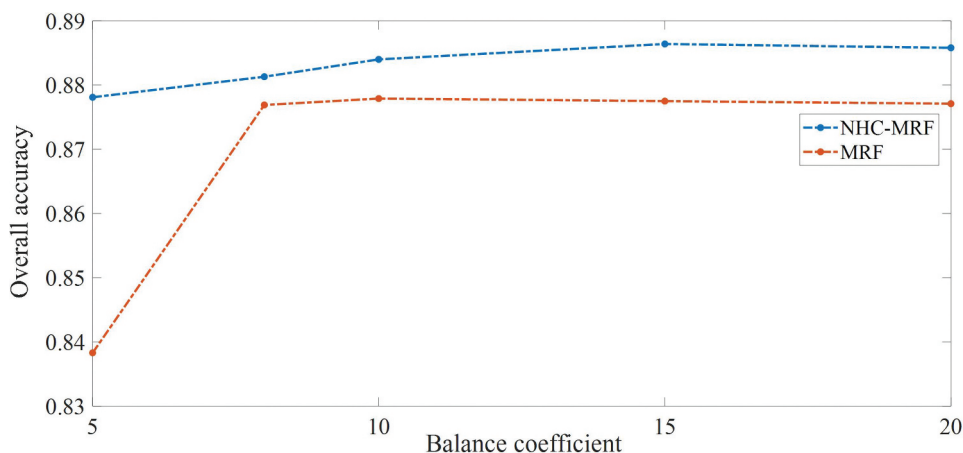

Figure 13. Optimization of the balance coefficients of the NHC-MRF and the conventional MRF for the Radarsat-2 Flevoland real SAR image.
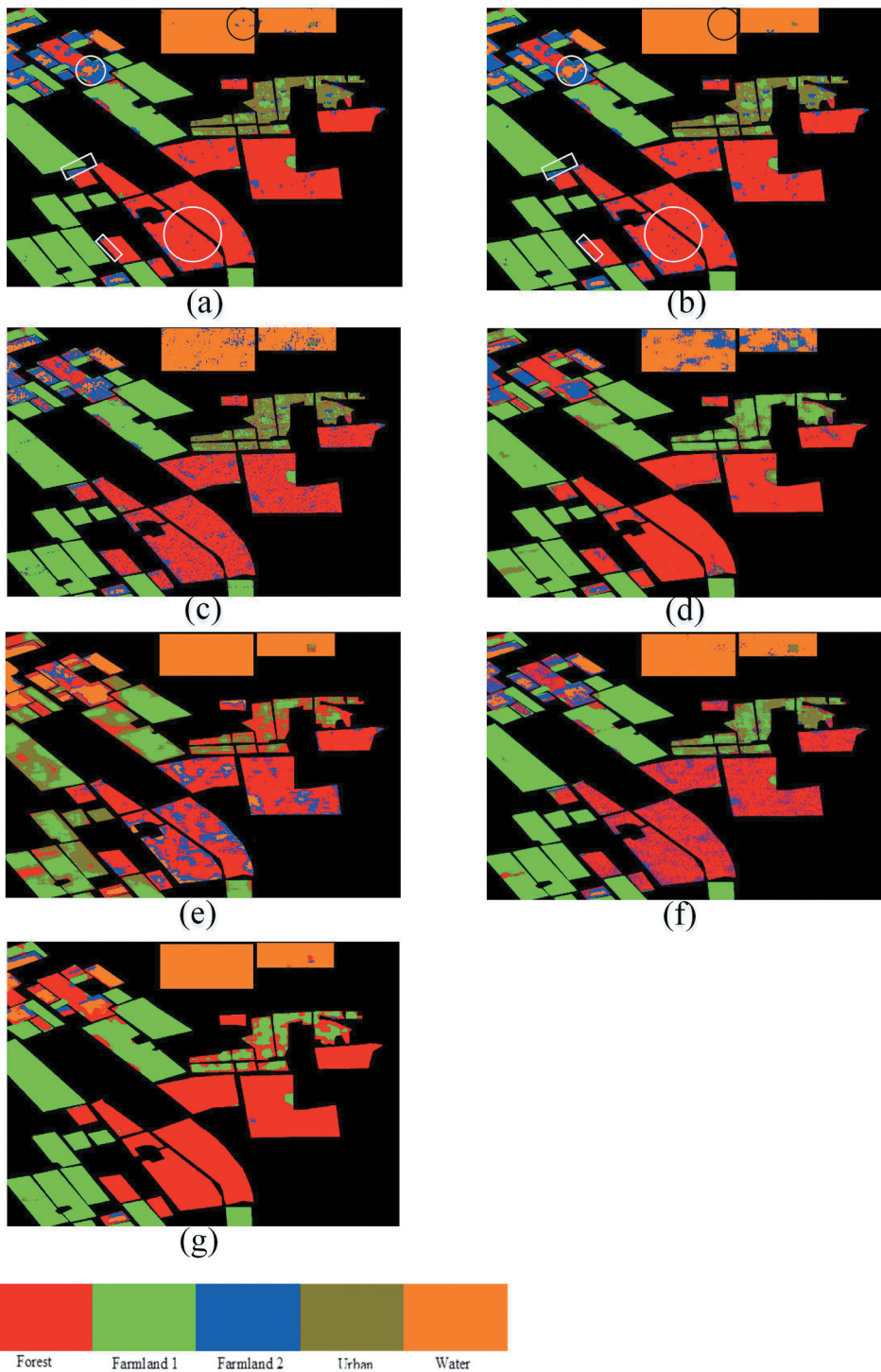
Figure 14. Classification results for the Flevoland SAR image: (a) results of the proposed algorithm; (b) results of the CNN-MRF; (c) results of the CNN; (d) results of the DBN; (e) results of the SAE; and (f) results of the SVM.

Table 5. Classification accuracies of all of the compared algorithms for the Radarsat-2 Flevoland real SAR image.

categories	Proposed algorithm(%)	CNN-MRF(%)	CNN(%)	DBN(%)	SAE (%)	SVM (%)	RF (%)
forest	94.99	92.76	88.68	96.59	86.64	87.38	99.17
farmland 1	97.63	97.58	96.75	96.18	82.33	96.87	97.76
farmland 2	42.32	41.50	40.07	29.35	10.26	32.64	9.73
urban	77.40	73.26	76.37	43.61	68.53	54.73	0.2
water	98.53	99.44	92.86	73.46	99.02	97.33	99.37
OA	89.11	87.79	85.70	81.30	81.51	82.71	81.05

can lead to regional-level misclassification. In addition, some boundaries also exist and can be used to evaluate the abilities of the proposed algorithm and the compared algorithms to retain boundaries. Consequently, this image is explored to evaluate the robustness of the proposed algorithm in eliminating regional-level misclassification and retaining boundaries for different scenes.

The balance coefficients are optimized before the evaluation to obtain the optimized performance of the proposed NHC-MRF, and the results of the optimization are presented in Figure 13. To enable fair comparison, the parameters of the compared algorithms are also optimized. The optimized balance coefficients of the MRF and the NHC-MRF are still 10 and 15, which demonstrates the robustness of the parameters of the proposed model. The optimized coefficients of the label coexisting probability term and the top2 label term are also 0.5 and 0.9.

The classification accuracies of all of the explored algorithms are presented in Table 5. The corresponding classification result maps of the compared algorithms are presented in Figure 14. To enable better discussion and visualization, the confusion matrices of the proposed algorithm and the CNN-MRF algorithm are presented in Figure 15. According to the classification accuracies listed in Table 5, the proposed algorithm still achieves the highest overall accuracy and the highest accuracies for almost of the types of terrain, which demonstrates the robustness of the performance of the proposed algorithm for a high-resolution real SAR image with a high level of speckle noise. The superiority of the proposed algorithm is derived from the ability of the NHC-MRF to eliminate regional-level misclassification and retain boundaries. In particular, of its ability to eliminate regional-level misclassification leads to clear enhancement of the classification accuracy of forest, farmland2, and urban land.
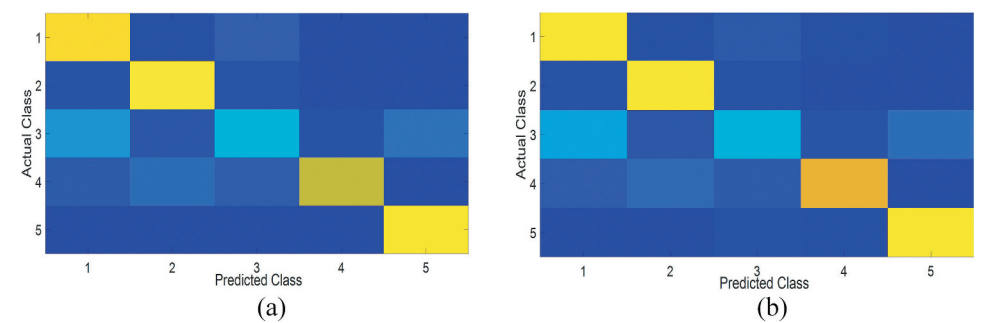

(a)   (b)

Figure 15. Confusion matrices of the proposed algorithm and the CNN-MRF for the radarsat-2 Flevoland SAR image.

By comparing Figures 15(a,b) it can be seen that the confusion between the farmland2, forest, and urban land is noticeably reduced, which leads to the superiority of the proposed algorithm.

This conclusion is also demonstrated by the result maps. As highlighted by the white circles in Figures 14(a,b) a series of regional-level misclassifications can be reduced by applying the NHC-MRF to classify the forest, farmland2, and urban land. This result demonstrates the superiority of the NHC-MRF to eliminate regional-level misclassification, which is derived from the use of refined spatial constraints. The advantage of the retention of the boundaries by the NHC-MRF is also demonstrated by the classification result maps. By comparing Figures 14(a,b) as highlighted by the white rectangles, some of the boundaries and details are also preserved better by the NHC-MRF model. This result demonstrates that the superiority of the NHC-MRF derives from the higher boundary preserving ability resulting from allowing heterogeneity in the neighbourhoods in the NHC-MRF model. The RF with 800 decision trees achieves 81.05% OA, which is about the same with that of SAE and DBN. The RF shows higher speckle noise immunity ability than DBN and SAE because that the classification results of the RF is smoother. However, the RF shows lower discriminating capability for terrains with high inter-class similar than CNN, DBN and SAE. This conclusion can be supported by the classification results of RF on the terrain of farmland2 and the urban. It can be seen that the results of RF on this terrain is almost wrong, which is much worse than that of DBN and SAE. Consequently, the overall accuracy of RF is close to that of DBN and SAE.

3.2.4. Discussion

In this section, we further illustrate the mechanism of the proposed NHC-MRF. In addition, an ablation experiment is conducted to evaluate the effect of the label coexisting probability and the top-2 label term of the NHC index.

(1) First, the classification accuracies and result maps of the proposed framework and the CNN-MRF with different balance coefficients are illustrated to demonstrate that the better performance of the NHC-MRF cannot be achieved by simply reducing the balance coefficient of the binary energy term. The values of the balance coefficient are set as 8, 10, and 15. The overall accuracies are listed in Table 6, and the result maps are presented in Figure. 16. It can be seen that the OA of the proposed algorithm maintains its superiority over the CNN-MRF for different balance coefficients. In addition, the highest OA of the CNN-MRF is still lower than the lowest value of the proposed algorithm, which demonstrates that the superiority of the NHC-MRF cannot be obtained by simply reducing the balance coefficient. This conclusion is further demonstrated by the result maps presented in Figure. 16. It can be seen that, in comparison with the CNN-MRF, the proposed NHC-MRF can always eliminate the regional-level misclassification, which demonstrates that the superiority of the NHC-MRF is derived from the mechanism of the NHC index.

(2) Second, an ablation experiment is performed to evaluate the effects of the label coexisting probability term and the top-2 term using the Flevoland dataset. In this experiment, the coexisting probability and the top-2 term are removed separately to evaluate the classification performance of the proposed algorithm. The classification accuracies are listed in Table 7.
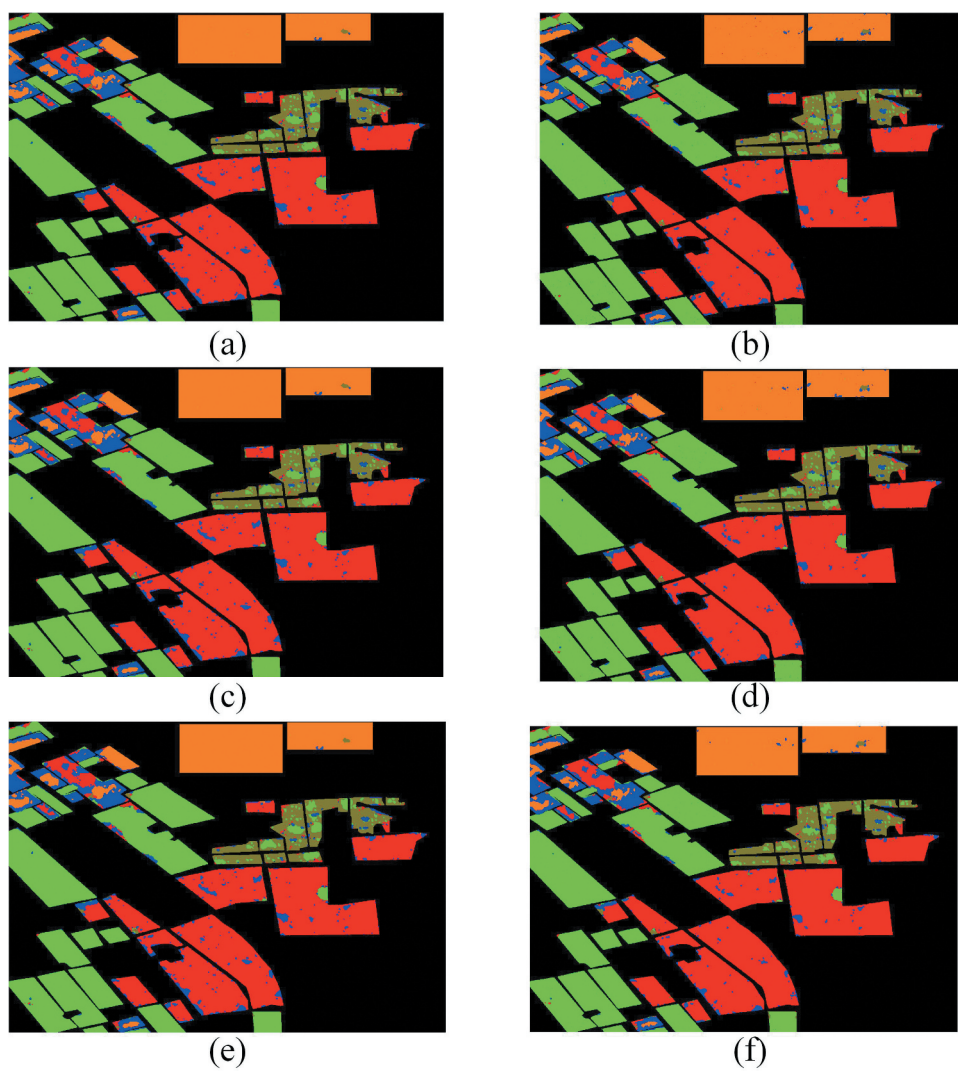
Figure 16. Classification result maps of the proposed algorithm and the CNN-MRF with different balance coefficient values.

Table 6. Oas of the proposed algorithm and the CNN-MRF for different balance coefficient values.

Methods	$\beta = 8$	$\beta = 10$	$\beta = 15$
Proposed	88.13%	88.40%	88.64%
CNN-MRF	87.75%	87.79%'	87.69%'

It can be seen from Table 7 that the label coexisting probability term makes a greater contribution to the enhancement of the classification accuracy. In contrast, no enhancement of the OA can be achieved by simply using the top-2 label term because this term is used to refine the label coexisting probability term. However, the top-2 label term can lead to enhancement of the classification
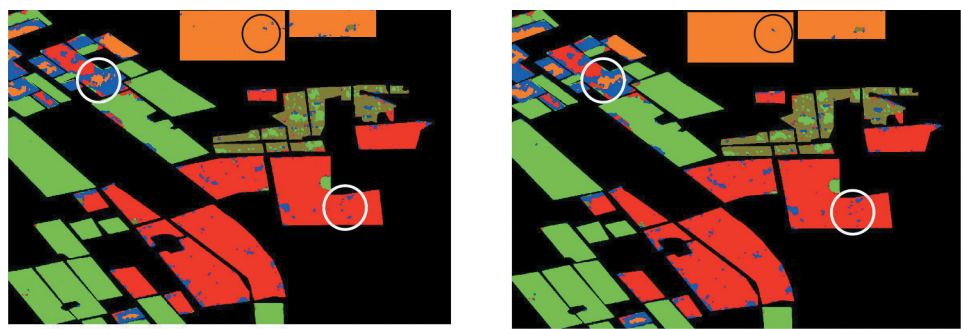
Figure 17. Effect of utilizing the top-2 label term.

Table 7. Results of the ablation experiment.

coexisting probability term	top-2 label term	Accuracy
√	√	85.70%
√	×	85.23%′
×	√	84.79%

accuracy when the label coexisting probability term is utilized. For some regions, the probability distributions output by the CNN are not reliable, which leads to misclassification caused by the unsuitable label coexisting probability. By using the top-2 label term, the misclassification can be refined (highlighted by the white circles in Figure 17). In some cases such as the results highlighted by the black circles, the assumption that the top-2 labels term is based on (see Section 3.3.2) may be improper. Consequently, some misclassification will be caused by using the top2 labels term. However, the top2 label term can lead to the elimination of misclassification in most cases.

(3) The proposed NHC-MRF is compared with the RCC-MRF to further demonstrate the effectiveness of the NHC-MRF model. The synthetic SAR image is utilized to evaluate the NHC-MRF and the RCC-MRF because this image is suitable for evaluating
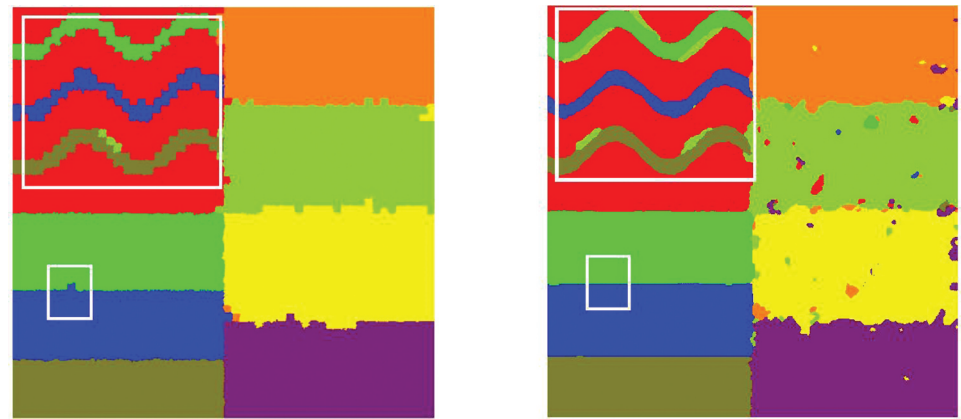

Figure 18. Comparison of the NHC-MRF and the RCC-MRF.

the ability to retain the boundaries due to the exactly known boundaries. The initial results of these two models are obtained using the same CNN model. The classification result maps of the NHC-MRF and RCC-MRF are presented in Figs. 19(a) and 19(b). The RCC-MRF achieves a slightly lower overall accuracy (96.08%) than the NHC-MRF. The superiority of the RCC-MRF is derived from the higher level of smoothing, in which the super-pixel is regarded as the unit of classification. However, the ability of the RCC-MRF to retain boundaries is much lower than that of the NHC-MRF, which offsets the superiority of the super-pixel based smoothing. As highlighted by Figures 18(a,b) the NHC-MRF can better retain the boundaries of the curved narrow regions than the RCC-MRF. Based on this classification result, it can be concluded that the superiority of the NHC-MRF is derived from the combination of boundary retention and regional-level misclassification elimination.

4. Conclusions

In this study, a novel SAR image classification framework was developed. This framework fuses the CNN and the newly proposed NHC-MRF models. The NHC-MRF contains the NHC index in the binary energy term, which was developed to improve the conventional neighbourhood consistency (homogeneous) spatial constraint by adaptively allowing the existence of heterogeneity in the neighbourhoods. This NHC index consists of a label coexisting probability term and a top-2 label term, which are obtained based on the outputs of the CNN. The label coexisting probability term describes the confidence of the heterogeneity in the neighbourhood, and the top-2 label term is used as the correction of the label coexisting probability term. Due to the effect of the label coexisting probability term, the details can be protected by allowing heterogeneity, and the regional-level misclassification can be eliminated by changing the impact of the neighbourhoods by using coexisting probability based weighting. Experiments were conducted using three SAR datasets, and the superiority of the proposed framework was demonstrated. Analysis and an ablation experiment were also conducted to further illustrate the mechanisms of the NHC-MRF model. Finally, by comparing the proposed model with the RCC-MRF based framework, the superiority and limitations of the proposed framework were investigated. In future work, the NHC index will be developed for regional-level SAR image classification frameworks.

Acknowledgements

Disclosure statement

Funding

ORCID

Lu Jia http://orcid.org/0000-0002-7982-797X

References

Anjun, Z., Yang, X., Jia, L., Ai, J., and Xia, J. (2019). SRAD-CNN for adaptive synthetic aperture radar image classification. International Journal of Remote Sensing, 40(9), 3461–3485. https://doi.org/10.1080/01431161.2018.1547450 .

Cao, Y., Y. Wu, P. Zhang, W. Liang, and M. Li. 2019. "Pixel-Wise PolSar Image Classification via a Novel Complex-Valued Deep Fully Convolutional Network." Remote Sensing 11 (22): 2653. https://doi.org/10.3390/rs11222653.

Chapelle, O., P. Haffner, and V. N. Vapnik. 1999. "Support Vector Machines for Histogram-Based Image Classification." IEEE Transaction on Neural Network 10 (5): 1055–1064. https://doi.org/10.1109/72.788646.

Dekker, R. J. 2003. "Texture Analysis and Classification of ERS SAR Images for Map Updating of Urban Areas in the Netherlands." IEEE Transactions on Geoscience and Remote Sensing 41 (9): 1950–1958. https://doi.org/10.1109/TGRS.2003.814628.

Duan, Y., F. Liu, L. Jiao, P. Zhao, and L. Zhang. 2017. "SAR Image Segmentation Based on Convolutional-Wavelet Neural Network and Markov Random Field." Pattern Recognition 64 (C): 255–267. https://doi.org/10.1016/j.patcog.2016.11.015.

Ferro-Famil, L., E. Pottier, and J. S. Lee. 2000. "Unsupervised Classification of Multifrequency and Fully Polarimetric SAR Images Based on the H/a/alpha-Wishart Classifier." IEEE Transactions on Geoscience and Remote Sensing 39 (11): 2332–2342. https://doi.org/10.1109/36.964969.

Fukuda, S., and H. Hirosawa. 1999. "A Wavelet-Based Texture Feature Set Applied to Classification of Multifrequency Polarimetric SAR Images." IEEE Transactions on Geoscience and Remote Sensing 37 (5): 2282–2286. https://doi.org/10.1109/36.789624.

Gao, F., T. Huang, J. Wang, J. Sun, A. Hussain, and E. Yang. 2017. "Dual-branch deep convolution neural network for polarimetric SAR image classification." Applied Sciences 7 (5): 447. https://doi.org/10.3390/app7050447.

Geng, J., H. Wang, J. Fan, and X. Ma. 2017. "Deep Supervised and Contractive Neural Network for SAR Image Classification." IEEE Transactions on Geoscience and Remote Sensing 55 (4): 2442–2459. https://doi.org/10.1109/TGRS.2016.2645226.

He, S., R. Lau, W. Liu, Z. Huang, and Q. Yang. 2015. "-SuperCNN: A Superpixelwise convolutional neural network for salient object detection." International Journal of Computer Vision 115 (3): 330–344. https://doi.org/10.1007/s11263-015-0822-0.

He, K., X. Zhang, S. Ren, and J. Sun 2016. Deep Residual Learning for Image Recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770–778).

Jiang, Y., M. Li, P. Zhang, X. Tan, and W. Song. 2021. "Hierarchical Fusion Convolutional Neural Networks for Sar Image Segmentation." Pattern Recognition Letters 174 (115–123): 115–123. https://doi.org/10.1016/j.patrec.2021.04.005.

Krizhevsky, A., I. Sutskever, and G. Hinton 2012 ImageNet Classification with Deep Convolutional Neural Networks. *International Conference on Neural Information Processing Systems*. (vol. 25, pp. 1097–1105). LakeTahoe, USA.

Lécun, Y., L. Bottou, Y. Bengio, and P. Haffner. 1998. "Gradient-Based Learning Applied to Document Recognition." *Proceedings of the IEEE* 86 (11): 2278–2324. https://doi.org/10.1109/5.726791.

Liang, W., Y. Wu, M. Li, Y. Cao, and X. Hu. 2021. "High-Resolution SAR Image Classification Using Multi-Scale Deep Feature Fusion and Covariance Pooling Manifold Network." *Remote Sensing* 13 (2): 328. https://doi.org/10.3390/rs13020328.

Liu, H., S. Yang, S. Gou, D. Zhu, R. Wang, and L. Jiao. 2017. "Polarimetric SAR Feature Extraction with Neighbourhood Preservation-Based Deep Learning." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 10 (4): 1456–1466. https://doi.org/10.1109/JSTARS.2016.2618891.

Lv, Q., Y. Dou, X. Niu, J. Xu, and B. Li 2014. "Classification of Land Cover Based on Deep Belief Networks Using Polarimetric RADARSAT-2 Data. *IEEE Geoscience and Remote Sensing Symposium* (pp. 4679–4682). Quebec, Canada.

Maryam, I. 2022. "Two-Step Discriminant Analysis Based Multi-View Polarimetric SAR Image Classification with High Confidence." *Scientific Report* 12 (1): 5984. https://doi.org/10.1038/s41598-022-09871-w.

McNairn, H., A. Kross, D. Lapen, R. Caves, and J. Shang. 2014. "Early Season Monitoring of Corn and Soybeans with TerraSAR-X and RADARSAT-2." *International Journal of Applied Earth Observation and Geoinformation* 28:252–259. 10.1016/j.jag.2013.12.015.

Modava, M., G. Akbarizadeh, and M. Soroosh. 2019. "Hierarchical Coastline Detection in Sar Images Based on Spectral-Textural Features and Global-Local Information." *IET Radar Sonar & Navigation* 13 (12): 2183–2195. https://doi.org/10.1049/iet-rsn.2019.0063.

Ressel, R., A. Frost, and S. Lehner. 2015. "A Neural Network-Based Classification for Sea Ice Types on X-Band SAR Images." *IEEE Journal of Selected Topics in Applied Earth Observations & Remote Sensing* 8 (7): 3672–3680. 10.1109/JSTARS.2015.2436993.

Szegedy, C., W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, et al. 2015 Going Deeper with Convolutions. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1–9. https://doi.org/10.1109/CVPR.2015.7298594.

Tison, C., J. M. Nicolas, F. Tupin, and H. Maitre. 2004. "A New Statistical Model for Markovian Classification of Urban Areas in High-Resolution SAR Images." *IEEE Transactions on Geoscience and Remote Sensing* 42 (10): 2046–2057. https://doi.org/10.1109/TGRS.2004.834630.

Tzeng, Y. C., and K. S. Chen. 1998. "-A fuzzy neural network to SAR image classification." *IEEE Transactions on Geoscience and Remote Sensing* 36 (1): 301–307. https://doi.org/10.1109/36.655339.

Xiang, D., T. Tang, C. Hu, Y. Li, and Y. Su. 2014. "A Kernel Clustering Algorithm with Fuzzy Factor: Application to SAR Image Segmentation." *IEEE Geoscience and Remote Sensing Letters* 11 (7): 1290–1294. https://doi.org/10.1109/LGRS.2013.2292820.

Zhang, P., X. Tan, B. Li, Y. Jiang, W. Song, M. Li, and Y. Wu. 2021. "PolSar Image Classification Using Hybrid Conditional Random Fields Model Based on Complex-Valued 3-D CNN." *IEEE Transactions on Aerospace and Electronic Systems* 57 (3): 1713–1730. https://doi.org/10.1109/TAES.2021.3050648.

Zhang, Z., H. Wang, F. Xu, and Y. Q. Jin. 2017. "Complex-Valued Convolutional Neural Network and Its Application in Polarimetric SAR Image Classification." *IEEE Transactions on Geoscience and Remote Sensing* 55 (12): 7177–7188. https://doi.org/10.1109/TGRS.2017.2743222.

Zhang, A., X. Yang, S. Fang, and J. Ai. 2020. "Region Level SAR Image Classification Using Deep Features and Spatial Constraints." *ISPRS Journal of Photogrammetry and Remote Sensing* 163:36–48. https://doi.org/10.1016/j.isprsjprs.2020.03.001.

Zhang, A., X. Yang, L. Jia, J. Ai, and J. Xia. 2019. "SRAD-CNN for Adaptive Synthetic Aperture Radar Image Classification." *International Journal of Remote Sensing* 40 (9): 3461–3485. https://doi.org/10.1080/01431161.2018.1547450.

Zhao, W., and S. Du. 2016. "Spectral–Spatial Feature Extraction for Hyperspectral Image Classification: A Dimension Reduction and Deep Learning Approach." *IEEE Transactions on Geoscience and Remote Sensing* 54 (8): 4544–4554. https://doi.org/10.1109/TGRS.2016.2543748.

Zhou, Y., H. Wang, F. Xu, and Y. Jin. 2016. "Polarimetric SAR Image Classification Using Deep Convolutional Neural Networks." *IEEE Geoscience and Remote Sensing Letters* 13 (12): 1935–1939. https://doi.org/10.1109/LGRS.2016.2618840.