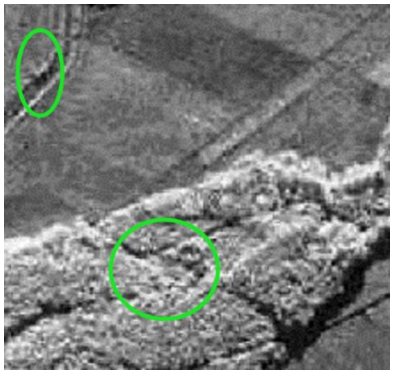
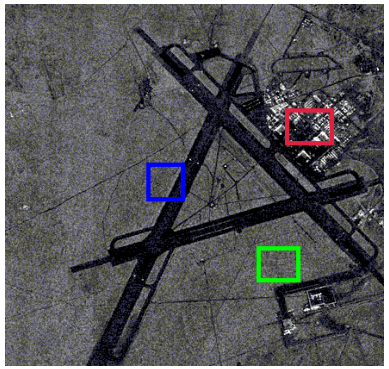
Hierarchical multinomial latent model with G^0 distribution for remote sensing image segmentation

Yiping Duan, Xiaoming Tao, Chaoyi Han, Jianhua Lu
Department of Electronic Engineering
Tsinghua University
Beijing, China

Motivation

- Background: Remote sensing images are essentially rich in scales, structures, objects and so on. This makes that semantic segmentation is still a challenging problem. Our paper focus on the SAR remote sensing images.
- Two key factors: The labeling consistency and detail preservations are equally important for remote sensing image semantic segmentation.
- Bottleneck: multinomial latent model with amplitude and texture feature is a popular contextual segmentation method. The single spatial contextual relationship is difficult to deal with the heterogamous structures of remote sensing images.

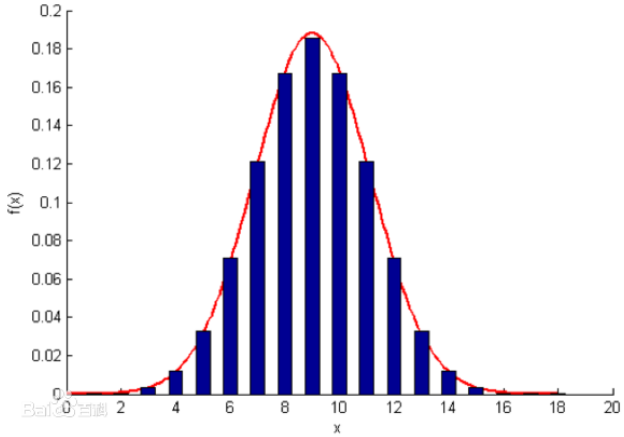




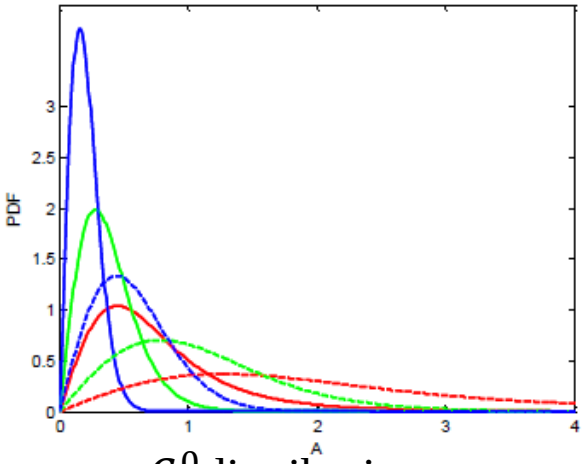
- Motivation: The single scale model limits the ability of the model to capture the global information. The multiscale information should be incorporated into our model.

Motivation

- Scattering characteristic: the scattering statistics reflects the roughness of the remote sensing images.
- Bottleneck: Due to the special principle of remote sensing images, especially the SAR images, the Gaussian based distribution is generally no longer confirmed.
- Motivation: Non-stationary G^0 distribution can flexibly model the regions with different characteristics, homogeneous, heterogeneous and extremely heterogeneous regions, which reflects the structural information of the remote sensing images.



Gaussian distribution



G^0 distribution

A. C. Frery, H. J. Muller, C. C. F. Yanasse and S. J. S. Sant'Anna, "A model for extremely heterogeneous clutter," IEEE Trans. Geosci. Remote Sens., vol. 35, no. 3, pp. 648-659, May. 1997.

Scattering statistics of the SAR images

- Different distributions and properties of the SAR images

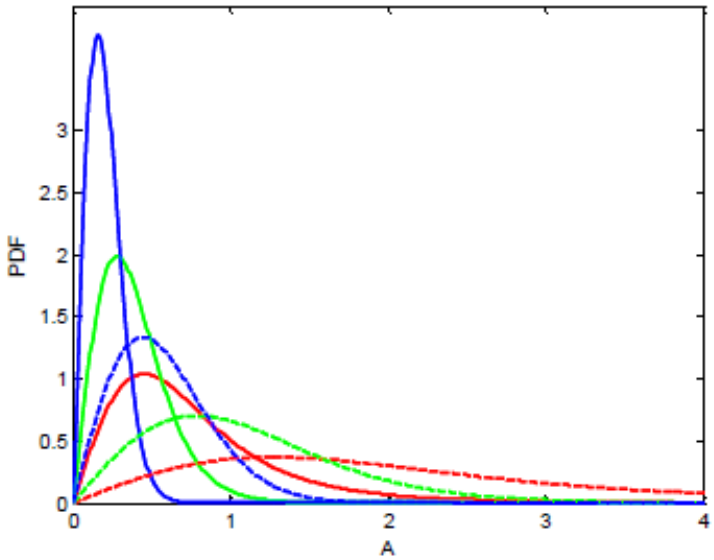
Distribution	PDF	Property
Lognormal [3]	$p(A;n,\sigma)=\frac{1}{\sigma A\sqrt{2\pi}}\exp\left[-\frac{(\ln A-n)^2}{2\sigma^2}\right], n\in\mathbb{R}, \sigma>0, A>0$	Single look homogeneous
Weibull [3]	$p(A;\eta,\mu)=\frac{\eta}{\mu^{\eta}}A^{\eta-1}\exp\left[-\left(\frac{A}{\mu}\right)^{\eta}\right], \eta,\mu>0, A\geq 0$	Single look, Homogeneous,
Gamma [3]	$p(A;L,\mu)=\frac{1}{\Gamma(L)}\left(\frac{L}{\mu}\right)^{L}A^{L-1}\exp\left(-\frac{LA}{\mu}\right), L,\mu>0, A\geq 0$	Multi look, Homogeneous,
Nakagami [3]	$p(A;L,\lambda)=\frac{2}{\Gamma(L)}(\lambda L)^{L}A^{2L-1}\exp\left(-\lambda LA^2\right), L,\lambda>0, A\geq 0$	Multi look, Homogeneous,
K -distribution [27]	$p(A;\mu,L,n)=\frac{2}{\Gamma(L)\Gamma(n)}A^{\frac{L+n}{2}-1}C^{L+n}K_{n-L}\left(2A^{1/2}\right)$, $C=\left(\frac{Ln}{\mu}\right)^{1/2}, \mu>0, 0<L<n, A\geq 0$	Multi look, Heterogeneous,
G^0 -distribution [27]	$p(A;\alpha,\gamma,n)=\frac{2n^{n}\Gamma(n-\alpha)\gamma^{-\alpha}A^{2n-1}}{\Gamma(n)\Gamma(-\alpha)\left(\gamma+nA^2\right)^{n-\alpha}}, -\alpha,\gamma,n,A>0$	Multi look, Homogeneous, Heterogeneous, Extremely heterogeneous.

G^0 distribution

- The density function of G^0 distribution is described as

$$p(A;\alpha,\gamma,n)=\frac{2n^n\Gamma(n-\alpha)\gamma^{-\alpha}A^{2n-1}}{\Gamma(n)\Gamma(-\alpha)(\gamma+nA^2)^{n-\alpha}},-\alpha,\gamma,n,A>0$$

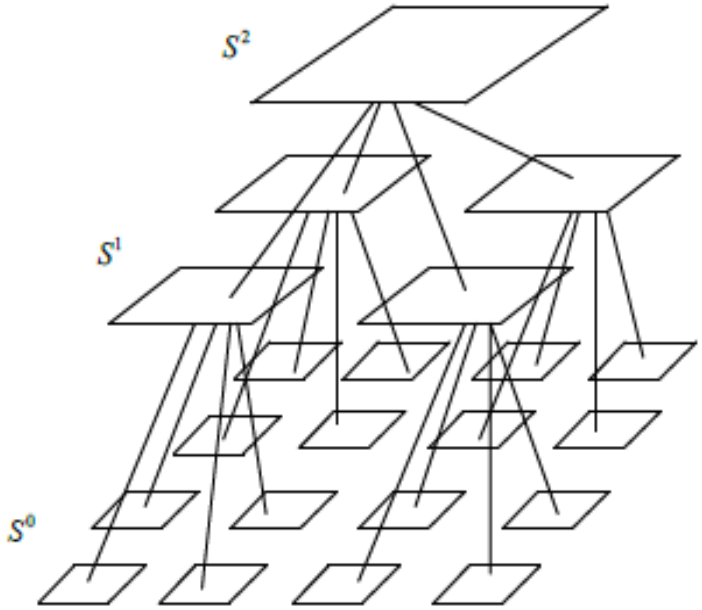
where the parameter n is the number of looks. The parameter α and γ are used to characterize the roughness and scale of SAR image, respectively



Densities of distribution with different parameters.

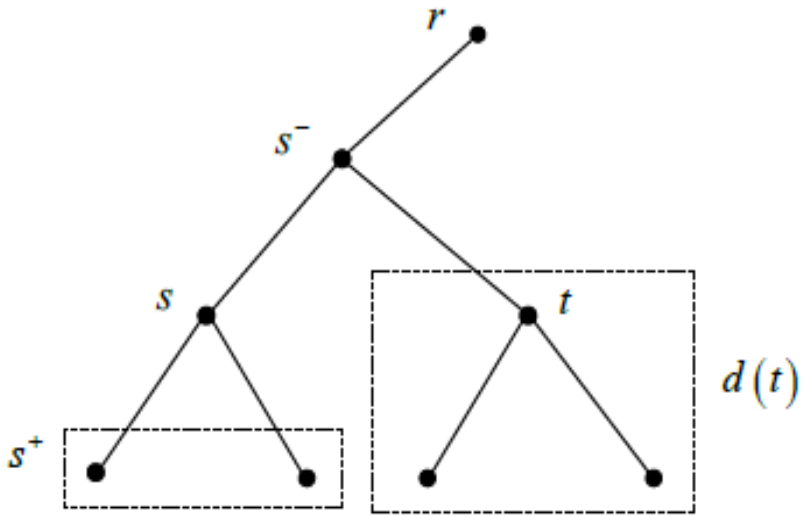
Hierarchical multinomial latent model with G^0 distribution

- Our model is constructed on the quad-tree structure



(a)

Quad-tree structure



(b)

Quad-tree notations

The quad tree is obtained by two layer wavelet decomposition. $s^0$is the finest scale and $s^2$is the coarsest scale. The local and global information can be captured by this structure.

Hierarchical multinomial latent model with G^0 distribution

- Our model is defined as follows

We define that $Y = (Y_s)_{s \in S}$ is the observed field and $X = (X_s)_{s \in S}$ is the label field, where $S = S^0 \cup S^1 \cup S^2$. Let $\Omega = \{1,2, \ldots, K\}$ be a finite set, the labeling is performed on this set. Our goal is to estimate labels X given a set of observed data Y and the image segmentation task can be formulated as the modes of posterior marginal (MPM)

$$\forall s \in S, \hat{x}_s = \arg\max_{x_s \in \Omega} p(X_s = x_s \mid Y = y)$$
$$\triangleq \arg\max_{x_s \in \Omega} p(x_s \mid y)$$

The model include three parts: the likelihood model, spatial contextual model and transition Probabilities. The three portions are introduced in the following part. Then, MPM are used to obtain the segmentation result by bottom-up and up-down way.

Likelihood model: combine amplitude and texture

- The likelihood model combine amplitude and texture features

$$p(y_s \mid \theta_k) = p_A(y_s \mid \theta_k)\, p_T(y_{\partial s} \mid \theta_k)$$

Amplitude feature

Texture feature

- Texture density is described as t distribution

$$p_T(\mathbf{y}_{\partial s} \mid y_s, \boldsymbol{\mu}_k, \beta_k, \varphi_k) = \frac{\Gamma\left((1+\beta_k)/2\right)}{\Gamma(\beta_k/2)(\pi\beta_k\varphi_k)^{1/2}} \times \left[1 + \frac{\left(y_n - \mathbf{y}_{\partial s}^T \boldsymbol{\mu}_k\right)^2}{\beta_k \varphi_k}\right]^{-\frac{\beta_k+1}{2}}$$

where $y_{\partial s}$ is the surrounding pixels of y_s. μ_k, β_k, φ_k are the parameters of the distribution.

- Amplitude density is described as the G^0 distribution, which can describe the homogeneous, heterogeneous and extremely heterogeneous regions

Spatial contextual model

- Spatial context model is also the prior model. It is the process to estimate the prior probabilities on the obtained segmentation map. We resort to the single-scale multinomial latent model which takes into account the smooth label constraints. By employing the logistic function, the prior density is written as:

$$p(x_s \mid x_t) = \frac{\exp\left[\eta\left(1+\sum_{x_t \in \eta_s} \delta(x_t = l)\right)\right]}{\sum_{m=1}^{K}\left[\eta\left(1+\sum_{x_t \in \eta_s} \delta(x_t = m)\right)\right]}$$

$$\delta(x_t = m) = \begin{cases} 1, x_t = m \\ 0, x_t \neq m \end{cases}$$

This term captures the relationships between the pixel label and the surrounding labels. The local spatial relationships are well described.

Transition Probabilities

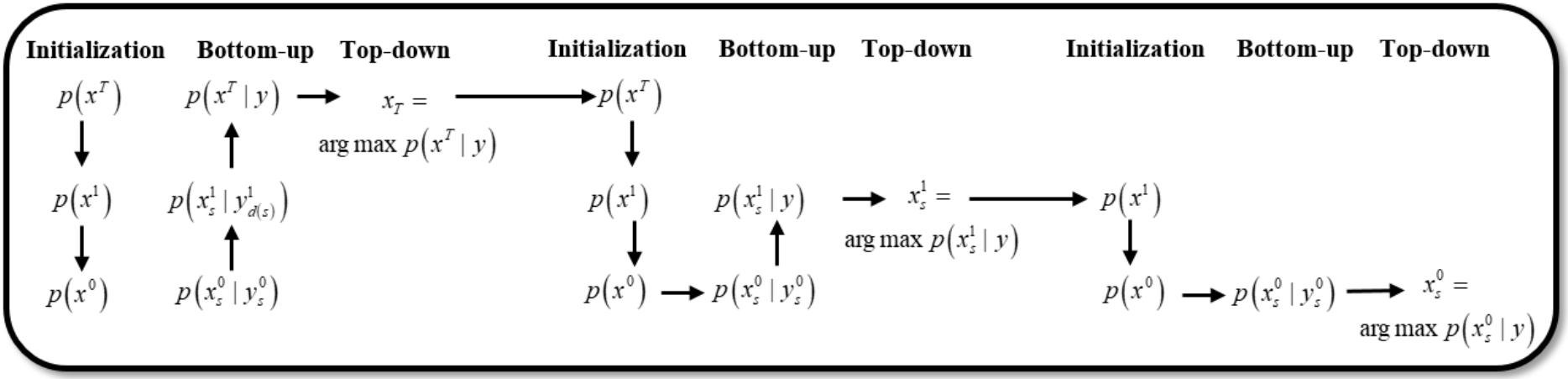
- The transition probabilities represent the causalities of the statistical interactions between the different scales. It plays a decisive role in hierarchical multinomial latent model. We use the transition probability as follows:

$$\forall s \in S, p\left(x_s = j \mid x_{s^-} = i\right) = \begin{cases} \lambda_t, & if\ i = j, \\ \dfrac{1-\lambda_t}{K-1}, & otherwise. \end{cases}$$

where$x_s = \mathrm{j}$ represents that x_s belongs to the jth class. λ_t is model parameter with $\lambda_t > \frac{1}{K}$. In addition, λ_t is independent of the scale. This model favors identify between parent and children. In our approach, λ_t is set as 0.9.

MPM labeling consistency

- Two-pass Computation of Posterior Probabilities



- Bottom up pass:

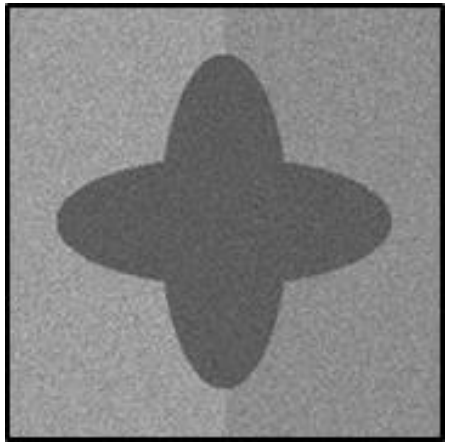
$$p(x_s \mid y_{d(s)}) \propto p(y_s \mid x_s) \prod_{\omega \in s^+} \sum_{x_\omega} \left[\frac{p(x_\omega \mid y_{d(\omega)})}{p(x_\omega)} p(x_\omega \mid x_s) \right]$$

- Top down pass:

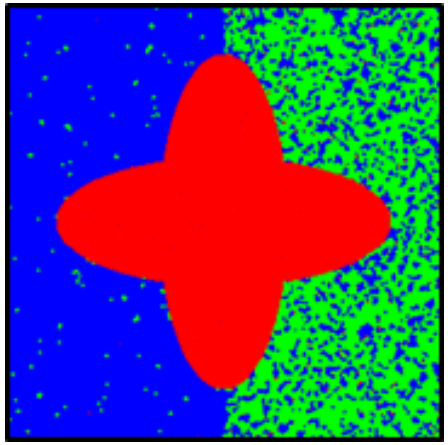
$$\forall s \in S^T, \hat{x}_s = \arg\max_{x_s} p(x_s \mid y)$$

Experiments

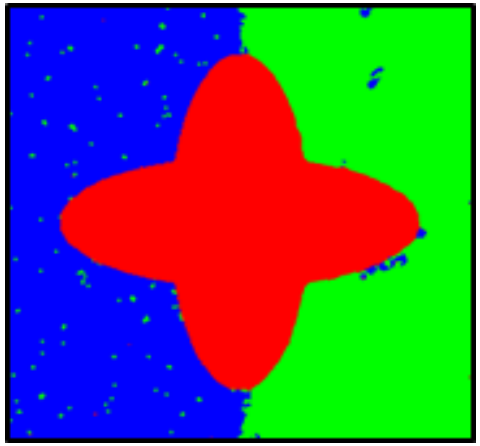
- Experimental results on the synthetic SAR images.



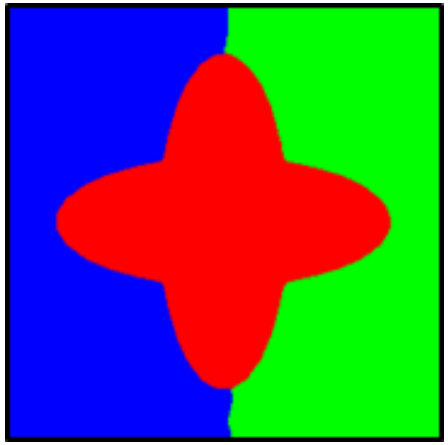
Original image



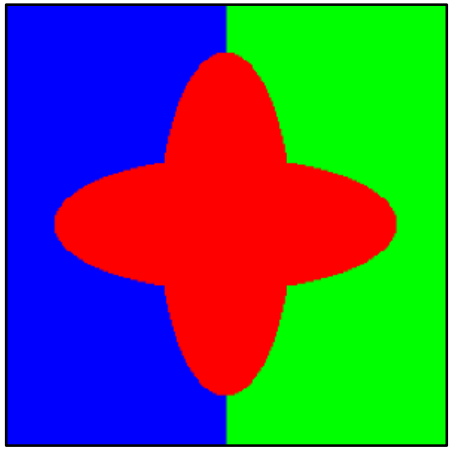
MRF



ATML-CEM



Our



Ground truth

Our approach can improve the labeling consistency and preserve the details very well. Our experimental result is the closet to the ground truth segmentation.

Experiments

- Experimental results on the synthetic SAR images.

Table I KAPPA AND GCE MEASURES OF SYNs

	MRF		HMRF		ATML-CEM		HML-G^0	
	KAPPA	GCE	KAPPA	GCE	KAPPA	GCE	KAPPA	GCE
1LOOK	0.0646	0.3897	0.4064	0.1425	0.3559	0.1794	**0.4116**	**0.1275**
2LOOK	0.4535	0.2561	0.4629	**0.1654**	0.6257	0.1820	**0.6585**	0.1726
3LOOK	0.5458	0.2342	0.5150	0.1783	**0.8061**	**0.1167**	0.7541	0.1379
4LOOK	0.5875	0.2125	0.4860	0.1734	0.8735	0.0786	**0.8870**	**0.0691**
5LOOK	0.6872	0.1909	0.5220	0.1817	0.9632	0.0211	**0.9854**	**0.0096**
6LOOK	0.6775	0.1786	0.6137	0.1793	0.9696	0.0243	**0.9872**	**0.0084**
7LOOK	0.7192	0.1626	0.6203	0.1795	0.9721	0.0184	**0.9907**	**0.0062**
8LOOK	0.7457	0.1492	0.6995	0.1576	0.9895	0.0076	**0.9939**	**0.0041**
9LOOK	0.7963	0.1251	0.7788	0.1261	0.9933	0.0052	**0.9964**	**0.0026**
10LOOK	0.8181	0.1155	0.8527	0.0886	0.9935	0.0047	**0.9957**	**0.0029**

[1] H. Derin and H. Elliott, "Modeling and segmentation of noisy and textured images using Gibbs random fields," IEEE Trans. Pattern Anal. Mach. Intell., vol. 9, no. 1, pp. 39-55, Jan. 1987.

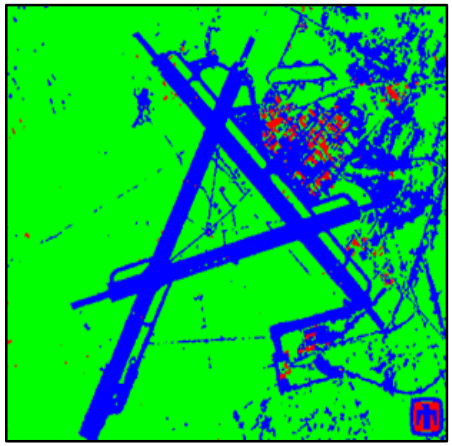
[2] K. Kayabol and J. Zerubia, "Unsupervised Amplitude and Texture Classification of SAR Images with Multinomial Latent Model," IEEE Trans. Image Process., vol. 22, no. 2, pp. 561-572, Feb. 2013.

[3] A. Voisin, et al., "Classification of very high resolution SAR images of urban areas using copulas and texture in a hierarchical markov random field model," IEEE Geosci. Remote. Sens. Lett., vol. 10, no. 1, Jun. 2013.

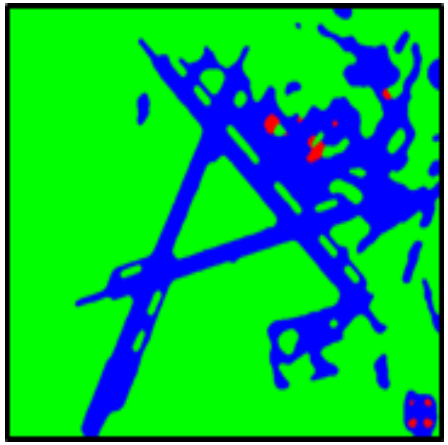
Experiments

- Experimental results on the real SAR images.

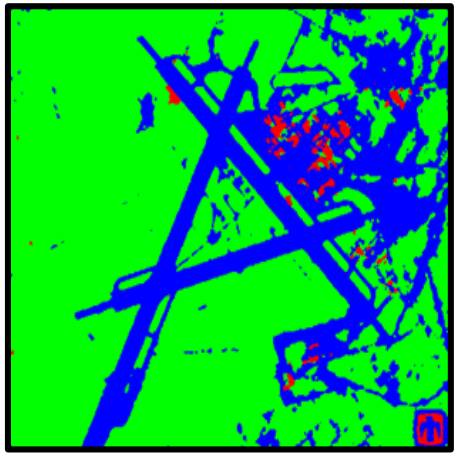
Original image



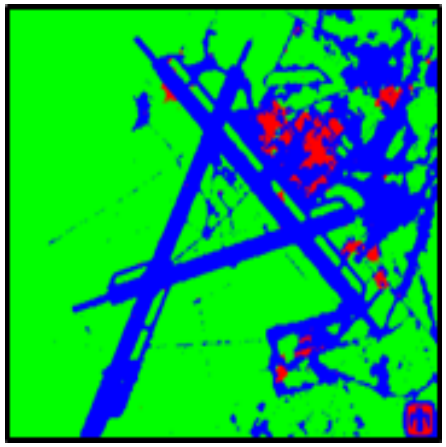
MRF



ATML-CEM



HMRF



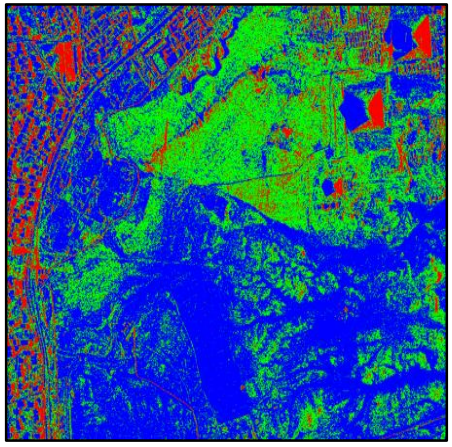
Our

Ground truth

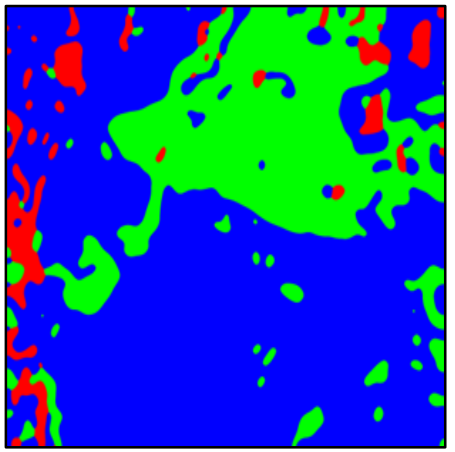
Experiments

- Experimental results on the real SAR images.

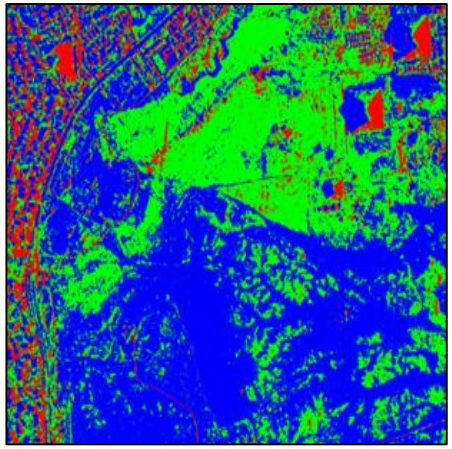
Original image



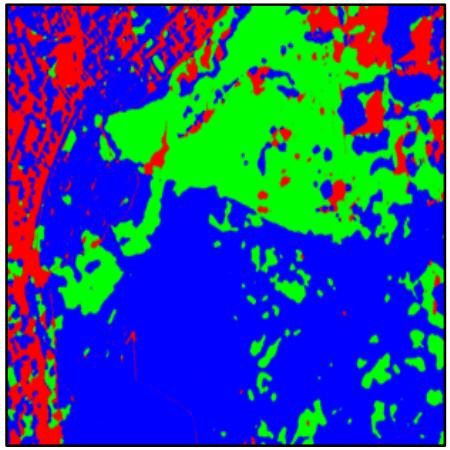
MRF



ATML-CEM



HMRF



Our

Ground truth

thank you!