Bayes-SAR Net: Robust SAR Image Classification with Uncertainty Estimation Using Bayesian Convolutional Neural Network

Dimah Dera*, Ghulam Rasool*, Nidhal C. Bouaynaya*, Adam Eichen†, Stephen Shanko†, Jeff Cammerata† and Sanipa Arnold†
*Department of Electrical and Computer Engineering — Rowan University, Glassboro, NJ, USA
†Lockheed Martin Rotary and Mission Systems — Moorestown, NJ, USA

***Abstract*—Synthetic aperture radar (SAR) image classification is a challenging problem due to the complex imaging mechanism as well as the random speckle noise, which affects radar image interpretation. Recently, convolutional neural networks (CNNs) have been shown to outperform previous state-of-the-art techniques in computer vision tasks owing to their ability to *learn* relevant features from the data. However, CNNs in particular and neural networks, in general, lack uncertainty quantification and can be easily deceived by adversarial attacks. This paper proposes *Bayes-SAR Net*, a Bayesian CNN that can perform robust SAR image classification while quantifying the uncertainty or confidence of the network in its decision. *Bayes-SAR Net* propagates the first two moments (mean and covariance) of the approximate posterior distribution of the network parameters given the data and obtains a predictive mean and covariance of the classification output. Experiments, using the benchmark datasets Flevoland and Oberpfaffenhofen, show superior performance and robustness to Gaussian noise and adversarial attacks, as compared to the SAR-Net homologue. *Bayes-SAR Net* achieves a test accuracy that is around 10% higher in the case of adversarial perturbation (levels $\geqslant 0.05$).**

I. Introduction

Synthetic aperture radar (SAR) is an important remote sensing tool that provides high-resolution images of earth surface under all-weather and day-and-night conditions [1]. The image data from SAR is used in various applications ranging from environmental and earth system monitoring [2], geoscience and climate change research [3], 2-D and 3-D mapping [4], 4-D mapping (space and time) [5] and security and screening [1]. One of the advanced types of SAR is the polarimetric SAR (PolSAR), which can penetrate observed objects to a certain extent and record the complete scattering information of these objects [6]. The fully polarimetric waveforms enable one to capture the scatter characteristics/matrix of objects/targets [6]. Given the wide range of applications of SAR and PolSAR images, it is crucial to process (e.g., classify) these images using robust and reliable machine learning algorithms [7], [8]. However, due to the unique nature, complex imaging mechanism, and random speckle noise in SAR/PolSAR images, the robust multi-class classification remains a challenging problem [9], [10].

Recently, convolutional neural networks (CNNs), a special class of deep neural networks (DNNs) that specialize in processing grid data, have achieved human-level classification performance on object recognition from images of natural scenes [11]. In the field of SAR image processing and interpretation, CNN-based models have recently demonstrated superior performance [12] as compared to traditional machine learning techniques, such as support vector machines (SVM) [13], Wishart maximum likelihood (Wishart ML) [14] and random forest [15]. Traditional classification algorithms depend on hand-engineered features that usually require significant prior knowledge and domain expertise. In contrast, CNNs can learn hierarchical features and appropriate representations automatically from the data without requiring explicit expert human intervention [16]. However, the features extracted by CNNs may not be robust due to many challenges including over-fitting and the learned classifiers may fail under small perturbations in the data, i.e., adversarial noise [17]. For images, such perturbations are often too small to be perceptible, yet they completely fool the models. We argue that this vulnerability to designed perturbation attacks could be mitigated by propagating uncertainty across the network. In general, the quantification of uncertainty in the prediction is pivotal for the deployment of these algorithms in real-world scenarios, including SAR/PolSAR applications.

Bayesian neural networks are DNNs that provide a principled approach to reason about uncertainty by introducing probability distributions over the unknown parameters (i.e., network weights and biases). In a Bayesian setting, all information about the unknown parameters is provided by their posterior distribution given the data. Since it is very hard to obtain the posterior distribution of the network parameters given the training data, variational inference (VI) advances an approximation technique that changes the problem of density inference to an optimization problem [18]–[20]. Specifically, VI proposes to minimize a measure, e.g., the Kullback-Leibler (KL) divergence, between an approximating family of distributions and the true unknown posterior density function.

In this paper, we propose a novel Bayesian CNN, referred to as *Bayes-SAR Net*, which performs SAR image classification and uncertainty estimation in a unique framework. *Bayes-SAR Net* is built upon the extended VI framework proposed in [21] for propagating uncertainty in CNNs. In *Bayes-SAR Net*, the convolutional kernels are considered as random fields, and their first two moments are propagated through all layers (convolution, max-pooling and fully-connected) using first-order Taylor series approximations.

The proposed *Bayes-SAR Net* performs three distinct tasks for the SAR image classification problem: 1) learning hierarchical representations of the features from the SAR images through the multistage architecture of the convolutional layers, 2) performing multi-class SAR image classification using the fully-connected layers and the final Soft-Max layer, and 3) estimating the confidence or uncertainty in SAR image classification. Our experiments using two PolSAR datasets, acquired over Flevoland in The Netherlands and Oberpfaffenhofen in Germany, showed that *Bayes-SAR Net* achieves superior robustness to Gaussian noise as well as adversarial attacks compared to its classical (deterministic) CNN homologue [9].

II. *Bayes-SAR Net*

This section introduces the general framework of *Bayes-SAR Net*, which is based on the extended VI approach [21].

A. *Network Architecture*

The architecture of *Bayes-SAR Net* is based on the classical CNN, i.e., an input layer, several convolutional layers, max-pooling, fully-connected layers, and a final classification layer (e.g., Soft-Max classifier). The input data (e.g., an image) is processed using multiple learnable convolutional kernels, that are part of the convolutional layer, to extract distinguishing features from the input data. The convolution operation is followed by a nonlinear activation function, e.g., a rectified linear unit (ReLU). The resulting features are then processed using the max-pooling operation to reduce feature dimensions. Generally, a CNN has multiple layers of the convolution operation, ReLU function, and max-pooling before performing classification. The convolutional kernels, which are responsible for extracting distinguishing features from the input data, are unknown and are estimated using a gradient descent-based algorithm by minimizing a loss function defined between the true and estimated labels. Note here that, in the classical CNN framework, the unknown parameters or weights are real-valued deterministic quantities.

B. *Variational Inference Framework*

In our setting, i.e., *Bayes-SAR Net*, the unknown weights of the network $\mathbf{\Omega} = \big\{\{\{\mathcal{W}^{(k)}\}_{k=1}^{K}\}_{c=1}^{C}, \{\mathbf{W}^{(l)}\}_{l=1}^{L}\big\}$ are defined as random variables with a prior probability distribution $p(\mathbf{\Omega})$, where $\{\{\mathcal{W}^{(k)}\}_{k=1}^{K}\}_{c=1}^{C}$ is the set of C convolutional layers, and $\{\mathbf{W}^{(l)}\}_{l=1}^{L}$ is the set of L fully-connected layers. Once the training data samples, i.e, $\mathcal{D} = \{\mathcal{X}^{(i)}, \mathbf{y}^{(i)}\}_{i=1}^{N}$ are observed, the posterior distribution of the weights given the training data $p(\mathbf{\Omega}|\mathcal{D})$ can be approximated by minimizing its Kullback-Leibler (KL) divergence with a proposed distribution $q_\theta(\mathbf{\Omega})$ that is easy to compute, i.e.,

$$\begin{aligned} \theta^* &= \operatorname{argmin} \mathrm{KL}\left[q_\theta(\mathbf{\Omega}) \| p(\mathbf{\Omega}|\mathcal{D})\right] \qquad (1) \\ &= \operatorname{argmin} \int q_\theta(\mathbf{\Omega}) \log \frac{q_\theta(\mathbf{\Omega})}{p(\mathbf{\Omega}) p(\mathcal{D}|\mathbf{\Omega})} d\mathbf{\Omega} \\ &= \operatorname{argmin} \mathrm{KL}\left[q_\theta(\mathbf{\Omega}) \| p(\mathbf{\Omega}\right] - E_{q_\theta(\mathbf{\Omega})}\left\{\log p(\mathcal{D}|\mathbf{\Omega})\right\}. \qquad (2) \end{aligned}$$

Thus, the optimal posterior approximation is obtained by maximizing the following objective function, also known as evidence lower bound (ELBO):

$$\mathfrak{L}(\theta; \mathbf{y}|\mathcal{X}) = E_{q_\theta(\mathbf{\Omega})}(\log p(\mathbf{y}|\mathcal{X}, \mathbf{\Omega})) - \mathrm{KL}(q_\theta(\mathbf{\Omega} \| p(\mathbf{\Omega})). \quad (3)$$

In the proposed framework, all convolutional kernels are assumed to be independent of each other and of the weights in the fully connected layers. This assumption is not only reasonable but also desirable as it ensures that the convolutional kernels extract “independent features” and also helps avoid redundancy in the unknown parameters. The ELBO objective function in Eq. (3) consists of the expected log-likelihood of the training data given the weights and the regularization on the weights of the convolutional and fully connected layers. We define the expected log-likelihood as a multivariate Gaussian distribution with the mean vector and variance-covariance matrix estimated by propagating the mean and covariance of the approximating distribution $q_\theta(\mathbf{\Omega})$ through the layers of the CNN.

C. *Propagating Moments*

Figure 1 shows the architecture of *Bayes-SAR Net*. The ultimate goal would be to propagate the probability distributions defined over the unknown parameters through the CNN layers and subsequently find the predictive distribution of the output, i.e., classification decision. However, propagating distributions through non-linearities is mathematically intractable. Therefore, we propose to propagate only the first two moments, i.e., means and covariances using first-order Taylor series approximations. By doing so, we are actually propagating approximate Gaussian distributions across the network layers. A detailed derivation of the moments propagation across the different layers (convolution, non-linearity, max pooling and fully connected) is provided below.

1) Convolution: Convolution is a linear operation between the convolutional kernels $\mathcal{W}^{(k)} \in \mathbb{R}^{r_1 \times r_2 \times K}$ and the input multi-channel image, such that $\mathbf{z}^{(k)} = \mathbf{X}\ \mathrm{vec}(\mathcal{W}^{(k)})$, where vec is the vectorization operation. The matrix $\mathbf{X}$ represents the input image with every row representing a vectorized sub-tensor, $\mathcal{X}_{i:i+r_1-1, j:j+r_2-1}$, selected from the input image $\mathcal{X}$ and having the same size as that of the kernels. We define multivariate Gaussian distributions over the convolutional kernels, i.e., $\mathrm{vec}(\mathcal{W}^{(k)}) \sim \mathcal{N}\left(\mathbf{m}^{(k)}, \mathbf{\Sigma}^{(k)}\right)$. It follows that, $\mathbf{z}^{(k)} \sim \mathcal{N}\left(\mathbf{X}\mathbf{m}^{(k)},\ \mathbf{X}\mathbf{\Sigma}^{(k)}\mathbf{X}^T\right)$.

2) Non-linear Activation Function: The convolution results are fed to an element-wise, nonlinear activation function φ, such that: $\mathbf{g}_i^{(k)} = \varphi\big(\mathbf{z}_i^{(k)}\big)$ (Fig. 1). We use a Taylor series expansion of φ around the mean $\mu_{\mathbf{z}_i^{(k)}}$:

$$\begin{aligned} \varphi\big(\mathbf{z}_i^{(k)}\big) =& \varphi\big(\mu_{\mathbf{z}_i^{(k)}}\big) + \big(\mathbf{z}_i^{(k)} - \mu_{\mathbf{z}_i^{(k)}}\big) \frac{d\varphi(\mu_{\mathbf{z}_i^{(k)}})}{d\mathbf{z}_i^{(k)}} \\ &+ \frac{1}{2!}(\mathbf{z}_i^{(k)} - \mu_{\mathbf{z}_i^{(k)}})^2 \frac{d^2\varphi(\mu_{\mathbf{z}_i^{(k)}})}{d(\mathbf{z}_i^{(k)})^2} + \cdots . \end{aligned} \qquad (4)$$

By computing the expected value and the variance of both sides of Eq. (4), we can approximate the mean and covariance

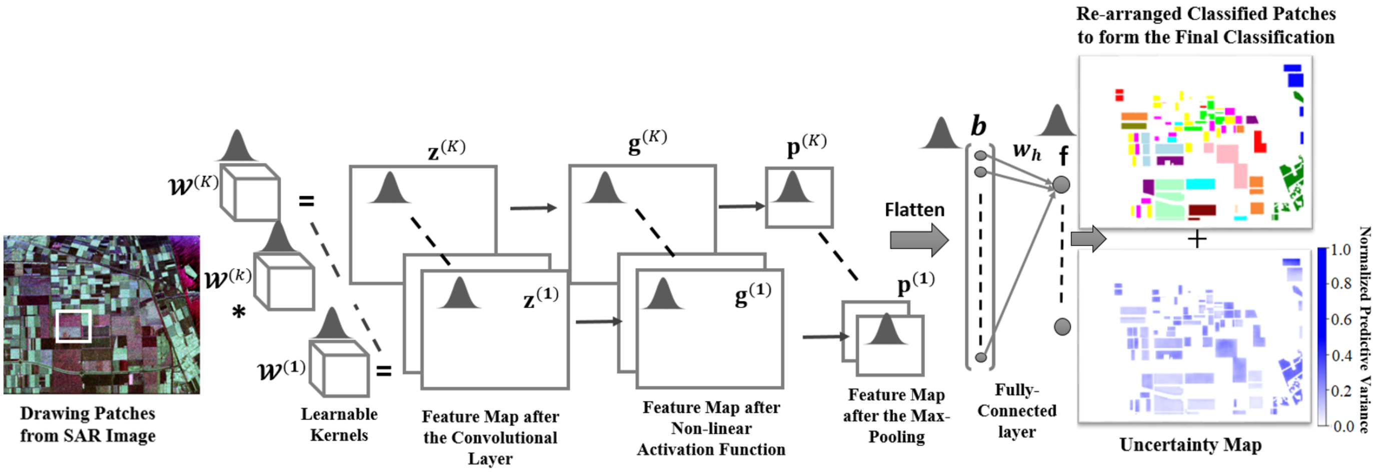

Fig. 1. The architecture of *Bayes-SAR Net* depicting convolution kernels, activation function, max-pooling operation, and fully connected layers. We define probability distributions over all unknown parameters and propagate the first two moments through these layers. The inputs of the network are patches from a SAR image and the output encompasses: i) the classification decision and ii) the associated uncertainty map generated using the predictive covariance matrix.

of the features after the activation function, $\boldsymbol{\mu}_{\mathbf{g}^{(k)}}$ and $\boldsymbol{\Sigma}_{\mathbf{g}^{(k)}}$, as follows:

$$\mathrm{E}\big(\mathbf{g}_i^{(k)}\big) \approx \varphi\Big(\mathrm{E}\big(\mathbf{z}_i^{(k)}\big)\Big) \tag{5}$$

$$\mathrm{Var}\big(\mathbf{g}_i^{(k)}\big) \approx \sigma^2_{\mathbf{z}_i^{(k)}} \left(\frac{d\varphi(\mu_{\mathbf{z}_i^{(k)}})}{d\mathbf{z}_i^{(k)}}\right)^2 \tag{6}$$

$$\mathrm{Cov}\big(\mathbf{g}_i^{(k)}, \mathbf{g}_j^{(k)}\big) \approx \sigma_{\mathbf{z}_i^{(k)}\mathbf{z}_j^{(k)}} \left(\frac{d\varphi(\mu_{\mathbf{z}_i^{(k)}})}{d\mathbf{z}_i^{(k)}}\right)\left(\frac{d\varphi(\mu_{\mathbf{z}_j^{(k)}})}{d\mathbf{z}_j^{(k)}}\right), \tag{7}$$

where $i \neq j$.

3) Max-Pooling: We approximate the mean and covariance at the output of the max-pooling layer by applying the max-pooling operation on $\boldsymbol{\mu}_{\mathbf{g}^{(k)}}$, i.e., $\boldsymbol{\mu}_{\mathbf{p}^{(k)}} = pool\big(\boldsymbol{\mu}_{\mathbf{g}^{(k)}}\big)$, and downsampling the covariance $\boldsymbol{\Sigma}_{\mathbf{g}^{(k)}}$ by keeping only the rows and columns that correspond to the pooled means.

4) Fully-Connected Layer: The vectorized feature map at the output of the max-pooling layer forms an input vector $\mathbf{b}$ to the fully-connected layer, such that $\mathbf{b} = \big[\mathbf{p}^{(1)T}, \cdots, \mathbf{p}^{(K)T}\big]^T$. Hence, $\mathbf{b}$ has the mean and covariance matrix,

$$\boldsymbol{\mu}_{\mathbf{b}} = \begin{bmatrix} \boldsymbol{\mu}_{\mathbf{p}^{(1)}} \\ \vdots \\ \boldsymbol{\mu}_{\mathbf{p}^{(K)}} \end{bmatrix}, \boldsymbol{\Sigma}_{\mathbf{b}} = \begin{bmatrix} \boldsymbol{\Sigma}_{\mathbf{p}^{(1)}} & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \boldsymbol{\Sigma}_{\mathbf{p}^{(K)}} \end{bmatrix}. \tag{8}$$

We define multivariate Gaussian distributions over the weight vectors of the fully-connected layer, i.e., $\mathbf{w}_h \sim \mathcal{N}(\mathbf{m}_h, \boldsymbol{\Sigma}_h)$, where $h = 1, \cdots, H$, and H is the number of output neurons. The output of the fully connected layer, i.e., $\mathbf{f}$, is the result of multiplying two independent random vectors $\mathbf{b}$ and $\mathbf{w}_h$. Thus, the elements of $\boldsymbol{\mu}_{\mathbf{f}}$ and $\boldsymbol{\Sigma}_{\mathbf{f}}$ are derived as:

$$\mathrm{E}(f_h) = \mathbf{m}_h^T \boldsymbol{\mu}_{\mathbf{b}}, \tag{9}$$

$$\mathrm{Var}(f_h) = \mathrm{tr}\left(\boldsymbol{\Sigma}_h \boldsymbol{\Sigma}_{\mathbf{b}}\right) + \mathbf{m}_h^T \boldsymbol{\Sigma}_{\mathbf{b}} \mathbf{m}_h + \boldsymbol{\mu}_{\mathbf{b}}^T \boldsymbol{\Sigma}_h \boldsymbol{\mu}_{\mathbf{b}}, \tag{10}$$

$$\mathrm{Cov}(f_{h_i}, f_{h_j}) = \mathbf{m}_{h_i}^T \boldsymbol{\Sigma}_{\mathbf{b}} \mathbf{m}_{h_j}, i \neq j. \tag{11}$$

5) Soft-max Classifier: Similar to the activation function, we use first-order Taylor series to approximate the mean and covariance matrix passing through the soft-max non-linearity.

D. Uncertainty Maps of Bayes-SAR Net

After each forward pass, we find the mean vector and covariance matrix of the output layer, which are used to compute the objective (loss) function defined in Eq. (3). During back-propagation, we compute the gradient of the objective function w.r.t the variational parameters $\theta = \left\{\left\{\mathbf{m}^{(k)}, \boldsymbol{\Sigma}^{(k)}\right\}_{k=1}^{K}, \left\{\mathbf{m}_h, \boldsymbol{\Sigma}_h\right\}_{h=1}^{H}\right\}$ and update θ using gradient descent. Once training is complete, we use the covariance matrix of the classification decision to generate the associated *uncertainty map*, which corresponds to the output variance of every classified pixel in the input image.

III. Experimental Results and Discussion

A. Implementation

We assess the performance of the proposed *Bayes-SAR Net* on two PolSAR datasets, i.e., Airborne SAR (AIRSAR) data of agricultural area over Flevoland in The Netherlands [22] and the electronically steered array radar (ESAR) data collected over Oberpfaffenhofen, Germany [22]. Following the work in [9], we adopt a region-based classification approach, where we randomly select patches of size $m \times m$ from SAR images and use these patches as inputs to *Bayes-SAR Net*. We start by randomly selecting 3×3 sub-patches from the SAR image that share the same ground truth label. Then, we use these 3×3 sub-patches as the center of the $m \times m$ input patches. The label of every input patch is manually set as the label of the center sub-patch [9]. The sampled patches are balanced over all classes by using the ground truth information during sampling. We conducted a sensitivity analysis to establish the optimal size of the input patches for Flevoland dataset. For both PolSAR datasets, we use the following *Bayes-SAR Net* architecture for training and testing: two convolutional layers each followed by a ReLU activation function and a max-pooling layer, and one fully connected layer. The pooling has a size of 2×2 and a stride of 2 pixels. The first convolutional layer has 64 kernels (filters) with size $3 \times 3 \times 6$, and the second layer has 128 kernels with size $2 \times 2 \times 64$. The soft-max non-linearity is used as the final output layer. Our metric for comparing the performance of *Bayes-SAR Net* to the deterministic CNN, referred to as *SAR Net*, is the overall classification accuracy which is the accuracy on the entire SAR image [9]. We compared the performance of both networks at various levels of additive Gaussian noise and adversarial attacks. The adversarial attacks were generated using the fast gradient sign method (FGSM) [23], and the level of noise was measured by the highest conceivable value (HCV), which is equal to 3 standard deviation [24].

B. Experiment on Flevoland Dataset

Flevoland dataset is a subset of an L-band, full PolSAR image, acquired by the NASA/Jet Propulsion Laboratory AIRSAR platform in 1989 during MAESTRO-1 Campaign [22]. The image size is 750×1024 pixels with 6 channels $(T_{11}, T_{12}, T_{13}, T_{22}, T_{23}, T_{33})$ [9]. There are in total 15 identified classes including stembeans, peas, forest, lucerne, three types wheat, beet, potatoes, bare soil, grass, rapeseed, barley, water, and a small number of buildings. The ground truth class color codes are presented in Fig. 2(f). The sampling rate is set to 22% which provides $30,000$ samples, 90% for training and 10% validation, while the test accuracy is computed for the entire SAR image (i.e., $156,741$ patches). Figure 2 shows (a) the Flevoland SAR image, (b) the ground truth, (c) and (d) the classification results of the deterministic *SAR Net* and the proposed *Bayes-SAR Net*, respectively and (e) the uncertainty map produced by *Bayes-SAR Net*.

Table I presents the overall accuracy of *Bayes-SAR Net* and its deterministic homologue *SAR Net* on Flevoland SAR dataset before and after adding three levels of adversarial noise. The adversarial attacks target the class label "lucerne"; thus trying to fool the network into classifying all patches as "lucerne". The table shows the results for three different patch sizes, 8×8, 16×16 and 32×32. We note that the two networks perform well on noise-free SAR data; however, *Bayes-SAR Net* maintains significantly higher accuracy under adversarial attacks compared to *SAR Net*. Note that the network produces similar accuracy values for the three different patch sizes.

Figure 3 shows the classification results and uncertainty maps of *Bayes-SAR Net* for three levels of adversarial noise, where Fig. 3(a) is the ground truth and Fig. 3(b-d) are the classification results and uncertainty maps for HCV $= 0.1, 0.2$ and 0.3, respectively (patch size = 8×8). Observe that the uncertainty in the classification results increases as the level of noise increases. The class label "lucerne", which is the target of the attack, is represented in cyan color in the ground truth image. The arrows, in Fig. 3, refer to the pixels that are misclassified as "lucerne" and to the uncertainty associated with those pixels in the uncertainty map.

TABLE I
THE OVERALL ACCURACY OF THE PROPOSED *Bayes-SAR Net* AND DETERMINISTIC *SAR Net* ON THE FLEVOLAND SAR DATASET WITH AND WITHOUT ADDING THREE LEVELS OF ADVERSARIAL NOISE.

Adversarial Noise	*Bayes-SAR Net*			SAR Net		
	Patch Size			Patch Size		
	8×8	16×16	32×32	8×8	16×16	32×32
0.1	**83.6%**	**83.0%**	**83.3%**	73.6%	73.6%	73.9%
0.2	**68.8%**	**67.8%**	**67.7%**	56.8%	58.2%	57.8%
0.3	**56.1%**	**55.1%**	**56.7%**	47.4%	48.7%	46.9%
Zero noise	96.5%	98.5%	97.6%	96.2%	98.9%	98.8%

TABLE II
CLASSIFICATION ACCURACY OF THE PROPOSED *Bayes-SAR Net* AND *SAR Net* FOR FOUR DIFFERENT LEVELS OF GAUSSIAN NOISE ADDED TO THE FLEVOLAND SAR IMAGE.

Gaussian Noise	0.01	0.1	0.2	0.3
Bayes-SAR Net	98.1%	90.8%	83.9%	77.8%
SAR Net	95.3%	88.1%	77.9%	69.9%

Figure 4 shows the output variance, averaged over all pixels of the SAR image, versus the noise level measured by the HCV. The three curves represent the output variance (representing the uncertainty in the classification decision) for three different patch sizes. The output variance increases when the noise level increases, indicating that the network is less and less confident in its decision. This monotonic behavior is observed for all patch sizes. Interestingly, the output variance corresponding to a patch size of 16×16 has higher values than the other two. For small level attacks, all 3 patch sizes correspond to similar uncertainty behavior. For high level attacks, a patch size of 16×16 seems to be quite sensitive to attacks; and thus may be desirable at deployment. More simulations on additional SAR images need to be performed to understand the exact effect of the patch size on the output variance.

Table II displays the classification accuracy of *Bayes-SAR Net* and *SAR Net* for four different levels (variance values) of Gaussian noise added to the SAR image. *Bayes-SAR Net* is clearly more robust to additive Gaussian noise. This robustness to both Gaussian noise and adversarial attacks can be intuitively linked to the uncertainty information estimated by *Bayes-SAR Net*. In particular, the variance associated with every convolutional kernel may quantify the confidence in the features learned by that kernel. Since these confidence values are propagated across the network layers, *Bayes-SAR Net* could be weighing the features according to their confidence, and thus able to better resist noise and attacks.

C. Experiment on Oberpfaffenhofen Dataset

Figure 5(a) shows the ESAR L-band, multi-look data over Oberpfaffenhofen, Germany [22]. The size of the image is 1300×1200 pixels with 6 channels. The ground truth in Fig. 5(b) shows three distinct classes: built-up areas (red), wood land (green), and open areas (yellow). Figs. 5(c) and 5(d) show classification results of *SAR NeT* and *Bayes-SAR Net*, respectively. The uncertainty map of *Bayes-SAR Net* is shown in Fig. 5(e). The sampling rate is set to 10% for Oberpfaffenhofen dataset which provides $100,000$ samples, 95% for training and 5% for validation. The test accuracy is computed for the entire SAR image (i.e., $1,303,960$ patches).

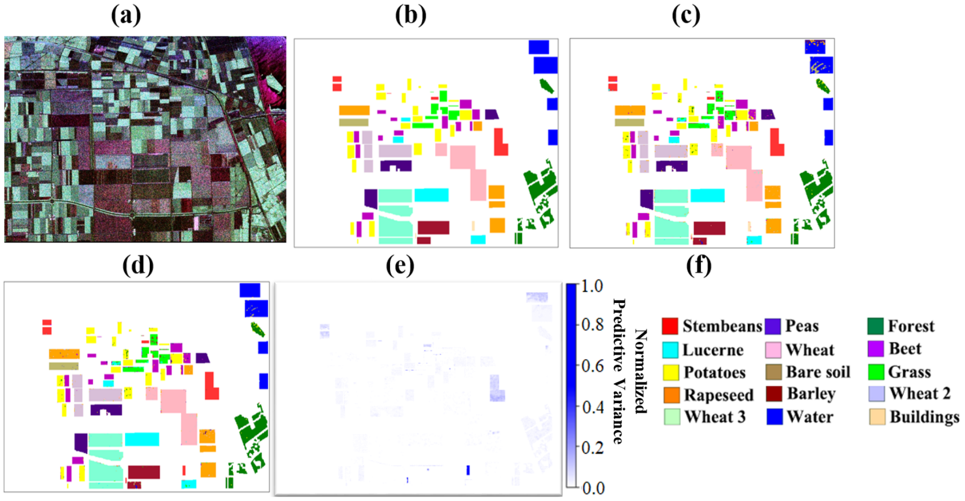

Fig. 2. Flevoland dataset. (a) The original RGB pseudo-color PolSAR image of Flevoland. (b) The ground truth map of the target scene. (c) and (d) The classification results of deterministic SAR Net and the proposed *Bayes-SAR Net*, respectively. (e) The uncertainty map produced by *Bayes-SAR Net*. (f) The legend for the ground truth.

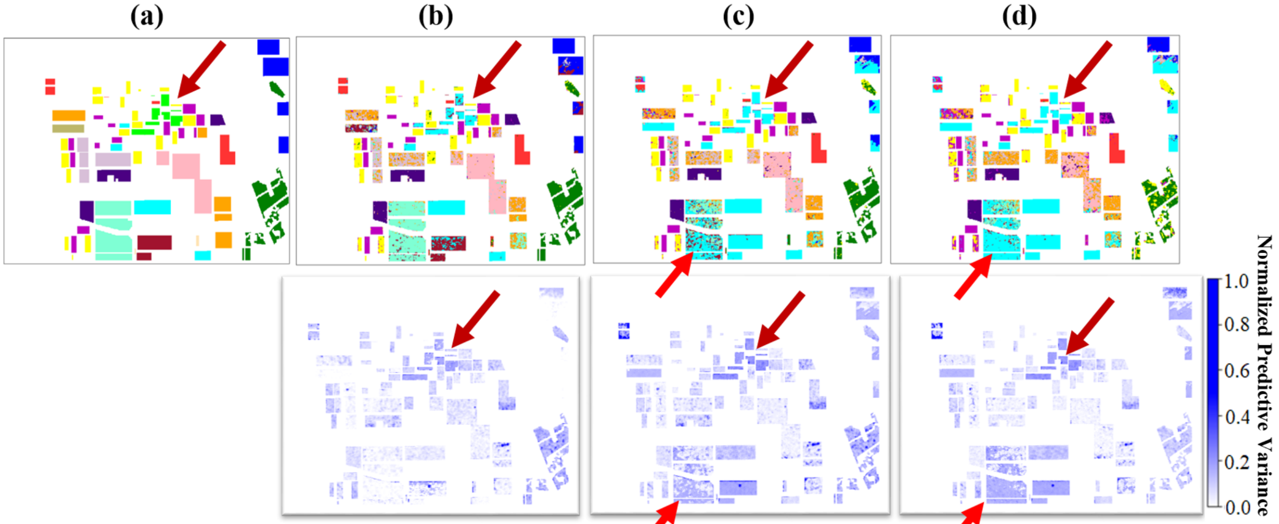

Fig. 3. The classification results and uncertainty maps of *Bayes-SAR Net* for three levels of adversarial noise corrupting the Flevoland SAR dataset. (a) The ground truth image, (b-d) the classification results and uncertainty maps for attack level $= 0.1$, 0.2, and 0.3, respectively. The patch size is set to 8×8.

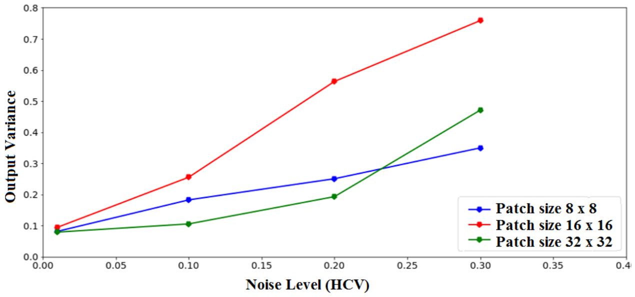

Fig. 4. The output variance averaged over all Flevoland SAR image plotted against the noise level measured by the HCV for three different patch sizes, $8 \times 8, 16 \times 16$ and 32×32.

Similar to the Flevoland dataset, we evaluate the performance of the proposed method on Oberpfaffenhofen dataset for three levels of adversarial noise, i.e., HCV $= 0.01, 0.05$ and 0.1, targeting the class label “open areas”. Table III illustrates the overall accuracy of *Bayes-SAR Net* and *SAR Net* on the Oberpfaffenhofen dataset before and after adding the three levels of adversarial noise. The patch size is set to 8×8. For higher levels of attacks, i.e., HCV = 0.05 and 0.1, *Bayes-SAR Net* achieves higher accuracy compared to its deterministic homologue. When the attack level is very low, both networks enjoy similar high accuracy.

IV. Conclusion

We introduced *Bayes-SAR Net*, a novel Bayesian convolutional neural network for synthetic aperture radar (SAR) image classification with uncertainty estimation. The proposed *Bayes-SAR Net* estimates the uncertainty in the classification decisions by propagating the first two moments of the approximating

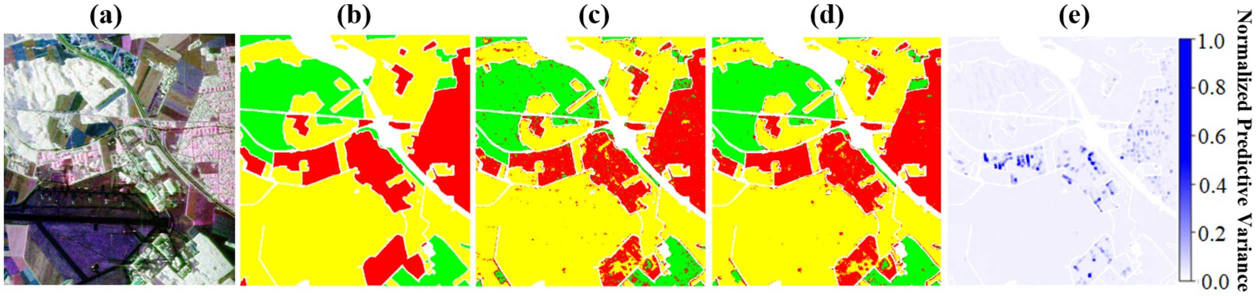

Fig. 5. Oberpfaffenhofen dataset. (a) The original RGB PolSAR image. (b) The ground truth with three distinct classes: built-up areas (red), wood land (green), and open areas (yellow). (c and d) The classification results of the deterministic *SAR Net* and *Bayes-SAR Net*, respectively. (e) *Bayes-SAR Net's* uncertainty map.

TABLE III
THE OVERALL ACCURACY OF THE PROPOSED *Bayes-SAR Net* AND *SAR Net* ON THE OBERPFAFFENHOFEN DATASET BEFORE AND AFTER ADDING FGSM ADVERSARIAL NOISE, AT LEVELS $0.01, 0.05$ AND 0.1. THE PATCH SIZE IS SET TO 8×8.

Adversarial Noise	Zero noise	0.01	0.05	0.1
Bayes-SAR Net	94.2%	89.7%	74.1%	67.7%
SAR Net	94.5%	89.5%	64.2%	59.3%

posterior distribution of the parameters given the training data. *Bayes-SAR Net* outputs: 1) the classification decision, 2) an uncertainty (or confidence) map associated with the classification. In addition, we showed that *Bayes-SAR Net* is robust to additive Gaussian noise as well as FGSM adversarial attacks as compared to its deterministic homologue, *SAR Net*, on two benchmark PolSAR datasets.

V. Acknowledgment

This work was supported by the National Science Foundation Awards NSF ECCS-1903466 and NSF CCF-1527822, as well as Lockheed Martin Inc. We are also grateful to UK EPSRC support through EP/T013265/1 project NSF-EPSRC: ShiRAS. Towards Safe and Reliable Autonomy in Sensor Driven Systems.

References

[1] A. Moreira, P. Prats-Iraola, M. Younis, G. Krieger, I. Hajnsek, and K. P. Papathanassiou, "A tutorial on synthetic aperture radar," *IEEE Geoscience and remote sensing magazine*, vol. 1, no. 1, pp. 6–43, 2013.

[2] C. Colesanti, A. Ferretti, F. Novali, C. Prati, and F. Rocca, "SAR monitoring of progressive and seasonal ground deformation using the permanent scatterers technique," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 41, no. 7, pp. 1685–1701, 2003.

[3] H.-D. Guo, L. Zhang, and L.-W. Zhu, "Earth observation big data for climate change research," *Advances in Climate Change Research*, vol. 6, no. 2, pp. 108–117, 2015.

[4] A. K. Gabriel, R. M. Goldstein, and H. A. Zebker, "Mapping small elevation changes over large areas: differential radar interferometry," *Journal of Geophysical Research: Solid Earth*, vol. 94, no. B7, pp. 9183–9191, 1989.

[5] G. Fornaro, D. Reale, and F. Serafino, "Four-dimensional SAR imaging for height estimation and monitoring of single and double scatterers," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 1, pp. 224–237, 2008.

[6] R. Touzi, W. Boerner, J. Lee, and E. Lueneburg, "A review of polarimetry in the context of synthetic aperture radar: Concepts and information extraction," *Canadian Journal of Remote Sensing*, vol. 30, no. 3, pp. 380–407, 2004.

[7] B. Hou, H. Kou, and L. Jiao, "Classification of polarimetric SAR images using multilayer autoencoders and superpixels," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 9, no. 7, pp. 3072–3081, 2016.

[8] F. Liu, L. Jiao, B. Hou, and S. Yang, "POL-SAR image classification based on wishart dbn and local spatial information," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 6, pp. 3292–3308, 2016.

[9] Y. Zhou, H. Wang, F. Xu, and Y.-Q. Jin, "Polarimetric SAR image classification using deep convolutional neural networks," *IEEE Geoscience and Remote Sensing Letters*, vol. 13, no. 12, pp. 1935–1939, 2016.

[10] J.-S. Lee, M. R. Grunes, and S. A. Mango, "Speckle reduction in multipolarization, multifrequency SAR imagery," *IEEE Transactions on Geoscience and remote sensing*, vol. 29, no. 4, pp. 535–544, 1991.

[11] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[12] J.-S. Lee, M. Grunes, E. Pottier, and L. Ferro-Famil, "Automated terrain classification using polarimetric synthetic aperture radar," NAVAL RESEARCH LAB WASHINGTON DC REMOTE SENSING DIV, Tech. Rep., 2005.

[13] C. Lardeux, P.-L. Frison, C. Tison, J.-C. Souyris, B. Stoll, B. Fruneau, and J.-P. Rudant, "Support vector machine for multi-frequency SAR polarimetric data classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 12, pp. 4143–4152, 2009.

[14] J.-S. Lee, M. R. Grunes, T. L. Ainsworth, L.-J. Du, D. L. Schuler, and S. R. Cloude, "Unsupervised classification using polarimetric decomposition and the complex wishart classifier," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 37, no. 5, pp. 2249–2258, 1999.

[15] L. Loosvelt, J. Peters, H. Skriver, B. De Baets, and N. E. Verhoest, "Impact of reducing polarimetric SAR input on the uncertainty of crop classifications based on the random forests algorithm," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 10, pp. 4185–4200, 2012.

[16] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[17] N. Akhtar and A. Mian, "Threat of adversarial attacks on deep learning in computer vision: A survey," *IEEE Access*, vol. 6, pp. 14 410–14 430, 2018.

[18] C. Blundell, J. Cornebise, K. Kavukcuoglu, and D. Wierstra, "Weight uncertainty in neural networks," in *Proceedings of the 32nd International Conference on International Conference on Machine Learning, (ICML)*, vol. 37, 2015, pp. 1613–1622.

[19] K. Shridhar, F. Laumann, A. Llopart Maurin, and M. Liwicki, "Bayesian convolutional neural networks," *arXiv preprint arXiv:1806.05978*, 2018.

[20] W. Roth and F. Pernkopf, "Variational inference in neural networks using an approximate closed-form objective," in *Neural Information Processing Systems, (NIPS) workshop*, 2016.

[21] D. Dera, G. Rasool, and N. Bouaynaya, "Extended variational inference for propagating uncertainty in convolutional neural networks," in *IEEE International Workshop on Machine Learning for Signal Processing*, 2019.

[22] Earth Online. PolSAR (The Polarimetric SAR Data Processing and Educational Tool). [Online]. Available: https://earth.esa.int/web/polsarpro/airborne-data-sources

[23] Y. Liu, X. Chen, C. Liu, and D. Song, "Delving into transferable adversarial examples and black-box attacks," in *Proceedings of 5th International Conference on Learning Representations*, 2017.

[24] J. M. Duncan, "Factors of safety and reliability in geotechnical engineering," *Journal of geotechnical and geoenvironmental engineering*, vol. 126, no. 4, pp. 307–316, 2000.