Abstract—In a complex-valued convolutional neural network, its elementary unit consists of a complex-valued convolution layer and a complex pooling layer. The pooling layer has a variety in its dynamics. In this paper, we propose complex absolute-value max pooling to extract complex-amplitude feature patterns meaningful for discovery and/or adaptive classification of land form in interferometric synthetic aperture radar (InSAR). Experimental examination into amplitude and phase values in convolutional kernels reveals that useful land-shape features emerge through self-organization in high-magnitude kernels, which suggests that the proposed dynamics is successful in extracting important features.

Index Terms—Complex-valued neural network, interferometric synthetic aperture radar (InSAR), earth artificial intelligence (AI)

I. INTRODUCTION

Many ideas have been proposed in these years to discover and/or classify adaptively local features in earth observation data [1]–[11]. They form the earth artificial intelligence (Earth AI) field. There, complex-valued convolutional neural networks are very useful for dealing with complex-amplitude local features consistently in interferometric synthetic aperture radar (InSAR) [12], [13]. In the near future earth AI systems, they will discover and continuously monitor specific land forms such as volcanos, floodable regions and glacier areas, which is significantly important for disaster prevention/reduction, global warming monitoring, and various other human life security. Related big-data processing frameworks have also been investigated widely and intensively [14], [15].

There are a variety of possible dynamics in a complex-valued elementary pair of a convolutional layer and pooling layer [16]–[19]. In particular, we have to design a pooling layer by considering what kind of features we need to extract, examining its available forward processing dynamics and learning algorithms.

In this paper, we propose complex absolute-value max pooling in the complex-valued convolutional neural networks to enhance the convolution signal-to-noise ratios by paying more attention to strong backscatterers and resulting convolutional outputs. Experiments demonstrate that this pooling realizes the self-organization of meaningful information in the convolutional kernels based on the assumption that high-intensity convolutional outputs originating from strong backscattering convey more information of the local target area by suppressing thermal and other noise. We find that this dynamics fits the purpose to discover and/or classify land-form features in InSAR.
II. INTERFEROMETRIC SYNTHETIC APERTURE RADAR OBSERVATION

Fig. 1 shows rough geometry of interferometric earth observation by satellite-borne synthetic aperture radar (SAR) [20]. A SAR system illuminates the earth with an electromagnetic wave having a wavelength of $\lambda$, and receives backscattered wave as the amplitude and phase, i.e., complex-amplitude. With the observation data at two satellite positions $S_m$ and $S_s$, corresponding to master and slave observations, we can obtain an interferogram to calculate the land height changes $\Delta H$ by using their phase difference $\Delta\Phi$ as

$$\Delta H = \frac{\lambda R_m \sin \theta}{4\pi B_{CT} \cos(\theta - \gamma_{CT})} \Delta\Phi$$  \hspace{1cm} (1)

Then, we can make a height map by accumulating the phase changes pixel by pixel. This accumulation process is called "unwrapping" since the phase value is "wrapped" into the range of $[-\pi, \pi]$. This calculated high-accuracy height map is called digital elevation model (DEM), which is obtained globally, spatially continuously and temporally frequently, and is very useful in human security.

Fig. 2 shows an (a) amplitude and (b) phase example of an InSAR interferogram obtained by Advanced Land Observing Satellite 2 (ALOS-2) of Japan Aerospace Exploration Agency (JAXA) for a Mt. Fuji area in Japan. The amplitude represents the magnitude of backscattering while the phase shows the height in a $2\pi$-wrapped manner as expressed by (1). In this paper, we do not unwrap the phase image. Instead, we make spatial difference images in east-west and north-south directions so that we deal with land forms as slopes rather than its absolute height.

III. PROPOSED COMPLEX-VALUED CONVOLUTIONAL NEURAL NETWORK

Fig. 3 shows the construction of the complex-valued convolutional neural network we propose for learning and processing InSAR local data [12]. Complex-valued neural networks are suitable for processing complex-amplitude data including InSAR interferograms [21]–[27]. The two-direction difference images are fed to the network as two-channel inputs ($C=2$) and processed by one or more units of convolution and pooling layers. The image sizes and parameters are summarized as follows.

- Total data image size: $0^x I \times 0^y J = 1,200 \times 1,200$ pixels
- Local window size: $28 \times 28$ pixels
- Convolution kernel number: $K=9$ for each channel, size: $1^x_{\text{Conv}} \times 1^y_{\text{Conv}} = 27 \times 27$ pixels
- Pooling window size: $1^x_{\text{Pool}} \times 1^y_{\text{Pool}} = 2 \times 2$ pixels

After a single or multiple convolution-and-pooling layer processing, we have a full connection layer to obtain total network outputs. In the present classification, we prepare the same number of output neurons as that of classes we intend.

A. Complex-valued convolution processing

In the following experiments, we feed the two complex-valued spatially-differential images mentioned above, denoted as $i=0^x x^w$ and $i=0^x x^m$ having common coordinates, to the neural network. A neural elementary unit is composed of a convolution layer and a pooling layer, and indexed by $i$ in Fig. 3(a). Input terminal layer is labeled as $i=0$, while the last layer is $i=L$, which is a fully connected network to work for decision.

We have single or multiple input signals (images) $0^x x^c$ generally, where $c$ is the image label and 0 means the input
layer. The output signal vector of the first convolution layer
\[ 1v_k = [1v_{kij}] \]
expressed for input signal vectors
\[ 0x = [0x_{i+p, j+q}] \]
as
\[ 1v = f(1u) \]  \hspace{1cm} (2)
\[ 1u_{kij} = \sum_{c=1}^{C} \sum_{p=1}^{P} \sum_{q=1}^{Q} 1w_{k,p,q} \ast 1x_{i+p, j+q} \]  \hspace{1cm} (3)
\[ f(1u_{kij}) = \tanh(1u_{kij}) \exp (j \arg(1u_{kij})) \]  \hspace{1cm} (4)
where \( 1W = w_{k,p,q} \) is the neural weight connecting an input signal at position \((p, q)\) in image \(c\) and a neuron at position \((p, q)\) in convolutional kernel \(k\) in the first layer, and \(C, P\) and \(Q\) are the numbers of the total input channels, kernel vertical size and kernel horizontal size, respectively. \(f(\cdot)\) is the activation function, and \(1u\) is internal state of \(l\)-th layer neuron.

**B. Forward processing and teacher-signal-backpropagation learning dynamics**

We determine the neural dynamics as follows. The \((l-1)\)-th layer output signals \(1x\) are processed in \(l\)-th layer to generate
respectively.

where $\epsilon$ signals, instead of errors, to update connection weights for $l$-th layer.

The teacher signal of $(l-1)$-th layer has

$$f(l^{-1}x) = \text{tanh}(l^{-1}x)$$

where $l^{-1}x$ as

$$l^{-1}x = f(l^{-1}u)$$

$$l^{-1}u = l^{-1}W l^{-1}x$$

where $l^{-1}W = [l^{-1}w_{nm}]$ is the connection weight between $m (= iJ+j)$-th neuron in $(l-1)$-th layer and $n (= pQ+q)$-th neuron in $l$-th layer.

We also determine the backpropagation dynamics of teacher signals, instead of errors, to update connection weights for supervised learning as follows. We calculate teacher signal vectors of each layers $l^{-1}t$ backwards from the output layer. The teacher signal of $(l-1)$-th layer $l^{-1}t$ is obtained from the teacher signal of $l$-th layer $l^{-1}t$ as [26]

$$l^{-1}t = (f(l^{-1}t \cdot l^{-1}W))^*$$

Then, we update $l^{-1}W$ by using $l^{-1}t$ as [28]–[30]

$$\|l^{-1}w_{nm}\|_{l^{-1}w_{nm}} = \epsilon_a \{ \epsilon_b \{ (1 - |l^{-1}x_n|) (l^{-1}x_n - l^{-1}t_n) \cos arg(l^{-1}x_n - l^{-1}t_n) \} |l^{-1}x_n| \sin arg(l^{-1}x_n - l^{-1}t_n) \}$$

$$ \text{arg}(l^{-1}w_{nm}) \leftarrow \text{arg}(l^{-1}w_{nm})$$

$$\|l^{-1}\theta_{rot}\|_{l^{-1}\theta_{rot}} = \epsilon_p \{ \epsilon_q \{ (1 - |l^{-1}x_n|) (l^{-1}x_n - l^{-1}t_n) \cos arg(l^{-1}x_n - l^{-1}t_n) \} |l^{-1}x_n| \sin arg(l^{-1}x_n - l^{-1}t_n) \}$$

$$\text{arg}(l^{-1}\theta_{rot}) \leftarrow \text{arg}(l^{-1}\theta_{rot})$$

where $\epsilon_a$ and $\epsilon_p$ are learning rates of amplitude and phase respectively.

C. Complex absolute-value max pooling

The output of the pooling layer is invariant for small position changes in the output patterns of the former layer. In other words, a pooling layer absorbs small position changes and small rotations of input images. Here, we determine the complex-valued pooling process as

$$l_{yk\cdot j} = l_{y} \arg\max_{p\cdot q} |l_{yp\cdot q}|$$

That is, the output of a pooling layer is a value that has maximum norm in the pooling window. If the local pattern in the input image is similar to the complex-amplitude pattern in a kernel of a convolutional layer, the norm of the output signal becomes large, and then the following pooling layer extracts this signal. At the same time, as shown in Fig. 3(b), this process enhances high-amplitude signals in the convolution process, resulting in extraction of phase information in high-amplitude input signals. This treatment reflects the thought that high-intensity signals are more meaningful, and should be extracted intensively than small or noise components.

Then we design the dynamics of the teacher-signal backpropagation in the pooling layer as follows. We make teacher signals backpropagate in order to update only the neurons that pass the signals to the following layer. Specifically, we get teacher signals of the convolution layer $l^{-1}t$ from teacher signals for the pooling layer $l^{-1}t$ as

$$l_{yk\cdot p\cdot q} = \begin{cases} l_{yk\cdot j} & \text{if } l_{yk\cdot j} = l_{yk\cdot p\cdot q} \\ l_{yk\cdot p\cdot q} & \text{otherwise} \end{cases} \quad (13)$$

D. Fully connected network for decision

For classification, the output layer of the complex-valued convolutional neural network is composed as a fully connected network, just like a usual complex-valued neural network. The input signal of the fully connected layer $L^{-1}x = [L^{-1}x_n]$ is obtained from the preceding pooling layer as

$$L^{-1}x_{k\times I \times J+i\times J+j} = L^{-1} y_{k\cdot i\cdot j} \quad (14)$$

where $k \times I \times J + i \times J + j (= n)$ is represented one dimensionally when the size of the output image of the pooling layer has $K$ images of $I \times J$ pixels.

IV. EXPERIMENTS AND RESULTS

Fig. 4 presents seven teacher zones from which we cut out local teacher signals for seven classes (categories), namely, north-facing slope, east-facing slope, south-facing slope, west facing slope, flat plain, sea, and Omuroyama (scoria cone). Omuroyama is a Japanese word meaning small eruptive crater. By including Omuroyama as a class, we intend to find volcanos of a similar shape for disaster prediction/mitigation. In the following experiments, we used a convolutional network having a single pair of convolution and pooling layers. We determined the teacher signals used at the full-connection network as $[1 -1 -1 -1 -1 -1 -1]^T$ for north-facing slope, $[-1 1 -1 -1 -1 -1 -1]^T$ for east-facing slope, and so on, where $[\cdots]^T$ stands for transpose.
Fig. 5 presents the classification result for the whole target region after the learning is completed. In this paper, we focus on the extraction of Omuroyama-like land form. In the figure, small red areas indicate the decision of Omuroyama shape. We can see that they are scattered widely.

Let’s investigate into Areas (i), (ii) and (iii) in Fig. 5. Fig. 6(a) shows Area (i) with red patches at the Omuroyama decisions. Besides the teacher Omuroyama, we find another Omuroyama class patch in the east-south direction, which is an eruptive crater named Katafutayama. It is smaller in its size than the teacher Omuroyama, but we can see in the photo in Fig. 6(b) on the right-hand side that the shape is homothetic. That is, the complex-valued convolutioanal network succeeded in finding a similar land shape patch.

Fig. 7(a) shows the result for Area (ii), in which we find another Omuroyama. (The mountain is named also Omuroyama, and this is a typical Omuroyama in Japan in its shape.) In the photo in Fig. 7(b), we can confirm the similar shape. Another patch shows another mountain having similar land form.

Fig. 8(a) shows Area (iii), Hakone area, a popular sightseeing area located in the Fuji-Hakone National Park. The marked two mountains are “Futagoyama,” meaning twin mountains. In Fig. 8(b), they are found surely to have also Omuroyama-like shapes. However, these mountains are much bigger than the teacher Omuroyama, in particular in its height. Since there are no high mountains around them, it is not widely known that these mountains have the Omuroyama-like shape. In this way, however, they are discovered as Omuroyama-class mountains surprisingly.
V. SELF-ORGANIZED IMAGES IN THE CONVOLUTIONAL KERNELS

Fig. 9 presents the convolution kernels before and after the learning process. They are found to self-organize to be meaningful kernels as follows. Gray-scale images show the amplitude values while color images represent phase values. The initial values before the learning is shown in Fig. 9(a). Both the amplitude and the phase are at random, presenting no specific features. In contrast, Fig. 9(b) presents the emerged kernels. Both in \( W^{c=1} \) and \( W^{c=2} \), the amplitude values are large in four kernels, i.e., \( k = 1, 2, 6 \) and 9. Other kernels are relatively small. The large ones indicate significant meanings.

The kernel \( k=2 \) of \( W^{c=1} \) in Fig. 9(b) shows large values in the amplitude, and its phase values are pink in the left half while green in the right half. This spatial pattern represents an eastward up-and-down, that is, a ridge in the north-south direction. Similarly, Kernel \( k=9 \) in \( W^{c=2} \) also shows large amplitude values, and its phase has green values in the upper half while pink values in the lower half. It means a northward up-and-down, or a ridge in the east-west direction. In addition, the combination of these two features represents a mountain shape having a peak at around the center. In this way, the kernels self-organize into meaningful spatial information representation. The correspondence to the large amplitude reveals the successful feature extraction with the complex-absolute-value max pooling in the complex domain.

VI. SUMMARY

This paper proposed the complex absolute-value max pooling to extract land form features adaptively by dealing with local complex-amplitude in interferometric synthetic aperture radar. The network discovers and classifies adaptively special land forms such as Omuroyama. We confirmed in the experiment that meaningful features emerge in the kernels in the convolutional layers by investigating the complex-amplitude spatial patterns.

REFERENCES

Fig. 9. Amplitude (gray) and phase (color) of convolutional weights for kernel number $k$ (a) before and (b) after the learning process ($W^c=1$ for east-west difference data, $W^c=2$ for north-south difference data).


