SCIENCE CHINA
Information Sciences

• LETTER •

February 2021, Vol. 64 129101:1–129101:3
https://doi.org/10.1007/s11432-018-1513-5

RGA-CNNs: convolutional neural networks based on reduced geometric algebra

Rui WANG[1], Miaomiao SHEN[1], Xiangyang WANG[1] & Wenming CAO[2*]

[1]*School of Communication and Information Engineering, Shanghai University, Shanghai* 200444*, China;*
[2]*College of Information Engineering, Shenzhen University, Shenzhen* 518060*, China*

Received 10 December 2018/Revised 23 March 2019/Accepted 22 June 2019/Published online 16 July 2020

Citation Wang R, Shen M M, Wang X Y, et al. RGA-CNNs: convolutional neural networks based on reduced geometric algebra. Sci China Inf Sci, 2021, 64(2): 129101, https://doi.org/10.1007/s11432-018-1513-5

Dear editor,

Recently, convolutional neural networks (CNNs) have exhibited high performance particularly in object detection [1], face recognition [2], and image classification [3]. However, there has been little work on CNN models for multidimensional data, such as the three-dimensional (3D) data, which are typically presented as color images [4]. Traditional real-valued CNN models [4] have achieved state-of-the-art results for gray-scale images. However, color images are handled by simply treating them as three gray-scale images to apply real-valued CNN models independently, which potentially leads to the loss of fusion information of different color channels. Guberman [5] proposed the idea of complex-valued CNN with complex-valued inputs and parameters, showing that they can be seen as restricted forms of larger real-valued CNNs, and as such have the potential to mitigate over-fitting issues.

Kominami et al. [6] extended CNNs to the quaternionic domain, namely quaternion-valued CNNs (QCNNs) that can efficiently handle problems with multiple inputs and multiple outputs. Traditional real-valued neural networks cannot directly represent data in multidimensional spaces, particularly 3D signals such as color images. On the contrary, QCNNs can efficiently deal with color images but only by encoding the three color channels into the three imaginary parts, due to the noncommutativity of quaternionic multiplication. Unfortunately, it may inevitably produce substantial data redundancy and a more complicated network structure, and is not conducive to the design of fast algorithms.

Existing CNN models represent color image pixels either as scalar [4], resulting in the loss of some inherent color structure, or as a quaternion vector matrix [6], significantly increasing the computational complexity. Herein, we present a novel type of CNN based on reduced geometric algebra (RGA), which we call an RGA-CNN. This represents both the input color image pixels and convolution kernels as RGA multivectors, which enables it to capture the inherent color structures, eliminates data redundancy, and simplifies the network.

The theory of RGA with commutative multiplication properties is introduce in [7]. Our model utilizes RGA theory to represent the input image, neurons, convolution kernels, learning algorithm, and all related computations within the RGA framework, which is used to define the basic operational rules, properties, and convolution operations. In this way, we can fully preserve joint channel information and consequently reduce the network complexity. Finally, we conduct experiments to evaluate the effectiveness and feasibility of proposed model, showing that it can preserve the inherent color structures and achieve higher learning performance at a lower training cost.

RGA-CNNs. We propose a novel type of CNN based on RGA, which we call an RGA-CNN. Its basic structure is presented in Appendix B, and its operation process can be summarized as follows.

The input matrix and convolution kernel are represented as RGA multivectors $A \in (\mathbb{G}_2^R)^{M\times M}$ and $K \in (\mathbb{G}_2^R)^{N\times N}$, respectively. In each convolution layer (shown in Appendix B), the output $B \in (\mathbb{G}_2^R)^{(M-N+1)\times(M-N+1)}$ is given by

$$b = f\left(\sum \frac{ka}{|k|} + \varphi\right). \tag{1}$$

The notation used in (1) is defined in Appendix B.

Then, we extend the conventional max-pooling method from the real domain to the RGA domain, as follows:

$$c_{ij} = \max_{(k,l)\in C_{ij}} b_{kl}. \tag{2}$$

Again, the notation used in (2) is defined in Appendix B.

Generally, after a series of convolution and max-pooling layers, the final outputs are used as input to an RGA-based multilayer perception (RGA-MLP). RGA-MLPs and their associated learning algorithm are defined in Appendix B. The learning rate and training loss play important roles in the training of CNN models. Choosing a suitable learning rate enables the objective function to converge to the local minimum within a reasonable number of iterations. If the

* Corresponding author (email: wmcao@szu.edu.cn)

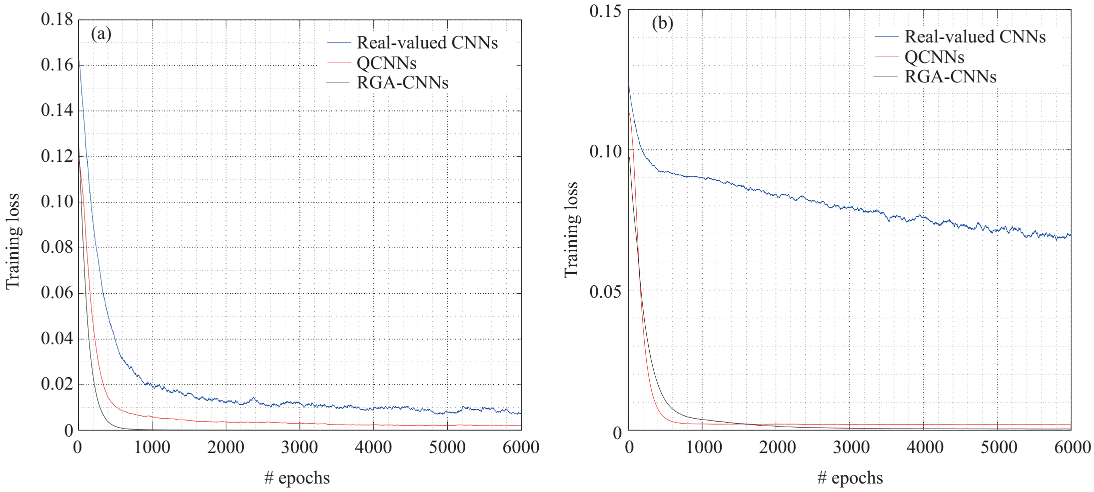

Figure 1 (Color online) Training loss curves achieved by the real-valued CNN, QCNN, and proposed RGA-CNN for (a) the 3D geometrical shape and (b) the color image classification tasks.

learning rate is too low, convergence will be very slow. By contrast, if the learning rate is too large, the loss may oscillate around the minimum value, or even fail to converge. Using a fixed learning rate means the loss will likely oscillate within a larger range near the optimal value, while decreasing the learning rate during later epochs means it will converge to oscillate within a smaller area near the minimum value. This shows that choosing an appropriate learning rate is very important for training CNN models.

The loss function is typically used to estimate the parameters, and is some function of the difference between the estimated and true values for the training examples. During training, it specifies how differences between the predicted (output) and true labels are penalized, and is normally based on the neural network's final layer.

Results and discussion. Now, we perform classification experiments with 3D geometrical shape and color image datasets to compare the practical performance and computational efficiency of our proposed RGA-CNN model with those of existing real-valued CNN [4] and QCNN [6] models, both quantitatively and visually. Figure 1(a) and (b) show the training loss curves achieved by the real-valued CNN, QCNN, and proposed RGA-CNN for the 3D geometrical shape and color image classification tasks, respectively. Here, we can clearly see that our proposed RGA-CNN achieve significantly higher performance, namely lower training losses and higher test-set accuracies, than the traditional real-valued network. This is due to the proposed RGA-CNN's ability to learn and explore the inherent color structures of the color images, which play an important role in such classification tasks. Appendix C presents additional simulation results.

Currently, CNNs have been developed that are based on real-valued, complex-valued, and quaternion-valued domains. Many improved schemes have been developed for real-valued CNNs, enabling them to achieve state-of-art performance. For example, Graham [8] proposed an effective CNN structure and formulated a fractional max-pooling approach where the multiplicative factor α is allowed to take on non-integer values, demonstrating that this could reduce overfitting on a variety of datasets, such as CIFAR-100. Springenberg et al. [9] also presented an improved CNN architecture, consisting solely of convolutional layers, that yielded competitive or state-of-the-art performance on several object recognition datasets (CIFAR-10, CIFAR-100, and ImageNet).

By contrast, we focus on utilizing RGA theory to extend existing real-valued CNNs or quaternion-valued CNNs (QCNNs) to the RGA domain, in order to develop what we call RGA-CNNs. Our aim is to construct CNNs that can effectively handle multidimensional signals with higher performance and lower data redundancy and computation complexity than real-valued CNNs or QCNNs. To demonstrate the superiority of the proposed RGA-CNNs over real-valued CNNs and QCNNs, we conduct classification experiments by constructing CNNs of each of these three types with typical structures.

In future work, we will explore potential extensions to our proposed RGA-CNNs, aiming to improve their architectures and achieve greater performance. Inspired by the work in [8, 9], we also plan to further improve our algorithm. This approach holds great promise for improving multi-dimensional signal processing.

Conclusion. In this study, we have proposed a new type of convolutional neural network (CNN) based on RGA, which we call an RGA-CNN. This successfully extends traditional real-valued CNNs into RGA spaces. RGA-CNNs can process multidimensional data in a holistic manner, retaining the relationships between different dimensions. To create RGA-based networks, we have redefined the convolution, pooling, hidden, and output layers based on RGA theory. To demonstrate the effectiveness of the proposed RGA-CNNs, we have also performed classification experiments using 3D geometric shape and color image datasets, comparing the RGA-CNN's classification results with those of a real-valued CNNs and a QCNN. This results indeed show that the proposed RGA-CNN yields the best performance, with lower training losses, higher test-set accuracies, and faster convergence than the real-valued CNN and faster computation than the QCNN.

Acknowledgements This work was supported by National Natural Science Foundation of China (Grant Nos. 61771299, 61771322, 61375015, 61301027)

Supporting information Appendixes A–C. The supporting information is available online at info.scichina.com and link.

springer.com. The supporting materials are published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.

References

1 Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014. 580–587

2 Taigman Y, Yang M, Ranzato M, et al. DeepFace: closing the gap to human-level performance in face verification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014. 1701–1708

3 Krizhevsky A, Sutskever L, Hinton G E. Imagenet classification with deep convolutional neural networks. In: Proceedings of Advances in Neural Information Processing Systems, 2012. 1097–1105

4 Zhang F, Cai N, Wu J X, et al. Image denoising method based on a deep convolution neural network. IET Image Process, 2018, 12: 485–493

5 Guberman N. On complex valued convolutional neural networks. 2016. ArXiv: 1602.09046

6 Kominami Y, Ogawa H, Murase K. Convolutional neural networks with multi-valued neurons. In: Proceedings of International Joint Conference on Neural Networks, Anchorage, 2017. 2673–2678

7 Shen M M, Wang R, Cao W M. Joint sparse representation model for multi-channel image based on reduced geometric algebra. IEEE Access, 2018, 6: 24213–24223

8 Graham B. Fractional max-pooling. 2015. ArXiv: 1412.6071

9 Springenberg J T, Dosovitskiy A, Brox T, et al. Striving for simplicity: the all convolutional net. In: Proceedings of International Conference on Learning Representations (ICLR), 2015. 1–14