Split-Complex Convolutional Neural Networks

Stanford ENGINEERING
Electrical Engineering

Timothy Anderson
timothy.anderson@stanford.edu
Department of Electrical Engineering, Institute for Computational and Mathematical Engineering

ICME

Introduction

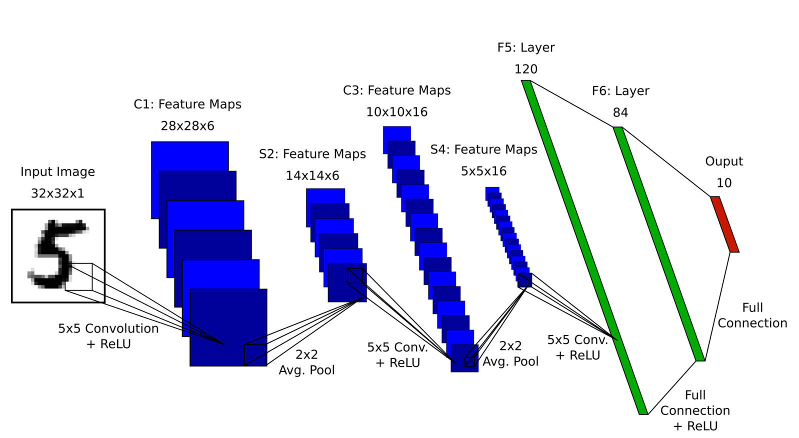
- Clifford algebras have long history in neural networks [4], but have only recently received renewed attention [1]
- Motivation: complex numbers have rotational structure, so complex-valued convolutional neural networks potentially have rotational invariance [1]
- Most recent work has focused on complex-valued networks [2, 5]
  - Virtually no work on split-complex numbers in neural networks

Mathematical Framework

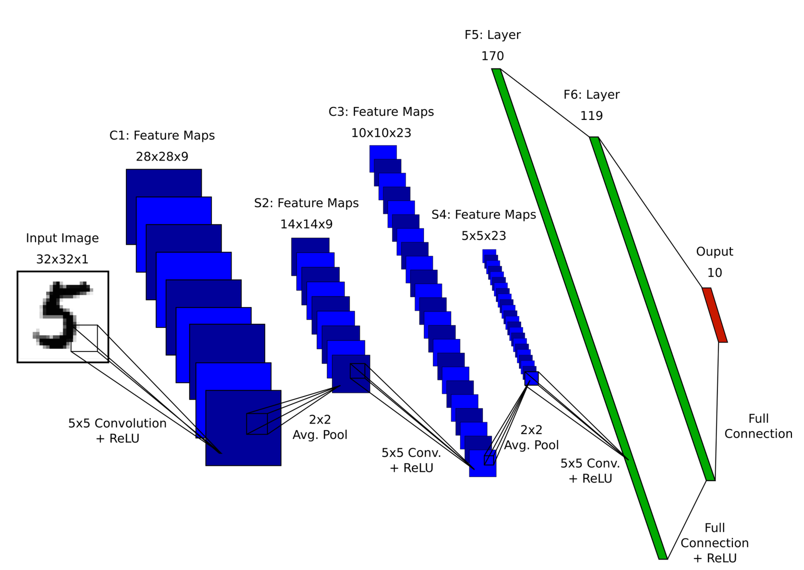
- Complex number $\mathbb{C}$ member of two-dimensional algebras
  - $x \in \mathbb{C}$ has form $x = a + bi$ with $a, b \in \mathbb{R}$ and $i^2 = -1$
- Split-complex numbers $\mathbb{S}$ have same form, formed by imposing $i^2 = +1$

Setup

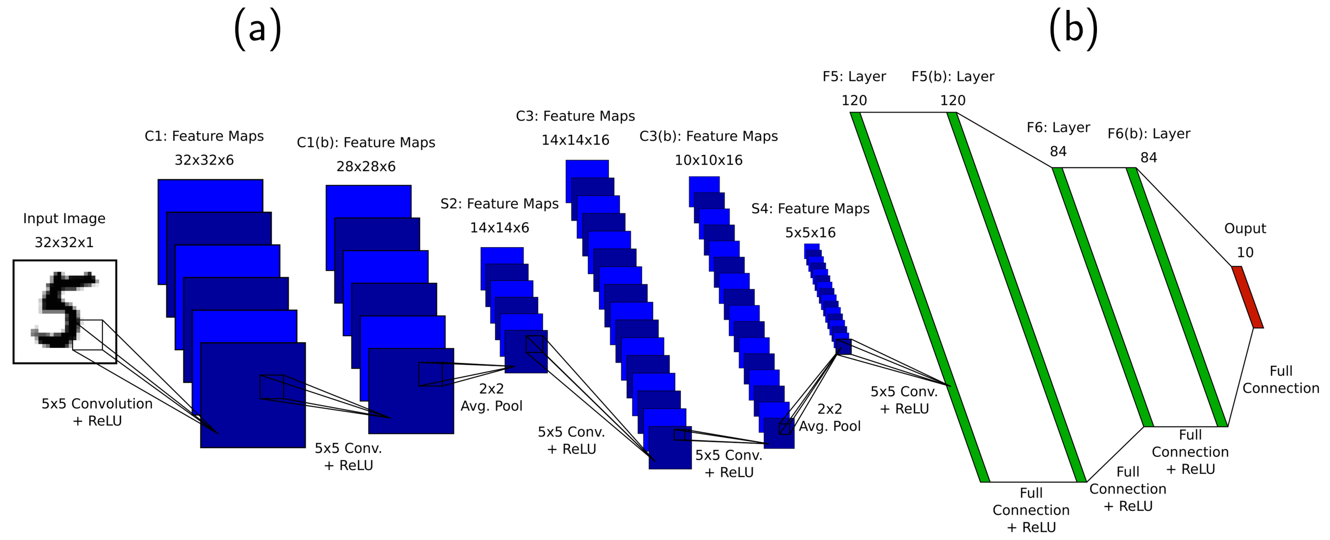
- Followed approach of [5] to compare with complex network
- Network architectures based on LeNet-5 [3]
- Tested wide and deep architectures to compare complex-valued networks with doubling number of real-valued parameters
- Compared regularized and unregularized models



(a)



(b)



(c)

(a) Baseline LeNet-5 architecture, (b) "Wide" network architecture. The number of filters or neurons at each layer increased by $\sim \sqrt{2}$ so number of parameters is approximately doubled. (c) "Deep" network architecture. Each layer from the baseline architecture is repeated to double number of parameters.

Implementation Overview

- Complex and split-complex numbers are commutative fields
- Represent real and imaginary components as separate parameters and implement complex arithmetic via parameter sharing in computational graph
  - *Ex:* Split-complex valued convolution:

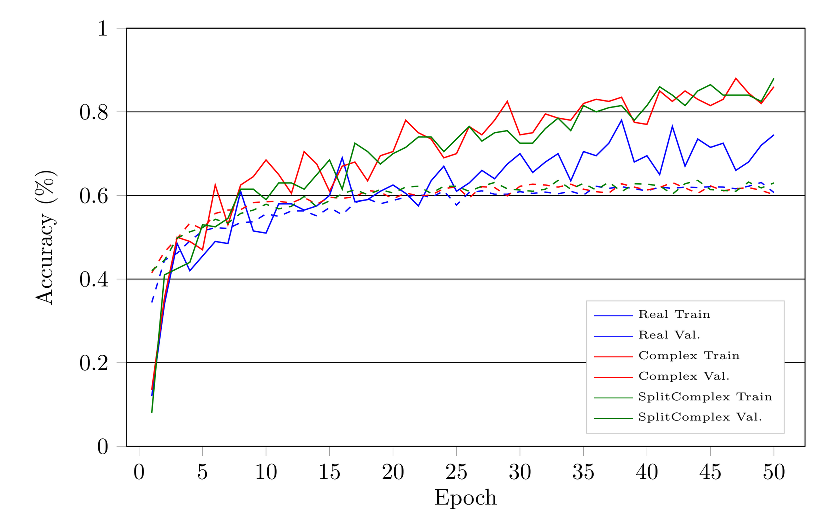
$$\mathbf{X}_i = \mathbf{X}_{i,R} + \mathbf{X}_{i,I} i, \quad \mathbf{W} = \mathbf{W}_R + \mathbf{W}_I i$$
$$\mathbf{W} * \mathbf{X} = \mathbf{W}_R * \mathbf{X}_{i,R} + \mathbf{W}_I * \mathbf{X}_{i,I} + (\mathbf{W}_R * \mathbf{X}_{i,I} + \mathbf{W}_I * \mathbf{X}_{i,R}) i$$

- Generalized ReLU activation function:

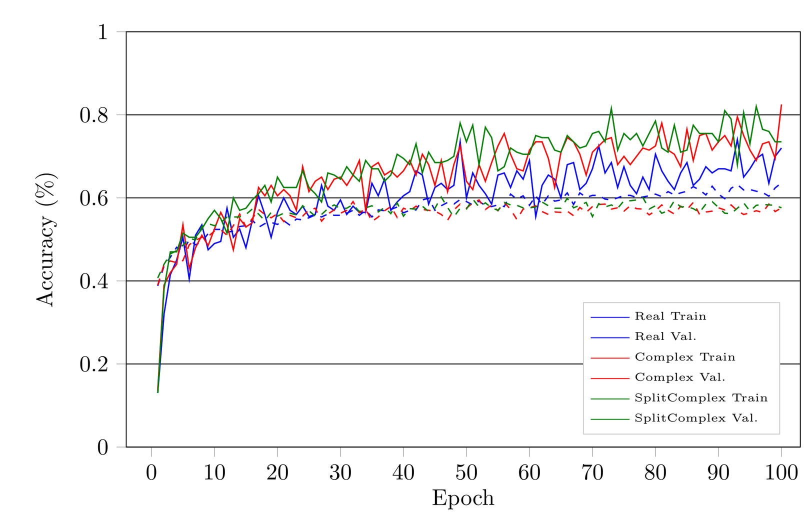
$$ReLU(x) = \begin{cases} x & \Re(x) \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

Training Curves

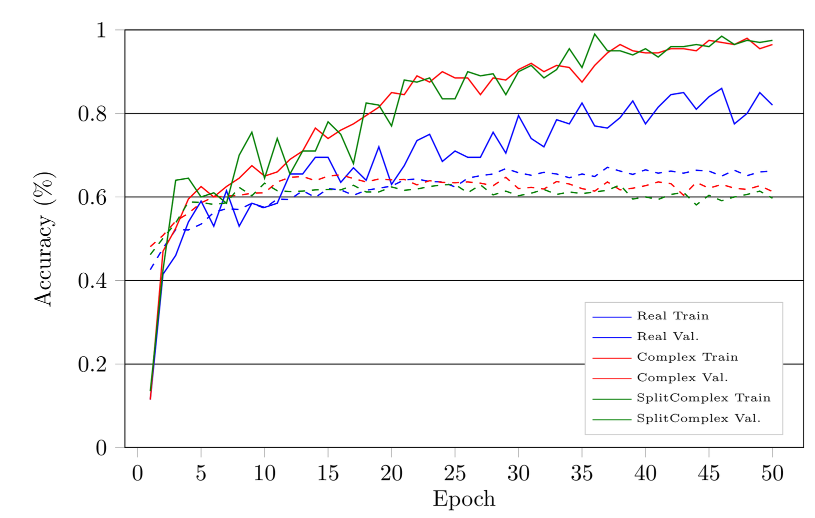
Training curves for CIFAR-10.



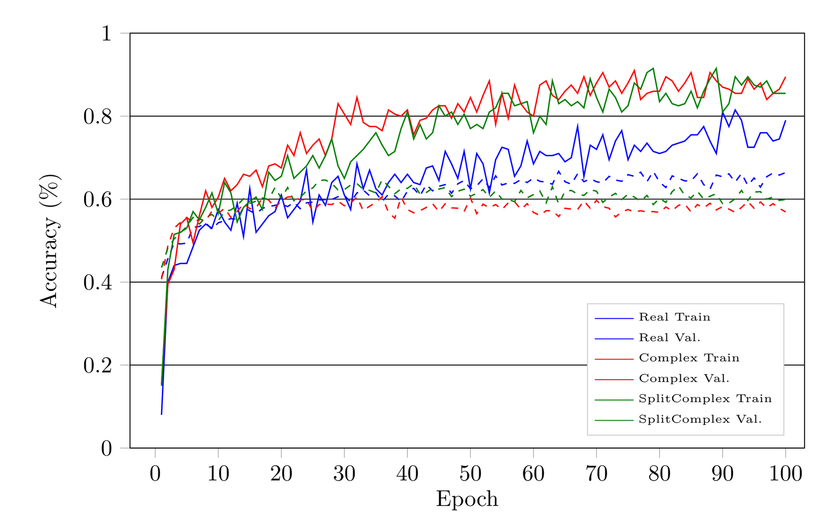
(a) Baseline model



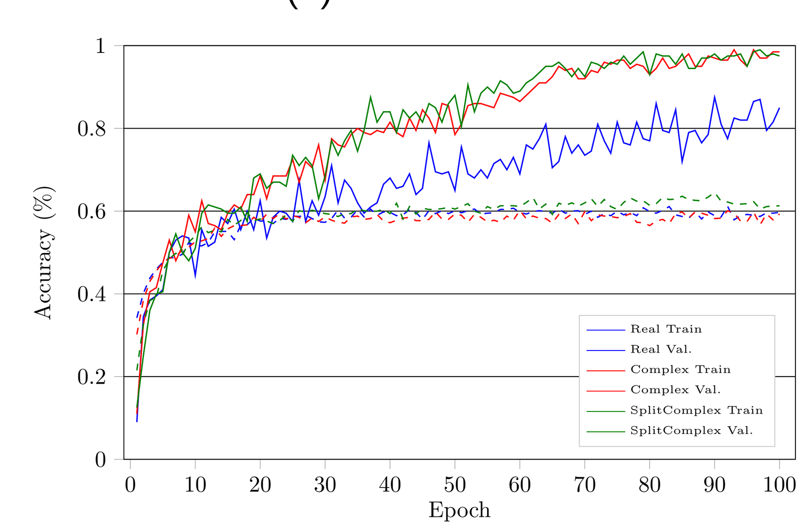
(b) Baseline model (regularized)



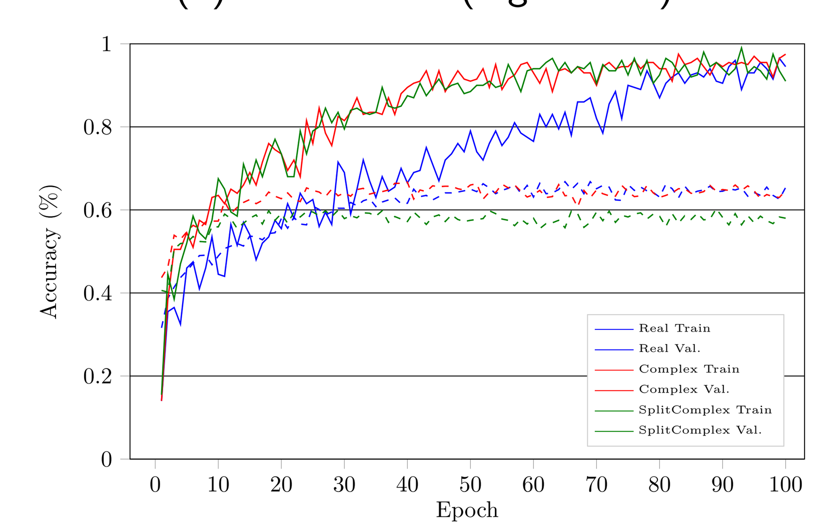
(c) Wide model



(d) Wide model (regularized)



(e) Deep model



(f) Deep model (regularized)

Results

Test set error (%) results from visual recognition experiments.

Architecture	MNIST	CIFAR-10	CIFAR-10 ($+L_2$ reg)
Real	1.1	38.3	39.0
Complex	1.1	40.6	41.4
Split-Complex	1.1	38.7	43.3
Real (Wide)	0.9	**35.1**	**35.9**
Complex (Wide)	1.0	38.7	43.6
Split-Complex (Wide)	**0.7**	38.9	41.2
Real (Deep)	**0.7**	42.2	37.9
Complex (Deep)	1.3	40.5	36.3
Split-Complex (Deep)	1.0	38.9	42.6

Discussion

- Complex and split-complex weights do not improve accuracy as much as changing network topology
  - Adding depth or width to the network seems to have greater effect
- Complex networks not self regularizing (as proposed in [2])
  - Complex/split-complex networks appear more susceptible to overfitting

Conclusion

- Locally increasing complexity of computational graph with complex arithmetic does not appear as effective as increasing depth or width

Future Work

- Improve regularization techniques for complex/split-complex networks
- Apply to contexts with complex-valued data
- Extend other Clifford algebras (e.g. quaternions) to neural networks

Acknowledgements

Many thanks to Dr. Monica Martinez-Canales (Intel Corporation) for her invaluable guidance and support on this project.

Citations

[1] J. Bruna, S. Chintala, Y. LeCun, S. Piantino, A. Szlam, and M. Tygert. A mathematical motivation for complex-valued convolutional networks. (2014):1–11, 2015.

[2] N. Guberman. On Complex Valued Convolutional Neural Networks. *arXiv*, 2016.

[3] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

[4] J. Pearson, D. Bisset, and IEEE. Neural networks in the Clifford domain. 3:1465–1469, 1994.

[5] C. Trabelsi, O. Bilaniuk, D. Serdyuk, S. Subramanian, J. F. Santos, S. Mehri, N. Rostamzadeh, Y. Bengio, and C. J. Pal. Deep Complex Networks. (Nips), 2017.