

A SUPPLEMENTARY

In this supplementary, we first demonstrate the benefits of incorporating phase information for image classification on complex-valued MRI patches.

Then, we investigate the robustness of CSR to deformations and compare it with real-valued pre-determined scattering (S) (Bruna & Mallat, 2013) and learnable scattering (LS) (Gauthier et al., 2022).

Furthermore, we provide supplementary analysis of H-CReLU components on various datasets to complement the findings presented in § 4.2.

Finally, we expand the ablation studies, analyzing the impact of various H-CReLU dimensions and presenting more ablation results in supplementary to § 4.3.

A.1 PHASE INFORMATION FOR MRI PATCH CLASSIFICATION

In § 4, we introduce a novel dataset for classifying complex-valued MRI patches. The goal is to train a classifier that can accurately identify the anatomical orientation of the complex-valued input patches. This section aims to analyze whether the inherent phase information can enhance patch classification performance.

Method	Phase	Aug.	MRI Patch	
			100 samples	500
CSR+ LL	\times	-	75.97 \pm 0.63	89.33 \pm 0.40
	\checkmark	\times	74.22 \pm 0.57	91.73 \pm 0.33
	\checkmark	\checkmark	78.53\pm0.35	92.25 \pm0.30
CSR+ CDS [†]	\times	-	85.60 \pm 0.73	95.60 \pm 0.19
	\checkmark	\times	84.80 \pm 1.06	99.18 \pm 0.15
	\checkmark	\checkmark	88.95\pm0.78	99.32\pm0.11

[†]: CDS type-I (Singhal et al., 2022); Aug.: random and constant phase augmentation

Table S1: **Classification accuracy of complex-valued MRI patch dataset with and without phase information.** To mitigate the sensitivity to phase, we also incorporate a model trained on complex-valued inputs with phase augmentation, resulting in the highest accuracy and demonstrating the importance of phase information in MRI patch classification.

Table S1 compares the performance of the CSR+LL and CSR+CDS models trained with magnitude-only MRI patches against the model trained using complex-valued input data.

Due to physics and acquisition factors, complex-valued MRI images can have random phase offsets. Therefore, two phase maps, ϕ_1 and $\phi_2 = \phi_1 + \alpha$ (where α is a constant phase ranging from $[0, 2\pi)$), provide the same phase information. To reduce the phase sensitivity, during training, we augment the phase of the input patches on the fly by multiplying each patch with random constant phase $e^{j\theta}$, where θ is uniformly distributed between $[0, 2\pi)$. We include the results in table S1.

The results indicate that when trained on only 100 samples, models trained on complex-valued inputs without phase augmentation (second row) show slightly lower accuracy than those trained on magnitude-only patches (first row) due to phase sensitivity. However, incorporating phase augmentation can considerably enhance the performance of models trained on complex-valued inputs (third row), resulting in the highest accuracy and highlighting the significance of phase information.

When moving to a larger training set of 500 samples, increased data diversity inherently mitigates the phase sensitivity. In this scenario, models trained on complex-valued inputs without phase augmentation already outperform those trained on magnitude-only images by a large margin, further highlighting the advantages of incorporating phase information in patch classification.

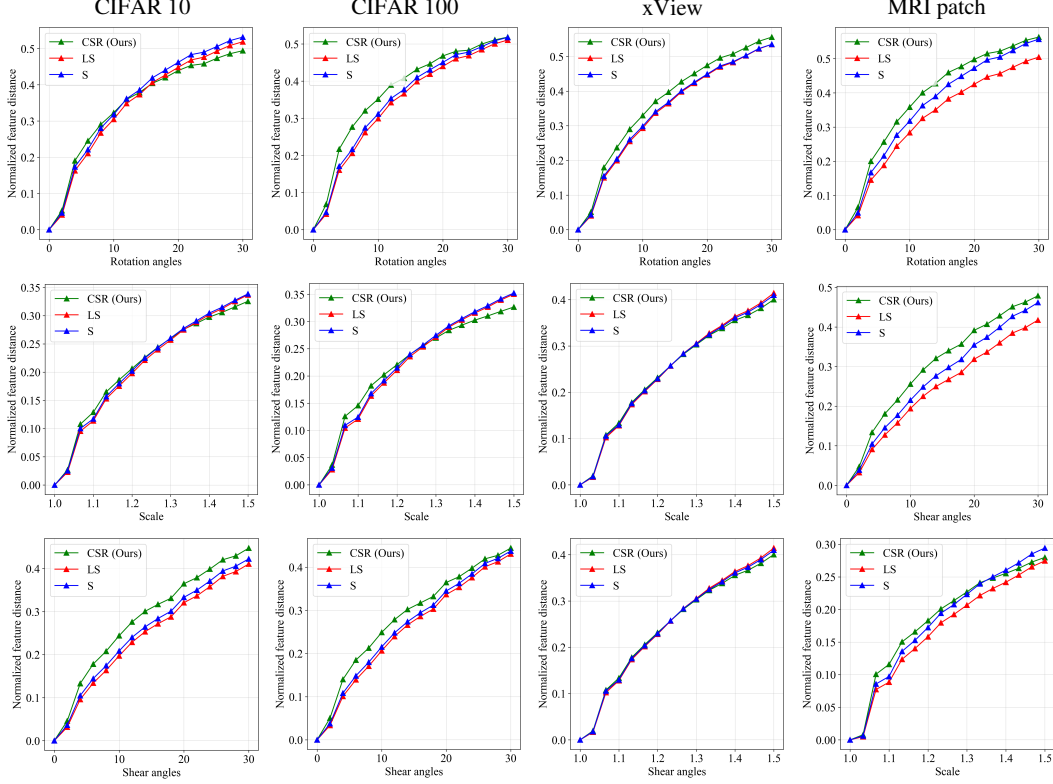


Figure S1: **Normalized distance comparisons of different deformations (i.e., rotating, scaling, shearing).** We compare the deformation stability of CSR with LS (Gauthier et al., 2022) and S (Bruna & Mallat, 2013) evaluated on various datasets. From left to right, we evaluate CIFAR 10, CIFAR 100, xView, and MRI patch classification. The plots illustrate the change in normalized distances with respect to deformation levels. CSR roughly matches the deformation stabilities of LS and S.

A.2 ROBUSTNESS TO GEOMETRIC DEFORMATIONS

Scattering transforms (Bruna & Mallat, 2013; Gauthier et al., 2022) have been shown to be stable to small deformations as a built-in feature. As described in § 3.1, CSR can be theoretically proved to be stable to deformations and invariant to local translations.

This section explores empirical results on the stability of CSR to a wider range of geometric deformations and compares it with S (Bruna & Mallat, 2013) and LS (Gauthier et al., 2022). Following Gauthier et al. (2022), we include rotation, scaling, and shearing as our deformations.

To study the robustness of deformations, we apply deformations of strength l to a given image I , resulting in a deformed image denoted as $\tilde{I}(l)$. When evaluating rotation and shearing, the deformation angle l ranges from $[0, 30^\circ)$, while the scale parameter l ranges from $[1, 1.5]$ when assessing scaling. All the deformations are implemented using torchvision.

We then compute the normalized Euclidean distance $D(l)$ between CSRs of I and $\tilde{I}(l)$ as a function of l :

$$D(l) = \frac{\|\text{CSR}(I) - \text{CSR}(\tilde{I}(l))\|_2}{\|\text{CSR}(I)\|_2}. \quad (1)$$

We compute the average $D(l)$ across the entire dataset and plot it against l . Figure S1 depicts the results obtained from all four datasets using CSR+LL: CIFAR 10, CIFAR 100, xView, and MRI patch. We include the results from S and LS for comparison.

Our observation suggests that CSR is generally on par with LS and S in terms of deformation stability across datasets. More specifically, CSR exhibits slightly better deformation robustness (lower distance values) in CIFAR 10 rotation, CIFAR 10 scaling, CIFAR 100 scaling, xView scaling, and shearing, while it shows slightly worse stability in some other scenarios.

A.3 ANALYSIS OF H-CReLU

In this section, we provide a comprehensive visual analysis of how H-CReLU works for complex number to complement the findings presented in § 4.2.

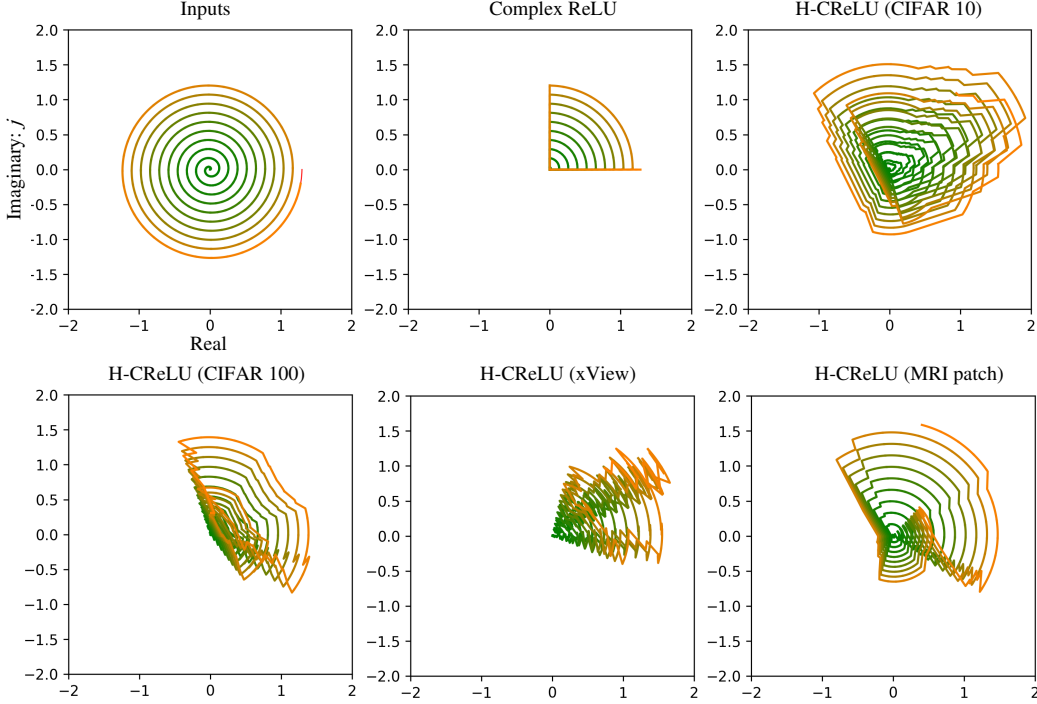


Figure S4: **Visualization of H-CReLU in mapping complex-valued points.** We generate an initial set of points on a spiral trajectory on the complex plane, where each point corresponds to a unique complex number. We then visualize how H-CReLU maps the input complex numbers to their outputs on all four datasets (*i.e.*, CIFAR 10, CIFAR 100, xView, MRI patch). The same color corresponds to the same points across the figures. CReLU results in certain input points collapsing into each other, while H-CReLU successfully avoids information loss.

Figure S4 visualizes how our learned H-CReLU $f_w(\cdot)$ maps complex numbers. Following Sandler et al. (2018), we generate an initial set of points on a spiral trajectory on the complex plane. Each point of coordinate (x, y) corresponds to a unique complex number $x + jy$. Next, we apply two activation functions, CReLU and H-CReLU, to the initial set of points. These functions transform the input points into output points, which we plot on the complex plane.

H-CReLUs are obtained from CSRs trained with LL on different datasets (*i.e.*, CIFAR 10, CIFAR 100, xView, MRI patch). The figure indicates that H-CReLU obtained from different datasets exhibit very distinct patterns. The outputs of H-CReLU from CIFAR 10 provide complete coverage of the complex plane, whereas those from xView have a limited angular span.

As a next step, we plan to improve our interpretation of the mapping results and gain a better understanding of the significance of different patterns.

Method	L. F.	H-C.	CIFAR 100	xView
CSR + LL	-	-	41.66	75.77
	✓	-	45.88 $\uparrow 4.22$	78.58 $\uparrow 2.81$
	-	✓	44.34 $\uparrow 2.68$	78.20 $\uparrow 2.43$
	✓	✓	47.81 $\uparrow 6.15$	80.04 $\uparrow 4.27$
CSR + CDS [†]	-	-	41.01	80.09
	✓	-	44.30 $\uparrow 3.29$	82.97 $\uparrow 2.88$
	-	✓	44.17 $\uparrow 3.16$	82.33 $\uparrow 2.24$
	✓	✓	46.80 $\uparrow 5.79$	84.13 $\uparrow 4.04$

[†]: CDS type-I (Singhal et al., 2022); L.F.: Learnable filters; H-C.: High-dimensional C-ReLU (H-CReLU)

Table S2: **Ablation studies of different CSR components.** We analyze the contributions of learnable filtering and H-CReLU for CSRs on CIFAR 100 and xView datasets.

Method	Activation	CIFAR 10	CIFAR 100	xView	MRI Patch
CSR + LL	Modulus (Bruna & Mallat, 2013)	71.23	45.88	78.58	68.85 \pm 0.68
	CReLU	70.88 $\downarrow 0.35$	44.34 $\downarrow 1.55$	76.52 $\downarrow 2.06$	65.33 \pm 0.88 $\downarrow 3.52$
	GTRReLU (Singhal et al., 2022)	71.04 $\downarrow 0.19$	45.65 $\downarrow 0.23$	78.01 $\downarrow 0.57$	69.08 \pm 0.45 $\uparrow 0.24$
	H-CReLU (non-learned)	72.39 $\uparrow 1.16$	46.23 $\uparrow 0.35$	78.20 $\downarrow 0.38$	70.12 \pm 0.71 $\uparrow 1.27$
	H-CReLU (learned)	74.30 $\uparrow 3.07$	47.81 $\uparrow 1.93$	80.03 $\uparrow 1.45$	74.22\pm0.57 $\uparrow 5.37$
CSR + CDS	Modulus (Bruna & Mallat, 2013)	77.60	44.30	82.97	74.03 \pm 0.56
	CReLU	75.08 $\downarrow 2.52$	42.02 $\downarrow 2.28$	80.02 $\downarrow 2.95$	71.01 \pm 0.90 $\downarrow 3.02$
	GTRReLU (Singhal et al., 2022)	78.24 $\uparrow 0.64$	44.03 $\downarrow 0.27$	81.44 $\downarrow 1.53$	76.40 \pm 0.78 $\uparrow 2.37$
	H-CReLU (non-learned)	80.63 $\uparrow 3.03$	45.52 $\uparrow 1.22$	82.40 $\downarrow 0.57$	81.45 \pm 0.83 $\uparrow 7.42$
	H-CReLU (learned)	81.52 $\uparrow 3.92$	46.80 $\uparrow 2.50$	84.13 $\uparrow 1.16$	84.80\pm1.06 $\uparrow 10.07$

Table S3: **Ablation studies of different non-linear activation functions.** We compare our H-CReLU (non-learned and learned) with other complex-valued activation functions: complex modulus, CReLU, and GTRReLU from Singhal et al. (2022). H-CReLU yields the best results.

A.4 ABLATION STUDIES

§ 4.3 evaluates the contribution of CSR components (learnable filters and H-CReLU) on CIFAR 10 and MRI patch datasets. Here, Table S2 further provides results evaluated on CIFAR 100 and xView datasets. Consistent with our previous observations, both learnable filters and H-CReLU improve classification accuracy. We observe an approximate 6% performance gain in CIFAR 100 and around 4% accuracy improvement in xView, further highlighting the effectiveness of CSR components.

To gain a deeper understanding of the H-CReLU module, we create a non-learned H-CReLU for benchmark comparisons, where the matrix $\text{DOWN}N_h$ and $\text{UP}N_h$ are not learnable. Instead, we construct the matrix such that the real and imaginary parts of each element are random numbers uniformly distributed from $[-1, 1]$.

We compare our H-CReLU (both non-learned and learned) with other complex-valued activation functions, including complex modulus, CReLU, and GTRReLU (Singhal et al., 2022), on various datasets, as reported in Table S3.

Our results indicate that H-CReLU with learnable filters produces the highest classification accuracy on all datasets. Notably, we observe that H-CReLU with non-learned random filters can outperform other counterparts on all datasets, except on the xView dataset, indicating that with high-dimensional

lifting, C-ReLU functions as a better non-linear function, even without learning. The next section will further discuss how dimensionality impacts the results.

A.5 DIMENSIONALITY OF H-CReLU

Method	H-CReLU	CIFAR 10					
		$N_h = 2$	4	8	16	32	64
CSR+LL	non-learned	66.62	68.25	70.54	72.39	72.48	70.77
	learned	72.89	73.28	73.79	74.30	73.75	73.40
CSR+CDS	non-learned	70.77	74.56	77.99	80.63	80.52	80.80
	learned	79.18	80.86	81.40	81.52	81.05	80.87

Table S4: **Ablation studies on different H-CReLU dimensionalities.** We compare the classification results of H-CReLU with varying dimensionalities (N_h), including non-learned and learned H-CReLU using CSR+LL and CSR+CDS on the CIFAR 10 dataset. **Bold** indicates best result in each row.

Thus far, we have set $N_h = 16$ for H-CReLU in all our experiments. In this section, we explore the impact of N_h on the classification results, with a particular focus on the CIFAR 10 dataset. We conduct experiments with both non-learned and learned H-CReLU with $N_h = 2, 4, 8, 16, 32, 64$.

Table S4 presents the classification results of CSR+LL and CSR+CDS on the CIFAR 10 dataset. We note that, for learned H-CReLU, $N_h = 16$ yields the highest accuracy. Remarkably, the results of $N_h = 2$ are only 1.41% lower than $N_h = 16$ (CSR+LL), and still surpass other real-valued scattering counterparts and complex-valued networks discussed in § 4. This highlights the benefits of H-CReLU when N_h is low. We note that, when $N_h > 16$, the accuracy starts to saturate and decrease. We hypothesize that H-CReLU is more susceptible to overfitting when N_h is high.

On the other hand, for non-learned H-CReLU, we observe a significant gap in accuracy across different dimensionalities, where $N_h = 2, 4$ yields markedly poorer results. We note that the accuracy begins to saturate when $N_h \geq 16$. Specifically, for CSR+LL, $N_h = 32$ yields the highest results, while for CSR+CDS, $N_h = 64$ produces the highest accuracy.

REFERENCES

- Joan Bruna and Stéphane Mallat. Invariant scattering convolution networks. *IEEE transactions on pattern analysis and machine intelligence*, 35(8):1872–1886, 2013.
- Shanel Gauthier, Benjamin Thérien, Laurent Alsene-Racicot, Muawiz Chaudhary, Irina Rish, Eugene Belilovsky, Michael Eickenberg, and Guy Wolf. Parametric scattering networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5749–5758, 2022.
- Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4510–4520, 2018.
- Utkarsh Singhal, Yifei Xing, and Stella X Yu. Co-domain symmetry for complex-valued deep learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 681–690, 2022.