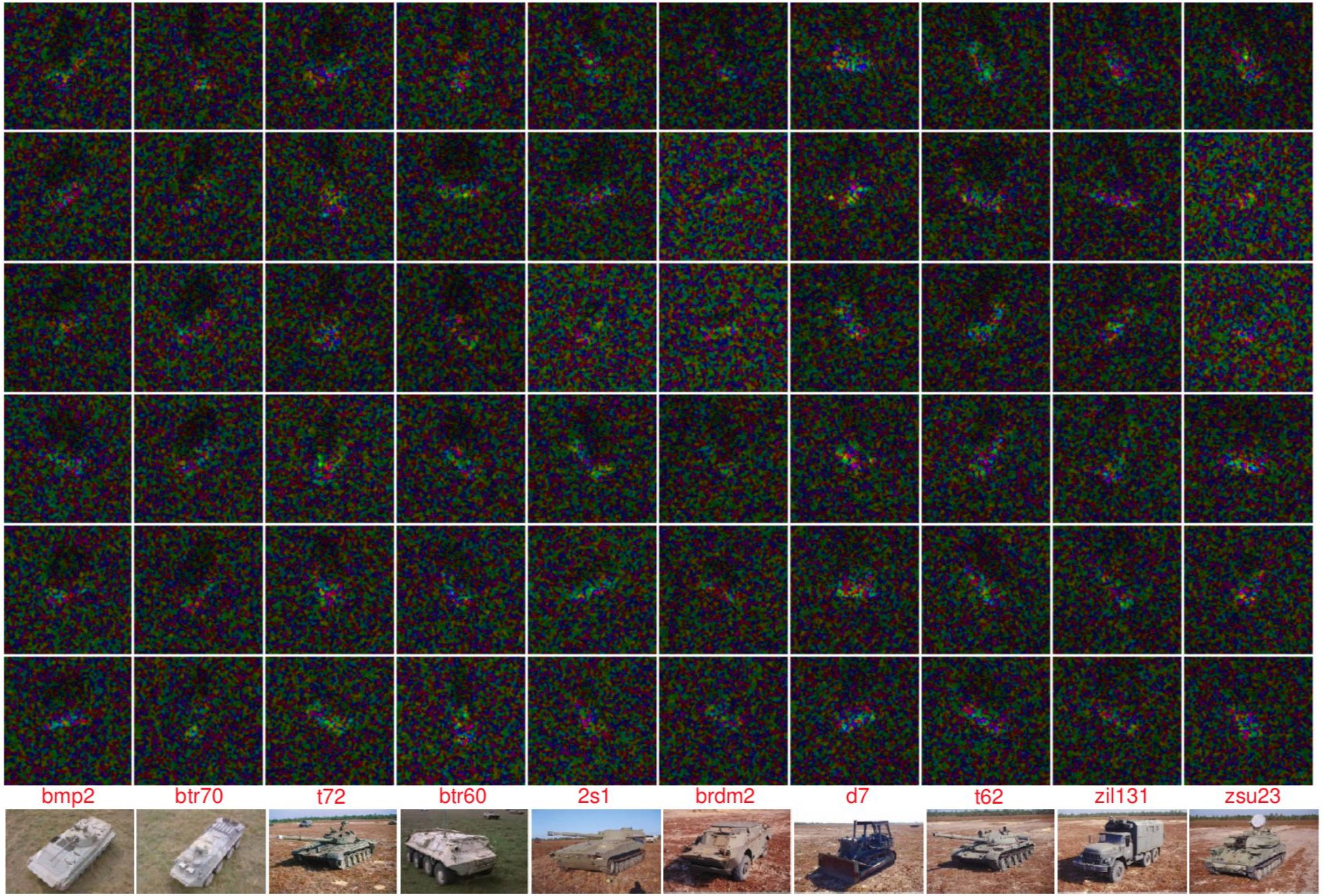
C-SURE: Shrinkage Estimator & Prototype Classifier for Complex-Valued Deep Learning

Yifei Xing Rudrasis Chakraborty Minxuan Duan Stella Yu

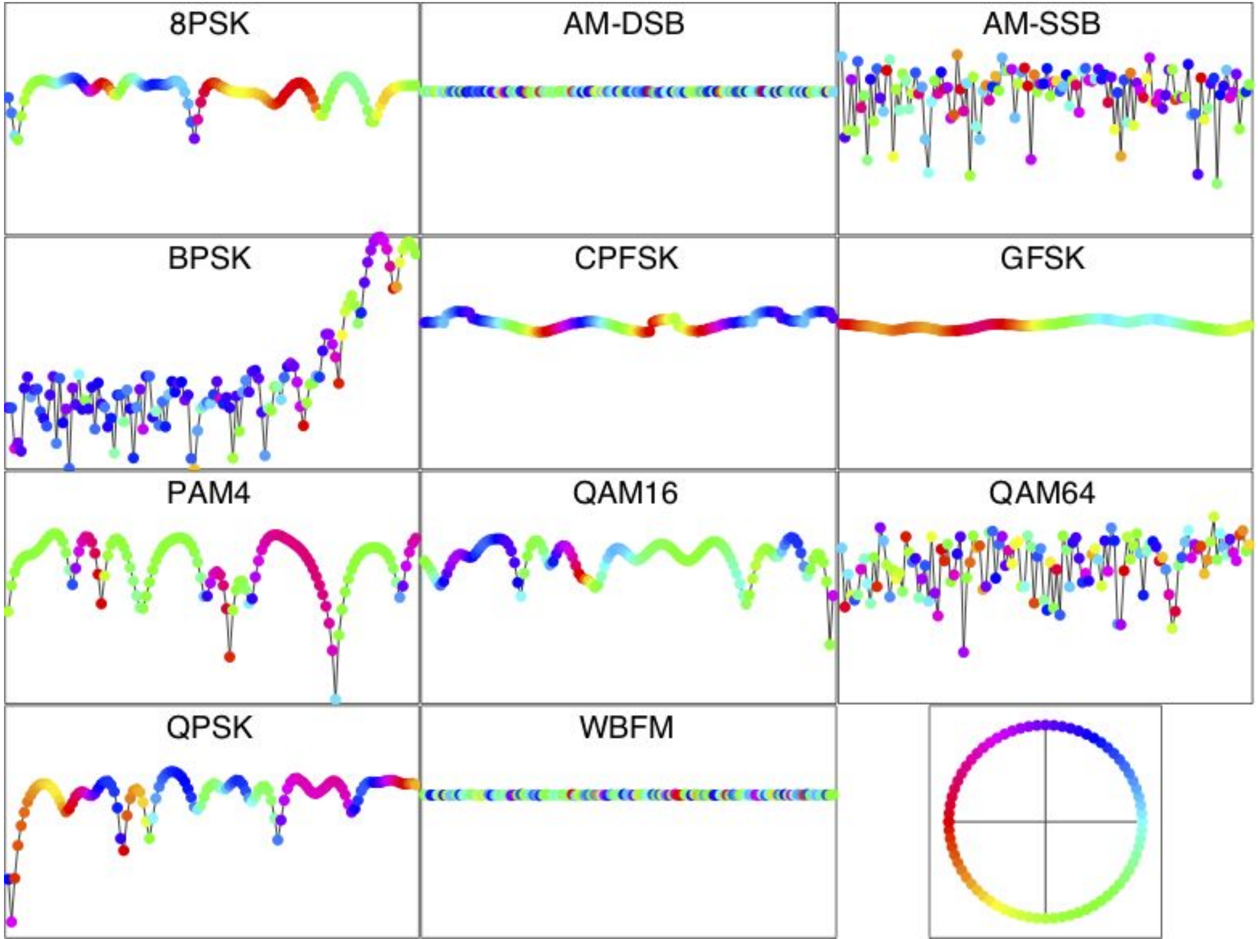
UC Berkeley / ICSI

CVPRW-PBVS
06/ 14/ 2020

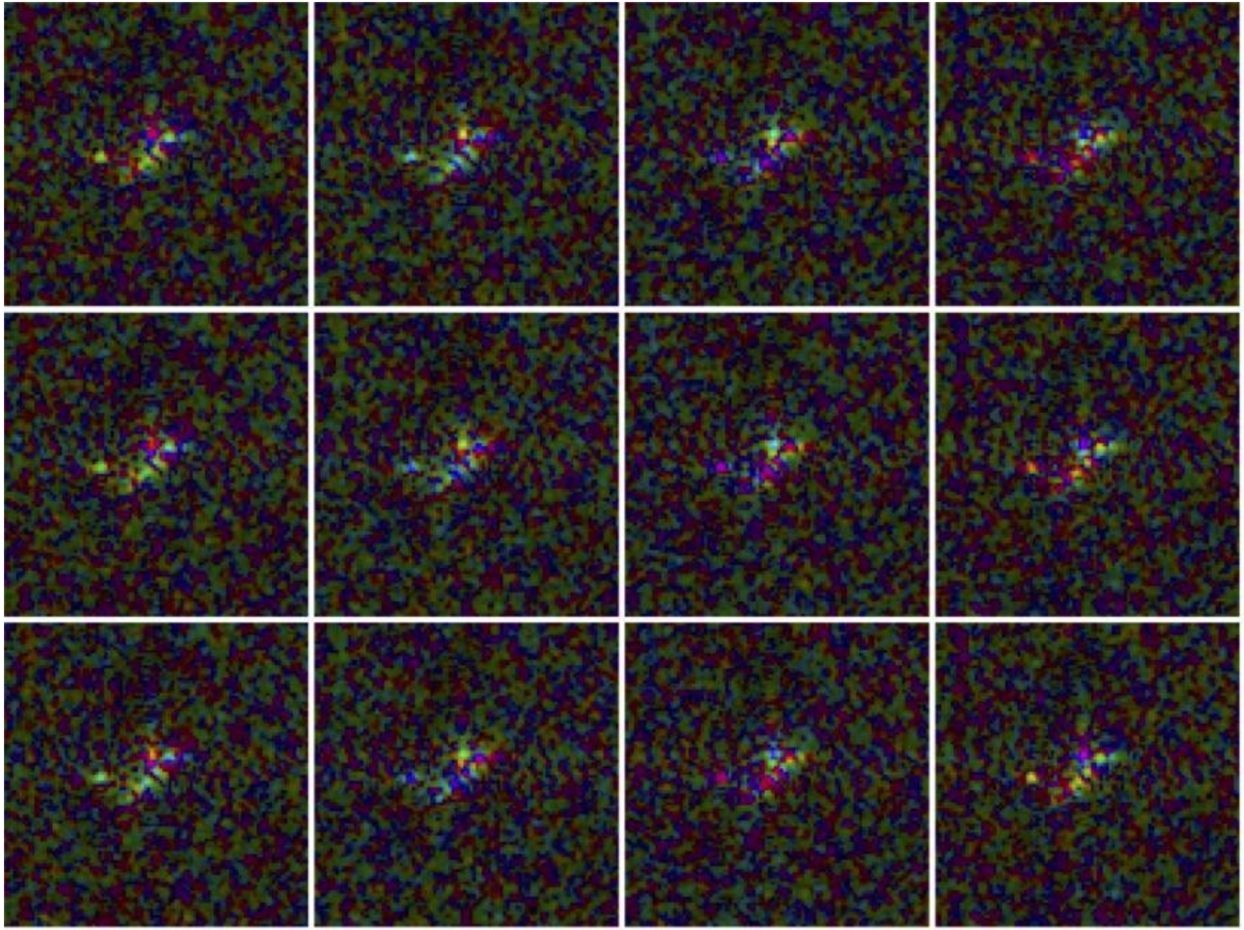
Complex-Valued Synthetic Aperture Radar Images



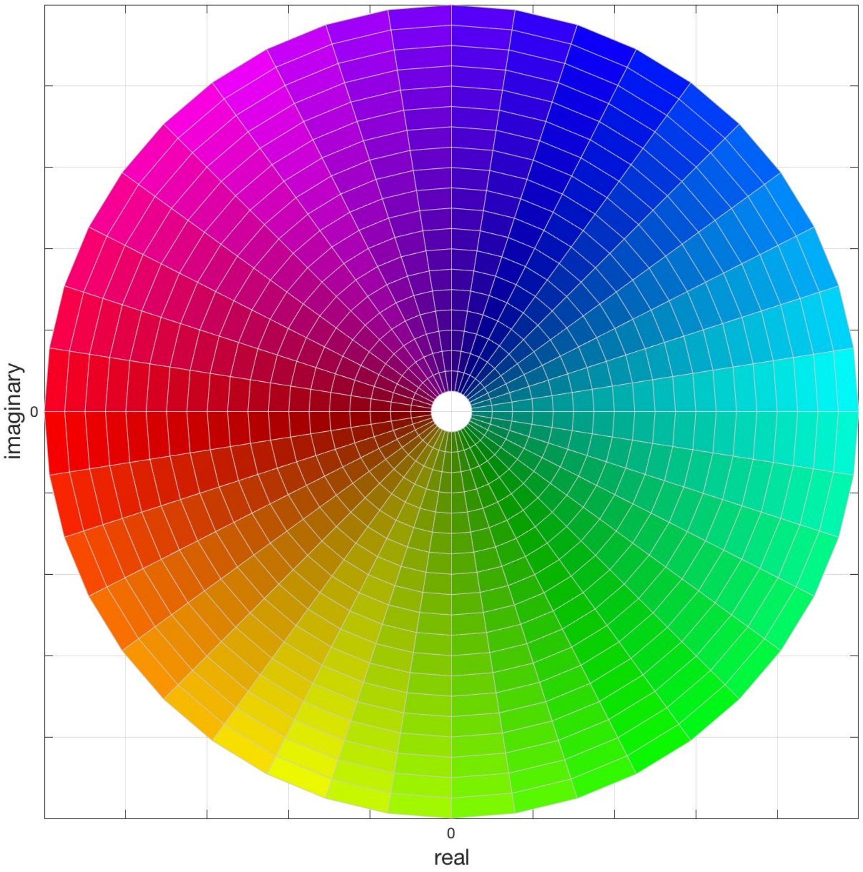
Complex-Valued Radio Frequency Signals



Complex-Valued Scaling Ambiguity



Complex Plane as a Riemannian Manifold



$$\mathbf{z} = x + iy = r * e^{j\theta}$$

$$\Leftrightarrow$$

$$(r, R(\theta)) = \left(r, \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}\right)$$

$$\Leftrightarrow$$

$$\mathbf{C} \setminus \{\mathbf{0}\} \iff \mathbf{R}^{+} \times \mathrm{SO}(2)$$

SurReal CNN Classifier Review

1. New complex-valued convolution, equivariant to complex scaling
   - Weighted Frechet mean on the complex manifold
2. New fully-connected layer function, invariant to complex scaling
   - Distance transform on the complex manifold
3. Both require mean estimation on the complex manifold

Contributions

- James-Stein estimator: A better alternative than MLE
  - Extension to our complex manifold
- SURE estimate of the James-Stein Shrinkage Estimator
  - Dominance over Fréchet mean
- Incorporate C-SURE Into a CNN classifier
  - Prototype-based classifier
  - Experimental validation

James Stein Estimator

- Hierarchical Bayesian model:

$$\mathbf{X} \sim N(\boldsymbol{\theta_i}, \sigma^2 I)$$

$$\boldsymbol{\theta}_{\mathbf{i}} \sim N(\boldsymbol{\mu}, \boldsymbol{\tau} I)$$

- Maximum *a posteriori* probability of θ :

$$\hat{\theta}(\mu, \tau; \sigma) = \frac{\tau^2}{\tau^2 + \sigma^2} X + \frac{\sigma^2}{\tau^2 + \sigma^2} \mu.$$

SURE Estimate

- Minimum squared error risk for any estimator:

$$\mathrm{MSE}(h) = \mathrm{E}_\theta[\| h(x) - \theta \|^2]$$

- SURE is unbiased in terms of the MSE risk:

$$\mathrm{E}_\theta\{\mathrm{SURE}(h)\} = \mathrm{MSE}(h)$$

- SURE does not depend on the unknown θ:

$$\mathrm{SURE}(\mu, \tau) = -p\sigma^2 + \|\hat{\theta} - X\|^2 + 2\sigma^2 \sum_{i=1}^{p} \frac{\partial \hat{\theta}}{\partial X_i}$$

- SURE as a proxy for hyperparameter selection:

$$\hat{\mu}^{\mathrm{SURE}}, \hat{\tau}^{\mathrm{SURE}} = \arg\min_{\mu,\tau} \mathrm{SURE}(\mu, \tau)$$

Our C-SURE Estimator for Complex Manifold

- Assume each class a mixture of Gaussians on complex manifold:

$$X_i | M_i \overset{ind}{\sim} \mathrm{LN}\left(M_i, \nu I\right), i = 1, \ldots, p$$

$$M_i \overset{i.i.d}{\sim} \mathrm{MLN}(w, \mu, D).$$

- C-SURE estimate for each component:

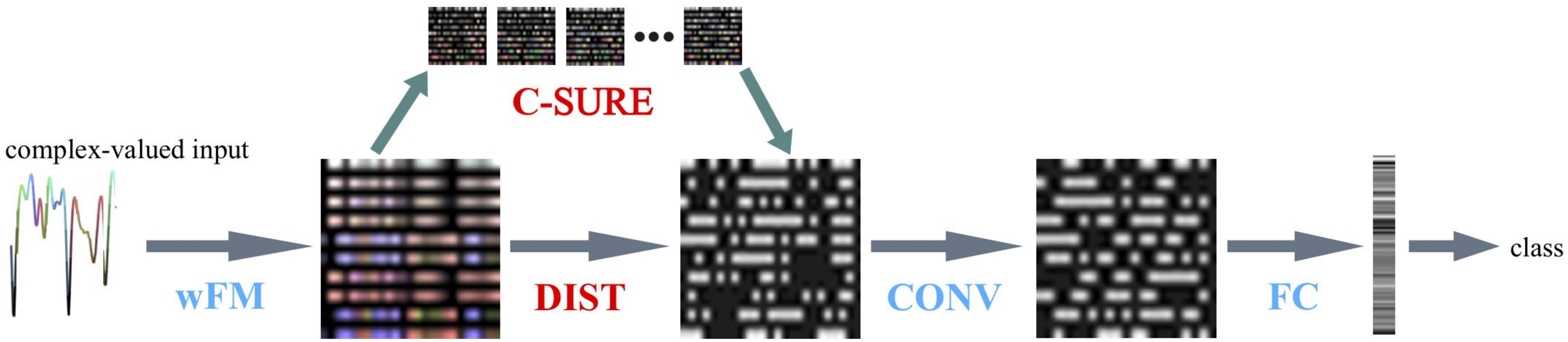
$$\left(\hat{\mu}_k^{\mathrm{SURE}}, \hat{\lambda}_k^{\mathrm{SURE}}\right) = \arg\min_{\mu_k, \lambda_k} \sum_{i=1}^{p} \frac{\nu}{(\lambda_k + \nu)^2} \left(\nu \left\| \log \overline{X}_i^{\mathrm{LE}} - \log \mu_k \right\|^2 + \frac{p\left(\lambda_k^2 - \nu^2\right)}{N} \right)$$

- C-SURE shrinkage estimator for the mixture of components:

$$\widehat{M}_i^{\mathrm{SURE}}(w) = \sum_{k=1}^{K} \exp\left(w_k \left(\frac{\hat{\lambda}_k^{\mathrm{SURE}}}{\hat{\lambda}_k^{\mathrm{SURE}} + \nu} \log \overline{X}_i^{\mathrm{LE}} + \frac{\nu}{\hat{\lambda}_k^{\mathrm{SURE}} + \nu} \log \hat{\mu}_k^{\mathrm{SURE}} \right) \right)$$

Old Model: SurReal Discriminator Classifier, 2019

New Model: C-SURE Prototype Classifier

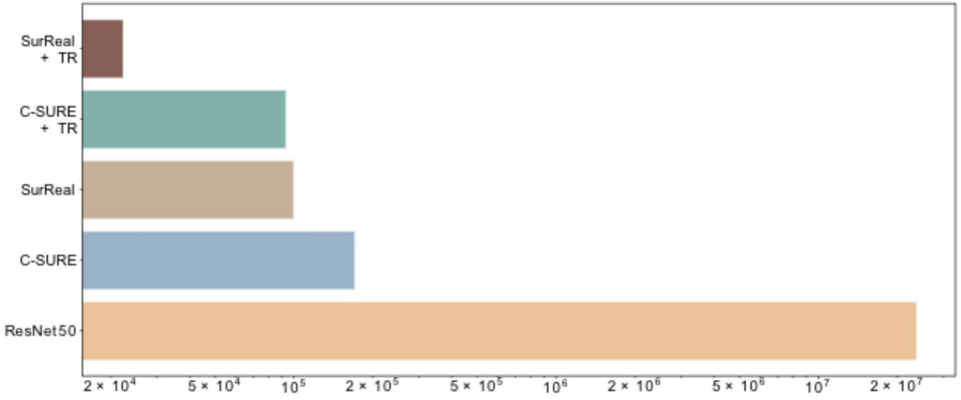


1. New distance transform layer:
    a. Compute the C-SURE estimate per class
    b. Compute the min distance to all the class means per feature
2. Continue real-valued classification upon the distance feature

C-SURE Outperforms Real-Valued and SurReal CNNs

Dataset	Real-Valued	SurReal	C-SURE
MSTAR-L	99.1%	**99.2%**	**99.2%**
MSTAR-S	97.4%	97.7%	**98.1%**
RadioML	75.8%	78.4%	**81.6%**

1. More accurate
2. Much smaller than Real



3. More robust, stable, fast converging than SurReal
4. Better than MLE for the prototype CNN classifier