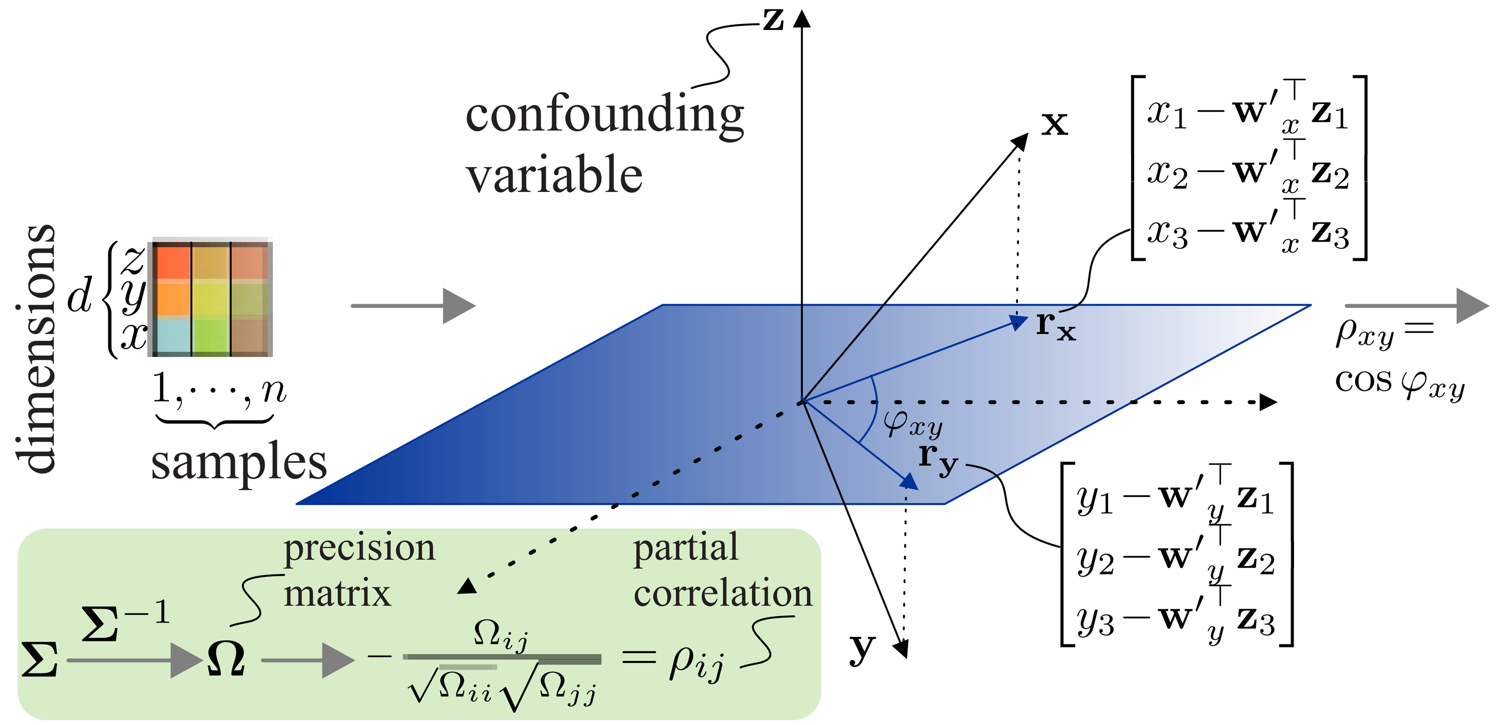
Learning Partial Correlation based Deep Visual Representation for Image Classification

Saimunur Rahman[1,2], Piotr Koniusz[1,3], Lei Wang[2], Luping Zhou[4], Peyman Moghadam[1,5], Changming Sun[1]
[1]CSIRO Data61, [2]University of Wollongong, [3]Australian National University, [4]University of Sydney, [5]Queensland University of Technology

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Key highlights:

- Pairwise correlation computed from CNN channels becomes "contaminated" once there is another channel correlating with both channels of interest, resulting in the "confounding" effect.



- "Partial correlation" removes the confounding effect. Sparse inverse covariance estimation (SICE) also removes confounding effect and is more robust to estimation given limited data.

SICE estimation with CNN:

- Partial correlation estimation from low spatial resolution and large number of CNN feature channels is challenging.

SICE is defined as follows:

$$S^* = \arg\max_{S \succ 0} \operatorname{logdet}(S) - \operatorname{trace}(\mathbf{\Sigma S}) - \lambda \|\mathbf{S}\|_1$$

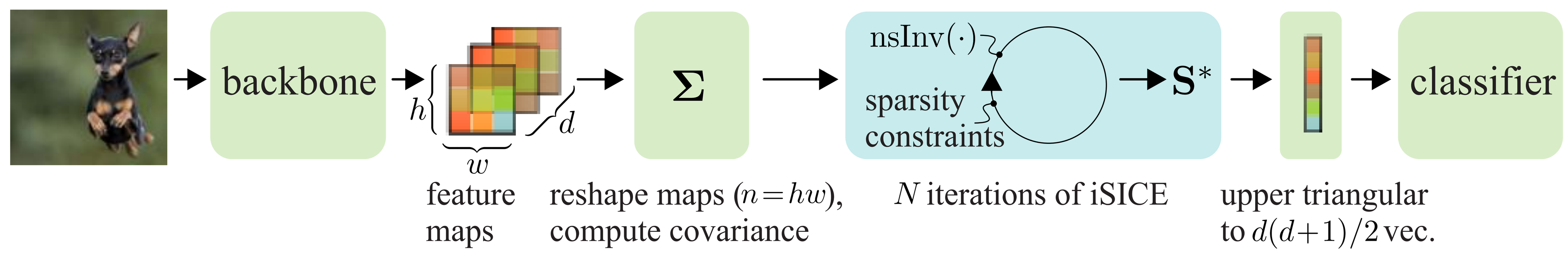
where Σ is a sample-based covariance matrix, and $\det(\cdot)$, $\operatorname{trace}(\cdot)$ and $\|\cdot\|_1$ denote the determinant, trace and $\ell 1$-norm, respectively.

- The is no suitable solver available for running SICE in CNN.
  - Convex solvers such as CVXPY and GLASSO do not support backprop.
  - CVXPYLayers is inefficient in handling large SICE problems.
- We propose an effective method for SICE based on NS iterations.

Proposed iterative SICE (iSICE):

$$\frac{\partial J}{\partial \mathbf{S}} = \frac{\partial}{\partial \mathbf{S}} \operatorname{logdet}(\mathbf{S}) - \frac{\partial}{\partial \mathbf{S}} \operatorname{trace}(\mathbf{\Sigma S}) - \lambda \frac{\partial}{\partial \mathbf{S}} \|\mathbf{S}\|_1$$
$$= \mathbf{S}^{-1} - \mathbf{\Sigma} - \lambda \left(\frac{\partial}{\partial \mathbf{S}} \mathbf{S}^{+} - \frac{\partial}{\partial \mathbf{S}} \mathbf{S}^{-} \right)$$
$$= \mathbf{S}^{-1} - \mathbf{\Sigma} - \lambda (\operatorname{sign}(\mathbf{S}^{+}) - \operatorname{sign}(\mathbf{S}^{-}))$$

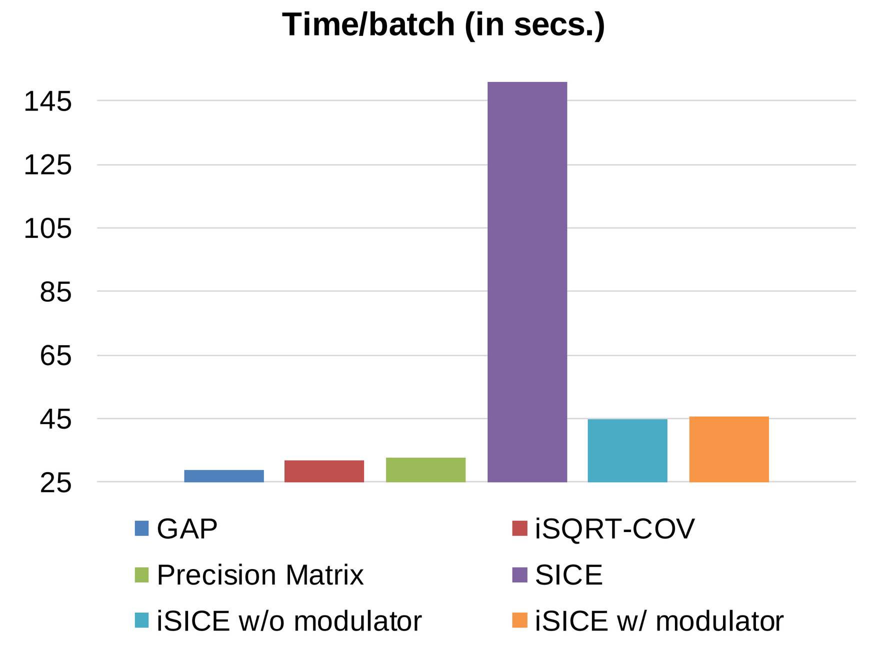
Proposed iSICE integration with CNN:

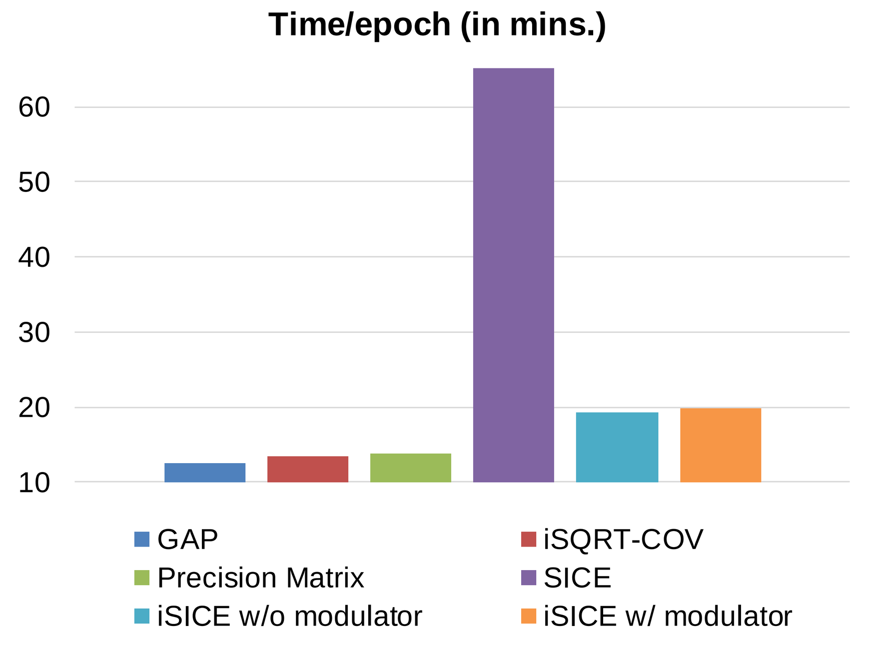


iSICE vs. its covariance counterparts:

Method	Matrix Dim.	MIT		Airplane		Birds		Cars		Average	
		VGG	ResNet	VGG	ResNet	VGG	ResNet	VGG	ResNet	VGG	ResNet
iSQRT-COV	256 × 256	76.1	78.8	90.0	90.9	84.5	84.3	91.2	92.1	85.5	86.5
	512 × 512	76.9	**82.8**	91.5	91.1	85.0	84.5	92.2	92.1	86.4	87.6
Precision **Ω**	256 × 256	**80.2**	**80.8**	89.4	91.2	83.4	84.7	92.0	92.0	86.3	87.1
	512 × 512	80.7	82.7	90.1	91.5	84.9	84.0	92.5	92.6	87.0	87.7
SICE	128 × 128	71.0	73.1	85.5	86.9	77.3	78.0	87.0	87.9	80.2	81.5
	256 × 256	73.7	75.4	87.9	89.2	79.7	80.3	89.5	89.3	82.7	83.6
iSICE	256 × 256	78.7	80.5	**92.2**	**92.7**	**86.5**	**85.9**	**94.0**	**93.5**	**87.9**	**88.2**
	512 × 512	**81.1**	81.7	**92.9**	**92.6**	**86.8**	**86.0**	**94.6**	**93.8**	**88.9**	**88.5**

Computation cost:

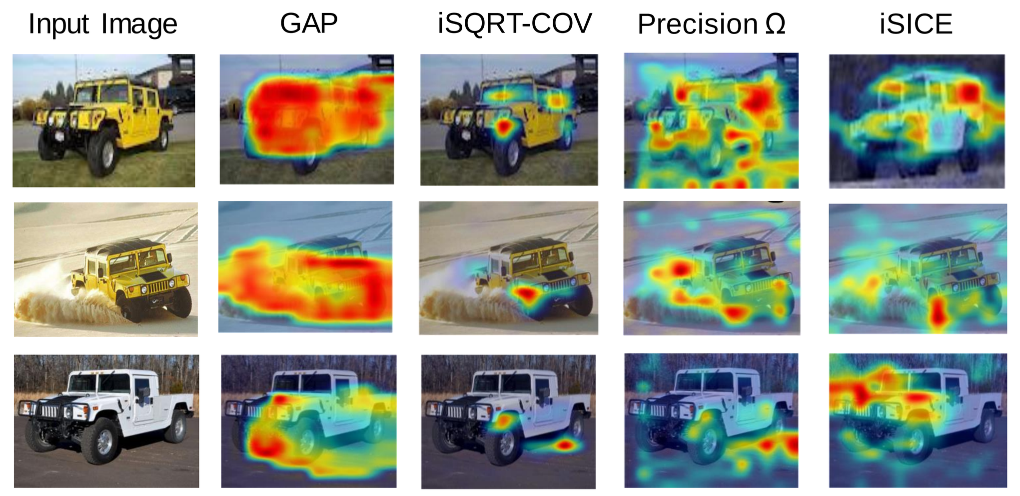




Comparison with state-of-the-art:

Method	Backbone	MIT	Airplane	Birds	Cars	DTD	iNatuarlist	mini-ImageNet
GAP [37]	VGG-16	–	76.6	70.4	79.8	–	–	–
NetVLAD [2]		–	81.8	81.6	88.6	–	–	–
NetFV [28]		–	79.0	79.9	86.2	–	–	–
BCNN [27]		77.6	83.9	84.0	90.6	70.6	–	–
CBP [11]		76.2	84.1	84.3	91.2	67.7	–	–
LRBP [17]		–	87.3	84.2	90.9	–	–	–
KP [6]		–	86.9	86.2	92.4	–	–	–
HIHCA [4]		–	88.3	85.3	91.7	–	–	–
Improved BCNN [25]		–	88.5	85.8	92.0	–	–	–
SMSO [46]		79.5	–	85.0	–	–	–	–
MPN-COV [43] (reproduced)		–	86.1	82.9	89.8	–	–	–
iSQRT-COV [23] (reproduced)		76.1	90.0	84.5	91.2	71.3	56.2	76.2
DeepCOV [9]		79.2	88.7	85.4	91.7	–	–	–
DeepKSPD [9]		**81.0**	90.0	84.8	91.6	–	–	–
RUN [47]		80.5	91.0	85.7	–	–	–	–
FCBN [48]		80.3	90.5	85.5	–	–	–	–
TKPF [49]		80.5	91.4	86.0	–	–	–	–
Precision **Ω**		80.2	89.4	83.4	92.0	74.0	57.9	81.0
iSICE (ours)		78.7	**92.2**	**86.5**	**94.0**	**74.7**	**59.8**	**78.7**
CBP [11]	ResNet-50	–	81.6	81.6	88.6	–	–	–
KP [6]		–	85.7	84.7	91.1	–	–	–
SMSO [46]		79.7	–	85.8	–	–	–	–
iSQRT-COV [23] (reproduced)		78.8	90.9	84.3	92.1	73.0	57.7	70.7
DeepCOV-ResNet [34]		83.4	83.9	**86.0**	85.0	–	–	–
TKPF [49]		**84.1**	92.1	85.7	–	–	–	–
Precision **Ω**		80.8	91.2	84.7	92.0	73.7	59.6	65.6
iSICE (ours)		80.5	**92.7**	85.9	**93.5**	**75.7**	**60.7**	**72.0**
iSQRT-COV [23]	VGG-19	76.3	90.3	84.1	91.4	71.8	56.9	75.4
Precision **Ω**		79.6	91.1	83.2	92.2	74.2	57.3	73.8
iSICE (ours)		**80.6**	**92.5**	**86.6**	**93.9**	**74.9**	**59.6**	77.1
iSQRT-COV [23]	ResNet-101	79.3	91.0	84.4	92.3	73.0	70.6	73.9
Precision **Ω**		77.9	90.1	83.3	91.4	71.2	69.8	73.0
iSICE (ours)		**81.0**	**92.9**	**86.6**	**93.6**	**75.4**	**72.0**	**78.0**
iSQRT-COV [23]	ResNeXt-101	81.6	91.3	86.2	92.4	75.7	72.2	76.1
Precision **Ω**		85.7	90.2	84.6	89.9	76.9	72.3	77.6
iSICE (ours)		**86.3**	**94.6**	**87.2**	**94.5**	**78.7**	**73.8**	**81.0**
iSQRT-COV [23]	ConvNext-T	77.8	88.1	83.5	89.4	84.7	61.5	82.0
Precision **Ω**		78.5	81.2	83.7	92.2	83.9	59.3	83.6
iSICE (ours)		**85.4**	**90.4**	**86.7**	**93.1**	**88.9**	**65.0**	**85.1**
iSQRT-COV [23]	Swin-T	82.1	87.6	85.1	89.7	86.1	58.1	67.7
Precision **Ω**		82.5	88.2	84.9	90.5	86.5	59.1	65.6
iSICE (ours)		**85.9**	**89.6**	**86.5**	**91.3**	**88.3**	**61.9**	**69.1**
iSQRT-COV [23]	Swin-B	86.6	91.3	88.0	92.0	79.4	69.7	64.9
Precision **Ω**		87.0	90.7	87.7	93.1	**80.1**	67.3	66.4
iSICE (ours)		**87.6**	**92.9**	**88.3**	**93.3**	79.8	**72.4**	**68.4**

Feature visualisation:



Paper website & code:

References:

1. Zhang et al. "Beyond covariance: SICE and kernel based visual feature representation." In IJCV 129 (2021): 300-320.
2. Peihua, et al. "Towards faster training of GCP networks by iterative matrix square root normalization." *In CVPR* 2018.
3. Lin and Maji. "Improved bilinear pooling with CNNs." In BMVC 2017.
4. Friedman et al. "Sparse inverse covariance estimation with the graphical lasso. Biostatistics, 9(3), 432-441.