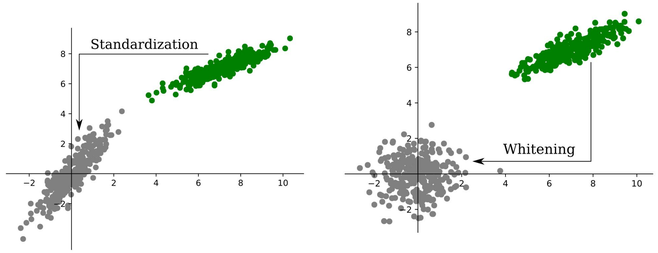
Whitening Coloring Batch Transform for GANs

Aliaksandr Siarohin, Enver Sangineto, Nicu Sebe
DISI, University of Trento, Italy

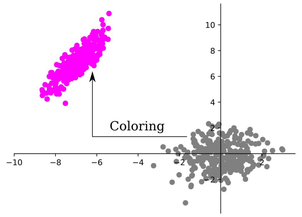
UNIVERSITÀ DEGLI STUDI DI TRENTO
Dipartimento di Ingegneria
e Scienza dell'Informazione

Whitening Coloring Batch Transform:

- Whitening is a normalization technique. Contrary to batch norm (standardization), whitening also decorelates features:

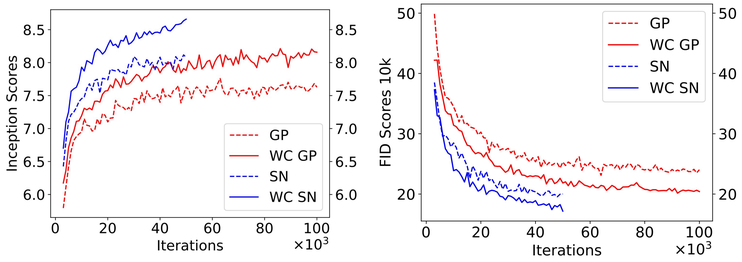


- Whitening reduces the network capacity. To restore the original capacity coloring is introduced:



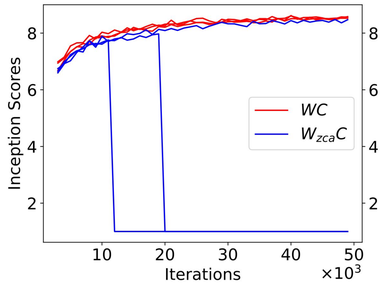
Cifar10	IS	FID
W-only	6.63	36.8
WC-diag	7.00	34.1
WC	**8.66**	**17.2**

- Whitening improves the conditioning number of the generator Jacobian. Controlling the conditional number the Jacobian plays an important role in the improve GAN training[1]:



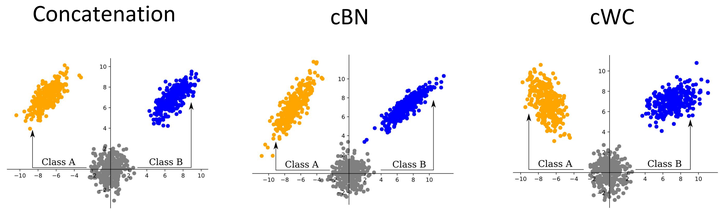
Cholesky Whitening:

- We propose Cholesky based whitening. It has marginal runtime overhead (32%). ZCA whitening [2] is an order of magnitude slower. ZCA whitening also has unstable gradient computation:

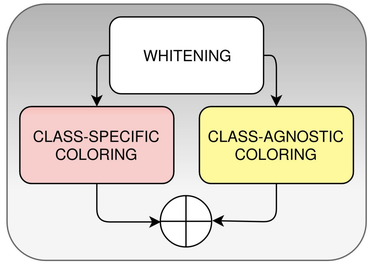


Conditional Coloring (cWC):

- Conditional coloring is similar to cBN. Coloring parameters depend on the input condition (class label in our case). Conditional coloring is the more expressive than other commonly used methods:



- Conditional coloring is only works jointly with whitening and with a class-agnostic branch:



Cifar10	IS	FID
c-std-C	7.92	24.4
cWC-cls-only	8.10	28.0
WC	**8.97**	**13.5**

Parameter-efficient conditional coloring (cWC_{sa}):

- A dictionary of coloring filters is learned, a filter for each class is a linear combination of the dictionary filters. Another view: a single layer predicts coloring filters from one-hot class embedding.
- This technique scales up to ImageNet. We can obtain better performance than cBN based networks, with significantly less parameters in the generator:

Imagenet	#params	IS
cBN	45M	29.7
cWC_{sa}	**6M**	**34.4**

Discriminative experiments:

- WC usually achieves a test error slightly better than BN. Hoewer, WC has significantly higher train error, thus can be trained for more iteration than BN:

	Cifar10		Cifar100	
	ResNet-32	ResNet-56	ResNet-32	ResNet-56
BN	7.31	7.21	31.41	30.86
WC	7.30	6.33	29.50	28.69
WC x2	**6.37**	**5.95**	**29.00**	**27.21**

Acknowledgment:

- Many thanks to Anirudh Goyal, who kindly presented this poster.

1. Odena et al. Is generator conditioning causally related to gan performance?
2. Huang et al. Decorrelated batch normalization.
3. Dumoulin, et al., "Feature-wise transformations"

Our code is publicly available:
https://github.com/AliaksandrSiarohin/wc-gan