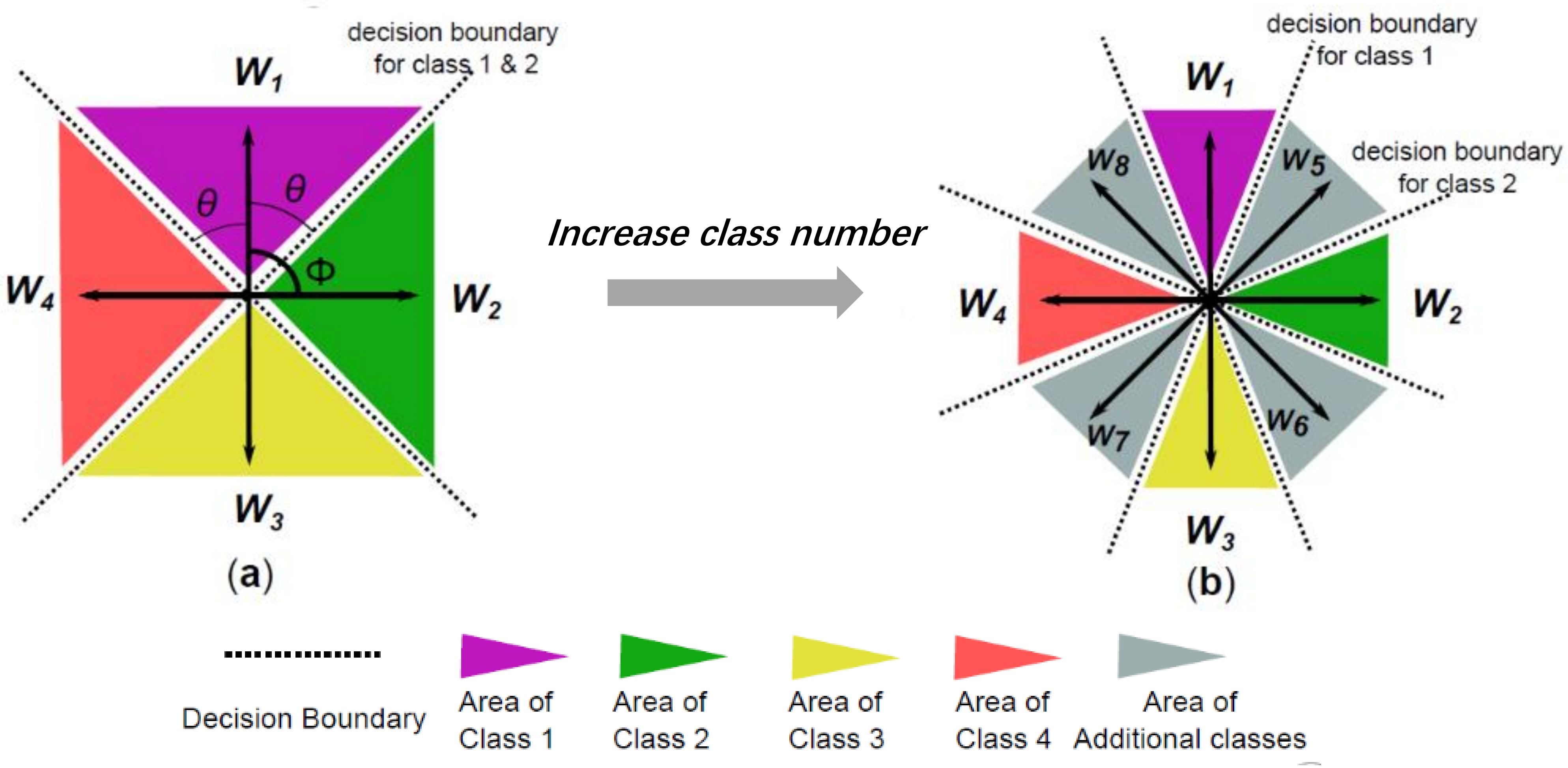
Virtual Class Enhanced Discriminative Embedding Learning

Binghui Chen, Weihong Deng, Haifeng Shen
BUPT & DiDi

32nd Conference on Neural Information Processing Systems (NeurIPS), 2018, Montréal, Canada.

Observation & Motivation

- For d-dimensional feature space under Softmax classifier, the feature region of each class is inversely proportional to the number of class.



Virtual Softmax :

Learning towards discriminative image features

- Formulation: inject a dynamic virtual negative class

$$L = -\frac{1}{N}\sum_{i=1}^{N}\log\frac{e^{W_{y_i}^T X_i}}{\sum_{j=1}^{C} e^{W_j^T X_i} + e^{W_{virt}^T X_i}}$$

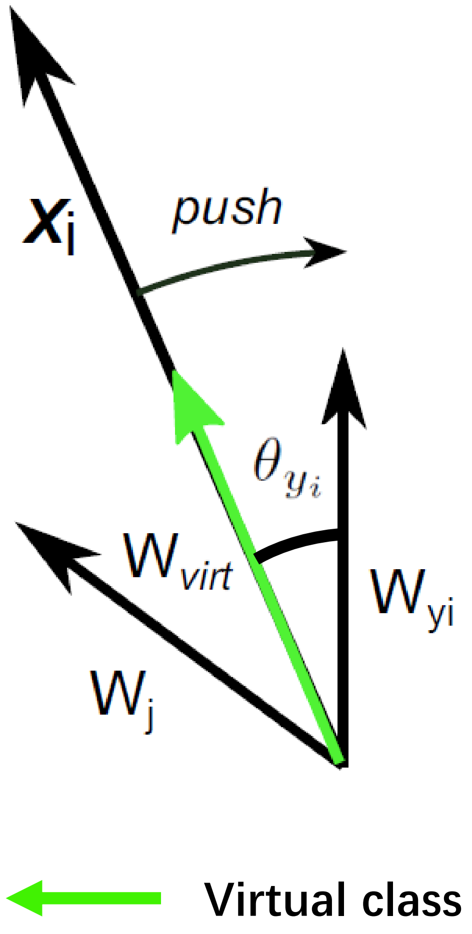
where $W_{virt} = \frac{\|W_{y_i}\| X_i}{\|X_i\|}$

- Optimization goal:

$$W_{y_i}^T X_i \geq \max_{j\in C+1}(W_j^T X_i) = W_{virt}^T X_i$$

i.e. $\|W_{y_i}\|\|X_i\|\cos\theta_{y_i} \geq \|W_{y_i}\|\|X_i\|$

$$\Rightarrow \theta_{y_i} = 0$$



- Optimization goal of Virtual Softmax:

$$W_{y_i}^T X_i \geq \max_{j \in C+1}(W_j^T X_i) = W_{virt}^T X_i$$

$$\text{i.e.} \quad \|W_{y_i}\|\|X_i\| \cos\theta_{y_i} \geq \|W_{y_i}\|\|X_i\|$$

$$\longrightarrow \quad \theta_{y_i} = 0$$

- The conventional Softmax learns towards a weaker goal:

$$\|W_{y_i}\| \cos\theta_{y_i} \geq \max_{j \in C}\left(\|W_j\| \cos\theta_j\right)$$

$$\longrightarrow \quad \theta_{y_i} \leq \min_{j \in C}\left(\arccos\left(\frac{\|W_j\|}{\|W_{y_i}\|} \cos\theta_j\right)\right)$$

Objective comparison

Discussion：

- Interpretation from *Coupling Decay:*

$$L_i = -\log \frac{e^{W_{y_i}^T X_i}}{\sum_{j=1}^{C} e^{W_j^T X_i} + e^{W_{virt}^T X_i}} \qquad (1)$$

$$L_i = -W_{y_i}^T X_i + \log\left(\sum_{j=1}^{C} e^{W_j^T X_i} + e^{\|W_{y_i}\| \|X_i\|}\right) \qquad (2)$$

perform the first order Taylor Expansion for the second log-term in Eq.2, a term of $\|W_{y_i}\|\|X_i\|$ shows up. Therefore, minimizing the above equation is to minimize $\|W_{y_i}\|\|X_i\|$ to some extend, and this can be viewed as a coupling decay term, i.e. **Data-Dependent Weight Decay** and **Weight-Dependent Data Decay**.

- Interpretation from *Feature Update:*

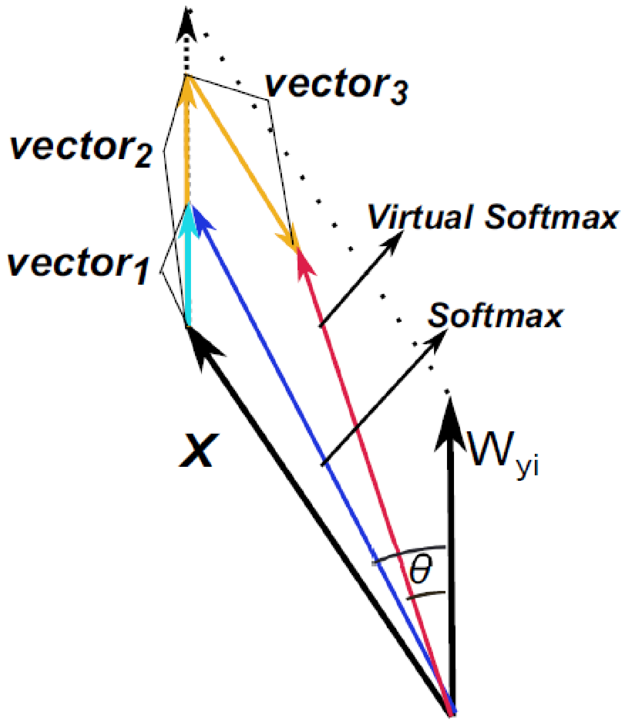
For a linear neural layer, the *Feature Update* by Softmax and our Virtual Softmax is like:

Softmax:

$$X' = X + \underbrace{\beta(W_{y_i} - \frac{e^{W_{y_i}^T X} W_{y_i}}{\sum_{j=1}^{C} e^{W_j^T X}})}_{vector_1}$$

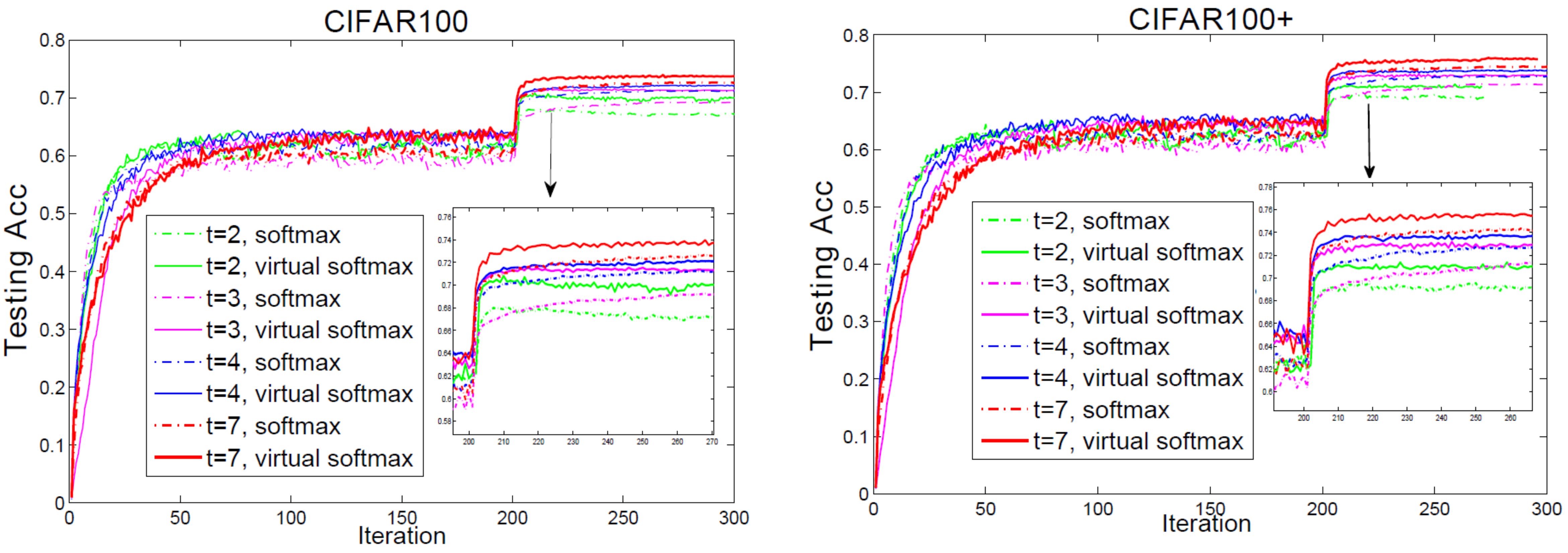
Virtual Softmax:

$$X' = X + \underbrace{\beta(W_{y_i} - \frac{e^{W_{y_i}^T X} W_{y_i}}{\sum_{j=1}^{C} e^{W_j^T X} + e^{W_{virt}^T X}})}_{vector_2} \underbrace{-\beta \frac{e^{W_{virt}^T X} W_{virt}}{\sum_{j=1}^{C} e^{W_j^T X} + e^{W_{virt}^T X}}}_{vector_3}$$



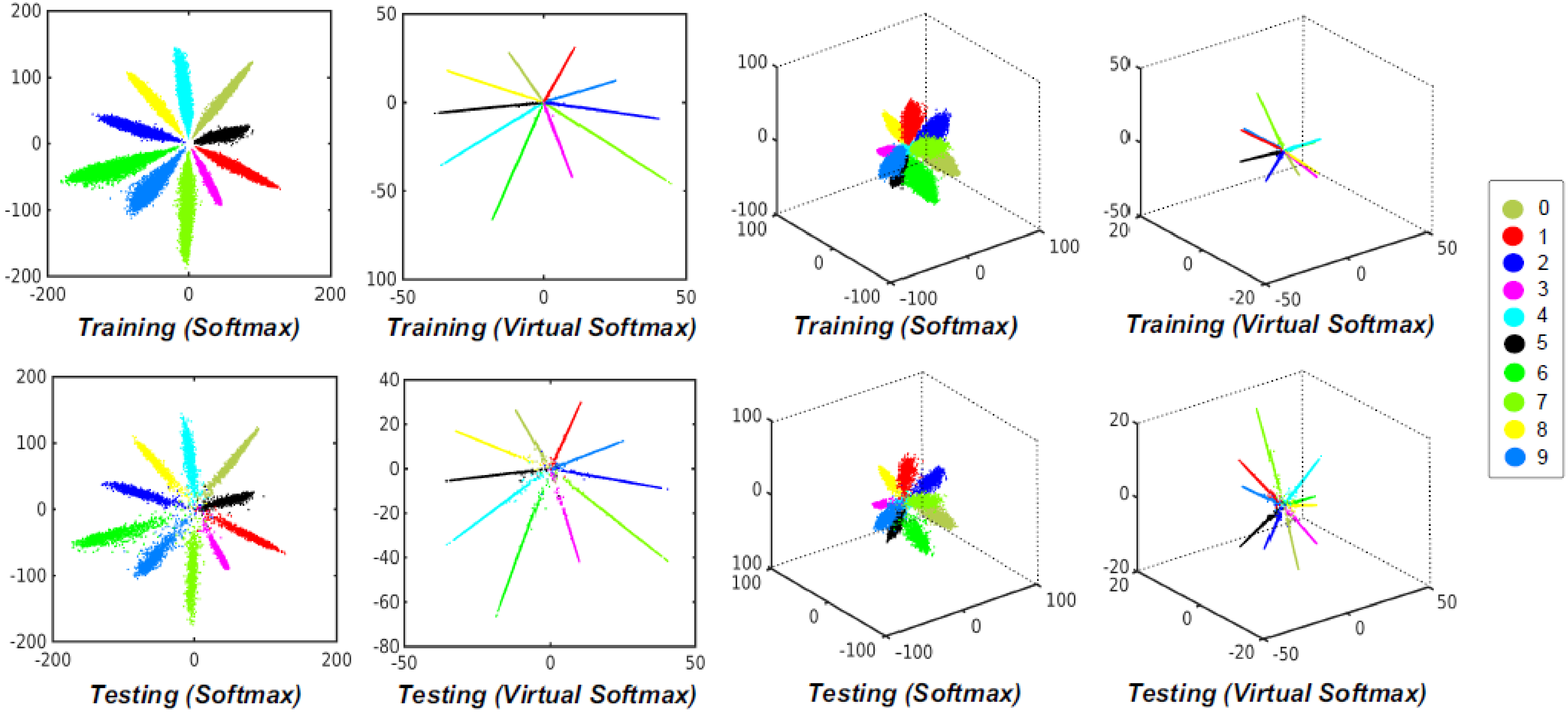
Experiments：

- Similar convergence and higher accuracy on CIFAR100:



Experiments :

- Visualization of Feature Compactness and Separability on MNIST :



Testing Accuracy: 98.91%

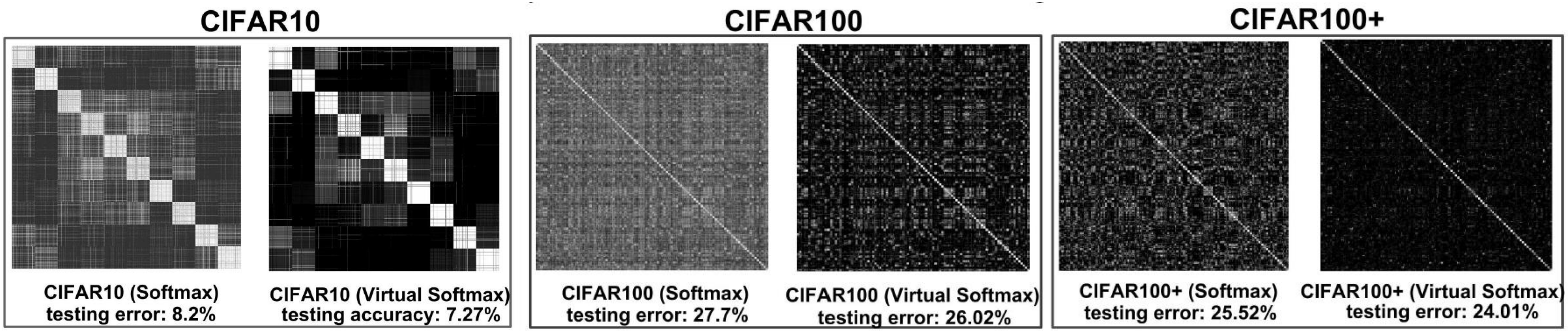
Testing Accuracy: 99.2%

Testing Accuracy: 99.13%

Testing Accuracy: 99.38%

Experiments :

- Visualization of intra-class and inter-class similarities on CIFAR10, CIFAR100:



Experiments :

- Performances on small-scale object classification datasets:

Method	MNIST(%)	SVHN(%)
Maxout [7]	0.45	2.47
DSN [19]	0.39	1.92
R-CNN [20]	0.31	1.77
WRN [39]	-	1.85
DisturbLabel [38]	0.33	2.19
Noisy Softmax [3]	0.33	-
L-Softmax [22]	0.31	-
Softmax	0.35	2.11
NS*[3]	0.32	2.04
LS*[22]	0.30	2.01
AS*[21]	0.31	2.04
Virtual Softmax	***0.28***	***1.93***

Table 3: Recognition error rates on MNIST and SVHN. * denotes our reproducing.

Method	CIFAR10(%)	CIFAR100(%)	CIFAR100+(%)
GenPool [18]	7.62	32.37	-
DisturbLabel [38]	9.45	32.99	26.63
Noisy Softmax [3]	7.39	28.48	-
L-Softmax [22]	7.58	29.53	-
ACU [13]	7.12	27.47	-
ResNet-110 [9]	-	-	25.16
Densenet-40 [11]	7.00	27.55	24.42
Softmax	7.15	27.7	25.52
NS*[3]	6.91	26.33	25.20
LS*[22]	6.77	26.18	24.32
AS*[21]	6.83	26.09	24.11
Virtual Softmax	***6.68***	***26.02***	***24.01***

Table 4: Recognition error rates on CIFAR datasets. + denotes data augmentation. * denotes our reproducing.

- Performances on large-scale object classification and face verification datasets:

Method	Top1	Top5
Softmax	47.63	73.14
NS*[3]	47.96	73.25
LS*[22]	48.59	73.82
AS*[21]	48.66	73.57
Virtual Softmax	***48.84***	***74.06***

Table 7: Acc (%) on ImageNet32

Softmax	1	99.10	94.59
NS*[3]	1	99.16	94.75
LS*[22]	1	99.37	95.58
AS*[21]+Normface*[33]	1	***99.57***	***96.45***
Virtual Softmax	1	99.46	95.85

Table 6: Verification results (%) on LFW/SLLFW. * denotes our reproducing.

Thanks!

http://www.bhchen.cn