Collaborative Layer-wise Discriminative Learning (CLDL) in Deep Neural Networks

Xiaojie JIN, Yunpeng CHEN, Jian DONG, Jiashi FENG, Shuicheng YAN

1. Motivation

- Intermediate features learned at different layers in a CNN are suitable for discriminating objects of different complexities.
- In the training phase, if a sample has already been correctly classified at a specific layer with high confidence, the rest layers should focus on classifying other difficult samples.

2. Main idea

- First, introduce multiple classifiers on top of multiple layers.
- Second, each classifier coordinates with other classifiers to jointly maximize the final classification performance.

3. CLDL

(A) Cross-layer heterogeneities

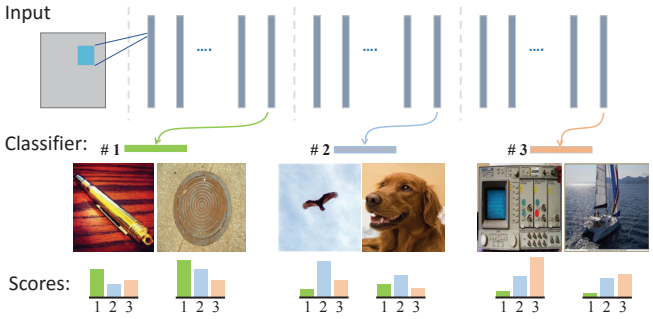
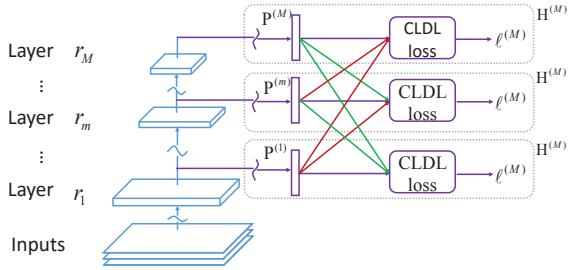


Fig. 1: Three classifiers introduced at bottom/top layers of a deep model can correctly classify simple/complex samples.

(B) Architecture of CLDL-DNN



$H^{(m)}$: Classifier at the top of layer r_m

$P^{(m)}$: Categorical probability scores over all categories, output by $H^{(m)}$

$\ell^{(m)}$: The loss function of $H^{(m)}$

$\mathcal{W}$: All the learnable weights in CLDL

The definition of loss function for each classifier

$$\ell^{(m)}(\mathbf{x}, y^*, \mathcal{W}) = \underbrace{-\log \mathrm{P}^{(m)}(y^*)}_{C^{(m)}} \underbrace{\prod_{t=1, t\neq m}^{M} \left[1 - \mathrm{P}^{(t)}(y^*)\right]^{\frac{1}{M-1}}}_{T^{(m)}}$$

$C^{(m)}$: Conventional cross-entropy loss

$T^{(m)}$: Geometric mean of the prediction scores on target class output by classifiers other than $H^{(m)}$

The overall objective function of CLDL

$$L^{(\mathrm{Net})}(\mathbf{x}, y^*, \mathcal{W}) = \sum_{m=1}^{M} \lambda_m \ell^{(m)} + \alpha \|\mathcal{W}\|_2$$

Inference strategy

$$y^* = \mathrm{argmax}_y L^{(\mathrm{Net})}(\mathbf{x}, y^*, \mathcal{W})$$

(C) Explanation of CLDL

- $T^{(m)}$ measures how well those "companion" classifiers perform on classifying the input sample
- Consider $C^{(m)}$ together with $T^{(m)}$ distinguishes CLDL loss from conventional loss: each classifier considers the performance of other classifier when trying to classify input
- CLDL can be viewed as a simplified version of conditional random field (CRF) model. Please refer to proof in paper.

4. Results

- The effect of classifier number in CLDL on the classification accuracy

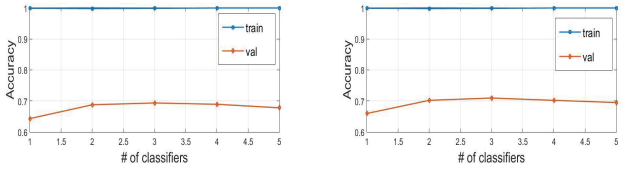


Fig. 2: Evaluation of NIN model on CIFAR-100 with different number of classifiers in CLDL. **Left**: without data augmentation. **Right**: with data augmentation

- Result for object classification
  - For CIFAR-100 and MNIST, NIN are used as baseline
  - For ImageNet, GoogLeNet is used as baseline
  - Three classifiers are used in CLDL

Dataset	Maxout	NIN	DSN	GoogLeNet	CLDL
CIFAR-100	38.57	35.96	34.57	-	**30.40**
CIFAR-100*	-	32.75	-	-	**29.05**
MNIST	0.47	0.42	0.39	-	**0.28**
ImageNet	-	-	-	11.1	**10.21**

*with data augmentation

- Result for scene classification
  - VGGNet with 11/16/11 convolutional layers are used as baseline for MIT67, SUN397, Places205, respectively.
  - Three classifiers are used in CLDL

Dataset	Places [5]	Deep19 [16]	DAG-CNN [16]	Places-AlexNet [5]	Places-CNDS8 [49]
MIT67	54.32	70.80	77.50	68.20	76.10
SUN397	68.24	51.90	56.20	54.30	60.70
Places205	50.00	-	-	80.90	84.10

Dataset	Places-GoogLeNet [48]	Places205-VGG11 [50]	Places205-VGG13 [50]	Places205-VGG16 [5]	CLDL
MIT67	76.30	82.00	81.90	81.20	**84.69**
SUN397	61.10	65.30	66.70	66.90	**70.40**
Places205	85.41	87.60	88.10	88.50	**88.67**