Learning Discriminative Features via Label Consistent Neural Network

Zhuolin Jiang, Yaming Wang, Larry Davis,

Walt Andrews, Viktor Rozgic

03-27-2017

Raytheon BBN Technologies

University of Maryland 18 56

Overview

- Motivations
  - Feature learning of hidden layers receive no direct guidance on class information
  - Early hidden layers of a CNN tend to capture low-level features shared across categories such as edges and corners, while late hidden layers are more class-specific
- Our contributions
  - We propose a supervised feature learning method, ***Label Consistent Neural Network*** **(LCNN)**, which enforces direct supervision in late hidden layers
  - LCNN can learn class-specific neurons or discriminative features

Discriminative Representations

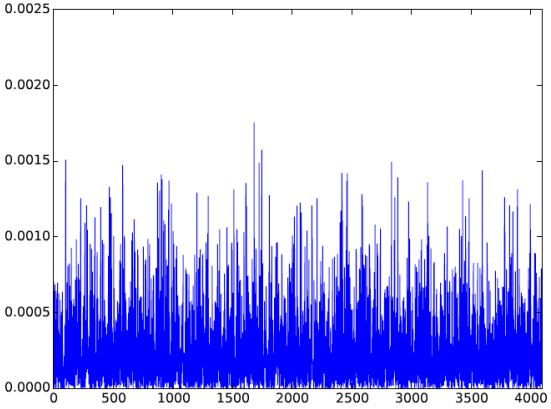
FC6 layer representations from VGGNet-16

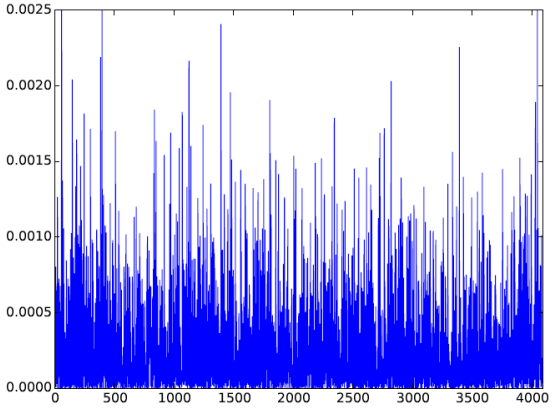
FC7 layer representations from VGGNet-16

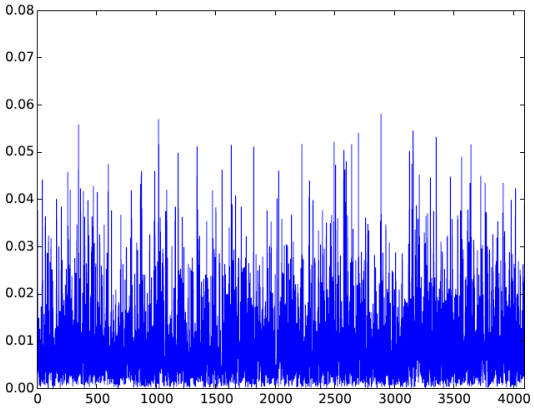
Spatial stream

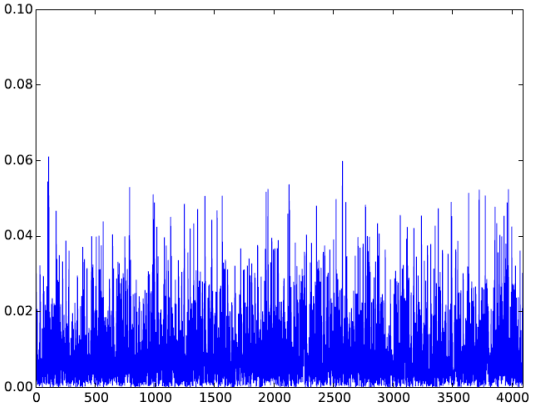
Class 4, baby-crawling, 35 testing videos from UCF101 dataset

Temporal stream







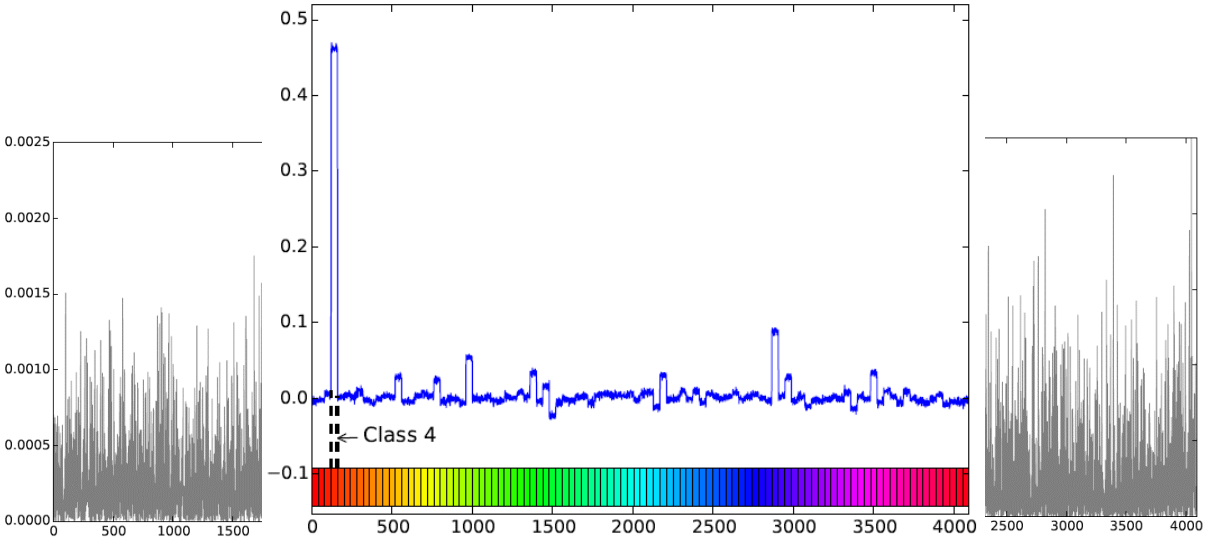


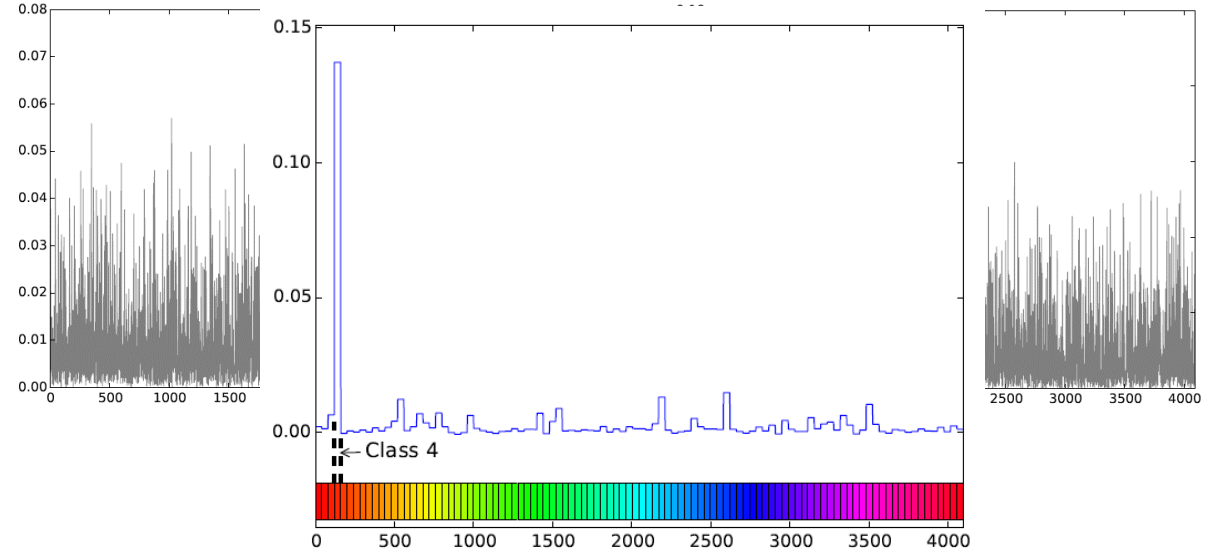
Discriminative Representations

Spatial stream

Class 4, baby-crawling, 35 testing videos from UCF101 dataset

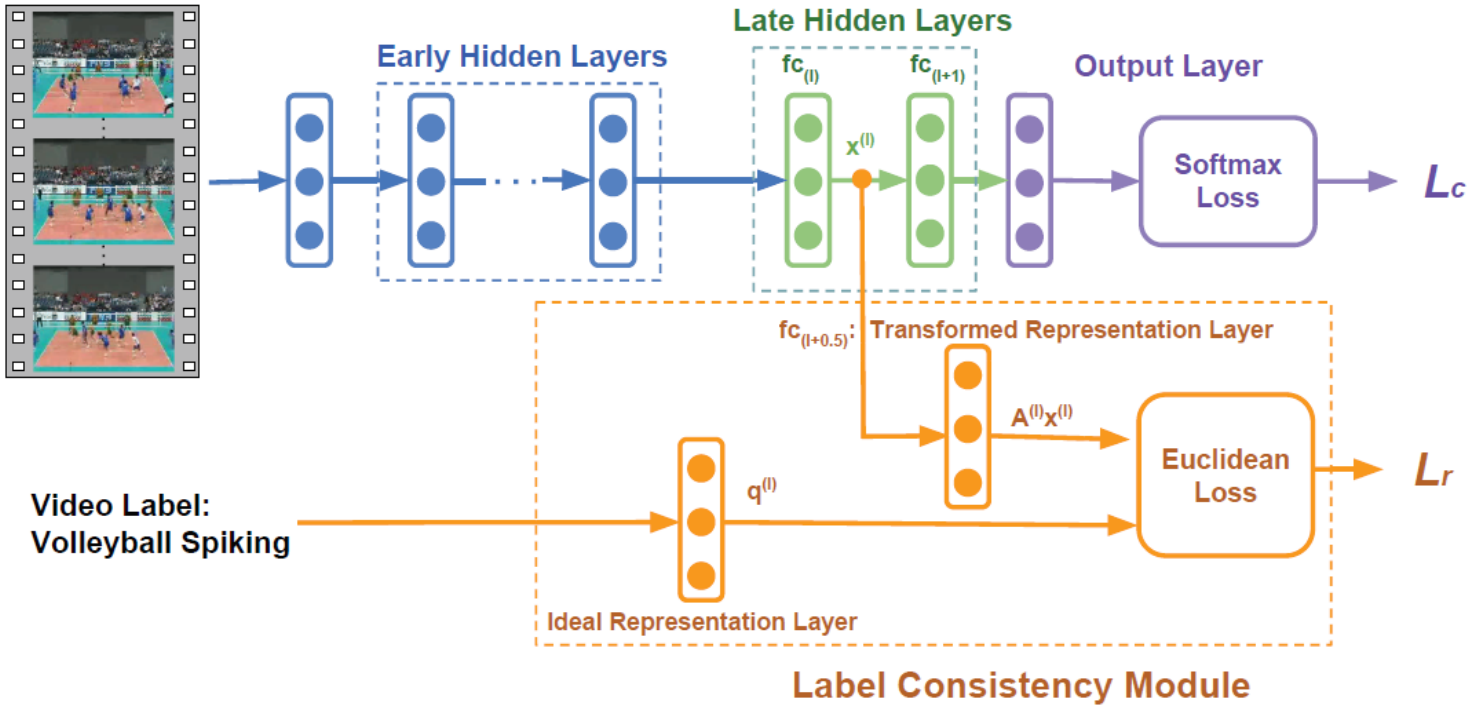
Temporal stream





Label Consistent Neural Network

- Overview
  - Explicit supervision of a hidden layer
  - Encourages the hidden layer representation to approximate “ideal discriminative representation”



- **Overall objective function of LCNN**

$$L = L_c + \alpha L_r$$

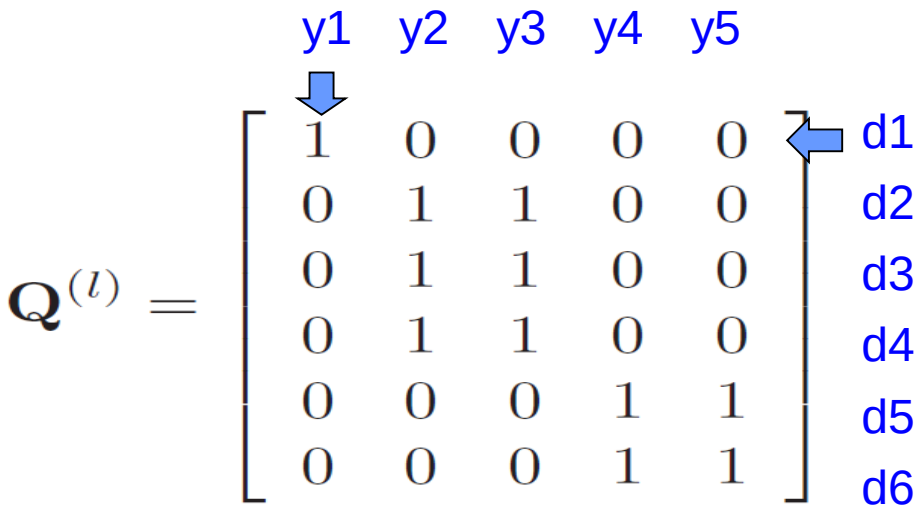
where L_r is the discriminative representation error: ,

$$L_r = L_r(\mathbf{x}^{(l)}, y, \mathbf{A}^{(l)}) = \|\mathbf{q}^{(l)} - \mathbf{A}^{(l)}\mathbf{x}^{(l)}\|_2^2$$

where $q^{(l)}$ is the ideal discriminative representation

- **Ideal discriminative representations**

Given six neurons {$d_1 \ldots d_6$} and five samples {$y_1 \ldots y_5$}



$$\mathbf{Q}^{(l)} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$

Experiments

- UCF-101

Network Architecture	Spatial	Temporal	Both
ClarifaiNet [28]	72.7	81	87
VGGNet-19 [41]	75.7	78.3	86.7
VGGNet-16 [36]	79.8	85.7	90.9
VGGNet-16* [36]	-	85.2	-
baseline	77.48	83.71	-
LCNN-1	80.1	85.59	89.87
LCNN-2 (argmax)	80.7	85.57	91.12
LCNN-2 (k-NN)	81.3	85.77	89.84

- THUMOS15

Network Architecture	Spatial	Temporal	Both
VGGNet-16 [36]	54.5	42.6	-
ClarifaiNet [28]	42.3	47	-
GoogLeNet [32]	53.7	39.9	-
baseline	55.8	41.8	-
LCNN-1	56.9	45.1	59.8
LCNN-2 (argmax)	57.3	44.9	61.7
LCNN-2 (k-NN)	58.6	45.9	62.6

- Cifar-10

Method (Without Data Augment.)	Test Error (%)
Stochastic Pooling [42]	15.13
Maxout Networks [7]	11.68
DSN [21]	9.78
baseline	10.41
LCNN-2 (argmax)	9.75

Method (With Data Augment.)	Test Error (%)
Maxout Networks [7]	9.38
DropConnect [33]	9.32
DSN [21]	8.22
baseline	8.81
LCNN-2 (argmax)	8.14

- Caltech-101

Method	Accuracy(%)
LC-KSVD [14]	73.6
Zeiler [43]	86.5
Dosovitskiy [4]	85.5
Zhou [45]	87.2
He [9]	91.44
baseline	87.1
LCNN-1 (k-NN)	88.51
LCNN-2 (argmax)	90.11
LCNN-2 (k-NN)	89.45
baseline*	92.5
LCNN-2* (argmax)	93.7
LCNN-2* (k-NN)	93.6

Thank you!