More Correlations Better Performance： Fully Associative Networks for Multi-label Image Classification

YaNing Li

Liu Yang

1 Background

2 Related work

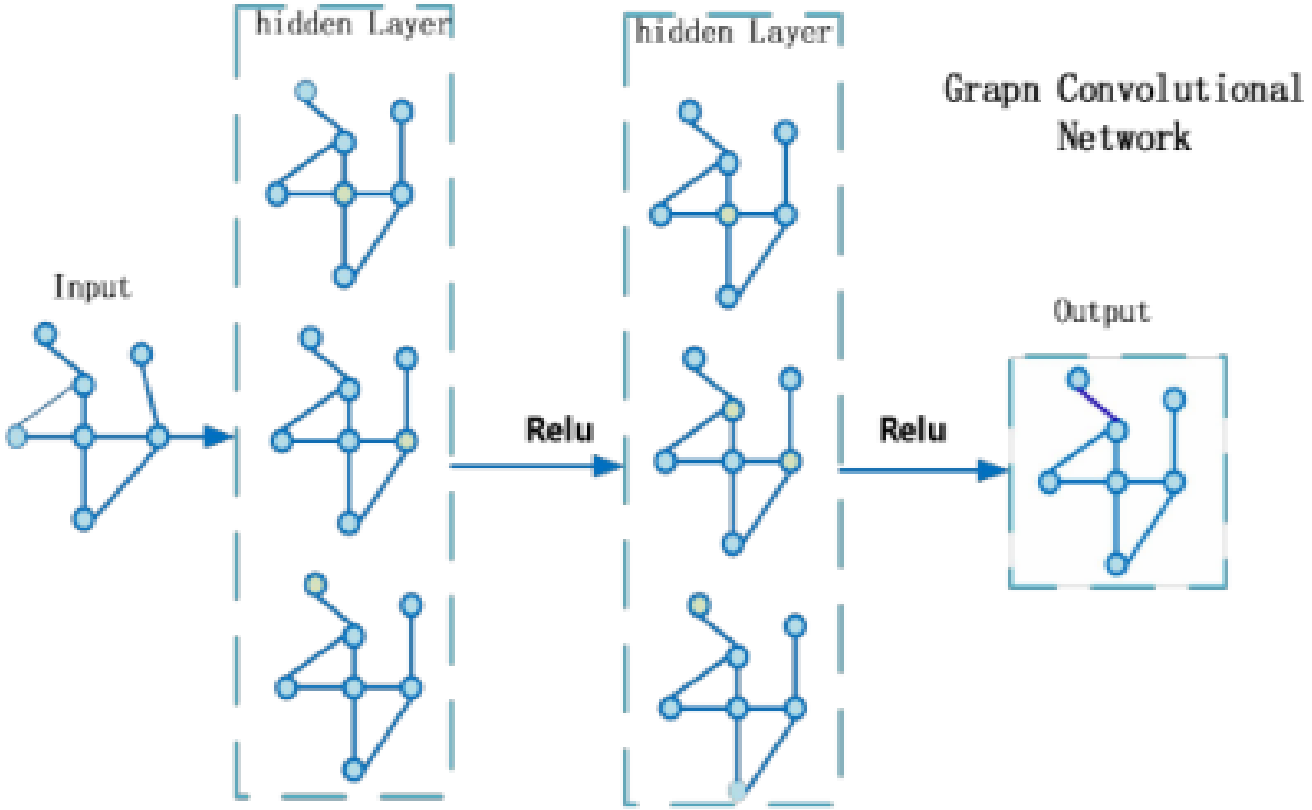
3 Model

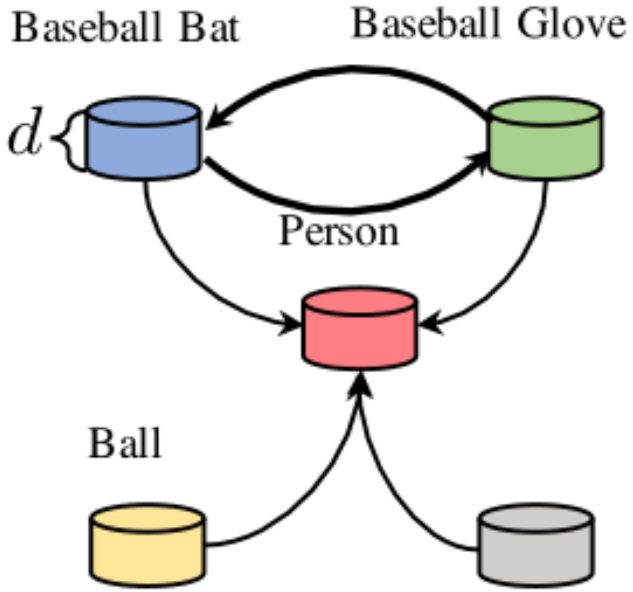
4 Experiment

1 Background

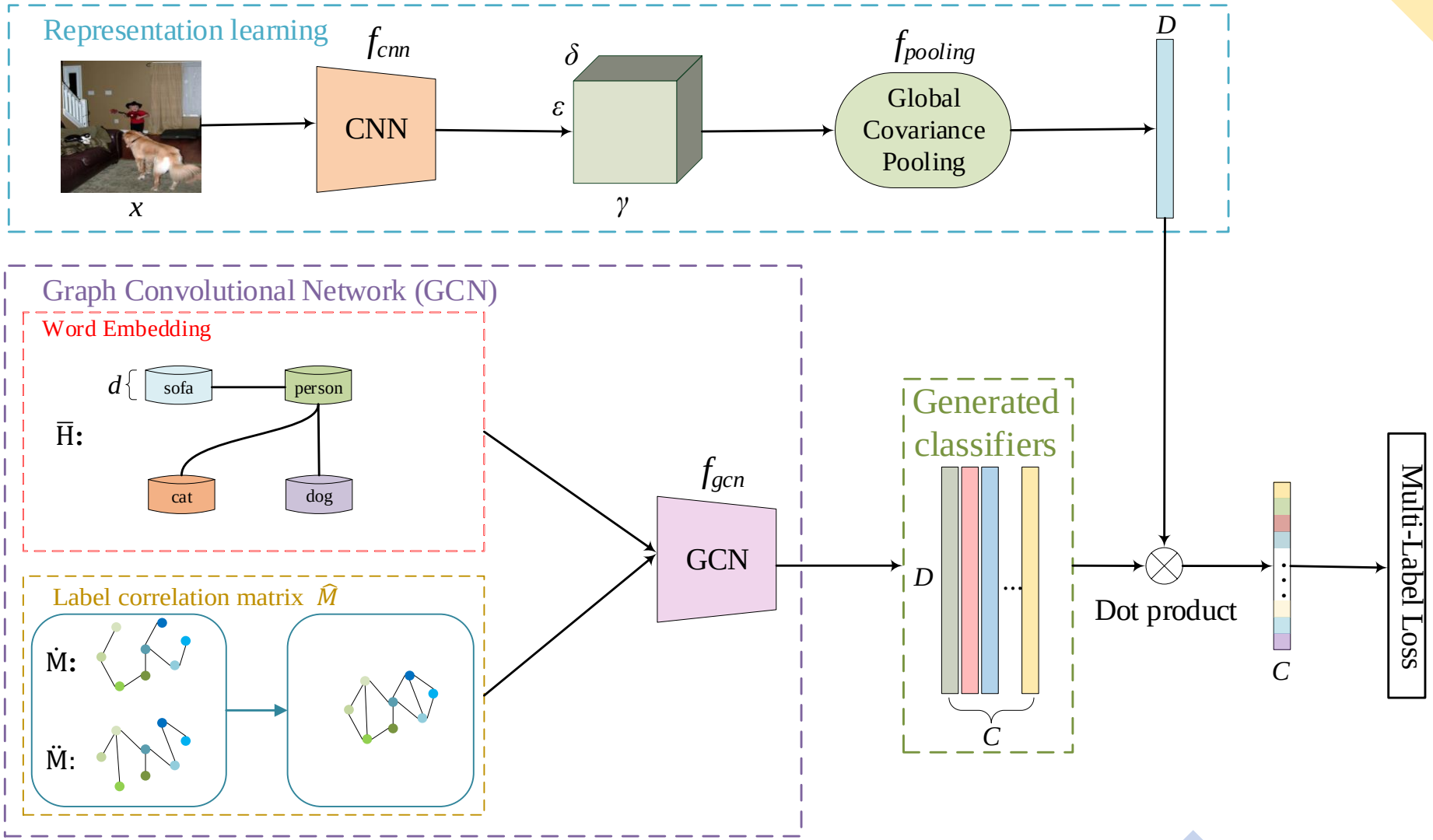
- Video classification
- Image annotation
- Medical diagnosis

2 Related work

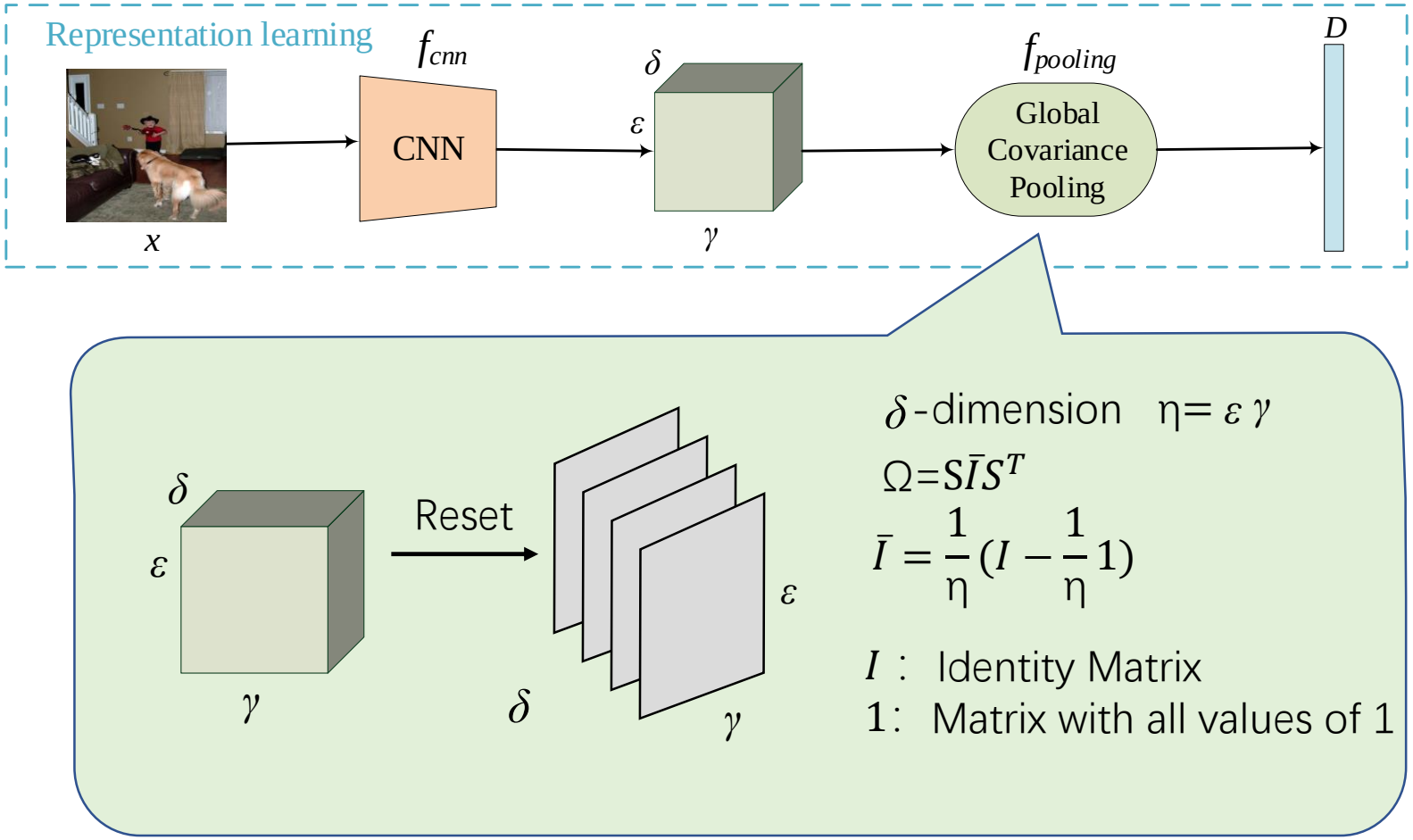




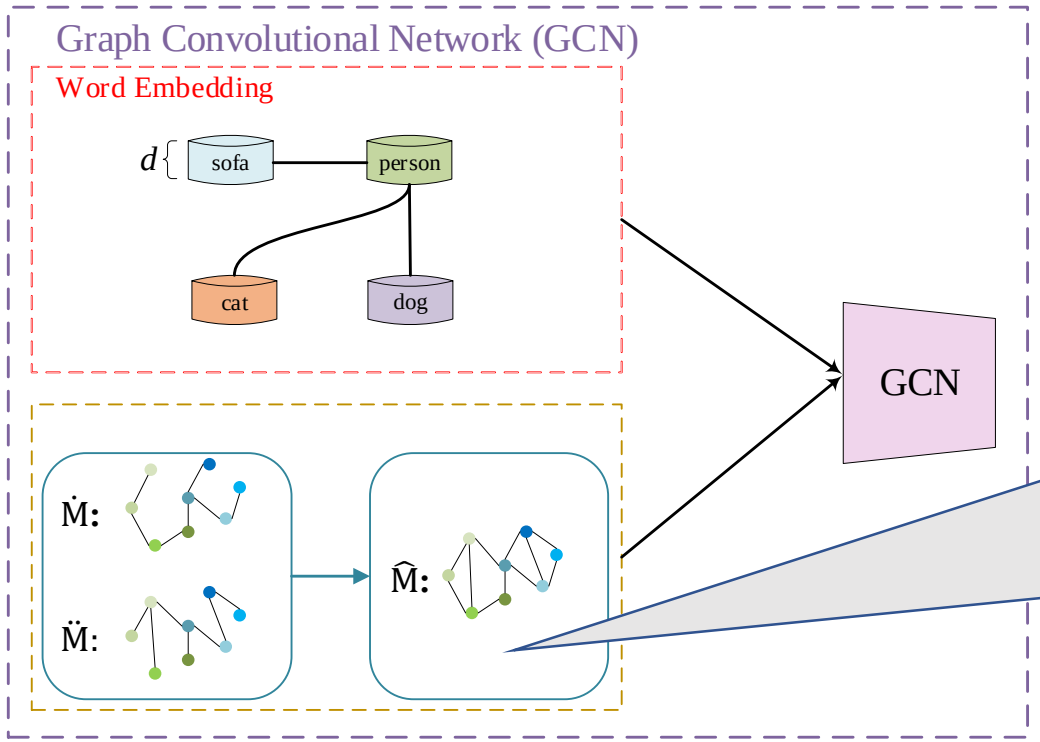
3 Model


Representation learning
f_{cnn}
CNN
x
δ
ε
γ
$f_{pooling}$
Global Covariance Pooling
D
Graph Convolutional Network (GCN)
Word Embedding
d
sofa
person
cat
dog
$\bar{H}$:
Label correlation matrix $\hat{M}$
$\dot{M}$:
$\ddot{M}$:
f_{gcn}
GCN
Generated classifiers
D
C
Dot product
C
Multi-Label Loss

3 Model


Representation learning
f_{cnn}
CNN
δ
ε
γ
$f_{pooling}$
Global Covariance Pooling
D
x
Reset
δ-dimension $\eta = \varepsilon\gamma$
$\Omega = S\bar{I}S^T$
$\bar{I} = \frac{1}{\eta}(I - \frac{1}{\eta}1)$
I : Identity Matrix
1: Matrix with all values of 1

3 Model

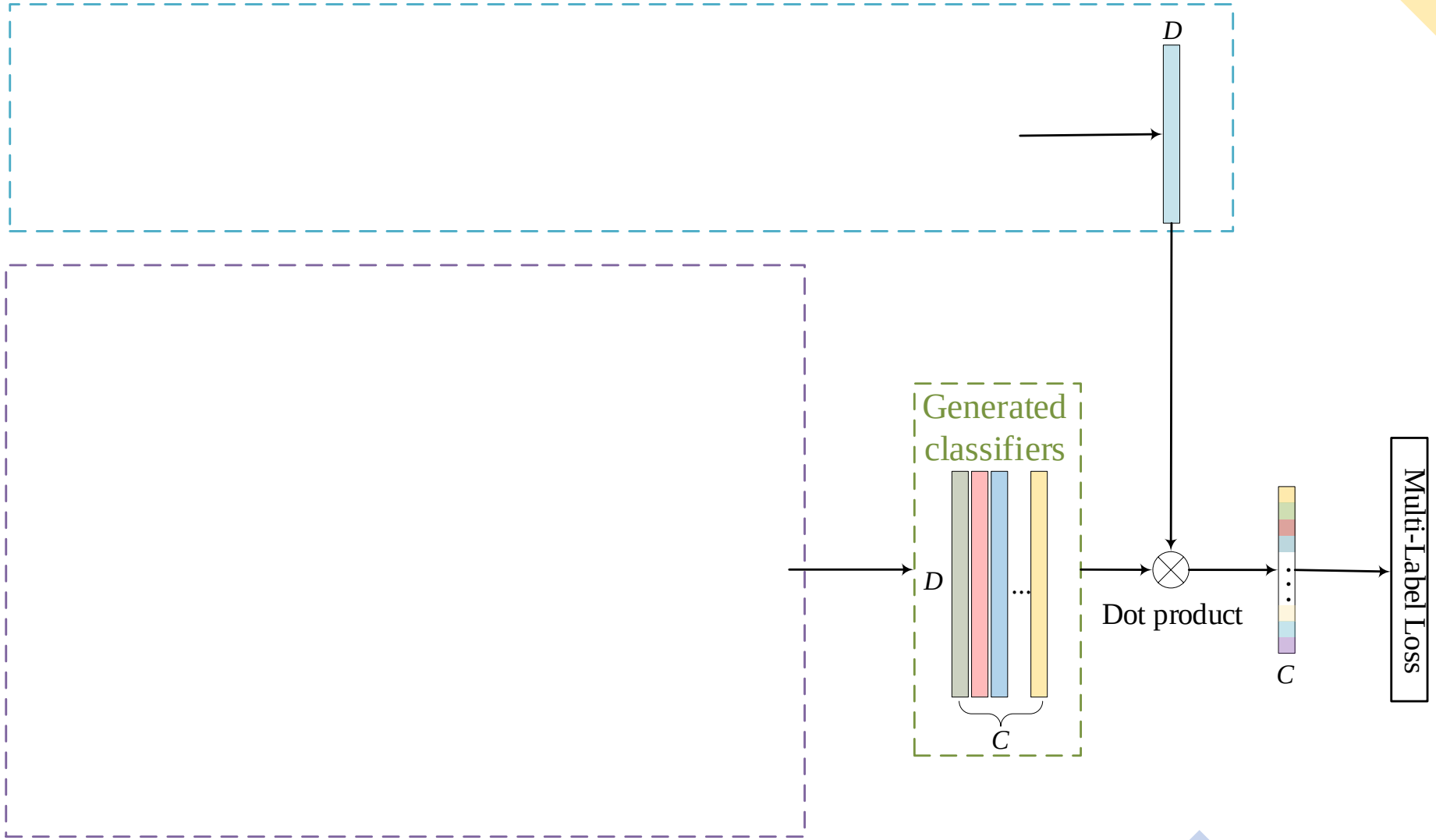


$\dot{M}$: The statistical pairs of training labels appear in the matrix

$$\ddot{M} = Y^T \tilde{X} \tilde{X} Y$$

$$\tilde{M} = \alpha \dot{M} + (1 - \alpha) \ddot{M}$$

3 Model



4 Experiment

Datasets

- VOC 2007 samples 9963 label 20
- MS-COCO samples 123287 label 80

Comparisons

- ResNet-101
- Vgg16
- HCP
- Order-Free
- SRN
- RNN-Attention
- RLSD
- ML-GCN
- CNN-RNN

Metrics

- Loss
- Accuracy
- F1-score
- Precison
- Recall

4 Experiment

Comparisons (in %)of different methods in terms of ap and map on VOC 2007 dataset.

Methods	aero	bike	bird	boat	bottle	bus	car	cat	chair	cow	table	dog	horse	motor	person	plant	sheep	sofa	train	tv	mAP↑
CNN-RNN	96.7	83.1	94.2	92.8	61.2	82.1	89.1	94.2	64.2	83.6	70.0	92.4	91.7	84.2	93.7	59.8	93.2	75.3	99.7	78.6	84.0
RLSD	96.4	92.7	93.8	94.1	71.2	92.5	94.2	95.7	74.3	90.0	74.2	95.4	96.2	92.1	97.9	66.9	93.5	73.7	97.5	87.6	88.5
VeryDeep	98.9	95.0	96.8	95.4	69.7	90.4	93.5	96.0	74.2	86.6	87.8	96.0	96.3	93.1	97.2	70.0	92.1	80.3	98.1	87.0	89.7
HCP	98.6	97.1	98.0	95.6	75.3	94.7	95.8	97.3	73.1	90.2	80.0	97.3	96.1	94.9	96.3	78.3	94.7	76.2	97.9	91.5	90.9
RNN-Attention	98.6	97.4	96.3	96.2	75.2	92.4	96.5	97.1	76.5	92.0	87.7	96.8	97.5	93.8	98.5	81.6	93.7	82.8	98.6	89.3	91.9
ML-GCN	98.6	94.6	97.7	**97.1**	**84.9**	89.5	96.0	96.8	86.1	95.4	87.9	**97.3**	**98.2**	94.0	98.3	86.2	91.9	87.6	**97.6**	89.9	93.2
FAN (Ours)	**99.5**	**95.2**	**98.6**	96.8	84.7	**96.8**	**97.8**	**98.6**	**87.1**	**97.2**	**90.1**	95.8	97.6	**97.1**	**98.7**	**90.6**	**93.4**	**90.1**	97.5	**92.1**	**94.8**

Comparisons of state-of-the art methods on MS-COCO dataset.

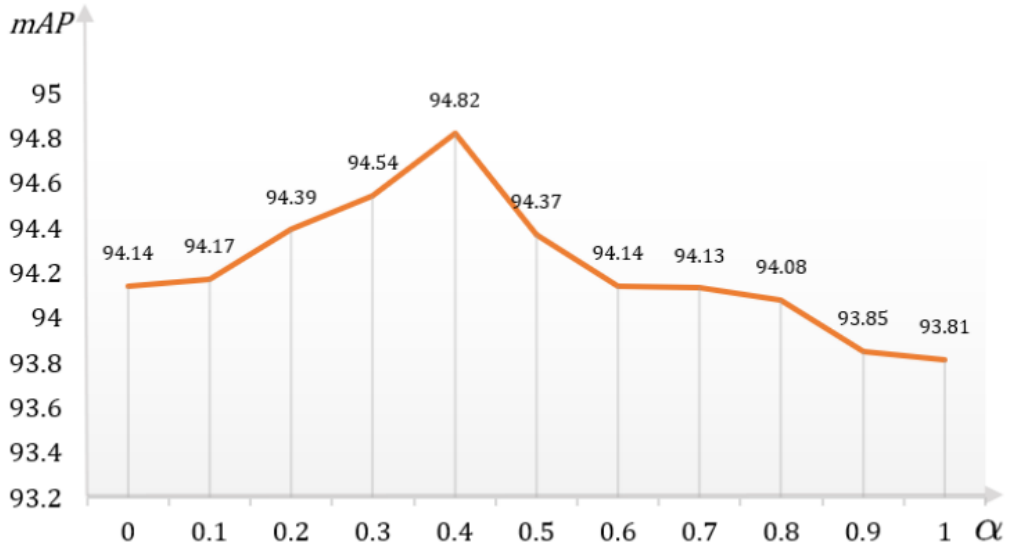
Methods	All		Top-3					
	loss↓	mAP↑	CP↑	CR↑	CF1↑	OP↑	OR↑	OF1↑
ResNet-101	-	77.3	84.1	59.4	69.7	89.1	62.8	73.6
CNN-RNN	-	61.2	66.0	55.6	60.4	69.2	66.4	67.8
RNN-Attention	-	-	79.1	58.7	67.4	84.0	63.0	72.0
Order-Free RNN	-	-	71.6	54.8	62.1	74.2	62.2	67.7
SRN	-	77.1	85.5	58.8	67.4	87.4	62.5	75.2
ML-GCN	0.0812	82.1	**87.5**	64.2	73.4	83.1	60.7	69.5
FAN (Ours)	**0.0807**	**83.2**	85.1	**64.6**	**73.4**	**90.5**	**66.8**	**76.8**

4 Experiment

Comparisons loss, map and outher evaluation metrics with ml-gcn on voc 2007

Methods	All							
	loss↓	mAP ↑	CP↑	CR ↑	CF1↑	OP↑	OR↑	OF1↑
ML-GCN	0.0543	93.2	88.1	84.8	86.4	89.1	85.6	87.9
PAN	0.0537	93.3	**88.2**	83.3	85.8	89.0	86.1	87.5
FAN	**0.0462**	**94.8**	86.9	**88.4**	**86.9**	**90.0**	**86.8**	**88.4**

Different α in $\tilde{M} = \alpha\dot{M} + (1-\alpha)\ddot{M}$



4 Experiment

Different GCN architectures on voc 2007

Architectures of GCN	VOC 2007				
	All			Top-3	
	mAP↑	CF1↑	OF1↑	CF1↑	OF1↑
2-layer 300→2048→8256	**94.8**	**86.9**	**88.4**	**86.7**	**88.3**
3-layer 300→1024→2048→8256	93.2	85.8	87.0	85.8	87.1
4-layer 300→1024→2048→4096→8256	93.1	86.1	86.5	85.8	86.6

Different strategies of dimensionality reduction on voc 2007

Dimensionality Reduction	VOC 2007					
	All				Top-3	
	loss↓	mAP↑	CF1↑	OF1↑	CF1 ↑	OF1↑
2048 → 128	0.0576	93.2	83.6	85.7	83.5	85.7
2048 → 512 → 128	**0.0462**	**94.8**	**86.9**	**88.4**	**86.7**	**88.3**

Thank you !