Collaborative Networks for Person Verification

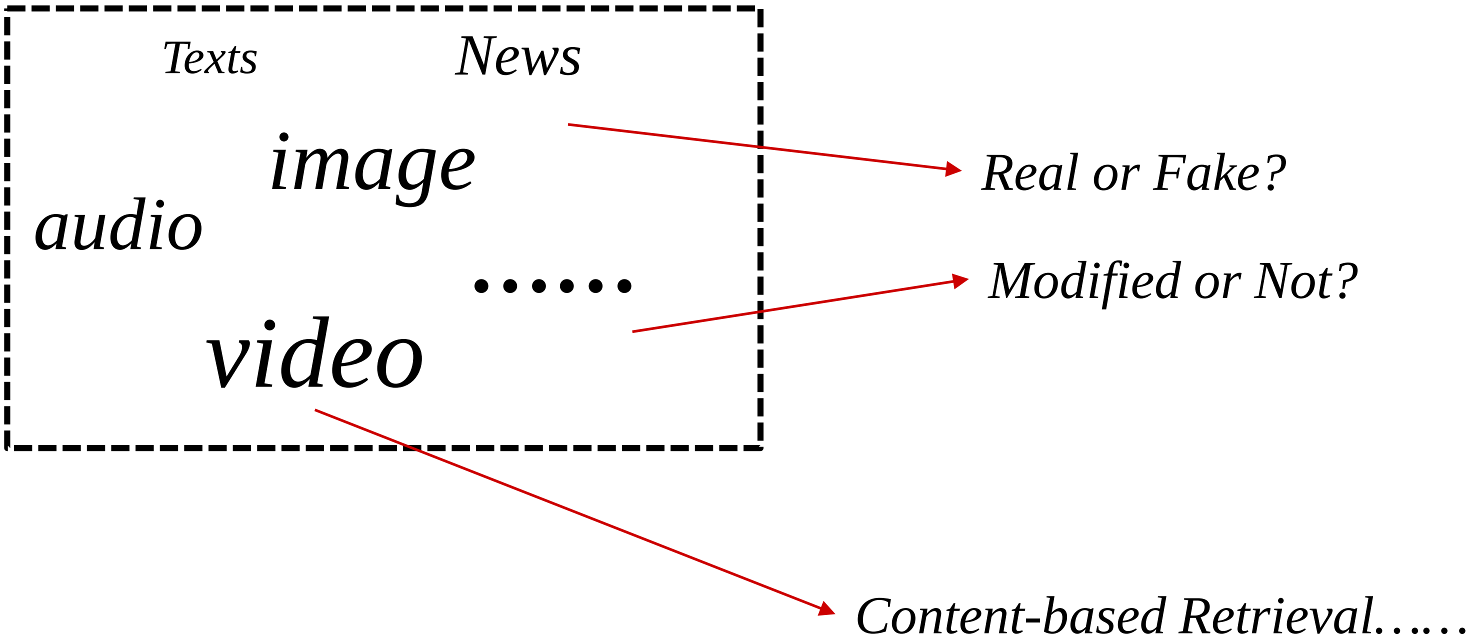
School of Electronic and Computer Engineering, Peking University
Yihao Zhang, Wenmin Wang*, Jinzhuo Wang

Overview

- Introduction
- The proposed method
  - Pipeline
  - Architecture
  - Improved Siamese Network (iSN)
  - Deep Discriminative Network (DDN)
  - Collaborative Learning
- Experiments
- Results

Definition

Multimedia Verification



Multimedia Verification

Texts *News*

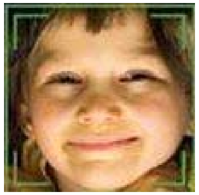
image

audio

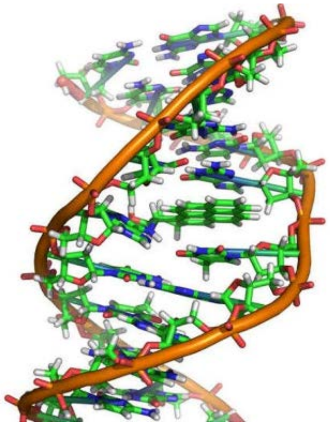
……

video

Person Verification



……



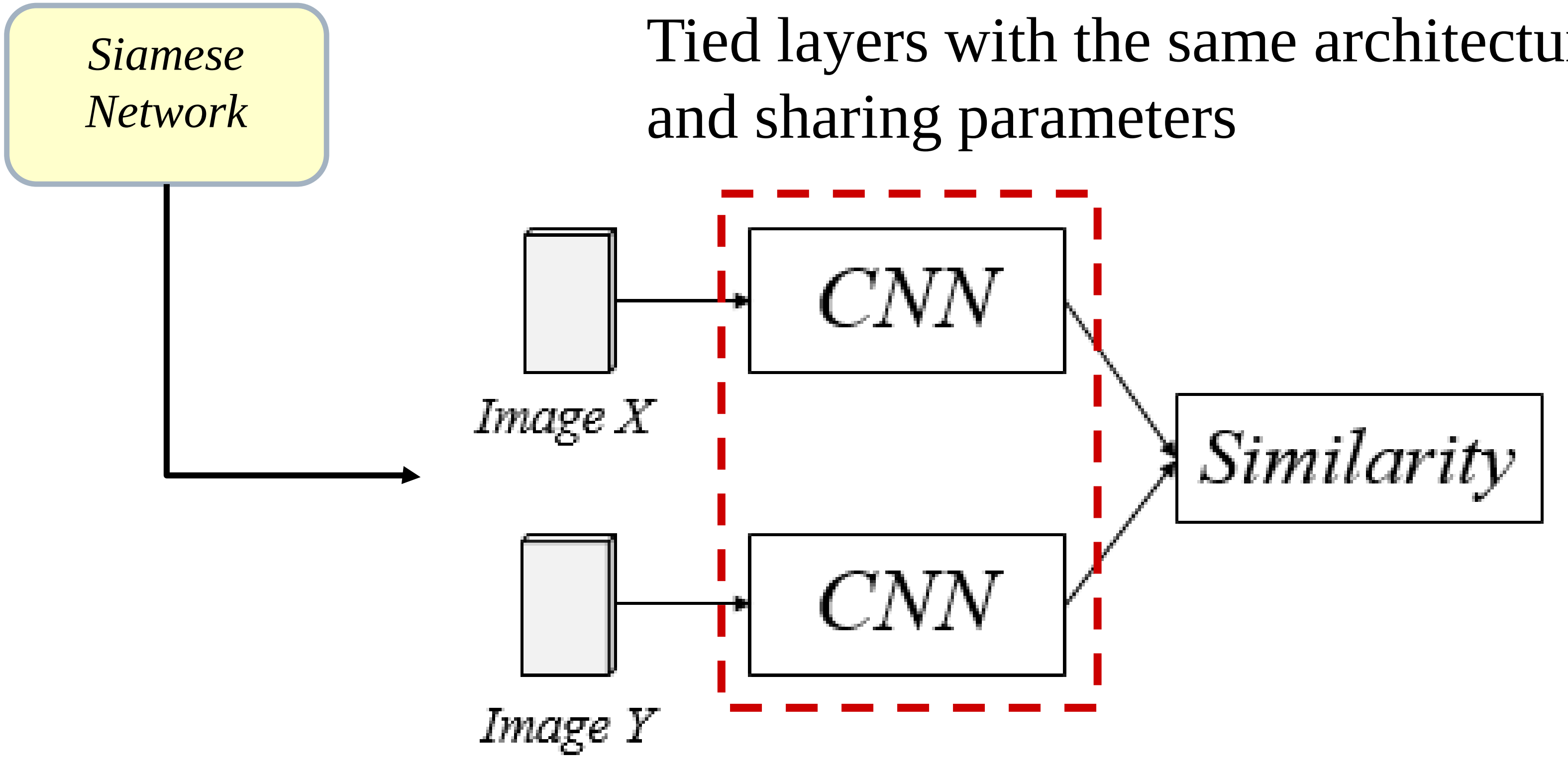
Challenges

- Biometrics cues:
  - face, fingerprint, iris, DNA, etc.
- Different settings of camera **parameters**
- **Illumination** changes across different cameras and capturing time
- **Alignment** problems caused by the posture and shooting angle
- Body **occlusion** and background **cluttering**
- **Similarity** across different persons
- ……

Background

1) Handcrafted feature + metric learning
2) Deep Methods …

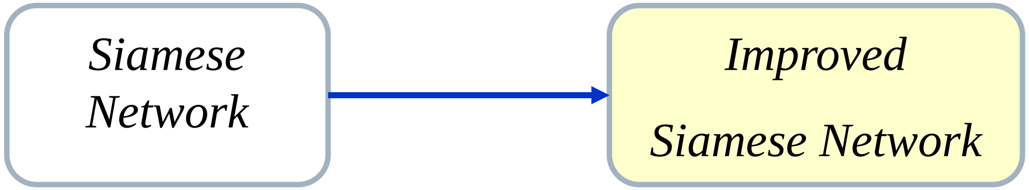
Tied layers with the same architecture and sharing parameters



① The prevalent framework which we based

[11] Y. Dong, L. Zhen, S. Liao, and Stan Z. Li. 2014. Deep Metric Learning for Person Re-identification. In International Conference on Pattern Recognition. 34-39.

The Proposed Method



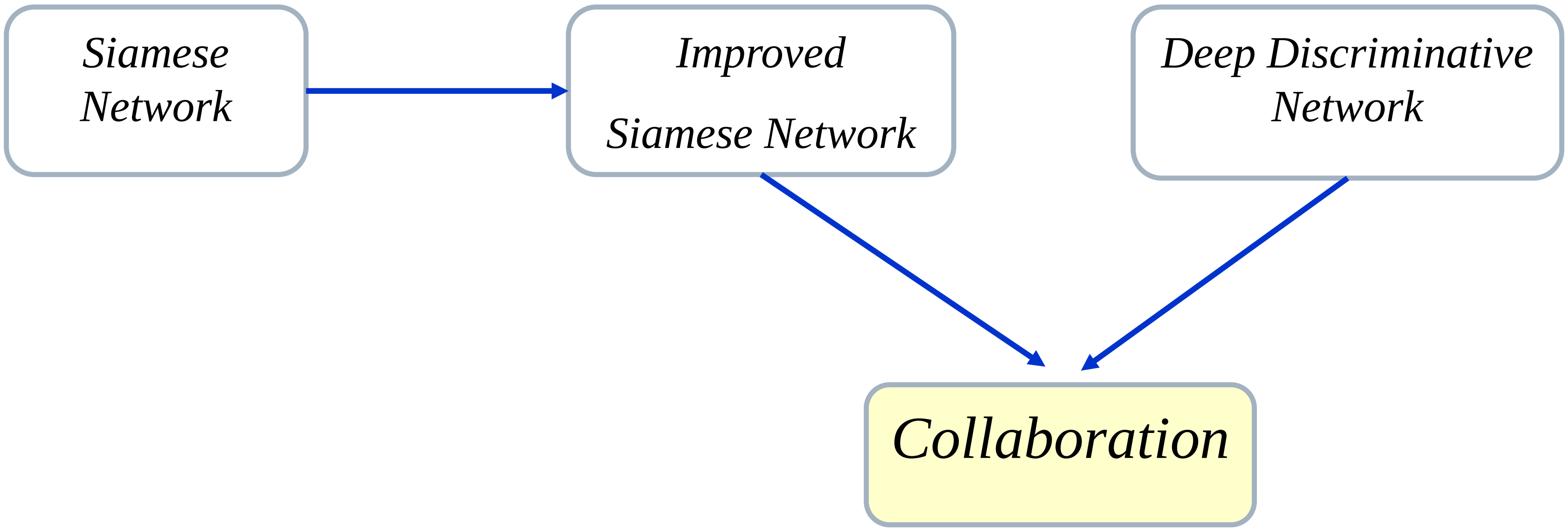
① The prevalent framework which we based

② We improve the Siamese Network

The Proposed Method

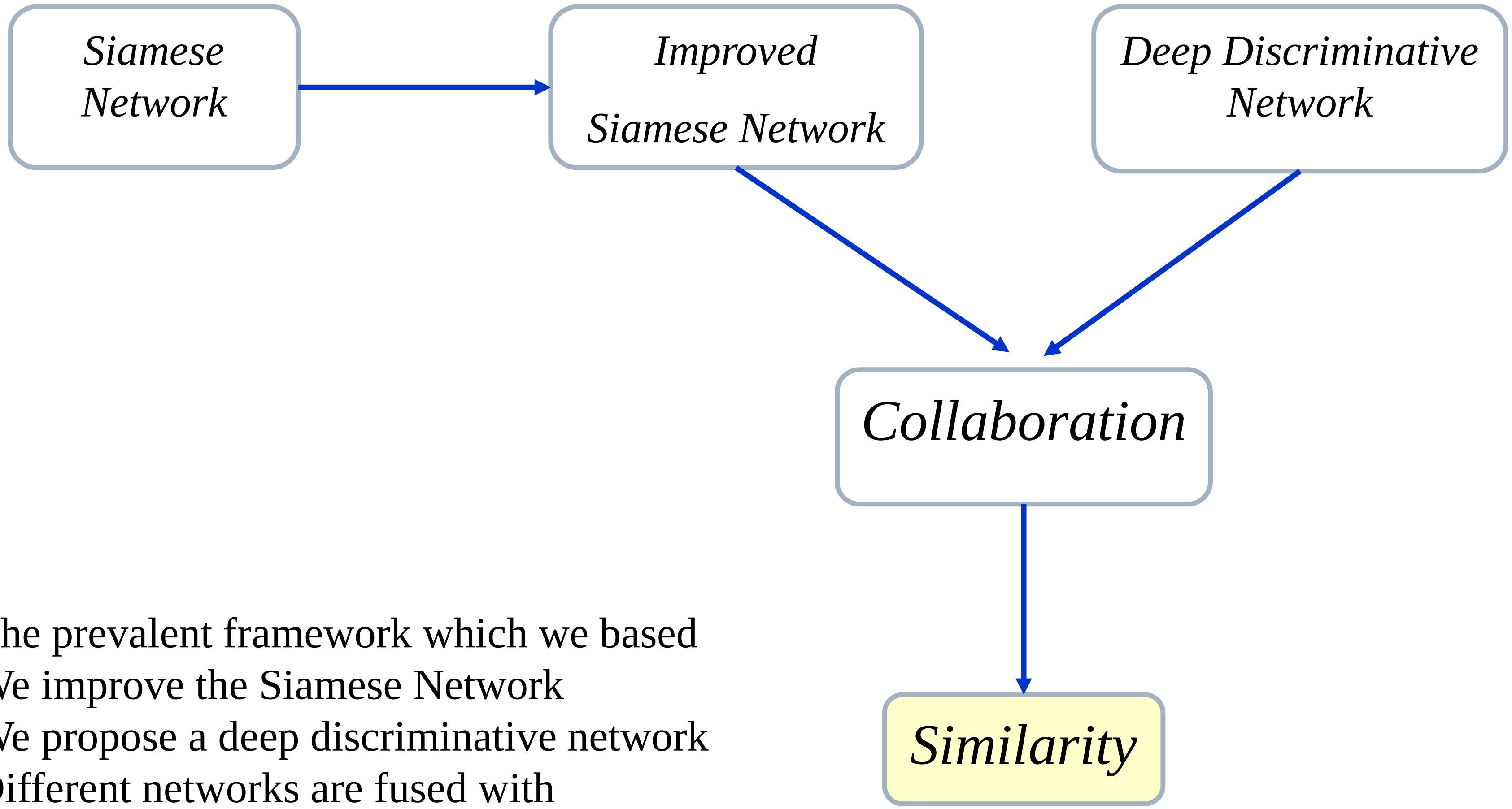
① The prevalent framework which we based
② We improve the Siamese Network
③ We propose a deep discriminative network

The Proposed Method



① The prevalent framework which we based
② We improve the Siamese Network
③ We propose a deep discriminative network
④ Different networks are fused with collaborative strategy

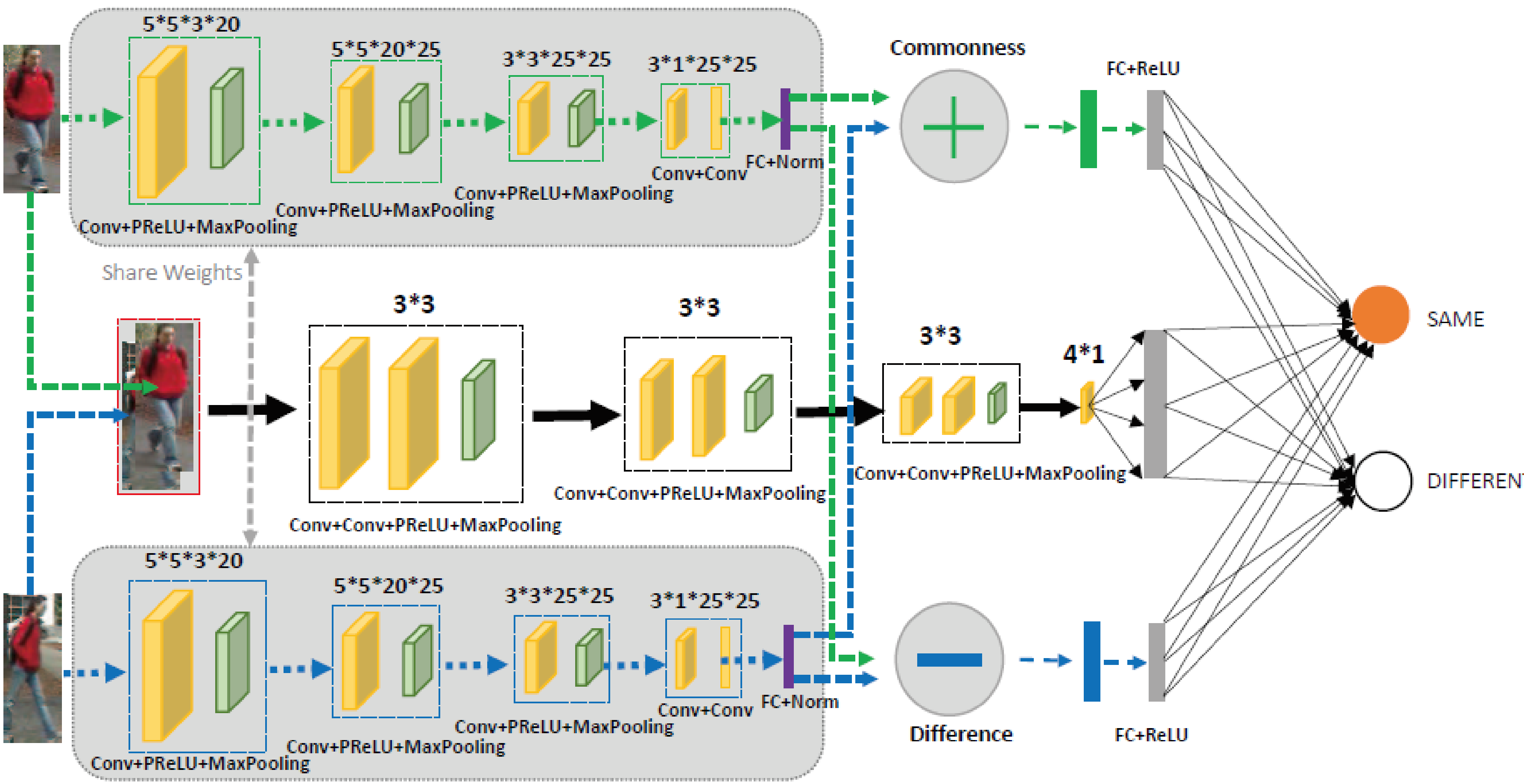
The Proposed Method



① The prevalent framework which we based
② We improve the Siamese Network
③ We propose a deep discriminative network
④ Different networks are fused with collaborative strategy

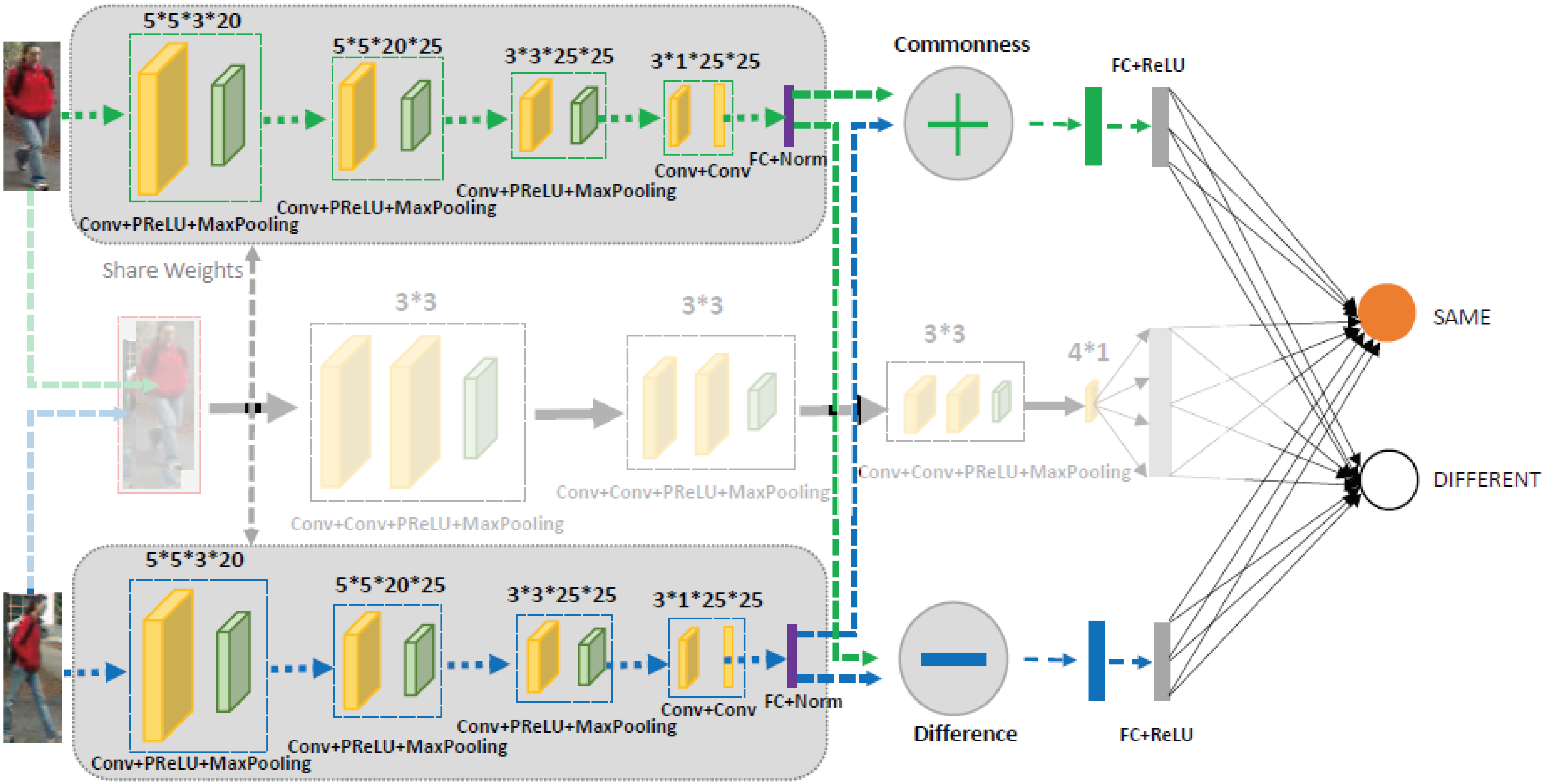
The Framework

Collaborative Networks for Person Verification



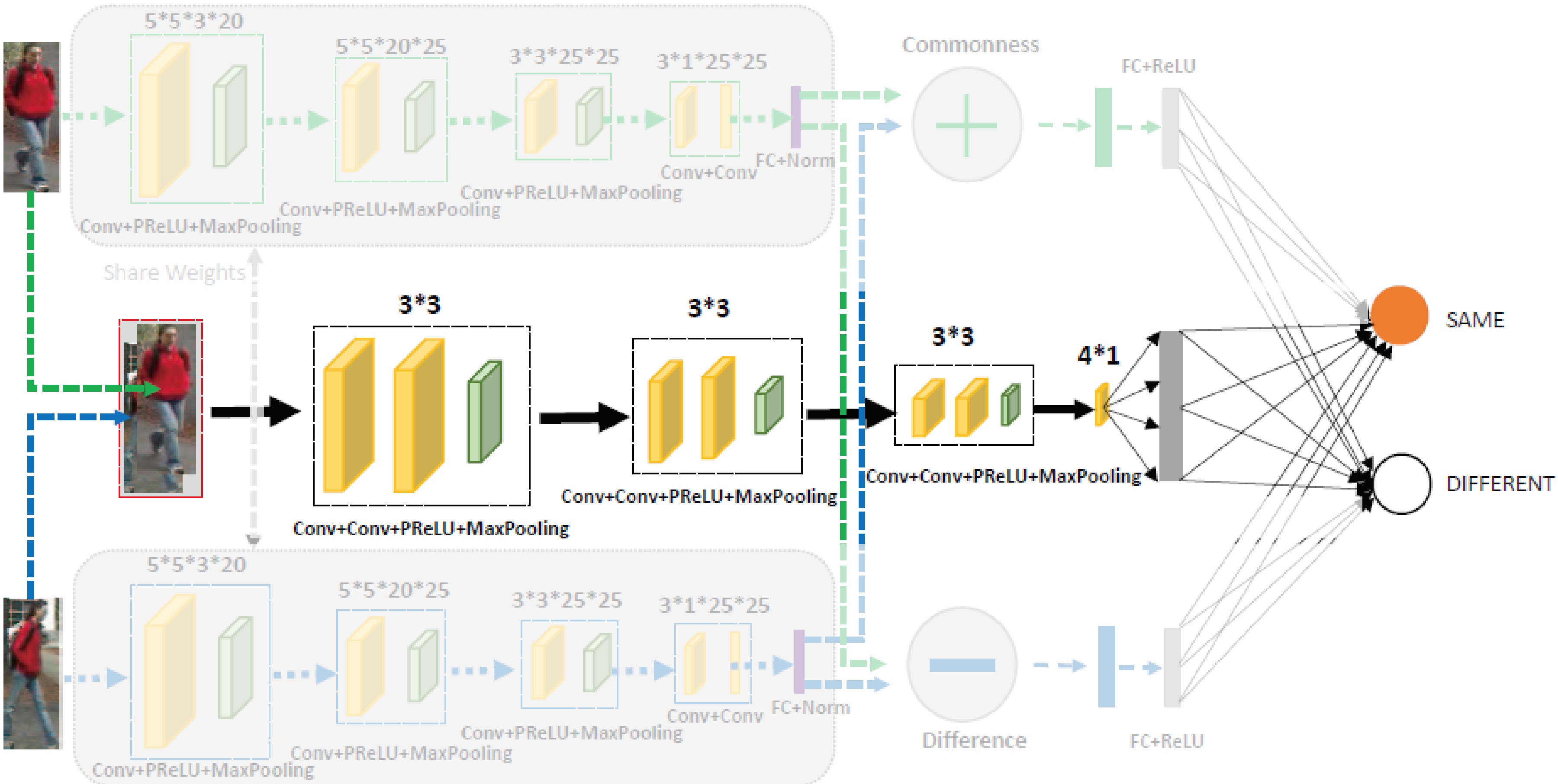
The Framework

improved Siamese Network (iSN)



The Framework

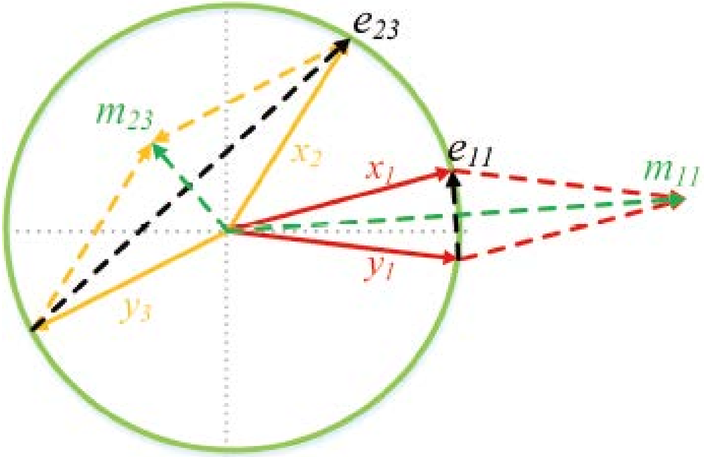
Deep Discriminative Network (DDN)



improved Siamese Network (iSN)

- For similar vectors $\overrightarrow{x_1}$ and $\overrightarrow{y_1}$
  - $|\overrightarrow{e_{11}}| = |\overrightarrow{x_1} - \overrightarrow{y_1}|$ is supposed to be small
  - $|\overrightarrow{m_{11}}| = |\overrightarrow{x_1} + \overrightarrow{y_1}|$ is supposed to be large

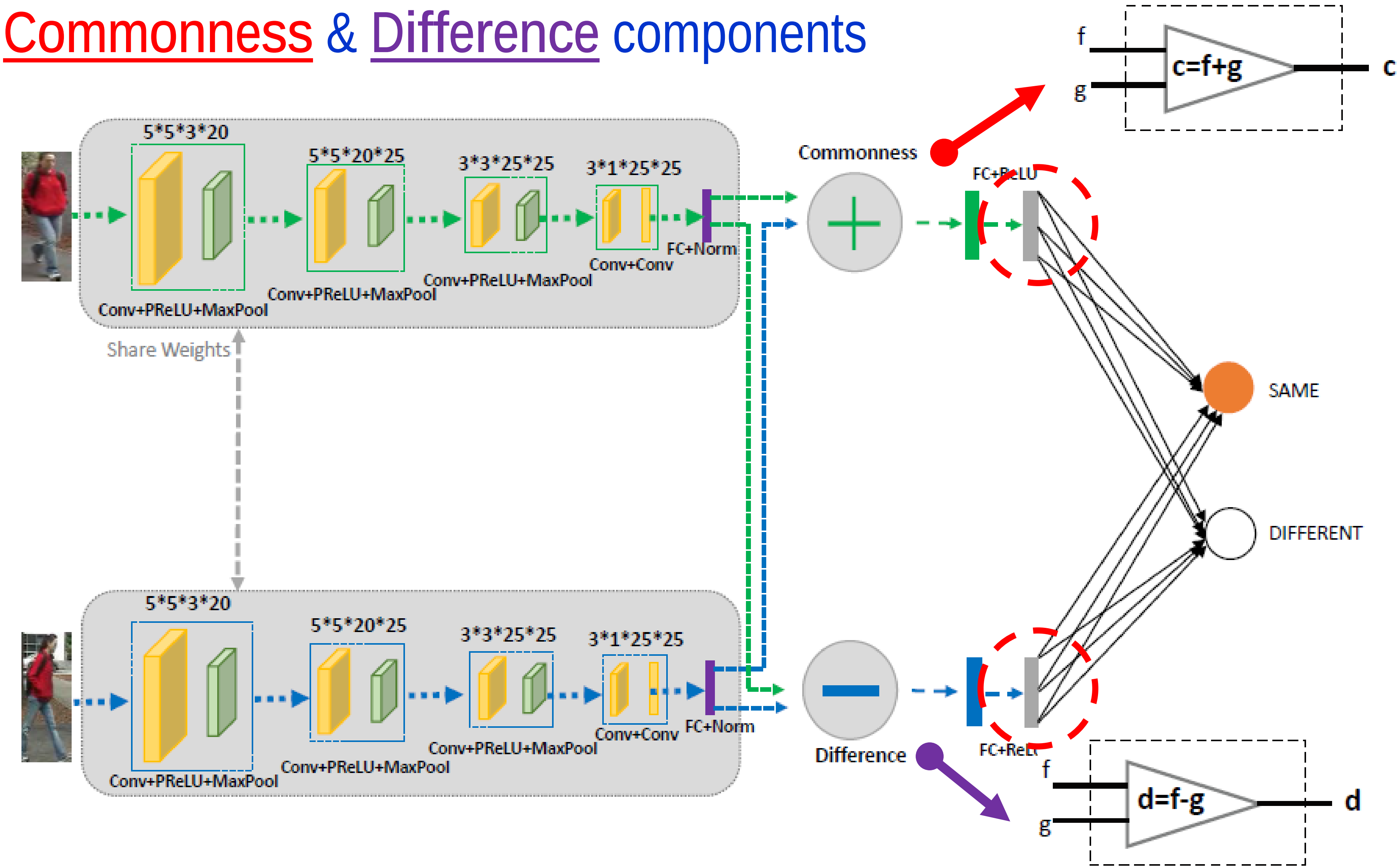
- For dissimilar vectors $\overrightarrow{x_2}$ and $\overrightarrow{y_3}$
  - $|\overrightarrow{e_{23}}| = |\overrightarrow{x_2} - \overrightarrow{y_3}|$ is supposed to be large
  - $|\overrightarrow{m_{23}}| = |\overrightarrow{x_2} + \overrightarrow{y_3}|$ is supposed to be small



[41] Y. Yang, S. Liao, Z. Lei, and Stan Z Li. 2016. Large scale similarity learning using similar pairs for person verification. In AAAI. 3655-3661.

improved Siamese Network (iSN)

Commonness & Difference components



improved Siamese Network (iSN)

The effect of summary layer

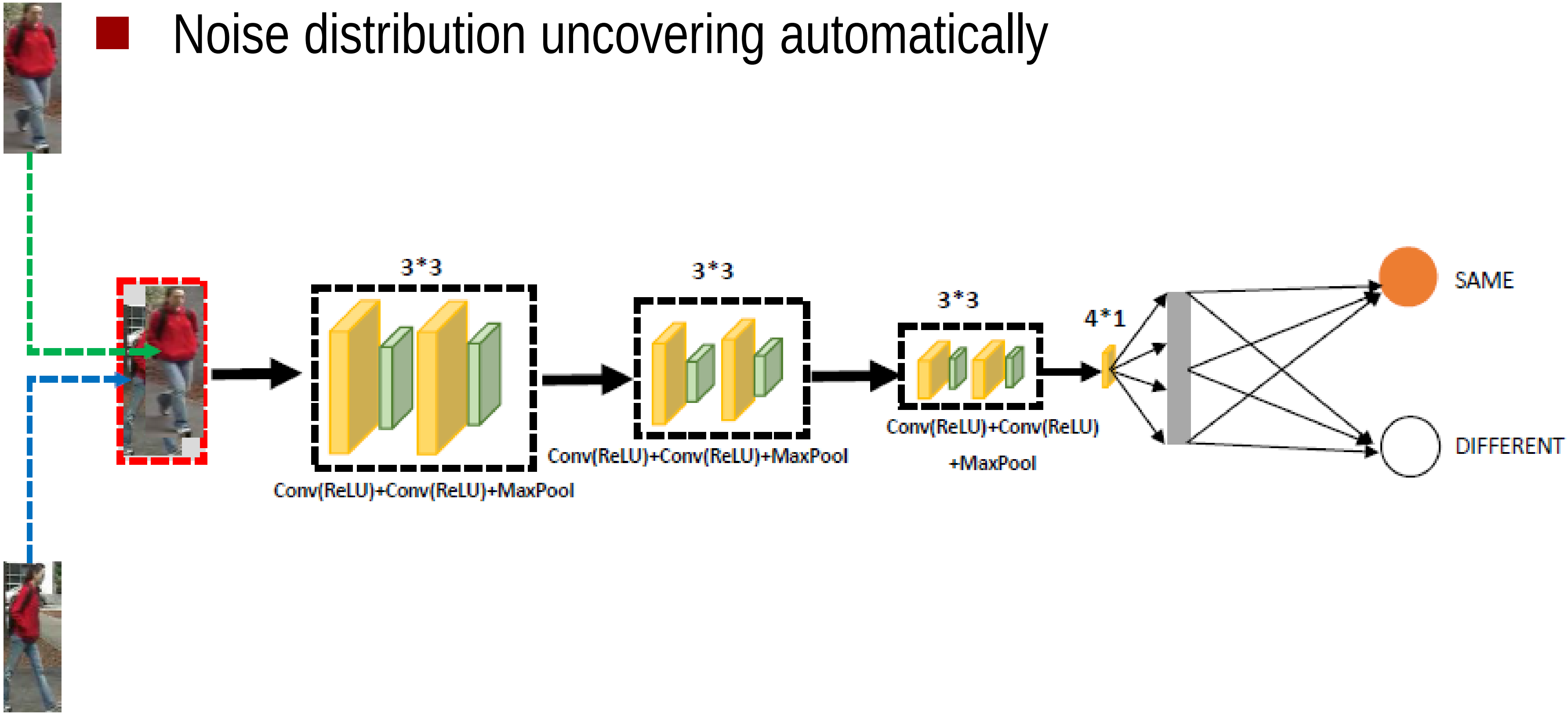
Table 1: The performance (%) of iSN with and without summary layer after the *commonness* and *difference* components on QMUL GRID dataset.

Type	rank 1	rank 5	rank 10	rank 20
without summary	0.80	4.00	8.00	16.00
with summary	**16.00**	**36.00**	**51.20**	**69.60**

Deep Discriminative Network (DDN)

Motivation

- Original difference space
- No feature extraction of individual image
- Noise distribution uncovering automatically

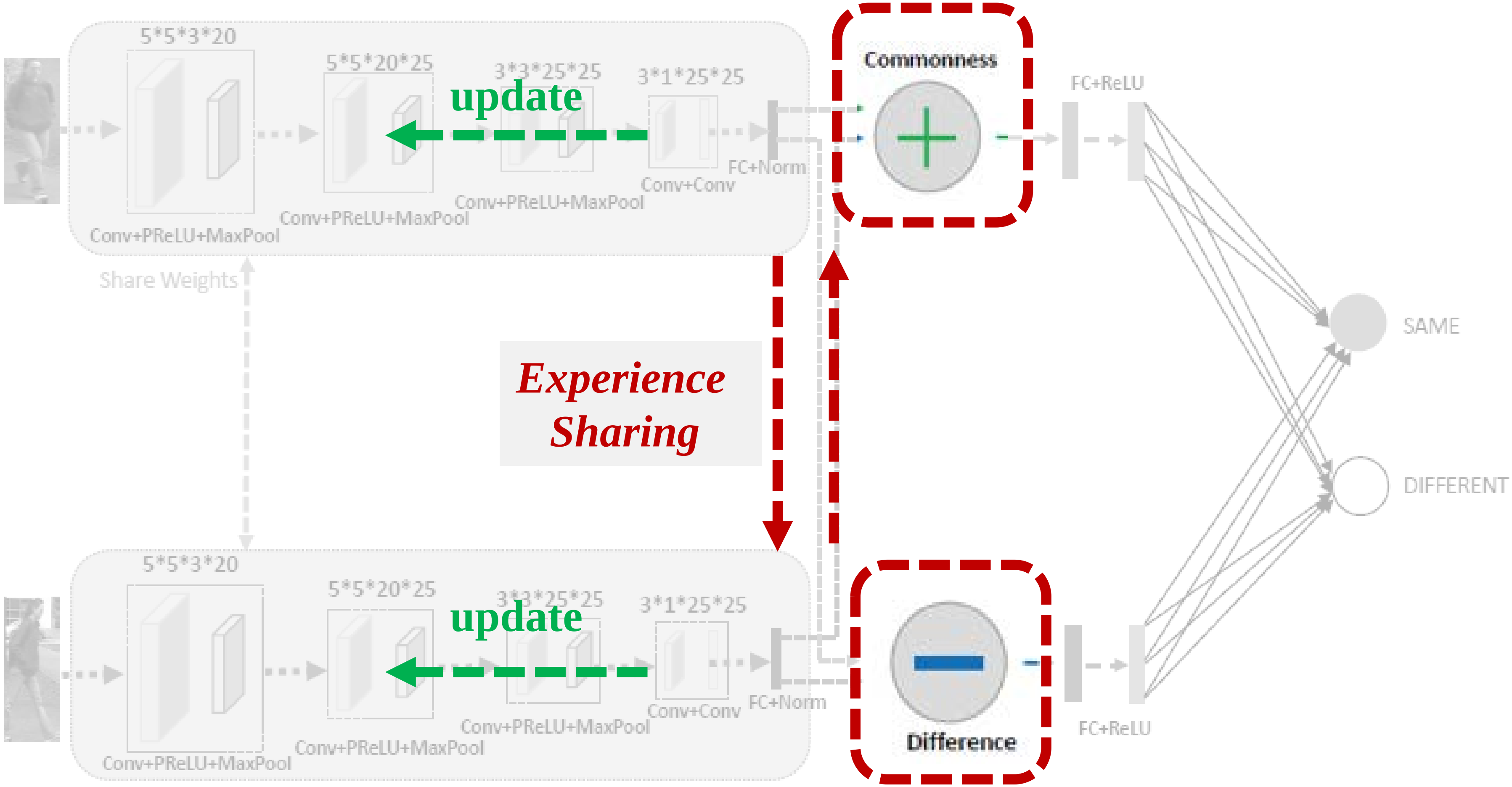


Collaborative Learning

- Agent Definition
  - Independent and complete network
- To improve: Competition or Cooperation?
- Collaborative learning strategy:
  - Agents interacting with each other by sharing different experience
    - *Commonness and difference components*

Collaborative Learning

Experience sharing

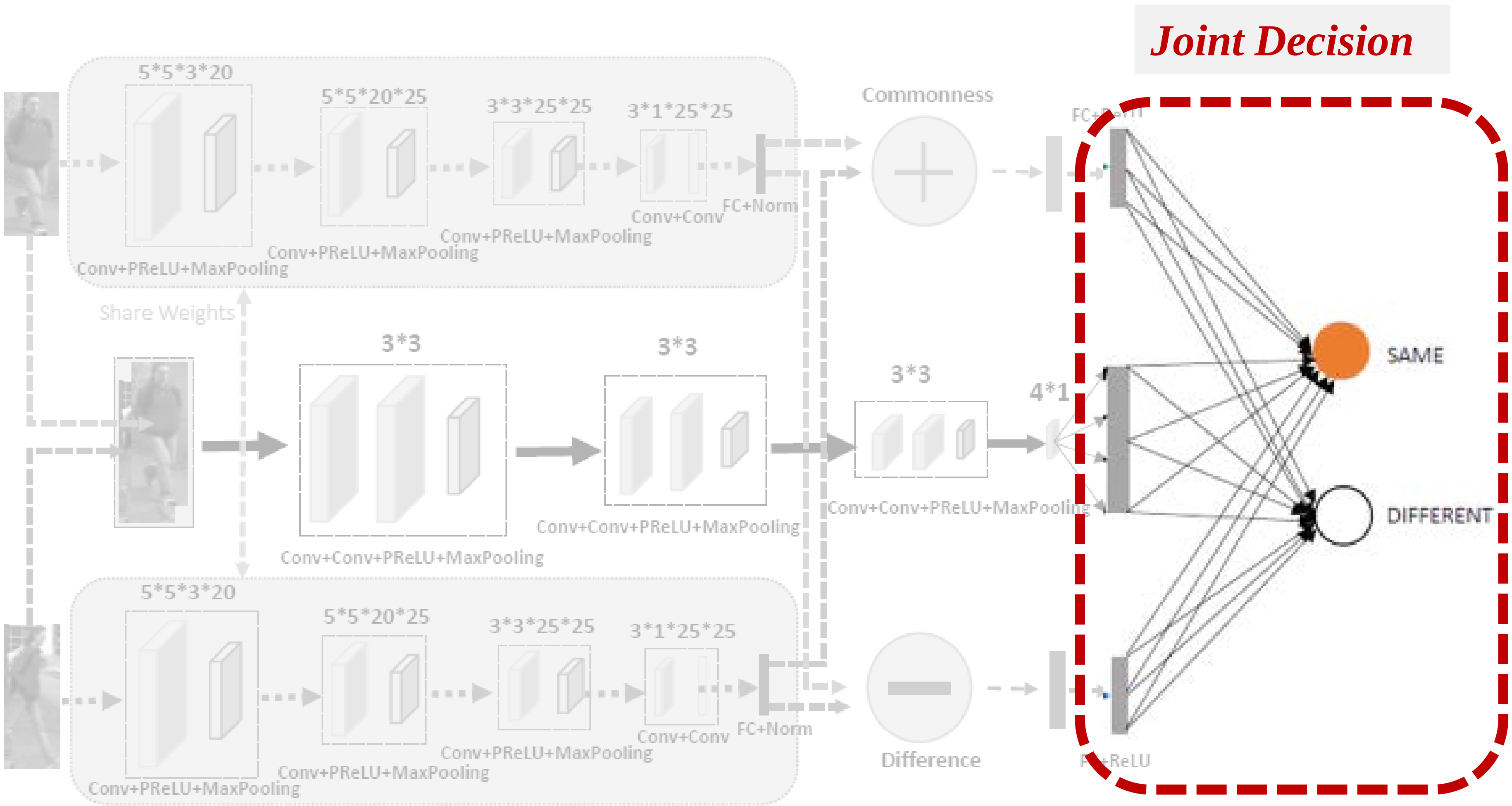


Collaborative Learning

- Agent Definition
  - Independent and complete network
- To improve: Competition or Cooperation?
- Collaborative learning strategy:
  - Agents interacting with each other by sharing different experience
    - *Commonness and difference components*
  - Learn mistakes from another agents
    - *Joint decision of the probability*

Collaborative Learning

Learning mistakes from one another



Experiments

- Evaluation on four datasets
  - CUHK01, CUHK03, PRID2011, QMUL GRID
- Implementation Details
  - Images resized to 160*60
  - Mini-batches of 128
  - Learning rate:0.05
  - Learning rate decay: 0.0001
  - Momentum: 0.9
  - Weight decay: 0.0005
  - Implemented on Torch7

Results

Datasets: CUHK03

- Total pedestrians: 1360
- Total images: 13164
- Source: 6 different surveillance cameras
- Training person#:1260
- Testing person#:100
- Additional gallery: NO

Method	Rank 1	Rank 5	Rank 10	Rank 20
eSDC [44]	8.76	24.07	38.28	53.44
LDML [13]	13.51	40.73	52.13	70.81
KISSME [15]	14.17	48.54	52.57	70.53
FPNN [21]	20.65	51.50	66.50	80.00
LOMO+XQDA [22]	52.20	82.23	92.14	96.25
IDLA [1]	54.74	86.50	93.88	98.10
LOMO+MLAPG [23]	57.96	87.09	94.74	98.00
ensembles [30]	62.10	89.10	94.30	97.80
PersonNet [38]	64.80	89.40	94.92	98.20
NullReid [43]	62.55	90.05	94.80	98.10
MTDnet [5]	**74.68**	**95.99**	**97.47**	–
Ours	62.63	91.63	97.00	**99.65**

Results

Datasets: QMUL GRID

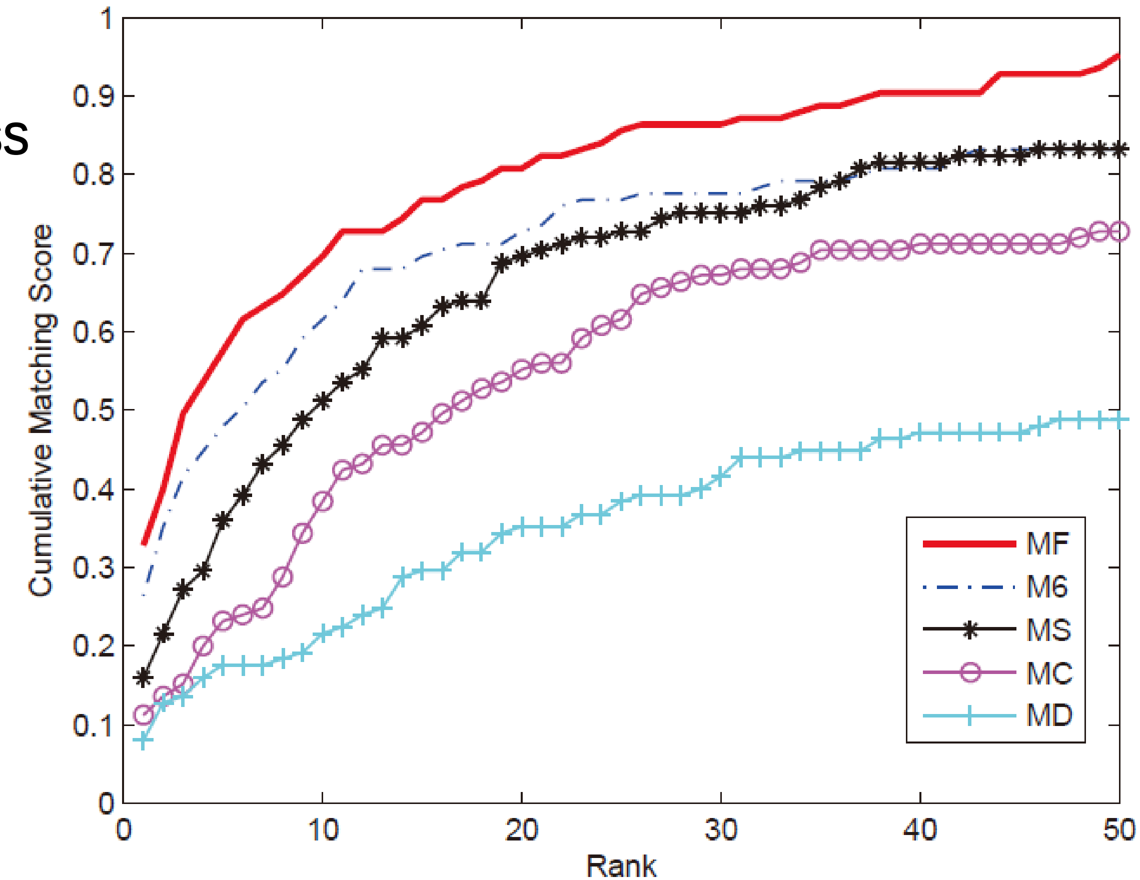
- Total pedestrians: 1150
- Total images: 1400
- Source: 8 different surveillance cameras
- Training person#:125
- Testing person#:125+775
- Additional gallery: 775

Method	Rank 1	Rank 5	Rank 10	Rank 20
LOMO+XQDA [22]	16.56	33.84	41.84	47.68
KEPLER [28]	18.40	39.12	50.24	57.04
Norm X-Corr [33]	19.20	38.40	53.60	66.40
NLML [17]	24.54	35.86	43.53	–
SSDAL+XQDA [32]	22.40	39.20	48.00	–
DR-KISS [34]	20.60	39.30	51.40	–
SCSP [3]	24.24	44.56	54.08	59.68
Ours	**32.80**	**57.60**	**69.60**	**80.80**

Results

Ablation Experiments

- MF: the complete framework
- M6: only with DDN
- MS: only with **iSN**
- MC: only with **commonness** component in iSN
- MD: only with **difference** component in iSN



Results

Datasets: PRID2011

- Total pedestrians: 934
- Total images: 1134
- Source: 2 static cameras
- Training person#:100
- Testing person#:100+549
- Additional gallery: 549

Method	Rank 1	Rank 5	Rank 10	Rank 20
ITML [9]	12.00	–	36.00	47.00
KISSME [15]	15.00	–	39.00	52.00
kLFDA [39]	22.40	46.60	58.10	–
DML [11]	17.90	37.50	45.90	–
NullReid [43]	29.80	52.90	66.00	76.50
Ensembles [30]	17.90	40.00	50.00	62.00
ImpTrpLoss [7]	22.00	–	47.00	57.00
MTDnet [5]	32.00	51.00	62.00	–
Ours	**43.00**	**76.00**	**85.00**	**95.00**

Results

Matching results on PRID2011

Conclusion

- First, we propose an improved siamese network for person verification, which is composed by commonness and difference components;
- Second, we propose a deep discriminative network for person verification, which learns how to discriminate the difference in the original difference space;
- Third, a collaborative learning strategy is put forward and different networks are fused into a unified architecture;
- Finally, experiments are conducted on several datasets and our method achieves competitive or superior performance compared to the state-of-the-art methods.

Q&A