ICPR 2016

23rd International Conference on Pattern Recognition

WeAT4.5

Person Re-Identification Using CNN Features Learned from Combination of Attributes

Tetsu Matsukawa Einoshin Suzuki

Kyushu University, JAPAN

Person Re-Identification

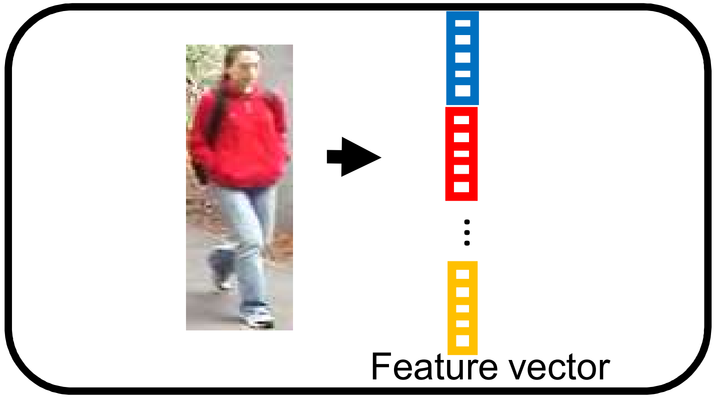
- Task : Find the same person in different camera view

- Challenges : Large intra-personal variations
  e.g. illumination/pose/occlusion/background

Approach for Re-Identification

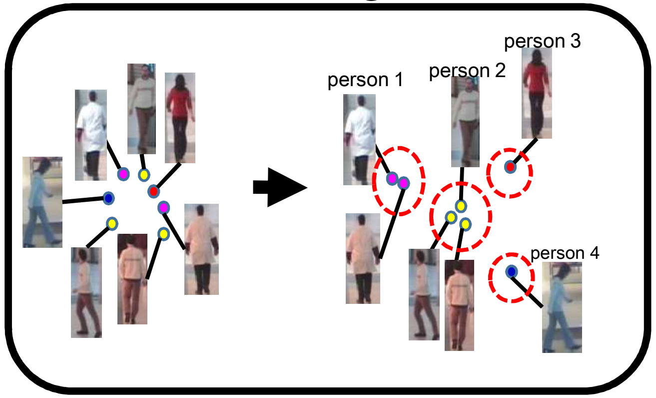
- Traditional Approach

1. Feature extraction



+

2. Metric learning



- Covolutional Neural Network (CNN)
  - ➢ Unify feature extraction and distance metric learning [Yi ICPR2014]
  - ➢ Require a large amount of annotated training data
  - ➢ Most of person re-identification datasets are small (typically below than 1K training images)

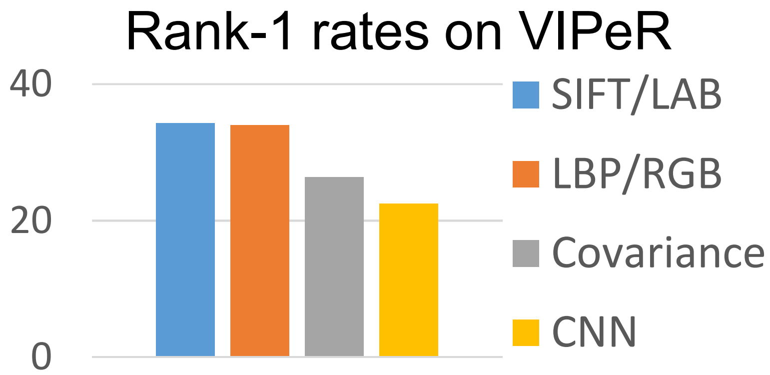
CNN Features

- Neural activations of top layers of a pre-trained CNN
  - Strong off-the-shelf feature descriptors
    [Donahue ICML2014, Razavian CVPRWS 2014, Azizpour PAMI2016]
  - A large training data is required only for feature learning

- Problem
  - Pre-trained CNN features perform poor in person re-identification due to large-disparity from pre-trained task

Pre-training task

Target task

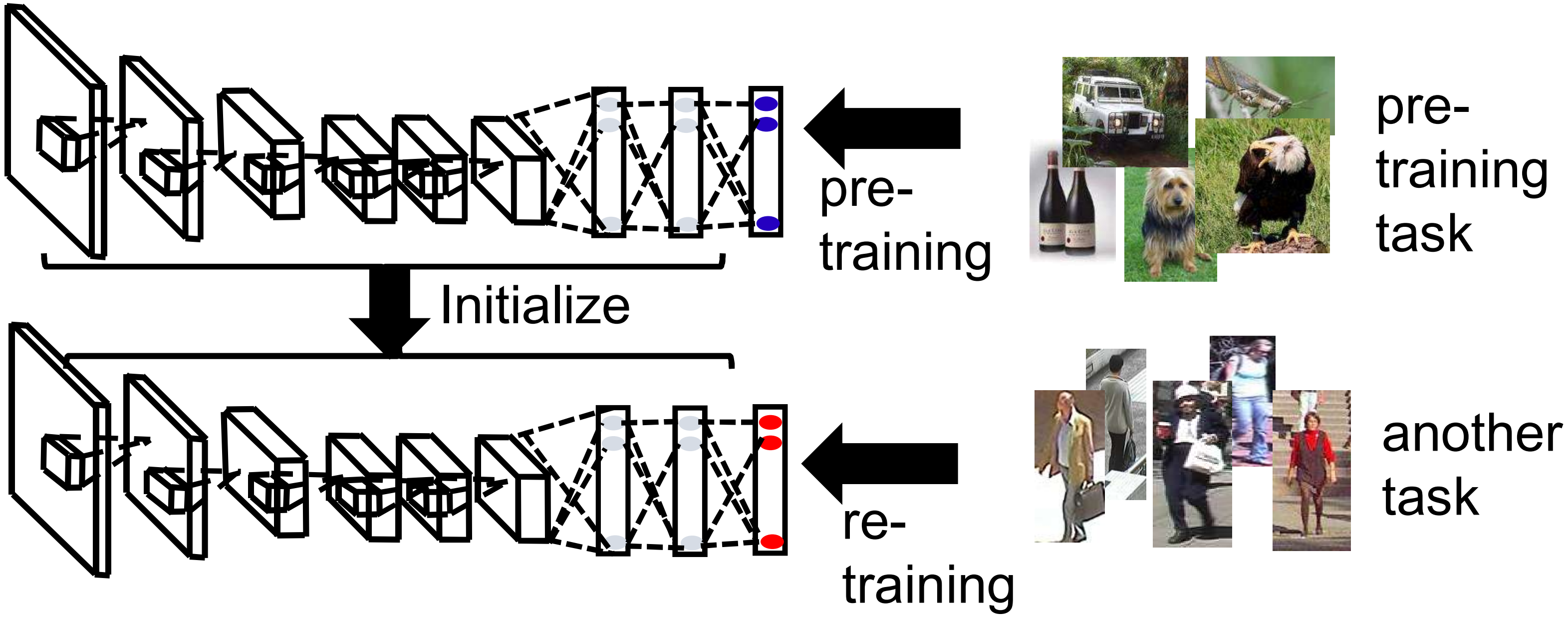


[Paisitkringkrai CVPR2015]

Kyushu University

Fine-tuning

- Re-training pre-trained CNN with different dataset
  - Transfer knowledge of pre-trained data
  - Significantly improve the recognition accuracies on another task [Oquab CVPR2014, Chatfiled BMVC2014 etc]



Contribution(1): Fine-tuning by Attribute Classification Task

- Task: person image recognition by semantic attributes
  eg. Gender, Luggage, Clothing.

male | female | backpack | luggage case | sweater | tshirt | short skirt | hot pants

- Advantage
  - Easy to be labeled by human annotator
  - Large number of training samples per attributes
- Large-scale dataset
  - eg. PETA[Deng ACMMM2014], RAP[Li ArXiv2016]

Motivation

- Annotated attributes in existing datasets are coarse to determine specific person

Tshirt

backpack

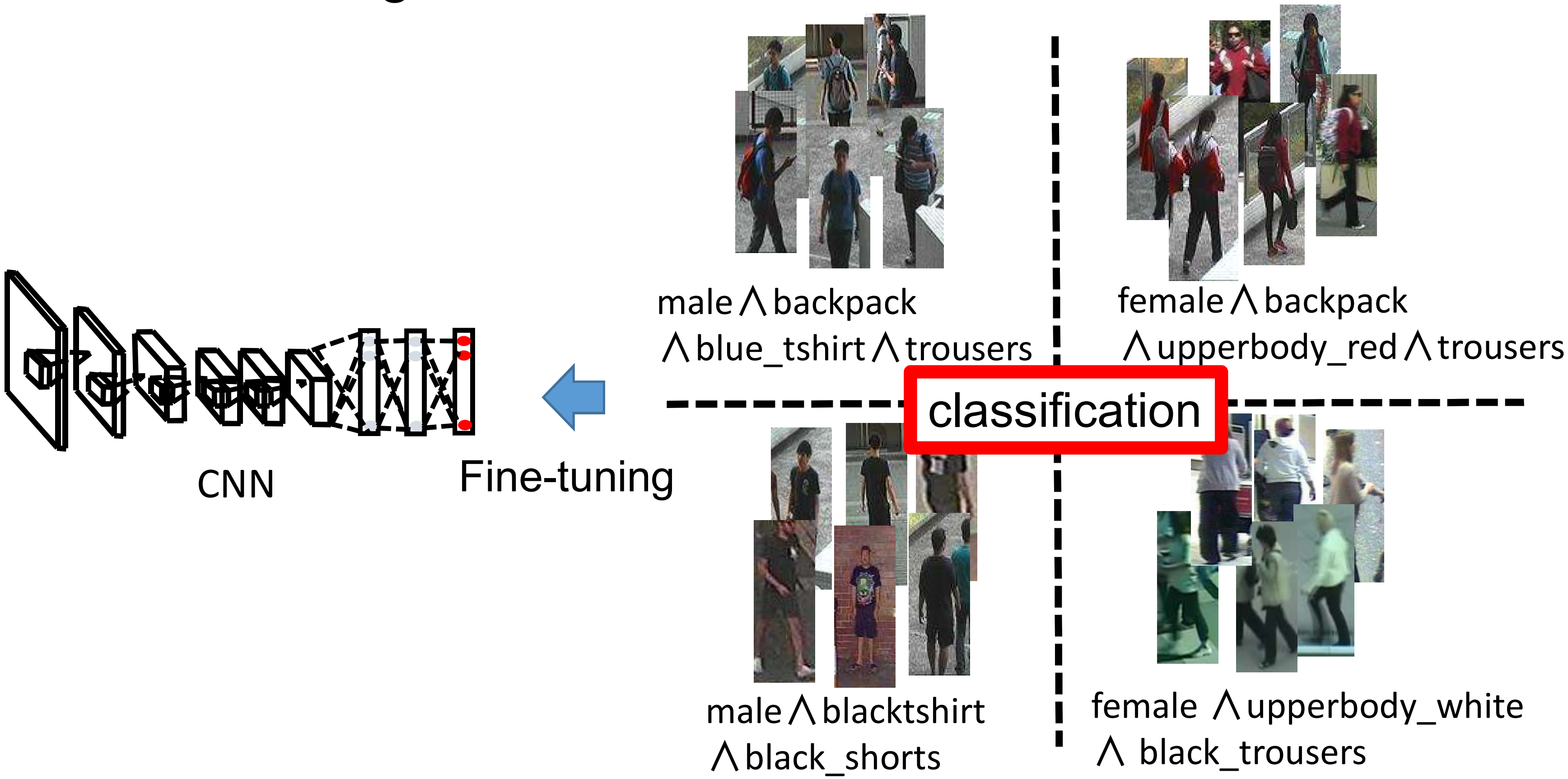
➢ Concern: Discriminative power of CNN features solely fine-tuned on attribute classification task would be insufficient

- Combination of attribute is more person specific

Tshirt ∧ backpack
∧ upper body white
∧ trouser

Contribution(2): Fine-tuning by Classification of Combination Attributes

- Fine-tuning task: Classification of attribute-combination

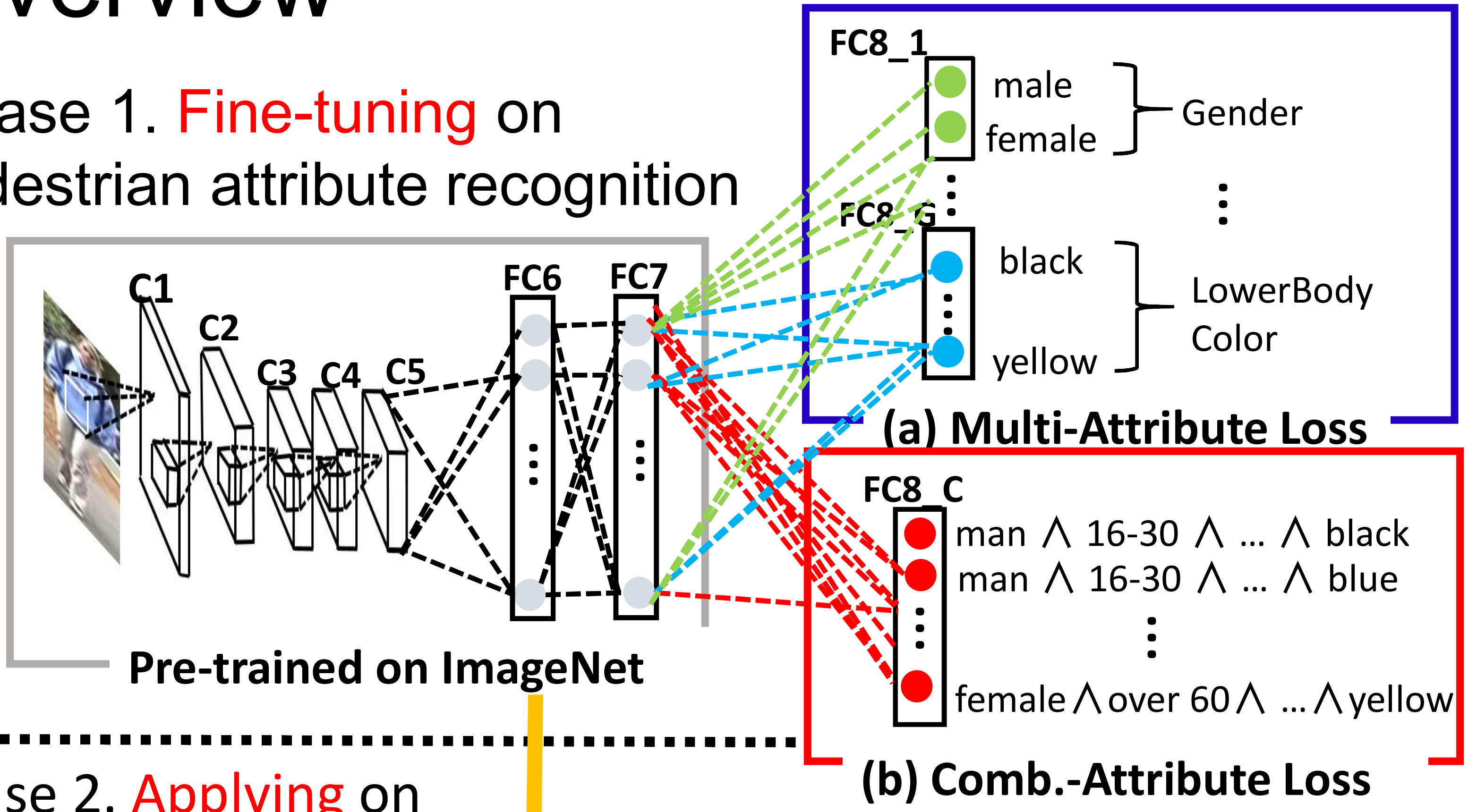


- Expectation : Emerge discriminative features within CNN

ICPR 2016

Overview

Phase 1. Fine-tuning on pedestrian attribute recognition

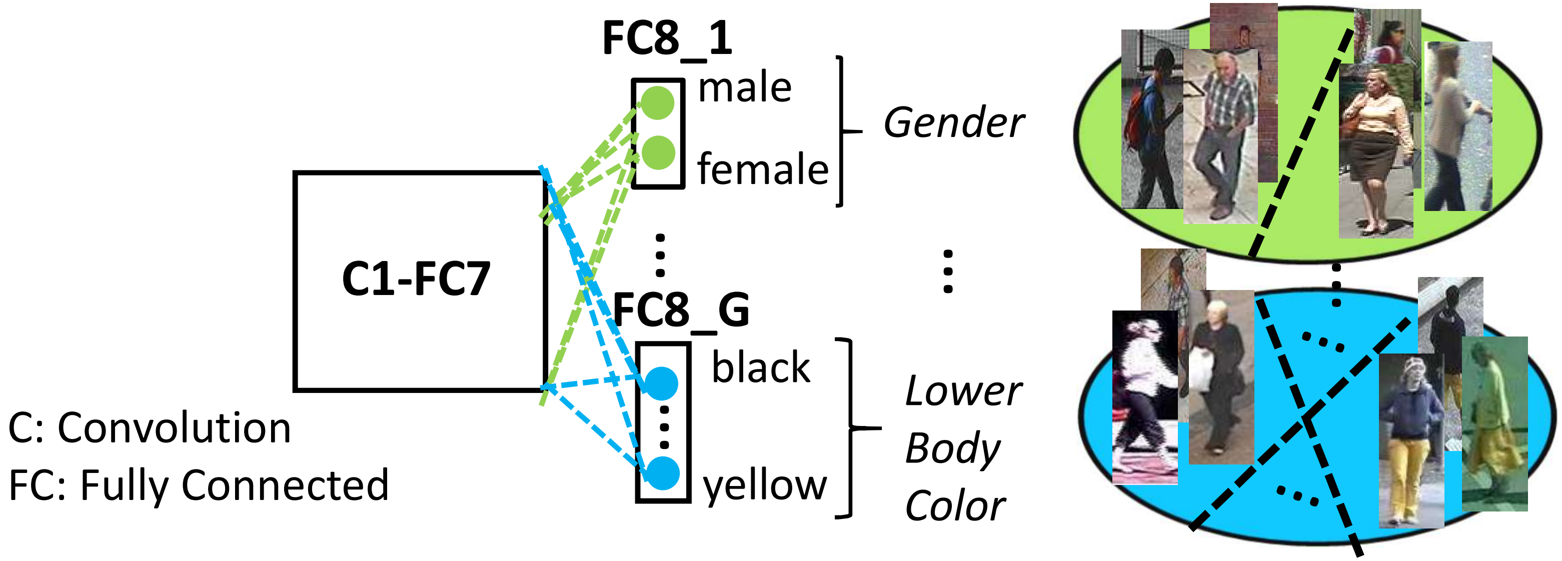


Phase 2. Applying on person re-identification dataset

CNN Feature + Metric Learning

(a) Multi-Attribute Classification

- Prepare: Dataset with mutually exclusive attribute labels in G groups
- Attach multi-class classification layer for each group



- ➢ Share the lower layers of CNN inspired by [Li ACPR2015, Zhu ICB2015, Sudowe ICCVWS 2015]

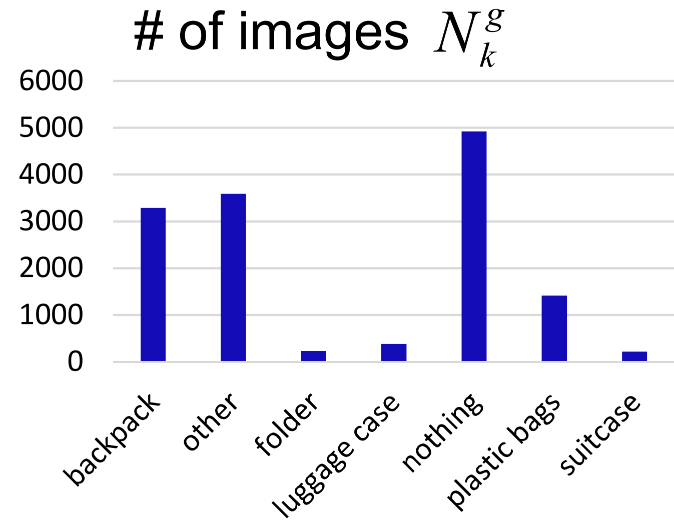
(a) Loss Function

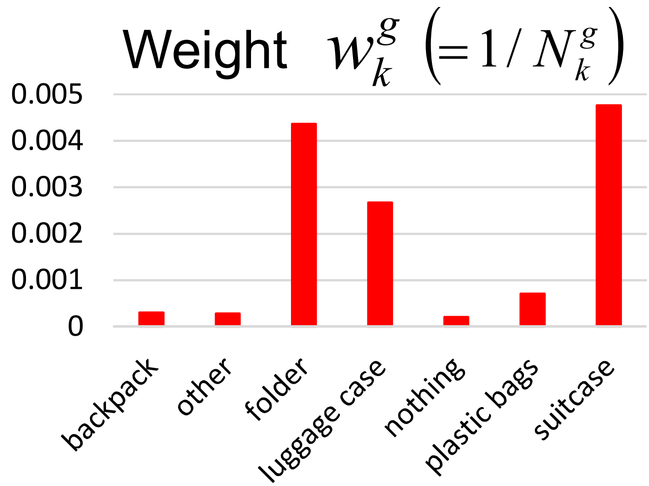
- #of training samples for each attribute is largely imbalanced

- Use weighted cross entropy loss function
  - Loss for g-th attribute group

$$L^g = -\frac{1}{N^g} \sum_{i=1}^{N} \sum_{k=1}^{K^g} w_{k(i)}^g l_{i,k}^g \log p_{i,k}^g$$

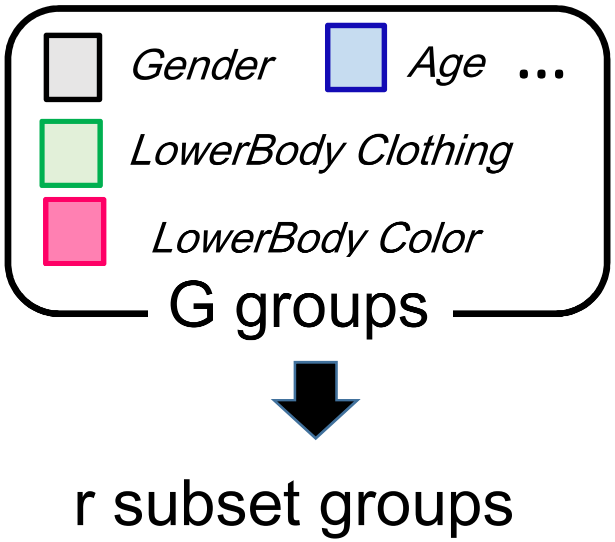
w_k^g	: weight for k-th attribute
$l_{i,k}^g$	: correct label
$p_{i,k}^g$	: output of softmax function





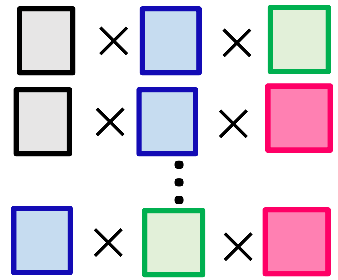
(b) Combination of Groups

- Combination only among different groups is required since attributes in each group is mutually exclusive
- Combination among r subset groups can be considered

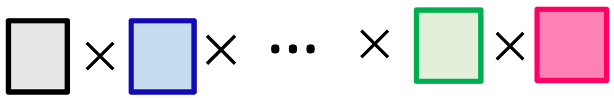


➢ If r < G :
Selection from many possible group subsets required

e.g. r = 3



➢ <u>If r = G:</u>
<u>No need to select subset groups</u>
(<u>use this case as default</u>)



Kyushu University

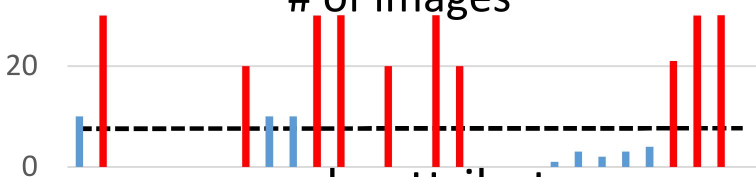
(b) Comb.-Attribute Classification

- All attribute combinations among different groups

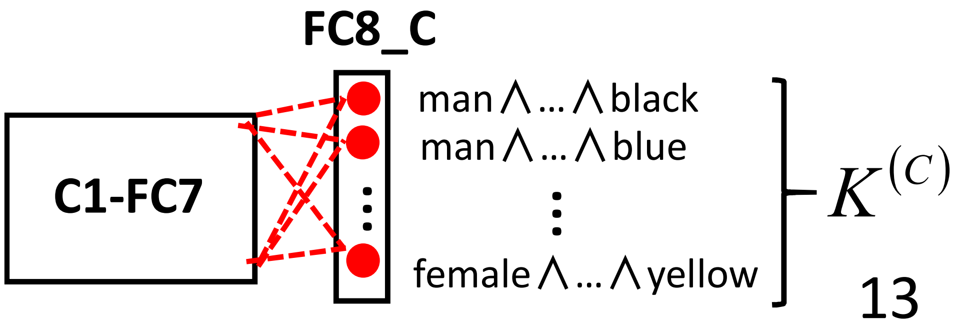
$$K^{(C')} = \underbrace{\begin{bmatrix}\text{male} \\ \text{female}\end{bmatrix}}_{K^{(1)}} \times \underbrace{\begin{bmatrix}\text{less15} \\ \vdots \\ \text{over60}\end{bmatrix}}_{K^{(2)}} \times \cdots \times \underbrace{\begin{bmatrix}\text{black} \\ \vdots \\ \text{yellow}\end{bmatrix}}_{K^{(G)}}$$

$K^{(g)}$: # of attributes in g-th group

- Select $K^{(C)} \left(\leq K^{(C')}\right)$ frequent comb.-attributes



- Attach $K^{(C)}$ class classification layer
  - Minimize the weighted cross entropy loss L^C



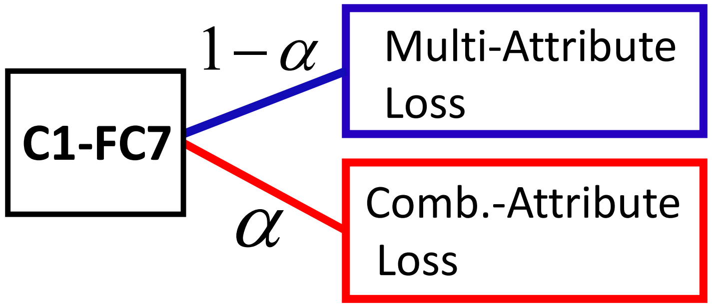
(a)+(b) Total Loss Function

- Some training data do not have the comb.-attribute label
  - ➢ Missing labels in some attribute group
  - ➢ Discard too rare combinations

⬇

- Jointly minimize with multi-attribute classification losses

$$L = \alpha L^{C} + (1-\alpha)\frac{1}{G}\sum_{g=1}^{G} L^{g}$$



$0 \le \alpha \le 1$: Contribution of combination attribute loss

Default: $\alpha = 0.5$

Database for Fine-tuning

- PETA [Deng ACMMM2014]: 19,000 images with 61 annotated attributes
- We manually selected 7 attributes groups

( ) : # of attributes

Experiment

- Setup
  - Network architecture: AlexNet [Krizhevsky NIPS2012]
  - Extract 4,096 dim. feature vector from FC6 layer
    (Applied L2 norm normalization)
  - Metric Learning: Cross-view Quadric Discriminant Analysis (XQDA) [Liao CVPR15]

- Four person re-identification dataset

VIPeR

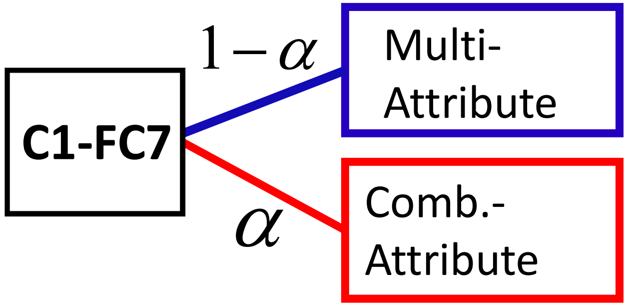
CUHK01

PRID450S

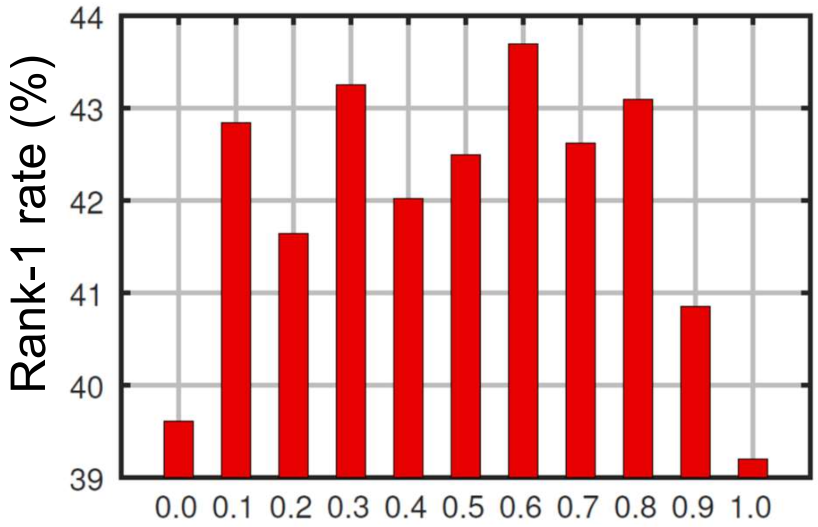
GRID

ICPR 2016

Performance analysis on VIPeR (1/3)

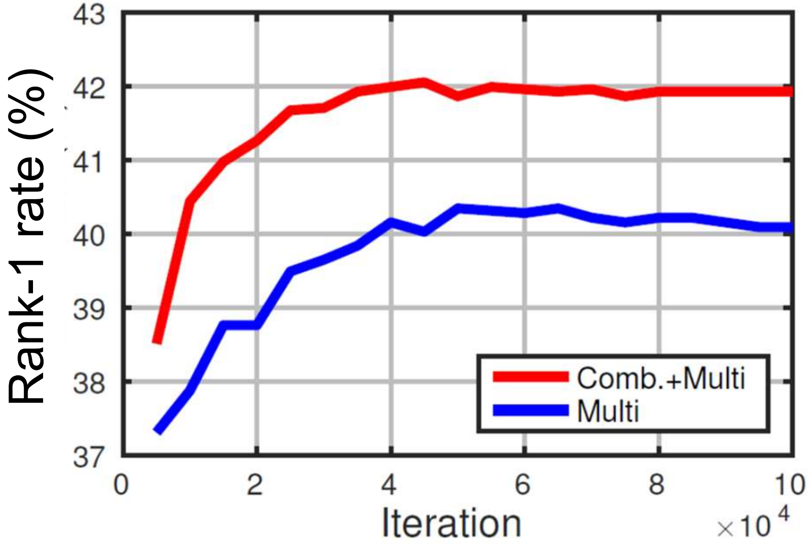


Parameter α



Multi.-Attribute ← High Contribution → Comb.-Attribute

# of iteration (backpropagation)



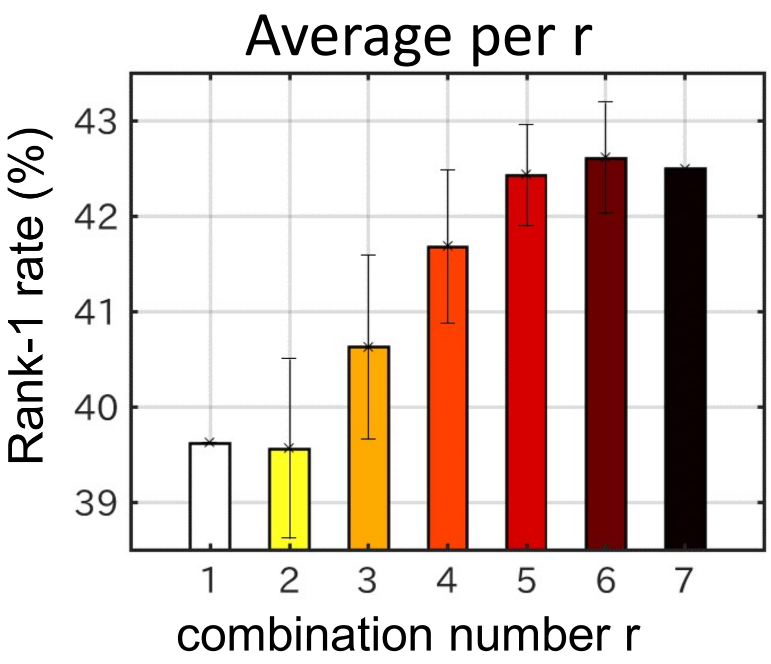
- Using both losses improves accuracies

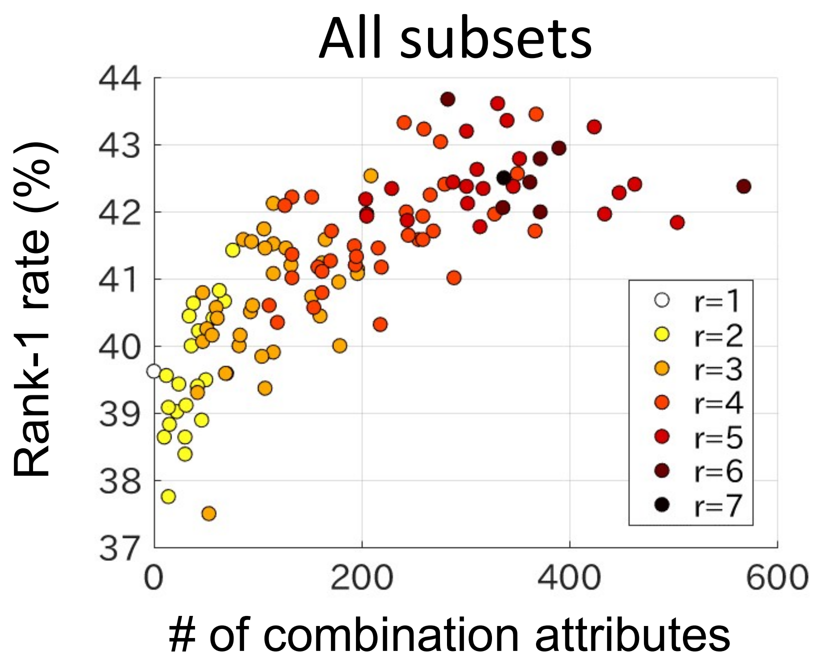
- Converged about 5k iterations
- Comb.+Multi consistently better

Performance analysis on VIPeR (2/3)

Gender *Age*
…
LowerBody
Color

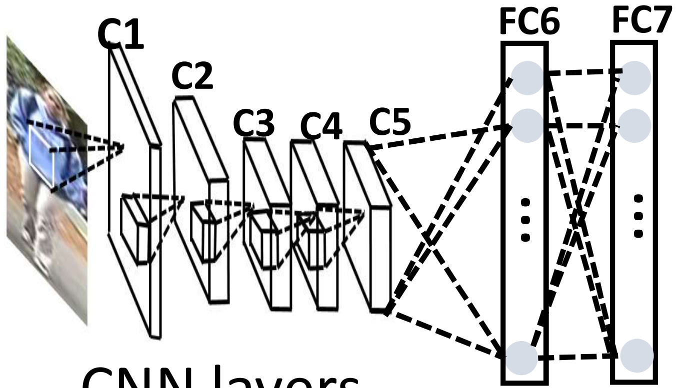
r groups



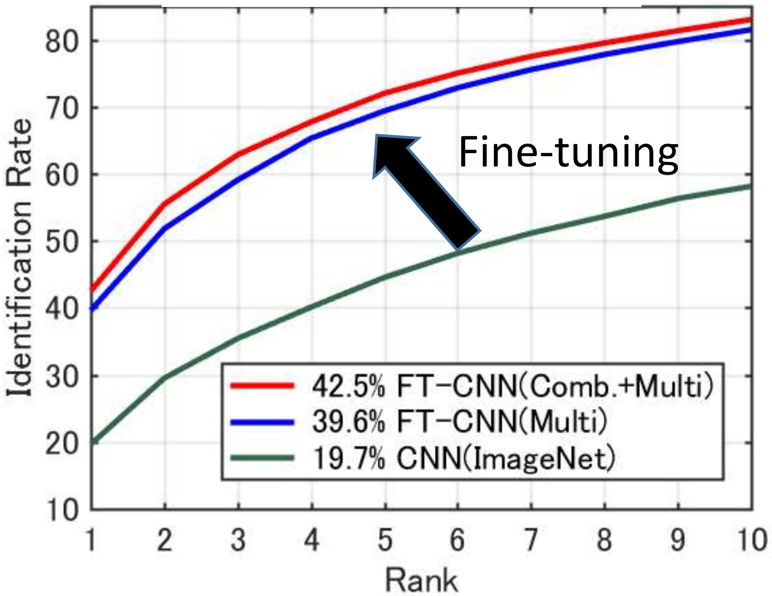


- Performance increases as to increase the combined groups
- Larger r produce large number of combination attributes

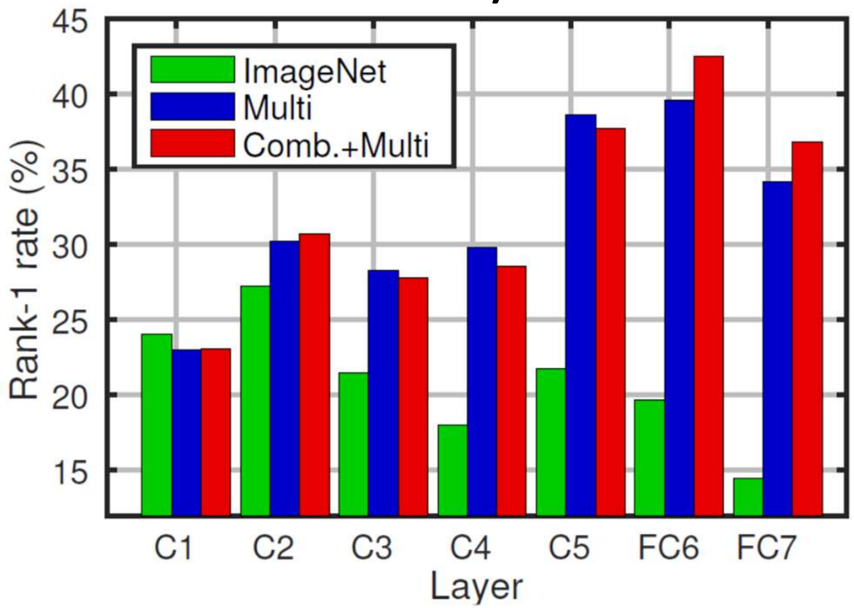
Performance analysis on VIPeR (3/3)



CMC Curve (FC6)



CNN layers

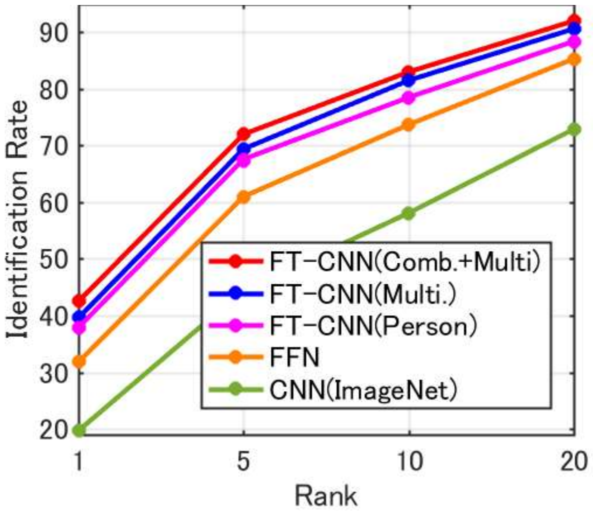


- Multi-attribute classification loss improves 19.9 % in rank-1
- Combination attribute loss further improves 2.9 % in rank-1.

- Lower layer better when fine-tuning is not conducted
- FC6 layer is better than FC7 layer

Performance Comparison (1/2)

CMC Curve (VIPeR)



- Compared CNN features
  - FT-CNN: Fine-tuned
    - Comb.+Multi: Ours
    - Multi.: Only multi-attribute loss
    - Person: Person identity of PETA
  - FFN: Feature Fusion Net [Wu WACV2016]
  - ImageNet: Pre-trained CNN

Rank-1 rates (%)

Methods	VIPeR	CUHK01	PRID450S	GRID
FT-CNN(Comb.+Multi)	42.5	46.8	58.2	25.2
FT-CNN(Multi.)	39.6	44.8	55.8	24.6
FT-CNN(Person)	37.9	44.0	56.4	23.9
FNN	31.8	32.4	51.6	-
CNN(ImageNet)	19.7	28.5	38.0	8.2

Performance Comparison (2/2)

State-of-the-art — Rank-1 rates (%)

Methods	Ref.	VIPeR	CUHK01	PRID450S	GRID
GOG	CVPR2016	49.7	57.8	68.4	24.7
FT-CNN	Ours	42.5	46.8	58.2	25.2
LOMO	CVPR2015	40.0	50.0	61.4	16.6
Improved Deep	CVPR2015	34.8	47	[illegible]	[illegible]
SCNCD	ECCV2014	37.8	-	41.6	-
DALF	ICPR2014	35.4	-	-	18.1

➢ Competitive to LOMO

CNN + hand-crafted descriptors — Rank-1 rates (%)

Methods	Ref.	VIPeR	CUHK01	PRID450S	GRID
FT-CNN+LOMO	Ours	52.1	62.3	71.5	29.1
FFN+LOMO	WACV2016	51.1	[illegible]	[illegible]	[illegible]
Metric Ensemble	CVPR2015	45.9	53.4	-	-

➢ Improved the accuracies

ICPR 2016

Conclusion

- CNN fine-tuning on pedestrian attributes dataset
  - High performance gains by conducting fine-tuning with multi-attribute classification loss (16.3-19.9% in CMC@rank-1)
  - Combination of attributes loss further improve performances (0.6-2.9% in CMC@rank-1)
  - Achieved competitive performance to hand-crafted descriptors

- Future work
  - Increase number of training samples
  - Combination with a classification loss of person identity