Person Re-Identification by Joint Learning of Multi-Loss Classification

Wei Li, Xiatian Zhu, Shaogang Gong

Queen Mary University of London

{w.li, xiatian.zhu, s.gong}@qmul.ac.uk

- Introduction
- Model Design
- Experiments
- Conclusions

Queen Mary University of London

- **Introduction**
- Model Design
- Experiments
- Conclusions

Introduction

What is person re-identification?

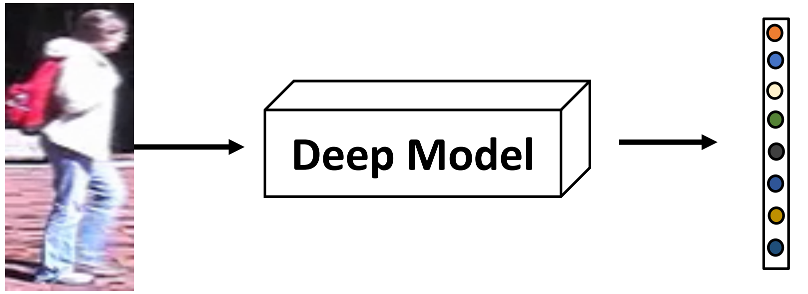
Matching person identities from non-overlapping camera views.

- ❑ **Challenges**: appearance changes in human pose, illumination, occlusion, and background clutter [Gong et al., 2014];
- ❑ **Objective**: the view- and location-invariant (cross-domain) representation or good matching metric.

Introduction

Person re-ids

Cat	Method	Feature		Metric	
		Hand-Crafted	DL	CPSL	Generic
A	**XQDA** [Liao *et al.*, 2015]	LOMO	-	XQDA	-
	GOG [Matsukawa *et al.*, 2016b]	GOG	-	XQDA	-
	NFST [Zhang *et al.*, 2016]	LOMO, KCCA	-	NSFT	-
	SCS [Chen *et al.*, 2016]	CHS	-	SCS	-
B	**DCNN+** [Ahmed *et al.*, 2015]	-	DCNN+	DVM	-
	X-Corr [Subramaniam *et al.*, 2016]	-	X-Corr	DVM	-
	MTDnet [Chen *et al.*, 2017a]	-	MTDnet	DVM, L2	-
C	**S-CNN** [Varior *et al.*, 2016]	-	S-CNN	-	L2
	DGD [Xiao *et al.*, 2016]	-	DGD	-	L2
	MCP [Cheng *et al.*, 2016]	-	MCP	-	L2
	JLML (Ours)	-	JLML	-	L2



- To construct good features for person images [feature learning]
- To achieve good distance metric for matching task [metric learning]
  - ➢Focused on the **deep feature learning** while using **L2** metric only.

Introduction

Observations

- Person images have some kind of unified body structure;
- Low inter-class and high intra-class variance caused by appearance changes;
- Human visual systems leverage both **global** (contextual) and **local** (saliency) information concurrently;
- Either local or global feature learning alone is suboptimal .

Introduction

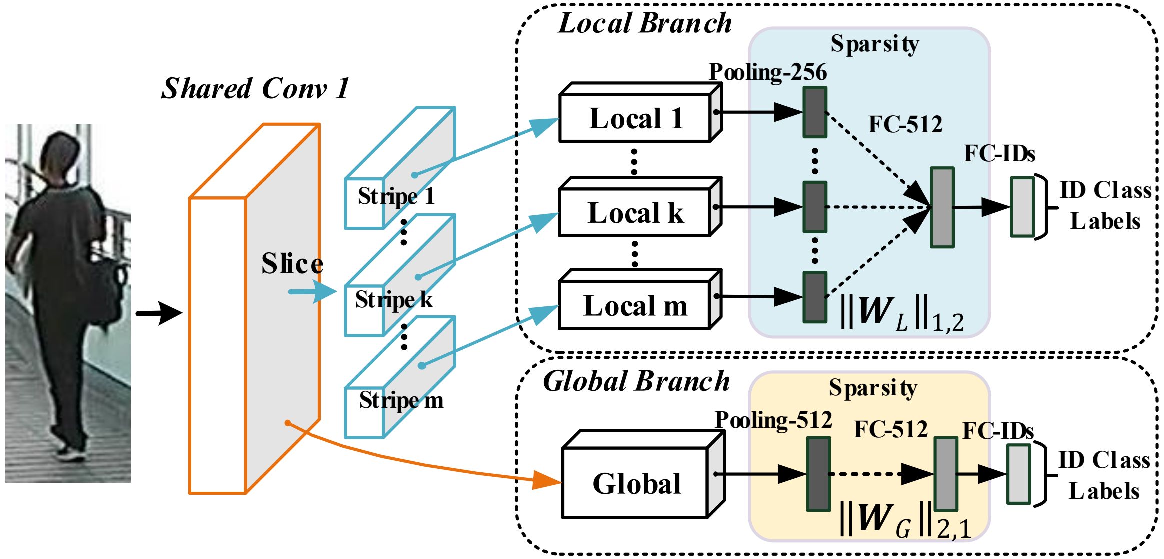
Problem Definition

How to discover and capture concurrently **complementary discriminative** information for both local and global visual features of person images?

- Introduction
- **Model Design**
- Experiments
- Conclusions

Model Design

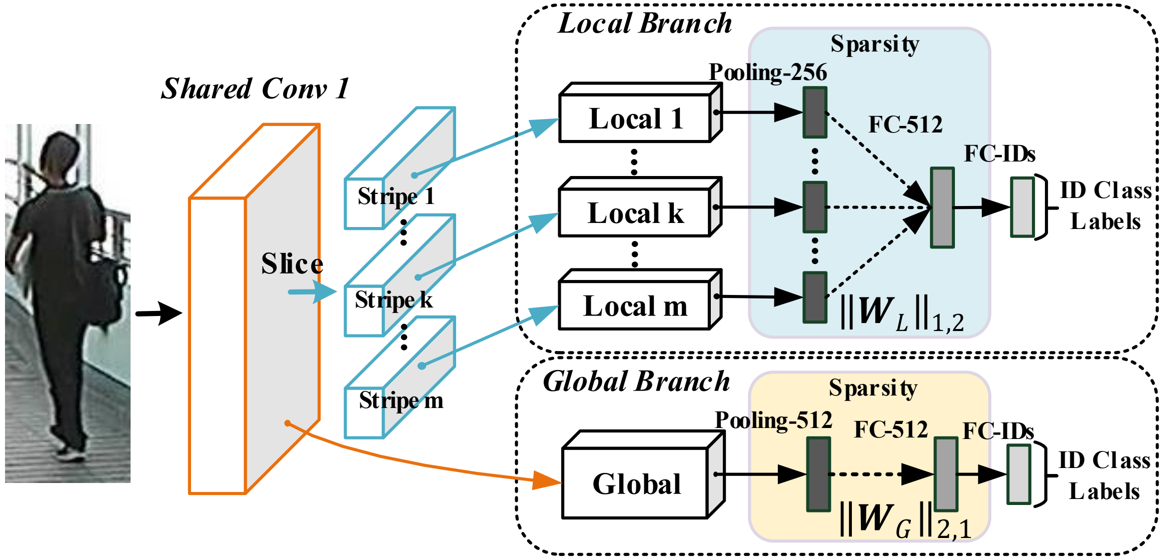
Joint Learning of Multi-Loss



- **Local branch** of m streams learning the discriminative local visual features for m local image regions; (saliency)
- **Global branch** responsible for learning the most discriminative global level features from the entire person image. (contextual)

Model Design

Joint Learning of Multi-Loss

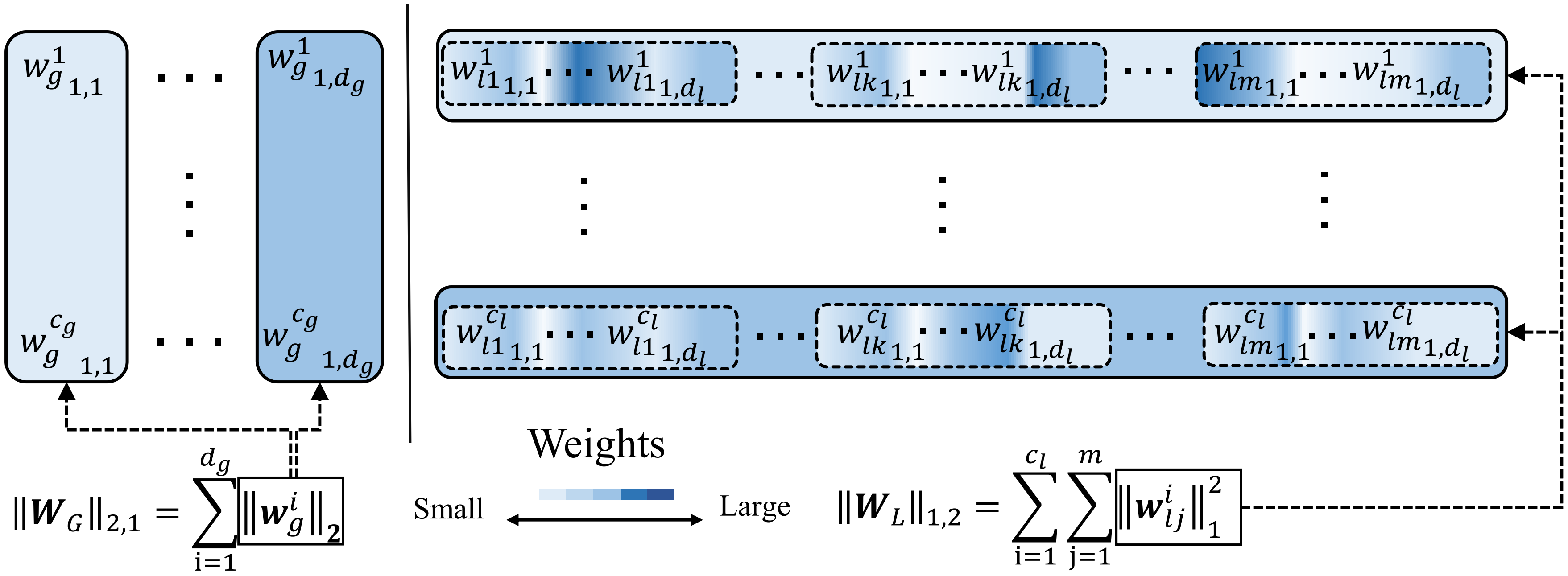


- Shared low-level features;
- Multi-task independent learning subject to shared label constraints;
- Adopt the Residual CNN unit [He et al., 2016] as the JLML's building blocks.

Model Design

Feature Selections:

Learning robustness against noise and diverse data source.



Model Design

Loss function

Learning robustness against noise and diverse data source.

$$l = -\frac{1}{n_{\text{bs}}} \sum_{i=1}^{n_{\text{bs}}} \log \left(p(\tilde{y}_i = y_i | \boldsymbol{I}_i) \right)$$

- Significantly simplified training data batch construction;
- More scalable in real-world applications with very large training population sizes when available;
- Representations optimised for classification tasks can generalise well to new categories

- Introduction
- Model Design
- **Experiments**
- Conclusions

Experiments

State-of-the-art re-id results on CUHK01 and CUHK03

Table 7: CUHK01 evaluation. $1^{st}/2^{nd}$ best in bold/typewriter.

Cat	Split	871/100 split				486/485 split			
	Rank (%)	R1	R5	R10	R20	R1	R5	R10	R20
	Single-Shot Testing Setting								
A	GOG	-	-	-	-	57.8	79.1	86.2	92.1
B	DCNN+	65.0	-	-	-	47.5	71.6	80.3	87.5
	X-Corr	81.2	97.3	-	98.6	65.0	89.7	-	94.4
	MTDnet	78.5	96.5	97.5	-	-	-	-	-
C	DGD	-	-	-	-	66.6	-	-	-
	MCP	-	-	-	-	53.7	84.3	91.0	96.3
	JLML	87.0	97.2	98.6	99.4	69.8	88.4	93.3	96.3
	Multi-Shot Testing Setting								
A	XQDA	-	-	-	-	63.2	83.9	90.0	94.2
	GOG	-	-	-	-	67.3	86.9	91.8	95.9
	NFST	-	-	-	-	69.1	86.9	91.8	95.4
C	**JLML**	91.2	98.4	99.2	99.8	76.7	92.6	95.6	98.1

Table 5: CUHK03 evaluation. $1^{st}/2^{nd}$ best in red/blue.

Cat	Annotation	Labelled				Detected			
	Rank (%)	R1	R5	R10	R20	R1	R5	R10	R20
A	XQDA	55.2	77.1	86.8	83.1	46.3	78.9	83.5	93.2
	GOG	67.3	91.0	96.0	-	65.5	88.4	93.7	-
	NSFT	62.5	90.0	94.8	98.1	54.7	84.7	94.8	95.2
B	DCNN+	54.7	86.5	93.9	98.1	44.9	76.0	83.5	93.2
	X-Corr	72.4	95.5	-	98.4	72.0	96.0	-	98.2
	MTDnet	74.7	96.0	97.5	-	-	-	-	-
C	S-CNN	-	-	-	-	68.1	88.1	94.6	-
	DGD	75.3	-	-	-	-	-	-	-
	JLML	83.2	98.0	99.4	99.8	80.6	96.9	98.7	99.2

Experiments

State-of-the-art re-id results on GRID and Market-1501

Table 9: GRID evaluation. 1st/2nd best in red/blue.

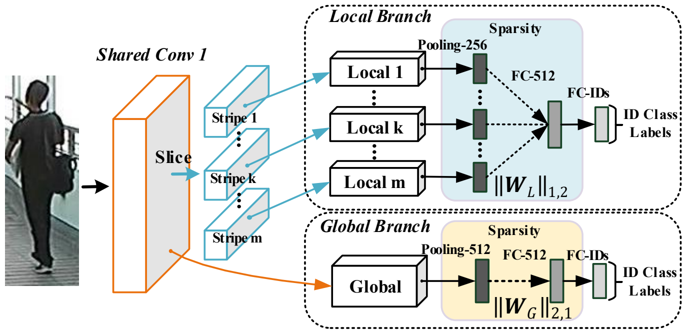
Cat	Rank (%)	R1	R5	R10	R20
A	XQDA	16.6	33.8	41.8	52.4
	GOG	24.7	47.0	58.4	69.0
	SCS	24.2	44.6	54.1	65.2
B	X-Corr	19.2	38.4	53.6	66.4
C	**JLML**	37.5	61.4	69.4	77.4

Table 6: Market-1501 evaluation. 1st/2nd best in red/blue. All person bounding box images were auto-detected.

Cat	Query Type	Single-Query		Multi-Query	
	Measure (%)	R1	mAP	R1	mAP
A	XQDA	43.8	22.2	54.1	28.4
	SCS	51.9	26.3	-	-
	NFST	61.0	35.6	71.5	46.0
C	S-CNN	65.8	39.5	76.0	48.4
	JLML	85.1	65.5	89.7	74.5

Experiments

Ablation Study



Loss	Query Type	Single-Query		Multi-Query	
	Measure (%)	R1	mAP	R1	mAP
UniLoss	Global Feature	58.3	31.7	70.4	43.2
	Local Feature	46.3	26.3	58.0	34.0
	Full	76.1	52.2	83.7	62.8
MultiLoss	Global Feature	77.4	56.0	85.0	66.0
	Local Feature	78.9	57.8	86.4	68.4
	Full	85.1	65.5	89.7	74.5

- ✓ Importance of branch independence (Multi-loss matters).
- ✓ Complementary benefits of global and local features.

Experiments

Ablation Study

Query Type	Single-Query		Multi-Query	
Measure (%)	R1	mAP	R1	mAP
Without Shared Feature	83.2	63.1	88.3	72.1
With Shared Feature	85.1	65.5	89.7	74.5

✓ Benefits from shared low-level features.

Query Type	Single-Query		Multi-Query	
Measure (%)	R1	mAP	R1	mAP
Without SFL	83.4	63.8	88.7	72.9
With SFL	85.1	65.5	89.7	74.5

✓ Effects of selective feature learning (SFL) .

Experiments

Further Analysis

Query-Type	Single-Query		Multi-Query	
Measure (%)	R1	mAP	R1	mAP
KISSME	82.1	61.4	87.5	70.2
XQDA	82.6	63.2	88.2	72.4
CRAFT	77.9	56.4	-	-
L2	85.1	65.5	89.7	74.5

✓ Complementary of JLML features and metric learning .

Model	FLOPs	PN (million)	Depth	Stream #
AlexNet	7.25×10^{8}	58.3	7	1
VGG16	1.55×10^{10}	134.2	16	1
ResNet50	3.80×10^{9}	23.5	50	1
GoogLeNet	1.57×10^{9}	6.0	22	1
JLML-ResNet39	1.54×10^{9}	7.2	39	5

✓ Comparisons of model size and complexity.

Experiments

Further Analysis

Query-Type	Single-Query		Multi-Query	
Measure (%)	R1	mAP	R1	mAP
2	83.9	64.4	88.8	72.9
4	85.1	65.5	89.7	74.5
6	83.4	62.6	88.5	71.8
8	82.3	61.3	87.4	70.7
10	81.7	60.4	87.2	69.8

✓ Effect of body parts.

✓ 4 body-parts: head + shoulder, upper-body, upper-leg and lower-leg.

- Introduction
- Model Design
- Experiments
- **Conclusions**

Conclusions

- ✓ An idea of learning concurrently both **local** and **global** discriminative feature selections in different context;
- ✓ A novel Joint Learning Multi-Loss (**JLML**) CNN model by optimising multiple classification losses;
- ✓ A structured sparsity based **feature selection** learning mechanism for improving JLML robustness w.r.t noise and data co-variance.

Thanks