

DOCTORAL THESIS

Open-world person re-identification

Ye, Mang

Date of Award:
2019

[Link to publication](#)

General rights

Copyright and intellectual property rights for the publications made accessible in HKBU Scholars are retained by the authors and/or other copyright owners. In addition to the restrictions prescribed by the Copyright Ordinance of Hong Kong, all users and readers must also observe the following terms of use:

- Users may download and print one copy of any publication from HKBU Scholars for the purpose of private study or research
- Users cannot further distribute the material or use it for any profit-making activity or commercial gain
- To share publications in HKBU Scholars with others, users are welcome to freely distribute the permanent URL assigned to the publication

Abstract

With the increasing demand of intelligent video surveillance systems, person re-identification (re-ID) plays an important role in intelligent video analysis, which aims at matching person images across non-overlapping camera views. It has gained increasing attention in computer vision community. With the advanced deep neural networks, existing methods have achieved promising performance on the widely-used re-ID benchmarks, even outperform the human-level rank-1 matching accuracy. However, most of the research efforts are conducted on the closed-world settings, with large-scale well annotated training data and all the person images are from the same visible modality. As a prerequisite in practical video surveillance application, there is still a large gap between the closed-world research oriented setting and the practical open-world settings. In this thesis, we try to narrow the gap by studying three important issues in open-world person re-identification, including 1) *unsupervised learning with large-scale unlabelled training data*; 2) *learning robust re-ID model with label corrupted training data* and 3) *cross-modality visible-thermal person re-identification with multi-modality data*.

For unsupervised learning with unlabelled training data, we mainly focus on video-based person re-identification, since the video data is usually easily obtained by tracking algorithms and the video sequence provides rich weakly labelled samples by assuming the image frames within the tracked sequence belonging to the same person identity. Following the cross-camera label estimation approach, we formulate the cross-camera label estimation as a one-to-one graph matching problem, and then propose a novel dynamic graph matching framework to estimate cross-camera labels.

However, in a practical wild scenario, the unlabelled training data usually cannot satisfy the one-to-one matching constraint, which would result in a large proportion of false positives. To address this issue, we further propose a novel robust anchor embedding method for unsupervised video re-ID. In the proposed method, some anchor sequences are firstly selected to initialize the CNN feature representation. Then a robust anchor embedding method is proposed to measure the relationship between the unlabelled sequences and anchor sequences, which considers both the scalability and efficiency. After that, a top- k counts label prediction strategy is proposed to predict the labels of unlabelled sequences. With the newly estimated sequences, the CNN representation could be further updated.

For robust re-ID model learning with label corrupted training data, we propose a two-stage learning method to handle the label noise. Rather than simply filtering the falsely annotated samples, we propose a joint learning method by simultaneously refining the falsely annotated labels and optimizing the neural networks. To address the limited training samples for each identity, we further propose a novel hard-aware instance re-weighting strategy to fine-tune the learned model, which assigns larger weights to hard samples with correct labels.

For cross-modality visible-thermal person re-identification, it addresses an important issue in night-time surveillance applications by matching person images across different modalities. We propose a dual-path network to learn the cross-modality feature representations, which learns the multi-modality sharable feature representations by simultaneously considering the modality discrepancy and commonness. To guide the feature representation learning process, we propose a dual-constrained top-ranking loss, which contains both cross-modality and intra-modality top-ranking constraints to reduce the large cross-modality and intra-modality variations.

Besides the open-world person re-identification, we have also studied the unsupervised embedding learning problem for general image classification and retrieval. Motivated by supervised embedding learning, we propose a data augmentation invariant and instance spread-out feature. To learn the feature embedding, we propose

a instance feature-based softmax embedding, which optimizes the embedding directly on top of the real-time instance features. It achieves much faster learning speed and better accuracy than existing methods.

In short, the major contributions of this thesis are summarized as follows.

- A dynamic graph matching framework is proposed to estimate cross-camera labels for unsupervised video-based person re-identification.
- A robust anchor embedding method with top- k counts label prediction is proposed to efficiently estimate the cross-camera labels for unsupervised video-based person re-identification under wild settings.
- A two-stage PurifyNet is introduced to handle the label noise problem in person re-identification, which jointly refines the falsely annotated labels and mines hard samples with correct labels.
- A dual-constrained top-ranking loss with a dual-path network is proposed for cross-modality visible-thermal person re-identification, which simultaneously addresses the cross-modality and intra-modality variations.
- A data augmentation invariant and instance spread-out feature is proposed for unsupervised embedding learning, which directly optimizes the learned embedding on top of real-time instance features with softmax function.

Keywords: Person re-identification, open-world, unsupervised learning, cross-modality, label noise

Table of Contents

| | |
|---|------------|
| DECLARATION | i |
| Abstract | ii |
| Acknowledgements | iv |
| Table of Contents | vii |
| List of Tables | xi |
| List of Figures | xiv |
| Chapter 1 Introduction | 1 |
| 1.1 Background and Motivation | 1 |
| 1.1.1 Background | 1 |
| 1.1.2 Motivations of this project | 3 |
| 1.2 Review of Related Person Re-identification Methods | 8 |
| 1.2.1 Closed-world Person Re-Identification | 8 |
| 1.2.2 Open-world Person Re-Identification | 10 |
| 1.3 Contributions of This Thesis | 12 |
| 1.4 Thesis Overview | 14 |
| Chapter 2 Dynamic Graph Co-Matching for Unsupervised Video- based Person Re-Identification | 16 |
| 2.1 Introduction | 16 |

| | | |
|-------|---|----|
| 2.2 | Related Work | 19 |
| 2.3 | Graph Matching for Video Re-ID | 21 |
| 2.3.1 | Why Graph Matching is Good for Label Estimation in Re-ID? | 21 |
| 2.3.2 | Problem Formulation | 23 |
| 2.4 | Dynamic Graph Co-Matching | 25 |
| 2.4.1 | Positive Re-weighting | 27 |
| 2.4.2 | Metric Learning with Re-weighted Labels | 30 |
| 2.4.3 | Dynamic Updating | 30 |
| 2.4.4 | Extension with Co-Matching | 31 |
| 2.4.5 | Convergence and Complexity Analysis | 32 |
| 2.5 | Experimental Results | 34 |
| 2.5.1 | Experimental Settings | 34 |
| 2.5.2 | Self Evaluation | 36 |
| 2.5.3 | Comparison with State-of-the-art Label Estimation | 41 |
| 2.5.4 | Comparison with State-of-the-art Unsupervised re-ID | 42 |
| 2.5.5 | Robustness in the Wild | 44 |
| 2.6 | Conclusion | 45 |

Chapter 3 Robust Anchor Embedding for Unsupervised Video Person

| | | |
|-------|--------------------------------------|-----------|
| | Re-Identification in the Wild | 47 |
| 3.1 | Introduction | 47 |
| 3.2 | Related Work | 50 |
| 3.3 | Proposed Method | 51 |
| 3.3.1 | Overview | 51 |
| 3.3.2 | Anchor Initialization | 52 |
| 3.3.3 | Label Estimation | 53 |
| 3.3.4 | Model Updating | 57 |
| 3.4 | Experimental Results | 57 |
| 3.4.1 | Experimental Settings | 57 |
| 3.4.2 | Detailed Analysis | 59 |

| | | |
|--|---|-----------|
| 3.4.3 | Comparison with State-of-the-arts | 62 |
| 3.4.4 | Robustness in the Wild | 64 |
| 3.5 | Conclusion | 65 |
| Chapter 4 PurifyNet: A Robust Person Re-Identification Model with | | |
| | Noisy Labels | 66 |
| 4.1 | Introduction | 66 |
| 4.2 | Related Work | 69 |
| 4.3 | Proposed Method | 70 |
| 4.3.1 | Baseline Model | 71 |
| 4.3.2 | Joint Learning with Label Refinement | 72 |
| 4.3.3 | Hard-aware Instance Re-weighting | 73 |
| 4.4 | Experimental Results | 76 |
| 4.4.1 | Datasets and Experimental Settings | 76 |
| 4.4.2 | Self Evaluation | 78 |
| 4.4.3 | Comparison with State-of-the-Arts | 81 |
| 4.4.4 | Application in Unsupervised Re-ID | 82 |
| 4.4.5 | Application in Video Re-ID | 82 |
| 4.5 | Conclusion | 83 |
| Chapter 5 Bi-directional Dual-Constrained Top-Ranking for Visible | | |
| | Thermal Person Re-Identification | 85 |
| 5.1 | Introduction | 85 |
| 5.2 | Related Work | 89 |
| 5.3 | Proposed Method | 91 |
| 5.3.1 | Dual-path Network | 91 |
| 5.3.2 | Bi-directional Dual-Constrained Top-Ranking | 93 |
| 5.3.3 | Bi-directional Center-constrained Top-Ranking | 96 |
| 5.3.4 | Online Batch Sampling Training Strategy | 99 |
| 5.4 | Experimental Results | 99 |

| | | |
|---|--|------------|
| 5.4.1 | Experimental Settings | 99 |
| 5.4.2 | Comparison With Baseline Methods | 102 |
| 5.4.3 | Comparison with the State-of-the-arts | 104 |
| 5.4.4 | Further Evaluation and Analysis | 106 |
| 5.5 | Conclusion and Future Work | 111 |
| Chapter 6 Unsupervised Embedding Learning via Invariant and Spreading Instance Feature | | 113 |
| 6.1 | Introduction | 113 |
| 6.2 | Related Work | 116 |
| 6.3 | Proposed Method | 117 |
| 6.3.1 | Instance-wise Softmax Embedding | 118 |
| 6.3.2 | Softmax Embedding on ‘Real’ Instance Feature | 119 |
| 6.3.3 | Rationale Analysis | 121 |
| 6.3.4 | Training with Siamese Network | 121 |
| 6.4 | Experimental Results | 122 |
| 6.4.1 | Experiments on Seen Testing Categories | 122 |
| 6.4.2 | Experiments on Unseen Testing Categories | 126 |
| 6.4.3 | Ablation Study | 129 |
| 6.4.4 | Understanding of the Learned Embedding | 131 |
| 6.5 | Conclusion | 132 |
| Chapter 7 Conclusions | | 133 |
| 7.1 | Summary | 133 |
| 7.2 | Future Work | 136 |
| Bibliography | | 138 |
| CURRICULUM VITAE | | 163 |