# Viewpoint-Aware Loss with Angular Regularization for Person Re-Identification

**Poster** · November 2020



1 author:

Zhu Zhihui
PINDUODUO

# Viewpoint-Aware Loss with Angular Regularization for Person Re-Identification

Zhihui Zhu[12], Xinyang Jiang[1], Feng Zheng[1], Xiaowei Guo[1], Feiyue Huang[1], Weishi Zheng[2], Xing Sun[1]
[1]Tencent Youtu Lab, Shanghai, China
[2]Sun Yat-sen University, Shenzhen, China
[2]Southern University of Science and Technology, Shenzhen, China

## Introduction

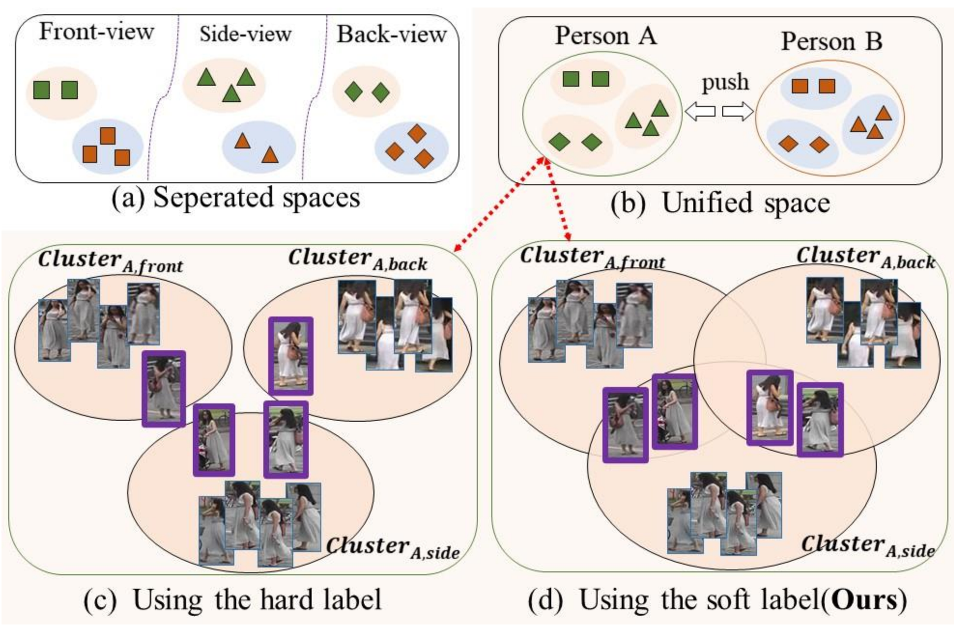


Fig.1 Comparisons of different methods

**Motivation**

- Viewpoint variations in this work, one of the most important and difficult challenges for Re-ID research and practical application.
- Existing methods treat viewpoint learning and identity discrimination as two separate progresses. In such a case, it is not a principle way to learn optimal identity classification under various viewpoint variations.
- Existing methods cast viewpoints of persons as hard labels, while in reality the viewpoint of probe image is ambiguous.

**To address these issues, we in this paper**

- ✓ propose a novel Viewpoint-Aware Loss with Angular Regularization(**VA-reID**), which projects the feature from different viewpoints into a unified hyper-sphere and effectively models the distribution on both identity-level and viewpoint-level.
- ✓ introduce a soft label learning method called viewpoint-aware adaptive label smoothing method (VALSR) for obtaining adaptive soft label for identity-level and viewpoint-level feature representation.

## Adaptive Label Smoothing

Given a training image with identity label $y_i$, the softmax loss of classification is

$$L_y = \frac{1}{K}\sum_{j}^{K} p_j log(q(j))$$

where p is the classification label,

$$p_j = \begin{cases} 1 & \text{if } i = y_i \\ 0 & \text{otherwise} \end{cases}$$

An improvement of labelling is LSR(label smoothing):

$$p_j = \begin{cases} 1-\varepsilon & \text{if } j = y \\ \dfrac{\varepsilon}{K-1} & \text{otherwise} \end{cases}$$

where ε is a manual value. However, this results in the same expected probability of ground-truth category for every input sample and so do other categories. A better idea is the adaptive label:

$$p_j = \begin{cases} 1-\alpha(1-q(j)) & \text{if } j = y \\ \dfrac{\alpha(1-q(j))}{K-1} & \text{otherwise} \end{cases}$$

The new smoothing parameter ε= α(1-q(j)) is related to the prediction probability of network. α is a multiplicative scaling coefficient.

## Framework

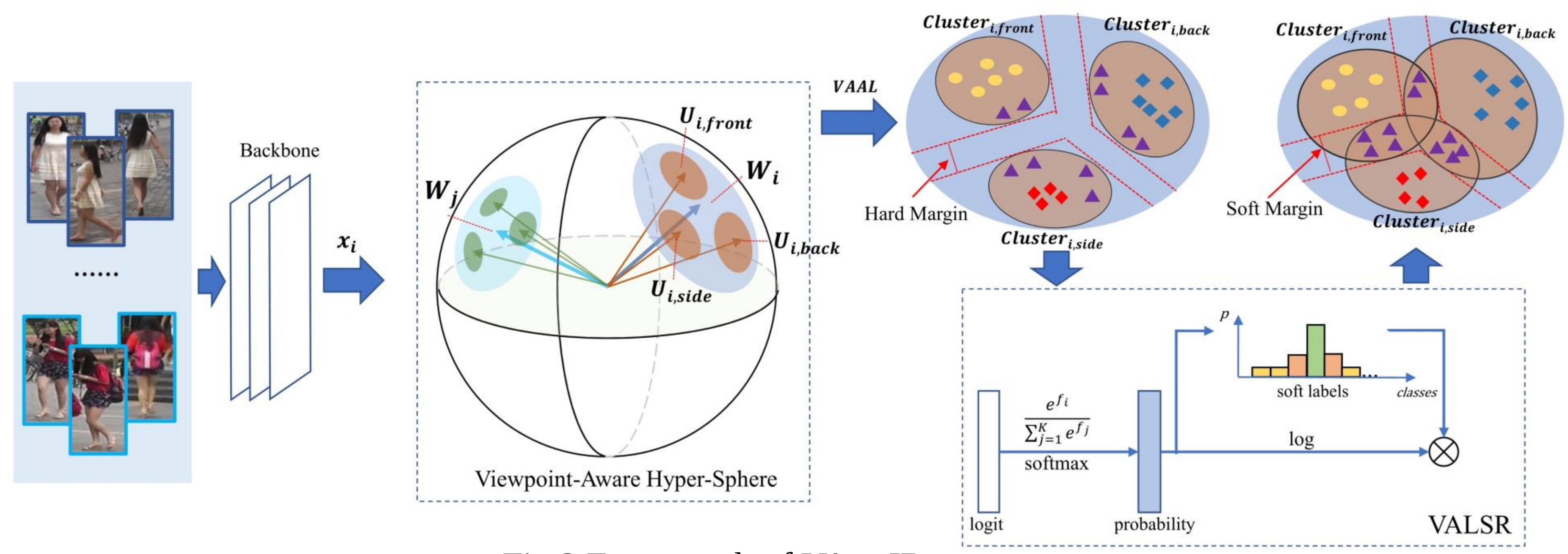


Fig.2 Framework of VA-reID

- **Viewpoint-Aware Adaptive Label Smoothing.**

In order to model the distribution of different viewpoints, each identity class is further classified into 3 subclass, corresponding to 3 viewpoints (i.e., front, side and back viewpoint). The softmax loss of viewpoint classification is

$$L_v = \frac{1}{K}\sum_{j}^{K}\sum_{k}^{3} t_{j,k} log(r(j,k))$$

Similarly, we apply the Viewpoint-Aware Adaptive Label :

$$t_{jk} = \begin{cases} 1-\varepsilon_1-\varepsilon_2 & \text{if } j = y_i, k = v_i \\ \dfrac{\varepsilon_2}{2} & \text{if } j = y_i, k \neq v_i \\ \dfrac{\varepsilon_1}{K-3} & \text{otherwise} \end{cases}$$

where $\varepsilon_1 = \alpha(1 - \sum_{v=1}^{3} r(y_i, v))$

$\varepsilon_2 = \alpha(1 - r(y_i, v_i))$.

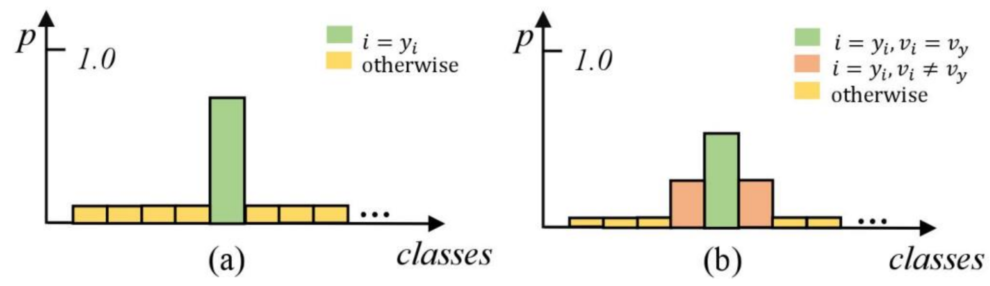


Fig.3 Illustration of adaptive soft label

- **Overall loss.**

We models the distribution on both identity-level and viewpoint-level. The overall loss:

$$L_{va} = L_y + L_v + \beta L_R$$

Specifically, to maintain the visual similarity between features from the same person but with different viewpoints, we propose to add an extra regularization term to pull the 3 viewpoint centers $W_{ij}$ closer to the corresponding identity center:

$$L_R = \sum_{k=1}^{K}\sum_{j=1}^{3} \frac{W_j^T \cdot U_{jk}}{||W_j|| \cdot ||U_{jk}||}$$

## Results

| Category | Method | Market-1501 | | | DukeMTMC-reID | | |
|---|---|---|---|---|---|---|---|
| | | mAP | Rank-1 | Rank-5 | mAP | Rank-1 | Rank-5 |
| stripe based | PCB (Sun et al. 2018) | 77.4 | 92.3 | 97.2 | 66.1 | 81.7 | 89.7 |
| | PCB+RPP (Sun et al. 2018) | 81.6 | 93.8 | 97.5 | 69.2 | 83.3 | 90.5 |
| | MGN (Wang et al. 2018) | 86.9 | 95.7 | - | 78.4 | 88.7 | - |
| attention based | HA-CNN (Li, Zhu, and Gong 2018) | 75.7 | 91.2 | - | 63.8 | 80.5 | - |
| | ABD-Net (Chen et al. 2019) | 88.28 | 95.60 | - | 78.59 | 89.00 | - |
| human parsing | SPReID (Kalayeh et al. 2018) | 83.36 | 93.68 | 97.57 | 73.34 | 85.95 | 92.95 |
| | DSA-reID (Zhang et al. 2019) | 87.6 | 95.7 | 98.4 | 74.3 | 86.2 | |
| metric learning | Pyramid (Zheng et al. 2019a) | 88.2 | 95.7 | 98.4 | 79.0 | 89.0 | - |
| | SRB(ResNet50) (Luo et al. 2019) | 85.9 | 94.5 | - | 76.4 | 86.4 | - |
| | SRB(SeResNext101) (Luo et al. 2019) | 88.0 | 95.0 | - | 79.0 | 88.4 | - |
| | HPM (Fu et al. 2019) | 82.7 | 94.2 | 97.5 | 74.3 | 86.6 | - |
| pose/view related | OSCNN (Chen et al. 2018) | 73.5 | 83.9 | - | - | - | - |
| | PDC (Su et al. 2017b) | 63.41 | 84.14 | - | - | - | - |
| | PN-GAN (Qian et al. 2018) | 72.58 | 89.43 | - | 53.20 | 73.58 | - |
| | PIE (Zheng et al. 2019b) | 69.25 | 87.33 | 95.56 | 64.09 | 80.84 | 88.30 |
| | PGR (Li et al. 2019) | 77.21 | 93.87 | 97.74 | 65.98 | 83.63 | 91.66 |
| **This work** | **Ours** | **91.70** | **96.23** | **98.69** | **84.51** | **91.61** | **96.23** |
| | **Ours+reranking** | **95.43** | **96.79** | **98.31** | **91.82** | **93.85** | **96.50** |

Tab.1 Performance comparisons to the SOTA results on Market-1501 and DukeMTMC-reID.

We evaluate our proposed VA-reID model with the state-of-the-art models based on deep learning. Our proposed VA-reID model outperforms the state-of-the-art methods.

Tab.2 Ablation Study

| Method | Market-1501 | | DukeMTMC-reID | |
|---|---|---|---|---|
| | mAP | Rank-1 | mAP | Rank-1 |
| Xent | 86.30 | 94.31 | 76.70 | 86.94 |
| LSR | 86.72 | 94.35 | 77.47 | 87.43 |
| $L_y$(baseline) | 86.97 | 94.7 | 77.99 | 87.39 |
| $L_y$+$L_v$ | 88.67 | 95.37 | 78.86 | 88.38 |
| $L_y$+$L_R$ | 88.25 | 95.25 | 78.25 | 87.84 |
| VA-reID | 89.97 | 95.87 | 81.48 | 91.11 |
| VA-reID+RR | 95.09 | 96.32 | 90.66 | 92.46 |
| VA-reID+local | 91.70 | 96.23 | 84.51 | 91.61 |
| VA-reID+local+RR | 95.43 | 96.79 | 91.82 | 93.85 |

Fig.4 Visualization