Inferability-guided Learning for Person Re-identification

Yunjeong Choi Seungwoo Jung Jinheock Choi
Seoul National University
{racheal0, tmddn833, jinheock}@snu.ac.kr

Abstract

Person re-identification (ReID) focuses on identifying a specific individual from a collection of images taken by different cameras. Although significant progress has been made, current methodologies apply same strength of supervision to all galleries regardless of inferability which can lead models to overfit on domain-specific data. To address this, we introduce a novel inferability-guided learning approach that adjusts the supervision intensity based on the inferability of each viewpoint. We propose two techniques for measuring inferability: one assessing the angle of visibility and another evaluating the length of the overlapping visible arc modeled as an ellipse. These techniques help the model emphasize reliably inferable features, thereby reducing dependency on uncertain information. Experimental results show that our method aligns the feature space with inferability, enhancing interpretability and explainability. However, this alignment does not necessarily improve generalization performance in ReID tasks, indicating a trade-off between using reliable and potentially useful but domain-specific data. Our contributions offer a versatile framework for viewpoint-based supervision, aligning feature space with infearability which further improves model interpretability and explainability.

1. Introduction

Person re-identification(ReID) aims to retrieve a queried person from a set of images captured by disjoint cameras. With the significant advancements in visual information extraction, person ReID models have achieved impressive performance [2, 4, 9, 25]. However, it grapples with challenges such as limited datasets due to privacy concerns, resulting in fewer resources compared to other domains [10, 11, 15, 17]. This scarcity, coupled with dataset distribution disparities, poses complexities in training models for domain generalizable (DG) ReID. Despite these hurdles, substantial research [7, 18, 20, 22] is being conducted to enable models to exhibit robust performance across diverse domains despite being trained on a single domain.

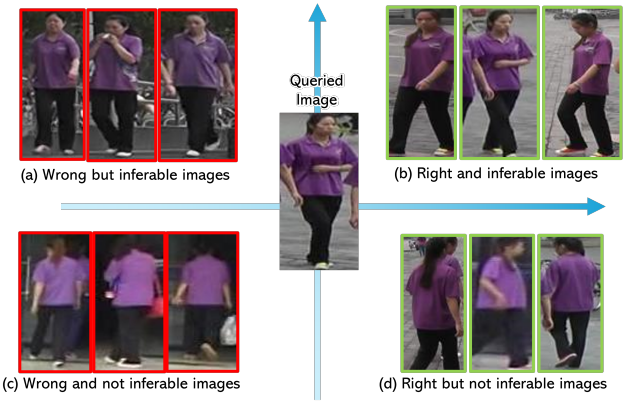


Figure 1. **Plausibility and Inferability.** All the photos in this figure are from the gallary that are "plausible" matches to the queried image. These photos can be categorized as to whether they match the queried image (left or right) or whether they can be inferred from the queried image (top or bottom). In this case, even though the human reidentification model is trained using images with matching identities, it learns the back that cannot be inferred from the front as a ground truth pair. As a result, the model will proceed with supervised learning without sufficient evidence when matching people, and it will learn domain-specific data.

However, we observed that the prevalent supervised learning scheme in current ReID systems is inevitably prone to overfit on training domain. As depicted in Fig. 1, there exists an inherent limitation in inferring the opposite side (*i.e* 180° difference, such as in front and back), due to the absence of visual cues from the opposite viewpoint. For instance, when the queried person is facing the camera, it is impossible to precisely determine the *correct* back view image, though it may be *plausible*. That is, not all images in the ground truth gallery can be inferred in the same manner.

This intrinsic problem of having multiple plausible galleries for a single viewpoint exacerbates the tendency of models to overfit on the training domain. Since viewpoints that are nearly impossible to infer from a given image lack sufficient visual cues, it becomes inevitable to rely on the data used during training. Continuously applying supervision to only the opposite view available within the train-

ing domain compels the models to adopt to domain-specific data. This leads to limit the ability to generalize in unseen domains, as it relies on the assumption that a *single* designated opposite view is correct.

A natural question that arises from this observation is: *Could we reduce dependency on the training domain by relying less on non-inferable information and more on inferable information?* To address this issue, we propose a novel inferability-guided learning scheme. Our approach modifies the supervision strategy by weighting the inferential possibility of each viewpoint. Views that are almost impossible to infer receive lighter supervision, acknowledging the higher uncertainty and variability in these cases. Conversely, viewpoints that are strongly inferable are subjected to stronger supervision. This method not only reduces the risk of overfitting by not forcing the model to learn from highly uncertain or inaccurate assumptions but also enhances the generalization capabilities across diverse arbitrary domains.

To achieve this, we propose two methodologies for quantifying *inferability* using the heading direction of images inferred by 3D body keypoints. The first method evaluates the degree of overlap based on the angle visible to the camera. This approach considers how much of the person's body is visible from the camera's perspective and calculates inferability periodically based on the extent of this overlap. The second method models human body as an ellipse, and quantifies inferability by measuring the length of the overlapping arc of the exposed surface of the person as seen from the camera. By assessing the proportion of the person's surface area that is visible and comparing it across different viewpoints, we can determine the inferability more precisely.

These two quantification methods provide a robust framework for guiding the supervision process, ensuring that the model focuses more on reliably inferable features while avoiding overreliance on uncertain and non-inferable information. Our experiments demonstrate that injecting our inferability-guided supervision methodology aligns the feature space with inferability. Specifically, within the same person's galleries, images with similar viewpoints are closer to each other than those with opposite viewpoints.

However, thorough experiments revealed that aligning the feature space with inferability does not necessarily improve generalization performance, particularly in person ReID tasks. This may be attributed to the trade-off involved in utilizing less information from the galleries, which, although potentially unreliable in other domains, could be beneficial for generalization. Nonetheless, our work significantly contributes to creating a feature space that is more coherently aligned with viewpoints, thereby enhancing the interpretability and explainability of the model.

Our major contributions are:

- *We propose a novel inferability-guided learning scheme that adjusts the supervision strategy based on the inferability of each viewpoint.*
- *We introduce two novel methods for quantifying inferability, providing a versatile metric that can be applied to any viewpoint-related images.*
- *We demonstrate that our inferability-guided supervision methodology aligns the feature space with inferability, resulting in images with similar viewpoints being closer to each other within the same person's galleries. This alignment enhances interpretability and explainability, offering global applicability.*

2. Related Work

Closed-World Person Re-Identification. Closed-World Person Re-Identification focuses on identifying individuals within a controlled, experimental setting, where it is assumed that all test subjects are included in the training database. This setting typically utilizes images or videos captured by fixed surveillance cameras, with data clearly represented within well-defined bounding boxes. [24] significantly enhanced the efficiency of Re-ID systems by integrating deep learning techniques within the closed-world setting, influencing subsequent developments in Re-ID technologies profoundly.

Many works have been conducted on feature representation learning, deep metric learning, and ranking optimization. [12] proposed effective methods for combining global and local features to enhance accuracy in identity recognition. Additionally, [14] tackled the limitation of relying solely on visual cues by incorporating attribute-based learning to enrich the feature set with semantic information.

In the area of metric learning, [8] utilized triplet loss functions to enable more sophisticated identity distinctions in Re-ID systems. This approach significantly enhances system performance by better capturing the complex variations in human appearance than conventional methods. For ranking optimization, the angular loss-based metric learning proposed by [1] proved effective. These technologies contribute importantly to accurately identifying and matching query subjects within a closed-world setting. [19] specifically designed to improve the ranking process by focusing on embedding distances. This methodology not only optimizes the accuracy of the top-ranked retrieval results but also ensures that the differences in the feature space translate into meaningful distinctions between different identities. By effectively reducing the intra-class variability while maximizing the inter-class differences, this approach has set a new standard in ranking optimization, ensuring that the system's responses are both relevant and reliable.

Open-World Person Re-Identification. Open-world person re-identification (Re-ID) focuses on identifying indi-

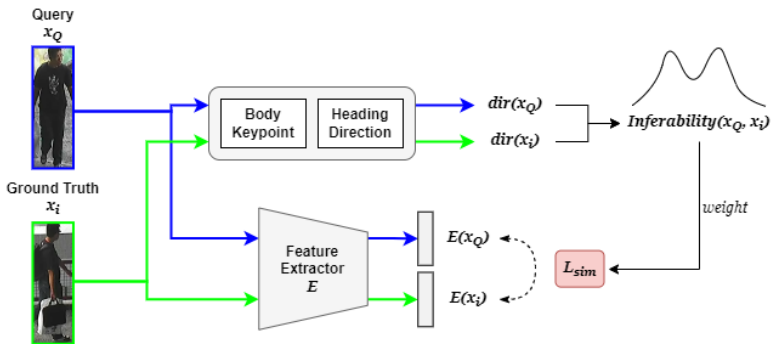


Figure 2. **The Overall Architecture of Proposed Training Scheme.** Given a query image, the pose-conditioned generator G creates a novel plausible opposite-viewpoint images. The augmented images, along with the original ground truth x_i and query x_Q, are processed by the feature extractor E, which discriminative features. The extracted features are then utilized in the inferability-guided learning process, where the similarity between the query and ground truth images is weighted by their inferability score, derived from the difference in their heading directions as determined by body keypoints. The entire process aims to reduce reliance on training domain via enriched the visual cues of the non-inferable viewpoints and adjusted supervision.

viduals in unpredictable environments, where it is required to recognize persons not previously included in the training dataset. Unlike closed-world scenarios, open-world Re-ID faces significant variability due to changes in background, lighting, clothing, and appearance. [6] discuss how these systems can dynamically update their models to incorporate new data as it becomes available, continuously learning and improving. This approach addresses the critical issue of model obsolescence in rapidly changing scenarios.

Additionally, [16] introduce an unsupervised learning approach which enables the identification of individuals without relying on pre-labeled data. This is particularly valuable in open-world settings where obtaining comprehensive labeled datasets is impractical. The system adapts using pseudo-labels generated from the data itself, refining these labels as more data becomes available. [3] present a novel strategy that combines multi-domain learning techniques with a hardness-aware loss function. This function effectively distinguishes between in-domain and out-of-domain data, enhancing the robustness of the Re-ID system against the diversity of the open world. These methodologies significantly contribute to bridging the gap between the controlled environments of closed-world Re-ID and the unpredictable nature of open-world scenarios.

3. Methods

problem formulation. Given a set of images $X = \{x_1, x_2, \ldots, x_n\}$ and corresponding labels $Y = \{y_1, y_2, \ldots, y_n\}$, the goal of person ReID is to identify the most similar image $\hat{x}$ from X that matches the query x_Q, excluding the query itself ($\hat{x} \neq x_Q$). This can be formalized as finding the image that maximizes the similarity with respect to the query:

$$\hat{x} = \arg \max_{x \in X \setminus \{x_Q\}} sim(E(x), E(x_Q)) \tag{1}$$

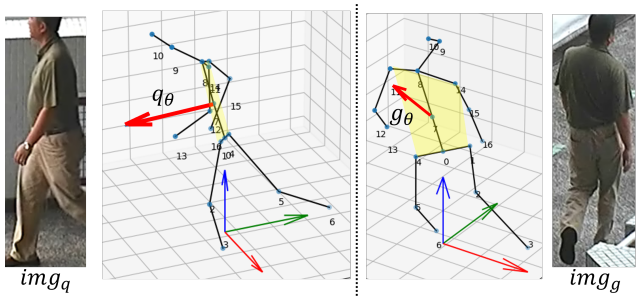
where E is the feature extractor that maps an image tod a feature space. In this work, we aim to train a feature extractor E to extract more similar features for mutually inferable targets. To achieve this, we model inferability estimation based on heading direction and reflect it in triplet loss to perform inferability guided learning. Proposed architecture is illustrated in Fig. 2, described in the following sections. Our methodology can be used as an add-on to algorithms that typically use distance between features in Person Re-Identification, regardless of the structure of the feature extractor.

3.1. Inferability Estimation

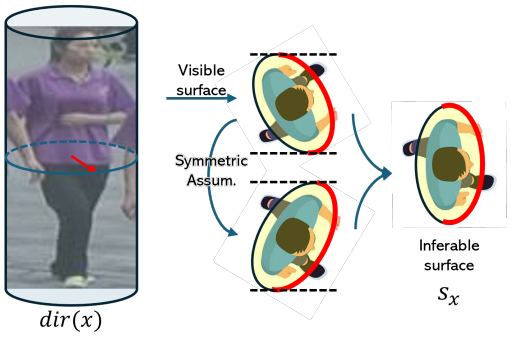
As the first step to quantify the *inferability* between two images, we need to determine the heading direction of it. To do so, the 3D keypoints are estimated by a pose estimation algorithm [5]. Out of all the 3D keypoints found, seven key body points are used to determine the heading angle: the left and right shoulders, hips, and a point on the 2D centerline. These keypoints, represented as $K = \{\{p_j\}_{j=1}^{7}\}_{i=1}^{n}$. Using these points, a plane is found using least square fitting, and the normal vector of this plane is defined as the heading vector $dir(x_i)$.

$$dir(x_i) = arg \min_{v \in \mathbb{R}^3} \|K_i^T v - \mathbb{1}\|_2 \tag{2}$$

The heading vector is projected onto the 2D plane to get the relative angle between the image view frame and the direction of the person. In addition, a possible assumption for the inferability aspect of Persion re-id is that people are

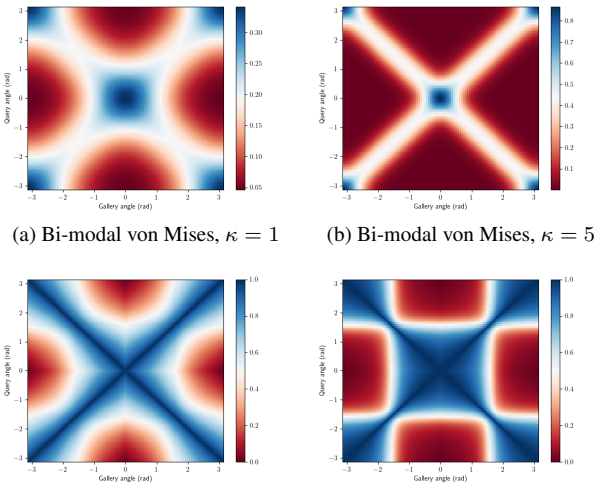


(a) Heading direction with pose estimation.



(b) Infereble surface with symmetric assumption.

Figure 3. **Illustration of Inferable Surface Estimation.** Using MMPose [5], the 3D coordinates of a keypoint on the torso that determines the orientation of the body is obtained as shown in (a). The heading vector(bold red arrow) is from normal vector of plane fitted by least square from the keypoint on the torso. In (b), the infeasible surface (red curve) is found through the heading angle (Dir(x)) and symmetrical assumption of a person body. Images are from MSMT17 [21].



(a) Bi-modal von Mises, $\kappa = 1$ (b) Bi-modal von Mises, $\kappa = 5$

(c) Jaccard coefficient, $b/a = 1.5$ (d) Jaccard coefficient, $b/a = 3.5$

Figure 4. **Inferability with relative heading direction.** Each image shows the values for two inferability designs that satisfy periodicity and symmetry, given the two heading directions that we want to compare. For each design, there are parameters kappa and b/a that determine how sensitive it is to angular differences. The larger these values are, the more sensitive the variation is to the difference between the two angles.

symmetrical on the left and right. In Figure 3b, the infeasible surface is found through the previously obtained heading angle using the assumption for symmetrical person. To formalize inferability, there are two conditions that must be satisfied, symmetry and periodicity. Symmetry reflects the symmetry of the left and right sides of a person, and periodicity is required to show a relationship between two heading directions. To reflect these characteristics, we design the bi-modal von Mises distribution and the Jaccard coefficient of the inferable surface with the assumption that the person have shape of ellipse pole. The values for each of the designed formulas are depicted in Figure 4.

Bi-modal von Mises distribution Since the inferability must be in the form of a periodic function for the heading direction, we use the von Mises probability density function, which is a probability function with a periodic form as the basis function. Also, it should follow a symmetric distribution with respect to 0 degrees, so we define it as follows:

$$f(x \mid \mu, \kappa) = \frac{\exp(\kappa \cos(x - \mu))}{2\pi I_0(\kappa)}$$
$$inf(x_i, x_j) = \frac{1}{2} f(x_j \mid dir(x_i), \kappa) + \frac{1}{2} f(x_j \mid flip(dir(x_i)), \kappa)) \quad (3)$$

where $dir(x_i)$ and $dir(x_j)$ represent the heading directions of the respective images. A larger difference in heading directions implies lower *inferability*, due to less overlap in visual cues. This metric quantitatively assesses the challenge in inferring one image from the another when their orientations diverge significantly. κ is a hyperparameter that controls the sensitivity of the supervision strength to differences in heading directions. A higher value of κ results in a sharper decrease in weight for larger directional differences, thus mitigating the impact of highly uncertain matches on the learning process.

Jaccard coefficient with ellipse Given the person's heading direction $dir(x_i)$, as shown in figure 3b, we can find the area visible in the view frame. By flipping this with respect to the person's front axis, we can get the entire inferable surface, which is called s_{x_i}. Similarly, for the inferable surface with x_j, we get another s_{x_j}. The more intersections of inferable surfaces, the more inferable it is. Therefore, the Jaccard coefficient, which is the ratio of the intersection to the union of s_{x_i} and s_{x_j}, is used to define inferability as

equation 4. The $\mid s_x \mid$ can be simply computed by length of ellipse with a ratio b/a between the major(b) and minor(a) axes.

$$inf(x_i, x_j) = \frac{\mid s_{x_i} \cap s_{x_j} \mid}{\mid s_{x_i} \cup s_{x_j} \mid} \tag{4}$$

Similar to von Mises' κ, the larger the value of b/a, the more sensitive it is to angular disparity. This allows you to control the sensitivity between heading direction differences and inferability between two images.

3.2. Inferability-guided learning

In contrast to the conventional supervised learning scheme where all ground truths equally contributes during training, our proposed inferability-guided learning incorporates the *inferability* to adjust the contribution. That is, non-inferable viewpoints are weakly supervised while highly-inferable viewpoints are strongly supervised.

Learning Objective.

$$L_{sim} = \sum_{(i,j),i \neq j} inf(x_i, x_j) \cdot l(E(x_i;\theta), E(x_j;\theta), y_i, y_j) \tag{5}$$

where $inf(\cdot,\cdot)$ is *inferability* function of either Eq. (3) or Eq. (4), E is a feature extractor, l is conventional loss function in person ReID, and y_i is an identity label of x_i. Here, the loss function was defined as triplet loss.

This dynamic supervision strategy ensures that pairs of images with high inferability contribute more to the model's training, encouraging the learning of robust and generalizable features. Conversely, pairs with low inferability have a reduced impact, preventing the model from overfitting to ambiguous and less informative data.

4. Experiments

4.1. Datasets and Evaluation Metrics

Baseline. To evaluate our proposed learning scheme, we chose PAT [18], the state-of-the-art generalizable person re-id model as our baseline.

Datasets. We conducted experiments with the three most commonly used datasets (Market1501 [23], MSMT17 [21], CUHK03-NP [13]) to evaluate the performance of proposed algorithm both in various environments and outside of the training domain. Additionally, we have restructured the existing dataset by adding 3D pose information in two ways: discretely and continuously. We used the dataset with discretely added pose information to determine model overfitting, while the dataset with continuously added pose information was used for training the model.

Evaluation. To ensure a fair comparison with baseline algorithm [18], we will evaluate performance using mAP and R1-score, traditionally used metrics in person re-identification tasks. We will also use a newly processed dataset—where query-groundtruth pairs that cannot be inferred are no longer recognized as groundtruth pairs—to assess whether each algorithm is overfitting on the training data.

Train Domain	Market1501 [23]					
Test Domain	Market1501 [23]		CUHK03-NP [13]		MSMT17 [21]	
Metric	R1	mAP	R1	mAP	R1	mAP
baseline	**92.5**	**81.4**	**27.2**	**25.9**	**39.6**	**16.3**
kappa = 1	90.4	75.0	19.8	20.0	30.9	12.1
kappa = 3	75.4	52.0	4.9	5.4	11.7	3.8
kappa = 4	90.6	75.0	11.3	12.4	27.4	10.0
kappa = 5	89.2	72.4	16.7	17.3	29.8	11.1
kappa = 6	89.5	74.5	20.6	21.3	34.2	13.8
kappa = 7	90.5	75.5	20.6	21.2	33.0	13.2
kappa = 8	89.5	73.9	16.7	17.2	29.7	11.4

Table 1. **Performance comparison for different values of kappa.** The shaded values represent in-domain evaluations, while the others are assessed in cross-domain.

4.2. Viewpoint based Inferability

Distribution Modeling. Using a periodic Von Mises Distribution, we modeled the inferability of objects in images based on their viewpoints. Since the hyperparameter kappa is crucial in the Von Mises Distribution, and we conducted ablation study with different kappa values to find the optimal one. As shown in Tab. 1, the selected range of kappa values is between 4 and 7. When employing these kappa values, the inferability becomes zero if the viewpoint difference between images of the same identity exceeds approximately 100 degrees, which is intuitively reasonable. Injecting the inductive bias of that humans are mostly symmetrical, we used another Von Mises Distribution that is symmetrical to 0 degrees along with the generated distribution. (We set the camera lens direction to 0 degrees, with counterclockwise rotation being positive, and represented the rotation angle from -180 degrees to 180 degrees.)

Train Domain	Market1501 [23]					
Test Domain	Market1501 [23]		CUHK03-NP [13]		MSMT17 [21]	
Metric	R1	mAP	R1	mAP	R1	mAP
baseline	92.5	**81.4**	**27.2**	**25.9**	39.6	16.3
min = 0.2	90.7	76.2	21.3	20.9	35.6	14.3
min = 0.5	92.2	79.2	24.3	24.4	40.5	**16.8**
min = 0.8	**92.8**	79.8	25.3	24.8	**40.6**	**16.8**

Table 2. **Performance comparison for different rescaling values (kappa=4).**

For galleries deemed non-inferable due to a direction difference exceeding 100 degrees, applying the von Mises

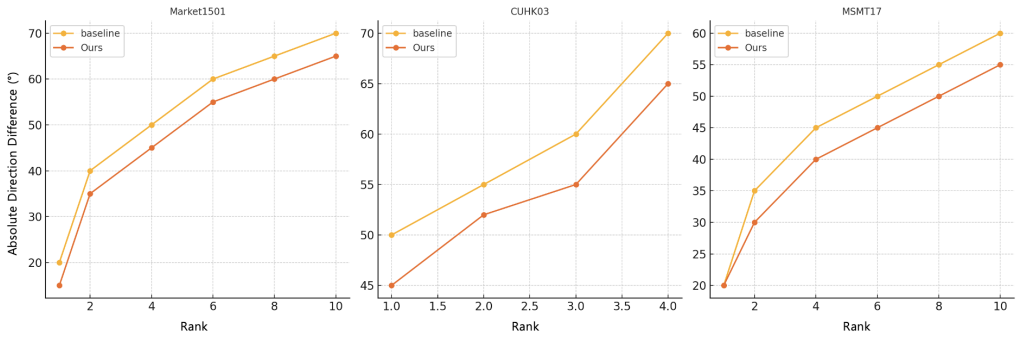


Figure 5. **Average difference in heading direction between query and gallery.** The results show the average angular difference between the query image and the images ranked by an algorithm trained on the Market1501 dataset with kappa = 5 and minimum inferability of 0.1. Overall, as the rank increases, the angular difference also increases. Although the angular difference is larger in the untrained domain compared to the trained domain, it still shows an increasing trend with higher ranks. The reason for different ranks across datasets is due to the varying number of gallery images matching the query in each dataset.

distribution would set inferability to a value close to zero. However, this approach is inappropriate as it completely severs the relationship between queries and galleries with a direction difference greater than 100 degrees. Therefore, the minimum value of inferability was experimentally determined and applied as shown in Tab. 2.

Viewpoint-aware Ranking. We trained our model to prioritize images with smaller viewpoint differences from the query image, among images with the same identity. As shown in Fig. 5, the viewpoint difference between the images and the query increases with rank. Notably, the average direction difference within the same rank is reduced in our method compared to the baseline. This demonstrates that our inferability-guided supervision effectively aligns the feature space with inferability, thereby ranking more inferable images higher. The feature space, enriched with inferability, is more explainable and interpretable. The persistence of this tendency in cross-domain evaluation indicates that our algorithm can consider inferability when learning or performing inference in an open-world setting.

4.3. Domain Generalizability

In-Domain and Cross-Domain Evaluation.

Train Domain	Market1501 [23]					
Test Domain	Market1501 [23]		CUHK03-NP [13]		MSMT17 [21]	
Metric	R1	mAP	R1	mAP	R1	mAP
baseline	92.5	**81.4**	**27.2**	**25.9**	39.6	16.3
Bi-modal von Mises distribution	**92.8**	79.8	25.3	24.8	**40.6**	**16.8**
Jaccard coefficient with ellipse	89.9	75.0	21.5	21.7	32.8	13.4

Table 3. **Performance comparison for different inferability measure.**

For the inferability measures we proposed, bi-modal von Mises distribution method was superior to the jaccard coefficient with ellipse method. (Tab. 3 Unfortunately, in all the experiments conducted in sections 4.2 and 4.3, no single model significantly outperformed the baseline across all metrics. Further research is needed to achieve more generalizable performance.

4.4. Time Consumption

Calculating inferability requires the extraction of 3D body keypoints, which can be time-consuming. However, note that inferability is applied only during training and not during inference. As a result, while the training process takes approximately five times longer for the same architecture, the inference time remains unaffected. Additionally, by pre-extracting keypoints for the training dataset, the overall training time can be unaffected.

	Online Train	Pre-extract Train	Inference
Ours	5x	1x	1x

Table 4. **Time consumption comparison.** Inference time consumption remains unaffected.

5. Conclusion

We proposed a novel inferability-guided learning approach for generalizable person re-identification. By leveraging differences in heading direction, our method encourages the model to focus less on domain-specific knowledge and more on domain-general inferable knoweldge. Although the performance improvement in cross-domain settings was not significant, our approach effectively constructs an inferability-enriched feature space. This enhanced feature space provides greater interpretability and explainability, as demonstrated in our experiments.

6. Limitations and Future Work

It was challenging to significantly improve performance with partial modifications to the baseline algorithm's loss function, rather than a comprehensive overhaul. While we hoped that our modified loss would balance well with the existing loss function during training, it was insufficient. Therefore, for future work, we believe that proposing a new overall structure and redefining the loss function could lead to groundbreaking performance improvements.

The overfitting issue we identified—specifically, the problem of finding rear images based only on front images—could result in a decline in algorithm performance in real-life open-world scenarios. This issue arises because training and testing are both conducted within limited domains, and there is no metric for overfitting. To clearly identify this overfitting phenomenon, it is necessary to sample similar images from currently available datasets, perform inference among the sampled images, and explore methods to address this problem.

References

[1] Song Bai, Peng Tang, Philip HS Torr, and Longin Jan Latecki. Re-ranking via metric fusion for object retrieval and person re-identification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, page 740–749, 2019. 2

[2] Xiang Bai, Mingkun Yang, Tengteng Huang, Zhiyong Dou, Rui Yu, and Yongchao Xu. Deep-person: Learning discriminative deep features for person re-identification. *Pattern Recognition*, 98:107036, 2020. 1

[3] Slawomir Bak, Peter Carr, and Jean-Francois Lalonde. Domain adaptation through synthesis for unsupervised person re-identification. In *Proceedings of the European conference on computer vision (ECCV)*, page 189–205, 2018. 3

[4] Weihua Chen, Xianzhe Xu, Jian Jia, Hao Luo, Yaohua Wang, Fan Wang, Rong Jin, and Xiuyu Sun. Beyond appearance: a semantic controllable self-supervised learning framework for human-centric visual tasks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 15050–15061, 2023. 1

[5] MMPose Contributors. Openmmlab pose estimation toolbox and benchmark. https://github.com/open-mmlab/mmpose, 2020. 3, 4

[6] Abir Das, Rameswar Panda, and Amit K Roy-Chowdhury. Continuous adaptation of multi-camera person identification models through sparse non-redundant representative selection. *Computer Vision and Image Understanding*, 156:66–78, 2017. 3

[7] Zhaopeng Dou, Zhongdao Wang, Yali Li, and Shengjin Wang. Identity-seeking self-supervised representation learning for generalizable person re-identification. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 15847–15858, 2023. 1

[8] Alexander Hermans, Lucas Beyer, and Bastian Leibe. In defense of the triplet loss for person re-identification. *arXiv preprint arXiv:1703.07737*, 2017. 2

[9] Fabian Herzog, Xunbo Ji, Torben Teepe, Stefan Hörmann, Johannes Gilg, and Gerhard Rigoll. Lightweight multi-branch network for person re-identification. In *2021 IEEE International conference on image processing (ICIP)*, pages 1129–1133. IEEE, 2021. 1

[10] Xuemei Jia, Xian Zhong, Mang Ye, Wenxuan Liu, and Wenxin Huang. Complementary data augmentation for cloth-changing person re-identification. *IEEE Transactions on Image Processing*, 31:4227–4239, 2022. 1

[11] Kajal Kansal and A Venkata Subramanyam. Autoencoder ensemble for person re-identification. In *2019 IEEE Fifth International Conference on Multimedia Big Data (BigMM)*, pages 257–261. IEEE, 2019. 1

[12] Dangwei Li, Xiaotang Chen, Zhang Zhang, and Kaiqi Huang. Learning deep context-aware features over body and latent parts for person re-identification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 384–393, 2017. 2

[13] Wei Li, Rui Zhao, Tong Xiao, and Xiaogang Wang. Deepreid: Deep filter pairing neural network for person re-identification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 152–159, 2014. 5, 6

[14] Yutian Lin, Liang Zheng, Zhedong Zheng, Yu Wu, Zhilan Hu, Chenggang Yan, and Yi Yang. Improving person re-identification by attribute and identity learning. *Pattern recognition*, 95:151–161, 2019. 2

[15] Zhouchi Lin, Chenyang Liu, Wenbo Qi, and Shing-Chow Chan. A color/illuminance aware data augmentation and style adaptation approach to person re-identification. *IEEE Access*, 9:115826–115838, 2021. 1

[16] Zimo Liu, Dong Wang, and Huchuan Lu. Stepwise metric promotion for unsupervised video person re-identification. In *Proceedings of the IEEE international conference on computer vision*, pages 2429–2438, 2017. 3

[17] Niall McLaughlin, Jesus Martinez Del Rincon, and Paul Miller. Data-augmentation for reducing dataset bias in person re-identification. In *2015 12th IEEE International conference on advanced video and signal based surveillance (AVSS)*, pages 1–6. IEEE, 2015. 1

[18] Hao Ni, Yuke Li, Lianli Gao, Heng Tao Shen, and Jingkuan Song. Part-aware transformer for generalizable person re-identification. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 11280–11289, 2023. 1, 5

[19] Florian Schroff, Dmitry Kalenichenko, and James Philbin. Facenet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 815–823, 2015. 2

[20] Shouhong Wan, Peiyi Zhang, Peiquan Jin, and Pengcheng Ding. A part invariance network for cross-domain person re-identification. In *2021 IEEE 33rd International Conference on Tools with Artificial Intelligence (ICTAI)*, pages 575–581. IEEE, 2021. 1

[21] Longhui Wei, Shiliang Zhang, Wen Gao, and Qi Tian. Person transfer gan to bridge domain gap for person re-identification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 79–88, 2018. 4, 5, 6

[22] Yue Zhang, Fanghui Zhang, Yi Jin, Yigang Cen, Viacheslav Voronin, and Shaohua Wan. Local correlation ensemble with gcn based on attention features for cross-domain person re-id. *ACM Transactions on Multimedia Computing, Communications and Applications*, 19(2):1–22, 2023. 1

[23] Liang Zheng, Liyue Shen, Lu Tian, Shengjin Wang, Jingdong Wang, and Qi Tian. Scalable person re-identification: A benchmark. In *Computer Vision, IEEE International Conference on*, 2015. 5, 6

[24] Liang Zheng, Yi Yang, and Alexander G Hauptmann. Person re-identification: Past, present and future. *arXiv preprint arXiv:1610.02984*, 2016. 2

[25] Zhihui Zhu, Xinyang Jiang, Feng Zheng, Xiaowei Guo, Feiyue Huang, Xing Sun, and Weishi Zheng. Aware loss with angular regularization for person re-identification. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pages 13114–13121, 2020. 1