Gait-X: Exploring X modality for Generalized Gait Recognition

Zengbin Wang[1] Saihui Hou[2] Junjie Li[1] Xu Liu[3] Chunshui Cao[3]
Yongzhen Huang[2,3] Siye Wang[1] Man Zhang[1,4*]
[1]Beijing University of Posts and Telecommunications [2]Beijing Normal University
[3]WATRIX.AI [4]Qinghai Institute of Technology

Abstract

Modality exploration has been repeatedly mentioned in gait recognition, evolving from silhouette to parsing, mesh, point clouds, etc. These latest modalities agree that silhouette is less affected by background and clothing noises, but argue it loses too much valuable discriminative information. They seek to retain the strengths of silhouette while extracting more semantic or structural information through upstream estimation for better recognition. We agree with this principle but argue that these upstream estimations are usually unstable and the resulted modalities rely on pre-defined design. Moreover, the crucial aspect of modality generalization remains underexplored. To address this, we propose Gait-X to explore how to flexibly and stably develop a gait-specific generalized X modality from a frequency perspective. Specifically, ***1)*** *We replace upstream estimation with stable frequency decomposition and conduct a comprehensive analysis of how different frequencies impact modality and within-/cross-domain performance;* ***2)*** *To enable flexible modality customization and mitigate the influence of noise and domain variations, we propose to remove irrelevant low-frequency noise and suppress high-frequency domain-specific information to form our X modality;* ***3)*** *To further improve model generalization, we expand the representation across multiple frequencies to guide the model in balancing whole frequencies for enhanced generalization. Extensive experiments on CCPG, SUSTech1K, and CASIA-B datasets show superior generalization.*

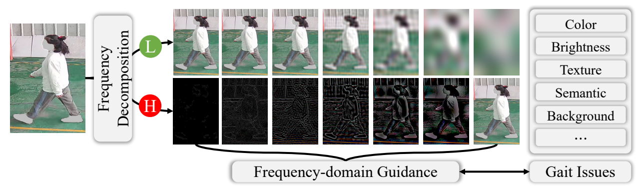


Figure 1. Motivation. Frequency decomposition is a *stable* tool to transform RGB image into *discrete* frequency representations in a *near-lossless* manner. They are closely related to current gait issues and can serve as *guidance* to selectively process irrelevant frequencies and balance them for better recognition and further generalization. L and H indicate low and high frequency.

1. Introduction

Gait recognition identifies individuals through their appearance and unique walking patterns, which has obtained explosive growth in recent years. The long-distance and non-invasive nature of gait sequence data collection makes it reliable and suitable for public security applications [47, 50].

Early appearance-based gait recognition widely adopted binary silhouette modality as input [2, 6, 35, 53]. The main advantage is its ability to eliminate background and clothing texture noise, allowing researchers to focus on body shape and shape variations across frames for recognition. Some early studies have also explored RGB modality as input, leveraging its rich information and ability to feature disentanglement [24, 78, 79]. With cloth-changing becoming the key challenge in gait recognition, they all have recognized limitations in their input modalities: 1) Silhouette-based methods are sensitive to shape variations caused by clothing changes, while the simplistic binary silhouette lacks additional discriminative features [51, 64, 68, 82], making the cloth-changing task more challenging; 2) RGB-based methods can be confused by clothing colors and textures, and have proven to rely on limited unchanged features like head or shoes [13, 33] for recognition.

Recently, gait recognition has turned to more promising modalities, such as parsing [64, 65, 82, 83], 3D mesh [68, 81], point cloud [14, 51], and optical flow [15, 39]. 1) Compared with silhouette, researchers argue that binary silhouette has low information entropy [82], while the added semantics or structural information of these modalities can provide additional discriminative features to improve recognition [51, 64, 81, 82]. 2) Compared with RGB, these modalities help reduce clothing or background noise, and guide the model to focus on more discriminative features like semantics or structure. Additionally, recent stud-

*Corresponding author (zhangman@bupt.edu.cn).

ies have also made initial attempts to obtain intermediate modalities from RGB through learning-based method (*e.g.*, DINO-based gait representation in BigGait [73]). To summarize, the current paradigm in gait modality exploration is trending toward enriching silhouette or denoising RGB.

We fully appreciate these valuable works to build a better gait modality. However, in some cases, we find that these advanced modalities or intermediate modalities often rely on upstream estimations or black-box learning processes, which inevitably suffer from instability or lack of controllability. For example, the upstream estimation results are often pre-defined and variations in estimation methods can lead to significant performance differences in downstream recognition [64, 82]. Another example is that the representations in learning-based methods are often highly coupled with all-purposed information in high dimensions [73]. Furthermore, the critical aspect of modality generalization (*e.g.*, within-domain and cross-domain evaluations) remains underexplored. Thus, a natural question arises: **How can we develop a gait-specific modality with strong generalization in a stable and flexible manner?**

To answer this question, a straightforward idea is to decompose image into multiple discrete components and remove the noise components for better recognition. Fortunately, we find that frequency decomposition (*e.g.*, Discrete Cosine Transformation [30, 84]) is a *stable* tool that enables near lossless transformation of image into multiple frequencies. More importantly, recent frequency analysis in computer vision (*e.g.*, JPEG compression [57]) has shown that different frequencies are strongly correlated with image features such as color or texture. By selectively removing or suppressing certain frequencies, it is possible to control the image's final representation in a *flexible* manner. As in Figure 1, we also find some relations between multiple frequencies, gait modalities, and gait issues. Additionally, recent progress in frequency-domain model training [16, 71] also suggests that properly modulating low and high frequencies is effective in improving model generalization.

Based on the above motivations, we propose **Gait-X** to explore developing a better gait intermediate modality (*i.e.*, X modality) and model training strategy toward generalized gait recognition, all from a frequency-domain decomposition perspective. Specifically, ***1) For gait intermediate modality input***, we find that extremely low frequencies often contain more noises that impact recognition, and higher frequencies contain more within-domain information that is harmful to generalization. *We empirically propose to remove extremely low frequencies and progressively suppress high frequencies to develop a more generalized X modality.* ***2) For gait model training strategy***, we find that the model initially focuses on low frequencies and gradually shifts to high frequencies during training. However, overfitting to high frequencies leads to poor generalization. *We propose a frequency decomposition training strategy to guide the gait model to consider a broader frequency range to mitigate this issue and improve generalization.* Notably, Gait-X balances both within- and cross-domain performance, providing a practical solution toward generalized gait recognition.

The main contributions can be summarized as follows.

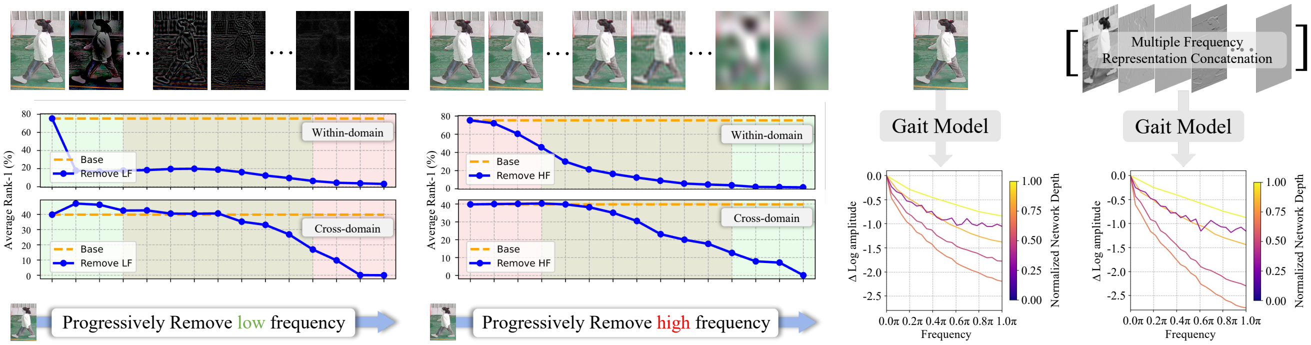
- We pioneer frequency-guided modality exploration and model training strategy in gait recognition, providing a comprehensive analysis of why and how to design strategies toward generalized gait recognition.
- The proposed gait-specific X modality enables stable generation process through flexible frequency modulation, serving as a fresh attempt to enrich gait modality library.
- The proposed frequency decomposition training strategy provides a simple yet effective way to adjust the model's fitting direction to a broader range of frequencies for better generalization.
- Gait-X achieves superior performance in within-domain and cross-domain evaluations. Compared to RGB modality, Gait-X achieves +13.4% in within-domain CCPG and +22.6% in cross-domain CCPG→CASIA-B.

2. Related Work

2.1. Gait Recognition

Modality Exploration in Gait Recognition has been expanded from the widely used silhouette [2, 6, 18–21, 25, 35, 53, 58, 59, 66, 75] to other advanced modalities through upstream estimation models and have proven effective, such as parsing [64, 82, 83, 87], skeleton [9, 55, 56, 77], 3D mesh [68, 81, 85], point clouds [14, 51], optical flow [15, 39], *etc*. One common trait of them is that they inherit two key advantages of silhouette: background noise invariance and clothing texture invariance. Another common trait is that they seek to improve the information entropy of gait by introducing more discriminative information (*e.g.*, semantics, body structure) for better recognition. Some studies have also utilized RGB modality as input and adhering to the above principles. They have explored end-to-end methods [34, 78, 79] to optimize the intermediate gait representation or LVM-based methods to obtain universal representation with further noise refinement [73]. In this paper, Gait-X replaces upstream estimations or LVM with stable frequency decomposition, innovatively constructing a gait-specific X modality through appropriately handling different frequencies.

Frequency Analysis in Past Gait Recognition mainly focuses on temporal frequency analysis to extract gait signals [40–42, 62, 74], such as gait cycle periodicity, step frequency, and spectral variations among individuals. For example, Mobasseri *et al.* [42] use temporal spectrograms to represent micro-Doppler signals of gait. Yu *et al.* [74] demonstrate that key Fourier descriptors (KFDs) from hu-



(a) Impact of Frequencies on Gait Modality and Modality Generalization. (b) Frequency Distributions of Gait Model Layers.

Figure 2. **Different frequencies are highly correlated with gait modality, modality generalization, and model optimization direction.** **(a)** *Top*: Lower frequencies are related to color, brightness, and smooth textures. Higher frequencies contain more imperceptible information. *Bottom*: Within-domain prefers extremely low frequencies and high frequencies. Cross-domain is unaffected by high frequencies, but the removal of extremely low frequency brings more benefits. **(b)** *Top*: Expanding original gait image into multiple frequency bands and obtain their corresponding representations. *Bottom*: Relative log amplitudes of feature map in gait model's different layers show that gait model initially fit low frequency and then shift to high frequency. Taking multiple frequency representation concatenation as input can better suppress high frequencies during the early training. (Other details: The colors in the bottom of (a) indicate low, medium, high frequencies. Within-domain: CCPG, average rank-1 under gait evaluation protocol. Cross-domain: CCPG→CASIA-B, average rank-1.)

man silhouettes provide more discriminative power and reduce computational cost compared to traditional methods. Makihara *et al*. [40] construct a gait silhouette volume (GSV) from infrared gait images, extracting temporal frequency-domain features to capture gait periodicity. In this paper, we shift from temporal frequency analysis toward spatial dimension, and explore to build a frequency-guided gait intermediate modality and a frequency-guided model training strategy for generalized gait recognition.

2.2. Frequency Domain Learning

Frequency-domain learning has been widely used in traditional digital image processing for decades [1, 49]. Recently, it has shown benefits in four main areas by either applying frequency-domain transformations to input images or using frequency as guidance to improve model training, involving: 1) Accelerating training or facilitating the optimization of network [31, 36, 45, 52]; 2) Providing effective data augmentation [16, 23, 69]; 3) Learning informative representations of non-local receptive fields [22, 48]; 4) Assisting in the understanding of certain DNN behaviors [70–72]. In this paper, Gait-X leverages the strengths of frequency decomposition to reveal the relation between frequencies and gait modalities, exploring a better gait intermediate modality and a model training strategy toward generalized gait recognition.

3. Preliminaries

In this section, we propose to explore the impact of different frequencies on gait recognition. Specifically, we investigate how to manipulate frequency at gait modality (input level) and gait model training strategy (training level) toward better generalized gait recognition. We meticulously adopt Discrete Cosine Transform (DCT) [30, 84] as the frequency decomposition tool for following analysis.

3.1. How Frequency Impact Gait Modality?

For gait modality representation, as shown in the top of Figure 2(a), we progressively remove low or high frequencies to observe the changes in the resulting gait modality. Some intuitive findings show that: 1) Extremely low frequency mainly captures color, brightness, and some smooth textures; 2) Low-to-mid frequency begins to reveal the overall structure; 3) Mid-to-high frequency provides more edge and texture details; 4) Extremely high frequency contains more imperceptible information, often accompanied by domain-specific information, as mentioned in other studies [23, 32].

For modality generalization, as shown in the bottom of Figure 2(a), we train the gait model [7] with RGB from scratch, and then test it after progressively removing low or high frequencies. The within-domain (CCPG, average rank-1 under gait evaluation protocol) and cross-domain (CCPG→ CASIA-B, average rank-1 results show that:

- Within-domain tends to overfit extremely low and high frequencies since they involve more domain-specific information. However, the extremely low frequencies contain some unreliable factors like color and brightness, while the higher frequencies contain more imperceptible domain-specific information. Although these may improve within-domain performance, they largely hinder the broader goal of cross-domain generalization.
- Cross-domain is not affected by higher frequencies and

hindered by low frequencies. It favors more mid-to-low frequencies that carry more domain-independent information, such as structure and semantics.

Conclusion for Gait Modality. Based on the above analysis, we realize that a direct attempt to construct a gait-specific modality for generalized gait recognition by modulating low and high frequencies could be: **Remove the noisy extremely low frequencies and appropriately suppress high frequencies.**

The reasons for selecting removal for extremely low frequencies while suppression for high frequencies are threefold: **1)** Extremely low frequencies negatively affect both within-domain and cross-domain performance, while high frequencies are crucial for within-domain performance but do not benefit cross-domain performance. The degree of suppression applied to high frequencies allows for a trade-off between within-domain and cross-domain performance. **2)** The characteristic of frequency decomposition is that extremely low frequencies are always concentrated, while high frequencies are spread more widely. It is easy to determine the extremely low frequencies, while defining the boundaries of high frequencies is challenging. **3)** The intuitive findings in Figure 2(a) also show that extremely low frequency contains rich within-domain information (*e.g.*, color, brightness, smooth textures), making removal a better choice than suppression.

3.2. How Frequency Impact Model Training?

Previous research reveals three key findings of the relations between frequencies and CNNs during training [70–72]: **1)** CNNs initially fit low frequency, gradually shift to high frequency, and finally emphasize more high-frequency components; **2)** Deep layers in CNNs focus more on high-frequency information compared to shallow layers. **3)** Excessive focus on high frequencies tends to overfit domain-specific features, leading to reduced generalization (*e.g.*, in classification task, a cat with the high-frequency texture of an elephant may be misclassified as an elephant [10]).

Considering that gait models adopt similar CNN architectures, we find that they exhibit similar tendencies in gait models. As in the left part of Figure 2(b), we feed the original RGB image into gait model [7], and plot the relative log amplitudes of frequency transformed feature map in each layer of gait model. The x-axis indicates discrete frequency components, ranging from low to high frequencies (left to right). We find that the value of high-frequency component first decrease and then increase with the increase in model layer depth. This suggests that gait model shares the same conclusion with current CNNs, *i.e.*, the model first fits low frequencies and then high frequencies, ultimately leading to high-frequency dominance, which impacts generalization.

Conclusion for Model Training Strategy. As discussed above, a direct attempt is to **guide the model towards a broader frequency range, rather than letting high frequencies dominate.** More importantly, the guidance should be applied at the shallow layers, as the training objective of the model will always push deeper layers to better fit within-domain information.

For details, as shown in the right part of Figure 2(b), we attempt to decompose the input into multiple frequency bands and obtain the corresponding modality representations. These are then concatenated along the channel dimension and fed into the gait model. The empirical results show that the learned features shift towards reducing high frequencies, leading to improved performance.

4. Methodology

In this section, we leverage the above conclusions to develop the final solution. As shown in Figure 3, we employ standard Discrete Cosine Transform (DCT) as the tool and specifically refine it for gait recognition (Sec. 4.1), focusing on two aspects: building a better gait modality (termed X modality) (Sec. 4.2) and designing improved model training strategies with frequency decomposition (Sec. 4.3) for generalizable gait recognition.

4.1. Modifications of DCT for Gait Recognition

To enhance the utility of frequency decomposition tool for gait recognition, as shown in Figure 3, we follow JPEG compression [5, 57] and introduce two modifications:

Color Transformation. We first convert the image $I(x, y)$ from RGB to YCbCr color space (denoted as $\varphi(\cdot)$ in Eq. (1)), effectively separating the input into luminance (Y, brightness) and chrominance (Cb/Cr, color) spaces. This modification enables subsequent *targeted suppression* in Sec. 4.2 to mitigate brightness and color issues.

Patch-based DCT (PDCT). After the color transformation, we adjust the value range of YCbCr from standard [0, 255] to a symmetric range of [-128, 127] to facilitate DCT process [5, 57]. The image is then divided into a series of non-overlapping patches P_i with sizes of $K \times K$, and DCT is applied to obtain spectral patches F_{P_i} as follows:

$$\text{patches}: \{P_i\}_{i=1}^{H/K \times W/K} = \text{Patchify}(\varphi(I) - 128), \quad (1)$$

$$\text{spectral patches}: \{F_{P_i}\}_{i=1}^{H/K \times W/K} = \text{DCT}(P_i). \quad (2)$$

Discussion about Patch-based DCT modification. We select patch-based strategy since it contains two benefits: *1) Lower Computational Complexity.* DCT's complexity grows quadratically with image size (*i.e.*, $O(H \times W)$), making it inefficient for large images. Patches with small sizes $(K \times K)$ are suitable for efficiency. *2) Better Local Detail Preservation.* Applying DCT to the whole image may obscure important local details with global frequency information. Multiple patches retain crucial local information that is proven essential for accurate gait recognition [6, 35, 60].

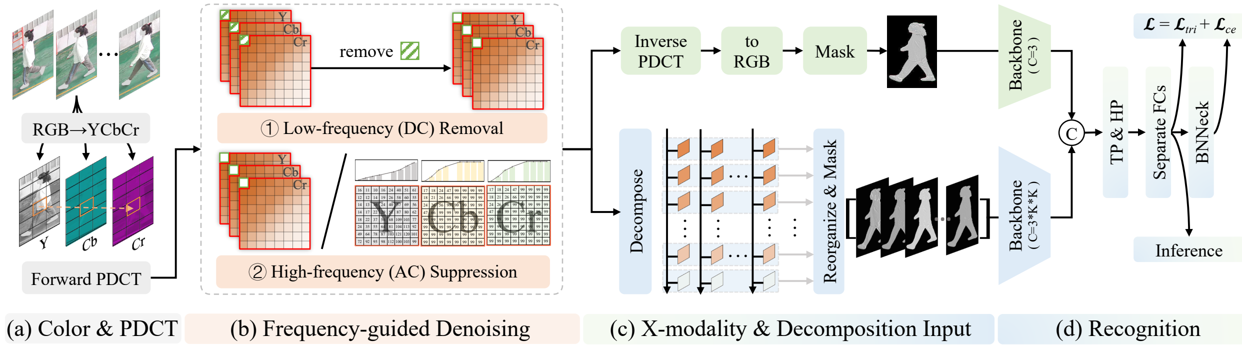


Figure 3. Gait-X introduces: **(a)** color transformation to separate brightness (Y) and color (Cb/Cr), and Patch-based DCT (patch size: K×K=8×8) for frequency decomposition; **(b)** For each patch, the upper-left pixel indicates the extremely low frequencies (known as direct component, DC), and the other frequencies are regarded as high frequencies (known as alternating component, AC). The frequency-guided denoising involves low-frequency removal to eliminate average brightness/color impact and high-frequency suppression (divided by the suppression matrix [43, 44], best viewed in Figure 5) to implement more suppression on higher frequencies and colors (Cb/Cr); **(c)** all processed patches undergo two distinct branches to generate X modality and build decomposition inputs for **(d)** final recognition as [7].

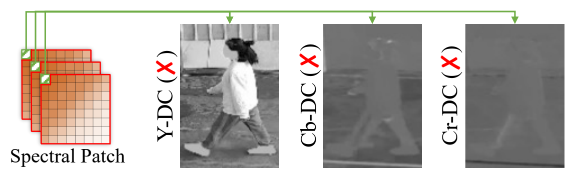


Figure 4. Visualization of *removed* DC components of Y/Cb/Cr.

Y:

16	11	10	16	24	40	51	61
12	12	14	19	26	58	60	55
14	13	16	24	40	57	69	56
14	17	22	29	51	87	80	62
18	22	37	56	68	109	103	77
24	35	55	64	81	104	113	92
49	64	78	87	103	121	120	101
72	92	95	98	112	100	103	99

Cb:

17	18	24	47	99	99	99	99
18	21	26	66	99	99	99	99
24	26	56	99	99	99	99	99
47	66	99	99	99	99	99	99
99	99	99	99	99	99	99	99
99	99	99	99	99	99	99	99
99	99	99	99	99	99	99	99
99	99	99	99	99	99	99	99

Cr:

17	18	24	47	99	99	99	99
18	21	26	66	99	99	99	99
24	26	56	99	99	99	99	99
47	66	99	99	99	99	99	99
99	99	99	99	99	99	99	99
99	99	99	99	99	99	99	99
99	99	99	99	99	99	99	99
99	99	99	99	99	99	99	99

Figure 5. Visualization of standard suppression matrix of Y/Cb/Cr in JPEG compression [43, 44]. It uses a *progressive suppression strategy* to suppress high frequencies, with stronger suppression applied to higher frequencies. Moreover, stronger suppression is applied to the Cb/Cr color components to reduce color impact.

4.2. Frequency-guided X Gait Modality

Based on the conclusion in Sec. 3.1 and modifications in Sec. 4.1, we introduce to remove and suppress specific frequencies that affect recognition and generalization at the input level, aiming to construct a gait-specific X modality toward generalized gait recognition.

Low Frequency (DC) Removal. PDCT facilitates energy compaction, gathering visually relevant information in the upper left corner of each patch, known as Direct Component (DC, Figure 4) [5]. DC represents the average value of each channel. For YCbCr image input, the DC in Y channel indicates average brightness while Cb/Cr channels indicate average color. Unlike JPEG, which retains the DC for visual quality, we follow our conclusion from Sec. 3.1 to remove all DC components from Y/Cb/Cr channels, efficiently reducing environmental lighting, color, and noise effects.

High Frequency (AC) Suppression. To guide gait representation to focus on more mid-to-low frequencies for better generalization at the input level, as shown in Figure 3(b), we introduce a *progressive frequency suppression* strategy that gradually suppresses other frequencies (Alternating Component, AC) from low to high, with stronger suppression applied to higher frequencies. Additionally, due to the potentially negative impact of chrominance (Cb/Cr) components on recognition, we apply more intense suppression on chrominance channels to learn more chroma-independent discriminative information. The suppression matrixes (Q_{Y}, Q_{Cb}, Q_{Cr}) are shown in Figure 5.

$$\{\hat{F_{P_i}}\}^{\hat{\mathrm{Y}}} = \frac{\{F_{P_i}\}^{\mathrm{Y}}}{Q_{\mathrm{Y}} \cdot \gamma},\ \{\hat{F_{P_i}}\}^{\hat{\mathrm{Cb}}} = \frac{\{F_{P_i}\}^{\mathrm{Cb}}}{Q_{\mathrm{Cb}} \cdot \gamma},\ \{\hat{F_{P_i}}\}^{\hat{\mathrm{Cr}}} = \frac{\{F_{P_i}\}^{\mathrm{Cr}}}{Q_{\mathrm{Cr}} \cdot \gamma}. \tag{3}$$

where γ is used to control the suppression level.

Inverse PDCT & Color & Mask Operation. Finally, we apply Inverse PDCT (IPDCT) and color transformation ($\varphi'(\cdot)$) to convert all patches back to RGB image. Then we adopt silhouette as mask to filter unnecessary background noise to form the final X gait modality. As in Figure 3(c)(d), it will be directly sent to the backbone for feature extraction.

$$X = \mathrm{Mask}(\varphi'(\mathrm{IPDCT}(\{\hat{F_{P_i}}\}_{i=1}^{H/K \times W/K}))) \tag{4}$$

4.3. Frequency Decomposition Model Training

In this subsection, we follow the conclusion and solution in Sec. 3.2, and propose a Frequency Decomposition Training Strategy to expand the input to multiple frequency representations. This operation can guide the gait model to focus on a broader range of frequencies at early stages, thereby

Table 1. Within-domain results on CCPG [33] dataset. The best and second-best are **bolded** and <u>underlined</u> respectively.

Modality	Method	Venue	ReID Evaluation Protocol					Gait Evaluation Protocol				
			CL	UP	DN	BG	Mean	CL	UP	DN	BG	Mean
Skeleton	GaitGraph2 [56]	CVPRW'22	5.0	5.7	7.3	8.8	6.7	5.0	5.3	5.8	6.2	5.6
	GaitTR [77]	ES'23	24.3	28.7	31.1	28.1	28.1	15.7	18.3	18.5	17.5	17.5
	GPGait [9]	ICCV'23	54.8	65.6	71.1	-	-	-	-	-	-	-
	SkeletonGait [8]	AAAI'24	52.4	65.4	72.8	80.9	67.9	40.4	48.5	53.0	61.7	50.9
Silhouette	GaitSet [2]	AAAI'19	75.9	84.3	83.1	87.3	82.6	59.7	65.9	65.3	71.3	65.6
	GaitPart [6]	CVPR'20	74.9	83.4	82.3	85.9	81.6	60.5	66.2	65.0	70.5	65.6
	AUG-OGBase [33]	CVPR'23	84.7	88.4	89.4	-	-	-	-	-	-	-
	GaitBase [7]	CVPR'23	89.1	93.4	92.5	94.3	92.3	74.0	77.7	79.3	83.5	78.6
RGB	AP3D [12]	ECCV'20	62.6	67.6	82.0	97.3	77.4	53.4	57.3	69.7	91.4	67.8
	PSTA [63]	ICCV'21	51.9	62.0	72.3	94.1	70.1	42.2	52.2	60.3	84.5	59.8
	PiT [76]	TII'22	49.1	56.2	78.0	96.9	70.1	41.0	47.6	64.3	91.0	61.0
	BigGait [73]	CVPR'24	<u>89.6</u>	<u>93.2</u>	<u>95.2</u>	<u>97.2</u>	<u>93.8</u>	**82.6**	<u>85.9</u>	<u>87.1</u>	<u>93.1</u>	<u>87.2</u>
	Gait-X (Ours)		**91.5**	**94.3**	**96.8**	**98.1**	**95.2**	<u>82.0</u>	**86.0**	**90.7**	**95.7**	**88.6**

avoiding premature overfitting to high frequencies for better generalization. All involve two steps:

Decomposition. After applying the similar processes as in Sec. 4.2 (*i.e.*, PDCT, Low Frequency Removal, and High Frequency Suppression), we obtain P spectral patches ($P{=}\frac{H}{K}\times\frac{W}{K}$) with their frequencies arranged from low to high. Each patch includes $K\times K$ elements and each element represents a specific frequency of that local area.

Reorganization and Training. As shown in Figure 3(c), we reorganize the elements from the same position across different patches, combining them into an image-like version through numerical min-max normalization and adjusting the numerical range to [0, 255].

$$\overbrace{\underbrace{H\times W}_{\text{input}}}^{\text{image size}} \Rightarrow \underbrace{\overbrace{\frac{H}{K}\times\frac{W}{K}}^{\text{numbers}}\times\overbrace{K\times K}^{\text{patches}}}_{\text{decomposition}} \Rightarrow \underbrace{\overbrace{K\times K}^{\text{numbers}}\times\overbrace{\frac{H}{K}\times\frac{W}{K}}^{\text{image size}}}_{\text{reorganization}} \tag{5}$$

As a result, we obtain $K\times K$ spectral images, and each is relevant to original image. As in Figure 3(c)(d), we also adopt silhouette as mask to remove background noise, concatenate them along channel dimensions, and feed them to backbone to expand the model's focus to a broader range of frequencies for better generalization.

5. Experiments

5.1. Datasets and Implementation Details

Datasets. We utilize CCPG [33], SUSTech1K [51], and CASIA-B [75] to evaluate our Gait-X. Additionally, we strictly follow the official evaluation protocols among these gait datasets [33, 51, 75] for a fair comparison. We do not select other datasets like Gait3D [81], GREW [86], or OUMVLP [54] as they lack the required RGB modality. Further experiments can be conducted when RGB becomes available. Although frequency decomposition can also be applied to other modalities like silhouette, it does not fulfill one of our goals of supplementing more discriminative information than silhouette for recognition. Notably, all selected datasets provide challenging tasks like clothing changes, multiple views, and varied walking types. We believe these datasets are sufficient for evaluation since recently published papers like BigGait [73] also utilize our selected datasets for evaluation.

Implementation Details. During training, we randomly sample 8 individuals and 8 sequences, each randomly sampled fixed 30 frames. The training utilizes SGD optimizer, with learning rate 0.1 and momentum 0.9. For CCPG and CASIA-B, we train for 60k iterations and reduce the learning rate by 0.1 at 20k, 40k, and 50k milestones. For SUSTech1K, the training lasts 40k iterations with rate reductions at 20k and 30k iterations. We adopt the latest ResNet-like Backbone (GaitBase [7]) and follow its loss function settings (*i.e.*, triplet and cross-entropy loss). The X modality and decomposition training will undergo two distinct branches and simply concatenate along the height dimension after backbone like GaitSet [2]. More streamlined structures are worth proposing in the future. None of data augmentations are used for simplicity.

Other Details. Following JPEG [57, 84], the patch size is set as $K = 8$. The γ in Eq. (3) is set as $1/16$. Following the patch-size output of decomposition and referencing [29], we perform K-fold up-sampling on the original image, ensuring minimal modifications to the Backbone.

5.2. Comparisons of Within-domain Evaluation

Comparisons on CCPG. While CCPG provides the most challenging cloth-changing settings, as shown in Table 1, Gait-X significantly improves the average rank-1 by +37.7%, +10.0%, +1.4% compared to skeleton, silhouette, and RGB under the gait evaluation protocol. These results confirm that our Gait-X owns more discriminative features, enabling consistently improving single-domain performance across various cloth-changing settings.

Table 2. Within-domain results on SUSTech1K [51] dataset. The best and second-best are **bolded** and underlined respectively.

Modality	Method	Venue	Normal	Bag	Clothing	Carrying	Umbrella	Uniform	Occlusion	Night	Overall
Skeleton	GaitGraph2 [56]	CVPRW'22	22.2	18.2	6.8	18.6	13.4	19.2	27.3	16.4	18.6
	GaitTR [77]	ES'23	33.3	31.5	21.0	30.4	22.7	34.6	44.9	23.5	30.8
	SkeletonGait [8]	AAAI'24	67.9	63.5	36.5	61.6	58.1	67.2	79.1	50.1	63.0
Silhouette	GaitSet [2]	AAAI'19	76.3	76.8	46.8	73.9	73.1	71.0	73.6	28.3	73.8
	GaitPart [6]	CVPR'20	76.5	75.6	48.9	73.0	72.0	72.2	77.0	26.5	73.3
	GiatGL [35]	ICCV'21	79.2	79.1	60.0	76.0	76.8	76.9	76.4	21.5	76.7
	GaitBase [7]	CVPR'23	84.7	83.8	59.8	81.9	82.1	82.4	88.0	30.3	82.1
Lidar	PointNet++ [46]	NeurIPS'17	55.9	52.2	41.6	49.6	47.8	45.9	54.2	52.5	50.8
	PointTransformer [80]	ICCV'21	53.2	48.1	32.1	43.2	39.1	41.8	47.9	47.1	44.4
	SimpleView [11]	ICML'21	72.3	68.8	57.2	63.3	49.2	62.5	79.7	66.5	64.8
	LidarGait [51]	CVPR'23	91.8	88.6	74.6	89.0	67.5	80.9	94.5	90.4	86.8
	HMRNet [14]	MM'24	92.7	92.3	79.6	90.3	83.1	86.2	95.2	90.4	90.2
RGB	AP3D [12]	ECCV'20	94.4	-	82.7	-	95.3	91.6	-	**89.4**	96.8
	PSTA [63]	ICCV'21	92.9	-	72.3	-	92.1	83.2	-	79.9	93.6
	BigGait [73]	CVPR'24	96.1	-	73.3	-	96.0	93.2	-	85.3	96.2
	Gait-X (Ours)		**98.7**	**97.4**	**85.2**	**97.1**	**97.5**	**96.9**	**99.4**	71.9	**96.9**

Table 3. Within-domain results on CASIA-B [75] dataset. The best and second-best are **bolded** and underlined respectively. $+$ indicates using 128×88 resolution.

Modality	Method	Venue	NM	BG	CL	Mean
Skeleton	GaitGraph2 [56]	CVPRW'22	80.3	71.4	63.8	71.8
	GaitTR [77]	ES'23	94.7	89.3	86.7	90.2
	GPGait [9]	ICCV'23	93.6	80.2	69.3	81.0
	CAG [27]	TIP'23	96.4	91.8	90.6	92.9
Silhouette	GaitSet [2]	AAAI'19	95.0	87.2	70.4	84.2
	GaitSet$^+$ [2]	AAAI'19	95.6	91.5	75.3	87.5
	GaitPart [6]	CVPR'20	96.2	91.5	78.7	88.8
	GLN$^+$ [17]	ECCV'20	96.9	94.0	77.5	89.5
	CSTL [26]	ICCV'21	97.8	93.6	84.2	91.9
	CSTL$^+$ [26]	ICCV'21	98.0	95.4	87.0	93.5
	3D Local [28]	ICCV'21	97.5	94.3	83.7	91.8
	3D Local$^+$ [28]	ICCV'21	98.3	95.5	84.5	92.8
	GaitGL [35]	ICCV'21	97.4	94.5	83.6	91.8
	MetaGait [3]	CVPR'22	98.1	95.2	86.9	93.4
	DANet [37]	CVPR'23	98.0	95.9	89.9	94.6
	GaitBase [7]	CVPR'23	97.6	94.0	77.4	89.7
	GaitBase$^+$ [7]	CVPR'23	99.0	96.1	83.8	93.0
	GaitGCI [4]	CVPR'23	98.4	96.6	88.5	94.5
	DyGait [61]	ICCV'23	98.4	96.2	87.8	94.1
	VPNet [38]	CVPR'24	98.3	96.3	90.0	94.9
	CLTD [67]	ECCV'24	98.6	96.4	89.3	94.8
RGB	GaitEdge [34]	ECCV'22	97.9	96.1	86.4	93.5
	AP3D [12]	ECCV'20	99.8	99.4	87.6	95.6
	PSTA [63]	ICCV'21	98.2	96.5	54.2	83.0
	BigGait [73]	CVPR'24	**100.0**	99.6	90.5	96.7
	Gait-X (Ours)		99.8	**99.7**	**93.7**	**97.7**

Comparisons on SUSTech1K. SUSTech1K includes more diverse conditions. As shown in Table 2, Gait-X achieves outstanding performance in the most settings, with overall improvements of +33.9%, +14.8%, +6.7%, +0.1% over skeleton, silhouette, lidar, and RGB modalities. Moreover, Gait-X reaches an impressive 85.2% in the most challenging cloth-changing condition. The night condition is not good as it cannot compare to lidar due to its inherent advantages, and the silhouette in mask operation often fails to accurately segment in the night condition.

Comparisons on CASIA-B. CASIA-B has been widely studied since 2006, with the best performance reaching 96.7% (average rank-1) in BigGait [73]. However, the most challenging CL setting still only achieves 90.5%. In comparison, our Gait-X achieves 97.7% average rank-1 and a 3.2% improvement under the CL condition.

5.3. Comparisons of Cross-domain Generalization

Table 4 shows that Gait-X achieves strong generalization across three gait datasets. For example, Gait-X achieves +1.6% (average rank-1) in CCPG→CASIA-B and +21.5% (overall) in CCPG→SUSTech1K evaluations. 1) Compared to silhouette, this confirms that our X modality provides more sufficient and stable information for recognition, such as semantics. 2) Compared to RGB, our X modality can mitigate potential noise and retain sufficient information for recognition. Although certain cases, like night in SUSTech1K, show lower performance due to poor-quality segmentation in mask operation, this does not affect the border performance improvements achieved by Gait-X.

5.4. Ablation Study

If not specified, we conduct ablation experiments using 8 frames to save time and memory. The comparisons are around within-domain evaluation on CCPG under gait evaluation protocol and cross-domain evaluation from CCPG to CASIA-B. The ablation of frames is shown in Table 5.

Impact of X modality & Decompose Training. As shown in the Table 6 with gray color, single X modality or decomposition training improves average rank-1 by +7.8%, +8.2% in within-domain evaluation, and +17.5%, +25.1% in cross-domain evaluation compared to Baseline. Gait-X combines them and achieves more improvements.

Table 4. Cross-domain results across CCPG [33], SUSTech1K [51], and CASIA-B [75] datasets. The best and second-best are **bolded** and underlined respectively. The comparisons mainly include the recent silhouette-based method [7] and RGB-based methods [12, 63, 73].

Method	Source Dataset: CCPG					
	to CASIA-B			to SUSTech1K		
	NM	BG	CL	Clothing	Night	Overall
GaitBase	65.3	59.2	27.2	10.0	14.4	18.7
AP3D	53.7	46.2	11.9	36.2	**51.6**	55.3
PSTA	49.7	42.3	8.8	25.7	29.4	40.6
BigGait	77.4	71.5	**33.6**	43.7	44.8	56.4
Gait-X	**81.8**	**76.9**	28.5	**47.9**	43.4	**73.4**

Method	Source Dataset: SUSTech1K						
	to CASIA-B			to CCPG			
	NM	BG	CL	CL	UP	DN	BG
GaitBase	77.9	70.7	28.0	**14.7**	**19.7**	**25.1**	36.6
AP3D	56.7	48.1	15.3	5.5	7.9	13.9	35.1
PSTA	31.2	27.7	10.6	3.7	5.7	9.5	26.5
BigGait	91.1	85.8	18.7	4.5	11.5	11.9	45.5
Gait-X	**92.0**	**86.8**	**28.3**	7.2	15.1	22.2	**56.0**

Method	Source Dataset: CASIA-B						
	to SUSTech1K			to CCPG			
	Clothing	Night	Overall	CL	UP	DN	BG
GaitBase	9.7	14.3	16.5	**12.8**	17.0	21.1	29.2
AP3D	29.3	47.4	48.3	2.1	2.9	3.9	6.1
PSTA	19.9	37.5	34.6	1.7	1.9	3.4	5.0
BigGait	36.9	**60.2**	64.8	7.5	**19.5**	14.2	43.0
Gait-X	**43.9**	48.6	**71.4**	10.8	17.2	**23.2**	**47.1**

Table 5. Ablation of the number of frames.

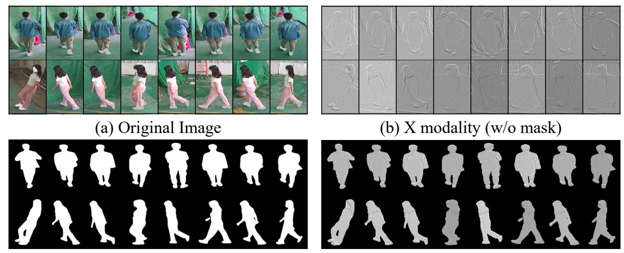
Setting	Within-domain (CCPG, Gait)					Cross-domain (to CASIA-B)			
	CL	UP	DN	BG	Mean	NM	BG	CL	Mean
Baseline$_{\times 30 \text{ frames}}$	67.8	72.2	78.0	82.9	75.2	56.5	48.0	15.0	39.8
Gait-X$_{\times 30 \text{ frames}}$	82.0	86.0	90.7	95.7	88.6	81.8	76.9	28.5	62.4
Baseline$_{\times 8 \text{ frames}}$	67.8	72.8	80.0	86.3	76.7	54.2	47.4	13.6	38.4
Gait-X$_{\times 8 \text{ frames}}$	79.0	84.4	89.0	96.8	87.3	85.9	80.0	28.8	64.9

Table 6. Ablation of different modules in Gait-X. "Mask", "DC" and "AC" indicate background removal, low-frequency (DC) removal, and high-frequency (AC) suppression.

Setting	Mask	DC	AC	Within-domain (CCPG, Gait)					Cross-domain (to CASIA-B)			
				CL	UP	DN	BG	Mean	NM	BG	CL	Mean
Baseline	×	×	×	67.8	72.8	80.0	86.3	76.7	54.2	47.4	13.6	38.4
Single X Modality	✓	×	×	67.2	74.0	84.5	93.1	79.7	77.7	69.7	18.9	55.4
	✓	×	✓	68.3	76.0	84.7	93.6	80.7	78.2	72.3	20.6	57.0
	✓	✓	×	70.1	78.1	82.3	91.7	80.6	71.3	64.7	25.4	53.8
	✓	✓	✓	74.7	81.6	86.3	95.2	84.5	73.7	68.6	25.5	55.9
Single Decompose Training	✓	×	×	68.4	75.4	84.7	94.1	80.7	76.2	69.9	21.4	55.8
	✓	×	✓	69.7	75.6	85.6	94.6	81.4	76.9	70.5	22.3	56.6
	✓	✓	×	71.9	79.2	85.9	94.8	83.0	84.9	77.4	21.8	61.4
	✓	✓	✓	75.3	80.2	87.6	96.3	84.9	**87.1**	79.2	24.3	63.5
Mask	×	×	×	69.4	74.9	82.3	88.8	78.9	63.6	57.3	16.2	45.7
	✓	×	×	72.8	79.4	87.7	96.3	84.1	80.5	74.3	21.9	58.9
	×	✓	✓	78.9	82.9	88.4	94.0	86.1	75.3	67.6	19.0	54.0
Gait-X	✓	✓	✓	**79.0**	**84.4**	**89.0**	**96.8**	**87.3**	85.9	**80.0**	**28.8**	**64.9**

Impact of DC and AC operations. As shown in the second and third parts of Table 6, the low-frequency (DC) removal and high-frequency (AC) suppression significantly improve within-domain performance and generalization in both single X modality or single decomposition training. Although there are occasional slight drops under NM and BG settings, both operations consistently improve performance in the more critical cloth-changing (CL) setting. Take the single X modality for example, +7.5% in within-domain results of CCPG and +6.6% in CCPG→CASIA-B.

Impact of mask operation. As shown in the last three rows in Table 6, the mask operation allows to remove the background noise. It is worth noting that the mask operation only provides a slight improvement in within-domain evaluation. The cross-domain evaluation benefits more from mask due to the different scene noises, and our Gait-X can further bring 6% performance gain.



(a) Original Image (b) X modality (w/o mask) (c) Mask (Silhouette) (d) X modality (w/ mask)

Figure 6. Examples of X modality. Best viewed (4×) in color.

Visualization of the final X modality. As shown in Figure 6, with the help of our proposed low-frequency (DC) removal, high-frequency (AC) suppression, and the mask to remove background noise, the final X modality can effectively remove or suppress brightness, color, and clothing texture, serving as a novel modality for further gait research.

Limitation and Discussion. 1) Our frequency analysis in gait recognition is an initial attempt that focuses on modality exploration and model training. More about model structure design is worth further exploration. 2) The removal of extremely low frequencies and suppression of high frequencies may inevitably lose little valuable information. Exploring a more strict solution is meaningful. 3) This paper focuses on mitigating the noise in human body instead of background. The introduced mask operation is a simple way to remove background noise. More mask-free methods are worth further exploration.

6. Conclusion

In conclusion, we propose to explore a gait-specific X modality and model train strategy for generalized gait recognition, named Gait-X. We systematically analyze the impact of frequency on gait modality and explore how to filter domain-specific information while emphasizing domain-independent information to improve both gait modality and model generalization. Gait-X offers a fresh frequency perspective and achieves superior performance. We believe this work will open new avenues through frequency analysis and inspire future gait research.

7. Acknowledgements

This work is jointly supported by the National Natural Science Foundation of China (62276031, 62276025, 62206022, 62476027), the National Key Research and Development Program of China (2023YFF0904700), the Doctoral Innovation Fund of Beijing University of Posts and Telecommunications (CX20242083), and the Fundamental Research Funds for the Central Universities (2253200026).

References

[1] Tinku Acharya and Ajoy K Ray. *Image Processing: Principles and Applications*. John Wiley & Sons, 2005. 3

[2] Hanqing Chao, Yiwei He, Junping Zhang, and Jianfeng Feng. GaitSet: Regarding gait as a set for cross-view gait recognition. In *AAAI*, 2019. 1, 2, 6, 7

[3] Huanzhang Dou, Pengyi Zhang, Wei Su, Yunlong Yu, and Xi Li. MetaGait: Learning to learn an omni sample adaptive representation for gait recognition. In *ECCV*, 2022. 7

[4] Huanzhang Dou, Pengyi Zhang, Wei Su, Yunlong Yu, Yining Lin, and Xi Li. GaitGCI: Generative counterfactual intervention for gait recognition. In *CVPR*, 2023. 7

[5] Max Ehrlich, Larry Davis, Ser-Nam Lim, and Abhinav Shrivastava. Quantization guided jpeg artifact correction. In *ECCV*, 2020. 4, 5

[6] Chao Fan, Yunjie Peng, Chunshui Cao, Xu Liu, Saihui Hou, Jiannan Chi, Yongzhen Huang, Qing Li, and Zhiqiang He. GaitPart: Temporal part-based model for gait recognition. In *CVPR*, 2020. 1, 2, 4, 6, 7

[7] Chao Fan, Junhao Liang, Chuanfu Shen, Saihui Hou, Yongzhen Huang, and Shiqi Yu. OpenGait: Revisiting gait recognition towards better practicality. In *CVPR*, 2023. 3, 4, 5, 6, 7, 8

[8] Chao Fan, Jingzhe Ma, Dongyang Jin, Chuanfu Shen, and Shiqi Yu. SkeletonGait: Gait recognition using skeleton maps. *arXiv preprint arXiv:2311.13444*, 2023. 6, 7

[9] Yang Fu, Shibei Meng, Saihui Hou, Xuecai Hu, and Yongzhen Huang. GPGait: Generalized pose-based gait recognition. In *ICCV*, 2023. 2, 6, 7

[10] Robert Geirhos, Patricia Rubisch, Claudio Michaelis, Matthias Bethge, Felix A. Wichmann, and Wieland Brendel. Imagenet-trained cnns are biased towards texture; increasing shape bias improves accuracy and robustness. In *ICLR*, 2018. 4

[11] Ankit Goyal, Hei Law, Bowei Liu, Alejandro Newell, and Jia Deng. Revisiting point cloud shape classification with a simple and effective baseline. In *ICML*, 2021. 7

[12] Xinqian Gu, Hong Chang, Bingpeng Ma, Hongkai Zhang, and Xilin Chen. Appearance-preserving 3d convolution for video-based person re-identification. In *ECCV*, 2020. 6, 7, 8

[13] Xinqian Gu, Hong Chang, Bingpeng Ma, Shutao Bai, Shiguang Shan, and Xilin Chen. Clothes-changing person re-identification with rgb modality only. In *CVPR*, 2022. 1

[14] Xiao Han, Yiming Ren, Peishan Cong, Yujing Sun, Jingya Wang, Lan Xu, and Yuexin Ma. Gait recognition in large-scale free environment via single lidar. In *ACM MM*, 2024. 1, 2, 7

[15] Ahmed R Hawas, Heba A El-Khobby, Mohammed Abd-Elnaby, and Fathi E Abd El-Samie. Gait identification by convolutional neural networks and optical flow. *Multimedia Tools and Applications*, 2019. 1, 2

[16] Keji He, Chenyang Si, Zhihe Lu, Yan Huang, Liang Wang, and Xinchao Wang. Frequency-enhanced data augmentation for vision-and-language navigation. In *NeurIPS*, 2024. 2, 3

[17] Saihui Hou, Chunshui Cao, Xu Liu, and Yongzhen Huang. Gait lateral network: Learning discriminative and compact representations for gait recognition. In *ECCV*, 2020. 7

[18] Saihui Hou, Xu Liu, Chunshui Cao, and Yongzhen Huang. Set residual network for silhouette-based gait recognition. *IEEE TBIOM*, 2021. 2

[19] Saihui Hou, Xu Liu, Chunshui Cao, and Yongzhen Huang. Gait quality aware network: Toward the interpretability of silhouette-based gait recognition. *IEEE TNNLS*, 2022.

[20] Saihui Hou, Panjian Huang, Xu Liu, Chunshui Cao, and Yongzhen Huang. Cloth-imbalanced gait recognition via hallucination. *IEEE TCSVT*, 34(7):5665–5676, 2024.

[21] Saihui Hou, Zengbin Wang, Man Zhang, Chunshui Cao, Xu Liu, and Yongzhen Huang. Edge-oriented adversarial attack for deep gait recognition. *IJCV*, 133(4):1549–1563, 2025. 2

[22] Wei Hu, Zeqing Fu, and Zongming Guo. Local frequency interpretation and non-local self-similarity on graph for point cloud inpainting. *IEEE TIP*, 2019. 3

[23] Jiaxing Huang, Dayan Guan, Aoran Xiao, and Shijian Lu. FSDR: Frequency space domain randomization for domain generalization. In *CVPR*, 2021. 3

[24] Panjian Huang, Saihui Hou, Chunshui Cao, Xu Liu, Xuecai Hu, and Yongzhen Huang. Integral pose learning via appearance transfer for gait recognition. *IEEE TIFS*, 2024. 1

[25] Panjian Huang, Yunjie Peng, Saihui Hou, Chunshui Cao, Xu Liu, Zhiqiang He, and Yongzhen Huang. Occluded gait recognition with mixture of experts: An action detection perspective. In *ECCV*, 2025. 2

[26] Xiaohu Huang, Duowang Zhu, Hao Wang, Xinggang Wang, Bo Yang, Botao He, Wenyu Liu, and Bin Feng. Context-sensitive temporal feature learning for gait recognition. In *ICCV*, 2021. 7

[27] Xiaohu Huang, Xinggang Wang, Zhidianqiu Jin, Bo Yang, Botao He, Bin Feng, and Wenyu Liu. Condition-adaptive graph convolution learning for skeleton-based gait recognition. *IEEE TIP*, 2023. 7

[28] Zhen Huang, Dixiu Xue, Xu Shen, Xinmei Tian, Houqiang Li, Jianqiang Huang, and Xian-Sheng Hua. 3d local convolutional neural networks for gait recognition. In *ICCV*, 2021. 7

[29] Jiazhen Ji, Huan Wang, Yuge Huang, Jiaxiang Wu, Xingkun Xu, Shouhong Ding, ShengChuan Zhang, Liujuan Cao, and Rongrong Ji. Privacy-preserving face recognition with learnable privacy budgets in frequency domain. In *ECCV*, 2022. 6

[30] Xin Jin, Tianyu He, Kecheng Zheng, Zhiheng Yin, Xu Shen, Zhen Huang, Ruoyu Feng, Jianqiang Huang, Zhibo Chen, and Xian-Sheng Hua. Cloth-changing person re-identification from a single image with gait prediction and regularization. In *CVPR*, 2022. 2, 3

[31] Jong Hwan Ko, Burhan Mudassar, Taesik Na, and Saibal Mukhopadhyay. Design of an energy-efficient accelerator for training of convolutional neural networks using frequency-domain computation. In *ADAC*, 2017. 3

[32] Qiufu Li, Linlin Shen, Sheng Guo, and Zhihui Lai. Wavelet integrated cnns for noise-robust image classification. In *CVPR*, pages 7245–7254, 2020. 3

[33] Weijia Li, Saihui Hou, Chunjie Zhang, Chunshui Cao, Xu Liu, Yongzhen Huang, and Yao Zhao. An in-depth exploration of person re-identification and gait recognition in cloth-changing conditions. In *CVPR*, 2023. 1, 6, 8

[34] Junhao Liang, Chao Fan, Saihui Hou, Chuanfu Shen, Yongzhen Huang, and Shiqi Yu. GaitEdge: Beyond plain end-to-end gait recognition for better practicality. In *ECCV*, 2022. 2, 7

[35] Beibei Lin, Shunli Zhang, and Xin Yu. Gait recognition via effective global-local feature representation and local temporal aggregation. In *ICCV*, 2021. 1, 2, 4, 7

[36] Jinhua Lin, Lin Ma, and Yu Yao. A fourier domain acceleration framework for convolutional neural networks. *Neurocomputing*, 2019. 3

[37] Kang Ma, Ying Fu, Dezhi Zheng, Chunshui Cao, Xuecai Hu, and Yongzhen Huang. Dynamic aggregated network for gait recognition. In *CVPR*, 2023. 7

[38] Kang Ma, Ying Fu, Chunshui Cao, Saihui Hou, Yongzhen Huang, and Dezhi Zheng. Learning visual prompt for gait recognition. In *CVPR*, 2024. 7

[39] Zohra Mahfouf, Hayet Farida Merouani, Imed Bouchrika, and Nouzha Harrati. Investigating the use of motion-based features from optical flow for gait recognition. *Neurocomputing*, 2018. 1, 2

[40] Yasushi Makihara, Ryusuke Sagawa, Yasuhiro Mukaigawa, Tomio Echigo, and Yasushi Yagi. Gait recognition using a view transformation model in the frequency domain. In *ECCV*, 2006. 2, 3

[41] Nyan M.N., Tay F.E.H., Seah K.H.W., and Sitoh Y.Y. Classification of gait patterns in the time–frequency domain. *Journal of biomechanics*, 2006.

[42] Bijan G Mobasseri and Moeness G Amin. A time-frequency classifier for human gait recognition. In *Optics and Photonics in Global Homeland Security V and Biometric Technology for Human Identification VI*, 2009. 2

[43] Mingyu Ouyang and Zhenzhong Chen. JPEG quantized coefficient recovery via dct domain spatial-frequential transformer. *IEEE Transactions on Image Processing*, 2024. 5

[44] William B Pennebaker and Joan L Mitchell. *JPEG: Still Image Data Compression Standard*. Springer Science & Business Media, 1992. 5

[45] Anish Prabhu, Ali Farhadi, Mohammad Rastegari, et al. Butterfly transform: An efficient fft based neural architecture design. In *CVPR*, 2020. 3

[46] Charles Ruizhongtai Qi, Li Yi, Hao Su, and Leonidas J Guibas. PointNet++: Deep hierarchical feature learning on point sets in a metric space. In *NeurIPS*, 2017. 7

[47] Veenu Rani and Munish Kumar. Human gait recognition: A systematic review. *Multimedia Tools and Applications*, 2023. 1

[48] Yongming Rao, Wenliang Zhao, Zheng Zhu, Jiwen Lu, and Jie Zhou. Global filter networks for image classification. In *NeurIPS*, 2021. 3

[49] GP Obi Reddy. Digital image processing: Principles and applications. *Geospatial Technologies in Land Resources Mapping, Monitoring and Management*, 2018. 3

[50] Chuanfu Shen, Shiqi Yu, Jilong Wang, George Q Huang, and Liang Wang. A comprehensive survey on deep gait recognition: Algorithms, datasets and challenges. *arXiv preprint arXiv:2206.13732*, 2022. 1

[51] Chuanfu Shen, Chao Fan, Wei Wu, Rui Wang, George Q Huang, and Shiqi Yu. LidarGait: Benchmarking 3d gait recognition with point clouds. In *CVPR*, 2023. 1, 2, 6, 7, 8

[52] Xiang-Jun Shen, Zhaorui Xu, Liangjun Wang, and Zechao Li. Time and memory efficient large-scale canonical correlation analysis in fourier domain. In *ACM MM*, 2022. 3

[53] Kohei Shiraga, Yasushi Makihara, Daigo Muramatsu, Tomio Echigo, and Yasushi Yagi. GEINet: View-invariant gait recognition using a convolutional neural network. In *ICB*, 2016. 1, 2

[54] Noriko Takemura, Yasushi Makihara, Daigo Muramatsu, Tomio Echigo, and Yasushi Yagi. Multi-view large population gait dataset and its performance evaluation for cross-view gait recognition. *IPSJ transactions on Computer Vision and Applications*, 2018. 6

[55] Torben Teepe, Ali Khan, Johannes Gilg, Fabian Herzog, Stefan Hörmann, and Gerhard Rigoll. GaitGraph: Graph convolutional network for skeleton-based gait recognition. In *ICIP*, 2021. 2

[56] Torben Teepe, Johannes Gilg, Fabian Herzog, Stefan Hörmann, and Gerhard Rigoll. Towards a deeper understanding of skeleton-based gait recognition. In *CVPR*, pages 1569–1577, 2022. 2, 6, 7

[57] Gregory K Wallace. The jpeg still picture compression standard. *Communications of the ACM*, 1991. 2, 4, 6

[58] Jilong Wang, Saihui Hou, Yan Huang, Chunshui Cao, Xu Liu, Yongzhen Huang, and Liang Wang. Causal intervention for sparse-view gait recognition. In *ACM MM*, pages 77–85, 2023. 2

[59] Jilong Wang, Saihui Hou, Yan Huang, Chunshui Cao, Xu Liu, Yongzhen Huang, and Liang Wang. Free lunch for gait recognition: A novel relation descriptor. In *ECCV*, 2024. 2

[60] Lei Wang, Bo Liu, Fangfang Liang, and Bincheng Wang. Hierarchical spatio-temporal representation learning for gait recognition. In *ICCV*, 2023. 4

[61] Ming Wang, Xianda Guo, Beibei Lin, Tian Yang, Zheng Zhu, Lincheng Li, Shunli Zhang, and Xin Yu. DyGait: Exploiting dynamic representations for high-performance gait recognition. In *ICCV*, 2023. 7

[62] Wei Wang, Alex X Liu, and Muhammad Shahzad. Gait recognition using wifi signals. In *ACM UbiComp*, 2016. 2

[63] Yingquan Wang, Pingping Zhang, Shang Gao, Xia Geng, Hu Lu, and Dong Wang. Pyramid spatial-temporal aggregation for video-based person re-identification. In *ICCV*, 2021. 6, 7, 8

[64] Zengbin Wang, Saihui Hou, Man Zhang, Xu Liu, Chunshui Cao, and Yongzhen Huang. GaitParsing: Human semantic parsing for gait recognition. *IEEE TMM*, 2023. 1, 2

[65] Zengbin Wang, Saihui Hou, Man Zhang, Xu Liu, Chunshui Cao, Yongzhen Huang, and Shibiao Xu. LandmarkGait: Intrinsic human parsing for gait recognition. In *ACM MM*, 2023. 1

[66] Zengbin Wang, Saihui Hou, Man Zhang, Xu Liu, Chunshui Cao, Yongzhen Huang, Peipei Li, and Shibiao Xu. QAGait: Revisit gait recognition from a quality perspective. In *AAAI*, 2024. 2

[67] Haijun Xiong, Bin Feng, Xinggang Wang, and Wenyu Liu. Causality-inspired discriminative feature learning in triple domains for gait recognition. *arXiv preprint arXiv:2407.12519*, 2024. 7

[68] Chi Xu, Yasushi Makihara, Xiang Li, and Yasushi Yagi. Occlusion-aware human mesh model-based gait recognition. *IEEE TIFS*, 2023. 1, 2

[69] Qinwei Xu, Ruipeng Zhang, Ziqing Fan, Yanfeng Wang, Yi-Yan Wu, and Ya Zhang. Fourier-based augmentation with applications to domain generalization. *Pattern Recognition*, 2023. 3

[70] Zhiqin John Xu and Hanxu Zhou. Deep frequency principle towards understanding why deeper learning is faster. In *AAAI*, 2021. 3, 4

[71] Zhi-Qin John Xu, Yaoyu Zhang, Tao Luo, Yanyang Xiao, and Zheng Ma. Frequency principle: Fourier analysis sheds light on deep neural networks. *arXiv preprint arXiv:1901.06523*, 2019. 2

[72] Zhi-Qin John Xu, Yaoyu Zhang, and Yanyang Xiao. Training behavior of deep neural network in frequency domain. In *ICONIP*, 2019. 3, 4

[73] Dingqiang Ye, Chao Fan, Jingzhe Ma, Xiaoming Liu, and Shiqi Yu. BigGait: Learning gait representation you want by large vision models. In *CVPR*, 2024. 2, 6, 7, 8

[74] Shiqi Yu, Liang Wang, Weiming Hu, and Tieniu Tan. Gait analysis for human identification in frequency domain. In *ICIG*, 2004. 2

[75] Shiqi Yu, Daoliang Tan, and Tieniu Tan. A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition. In *ICPR*, 2006. 2, 6, 7, 8

[76] Xianghao Zang, Ge Li, and Wei Gao. Multidirection and multiscale pyramid in transformer for video-based pedestrian retrieval. *IEEE TII*, 2022. 6

[77] Cun Zhang, Xing-Peng Chen, Guo-Qiang Han, and Xiang-Jie Liu. Spatial transformer network on skeleton-based gait recognition. *Expert Systems*, 2023. 2, 6, 7

[78] Ziyuan Zhang, Luan Tran, Xi Yin, Yousef Atoum, Xiaoming Liu, Jian Wan, and Nanxin Wang. Gait recognition via disentangled representation learning. In *CVPR*, 2019. 1, 2

[79] Ziyuan Zhang, Luan Tran, Feng Liu, and Xiaoming Liu. On learning disentangled representations for gait recognition. *IEEE TPAMI*, 2020. 1, 2

[80] Hengshuang Zhao, Li Jiang, Jiaya Jia, Philip HS Torr, and Vladlen Koltun. Point transformer. In *ICCV*, 2021. 7

[81] Jinkai Zheng, Xinchen Liu, Wu Liu, Lingxiao He, Chenggang Yan, and Tao Mei. Gait recognition in the wild with dense 3d representations and a benchmark. In *CVPR*, 2022. 1, 2, 6

[82] Jinkai Zheng, Xinchen Liu, Shuai Wang, Lihao Wang, Chenggang Yan, and Wu Liu. Parsing is all you need for accurate gait recognition in the wild. In *ACM MM*, 2023. 1, 2

[83] Jinkai Zheng, Xinchen Liu, Boyue Zhang, Chenggang Yan, Jiyong Zhang, Wu Liu, and Yongdong Zhang. It takes two: Accurate gait recognition in the wild via cross-granularity alignment. In *ACM MM*, 2024. 1, 2

[84] Yangming Zhou, Chengyou Wang, and Xiao Zhou. DCT-based color image compression algorithm using an efficient lossless encoder. In *ICSP*, 2018. 2, 3, 6

[85] Haidong Zhu, Zhaoheng Zheng, and Ram Nevatia. Gait recognition using 3-d human body shape inference. In *WACV*, 2023. 2

[86] Zheng Zhu, Xianda Guo, Tian Yang, Junjie Huang, Jiankang Deng, Guan Huang, Dalong Du, Jiwen Lu, and Jie Zhou. Gait recognition in the wild: A benchmark. In *ICCV*, 2021. 6

[87] Shinan Zou, Chao Fan, Jianbo Xiong, Chuanfu Shen, Shiqi Yu, and Jin Tang. Cross-covariate gait recognition: A benchmark. *arXiv preprint arXiv:2312.14404*, 2023. 2