Siamese Neural Network based Gait Recognition for Human Identification

Cheng Zhang, Wu Liu, Huadong Ma, Huiyuan Fu
Beijing University of Posts and Telecommunications

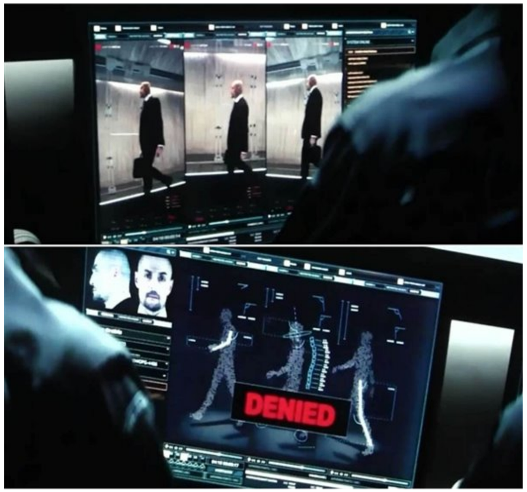
ICASSP 2016

Outline

- Introduction
- Proposed method
  - Conventional CNN based Gait Recognition
  - Siamese Network based Gait Recognition
- Experiments
- Conclusions

Definition

- ***Gait analysis*** is the systematic study of animal locomotion, more specifically the study of human motion, using the eye and the brain of observers, augmented by instrumentation for measuring body movements, body mechanics, and the activity of the muscles.

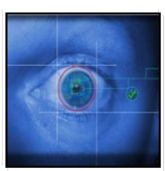


Background

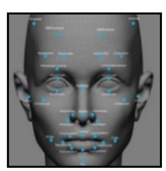
- Social security
  - Video big data and camera network
  - Remote surveillance
  - Identification and attribute classification
- Biometric authentication techniques
  - Facial recognition
  - Iris recognition
  - Fingerprint technologies
  - Voice verification
  - Hand geometry

Characteristics

- Remote accessed
  - It can identify subjects from a distance without interrupting the subject
- Robust
  - Even in low resolution videos, the gait still works well
- Security
  - It is difficult to imitate or camouflage human gait

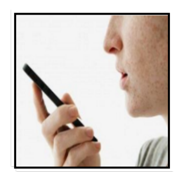


iris

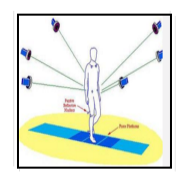


face

fingerprint



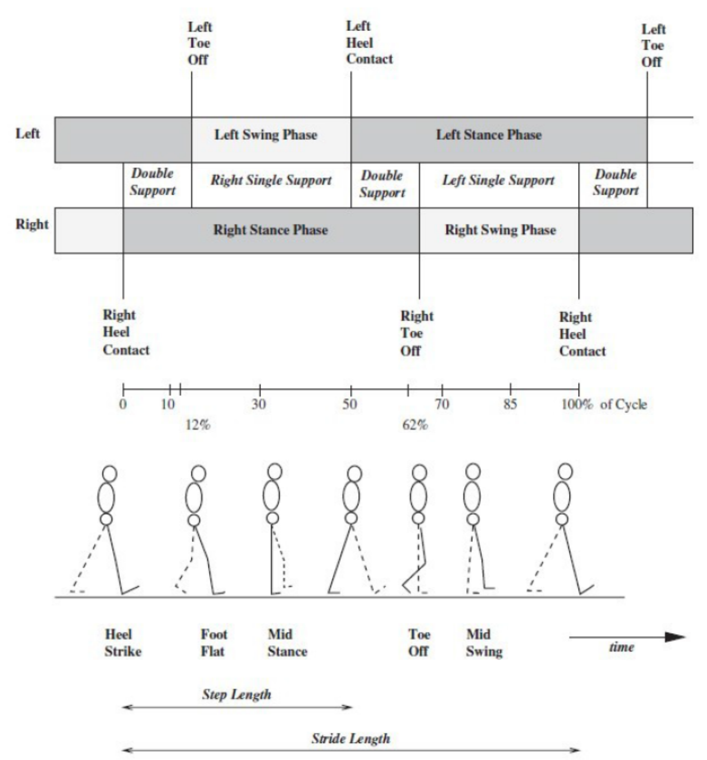
voice

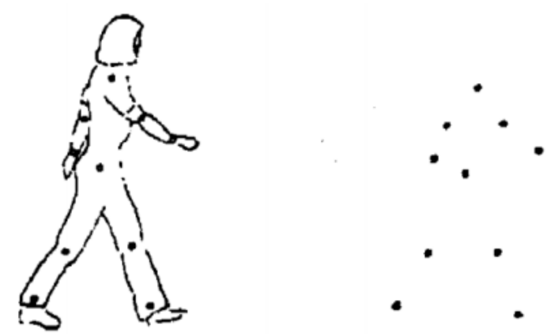


gait

Why gait works ?

- A plethora of technique and data continue to show that a person's walking is indeed unique





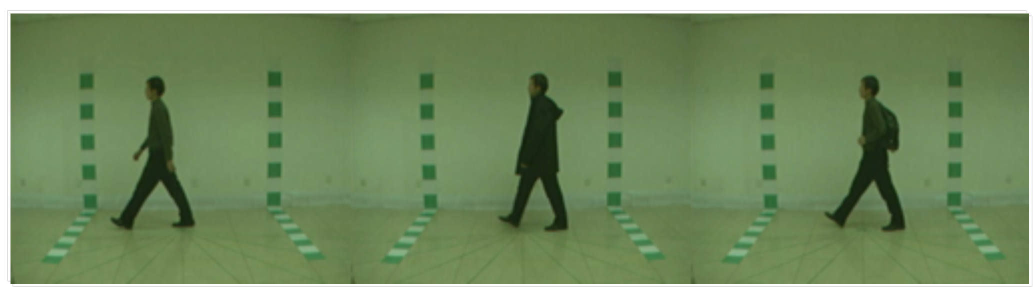
Murray M P, Drought A B, Kory R C. *Walking patterns of normal men*[J]. The Journal of Bone & Joint Surgery, 1964, 46(2): 335-360.
Johansson G. Visual perception of biological motion and a model for its analysis[J]. Attention, Perception, & Psychophysics, 1973, 14(2): 201-211.

Challenges

- Inconspicuous inter-class difference from the different people

Gait silhouettes of different subject

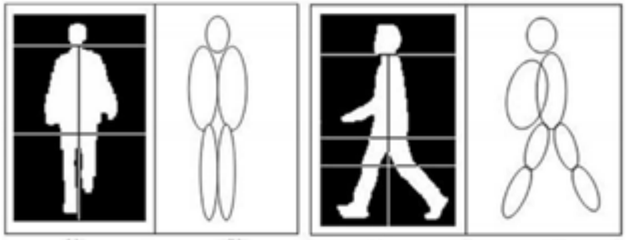
- The large intra-class variations from the same person
  - Walking speeds
  - Viewpoints
  - Clothing
  - Belongings
  - Occasion

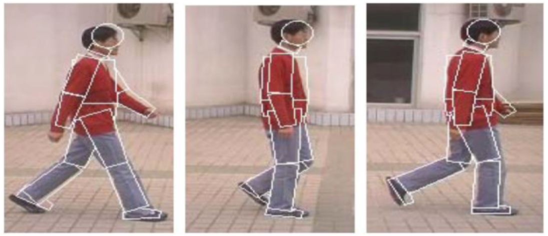


normal clothes backpack

Recent Efforts and Major Drawback

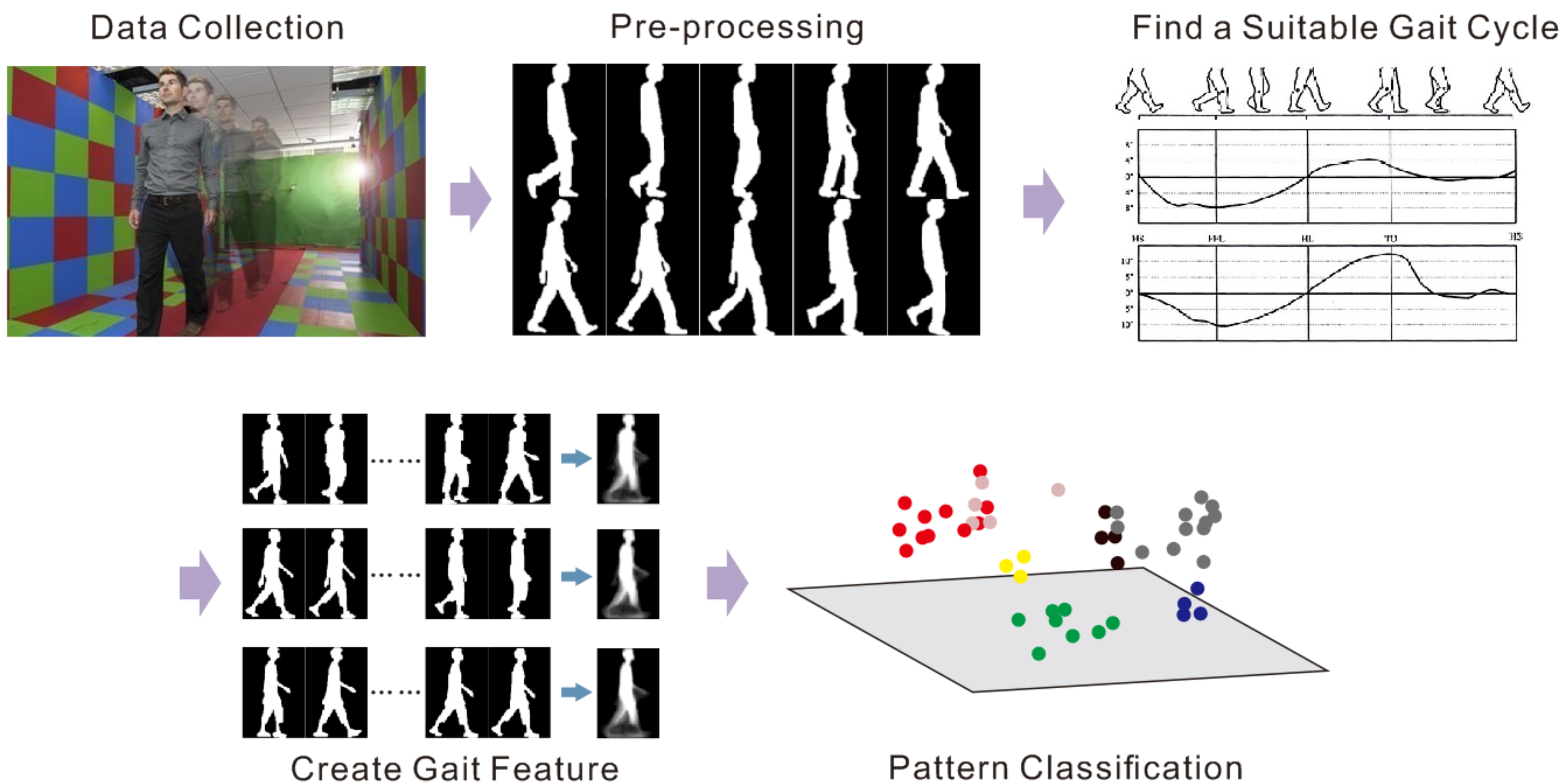
- Model-based methods
  - Extracting human body structure from the images
  - Requiring a high resolution as well as higher computational cost and are not yet suitable for outdoor surveillance





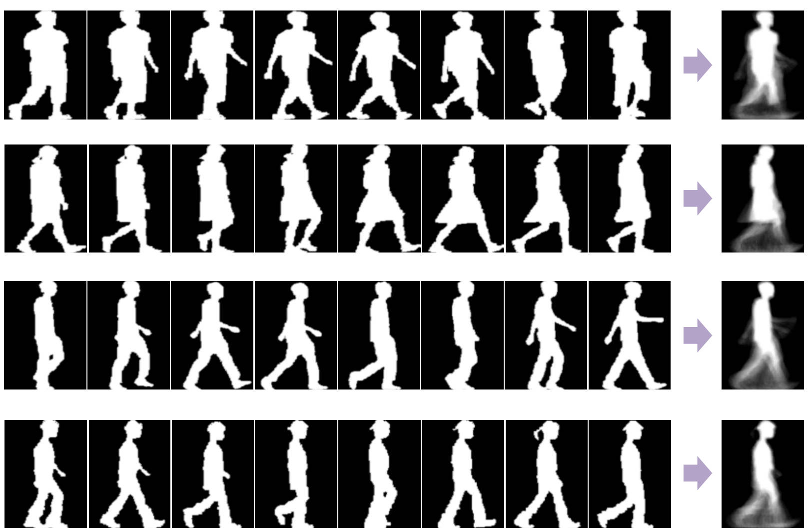
- Model-free methods
  - Using the whole motion pattern/features of the human body, and performing recognition at lower resolutions
  - Human-crafted gait features can extremely hard to break through feature representation bottleneck when facing with the gait and appearance changes

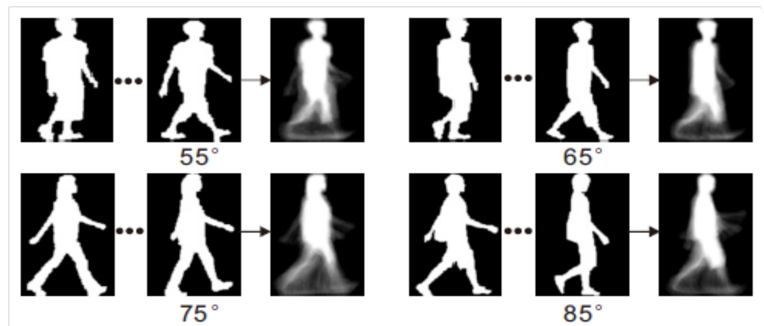
General Steps of Our System



Gait Energy Image

- Averaging of silhouette over one gait cycle
  - Represent a human motion sequence in a single image while preserving temporal information
  - Robust to incidental silhouette errors in individual image

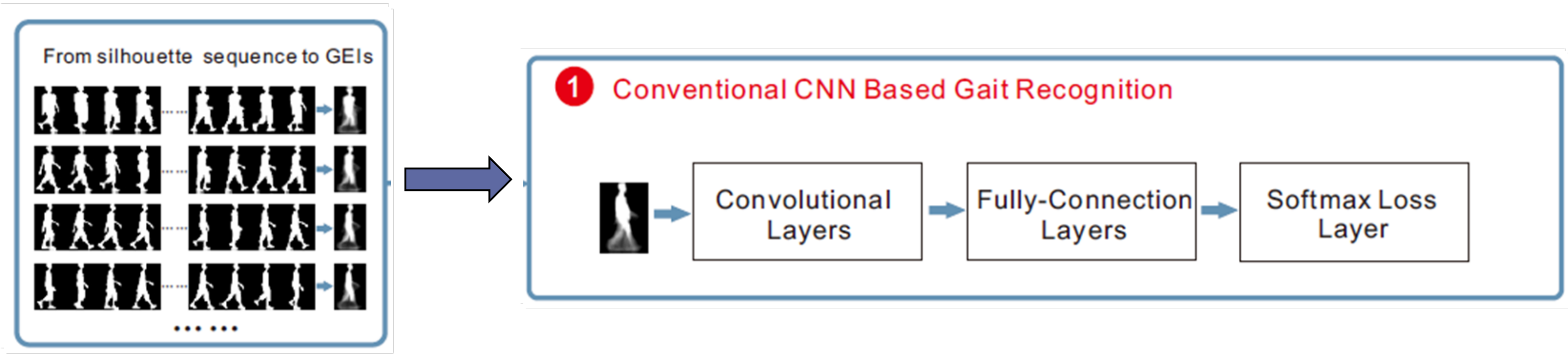




$$G(x,y) = \frac{1}{N}\sum_{t=1}^{N} I(x,y,t)$$

Conventional CNN based GR

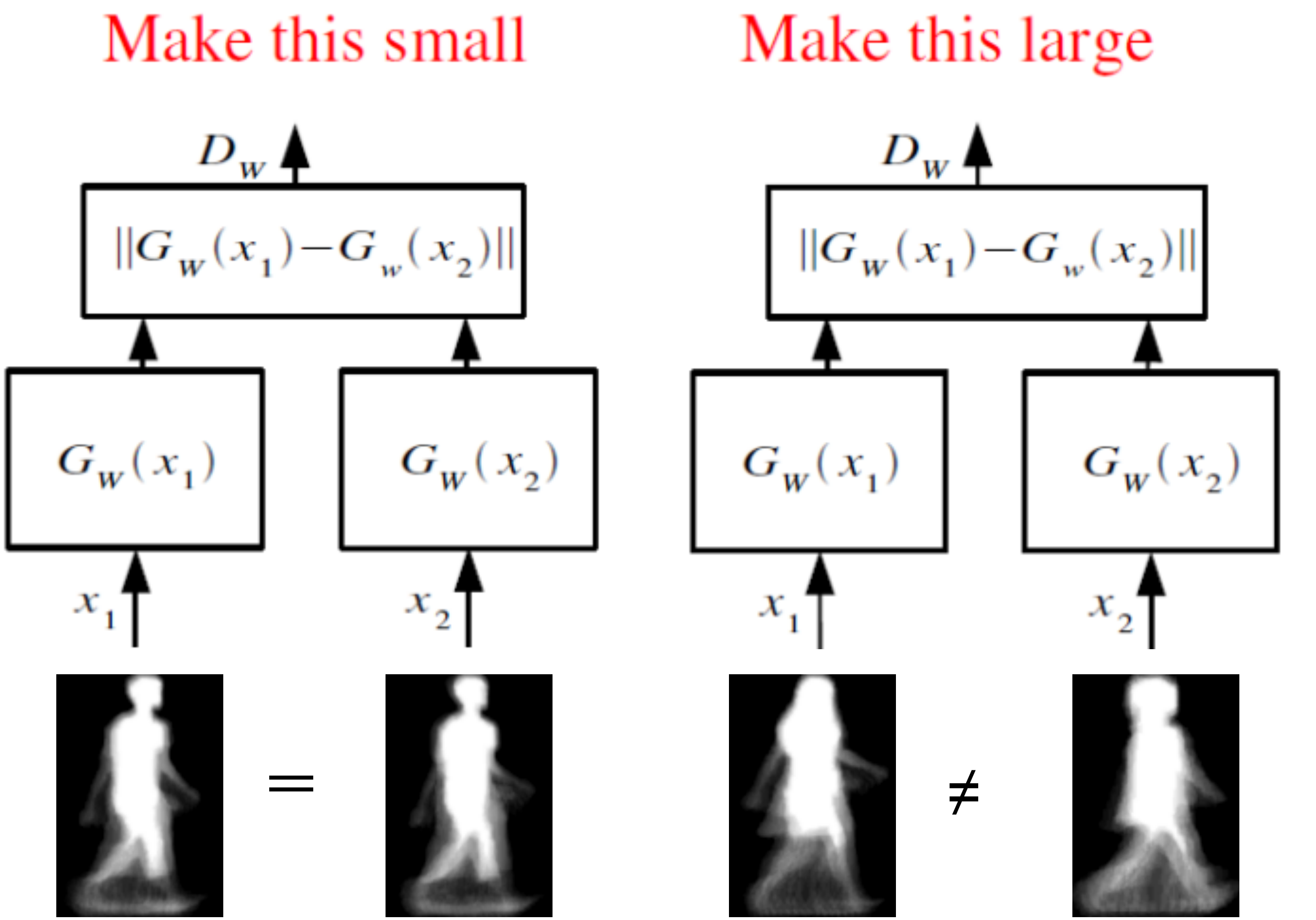
- Retrain the CNNs on the gait dataset
  - CNNs are able to learn discriminative features
  - Fine-tuning from a pre-trained model (e.g., AlexNet) is a good solution to solve the data limitation problem and speed up the convergence of new model
  - Employ the AlexNet and only change the 1,000 label output to the number of subjects in gait dataset



Problems

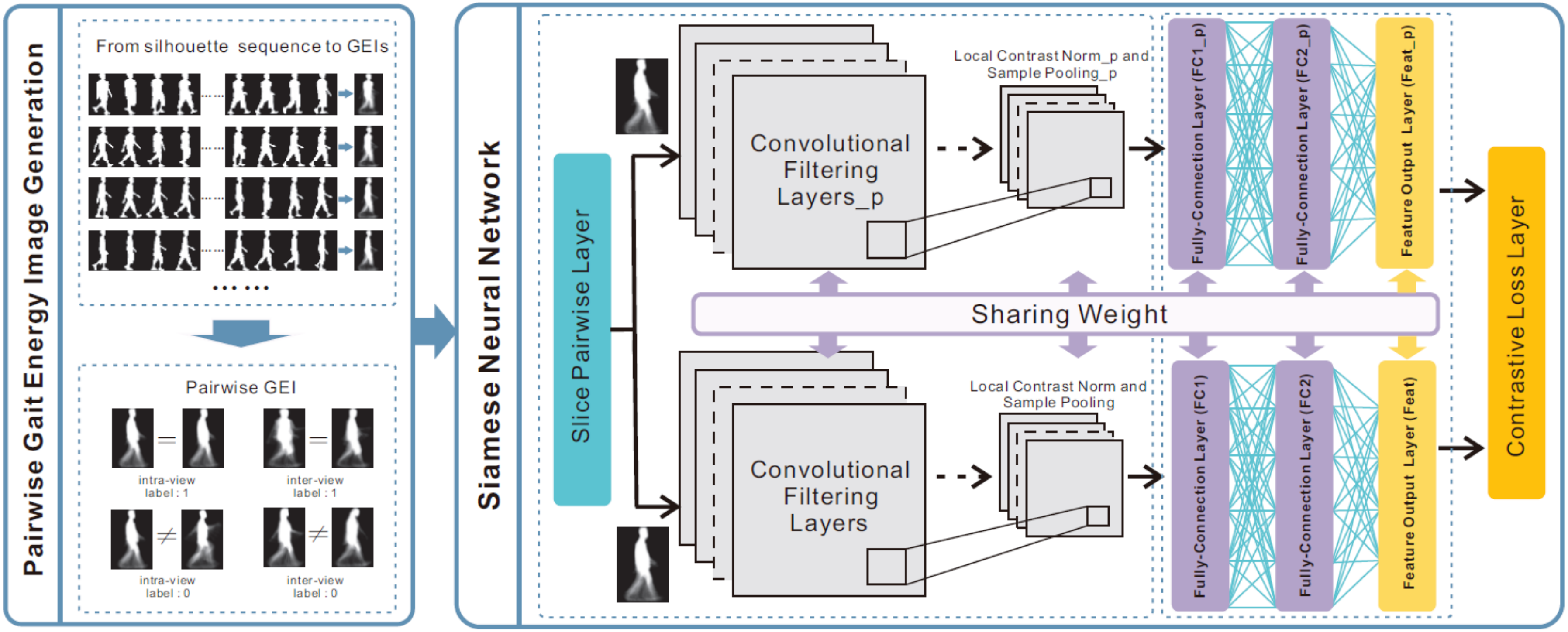
- Data limitation
  - To learn sufficient features, the CNN requires a mass of training data for all categories
  - For gait recognition, the number of subjects can be large, while with only a few examples per subject in public database
- Domain gap
  - Gait recognition for human identification is essentially a search problem but not classification

Metric Learning



Proposed Framework

- Siamese Neural Network based gait recognition



Sampling

- Training data is highly unbalanced
  - Using a sampler to generate equal number of positive and negative in each mini-batch, avoid overly biased towards to negative decisions
  - Using a sampler to enforce variety to prevent overfitting to a limited negative set
- Specially, the training set is selected from OULP-C1V1-A-Gallery dataset, with 20,000 similar GEI pairs and randomly selected 20,000 dissimilar pairs

Loss Function

- The distance $E_W(x_1, x_2)$ between a pair of GEIs can be measured by:

$$E_W(x_1, x_2) = ||S_W(x_1) - S_W(x_2)||_2^2$$

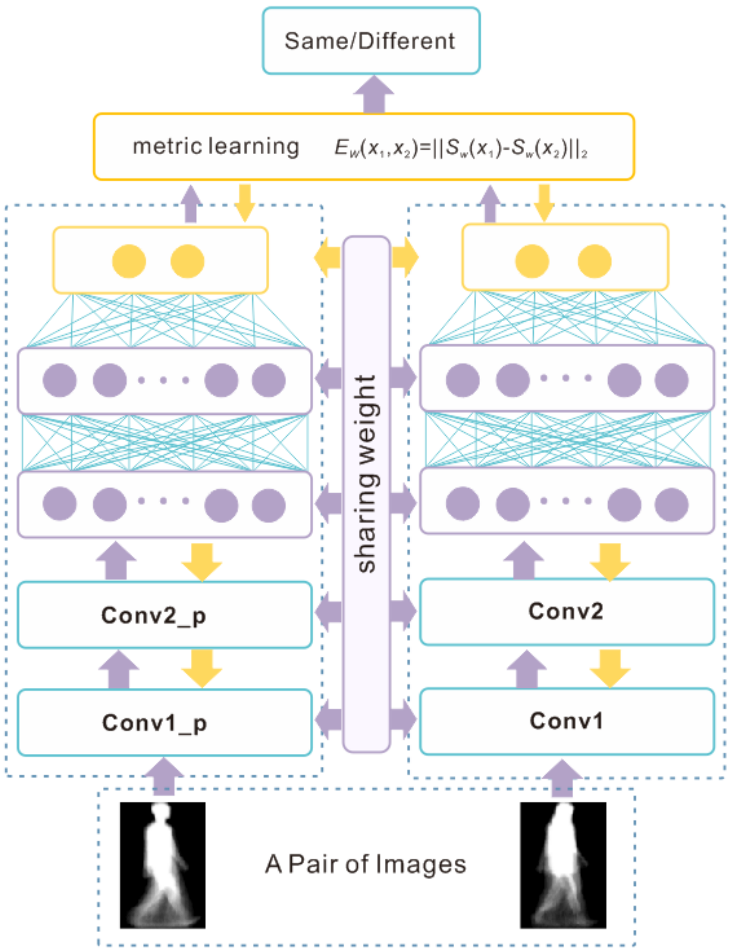
- We can define the contrastive function as follows:

$$\mathcal{L}(W) = \sum_{i=1}^{P} L(W, (y, x_1, x_2)^i)$$

$$\begin{aligned} L(W, (y, x_1, x_2)^i) = & (1 - y) \cdot max(m - E_W(x_1, x_2)^i, 0) \\ & + y \cdot E_W(x_1, x_2)^i \end{aligned}$$

Training and Feature Extraction

- Supervised setting
- Minimized the contrastive loss function over a training set of N patch pairs using stochastic gradient descent
- Experimented with different parameters and gave the best performance of feature representation



Experiments

- **Database**: OU-ISIR Large Population
- **Evaluation**: Rank-1 and Rank-5 identification rates
- **Baselines**: STOA gait recognition methods, i.e., GEI, FDF, HWLD, VTM, and RankSVM
- **Pipeline**: Background segmentation -> Periodic identification -> GEIs generation -> DNN training -> DNN feature extraction -> K-Nearest-Neighbor searching

Database

- OU-ISIR Large Population Gait Database
  - Contains the world's largest number of subjects
  - Records two sequences for each subject: probe (query) and gallery (source) sequence, offers fair comparison test bed

	Observation angle					
Dataset	55 [deg]	65 [deg]	75 [deg]	85 [deg]	All	Total
LP-C1V1-A	3,706	3,770	3,751	3,249	3,141	**3,835**
	(1,977/1,729)	(2,007/1,763)	(1,995/1,756)	(1,688/1,561)	(1,644/1,497)	(2,032/1,803)
LP-C1V1-B	3,998	4,005	4,002	3,923	3,904	**4,007**
	(2,129/1,869)	(2,133/1,872)	(2,133/1,869)	(2,073/1,850)	(2,061/1,843)	(2,135/1,872)

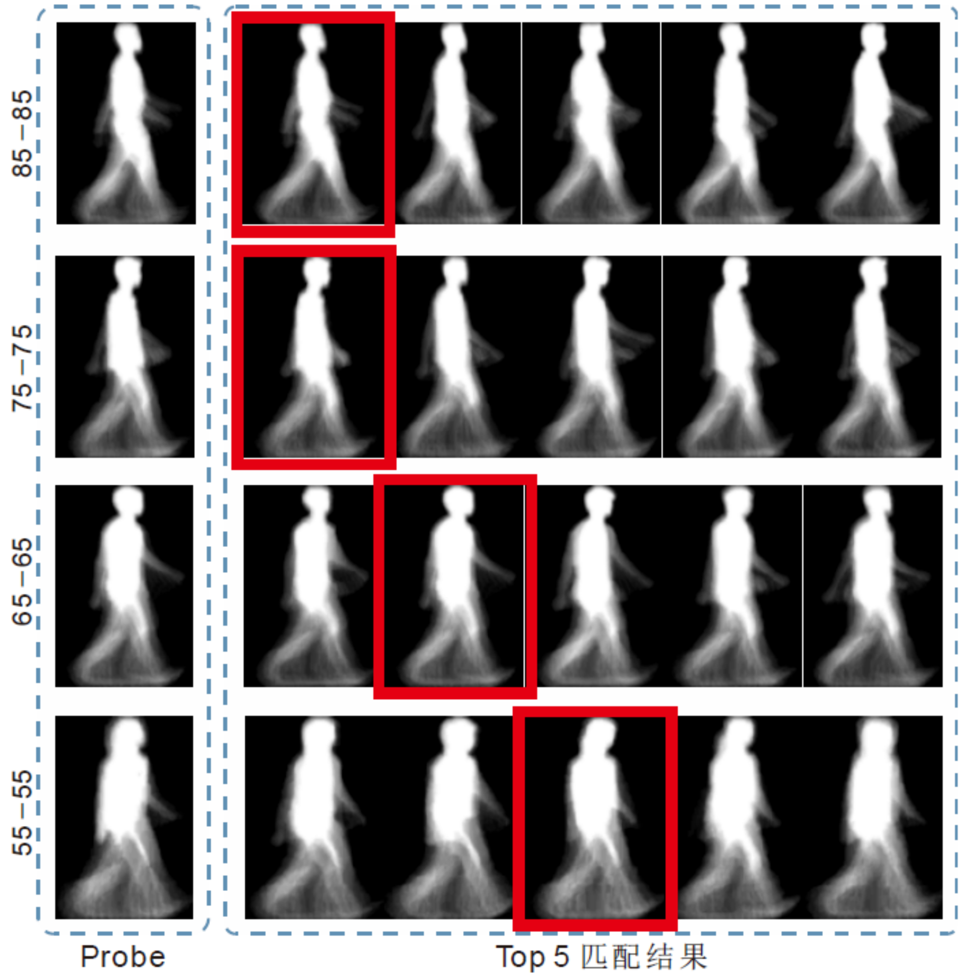
Intra-view recognition

	Rank-1 Identification Rate (%)					Rank-5 Identification Rate (%)				
Method	55	65	75	85	All	55	65	75	85	All
HWLD [7]	—	—	—	87.70	95.50	—	—	—	94.70	**98.50**
GEI [9]	84.70	86.63	86.91	85.72	94.24	92.39	92.84	92.78	93.01	97.13
FDF [9]	83.89	85.49	86.59	85.90	94.17	91.53	92.81	92.88	92.83	97.10
CNN.FC1	73.96	76.71	77.87	78.82	86.09	86.64	88.67	89.39	90.09	93.56
SiaNet.FC	**90.12**	**91.14**	**91.18**	**90.43**	**96.02**	**94.98**	**95.90**	**95.92**	**95.97**	98.31

Table 1. Comparison results of different methods in term of the Rank-1 and Rank-5 Identification Rates

[9] H. Iwama, M. Okumura, Y. Makihara, and Y. Yagi, “The ou-isir gait database comprising the large population dataset and performance evaluation of gait recognition,” *IEEE TIFS*.
[7] Sivapalan, D. Chen, S. Denman, S. Sridharan, and C. Fookes, “Histogram of weighted local directions for gait recognition,” in *CVPR*W, 2013.

Some Results



Inter-view recognition

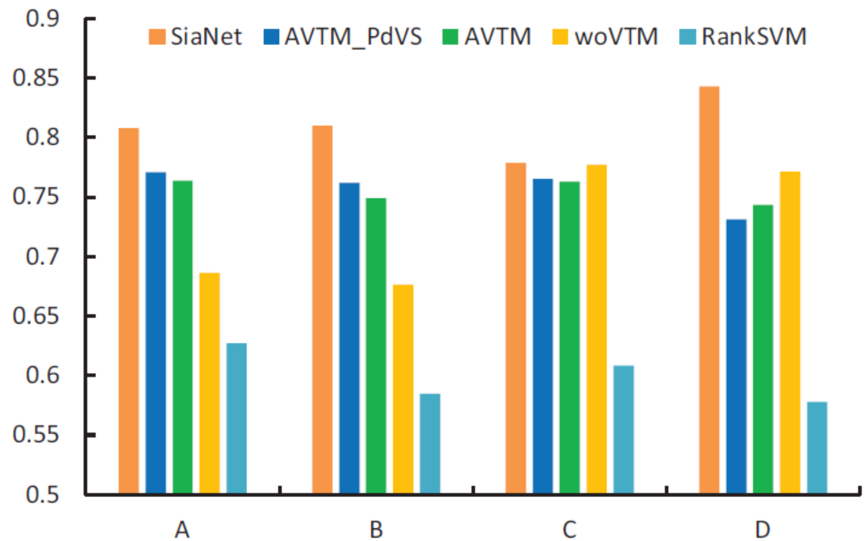


Fig. 2. Comparison of the cross-view matching approaches on different types of inter-degree test (in terms of rank-1 identification rate). Group A∼D stand for (65,75), (75,65), (75,85) and (85,75).

D. Muramatsu, A. Shiraishi, Y. Makihara, M. Uddin, and Y. Yagi, “Gait-based person recognition using arbitrary view transformation model,” *IEEE TIP*, 2015.
R. Mart´ ın-F´ elez and T. Xiang, “Gait recognition by ranking,” in *ECCV*, 2012.

Conclusions

- We present one of the first attempts to study the deep neural network based gait recognition for human identification with distance metric learning
- In the end-to-end framework, we leverage the competitive GEI presentation as the input of network while holistically exploit the Siamese neural network to learn effective feature representations for human identification
- The comprehensive evaluations show that we impressively outperform the state-of-the-arts on the world's largest challenge gait benchmark dataset

Future Works

- 3-Dimensional Siamese neural network
- Quasi-periodic or sub-frame gait recognition
- Unconstrained environment, like illumination changes, dark illumination, cluttered background, motion blur, and image compression noise
- …

"High'st Queen of state, Great Juno comes; I know her by her gait"

—— The Tempest [Act 4 Scene 1], Shakespeare

Any questions ?