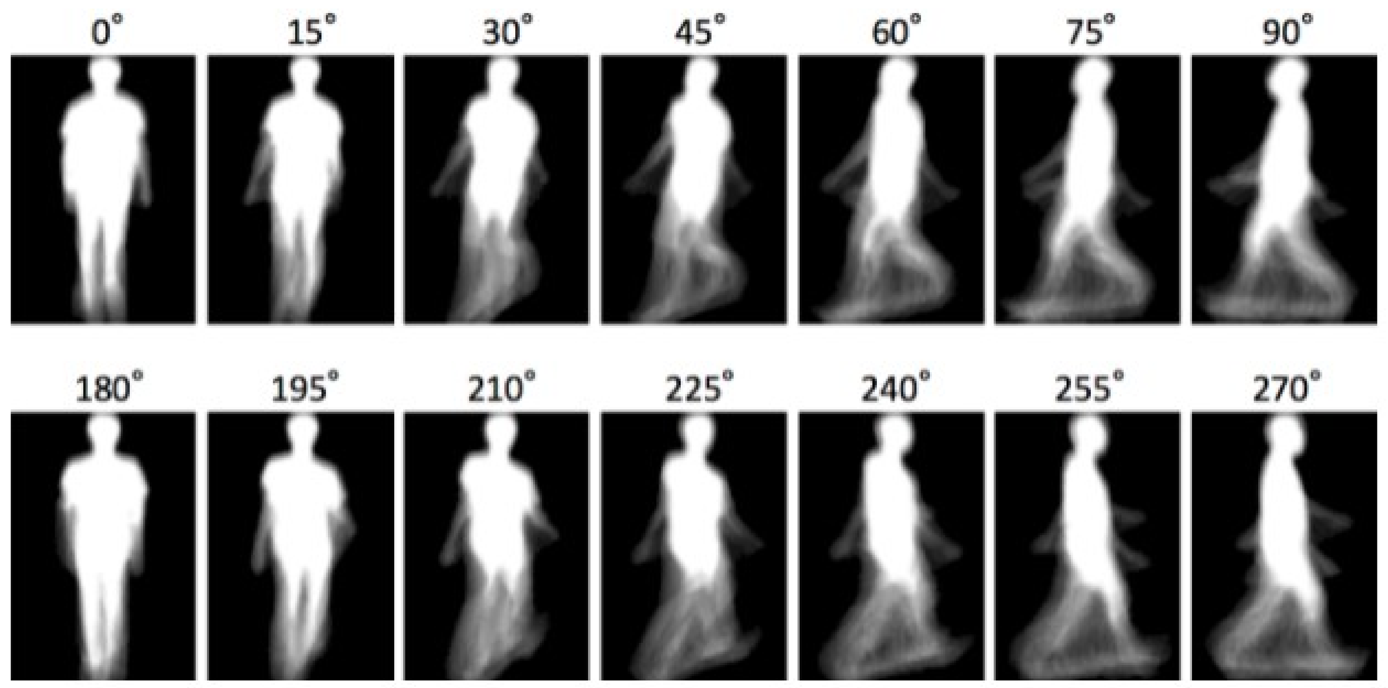
Silhouette based View embeddings for Gait Recognition under Multiple Views

Tianrui Chai, Xinyu Mei, Annan Li, Yunhong Wang
State Key Laboratory of Virtual Reality Technology and Systems,
School of Computer Science and Engineering, Beihang University, Beijing 100191, China.
{trchai,xymei,liannan,yhwang}@buaa.edu.cn

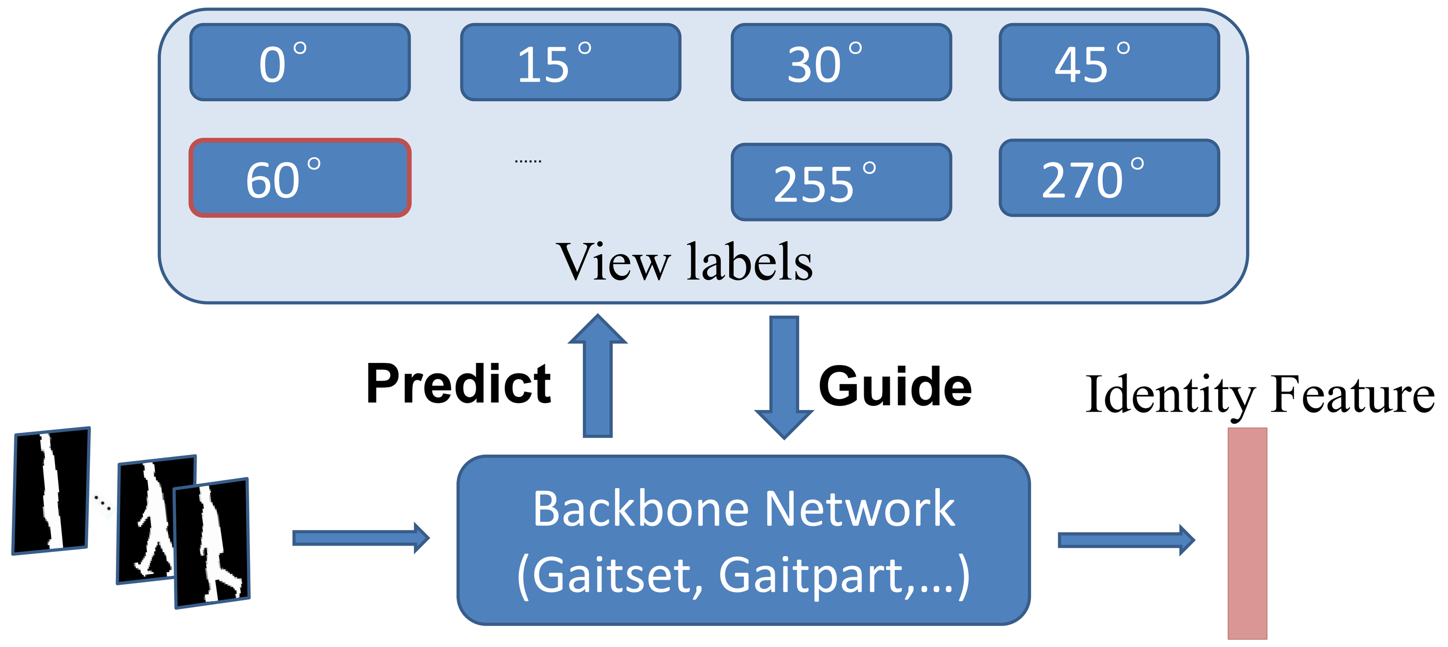
Paper ID:2121
IEEE ICIP 2021

Introduction:

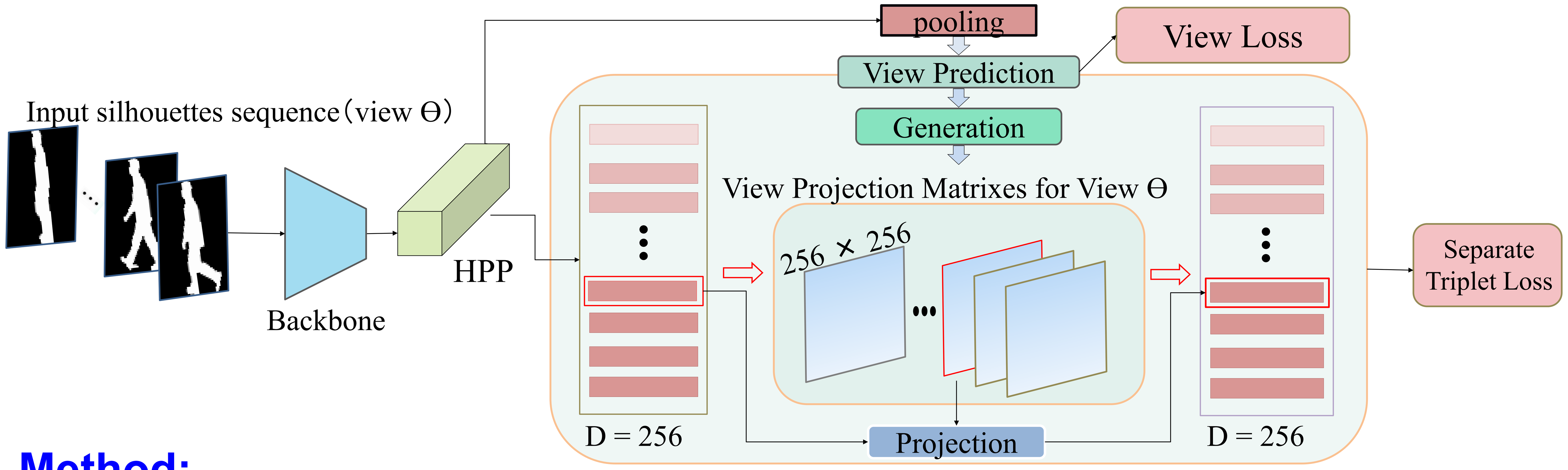
- Gait is a biometric presenting the walking style of people and has an edge over other biometrics because it can be recognized at a distance with much less cooperation.
- Viewpoint differences is a very tricky problem, because it may bring greater visual differences than the identity.



- Many methods have been proposed based on silhouette image sequences, for example, GaitSet, GaitGL, GaitPart.
- However, all of these methods mainly focus on how to extract features through spatial-temporel modeling. None of them take view itself into consideration.



- In this paper, we proposed a general framework for multi-view gait recognition by explicit view angle embedding, based on which, two state-of-the-art gait recognition backbones, i.e.Gaitset and GaitGL are enhanced. Compared with the original ones, the enhanced ones, improve the performance. The effectiveness is well demonstrated by the experiments on CASIA-B and OUMVLP datasets.



Method:

- The feature map X_f is calculated form the input X_{in} using the backbone E as well as the view feature f_v:

$$X_f = E(X_{in}) \;\; and \;\; f_v = F(P_{Global\_Avg}(X_f))$$

- Especially for Gaitset, there is another feature map X_g, then the view feature is defined as:

$$f_v = F(P_{Global\_Avg}([X_f; X_g]))$$

- Where F denotes a fc layer. Then the predicted view can be expressed as:

$$\hat{p} = W_{view} f_v + B_{view} \;\; and \;\; \hat{y} = \arg\max_i \hat{p}_i$$

- Where $\hat{y} \in \{0,1,2,\dots M\}$ and M is the number of views (10 for CASIA-B and 14 for OU-MVLP). The corresponding set of projection matrixes is:

$$Z_{\hat{y}}| = \{W_i | i = 1,2,\dots,n\}$$

- Assuming the feature obtained after HPP module are:

$$f_{HPM} \in \mathbb{R}^{n \times D}$$

- Then the final Identity feature is:

$$f_{final,i} = W_i f_{HPM,i}$$

$$f_{final} = [f_{final,1}, f_{final,2} \dots, f_{final,n}]$$

- The joint loss is

$$\mathcal{L}_{CE} = -\sum_{i=1}^{N}\sum_{i=1}^{M} y_j log(p_{ji}) \;\; w.r.t. \;\; p_{ji} = \frac{e^{\hat{p}_{ji}}}{\sum_{i=1}^{M} e^{\hat{p}_{ji}}}$$

$$\mathcal{L}_{trip} = \frac{1}{K}\sum_{i=1}^{K}\sum_{j=1}^{n} max(m - d_{ij}^{-} + d_{ij}^{+}, 0)$$

$$\mathcal{L} = \lambda_{CE}\mathcal{L}_{CE} + \lambda_{trip}\mathcal{L}_{trip}$$

Experiments:

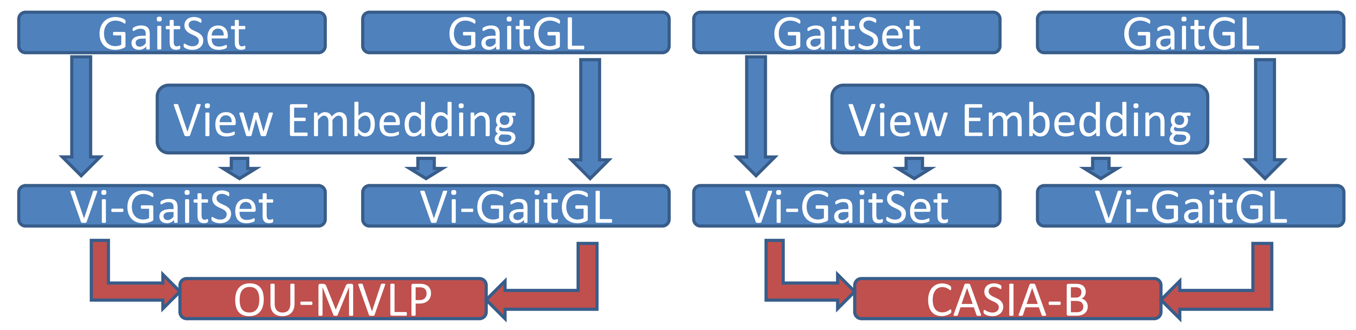


Table 2. Rank-1 accuracy (%) on OU-MVLP under 14 probe views excluding identical-view cases.

Probe angle	Gallery All 14 views					
	GEINet	Gaitset	Vi-Gaitset	GaitPart	GaitGL	Vi-GaitGL
0°	11.4	79.5	**81.8**	82.6	84.3	**85.6**
15°	29.1	87.9	**89.2**	88.9	89.8	**90.2**
30°	41.5	89.9	**90.5**	90.8	90.8	**91.2**
45°	45.5	90.2	**90.5**	91.0	91.0	**91.5**
60°	39.5	88.1	**89.2**	89.7	90.5	**91.1**
75°	41.8	88.7	**89.5**	89.7	90.5	**90.9**
90°	38.9	87.8	**89.0**	89.9	90.3	**90.4**
180°	14.9	81.7	**83.9**	85.2	88.1	**88.3**
195°	33.1	86.7	**88.1**	88.1	87.9	**88.7**
210°	43.2	89.0	**89.7**	90.0	89.6	**90.6**
225°	45.6	89.3	**89.8**	90.1	89.8	**90.6**
240°	39.4	87.2	**88.6**	89.0	88.9	**90.1**
255°	40.5	87.8	**88.5**	89.1	88.9	**89.9**
270°	36.3	86.2	**87.6**	88.2	88.2	**89.4**
mean	35.8	87.1	**88.3**	88.7	89.1	**89.9**

（More experiments of dataset CASIA-B can be find in the paper)

Conclusion

- Combined the task of view prediction and gait recognition.
- Proposed a general view embedding framework for improving multi-view gait recognition.
- The proposed framework with GaitSet and GaitGL as the backbone meets the state-of-the-art on two large-scale public gait datasets.