

# Supplementary Materials: AerialGait: Bridging Aerial and Ground Views for Gait Recognition

ANONYMOUS AUTHORS

## 1 ARCHITECTURES OF FEATURE EXTRACTION BACKBONE, $f_\mu(\cdot)$ , AND $f_\sigma(\cdot)$

We provide the architectures of the feature extraction backbone, along with the functions  $f_\mu(\cdot)$  and  $f_\sigma(\cdot)$  within the Feature-Oriented Uncertainty Learning module, as detailed in Table 4. To address frame discontinuity, we employ a 2D convolutional backbone. Moreover, the architectures of  $f_\mu(\cdot)$  and  $f_\sigma(\cdot)$  are designed to be light-weighted. Specifically, "Conv2D( $3 \times 3$ , 128, 256)" represents a 2D convolutional layer whose kernel size is  $3 \times 3$ , input channel is 128 and the output channel is 256. "Maxpool( $2 \times 2$ )" refers to a max pooling layer with a kernel size  $2 \times 2$ . "BN" denotes the batch normalization layer. "ReLU" signifies the rectified linear unit function.

Table 4. Architectures of Feature Extraction Backbone,  $f_\mu(\cdot)$ , and  $f_\sigma(\cdot)$

Module	Architecture	Module	Architecture
Feature Extraction Backbone	Conv2D( $3 \times 3$ , 1, 64)	$f_\mu(\cdot)$	Conv2D( $3 \times 3$ , 256, 512)
	-> BN()-> ReLU()		-> BN()-> ReLU()
	-> Conv2D( $3 \times 3$ , 64, 64)		Conv2D( $3 \times 3$ , 512, 512)
	-> BN()-> ReLU()		-> BN()-> ReLU()
	-> MaxPool( $2 \times 2$ )		
	-> Conv2D( $3 \times 3$ , 64, 128)	$f_\sigma(\cdot)$	Conv2D( $3 \times 3$ , 256, 512)
	-> BN()-> ReLU()		-> BN()-> ReLU()
	-> Conv2D( $3 \times 3$ , 128, 128)		
	-> BN()-> ReLU()		
	-> MaxPool( $2 \times 2$ )		
	-> Conv2D( $3 \times 3$ , 128, 256)		
	-> BN()-> ReLU()		
	-> Conv2D( $3 \times 3$ , 256, 256)		
	-> BN()-> ReLU()		