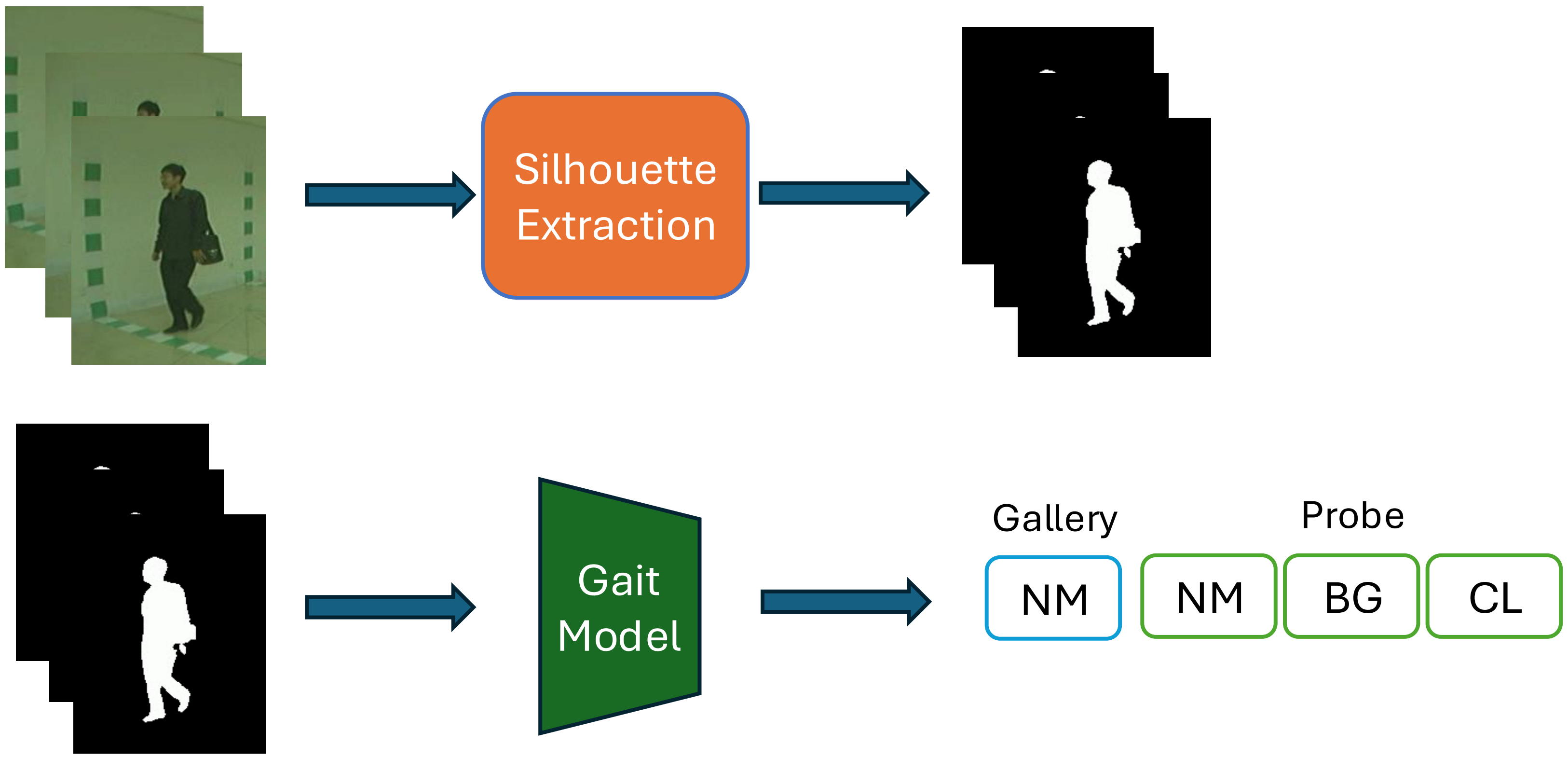
RobustGait: Robustness Analysis for Appearance-Based Gait Recognition

Reeshoon Sayera, Akash Kumar, Sirshapan Mitra, Prudvi Kamtam, Yogesh S Rawat
University of Central Florida

WACV 2026
TUCSON, AZ
3/6 – 3/10

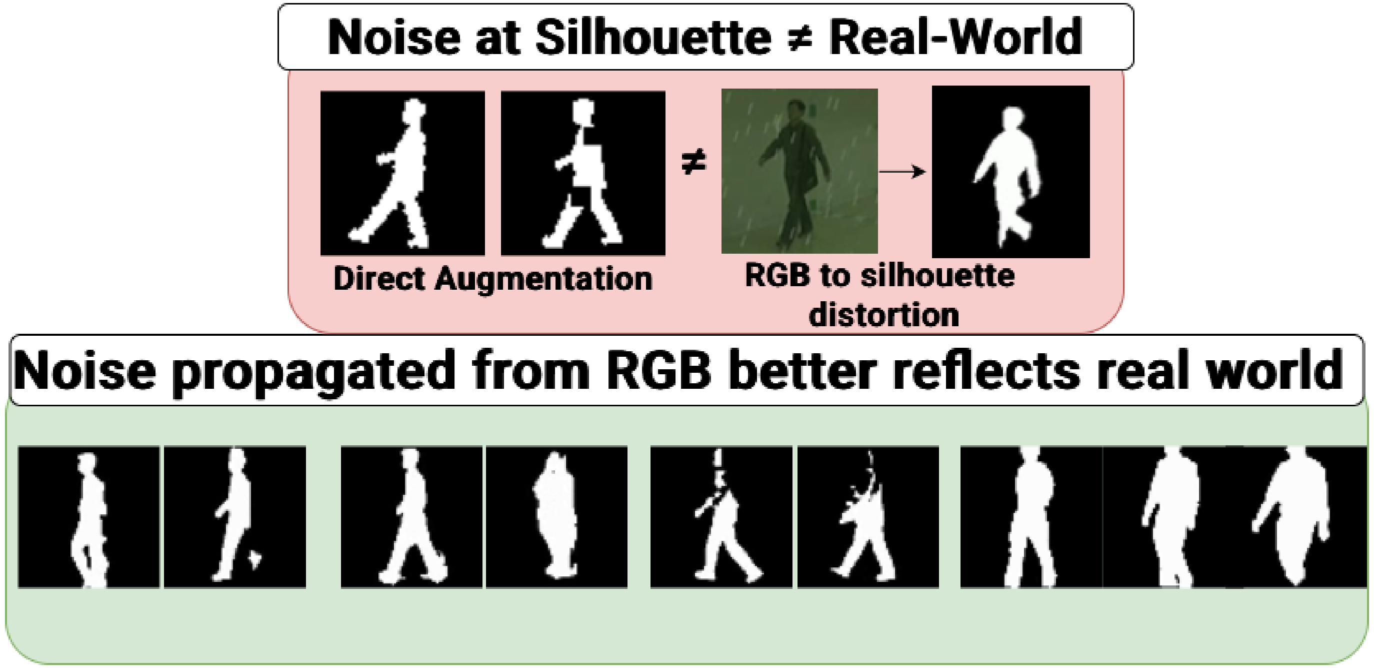
Problem Statement

- **Gait recognition** involves a **two-stage** system.



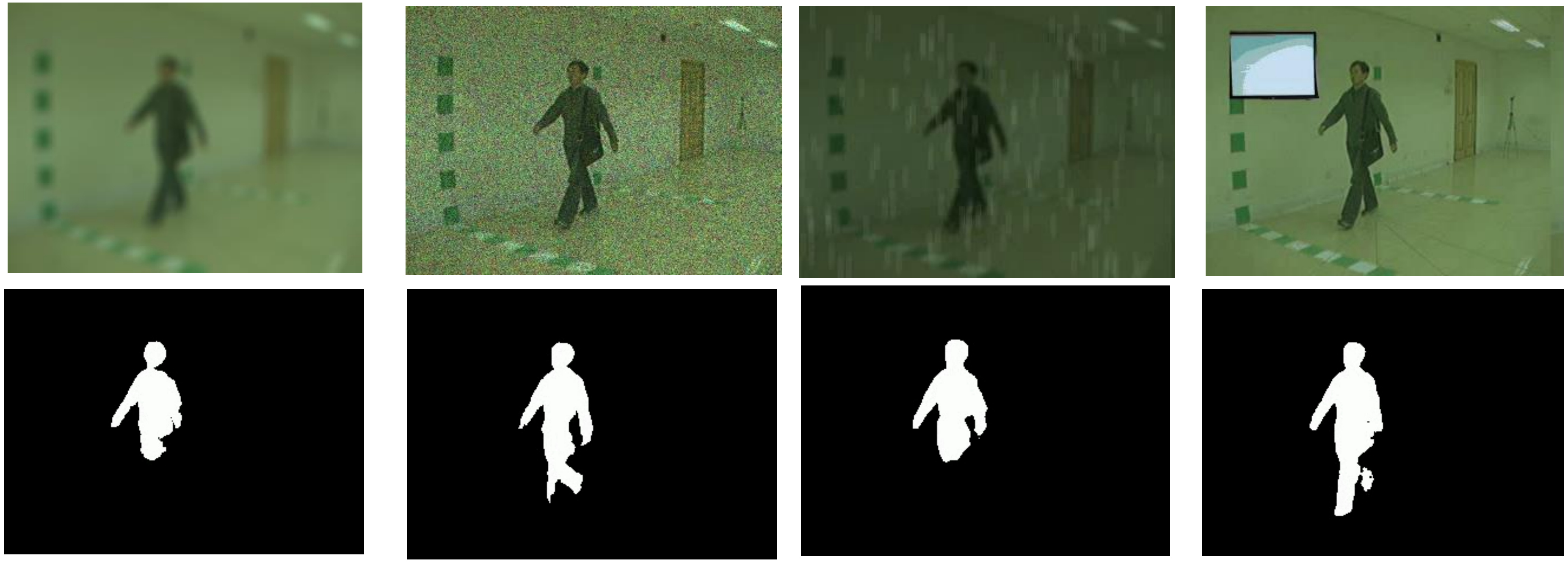
Problem Statement

- Directly corrupting silhouettes limits perturbations to simple augmentations like flip, rotate and random erasing.
- Applying noise at the RGB level enables realistic **temporal**, **environmental**, and **digital** degradations to naturally propagate through silhouette extraction.



Problem Statement

- **Noise in RGB** data can arise from various sources.

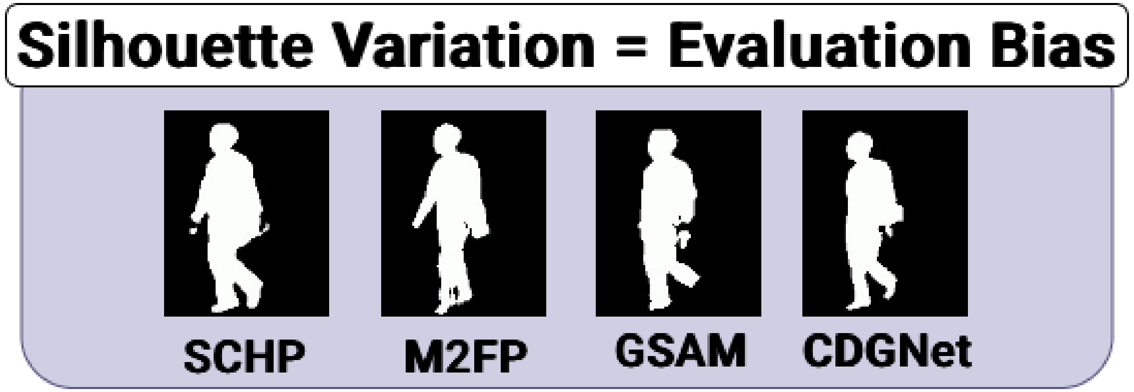


- Gait models are ***rarely*** tested under such distribution shifts.

Evaluating RGB noise propagation to silhouettes is vital for gait recognition robustness.

Problem Statement

- Variation across silhouette extractors leads to **evaluation bias** due to **variable silhouette quality.**



This highlights the need for standardized extraction methods across benchmarks.

Overview

- **Benchmark to evaluate gait recognition robustness under real-world RGB corruptions and silhouette extractor bias.**

- The benchmark covers:
  - **3** major gait datasets.
  - Covers **15 real-world corruptions** types across **5 severity levels.**
  - Benchmarks **6 state-of-the-art gait recognition models** spanning diverse architectures and **4 silhouette extraction methods**.

Benchmark Setup

Perturbations

- **Digital**: Noise introduced by *camera sensor*


Gaussian noise


Shot noise


Defocus blur


Original


Impulse noise


Speckle noise


Zoom blur

Perturbations

- **Temporal**: Affect *frame consistency*

Freeze

Sampling

Zoom In

Perturbations

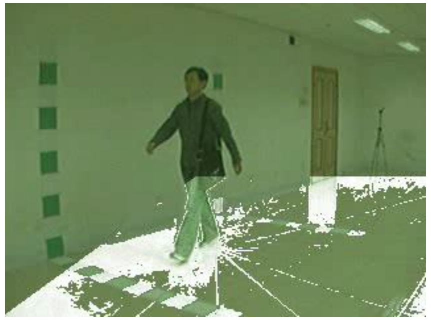
- **Environmental**: Mimic weather effects

Original

Low light

Fog

Rain

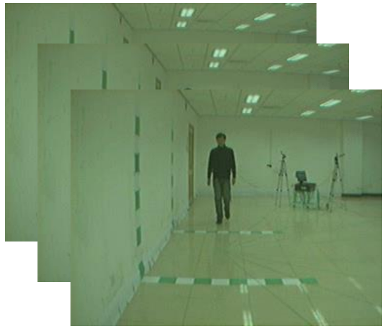


Snow

Perturbations

- **Occlusion**: Simulates real world obstructions by *introducing static objects* into scenes

Datasets



CASIA-B

SUSTech1k

CCPG

DATASET	ENVIRONMENT	#ID	CONDITIONS	#CAM	#SEQ
CASIA-B	Indoor	124	Normal, Bag, Clothing	11	13.6k
SUSTech1k	Outdoor	1050	Clothing variations	12	25.2k
CCPG	Hybrid	200	Occlusions, Illumination, etc.	10	16k

Models

Silhouette Extraction Models

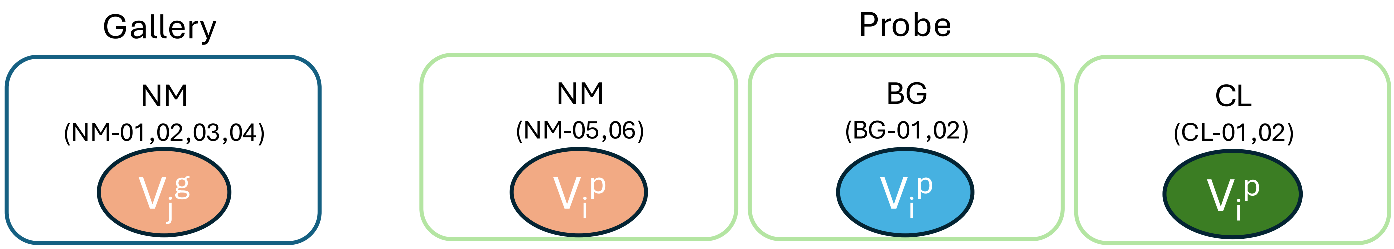
- Single Human (**SCHP, CDGNet**)
- Multiple Human (**M2FP**)
- Foundation Models (**Grounded SAM**)

Gait Recognition Models

- Smaller Capacity *(11-23M parameters)*
  - **GaitSet**
  - **GaitPart**
- Medium Capacity *(23-43M parameters)*
  - **GaitBase**
  - **DeepGaitV2**
- Large Capacity *(50M+ parameters)*
  - **SwinGait**

Evaluation Metrics

- Gait Recognition
  - **Single View Rank-1 accuracy**



Gallery Subset	**Probe Subset**	**View Pair**	**Walking Condition Pair**	**Rank-1 Accuracy**
V_j^g	V_i^p	(V_{ip} , V_{jg})	NM-NM	NM@1
	V_i^p		NM-BG	BG@1
	V_i^p		NM-CL	CL@1

Evaluation Metrics

- Robustness Analysis
  - **Absolute robustness:** Overall performance degradation (% of total scale).

$$\delta_a = 1 - \frac{D_c - D_p}{100}$$

  - **Relative robustness:** Measures proportional performance drop relative to clean data.

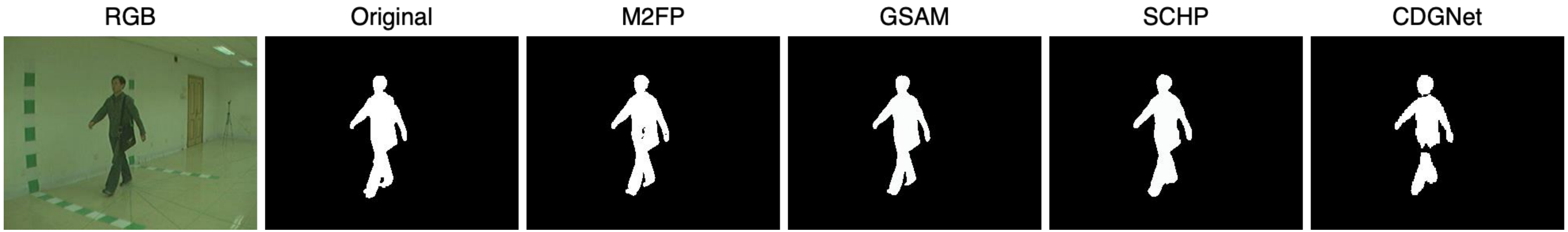
$$\delta_r = 1 - \frac{D_c - D_p}{D_c}$$

Benchmark Analysis

- RobustGait evaluation spans four dimensions:
  - the silhouette extraction method(segmentation and parsing networks)
  - the type of perturbation (digital, environmental, temporal, occlusion)
  - the architectural capacities of gait recognition models
  - various deployment scenarios

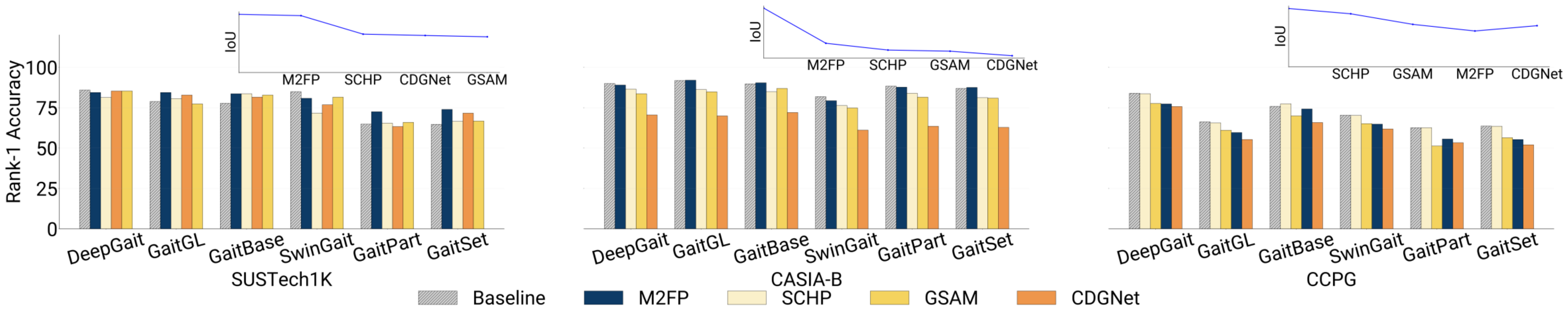
Impact of Silhouette Extraction

- Different gait datasets use different models for silhouette extraction.
- **CASIA-B** initially used a basic **background subtraction algorithm** for silhouette extraction.
- **CCPG** uses **U-Net,** while **SUSTech1k** uses **PaddleSeg** for segmentation.
- This **inconsistency** leads to **unfair comparison of models across datasets.**



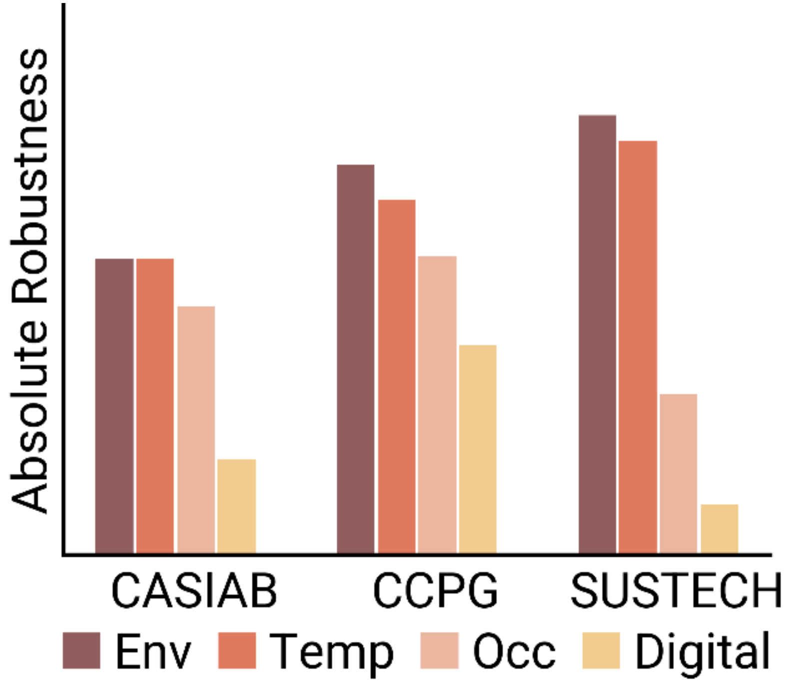
Silhouette Quality vs Recognition Performance

- Higher segmentation IoU (mask quality) corresponds to higher Rank-1 recognition accuracy across gait models.



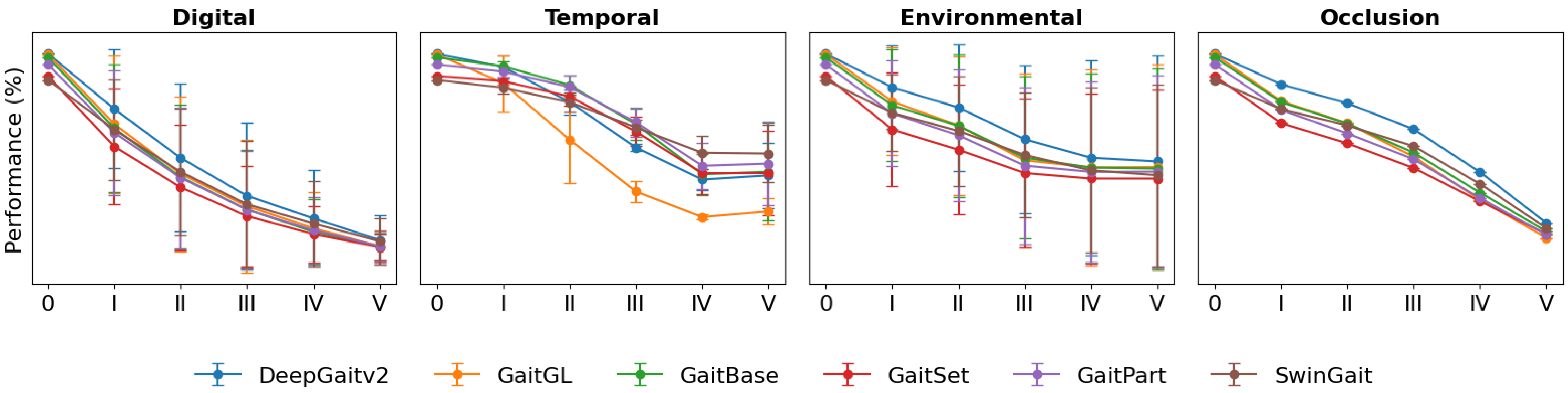
Robustness by Noise Types

- **Digital Noise significantly impairs gait recognition.**
  - Models are most robust to temporal and environmental noise.
  - Digital noise distorts pixel values while temporal noise only affects part of the gait sequence.



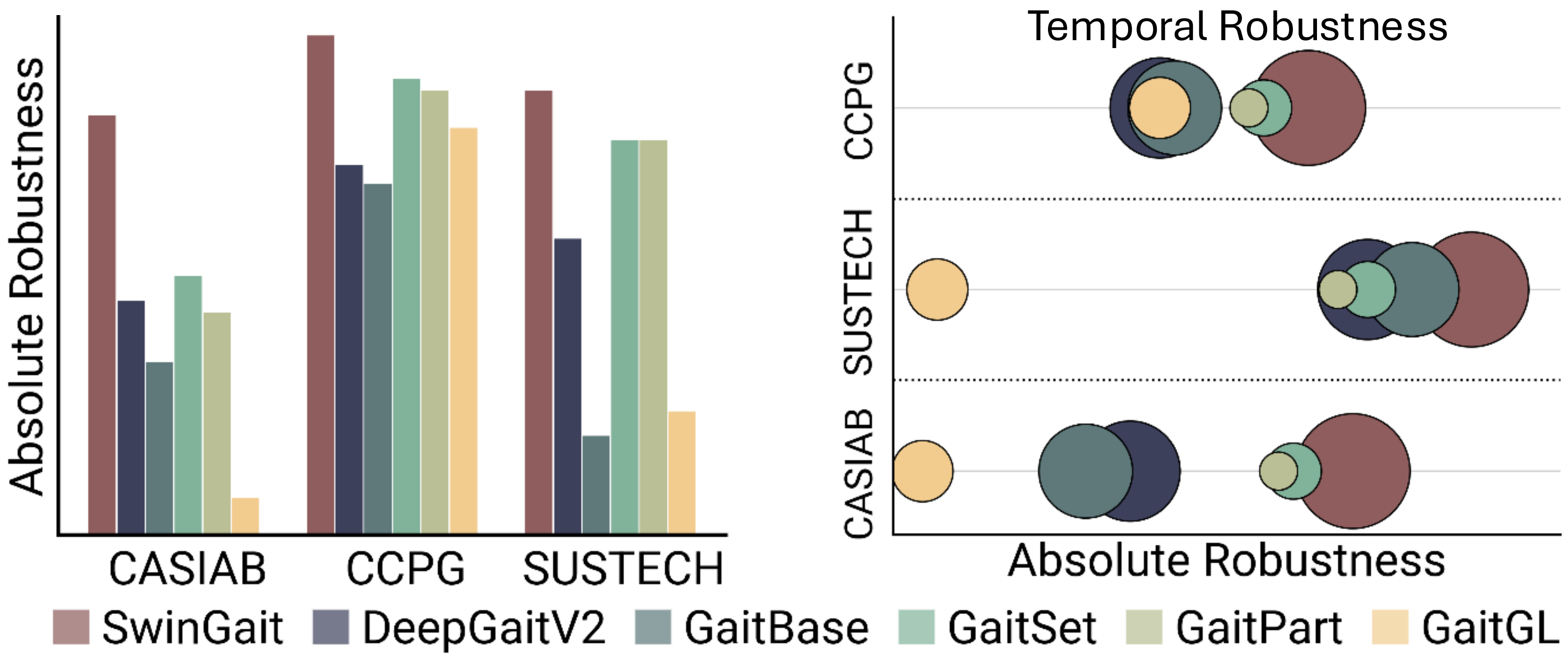
Robustness by Noisy Types

- **Digital and Occlusion severity impacts most.**



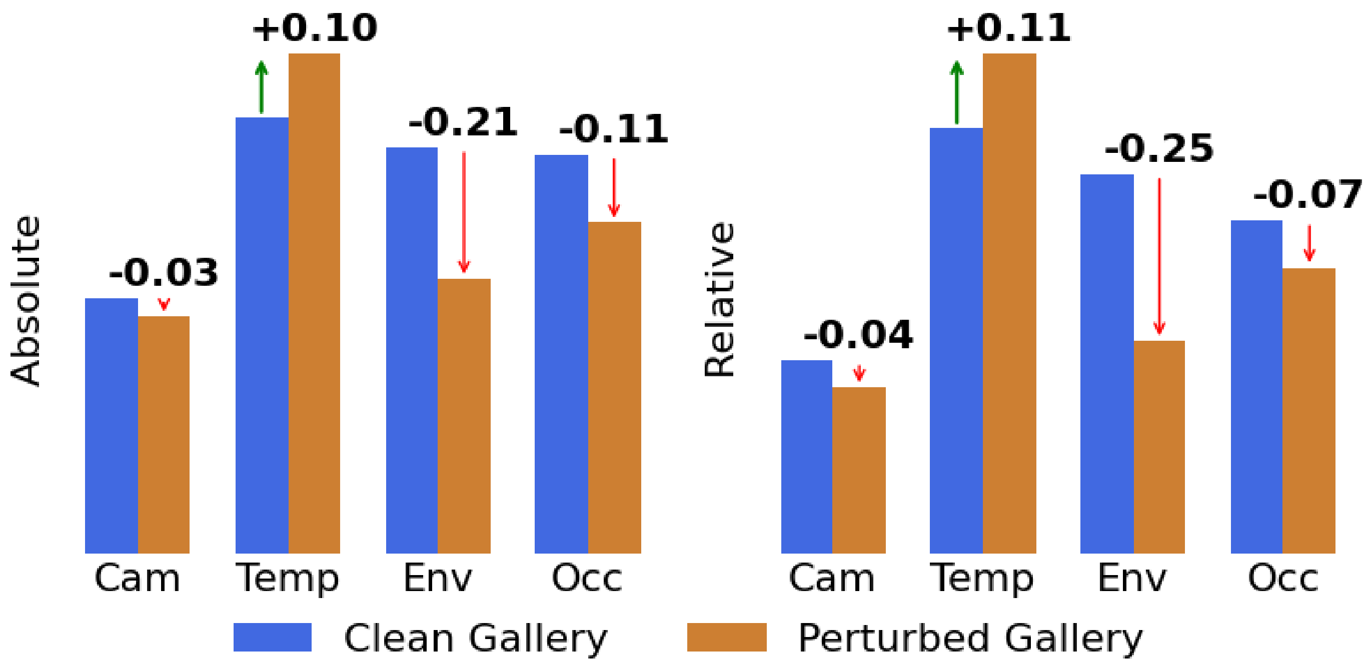
Robustness by Model Capacities

- Transformers outperform CNNs in resilience to perturbations.
- Smaller capacity set-based models are more robust than larger capacity models to temporal noise.



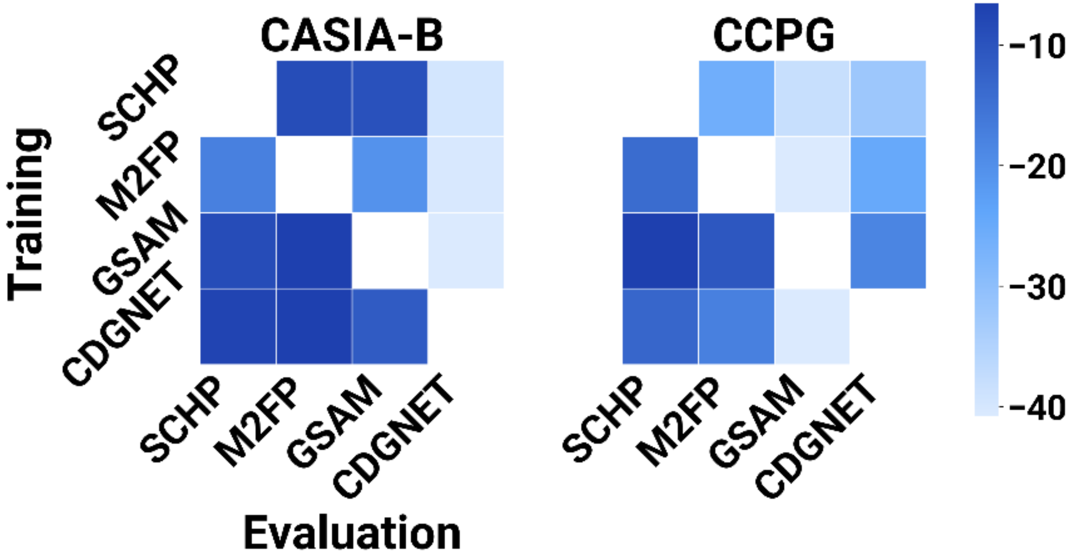
Robustness in Deployment Scenarios

- **Scenario 1:** Both probe and gallery are noisy to simulate real-world surveillance conditions.
- **Observation:** Models trained on clean data overfit to clean features, struggling in noisy probe-gallery settings.



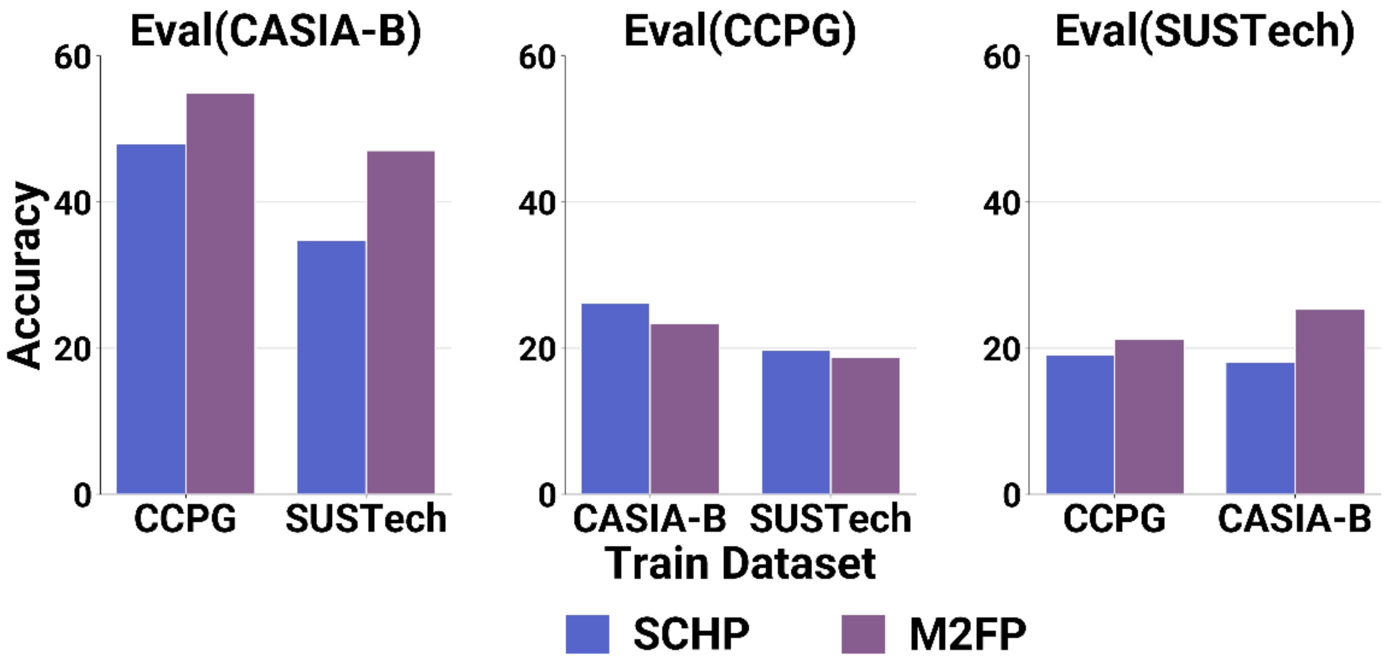
Robustness in Deployment Scenarios

- **Scenario 2:** Cross-extractor evaluation: train on one, test on another.
- **Observation**: Accuracy drops when training and testing use different extraction pipelines.



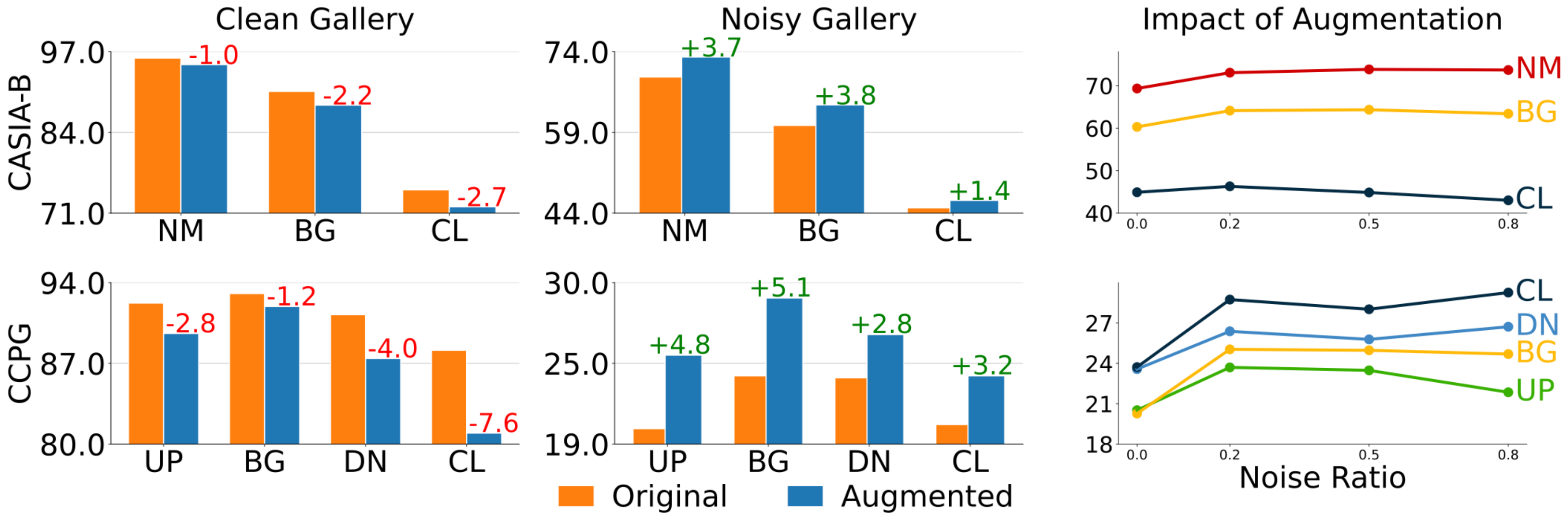
Robustness in Deployment Scenarios

- **Scenario 3:** Same silhouette extractor, but different training and evaluation datasets.
- **Observation:** M2FP performs better on CASIA-B and SUSTech, while SCHP excels on CCPG.



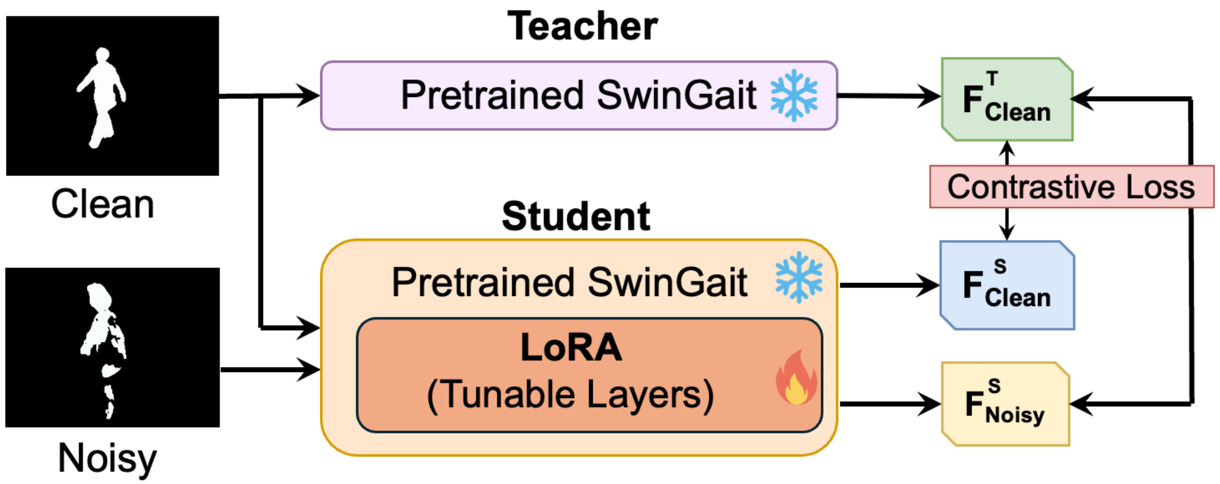
Improving Robustness: Noise Aware Training

- Training with noisy silhouettes improves robustness.
- Slight drop in clean accuracy (forgetting).
- ~25–30% noise is sufficient (diminishing returns).



Improving Robustness: Noise Aware Training

- LoRA-based distillation improves noise robustness.
- Preserves clean accuracy (no forgetting).
- Achieves stronger robustness than noise only training



Thank You