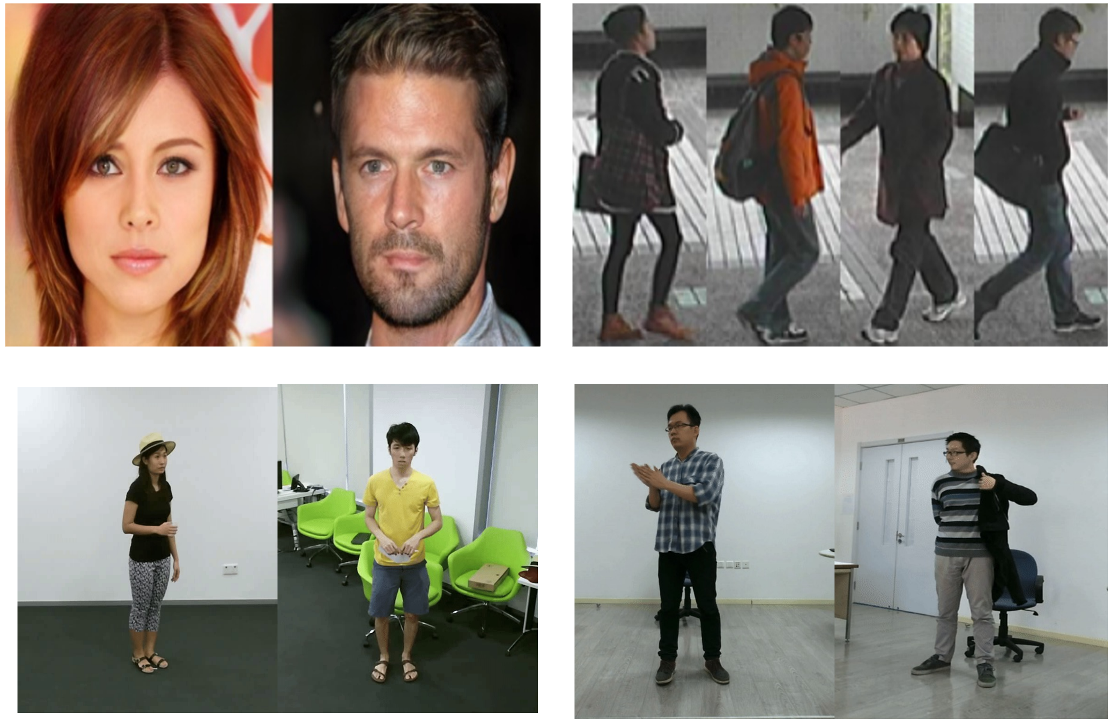
Activity-Biometrics: Person Identification from Daily Activities

Shehreen Azad, Yogesh Singh Rawat

Center for Research in Computer Vision, University of Central Florida

Motivation

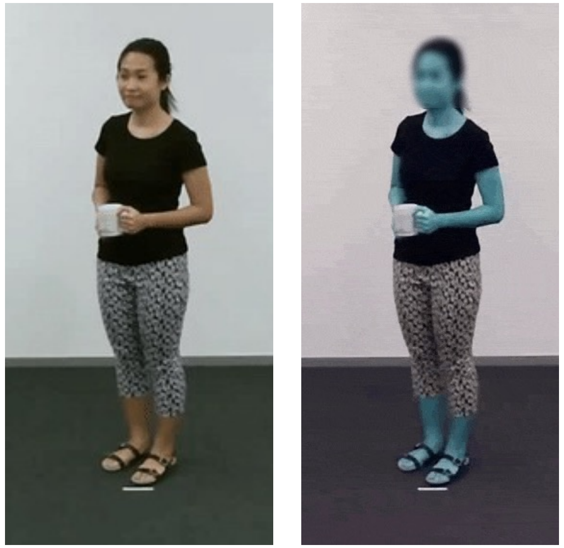
- Existing works in person identification
  - Face recognition
  - Whole body recognition
  - Gait recognition
- How to generalize?
  - Identify person from activities
  - Daily activities



Top - CelebA-HQ, CUHK-01
Bottom - NTU RGB+D 120, PKU MMD

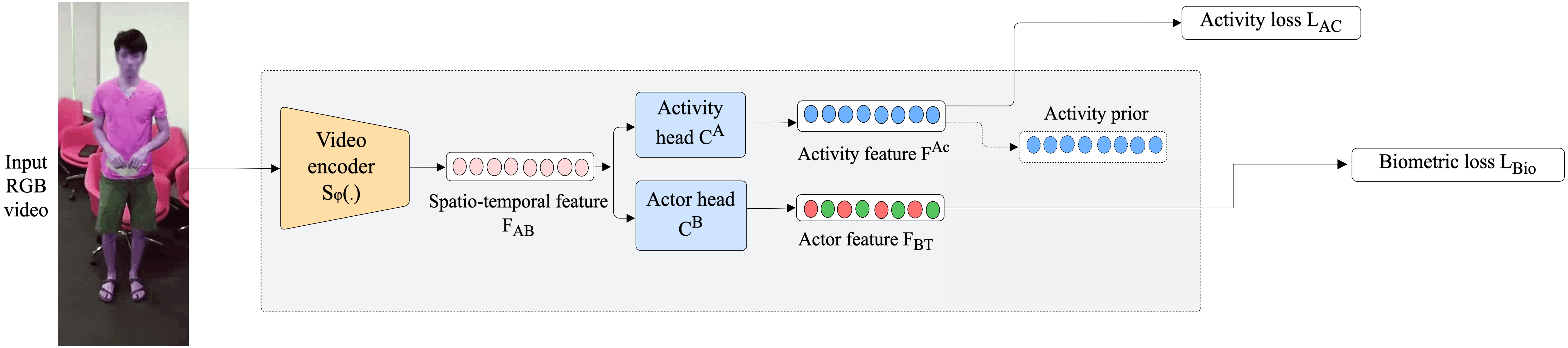
Our contribution

- Novel problem statement
- Simple joint learning framework using disentanglement of biometrics and non-biometrics features
  - Bias-less knowledge distillation
  - Bias learning using distortion
- Benefits of using activity prior for person identification



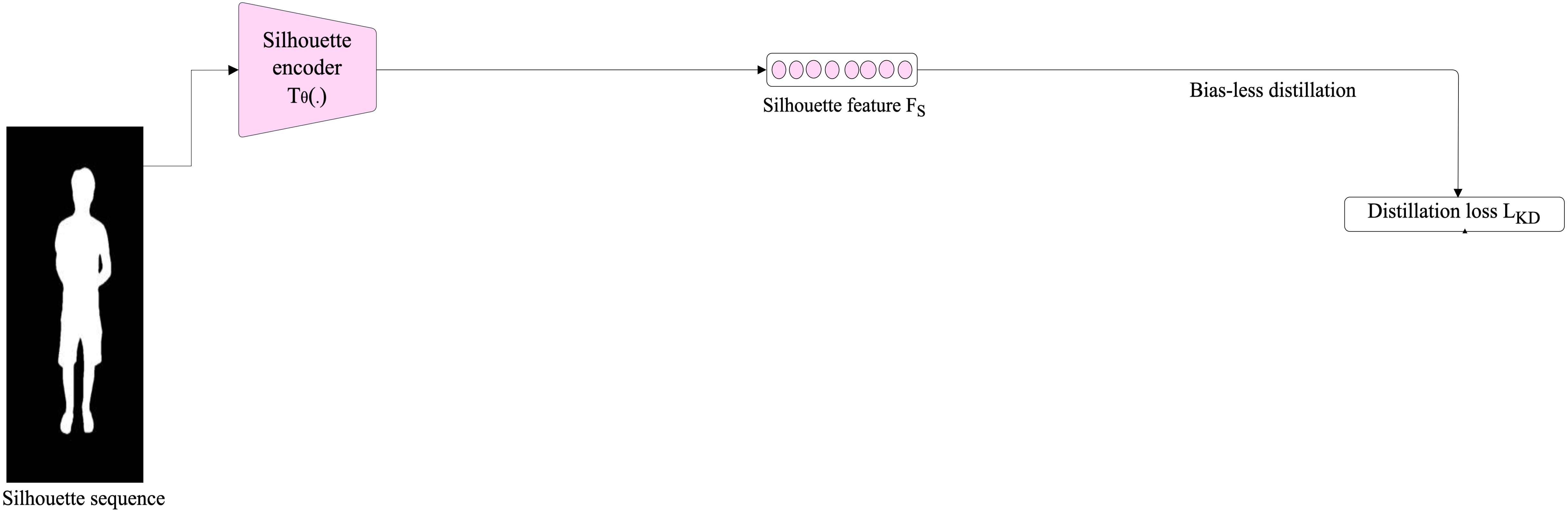
Original and its corresponding augmented sample from NTU RGB-D 120

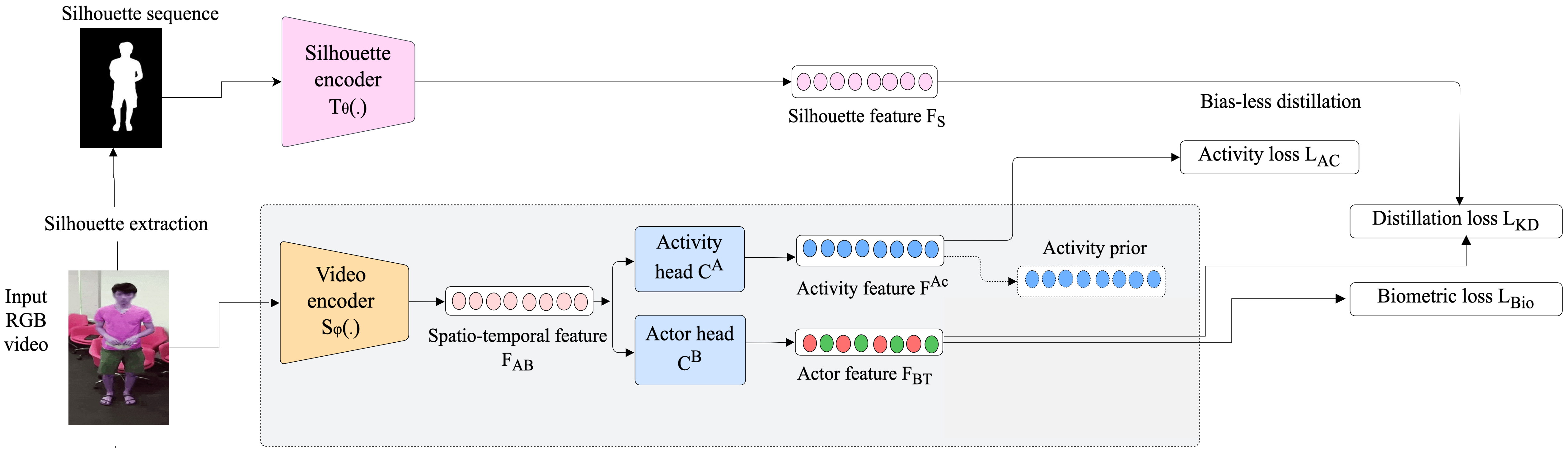
Proposed Methodology

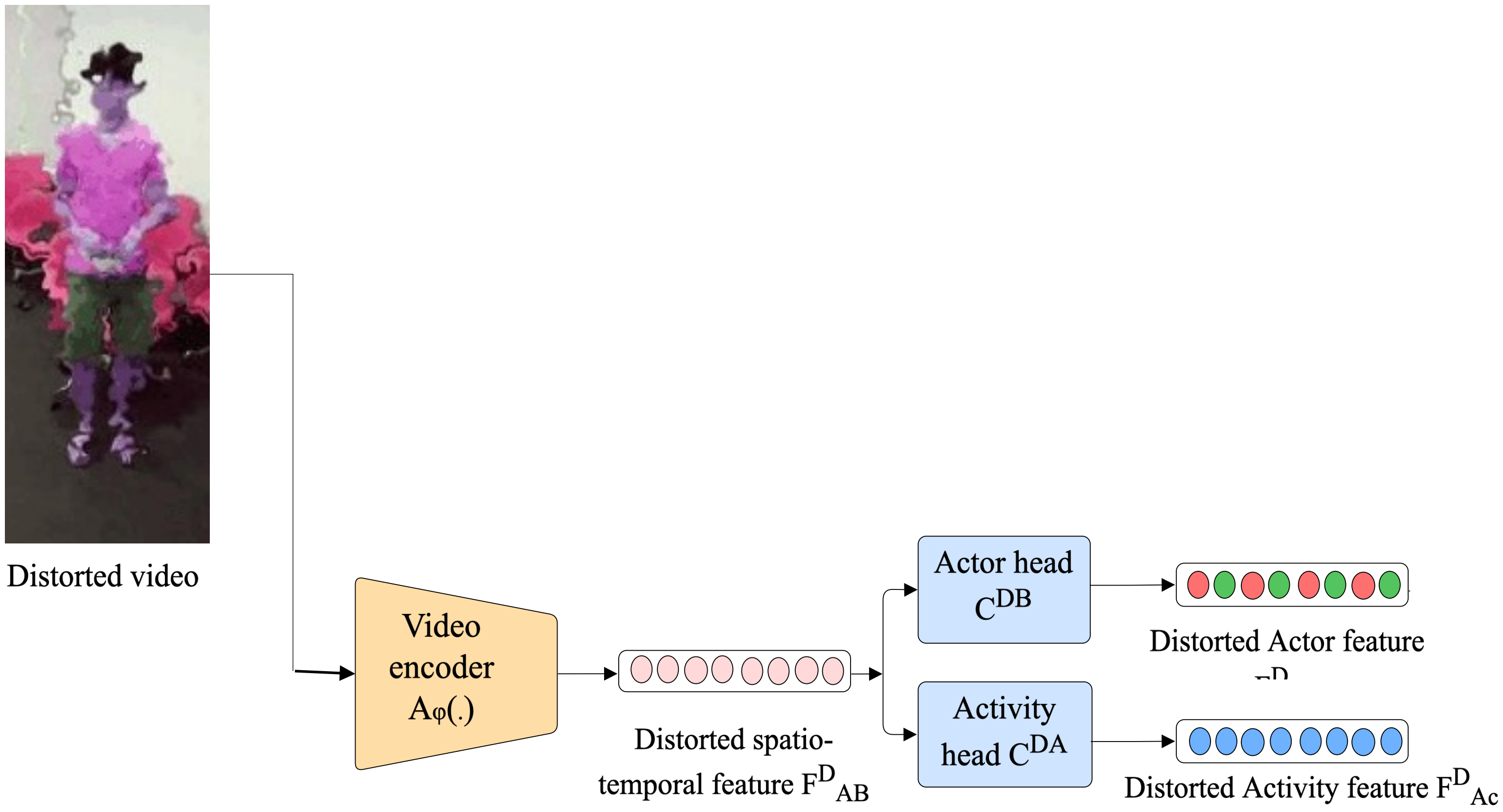


CVPR JUNE 17-21, 2024 SEATTLE, WA

UCF CENTER FOR RESEARCH IN COMPUTER VISION

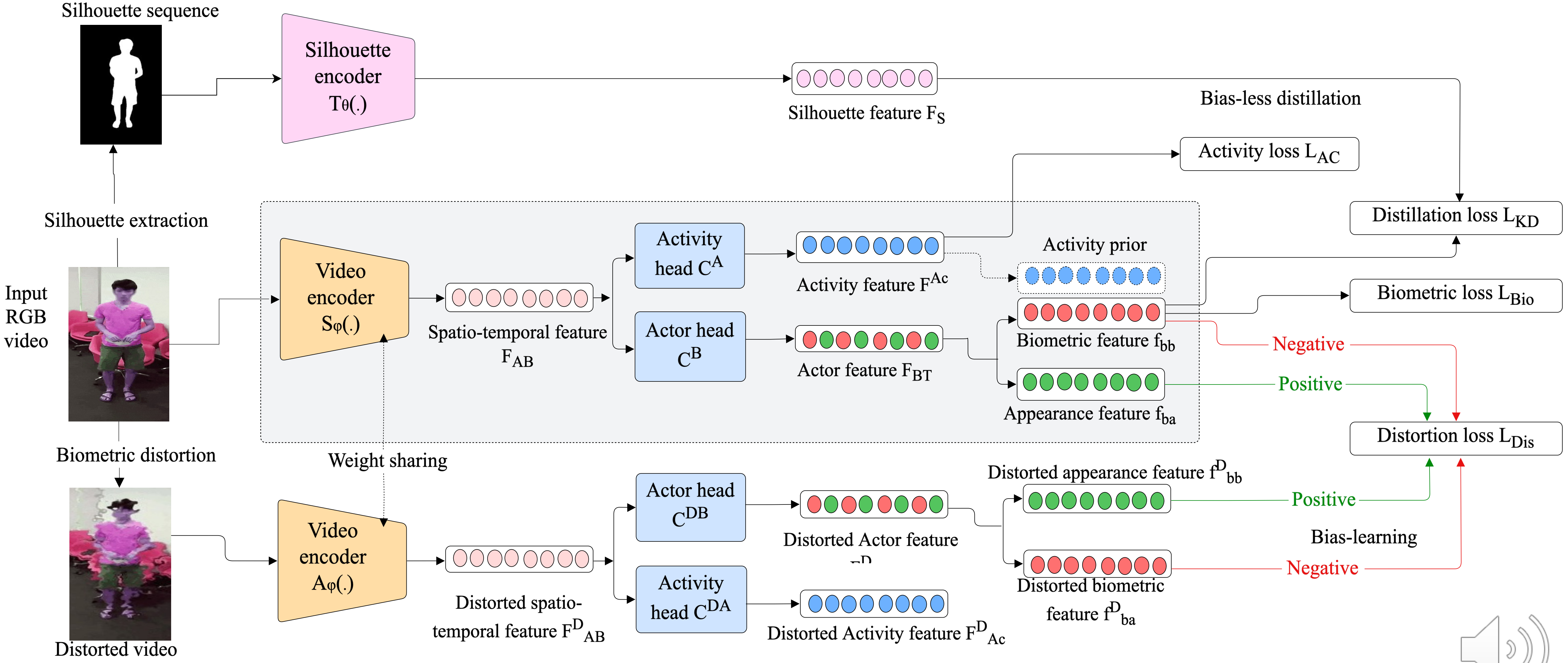



Silhouette sequence
Silhouette encoder $T_\theta(.)$
Silhouette feature F_S
Bias-less distillation
Activity loss L_{AC}
Distillation loss L_{KD}
Silhouette extraction
Input RGB video
Video encoder $S_\varphi(.)$
Spatio-temporal feature F_{AB}
Activity head C^A
Actor head C^B
Activity feature F^{Ac}
Activity prior
Actor feature F_{BT}
Biometric loss L_{Bio}


Distorted video
Video encoder $A_\varphi(.)$
Distorted spatio-temporal feature F^{D}_{AB}
Actor head C^{DB}
Activity head C^{DA}
Distorted Actor feature
Distorted Activity feature F^{D}_{Ac}

CVPR JUNE 17-21, 2024 SEATTLE, WA

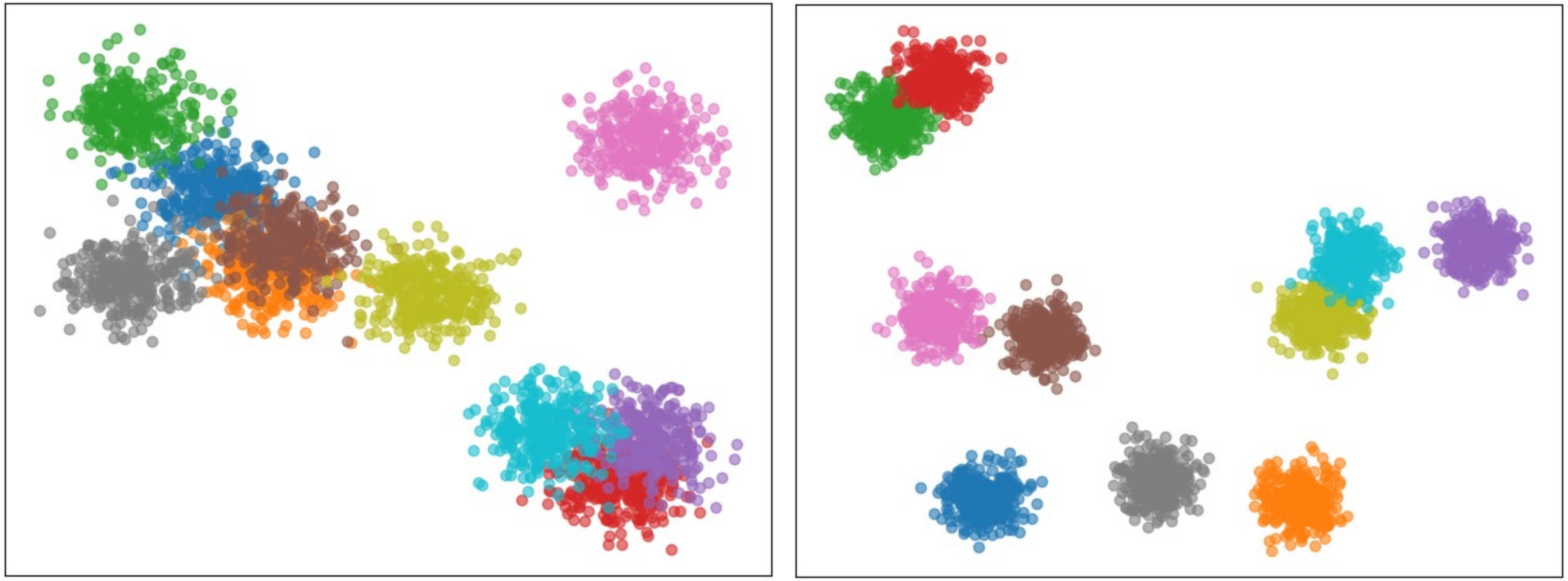
UCF CENTER FOR RESEARCH IN COMPUTER VISION



Comparison with prior work

	Methods	Venue	NTU RGB-AB		PKU MMD-AB		Charades-AB		ACC-MM1-Activities	
			Rank 1	mAP	Rank 1	mAP	Rank 1	mAP	Rank 1	mAP
Image	CAL [16]	CVPR22	73.79	28.40	81.31	49.45	43.84	25.81	69.83	42.81
	PSTR [5]	CVPR22	69.14	34.14	84.33	47.52	37.15	24.69	57.41	34.48
	SCNet [17]	ACM MM23	69.89	31.47	79.53	43.55	31.73	21.89	64.68	39.79
	AIM [43]	CVPR23	71.37	35.41	82.52	48.89	40.13	28.31	74.79	49.14
Video	TSF [23]	AAAI20	71.79	31.80	76.43	37.50	35.38	21.89	49.41	29.73
	VKD [35]	ECCV20	67.41	35.63	78.35	38.54	36.31	20.71	55.38	29.57
	BiCnet-TKS [21]	CVPR21	72.71	34.45	80.79	38.52	40.31	27.34	60.44	32.79
	STMN [12]	ICCV21	72.98	35.08	76.55	47.92	38.72	24.49	59.44	39.68
	PSTA [40]	ICCV21	67.41	34.78	77.44	50.42	42.89	28.32	71.41	50.31
	SINet [4]	CVPR22	69.41	30.68	79.58	40.80	40.31	26.90	65.39	45.41
	Video-CAL [16]	CVPR22	75.49	39.86	79.59	49.42	43.91	28.51	77.48	50.08
Baselines	GaitGL [28] †	-	61.51	28.89	65.38	33.78	18.43	6.81	39.41	18.51
	ResNet3D-50 [18]	-	64.23	26.89	69.70	32.64	32.25	17.42	44.31	22.54
	MViTv2 [26]	-	63.87	26.41	68.37	28.52	28.51	15.39	40.59	21.52
	ABNet (ours)	-	**78.76**	**40.31**	**86.83**	**57.31**	**45.84**	**31.58**	**80.43**	**52.71**

Effect of disentanglement on feature space



Colors denote actor ID

(Left – right) Feature space of backbone and AB-Net on NTU RGB-AB

Need good silhouette extractor

- Silhouettes are only needed during training
- Training with better silhouettes does give better performance

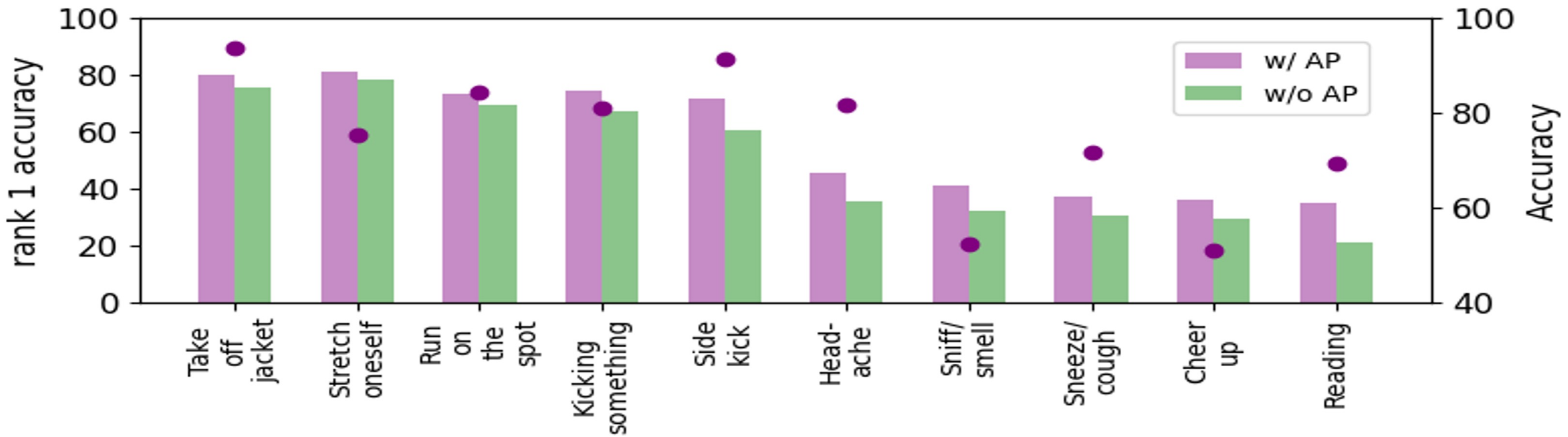
Silhouette extractor	Rank 1 accuracy
Mask2Former	85.2
Grounded-SAM	87.8 (2.6%)

Evaluation on NTU RGB-AB subset

NTU RGB-AB sample extracted with GroundedSAM

Effect of activity prior

- The knowledge of easily recognizable activities do not introduce bias



Evaluation on NTU RGB-AB across top 5 easiest hardest activities. Left axis: biometrics rank 1 accuracy. Right axis: activity recognition accuracy.

Code

Thank you for listening

For any questions, contact: Shehreen.Azad@ucf.edu