The Role of Cycle Consistency for Generating Better Human Action Videos from a Single Frame

Runze Li Bir Bhanu

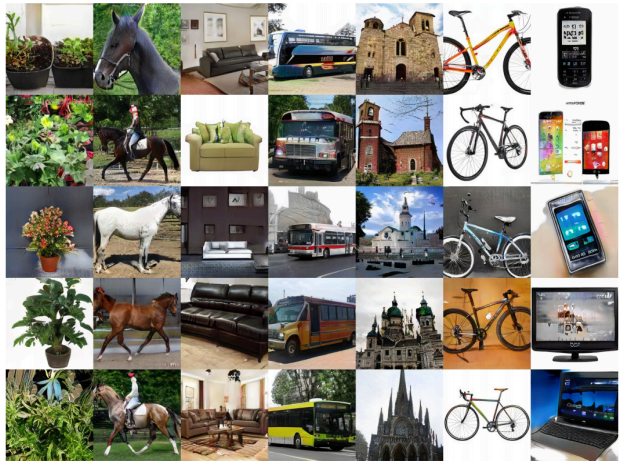
Visualization and Intelligent Systems Laboratory
University of California Riverside

Motivation

Image Generation

GAN

StyleGAN



Progressive-growing GAN

Video Generation

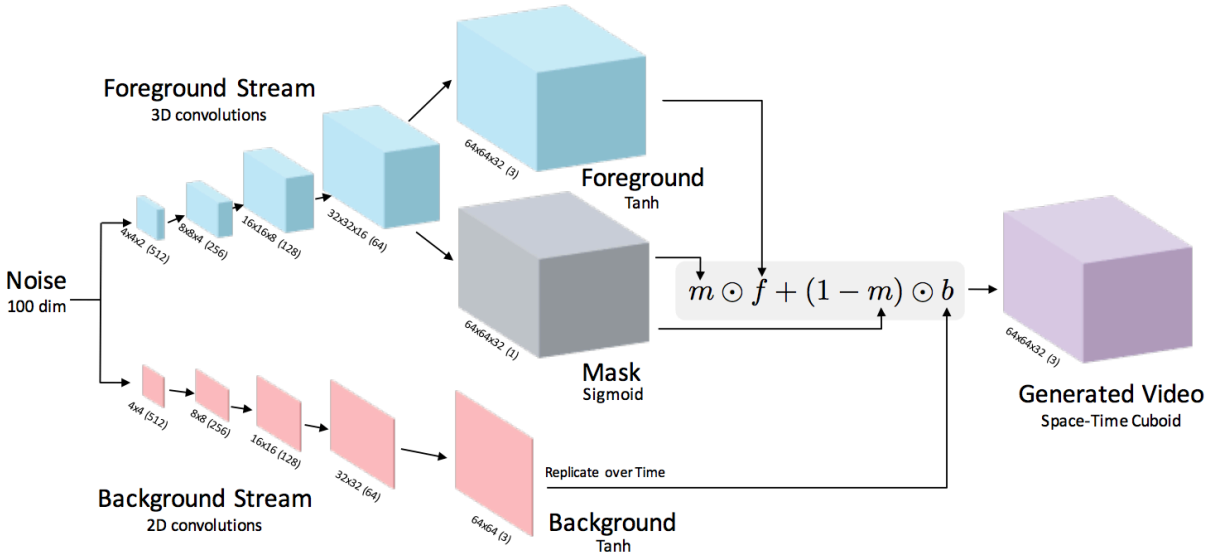
- Generate videos with dynamics
- Generate videos with human action semantics

[1] Goodfellow, Ian, et al. “Generative Adversarial Nets.” Advances in Neural Information Processing Systems 27, 2014.
[2] Karras, Tero, et al. “A Style-Based Generator Architecture for Generative Adversarial Networks.” Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019.
[3] Karras, Tero, et al. “Progressive Growing of GANs for Improved Quality, Stability, and Variation”. The Sixth International Conference on Learning Representations, 2018.
[4] Vondrick, Carl, et al. “Generating Videos with Scene Dynamics”. Advances in Neural Information Processing Systems 29, 2016.

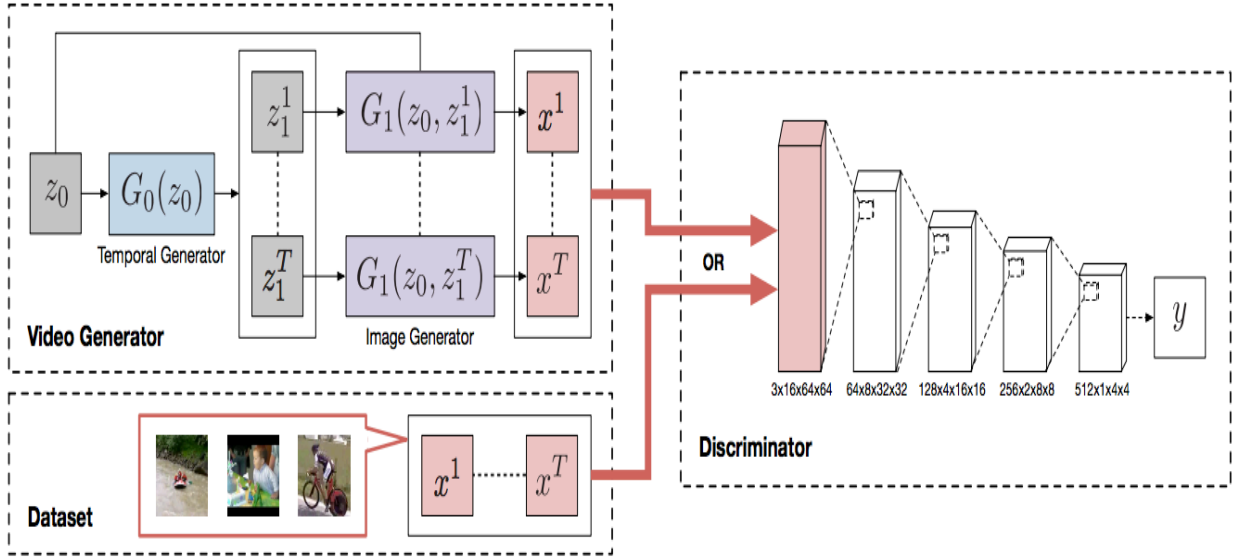
Related Work

Video Generation

- Two-stream network



Vondrick, Carl, et al.



Saito, Masaki, et al.

- Two-stage network
  - Generate motions, then generate videos
  - Generate human poses, then generate videos

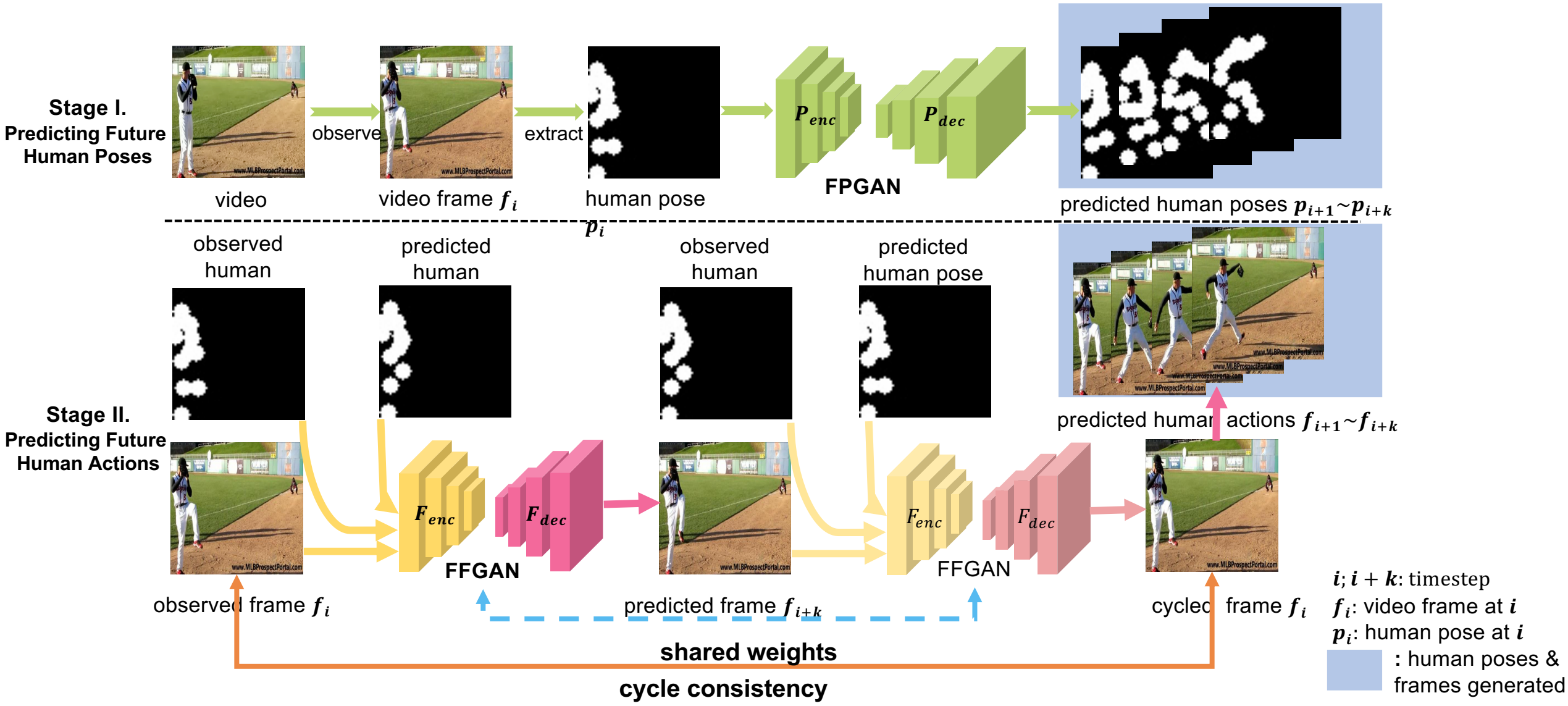
Key Ideas

- **Two-stages video prediction**
  Adapt a two-stage method to predict human poses and then to predict videos involving human action semantics in the future after observing a single frame.

- **Cycle-consistency constraints**
  Enforce the appearance and motion constraints via cycle consistency for generating human actions in the future.

- **Extensive experiments**
  Conduct thorough qualitative and quantitative evaluations on both simple and complicated human action datasets.

Framework


Stage I.
Predicting Future Human Poses
observe
extract
video
video frame f_i
human pose p_i
P_{enc}
P_{dec}
FPGAN
predicted human poses p_{i+1}~p_{i+k}
Stage II.
Predicting Future Human Actions
observed human
predicted human
observed human
predicted human pose
F_{enc}
F_{dec}
FFGAN
F_{enc}
F_{dec}
FFGAN
predicted human actions f_{i+1}~f_{i+k}
observed frame f_i
predicted frame f_{i+k}
cycled frame f_i
shared weights
cycle consistency
i; $i+k$: timestep
f_i: video frame at i
p_i: human pose at i
: human poses & frames generated

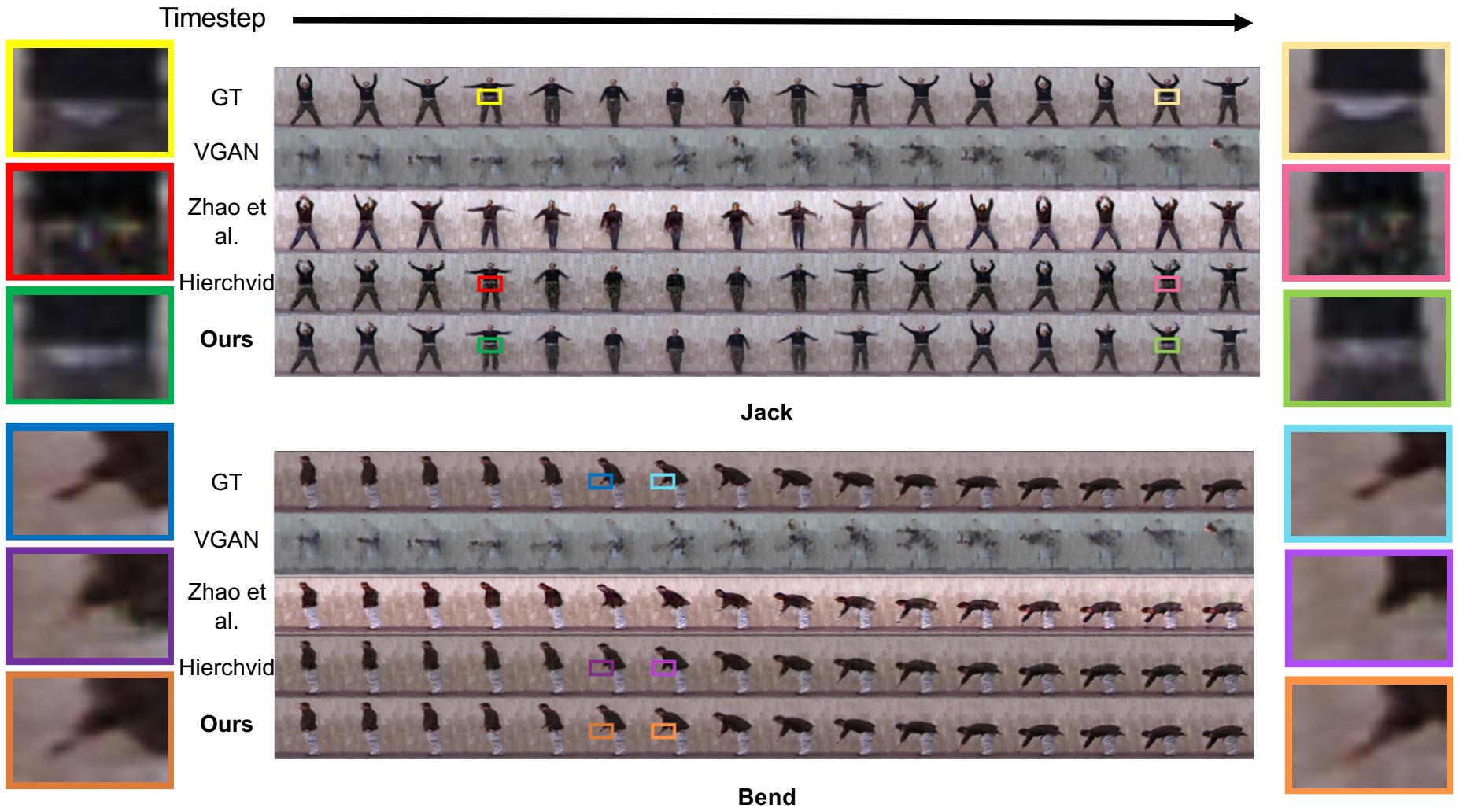
Experimental Results– Weizmann Dataset

- Quantitative Results
- Evaluation Metrics
  - SSIM: structural similarity index measure.
  - MSE: mean squared error.
  - PSNR: Peak signal-to-noise ratio.

Method	SSIM ↑	MSE ↓	PSNR ↑
VGAN (CVPR 2017)	0.1547	0.0628	12.0488
Zhao et al. (ECCV 2018)	0.787	0.005	22.198
Hierchvid (ICML 2017)	0.842	0.0026	25.7213
Ours	**0.9409**	**0.0018**	**28.6414**

Experimental Results– Weizmann Dataset

- Qualatitive Results



Experimental Results– Penn Action & UCF-101 Dataset

- Quantitative Results
- Evaluation Metrics
  - SSIM, MSE, PSNR.
  - IS: Inception score.
  - FID: Fréchet Inception Distance.

	Penn Action					UCF-101				
Method	SSIM ↑	MSE ↓	PSNR ↑	IS ↑	FID ↓	SSIM ↑	MSE ↓	PSNR ↑	IS ↑	FID ↓
Zhao et al.	-	0.023	18.25	-	-	-	-	-	-	-
Hierchvid	-	0.03	15.875	-	-	-	-	-	-	-
SCGAN-gen	-	-	-	-	-	0.73	-	-	-	-
SCGAN-full	-	-	-	-	-	0.87	-	-	5.7	-
Zhao et al.	0.568	0.063	12.372	3.012	34.39	0.73	0.065	12.247	3.646	61.729
Hierchvid	0.57	0.0456	13.5348	3.1019	37.96	0.7	0.07	12	3.7	50
Ours	**0.799**	**0.016**	**18.292**	**3.247**	**19.315**	**0.75**	**0.03627**	**13.8408**	**4.59**	**25.3384**

[1] Zhao et al.,ECCV 2018.
[2] Hierachvid, ICML 2017.
[3] SCGAN, ECCV 2018.

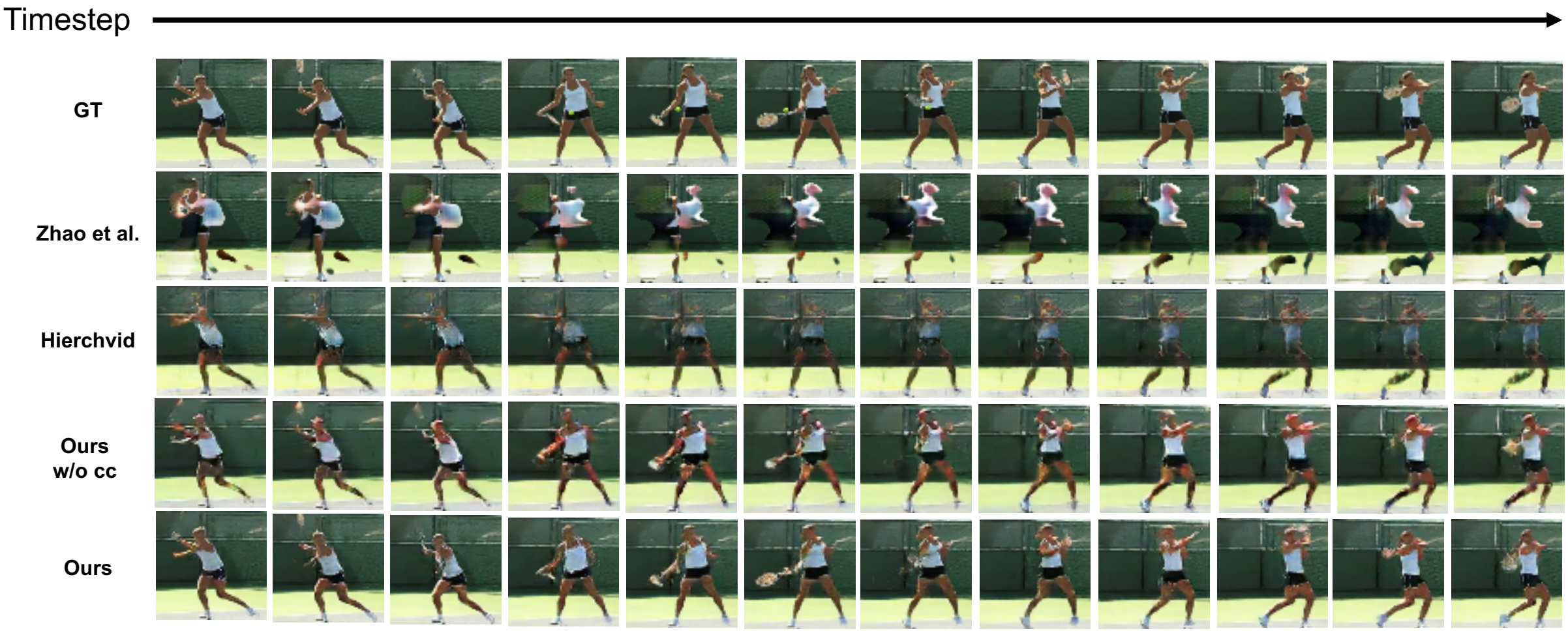
Experimental Results– Penn Action Dataset

- Qualatitive Results

Action: baseball pitch

Experimental Results– Penn Action Dataset

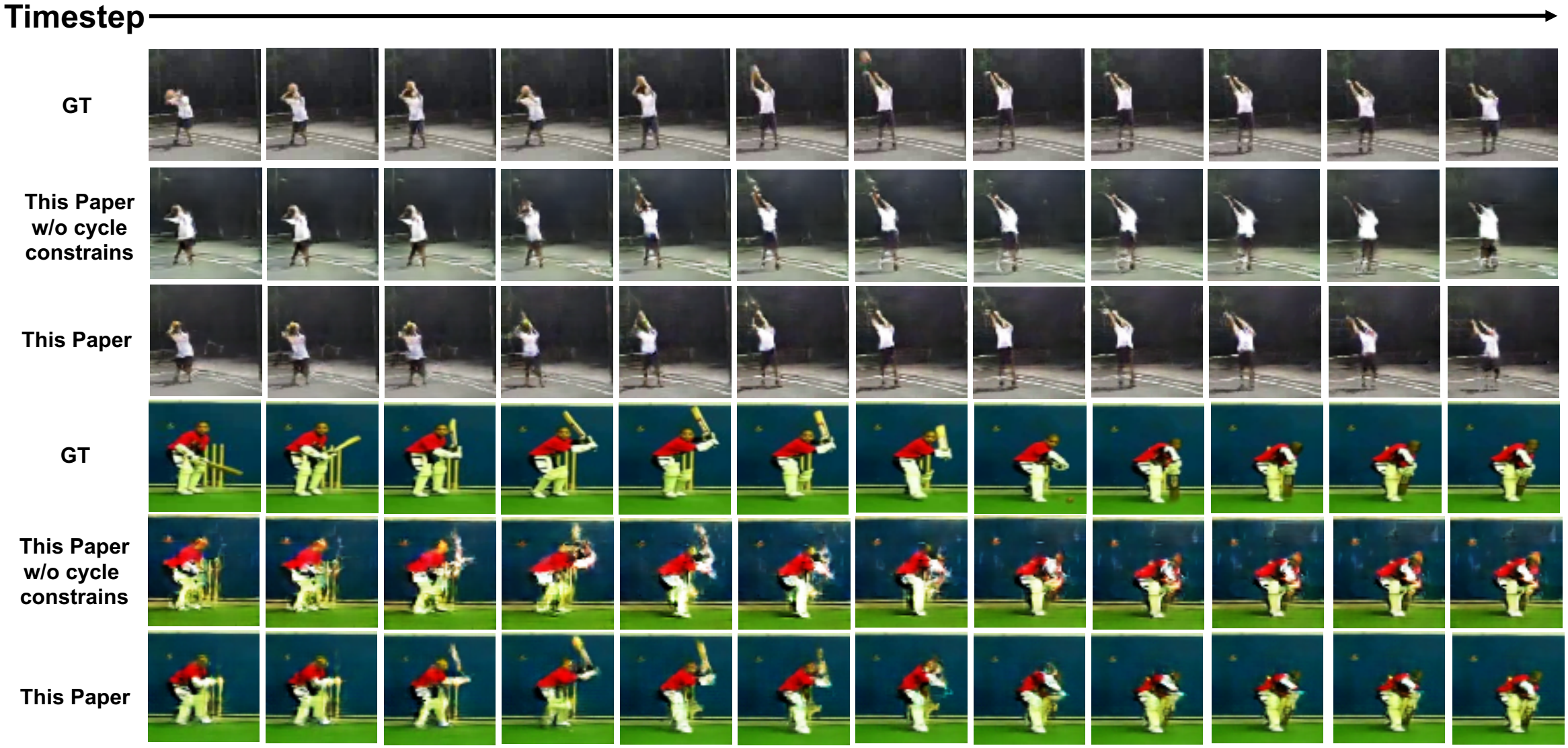
- Qualatitive Results



Action: tennis swing

Experimental Results – UCF-101 Dataset

- Qualatitive Results



Action: basketball & criket swing

Summary

- **Generating human action videos with a single shot**
- **Employing a two-stage network, predict human poses then predict human action videos**
- **Maintaining appearance and motion consistency across generated human action videos**
- **Performing quantitative and qualitative experiments**

Acknowledgements

This material is based upon work supported in part by the National Science Foundation under Grant No. 1911197.