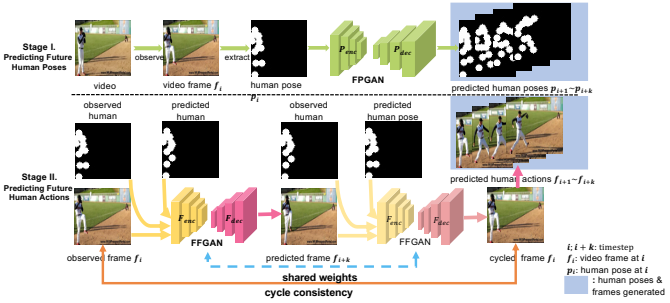
The Role of Cycle Consistency for Generating Better Human Action Videos from a Single Frame

Runze Li and Bir Bhanu
Visualization and Intelligent Systems Laboratory, UC Riverside at Riverside, CA, USA

Introduction

- **Problem:** Video prediction with human action semantics from a single shot.
- **Solution:** Use a two-stage network first by predicting human poses and then predicting human actions in the future; enforce appearance and motion constraints via cycle consistency; conduct thorough qualitative and quantitative experiments.

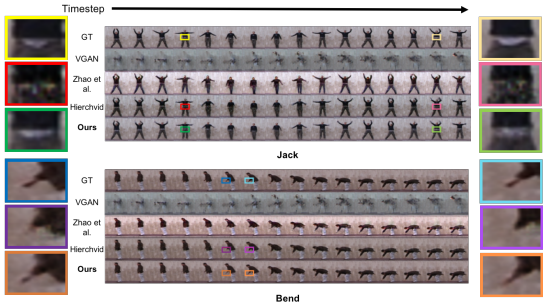
Framework



Experiments

- Weizmann Dataset

Method	SSIM ↑	MSE ↓	PSNR ↑
VGAN	0.1547	0.0628	12.0488
Zhao et al.	0.787	0.005	22.198
Hierchvid	0.842	0.0026	25.7213
Ours	**0.9409**	**0.0018**	**28.6414**

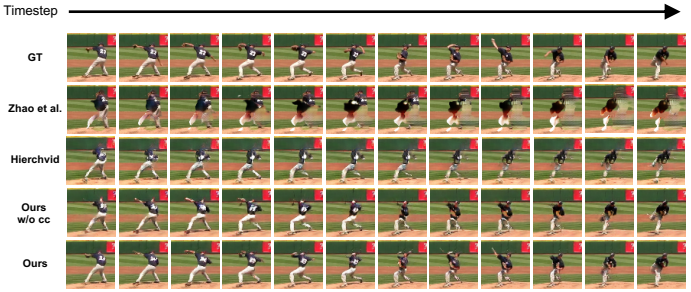


Experiments

- Penn Action Dataset

Method	SSIM ↑	MSE ↓	PSNR ↑	IS ↑	FID ↓
Zhao et al.	-	0.023	18.25	-	-
Hierchvid	-	0.03	15.875	-	-
SCGAN-gen	-	-	-	-	-
SCGAN-full	-	-	-	-	-
Zhao et al.	0.568	0.063	12.372	3.012	34.39
Hierchvid	0.57	0.0456	13.5348	3.1019	37.96
Ours	**0.799**	**0.016**	**18.292**	**3.247**	**19.315**

➢ Action: baseball swing

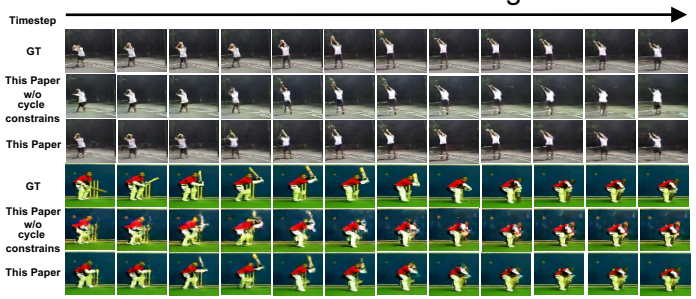


Experiments

- UCF-101 Dataset

Method	SSIM ↑	MSE ↓	PSNR ↑	IS ↑	FID ↓
Zhao et al.	-	-	-	-	-
Hierchvid	-	-	-	-	-
SCGAN-gen	0.73	-	-	-	-
SCGAN-full	0.87	-	-	5.7	-
Zhao et al.	0.73	0.065	12.247	3.646	61.729
Hierchvid	0.7	0.07	12	3.7	50
Ours	**0.75**	**0.03627**	**13.8408**	**4.59**	**25.3384**

➢ Action: basketball and criket swing



Conclusions

- Introduced cycle consistency to maintain the appearance and motion constraints for generating human actions in the future
- Both quantitative and qualitative results demonstrate the effectiveness of the proposed approach on Weizmann, Penn Action and UCF-101 Datasets

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