We Want to Dance Together

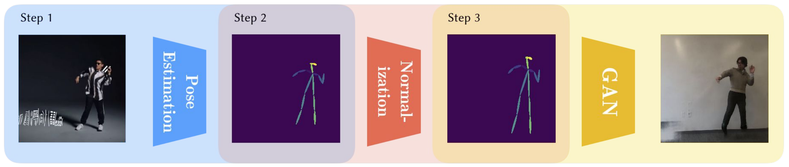
Zejiang Shen, Yueyi Sun, Ruochen Zhang, James Tompkin | zejiang_shen@brown.edu, yueyi_sun@brown.edu, ruochen_zhang@brown.edu, james_tompkin@brown.edu

BROWN Computer Science

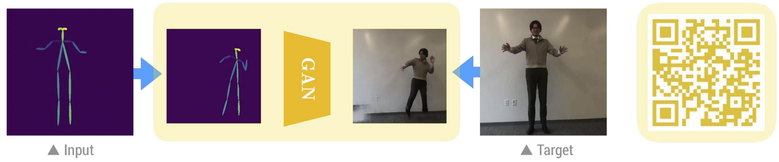
- Motivation -

- Dancing in coordination with another person is ubiquitous.
- State-of-the-art techniques are available for transferring dancing motion from a single source subject to another. However, only a few can meet this desire of transferring the high-level coordination required for dual dance.
- Based on our experiments, when there is more than one person in the video capture, directly transferring the multi-person motion can be problematic.

- Methods -



▲ The 3-step dance transferring method proposed by [1]



▲ Training the GAN with targets' pose and dance

▲ **Adding Segmentation in the GAN input** to hint the model

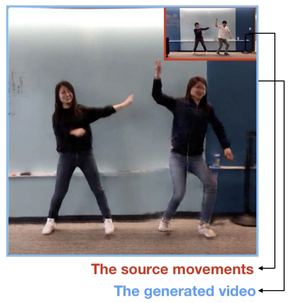
- The instance segmentation mask is added as an extra channel to the input. It's generated by MaskRCNN[2].
- Different people have different masks labels.

- Contribution -

- We propose an instance segmentation mask based improvement method based on [1] that enables a GAN to **transfer and generate multiple people's dance**.
- We verify our approach trained a model that can successfully transfer 2 students' motion simultaneously.
- We implement extensive experiments to discover what the model learns and identify some limitations.

- Results -

- We train a Pix2PixHD GAN model to generate dance from pose estimation and instance segmentation masks.
- The training data is a 1-minute 60fps video of 2 people doing random movements. The source motion is from another similar video of the other two people.
- Check our results by scanning the QR code! In the video, the top left is the source movements while the main video is generated by our GAN model.



Scan me

- References -

[1] Chan, C., Ginosar, S., Zhou, T. and Efros, A.A., 2018. Everybody dance now. arXiv preprint arXiv:1808.07371.

[2] He, K., Gkioxari, G., Dollár, P. and Girshick, R., 2017. Mask r-cnn. In Proceedings of the IEEE international conference on computer vision (pp. 2961-2969).

- Experiments -

※ Experiment I: Switch the Mask Labels ※

- Hypothetically, Switching the labels of the segmentation mask will change the generated person in the corresponding region. And the experiment verifies our idea: the generated subject is switched.

Scan me

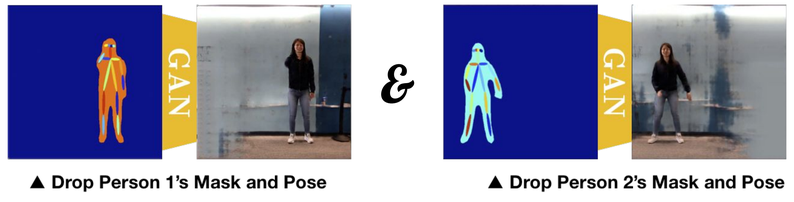
※ Experiment II: Drop one Person's Mask ※

▲ Drop Person 1's Mask ▲ Drop Person 2's Mask

- Ideally, dropping someone's mask should lead to a total disappearance of that one in the generated video. However, the subject still appears when her figure's mask is dropped, though the figure of the dropped the subject becomes blurry.

Scan me

※ Experiment III: Drop one Person's Mask & Pose ※



▲ Drop Person 1's Mask and Pose ▲ Drop Person 2's Mask and Pose

- After dropping one's mask and pose estimation, she totally disappears from the generated image. However, the output quality is limited, and it seems the model cannot discriminate persons herein.

Scan me