

TL;DR SteadyDancer reframes human image animation around an image-to-video pipeline so generation starts from a stable first frame instead of loosely binding identity in a reference-to-video setup. The project ships open inference code, 14B weights, and a dedicated X-Dance benchmark aimed at harder source and driving mismatches.

---

## What is SteadyDancer?

SteadyDancer is the official implementation of the paper *SteadyDancer: Harmonized and Coherent Human Image Animation with First-Frame Preservation*. It targets a common production problem: when the source image and driving video are not perfectly aligned, identity and structure drift over time.

The core design choice is to prefer an image-to-video (I2V) generation path over a reference-to-video (R2V) path. In the authors' framing, this helps preserve first-frame appearance and reduces abrupt transitions when source identity and driving motion are mismatched.

Links:

- Repo: <https://github.com/MCG-NJU/SteadyDancer>
  - Paper: <https://arxiv.org/abs/2511.19320>
  - Project page: <https://mcg-nju.github.io/steadydancer-web>
  - Model weights (HF): <https://huggingface.co/MCG-NJU/SteadyDancer-14B>
  - X-Dance benchmark: <https://huggingface.co/datasets/MCG-NJU/X-Dance>
- 

## Why this matters for teams shipping AI video

Most identity-preserving pipelines look good on clean demos but break on realistic inputs where:

- The source portrait and driving performer differ in framing, body layout, or garment structure.
- Motion starts at a different temporal state than the source frame.
- Real footage includes blur, occlusion, and non-trivial camera changes.

SteadyDancer is explicitly built around those failure cases. The repo positions this as a practical response to spatio-temporal misalignment, not just a benchmark-only increment.

---

## Release snapshot (exact dates)

Based on the public README timeline:

- 2025-11-24: paper published on arXiv, inference code and 14B weights released.
- 2025-11-24: X-Dance benchmark released.
- 2025-11-27: multi-GPU inference support added (FSDP + xDiT USP).
- 2025-12-04: GGUF format weights released.
- 2025-12-08: ModelScope weights released.
- 2025-12-11 and 2025-12-12: additional ComfyUI workflows added, including multi-person workflow support.

For planning purposes, this means SteadyDancer is not only a paper release: it also has an actively updated inference and tooling path.

---

## Inference pipeline at a glance

The repo workflow can be thought of in two stages:

1. Preprocess the driving signal into aligned positive and negative pose conditions.
2. Run `generate_dancer.py` with reference image, prompt, and both pose condition folders.

Minimal command sketch:

```
# 1) pose extraction + alignment
python preprocess/pose_align.py --imgfn_refer <ref.png> --vidfn <drive.mp4> ...
python preprocess/pose_align_withdiffaug.py --imgfn_refer <ref.png> --vidfn <drive.mp4> ...

# 2) generation (single GPU)
CUDA_VISIBLE_DEVICES=0 python generate_dancer.py \
--task i2v-14B \
--size 1024*576 \
--ckpt_dir ./SteadyDancer-14B \
```

```
--prompt "<prompt>" \  
--image <ref.png> \  
--cond_pos_folder <positive_pose_dir> \  
--cond_neg_folder <negative_pose_dir>
```

The authors also provide multi-GPU launch examples, but they note single-GPU runs are better for reproducibility when exact output matching matters.

---

## Practical adoption notes

- Hardware and dependencies are non-trivial: the stack includes PyTorch, FlashAttention, xformers, and OpenMMLab pose tooling (mmdcv, mmdet, mmpose). Plan environment setup time.
  - Start with one controlled use case (for example, short dance snippets with fixed framing) before broad rollout.
  - Treat preprocessing as part of your production system, not an optional step. Pose extraction and alignment quality strongly affects final stability.
  - If your team already uses ComfyUI-based pipelines, the newer workflow JSONs in the repo can speed up first integration.
- 

## Positioning vs alternatives

SteadyDancer is strongest when your source identity and driving motion are not perfectly matched and you need better first-frame continuity across the clip. If your current workflow already has tightly aligned source-driving pairs, gains may be smaller and setup overhead may dominate.

The right evaluation is task-specific: run side-by-side tests on your own footage and score identity retention, temporal smoothness, and failure rate across hard edge cases (occlusions, fast turns, partial body shots).

---

## References

- GitHub repository: <https://github.com/MCG-NJU/SteadyDancer>
- arXiv paper: <https://arxiv.org/abs/2511.19320>
- Project page: <https://mcg-nju.github.io/steadydancer-web>
- HF model: <https://huggingface.co/MCG-NJU/SteadyDancer-14B>

- HF dataset (X-Dance): <https://huggingface.co/datasets/MCG-NJU/X-Dance>

Notes: This draft reflects the public SteadyDancer README and linked artifacts available as of 2026-02-13.