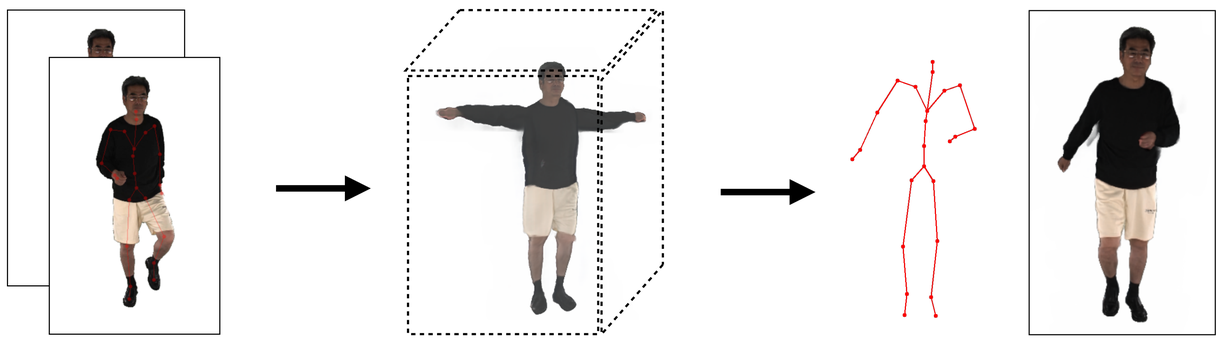
Generalized Dynamic Human Neural Fields from Few Views

Jakub Zadrozny, Hakan Bilen
The University of Edinburgh

Dynamic Free-Viewpoint Human Rendering

Task: given a few observed frames of an actor performing a motion, synthesize their model, which can be animated to a given target pose and rendered from an arbitrary viewpoint.



Input: a few monocular views with pose parameters

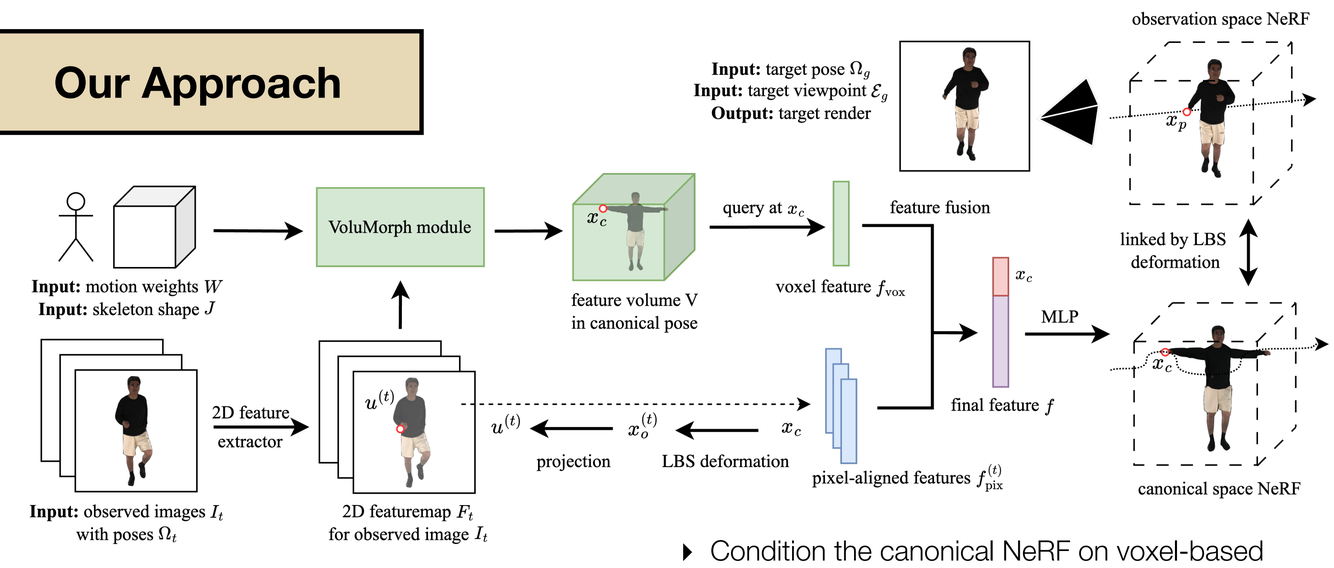
Estimate: subject NeRF model in canonical pose

Input: target pose parameters

Output: render in target pose

- **Subject-specific approaches**
  - − requires test-time optimization and extensive observations (typically a ~20s video)
  - − no prior, cannot in-paint unobserved regions
  - + usually most accurate (especially on soft elements like clothing)
  - + does not require a large training set
- **Generalized approaches (ours)**
  - + only feed-forward passes during inference
  - + requires less observed views (1-3 for our model)
  - + learns a prior, in-paints unobserved details
  - − typically less accurate compared to subject-specific (especially on soft elements)
  - − requires a multi-subject training dataset

Our Approach



- Produce the target render by volumetric rendering of a NeRF in target pose
- Always query the canonical NeRF instead
- Find corresponding points by inverse linear blend skinning (LBS) deformation

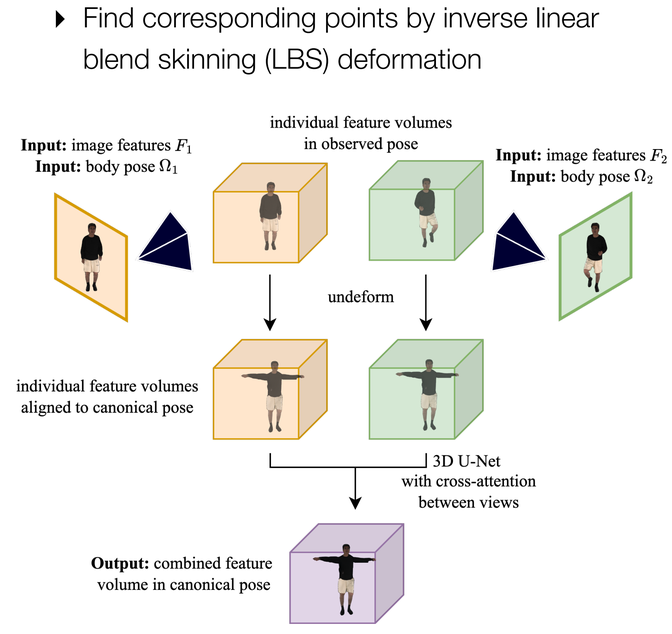


Diagram of our VoluMorph module.

- Condition the canonical NeRF on voxel-based and pixel-aligned features
- Voxel features are coarse, but they can resolve occlusions, inject prior and compensate for pose parameter errors
- Pixel-aligned features offer complementary fine details at high-resolution but are not always relevant (e.g. due to occlusions)
- Voxel features extracted using our *VoluMorph* module, which combines sets of posed images into a complete model in canonical pose
- Motion weights: initialized heuristically from skeleton shape, we then *learn* a correction directly from observed views

Key ideas:

- combine all available observed views in 3D using dense processing
- learn a correction for the LBS motion weights
- improve robustness to inaccuracies in body parameters

Results and Conclusions

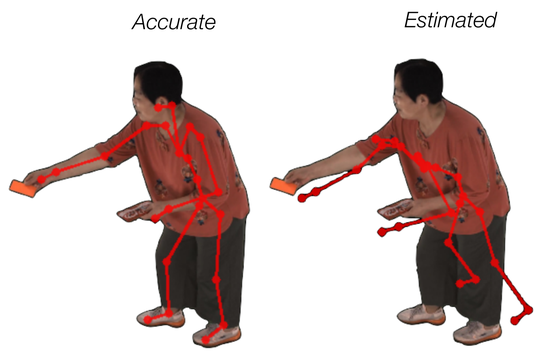
- Evaluate on hold-out, **unseen identities**
- Evaluate **novel pose** rendering by splitting motion sequences in half — select observed frames from the 1st half, evaluate on 2nd half
- Note that PSNR favours oversmoothed results over even slight misalignments
- LPIPS* $= \text{LPIPS} \times 10^3$

- Our method is competitive given 1 view, with a significantly better perceptual (LPIPS) score
- We outperform the baseline on all metrics given 2 (HuMMan) or 3 (DNA-Rendering) views
- Our rendering quality saturates given 3 views
- Our method shows better robustness to estimated body parameters
- Additional views currently do not help when using estimated body parameters

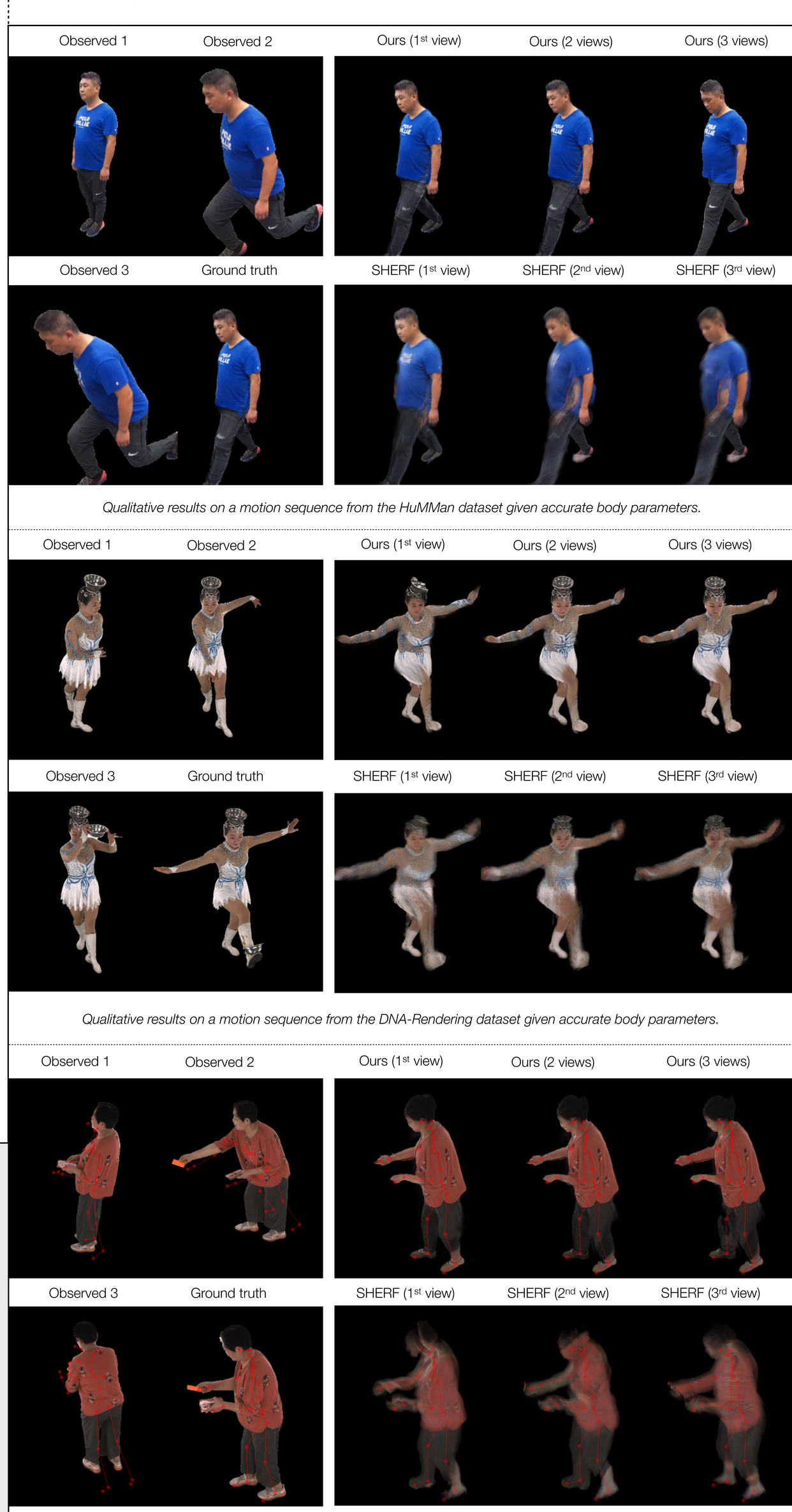
Method	HuMMan		DNA-Rendering			
	accurate body params		*accurate body params*		*estimated body params*	
	PSNR↑	LPIPS*↓	PSNR↑	LPIPS*↓	PSNR↑	LPIPS*↓
Ours (1 view)	26.62	34.24	27.83	40.27	**27.43**	**44.12**
Ours (2 views)	27.31	30.90	28.31	37.80	26.87	48.26
Ours (3 views)	**27.56**	29.59	**28.59**	**36.68**	27.04	47.22
Ours (4 views)	**27.56**	**29.53**	28.58	36.82	27.06	47.10
SHERF	26.87	45.03	28.49	48.22	26.93	61.97

Body Shape and Pose

- Existing methods rely heavily on linear blend skinning (LBS) deformations of the SMPL mesh to link points across different body poses
- The required SMPL parameters are typically accurately estimated from synchronized multi-view camera setups → impractical
- We report results with SMPL shape and pose parameters estimated directly from the observed views using an off-the-shelf HybrIK model



Frame from the DNA-Rendering dataset with annotated skeleton using accurate (left) and estimated (right) body shape and pose parameters.



Qualitative results on a motion sequence from the HuMMan dataset given accurate body parameters.

Qualitative results on a motion sequence from the DNA-Rendering dataset given accurate body parameters.

Qualitative results on a motion sequence from the DNA-Rendering dataset given estimated body parameters.

Dataset	Motion seqs. (train — test)	Actors	Cameras
HuMMan	339 (317 — 22)	153	10
DNA-Rendering	436 (372 — 64)	136	48

Dataset characteristics for HuMMan and DNA-Rendering.