Pose Splatter: A 3D Gaussian Splatting Model for Quantifying Animal Pose and Appearance

Jack Goffinet*, Youngjo min*, Carlo Tomasi, David Carlson

Duke University

(* Equal Contribution)

Duke

Paper

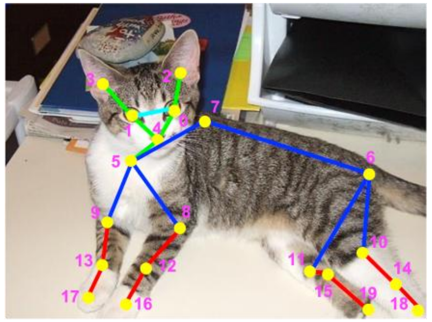
Code

Motivation

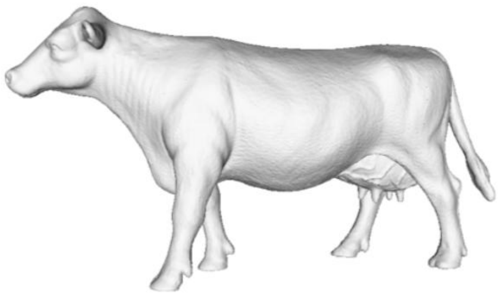
Why 3D Animal Pose Matters

- Understanding behavior is key to studying neural and genetic processes.
- 3D pose reveals walking, balance, and interactions
- **Keypoint methods:** need manual labels, too sparse for full shape or texture.
- **Mesh methods:** require per-frame optimization and species templates.

➤ **For a large-scale analysis, we need a method that is annotation-free, template-free, and fast.**



Keypoints [1]



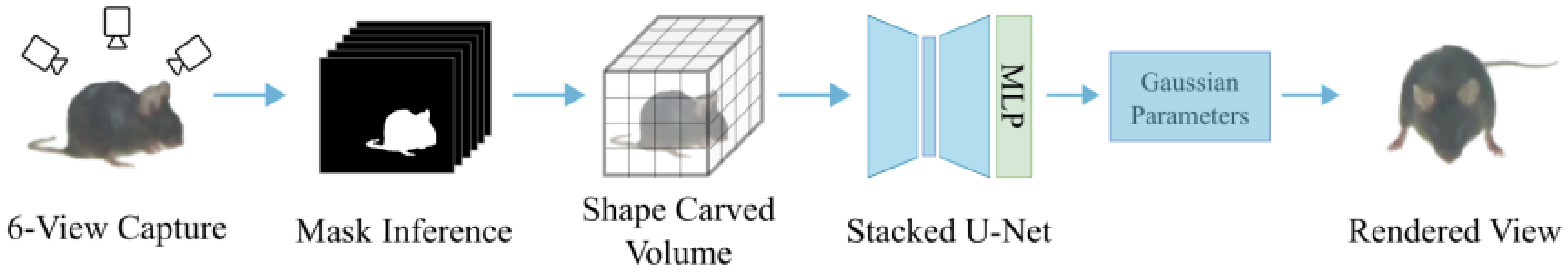
Mesh [2]

Duke

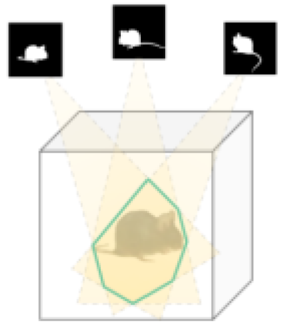
Method

- **Pose Splatter:** A Feed-Forward 3D Gaussian Splatting Framework
- **Goal:** Model full 3D pose & appearance of animals without labels or templates.
- **Pipeline:**

1. Multi-view images + SAM2 masks → shape-carved voxel volume.
2. Stacked 3D U-Net → refines volume into feature map.
3. MLP → Gaussian parameters (position, covariance, color, opacity).
4. Render via 3DGS with L_1 + IoU losses.



Pipeline



Shape Carving

Duke

Method

- **Advantages:**

1. Feed-forward inference (no per-frame optimization).
2. Annotation-free and template-free.
3. Lightweight (≈ 2.5 GB VRAM, 30 ms per frame).

Method

- **Rotation-Invariant Visual Embedding**
- **Goal:** A compact, rotation-invariant descriptor of pose & appearance
- **Pipeline:**

1. **Rendering:** 32 virtual views on a sphere around the animal
2. **Encoding:** each 224×224 render → 512-D feature via ResNet-18.
3. **Spherical Harmonics:**
   - Treat $f(\theta, \phi)$ (feature) as a function on the sphere.
   - Expand in basis Y_{lm} ($L = 3$) and keep $\|\hat{f}_{lm}\|^2$ → rotation invariance.
4. **Adversarial PCA:**
   - Further remove azimuth bias (light / view differences).
   - Produce final 50-D pose embedding.

Duke

Experiments

- **Datasets**

🐭 **Mouse:** 6 synchronized 30-min videos (1536×2048 @ 30 FPS) of a freely-moving mouse in a 28 cm arena (324,000 frames).

🐦 **Finch:** 20-min 6-view recording of a freely moving zebra finch (216,000 frames).

🐀 **Rat7M** [3]**:** 6 camera angles with partial occlusions (tail, feet) and uneven lighting.

Mouse

Finch

Rat

Duke

Experiments

- **Quantitative Results**

Method		Mouse				Finch			
		IoU↑	L1↓	PSNR↑	SSIM↑	IoU↑	L1↓	PSNR↑	SSIM↑
Per-Scene Optimization	3DGS	0.502	0.742	25.9	0.969	0.513	0.689	26.4	0.975
	FSGS	0.462	0.923	25.3	0.975	0.454	0.925	25.6	0.981
	GO	<u>0.732</u>	**0.628**	<u>28.8</u>	<u>0.977</u>	<u>0.819</u>	<u>0.382</u>	<u>34.1</u>	<u>0.990</u>
Feed-Forward	PixelSplat	0.424	0.921	25.2	0.968	0.428	0.858	26.2	0.971
	MVSplat	0.417	0.887	25.5	0.966	0.461	0.893	25.9	0.970
	Ours	**0.760**	<u>0.632</u>	**29.0**	**0.982**	**0.848**	**0.345**	**34.5**	**0.992**

Comparison with sparse-view 3DGS [4-8]

Method	Mouse (4 cam)				Finch (4 cam)			
	IoU↑	L1↓	PSNR↑	SSIM↑	IoU↑	L1↓	PSNR↑	SSIM↑
3DGS	0.447	0.786	25.8	0.967	0.459	0.754	26.1	0.973
FSGS	0.414	0.982	24.9	0.974	0.423	0.891	25.4	0.980
GO	<u>0.706</u>	**0.745**	**28.5**	<u>0.981</u>	<u>0.725</u>	**0.657**	**30.4**	**0.985**
Ours	**0.721**	<u>0.753</u>	<u>28.2</u>	**0.982**	**0.731**	<u>0.685</u>	<u>29.0</u>	<u>0.981</u>

Comparison with per-scene optimization 3DGS [4-6]

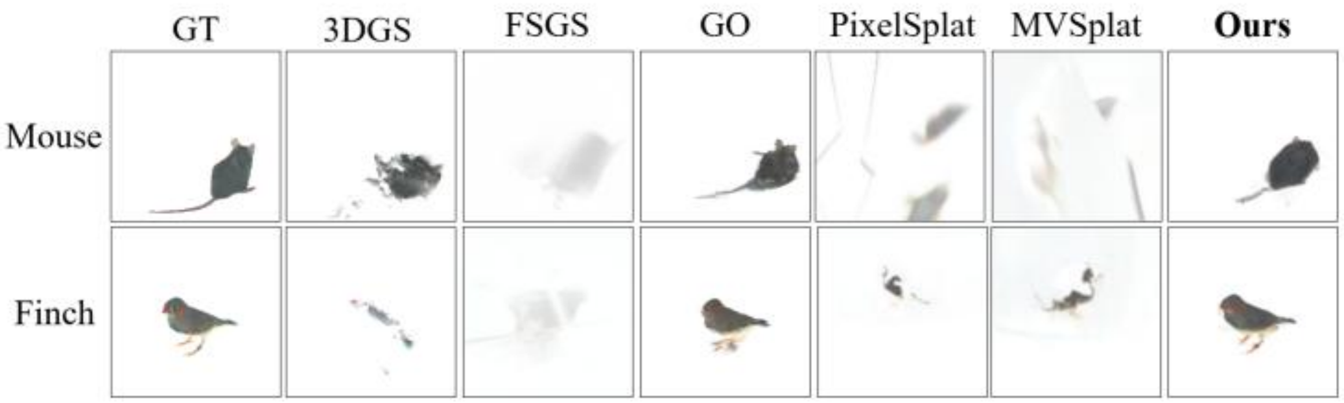
	IoU↑	L1↓	PSNR↑	SSIM↑
Mouse → Rat	**0.658**	**1.014**	**25.1**	**0.972**
Finch → Rat	0.545	1.200	24.0	**0.972**
Mouse → Finch	0.719	0.625	31.1	0.988
Finch → Mouse	0.736	0.609	29.3	0.982

5-camera cross-species generalization

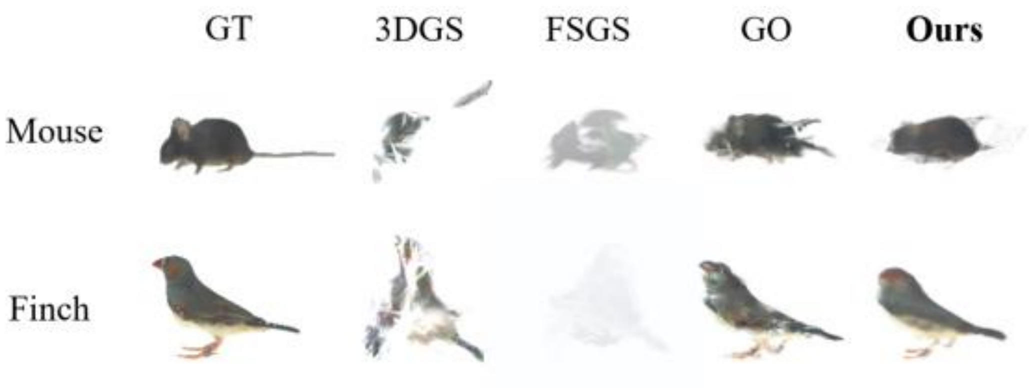
Duke

Experiments

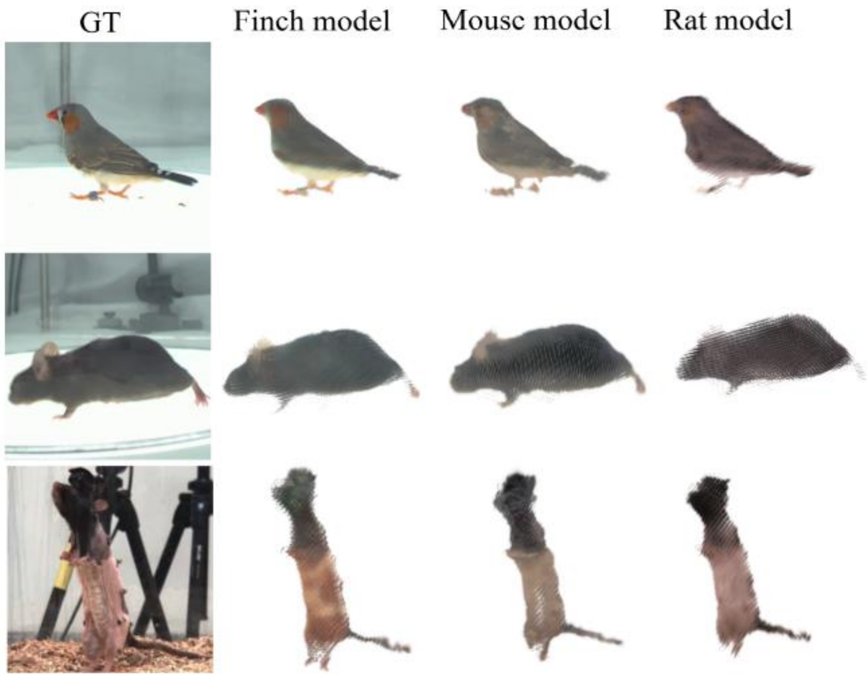
- **Qualitative Results**



Comparison with sparse-view 3DGS [4-8]



Comparison with per-scene optimization 3DGS [4-6]

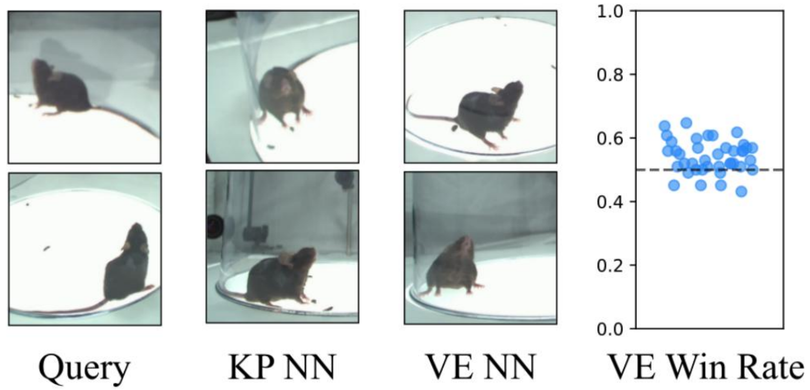


Cross-species renderings

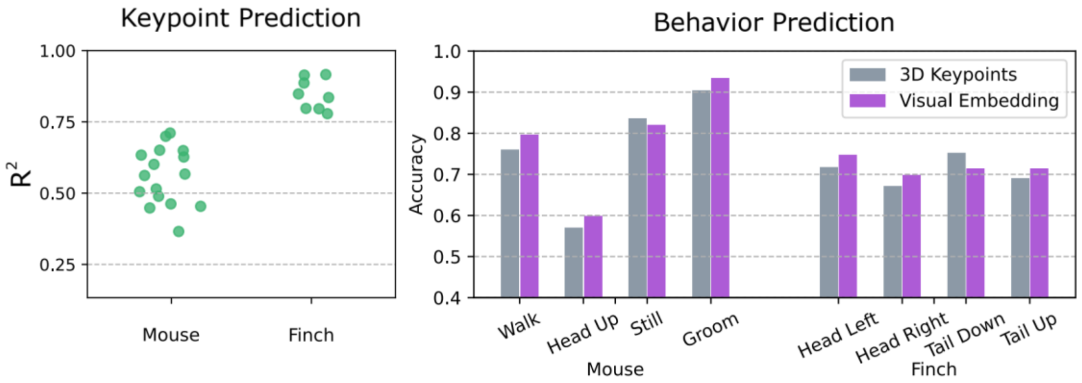
Duke

Experiments

- **Visual Embedding Results**



Nearest-neighbor preference study (vs. 3D Keypoints)



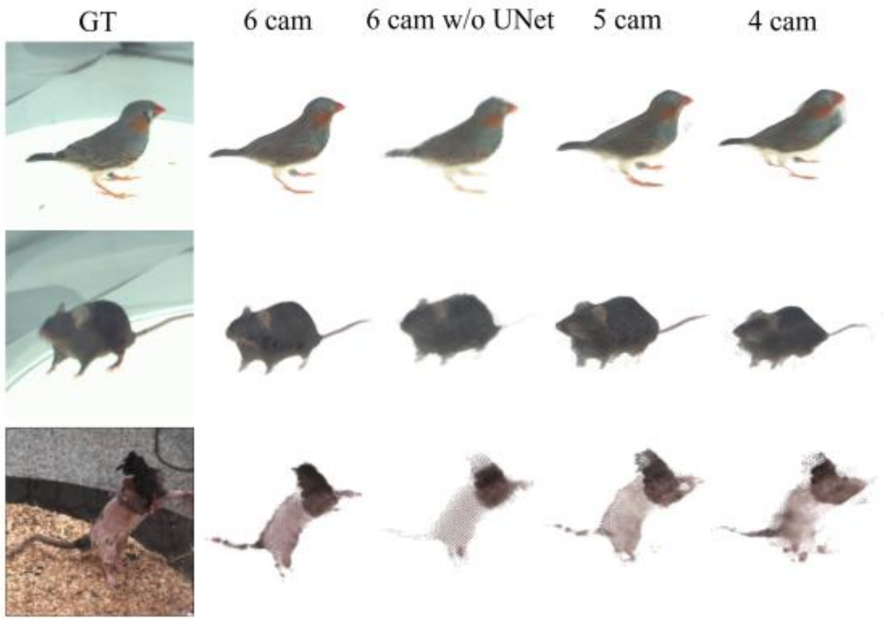
Behavior Prediction (vs. 3D keypoints)

Duke

Experiments

- **Ablation Study**

Method	Mouse				Finch				Rat			
	IoU↑	L1↓	PSNR↑	SSIM↑	IoU↑	L1↓	PSNR↑	SSIM↑	IoU↑	L1↓	PSNR↑	SSIM↑
6 cam	**0.868**	**0.317**	**33.5**	**0.989**	**0.913**	**0.231**	**36.4**	**0.991**	**0.797**	**0.658**	**26.9**	**0.975**
6 cam$^-$	0.825	0.380	32.2	0.987	0.876	0.308	34.5	0.990	0.664	0.849	25.5	0.971
5 cam	**0.760**	**0.632**	**29.0**	0.982	**0.848**	**0.345**	**34.5**	**0.992**	**0.794**	**0.628**	**27.6**	**0.981**
5 cam$^-$	0.748	0.663	28.8	**0.983**	0.838	0.421	33.7	0.991	0.688	1.16	24.6	0.970
4 cam	**0.721**	**0.753**	28.2	**0.982**	**0.731**	**0.685**	**29.0**	**0.981**	**0.651**	**1.16**	**24.4**	**0.967**
4 cam$^-$	0.701	0.737	**28.4**	0.982	0.675	0.874	28.0	0.979	0.579	2.01	23.5	0.955



Duke

Discussion & Conclusion

- **Key Achievements**

  1️⃣ **Full 3D pose & appearance:** reconstruction without any manual annotation or species-specific templates.
  2️⃣ **Feed-forward** model — no per-frame optimization; fast (≈ 30 ms/frame).
  3️⃣ **Lightweight** (≈ 2.5 GB VRAM) and scalable for large behavioral datasets.
  4️⃣ Introduces **rotation-invariant visual embedding** that supports downstream behavior analysis.

- **Limitations**

  1️⃣ Requires ≥ 4 calibrated cameras for stable performance.
  2️⃣ Struggles with multi-animal occlusions and crowded scenes.
  3️⃣ Visual embedding interpretability can be improved.

Duke

References

[1] Cao, J., Tang, H., Fang, H. S., Shen, X., Lu, C., & Tai, Y. W. (2019). Cross-domain adaptation for animal pose estimation. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 9498-9507).
[2] Zuffi, S., Kanazawa, A., Jacobs, D. W., & Black, M. J. (2017). 3D menagerie: Modeling the 3D shape and pose of animals. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 6365-6373).
[3] Marshall, Jesse D.; Aldarondo, Diego; Wang, William; P. Ölveczky, Bence; Dunn, Timothy (2021). Rat 7M. figshare. Collection. https://doi.org/10.6084/m9.figshare.c.5295370.v3
[4] Kerbl, B., Kopanas, G., Leimkühler, T., & Drettakis, G. (2023). 3D Gaussian splatting for real-time radiance field rendering. *ACM Trans. Graph.*, *42*(4), 139-1.
[5] Zhu, Z., Fan, Z., Jiang, Y., & Wang, Z. (2024, September). Fsgs: Real-time few-shot view synthesis using gaussian splatting. In European conference on computer vision (pp. 145-163). Cham: Springer Nature Switzerland.
[6] Yang, C., Li, S., Fang, J., Liang, R., Xie, L., Zhang, X., ... & Tian, Q. (2024). Gaussianobject: High-quality 3d object reconstruction from four views with gaussian splatting. arXiv preprint arXiv:2402.10259.
[7] Charatan, D., Li, S. L., Tagliasacchi, A., & Sitzmann, V. (2024). pixelsplat: 3d gaussian splats from image pairs for scalable generalizable 3d reconstruction. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 19457-19467).
[8] Chen, Y., Xu, H., Zheng, C., Zhuang, B., Pollefeys, M., Geiger, A., ... & Cai, J. (2024, September). Mvsplat: Efficient 3d gaussian splatting from sparse multi-view images. In *European Conference on Computer Vision* (pp. 370-386). Cham: Springer Nature Switzerland.
[9] Ravi, N., Gabeur, V., Hu, Y. T., Hu, R., Ryali, C., Ma, T., ... & Feichtenhofer, C. (2024). Sam 2: Segment anything in images and videos. arXiv preprint arXiv:2408.00714.

Duke

QR Codes

Paper (arxiv)

Code (github)

Duke