Sim2real Cattle Joint Estimation in 3D point clouds

Mohammad Okour, Raphael Falque and Alen Alempijevic
Robotics Institute, University of Technology Sydney, NSW, Australia
{mohammad.okour}@student.uts.edu.au,
{raphael.falque, alen.alempijevic}@uts.edu.au

Abstract— **Understanding the well-being of cattle is crucial in various agricultural contexts. Cattle's body shape and joint articulation carry significant information about their welfare, yet acquiring comprehensive datasets for 3D body pose estimation presents a formidable challenge. This study delves into the construction of such a dataset specifically tailored for cattle. Leveraging the expertise of digital artists, we use a single animated 3D model to represent diverse cattle postures. To address the disparity between virtual and real-world data, we augment the 3D model's shape to encompass a range of potential body appearances, thereby narrowing the "sim2real" gap. We use these annotated models to train a deep-learning framework capable of estimating internal joints solely based on external surface curvature. Our contribution is specifically the use of geodesic distance over the surface manifold, coupled with multilateration to extract joints in a semantic keypoint detection encoder-decoder architecture. We demonstrate the robustness of joint extraction by comparing the link lengths extracted on real cattle mobbing and walking within a race. Furthermore, inspired by the established allometric relationship between bone length and the overall height of mammals, we utilise the estimated joints to predict hip height within a real cattle dataset, extending the utility of our approach to offer insights into improving cattle monitoring practices.**

I. INTRODUCTION

In modern agriculture, robotics and automation promise a transformative future, streamlining labour-intensive tasks and enhancing productivity [1]. Perception systems are critical for livestock production systems, the animal body structure has an impact on behavior, well-being, and fertility [2], [3]. Specifically, observing cattle during locomotion is crucial for identifying health issues in livestock, such as structural soundness and lameness [3], [4], [5]. Crucial in this assessment is identifying the pose of joints and limb actuation, allowing for body pose estimation. Advancements in perception offer invaluable insights for optimising livestock management practices and improving overall herd performance.

Human pose detection and tracking frameworks have garnered significant attention for their versatile applications in human-computer interaction and activity recognition [6]. However, the scarcity of annotated animal pose data presents a major obstacle to developing animal pose estimation approaches. Animals, unlike humans, lack the capability to cooperate during data collection, resulting in significant difficulty in coordinating them in the process. Available animal data sets lack ground truth for joint position and solely contain human-annotation [7], [8], [9].
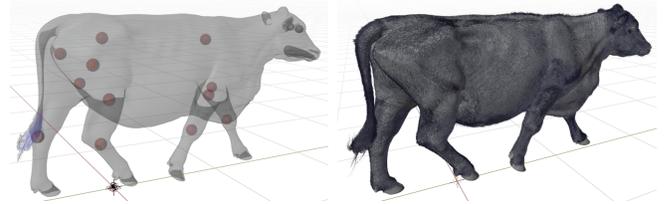
Fig. 1: Each model is annotated by twelve joints: two joints at each of the front legs, two at each of the back legs, two at each side of the hip bones, and two at either end of the spine.

Synthetic training data for 3D animal pose estimation has been explored, adapting techniques from human pose estimation [10]. This approach generates RGB images for network training, focusing on minimising distributional differences between synthetic and real animal data. However, the approach generally considers isolated animals rather than herds and does not fully exploit joint and shape information for comprehensive animal assessment. Moreover, existing simulations lack intra-species variability [10], limiting their ability to address *sim2real* challenges, particularly regarding shape deformation during animal movement [11]. Our previous work [12], [13] captures from multi-depth cameras capture 3D data of cattle in high fidelity while they are travelling through a race.

Extending our work on estimating joint coordinates from 3D point cloud data [13], we propose a methodology to utilise the manifold defined by the surface while leveraging animated 3D models to represent diverse cattle postures. Our contribution is specifically the use of geodesic distance over the surface manifold, coupled with multilateration to extract joints in a semantic keypoint detection encoder-decoder architecture. This enables the extraction of joint locations, where we demonstrate the robustness of our method to estimate joint locations on real cattle while they are in motion. We utilise joint information to estimate the hip height of cattle, drawing from work [14] that demonstrates an allometric relationship.

II. RELATED WORK

Pose estimation plays a crucial role across various disciplines, whether for humans or animals. In human-centric applications, such as virtual reality, gaming consoles, and human activity recognition, accurate pose estimation is

paramount. Early research by Shotton et al. [15] demonstrated the potential of depth data for human pose estimation, primarily in controlled environments like motion capture rooms. Subsequently, RGB-based approaches have gained momentum, with researchers like Pavlakos et al. [16] and Mathis et al. [17] employing deep-learning frameworks to estimate body joint poses. Notably, OpenPose [18] marked a significant milestone by introducing the first real-time multi-person system for body keypoints estimation, setting a benchmark for future research in human pose estimation. Additionally, Zhang et al. [19] delved deeper into the extraction of human keypoints in natural settings without human labels, utilizing 3D point clouds.

Animal pose estimation, drawing inspiration from human pose estimation techniques, has advanced significantly in assessing various health traits and behaviours in animals. However, the field encounters challenges due to the scarcity of fully annotated datasets, hindering the effective application of deep learning methodologies.

Existing animal datasets, including those referenced in [7], [5], predominantly consist of 2D images or sequences without ground truth annotations for joint positions. While custom animal datasets like AwA for quadrupeds [20] exist, they heavily rely on manual annotation, limiting their suitability for autonomous training of deep learning models.

Due to the scarcity of datasets in the animal field, researchers have explored various strategies to augment data input. One approach involves adapting learning models from other domains, such as humans or different animal species, and fine-tuning them [21], [22], [23], [24]. Other methods employ synthetic data augmentation processes [25], [26], [27], [28], [29].

While some of these approaches extract 3D keypoints, they may not precisely correspond to actual joints [30], or the resulting 3D models may lack an evaluation of mesh quality [31], [28]. Difficulties also arise in accurately assigning keypoint locations, necessitating interpretable relationships between the model and the actual animal shape [29]. Furthermore, although visually appealing 3D shapes may be generated, they may not be geometrically consistent with the animal's structure [32].

Recent investigations into synthetic animated training data for 3D pose estimation of animals draw inspiration from human pose estimation methodologies [10]. This process involves retraining established networks like OpenPose and Pose3D using simulated RGB and joint pose data generated under controlled conditions. The primary challenge is aligning the distribution of synthetic training data with that of real-world data collected from wildlife. While successful in addressing the difficulties of obtaining relevant anatomical keypoints on animal joints, this approach necessitates the conversion of RGB images into realistic environments. However, there have been no efforts to verify the differences between synthetic and real-world keypoint locations utilising 3D data for animal assessment.

Similarly, recent endeavours in extracting keypoints in 3D space, such as those proposed by other researchers [33], involve applying methods like heat kernel and geodesic distance. These keypoints are utilised for body segmentation, albeit positioned on the surface rather than within the joints. Alternatively, another work by [12] suggests using multi-depth-camera systems and PointNet++ for extracting semantic keypoints. However, the keypoints in their dataset are manually annotated and primarily located on the surface rather than internally within the body.

Our work seeks to quantify the robustness of estimating joint locations, informed by a synthetic dataset containing a variety of shapes and animal poses. We do so by examining joint data on real cattle while they are in motion, as well as establishing a relationship between joint data and hip height inspired by [14].

III. Methodology

Given that joints are not inherently situated on the surface, their locations must be inferred from curvature on the surface via a deep-learning model. The system architecture, presented in Figure 2, comprises four principal modules: point cloud creation, data preprocessing, a deep-learning model (PointNet++) tasked with estimating the nearest point cloud to each joint, and joint estimation. We also include an MLP model utilising joint information for hip height estimation. Apart from the annotated joints listed in [13], which include Carpal joint left, Elbow joint left, Carpal joint right, Elbow joint right, Tarsal joint left, Stifle joint left, Hip joint left, Tarsal joint right, Stifle joint right, Hip joint right, and Front spine, a new keypoint positioned at the tip of the cattle's hip, termed the Illium joint, has been introduced. This is depicted in Figure 3. This new keypoint is located on the tip of the cattle's hip. This results in a total of twelve keypoints per animal, as illustrated in Figure 1, which are subject to prediction.

The PointNet model takes the point cloud location as input. As the joints lie outside the mesh surface, this necessitates the estimation of their positions from distances inferred by the deep learning model via multilateration. This method utilises Euclidean distance to determine the requisite points.

Small Euclidean distances between annotated joints within a point cloud can introduce significant errors during the PointNet estimation. In such cases, the model may struggle to distinguish between closely spaced joints, leading to inaccurate estimations. To address this, geodesic distances are computed as an additional input, incorporating information on the distance over the surface in the learning framework.

Data processing involves two primary steps. Firstly, geodesic distances are obtained, representing the distance on the manifold from each joint utilising the heat kernel method. Simultaneously, Euclidean distances from each joint to the generated point cloud are computed. Secondly, the maximum value between these two distance metrics is determined. The loss function is then computed based on the logarithmic value of the maximum distances carried by each point cloud, relative to a known mesh.

To compute the geodesic distances, we first downsample the generated point clouds and then utilise the heat kernel
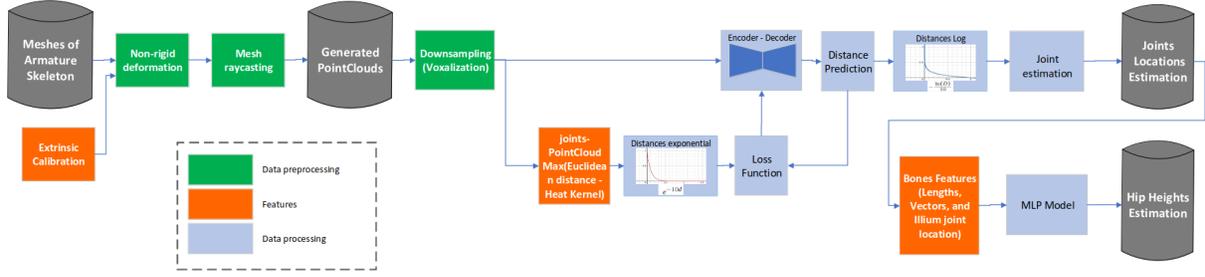
Fig. 2: Method overview: from the simulated model, the armature undergoes rigid scaling and meshes a non-rigid deformation. Through raycasting over a number of cameras, several point clouds are generated and merged. At inference time, the merged point cloud is passed into an encoder-decoder architecture (Pointnet++ [34]) to extract the keypoints. During training, the dataset uses keypoints from the armature and the distances on the manifold are pre-computed. The encoder-decoder inputs are $n \times 3$ points, and the outputs are the $n \times 13$ distances to the 13 joints keypoints.
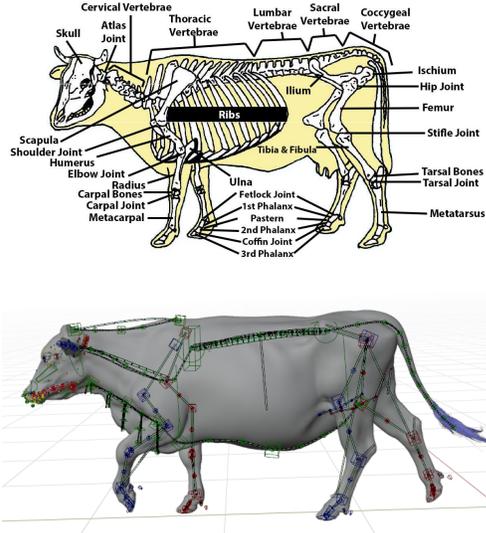
Fig. 3: Top: Cattle skeleton from [35] containing information of all the joints and bones. Bottom: The annotated model used in this work containing rigging and joints indicated by blue squares

Fig. 4: Barycetnric diagram of α, β, and γ. Where D_{g1}, D_{g2}, and D_{g3} are the heat kernel distances to a point in space.

$$D = \sum (B \times \begin{bmatrix} D_{g1} \\ D_{g2} \\ D_{g3} \end{bmatrix}) \qquad (2)$$

To accomplish this, during the creation of point clouds from the simulated model, we extract the hit face number for each point in the point cloud generated by ray casting. Subsequently, we consider the maximum distance (D_{max}) between the heat kernel distance on the manifold D and the Euclidean distance from a joint to each point in the point clouds, to be applied in the loss in the encoder-decoder network. We proposed to take the maximum distances to achieve a more accurate estimation of joint positions in cases where the joints are very close together, such as during walking. The logarithmic transformation is then applied to the maximum values to increase the penalty in the loss function as illustrated in equation 3. The effect of this transformation on one of the instances can be shown in Figure 5

$$L = -ln(D_{max})/10 \qquad (3)$$

The encoder-decoder network takes as input the point cloud (i.e., a matrix I of size $n \times 3$) and outputs a matrix $\hat{D}$ of size $n \times m$, where m represents the number of joints to be predicted.

The final step in predicting joints involves utilising the distance predictions from the PointNet++ model to estimate the positions of the joints. In line with prior research [38], we propose employing the multilateration technique to enhance joint estimation accuracy as in equation 4. Specifically, we

method [36] to precompute the distances on the manifold D_g. This computation leverages the tufted Laplacian [37] derived from the mesh structure. Subsequently, we employ a barycentric calculation, as depicted in Equation 1, to determine the geodesic distance of each point cloud to the face vertices as shown in figure 4. The output is used to find the final heat kernel distance D as illustrated in equation 2.

$$
\begin{aligned}
\alpha &= \frac{(\vec{AC} \cdot \vec{AC})(\vec{AB} \cdot \vec{AP}) - (\vec{AB} \cdot \vec{AC})(\vec{AC} \cdot \vec{AP})}{(\vec{AB} \cdot \vec{AB})(\vec{AC} \cdot \vec{AC}) - (\vec{AB} \cdot \vec{AC})^2} \\
\beta &= \frac{(\vec{AB} \cdot \vec{AB})(\vec{AC} \cdot \vec{AP}) - (\vec{AB} \cdot \vec{AC})(\vec{AB} \cdot \vec{AP})}{(\vec{AB} \cdot \vec{AB})(\vec{AC} \cdot \vec{AC}) - (\vec{AB} \cdot \vec{AC})^2} \\
\gamma &= 1 - \alpha - \beta \\
B &= \begin{bmatrix} \alpha \\ \beta \\ \gamma \end{bmatrix}
\end{aligned}
\qquad (1)
$$

designate the first point within the area of interest of the point cloud as the anchor and apply the least squares method to determine the joint positions. This area of interest constitutes a subset of points with the lowest estimated D_{max}, as illustrated in Figure 9. Formulations 5 and 6 allow predicting joints' position outside of the point cloud in the case where the underlying shape is convex, where d_n is the D_{max}. A sample of the joint detection using multilateration of the nearest point cloud group to a joint is displayed in Figure 9.

$$\varepsilon = \sum_{i=1}^{N} (\sqrt{(x_i - x)^2 + (y_i - y)^2 + (z_i - z)^2)} - \tilde{d}_i)^2 \quad (4)$$

$$\begin{bmatrix} 2x_2 & 2y_2 & 2z_2 \\ \vdots & \vdots & \vdots \\ 2x_n & 2y_n & 2z_n \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} x_2^2 + y_2^2 + z_2^2 - d_2^2 + d_1^2 \\ \vdots \\ x_n^2 + y_n^2 + z_n^2 - d_n^2 + d_1^2 \end{bmatrix} \quad (5)$$

$$\hat{x} = (H^T H)^{-1} H^T \tilde{b} \quad (6)$$

Where H = $\begin{bmatrix} 2x_2 + 2y_2 + 2z_2 \\ \vdots \\ 2x_n + 2y_n + 2z_n \end{bmatrix}$,

$\tilde{b} = \begin{bmatrix} x_2^2 + y_2^2 + z_2^2 - d_2^2 + d_1^2 \\ \vdots \\ x_n^2 + y_n^2 + z_n^2 - d_n^2 + d_1^2 \end{bmatrix}$, and $\hat{x} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}$

We thereafter employ a Multilayer Perceptron (MLP) model, consisting of 3 hidden layers (with 9, 7, and 5 neurons for each layer) with ReLU activation functions, to estimate the hip height. The model is trained using the Adam optimiser on a dataset comprising 175 instances of real animals, applying the leave-one-out technique to ensure fair evaluation. The features used for this prediction contain the coordinates of the keypoint related to the hip of the animal and information related to the backbones (i.e., their length and the vector from joint to joint). The backbones used in this module include Femur, Tibia, and Fibula. While the hip keypoint is the joint between the Ilium and Lumbar Vertebrae bones, as indicated in figure 3.

In summary, we employ the heat kernel method [36] to compute the maximum value between the distance on the manifold $\boldsymbol{D}$, obtained via the tufted Laplacian [37] on the mesh, and the euclidean distance from each joint to the point clouds. Subsequently, we utilise a barycentric calculation to derive the geodesic distance of each point cloud to the vertices of each triangular face of the mesh. This distance on the manifold is then treated as a feature and learned using an encoder-decoder network. The inputs of the encoder-decoder consist of the point cloud size, represented by a matrix $\boldsymbol{I}$ of size $n \times 3$, while the outputs correspond to the number of joints to be predicted, represented by a matrix $\hat{\boldsymbol{D}}$ of size $n \times m$.

IV. EXPERIMENTS

Utilising the annotated dataset derived from [13] and employing the joint prediction methodology detailed in Sec-
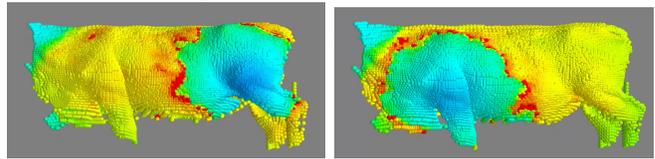
Fig. 5: Predicted distance on the manifold. Points coloured in blue represent the nearest points to a joint. Left the rear leg and right the front leg are being evaluated.
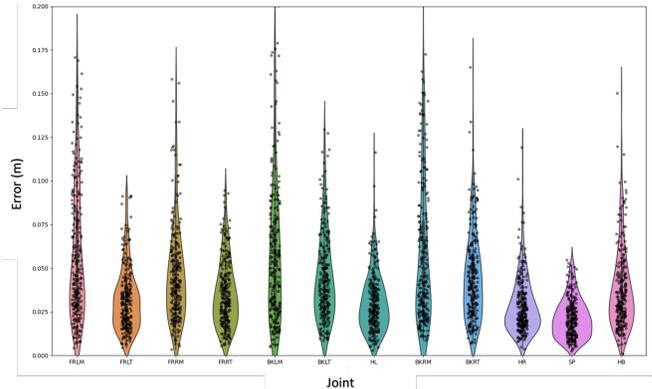
Fig. 6: Joints estimation error of synthetic data set. Joint names starting from left to right: Carpal joint left, Elbow joint left, Carpal joint right, Elbow joint right, Tarsal joint left, Stifle joint left, Hip joint left, Tarsal joint right, Stifle joint right, Hip joint right, Front spine, and Illium joint

tion III, we calculate the joints estimation error based on a simulated test dataset, as illustrated in Figure 6. The analysis reveals a mean error in joints estimation of 0.03 and a standard deviation of 0.01.

We assess the *sim2real* gap, resulting from training the network on a simulated dataset, via the consistency of joint predictions on a real animal walking through the race. We analyse the variation in bone length estimations and distance between joints across eighteen consecutive frames (instances); the findings of this evaluation are detailed in Figure 7, shedding light on the stability and reliability of the model's predictions of joint location in real-world scenarios.

A further quantitative evaluation is conducted by estimating the hip height of 175 real animals where the cattle were subsequently restrained to manually measure hip height. This evaluation is depicted in Figure 8. As detailed in Section III, the features employed for this prediction encompass the length of the back bones, the geometric vectors associated with these bones, and the highest point among the estimated nearest points of the hip keypoint.

V. DISCUSSION

The results obtained from the joint estimation error analysis reveal a mean error in joint estimation of 0.026 and a standard deviation of 0.012, with the majority of errors falling within the range of 100 mm. This indicates a certain level of accuracy in the joint prediction methodology utilised,
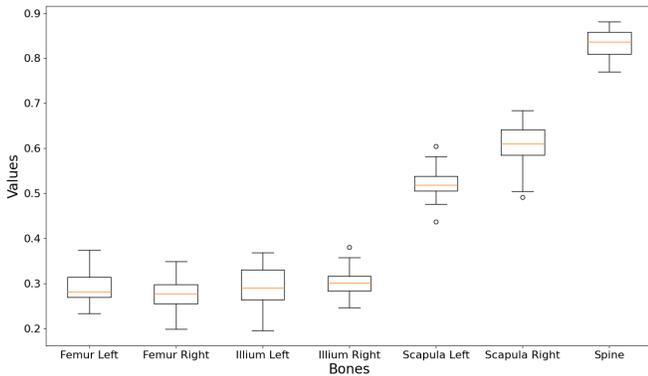
Fig. 7: Bone length estimation on a walking animal. This estimation was performed over eighteen consecutive frames while the animal walked through the race. For each bone length, we denote the mean, standard deviation and minimum/maximum estimate on the box plot. The scapula and spine are the longest bone structures.

Fig. 8: Predicted vs actual values of hip height estimation for a dataset comprising 175 instances of real animals, using the leave-one-out technique. The R^2 score is 0.64, and the Root Mean Square Error (RMSE) is 2.97. The green dashed line represents the ideal output, while the red dashed lines indicate a margin of ± 3.

(a)

(b)

Fig. 9: (a): The blue points represent the complete set of the lowest estimated D_{max} values. The area of interest, shown in black, highlights a subset of these points with the lowest estimated D_{max}. This output corresponds to the front left joint. (b) Represents joints prediction of an instance (red spheres are the ground truth and green are the estimated joints) with the error between both in meters

spine, underscores the precision of the model's estimations. These results contribute to the understanding of how predictive models can effectively estimate bone lengths in dynamic scenarios, such as walking, in cattle.

Overall, these results demonstrate the effectiveness of the employed methodologies in estimating joint positions, hip height, and bone lengths, with some limitations and opportunities for further refinement.

VI. CONCLUSION

This study presents a novel method for predicting joints using synthetic auto-annotated datasets. By estimating keypoints outside the mesh and utilising created point clouds, we address the challenge of bridging the gap between training deep models with simulated datasets and real-world data for this specific application.

Although our findings demonstrate promising results in joint prediction, further exploration is warranted to evaluate the potential applicability of this approach to other domains, such as health-related metrics. Additionally, future research could explore the development of anatomically accurate models for muscle and fat layers attached to the joints. Creating such models is a complex task typically undertaken by only a select few highly skilled 3D artists. Moreover, making these models parametric to encompass various body conditions for extensive data augmentation would introduce additional challenges to the process.

although there is room for improvement, particularly in reducing the maximum error.

Moving on to the quantitative evaluation of the *sim2real* gap through hip height estimation, we observe a coefficient of determination (R^2) of 0.64 and a root mean square error (RMSE) of 2.97. Despite the absence of certain segments within the real animal testing dataset, these metrics suggest a moderate level of predictive capability, implying that the trained network can reasonably estimate hip height based on the features employed.

Figure 7 illustrates the model's predictions of bone lengths for 18 consecutive walking cattle, offering insights into the accuracy and reliability of these predictions. Our findings reveal a smaller difference between the minimum and maximum values, indicating minimal variation in bone lengths during walking motion and suggesting the model's accuracy in predicting such changes. This consistency across diverse bones, including the left and right femur, ilium, scapula, and

REFERENCES

[1] L. N. Duong, M. Al-Fadhli, S. Jagtap, F. Bader, W. Martindale, M. Swainson, and A. Paoli, "A review of robotics and autonomous systems in the food industry: From the supply chains perspective," *Trends in Food Science & Technology*, vol. 106, pp. 355–364, 2020.

[2] H. M. Saad, R. M. Enns, M. G. Thomas, L. L. Leachman, and S. E. Speidel, "Foot scores genetic parameters estimation in beef cattle," *Translational Animal Science*, vol. 5, no. Supplement_S1, pp. S180–S184, 2021.

[3] M. B. Sadiq, S. Z. Ramanoon, W. M. Shaik Mossadeq, R. Mansor, and S. S. Syed-Hussain, "Association between lameness and indicators of dairy cow welfare based on locomotion scoring, body and hock condition, leg hygiene and lying behavior," *Animals*, vol. 7, no. 11, p. 79, 2017.

[4] M. Bell, M. Maak, M. Sorley, and R. Proud, "Comparison of methods for monitoring the body condition of dairy cows," *Frontiers in Sustainable Food Systems*, vol. 2, 11 2018.

[5] H. Russello, R. van der Tol, and G. Kootstra, "T-leap: Occlusion-robust pose estimation of walking cows using temporal information," *Computers and Electronics in Agriculture*, vol. 192, p. 106559, 2022.

[6] J. Wang, S. Tan, X. Zhen, S. Xu, F. Zheng, Z. He, and L. Shao, "Deep 3d human pose estimation: A review," *Computer Vision and Image Understanding*, vol. 210, p. 103225, 2021.

[7] H. Yu, Y. Xu, J. Zhang, W. Zhao, Z. Guan, and D. Tao, "Ap-10k: A benchmark for animal pose estimation in the wild," *arXiv preprint arXiv:2108.12617*, 2021.

[8] L. Jianguo, L. Weiyao, H. Tang, M. Greg, and D. Joachim, "Iccv 2019 workshop & challenge on computer vision for wildlife conservation (CVWC)," 2019. [Online]. Available: https://cvwc2019.github.io/challenge.html

[9] D. Joska, L. Clark, N. Muramatsu, R. Jericevich, F. Nicolls, A. Mathis, M. W. Mathis, and A. Patel, "Acinoset: a 3d pose estimation dataset and baseline models for cheetahs in the wild," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021.

[10] A. S. Fangbemi, Y. F. Lu, M. Y. Xu, X. W. Luo, A. Rolland, and C. Raissi, "Zoobuilder: 2d and 3d pose estimation for quadrupeds using synthetic data," 2020. [Online]. Available: https://arxiv.org/abs/2009.05389

[11] S. Höfer, K. Bekris, A. Handa, J. C. Gamboa, M. Mozifian, F. Golemo, C. Atkeson, D. Fox, K. Goldberg, J. Leonard *et al.*, "Sim2real in robotics and automation: Applications and challenges," *IEEE transactions on automation science and engineering*, vol. 18, no. 2, pp. 398–400, 2021.

[12] R. Falque, T. Vidal-Calleja, and A. Alempijevic, "Semantic keypoint extraction for scanned animals using multi-depth-camera systems," *arXiv preprint arXiv:2211.08634*, 2022.

[13] M. Okour, R. Guenot-Falque, T. Vidal Calleja, and A. Alempijevic, "Sim2real cattle pose prediction in 3d pointclouds," in *Australasian Conference on Robotics and Automation*. ARAA, 2022.

[14] T. E. Ali, E. Burnside, and L. Schaeffer, "Relationship between external body measurements and calving difficulties in canadian holstein-friesian cattle," *Journal of Dairy Science*, vol. 67, no. 12, pp. 3034–3044, 1984.

[15] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake, "Real-time human pose recognition in parts from single depth images," in *CVPR 2011*, 2011, pp. 1297–1304.

[16] G. Pavlakos, X. Zhou, and K. Daniilidis, "Ordinal depth supervision for 3d human pose estimation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 7307–7316.

[17] A. Mathis, P. Mamidanna, K. M. Cury, T. Abe, V. N. Murthy, M. W. Mathis, and M. Bethge, "Deeplabcut: markerless pose estimation of user-defined body parts with deep learning," *Nature neuroscience*, vol. 21, no. 9, pp. 1281–1289, 2018.

[18] Z. Cao, G. Hidalgo Martinez, T. Simon, S. Wei, and Y. A. Sheikh, "Openpose: Realtime multi-person 2d pose estimation using part affinity fields," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019.

[19] D. Zhang, Y. Wu, M. Guo, and Y. Chen, "Deep learning methods for 3d human pose estimation under different supervision paradigms: a survey," *Electronics*, vol. 10, no. 18, p. 2267, 2021.

[20] P. Banik, L. Li, and X. Dong, "A novel dataset for keypoint detection of quadruped animals from images," *arXiv preprint arXiv:2108.13958*, 2021.

[21] A. Sanakoyeu, V. Khalidov, M. S. McCarthy, A. Vedaldi, and N. Neverova, "Transferring dense pose to proximal animal classes," in *Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition*, 2020, pp. 5233–5242.

[22] A. Mathis, T. Biasi, S. Schneider, M. Yuksekgonul, B. Rogers, M. Bethge, and M. W. Mathis, "Pretraining boosts out-of-domain robustness for pose estimation," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2021.

[23] T. D. Pereira, N. Tabris, J. Li, S. Ravindranath, E. S. Papadoyannis, Z. Y. Wang, D. M. Turner, G. McKenzie-Smith, S. D. Kocher, A. L. Falkner *et al.*, "Sleap: Multi-animal pose tracking," *BioRxiv*, 2020.

[24] J. Cao, H. Tang, H.-S. Fang, X. Shen, C. Lu, and Y.-W. Tai, "Cross-domain adaptation for animal pose estimation," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 9498–9507.

[25] J. Mu, W. Qiu, G. D. Hager, and A. L. Yuille, "Learning from synthetic animals," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 12 386–12 395.

[26] L. Del Pero, S. Ricco, R. Sukthankar, and V. Ferrari, "Behavior discovery and alignment of articulated object classes from unstructured video," *International Journal of Computer Vision*, vol. 121, no. 2, pp. 303–325, 2017.

[27] ——, "Articulated motion discovery using pairs of trajectories," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 2151–2160.

[28] C. Li and G. H. Lee, "From synthetic to real: Unsupervised domain adaptation for animal pose estimation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 1482–1491.

[29] S. Zuffi, A. Kanazawa, T. Berger-Wolf, and M. J. Black, "Three-d safari: Learning to estimate zebra pose, shape, and texture from images" in the wild"," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 5359–5368.

[30] M. Badger, Y. Wang, A. Modh, A. Perkes, N. Kolotouros, B. G. Pfrommer, M. F. Schmidt, and K. Daniilidis, "3d bird reconstruction: a dataset, model, and shape recovery from a single view," in *European Conference on Computer Vision*. Springer, 2020, pp. 1–17.

[31] Y. Yao, Y. Jafarian, and H. S. Park, "Monet: Multiview semi-supervised keypoint detection via epipolar divergence," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 753–762.

[32] C.-H. Yao, W.-C. Hung, Y. Li, M. Rubinstein, M.-H. Yang, and V. Jampani, "Lassie: Learning articulated shapes from sparse image ensemble via 3d part discovery," *arXiv preprint arXiv:2207.03434*, 2022.

[33] H. Yang, K. Hao, and Y. Ding, "Semantic segmentation of human model using heat kernel and geodesic distance," *Mathematical Problems in Engineering*, vol. 2018, pp. 1–13, 2018.

[34] C. R. Qi, L. Yi, H. Su, and L. J. Guibas, "Pointnet++: Deep hierarchical feature learning on point sets in a metric space," *Advances in neural information processing systems*, vol. 30, 2017.

[35] C. M. Gatton, "Beef cattle discovery - skeletal," https://afs.ca.uky.edu/livestock/beef/skeletal/summary, [Accessed 25-02-2024].

[36] K. Crane, C. Weischedel, and M. Wardetzky, "The heat method for distance computation," *Commun. ACM*, vol. 60, no. 11, pp. 90–99, Oct. 2017.

[37] N. Sharp and K. Crane, "A Laplacian for Nonmanifold Triangle Meshes," *Computer Graphics Forum (SGP)*, vol. 39, no. 5, 2020.

[38] P. Tarrío, A. M. Bernardos, and J. R. Casar, "Weighted least squares techniques for improved received signal strength based localization," *Sensors*, vol. 11, no. 9, pp. 8569–8592, 2011.