

# Avances en Reconstrucción 3D de Cuerpos Humanos mediante Deep Learning

Magda Alexandra Trujillo-Jiménez<sup>1,2</sup>[0000-0001-5506-3496], José Pablo Navarro<sup>1,2</sup>[0000-0003-2180-449X], and Rolando González-José<sup>2</sup>[0000-0002-8128-9381]

<sup>1</sup> Laboratorio de Ciencias de las Imágenes, Departamento de Ingeniería Eléctrica y Computadoras, Universidad Nacional del Sur y CONICET, Bahía Blanca, B8000, Argentina

<sup>2</sup> Instituto Patagónico de Ciencias Sociales y Humanas, Centro Nacional Patagónico, CCT CENPAT CONICET, Puerto Madryn, U9120, Argentina  
{mtrujillo, pnavarro, rolando}@cenpat-conicet.gob.ar

**Abstract.** En campos como la salud, la antropología, el deporte, el diseño de indumentaria y otros ámbitos relacionados, la obtención de medidas antropométricas precisas es esencial para diversas aplicaciones prácticas, abarcando desde la forma y el tamaño del cuerpo humano hasta la medición de perímetros y volúmenes. El avance de métodos basados en redes neuronales profundas y técnicas de Machine Learning ha revolucionado la capacidad para crear modelos tridimensionales precisos, especialmente en el análisis de mallas 3D, con un enfoque particular en cuerpos humanos. Se presentará una revisión exhaustiva de los avances más recientes en este campo, destacando los modelos y técnicas utilizados para la generación y análisis de mallas 3D aplicados a la reconstrucción antropométrica utilizando modelos basados en Deep Learning. Se abordarán métodos innovadores de segmentación y clasificación, así como técnicas para la estimación precisa de la pose y forma humana, enfatizando los modelos basados en Deep Learning que han simplificado estas tareas. Además, se compartirán resultados preliminares sobre la reconstrucción 3D de cuerpos humanos a partir de video, junto con una discusión sobre los desafíos continuos en términos de precisión y calidad de las reconstrucciones, así como la importancia de optimizar los recursos computacionales para avanzar en este campo prometedor.

**Keywords:** 3D · Reconstrucción corporal · Deep Learning.

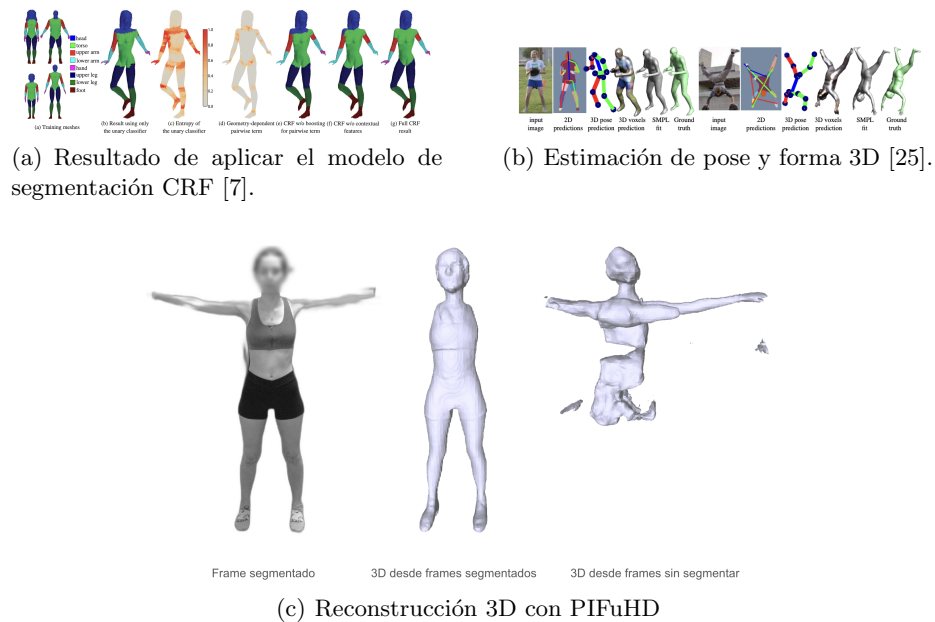
## 1 Introducción

En diversos dominios como la medicina, el deporte, la moda y áreas afines, existe una demanda constante de medidas antropométricas precisas que abarcan la forma, el tamaño y otros aspectos morfológicos del cuerpo humano [15,20,14]. Sin embargo, la obtención de modelos tridimensionales (3D) de cuerpos humanos, principalmente a través de escaneos corporales, presenta desafíos significativos. Aunque se requiera que el sujeto permanezca estático dentro del escáner, los

datos resultantes suelen ser superficiales e incompletos debido a oclusiones y movimientos involuntarios [1]. Además, los altos costos y la complejidad asociada con tales escáneres limitan su accesibilidad para la mayoría de los profesionales médicos, especialistas e investigadores. El análisis de modelos tridimensionales, especialmente en el contexto de cuerpos humanos, desempeña un papel crucial en la superación de estos desafíos. La habilidad para interpretar y manipular estos modelos con precisión es fundamental para el progreso en estos campos [11,23,17]. El avance significativo de las redes neuronales profundas ha sido un catalizador clave en el desarrollo de esta disciplina, impulsando innovaciones y soluciones efectivas para abordar los desafíos mencionados. Los principales avances en métodos basados en Deep Learning se centran en la segmentación, clasificación y generación de mallas 3D de cuerpos humanos [7,25]. Desde la identificación de regiones anatómicas hasta la estimación de la forma y postura humanas [1], estos avances están transformando nuestra capacidad para interpretar y utilizar datos tridimensionales de manera efectiva. A su vez, el desarrollo de redes neuronales profundas y técnicas de aprendizaje automático ha mejorado significativamente la capacidad para crear modelos tridimensionales precisos, particularmente en la reconstrucción de mallas 3D de cuerpos humanos [6,19,2]. En este artículo, se presenta una revisión de los avances más recientes en este ámbito, destacando modelos y técnicas empleadas para la generación y análisis de mallas 3D aplicadas a la reconstrucción antropométrica. Específicamente, se presentan resultados preliminares de la reconstrucción 3D de cuerpos humanos a partir de imágenes 2D, resaltando los desafíos persistentes en términos de precisión y calidad de las reconstrucciones, así como la necesidad de optimizar recursos computacionales.

## 2 Métodos de Segmentación y Estimación de Pose

En la investigación actual sobre la reconstrucción 3D de cuerpos humanos mediante Deep Learning, se han desarrollado varios enfoques innovadores para abordar la segmentación de mallas tridimensionales [18,29,7,4]. Entre estos enfoques, destacan modelos como CRF [7], MultiBodySync [4], que han demostrado una eficacia notable. Estos modelos están diseñados para aprender de manera automática a identificar regiones específicas dentro de una malla o nube de puntos 3D y asignarles etiquetas correspondientes. En CRF, el algoritmo utiliza cientos de características geométricas y contextuales de las etiquetas y aprende distintos tipos de segmentaciones para distintas tareas (ver Fig. 1(a)), sin necesidad de ajustar manualmente los parámetros, produciendo a menudo segmentaciones y etiquetados comparables a los producidos por humanos [7]. MultiBodySync, incorpora la sincronización espectral en una red declarativa profunda iterativa, al separar explícitamente los módulos de estimación de la correspondencia y de la segmentación del movimiento, se logra una gran generalizabilidad entre diferentes categorías de objetos (incluyendo cuerpos humanos). Al poder identificar y etiquetar automáticamente estructuras anatómicas relevantes, estos modelos facilitan tanto el análisis como la posterior manipulación de los datos, lo que



**Fig. 1.** Modelos basados en Deep Learning para la segmentación corporal, estimación de pose y forma, y reconstrucción 3D.

resulta fundamental en numerosas aplicaciones clínicas y de investigación. Por su parte, la estimación de la pose y la forma humana a partir de modelos 3D sigue siendo un desafío en la visión por computadora. Modelos como los de SMPL [12] han facilitado esta tarea al aprender la forma del cuerpo humano y su variación en función de la pose [25,16,27]. Por ejemplo, el método presentado en BodyNet [25] tiene una arquitectura de red multitarea totalmente automática de extremo a extremo que predice la forma del cuerpo humano en 3D a partir de una sola imagen (ver Fig. 1(b)). La imagen RGB de entrada pasa primero por subredes para la estimación de la pose 2D y la segmentación de las partes del cuerpo en 2D. Estas predicciones, combinadas con la RGB, se introducen en otra red que predice la pose 3D. Todas las subredes se combinan en una red final para inferir la forma volumétrica. Sin embargo, persisten desafíos en la generación de reconstrucciones con un nivel de detalle adecuado, especialmente cuando implica mediciones antropométricas precisas. Los métodos de segmentación y estimación de forma y pose operan tanto con modelos obtenidos de escáneres como con diseños computacionales. Sin embargo, la variabilidad morfológica entre individuos, particularmente en las estructuras anatómicas humanas, plantea desafíos singulares que demandan enfoques innovadores y adaptativos. Por consiguiente, es crucial evaluar estos métodos en modelos basados en cuerpos humanos reales, como aquellos derivados de la reconstrucción 3D. Es importante destacar que en particular las poblaciones latinoamericanas están subestimadas en las bases

de datos preexistentes, lo que subraya la necesidad de incluir una representación más diversa para mejorar la precisión y aplicabilidad de estos métodos en una gama más amplia de contextos culturales y biológicos.

### 3 Generación de Mallas 3D

El uso de redes neuronales convolucionales ha permitido avances significativos en la generación de mallas 3D a partir de imágenes 2D. Modelos como Pixel2Mesh [26] y Point2Mesh [3] han demostrado la capacidad de inferir la estructura tridimensional de objetos a partir de imágenes o datos de nube de puntos. Para la reconstrucción de cuerpos humanos, algunos métodos recientes usan diferentes técnicas de Deep Learning que permiten la recuperación completa de formas o posturas humanas en 3D a partir de vídeos o imágenes 2D [22,9,13]. En particular, PIFuHD [19], es un modelo que utiliza una función implícita alineada por píxeles para generar mallas tridimensionales con buena calidad y detalle. Sin embargo, mantiene algunas limitaciones como el tamaño de entrada ( $512 \times 512$ ) y resolución ( $128 \times 128$ ), e imprecisiones al inferir la profundidad de forma. Presentamos algunos ejemplos de resultados preliminares de reconstrucciones de cuerpos humanos usando esta red, y comparaciones cuando usamos imágenes de entrada segmentada y sin segmentar (ver Fig. 1(c)). Se confirma que, al realizar una segmentación previa, se tiene la certeza de que se elimina el ruido de fondo presente en la escena [22] y se obtienen resultados de reconstrucción más fiables. El código utilizado para probar el modelo PIFuHD se encuentra disponible en el repositorio [https://github.com/aletrujim/3Dtest/blob/main/PIFu\\_test.ipynb](https://github.com/aletrujim/3Dtest/blob/main/PIFu_test.ipynb). Finalmente, la reciente bibliografía sobre la reconstrucción de humanos en 3D basada en Gaussian Splatting destaca en la creación de avatares humanos detallados y animables. Modelos como GaussianBody [8], GVA [10] y HAHA [21] utilizan esta técnica para generar avatares articulados a partir de videos, superando los desafíos de consistencia temporal y precisión geométrica. Además, propuestas como GoMAvatar [28] y SplatArmor [5] ofrecen soluciones eficientes para la animación de estos modelos, integrando Gaussians en estructuras de malla y mejorando su adaptabilidad a diversos escenarios. Estos estudios no solo demuestran el potencial de Gaussian Splatting en la reconstrucción de humanos, sino que también establecen un marco robusto para futuras aplicaciones en animación y realidad virtual.

### 4 Conclusiones

A pesar de los avances logrados, persisten desafíos significativos en este campo. El etiquetado de partes específicas en mallas 3D y la segmentación precisa de objetos tridimensionales continúan siendo áreas de investigación abiertas. La complejidad inherente a los modelos 3D, debido a sus grandes tamaños, requiere mucho espacio en memoria, alta capacidad de computación y tiempos prolongados de entrenamiento [7], lo que limita la escalabilidad y robustez de los modelos [4]. Los avances en reconstrucción 3D de cuerpos humanos mediante Deep Learning

han abierto nuevas posibilidades en diversas aplicaciones. Sin embargo, persisten desafíos en términos de precisión y calidad de las reconstrucciones, así como en la optimización de recursos computacionales. En particular, como ya se ha presentado [24], será posible verificar que la morfometría del cuerpo humano es un predictor robusto de fenotipos biomédicos relacionados con afecciones, como la obesidad y el sobrepeso, así como validar su utilidad en la práctica clínica rutinaria.

## References

1. Anguelov, D., Srinivasan, P., Koller, D., Thrun, S., Rodgers, J., Davis, J.: Scape: shape completion and animation of people. In: ACM SIGGRAPH 2005 Papers, pp. 408–416 (2005)
2. Guo, C., Jiang, T., Chen, X., Song, J., Hilliges, O.: Vid2avatar: 3d avatar reconstruction from videos in the wild via self-supervised scene decomposition. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 12858–12868 (2023)
3. Hanocka, R., Metzer, G., Giryas, R., Cohen-Or, D.: Point2mesh: A self-prior for deformable meshes. arXiv preprint arXiv:2005.11084 (2020)
4. Huang, J., Wang, H., Birdal, T., Sung, M., Arrigoni, F., Hu, S.M., Guibas, L.J.: Multibodysync: Multi-body segmentation and motion estimation via 3d scan synchronization. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 7108–7118 (2021)
5. Jena, R., Iyer, G.S., Choudhary, S., Smith, B., Chaudhari, P., Gee, J.: Splatarmor: Articulated gaussian splatting for animatable humans from monocular rgb videos. arXiv preprint arXiv:2311.10812 (2023)
6. Jiang, H., Cai, J., Zheng, J.: Skeleton-aware 3d human shape reconstruction from point clouds. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 5431–5441 (2019)
7. Kalogerakis, E., Hertzmann, A., Singh, K.: Learning 3d mesh segmentation and labeling. In: ACM SIGGRAPH 2010 papers, pp. 1–12 (2010)
8. Li, M., Yao, S., Xie, Z., Chen, K., Jiang, Y.G.: Gaussianbody: Clothed human reconstruction via 3d gaussian splatting. arXiv preprint arXiv:2401.09720 (2024)
9. Liu, J., Akhtar, N., Mian, A.: Deep reconstruction of 3d human poses from video. IEEE Transactions on Artificial Intelligence (2022)
10. Liu, X., Wu, C., Liu, X., Liu, J., Wu, J., Zhao, C., Feng, H., Ding, E., Wang, J.: Gea: Reconstructing expressive 3d gaussian avatar from monocular video. arXiv preprint arXiv:2402.16607 (2024)
11. Löffler-Wirth, H., Willscher, E., Ahnert, P., Wirkner, K., Engel, C., Loeffler, M., Binder, H.: Novel anthropometry based on 3d-body-scans applied to a large population based cohort. PloS one **11**(7), e0159887 (2016)
12. Loper, M., Mahmood, N., Romero, J., Pons-Moll, G., Black, M.J.: Smpl: A skinned multi-person linear model. In: Seminal Graphics Papers: Pushing the Boundaries, Volume 2, pp. 851–866 (2023)
13. Lu, Y., Yu, H., Ni, W., Song, L.: 3d real-time human reconstruction with a single rgbd camera. Applied Intelligence **53**(8), 8735–8745 (2023)
14. Ma, Q., Yang, J., Ranjan, A., Pujades, S., Pons-Moll, G., Tang, S., Black, M.J.: Learning to dress 3d people in generative clothing. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 6469–6478 (2020)

15. Maessen, M.F., Eijsvogels, T.M., Verheggen, R.J., Hopman, M.T., Verbeek, A.L., Vegt, F.d.: Entering a new era of body indices: the feasibility of a body shape index and body roundness index to identify cardiovascular health status. *PloS one* **9**(9), e107212 (2014)
16. Mehta, D., Sotnychenko, O., Mueller, F., Xu, W., Elgharib, M., Fua, P., Seidel, H.P., Rhodin, H., Pons-Moll, G., Theobalt, C.: Xnect: Real-time multi-person 3d motion capture with a single rgb camera. *Acm Transactions On Graphics (TOG)* **39**(4), 82–1 (2020)
17. Navarro, P., Ramallo, V., Cintas, C., Ruderman, A., de Azevedo, S., Paschetta, C., Pérez, O., Pazos, B., Delrieux, C., González-José, R.: Body shape: Implications in the study of obesity and related traits. *American Journal of Human Biology* **32**(2), e23323 (2020)
18. Omran, M., Lassner, C., Pons-Moll, G., Gehler, P., Schiele, B.: Neural body fitting: Unifying deep learning and model based human pose and shape estimation. In: 2018 international conference on 3D vision (3DV). pp. 484–494. IEEE (2018)
19. Saito, S., Simon, T., Saragih, J., Joo, H.: Pifuhd: Multi-level pixel-aligned implicit function for high-resolution 3d human digitization. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 84–93 (2020)
20. Santos, D.A., Dawson, J.A., Matias, C.N., Rocha, P.M., Minderico, C.S., Allison, D.B., Sardinha, L.B., Silva, A.M.: Reference values for body composition and anthropometric measurements in athletes. *PloS one* **9**(5), e97846 (2014)
21. Svitov, D., Morerio, P., Agapito, L., Del Bue, A.: Haha: Highly articulated gaussian human avatars with textured mesh prior. *arXiv preprint arXiv:2404.01053* (2024)
22. Trujillo-Jiménez, M.A., Navarro, P., Pazos, B., Morales, L., Ramallo, V., Paschetta, C., De Azevedo, S., Ruderman, A., Pérez, O., Delrieux, C., et al.: Body2vec: 3d point cloud reconstruction for precise anthropometry with handheld devices. *Journal of Imaging* **6**(9), 94 (2020)
23. Tsoli, A., Loper, M., Black, M.J.: Model-based anthropometry: Predicting measurements from 3d human scans in multiple poses. In: IEEE winter conference on applications of computer vision. pp. 83–90. IEEE (2014)
24. Utkualp, N., Ercan, I., et al.: Anthropometric measurements usage in medical sciences. *BioMed research international* **2015** (2015)
25. Varol, G., Ceylan, D., Russell, B., Yang, J., Yumer, E., Laptev, I., Schmid, C.: Bodynet: Volumetric inference of 3d human body shapes. In: Proceedings of the European conference on computer vision (ECCV). pp. 20–36 (2018)
26. Wang, N., Zhang, Y., Li, Z., Fu, Y., Liu, W., Jiang, Y.G.: Pixel2mesh: Generating 3d mesh models from single rgb images. In: Proceedings of the European conference on computer vision (ECCV). pp. 52–67 (2018)
27. Weinzaepfel, P., Brégier, R., Combaluzier, H., Leroy, V., Rogez, G.: Dope: Distillation of part experts for whole-body 3d pose estimation in the wild. In: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXVI 16. pp. 380–397. Springer (2020)
28. Wen, J., Zhao, X., Ren, Z., Schwing, A.G., Wang, S.: Gomavatar: Efficient animatable human modeling from monocular video using gaussians-on-mesh. *arXiv preprint arXiv:2404.07991* (2024)
29. Xu, Y., Zhu, S.C., Tung, T.: Denserac: Joint 3d pose and shape estimation by dense render-and-compare. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 7760–7770 (2019)