

Jornadas de Automática

Sistema de Rehabilitación de Muñeca con Realidad Aumentada mediante Eliminación Visual del Robot utilizando inpainting

Benavides, Diego^a, Ciscal, Ana^{a,*}, de la Fuente, Eusebio^a, Fraile, Juan Carlos^a, Pérez Turiel, Javier^a

^a Instituto de las Tecnologías Avanzadas de la Producción, Escuela de Ingenierías Industriales, Universidad de Valladolid, Paseo Prado de la Magdalena 3-5, 47011, Valladolid, España.

To cite this article: Benavides, Diego, Ciscal, Ana, de la Fuente, Eusebio, Fraile, Juan Carlos, Pérez Turiel, Javier. 2025. Immersive AR-Based Wrist Rehabilitation with Visual Robotic Removal via Image Inpainting Jornadas de Automática, 46. <https://doi.org/10.17979/ja-cea.2025.46.12152>

Resumen

Este trabajo presenta un sistema innovador para la rehabilitación de la muñeca que combina un dispositivo robótico de asistencia con una plataforma de realidad aumentada inmersiva. Para mejorar la experiencia del paciente y favorecer los mecanismos de neuroplasticidad, se presenta un algoritmo de eliminación visual del robot mediante técnicas de segmentación y relleno de imagen (inpainting). El sistema integra un modelo de segmentación basado en UNet y un modelo de inpainting entrenado específicamente, logrando reconstruir la escena de forma realista tras eliminar el dispositivo robótico. Se ha evaluado el desempeño de distintos modelos de inpainting mediante métricas tradicionales y perceptuales, demostrando mejoras significativas en la calidad de las imágenes reconstruidas y en la inmersión del entorno de terapia. Estos resultados abren la puerta a una rehabilitación más eficaz y motivadora, combinando los beneficios funcionales de la robótica con el realismo de la realidad aumentada.

Palabras clave: Percepción y detección, Procesamiento y sistemas de imágenes biomédicas y médicas, Robótica, Tecnología asistencial e ingeniería de rehabilitación, Trabajo en entornos reales y virtuales.

Immersive AR-Based Wrist Rehabilitation with Visual Robotic Removal via Image Inpainting

Abstract

This work presents an innovative system for wrist rehabilitation that combines an assistive robotic device with an immersive augmented reality (AR) platform. To enhance the patient's experience and support neuroplasticity mechanisms, a visual robot removal algorithm is proposed using image segmentation and inpainting techniques. The system integrates a UNet-based segmentation model and a specifically trained inpainting model, effectively reconstructing the scene realistically after removing the robotic device. The performance of different inpainting models has been evaluated using both traditional and perceptual metrics, showing significant improvements in the quality of the reconstructed images and the immersive experience of the therapy environment. These results pave the way for more effective and engaging rehabilitation, combining the functional benefits of robotics with the realism of augmented reality.

Keywords: Assistive technology and rehabilitation engineering, Biomedical and medical image processing and systems, Perception and sensing, Robotics, Work in real and virtual environments

1. Introducción

En 2010, aproximadamente 1000 millones de personas en todo el mundo convivían con algún tipo de discapacidad moderada o severa, lo que suponía aproximadamente el 15 % de la población mundial (World Health Organization and World Bank, 2011). Extrapolando este porcentaje hasta nuestros días, el número de personas que sufre discapacidad severa o moderada asciende a unos 1200 millones. Solamente en España, más 4 millones de personas sufren algún tipo de discapacidad, de las cuales más del 10 % están relacionadas con el Daño Cerebral Adquirido (DCA) (INE, 2020).

El tratamiento del DCA requiere un equipo multidisciplinar compuesto por médicos rehabilitadores, fisioterapeutas, logopedas, neuropsicólogos y terapeutas ocupacionales, además de un seguimiento social a cargo de psicólogos, trabajadores sociales, educadores y cuidadores. La participación de tantos profesionales eleva los costes y limita la cantidad de pacientes que pueden beneficiarse de estos cuidados (Sánchez et al., 2022).

Concretamente, la rehabilitación es clave en la recuperación, ya que el 75 % de los pacientes que sufren un ACV presentan algún grado de paresia en las extremidades superiores, especialmente en manos y muñecas (Rathore et al., 2002). Una rehabilitación efectiva mejora los resultados funcionales y la calidad de vida, pero los enfoques tradicionales enfrentan desafíos como sus altos costes, la necesidad de una participación continua del fisioterapeuta y la pérdida de motivación del paciente. Por ello, es crucial desarrollar soluciones que reduzcan costes, mantengan la eficacia y aumenten la adherencia del paciente a la terapia.

Los robots de rehabilitación son una solución prometedora, ya que pueden asistir al movimiento sin intervención continua de un profesional sanitario, permitiendo tratar a varios pacientes simultáneamente y reduciendo sus costes (Polygerinos et al., 2015). Son especialmente útiles en las primeras etapas de recuperación, cuando los pacientes requieren una movilización pasiva. Al automatizar estas tareas, disminuyen la carga de los terapeutas y permiten sesiones más frecuentes e intensas, mejorando la recuperación. Además, se adaptan al progreso del paciente, ofreciendo una terapia personalizada.

Otro enfoque transformador en la rehabilitación es la incorporación de videojuegos, que ha demostrado mejorar la motivación y el compromiso del paciente en comparación a las terapias tradicionales (Masmoudi et al., 2024). El uso de *Serious Games* permiten la inclusión de desafíos, recompensas y seguimiento del progreso, personalizando los objetivos a cada paciente y convirtiendo los ejercicios repetitivos en una experiencia atractiva. Este aumento en la motivación se ha relacionado con una mejor adherencia al tratamiento y mejores resultados en la rehabilitación. Además, incluyendo tareas de memorización o tomas de decisiones, este tipo de terapia se han demostrado útil también en el aspecto cognitivo (Rego et al., 2017).

Para implementar estos videojuegos, habitualmente se hace uso de la realidad virtual (VR), que permite crear entornos inmersivos con los que el paciente puede interactuar. Sin embargo, la realidad aumentada (AR), que combina elementos digitales y reales, permiten que los

pacientes vean e interactúen con su propio cuerpo dentro del entorno virtual, que deriva en una experiencia más natural y realista, y finalmente en mayores beneficios terapéuticos. Por desgracia, no es común combinar terapias basadas en AR con robots, ya que la presencia del dispositivo resta inmersividad a la terapia y puede afectar a los mecanismos de neuroplasticidad relacionados con la recuperación del paciente (Hao et al., 2022).

Este trabajo propone un sistema innovador que integra la AR con un dispositivo robótico de asistencia para la rehabilitación de la muñeca, combinando las ventajas motivacionales e inmersivas de la AR con los beneficios funcionales de la robótica. Sin embargo, la presencia del robot en el campo de visión del paciente puede afectar la inmersión y reducir la sensación de movimiento autónomo. Para abordar este problema, se ha desarrollado un algoritmo que segmenta y elimina los componentes robóticos de la imagen mediante un proceso de *image inpainting*, reconstruyéndola para mostrar únicamente la extremidad del paciente y mejorar la experiencia inmersiva. Este algoritmo consta de una fase de segmentación del robot y una fase de rellenado realista de la imagen, siendo esta última el foco principal del desarrollo.

Para realizar tareas de *inpainting* se han propuesto varios tipos de modelos en los últimos años, que pueden clasificarse según tres características principales, conocidas como el trilema del aprendizaje generativo (Xiao et al., 2021): velocidad de generación, calidad de imagen y diversidad y cobertura de generación.

Entre los modelos que proporcionan una buena calidad de imagen los basados en *Denosing Diffusion Probabilistic Models* (DDPM) (Ho et al., 2020) han logrado resultados impresionantes en tareas de generación de imágenes en términos de calidad y diversidad de generación (Lugmayr et al., 2022). Estos algoritmos han sido diseñados para estimar y eliminar el componente de ruido en una imagen con el fin de eliminarlo. De manera simplificada, el entrenamiento de estos modelos consiste en agregar ruido gradualmente a la imagen original x_0 , generando estados intermedios en cada paso de tiempo τ (x_{T-1} , x_{T-2} , ...) hasta una imagen completamente ruidosa x_T . El modelo aprende a estimar el ruido en cada imagen y lo elimina de manera iterativa para reconstruir la imagen previa, hasta finalmente alcanzar la imagen original. Sin embargo, este proceso iterativo es demasiado lento, lo que los hace inadecuado para su uso en tiempo real.

Por otro lado, los modelos basados en *Generative Adversarial Networks* (GANs) (Goodfellow et al., 2014), más tradicionales, combinan una buena calidad de imagen con una velocidad de generación conveniente, lo que potencialmente les permite producir imágenes en tiempo real. Estos algoritmos constan de dos modelos: un generador, cuya función es crear nuevas partes de la imagen que coincidan con la imagen original, y un discriminador, encargado de distinguir las imágenes reconstruidas de las imágenes originales. Ambos modelos compiten entre sí, de manera que el generador intenta engañar al discriminador generando imágenes más realistas, mientras que el discriminador mejora su capacidad para diferenciar las imágenes reconstruidas de las reales. No obstante, estas redes fallan al generar imágenes con diferencias sustanciales respecto a las imágenes de entrenamiento, una tarea en la que los DDPMs sí destacan



Figura 1: Plataforma de realidad aumentada M3Display (Cisnal et al., 2023).

(Lugmayr et al., 2022). Para nuestros propósitos, el escenario de la terapia estará controlado, por lo que la diversidad de imágenes podría no ser un factor crítico si el modelo se entrena con una cantidad suficiente de imágenes.

En resumen, este trabajo presenta un sistema que permite a los pacientes realizar ejercicios de rehabilitación con asistencia robótica mientras mantienen un entorno de AR altamente inmersivo. Al garantizar que solo la extremidad del paciente sea visible, el sistema propuesto busca mejorar la motivación del paciente, facilitar la neuroplasticidad y maximizar los resultados de la rehabilitación.

2. Materiales y métodos

La Figura 1 muestra el diagrama conceptual del proceso completo de tratamiento de la imagen. Este proceso se inicia con la captura de la imagen del miembro parético utilizando la plataforma de rehabilitación. A continuación, la imagen es procesada por el módulo de segmentación, encargado de identificar y segmentar la plataforma robótica para su eliminación del campo visual. Finalmente, la imagen resultante es introducida en el módulo de *inpainting*, cuya función es reconstruir de manera realista la región previamente eliminada, restaurando la apariencia original de la escena.

2.1. Plataforma de rehabilitación

La plataforma de realidad aumentada M3Display (Figura 2), está formada por un monitor de 27 pulgadas colocado perpendicularmente a la línea de visión del paciente, situado entre sus ojos y sus manos. En el dorso del monitor se encuentra un espejo de 380 x 550 mm que permite la reflexión de las imágenes capturadas por una cámara (1080p/30 fps, 3 MP, 78° dFoV). La cámara está montada en un soporte que permite capturar las imágenes con un ángulo específico de 39° respecto al espejo para garantizar una representación sin distorsiones de las manos en el monitor. Esto permite que el sistema cree la ilusión de un brazo que se extiende de manera continua a través del monitor. Además, la plataforma es versátil y puede utilizarse de forma independiente para la rehabilitación del miembro superior, especialmente con realidad aumentada (Cisnal et al., 2023).

Este trabajo propone el uso conjunto de la plataforma M3Display y el dispositivo de rehabilitación de muñeca M3Rob (Alonso-Linaje et al., 2024), diseñados para complementar sus funcionalidades: mientras que M3Rob asiste al movimiento del miembro parético mediante un mecanismo de tres articulaciones activas en serie — pronación/supinación (PS), flexión/extensión (FE) y radial/ulnar (RU)—, la plataforma M3Display proporciona el entorno de realidad aumentada inmersiva. El efector final del robot está equipado con un mango que incorpora un sensor de fuerza, permitiendo modos de terapia activa. El dispositivo (Figura 3) admite un rango de movimiento de

180° para PS, 135° para FE, y 110° para RU, garantizando el rango completo de movimiento de la muñeca.

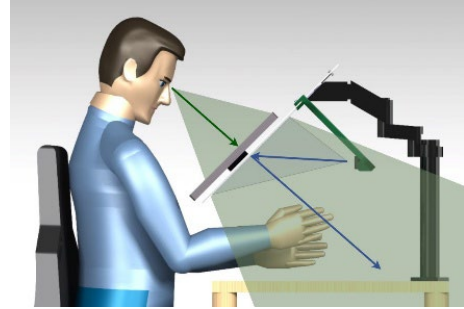


Figura 2: Plataforma de realidad aumentada M3Display (Cisnal et al., 2023).



Figura 3: Imagen virtual del dispositivo M3Rob (Alonso-Linaje et al., 2024).

2.2. Módulo de segmentación

El modelo de segmentación empleado es MobileNet, basado en una arquitectura UNet (Ronneberger et al., 2015). Esta arquitectura consta de un *encoder*, que extrae características reduciendo progresivamente la resolución espacial, y un *decoder*, que reconstruye la máscara de segmentación mediante la combinación de las salidas intermedias del encoder y las capas previas del decoder, incrementando la resolución en cada etapa. La elección de este modelo se fundamenta en un estudio previo (Brizuela et al., 2022), donde se evaluaron tres variantes de UNet (MobileNet, ResNet50 y EfficientNet-B4) sobre imágenes de 1280x704 píxeles, seleccionando la que obtuvo mejor rendimiento.

2.3. Módulo de inpainting

Tras la eliminación del robot de la imagen mediante el módulo de segmentación, el módulo de *inpainting* reconstruye la imagen de forma realista utilizando un modelo basado en GANs, entrenado con un conjunto de datos específico para este problema.

2.3.1. Conjunto de datos

El conjunto de datos cuenta con 10367 imágenes capturadas en entornos similares a aquellos donde se llevarán a cabo las terapias. Dado que el objetivo del modelo es reconstruir imágenes sin el dispositivo mecatrónico, el cual debe ser eliminado completamente por el modelo de

segmentación, en ellas el robot no está presente y aparece únicamente el brazo del usuario. Posteriormente se aplicó una máscara con la misma forma que el robot, de modo que el modelo entrenado con nuestro conjunto de datos debería ser capaz de reconstruir imágenes incompletas en las que el área desconocida tiene una forma robótica tras la segmentación, de la misma manera en que puede reconstruir una imagen de entrenamiento con una máscara artificial con forma de robot (Figura 4).

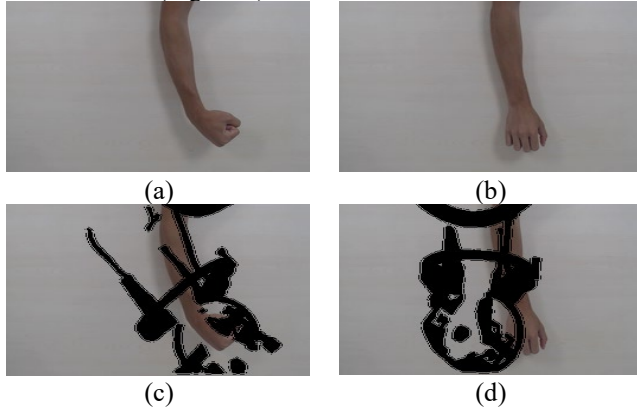


Figura 4: Muestras del ejemplo del dataset empleado: imágenes originales (a,b) y sus correspondientes máscaras asignadas aleatoriamente (b, d)

Las máscaras robóticas se obtuvieron tanto del conjunto de entrenamiento del módulo de segmentación como de sus predicciones para aumentar la variabilidad de este tipo de imágenes. Todas las imágenes y máscaras fueron reescaladas a 256x256 píxeles con el objetivo de mantener una alta velocidad de inferencia y reducir el consumo de memoria. Por ello el tamaño de *batch* que se utilizó fue de 4.

2.3.2. Modelo Inpainting

Se emplea el modelo LaMa (Suvorov et al., 2022) que utiliza estructuras llamadas *Fast Fourier Convolutions* (FFCs), que procesan imágenes en paralelo mediante convoluciones tradicionales y convoluciones en el dominio de la frecuencia a través de la *Fast Fourier Transform* (FFT). El uso de FFCs permite procesar campos receptivos más grandes con menos parámetros. En el dominio de la frecuencia, los *kernels* de convolución afectan toda la imagen desde las primeras capas, capturando detalles finos antes de que sean abstraídos en capas más profundas, reduciendo la cantidad de parámetros necesarios y haciendo la red más compacta y rápida. Además, las FFCs muestran un rendimiento excepcional en el procesamiento de estructuras periódicas o repetitivas, patrones comunes en objetos artificiales. Finalmente, el modelo LaMa ofrece una calidad de generación de imágenes comparable a la de los Denoising Diffusion Probabilistic Models (DDPMs), dado que la mayoría de estos modelos utilizan lo como referencia para sus comparaciones (Rombach et al., 2022), (Corneanu et al., 2024), (Chen et al., 2024).

Específicamente, se utilizó una modificación preentrenada de LaMa llamada *BigLaMa*. La principal diferencia es que el generador en *BigLaMa* contiene 18 bloques residuales, en comparación con los 9 del modelo original. Además, este modelo fue entrenado con 4,5 millones de imágenes del *Places Challenge* utilizando un tamaño de *batch* de 120.

A partir de este modelo preentrenado, se ajustaron tres modelos adicionales con el conjunto de datos de rehabilitación: ***BigLaMa Random***, entrenado únicamente con máscaras aleatorias, como el modelo original *BigLaMa*; ***BigLaMa Custom***, entrenado exclusivamente con máscaras robóticas para garantizar que el modelo aprenda la forma del robot; y ***BigLaMa Mixed***, entrenado con una combinación de máscaras aleatorias (75 %) y máscaras personalizadas (25 %). Todos los modelos ajustados fueron entrenados durante 200 épocas con un tamaño de *batch* de 4, tomando aproximadamente 67 horas en una GPU NVIDIA RTX 3060.

2.4. Métricas de rendimiento

La evaluación del modelo se realiza mediante cinco métricas:

- **Peak Signal to Noise Ratio (PSNR):** Evalúa la similitud entre imágenes, donde valores más altos indican una mayor similitud (Horé and Ziou, 2010).
- **Structural Similarity Index Measure (SSIM):** Evalúa la similitud entre imágenes relacionándola estrechamente con la percepción visual humana (Wang et al., 2004). Cuantifica la distorsión de la imagen a través de tres componentes independientes: distorsión de luminancia, distorsión de contraste y pérdida de correlación estructural.
- **Fréchet Inception Distance (FID):** Evalúa modelos generativos comparando imágenes generadas y reales aprovechando modelos de redes convolucionales (CNNs) para alinearse mejor con la percepción humana (Heusel et al., 2017). Se basa en el *Inception Score* (Salimans et al., 2016), que utiliza una CNN *Inception* para extraer características a múltiples escalas (Szegedy et al., 2014). El FID mejora este índice calculando la distancia de Fréchet de las distribuciones de datos reales y generados con sus propiedades estadísticas (media y covarianza), asumiendo una aproximación Gaussiana, donde valores más bajos indican mayor similitud.
- **Learned Perceptual Image Patch Similarity (LPIPS):** Similar al FID, LPIPS compara no solo las salidas finales del clasificador, sino también las activaciones intermedias en múltiples capas de un modelo preentrenado, como VGG16 (Zhang et al., 2018). La similitud se calcula mediante parches de las imágenes reales y generadas, cuyas activaciones se normalizan, se ponderan con los pesos específicos de cada capa y se calcula su norma para medir la distancia. Este enfoque captura diferencias perceptuales en distintos niveles de abstracción y se correlaciona estrechamente con la percepción humana, donde distancias menores indican mayor similitud.
- **F1-score:** Esta métrica adapta el *F1-Score* reemplazando *precision* y *recall* por FID y LPIPS (Van Rijsbergen, 1979). La media armónica penaliza los desequilibrios entre estas métricas, proporcionando una evaluación tanto de la similitud perceptual como de la calidad estadística de la imagen generada.

3. Resultados

Los modelos han sido evaluados desde un enfoque cuantitativo (Tabla 1) y cualitativo (Figura 5), con el objetivo de valorar su rendimiento tanto de forma objetiva como desde la perspectiva del usuario final.

La Tabla 1 presenta los resultados para las distintas

Tabla 1: Resultados cualitativos de los distintos modelos entrenados en comparación con el modelo preentrenado *BigLaMa Default*.

	<i>BigLaMa Default</i>	<i>BigLaMa Random</i>	<i>BigLaMa Mixed</i>	<i>BigLaMa Custom</i>
PSNR (dB)	42,659	41,435 (↓ 2,68 %)	42,012 (↓ 1,52 %)	38,504 (↓ 9,74 %)
SSIM	0,9803	0,9815 (↑ 0,12 %)	0,9791 (↓ 0,12 %)	0,9749 (↓ 0,55 %)
FID	0,4793	0,3016 (↓ 37,1 %)	0,3356 (↓ 30,0 %)	0,4595 (↓ 4,10 %)
LPIPS	0,1151	0,1031 (↓ 10,4 %)	0,0815 (↓ 29,2 %)	0,1546 (↑ 34,3 %)
F1 _{FID/LPIPS}	0,6552	0,7848 (↑ 19,8 %)	0,7706 (↑ 17,6 %)	0,6590 (↑ 0,58 %)

Entre paréntesis se indica el porcentaje de mejora (en verde) o empeoramiento (en rojo) de cada métrica respecto al modelo *BigLaMa Default*. Nótese que un mayor valor en PSNR, SSIM y F1_{FID/LPIPS} refleja una mejor calidad, mientras que en FID y LPIPS una menor puntuación indica mejor rendimiento.

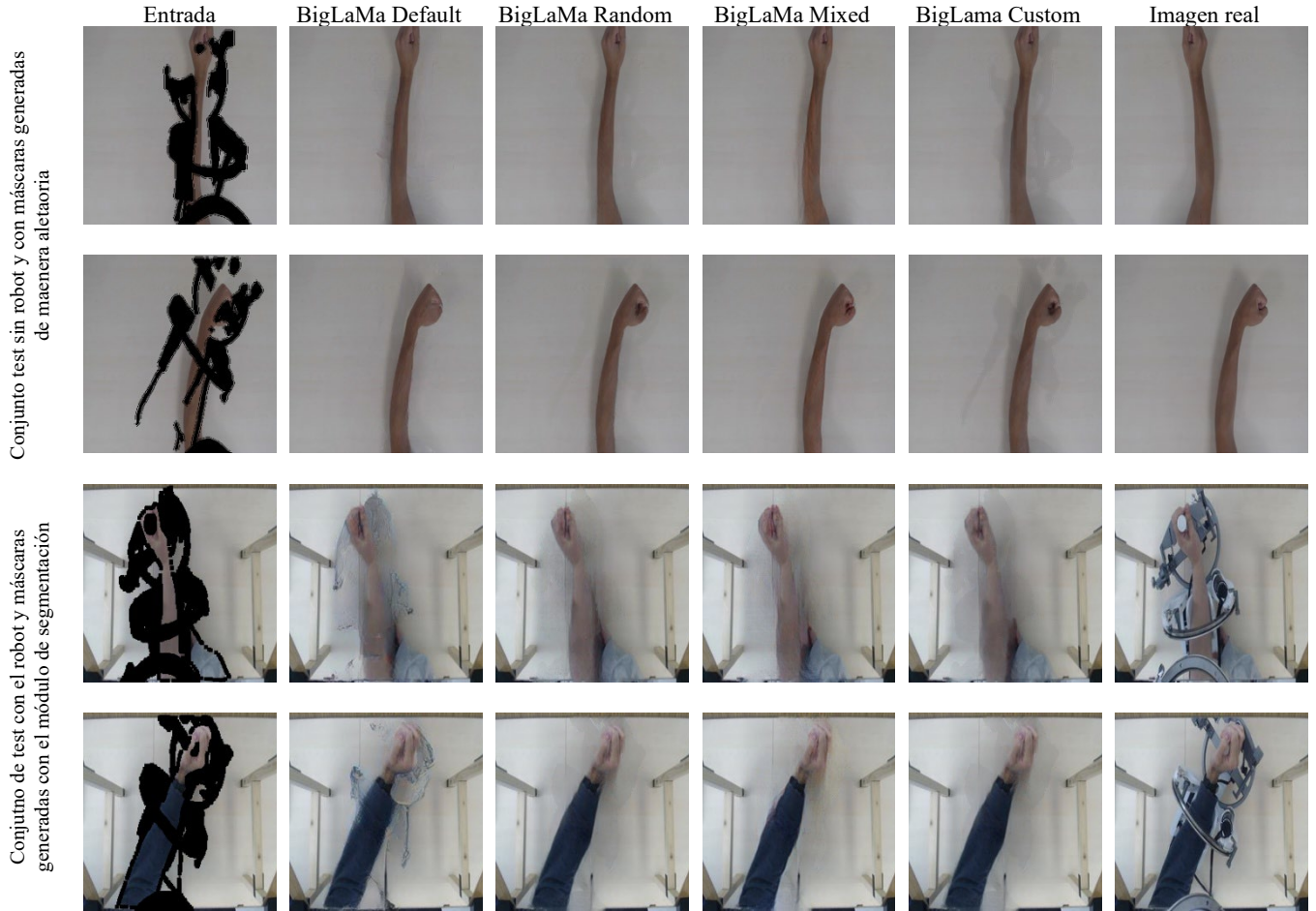


Figura 5: Ejemplo de reconstrucciones de los distintos modelos en el conjunto de test sin robot y máscaras aleatorias (filas 1 y 2) y con el robot y máscaras generadas con el módulo de segmentación (filas 3 y 4). Todas las imágenes se redimensionaron a 256x256 píxeles para su procesamiento y visualización.

métricas comparándolos con el modelo *BigLaMa* sin reentrenar (*BigLaMa Default*). Para aquellas más tradicionales (PSNR y SSIM), el reentrenamiento apenas aporta mejoras e incluso empeora ligeramente los valores. En cambio, para las métricas basadas en redes convolucionales, más correlacionadas con la percepción visual humana, los tres modelos reentrenados logran un mejor desempeño respecto al modelo original. La Figura 5 ilustra cuatro ejemplos de reconstrucciones generadas por los modelos evaluados en imágenes del conjunto de test de entrenamiento y en imágenes reales, permitiendo apreciar visualmente las diferencias entre ellos. En términos de rendimiento, se alcanzó una tasa de 15,76 FPS con imágenes de 256x256 píxeles en una GPU RTX3060.

3. Discusión y conclusiones

El modelo *Default* mejora significativamente al reentrenarse con imágenes similares a las usadas en terapia, mostrando un mejor rendimiento en las métricas basadas en CNNs frente a métricas tradicionales. Los resultados visuales también evidencian mejoras en los tres modelos reentrenados. *BigLaMa Random* y *BigLaMa Mixed* consiguen reconstrucciones más realistas, mientras que *BigLaMa Custom* presenta una ligera sombra en la zona segmentada, lo que explica su menor rendimiento cuantitativo, posiblemente debido a sobreajuste por el número limitado de máscaras empleadas en su entrenamiento. No obstante, *BigLaMa Custom* genera bordes más definidos en el brazo al adaptarse mejor al tipo de máscara utilizada. Esto sugiere que su rendimiento pondría mejorar con un mayor número de máscaras e imágenes. Todos los modelos funcionan mejor en

el conjunto de test sin robot y con máscaras aleatorias (más similar al conjunto de entrenamiento) que en el conjunto con robot y máscaras generadas mediante segmentación (que presenta mayor variabilidad).

En cuanto al rendimiento, aunque es limitado para estas aplicaciones en tiempo real, es previsible que tanto la tasa de refresco como el tamaño de imagen aumenten utilizando una GPU de mayor capacidad o configuraciones en paralelo.

En conclusión, el modelo desarrollado representa un punto de partida prometedor como alternativa para integrar dispositivos robóticos en terapias de rehabilitación basadas en realidad aumentada, permitiendo su eliminación visual y aumentando la sensación de inmersión percibida por el paciente en tiempo real. Aunque aún presenta limitaciones en rendimiento y calidad de imagen, este enfoque de *inpainting* ha demostrado su potencial para procesar este tipo de imágenes y ofrecer resultados funcionales.

Agradecimientos

Este trabajo ha sido financiado por el Ministerio de Ciencia e Innovación de España a través de los proyectos RTC2019-007350-1 y PID2022-138206OB-C33. También se contó con el apoyo de la empresa TICCYL Digital S.L.U.

Referencias

- Alonso-Linaje, G., Císnal, A., Fraile, J. C., Turiel, J. P., 2024. Design and analysis of the m3rob: A robotic platform for wrist and hand rehabilitation. *IEEE Access* 12, 30472–30481. DOI: 10.1109/ACCESS.2024.3369908
- Brizuela, G. S., Rodríguez, D. S., de la Fuente López, E., Fraile, J. C., Turiel, J. P., 2022. Segmentación semántica de elementos mecánicos en entornos aumentados de rehabilitación. In: *Jornadas de Robótica, Educación y Bioingeniería*.
- Chen, T., Zhang, J., Hong, Y., Zhang, Y., Zhang, L., 2024. Assessing image inpainting via re-inpainting self-consistency evaluation. In: *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management. CIKM '24*. Association for Computing Machinery, New York, NY, USA, p. 291–300. DOI: 10.1145/3627673.3679693
- Císnal, A., Alonso-Linaje, G., Veganzones, M., and Juan Carlos Fraile, J. P. T., 2023. M3display: Sistema de realidad aumentada para la rehabilitación de la función motora del miembro superior. In: *XLI Congreso Anual de la Sociedad Española de Ingeniería Biomédica*. URL: <http://hdl.handle.net/10317/13479>
- Comeanu, C., Gadde, R., Martínez, A. M., January 2024. Latentpaint: Image inpainting in latent space with diffusion models. In: *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*. pp. 4334–4343.
- Fernández-Sánchez, M., Aza-Hernández, A., Verdugo-Alonso, M., 4 2022. Models of public care for the population with acquired brain injury in Spain: a study of the situation by spanish autonomous communities.
- Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y., 6 2014. Generative adversarial networks. *Science Robotics* 3, 2672–2680. DOI: 10.48550/arXiv.1406.2661
- Hao, J., Xie, H., Harp, K., Chen, Z., Siu, K.-C., 2022. Effects of virtual reality intervention on neural plasticity in stroke rehabilitation: A systematic review. *Archives of Physical Medicine and Rehabilitation* 103 (3), 523–541. DOI: 10.1016/j.apmr.2021.06.024
- Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., Hochreiter, S., 6 2017. Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Advances in Neural Information Processing Systems* 2017- December, 6627–6638. DOI: 10.18034/ajase.v8i1.9
- Ho, J., Jain, A., Abbeel, P., 6 2020. Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems* 2020-December. DOI: <https://doi.org/10.48550/arXiv.2006.11239>
- Horé, A., Ziou, D., 2010. Image quality metrics: PSNR vs. SSIM. In: *2010 20th International Conference on Pattern Recognition*. pp. 2366–2369. DOI: 10.1109/ICPR.2010.579
- INE, 2020. Encuesta de discapacidad, autonomía personal y situaciones de dependencia. 2020. <https://www.ine.es/jaxi/Datos.htm?tpx=51614>.
- Lugmayr, A., Danelljan, M., Romero, A., Yu, F., Timofte, R., Van Gool, L., June 2022. Repaint: Inpainting using denoising diffusion probabilistic models. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. pp. 11461–11471.
- Masmoudi, M., Zenati, N., Izountar, Y., Benbelkacem, S., Haicheur, W., Guerroudjii, M. A., Oulefki, A., Hamitouche, C., 2 2024. Assessing the effectiveness of virtual reality serious games in post-stroke rehabilitation: a novel evaluation method. *Multimedia Tools and Applications* 83, 36175–36202. DOI: 10.1007/s11042-023-17980-5
- Polygerinos, P., Wang, Z., Galloway, K. C., Wood, R. J., Walsh, C. J., 2015. Soft robotic glove for combined assistance and at-home rehabilitation. *Robotics and Autonomous Systems* 73, 135–143, *wearable Robotics*. DOI: 10.1016/j.robot.2014.08.014
- Rathore, S. S., Hinn, A. R., Cooper, L. S., Tyroler, H. A., Rosamond, W. D., 2002. Characterization of incident stroke signs and symptoms. *Stroke* 33 (11), 2718–2721. DOI: 10.1161/01.STR.0000035286.87503.31
- Rego, P. A., Rocha, R., Faria, B. M., Reis, L. P., Moreira, P. M., 1 2017. A serious games platform for cognitive rehabilitation with preliminary evaluation. *Journal of Medical Systems* 41, 1–15. DOI: 10.1007/s10916-016-0656-5
- Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B., Jun. 2022. High-Resolution Image Synthesis with Latent Diffusion Models. In: *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE Computer Society, Los Alamitos, CA, USA, pp. 10674–10685. DOI: 10.1109/CVPR52688.2022.01042
- Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. In: *Navab, N., Hornegger, J., Wells W. M., Frangi, A. F. (Eds.), Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. Springer International Publishing, Cham, pp. 234–241.
- Salimans, T., Goodfellow, I., Zaremba, W., Cheung, V., Radford, A., Chen, X., 6 2016. Improved techniques for training gans. *Advances in Neural Information Processing Systems*, 2234–2242. DOI: <https://doi.org/10.48550/arXiv.1606.03498>
- Suvorov, R., Logacheva, E., Mashikhin, A., Remizova, A., Ashukha, A., Silvestrov, A., Kong, N., Goka, H., Park, K., Lempitsky, V., 2022. Resolution-robust large mask inpainting with fourier convolutions. In: *2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*. pp. 3172–3182. DOI: 10.1109/WACV51458.2022.00323
- Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan D., Vanhoucke, V., Rabinovich, A., 9 2014. Going deeper with convolutions. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* 07-12-June-2015, 1–9. DOI: 10.1109/CVPR.2015.7298594
- Van Rijsbergen, C. J., 1979. Information retrieval. 2nd. newton, ma. Wang, Z., Bovik, A., Sheikh, H., Simoncelli, E., 2004. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing* 13 (4), 600–612. DOI: 10.1109/TIP.2003.819861
- World Health Organization, World Bank, 2011. World Report on Disability 2011. World Health Organization, Geneva, nLM classification: HV 1553. URL: <https://www.who.int/publications/i/item/9789241564182>
- Xiao, Z., Kreis, K., Vahdat, A., 12 2021. Tackling the generative learning trilemma with denoising diffusion gans. *ICLR 2022 - 10th International Conference on Learning Representations*. DOI: <https://doi.org/10.48550/arXiv.2112.07804>
- Zhang, R., Isola, P., Efros, A. A., Shechtman, E., Wang, O., 1 2018. The unreasonable effectiveness of deep features as a perceptual metric. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 586–595. DOI: 10.1109/CVPR.2018.