

**PONTIFICIA UNIVERSIDAD
CATÓLICA DEL PERÚ**
Escuela de Posgrado



Evaluación de método para la detección automática de puntos de referencia (landmark detection) en imágenes en dos dimensiones de huellas plantares para el diseño de una plantilla ortopédica

Trabajo de investigación para obtener el grado académico de Maestro en Informática con mención en Ciencias de la Computación que presenta:

Gustavo Miguel Donayre Gamboa

Asesor:

Pablo Alejandro Fonseca Arroyo

Lima, 2024

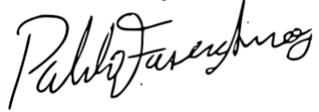
Informe de Similitud

Yo, Pablo Alejandro FONSECA ARROYO, docente de la Escuela de Posgrado de la Pontificia Universidad Católica del Perú, asesor de el trabajo de investigación titulado *Evaluación de método para la detección automática de puntos de referencia (landmark detection) en imágenes en dos dimensiones de huellas plantares para el diseño de una plantilla ortopédica* de el autor Gustavo Miguel DONAYRE GAMBOA, dejo constancia de lo siguiente:

- El mencionado documento tiene un índice de puntuación de similitud de 11%. Así lo consigna el reporte de similitud emitido por el software *Turnitin* el 11/04/2024.
- He revisado con detalle dicho reporte y la tesis de investigación, y no se advierte indicios de plagio.
- Las citas a otros autores y sus respectivas referencias cumplen con las pautas académicas.

Lugar y fecha:

San Miguel, 11 de Abril de 2024.

Apellidos y nombres del asesor / de la asesora: Fonseca Arroyo, Pablo Alejandro	
DNI: 44695174	Firma 
ORCID: 0000-0002-0208-2842	

RESUMEN

El presente trabajo de investigación evalúa la técnica de regresión de mapas de calor (heatmap regression - HR) para la detección automática de puntos de referencia (landmark detection) en imágenes médicas, específicamente en las imágenes de huellas plantares en dos dimensiones. El estudio se basa en la regresión de mapas de calor con aprendizaje profundo, una técnica que ha demostrado ser efectiva en la detección de puntos en rostros y en la estimación de la pose humana. Se propone un método automático para la detección de 8 puntos en las imágenes digitalizadas de huellas plantares que servirán de referencia para el diseño base de una plantilla ortopédica bidimensional, buscando así mejorar el proceso de fabricación de plantillas ortopédicas, que actualmente se realiza de forma manual y artesanal en la mayoría de los países de América Latina. La detección automática de estos puntos de referencia en las huellas plantares tiene el potencial de agilizar este proceso y mejorar la precisión de las plantillas.

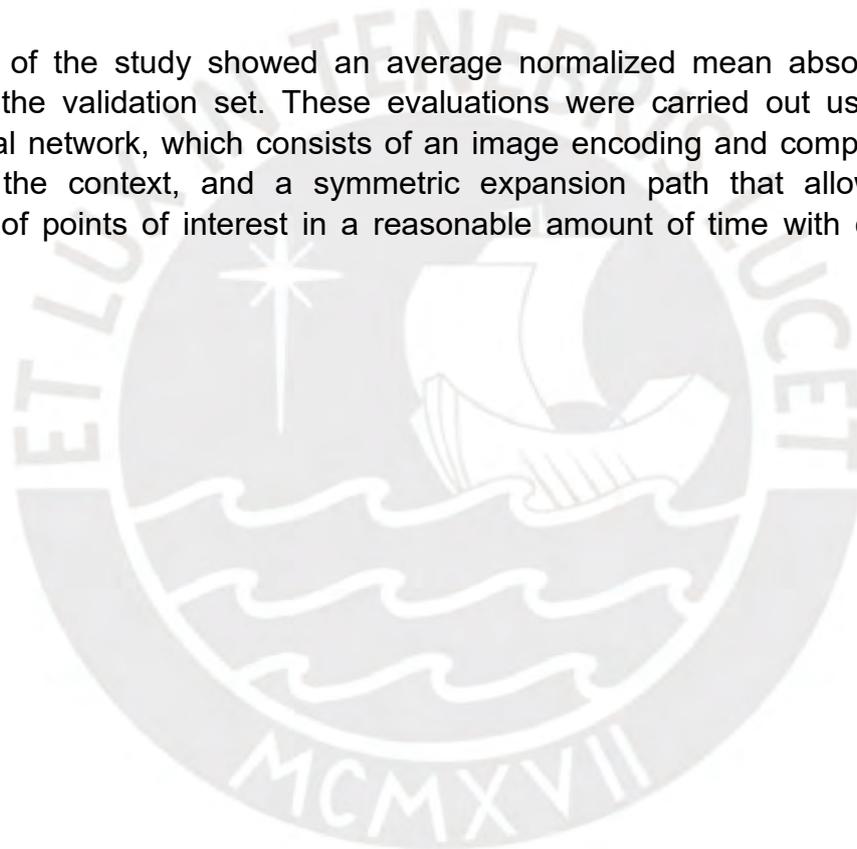
Los resultados del estudio mostraron un error absoluto promedio normalizado de 0.01017 en el conjunto de validación. Estas evaluaciones se llevaron a cabo utilizando una red convolucional U-Net, la cual consta de una ruta de codificación y compresión de imágenes para capturar el contexto, y una ruta de expansión simétrica que permite una localización precisa de puntos de interés en un tiempo razonable gracias al uso de los procesadores GPU actuales.

Palabras clave — Machine Learning, Deep Learning, Heatmap regression, footprint.

ABSTRACT

This paper evaluates the heatmap regression (HR) technique for landmark detection in medical images, specifically in two-dimensional footprint images. The study is based on heatmap regression with deep learning, a technique that has proven to be effective in face landmark detection and human pose estimation. We propose the evaluation of an automatic method for the detection of 8 points in the digitized images of plantar footprints that will serve as a reference for the base design of a two-dimensional orthopedic insole, thus seeking to improve the orthopedic insole manufacturing process, which is currently handmade and handcrafted in most Latin American countries. The automatic detection of reference points in the plantar footprints would speed up this process and improve the accuracy of the insoles.

The results of the study showed an average normalized mean absolute error of 0.01017 in the validation set. These evaluations were carried out using a U-Net convolutional network, which consists of an image encoding and compression path to capture the context, and a symmetric expansion path that allows accurate localization of points of interest in a reasonable amount of time with current GPU processors.



ÍNDICE DE CONTENIDO

RESUMEN.....	i
ABSTRACT.....	ii
ÍNDICE DE CONTENIDO	iii
ÍNDICE DE TABLAS	iv
ÍNDICE DE FIGURAS.....	v
SECCIÓN I.....	1
INTRODUCCIÓN.....	1
SECCIÓN II	3
TRABAJOS RELACIONADOS	3
SECCIÓN III	4
MÉTODOS	4
SECCIÓN IV	13
RESULTADOS	13
SECCIÓN V.....	14
CONCLUSIÓN Y DISCUSIÓN.....	14
TRABAJOS FUTUROS	15
REFERENCIAS BIBLIOGRÁFICAS	16

ÍNDICE DE TABLAS

Tabla 1 Resultado de los modelos propuestos sobre el conjunto de imágenes de validación.	13
---	----



ÍNDICE DE FIGURAS

Figura 1: Ejemplo de toma de imagen de huella plantar	1
Figura 2: Medidas y puntos de referencia de la huella plantar	2
Figura 3: Proceso para la detección de 8 puntos de interés	4
Figura 4: Izquierda - Imagen original. Derecha - Copia espejo huella plantar	6
Figura 5: Anotación huellas plantares en VGG Image Annotator	7
Figura 6: Inversa de la imagen de huella plantar	8
Figura 7: Imágenes de 333x256 px con 8 anotaciones y mapas de calor por cada anotación.....	9
Figura 8: Estructura de la red U-Net con base en una ResNet50 pre entrenada en la parte de la codificación (izquierda de la U-Net).....	11
Figura 9: Resultados de la red U-Net (objetivo, predicción, mapa de calor de la predicción).....	12
Figura 10: Entrenamiento por 26 épocas de la red U-Net (ResNet50) con imágenes de 333x256 px.....	12
Figura 11: Error absoluto promedio normalizado - NMAE.....	14



SECCIÓN I

INTRODUCCIÓN

En Perú, así como en la mayoría de los países de América Latina, el proceso de fabricación de plantillas ortopédicas para pacientes con diversas patologías, incluida la diabetes, se realiza mayoritariamente de forma manual y artesanal [1], como se puede observar en la fig. 1. Para poder realizar el estudio previo a la fabricación de las plantillas, se recopila información sobre la forma y las presiones que ejerce el pie del paciente, para lo cual se utiliza el podómetro. Los podómetros se pueden clasificar en dos grandes grupos: los cualitativos y los cuantitativos. En este trabajo, no se abordan los podómetros cuantitativos (con sensores electrónicos u otros), los escáneres láser en tres dimensiones y las mallas en tres dimensiones a partir de fotografías.



Figura 1: Ejemplo de toma de imagen de huella plantar

Los podómetros cualitativos producen una impresión de la huella plantar en una hoja de papel. Sin embargo, esta hoja impresa presenta desafíos logísticos, como la necesidad de ser transportada al lugar de fabricación de la plantilla, así como problemas relacionados con la falta de técnicos especializados para la confección de la plantilla destinada al paciente. Para elaborar la plantilla ortopédica, es necesario detectar puntos específicos y tomar medidas sobre la impresión de la huella plantar, como lo señala Kimura [2]. Estos puntos y medidas necesarios se ilustran en la fig. 2.

Los puntos de interés semánticos son conjuntos de puntos o píxeles en imágenes que proporcionan información sobre la estructura o forma, como rostros, manos, cuerpos humanos y objetos cotidianos. Por lo tanto, la identificación de estos puntos de interés semánticos es crucial para diversas aplicaciones en el campo de la visión por computadora [3].

Este trabajo propone evaluar un método automático para la detección de 8 puntos en las imágenes digitalizadas de huellas plantares, que servirán como referencia para el diseño base de una plantilla ortopédica en dos dimensiones.

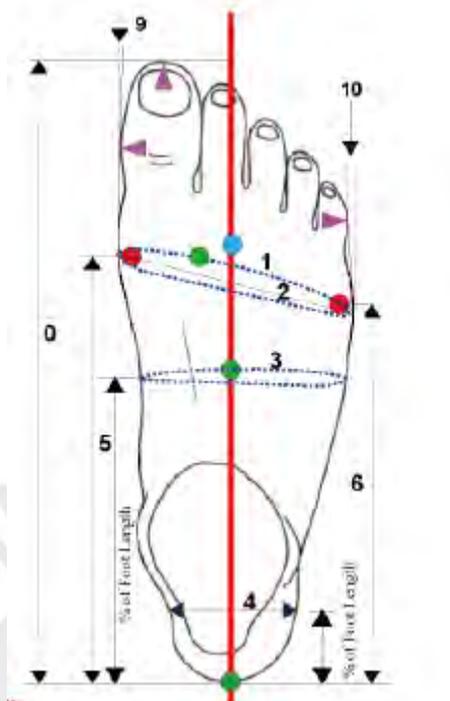


Figura 2: Medidas y puntos de referencia de la huella plantar

La regresión de mapas de calor es un método ampliamente utilizado para la localización de puntos de interés basándose en redes neuronales profundas [5]. Este método busca predecir un mapa de calor en lugar de una coordenada numérica mediante una capa totalmente conectada, donde el punto máximo de activación corresponde al punto de interés semántico de la imagen de entrada [3].

Este artículo se estructura en varias secciones: la Sección II aborda los trabajos relacionados, la Sección III describe los métodos utilizados para la evaluación del modelo empleado, la Sección IV analiza los resultados obtenidos, la Sección V presenta las conclusiones y la Sección VI señala los trabajos futuros.

SECCIÓN II

TRABAJOS RELACIONADOS

En los últimos años, las redes neuronales convolucionales [4] han tenido mucho éxito en superarse continuamente en diversas tareas de reconocimiento visual. No obstante, su éxito es limitado por el tamaño de los conjuntos de datos de entrenamiento y el tamaño de las redes neuronales resultantes. El uso típico de las redes neuronales convolucionales son las tareas de clasificación, donde la salida de la red es una etiqueta de una clase a la que pertenece la imagen; sin embargo, en muchos otros casos, especialmente en el procesamiento de imágenes médicas, la salida debe incluir la información de localización, por ejemplo, indicar qué etiqueta de clase debe estar asignada a cada píxel de una imagen radiográfica.

Para hacer frente a esta limitación en las imágenes médicas, se han desarrollado redes que entrenaban con una configuración de una ventana deslizante para predecir la etiqueta de cada píxel de una región local (recorte de la imagen) alrededor del mismo píxel, que es utilizado como entrada del modelo [5]. Posteriormente, esta red neuronal se ha mejorado, haciéndola más rápida y evitando en lo posible la redundancia de los recortes de imágenes a procesar. En ese sentido, la arquitectura U-Net logró un desempeño destacado en diferentes aplicaciones de segmentación en imágenes biomédicas [6].

Los algoritmos basados en redes neuronales convolucionales suelen utilizar la salida de la última capa como representación de características. Sin embargo, la información de esta capa puede ser espacialmente demasiado amplia para permitir una localización precisa. Por otro lado, las capas anteriores pueden ser precisas en la localización, pero carecer de la capacidad para capturar la semántica. Para resolver este problema, se desarrollaron soluciones que utilizan hiper-columnas como descriptores de píxeles [7].

Sin embargo, como se observó con los métodos para la estimación de la pose humana, la regresión directa de las coordenadas implica un mapeo altamente no lineal de las imágenes de entrada a coordenadas puntuales. En lugar de realizar la regresión de coordenadas, se propuso como alternativa un enfoque más sencillo: un mapeo de imagen a imagen basado en la regresión de mapas de calor, los cuales codifican la pseudo probabilidad de que un punto de referencia se encuentre en una posición de píxel determinada. De este modo, la red neuronal encargada de la estimación de la pose humana aprende a generar respuestas con valores altos en ubicaciones cercanas a la referencia objetivo, mientras que las respuestas en ubicaciones erróneas son suprimidas [8].

SECCIÓN III

MÉTODOS

Para realizar este trabajo se ha realizado los siguientes pasos, como se puede observar en la figura 3

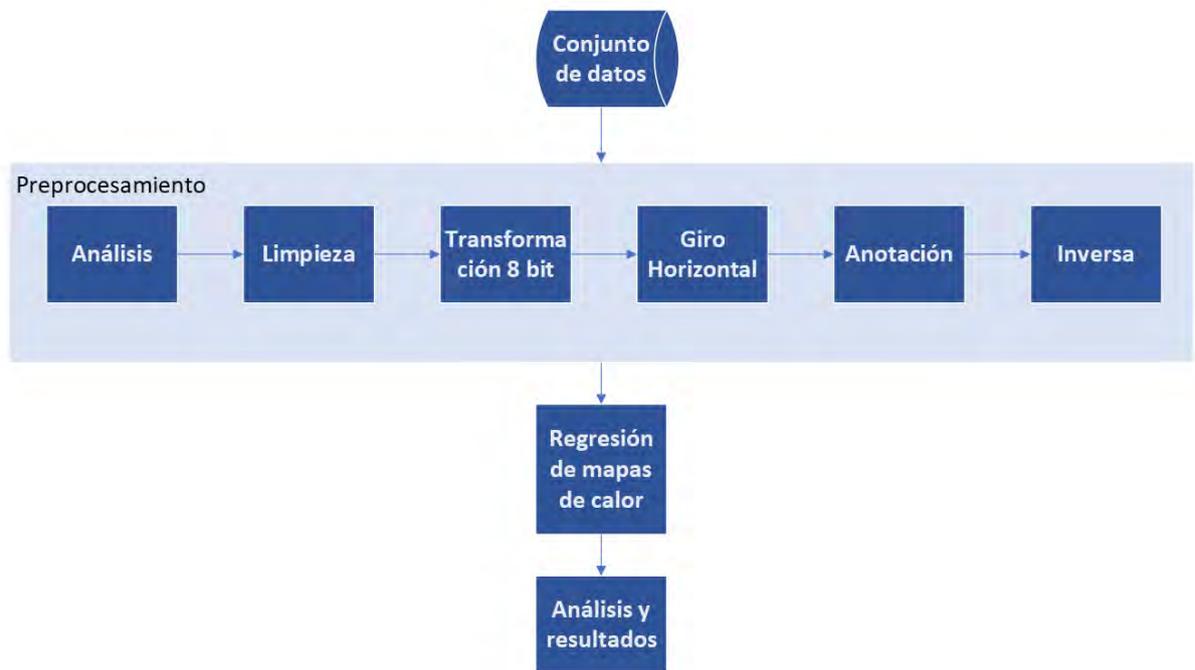


Figura 3: Proceso para la detección de 8 puntos de interés

A. Conjunto de datos

1. *Base de datos de imágenes de huellas plantares en la plataforma de ciencia de datos Kaggle:* En este conjunto de datos se tienen imágenes de la impresión con tinta de huellas plantares de 32 personas, se incluyen las imágenes del pie izquierdo y derecho. Las imágenes están en formato JPEG, escaneadas desde una hoja de papel con una resolución de 300 puntos por pulgada y 32 bits RGB. Este conjunto de datos tiene en total 100 imágenes, 60 del pie izquierdo y 40 del pie derecho [9].
2. *Base de datos de imágenes de huellas plantares de la ONG Pies Felices – Trujillo - Perú:* Este conjunto de datos fue entregado por la Organización No Gubernamental (ONG) Pies Felices, con sede en la ciudad de Trujillo, al norte del Perú. Las imágenes fueron escaneadas a una resolución de 300 puntos por pulgada y 32 bits RGB. El conjunto de datos comprende 65 imágenes de la impresión con tinta de huellas plantares, distribuidas en 32 del pie izquierdo y

33 del pie derecho.

3. *Base de datos de imágenes de huellas plantares en IEEEDataPort - Datasets*: El conjunto de datos BIOMETRIC 220X6 HUMAN FOOTPRINT pretende dotar a la huella humana de capacidad jurídica al ser usada como identificación biométrica. Este conjunto de datos, creado utilizando el escáner EPSON 5500, consta de 6 huellas multispectrales del lado derecho por persona, obtenidas de 220 voluntarios en diferentes periodos de tiempo, lo que suma un total de 1320 imágenes. [10].

B. Preprocesamiento de imágenes

- *Análisis*: Se analizaron las imágenes, todas con un tamaño variable dependiendo del conjunto de datos de procedencia. Los tamaños de las imágenes variaban desde 3507 píxeles de alto por 2550 píxeles de ancho hasta 666 píxeles de alto por 256 píxeles de ancho.
- *Limpieza*: Se llevó a cabo una limpieza manual para eliminar anotaciones realizadas sobre las hojas con bolígrafos u otros utensilios, que contenían información diversa como nombres, codificación, etc., así como algunos errores como manchas de tinta o marcas en el papel. También se descartaron las imágenes incompletas o con errores materiales, como manchas muy grandes que cubrían parte de la huella plantar.
- *Transformación a 8 bits*: Se procedió a transformar las imágenes a escala de grises, con 8 bits por píxel, para obtener valores de intensidad de negro entre 0 y 255 para cada píxel.
- *Giro horizontal*: Se creó una copia espejo horizontal de las imágenes de las huellas plantares del pie izquierdo, de modo que todas las imágenes tuvieran la configuración de una huella del pie derecho. Se realizó este procedimiento para tener una única configuración de huella plantar. Como se puede observar en la figura 4, la imagen de la izquierda corresponde al pie izquierdo, mientras que la imagen de la derecha es la copia espejo que ya tiene la configuración de un pie derecho.



Figura 4: Izquierda - Imagen original. Derecha - Copia espejo huella plantar

- *Anotación de puntos de interés:* Se llevó a cabo la anotación de ocho (8) puntos de interés en cada imagen, como se puede observar en la figura 5. Esta anotación se realizó sobre las imágenes con la resolución original para garantizar una mayor precisión en la tarea. Los puntos del 1 al 5 se utilizaron para obtener las coordenadas del píxel superior de la huella plantar, mientras que los puntos 6, 7 y 8 se emplearon para obtener las coordenadas de la izquierda, derecha e inferior de la huella plantar, así como los puntos de apoyo del pie.

Estos puntos se detallan a continuación:

- Punto 1: Corresponde a la parte superior del primer dedo, hallux o dedo gordo.
- Punto 2: Corresponde a la parte superior del segundo dedo.
- Punto 3: Corresponde a la parte superior del tercer dedo.
- Punto 4: Corresponde a la parte superior del cuarto dedo.
- Punto 5: Corresponde a la parte superior del quinto dedo.
- Punto 6: Corresponde a la parte izquierda de la cabeza del primer metatarso, o la parte más izquierda de la huella plantar.
- Punto 7: Corresponde a la parte derecha de la cabeza del quinto metatarso, o la parte más derecha de la huella plantar.

- Punto 8: Corresponde a la parte inferior del calcáneo, talón, o parte inferior de la huella plantar.

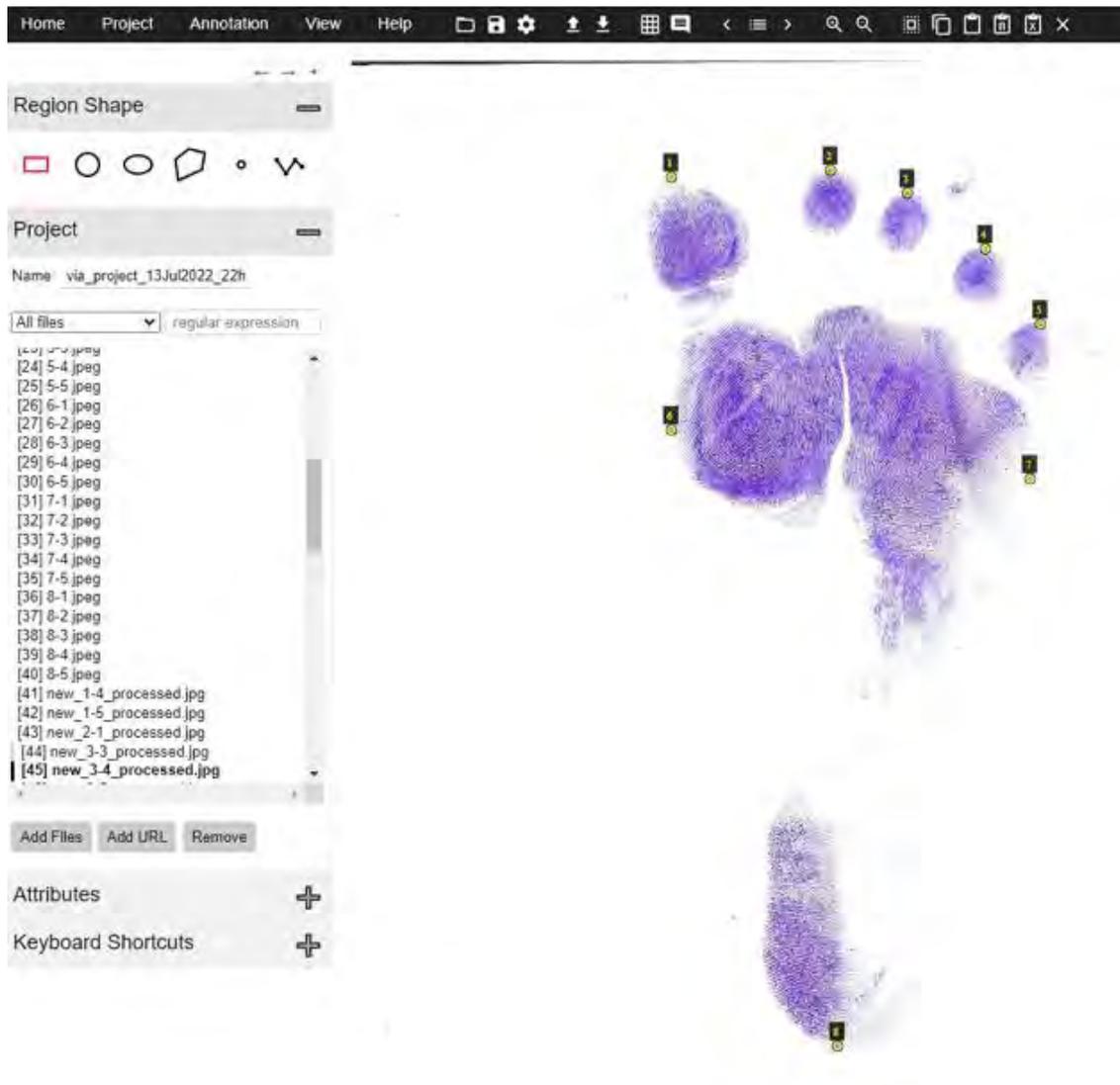


Figura 5: Anotación huellas plantares en VGG Image Annotator

- *Inversión*: Se procedió a invertir la imagen cambiando el fondo de color blanco o valores cero (0) a negro o valores doscientos cincuenta y cinco (255), de modo que la huella pudiera tener valores positivos mayores a cero (0), como se puede observar en la figura 6.



Figura 6: Inversa de la imagen de huella plantar

Para el proceso de anotación de los ocho (8) puntos en las imágenes, se utilizó el programa de código abierto VGG Image Annotator (VIA) [11]. Para todos los procesos de modificación y transformación en las imágenes, se empleó el programa de código abierto Fiji [12].

C. Arquitectura

Regresión de mapas de calor: Para este método se utilizó la plataforma Google Colab. Esta implementación se realizó con la opción de una GPU NVIDIA A100 con 40 GB de RAM, y el tiempo de entrenamiento fue de aproximadamente entre 15 a 45 minutos. El tamaño total de las redes neuronales oscila entre 150 MB y 2 GB. Se han utilizado el marco de trabajo de software libre para aprendizaje automático basado en el lenguaje de programación Python, Pytorch, así como las librerías Torch y fastai, que son librerías de aprendizaje profundo. El código utilizado es una adaptación de la implementación oficial en Pytorch del trabajo “Aprendizaje profundo de representaciones de alta resolución para la estimación de la pose humana” [13]. Se entrenó una red convolucional U-Net [6], la cual consiste en una red con una ruta que codifica y contrae las imágenes para capturar contexto, y otra ruta de expansión simétrica que permite una localización de puntos de interés de manera bastante precisa y en un tiempo aceptable utilizando los procesadores GPU actuales. Nuestro método se basa en la regresión de imágenes de mapas de calor [14], que codifican la probabilidad de que un punto de interés se encuentre en una posición de píxel determinada. Al permitir un mapeo de imagen a imagen, nos beneficiamos del uso de las redes convolucionales, ya que se reduce el número de pesos de la red y, por tanto, la complejidad computacional total.

D. Experimentación

Para realizar la evaluación de las imágenes con este método, primero se modificó el tamaño de las imágenes. Inicialmente, se cambiaron a un tamaño de 333x256 píxeles y posteriormente, en un segundo experimento, a 333x128 píxeles. También se escaló a estos nuevos tamaños cada una de las 8 anotaciones, y se aplicó aumento de datos de entrenamiento girando la imagen hasta en 3 grados en sentido de las manecillas del reloj o en sentido contrario, de manera aleatoria. Posteriormente, se crearon los mapas de calor para cada una de las anotaciones, como se muestra en la figura 7.

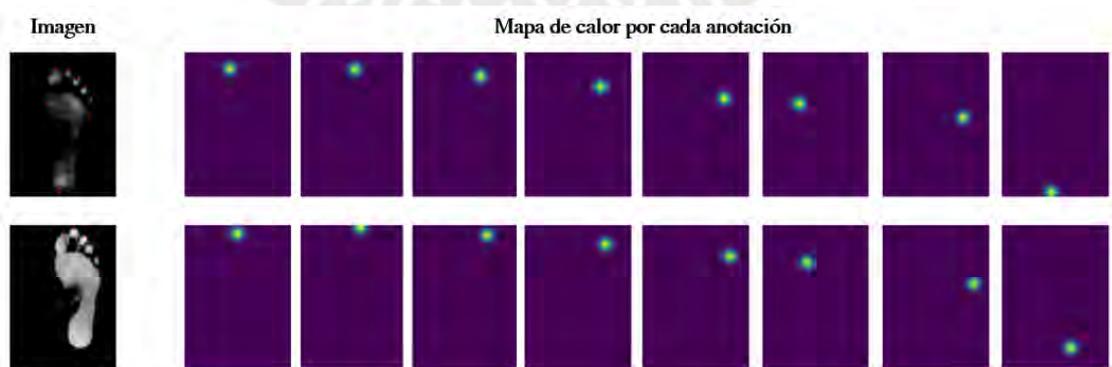


Figura 7: Imágenes de 333x256 px con 8 anotaciones y mapas de calor por cada anotación

Se utiliza el error absoluto promedio normalizado (Normalized Mean Absolute Error – NMAE por sus siglas en inglés), que es el error absoluto porcentual

de cada coordenada inferida respecto a la real. En el presente trabajo, la función utilizada recibe como parámetros dos mapas de calor, que se convierten en puntos de referencia para realizar el cálculo del NMAE, como se puede observar en la ecuación 1, para evaluar tanto la etapa de entrenamiento como los resultados. Como función de pérdida (Loss function) se utiliza el error medio cuadrado (MSE por sus siglas en inglés) de dos mapas de calor.

$$NMAE = \frac{1}{n} \sum_{i=1}^n \frac{|y_i - y_i^{\wedge}|}{|y_i|} \quad (1)$$

Donde

- n es el número total de muestras en el conjunto de datos.
- y_i es el valor real en la posición i .
- y_i^{\wedge} es el valor predicho por el modelo en la posición i .

El proceso de entrenamiento consiste en una combinación de operaciones de codificación y decodificación de mapas de calor. Para llevar a cabo esta tarea, se han configurado dos experimentos utilizando dos redes neuronales del tipo U-Net basadas en redes ResNet, preentrenadas con las imágenes de la base de datos Imagenet. Se ha modificado la última capa de estas redes para que la salida sea una de las 8 opciones de los puntos de referencia. En la Figura 8 se describe la estructura de la red neuronal U-Net.

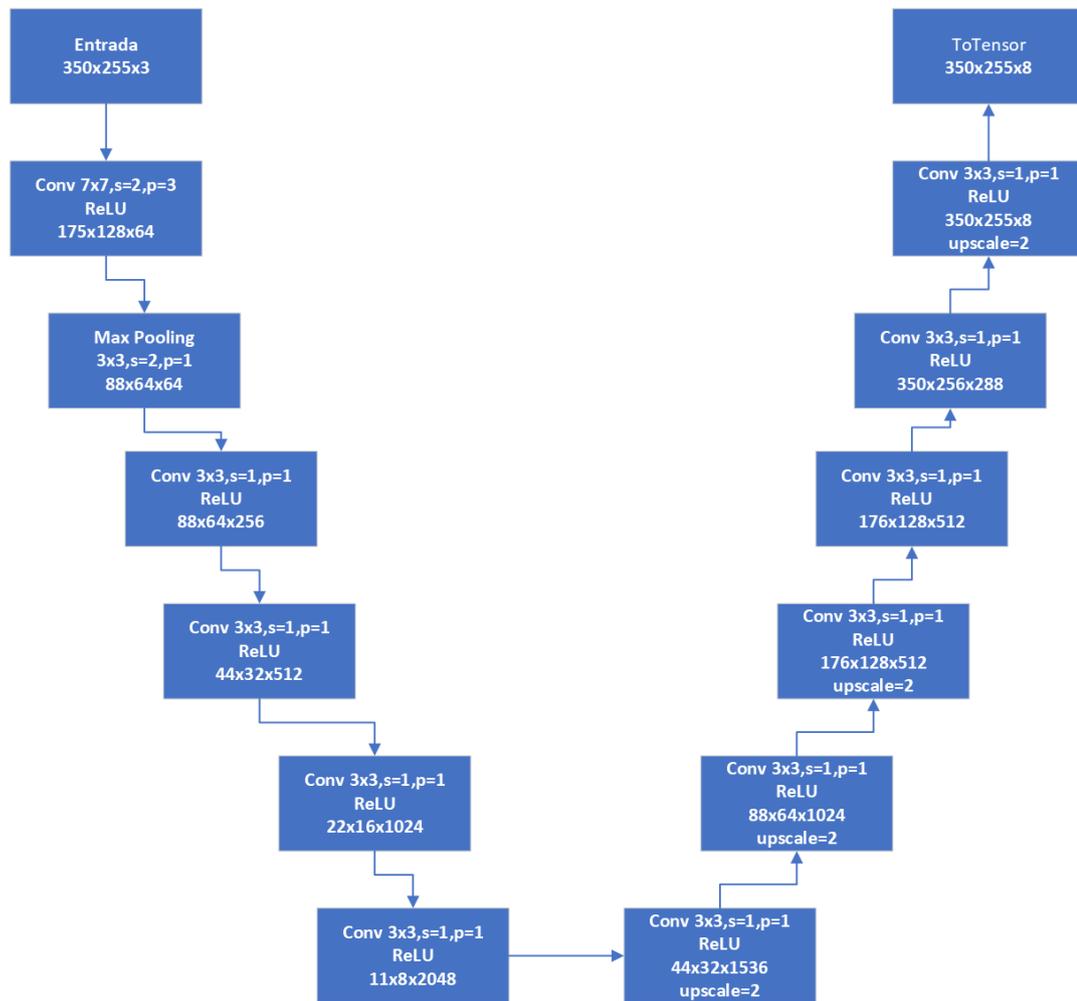


Figura 8: Estructura de la red U-Net con base en una ResNet50 pre entrenada en la parte de la codificación (izquierda de la U-Net)

Se dividieron las imágenes en 1056 para entrenamiento, 264 para validación del modelo y 132 para pruebas. Se entrenó la red neuronal por 26 épocas, con una tasa de aprendizaje (learning rate) de $1e-4$ y se obtuvieron los resultados de la figura 10, también se puede observar la predicción sobre las imágenes de validación en la figura 9.

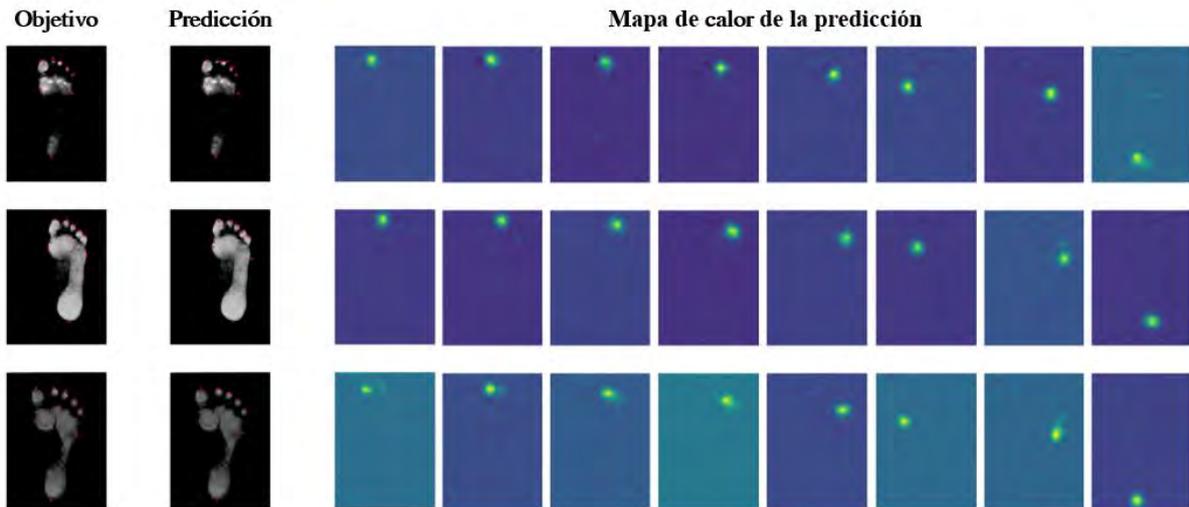


Figura 9: Resultados de la red U-Net (objetivo, predicción, mapa de calor de la predicción)

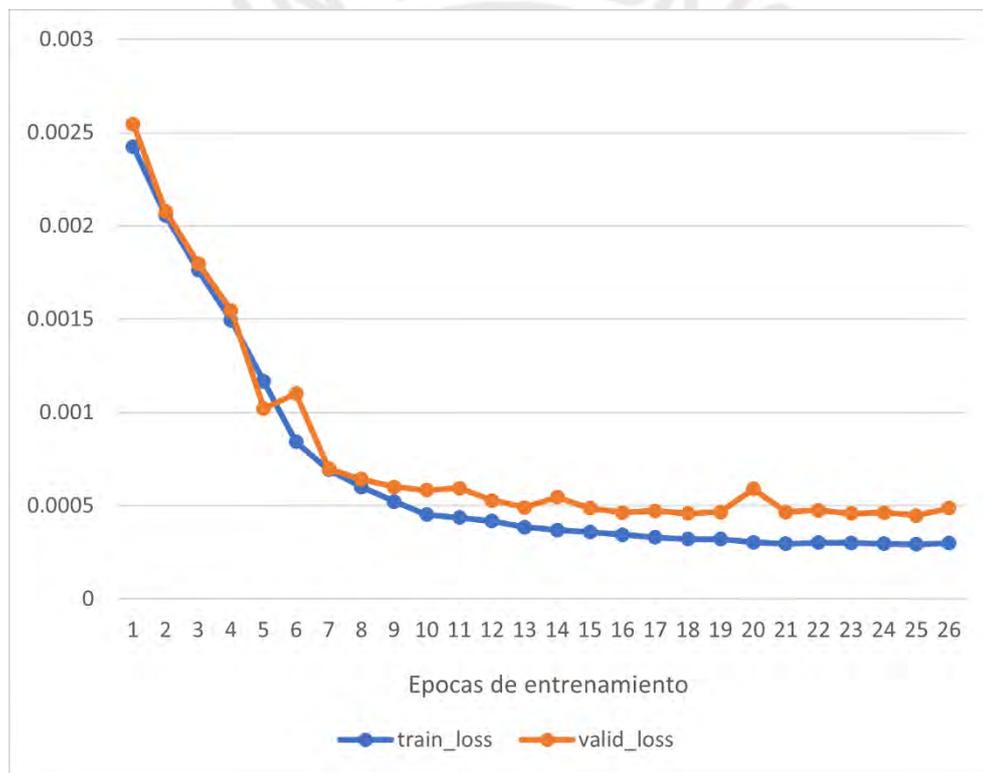


Figura 10: Entrenamiento por 26 épocas de la red U-Net (ResNet50) con imágenes de 333x256 px.

SECCIÓN IV

RESULTADOS

De los diversos experimentos realizados, se concluye que la configuración de red U-Net basada en ResNet50 preentrenada, utilizando imágenes de 333x256 píxeles, arroja el menor valor en el conjunto de validación.

Otro punto para tener en cuenta es el tiempo de entrenamiento: la red U-Net con base en ResNet50 (con un total de 339 millones de parámetros, de los cuales 315 millones son entrenables) requiere aproximadamente el doble de tiempo para completar el entrenamiento en comparación con la U-Net basada en una ResNet18 (con un total de 31 millones de parámetros, de los cuales 19 millones son entrenables).

Para el presente trabajo, el entrenamiento tomó alrededor de 15 minutos para la red ResNet18 y 45 minutos para la red ResNet50, utilizando una tarjeta gráfica GPU NVIDIA A100 con 40 GB de RAM. Por otro lado, el tamaño de las redes neuronales resultantes es otro factor para tener en cuenta: las redes neuronales basadas en ResNet18 tienen un tamaño final de aproximadamente 150 MB, mientras que las redes basadas en ResNet50 tienen un tamaño final de 1500 MB, siendo diez veces en tamaño más grandes que las anteriores. El resumen de los resultados se presenta en la Tabla 1.

Modelos	Tamaño de la imagen	NMAE
U-Net (ResNet18)	333x256	0.011914
	333x128	0.010470
U-Net (ResNet50)	333x256	0.010170
	333x128	0.011914

Tabla 1 Resultado de los modelos propuestos sobre el conjunto de imágenes de validación.

SECCIÓN V

CONCLUSIÓN Y DISCUSIÓN

En este trabajo de investigación se han experimentado con diversas variaciones de un método propuesto conocido como regresión con mapas de calor, con el objetivo de detectar puntos de interés sobre una imagen de una huella plantar para su uso en la fabricación de una plantilla ortopédica. La red neuronal U-Net, basada en una red preentrenada ResNet50 y utilizando imágenes de 333x256 píxeles, presenta resultados que se asemejan en precisión de la ubicación de puntos de interés con el trabajo que realizan los ortopedistas en el proceso manual de desarrollo de unas plantillas basadas en las imágenes de huellas plantares, como se puede evidenciar en la figura 11. Esta red neuronal generada puede ser utilizada para resolver el problema planteado, que consiste en la detección automática de puntos de interés para la diagramación de una plantilla ortopédica, lo que permitirá agilizar significativamente el proceso de creación de estas plantillas.



Figura 11: Error absoluto promedio normalizado - NMAE

TRABAJOS FUTUROS

- Se explorarán alternativas para procesar imágenes de mayor tamaño que 333x256 píxeles, considerando las limitaciones de recursos computacionales asociadas al método de regresión de mapas de calor.
- Se contemplaría investigar un segundo método para la detección de puntos de interés mediante el uso de transformadores de visión (vision transformers).
- Se llevaría a cabo una investigación sobre la diagramación y codificación de una plantilla ortopédica, con el objetivo de facilitar su envío a una impresora 3D o a un torno CNC computarizado para su fabricación.



REFERENCIAS BIBLIOGRÁFICAS

- [1] P. G. Peña Montoya, "Análisis mediante elementos finitos a órtesis de pie, plantillas ortopédicas, y comparación de los modelos en base a resultados obtenidos de un sistema de medición de presiones plantares," Master's thesis, 2018.
- [2] K. Kimura, T. Utsumi, M. Kouchi, and M. Mochimaru, "3d foot scanning system in foot-automated anatomical landmark detection and labeling," in *Asian Workshop on 3D Body Scanning Technologies, Tokyo, Japan*, pp. 17-18, 2012.
- [3] B. Yu and D. Tao, "Heatmap regression via randomized rounding," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021.
- [4] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, pp. 436-444, 2015.
- [5] D. Ciresan, A. Giusti, L. Gambardella, and J. Schmidhuber, "Deep neural networks segment neuronal membranes in electron microscopy images," *Advances in neural information processing systems*, vol. 25, 2012.
- [6] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, pp. 234-241, Springer, 2015.
- [7] B. Hariharan, P. Arbeláez, R. Girshick, and J. Malik, "Hypercolumns for object segmentation and fine-grained localization," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 447-456, 2015.
- [8] J. J. Tompson, A. Jain, Y. LeCun, and C. Bregler, "Joint training of a convolutional network and a graphical model for human pose estimation," *Advances in neural information processing systems*, vol. 27, 2014.
- [9] R. Khokher and R. C. Singh, "Footprint-based personal recognition using dactyloscopy technique," in *Industrial Mathematics and Complex Systems*, pp. 207-219, Springer, 2017.
- [10] K. Nagwanshi and S. Dubey, "Biometric 220x6 human footprint," DOI: <https://doi.org/10.21227/7gmx-jq63>, 2019.
- [11] A. Dutta and A. Zisserman, "The VIA annotation software for images, audio and video," in *Proceedings of the 27th ACM International Conference on Multimedia, MM '19, (New York, NY, USA), ACM*, 2019.
- [12] J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, et al., "Fiji: an open-source platform for biological-image analysis," *Nature methods*, vol. 9, no. 7, pp. 676-682, 2012.
- [13] K. Sun, B. Xiao, D. Liu, and J. Wang, "Deep high-resolution representation learning for human pose estimation," in *CVPR*, 2019.
- [14] C. Payer, D. Stern, H. Bischof, and M. Urschler, "Integrating spatial configuration into heatmap regression based cnns for landmark localization,"

