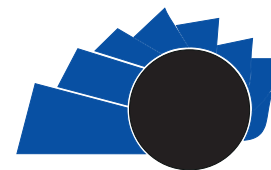




UNIVERSIDAD DISTRITAL
FRANCISCO JOSÉ DE CALDAS

Visión Electrónica *Más que un estado sólido*

<https://revistas.udistrital.edu.co/index.php/visele>



Visión Electrónica

VISIÓN INVESTIGADORA

Red neuronal convolucional para discriminar herramientas en robótica asistencial

Convolutional neural network to discriminate tools for assistance robotics

Robinson Jiménez Moreno¹, Oscar Avilés S.², Diana Marcela Ovalle³

INFORMACIÓN DEL ARTÍCULO

Historia del artículo:

Enviado: 19/07/2018

Recibido: 13/08/2018

Aceptado: 31/08/2018

Palabras clave:

Aprendizaje de máquina

Aprendizaje profundo

Reconocimiento de patrones

RNC

Robótica asistencial

Visión de máquina

Open access



Keywords:

Machine learning

Deep learning

Pattern recognition

CNN

Assisted robotics

Machine vision

RESUMEN

En el presente artículo se expone el entrenamiento de una Red Neuronal Convolutiva (RNC) para discriminación de herramientas de uso común en tareas de mecánica, electricidad, carpintería y similares. Para el caso, se toman como objetivos de entrenamiento pinzas, destornilladores, tijeras y alicates, los cuales puedan ser identificados por la red, y permite dotarle a un brazo robótico la facultad de identificar una herramienta deseada—de entre las anteriores—para su posible entrega a un usuario. La arquitectura neuro convolutiva empleada para la red presenta un porcentaje de acierto del 96% en la identificación de las herramientas entrenadas.

ABSTRACT

This article exposes the training of a Convolutional Neuronal Network (CNN) for discrimination of tools, commonly used in mechanical, electrical, carpentry and similar tasks. For this purpose, training objectives include nippers, screwdrivers, scissors and pliers, in order to be classified by the network, allowing a robotic arm to identify a desired tool for its possible delivery to a user. The CNN architecture used for the network presents a 96% success rate in the identification of tools.

¹ Ing. Electrónico, Universidad Distrital Francisco José de Caldas, Colombia. MSc. En Automatización Industrial, Universidad Nacional de Colombia, Colombia. Afiliación institucional: Universidad Militar Nueva Granada, Colombia. Correo electrónico: robinson.jimenez@unimilitar.edu.co. ORCID: <https://orcid.org/0000-0002-4812-3734>

² Ing. Electrónico, Universidad Antonio Nariño, Colombia. MSc. En Sistemas Automáticos de Producción, Universidad Tecnológica de Pereira, Colombia. PhD. En Ingeniería Mecánica, Universidade Estadual de Campinas, Brasil. Afiliación institucional: Universidad Militar Nueva Granada, Colombia. Correo electrónico: oscar.aviles@unimilitar.edu.co. ORCID: <https://orcid.org/0000-0001-8676-9926>

³ Ing. Electrónica, Universidad Distrital Francisco José de Caldas, Colombia. MSc. En Ingeniería Eléctrica, Universidad de los Andes, Colombia. PhD. En Tecnologías Industrial, Universidad Politécnica de Cartagena, España. Lugar de trabajo o Afiliación institucional: Universidad Distrital Francisco José de Caldas, Colombia. Correo electrónico: dmovalle@gmail.com. ORCID: <https://orcid.org/0000-0002-3650-4273>

1. Introducción

Los agentes robóticos hacen parte común en labores de ensamble en múltiples empresas donde se realizan labores repetitivas y preprogramadas las cuales hoy en día se están potencializando mediante técnicas de inteligencia artificial, [1]. Estas técnicas involucran algoritmos de visión de máquina que en robots industriales permiten integrar múltiples factores de autonomía y precisión, como se presenta en [2]. Pero, aún más, actualmente se orientan en asegurar la interacción de personas en el área de trabajo del robot evitando colisiones entre ambos, [3]. Lo anterior ha conllevado al desarrollo de múltiples aplicaciones en la interacción de ambos campos: la robótica y la visión de máquina; por ejemplo en [4] se presenta una aplicación robótica para detección de enfermedades en cultivos basada en procesamiento de imagen y redes neuronales.

Sin embargo, tareas de control robótico por visión de máquina implican un eficiente reconocimiento de objetos permitiendo la ejecución de tareas como la navegación autónoma del robot por seguimiento, como propone [5]. Dicha tarea de reconocimiento requiere de un algoritmo de aprendizaje de máquina que acompañe a los de procesamiento de imágenes, sea por redes neuronales como se menciona en [4], métodos de clasificación bayesianos por vecinos cercanos como se expone en [6], o sistemas de inferencia difusa, por ejemplo para aplicaciones de parqueo, como se expone en [7]. La integración de estas técnicas de inteligencia artificial dan lugar a lo que hoy se denomina robots inteligentes, como el descrito en [8] que cuenta con múltiples grados de libertad, una cámara de profundidad y algoritmos de procesamiento orientados al reconocimiento de objetos.

De otra parte, recientes técnicas de inteligencia artificial han mostrado su eficiencia en el reconocimiento de objetos en imágenes, tal es el caso de las redes RNC, [9-10]. Estas redes toman directamente la imagen y entregan la predicción del objeto presentado, mostrando mayor robustez que las técnicas de procesamiento de imagen tradicionales, por lo que se hacen ideales para el desarrollo de robots inteligentes, aspecto poco abordado pero donde ya se encuentran trabajos en el área, [11].

Por lo anterior, la fortaleza de las RNC ha derivado en múltiples aplicaciones en el campo de la medicina [12], [13], [14], al ser empleada, por ejemplo, en imágenes diagnósticas. Igualmente en el campo

de reconocimiento de objetos, como se evidencia en [15], [16]; e incluso en extracción de características para determinación de patrones de precisión bajo medidas antropométricas, [17]. Derivado de todo lo anterior, la presente investigación se centra en evaluar este tipo de redes en el entrenamiento de la detección de herramientas encaminadas al soporte de actividades mediante un robot asistencial.

El artículo se estructura en tres secciones: la primera en el entrenamiento de la RNC para la identificación de un grupo de herramientas puntuales, pero de uso general, lo que abre la posibilidad de integración de este desarrollo a aplicaciones de robótica asistencial en múltiples escenarios (robots inteligentes). La segunda sección corresponde al análisis de resultados frente a la predicción de la red a través de una matriz de confusión que relaciona el aprendizaje de cada categoría de las herramientas empleadas; y la última sección expone las conclusiones alcanzadas.

2. Entrenamiento de la RNC

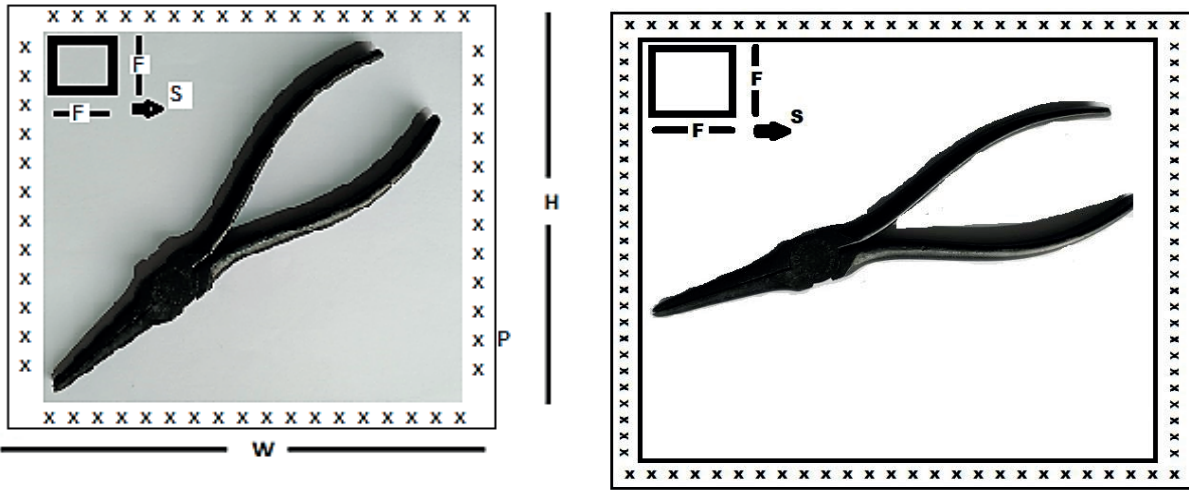
Las RNC cuentan con una estructura básica vista como un núcleo típico basado en tres capas: una de convolución, una de rectificación lineal (RELU) y una denominada pooling, [10]. La red puede incluir n de esos núcleos y variar el orden de dichas capas, siendo necesario establecer los parámetros básicos de cada una. Tales parámetros se caracterizan al determinar el tamaño estándar de las imágenes de entrada; es decir, todas las imágenes de la base de datos deberán tener la misma dimensión, un alto (H) y ancho conocido (W), así como una profundidad (D). La profundidad está determinada en 1 para imágenes en escala de grises y en 3 para imágenes en color. La Figura 1 ilustra los diferentes hiperparámetros a establecer en el entrenamiento, los cuales se describen a continuación.

Filtro (F): corresponde inicialmente al elemento de convolución, posee unas dimensiones de alto y ancho basadas en la imagen; la profundidad del filtro depende también si la imagen es a color o escala de grises.

Stride (S): corresponde al intervalo de desplazamiento para aplicar los filtros sobre toda la imagen de entrada.

Padding (P): o relleno lateral, extiende los lados usualmente con ceros, pero puede replicarse las filas y/o columnas adyacentes; se define acorde al tamaño

Figura 1. Hiperparámetros de convolución



Fuente: elaboración propia.

del filtro, la imagen y el Stride, para cubrir todos los píxeles dentro del proceso de convolución.

Dichos parámetros interrelacionan las capas entre sí: la imagen de entrada con el filtro de convolución, el resultado de la convolución con la capa rectificación lineal, etc. Los cálculos de las dimensiones que debe tener cada conjunto de capas se relacionan mediante (1), (2) y (3).

$$W_n = \frac{(W_{n-1} - F + 2P)}{S} + 1 \tag{1}$$

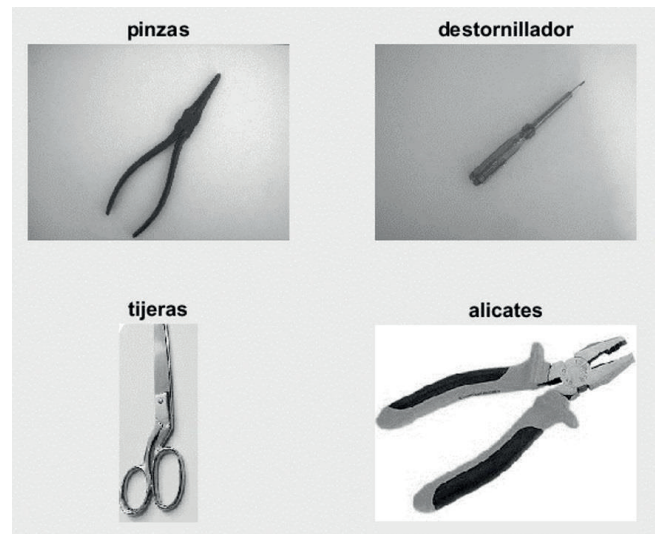
$$H_n = \frac{(H_{n-1} - F + 2P)}{S} + 1 \tag{2}$$

$$D_n = k \tag{3}$$

Como objeto de reconocimiento de la RNC a entrenar se establecen herramientas de uso común en tareas de mecánica, electricidad, carpintería y similares. Para el caso se toman como objetivos de entrenamiento pinzas, destornilladores, tijeras y alicates, todas a color. La Figura 2 ilustra una muestra de cada clase a identificar, y en la Tabla 1 se relaciona la cantidad de imágenes empleadas; el

conjunto de la base de datos se parte en dos: las de entrenamiento y las de validación, en una relación de 70/30 por ciento respectivamente.

Figura 2. Herramientas de entrenamiento



Fuente: elaboración propia.

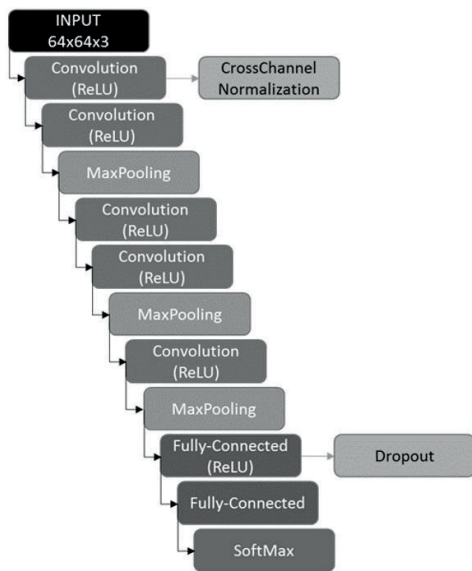
Tabla 1. Base de datos de entrenamiento.

IMAGENES	Pinzas	Alicates	Tijeras	Destornillador
Entrenamiento	75	30	53	64

Fuente: elaboración propia.

La Figura 3 ilustra la arquitectura de la RNC empleada que, como se mencionó, permite observar las diferentes combinaciones convolución-RELU-Pooling a que haya lugar. A su vez, se puede apreciar las capas de salida de la red (Fully – Connected y Softmax) típicas también de cualquier estructura neuro convolucional y que entregan la predicción final de la red respecto a una imagen de entrada sometida a los filtros aprendidos. El establecimiento de la arquitectura final a emplear se realizó mediante métodos heurísticos en función a las ecuaciones y núcleo base mencionados.

Figura 3. Arquitectura convolucional empleada



Fuente: elaboración propia.

3. Resultados y Análisis

El entrenamiento fue llevado a cabo mediante MATLAB®, en un equipo de cómputo con procesador de 2,7 Ghz y 16GB de memoria RAM, sin unidad de procesamiento grafico (GPU), permitiendo tiempos de entrenamiento de 5 minutos en promedio. Esta aclaración se realiza pues diferentes arquitecturas y aplicaciones de RNC pueden implicar el uso de GPU y considerable tiempo de entrenamiento. En

la Tabla 2 se presenta la matriz de confusión que relaciona el aprendizaje de cada categoría de las herramientas empleadas, permitiendo concluir que el 96% alcanzado por la red es un desempeño suficiente para la aplicación propuesta. El tiempo de clasificación de la red entrenada es promedio de 0,23 segundos.

Se puede observar a su vez que, hay una relación directa entre las características de la herramienta y la cantidad de imágenes empleadas. Para el caso, a pesar de se emplearon menor número de imágenes para los alicates, las características de esta herramienta frente a las demás, son notorias, de ahí que presenten el segundo mejor índice de identificación. Por su parte las pinzas, con mayor número de imágenes de entrenamiento y variación en su ubicación y rotación en la imagen, fueron las mejor aprendidas por la red. Sin embargo, el destornillador que posee una estructura simple, con bajas características presenta un mejor aprendizaje que las tijeras, ya que estas son más fácilmente confundidas con las pinzas.

En la Figura 4 se pueden apreciar algunos de los resultados de la clasificación de la red neuroconvolucional entrenada, donde se evidencian errores tanto para la categoría destornillador como para la de tijeras. Para el caso del destornillador, la imagen de prueba presenta un acercamiento a la herramienta, mientras que las imágenes usadas en la base de datos están más alejadas y se aprecia mucho más el fondo, esto resalta la parte de agarre del destornillador confundiéndolo con la estructura más gruesa de los alicates. En el segundo caso se logró establecer que en la base de datos de entrenamiento no había tijeras orientadas hacia abajo, lo cual, si ocurre con las pinzas, de ahí la interpretación errónea de la red. Al evaluar la red con objetos diferentes a los entrenados, es claro que los clasificaría dentro de una de las únicas cuatro categorías que conoce la red, evidenciando una tendencia hacia las pinzas, aspecto derivado de que es la herramienta con más imágenes en la base de datos y que incluye más variaciones en relación con traslación y rotación en la escena.

Tabla 2. Matriz de confusión predicción.

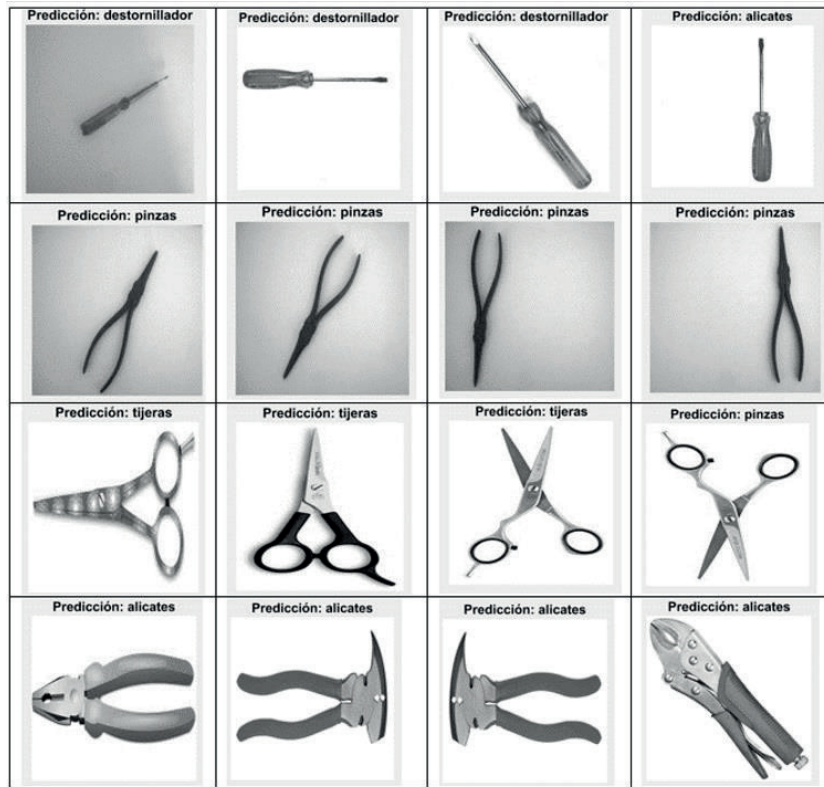
	Pinzas	Alicates	Tijeras	Destornillador	PROMEDIO
Verdadero positivo	98,6%	96,8%	93,6%	95,4%	96,1%
Falso positivo	1,4%	3,2%	6,4%	4,6%	3,9%

Fuente: elaboración propia.

Finalmente, se agregó una clase adicional al entrenamiento denominada otros, para validar la identificación de la red frente a herramientas conocidas y el resto posible. Dicha categoría consistió en 50 imágenes de 4 herramientas diferentes a las

deseadas, permitiéndole ahora a la red ser más selectiva sin comprometer el desempeño previo obtenido. La Figura 5 ilustra el resultado de clasificación de la red para este caso, que para el caso tarda 0,241 segundos.

Figura 4. Resultados de predicción de la red



Fuente: elaboración propia.

Figura 5. Resultados de predicción de la red con elementos no deseados



Fuente: elaboración propia.

4. Conclusiones

Para dotar un agente robótico con la capacidad de entregar herramientas se requiere implementar sobre su sistema de control algoritmos de inteligencia artificial, como son las RNC. En esta investigación se logró obtener una RNC eficiente (96%) para la discriminación de herramientas, demostrando así la versatilidad de esta técnica en el reconocimiento de objetos y sus aplicaciones en Robótica.

De otra parte, se logró evidenciar la importancia de la diversidad de imágenes en la base de datos de entrenamiento, ya que la predicción de la red se sesgó por efecto de las de mayores características presentes. A su vez, se evidenció que no solo inciden las traslaciones y rotaciones del objeto en la escena sino también su cercanía, es decir: alejar o acercar el objeto a detectar permite exhibir más características, o no, a ser aprendidas por la red, lo cual incide en la precisión de la clasificación final.

El aumentar las categorías de la red evidencia un requerimiento adicional del tiempo de entrenamiento, que no genera un costo computacional significativo (incremento del 5%), en la clasificación de la red.

5. Reconocimientos

La investigación de la cual deriva este artículo fue soportada por el grupo de investigación IDEAS de la Universidad Distrital Francisco José de Caldas.

Referencias

- [1] P. Sincak, P. Smolar y J. Ondo, “The major challenges in intelligent robotics (Towards cloud robotics and its impact to crowdsourced image processing for intelligent agents serving for robots)” 2015 13th International Conference on Emerging eLearning Technologies and Applications (ICETA), Stary Smokovec, 2015, pp. 1-6, <https://doi.org/10.1109/ICETA.2015.7558510>
- [2] S. Hu, S. Chen, S. Chen, G. Xu y D. Sun, “Automated Transportation of Multiple Cell Types Using a Robot-Aided Cell Manipulation System With Holographic Optical Tweezers” in *IEEE/ASME Transactions on Mechatronics*, vol. 22, n°. 2, pp. 804-814, April 2017.
- [3] S. D. Lee y J. B. Song, “Collision detection of humanoid robot arm under model uncertainties for handling of unknown object” 2015 IEEE-RAS 15th International Conference on Humanoid Robots (Humanoids), Seoul, 2015, pp. 718-721, <https://doi.org/10.1109/HUMANOIDS.2015.7363433>
- [4] S. K. Pilli, B. Nallathambi, S. J. George y V. Diwanji, “eAGROBOT — A robot for early crop disease detection using image processing” 2015 2nd International Conference on Electronics and Communication Systems (ICECS), Coimbatore, 2015, pp. 1684-1689, <https://doi.org/10.1109/ECS.2015.7124873>
- [5] S. Z. Murshed et al., “Controlling an embedded robot through image processing based object tracking using MATLAB” 2016 10th International Conference on Intelligent Systems and Control (ISCO), Coimbatore, 2016, pp. 1-6 <https://doi.org/10.1109/ISCO.2016.7726922>
- [6] S. Jia, Z. Ju, T. Xu, H. Zhang y X. Li, “Mechanical arm grasping and target recognition strategy based on the coarse-to-fine algorithm” 2016 IEEE International Conference on Information and Automation (ICIA), Ningbo, 2016, pp. 1812-1817, <https://doi.org/10.1109/ICInfA.2016.7832112>
- [7] Y. Y. Aye, K. Watanabe, S. Maeyama y I. Nagai, “Image-based fuzzy control of a car-like mobile robot for parking problems” 2016 IEEE International Conference on Mechatronics and Automation, Harbin, 2016, pp. 502-507, <https://doi.org/10.1109/ICMA.2016.7558614>
- [8] M. M. Ali, H. Liu, N. Stoll y K. Thurow, “Intelligent arm manipulation system in life science labs using H2O mobile robot and Kinect sensor” 2016 IEEE 8th International Conference on Intelligent Systems (IS), Sofia, 2016, pp. 382-387, <https://doi.org/10.1109/IS.2016.7737449>
- [9] A. Krizhevsky, I. Sutskever y G. E. Hinton, “Imagenet classification with deep convolutional neural networks”, NIPS’12 Proceedings of the 25th International Conference on Neural Information Processing Systems, 2012 – vol. 1 Pages 1097-1105.
- [10] M.D., Zeiler, R. Fergus. “Visualizing and Understanding Convolutional Networks”. In: Fleet D., Pajdla T., Schiele B., Tuytelaars T. (eds) Computer Vision – ECCV 2014. ECCV 2014. Lecture Notes in Computer Science, vol 8689. Springer, Cham,

- [11] Z. Wang, Z. Li, B. Wang, H. Liu. "Robot grasp detection using multimodal deep convolutional neural networks". *Advances in Mechanical Engineering*, vol. 8, Issue 9: September-23-2016.
- [12] M. Halicek, G. Lu, J. V. Little et al; "Deep convolutional neural networks for classifying head and neck cancer using hyperspectral imaging". *J. Biomed. Opt.* vol. 22, n° 6, <https://doi.org/10.1117/1.JBO.22.6.060503>
- [13] Z. Xiaolong, J. Beth, C. Rosa, T. Chung "Self-Recalibrating Surface EMG Pattern Recognition for Neuroprosthesis Control Based on Convolutional Neural Network". *Frontiers in Neuroscience*. Vol. 11, 2017, p 379, <https://doi.org/10.3389/fnins.2017.00379>
- [14] S. Vieira, H. L. Walter, A. Mechelli, "Using deep learning to investigate the neuroimaging correlates of psychiatric and neurological disorders: Methods and applications", *Neuroscience & Biobehavioral Reviews*, vol. 74, 2017, pp. 58-75, <http://dx.doi.org/10.1016/j.neubiorev.2017.01.002>.
- [15] Z. Deng, L. Lei, H. Sun, H. Zou, S. Zhou y J. Zhao, "An enhanced deep convolutional neural network for densely packed objects detection in remote sensing images" 2017 International Workshop on Remote Sensing with Intelligent Processing (RSIP), Shanghai, 2017, pp. 1-4, <https://doi.org/10.1109/RSIP.2017.7958800>
- [16] A. Ding, Q. Zhang, X. Zhou y B. Dai, "Automatic recognition of landslide based on CNN and texture change detection" 2016 31st Youth Academic Annual Conference of Chinese Association of Automation (YAC), Wuhan, 2016, pp. 444-448, <https://doi.org/10.1109/YAC.2016.7804935>
- [17] I. Marras, P. Palasek y I. Patras, "Deep Refinement Convolutional Networks for Human Pose Estimation" 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017), Washington, DC, 2017, pp. 446-453, <https://doi.org/10.1109/FG.2017.148>