








# Aprendizaje Automático Profundo y Visión por Computadora. Aplicaciones en el Reconocimiento de Lengua de Señas e Imágenes Astronómicas

L. Lanzarini<sup>1,2</sup> , F. Ronchetti<sup>1,2,3</sup> , F. Quiroga<sup>1,2,4</sup> , G. Rios<sup>1,3</sup>, U. Cornejo Fandos<sup>1,5</sup>, K. Canaza<sup>1</sup>, Pedro Dal Bianco<sup>1</sup>, Ivan Mindlin<sup>1</sup>, F. Ravettino<sup>1</sup>, A. Rosete<sup>6</sup> , R. Gamen<sup>7</sup> , D. Puig Valls<sup>9</sup> , J. Torrents-Barrena<sup>9</sup> , Y. Aidelman<sup>7,8</sup>, C. Escudero<sup>8</sup>.

<sup>1</sup> Instituto de Investigación en Informática LIDI, Facultad de Informática, Universidad Nacional de La Plata, La Plata, Argentina.\*

<sup>2</sup> Facultad de Informática, Universidad Nacional de La Plata, La Plata, Argentina

<sup>3</sup> Comisión de Investigaciones Científicas de la Pcia. De Bs. As. (CIC)

<sup>4</sup> Becario postgrado UNLP <sup>5</sup> Becario CIN

<sup>6</sup> Universidad Tecnológica de La Habana “José Antonio Echeverría” (CUJAE), La Habana, Cuba

<sup>7</sup> Facultad de Ciencias Astronómicas y Geofísicas, Universidad Nacional de La Plata, La Plata, Argentina

<sup>8</sup> Instituto de Astrofísica de La Plata (IALP CONICET), La Plata, Argentina

<sup>9</sup> Universitat Rovira i Virgili (URV), Tarragona, España

\* Centro asociado de la Comisión de Investigaciones Científicas de la Pcia. De Bs. As. (CIC)

Contacto: laural@lidi.info.unlp.edu.ar

## CONTEXTO

Esta presentación corresponde a las tareas de investigación que se llevan a cabo en el III-LIDI en el marco del proyecto F025 “Sistemas inteligentes. Aplicaciones en reconocimiento de patrones, minería de datos y big data” perteneciente al Programa de Incentivos (2018-2021).

## RESUMEN

Esta línea de investigación se centra en el estudio y desarrollo de Sistemas Inteligentes para la resolución de problemas de reconocimiento de patrones en imágenes y video, utilizando técnicas de Aprendizaje Automático clásicas, junto con Redes Neuronales Convolucionales y Aprendizaje profundo. El trabajo presentado describe diferentes casos de aplicación en visión por computadora.

Una de las líneas de investigación principales que se continúa desarrollando es el reconocimiento de lengua de señas. Este es un problema complejo y multidisciplinar, que presenta diversos subproblemas a resolver como el reconocimiento del intérprete, la segmentación de manos, la clasificación de diferentes configuraciones y de un gesto dinámico, entre otros.

En esta área se está estudiando la forma de reconocer formas de mano de la Lengua de Señas con conjuntos de datos de tamaño reducido, dada la falta de datos de entrenamiento para este dominio.

Además, se están comenzando a utilizar Redes Generativas Adversarias (GANs) para aumentar bases de datos de formas de mano, con el objetivo de complementar desde otro enfoque el entrenamiento de modelos para su clasificación.

Por otro lado, se está estudiando la forma en que las redes neuronales codifican la invarianza a las transformaciones y otras propiedades transformacionales, con el objetivo de poder analizar y comparar estos modelos, y finalmente mejorarlos. De esta forma se espera poder mejorar los modelos de clasificación de objetos transformados, en particular, de formas de mano.

Siguiendo con la línea de reconocimiento de patrones en imágenes, se está llevando a cabo una colaboración con investigadores de la Facultad de Astronomía y Geofísica de la UNLP para crear modelos de clasificación de imágenes de objetos celestes. Además, se está desarrollando un sistema para recuperar la información de placas espectrográficas astronómicas antiguas, con el objetivo de recuperar los espectrogramas que contienen.

Por último, se está diseñando un prototipo para crear experiencias interactivas multimedia que incorpore técnicas de aprendizaje profundo tanto para la entrada de datos por parte de los usuarios, como para la generación de una salida en forma de imágenes, videos, sonido y texto.

**Palabras clave:** Aprendizaje Automático, Visión por Computadoras, Lengua de Señas, Redes Generativas Adversarias, Invarianza, Equivarianza, Imágenes Astronómicas, Interacción Inteligente.

## 1. INTRODUCCION

El Instituto de Investigación en Informática LIDI (III-LIDI) tiene una larga trayectoria en el estudio, investigación y desarrollo de Sistemas Inteligentes basados en distintos métodos de Aprendizaje Automático y Redes Neuronales

Como resultado de estas investigaciones se han diseñado e implementado técnicas originales aplicables a la clasificación y el análisis de características de objetos en imágenes, generación de imágenes para aumentación de datos, y estudio del funcionamiento de las redes neuronales. En relación con esta línea, actualmente se están desarrollando los siguientes temas:

### 1.1. Clasificación de formas de mano de la Lengua de Señas

Las lenguas de señas utilizan un conjunto finito de formas de mano que, en combinación con movimientos de las manos y el cuerpo, y expresiones faciales, se utilizan para señalar.

Una etapa fundamental en el reconocimiento de la lengua de señas es la clasificación de estas formas de mano, y por ende un área de investigación prioritaria para mejorar el reconocimiento.

Por otro lado, las Redes Neuronales Convolucionales (CNN, por sus siglas en inglés) son actualmente el estado del arte en el área. En un estudio previo, se realizaron experimentos comparando diversas arquitecturas de CNN para clasificar formas de manos, pero sólo con dos bases de datos,

LSA16 y RWTH (ver Figura 1). Los resultados de estas investigaciones fueron publicados en [1] advirtiendo que el desempeño de los modelos utilizados estuvo limitado por la falta de datos.



Figura 1. Ejemplos de las bases de datos de formas de mano RWTH y LSA16.

Por ende, se realizó una comparación de distintos modelos de Redes Neuronales diseñados específicamente para trabajar con conjuntos de datos de tamaño reducido [2], extendiendo también el análisis a otras bases de datos. Estos modelos específicos demostraron una mejora significativa respecto a los modelos tradicionales. Actualmente, se está continuando este trabajo incrementando la variedad de bases de datos y modelos a evaluar.

### 1.2. Redes Generativas Adversarias para la Lengua de Señas.

Recientemente, las Redes Generativas Adversarias han mostrado resultados satisfactorios al crear imágenes artificiales a partir de datos aleatorios, o de una imagen para conseguir una transformación de la misma [6]. Actualmente, se está comenzando a investigar cómo funcionan estas redes para generar videos artificiales.

Por otro lado, las bases de datos de Lengua de Señas, dada la naturaleza compleja que poseen, o bien poseen pocos datos o los datos que las forman no presentan la suficiente diversidad. Además, la creación de datos artificiales no es trivial como en otros dominios, lo que hace muy difícil aplicar técnicas clásicas como *Data Augmentation*.

Por tal motivo, en el III LIDI se están comenzando a utilizar Redes Generativas Adversarias (GANs) para generar imágenes y

videos artificiales relacionados con la lengua de señas. Este tipo de redes permitirá aumentar las bases de datos de formas de mano, con el objetivo de complementar desde otro enfoque el estudio de modelos y algoritmos de clasificación para bases de datos con pocos datos etiquetados. Estas investigaciones son llevadas a cabo en marco de una tesis doctoral financiada por la UNLP a través de una beca de postgrado.

### 1.3. Métricas de Equivarianza

Una de las propiedades deseables en un modelo de clasificación de formas de mano es que sea invariante a la rotación y otras transformaciones afines, debido a que la forma de mano se considera la misma independientemente de la orientación que tenga, su tamaño, y su posición.

En los últimos años, varios modelos fueron propuestos para añadir invarianza a la rotación y otras transformaciones en CNNs [3]. No obstante, no está claro como estos modelos impactan en el aprendizaje usual de los pesos de la red.

Por este motivo, se ha trabajado en definir y evaluar diversas métricas de las propiedades Invarianza y Auto-Equivarianza, casos especiales de la Equivarianza. Estas métricas permiten la caracterización de la forma en que los distintos modelos invariantes o equivariantes codifican dichas propiedades [4]. Se está trabajando en mejorar el desempeño y la usabilidad de las métricas para popularizar su uso entre investigadores del área.

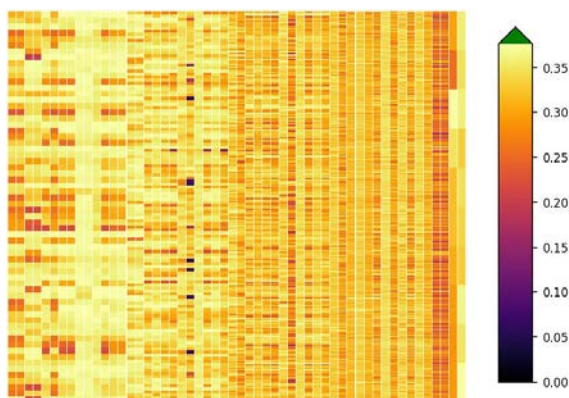


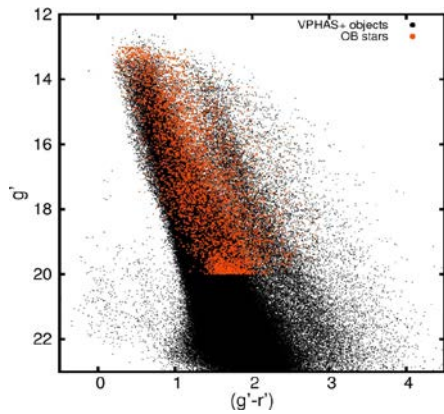
Figura 2. Invarianza por capas y unidades de una CNN con arquitectura ResNet.

### 1.4. Análisis de Imágenes Astronómicas

En los últimos años, la cantidad de información astronómica disponible se ha multiplicado de forma exponencial. En consecuencia, diversas tareas que anteriormente se realizaban de forma manual o semi-manual deben ahora automatizarse aún más. En ocasiones, también requieren de sistemas inteligentes para realizar tareas cuyos criterios no son objetivos, como la clasificación o identificación de objetos celestes.

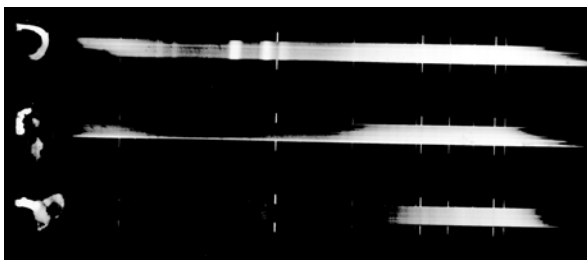
En este ámbito, en el III-LIDI se están desarrollando dos proyectos. El primero consiste en determinar modelos de clasificación a partir de información fotométrica de estrellas Be [5]. Estas estrellas son estrellas no-supergigantes cuyo espectro exhibe emisión en línea  $H\alpha$ . La técnica que generalmente se utiliza para detectar estrellas Be es el uso de diagramas fotométricos de dos colores. No obstante, estos diagramas pueden identificar otras fuentes astrofísicas, como por ejemplo estrellas supergigantes o estrellas Wolf-Rayet, entre otras. Por otro lado, como ya se mencionó, la cantidad de información recolectada por los telescopios actuales es enorme, lo que lleva a un tiempo extremadamente alto para analizar de forma manual estos diagramas.

En este sentido, se están implementando modelos clásicos de Aprendizaje Automático que permitan una detección de estrellas Be candidatas, para luego ser evaluadas por un experto [7]. Para esto se está comparando la información proveniente del catálogo público de MohrSmith, el cual se encuentra etiquetado, con otros catálogos no etiquetados. Además, se está investigando qué variables son realmente relevantes para analizar estos cuerpos celestes.



**Figura 3. Objetos celestes en la región Carina's Arm.**

El segundo proyecto que se está desarrollando en el III-LIDI se refiere al diseño e implementación de un sistema de procesamiento automático de placas espectrográficas astronómicas antiguas. La espectroscopia tiene un uso especial en astronomía ya que se puede usar para averiguar muchas propiedades de estrellas y galaxias distantes. En el caso particular de las placas espectrográficas, se están utilizando técnicas de procesamiento de imágenes inteligente para corregir algunos defectos propios de la tecnología de la época. De esta forma, se podrán recuperar los espectrogramas que contienen información nunca antes analizada ni tampoco capturada, ya que no existían otros telescopios mirando la misma porción del cielo en esas épocas.



**Figura 4. Placa espectrográfica digitalizada.**

### 1.5. Multimedia interactiva con Aprendizaje Profundo

En el marco del proyecto de investigación, desarrollo e innovación “Generación de contenido multimedia mediante IA para entornos interactivos” evaluado y financiado por la Facultad de Informática de la UNLP y

continuando la línea de trabajos anteriores de reconocimiento y localización de objetos en video, se está trabajando en el desarrollo de aplicaciones que combinen herramientas de inteligencia artificial como el Aprendizaje Profundo e interacción multimedial.

De esta forma, se busca utilizar la salida de modelos de reconocimiento de acciones y emociones, localización de personas y otros, como entradas a un sistema de generación de imágenes, videos y sonido de forma interactiva.

Para llevar esto a cabo, se están analizando diferentes modelos de generación de texto, audio e imágenes, basados en redes recurrentes y generativas [8][9].

## 2. LINEAS DE INVESTIGACIÓN Y DESARROLLO

- Definición de métricas de invarianza y equivarianza en modelos de redes neuronales
- Análisis de modelos de redes neuronales y sus características en términos de invarianza y auto-equivarianza.
- Representación y clasificación de configuraciones de manos para el lenguaje de señas.
- Generación de imágenes de formas de mano con GANs.
- Modelos de clasificación de imágenes de estrellas.
- Desarrollo de un sistema de procesamiento automático de placas espectrográficas astronómicas antiguas.
- Sistemas interactivos multimediales inteligentes.

## 3. RESULTADOS OBTENIDOS/ESPERADOS

- Comparación de modelos especializados para bases de datos con pocas muestras para la clasificación de formas de mano.

- Desarrollo de métricas de invarianza y auto-equivarianza para redes neuronales
- Análisis de los mecanismos de las redes neuronales para adquirir invarianza y autoequivarianza a las transformaciones afines.
- Redes generativas para la creación de datos artificiales en la Lengua de Señas.
- Modelo de clasificación de estrellas Be generalizable a distintos conjuntos de datos.
- Sistema de procesamiento automático de placas espectrográficas astronómicas.
- Prototipo de interacción multimedia utilizando IA.

#### 4. FORMACIÓN DE RECURSOS HUMANOS

El grupo de trabajo de la línea de I/D aquí presentada está formado por: 1 profesor con dedicación exclusiva, 1 investigador CIC-PBA, 1 becario de posgrado de la UNLP con dedicación docente, 1 becario CIC, 1 becario CIN, 5 tesistas, 3 profesores extranjeros, y 3 investigadores externos.

Dentro de los temas involucrados en esta línea de investigación, en los últimos dos años se han finalizado 2 tesis de doctorado, 1 tesis de especialización, y 7 tesinas de grado de Licenciatura.

Actualmente se están desarrollando 1 tesis de doctorado, 1 tesis de especialista, 3 tesinas de grado de Licenciatura y 4 trabajos finales de Ingeniería en Computación. También participan en el desarrollo de las tareas becarios y pasantes del III-LIDI.

#### 5. REFERENCIAS

[1] Quiroga, F., Antonio, R., Ronchetti, R., Lanzarini, L., Rosete, A. A Study of Convolutional Architectures for Handshape Recognition applied to Sign Language, publicado en el XXIII Congreso Argentino de Ciencias de la Computación (CACIC 2017) (pp. 13-22).

[2] Cornejo Fandos, U., Rios, G., Ronchetti, F., Quiroga, F., Hasperué, W., Lanzarini, L. Recognizing Handshapes using Small Datasets, publicado en el XXV Congreso Argentino de Ciencias de la Computación (CACIC 2019, Rio Cuarto) (pp. 105-114).

[3] Quiroga F., Ronchetti F., Lanzarini L., Fernandez-Bariviera A. Revisiting Data Augmentation for Rotational Invariance in Convolutional Neural Networks. International Conference on Modeling and Simulation in Engineering, Economics and Management (MS'2018 GIRONA).

[4] Quiroga, F., Torrents-Barrena, J., Lanzarini, L., & Puig, D. (2019, June). Measuring (in) variances in Convolutional Networks. In Conference on Cloud Computing and Big Data (pp. 98-109). Springer, Cham.

[5] Aidelman Y., Escudero C. Using Machine-Learning to obtain Be star candidates. IX La Plata International School (LAPIS) on Astronomy and Geophysics. Febrero 2020.

[6] Goodfellow I. J., Pouget-Abadie j., Mirza M., Xu B., Warde-Farley D., Ozair S., Courville A., Bengio Y. Generative Adversarial Networks. NIPS'14 Proceedings of the 27th International Conference on Neural Information Processing Systems. v2. pp 2672-2680. 2014.

[7] Jaschek M., Slettebak A., Jaschek C. *Be star terminology*. Be Star Newsletter. 1981.

[8] Vondrick C., Pirsiavash H., Torralba A. *Generating Videos with Scene Dynamics*. 29th Conference on Neural Information Processing Systems (NIPS 2016). 2016.

[9] Zhu Y., Lan Z., Newsam S., Hauptmann A. G.. *Hidden Two-Stream Convolutional Networks for Action Recognition*. Accepted at ACCV 2018. arXiv:1704.00389. 2017.