

Deep Learning Para Visión por Computadora.

Facundo Quiroga^{1,2}, Franco Ronchetti^{1,2}, Gastón Ríos^{1,3}, Pedro Dal Bianco^{1,3},
Santiago Ponte Ahon¹, Oscar Stanchi¹, Juan Seery¹, Tatiana Badaracco⁵, Federico Rabinovich⁵, Javier
Vazquez⁵, Aurelio Fernandez-Bariviera⁴, Saif Khalid⁴, Hatem Rashwan⁴, Domenec Puig Valls⁴, Yael Aidelman,
Roberto Gamen, Laura Lanzarini¹, Waldo Hasperué¹

¹ Instituto de Investigación en Informática LIDI, Facultad de Informática, Universidad Nacional de La Plata, La Plata, Argentina.*

² Comisión de Investigaciones Científicas de la Pcia. De Bs. As. (CICPBA)

³ Becario postgrado UNLP

⁴ Universitat Rovira i Virgili, Tarragona, España

⁵ Universidad de Buenos Aires, Buenos Aires, Argentina

⁶ Instituto de Astronomía de La Plata, Universidad Nacional de La Plata, La Plata, Argentina.

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

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

CONTEXTO

Esta presentación corresponde a algunas de las tareas de investigación que se llevan a cabo en el III-LIDI en el marco del proyecto “Inteligencia de Datos. Técnicas y Modelos de Machine Learning” perteneciente al Programa de Incentivos (2023-2026).

- Predicción de series temporales, incluyendo series de precios de acciones de la bolsa.

Palabras clave: Redes Neuronales, Redes Convolucionales, Redes Recurrentes, Visión por Computadoras, Lengua de Señas, Bases de datos, *Crowdsourcing*, Redes Generativas Adversarias, Invarianza, Equivarianza.

RESUMEN

Esta línea de investigación se centra en el estudio y desarrollo de sistemas y modelos de Deep Learning para resolución de problemas de Visión por computadora, incluyendo técnicas de Aprendizaje Automático clásicas, y Aprendizaje profundo con Redes Neuronales Convolucionales (CNNs), Recurrentes (RNNs) y Transformers. En este contexto, los autores del presente trabajo aplican estas técnicas a diversos problemas:

- Traducción de la Lengua de Señas (TLS), en particular la Lengua de Señas Argentina (LSA).
- Interpretabilidad de modelos de Redes Neuronales, incluyendo análisis de invarianza y equivarianza y de filtros convolucionales, con énfasis en modelos que trabajan con imágenes médicas.
- Análisis y procesamiento de datos astronómicos, enfocados en la recuperación de trabajos de observación históricos.

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. Traducción de lengua de señas

La traducción de la lengua de señas (TLS) es un campo de estudio activo que abarca la interacción humano-computadora, la visión por computadora, el procesamiento de lenguaje natural y el aprendizaje automático.

Este es un problema complejo y multidisciplinar, que presenta diversos subproblemas a resolver como el reconocimiento del intérprete que realiza una seña, la segmentación de manos, la clasificación de diferentes configuraciones y de un gesto dinámico, entre otros.

La falta de datos es la dificultad principal para lograr modelos de TLS de alto desempeño.

1.1.1 LSA-T: Base de datos de LSA para TLS

La creación de un conjunto de datos apropiado para la LSA resulta esencial para la creación de modelos específicos. Por este motivo, se finalizó con el desarrollo y la validación de la base de datos LSA-T [1]. la más grande al momento con señas de la LSA y una de las más grandes del mundo en general. Además, se puso en marcha la plataforma **seni.ar** para permitir que la comunidad sorda contribuya con nuevos videos de LSA, y agregue información sobre los existentes.

En base a esta base de datos, esta línea de investigación continúa mejorando los resultados con modelos básicos obtenidos previamente, y busca paliar la falta de datos de dos formas:

- En el dominio de la entrada, mediante videos de otras lenguas de señas que si bien no comparten la semántica si comparten poses y movimientos
- En el dominio de la salida, mediante la aumentación de datos con modelos grandes de lenguajes (LLMs) para generar traducciones alternativas.

1.1.2 Generación de imágenes, videos y poses de la LSA

Se desarrollaron modelos basados en Redes Generativas Adversarias (GANs) para la generación de imágenes de formas de mano que puedan asistir a entrenar o preentrenar modelos de TLS para la LSA. Los modelos obtenidos permiten condicionar al generador

utilizando la pose mediante un módulo SPADE, y de esta forma generar imágenes de formas de mano arbitrarias y mejoran el desempeño de los modelos



7. Figura X. Comparación de imágenes reales y generadas con GANs.

Esta línea de investigación se continúa actualmente buscando generar secuencias de poses o videos de señas completas para asistir en el entrenamiento de modelos de TLS para la LSA.

1.2 Interpretabilidad

Las redes neuronales son modelos tradicionalmente considerados como de caja negra. En los años recientes, se han realizado varios esfuerzos para comprender su funcionamiento de forma tal que el mismo sea más predecible o modulable.

1.2.1 Modelos interpretables de predicción de calidad de imágenes de fondos de ojos.

La retinopatía diabética es una enfermedad aguda pero que puede tratarse con detección temprana. Para ello, sería de gran utilidad contar con modelos que permitan automatizar el análisis de imágenes de fondo de ojos. En particular, sería útil tener un modelo que permita determinar la calidad de una imagen para asistir la captura de las mismas en tiempo real, y además que sea interpretable para poder otorgar feedback al usuario. Por este motivo, se presentó un modelo basado en la CNN UNet de

predicción de calidad de imagen que se denominó VISTA. El modelo contiene un módulo de Concept Whitening, una técnica de interpretabilidad intrínseca, que indica en base a qué conceptos la red realiza sus predicciones [3].

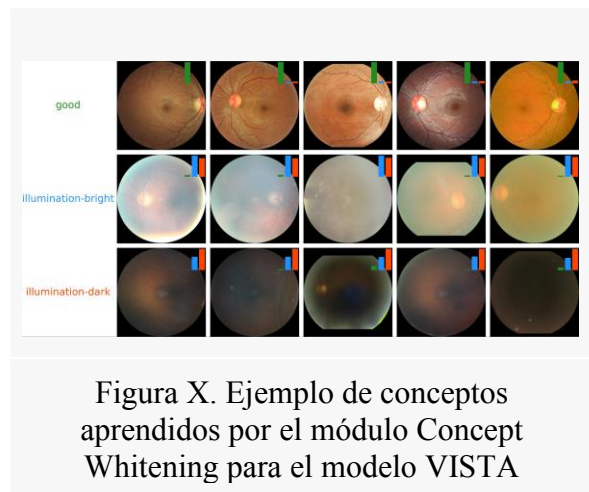


Figura X. Ejemplo de conceptos aprendidos por el módulo Concept Whitening para el modelo VISTA

1.2.2 Invarianza y Equivarianza de CNNs conocidas

La invarianza y equivarianza a las transformaciones son propiedades deseables en varios modelos de redes debido a que nos permiten razonar más fácilmente respecto a su funcionamiento. Por este motivo, en base a métricas previamente definidas por este grupo de investigación [4], se está trabajando en la caracterización de modelos de redes neuronales reconocidos. Se está trabajando con los modelos EfficientNet y ViT para comprender la distribución de su invarianza en sus distintas capas, y cómo distintas estrategias de entrenamiento pueden afectar dicha distribución (Figura 1).

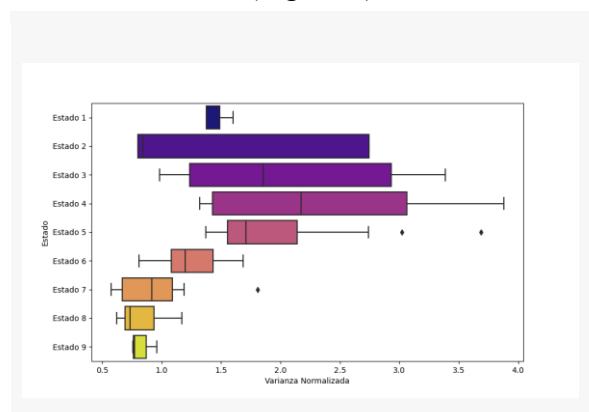


Figura X. Varianza normalizada de los 9 bloques de la CNN EfficientNetV2-L.

1.2.3 Interpretabilidad de filtros convolucionales mediante entrenamiento no supervisado de CNNs

Los filtros convolucionales de una red entrenada con buen desempeño suelen usarse para realizar transferencia de aprendizaje. No obstante, su interpretabilidad depende del proceso de entrenamiento y requiere un análisis post-hoc. Por este motivo, se desarrolló un método de entrenamiento de filtros convolucionales basado en clustering mediante k-medias o componentes principales. Los filtros obtenidos permiten un desempeño igual o mejor que el de las redes entrenadas tradicionalmente con menos iteraciones de descenso de gradiente [6]



Figura X. Los filtros inicializados mediante el método propuesto requieren pocos cambios durante el entrenamiento.

1.3 Predicción de precios de acciones

Los modelos de predicción de precios de una acción particular suelen entrenarse solo con datos de una acción. Alternativamente, pueden incluirse datos de otras acciones o del mercado en general. Mediante un enfoque de clustering de series de precios y de balances, se determinaron acciones accesorias a otras para asistir en el entrenamiento de un modelo de predicción de precios. Mediante dicho modelo, basado en RNN, se obtuvo un algoritmo de inversión con una ganancia trimestral promedio del 7% en 2021 y 2022 [7].

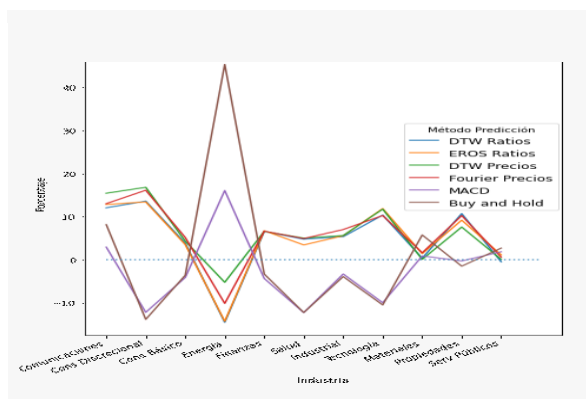


Figura X. Rendimientos de inversión con las predicciones de la red, por segmento, y para cada esquema de clustering/inversión.

1.4 Sistema de análisis de placas espectrográficas antiguas

Continuando el desarrollo de un sistema de análisis automatizado de placas espectrográficas del proyecto ReTroH[2], se está desarrollando el módulo de alineamiento de lámparas de comparación empíricas y teóricas. Se experimentó con varios métodos de base para ser comparados con los métodos a desarrollar basados en CNNs.

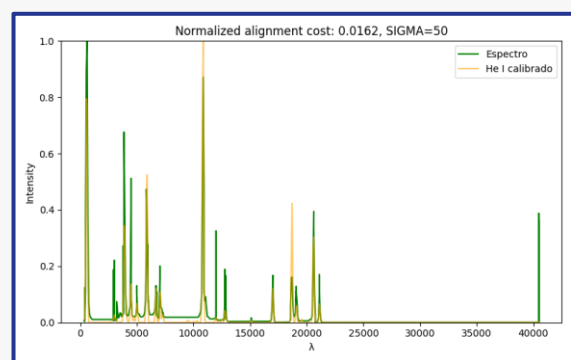


Figura X. Alineamiento entre lámpara de Helio empírica y teórica

2. LINEAS DE INVESTIGACIÓN Y DESARROLLO

- Redes neuronales profundas, convolucionales y transformers.
- Traducción de Lengua de Señas.

- Generación de imágenes y poses con GANs para paliar la falta de datos y desbalance de clases en Lengua de Señas
- Herramientas de interpretabilidad para visión por computadora
- Interpretabilidad y entrenamiento no supervisado de Redes Convolucionales.
- Invarianza y auto-equivarianza en redes neuronales, en particular de modelos ampliamente difundidos.
- Mejora de la predicción de series temporales con Redes Neuronales profundas mediante agrupamiento de los datos de entrenamiento.

3. RESULTADOS OBTENIDOS/ESPERADO

- Generador de imágenes de formas de mano basado en poses generalizable a cualquier conjunto de datos y poses.
- Interpretabilidad mediante invarianza de modelos EfficientNet y ViT
- Validación mediante métodos de interpretabilidad de un modelo de predicción de calidad de imágenes de retinopatía diabética basado en UNet.
- RNN para predicción de precios de acciones con una ganancia estimada de 7% trimestral de 2017 a 2022.
- Ampliación de LSA-T mediante sistema web de Crowdsourcing y métodos de aumentación de datos mediante LLMs.
- Algoritmo de alineamiento de lámparas de comparación en imágenes de espectros históricos basado redes neuronales.

4. FORMACIÓN DE RECURSOS HUMANOS

El grupo de trabajo de la línea de I/D aquí presentada está formado por: 3 profesores con dedicación exclusiva, un JTP dedicación

exclusiva, 2 investigadores CIC-PBA, 4 becarios de posgrado de la UNLP con dedicación docente, 3 tesistas de maestría de la UBA, y 4 profesores extranjeros.

Actualmente se están desarrollando 4 tesis de doctorado y 3 de magíster. También participan en el desarrollo de las tareas becarios y pasantes del III-LIDI.

5. REFERENCIAS

- [1] Dal Bianco, P., Ríos, G., Ronchetti, F., Quiroga, F., Stanchi, O., Hasperué, W., & Rosete, A. (2023, January). LSA-T: The First Continuous Argentinian Sign Language Dataset for Sign Language Translation. In *Advances in Artificial Intelligence—IBERAMIA 2022: 17th Ibero-American Conference on AI*, Cartagena de Indias, Colombia, November 23–25, 2022, Proceedings (pp. 293-304). Cham: Springer International Publishing.
- [2] Ríos, G., Dal Bianco, P., Ronchetti, F., Quiroga, F., Stanchi, O., Hasperué, W. Generación de gestos de lengua de señas con redes neuronales generativas basadas en poses y etiquetas. CACIC 2023, Junin, Buenos Aires
- [3] Khalid, S., Rashwan, H. A., Abdulwahab, S., Abdel-Nasser, M., Quiroga, F. M., & Puig, D. (2024). FGR-Net: Interpretable fundus image gradeability classification based on deep reconstruction learning. *Expert Systems With Applications*, 238, 121644.
- [4] Quiroga, F., Torrents-Barrena, J., Lanzarini, L., & Puig, D. Measuring (in) variances in Convolutional Networks. In *Conference on Cloud Computing and Big Data* (pp. 98-109). Springer, Cham. 2019.
- [6] Rabinovich, F. Métodos de Entrenamiento basados en Agrupamiento para Capas Convolucionales en Redes Neuronales. Tesis de Maestría, UBA. Directores: Quiroga, F, Ronchetti, R.
- [7] Sáenz, J. V., Quiroga, F. M., & Bariviera, A. F. (2023). Data vs. information: Using clustering techniques to enhance stock returns forecasting. *International Review of Financial Analysis*, 88, 102657.
- [8] Pereyra, N., Ponte Ahon, S., Aidelman, Y. J., Ronchetti, F., Quiroga, F., Gamen, R., & Cidale, L. (2023). Software de procesamiento automático de placas espectrográficas. *Boletín de la Asociación Argentina de Astronomía La Plata Argentina*, 64, 332-334.