











Deep Learning. Aplicaciones en Reconocimiento de Lengua de Señas, Generación de Lenguaje Natural e Imágenes Astronómicas.

F. Ronchetti^{1,2,3} , F. Quiroga^{1,2,4} , G. Rios^{1,4} , P. Dal Bianco¹ , I. Mindlin¹ , L. Lanzarini^{1,2} , A. Rosete⁶ , R. Gamen⁷ , D. Puig Valls⁹ , J. Torrents-Barrena⁹ , Y. Aidelman^{7,8}, C. Escudero⁸, A. Granada¹⁰.

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

² Facultad de Informática, Universidad Nacional de La Plata, La Plata, Argentina

³ Comisión de Investigaciones Científicas de la Pcia. De Bs. As. (CIC)

⁴ Becario postgrado UNLP ⁵ Becario CIN

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

⁷ Facultad de Ciencias Astronómicas y Geofísicas, Universidad Nacional de La Plata, La Plata, Argentina

⁸ Instituto de Astrofísica de La Plata (IALP CONICET), La Plata, Argentina

⁹ Universitat Rovira i Virgili (URV), Tarragona, España

¹⁰ Universidad Nacional de Río Negro (UNRN), Río Negro, 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 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, video y texto, 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 y procesamiento de lenguaje natural.

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 utilizando Redes Recurrentes para reconocer señas dinámicas, utilizando la base de datos LSA64 de Lengua de Señas Argentina [2].

Por último, se están utilizando 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. 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.

Por último, se creó un modelo neuronal capaz de generar texto artificial que se adapte al género *freestyle*.

Palabras clave: Redes Neuronales, Redes Convolucionales, Redes Recurrentes, Visión por Computadoras, Lengua de Señas, Redes Generativas Adversarias, Invarianza, Equivarianza, Imágenes Astronómicas.

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. Librería para experimentación con Formas de Mano de 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.

En un estudio previo, se realizaron experimentos comparando diversas arquitecturas basadas en redes neuronales para clasificar formas de manos [1] pero advirtiendo que el desempeño de los modelos utilizados estuvo limitado por la falta de datos.

Por este motivo, se desarrolló una librería para facilitar la descarga y combinación de distintas bases de datos de formas de mano con el objetivo de mejorar los sistemas reconocedores de lengua de señas.



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

Dicha librería, de código abierto y disponible públicamente¹, permite la descarga de diez conjuntos de datos de imágenes de formas de mano, e incluye metadatos que indica la correspondencia entre las formas de mano de las distintas bases de datos. De este modo, se logra simplificar y unificar las tareas de preprocesamiento y la comparación de distintas bases de datos, así como la transferencia de aprendizaje entre ellas.

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].

Por otro lado, las bases de datos de Lengua de Señas, dada su compleja naturaleza, 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 artificiales relacionadas con la lengua de señas. Este tipo de redes permitirá aumentar

¹https://github.com/midusi/handshape_datasets

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. Reconocimiento de Lengua de Señas con Redes Recurrentes y Convolucionales

En esta línea de investigación, el objetivo es la clasificación de gestos en videos de Lengua de Señas. En particular, estamos trabajando con señas de la base de datos LSA64. Esta consiste en un registro de 64 señas de la Lengua de Señas Argentina. Para la tarea se compararon diversas técnicas en el estado del arte del Aprendizaje Automático basadas en Redes Neuronales. Específicamente, se compararon tanto arquitecturas basadas en Redes Recurrentes y Convolucionales, como distintas estrategias de preprocesamiento para optimizar la calidad del reconocimiento. También se están analizando los modelos entrenados para comprender mejor el impacto de estas estrategias de preprocesamiento y de la forma de representación lograda por cada tipo de arquitectura.



Figura 2: Activaciones de las capas convolucionales de una Red Recurrente para reconocer señas en video.

1.4. Métricas de Equivarianza

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.

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.

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 continua con la utilización de las métricas previamente definidas [4] para caracterizar modelos de redes neuronales típicos, ya sea desde capas muy utilizadas como Batch Normalization, Dropout, Max-Pooling, arquitecturas completas como Residual Networks, VGG o AllConvolutional, y arquitecturas especializadas como TI-Pooling.

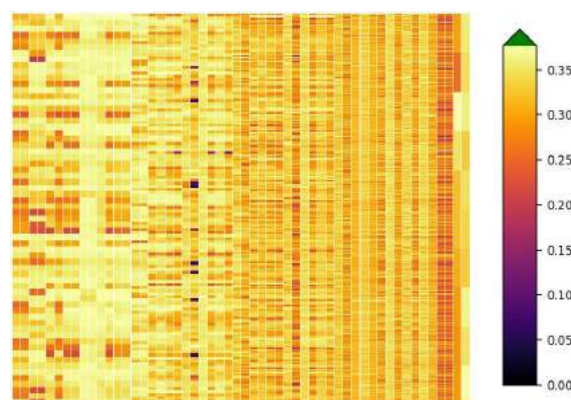


Figura 3. Invarianza por capas y unidades de una CNN con arquitectura ResNet.

1.5. 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 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 $H\alpha$. Para el análisis de muchos objetos astronómicos, los datos fotométricos son relativamente más fáciles de obtener debido al menor tiempo de uso del telescopio. Por lo tanto, existe una necesidad creciente de utilizar información fotométrica para identificar automáticamente objetos para estudios más detallados, especialmente estrellas con líneas de emisión $H\alpha$ como las estrellas Be. En este trabajo se evaluó el uso de redes neuronales para identificar candidatos a estrella Be a partir de un conjunto de estrellas tipo OB. Las redes se entrenaron utilizando un subconjunto etiquetado de las bases de datos VPHAS + y 2MASS, con filtros u, g, r, $H\alpha$, i, J, H y K. Para evitar el efecto del enrojecimiento, se propuso y se evaluó el uso de índices Q para mejorar la generalización del modelo a otras bases de datos. Para validar el enfoque, se etiquetó manualmente un subconjunto de la base de datos y se usó para evaluar modelos de identificación de candidatos. También se etiquetó un conjunto independiente de datos utilizando validación cruzada [7]. Los modelos entrenados lograron obtener un 25% de Sensibilidad (*Recall*) fijando la precisión (*Precision*) al 99%. Actualmente se está trabajando en agregar nuevas bases de datos de otros investigadores y así como otras áreas del cielo.

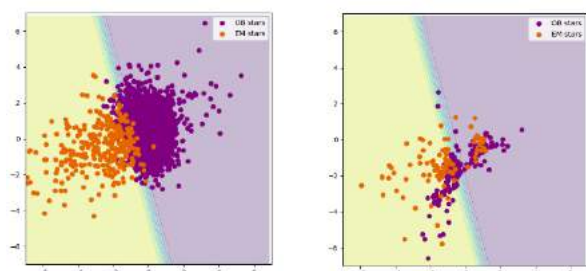


Figura 4. Reducción a 2 features para los conjuntos de datos de Mohr-Smith (izquierda) y Liu (derecha). Las líneas de decisión corresponden a una función sigmoidea de un regresor logístico entrenado en Mohr-Smith.

1.6. Generación de letras de rap con Redes Neuronales Recurrentes

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 de patrones utilizando Deep Learning, se realizó un proyecto de generación automática de letras de rap con estilo freestyle.

Para llevar esto a cabo, se analizaron diferentes modelos de generación de texto basados en redes recurrentes y generativas [8][9]. Se confeccionaron dos bases de datos específicas de este tipo de letras en español, inexistentes hasta la fecha. Se analizaron diferentes sistemas de preprocesamiento del texto incluyendo tokenización a nivel de palabra, de sílaba y el algoritmo BPE. Se realizaron diversos experimentos con redes recurrentes bidireccionales tipo LSTM y se realizó un diccionario específico de rimas en español para controlar cómo el texto las genera.

2. LINEAS DE INVESTIGACIÓN Y DESARROLLO

- Redes neuronales profundas, convolucionales y recurrentes.
- Invarianza y auto-equivarianza en redes neuronales.
- Reconocimiento de lenguaje de señas.
- Generación de imágenes con GANs.
- Análisis de datos astronómicos.
- Generación de lenguaje natural.

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 modelos para el reconocimiento de lengua de señas en video.

- Librería para facilitar la experimentación con imágenes de formas de manos para el lenguaje de señas.
- 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.
- Modelo generativo de texto artificial para batallas de *freestyle*.

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, 2 becarios de posgrado de la UNLP con dedicación docente, 1 becario CIC, 1 becario CIN, 6 tesis, 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, 2 tesis de especialización, y 5 tesinas de grado de Licenciatura.

Actualmente se están desarrollando 2 tesis de doctorado, 1 tesis de especialista, 5 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., Ronchetti F., Quiroga F., Lanzarini L. Reddening-Free Q Indices to Identify Be Star Candidates. Cloud Computing, Big Data & Emerging Topics. pp 111-123. Springer International Publishing. 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] Potash, P., Romanov, A., and Rumshisky, A. Ghostwriter: Using an lstm for automatic rap lyric generation. In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, 1919-1924. 2015.
- [9] Young, T., Hazarika, D., Poria, S., and Cambria, E. Recent trends in deep learning based natural language processing. *iee Computational intelligence magazine*, 13(3):55. 2018.