

Mar del Plata, 25 de octubre de 2019.-

RESOLUCIÓN DEL RECTORADO N° 382/19

VISTO:

El proyecto de investigación "*Inteligencia Artificial en FPGA. IA-FPGA*" presentado por los docentes de la Facultad de Ingeniería, Ing. Martín Osvaldo Vázquez, Ing. Bruno Eduardo Nicolás Constanzo, Mg. Ing. Elías Lucas Leiva, Ing. Gonzalo Matías Ruiz de Angeli, Mg. Jordina Torrens Barrena, TI Denis Adriel Chambi y Sr. Ayrton Betti tramitado bajo el expediente de Investigación 201-2019; y

CONSIDERANDO:

La presentación formal del proyecto se hizo de acuerdo a lo establecido en el art. 31 de la Resolución de Rectorado N° 053/18 mediante expediente abierto a solicitud de la Secretaria de Investigación de la Facultad de Ingeniería, con fecha 22 de octubre de 2019;

La evaluación del especialista Dr. Ing. Silvano Rossi (Fs. 20 – 22) y del Dr. Ruben Wainschenker (Fs. 23 – 24) del expediente 201-2019);

Que la Secretaria de Investigación de la Universidad FASTA, la Secretaria de Investigación de la Facultad de Ingeniería y el Decano de la Facultad de Ingeniería, han propuesto al Ing. Martín Osvaldo Vázquez como Director e Investigador Adjunto categorizado por la UFASTA;

Que el investigador propuesto acredita una dilatada y rica trayectoria científica, profesional y académica que permite destacar su testimonio y vocación de servicio y que cualifica indudablemente el claustro universitario;

Que la Secretaria de Investigación de la Universidad FASTA, la Secretaria de Investigación de la Facultad de Ingeniería y el Decano de la Facultad de Ingeniería, han propuesto al Ing. Bruno Eduardo Nicolás Constanzo como Investigador Adjunto categorizado por la UFASTA;

Que el investigador propuesto acredita una dilatada y rica trayectoria científica, profesional y académica que permite destacar su testimonio y vocación de servicio y que cualifica indudablemente el claustro universitario;

Que la Secretaria de Investigación de la Universidad FASTA, la Secretaria de Investigación de la Facultad de Ingeniería y el Decano de la Facultad de Ingeniería, han propuesto como auxiliares de investigación al Mg. Ing. Elías Lucas Leiva, Ing. Gonzalo Matías Ruiz de Angeli, Mg. Jordina Torrens Barrena, TI Denis Adriel Chambi y Sr. Ayrton Betti;



Lo dispuesto por la Resolución del Rectorado N° 053/18, en sus artículos 8, 9, 10, 17 y concordantes;
Por ello, y en uso de las atribuciones que le confieren los Arts. 28° inc. d) y concordantes del Estatuto Universitario

**EL RECTOR DE LA UNIVERSIDAD FASTA
DE LA FRATERNIDAD DE AGRUPACIONES SANTO TOMÁS DE AQUINO**

R E S U E L V E :

Artículo 1°: Aprobar el proyecto de investigación “*Inteligencia Artificial en FPGA. IA-FPGA*” que se adjunta en Anexo I de a presente.-

Artículo 2°: Designar al **Ing. Martín Osvaldo VÁZQUEZ** (DNI 23.617.506) como Director e Investigador Adjunto categorizado por la UFASTA del Proyecto “*Inteligencia Artificial en FPGA. IA-FPGA*” por el término de 30 meses a partir del 1° de noviembre de 2019, según los alcances del art. 32 y cc. de la Resolución Rectoral N° 053/18.-

Artículo 3°: Designar al **Ing. Bruno Eduardo Nicolás CONSTANZO** (DNI 33.566.742) como Investigador Adjunto categorizado por la UFASTA del Proyecto “*Inteligencia Artificial en FPGA. IA-FPGA*” por el término de 30 meses a partir del 1° de noviembre de 2019.-

Artículo 4°: Designar al **Mg. Ing. Elías Lucas LEIVA** (DNI 27.539.740) como Auxiliar de Investigación Graduado categorizado por la UFASTA del Proyecto “*Inteligencia Artificial en FPGA. IA-FPGA*” por el término de 30 meses a partir del 1° de noviembre de 2019.-

Artículo 5°: Designar al **Ing. Gonzalo Matías RUIZ DE ANGELI** (DNI 31.821.160) como Auxiliar de Investigación Graduado categorizado por la UFASTA del Proyecto “*Inteligencia Artificial en FPGA. IA-FPGA*” por el término de 30 meses a partir del 1° de noviembre de 2019.-

Artículo 6°: Designar a la **Mg. Jordina TORRENS BARRENA** (DNI 48.014.960-E) como Auxiliar de Investigación Graduado categorizado por la UFASTA del Proyecto “*Inteligencia Artificial en FPGA. IA-FPGA*” por el término de 30 meses a partir del 1° de noviembre de 2019.-



Artículo 7°: Designar al **TI Denis Adriel CHAMBI** (DNI 32.875.131) como Auxiliar de Investigación Alumno categorizado por la UFASTA del Proyecto “*Inteligencia Artificial en FPGA. IA-FPGA*” por el término de 30 meses a partir del 1° de noviembre de 2019.-

Artículo 8°: Designar al **Sr. Ayrton BETTI** (DNI 40.373.776) como Auxiliar de Investigación Alumno Externo categorizado por la UFASTA del Proyecto “*Inteligencia Artificial en FPGA. IA-FPGA*” por el término de 30 meses a partir del 1° de noviembre de 2019.-

Artículo 9° Dése a conocer, remítase copia a la Secretaría de Investigación de la Universidad FASTA, al Decanato de la Facultad de Ingeniería y a los investigadores designados, archívese.



PROF. MARCELA S. GRECA DE GASCÓN
SECRETARÍA GENERAL
UNIVERSIDAD FASTA



DR. JUAN CARLOS MENA
RECTOR
UNIVERSIDAD FASTA

ANEXO

RESOLUCIÓN DEL RECTORADO N° 382/19

Proyecto de Investigación “*Inteligencia Artificial en FPGA. IA-FPGA*”

”

1. EL PROYECTO DE INVESTIGACIÓN

NO. DE PRESENTACIÓN DE PROYECTO:	
TÍTULO O DENOMINACIÓN DEL PROYECTO INTELIGENCIA ARTIFICIAL EN FPGA	
ACRÓNIMO IA- FPGA	
MES Y AÑO DE INICIO: 10/2019	
MES Y AÑO DE FINALIZACIÓN: 03/2022	
ÁREA DE CONOCIMIENTO ⁽ⁱ⁾ : INGENIERÍAS Y TECNOLOGÍAS	
SUB-ÁREA DE CONOCIMIENTO ⁽ⁱⁱ⁾ : HARDWARE Y ARQUITECTURA DE COMPUTADORAS	

2. INSTITUCIONES PARTICIPANTES

INSTITUCIÓN/ES EJECUTORA/S DEL PROYECTO: UNIVERSIDAD FASTA
CENTRO/S DE INVESTIGACIÓN EJECUTOR/ES: GRUPO DE INVESTIGACIÓN SISTEMAS EMBEBIDOS
INSTITUCIÓN QUE PRESENTA EL PROYECTO: UNIVERSIDAD FASTA
ENTIDAD/ES FINANCIADORA/S DEL PROYECTO: UNIVERSIDAD FASTA
INSTITUCIÓN/ES ADOPTANTE/S DEL PROYECTO: NO CORRESPONDE
INSTITUCIÓN/ES DEMANDANTE/S DEL PROYECTO (si la/s hubiera): NO CORRESPONDE
INSTITUCIÓN/ES PROMOTORA/S DEL PROYECTO (si la/s hubiera): NO CORRESPONDE

3. DIRECTOR

NOMBRE Y APELLIDO DEL DIRECTOR DEL PROYECTO: MARTÍN OSVALDO VÁZQUEZ
DIRECCIÓN DE CONTACTO DEL DIRECTOR: movazquez@ufasta.edu.ar ; martin.o.vazquez@gmail.com
NOMBRE Y APELLIDO DEL CO-DIRECTOR: BRUNO CONSTANZO
DIRECCIÓN DE CONTACTO DEL CO-DIRECTOR: bconstanzo@ufasta.edu.ar

4. EQUIPO DE TRABAJO

NOMBRE Y APELLIDO	INSTITUCIÓN –UNIDAD/ES ACADÉMICA/S	FUNCIÓN
MG. LUCAS LEIVA	FACULTAD DE INGENIERÍA	AUX. INV. GRADUADO
MG. JORDINA TORRENTS-BARRENA	FACULTAD DE INGENIERÍA	AUX. INV. GRADUADO
MG. GONZALO RUIZ DE ANGELIS	FACULTAD DE INGENIERÍA	AUX. INV. GRADUADO
ADRIEL CHAMBI	FACULTAD DE INGENIERÍA	AUX. INV. ALUMNO
AYRTON BETTI	FACULTAD DE INGENIERÍA DE UNMDP	AUX. INV. ALUMNO EXTERNO

5. CARACTERIZACIÓN DEL PROYECTO
PROBLEMA O NECESIDAD A RESOLVER

El modelo de Aprendizaje Profundo de Redes Neuronales Convolucionales (CNN) ha empujado los límites de varias tareas de Inteligencia Artificial. En algoritmos típicos de Machine Learning (ML), las CNN involucran dos fases: una fase de entrenamiento y otra fase de inferencia o clasificación. En los últimos años han surgido diversos frameworks que favorecen el diseño e implementación de sistemas de Inteligencia Artificial, como PyTorch, TensorFlow, Lasagne basado en Theano , Keras, y Chainer, entre otros. Estos frameworks proporcionan unidades de red neuronal, funciones de costo y optimizadores para ensamblar y entrenar modelos.

Se aprecian dos aspectos fundamentales a considerar en cuanto al desarrollo involucrado en CNN. Por un lado la aceleración hardware del entrenamiento de la red. Por otro lado, en el caso que se requiera, la implementación de una solución final completa en un sistema embebido, el cuál debe considerar restricciones duras de tiempo real, ocupación de área,



consumo de potencia energética, etc.

El empleo de la tecnología FPGA (Field Programmable Gate Array) aparece como una buena alternativa en el momento de utilizarse como plataforma para la implementación de arquitecturas de redes neuronales complejas, o para ser utilizadas como co-procesadores hardware dedicados, ejecutando las rutinas críticas de los sistemas de inteligencia artificial que dominan actualmente el mercado.

PRODUCTO O PROCESO A GENERAR

Si bien es cierto que las principales compañías de FPGAs ofrecen algunas herramientas que posibilitan el desarrollo de CNNs en FPGAs, como así también existen soluciones publicadas desde el ámbito científico y académico, también es cierto que hay mucho por desarrollar y aportar en este campo. En este proyecto se espera establecer un relevamiento y evaluación de las herramientas existentes para el desarrollo de inferencia de Deep Learning, y poder contribuir con el desarrollo de técnicas y mecanismos para el diseño e implementación de redes CNN en plataformas FPGAs para cualquiera de los dos propósitos fundamentales establecidos, estos son: i) aceleración hardware del entrenamiento, ó ii) solución final embebida dedicada a la aplicación.

No solo se validarán y evaluarán las soluciones existentes en el mercado y en el mundo académico en esta tecnología, sino además se generará nuevos diseños e implementaciones.

RESUMEN, DETALLANDO OBJETIVOS Y ACTIVIDADES DEL PROYECTO

El interés primario está en la implementación en el ámbito Universitario de actividades de desarrollo tecnológico de punta. El interés secundario es que se desarrollen en Argentina industrias vinculadas con la electrónica y el diseño de sistemas digitales, en las principales tendencias globales de investigación. En paralelo se trata de formar recursos humanos genuinos a nivel de investigación y postgrado en la Universidad FASTA.

Durante la ejecución del proyecto se implementarán arquitecturas hardware que permitan la aceleración de técnicas de inteligencia artificial, principalmente de Deep Learning basado en CNN. Dentro de las características se pretende:

- Obtener tiempos de ejecuciones significativamente inferiores a alternativas software (10-20x)
- Permitir la ejecución en tiempo real
- Implementar precisiones parametrizables, a medida de la aplicación
- Alcanzar un ahorro en el consumo de energía
- Permitir la ejecución en sistemas embebidos y/o portables.
- Aplicar optimizaciones que se adecúen a la plataforma destino.

Como objetivos particulares se encuentra el diseño e implementación de:



- Plataformas de Inteligencia Artificial basadas en FPGAs de Xilinx e Intel FPGA.
- Co-procesadores hardware basados en lógica programable, que aceleren las técnicas de Inteligencia Artificial basadas en Deep Learning, tanto en las fases de entrenamiento, como en clasificación.
- Arquitecturas de CNN aplicadas a casos de estudio particulares.
- Generalización de las arquitecturas e integración con Frameworks de Deep Learning.

El proyecto comenzará con el perfeccionamiento de los integrantes del grupo en las tecnologías de síntesis en los ambientes de diseño provistos por los fabricantes de la tecnología objetivo.

Luego se estudiarán técnicas de implementación de técnicas de Deep Learning, particularmente implementación hardware de redes CNN. En este período se planea profundizar los estudios realizados en proyectos anteriores, y fortalecer las vinculación con otros institutos de investigación

Se profundizará el análisis de los algoritmos utilizados en Deep Learning y se aplicarán las arquitecturas generadas a casos de estudios particulares.

A continuación se generalizarán de las arquitecturas implementadas e integrarán con Frameworks de Deep Learning.

Por último, los detalles de las implementaciones se consignaran en informes y eventualmente publicarán y/o presentarán en revistas y congresos afines.

NOVEDAD U ORIGINALIDAD LOCAL EN EL CONOCIMIENTO

En el ámbito académico argentino no se han desarrollado demasiados proyectos sobre sistemas digitales. Asimismo, si bien el uso de técnicas de Deep Learning ha sido objeto de un gran interés por investigadores a nivel nacional, no se ha puesto especial énfasis en su aceleración. Por otro lado, la aplicación de la tecnología FPGA para el desarrollo de coprocesadores aceleradores y/o sistemas embebidos en el dominio de Deep Learning, se encuentra en un estado muy insipiente a nivel nacional, con lo cuál presenta una gran oportunidad para realizar contribuciones y ofrecer soluciones innovadoras.

En FASTA este proyecto es innovador y apunta a sensibilizar los recursos humanos locales acerca de la apertura futura de la Argentina en los mercados de productos vinculados a la aceleración de aplicaciones cómputo intensivo, a los sistemas embebidos y a los sistemas digitales. Se espera que el conocimiento generado durante la ejecución del proyecto, permita generalizar una plataforma de soporte de inteligencia artificial (particularmente CNN) en



hardware aplicable en diferentes campos de aplicación, resultando un producto conveniente a otros investigadores y empresas

GRADO DE RELEVANCIA

La Inteligencia Artificial (IA) ha tenido, desde hace algunos años, un gran impacto en la sociedad y en la economía, entre otras áreas. Hoy en día muchas organizaciones utilizan estas herramientas para apalancar su crecimiento. Según algunos reportes de consultoras, la utilización de IA podría resultar en un incremento de más del 35% de producto bruto interno de algunos países.

En el contexto nacional, se cuentan con varias iniciativas orientadas a mejorar la formación de recursos humanos en la informática, como son: la reciente Ley De Economía del Conocimiento¹ para aumentar la productividad y la generación de empleo de calidad, que incluye a las actividades relacionadas con el software; y el plan industria 4.0² para favorecer la competitividad del sector productivo a través de la incorporación de tecnologías emergentes como la IA, entre otras. En particular se desea mencionar el Plan Nacional de Inteligencia Artificial, que es un plan estratégico que se encuentra en la Agenda Digital 2030³.

Por otro lado, en los países en vía de desarrollo como es el caso de Argentina, la utilización de tecnologías como FPGAs para el desarrollo de hardware se presenta como una alternativa muy auspiciosa. Las razones fundamentales se encuentran asociadas a los costos en comparación con desarrollos ASICs, tanto en lo que respecta al chip adquirido como también al diseño de estos circuitos. Los FPGAs son una tecnología accesible que ofrece la posibilidad de realizar implementaciones a partir de descripciones basadas en software.

GRADO DE PERTINENCIA

La propuesta se orienta al diseño de circuitos digitales y sistemas embebidos en FPGAs, para la implementación de diferentes funciones que intervienen en las redes CNN. Dichas funciones se implementan típicamente mediante software con prestaciones en tiempo bastante inferiores a las soluciones hardware. Las plataformas que se proyectan utilizar son las pertenecientes a los principales fabricantes, tales como Xilinx e Intel-Altera. Estas compañías poseen programas universitarios en donde ofrecen donaciones de productos y herramientas para su uso gratuito en el ámbito académico.

¹ <https://www.argentina.gob.ar/noticias/el-congreso-aprobo-la-ley-que-promueve-la-economia-del-conocimiento>

² <https://www.argentina.gob.ar/noticias/el-gobierno-creo-el-plan-industria-argentina-40-para-favorecer-la-transformacion>

³ <https://www.casarsoda.gob.ar/informacion/actividad-oficial/9-noticias/44081-el-gobierno-presento-lanueva-agenda-digital-2030>

Esta tecnología posibilita tener rendimientos compatibles con especificaciones de tiempo real existente, usando los dispositivos de lógica programable como coprocesadores hardware ad-hoc, o bien como una solución final completa de un sistema embebido.

Los integrantes del proyecto poseen una vasta experiencia en los temas afines a esta propuesta, tanto en lo que respecta a diseño digital en FPGA, como en los que respecta a Deep Learning y desarrollos/aplicaciones de CNNs

GRADO DE DEMANDA**NO CORRESPONDE****BIBLIOGRAFIA**

- [1] Kehtarnavaz, Nasser, and Mark Gamadia. "Real-time image and video processing: from research to reality." *Synthesis Lectures on Image, Video & Multimedia Processing 2.1*: 1-108, 2006.
- [2] Automated Imaging Association, <http://www.visiononline.org/>
- [3] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer, "Automatic differentiation in PyTorch," in *NIPS Workshop Autodiff*, 2017.
- [4] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, "TensorFlow: Large-scale machine learning on heterogeneous systems," 2015. Software available from tensorflow.org.
- [5] R. Al-Rfou, et. al, "Theano: A python framework for fast computation of mathematical expressions," *arXiv preprint arXiv:1605.02688*, 2016.
- [6] S. Dieleman, J. Schlüter, C. Raffel, E. Olson, S. K. Sønderby, D. Nouri, D. Maturana, M. Thoma, E. Battenberg, J. Kelly, J. D. Fauw, M. Heilman, D. M. de Almeida, B. McFee, H. Weideman, G. Takács, P. de Rivaz, J. Crall, G. Sanders, K. Rasul, C. Liu, G. French, and J. Degraeve, "Lasagne: First release.," Aug. 2015.
- [7] F. Chollet et al., "Keras." <https://keras.io>, 2015.
- [8] S. Tokui, K. Oono, S. Hido, and J. Clayton, "Chainer: a next-generation open source framework for deep learning," in *NIPS Workshop Machine Learning Systems*, 2015.
- [9] Rodríguez-Andina, Juan J., Maria D. Valdes-Pena, and Maria J. Moure. "Advanced features and industrial applications of FPGAs—A review." *IEEE Transactions on Industrial Informatics* 11.4 (2015): 853-864.



- [10] Yann LeCun, Yoshua Bengio, and Georey Hinton. Deep learning. *Nature*, 521(7553):436–444, 2015.
- [11] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, and others. Imagenet large scale visual recognition challenge. *International Journal of Computer Vision*, 115(3):211–252, 2015.
- [12] Ross Girshick. Fast R-CNN. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition- CVPR '15*, pages 1440–1448, 2015.
- [13] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully Convolutional Networks for Semantic Segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition - CVPR '15*, pages 3431–3440, 2015.
- [14] Ying Zhang, Mohammad Pezeshki, Philémon Brakel, Saizheng Zhang, Cesar Laurent Yoshua Bengio, and Aaron Courville. Towards end-to-end speech recognition with deep convolutional neural networks. *arXiv preprint, arXiv:1701*, 2017.
- [15] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint, arXiv:1409:1–14*, 2014.
- [16] Eriko Nurvitadhi, Suchit Subhaschandra, Guy Boudoukh, Ganesh Venkatesh, Jaewoong Sim, Debbie Marr, Randy Huang, Jason OngGeeHock, Yeong Tat Liew, Krishnan Srivatsan, and Duncan Moss. Can FPGAs Beat GPUs in Accelerating Next-Generation Deep Neural Networks? In *Proceedings of the ACM/SIGDA International Symposium on Field-Programmable Gate Arrays - FPGA '17*, pages 5–14, 2017.
- [17] Kalin Ovtcharov, Olatunji Ruwase, Joo-young Kim, Jeremy Fowers, Karin Strauss, and Eric Chung. Accelerating Deep Convolutional Neural Networks Using Specialized Hardware. *White paper*, pages 3–6, 2, 2015.
- [18] Jiantao Qiu, Jie Wang, Song Yao, Kaiyuan Guo, Boxun Li, Erjin Zhou, Jincheng Yu, Tianqi Tang, Ningyi Xu, Sen Song, Yu Wang, and Huazhong Yang. Going Deeper with Embedded FPGA Platform for Convolutional Neural Network. In *Proceedings of the ACM/SIGDA International Symposium on Field-Programmable Gate Arrays - FPGA '16*, pages 26–35, New York, NY, USA, 2016. ACM.
- [19] Intel FPGA. Intel® Stratix® 10 Variable Precision DSP Blocks User Guide. Technical report, Intel FPGA Group, 2017.

ⁱ Debe indicarse el área de conocimiento principal del proyecto según la clasificación OCDE-UNESCO versión 2010.

ⁱⁱ Debe indicarse la sub-área de conocimiento principal del proyecto según la clasificación OCDE-UNESCO versión 2010.