

Modelo de Redes Neuronales Convolucionales Profundas para la Clasificación de Lesiones en Ecografías Mamarias

Hugo Chanampe¹, Silvana Aciar², Martin de la Vega³, José Luis Molinari Sotomayor,
German Carrascosa, Alejandro Lorefice

Centro de Investigación y Desarrollo Informático (CIDI)¹³

Universidad Nacional de La Rioja¹³

Universidad Nacional de San Juan²

¹hugochanampe@unlar.edu.ar

²saciar@unsj-cuim.edu.ar

³martindelavega@unlar.edu.ar

Resumen

El cáncer de mama es una de las principales causas de muerte en mujeres mayores en Argentina. Nuestro estudio apunta a utilizar técnicas Deep Learning para clasificar distintos tipos de lesiones en ecografías mamarias. Se muestra el actual estado del arte de las arquitecturas de Deep Learning en la clasificación de imágenes y sus distintas estrategias de entrenamiento. Usando Redes Neuronales Convolucionales Profundas se desarrolló un modelo que permite clasificar diferentes tipos de lesiones. El modelo logro alcanzar una precisión del 62%, sobre la base datos mini-MIAS

1. Contexto

Este artículo se encuentra enmarcado dentro del proyecto de investigación “Deep Learning para el Reconocimiento de Patrones”, aprobado según resolución CICYT 056/17 perteneciente al Departamento Académico de Ciencias Exactas, Físicas y Naturales de la Universidad Nacional de La Rioja. Su ejecución se está llevando a cabo en el Centro de Investigación y Desarrollo Informático (CIDI) perteneciente a este mismo departamento.

2. Introducción

Hoy en día, las técnicas de Deep Learning (DL) o aprendizaje profundo vienen logrando exitosos resultados en el procesamiento de imágenes naturales, reconocimiento de objetos dentro de las mismas y clasificación de imágenes. Todo esto es posible gracias a las grandes capacidades de cómputo actuales y también

gracias a la convivencia de las potentes GPU de la actualidad.

Uno de los métodos de DL que se emplea para el análisis de imágenes son las *Redes Neuronales Convolucionales Profundas* (DCNN) [1], que puede descubrir las características en las imágenes, como detección de bordes y esquinas hasta llegar a detectar características más complejas, a través del entrenamiento con grandes conjuntos de datos. Esto también se está empleando en el análisis de imágenes médicas con prometedores resultados sobre varias de sus aplicaciones, segmentación de órganos, detección y clasificación de distintos tipos lesiones en los órganos, etc. [2]

El *Cáncer de Mama* (CDM) es una de las primeras causas de muerte en Argentina, en mujeres de entre 20 y 59 años de edad y la segunda para mujeres mayores de 59 años [3]. El diagnóstico y el tratamiento de esta enfermedad en etapas tempranas es vital para prevenir el progreso de la enfermedad y reducir los índices de mortalidad [4]. La revisión de mama se realiza a través del estudio de un tipo de imagen, como lo es *Ecografía Mamaria* (EM), que es el examen básico para el CDM.

La implementación de los métodos de DL, como las CNN puede ayudar a mejorar la clasificación de distintos tipos de lesiones dentro de las EM.

El propósito de este trabajo es desarrollar un modelo que use DL para clasificar los distintos tipos de lesiones en ecografías mamarias.

Este artículo está estructurado de la siguiente forma: En la sección 3 se presentan algunos conceptos de DL y una revisión de las diferentes arquitecturas de DL para el análisis de imágenes y diferentes estrategias de entrenamiento. En la sección 4 se presenta el modelo propuesto en este artículo. En la sección 5 se presentan los resultados obtenidos. En la sección 6 conclusiones y

trabajos futuros. Por último, en la sección 7 formación de recursos humanos y estructura del equipo de trabajo del proyecto de investigación.

3. Deep Learning o Aprendizaje Profundo

Los métodos de DL buscan un conjunto de transformación directamente desde los datos mismos. Esta mirada está logrando excelentes resultados, particularmente en los problemas que pertenecen al campo de *Visión de las Computadoras* (VC), como la clasificación de escenas naturales y la detección de objetos [5]. Los modelos de DL están siendo adaptados también a diferentes tareas médicas, como ayuda en el diagnóstico temprano del Alzheimer [6], segmentación automática en lesiones provocadas por la Esclerosis Múltiple [7], detección de tumores en ecografías mamarias (EM) [8] y muchos más. Sin embargo, únicamente unos pocos trabajos tienen explorados los métodos de DL en dirección de la clasificación automática de lesiones identificadas en EM [9].

Los métodos tradicionales de *Aprendizaje Automático* (ML) como las máquinas de soporte vectorial, el método del vecino más próximo y el análisis de discriminación lineal fueron usados como clasificadores para una gran variedad de desarrollos para el diagnóstico temprano del cáncer de mama [10] [11]. Sin embargo, tienen una capacidad limitada para procesar los datos en su forma original.

DL, sin olvidar que es un subcampo de ML, tiene métodos como lo son las arquitecturas de DCNN, que funcionan aprendiendo características relevantes directamente desde las grandes bases de datos (datos en bruto). Un aspecto clave de DL es que el aprendizaje de las características de los datos, no fue diseñado por un ingeniero experto. Estas fueron aprendidas desde los datos usando un procedimiento de aprendizaje general.

3.1. Antecedentes de las Redes Neuronales Convolucionales

Las DCNN se han convertido en una de las técnicas más populares dentro del campo de ML, que es muy usada en el reconocimiento de imágenes. La capacidad de las DCNNs puede ser controlada variando su profundidad y amplitud, también estas redes pueden hacer generalmente suposiciones correctas [12].

La clasificación de imágenes o clasificación de exámenes fue uno de los primeros campos donde DL hizo una mayor contribución en el análisis de imágenes médicas. En un examen de clasificación, por lo general se tiene una o varias imágenes como entrada y una sola variable de salida, el diagnóstico (ej.: presencia de enfermedad o ausencia). En [12] se representó un punto de inflexión para el reconocimiento de objetos a gran escala

cuando se pudo entrenar una de las primeras DCNN sobre un gran base de datos como ImageNet.

Las DCNN requieren una gran cantidad de datos de entrenamiento etiquetados, un aspecto crítico que puede ser un obstáculo difícil de sortear en el dominio de las imágenes médicas. Sin embargo, también el entrenamiento de las DCNN involucra muchos recursos de procesamiento y memoria, esto a veces podría consumir bastante tiempo. A veces el entrenamiento de este tipo de arquitecturas es complicado también por los problemas sobreajuste (overfitting) y bajo ajuste (underfitting) cuando hay datasets muy pequeños para el entrenamiento y cuando hay poco tiempo de entrenamiento [13]. Por eso el entrenamiento desde de este tipo de arquitecturas puede ser tedioso y consumir mucho tiempo, exigiendo una gran experticia y paciencia [9] [14].

Las DCNN es la técnica de DL más exitosa aplicada a la compresión de las imágenes [15]. Estas tuvieron su primera aparición exitosa en el mundo real en la aplicación de LeNet [16], para el reconocimiento de dígitos escritos a mano. A pesar de este éxito inicial no cobro impulso hasta que no se desarrollaron varias técnicas nuevas para el entrenamiento eficiente de las redes profundas. La clave fue el aporte realizado por [12] a fines del 2012 en la competencia de ImageNet Large-Scale Visual Recognition Challenge (ILSVRC). La DCNN que presentaron en esa oportunidad, llamada AlexNet, ganó la competición por un amplio margen. En el 2014 el Visual Geometry Group a través de su modelo propuesto (VGG-Net), cuyo principal aporte fue una evaluación a fondo de las redes de profundidad creciente usando una arquitectura muy pequeña de 3x3 filtros convulsionados (convolutional filters), que alcanzaron una significativa mejora, esto se logró configurando profundidades de entre 16 y 19 capas de pesos (weigh layers) [17]. Cabe destacar que este modelo obtuvo el segundo lugar en la competencia de ILSVRC. También en el ILSVRC 2014 También en el ILSVRC 2014 la arquitectura ganadora propuso una Deep Convolutional Neural Network, que fue llamada GoogLeNet, que tenía una red con 22 capas de profundidad, que contiene dos convolutional layers, tres pooling layers y nueve inception layers, pero reduciendo el número de parámetros a 4 millones. Cada inception layer consta de seis convolutional layers y una pooling layer. Pero su marca principal, además de mejorar el estado del arte en la clasificación y detección en la competencia, fue la mejora en la utilización de los recursos computacionales dentro de la red [18]. En el 2015 se desarrolló Inception-V2, la primera versión de Inception fue GoogLeNet anteriormente mencionada. Los cambios que se introdujeron en esta nueva versión fue la implementación del método llamado Batch Normalization (BN) que permite usar índices de aprendizajes mucho más altos, también actúa como un regularizador en las capas de entrada y hasta puede evitar uso de la técnica de Dropout

en la fase entrenamiento [19]. Usando el método BN se alcanzó la misma precisión que su antecesor, pero en menos pasos de entrenamiento de la red. Usando el modelo de Inception-V2 se logró un índice de error de validación 4.9 % sobre ImageNet.

Tabla 1 - Arquitecturas de Redes Neuronales Convolucionales Profundas

Algoritmos	Nº de Capas	Índice de Error	Posición en Competencia ILSVRC	Año
AlexNet	8	15.3 %	Primer Lugar	2012
VGGNet	19	7.3 %	Segundo Lugar	2014
GoogleNet	22	6.67 %	Primer Lugar	2014
ResNet	152	3.57 %	Primer Lugar	2015

Luego en la tercera versión, llamada Inception-V3, se mejoró la arquitectura agregando la idea de factorización que consiste en disminuir el número de parámetros en cada módulo inception, logrando reducir el costo computacional como se muestra en el trabajo [20]. Donde se presentó el modelo Inception-V3, que obtuvo mejores resultados sobre la misma base de datos con un índice de error sobre el set de validación de 3.58%. En el mismo año surge el problema de la degradación en las arquitecturas de redes neuronales profundas, que es cuando se aumenta la profundidad de una red neuronal (incremento en la cantidad de capas). Pero esto conlleva un perjuicio en la tasa de error del modelo, aumentando considerablemente como se reportó en [21] [22]. Aquí es donde aparecen las Deep Residual Nets que pueden fácilmente ganar precisión incrementando la profundidad en su arquitectura. La arquitectura que ganó en 2015 la competencia ILSVRC fue ResNet-34. Esta arquitectura es muy similar a la VGG-Net, excepto porque a cada filtro de 3x3 se le agrega una conexión de acceso directo, lo que la convierte a la red en una versión residual de la anterior. Esta red alcanzó una considerable mejora en la tasa de error en el entrenamiento, esto indica que el problema de la degradación está bien direccionado, permitiéndole ganar en precisión y profundidad en la red [23]. En la *Tabla 1*, se muestra la progresión de las arquitecturas que fueron más trascendentales en estos últimos años, por los resultados alcanzados en competencia ImageNet, destacándose la disminución de los índices de error.

Hoy por hoy, las DCNNs están siendo más ampliamente usadas en el análisis de imágenes médicas

3.2. Estrategias de Entrenamiento en las Redes Neuronales Convolucionales

3.2.1. Pre-Training y Fine-Tuning

Una prometedora alternativa al entrenamiento desde cero de las DCNN es la transferencia de aprendizaje, que es esencialmente el uso de redes pre-entrenadas desde campos de aplicación diferentes, por lo general se lleva a cabo sobre imágenes naturales. Los modelos pre-entrenados están siendo aplicados de manera exitosa en varias tareas dentro del campo de visión de las computadoras (VC), que puede ser como un generador de características o como una base para el aprendizaje mediante transferencia [24] [25]. Se tiene de esta manera dos estrategias aplicadas a la transferencia de aprendizaje que se pueden identificar: (1) uso de las redes pre-entrenadas (pre-training) solamente como extractor de características y (2) hacer pequeños ajustes (fine-tuning), sobre los pesos de una red pre-entrenada, para lograr una mejor performance sobre los datos a procesar. Pre-training significa inicializar las redes con parámetros pre-entrenados en lugar de establecer parámetros aleatoriamente. Es bastante popular en los modelos basados en DCNN, debido a que da la ventaja de permitir acelerar el proceso de aprendizaje y mejorar la capacidad de generalización. Como AlexNet, logro excelentes resultados [12], numerosos enfoques eligen a este modelo entrenado sobre ImageNet 2012 como su modelo profundo de referencia [26] [27] y utilizan fine-tuning para el ajuste de los parámetros, en concordancia con la tarea específica que tienen como objetivo.

3.2.2. Data Augmentation

El aumento de datos es el método más sencillo para reducir el sobreajuste cuando trabajamos con imágenes como datos. Consiste en utilizar las mismas imágenes del dataset que tenemos y conservando sus etiquetas se aplican diferentes transformaciones en las imágenes para generar datos adicionales sin introducir costos adicionales de etiquetado. Por ejemplo en el modelo AlexNet [12], tiene dos formas distintas de aumento de datos: la primera consiste en generar desplazamientos horizontales de las imágenes y reflejos horizontales, y la segunda forma consiste en alterar la intensidad de los canales de RGB en las imágenes de entrenamiento [13]. También se tomó AlexNet como base y se agregó transformaciones adicionales que mejora la cualidad invariable del desplazamiento y el color así extendiendo con un montón de imágenes con píxeles extras y agregando manipulaciones de color adicionales [28].

3.2.3. DropOut

La implementación de este procedimiento de entrenamiento consiste en desconectar un porcentaje de las neuronas en cada iteración del entrenamiento. El termino Dropout [29] se refiere a sacar de una red neuronal, unidades ocultas o visibles. También esta técnica es utilizada para evitar el sobreajuste, que es cuando tiene un conjunto de datos reducidos para entrenar nuestro modelo. Esto ocasiona muchas veces que se produzca un sobreajuste sobre este set de datos por parte del modelo.

A continuación, en la *Tabla 2*, se muestra las diferentes estrategias de entrenamiento de las DCNN y su principal característica.

Tabla 2 - Estrategias de Entrenamiento de las Redes Neuronales Convolucionales Profundas

Técnicas	Características
Pre-training	Uso de las de redes solamente como extractor de características
Fine-Tuning	Se hacen pequeños ajustes sobre los pesos de una red.
Data Augmentation	Esta estrategia de entrenamiento se utiliza para agrandar el conjunto de datos de entrenamiento de manera artificial sobre imágenes.
DropOut	Permite reducir el overfitting, desconectando unidades visible como ocultas de una red neuronal.

4. Modelo Propuesto

En esta sección se propone el desarrollo de un modelo que empleará como datos entrada las EM y aplicará una o varias arquitecturas de DCNN, vistas en las secciones anteriores. Permitiendo obtener una salida la cual será discriminar los distintos tipos de lesiones que se detectaron en las imágenes de entrada del modelo.

4.1. Base de Datos de Mamografías

La base de datos que se utilizará para el análisis de distintos tipos de lesiones será mini-MIAS [30]. Los tipos de anomalías que tiene este conjunto de dato son: 1) calcificación (CALC), 2) masas bien definidas (CIRC), 3) masas espiculadas (SPIC), 4) otras masas mal definidas (MISC), 5) Distorsión Arquitectónica (ARCH), 6) Asimetría (ASYM) y 7) Normal (NORM).

La base de datos contiene 322 imágenes que tienen una resolución de 1024x1024. Este base de datos es bastante pequeña en el contexto de DL y VC, pero es ampliamente usada en la literatura [31].

4.2. Procesamiento de las Imágenes

Se procederá a dividir el cien por ciento de las imágenes en tres: a) set de datos de entrenamiento (80%), b) set de datos de validación (10%) y, por último, el set de datos de prueba con el 10% restante. Para cada grupo se escogerán los datos cuidadosamente, dejando un ejemplo de cada uno de los tipos de lesiones que contenga el dataset.

El set de datos de prueba será usado solo para evaluar el modelo propuesto, esto implica que estos datos nunca serán utilizados para entrenar el modelo.

4.2.1. Aumento del Set de Datos de Entrenamiento

Al set de entrenamiento y validación se le aplicaran distintas técnicas de *data augmentation* como desplazamientos, rotaciones y giros aleatorios. Todas estas transformaciones se harán a pequeña escala, para no alterar la naturaleza de las imágenes. También se hará un análisis para observar la distribución de la muestra y normalizar los datos para que cada tipo de lesión tenga el mismo número de imágenes.

Por el ultimo se deberá cambiar el tamaño de las imágenes a 224x224 que es tamaño por defecto que usan la mayoría de las arquitecturas más populares vistas.

Esto permitirá aumentar el tamaño del set de datos de entrenamiento y validación.

4.3. Modelo

Se entrenará una arquitectura de DCNN para la clasificación lesiones en las ecografías mamarias y se analizara la precisión del modelo con distintas métricas de evaluación [32]. Esto se describe a continuación.

4.3.1. Arquitectura de DCNN y Entrenamiento

Se seleccionará y evaluará una arquitectura DCNN: en este caso de estudio se eligió *VGGNet-16*. Se seleccionó este modelo debido a que es una arquitectura poco profunda, en comparación con las más recientes, y no requiere gran capacidad de computo, ya que la disponibilidad de recursos es limitada para este proyecto.

Para llevar a cabo el entrenamiento se realizarán algunas modificaciones en la arquitectura original de VGGNet-16, pre entrenada sobre base de datos de Imagenet. Entre esas modificaciones están quitar el clasificador original. Luego se insertará un nuevo clasificador, acorde a la base de datos seleccionada. Por último, se procederá a entrenar el modelo.

5. Resultados Obtenidos

En la *tabla 3* que se muestra se muestran los resultados obtenidos. La precisión alcanzada por el modelo propuesto fue de un 62%.

Tabla 3 - Resultados

Tipo de Lesiones	Precision	Recall	f1-score	support
ARCH	0.00	0.00	0.00	5
ASYM	0.00	0.00	0.00	3
CALC	0.32	0.75	0.44	8
CIRC	0.20	0.33	0.25	6
MISC	0.14	0.25	0.18	4
NORM	0.86	0.48	0.62	52
SPIC	0.50	0.40	0.44	5
promedio / total	0.62	0.43	0.48	0.83

6. Conclusión y Trabajos Futuros

En nuestro estudio, se presentó un modelo de DL que implementa una arquitectura de DCNN para clasificar distintos tipos de lesiones directamente desde las ecografías mamarias. Demostró que es difícil alcanzar una alta precisión cuando no se posee una base de datos con gran número de ejemplos para cada tipo de lesión. Pero si se pudo observar cómo influyen las diferentes estrategias de entrenamiento para contrarrestar esta situación.

Futuros trabajos, podrían incluir explorar otras arquitecturas y probar este modelo con datasets que sean de mayor tamaño o utilizar múltiples fuentes de datos.

7. Formación de Recursos Humanos

El proyecto de investigación se encuentra estructurado de la siguiente manera:

- Director Consultor: Dra. Silvana Aciar
- Director Ejecutivo: Ing. Hugo Chanampe, se encuentra realizando su trabajo final de Maestría.
- Co-Director: Ing. Martin de la Vega, se encuentra realizando su trabajo final de Maestría.
- Integrantes:

- Molinari Sotomayor, José Luis
- Carrascosa, German
- Lorefice, Alejandro Martin

Todos los integrantes son alumnos avanzados de las carreras de sistemas de la Universidad Nacional de La Rioja.

8. Referencias

- [1] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [2] W. Zhang, R. Li, H. Deng, L. Wang, W. Lin, S. Ji, and D. Shen, "Deep convolutional neural networks for multi-modality iso-intense infant brain image segmentation," *Neuroimage*, vol. 108, pp. 214–224, 2015.
- [3] "Instituto Nacional del Cancer -Mortalidad," 2015. [Online]. Available: Mortalidad. [Accessed: 13-Apr-2018].
- [4] R. a. Smith, V. Cokkinides, and H. J. Eyre, "American Cancer Society Guidelines for the Early Detection of Cancer, 2004," *CA. Cancer J. Clin.*, vol. 54, no. 1, pp. 41–52, 2004.
- [5] Y. Bengio, A. Courville, P. V.-I. transactions on pattern, and undefined 2013, "Representation learning: A review and new perspectives," *ieeexplore.ieee.org*.
- [6] S. Liu, S. Liu, W. Cai, S. Pujol, R. Kikinis, and D. Feng, "Early diagnosis of Alzheimer's disease with deep learning," in *2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI)*, 2014, pp. 1015–1018.
- [7] A. Birenbaum and H. Greenspan, "Longitudinal Multiple Sclerosis Lesion Segmentation Using Multi-view Convolutional Neural Networks," Springer, Cham, 2016, pp. 58–67.
- [8] A. Akselrod-Ballin, L. Karlinsky, S. Alpert, S. Hasoul, R. Ben-Ari, and E. Barkan, "A Region Based Convolutional Network for Tumor Detection and Classification in Breast Mammography," Springer, Cham, 2016, pp. 197–205.
- [9] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. W. M. van der Laak, B. van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Med. Image Anal.*, vol. 42, no. December 2012, pp. 60–88, 2017.
- [10] H. Cai, Y. Peng, C. Ou, M. Chen, and L. Li, "Diagnosis of Breast Masses from Dynamic Contrast-Enhanced and Diffusion-Weighted MR: A Machine Learning Approach," *PLoS One*, vol. 9, no. 1, p. e87387, Jan. 2014.
- [11] M. Krishnan, S. Banerjee, ... C. C.-E. S. with, and undefined 2010, "Statistical analysis of mammographic features and its classification using support vector machine," *Elsevier*.
- [12] A. Krizhevsky, I. Sutskever, ... G. H. information processing, and undefined 2012, "Imagenet classification with deep convolutional neural networks," *papers.nips.cc*.
- [13] Y. Guo, Y. Liu, A. Oerlemans, S. Lao, S. Wu, and M. S. Lew, "Deep learning for visual understanding: A

- review,” *Neurocomputing*, vol. 187, pp. 27–48, 2016.
- [14] R. Girshick, J. Donahue, T. Darrell, and J. Malik, “Region-Based Convolutional Networks for Accurate Object Detection and Segmentation,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 1, pp. 142–158, 2016.
- [15] J. S.-N. networks and undefined 2015, “Deep learning in neural networks: An overview,” *Elsevier*.
- [16] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2323, 1998.
- [17] K. Simonyan and A. Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition,” *Int. Conf. Learn. Represent.*, pp. 1–14, 2015.
- [18] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, “Going deeper with convolutions,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2015, vol. 07–12–June, pp. 1–9.
- [19] S. Ioffe and C. Szegedy, “Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift,” Feb. 2015.
- [20] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, “Rethinking the Inception Architecture for Computer Vision.” pp. 2818–2826, 2016.
- [21] K. He, J. S.-P. of the I. conference on computer, and undefined 2015, “Convolutional neural networks at constrained time cost,” *cv-foundation.org*.
- [22] R. K. Srivastava, K. Greff, and J. Schmidhuber, “Highway Networks,” May 2015.
- [23] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.
- [24] A. S. Razavian, H. Azizpour, J. Sullivan, and S. Carlsson, “CNN features off-the-shelf: An astounding baseline for recognition,” in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 2014, pp. 512–519.
- [25] O. A. B. Penatti, K. Nogueira, and J. A. Dos Santos, “Do deep features generalize from everyday objects to remote sensing and aerial scenes domains?,” in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 2015, vol. 2015–Octob, pp. 44–51.
- [26] K. He, X. Zhang, S. Ren, and J. Sun, “Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 9, pp. 1904–1916, 2015.
- [27] S. Ren, K. He, R. Girshick, J. Sun, J. Donahue, T. Darrell, J. Malik, W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, R. Girshick, D. Impiombato, S. Giarrusso, T. Mineo, O. Catalano, C. Gargano, G. La Rosa, F. Russo, G. Sottile, S. Billotta, G. Bonanno, S. Garozzo, A. Grillo, D. Marano, G. Romeo, D. Erhan, C. Szegedy, A. Toshev, and D. Anguelov, “Rich feature hierarchies for accurate object detection and semantic segmentation,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 794, pp. 1–15, 2015.
- [28] A. G. Howard, “Some Improvements on Deep Convolutional Neural Network Based Image Classification,” *arXiv Prepr. arXiv1312.5402*, pp. 1–6, 2013.
- [29] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, “Dropout: A Simple Way to Prevent Neural Networks from Overfitting,” *J. Mach. Learn. Res.*, vol. 15, pp. 1929–1958, 2014.
- [30] “The mini-MIAS database of mammograms.” [Online]. Available: <http://peipa.essex.ac.uk/info/mias.html>. [Accessed: 17-Aug-2018].
- [31] N. Dhungel, G. Carneiro, and A. P. Bradley, “Automated Mass Detection in Mammograms Using Cascaded Deep Learning and Random Forests,” *2015 Int. Conf. Digit. Image Comput. Tech. Appl.*, pp. 1–8, 2015.
- [32] M. Sokolova and G. Lapalme, “A systematic analysis of performance measures for classification tasks,” *Inf. Process. Manag.*, vol. 45, no. 4, pp. 427–437, Jul. 2009.