Clasificación de cálculos renales con técnicas de Deep Learning

Víctor Asensio-Casas

Resumen — En este proyecto se propone una herramienta de clasificación automática de imágenes basada en algoritmos de visión por computador. El objetivo es obtener una clasificación que permita identificar el tipo de un cálculo renal. Para realizar dicho clasificador se han estudiado diferentes técnicas relacionadas con los últimos avances en análisis de cálculos renales, y analizado el potencial de las *Convolutional Neural Network* (CNN). Posteriormente se explica el desarrollo del clasificador que consta de tres partes principales: aplicación de la técnica *fine*-tune a una red neuronal, preprocesamiento de las imágenes y pruebas de clasificación. Los resultados obtenidos a lo largo del desarrollo de este proyecto son positivos para el futuro, aunque no se hayan obtenido unos resultados excesivamente buenos, hemos podido comprobar que, una *Convolutional Neural Network,* con multitud de imágenes puede llegar a mejorar los resultados obtenidos con los clasificadores que se usaban previamente.

Palabras clave— Piedras de riñón; myStone; Dispositivo médico; *Convolutional Neural Network*; *Deep Learning*; *AlexNet*; *Fine-tune*; clasificación

Abstract—This project proposes an automated image classification tool based in computer vision algorithms. The final objective is to classify kidney stones. To develop this tool we have first carried out a research on the State-of-the-art of renal calculi analysis methods and the efficiency of Convolutional Neural Networks (CNN). After that, we propose our classification methodology which consists in three main stages: applying the fine-tune technique, image preprocessing, and testing the classification. The results of this project are very hopeful even though the results have not been too high because the Convolutional Neural Network results would increase as we increase the number of images.

IndexTerms— Kidney Stones; myStone; Medical Device; Convolutional Neural Network; Deep Learning; AlexNet; Fine-tune; Classification

1 Introducción

LOS cálculos renales afectan a una pequeña parte de la sociedad en algún momento de su vida. No obstante, es un problema altamente recurrente. La única forma de mitigar el riesgo de volver a padecer un episodio litiásico es identificar el tipo de cálculo y seguir el tratamiento estipulado para dicha clase.

Actualmente el análisis de los cálculos se hace en el laboratorio, aplicando una combinación de diversas técnicas químicas.

MyStone es un proyecto en el que se pretende, mediante un dispositivo hardware que, incluye una cámara y diversos leds, simular estas técnicas con procedimientos de visión artificial.

Apoyándose en las técnicas que ofrece el *Deep Learning*, más concretamente en las *Convolutional Neural Network* (CNN), este trabajo propone un clasificador que simule los procedimientos típicos de un experto en el análisis de un cálculo.

Previamente, en el campo de la visión, se ha hecho uso de *Suport Vector Machines* (SVM) para clasificar los cálculos [11]. Los resultados obtenidos no son satisfactorios, por lo que se requiere el análisis del problema con otro enfoque, en este caso desde el punto de vista del *Deep Learning*.

Los distintos tipos de imágenes que permite capturar el dispositivo de *myStone* se muestran en la Figura 1.

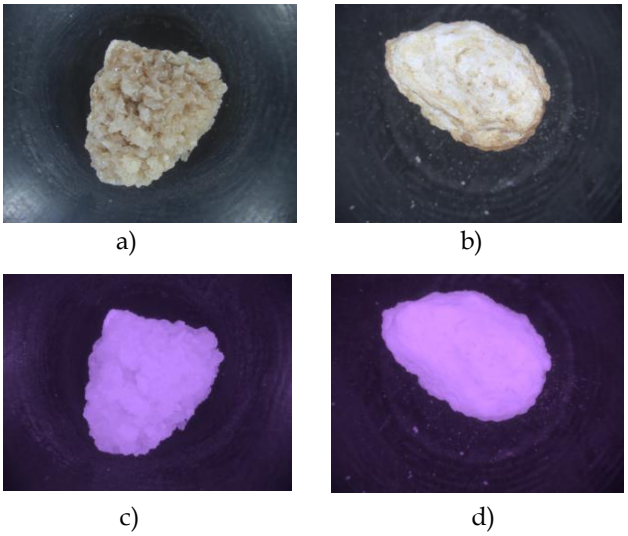


Figura 1. Imágenes tomadas con el dispositivo myStone. a) luz blanca, vista exterior, b) luz blanca, vista interior, c) IR, vista exterior, d) IR, vista interior.

- *E-mail de contacto: Victor.asensio@e-campus.uab.cat*
- *Mención realizada: Computación*
- *Trabajo tutorizado por: Felipe Lumbreras Ruiz (Ciencias de la Computación y Centro de Visión por Computador)*
- *Curso 2015/16*

El artículo está estructurado de la siguiente manera. En la Sección 2 se presentan las motivaciones personales por las que se ha elegido este tema. La Sección 3 muestra los objetivos que se quieren alcanzar al finalizar el proyecto. La distribución de horas respecto a los objetivos está en la Sección 4. Posteriormente en la Sección 5 se trata el estado del arte donde se profundiza el estudio de los métodos que se utilizan en otros sectores, los que se han aplicado en el proyecto *myStone* previamente y finalmente un estudio sobre las *CNN*. En la sección 6, se explica la metodología llevada a cabo para hacer el proyecto. La Sección 7 incluye los resultados. Finalmente, la Sección 8 contiene las conclusiones y líneas futuras.

2 Motivaciones

El 5% de la población mundial producirá cálculos renales a lo largo de su vida. Este es un problema de salud [15] que afecta a nivel mundial y puede llegar a provocar que se tengan que seguir tratamientos dolorosos, cirugías para la extracción de los cálculos, e incluso se pueden llegar a producir fallos renales provocando la muerte del paciente (3% de los casos). A pesar de no poder evitar la formación de cálculos renales, una vez expulsados, tenemos la oportunidad de identificar qué tipo de cálculo ha generado el paciente y por lo tanto reducir la posibilidad de que vuelva a generar más piedras [12] de ese tipo, sugiriéndole un tratamiento específico.

Los métodos que se utilizan actualmente para identificar los cálculos son todos excesivamente caros y lentos. Por otro lado el proyecto *myStone* pretende simular el proceso llamado análisis morfoconstitucional, explicado en la Sección 5, donde a partir de imágenes de los cálculos es capaz de identificar el tipo.

3 Objetivos

El principal objetivo de este proyecto es estudiar cómo aplicar una *CNN* en nuestro problema. La base de datos para realizar los experimentos está conformada por un conjunto de datos con imágenes en color e infrarrojo.

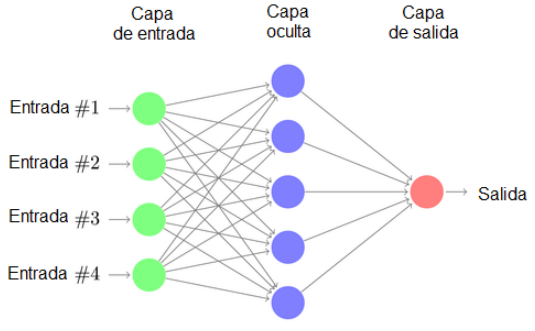


Figura 2. Estructura de una CNN

Las *CNN* se componen de tres tipos de capas distintas, donde las intermedias u ocultas son las más abundantes y las que permiten realizar el aprendizaje de la red. En el caso de las *CNN* la capa de entrada corresponde con la imagen de entrada, mientras que la capa de salida tiene un número de nodos igual a las clases que se quieren clasificar.

Debido a la extensión del proyecto hemos decidido separar en bloques los objetivos para poder alcanzar el objetivo final de forma más eficiente.

- Estudiar estado del arte.
  - Avances en el campo de las *CNN*.
  - Aplicación de la técnica *fine-tune* que se detalla en la sección 5.
- Generar las bases de datos.
  - Adaptar imágenes a la red.
  - Generar imágenes virtuales.
  - Generar base de datos de píxeles.
- Análisis del comportamiento de clasificación con una *Super Vector Machine* (SVM). Estos resultados son tomados como referencia para la evaluación final.
- Análisis de pruebas aplicadas a la base de datos de forma **global**.
  - Aplicación directa de la técnica *fine-tune*.
- Evaluar los resultados y compararlos con los resultados referencia del mejor clasificador (*SVM* o *Random-Forest*).
- Estudio y aplicación de la herramienta MatConvNet [2].
  - Estudiar la aplicación de *fine-tune* en la red AlexNet.

El proyecto parte con una base de datos de 1500 imágenes de cálculos de riñón aproximadamente, aun sabiendo que no son suficientes para poder hacer *Deep Learning* eficazmente se pretende analizar el comportamiento de las redes.

4 Planificación

Se pretende completar siguiendo los *timings* establecidos.

Tabla 1: Distribución de horas.

Bloques	horas
Estado del arte	40
Generación de base de datos	40
Pruebas globales	160
Evaluación	60
TOTAL	**300**

La mayor parte del tiempo se ha dedicado a la realización de pruebas, puesto que el éxito del trabajo consiste en completar los experimentos suficientes que permitan justificar una posible mejora en los métodos actuales de clasificación de cálculos renales.

La evaluación permite analizar si hemos cumplido con los objetivos establecidos.

La generación de la base de datos ha consistido en adaptar la que se utilizó en [13].

5 ESTADO DEL ARTE

El problema de identificar cálculos de riñón es actualmente muy frecuente, aun así no existe ninguna técnica en el campo de visión por computador que lo solucione. Las únicas formas de clasificar un cálculo son mediante el uso de técnicas químicas como el uso de un kit de análisis químico, difracción de rayos X, microscopio electrónico (SEM), espectrometría de infrarrojos u otro tipo de técnicas[5].La aplicación de la técnica SEM es la que ofrece resultados más exactos y precisos. Permite identificar ciertos patrones a vista microscópica, pero se requiere a un experto para interpretar dichos resultados.

Actualmente muy pocas personas son capaces de identificar cálculos a simple vista, pero los resultados muchas veces están sujetos a la subjetividad del observador (cuando la decisión es entre dos clases similares) y pueden diferir entre distintos expertos analizando la misma muestra.

Vamos a explicar las técnicas que se utilizan en el sector químico, óptico, las que se han estado utilizando hasta ahora en el proyecto *myStone* y por último las que queremos aplicar en este proyecto.

5.1 Métodos químicos

Kit de análisis químico

Esta técnica es la más usada en laboratorios para el análisis de piedras porque es las más rápida y fácil de aplicar, aun así solo detecta la presencia de iones y radicales individuales, los cuales no permiten diferenciar de forma específica la composición en algunos tipos de piedras o mezclas. Los resultados no son muy precisos.

Análisis de termogravimetría

Técnica viable, rápida y fácil de aplicar. Se basa en llevar el cálculo progresivamente a 1000 ºC en una atmósfera de oxígeno y observar cuando empieza a perder peso. Esto se usa como una característica y nos permite diferenciar entre distintos tipos. Además nos indica la proporción de los diferentes componentes químicos que la componen.

Difracción de rayos X

Técnica que usa rayos X monocromáticos con tal de identificar cálculos renales basándose en su constitución. Lanza un patrón de difracciones únicas producidas por el material cristalino y estas difracciones atraviesan pequeñas fracciones de cristales del cálculo o son reflejadas en forma de patrones particulares [17].

Espectrometría infrarroja

Es la que mejores resultados proporciona si queremos saber el tipo de cálculo, ya que permite la identificación molecular de ciertos materiales debido al efecto del infrarrojo. Este método es más preciso y sofisticado que utilizar el kit de análisis químico, más rápido que la difracción de rayos X y más barato que las técnicas ópticas que se comentan más adelante, pero siguen existiendo ambigüedades entre algunas clases que, sin la ayuda de los métodos ópticos no podríamos resolver. Además estos métodos tienen un inconveniente, en cálculos blandos deshacen la estructura de la muestra y se pierde el morfo. [14]

5.2 Métodos ópticos

Microscopio electrónico de barrido (SEM)

Permite generar imágenes de alta calidad, donde se pueden apreciar los conglomerados minúsculos, los cuales están contenidos en las piedras, imposibles de ver a simple vista. Éstos dependiendo del patrón que sigan nos permiten identificar de forma precisa el tipo de piedra que estamos tratando. Este método resuelve las ambigüedades descritas en otros métodos ofreciendo resultados más exactos y precisos.

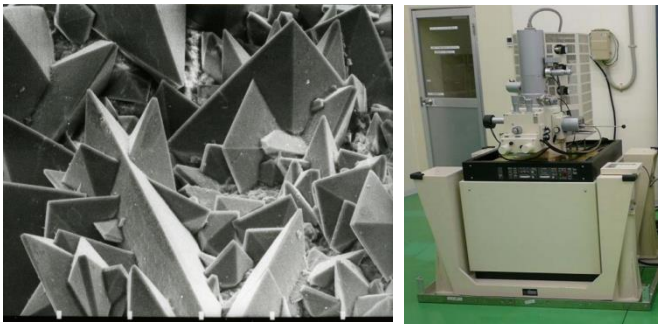


Figura 3. A la izquierda se muestra una imagen tomada con el SEM, donde se muestra la estructura de un COD. A la derecha podemos observar el Hitachi S-4100 T SEM.

Análisis morfoconstitucional

Técnica que consiste en describir un cálculo usando un microscopio estereoscópico [4, 8]. Es necesario que un experto realice el análisis. La piedra debe ser cortada por la mitad con el fin de permitir al experto analizar las partes internas y externas de cada fragmento. La visualización de cristales, lobulaciones y otras características de forma, dureza o color en el cálculo, debe ser anotada debido a que puede llegar a aportar información sobre la clase del cálculo. Este método es el que se intenta emular en el proyecto *myStone*.

5.3 Caracterización de los cálculos

Existen varias etiquetas para clasificar las vistas de las piedras y a continuación vamos a precisar en el concepto de los cálculos y la composición de cada uno, apoyándonos en la clasificación realizada en [10]:

1. **COM**. Oxalato de calcio monohidrato
2. **COD**. Oxalato de calcio dihidrato
3. **HAP**. Hidroxiapatita
4. **STR**. Estruvita (Fosfato de amonio de magnesio)
5. **BRU**. Brushita (Fosfato de calcio hidrogenado
6. **CYS**. Cistina
7. **AU**. Ácido úrico
8. **AUD.** Ácido úrico dihidrato
9. **AUA.** Ácido úrico anhidro

Existen cuatro tipos de cálculos que están constituidos por la combinación de diversas substancias químicas:

10. **TRA**. Oxalato de calcio dihidrato transformada en

Oxalato de calcio monohidrato.

11. **MXD**. Mixta de oxalato de calcio dihidrato con hidroxiapatita.
12. **AU-CO**. Ácido úrico con núcleo de oxalato de calcio.
13. **COD-HAP**. Oxalato de calcio dihidrato e hidroxiapatita.
14. **COM-HAP.** Oxalato de calcio monohidrato e hidroxiapatita.

Estas etiquetas [13] nos permiten definir los tipos de vista de los cálculos de riñón.

Además de los tipos de vista de los cálculos hay otro tipo de clasificación más profunda y necesaria para poder identificar el cálculo de forma correcta: por clases de tubo.

Esta clasificación es más artificial y viene dada por expertos del sector:

- 2: COM
- 2b:COM con core de HAP
- 2cod: COM con depósitos superficiales de COD
- 3: COD
- 3t: Transformada (TRA) de COD a COM
- 3b: Transformada de COD a COM con HAP minoritario.
- 4: Mixto COD y HAP.
- 5: HAP
- 5b: HAP y COD minoritario.
- 6: Estruvita
- 7: Brushita
- 8: AUA
- 8b: AUD
- 9: AU y CO
- 10: CYS

Esto nos define el tipo de tubo a partir de 4 imágenes correspondientes a la parte exterior e interior de la piedra. Además se debe destacar que algunas clases de vista contienen pocas muestras y se combinaron con otras similares para aumentar el rendimiento del clasificador. La Tabla 2 muestra los cambios realizados.

Tabla 2: Cambios tipos de vista.

Clase vista original	Clase vista destino
AUA/AUD	AU
COM-HAP	COM

5.4 Trabajo previo

Actualmente el proyecto *myStone* está avanzado y lo que se propone es, mediante un hardware que consiste en una cámara, ocho leds blancos (cuatro a 45 º y los otros cuatro a 90º) y ocho leds IR, intentar clasificar las piedras a partir de la toma de 8 imágenes por vista.

En la Figura 4 podemos visualizar la estructura del hardware que usamos para tomar las imágenes.

La placa de iluminación de la Figura 5 es la que permite generar una iluminación artificial adecuada para el problema y la visión óptima de la piedra.

Este método permite combinar varias configuraciones que, incluyen la activación de distintos leds con el fin de iluminar la muestra desde varios ángulos y con distintas iluminaciones.

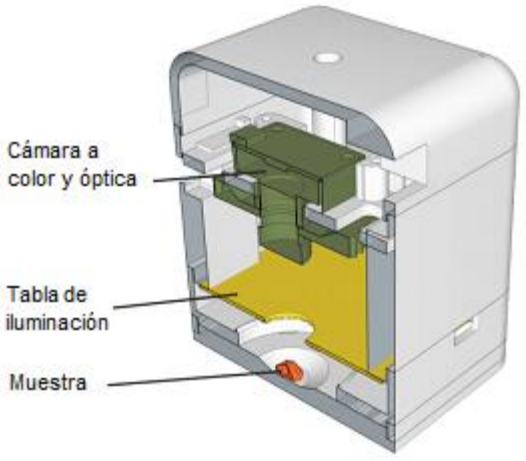


Figura 4. Sección donde se muestra la estructura del hardware myStone que se encarga de tomar las imágenes. El modelo de la cámara utilizado es el Basler dart 14uc a color de 4 MPix.

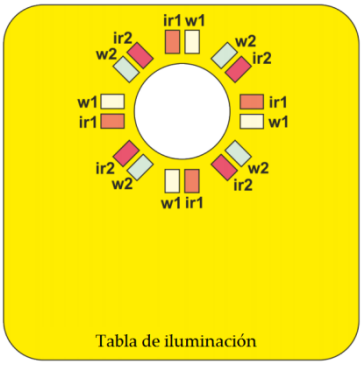


Figura 5. Configuración de los LED blancos e infrarrojos. IR1 trabaja a 860 nm e IR2 a 940 nm.

La clasificación de los cálculos se realiza de la siguiente forma:

1. Colocamos la muestra en la base.
2. Cerramos el dispositivo colocándole la cámara encima.
3. Adquirimos las imágenes. (32 imágenes, 8 por cada vista)
4. Recortamos la zona de la piedra. (Segmentar)
5. Extraemos características de las imágenes aplicando *local binary pattern*(LBP) e histogramas a color (YCbCr).
6. Clasificamos la piedra mediante la técnica de *Super Vector Machine* y se obtiene un informe con el tratamiento que el paciente ha de seguir.

Todo este proceso nos da los siguientes resultados a nivel de vista con la técnica SVM:

Tabla 3: Precisión a nivel de vistas.

Clases (etiquetas)	Muestras	Precisión
COM	239	75%
COD	126	50%
COD-HAP	125	38%
TRA	203	32%
MXD	153	40%
HAP	64	6%
STR	123	59%
BRU	55	4%
AU	239	82%
AU-CO	41	17%
CYS	24	0%
TOTAL	**1392**	**50 %**

Como podemos observar en la Tabla 3 en general son resultados bastante bajos ya que los especialistas necesitan porcentajes más altos de acierto (>90%) para poder usar este dispositivo. Generalmente podemos observar que tenemos tipos de vistas con muy pocas imágenes que son las que menos aciertos nos proporcionan, como es el caso de la CYS que solamente tenemos 24 imágenes para procesar. Por otro lado tenemos muestras con más imágenes, como COM y AU dónde los porcentajes mejoran substancialmente.

5.5 Convolutional neural network

Son herramientas utilizadas con mucha frecuencia dentro del ámbito del *Machine Learning* y cada vez va más en aumento su uso, llegando a mejorar prácticamente en todos los casos los métodos clásicos. El objetivo de éstas se relaciona con el reconocimiento de imágenes.

Las capas de la red neuronal pueden tener funciones de activación distintas, esto nos permite ajustar nuestro problema adecuadamente haciendo diversas combinaciones de estas capas con tal de llegar a soluciones óptimas.

Además permiten aplicar la técnica de*fine-tune*, esto significa que podemos coger una red previamente entrenada, especializada en algún dominio concreto y hacer que ésta sea capaz de clasificar dentro de cualquier otro dominio, enseñándole imágenes del nuevo dominio a aprender. Se ha decidido enfocar el proyecto hacía la aplicación del *fine-tune* debido a que, tenemos muy pocas muestras y esto haría que el aprendizaje no fuero robusto.

5.6 Toolboxes

Las herramientas más utilizadas actualmente para usar *CNN* son:

MatConvNet

Herramienta eficiente y simple. Implementada en Matlab permite el uso de redes entrenadas previamente y la generación desde cero de las mismas. Esta herramienta ha sido utilizada para desarrollar este proyecto.

Theano

Librería Python que permite eficacia a la hora de evaluar expresiones matemáticas que involucran matrices multidimensionales. Es una librería compleja y poco adecuada para una primera toma de contacto con estos problemas.

Caffe

Framework de *Deep Learning* que permite alta velocidad de ejecución y modularidad. Herramienta que se utiliza en los experimentos donde se requiere de alta velocidad y un control muy fino, pero no es adecuada para una primera aproximación al problema.

6 METODOLOGÍA

En este apartado se explican los distintos procesos que se le han aplicado a las imágenes para realizar los experimentos. Como hemos generado una base de datos compatible con la *CNN* que vamos a utilizar, se analiza el proceso de calibrar las distintas cámaras y la importancia de utilizar la potencia de la tarjeta gráfica en vez del procesador para trabajar de manera más eficiente.

6.1 Métricas

Para valorar los resultados obtenidos y poder compararlos con los valores de referencia se utiliza el valor *precision*. Esta métrica marca el porcentaje de acierto contemplando solamente los *true positive*.

$$Precision = \frac{\#\ true\ positive}{\#\ samples} * 100$$

Fórmula 1: Cálculo de precisión.

Los *true positive* son los aciertos, es decir cuando el clasificador dice que es una clase y coincide con el *groundtruth*. Los *samples* son el número total de muestras que se utilizan para realizar el *test*.

6.2 Generación de bases de datos

La base de datos de referencia ha sido la de myStone (VALTEC) [13].

Nuestro trabajo en este apartado fue dedicado a preparar el conjunto de imágenes de la base de datos para poder introducirlas en *AlexNet*, la *CNN* que elegimos para la resolución de este problema.

Los datos por tubo estaban organizados exactamente en 32 imágenes, donde cada 8 imágenes pertenecen a una vista y el tamaño de cada imagen era de 2592x1944x3. Además de cada vista tenemos su respectiva máscara.

Todos los tubos con los que trabajamos tenían esta configuración, y eran 348 en total.

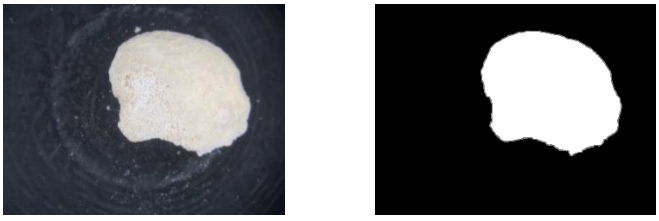


Figura 6: Imagen izquierda muestra el cálculo bajo luz visible y la derceha muestra la máscara para segmentar.

Datos en la red neuronal

Las redes neuronales en MatConvNet trabajan con una estructura de datos bien definida llamada `imdb`. Nuestro trabajo se basa en utilizar una red neuronal ya entrenada, configurada con 21 capas, en el caso de AlexNet [1] que solo permite la entrada de imágenes con una dimensión de 227x227x3.La estructura `imdb` contiene los datos (entrenamiento y validación), la media de los datos (en nuestro caso al usar una red preentrenada consiste en utilizar la imagen normalizada de la red), las etiquetas de *groundtruth* de las vistas, un conjunto de datos que define qué imágenes serán para validar y cuales para entrenar y por último incluye una parte de meta datos que define las clases y sus nombres.

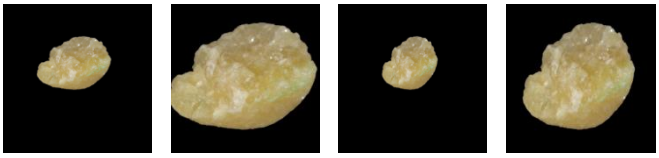
Centrándonos en los datos, que es la parte más importante en esta sección, los preparamos de tal forma que una matriz incluya todas las imágenes con un tamaño de 227x227x3x348, donde el último valor depende directamente del número de muestras. La preparación de estos datos, en primera instancia se pensó almacenarlos en disco, de forma que, se tuviera acceso a los datos generados posteriormente. Una vez realizado este proceso, la clasificación consistiría solamente en cargar ficheros a memoria. Sin embargo esto dio resultados pésimos en cuanto a tiempo de ejecución, por lo que se optó por manipular estos datos de forma dinámica y mejoró notablemente el tiempo de ejecución.

Finalmente, lo único que se almacena en disco es la base de datos de imágenes que se genera previamente a `imdb`.

6.3 Pre-procesado de las imágenes

Las imágenes del dataset seleccionado han de adaptarse a la CNN, lo cual obliga a redimensionarlas. Redimensionar una imagen implica perder mucha información, por lo tanto decidimos realizar varios experimentos tratando la imagen de forma distinta para estudiar la forma más eficaz de tratar los datos. Además observamos que las imágenes, al ser redimensionadas a la dimensión adecuada, no mantenían el *aspect ratio* debido a que la imagen origen no tenía una dimensión cuadrada y por lo tanto añadía más confusión a la red. Con tal de solucionarlo tuvimos que añadir un preprocesamiento que mantuviese el *aspect ratio* y por lo tanto generara imágenes cuadradas, eliminando de forma horizontal parte de la imagen que no nos interesará.

El primer experimento que se llevó a cabo fue introducir la imagen aplicando únicamente la máscara. Posteriormente realizamos el segundo experimento, en el que lo único que hacíamos era quedarnos con la imagen escalada, es decir quedarnos la mitad de la imagen o ¼ de la misma. Este experimento no ofreció buenos resultados, debido al tamaño de algunas piedras o que no estaban siempre en el centro de la imagen y por lo tanto la red aprendía imágenes completamente en negro, lo cual no aportaba información. Para solucionarlo hicimos un experimento en el que buscábamos el centro de masas de la imagen (la posición central de la piedra) para poder quedarnos con una parte de la imagen sin perder las piedras de vista.



a) b) c) d)

Figura 7: a) y b) mantienen el aspect ratio. Se observa la deformación de la piedra en las imágenes c) y d).

Con el algoritmo recortamos la imagen y nos quedamos con la imagen escalada. A partir de factores de escala de ½, un ¼ y así consecutivamente hasta quedarnos con la parte de la imagen que se ajustara al tamaño permitido por la CNN, es decir 227x227.El mejor clasificador resultó con un factor de escala de ½.

Todos los experimentos posteriores fueron realizados con estos ajustes, es decir manteniendo el *aspect ratio*, buscando el centro de masas y por último quedándonos con la mitad de la imagen.

Tabla 4: Precisión variando el tamaño.

Tamaño imágenes	Media (%)
1944x1944	37.03
972x972	45.99
486x486	27.86
243x243	25.112
227x227	16.94

6.4 Uso de GPU

Para poder llevar a cabo este proyecto necesitábamos hacer el mayor número de pruebas posibles para poder comparar y saber qué dirección era la más adecuada para cumplir los objetivos marcados. Observamos que realizar muchos experimentos con la CPU (Intel Core i7-4770 3.40Ghz) era prácticamente imposible, debido al excesivo tiempo de procesamiento que necesitaba para poder ejecutar el algoritmo de entrenamiento. Para solucionar este problema relacionado con el tiempo de ejecución decidimos investigar cómo podíamos hacer uso de la tarjeta gráfica o GPU (NVIDIA GeForce GTX 750 Ti) y probamos varias configuraciones para que MatConvNet reconociera el Toolkit de CUDA.

Finalmente instalamos una versión de CUDA (6.5), compatible con nuestro sistema formado por Windows 8.1, Visual Studio 2013 y Matlab2014a.

Tabla 5: Tiempos de ejecución del aprendizaje de una muestra.

Devices (10 epochs)	Tiempo de ejecución (min.)
CPU	~75.43
GPU	~19.32

Como se muestra en la Tabla 5 el uso de la GPU nos proporcionó una reducción de tiempo de un 74.38%.

6.5 Calibrar cámaras

Un aspecto muy importante es la calibración de las cámaras. El modelo utilizado es el **Baslerdart uc-14** a color, dispositivo con el que trabaja *myStone*. Es la cámara encargada de la toma de imágenes. Actualmente el grupo de **Valtec** tiene 6 cámaras de este tipo y todas ellas como hemos podido analizar durante el desarrollo del proyecto toma las imágenes de forma distinta, por lo tanto para hacer más robusta la clasificación de las redes neuronales se necesitaban calibrar todas las cámaras.

Para realizar este experimento usamos las 348 muestras disponibles y fuimos variando la ganancia RGB entre los factores 0.70 y 1.30.

Si el cambio de intensidad es igual en todos los canales el resultado del clasificador no varía excesivamente del de referencia, en cambio si los cambios son de carácter específico en uno de los tres canales puede llegar a producir resultados totalmente erróneos y bajar el ratio de acierto de forma excesiva. Esto quiere decir que las cámaras se han de calibrar todas para que los resultados del clasificador sean robustos.

Calibrar una cámara significa que al tomar una imagen con dos cámaras distintas, éstas mantengan la composición de colores RGB de la imagen para que el clasificador sea capaz de clasificarlas igual, siendo la misma muestra.

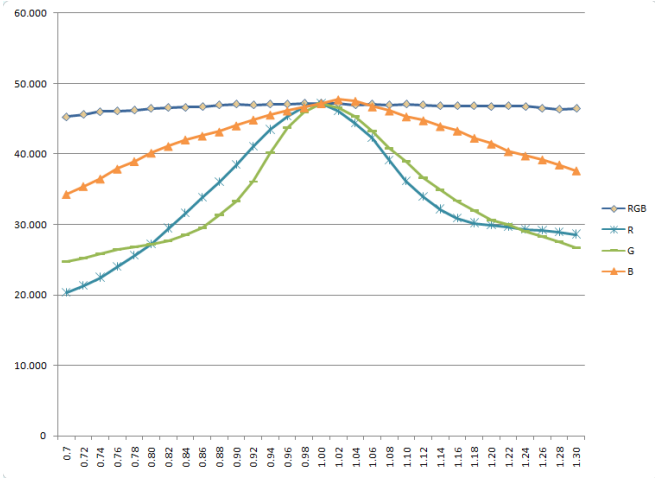


Figura 8:Comportamiento del clasificador CNN con distintas intensidades.

7 Resultados

7.1 Parametrizando las CNN

AlexNet es la red que usamos para clasificar los cálculos, contiene 21 capas [1] y la penúltima es la que contiene las clases de los cálculos. La técnica de *fine-tune* consiste en cambiar las dos últimas capas y hacer que aprenda una nueva base de datos cambiando el dominio de la red. Como hemos comentado, la penúltima capa contiene las clases que esperamos que identifique, es decir en este caso constará de 11nodo correspondientes a los tipos de vista explicados en la Sección 5.3.La última capa tendrá una función de activación *Softmaxloss* que es la que permite entrenar. Una vez la red está entrenada, si queremos clasificar y predecir tenemos que cambiar la última capa, designándole una capa *Softmax*.

La red tiene un conjunto de parámetros que se pueden variar y repercuten directamente en los resultados de clasificación final. Los parámetros que definen el aprendizaje de una CNN son los *epochs*, los cuales marcan el número de iteraciones a realizar, incluidos en estos tenemos los *batches* que marcan el subconjunto en el que se dividirán el conjunto de imágenes. Nosotros hemos utilizado 50 *batches*. Además hay otros parámetros relevantes como el *learning rate* o el *momentum*.

El *learning rate* es el factor que permite hacer que la red aprenda de forma más ágil pero con más probabilidad de equivocarse o de forma más lenta pero más robusta. Es un factor que en nuestro caso se le ha asignado un valor estático de 0.001 pero hay algoritmos que permiten adaptarlo de forma dinámica en función de la evolución del aprendizaje de la red.

El *momentum* es el factor que permite suavizar las oscilaciones del aprendizaje hasta hacerlo converger, el valor recomendado es 0.9 y es el que se utiliza en este proyecto.

7.2 Clasificación de vista

El clasificador de vistas permite identificar el tipo de cálculo a partir de una imagen.

Primeras pruebas

Los primeros experimentos fueron enfocados para organizar las muestras disponibles de forma que obtuviésemos una separación de los datos robusta y óptima. Primero decidimos dividir las muestras de forma simple, es decir, dividiendo todas las muestras en un conjunto de *train* utilizado para el aprendizaje de la red y otro conjunto de *test* utilizado para comprobar la precisión de acierto.

El mejor resultado se produjo en la Tabla 7 cuando dividimos las muestras 90 % *train* y 10 % *test*.

Tabla 6: Precisión porcentajes.

Train (%)	Precisión (%)
10	23.18
30	31.82
50	35.42
70	36.98
90	37.13

Además realizamos un estudio del *k-fold* para poder analizar su comportamiento respecto al anterior.

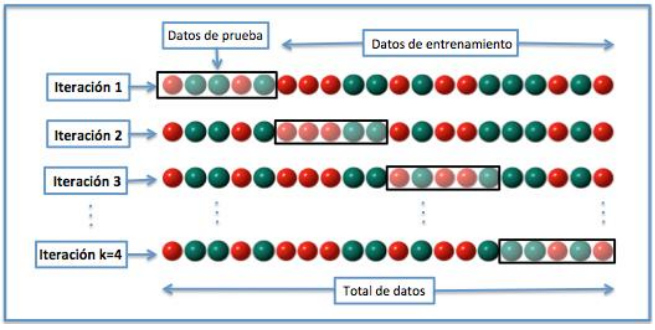


Figura 9: Técnica k-fold.

Esta técnica consiste en partir el conjunto de datos en k bloques de manera que los datos son divididos en conjuntos de *test* y *train* que se van alternando.

Se han realizado dos experimentos con esta técnica para analizar el comportamiento asignando a un valor de k de 5 y 10.

Tabla 7: Precisión 5-fold.

Tubos a clasificar	Precisión (%)
1-70	41.58
71-140	46.83
141-210	46.62
211-280	47.71
281-349	47.74
Media	45.99

Los resultados obtenidos fueron superiores a los de la Tabla 6.

Los experimentos realizados con el 10-fold fueron llevados a cabo pero dieron resultados menores que la media de la Tabla 7 y asimismo el tiempo de ejecución era el doble en un 10-fold, por lo que decidimos tomar como referencia el 5-fold para experimentos posteriores.

Imágenes virtuales

Partiendo del resultado referencia del 5-fold, 45.99 % de acierto decidimos hacer varios experimentos con el mismo conjunto de datos, pero aplicando cambios en las imágenes con tal de tener una base de datos más amplia.

A estos experimentos les fue asignada la letra B como identificador y por lo tanto el 5-fold fue denominado B0 como resultado referencia.

- B1: Conjunto de imágenes virtuales generadas con rotaciones de 90º, 180º, 270º, *flip* en el eje x y *flip* en el eje y.
- B2: Conjunto de imágenes virtuales generadas con rotaciones de 45º, 135º, 180º, 225º y 315º.
- B3: Conjunto de imágenes virtuales generadas con un leve enfoque y desenfoque.
- B4: Conjunto de imágenes virtuales generadas con escalados de 0.9, 0.95, 1.05, 1.10 y 1.15.
- B5: Conjunto de imágenes generada a partir de la combinación de los canales *red*, *green* de la imagen original y *green* de la imagen IR.
- B6: Conjunto de imágenes añadiendo todas las imágenes IR.
- B7: Conjunto de imágenes generado por la designación de centros aleatorios dentro del conjunto de la piedra y quedándonos con pequeños *patches* de 227x227.
- B8: Conjunto de imágenes generado por la designación de centros fijos dentro del conjunto de la piedra y quedándonos con pequeños *patches* de 227x227. En este experimento solo se predecía el *patch* central.
- B9: Conjunto de imágenes generado igual que en B8 pero en esta ocasión se predicen todos los *patches*.
- B10: Conjunto de imágenes aumentado por elevar el número de muestras.
- B11: Conjunto de imágenes generado quedándonos con la mitad de la información, nos quedábamos con un pixel cada dos(*Binning*)
- B12: Conjunto de imágenes generado quedándonos con la mitad de la información pero aquí añadimos 4 imágenes de la misma vista. (*Binning* parcial)
- B13: Conjunto de imágenes igual que B12 pero hacemos un sumatorio de las 4 imágenes con un rango máximo de 255.
- B14: Conjunto de imágenes igual que B12 pero hacemos un sumatorio de las 4 imágenes con un rango máximo de 255x4.

Tabla 8: Precisión con tipo B.

Tipo	Precisión (%)
B0	45.99
B1	48.098
B2	49.294
B4	46.84
B7	42.56
B13	47.011

Como se muestra en la tabla 8B2 es el experimento más robusto y nos ofrece los mejores resultados respecto a B0. B7 destaca por mejorar los resultados mostrados en la Tabla 4, donde los experimentos fueron realizados con secciones de las imágenes de tamaño 227x227x3 (16.94% de acierto).Estos resultados demuestran que añadir imágenes generadas de forma artificial a la CNN nos permite aumentar el porcentaje de acierto a la hora de clasificar. B2 además nos indica que empezamos a acercarnos al valor presentado en el estudio del arte (50 % de acierto).

Reducir el número de imágenes

La configuración física del computador para realizar el procesamiento de las imágenes, el aprendizaje de la red y las predicciones era limitada y no permitía trabajar con grandes volúmenes de imágenes, por lo que nos vimos obligados a reducir el número de imágenes base. El dispositivo *myStone* toma 8 capturas por vista y nosotros decidimos quedarnos solamente con una para poder así aumentar el número de imágenes generadas de forma artificial. Siguiendo esta línea, utilizamos las 4 imágenes con leds visibles de forma individual.

Tabla 9: Precisión

Tipo de imagen	Precisión (%)
Visible 1, baja exposición	45.652
Visible 1, alta exposición	46.667
Visible 2, baja exposición	44.203
Visible 2, alta exposición	44.928

Las imágenes con visible 1 y alta exposición fueron las que mejor resultados dieron tal como se aprecia en la Tabla 9, por lo tanto estas fueron las que elegimos para seguir experimentando. El porcentaje de acierto muestra

que el conjunto total de imágenes añadía una pequeña confusión al clasificador y por eso el pequeño aumento de la precisión en referencia al valor de B0. (45.99 % de acierto).

Los experimentos siguientes fueron etiquetados con la letra D, porque se consideran un avance, ya que permitían aumentar la capacidad de las imágenes virtuales en memoria.

- D1: Conjunto de imágenes generadas mediante rotaciones aleatorias. Utilizamos 30 imágenes por vista.
- D2: Conjunto de imágenes generadas mediante rotaciones fijas. Utilizamos 30 imágenes por vista.
- D4: Conjunto de imágenes generadas mediante escalados aleatorios entre 0,8 y 1,2. Utilizamos 30 imágenes por vista.
- D7: Conjunto de imágenes generadas mediante traslaciones aleatorias respecto al punto de masas de la imagen. Utilizamos 30 imágenes por vista.

Tabla 10: Precisión con tipo D

Tipo	Precisión (%)
D0	46.667
D1	51.039
D4	44.803
D7	46.81

El valor del experimento D1 es superior al del experimento anterior B2 (49.294 % de acierto) y finalmente conseguimos superar también el valor de referencia mencionado en el estado del arte (50 % de precisión). Esto demuestra que el aumento de imágenes de forma artificial, implica aumentar el valor de la precisión del clasificador.

Posteriormente hicimos diversas pruebas relacionadas con la técnica de clasificación *k-fold* para ver su evolución conforme aumentábamos el valor de k. El objetivo de estas pruebas era analizar la tendencia de la curva de precisión y obtuvimos los resultados esperados.

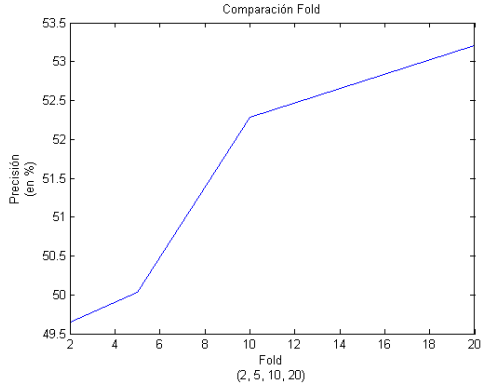


Figura 10:Comportamiento K-fold, k = 2, 5, 10, 20

Como se puede apreciar en la Figura 10, la tendencia es positiva, por lo tanto podemos esperar que aplicando la técnica de *leave one out* o haciendo un *k-fold* con la k igual al número de muestras, la tendencia sería aumentar la precisión del clasificador.

7.3 Clasificador de tubos

El clasificador de vistas, analizado hasta este punto es muy importante para poder dar por cerrado el proyecto e incluso hemos observado que ha ido mejorando conforme se hacían experimentos. Aun así para cerrar el proyecto era requisito indispensable un clasificador de tubos, es decir a partir de las 4 imágenes correspondientes a cada vista extraer cual es el tipo de tubo analizado. Es muy importante este clasificador ya que es el que se entregará al médico, es decir, el que identifica el tubo analizado y estipula el tratamiento adecuado.

Clasificador conjunto

El proceso que seguimos para hacer este clasificador fue coger las 4 imágenes correspondientes a cada vista, utilizar la CNN para clasificarlas, obteniendo una probabilidad de pertenecer a cada tipo de tubo y finalmente sumando estas probabilidades nos quedábamos con el más probable.

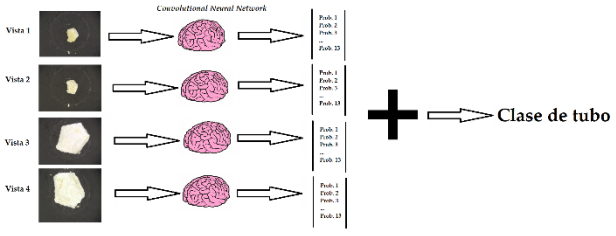


Figura 11: Clasificación de tubo

Los resultados obtenidos no fueron muy buenos pero ya éramos conscientes de que esto podía suceder porque faltan muchas pruebas a realizar con este clasificador, que es muy diferente al de vistas. Aun así aplicamos el mismo proceso que con el clasificador de vistas, es decir añadimos imágenes virtuales, generadas a partir de la rotación de las imágenes base y la Tabla 11 indica que se obtuvo un 43.403 % de acierto.

Tabla 11: Precisión clasificador de tubo

Tipo	Precisión (%)
T0	41.655
T1	43.403

Combinando SVM con CNN

Para cerrar el proyecto hicimos un último experimento mezclando las técnicas SVM y CNN, con tal de analizar qué resultados podíamos llegar a obtener.

Con la herramienta SVM, en cuanto a tubos, hemos llegado a obtener un 52 % de precisión de acierto, mientras que con las CNN solo hemos llegado a un 43 %.

La idea de mezclar estas técnicas se basa en el aprendizaje de las muestras como hemos hecho en el apartado anterior y a partir de ahí a la hora de predecir una muestra, eliminamos las capas de la red que no nos interesen y

nos quedamos con las características calculadas por esta. Una vez tenemos las características calculadas, las juntamos por tubo y finalmente usamos una SVM para hacer que aprendan estas muestras.

Las características las hemos extraído desde distintas capas con el fin de analizar cual nos ofrecía más robustez.

Tabla 12: Precisión clasificador SVM + CNN

Capa	Precisión (%)
18	26.739
16	25.336
20	20.714
18 + 16	26.471

El resultado de la Tabla 12 es inferior a lo que hemos visto en el clasificador anterior (43 % de precisión), por lo tanto concluimos que está no es la mejor línea de enfocar el clasificador de tubos.

8 Conclusiones

El desarrollo del proyecto ha cumplido los plazos de tiempo impuestos al empezar. Aun así el proyecto ha ido sufriendo pequeños cambios durante su transcurso, como el hecho de eliminar las pruebas del clasificador local para otorgar más tiempo de pruebas al global entre otros. Todos estos cambios los hemos realizado debido a que necesitábamos analizar a fondo los aspectos que ya teníamos desarrollados.

Se han alcanzado los objetivos principales de manera satisfactoria, puesto que nos hemos introducido en el mundo del *Deep Learning* y hemos conseguido aplicar la técnica de *fine-tune* a una *CNN* con el objetivo de que ésta aprendiese las muestras. Por otro lado hemos tenido muchas limitaciones, la que más ha pesado ha sido el tiempo, ya que el aprendizaje de una CNN es bastante lento.

Finalmente, hemos podido analizar el comportamiento de una *CNN* y los resultados han sido optimistas. Los clasificadores han ido aumentando la precisión de acierto en el transcurso del proyecto ya que hemos utilizado técnicas para aumentar el número de imágenes en el aprendizaje que han dado muy buenos resultados. Empezamos con un 45 % de *precisión* y hemos sido capaces de aumentar este valor un 7 % dejándolo en un 52 % de acierto aproximadamente.

8.1 Líneas futuras

Proponemos la mejora del sistema utilizado para aprender las muestras y por lo tanto, eliminar las limitaciones físicas que implica utilizar 16 GB de RAM y una tarjeta gráfica de gama media/alta. Con todos estos cambios a nivel hardware probar el rendimiento de otras *CNN* (hasta ahora estábamos limitados por la tarjeta gráfica), añadir más muestras de cálculos renales, con el aumento de capacidad de la RAM aumentar el número de imágenes virtuales generadas y profundizar en el estudio de un clasificador por tubos.

AGRADECIMIENTOS

Este trabajo ha sido realizado dentro del proyecto VALTEC13-1-0148.

Agradezco profundamente el apoyo recibido por todos los compañeros del CVC, pero principalmente a mi tutor Felipe Lumbreras, a Joan Serrat y a Gemma Rotger que son los que se han encargado de ayudarme siempre que he encontrado una dificultad, ofreciéndome todo su apoyo.

También agradezco a toda mi familia las veces que me he quedado bloqueado, que hayan tenido paciencia y me hayan apoyado en todo, especialmente a Pilar Sánchez que ha sido un gran apoyo moral durante el transcurso del proyecto.

Bibliografía

[1] A. Krizhevsky, I. Sutskever, Geoffrey E. Hinton. *ImageNet Classification with Deep Convolutional Neural Networks*, 2012.

[2] A. Vedaldi and K. Lenc, Proc. of the ACM Int. Conf. *On Multimedia. MatConvNet – Convolutional Neural Networks for MATLAB*, 2015.

[3] A. Vedaldi, K. Chatfield, K. Simonyan, A. Zisserman. *Return of the devil in the Details: Delving Deep into Convolutional Nets*, 2014.

[4] A. Puigvert. *La litogenesis del riñón. Anales de medicina y cirugía*. (1982) Vol. 60 No. 262.

[5] A Basiri, M Taheri and F Taheri *What is the state of the stone analysis techniques in urolithiasis?* Urology journal, 2012.

[6] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Angue-lov, D. Erhan, V. Vanhoucke, A. Rabinovich. *Going Deeper with Convolutions*, 2015.

[7] D. C. Ciresan, U. Meier, J. Masci, L. M. Gam-bardella, J. Schmidhuber. Flexible, *High Performance Convolutional Neural Networks for Image Classification*.

[8] F. Grases, B. Isern, P. Sanchis, J. Perello, J. J. Torres and A. Costa-Bauza. *Phytate acts as an inhibitor in formation of renal calculi. Frontiers in pilot study in Wistar rats*. Life Sciences. Volume 75, Issue 1, 21 May 2004, Pages 11-19.

[9] F. Grases, A. Costa-Bauzat, R. M Prieto. *Renal lithiasis and nutrition. Nutrition Journal* 2006, 5. 23.

[10] F. Grases, A. Costa-Bauza, M. Ramis, V. Montesinos, A. Conte. *Simple Classification of renal calculi closely related to their micromorphology and etiology*, 2002.

[11] F. Blanco. *Chemical Speciation on Urinary Lithiasis, Image Analysis and Separation Techniques for the study of Lithogenesis*, 2014.

[12] G. Bihl, A. Meyers. *Recurrent renal stone disease – advances in pathogenesis and clinical management*. The lan-cet. Volume 358, Issue 9282, 25 August 2001, Pages 651–656

[13] G. Rotger Moll. *A comparative study of renal calculi classification according to their morphology using machine vision techniques*, 2015.

[14] G. Kravda, D. Helgo y M.K. Moe. *Infrared spectrometry is the gold standard for kidney Stone analysis Tidskr Nor Legeforen 2015; 135-313 - 4*

[15] S. Charafi, M. Mbarki, A. Costa-Bauza, R. M. Prieto, A. Oussama, F. Grases. *A comparative study of two renal Stone analysis methods*, 2010.

[16] V. Romero, H. Akpinar and D. G. Assimos.*Kidney Stones: A Global Picture of Prevalence, Incidence and Associated Risk Factors*. Reviews in Urology, Vol 12, No 2/3 2010.

[17] V. B. Nalbandyan. *X-Ray diffraction analysis of urinary calculi: need for heat treatment.* Urol Res. 2008 Oct; 36(5):247-9.

[18] Y. Jia, E. Shelhamer, J. Donahue, S. Kara-yev, J. Long, R. Girshick, S. Guadarrama, T. Darrell. *Caffe: Convolutional Architecture for Fast Feature Embedding*, 2014.

APÉNDICES

A1. CONCEPTOS

- **Capa *Softmax:*** Permite normalizar los resultados evitando valores extremos u *outliers*, en el ámbito de las CNN se utiliza en la última capa para la clasificación. Durante el proyecto se utiliza para clasificar las muestras.
- **Capa *Softmaxloss:*** Conceptualmente es igual que la capa anteriormente descrita, pero aporta más estabilidad en el gradiente. Se utiliza para aprender las muestras. Durante el proyecto se utiliza para aprender el conjunto de muestras.
- ***Epochs***: Medida que permite definir cuantas veces utilizará la CNN los datos de aprendizaje para actualizar los pesos de la red.
- ***Batches***: Subconjunto de los datos, que se le enseñan a la CNN de forma simultánea, es decir si hay 100 muestras y este valor es de 10, dentro de un *epoch* se le enseñarán 10 lotes distintos de datos a la CNN.
- ***Learning rate:*** Factor que permite el aprendizaje de la red de una forma más ágil pero menos robusta, o más lenta pero más robusta.
- ***Momentum:*** Factor que suaviza las oscilaciones en el aprendizaje hasta que hacerlo converger.