

Detección de Imágenes Pornográficas mediante Redes Neuronales Artificiales.

José Carlos Cortizo Pérez y Diego Expósito Gil

Artificial Intelligence & Network Solutions, C/Bélgica 7 Fuenlabrada (Madrid)
{jccp, deg}@ainetsolutions.com

Resumen. El crecimiento exponencial del uso de Internet así como las innatas capacidades de los niños para el manejo de los ordenadores y el uso creciente de Internet en los puestos de trabajo ha desatado la necesidad de filtrar ciertos datos accesibles a través de Internet. Filtrar los textos referentes a pornografía es una tarea relativamente sencilla por el lenguaje utilizado y porque el tratamiento estadístico del lenguaje resulta muy efectivo en estos ámbitos, pero al tratar con imágenes todo cambia pues el concepto de una fotografía pornográfica es un concepto muchísimo más abstracto y las técnicas estadísticas no sirven aquí para clasificar a las imágenes en pornográficas o no pues sus unidades básicas, los píxeles, no aportan información de por sí.

1 Introducción

Internet ha pasado, en un periodo relativamente corto de tiempo, a ser una fuente inagotable de todo tipo de contenidos, desde investigaciones, noticias, etc. hasta contenidos violentos, xenófobos o pornográficos. A la par, se ha ido instaurando en los trabajos y en las casas de una forma exponencial por lo que, hoy por hoy, en la mayoría de los puestos de trabajo y en la mayoría de las casas se utiliza Internet con frecuencia o, por lo menos, existe la posibilidad de acceso. Esto ha marcado nuevas necesidades pues en los puestos de trabajo se busca rendimiento y, por tanto, se quiere evitar que el trabajador acceda a según que contenidos que no aumentan la productividad, como es el caso de la pornografía. En las casas los más jóvenes son los que más utilizan las nuevas tecnologías y, en concreto Internet; así pues es conveniente limitar el acceso a ciertos contenidos, no aptos para esas edades.

Los datos disponibles a través de Internet suelen ser de dos clases: textuales o audiovisuales. El tratar datos textuales es más sencillo pues mediante técnicas estadísticas y de Aprendizaje Automático, es relativamente sencillo el saber si una web es pornográfica o no (en cuanto al texto). Pero con las imágenes la problemática es muy distinta porque la misma estructura de una imagen es algo realmente abstracto (una matriz de puntos de color) y, además, el concepto de lo que es pornografía en cuestión de imágenes es bastante difícil de definir. Si una página web contiene la palabra “sexo”, muy probablemente sea pornográfica pero para determinar si una imagen es pornográfica han de aparecer personas humanas manteniendo relaciones sexuales. Y explicarle lo que es una relación sexual a una máquina es algo realmente inviable.

El tratamiento estadístico de los píxeles (podría decirse que son a la imagen lo que las palabras al texto) no puede ofrecer un resultado viable pues definiría un porcentaje de zonas de color y de colores lo cuál no representa mucha información pues no sería capaz de distinguir una mujer desnuda de una mujer con bañador o de una piedra de un color y forma similares. Para poder conocer si una imagen es pornográfica o no, se ha de analizar con las medidas y herramientas propias de las imágenes y posteriormente realizar el aprendizaje o selección sobre estos datos. Así pues, los datos serán generalmente numéricos y muy probablemente representarán una buena cantidad por cada imagen.

Contando con lo anterior y con que básicamente se trata de hacer una tarea de categorización una de las posibilidades es el Aprendizaje Neuronal, que es el finalmente elegido para el estudio de las imágenes y la valoración de si son pornográficas o no. Cabe reseñar que el uso de las redes de neuronas en trabajos de aprendizaje sobre imágenes está extendido y algunos de los resultados obtenidos son alentadores como los de [mitchell97] o [bishop96].

2 Trabajos Relacionados

La pornografía en Internet es un tema ciertamente preocupante y muestra de ello son las innumerables investigaciones que se realizan sobre el filtrado de estos contenidos. De hecho solo en el periodo 1999-2002, la Unión Europea ha invertido más de 9 millones de euros en investigaciones dedicadas a obtener filtros para este tipo de contenidos, así como contenidos racistas y xenófobos. A nivel europeo, uno de los proyectos con más éxito es POESIA (Public Opensource Environment for a Safer Internet Access) descrito en [gomez02]. POESIA ha supuesto la creación de un filtro de contenidos pornográficos, violentos y xenófobos open-source basado en el filtrado tanto de contenidos textuales como de las imágenes. En cuanto al texto las técnicas utilizadas se basan en técnicas típicas de procesamiento del lenguaje natural, descritas en [gomez03]. En cuanto al filtrado de información contenida en imágenes POESIA presenta un acercamiento pobre pues, en cuanto a pornografía, se centra en el uso exclusivo de filtros de detección de piel [fleck96] que realmente no funcionan con un alto grado de fiabilidad (43% de efectividad según [forsyth96]).

Como continuación a POESIA, que es un proyecto orientado a filtros para entornos escolares, han surgido dos proyectos que continúan el cometido de POESIA pero orientándolo a entornos empresariales, son TEFILA (Técnicas de Filtrado basadas en Ingeniería del Lenguaje, Aprendizaje automático y Agentes) y TEFILA2, bajo financiación del programa de fomento de la investigación técnica (PROFIT). En TEFILA el reconocimiento de imágenes cobra mayor relevancia pues POESIA ha arrojado unos resultados muy satisfactorios en cuanto al filtrado de texto, por lo que la asignatura pendiente de los filtros queda reducida a las imágenes. Para intentar solventar este problema, en TEFILA se han generado unos vectores de características por cada imagen a partir de los trozos de piel detectados en las imágenes, a partir de varios filtros, y a continuación se pretende utilizar estos vectores para conseguir discernir entre fotografías pornográficas y no pornográficas, pero para ello se han de utilizar técnicas de aprendizaje automático más rápidas que las redes de neuronas

debido a que el tiempo de respuesta para un filtro real ha de ser muy rápido. Por ahora TEFILA sigue en marcha y está en el punto de aprender de esos vectores.

Existen algunos otros proyectos dedicados a la detección de pornografía pero casi todos ellos basados en el uso de detectores de piel, estableciendo criterios como que si tiene más de un X% de piel, la fotografía es pornográfica.

3 Procesado de las Imágenes

Antes siquiera de pensar en las redes a usar y en las topologías a definir, hay que comprender muy bien la naturaleza de una fotografía para estudiar que tipo de entradas hay que introducirle a la red para que pueda aprender sobre esa entrada.

Básicamente una fotografía es una lista de tríos de valores entre 0 y 255; cada trío representa un píxel de la foto y dentro de ese conjunto hay un valor para el componente de rojo (R), otro para el verde (G) y otro para el azul (B). Los valores van entre 0 (ausencia de ese color) y 255 (presencia total de ese color). Así pues:

$$pixel = \{R, G, B\} \quad R, G, B = \{0 \dots 2^8 - 1\}$$

Con esto se puede afirmar que existen un total de $2^8 \times 2^8 \times 2^8 = 2^{24} = 16,777,216$ colores por cada píxel lo cuál es una gran cantidad de información. Debido a este hecho se ha descartado el ofrecer una entrada a la red que consista en una serie de píxeles y, por cada uno, su valor de color obtenido mediante la correspondencia entre sus colores y el valor para el píxel pues al normalizar al rango entre 0 y 1, la distancia entre colores sería demasiado pequeña. En este caso para obtener el valor del píxel habría que hacer:

$$pixel = dec(R \cdot G \cdot B) / 255$$

De esta forma, teniendo los valores del píxel (243, 126, 35), bastaría con concatenar las representaciones binarias de los colores del píxel y normalizarlo en el intervalo 0, 1. El resultado para este caso sería de $11110011 \cdot 01111110 \cdot 00100011 = 111100110111111000100011$ que representado en decimal daría el valor 15957539 y normalizado al intervalo de entrada de una red neuronal 0,951143443584442138671875. El problema es que el siguiente color está a menos de 0,0000000625 de este mismo por lo que la capacidad de discernir las entradas de la red sería muy reducida.

Para solventar este problema se han buscado representaciones más simples. La primera y más sencilla es dar como entrada una lista formada por $3 \times N \times M$ (siendo N y M las dimensiones en píxeles de la imagen) entradas y cada una de ellas representando el valor correspondiente para el rojo, verde o azul (normalizados dividiendo por 255). Una vez procesado así hay que tener en cuenta el gran número de entradas pues las fotografías suelen tener unas dimensiones mínimas de 300x300 píxeles (lo cuál lleva a 270,000 entradas) y además estas dimensiones son variables. Para poder solucionar esto lo que se ha hecho es tomar 2 tamaños estándar, uno a

30x30 y otro a 90x90 y redimensionar todas las imágenes a esos dos tamaños para que las entradas de la red sean homogéneas (ver figura 1).

Este primer acercamiento a la entrada de la red es un acercamiento bastante trivial pues da como entrada toda la información presente en una fotografía, la cuál es mucha y puede, incluso, llegar a ser demasiada información como para que se generalicen los datos. Con esto en mente se han buscado otras formas de representar las imágenes para ofrecerlas como entradas a la red.

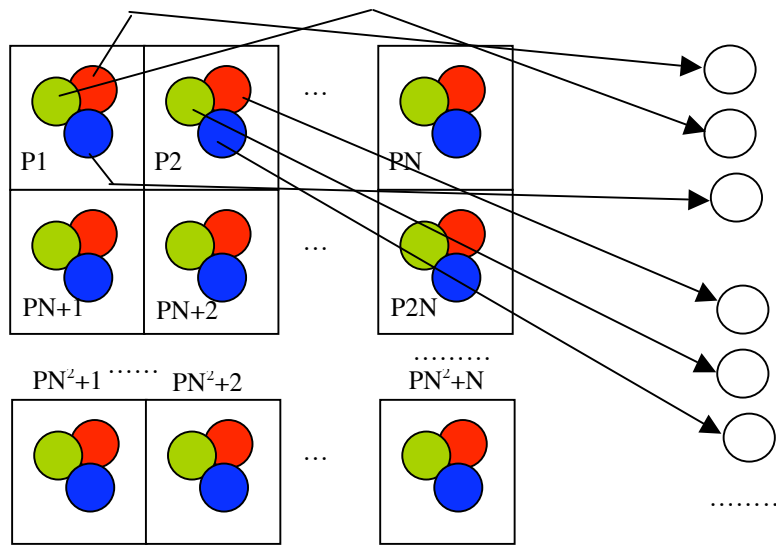


Fig. 1. Representación de la entrada a una red a partir del formateo de las imágenes como una serie de tríos R,G,B por cada píxel después de escalar las imágenes a dimensión $N \times N$.

Como consecuencia lógica de la anterior representación surge el simplificarla para agrupar los 3 valores de cada píxel en uno pero de una forma simplificada para no caer en el error comentado anteriormente. De esta forma se crea un proceso denominado Segmentación y Ensamblaje de los colores. La idea consiste tanto en reducir el número de entradas pasando de 3 a 1 por cada píxel, como reducir el número total de colores de las imágenes para que la diversidad de cada entrada sea menor, y así, que exista una mayor separación entre cada valor de la entrada para que la red pueda discernir mejor entre las entradas.

Para lograr eso se ha realizado un proceso de selección de rangos en cada canal de color para, posteriormente, ensamblar el resultado en un único valor por píxel. Para cada píxel se coge sus valores de R, G y B, cada uno de los cuales puede valer entre 0 y 255. Estos rangos se segmentan en M rangos, pudiendo M valer 2, 3 o 5 (estas son las opciones elegidas). Una vez determinado el valor de M , se divide el rango total de un color (256 valores) en M partes, estableciendo un valor determinado para cada

parte del rango (entre 0 y M-1). Esto se hace para cada componente de color del píxel, obteniendo pues 3 números en base M, uno para R, otro para G y otro para B. Para componer el valor del píxel global, basta con asignar a cada combinación de valores de R, G y B un único valor, para ello:

$$pixel = R \times M^2 + G \times M + B$$

Básicamente este proceso consiste en crear un número en base M con las 3 componentes y pasarlo a decimal. Una vez obtenido el nuevo número en decimal basta con normalizar el resultado al rango de entradas de la red, para ello se divide entre el valor máximo posible que será 7 para segmentación 2, 26 para segmentación 3 y 124 para segmentación 5. Una visión gráfica de este proceso se puede ver en la figura 2.

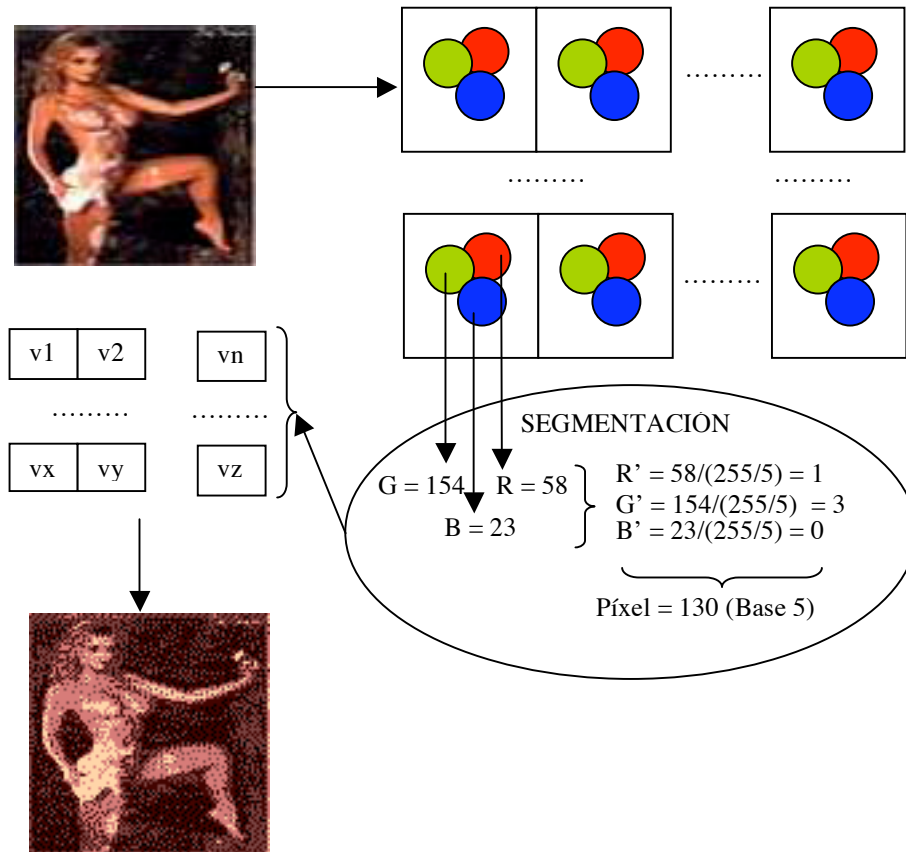


Fig. 2. Representación del proceso de segmentación de una fotografía pornográfica mediante y su resultado al aplicar sobre todos los píxeles de la foto ($M = 5$).

Así pues, con las segmentaciones se obtienen 3 tipos de entradas, con M valiendo 2, 3 o 5. Además esto se aplica tanto con N valiendo 30 como con N valiendo 90, lo cuál genera 6 posibles combinaciones de pares (N, M) que luego serán pruebas con las redes: (30, 2), (30, 3), (30, 5), (90, 2), (90, 3), (90, 5).

Hasta aquí se ha tratado de transformaciones más o menos simples de las imágenes, al menos son transformaciones conservativas pues conservan la foto en mayor o menor grado y se la dan como entrada a la red. Otras transformaciones pueden dar información de la imagen pero de forma implícita. De hecho una de las técnicas que más datos da sobre una fotografía es el Histograma, o lo que es lo mismo, el porcentaje de píxeles que existen en la foto con cada uno de los 256 posibles valores distintos para la luminosidad.

Para darle esta información a la red primero se procesan las fotografías extrayendo valores de luminosidad en lugar de colores, según la fórmula de cálculo de luminosidad (Y) de Craig Markwart que dice que la luminosidad se puede hallar a partir de los colores del píxel de la siguiente forma:

$$Y=0.3 \times R + 0.59 \times G + 0.11 \times B$$

Una vez realizados estos cálculos para todos los píxeles basta con examinar la fotografía al completo calculando el número de píxeles para cada valor de la luminosidad (256 valores distintos) y, posteriormente, transformar esos valores a porcentajes. De esta forma se consigue, además, un número de entradas fijo y reducido para cualquiera que sea el tamaño de la fotografía. Se puede ver en la figura 3 la relación entre una fotografía y su histograma.

Además de ofrecer el histograma completo a la red neuronal, también se puede ofrecer simplemente los datos de mayor interés como son los 10 valores de luminosidad con mayor porcentaje o los porcentajes de luminosidad segmentados en 10 rangos de forma que se acumulan los valores de cada uno de los rangos.

2 Tipos de Redes Neuronales Utilizadas

Una vez determinadas las entradas a la red hay que ver qué tipos de redes neuronales conviene utilizar para resolver el problema. Ante todo se busca la resolución de un problema en el que las determinaciones sobre la clasificación tienen una componente subjetiva muy importante. De hecho una fotografía podría ser considerada o no pornografía dependiendo del país dónde se realice la clasificación, o las implicaciones religiosas o morales de la persona en cuestión, o incluso el momento socio-político. Teniendo en cuenta todo esto, se pretende que la red aprenda a partir de la clasificación determinada de antemano, con las implicaciones culturales de los que realizan la clasificación pues se ha realizado de acuerdo al entorno sobre el que

trabajaría el filtro que podría ser escolar o empresarial. Así pues el tipo de red neuronal ha de poder trabajar de modo supervisado, por lo que se han descartado los mapas autoorganizativos porque realizan un clustering según sus criterios aunque podría haber sido interesante probarlos.

Los tipos de redes finalmente seleccionadas han sido el perceptrón multicapa retroalimentado ([fausett94], [gurney97], [haykin99]) y las redes de base radial ([broomhead88], [lee88]) mediante el simulador [arias], por ofrecer modalidades de aprendizaje supervisado y minimizar el gradiente del error con unos resultados aceptables en algunos otros entornos que también trabajaban sobre fotografías ([mitchell97], [pankiewicz97]).

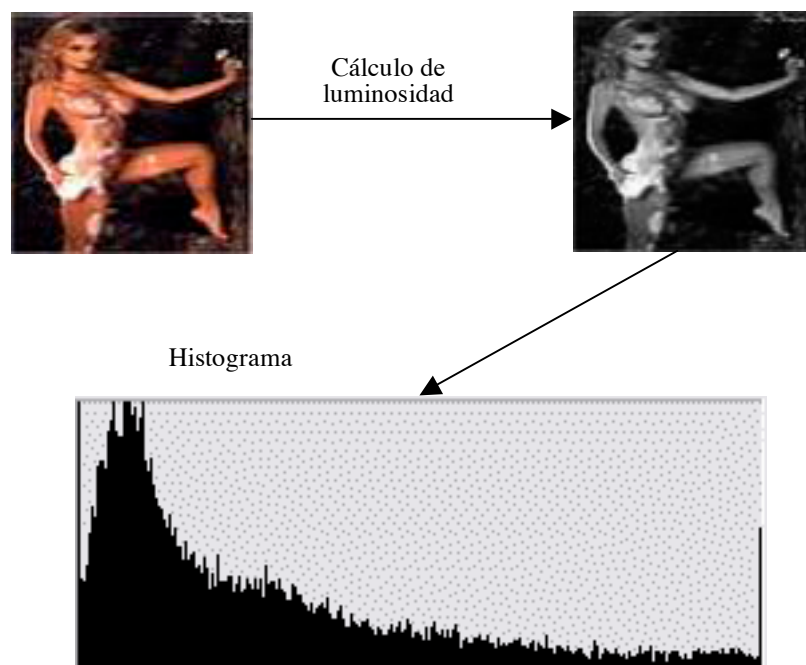


Fig. 3. Proceso de extracción del histograma de una imagen, mediante conversión previa a escala de luminosidad

Dentro de cada tipo de red existen múltiples parámetros sobre los cuáles experimentar para estudiar los resultados, las variaciones para estos parámetros se estudiarán en la fase experimental para verlos junto con sus resultados.

5 Conjuntos de Datos

A la hora de realizar el aprendizaje es cuestión fundamental definir los datos sobre los que se van a aprender y evaluar los resultados. Con anterioridad se ha estudiado como se iban a procesar los datos para que la red pudiera aprender pero en ahora lo que importa es qué tipo de datos, qué tipo de fotografías, qué numero de ellas y en qué proporción se van a utilizar.

Se han recopilado de forma manual 2300 fotografías de distintos medios y de distintas dimensiones y formatos gráficos y se han clasificado también de forma manual en fotografías pornográficas y no pornográficas. El criterio utilizado para la selección ha sido decir que pornografía es todo aquello en lo que aparecen cuerpos desnudos sea cuál sea la finalidad de la fotografía. Todo lo que no entre dentro de estas características será considerado no pornográfico, por lo que en estas fotografías se puede encontrar paisajes, personas, etc. (ver figura 4).



Fig. 4. Detalle de distintas fotografías consideradas como no pornográficas escaladas a 90x90, antes de ser tratadas para poderlas dar como entrada a la red.

Evidentemente lo que se considera pornografía incluye un gran rango de vertientes tales como fotografía de desnudos artísticos, nudismo, sexo, etc. Un ejemplo de distintas fotografías pornográficas se puede ver en la figura 5.

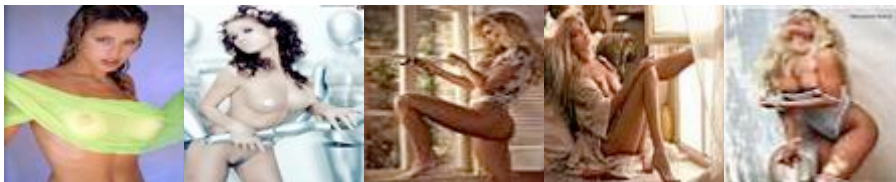




Fig. 5. Detalle de diversas fotografías consideradas como pornográficas antes de ofrecerlas como entrada a la red neuronal.

Las 2300 fotografías se dividen en 1200 pornográficas y 1100 no pornográficas. En cuanto a la división para el aprendizaje se han escogido un 70% para el entrenamiento y un 30% para la prueba de forma aleatoria, obteniendo para las fotografías de 30x30 1603 fotografías para entrenamiento y 689 para prueba y, para las fotografías de 90x90, 1489 fotografías para entrenamiento y 638 para la prueba.

6 Experimentación

Una vez definidas las entradas a la red y los tipos de redes a utilizar, queda probar como de bien clasifican estas redes las imágenes que se les ofrecen como entradas. Para ello se ha probado cada tipo de entrada a la red con cada red y, a su vez, estas con distintos tipos de parámetros y topologías. Hay que tener en cuenta que se han preparado once tipos diferentes de entrada (todos los colores para todos los píxeles para 30x30, lo mismo para 90x90, segmentación a 2, 3 y 5 niveles tanto para 30x30 como para 90x90, el histograma, la segmentación del histograma a 10 niveles y los 10 mejores valores) y que estas 11 combinaciones hay que probarlas con cada red y a su vez cada red con múltiples variantes en valores y topologías.

En cuanto al proceso de experimentación, lo que se ha hecho es repetir cada experimento varias veces (tres veces) para comprobar que los resultados finales son más o menos equivalentes, si los resultados eran demasiado variables se han descartado por no ofrecer consistencia.

6.1 Experimentos con Redes de Base Radial

Para las Redes de Base Radial basta con definir el número de neuronas de entrada, el número de neuronas ocultas y el número de neuronas de salida. Se pueden modificar algunos parámetros como la tasa de aprendizaje pero se han tomado valores estándar [patterson96] para no realizar un crecimiento exponencial del número de experimentos. Los parámetros que sí varían en este tipo de red son el número de neuronas en la capa oculta y, desde luego, el tipo de entrada ofrecida, en cuanto a la tasa de aprendizaje se maneja siempre el valor de 0,01.

6.1.1 Imágenes a 30x30 con todos los valores de color y 2 Neuronas ocultas

Empezando con las entradas con menor nivel de abstracción, se prueba el introducir a la red fotografías escaladas a 30x30 píxeles y ofrecer, por cada píxel, sus 3 componentes de color (R, G y B) normalizadas al intervalo (0,1). El número de neuronas ocultas seleccionado ha sido de 2, teóricamente una por cada posible clasificación, de forma que cada centro representaría el centro de masas de las fotografías de un tipo (pornográficas o no pornográficas).

Los resultados obtenidos permiten observar a simple vista (ver figura 6) que la red no generaliza demasiado bien pues el error de prueba sube cuánto más se entrena, pero también hay que entender que los resultados obtenidos desde un inicio son bastante aceptables (0.0728 de error en prueba y 0.054 de error en entrenamiento, lo cuál es, incluso, inferior a las cotas deseadas).

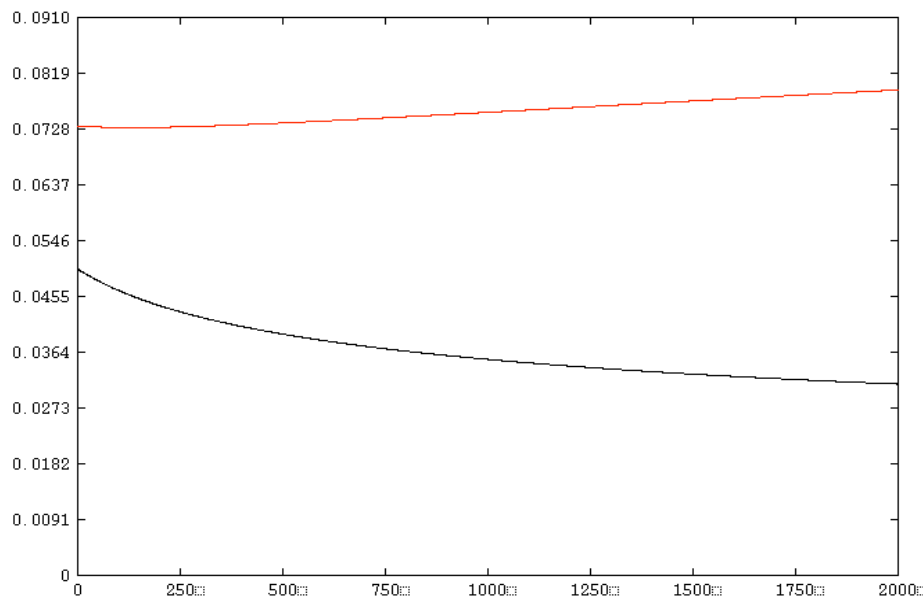


Fig. 6. Gráfica que representa la evolución del entrenamiento para este experimento. La línea roja denota los resultados en las pruebas y la negra en el entrenamiento.

Viendo la gráfica (Figura 6) se puede apreciar que la red aprende y da buenos resultados pero para valorar mejor estos conviene estudiar la siguiente tabla que muestra cómo se clasifican los ejemplos del conjunto de entrenamiento:

Tabla 1. Tabla que muestra cómo se clasifican los elementos de cada categoría. En cada fila se establece para cada categoría el número de elementos que el sistema ha puesto en cada categoría.

	No Porno	Porno
No porno	273	35
Porno	91	290

A partir de los datos de la Tabla 1 se concluye que se clasifican correctamente el 81,7126% de los elementos del conjunto de prueba. Con estos mismos datos se construye la Tabla 2 en la cuál se pueden estudiar los valores de recall y precisión de cada categoría lo cuál aporta información sobre la proporción de fotografías recuperadas por categoría sobre las fotografías totales de esa categoría (porcentaje de acierto por categoría) y la proporción de documentos relevantes recuperados sobre todos los recuperados, es decir, el porcentaje de elementos que el sistema dice que son de esa categoría que realmente lo son.

Tabla 2. Recall y precisión por cada una de las categorías.

	Recall	Precisión
No porno	0,886364	0,750000
Porno	0,761155	0,892308

6.1.2 Imágenes a 30x30 con todos los valores de color y 10 Neuronas ocultas

Una vez vistos los impresionantes resultados conseguidos con una simple red de dos neuronas ocultas, el siguiente paso es intentar aumentar el número de neuronas ocultas para aumentar el espacio de búsqueda. Para ello se han probado 10 neuronas ocultas para poder evaluar la mejora o no sobre el anterior experimento.

Tabla 3. Tabla que muestra cómo se clasifican los elementos de cada categoría. En cada fila se establece para cada categoría el número de elementos que el sistema ha puesto en cada categoría.

	No Porno	Porno
No porno	282	26
Porno	87	294

Los resultados obtenidos permiten observar a simple vista (ver figura 7) que la red generaliza bastante mejor. De hecho el error inicial ahora es mayor y la red aprende bastante bien, generalizando, moviéndose desde valores de error de 0,13 a 0,052 en entrenamiento y a un mínimo de 0,074 en prueba.

En la tabla 3 se puede apreciar la distribución que da el sistema de los ejemplos de prueba sobre cada categoría y en la tabla 4 los valores de recall y precisión. En cuanto a la precisión total del sistema, en este experimento aumenta hasta valores de 83,5994 % lo cuál es un resultado muy bueno.

A partir de estos valores se puede extraer una tabla de recall y precisión que permite estudiar mejor la correcta clasificación por cada categoría (ver Tabla 4).

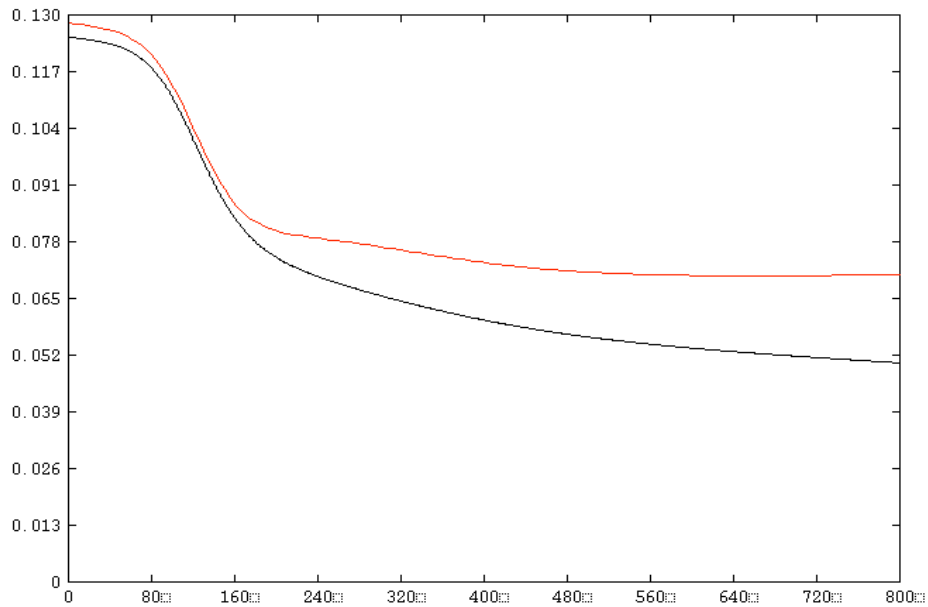


Fig. 7. Gráfica que representa la evolución del entrenamiento para este experimento. La línea roja denota los resultados en las pruebas y la negra en el entrenamiento.

Tabla 4. Recall y precisión por cada una de las categorías.

	Recall	Precisión
No porno	0,915584	0,764228
Porno	0,771653	0,918750

6.1.3 Imágenes a 30x30 con todos los valores de color y 20 Neuronas ocultas

Al analizar la mejora de la clasificación al aumentar las neuronas ocultas se decidió seguir aumentando su número, ahora serían 20 neuronas ocultas. Todo lo restante se mantiene igual.

Los resultados obtenidos permiten observar a simple vista (ver figura 7) que la red a pesar de generalizar no clasifica tan bien, se queda en un 82,4383%.

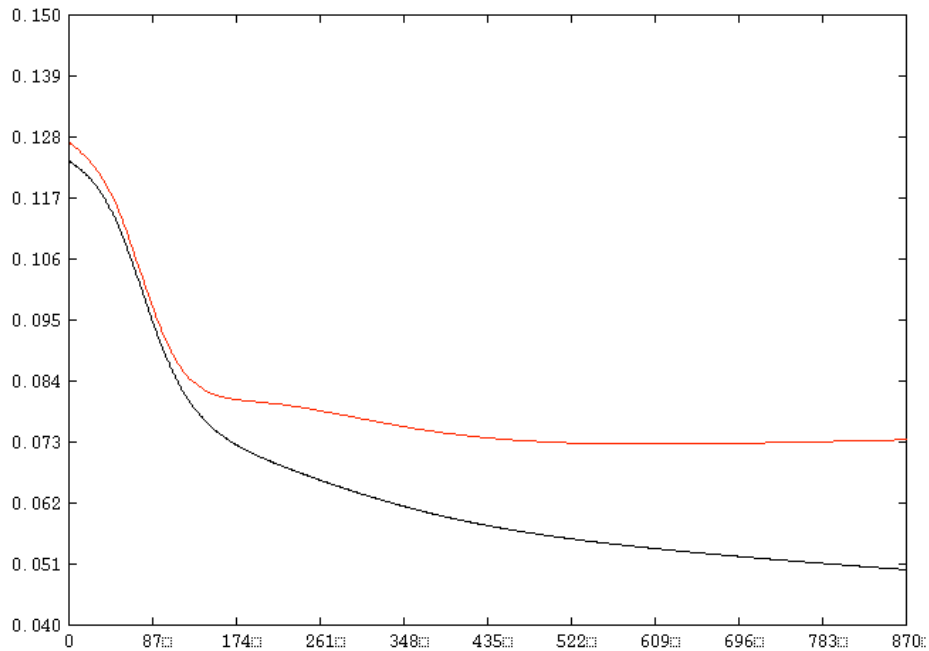


Fig. 8. Gráfica que representa la evolución del entrenamiento para este experimento. La línea roja denota los resultados en las pruebas y la negra en el entrenamiento.

Tabla 5. Tabla que muestra cómo se clasifican los elementos de cada categoría. En cada fila se establece para cada categoría el número de elementos que el sistema ha puesto en cada categoría.

	No Porno	Porno
No porno	283	25
Porno	96	285

Tabla 6. Recall y precisión por cada una de las categorías.

	Recall	Precisión
No porno	0,918831	0,746702
Porno	0,748031	0,918354

6.1.4 Imágenes a 30x30 segmentadas a 5 niveles por color y 2 Neuronas ocultas

Una vez probado que el sistema es capaz de clasificar a partir de los valores de los píxeles sin más, se va a tratar de simplificar la entrada de la red para evaluar las repercusiones que pueda entrañar el reducir el número de valores posibles de la entrada, así como el número de entradas.

Tabla 7. Tabla que muestra cómo se clasifican los elementos de cada categoría. En cada fila se establece para cada categoría el número de elementos que el sistema ha puesto en cada categoría.

	No Porno	Porno
--	-----------------	--------------

No porno	230	78
Porno	110	271

La red parece generalizar bien, aprende con rapidez en un principio oscilando desde valores de 0,2 de error hasta valores de 0,07 en entrenamiento y 0,089 en prueba, pero los resultados en cuanto a clasificación son bastante peores. La precisión obtenida es del 72,7141% de aciertos, que dista bastante del casi 84% de anteriores experimentos.

Tabla 8. Recall y precisión por cada una de las categorías.

	Recall	Precisión
No porno	0,746753	0,676470
Porno	0,711286	0,776504

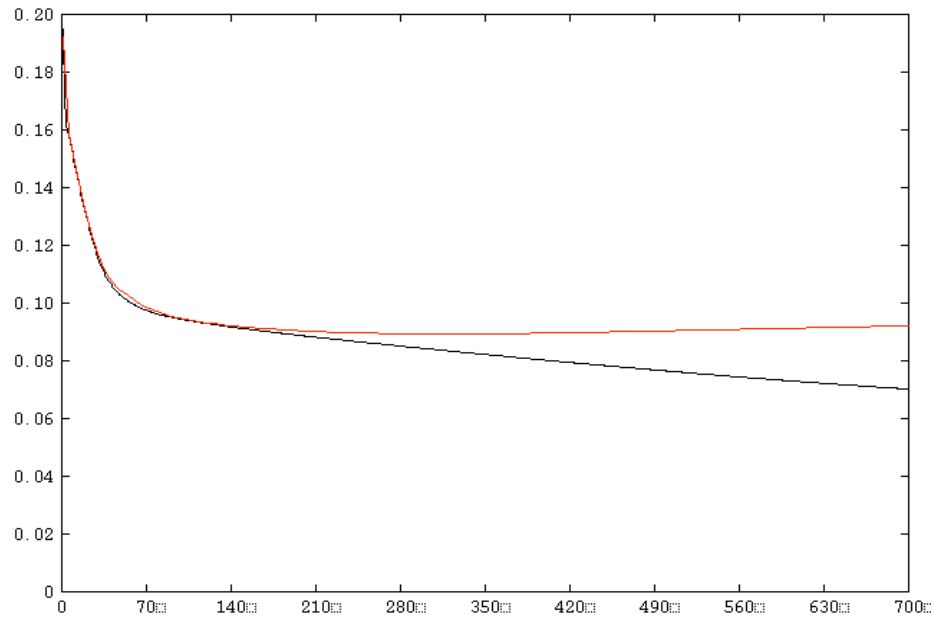


Fig. 9. Gráfica que representa la evolución del entrenamiento para este experimento. La línea roja denota los resultados en las pruebas y la negra en el entrenamiento.

6.1.5 Imágenes a 30x30 segmentadas a 5 niveles por color y 10 Neuronas ocultas

En anteriores experimentos aumentar el número de neuronas ocultas de 2 a 10 aumentó la calidad de la clasificación por lo que aquí también probaremos esto, para ver si se mantiene el mismo comportamiento.

Basta ver los resultados obtenidos para ver que no merece la pena, en este tipo de segmentación, aumentar las neuronas ocultas.

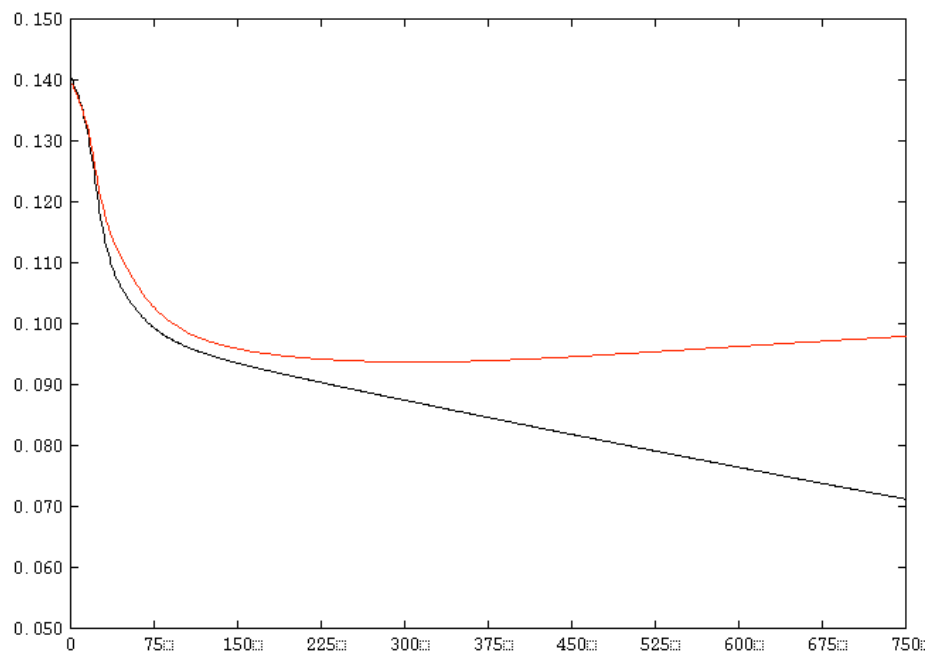


Fig. 10. Gráfica que representa la evolución del entrenamiento para este experimento. La línea roja denota los resultados en las pruebas y la negra en el entrenamiento.

Tabla 9. Tabla que muestra cómo se clasifican los elementos de cada categoría. En cada fila se establece, para cada categoría, el número de elementos que el sistema ha puesto en cada categoría.

	No Porno	Porno
No porno	232	76
Porno	114	267

Tabla 10. Recall y precisión por cada una de las categorías.

	Recall	Precisión
No porno	0,753246	0,670520
Porno	0,700787	0,778425

6.1.6 Imágenes a 30x30 segmentadas a 2 niveles por color y 2 Neuronas ocultas

Se reduce la segmentación para experimentar si aún con menos niveles en la entrada se consiguen resultados aceptables.

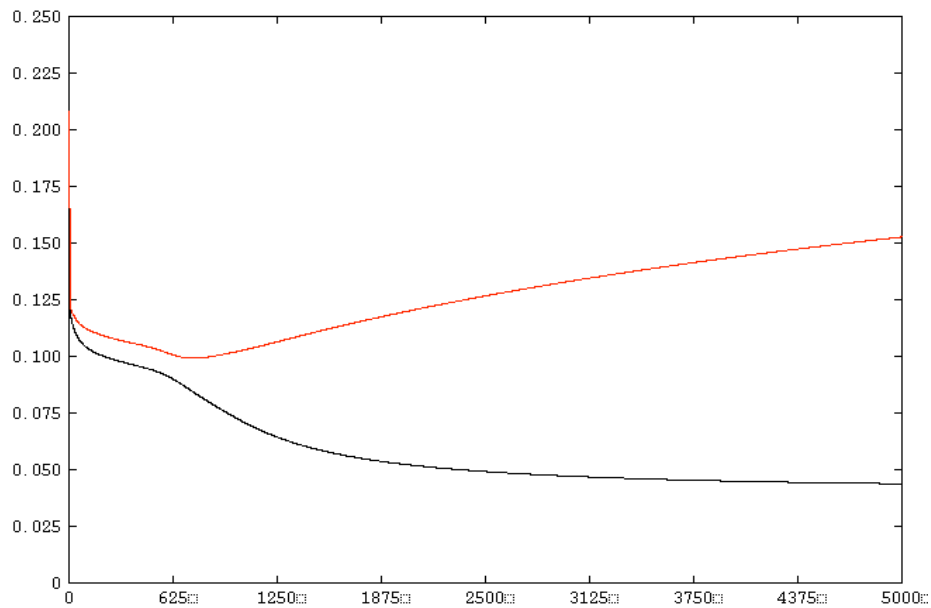


Fig. 11. Gráfica que representa la evolución del entrenamiento para este experimento. La línea roja denota los resultados en las pruebas y la negra en el entrenamiento.

Tabla 11. Tabla que muestra cómo se clasifican los elementos de cada categoría. En cada fila se establece, para cada categoría, el número de elementos que el sistema ha puesto en cada categoría.

	No Porno	Porno
No porno	198	110
Porno	137	244

Tabla 12. Recall y precisión por cada una de las categorías.

	Recall	Precisión
No porno	0,642857	0,591044
Porno	0,640419	0,689265

6.1.7 Imágenes a 90x90 segmentadas a 5 niveles por color y 2 Neuronas ocultas

Para fotografías más grandes la segmentación reduce notablemente el número de entradas y la carga computacional. El resultado es de 72,4138% de acierto en la clasificación.

Tabla 13. Tabla que muestra cómo se clasifican los elementos de cada categoría. En cada fila se establece, para cada categoría, el número de elementos que el sistema ha puesto en cada categoría.

	No Porno	Porno
No porno	224	84
Porno	92	238

Tabla 14. Recall y precisión por cada una de las categorías.

	Recall	Precisión
No porno	0,727272	0,708861
Porno	0,721212	0,739130

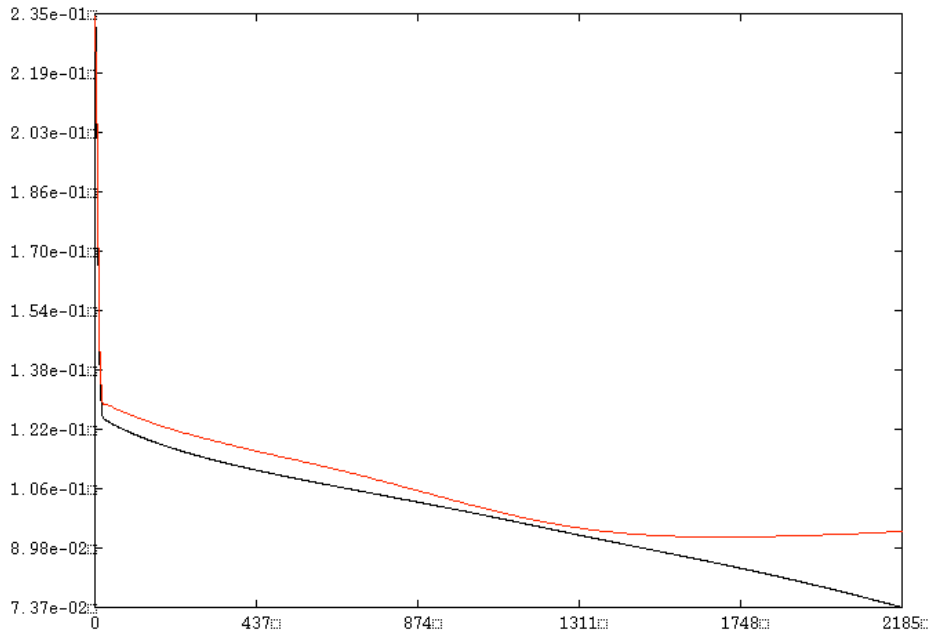


Fig. 12. Gráfica que representa la evolución del entrenamiento para este experimento. La línea roja denota los resultados en las pruebas y la negra en el entrenamiento.

6.1.8 Imágenes normales, Histograma simple, 2 Neuronas ocultas.

Como se ha comentado anteriormente, otra forma de pasar información de las imágenes pero sin meter dichas imágenes como entradas directas, es procesar y disponer como entrada el histograma que estas formas. Los resultados son bastante peores, un 49% de acierto en la clasificación y realmente al ver la gráfica del simulador se comprende, pues los valores mínimos están en la cota 0,125.

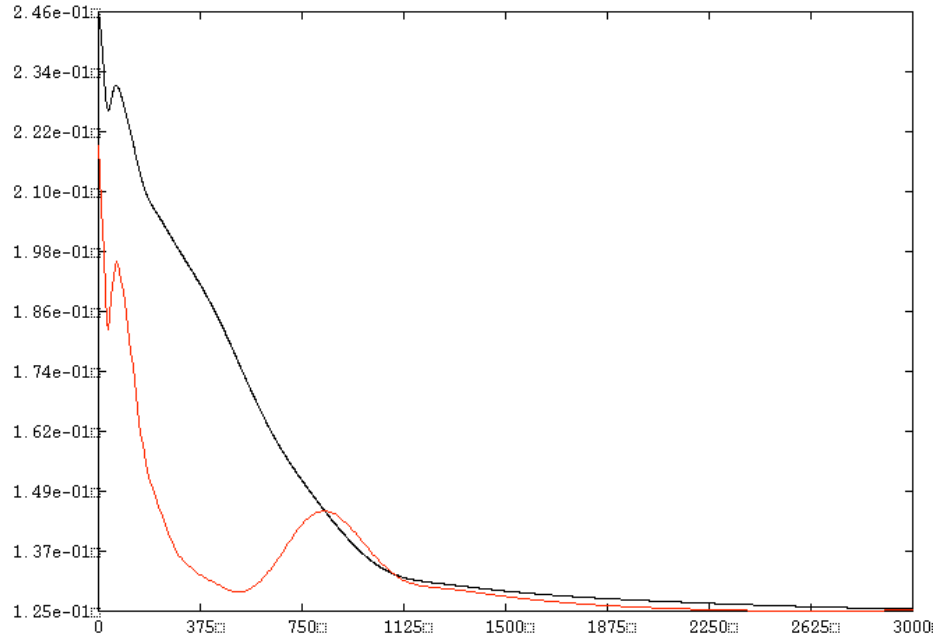


Fig. 13. Gráfica que representa la evolución del entrenamiento para este experimento. La línea roja denota los resultados en las pruebas y la negra en el entrenamiento.

Tabla 15. Tabla que muestra cómo se clasifican los elementos de cada categoría. En cada fila se establece, para cada categoría, el número de elementos que el sistema ha puesto en cada categoría.

	No Porno	Porno
No porno	181	127
Porno	196	134

Tabla 16. Recall y precisión por cada una de las categorías.

	Recall	Precisión
No porno	0,587662	0,480106

Porno	0,406060	0,513409
--------------	----------	----------

6.2 Experimentos con Perceptrón Multicapa con Retropropagación

Para las Redes formadas por perceptrones multicapa con retropropagación para efectuar el descenso de grandiete del error, hay que definir más parámetros que para las redes de base radial. En estos experimentos solo se han utilizado topologías en 3 capas, por lo que se han reducido los parámetros a los mismos que en redes de base radial.

6.2.1 Imágenes a 30x30 con todos los valores de color y 2 Neuronas ocultas.

Este experimento permite comparar los dos tipos de redes de forma directa pues los datos son todos exactamente los mismos. El resultado es que las redes de base radial funcionan mejor para estos datos pues el porcentaje de corrección en la clasificación en las redes de perceptrones con retropropagación es de 79,9709% de aciertos.

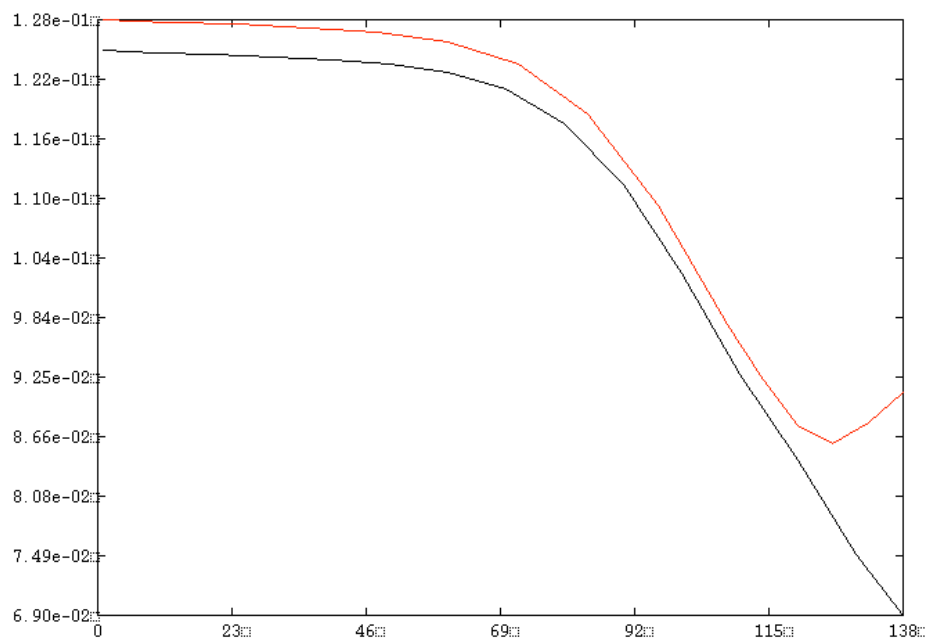


Fig. 14. Gráfica que representa la evolución del entrenamiento para este experimento. La línea roja denota los resultados en las pruebas y la negra en el entrenamiento. Cada ciclo corresponde a diez reales.

Tabla 17. Tabla que muestra cómo se clasifican los elementos de cada categoría. En cada fila se establece, para cada categoría, el número de elementos que el sistema ha puesto en cada categoría.

	No Porno	Porno
No porno	268	40
Porno	98	283

Tabla 18. Recall y precisión por cada una de las categorías.

	Recall	Precisión
No porno	0,870129	0,732240
Porno	0,742782	0,876160

6.2.2 Imágenes a 90x90 con todos los valores de color y 2 Neuronas ocultas.

Este experimento no se realizó con anterioridad para las redes de base radial, los resultados son parecidos a los experimentos con retropropagación a 30x30 aunque el tiempo de cómputo resultó mucho mayor. .

Tabla 19. Tabla que muestra cómo se clasifican los elementos de cada categoría. En cada fila se establece, para cada categoría, el número de elementos que el sistema ha puesto en cada categoría.

	No Porno	Porno
No porno	253	55
Porno	99	284

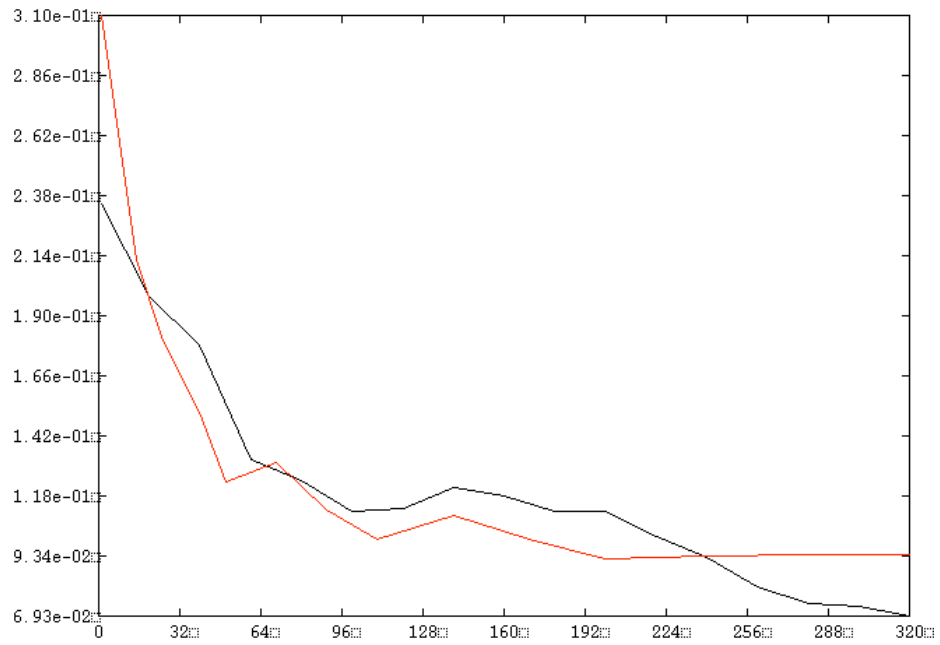


Fig. 15. Gráfica que representa la evolución del entrenamiento para este experimento. La línea roja denota los resultados en las pruebas y la negra en el entrenamiento.

Tabla 20. Recall y precisión por cada una de las categorías.

	Recall	Precisión
No porno	0,821428	0,71875
Porno	0,741514	0,837758

6.3 Relación de Todos los Experimentos

En la siguiente tabla se muestran los experimentos concluidos con sus respectivos resultados. En cuanto a los tipos de entrada, RGB 30 significa que las imágenes están escaladas a 30x30 y se tiene una entrada por cada color, RGB 90 similar pero con imágenes a 90x90, S5 30 indica imágenes a 30x30 pero con segmentación 5, equivalente para S3 30, S2 30, S5 90, S3 90, S2 90. H T indica que la entrada es el histograma completo de la imagen, H 10 M es que se introducen los 10 mejores valores y H 10 MS es que se introducen los 10 mejores segmentados.

#	Tipo Red	Entradas	Ocultas	Salida	%	Tipo Entrada
1	RBR	2700	2	1	82%	RGB 30
2	RBR	2700	10	1	84%	RGB 30
3	RBR	2700	20	1	82%	RGB 30
4	RBR	24300	2	1	N/C	RGB 90
5	RBR	900	2	1	73%	S5 30
6	RBR	900	10	1	73%	S5 30
7	RBR	900	2	1	67%	S3 30
8	RBR	900	2	1	64%	S2 30
9	RBR	8100	2	1	73%	S5 90
10	RBR	8100	4	1	73%	S5 90
11	RBR	8100	12	1	69%	S5 90
12	RBR	8100	2	1	68%	S5 90
13	RBR	8100	2	1	64%	S5 90
14	RBR	256	2	1	49%	H T
15	RBR	10	2	1	54%	H 10M
16	RBR	20	2	1	54%	H 10MS
17	PMR	2700	2	1	80%	RGB 30
18	PMR	24300	2	1	78%	RGB 90
19	PMR	900	2	1	71%	S5 30
20	PMR	8100	2	1	68%	S5 90
21	PMR	256	2	1	33%	H T

7 Conclusiones

Como se ha visto son grandes los esfuerzos que se han hecho en este campo de investigación y que seguro se seguirán haciendo. Existen grandes intereses en estos métodos de filtrado de material pornográfico. Empresas y centros educativos de toda índole son los demandantes más interesados en el desarrollo efectivo de este tipo de filtros. Este documento ha mostrado una primera aproximación al problema con resultados bastante alentadores.

Nuestro principal objetivo a la hora de estudiar el resultado de cada experimento ha sido el comprobar la relación de acierto/error en cada categoría (porno/no porno). Un sistema de filtrado de imágenes porno no puede conformarse con eliminar todas las fotos porno si con ello elimina un alto porcentaje de fotos no pornográficas. La calidad de filtrado reside, entonces, en la capacidad de discernir con exactitud que fotos son no porno y cuales sí lo son. Con un error aceptable y por otra parte, ciertamente esperable, se conseguirá eliminar un alto porcentaje de imágenes pornográfica con un coste en imágenes no pornográficas muy pequeño.

Contrariamente a lo pensado al inicio de esta investigación, las redes neuronales son capaces de aprender mejor (al menos sobre los datos que aquí se han ofrecido como elementos a clasificar, las fotografías) a partir de toda la información de la imagen, ya que al intentar minimizar la cantidad de datos a ofrecer a la red, se consigue que se pierda información valiosa y, por tanto, la red aprende peor. También ha resultado interesante comprobar la mejor clasificación ofrecida por las redes de base radial pues, a priori, se confiaba en un mejor funcionamiento de las redes de neuronas con retropropagación al tender las redes a usar demasiadas neuronas.

Tampoco conviene perder de vista que los acercamientos que reducen la información entregada a la red hacen que se reduzca el tiempo de aprendizaje y procesado, por lo que algunas de las soluciones con menor número de datos de entrada puede ofrecer resultados interesantes a un relativo bajo coste de ejecución.

No deja de ser interesante las connotaciones que esto tiene, la red neuronal puede aprender a partir de todos los datos de la entrada, por lo que realmente generaliza muy bien, abstrayendo la información a partir de datos tan realmente poco informativos a priori como los componentes de color de los píxeles. A una persona humana le resultaría imposible clasificar una fotografía a partir de esta información pues se encontraría con una matriz de píxeles, representado cada uno por tres valores numéricos comprendidos entre 0 y 255, muy poco descriptiva. Además hay que tener en cuenta que la entrada no está preparada para que las partes pornográficas se ubiquen en alguna zona en concreto, por lo cuál la propia red ha sabido distinguir en que partes está presente y en cuáles no, qué se puede considerar como pornográfico y qué no, etc. Todas estas cuestiones hacen ver cuán potente son las redes de neuronas en cuanto a generalización, a pesar que su salida no es nada explicativa pues no es capaz de “explicar” el por qué de la decisión de clasificar a una fotografía como pornográfica o como no pornográfica.

8 Líneas de Investigación

El acercamiento a la clasificación de imágenes pornográficas mediante redes de neuronas utilizado en este trabajo pretende sentar una base para otros acercamientos más sólidos. El ofrecerle a la red la fotografía como entrada (sea como sea la forma en la que se ofrezca), es el primer aspecto a probar pero subsiguientes acercamientos deberían tratar de procesar más las imágenes y ofrecer esos procesados como entradas como vectores de características.

Para generar estos vectores de características se pueden utilizar elementos típicos del tratamiento de imágenes tales como histogramas (ya utilizados como entrada

directa en este acercamiento), transformadas de Fourier, trozos de piel de la foto, etc. Con todos estos datos se pueden producir entradas a la red muy completas que permitan una mejor clasificación.

Los experimentos aquí desarrollados muestran la dificultad en la que se encuentra el investigador a la hora de escoger los distintos parámetros de una red. Realizar profundas investigaciones con el objetivo de alcanzar conclusiones útiles a cerca de qué valores utilizar para cada parámetro o establecer unos rangos demostrados de mayor efectividad para cada uno de estos valores supondría un avance enorme en el camino de encontrar un proceso de filtrado cada vez más eficiente.

9 Referencias

1. [arias] Arias, J. D.: Simulador de Redes de Base Radial.
2. [bishop96] Bishop, C. M., Bishop, C.: "Neural Networks for Pattern Recognition" Oxford University Press. 1996.
3. [broomhead88] Broomhead D. S., Lowe D.: Multivariate functional interpolation and adaptive networks. *Complex Systems*, 2, pp. 321-355.
4. [fausett94] Fausett L.: *Fundamentals of Neural Networks*. Prentice-Hall, 1994.
5. [fleck96] Margaret Fleck, David Forsyth, and Chris Bregler: Finding Naked People, 1996 European Conference on Computer Vision , Volume II, pp. 592-602.
6. [forsyth96] David A. Forsyth and Margaret M. Fleck: Identifying nude pictures, *IEEE Workshop on the Applications of Computer Vision 1996*, pp. 103-108.
7. [gomez02] Gómez Hidalgo, J.M., de Buenaga Rodríguez, M., Puertas Sanz, E., Carrero García, F.: Text Filtering at POESIA: a new Internet content filtering tool for educational environments. *Procesamiento del Lenguaje Natural*, 29, pp. 291-2, 2002.
8. [gomez03] Gómez Hidalgo, J.M., Giráldez Bretón, I., Buenaga Rodríguez, M. de.: Text Categorization for Internet Content Filtering (Categorización para los filtros de contenido en Internet). *Inteligencia Artificial - Revista Iberoamericana de Inteligencia Artificial*, Num. 22, Vol. 5, pp. 147-160, 2003
9. [gurney97] Gurney K.: *An Introduction to Neural Networks*. UCL Press, 1997.
- 10.[haykin99] Haykin S., *Neural Networks* , 2nd Edition, Prentice Hall, 1999.

- 11.[lee88] Lee, S., Kil, R.: Multilayer feedforward potencial function Networks. Proceedings of the IEEE Second International Conference on Neural Networks 1988. pp 161-171.
- 12.[mitchell97] Mitchell, T. M.: Machine Learning. McGraw-Hill International Editions (1997)
- 13.[pankiewicz97] Pankiewicz, G.S., 1997, Neural network classification of convective airmasses for a flood forecasting system, International Journal of Remote Sensing, (special issue on the use of neural networks in remote sensing).
- 14.[patterson96] Patterson, D. Artificial Neural Networks . 1996.