



Competición CAEPIA-App: Awesome Pic

Fernando Rubio

SIMD Lab, I3A

UCLM

Albacete, España

Email: fernando.rubio@uclm.es

M. Julia Flores

Departamento de Sistemas Informáticos

UCLM

España

Email: julia.flores@uclm.es

Jose M. Puerta

Departamento de Sistemas Informáticos

UCLM

España

Email: jose.puerta@uclm.es

I. INTRODUCCIÓN

Dentro de la visión artificial existen multitud de problemas relacionados con robótica, salud, seguridad, etc. En los últimos años se ha visto como el paradigma Deep Learning ha revolucionado estos ámbitos, pero de la misma forma otros nuevos problemas más complejos han empezado a surgir a partir de estas soluciones, problemas como el de la evaluación de la calidad estética en fotografía, donde se trata de medir o cuantificar lo atractiva o “bonita” que es una imagen.

I-A. Evaluación automática de la calidad estética

La evaluación automática de la calidad estética en fotografía es uno de los problemas más desafiantes de la visión artificial dado su carácter altamente subjetivo, ya que dos personas pueden tener opiniones muy distintas sobre la calidad de una imagen e incluso entre expertos puede haber diferencias. Sin embargo, esta dificultad no ha impedido que en los últimos años hayan aparecido gran cantidad de estudios tratando de dar una solución a esta tarea. Este interés se debe principalmente a las posibles aplicaciones, ya que cada día generamos una gran cantidad de imágenes debido, sobre todo, al “boom” de las redes sociales donde una posible ordenación automática de las fotos en base a su calidad es de gran utilidad. Pero también se podría utilizar para hacer recomendaciones de filtros para mejorar el aspecto de nuestras fotografías. Los desarrolladores de teléfonos tratan de ofrecer cada vez mejores cámaras y aplicaciones para sus dispositivos donde estas recomendaciones tienen cabida. Además, existen múltiples opciones en las galerías de las imágenes, desde el agrupamiento de fotografías, hasta la realización automática de videos temáticos a partir de secuencias de fotografías, donde sería muy interesante ordenar por estética las seleccionadas. Otro de los campos de explotación es la publicidad, donde el diseño de imágenes atrayentes para la gente es clave a la hora de hacer anuncios.

Como ya se ha comentado, se trata de un problema con una alta subjetividad, por lo que es difícil medir como de atractiva es una imagen. En la literatura la forma de abordar el problema viene definido en gran parte por las bases de datos disponibles. Generalmente constan de un conjunto de imágenes evaluadas por múltiples personas en base a un determinado ranking. La base de datos más utilizada actualmente es AVA [1] que contiene más de 250.000 imágenes, todas ellas con múltiples votos del 0 al 10 de distintas personas.

I-B. Problema de clasificación binaria

Gracias a estas bases de datos, es posible convertir la evaluación automática de la calidad estética en un problema de *machine learning*, más concretamente en un problema de aprendizaje supervisado o semi-supervisado, donde tenemos una entrada, que es la imagen, y la salida, que son los votos.

Sin embargo, esta salida es difícil de manejar por muchos algoritmos, además de que esos votos no es la información que espera el usuario final, ya que realmente lo que se busca es saber si una foto es buena o mala en base, quizás a una graduación. Es por ese motivo, que en la mayoría de la literatura las imágenes de las bases de datos son etiquetadas en “snapshots” o “professional shots” en base a estadísticos sobre los votos de las imágenes, como es la media y la moda. Siguiendo con la base de datos AVA, lo más frecuente es que una fotografía se etiquete como “professional shot” cuando la media de sus votos supere el 5, y como “snapshot” en caso contrario.

I-C. Técnicas de IA utilizadas

En este trabajo hemos seguido el mismo procedimiento, donde a partir de una imagen obtenemos una etiqueta de “snapshot” o “professional shot” y una probabilidad asociada a la clase. En [2] se resume que las técnicas de Transfer Learning sobre redes Deep Learning son las que mejores resultados han obtenido en el problema de evaluación de la calidad estética con el enfoque de clasificación binaria. El Transfer Learning consiste en utilizar una red entrenada para un problema diferente, pero relacionado. Cuando trabajamos con imágenes, el problema por excelencia es ImageNet, donde el objetivo es etiquetar las imágenes con objetos que aparecen en ellas, de entre 1.000 posibles “tags”. De este problema se han obtenido algunas de las arquitecturas de redes neuronales más importantes para el procesamiento de imágenes.

Para este problema, hemos utilizado la arquitectura de AlexNet [3] y los pesos para el problema de ImageNet, donde hemos aplicado *finetuning*, una técnica de Transfer Learning. El proceso consiste en sustituir las últimas capas de la red para ajustarla a nuestro problema, en este caso, hemos sustituido la capa de salida de 1.000 etiquetas de objetos por nuestra salida binaria (“snapshot” o “professional shot”) y por último hemos reentrenado con AVA sólo esta capa final.

Pero en este trabajo no sólo buscamos la mejor solución, si no una que se pueda aplicar a dispositivos móviles y ejecutar



Figura 1: Ejemplo de la clasificación *online* donde se puede observar el “emoji” en la esquina superior izquierda para la clase “professional shot”.



Figura 2: Ejemplo de la clasificación *offline* donde se puede observar en la parte inferior la puntuación obtenida por nuestra imagen en la red neuronal de nuestro servidor.

en tiempo real. En [4] se lleva a cabo un estudio del comportamiento de algunas de las técnicas y algoritmos más populares en los problemas de visión artificial aplicados al problema de la calidad estética. En estos casos, la estructura de las propuestas se basa en aplicar un filtro a la imagen para extraer características significativas y reducir la dimensionalidad de la entrada, para después utilizarlo en un clasificador como son algunos modelos probabilísticos, además de otro tipo de modelos.

De entre ellos hay algunas soluciones capaces de ejecutarse con unos requisitos de memoria y tiempo mínimos, como por ejemplo el histograma de grises, una de las técnicas más utilizadas en el procesamiento de imágenes. Mientras que en el lado de los clasificadores nos encontramos con que Naïve Bayes es el que mejores tiempo de respuesta nos ofrece, además de que sus resultados son aceptables dentro del estado del arte. Este enfoque se ha utilizado para esta primera versión de la app, pero una de las líneas de investigación que llevamos a cabo es el mejorar este procesamiento en tiempo real en base a redes Deep Learning mucho más eficientes y menos complejas que aún mantengan unos resultados significativos en sus predicciones.

II. DESCRIPCIÓN DE LA APP

Awesome Pic¹ es una aplicación de fotografía capaz de medir la calidad estética de las imágenes tomadas por la cámara. La app consta de una cámara, una pantalla de previsualización de la fotografía, una galería de imágenes y un menú de configuración. A continuación vamos a describir en más detalle cada una de estas pantallas:

- La cámara es similar a la que viene por defecto en los dispositivos móviles y que permite tomar capturas o imágenes digitales. Dentro de la cámara contamos con las opciones de tomar la imagen, navegar a la galería o salir. En la cámara también aparece información sobre la calidad estética en tiempo real u *online* de lo que se previsualiza y que se obtiene directamente por la app. Esto se puede observar en la Figura 1.

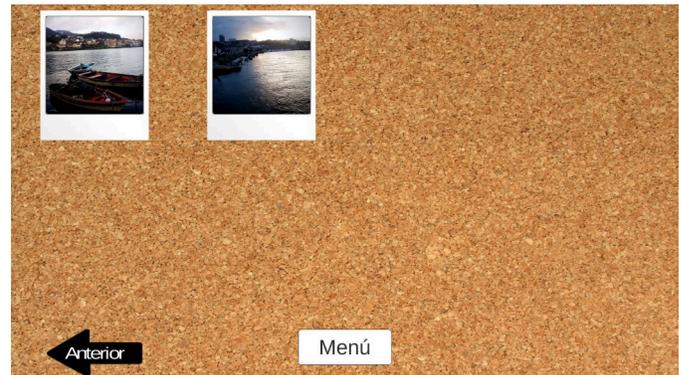


Figura 3: Galería de Awesome Pic.

- Una vez la foto es tomada, pasamos a la pantalla de previsualización como en la Figura 2, donde aparece la imagen capturada por la cámara en el momento que se pulsa el botón. En esta pantalla disponemos de opciones para almacenar la imagen en la galería, descartarla o hacer una evaluación más precisa *offline* de la calidad estética mediante un servicio externo.
- En la galería podemos visualizar las imágenes almacenadas, como se muestra en la Figura 3.
- En el menú de configuración podemos cambiar el idioma de la app, el tamaño de la fuente y aplicar un filtro de alto contraste.

II-A. Evaluación online

En la primera evaluación de la calidad estética, la clasificación se lleva a cabo en tiempo real y en un dispositivo móvil. Esto supone unas restricciones de tiempo y memoria muy altas para realizar el tratamiento de la imagen y su procesamiento por el modelo aprendido.

Como ya hemos comentado en las técnicas utilizadas, hemos seleccionado y aprendido un modelo Naïve Bayes gaussiano con la base de datos AVA convertida a histograma de grises de tamaño 255. El motivo de elegir esta combinación es lo rápido que son ambos procesos y que esto permite ejecutar la evaluación completa en tiempo real y sin consumir grandes recursos del dispositivo móvil.

¹<https://www.youtube.com/watch?v=uanPRdW7iTU>



La salida del modelo es la clasificación que puede ser una “snapshot” o “professional shot” y, dependiendo de esto, mostramos en la esquina superior izquierda un “emoji” como el que se muestra en la Figura 1 que, con el color y la expresión de la cara, nos indica el resultado de la clasificación:

- Rojo, cara de enfado: fotografía de mala calidad o “snapshot”.
- Verde, cara alegre: fotografía de calidad o “professional shot”.

Esta información se actualiza en cada frame mientras observamos con la cámara, por lo que los “emojis” varían hasta que se toma la foto.

II-B. Evaluación offline

La segunda clasificación se realiza mediante una consulta a un servicio externo implementado por nosotros para la aplicación. Una vez la imagen es capturada es posible enviarla a dicho servicio mediante un botón de lupa que aparece en la esquina superior derecha. En este caso no tenemos las restricciones de tiempo y recursos que en el caso anterior, ya que en el servidor contamos con una tarjeta gráfica dedicada al cómputo y de cierto tiempo para recibir la respuesta. Obviamente no podemos esperar horas a que se obtenga el resultado, pero si unos cuantos segundos.

Este servicio realiza una clasificación binaria con algoritmos más precisos, en este caso, el modelo AlexNet adaptado a la evaluación de la calidad estética con Transfer Learning. De la clasificación se obtiene también una probabilidad o certeza con la que se ha realizado dicha clasificación. Si nos quedamos con la probabilidad de la clase “professional shot” obtendremos un valor donde el 0 representa que se trata de una “snapshot” con un 100 % de certeza y que tiene una calidad pésima, mientras que una puntuación cercana a 100 representa una imagen de gran calidad. En resumen, cuanto más alto sea este valor, más probable será de que se trate de una “professional shot” y de que esa imagen sea atrayente a una mayor cantidad de gente.

Cuando la imagen ha sido procesada por el servidor, este devuelve la probabilidad antes mencionada que es mostrada en un recuadro en la parte inferior como vemos en la Figura 2, indicando así su calidad. Una vez hecho esto podremos elegir entre almacenar la foto en la galería o descartarla.

II-C. Tecnologías y herramientas

Para el desarrollo de la aplicación móvil se ha utilizado Unity3D [5], uno de los motores de juego más populares y más versátiles. El uso de esta herramienta permite crear apps multiplataforma, lo que incluye Android e iOS, y dota a las apps de un aspecto más atractivo y original. El lenguaje de programación utilizado es C# donde se implementan las funciones de procesamiento de imágenes y algunos de los algoritmos de aprendizaje automático para la evaluación de la calidad estética en tiempo real.

Además de la app se ha creado un servicio en Python para hacer consultas más precisas sobre la calidad y que requieren de un cómputo mucho mayor. Las llamadas a dicho servicio se gestionan a través de Flask [6] y la librería Caffé [7]. Por

último, el servicio se ejecuta a través de un contenedor de Docker [8].

III. UTILIDAD Y VIABILIDAD DE LA APP

Se trata de una primera versión de la app donde se aprecian dos tipos de recomendaciones sobre la calidad de las imágenes, una en tiempo real que toma dos valores y otra más precisa una vez la imagen ha sido tomada, donde se evalúa del 0 al 100.

Es indudable que cualquier aficionado a la fotografía puede beneficiarse de la aplicación y cualquier persona puede ver mejorada la calidad de sus imágenes, ya sea para subirlas a redes como Instagram o Facebook o para un uso más personal. Además, la puntuación obtenida por el servicio puede utilizarse para la ordenación de fotografías o la comprobación de que filtros benefician o perjudican a una determinada imagen. Se trata de una aplicación única, ya que no hemos sido capaces de encontrar otra que sea capaz de valorar la calidad estética de las imágenes tomadas.

La aplicación se encuentra disponible para todo el mundo de forma gratuita a través de la Play Store de Android² y no requiere de paquetes, ni de ningún tipo de aplicación adicional. En este sentido hace que su descarga y manejo sean muy sencillos, además de que dispone de varios idiomas. La mayor limitación viene del lado del servicio, ya que se encuentra desplegado en una máquina que no está preparada para un volumen de peticiones alto o una escalabilidad grande de usuarios.

IV. CONCLUSIONES

Se trata de la primera versión de la app **Awesome Pic** donde utilizando algoritmos de aprendizaje automático hemos sido capaces de resolver el problema de la evaluación de la calidad estética en fotografía y de plasmar dichos resultados en una cámara capaz de mostrar información sobre la calidad de las imágenes tanto *online* como *offline*.

El objetivo de esta aplicación es plasmar los resultados obtenidos en nuestro trabajo de investigación en este campo de la visión artificial. Nuestra intención es que conforme mejoren dichos algoritmos y técnicas, poder incluirlos en la app y mejorar su comportamiento, así como incluir nuevas funcionalidades, como puede ser la recomendación de filtros o un sistema de puntuación más precisa en tiempo real. Otra de las mejoras consiste en desplegar el servicio en la nube, más concretamente en AWS, mejorando el tiempo de respuesta, la capacidad de recibir más peticiones y de aumentar la escalabilidad o que incluso se pueda utilizar dicho servicio fuera de la aplicación con otros fines.

AGRADECIMIENTOS

Este trabajo ha sido parcialmente financiado con los fondos FEDER y el Gobierno Español (MICINN) a través del proyecto TIN2016-77902-C3-1-P.

²<https://play.google.com/store/apps/details?id=com.Frubio.AwesomePic>

REFERENCIAS

- [1] N. Murray, L. Marchesotti, and F. Perronnin, “Ava: A large-scale database for aesthetic visual analysis,” in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2012, pp. 2408–2415.
- [2] Y. Deng, C. C. Loy, and X. Tang, “Image aesthetic assessment: An experimental survey,” *IEEE Signal Processing Magazine*, vol. 34, no. 4, pp. 80–106, 2017.
- [3] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in *Advances in neural information processing systems*, 2012, pp. 1097–1105.
- [4] F. Rubio, M. J. Flores, and J. M. Puerta, “Drawing a baseline in aesthetic quality assessment,” in *Tenth International Conference on Machine Vision (ICMV 2017)*, vol. 10696. International Society for Optics and Photonics, 2018, p. 106961M.
- [5] U. G. Engine, “Unity game engine-official site. <http://unity3d.com> [accessed: September 6, 2018.]”
- [6] F. A Python Microframework, “<http://flask.pocoo.org> [accessed: September 6, 2018.]”
- [7] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, “Caffe: Convolutional architecture for fast feature embedding,” in *Proceedings of the 22nd ACM international conference on Multimedia*. ACM, 2014, pp. 675–678.
- [8] Docker, “Docker-official site. <https://www.docker.com> [accessed: September 6, 2018.]”