Variantes de Vectores de Fisher para la clasificación de imágenes de lesiones de piel mediante redes neuronales profundas residuales

Cristian Luciano Salto[1], Daniel Acevedo[2,3]

[1] Maestría en Explotación de Datos y Descubrimiento del Conocimiento, Facultad de Cs. Exactas y Naturales, Universidad de Buenos Aires
[2] Universidad de Buenos Aires, Fac. de Cs. Exactas y Naturales, Depto. de Computación. Buenos Aires, Argentina
[3] CONICET-UBA. Instituto de Investigación en Cs. de la Computación (ICC). Buenos Aires, Argentina dacevedo@dc.uba.ar

Resumen El presente trabajo investiga alternativas al problema de clasificación de imágenes dermatoscópicas utilizando redes neuronales residuales (ResNet) y codificando los descriptores de la misma con Vectores de Fisher. En primer lugar se realizó el reentrenamiento de un clasificador ResNet-50. Luego se aplicaron Vectores de Fisher a partir de los descriptores de distintas muestras de una imagen. Otra alternativa investigada fue generar vectores de Fisher sobre la base de los descriptores obtenidos como salida del quinto bloque convolucional de la red ResNet-50. Finalmente se realizó un ensamble de las aplicaciones de vectores de Fisher logrando resultados acorde a lo desarrollado en otros trabajos.

Keywords: ResNet · Vectores de Fisher · Imágenes dermatológicas

1. Introducción

El melanoma es un cáncer que surge de los melanocitos, células pigmentadas especializadas que se encuentran predominantemente en la piel [1]. La detección temprana es fundamental para disminuir la mortalidad, en este sentido, se están desarrollando muchas aplicaciones utilizando imágenes dermatoscópicas para el diagnóstico de melanomas haciendo que el proceso de evaluación sea cada vez más objetivo, más preciso y universalmente disponible [6]. A continuación se presentan una serie de aplicaciones relacionadas al diagnóstico de melanomas que utilizan redes neuronales profundas para extraer descriptores de las imágenes y codificarlos como vectores de Fisher (FV). Estos métodos varían en función de las redes utilizadas, la capa de la cual extraen los descriptores y el procesamiento de las imágenes. En [9] integran una red residual totalmente convolucional (FCRN) y otras redes residuales profundas para clasificación. En [11], [13] y [10] presentan marcos de clasificación híbridos para la evaluación de imágenes dermatoscópicas mediante la combinación de una red neuronal convolucional profunda (CNN), FV y una máquina de vectores de soporte (SVM, por sus siglas en inglés). Finalmente, para abordar el diagnóstico de melanomas mediante

imágenes, en [12] proponen un marco para el reconocimiento automático de lesiones cutáneas mediante la codificación basada en redes cruzadas de múltiples redes convolucionales.

En el presente trabajo se exponen tres alternativas para el problema de clasificación de melanomas. El primer enfoque involucra al entrenamiento de una red convolucional ResNet-50. Posteriormente se aplican FV a dos métodos para extraer los descriptores a partir de una imagen. El resto de este artículo está organizado de la siguiente manera. La Sección 2 describe los detalles de la metodología de los modelos propuestos. La Sección 3 presenta los resultados experimentales y finalmente en la Sección 4 las conclusiones del trabajo.

2. Metodología

En esta sección mencionamos la arquitectura de la red convolucional empleada y el algoritmo utilizado para calcular los FV. Seguidamente se presentan la aplicación de ambos conceptos a los experimentos realizados.

2.1. Redes Neuronales Residuales y Vectores de Fisher

En este trabajo, adoptamos la arquitectura de la red neuronal residual (ResNet-50) introducida en [3], que ocupa el puesto número uno en el desafío de reconocimiento visual a gran escala de ImageNet 2016 [7] para la extracción de características. ResNet adopta el aprendizaje residual a algunas capas apiladas.

El cálculo de los vectores de Fisher se realizó de acuerdo al algoritmo planteado en [8]. Dicho algoritmo tomó como valores de entrada los descriptores de una imagen generados por una red convolucional y los parámetros del modelo de mixtura de gaussiana (GMM) para generar como salida los Vectores de Fisher. A partir de los parámetros de entrada computa los estadísticos de orden 0, 1 y 2 para luego calcular los FV y finalmente normalizarlos mediante la normalización de potencia y la normalización L_2. Estas normalizaciones se presentan en [5], donde demostraron ser necesarias para obtener resultados competitivos cuando el FV se combina con un clasificador lineal.

2.2. ResNet-50 Fine tuning

A partir de la red entrenada con ImageNet, ver [3], se realizó transferencia de aprendizaje. Para adaptar este modelo a la aplicación actual fue necesario reemplazar la última capa totalmente conectada por una capa de una única neurona de salida con activación sigmoide. Luego la red fue entrenada por etapas descongelando los pesos de las neuronas desde la capa superior hasta las capas menores. Así, en primer lugar, durante 100 épocas con optimizador Adam, batch size 32 y learning rate 0.0001, la capa totalmente conectada fue entrenada. Posteriormente se descongelaron los pesos de las neuronas del quinto y cuarto bloque convolucional, para estos casos también se realizó un entrenamiento durante 100 épocas

con optimizador SGD, batch size 32 y learning rate 0.0001. En cuanto al preprocesamiento de imágenes se realizaron rotaciones en 90, 180 y 270 grados, reflejos y traslaciones sobre ambos ejes, escalado, normalización y estandarización por imagen. El tamaño de entrada de las imágenes fue de 224×224.

2.3. Fisher con muestras

Para obtener las representaciones de las imágenes a través de una red CNN y procesar las mismas con FV se siguió una metodología similar al planteado en [11]. En nuestro esquema, representamos una imagen de lesión cutánea como múltiples subimágenes y extraemos características profundas para estas subimágenes. Específicamente, adoptamos la red previamente entrenada en la Sección 2.2. Luego reemplazamos las capas de agrupación promedio y activación sigmoide por una capa plana. Cada una de las subimágenes, de tamaño 224×224, es procesada por la red y se obtienen las representaciones generadas por la capa plana. Por lo tanto, obtenemos un conjunto de N vectores de características de dimensión 2048 para cada imagen, siendo N la cantidad de subimágenes procesadas. Para un cálculo más eficiente, los vectores de características son escalados y su dimensionalidad es reducida mediante PCA. Luego se entrena un modelo de mixtura de gaussianas (GMM) con K componentes. Seguidamente a partir de los descriptores de una imagen y las componentes de GMM se obtienen los FV aplicando el algoritmo planteado en [8]. Tanto la cantidad de componentes principales como la cantidad de componentes gaussianas son hiperparámetros del modelo. Para crear estos modelos se trabajó con 32 muestras a partir de cada imagen, mientras que para generar los vectores de Fisher se utilizaron 64. Finalmente, se entrena un clasificador de máquinas de vectores de soporte (SVM) con validación cruzada de 5 iteraciones para obtener los mejores parámetros.

2.4. Fisher con descriptores

El proceso es similar al planteado en [10]. En primer lugar, eliminamos las capas de agrupación promedio y activación sigmoide, de la red entrenada en la Sección 2.2, para trabajar con los descriptores generados por el quinto bloque convolucional. Dada una imagen X_i se realizan rotaciones en 90°, 180° y 360° obteniendo cuatro muestras $\mathbb{X}_i = \{X_{i1}, X_{i2}, X_{i3}, X_{i4}\}$ de igual tamaño, en este caso 224×224. Cada una de estas muestras es procesada por la red obteniendo como salida del quinto bloque convolucional un vector de dimensión $(7 \times 7 \times 2048)$ para cada muestra. Para continuar con el proceso se decidió trabajar con vectores de dos dimensiones (49×2048), de esta manera, a partir de la imagen original se obtienen cuatro vectores $\mathbb{F}_i = \{F_{i1}, F_{i2}, F_{i3}, F_{i4}\} \in \mathbb{R}^{(49 \times 2048)}$. Con los vectores de todas las imágenes se entrena un modelo de mixtura de gaussianas, previo escalamiento y reducción de la dimensionalidad con PCA. Una vez entrenado un GMM, a partir del mismo se obtienen los FV de cada imagen con los descriptores siguiendo el algoritmo de [8]. Finalmente los parámetros del clasificador SVM fueron elegidos utilizando validación cruzada con 5 iteraciones.

3. Resultados

En las tres aplicaciones se utilizaron los datos de la competencia ISIC 2016 [2]. Este conjunto de datos se compone de 900 imágenes de entrenamiento y 379 imágenes de prueba. El primer conjunto se divide en 727 imágenes de lesiones benignas y 179 melanomas. En tanto que el conjunto de prueba se conforma a partir de 304 lesiones benignas y 75 melanomas. El Cuadro 1 contiene los resultados de los métodos propuestos, un ensamble entre las aplicaciones de FV (FV-Ensamble) y otros trabajos, que fueron detallados brevemente en la introducción, utilizando el mismo conjunto de datos de prueba [2]. Para comparar los distintos métodos se utilizaron las métricas: Precisión media (mAP), Accuracy (Acc) y Área bajo la curva ROC (AUC). De las tres alternativas planteadas anteriormente, los Vectores de Fisher procesados a partir de los descriptores del quinto bloque convolucional (ResNet-50-FV, ver Sección 2.4) son los que mejor rendimiento tienen analizando las tres métricas. Si bien el ensamble propuesto es superior en las métricas mAP y AUC a los modelos originales y al mejor clasificador de la competencia ISBI 2016 [9], existen otras aplicaciones desarrolladas que lograron mejores rendimientos a ambos trabajos [10,13,12].

Cuadro 1. Métodos propuestos y comparación con otras aplicaciones.

Método	Red	mAP	Acc	AUC
ResNet-50-FT	ResNet-50	58.52	82.58	79.62
ResNet-50-FV	ResNet-50	54.62	84.16	75.58
ResNet-50-FV-CM	ResNet-50	61.34	84.69	80.93
FV-Ensamble	ResNet-50	64.46	83.90	82.95
DCNN-FV [10]	ResNet-50	68.49	**86.81**	85.20
DCNN-FV [11]	AlexNet-FC6	59.50	83.09	79.57
LDF-FV [13]	ResNet-50	68.49	**86.81**	85.20
CUMED [9]		63.70	85.50	80.40
SLD [14]		68.10	86.30	82.20
M-CNN-FV [12]		**68.73**	**86.81**	**85.82**
VGG-16-FV [4]	Ensamble	68.13	80.29	82.11

4. Conclusiones

Concluimos que el rendimiento de los tres métodos propuestos se encuentran en el orden de magnitud de clasificación de las aplicaciones desarrolladas. Además el método de Fisher con los descriptores del quinto bloque convolucional incrementa sensiblemente la capacidad de predicción. El ensamble realizado entre las aplicaciones de FV logra mejorar las métricas de mAP y AUC pero no supera a otras aplicaciones realizadas para la clasificacion de imágenes dermatoscópicas. Los futuros trabajos pueden evaluar el ensamble desarrollado en otros conjuntos de datos y experimentar obtener los descriptores de otros bloques convolucionales.

Referencias

1. Gray-Schopfer, V., Wellbrock, C., Marais, R.: Melanoma biology and new target therapy. Nature **445**, 851–7 (03 2007)
2. Gutman, D., Codella, N.C.F., Celebi, E., Helba, B., Marchetti, M., Mishra, N., Halpern, A.: Skin lesion analysis toward melanoma detection: A challenge at the international symposium on biomedical imaging (isbi) 2016, hosted by the international skin imaging collaboration (isic) (2016)
3. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 770–778 (2016)
4. Liberman, G., Acevedo, D., Mejail, M.: Classification of melanoma images with fisher vectors and deep learning. In: Iberoamerican Congress on Pattern Recognition (CIARP). pp. 732–739 (2019)
5. Perronnin, F., Sánchez, J., Mensink, T.: Improving the fisher kernel for large-scale image classification. In: Computer Vision – ECCV 2010. pp. 143–156. Springer Berlin Heidelberg, Berlin, Heidelberg (2010)
6. Ring, C., Cox, N., Lee, J.B.: Dermatoscopy. Clinics in Dermatology (2021)
7. Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A.C., Fei-Fei, L.: ImageNet Large Scale Visual Recognition Challenge. International Journal of Computer Vision (IJCV) **115**(3), 211–252 (2015)
8. Sánchez, J., Mensink, T., Verbeek, J.: Image classification with the fisher vector: Theory and practice. International Journal of Computer Vision **105** (12 2013)
9. Yu, L., Chen, H., Dou, Q., Qin, J., Heng, P.A.: Automated melanoma recognition in dermoscopy images via very deep residual networks. IEEE Transactions on Medical Imaging **36**(4), 994–1004 (2017)
10. Yu, Z., Jiang, X., Zhou, F., Qin, J., Ni, D., Chen, S., Lei, B., Wang, T.: Melanoma recognition in dermoscopy images via aggregated deep convolutional features. IEEE Transactions on Biomedical Engineering **66**(4), 1006–1016 (2019)
11. Yu, Z., Ni, D., Chen, S., Qin, J., Li, S., Wang, T., Lei, B.: Hybrid dermoscopy image classification framework based on deep convolutional neural network and fisher vector. In: 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017). pp. 301–304 (2017)
12. Yu, Z., Jiang, F., Zhou, F., He, X., Ni, D., Chen, S., Wang, T., Lei, B.: Convolutional descriptors aggregation via cross-net for skin lesion recognition. Applied Soft Computing **92**, 106281 (2020)
13. Yu, Z., Jiang, X., Wang, T., Lei, B.: Aggregating deep convolutional features for melanoma recognition in dermoscopy images. In: Machine Learning in Medical Imaging. vol. 10541, pp. 238–246 (09 2017)
14. Zhang, J., Xie, Y., Wu, Q., Xia, Y.: Medical image classification using synergic deep learning. Medical Image Analysis **54**, 10–19 (2019)