EVALUACIÓN DE UN DETECTOR DE IMÁGENES FALSAS ENTRENADO EN UNA RED GAN CONVOLUCIONAL

Se plantea la situación hipotética de una aplicación de venta de ropa que tiene un problema con usuarios que suben prendas para vender con imágenes falsas, generadas artificialmente. Se evalúa la posibilidad de entrenar una red GAN convolucional para usar su detector como un método automático para detectar estas imágenes falsas

Andrés Ricardo Pérez Rojas
riperezro@unal.edu.co
Universidad Nacional de Colombia
Introducción a los Sistemas Inteligentes

UNIVERSIDAD NACIONAL DE COLOMBIA

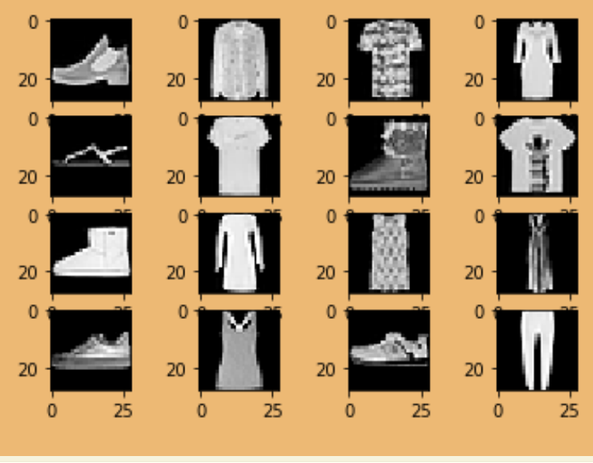
Introducción

El equipo detrás de la aplicación debe tomar una decisión sobre cómo manejar la situación de las imágenes falsas. Contratar moderadores que analicen detalladamente todas las fotos subidas por los usuarios es ineficiente, costoso, limitaría el crecimiento potencial de la aplicación y afectaría la experiencia de los usuarios al hacer mucho más lento el proceso de publicar un nuevo artículo.

Esto lleva al equipo a pensar en posibles soluciones automatizadas para detectar imágenes generadas o adulteradas. Una de las opciones para construir este detector es mediante las redes neuronales GAN (Generative Adversarial Networks). El objetivo es determinar la viabilidad de esta opción para filtrar las fotos que no sean reales.

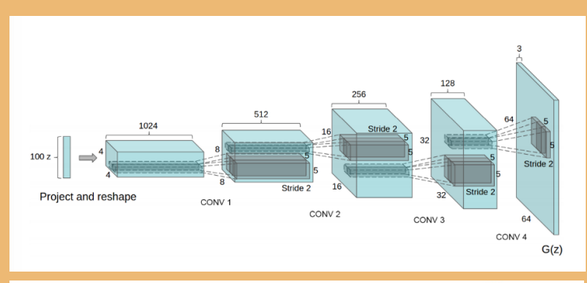
Dataset

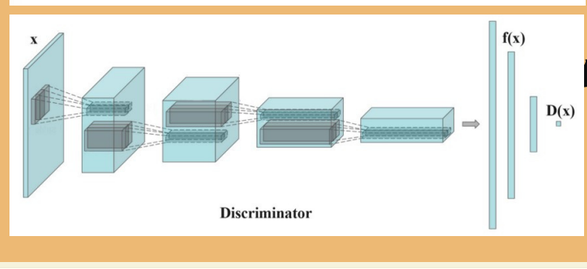
El proyecto se realizó sobre el dataset Fashion MNIST de Kaggle. Consta de imágenes de 28x28 en escala de grises, categorizadas en una de 10 clases que representan diferentes artículos de ropa. Estas son algunas de las imágenes del dataset:



Modelos

Las redes GAN convolucionales entrenan dos modelos a la vez: un generador de imágenes falsas y un discriminador que diferencia las imágenes reales de las generadas por el generador.

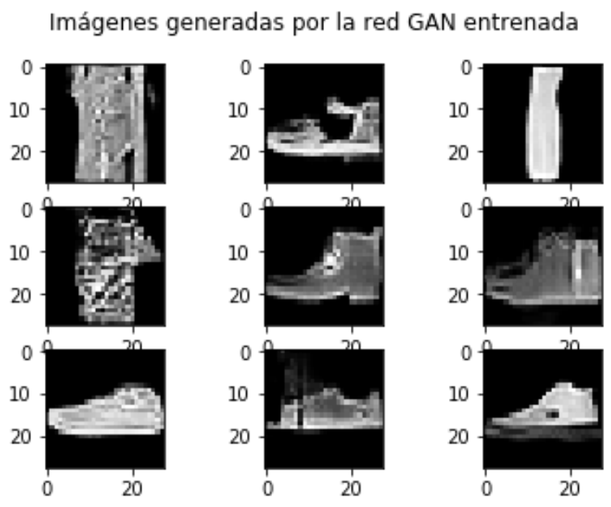


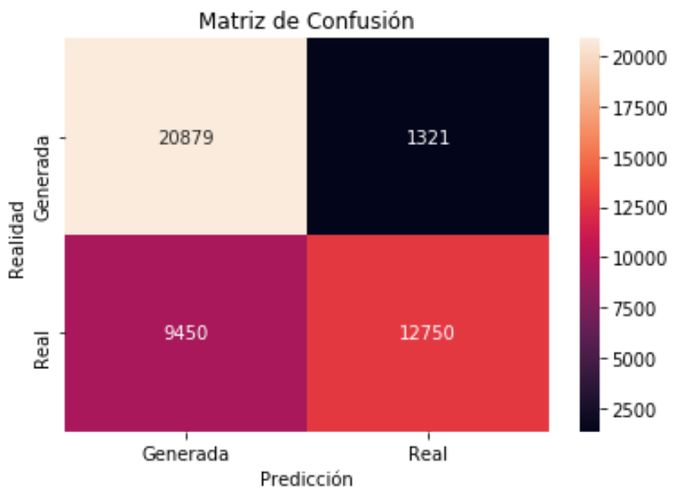


Resultados

Se experimentó con dos arquitecturas de red GAN convolucional encontradas en notebooks de Kaggle. Se muestran los resultados de la que tuvo mejor desempeño. En general las redes tuvieron un buen desempeño identificando imágenes falsas, pero tenían problemas con las imágenes reales, identificándolas muchas veces como falsas.

Se tomaron 3 de las 10 clases como anomalías para evaluar la generalización con resultados aceptables. La detección se considera buena para llevar a la práctica y realizar una inspección manual de las imágenes subidas que sean maracadas por la red neuronal como falsas





Conclusiones

El desempeño y la generalización encontrada deja pensar que si es posible entrenar una red GAN para que su detector ayude a detectar imágenes falsas. Los resultados sobre este dataset dejan pensar que se tendran bastantes falsos positivos pero pocos falsos negativos.

Se sugiere explorar una red que tome como entrada varios generadores GAN con las arquitecturas más usadas actualmente como entrada de discriminador, similar a como se realiza en Hsu et al.

Referencias:
[1]Zalando Research, "Fashion MNIST", Kaggle.com, 2017. [Online]. Available: https://www.kaggle.com/zalando-research/fashionmnist?select=fashion-mnist_train.csv. [Accessed: 15- Jan- 2022].

[2]Sayak, "Introduction to GANs on Fashion MNIST Dataset", Kaggle.com, 2020. [Online]. Available: https://www.kaggle.com/sayakdasgupta/introduction-to-gans-on-fashion-mnist-dataset. [Accessed: 15- Jan- 2022].

[3]A. Goel, "Introduction to GANs with Keras", Kaggle.com, 2020. [Online]. Available: https://www.kaggle.com/yushg123/introduction-to-gans-with-keras. [Accessed: 15- Jan- 2022].

[4]"Deep Convolutional Generative Adversarial Network | TensorFlow Core", TensorFlow, 2022. [Online]. Available: https://www.tensorflow.org/tutorials/generative/dcgan. [Accessed: 06- Feb- 2022].

[5]IBM, "IBM SSPS Modeler CRISP-DM Guide", Ibm.com, 2021. [Online]. Available: https://www.ibm.com/docs/en/spss-modeler/SaaS?topic=guide-introduction-crisp-dm. [Accessed: 15- Jan- 2022].

[6] A. Radford, L. Metz, S. Chintala. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. 2015. [Online]. Available: https://arxiv.org/abs/1511.06434. [Accessed: 06- Feb- 2022].

[7]C. Shorten, "Deeper into DCGANs", Medium, 2019. [Online]. Available: https://towardsdatascience.com/deeper-into-dcgans-2556dbd0baac. [Accessed: 06- Feb- 2022].

[8]J. Hui, "GAN — Ways to improve GAN performance", Medium, 2018. [Online]. Available: https://towardsdatascience.com/gan-ways-to-improve-gan-performance-acf37f9f59b. [Accessed: 06- Feb- 2022].

[9]C. Hsu, C. Lee and Y. Zhuang, Learning to Detect Fake Face Images in the Wild, 3rd ed. 2018. [Online]. Available: https://arxiv.org/abs/1809.08754. [Accessed: 06- Feb- 2022].