

SumajGAN: transferencia supervisada de maquillaje facial con redes generativas adversarias profundas

Pedro Jesús Guzmán-Ramos

pedrojesusguzmanramos@gmail.com / Universidad ESAN

Wilfredo Mamani-Ticona

wmamani@esan.edu.pe / Universidad ESAN

Recepción: 1/9/2020 Aceptación: 23/10/2020

RESUMEN. El reto de la transferencia de maquillaje de una imagen a otra ya está resuelto por los modelos BeautyGAN, PairedCycleGAN y BeautyGlow. Estos modelos lograron solucionar el reto mencionado mediante un enfoque de aprendizaje semisupervisado; lo cual resuelve el problema de obtener un *dataset* alineado de maquillaje, pero a costa de un alto poder de cómputo. Por este motivo, en esta investigación se creó un *dataset* de imágenes alineadas y adicionalmente se propuso un modelo de transferencia de maquillaje mediante un enfoque supervisado. El *dataset* está compuesto por 5400 grupos de imágenes, cada grupo de imágenes se encuentra conformado por una imagen sin maquillaje, una imagen con maquillaje de referencia y otra imagen con el maquillaje de la referencia y la identidad de la persona sin maquillaje. El modelo propuesto en esta investigación es llamado SumajGAN, el modelo se encuentra conformado por un discriminador de tipo PatchGAN y un generador de dos entradas inspiradas en un *autoencoder*. Se realizaron varios experimentos y el mejor resultado obtenido fue de 0,021658644 de error absoluto medio y alta resolución con una correcta transferencia de maquillaje. El modelo SumajGAN ha logrado realizar el objetivo planteado disminuyendo el tiempo de entrenamiento de modelos como BeautyGAN, PairedCycleGAN y BeautyGlow.

PALABRAS CLAVE: GAN / *autoencoder* / CNN / BeautyGAN / PairedCycleGAN / BeautyGlow

SumajGAN: Supervised Transfer of Face Makeup with Deep Generative Adversarial Networks

ABSTRACT. The challenge of transferring makeup from one image to another is already solved by BeautyGAN, PairedCycleGAN and Beauty Glow. These models managed to meet the aforementioned challenge using a semi-supervised learning approach which solves the problem of obtaining an aligned makeup dataset but at the expense of a high computing power. Therefore, in this research, an aligned image dataset was developed and, additionally, a makeup transfer model was proposed using a supervised approach. The dataset consisted of 5,400 groups of images: each group of images was composed of a person's no-makeup face image, a reference makeup face image, and a reference makeup face image transferred into the person's no-makeup face image. The model proposed in this research is called SumajGAN. It is made of a PatchGAN-type discriminator and a two-input generator based on an autoencoder. Several experiments were conducted, and the best result achieved a mean absolute error of 0.021658644 and a high-resolution makeup transfer. The SumajGAN model has managed to achieve the objective by reducing the training time of models such as BeautyGAN, PairedCycleGAN and Beauty Glow.

KEYWORDS: GAN / autoencoder / CNN / BeautyGAN / PairedCycleGAN / BeautyGlow

1. INTRODUCCIÓN

Un modelo capaz de realizar la transferencia de maquillaje basado en una referencia es el presentado por Li, Qian, Dong, Liu y Yan (2018), titulado *BeautyGAN: transferencia de maquillaje facial a nivel de instancia con red adversa profunda generativa*. BeautyGAN plantea desarrollar un modelo basado en las redes generativas adversarias, incluyendo dos generadores similares con doble entrada y salida cada una, el primero para imágenes con maquillaje y el segundo para imágenes sin maquillaje. BeautyGAN se encuentra basado en el enfoque semisupervisado y realiza su aprendizaje a nivel de dominio. Los resultados obtenidos en la investigación del BeautyGAN muestran una alta precisión para la transferencia de maquillaje y, según los estudios comparativos perceptivos, muestran un 61,84 % de aceptación frente a un 33,91 % de la investigación presentada por Liao, Yao, Yuan, Hua y Sing (2017), titulada *Transferencia de atributos visuales a través de una analogía de imagen profunda*.

Otro modelo capaz de realizar la transferencia de maquillaje dentro del grupo de las redes generativas adversarias es PairedCycleGAN, presentado por Chang, Lu, Yu y Finkelstein (2018), titulado *PairedCycleGAN: transferencia de estilo asimétrico para aplicar y quitar maquillaje*. El modelo PairedCycleGAN emplea el enfoque semisupervisado basado en la confrontación consistente de redes generativas. En el proceso de entrenamiento el modelo PairedCycleGAN en una primera fase se encarga de remover el maquillaje de una fotografía de referencia, mientras que en una segunda fase se encarga de aplicar maquillaje a una fotografía de entrada. Los resultados obtenidos en la investigación de PairedCycleGAN, según los estudios comparativos perceptivos, muestran un 65,7 % frente a un 34,3 % de la investigación presentada por Liao, Yao, Yuan, Hua y Sing (2017), titulada *Transferencia de atributos visuales a través de una analogía de imagen profunda*.

Finalmente, otro modelo específico en realizar la transferencia de maquillaje condicionada a una referencia, dentro del campo de las redes generativas basadas en flujo, es el modelo BeautyGlow presentado por Chen, Hui, Wang, Tsao, Shuai y Cheng (2019), titulado *BeautyGlow: marco de transferencia de maquillaje bajo demanda con red generativa reversible*. BeautyGlow resuelve el problema de la transferencia de maquillaje realizando una descomposición de la imagen de entrada y la referencia en vectores latentes, gracias al modelo Glow. Estos vectores latentes son unidos para formar una nueva imagen de entrada sin maquillaje con la aplicación del maquillaje de referencia, mediante la ayuda de funciones de pérdida de la estructura y el maquillaje. Los resultados muestran imágenes de alta calidad y de gran realismo. Por otro lado, según la aplicación realizada de un estudio perceptivo comparativo, el modelo BeautyGlow muestra un 55 % de aceptación frente al 45 % del modelo BeautyGAN.

BeautyGAN, PairedCycleGAN y BeautyGlow son modelos de alta precisión en la transferencia de maquillaje condicionada a una referencia. Estos modelos realizan su aprendizaje de acuerdo con innovadoras maneras marcadas dentro del enfoque semisupervisado. La investigación abordó el problema de la transferencia condicionada de maquillaje mediante el desarrollo

de un modelo enmarcado en el enfoque supervisado, llamado SumajGAN. Este modelo emplea un generador, basado en U-Net, de dos entradas y un discriminador, tipo PatchGAN. SumajGAN, al ser un modelo supervisado, tiene un *target* que permite calcular el resultado, basado en métricas.

2. METODOLOGÍA

La metodología propuesta en la investigación consta de cuatro etapas: 1) obtención del *dataset*, en el cual se detallan los procedimientos necesarios para obtener las imágenes; 2) preprocesamiento del *dataset*, en el que se realiza el tratamiento a las imágenes como estandarizar el formato, el tamaño y la focalización de los rostros para que el modelo pueda tener un mejor rendimiento; 3) diseño del modelo, en el cual se describe el fundamento teórico del modelo SumajGAN y aspectos necesarios para su desarrollo; 4) evaluación, en donde se detallan los criterios de evaluación del rendimiento del modelo, y 5) prueba, en el que se detalla la aplicación del mejor modelo.

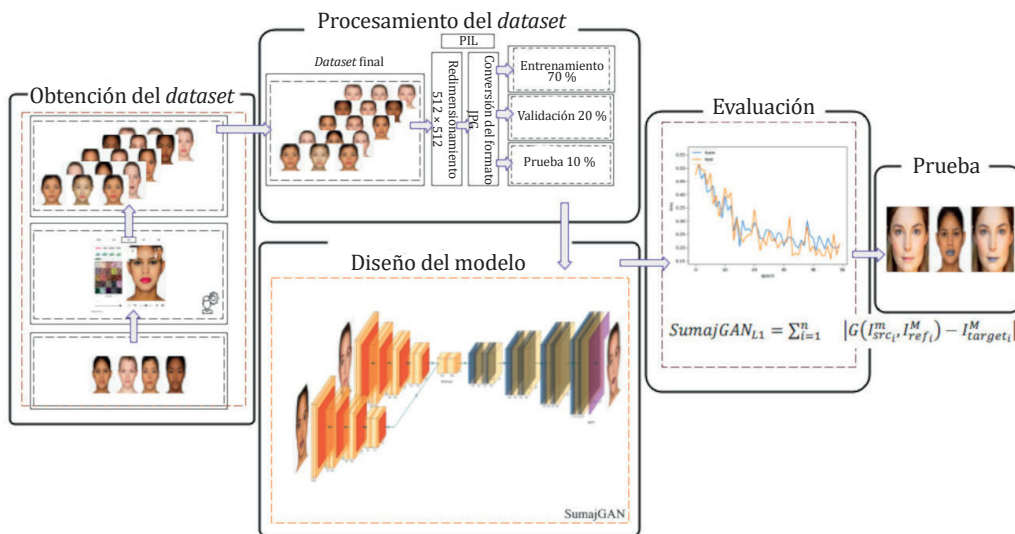


Figura 1. Metodología propuesta

Elaboración propia

2.1 Obtención del *dataset*

Se creó un *dataset* conformado por 5400 grupos de imágenes alineadas con diversos tipos de maquillaje. Un grupo de imágenes alineadas se encuentra conformado por tres imágenes (figura 2): *image source*, *image reference* e *image target*. Cada imagen del *dataset* tiene una

dimensión de 512×512 píxeles, y alta resolución. Se realizó de manera manual la alineación de cada imagen utilizando programas de edición para colocar un maquillaje similar a otro en cuanto a color y tono.



Figura 2. Bloque de imágenes alineadas (izquierda: *image source*, centro: *image reference*, derecha: *image target*)
Elaboración propia

2.2 Preprocesamiento del *dataset*

Este proceso permitió el redimensionamiento, a una escala de 512×512 píxeles; conversión a un formato JPG de todas las imágenes y separación del *dataset* en entrenamiento, validación y prueba. La proporción del *dataset* de entrenamiento, validación y prueba es de 70 %, 20 % y 10 % respectivamente del *dataset* final. Esta separación proporcional del *dataset* de entrenamiento, validación y prueba es realizada mediante el recorrido de todo el *dataset* y una asignación probabilística del 0,7, 0,2 y 0,1 de cada bloque de imágenes.

2.3 Diseño del modelo

La transferencia de maquillaje involucra tener dos imágenes de entrada. La primera, I_{src}^m , formada por la fotografía del usuario sin maquillaje. La segunda, I_{ref}^M , formada por una fotografía de referencia la cual cuenta con maquillaje. El modelo propuesto en esta investigación (SumajGAN) debe transformar las entradas I_{src}^m e I_{ref}^M y producir una nueva imagen I_{src}^M , manteniendo la identidad del cliente y el maquillaje de la imagen de referencia (figura 3).

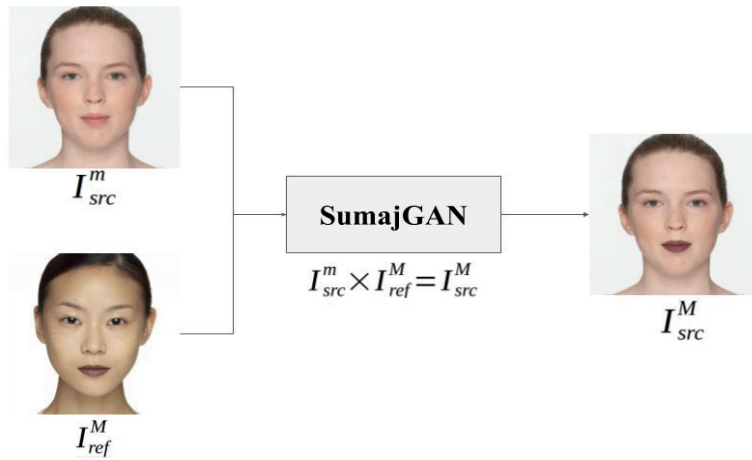


Figura 3. Diseño global del modelo SumajGAN

Elaboración propia

El modelo propuesto está conformado por una red generadora y otra discriminadora. La red discriminadora (figura 5) se encuentra basada en un discriminador de tipo PatchGAN; mientras que el generador (figura 4) se encuentra basado en el modelo U-Net con la particularidad de tener dos entradas convolucionales y estar unidos al *decoder* al nivel del *bottleneck* y capas al mismo nivel mediante las *skip-connection*. Un bloque de convolución está conformado por dos subbloques de convolución de *kernel* 3×3 , seguido por una función de activación no lineal ReLU y finalmente una capa *max-pooling* 2×2 . Un bloque de deconvolución está formado por una capa *up-pooling*, seguido por dos subbloques de convolución con su respectiva función de activación no lineal ReLU. La configuración es realizada con base en el modelo U-Net que emplea doble convolución en cada nivel para luego hacer un *pooling* o *up-pooling* de 2×2 .

SumajGAN al igual que el modelo pix2pix son modelos de entrenamiento supervisado y tienen como objetivo asemejarse lo más posible a un *target*. Asimismo, con un entrenamiento adecuado, aseguran un resultado óptimo como lo planteado en el *target*. Análogamente se puede decir que pix2pix es a CycleGAN como SumajGAN es a BeautyGAN. Al igual que pix2pix, busca solucionar el mismo problema que CycleGAN, pero de manera supervisada. SumajGAN busca solucionar el mismo problema de BeautyGAN de manera supervisada.

2.4 Evaluación

El modelo BeautyGAN propuesto por Li *et al.* (2018) plantea la métrica *identity loss* y *makeup loss* para cuidar la transferencia de la identidad de la *image source* y preservar el maquillaje de la *image reference*. La *identity loss* permite delimitar la generación de una imagen preservando la estructura del rostro de la persona de la *image source*. La *identity loss* es medida mediante la

diferencia absoluta entre la salida producida por el generador y la *image source* (ecuación 1). El *makeup loss* permite delimitar la generación de una imagen conservando atributos de la *image reference*. El *makeup loss* es medido mediante la diferencia absoluta entre la salida producida por el generador y la *image reference* (ecuación 2). El *target loss* permite medir la pérdida de la imagen generada frente al *target*. El *target loss* es medido mediante la diferencia absoluta de la *image target* y la imagen de salida del generador (ecuación 3). La pérdida del ciclo adversario es medida mediante la entropía cruzada de la *image target* y la imagen producida por el generador luego de pasar al discriminador (ecuación 4). La función de pérdida total será la sumatoria de las pérdidas mencionadas (ecuación 5).

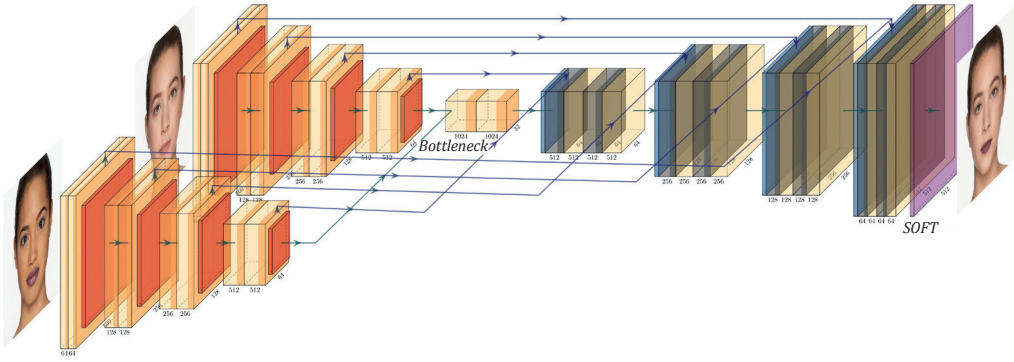


Figura 4. Modelo generador del SumajGAN
Elaboración propia

$$LOSS_{Identity}(G) = \alpha |G(I_{src}^m, I_{ref}^M) - I_{src}^m| \quad (1)$$

$$LOSS_{Makeup}(G) = \beta |G(I_{src}^m, I_{ref}^M) - I_{ref}^M| \quad (2)$$

$$LOSS_{Target}(G) = \theta |G(I_{src}^m, I_{ref}^M) - I_{target}^M| \quad (3)$$

$$LOSS_{GAN}(G, D) = \log(D(I_{target}^M)) + \log(1 - D(G(I_{src}^m, I_{ref}^M))) \quad (4)$$

$$LOSS_{Total}(G, D) = LOSS_{Identity} + LOSS_{Makeup} + LOSS_{Target} + LOSS_{GAN} \quad (5)$$

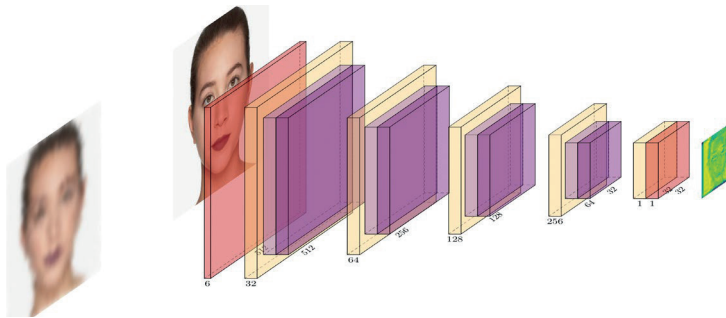


Figura 5. Modelo discriminador PatchGAN
Elaboración propia

El modelo SumajGAN sigue un enfoque supervisado, esto quiere decir que tiene un *target* que puede condicionar la salida del modelo. Con ello, se plantea evaluar el modelo empleando la pérdida L1, la cual medirá la diferencia absoluta de la salida producida por el generador con la *image target* (ecuación 6). La función de pérdida L1 será empleada tanto en el *dataset* de entrenamiento como en el *dataset* de validación para controlar el *overfitting* y permitir que el modelo logre generalizar en diversas situaciones reales. Los valores del dominio de la métrica se encuentran entre 0 y el $+\infty$, siendo los valores próximos a 0 los de mejor rendimiento.

$$SumajGAN_{L1} = \sum_{i=1}^n |G(I_{src_i}^m, I_{ref_i}^M) - I_{target_i}^M| \quad (6)$$

2.5 Prueba

La etapa de prueba permitió verificar la posibilidad de implementar el modelo en un ambiente de producción. Por otro lado, en esta etapa se realizaron diversos experimentos con la finalidad de obtener un modelo óptimo que permitiera transferir de la mejor manera el maquillaje de la referencia a la imagen del usuario.

3. RESULTADOS

El mejor experimento muestra una pérdida del generador (figura 6) sin tendencia a mejorar para el *dataset* de evaluación a partir de la época 30, con una pérdida para la época 57 de 3,118789. En la pérdida del discriminador (figura 7) se observa que el modelo tiene tendencia al *overfitting* a partir de la época 50. La métrica L1 (figura 8) muestra que a partir de la época 21 hasta la época 57 hay muy poca mejora en el aprendizaje. El valor L1 en la época 57 es 0,021658644. Por otro lado, el ejemplo generado en cada época (figura 9) muestra una mejora de la resolución en cada época; asimismo, desde las primeras épocas se evidencia la

transferencia de maquillaje en la imagen generada. En la función de pérdida L1 del generador (figura 8), se aprecia un cambio importante en las 20 primeras épocas, estas son referidas a la construcción de la estructura del rostro. Mientras que las siguientes épocas muestran cambios menores, ya que se refieren a cambios relativamente menores. Sin embargo, para el modelo propuesto es importante, ya que es la fijación del color y tono de los distintos maquillajes. Esto se verificó guardando ejemplos generados con el *dataset* de prueba, el cual evidencia que con cada iteración la fijación del maquillaje mejora.

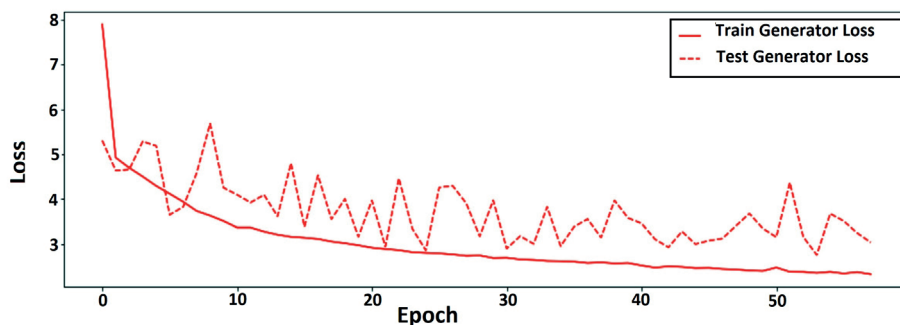


Figura 6. Pérdida del generador del mejor experimento

Elaboración propia

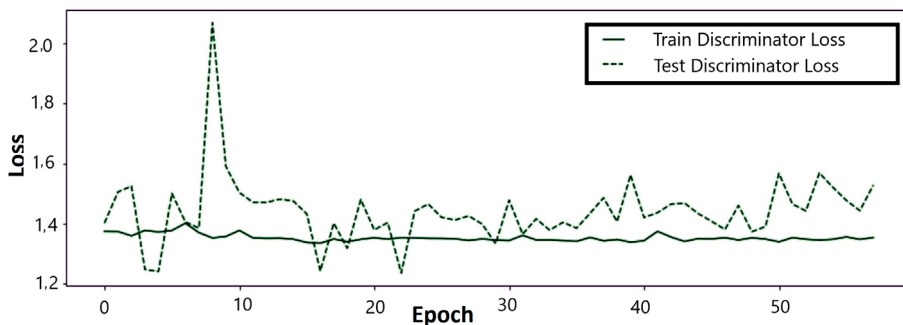


Figura 7. Pérdida del discriminador del mejor experimento

Elaboración propia

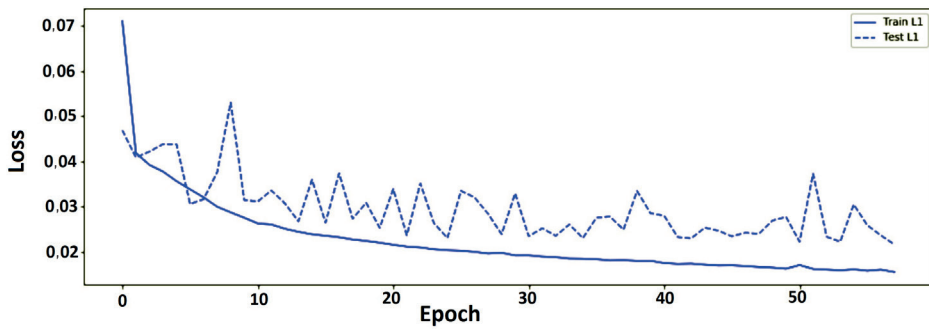


Figura 8. Métrica del mejor experimento
Elaboración propia



Figura 9. Métrica del mejor experimento
Elaboración propia

Las pruebas realizadas a las imágenes de las figuras 10 y 11 muestran imágenes generadas de muy alta resolución, muy similares a la *image source* de cada ejemplo. Asimismo, las pruebas muestran un alto rendimiento en la transferencia de maquillaje, en cuanto al color y tipo de maquillaje.

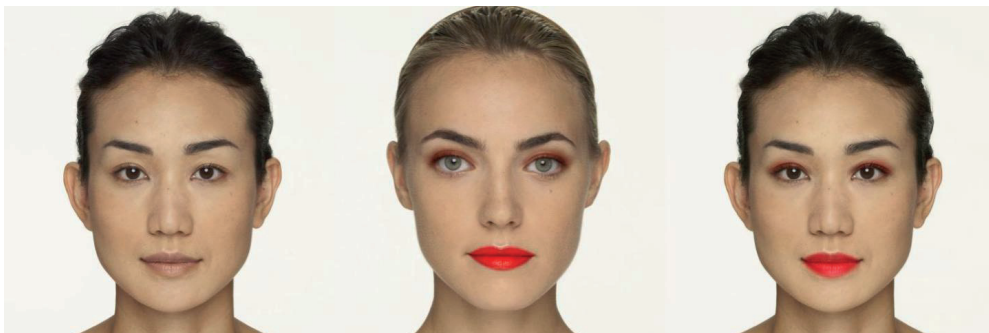


Figura 10. Métrica del mejor experimento
Elaboración propia

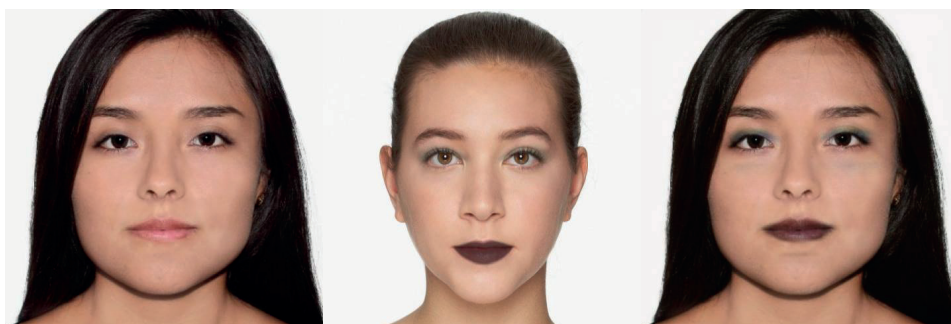


Figura 11. Métrica del mejor experimento
Elaboración propia

4. CONCLUSIONES

En la investigación se propuso un modelo denominado SumajGAN, el cual fue capaz de sintetizar una imagen con maquillaje condicionada a una imagen del usuario y una referencia de maquillaje. El modelo SumajGAN pertenece al grupo de las redes generativas adversarias, estas se encuentran formadas por dos redes, una conocida como generador y la otra como discriminador.

SumajGAN fue diseñado para transferir maquillaje de una fotografía de referencia a otra fotografía sin maquillaje. El entrenamiento del modelo SumajGAN ha sido realizado mediante el enfoque supervisado. Se creó un *dataset* de imágenes de rostros con maquillaje y rostros sin maquillaje condicionadas a una referencia; fue realizado mediante la edición manual de atributos de maquillaje como el color de los labios, aplicación de base en el rostro, delineador de los ojos y color de los párpados a una imagen sin maquillaje. Las imágenes obtenidas mediante esta técnica quizá no pueden ser muy atractivas, ya que requiere la disposición de mucho tiempo y los datos pueden encontrarse sesgados al no ser tomados del mundo real. Sin embargo, esta forma de obtener los datos es muchísimo más barata que tener que maquillar a muchas mujeres y tomarles una fotografía. Por otro lado, contar con un *dataset* alineado disminuye el tiempo de entrenamiento frente a modelos como PairedCycleGAN, BeautyGAN y BeautyGlow.

El modelo SumajGAN está conformado por una red generadora y otra discriminadora. La red discriminadora se encuentra basada en un discriminador de tipo PatchGAN; mientras que el generador se encuentra basado en el modelo U-Net con la particularidad de tener dos entradas convolucionales y estar unidos al *decoder* al nivel del *bottleneck* y capas al mismo nivel mediante las *skip-connection*.

El desempeño luego del entrenamiento del modelo SumajGAN según la métrica L1 fue de 0,021658644 para el *dataset* de evaluación. Por otra parte, las pruebas realizadas al modelo SumajGAN han mostrado imágenes de alta resolución y una correcta transferencia de maquillaje en cuanto al tipo y color de maquillaje. De esta manera, el modelo SumajGAN es, potencialmente, aplicable en entornos de producción de venta *online* de productos cosméticos para facilitar los procesos de compra de los clientes.

El modelo ha sido entrenado usando un CPU Core i7 con un tiempo estimado de tres horas por cada época; por ello es recomendable utilizar un procesador de tipo GPU o TPU para mejorar significativamente el tiempo de entrenamiento. Por otro lado, es recomendable probar con un *framework* como Pytorch, el cual es el más empleado para modelos generativos, y comparar los resultados en cuanto a la eficiencia de generación y tiempo de entrenamiento.

SumajGAN plantea un modelo de generación de imágenes condicionada a dos entradas, uno tomado como estructura y otro como referencia y siempre condicionado a un resultado previo, aprendizaje supervisado. Como trabajo futuro se plantea generalizar el modelo para otros objetivos como transferencia de prendas de vestir u otros atributos. Asimismo, se tiene pensado realizar un modelo que permite el ingreso de más de dos entradas que combine atributos, siempre dentro del campo supervisado.

REFERENCIAS

- Chen, H., Hui, K., Wang, S., Tsao, L., Shuai, H., y Cheng, W. (2019). *BeautyGlow: On-Demand Makeup Transfer Framework with Reversible Generative Network*. IEEE Xplore. <https://bit.ly/2wAPZDk>
- Chang, H., Lu, J., Yu, F., y Finkelstein, A. (2018) *PairedCycleGAN: Asymmetric Style Transfer for Applying and Removing Makeup* [Sesión de conferencia]. 2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). doi 10.1109/CVPR.2018.00012
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., y Bengio, Y. (2020). Generative Adversarial Networks. *Commun. ACM*, 63(11), 139-144. <https://doi.org/10.1145/3422622>
- Gulli, A., y Pal, S. (2017). *Deep Learning with Keras Implement Neural Networks with Keras on Theano and TensorFlow*. Packt Publishing.
- Guo, D., y Sim, T. (2009). Digital face makeup by example. *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 73-79. <https://doi.org/10.1109/CVPR.2009.5206833>

- Isola, P., Zhu J., Zhou T., y Efros, A. (2017). *Image-to-Image Translation with Conditional Adversarial Networks*. 2017 [Sesión de conferencia]. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), (pp. 5967-5976). 10.1109/CVPR.2017.632.
- Kingma, D., y Dhariwal, P. (2018). Glow: Generative Flow with Invertible 1×1 Convolutions. arXiv:1807.03039v2.
- Li, T., Qian, R., Dong, C., Liu, S., y Yan, Q. (2016). *Makeup Like a Superstar: Deep Localized Makeup Transfer Network* [Sesión de conferencia]. IJCAI'16 Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence), (pp. 2568-2575).
- Li, T., Qian, R., Dong, C., Liu, S., y Yan, Q. (2018). BeautyGAN: *Instance-level Facial Makeup Transfer with Deep Generative Adversarial Network* [Sesión de conferencia]. 2018 ACM Multimedia Conference on Multimedia Conference (ACMMM), (pp. 645-653). 10.1145/3240508.3240618.
- Zhu, J., Park, T., Isola, P., y Efros, A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. *ArXiv preprint* arXiv:1703.10593.