

# Eliminación de los efectos de humo en cirugía laparoscópica usando redes antagónicas generativas y el principio del canal oscuro

*Removal of smoke effects in laparoscopic surgery via adversarial neural network and the dark channel prior*

Sebastián Salazar-Colores<sup>1</sup>, Hugo A. Moreno<sup>1</sup>, Ulises Moya<sup>2</sup>, César J. Ortiz-Echeverri<sup>3</sup>, Luis A. Tavares de la Paz<sup>4</sup> y Gerardo Flores<sup>1\*</sup>

<sup>1</sup>Centro de Investigaciones en Óptica, León, Guanajuato; <sup>2</sup>Centro de Inteligencia Artificial, Gobierno del Estado de Jalisco y Universidad Autónoma de Guadalajara, Guadalajara, Jalisco; <sup>3</sup>Universidad Autónoma de Querétaro, Facultad de Informática, Querétaro, Querétaro; <sup>4</sup>Hospital Regional de Alta Especialidad del Bajío, León, Guanajuato. México

## Resumen

**Antecedentes:** Durante la cirugía laparoscópica, la calidad de la imagen puede verse gravemente degradada por el humo quirúrgico causado por el uso de herramientas de disección de tejidos que reducen la visibilidad de los órganos y tejidos.

**Objetivo:** Mejorar la visibilidad en cirugía laparoscópica mediante la combinación de técnicas de procesamiento de imágenes basadas en técnicas clásicas e inteligencia artificial. **Método:** Desarrollo de un enfoque híbrido para la eliminación de los efectos del humo quirúrgico, basado en la combinación del método del principio del canal oscuro (DCP, dark channel prior) y una arquitectura de red neuronal píxel a píxel conocida como red antagónica generativa (GAN, generative adversarial network). **Resultados:** Los resultados experimentales han demostrado que el método propuesto logra un mejor rendimiento que los resultados individuales de DCP y GAN en cuanto a calidad de la restauración, obteniendo (según las métricas de la proporción máxima de señal a ruido [PSNR, Peak Signal-to-Noise Ratio] y el índice de similitud estructural [SSIM, Structural Similarity Index]) mejores resultados que otros métodos relacionados. **Conclusiones:** El enfoque propuesto disminuye los riesgos y el tiempo de la cirugía laparoscópica, ya que una vez que la red está correctamente entrenada, el sistema puede mejorar la visibilidad en tiempo real.

**Palabras clave:** Laparoscopia. Mejoramiento de imágenes. Eliminación de humo. Procesamiento de imágenes. Red antagónica generativa.

## Abstract

**Background:** In laparoscopic surgery, image quality can be severely degraded by surgical smoke caused by the use of tissue dissection tools that reduce the visibility of the observed organs and tissues. **Objective:** Improve visibility in laparoscopic surgery by combining image processing techniques based on classical methods and artificial intelligence. **Method:** Development of a hybrid approach to eliminating the effects of surgical smoke, based on the combination of the dark channel prior (DCP) method and a pixel-to-pixel neural network architecture known as a generative adversarial network (GAN).

**Results:** Experimental results have shown that the proposed method achieves better performance than individual DCP and GAN results in terms of restoration quality, obtaining (according to PSNR and SSIM index metrics) better results than some

### Correspondencia:

\*Gerardo Flores

Lomas del Bosque 115

Lomas del Campestre

C.P. 37150, León, Guanajuato, México

E-mail: gflores@cio.mx

0009-7411/© 2020 Academia Mexicana de Cirugía. Publicado por Permanyer. Este es un artículo *open access* bajo la licencia CC BY-NC-ND (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Fecha de recepción: 31-08-2020

Fecha de aceptación: 06-12-2020

DOI: 10.24875/CIRU.20000951

Cir Cir. 2022;90(1):74-83

Contents available at PubMed

[www.cirugiacirujanos.com](http://www.cirugiacirujanos.com)

related state-of-the-art methods. **Conclusions:** The proposed approach decreases the risks and time of laparoscopic surgery because once the network is trained, the system can improve real-time visibility.

**Keywords:** Laparoscopy. Haze removal. Image processing. Generative adversarial network.

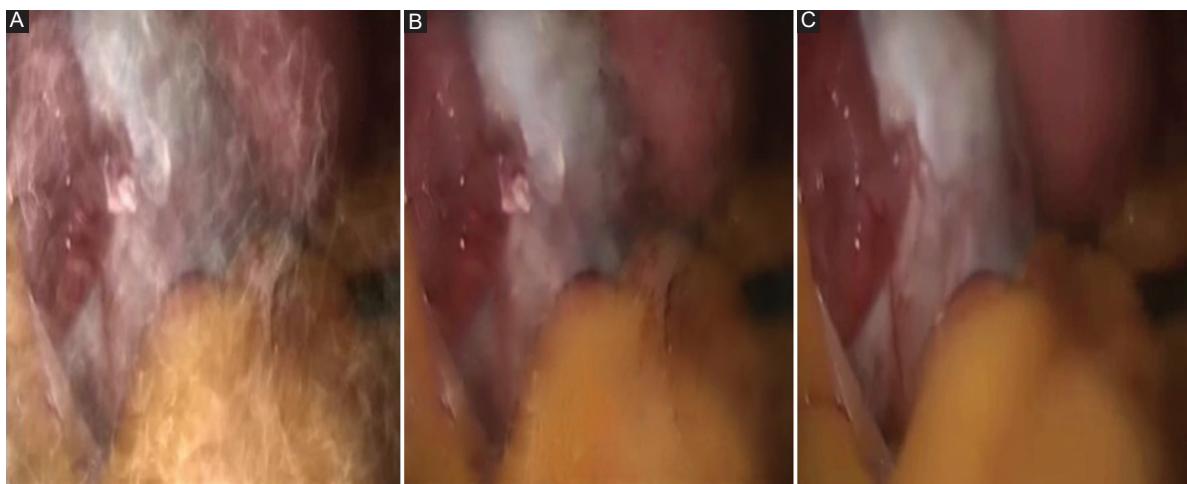
## Introducción

La cirugía laparoscópica consiste en la inserción de una cámara y dispositivos quirúrgicos a través de pequeñas incisiones. Ha reemplazado en gran medida a la cirugía tradicional en diversos campos quirúrgicos, como la cirugía abdominal, ginecología, torácica, etc., debido a que presenta una menor incidencia de infecciones y una recuperación más rápida, y generalmente deja menos secuelas que la cirugía tradicional. En este tipo de procedimientos se realizan cauterizaciones en los tejidos, lo cual origina vapor de agua y monóxido de carbono que disminuyen la visibilidad de las cámaras usadas. Esta falta de visibilidad aumenta el tiempo de la cirugía e incluso la probabilidad de errores cometidos por el cirujano, lo que puede derivar en consecuencias negativas para la salud del paciente<sup>1</sup>. Comúnmente, el proceso de eliminación de humo se realiza con instrumentos médicos, pero esta manera resulta costosa y poco práctica, por lo que se requiere otra opción para resolver el problema. Lo anteriormente mencionado ha motivado el desarrollo de investigación en procesamiento digital de imágenes, la cual se ha enfocado en la búsqueda de métodos para reducir los efectos del humo y, por tanto, aumentar la visibilidad.

Una de las técnicas clásicas más usadas para este propósito parte del modelo de dispersión atmosférica, en el que se estima el mapa de transmisión (el cual contiene información acerca de la densidad de humo) y las componentes de la luz en una imagen determinada<sup>2</sup>. Por ejemplo, Wang, et al.<sup>1</sup> proponen una técnica basada en el modelo físico de la dispersión del humo, similar al modelo usado en el procesamiento de imágenes para eliminar los efectos de niebla (*dehazing*), en el que se determina el nivel de humo considerando el hecho de que en los píxeles que lo contienen se presenta un bajo contraste, así como bajas diferencias entre canales. Basándose en esta observación, se define una función de costo y se resuelve utilizando un método de Lagrange. Kotwal, et al.<sup>3</sup> formulan la cuestión de la niebla y la eliminación de humo en las imágenes laparoscópicas como un problema de inferencia bayesiana, en el cual se usa un sistema probabilístico con modelos de las imágenes sin niebla, así como un mapa de transmisión que indica la atenuación del color

en presencia de humo. Gu, et al.<sup>4</sup> implementan un sistema digital basado en el principio del canal oscuro (DCP, *Dark Channel Prior*), en donde se observa que estadísticamente las imágenes libres de niebla en exteriores contienen algunos píxeles con muy baja intensidad al menos en uno de los canales de color.

Muchos problemas en el procesamiento de imágenes, gráficos y visión requieren la conversión de una imagen de entrada a una imagen de salida determinada. Hasta hace un par de años, como se ha comentado, la mayor parte de la investigación se realizaba mediante métodos diseñados «ingenierilmente», es decir, el investigador analizaba y buscaba patrones sobre los que diseñaba un algoritmo explícito. En la actualidad, con los nuevos avances del aprendizaje automático (*machine learning*), el cual es una parte de la inteligencia artificial y trata sobre modelos que mejoran con la experiencia, la situación ha cambiado radicalmente. Ahora, gracias sobre todo al aprendizaje profundo (*deep learning*) basado en redes neuronales convolucionales (CNN, *Convolutional Neural Network*) que son capaces de encontrar patrones y abstracciones autónomamente, la tarea del investigador se enfoca en adaptar o desarrollar y ajustar un modelo de aprendizaje automático. Las redes generativas antagonicas (GAN, *Generative Adversarial Network*) propuestas en 2014 están formadas por dos redes neuronales y son una solución de propósito general que ha obtenido buenos resultados en una amplia variedad de aplicaciones que requieren un mapeo de imagen a imagen, siendo actualmente la tecnología puntera en muchas de ellas<sup>5-9</sup>. Las GAN constan de dos partes principales: una red neuronal que actúa como una función generadora, en la cual a partir de una imagen de entrada se genera una imagen de salida con determinados cambios, y una función discriminadora que evalúa la imagen generada con una imagen real con el fin de clasificarla como real o falsa. La idea básica de las GAN es lograr la generación de imágenes tan fieles a las originales que el discriminador no logre encontrar diferencia alguna. Isola, et al.<sup>5</sup> demuestran que este enfoque es efectivo para sintetizar imágenes a partir de mapas de etiquetas, reconstruir objetos a partir de mapas de bordes y colorear imágenes originalmente en escala de grises. En



**Figura 1.** Funcionamiento del modelo DCP + GAN en una imagen quirúrgica laparoscópica. **A:** imágenes de entrada con humo de  $\text{CO}_2$  sintético. **B:** nuestro resultado con DCP + GAN. **C:** ground-truth.

trabajos recientes se han propuesto arquitecturas de aprendizaje profundo para reducir el efecto del humo en las imágenes laparoscópicas<sup>10,11</sup>. Chen, et al.<sup>12</sup> proponen el uso de las GAN para el aprendizaje no supervisado y la eliminación del humo en las imágenes. En este trabajo presentamos los resultados de aplicar una GAN y además proponemos un método híbrido que combina una técnica de procesamiento de imágenes con una GAN. En la figura 1 se muestra un ejemplo de la arquitectura aquí propuesta, en donde se usa el DCP como etapa de preprocessamiento bajo la hipótesis de la reducción en la complejidad de los patrones presentes en las imágenes que posteriormente ingresan a una arquitectura GAN.

Los experimentos fueron realizados sobre un conjunto de datos con imágenes de laparoscopia a las cuales se les generó artificialmente el efecto del humo mediante un *software* de edición de vídeo. De acuerdo con las métricas usadas para evaluar los métodos propuestos, el uso de GAN para reducir el efecto del humo mejora la reconstrucción de la imagen frente al enfoque basado en el DCP y el modelo de dispersión.

El presente trabajo está estructurado de la siguiente forma: en el apartado «II» (marco teórico) se presentan los fundamentos teóricos y conceptuales usados para la implementación de la red, así como las métricas para la evaluación de los resultados; en el apartado «III» (método propuesto) se explican las configuraciones usadas en el diseño del experimento; en el apartado «IV» (resultados) se presentan los resultados obtenidos, así como las tablas comparativas respecto a trabajos relacionados; y finalmente se incluyen una

discusión, unas conclusiones y una perspectiva de futuras investigaciones.

### Marco teórico

En este apartado se presentan los fundamentos teóricos necesarios para la implementación de la metodología propuesta.

### MODELO DE DISPERSIÓN ATMOSFÉRICA

La degradación atmosférica es un fenómeno físico causado por partículas en el medio atmosférico que absorben y dispersan la luz<sup>13</sup>. Esta degradación puede ser expresada, utilizando el modelo dicromático<sup>14</sup>, como:

$$I(x,y) = J(x,y)t(x,y) + A(1-t(x,y))$$

donde  $I(x,y)$  es la intensidad de la imagen de color en cada pixel  $(x,y)$  con presencia de niebla,  $J(x,y)$  es la intensidad de la imagen RGB en cada píxel  $(x,y)$  sin degradación,  $A$  es la matriz normalizada de la componente de la luz atmosférica en cada canal de la imagen y  $t(x,y)$  representa el mapa de transmisión en un atmósfera homogénea, modelado por:

$$t(x,y) = e^{-\beta d(x,y)}, 0 < t(x,y) < 1,$$

donde  $\beta$  es el coeficiente de dispersión, que depende principalmente del tamaño de la partícula de dispersión<sup>14</sup> y  $d(x,y)$  representa la distancia de la cámara a la escena en la posición del píxel  $(x,y)$ .

Utilizando el modelo dicromático es posible obtener la imagen sin los efectos de neblina  $J(x,y)$  siempre y cuando se pueda obtener una estimación precisa de  $t(x,y)$  y  $A$ <sup>13</sup>.

## CANAL OSCURO

El canal oscuro  $I^{dark}$  para cada píxel se define como:

$$I^{dark}(x,y) = \min c \in \{RGB\} \left( \min z \in \Omega(x,y) \frac{I^c(z)}{A^c} \right)$$

donde  $\Omega(x,y)$  es un kernel generalmente cuadrado centrado en la posición  $(x,y)$  (utilizar un kernel cuadrado no tiene mayor motivación que reducir al máximo el tiempo requerido para el procesamiento),  $I^c(z)$  son los elementos de la imagen en las posiciones  $z \in \Omega(x,y)$ , representa cada imagen en los respectivos canales RGB y es el valor de en cada una de las componentes  $c$ .

El DCP se basa en la observación de que, en imágenes sin neblina, al menos un canal de color tiene una intensidad muy baja en algunos píxeles. En otras palabras, la intensidad en una sección  $\Omega$  de la imagen debería tener un valor muy bajo en algún canal si no hay neblina entre la cámara y el objeto capturado. Esto se encuentra definido como la relación estadística entre el canal oscuro y las regiones sin neblina, donde se cumple:

$$I^{dark}(x,y) \rightarrow 0$$

La relación existente entre  $I^{dark}$  y  $t$  se expresa como:

$$t(x,y) = 1 - \omega I^{dark}(x,y)$$

donde  $0 < \omega < 1$  representa el nivel de restauración. De acuerdo con He, et al.<sup>15</sup>,  $\omega=0.95$ , argumentando que este valor produce un mejor aspecto en las imágenes restauradas. Por otra parte, la luz atmosférica puede ser estimada mediante la ecuación:

$$A_c = \max \sum_{c=1}^3 \left( \begin{array}{l} \text{argarg max } (x,y) \in 0.1\% \\ (h, \omega) (I^{dark}(x,y)) \end{array} \right)$$

donde  $h$  es la altura y  $\omega$  es el ancho de la imagen.

## REDES ANTAGÓNICAS GENERATIVAS

Las GAN, presentadas en 2014 por Goodfellow, et al.<sup>16</sup>, son una clase de métodos de inteligencia artificial de aprendizaje no supervisado, implementados por un sistema de dos redes neuronales que compiten mutuamente, y un generador (G) que se encarga de generar una imagen falsa a partir del entrenamiento de muchas imágenes originales, que aprende las características fundamentales de un tipo de imagen en particular; por ejemplo, si esta etapa es entrenada con muchos rostros diferentes, tendrá la capacidad de generar

rostros falsos con características reales. En una segunda etapa se conecta un discriminador (D) que tiene como objetivo identificar si una imagen es real o falsa. Al estar conectadas las dos etapas en un proceso iterativo, el trabajo G es generar imágenes falsas con características tan reales que logre «engaños» a la etapa D, momento en que se considera que la GAN está entrenada y en la aplicación se usa solo la etapa G. Tanto la etapa G como la etapa D se implementan usando capas convolucionales. En el caso de G, se realizan operaciones convolucionales para hacer una reducción de dimensionalidad en las imágenes y posteriormente una operación inversa o deconvolucional, hasta generar una imagen de iguales dimensiones a la imagen de entrada. La etapa D, por su parte, usa operaciones convolucionales hasta el punto de tener un clasificador binario que decida si la imagen generada es real (se asigna un valor de 1) o falsa (se asigna un valor de 0).

## CAPAS CONVOLUCIONALES

Las operaciones convolucionales en imágenes están inspiradas en el córtex visual del ojo humano, en donde algunas células se especializan en detectar formas particulares, como líneas y bordes. A lo largo de la jerarquía de la corteza, la complejidad de los estímulos se va incrementando. De manera análoga, las capas convolucionales conectadas en una jerarquía permiten que las primeras capas detecten características simples de la imagen y, a medida que se incrementan las capas, se van logrando niveles mayores de abstracción. Esta cualidad confiere a las redes neuronales convolucionales la capacidad de «ver» e identificar objetos. Específicamente, cada neurona de un mapa de características está conectada a una región de neuronas vecinas en la capa anterior. El nuevo mapa de características se puede obtener haciendo una convolución de la entrada y luego aplicando una función de activación.

## UNIDAD LINEAL RECTIFICADA

La unidad lineal rectificada (ReLU, Rectified Linear Unit) es una de las funciones de activación no saturada más notable, la cual se define como:

$$a_{i,j,k} = \max(z_{i,j,k}, 0)$$

donde  $z_{i,j,k}$  es la entrada de la función de activación en la ubicación  $(i,j)$  en el canal  $k$ . ReLU es una función lineal por partes que pone la parte negativa a cero y retiene la parte positiva. La simple operación

$\max(\cdot)$  de ReLU le permite computar mucho más rápido que las funciones de activación sigmoide o  $\tanh(\cdot)$ , y también induce la dispersión en las unidades ocultas y permite a la red obtener fácilmente representaciones dispersas. Se ha demostrado que las redes profundas pueden ser entrenadas eficientemente usando ReLU<sup>17</sup>.

### REDES CONDICIONALES GENERATIVAS ANTAGÓNICAS

Las GAN son modelos generativos que aprenden un mapeo a partir de un vector de ruido aleatorio  $z$  para una imagen de salida  $y$ ,  $G:z \rightarrow y$ . Por su parte, las redes condicionales generativas antagónicas (cGAN) aprenden a partir de la imagen de entrada  $x$  y el vector de ruido aleatorio  $z^5$ , es decir,  $G:x,z \rightarrow y$ . Este tipo de red es muy utilizado en métodos que requieren una traslación de imagen a imagen.

### Método

En este apartado se describe la metodología usada para la obtención de los datos y la implementación del algoritmo propuesto.

### Obtención de la base de datos

Las imágenes utilizadas en este trabajo se obtuvieron de vídeos de distintos repositorios públicos, así como de vídeos proporcionados por un grupo de médicos especialistas en cirugía laparoscópica. Puede verse un vídeo comparativo de los resultados de los métodos propuestos en un vídeo quirúrgico real en <https://www.youtube.com/watch?v=QvUKcHonCHw&feature=youtu.be>. De los vídeos disponibles se obtuvieron 6000 imágenes representativas de diferentes niveles de afectaciones por humo. Para generar los datos de entrada se simuló artificialmente el humo, usando el software de procesamiento gráfico de código abierto Blender, formando imágenes de entrada con dimensiones de 512 × 512 píxeles.

### Métricas

Para tener una visión del desempeño del método propuesto en este trabajo se evalúan los resultados empleando métricas ampliamente utilizadas en la literatura: la proporción máxima de señal a ruido (PSNR, *Peak Signal-to-Noise Ratio*) y el índice de similitud estructural (SSIM, *Structural Similarity*

*Index*). A continuación se explican brevemente estos conceptos:

- La PSNR es una medida cuantitativa de la calidad de una reconstrucción<sup>18</sup>. Se utiliza ampliamente en imágenes. Para definir la métrica PSNR es necesario definir el error cuadrático medio (MSE, *Mean Squared Error*), el cual, para dos imágenes monocromas  $I$  y  $J$  de tamaño  $m \times n$ , se define como:

$$MSE = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} \|I(i,j) - J(i,j)\|^2$$

Y la PSNR se obtiene mediante una relación logarítmica de base 10 dada por:

$$PSNR = 10 \log \left( \frac{MAX_I^2}{MSE} \right) = 20 \log \left( \frac{MAX_I}{\sqrt{MSE}} \right)$$

donde  $MAX = 2^B - 1$  y  $B$  es el número de bits utilizados en la imagen. Valores altos de PSNR indican mejores restauraciones.

- El SSIM es una métrica de similitud de imagen perceptiva que fue propuesta como alternativa al MSE y la PSNR para aumentar la correlación con la evaluación subjetiva. Para las imágenes originales y reconstruidas  $I$  y  $J$ , el SSIM se define como:

$$SSIM(I,J) = \frac{(2\mu_I\mu_J + C_1)(2\sigma_{IJ} + C_2)}{(\mu_I^2 + \mu_J^2 + C_1)(\sigma_I^2 + \sigma_J^2 + C_2)}$$

donde  $\mu$ ,  $\sigma$  y  $\sigma_{IJ}$  son la media, la varianza y la covarianza de las imágenes, respectivamente.

### Método propuesto

El método propuesto se basa en el supuesto de que una red neuronal cGAN tiene un mejor rendimiento en tanto su entrada sea más parecida a su salida esperada. Se espera entonces que al aplicar el canal oscuro antes de entrenar la red neuronal, el rendimiento de esta se vea incrementado. Para realizar un análisis del rendimiento del enfoque propuesto se propusieron tres experimentos:

- 1) Técnica basada en el canal oscuro:

El canal oscuro ha dado excelentes resultados en tareas de *dehazing*; no obstante, el problema abordado es distinto, ya que la atmósfera no es homogénea, por lo que ya no existe una relación entre el canal oscuro y la distancia de los objetos. Sin embargo, la distancia  $d(x)$  de la cámara al cuerpo humano es despreciable, por lo cual el canal oscuro

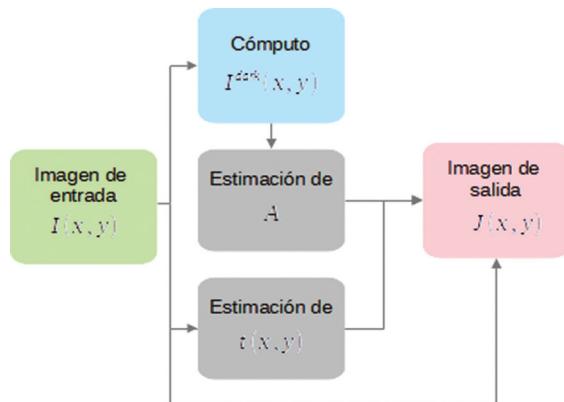


Figura 2. Método basado en DCP.

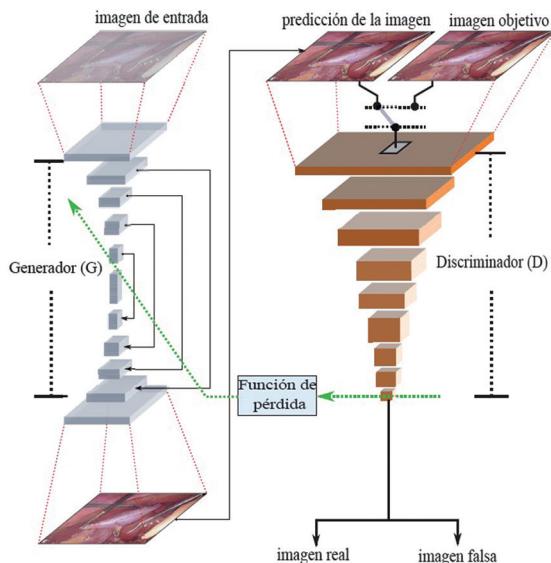


Figura 3. Método basado en DCP. Diagrama de la arquitectura cGAN. La imagen de entrada pasa a través de varias capas convolucionales hasta lograr una reducción de dimensionalidad en la capa intermedia de la estructura generadora (G). Seguidamente se realiza un proceso inverso con capas convolucionales transpuestas, hasta llegar a una imagen con igual dimensionalidad a la imagen de entrada, la cual será la predicción de la imagen sin ruido (humo de  $\text{CO}_2$ ). Por otra parte, un discriminador (D) basado en operaciones convolucionales compara la predicción de la imagen con la imagen objetivo.

tiene una correspondencia directa con el coeficiente de dispersión  $\beta$ . Por esta razón, el canal oscuro tiene también validez para esta condición. La relación entre el canal oscuro y la transmisión se puede expresar como:

$$t(x,y) = 1 - \omega I^{\text{dark}}(x,y)$$

Sustituyendo (1) en (5) y considerando además que la atmósfera no es homogénea, se obtienen diferentes coeficientes de dispersión para cada elemento  $(x,y)$ :

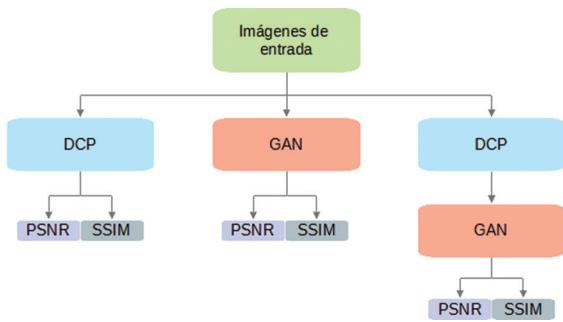


Figura 4. Metodología propuesta para la evaluación de desempeño.

$$e^{-\beta(x,y)d(x,y)} = 1 - \omega I^{\text{dark}}(x,y)$$

Por tanto:

$$I^{\text{dark}}(x,y) = \frac{1 - e^{-\beta(x,y)d(x,y)}}{\omega}$$

Suponiendo la distancia del cuerpo constante y  $\omega=1$ :

$$I^{\text{dark}}(x,y) = 1 - e^{-d\beta(x,y)}$$

Podemos decir que la relación del canal oscuro con la transmisión está en función ahora del coeficiente y la densidad de partículas  $\beta$  en cada uno de los píxeles, resultando útil para la eliminación de los efectos.

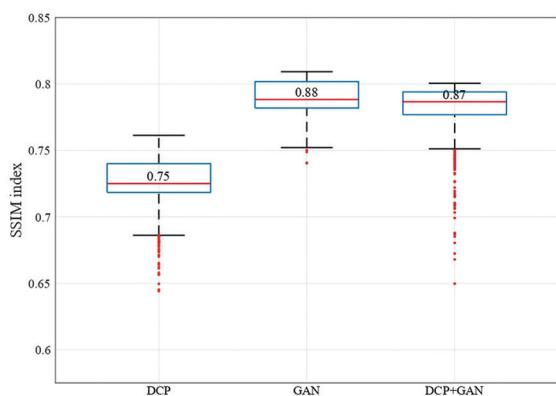
En la figura 2 se muestra el método basado en el DCP implementado.

### 2) Utilizando la cGAN propuesta:

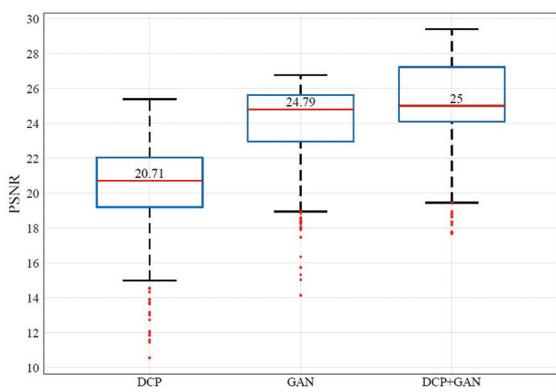
En la figura 3 se muestra un diagrama de la red neuronal utilizada. En las tablas 1 y 2 se muestran la arquitectura y los hiperparámetros del generador y del discriminador empleados. Como función de optimización en la red neuronal se empleó la estimación del momento adaptativo (ADAM, *Adaptive Moment Estimation*). Este algoritmo es una extensión del descenso de gradiente estocástico para actualizar los pesos de red de forma iterativa en función de los datos de entrenamiento; recientemente ha visto una adopción más amplia para aplicaciones de aprendizaje profundo en visión por computadora y procesamiento de lenguaje natural. Como métrica de pérdida se utiliza el MSE, el cual se usa como la función de pérdida para la regresión.

### 3) Usando una combinación de los enfoques mencionados:

Las imágenes de entrada en la figura 3 son las imágenes de salida  $J$  de la figura 2. La evaluación del desempeño para cada caso se representa en la figura 4, en la que se muestra la metodología



**Figura 5.** Comparación del desempeño de los métodos propuestos de acuerdo con el SSIM.



**Figura 6.** Comparación del desempeño de los métodos propuestos de acuerdo con la PSNR.

propuesta para evaluar el desempeño de las tres configuraciones mencionadas.

### Configuración de los experimentos

Los experimentos fueron realizados con una computadora con un procesador Ryzen Threadripper 1900, 128 Gb de memoria RAM y una tarjeta gráfica Nvidia RTX 2080 Ti, usando Linux Ubuntu 18.10, Python 3.5, Librerías OpenCV 3.3 y Keras 2.1.0.

### Resultados

En las figuras 5 y 6 se muestra la comparación de las métricas SSIM y PSNR para los métodos DCP, cGAN y DCP-cGAN. La primera observación importante en estas figuras es la mejora significativa de los métodos cGAN y DCP-cGAN frente a la reconstrucción basada únicamente en DCP, el cual alcanzó valores de SSIM =

**Tabla 1.** Arquitectura de la red generadora utilizada

| Capa | Conv | Kernel | Salto | Definición                       | Tamaño          |
|------|------|--------|-------|----------------------------------|-----------------|
| 1    | 64   | 4      | 2     | Conv->Batchnorm->Leaky ReLU      | (256, 256, 4)   |
| 2    | 128  | 4      | 2     | Conv->Batchnorm->Leaky ReLU      | (128, 128, 128) |
| 3    | 256  | 4      | 2     | Conv->Batchnorm->Leaky ReLU      | (64, 64, 256)   |
| 4    | 512  | 4      | 2     | Conv->Batchnorm->Leaky ReLU      | (32, 32, 512)   |
| 5    | 512  | 4      | 2     | Conv->Batchnorm->Leaky ReLU      | (16, 16, 512)   |
| 6    | 512  | 4      | 2     | Conv->Batchnorm->Leaky ReLU      | (8, 8, 512)     |
| 7    | 512  | 4      | 2     | Conv->Batchnorm->Leaky ReLU      | (4, 4, 512)     |
| 8    | 512  | 4      | 2     | Conv->Batchnorm->Leaky ReLU      | (1, 1, 512)     |
| 9    | 512  | 4      | 2     | Deconv->Batchnorm->ReLU->Dp(0.5) | (2, 2, 1024)    |
| 10   | 512  | 4      | 2     | Deconv->Batchnorm->ReLU->Dp(0.5) | (4, 4, 1024)    |
| 11   | 512  | 4      | 2     | Deconv->Batchnorm->ReLU->Dp(0.5) | (8, 8, 1024)    |
| 12   | 512  | 4      | 2     | Deconv->Batchnorm->ReLU->Dp(0.5) | (16, 16, 1024)  |
| 13   | 512  | 4      | 2     | Deconv->Batchnorm->ReLU          | (32, 32, 1024)  |
| 14   | 256  | 4      | 2     | Deconv->Batchnorm->ReLU          | (64, 64, 512)   |
| 15   | 128  | 4      | 2     | Deconv->Batchnorm->ReLU          | (128, 128, 256) |
| 16   | 64   | 4      | 2     | Deconv->Batchnorm->ReLU          | (256, 256, 128) |
| 17   | 3    | 4      | 2     | Tanh                             | (512, 512, 3)   |

0.75 y PSNR = 20.71, mientras cGAN alcanzó valores de SSIM = 0.88 y PSNR = 24.79. En cuanto al desempeño entre cGAN y DCP-cGAN, se observa que la métrica de PSNR fue ligeramente superior para DCP-cGAN, con un valor PSNR = 25.00, frente a PSNR = 24.79 logrado por cGAN. En el caso de la métrica SSIM, cGAN obtuvo un valor medio de SSIM = 0.88, frente a SSIM = 0.87 con DCP-cGAN. El hecho de agregar una etapa previa DCP antes de la cGAN abre un camino a una arquitectura de *deep learning* que enfoque el aprendizaje en las regiones de mayor incidencia de humo.



**Figura 7.** Comparación de los resultados con los métodos utilizados. **A:** imágenes de entrada con humo de  $CO_2$  sintético. **B:** DCP. **C:** GAN. **D:** DCP + GAN. **E:** ground-truth.

Los resultados actuales pueden ser mejorados ajustando los parámetros del canal oscuro, así como hiperparámetros como el número de capas ocultas, el número de convoluciones por capa y tamaño de los filtros en la convolución, entre otros.

## Discusión

La figura 7 muestra una comparación que permite observar cualitativamente el desempeño de los distintos métodos implementados. Si bien la técnica basada

**Tabla 2. Arquitectura de la red neuronal discriminadora utilizada**

| Capa | Conv | Kernel | Salto | Definición                         | Tamaño          |
|------|------|--------|-------|------------------------------------|-----------------|
| 1    | 64   | 4      | 2     | Conv->Batchnorm->Leaky ReLU        | (256, 256, 4)   |
| 2    | 128  | 4      | 2     | Conv->Batchnorm->Leaky ReLU        | (128, 128, 128) |
| 3    | 256  | 4      | 2     | Conv->Batchnorm->Leaky ReLU        | (64, 64, 256)   |
| 4    | 0    | 0      | 0     | ZeroPadding2D                      | (32, 32, 512)   |
| 5    | 512  | 4      | 1     | Conv                               | (16, 16, 512)   |
| 6    | 0    | 0      | 0     | BatchNorm->Leaky ReLU->ZeroPadding | (8, 8, 512)     |
| 7    | 1    | 4      | 1     | Conv                               | (4, 4, 512)     |

únicamente en DCP logra reducir de manera significativa la presencia de humo en las imágenes, se puede apreciar también una alteración en los colores originales. Por su parte, la técnica basada solo en GAN conserva mejor los colores originales de la imagen y logra una buena reducción del humo, pero algunos detalles finos, principalmente los bordes y los reflejos de luz, se ven alterados con respecto al *ground-truth*. Finalmente, el método propuesto DCP + GAN logra una reducción del humo conservando también los colores originales, así como una mejor recuperación de zonas con altos niveles de brillo y de bordes, en comparación con cada una de las técnicas por sí solas.

Los resultados cuantitativos y cualitativos indican que la combinación de cGAN con una etapa previa en la que se calcule el canal oscuro ayuda a la identificación de las regiones con presencia de humo, haciendo que la generación de imágenes sintéticas por parte de cGAN se enfoque más en dichas regiones que en el resto de la imagen. Algo muy importante que se debe mencionar es que, aunque se visualiza un mejor rendimiento con el método propuesto DCP + cGAN, la ventaja no es significativa en términos estadísticos, ya que como se puede ver en las figuras 5 y 6 existe un solapamiento en los *boxplots*; sin embargo, los resultados son un buen punto de inicio para la experimentación de una mezcla de conceptos más refinada.

## Conclusiones

En este artículo se propone un método para mejorar la reducida visibilidad en imágenes laparoscópicas causada por la obstrucción visual debida al humo generado por la combustión de los tejidos durante la

cirugía. El método propuesto se basa en la combinación de un método clásico de procesamiento, el DCP, y un método de inteligencia artificial, la cGAN. Este método es capaz de reducir la influencia del humo en la imagen de salida y recuperar la gama de colores similar a la imagen original, mostrando que un método híbrido puede incrementar el desempeño de otros métodos. Esta combinación de técnicas ha mostrado tener un alto desempeño en la reconstrucción de las imágenes objetivo, de acuerdo con las métricas PSNR e SSIM usadas en el presente trabajo.

Este artículo abre un nuevo panorama en la investigación relacionada al mostrar que la mezcla de un método clásico con uno de inteligencia artificial puede superar los resultados de ambos aislados. Como trabajo futuro, se espera mejorar el comportamiento de la cGAN al agregar una función de pérdida especialmente diseñada para la resolución del problema expuesto. Por otra parte, se plantea hacer una búsqueda exhaustiva de los parámetros en el modelo de dispersión con el fin de ajustar la etapa de preprocesamiento en la entrada de la cGAN.

## Agradecimientos

Hugo A. Moreno desea agradecer especialmente al Consejo Nacional de Ciencia y Tecnología por el apoyo financiero brindado para sus estudios de maestría.

## Conflictos de intereses

Los autores declaran que no existe conflicto de intereses.

## Responsabilidades éticas

**Protección de personas y animales.** Los autores declaran que para esta investigación no se han realizado experimentos en seres humanos ni en animales.

**Confidencialidad de los datos.** Los autores declaran que han seguido los protocolos de su centro de trabajo sobre la publicación de datos de pacientes.

**Derecho a la privacidad y consentimiento informado.** Los autores han obtenido el consentimiento informado de los pacientes y/o sujetos referidos en el artículo. Este documento obra en poder del autor de correspondencia.

## Bibliografía

1. Wang C, Cheikh FA, Kaaniche M, Beghdadi A, Elle OJ. (2018). Variational based smoke removal in laparoscopic images. *Biomed Eng Online*. 2018;17:139.

2. Xu Y, Wen J, Fei L, Zhang Z. Review of video and image defogging algorithms and related studies on image restoration and enhancement. *IEEE Access*. 2015;4:165-88.
3. Kotwal A, Bhalodia R, Awate SP. Joint desmoking and denoising of laparoscopy images. En: 2016 IEEE 13<sup>th</sup> International Symposium on Biomedical Imaging (ISBI). p. 1050-4.
4. Gu L, Liu P, Jiang C, Luo M, Xu Q. Virtual digital defogging technology improves laparoscopic imaging quality. *Surg Innov*. 2015;22:171-6.
5. Isola P, Zhu JY, Zhou T, Efros AA. (2017). Image-to-image translation with conditional adversarial networks. En: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2017. p. 1125-34.
6. Karras T, Laine S, Aila T. (2019). A style-based generator architecture for generative adversarial networks. En: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2019. p. 4401-10.
7. Chen Y, Jakary A, Avadiappan S, Hess CP, Lupo JM. Qsmgan: improved quantitative susceptibility mapping using 3d generative adversarial networks with increased receptive field. *NeuroImage*. 2020;207:116389.
8. Zhang Y, Gan Z, Fan K, Chen Z, Henao R, Shen D, et al. Adversarial feature matching for text generation. *arXiv:1706.03850 [stat.ML]*. Disponible en: <https://arxiv.org/abs/1706.03850>
9. Zhang H, Sindagi V, Patel VM. (2019). Image de-raining using a conditional generative adversarial network. *arXiv:1701.05957 [cs.CV]*. Disponible en: <https://arxiv.org/abs/1701.05957>
10. Wang C, Mohammed AK, Cheikh FA, Beghdadi A, Elle OJ. Multiscale deep desmoking for laparoscopic surgery. *Medical Imaging 2019: Image Processing*. International Society for Optics and Photonics. 2019;10949:109491Y.
11. Bolkar S, Wang C, Cheikh FA, Yildirim S. Deep smoke removal from minimally invasive surgery videos. En: 25<sup>th</sup> IEEE International Conference on Image Processing (ICIP); 2018. p. 3403-7.
12. Chen L, Tang W, John WN. Unsupervised learning of surgical smoke removal from simulation. The 11<sup>th</sup> Hamlyn Symposium on Medical Robotics Proceedings, June 2018. p. 24- 27. 2018.
13. Li B, Ren W, Fu D, Tao D, Feng D, Zeng W, et al. Benchmarking single-image dehazing and beyond. *IEEE Trans Image Process*. 2018;28:492-505.
14. Gibson KB, Nguyen TQ. Fast single image fog removal using the adaptive Wiener filter. En: 2013 IEEE International Conference on Image Processing; 2013. p. 714-8.
15. He K, Sun J, Tang X. Single image haze removal using dark channel prior. *IEEE Trans Pattern Anal Mach Intell*. 2011;33:2341-53.
16. Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. (2014). Generative adversarial nets. *arXiv:1406.2661 [stat.ML]*. Disponible en: <https://arxiv.org/abs/1406.2661>.
17. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. En: *Advances in Neural Information Processing Systems*. 2012. Disponible en: <https://proceedings.neurips.cc/paper/2012/file/c399862d3b9d6b76c8436e924a-68c45b-Paper.pdf>.
18. Huynh-Thu Q, Ghanbari M. Scope of validity of PSNR in image/video quality assessment. *Electronics Letters*. 2008;44:800-1.