

Quantification of Collagen Fiber Orientation and Organization by Singular Value Decomposition

Angel A. Zeitoune^{1,2}, Luciana A. Erbes^{1,2}, Victor H. Casco^{1,3} and Javier F. Adur^{1,2}

¹Laboratorio de Microscopía Aplicada a Estudios Moleculares y Celulares, Facultad de Ingeniería, Universidad Nacional de Entre Ríos, Oro Verde, Argentina

¹Biofotónica y Procesamiento de Información Biológica (ByPIB),

Centro de Investigación y Transferencia de Entre Ríos, UNER, CONICET, Oro Verde, Argentina

²Centro de Investigación y Transferencia de Entre Ríos, UNER, CONICET, Oro Verde, Argentina

Abstract— Second-harmonic generation microscopy is a new imaging tool that allows analyzing non-centrosymmetric structures, such as collagen. This method can be used to detect structural organization of human ovarian extracellular matrix. Collagen remodeling can be used as early indicator of disease states, such as tumor development and invasiveness. Scoring structural characteristics like main collagen orientation and organization do allow differentiating between healthy and cancerous tissue. In this report, it is proposed a new method to calculate the main orientation angle as well as the degree of collagen fibers organization. The algorithm is based on the analysis of image Fourier transform modulus, characterizing spectral distribution by using singular value decomposition. The preliminary results obtained in this work demonstrate that the method allows to discriminate healthy from cancerous tissue.

Keywords— SHG, microscopy, collagen, organization, tumor, SVD.

Resumen— La microscopía de generación de segundo armónico es una nueva herramienta de imágenes que permite analizar estructuras no centrosimétricas, como el colágeno. Este método puede usarse para detectar la organización estructural en el ovario humano. La remodelación del colágeno es un indicador temprano de los estados de enfermedad, como el desarrollo tumoral y la invasividad. Las características estructurales como la orientación principal y organización de las fibras de colágeno permiten diferenciar entre tejido sano y canceroso. En este informe, se propone un nuevo método para calcular el ángulo de orientación principal, así como el grado de organización de las fibras de colágeno. El algoritmo se basa en el análisis del módulo de transformación de Fourier de la imagen, que caracteriza la distribución espectral mediante el uso de la descomposición en valores singulares. Los resultados preliminares obtenidos en este trabajo demuestran que el método permite discriminar entre tejido sano y canceroso.

Palabras clave— SHG, microscopía, colágeno, organización, cáncer, SVD.

I. INTRODUCCIÓN

EL cáncer es una de las principales causas de mortalidad en el mundo por lo que se destinan ingentes esfuerzos en su estudio. La detección temprana aumenta considerablemente la eficacia de los tratamientos, la probabilidad de supervivencia se incrementa, la morbilidad se reduce y el tratamiento resulta más económico.

Estudios recientes demuestran que los desarrollos tumorales provocan remodelación del microambiente circundante, incluyendo cambios en la organización de las fibras colágenas de la matriz extracelular de los tejidos [?]. Su observación y cuantificación permite comprender mejor cómo la organización de la matriz, la densidad y la composición afectan tanto la formación como la progresión del tumor.

Debido al ordenamiento semicristalino y a la estructura no centrosimétrica de la molécula proteica de colágeno, las fibras pueden ser visualizadas mediante nuevas técnicas de microscopía no lineal como la microscopía de generación de

segundo armónico (SHG) [?]. Esta técnica permite capturar imágenes de alto contraste y resolución espacial de las fibras de colágeno; posee capacidad para penetrar especímenes gruesos sin la necesidad de tinción [?], [?]. Se basa en el efecto óptico no lineal producido por una luz láser intensa; cuando dos fotones combinan su energía para excitar a una molécula, esta emite un fotón con exactamente el doble de energía (el doble de frecuencia).

El objetivo de este trabajo es proponer una nueva metodología para cuantificar dos parámetros de la remodelación del colágeno: organización y orientación. El primero mide el grado de paralelismo existente entre las fibras de colágeno, mientras que el segundo determina el ángulo de inclinación en el cual el grupo mayoritario se encuentra orientado, las que en tejidos normales se orientan paralelas al epitelio suprayacente.

Actualmente, el método más extendido para el cálculo de la orientación principal, es a partir del análisis del módulo de la transformada de Fourier (FT) de la imagen. Este método fue introducido por Rossi *et al* [?] para demostrar cambios morfológicos en la córnea porcina. La imagen formada por

la FT, es ajustada a una elipse, donde la dirección del eje principal es perpendicular a la dirección de las fibras y la organización se cuantifica como el cociente entre la longitud del radio menor sobre el radio mayor [?], [?], [?]. También se puede estimar usando un ajuste lineal [?], [?], [?], transformada de Radón [?], función de autocorrelación [?], basado en un histograma en coordenadas polares [?], o a través de la matriz de covarianza [?].

Analizando directamente la imagen espacial, Hu *et al* [?] propusieron calcular la orientación principal mediante la estimación de la matriz de coocurrencia de niveles de gris, donde se busca el ángulo donde el estadístico correlación es máximo.

La organización de las fibras de colágeno se cuantifica como el cociente entre la longitud del radio menor sobre el radio mayor de la elipse ajustada en la imagen transformada [?], [?], [?], [?], [?], [?]. Otros autores la estiman mediante la frecuencia espacial máxima [?]. Fung *et al* la estima mediante el gradiente espectral de potencia [?]. En este trabajo se propone un método matemático y directo para calcular ambos parámetros utilizando descomposición en valores singulares.

II. MATERIALES

A. Material biológico

Las biopsias ováricas fueron obtenidas del archivo del Departamento de Patología del Hospital Universitario de Campinas, SP, Brasil. El protocolo experimental fue aprobado por el Comité de Ética de la Facultad de Ciencias Médicas, Unicamp - N° 1437/2001. Todos los procedimientos fueron realizados de acuerdo con la Declaración de Helsinki y los principios éticos de la comunidad médica.

Se prepararon muestras de tejido ovárico de 4 μm de espesor y luego se tiñeron con hematoxilina-eosina empleando técnicas estándar. Cada sección de tejido fue evaluada y diagnosticada por un patólogo experimentado. Se capturaron imágenes de SHG de cada una de las muestras utilizando un microscopio invertido Axio Observer Z.1 equipado con un cabezal Zeiss LSM 780-NLO confocal scan (Carl Zeiss AG, Alemania) de acuerdo al método descrito por Adur *et al* [?]. En este trabajo preliminar fueron empleadas 5 muestras de tejido ovárico normal y 5 muestras de tejido ovárico patológico del tipo adenocarcinoma mucinoso.

III. ALGORITMO

A. Función ventana

La FT de una imagen es tratada como parte de una señal periódica de imágenes idénticas. Esta imagen es extendida horizontal y verticalmente como un mosaico, creando discontinuidades en los bordes. Estas generan altas componentes frecuenciales en la imagen transformada, las que se visualizan como líneas verticales y horizontales que se asemejan al “signo más”. Esta distorsión puede atenuarse significativamente multiplicando la imagen espacial por una función que suavice los bordes al color negro (valor cero de fondo). En el método propuesto fue utilizada una ventana *Tukey*, con una relación de suavizado α de 0,25 [?].

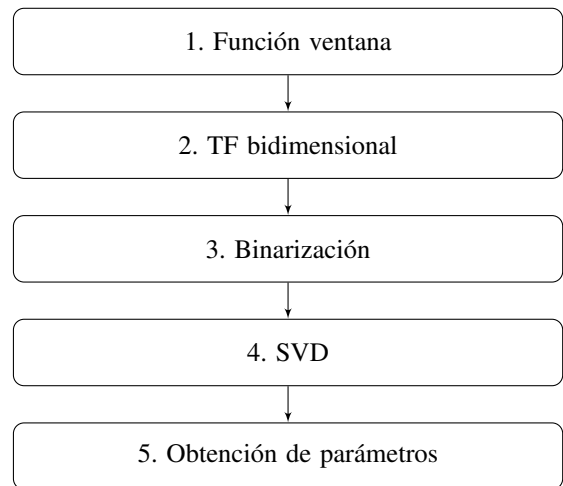


Fig. 1. Algoritmo para el cálculo de la orientación y organización de las fibras de colágeno.

B. Transformada de Fourier bidimensional

La transformada de Fourier bidimensional permite la conversión de la imagen en el dominio espacial en una representación en el dominio frecuencial, donde la frecuencia cero se encuentra centrada en la imagen y la distancia de cualquier punto al centro es proporcional a la frecuencia. Esta representación es la suma de exponenciales complejas de magnitudes, frecuencias y fases variables.

Una línea en el dominio espacial está compuesta por una amplia gama de componentes frecuenciales. Se puede observar que la distribución de las frecuencias se encuentran alineadas y su inclinación está relacionada con el ángulo de inclinación de la línea en la imagen espacial, independiente de su posición (Fig. 2). Específicamente la inclinación en el campo frecuencial es perpendicular a la línea en la imagen espacial.

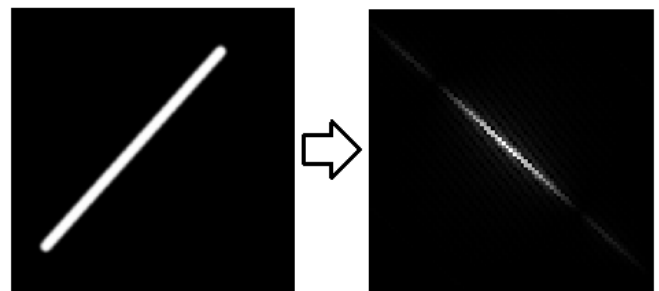


Fig. 2. Imagen espacial de una línea y su correspondiente transformada de Fourier

En el caso particular de las fibras de colágeno, el conjunto de fibras nunca se encuentra perfectamente alineado, por lo que la imagen frecuencial no forma una línea, sino que la distribución de frecuencias adopta una forma elíptica. En la Fig. 3 se pueden observar las fibras de colágeno de una muestra de tejido ovárico normal, de la cual se le obtiene la TF directamente y anteponiendo una función ventana. Conforme las fibras presenten una estructura desorganizada, es decir, exhiba patrones aleatorios, la distribución de frecuencias adoptará una forma circular. En la Fig. 4 se pueden observar las fibras de colágeno de una muestra de tejido ovárico patológico, de la

cual se le obtiene la TF.

La Fig. 3 y la Fig. 4 representan dos casos extremos donde las fibras de colágenos evidencian notables diferencias morfológicas. Esta diferencia puede ser imperceptible entre muestras de tejido normal y tejido patológico, o entre diferentes tipos de tejido patológico.

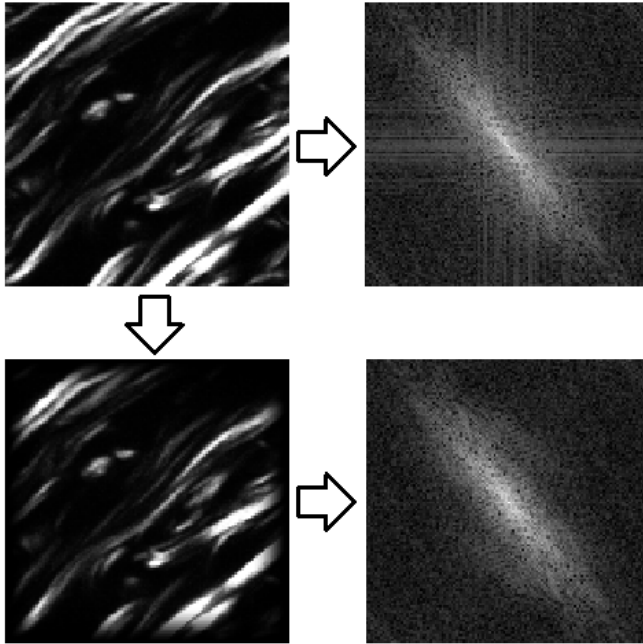


Fig. 3. Imagen de las fibras de colágeno de un tejido ovárico normal y su correspondiente transformada de Fourier. Comparación de las imágenes transformadas sin y con la multiplicación por la ventana de Tukey.

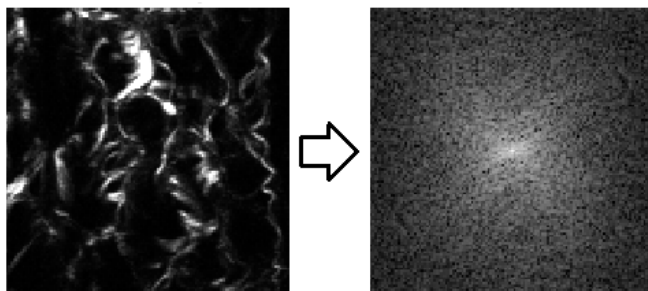


Fig. 4. Imagen de las fibras de colágeno de un tejido ovárico patológico y su correspondiente transformada de Fourier

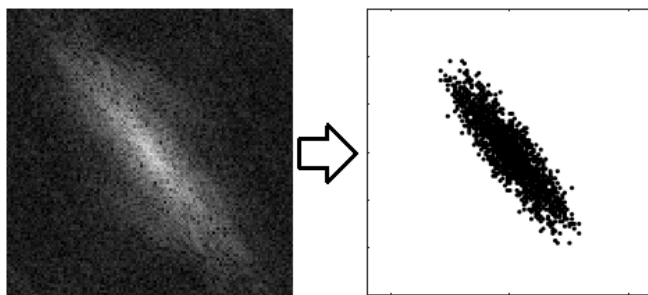


Fig. 5. Binarización del módulo de la transformada de Fourier.

C. Binarización

La distribución de las componentes frecuenciales en la TF adopta una forma elíptica que es extraída mediante la definición de un umbral que permite separar las componentes más relevantes respecto del fondo. La selección de este umbral es un factor importante para obtener correctamente la forma generada. Se utilizó en el presente trabajo, un umbral adaptativo definido previamente por Zeitoune *et al* [?].

La Fig. 5 muestra la binarización del módulo de la TF. Posteriormente la imagen fue convertida a un par ordenado de valores (x,y), formando una matriz $A \in \mathbb{R}^{m \times 2}$, donde m es la cantidad de elementos con valor 1.

D. Descomposición en valores singulares

La descomposición en valores singulares (SVD) de una matriz $A \in \mathbb{R}^{m \times n}$ es una factorización de la misma como el producto de matrices. Puede expresarse de la siguiente forma:

$$A = U \Sigma V^T \tag{1}$$

donde $U \in \mathbb{R}^{m \times m}$ y $V \in \mathbb{R}^{n \times n}$ son matrices ortogonales y $\Sigma \in \mathbb{R}^{m \times n}$ es una matriz diagonal que contiene los valores singulares $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r \geq 0, r = \min(m, n)$.

En este caso particular con $n=2$, dado que $m > n$, puede escribirse la versión reducida como $U \in \mathbb{R}^{m \times 2}$ y $\Sigma \in \mathbb{R}^{2 \times 2}$, entonces la matriz Σ sólo contendrá 2 valores singulares positivos no nulos.

Para la obtención de los dos parámetros (sección III-E) sólo se requiere calcular la matriz V y la matriz Σ . La primera está formada por los autovectores de $A^T A$, y la segunda por los valores singulares, que son la raíz cuadrada de los autovalores de $A^T A$.

$$V = [\mathbf{v}_1 \ \mathbf{v}_2] = \begin{bmatrix} v_{11} & v_{21} \\ v_{12} & v_{22} \end{bmatrix} \tag{2}$$

$$\Sigma = \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix} \tag{3}$$

E. Obtención de parámetros

Se obtendrán dos parámetros a partir de la descomposición en valores singulares de la matriz $A \in \mathbb{R}^{m \times 2}$: la orientación y organización de las fibras de colágeno. Como se verá posteriormente (sección IV-A), el autovector \mathbf{v}_1 representa la dirección del eje principal y el autovector \mathbf{v}_2 representa el eje secundario. Conociendo que la dirección de las fibras de colágeno son perpendiculares a la dirección del eje principal, al ser \mathbf{v}_1 y \mathbf{v}_2 ortogonales, \mathbf{v}_2 representará la dirección de las fibras de colágeno. Entonces, la orientación de las fibras de colágeno podrá calcularse como¹:

$$\theta = \arctg \left(\frac{v_{22}}{v_{21}} \right) \tag{4}$$

La medida de cualquier valor singular de la matriz Σ es indicativa de su importancia en la propagación de los datos en esa dirección. Específicamente, el valor singular es la desviación estándar de los datos en la dirección de cada

¹En un programa, el pixel (0,0) de la matriz que representa la imagen se encuentra en la esquina superior izquierda de la imagen, por lo que el resultado real es el valor negativo.

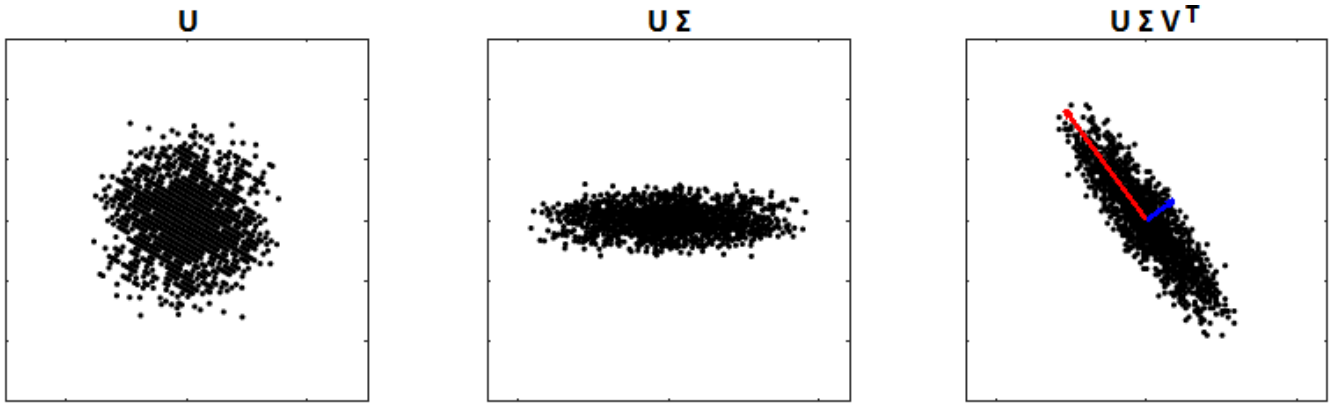


Fig. 6. Gráficas de la descomposición de la matriz A.

vector. El grado de organización de las fibras de colágeno será definido mediante la relación de aspecto (RA) como el cociente entre el valor singular en la dirección secundaria (σ_2) y el valor singular en la dirección primaria (σ_1):

$$RA = \frac{\sigma_2}{\sigma_1} \quad (5)$$

IV. RESULTADOS Y DISCUSIÓN

A. Interpretación geométrica

Dado que SVD es una reducción en matrices ortogonales, puede interpretarse que la matriz U es la proyección de la matriz A sobre los vectores ortogonales que forman la matriz S, multiplicada por un factor de dispersión Σ . Es decir, la matriz de datos U, es escalada y rotada para formar A. La matriz Σ es una matriz diagonal dada por la ecuación 3, por lo que σ_1 escala únicamente a la primera columna de U y σ_2 multiplica a la segunda columna de U. La matriz V tiene la estructura de una matriz de rotación en el espacio euclidiano²:

$$V = \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix} \quad (6)$$

Este análisis resulta evidente al observar la Fig. 6.

B. Distinción entre muestras de tejido normal y patológico

La Fig. 7 muestra la variabilidad en la organización de los tejidos. Se puede observar una diferencia significativa entre ambos tipos de tejido. La orientación de las fibras de colágeno adquiere importancia cuando estas son comparadas con el tejido epitelial de la capa adyacente. Es por ello, que se calculó la diferencia entre el ángulo del tejido epitelial y el ángulo principal de las fibras de colágeno. La Fig. 8 muestra una diferencia significativa entre ambos tipos de tejido. Principalmente el tejido patológico presenta mayor variabilidad.

V. CONCLUSIONES

En el presente trabajo se presenta un nuevo método para cuantificar dos parámetros importantes que caracterizan la distribución de las fibras colágenas, la organización y la orientación principal. La organización, se cuantificó mediante la relación de aspecto, en la que un valor cercano a cero, indica

²La posición del signo negativo se debe a que es una matriz que post-multiplica.

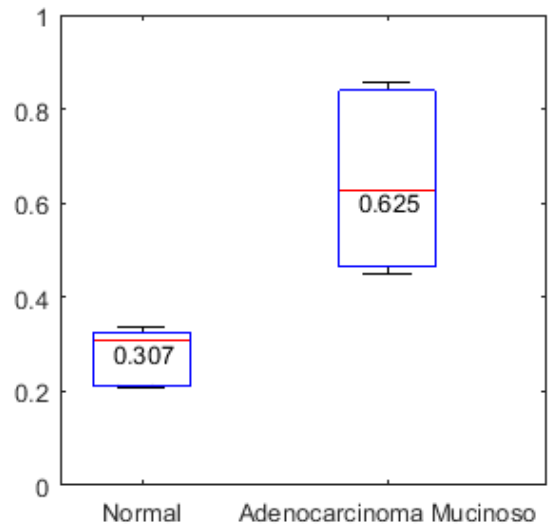


Fig. 7. Distribución de la relación de aspecto para las 5 muestras de tejido normal y 5 muestras de adenocarcinoma mucinoso

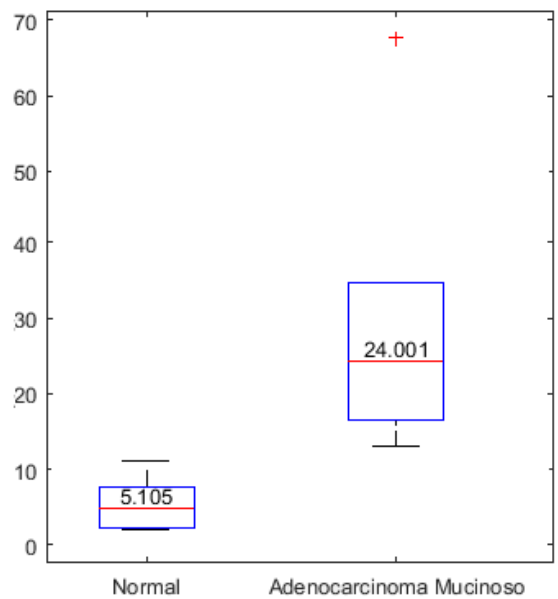


Fig. 8. Distribución de la diferencia entre el ángulo de inclinación de las fibras de colágeno respecto al tejido epitelial adyacente para las 5 muestras de tejido normal y 5 muestras de adenocarcinoma mucinoso

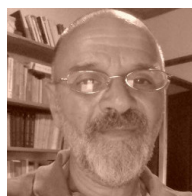
un alto paralelismo entre las fibras, mientras que un valor cercano a uno implica una organización aleatoria de las fibras de colágeno. Los resultados mostraron diferencia estadística entre los tejidos normales y los tejidos patológicos analizados, permitiendo establecer un umbral de separación. En función de estos resultados, se concluye que este parámetro posee un gran potencial como clasificador del estado de los tejidos en concordancia con trabajos anteriores.

La orientación principal de las fibras de colágeno indica la dirección en el cual el grupo mayoritario se encuentra orientado. Se calculó la diferencia entre este valor y el del tejido epitelial adyacente. Una diferencia entre 0° a 15° implica un alto paralelismo entre las fibras de colágeno y el tejido epitelial. Este valor junto a un bajo RA permitirían estimar un tejido normal.

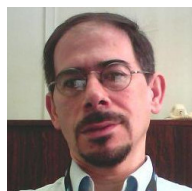
La descomposición en valores singulares se caracteriza por la simplicidad y la velocidad de cálculo. Permite reemplazar el ajuste manual de una elipse realizado sobre la imagen transformada por un cálculo matemático. Los autores pretenden extender este método a imágenes tridimensionales y aumentar el número de muestras en futuros trabajos.

REFERENCIAS

- [1] K. Lipnikov, G. Manzini, and M. Shashkov, "Mimetic finite difference method," *Journal of Computational Physics*, vol. 257, no. PB, pp. 1163–1227, 2014. [Online]. Available: <http://dx.doi.org/10.1016/j.jcp.2013.07.031>
- [2] M. Bonazzoli, V. Dolean, F. Rapetti, and P. H. Tournier, "Parallel preconditioners for high-order discretizations arising from full system modeling for brain microwave imaging," *International Journal of Numerical Modelling: Electronic Networks, Devices and Fields*, no. January, pp. 1–7, 2017.
- [3] N. Jiang, M. Kubacki, W. Layton, M. Moraiti, and H. Tran, "A Crank-Nicolson Leapfrog stabilization: Unconditional stability and two applications," *Journal of Computational and Applied Mathematics*, vol. 281, pp. 263–276, 2015. [Online]. Available: <http://dx.doi.org/10.1016/j.cam.2014.09.026>
- [4] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1*, ser. NIPS'12. USA: Curran Associates Inc., 2012, pp. 1097–1105. [Online]. Available: <http://dl.acm.org/citation.cfm?id=2999134.2999257>



Victor H. Casco se graduó de Licenciado en Biología en el año 1986, obtuvo su título de Dr. en Ciencias Biológicas en 1992. Actualmente es Profesor Titular Ordinario y Director del Laboratorio de Microscopia Aplicada a Estudios Moleculares y Celulares. Sus áreas de interés son Péptidos Vasoactivos, Moléculas de Adhesión Celular y Técnicas y Desarrollo de Sistemas de Microscopia.



Javier F. Adur es graduado en Bioingeniería (1997, UNER). Máster en Ingeniería Biomédica (2006-UNER) y Doctor en Ciencias Biológicas (2008-UNL). Desde el año 1997 es Docente/Investigador en la Facultad de Ingeniería-UNER; investigador adjunto del CONICET 2013; posee el cargo de Profesor titular Ordinario desde el 2017. Áreas de interés: Biofotónica. Microscopias lineales y no lineales. Su utilización como herramientas de diagnóstico precoz de enfermedades. Radiaciones No Ionizantes. Aplicaciones en diagnóstico y terapia.

Angel A. Zeitoune es Bioingeniero graduado de la Facultad de Ingeniería de la U.N.E.R. (2005). Es becario doctoral del CONICET realizando su investigación sobre la caracterización de las fibras de colágeno mediante procesamiento de imágenes microscópicas. Es miembro del grupo de Biofotónica y Procesamiento de Información Biológica (ByPIB) de la Facultad de Ingeniería de la UNER.



Luciana A. Erbes es Licenciada en Informática graduada de la Facultad de Ingeniería de la U.N.E.R. (2005). Es becaria doctoral del CONICET realizando su investigación sobre la cuantificación tridimensional aplicada al análisis morfológico y de autofluorescencia como potencial herramienta de diagnóstico óptico. Es miembro del grupo de Biofotónica y Procesamiento de Información Biológica (ByPIB) de la Facultad de Ingeniería de la UNER.

