

¿Por qué enseñar ciencia reproducible y cómo la revista IPOL puede ayudar a conseguirlo?

Nelson Monzón^a

^aCentre de Mathématiques et Leurs Applications (CMLA). École Normale Supérieure
Paris-Saclay. 61, Avenue du Président Wilson, 94235 Cachan cédex, París, Francia

ABSTRACT

En este trabajo, queremos destacar la importancia de inculcar ciencia reproducible al alumnado, especialmente en la formación de nuevos investigadores e investigadoras. La reproducibilidad es un requisito fundamental del método científico ya que permite confirmar que nuestros resultados se sustentan en evidencias contrastables. En este contexto, destacamos el potencial que tiene la revista *Image Processing On line* (IPOL) como recurso educativo gracias a su política editorial y su entorno de experimentación en línea dentro del campo del procesamiento de señales y las matemáticas.

Todas sus publicaciones sustentan sus aportaciones científicas con una descripción completa y precisa del algoritmo que implementa el método que exponen. Esta será validada durante el proceso de revisión por pares de la revista asegurando que coincide con el programa utilizado en los experimentos publicados. Dicho programa también será publicado y protegido por la correspondiente licencia. Además, cada artículo se complementa con un sistema de experimentación en línea que ejecuta el algoritmo publicado permitiendo a la comunidad científica realizar sus propias pruebas con datos y parámetros a su elección. Este sistema ofrece también un archivo completo y público de muchos de los experimentos realizados.

En virtud de esto, consideramos que IPOL puede ser un recurso valioso que ayuda al profesorado a enseñar matemáticas aplicadas de una manera sencilla, intuitiva y colaborativa (dado su entorno web y el archivo), convirtiendo las aulas en un entorno científico que favorezca la interacción, experimentación y el entendimiento del estudiantado mediante una metodología de aprendizaje constructivista.

Keywords: método científico, ciencia reproducible, matemáticas aplicadas, aprendizaje en red, entorno experimental.

1. INTRODUCCIÓN

1.1 ¿Por qué enseñar ciencia reproducible?

Enseñar investigación científica significa instruir al alumnado que, en su búsqueda del conocimiento, debe seguir un proceso reflexivo y sistemático que explore, entienda y demuestre sus teorías de una manera empírica, medible y veraz. Este proceso se conoce como método científico y nace de la formulación de hipótesis que se producen como consecuencia de la observación y estudio de fenómenos. Estas deben ser verificadas de forma fidedigna a partir de los resultados que se obtienen mediante un análisis experimental en entornos controlados. Podría decirse que “dónde no hay método científico, no hay ciencia”.¹

Estos resultados son sometidos a una confrontación rigurosa y escrutinio público cuando son publicados en conferencias y revistas científicas. Normalmente, una publicación describe, en mayor o menor medida, el proceso seguido en la recolección y análisis de datos de sus estudios. El problema es que, en demasiadas ocasiones, la información descrita no es suficiente como para garantizar que las conclusiones sean ciertas.² Esto dificulta que otros integrantes de la comunidad científica puedan reconstruir, bajo las mismas condiciones, los resultados que se proclaman en un artículo y pone en duda uno de los pilares que fundamentan el método científico: su reproducibilidad.

monzon@cmla.ens-cachan.fr, (928) 45 87 10

Por añadidura, un experimento que no se puede reproducir fielmente es también un experimento que no se puede refutar fácilmente ya que las únicas evidencias a analizar son aquellas que se han presentado en el estudio. Como consecuencia, sus propuestas no podrían considerarse como absolutamente verdaderas sino, en todo caso, como no refutadas. También debemos tener presente que, sin suficiente información del entorno experimental, se antoja complicado la repetibilidad del experimento con otros conjuntos de datos. Todas estas circunstancias limitan el impacto del estudio y pueden poner en entredicho la validez epistemológica del experimento.³

La comunidad académica, en el desempeño de la investigación, enseñanza y divulgación científica fundamenta sus actividades en principios de rigor, responsabilidad y honestidad.⁴ Por este motivo, consideramos que se debería formar al alumnado bajo la premisa de que su investigación debería ser reproducible y sus resultados basados en evidencias perfectamente contrastables. Un trabajo puede errar quizás en sus conclusiones pero, si es reproducible, será un trabajo honesto.⁵ Además, esta transparencia es una garantía de calidad y facilita la detección y corrección de errores durante la revisión previa a una publicación.⁶ Probablemente también ayude a su aceptación y reconocimiento en citas.^{7,8}

1.2 La iniciativa IPOL

La revista *Image Processing On Line* (IPOL) * promueve un entorno de difusión de ciencia abierta y reproducible en el campo del procesamiento de señales (imagen, vídeo, audio, 3D). Su objetivo es presentar trabajos matemáticos cuyos resultados, además de aportar valor científico, puedan ser confirmados con exactitud.

En el campo de las ciencias de la computación y de las matemáticas, muchos investigadores^{9,10} proclaman que una publicación científica debe necesariamente explicar en detalle los algoritmos que proponen. Además, el programa utilizado es una parte integral de la actividad científica y, por lo tanto, debería publicarse. Donoho *et al.*¹¹ evidenció que existe una crisis de credibilidad científica por estos motivos. Las causas son muchas incluidas las descripciones incompletas, la no publicación del código fuente o, simplemente, que el algoritmo publicado no se corresponde con el programa utilizado.

IPOL^{12,13} nace como respuesta a estas problemáticas proponiendo un sistema editorial que evalúa que sus publicaciones describan sus propuestas bajo el prisma de la reproducibilidad. Cada artículo publicado se compone de cuatro elementos: (i) un documento en PDF, (ii) un enlace para la descarga del programa utilizado, (iii) un sistema de demostración (demo) para su experimentación y (iv) un archivo de experimentos.

Al igual que en otras revistas, los artículos presentan un marco teórico, los experimentos realizados y toda la información que los autores consideren necesaria. La diferencia reside en que, en todos los casos, se espera una descripción completa de los detalles matemáticos y una explicación precisa del algoritmo, junto con su código fuente, mediante diagramas o una notación formal esquemática (pseudocódigo). En este sentido, el código debe ser transparente y no debe utilizar trucos de implementación a menos que estos se describan en el artículo. La aceptación de un artículo supone que durante la revisión por pares se ha verificado que el programa suministrado corresponde con su descripción.

Además, IPOL ofrece un sistema de demos en línea¹⁴ que permite que cualquier persona pueda realizar sus propios experimentos ejecutando el código fuente original. Así se puede experimentar con este de manera cómoda y sencilla desde cualquier navegador *web*. Cada demo propone unos parámetros por defecto que suelen ir acompañados de una breve explicación del efecto esperado acorde a sus valores. Estos parámetros pueden ser configurados antes de cada ejecución con el valor que prefieran utilizar. Por otro lado, ese experimento será archivado cada vez que se emplee datos propios. Esto ayuda a que numerosos investigadores e investigadoras puedan colaborar de forma remota.

Las demos en línea son una de las principales ventajas de la revista y ayudan a que, además de presentar estudios reproducibles, estos también sean refutables y repetibles. Esto se debe a que las demos permiten que otros integrantes de la comunidad científica puedan ejecutar con otros datos y distintas configuraciones de parámetros.

* <http://www.ipol.im/>

2. PROPUESTA DOCENTE

Un entorno de enseñanza que se limite a una recepción estática de información por parte de los alumnos y las alumnas producirá, en el mejor de los casos, un entendimiento parcial de los conceptos explicados, pero difícilmente se traducirá en la adquisición de habilidades que permitan una aplicación real de lo aprendido.¹⁵ Varios estudios^{16,17} destacan que, en el contexto de las matemáticas, el aprendizaje del alumnado debería apoyarse en actividades de experimentación. Con esto se consigue trabajar esta materia desde un enfoque aplicado y no se limita a una perspectiva teórica. Esta premisa debería ser primordial en lo que a formación de alumnado científico se refiere.

En este punto destacamos el interés de emplear un aprendizaje de tipo constructivista^{18,19} en este tipo de enseñanzas. Uno de sus objetivos principales es que durante el proceso de construcción (aprendizaje) de los conocimientos se enfoque en la resolución de tareas relevantes y útiles frente a problemáticas reales. A contar se aprende contando, a leer se aprende leyendo y a experimentar se aprende experimentando.

IPOLE ofrece un sistema de experimentación compartido y orientado a la investigación reproducible que puede ser usado como recurso educativo. Consideramos esto útil desde dos perspectivas. La primera nace de nuestra creencia de que la ciencia reproducible debería ser inculcada a toda persona que quiera introducirse en el mundo científico. En segundo lugar, sus demos en línea aportan un proceso interactivo, participativo y dinámico que enseña matemáticas aplicadas desde un punto de vista científico y riguroso y favorece un aprendizaje constructivista. Respecto a esto, las demos IPOLE plantean casos reales de aplicación de los métodos que exponen. Esto ayuda a que su aprendizaje surja a partir de planteamientos de problemas basados en situaciones de la vida real que ayuden a ir construyendo el conocimiento matemático.

Si bien es cierto que hay distintas posturas respecto a como enfocar el aprendizaje constructivista,²⁰ todas ellas coinciden en que es la persona que aprende quién debe ser la principal protagonista. En este sentido, uno de los puntos fuertes del sistema es que cada estudiante puede realizar sus propias ejecuciones con los parámetros y datos que desee construyendo su conocimiento matemático mediante la experimentación y análisis de resultados obtenidos al finalizar cada experimento. Esto puede favorecer también que su aprendizaje sea significativo²¹ ya que cada experiencia nueva modifica y reajusta los conocimientos previamente adquiridos.

En estos momentos, se cuenta con más de 150 demos publicadas en numerosos ámbitos de la visión por computador, principalmente imagen pero también vídeo y audio. Entre sus trabajos el alumnado puede experimentar con programas para suavizar una imagen aplicando una gaussiana * o matrices de convolución 3×3 **. Por otro lado, IPOLE como revista científica ha publicado algoritmos avanzados en diversos temas como, por ejemplo, modelos de color,²² flujo óptico,²³ segmentación,²⁴ calibración de cámaras²⁵ o un análisis detallado del método SURF,²⁶ entre otros muchos más.

En este trabajo mostraremos un ejemplo del funcionamiento de una demo de reducción de ruido como recurso educativo inspirados en la aplicación práctica que utiliza el Dr. Guillermo Sapiro, profesor de Ingeniería Eléctrica e Informática en la Universidad de Duke, instruyendo a sus estudiantes en procesamiento de imágenes y vídeo en el curso *Image and Video Processing: From Mars to Hollywood with a Stop at the Hospital* ***.

Sapiro utiliza una metodología de aprendizaje activo y experimentación guiada enseñando el funcionamiento de métodos numéricos a partir de su experimentación. El docente orienta la experiencia para que el proceso de aprendizaje se relacione con elementos que permitan al alumnado construir su propio conocimiento.

3. RESULTADOS EXPERIMENTALES Y DISCUSIÓN

A continuación mostraremos cómo se puede enseñar, mediante experiencias reales, al estudiantado que esté interesado en aprender técnicas de mejora en imágenes afectadas por ruido. El ruido es una variación aleatoria del brillo o el color en imágenes digitales que no se corresponde con la realidad de la escena y que empeora notablemente la calidad de esta. Uno de los métodos más relevantes en la literatura que investiga la reducción de ruido es “*Non-Local Means Denoising*”²⁷ cuya demo **** (y su algoritmo) está publicada en Buades *et al.*²⁸

* <https://ipolcore.ipol.im/demo/clientApp/demo.html?id=77777000037>

** <http://ipolcore.ipol.im/demo/clientApp/demo.html?id=77777000022>

*** <https://www.coursera.org/learn/image-processing>

**** http://demo.ipol.im/demo/bcm_non_local_means_denoising/

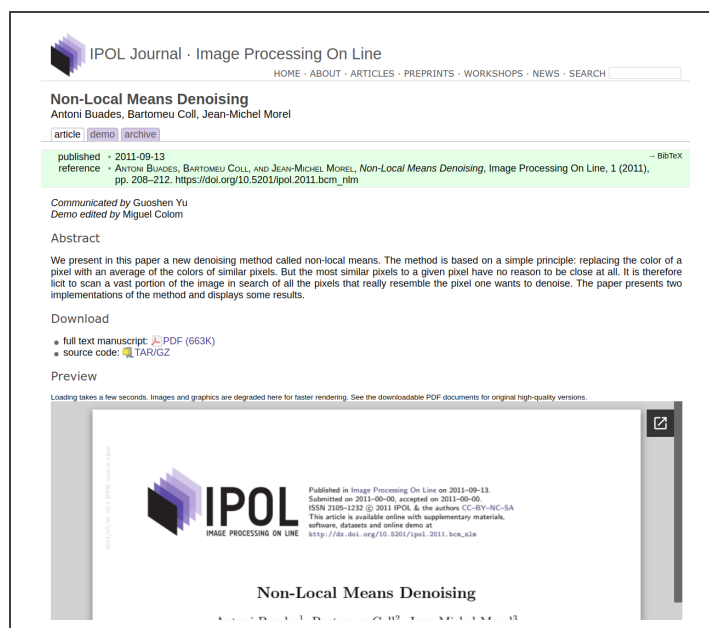


Figura 1. Página principal de la publicación “Non-Local Means Denoising”.

La figura 1 muestra la pantalla principal a la que se accede para iniciar la práctica. Aquí el alumnado puede ver información típica como sus autores y/o autoras, la fecha de publicación, el número de páginas o el resumen. Además, también pueden descargar el trabajo de investigación en diferentes resoluciones y el código fuente original que es exactamente el mismo que se está utilizando en la demostración de pruebas. Antes de usar la demo, recomendamos que el profesorado haya explicado los principales detalles y funcionamiento del método.

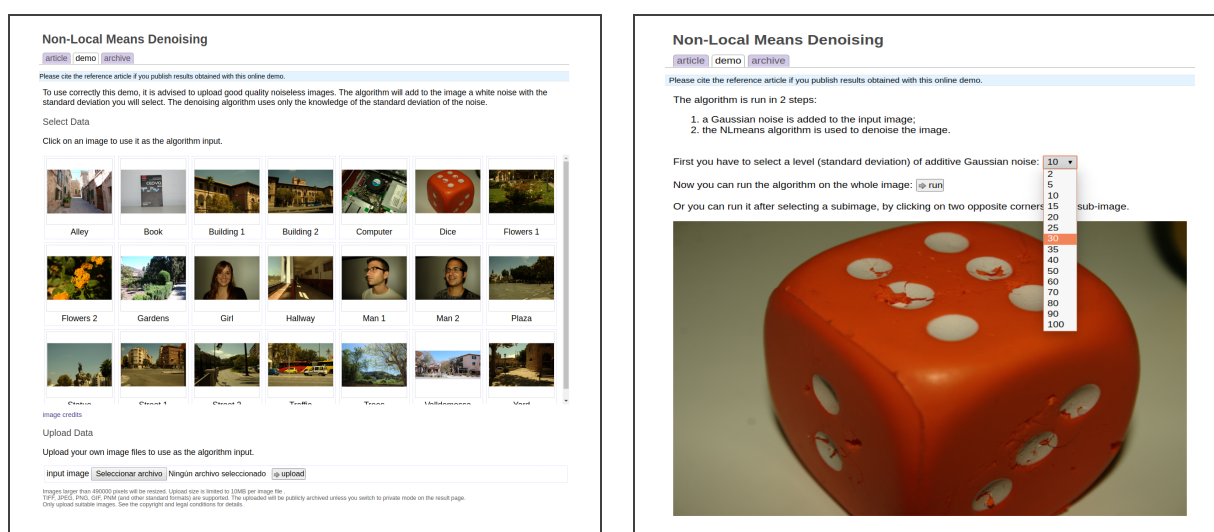


Figura 2. Demo asociada con la publicación “Non-Local Means Denoising”.

En la figura 2 se observa a la izquierda la interfaz principal de la demo dónde los alumnos y las alumnas pueden seleccionar entre imágenes estándar o propias (usando el botón de “subir archivo”) para sus experimentos. A la derecha se presenta la pantalla de selección de parámetros. En esta pantalla podemos ver también que la demo ofrece una breve explicación de los pasos necesarios para iniciar la ejecución. Una vez terminada, el alumnado debería explicar qué observa y aportar su propia interpretación en función de los conocimientos enseñados, así como su propia experiencia a partir del experimento.

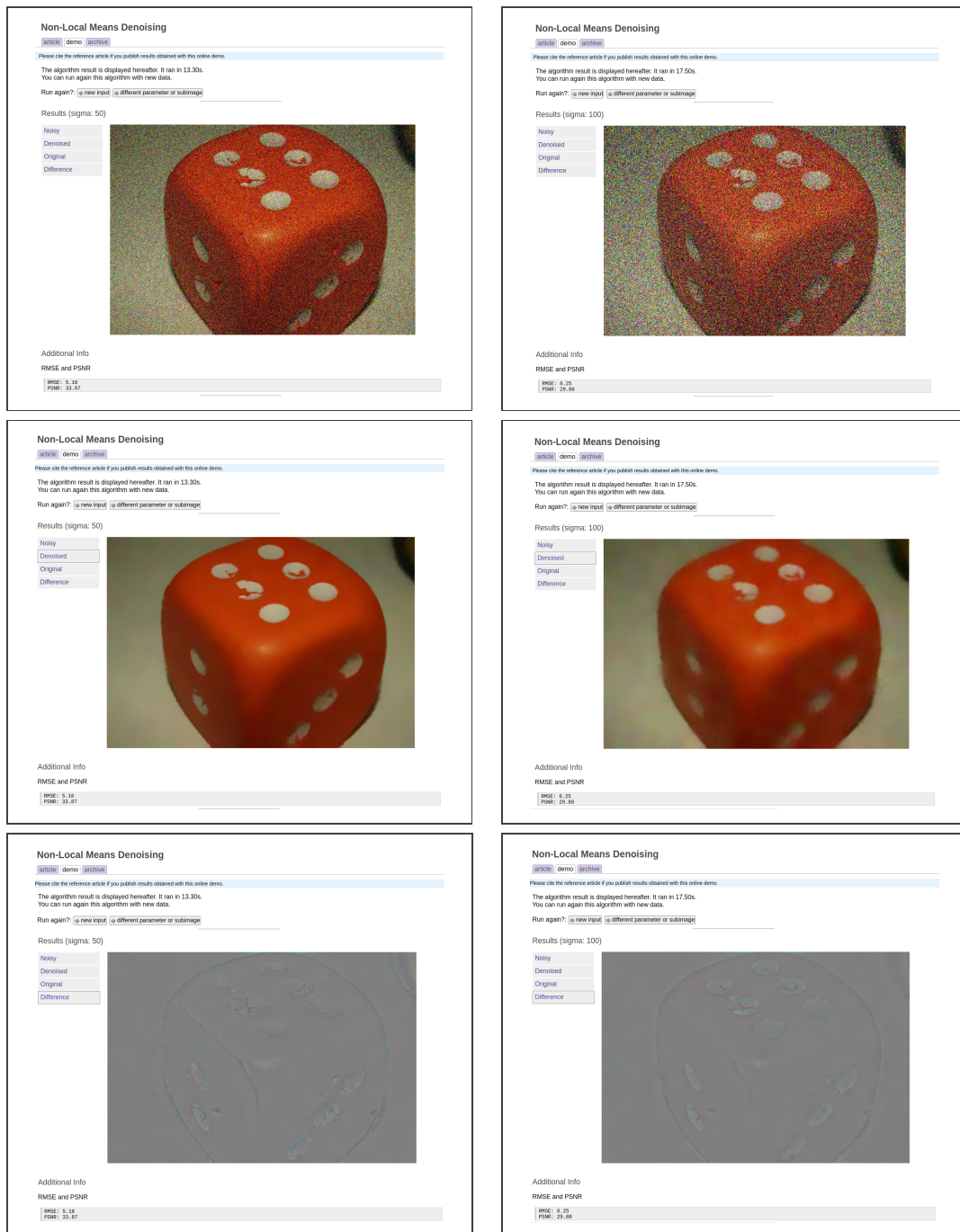


Figura 3. Resultados obtenidos para valores de ruido, $\sigma := 50, 100$. De arriba a abajo se observa la imagen con ruido, la resultante tras aplicar el método y la diferencia entre la original y la mejorada.

La figura 3 muestra los resultados que obtendría el estudiantado si usaran la imagen del dado y una distribución aleatoria de ruido de $\sigma := 50$ y $\sigma := 100$ (izquierda y derecha de la figura). De arriba abajo, vemos la imagen ruidosa, el resultado mejorado y la imagen diferencia entre la original y la mejorada. Los resultados muestran, en la parte inferior, el error calculado usando métricas estándar de reducción de ruido. Nótese que cada una de las imágenes de la figura pertenecen a la misma interfaz y se visualizan en la demo desplazándose por las pestañas de esta.

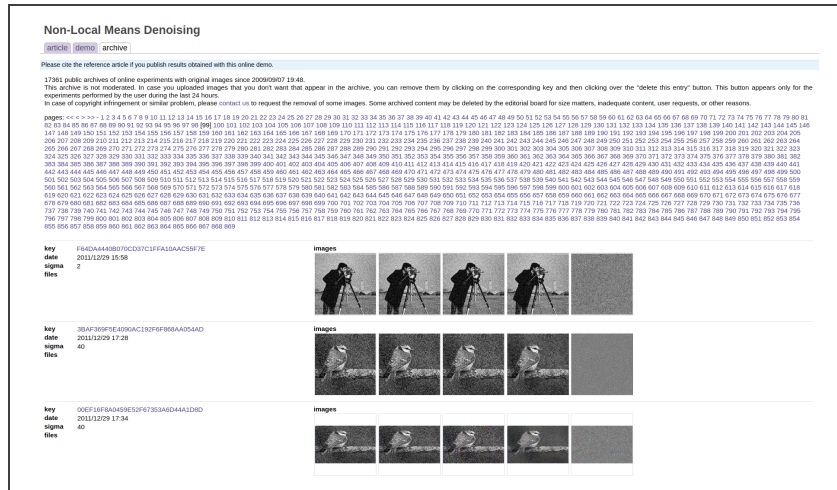


Figura 4. Historial de experimentos (archivo).

Recordemos que, si en vez de utilizar los datos que aparecen por defecto, se sube una imagen distinta mediante el botón de “subir archivo”, este experimento será almacenado en el histórico (ver figura 4). Esto permite al profesorado mostrar múltiples ejemplos con experimentos realizados por usuarios de todas partes del mundo. En esta figura se observa la enorme cantidad de ejemplos que se podrían explicar en clase. De esta manera, se puede garantizar la colaboración remota y la replicabilidad de los experimentos. Estos experimentos-ejemplos pueden ser analizados y contrastados debido a la accesibilidad de los datos que permite su replicabilidad.

Por último, nos gustaría recordar que, aunque el constructivismo puede tener distintas interpretaciones, todas ellas promueven la exploración libre de la persona que aprende en su búsqueda del conocimiento. En este sentido, recomendamos al profesorado que quiera hacer uso de IPOL que incluya en sus guías de actividades la exploración del sistema de archivo y que el alumnado utilice sus propias imágenes en algunos de sus experimentos.

4. CONCLUSIONES

Enseñar el manejo de herramientas y técnicas que faciliten la reproducibilidad de los datos es de gran ayuda en la formación de nuevos científicos y científicas, no sólo porque les será útil para garantizar que sus trabajos sean fiables, sino porque les facilitará abrir nuevas líneas de colaboración.

Es una oportunidad de mejorar sus aportaciones a la ciencia y facilitar su comprensión y reutilización para que otros miembros de la comunidad científica puedan aumentar su alcance. Probablemente, esto se traducirá en citas, reconocimiento y posibles co-autorías en publicaciones con otros grupos de investigación. Además, la publicación de código contrastado favorece su calidad y contribuye a acelerar el progreso científico.

En el campo del procesamiento de señales, la revista *Image Processing On Line* puede ser utilizada como recurso docente gracias a su sistema *web* de experimentación abierta y orientado a la investigación reproducible. Su sistema de demos favorece un aprendizaje constructivista del conocimiento por parte del alumnado ya que este se puede adquirir a partir de sus propias experiencias y experimentos. Hemos presentado dos ejemplos que permiten enseñar al alumnado las particularidades de un método de reducción de ruido y los resultados que se pueden obtener.

IPOL es una iniciativa en constante crecimiento e invita a la comunidad de docentes e investigadores a continuar publicando nuevos métodos y trabajos. Nuestro objetivo es que pueda ser de ayuda en la formación de nuevos científicos y científicas ya que

“La ciencia hace verdaderos progresos cuando una verdad nueva encuentra un ambiente preparado para acogerla.”

Piotr Alekséyevich Kropotkin

5. AGRADECIMIENTOS

Este trabajo ha sido parcialmente financiado por “Office of Naval research grant N00014-17-1-2552”. Me gustaría agradecer a los profesores Jean-Michel Morel y Luis Álvarez León por sus sugerencias y comentarios que han sido muy enriquecedores.

REFERENCES

- [1] Bunge, M., [*La ciencia, su método y su filosofía*], Buenos Aires. Ediciones Siglo XX (1987).
- [2] Ince, Darrel C. Hatton, L.-C. J., “The case for open computer programs,” *Nature*, 482–485 (2012).
- [3] Pigliucci, M., B. M., [*Philosophy of pseudoscience: reconsidering the demarcation problem*], University of Chicago Press, Chicago, IL, Estados Unidos (2013).
- [4] Bunge, M., [*La ética y la ciencia*], Buenos Aires. Ediciones Siglo XX (1972).
- [5] Peng, R., “Reproducible research in computational science,” *Science* (6060), 1226–1227 (2011).
- [6] Check-Hayden, E., “Rule rewrite aims to clean up scientific software.,” *Nature*, **520**, 276–277 (2015).
- [7] Piwowar, H. A., Day, R. S., and Fridsma, D. B., “Sharing detailed research data is associated with increased citation rate,” *PLOS ONE* (03 2007).
- [8] Vandewalle, P., “Code sharing is associated with research impact in image processing,” *Computing in Science Engineering*, **14** (2012).
- [9] Claerbout, J. F. and Karrenbach, M., [*Electronic documents give reproducible research a new meaning*], 601–604 (2005).
- [10] Stodden, V., “The legal framework for reproducible scientific research: Licensing and copyright,” *Computing in Science Engineering*, **11**, 35–40 (Jan 2009).
- [11] Donoho, D., Maleki, A., Rahman, I., Shahram, M., and Stodden, V., “Reproducible research in computational harmonic analysis,” *Computing in Science & Engineering*, **11** (1), 8–18 (2009).
- [12] Limare, N., *Reproducible research, software quality, online interfaces and publishing for image processing*, PhD thesis, École normale supérieure de Cachan-ENS Cachan (2012).
- [13] Colom, M., Kerautret, B., Limare, N., Monasse, P., and Morel, J.-M., “Ipol: a new journal for fully reproducible research; analysis of four years development,” in [*2015 7th International Conference on New Technologies, Mobility and Security (NTMS)*], 1–5, IEEE (2015).
- [14] Arévalo, M., Escobar, C., Monasse, P., Monzón, N., and Colom, M., “The ipol demo system: A scalable architecture of microservices for reproducible research,” in [*Reproducible Research in Pattern Recognition*], Kerautret, B., Colom, M., and Monasse, P., eds., 3–16, Springer International Publishing, Cham (2017).
- [15] López, G., “Pensamiento crítico en el aula,” *Docencia e Investigación*, 41–60 (Enero/Diciembre 2012).
- [16] Stenhouse, L., [*La investigación como base de la enseñanza*], Ediciones Morata. Colección Pedagogía. Madrid. (1987).
- [17] Hodson, D., “Hacia un enfoque más crítico del trabajo de laboratorio.,” *Enseñanza de las Ciencias*, **12** (1994).
- [18] Piaget, J., [*The child’s construction of reality*], Routledge and Kegan Paul, London (c1954, 1976 printing). Translation of *La construction du réel chez l’enfant*.
- [19] Vygotsky, L. S., [*Thinking and Speech*] (1934). Translation of *Myshlenie i rech’*: Psikhologicheskie issledovaniya.
- [20] Le Moigne, J. L., “From jean piaget to ernst von glasersfeld: An epistemological itinerary in review.,” *Constructivist Foundations*, **6** (2), 152–156.
- [21] Ausubel, D., [*The psychology of meaningful verbal learning*], Oxford, England: Grune & Stratton (1963).
- [22] Limare, N., Petro, A. B., Sbert, C., and Morel, J.-M., “Retinex Poisson Equation: a Model for Color Perception,” *Image Processing On Line*, **1**, 39–50 (2011).
- [23] Monzón, N., Salgado, A., and Sánchez, J., “Robust Discontinuity Preserving Optical Flow Methods,” *Image Processing On Line*, **6**, 165–182 (2016).
- [24] Grompone von Gioi, R., Jakubowicz, J., Morel, J.-M., and Randall, G., “LSD: a Line Segment Detector,” *Image Processing On Line*, **2**, 35–55 (2012).

- [25] Alvarez, L., Gomez, L., and Sendra, J. R., “Algebraic Lens Distortion Model Estimation,” *Image Processing On Line*, **1** , 1–10 (2010).
- [26] Oyallon, E. and Rabin, J., “An Analysis of the SURF Method,” *Image Processing On Line*, **5** , 176–218 (2015).
- [27] Buades, A., Coll, B., and Morel, J. ., “A non-local algorithm for image denoising,” in [*2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, **2**], 60–65 (June 2005).
- [28] Buades, A., Coll, B., and Morel, J.-M., “Non-Local Means Denoising,” *Image Processing On Line*, **1** , 208–212 (2011).