

Evaluación de técnicas de clasificación orientadas a la identificación automática de órganos del camarón a partir de imágenes histológicas

Angel Guale, Roberto Yoncon, Fanny A. Panchana, Bonny N. Bayot and Dennis G. Romero

Escuela Superior Politécnica del Litoral, ESPOL, Facultad de Ingeniería en Electricidad y Computación; Facultad de Ingeniería Marítima, Ciencias Biológicas, Oceánicas y Recursos Naturales, Campus Gustavo Galindo Km 30.5 Vía Perimetral, P.O. Box 09-01-5863, Guayaquil, Ecuador, {dgrmero,bbayot,panchanaf}@espol.edu.ec

Resumen– Detectar la presencia de enfermedades del camarón a partir de imágenes histológicas es un trabajo complejo que requiere la identificación inicial del órgano a analizar, debido a que enfermedades como el Síndrome de la Mancha blanca (WSS) o la Vibriosis se manifiestan en órganos específicos durante sus etapas iniciales. Este artículo presenta una evaluación de técnicas para la identificación automática de órganos del camarón basada en redes neuronales. Se presenta una comparación de resultados utilizando técnicas de clasificación basadas en extracción de características y redes neuronales convolucionales (CNN), obteniendo mejores resultados con CNN (93% de acierto considerando 7 órganos) usando Transfer Learning.

usadas en acuicultura, resaltando el uso de CNNs como un importante recurso a considerar para aplicaciones similares.

El artículo ha sido organizado de la siguiente manera: La sección II revisa distintas técnicas de visión por computador utilizadas para la identificación de puntos característicos, enfocado a la clasificación multiclase; la sección III presenta la metodología considerada en los experimentos; la sección IV muestra los resultados obtenidos y la correspondiente discusión. Finalmente, en la sección V se presentan las conclusiones del trabajo.

I. INTRODUCCIÓN

A nivel mundial, el cultivo de camarón es una de las principales actividades acuícolas de constante crecimiento. Sin embargo, las enfermedades virales y bacterianas son grandes limitantes de su producción, representando una amenaza para su sostenibilidad. La histología es una de las principales herramientas para el diagnóstico de enfermedades y es especialmente importante cuando no se dispone de herramientas moleculares para la detección específica de un patógeno. En un enfoque histológico se requiere identificar los órganos, antes de la observación de severidad de las lesiones provocadas por los agentes infecciosos. Diferenciar un órgano de otro en el camarón es una tarea compleja, debido a que las imágenes (capturadas por medio de un microscopio) presentan una apariencia altamente variable, lo cual dificulta mediante métodos clásicos de análisis y procesamiento de imágenes obtener un rendimiento aceptable de clasificación. Estudios anteriores han buscado categorizar automáticamente imágenes histológicas usando características basadas en *Scale-Invariant Features Transform* (SIFT) [1][2], entre otras [3][4][5].

Este trabajo evidencia la ventaja del uso de redes neuronales convolucionales (CNN) en relación a los métodos tradicionales de clasificación usados en acuicultura, identificando órganos de camarones a partir de imágenes de microscopio y realizando una comparación con otros métodos citados en la literatura. De esta forma, entre las principales contribuciones del presente trabajo se encuentra la comparación de una red neuronal convolucional (AlexNet) pre-entrenada con técnicas tradicionales de clasificación

Keywords: Acuicultura, Histología, Identificación de enfermedades, Redes neuronales convolucionales.

II. TRABAJOS RELACIONADOS

El uso de imágenes tomadas de microscópico ha incrementado el número de aplicaciones relevantes para la medicina y la bioquímica, haciendo del procesamiento digital de imágenes un enfoque necesario para el desarrollo de estos campos de estudio. Para obtener resultados acertados del análisis y procesamiento de imágenes es indispensable contar con datos de entrada que representen adecuadamente las categorías a los cuales pertenecen, además de un número suficiente de imágenes para el correcto entrenamiento de los clasificadores. A partir del pre-procesamiento de las imágenes histológicas, es necesario determinar las características más relevantes de cada clase, siendo este un problema general en reconocimiento de objetos y recuperación de imágenes, debido a que se requiere obtener representaciones discriminantes que favorezcan la clasificación. Varias técnicas han sido utilizadas para la clasificación de imágenes histológicas, obteniendo características relevantes de cada imagen, posterior a un proceso de selección, que busca determinar las características más significativas.

Entre las técnicas utilizadas en la literatura para la identificación de particularidades en imágenes se encuentran: *Scale-Invariant Features Transform* (SIFT), *Features from Accelerated Segment Test* (FAST), *Oriented FAST and Rotated BRIEF* (ORB), *Speeded Up Robust Features* (SURF), *Histogram of Oriented Gradient* (HOG) y *Local Binary*

Digital Object Identifier (DOI): <http://dx.doi.org/10.18687/LACCEI2017.1.1.011>
ISBN: 978-0-9993443-0-9
ISSN: 2414-6390

Patterns (LBP), esta última es utilizada generalmente para el análisis de texturas. El algoritmo de detección de esquinas FAST, ha ganado popularidad debido a su eficiencia en comparación con otros algoritmos usados en extracción de características como, por ejemplo, *Diferencia de Gaussianas* (DoG), usado por los detectores SIFT[6], SUSAN[7] y Harris[8].

Varios estudios mostraron que FAST tiene una respuesta rápida, útil para sistemas en tiempo real, aunque posee la debilidad de la mayoría de algoritmos de esta naturaleza: carece de eficacia en la detección de esquinas de objetos con un fondo de imagen poco uniforme [9].

La eficacia de SIFT en aplicaciones donde se requiere la detección de características visuales ha sido sobresaliente, incluyendo el reconocimiento de objetos [10], [11] y registro de imágenes [12]. No obstante, este algoritmo implica un gran costo computacional, en especial para sistemas de tiempo real o dispositivos de bajo poder computacional como teléfonos celulares, lo cual motivó el uso de algoritmos más eficientes, como SURF [13], el cual presenta un rendimiento apenas menor que SIFT, y otras alternativas como FAST y ORB, con menor costo computacional.

Un modelo interesante basado en SIFT ha ganado popularidad en los últimos 5 años, llamado *Bag-of-Words* (BoW), el cual ha atraído atención por parte de la comunidad de visión por computador y puede ser definido como una representación de histograma basada en características independientes [12]. Esta representación ha mostrado ser efectiva en diferentes escenarios de categorización de imágenes y tareas de reconstrucción [14], [15]. Algunas propiedades del modelo BoW son particularmente útiles para el análisis de imágenes histológicas y han sido aprovechadas exitosamente en aplicaciones que hacen uso de este tipo de imágenes [16][17][18]. Las imágenes histológicas tienen una estructura particular, con pocos colores y variada forma de sus bordes y texturas. Adicionalmente, el aumento utilizado por el microscopio puede incluir partes de otros órganos en una misma imagen, aumentando la complejidad del problema a resolver.

El uso de histogramas de características globales es presentado en [19] para la reconstrucción de una amplia variedad de imágenes médicas, mostrando que *bag of features* es también apropiado para este tipo de aplicación. Sin embargo, aunque estos y otros algoritmos existentes en la literatura pueden ser aplicados a imágenes de distinta naturaleza, los resultados podrían ser desfavorables dependiendo de la complejidad de las imágenes, aportando con características en patrones poco representativos [20].

Un enfoque que está ganando popularidad es el uso de las redes neurales convolucionales profundas. Estas redes neuronales artificiales han sido utilizadas en tareas como reconocimiento de dígitos [21], representación de imágenes [22], reconocimiento de objetos [23] y categorización de imágenes [24]. En un esquema de entrenamiento de estas

redes, conocido como *Deep Learning*, se disponen múltiples capas, ponderando automáticamente características representativas mediante operaciones de convolución, usando filtros no lineales [25]. Sin embargo, las DNN (*Deep Neural Networks*) se enfrentan a algunos inconvenientes, tales como el alto costo computacional [25] y la necesidad de un gigantesco conjunto de datos [25][26]. El entrenamiento de este tipo de redes requiere días, semanas o inclusive meses sobre CPU [26].

En la búsqueda de reducir el tiempo de procesamiento, el alto costo computacional ha sido manejado por herramientas que permiten ejecutar estos algoritmos de aprendizaje automático sobre los núcleos de una GPU, logrando resultados cientos de veces más rápido [26]. La necesidad de una gran cantidad de datos para entrenamiento es difícil de manejar con DNNs, ya que un conjunto pequeño de datos conlleva a problemas de resultados poco confiables, debido a la falta de ajuste de los datos al modelo [27]. Sin embargo, es posible utilizar un conjunto de datos limitado utilizando una técnica llamada *Transfer-Learning*, con la cual es posible utilizar una DNN (*Deep Neural Network*) pre-entrenada con otros datos de entrada y ajustar las últimas capas de la red utilizando nuevos datos. De forma complementaria, *Dropout* es una técnica que previene el *overfitting* y además provee una forma de combinar diferentes arquitecturas de redes neuronales eficientemente, ignorando ciertas conexiones dentro de la red [25][28]. Publicaciones recientes muestran el incremento de la popularidad del uso de DNNs para problemas de identificación en la acuicultura, una de ellas es la identificación de camarones con exoesqueleto blando (*soft shell shrimp*)[24].

III. METODOLOGÍA

A. Conjunto de datos

Se utilizó un conjunto de imágenes provistas por el Centro Nacional de Acuicultura e Investigaciones Marinas (CENAIM), las cuales se encuentran distribuidas respectivamente en 13 clases de órganos de camarón: branquias, corazón, músculo, hepatopáncreas, epitelio de estómago, epitelio de intestino, epitelio del cuerpo, órgano linfóide, tejido hematopoyético, cordón nervioso, glándula antenal, pleópodos y tejido conectivo del estómago, como se muestran en la Fig. 1. Estas imágenes histológicas fueron obtenidas en aumento microscópico de 40x. Las imágenes de órganos de camarón presentaron características muy variadas dentro de una misma clase, y a su vez, imágenes entre diferentes clases presentaron similitudes, lo cual dificultó el análisis de las mismas.

B. Clasificación usando Redes Neuronales Convolucionales

Se utilizó la técnica *Transfer Learning* para la identificación automática de imágenes histológicas de camarón, utilizando la conocida arquitectura *AlexNet* [27], la

cual se encuentra disponible en el framework Caffe Software Package, que pone a disposición un conjunto de herramientas

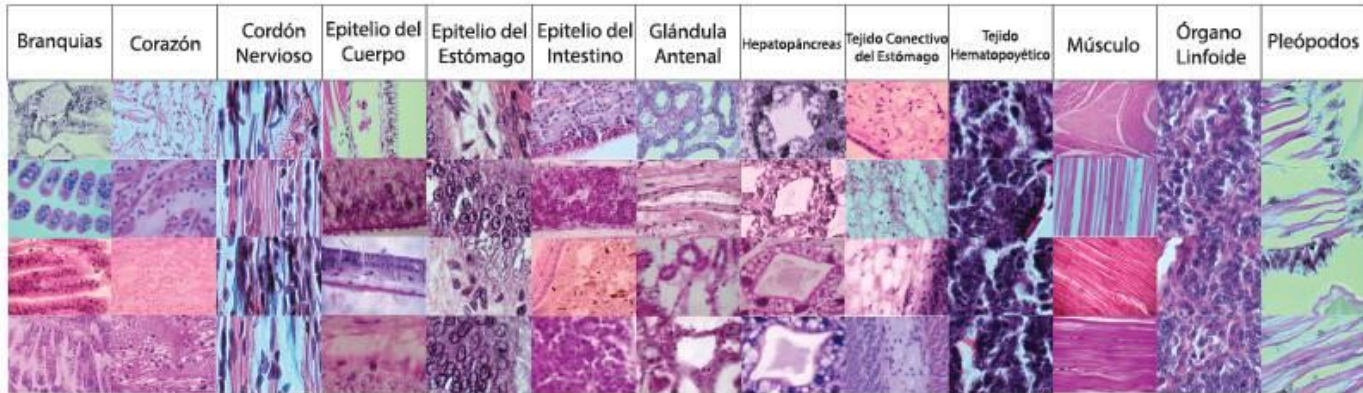


Fig. 1. Imágenes histológicas de los 13 órganos que forman parte del conjunto de datos. De derecha a izquierda se muestran imágenes histológicas de: branquias, corazón, cordón nervioso, epitelio del cuerpo, epitelio del estómago, epitelio del intestino, glándula antenal, hepatopáncreas, tejido conectivo del estómago, tejido hematopoyético, músculo, organo linfoide y pleópodos.

para entrenamiento, prueba, ajuste fino y despliegue de modelos con ejemplos bien documentados [29].

Específicamente, se utilizó la arquitectura *bvlc-reference-caffenet*, una réplica de *AlexNet*. La arquitectura de *AlexNet* posee ocho capas con sus respectivos parámetros de aprendizaje, cinco de ellas son capas convolucionales y las restantes son completamente conectadas (fully connected). Originalmente *AlexNet* fue diseñada para soportar 1000 clases, sin embargo, nuestro problema de clasificación de órganos de camarón considera 7 clases (las de mayor cantidad de imágenes), por lo que se adaptó la arquitectura *bvlc-reference-caffenet* a un número menor de salidas, siendo esto posible debido a la flexibilidad del *framework* para la modificación de la última capa de la red y en general, en cualquiera de ellas.

1) **Función de Activación:** La función de activación se encarga de convertir la entrada de una neurona en nuevos datos de salida, por lo que una función de activación con baja complejidad de cálculo agilizaría el proceso de entrenamiento de la red y, por esta razón redes como AlexNet utilizan una capa de neuronas llamada ReLu (*Rectified linear Units*) que hacen uso de la función de activación (1), logrando mejor rendimiento en relación con otras funciones de activación como *Sigmoid* y *Tanh*, también utilizadas en CNNs.

$$f(x) = \begin{cases} 0 & x \in (-\infty, 0) \\ x & x \in [0, +\infty) \end{cases} \quad (1)$$

2) **SoftMax Regression:** Se utilizó SoftMax Regression para asignar probabilidad a cada clase en la salida de la última capa de la red, debido a que ésta reduce el costo computacional al momento de entrenar la red neuronal, en

comparación con otras funciones de regresión logística. SoftMax se encarga de generar una distribución de probabilidad para las 7 clases.

3) **Data Augmentation:** Se empleó data augmentation sobre el conjunto de imágenes, con el objetivo de que la red pueda contar con más datos de entrada para su entrenamiento y pruebas. Debido a que las imágenes provistas por CENAIM poseen una alta resolución (1600 x 1600 píxeles, en promedio) se realizó cortes de 100 x 100 píxeles, y adicionalmente a cada corte se le aplicó transformaciones de rotación, con la finalidad de prevenir el overfitting.

4) **Transfer learning:** Dado que el proceso de obtención de imágenes histológicas de órganos del camarón requiere un extensivo trabajo manual, y por tanto el número de imágenes disponibles es limitado, es difícil contar con suficiente cantidad de datos como para entrenar una DNN con pesos aleatorios. Para sobreponer este inconveniente se utiliza el enfoque de Transfer-Learning, que consiste en utilizar una red pre-entrenada con otro conjunto de datos, de atributos similares. Se aplicó transfer-learning con el modelo pre-entrenado de AlexNet con la conocida base de datos ImageNet.

C. Clasificación usando características Bag of Words

Se utilizó una implementación de SIFT basada en el algoritmo propuesto por David Lowe en [6], en el enfoque BoW. En el proceso de *Bag of Words* se puede apreciar cuatro grandes etapas: detección de puntos característicos, obtención de los descriptores, la generación del diccionario de características [30] y finalmente la categorización [31]. De la localización de los puntos característicos y creación de los descriptores se encarga el algoritmo SIFT. Con los

descriptores se construye el diccionario de características, asignando una marca de clase a cada grupo de descriptores.

Finalmente, para la clasificación se empleó Support Vector Machine (SVM), usando la librería LIBSVM [32]. El algoritmo para la implementación de *Bag-of-Words* fue tomado de Minimal bag of visual Words classifier [33], una implementación escrita en Python.

1) Equipos utilizados: Tanto los modelos entrenados con características *Bag-of-Words* y CNN, fueron ejecutados en un computador virtual utilizando sólo procesamiento en CPU. Si bien es cierto, la ejecución en GPU es más rápida (y casi necesaria en estos casos), el tiempo de entrenamiento de los modelos no es considerado relevante para este estudio. Las especificaciones del equipo utilizado se detallan a continuación:

- Sistema Operativo: Ubuntu 16.04 LTS
- Procesador: Intel(R) Xeon(R) CPU E5-2673 v3 @2.4Ghz
- Memoria RAM: 55 GB
- Almacenamiento en disco duro: 512GB (SSD).

IV. RESULTADOS

Posterior al entrenamiento realizado con los detalles mencionados en la sección anterior, se obtuvo el modelo entrenado de cada clasificador, tanto para el enfoque de Bag-of-Words como CNN basada en AlexNet. Para cada uno de ellos, se realizó pruebas con un conjunto de aprox. 300 imágenes por clase: branquias, tejido conectivo, corazón, epitelio del cuerpo, glándula antenal, hepatopáncreas, músculo, obteniendo un archivo con los resultados de cada clasificador, por clase y globales. La Tabla 1 presenta los resultados obtenidos en ambos casos.

Para proveer una medida de la eficacia de clasificación, los resultados fueron calculados en función de un valor conocido en la literatura como medida-F (F-Measure), que combina precisión/precision (también llamado valor predictivo positivo) y recuperación/recall (conocida también como sensibilidad) y corresponde con la media armónica de ambos valores en términos de verdaderos positivos (Tp), Falsos positivos (Fp) y Falsos Negativos (Fn) [34].

TABLA I
RESULTADO DE LA CLASIFICACIÓN

| TIPO DE ÓRGANO | F (DNN) | F (SIFT-BOF) | TOTAL |
|---------------------|---------|--------------|-------|
| Tejido Conectivo | 0.934 | 0.693 | 305 |
| Corazón | 0.978 | 0.849 | 305 |
| Glándula Antenal | 0.995 | 0.478 | 295 |
| Hepatopáncreas | 0.934 | 0.706 | 284 |
| Epitelio del cuerpo | 0.997 | 0.885 | 291 |
| Músculo | 0.987 | 0.380 | 304 |
| Branquias | 0.993 | 0.864 | 295 |

El método de clasificación CNN obtuvo una considerable tasa de acierto de clasificación (93%) con un relativamente corto tiempo de entrenamiento. Se observó que el clasificador CNN mostró un valor de eficacia por clase mayor que el clasificador SVM usando SIFT-BOF, para imágenes histológicas de las clases evaluadas. Como se puede apreciar en la Fig. 2, el clasificador CNN alcanza una ventaja amplia sobre SVM con características SIFT-BOF en dos de las siete clases: glándula antenal y músculo, mientras que la eficacia posee una diferencia menor en las clases: branquias y epitelio del cuerpo.

Se había mencionado que la dificultad de trabajar con imágenes histológicas es la marcada similitud entre distintas clases, esto se evidenció en el clasificador SVM, el cual clasificó cerca de 200 imágenes de tejido conectivo como glándula antenal a partir de las características obtenidas con bag-of-words. De igual manera, considerando todas las clases, se evidencia un menor rendimiento de SVM considerando descriptores SIFT, las imágenes de glándula antenal albergan similitudes con las otras clases. Entre las clases con mayor número de aciertos está la glándula antenal y el epitelio del cuerpo, sin embargo, el valor de eficacia se vio afectado en glándula antenal, ante la gran cantidad de falsos positivos.

Usando el clasificador CNN, 18 imágenes de tejido conectivo fueron identificadas como hepatopáncreas y 14 imágenes de hepatopáncreas fueron clasificadas como tejido conectivo, lo cual sugiere una similitud existente entre estas dos clases bajo el criterio de los filtros convolucionales. Adicionalmente, se encontró similitud menor entre las glándulas y las branquias. Cinco de las clases evaluadas fueron clasificadas con más del 97% de acierto, mientras que las dos restantes, alcanzaron cerca del 93% de eficacia en clasificación. Estos resultados sugieren que, para aplicaciones en acuicultura el uso de redes neuronales convolucionales representan un valioso recurso a considerar en la identificación de particularidades de imágenes histológicas.

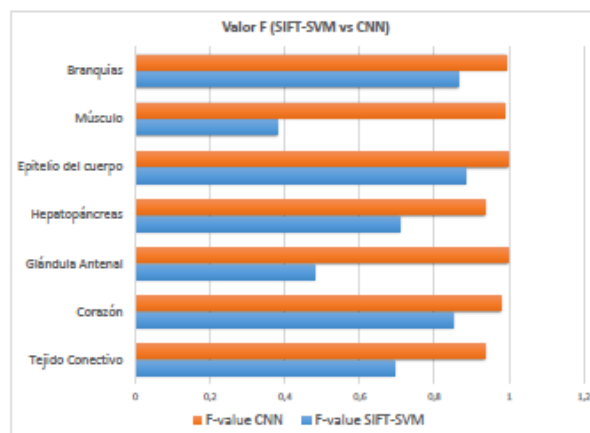


Fig. 2. Eficacia determinada utilizando medida-F en una prueba para cada clasificador con 300 imágenes histológicas

de 7 clases de órganos de camarón: branquias, corazón, epitelio del cuerpo, glándula antenal, hepatopáncreas, tejido conectivo del estómago, músculo.

V. CONCLUSIONES

Un estudio comparativo fue realizado para evaluar el rendimiento de dos enfoques de clasificación, orientado a la identificación automática de órganos del camarón a partir de imágenes histológicas. Dos métodos de clasificación fueron implementados, el primero utilizando técnicas basadas en descriptores de puntos característicos y máquina de vectores de soporte y, el segundo, utilizando una red neuronal convolucional con arquitectura *AlexNet*, considerando un modelo pre-entrenado. Se presentó el rendimiento obtenido por cada clasificador basado en el criterio de medida-F.

Como futuro trabajo se pretende analizar una mayor cantidad de imágenes, lo cual permitirá mostrar un resultado global de clasificación considerando un mayor número de órganos. Sin embargo, los resultados obtenidos a partir de 7 órganos de camarones, sugieren el uso de redes neuronales convolucionales para la identificación de imágenes histológicas en lugar de los métodos tradicionales de clasificación, motivando en la acuicultura el uso de CNNs como herramientas para la identificación de particularidades en imágenes.

AGRADECIMIENTOS

Este trabajo ha sido financiado por la ESPOL, mediante el Proyecto de Investigación “Reconocimiento de patrones: casos de estudio en la agricultura y acuicultura” M1-D1-2015. Modalidad de titulación proyecto de investigación, tema: “Taxonomía de técnicas de reconocimiento de patrones y sus aplicaciones”. Los autores desean expresar su agradecimiento al Centro de Investigación, Desarrollo e Innovación de Sistemas computacionales (CIDIS) y al Centro Nacional de Acuicultura e Investigaciones Marinas (CENAIM).

REFERENCIAS

- [1] M. M. M. Fouad, H. M. Zawbaa, N. El-Bendary, and A. E. Hassani, “Automatic Nile tilapia fish classification approach using machine learning techniques,” in *Hybrid Intelligent Systems (HIS)*, 2013 13th International Conference on. IEEE, 2013, pp. 173–178.
- [2] D. A. Lisin, M. A. Mattar, M. B. Blaschko, E. G. Learned-Miller, and M. C. Benfield, “Combining local and global image features for object class recognition,” in 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05)-Workshops. IEEE, 2005, pp. 47–47.
- [3] J. Lines, R. Tillett, L. Ross, D. Chan, S. Hockaday, and N. McFarlane, “An automatic image-based system for estimating the mass of freeswimming fish,” *Computers and Electronics in Agriculture*, vol. 31, no. 2, pp. 151–168, 2001.
- [4] M. A. Bueno-Ibarra, M. C. Chávez-Sánchez, and J. Álvarez-Borrego, “Nonlinear law spectral technique to analyze white spot syndrome virus infection,” *International Journal on Advances in Life Sciences Volume 2, Number 3 & 4*, 2010, 2010.

- [5] J. C. Caicedo and E. Izquierdo, “Combining low-level features for improved classification and retrieval of histology images,” *T. Mass-Data Analysis of Images and Signals*, vol. 2, no. 1, pp. 68–82, 2010.
- [6] D. G. Lowe, “Distinctive image features from scale-invariant keypoints,” *International journal of computer vision*, vol. 60, no. 2, pp. 91–110, 2004.
- [7] S. M. Smith and J. M. Brady, “Susan—a new approach to low level image processing,” *International journal of computer vision*, vol. 23, no. 1, pp. 45–78, 1997.
- [8] K. Mikolajczyk and C. Schmid, “An affine invariant interest point detector,” in *European conference on computer vision*. Springer, 2002, pp. 128–142.
- [9] M. Guerrero, “A comparative study of three image matching algorithms: Sift, surf, and fast,” 2011.
- [10] A. Suga, K. Fukuda, T. Takiguchi, and Y. Ariki, “Object recognition and segmentation using sift and graph cuts,” in *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*. IEEE, 2008, pp. 1–4.
- [11] F. A. Pavel, Z. Wang, and D. D. Feng, “Reliable object recognition using sift features,” in *Multimedia Signal Processing, 2009. MMSp’09. IEEE International Workshop on*. IEEE, 2009, pp. 1–6.
- [12] J. C. Caicedo, A. Cruz, and F. A. Gonzalez, “Histopathology image classification using bag of features and kernel functions,” in *Conference on Artificial Intelligence in Medicine in Europe*. Springer, 2009, pp. 126–135.
- [13] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, “Orb: An efficient alternative to sift or surf,” in 2011 International conference on computer vision. IEEE, 2011, pp. 2564–2571.
- [14] J. Yang, Y.-G. Jiang, A. G. Hauptmann, and C.-W. Ngo, “Evaluating bag-of-visual-words representations in scene classification,” in *Proceedings of the international workshop on Workshop on multimedia information retrieval*. ACM, 2007, pp. 197–206.
- [15] A. Bosch, X. Muñoz, and R. Martí, “Which is the best way to organize/classify images by content?” *Image and vision computing*, vol. 25, no. 6, pp. 778–791, 2007.
- [16] A. Cruz-Roa, J. C. Caicedo, and F. A. González, “Visual pattern mining in histology image collections using bag of features,” *Artificial intelligence in medicine*, vol. 52, no. 2, pp. 91–106, 2011.
- [17] T. Tommasi, F. Orabona, and B. Caputo, “Clef2007 image annotation task: An svm-based cue integration approach,” in *Proceedings of ImageCLEF 2007-LNCS*, no. LIDIAP-CONF-2007-041, 2007.
- [18] D. K. Iakovidis, N. Pelekis, E. E. Kotsifakos, I. Kopanakis, H. Karanikas, and Y. Theodoridis, “A pattern similarity scheme for medical image retrieval,” *IEEE Transactions on Information Technology in Biomedicine*, vol. 13, no. 4, pp. 442–450, 2009.
- [19] T. Deselaers, D. Keysers, and H. Ney, “Fire-flexible image retrieval engine: Imageclef 2004 evaluation,” in *Workshop of the Cross-Language Evaluation Forum for European Languages*. Springer, 2004, pp. 688–698.
- [20] Y. J. Zhang, “A survey on evaluation methods for image segmentation,” *Pattern recognition*, vol. 29, no. 8, pp. 1335–1346, 1996.
- [21] P. Y. Simard, D. Steinkraus, and J. C. Platt, “Best practices for convolutional neural networks applied to visual document analysis,” in *ICDAR*, vol. 3, 2003, pp. 958–962.
- [22] S. Behnke, *Hierarchical neural networks for image interpretation*. Springer Science & Business Media, 2003, vol. 2766.
- [23] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.
- [24] Z. Liu, F. Cheng, and W. Zhang, “Identification of soft shell shrimp based on deep learning,” in 2016 ASABE Annual International Meeting. American Society of Agricultural and Biological Engineers, 2016, p. 1.
- [25] N. Srivastava, G. E. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, “Dropout: a simple way to prevent neural networks from overfitting,” *Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929–1958, 2014.
- [26] D. Ciregan, U. Meier, and J. Schmidhuber, “Multi-column deep neural networks for image classification,” in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2012, pp. 3642–3649.

- [27] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.
- [28] N. Srivastava, "Improving neural networks with dropout," Ph.D. dissertation, University of Toronto, 2013.
- [29] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, "Caffe: Convolutional architecture for fast feature embedding," in *Proceedings of the 22nd ACM international conference on Multimedia*. ACM, 2014, pp. 675–678.
- [30] L. Fei-Fei and P. Perona, "A bayesian hierarchical model for learning natural scene categories," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 2. IEEE, 2005, pp. 524–531.
- [31] G. Csurka, C. Dance, L. Fan, J. Willamowski, and C. Bray, "Visual categorization with bags of keypoints," in *Workshop on statistical learning in computer vision, ECCV*, vol. 1, no. 1-22. Prague, 2004, pp. 1-2.
- [32] C.-C. Chang and C.-J. Lin, "Libsvm: a library for support vector machines," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 2, no. 3, p. 27, 2011.
- [33] L. S. Hackenberg. (2014) Minimal bag of visual words image classifier. [Online]. Available: <https://github.com/shackenberg/Minimal-Bag-of-Visual-WordsImage-Classifier>
- [34] D. M. Powers, "Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation," 2011.