2022 Lleida 27·1 junio·juny julio·juliol Cataluña Catalunya

8º CONGRESO FORESTAL ESPAÑOL

La **Ciencia forestal** y su contribución a los **Objetivos de Desarrollo Sostenible**

8CFE - 1037

Edita: Sociedad Española de Ciencias Forestales
Cataluña | Catalunya · 27 junio | juny - 1 julio | juliol 2022
ISBN 978-84-941695-6-4

Organiza

Sociedad Española de Ciencias Forestales SECF

Detección y segmentación automática de pilas de madera mediante Redes Neuronales Convolucionales y Visión Artificial

GARCÍA-PASCUAL, B.[1], ACUNA, M.[2]

[1] föra forest technologies, SLL.
[2] University of the Sunshine Coast, Australia.

Resumen

Una parte importante del coste de las explotaciones forestales madereras se deriva del método de cubicación empleado. En este sentido, la cubicación manual de madera apilada resulta ineficiente e imprecisa, siendo sus alternativas demasiado costosas. Por ello, han surgido tecnologías basadas en sensores ópticos de bajo coste que aplican algoritmos de Visión Artificial e Inteligencia Artificial para obtener estimaciones del volumen de madera.
En esta investigación, hemos aplicado una Red Neuronal Convolucional (CNN) para detectar y segmentar testas de *Pinus radiata* D. Don cargadas sobre camiones. Para ello, hemos entrenado el algoritmo Mask R-CNN usando una base de datos de 135 imágenes de cargamentos de madera (5.381 trozas) tomadas con orientación, iluminación y resolución variables. Estas imágenes se procesaron con el fin de incrementar la cantidad de datos disponibles de 135 a 418 imágenes, utilizándose el 60% de estas para entrenar el modelo, y el 40% restante para validarlo. Nuestros resultados preliminares muestran que el modelo ha detectado más del 95% de las testas con un error en la estimación de su superficie inferior al 3.6%.

Palabras clave

Automatización de procesos, planificación de operaciones, redes neuronales convolucionales, aprendizaje profundo.

1. Introduction

La estimación del volumen de madera apilada es un proceso forestal clave que tradicionalmente se ha realizado mediante medición manual, lo cual resulta ineficiente y altamente dependiente de la experiencia del operario. Por otra parte, es común emplear sensores laser para escanear las trozas una a una, obteniendo así mediciones más precisas, pero a costa de un tiempo de procesado que puede resultar excesivo cuando el número de trozas a medir es elevado (JANÁK, 2005; 2007; KNYAZ & MAKSIMOV, 2014).

Por otro lado, el auge de la Visión Artificial ha permitido el desarrollo de nuevas técnicas de cubicación basadas en el análisis de imágenes, las cuales se pueden dividir según el enfoque que toman en reconstrucción 3D de todo el cargamento, por un lado, y en detección y segmentación de las testas en imágenes 2D por otro.

Respecto a los modelos 3D, estos se generan empleando un tipo de algoritmo denominado *Structure from Motion* (SfM), el cual permite encontrar puntos comunes entre imágenes solapadas para generar una nube de puntos y reconstruir el objeto de estudio, obteniendo así una estimación de su volumen (SCHONBERGER & FRAHM, 2016). Si bien este enfoque ofrece resultados muy precisos, equiparables a los obtenidos mediante sensores láser, el coste computacional de generar los modelos 3D suele ser demasiado elevado (ACUNA & SOSA, 2019). Sin embargo, cabe destacar que algunos autores han

conseguido reducir dicho coste hasta el punto de ejecutar estos algoritmos en dispositivos móviles (HERBON et al., 2015).

En cuanto a la detección de objetos, estos métodos utilizan algoritmos menos exigentes para analizar las imágenes y detectar y segmentar los objetos contenidos en ellas (SOLEM, 2012), lo que posibilita la cubicación de pilas de madera si la longitud de las trozas es conocida. A pesar de que estas técnicas son menos precisas que las basadas en SfM, permiten realizar estimaciones del volumen de madera en condiciones operativas de manera rápida, lo que ha dado lugar a la emergencia de numerosas aplicaciones para la cubicación de pilas de madera (KÄRHÄ et al., 2019).

En los últimos años, las Redes Neuronales Convolucionales (CNN) (LECUN, 1989) se han convertido en los algoritmos predominantes en el campo de la Visión Artificial, particularmente en la detección y segmentación de objetos (ZHIQIANG & JUN, 2017; ALOYSIUS & GEETHA, 2018; KATTENBORN et al., 2021). A pesar de esto, solo unos pocos estudios se han centrado en la viabilidad de las CNN para detectar y segmentar las testas de trozas apiladas. Uno de estos estudios fue llevado a cabo SAMDANGDECH y PHIPHOBMONGKOL (2018), quienes usaron un método CNN multietapa para segmentar y detectar las testas de trozas remolcadas por camiones de carga. En primer lugar, detectaron el cargamento completo mediante un modelo CNN de detección de objetos, tras lo cual aplicaron un modelo CNN de segmentación semántica sobre las testas. Puesto que estos últimos no distinguen entre objetos de la misma clase, aplicaron operaciones morfológicas (MO) y etiquetado de elementos conectados (CCL) para separar cada una de las testas del cargamento.

No obstante, los modelos CNN de segmentación por instancias no poseen estas limitaciones, siendo capaces de separar objetos de la misma clase durante la segmentación (HARIHARAN et al., 2014). Un modelo de este tipo que ha ganado gran popularidad es Mask R-CNN (HE et al., 2020), el cual se ha utilizado en un amplio abanico de aplicaciones agroforestales. Algunos ejemplos incluyen: estimación de la biomasa en olivares mediante la toma de imágenes multiespectrales con drones (SAFONOVA et al., 2021), segmentación de copas mediante imágenes aéreas y satelitales (BRAGA et al., 2020) y detección automática de defectos en chapas de madera (LI et al., 2021). Resulta de especial interés el método desarrollado por WIMMER et al. (2021), ya que utilizaron Mask R-CNN para segmentar e identificar testas individuales a lo largo de toda la cadena de transporte. Con esto, demostraron que este modelo es capaz de detectar y segmentar trozas individuales de manera precisa sin necesidad de incluir pasos adicionales en el algoritmo.

2. Objetivos

En base a lo expuesto anteriormente, se han establecido los siguientes objetivos para el presente estudio:

1. Entrenar el modelo Mask R-CNN para que segmente testas de manera individual a partir de imágenes de trozas apiladas
2. Evaluar la bondad del modelo mediante el cálculo de una serie de métricas

Con esto, se pretende probar la hipótesis de que las CNN y, en particular, el modelo Mask R-CNN, pueden emplearse para detectar y segmentar de manera precisa las testas de trozas contenidas en cargamentos de madera.

3. Metodología

3.1. Modelo empleado

En el presente estudio se ha empleado Mask R-CNN (HE et al., 2020), un modelo CNN de segmentación por instancias especializado en detectar y segmentar de manera individualizada cada uno de los objetos contenidos en una imagen. Para ello, el modelo se divide en dos submodelos: el esqueleto y la cabeza. El primero extrae la información relevante de las imágenes, mientras que el segundo analiza esta información para proponer una serie de Regiones de Interés (RoI). Posteriormente, el modelo asigna una puntuación para estos RoIs según la probabilidad de que contengan un objeto, además de las coordenadas de un cuadro delimitador que encierra el objeto detectado, la clase a la que pertenece y una máscara que cubre su superficie (ver Figura 1). Estas predicciones se comparan entonces con la verdad terreno para calcular la función pérdida o coste siguiendo la Ecuación 1, que sirve para optimizar los parámetros del modelo (HE et al., 2020).

$$L = L_{cls} + L_{box} + L_{mask} \qquad \text{Ecuación 1}$$

Donde L es la pérdida total, L_{cls} es el error cometido durante la predicción de la clase a la que pertenece el objeto, L_{box} es el error cometido durante la estimación de las coordenadas del cuadro delimitador y L_{mask} es el error cometido durante la estimación de la superficie ocupada por el objeto.

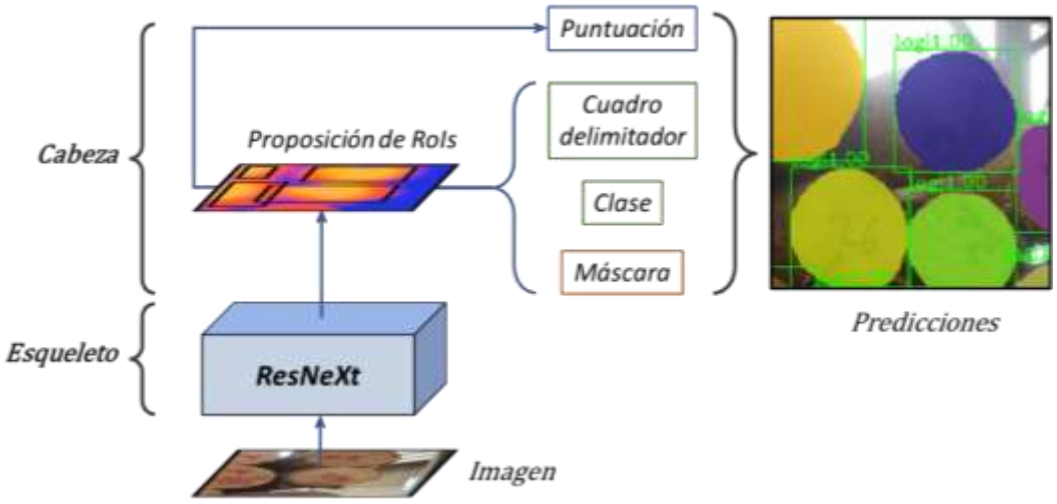

Figura 1. Funcionamiento interno de Mask R-CNN.

3.2. Procesado de imágenes

La base de datos utilizada se compone de 135 imágenes RGB de cargamentos de madera de *Pinus radiata* D. Don tomadas desde la parte trasera de los vehículos, de modo que las testas fuesen total o parcialmente visibles, sin importar el ángulo, la iluminación o la calidad de la imagen (ver Figura 2).

Figura 2. Imágenes empleadas para entrenar el modelo.

La verdad terreno correspondiente con estas imágenes se preparó delineando manualmente la superficie y el cuadro delimitador de cada una de las testas mediante el programa *Computer Vision Annotation Tool* (CVAT) (OPENVINO, 2017). Puesto que el número de imágenes disponibles era limitado, se utilizó el recorte de imágenes como técnica de aumento de datos para suplir esta carencia (SHORTEN & KHOSHGOFTAAR, 2019; ZOPH et al., 2020). En primer lugar, se definió una resolución de 1024x1024 píxeles para todas las imágenes, ya que entrenar estos modelos con resoluciones altas mejora los resultados (TAN & LE, 2019) y emplear el mismo tamaño para todas las imágenes facilita el proceso de aprendizaje (KANNOJIA & JAISWAL, 2018).

Cuando el tamaño de las imágenes fue menor de este valor, se añadieron píxeles con valor cero hasta que se alcanzó. Por otra parte, a las imágenes con una resolución mayor se les realizó cortes con un tamaño de 1024x1024 píxeles de manera sistemática, pero con un cierto solape, de modo que se extrajeran la mayor cantidad posible de parches. Además, cualquier parche que no contuviese ninguna troza fue descartado. Finalmente, la imagen original se redujo hasta los 1024x1024 píxeles y se expandió con píxeles negros para mantener la relación de aspecto. Esto permitió incrementar el número de muestras empleadas durante el entrenamiento de 135 a 418.

La base de datos resultante se dividió aleatoriamente en una proporción del 60% y el 40% entre los sets de entrenamiento y validación respectivamente (ver Tabla 1).

Tabla 1. Distribución de las imágenes de la base de datos entre los sets de entrenamiento y validación.

Set	*Imágenes*		*Trozas*	
	Número	Proporción	Número	Proporción
Entrenamiento	250	60%	4370	59,4 %
Validación	168	40%	2988	40,6 %
TOTAL	418	100%	7358	100 %

3.3. Entrenamiento del modelo

Para entrenar el modelo Mask R-CNN se empleó la plataforma de computación en la nube Google Colaboratory y el código puesto a disposición de manera libre por MMDetection (CHEN, K. et al., 2019). Para ello, el modelo se entrenó durante 12 épocas utilizando los hiperparámetros establecidos por defecto por MMDetection, a excepción del

ratio de aprendizaje y el tamaño de lote, los cuales tomaron un valor de 0.0025 y 2 respectivamente. Por otra parte, se hizo uso de técnicas de aprendizaje por transferencia o "transfer learning" (BOZINOVSKI & FULGOSI, 1976) con el fin de compensar la escasez de datos (ZOPH et al., 2020). Para ello, se utilizó el modelo CNN ResNeXt 101 64x4d (XIE et al., 2017) como esqueleto de Mask R-CNN. Además, durante el entrenamiento se aplicó como técnica de aumento de datos adicional el volteo horizontal de las imágenes con una probabilidad del 50%. Los pasos llevados a cabo se resumen en la Figura 3.

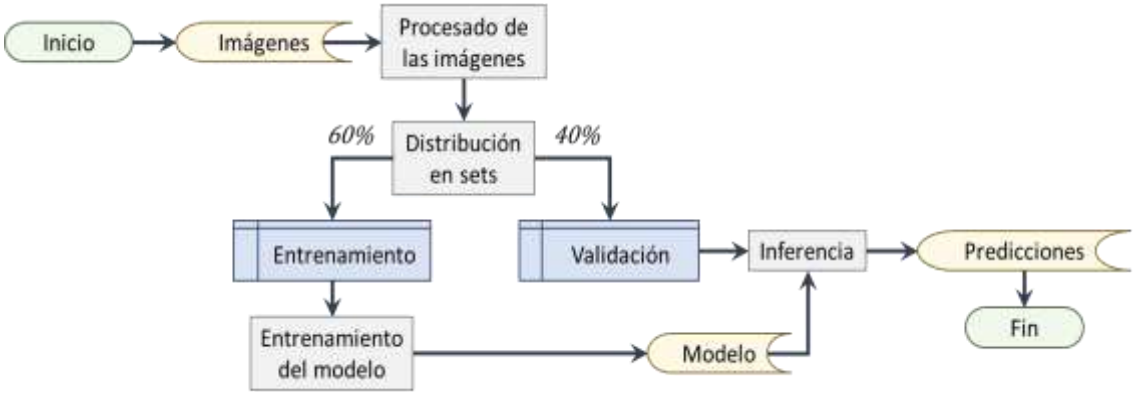

Figura 3. Pasos llevados a cabo para entrenar el modelo y generar las predicciones.

3.4. Validación del modelo

Una vez entrenado el modelo y generadas las predicciones en el set de validación, se procedió al cálculo del índice de Jaccard (IoU) entre los cuadros delimitadores de trozas y predicciones tal como lo describen PADILLA et al. (2020). En base a este índice se determinó si las testas fueron detectadas o no, estableciendo las siguientes categorías:

- **Positivo real (PR):** cuando la troza ha sido detectada correctamente y el valor del IoU es superior a 0,5.
- **Falso positivo (FP):** cuando la predicción no se corresponde con una troza real o se haya desplazada respecto de esta y el IoU es inferior a 0.5.
- **Falso negativo (FN):** cuando la troza no ha sido detectada y el valor del IoU es inferior a 0.5.

Puesto que el modelo empleado ofrece una cantidad de predicciones muy superior a la de objetos reales contenidos en una imagen, existe el riesgo de cometer un doble conteo de positivos reales y sobreestimar la precisión del modelo. Por ende, cuando un objeto real fue detectado por más de una predicción únicamente se consideró como positivo real aquella con la mayor puntuación.

La categorización anterior permitió evaluar la bondad del modelo mediante el cálculo de la proporción de predicciones que resultaron correctas, o precisión (Ecuación 2), y del ratio de positivos reales (RPR), o exhaustividad (Ecuación 3). En base a estas, se calculó el Valor-F como una medida equilibrada entre ambas (Ecuación 4).

$$Precisión = \frac{PR}{PR + FP} \qquad \text{Ecuación 2}$$

$$Exhaustividad\ (RPR) = \frac{PR}{PR + FN} \qquad \text{Ecuación 3}$$

$$Valor - F = \frac{2 \cdot Precisión \cdot RPR}{Precisión + RPR}$$ Ecuación 4

El Valor-F se empleó también para evaluar la bondad de las máscaras generadas por el modelo, para lo cual se compararon las máscaras de la verdad terreno y de las predicciones píxel a píxel.

Además, se evaluó el error relativo en la estimación de la superficie del conjunto de testas contenidas en cada imagen (ε_r) siguiendo la Ecuación 5, midiendo la superficie en número de píxeles.

$$\varepsilon_r^i = \frac{\left|\sum_{t=1}^{t} A_{testa_t} - \sum_{p=1}^{p} A_{pred_p}\right|}{\sum_{t=1}^{t} A_{testa_t}}$$ Ecuación 5

Donde i es la imagen considerada, t es una testa contenida en i, p es una predicción contenida en i, A_{testa} es la superficie de la testa y A_{pred} es la superficie de la predicción.

Por otra parte, con el fin de comparar nuestros resultados con los de otros autores (GUTZEIT & VOSKAMP, 2012), se calculó la desviación entre el número de trozas predichas y el número de trozas reales (N_{log} desv) siguiendo la Ecuación 6, así como el ratio de falsos positivos (RFP) siguiendo la Ecuación 7.

$$N_{trozas}\,\text{desv}^i = \frac{\left|N_{trozas} - N_{preds}\right|}{N_{trozas}}$$ Ecuación 6

Donde i es la imagen considerada, N_{trozas} el número de trozas contenidas en i y N_{preds} el número de predicciones generadas por el modelo para la imagen i.

$$RFP = \frac{FP}{PR + FN}$$ Ecuación 7

Por último, Como se mencionó anteriormente, la puntuación puede entenderse como una medida de la calidad de una predicción, ya que esta representa la probabilidad de encontrar un objeto dentro del cuadro delimitador propuesto por el modelo. Por tanto, esta variable puede emplearse para establecer un umbral de puntuación por debajo del cual se eliminan predicciones de baja calidad, mejorando así los resultados del modelo. Dado que el umbral de puntuación óptimo era desconocido, se procedió a considerar una serie de valores incluidos entre 0 y 0,99 con un intervalo de 0,05, evaluando las variables descritas para cada uno de ellos.

4. Resultados

Una vez entrenado el modelo, se mesuró su bondad de acuerdo con las variables descritas anteriormente empleando el set de validación (ver Tabla 2). Como resultado, se observó que el mejor desempeño del modelo se dio al eliminar todas aquellas predicciones con una puntuación inferior a 0,95. En este punto, el modelo fue capaz de detectar correctamente el 92,4% (σ=12,4) de las trozas con una precisión del 98,8% (σ=6,2), alcanzando un Valor-F del 95,1% (σ=9,0) y una desviación entre el número de trozas predichas y el número de trozas reales del 6,9% (σ=11,3). Además, el modelo estimó la superficie de las testas con un error relativo inferior al 3,6% (σ=9,4), alcanzando un Valor-F de las máscaras del 95,3% (σ=7,3).

Tabla 2. Bondad del modelo Mask R-CNN evaluado en el set de validación; valores expresados porcentualmente.

Precisión	RPR	Valor-F	N_{trozas} desv.	Valor-F másc.	ε_r
98,8 (σ=6,2)	92,4 (σ=12,4)	95,1 (σ=9,0)	6,9 (σ=11,3)	95,3 (σ=7,3)	3,6 (σ=9,4)

En términos generales, nuestro modelo fue capaz de ofrecer buenas predicciones para en un amplio abanico de situaciones, incluyendo imágenes de gran dificultad por la orientación en que se tomaron, por el grado de oclusión de las trozas o por las condiciones de iluminación (ver Figura 4, fila superior). No obstante, en imágenes demasiado complejas el modelo fue incapaz de ofrecer buenos resultados, bien porque la superficie de las testas era apenas visible o porque las imágenes fueron tomadas lateralmente al plano de las testas, lo que sugiere que el mejor desempeño del modelo se da al tomar las imágenes perpendicularmente al plano de las testas (ver Figura 4, fila inferior).

Figura 4. Imágenes complejas para las cuales el modelo ofreció buenos resultados (fila superior) y para las cuales el modelo falló debido su excesiva dificultad (fila inferior).

5. Discusión

En este estudio se pretende probar la capacidad de los modelos CNN para ofrecer buenas estimaciones de la superficie de las testas de trozas apiladas, de modo que sirvan para desarrollar nuevas técnicas de cubicación de pilas de madera. Para ello, se compararon nuestros resultados con los obtenidos por otros autores.

Respecto a la detección de objetos, HERBON et al. (2014) reportaron un RPR del 99,3% y un RFP del 0,4%, aunque no aportaron métricas que evaluasen la calidad de la segmentación de las testas. En cuanto a los demás autores (GUTZEIT & VOSKAMP, 2012; MEHRENTSEV & KRUGLOV, 2019), si bien detectaron una mayor proporción de las trozas que en nuestro caso, reportaron valores de RPR y de RFP mayores a los nuestros.

En lo que a segmentación se refiere, los resultados obtenidos por nuestro modelo fueron superiores a los reportados por GUTZEIT y VOSKAMP (2012) y por GALSGAARD et al.

(2015). Aun con todo, el método propuesto por SAMDANGDECH y PHIPHOBMONGKOL (2018) ofrece los mejores resultados, ya que reportan el mayor Valor-F de las máscaras (97%), además de la menor desviación entre el número de predicciones y el de trozas reales (5,5%). Sin embargo, su método se basa en el uso de dos modelos CNN, de operaciones morfológicas y de etiquetado de elementos conectados para obtener predicciones para cada troza, mientras que nuestro método no requirió de la aplicación de algoritmos intermedios.

Un ventaja importante del método empleado es que el modelo ignoró la corteza de las testas al realizar la segmentación, probablemente debido a que durante el etiquetado de las imágenes estas partes se excluyeron (ver Figura 5). Por el contrario, otros autores se vieron forzados a aplicar factores de corrección (KRUGLOV et al., 2017) o a aplicar pasos adicionales en sus algoritmos (PÁSZTORY & POLGÁR, 2016). Además, los modelos CNN no se ven afectados por la forma de las testas, a diferencia de algoritmos como la transformada de Hough.

Figura 5. Corteza de las trozas ignorada durante la segmentación.

No obstante, algunos autores reportan que los modelos CNN ofrecen peores resultados al tratar de detectar objetos pequeños, lo que puede limitar su viabilidad cuando las testas se encuentran demasiado alejadas o su diámetro es reducido (CHEN, C. et al., 2016). Además, la presencia de oclusiones en las imágenes supone una importante desventaja, ya que estos solamente son capaces de segmentar la superficie visible de los objetos (ver Figura 4, fila inferior).

Sin embargo, la principal limitación de los modelos CNN es que estos deben entrenarse con una gran cantidad de muestras. Además, estos modelos suelen perder precisión cuando se aplican en condiciones diferentes a las de entrenamiento, ya sea por aspectos climáticos, de iluminación o por el ángulo y la distancia a la que se toman las imágenes (ZOPH et al., 2020). En este sentido, cabe destacar que el principal obstáculo encontrado en el presente estudio fue la escasez de muestras, ya que no se dispuso de más de 250 imágenes durante el entrenamiento. Por tanto, en futuros trabajos se abordará la obtención de una base de datos mayor, así como la aplicación de algoritmos de aumento de datos más sofisticados (CUBUK et al., 2018; 2020).

Por último, los recursos computacionales empleados en este estudio fueron limitados, lo que impidió la aplicación de algoritmos de ajuste de hiperparámetros (YU & ZHU, 2020), así como el entrenamiento de modelos más complejos, como el propuesto por QIAO et al. (2020).

6. Conclusiones

Los avances en el campo de la Visión Artificial han dado lugar a la aparición de numerosos métodos de cubicación de pilas de madera basados en detección de objetos. Puesto que las CNN son lo más avanzado en este campo, ofrecen una oportunidad para el desarrollo de nuevas técnicas más precisas y eficientes.

Dado que algunos autores han empleado con éxito las CNN para detectar y segmentar las testas de trozas apiladas, el presente trabajo se llevó a cabo con el de objetivo probar la viabilidad del modelo Mask R-CNN para cubicar pilas de madera de manera automática. Así pues, este modelo se entrenó y validó empleando una base de datos de 418 imágenes de pilas de madera, obteniendo resultados comparables a los de otros autores. En primer lugar, el modelo fue capaz de detectar el 92,4% de las trozas con una probabilidad de acierto del 98,8%. Además, durante la segmentación de las testas se obtuvo un Valor-F de las máscaras del 95,1%, siendo el error relativo en la estimación de la superficie de las pilas de madera inferior al 3,6%.

Los resultados obtenidos en el presente estudio demuestran la capacidad de los modelos CNN para detectar y segmentar de manera automática las testas de trozas apiladas, lo que abre las puertas al desarrollo de herramientas de cubicación de pilas de madera. Esta posibilidad, así como el uso de una base de datos mayor y algoritmos más sofisticados se explorarán en futuros trabajos.

7. Bibliografía

ACUNA, M.; SOSA, A.; 2019. Automated volumetric measurements of truckloads through multi-view photogrammetry and 3D reconstruction software. *CROJFE*. 40(1). p. 151-162.

ALOYSIUS, N.; GEETHA, M.; 2018. A review on deep Convolutional Neural Networks. En: *ICCSP*. p. 588-592.

BOZINOVSKI, S.; FULGOSI, A.; 1976. The influence of pattern similarity and transfer learning upon training of a base perceptron B2. En: *Proc. Symp. Inform.* Bled.

BRAGA, J. R. G.; PERIPATO, V.; DALAGNOL, R.; FERREIRA, M. P.; TARABALKA, Y.; ARAGÃO, L. E. O. C.; CAMPOS VELHO, H. F. DE; SHIGUEMORI, E. H.; WAGNER, F. H.; 2020. Tree crown delineation algorithm based on a convolutional neural network. *Remote Sens.* 12(8). p. 1288.

CHEN, C.; LIU, M. Y.; TUZEL, O.; XIAO, J.; 2016. R-CNN for small object detection. En: *ACCV*. p. 214-230.

CHEN, K.; WANG, J.; PANG, J.; CAO, Y.; XIONG, Y.; LI, X.; SUN, S.; FENG, W.; LIU, Z.; XU, J.; ZHANG, Z.; CHENG, D.; ZHU, C.; CHENG, T.; ZHAO, Q.; LI, B.; LU, X.; ZHU, R.; WU, Y.; DAI, J.; WANG, J.; SHI, J.; OUYANG, W.; LOY, C. C.; LIN, D.; 2019. MMDetection: Open MMLab Detection Toolbox and Benchmark. *arXiv*. 1906.07155.

CUBUK, E. D.; ZOPH, B.; MANÉ, D.; VASUDEVAN, V.; LE, Q. V.; 2018. AutoAugment: Learning Augmentation Policies from Data. *CVPR*. Section 3. p. 113-123.

CUBUK, E. D.; ZOPH, B.; SHLENS, J.; LE, Q. V.; 2020. Randaugment: Practical automated data augmentation with a reduced search space. En: *CVPR*. p. 3008-3017.

GALSGAARD, B.; LUNDTOFT, D. H.; NIKOLOV, I.; NASROLLAHI, K.; MOESLUND, T. B.; 2015. Circular hough transform and local circularity measure for weight estimation of a graph-cut based wood stack measurement. En: *WACV*. p. 686-693.

GUTZEIT, E.; VOSKAMP, J.; 2012. Automatic segmentation of wood logs by combining detection and segmentation. En: *ISVC*. p. 252-261.

HARIHARAN, B.; ARBELÁEZ, P.; GIRSHICK, R.; MALIK, J.; 2014. Simultaneous detection and segmentation. En: *ECCV*. p. 297-312.

HE, K.; GKIOXARI, G.; DOLLÁR, P.; GIRSHICK, R.; 2020. Mask R-CNN. *IEEE PAMI*. 42(2). p. 2961-2969.

HERBON, C.; TÖNNIES, K. D.; OTTE, B.; STOCK, B.; 2015. Mobile 3D wood pile surveying. En: *MVA*. p. 422-425.

HERBON, C.; TÖNNIES, K.; STOCK, B.; 2014. Detection and segmentation of clustered objects by using iterative classification, Segmentation, And Gaussian mixture models and application to wood log detection. En: *GCPR*. Cham. p. 354-364.

JANÁK, K.; 2005. Differences in volume of round timber caused by Different determination methods. *Drv. Ind.* 56(4). p. 165-170.

JANÁK, K.; 2007. Differences in round wood measurements using electronic 2D and 3D systems and standard manual method. *Drv. Ind.* 58(3). p. 127-133.

KANNOJIA, S. P.; JAISWAL, G.; 2018. Effects of Varying Resolution on Performance of CNN based Image Classification An Experimental Study. *IJCSE*. 6(9). p. 451-456.

KÄRHÄ, K.; NURMELA, S.; KARVONEN, H.; KIVINEN, V. P.; MELKAS, T.; NIEMINEN, M.; 2019. Estimating the accuracy and time consumption of a mobile machine vision application in measuring timber stacks. *Comput. Electron. Agric.* 158. p. 167-182.

KATTENBORN, T.; LEITLOFF, J.; SCHIEFER, F.; HINZ, S.; 2021. Review on Convolutional Neural Networks (CNN) in vegetation remote sensing. *P&RS*. 173. p. 24-49.

KNYAZ, V. A.; MAKSIMOV, A. A.; 2014. Photogrammetric technique for timber stack volume contol. En: *ISPRS Archives*. p. 157-162.

KRUGLOV, A.; SHISHKO, E.; KOZHOVA, V.; ZAVADA, S.; 2017. Software for Round Timber Cubic Capacity Measurement through Photogrammetry. En: *ICCAIRO*. p. 288-293.

LECUN, Y.; 1989. Generalization and network design strategies. En: *Connect. Perspect.* p. 143-155.

LI, D.; XIE, W.; WANG, B.; ZHONG, W.; WANG, H.; 2021. Data augmentation and layered deformable Mask R-CNN-based detection of wood defects. *IEEE Access*. 9. p. 108162-108174.

MEHRENTSEV, A. V.; KRUGLOV, A. V.; 2019. The algorithm and software for timber batch measurement by using image analysis. En: *CCIS*. Cham. p. 56-65.

OPENVINO; 2017. Computer Vision Annotation Tool. Disponible en: https://github.com/openvinotoolkit/cvat. Accedido: 27/10/20.

PADILLA, R.; NETTO, S. L.; SILVA, E. A. B. DA; 2020. A Survey on Performance Metrics for Object-Detection Algorithms. En: *IWSSIP*. p. 237-242.

PÁSZTORY, Z.; POLGÁR, R.; 2016. Photo Analytical Method for Solid Wood Content Determination of Wood Stacks. *JOAAT*. 3(1). p. 54-57.

QIAO, S.; CHEN, L.-C.; YUILLE, A.; 2020. DetectoRS: Detecting Objects with Recursive Feature Pyramid and Switchable Atrous Convolution. *arXiv*. 2006.02334. p. 10213-10224.

SAFONOVA, A.; GUIRADO, E.; MAGLINETS, Y.; ALCARAZ-SEGURA, D.; TABIK, S.; 2021. Olive tree biovolume from UAV multi-resolution image segmentation with Mask R-CNN. *Sensors*. 21(5). p. 1617.

SAMDANGDECH, N.; PHIPHOBMONGKOL, S.; 2018. Log-End Cut-Area Detection in Images Taken from Rear End of Eucalyptus Timber Trucks. En: *JCSSE*. p. 1-6.

SCHONBERGER, J. L.; FRAHM, J. M.; 2016. Structure-from-Motion Revisited. En: *CVPR*. p. 4104-4113.

SHORTEN, C.; KHOSHGOFTAAR, T. M.; 2019. A survey on Image Data Augmentation for Deep Learning. *J. Big Data*. 6(1). p. 1-48.

SOLEM, J. E.; 2012. Programming Computer Vision with Python. O'Reilly Media, Inc. p. 264.

TAN, M.; LE, Q. V.; 2019. EfficientNet: Rethinking model scaling for convolutional neural networks. En: *ICML*. p. 6105-6114.

WIMMER, G.; SCHRAML, R.; HOFBAUER, H.; PETUTSCHNIGG, A.; UHL, A.; 2021. Two-Stage CNN-Based Wood Log Recognition. En: *ICCSA*. Springer. Cham. p. 115-125. Disponible en: https://link.springer.com/10.1007/978-3-030-87007-2_9.

XIE, S.; GIRSHICK, R.; DOLLÁR, P.; TU, Z.; HE, K.; 2017. Aggregated residual transformations for deep neural networks. En: *CVPR*. p. 1492-1500.

YU, T.; ZHU, H.; 2020. Hyper-Parameter Optimization: A Review of Algorithms and Applications. *arXiv*. 2003.05689.

ZHIQIANG, W.; JUN, L.; 2017. A review of object detection based on Convolutional Neural Network. En: *CCC*. p. 11104-11109.

ZOPH, B.; GHIASI, G.; LIN, T. Y.; CUI, Y.; LIU, H.; CUBUK, E. D.; LE, Q. V.; 2020. Rethinking pre-training and self-training. En: *NeurIPS*. Vancouver.