Aprendizaje profundo para la detección de árboles en imágenes aéreas de un cultivo de higo

Juan Luis Torres Olivares (CInC - UAEM)
Jorge Alberto Fuentes Pacheco (CONACyT - UAEM)
Edgar Francisco Román Rangel (ITAM)
Porfirio Juárez López (FCA - UAEM)
Jorge Hermosillo Valadez (CInC - UAEM)
Juan Manuel Rendón Mancha (CInC - UAEM)

Introducción

La Agricultura de Precisión tiene como objetivo aumentar la producción de los cultivos, reducir los costos de producción e impacto ambiental. En este contexto, un área de investigación activa es la segmentación de cultivos en imágenes digitales para clasificar, detectar estrés hidríco o plagas, donde segmentar la imagen constituye una etapa importante. Este problema se complica con los cultivos en campo abierto debido a diferentes factores como la iluminación, el clima y las prácticas agrícolas por parte de los agricultores [4].

Segmentar una imagen para hacer una detección de objetos de interés significa subdividirla en regiones que la constituyen donde el nivel de detalle de esta subdivisión depende del problema que se trate de resolver. Realizar una segmentación en imágenes complejas es una de las tareas mas difíciles en procesamiento de imágenes [5].

Usar redes neuronales profundas para tratar el problema de la segmentación es lo más adecuado, dado que se han hecho grandes avances en la resolución de problemas que antes se consideraban difíciles de resolver por parte de la comunidad científica. Además las redes neuronales usan los datos en su forma original sin necesitar un preprocesamiento que puede provocar la pérdida de información relevante [2]. La forma de aplicar las redes neuronales profundas es usando el aprendizaje supervisado que implica observar varios ejemplos (normalmente entre miles y millones de ejemplos) en un conjunto de datos (*dataset*) que están representados como un vector x y un vector y asociado con el objetivo de predecir y a partir de x [3].

Para realizar la detección de árboles de higo se propone una red neuronal convolucional profunda (CNN por sus siglas en inglés) con arquitectura *encoder-decoder*. En el *encoder* se usan las capas Convolución y *Maxpooling* y en el *decoder* se usan las capas Convolución y *Upsampling*. Esta CNN se entrena con un conjunto de datos propio.

Desarrollo

1. Realizar un plan de vuelo a baja altura en una parcela de cultivos de higo en campo abierto con un dron para tomar fotografías.

Figura 1: Izquierda, el dron usado. Derecha, el plan de vuelo sobre el cultivo.

2. Seleccionar diez imágenes tomadas por el dron, recortarlas en secciones de 2000×1500 pixeles y segmentarlas manualmente. Estas segmentaciones serán la segmentación correcta conocida como *Ground Truth* (GT).

Figura 2: Izquierda, imagen a color de tamaño 2000×1500. Derecha, segmentación manual.

3. Recortar las diez imágenes de 2000×1500 pixeles con su segmentación manual en secciones de tamaño 128×128 pixeles con un traslape del 70 %.

Figura 3: Seis recortes consecutivos de 128×128 pixeles con traslape del 70 %.

4. Diseñar una CNN que segmente usando las capas *Convolution*, *Maxpooling* y *Upsampling*. Agregar *batch normalization* a la CNN como otro experimento. Entrenar las CNNs con ocho imágenes de 2000×1500 recortadas de 128×128 durante 120 épocas.

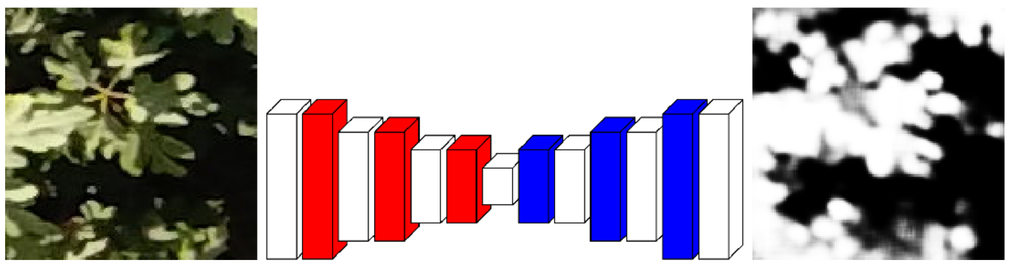

Figura 4: La CNN propuesta para segmentar. La entrada es una imagen RGB recortada y la salida es su segmentación.

5. Debido a que la salida de la CNN tiene valores entre 0 y 1, se hace un postprocesaminto binarizando las predicciones de la CNN como:

$$I_{binarizada}(i,j) = \begin{cases} 1 & \text{si } I_{predicha}(i,j) > 0.5 \\ 0 & \text{si } I_{predicha}(i,j) \leq 0.5 \end{cases}$$

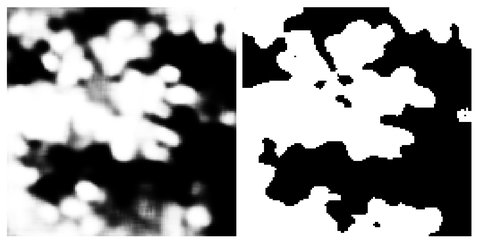

Figura 5: Izquierda, predicción de la CNN. Derecha, predicción binarizada.

6. Comparar los resultados de la CNN propuesta contra SegNet-Basic, una versión más pequeña de SegNet.

Resultados

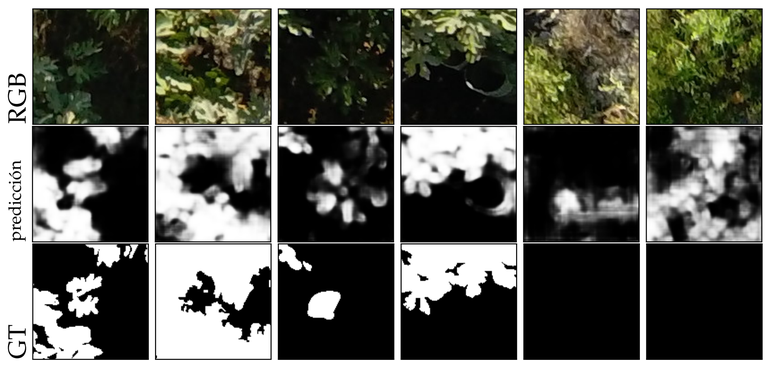

Figura 6: Algunos resultados de segmentar recortes de dos imágenes de prueba.

Métricas como exactitud, precisión y recuerdo [1] se usan para medir el desempeño de la CNN. Para calcular estas métricas primero se debe obtener el número de pixeles clasificados como correctos e incorrectos entre la salida de la CNN y el GT [1].

Clasificado como		*Ground truth* (GT)
Hoja de higo	NO hoja de higo	
VP	FN	Hoja de higo
FP	VN	NO hoja de higo

Cuadro 1: Casos posibles al clasificar los pixeles usando el *ground truth* para comparar. Es deseable que los valores de la diagonal compuesta por VP y VN sean altos y los valores de la diagonal compuesta por FN y FP sean bajos.

En la siguiente tabla se muestran los resultados númericos. El rango para las métricas están entre $[0, 1]$.

Nombre de la métrica	CNN propuesta	CNN + *batch normalization*	SegNet-Basic
Tasa de éxito global o **Exactitud**	0.9384	0.9410	0.9382
Tasa de verdaderos negativos o **Especificidad**	0.9279	0.9466	0.9433
Valor predictivo positivo o **Precisión**	0.9200	0.9385	0.9349
Tasa de verdaderos positivos, sensibilidad o **Recuerdo**	0.9505	0.9345	0.9322
Valor predictivo negativo (VPN)	0.9555	0.9431	0.9410
Medida f	0.9350	0.9365	0.9335

Cuadro 2: Métricas con su definición y valor obetnido, usadas para medir el desempeño de la CNN.

Conclusiones

En este trabajo se implementó una CNN que hace segmentación de imágenes aéreas de cultivos de higo. La CNN se puso a prueba con imágenes propias segmentadas manualmente. En general el desempeño fue por arriba del 90 % para las métricas de evaluación que se usaron cumpliendo con los objetivos y son comparables con el estado del arte.

La calidad de la segmentación apreciada de manera visual es muy buena aunque se usó un conjunto de datos relativamente pequeño. Las imágenes presentan características que dificultan su segmentación como son: diferentes niveles de iluminación, suelo de diferentes colores, maleza y objetos ajenos al cultivo de higo.

Lo mostrado en este trabajo muestra los resultados de usar el aprendizaje profundo para la segmentación de cultivos en campo abierto en condiciones de iluminación no controlados con un pequeño conjunto de datos de entrenamiento. El problema de segmentar las imágenes de árbol de higo usadas en este trabajo es inédito.

Referencias

[1] Tom Fawcett. An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8):861 – 874, 2006. ROC Analysis in Pattern Recognition.

[2] Yann LeCun, Yoshua Bengio y Geoffrey Hinton. Deep learning. *Nature*, 521:436–44, 05 2015.

[3] Ian Goodfellow, Yoshua Bengio y Aaron Courville. *Deep Learning*. MIT Press, 2016. `http://www.deeplearningbook.org`.

[4] Andreas Kamilaris y Francesc X. Prenafeta-Boldú. Deep learning in agriculture: A survey. *Computers and Electronics in Agriculture*, 147:70 – 90, 2018.

[5] Rafael C. Gonzales y Richard E. Woods. *Digital Image Processing*. Parson, 2008.

Los autores agradecen al Laboratorio Nacional de Supercómputo del Sureste de México perteneciente a la red de laboratorios nacionales CONACYT, por los recursos computacionales, el apoyo y la asistencia técnica y al CONACyT.

Artículo publicado