# Segmentación semántica para reconocimiento de escenas

**M. A. Olanda Prieto Ordaz, Dra. Graciela Ramírez Alonso, M.I. David Maloof Flores**
Universidad Autónoma de Chihuahua /Facultad de Ingeniería
FINGUACH Año 6, Núm. 19, marzo - mayo del 2019

En la actualidad, uno de los principales retos en el área de visión por computadora es realizar con mayor precisión las diferentes tareas de clasificar, localizar y etiquetar semánticamente los elementos que contiene una imagen con la finalidad de interpretar el contexto de la misma. A esta tarea se le conoce como reconocimiento de la escena (Shin *et al.*, 2016).

A causa del incremento de las aplicaciones utilizadas en *robots* móviles, automóviles autónomos, realidad virtual, por nombrar algunos, se ha generado un mayor interés en el área de reconocimiento de escena, lo que impulsa la necesidad de comprender imágenes complejas (Bassiouny & El-Saban, 2014).

La segmentación semántica, considerada como una tarea de alto nivel que encauza a una mayor comprensión de la escena, consiste en asignar una etiqueta semántica a cada píxel de una imagen que permite identificar a todos los elementos que constituyen la escena. Sin embargo, para realizar esta ardua tarea es necesario considerar que la segmentación semántica no es una actividad aislada en el área de visión por computadora, se deben realizar previamente una clasificación y localización de las diversas clases de elementos que la componen (Yu *et al.*, 2018), (Garcia-Garcia *et al.*, 2018).

La Figura 1 muestra un ejemplo de lo que realiza un algoritmo de segmentación semántica. Cada píxel es etiquetado en alguna de las clases que se han definido previamente.

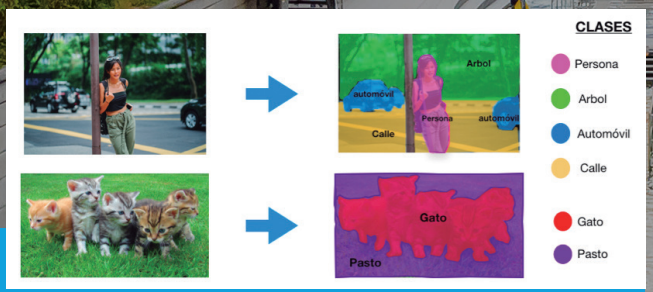


**Figura 1**. Segmentación semántica de una imagen.

Al partir de lo anterior, existen dos aspectos que contribuyen a la calidad en los resultados de una segmentación semántica. El primero consiste en diseñar una representación de características que permita diferenciar los objetos de varias clases. El segundo, va dirigido a cómo utilizar la información contextual para asegurar la consistencia entre las etiquetas de los pixeles, es decir cómo etiquetar cada pixel coherentemente (Yu *et al.*, 2018).

Algunos de los modelos computacionales que han sido ampliamente utilizados en este tipo de tareas están basados en arquitecturas de aprendizaje profundo, específicamente de Redes Neuronales Convolucionales (CNN). En términos generales una CNN es una red neuronal multicapa que toma varias entradas de un tamaño fijo y posteriormente produce una clasificación de toda la imagen. Las CNN fueron diseñadas al tomar como referencia un estudio acerca del funcionamiento de la corteza visual de gatos, en donde se identificó que las neuronas son organizadas jerárquicamente para recibir la información visual, partiendo de las células más simples y superficiales a células más complejas y conforme se adentran en profundidad responden a características de mayor nivel de dificultad (Hubel & Wiesel, 1968).

El proceso de abstraer información a partir de una imagen que permita realizar las tareas de clasificación y reconocimiento en una CNN, simula un proceso al publicado en el estudio realizado por Hubel & Wiesel, donde las capas más superficiales de la CNN obtienen información de características simples de la imagen y conforme se avanza en capas más profundas se obtienen características más complejas.

Las arquitecturas de CNN que principalmente han contribuido y que son punto de referencia en el área son: *AlexNet, VGG-16*, *GoogleLeNet* y *ResNet*, las cuales han surgido en ese orden cronológico a partir del 2012 y han destacado en el Reto del Reconocimiento Visual a Gran Escala de *ImageNet* (ILSVRC). Estas arquitecturas han sido modificadas por diferentes investigadores para generar mejoras en las técnicas de segmentación semántica. La red *Fully Convolutional Networks* o FCN propuesta por Shelhamer, Long, & Darrell en el año 2017, es un ejemplo de esto, en donde la salida del modelo se compone de mapas espaciales utilizados para decodificar la entrada produciendo píxeles etiquetados. Esta arquitectura de red es una de las más populares para tareas de segmentación semántica (Shelhamer, Long, & Darrell, 2017),(García-García *et al.*, 2018). La Figura 2 muestra el diagrama de la red FCN.

La segmentación semántica aporta una gran cantidad de información a la comprensión de una escena, sin embargo, queda un camino intrincado para definir las diferentes relaciones de las clases encontradas en la imagen que faciliten la comprensión de la escena. Aun y cuando diferentes autores han propuesto modelos que han logrado buenos resultados, todavía se considera como un problema abierto para aplicaciones del mundo real donde pueden ocurrir diferentes perturbaciones que afectan a la calidad de los resultados de la segmentación semántica.

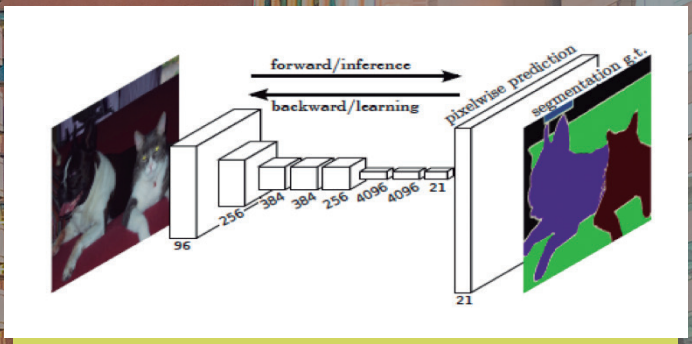


**Figura 2**. Imagen tomada de *Fully Convolutional Network* de *Long et al.*, 2017.

**Referencias:**

Bassiouny, A., & El-Saban, M. (2014). Semantic segmentation as image representation for scene recognition. In 2014 IEEE International Conference on Image Processing (ICIP) (pp. 981–985). IEEE. http://doi.org/10.1109/ICIP.2014.7025197

García-García, A., Orts-Escolano, S., Oprea, S., Villena-Martínez, V., Martinez-Gonzalez, P., & Garcia-Rodríguez, J. (2018). A survey on deep learning techniques for image and video semantic segmentation. *Applied Soft Computing Journal*, 70, 41–65. http://doi.org/10.1016/j.asoc.2018.05.018

Hubel, D. H., & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *The Journal of Physiology*, 195(1), 215–243. http://doi.org/10.1113/jphysiol.1968.sp008455

Shelhamer, E., Long, J., & Darrell, T. (2017). Fully Convolutional Networks for Semantic Segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(4), 640–651. http://doi.org/10.1109/TPAMI.2016.2572683

Shin, H.-C., Roth, H. R., Gao, M., Lu, L., Xu, Z., Nogues, I.; Summers, R. M. (2016). Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning. IEEE Transactions on Medical Imaging, 35(5), 1285–1298. http://doi.org/10.1109/TMI.2016.2528162

Yu, H., Yang, Z., Tan, L., Wang, Y., Sun, W., Sun, M., & Tang, Y. (2018). Methods and datasets on semantic segmentation: A review. Neurocomputing, 304, 82–103. http://doi.org/10.1016/j.neucom.2018.03.037