

Fundamentos de segmentación semántica

Contents

- Segmentación de Imágenes
- Referencias

Segmentación de Imágenes

La segmentación de imágenes es una técnica fundamental en el procesamiento digital de imágenes y visión por computador que consiste en dividir una imagen en regiones homogéneas con base en ciertas características visuales como color, textura, forma o intensidad. Su objetivo principal es simplificar o cambiar la representación de una imagen para hacerla más fácil de analizar e interpretar computacionalmente.

A diferencia de la detección de objetos, que produce coordenadas de cajas delimitadoras (bounding boxes), la segmentación se centra en asignar una etiqueta a cada píxel, clasificándolo como parte de una clase específica u objeto determinado.

Aplicaciones

1. Imagenología Médica:

- La segmentación delimita tejidos blandos, tumores, vasos sanguíneos y órganos.
- Permite construir mapas tridimensionales para cirugía guiada por imágenes.
- Mejora el análisis longitudinal de enfermedades crónicas mediante el seguimiento de cambios en tejidos segmentados.

2. Teledetección y Geomática:

- Clasificación de cobertura del suelo (urbano, agrícola, forestal, cuerpos de agua).
- Delimitación de cuencas, costas o zonas de riesgo por desastres naturales.
- Extracción de objetos como edificios, caminos o parcelas agrícolas.

3. **Conducción Autónoma:** Para que un vehículo autónomo navegue de forma segura, debe reconocer con exactitud su entorno inmediato. La segmentación semántica y de instancias transforma la entrada visual de cámaras en un mapa de ocupación semántica.

- Etiquetado por píxel de carriles, bordes, señales, peatones y vehículos.
- Separación entre objetos dinámicos y estáticos.

1. Agricultura de Precisión:

- Delimitación de lotes, surcos y especies vegetales.
- Segmentación de áreas afectadas por plagas, enfermedades o estrés hídrico.
- Clasificación de estados fenológicos.

Tipos de Segmentación de Imágenes según

La segmentación de imágenes puede clasificarse en función del tipo de información que proporciona como salida. Cada tipo responde a una necesidad diferente de interpretación visual y está adaptado a distintas aplicaciones prácticas.

1. Segmentación Semántica

Este método asigna a cada píxel de la imagen una clase predefinida, como "carretera", "vegetación", "cielo", "edificio" o "persona". El resultado es una máscara de clasificación por píxel, donde cada región visual es etiquetada según su categoría, sin diferenciar entre instancias distintas de una misma clase.

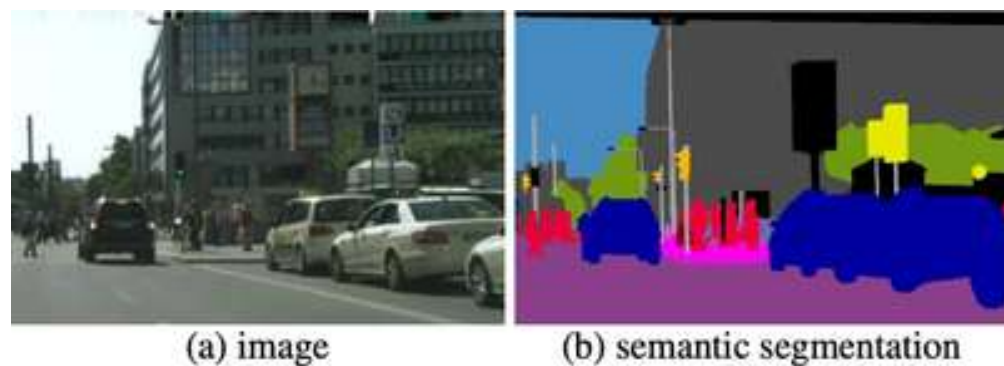


Fig.1. Imagen tomada de: Deep learning-based panoptic segmentation: Recent advances and perspectives - Scientific Figure on ResearchGate. Disponible en ResearchGate

La arquitectura comúnmente empleada es el Fully Convolutional Network (FCN), junto con variantes como U-Net, Feature Pyramid Network(FPN), LinkNet, Pyramid Scene Parsing Network (PSPNet), entre otras, que extraen características jerárquicas y reconstruyen el mapa de etiquetas a la resolución original.

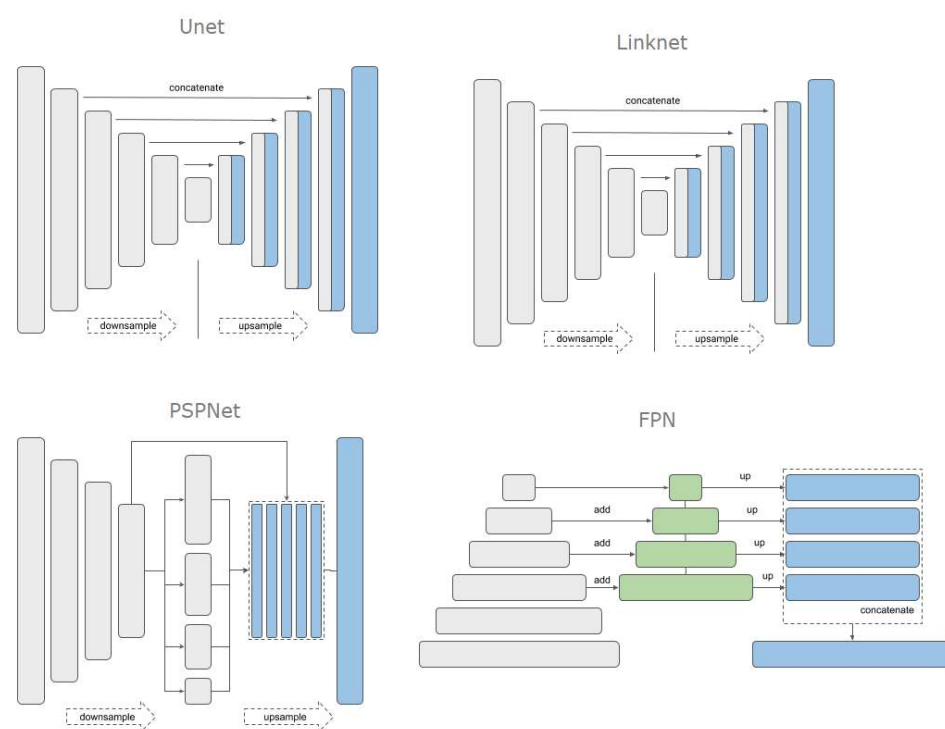


Fig.2. Imagen tomada de [repositorio de Qubvel](#)

Estructura del Dataset-

Un dataset típico de segmentación semántica está compuesto por:

- **Imágenes de entrada:** Estas pueden ser imágenes RGB, multiespectrales, térmicas o derivadas de sensores remotos como satélites o drones. En aplicaciones geoespaciales, por ejemplo, las imágenes pueden tener resolución submétrica o de decenas de metros, dependiendo de la fuente.
- **Máscaras de anotación por píxel:** Son imágenes que tienen la misma dimensión espacial (alto y ancho) que la imagen original, pero en lugar de contener valores de color, cada píxel almacena un número entero que representa una clase semántica. Por ejemplo: 0 = fondo, 1 = carretera, 2 = construcción y 3 = vegetación. Cada valor indica a qué clase pertenece cada píxel, permitiendo a la red aprender patrones de segmentación. El formato más común es una matriz de enteros de una sola capa (1 canal), donde los valores codifican las clases.



Fig.3 dataset

Métricas de evaluación

- **Precisión por píxel** se define como la proporción de píxeles que fueron correctamente clasificados respecto al total de píxeles en la imagen. Es una métrica general que da una visión global del desempeño del modelo, sin tener en cuenta el desequilibrio entre clases.

$$\text{Pixel Accuracy} = \frac{\sum_{i=1}^k TP_i}{\sum_{i=1}^k (TP_i + FP_i + FN_i)}$$

Donde:

- TP_i : Número de píxeles verdaderamente positivos de la clase (i).
- FP_i : Número de píxeles falsamente positivos de la clase (i).
- FN_i : Número de píxeles falsamente negativos de la clase (i).
- k : Número total de clases.
- **mIoU** (intersección sobre unión media) Se calcula como el promedio, sobre todas las clases, de la intersección entre la predicción y la verdad de terreno, dividido por su unión.

$$IoU_i = \frac{TP_i}{TP_i + FP_i + FN_i}$$

$$mIoU = \frac{1}{k} \sum_{i=1}^k IoU_i$$

Donde:

- TP_i : Píxeles correctamente clasificados como clase i
- FP_i : Píxeles incorrectamente clasificados como clase i
- FN_i : Píxeles de clase i que fueron mal clasificados
- k : Número total de clases

Esta métrica penaliza fuertemente tanto falsos positivos como falsos negativos, dando una mejor idea de la calidad espacial de la segmentación.

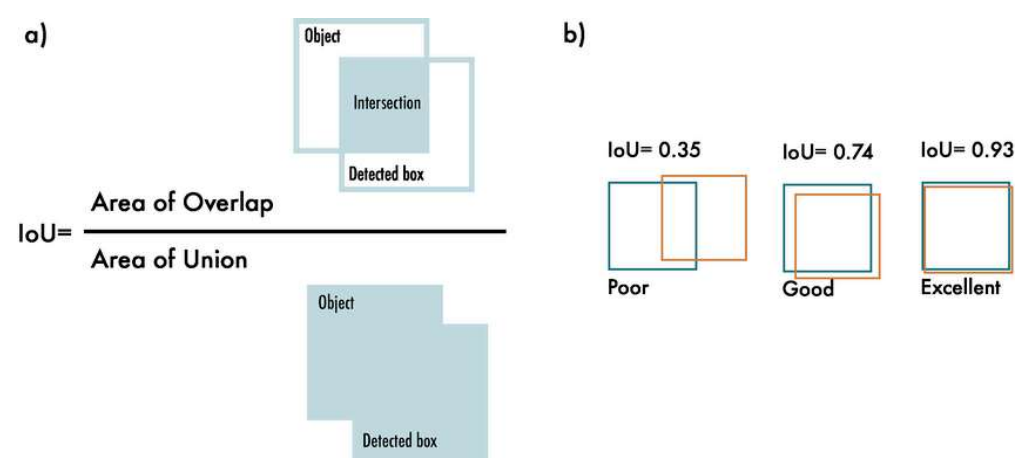


Fig.4. Loss Functions and Metrics in Deep Learning. A Review - Scientific Figure on ResearchGate.* > Disponible en: ResearchGate

2. Segmentación de Instancias

La segmentación de instancias introduce la capacidad de identificar y separar objetos individuales que pertenecen a la misma clase. Cada píxel es asignado no solo a una clase, sino también a una instancia única de esa clase. Por ejemplo, dos autos serán segmentados como “auto 1” y “auto 2”, aunque ambos pertenezcan a la clase “auto”. Este tipo de segmentación requiere una combinación de detección de objetos y segmentación precisa. Algoritmos como Mask R-CNN utilizan una arquitectura híbrida: primero detectan cajas delimitadoras (bounding boxes) para cada objeto y luego generan una máscara de segmentación para cada uno.

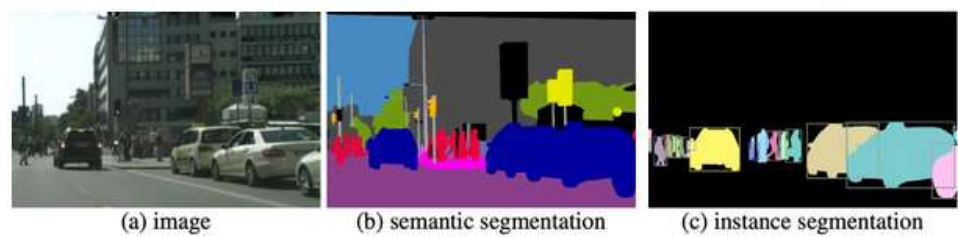


Fig.5. Imagen tomada de: Deep learning-based panoptic segmentation: Recent advances and perspectives - Scientific Figure on ResearchGate. Disponible en ResearchGate

Estructura del Dataset-

- **Máscara por instancia** es una imagen binaria (en escala de grises) donde: Los píxeles del objeto tienen valor 1, el fondo tiene valor 0.Para segmentación por instancias, no se hace una sola máscara para todos los objetos de la misma clase (como en segmentación semántica), sino una máscara independiente para cada objeto.
- **Etiqueta de clase por instancia**; Cada máscara debe estar asociada a una clase específica (por ejemplo, “via”, “vegetacion”, “pconstruccion”).
- **Bounding Boxes**: por instancia (opcional pero común).Muchas arquitecturas como Mask R-CNN también usan una caja delimitadora (bounding box) para cada objeto.

3. Segmentación panóptica

Su propósito es ofrecer una interpretación global y unificada de la escena, combinando simultáneamente la comprensión semántica del contenido visual con la identificación de objetos individuales.

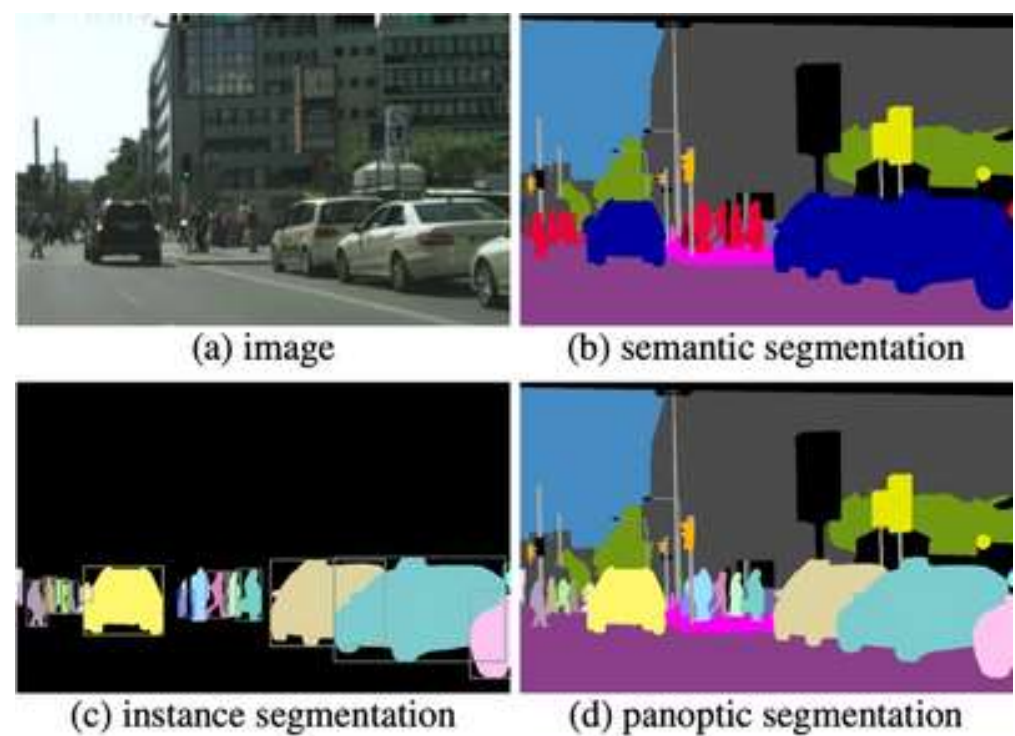


Fig.6. Imagen tomada de: Deep learning-based panoptic segmentation: Recent advances and perspectives - Scientific Figure on ResearchGate. Disponible en ResearchGate

Estructura del Dataset- Esta segmentación unifica los dos enfoques anteriores: segmenta instancias individuales de objetos “contables” (carros, construcciones) y clasifica regiones de “cosas no contables” (vegetación, agua) a través de segmentación semántica.

- Una máscara de segmentación semántica por píxel.
- Una colección de máscaras de instancias individuales.
- Cada píxel tiene una etiqueta de clase y (si corresponde) un ID de instancia.

4. Segmentación Interactiva

El usuario indica una región de interés mediante clics, trazos o puntos. A partir de estas señales, el algoritmo genera una segmentación del objeto deseado.

Modelos como SAM (Segment Anything Model) o Deep Extreme Cut aprenden a generalizar desde señales mínimas para segmentar objetos desconocidos. Estos modelos no requieren clases predefinidas y pueden segmentar cualquier tipo de objeto.



Fig.7. Imagen tomada de: Segment Anything

Referencias

- Qubvel. (2019). segmentation_models [Computer software]. GitHub. [qubvel/segmentation_models](https://github.com/qubvel/segmentation_models)
- Chuang, Y., Zhang, S., & Zhao, X. (2023). Deep learning-based panoptic segmentation: Recent advances and perspectives. IET Image Processing, 17, Article e12853. <https://doi.org/10.1049/ipr2.12853>
- Meta AI. (2023). Segment Anything. <https://segment-anything.com/>
- IBM. (s.f.). ¿Qué es la segmentación de imágenes? Recuperado, de [https://www.ibm.com/es-es/topics/image-segmentation:contentReference\[oaicite:1\]{index=1}](https://www.ibm.com/es-es/topics/image-segmentation:contentReference[oaicite:1]{index=1})