Clase 133 — Segment Anything (SAM / SAM 2): foundation model para segmentación

Parte: 2 — Deep Learning · Fuente: Kirillov et al. (2023) + Ravi et al. (2024) SAM 2. Duración estimada: 85 min.

Clase 133 — Segment Anything (SAM / SAM 2): foundation model para segmentación

> *Parte: 2 — Deep Learning · Fuente: Kirillov et al. (2023) + Ravi et al. (2024) SAM 2. Duración estimada: 85 min.*

Objetivo

Usar Segment Anything (Meta AI 2023) y SAM 2 (2024) — el foundation model para segmentación: entrenado en 11M imágenes + 1.1B máscaras (SAM 2 agrega video). Segmenta cualquier objeto dado un prompt (punto, caja, máscara, "todo"). Zero-shot — no requiere training para empezar.

Resultados de aprendizaje

Al finalizar, el estudiante podrá:

- Instalar y cargar SAM/SAM 2 con pip install 'git+https://github.com/facebookresearch/segment-anything-2'.
- Aplicar prompts: punto, caja, multi-punto positivo/negativo.
- Generar máscaras en modo "everything" (segmenta cada objeto automáticamente).
- Tracking de máscaras en video con SAM 2 (memoria temporal).
- Combinar SAM con detector (YOLO) para pipeline detection → segmentation.

Temas

- Arquitectura SAM: ViT encoder + prompt encoder + mask decoder.
- Promptable: una sola red, múltiples interfaces.
- SAM 2: adds memory + tracking en video.
- Variants: vit_h (mejor calidad), vit_l, vit_b (más rápido).
- Pipeline detección + segmentación: YOLO/Grounding DINO → boxes → SAM → masks.
- Fine-tuning SAM (rara vez necesario, casos médicos / dominios muy específicos).

Definiciones y características

- Promptable segmentation: input = imagen + prompt (punto/caja/mask) → output = mask.
- Mask decoder: producirá hasta 3 máscaras por prompt (handles ambigüedad).
- SamPredictor: API estándar para una imagen.
- SamAutomaticMaskGenerator: modo "everything" — segmenta toda la imagen automáticamente.
- SAM 2 memory: bank de features previas para tracking en video.

Dataset / recursos

- Imágenes propias o cualquier dataset visual.
- Modelos pretrained: <https://github.com/facebookresearch/segment-anything-2>.
- Librerías: segment-anything-2 (PyTorch).

Ejercicios

1. SAM setup: descargar checkpoint vit_h. Cargar con SamPredictor.
2. Punto prompt: predictor.set_image(img); masks, scores, _ = predictor.predict(point_coords=[[x,y]], point_labels=[1]).
3. Box prompt: pasar box=[x1,y1,x2,y2]; útil después de YOLO.
4. Everything mode: SamAutomaticMaskGenerator(sam).generate(img) → lista de masks.
5. SAM 2 video tracking: tomar un punto en frame 0, propagar a través del video.

Homework verificable

Pipeline YOLO + SAM para anotación semi-automática:

1. YOLOv11 detecta objects → boxes.
2. Para cada box, SAM produce máscara precisa.
3. Visualizar overlay sobre imagen.
4. Reportar tiempo por imagen.

Criterio de aceptación: las máscaras son visualmente correctas (alineadas a contornos); pipeline corre en GPU < 1s por imagen.

Errores comunes

Síntoma / mensaje	Causa y cómo arreglar
SAM lento en CPU	Es enorme. Fix: usar vit_b o GPU.
Máscara incluye background	Prompt ambiguo. Fix: agregar puntos negati
Out-of-memory con ViT-H	4-8 GB VRAM necesarios. Fix: vit_l o ViT-B
Mask decoder devuelve 3 masks, ¿cuál uso?	El score más alto suele ser la correcta. F
SAM 2 tracking pierde objeto en video	Re-prompt cada N frames.

Preguntas frecuentes

¿SAM reemplaza segmentación supervisada?

Para uso interactivo / anotación, sí. Para producción con clases específicas sin intervención humana, fine-tune un segmentador clásico (DeepLabV3+, YOLO-seg) es más eficiente.

¿Open vocabulary?

SAM solo: NO. Para "segmenta el perro" con texto: combinar con Grounding DINO (detector text-grounded) → caja → SAM → mask. O usar Grounded-SAM end-to-end.

¿SAM en mobile?

Hay versiones distilled: MobileSAM, EfficientSAM — 10× más rápidas, ligera pérdida de calidad.

¿Fine-tune SAM?

Posible pero rara vez vale la pena. Casos médicos / dominios muy distintos.

¿Licencia?

Apache 2.0. Comercialmente OK.

Referencias

- Kirillov et al. (2023), Segment Anything, ICCV.
- Ravi et al. (2024), SAM 2: Segment Anything in Images and Videos.
- Segment Anything project.
- Grounded-SAM.

Siguiente clase

Clase 134 — YOLOv11 práctico: detección, segmentación, pose, tracking

Apéndice: notebook (primer bloque)

SAM pesa ~2.4 GB (ViT-H). Fallback con SLIC superpixels de skimage que ilustra el concepto de máscara por prompt sobre imagen sintética 256x256.

```
USE_SAM = False
try:
    from segment_anything import sam_model_registry, SamPredictor
    USE_SAM = True
    print('segment_anything disponible')
except Exception as e:
    print('SAM no disponible. Fallback SLIC. Motivo:', type(e).__name__)

import numpy as np
import matplotlib.pyplot as plt
from skimage.segmentation import slic, mark_boundaries
np.random.seed(42)
```

Archivos complementarios

- notebook.ipynb