

Cegados por la IA

Anexo B: Datos, Entrenamiento y Métricas de Calidad

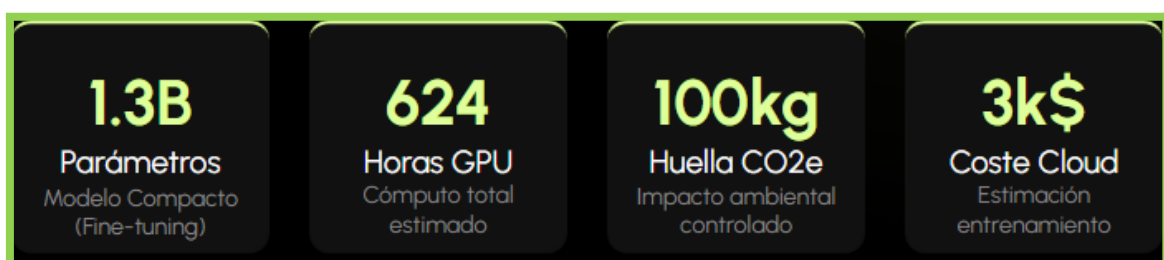


1. Dimensionamiento del Cómputo

Entrenar un modelo multimodal competente no es trivial. Hemos calculado los requisitos para realizar el *fine-tuning* de nuestro modelo compacto de 1.3B parámetros.

Cálculos de Entrenamiento:

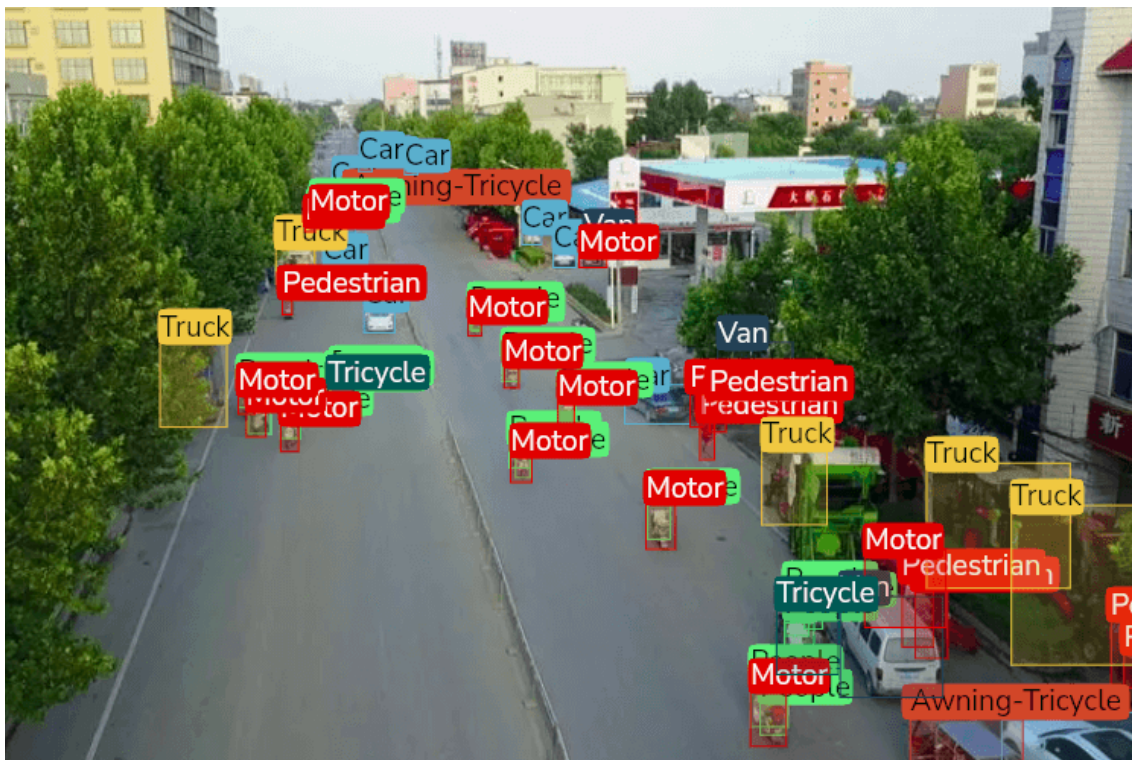
- **FLOPs estimados:** $2,3 \times 10^{20}$ FLOPs por ciclo de entrenamiento.
- **Tiempo de GPU:** Estimamos unas **624 horas de GPU** (equivalente a NVIDIA A100/H100) para cubrir el entrenamiento base y las iteraciones de corrección (factor x3 de seguridad).
- **Coste Financiero:** Utilizando instancias *spot* en la nube (~3 USD/hora), el presupuesto de cómputo ronda los **2.000 - 3.000 USD**.
- **Huella de Carbono:** ~100 kg CO₂e, un impacto bajo gracias al uso de modelos compactos.



2. Estrategia de Datos (Datasets)

No usamos datos de nuestros usuarios para entrenar. Nuestra política es clara: **modelos pre-entrenados + datasets públicos licenciados**.

Dataset	Uso Principal	Licencia
MS COCO / LAION-5B	Comprensión general del mundo y objetos comunes.	CC-BY / Apache
VizWiz	Crítico: Fotos reales tomadas por personas ciegas (desenfocadas, mal encuadradas).	CC-BY
TextCaps / COCO-Text	OCR en el mundo real (leer carteles, precios, menús).	Varias
Cityscapes	Navegación urbana (bordillos, semáforos, pasos de cebra).	Varias



3. Criterios de Aceptación (Go / No-Go)

No lanzaremos ninguna actualización que no pase estos umbrales automáticos y humanos:

1. **Seguridad Física:** Recall > 99% en obstáculos críticos (huecos, tráfico).
Cero falsos negativos permitidos.
2. **Calidad Descriptiva:** Puntuación CIDEr ≥ 90 en el benchmark VizWiz.
3. **Aceptación de Usuario:** Puntuación SUS (System Usability Scale) ≥ 80 en pruebas con voluntarios de la ONCE.

4. Referencias

- **Documentos de los sprints:**
 - *Entrenamiento v3.0.docx*
 - *Entrega - Sprint 3.docx* (Evaluación y calidad)
- **Referencias externas:**
 - MS COCO Captions: <https://cocodataset.org/#captions-2015>
 - LAION-5B Dataset: <https://laion.ai/blog/laion-5b/>
 - VizWiz (Datos de usuarios ciegos): <https://vizwiz.org/>
 - TextCaps (OCR): <https://textvqa.org/textcaps>
 - Cityscapes (Entorno urbano):
<https://www.cityscapes-dataset.com/>
 - Costes de GPU Cloud (Lambda Labs):
<https://lambdalabs.com/service/gpu-cloud>