

# **Segmentación Semántica con SegFormer: Un Enfoque con Transformers para Imágenes Médicas**

**Evelyn Cueva J.**

**Proyecto final, Aprendizaje Profundo, Maestría en Inteligencia Artificial**  
**08 de febrero de 2025**

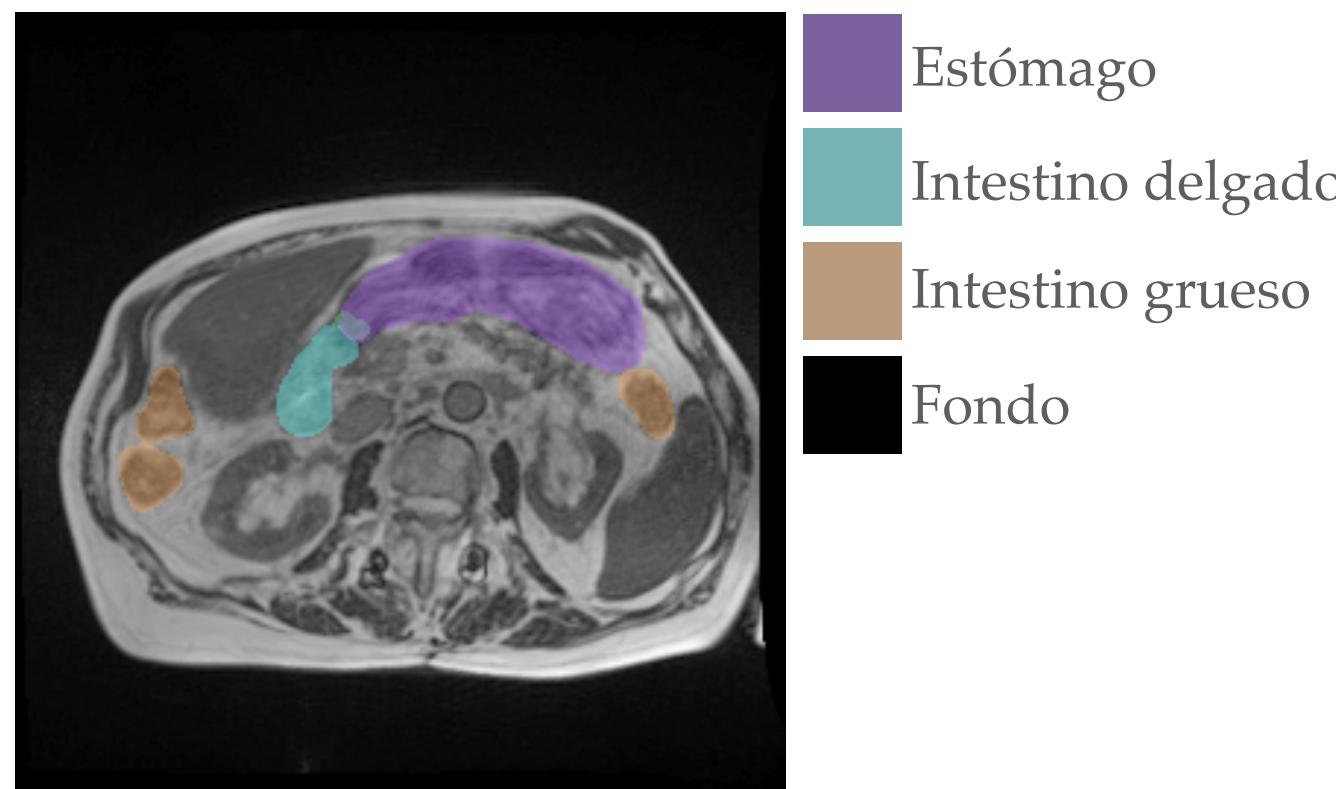


# Segmentación semántica

Identificación y diferenciación de regiones pixel por pixel.



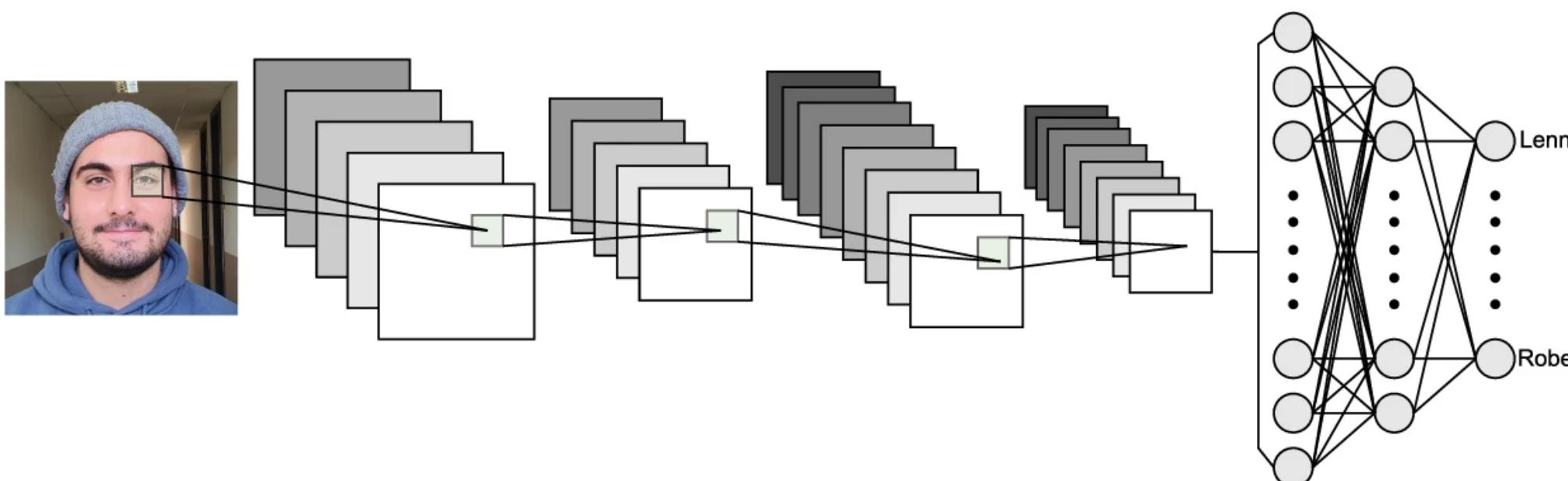
En imágenes médicas (MRI abdominal):  
Identificación del tracto intestinal luminal



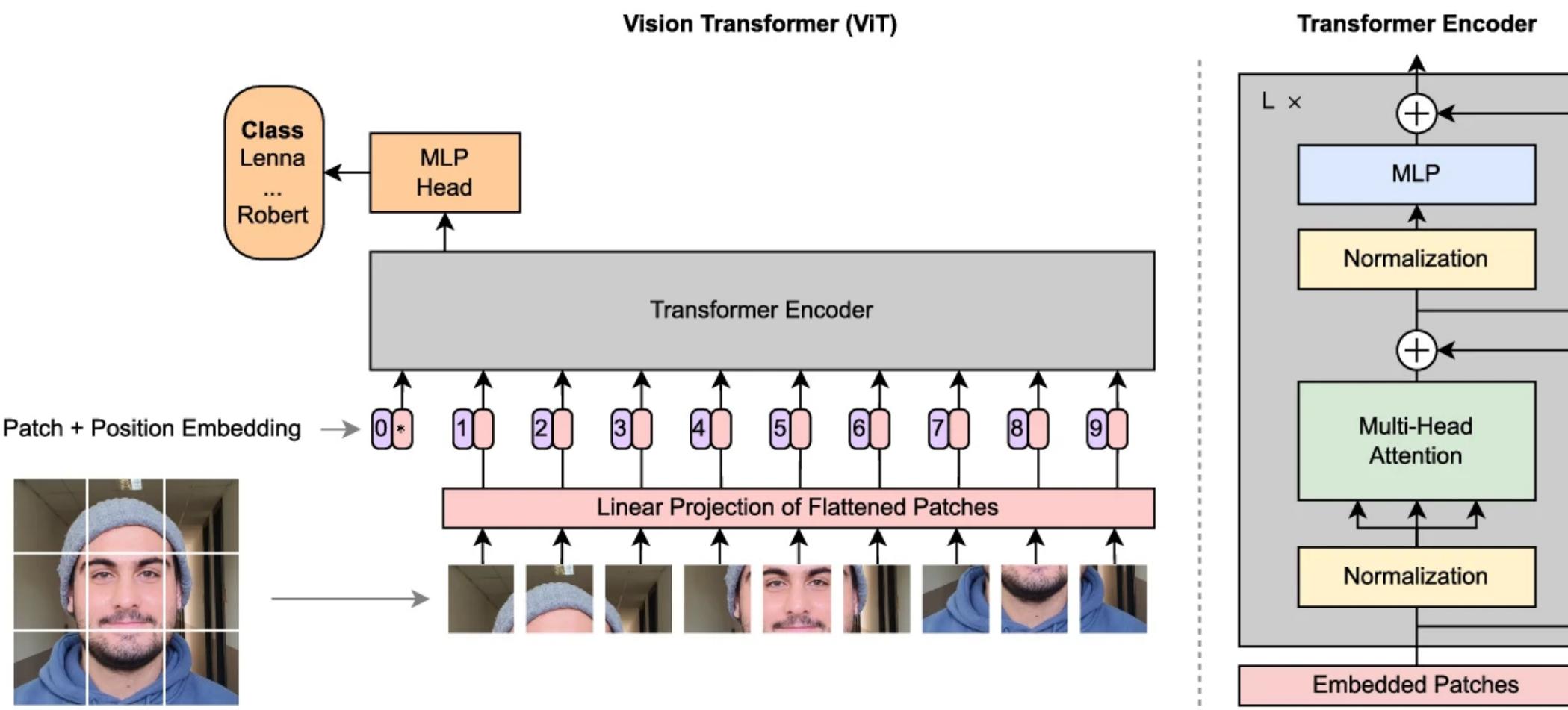
# Motivación

# CNN vs Transformers

## CNNs

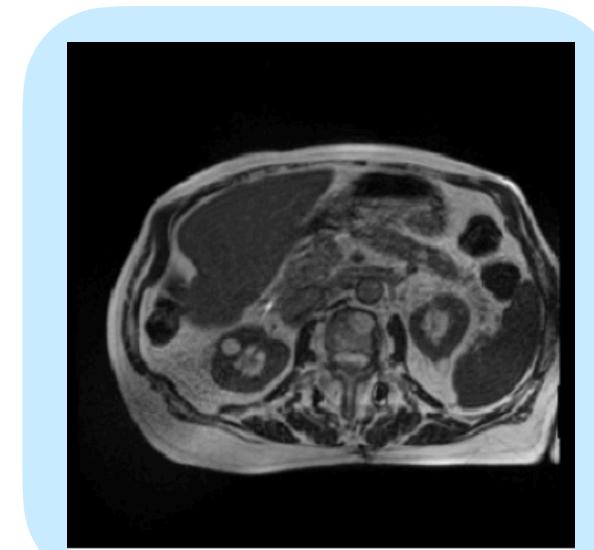


## Transformers



- **Convoluciones** para extraer **características locales**: bordes o texturas a partir de **regiones cercanas en la imagen**.
- **Procesan las imágenes** directamente **en su formato matricial (2D)**, manteniendo una estructura espacial estricta.
- **Mecanismos de atención** (Self-Attention) para capturar relaciones **globales** en toda la imagen, independientemente de la distancia espacial entre los píxeles.
- Dividen la imagen en **patches** (**fragmentos cuadrados**) que se convierten en vectores lineales (secuencias), con información adicional de posiciones relativas (**positional encoding**).

# Segformer: Segmentación + Transformer (2021)



Entrada

Imagen de tamaño  
 $H \times W$

Encoder

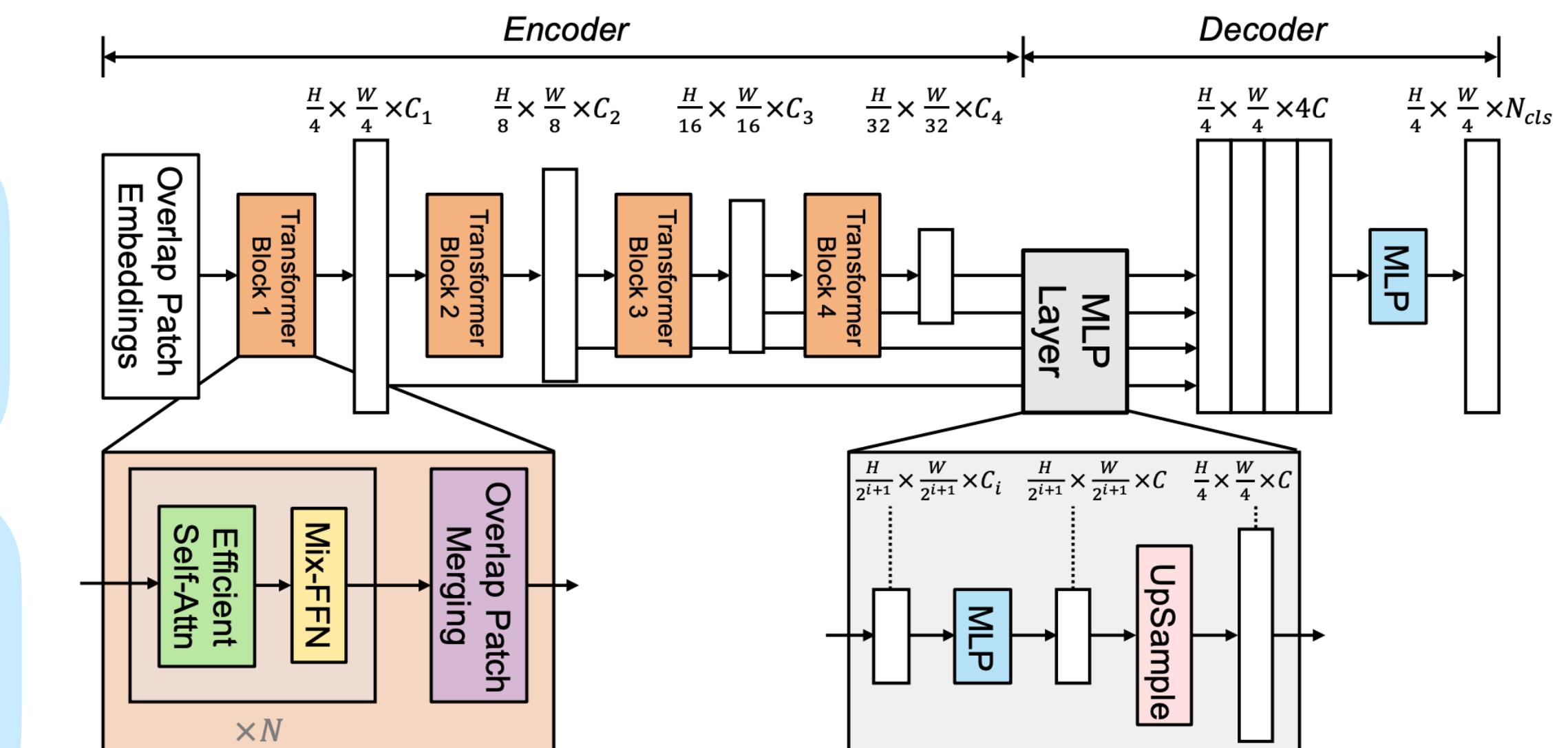
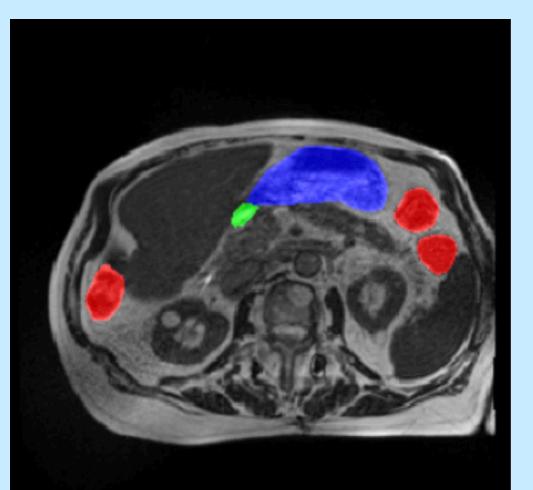
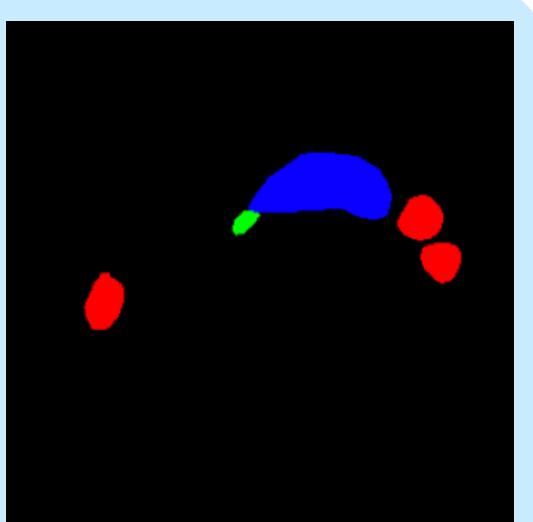
Produce características multiescala **sin codificación posicional**, evitando problemas de interpolación cuando las resoluciones del test y train difieren.

Decoder

Integra atención **local y global** para generar representaciones efectivas, eliminando la necesidad de decodificadores complejos.

Salida

Mapa de  
segmentación  
 $H/4 \times W/4 \times$   
clases



**SegFormer-B4** (50.3% mIoU en ADE20K con 64M parámetros):

- 5 veces más pequeño y 2.2% mejor en mIoU que **Swin Transformer**.
- Logra una eficiencia y escalabilidad significativamente superiores debido a su decodificador **MLP** simple y la ausencia de codificación posicional.

# Conjunto de datos y preprocesamiento

Este conjunto de datos fue usado para el desafío de Segmentación de Tracto Gastrointestinal UW-Madison auspiciado por Kaggle.

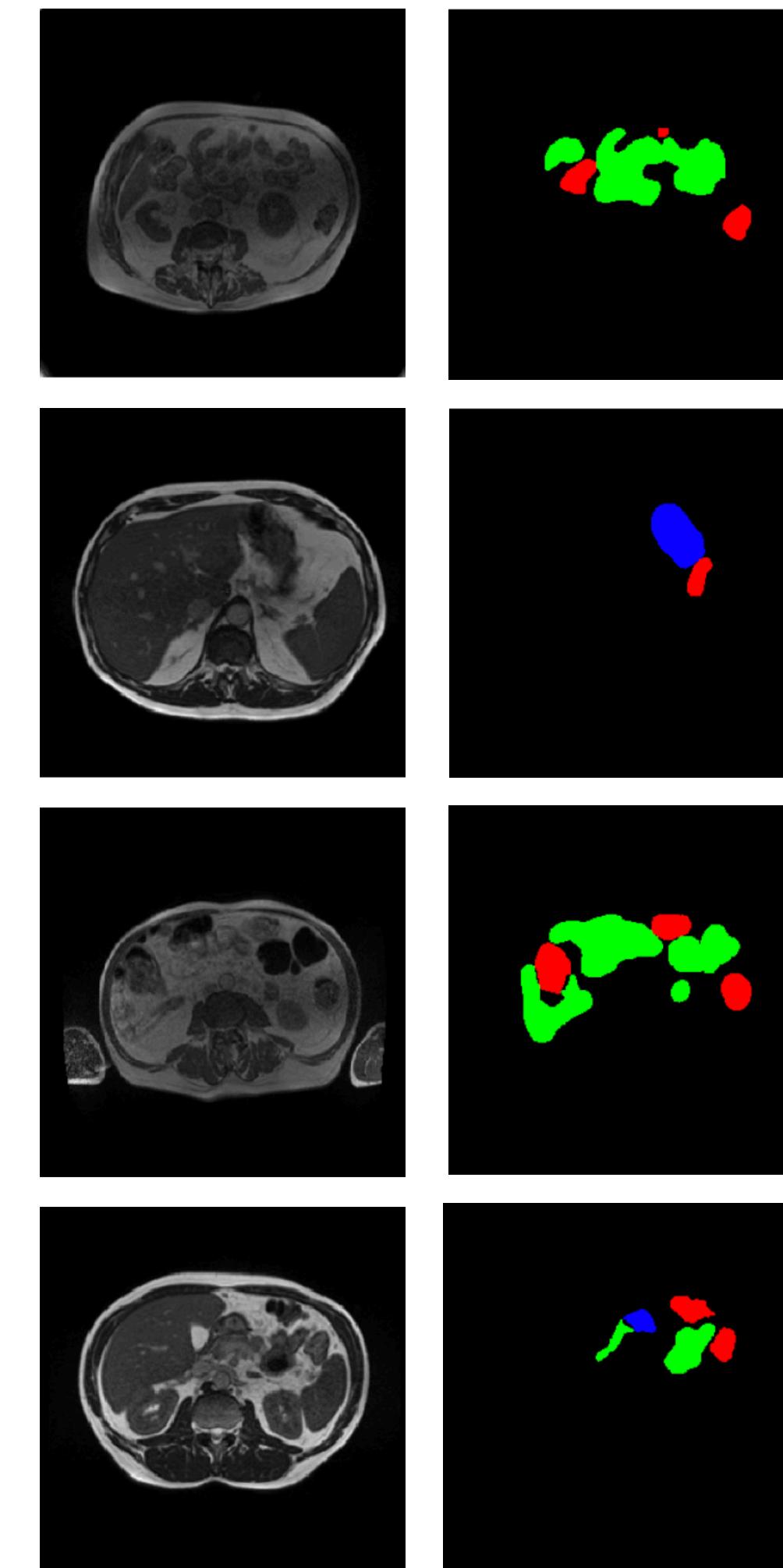
## Exploración

- El conjunto de datos contiene **467 imágenes de resonancia magnética (MRI)** de **107 pacientes**, cada uno sometido a **1-5 exploraciones seriadas** de MRI del abdomen.
- Cada exploración de MRI consiste en aproximadamente **144 cortes axiales** con una resolución promedio de píxeles de **1.5 × 1.5 × 3 mm**.

### Data Explorer

4.14 GB

```
• dataset
  • case101
    • case101_day20
      • contours
      • scans
      • thumbnail.png
    • case101_day22
    • case101_day26
    • case101_day32
  • case102
  • case105
  • case107
```



## Lectura

## Transformaciones

Rotaciones, traslaciones, escalamiento, reflejos horizontales y verticales, cambios de brillo y contraste

## Normalización

# Implementación y entrenamiento

## Herramientas

### Implementación



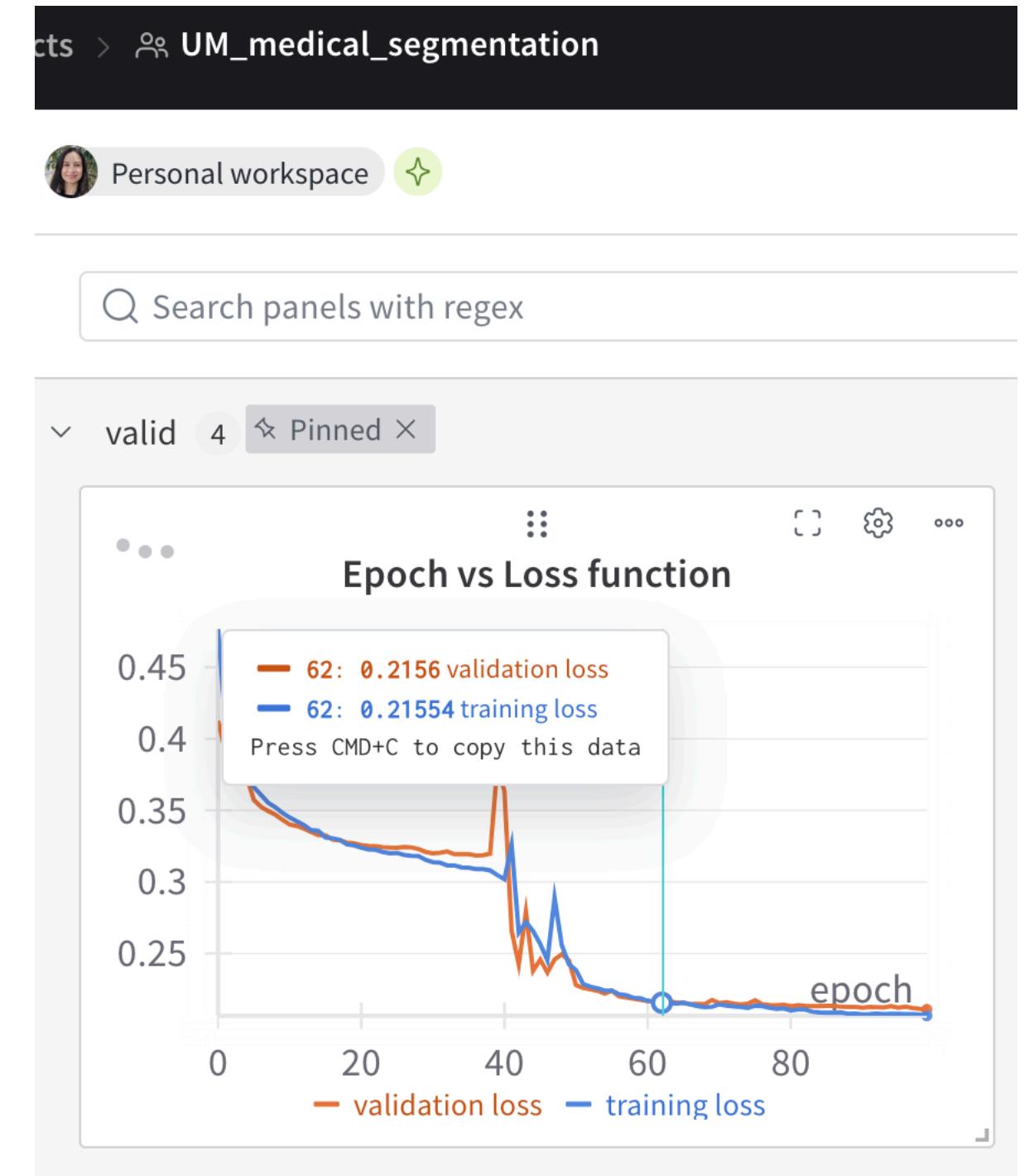
### Finetuning



### Hugging Face

A screenshot of the Hugging Face Model Hub. At the top, there's a search bar and a navigation menu with links like 'Models', 'Datasets', 'Spaces', 'Posts', 'Docs', 'Enterprise', 'Pricing', 'Log In', and 'Sign Up'. Below the menu, a specific model card is displayed for 'nvidia/segformer-b4-finetuned-ade-512-512'. The card includes details such as 'Image Segmentation', 'Transformers', 'PyTorch', 'TensorFlow', 'scene\_parse\_150', 'segformer', 'vision', 'Inference Endpoints', 'arxiv:2105.15203', and 'License: other'. It also shows a 'Model card', 'Files and versions', and a 'Community' section with 4 members. A summary states: 'SegFormer (b4-sized) model fine-tuned on ADE20k'. It mentions that the SegFormer model was fine-tuned on ADE20k at resolution 512x512. It was introduced in the paper 'SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers' by Xie et al. and first released in this repository. Downloads last month are listed as 9,842, accompanied by a small line graph. There are sections for 'Inference Providers' and 'Image Segmentation'.

### Control del entrenamiento



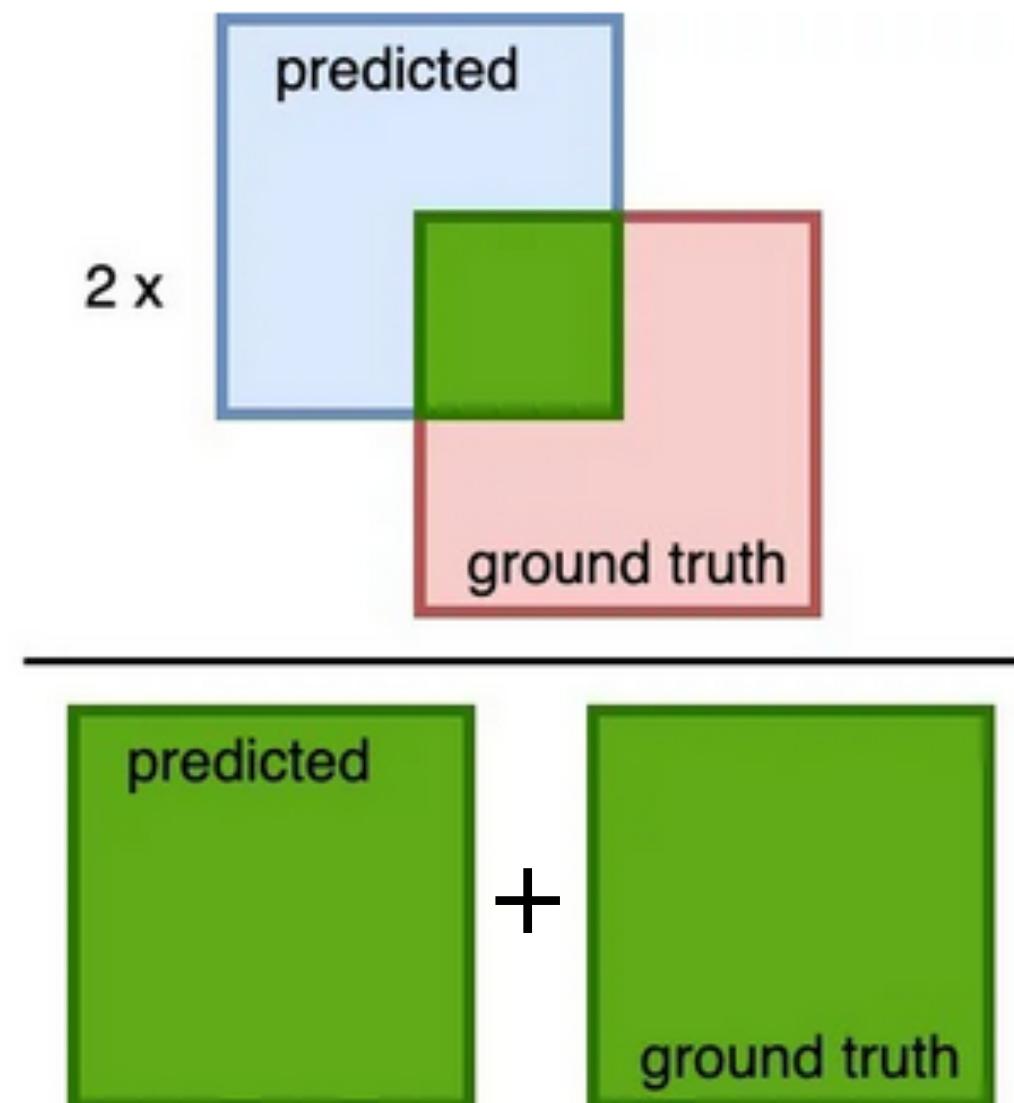
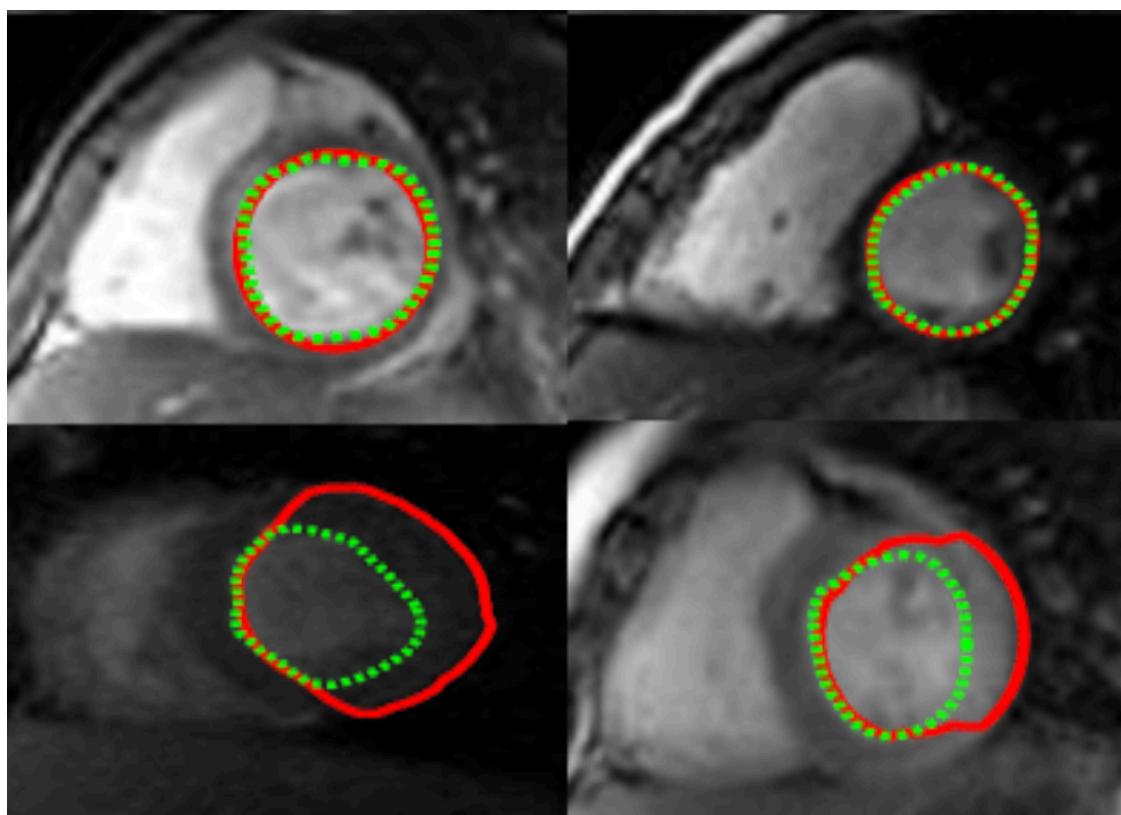
# Hiperparámetros

Hiperparámetro	Valor
Número de clases	4
Tamaño de la imagen	288 x 288
Tamaño de batch	8
Número de épocas	100
Tasa de aprendizaje inicial	3E-04
Optimizador	Adam
Model - Finetuning	nvidia/segformer-b4-finetuned-ade-512-512

# Métricas de evaluación y función de pérdida

## Coeficiente Dice

$$\text{Dice coefficient} = \frac{2x \text{ area of overlapped (green)}}{\text{total area (green)}} =$$



## Cross-Entropy

$$-\sum_{j=1}^M y_j \log(p(y_j))$$

Indicator variable

$\sum_{i=1}^N y_i \log(p(y_i)) + (1 - y_i) \log(1 - p(y_i))$

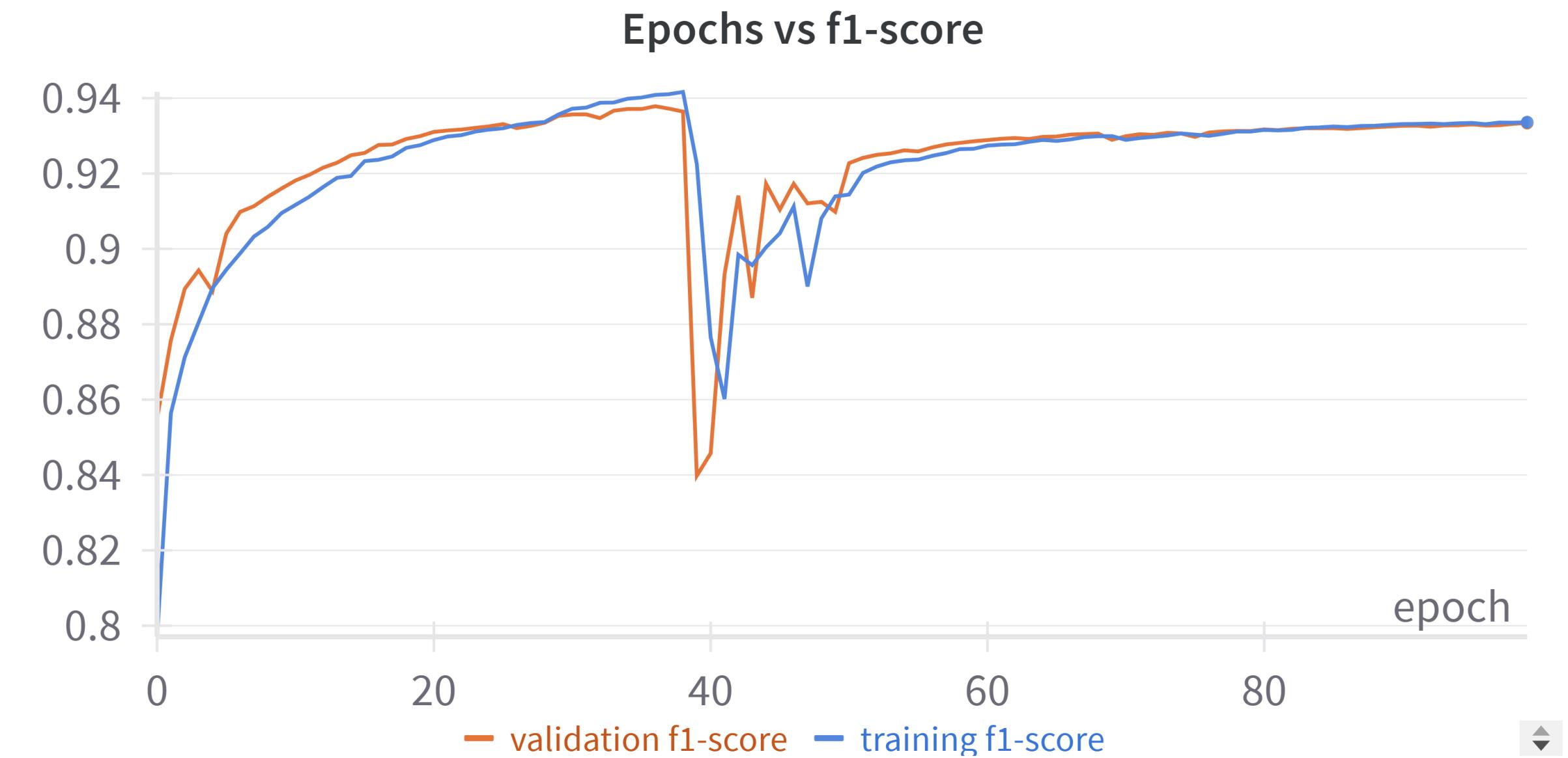
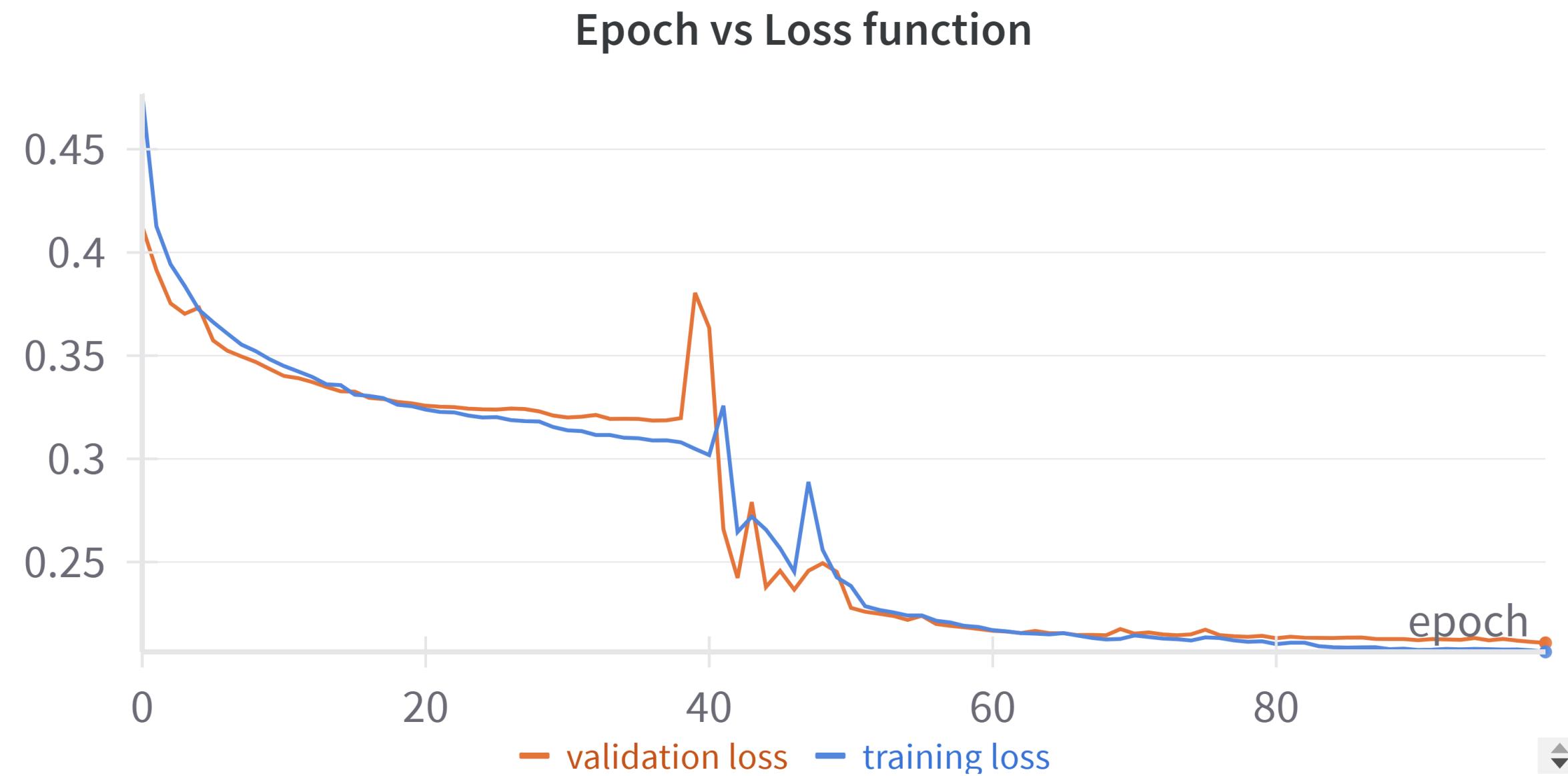
Sum over trials      Sum over classes

Label      Prob of positive class

Label      Prob of positive class

$$\text{Loss} = 1 - \text{Coeficiente Dice} + \text{Cross-Entropy}$$

# Resultados en train y test

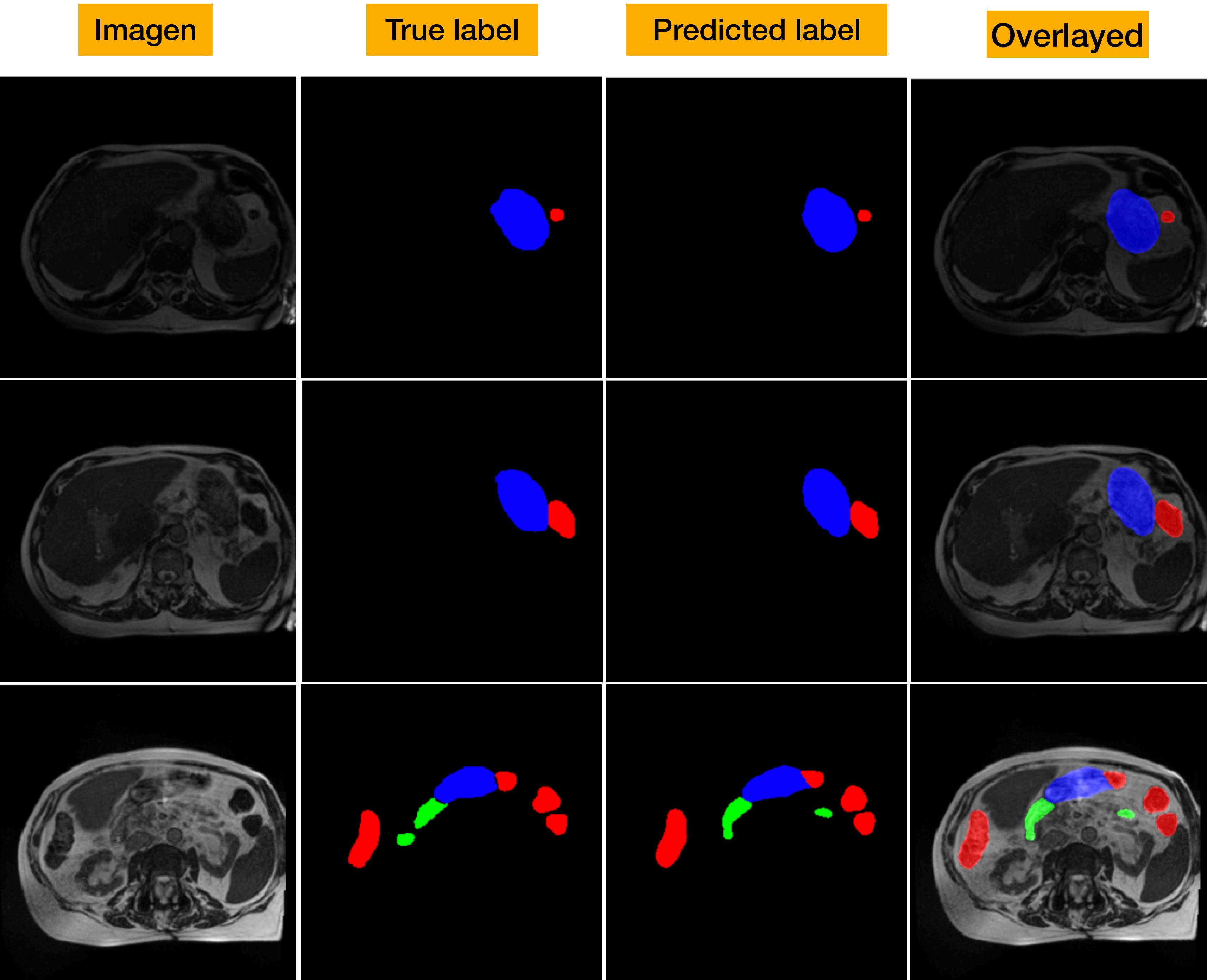


Métrica	Train	Validación
<b>Loss</b>	0.20646	0.31851
<b>F1-score</b>	0.93363	0.93786

Como referencia, con U-Net 2D- 3D (CNN) los valores f1-score son bastante similares.

# Predicciones

Color	Etiqueta
Azul	Estómago
Rojo	Intestino grueso
Verde	Intestino delgado
Negro	Fondo



# Conclusiones

1. **Desempeño Consistente:** El F1-Score en validación (0.93338) es muy similar al de entrenamiento (0.93363), lo que sugiere que el **modelo generaliza bien y no sufre de sobreajuste.**
2. **Pérdida baja:** Las pérdidas obtenidas en entrenamiento y validación son bajas (0.20646 y 0.21081, respectivamente), lo que indica que el modelo está bien optimizado.
3. **Fine-tuning Exitoso:** La combinación del coeficiente de Dice y cross-entropy, junto con el uso de **SegFormer**, permitió un entrenamiento efectivo para lograr segmentaciones precisas.

# Bibliografía



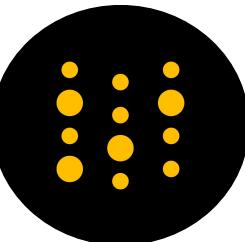
Xie, Enze, et al. "SegFormer: Simple and efficient design for semantic segmentation with transformers." *Advances in neural information processing systems* 34 (2021): 12077-12090.



He, Kelei, et al. "Transformers in medical image analysis." *Intelligent Medicine* 3.1 (2023): 59-78.



<https://github.com/evelyncueva/DeepLearningCourse>



[https://wandb.ai/evelyngcuevaj-escuela-polit-cnica-nacional/UM\\_medical\\_segmentation](https://wandb.ai/evelyngcuevaj-escuela-polit-cnica-nacional/UM_medical_segmentation)

# **Gracias**