

Segmentación Automática de Regiones de Fluido Patológico en Imágenes OCT usando UNet

Guillermo García Engelman, Antonio Serrano Rodríguez
Trabajo de Prácticas – Visión por Computador Avanzada
Grado en Inteligencia Artificial

Resumen—La segmentación de fluido patológico en imágenes de Tomografía de Coherencia Óptica (OCT) es una tarea relevante en el diagnóstico oftalmológico. En este informe se propone una metodología basada en UNet, evaluando distintas funciones de pérdida y mejoras técnicas como Batch Normalization y data augmentation. Se comparan experimentalmente distintas configuraciones usando métricas de evaluación relevantes y se discuten los resultados obtenidos.

I. INTRODUCCIÓN

La OCT permite capturar imágenes detalladas de la retina de forma no invasiva. La presencia de fluido patológico es un indicador importante de enfermedades oculares como la DMAE. La segmentación automática de estas regiones puede acelerar el diagnóstico y reducir la carga de trabajo de los especialistas.

II. DESCRIPCIÓN DEL DATASET

El conjunto de datos proporcionado consta de 50 imágenes OCT en escala de grises, con sus correspondientes máscaras binarias que indican las regiones con fluido patológico. Se realiza una división 70/15/15 en conjuntos de entrenamiento, validación y test. Se aplica una resolución fija de 416x624.

III. METODOLOGÍA

III-A. Arquitectura

Se emplea la red UNet debido a su eficacia comprobada en segmentación biomédica. Su estructura simétrica encoder-decoder con conexiones de salto facilita la preservación de detalles espaciales.

III-B. Funciones de Pérdida

Se evaluaron varias funciones:

- **Weighted BCE**: pondera la clase minoritaria.
- **DiceLoss**: basada en la superposición de máscaras.
- **BCEDiceLoss**: combinación de BCE y Dice.
- **FocalLoss**: enfoca en muestras difíciles.
- **TverskyLoss**: generaliza Dice ponderando FP y FN.

III-C. Mejoras Propuestas

- **Batch Normalization**: mejora la estabilidad durante el entrenamiento.
- **Data Augmentation**: se aplicaron rotaciones leves, espejado horizontal, ajustes de brillo/contraste y desenfoque.
- **Early Stopping**: evita sobreentrenamiento utilizando la pérdida de validación como criterio de parada.

IV. EXPERIMENTOS Y RESULTADOS

A continuación se detallan los resultados obtenidos para cada configuración de red. En cada caso se incluyen un resumen de la estrategia, descripción del entrenamiento, tabla de métricas y análisis.

IV-A. UNet + Weighted BCE

Resumen: Modelo base utilizando Weighted Binary Cross-Entropy para compensar el desbalance de clases.

Entrenamiento: Convergencia inestable. Dice en validación se mantuvo en cero durante 20 épocas. Mejora progresiva a partir de la época 21.

Cuadro I
MÉTRICAS - UNET + WEIGHTED BCE

Precision	Recall	IoU	Dice	Specificity
0.3650	0.7429	0.2903	0.4346	0.9861

Análisis: Alta sensibilidad pero con muchas falsas detecciones. Segmentaciones ruidosas.

IV-B. UNet + DiceLoss

Resumen: Se emplea la función de pérdida DiceLoss, basada en la superposición entre la máscara predicha y la real, lo que la hace especialmente adecuada para tareas de segmentación con clases desbalanceadas.

Entrenamiento: El modelo mostró una convergencia rápida, alcanzando valores de Dice superiores a 0,80 en el conjunto de validación antes de la época 20. El entrenamiento fue estable y sin oscilaciones significativas.

Cuadro II
MÉTRICAS - UNET + DICELOSS

Precision	Recall	IoU	Dice	Specificity
0.8123	0.8893	0.7355	0.8374	0.9987

Análisis: El modelo ofrece un buen equilibrio entre precisión y sensibilidad. Las segmentaciones resultantes son limpias y bien ajustadas a las regiones reales de fluido, con pocos falsos positivos y negativos. Es una de las configuraciones más estables y efectivas entre las evaluadas.

IV-C. UNet + BCE + Dice

Resumen: Esta configuración combina Binary Cross-Entropy (BCE), que penaliza píxeles individuales incorrectos, con DiceLoss, que favorece la superposición global de las máscaras. La combinación busca aprovechar la sensibilidad del BCE con la robustez del Dice frente al desbalance de clases.

Entrenamiento: El modelo presentó un estancamiento inicial durante las primeras 10 épocas. A partir de ahí, mostró una mejora progresiva y sostenida en las métricas de validación, alcanzando una convergencia razonable sin grandes oscilaciones.

Cuadro III
MÉTRICAS - UNET + BCE + DICE

Precision	Recall	IoU	Dice	Specificity
0.7273	0.8374	0.6254	0.7350	0.9976

Análisis: Esta configuración mejora notablemente respecto al modelo base con Weighted BCE. La cobertura de las regiones de fluido es adecuada, aunque con una ligera tendencia a la sobredetección, lo que sugiere una sensibilidad algo elevada a cambio de una precisión moderada.

IV-D. UNet + Focal Loss

Resumen: Se utiliza la función Focal Loss, diseñada para abordar el desbalance de clases acentuando el peso de los ejemplos difíciles y penalizando con mayor intensidad los errores cometidos sobre la clase minoritaria.

Entrenamiento: El proceso de entrenamiento fue lento. Durante las primeras 25 épocas no se observó una mejora significativa en la métrica Dice de validación. A partir de ese punto, el modelo comenzó a generalizar mejor, aunque sin alcanzar el rendimiento de otras configuraciones.

Cuadro IV
MÉTRICAS - UNET + FOCAL LOSS

Precision	Recall	IoU	Dice	Specificity
0.8958	0.6027	0.5745	0.7046	0.9996

Análisis: El modelo presenta una alta precisión, lo que indica una baja tasa de falsos positivos. Sin embargo, la sensibilidad es baja, lo que implica que muchas regiones de fluido no fueron detectadas. El resultado es una segmentación conservadora, adecuada en contextos donde se prioriza evitar sobresegmentación, pero subóptima para una cobertura completa.

IV-E. UNet + Tversky Loss

Resumen: En esta variante se utiliza Tversky Loss, una generalización de Dice que permite ponderar de forma asimétrica los falsos positivos y los falsos negativos. Es especialmente útil cuando se requiere ajustar el balance entre sensibilidad y precisión.

Entrenamiento: El modelo mostró una mejora progresiva a partir de la época 12, alcanzando una convergencia estable. No se observaron oscilaciones bruscas en las métricas de validación, lo que sugiere una buena estabilidad durante el entrenamiento.

Cuadro V
MÉTRICAS - UNET + TVERSKY LOSS

Precision	Recall	IoU	Dice	Specificity
0.8164	0.8297	0.7203	0.8198	0.9987

Análisis: Esta configuración logra un equilibrio sólido entre precisión y sensibilidad. Las máscaras segmentadas son precisas y con pocos errores de predicción. Es una de las opciones más consistentes del estudio, ofreciendo buenos resultados sin necesidad de mejoras adicionales como augmentación o normalización.

IV-F. UNet + Tversky + Data Augmentation

Resumen: Se emplea Tversky Loss junto con técnicas de data augmentation para aumentar la diversidad del conjunto de entrenamiento. Se aplicaron transformaciones como rotaciones suaves, espejado horizontal, ajustes de brillo/contraste y desenfoque leve.

Entrenamiento: El entrenamiento mostró una convergencia estable, con un desempeño óptimo en validación alcanzado en torno a la época 28. Las transformaciones no introdujeron oscilaciones ni degradación del rendimiento.

Cuadro VI
MÉTRICAS - UNET + TVERSKY + DATA AUGMENTATION

Precision	Recall	IoU	Dice	Specificity
0.7158	0.9223	0.6746	0.7727	0.9973

Análisis: El modelo presentó una sensibilidad muy alta, siendo capaz de detectar la mayoría de las regiones de fluido. No obstante, esta mejora vino acompañada de un leve incremento en los falsos positivos, lo que introdujo algo de ruido en las segmentaciones. Es una configuración útil cuando se prioriza la detección completa.

IV-G. UNet + Tversky + Batch Normalization

Resumen: Se introduce Batch Normalization en la arquitectura UNet con Tversky Loss, con el objetivo de mejorar la estabilidad del entrenamiento y acelerar la convergencia mediante la normalización intermedia de activaciones.

Entrenamiento: El progreso inicial fue lento, pero a partir de la época 15 se observó una mejora significativa en las métricas de validación. El entrenamiento fue estable, y las segmentaciones fueron progresivamente más limpias.

Cuadro VII
MÉTRICAS - UNET + TVERSKY + BATCH NORMALIZATION

Precision	Recall	IoU	Dice	Specificity
0.9128	0.7004	0.6277	0.7442	0.9988

Análisis: Esta configuración mostró una excelente precisión, aunque con una sensibilidad más baja. Las segmentaciones fueron conservadoras, lo que reduce los falsos positivos, pero puede dejar algunas regiones sin detectar. Es ideal en escenarios donde se requiere alta fiabilidad en cada predicción positiva.

IV-H. UNet + Tversky + Data Augmentation + Batch Normalization

Resumen: Configuración combinada que incluye tanto data augmentation como batch normalization sobre el modelo basado en Tversky Loss. El objetivo es aprovechar simultáneamente la robustez frente al desbalance, la mejora en la generalización y la estabilidad del entrenamiento.

Entrenamiento: El modelo comenzó a mostrar mejoras a partir de la época 10. A lo largo del entrenamiento, las métricas de validación fueron estables, con una evolución coherente hasta la época 66.

Cuadro VIII
MÉTRICAS - UNET + TVERSKY + DATA AUGMENTATION + BATCH NORMALIZATION

Precision	Recall	IoU	Dice	Specificity
0.8003	0.8266	0.6681	0.7931	0.9985

Análisis: El modelo alcanzó un equilibrio muy adecuado entre precisión y sensibilidad. Las máscaras generadas fueron limpias y detalladas, sin ruido significativo. Esta configuración se posiciona como una opción robusta y versátil en contextos clínicos donde se requiere tanto cobertura como fiabilidad.

IV-I. UNet + DiceLoss + Data Augmentation

Resumen: Se emplea DiceLoss junto con técnicas de data augmentation para mejorar la capacidad de generalización del modelo frente a un conjunto de datos reducido. Las transformaciones incluyen rotaciones leves, reflejos horizontales y ajustes de brillo y contraste.

Entrenamiento: El entrenamiento mostró una mejora continua desde la época 6, sin oscilaciones importantes. La convergencia se alcanzó alrededor de la época 43, con un rendimiento estable en validación.

Cuadro IX
MÉTRICAS - UNET + DICELOSS + DATA AUGMENTATION

Precision	Recall	IoU	Dice	Specificity
0.7386	0.7839	0.6424	0.7563	0.9980

Análisis: El modelo mostró una buena cobertura de las regiones de fluido patológico. Las segmentaciones fueron razonablemente precisas, aunque se observó un leve suavizado en los bordes y una ligera presencia de falsos positivos. Es una configuración sólida cuando se desea un rendimiento estable sin añadir complejidad estructural.

IV-J. UNet + DiceLoss + Batch Normalization

Resumen: Se combina DiceLoss con batch normalization para mejorar la estabilidad y la eficiencia del entrenamiento. Esta normalización de activaciones contribuye a una convergencia más robusta y a una menor sensibilidad a la inicialización.

Entrenamiento: El modelo presentó una evolución estable a lo largo de las épocas. En múltiples ocasiones se superó el valor de 0,87 en la métrica Dice en validación. La convergencia se produjo hacia la época 77, sin signos de sobreajuste.

Cuadro X
MÉTRICAS - UNET + DICELOSS + BATCH NORMALIZATION

Precision	Recall	IoU	Dice	Specificity
0.9110	0.8584	0.7833	0.8756	0.9989

Análisis: Esta fue la configuración con mejor rendimiento de todo el conjunto experimental. La segmentación fue muy precisa, sin presencia significativa de ruido, y con una excelente superposición con las máscaras reales. Es la opción más recomendable cuando se busca rendimiento máximo sin aplicar técnicas de augmentación.

IV-K. UNet + DiceLoss + Data Augmentation + Batch Normalization

Resumen: Configuración más completa, que incorpora DiceLoss, técnicas de data augmentation y batch normalization. El objetivo es maximizar la generalización y la estabilidad del entrenamiento combinando mejoras estructurales y de regularización.

Entrenamiento: La fase inicial fue inestable, con fluctuaciones en las métricas de validación. Sin embargo, a partir de la época 10 se observó una mejora progresiva sostenida, alcanzando estabilidad y buen rendimiento general alrededor de la época 67.

Cuadro XI
MÉTRICAS - UNET + DICELOSS + DATA AUGMENTATION + BATCH NORMALIZATION

Precision	Recall	IoU	Dice	Specificity
0.9002	0.7920	0.7151	0.8319	0.9990

Análisis: Este modelo logró una excelente precisión y una buena sensibilidad, generando segmentaciones limpias, bien ajustadas y sin exceso de ruido. Aunque no alcanzó los máximos valores del mejor modelo con BatchNorm solo, su comportamiento general fue muy robusto, especialmente útil en entornos clínicos más variados.

V. DISCUSIÓN

La elección de la función de pérdida resulta clave para manejar el desbalance de clases. Si bien TverskyLoss y BCEDiceLoss ofrecen resultados razonables, DiceLoss sola con normalización por lotes fue la combinación más efectiva.

Curiosamente, el uso de data augmentation no siempre mejoró el rendimiento.

VI. CONCLUSIONES

- DiceLoss demostró ser la función más robusta para el problema.
- BatchNorm mejora el rendimiento significativamente.
- Data Augmentation no fue siempre beneficioso, posiblemente por el reducido tamaño de datos.
- UNet es una arquitectura eficaz incluso en escenarios con pocos datos.

VII. TRABAJO FUTURO

Aplicar técnicas como transfer learning, ajustar los umbrales de segmentación con validación cruzada, y explorar arquitecturas más modernas como Attention UNet o TransUNet.

REFERENCIAS

- [1] Ronneberger et al., *U-Net: Convolutional Networks for Biomedical Image Segmentation*, MICCAI 2015.
- [2] Salehi et al., *Tversky Loss Function for Image Segmentation Using 3D Fully Convolutional Deep Networks*, MICCAI 2017.
- [3] Lin et al., *Focal Loss for Dense Object Detection*, ICCV 2017.