

Análisis de Redes Convolucionales y Técnicas de Explicabilidad para Arquitecturas Neuronales Profundas para la Segmentación de Instancias

Trabajo de Fin de Grado

Autor: Juan Manuel Rodríguez Gómez

Tutores: Francisco Javier Merí de la Maza, Pablo Mesejo Santiago,
Marilyn Bello García

26 de Junio de 2025

Doble Grado en Ingeniería Informática y Matemáticas



UNIVERSIDAD
DE GRANADA

Primera Parte: Análisis de Redes Convolucionales

- ① Introducción
- ② Teoría de Mallat sobre Redes de Dispersión
- ③ Generalización de la Teoría de Mallat

Redes neuronales convolucionales

- Gran rendimiento empírico.
- Bases teóricas aún en desarrollo.
- Buscamos profundizar en su **modelización matemática**.

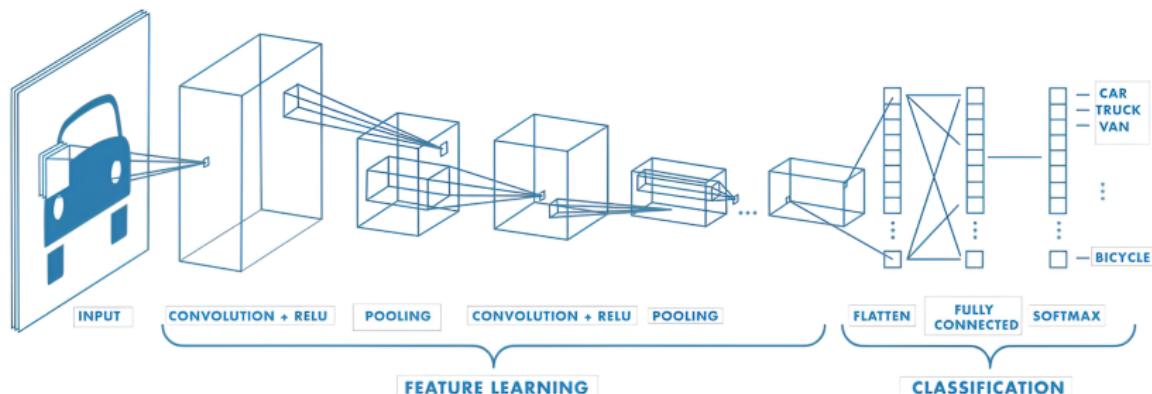


Figura 1: Diagrama de una red neuronal convolucional [Mat24b].

[Mat24b] MathWorks. Redes neuronales convolucionales. <https://es.mathworks.com/discovery/convolutional-neural-network.html>, 2024.

Primera propiedad: Invarianza por traslaciones

Definición

Sea \mathcal{H} un espacio de Hilbert. $\Phi : L^2(\mathbb{R}^d) \rightarrow \mathcal{H}$ es **invariante por traslaciones** si:

$$\Phi(T_t f) = \Phi(f), \quad \forall f \in L^2(\mathbb{R}^d), \quad \forall t \in \mathbb{R}^d,$$

donde $(T_t f)(x) := f(x - t)$.

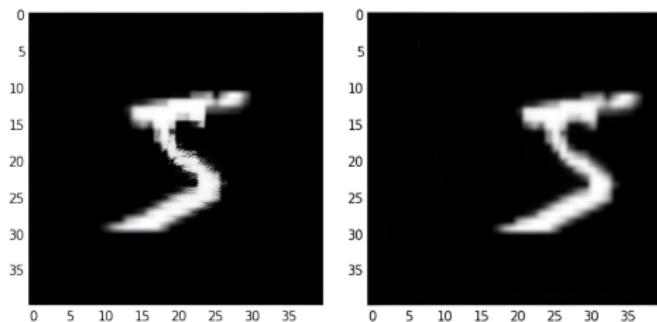


Figura 2: El dígito 5 se reconoce pese a su cambio de posición [KA18].

[KA18] E. Kauderer-Abrams. Quantifying translation-invariance in convolutional neural networks. *CoRR*, abs/1801.01450, 2018.

Segunda propiedad: Invarianza frente a pequeñas deformaciones

Definición

$\Phi : L^2(\mathbb{R}^d) \rightarrow \mathcal{H}$ es **Lipschitz-continuo** por la acción de difeomorfismos C^2 si existe una constante $C > 0$ tal que, para todo $f \in L^2(\mathbb{R}^d)$ y para todo $\tau \in C^2(\mathbb{R}^d)$, se cumple:

$$\|\Phi(f) - \Phi(T_\tau f)\| \leq C\|f\| (\|Jac(\tau)\|_\infty + \|Hess(\tau)\|_\infty),$$

donde $(T_\tau f)(x) := f(x - \tau(x))$.

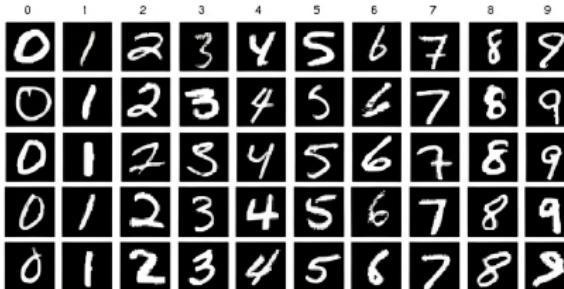


Figura 3: Los dígitos se reconocen pese a variaciones en su forma [LYP16].

[LYP16] S. Lim, S. Young, y R. Patton. An analysis of image storage systems for scalable training of deep neural networks. *ResearchGate*, 2016.

Objetivo

Desarrollar un operador Φ que sea invariante por translaciones y frente a pequeñas deformaciones, y que modelice el comportamiento de una red neuronal convolucional.

Limitaciones de la transformada de Fourier

Definición

La transformada de Fourier de $f \in L^1(\mathbb{R}^d)$ es:

$$\widehat{f}(\omega) := \int_{\mathbb{R}^d} f(x) e^{-2\pi i \langle x, \omega \rangle} dx$$

El **módulo** de la transformada de Fourier:

- Es invariante por traslaciones: $|\widehat{T_t f}| = |\widehat{f}|$.
- No es invariante frente a pequeñas deformaciones \Rightarrow Debemos buscar otro operador.

Ondículas como alternativa a la transformada de Fourier

Haremos uso de una **ondícula madre** del siguiente tipo:

$$\psi(x) = e^{i\langle x, \eta \rangle} \Theta(x)$$

- $\Theta(x)$: Función real con soporte en una bola de baja frecuencia centrada en $x = 0$ y con radio del orden de π .
- η : Vector que define la frecuencia central de la ondícula.

Definición

Una **ondícula madre escalada por un factor 2^j** , con $j \in \mathbb{Z}$, y **rotada por $r \in G$** , siendo G el grupo finito de rotaciones, se define como:

$$\psi_{2^j r}(x) = 2^j \psi(2^j r^{-1} x).$$

Transformada de Littlewood-Paley

Generamos la siguiente base ortonormal de ondículas:

$$\{\psi_\lambda(x)\}_{\lambda=2^j r \in 2^{\mathbb{Z}} \times G}$$

Definición

Sea $f \in L^2(\mathbb{R}^d)$ y sea $\{\psi_\lambda\}_{\lambda \in 2^{\mathbb{Z}} \times G}$ una familia de ondículas. Para cada $\lambda = 2^j r$, la **transformada de Littlewood-Paley** de f es:

$$(W[\lambda]f)(x) := (f * \psi_\lambda)(x) = \int_{\mathbb{R}^d} f(u) \psi_\lambda(x - u) du, \quad \forall x \in \mathbb{R}^d.$$

El operador $W[\lambda]f$ es Lipschitz-continuo bajo la acción de difeomorfismos pero **no es invariante por traslaciones**.

Problema con la transformada de Littlewood-Paley: La escala

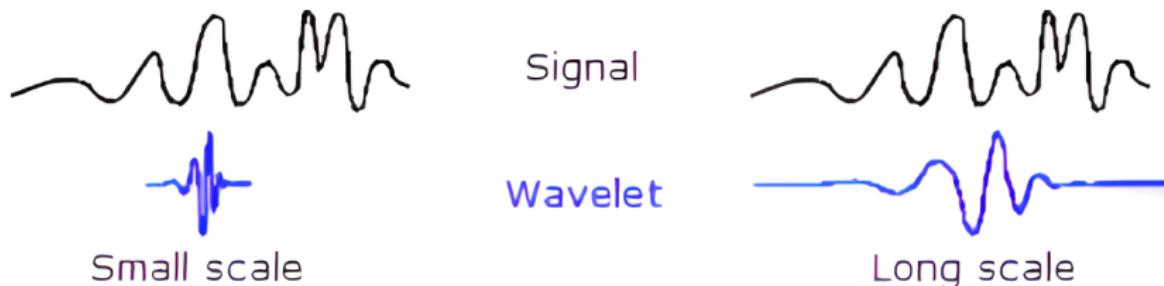


Figura 4: Ondículas a menor escala detectan mejor los cambios rápidos en la señal; a mayor escala, captan variaciones más globales [Mat24a].

A una cierta escala 2^J , con $J \in \mathbb{Z}$ fijo, solo se conservan las ondículas afectadas por un factor de escala $2^j > 2^{-J}$.

[Mat24a] MathWorks. Continuous wavelet transform and scale-based analysis. <https://es.mathworks.com/help/wavelet/gs/continuous-wavelet-transform-and-scale-based-analysis.html>, 2024.

Solución: Promediado de bajas frecuencias

Esto genera un umbral a partir del cual la base ortonormal de ondículas no podría identificar cambios de frecuencia.

Solución: Promediar las bajas frecuencias que no son captadas por estas ondículas en un dominio proporcional al factor 2^J .

$$A_J f = f * \phi_{2^J}, \quad \text{con} \quad \phi_{2^J}(x) = 2^{-J} \phi(2^{-J}x).$$

La transformada de ondículas, fijada una escala, es:

$$W_J f = \left\{ A_J f, (W[\lambda]f)_{\lambda \in \Lambda_J} \right\}$$

con $\Lambda_J = \{\lambda = 2^j r : r \in G^+, 2^j > 2^{-J}\}$, siendo G^+ el cociente de G sobre $\{-1, 1\}$.

Convenciones para el posterior desarrollo de este trabajo

- Trabajaremos en el espacio de funciones reales de $L^2(\mathbb{R}^d)$.
- $\widehat{\psi}$ es una función real que satisface $\widehat{\psi}(0) = \int \psi(x) dx = 0$ y $|\widehat{\psi}(r\omega)| = |\widehat{\psi}(\omega)|$, $\forall r \in G \Rightarrow W_J$ es unitario (preserva la norma y el producto escalar).
- $\widehat{\phi}(\omega)$ es real y simétrica, por lo que ϕ también lo es.
- Las derivadas de ϕ pertenecen a $L^1(\mathbb{R}^d)$.

Operador $U[\lambda]f$

Proposición

Si $U[\lambda]$ es un operador definido en $L^2(\mathbb{R}^d)$, no necesariamente lineal pero que commuta con traslaciones, entonces $\int_{\mathbb{R}^d} (U[\lambda]f)(x) dx$ es invariante por traslaciones si es finito.

Definimos entonces el operador:

$$U[\lambda]f = |f * \psi_\lambda|.$$

Propagador de dispersión

Un conjunto de elementos ordenados $p = (\lambda_1, \lambda_2, \dots, \lambda_m)$ con $\lambda_k \in \Lambda_\infty = 2\mathbb{Z} \times G^+$ es un **camino**. $p = \emptyset$ es el camino vacío.

Definición

Un **propagador de dispersión** es una composición de operadores no conmutativos $U[\lambda]$ ordenada por un camino.

$$U[p]f = U[\lambda_m] \cdots U[\lambda_2] U[\lambda_1]f = ||| f * \psi_{\lambda_1} | * \psi_{\lambda_2} | \cdots * \psi_{\lambda_m} |,$$

siendo $U[\emptyset] = \text{Id}$.

Tenemos que $\int_{\mathbb{R}^d} (U[p]f)(x) dx < \infty$.

Transformada de dispersión de ventana

Sea $J \in \mathbb{Z}$ y \mathcal{P}_J el conjunto de caminos finitos
 $p = (\lambda_1, \lambda_2, \dots, \lambda_m)$, con $\lambda_k \in \Lambda_J$ y $|\lambda_k| = 2^k > 2^{-J}$.

Definición

Definimos la **transformada de dispersión de ventana** como:

$$(S_J[p]f)(x) = ((U[p]f) * \phi_{2^J})(x), \quad \forall p \in \mathcal{P}_J.$$

Es decir:

$$(S_J[p]f)(x) = (||| f * \psi_{\lambda_1} | * \psi_{\lambda_2} | \cdots * \psi_{\lambda_m} | * \phi_{2^J}) (x).$$

En particular, $S_J[\emptyset]f = f * \phi_{2^J}$.

Comparativa con una red neuronal convolucional

- **Similitudes:**
 - Convolución: operador $W[\lambda]$.
 - Función de activación: Módulo.
 - Si p tiene longitud m : los coeficientes $S_J[p]f$ corresponden a los mapas de activación tras la capa m de la red.
- **Diferencias:**
 - En las redes neuronales convoluciones, los pesos se aprenden por retropropagación. En esta modelización, las ondículas son fijas y no se aprenden.
 - No se incluyen capas de *pooling* en esta modelización.

Ondículas admisibles

Definición

Una ondícula de dispersión ψ se considera **admisible** si existe $\eta \in \mathbb{R}^d$ y $\rho \geq 0$, con $|\widehat{\rho}(\omega)| \leq |\widehat{\phi}(2\omega)|$ y $\widehat{\rho}(0) = 1$, tal que $\widehat{\Psi}(\omega)$ satisface:

$$\alpha = \inf_{1 \leq |\omega| \leq 2} \sum_{j=-\infty}^{\infty} \sum_{r \in G} \widehat{\Psi}(2^{-j} r^{-1} \omega) \left| \widehat{\psi}(2^{-j} r^{-1} \omega) \right|^2 > 0.$$

Conservación de la norma y no-expansividad en conjuntos de caminos de la transformada de dispersión de ventana

Proposición

Si la ondícula es admisible, entonces:

$$\lim_{m \rightarrow \infty} \sum_{n=m}^{\infty} \|S_J[\Lambda_J^n]f\|^2 = 0, \quad y, \quad \|S_J[\mathcal{P}_J]f\| = \|f\|, \quad \forall f \in L^2(\mathbb{R}^d).$$

Conservación de la norma y no-expansividad en conjuntos de caminos de la transformada de dispersión de ventana

Proposición

Si la ondícula es admisible, entonces:

$$\lim_{m \rightarrow \infty} \sum_{n=m}^{\infty} \|S_J[\Lambda_J^n]f\|^2 = 0, \quad y, \quad \|S_J[\mathcal{P}_J]f\| = \|f\|, \quad \forall f \in L^2(\mathbb{R}^d).$$

Proposición

Para todo $f, h \in L^2(\mathbb{R}^d)$ y $J \in \mathbb{Z}$ se cumple que:

$$\|S_{J+1}[\mathcal{P}_{J+1}]f - S_{J+1}[\mathcal{P}_{J+1}]h\| \leq \|S_J[\mathcal{P}_J]f - S_J[\mathcal{P}_J]h\|.$$

Invarianza horizontal por traslaciones y frente a pequeñas deformaciones del operador $S_J[p]f$

Teorema

Para ondículas de dispersión admisibles, se verifica que:

$$\lim_{J \rightarrow \infty} \|S_J[\mathcal{P}_J]f - S_J[\mathcal{P}_J](T_t f)\| = 0, \quad \forall f \in L^2(\mathbb{R}^d), \quad \forall t \in \mathbb{R}^d.$$

Invarianza horizontal por traslaciones y frente a pequeñas deformaciones del operador $S_J[p]f$

Teorema

Para ondículas de dispersión admisibles, se verifica que:

$$\lim_{J \rightarrow \infty} \|S_J[\mathcal{P}_J]f - S_J[\mathcal{P}_J](T_t f)\| = 0, \quad \forall f \in L^2(\mathbb{R}^d), \quad \forall t \in \mathbb{R}^d.$$

Teorema

Existe una constante $C \in \mathbb{R}$ tal que, para todo $f \in L^2(\mathbb{R}^d)$, con $\|U[\mathcal{P}_J]f\|_1 < \infty$, y todo $\tau \in C^2(\mathbb{R}^d)$, con $\|\text{Jac}(\tau)\|_\infty \leq \frac{1}{2}$, se cumple que:

$$\|S_J[\mathcal{P}_J](T_\tau f) - S_J[\mathcal{P}_J]f\| \leq C \|U[\mathcal{P}_J]f\|_1 K(\tau),$$

donde:

$$K(\tau) = 2^{-J} \|\tau\|_\infty + \|\text{Jac}(\tau)\|_\infty \left(\log \left(\frac{\|\Delta\tau\|_\infty}{\|\text{Jac}(\tau)\|_\infty} \vee 1 \right) \right) + \|\text{Hess}(\tau)\|_\infty.$$

Generalización de la teoría de Mallat

Buscamos extender la teoría de Mallat sobre redes de dispersión para abarcar arquitecturas más generales:

- Incluir funciones de activación distintas del módulo (ReLU, sigmoide...).
 - Considerar diferentes tipos de filtros (no solo ondículas, también curvículas, crestículas...).
 - Modelizar también las capas de *pooling*.

Marcos semi-discretos

Definición

Sea $\{g_\lambda\}_{\lambda \in \Lambda} \subset L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$ un conjunto de funciones indexado por un conjunto numerable Λ . La colección:

$$\Psi_\Lambda := \{ T_t(lg_\lambda) \}_{t \in \mathbb{R}^d, \lambda \in \Lambda}$$

es un **marco semi-discreto** para $L^2(\mathbb{R}^d)$ si existen constantes $A, B > 0$ tales que:

$$A\|f\|^2 \leq \sum_{\lambda \in \Lambda} \|f * g_\lambda\|^2 \leq B\|f\|^2, \quad \forall f \in L^2(\mathbb{R}^d),$$

Las funciones $\{g_\lambda\}_{\lambda \in \Lambda}$ se denominan **átomos** del marco Ψ_Λ , y los valores $\langle f, T_t(Ig_\lambda) \rangle = (f * g_\lambda)(t)$ se llaman **coeficientes** del marco Ψ_Λ .

Operador de *pooling*

Definición

Sea $M_n(f * g_{\lambda_n}) \in L^2(\mathbb{R}^d)$ el resultado de aplicar una no-linealidad Lipschitz-continua sobre una convolución en la capa n-ésima. Esta se puede someter a un proceso de ***pooling*** definido como:

$$M_n(f * g_{\lambda_n}) \mapsto S_n^{d/2} P_n(M_n(f * g_{\lambda_n}))(S_n \cdot),$$

donde:

- $S_n \geq 1$ es el factor de *pooling*.
- P_n es un operador Lipschitz-continuo que satisface que $P_n f = 0$ cuando $f = 0$.

Sucesión de módulos y operador $U_n[\lambda_n]f$

Definición

Una sucesión de módulos es:

$$\Omega := ((\Psi_n, M_n, P_n))_{n \in \mathbb{N}},$$

- Ψ_n es un marco semi-discreto para $L^2(\mathbb{R}^d)$.
- M_n y P_n son operadores Lipschitz-continuos de $L^2(\mathbb{R}^d)$ que preservan el origen.

Sucesión de módulos y operador $U_n[\lambda_n]f$

Definición

Una sucesión de módulos es:

$$\Omega := ((\Psi_n, M_n, P_n))_{n \in \mathbb{N}},$$

- Ψ_n es un marco semi-discreto para $L^2(\mathbb{R}^d)$.
- M_n y P_n son operadores Lipschitz-continuos de $L^2(\mathbb{R}^d)$ que preservan el origen.

Definición

Sea $\Omega = ((\Psi_n, M_n, P_n))_{n \in \mathbb{N}}$ una sucesión de módulos. Sean $\{g_{\lambda_n}\}_{\lambda_n \in \Lambda_n}$ los átomos del marco Ψ_n , y sea $S_n \geq 1$ el factor de *pooling* asociado a la capa n -ésima de la red. Se define:

$$U_n(\lambda_n, f) := U_n[\lambda_n]f := S_n^{d/2} P_n(M_n(f * g_{\lambda_n}))(S_n \cdot).$$

Operador buscado: Extractor de características convolucional general

Definición

Sea $\Omega = ((\Psi_n, M_n, P_n))_{n \in \mathbb{N}}$ una sucesión de módulos. El **extractor de características** basado en Ω es:

$$\Phi_\Omega(f) := \bigcup_{n=0}^{\infty} \Phi_\Omega^n(f),$$

donde:

$$\Phi_\Omega^n(f) := \{(U[q]f) * \chi_n\}_{q \in \Lambda_1^n}, \quad \forall n \in \mathbb{N},$$

siendo χ_n el **átomo generador de salida** de la capa n .

Sucesión de módulos admisible

Definición

Sea $\Omega = ((\Psi_n, M_n, P_n))_{n \in \mathbb{N}}$ una sucesión de módulos, con cotas superiores del marco $B_n > 0$ y constantes de Lipschitz $L_n, R_n > 0$ para los operadores M_n y P_n , respectivamente. La sucesión de módulos se dice que es **admisible** si:

$$\max \{B_n, B_n L_n^2 R_n^2\} \leq 1, \quad \forall n \in \mathbb{N}.$$

Invarianza vertical por traslaciones del operador Φ_Ω

Teorema

Sea $\Omega = ((\Psi_n, M_n, P_n))_{n \in \mathbb{N}}$ una sucesión de módulos admisible, $S_n \geq 1$, $n \in \mathbb{N}$, el factor de *pooling*, y supongamos que los operadores $M_n, P_n : L^2(\mathbb{R}^d) \rightarrow L^2(\mathbb{R}^d)$ conmutan con el operador de traslación T_t . Entonces:

- ① $\Phi_\Omega^n(T_t f) = T_{t/(S_1 \dots S_n)}(\Phi_\Omega^n(f)), \quad \forall f \in L^2(\mathbb{R}^d), t \in \mathbb{R}^d, n \in \mathbb{N}.$
- ② Si además existe una constante $K > 0$ (que no depende de n) tal que:

$$|\widehat{\chi}_n(\omega)| |\omega| \leq K, \quad \text{c.p.d. } \omega \in \mathbb{R}^d, \quad \forall n \in \mathbb{N}_0,$$

entonces:

$$\|\Phi_\Omega^n(T_t f) - \Phi_\Omega^n(f)\| \leq \frac{2\pi|t|K}{S_1 \dots S_n} \|f\|, \quad \forall f \in L^2(\mathbb{R}^d), t \in \mathbb{R}^d.$$

Segunda Parte: Técnicas de Explicabilidad para Arquitecturas Neuronales Profundas para la Segmentación de Instancias

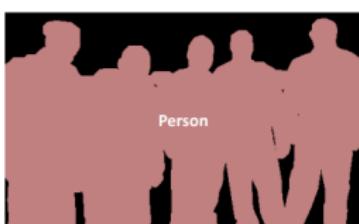
- ① Introducción
- ② Fundamentos y Estado del Arte
- ③ Método Propuesto
- ④ Experimentación
- ⑤ Conclusiones

Segmentación de instancias

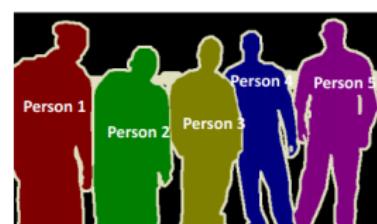
- Identificar y delimitar cada objeto individual en una imagen, proporcionando una máscara específica para cada instancia detectada.



Object Detection



Semantic Segmentation



Instance Segmentation

Figura 1: La segmentación de instancias combina la detección de objetos y la segmentación semántica [Liu18].

[Liu18] Y. Liu. The confusing metrics of AP and MAP for object detection / instance segmentation. <https://yanfengliu.medium.com/the-confusing-metrics-of-ap-and-map-for-object-detection-3113ba0386ef>, 2018.

Inteligencia Artificial Explicable (XAI)

- **Objetivo:** Hacer comprensibles y confiables modelos de IA (especialmente cajas negras).
- **Tipos de explicabilidad:**
 - **Según el momento:**
 - **Intrínseca:** El modelo en sí es explicable (árbol de decisión).
 - **Post-hoc:** Se explica tras el entrenamiento (Grad-CAM).
 - **Según el alcance:**
 - **Global:** Explica todo el modelo.
 - **Local:** Explica predicciones concretas.
- **Propiedades deseables:** Comprensibilidad, fidelidad, estabilidad, justicia.

Objetivos

- ① Estudiar la **segmentación de instancias** y el funcionamiento de Mask R-CNN.
- ② Aplicar Mask R-CNN a **radiografías dentales panorámicas**.
- ③ Investigar el campo de la **Inteligencia Artificial Explicable**.
- ④ Implementar una **adaptación de Grad-CAM** en Mask R-CNN.
- ⑤ **Evaluar** la efectividad de Grad-CAM en Mask R-CNN.

Mask R-CNN

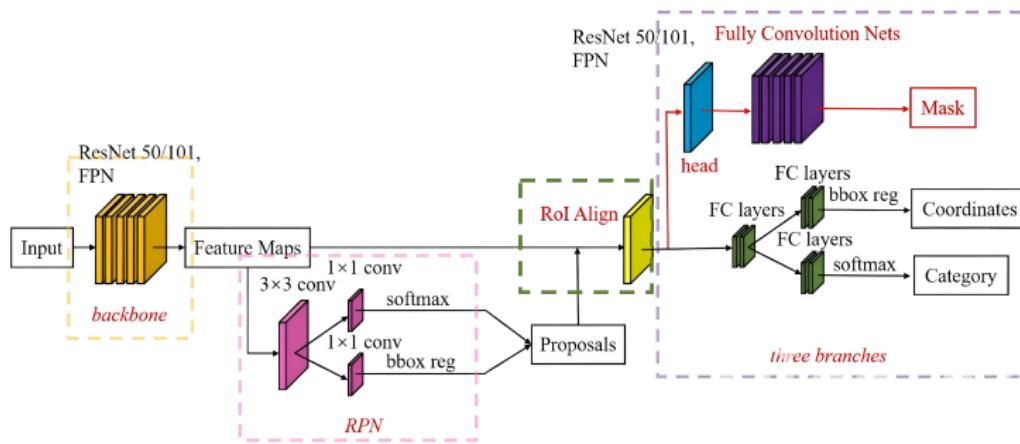


Figura 2: Arquitectura de Mask R-CNN [SKY23].

[SKY23] SKY ENGINE AI. What is Mask R-CNN? <https://skyengine.ai/se/skyengine-blog/119-what-is-mask-r-cnn>, 2023.

Mask R-CNN

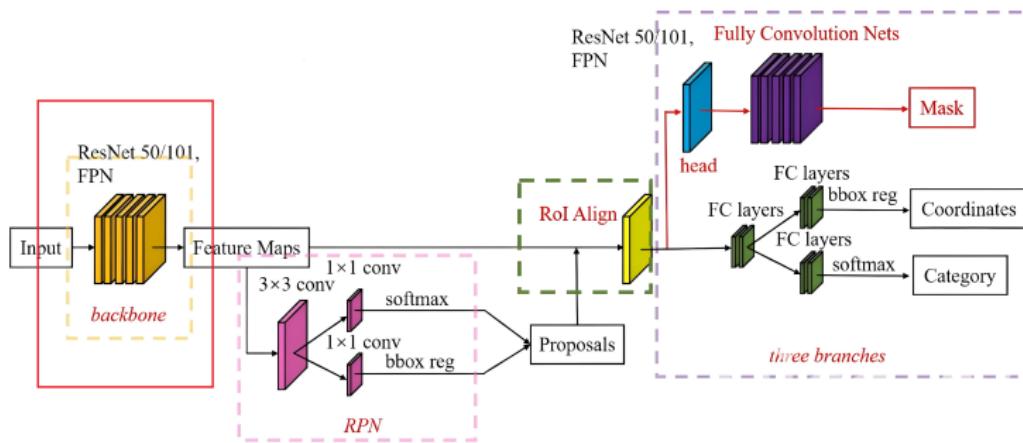


Figura 2: Arquitectura de Mask R-CNN [SKY23].

[SKY23] SKY ENGINE AI. What is Mask R-CNN? <https://skyengine.ai/se/skyengine-blog/119-what-is-mask-r-cnn>, 2023.

Mask R-CNN

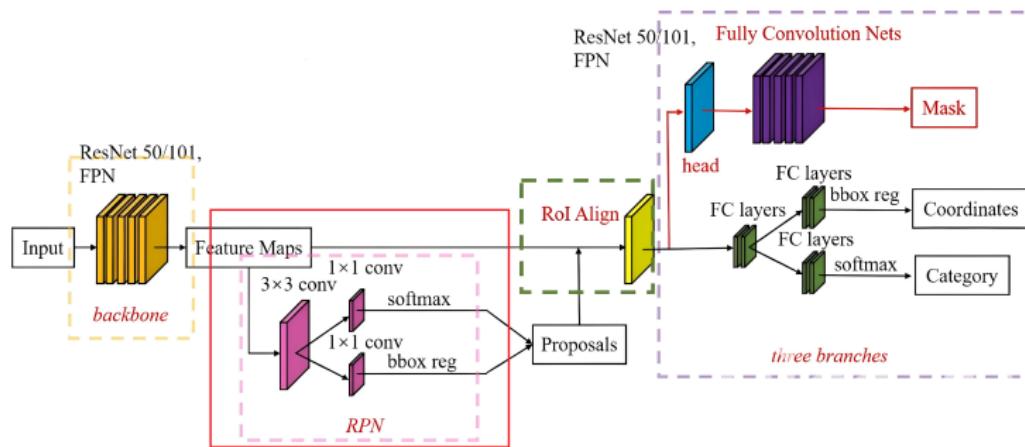


Figura 2: Arquitectura de Mask R-CNN [SKY23].

[SKY23] SKY ENGINE AI. What is Mask R-CNN? <https://skyengine.ai/se/skyengine-blog/119-what-is-mask-r-cnn>, 2023.

Mask R-CNN

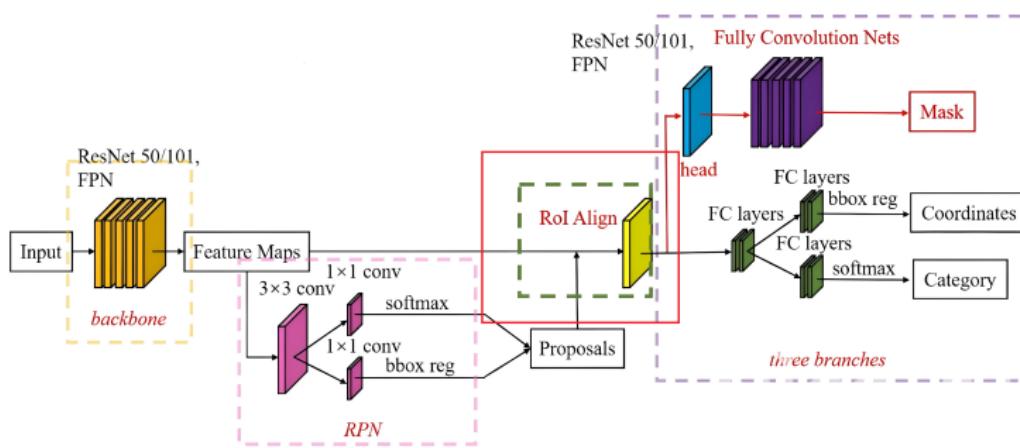


Figura 2: Arquitectura de Mask R-CNN [SKY23].

[SKY23] SKY ENGINE AI. What is Mask R-CNN? <https://skyengine.ai/se/skyengine-blog/119-what-is-mask-r-cnn>, 2023.

Mask R-CNN

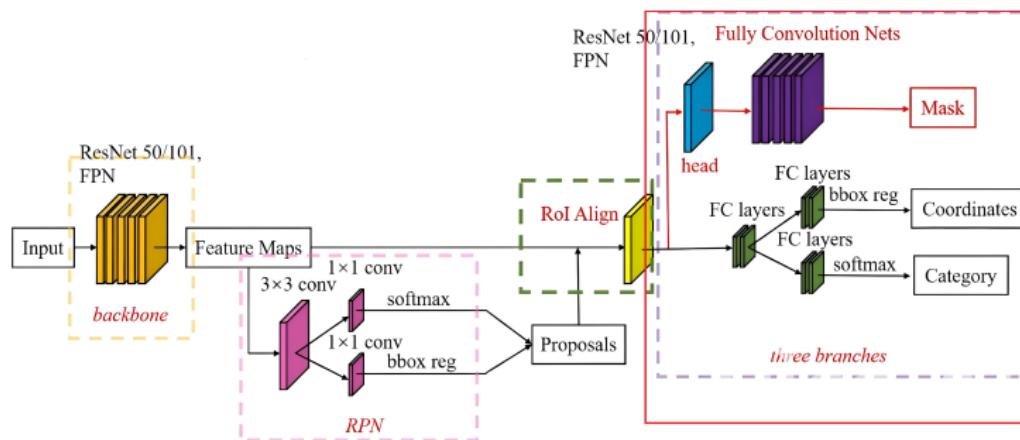


Figura 2: Arquitectura de Mask R-CNN [SKY23].

[SKY23] SKY ENGINE AI. What is Mask R-CNN? <https://skyengine.ai/se/skyengine-blog/119-what-is-mask-r-cnn>, 2023.

Grad-CAM

- Técnica de interpretabilidad **post-hoc** para redes neuronales convolucionales.
- Permite generar **mapas de calor visuales** que muestran las regiones más relevantes de una imagen para una clase específica.
- Se basa en los **gradientes** de la clase objetivo respecto a los mapas de características de una capa convolucional.
- **No requiere modificar ni reentrenar el modelo.**

Funcionamiento de Grad-CAM

- ① Elegir capa convolucional profunda:
Activaciones A^k .
- ② Calcular **gradientes de la clase objetivo y^c** respecto a A^k .
- ③ Realizar promedio global de gradientes: **Pesos α_k^c .**
- ④ $L_{\text{Grad-CAM}}^c = \text{ReLU} \left(\sum_k \alpha_k^c A^k \right)$.
- ⑤ **Interpolan y superponer** el mapa de calor sobre la imagen original.

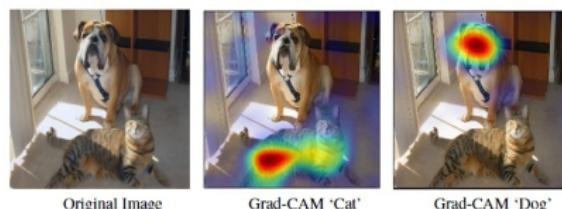


Figura 3: Mapas de calor de Grad-CAM para clase 'Cat' y 'Dog' [SCD+17].

[SCD+17] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, y D. Batra. Grad-CAM: Visual explanations from deep networks via gradient-based localization. *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, páginas 618–626, 2017.

Búsqueda Scopus

Segmentación dental

- 1992 publicaciones.
- Crecimiento desde 2015 (auge del Deep Learning).

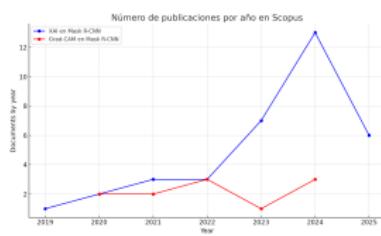
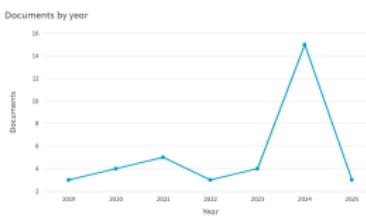
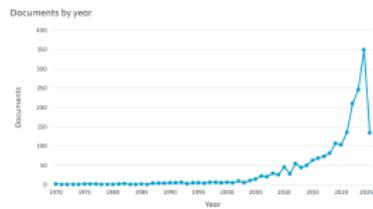
Mask R-CNN en seg. dental

- Solo 37 publicaciones.
- Línea de investigación reciente.

XAI aplicada a Mask R-CNN

- 35 publicaciones.
- Solo 11 aplican Grad-CAM.

Línea emergente con alto potencial



Adaptación de Grad-CAM a Mask R-CNN

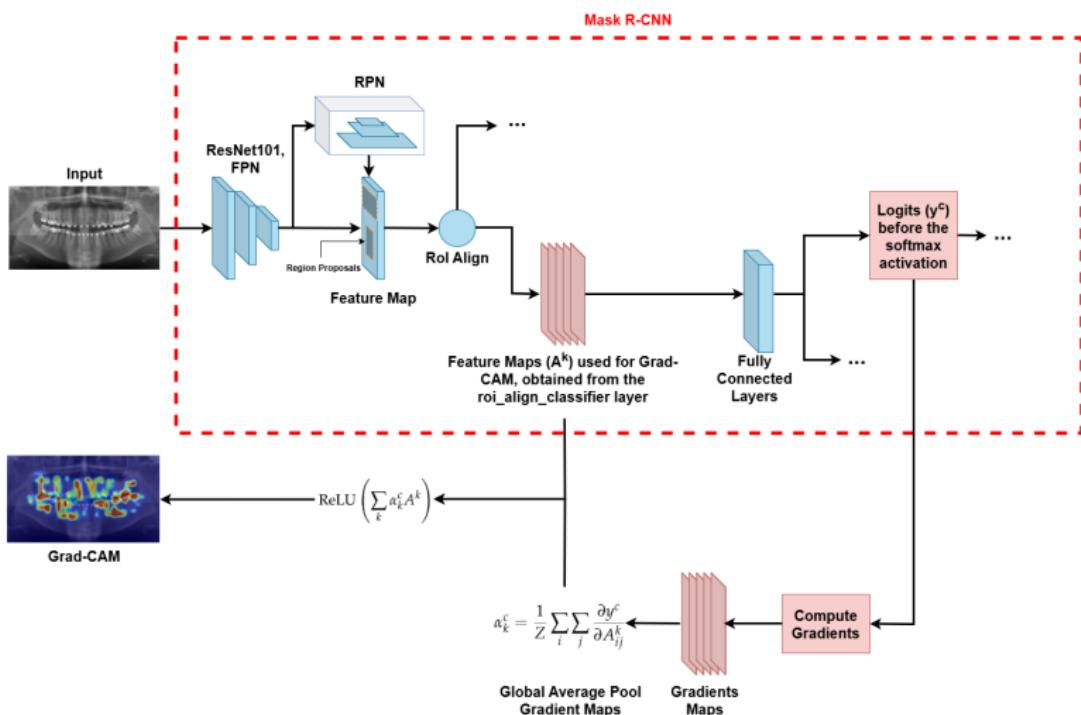
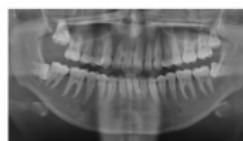
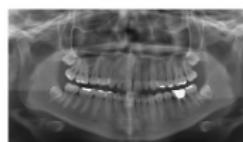
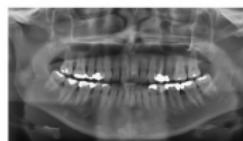


Figura 4: Proceso de adaptación de Grad-CAM a Mask RCNN.

Conjunto de datos DNS Panoramic Images

- **543 imágenes** de radiografías dentales panorámicas, junto con su respectiva máscara dental.
- **Diversidad:** dientes completos, restaurados, con aparatos o faltantes (**10 categorías**).
- **Partición:** 376 entrenamiento (70 %), 87 validación (15 %) y 80 test (15 %), distribuidas equilibradamente por categoría.



Métricas empleadas

Tipo	Métrica	Descripción
Detección	Precision	% de predicciones positivas correctas.
	Recall	% de objetos reales correctamente detectados.
	F1-Score	Media armónica entre Precision y Recall.
	AP@50	Área bajo curva P-R con IoU ≥ 0.5 .
Segmentación	DSC	Superposición entre máscara predicha y real.
	IoU	Solapamiento entre máscara predicha y real.
	Hausdorff Distance	Distancia máx. entre los bordes de las máscaras.
Explicabilidad	Deletion	Disminución de la confianza al eliminar píxeles.
	Insertion	Aumento de la confianza al insertar píxeles.
	Stability	Estabilidad ante perturbaciones.
	IoU _{Grad-CAM}	Solapamiento atención/segmentación.

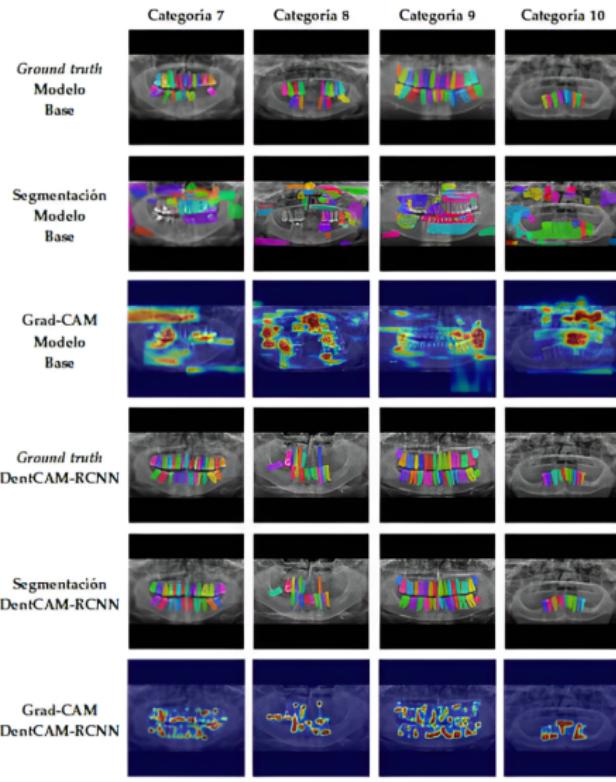
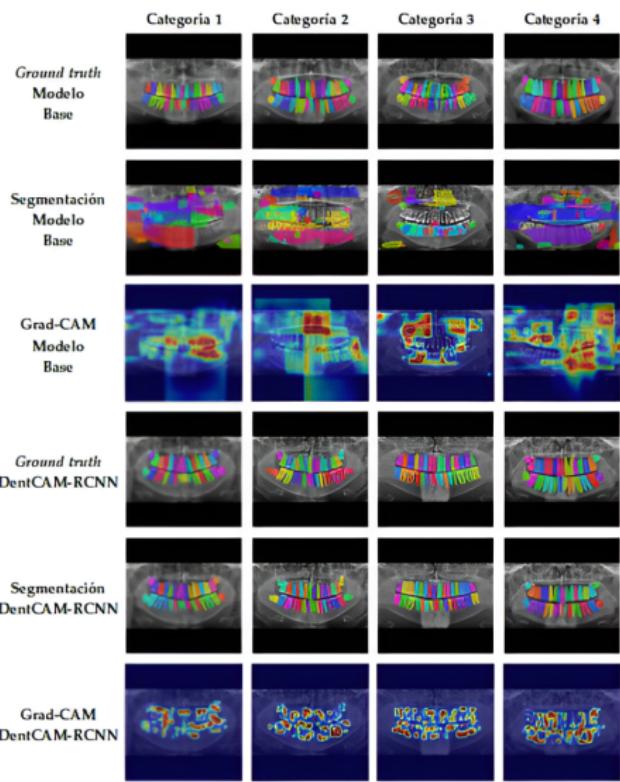
Análisis Cuantitativo

- **Modelos utilizados:**
 - **Modelo base:** Mask R-CNN preentrenado en COCO, sin adaptación al dominio dental.
 - **Modelo ajustado (DentCAM-RCNN):** Mask R-CNN con ajuste fino en sus capas superiores sobre DNS Panoramic Images.
- Los resultados muestran valores **promedio por categoría dental en el conjunto de test.**

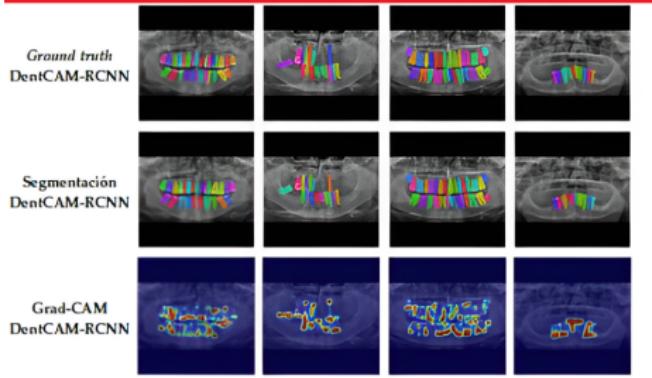
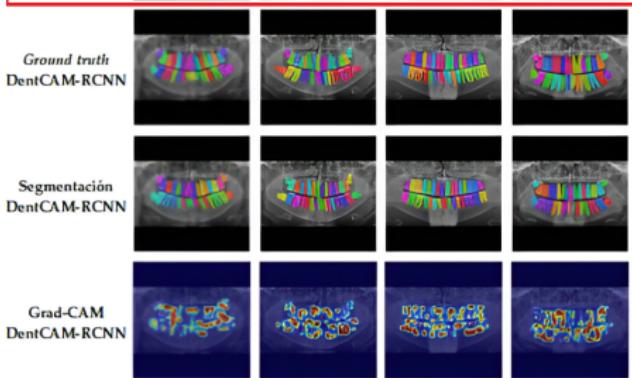
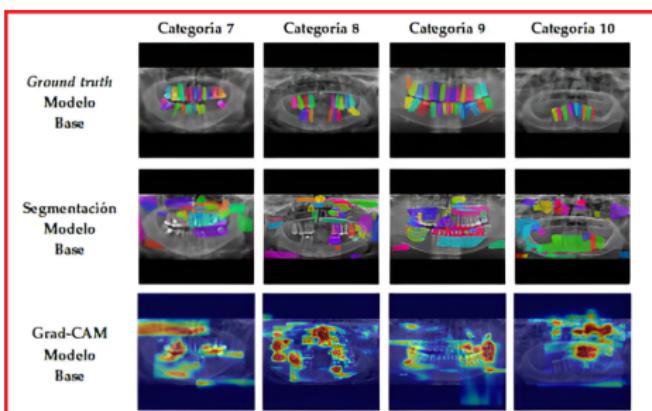
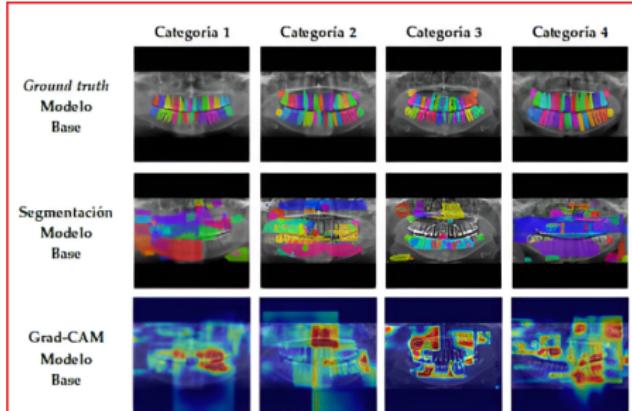
Modelo	Precision	Recall	F1	AP@50	Modelo	DSC	IoU	HD
Base	0.0523	0.0628	0.0571	0.0109	Base	0.0125	0.0097	30.73
Ajustado	0.9736	0.9480	0.9604	0.9398	Ajustado	0.8776	0.7923	12.26

Modelo	Deletion	Insertion	Stability	IoU _{Grad-CAM}
Base	0.9938	0.9936	0.2472	0.1347
Ajustado	0.1595	0.9965	0.8891	0.4365

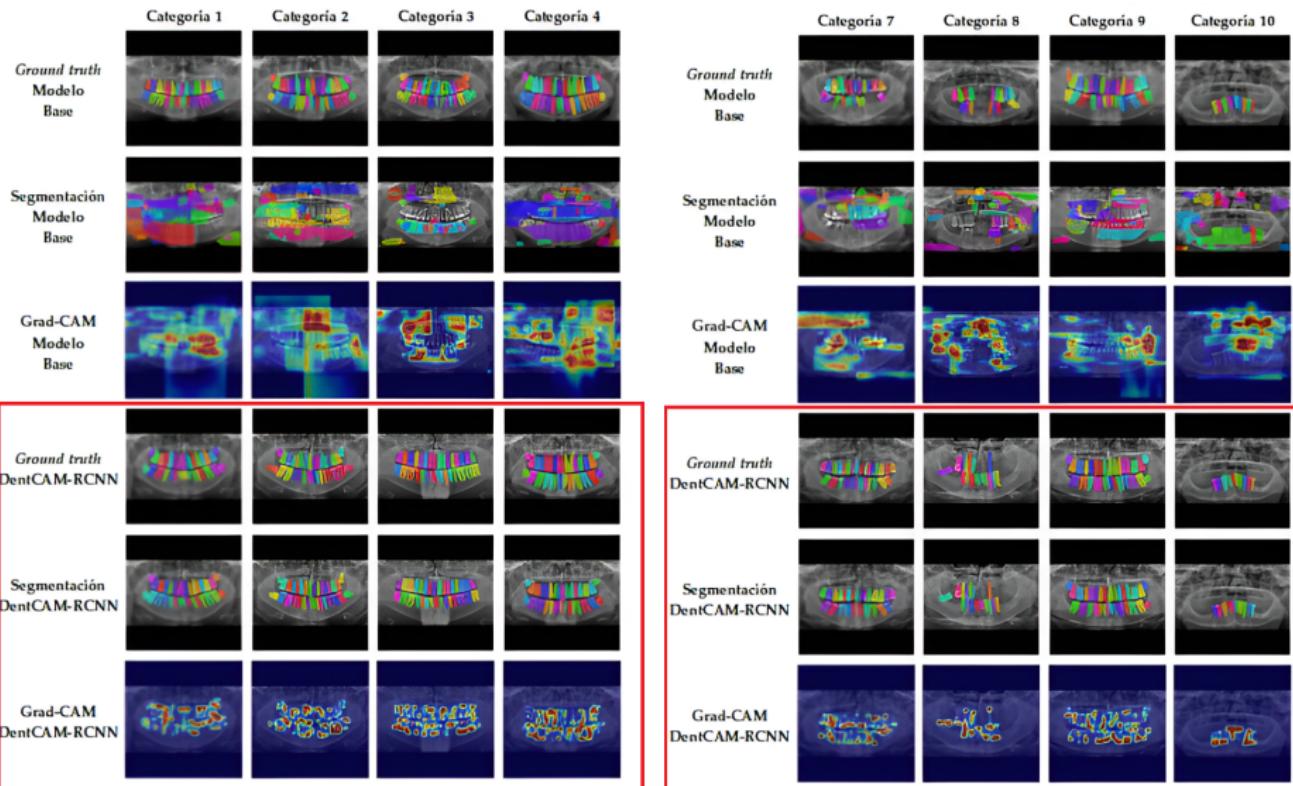
Análisis Cualitativo



Análisis Cualitativo



Análisis Cualitativo



Comparación con métodos clásicos de segmentación

- **Métodos clásicos de segmentación:** Rendimiento muy variable dependiente del enfoque utilizado. Resultados extraídos de [JFR+18].
- **DentCAM-RCNN:** Supera ampliamente a estos métodos.

Método	Precision	Recall	F1-Score
Region growing	0.3500	0.6300	0.4400
Splitting/merging	0.8100	0.0800	0.1400
Global thresholding	0.5200	0.6900	0.5600
Niblack method	0.5100	0.8200	0.6100
Fuzzy C-means	0.6100	0.4500	0.4900
Canny	0.4500	0.1100	0.1700
Sobel	0.6600	0.0300	0.0600
Active contours w/o edges	0.5100	0.5700	0.5200
Level set method	0.4800	0.6800	0.5200
Watershed	0.4800	0.8200	0.5800
DentCAM-RCNN	0.9776	0.9473	0.9622

[JFR+18] G. Jader et al. Deep instance segmentation of teeth in panoramic x-ray images. *SIBGRAPI*, 2018.

Comparación con estudios recientes de segmentación de instancias dentales en 2D

- **DentCAM-RCNN:** Alta eficacia con un volumen de datos intermedio.

Estudio	Imágenes	F1-Score	DSC	IoU
Jader et al. [JFR+18]	193	0.8800	–	–
Lee et al. [LHK+20]	1024	0.8750	–	0.8770
Brahmi y Jdey [BJ23]	107	0.6300	–	–
Rubiu et al. [RBC+23]	1000	–	0.8700	–
DentCAM-RCNN	543	0.9604	0.8776	0.7923

[JFR+18] G. Jader et al. Deep instance segmentation of teeth in panoramic x-ray images. *SIBGRAPI*, 2018.

[LHK+20] J.-H. Lee et al. Application of a fully deep convolutional neural network to the automation of tooth segmentation on panoramic radiographs. *Oral Surg. Oral Med. Oral Pathol. Oral Radiol.*, 2020.

[BJ23] W. Brahmi e I. Jdey. Automatic tooth instance segmentation and identification from panoramic x-ray images using deep cnn. *Multimedia Tools and Applications*, 2023.

[RBC+23] G. Rubiu et al. Teeth segmentation in panoramic dental x-ray using mask regional convolutional neural network. *Appl. Sci.*, 2023.

Conclusiones

- Se ha desarrollado **DentCAM-RCNN**: Modelo basado en Mask R-CNN ajustado al dominio dental, en el que se ha integrado Grad-CAM para dotarlo de explicabilidad visual.
- Este modelo **superá a métodos clásicos de segmentación, así como a estudios recientes.**
- **Todos los objetivos planteados al inicio han sido cumplidos con éxito.**
- **Código y resultados disponibles en GitHub:**
 - https://github.com/juanmaarg6/DentCAM_RCNN.

Fin de la Presentación

Muchas gracias por su atención.

¿Preguntas?

Análisis de Redes Convolucionales y Técnicas de Explicabilidad para Arquitecturas Neuronales Profundas para la Segmentación de Instancias

Trabajo de Fin de Grado

Autor: Juan Manuel Rodríguez Gómez

Tutores: Francisco Javier Merí de la Maza, Pablo Mesejo Santiago,
Marilyn Bello García

26 de Junio de 2025

Doble Grado en Ingeniería Informática y Matemáticas



UNIVERSIDAD
DE GRANADA