Evaluación de detectores convolucionales para el etiquetado de lesiones de caries en radiografías Bitewing

Ignacio Haeussler
Universidad de Chile
Santiago, Chile
ignacio.haeussler@ing.uchile.cl

ABSTRACT

En el campo de la salud dental el Deep Learning ha sido utilizado recientemente en la detección y clasificación de caries dentales [2, 32, 59], confirmando su potencial en la asistencia y disminución de la carga laboral en el diagnóstico odontológico. Sin embargo, los resultados obtenidos no han empleado métodos de detección de objetos o de segmentación de instancias diseñados para la obtención de información posicional de los objetos identificados, la que es necesaria en el diagnóstico profesional estándar. El presente trabajo de tesis se propone evaluar los principales métodos de detección de objetos, con el objetivo de demostrar su efectividad en esta tarea y desarrollar un sistema en el estado del arte, optimizado para la detección de lesiones de caries en radiografías Bitewing. Un total de 3000 radiografías previamente etiquetadas por dentistas radiólogos (Figura 1) serán utilizadas para entrenar los métodos Faster R-CNN [49], YOLOv3 [48], RetinaNet [38], Light-Head R-CNN [35], TridentNet [34] y HTC [6] y compararlos con el actual estado del arte en esta tarea, el método de segmentación semántica U-Net [51]. Serán presentadas, tanto para los métodos a evaluar como para el actual estado del arte, las tradicionales métricas de detección de objetos AP, AP50 y AP75 sobre un subconjunto de testeo de las imágenes mencionadas. Se presentan hipervínculos a las implementaciones en código abierto para cada uno de los métodos a ser evaluados.

CCS CONCEPTS

• **Computing methodologies** → **Artificial intelligence**; • **Computing methodologies** → **Machine learning**; • **Information systems** → *Information systems applications*;

KEYWORDS

Deep Learning, caries dentales, información posicional, diagnóstico profesional estándar, detección de objetos, radiografías Bitewing

ACM Reference format:
Ignacio Haeussler. 2019. Evaluación de detectores convolucionales para el etiquetado de lesiones de caries en radiografías Bitewing. In *Proceedings of Junio 2019, Santiago de Chile, Universidad de Chile,* 8 pages.
https://doi.org/10.1145/nnnnnnn.nnnnnnn

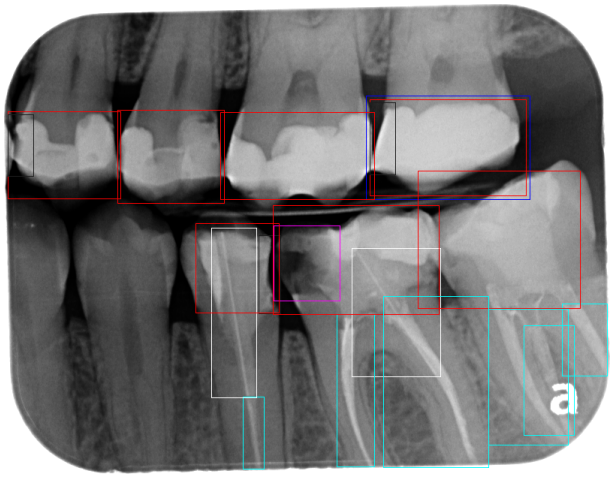

Figura 1: Radiografía Bitewing. Muestra desde las 3000 imágenes etiquetadas por dentistas radiólogos.

1. INTRODUCCIÓN

La visión computacional mediante Deep Learning ha demostrado una gran exactitud y eficiencia en la detección y clasificación de retinopatía diabética, cáncer de piel y tuberculosis pulmonar [14, 20, 29], superando a los métodos previos basados en características manualmente diseñadas. Más recientemente ha sido aplicada con éxito en la detección y clasificación de caries en radiografías dentales [2, 32], logrando superar la precisión odontológica promedio [59].

Estos resultados permitirían la asistencia y automatización parcial del diagnóstico odontológico, reduciendo tanto la carga de trabajo de los profesionales como la frecuencia de diagnósticos erróneos, la que presenta actualmente valores cercanos al 40 % [13, 59]. En este contexto es que la detección automática de estructuras y patologías dentales es relevante para la ciencia del cuidado de la cavidad oral.

Los métodos de visión computacional principalmente aplicados en ciencias de la salud tienen como objetivo la clasificación, detección de objetos y segmentación. Sin embargo, los métodos de

Junio 2019, Universidad de Chile, Santiago de Chile

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM.
https://doi.org/10.1145/nnnnnnn.nnnnnnn

clasificación no permiten una natural obtención de información posicional de las estructuras y patologías relevantes para el diagnóstico odontológico, por lo que métodos de detección de objetos o de segmentación serían requeridos para posibilitar una asistencia satisfactoria de un sistema de visión computacional. Hasta ahora esto solo ha sido realizado por Srivastava et al. [59], quienes a pesar de obtener resultados que confirman la efectividad de la visión computacional para la detección de caries dentales en radiografías Bitewing (Figura 1), emplean métodos basados únicamente en U-Net[51], los que han sido aventajados en forma consistente por variados trabajos posteriores. En segundo lugar, los distintos grados de avance de las lesiones de caries poseen distintos tamaños promedio, por lo que la precisión de detección podría verse beneficiada por un análisis multiescala, ausente en U-Net pero imprescindible en el estado del arte de la detección de objetos. En tercer lugar, el sistema desarrollado solo permite detectar la presencia de caries, no su tipo ni estado de avance, lo que también es esencial para el diagnóstico odontológico profesional estándar. Por último, los resultados obtenidos son escasos, e impiden estimar adecuadamente la dificultad y los desafíos enfrentados a la hora de detectar lesiones de caries en radiografías dentales.

Con respecto al primer punto, el método utilizado por el estado del arte para esta tarea está basado en FCN [42], el cual está diseñado para segmentación semántica. Por ello no es directamente aplicable a los conjuntos de datos estándar de detección de objetos. Sin embargo, su mAP de segmentación en Pascal VOC 2012 [15] es de 67.2 %, mientras que el estado del arte es de 89.0 % [8]. Un desarrollo similar ha ocurrido en la detección de objetos en el conjunto de imágenes COCO [39] desde la aparición de FCN en 2015, donde el mismo año Faster R-CNN obtuvo una mAP de detección de 34.7 %[28] y en la actualidad HTC alcanza una de 50.7 %. [6]. Estos resultados sugieren que la precisión alcanzada por una adaptación de FCN para la detección de objetos habría sido superada en forma similar o mayor a la experimentada por Faster R-CNN, dada la esencial optimización del último para tal tarea.

Con respecto al segundo punto, la variación en el tamaño promedio de los distintos grados de avance de las lesiones de caries sugiere que los avances de los métodos de detección de objetos en la detección multiescala permitirían generar mejores etiquetados. Inversamente, la detección de lesiones de caries podría constituir un medio por el que evaluar y comparar tanto la efectividad de los avances en detección multiescala (FPN vs convoluciones dilatadas [34]) como de otras propiedades de gran importancia para los detectores de objetos, como la efectividad de una pérdida focal [38] versus la aplicación de RPNs en los métodos de dos etapas, la efectividad del aprendizaje multitarea [6] para aumentar la precisión de la detección, y el reemplazo de la capa de agrupación de regiones de interés [49] por la obtención mapas regionales de puntajes [35].

Por lo tanto, el trabajo de tesis presentado se propone evaluar los principales métodos de detección de objetos convolucionales para el etiquetado de lesiones de caries en radiografías dentales, mediante un conjunto de 3000 radiografías Bitewing obtenidas desde clínicas dentales y centros radiológicos , anonimizadas para proteger los derechos de los pacientes y etiquetadas sistemáticamente por dentistas radiólogos. Además, 9000 radiografías Bitewing no etiquetadas podrían ser utilizadas para mejorar el aprendizaje mediante métdodos de aumentación de datos y de aprendizaje no supervizado.

Adicionalmente el trabajo presentado se propone el diseño de un sistema eficaz y en el estado del arte que detecte tanto la posición como el grado de avance en lesiones de caries en radiografías Bitewing. Se presentan en los apéndices resultados preeliminares obtenidos mediante Faster R-CNN [49]

2. BACKGROUND

Deep Learning en ciencias de la salud. En últimos 5 años, el Deep Learning basado en redes neuronales profundas, particularmente en redes neuronales convolucionales (CNNs), se ha incorporado vigorosamente en el análisis de imágenes médicas en las tareas de detección, clasificación y segmentación [40]. Algunos ejemplos de este tipo de tarea son el diagnóstico diferenciado entre las enfermedades de Alzheimer y de Huntington en datos de resonancia magnética [52, 61], clasificación [45] y segmentación multiescala de tumores [68], detección multiescala de lesiones cerebrales [17] y segmentación del estriado [9].

En el campo de la salud dental las CNNs han sido utilizado eficazmente en la detección de caries [2, 32, 59], dientes [5, 43, 65], la presencia de enfermedad periodontal [33] y en métodos para mejorar la resolución de imágenes dentales [21].

En la detección de lesiones de caries mediante Deep Learning, distintos tipos de imágenes dentales han sido utilizadas: Retroalveolar, Bitewing, Panorámica y Periapical. Lee et al. [32] realizan la detección en imágenes Periapicales, Srivastaba et al. [59] en Bitewing y Ali et al. [2] no especifica el tipo de imagen utilizada. Por otro lado, distintos métodos de detección han sido empleados. Tanto Ali et al. [2] como Lee et al. [32] recortan las imágenes de modo que se presente un único diente por imagen, las redimensionan a 64x64 y 299x299 píxeles respectivamente, y detectan solo la presencia o ausencia de caries (0 o 1), mientras que Srivastaba et al. [59] utiliza imágenes Bitewing completas, las redimensiona a 299x299 píxeles y realiza detección objetos sin clasificación, es decir, predicción de cuadros delimitadores sin especificar tipo para cada lesión de caries identificada, mediante métodos basados en redes totalmente convolucionales [42, 51]. No se han publicado otros trabajos que empleen métodos de detección de objetos para el etiquetado de lesiones de caries en radiografías dentales.

En la ciencia de la salud dental se reconocen distintas clasificaciones para los distintos grados de avance de las lesiones de caries relevantes para el diagnóstico odontológico, lo que no ha sido considerado en los trabajos previos. Por otro lado, solo Srivastaba et al. [59] utiliza métodos que permiten la obtención de información posicional de los elementos identificados, la que es requerida para un diagnóstico de utilidad profesional de estas patología. El sistema desarrollado emplea U-net [51], la que está basada en redes neuronales totalmente convolucionales (FCN) [42]. Sin embargo, las FCN están diseñadas para realizar segmentación semántica, y a pesar de que es posible utilizarla para realizar detección de objetos no está naturalmente diseñada para ello, y se ha visto superada consistentemente por diversos avances en el área [7, 8, 64, 67]. La U-net empleada por Srivasta et al. tampoco identifica el grado de avance de las lesiones detectadas.

Detectores de objetos profundos. Los métodos de detección de objetos basados en Deep Learning han mostrado recientemente avances dramáticos tanto en precisión como en eficiencia. Como

uno de los enfoques predominantes, los métodos de dos etapas [4, 11, 18, 19, 35, 49] proponen inicialmente un conjunto de regiones de interés imprecisas, las que son posteriormente refinadas mediante CNNs. En [19], R-CNN propone regiones mediante Selective Search [62] y luego clasifica y refina las regiones recortadas desde la imagen original mediante una CNN estándar de forma secuencial e independiente. Para reducir la redundancia computacional de la extracción de características en R-CNN, SPPNet [23] y Fast R-CNN [18] extraen una vez las características de cada imagen, y luego generan características regionales mediante capas de agrupación piramidal y de regiones de interés, respectivamente. La capa de agrupación de regiones de interés es posteriormente mejorada mediante una capa de alineamiento de regiones de interés [22] para solucionar el problema de la cuantización espacial gruesa.

Faster R-CNN [49] es el primero en proponer un marco de trabajo unificado para la detección de objetos, introduciendo una red de proposición de regiones de interés (RPN) que comparte la misma columna de red con la cabeza de detección (Figura 2), para reemplazar los aislados y poco eficientes métodos anteriores. Los siguientes trabajos aumentan la precisión y/o la eficiencia de Faster R-CNN. Con respecto a la precisión, Dai et al. [12] incorporan redes convolucionales deformables que aprenden desplazamientos sin supervisión para modelar transformaciones geométricas. Lin et al. [37] agregan redes de pirámides de características (FPN), las que explotan las jerarquías multi-escala inherentes a las CNNs para construir pirámides de características (Figura 3). Sobre FPN, Mask R-CNN [22] agrega una rama para predecir máscaras de segmentación, demostrando que un análisis multitarea mejora el aprendizaje tanto de la segmentación como de la detección de objetos. Más adelante, TridentNet [34] superaría tanto a FPN como a los métodos entrenados mediante pirámides de imágenes como SNIP [56] y SNIPER [57], reemplazando la pirámide de características por una predicción simultánea de múltiples escalas basada en la adición de ramas paralelas a la cuarta etapa de ResNet [24], con parámetros compartidos pero con distintos valores de dilatación en sus convoluciones (Figura 4).

Para refinar los cuadros delimitadores presentados por RPN, Cascade R-CNN [4] agrega múltiples etapas, donde la salida de cada etapa es entregada a la siguiente para un refinamiento de mayor calidad, aumentando simultáneamente los umbrales de IoU aceptados. Posteriormente, HTC [5] se encargaría de mejorar estas etapas, incluyendo tareas de segmentación semántica y de instancia simultáneamente a la detección de objetos (Figura 5), aumentando significativamente la precisión y logrando resolver los problemas asociados imágenes con un alto desorden de fondo (background clutters) en métodos anteriores.

Por otro lado, para aumentar la eficiencia de Faster R-CNN, R-FCN [11] construye mapas de puntuación sensibles a la posición a través de la red totalmente convolucional, logrando evitar las costosas cabezas de red dependientes de capas de agrupación de regiones de interés (Figura 6). Para evitar grandes mapas de puntuación en R-FCN, Light-Head R-CNN [35] es diseñada usando mapas de características delgados y una subred R-CNN pequeña para construir un detector de dos etapas aun más eficiente, permitiendo a los métodos de dos etapas lograr simultáneamente una precisión y eficiencia significativamente superiores a los métodos de una etapa.

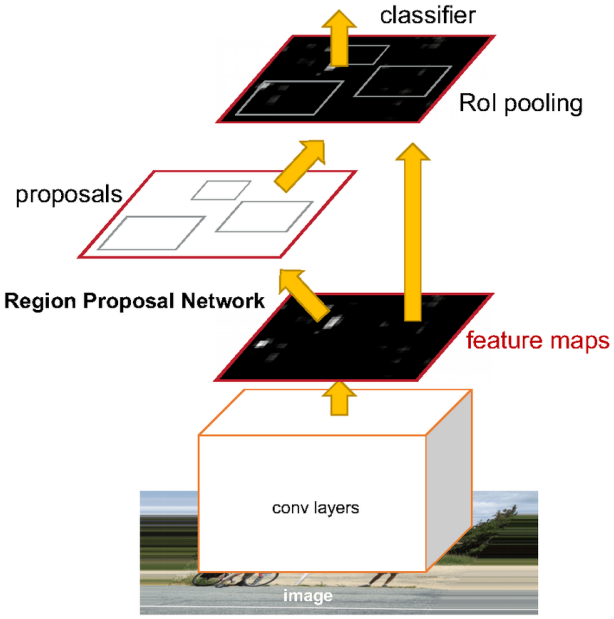

Figura 2: Faster R-CNN. Una columna convolucional extractora de características es compartida por la red generadora de regiones de interés la red clasificadora. La red clasificadora recibe tanto las características extraídas como las regiones propuestas. Ref: https://arxiv.org/pdf/1506.01497.pdf

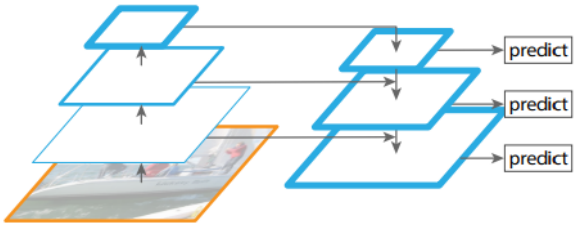

Figura 3: FPN. Las salidas de distintas capas son procesadas y combinadas para generar una pirámide de características desde la que se predicen objetos de distintas escalas, aprovechando así el natural decrecimiento de la resolución y aumento de la densidad semántica en las redes usadas en visión computacional. Ref: https://arxiv.org/pdf/1612.03144.pdf

Por otro lado, iniciados por Overfeat [53] y popularizados por YOLO [46–48] (Figura 7) y SSD [41] (Figura 8), buscan ser más eficientes por medio de clasificar directamente cuadros de anclaje predefinidos y refinarlos posteriormente mediante CNNs, sin el paso de proposición de regiones. SSD al igual que la última versión de YOLO [48] (Figura 9) busca mejorar su predicción multi-escala similarmente a como lo hace FPN, utilizando la información posicional de mapas de características de capas de mayor resolución. Basado en el módulo de predicción multicapa de SSD, DSSD [16] aumenta la precisión introduciendo información contextual mediante operadores deconvolucionales. RetinaNet [38] modifica a SSD para

Figura 4: TridentNet. La pirámide de características es reemplazada por tres ramas que comparten parámetros pero que utilizan distintos grados de dilatación en sus convoluciones, facilitando a la red el aprendizaje de distintas escalas. Las tres ramas son insertadas en la etapa más avanzada y con menor resolución de la red, maximizando el efecto de la dilatación. Ref: https://arxiv.org/pdf/1901.01892.pdf

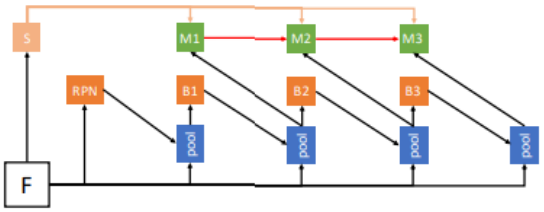

Figura 5: HTC. Hybrid Task Cascade agrega varias etapas a partir a la única previamente utilizada en RPN de Faster R-CNN. A partir de las características (features) extraídas de la etapa 4 de ResNet [24], HTC realiza segmentación semántica para detectar el fondo y lo utiliza en las distintas etapas de segmentación de instancia para refinar la identificación de los objetos, lo que simultáneamente permite refinar los cuadros delimitadores de las etapas de detección de objetos. Ref: https://arxiv.org/pdf/1901.07518v2.pdf

incorporar las pirámides de características de los métodos de dos etapas basados en FPN, y propone una nueva pérdida focal para resolver el problema de la extrema desproporción entre la clase de fondo y las de objetos, el que destaca como dificultad central en los detectores de una etapa. También heredando méritos de los enfoques de dos etapas, RefineDet [66] propone un módulo para refinar los cuadros de anclaje, filtrando los negativos y ajustando gruesamente los restantes previamente al módulo de detección.

3. PROBLEMA A RESOLVER

El presente trabajo propone la evaluación y comparación de los principales métodos de detección de objetos para la detección y clasificación de lesiones de caries en radiografías Bitewing. Actualmente se cuenta con 3000 imágenes Bitewing etiquetadas por dentistas radiólogos para etiquetas distribuidas entre 48 clases distintas. Entre ellas se encuentran 5 clases de caries para las cuales se tiene la distribución expuesta en la Figura 10. Puede apreciarse un desproporción de clases importante, sin embargo esto es un factor común en la aplicación de deep learning en ciencias de la salud [3, 40, 45]. El objetivo principal es evaluar los principales métodos de detección de objetos para identificar las caries presentes en un

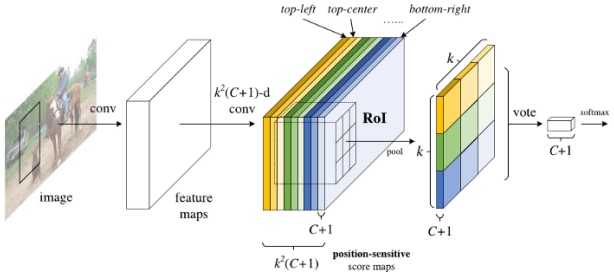

Figura 6: R-FCN. Reemplaza la costosa capa de agrupación de regiones de interés por una capa que calcula mapas de puntaje para cada clase (más clase de fondo) y zona en una grilla KxK que dividiría uniformemente cada región propuesta. Es decir, cada mapa estaría diseñado para detectar la presencia de una de las K^2 zonas para una clase específica (por ejemplo, la parte superior izquierda de una puerta, o la parte inferior derecha de un arco de fútbol). Luego, estos mapas son consultados en base a las propuestas de región entregadas por la RPN, dividiendo cada región en KxK zonas y obteniendo su puntaje para cada zona en las ubicaciones correspondientes de los mapas de puntaje. Con ello se obtienen un valor para cada zona de la región de interés para cada clase. Posteriormente, se promedian todas las zonas para cada clase y se obtiene un puntaje para cada clase. Ref: https://arxiv.org/pdf/1605.06409.pdf

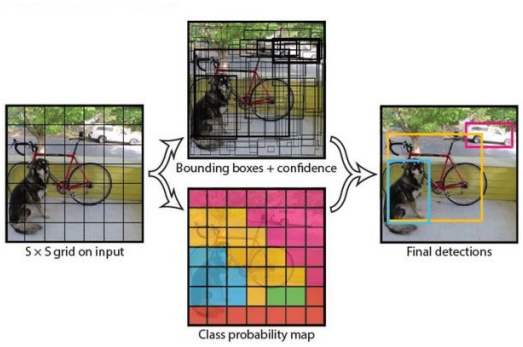

Figura 7: YOLO. Divide la imagen en SxS celdas, las que solo pueden predecir una única clase. Además, cada celda predice B cuadros delimitadores, cada uno con sus respectivas probabilidades de contener a un objeto. Utilizando la clase de la celda correspondiente, YOLO predice un objeto para el cuadro delimitador con mayor probabilidad, generalmente solo si ésta es mayor a 0.6 [55]. Ref: https://arxiv.org/pdf/1506.02640.pdf.

subset de testeo de las imágenes etiquetadas, habiéndose previamente entrenado cada uno de ellos en un subset de entrenamiento con intersección nula con el anterior. En particular, se pondrán a prueba al menos los siguientes métodos de detección de objetos:

- **Faster R-CNN [49]:** Debido a ser el primer método que propuso un marco de trabajo unificado para la detección

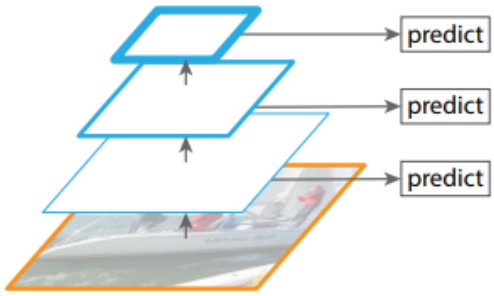

Figura 8: SSD. A diferencia de FPN, SSD utiliza directamente mapas de características de distinta resolución para una predicción multi-escala. Ref: https://arxiv.org/pdf/1612.03144.pdf.

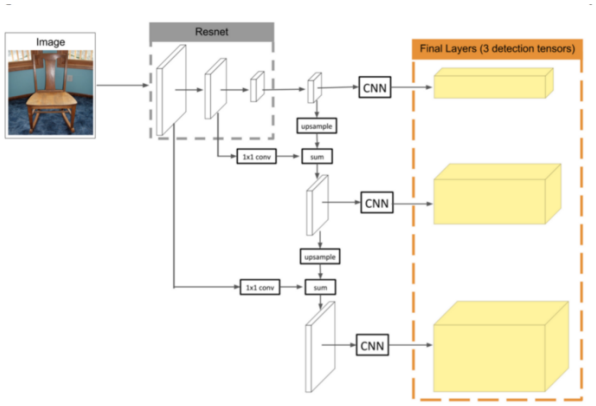

Figura 9: YOLOv3. La tercera versión de YOLO también utiliza el mecanismo de FPN para realizar una predicción multi-escala. Ref: [25]

de objetos y ser en el cual se han basado la mayoría de los trabajos posteriores, se considerará en este trabajo como uno de relevancia para la completitud de este trabajo. Código: https://github.com/facebookresearch/Detectron.

- **Light-Head R-CNN [35]:** Con 30.7 AP a 102 FPS en MSCOCO [39] y superando a todos los métodos de una etapa, Light-Head R-CNN se presenta como el más eficiente de los métodos de dos etapas y el poseedor del mejor trade-off entre eficiencia y precisión para aplicaciones en tiempo real (Figura 11). Código: https://github.com/zengarden/light_head_rcnn.
- **TridentNet [34]:** En la principal tarea de detección multi-escala, TridentNet se muestra como el exponente más importante, habiendo demostrado su superioridad frente a las pirámides de características de los influyentes FPN, Mask R-CNN, RetinaNet y las pirámides de imágenes de SNIP y SNIPER. Código: https://github.com/TuSimple/simpledet/tree/master/models/tridentnet.
- **YOLOv3 [48]:** Como el más conocido e influyente de los métodos de una etapa se consideró importante su evaluación. Su radical inclinación por la eficiencia frente a la precisión permitiría explorar más efectivamente la dificultad de la detección de lesiones de caries en radiografías Bitewing. Código: https://github.com/YunYang1994/tensorflow-yolov3.
- **RetinaNet [38]:** Se consideró que debido a ser el método de una etapa con mejor trade-off entre eficiencia y precisión y el mejor antes de la aparición de Light-Head R-CNN, RetinaNet debía ser incorporado para obtener una evaluación completa de los principales detectores convolucionales de objetos, tanto con respecto a los de una como a los de dos etapas. Código: https://github.com/facebookresearch/Detectron.
- **HTC [5]:** En base a sus etapas de refinamiento de cuadros delimitadores de Cascade R-CNN, Hybrid Task Cascade (HTC) presenta un desarrollo ortogonal y complementario tanto a la eficiencia de los mapas de puntaje de Light-Head R-CNN como a la eficiencia de la predicción multi-escala de TridentNet. Debido a ser el actual estado del arte [10], se considera principal su evaluación. Código: https://github.com/open-mmlab/mmdetection.

Adicionalmente se evaluará a U-Net [51] (Código: https://github.com/zhixuhao/unet) como línea de base al ser el único método hasta ahora utilizado en la detección y obtención de información posicional de lesiones de caries en radiografías dentales.

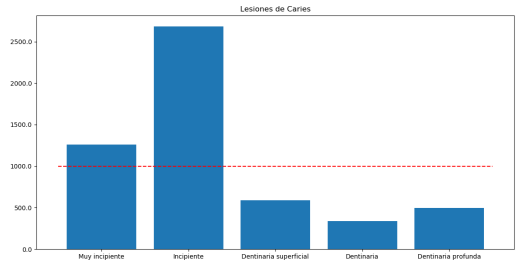

Figura 10: Clases de lesiones de caries etiquetadas: 1258 muy incipientes, 2682 incipientes, 590 dentinarias superficiales, 337 dentinarias, 495 dentinarias profundas

4. PREGUNTAS DE INVESTIGACIÓN

Las preguntas de investigación están enfocadas en obtener mediciones de la dificulta del problema, comparar los distintos enfoques y adquirir un conocimiento más profundo de sus distintas propiedades.

- ¿En qué rangos estará la precisión (en AP [39]) alcanzada? ¿Más cercana al estado del arte de MSCOCO o de Pascal VOC [15]?
- ¿El ranking de los métodos seguirá el mismo orden que el de MSCOCO o variara debido al cambio de dominio?
- ¿Podría cada uno de los métodos superar a la U-Net empleada inicialmente para la detección de caries en imágenes Bitewing?
- ¿Podrían beneficiarse los métodos de la utilización de avances incorporados por los demás, como reemplazar el uso de FPN en HTC por las tres ramas multi-escala de TridentNet, o Light-Head R-CNN de las etapas de refinación de HTC?

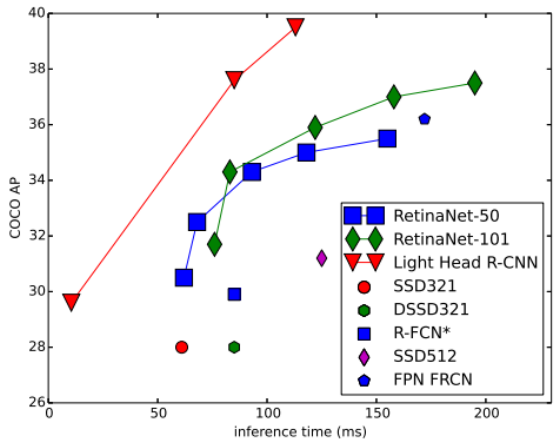

Figura 11: Trade-off entre eficiencia y precisión. Light-Head R-CNN se presenta como el método de dos etapas más eficiente, superando a todos los métodos de una etapa: RetinaNet, SSD y YOLOv3 (ausente en figura). Ref: https://arxiv.org/pdf/1711.07264.pdf

5. HIPÓTESIS

Los métodos actuales de detección de objetos son capaces de superar la precisión alcanzada por el estado del arte [51, 59] en la detección de lesiones de caries en radiografías Bitewing.

6. OBJETIVOS

6.1. Objetivo general

Evaluar los principales detectores de objetos convolucionales en el etiquetado automático de lesiones de caries en radiografías Bitewing y compararlos con el actual estado del arte.

6.2. Objetivos específicos

- Obtención de métricas de precisión en detección de objetos estándar para los detectores evaluados.
- Creación de un Benchmark para métodos de detección de objetos, a partir de un conjunto etiquetado de radiografías Bitewing.
- Desarrollo de un sistema para la detección de lesiones de caries en el estado del arte.
- Identificación de las principales dificultades que tienen los métodos evaluados en la detección de lesiones de caries en las imágenes utilizadas.
- Identificación parámetros de Data Augmentation [44] que permitan mejorar el aprendizaje de detectores de objetos en el conjunto de imágenes empleado.

7. METODOLOGÍA

Se entrenarán los métodos de detección de objetos mediante aprendizaje supervisado utilizando 3000 imágenes Bitewing anonimizadas y previamente etiquetadas por 3 grupos de dentistas radiólogos. Los ejemplos de entrenamiento efectivos consistirán en sub-imágenes correspondientes a cuadros delimitadores de los dientes presentes en las imágenes. Los cuadros delimitadores de dientes seleccionados dependerán de si poseen lesiones de un tipo, de varios tipos o no poseen restauraciones.

A continuación se enumeran y detallan los sucesivos pasos que se llevarán a cabo para el preprocesamiento de los datos, el entrenamiento de los modelos y la obtención de los resultados esperados:

1. En primer lugar, las imágenes deben ser preparados en una carpeta la que será accedida por los métodos, posiblemente transformándolas previamente a un formato binario para mayor eficiencia (como TFRecod de Tensorflow [1]).
2. Luego, el módulo encargado de preparar los batches de entrenamiento en tiempo real (basado en la API tf.data de Tensorflow) se configurará para recolectar las imágenes, revolverlas (de modo que su orden no siga un patrón deducible posteriormente por las redes neuronales), repetirlas NE veces (en forma 'lazy' sin realizar copias, con el objetivo de que el entrenamiento se realice una cantidad NE de épocas), parsear su contenido (obtener imagen y cuadros delimitadores asociados), aplicar Data Augmentation [44, 54, 63] (la que en principio consistirá en la comúnmente aplicada a CIFAR10 [24, 26, 27, 30, 31, 36, 50, 58, 60]), normalización (restar el promedio y dividir por desviación estándar de todas las imágenes) y almacenar el batch generado hasta su utilización. Es decir, el entrenamiento en TPU/GPU de un batch de imágenes es realizado en paralelo a la preparación en CPU del siguiente batch (esto puede ser realizado mediante la instrucción de tensorflow tf.data.Dataset.prefetch(n_batches), para mantener n_batches preprocesados, lo que puede ser útil en caso de que para algunos batches el preprocesamiento tarde más que su entrenamiento).
3. El entrenamiento será realizado con una cantidad de épocas, tasas de aprendizajes, tamaños de batch, optimizadores y arquitecturas de red variables y se informarán para cada uno de los resultados obtenidos.
4. Se obtendrán las métricas de precisión para cada uno de los métodos mediante su evaluación en el conjunto de testeo.
5. Se iterará sobre los pasos previos la cantidad de veces que se estimen convenientes, buscando descartar la hipótesis nula y obtener los resultados propuestos en la siguiente sección.

8. RESULTADOS ESPERADOS

Se presentarán al menos las AP, AP50 y AP75 para cada método, para cada tipo de carie, para las mejores configuraciones de los métodos empleados y para las configuraciones de parámetros de Data Augmentation que se estimen relevantes. Además, en el caso de probar distintas redes neuronales, se informarán las AP para las distintas redes empleadas. Posteriormente se realizará un análisis y comparación de los métodos empleados, con énfasis en dar respuestas a las preguntas de investigación planteadas.

Las principales dificultades previstas estarían asociados a la implementación de los distintos flujos de entrenamiento y al desarrollo los métodos donde la implementación provista por los autores entregue complicaciones (cada uno de los métodos a evaluar posee una implementación pública). Además, los métodos podrían requerir variaciones en sus hiperparámetros.

9. APORTES DE LA TESIS

Los principales aportes del trabajo una vez finalizado son:

- Evaluación y comparación de los principales detectores convolucionales en el etiquetado de lesiones de caries en radiografías Bitewing.
- Creación de un Benchmark para métodos de detección de objetos, a partir de un conjunto etiquetado de radiografías Bitewing.
- Un informe de las principales dificultades identificadas a la hora de detectar lesiones de caries en este tipo de imágenes.
- Un prototipo funcional que se pondrá en producción en el centro radiológico Cimex: http://cimexradiologia.cl/.

REFERENCIAS

[1] Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., et al. Tensorflow: Large-scale machine learning on heterogeneous distributed systems. *arXiv preprint arXiv:1603.04467* (2016).

[2] Ali, R. B., Ejbali, R., and Zaied, M. Detection and classification of dental caries in x-ray images using deep neural networks. In *Int. Conf. on Software Engineering Advances (ICSEA)* (2016), p. 236.

[3] Becker, A. S., Marcon, M., Ghafoor, S., Wurnig, M. C., Frauenfelder, T., and Boss, A. Deep learning in mammography: diagnostic accuracy of a multipurpose image analysis software in the detection of breast cancer. *Investigative radiology 52*, 7 (2017), 434–440.

[4] Cai, Z., and Vasconcelos, N. Cascade r-cnn: Delving into high quality object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (2018), pp. 6154–6162.

[5] Chen, H., Zhang, K., Lyu, P., Li, H., Zhang, L., Wu, J., and Lee, C.-H. A deep learning approach to automatic teeth detection and numbering based on object detection in dental periapical films. *Scientific reports 9*, 1 (2019), 3840.

[6] Chen, K., Pang, J., Wang, J., Xiong, Y., Li, X., Sun, S., Feng, W., Liu, Z., Shi, J., Ouyang, W., et al. Hybrid task cascade for instance segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2019), pp. 4974–4983.

[7] Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., and Yuille, A. L. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence 40*, 4 (2017), 834–848.

[8] Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., and Adam, H. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Proceedings of the European conference on computer vision (ECCV)* (2018), pp. 801–818.

[9] Choi, H., and Jin, K. H. Fast and robust segmentation of the striatum using deep convolutional neural networks. *Journal of neuroscience methods 274* (2016), 146–153.

[10] Code, P. W. Object Detection on COCO. https://paperswithcode.com/sota/object-detection-on-coco, 2019. [Online; accessed 12-July-2019].

[11] Dai, J., Li, Y., He, K., and Sun, J. R-fcn: Object detection via region-based fully convolutional networks. In *Advances in neural information processing systems* (2016), pp. 379–387.

[12] Dai, J., Qi, H., Xiong, Y., Li, Y., Zhang, G., Hu, H., and Wei, Y. Deformable convolutional networks. In *Proceedings of the IEEE international conference on computer vision* (2017), pp. 764–773.

[13] Estay, J., Bersezio, C., Arias, R., Fernández, E., Junior, O. B. O., de Andrade, M. F., Núñez, C. C., ESTAY, J., BERSEZIO, C., ARIAS, A., et al. Effect of clinical experience on accuracy and reliability of radiographic caries detection. *Int. J. Odontostomat 11*, 3 (2017), 347–352.

[14] Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., and Thrun, S. Dermatologist-level classification of skin cancer with deep neural networks. *Nature 542*, 7639 (2017), 115.

[15] Everingham, M., Van Gool, L., Williams, C. K., Winn, J., and Zisserman, A. The pascal visual object classes (voc) challenge. *International journal of computer vision 88*, 2 (2010), 303–338.

[16] Fu, C.-Y., Liu, W., Ranga, A., Tyagi, A., and Berg, A. C. Dssd: Deconvolutional single shot detector. *arXiv preprint arXiv:1701.06659* (2017).

[17] Ghafoorian, M., Karssemeijer, N., Heskes, T., Bergkamp, M., Wissink, J., Obels, J., Keizer, K., de Leeuw, F.-E., van Ginneken, B., Marchiori, E., et al. Deep multi-scale location-aware 3d convolutional neural networks for automated detection of lacunes of presumed vascular origin. *NeuroImage: Clinical 14* (2017), 391–399.

[18] Girshick, R. Fast r-cnn. In *Proceedings of the IEEE international conference on computer vision* (2015), pp. 1440–1448.

[19] Girshick, R., Donahue, J., Darrell, T., and Malik, J. Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (2014), pp. 580–587.

[20] Gulshan, V., Peng, L., Coram, M., Stumpe, M. C., Wu, D., Narayanaswamy, A., Venugopalan, S., Widner, K., Madams, T., Cuadros, J., et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *Jama 316*, 22 (2016), 2402–2410.

[21] Hatvani, J., Horváth, A., Michetti, J., Basarab, A., Kouamé, D., and Gyöngy, M. Deep learning-based super-resolution applied to dental computed tomography. *IEEE Transactions on Radiation and Plasma Medical Sciences 3*, 2 (2018), 120–128.

[22] He, K., Gkioxari, G., Dollár, P., and Girshick, R. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision* (2017), pp. 2961–2969.

[23] He, K., Zhang, X., Ren, S., and Sun, J. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE transactions on pattern analysis and machine intelligence 37*, 9 (2015), 1904–1916.

[24] He, K., Zhang, X., Ren, S., and Sun, J. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (2016), pp. 770–778.

[25] Hsin, C. Yolo Object Detectors: Final Layers and Loss Functions. https://medium.com/oracledevs/final-layers-and-loss-functions-of-single-stage-detectors-part-1-4abbfa9aa71c/, 2019. [Online; accessed 11-July-2019].

[26] Huang, G., Liu, Z., Van Der Maaten, L., and Weinberger, K. Q. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (2017), pp. 4700–4708.

[27] Huang, G., Sun, Y., Liu, Z., Sedra, D., and Weinberger, K. Q. Deep networks with stochastic depth. In *European conference on computer vision* (2016), Springer, pp. 646–661.

[28] Huang, J., Rathod, V., Sun, C., Zhu, M., Korattikara, A., Fathi, A., Fischer, I., Wojna, Z., Song, Y., Guadarrama, S., et al. Speed/accuracy trade-offs for modern convolutional object detectors. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (2017), pp. 7310–7311.

[29] Lakhani, P., and Sundaram, B. Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks. *Radiology 284*, 2 (2017), 574–582.

[30] Larsson, G., Maire, M., and Shakhnarovich, G. Fractalnet: Ultra-deep neural networks without residuals. *arXiv preprint arXiv:1605.07648* (2016).

[31] Lee, C.-Y., Xie, S., Gallagher, P., Zhang, Z., and Tu, Z. Deeply-supervised nets. In *Artificial Intelligence and Statistics* (2015), pp. 562–570.

[32] Lee, J.-H., Kim, D.-H., Jeong, S.-N., and Choi, S.-H. Detection and diagnosis of dental caries using a deep learning-based convolutional neural network algorithm. *Journal of dentistry 77* (2018), 106–111.

[33] Lee, J.-H., Kim, D.-h., Jeong, S.-N., and Choi, S.-H. Diagnosis and prediction of periodontally compromised teeth using a deep learning-based convolutional neural network algorithm. *Journal of periodontal & implant science 48*, 2 (2018), 114–123.

[34] Li, Y., Chen, Y., Wang, N., and Zhang, Z. Scale-aware trident networks for object detection. *arXiv preprint arXiv:1901.01892* (2019).

[35] Li, Z., Peng, C., Yu, G., Zhang, X., Deng, Y., and Sun, J. Light-head r-cnn: In defense of two-stage object detector. *arXiv preprint arXiv:1711.07264* (2017).

[36] Lin, M., Chen, Q., and Yan, S. Network in network. *arXiv preprint arXiv:1312.4400* (2013).

[37] Lin, T.-Y., Dollár, P., Girshick, R., He, K., Hariharan, B., and Belongie, S. Feature pyramid networks for object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (2017), pp. 2117–2125.

[38] Lin, T.-Y., Goyal, P., Girshick, R., He, K., and Dollár, P. Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision* (2017), pp. 2980–2988.

[39] Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., and Zitnick, C. L. Microsoft coco: Common objects in context. In *European conference on computer vision* (2014), Springer, pp. 740–755.

[40] Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., Van Der Laak, J. A., Van Ginneken, B., and Sánchez, C. I. A survey on deep learning in medical image analysis. *Medical image analysis 42* (2017), 60–88.

[41] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., and Berg, A. C. Ssd: Single shot multibox detector. In *European conference on computer vision* (2016), Springer, pp. 21–37.

[42] Long, J., Shelhamer, E., and Darrell, T. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (2015), pp. 3431–3440.

[43] Miki, Y., Muramatsu, C., Hayashi, T., Zhou, X., Hara, T., Katsumata, A., and Fujita, H. Classification of teeth in cone-beam ct using deep convolutional neural network. *Computers in biology and medicine 80* (2017), 24–29.

[44] Mikołajczyk, A., and Grochowski, M. Data augmentation for improving deep learning in image classification problem. In *2018 international interdisciplinary PhD workshop (IIPhDW)* (2018), IEEE, pp. 117–122.

[45] Pan, W., Gu, W., Nagpal, S., Gephart, M. H., and Quake, S. R. Brain tumor mutations detected in cerebral spinal fluid. *Clinical chemistry 61*, 3 (2015), 514–522.
[46] Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (2016), pp. 779–788.
[47] Redmon, J., and Farhadi, A. Yolo9000: better, faster, stronger. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (2017), pp. 7263–7271.
[48] Redmon, J., and Farhadi, A. Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767* (2018).
[49] Ren, S., He, K., Girshick, R., and Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. In *Advances in neural information processing systems* (2015), pp. 91–99.
[50] Romero, A., Ballas, N., Kahou, S. E., Chassang, A., Gatta, C., and Bengio, Y. Fitnets: Hints for thin deep nets. *arXiv preprint arXiv:1412.6550* (2014).
[51] Ronneberger, O., Fischer, P., and Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention* (2015), Springer, pp. 234–241.
[52] Sarraf, S., and Tofighi, G. Classification of alzheimer's disease using fmri data and deep learning convolutional neural networks. *arXiv preprint arXiv:1603.08631* (2016).
[53] Sermanet, P., Eigen, D., Zhang, X., Mathieu, M., Fergus, R., and LeCun, Y. Overfeat: Integrated recognition, localization and detection using convolutional networks. *arXiv preprint arXiv:1312.6229* (2013).
[54] Shijie, J., Ping, W., Peiyi, J., and Siping, H. Research on data augmentation for image classification based on convolution neural networks. In *2017 Chinese Automation Congress (CAC)* (2017), IEEE, pp. 4165–4170.
[55] Shivaprasad, P. A Comprehensive Guide To Object Detection Using YOLO Framework — Part II (Implementing using Python). https://medium.com/@pratheesh.27998/object-detection-part2-6a265827efe1, 2019. [Online; accessed 12-July-2019].
[56] Singh, B., and Davis, L. S. An analysis of scale invariance in object detection snip. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (2018), pp. 3578–3587.
[57] Singh, B., Najibi, M., and Davis, L. S. Sniper: Efficient multi-scale training. In *Advances in Neural Information Processing Systems* (2018), pp. 9310–9320.
[58] Springenberg, J. T., Dosovitskiy, A., Brox, T., and Riedmiller, M. Striving for simplicity: The all convolutional net. *arXiv preprint arXiv:1412.6806* (2014).
[59] Srivastava, M. M., Kumar, P., Pradhan, L., and Varadarajan, S. Detection of tooth caries in bitewing radiographs using deep learning. *arXiv preprint arXiv:1711.07312* (2017).
[60] Srivastava, R. K., Greff, K., and Schmidhuber, J. Training very deep networks. In *Advances in neural information processing systems* (2015), pp. 2377–2385.
[61] Suk, H.-I., Lee, S.-W., Shen, D., Initiative, A. D. N., et al. Deep ensemble learning of sparse regression models for brain disease diagnosis. *Medical image analysis 37* (2017), 101–113.
[62] Uijlings, J. R., Van De Sande, K. E., Gevers, T., and Smeulders, A. W. Selective search for object recognition. *International journal of computer vision 104*, 2 (2013), 154–171.
[63] Wang, J., and Perez, L. The effectiveness of data augmentation in image classification using deep learning. *Convolutional Neural Networks Vis. Recognit* (2017).
[64] Yu, F., and Koltun, V. Multi-scale context aggregation by dilated convolutions. *arXiv preprint arXiv:1511.07122* (2015).
[65] Zhang, K., Wu, J., Chen, H., and Lyu, P. An effective teeth recognition method using label tree with cascade network structure. *Computerized Medical Imaging and Graphics 68* (2018), 61–70.
[66] Zhang, S., Wen, L., Bian, X., Lei, Z., and Li, S. Z. Single-shot refinement neural network for object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2018), pp. 4203–4212.
[67] Zhao, H., Shi, J., Qi, X., Wang, X., and Jia, J. Pyramid scene parsing network. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (2017), pp. 2881–2890.
[68] Zhao, L., and Jia, K. Multiscale cnns for brain tumor segmentation and diagnosis. *Computational and mathematical methods in medicine 2016* (2016).

10. APÉNDICES

Se han obtenido resultados preliminares utilizando un subconjunto de imágenes utilizando una menor resolución, donde Faster R-CNN logra identificar lesiones de caries dentinarias y dentinarias profundas.

Figura 12: Radiografía Bitewing. Lesión de caries dentinaria profunda.

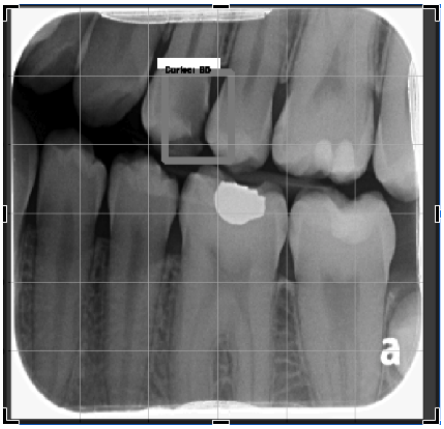

Figura 13: Radiografía Bitewing. Lesión de caries dentinaria profunda.

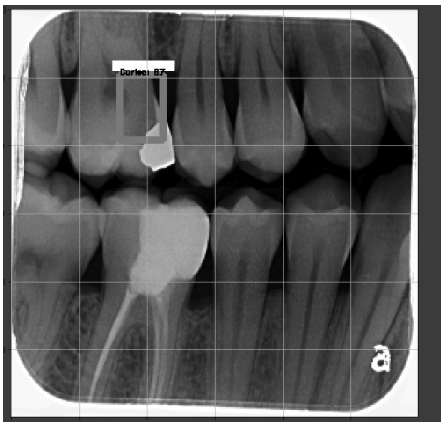

Figura 14: Radiografía Bitewing. Lesión de caries dentinaria.