

# 1 Functional Map of The World Dataset

El dataset usado para el entrenamiento de los modelos fue el llamado Functional Map of the World Dataset. El conjunto de datos "Functional Map of the World" (FMOW) es un conjunto de datos público de imágenes satelitales que se utiliza para tareas de clasificación de objetos y detección de objetos. El conjunto de datos contiene alrededor de 1 millón de imágenes de alta resolución de todo el mundo, que se han etiquetado manualmente con información sobre las clases de objetos presentes en la imagen.

El FMOW se divide en dos partes principales: una parte de entrenamiento y una parte de prueba. La parte de entrenamiento consta de alrededor de 900,000 imágenes etiquetadas, mientras que la parte de prueba contiene alrededor de 100,000 imágenes no etiquetadas. Las imágenes en el conjunto de datos muestran una variedad de paisajes y entornos, incluidas áreas urbanas y rurales, y se capturaron en diferentes momentos del día y en diferentes condiciones climáticas.

Las etiquetas de clase en el FMOW se basan en una taxonomía de objetos llamada "Functional Map of the World" (FMoW). La taxonomía FMoW se centra en las funciones que cumplen los objetos en lugar de en su apariencia física, lo que permite una clasificación más precisa y consistente de los objetos en diferentes contextos y entornos. Las clases de objetos incluyen cosas como edificios, vehículos, cuerpos de agua, cultivos y áreas verdes.

# 2 Modelo Faster R-CNN

Faster R-CNN es un algoritmo popular de detección de objetos que fue introducido por Shaoqing Ren, Kaiming He, Ross Girshick y Jian Sun en 2015. Es una extensión del modelo R-CNN original (Convolutional Neural Network basado en regiones), que fue introducido por Ross Girshick et al. en 2014.

La detección de objetos es una tarea de visión por computadora que implica identificar la presencia y ubicación de objetos en una imagen. En Faster R-CNN, esto se logra primero generando un conjunto de propuestas de región (es decir, ubicaciones de objetos candidatos) utilizando una Red de Proposición de Regiones (RPN), y luego clasificando estas propuestas utilizando una red de clasificación.

La RPN es una red neuronal completamente convolucional que toma una imagen como entrada y produce un conjunto de propuestas de objetos rectangulares, cada una con una puntuación de objetividad asociada. Estas propuestas se generan deslizando una pequeña red sobre el mapa de características convolucionales producido por una red de base pre-entrenada (típicamente una red VGG o ResNet). La RPN se entrena de extremo a extremo con la red de clasificación, utilizando una función de pérdida de múltiples tareas que combina una pérdida de clasificación binaria para la objetividad y una pérdida de regresión para las coordenadas del cuadro delimitador.

La red de clasificación toma cada propuesta generada por la RPN y realiza la clasificación de objetos y la regresión del cuadro delimitador. La red consta de

una serie de capas completamente conectadas que toman las características de cada propuesta como entrada y producen una etiqueta de clase y coordenadas del cuadro delimitador.

Una de las principales ventajas de Faster R-CNN sobre su predecesor, R-CNN, es su velocidad. R-CNN era un algoritmo lento, ya que implicaba ejecutar un modelo de aprendizaje profundo separado para cada propuesta de región. En contraste, Faster R-CNN utiliza un solo mapa de características convolucionales compartido para todas las propuestas, lo que le permite procesar imágenes mucho más rápidamente.

Otra ventaja de Faster R-CNN es su precisión. Se ha demostrado que logra un rendimiento de vanguardia en una serie de benchmarks de detección de objetos, incluidos los conjuntos de datos PASCAL VOC y MS COCO.