Detección de Edificaciones con Machine Learning

Jesús Santos Capote, Kenny Villalobos Morales, Jorge Soler González, Abraham González Rivero, Rainel Fernández Abreu, Eduardo García Maleta

Facultad de Matemática y Computación, Universidad de La Habana, La Habana, Cuba

Keywords: Detección de Objetos · Clasificación de Imágenes · Machine Learning · Pytorch · Keras · Tensorflow · Faster-RCNN

1. Intoducción

El uso de imágenes satelitales ha permitido avances significativos en la identificación y caracterización de diferentes tipos de construcciones, lo cual puede ser de gran utilidad en diversas áreas como la planificación urbana, la gestión de recursos naturales y la seguridad nacional, entre otras. Los autores proponen tres modelos, dos de clasificación de imágenes y uno de detección de objetos con el fin de identificar en imágenes satelitales edificaciones con una topología específica.

2. Dataset

2.1. 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.

Actualmente se encuentra libre para su descarga en Amazon S3.

3. Modelos Utilizados

3.1. 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 con Faster-RCNN 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.

La implementación de este modelo se realizó con la utilización de la biblioteca Pytorch de python y se realizó el entranmiento en Google Colab con la clase Stadium del dataset. Es decir el modelo se entrenó para detectar estadios en imágenes satelitales. De igual forma se puede entrenar para detectar otro tipo de edificación.

3.2. InceptionV3 + DNN

El modelo de machine learning que se presenta utiliza la arquitectura InceptionV3 de redes neuronales convolucionales (CNN) para la extracción de características, seguida de una red neuronal densa (DNN) para la clasificación.

El modelo InceptionV3 es una red neuronal convolucional profunda que ha demostrado ser altamente eficaz en la clasificación de imágenes. Fue desarrollado por Google y se encuentra disponible en la librería de aprendizaje profundo Keras, lo que lo hace accesible para su uso en proyectos de Machine Learning.

El modelo InceptionV3 se caracteriza por su arquitectura en forma de ïnception module", que se basa en la idea de combinar diferentes tamaños de filtros dentro de la misma capa convolucional. Esta técnica permite reducir la cantidad de parámetros que debe aprender el modelo, lo que a su vez reduce el riesgo de sobreajuste y mejora su capacidad de generalización.

Además, InceptionV3 utiliza técnicas como la regularización L2 y el dropout para prevenir el sobreajuste, y utiliza la función de activación ReLU para acelerar el entrenamiento de la red. El modelo también utiliza capas de agrupamiento máximo (max pooling) para reducir el tamaño de las características de la imagen y facilitar su procesamiento.

La red neuronal densa que se agrega a la arquitectura InceptionV3 se utiliza para la clasificación de las características extraídas por la red convolucional. La red consta de dos capas densas completamente conectadas con 128 y 2 neuronas respectivamente. La capa final de la red densa utiliza la función de activación softmax, que normaliza las salidas de la capa anterior para que representen probabilidades de clase para la clasificación multiclase. La función de activación ReLU se utiliza en la capa anterior para introducir no linealidad en la red y mejorar su capacidad de aprendizaje.

El modelo se entrenó para la clasificación de imágenes con aeropuertos y zoológicos.

3.3. DenseNet + ResNet

El modelo consciste en las redes neuronales convolucionales DenseNet121 y ResNet152, ambos pre-entrenados en el conjunto de datos de ImageNet. Las salidas de los modelos se concatenan y se alimentan a una capa completamente conectada y una capa de dropout para reducir el sobreajuste. Finalmente, se agrega una capa de salida con una función de activación softmax para realizar la clasificación en el número de clases especificado.

DenseNet121 es una arquitectura de red neuronal convolucional profunda que se caracteriza por su alta eficiencia en el uso de los parámetros. En lugar de concatenar las salidas de las capas anteriores a través de capas de conexión, como en las redes neuronales convolucionales tradicionales, DenseNet121 conecta todas las capas de la red en una estructura densamente conectada. Esto significa que cada capa recibe como entrada las salidas de todas las capas anteriores, lo que permite que la información fluya de manera más eficiente a través de la red y evita el problema de la desaparición del gradiente. DenseNet121 se utiliza comúnmente en aplicaciones de clasificación de imágenes y ha demostrado un alto rendimiento en conjuntos de datos como ImageNet.

ResNet, por otro lado, es una arquitectura de red neuronal convolucional profunda que se caracteriza por su capacidad para resolver el problema de la degradación de la precisión en redes neuronales muy profundas. La degradación de la precisión se refiere al hecho de que a medida que se agregan más capas a una red neuronal, la precisión de la red comienza a disminuir en lugar de mejorar. Para resolver este problema, ResNet introduce el concepto de çonexiones de salto"(skip connections), que permiten que la información fluya directamente desde las capas anteriores a las capas posteriores, evitando que la información se pierda a medida que se profundiza en la red. ResNet se utiliza comúnmente en tareas de clasificación de imágenes y ha demostrado un alto rendimiento en conjuntos de datos como ImageNet.

EL modelo fue entrenado para identificar las clases: airport terminal, burial site, park, stadium, zoo.

Referencias

1. K. He et al., "Deep residual learning for image recognition," arXiv 1512.03385, Dec 2015.
2. G. Huang, "Dense connected convolutional neural networks," IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), 2017.