Conceptos básicos:

Imágenes aéreas: Las imágenes aéreas son capturadas desde una perspectiva elevada, generalmente por medio de sensores montados en aviones, satélites y drones. Estas imágenes proporcionan una vista panorámica de una amplia área geográfica y se utilizan en diversas aplicaciones, como la cartografía, la monitorización del terreno y la planificación urbana.

Píxel: Es la unidad básica de una imagen digital. Cada píxel representa un punto en la imagen y contiene información de color y/o intensidad. En la segmentación semántica, el objetivo es asignar una etiqueta semántica a cada píxel.

Etiqueta semántica: Es una clase o categoría predefinida a la que se asigna un píxel en la segmentación semántica. Por ejemplo, las etiquetas pueden representar clases como edificios, árboles, carreteras, agua, cultivos, etc.

Segmentación: Consiste en dividir una imagen en regiones o segmentos más pequeños. En la segmentación semántica, el objetivo es asignar una etiqueta semántica a cada segmento, es decir, a cada región o conjunto de píxeles contiguos.

Red Neuronal Convolucional: Es un modelo de aprendizaje profundo diseñado específicamente para procesar datos estructurados en forma de una cuadrícula.

La idea principal detrás de una CNN es aprovechar la estructura espacial de los datos de entrada mediante el uso de capas con filtros aprendibles, llamadas capas convolucionales. Estos filtros convolucionan (se deslizan) sobre los datos de entrada, extrayendo características locales en diferentes ubicaciones espaciales. Al aprender estos filtros a través del proceso de entrenamiento, las CNN pueden capturar automáticamente representaciones jerárquicas de los datos de entrada, detectando características de bajo nivel (por ejemplo, bordes, esquinas) en las capas iniciales y características más complejas (por ejemplo, formas, texturas) en capas más profundas. El proceso de entrenamiento de una CNN implica la propagación hacia adelante, donde los datos de entrada se pasan a través de la red, y la propagación hacia atrás (backpropagation), donde la red aprende ajustando sus pesos en función de los errores calculados. Este proceso se realiza típicamente utilizando algoritmos de optimización como el descenso de gradiente estocástico (SGD) y sus variantes.

Una CNN, o Red Neuronal Convolucional, es un tipo de modelo de aprendizaje profundo utilizado principalmente en tareas de visión por computadora. Su funcionamiento se basa en la idea de aprovechar la estructura espacial de los datos de entrada, como imágenes, mediante el uso de capas convolucionales.

El funcionamiento de una CNN se puede resumir en los siguientes pasos:

Entrada de datos: Se proporciona una imagen o un conjunto de imágenes como entrada a la red.

Capas convolucionales: Estas capas contienen filtros o kernels que se deslizan por la imagen para extraer características locales en diferentes ubicaciones. Cada filtro convolucional genera un mapa de características que resalta una característica específica, como bordes, texturas o formas.

Capas de agrupamiento (pooling): Después de las capas convolucionales, se aplican capas de agrupamiento para reducir la dimensionalidad y resumir la información. El pooling reduce el tamaño de los mapas de características, preservando las características más relevantes y proporcionando cierta invariancia espacial.

Capas totalmente conectadas: Después de las capas convolucionales y de agrupamiento, se suelen agregar capas totalmente conectadas que actúan como una red neuronal tradicional. Estas capas aprenden a mapear las características extraídas a las clases o categorías de salida deseadas.

Funciones de activación: Se utilizan funciones de activación, como ReLU (Rectified Linear Unit), para introducir no linealidad en la red y permitir que la CNN aprenda representaciones más complejas.

Propagación hacia adelante: Durante la propagación hacia adelante, los datos de entrada pasan a través de las capas de la red, y las operaciones de convolución y agrupamiento se aplican sucesivamente para extraer y transformar las características.

Propagación hacia atrás (backpropagation): Después de la propagación hacia adelante, se calcula la pérdida o el error entre la salida predicha y la salida deseada. Luego, se utiliza el algoritmo de backpropagation para ajustar los pesos de la red y minimizar la pérdida, utilizando métodos de optimización como el descenso de gradiente.

Entrenamiento y ajuste de pesos: El proceso de backpropagation y el ajuste de pesos se repiten iterativamente a través del conjunto de datos de entrenamiento para que la red aprenda a mejorar su capacidad de predicción.

En resumen, una CNN utiliza capas convolucionales para extraer características espaciales de los datos de entrada y aprende a reconocer patrones y representaciones jerárquicas en imágenes. Esto la hace especialmente adecuada para tareas de visión por computadora, como clasificación de imágenes, detección de objetos y segmentación semántica.

VIDEOS CNN:

<https://www.youtube.com/watch?v=YRhxdVk_sIs>

<https://www.youtube.com/watch?v=QzY57FaENXg>

Descripción general:

La Segmentación Semántica de Imágenes Aéreas es una técnica dentro del campo del procesamiento de imágenes que tiene como objetivo asignar una etiqueta semántica a cada píxel de una imagen aérea. En otras palabras, se trata de dividir una imagen en diferentes regiones o segmentos y asignar a cada uno de ellos una etiqueta que indique la clase o categoría a la que pertenece.

La segmentación semántica es una tarea fundamental en el análisis de imágenes, ya que proporciona información detallada sobre la distribución espacial de diferentes objetos y elementos presentes en una imagen aérea, como edificios, árboles, carreteras, cuerpos de agua, entre otros. Esta información es de gran utilidad en diversas aplicaciones, como la cartografía, la vigilancia, la agricultura de precisión, la planificación urbana y la detección de cambios.

Para lograr la segmentación semántica, se utilizan diferentes enfoques y técnicas, que van desde métodos clásicos basados en características y clasificadores, hasta enfoques más modernos basados en el aprendizaje profundo (deep learning) y las redes neuronales convolucionales (CNNs). Estos últimos enfoques han demostrado un rendimiento muy superior en la segmentación semántica de imágenes aéreas debido a su capacidad para aprender características y patrones complejos directamente de los datos.

En resumen, la segmentación semántica de imágenes aéreas es una tarea crucial en el procesamiento de imágenes, que permite asignar etiquetas semánticas a cada píxel de una imagen, lo que proporciona información detallada sobre la distribución espacial de objetos y elementos en la escena. Esto tiene aplicaciones prácticas en diversos campos y se ha beneficiado del avance de las técnicas de aprendizaje profundo.

MODELO U-NET

Una de las ventajas de utilizar el modelo U-Net para la segmentación semántica es su capacidad para capturar tanto información local como contextual en las imágenes. Esto se debe a su estructura en forma de U, que incluye conexiones de salto o skip connections entre las capas de la red. Estas conexiones permiten que la información de nivel superior, que contiene características contextuales más amplias, se combine con la información de nivel inferior, que contiene detalles locales más finos.

Al combinar características de diferentes escalas y niveles de abstracción, el modelo U-Net puede capturar tanto las características de bajo nivel, como bordes y texturas, como las características de alto nivel, como formas y objetos completos. Esto le otorga al modelo una capacidad excepcional para realizar segmentaciones precisas y preservar los detalles finos de las regiones segmentadas.

Además, el modelo U-Net ha demostrado ser efectivo incluso en escenarios con conjuntos de datos pequeños debido a su arquitectura compacta y su capacidad para aprender características relevantes de manera eficiente. Esto lo hace adecuado para aplicaciones en las que la disponibilidad de datos etiquetados puede ser limitada.

En resumen, la ventaja del modelo U-Net radica en su capacidad para capturar información local y contextual de manera efectiva, lo que resulta en segmentaciones semánticas precisas y detalladas.

Técnicas y enfoques más utilizados en la segmentación semántica de imágenes aéreas:

Métodos basados en características y clasificadores: Estos métodos clásicos se basan en la extracción de características de los píxeles, como el color, la textura, la forma, entre otros. Luego se utilizan algoritmos de clasificación, como máquinas de vectores de soporte (SVM) o clasificadores bayesianos, para asignar una etiqueta semántica a cada píxel o segmento. Estos enfoques dependen en gran medida de la selección de características relevantes y pueden requerir una calibración cuidadosa de los clasificadores.

Métodos de segmentación basados en regiones: Estos métodos agrupan los píxeles vecinos en regiones o segmentos basados en criterios de similitud, como el color, la textura o la conectividad espacial. Luego, se asigna una etiqueta semántica a cada segmento utilizando técnicas de clasificación o reglas de decisión.

Enfoques de aprendizaje automático: Los métodos de aprendizaje automático, como las redes neuronales convolucionales (CNN), se han vuelto muy populares para la segmentación semántica de imágenes aéreas debido a su capacidad para aprender características y patrones complejos directamente de los datos. Estas redes son entrenadas utilizando grandes conjuntos de datos etiquetados y son capaces de aprender representaciones jerárquicas que capturan información semántica y espacial. Ejemplos de arquitecturas de redes neuronales utilizadas para la segmentación semántica incluyen U-Net, SegNet, Fully Convolutional Networks (FCN) y DeepLab.

Enfoques de segmentación basados en grafos: Estos enfoques modelan la imagen como un grafo, donde los píxeles son los nodos y las conexiones entre ellos representan las relaciones espaciales. Se utilizan técnicas de segmentación de grafos, como el corte mínimo o la propagación de etiquetas, para asignar etiquetas semánticas a los nodos o segmentos del grafo.

Métodos de fusión de información multi-escala: Dado que las imágenes aéreas contienen información a diferentes escalas, se utilizan técnicas de fusión de información multi-escala para mejorar la segmentación semántica. Estos métodos combinan características extraídas a diferentes niveles de detalle para capturar tanto información local como contextual.

Transferencia de conocimiento y entrenamiento con datos sintéticos: Debido a la escasez de grandes conjuntos de datos etiquetados para imágenes aéreas, se utilizan técnicas de transferencia de conocimiento y entrenamiento con datos sintéticos. La transferencia de conocimiento consiste en preentrenar una red neuronal en un conjunto de datos grande y luego afinarla en un conjunto de datos más pequeño y específico. Además, se generan datos sintéticos para aumentar la cantidad de datos de entrenamiento y mejorar el rendimiento del modelo.

<https://www.superannotate.com/blog/guide-to-semantic-segmentation#what-is-semantic-segmentation>

FÓRMULAS MATEMÁTICAS:

En la segmentación semántica, se utilizan varios conceptos y fórmulas matemáticas para realizar el procesamiento de imágenes y extraer características relevantes. A continuación, se presentan algunos de los conceptos y fórmulas comunes:

Histograma de color: El histograma de color es una representación gráfica de la distribución de los valores de color en una imagen. Se calcula contando el número de píxeles que tienen un valor de color específico en cada canal de color (rojo, verde y azul). El histograma de color se utiliza a menudo como una característica para diferenciar diferentes regiones en la imagen.

Transformaciones de color: Las transformaciones de color, como la transformación de espacio de color, se utilizan para cambiar la representación de color de una imagen. Ejemplos comunes son la conversión de RGB a escala de grises, la conversión a espacios de color como HSV (matiz, saturación, valor) o LAB (luminosidad, a, b), que pueden proporcionar características más distintivas para la segmentación.

Filtros de imagen: Los filtros de imagen, como los filtros de suavizado (por ejemplo, filtro Gaussiano) y los filtros de realce de bordes (por ejemplo, el operador de Sobel), se utilizan para mejorar la calidad de la imagen y resaltar características relevantes, como bordes y texturas.

Operaciones morfológicas: Las operaciones morfológicas, como la dilatación y la erosión, se utilizan para modificar la forma y la estructura de los objetos en una imagen. Estas operaciones son útiles para eliminar ruido, rellenar huecos y resaltar características específicas en la imagen.

Funciones de distancia: Las funciones de distancia, como la distancia euclidiana o la distancia de Mahalanobis, se utilizan para medir la similitud entre dos características o vectores de características. Estas funciones de distancia se emplean a menudo en algoritmos de clasificación y segmentación para asignar píxeles o regiones a clases específicas.

Algoritmos de clustering: Los algoritmos de clustering, como el algoritmo de k-medias o el algoritmo de segmentación por crecimiento de regiones, se utilizan para agrupar píxeles o regiones similares en segmentos o clases. Estos algoritmos se basan en medidas de similitud y buscan agrupar elementos que sean similares y separar aquellos que sean diferentes.

Funciones de energía y optimización: En algunos enfoques de segmentación semántica, se utilizan funciones de energía y técnicas de optimización para encontrar la configuración óptima de etiquetas semánticas. Estas funciones de energía consideran la coherencia espacial y la similitud de características entre píxeles o regiones vecinas, y se optimizan utilizando métodos como la minimización de cortes o la programación dinámica.

Estos son solo algunos ejemplos de los conceptos y fórmulas matemáticas utilizados en la segmentación semántica. Dependiendo del enfoque específico y del algoritmo utilizado, puede haber otros conceptos y fórmulas involucrados.

LIBRO PAGINA 966