

Recibido el 1 de enero de 2020, aceptado el 16 de enero de 2020, fecha de publicación 13 de febrero de 2020, fecha de la versión actual 18 de marzo de 2020.

Identificador de objeto digital 10.1109/ACCESS.2020.2973625

DeepFood: Análisis de imágenes de alimentos y dietética

Evaluación mediante modelo profundo

LANDU JIANG^{1,2}, (Miembro, IEEE), BOJIA QIU^{1,2}, XUE LIU^{1,2}, (Miembro, IEEE),

CHENXI HUANG Y KUNHUI LIN¹ Escuela de Informática, ¹

Universidad de Xiamen, Xiamen 361005, China ²Escuela de Ciencias de la Computación, Universidad McGill, Montreal, QC H3A 2A7,

Canadá Autores correspondientes: Xue Liu (xueliu@cs.mcgill.ca), Chenxi Huang (supermonkeyy@xmu.edu.cn) y Kunhui Lin (linkunhuixmu@163.com)

RESUMEN La comida es esencial para la vida humana y ha sido la preocupación de muchas convenciones de salud.

Hoy en día, las nuevas herramientas de evaluación dietética y análisis nutricional brindan más oportunidades para ayudar a las personas a comprender sus hábitos alimentarios diarios, explorar patrones nutricionales y mantener una dieta saludable. En este artículo, desarrollamos un sistema de evaluación dietética y reconocimiento de alimentos basado en un modelo profundo para estudiar y analizar alimentos a partir de imágenes de comidas diarias (por ejemplo, capturadas por un teléfono inteligente). Específicamente, proponemos un algoritmo de tres pasos para reconocer imágenes de múltiples elementos (alimentos) mediante la detección de regiones candidatas y el uso de una red neuronal convolucional profunda (CNN) para la clasificación de objetos. El sistema primero genera múltiples regiones de propuestas en imágenes de entrada aplicando la Red de propuestas de región (RPN) derivada del modelo Faster R-CNN. Luego identifica cada región de propuestas al mapearlas en mapas de características y las clasifica en diferentes categorías de alimentos, además de ubicarlas en las imágenes originales. Finalmente, el sistema analizará los ingredientes nutricionales en función de los resultados del reconocimiento y generará un informe de evaluación dietética calculando la cantidad de calorías, grasas, carbohidratos y proteínas. En la evaluación, llevamos a cabo experimentos extensos utilizando dos conjuntos de datos de imágenes de alimentos populares: UEC-FOOD100 y UEC-FOOD256. También generamos un nuevo tipo de conjunto de datos sobre alimentos basado en FOOD101 con límites. El modelo se evalúa a través de diferentes métricas de evaluación. Los resultados experimentales muestran que nuestro sistema es capaz de reconocer los alimentos con precisión y generar el informe de evaluación dietética de manera eficiente, lo que beneficiará a los usuarios con una visión clara de una dieta saludable y guiará su receta diaria para mejorar la salud y el bienestar del cuerpo.

ÍNDICE TÉRMINOS Computación de alimentos, salud, evaluación dietética, análisis nutricional, reconocimiento de imágenes, visión por computadora.

1. INTRODUCCIÓN

Según el comunicado de la Organización Mundial de la Salud (OMS) [47], la obesidad y el sobrepeso se definen como una acumulación anormal o excesiva de grasa que presenta un riesgo para la salud.

Afirma que la causa fundamental de estos problemas es un desequilibrio energético entre las calorías consumidas y gastadas. Desde 2016, ya hay más de 1.900 millones de adultos con sobrepeso en el mundo, y la epidemia de obesidad ha ido creciendo de manera constante, pero hasta ahora ningún país ha podido revertirla. En los Estados Unidos, en 2019, las tasas de obesidad en adultos ahora superan el 35% en nueve estados y el 30% en 31 estados, el costo médico anual de la atención médica para enfermedades relacionadas con la obesidad, incluidas enfermedades cardíacas, accidentes cerebrovasculares, diabetes tipo 2 y ciertos tipos de cáncer, es a

El editor asociado que coordina la revisión de este manuscrito y Quien lo aprobó para su publicación fue Ying Song.

La asombrosa cifra de 190,2 mil millones de dólares estadounidenses [16], y el costo médico para las personas con obesidad fue \$1,429 más alto que el de las personas con peso normal [22]. Aunque hay varios factores que pueden causar obesidad, como ciertos medicamentos, problemas emocionales como el estrés, menos ejercicio y mala calidad del sueño, el comportamiento alimentario: qué y cómo comen las personas es siempre el principal problema que resulta en el aumento de peso.

Las calorías, así como otros ingredientes nutricionales como grasas, carbohidratos y proteínas, son medidas de energía [6], [59]. Cada vez hay más personas a las que les gustaría realizar un seguimiento de lo que comen y de la cantidad de contenido nutricional que reciben todos los días para ver si llevan una dieta saludable. Por lo tanto, una estimación precisa de la ingesta calórica de la dieta será muy importante para el bienestar. Además, el rápido desarrollo del Internet de las cosas (IoT) y la explosión de datos mejoran la experiencia del usuario de las redes sociales [8].

La gente está dispuesta a grabar, cargar y compartir imágenes de alimentos en sitios web como Yelp [58], Dianping¹ y Yummly², por lo que ahora es más conveniente que nunca encontrar una gran cantidad de datos (imágenes y videos) relacionados con los alimentos. Es aún más necesario que nunca desarrollar herramientas de detección y reconocimiento de alimentos a partir de imágenes para promover el análisis de los ingredientes nutricionales de los recibos y realizar un seguimiento del hábito personal de comer y beber.

Actualmente, existen tres tipos de métodos más utilizados [1] para evaluar manualmente la ingesta dietética, incluidos registros dietéticos [4], recordatorios de 24 horas [41] y cuestionario de frecuencia de alimentos (FFQ) [42]. Para los registros de dieta, los sujetos deben registrar los alimentos y bebidas consumidos durante tres días consecutivos (dos días laborables y un día de fin de semana). Personal capacitado debe proporcionar instrucciones detalladas sobre cómo registrar la ingesta y los registros completos deben ingresarse en una aplicación como el Sistema de datos nutricionales para la investigación (NDSR) para su análisis. Al aplicar el recordatorio de 24 horas, se pide a los sujetos que informen sobre todos los alimentos consumidos en las últimas 24 horas, lo que se puede hacer mediante una llamada telefónica o una entrevista cara a cara. Es necesario recopilar y analizar los datos de los sujetos; personal capacitado realizará una entrevista para obtener más detalles [57].

A los sujetos que utilizan el método FFQ se les pide que informen con qué frecuencia consumieron ciertos alimentos y bebidas durante un período de tiempo específico (p. ej., 1 año). La mayoría de los FFQ están disponibles en formato impreso o electrónico y enumeran preguntas generales sobre la dieta diaria y las prácticas culinarias. Se implementan programas de software para calcular la ingesta de nutrientes multiplicando la frecuencia informada de cada alimento por la cantidad de nutrientes en cada alimento [62].

Aunque ya contamos con estos métodos estándar para informar información sobre la dieta, existe al menos un inconveniente que no podemos ignorar: dichos métodos todavía adolecen de sesgos, ya que se requiere que el sujeto estime su ingesta dietética por sí mismo. La evaluación dietética realizada por los participantes puede dar lugar a una subestimación y sobrestimación de la ingesta de alimentos [48]. Para eliminar el sesgo y mejorar la precisión del autoinforme, se han propuesto muchos sistemas de seguimiento de la alimentación automáticos o semiautomáticos. Además, recientemente hay un número cada vez mayor de aplicaciones creadas en plataformas móviles (es decir, teléfonos inteligentes) para el análisis de alimentos.

Por ejemplo, Zhu et al. [65] propusieron un método de clasificación de alimentos basado en segmentación para la evaluación dietética. Su objetivo es determinar las regiones donde se ubica un alimento en particular en una imagen donde se ubica un alimento en particular e identificarlas correctamente mediante técnicas de visión por computadora.

Pouladzadeh et al. desarrollaron otro sistema de medición de calorías de alimentos basado en la nube. [49] utilizando Support Vector Machine (SVM) para reconocer alimentos y calcular las calorías de cada artículo.

Por otro lado, con el rápido desarrollo de la inteligencia artificial y los algoritmos de aprendizaje profundo, es más preciso y eficiente analizar imágenes de alimentos utilizando modelos profundos.

El trabajo de Pouladzadeh et al. [50] describen un sistema de asistencia de medición de calorías para pacientes y médicos. El sistema logra una buena precisión de reconocimiento en porciones individuales de alimentos aprovechando redes neuronales convolucionales profundas.

Es capaz de registrar y medir la ingesta diaria de calorías del usuario de una manera conveniente e inteligente. En este artículo presentamos un sistema automático de reconocimiento de alimentos y análisis nutricional. Con una sola foto tomada con dispositivos móviles, nuestro sistema puede reconocer los alimentos y analizar el contenido nutricional de la comida, además de proporcionar al usuario un informe completo de evaluación dietética de lo que tiene. Con este sistema automático, ayudará al usuario a medir su ingesta diaria y le permitirá garantizar su salud a largo plazo. Nuestra contribución se resume en lo siguiente:

- En este artículo, exploramos un enfoque profundo basado en modelos para el reconocimiento de alimentos y la evaluación dietética. Específicamente, diseñamos e implementamos un sistema para el análisis de imágenes de alimentos: generamos la cantidad de ingredientes nutricionales de cada alimento a partir de imágenes capturadas diariamente. Se generará un informe de evaluación dietética exhaustivo en función de lo que ingiera durante la comida.
- Aprovechamos modelos de redes neuronales convolucionales profundas (por ejemplo, Faster R-CNN) para la detección e identificación de alimentos. En primer lugar, utilizamos la red de propuestas regionales para generar miles de propuestas regionales a partir de la imagen de entrada. Luego, una red neuronal convolucional profunda de última generación (es decir, VGG-16) para extraer los mapas de características de cada propuesta y clasificarlos como diferentes alimentos. Para reducir el tiempo de procesamiento, también aplicamos un módulo de regresión para ubicar cada alimento en la imagen.
- Generamos/revisamos un nuevo conjunto de datos relacionados con alimentos, denominado FOOD20-with-bbx, basado en el conjunto de datos existente FOOD101 [11]. Dado que el conjunto de datos original (FOOD101) solo contiene una etiqueta de categoría de imagen. Seleccionamos las 20 categorías más comunes del conjunto de datos y etiquetamos manualmente cada alimento en la imagen con información del cuadro delimitador. Al agregar esta información, nuestro conjunto de datos se puede utilizar en la evaluación

de otros modelos de detección. • Realizamos experimentos extensos para probar nuestro modelo de detección de alimentos y sistema de análisis nutricional. Utilizamos los conjuntos de datos UEC-FOOD100 [38] y UEC-FOOD256 [32] con información del cuadro delimitador para la evaluación de la detección. Comparamos el rendimiento de diferentes modelos utilizando la precisión media (mAP) y el retraso/velocidad de detección. También entrenamos nuestro modelo con nuestro nuevo conjunto de datos FOOD20 con cuadro delimitador para probar la solidez de nuestro modelo en escenarios con una gama más amplia de categorías de alimentos.

El resto del documento está organizado de la siguiente manera. En la sección II, presentamos una revisión exhaustiva de la literatura sobre diferentes enfoques para la detección de alimentos y el análisis nutricional. Se introducen varios sistemas de reconocimiento de alimentos y análisis de dietas de última generación. En la sección III, discutimos la motivación del proyecto y los desafíos de diseño del sistema propuesto. En el apartado IV elaboramos toda la estructura de los alimentos.

¹<http://www.dianping.com/>

²<https://www.yummly.com/>

sistema de análisis, así como el diseño detallado y la implementación de módulos de aprendizaje profundo. En la sección V, evaluamos nuestro sistema propuesto utilizando diferentes tipos de conjuntos de datos de imágenes. En la sección VI concluimos este artículo y discutimos posibles trabajos futuros.

II. TRABAJOS RELACIONADOS

A. RECONOCIMIENTO DE ALIMENTOS BASADO EN CARACTERÍSTICAS GEOMETRÍAS

El reconocimiento y análisis de categorías de alimentos ha sido un área de investigación popular en el campo del estudio de la nutrición. Sin embargo, es relativamente difícil porque los alimentos son objetos deformables con variaciones significativas en su apariencia. Los alimentos pueden tener una alta variación intracase (alimentos similares, como la carne de res y el bistec, se ven muy diferentes según cómo se cocinan), o una baja variación entre clases (diferentes alimentos como el pescado y el cerdo se ven muy similares).

Se han propuesto diferentes enfoques para reconocer alimentos en imágenes utilizando características geométricas como el descriptor SIFT [37], histogramas de color o GIST [46] y contexto de forma [9]. Además, Felzenszwalb [21] utiliza polígonos triangulados para representar una forma deformable para la detección y Jiang et al. [30] propone aprender una forma media de la clase objetivo basándose en la parametrización spline de placa delgada.

Además, Belongie [9] elige n píxeles de los contornos de una forma y luego forma $n - 1$ vectores como descripción de la forma a nivel de píxel. Aunque los enfoques basados en características geométricas funcionan bien en la detección de objetos para ciertos tipos de artículos, existen dos problemas principales para las tareas relacionadas con los alimentos.

El primer problema es que los métodos basados en características geométricas necesitan detectar características como bordes, contadores y puntos clave o puntos de referencia, que pueden no estar disponibles en las imágenes de alimentos. El otro problema es que es difícil describir la forma de un alimento en el mundo real, por lo que calcular la similitud de formas es muy difícil.

B. RECONOCIMIENTO DE ALIMENTOS BASADO EN MÉTODOS DE CARACTERÍSTICAS ESTADÍSTICAS

Para superar los problemas descritos anteriormente, se proponen enfoques que utilizan características estadísticas. En lugar de bordes o puntos clave, los métodos se centran en características estadísticas locales, como pares de píxeles. Dado que la distribución estadística de características locales por pares podría extraer importantes características de forma y relaciones espaciales entre los ingredientes alimentarios, facilitando así resultados más precisos en el reconocimiento de objetos.

Por ejemplo, Yang et al. [64] exploran las relaciones espaciales entre diferentes ingredientes (por ejemplo, verduras y carne en una comida) empleando un clasificador discriminativo de varios pasos. A cada píxel de la imagen se le asigna un vector que indica la probabilidad de que el píxel pertenezca a nueve ingredientes alimentarios [54]. Se genera un histograma multidimensional utilizando características locales estadísticas por pares, luego el histograma se pasa a una SVM multiclase para la clasificación de imágenes.

C. RECONOCIMIENTO DE ALIMENTOS BASADO EN MÉTODOS DE APRENDIZAJE MÁQUINA

Recientemente, ha habido un número cada vez mayor de investigaciones que realizan experimentos e investigaciones en los campos de la clasificación de alimentos, aprovechando algoritmos de aprendizaje automático/aprendizaje profundo.

Aizawa et al. [5] propusieron un enfoque basado en un marco bayesiano para facilitar el aprendizaje incremental tanto para la detección de alimentos como para la estimación del equilibrio alimentario. Bossard et al. [11] utilizó Random Forest en el conjunto de pruebas Food-101 logrando una precisión de clasificación del 50,67% mediante la extracción de componentes discriminativos. El modelo de bosque aleatorio se utiliza para agrupar los superpíxeles del conjunto de datos de entrenamiento [13]. En el trabajo también se aplicaron otras técnicas de clasificación avanzadas, incluidos los vectores de pesca mejorados (IFV) [52], el histograma de bolsa de palabras (BOW) [36], los bosques de agrupamiento aleatorio (RCF) [40] y el sistema discriminativo de nivel medio. Superpíxeles (MLDS) [56].

A medida que el poder computacional se fortalece, la red neuronal convolucional (CNN) y modelos más profundos también se utilizan ampliamente en el reconocimiento de alimentos y brindan un mejor rendimiento. Kagaya et al. [31] aplicaron el modelo CNN en la clasificación de imágenes de alimentos. Lograron una precisión muy alta del 93,8 % en la detección de alimentos y artículos no alimentarios. Los resultados experimentales sobre el reconocimiento de alimentos mostraron que la solución CNN propuesta superó a todos los demás métodos de referencia: logró una precisión promedio del 73,7 % en 10 clases.

Es más, en el trabajo se utiliza un modelo AlexNet ajustado [63]. El método logró resultados prometedores en conjuntos de datos de imágenes públicas de alimentos hasta el momento, con una precisión máxima del 67,7% para UEC-FOOD-256. Hassannejad et al. [26] aplican un modelo CNN de 54 capas para evaluar la efectividad del modelo profundo en la clasificación de imágenes de alimentos. El modelo se basa en las especificaciones de la arquitectura de reconocimiento de imágenes de Google: Inception. Además, GoogLeNet [60] se utilizó en [39] para el reconocimiento de alimentos para construir un sistema Im2Calories en el conjunto de datos Food-101.

Además, los investigadores comienzan a investigar qué características y modelos son más adecuados para el reconocimiento de alimentos y los incorporan al sistema de análisis de alimentos para calcular las calorías [7], [23], [33], [34]. Para estimar automáticamente las calorías de los alimentos a partir de una imagen de alimentos, se utilizan redes neuronales convolucionales multitarea para el aprendizaje simultáneo de calorías, categorías e ingredientes de los alimentos [18]. Es más, también se propone un enfoque de red generativa adversaria para el análisis de imágenes de alimentos [20].

Aunque el reconocimiento de los alimentos y el análisis del contenido nutricional han sido bien discutidos en el trabajo anterior, persisten dos desafíos básicos. En primer lugar, la mayoría de los enfoques se ocupan de la imagen de un solo alimento. En segundo lugar, todavía lleva mucho tiempo (2 segundos en general) detectar y clasificar los alimentos en imágenes. En este artículo, nuestro objetivo es abordar estos problemas y proponer un sistema automático de reconocimiento de alimentos para identificar los alimentos a partir de imágenes y generar informes de evaluación dietética para un plan de atención médica a largo plazo.

III. CONSIDERACIÓN DEL DISEÑO En esta

sección, mostramos nuestra consideración del diseño del sistema de reconocimiento de alimentos y evaluación dietética. Primero presentamos nuestra motivación y objetivos, luego exploramos los principales desafíos técnicos para construir el sistema.

A. MOTIVACIÓN

1) POR QUÉ NECESITAMOS UN SISTEMA AUTOMÁTICO Con

el rápido desarrollo de la informática inteligente y el Internet de las cosas (IoT), ahora tenemos una enorme cantidad de datos de redes sociales y redes móviles todos los días. La gente sigue subiendo, compartiendo y grabando lo que hacen todos los días en caso de perder la oportunidad de utilizarlos para mejorar nuestra vida diaria. Las imágenes de alimentos, las recetas y los diarios de alimentos se convierten en la información más popular para compartir; podemos aprender las implicaciones de construir un sistema automático de análisis nutricional aprovechando conjuntos de datos a gran escala. Con la ayuda de sistemas de reconocimiento y análisis de alimentos, los usuarios pueden registrar sus comidas diarias y evaluar hábitos dietéticos, además de promover su salud.

2) LA IDENTIFICACIÓN DE ALIMENTOS NO ES

FÁCIL Los métodos de visión por computadora hacen que el proceso de análisis de alimentos sea más confiable y preciso. Pero aún queda un desafío por abordar: cómo reconocer correctamente los diferentes tipos de alimentos en un plato/imagen. Aunque hay una gran cantidad de herramientas de reconocimiento de imágenes disponibles, los métodos para la identificación de alimentos todavía se basan en gran medida en las ingestas dietéticas autoinformadas. Esto se debe principalmente a que los alimentos suelen ser deformables en comparación con otros objetos del mundo real. Es difícil definir la estructura de un alimento, y también existe una varianza alta dentro de una clase (los alimentos similares se ven muy diferentes) y una baja entre clases (los diferentes alimentos se ven muy similares). En la Figura 1 se muestra un ejemplo de baja variación entre clases (Fuente: <https://www.cupcakeproject.com/cupcake-vs-muffin-update/>).



FIGURA 1. Ejemplo de baja varianza entre clases. Muffin y Cupcake son dos tipos diferentes de alimentos, pero se parecen en apariencia (por ejemplo, en sus características de forma).

3) CLASIFICACIÓN VS DETECCIÓN

Cómo definir el objetivo correcto: la clasificación de imágenes o la detección de objetos es muy importante, especialmente para el reconocimiento de alimentos

y escenarios de análisis nutricional en nuestro proyecto. Necesitaremos detección de objetos si nuestro objetivo es identificar los objetos en una imagen de comida, por ejemplo, contar la cantidad de manzanas en la imagen. Una imagen puede contener diferentes categorías de alimentos, y las técnicas de detección nos permiten reconocer qué tipo de alimentos tenemos en la comida. Con base en dichos resultados de detección, podemos calcular las calorías y analizar la nutrición de los alimentos.

B. DESAFÍOS

Aunque la idea de nuestro proyecto parece simple, en la práctica surgen muchos desafíos. A continuación se abordan tres desafíos principales en el reconocimiento y análisis de imágenes de alimentos reales:

1) REGIÓN DE INTERÉS En

este proyecto, necesitamos reconocer varios alimentos a partir de una sola imagen. Como se ilustra en la Figura 2 [38], una imagen de una comida real puede contener varios tipos de alimentos.

Por lo tanto, es deseable un esquema de identificación que detecte y reconozca múltiples alimentos a partir de una única imagen.

Hoy en día, la mayoría de las aplicaciones de análisis de alimentos extraen las características visuales de diferentes alimentos de la imagen completa.

Dado que el fondo puede contener objetos no alimentarios con características de forma o color similares a los de los alimentos. Dichos métodos incluyen información de fondo/características que inevitablemente engañarían al detector y afectarían la precisión de la detección. Se han aplicado modelos basados en aprendizaje profundo basados en regiones y han logrado un gran éxito en la detección de objetos. La idea básica de estas soluciones es utilizar diferentes tipos de métodos de propuesta de región para generar las regiones de interés (RoI). Estos RoI disociarán los objetos de destino del fondo, lo que hará mucho más fácil extraer las características para su reconocimiento. Sin embargo, los numerosos modelos de última generación que se implementaron en concursos populares como ImageNet, Large Scale Visual Recognition Challenge (ILSVRC) y MS-COCO aún no se han examinado ampliamente para conjuntos de datos de imágenes de alimentos.



FIGURA 2. Ejemplo de imagen de alimento a analizar.

2) EL RETRASO EN EL RECONOCIMIENTO DE LOS

ALIMENTOS El segundo desafío de este proyecto es reducir el tiempo de procesamiento para la detección y clasificación de los alimentos.

Los algoritmos de detección de objetos y las tareas de reconocimiento siempre consumen mucho tiempo. Básicamente, cuantos más objetos haya en una imagen, más tiempo puede tardar el sistema de clasificación. Es muy importante controlar el tiempo de procesamiento para una aplicación sensible al retraso.

Porque a los usuarios no les gustaría esperar uno o dos minutos para saber cuál es la cantidad de calorías que acaban de ingerir.

Por lo tanto, es necesario diseñar un sistema que procese todas las imágenes con diferentes números de objetos (alimentos) a una velocidad aceptable y que cueste menos tiempo.

3) INFORMACIÓN INSUFICIENTE SOBRE EL CONTENIDO NUTRICIONAL PARA EVALUACIÓN DIETÉTICA

Como sistema de evaluación dietética, identificar una gran cantidad de clases de alimentos y analizar con precisión el contenido nutricional es un gran desafío. Existen varias aplicaciones para teléfonos inteligentes para el análisis de alimentos, como MyFitnessPal [2] y SHealth [3], para ayudar a los usuarios a registrar su ingesta de alimentos. Algunos de los sistemas permiten a los usuarios utilizar cámaras de teléfonos inteligentes para tomar fotografías de los alimentos y reconocerlos. Este trabajo podría ayudar a los usuarios a alcanzar objetivos dietéticos como perder peso, controlar las alergias o mantener una dieta saludable. Todavía no hay suficiente información para realizar un estudio exhaustivo sobre lo que comemos: el número de clases de alimentos (recopiladas/registradas) es limitado. Aunque el crowdsourcing [44] puede convertirse en una solución, el modelo en sí cuesta una cantidad significativa de energía y, por lo tanto, impide su implementación generalizada.

C. EVALUACIÓN ALIMENTARIA Y ANÁLISIS NUTRICIONAL

Según un estudio [19] lanzado por el Instituto Internacional de Ciencias de la Vida (ILSI), se evalúan 43 nuevas herramientas de evaluación dietética basadas en tecnología entre 2011 y 2017, incluidos programas basados en la web, aplicaciones móviles y dispositivos portátiles. Los resultados muestran que la mayoría de las herramientas (79%) se basaron en ingestas dietéticas autoinformadas. Mientras que el 91% de ellos utilizó la entrada de texto y el 33% utilizó imágenes digitales para ayudar a identificar los alimentos. Sólo el 65% de las herramientas tenían bases de datos integradas para estimar energía o nutrientes, y menos del 50% contenían funciones de personalización además de generar informes automáticos.

Debido a las limitaciones de los métodos de autoinforme en las herramientas basadas en la web y en dispositivos móviles, se ha demostrado que aprovechar las imágenes de los alimentos capturadas por las cámaras de los teléfonos inteligentes es una forma eficiente de registrar las comidas diarias y analizar los componentes. Los métodos basados en imágenes o visión por computadora para la evaluación dietética se refieren a cualquier modelo que utilice imágenes/videos de episodios alimentarios para mejorar los esquemas de autoinforme en trabajos anteriores o como forma principal de registrar la ingesta dietética. Estos sistemas están ampliamente probados y validados. Básicamente, las características de las imágenes se extraen para el reconocimiento de alimentos y el programa verifica los alimentos con una imagen que se puede buscar. Luego, los tipos de alimentos y el tamaño de las porciones se comparan con la base de datos de alimentos y nutrientes para estudios dietéticos.

Según el registro, los métodos basados en imágenes podrían hacer que el progreso del análisis de alimentos sea más confiable y preciso. En este artículo, nuestro objetivo es proporcionar un sistema automático de reconocimiento de alimentos y evaluación dietética que tome una fotografía de los alimentos como entrada. Y generar la cantidad de ingredientes nutritivos de

cada alimento de la imagen. Abordaremos los problemas discutidos anteriormente para mejorar la precisión y la velocidad del reconocimiento de alimentos.

IV. DISEÑO DEL SISTEMA

En esta sección, primero brindamos una descripción general de nuestro sistema basado en modelos profundos propuesto para el reconocimiento de alimentos y el análisis nutricional. Luego discutimos los detalles de implementación de cada componente del sistema. Para obtener más información general y detalles técnicos sobre nuestro sistema de análisis de alimentos, remitimos a los lectores a la tesis de este proyecto [10].

A. DESCRIPCIÓN GENERAL DEL SISTEMA

En este artículo, proponemos un sistema basado en aprendizaje profundo para la detección de alimentos y analizamos los componentes nutricionales de cada imagen de comida. Como se muestra en la Figura 3, nuestro modelo consta de tres pasos principales.

- Primero extraemos las regiones de interés (ROI) aplicando la Red de propuesta de región derivada del modelo Faster R-CNN. Los ROI ayudarían a separar los alimentos del fondo y mejorarían la eficiencia del modelo de detección.

3Partes de este trabajo se han utilizado para cumplir parcialmente los requisitos de la tesis de maestría de Bojia Qiu (coautor de este artículo) en la Universidad McGill.

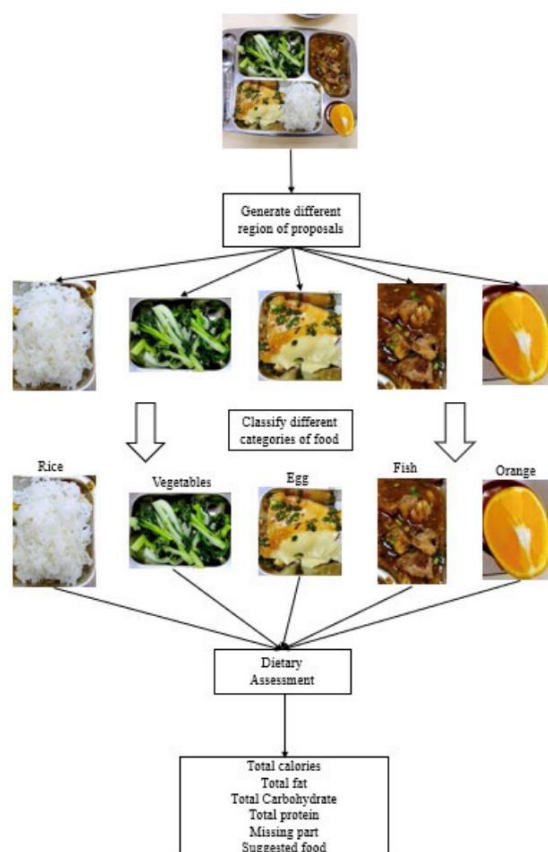


FIGURA 3. El sistema automático de tres pasos de reconocimiento de alimentos y análisis nutricional.

- El segundo paso es aplicar una red neuronal convolucional (CNN) bien diseñada en RoI seleccionadas y clasificarlas en diferentes categorías de alimentos.

Mientras tanto, también se utiliza un módulo de regresión para localizar las coordenadas de los alimentos en la imagen. •

El paso final es utilizar herramientas de evaluación dietética basadas en tecnología moderna para el análisis de la nutrición de los alimentos y generar un informe de salud para los usuarios basado en las imágenes de sus comidas.

Durante la prueba, examinamos varias arquitecturas de redes neuronales profundas diferentes para reemplazar la parte CNN en Faster R-CNN para realizar comparaciones de rendimiento. Por ejemplo, probamos 1. la arquitectura AlexNet [35] con 7 capas para detectar y clasificar objetos conjuntamente, 2. la red VGG-16 [55] para el reconocimiento de objetos desarrollada y entrenada por el renombrado Visual Geometry Group de Oxford, así como la Red Residual Red neuronal (ResNet) [28], que se basa en las construcciones conocidas de las células piramidales de la corteza cerebral.

B. DETECCIÓN BASADA EN LA REGIÓN

Como se analiza en la sección III-B, los enfoques de detección de objetos basados en regiones poseen una precisión líder en el reconocimiento de objetos. Tal como se define, propone diferentes regiones de la imagen de entrada y las clasifica en diferentes categorías.

Los métodos tradicionales de obcción de objetos basados en regiones utilizan una ventana deslizante que recorre la imagen. Hará que todo el proceso sea extremadamente lento, especialmente cuando se emplean modelos profundos de CNN.

Las CNN de detección temprana basadas en regiones (por ejemplo, R-CNN [25] y Fast R-CNN [24]) se basan en propuestas de regiones genéricas de entrada, como la búsqueda selectiva [61] y EdgeBox [17], etc. El proceso lleva mucho tiempo debido a la carga computacional de la generación de propuestas. Para abordar esta cuestión, Ren et al. [51] encontró una manera de hacer que la propuesta de la región sea más eficiente, llamada Faster R-CNN. Faster R-CNN tiene 4 partes principales que incluyen la extracción de características (un módulo convolucional básico con capa de convolución, función de activación relu y capa de agrupación), propuesta de región (anclajes clasificados como región de primer plano o región de fondo), regresión del cuadro delimitador (fijar la ubicación de los anclajes) y clasificación.

En este artículo, aplicamos el modelo Faster R-CNN para detectar los alimentos a partir de las imágenes. Esta sección presenta solo brevemente los aspectos clave de Faster R-CNN; para obtener más detalles técnicos remitimos a los lectores al artículo original [51].

C. DETECCIÓN DE ALIMENTOS

Como se mencionó, para extraer alimentos del fondo de las imágenes, aplicamos Faster R-CNN para detectar las regiones relacionadas con los alimentos. Dado que solo nos enfocamos en objetos de comida, los mapas de características serán menos complejos y resistentes al ruido de fondo. Sin embargo, el modelo de Faster R-CNN que utilizamos fue entrenado con la base de datos VOC2007, que solo contiene 20 tipos comunes de objetos no alimentarios. Por lo tanto, primero seleccionamos una base de datos de imágenes de alimentos, FOOD100 [38], para ajustar el modelo Faster R-CNN previamente entrenado. Luego usamos el modelo para extraer los RoI en cada imagen de comida. La función de pérdida se muestra en la ecuación 1.

$$L(p_i, t_i) = \frac{1}{N_{cls}} \sum_i L_{cl}(p_i, p_{i-1}) + \lambda \frac{1}{N_{reg}} \sum_i L_{reg}(t_i, t_{i-1}) \quad (1)$$

donde i es el índice del ancla en cada lote, p_i es la probabilidad prevista del ancla i . Si el ancla está en primer plano, es igual a 0. t_i representa el cuadro de verdad sobre el terreno para el ancla de primer plano. p_{i-1} es igual a 1, en caso contrario p_i es el correspondiente coordenadas del cuadro delimitador previsto. t_{i-1} es el correspondiente

1) DETALLES DE IMPLEMENTACIÓN

En la implementación, primero cambiamos el tamaño de la imagen ($s = 600$ píxeles y $l = 1000$ píxeles) para que el modelo VGG16 encaje en la memoria de la GPU durante el ajuste fino [24]. También configuramos 3 escalas diferentes para los anclajes en la detección - 128, 256 y 512 píxeles, y las relaciones de aspecto son 1: 1, 1: 2 y 2: 1. Resulta en 9 anclajes diferentes para cada punto en el mapa de características.

Para controlar el número total de anclajes en una imagen para que el modelo sea más fácil de converger, eliminamos los anclajes transfronterizos para que no contribuyan al cálculo de pérdidas en el proceso de entrenamiento. Después del filtrado, quedaron alrededor de 6000 anclajes en una imagen para entrenamiento. Pero aún aplicamos el RPN totalmente convolucional a toda la imagen durante el período de prueba. Aunque reducimos el número de anclajes a 6000, algunos de ellos todavía están muy superpuestos entre sí. Adoptamos la supresión no máxima (NMS) [43] en las regiones propuestas en función de su puntuación de clasificación para reducir aún más la redundancia. El algoritmo NMS escanea la imagen junto con la dirección del gradiente de la imagen para eliminar puntos (establecidos en cero) que no se encuentran en bordes importantes (no en los máximos locales). Después de NMS, una imagen puede tener alrededor de 2000 regiones propuestas para su detección.

Para compartir las capas convolucionales, utilizamos la estrategia de entrenamiento alternativo de 4 pasos en [51] para entrenar el modelo en lugar de aprender dos redes para RPN y el módulo de detección. Específicamente, primero entrenamos el módulo RPN aprovechando un modelo previamente entrenado. Para cada imagen, se selecciona aleatoriamente una muestra de 256 anclajes para la función de pérdida de un mini lote. En segundo lugar, entrenamos el modelo de detección con los anclajes de propuesta generados a partir de RPN. Dos modelos/redes no comparten capas convolucionales en esta etapa. En tercer lugar, utilizamos el modelo de dección para inicializar el entrenamiento RPN con capas convolucionales compartidas fijas. Sólo se ajustan las capas que pertenecen a RPN.

En el cuarto paso, realizamos un entrenamiento similar al modelo de detección usando RPN, por lo que estos dos modelos comparten las mismas capas convolucionales.

2) AUMENTO DE DATOS

Además, también utilizamos el aumento de datos para enriquecer las muestras de entrenamiento y mejorar el rendimiento del sistema. Primero volteamos las imágenes horizontalmente pero no verticalmente. Dado que la mayoría de los alimentos en el mundo real no se voltearán verticalmente. Luego aplicamos la rotación de la imagen a los datos con 90 grados, 180 grados y 270 grados.

D. CLASIFICACIÓN DE ALIMENTOS

1) CLASIFICACIÓN DE OBJETOS EI

módulo de clasificación utiliza los mapas de características propuestos calculados para la clasificación y calcula la puntuación para cada clase. Al mismo tiempo, aplica nuevamente la regresión del cuadro delimitador para mejorar la precisión de la localización de la región de la propuesta.

En este artículo, utilizamos VGGNet como modelo CNN para extraer el mapa de características de las regiones propuestas y realizar una clasificación para el reconocimiento de alimentos. El número máximo de alimentos en la base de datos UECFOOD100 es 5, por lo que establecemos la puntuación más alta de las regiones de cuadros delimitadores en 5. En el módulo de clasificación, utilizamos el modelo VGG-16, que contiene 16 capas de peso en la red con capas completamente conectadas. Se utilizará un vector de características de 4096 dimensiones para clasificar con precisión el objeto en categorías. Dado que hay 100 categorías de alimentos seleccionadas en UEC-FOOD100, la capa softmax contiene 100 unidades.

También cambiaremos la estructura de la capa correspondiente a diferentes conjuntos de datos con diferente número de categorías de alimentos respectivamente.

Además, todos los modelos profundos utilizados en este trabajo se entrenarán previamente con los conjuntos de datos de imágenes naturales y luego se entrenarán utilizando los conjuntos de datos de imágenes de alimentos que propusimos. Las muestras de imágenes se dividieron aleatoriamente en conjuntos de entrenamiento y validación.

2) REGRESIÓN DEL CAJA LÍMITE Para la

regresión del cuadro delimitador, es porque es posible que no podamos proponer una región que cubra perfectamente el objeto. Por lo tanto, necesitamos ajustar el cuadro delimitador inicial para que sus coordenadas coincidan con la verdad fundamental.

Como se muestra en la Figura 4, el cuadro verde G representa la ubicación de la verdad fundamental, mientras que los anclajes de primer plano originales son el cuadro azul P. La regresión del cuadro delimitador es encontrar una transformación que convierta P a G que esté más cerca del límite de la verdad fundamental. cuadro G. Una forma sencilla de lograr esto es, en primer lugar, utilizar el método de traducción en las ecuaciones 2 y 3.

$$G_x = A_w \cdot dx(A) + A_x \quad (2)$$

$$G_y = A_h \cdot dy(A) + A_y \quad (3)$$

Y luego acerque o aleje según las ecuaciones 4 y 5

$$G_w = A_w \cdot \exp(dw(A)) \quad (4)$$

$$G_h = A_h \cdot \exp(dh(A)) \quad (5)$$

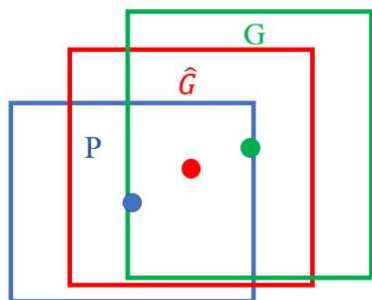


FIGURA 4. Regresión del cuadro delimitador [15].

En este artículo, asumimos que todas las operaciones anteriores son lineales.

Por lo tanto, podemos aprender $dx(A)$, $dy(A)$, $dw(A)$, $dh(A)$ utilizando la regresión lineal para ajustar las coordenadas del cuadro delimitador. La regresión lineal se utiliza para aprender los pesos W para que el vector de características de entrada A coincida con la verdad fundamental t . Donde $t = W \cdot A$.

La pérdida se define como se muestra en la ecuación 6.

$$\text{Pérdida} = \sum_i \left(t^i - \text{PESO} \cdot A^i \right)^2 \quad (6)$$

E. EVALUACIÓN DIETÉTICA

Después del reconocimiento del alimento, el sistema debería poder realizar una evaluación dietética que analice la nutrición de la comida. En este artículo, nos centramos principalmente en calcular el contenido de calorías, grasas, carbohidratos y proteínas de cada imagen de comida. Además, las verduras, los frutos secos y los alimentos integrales también son alimentos saludables que se recomiendan.

En un escenario del mundo real, el sistema debería solicitar al usuario información básica sobre su cuerpo, como edad, sexo, peso, altura y nivel de actividad, etc. Según el perfil, podemos estimar la cantidad adecuada de alimentos y nutrición que necesitan para mantener una vida saludable. dieta. Los usuarios pueden simplemente usar sus teléfonos inteligentes para tomar una fotografía de lo que comen en la comida (por supuesto, antes de comérselo). El sistema reconocerá el alimento en la imagen y luego estimará el estado nutricional de la comida. Se creará un registro de dieta para facilitar el seguimiento y recordar a los usuarios que calculen el número que dejaron.

¿Cómo determinar si el usuario tiene suficiente energía y tipos específicos de alimentos? Utilizamos la fuente estándar para la evaluación dietética: la Base de datos nacional de nutrientes (NNDB) del USDA [45], que contiene datos nutricionales sobre 8618 alimentos básicos. Podemos crear una tabla de referencia de tablas de información nutricional basada en los datos del USDA, incluidos todos los alimentos en nuestros conjuntos de datos. Y luego calcule la cantidad de ingredientes nutricionales para los alimentos capturados por los usuarios asignando los alimentos detectados a la tabla de referencia.

La Tabla 1 muestra un ejemplo de nuestra tabla de referencia de nutrición alimentaria. Para cada alimento, asumimos que el peso es de 400 gramos,

TABLA 1. Fragmento de la tabla de información nutricional de referencia. Cada fila contiene el valor de calorías, grasas, carbohidratos y proteínas.

Food(400g)	Calories	Fat(g)	Carbohydrate(g)	Protein(g)
Steak	1365	63	0	187.3
Ramen	760	29	78	35
Miso Soup	81	3.3	9.8	6.5
Fried Rice	619	12.8	106.8	12.8
Sushi	536	7.7	103.3	13.4
French Fries	428	21.4	57.1	4.8
Takoyaki	1264	92.8	67.2	38.4
Pizza	690	20.4	103.9	26
Hamburger	1086	73.7	0	99.1
...				

que es un tamaño normal para una sola porción. Cada fila de la tabla de referencia presenta la cantidad de calorías, grasas, carbohidratos y proteínas del alimento. Hay 100 categorías de alimentos en el conjunto de datos UEC-FOOD100, por lo que la tabla tendrá 100 filas.

Según la información del perfil del usuario, podemos proporcionar una calculadora de dieta. Por ejemplo, un hombre de 24 años, con un peso de 60 kilogramos y una altura de 170 centímetros, con un nivel de actividad moderado, una cantidad adecuada de ingredientes nutricionales para una dieta saludable será 2399 calorías por día, 311 gramos de carbohidratos, 109 gramos de proteína y 80 gramos de grasa.

V. EVALUACIÓN

En esta sección, evaluamos nuestros modelos y metodologías propuestas, que consta de tres partes: preprocesamiento de datos, entrenamiento de modelos y evaluación del desempeño.

A. CONJUNTOS DE DATOS

En este artículo, utilizamos dos conjuntos de datos del mundo real: UEC-FOOD100 [38], UEC-FOOD256 [32] para evaluar los modelos propuestos. UEC-FOOD100 contiene 12740 imágenes en un total de 100 categorías de diferentes alimentos, mientras que UEC-FOOD256 contiene 31395 imágenes de alimentos dentro de 256 categorías. La mayoría de las imágenes de comida en UEC-FOOD100 son comida japonesa. Hay más de 100 imágenes de cada categoría de alimentos en ambos conjuntos de datos, y la información del cuadro delimitador indica la ubicación de los alimentos en cada foto. La información básica del conjunto de datos UEC-FOOD100 se resume en la Tabla 2.

TABLA 2. Estadísticas del conjunto de datos UEC-FOOD100.

Total number of categories	100
Total number of images	12740
Number of images with single food item	11566
Number of images with multiple food items	1174
Category with largest amount of images	Miso Soup (729)
Category with least amount of images	Chicken Rice (101)

Aunque los conjuntos de datos UEC-FOOD100 y UEC-FOOD256 contienen una gran cantidad de imágenes de alimentos, la mayoría de las muestras son alimentos asiáticos que no pueden usarse para entrenar un modelo más generable para todas las imágenes de alimentos diferentes. Para construir un sistema robusto que funcione para una gama más amplia de categorías de dietas, elegimos otro conjunto de datos muy utilizado: FOOD101 [11], que contiene 101 categorías de alimentos, con 101.000 imágenes. Sin embargo, el conjunto de datos FOOD101 no incluye información del cuadro delimitador en las imágenes. No podemos entrenar el modelo propuesto utilizando muestras en FOOD101. Por lo tanto, generamos manualmente un cuadro delimitador para una parte de las imágenes en el conjunto de datos FOOD101: al menos 300 fotos dentro de cada clase seleccionada. Nombramos subconjuntos del conjunto de datos de imágenes FOOD101 con cuadro delimitador como FOOD20-with-bbx. En la Figura 5 se muestran varios ejemplos de imágenes tomadas de FOOD20-with-bbx.



FIGURA 5. Algunos ejemplos de alimentos en el conjunto de datos FOOD20-with-bbx. Este conjunto de datos incluye un solo alimento.

B. MÉTRICAS DE EVALUACIÓN

En este artículo, utilizamos la precisión promedio media (mAP) para evaluar nuestro modelo de detección. Al igual que en la detección de alimentos, la evaluación no es tan sencilla como otros modelos, porque hay dos tareas distintas que medir al mismo tiempo.

- Clasificación : determinar si un objeto existe en la imagen. •
- Regresión : determinación de la ubicación del objeto.

Además, hay muchas clases de alimentos y la distribución es desequilibrada. Por ejemplo, la cantidad de imágenes en la categoría de arroz en UEC-FOOD100 y UEC-FOOD256 es de cuatro a cinco veces mayor que en otras. Una simple precisión de medición métrica puede introducir sesgos. Por lo tanto, una "puntuación de confianza" asociada a cada cuadro delimitador podría ayudar a evaluar el modelo.

1) EVALUACIÓN DE CLASIFICACIÓN

La precisión promedio (AP) es una métrica que se utiliza frecuentemente para medir la precisión de los detectores de objetos [29]. Calcula el valor de precisión promedio para el valor de recuperación de 0 a 1.

Para cada clase, AP (7) es básicamente el área bajo la curva de recuperación de precisión, mientras que la curva de recuperación de precisión se calcula a partir de la salida de detección del modelo. Para simplificar, dividimos los valores de recuperación de 0 a 1,0 en 11 puntos: 0, 0,1, 0,2,..., 1,0, y calcular la precisión máxima medida en cada conjunto (8).

$$PA = \frac{1}{0}$$

Precisión (recuperación) d (recuperación)

(7)

$$PA = \frac{1}{11}$$

Precisión(Recalli)

(8)

Como la precisión y la recuperación están siempre entre 0 y 1. Por lo tanto, AP se encuentra entre 0 y 1.

C. EVALUACIÓN DE REGRESIÓN

Para evaluar qué tan bien el modelo de regresión estima la ubicación del objeto, utilizamos la intersección sobre

Métrica de unión (IoU) para medir la superposición entre dos límites [29]. Como se ilustra en 9, calculamos el área de superposición entre el cuadro delimitador previsto y la verdad fundamental (el límite del objeto real). Y el área de unión es el área abarcada por dos fronteras.

$$\text{IoU} = \frac{\text{Área de superposición}}{\text{Área de unión}} \quad (9)$$

En la prueba se utiliza un umbral IoU predefinido para determinar si la detección del objeto es positiva o negativa. Los umbrales de IoU pueden variar para diferentes tareas y establecimos el umbral de IoU en 0,5 en este proyecto. Finalmente, la puntuación media de precisión promedio (mAP) se puede calcular tomando la AP media de todas las clases de objetos con un umbral de IoU de 0,5.

D. CONFIGURACIÓN DEL EXPERIMENTO

1) PLATAFORMA EXPERIMENTAL

Implementamos nuestro modelo profundo utilizando Keras [14], que es una API de red neuronal de alto nivel escrita en Python. En el back-end utilizamos la plataforma TensorFlow [12] diseñada por Google.

E. RESULTADOS DEL EXPERIMENTO

1) LÍNEAS DE BASE Y AJUSTES DEL EXPERIMENTO

Seleccionamos dos modelos de referencia para la detección de alimentos: R-CNN [25] y un modelo de segmentación de imágenes de alimentos basado en CNN propuesto por Shimoda et al. [53]. Seguimos la configuración del entorno como trabajo de Shimoda para proporcionar una comparación razonable. Dividimos el conjunto de datos UEC-FOOD100 con un 80% como datos de entrenamiento y el 20% restante como conjunto de prueba con un impulso de 0,9 y una tasa de disminución de peso de 0,0005.

2) RESULTADOS DE LA DETECCIÓN a:

CONJUNTO DE DATOS UEC-FOOD100

Debido a las limitaciones de nuestros recursos informáticos, no podemos igualar la misma configuración: 40000 iteraciones que lo hizo el modelo de referencia [53]. Así, la pérdida de entrenamiento durante las primeras 239 épocas de entrenamiento se muestra en la Figura 6.

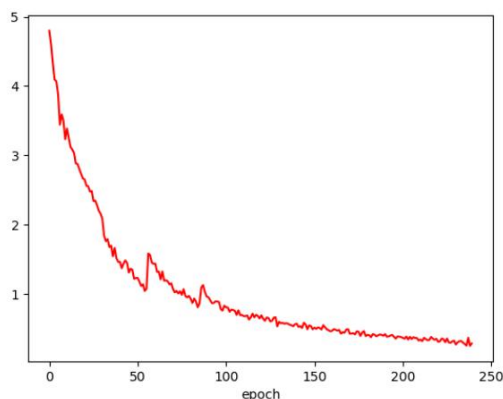


FIGURA 6. Pérdida de entrenamiento para las 239 épocas durante el paso de entrenamiento para el conjunto de datos UEC-FOOD100.

Calculamos el mapa de 100 clases de alimentos en el conjunto de datos. Sin embargo, para comparar con el modelo de referencia [53], seleccionamos 53 categorías de alimentos que han sido probados.

Cada categoría contiene al menos 10 imágenes y más de 11 categorías contienen más de 50 imágenes. El resultado de la detección de UEC-FOOD100 se muestra en la Tabla 3. Adquirimos los resultados del modelo de referencia del artículo/informe original [53]. Podemos ver en la Tabla 3 que, aunque tenemos muchas menos iteraciones en la sesión de entrenamiento, nuestro modelo propuesto aún supera al modelo R-CNN en ciertas categorías de alimentos. El conjunto que obtiene mejores resultados (el Conjunto 3 en comparación con el Conjunto 1) se debe a que hay más imágenes de muestra en el conjunto de datos, por lo tanto, con un tamaño de conjunto de datos grande con más datos de entrenamiento, nuestro modelo podría ser mucho más alto. exactitud.

TABLA 3. Los resultados en el conjunto de datos UEC-FOOD100. El conjunto 1 es el experimento que utilizó todas las clases de alimentos. El conjunto 2 es el experimento que prueba 53 clases (más de 10 elementos en el conjunto de datos de prueba). El conjunto 3 es el experimento que prueba con 11 clases (más de 50 elementos en el conjunto de datos de prueba).

mAP(%)	Set 1	Set 2	Set 3
RCNN-based model	26.0	21.8	25.7
BP-based model	49.9	55.3	55.4
Our model	17.5	23.1	25.5

La razón principal por la que nuestro modelo da como resultado un mAP bajo en estos experimentos es la limitación de los recursos informáticos, ya que hicimos un equilibrio entre el tiempo de entrenamiento (número de iteraciones) y el rendimiento del sistema. Básicamente, otros modelos pueden requerir más de 20000 iteraciones para entrenar un R-CNN más rápido, mientras que nosotros solo realizamos entre 250 y 400 iteraciones. Por lo tanto, somos positivos sobre nuestra solución propuesta. El modelo podría lograr una mayor precisión con mejores resultados de localización del método RoI. Por ejemplo, la precisión de la clasificación podría alcanzar el 98 % en la capacitación cuando la red de propuesta de región pudiera generar cuadros delimitadores correctos. Creemos que con recursos computacionales más potentes, podemos lograr un resultado mucho mejor con un número de interacción mucho mayor.

En la Figura 7 se muestran varios ejemplos de elementos detectados con éxito. Además, también probamos la velocidad de nuestro modelo. Dado que el tiempo de procesamiento depende tanto del diseño de la implementación como del hardware utilizado, solo presentamos una idea general de que nuestro modelo propuesto superó a los otros dos enfoques básicos. El R-CNN tiene el mayor retraso ya que genera las propuestas de región utilizando el algoritmo de búsqueda selectiva [61], lo que consume una cantidad considerable de tiempo. Si bien el modelo de Shimoda es mucho más rápido en la clasificación, todavía utilizan la búsqueda selectiva que puede tener un cuello de botella.



FIGURA 7. Ejemplos del resultado de la detección de UEC-FOOD100.

En este artículo, utilizamos la red de propuesta de región (RPN) [51] en lugar del algoritmo de búsqueda selectiva y simultáneamente entrenamos el modelo para la clasificación.

b: CONJUNTO DE DATOS UEC-FOOD256

También utilizamos el conjunto de datos UEC-FOOD256 (contiene 256 categorías de alimentos) para examinar el rendimiento de nuestro modelo propuesto. Como se muestra en la Tabla 4, hay 132 categorías con más de 10 artículos y 21 categorías con más de 50 artículos alimenticios en el conjunto de datos UEC-FOOD256. Podemos ver en la tabla que los valores de mAP para los tres conjuntos son peores en comparación con el conjunto de datos UEC-FOOD100. Puede deberse al aumento del número de clases de alimentos. Además, hay muchas categorías de alimentos que tienen características similares (por ejemplo, forma y color), como la sopa de miso y la sopa de miso con carne, que hacen que sea aún más difícil distinguirlas.

TABLA 4. Los resultados en el conjunto de datos UEC-FOOD256. La categoría 1 es el experimento que prueba con todas las clases de alimentos. La categoría 2 es el experimento que prueba con 132 clases (más de 10 elementos en el conjunto de datos de prueba). La categoría 3 es el experimento que prueba con 21 clases (más de 50 elementos en el conjunto de datos de prueba).

mAP(%)	Set 1	Set 2	Set 3
Our model	10.5	13.3	18.3

c: CONJUNTO DE DATOS FOOD20-CON-BBX

Además, también examinamos nuestro modelo utilizando nuestro conjunto de datos automodificado, FOOD20-with-bbx. Los datos de FOOD20 [11] contienen una gran cantidad de imágenes de alimentos occidentales que podrían usarse para probar la solidez de nuestro modelo propuesto. En FOOD20-with-bbx, seleccionamos 20 categorías de alimentos delimitados por cuadros delimitadores con coordenadas.

Para ahorrar tiempo, utilizamos el modelo previamente entrenado por el conjunto de datos UEC-FOOD100. El 80% de las muestras de imágenes en FOOD20-with-bbx se utilizaron en la sesión de entrenamiento y el 20% restante de los datos se utilizó para examinar los valores de mAP. Los resultados del experimento se muestran en la Tabla 5. Con 500 iteraciones, nuestro modelo propuesto podría alcanzar la precisión superior en el conjunto de datos FOOD20-with-bbx con un 71,7% y la precisión entre los cinco primeros llega al 93,1%.

TABLA 5. Los resultados en FOOD20-with-bbx.

# of iteration	Top-1 accuracy (%)	Top-5 accuracy (%)
100	28.8	61.1
250	43.6	81.2
500	71.7	93.1

Los resultados del reconocimiento de alimentos en FOOD20-with-bbx son relativamente mejores que los otros dos conjuntos de datos. Esto se debe principalmente a que la mayoría de las imágenes en el conjunto de datos FOOD20-with-bbx solo contienen un solo alimento, y la mayoría de los alimentos son de estilo occidental, lo que hace que sea más fácil distinguirlas según las características de forma y textura.

F. EVALUACIÓN DIETÉTICA

Para promover una dieta saludable, el paso final del sistema propuesto es analizar el contenido de nutrientes de cada imagen de alimento. Como se analizó en la sección anterior, para cada alimento detectado en la imagen, asumimos que el peso básico es de 400 gramos por artículo, que es un tamaño normal para una sola porción. La Figura 8 es un ejemplo de evaluación dietética en nuestro sistema. Con base en la información recopilada de la imagen, podríamos resumir el informe de evaluación dietética para el usuario todos los días.

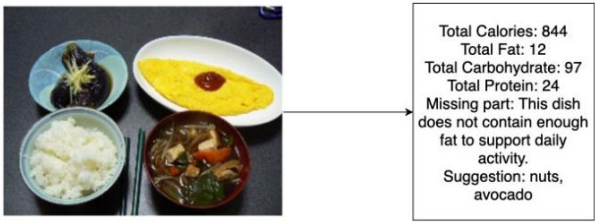


FIGURA 8. Ejemplo de resultado del sistema de evaluación dietética.

G. DISCUSIÓN

Aunque nuestro modelo propuesto funciona bien en diferentes tipos de conjuntos de datos, todavía hay margen de mejora en comparación con algunos de los modelos más modernos.

1) COMPLEJIDAD DEL MODELO

La complejidad del modelo es siempre un factor importante que afecta el rendimiento del modelo de aprendizaje profundo. Necesitamos hacer un equilibrio entre el tiempo de procesamiento y la precisión del sistema debido a las limitaciones del hardware. En los experimentos, utilizamos una red neuronal VGG-16 para la extracción de características y clasificación de elementos.

Debido a la arquitectura (nodos completamente conectados) de VGG con 16 capas, el proceso de capacitación podría llevar un tiempo considerable: varios días durante no más de 300 épocas para completar la sesión de capacitación. Es posible que probemos diferentes modelos en el futuro, como redes residuales profundas (ResNet) [28]. "ResNet tiene una complejidad temporal menor que VGG-16/19", se afirma en la presentación de Kaiming He [27]. Dado que ResNet está construido por varios bloques de construcción. Debido al uso de capas de agrupación promedio global en lugar de capas completamente conectadas, el tamaño del modelo se vuelve sustancialmente más pequeño. Por ejemplo, ResNet con 50 capas pesa 102 MB, mientras que nuestro modelo VGG-16 propuesto en la prueba pesa más de 533 MB.

2) RECOPIACIÓN DE DATOS

El segundo desafío es encontrar/generar un buen conjunto de datos que podamos utilizar para capturar imágenes de alimentos de las comidas diarias. Como problema que encontramos en la evaluación, aunque tenemos dos conjuntos de datos de imágenes populares, UEC-FOOD100 y UEC-FOOD256, la mayoría de las imágenes en dos conjuntos de datos son alimentos japoneses/Asain.

Mientras que el conjunto de datos FOOD101 tiene imágenes de comida de estilo occidental pero no tiene información del cuadro delimitador. Además, varios productos alimenticios pueden tener una variación intraclase alta o una variación interclase baja. Los artículos de la misma categoría que tienen una alta variación dentro de la clase pueden verse diferentes y los dos tipos diferentes de alimentos con

la baja variación entre clases puede tener una apariencia similar. Tanto la alta varianza intracase como la baja varianza entre clases pueden afectar significativamente la precisión de los modelos de detección.

Para abordar este problema, necesitamos buscar más conjuntos de datos como FOOD101 [11], para crear más conjuntos de datos FOOD20 con bbx.

En el futuro, continuaremos etiquetando nuestro conjunto de datos FOOD20-with-bbx para ampliar este conjunto de datos a una gama más amplia de categorías de alimentos. Es deseable la combinación con otros conjuntos de datos para crear un conjunto de datos sobre alimentos más diverso.

VI. CONCLUSIÓN Y TRABAJO FUTURO

En este artículo, exploramos el problema del reconocimiento de alimentos y la evaluación dietética aprovechando técnicas de aprendizaje profundo. En particular, para tener una mejor comprensión de la detección de objetos y el análisis nutricional, aplicamos el modelo Faster R-CNN de última generación para generar Rol y utilizamos una red neuronal profunda para extraer el mapa de características para el reconocimiento de alimentos. Analizamos la nutrición de los alimentos detectados y resumimos el informe de la comida con base en herramientas de evaluación dietética basadas en tecnología moderna. Llevamos a cabo extensos experimentos para evaluar la eficiencia y eficacia de nuestro sistema.

Los resultados muestran que nuestra solución propuesta logró un rendimiento comparable y tiene un gran potencial para promover una dieta saludable y un asesoramiento viable.

En el futuro, continuaremos nuestro trabajo para mejorar la precisión de nuestro sistema de detección y reducir el tiempo de procesamiento. Es deseable un esquema de análisis de alimentos más completo, como la predicción del peso. Además, para proporcionar una dieta saludable, el plan incluye una calculadora de dieta automática.

REFERENCIAS

- [1] Instituto de Ciencias Clínicas y Traslacionales de Colorado. (2014). Métodos de evaluación dietética más comunes. [En línea]. Disponible: <https://cctsi.cuanschutz.edu/> [2] MyFinessPal, 2018.
- [3] Samsung Salud, 2018.
- [4] Sistema de datos nutricionales para la investigación, 2019.
- [5] K. Aizawa, Y. Maruyama, H. Li y C. Morikawa, "Estimación del equilibrio alimentario mediante el uso de tendencias dietéticas personales en un registro de alimentos multimedia", *IEEE Trans. Multimedia*, vol. 15, núm. 8, págs. 2176–2185, diciembre de 2013.
- [6] D. Albanes, "Calorías totales, peso corporal e incidencia de tumores en ratones". *Cáncer Res.*, vol. 47, núm. 8, págs. 1887–1892, 1987.
- [7] S. Ao y CX Ling, "Adaptación de nuevas categorías para el reconocimiento de alimentos con representación profunda", en *Proc. IEEE Internacional. Conf. Taller de minería de datos (ICDMW)*, noviembre de 2015, págs. 1196–1203.
- [8] K. Ashton, "Esa cosa del 'Internet de las cosas'", *RFID J.*, vol. 22, núm. 7, págs. 97–114, 2009.
- [9] S. Belongie, J. Malik y J. Puzicha, "Contexto de forma: un nuevo descriptor para la coincidencia de formas y el reconocimiento de objetos", en *Proc. Adv. Inf. neuronal. Proceso. Syst.*, 2001, págs. 831–837.
- [10] Q. Bojia, Reconocimiento de alimentos y análisis nutricional utilizando CNN profundas. Montreal, QC, Canadá: McGill Univ., 2019.
- [11] L. Bossard, M. Guillaumin y L. Van Gool, "Food-101: minería de componentes discriminativos con bosques aleatorios", en *Proc. EUR. Conf. Computadora. Vis. Cham, Suiza: Springer*, 2014, págs. 446–461.
- [12] M. Abadi et al., "TensorFlow: un sistema para el aprendizaje automático a gran escala", en *Proc. 12° Simposio USENIX. Ópera. Sistema. Implemento de diseño. (OSDI)*, 2016, págs. 265–283.
- [13] L. Breiman, "Bosques aleatorios", *Mach. Aprende.*, vol. 45, núm. 1, págs. 5–32, 2001.
- [14] F. Chollet. (2015). Keras. [En línea]. Disponible: <https://keras.io> [15] Cuadro delimitador, CNblogs, 2019.
- [16] CM Hales, MD Carroll, CD Fryar y CL Ogden, "Prevalencia de la obesidad entre adultos y jóvenes: Estados Unidos, 2015–2016", *Centros de Prevención y Control de Enfermedades, Atlanta, GA, EE. UU.*, Tech. Rep. 288, 2016.
- [17] P. Dollar y CL Zitnick, "Detección rápida de bordes utilizando bosques estructurados", *IEEE Trans. Patrón Anal. Mach. Intel.*, vol. 37, núm. 8, págs. 1558–1570, agosto de 2014.
- [18] T. Ege y K. Yanai, "Estimación de calorías de alimentos basada en imágenes utilizando conocimientos sobre categorías de alimentos, ingredientes e instrucciones de cocción", en *Proc. Talleres temáticos Talleres temáticos multimedia ACM*, 2017, págs.
- [19] A. Eldridge, C. Piernas, A.-K. Illner, M. Gibney, M. Gurinović, JD Vries y J. Cade, "Evaluación de nuevas herramientas basadas en tecnología para la evaluación de la ingesta dietética: evaluación del grupo de trabajo de exposición e ingesta dietética de ILSI Europe", *Nutrients*, vol. 11, núm. 1, pág. 55, 2019.
- [20] S. Fang, Z. Shao, R. Mao, C. Fu, EJ Delp, F. Zhu, DA Kerr y CJ Boushey, "Estimación de porciones de alimentos de vista única: aprendizaje de imagen a energía Mapeos utilizando redes generativas adversarias", en *Proc. 25° IEEE Internacional. Conf. Proceso de imagen. (ICIP)*, octubre de 2018, págs. 251–255.
- [21] PF Felzenszwalb, "Representación y detección de formas deformables", *IEEE Trans. Patrón Anal. Mach. Intel.*, vol. 27, núm. 2, págs. 208–220, febrero de 2005.
- [22] EA Finkelstein, JG Trogdon, JW Cohen y W. Dietz, "Gasto médico anual atribuible a la obesidad: estimaciones específicas del pagador y del servicio", *Health Affairs*, vol. 28, núm. 5, págs. w822–w831, enero de 2009.
- [23] Z. Ge, C. McCool, C. Sanderson y P. Corke, "Modelado de características de redes neuronales convolucionales profundas locales para mejorar la clasificación de imágenes de grano fino", en *Proc. IEEE Internacional. Conf. Proceso de imagen. (ICIP)*, septiembre de 2015, págs. 4112–4116.
- [24] R. Girshick, "Fast R-CNN", en *Proc. IEEE Internacional. Conf. Computadora. Vis. (ICCV)*, diciembre de 2015, págs. 1440–1448.
- [25] R. Girshick, J. Donahue, T. Darrell y J. Malik, "Jerarquías de funciones ricas para una detección precisa de objetos y segmentación semántica", en *Proc. Conferencia IEEE. Computadora. Vis. Reconocimiento de patrones*, junio de 2014, págs. 580–587.
- [26] H. Hassannejad, G. Matrella, P. Ciampolini, I. De Munari, M. Mordonini y S. Cagnoni, "Reconocimiento de imágenes de alimentos mediante redes convolucionales muy profundas", en *Proc. 2do Int. Taller Asistencia Multimedia. Manejo dietético. (MADiMa)*, 2016, págs. 41–49.
- [27] K. He, X. Zhang, S. Ren y J. Sun, "Mapeos de identidad en redes residuales profundas", en *Proc. EUR. Conf. Computadora. Vis. Springer*, 2016, págs. 630–645.
- [28] K. He, X. Zhang, S. Ren y J. Sun, "Aprendizaje residual profundo para el reconocimiento de imágenes", en *Proc. Conferencia IEEE. Computadora. Vis. Reconocimiento de patrones. (CVPR)*, junio de 2016, págs. 770–778.
- [29] J. Hui. (2018). mAP (Precisión media media) para la detección de objetos. [En línea]. Disponible: <https://medium.com/@jonathanhui/map-mean-average-precision-for-object-detection45c121a31173>
- [30] T. Jiang, F. Jurie y C. Schmid, "Aprendiendo modelos previos de forma para objetos coincidencia", en *Proc. Conferencia IEEE. Computadora. Vis. Reconocimiento de patrones*, junio de 2009, págs. 848–855.
- [31] H. Kagaya, K. Aizawa y M. Ogawa, "Detección y reconocimiento de alimentos mediante una red neuronal convolucional", en *Proc. ACM Internacional. Conf. Multimedia (MM)*, 2014, págs. 1085–1088.
- [32] Y. Kawano y K. Yanai, "Expansión automática de un conjunto de datos de imágenes de alimentos aprovechando las categorías existentes con adaptación de dominio", en *Proc. Taller ECCV Transferencia Adaptación del conocimiento fuente. Computadora. Vis. (TASK-CV)*, 2014, págs. 3–17.
- [33] Y. Kawano y K. Yanai, "Sistema móvil de reconocimiento de alimentos en tiempo real", en *Proc. Conferencia IEEE. Computadora. Vis. Reconocimiento de patrones. Talleres*, junio de 2013, págs. 1–7.
- [34] Y. Kawano y K. Yanai, "FoodCam-256: Un sistema móvil de reconocimiento de alimentos en tiempo real a gran escala que emplea características de alta dimensión y compresión de pesos clasificadores", en *Proc. ACM Internacional. Conf. Multimedia (MM)*, 2014, págs. 761–762.
- [35] A. Krizhevsky, I. Sutskever y GE Hinton, "Clasificación de ImageNet con redes neuronales convolucionales profundas", en *Proc. Adv. Inf. neuronal. Proceso. Syst.*, 2012, págs. 1097–1105.
- [36] S. Lazebnik, C. Schmid y J. Ponce, "Más allá de las bolsas de características: coincidencia de pirámide espacial para reconocer categorías de escenas naturales", en *Proc. Computación IEEE. Soc. Conf. Computadora. Vis. Reconocimiento de patrones. (CVPR)*, vol. 2, junio de 2006, págs. 2169–2178.
- [37] DG Lowe, "Características distintivas de la imagen a partir de puntos clave invariantes de escala", *Int. J. Computación. Vis.*, vol. 60, núm. 2, págs. 91–110, noviembre de 2004.
- [38] Y. Matsuda, H. Hoashi y K. Yanai, "Reconocimiento de imágenes de múltiples alimentos mediante la detección de regiones candidatas", en *Proc. IEEE Internacional. Conf. Exposición multimedia*, julio de 2012, págs. 25–30.

- [39] A. Myers, N. Johnston, V. Rathod, A. Korattikara, A. Gorban, N. Silberman, S. Guadarrama, G. Papandreou, J. Huang y K. Murphy, "Im2Calories: Hacia un diario de alimentos automatizado con visión móvil", en *Proc. IEEE Internacional. Conf. Computadora. Vis. (ICCV)*, diciembre de 2015, págs. 1233-1241.
- [40] F. Moosmann, E. Nowak y F. Jurie, "Bosques de agrupación aleatoria para clasificación de imágenes", *IEEE Trans. Patrón Anal. Mach. Intel.*, vol. 30, núm. 9, págs. 1632-1646, septiembre de 2008.
- [41] Instituto Nacional de Salud (NIH). (2020). Recordatorio dietético de 24 horas (24HR): manual de evaluación dietética. [En línea]. Disponible: <https://dietassessmentprimer.cancer.gov/profiles/recall/>
- [42] Instituto Nacional de Salud (NIH). (2020). Cuestionario de frecuencia alimentaria: manual de evaluación dietética. [En línea]. Disponible: <https://dietassessmentprimer.cancer.gov/profiles/questionnaire/>
- [43] A. Neubeck y L. Van Gool, "Efficient non-maximum suppression", en *Proc. 18^o Int. Conf. Reconocimiento de patrones. (CIPR)*, 2006, págs. 850-855.
- [44] J. Noronha, E. Hysen, H. Zhang y KZ Gajos, "Platemate: Análisis nutricional de crowdsourcing a partir de fotografías de alimentos", en *Proc. 24^o año. Síntoma ACM. Software de interfaz de usuario. Technol.*, 2011, págs. 1-12.
- [45] Base de datos nacional de nutrientes de Usda, FNDDS, Departamento de Agricultura de Estados Unidos, Washington, DC, EE. UU., 2014.
- [46] A. Oliva y A. Torralba, "Modelado de la forma de la escena: una representación holística de la envolvente espacial", *Int. J. Computación. Vis.*, vol. 42, núm. 3, págs. 145-175, 2001.
- [47] Sobre peso y obesidad, Organización Mundial de la Salud, Ginebra, Suiza, 2020.
- [48] K. Poslusna, J. Ruprich, JHM de Vries, M. Jakubikova y P. van't Veer, "Declaración errónea de la ingesta de energía y micronutrientes estimada mediante registros de alimentos y recordatorios de 24 horas, métodos de control y ajuste en la práctica", *británico. J. Nutrición*, vol. 101, núm. T2, págs. S73-S85, julio de 2009.
- [49] P. Pouladzadeh, P. Kuhad, SVB Peddi, A. Yassine y S. Shirmoham-madi, "Medición móvil de calorías de alimentos basada en la nube", en *Proc. IEEE Internacional. Conf. Talleres de exposición multimedia (ICMEW)*, julio de 2014, págs. 1-6.
- [50] P. Pouladzadeh, P. Kuhad, SVB Peddi, A. Yassine y S. Shirmohammadi, "Medición de calorías alimentarias mediante una red neuronal de aprendizaje profundo", en *Proc. IEEE Internacional. Instrumento. Medidas. Tecnología. Conf.*, mayo de 2016, págs. 1-6.
- [51] S. Ren, K. He, R. Girshick y J. Sun, "Faster R-CNN: Hacia la detección de objetos en tiempo real con redes de propuesta de región", en *Proc. Adv. Inf. neuronal. Proceso. Syst.*, 2015, págs. 91-99.
- [52] J. Sánchez, F. Perronnin, T. Mensink y J. Verbeek, "Clasificación de imágenes con el vector de Fisher: teoría y práctica", *Int. J. Computación. Vis.*, vol. 105, núm. 3, págs. 222-245, diciembre de 2013.
- [53] W. Shimoda y K. Yanai, "Segmentación de imágenes de alimentos basada en CNN sin anotación de píxeles", en *Proc. En t. Conf. Imagen Anal. Proceso. Cham, Suiza: Springer*, 2015, págs. 449-457.
- [54] J. Shotton, M. Johnson y R. Cipolla, "Bosques de textos semánticos para categorización y segmentación de imágenes", en *Proc. Conferencia IEEE. Computadora. Vis. Reconocimiento de patrones*, junio de 2008, págs. 1-8.
- [55] K. Simonyan y A. Zisserman, "Redes convolucionales muy profundas para el reconocimiento de imágenes a gran escala", 2014, arXiv:1409.1556. [En línea]. Disponible: <http://arxiv.org/abs/1409.1556>
- [56] S. Singh, A. Gupta y AA Efros, "Descubrimiento no supervisado de parches discriminativos de nivel medio", en *Computer Vision—ECCV. Berlín, Alemania: Springer*, 2012, págs. 73-86.
- [57] N. Slimani, P. Ferrarì, M. Ocké, A. Welch, H. Boeing, M. van Liere, V. Pala, P. Amiano, A. Lagiou, I. Mattisson, C. Stripp, D. Engeset, R. Charrondière, M. Buzzard, W. van Staveren y E. Riboli, "Estandarización del método de calibración de recordatorio de dieta de 24 horas utilizado en la Investigación prospectiva europea sobre el cáncer y la nutrición (EPIC): conceptos generales y preliminares. resultados", *Eur. J. Clin. Nutrición*, vol. 54, núm. 12, págs. 900-917, diciembre de 2000.
- [58] J. Stoppelman. Gafido. [En línea]. Disponible: <https://www.yelp.com/>
- [59] S. Subramanian y A. Deaton, "La demanda de alimentos y calorías", *J. Political Economy*, vol. 104, núm. 1, págs. 133-162, febrero de 1996.
- [60] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke y A. Rabinovich, "Profundizando con las convoluciones", en *Proc. Conferencia IEEE. Computadora. Vis. Reconocimiento de patrones. (CVPR)*, junio de 2015, págs. 1-9.
- [61] JRR Uijlings, KEA van de Sande, T. Gevers y AWM Smeulders, "Búsqueda selectiva de reconocimiento de objetos", *Int. J. Computación. Vis.*, vol. 104, núm. 2, págs. 154-171, septiembre de 2013.
- [62] WC Willett, L. Sampson, MJ Stampfer, B. Rosner, C. Bain, J. Witschi, CH Hennekens y FE Speizer, "Reproducibilidad y validez de un cuestionario semicuantitativo de frecuencia alimentaria", *Amer. J. Epidemiol.*, vol. 122, núm. 1, págs. 51 a 65, 1985.
- [63] K. Yanai y Y. Kawano, "Reconocimiento de imágenes de alimentos utilizando una red convolucional profunda con entrenamiento previo y ajuste fino", en *Proc. IEEE Internacional. Conf. Talleres de exposición multimedia (ICMEW)*, junio de 2015, págs. 1-6.
- [64] S. Yang, M. Chen, D. Pomerleau y R. Sukthankar, "Reconocimiento de alimentos mediante estadísticas de características locales por pares", en *Proc. Computación IEEE. Soc. Conf. Computadora. Vis. Reconocimiento de patrones*, junio de 2010, págs. 2249-2256.
- [65] F. Zhu, M. Bosch, T. Schap, N. Khanna, DS Ebert, CJ Boushey y EJ Delp, "Clasificación de alimentos asistida por segmentación para la evaluación dietética", *Proc. SPIE*, vol. 7873, febrero de 2011, art. No. 78730B.



LANDU JIANG (Miembro, IEEE) recibió el B.Eng. Licenciado en ingeniería de seguridad de la información de la Universidad Jiao Tong de Shanghai, el M.Sc. Licenciatura en Gestión de la Construcción, el M.Sc. Licenciatura en Ciencias de la Computación de la Universidad de Nebraska-Lincoln y Ph.D. Licenciado por la Facultad de Ciencias de la Computación de la Universidad McGill. Actualmente ocupa el puesto postdoctoral Mitacs Accelerate Industrial en Aerial Technology y la Universidad McGill. Sus intereses de investigación incluyen visión por computadora, aprendizaje automático, detección inteligente, computación portátil y portátil, sistemas ciberfísicos, soluciones de energía verde y redes sociales en línea.



BOJIA QIU obtuvo su licenciatura de la Universidad de Zhejiang y su maestría de la Facultad de Ciencias de la Computación de la Universidad McGill. Sus intereses de investigación incluyen la aplicación de técnicas de aprendizaje profundo para tareas de detección y reconocimiento de objetos. También está interesado en optimizar la estructura de las redes neuronales y encontrar mejores hiperparámetros para mejorar el rendimiento del modelo profundo.



XUE LIU (Miembro, IEEE) recibió el Ph.D. Licenciado en Ciencias de la Computación por la Universidad de Illinois en Urbana-Champaign. Trabajó como profesor asociado presidido por Samuel R. Thompson en la Universidad de Nebraska-Lincoln y HP Labs, Palo Alto, CA, EE. UU. Actualmente es profesor titular en la Facultad de Ciencias de la Computación y William Dawson Scholar (profesor titular) en la Universidad McGill. También es vicepresidente de I+D, científico jefe y codirector del Samsung AI Centre-Montreal. También es profesor (nombramiento de cortesía) de

matemáticas y estadística en la Universidad McGill. También es miembro docente y miembro asociado del Instituto de Algoritmos de Aprendizaje de Montreal (MILA). Se desempeñó como científico jefe de Tinder Inc. Sus intereses de investigación incluyen redes informáticas y de comunicación, sistemas integrados y en tiempo real, sistemas distribuidos, sistemas ciberfísicos, informática ecológica y energías inteligentes. Ha publicado más de 280 artículos de investigación en importantes revistas académicas internacionales revisadas por pares y actas de congresos de gran reputación, incluidos ACM MobiCom, IEEE INFOCOM, ICNP, IEEE Security and Privacy (Oakland), IEEE RTSS, IEEE RTAS, ACM. /IEEE ICCPS, WWW, ACM UbiComp, ACM KDD, IEEE ICDE, etc. y varias TRANSACCIONES IEEE/ACM. Es o ha sido editor/editor asociado de IEEE/ACM TRANSACTIONS ON NETWORKING

(ToN), Transacciones ACM en Sistemas Ciberfísicos (TCPS), TRANSACCIONES IEEE EN TECNOLOGÍA VEHICULAR (TVT), ENCUESTAS Y TUTORIAL DE COMUNICACIONES IEEE (COMST) y TRANSACCIONES IEEE EN SISTEMAS PARALELOS Y DISTRIBUIDOS (TPDS). Formó parte de varios comités o paneles de revisión de subvenciones nacionales e internacionales.



CHENXI HUANG es actualmente profesor asistente en la Universidad de Xiamen. Sus intereses de investigación incluyen procesamiento de imágenes, reconstrucción de imágenes, fusión de datos, visualización tridimensional y aprendizaje automático, etc. Ha sido editor asociado del Journal of Medical Imaging y Informática de la Salud, desde 2019.



KUNHUI LIN es actualmente profesor con Universidad de Xiamen. Sus intereses de investigación incluyen sistemas de bases de datos, redes informáticas, aplicaciones multimedia, sistemas de información de inteligencia, etcétera.

...