OBJECTS RECOGNITION IN A GIVEN IMAGE USING ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

PRASAD H M
Assistant Professor, Department of Computer Science,
Vedavathi Government First Grade College,
Hiriyur, Chitradurga District.
Mail address: prasadworld8@gmail.com

Abstract:

Image recognition, also referred to as Computer Vision or object recognition, is a subfield of Machine Learning and Artificial Intelligence that deals with the ability of a computer system or model to identify and classify objects or features within digital images. The primary goal of image recognition is to teach machines to emulate the human visual system, allowing them to extract useful information from images or videos for various applications such as object detection, facial recognition, and autonomous vehicle navigation. Efficient and accurate object detection has been a important topic in advancement of computer vision systems. With the advent of deep learning techniques, the cure for object detection has increased drastically. The project aims to incorporate state-of-the-art technique for object detection with the goal of achieving high accuracy with a real-time performance. A major challenge in many of the object detection systems is the dependency on other computer vision techniques for helping the deep learning based approach, which leads to slow and non-optimal performance. In this project, we use a completely deep learning approach to solve the problem of object detection in an end-to-end fashion. The network is trained on the most challenging publicly available dataset (PASCAL VOC), on which a object detection challenge is conducted annually. The resulting system is fast and accurate, thus alling those applications which require object detection.

Keywords: **Objects recognition, image using, Artificial Intelligence and Machine Learning**

Introduction:

Amazon Rekognition makes it easy to add image and video analysis to your applications. You just provide an image or video to the Amazon Rekognition API, and the service can identify objects, people, text, scenes, and activities. It can detect any inappropriate content as well. Amazon Rekognition also provides highly accurate facial analysis, face comparison, and face search capabilities. You can detect, analyze, and compare faces for a wide variety of use cases, including user verification, cataloging, people counting, and public safety. Amazon Rekognition is based on the same proven, highly scalable, deep learning technology developed by Amazon's computer vision scientists to analyze billions of images and videos daily. It requires no machine learning expertise to use. Amazon Rekognition includes a simple, easy-to-use API that can quickly analyze any image or video file that's stored in Amazon S3. AWS Rekognition is a service that lets developers working with Amazon Web Services add image analysis to their applications. With AWS Rekognition your apps can detect, remember and recognize objects, scenes, and faces in images.

The human brain has a unique ability to immediately identify and differentiate items within a visual scene. Take, for example, the ease with which we can tell apart a photograph of a bear from a bicycle in the blink of an eye. When machines begin to replicate this capability, they approach ever closer to what we consider true artificial intelligence.

Computer vision aims to emulate human visual processing ability, and it's a field where we've seen considerable breakthrough that pushes the envelope. Today's machines can recognize diverse images, pinpoint objects and facial features, and even generate pictures of people who've never existed.

It's hard to believe, right? In this regard, image recognition technology opens the door to more complex discoveries. Let's explore the list of AI models along with other ML algorithms highlighting their capabilities and the various applications they're being used for.

Object detection is an important, yet challenging vision task. It is a critical part in many applications such as image search, image auto-annotation and scene understanding; however it is still an open problem due to the complexity of object classes and images. Current approaches to object detection can be categorized by top-down, bottom-up or combination of the two. Top-down approaches often include a training stage to obtain class-specific model features or to define object configurations. Hypotheses are found by matching models to the image features. Bottom-up approaches start from low-level or mid-level image features, i.e. edges or segments. These methods build up hypotheses from such features, extend them by construction rules and then evaluate by certain cost functions. The third category of approaches combining top-down and bottom-up methods have become prevalent because they take advantage of both aspects. Although top-down approaches can quickly drive attention to promising hypotheses, they are prone to produce many false positives when features are locally extracted and matched. Features within the same hypothesis may not be consistent with respect to low-level image segmentation. On the other hand, bottom-up approaches try to keep consistency in low level image segmentation, but usually need much more efforts in searching and grouping.

ML and AI models for image recognition

One can't agree less that people are flooding apps, social media, and websites with a deluge of image data. For example, over 50 billion images have been uploaded to Instagram since its launch. This explosion of digital content provides a treasure trove for all industries looking to improve and innovate their services.

Image recognition technology enables computers to pinpoint objects, individuals, landmarks, and other elements within pictures. This niche within computer vision specializes in detecting patterns and consistencies across visual data, interpreting pixel configurations in images to categorize them accordingly.

Given that this data is highly complex, it is translated into numerical and symbolic forms, ultimately informing decision-making processes. Every AI/ML model for image recognition is trained and converged, so the training accuracy needs to be guaranteed.

Convolutional neural networks (CNNs) in image recognition

CNNs are deep neural networks that process structured array data such as images. CNNs are designed to adaptively learn spatial hierarchies of features from input images.

During training, a CNN learns the values of the filters and weights through a backpropagation algorithm, adjusting them to recognize patterns and features in images, such as edges, textures, or object parts, which then contribute to recognizing the whole object within the image.

By stacking multiple convolutional, activation, and pooling layers, CNNs can learn a hierarchy of increasingly complex features.

Lower layers might learn to detect colors and edges, intermediate layers could learn to detect more complex structures like eyes or wheels, and deeper layers can detect high-level features like faces or entire objects, which is critical for image recognition tasks.

Overview of popular image recognition algorithms

A few image recognition algorithms are a cut above the rest. All of them refer to deep learning algorithms, however, their approach toward recognizing different classes of objects differs.

Faster Region-based CNN (Region-based Convolutional Neural Network) is the star in the R-CNN cluster considered as the best among machine learning models for image classification tasks.

It leverages a Region Proposal Network (RPN) to detect features together with a Fast RCNN representing a significant improvement compared to the previous image recognition models. Faster RCNN processes images of up to 200ms, while it takes 2 seconds for Fast RCNN. (The process time is highly dependent on the hardware used and the data complexity).

Single Shot Detector (SSD) divides the image into default bounding boxes as a grid over different aspect ratios. Then, it merges the feature maps received from processing the image at the different aspect ratios to handle objects of differing sizes. This approach makes SSDs very flexible and easy to train. With this AI model image can be processed within 125 ms depending on the hardware used and the data complexity.

Best image recognition models

Image recognition models use deep learning algorithms to interpret and classify visual data with precision, transforming how machines understand and interact with the visual world around us.

Let's look at the three most popular machine learning models for image classification and recognition.

- **Bag of Features Model**: BoF takes the image to be scanned and a sample photo of the object to be found as a reference. The model tries pixel-matching the features from the sample picture to various parts of the target image to identify any matches.

- **Viola-Jones Algorithm**: Viola-Jones scans faces and extracts features passed through a boosting classifier. As a result, a number of boosted classifiers are generated to check test images. A test image should generate a positive result from each classifier to find a match.

Transfer Learning

Transfer learning is a machine learning method where a model developed for a particular task is reused as the starting point for a model on a second task.

It is an effective technique, especially when the first task involves a large and complex dataset and the second task does not have as much data available. Here's how it works and why it is beneficial for image recognition:

- **Pre-trained Models**: In transfer learning, a model is initially trained on a large dataset with a wide range of images, like ImageNet, which has millions of images classified into thousands of categories. This initial model learns many features from the comprehensive dataset, making it a robust feature extractor for image data.

- **Feature Transfer**: When the pre-trained model is used for a new task, it can transfer the learned features (weights and biases) to the new task. Since the initial layers of CNNs often learn to recognize basic shapes and textures, while the later layers learn more specific details, the features in the initial layers are generally useful for many image recognition tasks.

- **Fine-Tuning**: The final layers of the model can be fine-tuned with a smaller dataset for a specific task. Fine-tuning involves re-training these layers to adjust the weights to better suit the specific task at hand, while the earlier layers may remain frozen on the weights learned from the original dataset.

Popular pre-trained models

MobileNet: Like many models, it was trained using the ImageNet dataset. Nonetheless, its design is tailored to be resource-efficient for mobile and embedded devices without significantly compromising accuracy. Its design renders it perfect for scenarios with computational limitations, such as image recognition on mobile devices, immediate object identification, and augmented reality experiences.

EfficientNet is a cutting-edge development in CNN designs that tackles the complexity of scaling models. It attains outstanding performance through a systematic scaling of model depth, width, and input resolution yet stays efficient.

Trained on the extensive ImageNet dataset, EfficientNet extracts potent features that lead to its superior capabilities. It is recognized for accuracy and efficiency in tasks like image categorization, object recognition, and semantic image segmentation.

Inception-v3, a member of the Inception series of CNN architectures, incorporates multiple inception modules with parallel convolutional layers with varying dimensions. Trained on the expansive ImageNet dataset, Inception-v3 has been thoroughly trained to identify complex visual patterns.

The architecture of Inception-v3 is designed to detect an array of feature scales, enabling it to perform various computer vision tasks, including but not limited to image recognition, object localization, and detailed image categorization.

Challenges in image recognition

The rapid progress in image recognition technology is attributed to deep learning, a field that has thrived due to the creation of extensive datasets, the innovation of neural network models, and the discovery of new tech opportunities.

Deep neural networks, engineered for various image recognition applications, have outperformed older approaches that relied on manually designed image features. Despite these achievements, deep learning in image recognition still faces many challenges that need to be addressed.

Model generalization

The major challenge lies in model training that adapts to real-world settings not previously seen. So far, a model is trained and assessed on a dataset that is randomly split into training and test sets, with both the test set and training set having the same data distribution.

Nevertheless, in real-world applications, the test images often come from data distributions that differ from those used in training. The exposure of current models to variations in the data distribution can be a severe deficiency in critical applications.

Scene understanding

Apart from data training, complex scene understanding is an important topic that requires further investigation. People are able to infer object-to-object relations, object attributes, 3D scene layouts, and build hierarchies besides recognizing and locating objects in a scene.

A wider understanding of scenes would foster further interaction, requiring additional knowledge beyond simple object identity and location. This task requires a cognitive understanding of the physical world, which represents a long way to reach this goal.

Modeling relationships

It is critically important to model the object's relationships and interactions in order to thoroughly understand a scene. Imagine two pictures with a man and a dog.

If one shows the person walking the dog and the other shows the dog barking at the person, what is shown in these images has an entirely different meaning. Thus, the underlying scene structure extracted through relational modeling can help to compensate when current deep learning methods falter due to limited data. Now, this issue is under research, and there is much room for exploration.

CONCLUSION: This work aims to incorporate stateof-the-art technique for object detection with the goal of achieving high accuracy with a real-time performance. A major challenge in many of the object detection systems is the dependency on other computer vision techniques for helping the deep learning based approach, which leads to slow and non-optimal performance.

REFERENCES:

1. Viola, P.A., Jones, M.J.: Rapid object detection using a boosted cascade of simple features. In: CVPR (2001)

2. Ferrari, V., Tuytelaars, T., Gool, L.J.V.: Object detection by contour segment networks. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3951, Springer, Heidelberg (2006)

3. Zhao, L., Davis, L.S.: Closely coupled object detection and segmentation. In: ICCV (2005)

4. Kokkinos, I., Maragos, P.Yuille, A.L.: Bottom-up & top-down object detection using primal sketch features and graphical models

5. Object Detection and Recognition in Digital Images: Theory and Practice by Boguslaw Cyganek in 2013.