The Rapid Rise of Deep Learning Computer Vision Technology

Last year, *ProgrammableWeb* published an article covering the rapid rise of computer vision technology and the increasing number of companies developing image recognition platforms. Until recently, computer vision technology has been used primarily for detecting and recognizing faces in photos. While facial recognition remains a popular use of this technology, there has been a rapid rise in the use of computer vision for automatic photo tagging and classification. This increase is largely due to recent advances in artificial intelligence (AI), specifically the use of convolutional neural networks (CNNs) to improve computer vision methods.

A recent IEEE Spectrum interview with Yann LeCun, director of AI research at Facebook, contains this explanation:

> The current excitement about AI stems, in great part, from groundbreaking advances involving what are known as "convolutional neural networks." This machine learning technique promises dramatic improvements in things like computer vision, speech recognition, and natural language processing. You probably have heard of it by its more layperson-friendly name: "Deep Learning."

AI is very hot right now, with major tech companies like Pinterest, Google and Microsoft doing amazing AI research. Pinterest is experimenting with deep learning algorithms to enhance product recommendations by automatically recognizing specific objects contained within the image of a pin. This new, experimental visual discovery and search technology will help provide relevant product recommendations for Pinterest users.

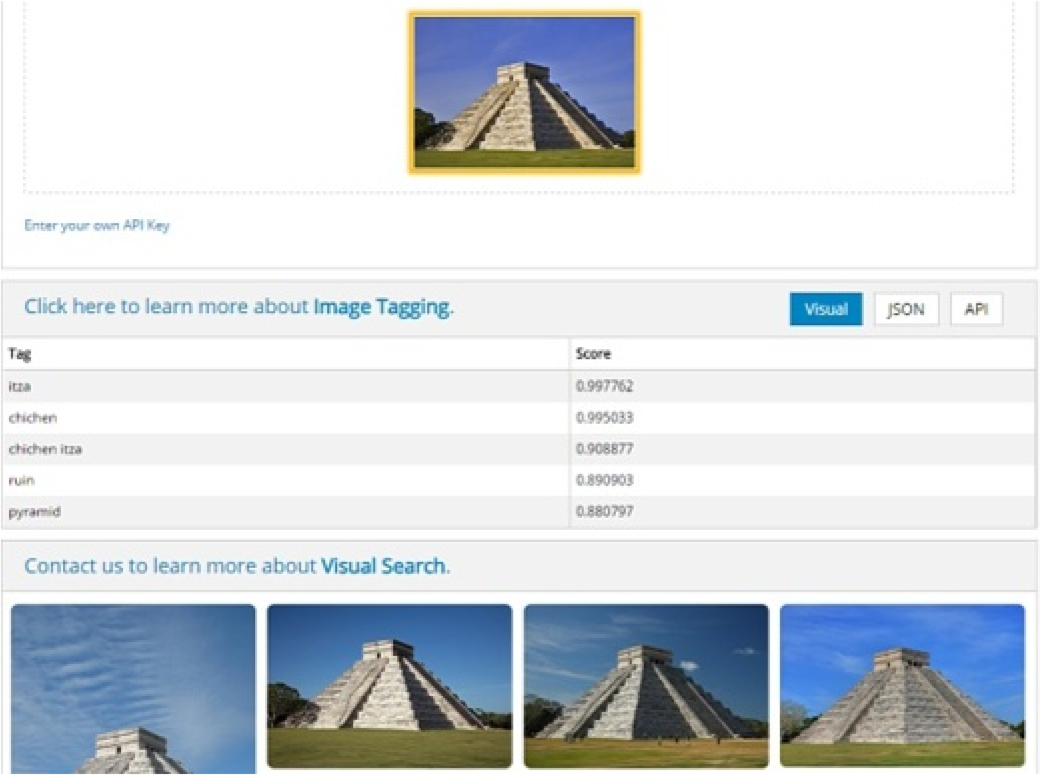
Google has been experimenting with CNNs for some time now, and the technology powers Android OS speech recognition and YouTube video recommendations. The recently announced Google Photos app uses deep learning-based computer vision to automatically recognize, classify and organize a user's photos. Google is also researching deep learning techniques to greatly improve object detection, classification and labeling. A recent Google research paper details the use of deep CNNs to automatically summarize a complex scene in a photo and generate captions that accurately describe the scene.

Microsoft Research has more than 1,000 scientists and engineers working on projects that involve machine learning, AI, computer vision, quantum computing, gaming and more. Examples of Microsoft Research projects include an approach for automatically generating image captions/descriptions; VC3 technology for keeping cloud-stored data safe; and the Traffic Prediction Project, which uses Bing traffic maps, road cameras, sensors and other data sources to predict traffic jams.

Major tech companies are not the only contributors to AI research. Academic institutions around the world are doing extensive research in AI and other areas of computer science. One particularly interesting research project, led by Babak Saleh and Ahmed Elgammal at Rutgers University, involves the use of computer vision algorithms to analyze digitized images of historical works of art. The algorithms are trained to detect and understand visual similarities in works of art as well as to automatically classify the paintings based on style, genre and artist.

This article highlights companies using deep learning-based computer vision technology and/or providing image recognition platforms featuring automatic photo tagging and classification. These companies were also chosen because they provide APIs and their websites feature live demos.

AlchemyAPI AlchemyVision



Enter your own API Key

Click here to learn more about Image Tagging.

Tag	Score
itza	0.997762
chichen	0.995033
chichen itza	0.908877
ruin	0.890903
pyramid	0.880797

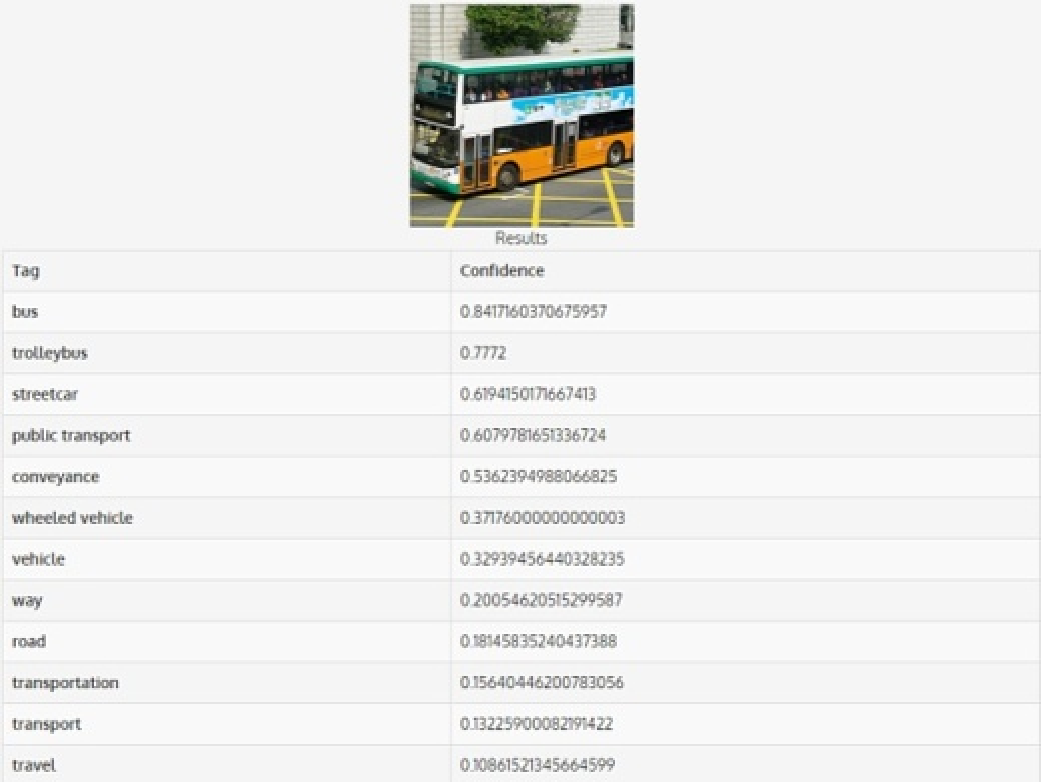
Contact us to learn more about Visual Search.

At the time of this writing, the AlchemyVision demo was the only one that returned a tag containing the formal name for this pyramid, "Chichen Itza." Try the demo. View API profile. Image credit: Flickr

AlchemyVision is an AlchemyAPI deep learning-based computer vision product that was launched in May 2014 and is capable of quickly analyzing, recognizing and understanding complex visual scenes. AlchemyVision features a set of APIs that includes a facial detection/recognition API, an image link extraction API and an image tagging API.

The image tagging API is capable of analyzing photos and understanding image context, summarizing scenes, identifying objects and returning up to 20 keywords with confidence scores. AlchemyVision is able to identify tens of thousands of objects in photos, including people, faces, animals, buildings, furniture and nature scenes.

AYLIEN Image Tagging API



Results

Tag	Confidence
bus	0.8417160370675957
trolleybus	0.7772
streetcar	0.6194150171667413
public transport	0.6079781651336724
conveyance	0.5362394988066825
wheeled vehicle	0.3717600000000003
vehicle	0.32939456440328235
way	0.20054620515299587
road	0.18145835240437388
transportation	0.15640446200783056
transport	0.13225900082191422
travel	0.10861521345664599

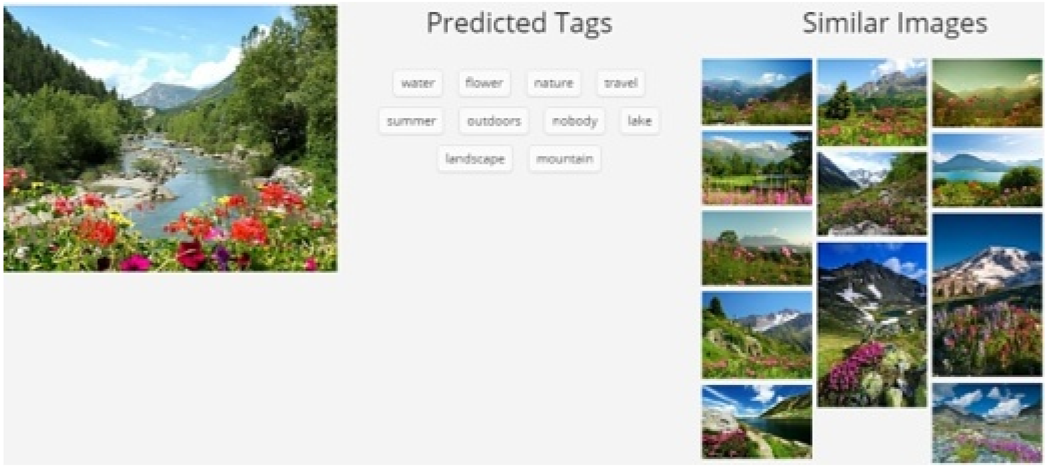
The AYLIEN demo returned dozens of tags along with confidence scores for many of the photos submitted. Try the demo. View API profile. Image credit: Flickr

AYLIEN offers a suite of APIs that provide capabilities such as sentiment analysis, classification, concept extraction and image tagging. Earlier this year, the company introduced a hybrid text and image analysis service through a partnership with Imagga. The API analyzes text and image content simultaneously in order to better understand the context of the document.

The recently released AYLIEN image tagging API leverages machine learning technology and deep learning algorithms to identify up to 6,000 objects, concepts and facial expressions. The API is capable of identifying people, faces, animals, food, buildings, vehicles, structures, and many other objects and concepts in photos.

Please note that the AYLIEN image tagging API is actually powered by Imagga (read the ProgrammableWeb article about the recent partnership formed by AYLIEN and Imagga.) There is also an image tagging demo on the Imagga website which demonstrates the capabilities of the Imagga API.

Clarifai Image and Video Recognition

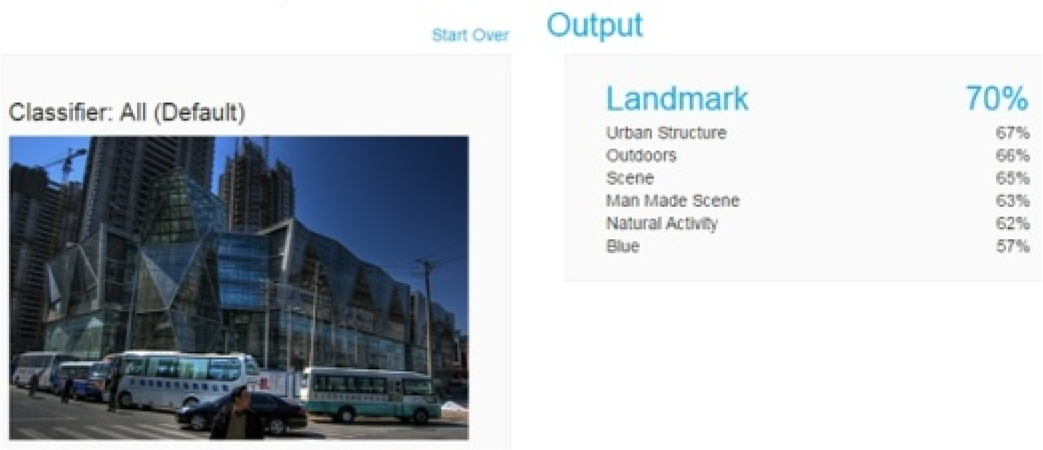


At the time of this writing, the Clarifai demo returned the greatest number of accurate tags for this nature photo. Try the demo. View API profile. Image credit: Flickr

Clarifai is a startup that provides deep learning-powered image and video recognition services. The platform utilizes CNNs that are able to learn complex visual concepts using massive amounts of data. Clarifai introduced the video recognition service at the beginning this year. According to the company announcement, the Clarifai system is capable of automatically analyzing video content recognizing 10,000 objects and concepts in real time.

The Clarifai API returns tags that classify objects contained in images and video along with probabilities/confidence scores. The API can also be used to search for and return similar images. The feedback endpoint allows end users to provide positive or negative feedback on the image tags returned by the API.

IBM Watson Visual Recognition Service



Classifier: All (Default)

Output

Landmark	70%
Urban Structure	67%
Outdoors	66%
Scene	65%
Man Made Scene	63%
Natural Activity	62%
Blue	57%

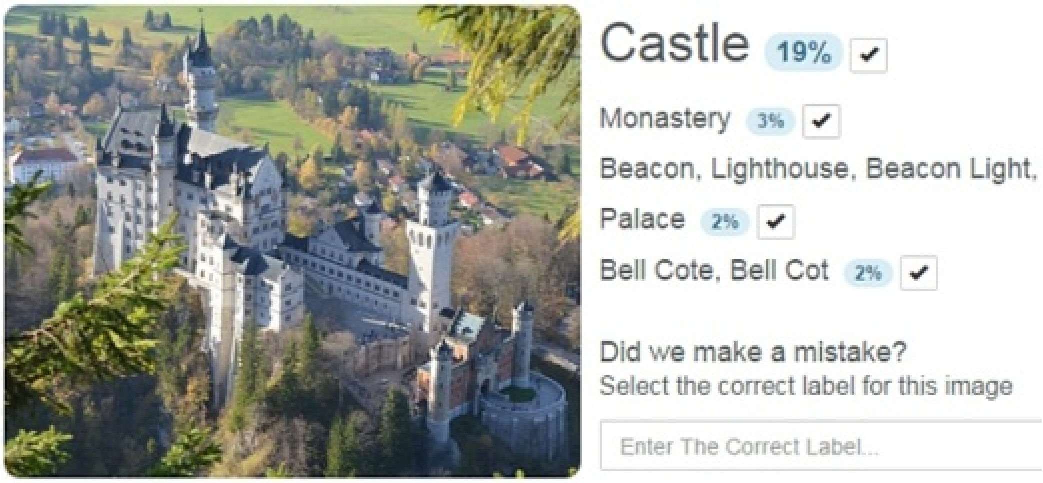
IBM's recent acquisition of AlchemyAPI should boost Watson's image tagging and classification capabilities. Try the demo. View API profile. Image credit: Flickr

Earlier this year, IBM introduced more than a dozen new services to Watson Developer Cloud, including speech to text, text to speech, concept insights, trade-off analytics and visual recognition.

The IBM Watson visual recognition service uses machine learning and visual models to analyze and understand the content of images and video frames. IBM Watson utilizes CNNs as semantic classifiers able to recognize settings, objects, events and other visual entities. At the time of this writing, there were more than 2,000 preset classifier and trained labels for categories such as animal, food, landmark, object, human, scene, sports and vehicle.

The IBM Watson Visual Recognition API allows images to be uploaded programmatically to the system, then analyzed against the labels specified. The API returns a list of tags/labels for the objects/concepts recognized within the image along with confidence scores.

MetaMind Vision



Castle 19% ☑

Monastery 3% ☑

Beacon, Lighthouse, Beacon Light,

Palace 2% ☑

Bell Cote, Bell Cot 2% ☑

Did we make a mistake?
Select the correct label for this image

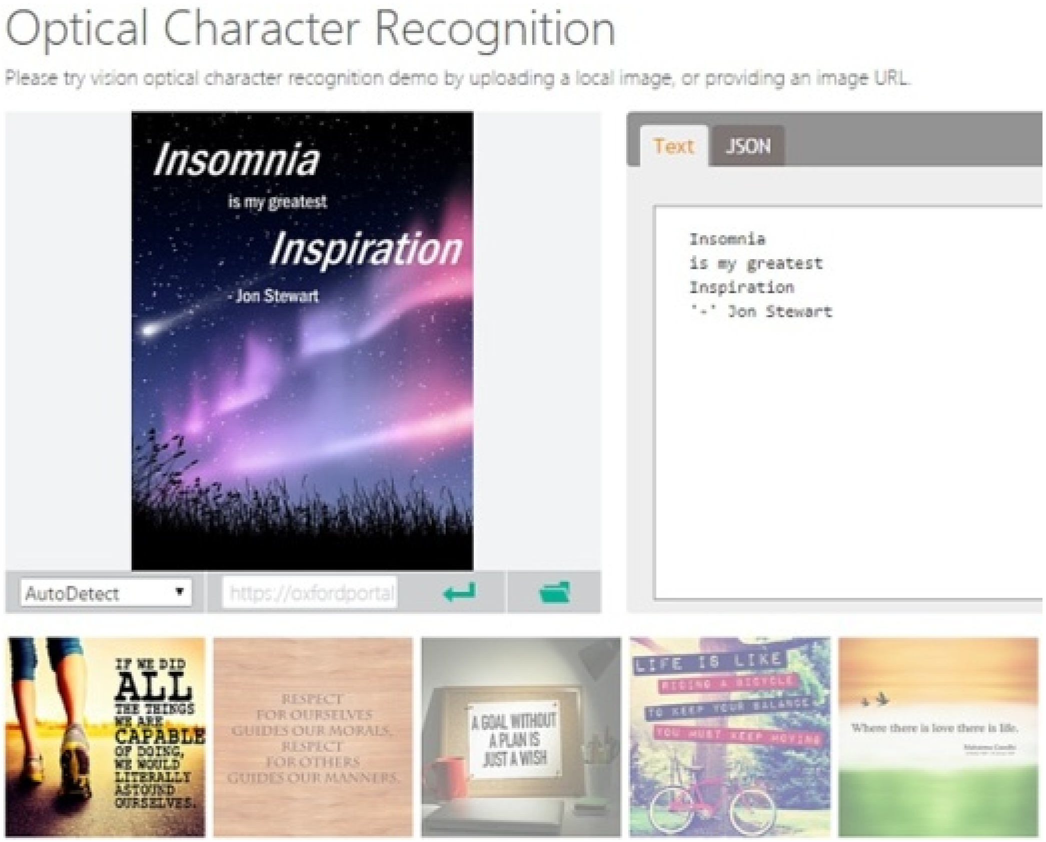
Enter The Correct Label...

At the time of this writing, MetaMind and AlchemyVision were the only demos that returned the tag "castle" for this photo of Neuschwanstein Castle. Try the General Image Classifier demo. Try the Food Classifier demo. View API profile. Image credit: Flickr

MetaMind was launched in December 2014 with the goal of providing artificial intelligence-as-a-platform that is easily accessible and easy to use. The company is focusing on the development of a new type of natural language processing, image understanding and knowledge base analytics platform. MetaMind utilizes groundbreaking technology called Recursive Deep Learning, which is based on deep learning research by the company's co-founder and CTO, Richard Socher.

The MetaMind API provides programmatic access to all the features of the platform, including image classification, text classification and semantic similarity. At the time of this writing, MetaMind Vision featured roughly 1,000 predefined classes and allowed users to create their own deep learning image classifiers.

Microsoft Project Oxford



Optical Character Recognition

Please try vision optical character recognition demo by uploading a local image, or providing an image URL.

```
Insomnia
is my greatest
Inspiration
'-' Jon Stewart
```

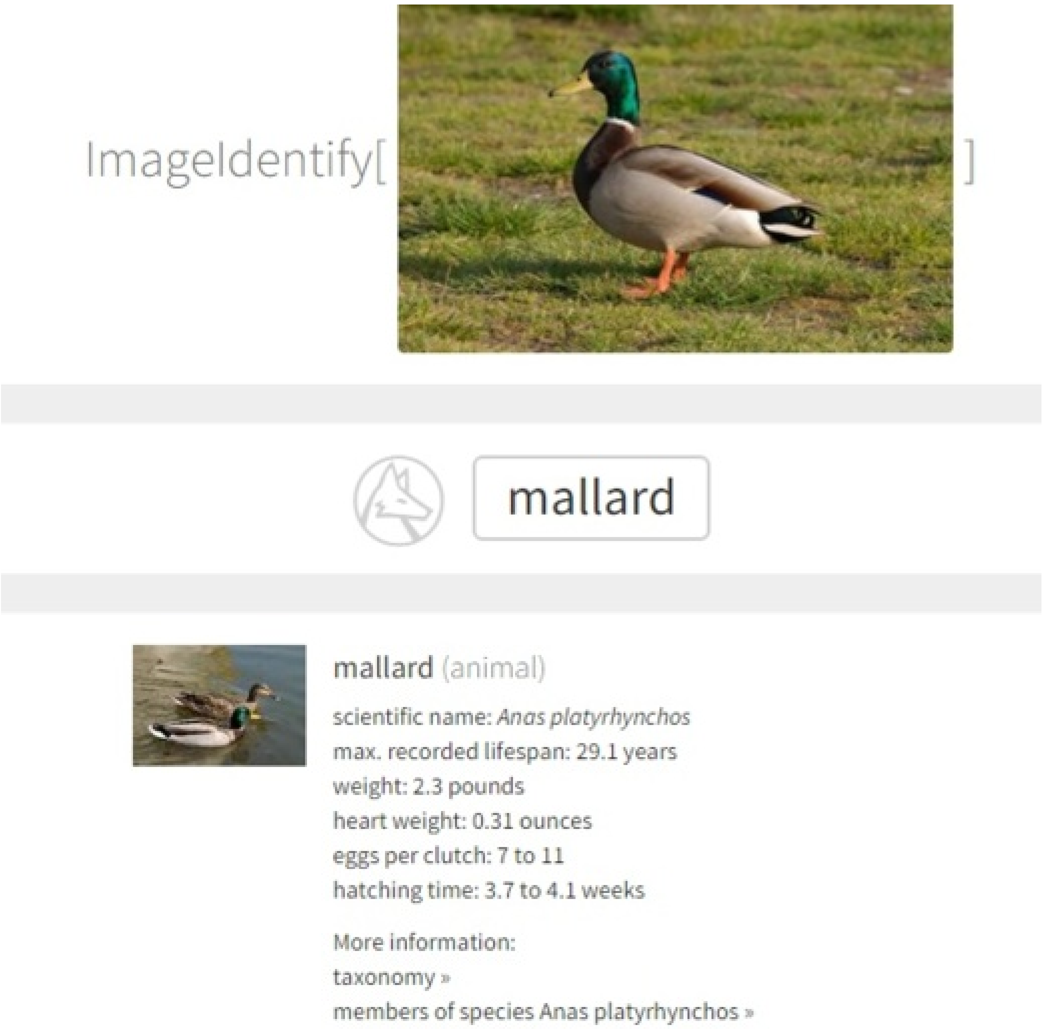
Insomnia
is my greatest
Inspiration
- Jon Stewart
Text
JSON
AutoDetect
https://oxfordportal
IF WE DID ALL THE THINGS WE ARE CAPABLE OF DOING, WE WOULD LITERALLY ASTOUND OURSELVES.
RESPECT FOR OURSELVES GUIDES OUR MORALS, RESPECT FOR OTHERS GUIDES OUR MANNERS.
A GOAL WITHOUT A PLAN IS JUST A WISH
LIFE IS LIKE RIDING A BICYCLE TO KEEP YOUR BALANCE YOU MUST KEEP MOVING
Where there is love there is life.

Project Oxford features optical character recognition that is able to recognize text in photos and extract the characters into a machine-usable format. Try the demos. View API profile.

A little over a month ago, Microsoft announced Project Oxford, a collection of APIs that developers can use to build applications that leverage Microsoft's natural language processing and/or computer vision technology. Project Oxford features computer vision APIs that use machine learning and image processing algorithms to analyze, recognize and understand visual content.

The image analysis API is capable of analyzing images and returning information such as image categories, predicted tags and dominant colors. The image thumbnail API generates high-quality thumbnails for inputted images and features smart cropping. The optical character recognition API is able to recognize text for a variety of languages and extracts the characters into a machine-usable format.

Wolfram Language Image Identification Project



ImageIdentify[ ]

mallard

mallard (animal)
scientific name: *Anas platyrhynchos*
max. recorded lifespan: 29.1 years
weight: 2.3 pounds
heart weight: 0.31 ounces
eggs per clutch: 7 to 11
hatching time: 3.7 to 4.1 weeks

More information:
taxonomy »
members of species Anas platyrhynchos »

The Wolfram ImageIdentify demo returns the top predicted tag for the photo along with a Wolfram|Alpha summary if the tag is for a proper name. Try the demo. View API profile. Image credit: Flickr

Wolfram launched the Image Identification Project just last month and announced that an ImageIdentify function is now built into the Wolfram Language. ImageIdentify uses deep learning-based computer vision technology to analyze, recognize and understand the content of an image. When the ImageIdentify function is applied to an image, it returns a symbolic entity that allows the Wolfram Language to perform additional computation. ImageIdentify can identify the content of images, provide probability scores, return definitions for objects and more.

The Wolfram Language Image Identification Project is an online application that showcases the ImageIdentify function and allows users to submit their own images to see how well ImageIdentify recognizes visual content. Developers who would like to access the ImageIdentify function via API can use the Wolfram Programming Cloud to create an ImageIdentify instant API. An instant API allows Wolfram Language code in the Wolfram Cloud to be called using a Web URL.

Conclusion

While the concept of convolutional neural networks has been around since the 1940s, it is only within the last few years that the use of CNNs has really taken off. CNNs are being used to significantly improve computer vision, speech recognition, natural language processing and other related technologies.

Today there are companies not only doing amazing research in the field of artificial intelligence, but also democratizing breakthroughs in AI. With so many advances in deep learning-based computer vision technology happening just within the last few years, it will be exciting to see what kind of breakthroughs are made in the field of computer vision in the not-too-distant future.