Fashion products classifier using Visual Recognition

Submitted by Nidhi Nigam

INTRODUCTION

Visual Recognition service uses deep learning algorithms to analyze images for scenes, objects, faces, and other content.

1.1 Overview

Visual Recognition features

1.**Classify virtually any visual content**

Visual Recognition understands the contents of images. Analyze images for scenes, objects, faces, colors, food, and other subjects that can give you insights into your visual content.

2.**Create your own classifiers**

It enables to create and train custom image classifiers using own image collections.

1.2 Purpose

How does Visual Recognition work?

A digital image represents a matrix of numerical values. These values represent the data associated with the pixel of the image. The intensity of the different pixels, averages to a single value, representing itself in a matrix format.

The information fed to the recognition systems is the intensities and the location of different pixels in the image. With the help of this information, the systems learn to map out a relationship or pattern in the subsequent images supplied to it as a part of the learning process.

After the completion of the training process, the system performance on test data is validated.

In order to improve the accuracy of the system to recognize images, intermittent weights to the neural networks are modified to improve the accuracy of the systems.

Some of the algorithms used in image recognition (Object Recognition, Face Recognition) are *SIFT (Scale-invariant Feature Transform), SURF (Speeded Up Robust Features), PCA (Principal Component Analysis), and LDA (Linear Discriminant Analysis)*.

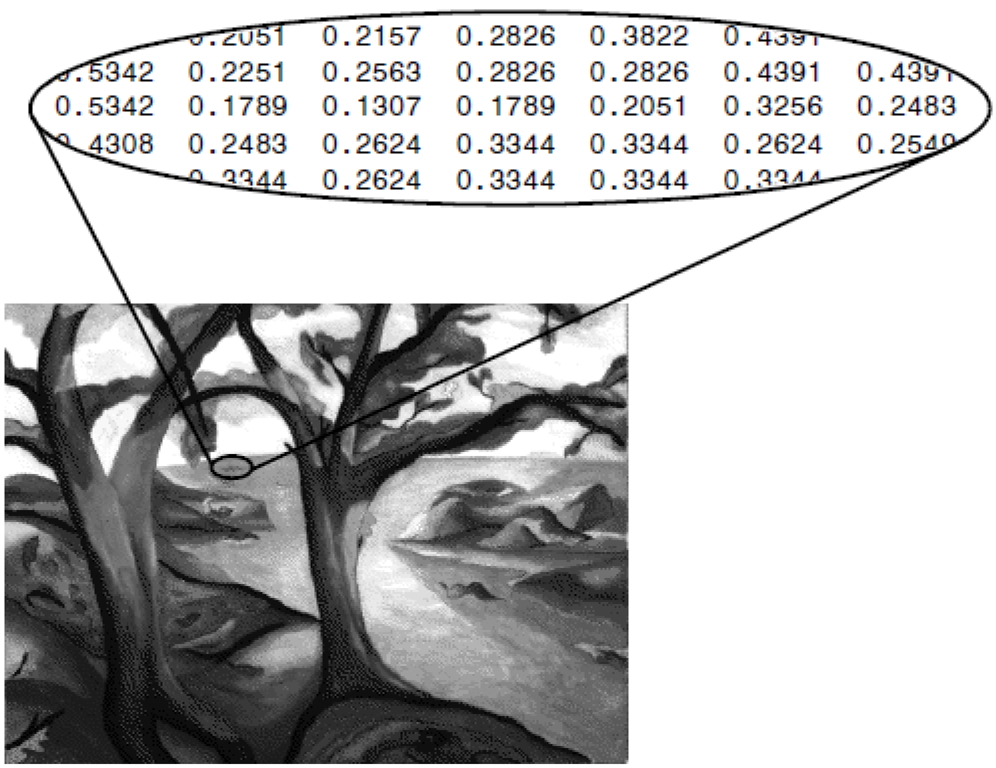


Figure 1: A Small Section Of An Image Represented In Matrix Format Note Values Of The Pixel Average To A Single Value

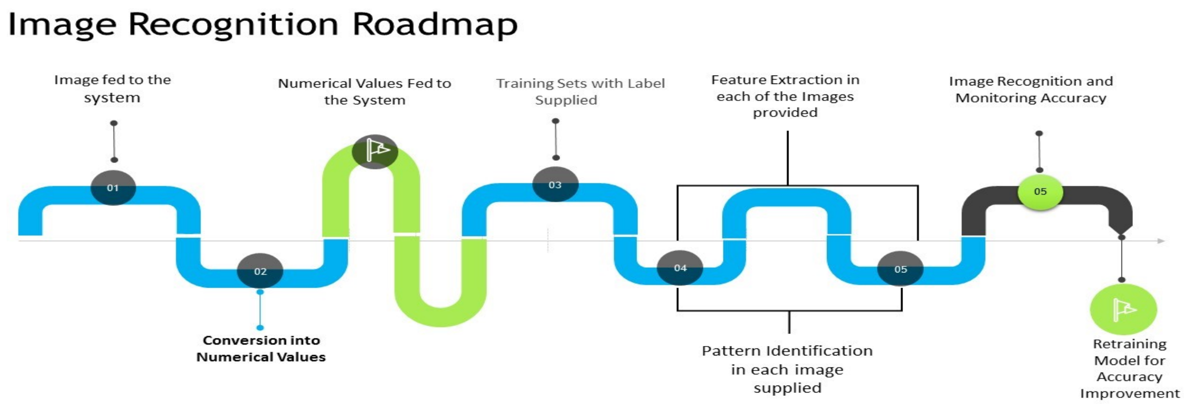


Figure 2 Path-Flow Of An Image

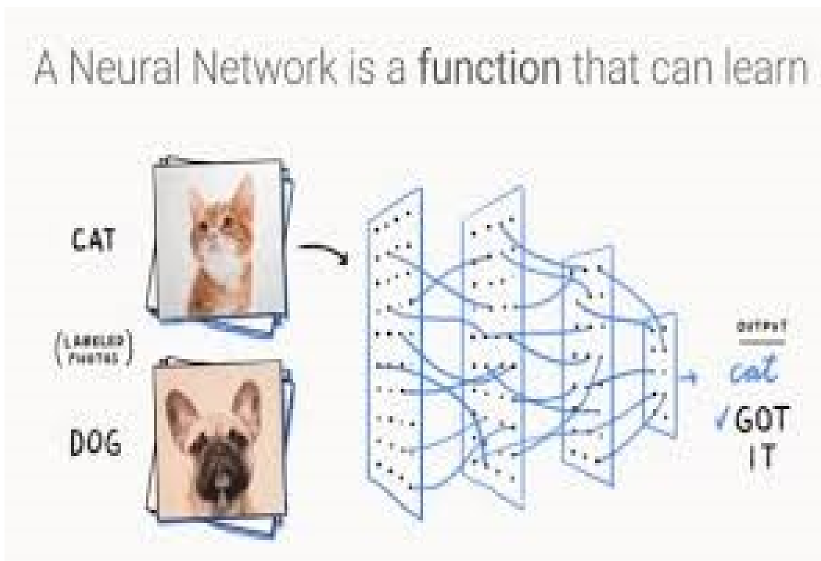


Figure 3: Output Of An Image: Cat. Input Image – Formulation Of Values – Identification Of Pattern – Training Of The Images- Test Data Validation

Challenges of Image Recognition

Viewpoint Variation: In a real world, the entities within the image are aligned in different directions and when such images are fed to the system, the system predicts inaccurate values. In short, the system fails to understand that changing the alignment of the image (left, right, bottom, top) will not make it different and that is why it creates challenges in image recognition.

Scale Variation: Variations in size affect the classification of the object. The closer you view the object the bigger it looks in size and vice-versa

Deformation: Objects do not change even if they are deformed. The system learns from the perfect image and forms a perception that a particular object can be in specific shape only. We know that in the real world, shape changes and as a result, there are inaccuracies when the system encounters a deformed image of an object.

Inter-class Variation: Certain object varies within the class. They can be of different shape, size, but still represents the same class. For example, buttons, chairs, bottles, bags come in different sizes and appearances.

-

LITERATURE SURVEY

Analytics India Magazine lists down the top 5 research papers in image classification

AlexNet (2012)

Dubbed as one of the milestones in deep learning, this research paper “ImageNet Classification with Deep Convolutional Neural Networks” started it all. Even though deep learning had been around since the 70s with AI heavyweights Geoff Hinton, Yann LeCun and Yoshua Bengio working on Convolutional Neural Networks, AlexNet brought deep learning into the mainstream. Authored by Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton, this 2012 paper won the ImageNet Large Scale Visual Recognition Challenge with a 15.4% error rate. In fact, 2012 marked the first year when a CNN was used to achieve a top 5 test error rate of 15.4% and the next best research paper achieved an error rate of 26.2. the paper was ground-breaking in its approach and brought the many concepts of deep learning into the mainstream.

GoogleNet (2015)

Inspired by the Inception thriller, GoogleNet proposes a deep convolutional neural network architecture codenamed “Inception”, which was responsible for setting the new state of the art for classification and detection in the ImageNet Large-Scale Visual Recognition Challenge. This Google project proposed a 22 layer convolutional neural network and was the winner of ILSVRC 2014 with an rate of 6.7%. according to experts, this CNN architecture was the first to propose a different approach from the general approach of simply stacking and pooling layers on top of each other.

Developed in response to index images, GoogleNet research project was undertaken by Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Drago Anguelov, Dumitru Erhan, Andrew Rabinovich and Christian Szegedy. At the core of the project was a reworked convolutional network architecture consisting of 100+ layers with a depth of 20 parameter layers) and is based Hebbian principle and scale invariance. Over the years, Google has been experimenting with neural networks to improve its image search ability and understand the content within Youtube videos.Google is leveraging these research advances and converting it into Google products such as in YouTube, image search and even self-driving cars.

ZFNet(2013)

This research paper authored by Matthew D Zeiler and Rob Fergus introduced a novel visualization technique that gave a peek into the functioning of intermediate feature layers and the operation of the classifier. This architecture was trained on 1.3 million images and it developed a visualization technique called De Deconvolutional Network that helped to examine different feature activations and their relation to the input space. The paper proposed to outperform Krizhevsky on the ImageNet classification benchmark.

Regularizing deep networks using efficient layerwise adversarial training (2017)

This research paper, authored by two University of Maryland researchers Rama Chellappa, Swami Sankaranarayanan and GE Global researchers Arpit Jain and Ser Nam Lim proposed a simple learning algorithm that leveraged perturbations of intermediate layer activation to provide a stronger regularization while improving the robustness of deep network to adversarial data. The research dealt with the behaviour of CNNs as related to adversarial data and the intrigue it had generated in computer vision. However, the effects of adversarial data on deeper networks had not been explored well.

The paper cited results of adversarial perturbations for hidden layer activations across different samples and leveraged this observation to devise an efficient adversarial training approach that could be used to train deep architectures.

Residual Attention Network for Image Classification (2017)

As the name implies, this latest research paper proposed a “Residual Attention Network” – a convolutional neural network that leverages attention mechanism which can incorporate feed forward network architecture in an end-to-end training fashion. Authored by Fei Wang, Mengqing Jiang, Chen Qian, Shuo Yang, Cheng Li, Honggang Zhang, Xiaogang Wang, Xiaoou Tang, their research methodology achieved a 0.6% top-1 accuracy improvement with 46% trunk depth and 69% forward FLOPs comparing to ResNet-200. The experiment also demonstrates that the neural network is robust vis-à-vis noisy labels.

2.2 Proposed solution

Prposed solution using IBM Watson Visual Recognition service to create a model and claasifying it on basis of provided data sets.The IBM Watson™ Visual Recognition service uses deep learning algorithms to analyze images for scenes, objects, and other content. The response includes keywords that provide information about the content.Proposed model can also take advantages of available models.

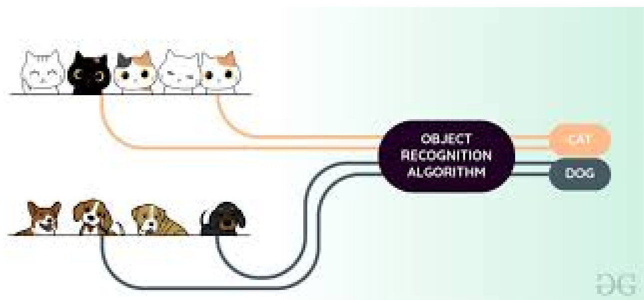
Available models

A set of built-in models provides highly accurate results without training:

- General model: Default classification from thousands of classes.
- Explicit model: Whether an image is inappropriate for general use.
- Food model: Specifically for images of food items.

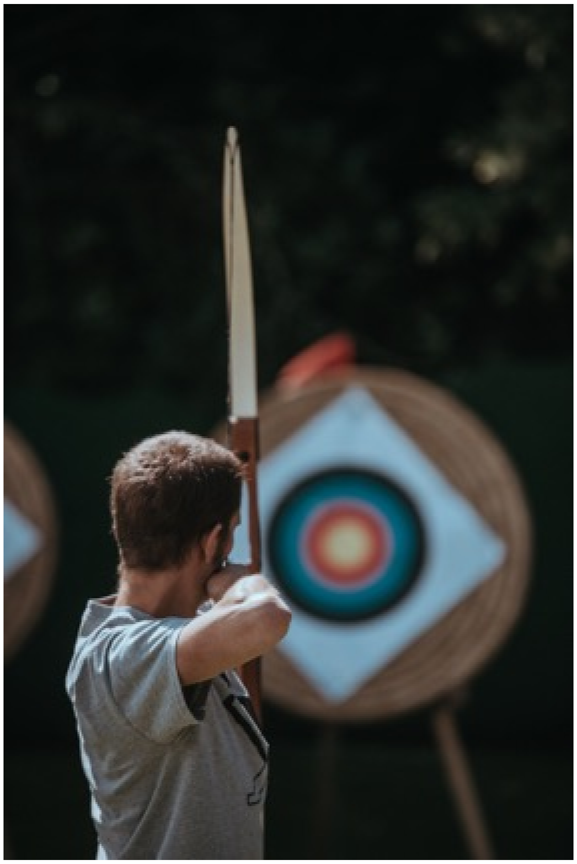
You can also train custom models to create specialized classes.

THEORITICAL ANALYSIS



Visual Recognition now uses a faster and more efficient deep learning network architecture for classification. The updated models also can differentiate more strongly between the top class and the rest of the classes. This approach might result in somewhat longer training times. The new architecture is used to train new custom models. When you retrain existing older models, the original architecture is used.

The following example shows the differentiation with the new architecture:



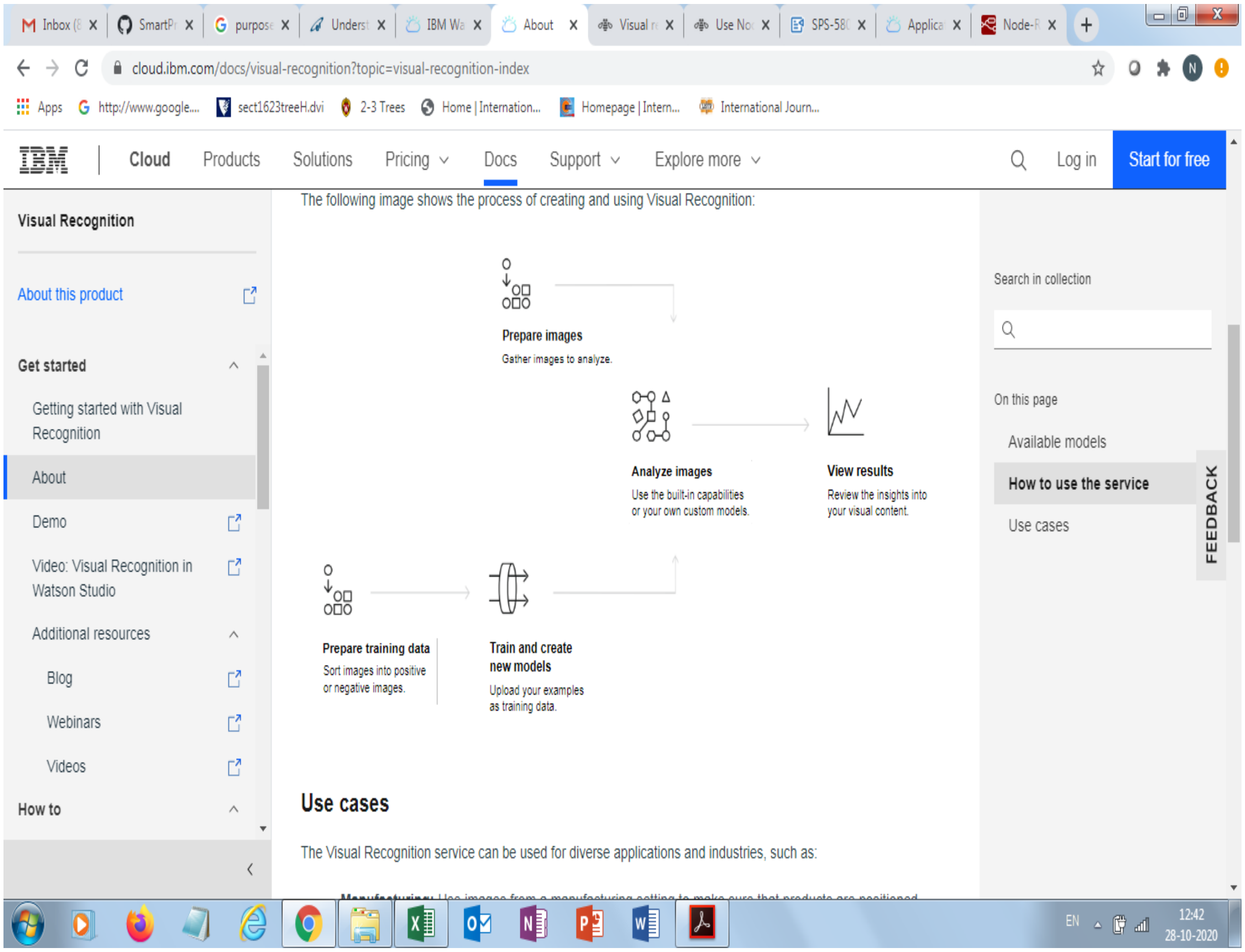
Original class and score	Updated class and score
Archery 0.99	archery 0.9
Auto Racing 0.996398	biking 0.004
Biking 0.0500174	fishing 0.001

Fishing 0.11029	golf 0.031
Golf 0.0980796	gymnastics 0.029
Gymnastics 0.964391	judo 0.021
Judo 0.339119	racing 0.002
Skating 0.0393602	skating 0.061
Skiing 0.0310527	skiing 0.003
Track and Field 0.208147	track 0.035

Experimental analysis

Experimental analysis results are shared via GitHub repo with all datasets.

Flowchart



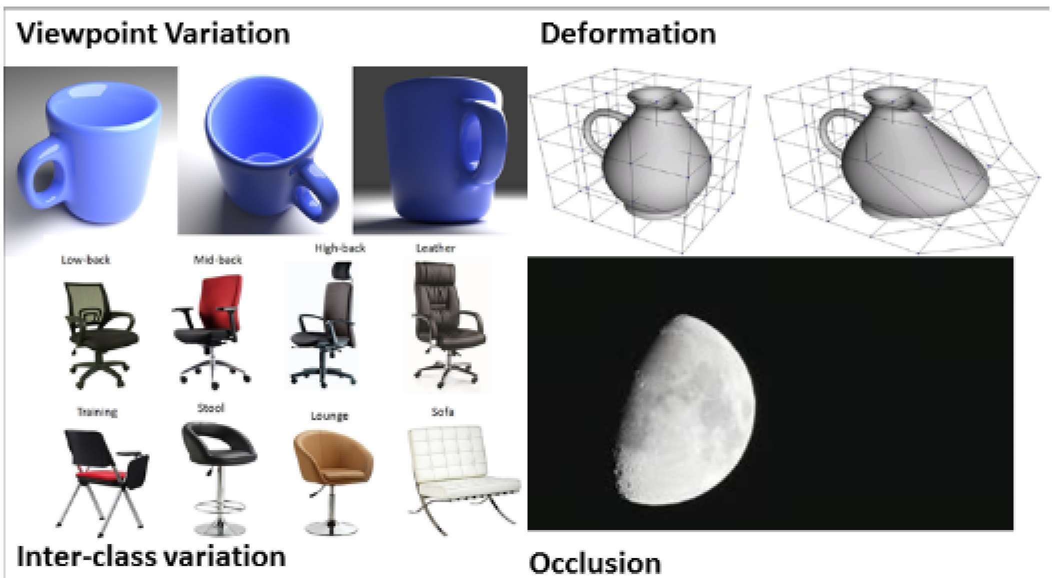
Advantages/Disadvantages

Limitations of Regular Neural Networks for Image Recognition

- The huge availability of data makes it difficult to process it due to the limited hardware availability.
- Difficulty in interpreting the model since the vague nature of the models

prohibits its application in a number of areas.

- Development takes longer time and hence, the flexibility is compromised with the development time. Although the availability of libraries like Keras makes the development simple, it lacks flexibility in its usage. Also, the Tensorflow provides more control, but it is complicated in nature and requires more time in development.



Applications

Uses of Image Recognition

- **Drones:**
  Drones equipped with image recognition capabilities can provide vision-based automatic monitoring, inspection, and control of the assets located in remote areas.

- **Manufacturing:**
  Inspecting production lines, evaluating critical points on a regular basis

within the premises. Monitoring the quality of the final products to reduce the defects. Assessing the condition of the workers can help manufacturing industries to have a complete control of different activities in the systems.

- **Autonomous Vehicles:**
  Autonomous vehicles with image recognition can identify activities on the road and take necessary actions. Mini robots can help logistics industries to locate and transfer the objects from one place to another. It also maintains the database of the product movement history to prevent the product from being misplaced or stolen.

- **Military Surveillance:**
  Detection of unusual activities in the border areas and automatic decision-making capabilities can help prevent infiltration and result in saving the lives of soldiers.

Forest Activities: Unmanned Aerial Vehicles can monitor the forest, predict changes that can result in forest fires, and prevent poaching. It can also provide a complete monitoring of the vast lands, which humans cannot access easily.

Conclusion

The IBM Watson Visual Recognition API is a powerful AI tool that identifies image content. The API comes with pretrained models that can accurately detect objects, scenes, colors, and foods, which can facilitate fast adoption and implementation. But the real power comes from the ability to train Watson to recognize custom classes.

With the power of custom classifiers, users helped California save water, performed infrastructure inspections with drones, and automated quality control inspections for automotive assembly lines. To get the most out of the Visual Recognition API, there are a number of techniques and optimizations that will help to maximize accuracy.