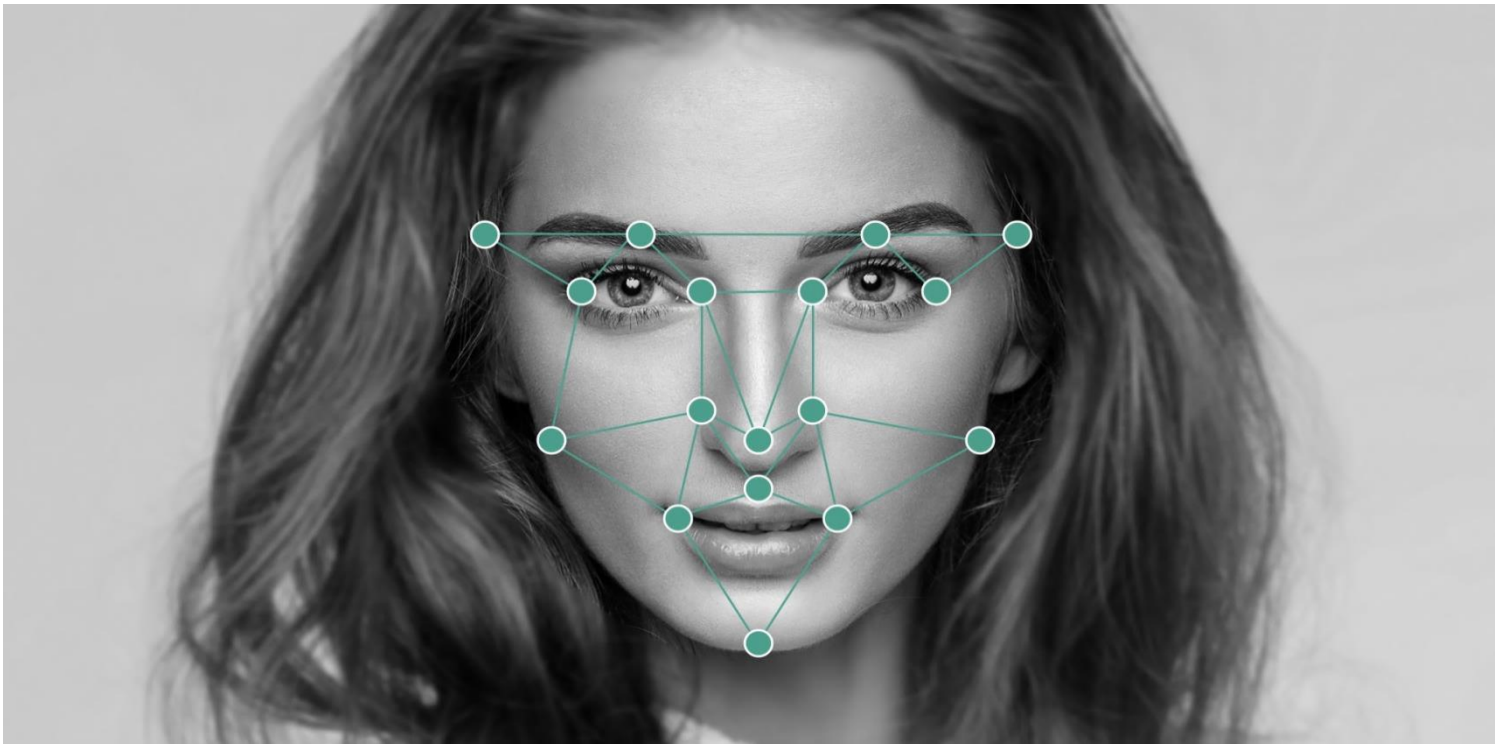

applexus
innovate to transform

AI-enabled Facial Recognition for Retail

White Paper

Introduction

With Forbes magazine and USA TODAY's recent stories about the alleged misuse of facial recognition by Facebook and Cambridge Analytica, Facial Recognition is a hot topic now. Rather than discussing the social, legal or customer opt-in/opt-out considerations of facial recognition, we will provide a foundation and go through the models and processes in this blog. This article gives you a high-level guidance around the technology, the processes involved and its practical applications in a retail or consumer-focused business.

AI-enabled facial recognition is vital for the retail industry as it can help them understand their customers and deliver better experiences. Recent advances in AI have significantly improved accuracy. This paper helps you understand the design of a deep learning based facial recognition system.

AI, Machine Learning and Deep Learning

Deep learning is an AI technology that has recently created breakthroughs in the field of computer vision speech and natural language processing. Before we go into details, let us understand what machine learning is. Machine Learning gives computers the ability to learn without being explicitly programmed by learning from data or experience to make predictions on unseen data. A Machine Learning model produces an output for a set of inputs, which is then compared with the desired output. Any non-conformance or error is fed back to make the output closer to the desired output or target.

Machine Learning is used in applications where there is no empirical relationship between the inputs and output. The objective of AI is to build machines that possess the same level of human intelligence though it still remains unachievable. Deep learning is based on the artificial neural network, which is a type of machine learning technique inspired from the structure and function of the human brain. It uses a large number of processing layers and large data sets as inputs to improve the model or prediction accuracy.

Deep Learning concept is not new and data scientists have strived hard to enhance prediction accuracy by adding more layers. With the confluence of several factors including availability of large data sets, powerful hardware – for example GPUs, and better algorithms and architectures, model accuracies have improved over time.

Facial Recognition Approaches

The problems in facial authentication have been studied in detail and researchers have proposed many methods to improve its accuracy rate.

Classical Approach

It involves handpicking features using domain knowledge of the data to create features, which are then classified using machine learning Algorithm. The approach works well for small data sets but fails for larger ones. Additionally, they are not effective on variations in pose, illumination or occlusions.

Modern Approach

In this approach, the neural network will find features itself. This works on large data sets and is invariant to pose, illuminations, occlusions, etc. Facebook's Deep Face and Google's Face Net use this approach. Face detection, landmarks detection and face alignment form the stages in pre-processing step. In the face recognition phase, we use the pre-processed images to identify a subject's face correctly. In the face detection stage, the system detects whether there is a face in the image or not and if there is, the facial landmarks are plotted and face alignment is carried out. The system then applies deep learning techniques to recognize the person.

Face Detection Using Histogram of Oriented Gradients (HOG)

The Histogram of Oriented Gradient uses visual attributes of the content in images, videos or applications to process images and detect faces. It spots image gradient or intensity change in localized portions of the image to extract features about the edges and shapes. HOG features are classified with a Support Vector Machine classifier for face detection.
After the system extracts the face from the bigger image, the images are aligned using landmark detection. This is then compared with mean landmarks on the reference image and aligned correctly using affine transformation. This is a linear mapping method that preserves point straight lines and planes without causing any distortion. The image thus created by affine transformation is used for facial recognition using deep learning.

There are basically two steps involved in deep learning:

1. Facial Learning

Consider a database with 1 million images of 1k users. The neural network having a deep learning architecture uses images to extract the image-specific features and labels. These features are then stored as embedded vectors, representing the face of each user.

2. Facial Matching

When a new input image is fed to the system, it extracts features from this image and compares it with a learned feature vector to perform a similarity measurement, which may be measured by Siamese, Cosine or Euclidian methods. The output decides whether there is a match or mismatch. Convolutional Neural Network (CNN) is the most widely used deep learning architecture in computer vision.

The reason:

- Rugged to shifts and distortions in the image
- Requires smaller memory as the same filter coefficients are used across different locations in the space
- Invariant to different poses, partial obstructions, horizontal or vertical shift
- Proven to work well in vision, speech and natural language Processing.

It is made of a convolutional layer, a non-linear activation function layer, pooling layer and fully connected layer. The function of the pooling layer is to reduce the spatial dimension of the image and the output is a fully connected neural network.

Convolutional Neural Network Simplified

The objective of the neural network is to adjust the parameters to make the training sample closer to the desired result. We define the parameters in terms of cost functions, the errors of which needs to be minimized as far as possible.

The filter parameters in the convolution layer and the synaptic weights in the neural network layer are the commonly adjusted parameters to minimize cost function. Stochastic Gradient Descent (SGD) based learning, popularly used for training CNN, enables faster training, gives better prediction accuracy compared to traditional methods and is more efficient on large datasets. Let's demystify the working of CNN with an example. Consider we input 5 × 5 image, which is convolved with a 3 × 3 filter matrix. We get a convolved feature map from the dot product (element-wise multiplication with matrices) of chunks of input image and filter image.

In simple terms, assume we are looking at an object through a smaller window. Just as we get different perspectives of an object when you move that window in different directions, you get feature maps or a combination of features when you slide a filter image over an input image.

ReLU Activation Function

As more of the real-world data is non-linear in nature, ReLu, also known as a rectified linear unit, introduces non-linearity in CNN. It selectively activates neuron by returning zero for negative pixel values in the input image. It also returns output value which is equal in intensity to the input value if it is greater than zero. Thus the rectified filter image has only non-negative values as shown in the figure below.

There are different ways of pooling, which summarizes the features in the feature map.

- **Average Pooling:** The input is divided into smaller portions and the average of the slice or full values are computed
- **Max Pooling:** In this layer, the spatial size is progressively reduced along with the parameters and the computational steps in the deep learning architecture

The abstracted form of the representation is achieved by dividing the input into smaller pooling regions and taking the maximum value in each region. In the example below, if we take 5, 11, 0 & 4, the output element contains the maximum of the 2×2 matrix, i.e. 11.
In the same manner, if we take a real image and pass it through a filter we get a convolved output. This is then passed through a rectified linear unit and pooling is performed over each map to get an output image.

Neural Network in Facial Recognition

Deep architecture is formed by stacking together a number of CNN building blocks. Deep learning procedure involves initializing the filters in the convolution randomly and automatically learning the most important parameters by the network.

Through back propagation using SVD, the network is trained end-to-end for all the global or local parameters to recognize a subject's face correctly. There is a natural progression from low level to the high-level structure as it passes through the different convolution layers.

As we go more in-depth to other convolution layers, the filters carry out dot product with the input of the previous convolution layers for classifying pixels to edges. Thus, the deep learning model performs hierarchical learning to combine the multistage outputs for accomplishing edge detection better. The deep learning architecture represents the face as a feature vector in an N × N matrix.

Integrating a Facial Recognition Model with a Client ling Software

Success of retail stores depend on how rapidly retailers respond to their customer's needs, since they are the new market-makers. To win in this customer age, retailers need to move from their traditional retailing software and adopt Clienteling software coupled with face recognition. This will help them identify their premium customers quickly, transform from an information source to points of engagement, and deliver the right product to their customers with the customized shopping experience.

Facial Recognition Cases/Instances:

- Clienteling: one-to-one personalized shopping experience
- In-store traffic analytics
- Dwell time at an aisle
- Visualize customer path in-store
- Emotion recognition at point of sale
- Order online and pick from store
- Payment and check out through face verification

About Applexus

Applexus Technologies (Applexus) is the global technology leader offering business consulting and SAP services to transform customers through digital innovation. We specialize in advisory, migration, implementation, and management of SAP S/4HANA and BW/4HANA solutions. Applexus delivers transformational business solutions for a marquee list of clients spanning retail, fashion, and consumer products industries. Applexus operates out of centers in North America, the United Kingdom, the Middle East, and India. For more information, visit us online at www.applexus.com.

Sources:

www.sap.com
blogs.sap.com
www.applexus.com
www.alliedmarketresearch.com
www.gartner.com
www.thebalance.com
www.nrf.com
www.statista.com
www.mckinsey.com