# Project Title Page

# Table of Contents

# 1. Introduction

A very compelling type artificial intelligence is that of the Computer Vision. Within the field of Artificial intelligence, Computer Vision is a sub field which deals with the identification, recognition and derivation of meaningful information from digital images and video data by teaching models and neural networks to accomplish this specific goals (IBM, n.d.).

The human eye over the years with training would be able to tell objects apart by their distinct characteristics and features. It is able to tell the distance from an object or whether an object is moving or still.

Computer vision with a lot of data are trained to do the same thing a human eye would be capable of doing with greater precision and in faster time with image data from cameras and algorithms to surpass that of the human capabilities. This systems have wide area of usage in the industries i.e manufacturing to the automotive industry and growing - the market is valued to be at about USD 20.31 billion expected to grow to about USD 175.72 billion by 2032 (Fortune Business Insights, n.d.).

# 2. Computer Vision

Computer Vision is a specialised field under Artificial intelligence that focuses on enabling machine to understand the visual world around them through the use of digital image and video data from cameras and deep learning. (SAS Institute, n.d.)

It focuses on replicating the complex function of recognition  of the human eye in machines allowing them to produce information from processing visual data (images and videos).

# 3. Architecture

A great deal of computer vision is identifying and understanding patterns from visual data, but before this can be archived the model would have to be trained just like a human would have to be trained to distinguish between different visual data. This training would require a lot of data which could already have been labeled (DataCamp, 2020).

To understand this image data being passed, the data would have to be analyzed because at its lowest level, computers only understand 1's and 0'. Convolutional neural networks help ML models see by fractionating images into pixels. Each pixel is given a label or tag to produce a matrix of this labels that the system can now interpret.

These labels are then collectively used to carry out convolutions, a mathematical process that combines two functions to produce a third function. Through this process, convolutional neural networks can process visual inputs (Ashtari, 2022).

The model would have to run analysis of data repeatedly till it can recognise patterns and distinctions between them to match the label as accurately as possible.

Typically Computer Vision would be handled in the following steps:

1. Acquiring training data (Images and Videos) (SAS Institute, n.d.).
2. Processing the data (SAS Institute, n.d.).
3. Understanding the data processed to acquire meaningfully information (SAS Institute, n.d.).

## 3.1. Architecture of Computer Vision Systems

Computer vision systems are rather complicated constructs that employ various elements and approaches to analyze visual data from the environment. In this section, the architecture of a conventional computer vision system is presented in order to describe the basic blocks that compose CV systems and their functions.

### *3.1.1. Image Acquisition :*

The first component of any vision system is the acquisition of images or videos. This involves:

Sensors and Cameras: Devices that records unprocessed video information about the surroundings. These vary from the general RGB cameras to depth cameras, thermal cameras along with LiDAR.

Image Preprocessing: Methods of improving the images captured using the designed system. This may include noise attainment, contrast enhancement and normalization (Gonzalez and Woods; 2018).

### *3.1.2. Image Processing:*

After acquiring the raw data, the system performs several preprocessing steps to prepare the data for further analysis:

Image Resizing and Cropping: Image resampling and resizing to fit the subsequent processing algorithms' input requirements.

Color Space Conversion: A common technique of converting images from one color space to another ( for instance from RGB to gray scale or HSV) for ease of processing.

Filtering and Edge Detection: To achieve the results implementing the filters to emphasize some peculiarities which maybe important, for example edges, contours, and textures. Sobel filters, Canny edge detection, or Gaussian blur are widely used (Szeliski, 2022).

### *3.1.3. Feature Extraction*

Feature extraction is crucial for identifying and describing important aspects of the image that will be used for recognition and classification:

Keypoint Detection: Localization of the points of interest in an image using SIFT (Scale-Invariant Feature Transform), SURF (Speeded-Up Robust Features) and ORB (Oriented FAST and Rotated BRIEF), etc. , (Lowe, 2004).

Descriptors: Calculating feature descriptors to characterize the image content in the small neighborhood of each keypoint. These are used for matching and classification as pointed out by Lowe (2004).

### *3.1.4. Model Architecture*

The core of any computer vision system is the model that processes the extracted features and performs tasks such as classification, and detection:

*Convolutional Neural Networks* (CNNs): A type of artificial neural network which is specially designed for solving problems of data representation and structure particularly image recognition systems. CNN structures are made up of several layers to comprise the convolutional layers, pooling layers, and fully connected layers (Goodfellow, Bengio, & Courville, 2016).

*Convolutional Layers* : The input image should be filtered in a way to find out the edges and texture in the provided image at propose by (LeCun, Bengio & Hinton, 2015).

*Pooling Layers* : Downsample the feature maps to decrease the dimensionality while preserving critical data; this is usually done using max pooling or average pooling (Goodfellow et al. , 2016).

*Recurrent Neural Networks* (RNNs): For the applications requiring the temporal characteristics: processing the sequential data, video analysis, etc. LSTM and GRU are the frequently used types (Hochreiter & Schmidhuber, 1997).

*Generative Models* : For instance, Generative Adversarial Networks (GANs) and Variational Autoencoders (VAEs) to generate images from learned data distribution (Goodfellow et al. , 2016).

### *3.1.5. Post-Processing*

After the model processes the input data, post-processing steps are often necessary to refine and interpret the results:After the model processes the input data, post-processing steps are often necessary to refine and interpret the results:

*Non-Maximum Suppression* (NMS): Regarding object detection, it eliminates multiple bounding boxes of objects and retains only the most productive ones (Ren et al. , 2015).

*Segmentation Refinement* : It is possible to enhance the boundaries and coherencies of the segmented objects using techniques such as conditional random files (Chen et al. , 2015).

*Result Interpretation* : Interpreting them into textual form or other usable forms, for example, providing descriptions of an image or its parts (GErrorException, objects and their locations detected in an image, areas of interest and so on) as illustrated below (Goodfellow et al. , 2016).

### *3.1.6. Deployment*

Deploying a computer vision system involves ensuring that the model runs efficiently in the target environment:

*Edge Devices* : As a sub-problem of edge computing, it can refer to the deployment of models on devices with restricted processing capabilities, such as smartphones, IoT devices, and drones. The following techniques such as model quantization, pruning and optimization can help in the achievement of this (Sze et al. 2017).

*Cloud-Based Systems* : Optimizing data into the cloud platforms, for real time scalability and high computational, useful for intensive computation and large data management (Dean et al. , 2012).

# 4. Computer Vision a Case sudy

One of the very important application of Computer Vision is object detection. At this use case, the goal is to precisely  locate object of interest in an image or video. The task would be to identify the position and boundaries of objects in images or videos (Wang, Sng, 2015).

## 4.1. Machine Vision in Language Translation

### *4.1.1. Google translate:*

This is an online service developed and provided by Google that allows a variety of translation services supporting over 100 languages using Neural Machine models for detection to precisely translate between languages and improve it self over time.

Visual Text translation offered by google translate uses Computer vision at it's core implementation, this done by the visual identification of the text character and precise handling of the information to give the end user a proper interpretation of the input data. Due to improvement in deep learning algorithms like convolutional neural networks google translate can not only detect an identify a dog in a picture but at it's currently trained level it is even able to detect and identify different dog breeds which older models would struggle to handle (Google AI Blog, 2018).

### *4.1.2. Google translate working*

The process of google translate identifying and accurately translating data for the user, a number of steps would taken to archive this.

1. Firstly, when the image is passed, google translate app tries to identify important parts of the image that might hold text to be translated, this involves looking for blobs of pixel that have similar color and near other similar blobs. This blobs searched would also be checked if they could be in a line so we can continuously read them.

2. After this blobs are read this is where google translate has to try recognizing what each letter actually is. This is done using the deep learning model, convolutional neural network. This network has been trained to accurately identify what different letters look like. This model has not just been trained on "precise-looking" letter images cause in the real world we'd be encountering all sort of dirty smudges or images marred by reflection, So to handle this cases the network was trained data generated by google to mimic this smudges and dirt that the real world input would include.

3. Knowing that the previous steps could potentially contain errors from recognition, the dictionary look-up of this recognised words would only be an approximation. Using this method, if an 'S' was read as '5' then the approximate word '5uper' would still be able to be looked up.

4. Finally the words translated from the dictionary look-up would be passed to the user either read-out or rendered to screen as the final result from all the computation of the previous steps.

## 5. Challenges of Computer Vision

1. Data Quality and Quantity

   *Data Annotation*: The first issue that we face in computer vision is that a lot of training data are often needed for models to learn from. Annotation of data in fields like object detection and segmentation is a tedious and costly process (Gonzalez & Woods, 2018; Szeliski, 2022).

   *Data Variability*: The change of light, occlusion, and shift in the perspective are some of the factors affecting the performance of the computer vision systems. It is still a challenge even to date to sustain such a form of resistance against such fluctuations (Szeliski, 2022).

2. Computational Complexity

   *Resource Requirements*: The deep learning models based on CNN generally require a significant amount of computational resources for training and to make predictions. This includes factors such as high-performance graphic processing units, large memory, and many others.

   *Real-Time Processing*: Real time computation for such jobs like video analysis and autonomous driving is challenging because of the higher demand for computational power (Krizhevsky, Sutskever & Hinton, 2012).

3. Model Interpretability

   *Black Box Nature*: Shared characteristics of the deep learning models mean that the models are said to make their decisions in a black box. This is difficult particularly when it comes to certain applications like diagnosis of images in medicine or in automobiles that are self-driven (Ribeiro, Singh, & Guestrin, 2016).

4. Generalization and transfer learning

*Domain Adaptation*: It's possible that some models, which work well in particular datasets, do not do so well when tested in other conditions – or on other datasets. This is why the concepts of domain adaptation and transfer learning exist, but they are still a thing to research (Pan & Yang, 2010).

5. Ethical and privacy Issues.

*Bias and Fairness*: Introducing prejudice in the training data is quite risky because it births prejudice in the models which then results to prejudice. Prejudice is another challenge that has been highlighted several times (Buolamwini & Gebru, 2018).

*Privacy Issues*: Computer vision in surveillance and facial recognition pose a great threat to the general publics right to privacy. Such development should be controlled in a manner that would still protect the personal freedoms of citizens including privacy (Harwell, 2019).

## 6. Possibilities

1. Advanced Healthcare Applications

Medical Imaging: Another way by which computer vision could transform the healthcare industry is through diagnosing illnesses in X-ray and MRI scans, and CT. Diagnostic assistance to the doctors is additional assistance that can be provided with the help of AI (Litjens et al. , 2017).

2. Autonomous Vehicles

Self-Driving Cars: Autonomous vehicle would not be possible without computer vision as it helps to interpret the environment and then move within it. This will comprise of tasks such as objects detection, lane detection and traffic sign detection (Chen et al. , 2015).

3. Enhanced Security and Surveillance

Intelligent Monitoring: Certain technologies like advanced computer vision systems can help in increasing security by performing monitoring/covert surveillance in real time as well as system anomaly detection. These systems can themselves detect and notify of emergent activities (Ren et al. , 2015).

4. Improved User Experience

*Augmented Reality* (AR): In this publication we talked about how computer vision aids in language translation in an application like google translate. Computer vision is core to AR since it uses the camera to place virtual objects on a real environment in real-time. It can be applied in game and vectoring and in education aids (Azuma, 1997).

*Facial Recognition*: Facial recognition application may improve user experience in the security aspects, socializing in the social media platforms, and customized services (Parkhi et al. , 2015).

5. Agricultural Automation

*Precision Agriculture*: Moreover, through computer vision, crop health can be continuously monitored, pests can be identified and harvesting can be mechanized which all lead to better yields in agriculture (Kamilaris & Prenafeta-Boldú, 2018).

# 7. Refrences


Ashtari, H. (2022, January 12). *What is computer vision? Meaning, examples, and applications in 2022.* Retrieved from https://www.spiceworks.com/tech/artificial-intelligence/articles/what-is-computer-vision/

Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. *Proceedings of Machine Learning Research*, 81, 77-91. Retrieved from http://proceedings.mlr.press/v81/buolamwini18a.html

Chen, C., Seff, A., Kornhauser, A., & Xiao, J. (2015). DeepDriving: Learning affordance for direct perception in autonomous driving. *Proceedings of the IEEE International Conference on Computer Vision*, 2722-2730. Retrieved from https://ieeexplore.ieee.org/document/7410757

DataCamp. (2020, July 8). *Seeing like a machine: A beginner's guide to image analysis in machine learning* [Blog Post]. Retrieved from https://www.datacamp.com/tutorial/seeing-like-a-machine-a-beginners-guide-to-image-analysis-in-machine-learning

Fortune Business Insights. (n.d.). *Computer vision market analysis*. Retrieved from https://www.fortunebusinessinsights.com/computer-vision-market-108827

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.

Google AI Blog. (2018, May 7). *How Google Translate squeezes deep learning onto a phone* [Blog Post]. Retrieved from https://research.google/blog/how-google-translate-squeezes-deep-learning-onto-a-phone/

Gonzalez, R. C., & Woods, R. E. (2018). *Digital image processing* (4th ed.). Pearson.

Harwell, D. (2019, January 15). Rights groups call for U.S. moratorium on use of facial recognition technology. *The Washington Post*. Retrieved from https://www.washingtonpost.com/technology/2019/01/15/rights-groups-call-us-moratorium-use-facial-recognition-technology/

IBM. (n.d.). *Computer vision*. Retrieved from https://www.ibm.com/topics/computer-vision

IBM. (n.d.). *Deep learning*. Retrieved from https://www.ibm.com/topics/deep-learning

Kamilaris, A., & Prenafeta-Boldú, F. X. (2018). Deep learning in agriculture: A survey. *Computers and Electronics in Agriculture, 147*, 70-90. doi:10.1016/j.compag.2018.02.016

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems, 25*, 1097-1105. Retrieved from https://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf

Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., ... & van Ginneken, B. (2017). A survey on deep learning in medical image analysis. *Medical Image Analysis, 42*, 60-88. doi:10.1016/j.media.2017.07.005

Papers with Code. (n.d.). *Object detection*. Retrieved from https://paperswithcode.com/task/object-detection#papers-list

Pan, S. J., & Yang, Q. (2010). A survey on transfer learning. I*EEE Transactions on Knowledge and Data Engineering, 22*(10), 1345-1359. doi:10.1109/TKDE.2009.191

Parkhi, O. M., Vedaldi, A., & Zisserman, A. (2015). Deep face recognition. *British Machine Vision Conference*. Retrieved from https://www.robots.ox.ac.uk/~vgg/publications/2015/Parkhi15/

Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster R-CNN: Towards real-time object detection with region proposal networks. *Advances in Neural Information Processing Systems, 28*, 91-99. Retrieved from https://papers.nips.cc/paper/5638-faster-r-cnn-towards-real-time-object-detection-with-region-proposal-networks.pdf

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?": Explaining the predictions of any classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1135-1144. doi:10.1145/2939672.2939778

SAS Institute. (n.d.). *Computer vision*. Retrieved from https://www.sas.com/en_us/insights/analytics/computer-vision.html

Szeliski, R. (2022). *Computer vision: Algorithms and applications* (2nd ed.). Springer.

Wang, L., & Sng, D. (2015). *Deep learning algorithms with applications to video analytics for a smart city*: A survey. arXiv preprint arXiv:1512.03131.

Zebra Technologies. (n.d.). *What is deep learning in machine vision*. Retrieved from https://www.zebra.com/us/en/resource-library/faq/what-is-deep-learning.html