**INDUSTRIAL TRAINING REPORT**

**Submitted by**

**ROLL NO: EC045**

MEET PARMAR (19ECUBG071)

SEMESTER VIII

****

Department of Electronics & Communication

Engineering Faculty of Technology.

Dharmsinh Desai University, Nadiad.

**WEEKLY REPORT ON THE INDUSTRIAL TRAINING**

* **Task Done till the date in the training period:**

1. In-Depth analysis of the research paper for the pre-foundation of the upcoming project.
2. Different datasets were reviewed and the usability of the same was observed.
3. Different Algorithms were analyzed which were used in the above-mentioned research paper.

* **Crux and the observations of the research papers that were analyzed:**

1. **TCR: Transformer-based Optical Character Recognition  
   with Pre-trained Models**

**Abstract:**

Existing approaches are usually built based on CNN for image understanding and RNN for char level text generation. In addition, another language model is usually needed to improve the overall accuracy as a postprocessing step. So an end-to-end text recognition approach with pre-trained image Transformer and text Transformer models, namely TrOCR, leverages the Transformer architecture for both image understanding and word piece-level text generation. The TrOCR model is simple but effective, and can be pre-trained with large-scale synthetic data and fine-tuned with human-labeled datasets.

**Introduction:**

Typically, an OCR system includes two main modules: a text detection module and a text recognition module. Text detection aims to localize all text blocks within the text image, either at word-level or textline-level. The text detection task is usually considered as an object detection problem.The text recognition task is usually framed as an encoderdecoder problem where existing methods leveraged CNNbased encoder for image understanding and RNN-based decoder for text generation. In this paper, we focus on the text  
recognition task for document images.

1. **Smart Reader Glass for Blind and Visually Impaired People**

**Abstract:**

Visually impaired people fail to read the text with existing technology. The proposed project targeted to design a spectacle with a camera by which the blind visually impaired people can read whatever they want to read based on contemporary OCR technique and text-to-speech (TTS) engines. This proposed smart reader will read any kind of documents like books, magazines and mobiles. People can access this novel technology with blindness and limited vision. Theearlier version of the proposed project was developed successfully with mobile readerwhich had certain drawbacks such as high cost due to the need of android mobile, not user friendly and improper focusing. To overcome these disadvantages, a spectacle type reader with camera is proposed in this project, which will be cost effective and more efficient.

Keywords: Micro camera, Python, Raspberry pi, Smart reader, Text-to-speech

**Introduction:**

This Project is aimed to design assistive smart glasses in wearable design format for the blind and visually impaired people. This smart spectacle with micro camera setup proposed in this project is designed to support reading printable version of any books, documents, mobile texts by converting text to audio,which can be heard by microphones or speakers. This portable and economical smartspectacle is programmed with raspberry pi module and the image processing techniquehelps in recognizing and extracting the text from the image [4, 5]. Finally, the extracted text is converted into speech and can be heard by blind and visually impaired people. Thefinal hardware model is tested with two test samples, the first one is with book page and another one is mobile document. The designed smart glass converted both test samples into a right audio format. This project is very affordable to all category of people and will be more useful.

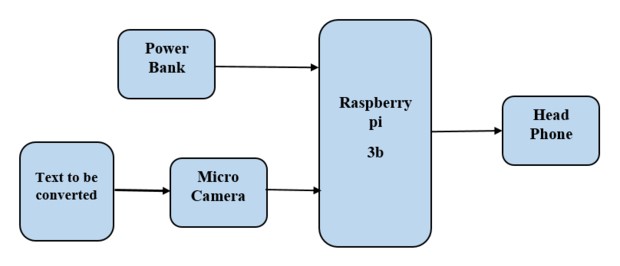


Fig 4. Block diagram of smart reader glass

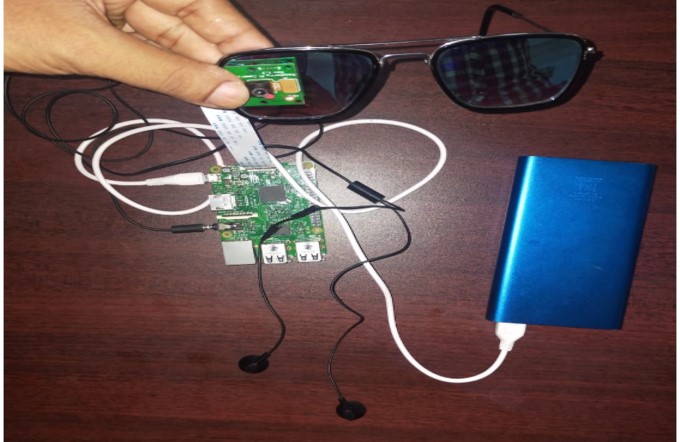
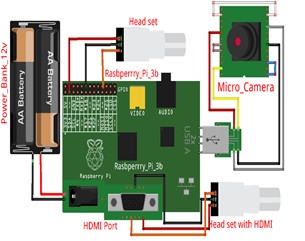


Fig 5. Circuit diagram of smart reader glass

**Results:**

The audio output can be heard using normal microphones used for mobile phone. The rigid assembling of hardware components ensure the lifetime of the proposed smart glass. Really, this smart reading spectacles will be an optimal choice for visually impaired peoples.

**Summary and Conclusion:**

The proposed project - Smart reader glass for blind and visually impaired people mainly aimed to assist the blind and visually impaired people to read the documents in the form of text book, mobile screen or computer screen. Even though some of the glasses are available in the market they are too costly and not affordable to common people. Our proposed project is designed with simple components which can be placed in shirt pocket. Hence this product is user friendly and will be available at very low price and compact size. The audio output can be heard using normal microphones used for mobile phone. The rigid assembling of hardware components ensure the lifetime of the proposed smart glass. Really, this smart reading spectacles will be an optimal choice for visually impaired peoples.

1. [**An Efficient End-to-End Neural Model for Handwritten Text Recognition**](https://arxiv.org/abs/1807.07965)

**Abstract:**

A novel approach that combines a deep convolutional network with a recurrent Encoder-Decoder network to map an image to a sequence of characters corresponding to the text present in the image. The entire model is trained end-to-end using Focal Loss [18], an improvement over the standard Cross-Entropy loss that addresses the class imbalance problem, inherent to text recognition. To enhance the decoding capacity of the model, Beam Search algorithm is employed which searches for the best sequence out of a set of hypotheses based on a joint distribution of individual characters. takes as input a down sampled version of the original image thereby making it both computationally and memory efficient

**Introduction:**

HTR has been formulated as a sequence matching problem: a sequence of features extracted from the input data is matched to an output sequence composed of characters from the text, primarily using Hidden Markov Models ( HMM ) [9][21]. However, HMMs fail to make use of the context information in a text sequence, due to the Markovian assumption that each observation depends only on the current state. This limitation was addressed by the use of Recurrent Neural Networks ( RNN ) which encode the context information in the hidden states.

1. **Zero-shot image restoration using denoising diffusion null-space model**

**Abstract:**

Denoising Diffusion Null-Space Model (DDNM), a novel zero-shot framework for arbitrary linear IR problems, including but not limited to image super-resolution, colorization, inpainting, compressed sensing, and deblurring. By refining only the null-space contents during the reverse diffusion process, we can yield diverse results satisfying both data consistency and realness.

**Introduction:**

a novel zero-shot solution for various IR tasks, which we call the Denoising Diffusion Null-Space Model (DDNM). By refining only the null-space contents during the reverse diffusion sampling, our solution only requires an off-the-shelf diffusion model to yield realistic and data-consistent results, without any extra training or optimization nor needing any modifications to network structures. Extensive experiments show that DDNM outperforms state-of-the-art zeroshot IR methods in diverse IR tasks, including super-resolution, colorization, compressed sensing, inpainting, and deblurring

* **OCR and handwriting datasets for machine learning:**

## 

[NIST Database](https://catalog.data.gov/dataset/nist-handprinted-forms-and-characters-nist-special-database-19): The US National Institute of Science publishes handwriting from 3600 writers, including more than 800,000 character images.

[MNIST Database](http://yann.lecun.com/exdb/mnist/): A subset of the original NIST data, has a training set of 60,000 examples of handwritten digits.

[Devangri Characters](http://www.iapr-tc11.org/mediawiki/index.php?title=Devanagari_Character_Dataset): A dataset of handwritten Devangari characters, composed of 1800 samples from 36 character classes obtained by 25 native writers.

[Mathematics Expressions](http://www.iapr-tc11.org/mediawiki/index.php?title=CROHME:_Competition_on_Recognition_of_Online_Handwritten_Mathematical_Expressions): More than 10,000 expressions, including more than 101 mathematical symbols.

[Chinese Characters](http://www.iapr-tc11.org/mediawiki/index.php?title=Harbin_Institute_of_Technology_Opening_Recognition_Corpus_for_Chinese_Characters_(HIT-OR3C)): A dataset of handwritten Chinese characters containing 909,818 images that corresponds to about 10 news articles.

[Arabic Printed Text](http://diuf.unifr.ch/diva/APTI/): Contains a lexicon of 113,284 words, and uses 10 Arabic fonts.

[Document database](http://www.iapr-tc11.org/mediawiki/index.php?title=IAM_Online_Document_Database_(IAMonDo-database)): Contains 941 online handwritten documents by 189 writers, and covers lists, tables, formulas, diagrams and drawings.

[Iam On-line Handwriting](http://www.fki.inf.unibe.ch/databases/iam-on-line-handwriting-database): Contains forms of handwritten English text acquired on a whiteboard, and includes more than 1700 entries.

[Street View Text](http://www.iapr-tc11.org/mediawiki/index.php?title=The_Street_View_Text_Dataset): The Street View Text dataset was harvested from Google Street View, and mostly deals with outdoor street level signs and boards.

[Street View House Numbers](http://www.iapr-tc11.org/mediawiki/index.php?title=The_Street_View_House_Numbers_(SVHN)_Dataset): Contains 73257 digits of house street numbers, taken from Google Street View.

[Natural Environment OCR](http://www.iapr-tc11.org/mediawiki/index.php?title=NEOCR:_Natural_Environment_OCR_Dataset): A dataset that contains 659 real world images with 5238 annotations of text.

[Scene Text](http://www.iapr-tc11.org/mediawiki/index.php?title=KAIST_Scene_Text_Database): Contains 3000 images captured in different environments, including outdoors and indoors scenes under different lighting conditions (clear day, night, strong artificial lights, etc).

[Text Detection](http://www.iapr-tc11.org/mediawiki/index.php?title=MSRA_Text_Detection_500_Database_(MSRA-TD500)): Contains 500 natural images, which are taken using a pocket camera. The indoor images are mainly signs, doorplates and caution plates while the outdoor images are mostly guide boards and billboards.

[Stanford OCR](http://ai.stanford.edu/~btaskar/ocr/): Contains handwritten words dataset collected by MIT Spoken Language Systems Group, published by Stanford.