PROJECT - II

Report on

OPTICAL CHARACTER RECOGNITION

Submitted in partial fulfillment of the requirements

of the degree of

Bachelor of Engineering
(Electronics and Telecommunication Engineering)

by

Tawade Shubham Ratnoji (15ET1105)
Singh Vishal Kumar (15ET1138)
Sonawane Riddesh Rajesaheb(15ET1043)

Supervisor

Mr. Jaswantsing Rajput

Department of Electronics and Telecommunication Engineering
Ramrao Adik Institute of Technology,
Sector 7, Nerul , Navi Mumbai
(Affiliated to University of Mumbai)
April 2019

Ramrao Adik Education Society's

Ramrao Adik Institute of Technology

(Affiliated to the University of Mumbai)

Dr. D. Y. Patil Vidyanagar, Sector 7, Nerul, Navi Mumbai 400 706.

Certificate of Approval

This is to certify that, the Project -II report entitled

"OPTICAL CHARACTER RECOGNITION"

is a bonafide work done by

Tawade Shubham Ratnoji (15ET1105)
Singh Vishal Kumar (15ET1138)
Sonawane Riddesh Rajesaheb(15ET1043)

and is submitted in the partial fulfillment of the requirement for the degree of

Bachelor of Engineering
(Electronics and Telecommunication Engineering)
to the
University of Mumbai.

Examiner	Supervisor

Project Coordinator	Head of Department	Principal

Ramrao Adik Education Society's
Ramrao Adik Institute of Technology
(Affiliated to the University of Mumbai)
Dr. D. Y. Patil Vidyanagar,Sector 7, Nerul, Navi Mumbai 400 706.

Declaration

We wish to state that work embodied in this dissertation entitled "**OPTICAL CHARACTER RECOGNITION**" has been carried out under the guidance of Mr. Jaswantsing Rajput at Department of Electronics and Telecommunication Engineering, Ramrao Adik Institute of Technology during 2018-2019.

We declare that the work being presented forms our own contribution and has not been submitted for any other Degree or Diploma of any University/Institute. Wherever references have been made to previous works of others, it has been clearly indicated. We understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

Tawade Shubham Ratnoji

Singh Vishal Kumar

Sonawane Riddesh Rajasaheb

Acknowledgments

We are very glad to thank to our project guide **Mr. Jaswantsing Rajput**, our project co-ordinators **Dr. Ashwini Naik** and **Mr. Bhushan Deore** and our **H.O.D Dr. M.D.Patil** for their encouragement and valuable guidance. We are also very much thankful to our colleagues for their constant support and time to time help. We have been fortunate to have received many useful suggestions from our colleagues which have greatly improved the clarity of our report. At the end special thanks to our **Principal Dr. Ramesh Vasappanavara**. We would like to appreciate suggestions and criticisms about the report from the readers.

Abstract

In India, quite 310 million folks use syllabic script for documentation. There has been a big improvement within the analysis associated with the popularity of written moreover as written syllabary text within the past few years. All feature-extraction techniques as well as training, classification and matching techniques useful for the recognition are used.

In any OCR or organization extracting discriminating feature is most vital and crucial step for its success. Accuracy of such system often depends on the good feature representation. In this project, we've got used convolutional neural network formula. during this half, initial image is preprocessed to get rid of noise, born-again to binary image, resized to mounted size so convert to grey scale image victimisation mask operation, it blurs the sides of the pictures. Finally network weight parameters area unit fine tuned by supervised back propagation learning to enhance the recognition performance.

Today there are certain areas where new methodologies are needed for the increasing desires. With these new methodologies the several techniques area unit gift for the character recognition of handprint Devnagari, Bangla, Tamil, China etc. but it is not used much. therefore during this project, we've got used a Minimum distance classifier technique for OCR System of written still as scanned newspaper Sanskritic language script.

Contents

List of Figures

Chapter 1

Introduction

1.1 Introduction to Devanagari Script

Devanagari could be a Northern Brahmic script associated with several alternative South Asian scripts as well as Gujarati, Bengali and Gurmukhi, and a lot of distantly, to variety of South-East Asian scripts as well as Thai, Bahasa Indonesia and Baybayin. The script is employed for over a hundred and twenty spoken Indo-Aryan languages, as well as Hindi, Nepali, Marathi, Maithili, Awadhi, Himalayish and Bhojpuri. It is also used for writing Classical Sanskrit texts. Generally the orthography of the script reflects the pronunciation of the language. Hindi is unremarkably spoken mistreatment a combination of around fifty two sounds, 10 vowels, forty consonants, nasalization and a kind of aspiration. These sounds area unit diagrammatically in the syllabary script by thirteen characters historically considered vowels and forty consonants.

Devanagari evolved from the Brahmi script. The word Devanagari has been mystery to students, there's a hypothesis that it'd be combination of 2 Sanskrit words Deva (God, king or Brahmans) and Nagari (city). virtually it combines to make City of Gods, Script of Gods.

Devanagari, a development of script system of acoustics, is that the solely script that has specific signs (grapheme) for the phonetically organized sounds of the human speech (phonemes), and it's versatile enough to write down foreign sounds by attaching marks to the nearer graphic symbol. The Roman, Greek, Hebrew, and Arabic alphabets have sure ancient names for indicating sound photos however there's no guarantee that one sign can have just one phonetic worth.

The Devanagari script is a crucial and wide used script of India. it's chiefly wont to write Hindi, Marathi, Nepali and Sanskrit languages. It conjointly is AN auxiliary script for alternative languages like Punjabi, Sanskrit and Kashmiri.

Devanagari is AN abugida. Consonant letters have AN inherent speech sound. Combining vowel-signs square measure hooked up to the consonant to point that a special vowel follows the consonant. See the table within the right-hand column for a short summary of options, taken from the Script Comparison Table.

Actually, Devanagari script is predicated on writing syllables, therefore the vowel-sign is if truth be told hooked up to the linguistic unit. AN writing linguistic unit includes clusters of consonants while not intervening vowel sounds. These clusters square measure generally portrayed as part incorporated forms, referred to as conjuncts.

The Devanagari script block contains additional characters than different Indo-Iranian language scripts, partially as a result of it is a pivot script for transliterations of different scripts.

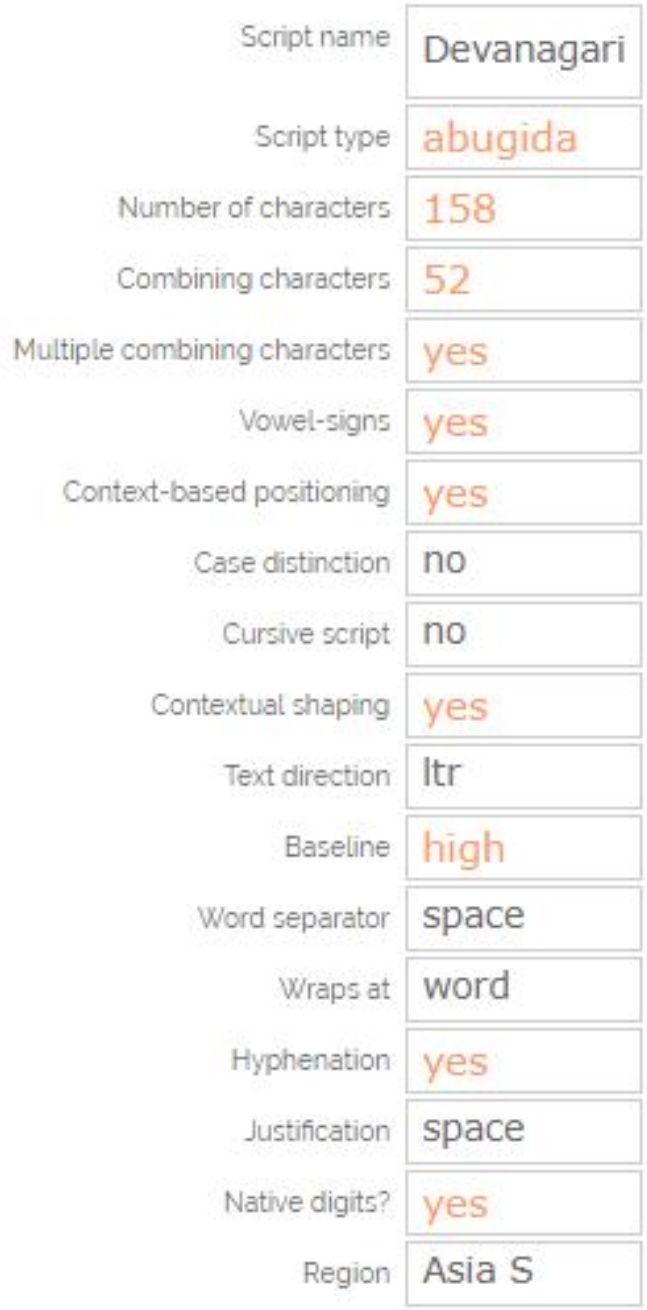


Figure 1.1: devanagari feature

Vowels

Indian phonograph is vowel dominant; each vowel is realizable in 3 scales Short, Long, Prolonged. All vowels can be pronounced in non-nasal and nasal modes. Which means each vowel can have 18 realizations (three divisions on the position of the particular organ in the mouth while pronouncing that letter, and two divisions of nasal or non-nasal. Since each group is independent of the others, the total number of pronouncements comes to 18). Since it is not possible to record these differences and they must be remembered by listening to it properly, the Vedas were not written.

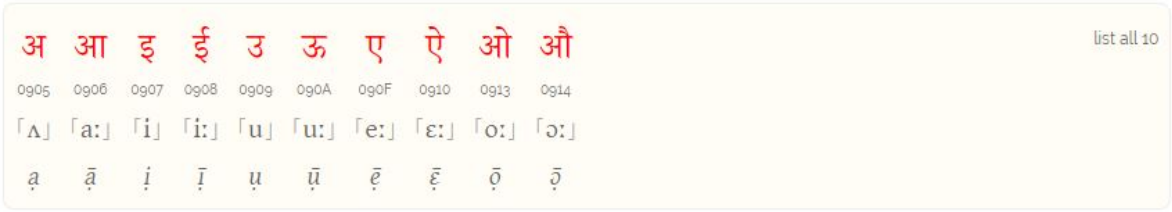


Figure 1.2: vowels in devanagari

Consonants

Figure 1.3: consonants in devanagari

All consonants designs either touch or cross the vertibar. There are exceptions only in the design of letters such as GA, NA, SHA. These letters do not touch the vertibar. This is a graphic peculiarity to point to the fact that the writing is a Ganesh Vidya.

According to the tradition of the scribes of the Ganapati School, one scribe came to write the copies text of Mahabharata for the author Vyasa Muni. Ganesha introduced vertibar of A vowel. This feature is high-lighted by Vyasa in Bhagwad Gita wherein Krishna says, I am the common factor of Aa kaar in all letters. Thus this feature was added in all consonantal designs. After the vertibar is drawn predominantly the graphics becomes Dev-Lipi.

1.2 Guideline for writting in devanagari script

i Devanagari characters hang from a horizontal line (called the head stroke) written at the top of the character. Unlike English letters which are written up from a line below them.

ii The body of the syllabic script characters ought to occupy regarding 2 thirds of the house between the lines.

iii In general the first stroke, or strokes, in a character are written from the left to the right and are then followed by any down strokes and finally the head stroke is added. Note that in some characters the head stroke is broken.

Chapter 2

Literature Survey

Optical Character Recognition is widely used for making editable soft copies of printed media. But when it comes to using OCR for documents written or printed in scripts other than Latin character segmentation becomes a very tough task , especially for indian languages.There are ancient documents written in devanagari script which are needed to be digitized in order to preserve and promote but because of its unique scripting digitization of such documents is very difficult.

There have been a few projects to use OCR for devanagri script. We can trace a very common method used in every single project for the purpose of segmentation of individual characters from words. In devanagari script every letter has a horizontal upper line so when a word is formed or written grouping together letters the upper line of every letter gets connected forming one single shitolekha. The first task in order to separate every single character is to remove shiroleka and it is done by providing the condition of a row having most number of dark pixels, then this row is inverted which separate out every letter. Once the letters are separated now we have to bound every sing letter. For this a bounding box is required and it is done by declaring a condition that specifies that the edge of the bound is defined by the column with all pixels white.

Chapter 3

Optical Character Recognition (OCR)

3.1 Definition of OCR

OCR Stands for Optical Character Recognition. OCR application is able to recognize and extract text information out of scanned document, such as PDF, TIFF, or other document image files. A PDF Converter with OCR ability can converts scanned PDF document into editable text.

Optical character recognition is that the mechanical or electronic conversion of scanned pictures of written, written or written text into machine-encoded text. it's extensively used as a tool for knowledge entry from some form of original paper knowledge supply, whether or not documents, sales receipts, mail, or any variety of written records. it's a general technique of digitizing written texts so they'll be electronically searched, keep a lot of efficiently, displayed on-line, and utilized in machine processes like AI, text-to-speech and text mining. OCR may be a field of analysis in pattern recognition, AI and laptop vision.

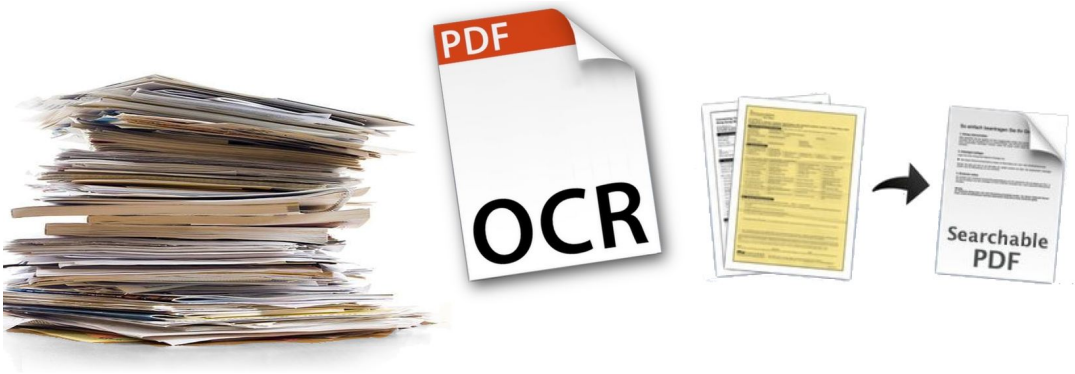


Figure 3.1: OCR

Optical Character Recognition, or OCR, may be a technology that allows you to convert differing kinds of documents, like scanned paper documents, PDF files or pictures captured by a camera into editable and search-able knowledge.

Imagine we got a paper document - as an example, article, brochure, or PDF contract your partner sent to you by email. Obviously, a scanner isn't enough to create this data out there for redaction, say in Microsoft Word. All a scanner will do is produce a picture or a snap of the document that's nothing over a set of black and white or color dots, called a formation image. so as to extract and re purpose knowledge from scanned documents, camera pictures or image-only PDFs, we would like AN OCR package that may single out letters on the image, place them into words so - words into sentences, so facultative you to access and edit the content of the initial document.

Probably the foremost documented use case for OCR is changing written paper documents into electronic text documents. Once a scanned paper document went through OCR process, the text of the document is emended with word processors like Microsoft Word or Google Docs. Before OCR technology was accessible, the sole choice to change written paper documents was to manually re-typing the text. Not solely was this massively time intense, it additionally came with quality and typewriting errors.

OCR is commonly used as a hidden technology, powering several acknowledge systems and services in our way of life. Less known, however as necessary, use cases for OCR technology embody information entry automation, assortment documents for search engines, automatic range plate recognition, in addition as helping blind and visually impaired persons.

OCR technology has tried vastly helpful in digitising historic newspapers and texts that have currently been regenerate into absolutely searchable formats and had created accessing those earlier texts easier and quicker.

3.2 History of OCR

Practically, character recognition could be a set of the pattern recognition.However, it is character recognition that gives the motive for creating pattern recognition and image processing in matured fields of science.

To implement the human features by machines, creating the machine ready to perform tasks like reading. The roots of character recognition can be traced back in 1870. This was the year that C.R.Carey of Hub of the Universe Massachusetts fictitious the tissue layer scanner that was a picture gear mechanism employing a mosaic of photocells. Two decades later the Polish P. Nipkow fictitious the successive scanner that was a serious breakthrough each for contemporary TV and reading machines. During the primary decades of the 19 the century many makes an attempt were created to develop devices to help the blind through experiments with OCR. However, the fashionable version of OCR failed to seem till the center of the 1940s with the event of the digital computer. The motivation for development from then on, was the attainable applications within the business world.

By 1950 the age was moving forward at a great speed, and data processing was changing into a very important field. information entry was performed through punched cards and a cheap manner of handling the increasing quantity of knowledge was needed. At

constant time the technology for machine reading was changing into sufficiently mature for application, and by the center of the 1950s OCR machines became commercially obtainable. The first true OCR reading machine was put in at Readers Digest in 1954. This instrumentation was accustomed convert typed sales reports into punched cards for input to the computer.

3.3 First Generation OCR

The industrial OCR systems showing within the amount from 1960 to 1965 is also known as the first generation of OCR.OCR machines of this generation were mainly characterised by the affected letter shapes browse. The symbols were particularly designed for machine reading, and therefore the initial ones did not look natural. After some time, multi-font machines started to browse up to 10 totally different fonts.The quantity of fonts were limited by the pattern recognition,model matching that compares the character image with a set of example pictures of every character of every font.

3.4 Second Generation OCR

The second generation reading machines came in middle of 1960s and 1970s. These machines were capable of recognising machine written as well as hand written characters. Once hand written characters were thought of, the list was strained to numerals and a couple of letters and symbols.

Figure 3.2: OCR

The first and famed system of this type was the IBM 1287, that was exhibited at the World honest in 1965. Also, during this amount Toshiba developed the primary automatic letter sorting machine for zip code numbers and Hitachi created the primary OCR machine for high performance and low value.

In this amount vital work was tired the world of standardization. In 1966, a thorough study of OCR needs was completed Associate in OCR list was defined OCR-A. This font was extremely conventionalized and designed to facilitate optical recognition, though still clear to humans. a ecu font was conjointly designed, OCR-B, which had additional natural fonts than the yank normal. Some tries were created to merge the 2 fonts into one normal, however instead machines having the ability to browse each standards appeared.

3.5 Third Generation OCR

For the third generation of OCR systems, showing within the middle of the 1970s, the challenge was documents of poor quality and enormous written and hand-written character sets. Low value and high performance were conjointly vital objectives, that were helped by the dramatic advances in hardware technology. Although a lot of subtle OCR-machines began to seem at the market straightforward OCR devices were still terribly helpful. within the amount before the private computers and optical maser printers began to dominate the realm of text production, writing was a special niche for OCR.

The uniform print spacing and tiny variety of fonts created merely designed OCR devices very helpful. Rough drafts may well be created on standard typewriters associated fed into the pc through an OCR device for final writing. during this approach word processors, which were an expensive resource at this point, may support many individuals and also the prices for instrumentation may well be cut.

3.6 OCR Today

Even though, Optical character recognition has become commercial but out there already within the 1950s, only a few thousand systems had been oversubscribed universally up till 1987. The most common reason being the cost of system.Since hardware was less expensive and OCR systems became out there as packages, the sale exaggerated significantly. Today up to a few thousands is that the range of systems oversubscribed weekly, and therefore the price of associate OCR has born with an element of 10 each alternative year for the last half dozen years.

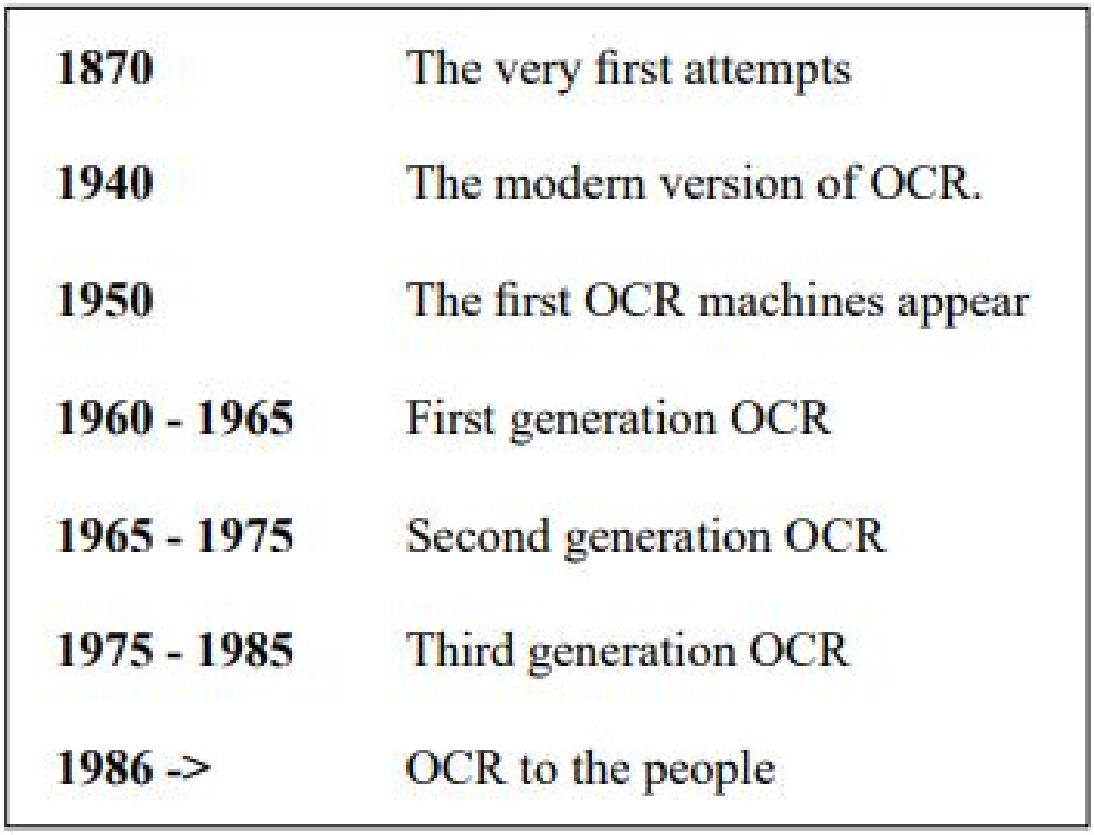


Figure 3.3: OCR Chronology

3.7 Need of OCR

Scanning a document and saving it in a picture file format (.tiff, .JPEG, etc.) is analogous to taking an image of a document with a camera or creating a duplicate of it by employing a copy machine, however instead we have a tendency to square measure employing a scanner. If we might prefer to edit the scanned image in our application program, we are going to ought to perform OCR on the image to convert the image into editable text. OCR may be a technology that converts documents that will—we will—we are able to scan into documents that our pc can scan. throughout the conversion, the document is analyzed, and characters and words square measure saved as editable text.

3.8 Methods of OCR

The primary concept in automatic recognition, is 1st to show the machine that groups of patterns that will appear and their appearance. In OCR the patterns are digits, letters and a few special characters. The training of machine is done by showing the samples of characters of the various categories. supported these examples the machine builds a epitome or an outline of every category of characters. Then, throughout recognition, the unknown characters are compared to the antecedent obtained descriptions, and allotted the category that offers the simplest match.

In most business systems for character recognition, the coaching method has been performed earlier. Some systems do but, embrace facilities for coaching within the case of inclusion of recent categories of characters.

3.9 Components of an OCR system

An OCR system has many elements. In given figure a standard setup is shown. The primary step within the method is to modify the document exploitation scanner. once the area containing text are settled, every image is taken through a method called segmentation. The symbols might then be pre-processed, removing noise, to help the extraction of options within the next stage.

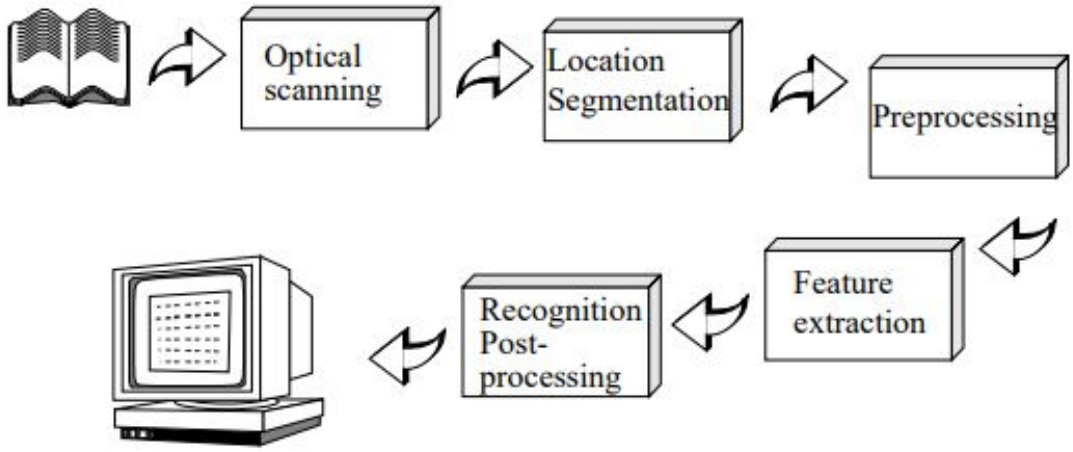


Figure 3.4: OCR Components

The identity of every image is obtained by scrutiny the obtained options with details of image categories obtained through past learning section. Finally discourse information is employed to reconstruct the words and numbers of the first text. within the next sections these steps and a few of the ways concerned are delineated in additional detail.

3.10 OCR for devanagari script

In this project, we've got used a theme to develop complete OCR system for explicit fonts and sizes of devanagari characters in order that we will use this technique for digitizing syllabic script document. we've got enforced steps of the OCR system like preprocessing, segmentation, feature extraction and classification.

Chapter 4

Working

4.1 Scanning

In OCR optical scanners are used that encompass transport mechanism and sensing device that converts strength into gray levels. written documents encompass black print on white background. once acting OCR structure image is reborn into bi-level black and white image. This method referred to as thresholding is performed on scanner to save lots of memory house and machine effort. The thresholding method is very important because the results of recognition are all passionate about quality of bi-level image.

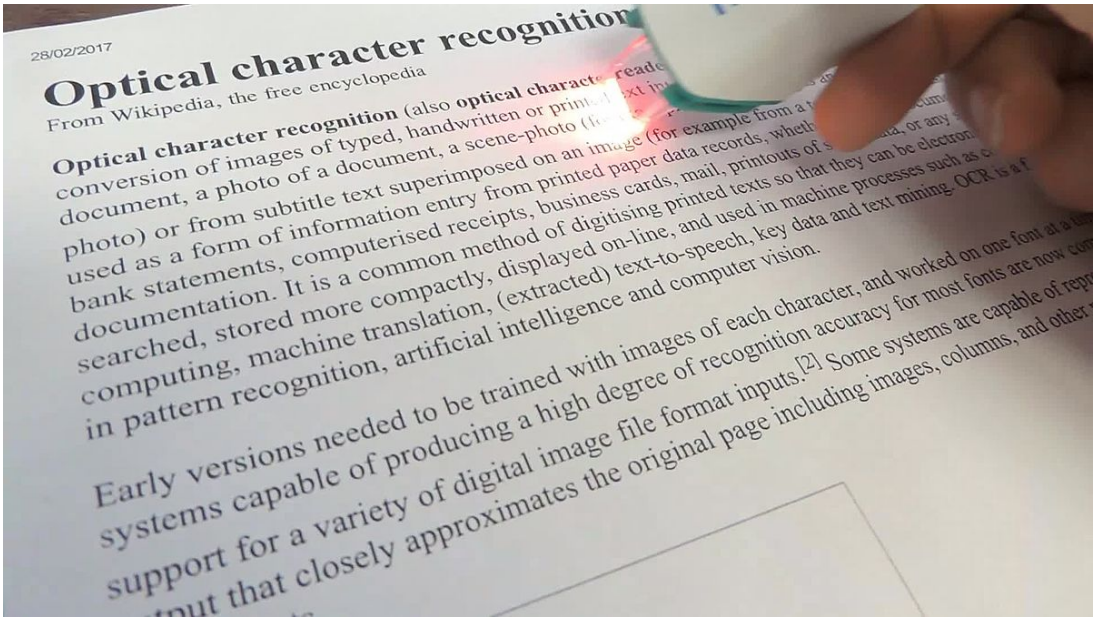


Figure 4.1: Scanning in OCR

A fixed threshold level is used where gray levels below this level are black otherwise levels above are considered white. For high contrast document with uniform background a pre-chosen fixed threshold can be sufficient. However, documents encountered in practice have rather large range. In these cases more sophisticated methods for thresholding are

required to obtain good results. The best thresholding methods vary threshold adapting to local properties of document such as contrast and brightness. However, such ways typically rely on structure scanning of document which needs additional memory and process capability.

4.2 Pre-Prcoessing

The third OCR element is pre-processing. The data counting on the info acquisition kind is subjected to variety of preliminary process steps to create it usable within the descriptive stages of character analysis. The image ensuing from scanning method could contain specific amount of noise. counting on the scanner resolution and also the inherent thresholding, the characters is also unclean or broken. a number of these defects which can cause poor recognition rates and square measure eliminated through pre-processor by smoothing digitized characters. Smoothing implies each filling and cutting. Filling eliminates little breaks, gaps and holes in digitized characters whereas cutting reduces breadth of line. the foremost common technique for smoothing moves a window across binary image of character and applies bound rules to the contents of window. Pre-processing conjointly includes standardization along with smoothing. The standardization is applied to get characters of uniform size, slant and rotation. the proper rotation is found through its angle. For turned pages and contours of text, variants of Hough remodel square measure ordinarily used for police investigation skew.

Some can argue that image pre-processing isn't an honest plan, since it distorts or changes verity nature of the information. However, intelligent use of image pre-processing will give edges and solve issues that ultimately cause higher feature detection. we tend to applied common ways for image enhancements and corrections which will have an effect on feature analysis downstream within the vision pipeline in each favorable and unfavorable ways that, reckoning on however the ways area unit utilized.

Pre-processing stage consists of compression and binarization steps.

1. Compression
2. Binarization

4.2.1 Compression

Scanners, digital cameras and different image capture devices typically turn out high resolution pictures of terribly high file sizes, and these got to be compressed into smaller sizes for efficient storage and retrieval. These devices produce the image files as fuss or raw format containing a great deal of redundant information, similarly as continual headers. pressure scanned pictures will take away this redundant information.

Several techniques exist for pressure scanned pictures. fuss pictures will simply be born-again into jpg, svg, gif, bmp or different such file formats that allow up to sixty p.c compression, looking on the resolution of the initial scanned image.

Scanned image typically contains a great deal of information that's repeated; as an example, a picture showing a bit of the sky can contain an oversized space that's uniformly blue. A compression algorithmic program, like the one utilized in a jog file, identifies this repetition and sets up a table containing one example from every continual information so it will reconstruct the image from these examples. additionally, it conjointly scans for patterns and extraneous information, so it will discard any further information and still contain all the essential info. during this method, it will compress our scanned pictures with negligible loss of quality. this enables the user to compress scanned pictures with smallest loss of quality and necessary information.

4.2.2 Binarization

The presence of background pictures or textures isn't the sole issue which will impair recognition quality. Low recognition quality brings additionally the low distinction of the first document and also the dynamic brightness of the background. For such documents the adaptive Binarization procedure is employed. It measures the brightness of the background and also the saturation of the black areas on the road so as to search out best binarization parameters for every separate line's fragment. As a result, the lines and words are properly detected and better recognition accuracy are reached.

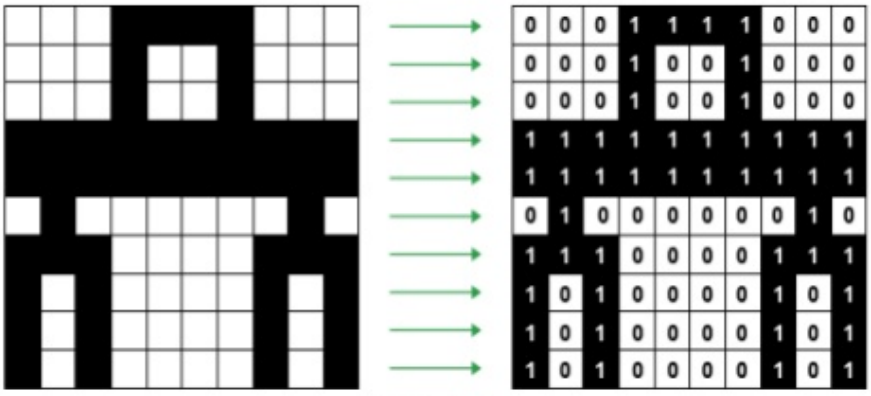


Figure 4.2: Binarization

With the help of bilateral filter applied on complex images multiple times, the input image is then converted to grey scale. A copy of the resultant grey scale image is inverted to handle various challenges with respect to font and background color. We will apply thresholding techniques like median blur followed by Gaussian blur on both the images and then we will add a black color border of one pixel width to both images and use flood fill technique to flood the pixels with white pixels. Now, will calculate black pixels in both the images and discard the one with less black pixels. The image is thus converted to a pure binary image that can be accessed as grey scale image with only two values 0 for black and 255 for white.

4.3 Image Noise

4.3.1 Noise sources in image

With the advancement in computers and digital imaging technologies, the prices of digital cameras and computers area unit lowering every years,and therefore these equipment are getting cheap of late.The usage of digital pictures in our everyday life is popping common. As a lot of data are often extracted from digital pictures, as compared to 1 dimensional signal,many analysis areas,including material researches,are currently utilizing digital images,such as microscopic pictures and X-ray pictures, collectively of their analysis tools. because the data from digital pictures area unit easier to be evaluated as compared with one dimensional signals, digital pictures area unit currently commonly employed in might analysis fields. sadly, kind of like alternative digital signals, digital pictures also are sometimes accidentally corrupted by unwanted signals, known as noise. Image may be a powerful medium to convey visual data. In digital Image process, removal of noise may be a extremely demanded space of analysis. Digital images area unit typically corrupted by noise throughout their acquisition and transmission. abuzz pictures are often found in many todays imaging applications. TV pictures area unit corrupted due to region interference and imperfections within the image reception. Noise is additionally introduced in digital artworks once scanning broken surfaces of the originals. Impulse noise corruption typically happens in digital image acquisition or transmission process as a results of photo-electronic device faults or channel bit errors.

Image transmission noise is also caused by numerous sources, like automobile ignition systems, industrial machines within the locality of the receiver, switching transients in power lines, lightning within the atmosphere and numerous unprotected switches. Digital images play a crucial role each in everyday life applications like television system, resonance Imaging, pc imaging additionally as in areas of analysis and technology like geographical information systems and natural philosophy .Digital pictures area unit typically corrupted by differing types of noise throughout its acquisition and transmission section. Such degradation negatively influences the performance of the many image processing techniques and a pre-processing module to filter the photographs is usually needed . Noise removal is one in all the main considerations within the field of pc vision and image process. Images are often contaminated by impulsive noise because of abuzz sensors or channel transmission errors or faulty storage hardware. The goal of removing impulsive noise is primarily to suppress the noise additionally on preserve the integrity of edges and elaborated data. within the field of digital image process, two applications of nice importance area unit noise filtering and image sweetening.

4.3.2 Different types of noises in images and their models

Impulsive Noise

One of the noises normally corrupting digital image is that the impulse noise. Therefore, impulse noise reduction has become one in all the active researches in these recent years. several impulse noise models are projected by analysisers for this research purpose. because of variety of non-idealities within the imaging method the noise usually corrupts pictures by exchange a number of the pixels of the initial image with new pixels having luminous values close to or capable the minimum or most of the allowable dynamic luminous vary. Impulse noise is caused by run-down pixels privately sensors, faulty memory locations in hardware, or transmission in an exceedingly noisy channel. Impulse noise is classified into 2 types: fixed-valued impulse noise and random-valued impulse noise. The fixed-valued impulse noise is additionally referred to as salt-and pepper noise wherever the grayscale price of a noisy element is either minimum or most in gray-scale pictures. The grayscale values of yelling pixels corrupted by random-valued impulse noise square measure uniformly distributed within the vary of [0,255] for grey scale images. Impulse noise is assumed as associate degree additive noise, and at random damages the element, at random positions.

Normally, impulse noise seems as black and white speckles on the image. Pixels corrupted by impulse noise are ordinarily having either extraordinarily high, or extraordinarily low intensity values. Usually, they need terribly high distinction towards their clean, uncorrupted swish encompassing areas. Therefore, impulse noise, even at a low level of degradation, can harm the looks of digital image per the distribution of rackety pixel values, impulse noise are often classified into 2 categories: Fixed-Valued impulse noise and Random valued impulse noise. The fastened Values impulse noise is additionally referred to as Salt and Pepper Noise since the component value of a loud component is either minimum or most worth in grayscale pictures. The values of rackety pixels corrupted by random valued impulse noise ar uniformly distributed within the vary of [0,255] for grayscale images. Removal of Random valued impulse noise is a lot of sophisticated thanks to the random distribution of the noise pixels.

The salt-and-pepper noise are referred to as shot noise, impulse noise or spike noise that's typically caused by faulty memory locations , bad component components within the camera sensors, or there are often temporal arrangement errors in the process of digitisation. within the salt and pepper noise there ar solely 2 doable values exists that's a and b and the chance of every is a smaller amount than zero.2.If the numbers bigger than this numbers the noise can swamp out image. For 8-bit image the standard worth for 255 for salt-noise and pepper noise is zero.

Gaussian Noise

An image gets corrupted with differing types of noise .Noise is also classified as substitutive noise (impulsive noise:salt and pepper noise,random valued impulse noise,etc.) and additive noise (e.g.additive white mathematician noise). In several occasions,noise in digital pictures is found to be additive in nature with uniform power within the whole information measure with mathematician chance distribution.such a noise is named as additive white mathematician noise(AWGN).

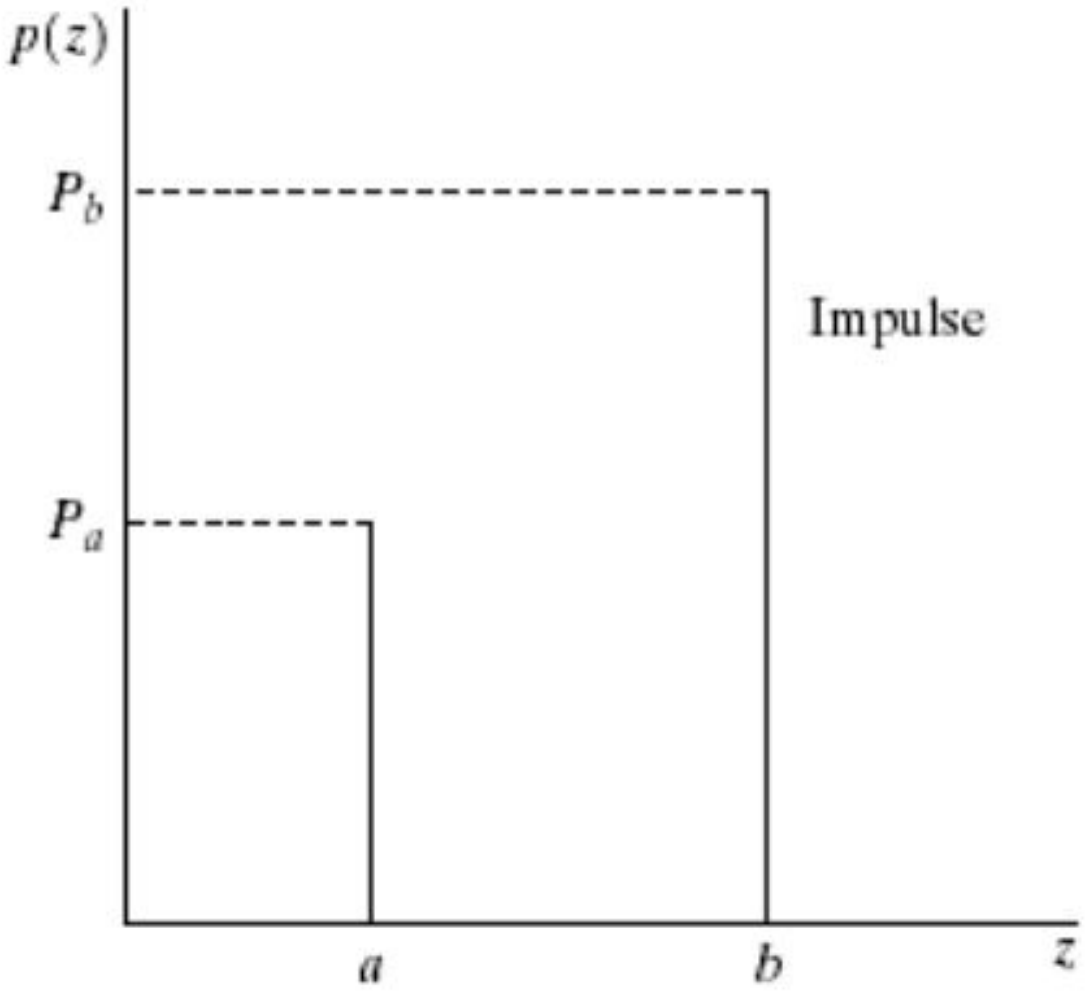


Figure 4.3: Impulse Noise Model

G(t) could be a chance variable that features a mathematician chance distribution .It is AN additive niose that's charactrsed by its variance . In , the uproarious image is described as a add of the initial uncorrupted image and also the mathematician distributed random noise G. once the variance of the random noise G is extremely low, G(x,y) is zero or terribly on the point of zero at several picture element locations.Under such circumstances,the uproarious image fAWGN is same or terribly on the point of the initial image f(x,y) at several picture element locations(x,y).

4.3.3 Image De-Noising

Image de-noising is extremely vital task in image process for the analysis of pictures. Ample image de-noising algorithms are accessible, however the simplest one ought to take away the noise utterly from the image, whereas protective the details. De-noising ways is linear yet as non-linear. wherever linear ways are quick enough, but they do not preserve the main points of the pictures, whereas the non- linear ways preserve the main points of the images.

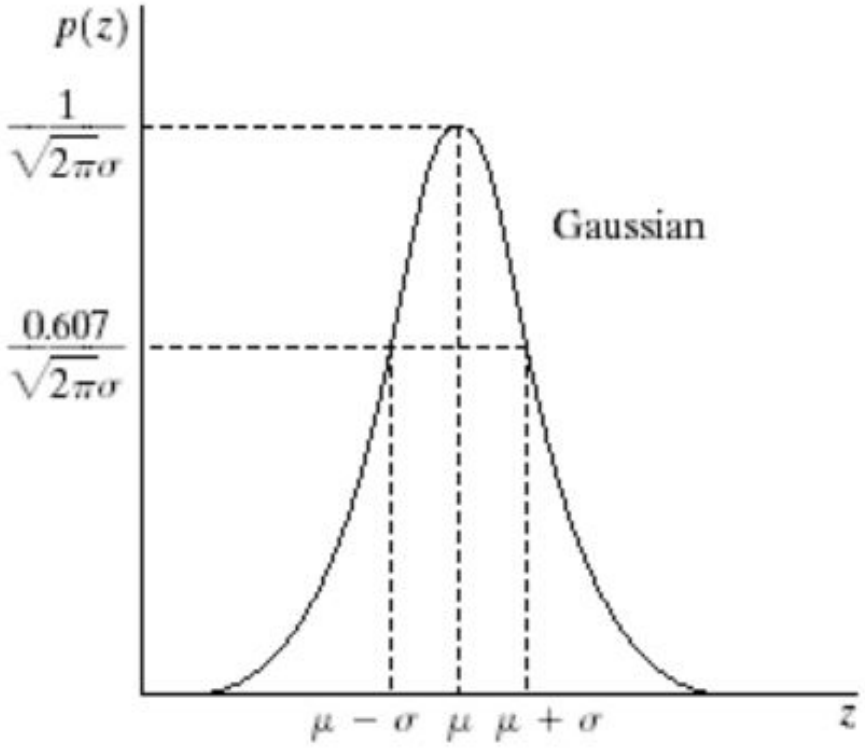


Figure 4.4: Gaussian Noise Model

Different types of filters used to remove the noise

Median Filter

Median filter may be a best order static, non- linear filter, whose response relies on the ranking of pel values contained within the filter region. Median filter is sort of common for reducing bound kinds of noise. Here the middle value of the pel is replaced by the median of the pel prices beneath the filter region .Median Filter may be a simple and powerful non-linear filter that relies order statistics. it's straightforward to implement methodology of smoothing pictures. Median filter is employed for reducing the number of intensity variation between one pel and the other pel. during this filter, we have a tendency to don't replace the pel price of image with the mean of all neighboring pel values, we have a tendency to replaces it with the norm. Then the median is calculated by initial sorting all the pel values into ascending order so replace the pel being calculated with the center pel price. If the neighboring pixel of image that is to be take into account contain an excellent numbers of pels, than the typical of the 2 middle pixel values is employed to switch. The median filter offers best result once the impulse noise proportion is a smaller amount than 0.1 percent. once the amount of impulse noise is accrued the median filter not offers best result. Median filter is good for salt and pepper noise. These filters are wide used as smooth for image process, likewise as in signal process. a serious advantage of the median filter over linear filters is that the median filter will eliminate the result of input noise values with extraordinarily giant magnitudes.

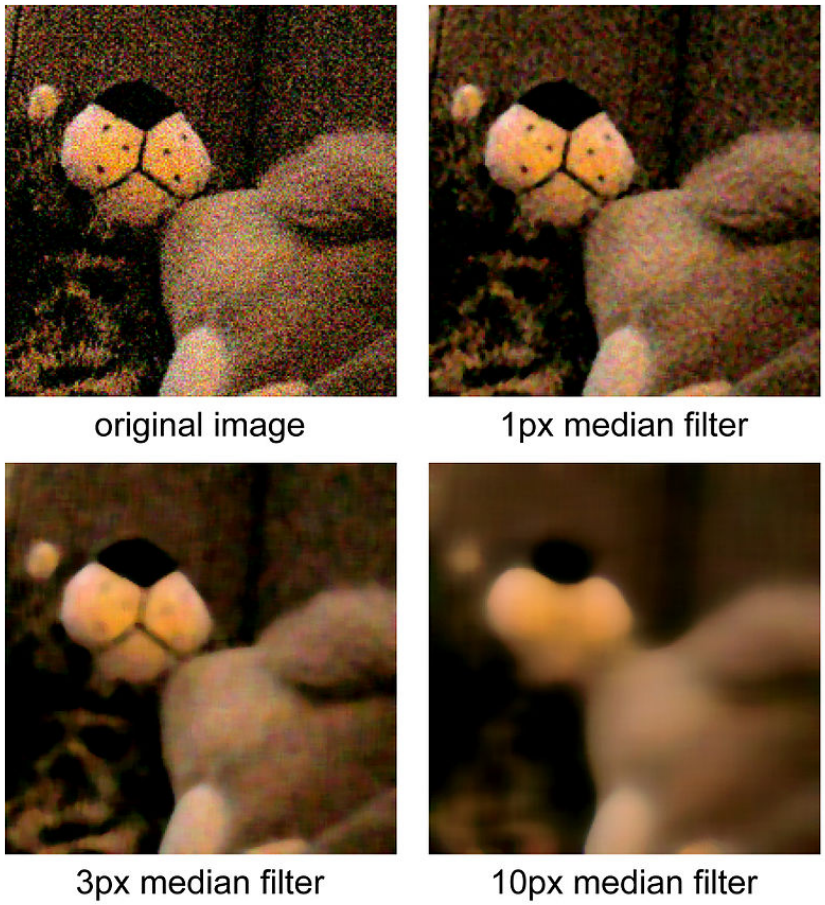


Figure 4.5: Median Filter

Mean Filter

Mean filter is associate degree averaging linear filter. Here the filter computes the typical price of the corrupted image in a very predecided space. Then the middle pel intensity price is replaced by that average price. The mean filter may be a simple abstraction filter. it's sliding-window filters that replace the middle price within the window. It replaces with the average mean of all the pel values within the kernel or window. The window is typically sq. however it is of any shape.

Gaussian blur

In image process, a mathematician blur (also referred to as mathematician smoothing) is that the results of blurring a picture by a mathematician operate (named once scientist and someone Carl Friedrich Gauss). it's a wide used result in graphics software system, usually scale back—to scale back—to cut back image noise and reduce detail. The visual result of this blurring technique could be a swish blur resembling that of viewing the image through a semitransparent screen, clearly totally different from the bokeh result made by AN out-of-focus lens or the shadow of AN object below usual illumination. mathematician smoothing is additionally used as a pre-processing stage in laptop vision algorithms so as to reinforce image structures at totally different scalessee scale house illustration and scale house implementation.

Figure 4.6: Gaussian Blur

Mathematically, applying a mathematician blur to a picture is that the same as convolving the image with a mathematician operate. this can be additionally referred to as a two-dimensional Weierstrass rework. in contrast, convolving by a circle (i.e., a circular box blur) would additional accurately reproduce the bokeh result. Since the Fourier rework of a mathematician is another mathematician, applying a mathematician blur has the result of reducing the image's high-frequency components; a mathematician blur is so an occasional pass filter.

Gaussian blurring is usually used once reducing the dimensions of a picture. once downsampling a picture, it's common to use a low-pass filter to the image before resampling. this can be to make sure that spurious high-frequency data doesn't seem within the downsampled image (aliasing). mathematician blurs have nice properties, like having no sharp edges, and so don't introduce ringing into the filtered image.

4.4 Image Segmentation

First, the captured document pictures square measure pre-processed for the attitude correction and noise removal.Then, the ultimate image is regenerate to grayscale and binarized victimisation Otsu segmentation technique for additional process. moreover, wanting at the mean horizontal run length of each black and white pixels the correct segmentation of foreground objects is checked. as an example, for the document pictures having dark background and lightweight foreground, the output of the binarization is reversed i.e. black background (represented as 0s) and white foreground (represented as 1s).

Figure 4.7: Image Segmentation

Segmentation of lines and words: The preliminary segmentation consists of the following steps:

1. We cypher the horizontal projection of the document image box. produce one vector during which all the columns in row are white pixels. And from that variety of rows line are separated from text.

2. We cypher the horizontal projection of the document image box. The row containing most variety of black pixels is taken into account to be the header line and take away it.

3. Separate character/symbol boxes of the image below the header line: to try and do this, we tend to create vertical projection of the image ranging from header line position to rock bottom row of the word image box. The columns that haven't any black pixels are treated as boundaries for extracting image boxes akin to characters.

4. We cypher the vertical projection of the image, ranging from the highest row of the image to the header Line Position. The columns that haven't any black pixels are used as delimiters for extracting prime modifier image boxes.

4.5 Feature Extraction

The next OCR element is feature extraction. the target of feature extraction is to capture essential characteristics of symbols. Feature extraction is accepted joined of the foremost tough issues of pattern recognition. the foremost uncomplicated means of describing character is by actual formation image. Another approach is to extract bound options that characterize symbols however leaves the unimportant attributes. The techniques for extraction of such options area unit divided into 3 groups viz. distribution of points transformations and series expansions and structural analysis.

The different teams of options area unit evaluated in step with their noise sensitivity, deformation, simple implementation and use. the standards utilized in this analysis are: hardiness in terms of noise, distortions, vogue variation, translation and rotation and sensible usage in terms of recognition speed, implementation quality and independence. a number of the usually used feature extraction techniques area unit example matching and correlation, transformations, distribution of points and structural analysis.

There are many features are extracted for the recognition of Marathi characters. For that consider features as follows-

1. Feature extraction using convolution

4.5.1 Convolution

Natural pictures have the property of being stationary, that means that the statistics of 1 half of the image square measure an equivalent as the other part. this means that the options that we tend to learn at one a part of the image may also be applied to alternative components of the image, and that we will use an equivalent options in the slightest degree locations.

More exactly, having learned options over tiny (say 8x8) patches sampled every which way from the larger image, we are able to then apply this learned 8x8 feature detector anyplace within the image. Specifically, we are able to take the learned 8x8 options and twist them with the larger image, therefore getting a distinct feature activation price at every location within the image.

To give a concrete example, suppose we've learned options on 8x8 patches sampled from a 96x96 image. Suppose additional this was finished associate auto-encoder that has one hundred hidden units. to urge the convolved options, for each 8x8 region of the 96x96 image, that is, the 8x8 regions beginning at (1, 1), (1, 2), . . . (89, 89), we might extract the 8x8 patch, and run it through our trained thin autoencoder to urge the feature activations. this might end in one hundred sets 89x89 convolved options.

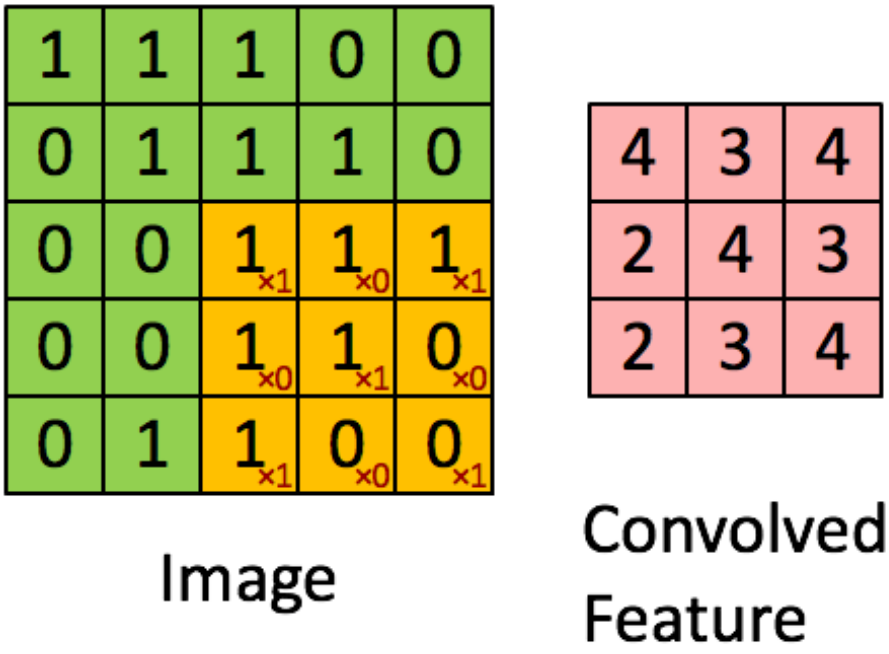


Figure 4.8: Convolution

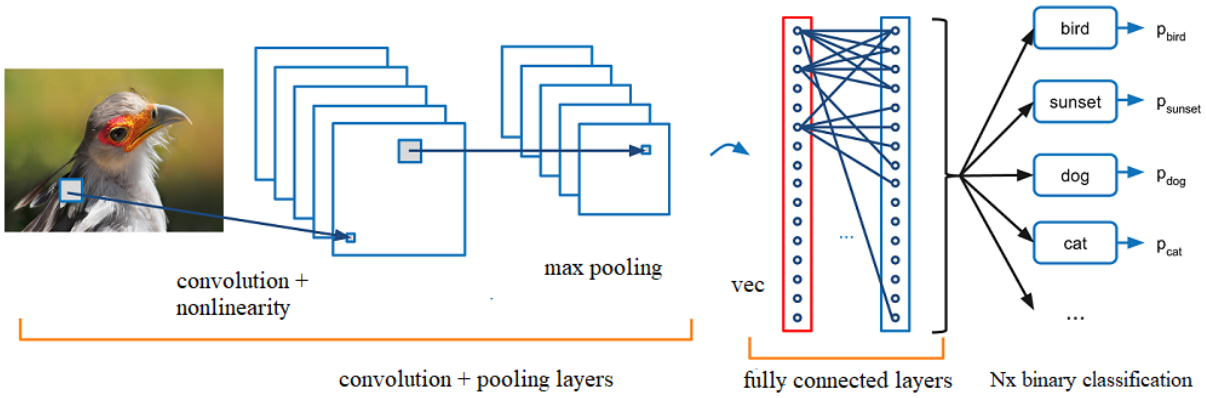


Figure 4.9: Convolution Neural Network Model

4.6 Classification and Recognition

4.6.1 Convolution Neural Network

Introduction

Convolution neural networks is similar to combination of mathematics and biology but these networks are variety of the foremost powerful innovations inside the sphere of laptop vision. 2012 was the first year that neural nets grew to prominence as Alex Krizhevsky used them to win that years ImageNet competition (basically, the annual contest of laptop vision), dropping the classification error record from twenty six p.c to fifteen p.c, associate astounding improvement at the time.Many big organisations started using deep learning since then. Facebook uses neural networks for automatic tagging,

Google uses it for searching images, Amazon for suggestions.Mostly these networks are used for image processing and object recognition from images

Convolutional neural networks are deep artificial neural networks that ar used primarily to classify pictures (e.g. name what they see), cluster them by similarity (photo search), and perform beholding inside frames. they're algorithms that may acknowledge faces, people, street signs, tumors, platypuses and plenty of alternative aspects of visual information. Convolutional networks perform optical character recognition (OCR) to modify text and build natural-language process attainable on analog and hand-written documents, wherever the pictures ar symbols to be transcribed. Convolution Neural Networks can even be applied to sound once it's diagrammatically visually as a photo. a lot of recently, convolutional networks are applied on to text analytics yet as graph information with graph convolutional networks. The effectiveness of convolutional networks in image recognition is one amongst the most reasons why we tend to are endeavor to extend the effectiveness of deep learning. they're powering major advances in laptop vision (CV), that incorporates a very important application for self-driving cars, robotics, drones, security, medical diagnoses, and coverings for the visually impaired.

When endeavour deep learning tasks, it's typically suggested to coach and run the models on raw inputs, without manually extracting any options before. the explanation for this can be that a network trained on the raw input might learn to extract these options on its own, however in distinction to operating with pre-built options, it'd even be able to additional optimize the feature extraction because the network improves. If the input is an image, it would, therefore, be fascinating to figure directly with its raw picture element values. Since a picture consists of the many pixels and every picture element is probably diagrammatically by multiple color values, the illustration of that image within the input layer can become extremely complicated. A full HD RGB image with 19201080 pixels would, as an example, need Associate in Nursing input layer consisting of concerning six million neurons. If one would use the simple fully-connected spec, every somatic cell within the succeeding layer would then be connected to concerning six million neurons and if the primary fully-connected layer would contain simply one thousand neurons, the total range of parameters would quantity to over six billion. Since the network needs to optimize all of those parameters, the coaching method might then become terribly time and storage intensive.

In order to unravel this procedure drawback, a distinct kind of spec is employed, known as Convolutional Neural Network (CNN). CNNs are specifically designed for working with pictures. For this reason, the neurons of a layer are organized across the 3 dimensions, height, width and depth, rather like the pixels in a picture wherever the depth dimension would differentiate the various color values. In addition thereto, CNNs introduce 2 new styles of hidden layers during which every somatic cell is just connected to a tiny low subset of the neurons within the previous layer, to stop the aforementioned drawback.

Convolutional Layers square measure employed in CNNs instead. What separates a convolutional layer from a fully-connected one is that each vegetative cell is simply connected to a tiny low, native set of the neurons within the previous layer, that could be a

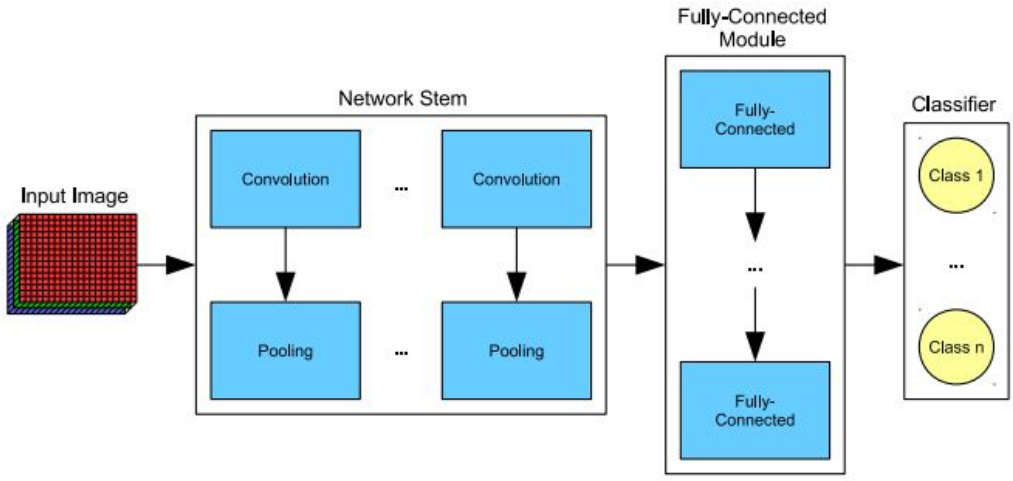


Figure 4.10: Convolution Model

sq. sized region across the peak and breadth dimensions. the scale of this sq. is a hyper-parameter named Receptive Field. For the depth dimension, there's no hyperparameter that needs to be outlined, as the convolutions square measure by default continuously performed across the whole depth. the rationale for this can be that the depth dimension of the input will generally outline the various colours of the image and it's typically necessary to mix them so as to extract any helpful info.

Neurons of the convolution operator will acknowledge sure local patterns of the previous layers output. Since the patterns that square measure recognized ought to be freelance of their position in the image, all neurons are going to be forced to acknowledge the same pattern by creating all of them share one single set of parameters. this idea is observed as Parameter Sharing. In order to currently acknowledge multiple completely different options among one layer, it's needed to possess many Filters, where each filter could be a cluster of neurons that acknowledge an explicit pattern at completely different locations within the image. within the convolutional layer, the depth dimension is then specifying to that filter a given neuron belongs.

Problem Space

Image classification is the process of extracting the features from input image that best describes the image. For humans recognition is one of the primary skill we have a tendency to learn from the instant we have a tendency to square measure born and comes naturally and effortlessly. while not even thinking double, we have a tendency tore able to quickly and seamlessly establish the surroundings we square measure in similarly because the objects that surround North American nation. After we see a picture or simply after we cross-check the globe around North American nation, most of the time we have a tendency to square measure able to like a shot characterize the scene and provides every object a tag, all while not even consciously noticing. These skills of having the ability to acknowledge patterns quickly, generalize from previous information, and adapt to completely different image environments square measure ones that we have a tendency to don't share with

our fellow machines.

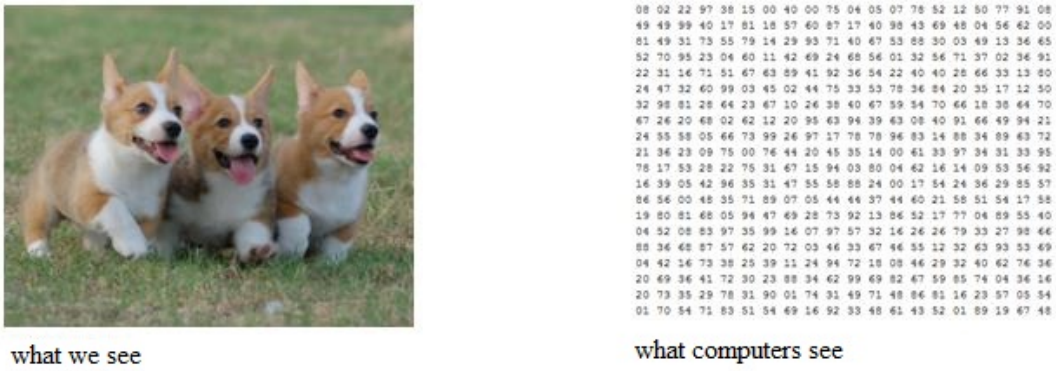


Figure 4.11: Problem Space

Inputs and Outputs

When a computer take an input an image, it will check associate degree array of component values. reckoning on image size and resolution, it'll see a thirty two by thirty two by three array of numbers (RGB values). Simply to drive home the purpose, to Illustrate we've a color image in JPG kind and its size is four-eighty by four-eighty. The representative array are four-eighty by four-eighty by three. every of those numbers is given a price from zero to 255 that defines the component intensity. These numbers, whereas nonsense to USA once we perform image classification, square measure the sole inputs accessible to the pc. the idea is that we tend to provide the pc these array of numbers and it will output numbers that describe image being a precise category (0.80 for dog, 0.15 for cat).

Expected Computer Performance

Now that we all know the matter additionally because the inputs and outputs, lets learn the way to handle this. What we would like the pc to try to to is to be ready to differentiate between all the pictures its given and puzzle out the distinctive options that build a dog a dog or that build a cat a cat. this can be the method that goes on in our minds subconsciously additionally. once we explore an image of a dog, we are able to classify it intrinsically if the image has ac knowledgeable options like paws or four legs. during a similar method, the pc is in a position perform image classification by searching for low level options like edges and curves, and so build up to a lot of abstract ideas through a series of convolutional layers. this can be a general summary of what a CNN will.

Biological Connection

When we 1st detected of the term convolutional neural networks, we tend to could have thought of one thing associated with neurobiology and that we would be correct to some extent. Visually CNNs take biological inspiration.The visual cortex—visual square measurea—cortical area—cortical region has tiny regions of cells that are sensitive to

specific regions of the visual view. this concept was distended upon by a desirable experiment by neuroscientist and historiographer in 1962 wherever they proved that some vegetative cell cells within the brain fired solely within the presence of edges of a precise orientation. for instance, some neurons pink-slipped once exposed to vertical edges and a few once shown horizontal or diagonal edges. neuroscientist and historiographer detected that each one of those neurons were organized during a columnar design which along, they were ready to manufacture beholding. this concept of specialised elements within a system having specific tasks (the vegetative cell cells within the visual area searching for specific characteristics) is one that machines use additionally, and is that the basis behind CNNs.

Structure

A a lot of elaborate summary of what CNNs do would be that we tend to take the image, pass it through a series of convolutional, nonlinear, pooling (downsampling), and absolutely connected layers, and acquire associate output. As we know, the output are often one category or a chance of categories that best describes the image. Now, the onerous half is knowing what every of those layers do.

First Layer - Mathematical Part

The first layer in a very CNN is often a Convolutional Layer. very first thing to form certain we tend to bear in mind is what the input to the current convolution layer is. Like we all know, the input may be a thirty two x thirty two x three array of component values. Now, the most effective thanks to justify a convolution layer is to assume an electric lamp that's lighting over the highest left side of the image. Lets say that the sunshine this electric lamp light covers a five by five space.Now, let us assume this electric lamp slippy across everywhere in the input image. In terms of machine learning concept, this electric lamp is termed a filter and therefore the area that it is lighting over is termed as receptive field. currently this filter is additionally associate array of numbers (the numbers square measure referred to as weights or parameters). mathematical terms convolution is perform derived by integration the impact of 1 perform on the form of different perform. just in case of image process input image is one such perform and therefore the applied mask or feature detector is another. we tend to apply this mask or feature detector on our input image to cut back the dimensions of image by removing redundant component values in image. This square measure the values that dont satisfy the conditions of mask. The method convolution works is we tend to place our filter mask on high left corner of our input image and acquire a price of the a part of image that is beneath filter. Once we tend to get price for that a part of image we tend to slide the filter one component to the correct ,we can decision this shift one component as a stride of 1 component. but in actual image process a stride of 1 component isn't sensible because it can lead to a way larger feature map. therefore a way larger stride is employed for correct pictures.

It is true that we tend to lose info when convolution. The feature map that we tend to find yourself with has fewer cells and thus less info than the first input image. However, the terribly purpose of the feature detector is to sift through the knowledge within the input image and filter the components that square measure integral thereto and exclude the remainder. think about it this fashion. What we tend to do is notice sure options,

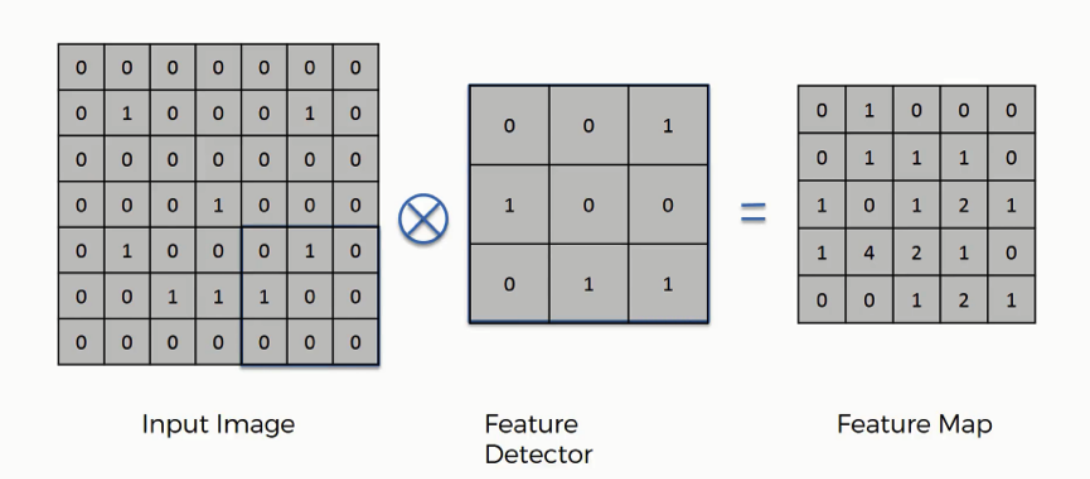


Figure 4.12: First Layers

say, their eyes and their nose, for example, and you straightaway apprehend WHO you're watching. These square measure the foremost revealing options, which is all of your brain must see so as to form its conclusion. Even these options square measure seen generally and not all the way down to their trivia. If your brain really had to method equally of knowledge that enters through your senses at any given moment, you'd 1st be unable to require any actions, and shortly you'd have a mental breakdown. Broad categorization happens to be a lot of sensible. Convolutional neural networks operate in barely an equivalent method. precisely like our we tend to brain we tend to derive the distinguished options of the image necessary for recognition and drop out the inessential component values of the image. This particle flip reduces the image size creating it easier for successive layer to extract the options to a a lot of minute and well outlined level.

A very necessary note is that the depth of this filter has got to be identical because the depth of the input (this makes positive that the maths works out), therefore the dimensions of this filter is five by five by three. Now, lets take the primary position the filter is certain example. it might be the highest left corner. because the filter is slippy, or convolving, round the input image, it's multiplying the values within the filter with the initial component—component—constituent—element values of the image. These multiplied square measure all added together. Thus currently we've got one range. This range is simply representation of once the filter is at the highest left of the image. Now, we tend to repeat this method for each location on the input volume. (Next step would be moving the filter to the proper by one unit, then right once more by one, and so on). each distinctive location on the input volume produces variety. once slippy the filter over all the locations, we will—we'll—we square measure going to resolve that what we tend to are left with may be a twenty eight by twenty eight by one array of numbers, that we tend to decision associate degree activation map or feature map. the rationale we tend to get a twenty eight by twenty eight array is that there square measure 784 completely different locations that a five by five filter will work on a thirty two by thirty two input image. These numbers square measure mapped to a twenty eight by twenty eight array. Lets say currently we tend to use 2 five by five by three filters rather than one. Then the output volume will be twenty eight by twenty eight by two. By exploitation additional filters, we tend to square measure able to preserve the abstraction dimensions higher.

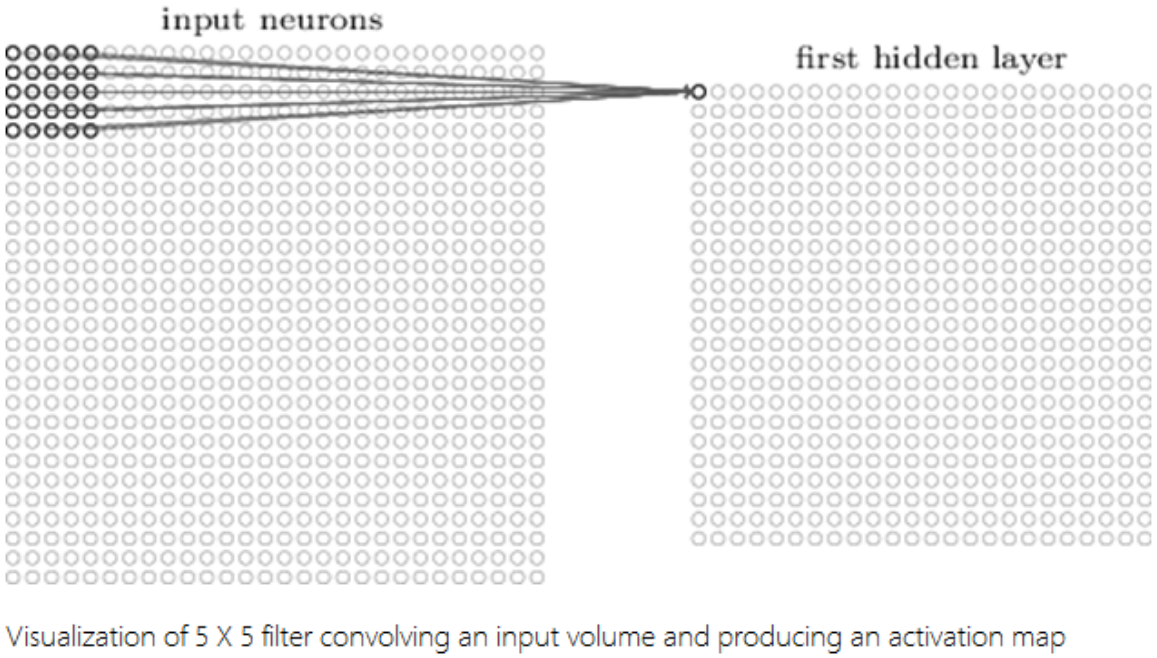


Figure 4.13: Activation Map

Mathematically, this is often whats occurring in an exceedingly convolutional layer.

First Layer High Level Perspective

Lets understand what convolution is performing at high level. Every of those filters are often understood of as feature identifiers. once we say options, we tend to area unit talking concerning things like straight edges, easy colours, and curves. deem the best characteristics that every one pictures have in common with one another. Lets say our 1st filter is seven x seven x three and goes to be a curve detector. (In this section, lets ignore the very fact that the filter is three units deep and solely take into account the highest depth slice of the filter and therefore the image, for simplicity.)As a curve detector, the filter can have a element structure during which there will be higher numerical values on the world that's a form of a curve (Remember, these filters that were talking concerning as simply numbers!).

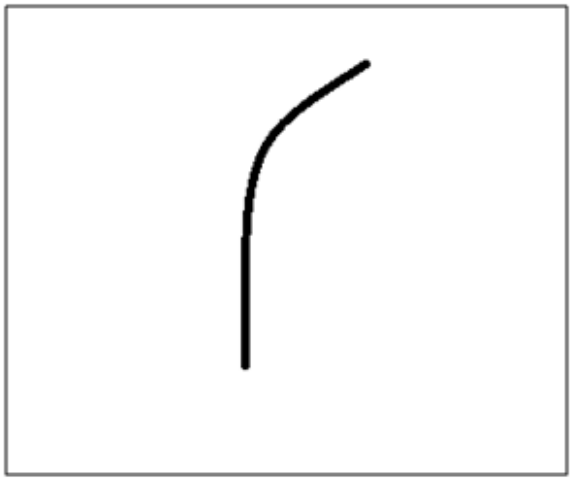
Now, lets return to visualizing this mathematically. once we have this filter at the highest left corner of the input volume, it's computing multiplications between the filter and element values at that region. currently lets take Associate in Nursing example of a picture that we would like to classify, and lets place our filter at the highest left corner.

Remember, what we have got to try and do is multiply the values within the filter with the first element values of the image.

Fundamentally, within the input image, if there's a form that typically looks similar to the curve that this filter is representing, then all of the multiplications summed along can end in an oversized value! currently lets see what happens once we move our filter.

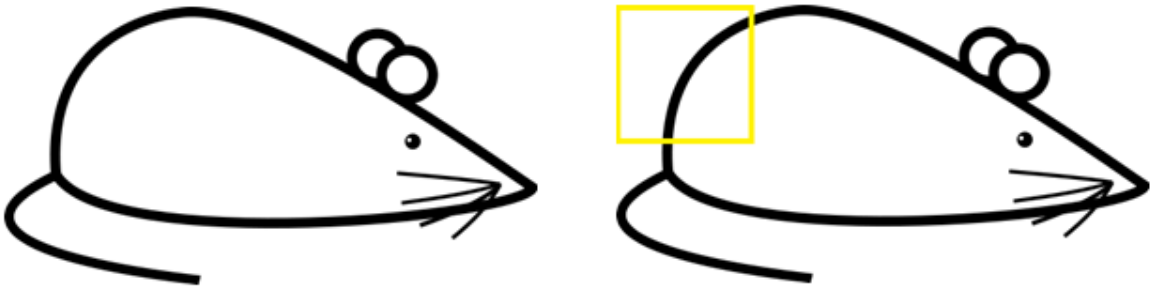
0	0	0	0	0	30	0
0	0	0	0	30	0	0
0	0	0	30	0	0	0
0	0	0	30	0	0	0
0	0	0	30	0	0	0
0	0	0	30	0	0	0
0	0	0	0	0	0	0

Pixel representation of filter



Visulaization of curve detector filter

Figure 4.14: Filter



Original Image

Visualization of filter on the image

Figure 4.15: Original Image And Filter

The value is far lower! this is often as a result of there was not something within the image section that more matured the curve detector filter. Remember, the output of this convolution layer is Associate in Nursing activation map. So, within the straightforward case of a 1 filter convolution (and if that filter could be a curve detector), the activation map can show the areas within which there at principally doubtless to be curves within the image. during this example, the highest left worth of our twenty six x twenty six x one activation map (26 due to the 7x7 filter rather than 5x5) are going to be 6600. This high worth means it's doubtless that there's some kind of curve within the input volume that caused the filter to activate. the highest right worth in our activation map are going to be zero as a result of there was not something within the input volume that caused the filter to activate (or a lot of merely aforementioned, there was not a curve in this region of the initial image). Remember, this is often only for one filter. this is often simply a filter

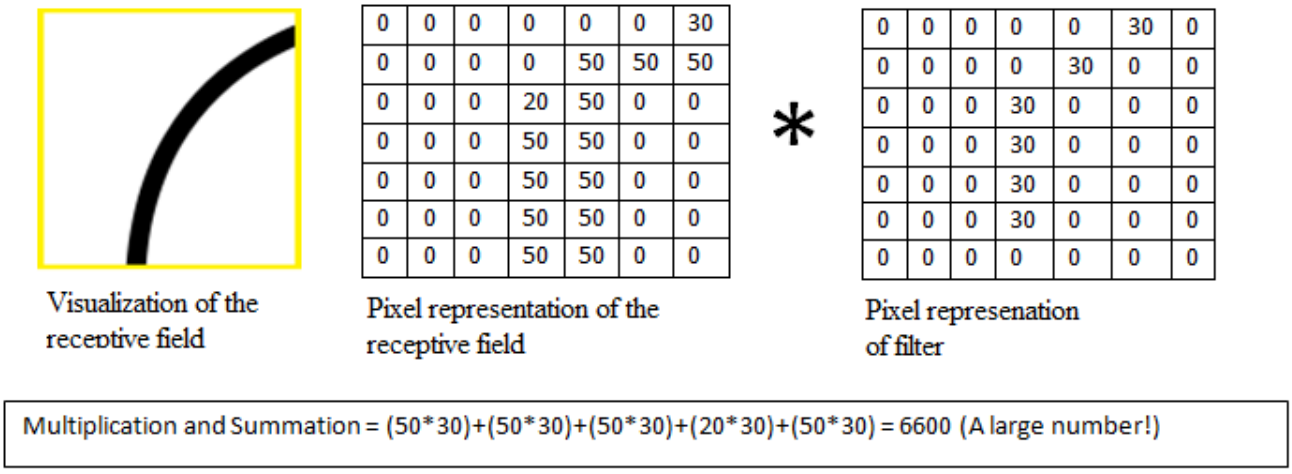


Figure 4.16: First Pixel Mulitiplication

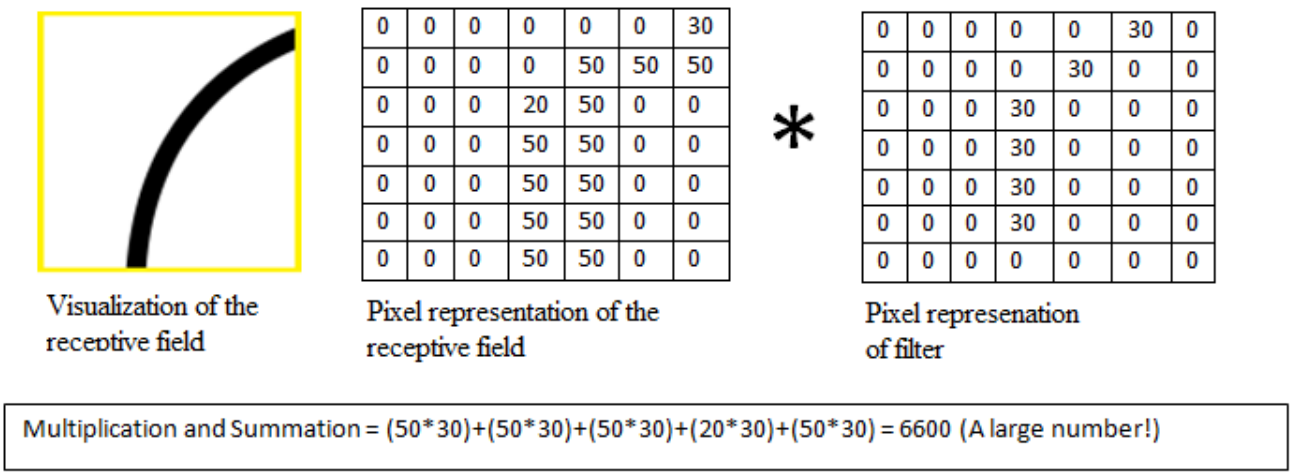


Figure 4.17: Second Mulitiplication

that's getting to notice lines that curve outward and to the proper. we are able to produce other filters for lines that curve to the left or for straight edges. The a lot of filters, the larger the depth of the activation map, and also the a lot of data we've concerning the input volume.

Going in Depth of Network

Now during a ancient convolutional neural specification, there area unit alternative layers that area unit interspersed between these convolution layers. during a general sense, they supply non-linearities and conservation of dimension that facilitate to enhance the lustiness of the network and management overfitting.

The last layer, however, is a crucial one and one that we are going to go in soon. Lets simply review what we have learned so far. we tend to knew concerning what the filters

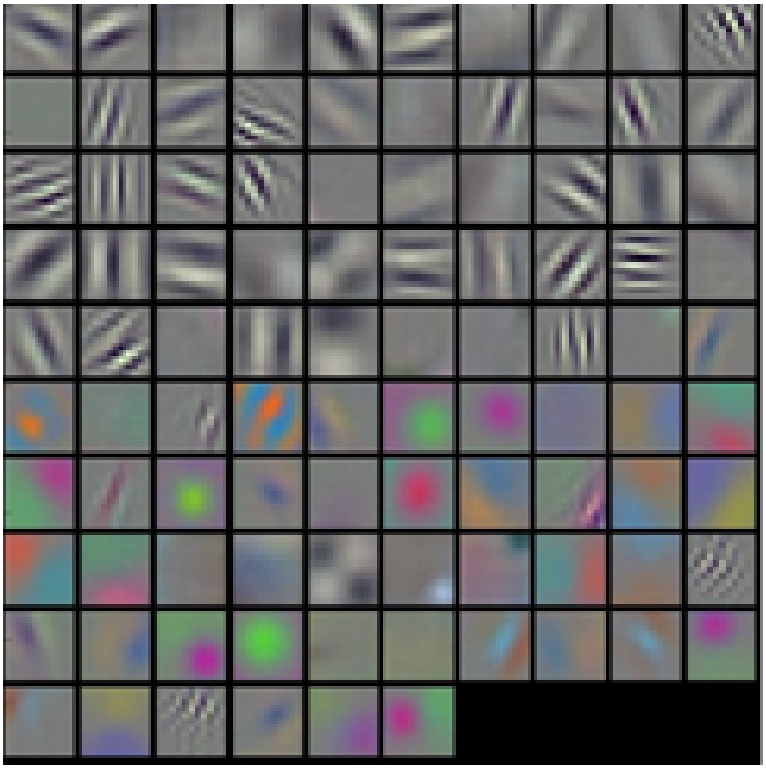


Figure 4.18: First Layers

Input -> Conv -> ReLU -> Conv -> ReLU -> Pool -> ReLU -> Conv -> ReLU -> Pool ->Fully Connected

Figure 4.19: CNN TABLE

within the 1st convolution layer area unit designed to observe. They observe low level options like curves and edges. We would assume, so as to find whether or not a picture could be a style of an object, we would like the network to be ready to acknowledge high level options like ears or paws. Thus lets have confidence what the output will be once the primary convolution layer. It would be a twenty eight by twenty eight by three volume (if we are using 3 five by five by three filters). once we bear another convolution layer, the output of the primary convolution layer becomes the input of the second convolution layer. Now, this can be a touch bit more durable to examine. However, once were concerning the second convolution layer, the input is that the activation maps that come from the primary layer. Thus every input layer is essentially explaining the location within the input image for wherever sure low level options seem. currently once we apply a group of filters on high of that, the output are activation that represent higher level options. styles of these options may be semicircles (combination of a curve and straight edge) or squares (Straight edges combination). As we tend to bear the network and bear additional convolution layers, we tend to get activation maps that represent additional and additional complicated options. By the tip of the network, we tend to might have some filters that

activate once there's handwriting within the image, filters that activate after they see pink objects, etc.

Rectified Linear Units (ReLU)

Activation operate is principally accustomed succeed 2 purposes:1)It helps our model to investigate and adapt to the interaction effects. This result will be explained as a amount A whose prediction is established by another amount B. we will take example of a model developed for predicting the result of sure body weights on accumulated risk of polygenic disorder , however during this case considering solely weight for predicting accumulated risk of polygenic disorder isn't a adequate parameter as a particular weight may be unhealthy for someone with smaller height however same weight will be healthy weight for a taller person. Hence, to predict the result of we tend on polygenic disorder we conjointly ought to take into account of that person associated and that we would say that weight and height have an interaction result. 2) facilitate a model account for non-linear effects. This simply means if I graph a variable on the horizontal axis, and my predictions on the vertical axis, it's not a line. Or same in a different way, the result of skyrocketing the predictor by one is completely different—completely different at different values of that predictor.

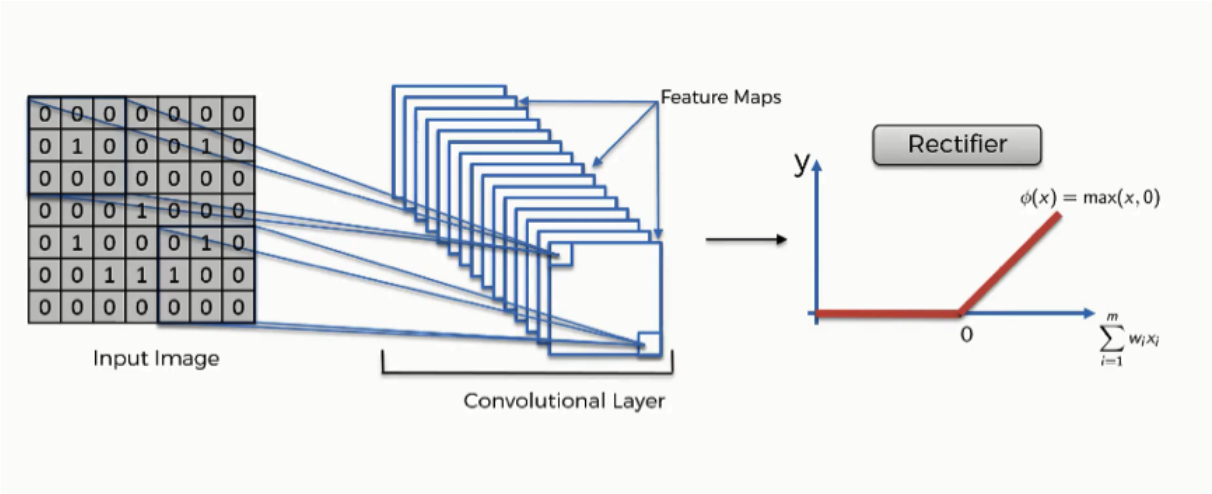


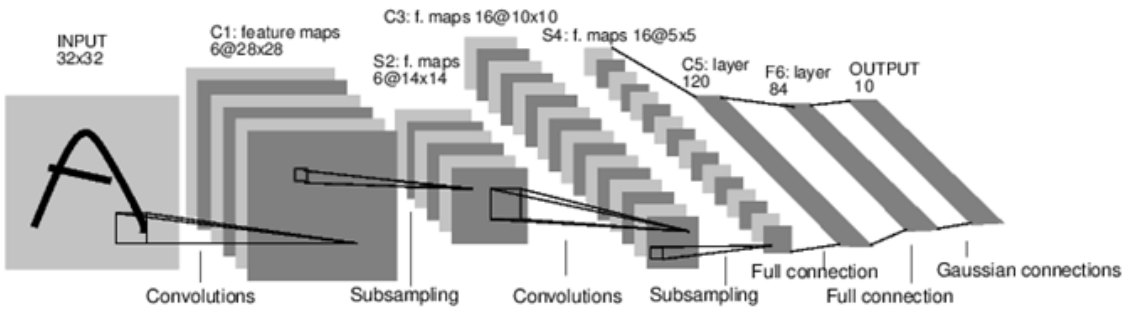
Figure 4.20: ReLU

We apply rectifier to the image so as to extend the non dimensionality in it. the explanation behind introducing non dimensionality is that pictures area unit naturally non linear. One will notice several non linear options. The rectifier serves to interrupt up the dimensionality even additional so as to create up for the dimensionality that we would impose a picture after we place it through the convolution operation. to examine however that really plays out, we will look into the subsequent image and see the changes that happen to that because it undergoes the convolution operation followed by rectification. The essential distinction between the non-rectified version of the image and also the corrected one is that the progression of colours. If you look closely at the primary one, you may notice elements wherever a white streak is followed by a gray one then a black one. a way ReLUs improve neural networks is by dashing up coaching. The gradient computation is extremely straightforward (either zero or one betting on the sign of xx).

Also, the process step of a ReLU is easy: any negative components area unit set to zero.0 – no exponentials, no multiplication or division operations.Gradients of logistical and hyperbolic tangent networks area unit smaller than the positive portion of the ReLU. this implies that the positive portion is updated earlier as coaching progresses. However, this comes at a price. The zero gradient on the left-hand facet is has its own downside, referred to as "dead neurons," within which a gradient update sets the incoming values to a ReLU specified the output is often zero; changed ReLU units like ELU (or Leaky ReLU, or PReLU, etc.) will ameliorate this.

Completely Connected Layer

Now we will observe these greater level options, the cake icing is attached to completely connected layer to tip of the network. This layer essentially takes associate degree input volume associate degree outputs an N dimension vector wherever N is that the variety of categories that the machine needs to select between.For instance, if we have a tendency to needed a number classification, N would be ten since there area unit ten digits. Every variety during this N dimensional vector represents the chance of a particular category. For instance, if the ensuing vector for a digit classification program is [0 0.1 0.1 0.75 0 0 0 0 0 0.05], then it shows one of ten p.c chance that the image could be a 1, a ten p.c chance that the image could be a a pair of, a seventy five p.c chance that the image could be a three, and a five p.c chance that the image could be a nine (There area unit different ways in which we will represent the output, however we have a tendency to area unit victimisation the softmax approach). The approach this absolutely connected layer works is that it's at the output of the previous layer (which as we have a tendency to keep in mind ought to represent the activation maps of high level options) and determines that features most correlate to a specific category. For instance, if the machine is predicting that an image could be a cat, it has high values in the activation maps which represents high level options sort of a paw or four legs, etc. Basically, a Fully connected layer appearance at what high level options most powerfully correlate to specific—a specific—a selected category and has particular weights in order that once we work out the merchandise between the weights and therefore the previous layer, we have a tendency to get the proper possibilities for the various categories.



A full convolution neural network (LeNet)

Figure 4.21: LeNet

4.7 Training and Recognition

The seventh OCR element is coaching and recognition. OCR systems extensively use the methodologies of pattern recognition that assigns Associate in Nursing unknown sample into a predefined category. The OCR square measure investigated in four general approaches of pattern recognition as steered in (a) guide matching (b) applied math techniques (c) structural techniques and (d) ANNs. These approaches square measure neither essentially freelance nor disjointed from one another.

The means the pc is in a position to regulate its filter values (or weights) is thru a coaching method referred to as back propagation.Before we have a tendency to get into back propagation, we have a tendency to should go behind and figure what neural network desires so as to figure. At the instant we have a tendency to all were born, our minds were recent. we have a tendency to didnt grasp what a bird,cat or dog was.In a very similar type of means, before the CNN begins, the filter values square measure is randomised. The filters dont grasp to appear for edges and curves. The filters within the higher layers dont grasp to appear for beaks and paws. As we have a tendency to grew older but, our oldsters and academics showed USA totally different footage and pictures and gave USA an appropriate label. The concept of being given an image and a label is that the coaching method that CNNs undergo. Before obtaining too into it, lets simply say that we've got a coaching set that has hundreds of pictures of cats,birds and dogs and every picture—the photographs includes a label of the picture of the animal.

So back propagation will be divided into four distinct sections, the passing game, the loss operate, the backward pass, and also the weight update. throughout the passing game, we have a tendency to take a coaching image that as we have a tendency to bear in mind could be a thirty two x thirty two x three array of numbers and pass it to the network. On our 1st coaching example, since all of the weights or filter values were willy-nilly initialized, the output can most likely be one thing like [0.1 0.1 0.1 0.1 0.1 0.1 0.1 0.1 0.1 0.1], essentially Associate in Nursing output that doesnt offer preference to any range specifically. The network, with its current weights, isnt ready to seek for those low level options or so isnt ready to create any cheap conclusion concerning what the classification may be. This goes to the loss operate a part of back propagation. bear in mind that what we have a tendency to square measure victimisation straight away is coaching information. This information has each a picture and a label. Lets say as an example that the primary coaching image inputted was a three. The label for the image would be [0 zero zero one zero zero zero zero zero 0]. A loss operate will be outlined in many alternative ways that however a typical one is MSE (mean square error), that is 0.5 times (actual - predicted) square.

$$E_{total} = \sum \frac{1}{2}(target - output)^2 \tag{4.1}$$

Lets say the variable L is capable that price. As we are able to imagine, the loss are extraordinarily high for the primary few coaching pictures. Now, lets simply deem this intuitively. we would like to induce to some extent wherever the anticipated label (output of the ConvNet) is that the same because the coaching label (This implies that our network got its prediction right).In order to induce there, we would like to attenuate the quantity

of loss we've got. Visualizing this as simply Associate in Nursing optimisation drawback in calculus, we would like to search out out that inputs (weights in our case) most directly contributed to the loss (or error) of the network.

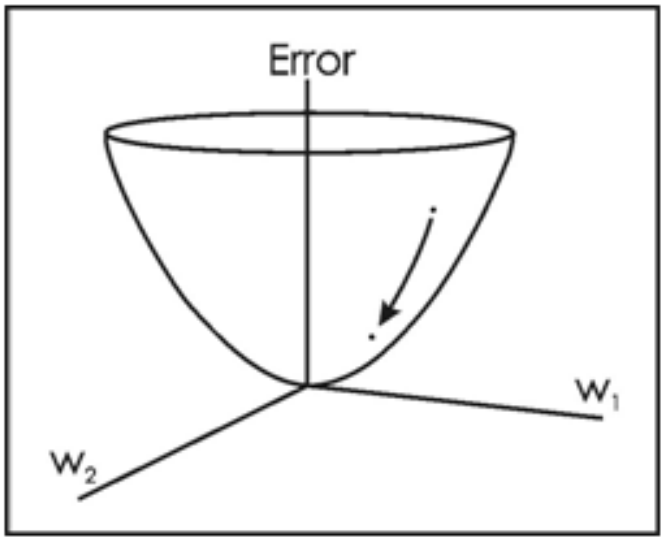


Figure 4.22: Loss

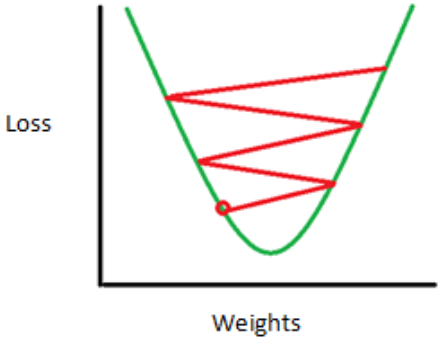
This can be represented mathematically as dL/dW. Now we try to do a backward experience of the network, that is determinant that weights supply most to the loss and finding solutions to regulate them in order such that loss decreases. Once we have a tendency to calculate this by-product, we have a tendency to then move to the last step that is that the weight update. this is often wherever we have a tendency to collect the weights of filters and modify them such that they alter within the other way of the gradient.

$$w = w_i - \eta \frac{dL}{dW} \tag{4.2}$$

w = weight
$w_i = initialweight$
$\eta = learningrate$

The learning rate may be a parameter that's chosen by the engineer. A high learning rate means larger steps square measure taken within the weight updates and so, it should take less time for the model to converge on Associate in Nursing best set of weights. However, a learning rate that's too high may end in jumps that square measure overlarge and not precise enough to succeed in the best purpose.

The process of passing game, loss perform, backward pass, and parameter update is one coaching iteration. The program can repeat this method for a hard and fast variety of iterations for every set of coaching pictures (commonly referred to as a batch). Once



Consequence of a high learning rate whre the jumps are too large and we are not able to minimize the loss.

Figure 4.23: HighLR

we have a tendency to end the attribute change on last coaching example, the network tends to be trained tolerably such that the weights of layers square measure properly.

Examining

In order visualize whether our CNN works, we've got a special set of pictures and tags and pass the pictures through the CNN. we have a tendency to compare the outcome to the bottom truth and check if the network is working or not.

4.8 Post-processing

The eighth OCR element is post-processing.A number of the unremarkable used post-processing activities embody grouping and error detection and correction. In rouping symbols in text square measure related to strings. The results of plain image recognition in text could be a set of individual symbols. However, these symbols don't typically contain enough data. These individual symbols square measure related to one another creating up words and numbers. The grouping of symbols into strings relies on symbols location in document. The symbols that square measure sufficiently shut square measure sorted along. For fonts with fastened pitch grouping method is straightforward as position of every character is thought. For print characters distance between characters square measure variable. the space between words square measure considerably giant than distance between characters and grouping is so doable. the issues occur for written characters once text is skew. till grouping every character is treated one by one, the context within which every character seems has not been exploited. However, in advanced optical text recognition issues, system consisting solely of single character recognition isn't comfortable.

Chapter 5

Result

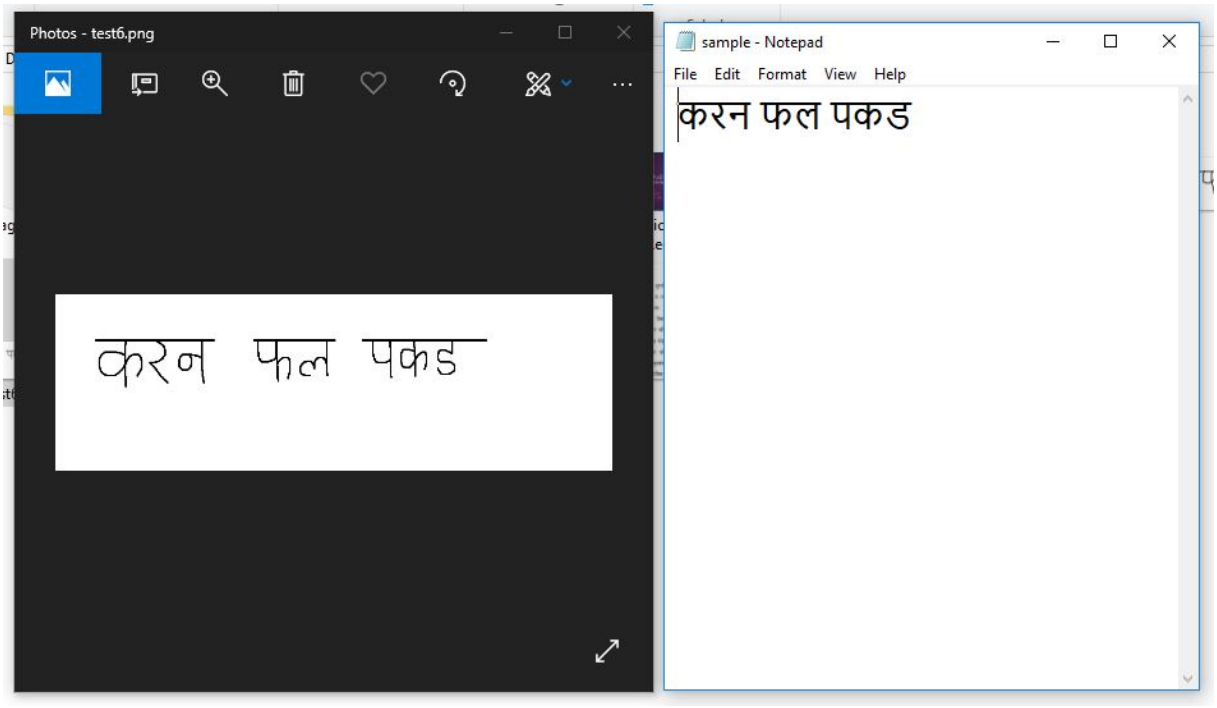


Figure 5.1: Result

5.1 Cost Analysis

Our project "Optical Character recognition of Devanagari script" is a software based project which uses Matlab simulation and a camera or any device which can capture images for image pre-processing and segmentation. For the next stage, we used Python IDE for training and recognition. The cost of our project is minimal as Matlab and Python IDE is readily available. We have implemented our project on Matlab software which was provided by the college and python IDE is free to download through the internet, so our project is completely free of cost.

Chapter 6

Future Scope

There are twenty two official languages in India that are written in thirteen completely different scripts and having a broad variance of 720 dialects. however out of all this languages 'Hindi' is that the most generally spoken and additionally national language in India. it's a language that springs from the oldest language i.e Sanskrit. The script used for Hindi and Indic is Nagari. the first motive for developing OCR for Nagari script is to make editable soft copies of documents written or written in devanagari script.

We have a really wealthy history of ancient literature chemical analysis back to around 1800 years B.C. So as to push this literature we want to form it additional universally clear.This ancient written literature has to be digitized and translated to alternative wide spoken languages for this purpose we are able to effectively use OCR. In an exceedingly country wherever we've twenty two official languages it's unreasonable to expect everyone to talk and perceive each single language. For this we have a tendency to do have Hindi as our national language then again we've around 720 completely different dialects creating it virtually not possible to own a 1 single language for communication. to get rid of this barrier of communication we are able to develop Bilingual or trilingual converter mistreatment OCR.

Once an entire OCR has been developed for 2 languages with font cryptography, spell checker and grammatical sentence check, then a converter may be enforced to convert sentences from one language to another through a written text and translation theme. Inspired from the modern setting we are able to develop a speech recognition application. The digital copy generated mistreatment OCR we are able to be recorded and by employing a voice synthesizer a speech output is generated. this is able to facilitate the blind to send and receive data.Similar to the previous application it's attainable to convert a speech to text too.

Chapter 7

Conclusion

Character recognition techniques associate a symbolic identity with the image of character.In this project an summary of assorted techniques of OCR has been conferred.An OCR includes varied phases like acquisition, pre-processing, segmentation, feature extraction, classification and post-processing.Each of the steps is mentioned intimately in this project.Using a combination of those techniques, an efficient OCR system is developed as a future work.

The OCR system also can be utilized in totally different sensible applications like number-plate recognition, smart libraries and varied different time period applications.The implementation of Optical character recognition technology is with efficiency wont to speed up translation of image primarily based documents that are presently straightforward to get, search and method.

Bibliography

1. B. Kondalarao, S. Sahoo and D. Pratihar, "Character Recognition Using Entropy Based Fuzzy c-means Clustering", In Proc. of Hybrid Intelligence for Image Analysis and Understanding,pp. 25-45,2017.
2. A. Choudhary, R. Rishi and S. Ahlawat, "Offline handwritten character recognition using features extracted from binarization technique", Elsevier AASRI Conference on Intelligent Systems and Control Procedia, vol. 4,pp. 306-312, 2013.
3. M. Gupta, N. Jacobson and E. Garcia, "OCR binarization and image preprocessing for searching historical documents", Elsevier Pattern Recognition, vol. 17,no. 3,pp. 1-6,2015.
4. S. Mori et. al, "Historical Review of OCR Research and Development", Proceeding IEEE, 80,vol. no 7, pp. 1029-1058, July 1992.
5. Veena Bansal and R.M.K. Sinha, "A Devnagari OCR and A Brief Overview of OCR for Indian Script ", PROC Symposium on Transaction support System (STRANS 2001), Feb. 15-17, 2001, Kanpur, India.
6. R. Jayadevan, Satish R. Kolhe, Pradeep M. Patil, and Umapada Pal "Offline Recognition of Devanagari Script: A Survey ", ieee transactions on systems, man, and cyberneticspart c: applications and reviews, vol. 41, no. 6, november 2011.
7. Mohanad Alata Mohammad Al-Shabi, "TEXT DETECTION AND CHARACTER RECOGNITION USING FUZZY IMAGE PROCESSING ", Journal of electrical engineering, VOL. 57, NO. 5, 2006, 258267
8. A. K. Pant, S. P. Panday, and S. R. Joshi, "Off-line nepali handwritten character recognition using multilayer perceptron and radial basis function neural networks, "in Internet (AH-ICI), 2012 Third Asian Himalayas International Conference on. IEEE, 2012, pp. 15.
9. H. Larochelle, Y. Bengio, J. Louradour, and P. Lamblin, "Exploring strategies for training deep neural networks,"The Journal of Machine Learning Research, vol. 10, pp. 140, 2009.
10. N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overtting, "The Journal of Machine Learning Research, vol. 15, no. 1, pp. 19291958, 2014.