

IMAGE STEGANOGRAPHY

USING MACHINE LEARNING

PAPER NAME: PROJECT 2 PAPER CODE: CS 892

ABHINABA PODDAR • ABHISEK DAS • ABHISHEK CHAKRABORTY •

ANSHUMAN SWAROOP DAS

Under the Supervision of Prof. PRATAP CHANDRA MANDAL

OUR OBJECTIVE

Building a tool
 that can can be
 used to hide a
 text in an image



REQUIREMENTS

System Requirements

- ► CPU: Dual core 64-bit 2.8 GHz 8.00GT/s CPUs.
- ► RAM: 2GB RAM (recommended 4GB RAM).
- ► Storage: 2GB for installation of Anaconda Navigator.

Software Requirements

- ► 0S: Windows 10, 8, 7.
- ► Anaconda Navigator v3.6.4 or higher.
- PyCharm and Jupyter Notebook IDEs.
- ► Any Browser like Google Chrome or Firefox.







TECHNOLOGY STACK

FRONT END >





BACK END

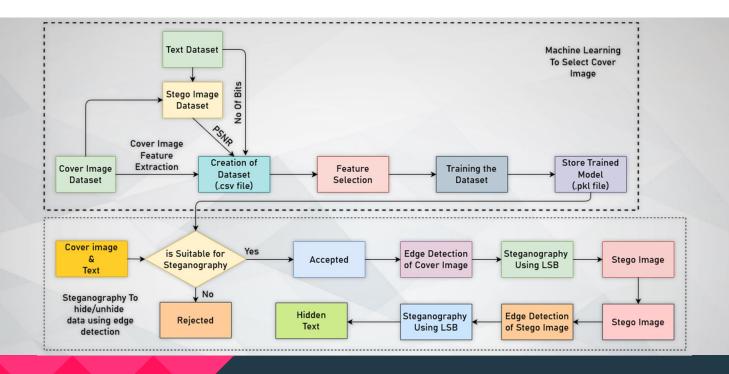








matpletlib

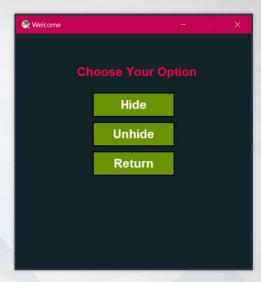


SCHEMATIC DIAGRAM

of our working model

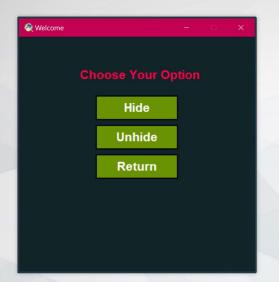


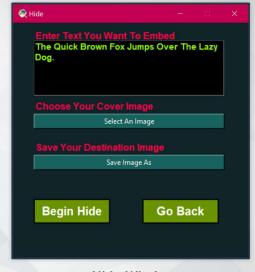




Options Window

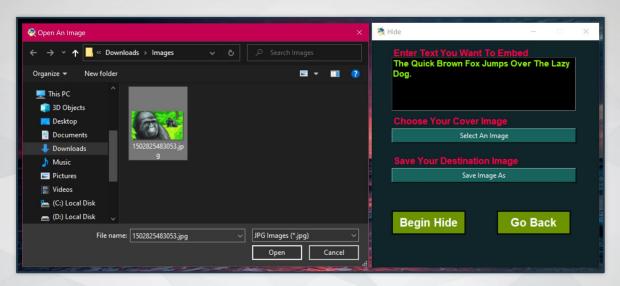




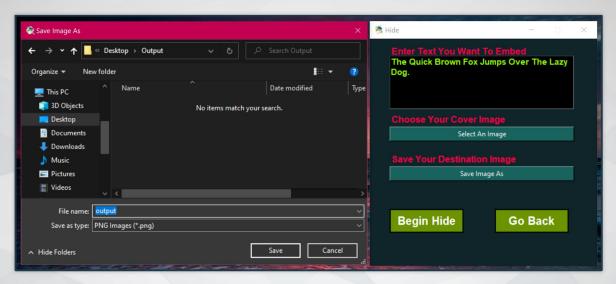


Options Window

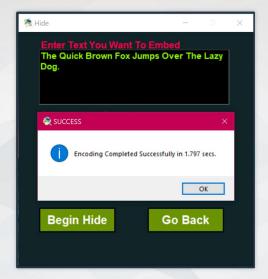
Hide Window



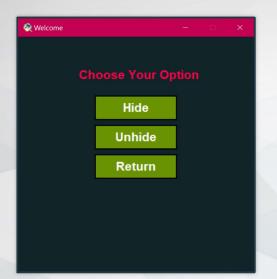
Choosing an Image for Hiding the Text

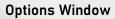


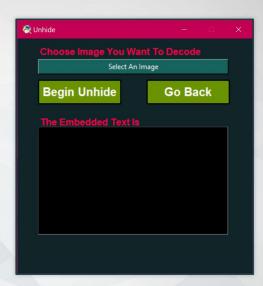
Choosing the path for saving the Image with hidden Text



After Successful Hiding of Text in the Image

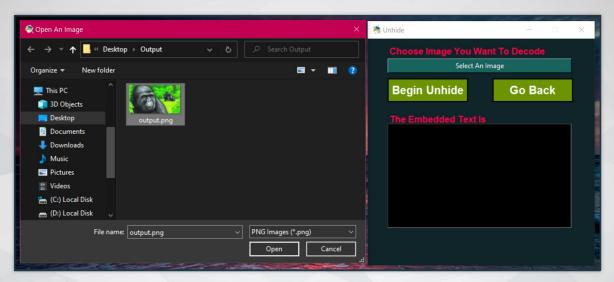




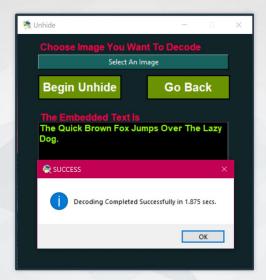


Unhide Window





Choosing an Image with a Text Embedded



After Successful Unhiding of Text in the Image

OUR DATASET

of cover images



Samples of Kaggle Steganography Dataset



Samples of TopLogo-10 Dataset

TEXT FILE

samples

Text3.txt - Notepac

WHO is bringing the world scientists and global health professionals together to accelerate the research and development process, and develop new norms and standards to contain the spread of the coronavirus pandemic and help care for those affected.

The solidarity of all countries will be essential to ensure equitable access to COVID-19 health products.

WHO is gathering the latest international multilingual scientific findings and knowledge on COVID-19. The global literature cited in the WHO COVID-19 database is updated daily (Monday through Friday) from searches of bibliographic databases, hand searching, and the addition of other expert-referred scientific articles. This database represents a comprehensive multilingual source of current literature on the topic. While it may not be exhaustive.new research is added regularly. The WHO evidence retrieval sub-group has begun collaboration with key partners to enrich the citations and build a more comprehensive database with inclusion of other content. The database is built by BIREME, the Specialized Center of PAHO/AMRO and part of the Regional Offices Department of Evidence and Intelligence for Action in Health. For further information or questions, please contact the WHO Library via email.

Ln 1, Col 1

100% Windows (CRLF)

Two roads diverged in a yellow wood, And sorry I could not travel both

And be one traveler, long I stood And looked down one as far as I could

To where it bent in the undergrowth;

Then took the other, as just as fair, And having perhaps the better claim, Because it was grassy and wanted wear; Though as for that the passing there Had worn them really about the same.

And both that morning equally lay In leaves no step had trodden black. Oh, I kept the first for another day! Yet knowing how way leads on to way, I doubted if I should ever come back.

I shall be telling this with a sigh Somewhere ages and ages hence: Two roads diverged in a wood, and I-I took the one less traveled by, And that has made all the difference.

CREATING A STEGO IMAGE

from cover image











Stego Image

FEATURES EXTRACTION

01

02

03

04

05

Brightness

Contrast

Pixels

Number of Bits

JPEG Quality

PSNR CALCULATION

We use these formulae to calculate the PSNR values using cover and stego images.

$$MSE = \left(\frac{1}{MN}\right) \sum_{i=1}^{M} \sum_{j=1}^{N} \left(X_{ij} - \overline{X_{ij}}\right)^{2} PSNR = 10.\log_{10} \frac{I^{2}}{MSE} db$$

$$PSNR = 10.\log_{10} \frac{I^2}{MSE} \quad db$$

Where,

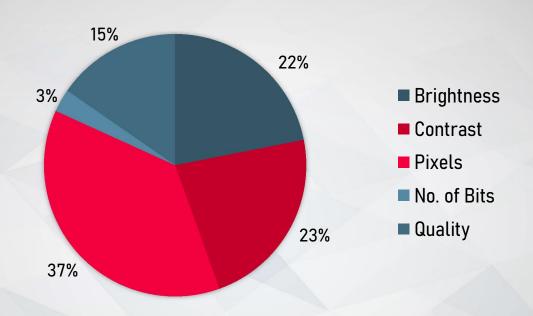
$$X_{ij}$$
 is the i^{th} row and the j^{th} column pixel in the original image, X_{ij} is the i^{th} row and the j^{th} column pixel in the reconstructed (stego) image,

- M and N are the height and the width of the image.
- I is the dynamic range of pixel values, or the maximum value that a pixel can take (equals to (255) for 8-bit images).

CREATING THE DATASET

Brightness	Contrast	Pixels	NoOfBits	Quality	Ans
0.270653	58.566865	1705600	1488	90	1
0.382019	88.083912	480200	1488	80	1
0.809317	81.171926	640000	1488	75	1
0.280167	53.957373	920320	1488	80	1
0.240583	44.726679	618240	1488	80	1

First Five Rows Generated from our CSV Dataset

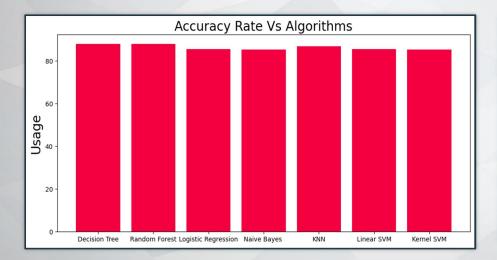


IMPORTANCE CHART of features

MACHINE LEARNING

- Decision Tree
- Random Forest
- Logistic Regression
- Naive Bayes
- K Nearest Neighbours
- Support Vector Machine (Linear)
- Kernel SVM (Radial Basis Function)

ACCURACY CHART of algorithms



Algorithms	Accuracy	
Decision Tree	87.85%	
Random Forest	87.84%	
Logistic Regression	85.39%	
Naive Bayes	85.36%	
K Nearest Neighbours	86.87%	
Linear SVM	85.39%	
Kernel SVM	85.31%	

ML GRAPHS & DATA

DECISION TREE RESULTS



Confusion Matrix: [[1887 162] [347 1792]]

Accuracy Score: 87.84622731614136%

AUC: 0.92

EDGE DETECTION

- Edges are significant local changes of intensity in an image.
- ► Edges typically occur on the boundary between two different regions in an image.
- Produce a line drawing of a scene from an image of that scene.
- Important features can be extracted from the edges of an image (e.g., corners, lines, curves).

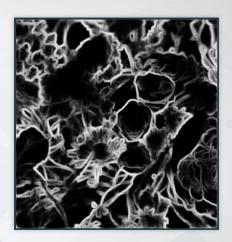
STEPS FOR EDGE DETECTION

- ➤ Smoothing Suppress as much noise as possible, without destroying the true edges.
- ► Enhancement
 Apply a filter to enhance the quality of the edges in the image.
- ► Localization

 Determine the exact location of an edge.



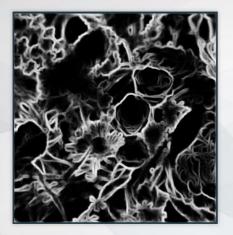
Original Image



HED Image

EDGE DETECTION

of cover image



HED Image

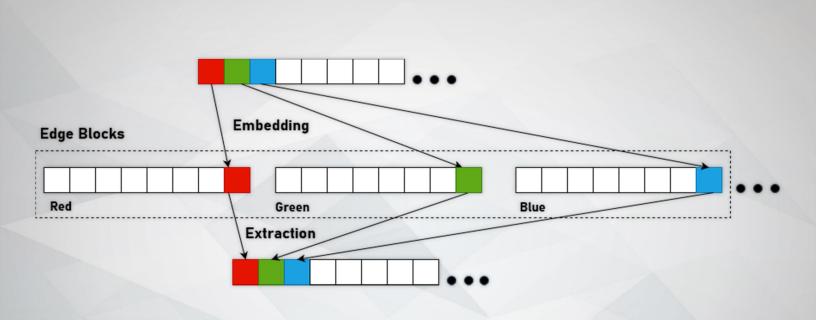


Canny Image

HED v CANNY comparison

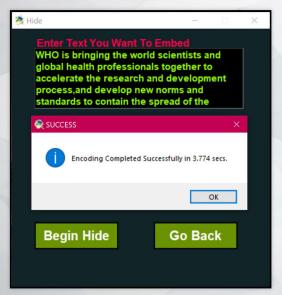
HIDING DATA IN EDGES

- We used Least Significant Bit (LSB) to embed message bits into the HED output.
- ► LSB replaces the least significant bit of the pixel with our message bit.
- Hiding the information in the edges of the image helps in nonsequential embedding so that it cannot be easily detected by attackers.

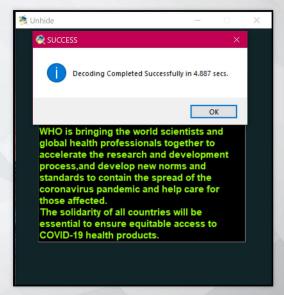


EDGE BASED LSB illustration

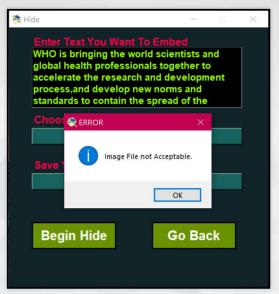
RESULTS of an acceptable image



Hide Process



Unhide Process



Unsuccessful Hiding Process

RESULTS of an unacceptable image

FUTURE ENHANCEMENTS

- Improving the accuracy by training a bigger dataset.
- ► Considering the higher number of features of the image.
- ▶ Improving the embedding of special characters in image.
- ► Enhancing Edge Detection Algorithm to detect Edges better and give better results while hiding information.

