IJCSE International Journal of Computer Sciences and Engineering Open Access
Review Paper Vol.-7, Issue-3, March 2019 E-ISSN: 2347-2693

Examining Robustness of Google Vision API Based on the Performance on Noisy Images

Akshat Pathak[1*], Aviral Ruhela[2], Anshul K. Saroha[3], Anant Bhardwaj[4]

[1,2,3,4]Dept. of Computer Science and Engineering, IMS Engineering College, Ghaziabad, Uttar Pradesh, India

Corresponding Author: akshat.p@imsec.ac.in Tel: +91 8587960403

DOI: https://doi.org/10.26438/ijcse/v7i3.8993 | Available online at: www.ijcseonline.org

Accepted: 14/Mar/2019, Published: 31/Mar/2019

Abstract— Google Cloud Vision is readily used for major purposes such as label detection face recognition mood analysis, object detection content filtering and that is to a certain extent. The efficiency of any system is based on the fact that how well the system is performing in suboptimal conditions in case of Google Cloud Vision the suboptimal working condition include the use of noisy images instead of perfect ones. This paper deals with how this Google Cloud Vision works under noisy images and how robust the system stays under these conditions. This API generates different outputs by adding different noises with different intensity in noise. It is clearly observed that with the mean value of 20% impulse noise and 0.1 variance Gaussian noise, the API can be easily misguided in predicting the actual label and text for the images. A better and accurate outcome can be obtained by pre-processing and validating the image for any noise and denoising an image up to some extent for a better and accurate outcome which could be more beneficial than updating the currently working algorithm.

Keywords— Google Cloud Vision, Robustness, Noisy Images, Gaussian Noise, Impulse Noise

I. Introduction

Traditionally convolutional neural networks were implemented for image and video recognition. Many image datasets were there for training; various techniques were implemented to improve the architecture. The solution which is now most acceptable is the ML algorithms which are implemented by various web services in which the leading is Google Cloud Vision and Amazon Rekognition.

Google Cloud Vision was introduced on 2nd December 2015 since then it's been continuously growing and developing. Image recognition feature was made first available on Google photos which totally relies on pattern recognition and matching algorithm for classification of images on the basis of landmark people faces and objects. Cloud Vision is a proprietary API which can be advantageous to various developers for developing applications for image analysis, by using the various REST APIs. There are features that are provided for image recognition including Identifying landmarks, Optical Character Recognition, Sage Search Detection, Facial Detection and Logo Detection. Which is further advanced to detect the emotions of the faces and works similar to entity detection like Google image search. Advantageous in the moderate content analysis along with language detection in OCR followed by image attributes.

All the aforementioned features were defined to work under noise-free conditions of the images and texts. Noise in the images could have produce completely different effect on the output, there is also a possibility of accepting the noise with images and providing an output that is same as with the original images. Such robustness cannot be obtained until a special training or Learning is created which from the past experiences of the images recognized. For major applications in real word, the system should be robust to all kind of input for better and accurate performance but the researches show a different reality which includes the vulnerability of adversarial images in such ML algorithm-based systems. This papers targets on examining the robustness of Google Vision API by attacking it with noisy data. This is done by adding Gaussian and Impulse noise up to an extent that the output of the image from Cloud Vision API is completely altered but humans will still able to recognize the same [1].

The paper is organized as follows: Section II is about the literature review of the past published papers, Section III is about the type of noises implemented for the research work, Section IV is all about the procedure used to attack the Google Cloud Vision API, Section V show the sample data used, Section VI is the final result followed by conclusion.

II. Literature Review

Majority of the algorithm for visual related task include convolutional neural networks. They are trained on very large image sets. Many papers concluded the fact that low confidence in the output is due to the distortion in images by various means [2][3]. Various architectures are also discussed for increasing the accuracy of such networks which have overcome the approach of retraining the dataset with noisy images [4]. The vulnerability of various algorithms is also observed for integrity attacks which includes adding small imperceptible force to miscalculate the outputs [5]. On contrary to the discussion above we majorly focus on generating adversely affected results by adding noises to our images and comparing the results with the original ones, which is in contrast to other papers. The noisy images can further denoise by using wiener filter [6] to validate the actual result and to verify that denoising could be a better option for Google Cloud Vision API instead of retraining the whole network with noisy or adverse data. Along with this, we are planning to attack the API with more random and discrete data.

III. Noise In Images

Unwanted distort signal corrupting the quality and information in the digital images is considered as noise. Disturbed background scenes, artefacts, absurd lines, corners, edges are some undesirable effects produce by noise [7]. Digitization and image transmission may sometimes arise noise in images. Physical factors include camera defects, light and sensor temperature. We are modifying our images by attacking them with Gaussian noise and salt and pepper noise (impulsive noise).

A. Gaussian Noise

Commonly known as electronic noise due to the fact that it arises from detectors and amplifiers. Possible natural sources are there such as thermal vibrations of an atom along with radiation of warm object that could be discrete [8]. The mathematical model of Gaussian noise consists of Probability density function (PDF) referred to normal distribution [9] also known as Gaussian distribution which is [10] and majorly disturbs the gray values of the image. The normalized histogram or the PDF is given by (1)

$$P(g) = \sqrt{\frac{e}{2\pi\sigma^2}}^{\frac{-(g-\mu)^2}{2\sigma^2}} \qquad (1)$$

Having g = gray value, σ = standard deviation and μ = mean. For default in terms of PDF we have the mean value zero with variance of 0.01 and 256 gray levels.

B. Impulsive noise

Commonly known as salt and pepper noise, independent noise, spike noise due to nature to drop the data values is also called data drop noise. Any sudden change in signal, error in digitization and transmission during synchronization, malfunctioning of camera cell. Natural causes are dust particles in air or in image acquisition source. A random pattern of black and white dots is seen in the images which can be identified as dark pixels over the light regions and light pixels over the dark regions [11], thus does not affect the whole image but some regions [12]. Given by (2)

$$P(z) = \begin{cases} P_a & for \quad z = a \\ P_b & for \quad z = b \\ 0 & othervise \end{cases} \qquad (2)$$

Where *P(z)* is a probability density function and P_a defines bright region and P_b defines dark region.

IV. Projected Attack On Google Vision API

This part is majorly focused on describing the planned attack on Google Cloud Vision API. The aim is to inflict the images along with noise, feeding them to Google Cloud Vision API and comparing the outputs provided by API for noisy and original images. Variety of image samples has been chosen for this experiment including a various kinds of Texts, Animals, Scenes, Objects, etc. We created our own dataset, which is divided into 5 categories

1. Human Faces
2. Animals
3. Objects
4. Surroundings
5. Text

The selected images are sent for analysis in the Application which we created using Google Cloud Vision API and it outputs Object Detection for Labels and Location, Sentiment Detection, Face Search, Logo and landmark detection for surroundings, OCR for Text images, but for the sake of accuracy and efficiency in the results we are directly sending our images to the Google Cloud Vision and recoding the output from that source only.

The methodology is followed by first testing the original image with our developed android application which uses Google Cloud Vision API and then recording the output returned by the API. Thenceforth the modified images with both Gaussian and Impulse noise separately are generated starting with very low variance value for Gaussian noise 0.01 and impulse noise density 5%. Updating Gaussian variance at each point by 0.01 and impulse noise density by 5% until we reach a point where we obtain completely different labels for our images with respect to the original. Our final aim is to find the point where we can proclaim our image as an adverse image. During the process, we also observed the confidence probability of labels diminishes during the course of increasing noises amplitude in the images.

Noises are introduced to the images by using MATLAB these images are then saved. These images are then processed in the Google Could Vision library which will provide us with the label detection along with other features which can be observed and analyse to find the traits how noise is affecting the processing and identification process.

V. Sample Data

A sample of 15 images with complete variated is tested for each category giving a total of 75 images which are to be tested for identification by adding cumulative noise till we reach a point where the images are humanly recognizable but the label detection fails to provide the correct labels for the images. So a total of 300 samples are tested manually for this research. Example image of each category can be seen in the Figure 1

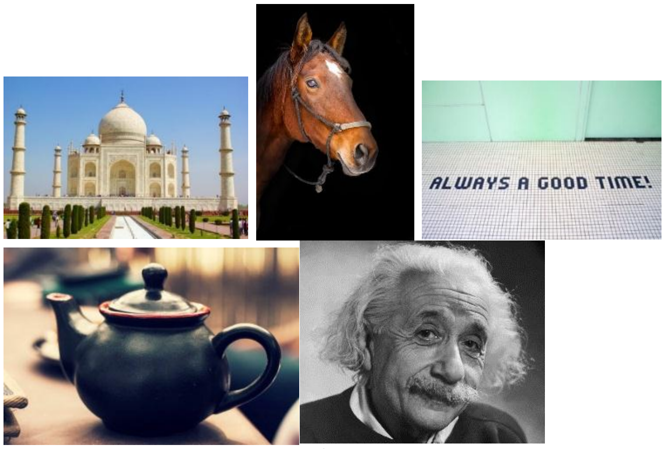


Figure 1 Sample Images for each Category

Figure 2 shows the sample images after adding noise with .01 variances in Gaussian noise and 5% of impulse then increasing the noise which shows how the texture and feature changes when noise is added. (a) with 5% impulse noise, (b) with 12% impulse noise, (c) with 20% noise

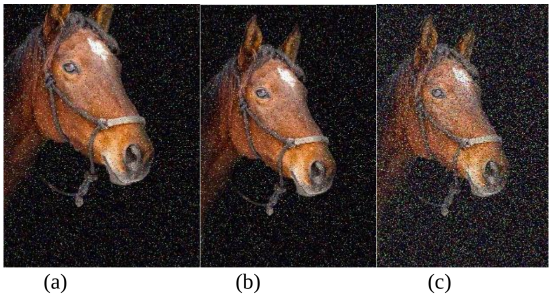
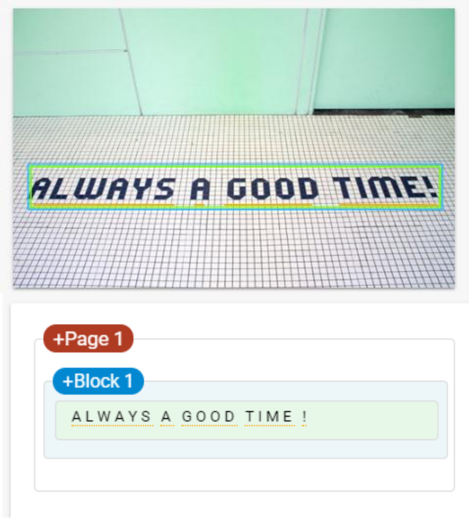

(a) (b) (c)

Figure 2 Increasing noise intensity level

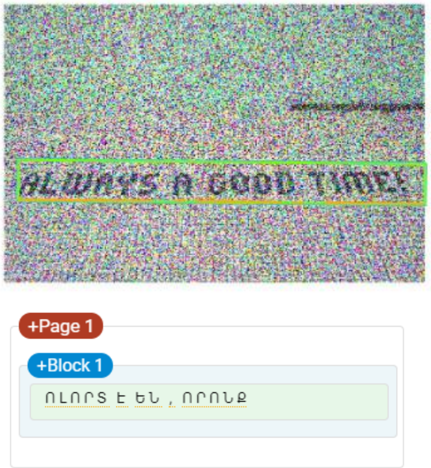
VI. Results

It is clearly observed that for different objects types and different categories and noise types the robustness varies vastly. It is found that for a dark background image the Gaussian and impulse noise intensity should be high to render Google Vision fail to identify. It is observed that the Google Cloud Vision is highly robust for text detection and is maximum for black background with white text and white background with black text containing capital letters mostly. For text the systems fail for 50% of impulse noise and 0.9 variance of Gaussian noise. In case of animal recognition, the system fails to identify with 20% impulsive noise and 0.1 variance Gaussian noise which is found same for monuments detection as well as for face detection while for objects the maximum objects are unidentified in 15% impulsive noise. We also found that with different noise the labels are also different. Meaning the variation in noise will also cause variation in label detection. For the same image Gaussian and impulse noise provide different labels as we can see in the figure 4(b) and figure 4(c).



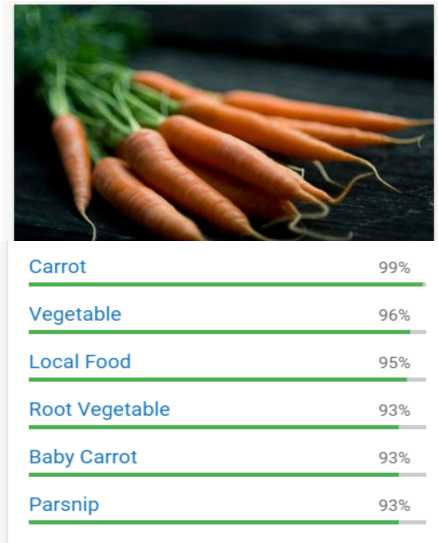
ALWAYS A GOOD TIME !

Figure 3(a). Original Text image



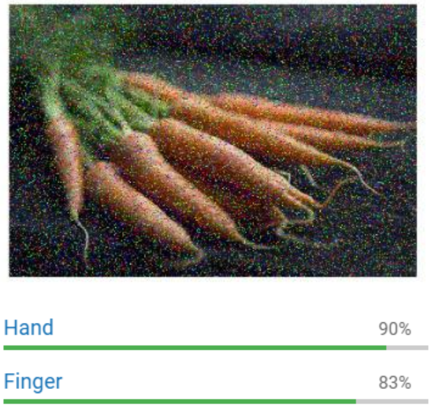
ոլորտ է են , որոնք

Figure 3(b). with 50% impulse noise



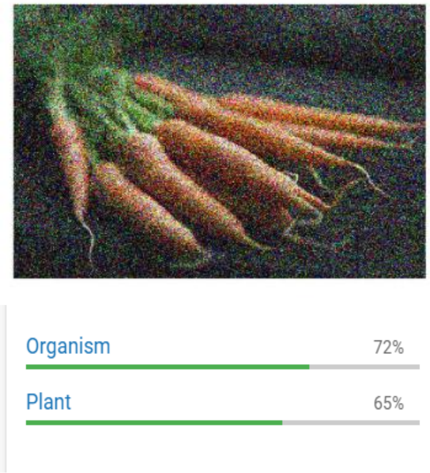
Carrot	99%
Vegetable	96%
Local Food	95%
Root Vegetable	93%
Baby Carrot	93%
Parsnip	93%

Figure 4(a). Original Image without any noise



Hand	90%
Finger	83%

Figure 4(b). With 20% impulse noise



Organism	72%
Plant	65%

Fig. 4(c) With 0.1 variance Gaussian noise

VII. Conclusion

We can finally conclude the fact that the Google Cloud Vision is not robust to noisy images and works very inefficiently under such conditions. We can also observe that the label changes with the nature of noise along with its intensity and finally becomes completely ineffective in text detection. We propose a pre-processing system which denoises the image (if noise is present) or a Cloud filter that rejects images with too much nose, to be installed before performing various detection operations.

References

[1] F. R¨ohrbein, P. Goddard, M. Schneider, G. James, and K. Guo, *"How does image noise affect actual and predicted human gaze allocation in assessing image quality?"* Vision research, vol. 112, pp. 11–25, 2015.

[2] K. Simonyan and A. Zisserman. *"Very deep convolutional networks for large-scale image recognition"*. In Proceedings of conference paper at ICL, 2015

[3] I. Vasiljevic, A. Chakrabarti, and G. Shakhnarovich, *"Examining the impact of blur on recognition by convolutional networks,"* arXiv preprint arXiv:1611.05760, 2016.

[4] S. Diamond, V. Sitzmann, S. Boyd, G. Wetzstein, and F. Heide, *"Dirty pixels: Optimizing image classification architectures for raw sensor data,"* arXiv preprint arXiv:1701.06487, 2017.

[5] Papernot, N., M Cdaniel, P., Jha, S., Fredrikson, M., Celik, Z. B.,and Swami, A. *"The limitations of deep learning in adversarial settings*". In Proceedings of the 1st IEEE European Symposium on Security and Privacy", arXiv preprint Xiv:1511.07528, 2016.

[6] Ratnesh Kumar Shukla, Ajay Agarwal, Anil Kumar Malviya, *"An Introduction of Face Recognition and Face Detection for Blurred and Noisy Images"*, International Journal of Scientific Research in Computer Science and Engineering, Vol.6, Issue.3, pp.39-43, 2018

[7] Dougherty G. *"Digital Image Processing for Medical Applications,"* second ed., Cambridge university press, 2010.

[8] Boyat, A. and Joshi, B. K. *"Image Denoising using Wavelet Transform and Median Filtering"*, 2013 Nirma University International Conference on Engineering (NUiCONE), Ahmedabad, pp. 1-6, 2013.doi: 10.1109/NUiCONE.2013.6780128

[9] Mandeep Kaur, Balkrishan Jindal, *"Improved Sparse matrix Denoising Techniques using affinity matrix for Geographical Images"*, International Journal of Scientific Research in Computer Science and Engineering, Vol.5, Issue.5, pp.51-56, 2017

[10] Priyanka Kamboj, Versha Rani,*" Brief study of various noise model and filtering techniques"*, Journal of Global Research in Computer Science, vol.4, No.4, pp.166-171, April 2013.

[11] Monika Raghav, and Sahil Raheja,*" Image Denoising Techniques: Literature Review"*, International Journal of Engineering and Computer Science, vol.3, pp. 5637-5641, Issue 5, May 2014.

[12] Joshi, A., Boyat, A. and Joshi, B. K. *"Impact of Wavelet Transform and Median Filtering on removal of Salt and Pepper noise in Digital Images,"* IEEE International Conference on Issues and Challenges in Intelligant Computing Teachniques, Gaziabad, India, 2014

Authors Profile

Mr. Akshat Pathak Scholar, Department of Computer Science and Engineering is pursuing Bachelor of Technology from IMS Engineering College, Ghaziabad, India. Has a keen research interest in image processing and genetic algorithm and worked on various projects and has published a paper on Cancer Biology.

Mr. Aviral Ruhela is pursuing Bachelor of Technology from IMS Engineering College, Ghaziabad with Computer Science and Engineering as his specialization. Has worked on projects related to Automation, Encryption and Image Processing and has developed desktop as well as mobile application for small enterprises.

Mr. Anshul Kumar Saroha Scholar, Department of Computer Science and Engineering is pursuing Bachelor of Technology from IMS Engineering College, Ghaziabad, India, has a keen research interest as well as completed various projects in Machine Learning and Image Processing.

Mr. Anant Bhardwaj Scholar, Department of Computer Science and Engineering is pursuing Bachelor of Technology from IMS Engineering College, Ghaziabad, India, has a keen research interest in image processing and web developing and worked on various projects on the same.