SIT 789: Edge orientation and morphology

1:

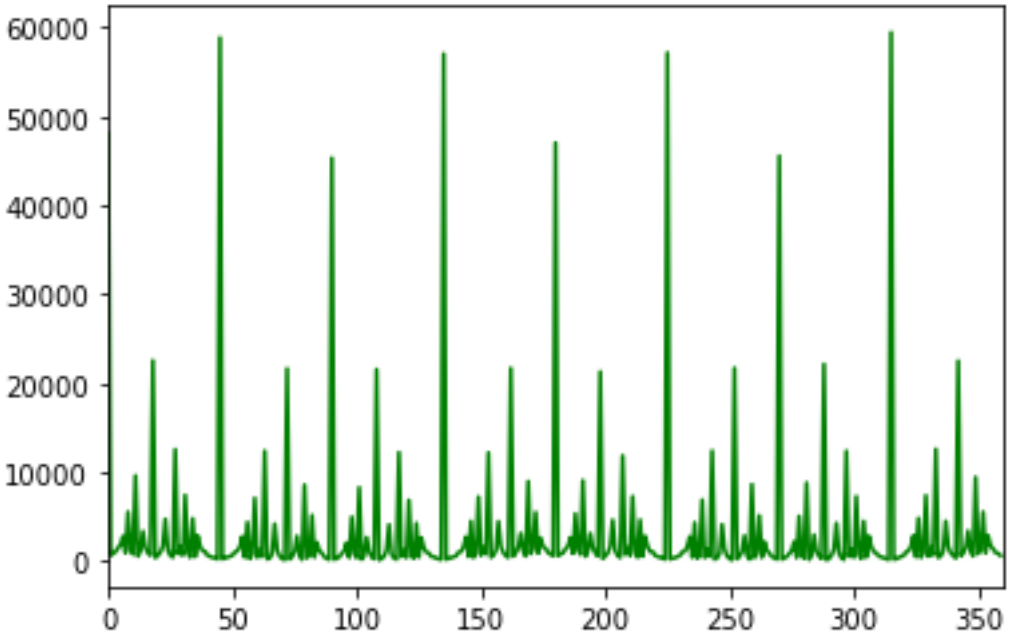


Fig 1: Histogram for fisherman.jpg

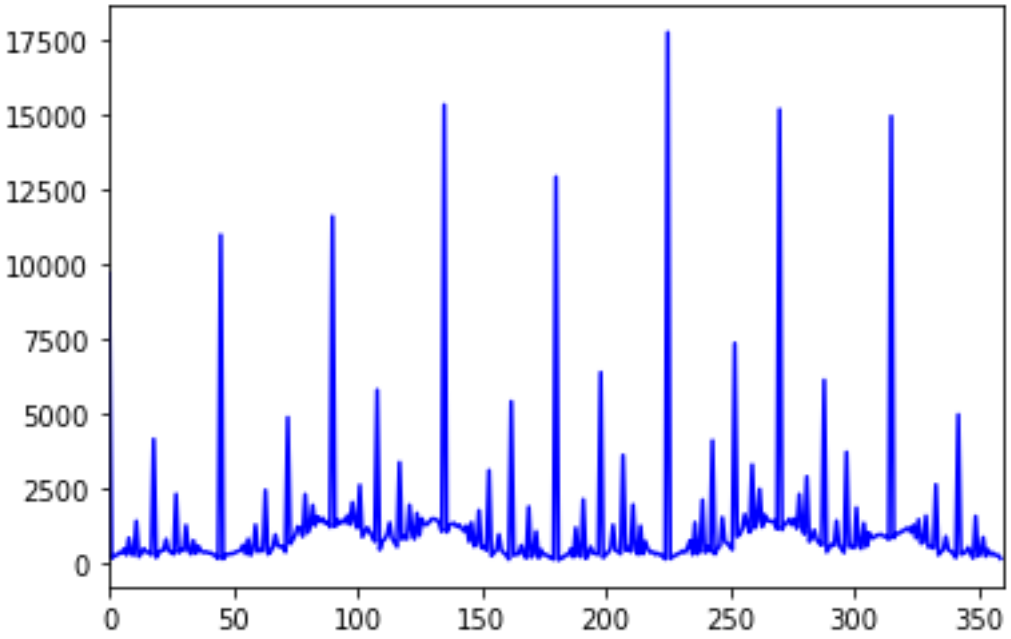


Fig 2: Histogram for empire.jpg

2:

When χ2 distance and KL Divergence is calculated between these two histograms then we get:

χ2 distance : [0.2154838]
KLDivergence: [0.5181271]
This shows that using χ2 distance histograms are more similar as the value is 0.21 whereas with KL Divergence the value is 0.51 which highlights that the histograms are less similar.

3:

Minimal Adversarial Examples for Deep Learning on 3D Point Clouds

Jaeyeon Kim[1] Binh-Son Hua[2,3] Duc Thanh Nguyen[4] Sai-Kit Yeung[1]
[1]Hong Kong University of Science and Technology [2]VinAI Research, Vietnam
[4]Deakin University [3]VinUniversity, Vietnam

Abstract

With recent developments of convolutional neural networks, deep learning for 3D point clouds has shown significant progress in various 3D scene understanding tasks, e.g., object recognition, semantic segmentation. In a safety-critical environment, it is however not well understood how such deep learning models are vulnerable to adversarial examples. In this work, we explore adversarial attacks for point cloud-based neural networks. We propose a unified formulation for adversarial point cloud generation that can generalise two different attack strategies. Our method generates adversarial examples by attacking the classification ability of point cloud-based networks while considering the perceptibility of the examples and ensuring the minimal level of point manipulations. Experimental results show that our method achieves the state-of-the-art performance with higher than 89% and 90% of attack success rate on synthetic and real-world data respectively, while manipulating only about 4% of the total points.

Figure 1: Left: input point cloud (1,024 points) classified correctly by PointNet [27] and 15 selected points (in green). Right: adversarial point cloud misclassified by PointNet and perturbed locations (in red).

1. Introduction

Deep learning has shown great potentials in solving a wide spectrum of computer vision tasks. In life-crucial applications, one concern is that deep neural networks can be vulnerable to adversarial examples, a special kind of inputs that can fool the networks to make undesirable predictions. Several adversarial attack techniques have been proposed to generate such examples. In contrast, adversarial defense methods have been developed to detect and neutralise adversarial examples. Therefore, understanding how adversarial attacks and defenses operate is of great importance to make deep learning techniques more reliable and robust.

With the growing popularity of low-cost 3D sensors and light-field cameras, the community has started investigating the vulnerability of deep networks on 3D data, especially 3D point clouds [40, 19, 20, 43, 36]. However, existing works focus on common scenarios, such as generating adversarial point clouds by perturbing points. While such approaches have high attack success rates, the perturbations are not imperceptible and can be identified easily by outlier detection or noise removal algorithms. In addition, existing adversarial attack methods do not perform optimally since all points in a point cloud are involved in the manipulation.

In this work, we study 3D adversarial attack in a more extreme but practical setting: how to generate an adversarial point cloud with minimum number of points perturbed from an original point cloud while maintaining the perceptibility of the original point cloud (see Figure 1). To address this problem, we propose a new formulation for adversarial point cloud generation that can be adapted to different attack strategies. The novelty of our work lies in both the research problem and proposed solution. Specifically, minimal 3D point cloud attack is an unexplored problem. We are also the first to propose a formulation that (i) considers both the perceptibilty and optimality of adversarial samples, and (ii) generalises both point perturbation and addition in a unified framework. In summary, our contributions include:

- A new technique to generate minimal 3D adversarial point clouds;
- A unified formulation that generalises two adversarial point cloud generation strategies: point perturbation and point addition;
- A vulnerability analysis on the relation between the perturbed points found by our method and the concept of

Fig 3: Result of binary image (Not inverted)

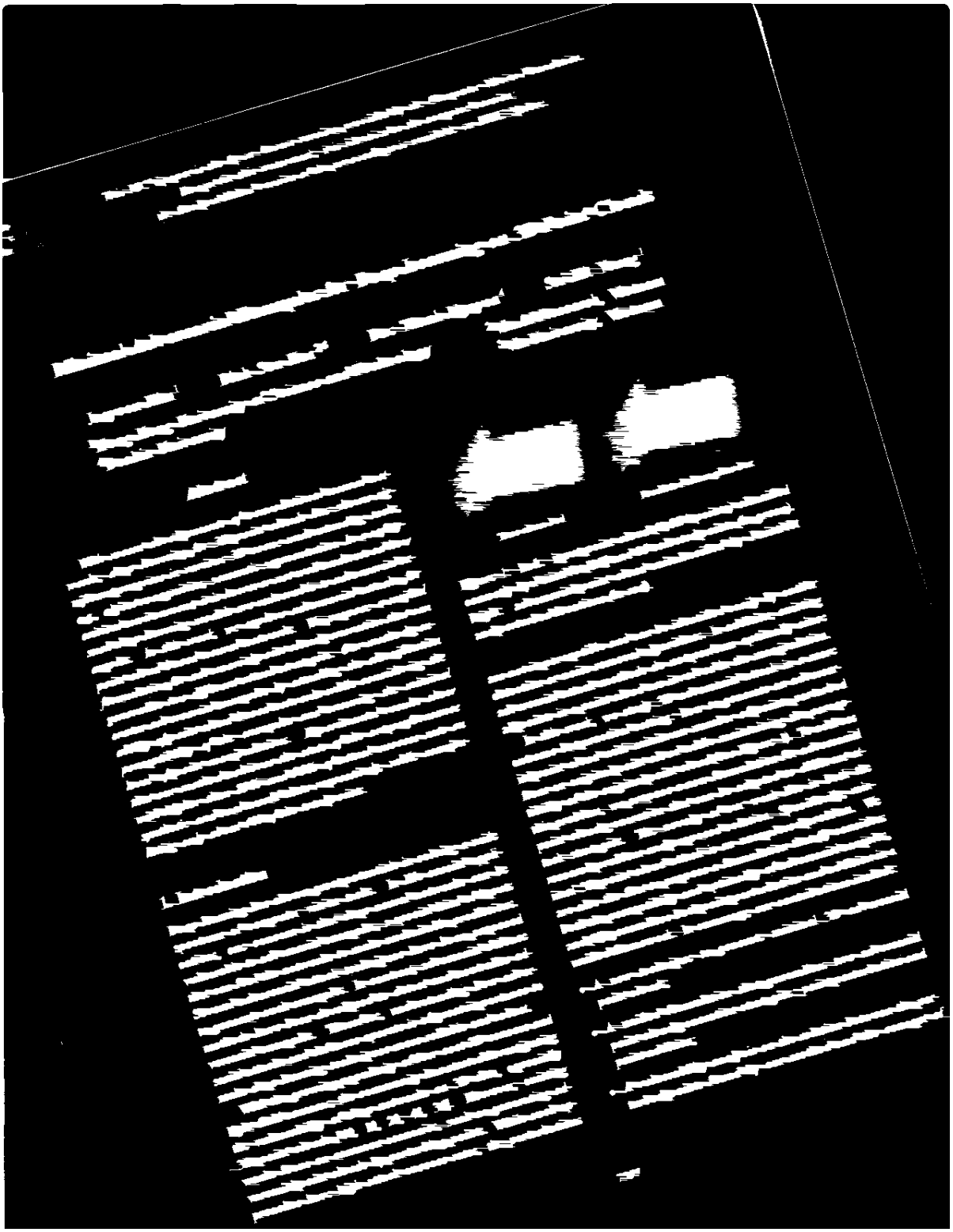


Fig 4: Result of closing using structuring element: (np.ones((1, 15), np.int)

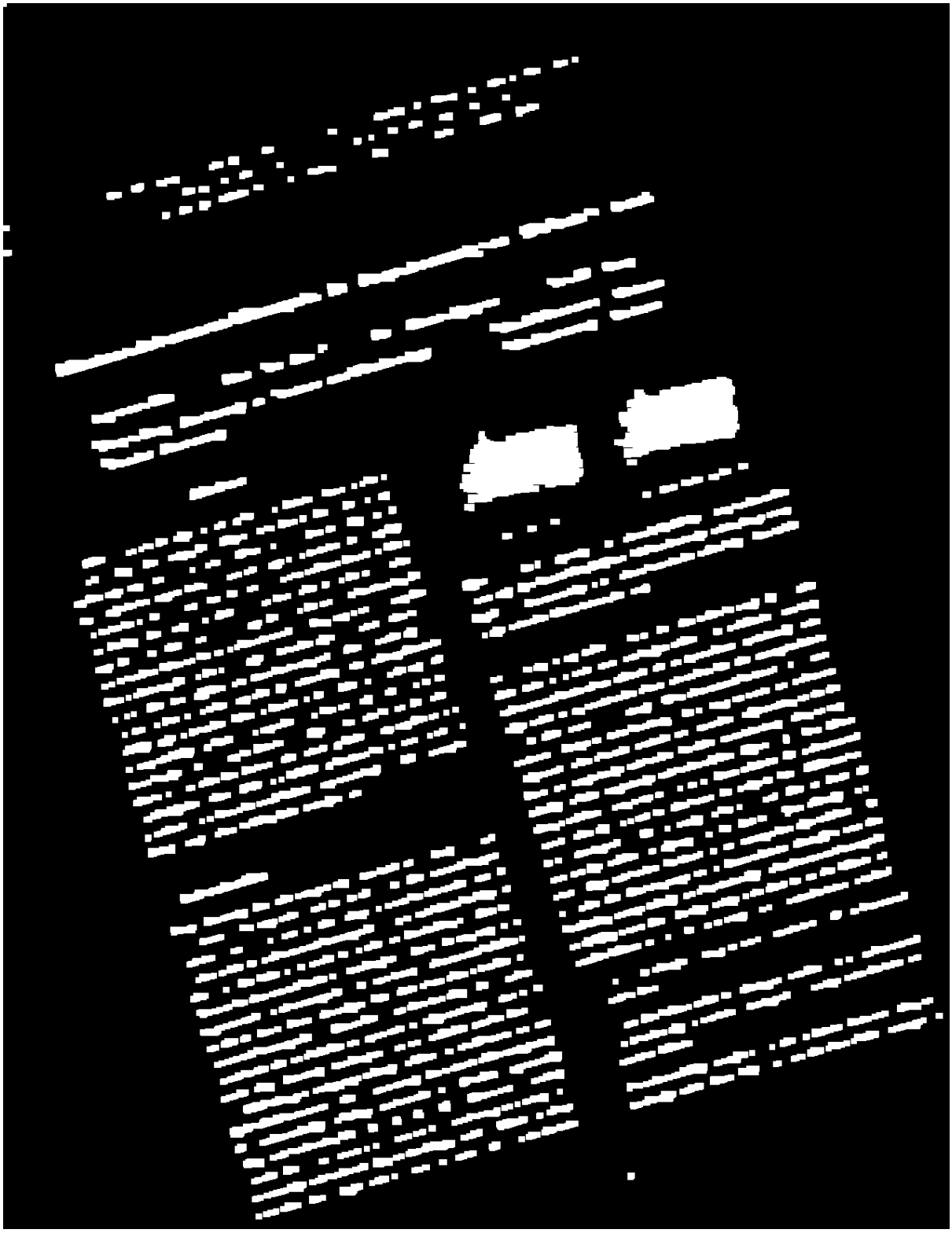


Fig 5: Result of opening using structuring element: (np.ones((8, 8), np.int)

4:

Minimal Adversarial Examples for Deep Learning on 3D Point Clouds

Jaeyeon Kim[1] Binh-Son Hua[2,3] Duc Thanh Nguyen[4] Sai-Kit Yeung[1]

[1]Hong Kong University of Science and Technology [2]VinAI Research, Vietnam
[3]VinUniversity, Vietnam
[4]Deakin University

Abstract

With recent developments of convolutional neural networks, deep learning for 3D point clouds has shown significant progress in various 3D scene understanding tasks, e.g., object recognition, semantic segmentation. In a safety-critical environment, it is however not well understood how such deep learning models are vulnerable to adversarial examples. In this work, we explore adversarial attacks for point cloud-based neural networks. We propose a unified formulation for adversarial point cloud generation that can generalise two different attack strategies. Our method generates adversarial examples by attacking the classification ability of point cloud-based networks while considering the perceptibility of the examples and ensuring the minimal level of point manipulations. Experimental results show that our method achieves the state-of-the-art performance with higher than 89% and 90% of attack success rate on synthetic and real-world data respectively, while manipulating only about 4% of the total points.

1. Introduction

Deep learning has shown great potentials in solving a wide spectrum of computer vision tasks. In life-crucial applications, one concern is that deep neural networks can be vulnerable to adversarial examples, a special kind of inputs that can fool the networks to make undesirable predictions. Several adversarial attack techniques have been proposed to generate such examples. In contrast, adversarial defense methods have been developed to detect and neutralise adversarial examples. Therefore, understanding how adversarial attacks and defenses operate is of great importance to make deep learning techniques more reliable and robust.

With the growing popularity of low-cost 3D sensors and light-field cameras, the community has started investigating the vulnerability of deep networks on 3D data, especially 3D point clouds [40, 19, 20, 43, 36]. However, existing works focus on common scenarios, such as generating adversarial point clouds by perturbing points. While such approaches have high attack success rates, the perturbations

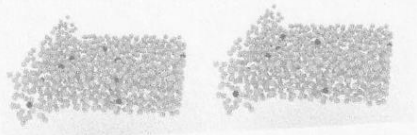


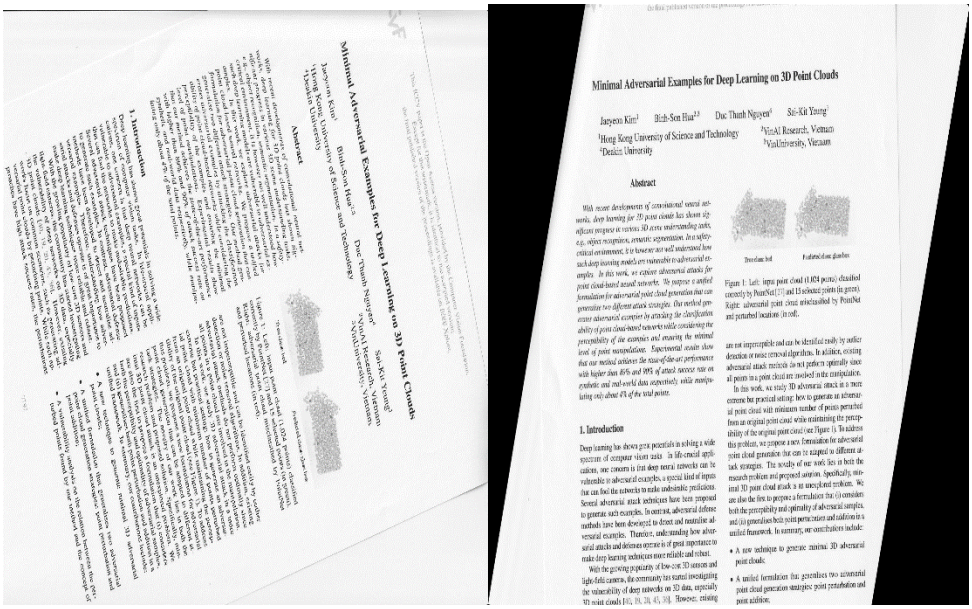
Figure 1: Left: input point cloud (1,024 points) classified correctly by PointNet [27] and 15 selected points (in green). Right: adversarial point cloud misclassified by PointNet and perturbed locations (in red).

are not imperceptible and can be identified easily by outlier detection or noise removal algorithms. In addition, existing adversarial attack methods do not perform optimally since all points in a point cloud are involved in the manipulation.

In this work, we study 3D adversarial attack in a more extreme but practical setting: how to generate an adversarial point cloud with minimum number of points perturbed from an original point cloud while maintaining the perceptibility of the original point cloud (see Figure 1). To address this problem, we propose a new formulation for adversarial point cloud generation that can be adapted to different attack strategies. The novelty of our work lies in both the research problem and proposed solution. Specifically, minimal 3D point cloud attack is an unexplored problem. We are also the first to propose a formulation that (i) considers both the perceptibilty and optimality of adversarial samples, and (ii) generalises both point perturbation and addition in a unified framework. In summary, our contributions include:

- A new technique to generate minimal 3D adversarial point clouds;
- A unified formulation that generalises two adversarial point cloud generation strategies: point perturbation and point addition;
- A vulnerability analysis on the relation between the perturbed points found by our method and the concept of

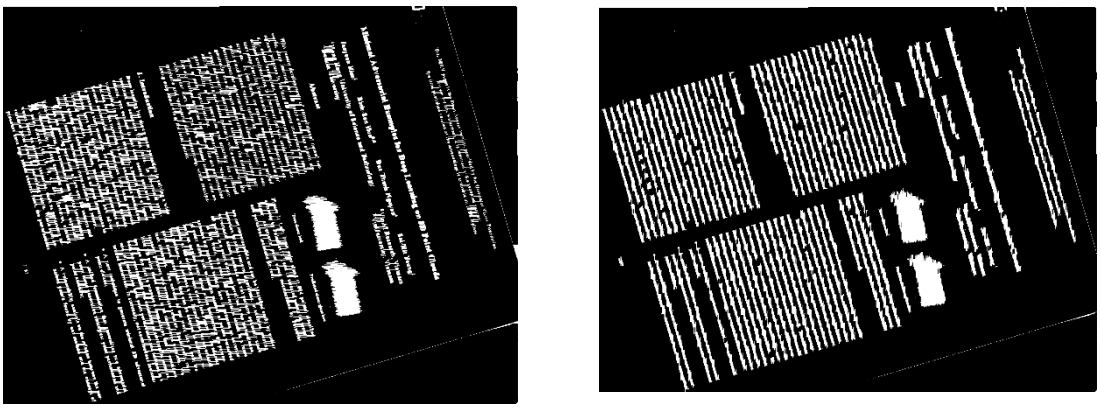
Fig 6: De-skewed document for doc.jpg



(a) (b)

Fig 7: (a) De-Skewed image using the structuring element (closing) used for doc.jpg; (b) De-skewed image using structuring element(np.ones((15, 1), np.int)) for closing.

We can see that using the same structuring element as in doc.jpg the approach does not work as the opening image (fig 8) links sentences from different lines into one. Whereas when we use a different structuring element which is perpendicular to the earlier structuring element then words from same line are linked as the document is horizontal. This makes using a perpendicular structuring element *np.ones(15, 1)* to the earlier structuring element optimal to use.



(a) (b)

Fig 8: Difference in (a) using same structuring element as in doc.jpg (b) opening image using different structuring element. np.ones(15, 1)