**Understanding classification failure in Machine Learning using  
CNNs and GANs**

Insert Subtitle Here

Sanket Mehrotra  
 Department of Computer Science  
 Colorado State University  
 Fort Collins, CO, USA  
 sanketm@colostate.edu

Rachit Dalal  
 Department of Computer Science  
 Colorado State University  
 Fort Collins, CO, USA  
rdalal@colostate.edu

ABSTRACT

In this sample-structured document, neither the cross-linking of float elements and bibliography nor metadata/copyright information is available. The sample document is provided in “Draft” mode and to view it in the final layout format, applying the required template is essential with some standard steps.

These steps, which should require generation of the final output from the styled paper, are mentioned here in this paragraph. First, users have to run “Reference Numbering” from the “Reference Elements” menu; this is the first step to start the bibliography marking (it should be clicked while keeping the cursor at the beginning of the reference list). After the marking is complete, the reference element runs all the options under the “Cross Linking” menu.

For accuracy check of the structured paper, user can run the option **Manuscript Validation**. It informs the user of the wrong or missing values in the paper. The user must correct the paper as per validation messages and rerun **Manuscript Validation**.

Now, to generate the required layout of the paper, the user should select one of the template styles under the Define Template Style option (choose the required layout design, i.e. choose between Journals and Proceedings).

Some specific values are required to create a standard layout by choosing a template for the journals or proceedings. So once the user chooses one of the template layout styles, the respective Journal or Conference details dialog box (i.e. **journal/conference acronym, DOI, ISBN, copyright, year, etc.**) will appear as a prompt during the Define Template Style functionality. The user should fill these values, after which the template creates the desired layout of the paper. The user can now create a PDF of his/her manuscript using the “*Save as PDF*” option.

If the user is adding any new data, they should make sure to style it as per the instructions provided in previous sections. Carry out the steps for Cross-linking, Fundref data, adding Document History (specific to journal submission), and finally, Manuscript validation and placing the respective while applying the required template.

CCS CONCEPTS

• Image Classification • Machine Learning Failure • Neural Networks • Convolutional Neural Networks • Generative Adversarial Networks

KEYWORDS

Image Misclassification, Adversarial Attacks, CNNs, Conditional GANs

ACM Reference format:

Sanket Mehrotra and Rachit R Dalal. 2021. Understanding classification failure in Machine Learning using CNNs and GANs. In *Proceedings of the Genetic and Evolutionary Computation Conference 2021 (GECCO ’21). ACM, New York, NY, USA, 2 pages.* https://doi.org/10.1145/1234567890

1. Introduction

Understanding how and why neural networks can be fooled into misclassifying images is an interesting topic that has been widely explored. This topic has gained much attention of researchers as we are rapidly entering the world of self-driving cars and these kinds of failures can causes major hurdles in bringing technologies to the mainstream society. Some papers [5] show that one-pixel change is enough to force a misclassification. This can be done using CNNs or GANs [5], and even uses evolutionary algorithms in some cases [6]. We want to try to this for ourselves and explore some possible reasons to explain this phenomenon. We plan on trying to work with GANs, something outside both of our machine learning experience.

Deep neural networks have been widely used for classifying images with the high accuracy. This project is our quest to explore the tolerance and identify particular weaknesses of image classifying networks to different types and degrees of structured, random and specialized noises. The addition of the noise is  
almost indistinguishable to the human eyes but surprisingly it can completely fool neural networks. Moreover, another important point which drives this research is its application in a real-time scenarios e.g. If a self-driving car just ignores or misclassifies a stop-sign or a pedestrian because its neural networks have mis-classified its sensors’ input images. With this we hypothesize that:

1. It should be possible to identify certain noise thresholds beyond which a network starts regularly failing,
2. Similarly, certain noise patterns may be more disruptive to the classification process of trained networks and even pre-trained networks.
3. Training GANs on random and structured noise datasets may help us understand the failure of the machine learning models or the solution to overcome this problem.



Figure 1: Figure Caption and Image above the caption [In draft mode, Image will not appear on the screen]

**Theorem/Proof/Lemma.** Insert text here for the enunciation or Math statement. Insert text here for the enunciation or Math statement. Insert text here for the enunciation or Math statement. Insert text here for the enunciation or Math statement. Insert text here for the enunciation or Math statement.

....Insert text here for the Quotation or Extract, Insert text here for the Quotation or Extract, Insert text here for the Quotation or Extract, Insert text here for the Quotation or Extract, Insert text here for the Quotation or Extract, Insert text here for the Quotation or Extract.

2. Background

In the below paragraph, it is explained how alt-txt value is placed in **MS Word 2010**. To add alternative text to a picture in Word 2010, follow these steps:

1. In a Word 2010 document, insert a picture.
2. Right click on the inserted picture and select the **Format Picture** option.
3. Select the **Alt Txt** option from the left-side panel options.
4. In the "Title:" and "Description:" text boxes, type the text you want to represent the picture, and then click "Close".

3. Motivation

When we started this project, we wanted to understand misclassification of supervised learning algorithms and neural nets. Over time we started work and focused on a subset of this problem, namely the effect of noise and it’s types on trained and pre-trained networks. Later on we stumbled upon the wide field of Adversarial attacks and generative adversarial networks.

4. Methodology

4.1 Dataset

German Traffic Sign Recognition Benchmark Dataset [2]  
Published by researchers at the Ruhr-Universität Bochum, Germany in 2011 for the International  
Joint Conference on Neural Networks (IJCNN). The dataset has the following properties:  
• Single-image, multi-class classification problem  
• More than 40 classes  
• More than 50,000 images in total  
The images in this dataset are ~ 32x32 images of road signs. Below are a few samples:

* + 1. Preprocessing

4.2 Experiments

4.2.1 Baseline Model

Pass.

4.2.2 Baseline Model

5. Results

*1.1.1.1 Heading Level 4.*Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here.

1. Conclusion

ACKNOWLEDGMENTS

Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here.

REFERENCES

[1] Patricia S. Abril and Robert Plant, 2007. The patent holder's dilemma: Buy, sell, or troll? *Commun. ACM* 50, 1 (Jan, 2007), 36-44. DOI: <https://doi.org/>10.1145/1188913.1188915.

[2] Sten Andler. 1979. Predicate path expressions. In *Proceedings of the 6th. ACM SIGACT-SIGPLAN Symposium on Principles of Programming Languages (POPL '79)*. ACM Press, New York, NY, 226-236. DOI:https://doi.org/10.1145/567752.567774

[3] Ian Editor (Ed.). 2007. *The title of book one* (1st. ed.). The name of the series one, Vol. 9. University of Chicago Press, Chicago. DOI:https://doi.org/10.1007/3-540-09237-4.

[4] David Kosiur. 2001. *Understanding Policy-Based Networking* (2nd. ed.). Wiley, New York, NY..

Conference Name:ACM Woodstock conference

Images

1. Architecture – CNN and GAN

Conference Short Name:WOODSTOCK’18

Conference Location:El Paso, Texas USA

ISBN:978-1-4503-0000-0/18/06

Year:2018

Date:June

Copyright Year:2018

Copyright Statement:rightsretained

DOI:10.1145/1234567890

RRH: F. Surname et al.

Price:$15.00

1. Dataset introduction grid – example of classes
2. Noise Grid
3. Baseline CNN training graphs