Year 12 Science Extension

Research Report

Improving the accuracy of a deep learning object detection model through altered colour saturation image training

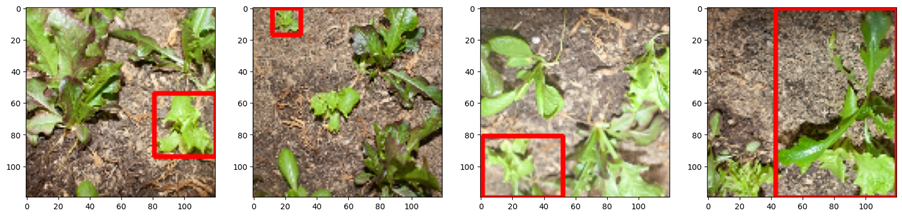
Student Number: 36225378

**Abstract**

Deep Learning Object Detection is crucial for computers to visualise their environment realistically. It is being increasingly applied in real-world applications. They require accurate detections; however output errors can arise. These errors can consist of false positives and hallucinations. This study evaluates the impact of altering colour saturation in training images to improve Deep Learning Object Detection performance. Three separate models were trained on standard, decreased and increased colour saturation images. Total Loss was used to evaluate the training performance. Substantially lower Total Loss values were found in a Deep Learning Object Detection model trained on images that lack colour saturation. This indicates improved accuracy with this model’s performance. Therefore current models could be retrained with images without colour saturation for significant improvements in accuracy. This simple and efficient method could prove a potential method of enhancing Deep Learning Object Detection model accuracy and deployment in a multitude of applications.

**Literature Review**

Object Detection is a process that allows computers to see their environments by visually identifying and tracking targets (Boesch, 2023). To function, it relies on the two components of localisation and classification simultaneously (Karagiannakos, 2019). Localisation is the core component where the computer vision model finds where an object is and draws a bounding box made from the intersection of two coordinates around it. Classification is where the model identifies what the object it has localised is. The combination of both processes allows object detection, for example a plant is in an image and the model knows it is a plant and not another object. Object detection is used everywhere in our lives and most people are unaware of its widespread use (Morgunov, 2020). There are many uses for this technology. These include autonomous driving algorithms, agricultural scanning, security surveillance systems, medical imaging and military tracking (Boesch, 2023).



A picture containing screenshot, map, text

Description automatically generated

A comparison of plants in dirt

Description automatically generated

Figure 1. Examples of Object Detection on lettuce at difference colour saturation levels. The red boxes are detected made the Object Detection model when trained on images of lettuce. The top selection is normal saturation level, the middle is a decreased saturation level and at the bottom is a selection with an increased saturation level.

As with any artificial intelligence, it has to "learn" based on a predetermined framework. Deep learning (DL) is preferred for object detection models compared to standard Machine Learning (ML). DL is a specialised branch within ML that mimics connections/neurons that occur in the human brain to learn effectively. Also known as deep neural networks, this efficient algorithm provides benefits such as feature automation, increased data size, faster learning and better analytics. (Sushman Biswas, 2021). DL congregates feature extraction and classification in the same neural processing task which allows for vast computing benefits (SuperAnnotate, 2021).

Although Deep Learning Object Detection (DLOD) models are powerful technology, there are certain output errors that can have a negative impact on accuracy. There include false positives and false negatives. (Figure 2.) (Kukil, 2023). An experimental study investigated ways to improve poor performance of DLOD models (Goswami, 2020). They conducted two experiments, one where they only included positive image detection which led to poor accuracy and one where they included a diverse training image input with both positive and negative image detections which improved accuracy.

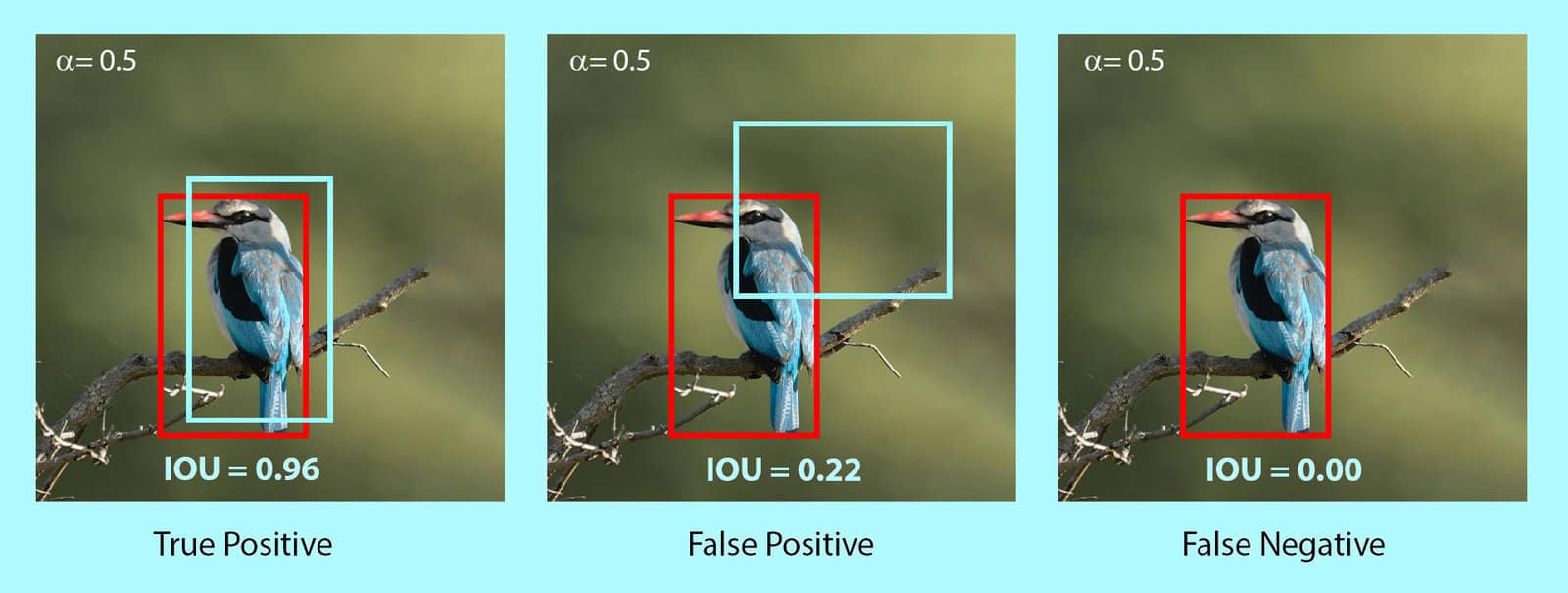


Figure 2. Types of detection. These include true positive, false positive and false negative. Note the IoU values denoting the accuracy of the detection. The higher the IoU value the more “correct” the detection (Kukil, 2023).

Hallucinations refer to a strong, confident false positive. An example of a hallucination in a DLOD model is when a bike wheel is detected in the place of a wheel, even when the wheel is not truly there (Figure 3.) This could be due to the training data including the wheel mount and when images were bound improperly. It has been confirmed that hallucinations were likely caused by poor, inaccurate labelling/annotations (Kayhan, Vredebregt and van Gemert, 2021). This put an emphasis the importance of clean, numerous, consistent image labelling.



Figure 3. Hallucination in DLOD Models. Note the high IoU for each hallucination indicating "confidence" when detecting for bike wheels. Faster RCNN and RetinaNet are different DOD algorithms, showing that each algorithm is susceptible to hallucinations (Kayhan, Vredebregt and van Gemert, 2021).

Other than adding additional images in different format for the training of a DLOD model, perhaps images could be visually altered before they are used for training. Illumination conditions are interesting as different visual parameters could potentially be used to tailor a model to a specific use-case. A limited amount of research is looking into this potential area of improvement. Significant improvements were recorded when investigating how to remove rain or snow effectively. Specifically, colour saturation was altered to increase the difference between the rain or snow from the rest of the image. This allowed the Object Detection model to accurately detect and remove the rain or snow compared to the baseline saturation (Soo-Chang Pei et al., 2014) (Figure 4.).

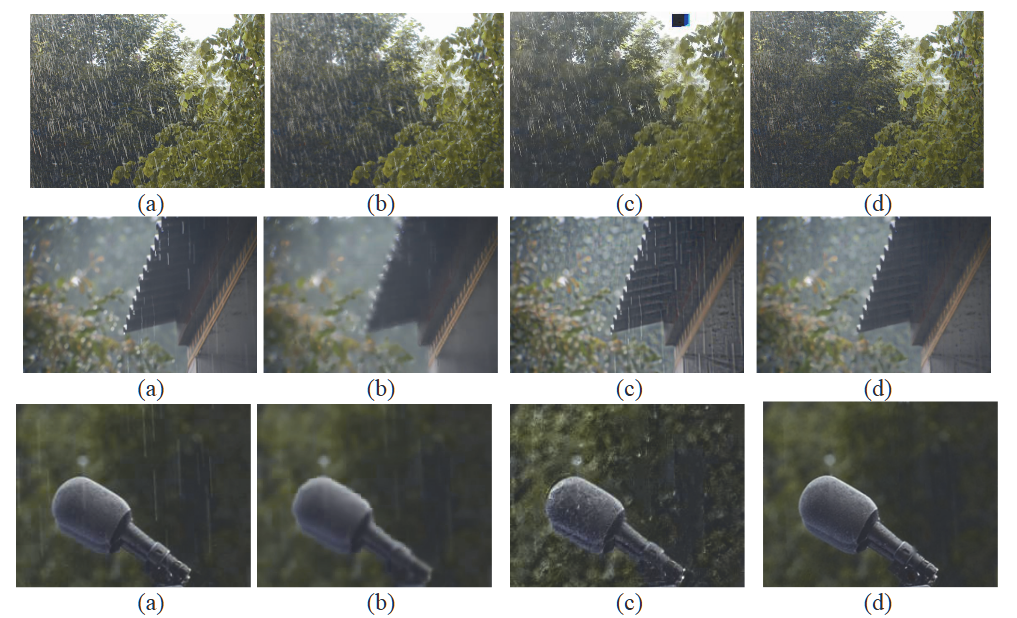


Figure 4. Images (d) used saturation altered data to effectively remove rain from images (a). Images (b) and (c) show rain removal strategies that are visually less effective (Soo-Chang Pei et al., 2014).

Altering training images can be done in various ways. For example, it has been identified that human brains primarily focus on saturation salience to detect objects and therefore improvements in DLOD models have been noticed through implementing a saliency map for image training. Research on image visualisation rendered better performance compared to state-of-the-art detection models. They specifically noted that, through positive feedback loops, their model was more effective against noise-sensitive backgrounds compared to unaltered models (Yan, 2017). Another example is improving the accuracy in a noisy scene by limiting the contrast to a smaller area of the detection object. With this visual parameter changed, they found that single model could outperform a multichannel model. This was because single channel models are more effective at predicting saliency compared to multichannel models when an object is not clearly defined or is covered by other objects. They concluded by mentioning that image contrast estimations for the target object improved the model. (Ahumada and Beard, 1997). More examples of altering the training images to improve performance include a 15 – 30% improvement by enhancing Signal over Noise Ratio (SNR) to add more information to the model per image (Becker and Cain, 2018) and 7.1% improvements by altering saturation salience at different image resolutions (Zhang, 2007).

In this project, I investigate whether altered colour saturation of training images could be a simple and effective way to improve the performance of a DLOD model. There are large impacts of improving DLOD models through visual parameter changes. It will improve edge-case detection and also allow models to be downsized to previous accuracy with faster training speeds.

**Research Question**

Is there a significant difference in total loss (TL) of a deep learning object detection (DLOD) model when trained using images with altered colour saturation?

**Research Hypothesis**

**Null Hypothesis**

There is no significant difference among the averages in the total loss metric of a deep learning object detection model when trained using images with altered colour saturation levels.

**Alternative Hypothesis**

There is a significant difference among the averages in the total loss metric of a deep learning object detection model when trained using images with altered colour saturation levels.

**Methodology**

*Overview*

Three DLOD models were trained. One model was trained on the unchanged images, another model was trained on images with decreased saturation images and a final model was trained with images with increased saturation images. Each model was retrained 10 times (10 repeats) and the Total Loss (TL) values were compared between each of the models to determine which colour saturation level-based model was the most accurate.

**Hardware and Image Collection**

The computer I used to develop and train the DLOD models for this investigation was a Windows 10 Pro Desktop PC. It has a dedicated GPU in the form of an NVIDIA 3090 Founders Edition Graphics Card with 24GB of onboard VRAM. The CPU is a Ryzen 7 5800X by AMD accompanied by 32GB of RAM. For input images, I grew lettuce to take pictures. I took the pictures with a Canon 50D. 12 images were collected and duplicated. Using Lightroom, I altered the colour saturation of these images. 12 images remained unchanged, 12 images had their saturation increased to 100 and 12 images had their saturation decreased to 100.

**Justification for Lettuce**

Lettuce plants were a suitable image subject due to numerous reasons. Firstly lettuces do not grow exactly the same way each time making their irregular shape, texture and structure difficult to detect consistently. This will allow a larger difference between strong performing DLOD models compared to weak models due to the subject difficulty. For this specific test colour saturation was altered and therefore the subject needed to display diverse colour saturation. Lettuces were well suited again as they have many colours in different patterns and arrangements.

**Coding the Model**

The development of the DLOD was conducted using the coding language Python, the learning software TensorFlow and a YouTube tutorial by Nicholas Renotte (Renotte, 2022). The first step was copying the template code from git-hub, accessing the code in the correct directory and connecting it to Jupyter Notebook version 7. Errors or incompatibilities in the code were fixed and resolved with the use of ChatGPT. The input images are imported into the [data][images] folder. The images were then scaled to 640480 pixels, and the lettuce labelled in the software called LabelMe. Labelling is the process of placing boxes to denote an object. The template code from Nicholas Renotte directly installs and opens LabelMe for seamless image labelling. Each lettuce plant or leaf in each image was labelled as accurately and precisely as possible (Figure 5). It was crucial to not label all the way to the edge of the image as the coordinates were likely to be removed when the image was transformed in the next steps.



Figure 5. Example of manual image labelling in LabelMe software.

Next the images and respective labels were partitioned into specific folders. 8 images into Train, 2 images into Val and 2 images into Test. These images were augmented to increase the amount of effective images. This process uses the labels and images to make slightly different images to increase the data size for training, validation and testing. I briefly looked into the augmented image and label sub folders in either the [train], [test] or [val] folders to ensure errors had not developed and that the images and labels correspond correctly. I used 500 epochs when training each model. This was a decent balance between training time and number of Total Loss values. Training took approximately 50 minutes for 500 epochs on my machine. After training is completed, the details of the training step are shown by running the hist. history cell. From there, the graphs were plotted directly in Jupyter Notebook to visualise the Total Loss values.

**Data Analysis**

Total Loss is the metric compared between each of the three DLOD models. 10 repeats were conducted for the training for each of the three models leading to 30 sets of TL values. Total loss is comprised of two loss metrics, being regularization and classification loss, therefore it is representative of a large part of the model’s training performance. Throughout training the model's parameters are adjusted to minimise loss. Loss quantifies the error or discrepancy between novel predictions made by the DLOD model and the true labels in the annotations. Basically, during the training step, the DLOD model repeatedly measures itself against a given benchmark and as it trains, its predictions should improve as the model trains, learns and repeats cycles. The lower the loss metric the better the training has performed as it means that the new detections are very similar to the given benchmarks.

The TL from the first 50 epochs of training is extremely variable and not highly representative of the final TL values. Instead of averaging the entire 500 epochs of TL values for each training cycle, only the last 200 epochs of TL values were averaged. By doing this, we can obtain a more reliable, representative and accurate value to measure its final performance. Only selecting the final value would also lead to inaccuracies as TL fluctuates randomly and can spike suddenly between one value and another.

TL values between three groups were compared using Analysis of Variance and a Tukey-HSD Post-hoc comparison was used to identify the difference between each of the groups in pairs. These analyses were completed using socsistatistics.com.

An ANOVA Test is the most suitable statistical analysis test for my data. Compared to a standard Student's t-Test an ANOVA Test is required for this research as the data contains more than 2 mean values (3 values). The assumptions for an ANOVA Test include random sampling, normal distribution, independent samples and equal variances.

**Results**

The average TL of the last 200 epochs for the standard saturation images was 0.00033419, for the images with no saturation was 0.000113372 and for the images with full saturation the average was 0.000427691. There is a significant difference between the averages showing that the TL values for the model trained with images without saturation are smaller than the other two models trained with standard images and fully saturated images.

|  |  |  |
| --- | --- | --- |
| Standard Saturation | No Saturation | Full Saturation |

Figure 4. Total Loss graphs for the 500 epochs of training for each DLOD model colour saturation level.

**Anova Statistical Analysis**

The p-value for the ANOVA Test was 0.001317 which is less than the 0.05 standard alpha value, therefore this test rejects the null hypothesis as there is significant difference between the mean TL values for each model (Table 1).

Table 1: ANOVA Test results.



**Post-HOC Comparison**

The Post-HOC t-Test was used to recognise specific differences between each of the data groups. With a standard alpha value of 0.05 greatly exceeding the p-value of 0.00115 when comparing the No Saturation model and the Full Saturation model. In this combination, the null hypothesis is rejected. The comparison between the Standard Saturation model and the Full Saturation model showed that there was not significant difference between the means of the two models with a large p-value of 0.46408 (Table 2). Therefore in this combination the alternative hypothesis must be rejected, and the null hypothesis accepted. When comparing the Standard Saturation model and the Full Saturation model the null hypothesis is rejected as the p-value 0.02281 is lower than the standard alpha value.

Table 2: Post-HOC t-Test results.



**Discussion**

**Discussion of Results**

This study confirmed the effectiveness of altering colour saturation for training images to improve accuracy of a DLOD model. The small p-value of 0.00131 from the Anova test is smaller than the 0.05 standard alpha threshold and therefore the null hypothesis can be rejected as the results show that there is a statistically significant difference amongst the model’s means. The Post-HOC comparsion further confirmed that, specifically the TL values from the No Saturation model was significantly different from the other models. Even when analysing the graphs visually there is a clear difference between each of the DLOD models.

The model trained with fully saturated images has the most spikes and irregularity in the line graph and also starts at the highest TL of 1.8 indicating poor performance relative to the other DLOD models. The model with the smoothest graph is the DLOD model trained with images without saturation. The No Saturation model TL values start at the lowest value of 0.8 compared to the other models. Deep Learning Object Detection Models are known for not using colour extensively when functioning. Instead they prioritise form of objects and texture. A forced-larger emphasis on colour has shown a sharp decrease in performance (Khan et al., 2012). Perhaps the removal of colour saturation from the training images simply allows the DLOD to focus more precisely on the other factors that affect its detections the most, namely shape/form, structural features, size and texture.

**Discussion of Methodology**

The methodology was challenging to reproduce exactly for each model. The main issue was that the image labelling had to be redone from scratch each time. This led to a different number of labels for each picture of potentially lettuce plants not being labelled entirely/properly. Since completing this part of the investigation, I have tested an improved way of labelling each image the exact same for each model. The images could be labelled thoroughly once and then these labels could be copied to other images and the images renamed to suit the exact labels. The data collection steps were also highly time consuming as the Jupyter notebook kernel had to be restarted and rerun each time for a new training cycle to occur. This was completed manually however it could be automated as successive cells in the code to allow it restart and retrain repeatedly without human intervention being required.

Altogether, even when considering the potential issues with the testing, training a DLOD model on images without saturation could be a simple and effective way of improving the accuracy. More testing would be required to confirm this method fully, however current literature supports the fact that colour is not significantly used in the models processing and could become a hinderance due to the increased data and therefore reduced processing speed.

**Conclusion**

The performance of a Deep Learning Object Detection model was tested when trained on images with altered colour saturation levels. This research study reveals that full saturation image training performed similarly to unchanged saturation image training. It also displayed that training on images without saturation performed significantly better than baseline when comparing the Total Loss metrics from training. This substantial improvement shows potential application to other Deep Learning Object Detection models that are used around the world. Further research into no-saturation image training with a different model architecture, larger sample size and larger model sizes could confirm whether these improvements can be found repeatedly. By leveraging similar techniques in other models, artificial intelligence systems around the world could be made more reliable, robust and efficient, benefiting many industries such as healthcare scanning, transportation optimisation, agriculture and surveillance.

**Appendix**

DLOD Code: <https://github.com/Jolenexe/Science-Extension-DLOD-Code>

The Total Loss metric graphs can also be found in the same GitHub repository under the name Total Loss Metrics from Training.docx

**References**

Boesch, G. (2023) *Object detection in 2023: The definitive guide*, *viso.ai*. Available at: <https://viso.ai/deep-learning/object-detection/#:~:text=Object%20detection%20is%20a%20key,in%20visual%20images%20or%20videos>. (Accessed: 10 July 2023).

Goswami, S. (2020). False Detection (Positives and Negatives) in object detection, *False Detection (Positives and Negatives) in Object Detection.* [online] p.15. doi: <https://doi.org/10.48550/arXiv.2008.06986>.

*Introduction to object detection with Deep Learning* (no date) *SuperAnnotate*. Available at: <https://www.superannotate.com/blog/object-detection-with-deep-learning> (Accessed: 10 July 2023).

Karagiannakos, S. (2019) *Localization and object detection with deep learning*, *Medium*. Available at: <https://towardsdatascience.com/localization-and-object-detection-with-deep-learning-67b5aca67f22> (Accessed: 10 July 2023).

Kukil (2023) *Intersection over union IOU in object detection segmentation*, *LearnOpenCV*. Available at: <https://learnopencv.com/intersection-over-union-iou-in-object-detection-and-segmentation/> (Accessed: 10 July 2023).

Pei, S.-C., Tsai, Y.-T. and Lee, C.-Y. (2014). Removing rain and snow in a single image using saturation and visibility features. *Removing rain and snow in a single image using saturation and visibility features’*. doi: <https://doi.org/10.1109/icmew.2014.6890551>.

Renotte, N. (2022). Build a Deep Face Detection Model with Python and Tensorflow | Full Course. [online] www.youtube.com. Available at: <https://www.youtube.com/watch?v=N_W4EYtsa10&t=4295s> [Accessed 21 Jul. 2023].

Morgunov, A. (2020). How to Train Your Own Object Detector Using TensorFlow Object Detection API. [online] neptune.ai. Available at: <https://neptune.ai/blog/how-to-train-your-own-object-detector-using-tensorflow-object-detection-api>.

Kayhan, O.S., Vredebregt, B. and van Gemert, J.C. (2021). Hallucination In Object Detection — A Study In Visual Part VERIFICATION. [online] IEEE Xplore. doi: <https://doi.org/10.1109/ICIP42928.2021.9506670>.

Pan, C. and Yan, W.Q. (2020). Object detection based on saturation of visual perception. Multimedia Tools and Applications, 79(27-28), pp.19925–19944. doi: <https://doi.org/10.1007/s11042-020-08866-x>.

Zhang, W., Zelinsky, G.J. and Samaras, D. (2007). Real-time Accurate Object Detection using Multiple Resolutions. Real-time Accurate Object Detection using Multiple Resolutions. doi: <https://doi.org/10.1109/iccv.2007.4409057>.

Becker, D. and Cain, S. (2018). Improved space object detection using short-exposure image data with daylight background. Applied Optics, 57(14), p.3968. doi: <https://doi.org/10.1364/ao.57.003968>.

Ahumada, A.J. and Beard, B.L. (1996). Object detection in a noisy scene. doi: <https://doi.org/10.1117/12.238715>.

OpenAI (2023). *ChatGPT*. [online] chat.openai.com. Available at: <https://chat.openai.com/>.