# Team 1 ECE 253 Final Project - Light influence on image processing for Neural Network

Yalu Ouyang
UC San Diego
yaouyang@ucsd.edu

Yin Lei
UC San Diego
y6lei@ucsd.edu

## Abstract

*Despite the popularity of image classification, currently little research conducted in the field takes the light conditions of images into account, and most image datasets neglect the importance of collecting pictures taken under different light settings. This bias could have dire influences if the light conditions indeed bear a significant influence on neural networks' performances. The goal of this project is to determine what effect various ambient lighting conditions have on state-of-the-art neural networks' ability to classify images, and whether their performance will improve on light-adjusted test sets. Having trained four Vision Transformer (ViT) models on datasets with different lighting characteristics, we evaluated and compared these models' performances on different test image sets. We then processed the test sets using three different image-processing algorithms and compared the performances of our networks on each of them. We found that ambient light indeed has a significant influence on neural networks' accuracies. We also found that the neural network trained with dark images performs much better on well-illuminated pictures than the neural network trained with well-lit images on dark pictures. However, pre-processing the test images proves to be quite time-consuming and does not always guarantee an improved result.*

## 1. Introduction

Image classification is an important and heavily-studied field in machine learning research due to its wide-range applications. Over the years, more and more advanced models that are trained on larger and larger datasets have been proposed that aim to improve accuracy. In the past decade, architectures such as ResNet [5], YOLO [4], and U-Net [9] have stood the test of time and become the most dominant CNN models in the field. In 2020, Dotsovitskiy [2] proposed applying transformer models for image classification, and since then Vision Transformer (ViT) models along with models using transformer backbones have also become alternative trending architectures for this task. However, despite the advancement of the field and the extensiveness of research, most benchmark datasets consist only of images that place objects in well-illuminated environments. The main reason for this biased choice is because pictures in abnormally lit settings are very difficult to collect. Besides, in most application scenarios for image classification, ambient light is always artificially provided. While this bias may be harmless in most situations, under certain circumstances light conditions can play a vital role. For example, in autonomous driving, night-vision object detection and classification are crucial for the passengers' safety. Night-vision classification is also important for detecting and classifying nocturnal animals in the wild. In these scenarios, the potential limitations with respect to these models' abilities to cope with under-lit images can lead to undesirable or dire consequences. In fact, in the field of traffic detection and traffic sign classification, a significant number of researchers are now devoting themselves to addressing this issue, and there have been night-time datasets arduously collected for this purpose. Despite their effort, the amount of dark image datasets is negligible in comparison to the huge volume of well-illuminated data, and virtually no images under other different light conditions systematically, except appearing here and there as outliers in well-illuminated datasets.

In our project, we seek to address this question by evaluating the influences of light conditions on performances of state-of-the-art classification models. We seek to not only know whether neural networks recognize dark pictures taken at night differently from the well-illuminated ones, but also test whether they will exhibit the same performance on a wider range of light setting, such as those taken on foggy or rainy days. We also want to test whether processing images to make them more correctly exposed before training and testing a neural network. To achieve this purpose, we collected over 15,000 images of wild animals over 11 categories. We then separated and processed these images to create in total 16 datasets for different training

and testing purposes. We then developed four ViT models on them: Glow-ViT, Glow-ViT-dark, Glow-ViT-illuminate, and Glow-ViT-mix. Finally, we processed the pictures of different test sets, each with 3 different algorithms, and compared the performances of our models on each of them. Our result shows that the light condition indeed plays an important role in machine learning's recognition of images, as each model perform drastically different to test sets with different light characteristics. However, we weren't able to find an ideal processing technique that will always guarantee an improved result when it is applied to the test set. We hope that our result shows that ambient light is an important factor that should be taken into account when training and testing one's models. We also suggest that it is possible to improve a model's accuracy and reduce the need for hard-to-get, unusual training images by applying an image processing technique to the test datasets without considerably increasing the processing time. Perhaps in the near future, more advanced image processing algorithms will emerge and help with the realization of this goal.

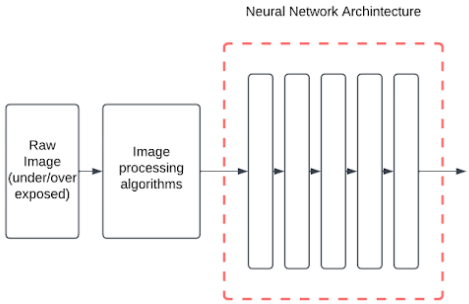


Figure 1. A proposed architecture that has the promise to facilitate training/testing and reduce cost

## 2. Previous Work

### 2.1. Image Classification

Image classification, the task of categorizing images into predefined labels based on their content, is perhaps one of the earliest and most thoroughly researched most learning topics. However, before the rise of deep learning algorithms, particularly convolutional neural networks (CNNs), image classification remained a field exclusive for image experts. Traditional machine learning algorithms required specifying important features in a bottom-up fashion, which required expert knowledge and were often computationally expensive. Neural networks, on the other hand, have the ability to automatically learn hierarchical features directly from raw pixel data, making them a powerful tool for image analysis. Since 2012, when the AlexNet was proposed by Krizhevsky et al. [1], CNN models have rapidly developed into the most popular approach and remained in the lead. It was only until recently that people began to apply transformer models for image recognition purposes as well. The innovating attempt was made by Dosovitskiy [2], who pointed out that one can divide an image into smaller patches and then treat them as one continuous vector as one would in natural language processing. While ViTs tend to outperform CNNs on larger datasets, CNNs are more efficient with memory and are better suited for resource-constrained environments. In fact, we tried to train ViT models for object detection at first (DeTR), but failed due to the limitation of computational resources.

### 2.2. Night-time Object Detection and Classification

Despite the development of image classification and object detection, very little research has been committed to the light effect of images on the performance of neural networks. The major reasons for this phenomenon are:1) most application scenarios of image recognitions, for example, facial recognition, have sufficient light exposure; and 2) pictures from abnormal light conditions are difficult to collect and hard to find. For example, none of the largest image datasets, including Cifar 10 [7] and ImageNet [12], contain even a portion of pictures taken under unusual light settings. This posed a huge challenge for us when we tried to find datasets to conduct experiments. The only datasets we can find that contain at least some unsually lit pictures are restricted to two domains: nighttime traffic and wild animal detection.

Traffic recognition is another fast growing field due to the growing demand for intelligent driving vehicles. In this field nighttime vision is paid a significant portion of attention, as driving safety requires the vehicles to be able to recognize surroundings as accurately as during the day and make the same sensible reactions. For this purpose, a handful of good-quality nighttime traffic datasets have been carefully collected, annotated and made available. Here we list two of the exemplary datasets: one is provided by Kaggle [6] and the other is pushed to Roboflow [3]. Besides attempts to collect datasets for training, other attempts have also been made to address this issue. For example, in a 2023 paper, Schutera et al. [10] proposed an online image-to-image translation method to convert nighttime traffic images to daytime ones to address the issue of overfitting of their datasets on daytime pictures. Others also attempted to artificially synthesizing nighttime traffic images via machine learning to enlarge the training datasets, and this gives inspiration to our own later approaches.

Besides traffic recognition, night-time vision is also emphasized in the field of wild animal recognition. This is the combined result of 1) very little artificial light in the wilderness and 2) the need to photograph and recognize nocturnal animals. Among wild animal datasets, there are two very good-quality ones we would like to mention.

Both of them are collected and annotated by an NGO called WildMe. The two datasets contain in total 9190 high-quality pictures of leopards and hyenas, and the light conditions of them vary greatly from picture to picture. These two datasets can be exceptionally good choices for future training purposes, and their information can be found below [14] [13].

Besides good-quality datasets, another way to facilitate animal recognition is through taking infrared photos at night, as gray scale pictures are an economic way that can bring out the contrast of the images that are originally blurred. In 2024, Wang et al. proposed a model based on Yolo-v8 to make sure the model can detect animals on gray scale pictures as well as on rgb pictures [11].

### 2.3. Under/overexposed image Processing

Image adjustment to make it correctly exposed has been a long-lasting research interest in the field of image processing. The most traditional method in this attempt is the well-known histogram equalization. By recording each pixel value and redistribute them according to their CDF, histogram equalization (HE) offers an easy and effective way to bring out the color contrast in the image. However, despite its merits, it can drastically amplify noises in the picture and distort the original color since it redistributes the color values globally in an even way. To improve on histogram equalization (HE), adaptive histogram equalization (AHE) and contrast-limited adaptive histogram equalization (CLAHE) have been later introduced. To address the problem that HE overlooks local color characteristics, these two AHE algorithms divide a whole pictures into smaller patches and perform local histogram equalization to them. Upon completion, the generated new tiles are "stitched" together using bilinear interpolation so that they will not look drastically different and unfit together. CLAHE proves to be a very effective way of processing images, and it remains a popular approach even at present.

With the development of neural networks, most state-of-the-art image processing algorithms rely on machine learning to help with image color adjustment. The merit of machine learning approaches is that the models can learn about the natural color representation during the training process, and thus it can not only bring out contrast of the image but also make it look more natural. Among all the models, the one we choose to replicate was proposed by Afifi et al. in 2021 [8]. The AFIFI model is "a coarse-to-fine deep neural network (DNN) model" trained on "over 24,000 images exhibiting the broadest range of exposure values to date with a corresponding properly exposed image". The wide scope of the training set gives the model a flexibility to handle images under various light conditions properly.

## 3. Methodology

Our starting idea is very simple: We want to create a variety of datasets, train several models on them, and test the performances of these models on a variety of test sets. By comparing the performances of each of these models on each dataset, we can gain insight into what each neural network has learned from the training dataset, and under what conditions they perform best.

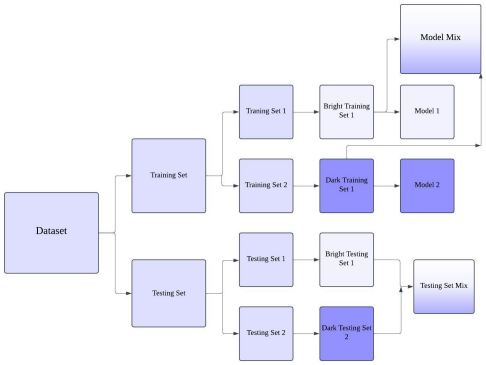


Figure 2. block-diagram illustration of our methodology

An important part of our project is creation of our own datasets. As is discussed previously, there are very few good-quality image datasets with a variety of light conditions. The only datasets are the two provided by WildMe [14] [13]. We first tried to use only these two datasets to train a model on object detections. Unfortunately, it turned out that DeTR models require much more computational resources than we could afford, so we had to abandon this attempt and try classification instead. If we try classification, however, there are simply not enough kinds of animals. Besides, while the aforementioned datasets contain unusually lit images, there are subtle differences among them and we thought that it would be over-general to put them into one category. As a result, we decided to find more animal datasets online that contain only well-illuminated pictures, and artificially synthesize the kind of datasets we want.

During our search for ideal pictures we found that most images with abnormal light conditions fell into four categories: there are *underexposed pictures*, where the pictures are shot in a bright environment but not enough light is taken by the camera, so the background is not dark, there is much color contrast, only the brightness is low; there are *less saturated pictures*, which are usually taken on foggy or rainy days. In these pictures the brightness is okay, but the colors are dull and less rich; there are *very dark pictures*: these pictures are usually taken at late night with no light around, so brightness, saturation, ambient color are all extremely low; and there are *grayscale pictures*.

To synthesize our algorithm, we first separated the

leopard and hyena dataset. We converted the pictures into gray scale and calculated the mean and median pixel values as indicators of brightness. If the mean value is extremely low (below 20), we categorized them as *very dark*. On the other hand, if the mean value is sufficiently high (above 80), we put them into the *well-illuminated* category. For the picures in the relatively dark region, we compared the difference between their mean value and median values and empirically labeled those whose mean and median pixel values differ by less than 15 *less saturated* and the other *underexposed*. For the rest of the dataset, we separated all the well-illuminated pictures according to a ratio of 0.6: 0.1: 0.1: 0.1: 0.1. we then filtered each of the smaller subsets with a standard photoshop patch action. To generate *underexposed images*, we only need to lower the brightness level; to obtain *less saturated images*, we applied a photo shop filter called *foggy night*–which dampened the color of the image drastically– and slightly lowered the saturation level. Synthesizing *very dark images* required us to first darken the whole surrounding (especially the background) and then bring down the brightness, saturation, and contrast, and we achieved this by applying a filter call *nightfromday*, lowering brightness level to the minimum, modifying the contrast value and the saturation level slightly, and finally applying a dark filter to offset the blue hue from the *nightfromday* mask. We then annotated these datasets and uploaded them to huggingface.

Having prepared our datasets, we trained four ViT models based on different image datasets. The details of the models will be provided in the subsequent section **simulations**. Finally, we processed the test datasets with three algorithms: HE, CLAHE, and AFIFI. Each algorithm has their pros and cons with differing light conditions. Again, these would be further discussed in the subsequnt sections.

## 4. Simulations

The following sections contain the performances of Glow-ViT variants on different datasets. The four Glow-ViT models were trained on *original*, *very-dark*, *well-illuminated*, and *mixed*. The main focus is on their performance on the following datasets with dark images of varying types : *very-dark*, *grayscale*, *less-saturated*, and *underexposed*.

### 4.1. Datasets

The pictures are mainly collected from three sources: WildMe, Image-CV, and kaggle. After collecting them we separated them and processed them to generate dataset with varied light conditions. The table below shows the details of our training and testing datasets.

Upon each test dataset, we performed three different image processing technique: HE, CLAHE, and AFIFI. Each method performs drastically different under different lighting conditions. In figure 5, we show a group of examples of


Figure 3. Well-lit images and 4 different kinds of dark images

comparison that illustrates the general pattern of the whole dataset after processing. For our very dark images, HE effectively brings out the color contrast, but it visibly introduces much noise and distorts the natural color pattern. CLAHE performs decently despite introducing minor noise. AFIFI generates very little noise. However, perhaps due to the fact that the model is not generally trained with such black images, the resulting image it generates is still relatively dark. Besides, it picked up the blue hue we created when we applied the *nightfromday* filter and restored it.

For the less saturated datasets, again HE makes the image unnaturally bright and introduces much noise. CLAHE, unlike its prior performance, is unable to brighten the picture since locally the color contrast is diminished. AFIFI, on the other hand, is able to restore the picture very well, bringing out the color contrast in a natural way.

Finally, for the underexposed images, HE bears a similar result. CLAHE does a decent job again, but the noise is tolerably visible. AFIFI again gives us the best result, restoring the original picture to its natural color while introducing very little noise.

In conclusion, AFIFI's performance far exceeds the other two in most cases. However, when the picture is too dark, it can create overly dark images due to prior training constraint. the performance of CLAHE is fairly decent, bringing out the color contrast at the expense of introducing tolerable noises. Nevertheless, when the image is less saturated, it may fail to pick up the local difference and produce dull images. HE is very effective at creating bright and rich pictures, but the colors are invariably unnatural and accompanied with significant noise.

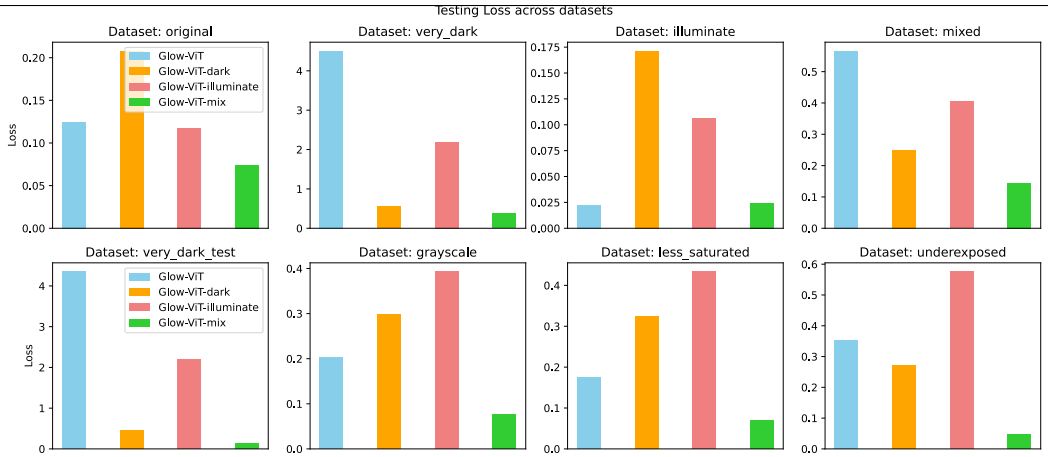


Figure 4. Glow-ViT model testing losses on all datasets

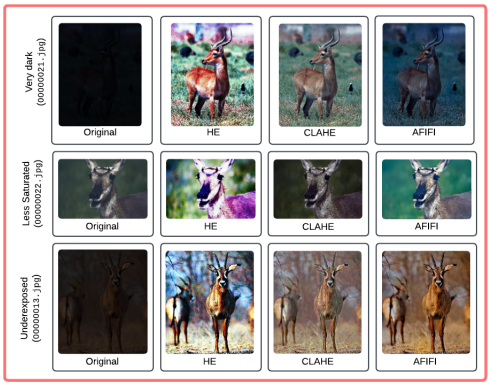


Figure 5. Examples of the three filtering methods (HE, CLAHE, AFIFI) we employed on the dark image datasets

| Dataset | Number of images | Size |
|---|---|---|
| Original | 12,101 | 1.2 GB |
| Very Dark | 2,340 | 36.4 MB |
| well-illuminated | 7,774 | 676 MB |
| mixed | 13,191 | 1.8 GB |
| verydark-test | 1,318 | 104 MB |
| grayscale | 1,825 | 673 MB |
| less-saturated | 1,137 | 229 MB |
| underexposed | 1,137 | 120 MB |

Table 1. Training and testing dataset information

### 4.2. Pure Glow-ViT models

The following table (2) and bar graph (4) illustrate the performance of all four Glow-ViT variants on these different datasets. A point of interest is that Glow-ViT-Dark's performance on bright image datasets was quite decent, with Cross Entropy losses staying below 1 and being in the same magnitude as the other models (all of which had bright images taking up a significant portion of their training set). This is especially surprising because the training set of darker images is much smaller both in image numbers and storage space due to the rarity of available images. A potential explanation for the fact that it generated very much comparable result to the other models trained on larger datasets is that Glow-ViT-Dark was able to grasp the fine difference patterns of pixel values in the image, so in bright images, when these patterns are amplified, the model is still able to grasp them despite occasional overfitting. On the other hand, Glow-ViT-Dark's performance on bright images was much better than Glow-ViT and Glow-ViT-Illuminate's performances on very dark images, as both logged Cross Entropy losses exceeding 1. The results suggest that the presence of dark images (compared to standard well-lit images) in the training set may have contributed to more robust ViT classification models, as shown by Glow-ViT-Dark and Glow-ViT-Mix.

| Dataset | Glow-ViT | Glow-ViT-Dark | Glow-ViT-Illuminate | Glow-ViT-Mix |
|---|---|---|---|---|
| Original | 0.1246 | 0.2078 | 0.1177 | 0.0744 |
| Very Dark | 4.5018 | 0.5554 | 2.1939 | 0.3715 |
| Illuminate | 0.0225 | 0.1712 | 0.1064 | 0.0243 |
| Mixed | 0.5657 | 0.2502 | 0.4042 | 0.1449 |
| Very Dark Test | 4.3495 | 0.4528 | 2.2027 | 0.1396 |
| Grayscale | 0.2021 | 0.2992 | 0.3929 | 0.0776 |
| Less Saturated | 0.176 | 0.3231 | 0.4334 | 0.0707 |
| Underexposed | 0.351 | 0.2702 | 0.5756 | 0.0472 |

Table 2. Evaluation losses across Glow-ViT models for different datasets.

### 4.3. Glow-ViT models with filtering

The following are evaluation results (6, 7, 3, 4, 5) we obtained on the same dark image datasets filtered with the following three methods: Histogram Equalization (HE), Contrast-limited Adaptive Histogram Equaliza-

tion (CLAHE), and the neural network model AFIFI.

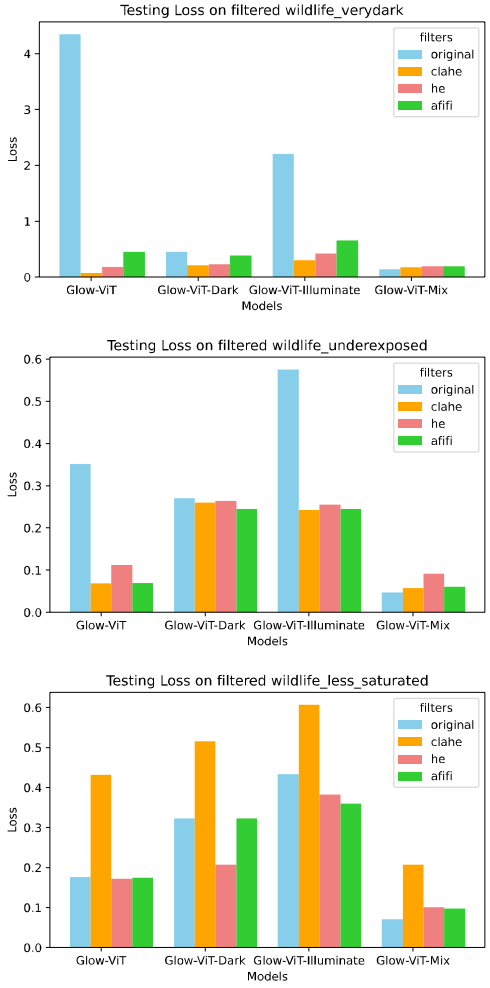


Figure 6. Testing loss of Glow-ViT models on filtered datasets

| Model | Dataset | Eval Loss | Eval Samples Per Second | Weighted F1-score |
|---|---|---|---|---|
| Glow-ViT | CLAHE | 0.075 | 18.797 | 0.9849 |
| Glow-ViT | HE | 0.1791 | 18.119 | 0.9538 |
| Glow-ViT | AFIFI | 0.4517 | 19.199 | 0.8959 |
| Glow-ViT-dark | CLAHE | 0.2089 | 18.775 | 0.9439 |
| Glow-ViT-dark | HE | 0.2324 | 18.860 | 0.9430 |
| Glow-ViT-dark | AFIFI | 0.3838 | 18.761 | 0.9144 |
| Glow-ViT-illuminate | CLAHE | 0.3039 | 19.033 | 0.9341 |
| Glow-ViT-illuminate | HE | 0.419 | 20.752 | 0.9051 |
| Glow-ViT-illuminate | AFIFI | 0.6543 | 20.576 | 0.8443 |
| Glow-ViT-mix | CLAHE | 0.1768 | 17.821 | 0.9644 |
| Glow-ViT-mix | HE | 0.1925 | 18.962 | 0.9584 |
| Glow-ViT-mix | AFIFI | 0.1952 | 18.662 | 0.9517 |

Table 3. Evaluation results for the Very-Dark-Test dataset.

As we can see, certain processing techniques brought significant gains in the accuracy of Glow-ViT models. By processing images beforehand to make it more correctly exposed, we observe that the performance of Glow-ViT-Illuminate increase drastically, resulting in a performance on par with the dark image models. However, when the model is trained on the mixed dataset, filtering techniques invariably diminish the performance of the Glow-ViT-Mix models.

| Model | Dataset | Eval Loss | Eval Samples Per Second | Weighted F1-score |
|---|---|---|---|---|
| Glow-ViT | CLAHE | 0.4322 | 19.202 | 0.9003 |
| Glow-ViT | HE | 0.1722 | 20.912 | 0.9590 |
| Glow-ViT | AFIFI | 0.1746 | 21.745 | 0.9520 |
| Glow-ViT-dark | CLAHE | 0.5156 | 20.770 | 0.8816 |
| Glow-ViT-dark | HE | 0.2069 | 22.236 | 0.9450 |
| Glow-ViT-dark | AFIFI | 0.3232 | 21.194 | 0.9233 |
| Glow-ViT-illuminate | CLAHE | 0.6073 | 20.399 | 0.8605 |
| Glow-ViT-illuminate | HE | 0.3827 | 20.900 | 0.9122 |
| Glow-ViT-illuminate | AFIFI | 0.3601 | 21.946 | 0.9199 |
| Glow-ViT-mix | CLAHE | 0.2070 | 18.978 | 0.9509 |
| Glow-ViT-mix | HE | 0.1004 | 21.674 | 0.9770 |
| Glow-ViT-mix | AFIFI | 0.0977 | 21.653 | 0.9778 |

Table 4. Evaluation results for the Less Saturated dataset.

| Model | Dataset | Eval Loss | Eval Samples Per Second | Weighted F1-score |
|---|---|---|---|---|
| Glow-ViT | CLAHE | 0.0688 | 18.855 | 0.9864 |
| Glow-ViT | HE | 0.1119 | 20.013 | 0.9753 |
| Glow-ViT | AFIFI | 0.0692 | 19.543 | 0.9802 |
| Glow-ViT-Dark | CLAHE | 0.2596 | 17.982 | 0.9436 |
| Glow-ViT-Dark | HE | 0.2636 | 19.576 | 0.9459 |
| Glow-ViT-Dark | AFIFI | 0.2450 | 20.645 | 0.9461 |
| Glow-ViT-illuminate | CLAHE | 0.2424 | 18.255 | 0.9397 |
| Glow-ViT-illuminate | HE | 0.2550 | 20.814 | 0.9412 |
| Glow-ViT-illuminate | AFIFI | 0.2446 | 20.759 | 0.9490 |
| Glow-ViT-Mix | CLAHE | 0.0577 | 19.283 | 0.9851 |
| Glow-ViT-Mix | HE | 0.0917 | 20.847 | 0.9780 |
| Glow-ViT-Mix | AFIFI | 0.0610 | 20.104 | 0.9877 |

Table 5. Evaluation results for the Underexposed dataset.

The speed of filtering images is also a major concern: Performing histogram equalization (HE) on these dataset (all around 1000 images) took 2 minutes, CLAHE and AFIFI took more than 45 and 90 minutes respectively to produce filtered results for a single dataset. This means that applying filtering to input images of neural network models will always result in a slower task completion time as opposed to only using a neural network.

## 5. Discussion

### 5.1. Review of results

We found that performance mainly increased for Glow-ViT variants trained on mostly bright images. However, simply applying these image processing algorithms on all image inputs does not guarantee a performance increase for the subsequent Glow-ViT models, and it actually lead to decreased accuracy in certain scenarios. For our well-trained model Glow-ViT-Mix, all processing algorithms (HE, CLAHE, AFIFI) led to less accuracy. We theorize that certain image characteristics may have been altered by these algorithms, thus leading to higher error for Glow-ViT models. The diagram below [8] offers a clear insight to changes image characteristics after it's been processed.

Another interesting phenomenon we notice is that neural networks do not "appreciate" the pictures the way we do. For example, while we think HE always gives the least correct restoration, in many cases its performance is best than the two more advanced techhniques. For example, it

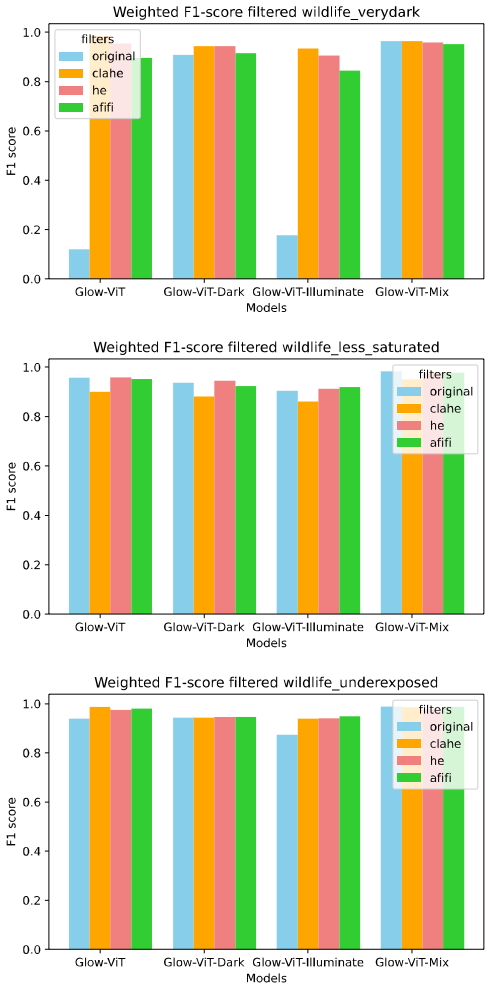


Figure 7. Weighted F1-scores of Glow-ViT models on the filtered datasets

performs better than AFIFI, our best model, in all competitions regarding the *verydark* dataset. Another phenomenon that first baffled us is that CLAHE actually hurt the performances of all models when it is applied for processing less saturated pictures. After some contemplation and discussion, we present the following hypothetical explanation: We suspect that, instead of paying most attention to the accuracy of colors and noise, the transformer model may have learned the drastic pixel value change patterns in a local area that constitute edges and boundaries. Therefore, it performs better when these patterns are amplified instead or diminished. For example, as is shown in figure 5 (5), CLAHE actually makes the *less saturated* pictures even less saturated, and AFIFI performs worst on the *verydark* test set, which it cannot bring to sufficiently bright color. However, to test this hypothesis, more research and experiments are required.

We also found the performance of Glow-ViT-Dark on bright image datasets to be a positive surprise. While it was worse than Glow-ViT and Glow-ViT-Illuminate, their overall performance are still in the same magnitude and it was much better than those two model's performance on the *very-dark* datasets. We hypothesize that this may be because, even in dark images, the main subject (or class) exhibits significantly more feature variation compared to the background, whose value range may be further compressed due to information loss caused by low lighting (think of it as a gray/black haze). Comparatively, the main subject has much richer information compared to the background, and our Glow-ViT-Dark was able to learn effectively from this.

### 5.2. Pros

Our project is the first attempt to systematically compare the influence of a variety of light conditions has on the performance of neural networks. Previously, the researchers either disregard the different lighting conditions as a determining factor, or rely solely on pictures under one setting (i.e., well-illuminated pictures), or only consider day and night pictures as different. Our result shows that light conditions indeed affect the neural networks, and that the influences are more subtle than what we previously believed to be. For example, our result clearly suggests that the models do not perform well on less-saturated images despite the fact that they are fairly bright.

One thing we did well was to train 4 different models on our synthesized datasets using the same ViT checkpoint. This allowed us to learn more about the behaviors of models, since the structure and parameters are all the same. Without doing so, we might not have discovered the significant speed advantage that training Glow-ViT-Dark had over the other models and the importance role that dataset size plays in this. Since darker images have overall smaller color range compared to bright, vivid images, this means that they can be more efficiently compressed, resulting in less storage space taken. This resulted a dataset with a fair amount of images that take up far less space than comparable bright images, thereby enabling models to train much faster and is less taxing on hardware computational power.

### 5.3. Cons

One drawback of our approach is that we only performed analysis on ViT classification models. Other models such as ResNet and DenseNet are also widely used for image classification tasks both in academia and industry, so our results here for the ViT architecture is not comprehensive enough and may not transfer to other models.

Another drawback of our approach is that although we obtained various results from Glow-ViT processing, we weren't able to extract much meaningful information that could provide answers on why certain processing algo-

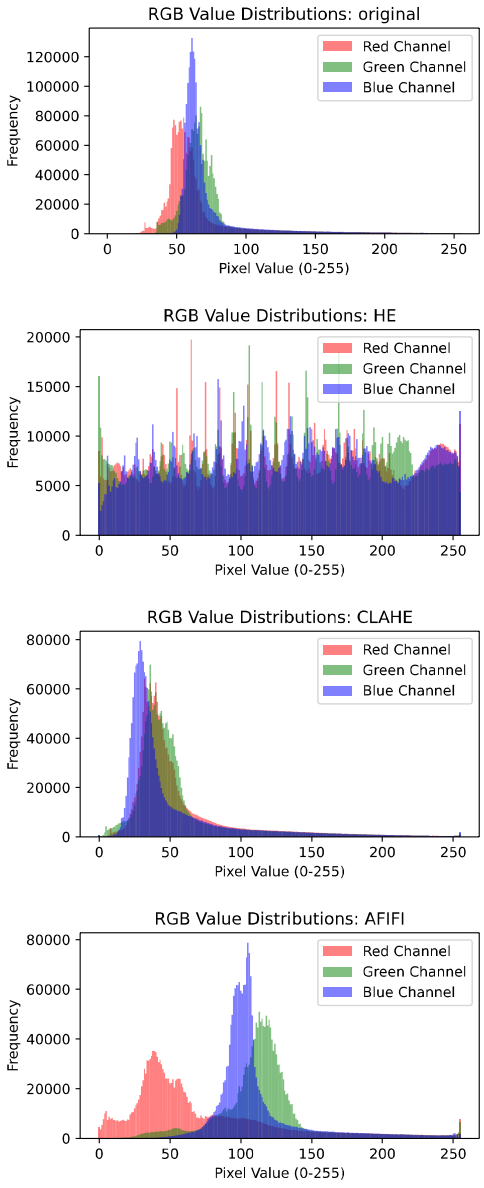


Figure 8. The RGB Value distribution of original image and filtered results.

rithms led to better/worse performance for Glow-ViT variants and how we should choose the most suitable processing algorithms for different image inputs.

### 5.4. Improvements

One area for future improvements is expanding the experiment to include other widely-used models like ResNet and DenseNet. Because these popular models mainly utilizes convolution instead of attention calculation in ViTs, they might exhibit different behaviors when presented with processed images.

Another area for future improvements is compressing the trained models with pruning. Pruning is the process of removing parameters from neural networks to make them more efficient whilst preserving accuracy. This should lead to reduced model size, improved computational efficiency, poentially faster inference (due to less parameters) and deployability on less performant hardware. Our initial goal for the project was to devise a method to improve neural network performance both in terms of accuracy as well as inference speed, so performing pruning would be very beneficial in that regards.

The preliminary outline for that experiment is as follows: Using our trained Glow-ViT models, first perform pruning on these models until it becomes unable to maintain a good accuracy whilst further trimming down parameters. These pruned models are used as the baseline, then in conjunction with image-processing algorithms, these baselines are pruned even further. This way, the experiment is structured to find maximum performance, time-wise as well as accuracy-wise.

## 6. Summary

Through the course of the project, we have identified that processing dark image inputs using techniques created to enhance image quality does not necessarily lead to better performance for the following neural network model. The existence of an extra filter sometimes even lead to diminished performance, and alongside the non-trivial - even slow - processing speed of the image-processing methods we tested, means that will be a better approach to introduce greater diversity and variation in the training dataset to enhance the neural network's performance on image classification.

The performance of Glow-ViT-Dark and Glow-ViT-Mix across all datasets highlighted the ability of neural networks to extract extra feature information from dark images. It signifies that careful evaluation on a statistical level, rather than solely on the subjective perception of the human eye, is needed when synthesizing an effective training data set for image classification.

Finally, the performance of Glow-ViT-Dark in combination with its small-footprint training dataset and speedy training makes it an ideal model for hardware-limited scenarios. Combined with histogram equalization, it offers a great blend of deployment speed from scratch and actual inference performance.

## 7. Contribution

Below are all of our team members and their contribution to the project:

- Yalu Ouyang - 50 %
- Yin Lei - 50 %

## References


[1] Krizhevsky A, Sutskever I, and Hinton GE. Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, 2012. 2

[2] Alexey Dosovitskiy. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint*, 2020. 1, 2

[3] Owen Hui. night traffic computer vision project, 2021. 2

[4] Redmon J, Divvala S, Girshick R, and Farhadi A. You only look once: Unified, real-time object detection. *arXiV preprint*, 2015. 1

[5] He K, Zhang X, and Sun J Ren S. Deep residual learning for image recognition. *arXiV preprint*, 2015. 1

[6] Kaggle.com. Nighttime driving dataset, 2021. 2

[7] Alex Krizhevsky. Cifar 10. https://www.cs.toronto.edu/~kriz/cifar.html, 2009. 2

[8] Afifi M, Derpanis K, Ommer B, and Brown M. Learning multi-scale photo exposure correction. *CVPR*, 2021. 3

[9] Ronneberger O, Fischer P, and Brox T. U-net: Convolutional networks for biomedical image segmentation. *arXiV preprint*, 2015. 1

[10] M. Schutera, M. Hussein, J. Abhau, R. Mikut, and M. Reischl. Night-to-day: Online image-to-image translation for object detection within autonomous driving by night. *IEEE Transactions on Intelligent Vehicles, vol. 6, no. 3, pp. 480-489*, 2023. 2

[11] Wang T, Ren S, and Zhang H. Nighttime wildlife object detection based on yolov8-night. *Electronics Letters. 2024;60(15)*, 2024. 3

[12] Stanford University. Imagenet. https://www.image-net.org/download.php, 2010. 2

[13] WildMe. lilawp. leopard id 2022, 2022. 3

[14] WildMe. lilawp. leopard id 2022, 2022. 3