Object Detection in Low Light : Humans vs Machines

Yash Rajbhar, Swati Khaundal, Sivashankaran S
Department of Humanities and Social Sciences, Indian Institute of Technology, Delhi

Abstract– This term paper presents the results of an investigation into the effect of varying exposure levels on the performance of an image classification model. The hypothesis was that training a model with a spectrum of black-and-white images with different exposure levels would improve its ability to classify images under low-light conditions. To test this hypothesis, we trained two models: one on a spectrum of images with varying exposure levels and another on low exposure images only. The results showed that the model trained on the spectrum of images achieved significantly better accuracy than the control model, indicating that exposure to diverse lighting conditions during training helped the model develop more robust representations of visual features. These findings have important implications for improving image classification algorithms in low-light environments and for understanding how the visual system adapts to changes in lighting conditions. The paper concludes with a discussion of the limitations of the study and suggestions for future research in this area.

I. Introduction

Human vision in low light conditions is primarily dependent on the activity of specialized cells in the retina called rods. Rods are more sensitive to light than the other type of cells in the retina called cones, which are responsible for color vision and visual acuity. In low light conditions, the iris of the eye dilates, allowing more light to enter the eye. This light is then absorbed by the rods, which generate electrical signals that are sent to the brain via the optic nerve. The brain then processes these signals to create a visual image. However, because rods are not very good at detecting fine detail or color, the resulting image is generally less sharp and less colorful than what we see in normal light conditions.

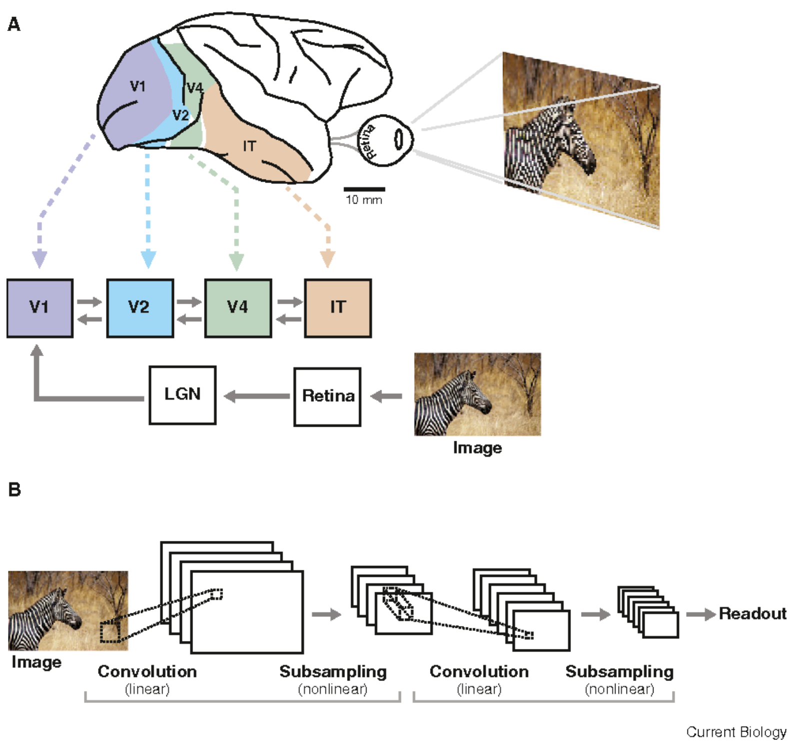
Artificial intelligence (AI) can also be trained to improve its ability to see in low light conditions, but it operates differently from the human visual system. AI algorithms can be trained using large datasets of images captured in low light conditions, which can be used to learn how to enhance and denoise images taken in similar lighting conditions. These algorithms use techniques like image processing, noise reduction, and super-resolution to enhance low light images.

Some AI systems are designed to use infrared or thermal imaging to improve their performance in low light conditions. Infrared and thermal imaging can be used to detect heat signatures and produce images based on the heat emitted by objects, which can be useful in low light conditions where visible light is limited. For example, some autonomous vehicles and drones use infrared sensors to navigate and detect obstacles in low light conditions. Similarly, some security cameras and surveillance systems use thermal imaging to detect and track people and vehicles in the dark. However, the scope of object recognition for such techniques is limited, since the set of objects that can be recognized through infrared or thermal imaging is sparse.

This calls for AI systems that are inherently better at object recognition in low light conditions without recourse to other tools.

II. Neuroscience of human vision

The different components of the neural architecture of human vision work together to process visual information and give rise to perception and object recognition. Each component, from the retina to the various visual processing areas in the brain, plays a critical role in our ability to perceive and interpret visual information.



Visual information is first processed in the retina, where photoreceptor cells detect light and convert it into electrical signals. These signals are then transmitted to the LGN in the thalamus, which relays the information to the primary visual cortex (V1) in the occipital lobe of the brain. V1 is responsible for processing basic visual features such as edges, lines, and orientation, and these features are then combined and processed further in higher-level visual areas such as V2, V4, and IT. These areas are involved in processing more complex visual features such as texture, color, and object recognition.

The interaction between these different visual processing areas is thought to be hierarchical, with information being processed and integrated at each level before being passed on to the next level for further processing. For example, information about basic visual features processed in V1 is integrated and processed further in V2, which in turn passes information on to V4 for processing of color and object recognition, and finally to IT for processing of more complex visual stimuli such as faces, objects, and scenes.

Mechanisms employed by the human visual system to adapt to low light conditions:

Contrast gain control is a process that involves the inhibition of neighboring neurons that are sensitive to different orientations or spatial frequencies, in order to enhance the contrast of objects and edges in the visual scene in low light conditions. It helps to ensure that the neurons respond optimally to a wide range of contrast levels, and it can enhance the visibility of low-contrast relevant stimuli in the visual scene.

Adaptive Pool is a mechanism that combines visual signals from multiple photoreceptor cells in the retina to improve the signal-to-noise ratio of visual information. In low light conditions, there is more noise in visual signals due to the reduced number of photons available, and Adaptive Pooling helps to reduce the impact of this noise by integrating signals from multiple photoreceptors.

Feedback Mechanism is a mechanism that involves the transmission of information from higher-level visual processing areas back to lower-level areas in the visual processing hierarchy. In low light conditions, Feedback Mechanism allows higher-level areas of the brain to provide feedback to lower-level areas, adjusting their sensitivity and enhancing their processing of visual information.

III. Importance of object detection technologies in low-light scenarios:

Object detection technologies in low light environments are an active area of research in computer vision. There are several approaches that have been proposed to address the challenges posed by low light conditions, including:

1. Image enhancement: Pre-processing the images to enhance their quality and visibility, for example by adjusting contrast, brightness, or color balance.
2. Image fusion: Combining information from multiple images with different exposure levels or from different sensors to improve the overall quality of the image and enhance the visibility of objects.
3. Infrared imaging: Using infrared sensors or thermal imaging to capture images in low light conditions where visible light is insufficient.
4. Deep learning: Using deep neural networks to learn robust representations of objects under varying lighting conditions and to improve the accuracy of object detection in low light environments.

All the techniques that has been proposed for enhancement of contrast and exposure in the images. But, there has been no attempts to understand or mimic how humans or anyother animals learn to detect objects in the dark.

There are huge number of benefits in developing better object detection technologies for low-light environments

Improved safety: Low light and night vision technologies can enhance safety in various contexts, such as driving at night or navigating through poorly lit environments. By enhancing the ability to detect and recognize objects in low light conditions, these technologies can help prevent accidents and improve overall safety.

Enhanced vision for individuals with impaired vision: People with impaired vision often have difficulty seeing in low-light conditions, impacting their daily activities. Low-light and night vision technologies can help these individuals see better in low-light conditions, improving their quality of life.

Better security: Low light and night vision technologies can be used in security and surveillance applications to improve detecting and recognizing objects in low light conditions. This can help prevent crime and improve overall security.

IV. Visual cortex and CNNs

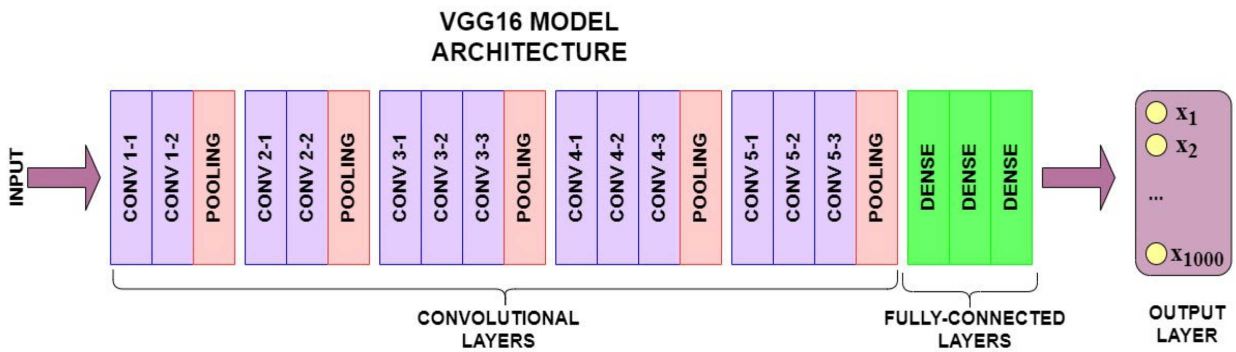
The simple cells in the visual cortex first detect lines, edges, and corners before the complex cells analyze other complex features (such as colors, shapes, and orientation), which have also been shown to have more spatial invariance (not dependent on orientation) in their responses. According to studies, complex cells combine visual information from numerous simple cells, each with a distinct preferred location. These two characteristics—selectivity to particular features and increasing spatial invariance through feedforward connection—distinguish artificial visual systems like CNNs from natural visual systems, just like how the cells in the cortex process visual information.

Numerous studies indicate that "core object recognition," or the quick identification of objects in the face of extreme variation in their appearance, is resolved in the brain by a series of reflexive, largely feedforward computations that result in a potent neuronal representation in the inferior temporal cortex. The algorithm underlying such intricate processes is, however, still largely unknown.

The idea that biological neurons behave this way or that our neural networks use the same feedforward and backpropagation model to learn is erroneous, even though the processing might appear similar in both systems. Visual perception is a linear process from the retina to other specialized processing regions. Machine learning (ML) has tried to mimic this feed-forward network but has yet to learn how biological synaptic weights are determined.

V. Vgg16

A popular neural network for image processing is the VGG-16, based on CNN. Convolutional, pooling, and fully-connected CNN layers make up the three layers of the 16-layer VGG-16. The convolution layer isolates the filter or kernel by extracting features from an input image. Convolved Feature/Feature Map has a filter size of 3X3 and learns the image features using small input data squares that preserve the connection between pixels. When the images are too large, pooling layers reduce the number of parameters, preventing the input data from being overfitting. A pooling layer follows each convolutional layer and cuts the input size in half before feeding it to the subsequent batch of convolutional layers.



VI. Data

The animals10 dataset contains information on 10 different types of animals. It contains about 28K medium quality animal images belonging to 10 categories: dog, cat, horse, spyder, butterfly, chicken, sheep, cow, squirrel, elephant. However, a sub-dataset has been created to focus specifically on only two of these animals. By focusing on only two animals, the sub-dataset allows for us to save time in testing the working of our model and our hypothesis. All of the images were in jpg or jpeg format. The two categories were randomly chosen : dog(4863 images) and cat(1668 images). For each category, a spectrum of black and white images were created with different exposure levels(100%, 70%, 50%, 30%). All images were resized to 128p*128p (p - pixels). Four dataloaders were created using tensorflow, one for each exposure level in black and white images.
Exempler Data:

VII. Model

The pretrained Keras VGG16 model was used to train an image dataset. The convolutional layers of the model had already been trained on the 'Imagenet' dataset. However, in this case, none of the layers were frozen, and all of the 14 lakh parameters were trained. This allowed for the model to be fine-tuned on the specific image dataset, potentially improving its performance on the specific classification task. By not freezing any layers, the model was able to learn and adjust the weights of the convolutional filters to better classify the images in the target dataset.

In this scenario, two models were trained to classify black-and-white images of different exposure levels. One model was trained on a spectrum of images with varying exposure levels, while the other control model was only trained on low exposure images. The hypothesis was that the model trained on the spectrum of images would perform better than the control model.
This hypothesis was based on the assumption that training the model on a diverse range of images with varying exposure levels would allow it to learn features that are robust to different lighting conditions. By contrast, the control model trained only on low exposure images may not have been exposed to enough variation to develop this level of robustness. The model was trained with an Adam optimizer with a learning rate = 0.0001. ‘Categorical cross entropy’ was chosen as the loss metric while ‘accuracy’ was chosen as the model’s metric. The model was trained for 3 epochs with 50 steps per epoch.

To evaluate the hypothesis, the performance of both models was compared on a held-out test set of images with low exposure images. If the model trained on a spectrum of images performs better than the control model, it could suggest that exposure to a range of lighting conditions during training helps the model develop more robust representations of visual features. This could be analogous to the process of dark adaptation in the human visual cortex, where exposure to a range of lighting conditions over time leads to changes in the sensitivity and responsiveness of the receptor cells, allowing for better vision in low-light conditions.

VIII. Results

The two models were trained to test the hypothesis that a model trained on a spectrum of black-and-white images with varying exposure levels would perform better than a control model trained only on low exposure images. The model trained on varying exposure levels achieved an accuracy of 90%, while the model trained on low exposure data achieved an accuracy of 75%.

The model trained on varying exposure levels performed significantly better than the control model, indicating that exposure to a diverse range of lighting conditions during training helped it develop more robust representations of visual features.

IX. Discussion

The VGG16 model can achieve a test accuracy of 92.7% in ImageNet, a dataset containing more than 14 million training images across 1000 object classes. While the VGG16 model is inspired by the visual cortex and is a simplified version of the visual processing in the human brain, it is still a highly simplified and abstract representation of this complex biological system. The convolutional layers of the VGG16 model act as the neuron layers in the visual cortex, where each neuron is responsible for processing a small patch of the visual field. However, the visual cortex comprises many different types of neurons that respond to different features and properties of visual stimuli, including color, orientation, motion, and more. In contrast, the VGG16 model used a fixed set of filters learned during training and applied to the entire image. Additionally, the fully connected layers in the VGG16 model act as the final decision-making layer, which outputs the predicted class or object. However, object recognition in the human visual system is not a pure feedforward process but involves feedback mechanisms and interactions between higher-level brain areas. These mechanisms help refine and improve object recognition by integrating prior knowledge and expectations and modulating the activity of lower-level visual processing areas.

X. Future Areas of Scope:

In humans, object recognition in low-light conditions is often aided by contextual information and prior knowledge. For example, if we are in a familiar environment, we may recognize objects even if they are poorly lit because we have prior knowledge of what objects should be present in that environment and what they should look like. Similarly, we may recognize an object’s shape even if it is partially obscured or poorly lit because we have prior knowledge of what that object should look like. Developing algorithms that utilize contextual information and prior knowledge to aid in object recognition in low light settings is a promising approach to improve object recognition accuracy in challenging lighting conditions. The key idea behind this approach is to use information about the scene and objects in the scene to guide the object recognition process. One way to incorporate contextual information is to use scene understanding techniques to segment the image into regions that are likely to contain objects of interest. Another approach is to use prior knowledge about object location and shape to guide object recognition. For example, if we know that a car is likely to be located in a certain part of the image, we can use this information to guide the object recognition process.

References

Erchova, I., Vasalauskaite, A., Longo, V., & Sengpiel, F. (2017). Enhancement of visual cortex plasticity by dark exposure.

Gandhi,T.K., Haritosh, A.,Human Visual Learning Inspired Effective Training Methods

Pokharel, M., Rajkarnikar, L., Karn, G., & Shrestha, S. (2021). Detection of Brain Tumor using VGG-16 Architecture.

Murugan, R.A., Sathyabama, B. Object Detection for Night Surveillance Using Ssan Dataset Based Modified Yolo Algorithm in Wireless Communication. Wireless Pers Commun 128, 1813–1826 (2023).

Peter J. Bex, Isabelle Mareschal, Steven C. Dakin; Contrast gain control in natural scenes. Journal of Vision 2007;7(11):12.

kaggle.com/datasets/alessiocorrado99/animals10