**INTERNSHIP RESEARCH REPORT**

**IMAGE CLASSIFICATION**

**AND**

**THEIR ROLE**

**IN DEVELOPING HIGH RANGE CAMERA SENSORS**

Submitted by

**NANDINI SAIN**

**INTERN**

**3RDITECH**

**Acknowledgement**

I would like to thank Ms Vrinda Kapoor, Founder of 3rditech for giving me the opportunity to do this internship within the organization.

I further thank Mr. Aakash Sir and Ritwik Sir for guiding and teaching me throughout the entire process. These skills would help me to expand my resume and advance my career.

**Information about the Internship position**

I joined 3rditech in the capacity of an Intern, at Research and Innovation Park, IIT Delhi, New Delhi. This internship was done in lieu of my status as a final year undergraduate student at Amity University, Noida.

I worked under the Embedded Systems team and AI/ML team which was led by Aakash sir and Ritwik Sir respectively. My main tasks were to design algorithms for the code, do connections, collect data and do research relevant information about the project that I was assigned to.

**INDEX**

1. Introduction
2. Objectives
3. Image Classification
4. Image Datasets
5. Challenges faced and Solutions
6. Natural Images vs Synthetic Images
7. Generative Artificial Intelligence
8. Generative AI models
9. An Introduction to Monochromatic Cameras for Night Vision
10. Features of night surveillance cameras
11. 3 major advantages of monochrome cameras over color cameras
12. Conclusion
13. Bibliography

**Aim**

Study on image datasets, their classification, and usage of current technologies.

**Objectives**

* Types of image datasets
* Image datasets characteristics
* Role of AI in image classification
* Usage of current technology camera sensor

**Introduction**

Since the time we have seen, we have always loved to capture the moments around us. Whether it is what we see or what we feel, we tend to look beyond the fact that a moment could be stilled. From analog cameras used by those privileged, to smart cameras on every mobile device, the rush of capturing has never settled. So naturally, the development of such cameras went just beyond from posing for pics. Over the span of 10 years, images have a greater value that a simple text, as images show the face value of an object (unless it’s AI generated but we would come back to it later.)

If one is to question how images are so valuable, the correct answers would be that images gives the viewer a visualization of what they can ought to imagine. Suppose we say to draw a tree, we won’t know how to draw it until we are given a view of it. Texts only tell us what an object is, and images inform us how a object looks like. This is the reason that in today’s time, we tend to trust images more that other sources of media.

**Image Classification.**

Images have a wide range of application in the field of computers, the most common being building machine learning algorithms which are trained on these image datasets. The most common objective of such algorithms is to have model or technology developed which work on the principle of image classification.

Image Classification is defined as the task of assigning a label or class to an entire image. Images are expected to have only one class for each image. Image classification models take an image as input and return a prediction about which class the image belongs to. This is crucial especially when the objects are assumed to some something else and are something else. it is used to identify an object that appears in an image. This task consists of labelling input images with a probability for the presence of a particular visual object class. To put it simply, it is a task where the system takes an input image and classifies it with an appropriate label. The label is always from a predefined set of possible categories.

Image classification differ from image recognition as Image recognition focuses on identifying and locating specific objects or patterns within an image, whereas image classification assigns an image to a category based on its content. In essence, image recognition is about detecting objects, while image classification is about categorizing images.

Image Classification is a concept of Hidden Markov Model. This approach considers feature vectors statistically dependent through an underlying state process assumed to be a Markov mesh, which has transition probabilities conditioned on the states of neighboring blocks from both horizontal and vertical directions. This allows the model to reflect dependency in two dimensions simultaneously.

Key Steps

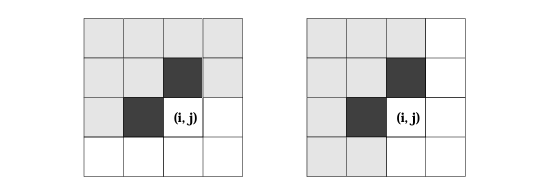
1. **Training:**

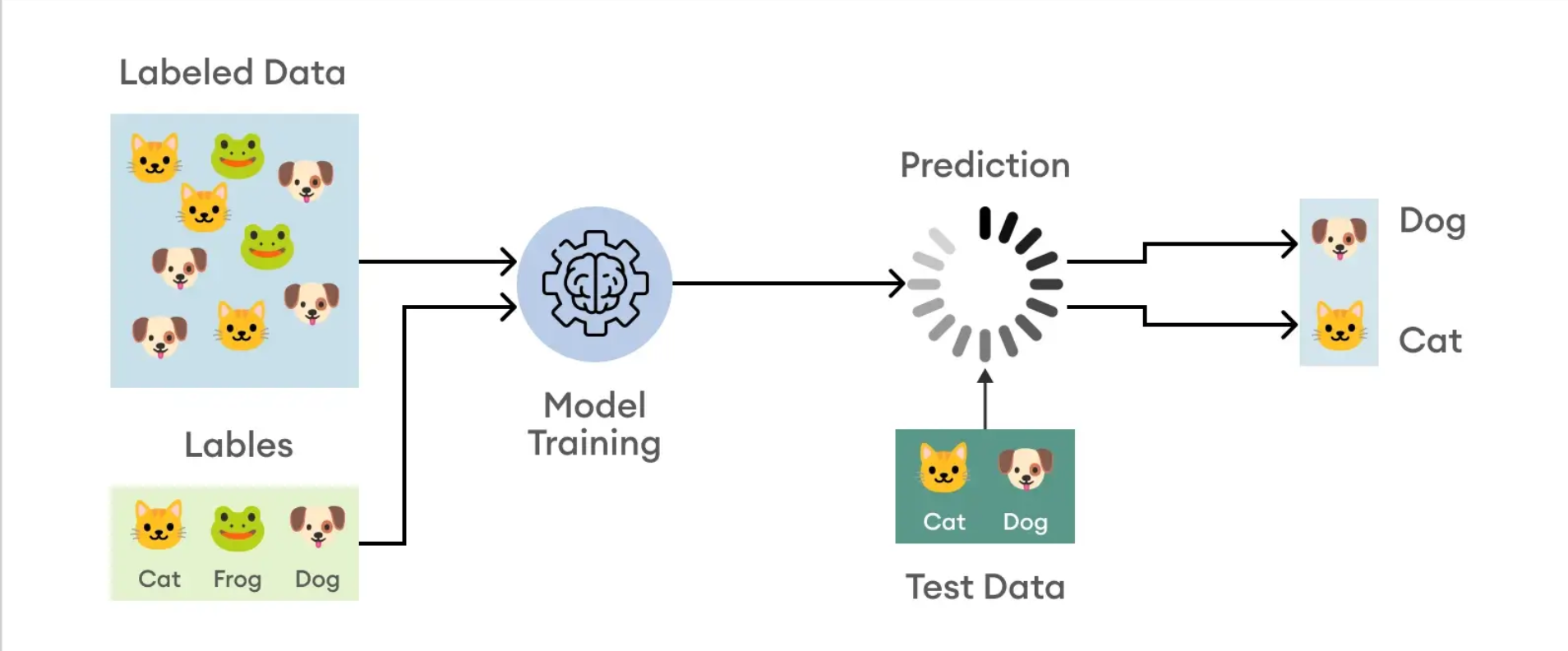
* Divide training images into non-overlapping blocks with equal size.
* Extract a feature vector for each block.
* Select the number of states for the 2D HMM.
* Estimate model parameters based on the feature vectors and their hand-labeled classes.

1. **Testing:**

* Generate feature vectors for the testing image.
* Search for the set of classes with maximum a posteriori probability given the feature vectors according to the trained 2D HMM

As in all blocks-based classification systems, an image to be classified is divided into blocks and feature vectors are evaluated as statistics of the blocks. The image is then classified according to the feature vectors. The 2-D HMM assumes that the feature vectors are generated by a Markov model which may change once every block. The classification algorithm attempts to end the optimal combination of classes jointly for many blocks at once. A one-dimensional approach of joint classification, assuming a scanning order in classification, is usually suboptimal.

****

****

**Image Datasets**

The most important component of image classification are the images itself. In layman terms, an image can be described as a visual representation of something. In computer science, an image is a visual representation of something that has been created or copied and stored electronically. It can be a photograph, graphic, or video frame. To a computer, an image is a series of ones and zeros, or binary representation of visual data.

Images are stored in the following format

1. JPEG (Joint Photographic Experts Group)
2. GIF (Graphics Interchange Format)
3. PNG (Portable Network Graphics)
4. SVG (Scalable Vector Graphics)
5. TIFF (Tag Image File Format)

An image dataset is a collection of images that are used to train machine learning models. It serves as the input data for algorithms, allowing them to learn patterns, features, and relationships within the images. Image datasets can range from a few hundred to millions of images, capturing different objects, scenes, or concepts. These datasets are annotated with labels or annotations, providing ground truth information for training and evaluation. A dataset in therefore assembles a collection of images that are labelled and used as references for objects in the world, to ‘point things out’ and name them.

Image datasets are commonly used in tasks such as image classification, image recognition, and object detection in the field of Computer Vision. These datasets can be obtained from two sources: online and manual collection based on specific task requirements. Image datasets are usually stored in image files such as jpg and png, and contain information such as images, annotation files, and semantic images. The most used image datasets for different tasks include

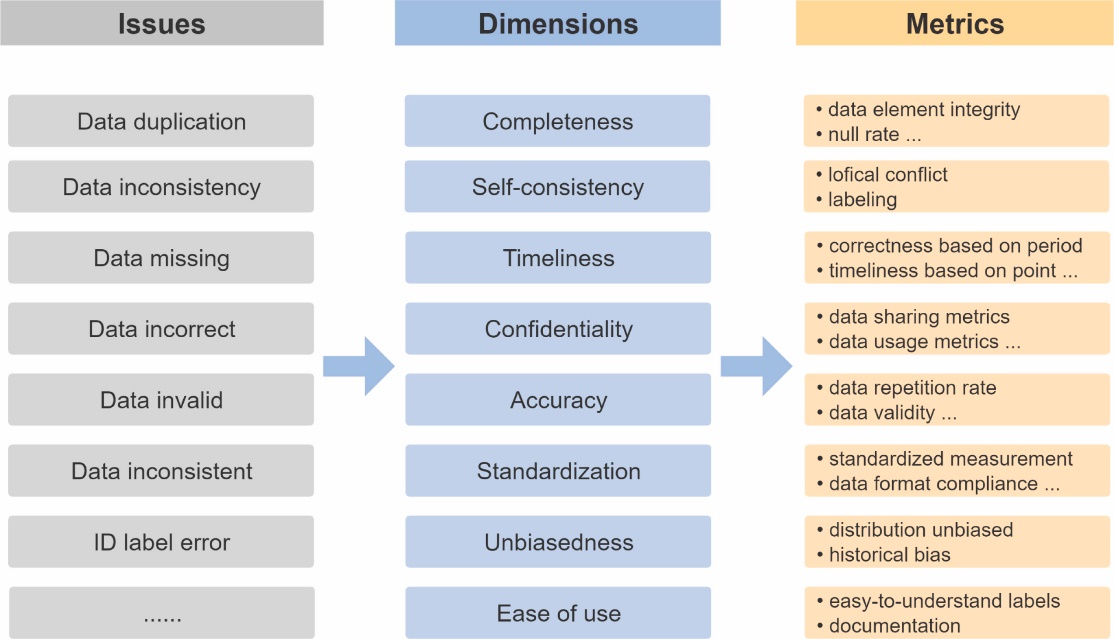
* Labelme datasets
* Pascal VOC datasets
* ImageNet
* Stanford canine datasets
* Places datasets
* CIFAR datasets

For image recognition tasks, commonly used datasets include

* ImageNet
* COCO datasets
* COIL100 datasets
* Places datasets
* CelebFaces datasets

Datasets such as ImageNet are built on an array of practices of mediation of photography: collecting, labelling, composing, assembling images and distributing them. The ImageNet project for instance is a collection of tens of millions of images manually annotated, sorted and organised according to certain conditions.

**Challenges faced and Solutions**



Data Images captured in real-world environments exhibit variability in factors such as lighting conditions, viewpoint, occlusion, and background clutter. Adapting models to handle such variability is essential for achieving robust performance across different scenarios.

The main challenges in image classification are the large number of images, the high dimensionality of the data, and the lack of labeled data. Images can be very large, containing a large number of pixels. The data in each image may be high-dimensional, with many different features. There is also a shortage of labeled data, which is data that has been annotated with the correct class or category.

The following are the most significant issues

* Intra-class variation
* Variation in Scale
* Variation in perspective
* Illumination
* Clutter in the background

Here are some effectives solutions to tackle the issues.

1. **Data quality**

One of the most important factors that affect the performance of image classification algorithms is the quality of the data. Data quality refers to the quantity, diversity, relevance, and cleanliness of the images used for training and testing the algorithms. If the data is insufficient, biased, noisy, or irrelevant, the algorithms might fail to learn the general patterns and features that distinguish different classes of images. For example, if the data is skewed towards a certain class, the algorithms might overfit to that class and ignore the others. If the data contains irrelevant or misleading information, such as watermarks, logos, or text, the algorithms might confuse them with the actual content of the images. To ensure data quality, you need to collect enough images that represent the classes of interest, balance the distribution of the classes, preprocess the images to remove noise and artifacts, and augment the data with transformations such as cropping, rotating, scaling, or flipping.

1. **Model complexity**

Another challenge that image classification algorithms face is finding the right balance between model complexity and generalization. Model complexity refers to the number and size of the parameters that define the algorithm, such as the layers, filters, and neurons in a neural network. A more complex model can capture more details and nuances of the images, but it also requires more data and computational resources to train and test. Moreover, a more complex model might overfit to the training data and fail to generalize to new or unseen images. A less complex model, on the other hand, might underfit to the data and fail to capture the essential features and patterns of the images. To find the optimal model complexity, you need to use techniques such as cross-validation, regularization, dropout, and early stopping.

1. **Image variability**

A third challenge that image classification algorithms face is dealing with the variability and diversity of the images in real-world scenarios. Image variability refers to the changes and variations that occur in the images due to factors such as lighting, perspective, occlusion, background, scale, pose, and expression. For example, an image of a dog might look different depending on the angle, distance, illumination, and surroundings of the camera. Image variability makes it harder for the algorithms to recognize and classify the images based on their content. To cope with image variability, you need to use techniques such as feature extraction, feature normalization, feature selection, and feature fusion.

1. **Evaluation metrics**

A fourth challenge that image classification algorithms face is choosing and interpreting the appropriate evaluation metrics. Evaluation metrics are used to measure and compare the performance of the algorithms on a given dataset or task. However, different metrics might have different assumptions, limitations, and trade-offs. For example, accuracy is a common metric that measures the percentage of correctly classified images, but it might not be suitable for imbalanced datasets or multiclass problems. Precision and recall are other metrics that measure the ability of the algorithms to identify the relevant images and avoid false positives and negatives, but they might conflict with each other. F1-score is a metric that combines precision and recall, but it might not reflect the cost or importance of different errors. To choose and interpret the evaluation metrics, you need to consider the characteristics of the dataset, the objective of the task, and the expectations of the users.

1. **Ethical and social implications**

A fifth challenge that image classification algorithms face is addressing the ethical and social implications of their use and impact. Ethical and social implications refer to the potential benefits and harms that the algorithms might have on individuals, groups, and society as a whole. For example, image classification algorithms might enable positive applications such as improving health care, enhancing education, or protecting wildlife. However, they might also enable negative applications such as invading privacy, discriminating against minorities, or spreading misinformation. To address the ethical and social implications, you need to follow principles such as fairness, accountability, transparency, and privacy.

**Natural Images vs Synthetic Images**

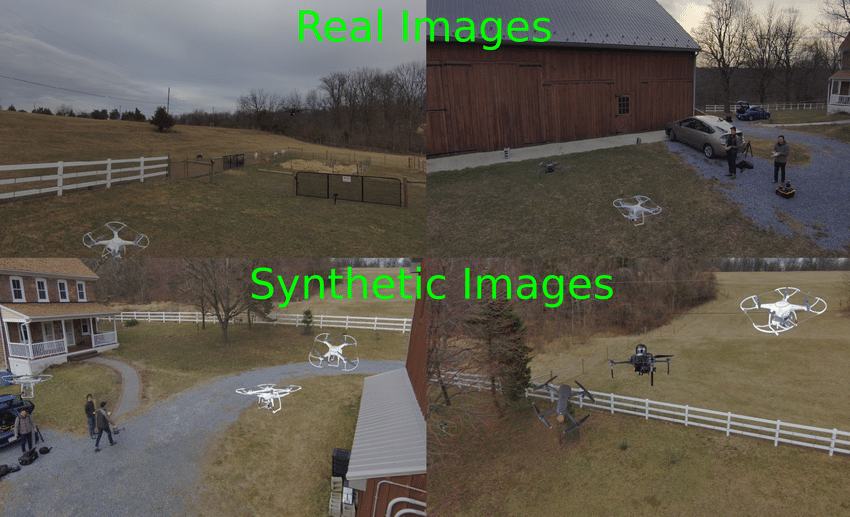
With the advancement of AI, images are nowadays classified based on the how they were created or originated.

**Natural Images**

* Captured by Camera: Natural images are captured by digital cameras, which convert light from a scene into pixel values.
* Real-World Data: Natural images are based on real-world data, including noise, artifacts, and imperfections due to the camera's sensor and processing.
* Metadata: Natural images often include metadata such as EXIF data, which contains information about the camera settings, date, and time of capture.

**Synthetic Images**

* Generated by Computer: Synthetic images are generated using computer algorithms, which simulate the laws of optics and physics to create realistic images.
* Pre-annotated: Synthetic images are often pre-annotated with labels, making them useful for training machine learning models.
* Controlled Variability: Synthetic images can be designed to have controlled variability, such as different lighting conditions, backgrounds, and scenarios.



Key Differences

* Origin: Natural images are captured by cameras, while synthetic images are generated by computers.
* Properties: Natural images have unique properties such as noise and artifacts, while synthetic images are designed to be realistic and consistent.
* Metadata: Natural images often include metadata, while synthetic images may not include metadata or may have controlled metadata.

Applications

* Computer Vision: Both natural and synthetic images are used in computer vision tasks such as object detection, segmentation, and recognition.
* Machine Learning: Synthetic images are used in machine learning tasks such as image classification, object detection, and image generation.
* Digital Forensics: Synthetic images are used in digital forensics to analyze and verify the authenticity of digital evidence.

Challenges

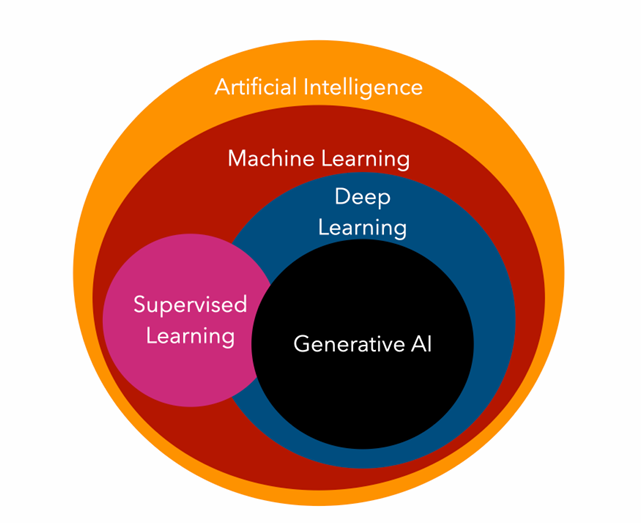
* Authenticity: Ensuring the authenticity and trustworthiness of synthetic images is a significant challenge.
* Detection: Developing methods to reliably detect and differentiate between real and synthetic images is an active area of research.

In summary, natural images are captured by cameras and reflect real-world scenes, while synthetic images are generated by computers and designed to be realistic. Synthetic images are useful for training machine learning models and can be used to create controlled variability, but ensuring their authenticity and detecting them reliably are significant challenges.

**Generative Artificial Intelligence**

Luckily, with the creation of Synthetic images, there also came a way to identify those images and how are they are trained to be like the original images.   
Generative Artificial Intelligence (Gen AI) is a type of artificial intelligence technology that can produce various types of content, including text, imagery, audio and synthetic data.

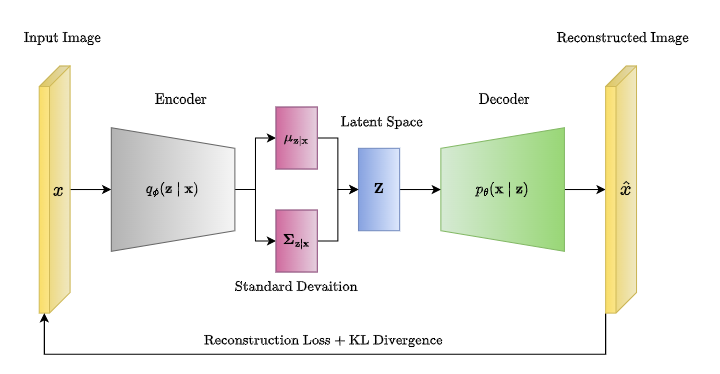
The recent buzz around generative AI has been driven by the simplicity of new user interfaces for creating high-quality text, graphics and videos in a matter of seconds. Generative AI starts with a prompt that could be in the form of a text, an image, a video, a design, musical notes, or any input that the AI system can process. Various AI algorithms then return new content in response to the prompt. Content can include essays, solutions to problems, or realistic fakes created from pictures or audio of a person.



**Generative AI models**

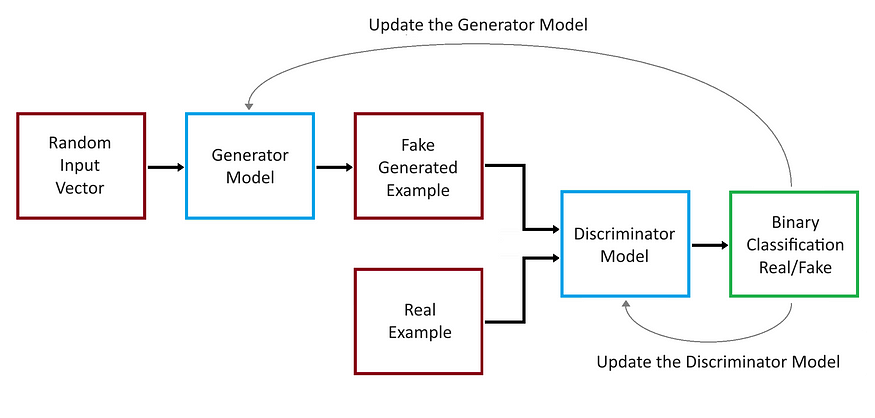
Generative AI models combine various AI algorithms to represent and process content. For example, to generate text, various natural language processing techniques transform raw characters (e.g., letters, punctuation and words) into sentences, parts of speech, entities and actions, which are represented as vectors using multiple encoding techniques. Similarly, images are transformed into various visual elements, also expressed as vectors. One caution is that these techniques can also encode the biases, racism, deception and puffery contained in the training data.

Once developers settle on a way to represent the world, they apply a particular neural network to generate new content in response to a query or prompt. Techniques such as GANs and variational autoencoders (VAEs) -- neural networks with a decoder and encoder -- are suitable for generating realistic human faces, synthetic data for AI training or even facsimiles of humans.



A Generative Adversarial Network (GAN) consists of two neural networks, namely the Generator and the Discriminator, which are trained simultaneously through adversarial training.

* Generator: This network takes random noise as input and produces data (like images). Its goal is to generate data that’s as close as possible to real data.
* Discriminator: This network takes real data and the data generated by the Generator as input and attempts to distinguish between the two. It outputs the probability that the given data is real.



Recent progress in transformers such as Google's Bidirectional Encoder Representations from Transformers (BERT), OpenAI's GPT and Google AlphaFold have also resulted in neural networks that can not only encode language, images and proteins but also generate new content.

**An Introduction to Monochromatic Cameras for Night Vision**

Monochromatic cameras for night vision are specialized cameras that use monochromatic sensors to capture images in low-light conditions. These cameras are designed to enhance visibility in dark environments by amplifying the available light, allowing for better image quality and increased visibility. Here are some key points about monochromatic cameras for night vision:

* Monochromatic Cameras for Night Vision
* Sensor Technology: Monochromatic cameras for night vision use monochromatic sensors, which capture images in shades of gray rather than color.
* Amplification: These cameras amplify the available light to enhance visibility in dark environments.
* Image Quality: Monochromatic cameras for night vision provide better image quality compared to color cameras in low-light conditions.
* Applications: These cameras are used in various applications such as surveillance, security, and military operations.

Many assume that color imaging is more advanced than black-and-white, or monochrome, imaging. But that’s not always true when it comes to machine vision.

Monochrome cameras combined with high-quality optical filters:

* Can offer higher contrast and better resolution
* Provide better signal-to-noise ratio
* Are sensitive to near-ultraviolet, visible and near-infrared spectrums
* Provide flawless results

These features give monochrome cameras a significant advantage when it comes to optical character recognition and verification, barcode reading, scratch or crack detection, wavelength separation and more. Depending on your application, monochrome cameras can be three times more efficient than color cameras.

**Features of night surveillance cameras**

A monochromatic night vision camera should have several relevant characteristics:

* Powerful IR illumination, preferably adaptive;
* Ability to switch from color mode to black and white;
* High light sensitivity of the sensor;
* ICR filter;
* Weatherproof housing.
* Backlight. The first point is mandatory if you need quality night photography. You should pay attention to the backlight parameters declared by the manufacturer:
* Illumination angle;
* Range of action;
* Execution.

Advantages

* Enhanced Visibility: Monochromatic cameras for night vision enhance visibility in dark environments by amplifying the available light.
* Better Image Quality: These cameras provide better image quality compared to color cameras in low-light conditions.
* Increased Sensitivity: Monochromatic cameras are more sensitive to light, allowing for better image capture in low-light conditions.

Challenges

* Color Information: Monochromatic cameras for night vision do not capture color information, which can be a limitation in certain applications.
* Noise: These cameras can be prone to noise in low-light conditions, which can affect image quality.

Examples

* Infrared Cameras: Infrared cameras are a type of monochromatic camera that use infrared radiation to capture images in low-light conditions.
* Night Vision Goggles: Night vision goggles are a type of monochromatic camera that use image intensification technology to enhance visibility in dark environments.

Future Directions

* Advancements in Sensor Technology: Advances in sensor technology are expected to improve the performance of monochromatic cameras for night vision.
* Integration with Other Technologies: Integration with other technologies such as artificial intelligence and machine learning can enhance the capabilities of monochromatic cameras for night vision.

In summary, monochromatic cameras for night vision are specialized cameras that use monochromatic sensors to capture images in low-light conditions. These cameras enhance visibility in dark environments by amplifying the available light and provide better image quality compared to color cameras in low-light conditions. However, they do not capture color information and can be prone to noise in low-light conditions.

**3 major advantages of monochrome cameras over color cameras**

In comparison with color camera, a monochrome camera has the following advantages:

* Monochrome cameras perform better in low lighting conditions
* Monochrome sensors have intrinsically higher frame rates
* Monochrome algorithms are well-tuned

**Conclusion**

* We have explored the fundamental differences between natural images and digital images, as well as the unique capabilities and applications of monochromatic cameras for night vision.
* Natural images are photographs captured by digital cameras, reflecting real-world scenes and objects. These images capture the inherent complexity and variability of the physical world, making them invaluable for a wide range of applications, from computer vision to digital forensics. Natural images are characterized by their unique properties, such as noise, artifacts, and imperfections, which are a result of the camera's sensor and processing.
* In contrast, digital images are computer-generated or digitally manipulated pictures, created through mathematical modeling and computational processes. Synthetic digital images offer several advantages, including the ability to control variability, pre-annotate data, and preserve privacy. However, ensuring the authenticity and trustworthiness of synthetic images remains a significant challenge, as does the reliable detection of synthetic images.
* Monochromatic cameras for night vision are a specialized type of digital imaging technology that leverages monochromatic sensors to enhance visibility in low-light conditions. These cameras amplify the available light, providing better image quality compared to color cameras in dark environments. While they offer enhanced visibility and increased sensitivity, monochromatic cameras for night vision do not capture color information, which can be a limitation in certain applications.
* As technology continues to evolve, we can expect to see further advancements in both natural and digital image capture, as well as the integration of monochromatic cameras with other emerging technologies, such as artificial intelligence and machine learning. These developments will undoubtedly lead to more robust, versatile, and reliable imaging solutions across a wide range of industries and applications.
* In conclusion, the distinctions between natural and digital images, as well as the unique capabilities of monochromatic cameras for night vision, highlight the ongoing evolution and diverse applications of imaging technology in our modern world.

**Bibliography**

* <https://www.analyticsvidhya.com/blog/2021/07/step-by-step-guide-for-image-classification-on-custom-datasets/>
* <https://cloud.google.com/vertex-ai/docs/image-data/classification/create-dataset>
* <https://developers.google.com/machine-learning/data-prep/construct/construct-intro>
* <https://www.linkedin.com/pulse/what-good-quality-training-dataset-machine-learning-tagx/>
* <https://selectstar-ai.medium.com/creating-the-best-quality-image-dataset-720f612944ed>
* <https://www.sciencedirect.com/science/article/pii/S0950584923001222>
* <https://www.kaggle.com/code/mekhdigakhramanian/creating-your-own-dataset-from-google-images>
* <https://chisw.com/blog/image-recognition-challenges/#:~:text=In%20image%20processing%2C%20key%20challenges,level%20accuracy%20across%20complex%20conditions>.
* <https://24x7offshoring.com/image-dataset-in-machine-learning/#:~:text=It%20is%20commonly%20used%20for,detection%2C%20segmentation%2C%20and%20captioning>.
* <https://unthinking.photography/articles/an-introduction-to-image-datasets>
* <https://medium.com/@FxisAi/what-are-the-key-challenges-in-developing-accurate-image-recognition-and-computer-vision-systems-f91d3a970c6f>
* <https://bayshoreintel.com/5-common-problems-with-working-on-real-world-image-data>
* <https://www.techtarget.com/searchenterpriseai/definition/generative-AI>
* <https://www.cmu.edu/intelligentbusiness/expertise/genai-principles.pdf>
* <https://www.eweek.com/artificial-intelligence/what-is-generative-ai/>
* <https://www.geeksforgeeks.org/hidden-markov-model-in-machine-learning/>
* <https://thelifesciencesmagazine.com/digital-image-analysis/>
* <https://www.geeksforgeeks.org/digital-image-processing-basics/>
* <https://www.geeksforgeeks.org/digital-image-processing-basics/>
* <https://medium.com/@globosetechnologysolution/building-and-using-image-datasets-for-machine-learning-f5eb6d5f9e2a>
* <https://unthinking.photography/articles/an-introduction-to-image-datasets#:~:text=A%20dataset%20in%20computer%20vision,examples%20contained%20in%20the%20dataset>.
* <https://www.linkedin.com/advice/0/what-some-common-challenges-pitfalls-image-classification>
* <https://www.researchgate.net/publication/325000268_Distinguishing_between_Natural_and_Computer-Generated_Images_Using_Convolutional_Neural_Networks>
* <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5948567/>
* <https://medium.com/@farihanur1438/image-classification-using-traditional-machine-learning-algorithms-332c14bb61b4>
* <https://kili-technology.com/data-labeling/computer-vision/image-annotation/image-recognition-with-machine-learning-how-and-why>
* <https://midopt.com/solutions/monochrome-imaging/>