FMDB *Transactions on Sustainable Computer Letters*

Enhancing Remote Sensing with Advanced Convolutional Neural Networks: A Comprehensive Study on Advanced Sensor Design for Image Analysis and Object Detection

R. P. Pranav[1,*], R. P. Prawin[2], R. Subhashni[3], Sonjoy Ranjon Das[4]

[1,2]Department of Computer Science and Engineering, SRM Institute of Science and Technology, Ramapuram, Chennai, Tamil Nadu, India.
[3]Department of Computer Science and Applications, St. Peter's Institute of Higher Education and Research, Chennai, Tamil Nadu, India.
[4]Department of Computer Science, Shipley College, Shipley, England, United Kingdom.
pranavramesh2004@gmail.com[1], prawinrp2002@gmail.com[2], subhashniraj2018@gmail.com[3], sanjoy.das@shipley.ac.uk[4]

Abstract: Advanced-Convolutional Neural Networks (A-CNNs) and modern computer vision technology are revolutionising remote sensing. This research revolutionises remote sensing visual identification by seamlessly merging cutting-edge sensor design and CNNs. We want to redefine remote sensing data use away from human-centric methods. Our substantial research into upgraded CNNs has given us unsurpassed image recognition accuracy and efficiency. Cutting-edge sensors that record stunning faraway images and provide contextual data for analysis have enhanced this achievement. Our extensive analysis includes sensor technology advances and the incorporation of CNN into remote sensing workflows. Thus, a comprehensive remote image analysis system for environmental monitoring, disaster management, and security enhancement has been created. The study showed great advances in distant picture processing and item detection. Our adaptable technique can recognise objects of interest, classify land cover, and track temporal change. Using sensor data efficiently with CNNs allows real-time, data-driven decision-making in many fields. In summary, our study ushers in a new era in distant sensing visual identification, entering undiscovered territory. With powerful CNNs and cutting-edge sensor design, remote image analysis and object detection can be improved across industries, ushering in a new era of innovation.

Keywords: Visual Recognition; Convolutional Neural Networks (CNNs); Sensor Design; Image Analysis; Satellite Imagery; Object Detection; Computer Vision; Environmental Monitoring; Machine Interpretability; Innovative Sensor Systems.

Received on: 16/05/2023, **Revised on:** 27/08/2023, **Accepted on:** 13/11/2023, **Published on:** 22/12/2023

Cite as: R. P. Pranav, R. P. Prawin, R. Subhashni, and S. Ranjon Das, "Enhancing Remote Sensing with Advanced Convolutional Neural Networks: A Comprehensive Study on Advanced Sensor Design for Image Analysis and Object Detection," FMDB Transactions on Sustainable Computer Letters., vol. 1, no. 4, pp. 255–266, 2023.

1. Introduction

Over the past decade, there has been an incredible increase in the gathering of visual data from above. These large amounts of data, known for being very large and having very specific details, have been used in many different fields, such as helping with disasters, watching the environment, and doing Intelligence, Surveillance, and Reconnaissance (ISR) [13]. However, managing and understanding these huge and complicated amounts of data has been difficult using only human analysis [14]. As a result of this difficulty, the field of picture data processing has undergone quick and significant change. Adopting advanced technologies such as Convolutional Neural Networks (CNNs) has received special attention [15]. These neural networks have

[*]Corresponding author.

shown to be game changers in image identification, allowing us to tackle previously thought intractable issues. The evolution of CNNs in image recognition is remarkable [16]. Their beginnings can be traced back to the late 1990s when Yann LeCun introduced the notion. However, it wasn't until Alex Krizhevsky's revolutionary work in 2012 that CNNs finally came to the fore [17]. Krizhevsky's groundbreaking deep architecture, which included several layers, was a watershed moment in computer vision. It proved deep learning's great potential, particularly when applied to challenging picture identification issues. We are solving an issue involving image recognition [18]. Imagine you have a collection of images (dataset D) taken with a sensor with specific settings (parameters P). Our goal is to identify the best settings (p) to optimize the performance of an image recognition model (m) on dataset D. We measure performance using criteria such as accuracy [19]. However, we must also consider the sensor's cost in a practical setting. Ultimately, we seek to balance performance and cost-effectiveness [20].

This research demonstrates using the latest image data and advanced visual recognition techniques to formulate an optimized solution to this problem [21]. We showcase findings for various problem settings. Our selected models $m \in M$ consist of CNN (Convolutional Neural Network) variants and image quality algorithms. However, it is crucial to emphasize that the approach presented here is versatile and can be applied in numerous scenarios [22]. Developing strong neural network software frameworks such as TensorFlow, Caffe, and PyTorch facilitates this advancement [23]. These frameworks have enabled a vibrant and dynamic community of researchers, engineers, and developers to work together on creating and implementing neural networks specifically designed for visual identification tasks [24]. They have enabled quick experimentation and innovation, allowing the industry to evolve astonishingly [25].

The importance of employing CNNs for visual recognition in the context of overhead imagery cannot be emphasized. Their extraordinary capabilities and the ever-increasing volume of image data highlight the importance of optimizing sensing systems to gather images that smoothly correlate with CNN-driven recognition [26]. These sensing devices must be built to provide copious data suitable for CNNs' sophisticated capabilities [27]. This optimization approach is not confined to a single image recognition model but may be applied to various models, including CNN-based models and picture quality algorithms. It is a flexible approach that may be used with various image recognition models, including the Human Visual System (HVS) [28]. One of the key aspects of this research is the comparative analysis of metrics. It delves into comparing metrics associated with human recognition and metrics associated with machine recognition [29]. This comparative approach offers valuable insights into sensor parameters' effectiveness in optimizing image recognition models' performance. The vital elements and configurations to solve the optimization problem [30]. These elements match the variables and settings in the optimization equation. We will explore the background in the following sections, covering image usefulness and sensor modelling [31]. We will also discuss the sensor simulator and our experimental approach [32].

Most importantly, our work incorporates the findings of multiple experiments [33]. These trials demonstrate the practicality and efficacy of the proposed optimization approach. Through these experiments, we provide insight into how sensor characteristics affect picture recognition performance, whether through classification, detection, retrieval, or any other recognition task. In conclusion, our research marks an important step toward optimizing sensor systems for image recognition [34]. It recognizes the growing importance of machine-based recognition in a world flooded with massive image databases. Incorporating CNNs and image quality algorithms highlights the adaptability of our method [35]. We discover significant patterns that link observable sensor parameters with recognition outcomes by comparing human and machine recognition measures [36]. Finally, our study has far-reaching consequences and provides novel advances in picture identification in the context of overhead images [37].

2. Literature Review

Sumbul et al. [1] address fine-grained object recognition and zero-shot learning in remote sensing imagery. The authors explore advanced techniques in image analysis for improved classification in geospatial data, contributing to the field's understanding of recognizing objects with high granularity and achieving zero-shot learning capabilities in remote sensing applications.

Song et al. [2] focus on intelligent object recognition of urban water bodies using deep learning on multi-source and multi-temporal high spatial resolution remote sensing imagery. The research aims to enhance the accuracy and efficiency of urban water body detection, contributing to advancements in intelligent image analysis for water resource management in urban areas.

Nica et al. [3] investigate the convergence of geospatial big data management, computer vision algorithms, remote sensing, image recognition, and event modelling in the metaverse. Their comprehensive approach sheds light on integrating advanced technologies and their implications for the virtual economy.

Cui et al. [4] contribute to remote sensing image analysis focusing on multi-scale semantic segmentation and spatial relationship recognition. Their work employs an attention model, enhancing the precision of identifying and understanding spatial features in remote sensing imagery, offering advancements in image processing for geospatial applications.

Lin et al. [5] introduce a novel secure satellite remote sensing approach by proposing a few-shot RF fingerprinting recognition system. The study addresses the need for enhanced security in satellite image processing, contributing to advancements in securing remote sensing data through innovative techniques in radio frequency fingerprinting.

Zhang et al. [6] comprehensively review deep learning applications in remote sensing image understanding. The study highlights the efficacy of deep learning methods for classification and feature extraction tasks, showcasing their potential to enhance the accuracy and efficiency of remote sensing data interpretation.

Alhilal et al. [7] focus on image-based object identification to enhance event-driven sensing in wireless multimedia sensor networks. Their work emphasizes efficient methods for identifying objects through images, contributing to advancements in sensor network capabilities for real-time event detection and monitoring.

Redondi et al. [8] present a testbed realization of a visual sensor network for object recognition. Their study focuses on the practical implementation of a network designed for efficient object recognition, contributing to developing robust and scalable systems in digital signal processing and visual sensor networks.

Mirdanies et al. [9] propose an object recognition system for remote-controlled weapon stations, employing the SIFT and SURF methods. Their study addresses the critical need for accurate and efficient object recognition in military applications, contributing to advancements in technology for enhancing the capabilities of remote-controlled weapon systems.

Che et al. [10] present a comprehensive review of the state of the art in object recognition, segmentation, and classification of mobile laser scanning point clouds. The study surveys current methodologies, highlighting advancements in processing techniques for extracting meaningful information from mobile laser scanning data, contributing to improved applications in various domains.

Martins et al. [11] introduce a Multi-scale Object-Based Convolutional Neural Network (multi-OCNN) for high-resolution remote sensing image classification. The study explores the effectiveness of this novel approach, highlighting its potential to enhance classification accuracy and efficiency in handling high spatial resolution imagery, contributing to advancements in remote sensing applications.

Mohan et al. [12] review remote sensing methods for landslide detection, focusing on machine and deep learning techniques. The study assesses the efficacy of these methods, offering insights into advancements in landslide detection technologies. Their work contributes to the evolving field of remote sensing applications for natural disaster monitoring.

3. Contextual Framework

When evaluating the quality of images in general, we can group these into two kinds of image quality assessment (IQA): full-reference IQA and no-reference IQA. With full-reference IQA, we can compare a target image with the original (undistorted) one. On the other hand, no-reference IQA doesn't have this luxury, so it has to evaluate the quality of an image without it being given a reference point. Numerous algorithms have been proposed for each category, all of which work toward taking an image and producing a single score representing its quality [38]. Full-reference IQA in evaluating compression effects like those created by JPEG typically use the Mean-Squared Error (MSE) and Peak Signal-to-Noise Ratio (PSNR) as their main metrics. For no-reference IQA, we don't have such a large amount of functions [39].

There are algorithms like BRISQUE and NIQE, for example [40]. The former is considered opinion-aware since it's developed using human evaluations, measuring the quality based on linear combinations of locally normalized luminance coefficients [41]. Its creator says it can capture image distortion and naturalness by doing so. On the other hand, the latter calculates its score by comparing features of the image to a multivariate Gaussian distribution model [42]. Comparing these IQA metrics with how humans perceive image quality and how CNNs recognize images, the team could determine their effectiveness and whether they could predict CNN performance across different sensor systems [43].

Specifically, we differentiate between the concepts of picture quality and utility, both of which are concerned with judging the worth of an image for particular recognition tasks. The National Imagery Interpretability Rating Scale (NIIRS) was developed by the IRARS group in 1974 with the purpose of quantifying the interpretability of imagery that is normally obtained from overhead platforms [44]. It employs a scale that ranges from 0 to 9, where 0 indicates that it is impossible for anyone to read it due to artefacts or blur, and nine indicates that you are able to differentiate between items at the centimetre scale. Despite the fact that there is no clear definition for "image quality" in the same way that there is for "image utility" in the context of NIIRS, image utility, as defined by NIIRS, is something that we are able to quantify, measure, and eventually improve [45]. We have

developed equations in order to establish a correlation between human-annotated NIIRS data and other quality characteristics of the most popular forms of imagery and the collecting techniques that are linked with them [46].

The GIQE5 optical analysis runs on the belief that an analyst will manipulate the image during the analysis progress, known as the softcopy scenario [47]. By merely using image-editing tools, the artist does not subscribe to postprocessing generic elements [48]-[50]. The GIQE5 involves three key variables The GIQE5 involves three key variables:

$$\mathrm{FN} = \frac{f}{D} \text{ and } \mathrm{Q} = \frac{\lambda \mathrm{FN}}{p} \tag{1}$$

In order to carry out the method, we have a plan in place. To begin, we make use of public photos, and then we proceed to incorporate any items that are already present in the image system [51]. After that, we remade these photographs so that they appeared to be real as if they had been obtained by a piece of different imaging equipment. However, we were the ones responsible for designing all of the parameters [52]. With regard to these sets, the process of image production is broken down into two groups, namely training standards and testing standards [53]. Following that, a CNN and a classifier model are trained with the help of these SVM simulations. Following the loading of the results of the tests, we analysed the images, and then we returned to the process of making the necessary adjustments to the various parameters [54]. Leakage and yield are two system metrics that are extremely important to consider when evaluating the performance indicators of a technology.

Let's consider a situation in designing an imaging system: for example, we can imagine a NASA-bound remote sensing system engineer whose decision leads to the diameter of the lens aperture selection being his concern [55]. This is achieved by studying GSD, SNR, output NIIRS value, and their contribution to the final price to find the fine line between the price and their true value [56]. An increase in aperture may cause a lower GD, a higher SNR, and finally, the highest NIIRS among all objects. On the other hand, there is a dimension after which the diameter of the aperture could rise, which is caused by the cost of building and launching the system [57]. If we connect this example to Equation 1, we can express it as:

$$\arg\min_{D} L(D; P, \mathcal{D}) \tag{2}$$

$$L = \alpha \cdot \mathrm{NIIRS}(D; P, \mathcal{D}) + \beta \cdot \mathrm{cost}(D; P) \tag{3}$$

In our proposed system, the letter 'D' represents the diameter of the system's aperture. The amount of monetary investment 'cost' respectively should be taken into account, 'NIIRS' means the potential NIIRS rating for the system and stands for the dataset used for the NIIRS model (in this case, the GIQE). The objective function we need to optimize, called 'L', is a complex function which includes coefficients' α' and 'β' for NIIRS scores and costs. Only the space around the sensing device and direction 'P' is not part of the camera's field of view, whose parameters are 'D'. Using night vision technology, we aim to optimize imagery for the Human Visual System to reach the highest Night Image Intensifier Rating System.

Convolutional Neural Networks, also known as CNNs, are a class of significant algorithms that are fast gaining popularity for decision-making procedures based on optical pictures obtained from above. Generally speaking, they are attempting to find solutions to issues with categorization, retrieval, detection, and segmentation. When it comes to image classification, the amount of different ways that deep CNN techniques can be applied varies depending on the architecture, particularly when it comes to scenes that are shot from above. In addition, there is an increasing tendency toward viewing and making enormous databases of publicly available surface images available to the public. For instance, the SpaceNet dataset, which was initially presented in 2016, is intended for use in classification applications. It is comprised of hundreds of thousands of annotated polygons that detect buildings in DigitalGlobe satellite imagery. The Functional Map of the World (fMoW) dataset, which was published in 2016, is another illustration. It contains one million DigitalGlobes that are organised into 63 categories. This increase in publicly available data using rectangular bounding boxes for segmentation, innovations in GPU hardware, and advancements in CNN architecture have all contributed to a significant increase in the use of CNN for surface image analysis. Image descriptions also contributed to this increase.

4. Methodology

We use a simple method to predict algorithm performance as image system parameters change. Initially, we create a code replicating imagery gathered by several overhead image systems. This simulator currently uses high-quality DigitalGlobe imagery. Next, we use the simulator to reproduce the target dataset while changing one of the system's parameters, like the optic focal length. Ultimately, we utilize this dataset to train and evaluate an image recognition model and track the evolution of validation performance to modified system parameters. A software framework was developed to regulate data processing, machine learning, parameter management, and result logging. Python 3 was used to write this framework, while PyTorch was

used to train and assess CNNs. The framework can facilitate the learning and image simulation procedures with parameter change.

5. Dataset

According to what was stated earlier, the fMoW dataset contains more than a million photographs, each of which is connected to a target that is a member of one of 62 classes or an additional false detection class. The photos that make up the dataset were captured in over 200 different countries using one of the four possible DigitalGlobe sensor platforms: GeoEye1, QuickBird2, WorldView2, or WorldView 3. The distribution of the total number of photographs taken at each station according to the classes. This dataset contains bounding boxes of objects that range in size from approximately 10 pixels at the smallest to over 700k pixels at the largest. The sizes of these bounding boxes are quite variable. In this way, it is distinguishable. A unique challenge arises when one considers how to pre-process the data in order to make it acceptable to a fixed input model. Within the context of this experiment, we investigate two fundamental approaches to this pre-processing, each of which is treated as an independent variable.

After being converted to TOA (or at-aperture) radiance, fMoW data has the potential to emulate the process of picture gathering from a variety of remote sensing devices. In this section, we will discuss this particular approach. First and foremost, we would want to bring to your attention the fact that one is limited by the capabilities of the picture system that was utilised in order to collect the initial dataset. We are only able to produce pictures of a lower quality than most other photographers (lower resolution, more noise). To provide a brief overview, the optical Modular Transfer Function (MTF) is utilised to execute a blurring function on the digital images that were initially captured. The data is resampled using the Fourier transform, and aliasing artefacts are incorporated into the process. The generated Fourier image is put back into picture space once an inverse transformation has been performed. Following this, a scalar value is computed that converts the radiance photos taken at the aperture into electron units based on the detector measurements. Initially, we provide the following definition of the optical cut-off frequency for an incoherent imaging system with diffraction limitation:

$$\nu_{\text{optcut}} = \frac{1}{\lambda \text{FN}} = \frac{D}{\lambda f} \tag{4}$$

Where f stands for focal length, D for aperture diameter, λ for light wavelength, and FN, or "f-number," is equal to f/D. This is the highest-frequency sinusoid an optical system produces on the picture plane. It is recommended to directly compare this number with the detector's Nyquist sampling limit.

The present use of the optical ground sample $(\text{GSS}_{\text{optics}})$ size complies with the Finite definition of:

$$\text{GSS}_{\text{optics}} = \lambda \frac{H}{D} \tag{5}$$

Where H represents the remote sensing system's height, what is meant by the ground-based optical cut-off frequency is:

$$\nu_{\text{optcut\_gnd}} = \frac{1}{2 \cdot \text{GSS}} = \frac{D}{2 \cdot \lambda \cdot H} \tag{6}$$

We used SqueezeNet, VGG, ResNet, and DenseNet as the four CNN architectures in our research. These designs were chosen to demonstrate the most advanced CNNs for visual recognition while allowing for considerable design and parameter customization. The PyTorch torch-vision library8 is used to implement all of the models, and either cross-entropy loss or triplet margin loss is utilized for training.

It is widely recognised that the VGG model, which is one of the earliest effective CNN architectures, is renowned for its numerous parameters and uncomplicated construction. A number of recurrent convolutional layers with dimensions of 3x3 are utilised in the VGG approach. Additionally, 2x2 max-pooling layers are utilised in order to reduce the spatial dimension between blocks. In spite of the fact that ResNet, DenseNet, and other designs have outperformed it in terms of performance, it is still valuable as a reliable baseline. The popular 16-layer variant of this model, which is referred to as VGG16, is the one that we use. VGG has a significant challenge in the form of the vanishing gradient problem, which occurs when additional layers are added. The skip connection structure, which is a fundamental block with its input appended to its output, is a solution that was provided by the ResNet architecture to address this problem. Because of this, it is now feasible to carry out optimization with a far more extensive network. While conducting tests, we make use of the ResNet152 architecture, which is comprised of 152 layers. DenseNet concatenates layer inputs to layer outputs rather than adding them, expanding on ResNet's skip connection approach. Because it connects all network levels directly, this paradigm promotes feature reuse. We use a 161-layer DenseNet (DensetNet161) architecture, showing promising results on the fMoW dataset and ImageNet. The basic architecture used by

the fMoW developers is DenseNet161, and every solution placed first in the competition used a DenseNet variation in some form or another. Although the design requires a lot of memory, it trains in a few epochs.

Transfer Learning: When referring to training a model on a sizable standard picture dataset and refining it using the relevant image data, possibly from a different domain, transfer learning is typically used for CNN models intended for visual recognition. In actuality, difficulties involving the visual identification of natural images are always resolved by transfer learning. Specifically, the target model is frequently pre-trained using the ImageNet dataset. Unless otherwise noted, all CNN models used in our studies were pre-trained on ImageNet. The torch vision library allowed for the download of these pre-trained weights.

6. Result and Discussion

Two different image recognition tasks are taken into consideration in order to evaluate the interpretability of machine images as a function of altering sensor attributes. The categorization problem is the first of these jobs, and it requires a specific instance to be placed into the appropriate class from a predetermined list of classes. The retrieval issue, the second of these tasks, calls for choosing examples from a group of instances that make up a gallery that matches a probe class. There are three main reasons why we employ categorization and retrieval. Initially, they are frequently applied to resolve practical issues. Therefore, the outcomes of these tests may be applied directly to evaluate real-world recognition software using overhead images. Secondly, these challenges are very easy compared to the detection and segmentation issues, which have more moving components. We aim to find a basic relationship to serve as a foundation for understanding more intricate recognition issues. Lastly, the retrieval problem, in particular, enables us to create a universal technique that may be used for classes that have not yet been discovered. To put it briefly, a salient feature extractor may be evaluated on fresh data with unknown classes after being trained on a known dataset of pictures, all without the need for retraining.

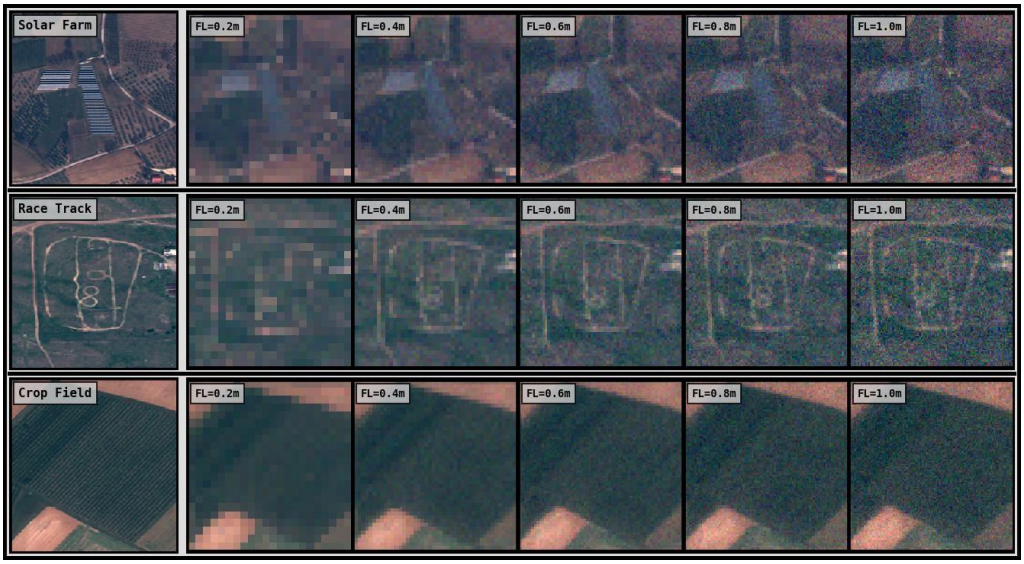

Figure 1: Sample image of three fMoW classes. [16]

Figure 1 shows how three sample picture classes from the fMoW subset are changed while maintaining constant values for the other parameters. The original photos are on the left, with their class designations in the top left corner, divided by a column.

In order to evaluate the classification task, we take into consideration the following four metrics: average precision (AP), top-1 accuracy, top-3 accuracy, and area under the ROC curve (AUC). Top-1 and top-k accuracy are the most fundamental measures available for evaluating categorization, and there is a good deal of comprehension surrounding these measurements. The process of designing sensors might, therefore, benefit from their utilisation. We came to the conclusion that top-3 accuracy was the most appropriate method after analysing the dataset and observing that certain groups of up to three objects were easily mislabeled or confused. It follows that a model that is capable of identifying three candidate classes, of which one will be accurate for the input image, is still something that we would be interested in. When it comes to machine learning applications, the Average Precision (AP) metric is widely utilised. This is due to the fact that it offers a consistent summary assessment of precision across all recall levels:

$$\mathrm{AP} = \int_0^1 \quad p(r)dr \qquad (7)$$

Precision is denoted by the symbol p(r) at a specific recall value, and recall is denoted by the symbol by itself. It is possible that in practice, precision and recall may be allocated different weights depending on the application; however, AP gives a helpful summary even in the absence of any prior information. Alternately, the Area Under the Receiver Operator Characteristic Curve (AUC) can be useful in situations when the application would benefit from a summary of the True Positive Rate (TPR), which is also referred to as recall versus False Positive Rate (FPR).

As the AP metric may be used to assess retrieval performance using any feature space as long as a distance metric can be determined between space components, we also use it to test retrieval performance. We examine an alternative interpretation of AP [55] for the retrieval issue, apart from the one presented in Equation 9. The distances between each gallery picture and the probing image are first measured to calculate the retrieval query's AP. The gallery items are then arranged to decrease or increase the distance from the probe. If the object category or class of a gallery image and a probing image are the same, they match. For the top-ranked m pictures, we quantify the accuracy as follows:

$$p_m = \frac{1}{m}\sum_{k=1}^{m} \quad x_k \tag{8}$$

If R represents the entire number of gallery photographs, then:

$$\text{AP} = \frac{1}{R}\sum_{i=1}^{n} \quad x_i p_i = \frac{1}{R}\sum_{i=1}^{n} \quad \frac{x_i}{i}\sum_{k=1}^{i} \quad x_k \tag{9}$$

Note that a CNN feature space is used to measure the distance between photos and that this distance is measured using a 2-norm distance. Through the use of AP, we are able to measure the usefulness of model features that have been fine-tuned and pre-trained. Using the retrieval problem in particular, we are able to evaluate the zero-shot case in situations when there have been no representations of the target classes seen during the training process. In spite of the fact that their labels do not correspond, it is still possible to observe that the probe and the samples from the highest-ranked gallery share similar qualities (Figure 2).

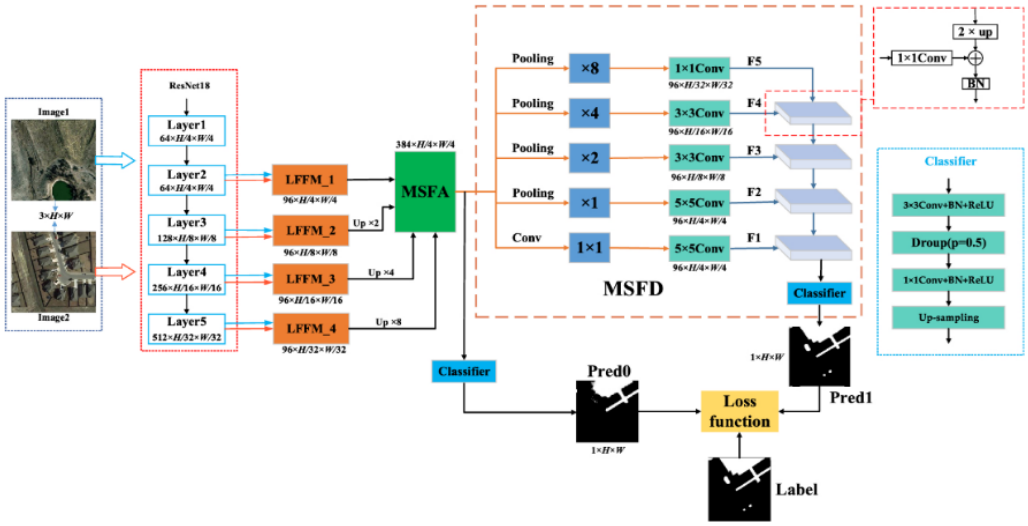

Figure 2: An architectural representation of A-CNN models with its layers [58]

Tracking urban expansion, deforestation, and other applications is made possible using pixel-wise change detection in remote sensing using MFPF-Net, which combines multi-scale features with spatial refinement for improved accuracy. The picture data underwent a substantial set of operations before and during the application of the simulation code. For the resizing procedure, an extra pre-processing step involved padding the edges with "reflected" values was needed. By doing this, boundary artefacts during the DFT conversions to and from the frequency domain were avoided. After cropping or scaling, the photos were converted into at-aperture radiance units and added to the sensor simulation algorithm. The generated images were expressed in TOA reflectance units.

After Simulation: Bilinear interpolation was again utilized to resample the images to 224 × 224 pixels because the imagery in TOA reflectance units can have different dimensions than the input pixels. The extra padding that was previously noted was removed in the case of scaled photos following this interpolation. Further domain normalization is not required because the TOA reflectance domain is near [0, 1]. Standard scores, which are calculated by subtracting the mean from the standard deviation and dividing it by the standard deviation, were utilised in the initial studies; however, this was only found to be

beneficial when a pre-trained model was not adjusted. The fine-tuning trials that were carried out on the fMoW subset did not contain any further data standardisation whatsoever.

Hardware and OS: Every experiment was run on a computer equipped with a Quadro M6000, Pascal Titan X, or Tesla K80 NVIDIA GPU. CUDA 10.0 and Ubuntu 18.04-based Docker containers were used for the studies. With the usage of two GPUs and an 88-person mini-batch size, the majority of tests took about 48 hours to finish.

CNN Experiments: The approaches outlined were used for three primary kinds of CNN trials. Ten-fold cross-validation with 35 classes was used for all trials; 900 photos from each class were used for training, and 100 images from each class were used for validation. The photos were arbitrarily given to each fold within each class.

Fundamental Training and Evaluation: Using the original dataset, the three CNN models were trained and verified to provide an upper bound on performance.

Degraded Training and Evaluation: The CNN models were trained and verified using the modified dataset, which was converted for every sensor parameter value.

IQA measures, including full-reference and no-reference versions, were evaluated and compared to the performance of the GIQE5 NIIRS and CNN evaluations. It was necessary to use all one thousand modified photographs from each of the thirty-five courses in order to calculate each metric for each and every simulator parameter setup. In order to pre-process the photos and normalise them to the range [0, 1], the "cropping" method that was stated earlier was utilised.

Full-Reference: According to what was said in the section under "Pre-Processing," a particular picture was utilised in order to assess the PSNR and SSIM metrics in comparison to one another before and after the simulation. The final result for SSIM was obtained by individually calculating the scores for each channel and then averaging the resulting scores. In order to carry out the testing, the sci-kit-image11 implementation of these metrics was utilised.

No-Reference: Only the photographs taken after the simulation were utilised in the calculation of the BRISQUE and NIQE measurements. In the same manner as SSIM, the scores for both measures were computed individually for each channel, and then the average was used to determine the final result. The BRISQUE implementation of NIQE was developed in 2013, while the sci-kit-video implementation was used in this study. Additional studies are necessary to establish whether or not retraining each model with data from the domain could potentially improve the performance of these algorithms.

Calibration: To define an upper bound on performance, we first train the four target CNN architectures on the original fMoW subset. This will represent an upper bound because the data modifications involve a rigorous degradation comprising down sampling and noise addition. Five measures have their highest values mentioned together with the training epoch during which they were attained. The four models underwent five epochs of fine-tuning after being pre-trained on ImageNet. It's interesting to see that, on the original ImageNet dataset, ResNet152 does somewhat better than the DenseNet161 design. This demonstrates once more how various topologies may function better on certain datasets. Understanding the performance connection between the many state-of-the-art designs presented is crucial since any of them may be the best for a particular dataset.

Baseline: We consider the connection between DenseNet161's rAP and optic focal length for our baseline experiment. The crop approach was used to pre-process the images, and five training epochs were used to train the model. This outcome is contrasted with NIIRS, which was calculated using the GIQE5:

$$\arg\max_{f} \mathrm{rAP}(f; P, m, \mathcal{D}) \qquad (10)$$

It would be simpler to recognise the gridded pattern of the solar farm if the focal length values were increased. This is due to the fact that spatial features carry greater weight than spectral features in situations where there is less light entering the system and more noise being produced. As an additional point of interest, certain classes have a greater range of rAP to focal length; for instance, the ratio of 0.4 for road bridges to 0.2 for agricultural areas is a greater example.

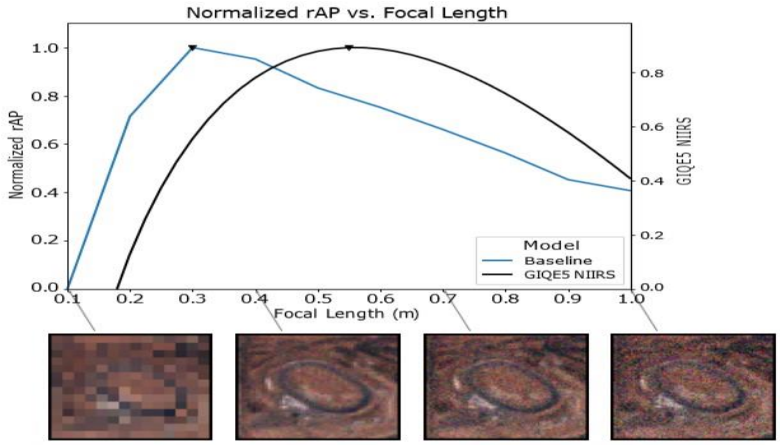

Figure 3: Normalized rAP versus Focal Length Graphic [16]

Figure 3 illustrates the baseline experiment for this study, where rAP is averaged across a variety of sensor system focal length values (D=0.05m) across validation partitions of the fMoW subset. To illustrate how the transformation affects the image's appearance, a racetrack image is displayed for various focal length values. For rAP, a distinct peak indicates the CNN's ideal trade-off between noise and resolution. The GIQE5 NIIRS curve has a distinct form and peaks at a higher focal length value for these focal length values. This indicates that the link between sensor system characteristics and CNN performance may differ for human interpretability. To properly compare the peaks between the two curves, note that GIQE5 NIIRS has not been adjusted, but rAP has been standardized to [0, 1]. Several metrics for the baseline model's validation set performance are displayed as a function of changing focal length. Although the normalized curves peak at various focal length values, you will see that they are almost identical. This variation's significance will vary depending on a particular application's cost and precision requirements. Plots of rAP were given for 10 different classes in the baseline experiment, which is in contrast to the averages that had been utilised in the trials that had been conducted before it. A lower bound for Rap is represented by the black dotted line in each panel. This lower bound is based on the assumption that the image embeddings are distributed randomly. Within the context of the baseline experiment, SqueezeNet1.1, VGG16, ResNet152, and DenseNet161 are evaluated and contrasted. You can see that the normalised curves are pretty similar to one another despite the fact that the recommended focal length for SqueezeNet1.1 is 0.6 metres, which is significantly longer than the other designs' 0.5 metres (Figure 4).

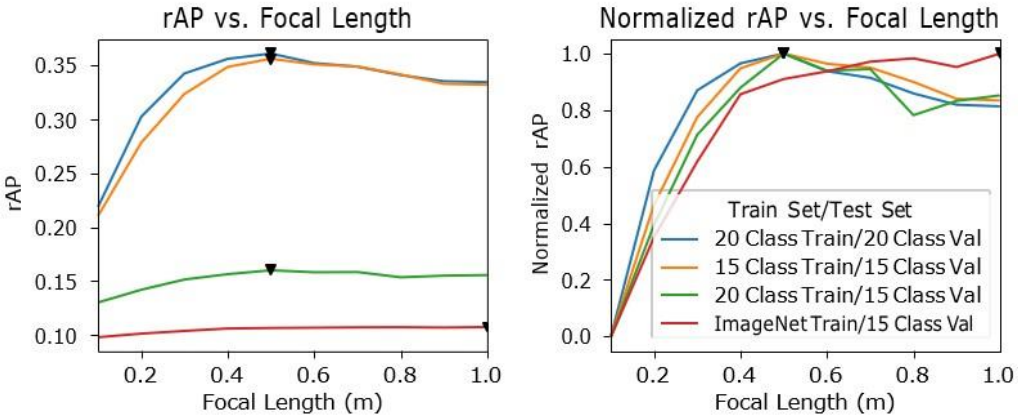

Figure 4: Normalised rAP versus Focal Length for Train/Test Set Models [16]

This chart displays the results of the zero-shot experiment that was conducted for the baseline circumstance. A total of four unique associations are displayed by these graphics: In the beginning, the partitions that are composed of 20 and 15 classes are investigated separately. Every single one of them was utilised for training and validation without being seen by the other. Following that, in the zero-shot scenario, the trained models from the 20-class partition were utilised in order to evaluate the 15-class partition by applying them. Examine the proximity of these three connections to one another, as well as the fact that their normalised rAP (which is displayed on the right) reaches its highest point at the same location. In addition, the performance

of the pre-trained ImageNet features is demonstrated by utilising a group of fifteen different classes. The conclusion is that there is a relationship that is distinct from the others, which suggests that these traits might not be significant for this matter, given the outcome.

7. Conclusion

In this work, we offer a method that may be used to optimise the settings of imaging systems that are used for remote sensing. The method is based on how well deep learning models perform on actual overhead picture data. In addition to conducting trials using a variety of sensor and learning settings, we also provide a tool that allows this study to be successfully implemented. This research demonstrates that CNNs are capable of performing visual recognition in a self-consistent manner across a variety of settings. In addition, we demonstrate how the performance of humans and machines in the field of visual identification can be quite different. To be more specific, we demonstrate that the most recent iteration of the GIQE algorithm does not correlate with the performance of CNN inside the parameter space that has been studied. Despite the fact that the GIQE is unable to compute NIIRS precisely in this parameter space, it is possible that CNN performance is more accurate than what is represented in NIIRS. This idea has a wide range of potential applications, such as sensing for autonomous vehicles and machine vision in industrial settings. When it comes to the development of automated sensing systems, we believe that sensor optimization for dealing with visual recognition issues will be an essential factor.

Acknowledgement: I am deeply grateful to my co-authors for their expertise and dedication, which greatly enrich this work, and I give special thanks to my friends for their unwavering support and encouragement throughout the research process.

Data Availability Statement: The data, graphs, and code supporting this study's findings are available from the corresponding author upon reasonable request.

Funding Statement: No funding has been obtained to help prepare this manuscript and research work.

Conflicts of Interest Statement: No conflicts of interest have been declared by the author(s). Citations and references are mentioned in the information used.

Ethics and Consent Statement: The consent was obtained from the organization and individual participants during data collection, and ethical approval and participant consent were received.

References

1. G. Sumbul, R. G. Cinbis, and S. Aksoy, "Fine-grained object recognition and zero-shot learning in remote sensing imagery," IEEE Trans. Geosci. Remote Sens., vol. 56, no. 2, pp. 770–779, 2018.
2. S. Song et al., "Intelligent object recognition of urban water bodies based on deep learning for multi-source and multi-temporal high spatial resolution remote sensing imagery," Sensors (Basel), vol. 20, no. 2, p. 397, 2020.
3. E. Nica and J. Vahancik, "Geospatial big data management and computer vision algorithms, remote sensing and image recognition technologies, and event modeling and forecasting tools in the virtual economy of the metaverse," Linguistic and Philosophical Investigations, vol. 22, pp. 9–25, 2023.
4. W. Cui et al., "Multi-scale semantic segmentation and spatial relationship recognition of remote sensing images based on an attention model," Remote Sens. (Basel), vol. 11, no. 9, p. 1044, 2019.
5. D. Lin, S. Hu, W. Wu, and G. Wu, "Few-shot RF fingerprinting recognition for secure satellite remote sensing and image processing," Sci. China Inf. Sci., vol. 66, no. 8, pp. 1-12, 2023.
6. L. Zhang, G.-S. Xia, T. Wu, L. Lin, and X. C. Tai, "Deep learning for remote sensing image understanding," J. Sens., vol. 2016, pp. 1–2, 2016.
7. M. S. Alhilal, A. Soudani, and A. Al-Dhelaan, "Image-based object identification for efficient event-driven sensing in wireless multimedia sensor networks," Int. J. Distrib. Sens. Netw., vol. 11, no. 3, p. 850869, 2015.
8. A. Redondi, L. Baroffio, A. Canclini, M. Cesana, and M. Tagliasacchi, "A visual sensor network for object recognition: Testbed realization," in 2013 18th International Conference on Digital Signal Processing (DSP), Fira, Greece, 2013.
9. M. Mirdanies, A. S. Prihatmanto, and E. Rijanto, "Object recognition system in Remote Controlled Weapon Station using SIFT and SURF methods," J. Mechatron. Electr. Power Veh. Technol., vol. 4, no. 2, pp. 99–108, 2013.
10. E. Che, J. Jung, and M. J. Olsen, "Object recognition, segmentation, and classification of mobile Laser Scanning point clouds: A state of the art review," Sensors (Basel), vol. 19, no. 4, p. 810, 2019.

11. V. S. Martins, A. L. Kaleita, B. K. Gelder, H. L. F. da Silveira, and C. A. Abe, “Exploring multi-scale object-based convolutional neural network (multi-OCNN) for remote sensing image classification at high spatial resolution,” ISPRS J. Photogramm. Remote Sens., vol. 168, pp. 56–73, 2020.
12. A. Mohan, A. K. Singh, B. Kumar, and R. Dwivedi, “Review on remote sensing methods for landslide detection using machine and deep learning,” Trans. Emerg. Telecommun. Technol., vol. 32, no. 7, pp. 1-10, 2021.
13. L. Zhang, L. Zhang, and B. Du, “Deep learning for remote sensing data: A technical tutorial on the state of the art,” IEEE Geosci. Remote Sens. Mag., vol. 4, no. 2, pp. 22–40, 2016.
14. Y. Long, Y. Gong, Z. Xiao, and Q. Liu, “Accurate object localization in remote sensing images based on convolutional neural networks,” IEEE Trans. Geosci. Remote Sens., vol. 55, no. 5, pp. 2486–2498, 2017.
15. K. Nogueira, O. A. B. Penatti, and J. A. dos Santos, “Towards better exploiting convolutional neural networks for remote sensing scene classification,” Pattern Recognit., vol. 61, pp. 539–556, 2017.
16. L. Jaffe, M. Zelinski, and W. Sakla, “Remote sensor design for visual recognition with convolutional neural networks,” IEEE Trans. Geosci. Remote Sens., vol. 57, no. 11, pp. 9090–9108, 2019.
17. A. A. Alfaifi and S. G. Khan, “Utilizing data from Twitter to explore the UX of ‘Madrasati’ as a Saudi e-learning platform compelled by the pandemic,” Arab Gulf Journal of Scientific Research, vol.39, no.3, pp. 200–208, 2022.
18. A. Ahmed Chhipa et al., “Adaptive Neuro-fuzzy Inference System Based Maximum Power Tracking Controller for Variable Speed WECS”,” Energies, vol. 14, no.19, p.6275, 2021.
19. A. Jafar, O. A. Alzubi, G. Alzubi, and D. Suseendran, “+ A Novel Chaotic Map Encryption Methodology for Image Cryptography and Secret Communication with Steganography,” International Journal of Recent Technology and Engineering, vol. 8, no. IC2, pp. 1122-1128, 2019.
20. A. K. Sharma, A. Panwar, P. Chakrabarti, and S. Viswakarma, “Categorization of ICMR Using Feature Extraction Strategy and MIR with Ensemble Learning” Procedia Computer Science, vol. 57, pp. 686–694, 2015.
21. A. Kumar, S. Singh, K. Srivastava, A. Sharma, and D. K. Sharma, “Performance and stability enhancement of mixed dimensional bilayer inverted perovskite (BA2PbI4/MAPbI3) solar cell using drift-diffusion model,” Sustain. Chem. Pharm., vol. 29, no. 10, p. 100807, 2022.
22. A. Kumar, S. Singh, M. K. A. Mohammed, and D. K. Sharma, “Accelerated innovation in developing high-performance metal halide perovskite solar cell using machine learning,” Int. J. Mod. Phys. B, vol. 37, no. 07, pp.1-14, 2023.
23. A. L. Karn et al., “B-lstm-Nb based composite sequence Learning model for detecting fraudulent financial activities,” Malays. J. Comput. Sci., vol.1, no.1, pp. 30–40, 2022.
24. A. L. Karn et al., “Designing a Deep Learning-based financial decision support system for fintech to support corporate customer’s credit extension,” Malays. J. Comput. Sci., vol. 2022, no.1, pp. 116–131, 2022.
25. A. Magare, M. Lamin, and P. Chakrabarti, “Inherent Mapping Analysis of Agile Development Methodology through Design Thinking”,” Lecture Notes on Data Engineering and Communications Engineering, vol. 52, pp. 527–534, 2020.
26. A. R. B. M. Saleh, S. Venkatasubramanian, N. R. R. Paul, F. I. Maulana, F. Effendy, and D. K. Sharma, “Real-time monitoring system in IoT for achieving sustainability in the agricultural field,” in 2022 International Conference on Edge Computing and Applications (ICECAA), Tamil Nadu, India, 2022.
27. G. A. Ogunmola, M. E. Lourens, A. Chaudhary, V. Tripathi, F. Effendy, and D. K. Sharma, “A holistic and state of the art of understanding the linkages of smart-city healthcare technologies,” in 2022 3rd International Conference on Smart Electronics and Communication (ICOSEC), Trichy, India, 2022.
28. G. Kannan, M. Pattnaik, G. Karthikeyan, Balamurugan, P. J. Augustine, and Lohith, “Managing the supply chain for the crops directed from agricultural fields using blockchains,” in 2022 International Conference on Electronics and Renewable Systems (ICEARS), Tuticorin, India, pp. 908-913, 2022.
29. H. Sharma and D. K. Sharma, “A Study of Trend Growth Rate of Confirmed Cases, Death Cases and Recovery Cases of Covid-19 in Union Territories of India,” Turkish Journal of Computer and Mathematics Education, vol. 13, no. 2, pp. 569–582, 2022.
30. I. Nallathambi, R. Ramar, D. A. Pustokhin, I. V. Pustokhina, D. K. Sharma, and S. Sengan, “Prediction of influencing atmospheric conditions for explosion Avoidance in fireworks manufacturing Industry-A network approach,” Environ. Pollut., vol. 304, no. 11, p. 119182, 2022.
31. J. A. Alzubi, O. A. Alzubi, A. Singh, and T. Mahmod Alzubi, “A blockchain-enabled security management framework for mobile edge computing,” Int. J. Netw. Manage., vol. 33, no. 5, pp. 1-13, 2023.
32. J. J. Lohith, A. Abbas, and P. Deepak, “A Review of Attacks on Ad Hoc On Demand Vector (AODV) based Mobile Ad Hoc Networks (MANETS),” International Journal of Emerging Technologies and Innovative Research, vol. 2, no. 5, pp. 1483–1490, 2015.
33. K. Shah, P. Laxkar, and P. Chakrabarti, “A hypothesis on ideal Artificial Intelligence and associated wrong implications”,” Advances in Intelligent Systems and Computing, vol. 989, pp. 283–294, 2020.
34. K. Verma, P. Srivastava, and P. Chakrabarti, “Exploring structure oriented feature tag weighting algorithm for web documents identification”,” Communications in Computer and Information Science, vol. 837, pp. 169–180, 2018.

35. L. J, A. Manoj, G. Nanma, and P. Srinivasan, “TP-Detect: trigram-pixel based vulnerability detection for Ethereum smart contracts,” Multimed. Tools Appl., vol. 82, no. 23, pp. 36379–36393, 2023.
36. J. J. Lohith and B. Cahkravarthi, “Intensifying the lifetime of Wireless Sensor Network using a centralized energy accumulator node with RF energy transmission,” in 2015 IEEE International Advance Computing Conference (IACC), Bangalore, India, pp. 180–184, 2015.
37. K. Lohith and B. Singh, “Digital forensic framework for smart contract vulnerabilities using ensemble models” Multimed,” Multimed. Tools Appl, 2023, Press.
38. M. Tiwari, P. Chakrabarti, and T. Chakrabarti, “Novel work of diagnosis in liver cancer using Tree classifier on liver cancer dataset (BUPA liver disorder)” Communications in Computer and Information Science, vol. 837, pp. 155–160, 2018.
39. N. Al-Najdawi, S. Tedmori, O. A. Alzubi, O. Dorgham, and J. A. Alzubi, “A Frequency Based Hierarchical Fast Search Block Matching Algorithm for Fast Video Video Communications,” International Journal of Advanced Computer Science and Applications, vol. 7, no. 4, pp.1-15, 2016.
40. N. Priyadarshi, A. K. Bhoi, A. K. Sharma, P. K. Mallick, and P. Chakrabarti, “An efficient fuzzy logic control-based soft computing technique for grid-tied photovoltaic system”,” Advances in Intelligent Systems and Computing, vol. 1040, pp. 131–140, 2020.
41. O. A. Alzubi, I. Qiqieh, and J. A. Alzubi, “Fusion of deep learning based cyberattack detection and classification model for intelligent systems,” Cluster Comput., vol. 26, no. 2, pp. 1363–1374, 2023.
42. P. Chakrabarti and P. S. Goswami, “Approach towards realizing resource mining and secured information transfer” International Journal of Computer Science and Network Security, vol. 8, no. 7, pp. 345–350, 2008.
43. P. P. Dwivedi and D. K. Sharma, “Application of Shannon entropy and CoCoSo methods in selection of the most appropriate engineering sustainability components,” Cleaner Materials, vol. 5, no. 10, p. 100118, 2022.
44. P. S. Venkateswaran, F. T. M. Ayasrah, V. K. Nomula, P. Paramasivan, P. Anand, and K. Bogeshwaran, “Applications of artificial intelligence tools in higher education,” in Advances in Business Information Systems and Analytics, IGI Global, USA, pp. 124–136, 2023.
45. P. Sindhuja, A. Kousalya, N. R. R. Paul, B. Pant, P. Kumar, and D. K. Sharma, “A Novel Technique for Ensembled Learning based on Convolution Neural Network,” in 2022 International Conference on Edge Computing and Applications (ICECAA), IEEE, Tamil Nadu, India, pp. 1087–1091, 2022.
46. R. S. Gaayathri, S. S. Rajest, V. K. Nomula, R. Regin, “Bud-D: Enabling Bidirectional Communication with ChatGPT by adding Listening and Speaking Capabilities,” FMDB Transactions on Sustainable Computer Letters., vol. 1, no. 1, pp. 49–63, 2023.
47. R. Singh et al., “Smart healthcare system with light-weighted blockchain system and deep learning techniques,” Comput. Intell. Neurosci., vol. 2022, pp. 1–13, 2022.
48. S. A. Haider et al., “Energy-Efficient Self-Supervised Technique to Identify Abnormal User over 5G Network for E-Commerce,” IEEE Transactions on Consumer Electronics, vol. 2024, pp. 1–1, 2024.
49. S. Abukharis, J. A. Alzubi, O. A. Alzubi, S. Alamri, and T. O. Tim O\’Farrell, “Packet error rate performance of IEEE802.11g under Bluetooth interface,” Res. J. Appl. Sci. Eng. Technol., vol. 8, no. 12, pp. 1419–1423, 2014.
50. S. Khan, “Artificial Intelligence Virtual Assistants (Chatbots) are Innovative Investigators,” International Journal of Computer Science Network Security, vol. 20, no. 2, pp. 93–98, 2020.
51. S. Khan, M. Altayar, “Industrial internet of things: Investigation of the applications, issues, and challenges,” Int. J. Adv. Appl. Sci., vol. 8, no. 1, pp. 104–113, 2021.
52. S. Parthasarathy, A. Harikrishnan, G. Narayanan, L. J., and K. Singh, “Secure distributed medical record storage using blockchain and emergency sharing using multi-party computation,” in 2021 11th IFIP International Conference on New Technologies, Mobility and Security (NTMS), Paris, France, 2021.
53. S. Samadi, M. R. Khosravi, J. A. Alzubi, O. A. Alzubi, and V. G. Menon, “Optimum range of angle tracking radars: a theoretical computing,” Int. J. Electr. Comput. Eng. (IJECE), vol. 9, no. 3, p. 1765, 2019.
54. A. Sholiyi, T. O’Farrell, O. A. Alzubi, and J. A. Alzubi, “Performance evaluation of turbo codes in high speed downlink packet access using EXIT charts,” Int. J. Future Gener. Commun. Netw., vol. 10, no. 8, pp. 1–14, 2017.
55. R. Sudheesh et al., “Bidirectional encoder representations from transformers and deep learning model for analyzing smartphone-related tweets,” PeerJ Comput. Sci., vol. 9, no.14, p. e1432, 2023.
56. V. Chunduri, A. Kumar, A. Joshi, S. R. Jena, A. Jumaev, and S. More,” Int. J. Data Sci. Anal., 2023, Press.
57. V. K. Nomula, R. Steffi, and T. Shynu, “Examining the Far-Reaching Consequences of Advancing Trends in Electrical, Electronics, and Communications Technologies in Diverse Sectors,” FMDB Transactions on Sustainable Energy Sequence, vol. 1, no. 1, pp. 27–37, 2023.
58. D. Lu, S. Cheng, L. Wang, and S. Song, “Multi-scale feature progressive fusion network for remote sensing image change detection,” Sci. Rep., vol. 12, no. 1, pp.1-19, 2022.