International Journal of Aquatic Science
ISSN: 2008-8019
Vol 12, Issue 03, 2021

UAV-Based Animal Monitoring and Analysis Using Deep Learning

Naresh Edla[1], Harsh Lohiya[2]

[1]Research Scholar, Department of Computer Science and Engineering, Sri SatyaSai University of Technology and Medical Sciences, Sehore Bhopal-Indore Road, Madhya Pradesh, India
[2]Research Guide, Department of Computer Science and Engineering,Sri SatyaSai University of Technology and Medical Sciences, Sehore Bhopal-Indore Road, Madhya Pradesh, India

Abstract: Many remote sensing applications have found success in using deep learning for data processing. Agriculture, geology, disaster assessment, urban planning, and other uses of remote-sensed imagery, are just some of the areas where this technology has found use. In this study, we examine the many forms of deep learning architectures and their applications in the analysis of remotely sensed data.

1. INTRODUCTION

Drones, or unmanned aerial vehicles (UAVs), are autonomous aircraft that can fly missions without a person on board [1]. UAVs, or unmanned aerial vehicles, come in a wide variety of forms and are used for a wide range of purposes [2]. Anti-aircraft target tactics, information collection, and observation of enemy areas were the first uses of the technology by the military [2,3,4,5]. UAV technology has also advanced beyond its original use. It has become more important in many areas of human endeavour in recent years. Drones are a thriving industry because of the many uses UAVs can be put to [6]. This is mostly due to the technology's user-friendliness. In livestock farming, for example, UAVs streamline a number of tasks, allowing for more effective animal management [7, 8, 9, 10]. Environmental, economic, technological, and strategic planning challenges have arisen for the cattle farming industry throughout time as a result of climate change, population increase, and fierce competition for land and other natural resources [11]. However, UAVs have lately seen broad use among livestock farmers due to the integration of cutting-edge technologies such as Artificial Intelligence (AI), the Internet of Things (IoT), Machine Learning (ML), state-of-the-art sensors, etc. Figure 1 is an example of a common conceptual design framework for a UAV in a cattle farming management system (LFMS). The system's four phases of development include the water inspection system, the network design based on the Long-Range Wide-Area Network (LoRaWAN), the drone equipped with sensors and cameras, and the optimization of the drone's flight route. Transceivers are placed around the necks of the livestock. In this way, data may be sent between the drone and the base.

Figure 1: UAV based farm animal monitoring

The majority of research on UAVs for animal monitoring [8] has focused on wildlife species such deer [9,10,11], elk [9], hippo [12], rhino [13], and elephant [14]. The number of cattle studies has increased in recent years, although they are still uncommon overall. Animal counting and detection, cattle roundup, feeding behaviour, animal identification, and health monitoring have all been the focus of a small number of academic studies on this topic. The vast majority of these research (and notably the more recent ones) use deep learning for animal identification.

All deep learning-based algorithms for detecting cattle in the literature fall into one of two categories. The first method employs convolutional neural networks (CNNs) to produce a probability heat map that presumably reveals the locations of animals [15]. The second method makes use of bounding box generation methods to define the region in which an item is located. Both of these research projects make advantage of the speed-optimized YOLO v2 architecture [7]. While all of these research have revealed highly hopeful findings, it is difficult to assess the true relevance of their findings due to limitations in the experimental methods. Under particular, it seems that the aerial photographs were always acquired in perfect circumstances; nevertheless, elements such as aircraft height and ground sample distance (GSD) related, lighting, time of day, presence of shadows, etc. are not explored. It may be difficult to avoid potentially hazardous situations due to the lack of guidelines or norms to guide mission planning, and as a consequence, accuracy under real operating circumstances may be much lower. It's also important to note that in the first method, the CNN architectures used were developed for the particular situation at hand, and therefore comparisons with more well-established architectures are either few or non-existent.

Most of the aforementioned experimental constraints are likely attributable to the fact that these are the first research specifically devoted to cattle detection, therefore the emphasis is on proof of concept rather than the finer points that would allow for practical operation. The primary aim of this effort is to increase knowledge on the topic since it is crucial for the widespread use of this technology in the future. While accurate techniques of animal detection are crucial to any attempt at estimating animal populations, it is also necessary to account for a wide variety of other variables (animal movement, image matching, etc.).

2. LITERATURE REVIEW

Literature reviews in the field of remote sensing are offered in a wide variety of settings. As a refresher, Zhang et al. (2016) compiled a collection of resources explaining the use of DL techniques in image classification at the time. Cheng and Han (2016) followed up on this work, but instead of focusing on cutting-edge ANN and ML techniques, they honed down on more conventional approaches. Ball et al. (2017) provided a comprehensive and thorough evaluation of DL theories, techniques, and its problems in dealing with remote sensing data in a survey. With data from their own trials, Cheng et al. (2017) updated the field of picture categorization. Zhu et al. (2017), who also honed emphasis on classification, summed together much of the existing data necessary to comprehend the DL approaches taken to this problem. In addition, the overall performance of DNNs in publically accessible datasets for image classification job was investigated in a study conducted by Li et al. (2018), which provided insight into several DL applications. According to the study conducted by Yao et al. (2018), DL is set to replace traditional methods of image categorization in the remote sensing sector.

While DL has shown some positive outcomes, much more research and testing is needed. The widespread interest in the many remote sensing applications that make use of hyperspectral imaging (HSI) data began about the same time. Perhaps the first survey of hyperspectral data was conducted in Petersson et al. (2017). A interdisciplinary overview of the widespread use of DL models to the processing of HSI datasets is provided in (Signoroni et al., 2019). Authors Signoroni et al. (2019) summarised the use of DL models to process highly detailed remotely sensed HSI data into classification tasks, object detection, semantic segmentation, and data enhancement, including denoising, spatial super-resolution, and fusion, as one of the most emerging areas of application. In a recent study, Ado et al. (2020) discuss the use of unmanned aerial vehicle (UAV)-based sensors to capture hyperspectral imagery for use in agriculture and forestry. They demonstrate the availability of a variety of DL methods for handling the complexity of HSI datasets.

Jia et al. (2021) give a more current review of DL for hyperspectral image classification with limited labelled data. They remark on the discrepancy between DL models and HSI datasets since DL models often need sufficient labelled samples, while it is typically difficult to gather enough samples in HSI datasets owing to the complexity and time-consuming nature of hand labelling. However, by combining deep learning strategies with complementary approaches like transfer learning and a lightweight model, the problems associated with small-sample sets may be clearly recognised. Deep learning is also a novel technique for the area of infrared thermal imaging processing to address diverse domains, notably in satellite-provided data. To name a few examples, convolutional layers can be used to detect potholes in roads using terrestrial imagery (Aparna et al., 2019), land surface temperatures can be detected using a combination of multispectral and microwave observations from orbital platforms (Wang et al., 2020b), and sea surface temperature patterns can be determined to identify ocean temperature extremes (Xavier Prochaska et al., 2021).

3. METHODOLOGY

Input Datasets: From those first photographs, two data sets were compiled. In the first, 224224 pixel squares were assigned to every animal in every picture, and they were used to create a precise bounding box around each individual. Though more than half of an animal's body was visible, it was counted even if it was on the image's edge. Although the animal of interest was always in the centre of the frame, other animals may have been partially visible in some blocks of images due to their closeness. There were a total of 8629 animal-containing squares chosen, and an equal number of backdrop squares. The goal of creating this dataset was to find the optimal classification accuracy when all items of interest are properly framed by the picture blocks. In the second data set, pictures were chopped up into blocks of 224 by 224 pixels on a regular grid. This number was used because blocks of this size contain almost all of the animals seen in the photos, and because 224 by 224 pixels is the usual input size for many of the CNNs evaluated here. To determine which blocks were cattle and which were not, we looked at how many pixels inside each block could be definitively connected with an animal without using any other block (or the complete picture) as reference. The volume and variety of pictures made this the most practical method, despite the criterion's subjectivity and potential for inter- and intra-rater disagreements [14]. The number of "non-cattle" blocks was significantly larger, but only 14,489 randomly chosen "non-cattle" blocks were utilised in the trials to prevent difficulties connected with class imbalance. One individual personally annotated both picture collections.

Recurrent Neural Network

A RNN is able to detect temporal patterns in time-series inputs. To combat the poor memory retention typical of RNNs, researchers have developed techniques like long short-term memory (LSTM) and gated recurrent unit (GRU). Both LSTM and GRU can handle lengthy sequences of data as input. There have been several applications of RNN and its variations in the processing of temporal data in the area of remote sensing, including picture classification and correction using satellite imagery. Figure 2 is an example of change detection conducted on a multi-temporal picture. RNN is able to learn from temporal data because it retains the sequence of inputs and the connection between them. Effective change detection is made easier by learning the interdependencies between the time-series inputs. RNN networks, because of their memory's portability, can quickly learn the transformation rules of time-series pictures. It is not only the bi-temporal shifts in land cover that RNN is able to learn, but also the multi-temporal shifts.

RNN is a powerful tool for identifying the impacts of natural catastrophes on crops and other areas and for analysing the data collected from such areas. While most remote sensing applications have used a single deep neural network, several recent efforts have focused on integrating multiple networks to handle input from many sensors. The second part of this article discusses these remote sensing hybrid architectures.

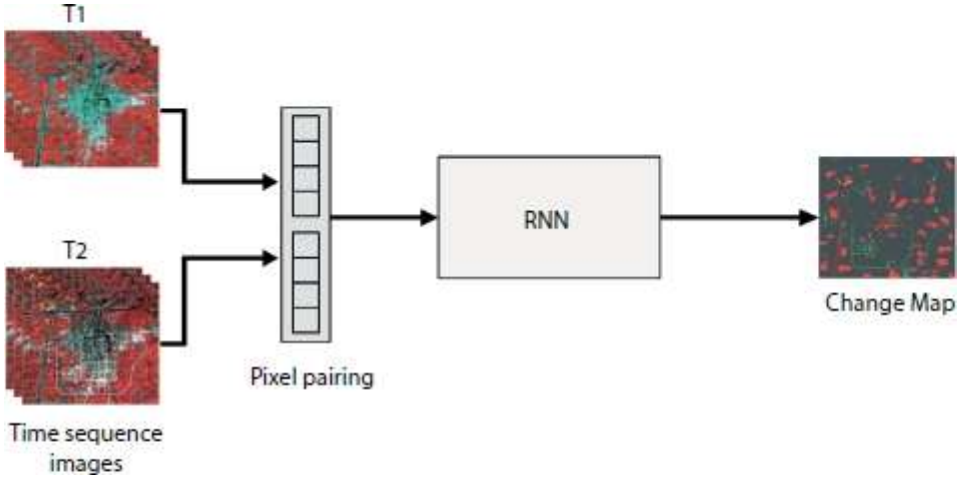

Figure 2: Change detection on multi-temporal images using RNN.

Generative Adversarial Network

To produce new instances that are similar to the input data, GAN is a generative model that automatically detects and learns patterns in the input data. Two convolutional neural networks (CNNs), a "generator" and "discriminator," are used to create a GAN. One of the primary applications of GAN in remote sensing is picture synthesis for data enhancement. The generation of high-resolution pictures is another use of GAN. Semi-supervised aircraft recognition from unlabeled airport pictures is achieved using GAN in [14]. Figure 3 demonstrates how GAN has been successfully used to pan-sharpening [2]. This technique, known as "pan-sharpening," combines high-resolution panchromatic and lower-resolution multispectral pictures into a single, high-resolution colour image. In [11], the writers attempted to use GAN image creation to fix the pan-sharpening issue. It has been taught to use an adaptive loss algorithm to provide crisp pan pictures.

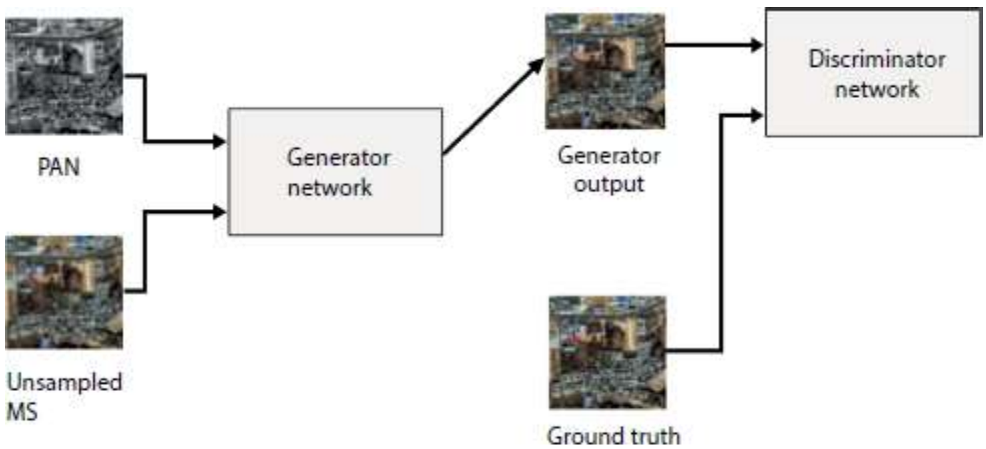

Figure 3: GAN for pan-sharpening with multispectral and panchromatic images**.**

Restricted Boltzmann Machine and Deep Belief Network

When given an input set, a restricted Boltzmann machine (RBM) may learn the probability distribution over that set. A bipartite graph is formed by the two sets of connections between the nodes, which may be either concealed or shown to the user. When RBMs are stacked, a deep belief network (DBN) is created. RBM is used for both classification and feature learning in remote sensing pictures. See Figure 4 for an example of RBM being used to classify hyper-spectral images. Nodes in the same layer in traditional RBM are linked to their counterparts in the other layer, while nodes in other layers are not interconnected. Instead of using binary visible layers, the authors of [5] use Gaussian ones, which allows for more accurate land cover categorization. Features are retrieved in an unsupervised manner using the Gaussian RBM, and then the features describing the land cover are supplied to a logistic regression unit.

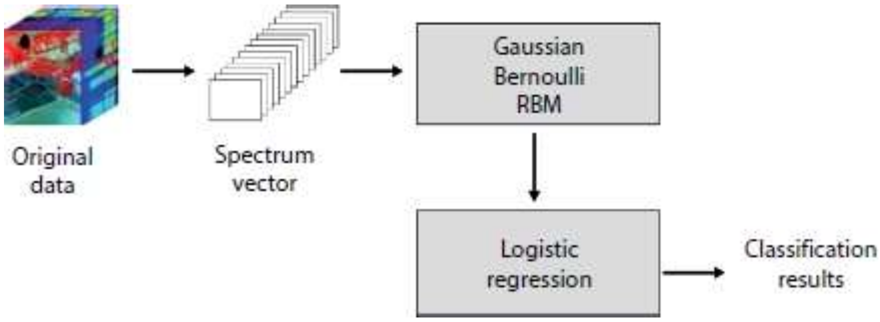

Figure 4: Gaussian-Bernoulli RBM for hyperspectral image classification.

Autoencoder

Known as an unsupervised deep learning model, autoencoder (AE) is capable of non-linear transformation. The unlabeled data is used by the non-linear function to discover the hidden surface[17]. The two networks in AE are the "encoder" and "decoder," with the former reducing the input's dimensionality and the latter reconstructing it. Dimensionality reduction [10] and feature variation [11] are two common uses of AE in remote sensing. Classification on unlabeled data is also a strength for AE [12]. Particularly, it is used for mapping snow cover [13]. Figure 5 depicts a stacked denoising autoencoder (SDAE) image classifier [14]. The idea of SDAE is to introduce noise into each encoder's input layer, therefore pushing the system to learn all of its crucial characteristics. It produces a noise-resistant network that is steady. The learning procedure in SDAE is bi-level. At the outset, the SDAE receives the unlabeled inputs. For unsupervised training, the features[18] are discovered by a greedy layer-by-layer strategy. The next step involves training a back propagation layer under supervision using labels.

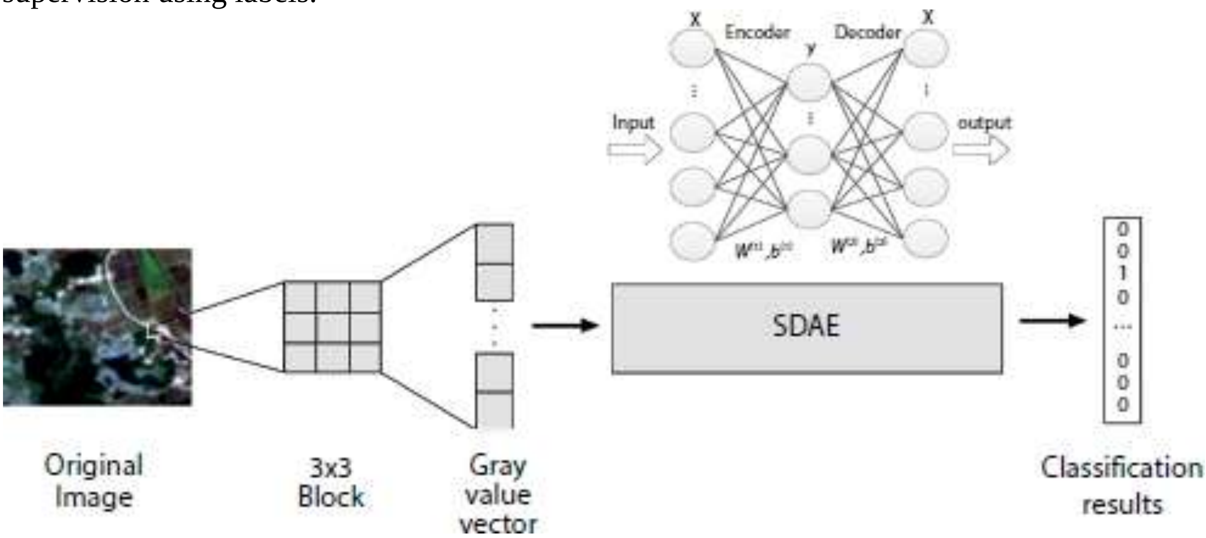

Figure 5: Remote sensing image classifier using stacked denoising autoencoder.

International Journal of Aquatic Science
ISSN: 2008-8019
Vol 12, Issue 03, 2021

Convolutional Neural Network

When it comes to picture classification and segmentation, the CNN has proven to be the most effective deep learning network to yet. As a result of its remarkable success in various computer vision tasks, the CNN has become a formidable resource. The proliferation of remote sensing satellites equipped with a wide range of sensors presents a potentially rich environment for CNN-based methods. The convolutional neural network (CNN) has been successfully used in remote sensing for object recognition and picture segmentation. CNN has been used successfully for detecting buildings [2, 3] and aeroplanes [1, 4]. For the most part, categorization is handled by the last completely linked layer in the network. Thus, categorization of land use and land cover using CNN has been more popular [4-7]. In Figure 6, we can see one example of a multi-scale classification performed using ResNet [8]. Objects on the ground, such as cities, farms, rivers, and woods, all have their own unique properties and resolutions, making it impossible to fully utilise these aspects using a single scale of observation. The ResNet-50 was chosen because it required less processing power to run than competing pre-trained models. ResNet-50 has an input size of 224 x 224 x 3 and 16 residual blocks with 3 convolutional layers. For accurate final projection of the land cover, we merge the boundary segmentation with the multi-scale patches and classify them with land cover category labels using a pre-trained ResNet-50. One of CNN's most useful functions is pan-sharpening [9].

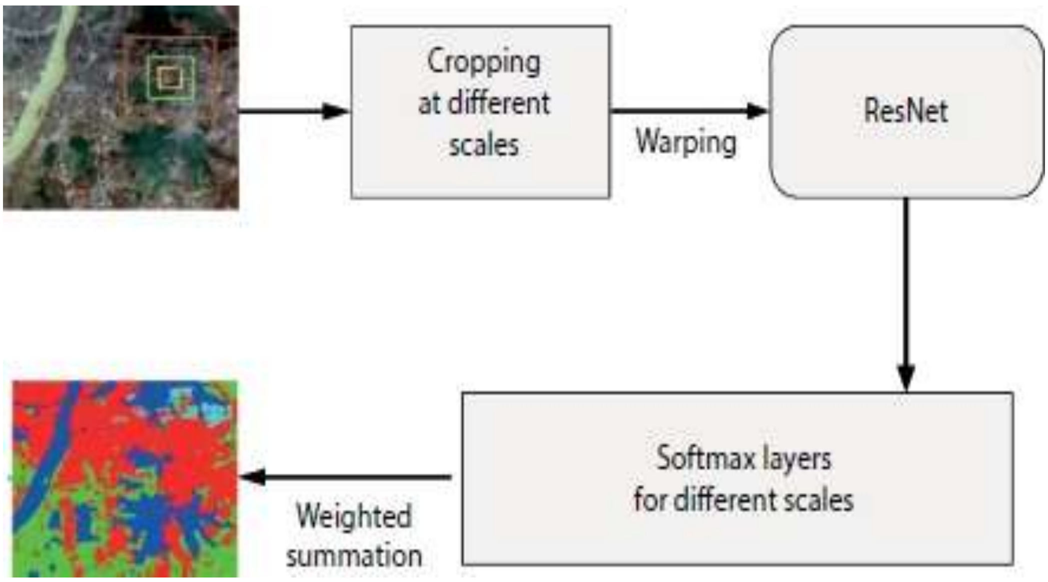

Figure 6: Land cover classification using CNN.

4. RESULTS

Using deep architectures on their own in remote sensing had shown mediocre success so far. Each deep learning architecture has advantages in certain use cases, therefore recent advancements in mixing them have produced greater outcomes. Change detection and categorization are two areas where hybrid architectures may shine. Additionally, picture fusion may be accomplished by using hybrid deep neural networks. Classification and change detection applications in remote sensing deal with enormous data sets, which significantly increases processing time. Different deep learning models are investigated, with the goal of effectively using their capabilities to solve this problem. The hybrid strategy leverages parallel processing across different networks to boost velocity and precision. Multiple hybrid deep architectures that can handle multi-sensor data are shown in Table 1.

Application of CNN-GAN to the problem of change detection has been reported [8]. For change detection, a generative learning approach may be used to learn the connection between the multi-temporal pictures and their respective change maps. Here is how the teaching and learning procedure works. Initial feature extraction using the CNN is used to clean up the multispectral pictures by getting rid of redundant data and noise. This mapping function is then learned using a GAN.

This hybrid design is also useful for clearing the sky of clouds. When there is cloud cover, important details in the picture are obscured. It's important to rebuild the missing data. To clear the skies in an optical picture, [9] suggests employing the matching SAR image in a two-step process including a convolutional neural network (CNN) and a generative adversarial network (GAN). First, the SAR picture is "transformed" into a simulated optical image that lacks fine detail and contains fewer colours. Second, a GAN is used to combine the simulated optical image with the SAR and hazy optical picture to generate a clear optical image.

When used to time-series image analysis, the RNN has demonstrated promising results. The LSTM network, a kind of RNN, excels at spotting shifts in data. Combining it with a common neural network has led to even greater gains. Here is how the CNN-LSTM network is trained to detect changes in the environment. In order to classify the input picture, the LSTM network uses the spatial and spectral properties extracted by the CNN. According to the chart, the total accuracy rate of 98.75% from the city expansion research performed using the CNN-LSTM hybrid network [10] is rather satisfying.

To create a unified picture from panchromatic and multispectral data, pan-sharpening has been accomplished by combining CNN and stacked autoencoder (SAE). According to [3], a hybrid architecture is developed that handles spatial and spectral information independently. SAE uses many layers to extract spectral characteristics from a multispectral picture. This is because CNN is able to extract the significant amounts of spatial information contained in panchromatic pictures. A fully linked layer is then fed the fused spatial and spectral data in order to classify water and urban regions. The hybrid design achieves a classification accuracy of 94.82% when applied to pan-sharpened images, which is much better than the accuracy achieved by any of the separate structures.

Hybrid architecture		**Inputs**	**Application**	**Results**
CNN	GAN	MS images	Change detection like farm monitoring	Overall error = 5.76%
CNN	GAN	SAR and MS	Cloud removal	SAM = 3.1621
CNN	LSTM	MS images	Change detection and animal monitoring	Overall accuracy = 98.75%
CNN	SAE	PAN and MS	Pan sharpening	Overall accuracy = 94.82%
CNN	GAN	Near-infrared (NIR) and short-wave infrared (SWIR)	Image Fusion	ERGAS = 0.69 SAM = 0.72
CNN	RNN	MS and PAN	To find oil spill in the ocean	mIoU = 82.1%

Table 1: Hybrid deep architectures for UAV.

SAM - Spectral Angle Mapper; ERGAS - Relative Dimensionless Global Error in Synthesis; mIoU - mean Intersection over Union.

To better identify oil spills in the water, a hybrid architecture was created by combining convolutional neural networks (CNNs) with recurrent neural networks (RNNs) [13]. The characteristics needed for oil spill segmentation can be extracted using a single deep CNN, however the model isn't very good at picking out the finer details. The hybrid architecture effectively catches the finer features of the oil spill in SAR pictures by adding conditional random fields in RNN. Mean intersection over union (mIoU), an assessment index for segmentation, is calculated to be 82.1 percent. It's a huge improvement over simple CNN setups. Additionally, the CNN-RNN network only requires 0.8 s to execute, which is much faster than the completely convolutional network architecture's 1.4 s.

Based on what has been said above, it is clear that a hybrid design provides much higher performance than a single architecture.

5. CONCLUSION

With the difficulties of remote sensing, such as the scarcity of ground truths and the difficulty of preprocessing, deep learning has become more important in recent years. This chapter provides a summary of the methods used to analyses remotely sensed images using deep learning. It has been suggested how remote sensing may benefit from the use of deep architectures like CNNs, RNNs, AEs, and GANs. Integration of data from many sensors using a hybrid of two separate architectures has also been proposed. Fusion's value in areas like as animal farm monitoring has been emphasized.

6. REFERENCES

[1] L. Zhang, L. Zhang, B. Du, "Deep learning for remote sensing data: A technical tutorial on the state of the art", IEEE Geosci. Remote Sens. Mag., 4 (2016), pp. 22-40.
[2] G. Cheng, J. Han, "A survey on object detection in optical remote sensing images" ISPRS J. Photogramm. Remote Sens., 117 (2016), pp. 11-28.
[3] Ball, J.E., Anderson, D.T., Chan, C.S., 2017, "A comprehensive survey of deep learning in remote sensing: Theories, tools and challenges for the community" arXiv 11.
[4] Cheng, G., Han, J., Lu, X., 2017, "Remote sensing image scene classification: Benchmark and state of the art". arXiv.
[5] X.X. Zhu, D. Tuia, L. Mou, G.S. Xia, L. Zhang, F. Xu, F. Fraundorfer, "Deep Learning in Remote Sensing: A Comprehensive Review and List of Resources" IEEE Geosci. Remote Sens. Mag., 5 (2017), pp. 8-36,
[6] Y. Li, H. Zhang, X. Xue, Y. Jiang, Q. Shen, "Deep learning for remote sensing image classification: A survey" Wiley Interdiscipl. Rev. Data Min. Knowl. Discov., 8 (2018), pp. 1-17.
[7] C. Yao, X. Luo, Y. Zhao, W. Zeng, X. Chen, "A review on image classification of remote sensing using deep learning" 2017 3rd IEEE International Conference on Computer and Communications, ICCC 2017 2018-Janua, 1947–1955 (2018).
[8] H. Petersson, D. Gustafsson, D. Bergström, "Hyperspectral image analysis using deep learning - A review" 2016 6th International Conference on Image Processing Theory, Tools and Applications, IPTA (2017).
[9] Signoroni, M. Savardi, A. Baronio, S. Benini, "Deep learning meets hyperspectral image analysis: A multidisciplinary review" J. Imag., 5 (2019).
[10] T. Ado, J. Hruka, L. Pdua, J. Bessa, E. Peres, R. Morais, J.J. Sousa, "Hyperspectral imaging: A review on uav-based sensors, data processing and applications for agriculture and forestry" Remote Sens., 12 (2020).
[11] S. Jia, S. Jiang, Z. Lin, N. Li, M. Xu, S. Yu, "A survey: Deep learning for hyperspectral image classification with few labeled samples" Neurocomputing, 448 (2021), pp. 179-204.
[12] Aparna, Bhatia, Y., Rai, R., Gupta, V., Aggarwal, N., Akula, A., 2019. Convolutional neural networks based potholes detection using thermal imaging. J. King Saud Univ. Comput. Inform. Sci. doi: https://doi.org/10.1016/j.jksuci.2019.02.004. URL
[13] S. Wang, J. Zhou, T. Lei, H. Wu, X. Zhang, J. Ma, H. Zhong, "Estimating land surface temperature from satellite passive microwave observations with the traditional

neural network, deep belief network, and convolutional neural network" Remote Sens., 12 (2020), 10.3390/RS12172691

[14] J. Xavier Prochaska, P.C. Cornillon, D.M. Reiman, "Deep learning of sea surface temperature patterns to identify ocean extremes" Remote Sens., 13 (2021), pp. 1-18, 10.3390/rs13040744

[15] Donahue, J.; Hendricks, L.A.; Guadarrama, S.; Rohrbach, M.; Saenko, K. Long-term recurrent convolutional networks for visual recognition and description. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 2625–2634

[16] Kumar, Ram and Sonaje, Vaibhav P and Jadhav, Vandana and Kolpyakwar, Anirudha Anil and Ranjan, Mritunjay K and Solunke, Hiralal and Ghonge, Mangesh and Ghonge, Mangesh, Internet Of Things Security For Industrial Applications Using Computational Intelligence (August 11, 2022). Available at SSRN: https://ssrn.com/abstract=4187998 or http://dx.doi.org/10.2139/ssrn.4187998

[17] Kumar, R., Singh, J.P., Srivastava, G. (2014). Altered Fingerprint Identification and Classification Using SP Detection and Fuzzy Classification. In: , et al. Proceedings of the Second International Conference on Soft Computing for Problem Solving (SocProS 2012), December 28-30, 2012. Advances in Intelligent Systems and Computing, vol 236. Springer, New Delhi. https://doi.org/10.1007/978-81-322-1602-5_139