Scene Illustration of Terrestrial Animals with Its Monitoring, Tracking and Recognizing Through Deep Learning in Relation with Granular Computing

Neelam Rawat, J.S. Sodhi, Rajesh K. Tyagi

Abstract: Combining Deep Learning Technique with Granular Computing employs an inductive paradigm for the terrestrial animal's elucidation. The proposed method frames the object (terrestrial animal) in arbitrary-shaped and sized granules rather than fixed and rectangular shaped, so that object can effectively mine and recognized. The goal is to present a formal model which automatically focus only on representative pixel of each granule rather than converting pixels from entire image through scanning. Thus, this work entails the process of recognizing not only the static animal in the background, but also depicts moving animal in foreground separately.
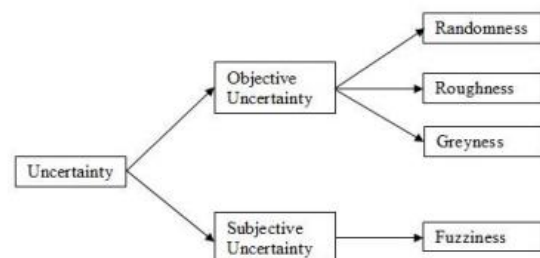
Keywords: Granular Computing (GrC), Deep Learning, Object Recognition, Object Tracking, CNN. GPRS

I. INTRODUCTION

From last decades, camera-trap images are the most powerful tool to evaluate and automatically extract the activities (e.g., eating, sleeping, running, drinking) which provides a valuable information to ecological studies [4]. In this paper, we extract to assess a specific hypothesis data that enable future ecological studies to have detailed and large-scale knowledge about the terrestrial animal movements in natural ecosystems that is even challenging for humans. In the monitoring and tracking process, frame rate goes high when detection process for target animal is with consecutive video frames. Even though complexity increases with addition of time and object orientation in tracking process [5]. The method of using both deep learning and granular computing can automate the process of monitoring and tracking of animals that dramatically improves the detection from those frames (granules).

A granule represents the subset of Universe of Discourse (UoD) which is composed of finer granules plotted by the functionality, similarity and distinguishability. Closeness and cohesion of any subset of objects can depict it as a granule. The concept of Granular Computing is to deal with uncertainty, imprecision, partial truth [7].

The process of representation and formation of granules called as Granulation (an operation performed on granules). Based on the application, granules can be fuzzy, rough fuzzy or crispy. [17]Process of decomposition makes larger granules into smaller or lower-level granules. Construction in granulation is bottom-up entrance or interrelationship whereas decomposition in granulation is top-down entrance or intra-relationship [3]. Granular Computing (GrC) models Uncertainty which may further categorize as Objective Uncertainty and Subjective Uncertainty (Figure 1.)

Figure 1: Dimensions of Uncertainty

- Models for subjective uncertainty
  - Fuzziness – fuzzy sets
- Models for objective uncertainty
  - Randomness – probability theory
  - Roughness – rough sets
  - Greyness – grey systems

Here, we have Objective Uncertainty, so we use rough set Granular Computing with Deep Learning Technique for recognition of animals and its motion detection [10].

The rest of the paper uphold with several sections as section 2 defines the fundamental understanding of utilized terms and techniques in this work. Section 3 extant the concept of deep learning concepts and the fusion of granular computing with deep learning as granulated deep learning exemplify in section 4. Proposed methodology and its result mentioned in section 5 and section 6 respectively followed be conclusions and references.

II. GRANULATION TECHNIQUE

In the proposed formal model, granular knowledge appears to focus on the representation of activities performed by animals with various activities likewise sleeping, eating, grazing, standing, waking and so on. Such granular frames, ascribes to be as granular models that quantify the source diversity of granular knowledge and reflect it as a higher order models.

* Correspondence Author
  Neelam Rawat, Research Scholar, Amity University, Noida, India. Email: neemarawat11@gmail.com
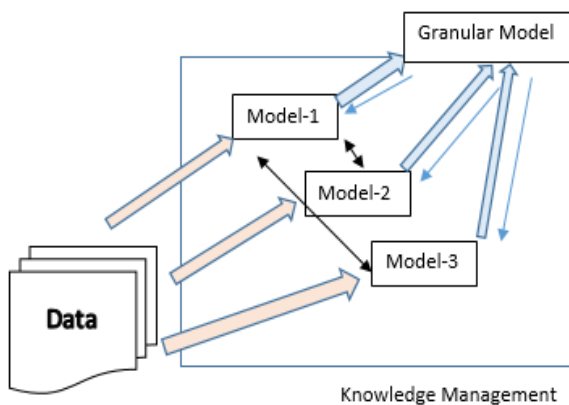 **J. S. Sodhi,** Amity University. Noida, India Email: jssodhi@akcgroup.com
  Rajesh K. Tyagi, Computer Science, Amity University, Gurugram, India. Email: profrajeshkumartyagi@gmail.com
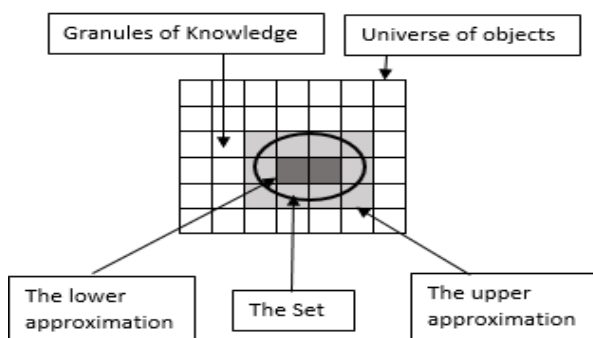
Several computational and algorithmic aspects associated with Granular Computing are presented and has shown a justifiable granularity principle for the construction of information granules which becomes opportune in the actualization of granular models [8][9].

A meaningful representation of an information granule in terms of numeric values instead single numeric entry would consider to be the actual essence of granularity. A rough description is required either through mean or median for representing a numeric data (which may also consider through standard deviation). For a certain probabilistic information granule, both mean and standard deviation imply as the principle aspects of its realization. Although probabilistic information granules are only a single possibility for constructing an information granule with numeric data [11][12]. Figure 2 shows how knowledge from different sources are encapsulated to form the granular model and all interactivities are represented through knowledge management space that evince all aspects of diversification at higher level of granularity.

Figure 2: Granular Model representation from diversity of knowledge.

Granulation and approximation in terms of formalism of knowledge granules can be illustrated as in Figure 3. For any object in an image of video sequence can be lower and upper approximation of that object. In simple way, a granule considers to be a collection of pixels which can represented as lower approximation of object if its pixels have its intensity value, other than that it represents upper approximation of objects in an image.

Figure 3: The definitions of approximations expressed regarding granules of knowledge

III. DEEP LEARNING TECHNIQUES

Deep Learning consider to be the form of more structured artificial neural network that uses the deep architecture of neurons which are linked together to replicate the human brain [6]. One of the deep neural network is CNN which is short for Convolutional Neural Network for image classification, recognition and detection.Deep CNN models technically train and test the images which gets around a sequence of layers that are fully connected with probabilistic values of 0 and 1 [18][19].For our work of recognizing terrestrial animals, five object detection models: R-CNN, Fast R-CNN, Faster R-CNN, Mask R-CNN and YOLO in Deep Learning are available [13][15].Among all models, we use mask R-CNN [1] in fusion with granular computing. The apprehension behind this is the wherewithal of image segmentation pixel-wise as segmentation at pixel-levelnecessitate more fine-grained alignment than bounding boxes which may further outlined to the regions of the original image more precisely and in better way.

IV. GRANULATED DEEP LEARNING

The artificial neural network (ANN) can be able to deal with all kinds of granular information of the real world. Multiple nodes in ANN represent the biological neurons in human brain which are linked and interact in such a way that nodes take the data and perform operations which further passed through neurons. As a result, activation of other nodes would form and the links have some weight which get altered by ANN to achieve adaptivity, speed, robustness and optimality. A popular neural network model – MLP, i.e., Multi-layer Perceptron that can update weights while training of nodes to prove its adaptivity. While comparing Deep learning (DL) considers to be the special form of Machine Learning where in CNN is the network used for images and videos.

Mask RCNN separates different objects in an image or video with an aiming to solve problem of instance segmentation and result to an object bounding box or mask that represent the object class. Mask R-CNN firstly generates regions from an input image and then predicts the class of that image with the help of refining the boundary box while generating a mask at pixel-level [2].

Given a segment composed of static and moving objects where static object consider the background. In recognition process, firstly the class of these objects is defined and then it tracks object movement. The method includes object background classification and object movement through granulated deep learning framework. Granulated background object (GB) can be denoted as:

$$G_B = B_U - (B_U \cap T)$$

Where, BU is the background which is approximated from the recognition process as in given algorithm 1.

Algorithm 1: Granulated Deep Learning for tracking object (terrestrial animal)

Input: V_i, $\forall i=1, 2, 3…,k$ where k is the total number of frames in a video V

Step 1: Consider the current video frame V_C and its previous frame V_{C-1}

Step 2: Calculate $T=| V_C - V_{C-1}|$ where T represents the temporal information.

Step 3: Execute granulation process on V_C frames.

Step 4: Calculate $B_U = \{\cup D_j, \forall_j=1, 2, 3….\}$ Such that $d_j<Th$ where $d_j\in D_j$ and Th is the Threshold for separating background and object.

Step 5: Calculate G_B, where G_B is Granulated model for background
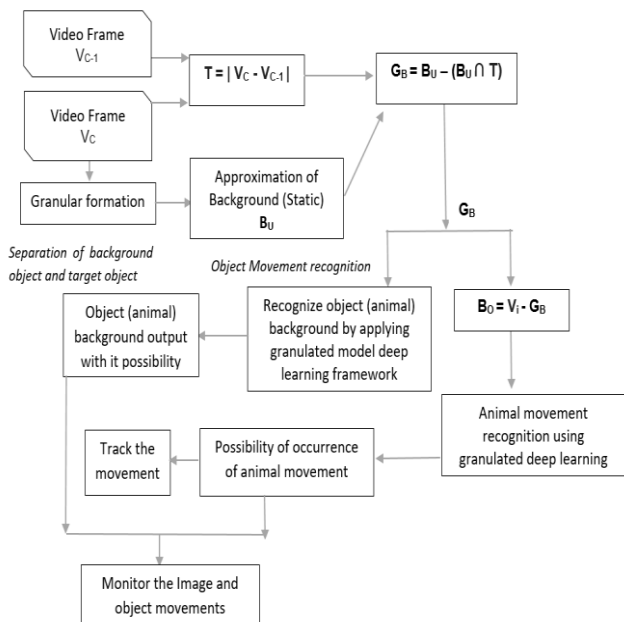
$$G_B = B_U – (B_U \cap T)$$

Step 6: Bolster G_B into the framework of granulated deep learning model so that it may detect the background (static) object.

Step 7: Calculate B_O as $B_O = V_i - G_B$, where B_O represent granulated object model

Step 8: Bolster B_O to the framework of granulated deep learning model

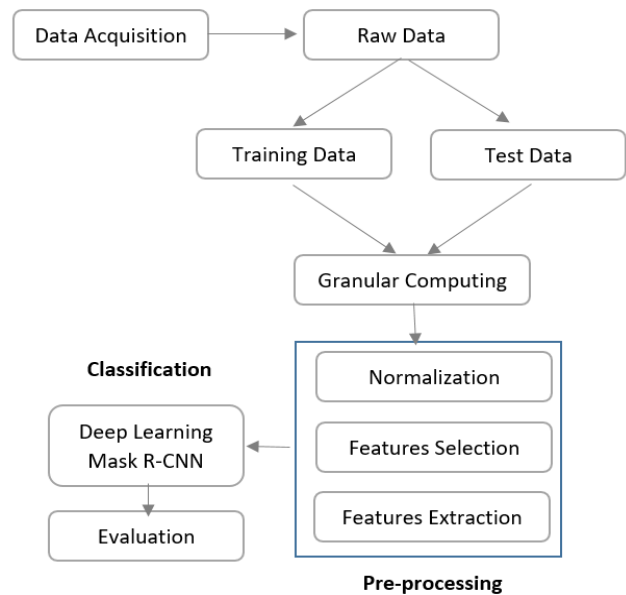Step 9: Place boundary boxes around object so that it gets tracked.

REPEAT steps for next video frame V_{C+1}

V. METHODOLOGY

Datasets: In our work, there is no benchmark, so instead of multiple species of terrestrial animals, we tackled with single-label classification (i.e., we took deer as one of the terrestrial animal). We consider a video of 2:10 minutes which captured the activities of set of deers. From the given video sequence, 261 frames have been formed. The dataset contains various activities of deer i.e., walking, standing, running and so on. The process of monitoring these activities and recognizing it would be defined in Figure 4.

Figure 4: Terrestrial animal activity monitoring and recognition process chart

Mask R-CNN identify outlines of an object at pixel-level which is called as instance segmentation i.e., given in Figure 7.

The whole process defines two stages where initially at first stage, Mask R-CNN scans input image to generate proposals i.e. define which contain objects. Further, at next stage, classification process is taken to outline the boundary boxes and masks. In compare to Faster R-CNN which is the successor of Mask R-CNN and most approved framework for detecting objects. Although, Mask R-CNN expands it with instance segmentation of an object from given video sequence. Mask R-CNN is composed of various modules at higher level: Initially detect corners and edges and subsequently detect objects that creates the backbone module of this process. [16] Region Proposal Network module represent as a very light-weighted neural network to scan the images and finds the area where objects are presented as given in Figure 9. Backbone features are scanned by RPN to define an outline of an object. In addition, RPN generates the following:

- Class for each object in the image
- Coordinates of bounding boxes for that image

In the module of Bounding box classifier for RoI, all classification inputs and outputs are proposed with a generation of bounding box of objects. In mask segmentation, regions of interests consider only positive region so that all masks get escalated.

Mask R-CNN works:

1. Consider an input image and move it to ConvNet to return the image feature map.
2. To achieve object proposals, RPN enforces on image feature maps so that it results objectness score of an image also.

3. Now to make the same size of each proposals, RoI pooling layer exercised on these object proposals.
4. To generate the mask for each proposal, fully connected layer has been applied so that it classify the object by defining the bounding boxes around it.
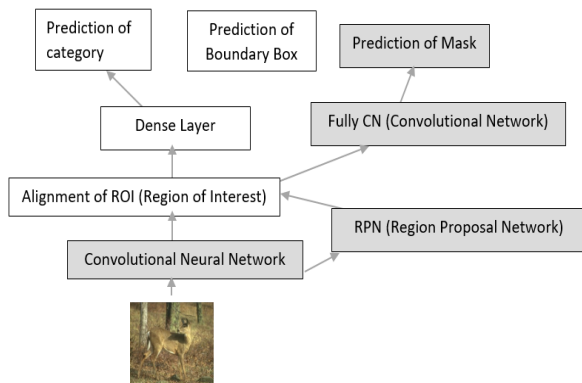

Figure 5: Mask R-CNN Model

Video Sequence Classification process steps:

1. Create validation and training dataset, scrutinize it. Training dataset is to train the object model and validation dataset to check the trained model.
2. From training and validation dataset, extract all frames from video sequences.
3. By using the frames from validation dataset to evaluate the object model. But before that, pre-process the frames to train it in the training dataset.
4. Steps remain continued until performance get satisfied. And thereafter, classify next video sequences for further training process.

Video Sequence Classification Model training process steps:

1. Apprehend all extracted video frames for training process.
2. Validation dataset is generated to evaluate the performance of unobserved data.
3. Model architecture then defined for object.
4. Find weights after model get trained.

Video Sequence Classification Model validation process steps:

1. Initially, re-define the architecture of object model and all weights get loaded.
2. Test data generation is defined.
3. Projects all test video frames for predictions of those test dataset.
4. Final object model gets determined and evaluated.

By using instance segmentation, we collect all the pixels of same properties and then create an outline which represents the mask at pixel level to make a clear picture of granularity of objects in image [14]. Images stored in matrix form and only vary according to pixels selected for the image. Matrix numbers represent the density of any image which easily quantifies the intensity of each pixel. Mask R-CNN successfully characterize the multi-level features so that to classify the target objects. Although, overall accuracies for

detection and monitoring would reflect a negative correlation but likely ways to enrich the training data to optimize the Mask R-CNN performance.

VI. RESULTS

In the proposed work, Figure 7 represents how an input image from a video sequence get classified, detected, segmented with a similarity index. In Figure 8, the instance segmentation has achieved from the following python code as sample code 1:

```
#ROI as region of object
region1 = cv2.imread(r'C:\Users\Neelam Rawat\.spyder-py3\image227.jpg')
region2 = cv2.cvtColor(region1,cv2.COLOR_BGR2HSV)
#target is the image we search in
target = cv2.imread(r'C:\Users\Neelam Rawat\.spyder-py3\image227.jpg')
region3 = cv2.cvtColor(target,cv2.COLOR_BGR2HSV)
# Find the outlines using Mask R-CNN
region4 = cv2.calcHist([region2],[0, 1], None, [180, 256], [0, 180, 0, 256] )
cv2.normalize(region4,region4,0,255,cv2.NORM_MINMAX)
R=cv2.calcBackProject([region3],[0,1],region4,[0,180,0,256],1)
disc = cv2.getStructuringElement(cv2.MORPH_ELLIPSE,(5,5))
cv2.filter2D(R,-1,disc,R)
R = np.uint8(R)
cv2.normalize(R,R,0,255,cv2.NORM_MINMAX)
ret,R = cv2.threshold(R,50,255,0)
result1=cv2.merge((R,R,R))
result2=cv2.bitwise_and(result1,target)
cv2.imshow('result2',res)
cv2.waitKey(0) & 0xFF
cv2.destroyAllWindows()
```
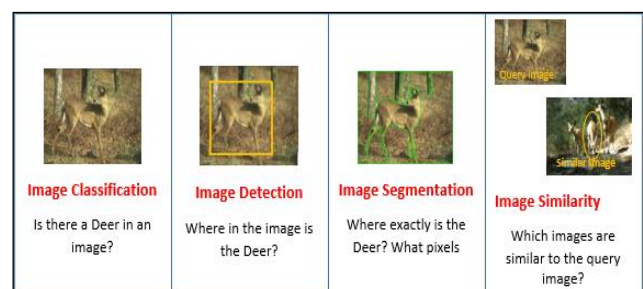
Code 1: Code for boundary boxes in targeted deer

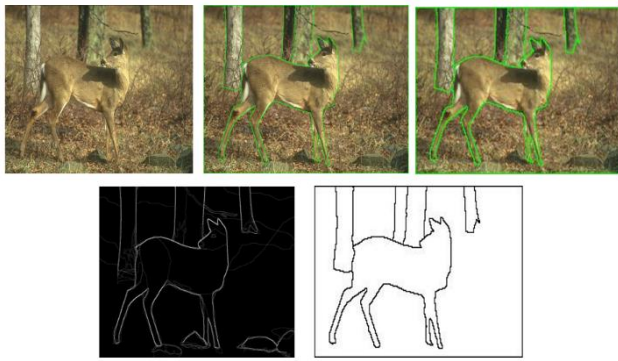**Figure 6: Detect Animals using Deep Learning with Granular Computing**

Figure 7: Deer detection through object outlined area

Sample codefor the detection of deer in Figure 8 is given as in code 2. Further, there is an arbitrary shape and size of image selection from the given video sequences is shown as in Figure 9.

```
# Initial Location
r,h,c,w = 250,90,600,225  # simply hardcoded the values
track_window = (c,r,w,h)

# set up the ROI for tracking
region1 = frame[r:r+h, c:c+w]
region_obj= cv2.cvtColor(frame,
    cv2.COLOR_BGR2HSV)
mask = cv2.inRange(region_obj, np.array((0., 60.,32.)),
    np.array((180.,255.,255.)))
region4 =
cv2.calcHist([region4],[0],mask,[180],[0,180])
cv2.normalize(region4,region4,0,255,cv2.NORM_MINM
    AX)
# Termination Criteria
criteria1 = ( cv2.TERM_CRITERIA_EPS |
    cv2.TERM_CRITERIA_COUNT, 10, 1 )
while(1):
ret ,frame = cap.read()
if ret == True:
obj_reg = cv2.cvtColor(frame, cv2.COLOR_BGR2HSV)
dst =
cv2.calcBackProject([obj_reg],[0],region4,[0,180],1)
```

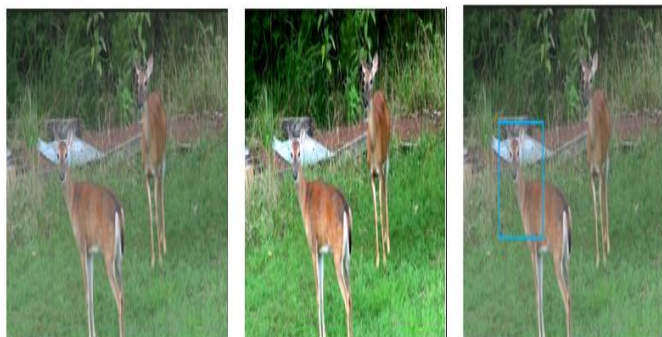Code 2: Code for defining boundary boxes of targeted deer

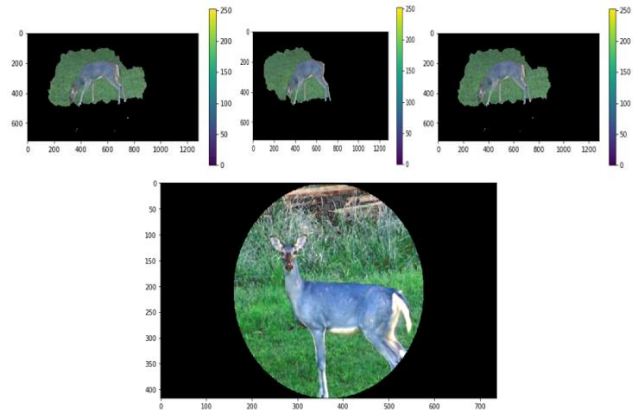**Figure 8: Region Proposal with boundary box in targeted deer from a video sequence**

Figure 9: Deer framing in an arbitrary fashion

VII.   CONCLUSIONS

In this work, a leading edge concept of granular deep learning framework has been defined to monitor, track and recognize terrestrial animals (i.e., deer in this case). Instead of treating convolutional layers as black box in earlier methods in deep learning, Mask R-CNN enforce distinguish layers in deep neural network form for image alignment. Overall, Higher resolution always improve object detection for objects that are far from cameras significantly and in mask R-CNN, number of proposals generated for region of interest is also considerably effective for monitoring and detecting movements. While comparing mask R-CNN with other detecting techniques likewise You Look Only Once (YOLO) or Single Shot Detectors (SSD) are also popular and have proved to be high performance. In addition, from the practical point of view, the proposed method needs a large scale with different range of terrestrial species.

ACKNOWLEDGEMENT

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

REFERENCES

1. Hu, Z., Fang, W., Gou, T., Wu, W., Hu, J., Zhou, S. and Mu, Y., 2019. A novel methodbased on a Mask
2. R-CNN model for processing dPCR images. Analytical Methods, 11(27), pp.3410-3418.
3. Hu, H., Pang, L. and Shi, Z., 2016. Image matting in the perception granular deep learning. Knowledge-Based Systems, 102, pp.51-63.
4. Lin, T.Y., 2003, May. Granular computing. In International Workshop on Rough Sets, Fuzzy Sets, Data Mining, and Granular-Soft Computing (pp. 16-24). Springer, Berlin, Heidelberg.
5. Li, Q., Qiu, Z., Yao, T., Mei, T., Rui, Y. and Luo, J., 2016, June. Action recognition by learning deep multi-granular spatio-temporal video representation. In Proceedings of the 2016 ACM on International Conference on Multimedia Retrieval (pp. 159-166). ACM.

Retrieval Number: B3830129219/2019©BEIESP
DOI: 10.35940/ijeat.B3830.129219
Journal Website: www.ijeat.org

2542

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

6. Lee, S.C. and Nevatia, R., 2014. Hierarchical abnormal event detection by real time and semi-real time multi-tasking video surveillance system. Machine vision and applications, 25(1), pp.133-143.
7. Wu, Z., Yao, T., Fu, Y. and Jiang, Y.G., 2016. Deep learning for video classification and captioning. arXiv preprint arXiv:1609.06782.
8. Rawat, N., Sodhi, J.S. and Tyagi, R., Wild Life Protection by Moving Object Data Mining-Discover with Granular Computing.
9. Rawat, N., Sodhi, J.S. and Tyagi, R.K., 2014, November. An algorithmic approach for analysis of animal movement with granular computing in relation with data mining. In 2014 International Conference on Contemporary Computing and Informatics (IC3I) (pp. 224-229). IEEE.
10. Rawat, Neelam; S. SODHI, J.; K. TYAGI, Rajesh. Analysis and tracking of animal movements through granulation of temporal domain (GTD). International Journal of Engineering & Technology, [S.l.], v. 7, n. 4.5, p. 501-505, sep. 2018. ISSN 2227-524X
11. PEDRYCZ, W., 2016. SYSTEM MODELING WITH FUZZY MODELS: FUNDAMENTAL DEVELOPMENTS AND PERSPECTIVES. Iranian Journal of Fuzzy Systems, 13(7), pp.1-14.
12. Loia, V., D'Aniello, G., Gaeta, A. and Orciuoli, F., 2016. Enforcing situation awareness with granular computing: a systematic overview and new perspectives. Granular Computing, 1(2), pp.127-143.
13. Loia, V., D'Aniello, G., Gaeta, A. and Orciuoli, F., 2016. Enforcing situation awareness with granular computing: a systematic overview and new perspectives. Granular Computing, 1(2), pp.127-143.
14. Rowcliffe, J.M., Field, J., Turvey, S.T. and Carbone, C., 2008. Estimating animal density using camera traps without the need for individual recognition. Journal of Applied Ecology, 45(4), pp.1228-1236.
15. Stern, U., He, R. and Yang, C.H., 2015. Analyzing animal behavior via classifying each video frame using convolutional neural networks. Scientific reports, 5, p.14351.
16. Torney, C.J., Lloyd-Jones, D.J., Chevallier, M., Moyer, D.C., Maliti, H.T., Mwita, M., Kohi, E.M. and Hopcraft, G.C., 2019. A comparison of deep learning and citizen science techniques for counting wildlife in aerial survey images. Methods in Ecology and Evolution, 10(6), pp.779-787.
17. Wang, X., Wang, H., Niu, S. and Zhang, J., 2019. Detection and localization of image forgeries using improved mask regional convolutional neural network.
18. Yao, Y.Y., 1998. A comparative study of fuzzy sets and rough sets. Information sciences, 109(1-4), pp.227-242.
19. Zhao, S., Liu, Y., Han, Y., Hong, R., Hu, Q. and Tian, Q., 2017. Pooling the convolutional layers in deep convnets for video action recognition. IEEE Transactions on Circuits and Systems for Video Technology, 28(8), pp.1839-1849.
20. Zeng, M., Nguyen, L.T., Yu, B., Mengshoel, O.J., Zhu, J., Wu, P. and Zhang, J., 2014, November. Convolutional neural networks for human activity recognition using mobile sensors. In 6th International Conference on Mobile Computing, Applications and Services (pp. 197-205). IEEE.

AUTHORS PROFILE

Neelam Rawat, Research Scholar from Amity University. Research Interests: Pattern Recognition, Artificial Intelligence, Machine Learning, Deep Learning. She has an experience of 17 years in academics and published papers in reputed journals and conferences. She aspires to motivate my faculties and students towards research and development program, who are the best resources of the Institute. Also, aspire to make work place as a center of excellence in traditional and innovation interdisciplinary, multidisciplinary research and development program.

Dr. J. S. Sodhi, CIO and Vice President of Amity Education Group and AKC Group of Companies. He has 25 years of Professional experience with Doctorate in Information Security. He is leading team of 150 IT professionals for overall IT management responsibility of Amity Institutions throughout 35 locations in India & 10 International locations and AKC Group of Companies. is also Executive Director of CCFIS (Cyborg Cyber Forensics and Information Security Pvt Ltd), a leading security research Organization under Amity Innovation Incubator. He has extreme caliber to motivate the human resources of the institute for best productivity by using value based persuasive talk.

Dr. Rajesh K. Tyagi, Professor in Computer Science and Engineering at AMITY University Gurgaon. He has committed to introduce the new state of art cutting edge labs facility from various funding agencies such as AICTE/ DST/ CSI/ IEEE/ ACM for getting new patents/ courses/ technologies in the Institute meticulously. He also has caliber to setup a cutting edge value added placements course such as dot net PHP/ Linux/ DBMS/ CCNA/ Android-based programs in work place for enhancement of learning and competitive attitude among students. He has made a link between an industry-academia for industry training as well as good placements.