International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

ISSN:2147-6799 **www.ijisae.org** **Original Research Paper**

Real-Time Object Detection and Classification Using Sparse Persistence Image-Based Color Directional Pattern (SPICDP) For Indoor Scenes

A.Sherly Alphonse[1], S.Abinaya[2], S.Abirami[3], G.Geo Jenefer[4], T.Benil[5]

Submitted: 10/09/2022 **Accepted:** 20/12/2022

Abstract*:* For the visual perception of mobile robots, object detection technique used in unseen real-world contexts is still a difficult task. Therefore, object recognition and localization, which are frequently referred to as detection, are crucial components of robot visual perception. The effectiveness of object detectors has significantly increased because of the quick development of deep learning networks. Topologically persistent characteristics, which rely on knowledge of an object's shape, are used in the suggested approach. Particularly, sparse persistence image (PI) feature types are retrieved in the proposed approach. Then, a Convolutional Neural Network (CNN) is trained to recognize objects using these properties. To do this, the system is first fed with the labelled training data. In contrast to prior object identification systems, the suggested approach takes input from videos or photos and classifies the objects using novel Sparse Persistence Image-Based Color Directional Pattern (SPICDP). The proposed approach achieves high accuracy than the other state-of-the-art methods.

Keywords: *Object, detection, Segmentation, persistent, feature.*

1.Introduction

The object detection, a field of computer science related to computer vision and image processing, finds instances of semantic objects belonging to a particular category in digital images and videos. Two well-studied detection domains are face and pedestrian detection. Deep learning has significantly influenced the world reacting to Artificial Intelligence (AI) during the previous limited years. Some of the well-known object identification methods include Faster RCNN, You Only Look Once (YOLO), and Region-based Convolutional Neural Networks (RCNN). When performance is more important than accuracy, YOLO excels Faster-RCNN and Single Shot Detector. Object recognition is used in many computer vision fields, including image retrieval and video surveillance. For the purpose of recognising objects, a novel method Sparse Persistence Image-based Color Directional Pattern (SPICDP) is projected in this work. The foremost contributions of the suggested technology are given below:

• The UW Indoor Scenes dataset is a novel dataset that is provided to assess the object recognition robustness in unobserved situations.
• It is demonstrated that topologically persistent features are more effective for recognition. They are more resilient to shifting surroundings than a cutting-edge cross-domain model for object detection.
• It is shown that Persistence Image (PI) features with less information have better recognition and performance compared to deep learning-based recognition using raw images.

The proposed technique SPICDP performs better than other object detection techniques when compared using recall and accuracy even in unfamiliar settings.

2.Related works

Redmon et al. [1] proposed YOLO, a novel method of object detection. Object detection has previously been accomplished using classifiers. They viewed object recognition as a regression issue to spatially discrete bounding boxes and associated class probabilities. In a single evaluation, a single neural network uses entire images to immediately forecast bounding boxes and probabilities of each class. Since the entire detection pipeline is made up of one network, detection performance may be adjusted throughout. Their seamless architecture is extremely fast. Their core YOLO model does real-time picture processing at a frame rate of 45 frames to each second. Fast YOLO, a scaled-down version of the network, processes just 155 frames each second while still outperforming other real-time detectors in mAP [1].

Thys et al. [2] proposed that in recent years, that an interest in adversarial attacks on machine learning models has grown. A Convolutional Neural Network's (CNN) output can be influenced to produce an entirely different outcome by making only little adjustments to the network's input. The initial process is by subtly altering the pixel values in an image to trick a classifier into producing the incorrect class. In other methods, "patches" that can be functional to an object to trick classifiers and detectors have been attempted to be learned. Some of these methods, such as altering an object and seizing it using a camera, have also

[1,]Vellore Institute of Technology, Chennai
9486452045, sherly.a@vit.ac.in
ORCID ID : 0000-0002-0019-9940
[2]Vellore Institute of Technology, Chennai
ORCID ID: 0000-0001-7957-7934
[3]Vellore Institute of Technology, Chennai
ORCID ID: 0000-0001-7467-5973
[4]St.Xavier's Catholic College of Engineering, Chunkankadai
ORCID ID: 0000-0002-0173-1417
[5]Vellore Institute of Technology, Chennai
ORCID ID: 0000-0002-2525-9081
** Corresponding Author Email: sherly.a@vit.ac.in*

demonstrated that the attacks are practicable in the real world. All of these methods, however, focusses on classes with essentially no intra-class variation [2].

Eriksen et al. [3] proposed that deep learning techniques have been used in recent years to generate outstanding improvements in object recognition. However, the difficulty of gathering and categorising training photos makes such techniques troublesome in real-world robotics applications. They provide a system that allows for the easy control of a robot to gather data that is pertinent to a given topic. The lifetime learning paradigm that this framework allows is that the robot gradually becomes more intelligent in its data collection and storage practises. This method is able to gather representative views of objects with less data requirements for long-term storage of datasets by iteratively training exclusively on image views that improve classifier performance.

Chen et al. [4] proposed that the H-divergence theory and the two domain adaptation components are implemented through the adversarial training of a domain classifier. In the Faster R-CNN model, the domain classifiers at various levels are more strengthened using consistency regularisation to create a Region Proposal Network (RPN). They tested their proposed methodology utilising a variety of datasets, such as Cityscapes, KITTI, SIM10K, etc. The outcomes show how well our suggested method works for reliable object detection in various domain shift scenarios.

Saito et al. [5] proposed that robust similarity of local features makes sense because it doesn't alter category-level semantics. Examples of such features are texture and colour. An unique detector adaption technique, based on robust local configuration and feeble global configuration is created. The main role is the global configuration, which places less emphasis on aligning globally dissimilar images and concentrates adversarial alignment loss on images which are globally similar. Moreover, they limit the strong model to only considering the feature map's local receptive fields. Additionally, all or a portion of the classes in the target domain that need to be detected are also present in the source domain. By adjusting the detector on two different kinds of intentionally and repeatedly created samples, they presented a dual-step advanced domain adaptation technique that starts with an object detector that has been pre-trained on the source domain. They evaluated their techniques using three-image domain datasets that they recently acquired, and saw an improvement in mean average precision (mAP) of between 5 to 20 percentage points when related to the basis that perform the best. Deep learning has had a significant impact on natural language processing, computer vision, movies, healthcare, machine learning, and 3D objects over the past ten years. Reininghaus et al. [6] proposed that topological data analysis provides a great source of insightful data for researching vision issues. But as of yet, they don't have a strong theoretical link to well-liked kernel-based learning methods like SVMs or PCA. To establish such a relationship, they developed a variety of scale-based kernels for use in persistence diagrams, a steady immediate depiction of topological features in data. It demonstrates the positive definiteness of this kernel and demonstrate its constancy with regard to the 1-Wasserstein distance. In tests on two datasets for 3D shape retrieval and texture identification, the proposed method outperforms a rival approach based on recently presented persistence settings. Deep learning has significantly influenced about the environment reacting to AI during the earlier limited years. Some of the well-known object identification methods include Faster RCNN, You Only Look Once (YOLO), Region-based Convolutional Neural Networks (RCNN), and Single Shot Detector (SSD). When performance is more important than accuracy, YOLO outclasses Faster-RCNN and SSD. Deep learning associates Mobile Nets with SSD to rapidly execute detection and tracking. However, all of these approaches currently in use typically take into account one resolution scale of appearance data using a common scale normalisation procedure. In addition to losing the possibility to find the associated appearance scales, this also eliminates the potentially relevant information. Its detection of rotated, masked, and distorted objects has limitations. Another drawback of the current technology is that it requires the input of photos or videos, which renders it unsuitable for real-time object detection. The proposed system overcomes all drawbacks and suggests a novel SPICDP for extraction of features from the images.

3.Proposed System

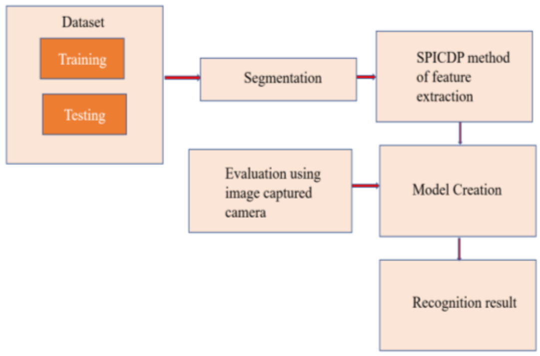

Fig.1. System architecture

For mobile robots' visual perception, object detection in unseen real-world contexts is still a difficult task. Therefore, object recognition and localisation, which are frequently referred to as detection, are crucial components of robot visual perception. The effectiveness of object detectors has significantly increased because of the quick development of deep learning networks for detecting tasks. Topologically persistent characteristics, which rely on knowledge of an object's shape, are used in the suggested approach. Amplitude and sparse persistence image (PI) are the two types of characteristics that are extracted. Then, a fully linked network is trained to recognise objects using these properties. The current cross domain object detection methodology is less resistant to shifting contexts than recognition using topologically persistent structures. Based on the data that is acquired using topologically persistent features, it is needed to identify the objects in the scene given an RGB image. Using a deep neural network, segmented images of the objects are created as in Fig.3. From the object segmentation maps, SPICDP features are extracted. The whole linked network is then fed these properties for recognition. Sample images from the dataset are given in Fig.2 [7].

3.1. Object Segmentation Map Generation

A segmentation method is used to create the segmentation maps for an input picture as in Fig.3. This work uses the cutting-edge DeepLabv3+ architecture [7] in particular, which is pre-trained on numerous classes and is thought to have a robust depiction of objects. Then, to select few images from the datasets, the network is trained using pixel-level foreground remarks. The outlines of the objects are used by a shape-based object identification technique to distinguish between various objects. Though, the segmentation is completed on photographs acquired from distances as much as two metres, the number of pixels is quite low. As a result, it is challenging for a segmentation prototype to accurately capture all of the fine details of an object's shape in a single frame [8]. Therefore, in order to maintain the objects' contour features, a dual-step segmentation framework is used. The segmentation prototype generates a segmentation map from the input image in the first stage. On this scene segmentation map, contour detection is used to produce the bounding boxes for the items.

Fig.2. Sample images from the Dataset

3.2 Generation of persistence-based features

In essence, segmentation maps are known as the binary pictures. Such binary images can be converted into a grayscale image using a variety of filtration functions, which is useful for creating a filtration of cubical complexes. A well-used filtering function called the height function is chosen, which generates a measurement that is sufficient to distinctively represent forms in R2 and surfaces in R3 in the persistent homology. Using transform bounding boxes, the scene image is split into a number of smaller images, each of which comprises a solo item. In step two, the same segmentation model that has been trained on these sub-images which is used to forecast the individual object segmentation maps [9].

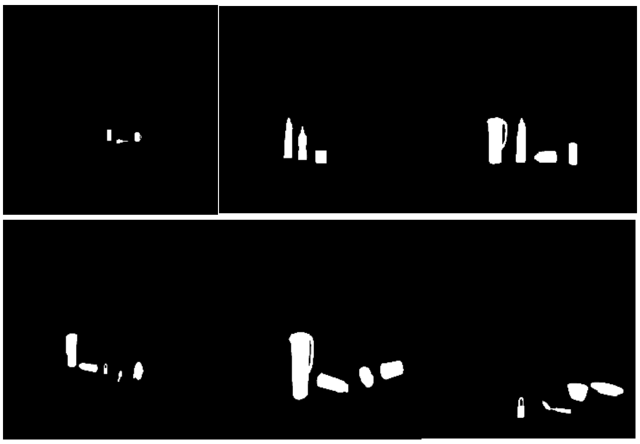

Fig.3. Sample images after segmentation

3.3 Sparse Persistence Image-based Color Directional Pattern (SPICDP) Features

Such a depiction is unsuitable for machine learning responsibilities because a PD's point density varies from shape to shape. Instead, some appropriate features are created for training the recognition network using the persistence image (PI) representation. However, topological information is weakly encoded in a small number of important pixel locations of the PIs that have nonzero entries. As a result, sparse sampling based on QR-pivoting is used to produce a Sparse PI. Each object segmentation mask's accompanying PDs are used to construct a set of d PIs for that mask. Fig. 4 shows sample persistence images (PIs). The group of PIs, calculated by means of height functions in various orientations on segmentation maps, take adequate data to differentiate among the two items [9].

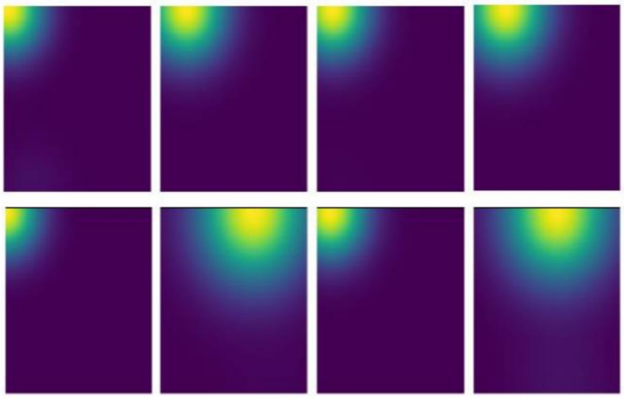

Fig.4. Persistence images

The SPICDP Pattern are obtained from the Persistence images for R, G and B channels. Edges are the areas of an image that correspond to the edges of objects. Edges are pixels where the brightness of an image changes suddenly. The behaviour of the image function in the pixel's neighbourhood is used to calculate an edge, a property that is associated to each individual pixel. The variable is a vector. The relationship between a pixel and its surrounding areas can be used to identify edge information in a picture. A pixel may show an edge point if its neighbours have highly disparate grey levels. Many use discrete approximations of differential operators and are implemented with convolution masks. The rate of variation in the image brightness is calculated by differential operations. A few operators provide orientation data. Other merely provide information on whether an edge exists at each point. Take one mask and rotate it in the eight principal compass directions: NW, N, SW, S, W, E, SE, and NE. The largest value obtained after each mask's convolution with the picture is the edge magnitude. The mask that yields the largest magnitude determines the edge direction.

The feature vector is created by SPICDP using the various channel combinations (R, G, B) of the skin photos after pre-processing. Three steps make up the feature extraction process: (i) employing compass masks to filter images, (ii) creating code images based on supreme response, and (iii) creating histograms and feature vectors. The feature vectors that were obtained as an output are used as the classification's input. Since the magnitudes got from the edges are extremely unchangeable to changes in light, this technique uses the edge data collected by means of the compass masks to generate the code picture and feature vector. Eight directional masks $\{D_{\theta_0}, D_{\theta_1} \ldots . D_{\theta_7}\}$, are utilised in this work, just as they are in fig. 3.3. For eight directions, the answer from each mask is regarded as $\{E_{\theta_0}, E_{\theta_1} \ldots . E_{\theta_7}\}$ accordingly [11]. Asymmetric Kirsch mask is taken into consideration in this work. The Kirsch mask is interchanged at an angle of 45 degrees to produce the eight masks shown in Fig.5. After that, the directional

Kirsch masks for eight different directions are used to convolute the three-by-three neighbourhoods of the image using the masks to filter out skin's edges. These masks result in eight solutions for every pixel. If complete responses are taken into account when creating the feature vector, the extent of the feature vector raises. Therefore, just the maximal response is used to create the feature vector in the proposed CLDP.

$$\begin{bmatrix}-3 & -3 & 5\\ -3 & 0 & 5\\ -3 & -3 & 5\end{bmatrix}\begin{bmatrix}-3 & 5 & 5\\ -3 & 0 & 5\\ -3 & -3 & -3\end{bmatrix}\begin{bmatrix}5 & 5 & 5\\ -3 & 0 & -3\\ -3 & -3 & -3\end{bmatrix}\begin{bmatrix}5 & 5 & -3\\ 5 & 0 & -3\\ -3 & -3 & -3\end{bmatrix}\begin{bmatrix}5 & -3 & -3\\ 5 & 0 & -3\\ 5 & -3 & -3\end{bmatrix}$$

$$D_{\theta_0} \qquad D_{\theta_1} \qquad D_{\theta_2} \qquad D_{\theta_3} \qquad D_{\theta_4}$$

$$\begin{bmatrix}-3 & -3 & -3\\ 5 & 0 & -3\\ 5 & 5 & -3\end{bmatrix}\begin{bmatrix}-3 & -3 & -3\\ -3 & 0 & -3\\ 5 & 5 & 5\end{bmatrix}\begin{bmatrix}-3 & -3 & -3\\ -3 & 0 & 5\\ -3 & 5 & 5\end{bmatrix}$$

$$D_{\theta_5} \qquad D_{\theta_6} \qquad D_{\theta_7}$$

Fig. 5. Eight directional Kirsch masks

Let $\{E_{\theta_0}, E_{\theta_1} \dots . E_{\theta_7}\}$ represent the eight responses that were received. The total number of positive and negative replies for each pixel are added together, and the highest response value out of the possible eight values for each and every pixel is selected to create the code image. Associated with the other existing ways, the proposed method greatly lowers the complexity of the code. Utilizing the direction data from the greatest response of respective pixel, a code picture, which is essentially a direction map, is formed. From this, a feature vector may be created by means of the histogram creation stage. This is what is specified in Eq. (1).

$$\text{Texture }(x, y) = \arg\ \max(E_{\theta_i}(x, y) |\ 0 \leq\ i\ \leq 7\ ) \qquad (1)$$

Where $D_{\theta_i}(x, y)$ denote the eight directions of a Kirsch mask, respectively, and $E_{\theta_i}(x, y)$ shows the response obtained using a directional mask $D_{\theta_i}(x, y)$ at specific position (x,y). The parameter i in this case represents the response's specific direction number. The highest response of respective pixel is utilized to make the Texture(x,y) of SPICDP, which is extremely robust and eliminates all of the noisy edge data.

Here, a feature vector is made and the histogram is formed. The ending feature vector, which captures the structural data, is created by concatenating the normalised histograms HS_i computed for each of the N grids that make up the code image. The combined histograms of each subregion, as shown in Eq. (2), make up the final feature vectors.

$$\text{feature vector of SPICDP} =< HS_1, HS_2 \dots \dots . HS_N > \qquad (2)$$

where N is how many smaller grids there are overall in the code image. The data of minor to major edges and corners can be extracted using this feature vector creation technique. When this SPICDP is functional to six channel groupings, this feature vector so collects colour data, edge data, and some texture data. The final feature vector used as input by CNN for classification is created by concatenating the feature vectors acquired using CLDP on all channel groupings, including SPICDP $_{R,R}$, SPICDP $_{G,G}$, SPICDP $_{B,B}$, SPICDP $_{R,G}$, SPICDP $_{R,B}$ and SPICDP $_{G,B}$. These features are then given to the CNN for classification [7] [12-15].

4. Results and discussion

This paper presents a topologically persistent features-based Real-Time Object Recognition system. It uses amplitude and sparse PI, two types of topologically persistent characteristics. Based on the shape data that is acquired using topologically persistent features, there is a need to identify complete objects in the scene given an RGB scene image. First, deep neural networks are used to create segmentation maps for the objects. The segmentation maps were then used to extract SPICDP features. A fully linked network is then fed these features for recognition. As recognition time is a crucial problem in practical object identification operations, the system features a straightforward architecture and great performance. The fact that the input for testing the model is acquired using the camera and used immediately in real-time for object detection is another significant benefit of the suggested approach. Several real photographs make up the dataset, which is separated into training and testing datasets with 70% and 30%, correspondingly. Images from the training dataset that have already been pre-processed and labelled are fetched throughout the training phase. By creating segmentation maps and extracting topologically persistent features, the suggested method's training is carried out. Topological data analysis (TDA) is a method for analysing datasets that uses topological methods. It can be difficult to extract data from datasets that are large-dimensional, partial, and noisy. TDA offers a broad framework for analysing data in such a way that is noise- and dimensionality-resistant and independent of the specific metric used. In addition, because of its topological character it is able to adapt to new mathematical techniques. The goal is to first examine the data's shape. TDA offers a promising link between topology and geometry because of its special characteristics. The machine is ready for testing after the training is finished and the model has been built. In order to prevent the training dataset from being overfit, the data that were removed from the larger dataset and used in the testing phase are called the testing dataset. The dataset is updated or removed based on the mistakes the machine made by repeatedly testing the objects. The trained model is put through an evaluation process once it has reached the necessary accuracy level. The evaluation data for the trained model is obtained by the proposed approach using a camera. Once the system receives the input, it detects the object that falls within the camera frame and displays its name on the screen. The proposed model's threshold level is 0.4, meaning that at this value, the recognition result will be correct as in Fig.6. The training and testing are done for both of our suggested algorithms five times using the identical groups of cropped items that were obtained using DeepLabv3+ [16] in Step 1. The implementations offered by Keras are used to train the networks for both ways. On the UW-IS dataset, the general presentation of both persistent features and object detection methods are analysed. The performance is compared against the commonly used Faster R-CNN object identification approach and Domain Adaptive Faster R-CNN. A pretrained model is utilized with the InceptionResNet-V2 feature extraction model and hyperparameters for Faster RCNN (also known as FR-CNN) [7]. A modified implementation is employed for Domain Adaptive Faster R-CNN, replacing the RoIpooling layer with the more well-liked RoI Align and the VGG backbone with ResNet-50. It has been demonstrated that the improved version outperforms other cutting-edge techniques for cross-domain object detection. This enhanced technique is known as DA-FR-CNN. The train and test of these techniques are done five times on both datasets, just as the proposed approaches. All of the training set photos with artificial lighting are utilised as the source domain in the DA-FR-CNN model, and all of the training set imageries with natural illumination are used as the target domain. According to performance studies, sparse PI-based recognition performs consistently in all environments and is significantly more accurate than amplitude-based prediction, whose accuracy drops by 4%. Without fine-tuning, the presentation of the current system Faster R-CNN significantly declines (accuracy falls by 25%). Additionally, the accuracy of the Domain Adaptive Faster RCNN (DA-FR-CNN) decreases significantly (by 18%). The SPICDP technique outdoes FR-CNN significantly and DA-FR-CNN by a tiny boundary as in Fig.7.

Fig.6. Recognition of Objects

In this work, it is suggested using topologically persistent traits to identify objects in enclosed spaces. From the object's binary segmentation maps, cubical complexes are created. There is a need to obtain several filtrations for every cubical complex by applying height functions in various directions. These filtrations are subjected to persistent homology to provide topologically persistent features that record the object's shape data. To train a fully linked identification network, SPICDP is proposed. In unknown situations, SPICDP features with CNN for classification outperform in terms of recognition performance compared to Faster R-CNN and its cutting-edge cousin DAFR-CNN.

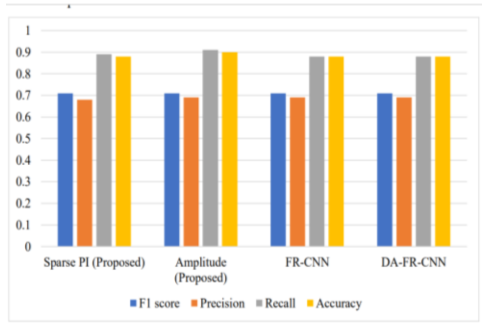

Fig.7. Performance Evaluation

5.Conclusion

In this paper a novel SPICDP feature is proposed which extracts the features from the persistent images. The SPICDP feature is extracted from the Red Green and the Blue Channels that helps to create a feature vector that comprises of effective texture and colour information. The SPICDP achieves good classification accuracy when combined with Convolutional Neural Network for classification. The comparison of results conclude that the proposed work achieves best performance than the other state-of-the art methods.

Author contributions

A.Sherly Alphonse: Conceptualization, Methodology, Software, Field study

S.Abinaya: Data curation, Writing-Original draft preparation, Software, Validation., Field study

S.Abirami: Visualization, Investigation, Writing-Reviewing.

G.Geo Jenefer: Supervision

T.Benil: Editing

Conflicts of interest

The authors declare no conflicts of interest.

References

[1] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You only look once: Unified, real-time object detection,” in IEEE Conf. Comput. Vis. Pattern Recognit., 2016, pp. 779–788.

[2] S. Thys, W. Van Ranst, and T. Goedem´e, “Fooling automated surveillance cameras: adversarial patches to attack person detection,” in IEEE Conf. Comput. Vis. Pattern Recognit. Workshop, 2019, pp. 0–0.

[3] C. Eriksen, A. Nicolai, and W. Smart, “Learning object classifiers with limited human supervision on a physical robot,” in IEEE Int. Conf. Robotic Comput., 2018, pp. 282–287.

[4] Y. Chen, W. Li, C. Sakaridis, D. Dai, and L. Van Gool, “Domain Adaptive Faster R-CNN for object detection in the wild,” in IEEE Conf. Comput. Vis. Pattern Recognit., 2018, pp. 3339–3348.

[5] K. Saito, Y. Ushiku, T. Harada, and K. Saenko, “Strong-weak

distribution alignment for adaptive object detection," in IEEE Conf. Comput. Vis. Pattern Recognit., 2019, pp. 6956–6965.
[6] J. Reininghaus, S. Huber, U. Bauer, and R. Kwitt, "A stable multi-scale kernel for topological machine learning," in IEEE Conf. Comput. Vis. Pattern Recognit., 2015, pp. 4741–4748.
[7] E.U.Samani, X.Yang, A.G.Banerjee, "UW Indoor Scenes (UW-IS) Dataset", Mendeley Data, V2, 2021,doi: 10.17632/dxzf29ttyh.2
[8] Ranadev, M. B. ., V. R. . Sheelavant, and R. L. . Chakrasali. "Predetermination of Performance Parameters of 3-Phase Induction Motor Using Numerical Technique Tools". International Journal on Recent and Innovation Trends in Computing and Communication, vol. 10, no. 6, June 2022, pp. 63-69, doi:10.17762/ijritcc.v10i6.5628.
[9] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770-778.
[10] E.U.Samani, X.Yang , A.G.Banerjee, "Visual object recognition in indoor environments using topologically persistent features," IEEE Robotics and Automation Letters. 2021 Jul 26;6(4):7509-16.
[11] S.Lazebnik, C.Schmid. J.Ponce, "Beyond Bags of Features: Spatial Pyramid Matching for Recognize Natural Scene Categories", in IEEE Conference on Computer Vision and Pattern Recognition, pp. 2169–2178. IEEE Press,2006 New York.
[12] T.Ojala, M.Pietikainen, T.Maenpaa, "Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns", in IEEE Transactions on Pattern Analysis and Machine Intelligence, 24(7),2002,pp. 971–987.
[13] Gupta, D. J. . (2022). A Study on Various Cloud Computing Technologies, Implementation Process, Categories and Application Use in Organisation. International Journal on Future Revolution in Computer Science &Amp; Communication Engineering, 8(1), 09–12. https://doi.org/10.17762/ijfrcsce.v8i1.2064
[14] Z.Yang ,L. Jin , D.Tao , "A comparative study of several feature extraction methods for person re-identification",In Chinese Conference on Biometric Recognition 2012 Dec 4 (pp. 268-277). Springer, Berlin, Heidelberg.
[15] A.KNsaif, S.H.Ali , K.N.Jassim , A.K.Nseaf, R.Sulaiman , A.Al-Qaraghuli , O.Wahdan , N.A.Nayan , "FRCNN-GNB: Cascade faster R-CNN with gabor filters and naïve Bayes for enhanced eye detection", IEEE Access. 2021 Jan 19; 9:15708-19.
[16] N.Jiang, J.Fang, Q.Yang, "Global-local domain adaptive object detection based on single shot multibox detector", Journal of Computer Applications. 2021 Feb 10;41(2):517.
[17] Andrie Dazlee, N. M. A., Abdul Khalil, S., Abdul-Rahman, S., & Mutalib, S. (2022). Object Detection for Autonomous Vehicles with Sensor-based Technology Using YOLO. International Journal of Intelligent Systems and Applications in Engineering, 10(1), 129–134. https://doi.org/10.18201/ijisae.2022.276
[18] G. Tauzin, U. Lupo, L. Tunstall, J. B. P´erez, M. Caorsi, A. MedinaMardones, A. Dassatti, and K. Hess, "giotto-tda: A topological data analysis toolkit for machine learning and data exploration," arXiv preprint arXiv:2004.02551, 2020.
[19] B.Baheti, S.Innani, S.Gajre, S.Talbar, "Semantic scene segmentation in unstructured environment with modified DeepLabV3+", Pattern Recognition Letters. 2020 Oct 1; 138:223-9.
[20] Linda R. Musser. (2020). Older Engineering Books are Open Educational Resources. Journal of Online Engineering Education, 11(2), 08–10. Retrieved from http://onlineengineeringeducation.com/index.php/joee/article/view/41
[21] A.S.Alphonse, D.Dharma, "A novel Monogenic Directional Pattern (MDP) and pseudo-Voigt kernel for facilitating the identification of facial emotion", Journal of Visual Communication and Image Representation. 2017 Nov 1;49:459-70.
[22] A.S.Alphonse, M.S.Starvin, "A novel maximum and minimum response-based Gabor (MMRG) feature extraction method for facial expression recognition", Multimedia tools and applications. 2019 Aug;78(16):23369-97.