A Review on Multi model-Deep Learning for Indoor - Outdoor Scene Recognition and Classification

[1] Pandit T. Nagrale, [2] Dr.Sarika Khandelwal

[1] PhD Scholar, Computer Science & Engineering, G.H.Raisoni University, Amravati.

[2] Associate Professor, Computer Science & Engineering, G.H.Raisoni College of Engineering, Nagpur.

[1] ptnagrale@gmail.com [2] sarikakhandelwal@gmail.com

Abstract

The advancement of computer vision to a new level has made it possible for laboratory robots to explore the physical environment. Even while this field makes strides, robots still have trouble figuring out their surroundings. Scene categorization is an important first step in gaining insight into the environment. A scene that can be classified at both the low and high levels. The primary function of automated surveillance applications such as indoor orientation, pedestrian identification, semantic categorizations, etc. is played by scene classification. DL approaches, and convolutional neural networks (CNN) in particular, have become a frequent option for scene categorization because they can remove features automatically and without the burden of doing so manually. CNNs with DL built in vastly outperform their predecessors. In order to create multi-layered learning models of neuron changes, it makes use of visual technology. In this work, a model for scene categorization that relies on characteristics extracted from deep CNN. The accuracy of scene classification has been improved by the application of transfer learning. To reduce parameter space and improve the feature quality, deep residual network is introduced as the feature extractor. From these kinds of non-technical encounters in the real world, it may gain insight into the reasons why transfer learning is possible. The goal of transfer learning is to apply the values of learned parameters from one model to another that has not yet been trained. This paper provides a brief introduction to Scene Classification, before moving on to Indoor-Outdoor Classification and the different deep learning techniques employed therein.

Index Terms—*Scene Classification, Indoor-Outdoor Classification, Transfer Learning, Deep Learning*

DOI Number: 10.14704/nq.2022.20.13.NQ88151 **Neuro Quantology 2022; 20(13):1187-1194**

Introduction:

The goal of scene understanding [2] is to provide an explanation for the content of the picture, the number of items in the scene, their positions in the scene, and the semantic connections between them. Natural sceneries with items like people, buildings, automobiles, highways, and shopping malls are easy for humans to recognise at a glance thanks to the scene identification process, which includes scene interpretation and scene classification. People generally have a fast, universal capacity for absorbing environmental sights. A scientific concept, "identification of scene" refers to the process of determining what kinds of scenes there are.

Understanding the scene from the speaker's point of view: a study of scene categorization and recognition algorithms [4] [5]. Image captioning is made easier thanks to scene recognition, which uses several approaches to guarantee that all relevant information is presented in a single, easily digestible word. The advancement of computer vision has opened up the world to robots previously confined to a lab. Although robots are making strides in this area, they still have a ways to go before they can fully comprehend their surroundings. The first step in gaining scene awareness is categorising the elements present.

Surveillance cameras, self-driving cars, home robots, and picture databases may all benefit from scene recognition technology [1-3] Cameras for monitoring activities have been installed in almost every public space. With the ever-increasing crime rate, it's crucial that everyone, from public spaces to private residences, have a camera installed. The need for operators to manage the camera's functions naturally rises. However, mistakes may be made, and not all data can be constantly watched.[7-12] Therefore, scene categorization may assist certain endeavours in overcoming these obstacles [6]. Image classification aims primarily to identify things within a picture. It's possible to tell different items apart using a number of techniques [8]. It classifies pictures into a few categories. Classifying photos into an indoor and outdoor category is the basis for image classification. Superb scene classification outcomes were accomplished by means of system mastery and in-depth knowledge acquisition strategies [11]. One of Computer Vision's outputs is scene detection, a Deep Learning implementation that calls for the development of an enhanced capacity for object identification. Larger datasets like ImageNet have substantially increased convolutional neural networks' (CNNs') ability to learn high-level features, but the quality of their scene identification output is still subpar. We create and deploy an automated system to recognise and identify the scenes in photos, a task that falls under the umbrella of visual recognition known as Scene Recognition[15-19]. One of the most popular current study directions in Deep Learning is automatic scene recognition and analysis. When we examine the development of Deep Learning over the last decade, we discover that Scene Recognition has been a major concern for scientists. The end game is to perform as well as the human brain. Scene recognition performance has not yet attained a high degree of efficiency, despite the availability of a few huge datasets like ImageNet, which have helped Convolutional Neural Networks (CNNs) in learning high-level features. We aimed to use thousands of labelled pictures from the Kaggle repository's Indoor scene image dataset to depict actual indoor locations and scenarios. To demonstrate how quantitative measurements like density and diversity can be used to forecast dataset biases and compare various data sets, we will utilise redesigned Convolutional Neural Networks (CNNs) techniques to attain maximum accuracy in scene recognition [22-27]. Our scene-centric database, Indoor Scenes by Kaggle, was developed since we found that ImageNet wasn't sufficiently challenging for Scene Recognition and related tasks. Our goal is to demonstrate state-of-the-art performance on all established scene criteria using the powerful technologies at our disposal. The use of Convolutional Neural Networks (CNNs) with combination features to boost accuracy performance on the Indoor Scenes dataset is a target. To demonstrate how these quantitative measurements may be used to forecast dataset biases and compare various data sets, we will apply re-engineered Convolutional Neural Networks (CNNs) techniques to attain maximum accuracy in scene recognition, as well as an unique measure of density and variety. Since scene-centric tasks like Scene Recognition struggle on image-centric datasets like ImageNet, we built Indoor Scenes by Kaggle. Our goal is to demonstrate state-of-the-art performance on all established scene criteria using the powerful technologies at our disposal. The goal is to use Convolutional Neural Networks (CNNs) with combination features to get better accuracy results on the Indoor Scenes dataset. Since deep Convolutional Neural Networks are designed to learn from and make use of a great deal of data, we want to build scene recognition methods that produce accurate results quickly. An integral aspect of Scene Recognition is the determination of where various items are sitting (e.g. bed, chair, sofa, light bulb, fan ...). Our goal is to influence people's ideas about utilising the category of locations by presenting a more detailed list of the indoor scenes in the photograph together with an explanation of their spatial relationships. To shorten the description and simplify the language, we can create a category of interior sceneries.

Literature Review

Yudi Da et al 2022 says that Human-centered 4D Scene Capture (HSC4D) may be used to correctly and rapidly build a living digital world that has expansive interior and outdoor settings, a wide range of human mobility, and complex interactions between people and their surrounds. Since it relies solely on inertial measurement units (IMUs) and light detection and ranging (LiDAR) placed on the user's body, HSC4D is unconstrained by physical location or the need for pre-built maps. In light of the fact that IMUs are capable of capturing human poses but always drift for long-period use and LiDAR is stable for global localization but rough for local positions and orientations, HSC4D makes both sensors complement each other by a joint optimization and achieves promising results for long-term capture.

Richard Strong Bowen et al 2021 introduce Foreground objects can be predicted using an object completion network (OCONet) based on their class. In order to forecast the texture and the extent of an object, OCONet employs an encoder-decoder architecture trained with adversarial loss. The latter is represented as a predicted signed-distance field. Background extension occurs in a separate process, and the object is composited on top using the anticipated mask. As demonstrated by our qualitative and quantitative findings, we outperform state-of-the-art picture extrapolation methods on difficult instances.

Adnan Ahmed Rafique et al 2020 proposed The recognition method is an unique segmentation framework that use statistical multi-object segmentation to create a robust scene model and then separate the items in the scene. The linear support vector machine is then used to extract the unique properties of the items that have been separated. At last, the scene recognition features and weights are fed into a multilayer perceptron. Significant gains were seen when comparing our method to the best of the current crop. This system works well in autonomous vision applications such as robotic vision, GPS-based position finders, sports, and security.

Abrar Ahmed et al 2019 says that It has become a critical area of study in computer vision to accurately segment and detect all mixed and obscured items in the complicated indoor/outdoor situations. Autonomous driving, traffic monitoring, security surveillance, human body part detection, object tracking, and scene recognition are all examples of crucial vision applications. Changes in lighting, occlusion, and orientation make it challenging to reliably identify every object in the image. Furthermore, multi-object detection accuracy is improved by segmenting the picture into its constituent pieces. In this work, we present the development of a system that utilises refined methods for precise multi-object segmentation and detection. To begin, we have combined the outcomes of two techniques for precise multi-object segmentation: I Decision trees, which label adjacent pixels and give each object in complex images its own colour, and (ii) Region-based segmentation, which detects multiple regions significantly and draws boundaries around each object. Second, detection is done by looking for items that have been given certain colours.

Aya M. Shaaban et al 2019 use With the use of a sizable dataset, a CNN is applied to the challenge of classifying interior spaces from their outside counterparts. The top levels of the Inception-v3 model are removed, and three more layers are added to accommodate the classification job, which is then utilised to extract scene data. The total 98.4 percent accuracy we achieve with our method demonstrates the superiority of CNN over the traditional methods.

Ahmad Jalal et al 2019 proposed Bowing, boxing, hand-shaking, high-fiving, hugging, kicking, patting, and pushing are just a few of the complex human actions that may be identified by a human activity recognition (HAR) system. To compare the efficacy of our system to that of other cutting-edge classifiers, we've developed a number of features algorithms, one of which is a convolutional neural network (CNN). While results from experiments show that the proposed method works well in complex realistic settings, it also has potential applications in areas such as security systems, e-learning, smart surveillance, violence detection, child abuse protection, elderly care, virtual games, intelligent video retrievals, and human-computer interaction.n.

3. Objective of Work

1. To present a technique for formulating low rank model for detected object
2. To study various dataset for indoor and outdoor image classifications for different environments
3. To study the existing machine learning and deep learning framework for their current efficiencies on indoor and outdoor scene recognition.
4. To make the novel multi deep learning framework which can be both used for indoor and outdoor environment
5. To have a comparative analysis our proposed framework with existing framework keeping the dataset constant.

4. Object detection

Many computer vision applications rely heavily on the ability to identify visual areas that correspond to moving objects. To put it simply, finding moving things in a digital picture sequence is what "object identification from video sequence" is all about. Identification and tracking of moving objects rely on motion detection. Detecting an item in motion over a series of photos is what we call "moving object detection." The capacity of a surveillance system to detect motion and record the next events is referred to as motion detection. When a moving item is detected, the surveillance camera is often triggered into action using a software-based monitoring system. Separating video streams into foreground and background content is valuable in and of itself, but identifying moving blobs also offers a focal point for identification, classification, and activity analysis, streamlining these procedures by ensuring that only "moving" pixels are taken into account. Temporal diffencing, background removal, and optical flow are the three traditional methods for detecting moving objects [9]. In dynamic settings, temporal differencing performs quite well, but it isn't great at isolating all of the important feature pixels. Despite providing the most comprehensive feature data, background removal is highly susceptible to dynamic scene changes brought on by lighting and other external factors. Independently moving objects may be detected in video with the help of optical flow, however most optical flow calculation methods are computationally demanding and cannot be applied to full-frame video streams in real time without dedicated gear.

4.1 Moving Object in real time environment:

Problems Several autonomous video surveillance systems rely heavily on the identification of moving objects. There is a lot more chaos and disorder in a real-time, out-of-control setting [1]. Several image-difference-based backdrop modelling approaches with great accuracy and real-time performance have emerged in recent years. Our goal here is to evaluate the efficacy of four different background modelling approaches—the Soft Impute Method, the Markov Random Field, graph cut segmentation, and the detection and removal of outlier Algorithm—using video sequences of outdoor scenes with wildly varying lighting and background motion due to weather-related outliers (any variation in illumination, distortion or noise in sequence of frames in video.). The data is analysed and provided with the intention of revealing promising avenues for strengthening motion detection methods in outdoor video surveillance systems [2]. We also analyse and examine motion in indoor and other contexts. [3] Provides an overview of the current state of video object detection research. This chapter delves into the many methods for detecting objects over several frames, including the optical flow approach, consecutive frame subtraction, and background subtraction.

5. The Proposed Methods

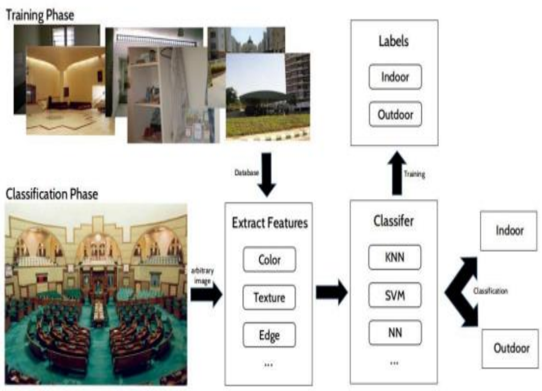
The purpose of this chapter is to discuss the actual implementation of the code, following up on the introduction of the project methodology and project tools and platform in the previous chapter. This document explains in depth how the system's many components function, as well as the processes involved in completing each module of the project. In this short report, we will briefly examine the project's code and explore the technologies and algorithms that were employed to bring this project to fruition. Moving The first stage in any video analysis process should always be object detection. Item detection is a prerequisite for every tracking technique, whether it be on a per-frame basis or only at the initial appearance of the tracked object against a static backdrop. Selecting a sample frame from a video can be useful when working with video data since the techniques can then be applied to the processing of all the frames in the film. This technique estimates the frame's foreground and background using a ranking parameter. Here, we examine the issue of moving object detection in low rank representation. To start, let's pretend there's no camera shake, only video outliers, and no camera motion at all. And hence, the project's design will involve three main stages.

Modules are as follows:

1. Moving Object Detection in Static Camera
2. Moving Object Detection in Moving Camera
3. Detect and Remove the Outliers

The pre-processing module [8], which includes all the fundamental steps needed to compute the data matrix and the transformation matrix, must be used by all three project modules. We'll begin developing the algorithm when we collect both input matrices. Figure 2.1 depicts the fundamental information flow in a system designed to identify moving objects. The object is segmented after undergoing preprocessing and a call to the soft-impute function. An RGB-to-grayscale video frame sequence is the algorithm's input. The technique provides a binary mask for each video frame. To create a binary mask, you'll give all of the background pixels a value of 0 and all of the other pixels a value of 1. The video frames needed for object detection from video are processed in the most fundamental ways by the pre-processing module.

6. Block Diagram



There are two main things in our methods

1) We describe the characteristics of scene coherence (SC) in an image of a scene and create a new loss, SCL, to use with feature extraction and classifiers. To train our proposed deep learning architecture, we use the SCL, the only loss specifically designed for scene identification.

2) In this paper, we offer a novel multi-deep learning architecture for integrating picture object and scene information. Our suggested method picks crucial characteristics and successfully fuses the two sets of features for training, in contrast to traditional fusion approaches in which the two features are simply concatenated and the classifier is created for the features.

First, it pre-processes all of the frames, beginning with the first and ending with the last. Then, it calculates the transform matrix for each frame, showing what changes, if any, have occurred in the object's location between frames. Each frame's preparatory work is depicted in this image. Because most variations appear in the middle of a movie, computing the transform matrix for a moving camera is different from the procedure used when working with a static image. Which transformation is used—a similarity transform, an affine transform, a rigid transform, or a projective transform—and therefore how many parameters are needed, is determined by the transformation matrix. [16].

Transformations are used:
• Position objects in a scene (modeling)
• Change the shape of objects
• Create multiple copies of objects
• Projection for virtual cameras

In the same way that one of the modules in the project is responsible for determining where items should be in a scene (modelling), another one must determine where an object is in a video. In order to register an item in a video sequence, an image registration technique is used. Find the transformation between pairwise points that are already in correspondence. Computation of transformation parameters from pairs of matching points requires first assuming that all pairs are connected by the same transformation. Converts a set of coordinates (x, y) from one system to a set of coordinates (x', y') in another.

7. FEATURES

To classify indoor and outdoor scenes, feature extraction is often thought to be significant. Developing and choosing discriminant characteristics is a popular area of study. Color, texture, and edge are all examples of low-level qualities that have been well studied. Some additional semantic characteristics or camera data have also been investigated to see if they can improve performance. Since the advent of convolutional neural networks (CNN), it has been widely held that CNN is best suited for learning deep scene characteristics, and therefore providing a system with superior performance in scene categorization.

7.1High-Level Feature

In the context of classifying indoor and outdoor scenes, it is typically used for low-level features. The spatial and structural information of the image's backdrop and objects is lost on low-level characteristics. It is important for scene image characteristics to not only define the entire image, but also to capture the information about the scene's backdrop and objects. As a result, a current direction in scene classification is study of novel methods for extracting scene features from photos. is now in route to somewhere[31-34].

7.2 Dynamic Scene

Certain robotics-related activities require dynamic scene categorization. Images that are continuous and time-correlated are considered dynamic. Most previous research has concentrated on situations from a single photograph. Classification methods for dynamic scenes differ greatly from those used for static images since the former tend to disregard potentially useful temporal and spatial clues. There is still a need to investigate dynamic indoor-outdoor scenario categorization.

7.3 Deep Learning

When the database's capacity approaches one million records, traditional methods based on low-level features have a hard time keeping up, while the deep learning-based approach has good performance. There has been a significant improvement in scene categorization thanks in large part to the deep convolution neural network. From a vast dataset of images, this convolution neural network can generalise to describe the picture. There has been an increasing convergence on the deep network's response properties as a universal representation of picture recognition[14]. There is still a lot of room for growth for convolutional neural networks in the computer vision industry. Scene categorization is sure to progress greatly in the near future, and convolution neural networks will be a key component of that progress..

8. CONCLUSIONS

In this study, we summarise the significant feature, classifier, and dataset developments during the past two decades in the field of Indoor-Outdoor scene categorization. There are still unresolved difficulties as the dataset grows and the needs of the various applications evolve, even though all the techniques with variable degrees of success stated classification accuracy might reach 88%. The deep learning-based method has recently shown promising results in a variety of computer vision applications. Yet, it remains challenging to articulate precisely how to categorise scenes as indoor or outdoor. Thus, additional breakthroughs in the Indoor-Outdoor scene categorization problem will require the attention of researchers from a variety of fields, including but not limited to neuroscience and machine learning, who may conduct interdisciplinary research to uncover new insights. Future plans involve constructing an Indoor-Outdoor dataset using previously collected data.

References:

[1] Pereira, Ricardo & Gonçalves, Nuno & Garrote, Luís & Silva, Tiago & Lopes, Ana & Nunes, Urbano. (2020). Deep-Learning based Global and Semantic Feature Fusion for Indoor Scene Classification. 67-73. 10.1109/ICARSC49921.2020.9096068.

[2] Ahmed, A.; Jalal, A.; Kim, K. A Novel Statistical Method for Scene Classification Based on Multi-Object Categorization and Logistic Regression. Sensors 2020, 20, 3871. [3]Matei, Alina & Glavan, Andreea & Talavera, Estefanía. (2020). Deep learning for scene recognition from visual data: a survey.

[3] Amlan Basu, Lykourgos Petropoulakis, Gaetano Di Caterina, John Soraghan, Indoor Home Scene Recognition Using Capsule Neural Networks, Procedia Computer Science, Volume 167, 2020, Pages 440-448, ISSN 1877-0509, [5]Glavan, Andreea & Talavera, Estefanía. (2021). InstaIndoor and Multi-modal Deep Learning for Indoor Scene Recognition.

[4] Binglin Niu, "Semantic Segmentation of Remote Sensing Image Based on Convolutional Neural Network and Mask Generation", Mathematical Problems in Engineering, vol. 2021, Article ID 2472726, 13 pages, 2021. https://doi.org/10.1155/2021/2472726

[5] Petrovska, B.; Zdravevski, E.; Lameski, P.; Corizzo, R.; Štajduhar, I.; Lerga, J. Deep Learning for Feature Extraction in Remote Sensing: A Case-Study of Aerial Scene Classification. Sensors 2020, 20, 3906. https://doi.org/10.3390/s20143906

[6] Sitaula, Chiranjibi & Xiang, Yong & Zhang, Yushu & Lu, Xuequan & Aryal, Sunil. (2019). Indoor Image Representation by High-Level Semantic Features. IEEE Access. PP. 1-1. 10.1109/ACCESS.2019.2925002.

[7] Zeng, Delu & Liao, Minyu & Tavakolian, Mohammad & Guo, Yulan & Zhou, Bolei & Hu, Dewen & Pietikäinen, Matti & Liu, Li. (2021). Deep Learning for Scene Classification: A Survey.

[8] Gongwei Chen; Xinhang Song; Bohan Wang; and Shuqiang Jiang.(2021).See More for Scene: Pairwise Consistency Learningfor Scene Classification.35th Conference on Neural Information Processing Systems (NeurIPS 2021).

[9] Ali, Hafiz & Kabir, Summiya & Ullah, Ghufran. (2021). Indoor Scene Recognition using ResNet-18. International Journal of Research Publications. 69. 10.47119/IJRP100691120211667.

[10] Oh, J.; Kim, M.; Ban, S.-W. Deep Learning Model with Transfer Learning to Infer Personal Preferences in Images. Appl. Sci. 2020, 10, 7641. https://doi.org/10.3390/app10217641

[11]Abeßer, J. A Review of Deep Learning Based Methods for Acoustic Scene Classification. Appl. Sci. 2020, 10, 2020. https://doi.org/10.3390/app10062020

[12] Jianjun Ni, Tao Gong, Yafei Gu, Jinxiu Zhu, Xinnan Fan, "An Improved Deep Residual Network-Based Semantic Simultaneous Localization and Mapping Method for Monocular Vision Robot", Computational Intelligence and Neuroscience, vol. 2020, Article ID 7490840, 14 pages, 2020. https://doi.org/10.1155/2020/7490840

[13] Mao Jiafa, Wang Weifeng, Hu Yahong & Sheng Weiguo (2019) A scene recognition algorithm based on deep residual network, Systems Science & Control Engineering, 7:1, 243-251, DOI: 10.1080/21642583.2019.1647576

[14] Priya Singla; Rajesh Mehra.Scene Recognition using Significant Feature Detection Technique.International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-9 Issue-3, January 2020.DOI: 10.35940/ijitee.C8653.019320 18

[15] Gopalapillai, R.; Gupta, D.; Zakariah, M.; Alotaibi, Y.A. Convolution-Based Encoding of Depth Images for Transfer Learning in RGB-D Scene Classification. Sensors 2021, 21, 7950. https://doi.org/10.3390/s21237950

[16] Sarfaraz Masood, Umer Ahsan, Fatima Munawwar, Danish Raza Rizvi, Mumtaz Ahmed, Scene Recognition from Image Using Convolutional Neural Network, Procedia Computer Science, Volume 167, 2020, Pages 1005-1012, ISSN1877-0509, https://doi.org/10.1016/j.procs.2020.03.400.

[17] Shaopeng Liu, Guohui Tian, "An Indoor Scene Classification Method for Service Robot Based on CNN Feature", Journal of Robotics, vol. 2019, Article ID 8591035, 12 pages, 2019. https://doi.org/10.1155/2019/8591035

[18] Huang, Shengyu & Usvyatsov, Mikhail. (2020). Indoor Scene Recognition in 3D. 8041-8048. 10.1109/IROS45743.2020.9341580.

[19] Sihua Sun, "Digital Audio Scene Recognition Method Based on Machine Learning Technology", Scientific Programming, vol. 2021, Article ID 2388697, 9 pages, 2021. https://doi.org/10.1155/2021/2388697

[20] Tadros, Timothy & Cullen, Nicholas & Greene, Michelle & Cooper, Emily. (2019). Assessing Neural Network Scene Classification from Degraded Images. ACM Transactions on Applied Perception. 16. 1-20. 10.1145/3342349.

[21] Shao, X.; Zhang, X.; Tang, G.; Bao, B. Scene Recognition Based on Recurrent Memorized Attention Network. Electronics 2020, 9, 2038. https://doi.org/10.3390/electronics9122038

[22] Yifei Zhang, Désiré Sidibé, Olivier Morel, Fabrice Mériaudeau. Deep multimodal fusion for semantic image segmentation: A survey. Image and Vision Computing, Elsevier, 2021, 105, pp.104042. 10.1016/j.imavis.2020.104042. hal-02963619

[23]Qiu, Y.; Satoh, Y.; Suzuki, R.; Iwata, K.; Kataoka, H. Indoor Scene Change Captioning Based on Multimodality Data. Sensors 2020, 20, 4761. https://doi.org/10.3390/s20174761

[24] B. Labinghisa and D. M. Lee, "A Deep Learning based Scene Recognition Algorithm for Indoor Localization," 2021

International Conference on Artificial Intelligence in Information and Communication (ICAIIC), 2021, pp. 167-170, doi: 10.1109/ICAIIC51459.2021.9415278.

[25] Anwer, R.M., Khan, F.S., Laaksonen, J., Zaki, N. (2019). Multi-stream Convolutional Networks for Indoor Scene Recognition. In: Vento, M., Percannella, G. (eds) Computer Analysis of Images and Patterns. CAIP 2019. Lecture Notes in Computer Science(), vol 11678. Springer, Cham. https://doi.org/10.1007/978-3-030-29888-3_16

[26] Albert Mosella-Montoro, Javier Ruiz-Hidalgo, 2D–3D Geometric Fusion network using Multi-Neighbourhood Graph Convolution for RGB-D indoor scene classification, Information 19 Fusion, Volume 76, 2021, Pages 46-54, ISSN 1566-2535, https://doi.org/10.1016/j.inffus.2021.05.002.

[27] Rui Zhang, Guangyun Li, Thomas Wunderlich, Li Wang, A survey on deep learning-based precise boundary recovery of semantic segmentation for images and point clouds, International Journal of Applied Earth Observation and Geoinformation, Volume 102, 2021, 102411, ISSN 0303-2434, https://doi.org/10.1016/j.jag.2021.102411.

[28] Haibin YAN, Yating LV, Venice Erin LIONG, Structure-aware fusion network for 3D scene understanding, Chinese Journal of Aeronautics, 2021, , ISSN 1000-9361, https://doi.org/10.1016/j.cja.2021.07.012.

[29] H. Seong, J. Hyun and E. Kim, "FOSNet: An End-to-End Trainable Deep Neural Network for Scene Recognition," in IEEE Access, vol. 8, pp. 82066-82077, 2020, doi: 10.1109/ACCESS.2020.2989863.

[30] B. Zhou, A. Lapedriza, A. Khosla, A. Oliva, and A. Torralba, “Places:A 10 million image database for scene recognition,” IEEE Transactionson Pattern Analysis and Machine Intelligence, vol. 40, no. 6, pp. 1452–1464, 2018

[31] J. Xiao, J. Hays, K. A. Ehinger, A. Oliva, and A. Torralba, “Sundatabase: Large-scale scene recognition from abbey to zoo,” in Proceedings of the IEEE Conference on Computer Vision and PatternRecognition (CVPR), 2010, pp. 3485–3492.

[32] A. Quattoni and A. Torralba, “Recognizing indoor scenes,” in Proceedings of the IEEE Conference on Computer Vision and PatternRecognition (CVPR), 2009, pp. 413–420.

[33] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein et al., “Imagenet large scale visual recognition challenge,” International Journal of Computer Vision, vol. 115, no. 3, pp. 211–252, 2015

[34] Vishakha Mehta, Sarika Khandelwal, Ashish Kumar Kumawat,” A Survey on Face Recognition algorithm” 2018 2nd International Conference on Trends in Electronics and Informatics (ICOEI) IEEE Xplore 02 December 2018