Bulletin of Business and Economics, 12(3), 1015-1019
https://bbejournal.com
https://bbejournal.com/index.php/BBE/article/view/530
ISSN Print: 2414-1771
ISSN Online: 2409-9368

An Automated Approach for 3D Objects Recognition Using Deep Convolution Neural Network in Business Applications

Amarah Abdullah[1]

Abstract
Object recognition, empowered by advanced pattern recognition techniques, has revolutionized various business sectors, transforming traditional processes with enhanced efficiency, accuracy, and security. In retail, object recognition facilitates dynamic inventory management and smart shelf optimization, while manufacturing benefits from automated quality control and streamlined assembly line processes. Robust surveillance and access control systems are fortified by sophisticated object recognition employing pattern recognition models, bolstering security measures. Healthcare integrates pattern recognition into medical imaging and patient monitoring systems, enabling accurate diagnoses and improved safety protocols. Object recognition refines marketing and advertising strategies, allowing businesses to analyze customer behavior and discern patterns in interactions and preferences using pattern recognition. The versatility of these technologies extends to real estate and finance, where object recognition is employed for property assessment, fraud detection, and document verification. The symbiotic relationship between object recognition and pattern recognition is emphasized, highlighting their combined potential to revolutionize traditional business processes. As businesses increasingly embrace these technologies, the seamless integration of object recognition with advanced pattern recognition algorithms is anticipated to usher in a new era marked by heightened efficiency, accuracy, and innovation across the business landscape.
Keywords: Computer Vision, Convolution Neural Network, Key Points; Object Classification, Object Recognition

1. Introduction
Object recognition is a fundamental task in the field of computer vision. Object recognition is the method of identifying different objects (like faces, vehicles or buildings) in digital images or videos. Latest research revolves around the development of learning algorithms for feature extraction and identification of entities and object classes [1] . Classifying objects into an apparent class comes under object recognition while localization of a precise object of interest in digital images or videos represents object detection [2]. All objects have specific features that characterize them to be a member of a particular class. These features also differentiate an object from other object classes. One can find many application areas of an object recognition; image recovery[3]; image based public safety and security [4]; crime investigations and forensics [5][6][7]; medical image analysis [8]; automated vehicle parking systems [9]; autonomous robots [10] and industrial inspection [11][12].
Several industries are currently working on projects of object recognition to address its needs and challenge [13]. Few of many research projects at France Telecom include face recognition [14][15][16] and video indexing [17]. Intel is conducting research on content-based image retrieval (CBIR) and object-based image retrieval [18][19]. Over the years significant progress is made by General Electric in the fields of video surveillance [20][21][22] and broadcast video [23][24][13]. Toyota developed techniques for 3D reconstruction of object shapes and separation of objects [25].
Many approaches for object recognition have been implemented over the decades. These approaches can be categorized as object models based [26], appearance based [27], feature based [28], genetic algorithm [29], deep learning based [30][31] and others [32]. Open challenges which were faced by various researchers in object recognition are: variation in scale, viewpoint, illumination and imaging conditions. Deformation of non-rigid objects, background clutter and large-scale image retrieval also pose computational complexities in object recognition [34].
Convolutional Neural Networks (CNN) is exploited in [35] for the recognition of various object classes. The proposed approach analyzes digital images for object recognition. Deep Learning or Deep Neural Networks alludes to Artificial Neural Networks (ANN) with multiple layers [35]. CNN has different layers; including convolutional, non-linearity, pooling and fully connected layers. In recent years, deep learning is viewed as one of the most integral learning approaches as it can deal with gigantic amount of data and still produce remarkable results [36] [37][38].
In this paper, CNN has been effectively used for the recognition of multi-class objects. A simple but effective CNN architecture containing multiple layers, is presented here for recognition of objects.
This paper proceeds as: details of the dataset and proposed approach is presented in Section 2. The results are achieved using the proposed approach and a detailed discussion based on comparison with other approaches are reported in Section 3. Finally, the paper is concluded in Section 4.

2. Experimental Methods
In this paper deep learning is used for object recognition from visual content. This section presents basic information and design of Deep Network for classification of objects. An image-based object recognition approach mainly consists of preprocessing, feature extraction and classification phases.
Preprocessing transforms the data into the format that can be easily and effectively processed for feature extraction. After preprocessing, features were extracted from the processed images. Afterwards the classifier was trained on this dataset and classification was performed. To begin with, we presented the details of the image dataset; utilized in our trials and its preprocessing. After that,

[1] School of Commerce and Accountancy, The University of Management and Technology, amarah.abdullah@umt.edu.pk

we presented the architecture of the proposed CNN classifier for business applications. Because recognition have major impact in business domain like in automation, healthcare, security, marketing, customer behavior analysis and etc.

Figure 1: Methodology Flow Chart

2.1. Dataset

Coil-100 dataset is employed to evaluate the effectiveness of the proposed approach. The dataset contains colored images of 100 different objects (boxes, bottles, cups, miniature cars, etc.) with uniform background rotated 360° about the vertical axis with an increment of 5° (72 images per object) [39]. Sample images of various objects of COIL-100 are shown in Fig. 2.

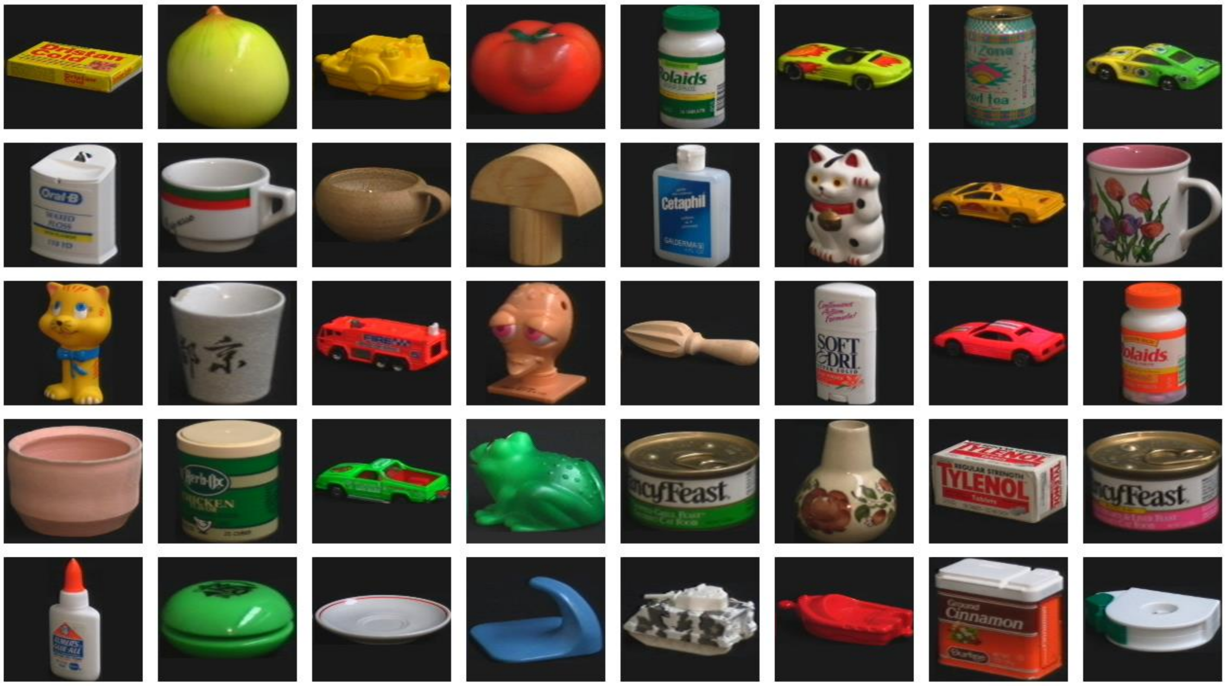

Figure 2: Images of the first 40 objects from Coil-100

In Coil-100, each image is in true colors. The dataset of Coil-100 comprises of 7,095 colored images of different objects which is taken from different geometrical angles spread across 100 labels. Coil-100 can be downloaded from Git Hub and Kaggle [40]. The complete setup for the collection of images is discussed in detail in [41]. These images have large variation in quality, quantity and scaled on a fixed size. Prior to the image analysis all the images are down sampled to the fixed resolution of 128x128.

Table 1: Number of images for each class. Each class of object labeled with digits (e.g. Obj1, Obj2, Obj3, etc.)

Classes	Number of Training Images	Number of Testing Images	Total Number of Images
Obj_1 to Obj_73	60	12	72
Obj_74	15	5	20
Obj_75	14	5	19
Obj_76 to Obj_100	60	12	72
Total Classes=100	*Total training Images=5,909*	*Total testing images=1,186*	*Total number of images=7095*

2.2. Architecture of the proposed CNN classifier

CNNs achieve remarkable performance in visual problems and prove to be competitive to the human performance in similar scenarios. This gives us an inspiration to evaluate the capability of CNNs for view invariant object recognition from images. The design of proposed CNN classifier differs from the conventional CNN as how the convolutional layers are arranged and how the nets are trained. The arrangement of convolutional layers of the proposed network is illustrated in Fig. 3.

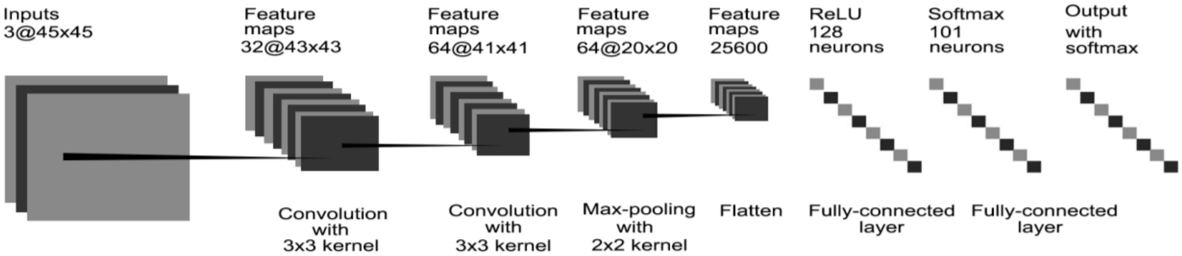

Figure 3: Defines the structure of CNN using different layers with dimensions

2.2.1. Convolution Layer 1

The principal convolution layer utilizes a convolution kernel. Convolution kernel of size 3x3 is convolved over the image with a depth size of 3 pixels. The kernel slides over the width and the height of the input to separate 32 (high dimension) features outline. The output of the first convolutional layer turns into the input for the next layer.

2.2.2. Convolution Layer 2

The second convolution layer takes the output of the previous convolution layer with steps utilizes the respective field of the size 3x3 with the 64 filters. Output of this layer becomes the input of the next pooling layer.

2.2.3. Max Pooling Layer 3

Pooling layer operates on each feature map independently. The 3rd Max pooling layer uses the maximum value from each of a cluster of neurons of the prior two layers with the size of 2x2.

2.2.4. Flatten Layer 4

After Max pooling, a flatten layer is used to merge or flatten layers to reduce the file size. Flattening combines all the layers into a single background layer. It multiplies the 64 filters with image 20x20 dimensions and provides the output shape 25600.

2.2.5. Dense Layer and Output Layer 5, 6

The 5th layer is a Dense layer. It uses ReLU activation function with 128 neurons. 6th layer is the output layer that has 101 Softmax neurons that corresponds to the 100 categories of the objects. At the end, Softmax classifier is applied which classifies all the100 categories of objects.

2.3 CNN training

In this section, learning of the parameter space of the proposed CNN classifier is explained. A dropout rate of 0.25 is utilized to abstain from overfitting and to calculate training and testing precision. The network is trained with the stochastic gradient descent (SGD) [42] algorithm with a categorical cross-entropy cost function. The dropout that eliminates a portion of neurons from the network is used to reduce the possible overfitting problem.

Training a CNN requires to choose an arrangement of hyper parameters, among which the learning rate (η) is the most basic one. It influences the training performance. A fixed learning rate over the whole training procedure appears problematic, since it dynamically assesses the training conducted. Here, an adaptive learning rate is utilized, that is an exponential capacity of cost $\eta = \eta_0 \times \exp(C)$ where η_0 is set to l.0 through preliminaries and errors and C is the training loss. Such learning rate updating schedule is directly related to the training performance. Initially, the training loss is huge, resulting in a high learning rate that accelerates the training procedure. Eventually, the learning rate is reduced with the loss that abstains from exceeding the best outcome.

3. Results and discussions

Results, their analysis, training and testing approaches used in this research are discussed in detail in this section. Effects of changes in various system designs and image representations and what does the system gain from the information are also discussed here. The convolutional neural system design (CNN) with results is presented in the Fig. 3.

3.1. Trial setup

Each analysis is done utilizing the online platform Kaggle.com and keras online [43]. Analyses are conducted on a PC with additional setup which is illustrated in above (Fig.3).

3.2. Architecture Training and Testing Approach

The proposed framework is implemented in Python. It is assessed using around 7095 object images. There is no explicit benchmark information for the object types. The dataset utilized in these analysis has been gathered with various variations (rotation, scale, illumination, and perspective and image blur) for each object. The dataset is partitioned into training and testing sets. Likewise, these images are utilized after resizing to 100x100 pixels. Object recognition in this work has been accomplished by training the CNN by allocating proper feature matrix and defining the target class based on the input class of data set. The dataset contains 72 views with various geometrical angles for each object of all classes. Training dataset contains 5909 images and 1186 images have been utilized for testing.

3.3. The structure of CNN

CNN utilizes the structure which includes Convolutional layer, Max pooling layer, ReLU layer, Flatten layer and Dense layer. In ordinary CNN engineering, each convolutional layer is trailed by a Rectified Linear Unit (ReLU) layer, at that point a Pooling layer then at least one convolutional layer lastly one Dense layer.
The input comprises of standard RGB pictures of size 100 x 100 pixels. A customized neural network transforms the input to a one dimensional array that makes the classifier less sensitive to positional changes.

Table 2: The CNN which we use in this research has the structure defined with layers, activation functions and dimensions

Layers	Activation Functions	Feature maps	Parameters
Convolutional Layer	ReLU	32@43x43	896
Convolutional Layer	ReLU	64@41x41	18496
Max Pooling Layer	--	64@20x20	--
Flatten Layer	--	25600	--
Dense Layer	ReLU	128 neurons	3276928
Dense Layer	Softmax	101 neurons	13029

3.4. Evaluation of the proposed model with multiple configurations

The image dataset is classified using various configurations. After numerous experimentations the best accuracy of 96.71% is achieved, where Softmax function is activated with dense layer.

Table 3: Evaluation of proposed model for object classification with multiple configurations

MODEL 1				
Layers	Name	Activation Function	Filters/Dropout	Accuracy%
01	Dense layer	ReLU	128 filters/0.1	88.11%
02	Dense layer	ReLU	64 filters/0.1	
03	Fully connected layer	Softmax	101	
MODEL 2				
01	Dense layer	ReLU	256 filters/ 0.05	85.91%
02	Dense layer	ReLU	128 filters/ 0.05	
03	Dense layer	ReLU	128 filters/ 0.05	
04	Dense layer	ReLU	128 filters/ 0.05	
05	Dense layer	ReLU	128 filters/ 0.05	
06	Fully connected layer	Softmax	101	

4. Conclusion

This research has demonstrated the deep learning-based approach for automatic object recognition that has given promising results in business domain. The proposed classification framework contains preprocessing, training and testing. The proposed model has achieved the classification accuracy of 96.71%. The proposed model can benefit from the inclusion of features in collaboration with the deep learning-based features with the goal that higher accuracy can be achieved. Further researches may experiment with these practices, furthermore attempt to discover new designs that give intriguing results.

References

[1] T. Guo, J. Dong, and H. Li, "Simple Convolutional Neural Network on Image Classification," pp. 721–724, 2017.

[2] X. Zhou, K. Yu, T. Zhang, and T. S. Huang, "Image Classification Using Super-Vector Coding of Local Image Descriptors," pp. 141–142, 2010.

[3] A. Gholipour, N. Kehtarnavaz, S. Member, R. Briggs, and M. Devous, "Brain Functional Localization : A Survey of Image Registration Techniques," vol. 26, no. 4, pp. 427–451, 2007.

[4] J. Flusser and B. Zitova, "Image registration methods : a survey," vol. 21, pp. 977–1000, 2003, doi: 10.1016/S0262-8856(03)00137-9.

[5] S. Garfinkel *et al.*, "Bringing Science to Digital Forensics with Standardized Forensic Corpora By Bringing science to digital forensics with standardized forensic corpora," 2009, doi: 10.1016/j.diin.2009.06.016.

[6] A. Valjarevic, "A Comprehensive and Harmonised Digital Forensic Investigation Process Model A Comprehensive and Harmonised Digital Forensic Investigation Process Model."

[7] Y. Yannikos, L. Graner, M. Steinebach, and C. Winter, "Chapter 21 DATA CORPORA FOR DIGITAL FORENSICS EDUCATION AND," pp. 309–325.

[8] J. Ker and L. Wang, "Deep Learning Applications in Medical Image Analysis," pp. 9375–9389, 2018.

[9] M. Hirz and B. Walzel, "Sensor and object recognition technologies for self-driving cars," vol. 15, no. 4, pp. 501–508, 2018.

[10] F. Ingrand, M. Ghallab, F. Ingrand, M. Ghallab, and A. Artificial, “Deliberation for autonomous robots : A survey To cite this version : HAL Id : hal-01137921 Deliberation for Autonomous Robots :,” pp. 10–44, 2017.
[11] Y. Xiong, B. Wang, C. Chu, and R. Gadh, “Electric Vehicle Driver Clustering using Statistical Model and Machine Learning.”
[12] J. Weston, F. Ratle, H. Mobahi, and R. Collobert, “Deep Learning via Semi-supervised Embedding,” pp. 639–640, 2012.
[13] Y. Hirano, C. Garcia, R. Sukthankar, and A. Hoogs, “Industry and Object Recognition : Applications , Applied Research and Challenges,” *Towar. Categ. Object Recognition, Lect. Notes Comput. Sci.*, vol. 4170, pp. 49–64, 2006.
[14] R. J. Micheals, P. Grother, and P. J. Phillips, “The NIST HumanID Evaluation Framework,” pp. 403–411, 2003.
[15] C. Liu, S. Member, H. Sako, and S. Member, “Effects of Classifier Structures and Training Regimes on Integrated Segmentation and Recognition of Handwritten Numeral Strings,” vol. 26, no. 11, pp. 1395–1407, 2004.
[16] P. Viola and M. Jones, “Rapid Object Detection using a Boosted Cascade of Simple Features,” 2001.
[17] A. F. Smeaton, “TRECVID : Benchmarking the Effectiveness of Information Retrieval Tasks on Digital Video 2 . The Shot Boundary Detection Task Work on algorithms for automatically recognizing and characterizing shot boundaries.”
[18] D. Hoiem, R. Sukthankar, H. Schneiderman, and L. Huston, “Object-Based Image Retrieval Using the Statistical Structure of Images,” 2004.
[19] S. C. S. Suite, “Efficient Near-duplicate Detection and Sub-image Retrieval Categories and Subject Descriptors.”
[20] A. G. A. Perera, C. Srinivas, A. Hoogs, and G. Brooksby, “Multi-Object Tracking Through Simultaneous Long Occlusions and Split-Merge Conditions,” 2006.
[21] R. Mazzon, S. F. Tahir, and A. Cavallaro, “Person re-identification in crowd,” *Pattern Recognit. Lett.*, vol. 33, no. 14, pp. 1828–1837, 2012, doi: 10.1016/j.patrec.2012.02.014.
[22] M. T. Chan, A. Hoogs, R. Bhotika, A. Perera, J. Schmiederer, and G. Doretto, “Joint Recognition of Complex Events and Track Matching,” 2006.
[23] A. Hoogs, J. Rittscher, G. Stein, and J. Schmiederer, “Video Content Annotation Using Visual Analysis and a Large Semantic Knowledgebase ∗,” 2003.
[24] D. J. Walter and R. S. Avila, “Model-Based Detection of Lung Nodules in Computed Tomography Exams 1 Thoracic Computer-Aided Diagnosis,” pp. 258–266, 2004, doi: 10.1016/S1076-6332(03)00729-3.
[25] J. Shi and J. Malik, “Normalized Cuts and Image Segmentation,” vol. 22, no. 8, pp. 888–905, 2000.
[26] M. J. Fenske, E. Aminoff, N. Gronau, and M. Bar, *Chapter 1 Top-down facilitation of visual object recognition: object-based and context-based contributions*, vol. 155 B. 2006.
[27] Z. Yu, “Intrinsic shape signatures: A shape descriptor for 3D object recognition,” in *2009 IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops 2009*, 2009, pp. 689–696, doi: 10.1109/ICCVW.2009.5457637.
[28] S. A. A. Shah, M. Bennamoun, F. Boussaid, and L. While, “Evolutionary Feature Learning for 3-D Object Recognition,” *IEEE Access*, vol. 6, pp. 2434–2444, 2017, doi: 10.1109/ACCESS.2017.2783331.
[29] G. Bebis, S. Louis, Y. Varol, and A. Yfantis, “Genetic object recognition using combinations of views,” *IEEE Trans. Evol. Comput.*, vol. 6, no. 2, pp. 132–146, 2002, doi: 10.1109/4235.996013.
[30] S. Hayat, S. Kun, Z. Tengtao, Y. Yu, T. Tu, and Y. Du, “A Deep Learning Framework Using Convolutional Neural Network for Multi-Class Object Recognition,” in *2018 3rd IEEE International Conference on Image, Vision and Computing, ICIVC 2018*, 2018, pp. 194–198, doi: 10.1109/ICIVC.2018.8492777.
[31] M. Rashid *et al.*, “A sustainable deep learning framework for object recognition using multi-layers deep features fusion and selection,” *Sustain.*, vol. 12, no. 12, 2020, doi: 10.3390/su12125037.
[32] M. Riesenhuber and T. Poggio, “Models of object recognition,” *Nat. Neurosci.*, vol. 3, no. 11s, pp. 1199–1204, 2000, doi: 10.1038/81479.
[34] L. Liu *et al.*, “Deep Learning for Generic Object Detection : A Survey,” *Int. J. Comput. Vis.*, 2019, doi: 10.1007/s11263-019-01247-4.
[35] S. Albawi and T. A. Mohammed, “Understanding of a Convolutional Neural Network,” 2018, no. April, doi: 10.1109/ICEngTechnol.2017.8308186.
[36] W. Huang and R. Kovacevic, “Development of a real-time laser-based machine vision system to monitor and control welding processes,” 2012, doi: 10.1007/s00170-012-3902-0.
[37] R. Collobert, J. Weston, and M. Karlen, “Natural Language Processing (almost) from Scratch,” vol. 1, pp. 1–34, 2000.
[38] Y. Goldberg, “A Primer on Neural Network Models for Natural Language Processing,” pp. 1–76, 2015.
[39] D. Koubaroulis and J. Kittler, “Evaluating Colour-Based Object Recognition Algorithms Using the SOIL-47 Database,” in *Database*, 2002, no. January, pp. 23–25.
[40] “GitHub - fqez_dataset-coil.” .
[41] S. A. Nene, S. K. Nayar, and H. Murase, “Columbia Object Image Library (COIL-100).”
[42] M. Convolutional, N. Network, and D. Optimizer, “Modified Convolutional Neural Network Based on Dropout and the Stochastic Gradient Descent Optimizer,” 2018, doi: 10.3390/a11030028.
[43] “Kaggle_ Your Home for Data Science.” .