iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** VII **Month of publication:** July 2023

DOI: https://doi.org/10.22214/ijraset.2023.54704

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue VII Jul 2023- Available at www.ijraset.com

A Review on Classification of Images with Convolutional Neural Networks

Mandeep Kaur[1], Pradip Kumar Yadava[2]
[1, 2]*Dept of Computer Science & Engineering, CT University, Ludhiana, Punjab India*

Abstract: ***Deep learning has recently been applied to scene labelling, object tracking, pose estimation, text detection and recognition, visual saliency detection, and image categorization. Deep learning typically uses models like Auto Encoder, Sparse Coding, Restricted Boltzmann Machine, Deep Belief Networks, and Convolutional Neural Networks. Convolutional neural networks have exhibited good performance in picture categorization when compared to other types of models. A straightforward Convolutional neural network for image categorization was built in this paper. The image classification was finished by this straightforward Convolutional neural network. On the foundation of the Convolutional neural network, we also examined several learning rate setting techniques and different optimisation algorithms for determining the ideal parameters that have the greatest influence on image categorization.***
Keywords: Convolutional neural network, Deep Learning, Transfer Learning, ImageNet, Image classification; learning rate, parametric solution.

I. INTRODUCTION

Computer vision image classification is crucial for our daily lives, employment, and education. A process that involves image preprocessing, picture segmentation, key feature extraction, and matching identification is used to categories images. We can now capture image data more quickly than ever before and use it in a variety of applications, such as face recognition, traffic identification, security, and medical equipment, thanks to the most up-to-date image classification techniques.

With the advent of deep learning, feature extraction and classifier have been combined into a learning framework in order to solve the drawbacks of the conventional approach of feature selection. In order to capture the more immaterial semantics of the information, high-level qualities are expected to be used to identify many layers of representation. Convolutional architectures are an essential part of deep learning when it comes to picture classification.

The convolutional neural network is based on the anatomy of the mammalian visual system. In 1962, Hubel and Wiesel proposed a model of the visual system based on the visual cortex of cats.

The concept of a receptive field has been introduced for the first time. Fukushima first proposed the hierarchical framework that Neo cognition would use to examine images in 1980. Neo cognition used the local connection between neurons to establish network translation invariance.

Many deep learning architectures are available. The classifier system model described in this paper was developed using convolutional neural networks, the most efficient and practical deep neural network for this type of data. Therefore, by applying these learning representations to tasks that require less training data, CNNs that have been trained on enormous datasets of images for recognition tasks may be utilised to their advantage.

Since 2006, numerous methods have been developed to circumvent the difficulties associated with training deep neural networks. Krizhevsky demonstrates a significant improvement over prior methods for the photo classification task and proposes a typical CNN architecture Alexnet. Given its success, numerous measures to improve Alexnet's performance have been suggested. It is advised to use VGGNet, GoogleNet, and ZFNet.

A. Hierarchical Feature Extraction

CNNs are excellent at learning hierarchical image representations. They are made up of several layers, such as pooling and convolutional layers, which gradually extract features at various levels of abstraction. By using a hierarchical approach, CNNs can more accurately recognise complex patterns and structures in images.

©IJRASET: All Rights are Reserved

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue VII Jul 2023- Available at www.ijraset.com

II. KEY REASONS FOR THE SIGNIFICANCE OF CNN

A. Translation Invariance

Because CNNs are created to be translation invariant, they can detect patterns anywhere in an image. Convolutional layers, which apply filters to an image and detect features regardless of their position, are used to do this. CNNs are more reliable and accurate in real-world applications thanks to this characteristic, which allows them to classify images regardless of their orientation or position.

B. Data Efficiency

Comparatively to conventional machine learning algorithms, CNNs require fewer training instances. Due to their capacity to identify pertinent aspects and extrapolate to previously unobserved data, they can learn from a limited number of samples. Because of this characteristic, CNNs are excellent in situations where there is a lack of huge volumes of labelled data.

C. Transfer Learning

CNNs are capable of transfer learning, which entails that they can take the knowledge they gain from one task and apply it to a related one. Pre-trained models, which are trained on big datasets and can be tailored for particular picture classification tasks, are used to do this.Transfer learning can significantly increase classification performance while requiring less training data..

D. Scalability

Due to their scalability, CNNs can be applied to a variety of challenging picture categorization problems. They can add or remove layers, vary the number of filters in each layer, and alter the size of the filters used in convolutional layers, which allows for this scalability. Because of its adaptability, CNNs can be used for a variety of tasks, from straightforward picture classification to more difficult ones like object recognition and segmentation.

To achieve this goal, the objectives below have been specified:

1) Create an architecture with a flexible gene encoding approach that does not limit the maximum length of the CNN building blocks. The improved architecture is anticipated to help CNNs attain good performance in handling various jobs at hand using this gene encoding approach.
2) Look into the connection weight encoding technique, which may effectively encode many connection weights. The suggested GA is anticipated to successfully optimize the weight connection initialization problem in CNNs using this encoding strategy. Create accompanying crossover, mutation, and selection (including environmental selection) operators that can handle the intended gene encoding techniques of both architectures and connection weights.
3) Suggest a reliable fitness metric for the people representing the various CNNs that doesn't call for a lot of processing power.
4) Examine whether the novel strategy significantly outperforms the already used techniques in terms of classification accuracy and weights.

III. METHODOLOGY OF EVALUATION

The main objective of our research is to understand the performance of networks with both static and real-time video streams. The first step in the following procedure is transfer learning on networks utilizing image datasets. Transfer learning on networks with picture datasets is the next step.

The next phase is then tested after that. The rate of the same item's prediction on still images and real-time video streams is then looked at.

The different accuracy rates are noted, noted down, and displayed in the tables given in the following sections. To determine whether there were any variations in prediction accuracy between the CNNs utilized in the study was the third important aspect for evaluating performance.

It must be made clear that videos are used as testing datasets rather than training datasets. We are looking for the best image classifier in which the object serves as the main criterion for scene category categorization.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

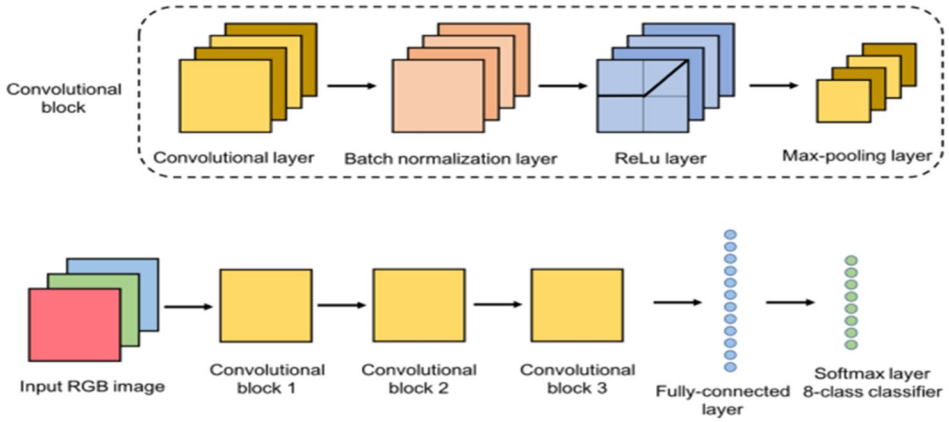


Fig. 1: CNN layers categorization

Different layers of the convolutional neural network used are:

1) *Input Layer:* Every CNN that is employed starts with a "input layer" that resizes images before passing them to later layers for feature extraction.
2) *Convolution Layer:* The subsequent layers, known as "Convolution layers," serve as filters for images, thereby extracting their features. They are also employed in testing to determine the match feature points.
3) *Pooling Layer:* The 'pooling layer' is then given the extracted feature sets. The most crucial information in large photographs is preserved while they are shrunk down in this layer. It maintains the highest value from The best fits of each feature within the window are retained for each window.
4) *Rectified Linear Unit Layer:* Every negative integer in the pooling layer is replaced with 0 by the subsequent "Rectified Linear Unit" or ReLU layer. This prevents learnt values from becoming stuck close to 0 or exploding towards infinity, helping the CNN maintain its mathematical stability.
5) *Fully Connected Layer:* The last layer, which is made up of all connected layers, converts the high-level filtered images into categories and labels.
6) *Basic CNN Components:* The three main types of convolutional neural network layers are the convolutional layer, the pooling layer, and the fully-connected layer.

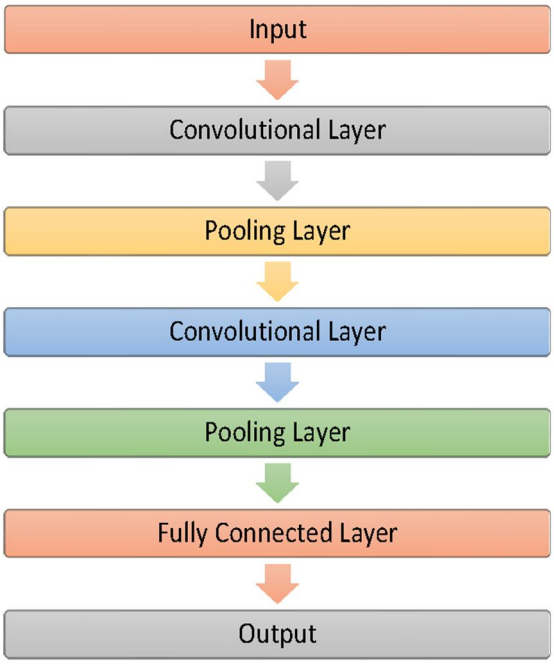


Fig. 2: Basic CNN components

The steps of proposed method are as follows:

a) *Creating Training and Testing Dataset:* The training and validation data sets are separated into two categories, i.e. the super classes images used for training are scaled to [224,244] pixels for AlexNet and [227,227] pixels for GoogleLeNet and ResNet50.
b) *Modifying CNNs Network:* Fully linked layer, softmax layer, and classification output layer should take the place of the final three network layers. Set the size of the final fully connected layer to correspond to the number of classes in the training set. To train a network more quickly, increase the fully linked layer's learning rate parameters.
c) *Train the Network:* Set the training parameters, such as learning rate, mini-batch size, and validation data, in accordance with the system's GPU specifications. Utilize the training data to train the network.
d) *Test the Accuracy of the Network:* Utilize the tuned network to classify the validation images, then determine the classification accuracy. Similarly, real-time video feeds are used to assess the fine-tuning network's accuracy.

IV.MODELS

Many clever pre-trained CNN exist; these CNN are capable of transmitting learning. As a result, all that is required for its input layer are the training and testing datasets. The basic building blocks and techniques used in the networks' design differ. The GoogleNet's Inception Modules carry out convolutions of varied sizes and combine the filters for the layer that follows. Instead of using filter concatenation, AlexNet uses the output of the layer before it as its input. Both networks employ the Caffe Deep Learning framework's implementation and have undergone independent testing. However, neural network training becomes more difficult as we move further away, and accuracy peaks before falling. Both of these problems are attempted to be addressed via residual learning. Many layers of a deep convolutional neural network are frequently overlaid and trained for the specified goal. The network picks up a variety of low-, mid-, and high-level features towards the end of its layers. In residual learning, a network tries to pick up some residual information rather than specific traits. Residual is only the feature that was taken in from the input of the layer and subtracted. ResNet does this using a shortcut connection, which connects some (n+x) of the input from one layer to another layer directly. Three current neural networks, the AlexNets, Google Nets, and ResNet50, are compared. The transfer learning concepts are then implemented, followed by the training of current networks and the construction of new networks for additional comparison. The new models perform very differently from the old networks despite having the same number of layers as the original models. The various accuracy rates that were determined using the same photographs are presented in the tables in the next section.

V. CONVOLUTIONAL NEURAL NETWORKS (CNNS) FOR IMAGE CLASSIFICATION OFADVANCEMENTS

1) *Attention Mechanisms:* Attention methods have recently been added to CNN topologies, allowing the network to concentrate on particular areas or elements of an image that are most important for categorization. You can investigate various attentional processes, such self- or spatial attention, and how these affect the accuracy and interpretability of CNN models.
2) *Transformer-Based Architectures:* The Transformer model has been modified for picture classification challenges as a result of its success in natural language processing. Convolutional layers are replaced with self-attention mechanisms in transformer-based architectures like Vision Transformers (ViTs), allowing the model to recognise global dependencies in images. You can compare these architectures' performance and scalability to those of conventional CNNs.
3) *Meta-Learning and Few-Shot Learning:* The goal of meta-learning techniques is to improve CNNs' capacity to learn from a small number of labelled examples by utilising prior information gained from comparable tasks or datasets. With a little amount of training data, few-shot learning methods like meta-learning, metric learning, or generative modelling help CNNs generalise to new classes. For image classification, you can research developments in meta-learning and few-shot learning and evaluate how they perform against conventional CNN models.
4) *AutoML and Neural Architecture Search:* Neural Architecture Search (NAS), a technique used in automated machine learning (AutoML), has drawn attention for its ability to automatically identify the best CNN structures for picture categorization. NAS techniques use evolutionary algorithms, gradient-based optimisation, or reinforcement learning to find architectures with better performance. You can talk about the advancements in NAS and AutoML and assess how well they work for finding better CNN designs.
5) *Explain ability and Interpretability:* Understanding how CNNs make decisions becomes increasingly important as these models become more complicated. The interpretability and explainability of CNNs for image categorization should be improved in future developments. Investigate techniques like attention visualisation, saliency maps, or class activation maps that reveal the areas of a picture that influence categorization decisions the most.

6) *Robustness and Adversarial Defense:* CNNs are vulnerable to adversarial assaults, in which minor changes to the input images can result in incorrect categorization. Future developments in CNN architectures should incorporate defenses against adversarial assaults to overcome the robustness and security issues. You can talk about various defense techniques and evaluate how well they work to make CNN models more robust.

VI. FUTURE DISCUSSIONS

In this section, we will discuss the fitness evaluation of the proposed EvoCNN method, weights-related parameters, and structures encoding strategies in more detail. Reviewing the experimental results may provide insight into the potential applications of the suggested EvoCNN technique. Crossover operators function as the local search whereas mutation operators serve as the exploratory search, or the global search. Only correctly enhancing both local and global searches could greatly improve performance because they should complement one another. The approaches for CNN weight optimisation that are frequently used are based on the gradient data. It is generally known that gradient-based optimizers are sensitive to the starting positions of the parameters that need to be improved. Without a good starting point, gradient-based approaches are prone to getting stuck in local minima. The large number of features makes it appear hard to find a better starting point for the connection weights using Gas. As we've shown, a significant number of parameters are incapable of being effectively optimised or encoded into the chromosomes. The suggested EvoCNN technique simply encodes the means and common derivations of the weights in each layer using an indirect encoding strategy. Methods now in use to find CNN architectures often take into account both an individual's fitness and the ultimate categorization accuracy. To get a final classification accuracy, the training approach typically needs many more epochs, which takes a lot of time.

A. Summary

An overview of convolutional neural networks (CNNs) for image categorization is given in this article. It opens by emphasising the limitations of conventional feature selection methods and the significance of picture classification in diverse fields. These constraints are addressed by the introduction of deep learning, in particular CNNs.

The article highlights that by using convolutional layers and layer pooling to extract features at various levels of abstraction, CNNs excel in learning hierarchical representations of images. Because of their ability to recognise patterns wherever they appear in a picture, CNNs have translation invariance. Due to their capacity to identify pertinent features and generalise to new data, they are also data-efficient, requiring fewer training samples.

The capacity of CNNs to use pre-trained models developed on substantial datasets and fine-tune them for particular image classification tasks is emphasised as a critical competency. This increases classification performance and lowers the amount of training data needed.

It is underlined as a crucial competency of CNNs to be able to leverage pre-trained models created on sizable datasets and optimise them for specific image classification applications. As a result, classification performance is improved and less training data is required.The process for assessing CNN performance is described in the text, and it entails training networks on static and live video streams, carrying out transfer learning, and testing accuracy. Input layers, convolution layers, pooling layers, rectified linear unit (ReLU) layers, and fully connected layers are among the various CNN layer types that are mentioned.

AlexNet, GoogLeNet, and ResNet50 are just a few of the models and architectures that are covered. The paper contrasts their performance and highlights developments in CNNs, such as neural architecture search, meta-learning, transformer-based designs, and attention methods. Additionally, it highlights the necessity of explainability, interpretability, and resilience against adversarial attacks in CNNs.

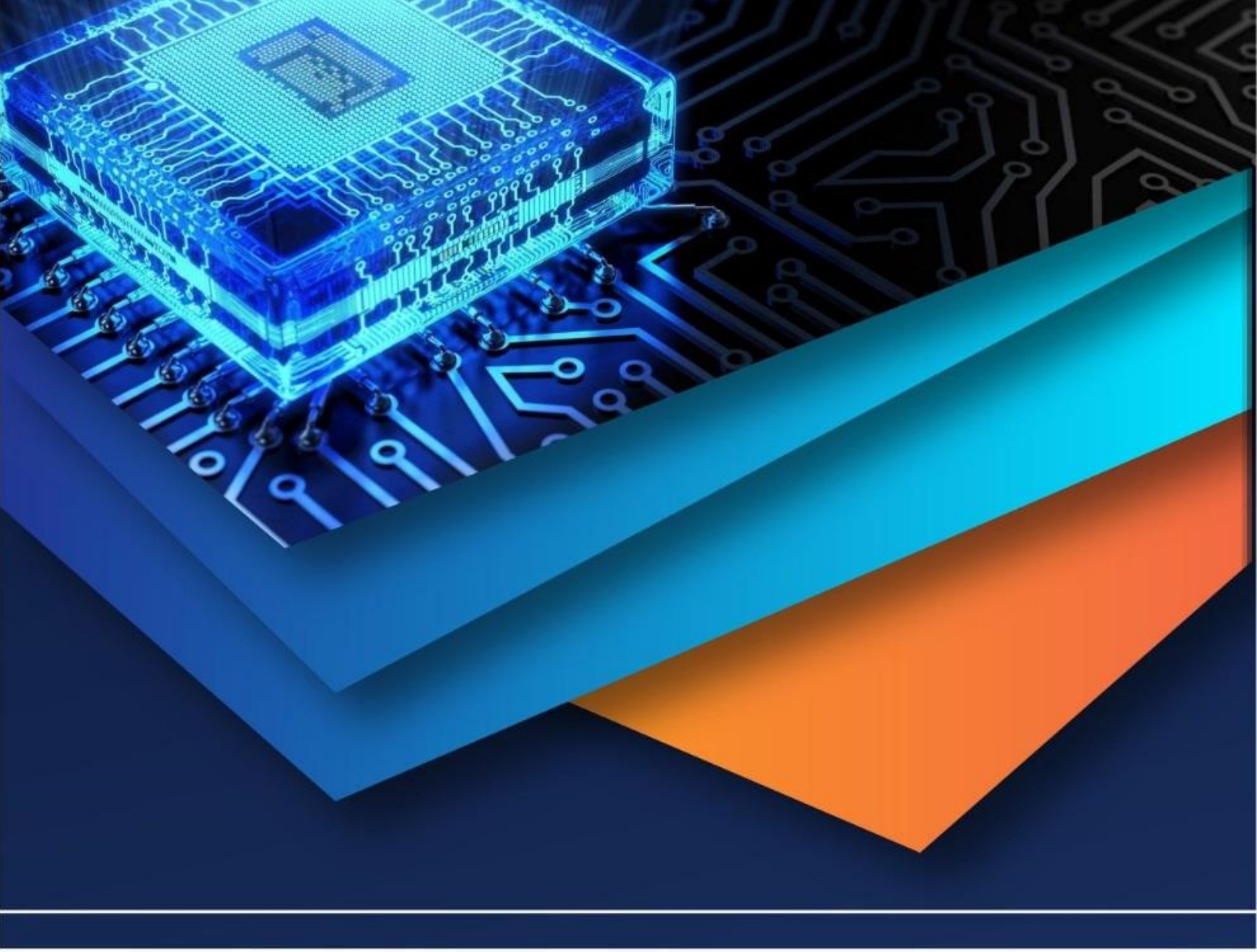
VII. CONCLUSION

In this study, a unique evolutionary technique is being developed to automatically evolve the topologies and weights of CNNs for image classification tasks. This objective has been achieved by proposing a new representation for weight initialization strategy, a new encoding scheme for variable-length chromosomes, a new genetic operator for chromosomes with different lengths, a slack binary tournament selection for choosing promising individuals, and a successful fitness evaluation method to speed up evolution. Given the restricted amount of training time, understanding deep learning is crucial. Evolutionary algorithms will be used in future research to handle the classification feature extraction difficulty and decrease the number of parameters needed for this operation, which will enhance our system.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

REFERENCES

[1] Browne JD, Boland DM, Baum JT, Ikemiya K, Harris Q, Phillips M, Neufeld EV, Gomez D, Goldman P, Dolezal BA. Lifestyle modification using a wearable biometric ring and guided feedback improve sleep and exercise behaviors: a 12-month randomized, placebo-controlled study. Front Physiol. 2021; 12:2094. https://doi.org/10.3389/fphys.2021.777874.
[2] Celik O, Yildiz BO. Obesity and physical exercise. Minerva Endocrinol (Torino). 2021;46(2):131–44. https://doi.org/10.23736/S2724-6507.20.03361-.
[3] Ege T, Yanai K. Image-based food calorie estimation using knowledge on food categories, ingredients and cooking directions. In Proceedings of the on Thematic Workshops of ACM Multimedia. 2017;367–75. https://doi.org/10.1145/3126686.3126742.
[4] Redmon J, and Angelova A, "Real-time grasp detection using convolutional neural networks", IEEEInternational Conference on Robotics and Automation, pp. 1316–1322, 2015.
[5] JasminPrafulBharadiyahttps://journaljerr.com/index.php/JE RR/article/view/858
[6] Hang Chang, Cheng Zhong, Ju Han, Jian-Hua Mao, "Unsupervised Transfer Learning via Multi-Scale Convolutional Sparse Coding for Biomedical Application." IEEE Transactions on Pattern Analysis and Machine Intelligence, 23 janvier 2017.
[7] JasminPrafulBharadiyahttps://journalajorr.com/index.php/AJ ORR/article/view/164
[8] Zhou, X., Yu, K., Zhang, T., & Huang, T. "Image classi¿cation using super-vector coding of localimage descriptors." In ECCV,2010.
[9] van de Sande, K. E. A., Gevers, T., and Snoek, C. G. M, "Evaluating color descriptors for object and scene recognition", IEEE Transactions on Pattern Analysisand Machine Intelligence." 1582– 1596. 2010.
[10] Howard, A. , "Some improvements on deep convolutional neural network based image classi¿cation." ICLR, 2014.
[11] Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L., "ImageNet: A large-scale hierarchical image database." In CVPR, 2009.
[12] Ahonen, T., Hadid, A., and Pietikinen, "M. Face description with local binary patterns: Application to face recognition." Pattern Analysis and Machine Intelligence, 2037–2041. 2016.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)