IT in Industry, Vol. 9, No.2, 2021

PublishedOnline12April 2021

SMART ELECTRICITY CONSERVATION SYSTEM USING EFFICIENTNET

Ushasukhanya S
Research scholar
ushasukhanya@gmail.com
Annamalai University

Jothilakshmi S
Associate Professor
jothi.sekar@gmail.com
Annamalai University

Ragupathy R
Associate Professor
cse_ragu@yahoo.com
Annamalai University

Abstract

Conservation of electric resource has been one of the important challenges over the decades. Worldwide, many nations are struggling to conserve and to bridge the gap between the demand and production of the resource. Though many measures like several Government acts, replacing existing products with energy conserving products and many solar based systems are being invented and used in practise, the demand and the need to preserve the resource still persists. Hence, this paper focuses on a novel technique to conserve the electric resource using a deep learning technique. The system uses Convolutional Neural Networks to identify and localize humans in the CCTV footages using EfficientNet, a deep transfer learning model. The classifier processes and yields its output to an embedded Arduino microcontroller, after detecting the presence/absence of human. The microcontroller enables/disables the electric power supply in the area of human's presence/absence, based on the classifier's output respectively. The system achieves an accuracy percentage of 84.2% in detecting humans in the footages with the subsequent enabling/disabling of electric power resource to conserve electricity.

Keywords: **Electricity Conservation, Human Detection, Convolutional Neural Networks, EfficientNet, Arduino.**

1 Introduction

Electricity conservation has been one of the biggest challenges in recent days. In spite of many measures taken to overcome this challenge, the demand still persists. Many sensor based products have been invented to reduce the consumption of electricity. Though the consumption of electricity is reduced by using these products, all these add an additional cost to the clients thereby making it an option for the customers to use. Hence conservation of electricity even after inventing these new products remains as a question mark. To overcome the issue, a new software based system is developed in this work which enables electricity only when required. The system detects human beings in CCTV footages using a deep learning technique. The footages are converted to frames and these frames are processed by Deep Convolutional Neural Network (CNN) to detect humans. The system uses the EfficientNetB0 model (transfer learning) for detecting humans in the frames and once humans are detected in

ISSN(Online):22031731
ISSN (Print): 2204-0595

the frames, the output is fed to an embedded Arduino microcontroller, which controls the connected power supply of that indoor environment.

The controller enables the power supply of the room (lights and fans) only if a human is detected by the classifier and disables it when the human is undetected by the classifier. The switching time between the enabling and disabling of the power is 8 seconds. This conserves about one third of the total electricity being consumed by the individual.

2 Literature Survey

Object detection has been an interesting topic of research since 1990. Numerous techniques have been invented to detect, identify, localize and recognize objects in images. Object detection has been used in various smart applications, content retrieval, security purposes, surveillance, and automated vehicle systems and so on. Conventional methods involve several independent steps for pre-processing, background subtraction [1], segmentation [2], feature extraction and object detection [3]. Recent techniques like machine learning and deep learning uses most of these steps together in a single step, which speeds up the process and also avoid manual intervention to a greater extent [4] [5]. Machine learning has been used in the field of computer vision for several years. As the recent real-time applications require enormous data to be processed for better prediction accuracy, human intervention in labelling the images for these cases becomes impossible. Hence, deep learning comes into place which can handle huge data without the necessity of human intervention [6].

Deep learning systems for object detection have been an increasing trend in computer vision. Various deep learning models are currently in practice which excels in predictions depends upon the data being used. For instance, Convolutional Neural Network (CNN) excels in handling image data [7] while Recurrent Neural Network (RNN) excels in handling text or speech data [8] [9]. As most of the data falls under these two categories, these are the two vital Neural Network models that classify most of the real-time objectives. Variants of CNN like Region Convolutional Neural Network (RCNN), Fast-Regional Convolutional Neural Network (fast R-CNN), Faster-Regional Convolutional Neural Network (Faster R-CNN), Single Shot Detector (SSD) and so on and so forth. These networks are implemented either by developing a model from the scratch or using transfer learning models. Transfer learning is another concept being widely used in this domain. Models like AlexNet, ResNet-50/101, VGG-16/19, CapsNet, Detnet and EfficientNet etc., are already created models which are trained on certain standard datasets. These models could be used for any real-time purposes, and the prediction accuracy is expected to be high, as the models are already trained and familiarized with certain features. Models are also developed specifically for a purpose and trained on the required dataset which performs better than transfer learning in many cases.

A system by Chenchen Zhu et al [10] was designed to detect faces of human using R-CNN. A model was developed considering multi-scale features. Synchronization of location specific features and semantic features of low-level

ISSN(Online):22031731
ISSN (Print): 2204-0595

and high-level layers are done to represent the face of human's present. The model outperformed the baseline methods while using standard face detection Data Set and Benchmark (FDDB) on validation process. Various researchers are working in optimizing the neural networks and one such system to solve these optimization problems, is developed by Sun et al [11]. This system called particle swarm optimization with binary encoding (BQPSO), developed by adjusting the evolution to discrete binary space, is created in the following equations. The movement of the particles M depends on the following equations:

$$mbest = \left(\frac{1}{M}\sum_{i=1}^{M} P_{i1}(t), \frac{1}{M}\sum_{i=1}^{M} P_{i2}(t), \dots, \frac{1}{M}\sum_{i=1}^{M} P_{iD}(t)\right) \tag{1}$$

$$a_{id}(t) = \rho \times p_{id} + (1 - \rho) \times p \tag{2}$$

$$x_{id}(t+1) = v_{id} \pm \text{б}|mbest_d - x_{id}(t)| * \ln\left(\frac{1}{u}\right) \tag{3}$$

where P is the particle, ρ and u are random numbers distributed uniformly on [0,1], б is a positive constant called Contraction Expansion Coefficient.

A system was developed by Hai huang [12] for marine organism detection and recognition. As the training samples were less, data augmentation techniques were used to enlarge the dataset. Four-step alternate training method with SGD was used for training the RPN. Unlike most papers that use affine transformation for data augmentation, perspective transformation is used for data augmentation in this paper. Experiment results confirm that the proposed three data augmentation methods increase the robustness of Faster R-CNN to marine turbulence variations, shooting angle variations, and uneven illumination variations, respectively. Transfer learning models were developed since 2012 by Krizhevsky et al for AlexNet in [13] is the first best transfer learning model that yielded best results in classifying images. This is taken up by GoogleNet, which achieved 74.8% accuracy in 2015 [14] subsequently by SENet in [15] with 82.7%. Many authors prefer to develop a model from the scratch to avoid the network's pre-learnt information unlike transfer learning.

A system was developed by MingXing Tan in [16] for classification of CIFAR datasets using EfficientNet. The authors have discovered that a better accuracy rate could be achieved by stabilizing the depth, width and resolution of the model. The system uses a compound co-efficient to evenly scale the three specified parameters of the model. The system initially uses a baseline model and then the model is scaled up to a cluster of models for enhancing the accuracy. The system was able to achieve an efficiency of 84.3% on ImageNet with increased processing speed. A similar system was developed for skin lesion classification in [17] on dermoscopic images. The authors developed a hybrid model by combining EfficientNets, ResNet and SENet which are selected by a search technique. Out of

ISSN(Online):22031731

ISSN (Print): 2204-0595

the various models used EfficientNet has outperformed other transfer learning models in classifying skin lesions.

A similar system for fruit recognition system was developed by Linh Duong et al in [18] for automatic harvesting using EfficientNet and MixNet. The system initially classifies fruits from other objects in the images using machine learning techniques. The system uses 48,905 images for training followed by 16,421 images for testing. The system proves that the transfer learning concept excels other methods in this case by yielding an average prediction accuracy of 99.8% in fruit detection. An automated system for the recent pandemic, Covid-19 was developed by Goncalo in [19] using EfficientNet. Initially, a binary classification was done by the images of Covid affected and un-affected patients with an average prediction accuracy of 99.6%. This is followed by multi-class classification in Covid-19, pneumonia and un-affected patients with a prediction rate of 97%. As the available dataset is not robust, the authors have claimed that the system could face a pitfall on scaling. Secondly the system outperforms only in the advanced stage of the disease and an early diagnosis system is suggested for future work

3 Proposed System

CNN plays a major role in detecting objects from videos/images. This system is developed based on a Deep CNN system to identify humans in CCTV footages. The videos of the footages are converted to frames at the rate of 6 frames per second (fps). These frames are processed by a CNN model in which the convolutional layer, which is the vital part of the model for detection is defined as

$$A_i = F_i(X_i) \tag{4}$$

where A is the output of the layer, X is the input image which is concatenated with the filter weights. For the real-time applications, certain sophisticated ConvNets are used in which there are multiple stages in the convolutional layers and in each stage except the initial stage, the structure of the ConvNet remains the same. This ConvNet is defined as

$$C \sum_{i=1 \; to \; n} F_i^L (X_{(b_i, d_i, c_i)}) \tag{5}$$

where b, d and c denotes the size of the image in each layer. The baseline CNN structure is inefficient for many real-time applications to process huge high dimensional data and hence certain modifications in the structure of the networks is being done in recent days to improve the model’s efficiency. One such modification is the scaling of the model either by their width or depth or by increasing the resolution of the image. Unfortunately, scaling of only one of these parameters could be done at a time and in some rare cases, two parameters are scaled but arbitrarily which does not guarantee enhanced overall efficiency of the model. Hence rather than scaling each and every parameter individually or arbitrarily, EfficientNet uses the concept of scaling all the parameters using a coefficient

ISSN(Online):22031731
ISSN (Print): 2204-0595

uniformly so that the difference factor between the parameters is maintained. This avoids certain important drawbacks of Deep Neural Networks like vanishing gradients and saturation of accuracies which occurs in the case of depth and width scaling respectively. Hence an EfficientNet-BO model is used in this system which scales the network in two phases as follows:

A compound coefficient **ϕ** is considered to manage the number of available resources followed by α, β, and γ parameters to allocate these resources to the network's depth, width and resolution respectively. the scaling now is done in the following way

$$N_d = \alpha^{\phi} \quad (6)$$

$$N_w = \beta^{\phi} \quad (7)$$

$$N_r = \gamma^{\phi} \quad (8)$$

Such that the dot product of the three parameters i.e.,

$$\propto \beta^2\gamma^2 \approx 2 \quad (9)$$

where α, β and γ is considered to be greater than 1

The structure of the model with 5.9M parameters is given in Figure 1.

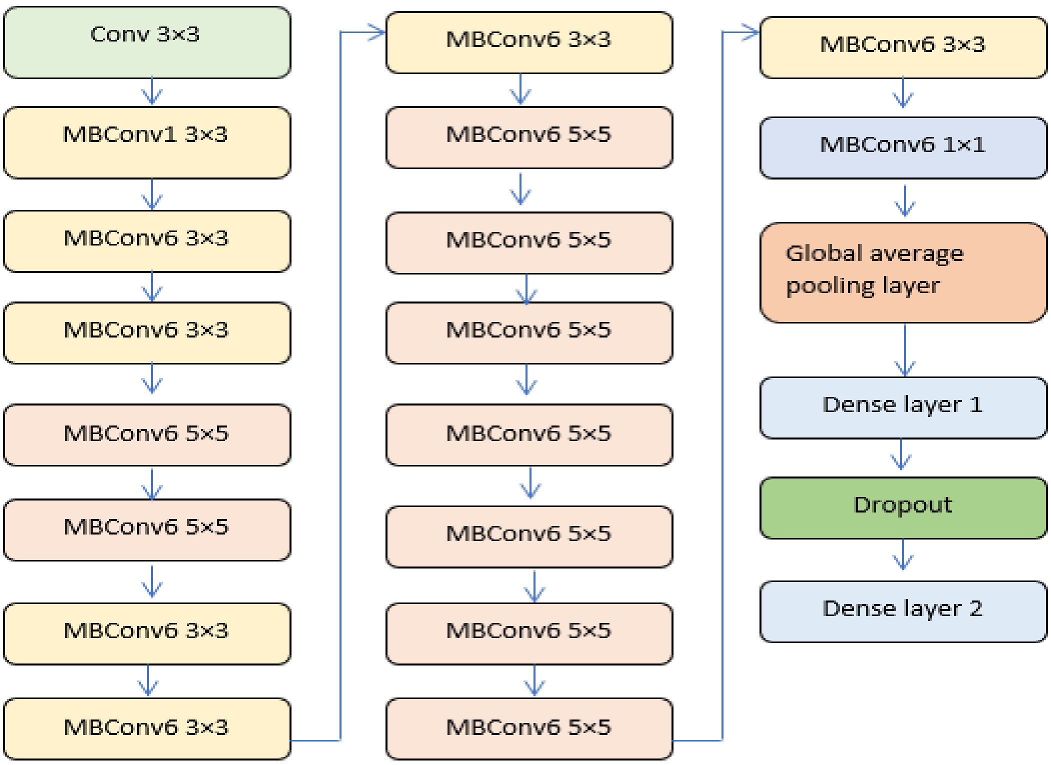

Figure 1 Architecture of the proposed system

ISSN(Online):22031731
ISSN (Print): 2204-0595

The global average pooling layer and a dropout of 20% is added at the end of this model to reduce the overfitting issues. As suggested by the authors, the optimal values for α, β and γ are 1.2, 1.1 and 1.15 respectively. These are the values used to scale the network. Two CCTV footages captured in different indoor environments, each in duration of 120 minutes and 140 minutes are converted to frames at the rate of 6 fps. Perspective transformation [20] of data augmentation is done to increase the dataset size. This technique mimics the images captured by the cameras. The resultant images are the augmented images at different angles and from various viewpoints, which is practically impossible for the camera to capture. The size of Datasets including footage frames and augmented images is 54000 frames for Dataset 1 and 58,200 for Dataset 2 which is then divided into training and testing set in a ratio of 80:20 respectively. An instance of the datasets with a positive image (Image with human) and a negative image (Image without human) is given in Figure 2.

Figure 2 Datasets Used for the Proposed System

The model is initially trained with 43,200 images in 48 epochs, followed by testing the model with the remaining 10,800 images. The model is finally optimized with Adam algorithm [21], since it is proved to be better than the RMSProp, to improve the overall efficiency rate of the network. The algorithm for the proposed system is given below

ISSN(Online):22031731

ISSN (Print): 2204-0595

Step 1	*Input*	*Dataset containing CCTV Footage images and augmented images*
Step 2	*Environment*	*Colab*
Step 3	*Configuration*	*Importing images with configuring the dataset for training, testing and validation*
Step 4	*Model used*	*EfficientNetB0 model (transfer learning concept is used)*
Step 5	*Modification*	*Addition of Global average pooling layer and dropout of 20%*
Step 6	*Optimization*	*Adam with a learning rate of 0.001*
Step 7	*Training &Testing*	*The model is trained with 43,200 images for 48 epochs, followed by testing with 10,800 images.*
Step 8	*Validation*	*Generate performance scores values*

Figure 3 Algorithm for the proposed system

The model yields an average accuracy of 91.2% in the validation set, which is a subset of the training data and 81.6% of accuracy on an average in the test data. The generalization issue is handled by the dropout regularization technique, which reduces the difference between the accuracies of the training and the testing set. The average prediction accuracy of the test data is 89.2% after the usage of dropout technique. The individual prediction accuracies of both the datasets for their validation and test datasets and the precision-recall trend is given in Figure 4 (a) and (b).

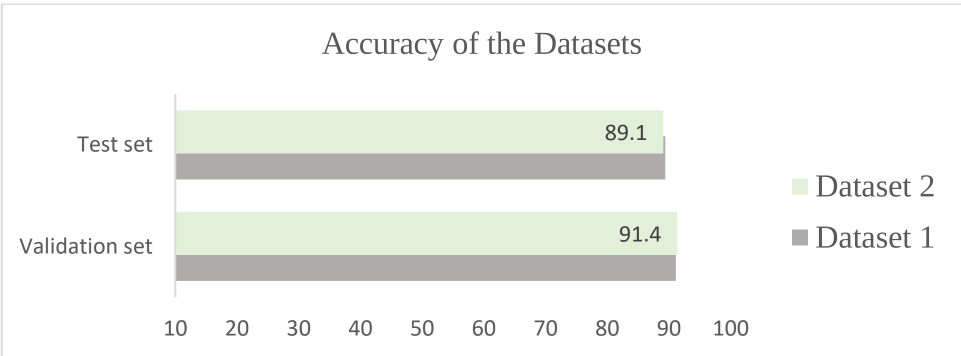

Figure 4(a) Accuracy of the Datasets

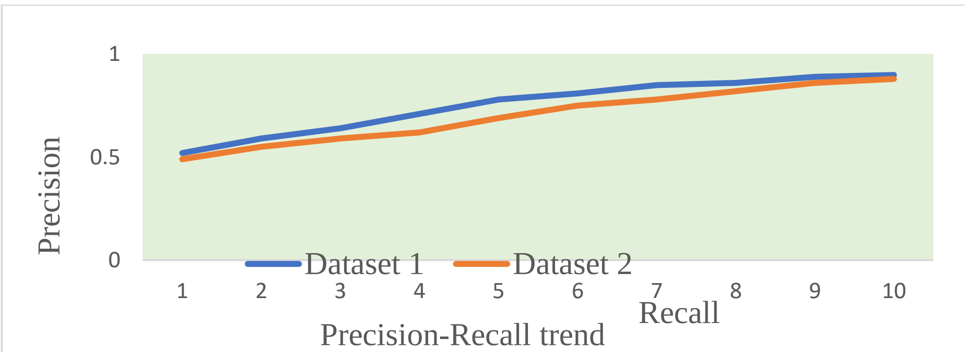

Figure 4(b) Precision-Recall curve

ISSN(Online):22031731
ISSN (Print): 2204-0595

Figure 5 depicts the detection of human in the frames by EfficientnetB0 model based on which the power supply will be managed by the microcontroller in the next session.

Figure 5 Detection of human in the frames of CCTV footages

4 Management of Electricity

The electric supply of the indoor environment where the resource has to be conserved is connected to this system. The output of the classifier is supplied to an embedded Arduino Microcontroller. Microcontroller has been used in many smart applications in recent days for power consumption [22] [23]. Hence, this system also uses a microcontroller which manages the power supply by enabling/disabling the electric power supply based on the human's presence or absence in the frame as detected by the classifier. The power consumption is therefore achieved by disabling the electric resource when it is actually not required without the usage of any additional equipment.

5 Conclusion and Future work

Thus, a smart electricity conservation system based on Deep CNN and Arduino Microcontroller is developed. The system detects humans in the CCTV footages by EfficientnetB0 model which is an upgraded version of CNN model. The classifier's output is given to the microcontroller which controls the power supply connected to it by enabling the power on human's presence and by disabling it on human's absence. This avoids the unnecessary wastage of electricity which saves up to one third of the total power consumption without the necessity of any additional equipment. The system has been designed only for indoor environments and hence electricity consumption in outdoor environment should be taken into consideration. Secondly, occlusion is not handled in this case which can also be considered for future scope.

References

1. Shaik Mohamed, Shahrizat & Tahir, Noorita & Adnan, Rahmi. (2010). Background modelling and background subtraction performance for object detection. pp. 1 - 6.
2. Cheng, Ming-Ming & Mitra, N & Huang, X & Torr, Philip & Hu, Shi-Min. (2011). Salient Object Detection and Segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence.
3. B.Wu and R. Nevatia, (2007). Simultaneous Object Detection and Segmentation by Boosting Local Shape Feature based Classifier,

ISSN(Online):22031731
ISSN (Print): 2204-0595

2007 IEEE Conference on Computer Vision and Pattern Recognition, pp. 1-8.

4. A. Raghunandan, Mohana, P. Raghav and H. V. R. Aradhya, (2018). Object Detection Algorithms for Video Surveillance Applications," 2018 International Conference on Communication and Signal Processing (ICCSP). pp. 0563-0568.
5. Gao, Chenqiang and Li, Pei & Zhang, Yajun & Liu, Jiang and Wang, Lan. (2016). People counting based on head detection combining Adaboost and CNN in crowded surveillance environment. Neurocomputing.
6. Najafabadi, Maryam & Villanustre, Flavio & Khoshgoftaar, Taghi & Seliya, Naeem & Wald, Randall & Muharemagic, Edin. (2015). Deep learning applications and challenges in big data analytics. Journal of Big Data.
7. Sharma, Neha & Jain, Vibhor & Mishra, Anju. (2018). An Analysis Of Convolutional Neural Networks For Image Classification. Procedia Computer Science. vol.132. pp. 377-384.
8. Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In NIPS, pp. 1097–1105, 2012.
9. Nowak, Jakub & Taspinar, Ahmet & Scherer, Rafal. (2017). LSTM Recurrent Neural Networks for Short Text and Sentiment Classification. pp. 553-562.
10. Zhu, Chenchen& Zheng, Yutong &Luu, Khoa &Savvides, Marios. (2017). CMS-RCNN: Contextual multi-scale region-based CNN for unconstrained face detection.
11. Tao Sun and Ming-hai Xu, (2017), A Swarm Optimization Genetic Algorithm Based on Quantum-Behaved Particle Swarm Optimization, Computational Intelligence and Neuroscience.
12. Zhou, Hao & Huang, Hai & Yang, Xu & Zhang, Lu & Qi, Lu. (2017). Faster R-CNN for Marine Organism Detection and Recognition Using Data Augmentation, pp. 56-62.
13. Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In NIPS, pp. 1097–1105.
14. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A. (2015). Going deeper with convolutions. CVPR, pp. 1–9.
15. Hu, J., Shen, L., and Sun, G. (2018). Squeeze-and-excitation networks. CVPR.
16. Tan, Mingxing & Le, Quoc. (2019). EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks.
17. Gessert, Nils & Nielsen, Maximilian & Shaikh, Mohsin & Werner, René & Schlaefer, Alexander. (2019). Skin Lesion Classification Using Ensembles of Multi-Resolution EfficientNets with Meta Data.

ISSN(Online):22031731
ISSN (Print): 2204-0595

18. Linh T. Duong, Phuong T. Nguyen, Claudio Di Sipio, Davide Di Ruscio. (2020). Automated fruit recognition using EfficientNet and MixNet, Computers and Electronics in Agriculture, vol. 171.
19. Gonçalo Marques, Deevyankar Agarwal, Isabel de la Torre Díez. (2020). Automated medical diagnosis of COVID-19 through EfficientNet convolutional neural network, Applied Soft Computing, vol. 96.
20. K. Wang, B. Fang, J. Qian, S. Yang, X. Zhou and J. Zhou, (2020). Perspective Transformation Data Augmentation for Object Detection," in IEEE Access, vol. 8, pp. 4935-4943.
21. Jais, Imran & Ismail, Amelia & Qamrun Nisa, Syed. (2019). Adam Optimization Algorithm for Wide and Deep Neural Network. Knowledge Engineering and Data Science. pp. 41-46.
22. X. Ren, H. Guo, G. He, X. Xu, C. Di and S. Li. (2016). Convolutional Neural Network Based on Principal Component Analysis Initialization for Image Classification. IEEE International Conference on Data Science in Cyberspace (DSC), pp. 329-334
23. Hu Q, Li F. (2013). Hardware design of smart home energy management system with dynamic price response. Smart Grid, IEEE Transactions. vol.4 (4). pp. 1878-87.

ISSN(Online):22031731
ISSN (Print): 2204-0595