Retraction

Retraction: Object Recognition in Images with Low-Resolution using Convolutional Neural Network (J. Phys.: Conf. Ser. 1916 012049)

Published 23 February 2022

This article (and all articles in the proceedings volume relating to the same conference) has been retracted by IOP Publishing following an extensive investigation in line with the COPE guidelines. This investigation has uncovered evidence of systematic manipulation of the publication process and considerable citation manipulation.

IOP Publishing respectfully requests that readers consider all work within this volume potentially unreliable, as the volume has not been through a credible peer review process.

IOP Publishing regrets that our usual quality checks did not identify these issues before publication, and have since put additional measures in place to try to prevent these issues from reoccurring. IOP Publishing wishes to credit anonymous whistleblowers and the Problematic Paper Screener [1] for bringing some of the above issues to our attention, prompting us to investigate further.

[1] Cabanac G, Labbé C and Magazinov A 2021 arXiv:2107.06751v1

Retraction published: 23 February 2022
Object Recognition in Images with Low-Resolution using Convolutional Neural Network

T Raghunathan ¹, G Sharmili ¹, T H Shreeja ¹, S Saru ¹
¹Department of Computer Science and Engineering, Sri Krishna College of Technology, Coimbatore, Tamil Nadu, India
Email: ²1tucs219@skct.edu.in, ³7tucs223@skct.edu.in, ⁴1tucs207@skct.edu.in, ⁵raghunathan.t@skct.edu.in

Abstract. Object recognition is a technology in computer vision that finds objects in an image or video series and identifies them. Phenomenal results have been recorded in object recognition studies using deep neural networks. But it has generally been deduced that sufficient image resolution and object size are obtainable, which cannot be assured in practical uses. Recognition of objects in lower resolution images is difficult. To overcome the stated problem, a Convolutional Neural Network (CNN) model for identifying objects in lower resolution images is proposed in this paper. In object recognition datasets, this approach outperforms the high recognition accuracy. In convolutional neural network models, both convolution and max-pooling layers are typically stacked. In the proposed approach, the pooling layer was substituted with a convolutional layer with an expanded phase without loss of precision in image recognition. The All Convolutional Neural Network with trained weights for recognizing lower resolution images is deployed. Through the obtained results, it is verified that the proposed model has high efficiency and accuracy.

1. Introduction
In machine learning, object recognition is a well-known challenge because it operates on deep learning strategy [1]. In many areas of computer vision, object recognition is applied, including image retrieval, surveillance systems, smartphone face recognition, security, driverless cars, machine inspection, and automated vehicle parking systems. Following the success in the paper, “ImageNet classification with deep convolutional neural networks” [2], the efficiency of deep neural networks in object recognition has improved intensely and rapidly. “In the previous studies, networks were built of minimal layers which created problems in the training process” [3]. “In the 2016 Large-Scale Visual Recognition Challenge dataset, the squeeze and excitation network reported a 2.25 percent error when categorizing images from 1000 classes and secured first place in the competition” [4]. “This is remarkable because the human error rate is 5.1 percent”. But this network has a poor emphasis on lower resolution object recognition than that on higher resolution imaging, despite these outcomes. Convolutional Neural Networks are meant to imitate the human visual system model thus, in the computer vision field, Convolutional Neural Networks have accomplished great achievements. To extract characteristics, the convolutional layers are used, and the pooling layers are used to increase the generalizability of the functionality. A common principle is applied to all the modern convolutional neural networks (CNNs) for their construction. “Generally,
convolution and pooling layers are used in the construction of the convolutional neural network, accompanied by a limited number of fully linked layers” [5]. The rectified linear activation function is used within each of these layers. During training using dropout, the networks are usually normalized to be broad and formalized. Components of convolutional neural networks are necessary for achieving good performance on object recognition datasets. Studies show that a similar network composed exclusively of convolutional layers, with a limited dimensionality of 2 strides, reaches the ultimate output without the use of complex functions, normalization, or pooling layers. Stochastic gradient descent algorithm was used to build this architecture. By performing ablation on the CIFAR-10 (Canadian Institute For Advanced Research) dataset, the impact of switching from a more common architecture to a simplified Convolutional neural network (CNN) model is studied. This CIFAR-10 dataset contains 10 classes of images. Each class has a separate category of images. Airplane, bird, automobile, deer, cat, frog, dog, ship, horse, truck are the ten categories of images. The proposed method classifies these images despite the low resolution. As suggested in “Very Deep Convolutional networks for large-scale Image Recognition” [6], the proposed method results confirm the efficacy of using limited convolutional layers and generate substantial new questions about the need to pool in Convolutional Neural Networks.

2. Related works
Studies using deep learning techniques on very lower resolution object recognition are in a developing state. “Studying Very Lower-Resolution Recognition Using Deep Networks” [7] used deep learning to identify very lower resolution objects and showed the possibility of using neural networks to recognize very lower resolution objects, as well as numbers, facial expressions, and symbols. Although with this approach, the efficiency has not been completely verified. “To leverage the visual similarity within a group of visually related object categories, a Novel Joint Dictionary Learning (JDL) algorithm was suggested” [8-13]. In this JDL algorithm, the recognition of objects is a problem whenever the total object classes are larger. suggested to train a typical network for object recognition by using lower resolution images. This trained network showed slightly higher efficiency on lower resolution images than on high resolution images. But its accuracy on higher resolution images was significantly worse. This made the necessity to build an efficient technique with images of different resolutions to achieve tremendous output. “Deep convolutional network attempted to solve performance on lower resolution images using image super-resolution (SR) techniques”. But this method failed to derive valuable information about objects from the images. “GenLR-Net and the resolution aware convolutional neural network have also attempted to address the issue of object recognition with very lower resolution”. “Deep Back-Projection Networks for Super-Resolution proposed Super-Resolution methods to retrieve visual data from input images, but this approach has low performance for lower resolution images [14]”. The proposed system achieves high performance on lower resolution images. The Convolutional Neural Network has about 1.3 million training parameters [15].

3. Proposed system
The model that is used in this experiment, varies in many aspects from standard CNNs. The pooling layers that are virtually existing in all traditional layers are eliminated and substituted with normal convolutional layers of two strides. [16-20] To understand this proposed process, the convolution and pooling layers of CNNs must be known. A feature-wise convolution performance is exhibited by the pooling layer. Furthermore, one may query why it is appropriate to incorporate such specific layers into the network. Although it is not easy to provide a full answer to this question, it is believed that there are typically three possible reasons why pooling will assist in CNNs.

1) The pooling layer allows the Convolutional neural network representation to be more flexible.
2) The decrease in spatial dimensionality carried out by pooling. So, the higher portions of the input can also be masked.

3) Optimization is done easier due to the feature-wise nature of the pooling layer. To achieve good efficiency with Convolutional neural networks, the reduction of dimensions carried out by pooling layers is important.

Now, this pooling layer can be eliminated from the network without the loss in efficiency by two means:

1. Each pooling layer can be removed and the convolutional layer stride can be increased accordingly.

2. The pooling layer can be substituted with a standard convolution with a stride greater than one.

The first way seems to have the drawback that it substantially decreases the correlation of the convolutional layer. It is similar to a pooling mechanism where only the higher level function result is taken into account and may lead to less detailed identification. The second method does not have this drawback, because all the existing convolutional layers remain constant, also it contributes to an enlargement in the total network parameters. [21] It is emphasized that this substitution can often be perceived as learning rather than fixing the mechanism of pooling layers, which was taken into consideration in previous studies. Although current studies involving these experiments on the substitution of convolution layers instead of pooling layers are not known, the principle of eliminating pooling layers is not totally uncommon. Experiments that used only convolution layers in an architecture corresponding to conventional CNNs have already been proposed. The second change in the proposed network model compared to the traditional CNNs is that the small convolutional layers are used which can significantly minimize the number of parameters in a network and thereby achieves optimization. This system additionally standardizes the network as it contains more parameters. [22-26] Therefore, the proposed system is a network consisting of only convolutional layers, activation and dropout layers to generate results over the entire image. The outline of the proposed architecture is represented in Figure 1.
Figure 1. Overall structure of the proposed system

4. Experimentation and results
The CIFAR-10 (Canadian Institute For Advanced Research) dataset is used to conduct the experiment to assess the efficiency of the proposed system. This CIFAR-10 dataset contains 60,000 color images, with 6000 images in each class in 10 classes. There are 50,000 training images and 10,000 test images in this dataset. Five training batches and one test batch constitute the dataset. The test batch consists of accurately 1000 random sample images within each class. And the remaining images are in the training batches in random order. The training batches comprise precisely 5000 images within each class. In the CIFAR-10 dataset, the resolution of the actual images is 32 x 32 pixels. With an average computational expense of 10 hours on a modern GPU, a broad model can be trained on this dataset. The proposed CNN model is built using simple 3 x 3 convolution layers. This is done because if a convolutional layer replaces max-pooling, then 3 x 3 is the minimal filter size required to construct the model. This deployed model is called the “All Convolutional Neural Network” model.

It is stated that this model is similar to the very in-depth models used in the ImageNet competition 2014. The All CNN model description is tabulated in Table 1.

Table 1. All Convolutional neural network Model description.

| All Convolutional neural network Model |
|----------------------------------------|
| 3 x 3 convolution 96 Rectified Linear Unit |
| 3 x 3 convolution 96 Rectified Linear Unit |
Learning rate, momentum and weight decay are chosen as the hyperparameters for training the model. A parameter that regulates the model is the learning rate. And every time the model weights are modified, the model is adjusted using this learning rate in relation to the expected error. Momentum is used to improve both training speed and accuracy. Additional training parameters such as epochs and batch size are defined. Dropout is a method used in training, where randomly chosen neurons are automatically dropped out. Dropout is an input layer that is used in the fully connected layers. For all layers, a dropout of 0.5 is added. Ultimately, a higher version of the all convolutional neural network is trained on the CIFAR-10 dataset to evaluate the network model using only convolution layers.

After the verification of the proposed model, it is found that this simple network achieved better performance and accuracy compared to other methods. This model achieved an accuracy of 90.8%. The comparison of algorithms and their accuracy on the CIFAR-10 dataset is represented in Table 2.

Table 2. Comparison of algorithms and its accuracy.

| CIFAR-10 classification error | Method                | Error (%) | #params |
|------------------------------|-----------------------|-----------|---------|
|                              | Maxout                | 11.689%   | >6M     |
|                              | Network in Network    | 10.412%   | ≅ 1M    |
|                              | Deeply Supervised     | 9.7%      | ≅ 1M    |
5. Conclusion and Future work
Very simple architectures can perform very well with advanced methods of training. The all convolutional neural network using only convolution layers outperforms the significant performance on the CIFAR-10 dataset. Thus, a new approach for object recognition is introduced for images with a low resolution with the help of a convolutional neural network model. Although the method is very effortless, it generates clearer visualizations of Regions of descriptive pictures than the conventional methods. This neural network can be used as a future work for other visual recognition challenges such as facial expressions, numbers, and alphabets.

References
[1] J. Seo and H. Park, Object Recognition in Very Low-Resolution Images Using Deep Collaborative Learning, in IEEE Access, vol. 7, pp. 134071-134082, 2019, doi: 10.1109/ACCESS.2019.2941005.
[2] D. Cai, K. Chen, Y. Qian, and J. Kamarainen, Convolutional low-resolution fine-grained classification, Pattern Recognit. Lett., vol. 119, pp. 166-171, Mar. 2019.
[3] S. P. Mudunuri, S. Sanyal and S. Biswas, GenLR-Net: Deep Framework for Very Low-Resolution Face and Object Recognition with Generalization to Unseen Categories, in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshop, 2018, pp. 602-611.
[4] S. D., & H, A. (2019). AODV Route Discovery and Route Maintenance in MANETs. 2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS). doi:10.1109/icaccs.2019.8728456
[5] H. Anandakumar and K. Umamaheswari, An Efficient Optimized Handover in Cognitive Radio Networks using Cooperative Spectrum Sensing, Intelligent Automation & Soft Computing, pp. 1–8, Sep. 2017. doi:10.1080/10798587.2017.1364931
[6] C. Szegedy, V. Vanhoucke, S. Ioffe, A. Shlens, and Z. Wojna, Rethinking the inception architecture for computer vision, in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2016, pp. 2818-2826.
[7] C. Szegedy, S. Ioffe, and V. Vanhoucke, (2016) Inception-v4, inception-resnet and the impact of residual connections on learning, in Proc. Int. Conf. Learn. Represent.Workshop, 2016.
[8] Z. Wang, S. Chang, Y. Yang, D. Liu and T. S. Huang, Studying Very Low-Resolution Recognition Using Deep Networks, in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2016, pp. 4792-4800.
[9] X. Peng, J. Hoffman, S. X. Yu and K. Saenko, Fine-to-coarse knowledge transfer for low-res image classification, in Proc. IEEE Int. Conf. Image Process., 2016, pp. 3683-3687.
[10] Dong, C., Loy, C. C., He, K., & Tang, X. (2016). Image Super-Resolution Using Deep Convolutional Networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 38(2), 295–307. doi:10.1109/tpami.2015.2439281.
[11] K. Simonyan and A. Zisserman, Very deep convolutional networks for large-scale image recognition, in Proc. Int. Conf. Learn. Represent., 2015, pp. 1–14.
[12] C. Szegedy et al., Going deeper with convolutions, in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2015, pp. 1–9.
[13] S. Ioffe and C. Szegedy, Batch normalization: Accelerating deep network training by reducing internal covariate shift, in Proc. Int. Conf. Mach. Learn., Lille, France, 2015, pp. 448-456.
[14] K. He, X. Zhang, S. Ren, and J. Sun, Delving deep into rectifiers: Surpassing human-level performance on imagenet classification, in Proc. IEEE Int. Conf. Comput. Vis., 2015, pp. 1026-1034.

[15] Estrach, Joan B., Szlam, Arthur, and Lecun, Yann. Signal recovery from pooling representations. In ICML, 2014.

[16] Zeiler, Matthew D. and Fergus, Rob. Visualizing and understanding convolutional networks. In ECCV, 2014.

[17] N. Srivastava, G. E. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. J. Mach. Learn. Res., vol. 15, no. 1, pp. 1929-1958, Jan. 2014.

[18] Lin, Min, Chen, Qiang, and Yan, Shuicheng. Network in network. In ICLR: Conference Track, 2014.

[19] Lee, Chen-Yu, Xie, Saining, Gallagher, Patrick, Zhang, Zhengyou, and Tu, Zhuowen. Deeply supervised nets. In Deep Learning and Representation Learning Workshop, NIPS, 2014.

[20] Goodfellow, Ian J., Warde-Farley, David, Mirza, Mehdi, Courville, Aaron, and Bengio, Yoshua. Maxout networks. In ICML, 2013.

[21] A. Krizhevsky, I. Sutskever, and G. E. Hinton. ImageNet classification with deep convolutional neural networks, in Proc. Adv. Neural Inf. Process. Syst., 2012, pp. 1097–1105.

[22] Hinton, Geoffrey E., Srivastava, Nitish, Krizhevsky, Alex, Sutskever, Ilya, and Salakhutdinov, Ruslan R. Improving neural networks by preventing co-adaptation of feature detectors 2012.

[23] Ning Zhou, Yi Shen, Jinye Peng, & Jianping Fan. (2012). Learning inter-related visual dictionary for object recognition, 2012 IEEE Conference on Computer Vision and Pattern Recognition. doi:10.1109/cvpr.2012.6248091.

[24] V. Nair and G. E. Hinton, Rectified linear units improve restricted boltzmann machines, in Proc. Int. Conf. Mach. Learn., Haifa, Israel, 2010, pp. 807-814.

[25] Jarrett, Kevin, Kavukcuoglu, Koray, Ranzato, Marc Aurelio, and LeCun, Yann. What is the best multi-stage architecture for object recognition? In ICCV, 2009.

[26] Behnke, Sven. Hierarchical neural networks for image interpretation. PhD thesis, 2003.