Volume: 4, Issue: 11
Page: 75-81
2020

ISSN 2520-4750 (Online) & ISSN 2521-3040 (Print)

International Journal of Science and Business

Deep Learning Models Based on Image Classification: A Review

Kavi B. Obaid, Subhi R. M. Zeebaree & Omar M. Ahmed

Abstract:
With the development of the big data age, deep learning developed to become having a more complex network structure and more powerful feature learning and feature expression abilities than traditional machine learning methods. The model trained by the deep learning algorithm has made remarkable achievements in many large-scale identification tasks in the field of computer vision since its introduction. This paper first introduces the deep learning, and then the latest model that has been used for image classification by deep learning are reviewed. Finally, all used deep learning models in the literature have been compared to each other in terms of accuracy for the two most challenging datasets CIFAR-10 and CIFAR-100.

IJSB
Accepted 13 October 2020
Published 20 October 2020
DOI: 10.5281/zenodo.4108433

Keywords: Deep Learning, Image Classification, Machine Learning, Models.

About Author (s)

Kavi B. Obaid, Computer Science Department, College of Science, University of Zakho, Iraq (e-mail: kavi.obaid@uoz.edu.krd).
Subhi R. M. Zeebaree, Duhok Polytechnic University, Iraq (e-mail: subhi.rafeeq@dpu.edu.krd).
Omar M. Ahmed, (Corresponding author),Information Technology Department, Zakho Technical Institute, Duhok Polytechnic University, Iraq (e-mail: omar.alzakholi@uoz.edu.krd).

1. Introduction

Recently, the field of machine learning has seen unprecedented growth due to a new wealth of data, increases in computational power, new algorithms, and a plethora of exciting new applications (Zeebaree et al., 2017; Mohammed et al., 2020). As researchers tackle more ambitious problems, the models they use are also becoming more sophisticated. However, the growing complexity of machine learning models inevitably comes with the introduction of additional hyperparameters. These range from design decisions such as the shape of neural network architecture, to optimization parameters such as learning rates, to regularization hyperparameters such as weight decay. The proper setting of these hyperparameters is critical for performance on difficult problems (Zebari et al., 2020; Ahmed, et al., 2020). With the increasing popularity of smartphones, digital, and surveillant camera, the number of images taken each year has increased (Mustafa et al., 2020; Salih et al., 2020). However, to make use of the trillions of images generated every year the use of supervised classifiers has become necessary (Ahmed & Sallow, 2017; Hussein & Abduallah, 2018). By using supervised classification techniques found in the machine learning space, we can see an overall improvement in the overall effectiveness of the categorization and classification of images (Dino et al., 2020; Zeebaree et al., 2018). Tasks such as the categorization and classification of species would be made much easier due to the subtleties picked up by high precision systems. In this paper, we aim to study the performance of the various classifier on the CIFAR-10 and CIFAR-100 datasets.

Image classification is an area that has roots in lots of technologies (Alzakholi et al., 2020; Haji et al., 2020). For this reason, the applications of these classifiers are varied from large-scale generalizations such as understanding a context of an image to smaller more detailed results such as looking for cancer on a patient test results (Zebari et al., 2019; Zeebaree, et al., 2020). Furthermore, there are also variations in the maximum time allowed for an image to be classified. As such, the classifiers chosen are varied in both complexity and computation time (Zeebaree et al., 2020; Dino et al., 2020). The most efficient classifiers that perform well in classifying images are that by using deep learning algorithms (Ahmed, 2020).

Deep learning is the hottest trend in machine learning. Although the theoretical concepts behind deep learning are not new, it has enjoyed a surge of interest over the past decade due to many factors. One example is that deep learning approaches have significantly outperformed state-of-the-art approaches in many tasks across different fields such as image processing, computer vision, speech processing, natural language processing (NLP), etc (Zhao et al., 2019; Zeebaree, et al., 2020). deep learning has been used to achieve unparalleled results across a variety of benchmark machine learning problems, and have been applied successfully throughout science and industry for tasks such as large-scale image and video classification (Ahmad, et al., 2020). To deal with the difficulties of training deep networks, some researchers have focused on developing better optimizers. Well-designed initialization strategies, in particular the normalized variance-preserving initialization for certain activation functions, have been widely adopted for training moderately deep networks (Ahmed, 2020; Ahmed & Abduallah, 2017).

2. Literature Review

Goodfellow et al. (2013) proposed a new activation function called maxout that is particularly well suited for training with dropout, and for which they have proven a universal approximation theorem. They have shown empirical evidence that dropout attains a good approximation to model averaging in deep models. They have shown that maxout exploits their model averaging behavior because the approximation is more accurate for maxout units than for tanh units. They have demonstrated that optimization behaves very differently in the context of dropout than in the pure stochastic gradient descent (SGD) case. By designing the maxout gradient to avoid pitfalls such as failing to use many of a model's filters, they can train deeper networks than is possible using rectifier units. they have also shown that maxout propagates variations in the gradient due to different choices of dropout masks to the lowest layers of a network, ensuring that every parameter in the model can enjoy the full benefit of dropout and more faithfully emulate bagging training.

Stollenga et al. (2014) proposed the DasNet, which is a deep neural network with feedback connections that are learned through reinforcement learning to direct selective internal attention to certain features extracted from images. After a rapid first shot image classification through a standard stack of feedforward filters, the feedback can actively alter the importance of certain filters "in hindsight", correcting the initial guess via additional internal "thoughts". DasNet successfully learned to correct image misclassifications produced by a fully trained feedforward Maxout network. Its active, selective, internal spotlight of attention enabled state-of-the-art results. Lin et al. (2014) proposed a novel deep network called "Network In Network" (NIN) for classification tasks. This new structure consists of MLPConv layers that use multilayer perceptrons to convolve the input and a global average pooling layer as a replacement for the fully connected layers in conventional CNN. MLPConv layers model the local patches better, and global average pooling acts as a structural regularizer that prevents overfitting globally. With these two components of NIN, they demonstrated state-of-the-art performance on many datasets. Through visualization of the feature maps, they demonstrated that feature maps from the last MLPConv layer of NIN were confidence maps of the categories, and this motivates the possibility of performing object detection via NIN.

Snoek et al. (2015) proposed deep networks for global optimization, or DNGO, which enables efficient optimization of noisy, expensive black-box functions. While their model maintains desirable properties of the Gaussian Processes (GP) such as tractability and principled management of uncertainty, it greatly improves its scalability from cubic to linear as a function of the number of observations. They demonstrate that while this model allows efficient computation, its performance is nevertheless competitive with existing state-of-the-art approaches for Bayesian optimization. They demonstrate empirically that it is especially well suited to massively parallel hyperparameter optimization. M. Liang & Hu (2015) proposed a recurrent convolutional neural network (RCNN) for (static) object recognition. The basic idea was to add recurrent connections within every convolutional layer of the feedforward CNN. This structure enabled the units to be modulated by other units in the same layer, which enhanced the capability of CNN to capture statistical regularities in the context of the object. The recurrent connections increased the depth of the original CNN while kept the number of parameters constant by weight sharing between layers. Experimental results demonstrated the advantage of RCNN over CNN for object recognition. Over four benchmark datasets, with fewer parameters, RCNN outperformed the state-of-the-art models. Increasing the number of parameters led to even better performance.

Graham (2015) they have trained convolutional networks with fractional max-pooling on several popular datasets and found substantial improvements in performance. Overlapping FMP seems to be better than disjoint FMP. Pseudorandom pooling regions seem to do better than random pooling regions when training data augmentation is used. It is possible that random pooling might regain the upperhand if they fine-tuned the amount of dropout used.

Clevert et al. (2016) proposed the exponential linear units (ELUs) for faster and more precise learning in deep neural networks. ELUs have negative values, which allows the network to push the mean activations closer to zero. Therefore ELUs decrease the gap between the normal gradient and the unit natural gradient and, thereby speed up learning. ELUs have a clear saturation plateau in its negative regime, allowing them to learn a more robust and stable representation. Experimental results show that ELUs significantly outperform many activation functions on different vision datasets. Further ELU networks perform significantly better than Rectified Linear Unit (ReLU) networks trained with batch normalization. ELU networks achieved one of the top 10 best reported results on CIFAR-10 and set a new state of the art in CIFAR-100 without the need for multi-view test evaluation or model averaging. Furthermore, ELU networks produced competitive results on the ImageNet in much fewer epochs than a corresponding ReLU network. Zhao et al. (2016) proposed a novel architecture, the "stacked what-where auto-encoders" (SWWAE), which integrates discriminative and generative pathways and provides a unified approach to supervised, semi-supervised and unsupervised learning without relying on sampling during training. An instantiation of SWWAE uses a convolutional network to encode the input, and employs a deconvolutional network to produce the reconstruction. The objective function includes reconstruction terms that induce the hidden states in the Deconvnet to be similar to those of the Convnet. Each pooling layer produces two sets of variables: the "what" which are fed to the next layer, and its complementary variable "where" that are fed to the corresponding layer in the generative decoder. Zhang et al. (2017) proposed a new Residual network of Residual network architecture (RoR), which was proved capable of obtaining a new state-of-the-art performance on CIFAR-10 and CIFAR-100 for image classification. Through empirical studies, their work not only significantly advanced the image classification performance but can also provide an effective complement to the residual-networks family in the future.

Huang et al. (2017) proposed a new convolutional network architecture, which they refer to as Dense Convolutional Network (DenseNet). It introduces direct connections between any two layers with the same feature-map size. They showed that DenseNets scale naturally to hundreds of layers while exhibiting no optimization difficulties. In their experiments, DenseNets tend to yield consistent improvement in accuracy with a growing number of parameters, without any signs of performance degradation or overfitting. Under multiple settings, it achieved state-of-the-art results across several highly competitive datasets. Moreover, DenseNets require substantially fewer parameters and less computation to achieve state-of-the-art performances. Because they adopted hyperparameter settings optimized for residual networks in their study, they believe that further gains in accuracy of DenseNets may be obtained by more detailed tuning of hyperparameters and learning rate schedules. Whilst following a simple connectivity rule, DenseNets naturally integrate the properties of identity mappings, deep supervision, and diversified depth. They allow feature reuse throughout the networks and can consequently learn more compact and, according to our experiments, more accurate models. Because of their compact internal representations and reduced feature redundancy. Yamada et al. (2018) roposed a new stochastic regularization method ShakeDrop which can be successfully applied to ResNet and its improvements as long as the residual blocks end with batch normalization (BN). Its effectiveness was confirmed through the experiments on CIFAR-10 and CIFAR-100 datasets. ShakeDrop achieved state-of-the-art performance in the CIFAR-10 and CIFAR-100 datasets. Nayman et al. (2019) proposed a novel optimization method for differential neural architecture search (XNAS), a Prediction with Experts Advice (PEA) principled optimization method for differential neural architecture search. Inner network architecture weights that govern operations and connections, i.e. experts, are learned via exponentiated-gradient back-propagation update rule. XNAS optimization criterion is well suited for architecture-selection since it minimizes the regret implied by sub-optimal selection of operations with a tendency for sparsity while enabling late bloomers experts to warm-up and takes over during the search phase. Regret analysis suggests the use of multiple learning rates based on the amount of information carried by the backward gradient. A dynamic mechanism for wiping out weak experts is used, reducing the size of the computational graph along with the search phase, hence reducing the search time and increasing the final accuracy. Liang et al. (2020) proposed Drop-Activation, a regularization method that introduces randomness on the activation function. Drop-Activation works by randomly dropping the nonlinear activations in the network during training and uses a deterministic network with modified nonlinearities for prediction. The advantage of the proposed method is two-fold. Firstly, Drop-Activation provides a simple yet effective method for regularization, as demonstrated by the numerical experiments. Furthermore, this is supported by their analysis in the case of one hidden-layer. they show that Drop-Activation gives rise to a regularizer that penalizes the difference between nonlinear and linear networks. Future direction includes the analysis of Drop-Activation with more than one hidden layer. Secondly, experiments verify that DropActivation improves the generalization in most modern neural networks and cooperates well with some other popular training techniques. Moreover, they show theoretically and numerically that Drop-Activation maintains the variance during both training and testing time, and thus Drop-Activation can work well with Batch Normalization.

In overall the Table 1 shows the classification accuracy for all models that are reviewed in literature based on image classification using CIFAR-10 and CIFAR-100 datasets.

Year	**Model**	**CIFAR-10**	**CIFAR-100**
2013	Maxout	90.65%	61.43%
2014	DasNet	90.78%	66.22%
2014	NIN	91.20%	64.32%
2015	DNGO	93.63%	72.60%
2015	RCNN	92.91%	68.25%
2015	FMP	96.53%	73.61%
2016	ELUs	93.45%	75.72%
2016	SWWAE	92.23%	69.12%
2017	RoR	96.23%	80.27%
2017	DenseNet	96.54%	82.82%
2018	ShakeDrop	97.69%	87.81%
2019	XNAS	98.40%	86.40%
2020	Drop-Activation	96.55%	83.80%

3. Conclusion

This paper reviews some current advance in deep learning models based on image classification. The model trained by the deep learning algorithm has made remarkable achievements in many large-scale identification tasks in the field of computer vision since its introduction. This paper first introduces the deep learning, and then the latest model that has been used for image classification by deep learning are reviewed. Finally, all used deep learning models in the literature have been compared to each other in terms of accuracy for the two most challenging datasets CIFAR-10 and CIFAR-100.

References

Ahmed, O. M. (2020). Gene Expression Classification Based on Deep Learning. *2019 4th Scientific International Conference Najaf (SICN), IEEE*, 145–149.

Ahmed, O. M., & Abduallah, W. M. (2017). A Review on Recent Steganography Techniques in Cloud Computing. *Academic Journal of Nawroz University; Vol 6 No 3 (2017)*. https://doi.org/10.25007/ajnu.v6n3a91

Ahmed, O. M., & Sallow, A. B. (2017). Android Security: A Review. *Academic Journal of Nawroz University; Vol 6 No 3 (2017)*. https://doi.org/10.25007/ajnu.v6n3a99

Alzakholi, O., Shukur, H., Zebari, R., Abas, S., & Sadeeq, M. (2020). Comparison Among Cloud Technologies and Cloud Performance. *Journal of Applied Science and Technology Trends*, *1*(2), 40–47.

Clevert, D.-A., Unterthiner, T., & Hochreiter, S. (2016). Fast and accurate deep network learning by exponential linear units (elus). *ArXiv Preprint ArXiv:1511.07289*.

Dino, H., Abdulrazzaq, M. B., Zeebaree, S. R. M., Sallow, A. B., Zebari, R. R., Shukur, H. M., & Haji, L. M. (2020). Facial Expression Recognition based on Hybrid Feature Extraction Techniques with Different Classifiers. *TEST Engineering & Management*, *83*, 22319–22329.

Dino, H. I., Zeebaree, S. R. M., Ahmad, O. M., Shukur, H. M., Zebari, R. R., & Haji, L. M. (2020). *Impact of Load Sharing on Performance of Distributed Systems Computations*.

Goodfellow, I., Warde-Farley, D., Mirza, M., Courville, A., & Bengio, Y. (2013). Maxout networks. *International Conference on Machine Learning*, 1319–1327.

Graham, B. (2015). Fractional max-pooling. *ArXiv Preprint ArXiv:1412.6071*.

Haji, L. M., Zebari, R. R., Zeebaree, S. R. M., WAFAA, M. A., Shukur, H. M., & Alzakholi, O. (2020). *GPUs Impact on Parallel Shared Memory Systems Performance–International Journal of Psychosocial Rehabilitation*. May.

Haji, Lailan M, Ahmad, O. M., Zeebaree, S. R. M., Dino, H. I., Zebari, R. R., & Shukur, H. M. (2020). *Impact of Cloud Computing and Internet of Things on the Future Internet*.

Haji, L. M., Zeebaree, S. R., Ahmed, O. M., Sallow, A. B., Jacksi, K., & Zeabri, R. R. (2020). Dynamic Resource Allocation for Distributed Systems and Cloud Computing. *TEST Eng. Manag*, *83*, 22417–22426.

Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 4700–4708.

Hussein, H. I., & Abduallah, W. M. (2018). A modified table lookup substitution method for hiding data in DNA. *2018 International Conference on Advanced Science and Engineering (ICOASE)*, 268–273.

Liang, M., & Hu, X. (2015). Recurrent convolutional neural network for object recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3367–3375.

Liang, S., Khoo, Y., & Yang, H. (2020). Drop-activation: Implicit parameter reduction and harmonic regularization. *ArXiv Preprint ArXiv:1811.05850*.

Lin, M., Chen, Q., & Yan, S. (2014). Network in network. *ArXiv Preprint ArXiv:1312.4400*.

Mohammed, A. J., Hassan, M. M., & Kadir, D. H. (2020). Improving Classification Performance for a Novel Imbalanced Medical Dataset using SMOTE Method. *International Journal*, *9*(3).

Mustafa, O. M., Haji, D., Ahmed, O. M., & Haji, L. M. (2020). Big Data: Management, Technologies, Visualization, Techniques, and Privacy. *Technology Reports of Kansai University*, *62*(05).

Nayman, N., Noy, A., Ridnik, T., Friedman, I., Jin, R., & Zelnik, L. (2019). Xnas: Neural architecture search with expert advice. *Advances in Neural Information Processing Systems*, 1975–1985.

Salih, A. A., Zeebaree, S. R. M., Abdulraheem, A. S., Zebari, R. R., Sadeeq, M. A. M., & Ahmed, O. M. (2020). Evolution of Mobile Wireless Communication to 5G Revolution. *Technology Reports of Kansai University*.

Shukur, H., Zeebaree, S., Zebari, R., Ahmed, O., Haji, L., & Abdulqader, D. (2020). Cache Coherence Protocols in Distributed Systems. *Journal of Applied Science and Technology Trends*, *1*(3), 92–97.

Shukur, H., Zeebaree, S., Zebari, R., Zeebaree, D., Ahmed, O., & Salih, A. (2020). Cloud Computing Virtualization of Resources Allocation for Distributed Systems. *Journal of Applied Science and*

Technology Trends, *1*(3), 98–105.
Snoek, J., Rippel, O., Swersky, K., Kiros, R., Satish, N., Sundaram, N., Patwary, M., Prabhat, M., & Adams, R. (2015). Scalable bayesian optimization using deep neural networks. *International Conference on Machine Learning*, 2171–2180.
Stollenga, M. F., Masci, J., Gomez, F., & Schmidhuber, J. (2014). Deep networks with internal selective attention through feedback connections. *Advances in Neural Information Processing Systems*, 3545–3553.
Yamada, Y., Iwamura, M., & Kise, K. (2018). *Shakedrop regularization*.
Zebari, D. A., Haron, H., Zeebaree, S. R. M., & Zeebaree, D. Q. (2019). Enhance the Mammogram Images for Both Segmentation and Feature Extraction Using Wavelet Transform. *2019 International Conference on Advanced Science and Engineering (ICOASE)*, 100–105.
Zebari, R., Abdulazeez, A., Zeebaree, D., Zebari, D., & Saeed, J. (2020). A Comprehensive Review of Dimensionality Reduction Techniques for Feature Selection and Feature Extraction. *Journal of Applied Science and Technology Trends*, *1*(2), 56–70.
Zeebaree, D. Q., Haron, H., & Abdulazeez, A. M. (2018). Gene Selection and Classification of Microarray Data Using Convolutional Neural Network. *2018 International Conference on Advanced Science and Engineering (ICOASE)*, 145–150. https://doi.org/10.1109/ICOASE.2018.8548836
Zeebaree, D. Q., Haron, H., Abdulazeez, A. M., & Zeebaree, S. R. M. (2017). Combination of K-means clustering with Genetic Algorithm: A review. *International Journal of Applied Engineering Research*, *12*(24), 14238–14245.
Zeebaree, S. R. M., Haji, L. M., Rashid, I., Zebari, R. R., Ahmed, O. M., Jacksi, K., & Shukur, H. M. (2020). Multicomputer Multicore System Influence on Maximum Multi-Processes Execution Time. *TEST Engineering & Management*, *83*(03), 14921–14931.
Zhang, K., Sun, M., Han, T. X., Yuan, X., Guo, L., & Liu, T. (2017). Residual networks of residual networks: Multilevel residual networks. *IEEE Transactions on Circuits and Systems for Video Technology*, *28*(6), 1303–1314.
Zhao, J., Mathieu, M., Goroshin, R., & Lecun, Y. (2016). Stacked what-where auto-encoders. *ArXiv Preprint ArXiv:1506.02351*.
Zhao, Z.-Q., Zheng, P., Xu, S., & Wu, X. (2019). Object detection with deep learning: A review. *IEEE Transactions on Neural Networks and Learning Systems*, *30*(11), 3212–3232.

Cite this article:

Kavi B. Obaid, Subhi R. M. Zeebaree & Omar M. Ahmed (2020). Deep Learning Models Based on Image Classification: A Review. *International Journal of Science and Business, 4*(11), 75-81. doi: https://doi.org/10.5281/zenodo.4108433

Retrieved from http://ijsab.com/wp-content/uploads/612.pdf

Published by

IJSAB International