An overview of deep learning models in computer vision applications

Conference Paper · June 2022

CITATIONS
0

READS
21

2 authors:

Sanjeev Kumar
Maharshi Dayanand University
1 PUBLICATION **0** CITATIONS
SEE PROFILE

Yogesh Kumar
Maharshi Dayanand University
3 PUBLICATIONS **1** CITATION
SEE PROFILE

An overview of deep learning models in computer vision applications

Sanjeev Kumar[1*], Yogesh Kumar[2]
[1]Department of Computer Science and Engineering, UIET, M.D. University, Rohtak
[2]Faculty of dept. of Computer Science and Engineering, UIET, M.D. University, Rohtak
[1]sanjeev.rs.uiet@mdurohtak.ac.in, [2]dryogeshkumar.uiet@mdurohtak.ac.in

Abstract: Demand and popularity of deep learning in day to day life in the applications like classification of images, detection of object and face detection develop a sense of excitement of understanding deep learning tools and techniques that can improve the functioning of above applications.This paper provides an insight on the research works performed so far of using deep learning tools and techniques in computer vision applications and also discussed specific CNNs models such as GoogLeNet, VGG S model, AlexNet CNN model etc. which are widely used model in computer vision and identify various issues like over fitting, learning rate capacity etc.

Keywords: *Deep learning, Convolutional NN, computer vision*

1. Introduction

In recent years, the influence of deep learning and artificial intelligence in the field of computer vision has become widely popular(H. Lee et al., 2009). Applications like object detection given by(Girshick et al., 2014) and (Lu et al., 2019),text classification(Lyu & Liu, 2021) and (Aggarwal & Zhai, 2013), face recognition(*M.Liu et al., 2013*),image classification(Gonzalez, 2007), gender classification(Makinen & Raisamo, 2008), scene classification(Zhou et al., 2014), traffic signs recognition(D. Ciregan et al. 2012) uses deep learning tools and techniques. Few applications for computer vision are AlexNet model(Gonzalez, 2007), VGG S model introduced by(Chatfield et al., 2014), gender classification model(Pandi et al., 2022), GoogLeNet model(Ge et al., 2015), Places-CNN model(Zhou et al., 2014), etc. These models raised issues like over fitting, learning rate and capacity, connectivity of neural network nodes etc.Working of deep learning considered various factors such as advanced GPU that reduces time for training and testing, availability of large data set for training etc.

This paper reviewed different computer vision applications like image detection, face detection, places-recognition etc. and associated deep learning models that will help the new researchers to give a picture of the "How much important deep learning for computer vision?"

1.1 Deep Learning

"Deep learning"(Voulodimos et al., 2018) is a subset of an artificial intelligence. It came into picture more precise in non-linear data analysis. Voice interface used in many automobiles are based on deep learning as it learns more as you talk to it. It uses a popular neural network algorithm called Long short term memory algorithm(LSTM) in voice recognition(Park et al.,

2019). Learning in computer means developing a database which consist of neural network for data processing.

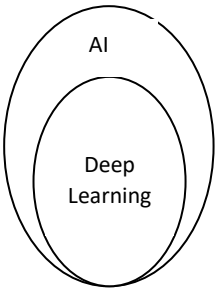

Fig.1. Deep learning as a subset of Artificial Intelligence

1.2.Artificial neural network

Human brain inspired parallel computational model that consist of large amount of neurons. The ANN is classified into single and multi linear neural networks and the architecture consists of input layer, output layers. These layers are associated with processing nodes that contains weights which results into series of output nodes(Sivanandam & Deepa, 2005).ANN consider various parameters like learning rate, activation function, weights etc. and adjust all these parameters well to produce required results.

1.2.1. Convolutional neural networks

CNN is a most widely used deep neural networks for applications like pattern recognition, image classification, facial recognition etc. CNN majorly consists of 3 layers known as convolutional layer, pooling layer and last one is fully connected layer that connects neuron nodes locally which results in reduction in the number of connections(Gu et al., 2018).

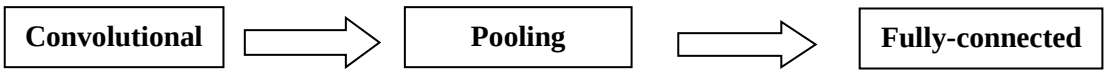

Fig.2. Layers of CNN

Layers of CNN

1. Convolutional layer determine the neuron outputs using local regions of input layers by calculating scalar product of weights and regions attached with input(O'Shea & Nash, 2015).
2. Pooling layer performs downsampling of given inputs and reducing parameters.
3. Fully-connected layer calculate the output from given activation function(Albawi et al., 2017).

Applications like object detection(Girshick et al., 2014), text classification(Lyu & Liu, 2021), face recognition(*M.Liu et al., 2013*),image classification(Gonzalez, 2007), gender

classification(Makinen & Raisamo, 2008), scene classification(Zhou et al., 2014), traffic signs recognition(D. Ciregan et al. 2012) etc. uses convolutional neural networks.

Table.1.Various CNN-models

Sno.	Authors	CNN-Model	Application
1.	(Russakovsky et al., 2015)	AlexNet-CNN	Image classification
2.	(Gliner et al., 2021)	GoogleNet-CNN	Image recognition
3.	(Simonyan & Zisserman, 2015)	VGGNet-CNN	Image classification
4.	Kaiming H et al. 2015	ResNet-CNN	Object detection(Lu et al., 2019)
5.	LeCun Y et al., 1996	LeNet-CNN	Digit recognition
6.	(Schroff et al., 2015)	FaceNet-CNN	Face recognition
7.	(Brandon Amos, Bartosz Ludwiczuk, 2016)	OpenFace-CNN	Face recognition
8.	(Gan et al., 2015)	DevNet-CNN	Event video recognition

2. Literature review

A multi-column neural network are used for traffic signs recognition(D. Ciregan et al. 2012). A DropConnect method proposed by(Hinton et al., 2012) on dropout and (C. Y. Lee et al., 2018) proposed DropConnect method for digit recognition. AU-aware networks were introduced for facial recognition(*M.Liu et al., 2013*) and these networks used Restricted Boltzmann Machines to extract more advanced features.(Gonzalez, 2007) introduced a newly CNN model for image recognition with less error rate over test data. (H. Lee, 2009) proposed a model for image recognition that can recognise large amount of images by using lesser number of features and also scale the image sizes according to requirement. (Dey et al., 2015) used Berkeley trained model in garment texture classification.(Zhou et al., 2014) uses VGG S-16, a CNN technique for scene recognition to identify the differences in images in terms of density and diversity. A visualisation method was proposed to understand the functionality of inner layers of features (Suah, 2017) and (Goodley & Tregaskis, 2006) proposed unsupervised CNN model for generic classification of images and successfully found out neural patterns. (Xia et al., 2015) Introduced discriminating auto encoder constructions using DRAE and trying to remove errors like noisy data. Optimisation of CNN in land-use classification can be achieved using a heuristic approach of multi view and unsupervised deep learning methods (Luus et al., 2015). A CNN architecture was proposed for gender and age prediction(Pandi et al., 2022) and a Deep Belief Networks(H. Lee et al., 2009) was proposed to learns the features of images, predicted more than GIST(Niko & Protzel, 2011).

Table.2. List of model based applications

Sno.	Authors	Application	Model
1.	(H. Lee et al., 2009)	Audio classification	CDB Networks
2.	(Girshick et al., 2014)	Object detection	Regional-CNN

3.	(Lyu & Liu, 2021)	Text classification	Recurrent CNN
4.	Krizhevsky et al. 2012	Image classification	Deep CNN
5.	(M.Liu et al., 2013)	Facial recognition	AU-aware Deep networks
6.	(Ge et al., 2015)	Plant classification	Deep-CNN
7.	(Park et al., 2019)	Voice interface	LSTM
8.	(D. Ciregan et al. 2012)	Image classification	Multi-column Deep-NN
9.	(H. Lee, 2009)	Image classification	Convolutional Deep Belief Networks
10.	(Pandi et al., 2022)	Age and gender classification	Deep-CNN
11.	(Luus et al., 2015)	Land-use classification	Multi view Deep-CNN
12.	(Makinen & Raisamo, 2008)	Gender classification	Support Vector Machine and adaboost
13.	(Zhou et al., 2014)	Scene recognition	Deep-CNN
14.	(Goodley & Tregaskis, 2006)	Image classification	AlexNet, VGG-19 neural networks
15.	(Helou & Nguyen, 2011)	Scene recognition	D-Belief networks(Hinton et al., 2006)
16.	(Dey et al., 2015)	Garment texture design recognition	Completed CENTRIST and Ternary CENTRIST methods.
17.	(Sabanci et al., 2017)	Food-grain classification	Levenberg–Marquardt (LM) algorithm(Hagan & Menhaj, 1994)
18.	(Toshev & Szegedy, 2014)	Deep pose estimation	DNN-based regressors
19.	(Taigman et al., n.d.)	Face recognition	Standard back-propagation on feed forward nets
20.	(Karpathy & Leung, 2014)	Video classification	Deep-CNN
21.	(Kitsikidis et al., 2014)	Human Pose classification	Hidden-state Conditional Random Fields (HCRF) classifier
22.	(Felzenszwalb & Huttenlocher, 2005)	Object recognition	Baysian network
23.	(Li et al., 2018)	Object detection	Saliency-guided Stacked Autoencoders(Srinivas et al., 2016)
24.	(Song et al., 2016)	Egocentric Activity recognition	Long Short-Term Memory (Kavi et al., 2016)
25.	(Makantasis et al., 2016)	Human behaviour recognition	MLP-CNN

3. Conclusion/Future work

Deep CNN proved to be powerful tools in performing computer-vision applications. Through this paper, we overviewed importance of Convolutional-NN in the sphere of computer vision. We studied various computer vision applications and different CNN models which are used

in computer vision applications. The research need to focus on reducing energy consumptions of CNNs in low-power devices. The fine-tuning of CNN layer can increase accuracy in the model and study on combination of DCNN and other neural network can be a further topic of research.

4. References

20141317351319923_2013_Fg_Myliu_Au-Aware Deep Networks for Facial Expression Recognition.Pdf. (n.d.).

Aggarwal, C. C., & Zhai, C. X. (2013). Mining text data. In *Mining Text Data* (Vol. 9781461432234). https://doi.org/10.1007/978-1-4614-3223-4

Albawi, S., Mohammed, T. A. M., & Alzawi, S. (2017). Layers of a Convolutional Neural Network. *Ieee*, 16.

Ponti, M. A., Ribeiro, L. S. F., Nazare, T. S., Bui, T., & Collomosse, J. (2017, October). Everything you wanted to know about deep learning for computer vision but were afraid to ask. In *2017 30th SIBGRAPI conference on graphics, patterns and images tutorials (SIBGRAPI-T)* (pp. 17-41). IEEE.

Brandon Amos, Bartosz Ludwiczuk, M. S. (2016). Openface: uniwersalna biblioteka rozpoznawania twarzy z aplikacjami mobilnymi. *CMU School of ...*, *2016 R.*, *16*(October 2015), 1–18. http://reports-archive.adm.cs.cmu.edu/anon/anon/usr0/ftp/2016/CMU-CS-16-118.pdf

Chatfield, K., Simonyan, K., Vedaldi, A., & Zisserman, A. (2014). Return of the devil in the details: Delving deep into convolutional nets. *BMVC 2014 - Proceedings of the British Machine Vision Conference 2014*, 1–11. https://doi.org/10.5244/c.28.6

Dey, E. K., Tawhid, M. N. A., & Shoyaib, M. (2015). An automated system for garment texture design class identification. *Computers*, *4*(3), 265–282. https://doi.org/10.3390/computers4030265

Felzenszwalb, P. F., & Huttenlocher, D. P. (2005). *Pictorial Structures for Object Recognition_International Journal of Computer Vision_2005_Felzenszwalb, Huttenlocher(冲突_WIN-FG4A7VV8SD8_2012-06-12 10-52-38).pdf*. *61*(1), 55–79.

Goel, A., Tung, C., Lu, Y. H., & Thiruvathukal, G. K. (2020, June). A survey of methods for low-power deep learning and computer vision. In *2020 IEEE 6th World Forum on Internet of Things (WF-IoT)* (pp. 1-6). IEEE.

Gan, C., Wang, N., Yang, Y., Yeung, D. Y., & Hauptmann, A. G. (2015). DevNet: A Deep Event Network for multimedia event detection and evidence recounting. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, *07-12-June-2015*, 2568–2577. https://doi.org/10.1109/CVPR.2015.7298872

Ge, Z. Y., McCool, C., Sanderson, C., & Corke, P. (2015). Content specific feature learning for fine-grained plant classification. *CEUR Workshop Proceedings*, *1391*.

Girshick, R., Donahue, J., Darrell, T., Malik, J., Berkeley, U. C., & Malik, J. (2014). 1043.0690. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, *1*, 5000. https://doi.org/10.1109/CVPR.2014.81

Gliner, J. A., Morgan, G. A., Leech, N. L., Gliner, J. A., & Morgan, G. A. (2021). Measurement Reliability and Validity. *Research Methods in Applied Settings*, 319–338.

https://doi.org/10.4324/9781410605337-29

Gonzalez, T. F. (2007). Handbook of approximation algorithms and metaheuristics. *Handbook of Approximation Algorithms and Metaheuristics*, 1–1432. https://doi.org/10.1201/9781420010749

Goodley, D., & Tregaskis, C. (2006). *Retrospective Accounts of Disabled Family Life*.

Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., Liu, T., Wang, X., Wang, G., Cai, J., & Chen, T. (2018). Recent advances in convolutional neural networks. *Pattern Recognition*, *77*, 354–377. https://doi.org/10.1016/j.patcog.2017.10.013

Hagan, M. T., & Menhaj, M. B. (1994). Training Feedforward Networks with the Marquardt Algorithm. *IEEE Transactions on Neural Networks*, *5*(6), 989–993. https://doi.org/10.1109/72.329697

Helou, A., & Nguyen, C. (2011). *Unsupervised Deep Learning for Scene Recognition*. 1–10.

Hinton, G. E., Osindero, S., & Teh, Y.-W. (2006). 2006 Dbn. *Neural Computation*, *18*, 1527–1554.

Hinton, G. E., Srivastava, N., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. R. (2012). *Improving neural networks by preventing co-adaptation of feature detectors*. 1–18. http://arxiv.org/abs/1207.0580

Islam, S. S., Rahman, S., Rahman, M. M., Dey, E. K., & Shoyaib, M. (2016, May). Application of deep learning to computer vision: A comprehensive study. In *2016 5th international conference on informatics, electronics and vision (ICIEV)* (pp. 592-597). IEEE.

Karpathy, A., & Leung, T. (2014). Karpathy_Large-scale_Video_Classification_2014_CVPR_paper. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 10–20. http://cs.stanford.edu/people/karpathy/deepvideo

Kavi, R., Kulathumani, V., Rohit, F., & Kecojevic, V. (2016). Multiview fusion for activity recognition using deep neural networks. *Journal of Electronic Imaging*, *25*(4), 043010. https://doi.org/10.1117/1.jei.25.4.043010

Kitsikidis, A., Dimitropoulos, K., Douka, S., & Grammalidis, N. (2014). Dance analysis using multiple kinect sensors. *VISAPP 2014 - Proceedings of the 9th International Conference on Computer Vision Theory and Applications*, *2*, 789–795. https://doi.org/10.5220/0004874007890795

Lee, C. Y., Gallagher, P., & Tu, Z. (2018). Generalizing Pooling Functions in CNNs: Mixed, Gated, and Tree. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *40*(4), 863–875. https://doi.org/10.1109/TPAMI.2017.2703082

Lee, H. (2009). Articulo "Convolutional Deep Belief Networks for Scalable Unsupervised Learning of Hierarchical Representations" Departamento de Ciencias Computacionales , Universidad Stanford. *Proceedings of the 26th International Confer- Ence on Machine Learning*, 8. http://robotics.stanford.edu/~ang/papers/icml09-ConvolutionalDeepBeliefNetworks.pdf

Lee, H., Yan, L., Pham, P., & Ng, A. Y. (2009). Unsupervised feature learning for audio classification using convolutional deep belief networks. *Advances in Neural Information Processing Systems 22 - Proceedings of the 2009 Conference*, 1096–1104.

Li, J., Xia, C., & Chen, X. (2018). A benchmark dataset and saliency-guided stacked autoencoders for video-based salient object detection. *IEEE Transactions on Image Processing*, *27*(1), 349–364. https://doi.org/10.1109/TIP.2017.2762594

Lu, S., Wang, B., Wang, H., Chen, L., Linjian, M., & Zhang, X. (2019). A real-time object detection algorithm for video. *Computers and Electrical Engineering*, *77*, 398–408. https://doi.org/10.1016/j.compeleceng.2019.05.009

Luus, F. P. S., Salmon, B. P., Van Den Bergh, F., & Maharaj, B. T. J. (2015). Multiview Deep Learning for Land-Use Classification. *IEEE Geoscience and Remote Sensing Letters*, *12*(12), 2448–2452. https://doi.org/10.1109/LGRS.2015.2483680

Lyu, S., & Liu, J. (2021). Convolutional recurrent neural networks for text classification. *Journal of Database Management*, *32*(4), 65–82. https://doi.org/10.4018/JDM.2021100105

Makantasis, K., Doulamis, A., & Doulamis, N. (2016). DEEP LEARNING BASED HUMAN BEHAVIOR RECOGNITION IN INDUSTRIAL WORKFLOWS Technical University of Crete , School of Production Engineering and Management , Chania , Greece National Technical University of Athens , School of Electrical and Computer Engineerin. *Ieee Icip*, 3–7.

Makinen, E., & Raisamo, R. (2008). Evaluation of gender classification methods with automatically detected and aligned faces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *30*(3), 541–547. https://doi.org/10.1109/TPAMI.2007.70800

Niko, S., & Protzel, P. (2011). *BRIEF-Gist – Closing the Loop by Simple Means*. 1234–1241.

O'Shea, K., & Nash, R. (2015). *An Introduction to Convolutional Neural Networks*. 1–11. http://arxiv.org/abs/1511.08458

Pandi, C., Adi Narayana Reddy, K., Alladi, R., Chandra Sekhar Reddy, V., & Sumithabhashini, P. (2022). Emotion and Gender Classification Using Convolution Neural Networks. *Lecture Notes in Networks and Systems*, *321*, 563–573. https://doi.org/10.1007/978-981-16-5987-4_57

Park, J., Son, H., Lee, J., & Choi, J. (2019). Driving Assistant Companion with Voice Interface Using Long Short-Term Memory Networks. *IEEE Transactions on Industrial Informatics*, *15*(1), 582–590. https://doi.org/10.1109/TII.2018.2861739

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C., & Fei-Fei, L. (2015). ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision*, *115*(3), 211–252. https://doi.org/10.1007/s11263-015-0816-y

Sabanci, K., Kayabasi, A., & Toktas, A. (2017). Computer vision-based method for classification of wheat grains using artificial neural network. *Journal of the Science of Food and Agriculture*, *97*(8), 2588–2593. https://doi.org/10.1002/jsfa.8080

Schroff, F., Kalenichenko, D., & Philbin, J. (2015). FaceNet: A unified embedding for face recognition and clustering. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, *07-12-June-2015*, 815–823. https://doi.org/10.1109/CVPR.2015.7298682

Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, 1–14.

Sivanandam, S. N., & Deepa, S. N. (2005). *Principles of S oft Computing About the Authors*.

Song, S., Chandrasekhar, V., Mandal, B., Li, L., Lim, J. H., Babu, G. S., San, P. P., & Cheung, N. M. (2016). Multimodal Multi-Stream Deep Learning for Egocentric Activity Recognition. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 378–385. https://doi.org/10.1109/CVPRW.2016.54

Srinivas, S., Sarvadevabhatla, R. K., Mopuri, K. R., Prabhu, N., Kruthiventi, S. S. S., & Babu, R. V. (2016). A taxonomy of deep convolutional neural nets for computer vision. *Frontiers Robotics AI*, *2*(JAN), 1–13. https://doi.org/10.3389/frobt.2015.00036

Suah, F. B. M. (2017). Preparation and characterization of a novel Co(II) optode based on polymer inclusion membrane. *Analytical Chemistry Research*, *12*, 40–46.

https://doi.org/10.1016/j.ancr.2017.02.001

Taigman, Y., Ranzato, M. A., Aviv, T., & Park, M. (n.d.). *Taigman_DeepFace_Closing_the_2014_CVPR_paper*. https://doi.org/10.1109/CVPR.2014.220

Toshev, A., & Szegedy, C. (2014). DeepPose_Human_Pose_2014_CVPR_paper.pdf. *2014 IEEE Conference on Computer Vision and Pattern Recognition*, 1653–1660. https://www.cv-foundation.org/openaccess/content_cvpr_2014/html/Toshev_DeepPose_Human_Pose_2014_CVPR_paper.html

Voulodimos, A., Doulamis, N., Doulamis, A., & Protopapadakis, E. (2018). Deep Learning for Computer Vision: A Brief Review. *Computational Intelligence and Neuroscience*, *2018*. https://doi.org/10.1155/2018/7068349

Xia, Y., Cao, X., Wen, F., Hua, G., & Sun, J. (2015). Learning discriminative reconstructions for unsupervised outlier removal. *Proceedings of the IEEE International Conference on Computer Vision*, *2015 Inter*, 1511–1519. https://doi.org/10.1109/ICCV.2015.177

Zhou, B., Lapedriza, A., Xiao, J., Torralba, A., & Oliva, A. (2014). Learning Deep Features for Scene Recognition using Places Database - Supplementary Materials. *NIPS'14 Proceedings of the 27th International Conference on Neural Information Processing Systems*, *1*, 487–495.