IJESMR

International Journal OF Engineering Sciences & Management Research

HYBRID DEEP LEARNING MODELS: BRIDGING THE GAP BETWEEN SUPERVISED AND UNSUPERVISED TECHNIQUES

[1]Sri Bhargav Krishna Adusumilli, [2]Harini Damancharla
[1]Senior Software Engineer, Research Scholar,
[2]Senior Software Engineer, Research Scholar,
[1]sribhargav09@gmail.com, [2]damanharini@gmail.com
[1]https://orcid.org/0009-0005-4059-387X, [2]https://orcid.org/0009-0000-3899-3325

Keywords: Deep Learning, Convolutional Neural Network, Recurrent neural networks , Restricted Boltzmann Machine, Auto encoders, Extreme Learning

ABSTRACT

Deep learning (DL), a branch of machine learning (ML) and artificial intelligence (AI) is nowadays considered as a core technology of today's Fourth Industrial Revolution (4IR or Industry 4.0). Due to its learning capabilities from data, DL technology originated from artificial neural network (ANN), has become a hot topic in the context of computing, and is widely applied in various application areas like healthcare, visual recognition, text analytics, cybersecurity, and many more. However, building an appropriate DL model is a challenging task, due to the dynamic nature and variations in real-world problems and data. Moreover, the lack of core understanding turns DL methods into black-box machines that hamper development at the standard level. This article presents a structured and comprehensive view on DL techniques including a taxonomy considering various types of real-world tasks like supervised or unsupervised. In our taxonomy, we take into account deep networks for supervised or discriminative learning, unsupervised or generative learning as well as hybrid learning and relevant others. We also summarize real-world application areas where deep learning techniques can be used. Finally, we point out ten potential aspects for future generation DL modeling with research directions. Overall, this article aims to draw a big picture on DL modeling that can be used as a reference guide for both academia and industry professionals.

INTRODUCTION

In the late 1980s, neural networks became a prevalent topic in the area of Machine Learning (ML) as well as Artificial Intelligence (AI), due to the invention of various efficient learning methods and network structures [52]. Multilayer perceptron networks trained by "Backpropagation" type algorithms, self-organizing maps, and radial basis function networks were such innovative methods [26, 36, 37]. While neural networks are successfully used in many applications, the interest in researching this topic decreased later on. After that, in 2006, "Deep Learning" (DL) was introduced by Hinton et al. [41], which was based on the concept of artificial neural network (ANN). Deep learning became a prominent topic after that, resulting in a rebirth in neural network research, hence, some times referred to as "new-generation neural networks". This is because deep networks, when properly trained, have produced significant success in a variety of classification and regression challenges [52].

Nowadays, DL technology is considered as one of the hot topics within the area of machine learning, artificial intelligence as well as data science and analytics, due to its learning capabilities from the given data. Many corporations including Google, Microsoft, Nokia, etc., study it actively as it can provide significant results in different classification and regression problems and datasets [52]. In terms of working domain, DL is considered as a subset of ML and AI, and thus DL can be seen as an AI function that mimics the human brain's processing of data. The worldwide popularity of "Deep learning" is increasing day by day, which is shown in our earlier paper [96] based on the historical data collected from Google trends [33]. Deep learning differs from standard machine learning in terms of efficiency as the volume of data increases, discussed briefly in Section "Why Deep Learning in Today's Research and Applications?". DL technology uses multiple layers to represent the abstractions of data to build computational models. While deep learning takes a long time to train a model due to a large number of parameters, it takes a short amount of time to run during testing as compared to other machine learning algorithms [127].

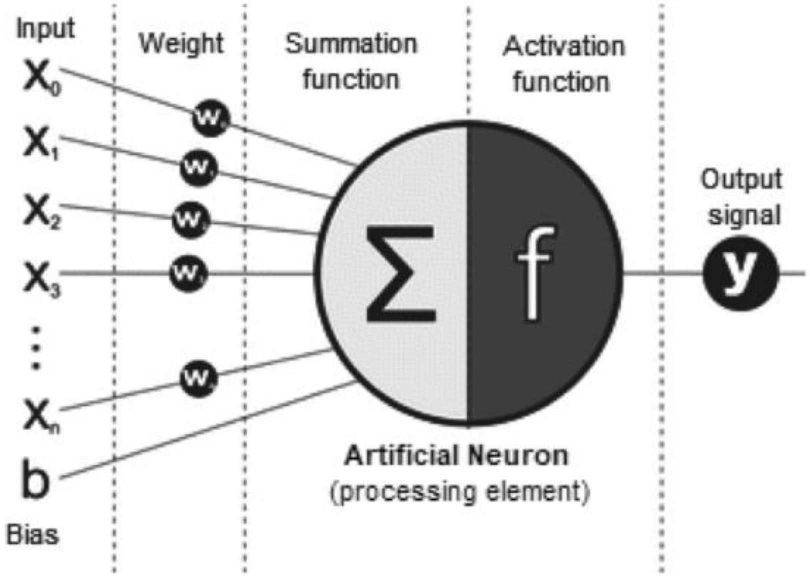
While today's Fourth Industrial Revolution (4IR or Industry 4.0) is typically focusing on technology-driven "automation, smart and intelligent systems", DL technology, which is originated from ANN, has become one of

 ISSN 2349-6193
Impact Factor: 2.805

IJESMR
International Journal OF Engineering Sciences & Management Research

the core technologies to achieve the goal [103, 114]. A typical neural network is mainly composed of many simple, connected processing elements or processors called neurons, each of which generates a series of real-valued activations for the target outcome. Figure 1 shows a schematic representation of the mathematical model of an artificial neuron, i.e., processing element, highlighting input (Xi), weight (w), bias (b), summation function ($\sum$), activation function (f) and corresponding output signal (y). Neural network-based DL technology is now widely applied in many fields and research areas such as healthcare, sentiment analysis, natural language processing, visual recognition, business intelligence, cybersecurity, and many more that have been summarized in the latter part of this paper.

Fig. 1



Schematic representation of the mathematical model of an artificial neuron (processing element), highlighting input (X_i), weight (w), bias (b), summation function ($\sum$), activation function (f) and output signal (y)

Although DL models are successfully applied in various application areas, mentioned above, building an appropriate model of deep learning is a challenging task, due to the dynamic nature and variations of real-world problems and data. Moreover, DL models are typically considered as "black-box" machines that hamper the standard development of deep learning research and applications. Thus for clear understanding, in this paper, we present a structured and comprehensive view on DL techniques considering the variations in real-world problems and tasks. To achieve our goal, we briefly discuss various DL techniques and present a *taxonomy* by taking into account three major categories: (i) deep networks for supervised or *discriminative learning* that is utilized to provide a discriminative function in supervised deep learning or classification applications; (ii) deep networks for unsupervised or *generative learning* that are used to characterize the high-order correlation properties or features for pattern analysis or synthesis, thus can be used as preprocessing for the supervised algorithm; and (ii) deep networks for *hybrid learning* that is an integration of both supervised and unsupervised model and relevant others. We take into account such categories based on the nature and learning capabilities of different DL techniques and how they are used to solve problems in real-world applications [97]. Moreover, identifying key research issues and prospects including effective data representation, new algorithm design, data-driven hyper-parameter learning, and model optimization, integrating domain knowledge, adapting resource-constrained devices, etc. is one of the key targets of this study, which can lead to "Future Generation DL-Modeling". Thus the goal of this paper is set to assist those in *academia and industry* as a reference guide, who want to *research and develop* data-driven smart and intelligent systems based on DL techniques.
The overall contribution of this paper is summarized as follows:

- This article focuses on different aspects of deep learning modeling, i.e., the learning capabilities of DL techniques in different dimensions such as supervised or unsupervised tasks, to function in an

 ISSN 2349-6193
Impact Factor: 2.805

IJESMR

International Journal OF Engineering Sciences & Management Research

automated and intelligent manner, which can play as a core technology of today's Fourth Industrial Revolution (Industry 4.0).

- We explore a variety of prominent DL techniques and present a taxonomy by taking into account the variations in deep learning tasks and how they are used for different purposes. In our taxonomy, we divide the techniques into three major categories such as deep networks for supervised or discriminative learning, unsupervised or generative learning, as well as deep networks for hybrid learning, and relevant others.
- We have summarized several potential real-world application areas of deep learning, to assist developers as well as researchers in broadening their perspectives on DL techniques. Different categories of DL techniques highlighted in our taxonomy can be used to solve various issues accordingly.
- Finally, we point out and discuss ten potential aspects with research directions for future generation DL modeling in terms of conducting future research and system development.

This paper is organized as follows. Section "Why Deep Learning in Today's Research and Applications?" motivates why deep learning is important to build data-driven intelligent systems. In Section" Deep Learning Techniques and Applications", we present our DL taxonomy by taking into account the variations of deep learning tasks and how they are used in solving real-world issues and briefly discuss the techniques with summarizing the potential application areas. In Section "Research Directions and Future Aspects", we discuss various research issues of deep learning-based modeling and highlight the promising topics for future research within the scope of our study. Finally, Section "Concluding Remarks" concludes this paper.

TECHNIQUES

Deep learning is a key area of research in the field of Image and Video processing, Computer vision [2,3] and Bioinformatics to name a few. This gave rise to the introduction and application of several variants of deep learning in the above mentioned fields. However, the techniques of deep learning generally are divided into three categories namely Convolutional Neural Networks(CNN), Restricted Boltzmann Machines(RBM) and Autoencoders. Additionally Recurrent Neural Networks and Extreme Learning are also a few techniques frequently used in this field. For better clarity, the architectural descriptions along with layer information and detailed functionalities of these techniques are described in a nutshell.

A. Convolutional Neural Network (CNNs)

When it comes to substantial training of multiple layers, the Convolutional Neural Network(CNN) is considered as the most momentous approach for a variety of applications [4]. The CNN architecture shown in the fig. 1, consists of mainly three types of layers, namely [1]-convolutional layers, pooling layer and fully connected layers. Training the network can be divided into forward and backward stages. In the forward stage first we classify the input image depending upon its weights and bias for each layer. The loss cost is calculated from the input data by using the predicted output. In the backward stage, depending on the loss cost measured, the gradients are calculated for each parameter. Using the gradients, it then updates the parameters for the next iteration. The training procedure can be halted after satisfactory number of repetitions. The functionalities of the said network is described as follows:

1) Convolutional Layers: In this layer, the CNN uses numerous filters to convolve the entire image including intermediate feature maps & generating different feature maps. The major privileges [5] of the convolution operation are-
• Reduce number of parameters using weight sharing mechanisms.
• Correlation between neighboring pixels are easy due to local connectivity.
• Location of object is fixed.
This advantages leads researchers to replace fully connected layers to put forward the learning process [6,14].

2) Pooling Layers: This layer is similar to convolution layer but minimizes the measurements of feature maps and also the parameters of the network. Generally average and max pooling are used. For average and max pooling, Boureu et al. [11] have demonstrated the theoretical details about their performances. Among all the three layers, the pooling layer happens to be the most profusely investigated. There exist mainly three

IJESMR
International Journal OF Engineering Sciences & Management Research

approaches, with varied usage, that are related to the pooling layers. All of these different pooling approaches are discussed as follows:

• Stochastic pooling- To overcome the problem of sensitivity for over-fitting the training set in max pooling [13], stochastic pooling is introduced. Here the activation in each pooling area is allocated randomly as per a multinomial distribution. In this method, similar kind of inputs are considered with very small variation to overcome the problem of over-fitting.

• Spatial pyramid pooling- As the input dimensions of an image is generally fixed while using CNN, the accuracy is compromised in case of of a variable sized image. This problem can be eliminated[15] by replacing the last layer of CNN architecture with spatial pyramid pooling layer. It takes an arbitrary input and gives a flexible solution with respect to size, aspect ratio, scales.

• Def pooling- Another type of pooling layer, that is sensitive to the distortion of visual patterns, known as def pooling is used for solving the ambiguity of deformation in computer vision.

3) Fully-connected layers: The final layer of CNN consists of 90% of the parameters. The feed forward network forms a vector of a particular length to follow up processing. Since these layers contain most of the parameters, there is a high computational burden while training the data.

B. Recurrent neural networks (RNNs)

Unlike CNN the Recurrent neural networks (RNNs) are generally used to deal with sequential data to represent time dependencies for various applications such as translating natural languages, music, time-series data, handwriting recognition, video processing etc. Deep convolutional RNN is a good candidate for frame prediction and object classification. As sequential data is trained using RNN, this network involves large number of complex parameters to train as compared to other neural networks. The architecture, optimization and training is complex in RNN model in order to train the sequential steps. As Hidden Markov Model depends upon the previous data it seems impractical to model the long time dependencies data. This problem is eliminated by RNN as it deals with the present data only with high accuracy. Now compared to other neural networks RNN gives good results for image captioning and analysis.

C. Restricted Boltzmann Machines(RBMs)

Minton et al. have in [11], proposed a generative random neural network technique called the Restricted Boltzmann machine. In RBM (a form of the Boltzmann Machine), the hidden and visible layer units are required to form a bipartite graph. As the RBM architecture is a double barreled graph, the hidden layer units H & the visible layer units V1 display conditional independence. Therefore,

P(HV1)=P(H1V1)P(H2V1)...P(HnV1)

In the above equation, given V1 as input, ’H’ can be derived using P(HV1). Likewise, V1 can be obtained through P(HV1). By altering the parameters,we can reduce the difference between V1 & V2, and resulting ’H’ will give us a fine lineament of V1. The practical implication along with detailed analysis is proposed by Minton [12]. It consists of flexible learning of an intensified gradient to resolve those problems. Because information passes via numerous layers of feature extractor, the model estimates the binary units corrupted by noise to conserve learned features. Many deep learning models [16] can be constructed by applying RBMs as learning models like Deep Boltzmmann Machine(DBMs), Deep Belief Networks(DBNs) and Deep Energy Models(DEMs).

Their description & their application to the field of Computer vision are respectively given as:

1) Deep Boltzmann Machines(DBMs): Like the RBM, the DBMs are also the part of the Boltzmann family. It consists of several layers of masked elements, where the odd and evennumbered layers are conditionally independent of each other. While training DBMs, all the supervised models are jointly trained. This type of training yields significant improvements in case of both likelihood [20] and classification. A major drawback of DBM is that the computational time requirement for approximate inference is comparatively greater than a Deep Belief Network, which in case of big datasets, renders the combined optimization of the parameters of Deep Boltzmann Machines, non-viable.

2) Deep Belief Networks(DBNs): Hinton et al. [17], introduced the deep belief network. This model is a probabilistic productive model that facilitates a collective expectation distribution over distinguishable labels and data. It is a layer-by-layer training approach which has the following benefits:

IJESMR
International Journal OF Engineering Sciences & Management Research

• Parameter selection is done for generation of appropriate initialization of the network, leading to a poor local optima till a certain level.
• Since it is a type of unsupervised learning, it makes a conclusion depending upon the clustered data without the need for any labelled data required for training. Although, it is a computationally exorbitant approach.

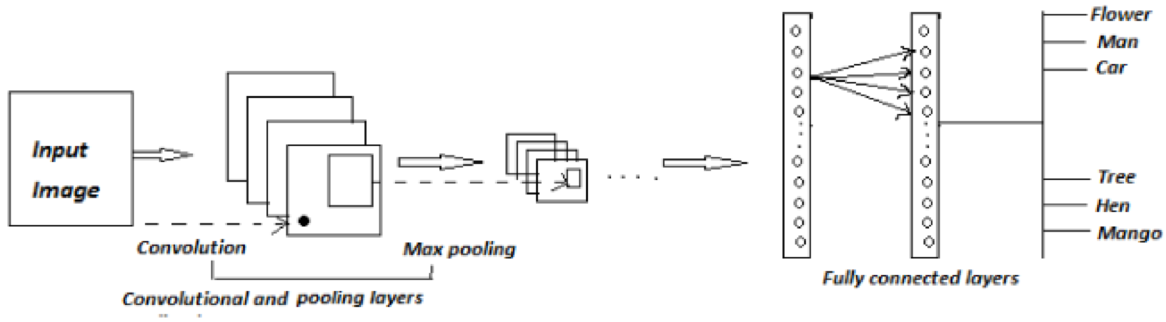


Fig. 1: Pipelined Architecture of Convolutional Neural Network

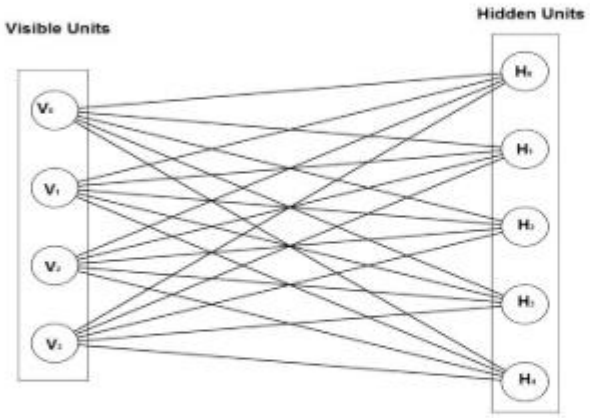


Fig. 2: Architecture of Restricted Boltzmann Machines

3) Deep Energy Models(DEMs): Ngiam et al. [19], introduced a deep Energy model, which is considered to be an approach to train the deep architectures. DEM is superior than DBMs and DBNs as it eliminates the issue of sharing the feature of having manifold random masked layers by only having one layer of random masked units. Thus DEM facilitates systematic training and a good conclusion about the data.

Although, CNNs are widely used over RBMs for many vision related tasks. Keeping this in mind, in order to shape both the local and global structure in face segmentation, Kae et al.[18], merged the CRF and the RBM, severely decreasing the error in face labelling.

D. Autoencoders
A notable variant of artificial neural networks, the autoencoder is used for training logical encodings [21]. The autoencoder learns to refurnish its own inputs instead of predicting any target value 'T' for given input 'S' by training the network accordingly. The learned feature is obtained by optimizing the reconstruction error and the correspondingcode which uses back propagation variant with same dimensionality. Some variants of autoencoders are presented as follows:
• Sparse autoencoder- It is used to [22,23] extract scattered features from input data. Sparse auto encoder is used. Sparse auto encoders are used as (i)For high dimension representation it is easy to divide it in different

IJESMR
International Journal OF Engineering Sciences & Management Research

classes depending upon their likelihood same as SVM. (ii)We can interpret the complex input data as we have many different classes depending upon their likelihood.(iii)Sparse encoders can be used for biological vision.

• De-noising autoencoder (DAE)- It is used to [24,25] amplify the resilience of the given model as it is able to work in the presence of random noise.

• Contractive autoencoder (CAE)- It is [26] the advanced version of DAE in which durability is achieved by adding contractive penalty to optimize reconstruction error used for unsupervised and transfer learning challenges.

E. Extreme Learning

A very recent topic in the vast areas of machine learning happens to be Extreme learning given by Guang-Bin Huang. It is a feed-forward neural network used for regression and classification tasks. It consists of a solitary layer of masked nodes in which the weights which are assigned as inputs to the masked nodes are random and are never corrected. In one step, the weights between the masked nodes and outputs are learned, which leads to learning of a linear model. These are better than networks trained by using back-propagation because of their faster learning ability and a good generalization capability.

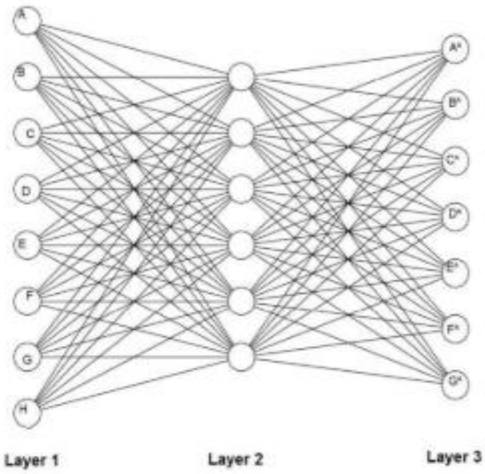


Fig. 3: Architecture of Autoencoder

Algorithm: Extreme learning Machine training algorithm learns a model of the following kind:

$\hat{Y} = W2\sigma(W1x)$

Where W1 is the matrix comprising of weights of the inputto-hidden layer, σ is an activation function and W2 is matrix of hidden to output layer weights. Given a design matrix 'X':

• W1 should be filled with Gaussian Random Noise.

• W2 should be estimated by least-square-fit to a matrix of response variable 'Y', with the pseudo inverse computed.

$W2=\sigma(W1X)+Y$

APPLICATIONS

Deep learning is a recent variant of artificial neural network. There are several potential applications of deep learning. Here we have presented some of the most notable applications related to image and video processing.

A. Image Classification

Depending upon the probability of its presence visual object class labeling can be done [27]. In deep learning most of the methods involve bags of visual words through which initially we get a histogram of quantized visual words then we proceed for classification. Most of the time sparse coding is used to recover the information loss.

B. Object Identification

ISSN 2349-6193
Impact Factor: 2.805

IJESMR

International Journal OF Engineering Sciences & Management Research

Object identification is somewhat related to image classification. In object identification process, it takes image as an input and then object estimation can be done depending upon the class and positions. The PASCAL VOC with 20 classes is often used for object identification. Deep learning (CNN based) is an advanced method after Deformable Part model (DPM) [28] for object identification. It is applicable for all scales unlike DPM. By using deep learning, we can divide the images into different classes using classifiers(SVM) and then check the presence of an object. Recurrent Convolutional Neural Networks(RCNNs) are used for fast object identification.

C. Image Retrieval

Image retrieval involves the collection of images having same object. Many CNN models are used and give good performance compared to traditional methods like VLAD and Fisher vector. Derived from the idea of Spatial Pyramid to extract patches for multiple scales Reverse SVM [29] is used. Without eliminating any spatial information it takes the complete image and then divides it into different scales. Using CNN models a large dataset can be trained and used for feature extraction and can also be applied for image retrieval to achieve better accuracy.

D. Semantic Segmentation

CNN models are used for semantic segmentation tasks, as it is potent of handling the pixel-level predictions. Output masks having a 2-dimensional spatial spread are required by semantic segmentation. The process of semantic segmentation is outlined below:

• Detection based Segmentation: This approach divides the images depending on object detection [30,31,32,33]. Recurrent Convolutional Neural Network along with simultaneous detection and segmentation generated the proposal for object identification and then applying traditional approach to divide the regions and to assign the pixels as activation vectors, which resulted in large improvements. Disadvantage of this approach is additional expense for object identification. Convolutional feature masking is an approach in which we proposals are personally obtained using the feature maps.

• FCN-CRFs Based Segmentation: In this approach, the fully connected layer is replaced by the fully convolutional layers. It is a commonly used technique for semantic segmentation [34,35]. A similar model is proposed by FCN integrated with conditional random fields(CRFs) into FCN [36] for elaborate boundary retrieval.

• Weakly supervised annotations: Other than improving the segmentation model, Weakly supervised segmentation is another area of interest for researchers [37]. Here it takes the segmentation of training data depending upon which it estimates the segmentation masks to update the neural network. Then it combines all small pixel levels so that we can get the complete image.

CONCLUSION

In this paper we have given an survey of Deep learning and its recent development. The analysis of prevailing deep learning architectures is done by developing a categorical layout. Deep learning algorithms are divided into three categories: Convolutional Neural Network, Restricted Boltzmann Machines, Autoencoder. Apart from that, RNN and extreme learning are also quite popular. In this paper we mainly dealt with the recent advancement of CNN dependent strategies, since it is mostly used for images. In some recent papers it has been reported CNN dependent strategies have surpassed the precision of human testers. In spite of the many researches reported till now, there is a large scope for additional advancements. As an example in theoretical foundation, it has not yet been shown that when and under what conditions they will perform well and an optimal structure for certain tasks. This paper summarizes and gives an idea to the new researchers to explore more in the vast yet young area of Deep learning.

REFERENCES

1. **Krizhevsky, A., Sutskever, I., & Hinton, G. E.** (2012). Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, 25, 1097-1105.
2. **Simonyan, K., & Zisserman, A.** (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.
3. **He, K., Zhang, X., Ren, S., & Sun, J.** (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).

ISSN 2349-6193
Impact Factor: 2.805

IJESMR

International Journal OF Engineering Sciences & Management Research

4. **Redmon, J., Divvala, S., Girshick, R., & Farhadi, A.** (2016). You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 779-788).
5. **Long, J., Shelhamer, E., & Darrell, T.** (2015). Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 3431-3440).
6. **Girshick, R., Donahue, J., Darrell, T., & Malik, J.** (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 580-587).
7. **Ren, S., He, K., Girshick, R., & Sun, J.** (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in Neural Information Processing Systems*, 28, 91-99.
8. **Goodfellow, I. J., Shlens, J., & Szegedy, C.** (2015). Explaining and harnessing adversarial examples. *arXiv preprint arXiv:1412.6572*.
9. **Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., ... & Houlsby, N.** (2021). An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*.
10. **Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q.** (2017). Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4700-4708).
11. **Chen, L. C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. L.** (2017). Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(4), 834-848.
12. **Radford, A., Metz, L., & Chintala, S.** (2016). Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*.
13. **Ronneberger, O., Fischer, P., & Brox, T.** (2015). U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* (pp. 234-241). Springer, Cham.
14. **Zhou, Z., Siddiquee, M. M. R., Tajbakhsh, N., & Liang, J.** (2018). Unet++: A nested u-net architecture for medical image segmentation. *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, 3-11.
15. **Lin, T. Y., Goyal, P., Girshick, R., He, K., & Dollar, P.** (2017). Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision* (pp. 2980-2988).
16. **Chen, X., Fang, H., Lin, T. Y., Vedantam, R., Gupta, S., Dollár, P., & Zitnick, C. L.** (2015). Microsoft coco captions: Data collection and evaluation server. *arXiv preprint arXiv:1504.00325*.
17. **Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., ... & Schiele, B.** (2016). The cityscapes dataset for semantic urban scene understanding. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 3213-3223).
18. **Paszke, A., Chaurasia, A., Kim, S., & Culurciello, E.** (2016). Enet: A deep neural network architecture for real-time semantic segmentation. *arXiv preprint arXiv:1606.02147*.
19. **Kingma, D. P., & Ba, J.** (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.
20. **Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., ... & Rabinovich, A.** (2015). Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1-9).
21. **Lecun, Y., Bengio, Y., & Hinton, G.** (2015). Deep learning. *Nature*, 521(7553), 436-444.
22. **Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L.** (2009). Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition* (pp. 248-255). Ieee.
23. **Jaderberg, M., Simonyan, K., Zisserman, A., & Kavukcuoglu, K.** (2015). Spatial transformer networks. *Advances in Neural Information Processing Systems*, 28.
24. **Chen, L. C., Papandreou, G., Schroff, F., & Adam, H.** (2017). Rethinking atrous convolution for semantic image segmentation. *arXiv preprint arXiv:1706.05587*.
25. **Zoph, B., Vasudevan, V., Shlens, J., & Le, Q. V.** (2018). Learning transferable architectures for scalable image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 8697-8710).