**Report**

**Abstract:**

The Inception network was a crucial milestone in the development of CNN Image classifiers. Prior to this architecture, most popular CNNs or the classifiers just used stacked convolution layers deeper and deeper to obtain better performance.

The Inception network, on the other hand, was heavily engineered and very much deep and complex. It used many different techniques to push its performance; both in terms of speed as well as accuracy.

Deep learning architecture is rapidly gaining steam as more and more efficient architectures emerge from research papers emerge from around the world. These research papers not only contain a ton of information but also provide a new way to the birth of new Deep learning architectures, they can often be difficult to parse through. And to understand these papers, one might have to go through that paper multiple times and perhaps even other dependent papers. Well! Inception is one of them.

**Introduction:**

Inception Network (ResNet) is one of the well-known deep learning models that was introduced by Christian Szegedy, Wei Liu, Yangqing Jia. Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich in their paper “Going deeper with convolutions” [1]in 2014.

Later the different versions of the Inception network are evolved. That was from Sergey Ioffe, Christian Szegedy, Jonathon Shlens, Vincent Vanhouck, and Zbigniew Wojna in their paper named “Rethinking the Inception Architecture for Computer Vision” [2] in 2015. The Inception model is categorized as one of the popular and most used deep learning models.

1. **Project Aims**

The proposal of few general design principles and optimization techniques proved to be useful for efficiently scaling up convolution networks.

Early in the network architecture, avoid representational bottlenecks.

The network will learn faster if it has more different filters which will have more different feature maps.

Spatial aggregation that is dimension reduction can be done over lower dimensional embeddings without much loss in representational power.

With the balance between width and depth, optimal performance of the network can be achieved.

1. **Project Objectves**

Process visual/spatial information at various scales and then aggregate

This is a bit optimistic, computationally

5×5 convolutions are especially expensive

1. **Background**

Using the inception module that is dimension-reduced inception module, a deep neural network architecture was built (Inception v1).

**Review of Literature:**

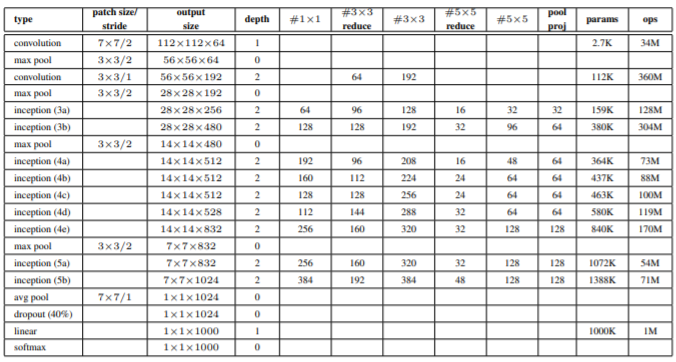
The basic idea of early traditional oracle bone inscription character recognition methods is to preprocess data first, then manually extract features around graph theory and topology, and then encode the features for matching and recognition. Li and Zhou et al. [2] regarded oracle bone inscription characters as an undirected graph composed of lines and points, so they extracted multilevel graph features based on graph theory, then recognized and classified them. Li et al. [3] abstracted oracle bone inscription characters and classified characters based on the graph isomorphism determination algorithm. S. Gu [4] considered that the topology of oracle bone inscription characters is more stable to some extent, and he used the minimum distance to judge the equivalence relationship of the topological structure coding of characters. These methods mainly focused on the font characteristics of oracle bone inscription characters and achieved meaningful results, but simple graph-theoretical characterizations with manual encoding are prone to underfitting in the case of a large number of data.

Many researchers have already implemented oracle bone inscription character recognition by neural networks and deep learning and achieved excellent results. Deep learning-based character recognition is supervised. It requires a large number of training data to enable deep neural networks to learn different patterns of oracle bone inscription characters and thus achieve automatic recognition of single character images. Lv et al. [5] proposed a curvature histogram-based Fourier descriptor to extract glyph features and then input the features to the classical support vector machine (SVM) model [6] for glyph classification. Guo et al. [7] proposed a multilevel oracle bone character representation method, combining sparse self-coding-based middle-level representation features and Gabor-based low-level representation features to describe oracle bone characters. Gao et al. [8] proposed a recognition method based on the Hopfield neural network with an analysis of the context for the recognition of fuzzy characters. Yongge Liu [9] extracted features by chunked histogram and introduced the classical SVM as their model for oracle bone inscription character classification. Meng et al. [10] extracted topological features by Hough transform as well as clustering and achieved recognition by calculating the distance between the actual image and the standard image. Liu et al. [11] created a convolutional neural network based on the classical SqueezeNet for the recognition of incomplete characters at the edges of oracle bones. Sun et al. [12] proposed a dual-view oracle bone character recognition system combining temporal-spatial psychovisual modulation (TPVM) and the character recognition algorithm. Zhang et al. [13] adopted an improved Siamese network to learn the similarity between an oracle bone character and the corresponding template typeset images. Fujikawa et al. [14] proposed a two-stage method that adopts the latest You Only Look Once (YOLO) model and MobileNet for character recognition. These methods introduced neural networks and deep learning, which make the model obtain a better ability of feature representation, so the accuracy of character recognition is improved significantly.

**Project Methods:**

Inception architecture uses the CNN blocks multiple times with different filters like 1×1, 3×3, 5×5, etc., so let us create a class for CNN block, which takes input channels and output channels along with batchnorm2d and ReLu activation.

Then create a class for inception module with dimension reduction, refer the figure above, which shows that is output from 1×1 filter, reduction 3×3, then output from 3×3 filter, reduction 5×5, then output from 5×5 and the out from 1×1 pool.



The original convolution block performs batch normalization after convolution, and then, the result is passed to the next layer by the activation function. We introduce an Inverted Bottleneck [18], which transforms a single convolution layer into a depthwise convolution layer and two pointwise convolution layers. We also refer to the property of the Multilayer Perceptron (MLP) block in the Transformer [19] that the hidden dimension is four times larger than the input dimension and design the pointwise convolution layer by setting the dimension of the middle hidden layer to four times the size of the input. At the same time, we set the size of the depthwise convolution layer to 7 × 7 to improve the accuracy. Since our operations complicate a simple convolution layer into a 3-layer convolution and a summation, we replace the batch normalization with a simpler layer normalization [20] to reduce the complexity. Figure [3](https://www.hindawi.com/journals/sp/2022/7490363/fig3/) shows the structure of the proposed new convolution block.

In the choice of activation function, we replace rectified linear unit (ReLU) with Gaussian error linear unit (GELU). The activation introduces the idea of random regularity, which is a probabilistic description of the neuronal input and is intuitively more in line with the natural understanding, and the experiment result is exactly better than ReLU.

**Project Activities:**

We introduce the Contextual Transformer (CoT) block [21], a novel Transformer-style module for visual recognition. There are a large number of heterogeneous characters in the oracle bone inscription, but they usually have the same features in some close positions. Convolution has no way to represent feature interactions at different spatial locations excellently, and a Transformer-style module is needed to improve this deficiency [22]. However, the conventional self-attention block ignores the rich contextual information among the nearest neighbors [23], while the CoT block can encode the context of the input keys by 3 × 3 convolution, which produces a static contextual representation of the input and better extracts the nearest-neighbor features. We place it after the initial 3-layer convolution block for feature learning of the nearest neighbor space. Figure [4](https://www.hindawi.com/journals/sp/2022/7490363/fig4/) shows the structure of the Contextual Transformer block.

**Design:**

We introduce the Convolutional Block Attention Module (CBAM) [24], which contains two modules of attention mechanism, the Channel Attention Module and the Spatial Attention Module.

The Channel Attention Module mainly processes the feature maps of different channels and tells the model in which maps should be paid more attention to. It first performs global MaxPooling or AveragePooling [25] for the feature maps on different channels and obtains a MaxPool Channel Attention vector and an AvgPool Channel Attention vector. Then input these two vectors into a weight-sharing multilayer perceptron (MLP) [26] with only one hidden layer to get two processed channel attention vectors. Finally, these two vectors are processed by element-wise summation and Sigmoid function, multiplied by the original feature map, and get a new feature map. It can be described as

The Spatial Attention Module mainly processes the feature regions on the feature maps and tells the model in which regions of the feature maps should be paid more attention to. It performs global MaxPooling and AvgPooling on the pixel values at the same locations on different feature maps in the axis direction, obtains two spatial attention maps, and concatenates them [27]. Then, the feature map passes through a 7 × 7 convolution and a sigmoid activation function to get a spatial attention matrix with the same dimension as the original feature map. Finally, the spatial attention matrix is multiplied by the original feature map and then outputs a new feature map. It can be described as

The CBAM block will derive the attention map by two different dimensions sequentially and then multiply the input feature map by the attention map to achieve adaptive feature optimization. Figure [5](https://www.hindawi.com/journals/sp/2022/7490363/fig5/) shows the structure of the CBAM. Using the attention mechanism, we can make our model more focused on the essential features and suppress unnecessary features [28]. We add CBAM blocks after the convolution layers in Inception-A, Inception-B, and Inception-D for extracting features in channel dimensions and spatial dimensions to improve the accuracy.

**Results:**

We divide the two oracle character image datasets into a training set and a test set in the ratio of 7 : 3, respectively, and then preprocess them after loading. Data loader loads the images and randomly crops them to different sizes and aspect ratios, resizes them to 299 × 299, and randomly rotates the images horizontally [29]. It enhances the diversity of the dataset, simulates the oracle characters that appear in different situations, and tests the robustness of the model.

We choose AdamW [30] as the optimizer, which can effectively improve the generalization performance and better avoid the parameter overfitting problem by decoupling the weight decay from the gradient update. The loss function we choose is cross-entropy loss function [31], and the cross-entropy describes the closeness of the actual output to the expected output. The smaller the cross-entropy is, the smaller the closeness is. The formula of the cross-entropy is as follows:where *p* is the expected output and *q* is the actual output. There are three model accuracy evaluation indicators, Top-1 Accuracy, Top-3 Accuracy, and Top-5 Accuracy. They refer to the probability that the expected result is among the top *n* in the classification ranking of the actual output, and Top-1 Accuracy is the probability of complete correct identification.

Our experimental environment is PyTorch 1.10.0, Python 3.8, and Cuda 11.3, and the hardware configuration is Intel(R) Xeon(R) Gold 5320 @ 2.20 GHz CPU and 16 GB NVIDIA RTX A4000 GPU. The learning rate is set to decrease as the epoch increases so that the objective function is fast enough to reach the local optimum. We plot the change curves of Top-1 Accuracy and Loss during the experimental process of the OBC306 dataset in Figures [6](https://www.hindawi.com/journals/sp/2022/7490363/fig6/) and [7](https://www.hindawi.com/journals/sp/2022/7490363/fig7/). They indicate that Top-1 Accuracy gradually increases and loss gradually decreases, both converge after the 30th epoch, and there is no overfitting, which preliminarily proves the effectiveness of our model. Figure [8](https://www.hindawi.com/journals/sp/2022/7490363/fig8/) shows the recognition results of some characters in the OBC306, which indicates that our model performs excellently on some character images that are difficult to recognize.

To further demonstrate the excellence of our model, we chose Inception-v3, AlexNet [32], and VGG-19 [33] neural network models to conduct experiments under the same conditions and compared the results.

AlexNet innovatively applies rectified linear unit as the activation function and uses dropout to randomly ignore a portion of neurons during training to avoid overfitting. It also proposes a local response normalization (LRN) layer to create a competition mechanism for the activity of local neurons. It suppresses other neurons with smaller feedback, and the values with larger responses become larger, enhancing the generalization ability of the model, which is suitable for characters recognition.

VGG-19 has thirteen convolutional layers and three fully connected layers, each of which further extracts more complex features, so each layer can be seen as an extractor of multiple local features. It has smaller convolutional kernels (3 × 3) and smaller pooling kernels (2 × 2), which enhance the depth of the network and improve the recognition accuracy while ensuring that it has the same receptive field. VGG-19 is widely used in the field of image feature extraction and recognition.

The results on the two datasets are shown in Tables [1](https://www.hindawi.com/journals/sp/2022/7490363/tab1/) and [2](https://www.hindawi.com/journals/sp/2022/7490363/tab2/), which indicate that our improved model outperforms other models on both datasets, reaching 98.171% Top-1, 99.837% Top-3, and 99.844% Top-5 Accuracy on HWOBC, and 87.732% Top-1, 94.847% Top-3, and 96.322% Top-5 Accuracy on OBC306. It proves the correctness and superiority of our proposed model.

**Conclusion:**

We propose an improved neural network model based on Inception-v3 for oracle bone inscription character recognition. We replace the original simple convolution layer with our new convolution block. We introduce the Contextual Transformer block between the convolution and pooling layers to improve the recognition ability of nearest-neighbor spatial features. We add the Convolutional Block Attention Module to three main inception modules to enhance the performance of character recognition. We apply the improved model to two oracle bone inscription character image datasets for character recognition and compare it with AlexNet, VGG-19, and Inception-v3 neural network models. The results indicate that our model achieves the best performance in both, reaching 98.171% Top-1, 99.837% Top-3, and 99.844% Top-5 Accuracy on dataset HWOBC, and 87.732% Top-1, 94.847% Top-3, and 96.322% Top-5 Accuracy on dataset OBC306, which prove the correctness and excellence of our proposed model.

For further work in the future, we plan to improve the model for those characters that rarely appear. Because the number of sample images of these characters is small and their glyphs are usually complex, our model cannot be trained sufficiently to recognize them correctly. We will also try to improve the accuracy of recognition under noise interference as much as possible, such as adopting effective image preprocessing methods to reduce the effect of noise.

**References:**

1. C. Cheung, “The Chinese history that is written in bone,” *American Scientist*, vol. 106, no. 3, pp. 133–135, 2018.View at: [Publisher Site](https://doi.org/10.1511/2018.106.3.133) | [Google Scholar](https://scholar.google.com/scholar_lookup?title=The%20Chinese%20history%20that%20is%20written%20in%20bone&author=C.%20Cheung&publication_year=2018)
2. F. Li and X. Zhou, “A graph-theoretic approach to automatic oracle recognition,” *Journal of Electronics and Information*, pp. 41–47, 1996.View at: [Google Scholar](https://scholar.google.com/scholar_lookup?title=A%20graph-theoretic%20approach%20to%20automatic%20oracle%20recognition&author=F.%20Li%20&author=X.%20Zhou&publication_year=1996)
3. Q. S. Li, Y. X. Yang, and A. M. Wang, “A graph isomorphism approach for oracle recognition,” *Computer Engineering and Applications*, vol. 47, no. 8, pp. 112–114, 2011.View at: [Google Scholar](https://scholar.google.com/scholar_lookup?title=A%20graph%20isomorphism%20approach%20for%20oracle%20recognition&author=Q.%20S.%20Li&author=Y.%20X.%20Yang&author=&author=A.%20M.%20Wang&publication_year=2011)
4. S. T. Gu, “A topological alignment-based method for oracle glyph recognition,” *Computers and Digital Engineering*, vol. 44, no. 10, pp. 2001–2006, 2016.View at: [Google Scholar](https://scholar.google.com/scholar_lookup?title=A%20topological%20alignment-based%20method%20for%20oracle%20glyph%20recognition&author=S.%20T.%20Gu&publication_year=2016)
5. X. Q. Lv, M. N. Li, K. W. Cai, X. Wang, and Y.-M. Tang, “A graph recognition-based classification method for oracle bones,” *Journal of Beijing University of Information Science and Technology*, pp. 92–96, 2010.View at: [Google Scholar](https://scholar.google.com/scholar_lookup?title=A%20graph%20recognition-based%20classification%20method%20for%20oracle%20bones&author=X.%20Q.%20Lv&author=M.%20N.%20Li&author=K.%20W.%20Cai&author=X.%20Wang&author=&author=Y.-M.%20Tang&publication_year=2010)
6. T. Evgeniou and M. Pontil, “Support vector machines: theory and applications,” *Advanced Course on Artificial Intelligence*, Springer, Berlin, Heidelberg, pp. 249–257, 1999.View at: [Google Scholar](https://scholar.google.com/scholar_lookup?title=Support%20vector%20machines:%20theory%20and%20applications&author=T.%20Evgeniou%20&author=M.%20Pontil&publication_year=1999)
7. J. Guo, C. Wang, E. Roman-Rangel, and H. Y. Chao, “Building hierarchical representations for oracle character and sketch recognition,” *IEEE Transactions on Image Processing*, vol. 25, no. 1, pp. 104–118, 2016.View at: [Publisher Site](https://doi.org/10.1109/tip.2015.2500019) | [Google Scholar](https://scholar.google.com/scholar_lookup?title=Building%20hierarchical%20representations%20for%20oracle%20character%20and%20sketch%20recognition&author=J.%20Guo&author=C.%20Wang&author=E.%20Roman-Rangel&author=&author=H.%20Y.%20Chao&publication_year=2016)
8. F. Gao, Q. Wu, Y. Liu, and X. Jing, “A semantic building block-based method for recognizing fuzzy oracle glyphs,” *Science Technology and Engineering*, vol. 30, pp. 67–70, 2014.View at: [Google Scholar](https://scholar.google.com/scholar_lookup?title=A%20semantic%20building%20block-based%20method%20for%20recognizing%20fuzzy%20oracle%20glyphs&author=F.%20Gao&author=Q.%20Wu&author=Y.%20Liu&author=&author=X.%20Jing&publication_year=2014)
9. Y. G. Liu and G. Y. Liu, “SVM-based oracle bone character recognition,” *Journal of Anyang Normal College*, vol. 2, pp. 54–56, 2017.View at: [Publisher Site](https://doi.org/10.23977/jeis.2017.22012) | [Google Scholar](https://scholar.google.com/scholar_lookup?title=SVM-based%20oracle%20bone%20character%20recognition&author=Y.%20G.%20Liu%20&author=G.%20Y.%20Liu&publication_year=2017)
10. L. Meng, “Recognition of oracle bone inscriptions by extracting line features on image processing,” in *Proceedings of the 6th International Conference on Pattern Recognition Applications and Methods*, pp. 606–611, Porto, Portugal, Febraury 2017.View at: [Publisher Site](https://doi.org/10.5220/0006225706060611) | [Google Scholar](https://scholar.google.com/scholar_lookup?title=Recognition%20of%20oracle%20bone%20inscriptions%20by%20extracting%20line%20features%20on%20image%20processing&author=L.%20Meng)
11. M. Liu, G. Liu, G. Liu, Y. Liu, and Q. Jiao, “Oracle bone inscriptions recognition based on deep convolutional neural network,” *Journal of Image and Graphics*, vol. 8, no. 4, pp. 114–119, 2020.View at: [Publisher Site](https://doi.org/10.18178/joig.8.4.114-119) | [Google Scholar](https://scholar.google.com/scholar_lookup?title=Oracle%20bone%20inscriptions%20recognition%20based%20on%20deep%20convolutional%20neural%20network&author=M.%20Liu&author=G.%20Liu&author=G.%20Liu&author=Y.%20Liu&author=&author=Q.%20Jiao&publication_year=2020)
12. W. Sun, G. Zhai, and Z. Gao, “Dual-view oracle bone script recognition system via temporal-spatial psychovisual modulation,” in *Proceedings of the 2020 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)*, pp. 193–198, IEEE, Guangdong, China, August 2020.View at: [Publisher Site](https://doi.org/10.1109/mipr49039.2020.00047) | [Google Scholar](https://scholar.google.com/scholar_lookup?title=Dual-view%20oracle%20bone%20script%20recognition%20system%20via%20temporal-spatial%20psychovisual%20modulation&author=W.%20Sun&author=G.%20Zhai&author=&author=Z.%20Gao)
13. C. Zhang, R. Zong, S. Cao, M. Yi, and B. Mo, “AI-powered oracle bone inscriptions recognition and fragments rejoining,” in *Proceedings of the Twenty-Ninth International Conference on International Joint Conferences on Artificial Intelligence*, pp. 5309–5311, Yokohama, Japan, January 2021.View at: [Google Scholar](https://scholar.google.com/scholar_lookup?title=AI-powered%20oracle%20bone%20inscriptions%20recognition%20and%20fragments%20rejoining&author=C.%20Zhang&author=R.%20Zong&author=S.%20Cao&author=M.%20Yi&author=&author=B.%20Mo)
14. Y. Fujikawa, H. Li, X. Yue, and C. V. G. A. L. Aravinda, “Recognition of oracle bone inscriptions by using two deep learning models,” *International Journal of Dental Hygiene*, 2022.View at: [Publisher Site](https://doi.org/10.1007/s42803-022-00044-9) | [Google Scholar](https://scholar.google.com/scholar_lookup?title=Recognition%20of%20oracle%20bone%20inscriptions%20by%20using%20two%20deep%20learning%20models&author=Y.%20Fujikawa&author=H.%20Li&author=X.%20Yue&author=&author=C.%20V.%20G.%20A.%20L.%20Aravinda&publication_year=2022)
15. B. Li, Q. Dai, F. Gao, W. Zhu, Q. Li, and Y. Liu, “HWOBC-a handwriting oracle bone character recognition database,” *Journal of Physics: Conference Series*, IOP Publishing, vol. 1651, no. 1, Article ID 012050, 2020.View at: [Publisher Site](https://doi.org/10.1088/1742-6596/1651/1/012050) | [Google Scholar](https://scholar.google.com/scholar_lookup?title=HWOBC-a%20handwriting%20oracle%20bone%20character%20recognition%20database&author=B.%20Li&author=Q.%20Dai&author=F.%20Gao&author=W.%20Zhu&author=Q.%20Li&author=&author=Y.%20Liu&publication_year=2020)
16. S. Huang, H. Wang, Y. Liu, X. Shi, and L. Jin, “Obc306: a large-scale oracle bone character recognition dataset,” in *Proceedings of the International Conference on Document Analysis and Recognition (ICDAR)*, pp. 681–688, IEEE, Australia, September 2019.View at: [Publisher Site](https://doi.org/10.1109/icdar.2019.00114) | [Google Scholar](https://scholar.google.com/scholar_lookup?title=Obc306:%20a%20large-scale%20oracle%20bone%20character%20recognition%20dataset&author=S.%20Huang&author=H.%20Wang&author=Y.%20Liu&author=X.%20Shi&author=&author=L.%20Jin)
17. S. Christian, V. Vincent, I. Sergey, J. Shlens, and Z. Wojna, “Rethinking the inception architecture for computer vision,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2818–2826, Nevada, ND, USA, June 2016.View at: [Google Scholar](https://scholar.google.com/scholar_lookup?title=Rethinking%20the%20inception%20architecture%20for%20computer%20vision&author=S.%20Christian&author=V.%20Vincent&author=I.%20Sergey&author=J.%20Shlens&author=&author=Z.%20Wojna)
18. Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, 2022, <https://arxiv.org/pdf/2201.03545>.
19. V. Ashish, N. Shazeer, N. Parmar et al., “Attention is all you need,” *Advances in Neural Information Processing Systems*, vol. 30, 2017.View at: [Google Scholar](https://scholar.google.com/scholar_lookup?title=Attention%20is%20all%20you%20need&author=V.%20Ashish&author=N.%20Shazeer&author=N.%20Parmar%20et%20al.&publication_year=2017)
20. J. Gu, Z. Wang, J. Kuen, and L. A. B. T. X. G. J. T. Ma, “Recent advances in convolutional neural networks,” *Pattern Recognition*, vol. 77, pp. 354–377, 2018.View at: [Publisher Site](https://doi.org/10.1016/j.patcog.2017.10.013) | [Google Scholar](https://scholar.google.com/scholar_lookup?title=Recent%20advances%20in%20convolutional%20neural%20networks&author=J.%20Gu&author=Z.%20Wang&author=J.%20Kuen&author=&author=L.%20A.%20B.%20T.%20X.%20G.%20J.%20T.%20Ma&publication_year=2018)
21. Y. Li, T. Yao, Y. Pan, and T. Mei, “Contextual transformer networks for visual recognition,” 2021, <https://arxiv.org/pdf/2107.12292>.View at: [Google Scholar](https://scholar.google.com/scholar_lookup?title=Contextual%20transformer%20networks%20for%20visual%20recognition&author=Y.%20Li&author=T.%20Yao&author=Y.%20Pan&author=&author=T.%20Mei&publication_year=2021)
22. C. Djork-Arné, T. Unterthiner, and S. Hochreiter, “Fast and accurate deep network learning by exponential linear units (elus),” 2015, <https://arxiv.org/pdf/1511.07289>.View at: [Google Scholar](https://scholar.google.com/scholar_lookup?title=Fast%20and%20accurate%20deep%20network%20learning%20by%20exponential%20linear%20units%20(elus)&author=C.%20Djork-Arn%C3%A9&author=T.%20Unterthiner&author=&author=S.%20Hochreiter&publication_year=2015)
23. E. Hinton Geoffrey, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov, “Improving neural networks by preventing co-adaptation of feature detectors,” 2012, <https://arxiv.org/pdf/1207.0580>.View at: [Google Scholar](https://scholar.google.com/scholar_lookup?title=Improving%20neural%20networks%20by%20preventing%20co-adaptation%20of%20feature%20detectors&author=E.%20Hinton%20Geoffrey&author=N.%20Srivastava&author=A.%20Krizhevsky&author=I.%20Sutskever&author=&author=R.%20R.%20Salakhutdinov&publication_year=2012)
24. S. Woo, J. Park, and J.-Y. I. S. Lee, “Cbam: convolutional block attention module,” *Computer Vision - ECCV 2018*, pp. 3–19, 2018.View at: [Publisher Site](https://doi.org/10.1007/978-3-030-01234-2_1) | [Google Scholar](https://scholar.google.com/scholar_lookup?title=Cbam:%20convolutional%20block%20attention%20module&author=S.%20Woo&author=J.%20Park&author=&author=J.-Y.%20I.%20S.%20Lee&publication_year=2018)
25. F. Kunihiko and S. Miyake, “Neocognitron: a self-organizing neural network model for a mechanism of visual pattern recognition,” *Competition and Cooperation in Neural Nets*, Springer, Berlin, Heidelberg, pp. 267–285, 1982.View at: [Google Scholar](https://scholar.google.com/scholar_lookup?title=Neocognitron:%20a%20self-organizing%20neural%20network%20model%20for%20a%20mechanism%20of%20visual%20pattern%20recognition&author=F.%20Kunihiko%20&author=S.%20Miyake&publication_year=1982)
26. Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.View at: [Publisher Site](https://doi.org/10.1109/5.726791) | [Google Scholar](https://scholar.google.com/scholar_lookup?title=Gradient-based%20learning%20applied%20to%20document%20recognition&author=Y.%20Lecun&author=L.%20Bottou&author=Y.%20Bengio&author=&author=P.%20Haffner&publication_year=1998)
27. K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, Las Vegas, NV, USA, July 2016.View at: [Publisher Site](https://doi.org/10.1109/cvpr.2016.90) | [Google Scholar](https://scholar.google.com/scholar_lookup?title=Deep%20residual%20learning%20for%20image%20recognition&author=K.%20He&author=X.%20Zhang&author=S.%20Ren&author=&author=J.%20Sun)
28. K. H. Kim, S. Hong, B. Roh, Y. Cheon, and M. Park, “Pvanet: deep but lightweight neural networks for real-time object detection,” 2016, <https://arxiv.org/abs/1608.08021>.View at: [Google Scholar](https://scholar.google.com/scholar_lookup?title=Pvanet:%20deep%20but%20lightweight%20neural%20networks%20for%20real-time%20object%20detection&author=K.%20H.%20Kim&author=S.%20Hong&author=B.%20Roh&author=Y.%20Cheon&author=&author=M.%20Park&publication_year=2016)
29. K. Hara, D. Saitoh, and H. Shouno, “Analysis of dropout learning regarded as ensemble learning,” in *Proceedings of the ICANN 2016, International Conference on Artificial Neural Networks*, pp. 72–79, Springer, Barcelona, Sepetember 2016.View at: [Publisher Site](https://doi.org/10.1007/978-3-319-44781-0_9) | [Google Scholar](https://scholar.google.com/scholar_lookup?title=Analysis%20of%20dropout%20learning%20regarded%20as%20ensemble%20learning&author=K.%20Hara&author=D.%20Saitoh&author=&author=H.%20Shouno)
30. D. P. Kingma and B. Jimmy, “Adam: a method for stochastic optimization,” 2014, <https://arxiv.org/abs/1412.6980>.View at: [Google Scholar](https://scholar.google.com/scholar_lookup?title=Adam:%20a%20method%20for%20stochastic%20optimization&author=D.%20P.%20Kingma%20&author=B.%20Jimmy&publication_year=2014)
31. Z. Zhang and M. Sabuncu, “Generalized cross entropy loss for training deep neural networks with noisy labels,” *Advances in Neural Information Processing Systems*, vol. 31, 2018.View at: [Google Scholar](https://scholar.google.com/scholar_lookup?title=Generalized%20cross%20entropy%20loss%20for%20training%20deep%20neural%20networks%20with%20noisy%20labels&author=Z.%20Zhang%20&author=M.%20Sabuncu&publication_year=2018)
    1. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” *Advances in Neural Information Processing Systems*, vol. 25, 2012.View at: [Google Scholar](https://scholar.google.com/scholar_lookup?title=Imagenet%20classification%20with%20deep%20convolutional%20neural%20networks&author=A.%20Krizhevsky&author=I.%20Sutskever&author=&author=G.%20E.%20Hinton&publication_year=2012)
32. K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” 2014, <https://arxiv.org/abs/1409.1556>.View at: Google Scholar.