International Journal of Emerging Trends in Science and Technology
7(1) June 2021, 25-30
http://www.publishingindia.com/ijetst

Architectural Structures in Convolutional Neural Network for Person Re-Identification

Elankeerthana R.[1*] and Vinotha R.[2]
[1]Assistant Professor, Department of Information Technology, M.Kumarasamy College of Engineering, Karur, Tamil Nadu, India. Email: elankeerthanar.it@mkce.ac.in
[2]Assistant Professor, Department of Information Technology, M.Kumarasamy College of Engineering, Karur, Tamil Nadu, India. Email: vinothar.it@mkce.ac.in
*Corresponding Author

Abstract: Convolutional network in deep learning algorithm has many architectural structures for the person re-identification to increase the network accuracy. Observation of the same person can be matched in different occasions like time, cameras is the Person recognition on appearance based classification. Many years researchers on computer vision realized Person re-identification has been tricky problem which video includes frame taken at different place, pose, condition, lighting condition, camera, occlusions, background and appearance. Here we implement different architectural structure in convolutional network in our dataset to give different accuracy and different error rate to analyses the best structural to recognize the person in different cameras. Tracking and finding a person deals with the security issue where many places like airports, streets, colleges, shopping malls, theaters and many other public places for identification of fraud cases.

Keywords: Convolutional neural network, Deep learning, Architectural Structure (LeNet, AlexNet, VGGNet, GoogLeNet, ResNet, ResNext, DenseNet), Person Recognition.

I. INTRODUCTION

In video tracking and Detection of person becomes an attractive purpose such as activities of recognition and monitoring of victims [1], surveillance on various places [2], in sports [3], etc. System which receives images and videos as input from multi cameras and toning different volumes of appearance between the people in image and video as the output is named as person re-identification. This task will make more complexity when increasing of number of cameras, due to multi cameras numerous people appearing in views point. A camera with views of non-overlapping is the issue in multi camera tracking. Extraction of feature and matching images are the two components in person re-identification. Biometric face or iris identification in method knows as traditional recognition cannot use due to squat resolution in surveillance camera. In person matching, using color and texture to improve the appearance feature for reliability and robustness because change of appearance is the most demanding trouble in mismatching.

Convolutional Neural Network (CNN) in Deep learning method is most successful method which is multilayer neural network with convolutional filter to transform the input image by processing to classify under convinced categories to some extent. There are classic and modern architectural structure are present in CNN for image analysis. LeNet, AlexNet, VGGNet are the classic architectural structure and GoogLeNet, ResNet, ResNext, DenseNet are the modern architectural structure in CNN. ILSVRC means ImageNet Large Scale Visual Recognition Challenge which the competition since 2010 is used to attain the highest accuracy in the network by implement the architectural structure of convolutional neural network (CNN) in given 14,197,122 of total images in data set.

LeNet is the first architectural structure implemented in CNN and successfully used on handwritten digits for zip code recognition in 1998. It is the first structure which gets the attention towards deep learning. **AlexNet** won the first place on ImageNet Challenge in 2012 which is similar to LeNet but it is larger, where networks are shallow in both. **VGGNet** won the second place on ImageNet Challenge in 2014 where it shows good

performance in critical factor. GoogLeNet won the first place on ImageNet Challenge in 2014 which has code name as Inception because it has inception module. ResNet won the first place on ImageNet Challenge in 2015 which is trained on very deep network up to 1,200 layers. ResNext won the second place on ImageNet Challenge in 2016 which is the extension of ResNet network which replaces the standard residual block to increase the cardinality. DenseNet has the dense layer in the network; it makes the pattern of loop connectivity where the information or data are passed from one layer to the other.

II. Related Work

A. Appearance

Appearance is the vital role in the identification of the facial expression, gesture, body movements, etc. Color and texture are the two factors of appearance use to identify the person in image. Color which address to the lighting variations. Capturing appearance of a person in the motionless state is very comfort one but when capturing of the person's image in articulation, monolithic 3D model, etc are difficult process. Image can be captured accurately by using the high resolution pixel camera, good detector, architectural structure to increase the efficacy in the network. For example, a person wears the bright red shirt can be easily detecting through the color in the multi-camera frames. When a person wears the plain shirt it is quite difficult in the frame to recognize the color. To get the exact picture off the human in the video or image the pixel resolution should be high to re-identify the person.

B. Multi-cameras:

Cameras are increasing due to its need in the society for safety and security purposes. Public places such as shopping malls, streets, theaters, railway stations, etc. and security concern activities like investigation bureau, jail, police station, houses, etc. to cover the evidence in the places. Due to need of more focus on the particular region or area multi-cameras are introduced. Data association and detection are the two approaches of multi-cameras tracking. Most algorithms focused on single camera than the multi-camera due to some criteria where multi-camera has two type scenarios. First type is look at same scene or fully overlapping with each other and the second type is partially or non overlap between each other.

III. Literature Review

R Aarthi* et al. In 2018, 'A Survey of Deep Convolutional Neural Network Applications in Image Processing' paper is about convolutional architecture are used for heavy problems in the computer vision using features. Application of CNN generally classified into [1] object detection due to challenging problems like expression variation, face verification in background and occlusion, [2] classification is used to trained data under various conditions, and quality enhancement is used to improve the resolution. Face recognition can be done by DCNN architecture contained by DeconvNet package. Neural network performed well in largest ImageNet dataset which contains images.

Yann LeCun* et al. In 1988, 'Gradient Based Learning Applied to Document Recognition' is about **LeNet** which is used for recognizing characters in the convolutional network. Recognizing individual character and separating the characters from the neighbor is challenging factor in the network structure. Heuristic Over-Segmentation is used to overcome the challenge face by the LeNet structure and this task is costly one and labeling is difficulty. LeNet has 60,000 parameters and several modules. Back-propagation is used to derive the character recognition. LeNet has various versions such as LeNet 1, LeNet 4, and Boosted LeNet are explained for the network module usage. LeNet is the first architectural structure in convolutional neural network in deep learning which is implemented for the recognition method and it is the classic neural network.

Alex Krizhevsky* et al. In 2012, 'ImageNet Classification with Deep Convolutional Neural Networks' is about **AlexNet** which participated in ImageNet contest known as ILSVRC-2010 which has 1000 various classes. Depth is very important factor to achieve the great result in the network. It has 60 million parameters, 65k neurons and 5 convolutional layers in it, which shows better performance on test data as 37.5% in top-1 error rate and 17.0% in top-5 error rate. Due to variation of model in ILSVRC-2012 has achieved 15.3% in top-5 error rate and won first place in the challenge. AlexNet architectural structure is mainly used to reduce the over-fitting in the network by using the dropout layer.

Christian Szegedy* et al. In 2017, 'Rethinking the Inception Architecture for Computer Vision' is about convolutional neural network gains attention in 2014 on various benchmarks. In 2014, VGGNet and **GoogLeNet** gains attention on ILSVRC competition where both architectural structure gains similar outperformance and VGGNet has simple structure and quite costly where as GoogLeNet work well under memory constrain and has computational budget. When comparing to AlexNet, VGGNet is thrice more in parameter and GoogLeNet is twelve time less in parameter as 5 million. Inception-v3 has the lowest error rate in both top-1 and top-5 accuracy as 17.2% and 3.58% respectively on the validation test. GoogLeNet is the winner of 2014 ILSVRC because it makes reduction the half of the error, 3.5% of top-5 error is detected in evaluation of multi-crop.

Kaiming He* et al. In 2015, 'Deep Residual Learning for Image Recognition' using image processing where the residual architectural structure is the used to ease the training on the deep networks than the other architectural structures. The 18/34-layer **ResNet** has lower accuracy than 50/101/152-layer ResNet on considering top 5 error test. ResNet architecture is 8 times deeper than the VGGNet structure. Depth in residual network 152 layer has lower complication than 16/19 layer in VGGNet. ResNet placed first in ILSVRC 2015 competition because it had achieved 3.57% of error rate on the ImageNet dataset. ResNet also achieved first place in various challenges such as COCO detection, COCO segmentation in 2015 competition. This powerful proof shows that the ResNet is applicable to various vision and non-vision troubles.

Saining Xie* et al. In 2017, 'Aggregated Residual Transformations for Deep Neural Networks' is the concept of **ResNext** which is more effective than the deeper neural network by increasing the capacity. ResNext is used to increase the cardinality to improve the network accuracy on classification. ResNext include more blocks like ResNet structure with cardinality as 32 paths. ResNext with 101- layer has 50% complexity and achieve better presentation accuracy than ResNet-200 layer. In dataset like ImageNet-5k and COCO object detection shows better performance than ResNet and has simpler design than inception module.

Gao Huang* et al. In 2018, 'Densely Connected Convolutional Networks' is about Dense Convolutional Network well known as **DenseNet**. Convolutional layer which is denser will give accurate value and efficiency to train the data when the layers are close enough to get the input and those close to get the output. They solve the fading gradient problem, toughen propagation features, support feature recycle, and the amount of parameters that diminishing considerably are the benefits of DenseNet architectural structure. While comparing to inception module, the DenseNet is simpler, more efficient and features are concatenate from dissimilar layers and comparing to ResNet, it increases the variation between different layers by getting the input that increase the efficiency.

IV. CNN Architecture In Deep Learning

There are several common architectural structure used widely for image recognition in network to recognize the person in the video frame with huge accuracy. In architecture, there is classic and modern structure used in image processing.

A. LeNet:

First real-world application which is implemented in convolutional neural network to increase the efficacy is LeNet. Even though LeNet is successful in initial, it did not gain more attention till other fresh techniques enter into the field of computer vision. LeNet is used for identify hand-written digit for zip code recognition in the service of postal and normally LeNet is used to read digits in image. LeNet network were shallow, contains about two and five layers of convolutional network, and has large kernel field in the layer of input and has small kernel field closer to output. LeNet has the activation function as hyperbolic tangent in the network. LeNet has the lower memory problem due to containing of single large kernel filter.

B. AlexNet:

AlexNet is the next network which makes the world to gain attention in deep learning network by ImageNet competition in 2012. In challenge, AlexNet took the first place by 15.3% top-5 error rate in the dataset. AlexNet is also has similar functions as LeNet where it network were also shallow, contains about two and five layers of convolutional network, and has large kernel field in the layer of input and has small kernel field closer to output. AlexNet has the activation function as rectified linear units in the network which makes difference between them. AlexNet has 60,000 parameters to train the data. AlexNet architectural structure is mainly used to reduce the over-fitting in the network by using the dropout layer. AlexNet gained attention even though got the smaller

computation, heavy memory, lower accuracy and could not battle hand engineered crafted solutions.

C. VGGNet:

VGGNet is the network won first runner up in 2014 ImageNet challenge in the dataset. VGGNet showed good performance in the critical factor of depth network. VGGNet gives 7.3% of top-5 error rate in dataset and has 138 million parameters. VGGNet has GPU problem in modest size because of huge requirement computation difficulty both in time and memory. Due to have the large sized kernel filters in the network it became inefficiency in the function. VGGNet has highest memory and most operation has to done to get the accuracy.

D. GoogLeNet:

GoogLeNet is popular because of its well usage in Google and had secure first place in 2014 ImageNet challenge in the dataset. GoogLeNet overcome the LeNet by making network efficiency, where GoogLeNet replace the numerous small sized kernels instead of single large sized kernel in the module. GoogLeNet has many modules known as Inception module which makes the dense construction of network into the normal dense construction by making approximately spare convolutional neural network. Inception module has split-transform-merge strategy which also present in ResNext network is vital general property, where Inception module can split the input into a small amount of lower-dimensional embedding and transform some filters and join together by concatenation. Due to the strategy the power of huge and dense layer, has computational complication low. GoogLeNet has code name as Inception which has many versions such as version 1 has 5 million parameters and version 3 has 23 million parameters and Nvidia GPU is used for the parameters. Inception V4 (ResNet + Inception) is the network which achieves 80% the highest accuracy of the entire network in the convolutional neural network in deep learning.

E. ResNet:

The popular network in the deep learning network is ResNet that boost the channel depth for mounting the overall capability of network which is the most efficient way. So, ResNet is the deepest neural network architecture which gives highest accuracy in the CNN as 95.51% of top-5 accuracy and 3.6% of top-5 error rate in the bulky dataset in 2015 ImageNet challenge and secure first place that gains the computer vision researchers to make deep learning to the next level. ResNet have achieved record-breaking performance not only in ImageNet but also in COCO object detection. ResNet improve its training by dropping the layers erratically in deep network. ResNet gives the moderate efficiency depending on the model used in the network and trained on very deep network nearly up to 1200 layers. ResNet contains ResNet-blocks is used to learn residual and mapping each layer that close to identity function.

Table 1. Comparison of Top 5 Error Rate and Top 5 Accuracy

Sr. No.	Architecture	Top-5 Error Rate	Top-5 accuracy
1	ResNet	3.6%	95.51%
2	VGG Net	7.35%	92.7%
3	AlexNet	15.3%	70.02%
4	GoogLeNet	6.67%	93.3%

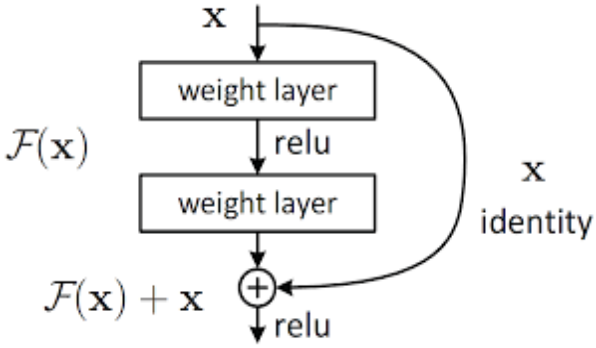

Fig 1 : RESNET - 50

F. DenseNet:

DenseNet is the architectural structure in CNN where it makes the loop connectivity pattern between every two layers in the network so, that each layer receives its signal from its preceding layer. Compared to ResNet, DenseNet achieve better performance with less complexity. The layer which is denser could not carry more information signal from one to next layer that causes the low information bottleneck problem. To avoid the problem, Dense and Slim is used to make the model more compact. The compact model gives high computational efficiency and parameter competence. Dense and Slim approach make the layer which is dense into thin layer to carry more information from one layer to the other for process of communication. The connectivity in DenseNet is each layer is connect to the next layer where first layer is

connected to second, third, fourth and till last layer and second layer connect to third, fourth and till last layer, so on. of the network while compared to ResNet, the connectivity is much easier than DenseNet -121 which first layer connect to second layer, second layer to third, third to fourth layer and till last layer. Advantage of DenseNet is more direct supervision. Comparison on test error between ResNet and DenseNet, where ResNet (110 layer, 1.7 M) gives 6.41%, ResNet (1001 layer, 10.2 M) gives 4.62%, DenseNet (100 layer, 0.8 M) gives 4.5%, DenseNet (250 layer, 15.3 M) gives 6.41%.

DenseNet-121:-
5+(6+12+24+16)*2 = 121

5 – Convolution and Pooling Layer
3 – Transition layers (6,12,24)
1 – Classification Layer (16)
2 – DenseBlock (1x1 and 3x3 conv)

Fig 2 : DenseNet – 121

TABLE II: Comparison of Error Rate between Resnet and Densenet

Sr. No.	Architecture	Layer	Parameter	Error-Rate
1	ResNet	110	1.7	6.41%
2	ResNet	1001	10.2	4.62%
3	DenseNet	100	0.8	4.5%
4	DenseNet	250	15.3	6.41%

G. ResNext:

ResNext is the extension of deep residual network which replace the residual block by a block of ResNext with cardinality in same complexity. ResNext also has common property with Inception module as "split-transform-merge" strategy which make residual block split into many ResNext block with the cardinality 32 to transform the data and merge entire residual block into one to form single ResNext. Compared to Inception module, ResNext has simpler design pattern. ImageNet-5k and COCO object detection dataset; ResNext shows better performance in accuracy than ResNet and secure second place in 2016 challenge. ResNext can create filters with full channel intensity of input, numbers of branches or groups as cardinality of ResNext cell, gains between mounting the cardinality, intensity and thickness of network. ResNext with 101- layer has 50% complexity and achieve better presentation accuracy than ResNet-200 layer.

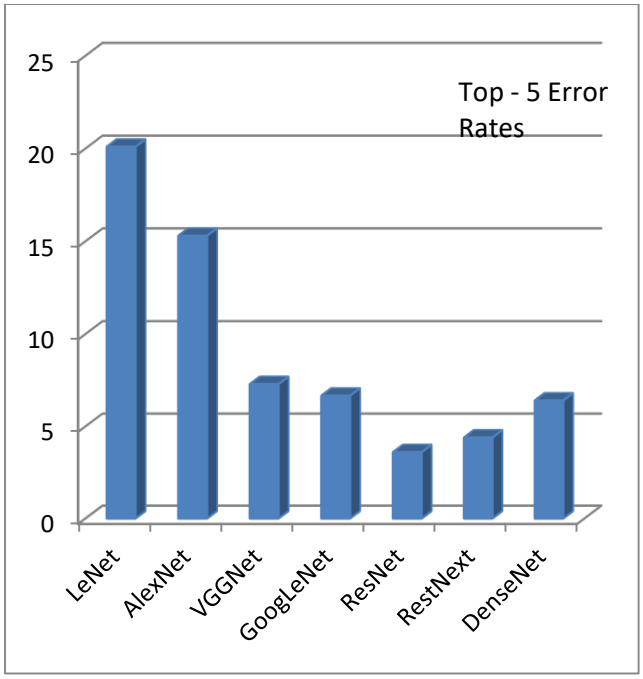

Fig 3: Top-5 Error Rate of Architectural Structure in Convolutional Neural Network

V.CONCLUSION

The architectural structures which are popular in Convolutional Neural Network (CNN) are presented in this paper. The architectural structures are the key to increase the network accuracy in the person re-identification. Even though cameras with high resolution, good quality detectors or trackers placed in the process of recognition, network is needed to qualify the image of the person. Each architectural structure has the high-quality accuracy depending on the model or dataset used to train on them. In the future work, one of the architectural structures is taken for the implementation process because nowadays security and safety became more essential not only in public places but also in private areas for own good.

REFERENCES

[1] Yinghao Cai and Matti Pietik¨ainen, "Person Re-identification Based on Global ColorContext" Springer-

Verlag Berlin Heidelberg, Part I, LNCS 6468, pp. 205–215, 2011.
[2] Kheir-Eddine Aziz, Djamel Merad, and Bernard Fertil, "Person Re-identification Using Appearance Classification" Springer-Verlag Berlin Heidelberg, ICIAR 2011, Part II, LNCS 6754, pp. 170–179, 2011.
[3] Sara Iodice and Alfredo Petrosino, "Person Re-identification Based on Enriched Symmetry Salient Features and Graph Matching" Springer-Verlag Berlin Heidelberg, MCPR 2013, LNCS 7914, pp. 155–164, 2013.
[4] R.Elankeerthana, K.Kokila,"Grassmann: Face Based Recognition Using Convolutional Neural Network In Deep Learning", International Journal of Scientific & Technology Research Volume 9, Issue 02, February 2020, ISSN 2277-8616.
[5] Chunxiao Liu1, Shaogang Gong2, Chen Change Loy3, and Xinggang Lin1, "Person Re identification: What Features Are Important?" Springer-Verlag Berlin Heidelberg, ECCV Ws/Demos, Part I, LNCS 7583, pp. 391–401, 2012.
[6] Ergys Ristani; Carlo Tomasi, "Features for Multi-target Multi-camera Tracking and Re- identification" 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, DOI: 10.1109/CVPR.2018.00632
[7] Zheng, Z.; Zheng, L.; and Yang, Y. 2017. Unlabeled samples generated by gan improve the person reidentification baseline in vitro. In Proceedings of the IEEE International Conference on Computer Vision.
[8] Liu, W.; Camps, O.; and Sznaier, M. 2017. Multi-camera multi-object tracking. arXiv preprint arXiv:1709.07065.
[9] Maksai, A.; Wang, X.; Fleuret, F.; and Fua, P. 2017. Non-markovian globally consistent multi-object tracking
[10] Andreas Specker, Daniel Stadler. el, "An Occlusion-aware Multi-target Multi-camera Tracking System" cvpv 2021
[11] P.Bergmann, T. Meinhardt, and L. Leal-Taixe. Tracking ´ without bells and whistles. In Int. Conf. Comput. Vis., pages 941–951, 2019.
[12] P. Khorramshahi, N. Peri, J.-c. Chen, and R. Chellappa. The devil is in the details: Self-supervised attention for vehicle re-identification. In Eur. Conf. Comput. Vis., pages 369–386, 2020.
[13] S. He, H. Luo, W. Chen, M. Zhang, Y. Zhang, F. Wang, H. Li, and W. Jiang. Multi-domain learning and identity mining for vehicle re-identification. In IEEE Conf. Comput. Vis. Pattern Recog. Worksh., pages 582–583, 2020.
[14] Y. He, J. Han, W. Yu, X. Hong, X. Wei, and Y. Gong. Cityscale multi-camera vehicle tracking by semantic attribute parsing and cross-camera tracklet matching. In IEEE Conf. Comput. Vis. Pattern Recog., pages 576–577, 2020
[15] H.-M. Hsu, T.-W. Huang, G. Wang, J. Cai, Z. Lei, and J.-N. Hwang. Multi-camera tracking of vehicles based on deep features re-id and trajectory-based camera link models. In IEEE Conf. Comput. Vis. Pattern Recog. Worksh., pages 416–424, 2019.
[16] H. W. Kuhn and Bryn Yaw. The hungarian method for the assignment problem. Naval Res. Logist. Quart., pages 83– 97, 1955.
[17] P. Kohl, A. Specker, A. Schumann, and J. Beyerer. The mta ¨ dataset for multi-target multi-camera pedestrian tracking by weighted distance aggregation. In IEEE Conf. Comput. Vis. Pattern Recog. Worksh., pages 1042–1043, 2020.
[18] X. Li, W. Wang, L. Wu, S. Chen, X. Hu, J. Li, J. Tang, and J. Yang. Generalized focal loss: Learning qualified and distributed bounding boxes for dense object detection. In Adv. Neural Inform. Process. Syst., pages 21002–21012, 2020
[19] Y. Qian, L. Yu, W. Liu, and A. G. Hauptmann. Electricity: An efficient multi-camera vehicle tracking system for intelligent city. In IEEE Conf. Comput. Vis. Pattern Recog. Worksh., pages 588–589, 2020.
[20] S. Qiao, L.-C. Chen, and A. Yuille. Detectors: Detecting objects with recursive feature pyramid and switchable atrous convolution. arXiv:2006.02334, 2020.
[21] A. Specker, A. Schumann, and J. Beyerer. A multitask model for person re-identification and attribute recognition using semantic regions. *In Art. Intell. and Mach. Learn. in Def.* Appl., 2020.
[22] Dr.T.Abirami, Ms.R.Elankeerthana, "Effective Detection Method for Fruit Recognition with Deep Learning", *International Journal of Advanced Science and Technology,* Vol. 29, No. 5, (2020), pp. 3512 – 3519
[23] D. Stadler and J. Beyerer. Improving multiple pedestrian tracking by track management and occlusion handling. *In IEEE Conf. Comput. Vis. Pattern Recog.*, 2021.
[24] D. Stadler, L. W. Sommer, and J. Beyerer. Pas tracker: Position-, appearance- and size-aware multi-object tracking in drone videos. *In Eur. Conf. Comput. Vis. Worksh.*, pages 604–620, 2020.
[25] Y. Sun, L. Zheng, Y. Yang, Q. Tian, and S. Wang. Beyond part models: Person retrieval with refined part pooling (and a strong convolutional baseline). *In Eur. Conf. Comput. Vis.*, pages 480–496, 2018