

Segmentation for Complex Background Images using Deep Learning Techniques



Ajay Bazil Issac, Manohar N, Varsha Kumari Jain M

Abstract—Segmentation is an important stage in any computer vision system. Segmentation involves discarding the objects which are not of our interest and extracting only the object of our interest. Automated segmentation has become very difficult when we have complex background and other challenges like illumination, occlusion etc. In this project we are designing an automated segmentation system using deep learning algorithm to segment images with complex background.

Keywords: Background Images, Deep Learning

I. INTRODUCTION

These days, image segmentation had turned out to be one of the main issue in the turf of computer vision. A lot of situation is in expanding interest for abstracting pertinent learning or semantic data from images, for example, self-driving, Robotics, etc. Each pixel in the image will be classified and divide the image to a meaningful region as a pre-processing step for the image analysis. In the past years in order to get good segmentation result researches were proposed several method such as “simplest pixel-level thresholding”, “clustering-based segmentation method”, and “graph partitioning segmentation method”. All of these methods, because of their less calculation difficulties and less parameters, they have more efficiency. Even though, there was some problems on their performance for the segmentation task without any non-natural extra information. When deep learning came to the world of computer vision, the image segmentation was implemented with the help of “Convolutional Neural Network” (CNN). And CNN was far better than the traditional methods used for the semantic segmentation.

II. RELATED WORK

In the last couple of years, number of researches proposed different techniques with “pixel-level thresholding strategies”, “clustering-based segmentation”, and “graph based segmentation” in order to create better output in the

semantic segmentation. But by the entrance of CNN, semantic segmentation got much better outcome when compared to the traditional techniques. A huge number of method were addressed for the improvement of semantic segmentation is FCN based methods. Chen et al introduced a deep lab model for pixel level classification. He uses a fully connected CRF as post-processing stage in their pipeline to make perfect the segmentation outcome. The unfortunate localization property in the deep networks were overcome by the basic version of the “DeepLab-v1”. It combines the reactions at the last FCN layer with a CRF for 1st time. Because of using this all pixel in the image even if it is far apart also taken to account, execution of this method were able to recuperate detailed arrangements in the segmentation that were lost due to the spatial invariance of the FCN. After this they improve their system and create the “DeepLab-v2 and DeepLab-v3”. This system was having exceptional features that were missing in many of the other proposals. Deep Lab systems are extraordinarily popular and successful in generating precise result. During this time a famous image pre-processing called “super pixel method” developed. There are 2 major different categories of super pixel segmentation are there, they are “graph-based method and gradient-ascent-based methods”. K-means clustering is an effective approach to generate super pixel. Jonathan Long, Evan Shelhamer and Trevor projected a way[12]. Their key insight was to develop a FCN which we give arbitrary sized input images and the system will produce a correspondingly sized output image. The system have a n efficient extrapolation and learning. In this work they adapt the classification network such as “Alexnet, VGGnet, GoogLeNet”. They handover the learned image by “fine-tuning” to the segmentation job. They create an architecture in order to generate an accurate segmentation. This architecture will combine the information from a “deep coarse layer” with appearance info from a shallow. Santosh Jangid and P S Bhatnagar[13] proposed a DCNN with multiple layers are projected. Multiple layers work to create an enhanced feature space. First layer studies 1st order features (e.g. edge). Second layer studies higher order features (combinations of edges, etc.). Some models study in an unsupervised mode and recognize general features of the input space. The last layer of features is fed into supervised layer and the network is frequently tuned using supervised training with initial weighting learned in the unsupervised phase. This DCNN gives a resolution in the form of color segmentation. Deep CNN based segmentation model shows 93% accurateness on BSDS300 dataset. “Gabriel L.

Revised Manuscript Received on 30 July 2019.

* Correspondence Author

Ajay Bazil Issa* Department of Computer Science, Amrita School of Arts and Sciences Mysuru Campus, India,

Manohar N, Department of Computer Science, Amrita School of Arts and Sciences Mysuru Campus, India,

Varsha Kumari Jain M, Department of Computer Science, Amrita School of Arts and Sciences Mysuru Campus, India,

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](https://creativecommons.org/licenses/by-nc-nd/4.0/) article under the CC-BY-NC-ND license <http://creativecommons.org/licenses/by-nc-nd/4.0/>

Segmentation for Complex Background Images using Deep Learning Techniques

Oliveira, Claas Bollen, Wolfram Burgard and Thomas Brox introduced three different architecture based on Up-Convolutional Networks. The first architecture called Part-Net. This is used for human body part segmentation in order to grab any specific problem. This architecture delivers robustness to overfitting. The next architecture they introduced is called Fast-Net. This architecture will reduce the computation load without dropping the representation power. This have the ability to run in the mobile GPUs also. The last and final architecture they build is called M-Net. In this M-Net through a multiresolution fusion they increase the robustness of semantic segmentation. They introduce a new dataset called “Freiburg City Dataset” which can be useful for the robotics circumstances. K. V. Jobin and C. V. Jawahar [15] proposed a method for segmentation in documents. In this document segmentation the pixel in the image will be classifies in to labels like text, comments, decorations and background. First they extract the features from the image. Then learn the image using a SVM classifier with the help of this features. In this learning they used “fisher vector encoded convolutional layer features (FV-CNN) “ and “fully connected layer features (FC-CNN)”. When compared to other approaches they got much better result in the document segmentation. Shubhra Aich, William van der Kamp, Ian Stavness [16] came with an architecture for image segmentation using the “depth-to-space (D2S) operation”. In this approach they are having a standard “convolutional neural network” and a final feature maps. This model decreases the calculations by half by removing the decoder portion. This model exhibit good performance while evaluating in the “Road segmentation dataset” with less computational cost.

III. PROPOSED METHODOLOGY

Our proposed methodology has 4 stages: pre-processing, ground truth, feature extraction and segmentation. Figure 1 shows the architecture of the proposed system.



Fig.1: - Basic architecture of the project.

Preprocessing

Image pre-processing is the process of enhancing the image by removing the unwanted distortion in order to perform certain image processing techniques. The image pre-processing help to improve the result of later processing. One of the digital filtering method called median filter is used here. It will remove the unwanted distortion from the images. Median filter is widely used method in the computerized image process because of, less bound conditions, it preserve edges though expelling commotion, conjointly having applications in flag process.

Median filter will decrease the commotion in the image by smoothening. Not at all like low pass separating, Median filter can safeguard discontinuities in a stage work and can smooth

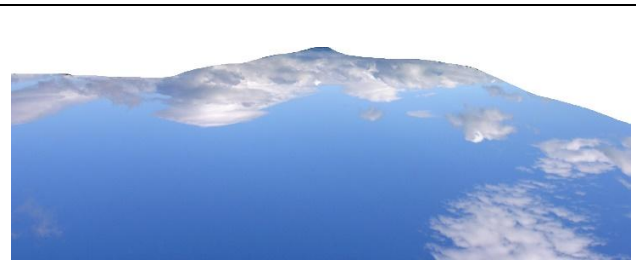
a couple of pixels whose qualities vary essentially from their surroundings without influencing different pixels.

A. Ground-truth

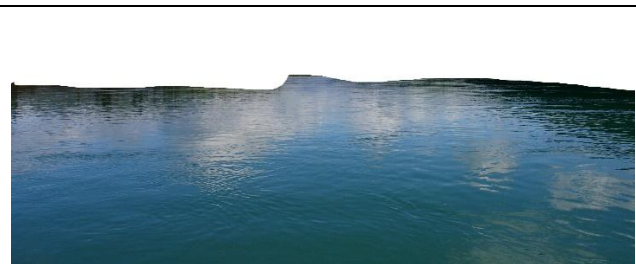
“In any segmentation algorithm creating ground truth is an important stage where the user will manually identify certain regions which occurs commonly in the images like grass, sand, water, sky etc. In this work we have created 12 different classes of ground truth namely, grass, green, leaf, plants, rock, sky, snow, soil, stone, tree, water, wood. These ground truths are later fed to feature extraction using AlexNet CNN, which will be further used to perform the segmentation. Figure 2.2 shows few samples of ground truth images of different class.



Class grass



Class sky



Class water

Fig.2.2: - Sample Ground truth images

B. Feature Extraction

In this stage we are extracting features for the ground truth images using AlexNet pre-trained convolutional neural network.

AlexNet

AlexNet is a convolutional neural network. It is a successful neural network in the “ILSVRC 2012”. It have the capacity of classifying thousands of classes.

The AlexNet have the architecture similar to the LeNet. It having filters in each layers and the convolutional layers are stacked. The ReLU activations were attached to every convolutional layer and FCN layers. The input image which we used in the AlexNet is 256*256, which means every image that we used as test image and preparation set should be 256*256. But in some cases if the image we used in processing is not the size of 256*256, the network will resize the images to 256*256 in order to utilize in the processing or training.

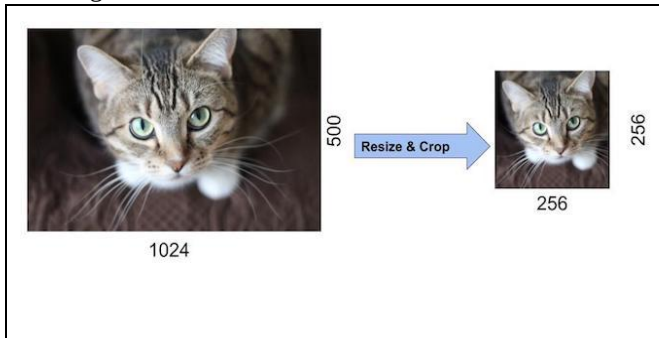


Fig.2.3: - Image resizing in to 256*256

AlexNet will take at least 5 days to train the data in “2 GTX 580 3GB GPUs” because of its 650000 neurons and sixty million parameters.

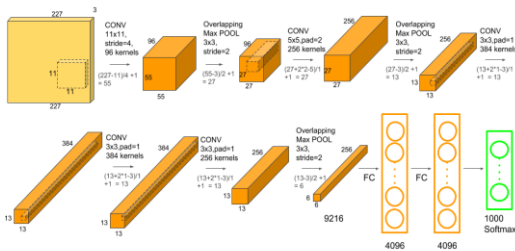


Fig.2.4: - AlexNet Architecture

AlexNet consist of five Convolutional Layers and three Fully connected Layers The features in the images will be extracted by multiple convolutional kernels. In a single convolutional layer itself consist of a number of number of kernels which are having same size. The first conv layer having 96 kernels of size 11*11*3. Kernels having same width, height and depth will be identical as unit's channels. After the first 2 convolutional layer AlexNet having overlapping max pooling layers(MPL). The next convolutional layers are (3rd, 4th, 5th) linked directly. Again after the 5th convolutional layer max pooling layers are there. After the first and second fully connector layer, a “softmax classifier” consist of thousand class labels were connected. After the first and second fully connector layer, a “softmax classifier” consist of thousand class labels were connected. ReLU nonlinearity is applied after the convolutional layer and fully connected layer. Before doing pooling, by nearby standardization the ReLU nonlinearity of first and second convolution layers are triled. But it is not much useful.

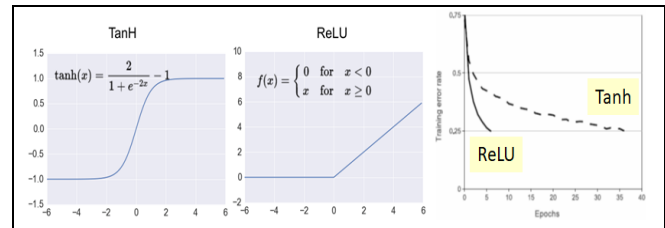


Fig.2.5: - Activation Function

Tanh was used before AlexNet. ReLU is introduced in AlexNet and it is 6 times faster than Tanh to get 25% training error rate. Max Pooling Layers (MLP) Max pooling is a simple discretization process. it is used to down-sample an input representation by dividing the image into pooling region and calculating the maximum of those region., with the exception of the neighboring windows over which the maximum is processed cover one another. The creators utilized “pooling windows” having size 3*3 with a walk of 2 amid the nearby windows. This covering behavior of pooling diminished the best “1 error rate by 0.4% and top-5 error rate by 0.3%” individually when contrasted with utilizing non-covering 2*2 pooling windows with a walk of 2 that would give same yield measurements. Dropout A neuron is dropped from the network with a likelihood of 0.5. At the point when a neuron is dropped, it doesn't add to any advance or in reverse propagation. So every information experiences an alternate network design. Subsequently, the learning weight parameters are progressively vigorous and don't get overfitted effectively. Amid testing, there is no dropout and the entire system is utilized, yet yield is scaled by a factor of 0.5 to represent the missed neurons while preparing. Dropout builds the quantity of emphases expected to join by a factor of 0.2, yet without dropout, AlexNet would overfit generously. “Feature extraction begins in image processing from an initial set of measured data and creates resulting values (features) proposed to be instructive and non-redundant, helping the following learning and generalization steps, and in some cases leading to better human clarifications. Feature extraction may be a spatiality reduction method, wherever associate degree initial set of raw variables is reduced to additional governable teams (features) for process, whereas still exactly and wholly describing the original data set.” We done feature extraction using AlexNet. We took the ground truth image classes and train each classes in AlexNet to extract the features. Like this we divide the input image into blocks and train each block separately.

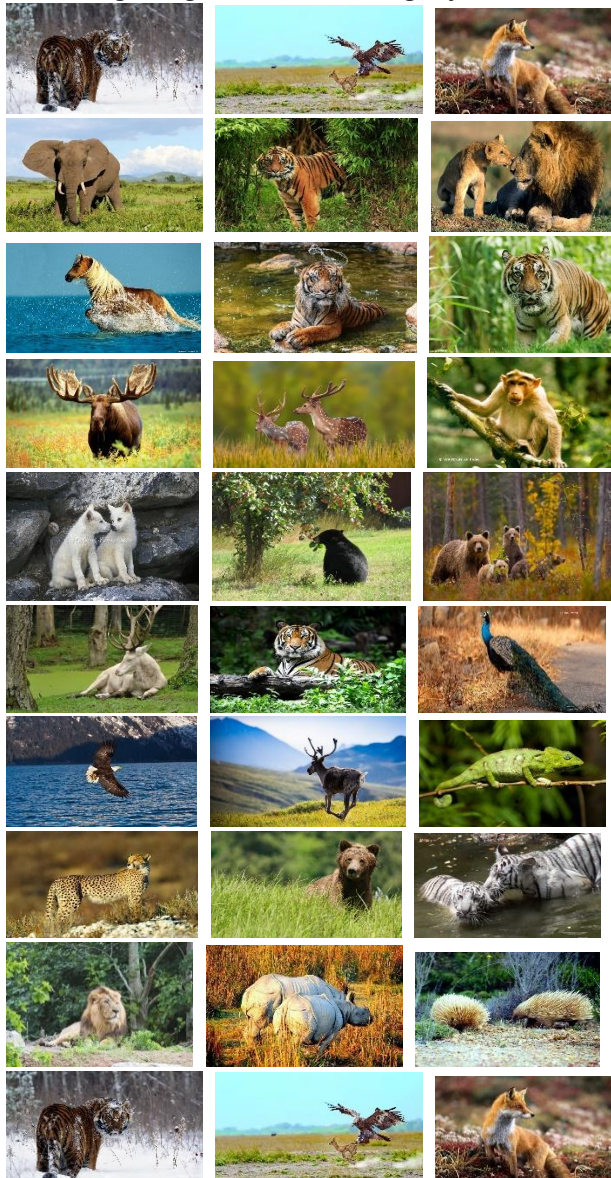
Variables - featuresTrain							
4089	4090	4091	4092	4093	4094	4095	4096
-12.0766	-7.6907	-9.2064	-5.3857	-2.4884	-6.8800	-2.2047	-0.2849
-3.7516	4.3704	-4.7379	-8.3923	-7.9218	-3.2925	0.8080	-3.3958
-3.7516	4.3704	-4.7379	-8.3923	-7.9218	-3.2925	0.8080	-3.3958
-5.0725	1.7160	-5.1334	-7.5691	-4.6482	-4.7046	-3.1459	-0.7051
-5.0725	1.7160	-5.1334	-7.5691	-4.6482	-4.7046	-3.1459	-0.7051
0.0572	1.5699	-4.9302	-9.0396	0.0010	0.2949	-1.2897	-5.2489
-1.1897	1.5899	-3.9757	-7.1195	-3.7655	-2.0205	1.0983	-3.4308
0.0828	1.9177	-3.3571	-4.4673	-1.0025	-4.2551	-4.2982	-2.0621
-2.4921	2.6722	-9.6251	-6.8940	-5.9369	-4.5534	-2.8511	-3.6618
-3.0247	2.4274	0.4950	-4.7505	-5.4886	-5.7378	1.1133	-1.7193
-3.0247	2.4274	0.4950	-4.7505	-5.4886	-5.7378	1.1133	-1.7193
-4.3180	-3.8942	-4.2382	-7.4004	-0.9486	-4.0001	-2.0264	0.3206
-7.7142	0.7227	-4.7784	-3.3044	-0.5807	-4.0766	-1.8130	1.7363

Class grass

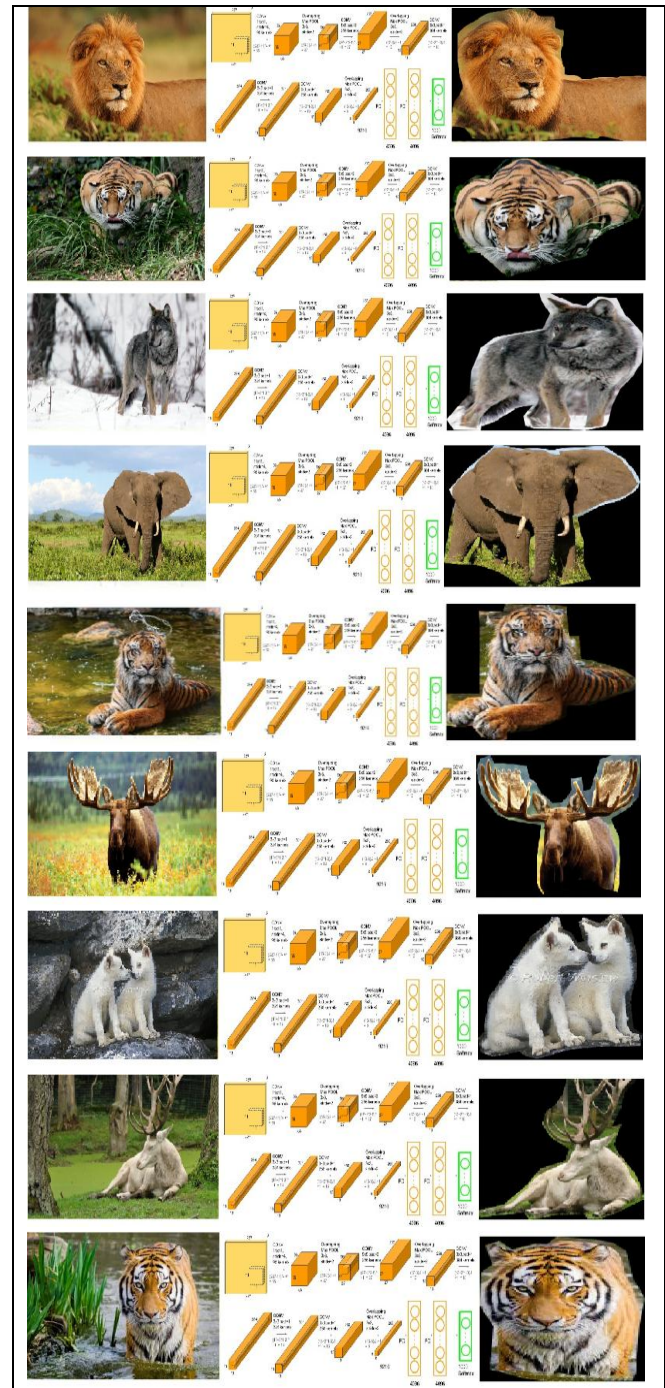
Fig.2.6: - Sample feature extraction

B. Segmentation

Image segmentation in pc vision is that the method of partitioning a digital image into multiple segments. The aim of segmentation is to represent image into something which is easier to analyze and meaningful. Segmentation is used to discover objects and boundaries in images. Initially the testing image is resized to 256*256 and applied median filters to enhance the image. After the pre-processing, the images are divided to different blocks like 4, 8, 16, 32. Then we will train each block of images using pre-trained Convolutional Neural Network called AlexNet. Later, each and every block in each case are compared with the ground truth class for checking similarity using Euclidean distance. After performing matching, if any block having 80% and above similarity with any of the ground truth class then it represents that, the respective block belongs to the background and we will make that entire block value to zero. The blocks which are having less than 80% matching then it concludes that the block is object of our interest and we will retain the block. Finally, we merge all the blocks to get final segmented image, after eliminating background and extracting object of our interest.



Sample input images



Segmented output images

REFERENCES

1. Ghiasi, G.; Fowlkes, C.C. Laplacian pyramid reconstruction and refinement for semantic segmentation. In Proceedings of the European Conference on Computer Vision, Amsterdam, the Netherlands, 8–16 October 2016; Springer: New York, NY, USA, 2016; pp. 519–534.
2. Lin, G.; Shen, C.; Hengel, A.V.D.; Reid, I. Exploring context with deep structured models for semantic segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *40*, 1352–1366. [CrossRef] [PubMed]
3. Noh, H.; Hong, S.; Han, B. Learning deconvolution network for semantic segmentation. In Proceedings of the IEEE International Conference on Computer Vision, Boston, MA, USA, 8–10 June 2015; pp. 1520–1528.
4. Chen, L.C.; Yang, Y.; Wang, J.; Xu, W.; Yuille, A.L. Attention to scale: Scale-aware semantic image segmentation. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition,

5. Las Vegas, NV, USA, 27 June–1 July 2016; IEEE: New York, NY, USA, 2016; pp. 3640–3649.
6. Zheng, S.; Jayasumana, S.; Romera-Paredes, B.; Vineet, V.; Su, Z.Z.; Du, D.L.; Huang, C.; Torr, P.H.S. Conditional random fields as recurrent neural networks. In Proceedings of the 2015 IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; IEEE: New York, NY, USA, 2015; pp. 1529–1537.
7. Chen, L.-C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. Semantic image segmentation with deep convolutional nets and fully connected crfs. *arXiv* **2014**, arXiv:1412.7062.
8. Chen, L.-C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Trans. Mach. Intell.* **2018**, *40*, 834–848. [[CrossRef](#)] [[PubMed](#)]
9. Chen, L.-C.; Papandreou, G.; Schroff, F.; Adam, H. Rethinking atrous convolution for semantic image segmentation. *arXiv* **2017**, arXiv:1706.05587.
10. Shi, J.; Malik, J. Normalized cuts and image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2000**, *22*, 888–905.
11. Felzenszwalb, P.F.; Huttenlocher, D.P. Efficient graph-based image segmentation. *Int. J. Comput. Vis.* **2004**, *59*, 167–181. [[CrossRef](#)]
12. Van den Bergh, M.; Boix, X.; Roig, G.; de Capitani, B.; Van Gool, L. Seeds: Superpixels extracted via energy-driven sampling. In Proceedings of the 12th European Conference on Computer Vision-Volume Part VII, Florence, Italy, 7–13 October 2012; Springer: New York, NY, USA, 2012; pp. 13–26.
13. 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).
14. Jangid, S., & Bhatnagar, P. S. (2018). Semantic Image Segmentation using Deep Convolutional Neural Networks and Super-Pixels. *International Journal of Applied Engineering Research*, 13(20), 14657-14663.
15. Oliveira, G. L., Bollen, C., Burgard, W., & Brox, T. (2018). Efficient and robust deep networks for semantic segmentation. *The International Journal of Robotics Research*, 37(4-5), 472-491.
16. Jobin, K. V., & Jawahar, C. V. (2018). Document image segmentation using deep features. In Computer Vision, Pattern Recognition, Image Processing, and Graphics: 6th National Conference, NCVPRIPG 2017, Mandi, India, December 16-19, 2017, Revised Selected Papers 6 (pp. 372-382). Springer Singapore.
17. Aich, S., van der Kamp, W., & Stavness, I. (2018, June). Semantic Binary Segmentation using Convolutional Networks without Decoders. In 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) (pp. 182-1824). IEEE.
18. Rani, N. S., Nithusha, V. K., & Roshna, T. P. (2015). Automatic recognition and verification of defective tablet blisters using entropy based filtering and histogram processing. *International Journal of Applied Engineering Research*, 10(5), 13155-13167.
19. Manohar, N., Subrahmanya, S., Bharathi, R. K., YH, S. K., & Kumar, H. (2016, August). Recognition and classification of animals based on texture features through parallel computing. In 2016 Second International Conference on Cognitive Computing and Information Processing (CCIP) (pp. 1-5). IEEE.



Ajay Basil Issac, MCA student at Amrita Vishwa Vidyapeetham, Mysore, Karnataka, India, received his Bsc in Computer Science from Calicut University, Kerala, India



Varsha Kumari Jain M, MCA student at Amrita Vishwa Vidyapeetham, Mysore, Karnataka, India, received his BCA from Mysore University, Karnataka, India

AUTHORS PROFILE



Dr. Manohar N, currently working as Assistant Professor in department of computer science Amrita Vishwa Vidyapeetham Mysuru Campus. Research interests are machine learning, computer vision and data mining. Over 6 years of teaching and research experience with more than 12 publications in international conferences and journals.