Bangladesh Army International University Of Science And Technology

Title : ArNet A Deep Learning Architecture For Pixel-wise Semantic Segmentation Of Images

Presented By:

Arik Md. Isthiaque
ID – 1103096
Section – B
Department of CSE, BAIUST

Supervised By:

Arifa Islam Champa
Lecturer
Department of CSE, BAIUST

Introduction

Segmentation - Process that classify object in image

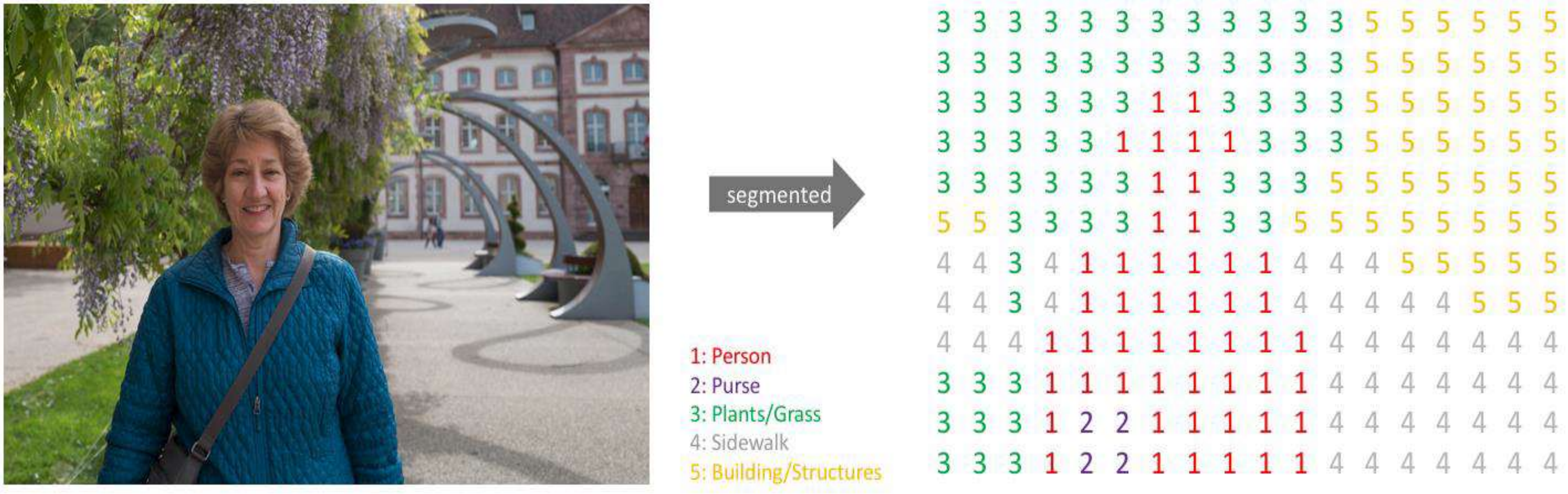


Figure 1: Visualization of how semantic segmentation works [1].

Challenges

- Building the DCNN
- Layer reduction
- Time reduction
- Add more classes
- Calculate the class weights

Motivation for Research

SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation [2]

- **Build to understand road scenes**
- **Model appearances and shapes**
- **Understand the spatial resolution between classes**
- **Used mostly for autonomous driving**

Literature Review

- **In *“Feedforward semantic segmentation with zoom-out features”* Mohammadreza Mostajabi, Payman Yadollahpour, and Gregory Shakhnarovich represented the statistical structure of a image by a purely feed forward network by mapping Superpixels using zoom-out features [3]**

- **In *“Convolutional Feature Masking for Joint Object and Stuff Segmentation ” Jifeng Dai, Kaiming He, and Jian Sun* image shape information was exploited by Convolutional Neural Network for object that are connected [4]**

Methodology

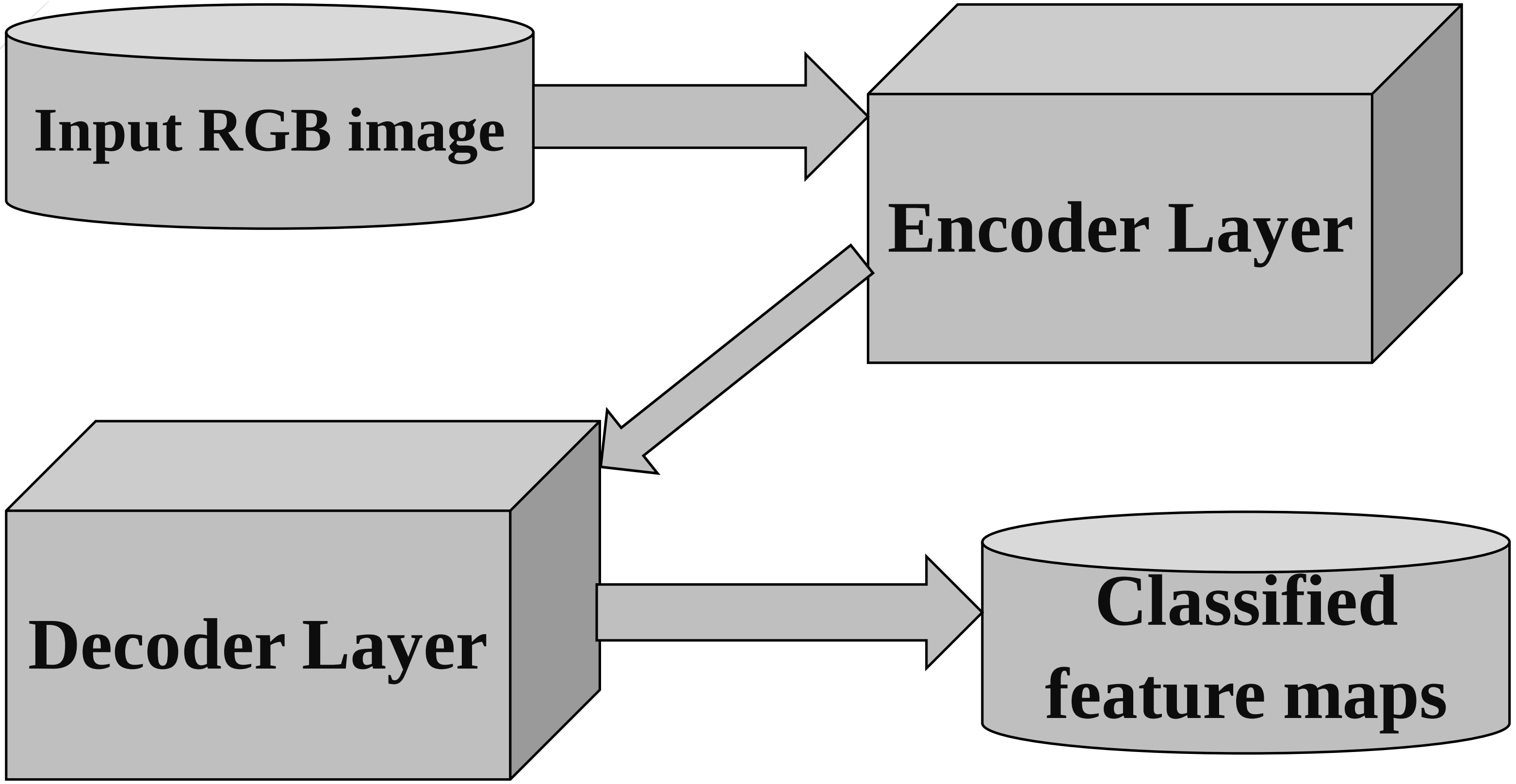


Figure 2: Overview of ArNet architecture.

Encoder Layer

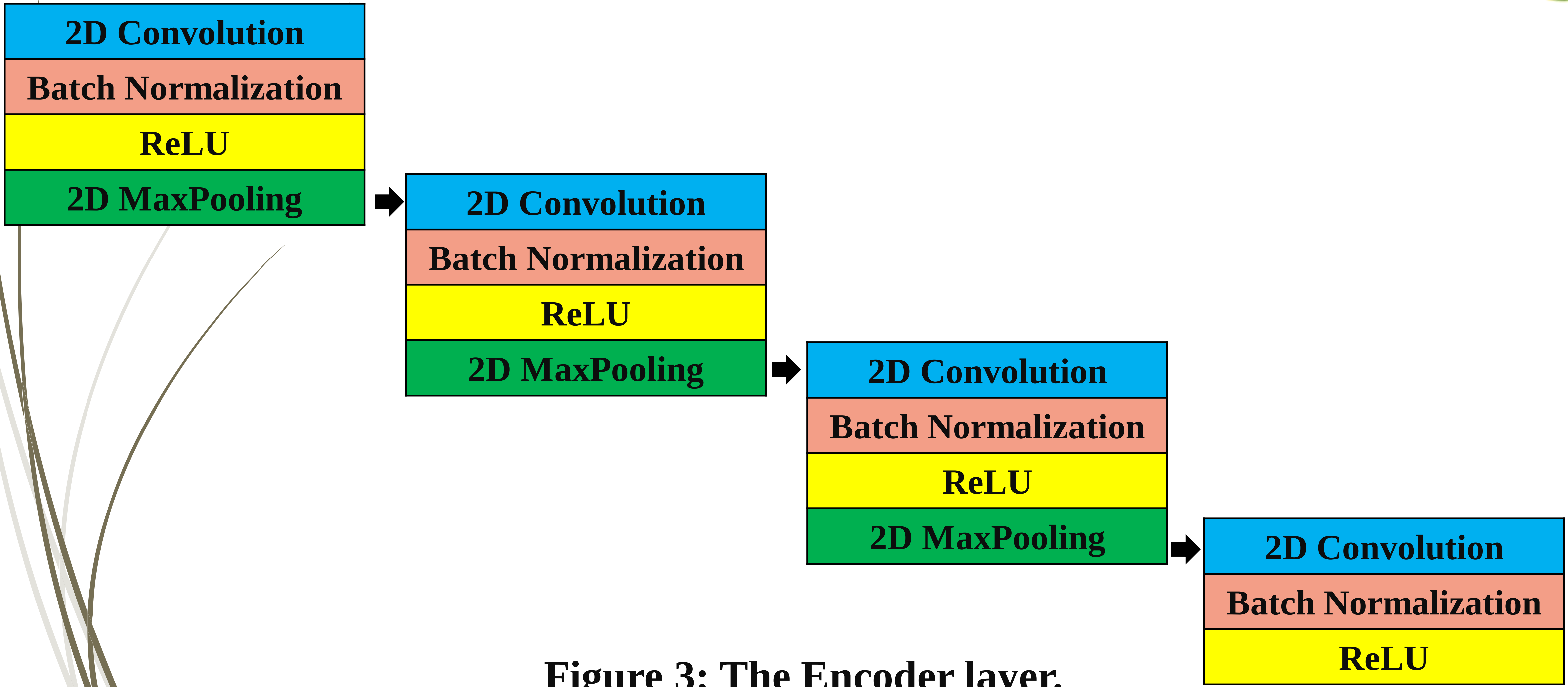
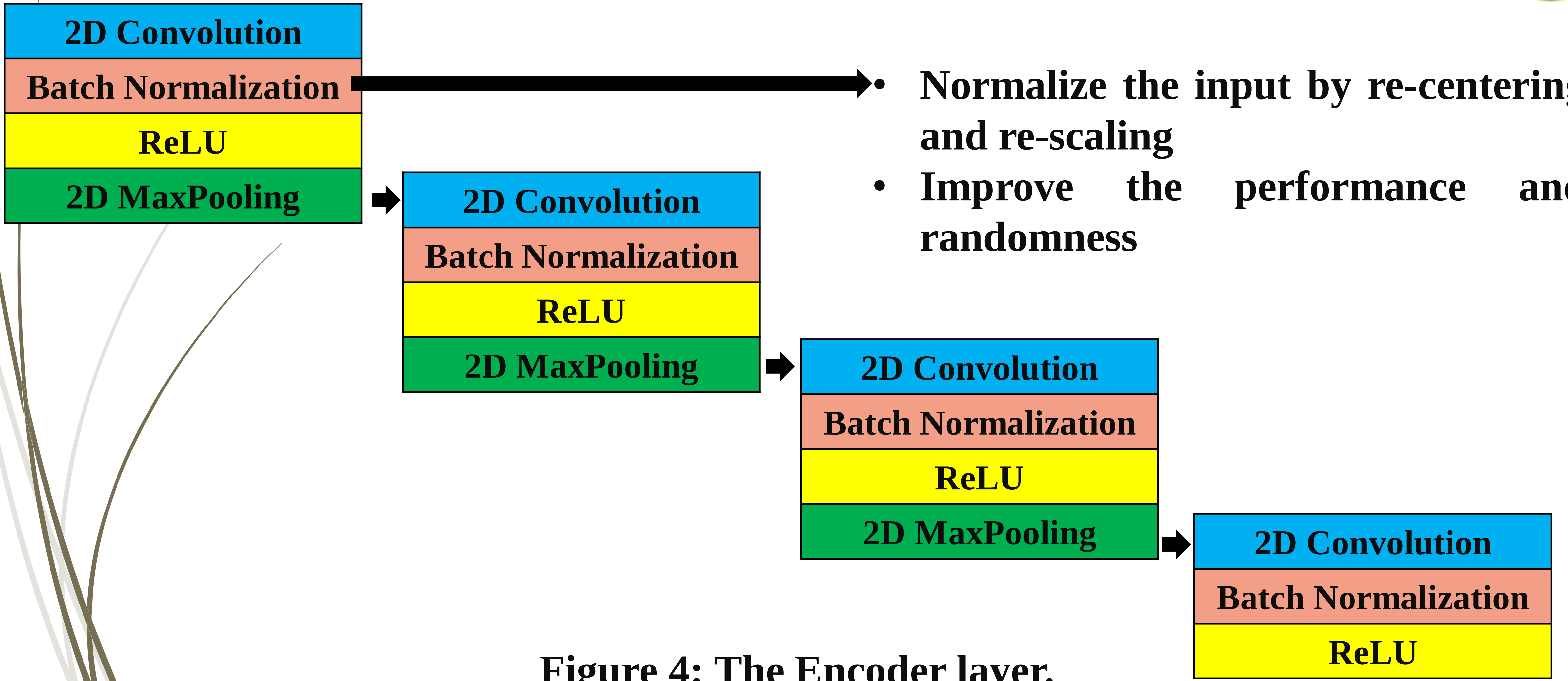


Figure 3: The Encoder layer.

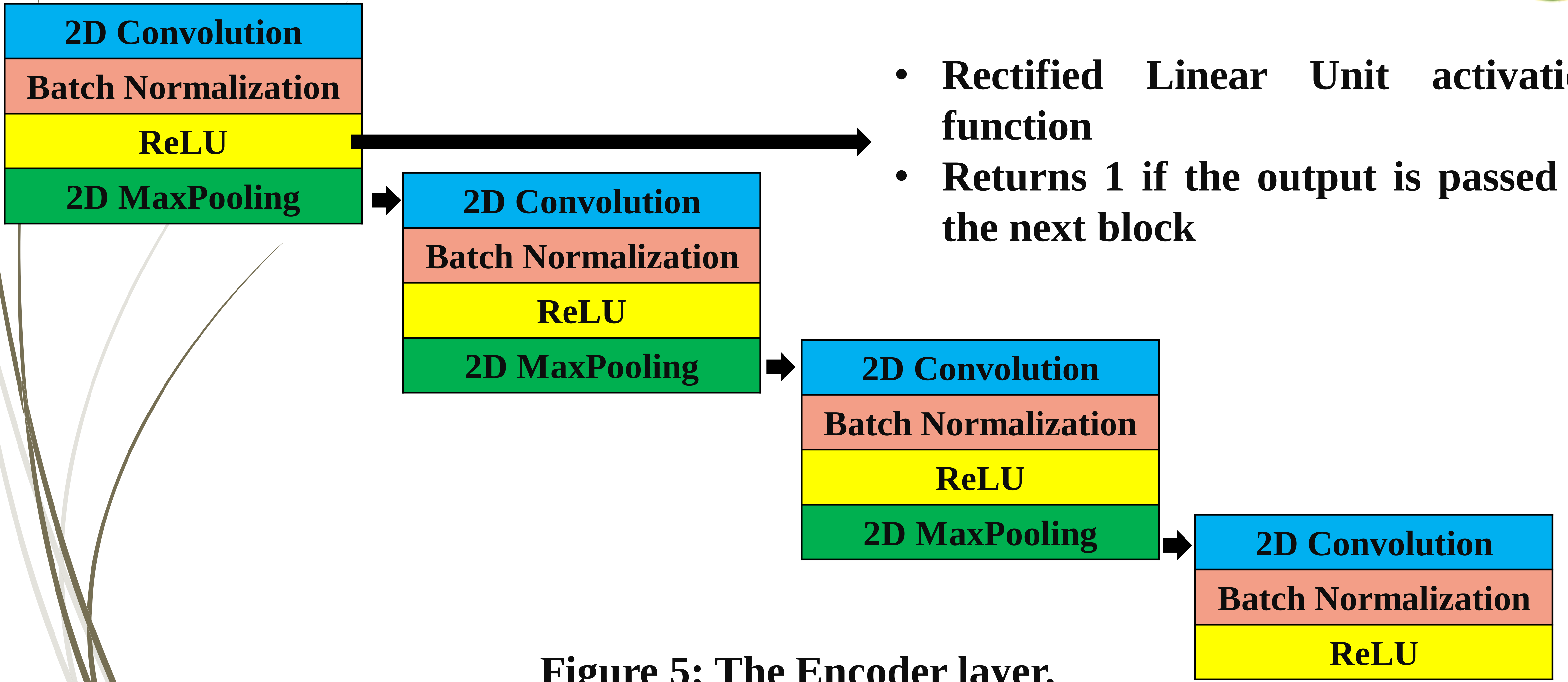
Encoder Layer



- **Normalize the input by re-centering and re-scaling**
- **Improve the performance and randomness**

Figure 4: The Encoder layer.

Encoder Layer



- Rectified Linear Unit activation function
- Returns 1 if the output is passed to the next block

Figure 5: The Encoder layer.

2D Convolution

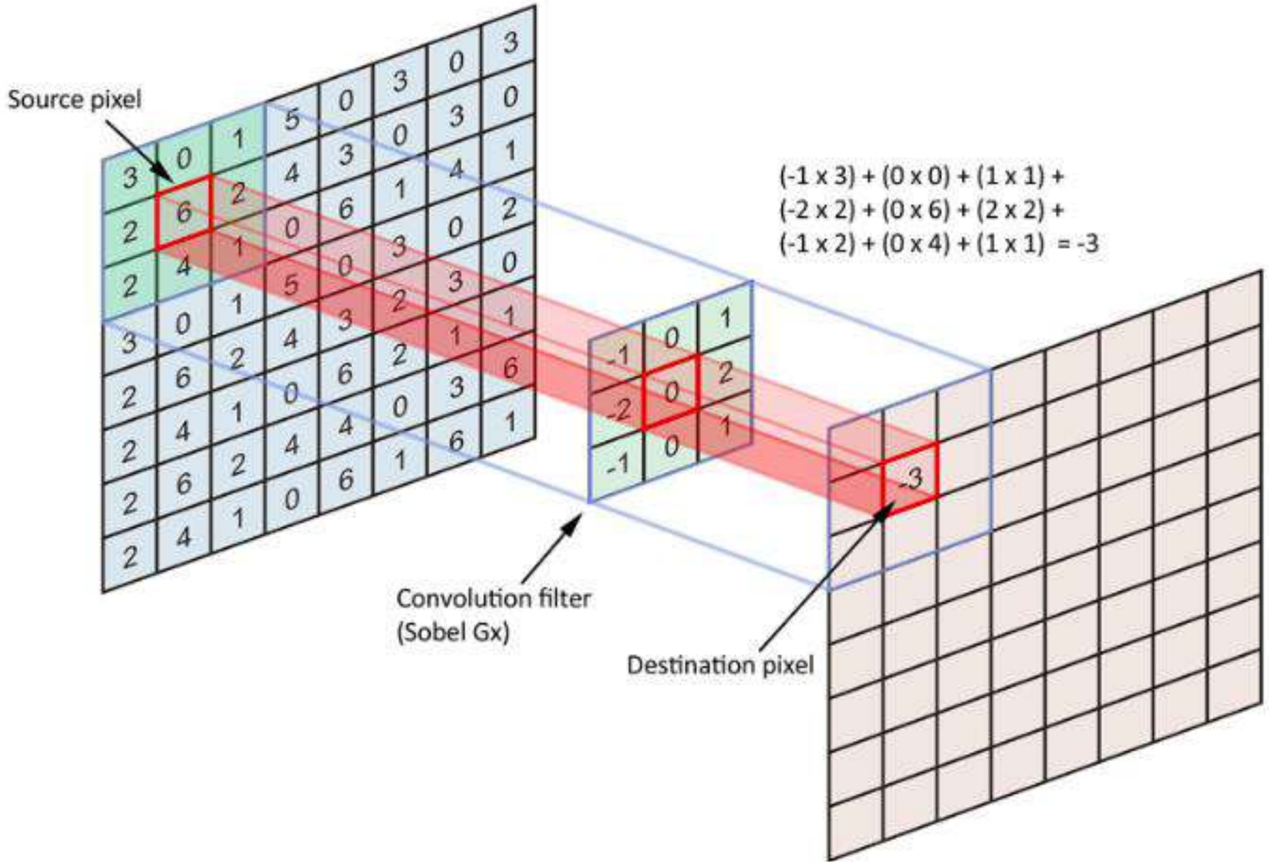
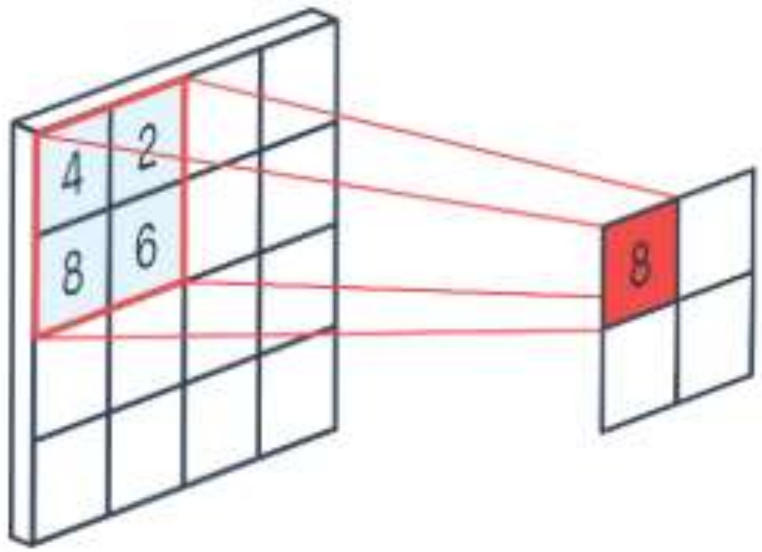


Figure 6: A 2D convolution operation [5].

2D MaxPooling



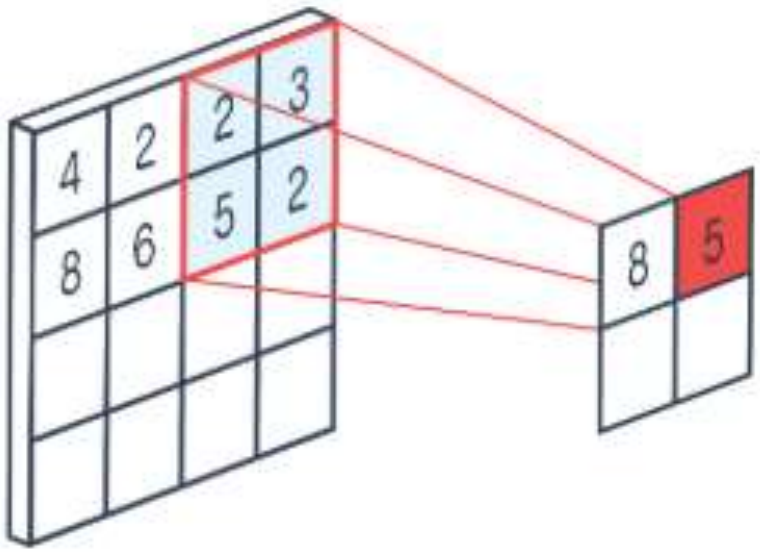


Figure 7: 2D MaxPooling [6].

Decoder Layer

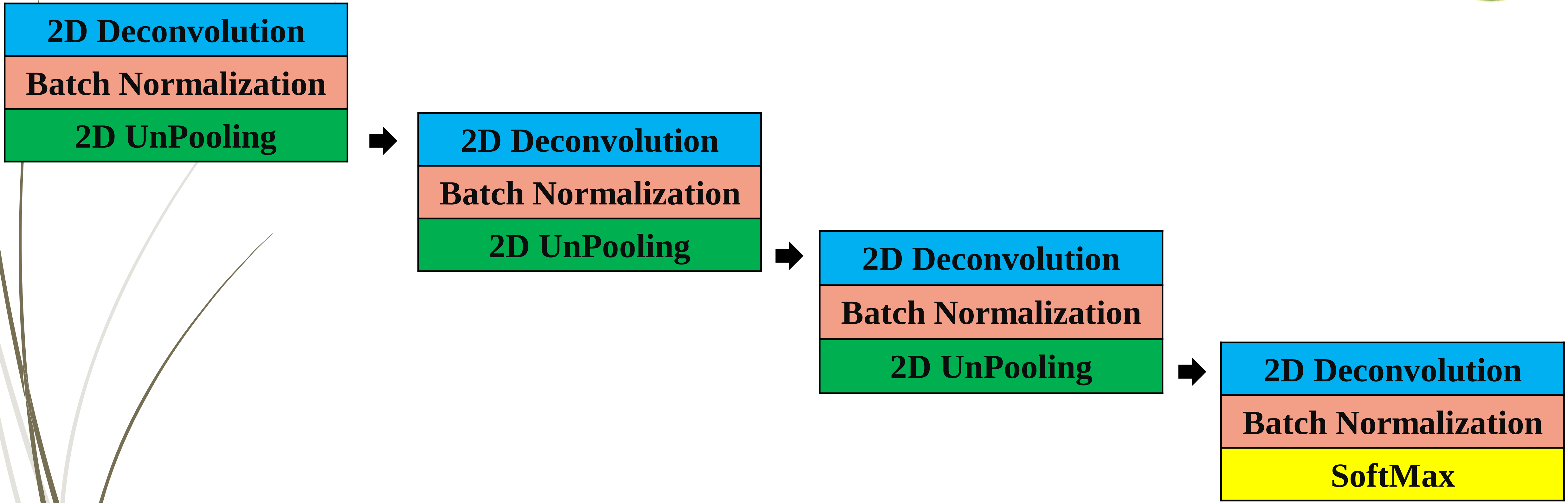


Figure 8: The Decoder layer.

Decoder Layer

- Exactly the opposite of 2D Convolution operation

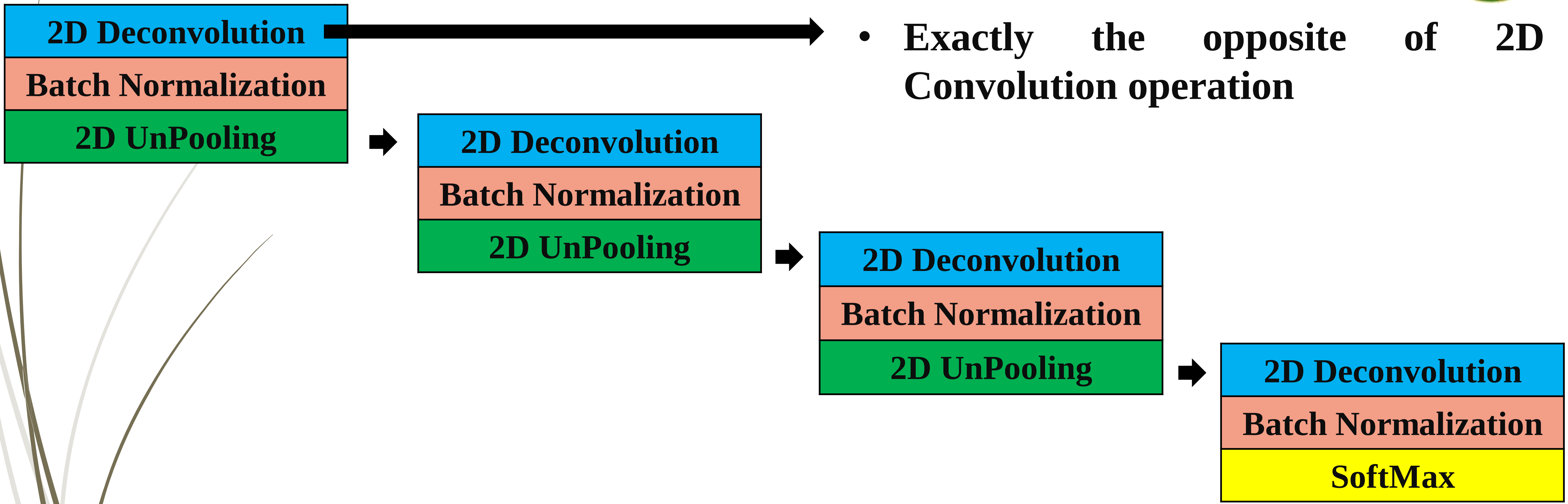


Figure 9: The Decoder layer.

Decoder Layer

2D Deconvolution
Batch Normalization
2D UnPooling

2D Deconvolution
Batch Normalization
2D UnPooling

2D Deconvolution
Batch Normalization
2D UnPooling

2D Deconvolution
Batch Normalization
SoftMax

- Exactly the opposite of 2D MaxPooling
- Remaining elements being set to zero

Figure 10: The Decoder layer.

Decoder Layer

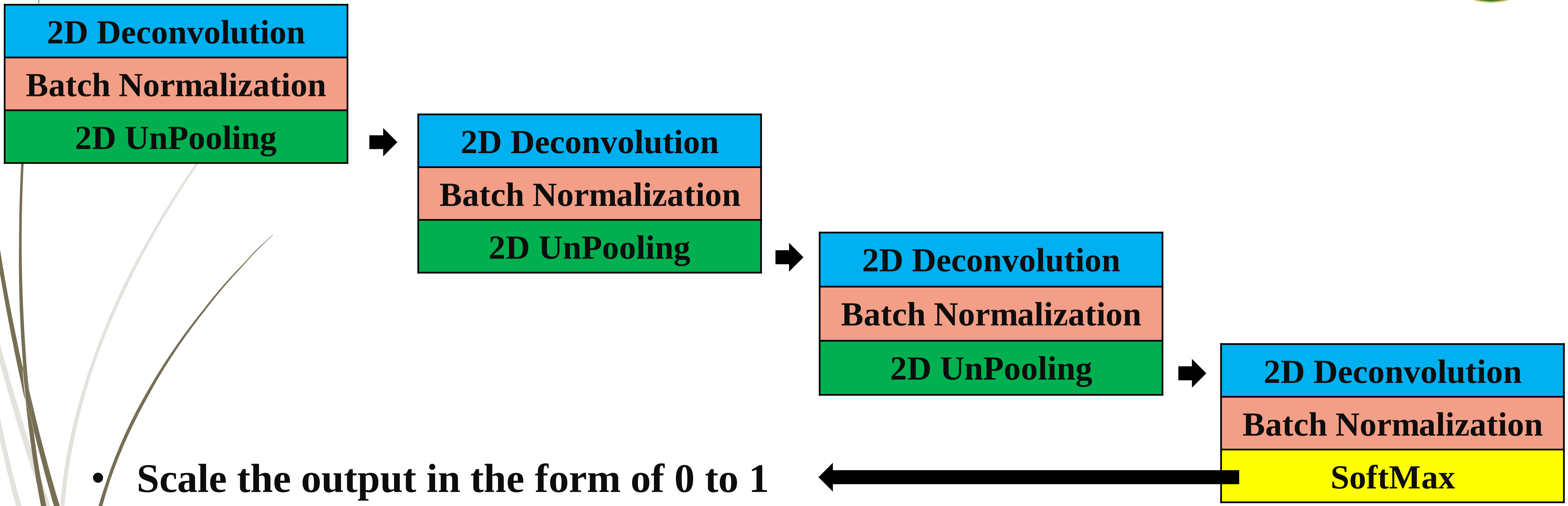


Figure 11: The Decoder layer.

Dataset

Figure 12: Some random image from the CamVid Dataset [2].

Result

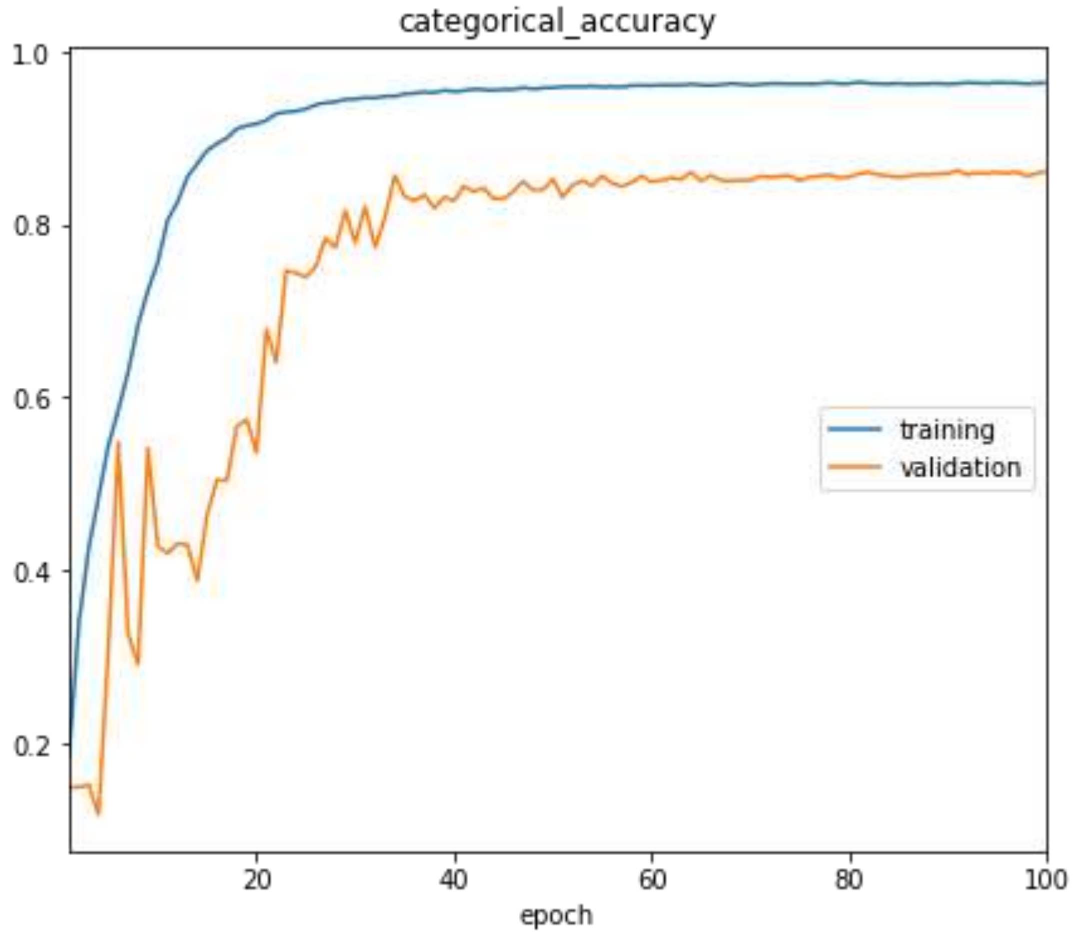
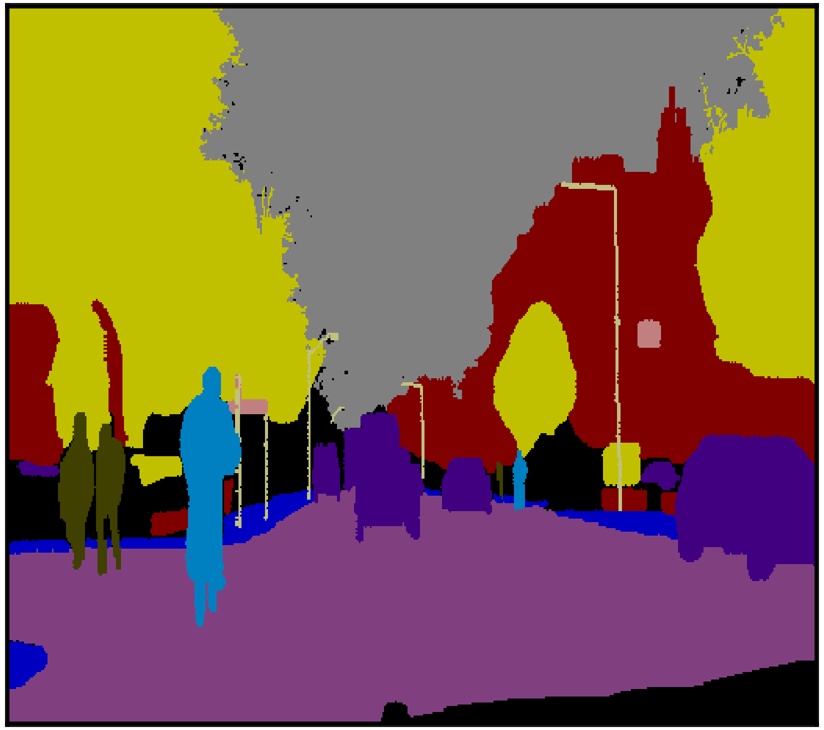


Figure 13: Categorical accuracy per epoch of the training and validation data.

Result

(a) Input Image



(b) Visualization of feature map

Figure 14: Visualizing the results of ArNet architecture.

Comparison

Table 1: Validation accuracy comparison between some the state-of-the-art architecture and ArNet.

Architecture	Validation Accuracy	Time Per Epoch (avg)
FCN8 [13]	62.5%	12 minutes
SegNet [18]	88.0%	1.6 hours
FC-DenseNet100 [19]	91.2%	13 hours (only one epoch was done completely)
ArNet	88.7%	32 minutes

Limitations

- Machine oriented architecture
- Very time consuming
  - 32 minutes per epoch on average
- Low number of dataset

Conclusion

- Improve machine visions
- Improve robotics
- Improve autonomous driving technology

Future Work

- Lower number of blocks
- Lower processing time
- More dataset
- Interface canny edge detector

References

1. A. A. Novikov, D. Lenis, D. Major, J. Hladůvka, M. Wimmer, K. Bühler, "Fully Convolutional Architectures for Multi-Class Segmentation in Chest Radiographs", Computer Vision and Pattern Recognition, (Submitted on 30 Jan 2017 (v1), last revised 13 Feb 2018 (this version, v4)).
2. V. Badrinarayanan, A. Kendall, R. Cipolla, Senior Member, IEEE, "*SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation* " [2017 TPAMI] h ttps://arxiv.org/abs/1511.00561.
3. M. Mostajabi, P. Yadollahpour, G. Shakhnarovich, "*Feedforward semantic segmentation with zoom-out features* ", Computer Vision and Pattern Recognition, (Submitted on 2 Dec 2014). arXiv:1412.0774.
4. J. Dai, K. He, J. Sun, "*Convolutional Feature Masking for Joint Object and Stuff Segmentation*" The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 3992-4000

References

5. D. Torres, Alejandro (Nov 2, 2010). "*Origin and history of convolution*". 41 pgs. http://www.slideshare.net/Alexdfar/origin-adn-history-of-convolution. Cranfield, Bedford MK43 OAL, UK. Retrieved March 13, 2013.

6. Q. Zhao, S. Lyu, B. Zhang, and W. Feng, "*Multiactivation Pooling Method in Convolutional Neural Networks for Image Recognition*", Big IoT Data Analytics in Fog Computing, Volume- 2018, Article ID- 8196906, https://doi.org/10.1155/2018/8196906.

Thank you…..